

YOU ARE WHERE YOU EAT FROM: PREDICTING SOCIOECONOMIC STATUS FROM
RESTAURANT MENUS AND USER REVIEWS

A THESIS SUBMITTED TO
THE GRADUATE SCHOOL OF INFORMATICS OF
THE MIDDLE EAST TECHNICAL UNIVERSITY
BY

ÖZGÜN OZAN KILIÇ

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF
MASTER OF SCIENCE
IN
THE DEPARTMENT OF INFORMATION SYSTEMS

SEPTEMBER 2020

**YOU ARE WHERE YOU EAT FROM: PREDICTING SOCIOECONOMIC STATUS FROM
RESTAURANT MENUS AND USER REVIEWS**

submitted by **ÖZGÜN OZAN KILIÇ** in partial fulfillment of the requirements for the degree of **Master
of Science in Information Systems Department, Middle East Technical University** by,

Prof. Dr. Deniz Zeyrek Bozşahin
Dean, **Graduate School of Informatics**

Prof. Dr. Sevgi Özkan Yıldırım
Head of Department, **Information Systems**

Assoc. Prof. Dr. Tuğba Taşkaya Temizel
Supervisor, **Information Systems, Middle East Technical University**

Examining Committee Members:

Assoc. Prof. Dr. Altan Koçyiğit
Information Systems, Middle East Technical University

Assoc. Prof. Dr. Tuğba Taşkaya Temizel
Information Systems, Middle East Technical University

Assoc. Prof. Dr. Aysu Betin Can
Information Systems, Middle East Technical University

Assist. Prof. Dr. Bilgin Avenoğlu
Software Engineering, TED University

Assist. Prof. Dr. Engin Demir
Computer Engineering, Hacettepe University

Date: 22.09.2020



I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Name, Surname: Özgün Ozan Kılıç

Signature :

ABSTRACT

YOU ARE WHERE YOU EAT FROM: PREDICTING SOCIOECONOMIC STATUS FROM RESTAURANT MENUS AND USER REVIEWS

Kılıç, Özgün Ozan

M.S., Department of Information Systems

Supervisor: Assoc. Prof. Dr. Tuğba Taşkaya Temizel

September 2020, 91 pages

Our culinary habits and traditions have always been related to our social norms, culture, and even personal details such as personality. In this study, restaurant data related to pita and pizza restaurants from Ankara found on the leading online food ordering system in Turkey, Yemeksepeti, are collected, enriched with other data, and analyzed. Cuisine-specific lexicons are created to vectorize restaurant menus; biterm topic modeling is used to model user reviews; various education rates or realty listing prices are used as a proxy for socioeconomic status. Statistical tests and random forest models are used to see the relationship between restaurants and their cuisine or location characteristics. The findings suggest that restaurant data can be used to some extent to make predictions about the district and the socioeconomic status of the district or the neighborhood. They also reveal that menu and statistical features are more distinguishing for pizza restaurants while comment and statistical features are more distinguishing for pita restaurants. It is found that pita restaurants are menu-wise less diverse and more similar to each other, neighborhood-wise more diverse, and they are more about getting the monetary value while pizza restaurants care more about menu curation. While side dish preferences and satisfactions for pita restaurants may be tied to the socioeconomic status, pizza restaurants respond more to user reviews when they are located in a higher status neighborhood. This study also provides some implications and further questions for future studies.

Keywords: Data Mining, Online Food Delivery System, Restaurant Menu, Food Ingredient, User Review, Topic Modeling

ÖZ

NEREDEN YİYORSAN OSUN: RESTORAN MENÜSÜ VE KULLANICI YORUMLARINDAN SOSYOEKONOMİK STATÜ TAHMİNİ

Kılıç, Özgün Ozan

Yüksek Lisans, Bilişim Sistemleri Bölümü

Tez Yöneticisi: Doç. Dr. Tuğba Taşkaya Temizel

Eylül 2020, 91 sayfa

Gastronomik alışkanlıklarımız ve geleneklerimiz her zaman sosyal normlarla, kültürle ve hatta kişilik gibi kişisel detaylarla alakalı olmuştur. Bu tez çalışmasında ülkenin lider çevrimiçi yemek sipariş sistemi olan Yemeksepeti üzerinden, Ankara konumlu pide ve pizza restoranlarının bilgileri toplanmış, başka verilerle zenginleştirilmiş ve analiz edilmiştir. Restoran menülerini vektörleştirmek için mutfaka özel sözlükler oluşturulmuş, kullanıcı yorumlarını modellemek için ikili kelime gruplarına dayanan konu modelleme kullanılmış ve sosyoekonomik statü göstergesi olarak çeşitli eğitim yüzdeleri veya emlak ilan fiyatları kullanılmıştır. Restoranlarla ait oldukları mutfak veya konumsal özellikleri arasındaki ilişkiyi görmek için istatistiksel testler ve rassal orman modelleri kullanılmıştır. Bulgular ilçe ve ilçe veya mahallenin sosyoekonomik statüsünü tahmin etmede restoran verisi kullanmanın bir dereceye kadar yararlı olduğunu göstermektedir. Pizza restoranları için menü ve istatistiksel özellikleri daha ayırt edici iken pide restoranları için yorum ve istatistiksel özelliklerin daha ayırt edici olduğu görülmüştür. Pide restoranlarının menü açısından daha az çeşitli olup daha çok birbirlerine benzediği, daha çeşitli tipte mahallelerde bulunduğu ve daha çok fiyat-performans odaklı bir anlayışa sahip olduğu bulunurken pizza restoranlarının menü sunumuna daha dikkat ettiği görülmüştür. Pide restoranlarında meze tercihi ve memnuniyetinin sosyoekonomik statü ile bağlantılı olabileceği görülürken pizza restoranlarının ise daha yüksek statülü mahallelerde bulunmaları durumunda kullanıcı yorumlarına daha sık cevap yazdığı görülmüştür. Bu çalışma ayrıca gelecek çalışmalar için de birtakım çıkarımlar ve cevaplanmak üzere daha fazla soru sunmaktadır.

Anahtar Kelimeler: Veri Madenciliği, Çevrimiçi Yemek Teslimi Sistemi, Restoran Menüsü, Yemek Malzemesi, Kullanıcı Yorumu, Konu Modelleme



To second chances

ACKNOWLEDGMENTS

First and foremost, I would like to thank my supervisor Assoc. Prof. Tuğba Taşkaya Temizel for her valuable time, wisdom, guidance, support, patience, and generosity. Her efforts have been crucial not only in my thesis research but also in my academic journey. I am immensely grateful that I was a part of her research team that pushed me to learn more and go beyond. I am grateful to Prof. Oğuz Işık for sharing his knowledge along with geographic and population-based datasets. I am grateful to everyone who taught me something or influenced me, especially professors in the Informatics Institute at Middle East Technical University, that lead me to this position. I am thankful to the examining committee members for their valuable feedback and suggestions.

I would like to acknowledge the support of my colleagues. I am thankful to Mehmet Ali Akyol for providing his realty listings dataset. I am also thankful to Ece Işık Polat and Yücelen Bahadır Yandık for helping me with manual labeling.

I would like to thank Chef Osman Özdemir for his occasional insights on ingredients and the food service sector in Turkey.

I am grateful to Yemeksepeti for providing such a useful platform and supporting my research, albeit unknowingly.

Finally, I would like to thank my family and friends for their support.

TABLE OF CONTENTS

ABSTRACT.....	iv
ÖZ.....	v
DEDICATION.....	vi
ACKNOWLEDGMENTS	vii
TABLE OF CONTENTS	viii
LIST OF TABLES	xi
LIST OF FIGURES	xiv
LIST OF ACRONYMS AND ABBREVIATIONS	xvi
CHAPTERS	
1 INTRODUCTION	1
1.1 Research Questions	2
1.2 Contributions of the Study.....	3
1.3 Organization of the Thesis.....	3
2 RELATED WORK	5
2.1 Restaurant Analysis	5
2.2 Ingredient and Dietary Habit Analysis	5
2.3 User Review Analysis	7

2.4	The Gap	8
3	METHODOLOGY AND RESULTS	9
3.1	Methodology	9
3.1.1	Data Source of Choice	9
3.1.2	Datasets	11
3.1.2.1	Primary	11
3.1.2.2	Secondary	11
3.1.3	Data Collection	12
3.1.4	Pre-processing	13
3.1.4.1	Word Labeling and Menu Vectorization	13
3.1.4.2	Comment Vectorization	15
3.2	Data Analysis	17
3.2.1	Descriptive Statistics	17
3.2.2	Menu Statistics and Socioeconomic Status	20
3.2.3	District Prediction with Menu Contents and Restaurant/Menu Statistics	21
3.2.4	Socioeconomic Status Prediction with User Reviews, Menu Contents, and Menu/Restaurant Statistics	26
4	DISCUSSION	43
4.1	Methodologies and Results	43
4.1.1	Menu Statistics and Socioeconomic Status	43
4.1.2	District Prediction	43
4.1.3	Socioeconomic Prediction	44
4.1.4	Pita versus Pizza	47

4.1.5	Predicting Place versus Attributes	48
4.1.6	Related Work	48
4.2	Implications of Use	49
5	CONCLUSION	51
5.1	Limitations and Assumptions	52
5.1.1	Data Collection	52
5.1.2	Data Analysis and Predictions	53
5.2	Future Work	55
	REFERENCES	57
	APPENDICES	
A	MENU LEXICONS AND ANNOTATIONS	63
A.1	Ingredient Lexicon	63
A.2	Side-dish Lexicon	65
A.3	Annotator Votes	66
B	EXTRA MATERIAL	70
C	COMMENT TOPICS	81

LIST OF TABLES

Table 1	Comparison of data sources	10
Table 2	Ankara's districts along with their number of restaurants that serve pizza or pita and their populations and university graduate rates	17
Table 3	Restaurants that serve pizza or pita along with their primary categories (cuisines)....	18
Table 4	Food-specific statistics for pizza-serving and pita-serving restaurants	19
Table 5	The districts used by the models during the classification along with their member sizes	22
Table 6	Classification scores for pizza and pita restaurants using different types of features ..	23
Table 7	Feature importances for pizza and pita restaurants, ranked by permutation-based feature importances	24
Table 8	Confusion matrix for pizza restaurants using the best model.....	25
Table 9	Confusion matrix for pita restaurants using the best model	25
Table 10	Model results for pizza restaurants with different feature sets	30
Table 11	Model results for pita restaurants with different feature sets	31
Table 12	Feature importances for square meter price prediction with pizza restaurants.....	32
Table 13	Feature importances for square meter price prediction with pita restaurants	33
Table 14	Point-biserial correlations between menu contents and square meter prices for pizza restaurants (N=67).	38
Table 15	Spearman correlations between education rates and square meter prices for pizza restaurants (N=67). Education rates correspond to the highest education obtained. See Table 23 for more information.	39
Table 16	Spearman correlations between education rates and square meter prices for pita restaurants (N=143). Education rates correspond to the highest education obtained. See Table 23 for more information.	39
Table 17	Point-biserial correlations between one-hot encoded binary district features and square meter prices for pizza restaurants (N=67).	39

Table 18	Point-biserial correlations between one-hot encoded binary district features and square meter prices for pizza restaurants (N=143).	40
Table 19	District clusters centers with their scaled education rates	40
Table 20	Pita restaurants' comment topic likelihoods that significantly differ between two education clusters, the Kruskal-Wallis test results with Bonferroni-corrected p-values	41
Table 21	Pizza and pita restaurants' menu term existences compared with two education clusters using Fisher's exact test. P-values are Bonferroni-corrected. An odds ratio higher than 1 indicates that the item is more common for the less educated districts' cluster while an odds ratio lower than 1 indicates that the item is more common for the more educated districts' cluster.	42
Table 22	Annotator votes for each lexicon item	66
Table 22	(continued)	67
Table 22	(continued)	68
Table 22	(continued)	69
Table 23	Ankara's education level percentages by district, rounded (Everyone is represented only under the highest education they have)	72
Table 24	Pizza restaurants' comment topic likelihoods compared with two education clusters using the Kruskal-Wallis test. P-values are Bonferroni-corrected.	74
Table 25	Pita restaurants' comment topic likelihoods compared with two education clusters using the Kruskal-Wallis test. P-values are Bonferroni-corrected.	75
Table 26	Pizza restaurants' menu term existences compared with two education clusters using Fisher's exact test. P-values are Bonferroni-corrected. An odds ratio higher than 1 indicates that the item is more common for the less educated districts' cluster while an odds ratio lower than 1 indicates that the item is more common for the more educated districts' cluster.	76
Table 26	(continued)	77
Table 26	(continued)	78
Table 27	Pita restaurants' menu term existences compared with two education clusters using Fisher's exact test. P-values are Bonferroni-corrected. An odds ratio higher than 1 indicates that the item is more common for the less educated districts' cluster while an odds ratio lower than 1 indicates that the item is more common for the more educated districts' cluster.	79
Table 27	(continued)	80
Table 28	Top 10 most relevant terms for each topic obtained from pizza restaurant reviews	82
Table 28	(continued)	83
Table 28	(continued)	84

Table 28 (continued)	85
Table 28 (continued)	86
Table 29 Top 10 most relevant terms for each topic obtained from pita restaurant reviews	87
Table 29 (continued)	88
Table 29 (continued)	89
Table 29 (continued)	90
Table 29 (continued)	91



LIST OF FIGURES

Figure 1	Top 50 most frequent term co-occurrences from pizza restaurants' user reviews. A thicker edge indicates higher frequency.	16
Figure 2	Top 50 most frequent term co-occurrences from pita restaurants' user reviews. A thicker edge indicates higher frequency.	16
Figure 3	Menu descriptive statistics with means shown as dashed vertical lines	20
Figure 4	District and square meter price distributions of pizza restaurants	27
Figure 5	District and square meter price distributions of pita restaurants	27
Figure 6	District and realty listings distributions of pizza restaurants	28
Figure 7	District and realty listings distributions of pita restaurants	28
Figure 8	Scatter plots of response percentages and square meter prices for each district ...	30
Figure 9	Top 10 most relevant terms of Topic 39 (from pizza restaurant reviews)	34
Figure 10	Top 10 most relevant terms of Topic 9 (from pizza restaurant reviews)	35
Figure 11	Top 10 most relevant terms of Topic 15 (from pizza restaurant reviews)	35
Figure 12	Top 10 most relevant terms of Topic 14 (from pita restaurant reviews)	36
Figure 13	Top 10 most relevant terms of Topic 70 (from pita restaurant reviews)	36
Figure 14	Top 10 most relevant terms of Topic 88 (from pita restaurant reviews)	37
Figure 15	Average silhouette width with different numbers of clusters	40
Figure 16	Districts clustered by their education rates	41
Figure 17	Top 10 most relevant terms of Topic 2 (from pita restaurant reviews)	42
Figure 18	Keçiören's (with a red border) population is denser and restaurants are fewer compared Çankaya (with a green border)	45
Figure 19	Districts of Ankara	70
Figure 20	Locations of restaurants that serve pizza, pita, or both on map along with the district borders	71

Figure 21	A screenshot of the custom web-based tool used to label lexicon items.	73
-----------	--	----



LIST OF ACRONYMS AND ABBREVIATIONS

BTM	Biterm Topic Modeling
CD	Comment dataset (see Section 3.1.2 for more information)
ED	Education dataset (see Section 3.1.2 for more information)
LDA	Latent Dirichlet Allocation
MD	Menu dataset (see Section 3.1.2 for more information)
ML	Menu lexicon dataset (see Section 3.1.2 for more information)
NLP	Natural language processing
PD	Population dataset (see Section 3.1.2 for more information)
RE	Real estate listings dataset (see Section 3.1.2 for more information)
RC	Restaurant coordinates dataset (see Section 3.1.2 for more information)
RD	Restaurant dataset (see Section 3.1.2 for more information)

CHAPTER 1

INTRODUCTION

Food has always had an importance beyond mere survival for the humankind. In some cultures, human and animal parts are ritually consumed to harness their strength [1]. In Asia, shark fin is consumed due to its rareness and luxury rather than its flavor [2]. Trying new tastes is a part of tourism; in fact, there is a phenomenon called "culinary tourism," which puts the focus on food in tourism [3].

Therefore, food is an important part of our social norms, culture, and identity. So much so that it is possible to gather a great deal of information about someone from their dietary habits, perhaps more than one would initially believe. According to studies, dietary habits do not only reflect religion, nationality, or class, but also personality and ethics [4–7].

Online food ordering is on the rise due to many reasons, such as convenience, societal pressure, and ease of use [8, 9]. They allow users to browse foods, make an order, make a payment, and rate the service from possibly many different restaurants without ever leaving a single website or application. Not only do they facilitate the whole food delivery process, but also they form digital data repositories with a good amount of quantitative and qualitative data that are open to the public. This thesis aims to analyze similarities and differences of districts of Ankara (shown in Figure 19), the capital city of Turkey, in relation to cuisines, restaurant menus, and restaurant reviews on Yemeksepeti.com, in a socioeconomic context.

Yemeksepeti.com¹ [10] is the leading online food ordering system in Turkey. It enables people to reach hundreds of different restaurants from a single application. A user can add an address, browse the restaurants that deliver to their location, order a meal from restaurants' menus, optionally pay for it online through Yemeksepeti, and later review their order (separately for speed, service, and taste) without ever leaving the web/mobile application. It expanded to other countries as well, and it was bought by Delivery Hero, an international competitor, in 2015 [11]. While it is not as popular, it also provides courier service for restaurants that do not have their own couriers.

By the end of 2019, Yemeksepeti had 36 thousand partner restaurants and 14 million total users while getting 340 million portion orders in the last twelve months [12, 13]. At that time, in the United Kingdom, a country that has a mature online shopping market, the market leader Just Eat had 35.7 thousand partner restaurants while getting 131 million orders in the last twelve months [14]. Considering Turkey's 62.07 million Internet users (indicating 74% of the whole population) and the UK's 65 million Internet users (indicating 96% of the whole population) at the beginning of 2020 [15], it can

¹ "Yemek sepeti" literally means "food basket" in Turkish. While the company's official name is "Yemek Sepeti," due to them referring to themselves as "Yemeksepeti" on most platforms, they will be referred as "Yemek-sepeti" from now on.

be argued that Yemeksepeti and online food delivery have a considerable penetration in Turkey. A study that compares Turkey and the United Kingdom also suggests that there is not a significant difference between British users' and Turkish users' online food purchase behaviors [16]. According to a consumer research, Turkey also leads customer retention rates in online food delivery platforms [8].

This study is restricted to pizza and pita restaurants. The reason only two main types of restaurants are chosen is comparing pizza/pita restaurants with other pizza/pita restaurants frees one from the additional noise of analyzing different types of restaurants and makes it possible to focus on differences between locations. On the other hand, choosing two types of restaurants instead of one makes it possible to compare two different cuisines with each other and understand how these two cuisines differ from each other in terms of variety and their customers. Pizza and pita restaurants are deliberately chosen. While pita in Turkish cuisine has a similar cooking method (basically placing various ingredients on a flat dough and oven baking it) to pizza and they are both fairly popular in Turkey, they also have contrasting origins. Pita comes from Middle Eastern cultures while pizza comes from Italy, which may be helpful to find interesting differences while minimizing method or ingredient differences.

1.1 Research Questions

The thesis aims to answer these questions:

RQ1 Is there a relationship between a restaurant's menu curation characteristics and its district's socioeconomic status²?

It is hypothesized that districts with higher socioeconomic status might have more diverse and elaborate menus, reflecting the districts' residents and their varying and higher expectations. As potential indicators of diversity and elaboration, statistics such as number of items, photograph coverage, description lengths, and number of ingredients will be analyzed.

RQ2 Is it possible to predict a restaurant's district by looking at its menu contents and/or statistics?

It is hypothesized that, due to cultural or economic reasons, restaurants in certain districts might typically have certain ingredients or side dishes that may distinguish them from other districts. To explore this, restaurant menus will be analyzed.

RQ3 Is it possible to predict a restaurant's surrounding area's socioeconomic status based on its user reviews?

It is hypothesized that people in different districts, in accord with their district demographics, might have different motivations, expectations, language characteristics, and reviewing behav-

² Socioeconomic status is a generic metric to group individuals, households, or locations according to their social and economic functions in the society. Its exact definition and formula vary. In [17], its factors are defined as "cultural possessions, effective income, material possessions, and participation in group activity in the community." In [18], a household-oriented approach is used to measure the status of students by stating that the main factors are family income, parents' educational levels, and parents' occupations while housing quality, family structure, and population density are some other indicators. According to an official report published by Kalkınma Ajansları Genel Müdürlüğü (General Directorate of Development Agencies) in Turkey, a socioeconomic development index of Turkey's cities is created using 52 variables under the categories of demographics, employment, education, healthcare, competitiveness and innovativeness, economy, accessibility, and life quality [19]. This study does not use any particular socioeconomic status metric and assumes that education and rental prices can be used as a proxy in a general sense, based on existing research [20–23].

iors. To find an answer to this question, restaurant reviews will be analyzed along with other data.

1.2 Contributions of the Study

The contributions of the study are as follows:

- Presenting example approaches on enriching data from an online food ordering system using external data
- Exploring how online food ordering systems, restaurants, and their customers can be used to make geolocational and socioeconomic predictions
- Exploring how restaurants, their menus, and their reviews differ in accordance with cuisine, location, or socioeconomics
- Analyzing/reviewing restaurants, their menus, and their reviews together to understand how these different factors affect online food delivery
- Providing an ingredient and side dish lexicon for Turkish pizza or pita restaurants/recipes

1.3 Organization of the Thesis

The organization of the thesis is as follows. Chapter 2 reviews the previous studies and explains how the thesis fits in the literature. Chapter 3 explains the data source of choice, used datasets, the methodology of collecting the data, and the methodology of certain analysis methods along with providing the results. Chapter 4 summarizes the results and provides further discussions. Chapter 5 concludes the thesis, mentions limitations, and points to directions for future work.

CHAPTER 2

RELATED WORK

In this chapter, related studies are given and later it is explained how this study fits in the literature. It is possible to group related work under three categories: restaurant analysis [24–27], ingredient and dietary habit analysis [4–6, 28–42], and user review analysis [43–59]. These categories are discussed in the following sections.

2.1 Restaurant Analysis

There are not many studies that focus on restaurant analysis, especially in a socioeconomic context. In [24], fast food restaurant distributions are analyzed along with census data. After controlling certain environmental factors such as highway density and median home values, it is found that low-income and black neighborhoods have more fast food restaurants.

There are studies that focus on Yemeksepeti. While [25] analyzes Yemeksepeti restaurants and their statistics such as service hours, restaurant reviews and responses, payment methods, cuisines, and average delivery times, these statistics are not grouped by cuisine or location, providing only an overview through sampling 1696 restaurants from Turkey. Another study on Yemeksepeti, [26] analyzes a global pizza chain's five branches in Eskişehir, Turkey, through 695 user reviews using ANOVA and Kruskal-Wallis tests. This study suggests that different branches of the same brand have significantly different scores for all categories (speed, service, and taste). It only considers user ratings and ignores user comments. Since it has a narrower focus, it also does not reflect on certain details such as why these locations differ from each other. In [27], it is stated that an online survey conducted with people who order food online shows that the participants' attitudes towards online food delivery differ according to their education, age, household size, and marital status. It is also stated that their attitudes do not significantly change with income, gender, and profession.

2.2 Ingredient and Dietary Habit Analysis

While there are studies that focus on restaurants' menu design, their concern is not how restaurant menus differ from each other but rather how restaurant menus can be optimized for profitability. For example, in [28], a quantitative method that ranks menu items based on data such as menu item costs, menu item prices, and menu item sales is proposed to find the most profitable menu items and potential changes that can increase the profitability. To the best of my knowledge, there are not any

studies that explore how competing restaurants' menus and ingredients differ from each other, especially in accordance with their locations and the corresponding socioeconomic statuses. However, there are studies that explore the relationship between menu item representations and their effects on the customers' perception. In [29], restaurant menu information's effects on users' attitudes are analyzed. It is found that what customers expect from restaurant menus differ based on the restaurants' price levels. For example, higher-end restaurants' customers expect menus to give more importance to nutritional values while lower-end restaurants' customers expect menus to give more importance to product characteristics such as quantity and brand. In [30], the survey results suggest menu item descriptions have an effect on perceptions of quality, price expectations, and purchase intentions. For example, more complex terminologies cause an increase in perceived quality and pricing expectations. It also mentions that Food Services of America suggests restaurants a selective approach to use more complex terminologies and photographs to make certain menu items more noticeable and attractive to customers.

While there are studies that analyze restaurants' menu item contents, they are mostly interested in nutritional contents and children's health. For example, in [31], only children's meals are collected and they are analyzed in a nutritional context. In [32], topic modeling is used with restaurant menus. However, it is aimed to make inter-cultural comparisons. For example, it is stated that certain ingredients such as soy sauce or sesame oil exclusively belong to certain Asian cuisines while Indian cuisine is remarkably different from the others. In [33], not restaurant menus but cooking recipes are collected along with several classifications for recipe recommendation. User reviews, ratings, and profiles are collected as well. Cooking methods and ingredients are extracted from recipes. While their focus is recipe recommendation, it is also found that different regions in the United States have significantly different cooking method preferences.

In [4], [5], and [6], through self-reported surveys, relationships between the Big Five personality traits [60, 61] and dietary habits are found. For example, healthier diets are linked to lower Neuroticism and higher Extraversion while more traditional diets are linked to less Openness. People who avoid meat describe themselves in ways that reflect Morality, Cooperativeness, Dutifulness, and Purposefulness. While [4] found no relationship between education level and dietary habits, [34] suggests that the educational system has an effect on students' personality traits. This indicates that the education level still may have a relationship with dietary habits through personality traits. In [35], it is suggested that, according to their survey responses, one of the factors of following a healthier diet is a high education level. Gender and living in multiperson households with children has a correlation with healthier diets as well. In [36], according to a consumer survey conducted in Turkey, having a work-life does not have a statistical effect on fast food consumption, but large households consume less fast food while households with children, high education, and high income consume more fast food. While it seemingly opposes [35] in terms of education levels and household characteristics, it may be caused by socioeconomic differences between the countries (Turkey and Denmark) in which these studies are conducted. However, [36] also points out there are other studies conducted in different countries that are in line with its findings [37–39]. In [40], survey respondents are categorized into seven different health lifestyle groups. It is found that there is a relationship between fast food consumption and being male, younger, black, or having a low-income. However, when the demographics are controlled, a significant difference among health lifestyle groups surprisingly could not be found about fast food consumption. In [41], a large scale analysis of American Twitter users, their network, and their tweets is done. Food mentions are detected and their calorie estimations are retrieved. Tweets and demographic data are mapped to zip codes. Overweightness is predicted from use of hashtags

and gender is predicted from usernames. It is found that there is a negative correlation between education and caloric values of tweets, and people who live in rural areas mention foods that have higher calories compared to people who live in urban areas. While urban people share more about alcohol and uncommon foods, rural people share more about pizza and desserts.

In [42], eating and drinking characteristics of Foursquare users around the world are analyzed through their check-ins in different types of restaurants. It is found that temporal patterns can be used to obtain clusters that are formed by countries that share certain cultural characteristics while not necessarily being neighbors. For example, although they belong to different regions, Turkey and Indonesia (sharing an Islamic culture and being labeled so) are clustered together.

2.3 User Review Analysis

There is a considerable amount of studies that focus on user reviews. In [43], Google reviews for local services (including but not limited to restaurants) are analyzed to build an aspect-based (based on categories such as service, value, etc.) summarization for those reviews. To do so, seed lexicon words are chosen, and the lexicon is extended through co-occurrences with those seed lexicon words. In [44], a method that extracts opinion sentences from product reviews and classifies those opinion sentence sentiments is proposed. This method focuses on adjectives and features that are defined by those adjectives. Similar to [43], a small set of seed words is used to create a sentiment lexicon. By finding synonyms and antonyms with the help of WordNet [45], this lexicon is extended and used to decide on feature and sentence sentiments.

Later in [46], it is stated that general-purpose lexicons such as SentiWordNet [47] are not good enough for restaurant reviews, and a custom lexicon is created through manually analyzing unigrams, bigrams, and sentiment patterns to classify restaurant review sentiments. In [48], Yelp restaurant reviews are analyzed, and a sentiment score for each term is calculated by using its point-wise mutual information value with positive and negative reviews in order to predict sentence sentiments. These restaurant reviews are also used to create a word-aspect association lexicon by finding terms that have moderate to high association with these aspects and similarly extending them later, which is used to calculate the likelihood of each word for five aspect categories (food, price, service, ambiance, and anecdotes). That way, reviews' aspect categories are predicted in a multi-class multi-label manner with Support Vector Machine classifiers.

There are also studies that prefer topic modeling approaches that commonly rely on Latent Dirichlet Allocation (LDA) [49], a hierarchical Bayesian model, or its variants/extensions. In [50], an aspect-based review summarization is applied using Multi-Grain Latent Dirichlet Allocation and a set of aspect-based sentiment predictors to force aspect-based cohesiveness of the topics. In [51] and [52], an LDA-based unsupervised (or rather weakly-supervised) model that detects both topic and sentiment from movie reviews using a word sentiment lexicon is proposed. To detect the sentiment as well, a sentiment layer is inserted between the document and topic layers. In [53], instead of detecting sentiment from reviews, a review rating layer is used. However, those individual review ratings are binarized to positive or negative to similarly indicate a review-level sentiment. In [54], user preferences and differences are also acknowledged and made a part of the model that works on user-review-item combinations. In [55] and [56], LDA models that make use of the domain knowledge is used. To do so, term co-occurrences that must/cannot exist in the same topic are provided to adjust topics.

In [57], it is stated that LDA does not work well with short texts such as tweets due to sparsity. To overcome this problem, Biterm Topic Modeling (BTM) is proposed. Instead of handling individual terms in a document (which can be the whole reviews or a sentence of it), BTM directly works with term co-occurrence patterns in documents within a given moving window range. BTM is used to model tweet topics, and it is found to capture more coherent topics compared to LDA. In [58], a Seeded Biterm Topic Model is used with product reviews. Similar to previous studies that use a seed lexicon, this study uses some domain-independent sentiment words such as "amazing" as seeds to eliminate unrelated biterns and discover aspects. Similar to some of the previous studies, this study also chooses to consider term co-occurrences in sentences instead of having a fixed window.

There are studies that apply deep learning to user reviews. In [59], a comparative review about using deep learning for aspect-based sentiment analysis is done. It states that while deep learning approaches have promising and comparable results in aspect extraction/categorization, approaches that combine aspect extraction and sentiment analysis tasks need improvement. Deep learning approaches are kept out of the scope of this study, and therefore not explored.

2.4 The Gap

It seems like there is not much focus in the literature on menu item and ingredient differences among restaurants. As mentioned in this chapter, there are various studies that come across the interests of this study. However, these interest points are scrutinized in different contexts and/or rather in isolation. This study aims to not only use restaurant menus and user reviews but also combine them and review them considering the socioeconomic context as well. It is believed that a greater deal of insight can be gained through analyzing restaurants, their menus, and customer reviews, and these insights can be used to explore socioeconomic and culinary reflections. While it is not the focus of this study, it also composes food-specific (pizza and pita) ingredient and side dish lexicons in Turkish, which could not be readily found in the literature.

CHAPTER 3

METHODOLOGY AND RESULTS

3.1 Methodology

In this chapter, datasets and the details of the methodologies used to collect and analyze data are explained.

3.1.1 Data Source of Choice

There are multiple ways through which one can collect data related to restaurants, their menus, and customer reviews. This study's choice of platform is Yemeksepeti to collect restaurant, menu, and review information. While some other websites provide information (such as exact location and reviewer details) that Yemeksepeti does not provide, certain aspects of it make it the obvious and best choice as the only popular online food ordering system. Listing a restaurant is not free on Yemeksepeti, and commissions are applied for each order. Therefore, Yemeksepeti has relatively higher entry barriers, which eliminates some of the noise (smaller restaurants with smaller resources). This may be seen as a potential risk, but Yemeksepeti has a national popularity and restaurant locations shown in Figure 20 demonstrate that Yemeksepeti has penetrated the whole urban areas in general in Ankara. Therefore, it is assumed that it would not cause a significant skewness in data. Since it is solely based on online orders and only people who recently made an online order can review a restaurant through their specific experience, reviews' credibility is relatively high. Moreover, since restaurants have limited delivery regions (meaning they mostly deliver to close locations), it is possible to make inferences about where the reviewers commonly reside (live/work/study), which enables cross-referencing using demographic data. Although Yemeksepeti only shows the comments that are written (and manually approved by Yemeksepeti) in the last six months, it is possible to find a couple hundred to a couple thousand comments for a restaurant. Other platforms typically have drastically fewer comments/reviews. Apart from being highly popular, Yemeksepeti has also implemented a gamification approach that encourages not only ordering more but also reviewing the restaurants more, which may have a role in the abundance of reviews. A major drawback is restaurant pages may not always be visible (due to working hours, crowdedness, etc.) even if one just wants to look at the menu, which requires web scraping in different days and/or times to catch the previously-unavailable information. A feature-wise comparison of Yemeksepeti along with other similar platforms at the moment is shown in Table 1. It should be noted that some platforms do not have the same features for every region and country. For example, Zomato does not have the online ordering feature and Google does not list menu items in Turkey as it started doing it in US.

Table 1: Comparison of data sources

Platform	Yemeksepeti	Google	Yelp	TripAdvisor	Zomato
Restaurant listing cost	Monthly fee	Free	Freemium	Freemium	Free
Restaurant categories	Detailed	Limited	Highly detailed (manually browsing them is problematic)	Highly detailed	Highly detailed
Directly accessible menu	Mandatory	Menu images can be separately uploaded	No dedicated space	No dedicated space	Menu images can be separately uploaded
Online order	Yes	No	No	No	No (for Turkey)
Open hours	Yes	Yes	Yes	Yes	Yes
Temporal crowdedness indicator	No (Average delivery time might be a substitute for the current crowdedness)	Yes (relative crowdedness graphs for days and hours)	No	No	No
Rating scale	Separate 10-point ratings for speed, service, and taste	5-point ratings	5-point ratings	Separate 5-point ratings for overall, food, service, value, and atmosphere	5-point (it is possible to give points with 0.5 increments)
Reviewed Reviewers	Each online order Members who ordered food recently	Restaurant Members	Restaurant Members	Restaurant Members	Restaurant Members
Reviewers	Not uniquely identifiable	Uniquely identifiable	Uniquely identifiable	Uniquely identifiable	Uniquely identifiable
Review length	300 characters at max	Can be long	Can be long	Can be long	Can be long
Photos with reviews	Not possible	Possible	Possible	Possible	Possible
Prices	Prices for the whole menu	Relative, overall	Relative, overall, price range	Relative, overall	Average meal prices
Address	District and/or neighborhood	Full address with pinned location on map	Full address with pinned location on map	Full address with pinned location on map	Full address with pinned location on map
Review count	Many (Only the ones written in the last six months are readable)	Few	Few	Few	Few
Data availability	Not always available (based on the restaurant's status)	Always available	Always available	Always available	Always available

3.1.2 Datasets

Various datasets are collected/created (primary ones) or acquired and appropriated (secondary ones) to be used. These datasets along with their short names are as follows:

3.1.2.1 Primary

- Restaurant dataset (RD): Scraped relevant restaurants (the ones that serve pizza or pita) along with many structured and semi-structured data such as district, category, open days/hours, available locations with their minimum costs, average category scores, approximate delivery time, etc. This dataset also includes the full menu information separate from the food-specific ones stored in the menu dataset (MD).
- Menu dataset (MD): Scraped relevant menu items of restaurants in RD. This dataset contains the restaurant names, the restaurant URLs, the item names, optional additional descriptions, and price values.
- Comment dataset (CD): Scraped user comments of restaurants listed in RD. This dataset contains the restaurant names, the restaurant URLs, review comment, restaurant's response (if any), and given scores. Orders that are cancelled by a customer or the restaurant itself are also automatically documented by Yemeksepeti in the form of a comment. If the cancellation is made by the restaurant (indicating they are not able to send the order for a reason), the amount of time passed between the order and the cancellation is included in minutes as well. These automated cancellation comments are also stored and distinctly labeled in this dataset.
- Menu lexicon dataset (ML): N-gram ingredient and side dish lexicons that are used with natural language processing (NLP) techniques (See Section 3.1.4.1) to extract menu information.
- Real estate listings (RE): Real estate prices all around Ankara, for rental and sale, scraped from one of the biggest online real estate website (Hürriyet Emlak [62]),
- Restaurant coordinates (RC): Restaurant coordinates that are later obtained in a semi-automated fashion mainly from Google Maps.

3.1.2.2 Secondary

- Population data of Ankara, Turkey (PD): A district-level dataset aggregated from a neighborhood-level dataset which is based on the official address-based population records from 2017. In the official records, neighborhoods that have a population smaller than 500 are not included. This dataset contains the population numbers of each age group for each district.
- Education status data of Ankara, Turkey (ED): A district-level dataset aggregated from a neighborhood-level dataset which is based on the official address-based population records from 2017. This dataset contains the population numbers for gender and age group pairs. District percentages are calculated from the raw numbers.

3.1.3 Data Collection

RD, MD, and CD are scraped from Yemeksepeti's website using a custom scraper written with Selenium in Python. All the restaurants' names and URLs in Ankara are scraped without filtering. Then, the scraper continuously ran to fetch restaurants' data when they are available. Scraping is spread over a week to avoid congesting the system and being restricted by the servers. On Yemeksepeti, restaurants are normally closed to visitors outside of working hours, but a restaurant can sometimes be viewed and accept timed orders. On the other hand, mostly due to a temporary congestion caused by high volumes of orders, a restaurant may not be viewed during its working hours as well. This iterative process also helped handling this limitation. No terms of use of Yemeksepeti [63] are violated during the process.

The scraper did not use restaurants' main categories as a filter since many restaurants that serve pizza or pita are listed in different categories, which would cause a significant information loss. On Yemeksepeti, each menu item must be listed under a menu category. The scraper analyzed the menu categories and detected categories dedicated to pizza and pita (or oven products since pita was frequently listed with other oven dishes). From those categories, "lahmacun" and "katıklı ekmek" (pita-like dishes that are much cheaper than pita) for pita and "pizza tost" (pizza sandwich similar to Hot Pockets sandwiches) are filtered out. For the rest, item name, description, price, and existence of a photo are collected along with the restaurant's name and URL. Menu data are stored in MD while general restaurant data (and unfiltered menu data) are stored in RD (57,770 menu items in total). Since most pizza restaurants serve the same dish in different sizes, only the first pizza category items are kept in MD to target smallest size pizza dishes in order to obtain more comparable menu items considering pita dishes are usually one-person dishes. However, since some menu items only existed in larger sizes, those menu items are retrieved from larger size categories to prevent ingredient information loss. Menu items that are listed but not available at the moment are collected as well since it was assumed that a permanent change on the menu would be reflected to the menu items as well. Review (comment) data (CD) from the relevant restaurants are also collected and saved separately. CD includes user rating(s) in speed, service, and taste categories (users can rate only the taste if the restaurant uses Yemeksepeti's own courier service), user comments (very short texts that can be at most 300 characters long), and the restaurant's response (if exists). Since it was not possible to uniquely identify reviewers, CD does not have the identity information. After an iterative process of separately collecting and analyzing these datasets, their final versions are collected as mentioned above in October 2019.

Administrative borders like districts and their generalized statistics do not necessarily apply to communities, but RD did not have specific coordinates (or even structured and standardized neighborhood data), so high-resolution geographic analysis was not possible. For further analysis, pinpoint restaurant locations and full addresses are later collected as well. The collection is firstly handled by a fully automatized process, through automated Google searches in polite and responsible frequencies. Levenshtein distance [64] is used to calculate the similarity ratio of the paired restaurants' names along with checking if the retrieved Google Places entity has the city, district, and neighborhood names in the full address and if the matched entity has already been matched with another restaurants before. It is discovered that restaurant addresses do not strictly follow the same format as the city, district, or neighborhood names are sometimes omitted. Therefore, a more flexible approach is implemented to handle these format issues. For each restaurant, if the retrieved entity does not have the required district in its address or does not have the required neighborhood in its address while not having a name close enough to the restaurant in question, the restaurants are not matched. To prevent false positive matches, after some test runs, the similarity threshold is initially set to 85%. A total number

of 423 restaurants are pinpointed. It is also discovered that there are subtle differences between the restaurant names listed on Yemeksepeti and on Google. The algorithm is run again, but this time, city/district/neighborhood names and common words related to the type of restaurant ("pizza," "pide," "kebab," "lahmacun," "kafe," "cafe," and "pastane") that tend to appear, disappear, or change place between platforms are removed during the comparison while the name similarity threshold is decreased to 70%. For the restaurants that do not satisfy the conditions, potentially matched restaurants are output together for human supervision. According to the human supervisor's input, restaurants are matched or passed. Some pizza branches had a few branches that were very close to each other that required a few restaurants to be manually matched. Eventually, 526 of 659 restaurants (79.82%) are matched with a pinpoint location in Ankara. These data are stored in RC and later added to RD as well. Restaurant locations in Ankara are shown in Figure 20.

During the location pinpointing process, it is found that a few restaurants were listed under different districts than their coordinates suggest. These restaurants were located in newly developed areas, close to the district borders. Some of them were also closer to the district center it was listed under, not by air distance but by the road network. These restaurants' districts are changed according to their coordinates and where those coordinates coincide. It is also found that some of the restaurants were located at the same place (manually confirmed through Google Street View), most probably operated by the same people. These seemingly separate restaurants are not merged together as they were actively operating on Yemeksepeti at the same time (based on the page accessibility and recent comments). Besides, they had different menus, delivery regions, and minimum costs. It is possible that these restaurants wanted to bypass cuisine filters that can be used by people on Yemeksepeti since a restaurant can only have one main cuisine category, except big brands that have their own cuisine category, or they wanted to liberate themselves from certain limitations. For example, a pita restaurant that serves döner can simply close its extension entity (that serves almost only döner) on Yemeksepeti when they sell out, instead of manually disabling ordering certain foods, which can be cumbersome.

For ML, menu items in RD are manually analyzed and separate n-gram lexicons are created for ingredients and side dishes. Alternatives and versions that have common misspellings are merged with the chosen form using another custom lexicon. Highly common, context-dependent, and problematic ingredients such as "süt" (milk) are intentionally left out since it is not practical and it does not give much information in this context. For example, while "süt" can occur in items made of milk such as "sütlâç" (rice pudding) and "sütlü tatlı" (pudding), it can also occur in items that are not made of milk, such as "süt danası" (calf) or "süt mısır" (sweet corn). Another such word is "tatlı" (dessert/sweet). Since "tatlı" also means sweet in Turkish and there is no possible distinction, it required more complex approaches to detect if "tatlı" means "dessert" or "sweet" (as in "sweet potato"). Therefore, it is simply not collected.

3.1.4 Pre-processing

3.1.4.1 Word Labeling and Menu Vectorization

To analyze restaurant menus based on their ingredients and side dishes, a lexicon was necessary. This section explains how the menu lexicon (ML) is created.

Ingredients and side dishes are extracted using Zemberek [65], which is an open source and comprehensive NLP framework written in Java for Turkish, and it has been widely used in academia and the industry. A list of words is initially formed by collecting Turkish noun words from the menu item descriptions. Normalization and lemmatization are also used. Additions, removals, and corrections are made through manually analyzing MD. Due to the agglutinative nature of Turkish, some terms such as "kaşar peyniri" are searched with some suffixes removed ("kaşar peynir") to catch occurrences such as "kaşar peynirli" ("with kaşar peyniri"). Alternative names for the same ingredients/side dishes are manually detected and used to create a correction list to automatically merge them before analysis for easier labeling and increased accuracy. "Bal" (honey) is searched as "ballı" (with honey) since all of its occurrence was as "ballı" and searching for "bal" can lead to errors as there are words such as "balık" (fish) that include the word "bal." Certain sauces such as barbeque sauce, marinara sauce, and bolognese sauce are found to exist without "sauce" (as "barbekü," "marinara," and "bolonez"), so they are searched without "sos" (sauce).

The word list is manually labeled firstly by the author and then two additional annotators in order to define whether a term is an ingredient, a side dish, both, or none of them. Annotators are also provided a "not sure" option that is mutually exclusive with being an ingredient or side dish. For each term, annotators are provided at most five example menu items from MD that include the specific term in their description. A screenshot of the custom web-based tool used by the annotators is shown in Figure 21. Fleiss' kappa score [66] is used to measure the agreement among the annotators. Indecisive ("not sure") votes are treated as a separate label. It is found that certain items could be served separately outside the context of MD, which caused discrepancies between the annotators' votes for specific items. These discrepancies in labeling are discussed with the annotators through solely considering the menu items found in MD and nothing else. After reconsiderations, kappa scores of 87.47% (for an item to be an ingredient) and 98.83% (for an item to be a side dish) are obtained. These labels are used to generate lexicons for ingredients and side dishes (ML). Majority vote is used to terminally classify items about which annotators were not in unanimous agreement. These lexicons are used to analyze variety and other characteristics of restaurants' menus. Lexicon items and annotator votes are provided in Appendix A.

Only one restaurant mentions serving ketchup directly with the dish. This is firstly deduced because they do not serve ketchup with any of the products or separately. Later, it is confirmed by the restaurant on phone. While ketchup and mayonnaise are separately included in meals in general, they are considered as ingredients since they are supposed to be eaten combined with the dish, just like other sauces. Besides, it is not common to consume them separately without other foods.

Some other ambiguous ingredient occurrences (such as lettuce on pizza) are confirmed by manually finding the phone numbers of the restaurants and asking them if they use those items as ingredients or serve them separately.

Whole grain and whole wheat are observed to be used as ingredients of an ingredient (dough) as some of the dishes were stated to be made from whole grain or whole wheat dough. However, dough are not categorized based on their own ingredients. Instead, whole grain and whole wheat are handled as separate ingredients. Similarly, items that can be made of different ingredients were treated separately. For example, a salami made from turkey is handled as (salami, turkey) while a salami made from beef is handled as (salami, beef). While this may have caused some information loss related to their co-existences, this is more convenient since these kind of items are function-wise rather similar and further separation would cause further sparsity. Certain seemingly similar ingredients such as "pastırma"

and "pastrami" are not combined due to the possibility that the occurrence of the non-traditional one (pastrami) might be helpful compared to the traditional one (pastırma).

Through these lexicons, two types of menu vectors are obtained for ingredients and side dishes of each restaurant; occurrence and existence vectors. The occurrence vectors simply hold the occurrences of each ingredient/side dish in the menu item name or description while the existence vectors hold their binarized version, indicating if a specific ingredient or side dish exists in a specific restaurant's menu. Some restaurants had menu items that had rather cryptic names such as "eski sevgilim pizza" ("my ex-lover pizza") that do not give much information without looking at the description while some restaurants had self-explanatory menu item names such as "sucuklu ananaslı pizza" ("pizza with sucuk and pineapple"), which caused certain ingredients to be counted twice for each menu item. Furthermore, restaurants had naturally varying numbers of menu items. Therefore, it is decided that using the existence vectors of ingredients or side dishes would be more reliable.

3.1.4.2 Comment Vectorization

To model user reviews, comment words that exist in the corresponding restaurant's name are removed from each comment and they are normalized, tokenized, and lemmatized. After these processes, punctuations, numbers, numbers in text, abbreviations, conjunctions, words that are measurement units, proper nouns, region/city/district names, words that exist in the restaurant's name (this is done twice to make sure they are removed), some auxiliary verbs, and words that are not morphologically explainable (words that do not stem from a known Turkish word from Zemberek's dictionary) are removed to reduce noise. Other than the recognized words, a small amount of emoticons that exist in the dataset and are recognized by Zemberek are collected as well (":-), " ::", " :(", " :/", and " :p"). The same processes are repeated for the responses given by the restaurants. However, processed responses are not used to obtain vectors. To filter out comments that do not provide enough information, comments that have less than three words before pre-processing or less than two words after pre-processing (two words are required to obtain at least one co-occurrence) are removed. After the filtering, 95,954 comments (44,396 for pizza restaurants, 44,848 for pita restaurants, and 6654 for restaurants that serve both) are obtained ready to be modeled. However, comments that belong to restaurants that serve both dishes are removed since those restaurants tend to have more variety in their menus which can confuse the model with irrelevant dishes.

To model user comments, BTM [57] is used to cluster comments without supervision. This prevents introducing a bias and missing comment topics that might not be considered without reading all of the comments. BTM is preferred over LDA due to its biterm (unordered word pairs that exist within a predefined window) based nature and its relative suitability for short texts such as these user reviews that do not exceed 300 characters. Pre-processed tokens of each comment are recombined in the correct order to obtain sentences formed with lemmas, and a window of four is used to extract word pairs. Special Turkish characters are simplified to prevent problems with visualization and processing. The number of topics is chosen as 100, an arbitrary number that should be big enough to let cohesive topics form. As a result, topics with highly explicable term co-occurrences are formed. Top 50 most frequent term co-occurrences for pizza and pita restaurants (out of 5047 and 4800 terms respectively) are shown in Figure 2 and Figure 2. It seems that the most frequent terms are about aspects of the restaurants, such as "hız" (speed), "servis" (service), and "lezzet" (taste); or adjectives that define these

aspects, such as "sicak" (hot), "iyi" (good), and "soguk" (cold). There are also terms that indicate situations about giving an order, arrival of the food, or thanking for the delivery.

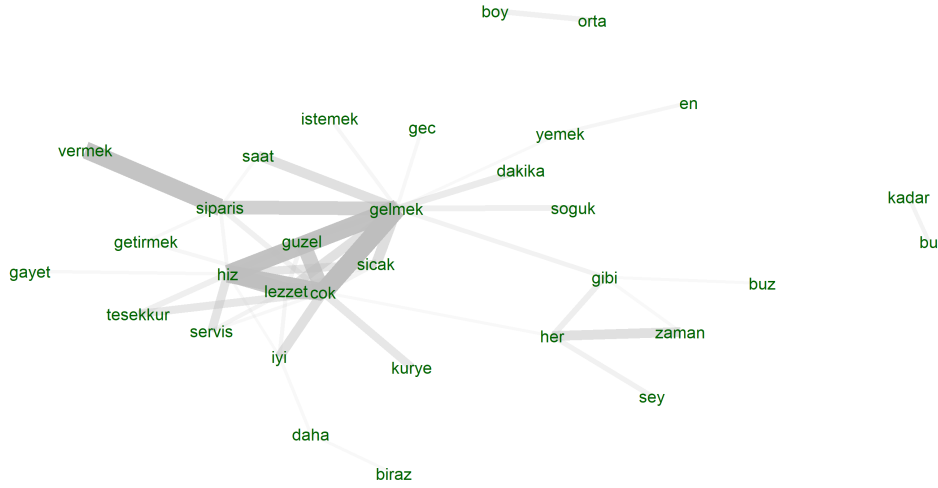


Figure 1: Top 50 most frequent term co-occurrences from pizza restaurants' user reviews. A thicker edge indicates higher frequency.

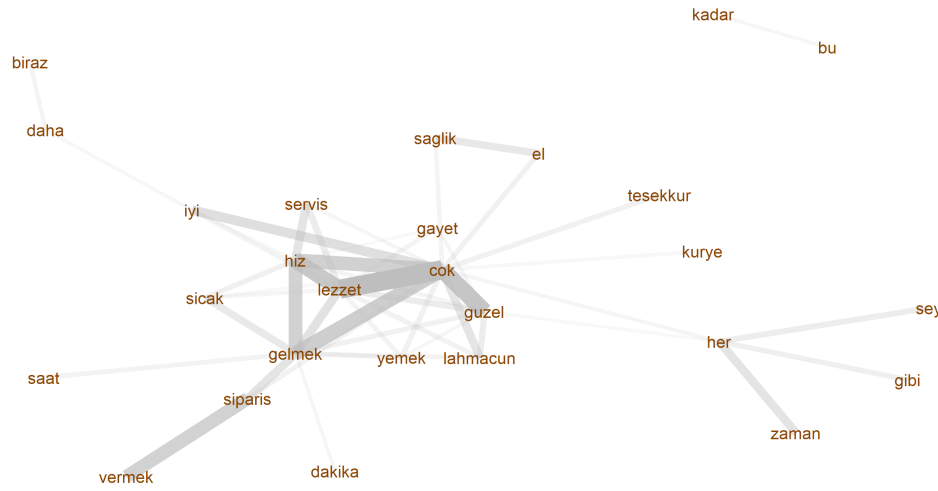


Figure 2: Top 50 most frequent term co-occurrences from pita restaurants' user reviews. A thicker edge indicates higher frequency.

Topic likelihood vectors are obtained for each review using the model. These vectors are grouped by restaurants and their topic-wise geometric means are calculated to obtain a single topic likelihood vector for each restaurant.

3.2 Data Analysis

3.2.1 Descriptive Statistics

1958 restaurants are found to be located in Ankara on Yemeksepeti. From these, 1105 are found to be menu-wise irrelevant. From the rest, 414 restaurants that serve pita and 294 restaurants that serve pizza (while 49 of them serving both) are identified. All the restaurants were tried to be scraped multiple times, but the rest were not accessible or successfully scraped, mostly due to long-term unavailability on the system. In total, after removing 1105 restaurants that are accessed and deemed irrelevant, out of 853 potentially relevant restaurants, 659 (77.26%) could be accessed and collected, which makes 33.66% of all the restaurants located in Ankara on Yemeksepeti. Unsurprisingly, more than half of the scraped restaurants were located in Çankaya, the district that has the largest population, the highest socioeconomic status, and the most educated people. These restaurants' district distributions are listed in Table 2. Eleven other districts that do not have restaurants that serve pizza or pita are not included. On Yemeksepeti, while there are more restaurants that serve pita, it seems like pizza is more widespread in terms of districts. Districts that have pita also have pizza while the other way around is not always true. Among the districts that have both, Pursaklar is the only one that has more pizza-serving restaurants than pita-serving restaurants. This is interesting, because Pursaklar is an underdeveloped district located towards the northeast part of the city (Ankara is growing towards west and southwest [67]) that also has a less educated population, which suggests one would expect to see more traditional restaurants.

Table 2: Ankara's districts along with their number of restaurants that serve pizza or pita and their populations and university graduate rates

District	Pizza	Pita	Population	Univ. grad. %
Altındağ	5	13	370729	13.68
Beypazarı	2	-	40358	12.37
Çankaya	144	193	920067	46.28
Çubuk	2	-	81853	9.53
Elmadag	1	-	43341	11.08
Etimesgut	34	52	565196	29.28
Gölbaşı	13	23	123267	25.05
Kahramankazan	1	-	46153	12.71
Keçiören	31	36	917226	18.21
Mamak	15	25	634763	13.50
Polatlı	3	4	110162	13.33
Pursaklar	5	3	141251	13.12
Sincan	6	18	519477	11.89
Yenimahalle	32	47	658907	28.31

A significant number of restaurants that serve pizza or pita were not listed under their respective categories, especially the ones that serve pizza. Categories under which the restaurants are listed are shown in Table 3. Three restaurants were found to also serve an item called "pizza pita." These items had relatively more ingredients compared to other pita dishes (hence "pizza pita"). Still, they are categorized as pitas due to having very similar prices with other restaurant-specific pitas, using "pizza" as an adjective, and not featuring pizza-specific characteristics, such as mozzarella cheese and tomato sauce.

Table 3: Restaurants that serve pizza or pita along with their primary categories (cuisines)

Category (cuisine)	Pizza	Pita
Börek (Pastry)	-	3
Cafe	41	17
Domino's Pizza (Brand)	52	-
Döner	7	27
Dünya Mutfağı (World Cuisine)	12	2
Ev Yemekleri (Homemade Food)	1	4
Fast Food & Sandwich	8	1
Kahvaltı (Breakfast)	1	-
Kebab & Türk Mutfağı (Kebab & Turkish Cuisine)	14	243
Kokoreç (Grilled Lamb Intestines)	1	-
Köfte (Meatball)	-	3
Little Caesars Pizza (Brand)	4	-
Papa John's Pizza (Brand)	9	-
Pasta & Tatlı (Cake & Dessert)	7	3
Pide (Pita)	12	101
Pilav (Cooked Rice)	-	1
Pizza & İtalyan (Pizza & Italian)	103	5
Pizza Hut (Brand)	6	-
Pizza Time (Brand)	16	-
Steak	-	2
Tavuk (Chicken)	-	2

9239 food-specific menu items (4674 pizzas and 4565 pitas) are collected. On Yemeksepeti, menu items can have individual descriptions (an extra text field that gives more information about the ingredients, menu items, or side dishes) and photographs. Menu-related statistics shown in Table 4 are calculated from these menu items. However, unlike pita restaurants, there are big pizza chains with many branches in the dataset. Since pizza chains have standardized menus, their menu contents do not differ much based on location, which homogenizes the data and affects the distributions. Restaurants that have their own cuisine types (such as Papa John's Pizza or Pizza Hut as seen in Table 3) are removed while keeping smaller chains that do not have their own categories on Yemeksepeti. These statistics are calculated from those non-chain pizza restaurants. It is seen that pizza restaurants' data generally have more variance and they have more elaborate menus with longer descriptions and more photographs. As expected, pizza restaurants have more ingredients, more menu items (possibly due to the number of ingredient combinations and different sizes), and slightly higher prices.

Table 4: Food-specific statistics for pizza-serving and pita-serving restaurants

Feature	Statistic	Pizza	Pita
N_ITEM	Mean	14.21	11.03
	Median	12.00	9.00
	Standard deviation	10.15	6.58
	Skewness	1.26	1.89
PRICE_MEDIAN	Mean	23.94	21.39
	Median	23.50	21.00
	Standard deviation	7.18	4.72
	Skewness	0.88	0.03
ITEM_AVG_DESC	Mean	9.61	3.79
	Median	9.60	3.81
	Standard deviation	2.27	2.34
	Skewness	1.86	0.60
N_INGREDIENT_SIDEDISH	Mean	25.99	8.45
	Median	25.00	8.00
	Standard deviation	12.30	4.46
	Skewness	0.57	1.20
PHOTO_PERCENTAGE	Mean	14.00	9.00
	Median	0.00	0.00
	Standard deviation	30.00	26.00
	Skewness	1.94	2.61

N_ITEM Number of items

PRICE_MEDIAN Median of menu item prices

ITEM_AVG_DESC Average description length (in words)

N_INGREDIENT_SIDEDISH Number of unique ingredients and side dishes

PHOTO_PERCENTAGE Percentage of items that have a photograph (%)

For 659 restaurants, 120,536 reviews (all written in the last 6 months) are collected. As explained in Section 3.1.2.2, automated comments indicating cancelled orders (1400 of them are cancelled by restaurants while 598 are cancelled by the customers) are collected. Apart from the menu item statistics, the number of comment, percentage of comments responded by the restaurant, number of orders cancelled by the users, number of orders cancelled by the restaurant, total number of order cancellations, geometric mean of after how many minutes the restaurant cancels the orders (this is automatically provided once an order is cancelled by the restaurant), and geometric mean of category scores (in speed, service, and taste) for each restaurant are calculated as well. Throughout this study, geometric mean is frequently preferred as a measure of central tendency due to its robustness against outliers.

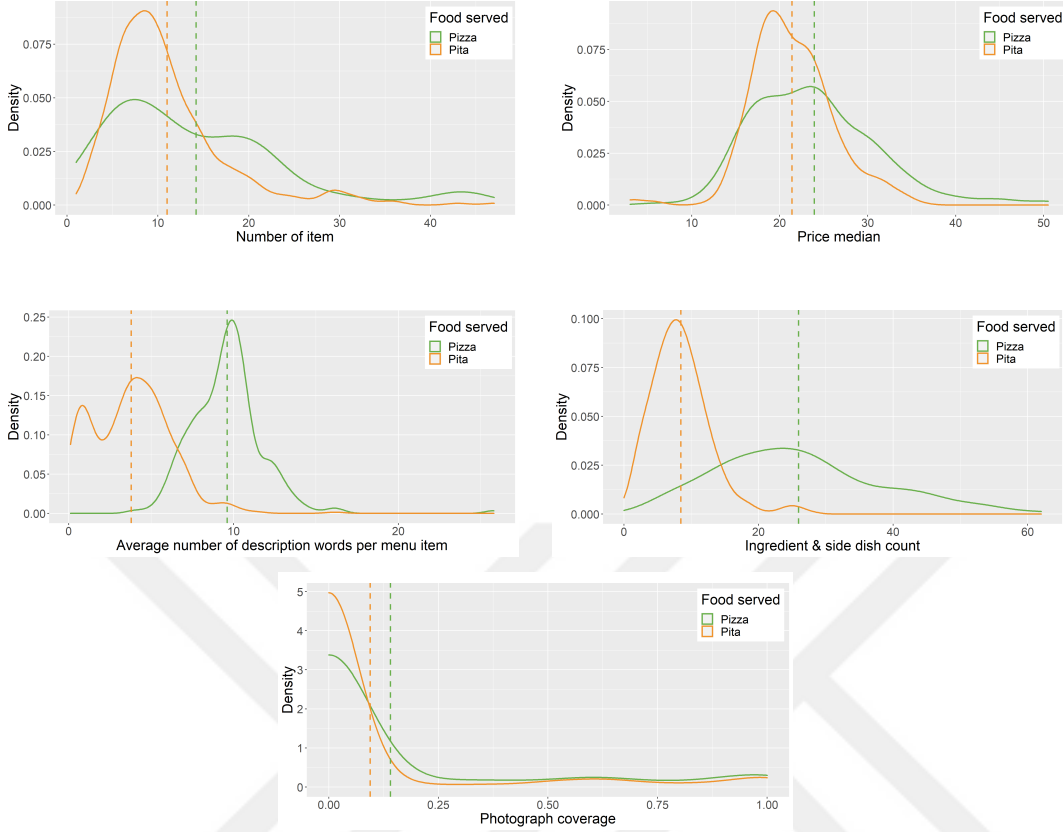


Figure 3: Menu descriptive statistics with means shown as dashed vertical lines

3.2.2 Menu Statistics and Socioeconomic Status

RQ1 Is there a relationship between a restaurant’s menu curation characteristics and its district’s socioeconomic status?

It is hypothesized that districts with higher socioeconomic status would have more diverse and elaborate menus or vice versa, reflecting the districts’ residents and their expectations. Therefore, these restaurants’ menus may have certain distinguishing qualities that reflect these differences as well. As indicators of diversity and elaboration, statistics such as number of items, photograph usage percentages, description lengths, and number of ingredients and side dishes are analyzed. Considering restaurants only have the district information that can be easily used to enrich the data, there are only a few features that could be used such as education rates, age-gender distributions, and land/realty prices. Education rates is chosen since it is more identifiable than the other features, it is based on the official numbers, a study on Ankara indicates that education is related to socioeconomic status [20], and education has a strong correlation with income in urban areas of Turkey [21]. Therefore, education rates from ED are used as a proxy of socioeconomic status.

Food-specific (considering only pizza or pita items) menu characteristics (average description length, number of items, number of unique ingredients and side dishes, percentage of items that have a photograph, and median item price) of restaurants are calculated from MD. These characteristics are then

compared with the education rates of each district (from MD) using the restaurants' locations. Pearson correlation coefficient is used as a correlation metric. Due to changes made in the education system, it has become rather complicated to interpret the education rates in Turkey. Table 23 shows the education rates of the districts based on their residents' highest completed degree. There are some categories that seem to overlap, but they actually do not share any resident overlaps. Schools that were known as primary schools (taking eight years to complete) are divided into elementary and middle schools that take four years each. Therefore, someone who had graduated from primary school can be at the same education level with someone who has recently graduated from middle school. To handle these confusions and deal with less categories, education rates are grouped. People who are illiterate and literate (able to read and write without completing a formal education) as "no formal education." People who completed elementary, middle, or primary school are grouped under "basic education." People who completed a secondary education are grouped (or rather renamed) under "high school education." Finally, people who have any kind of college degree are grouped under "university education."

For pita restaurants, it is found that there is a moderate positive correlation between restaurants' price medians and their districts' university graduate rates ($r=0.394$, $p<.001$, $N=414$). Similarly, a moderate positive correlation between price medians and university graduate rates is found for non-chain pizza restaurants ($r=0.314$, $p<.001$, $N=213$).

While price medians unsurprisingly have a correlation with districts/education rates, this relationship may not provide a means to confidently predict a restaurant's location or demographics of its customer base. Therefore, a more sophisticated approach is required.

3.2.3 District Prediction with Menu Contents and Restaurant/Menu Statistics

RQ2 Is it possible to predict a restaurant's district by looking at its menu contents and/or statistics?

It is hypothesized that, due to cultural or economic reasons, restaurants in certain districts might typically have certain ingredients or side dishes that may distinguish them from other districts.

A machine learning approach is implemented to see if it is possible to predict a restaurant's district based on its menu content, food-specific menu statistics mentioned in the previous subsection, and restaurant statistics such as restaurant score, order cancellation rates, latency to cancel an order by the restaurant, and review response percentages. To reflect the menu content, as explained in Section 3.1.4.1, ingredient and side dish vectors that indicate the presence/absence of each item (as categorical variables) for each restaurant are obtained. For pizza restaurants, only the ingredient lexicon is used while both ingredient and side dish lexicons are used for pita restaurants due to pitas being commonly served with side dishes. These are combined with the restaurant and menu statistics. Apart from the district and restaurant labels, including 13 statistical features that are food-specific menu and restaurant statistics (which will be also referred as "Statistics" as a feature set from now on), 146 features for pizza and 110 features for pita are used. The feature count difference stems from the fact that pizza and pita restaurants has different menu contents. These different types of features are used both separately and combined.

The model of choice is random forest due to its off-the-shelf applicability, simplicity, and its ability to handle high dimensional imbalanced datasets. Restaurants that serve both pizza and pita, or belong

to pizza chains¹ are removed to reduce noise and interference to districts. Listwise deletion is also applied to restaurants that had missing data (due to not having the data available). Then, districts that did not have at least five restaurants are eliminated as well. Unfortunately, this forces the models to work with districts that also happen to have relatively higher education rates, which is not surprising and possibly harmful for the model's success. This way, 139 pizza restaurants and 354 pita restaurants that are suitable are trained separately. The districts used by the models are shown in Table 5. Due to the relatively small size of datasets, leave-one-out cross-validation is used. Model parameters are optimized through grid search. However, it is found that although the test scores do not change drastically, the models tend to overfit through automated hyperparameter optimization using validation sets. To prevent this, tree depth is manually decreased although it leads to lower test scores.

Table 5: The districts used by the models during the classification along with their member sizes

District	Univ. grad. %	Pizza restaurants	Pita restaurants
Altındağ	13.68	-	12
Çankaya	46.28	83	170
Etimesgut	29.28	13	43
Gölbaşı	25.05	6	17
Keçiören	18.21	16	31
Mamak	13.50	7	22
Sincan	11.89	-	17
Yenimahalle	28.31	14	42

Due to the class imbalance, F1 score is used to evaluate a model and report its performance. ZeroR algorithm (which simply predicts everything as the most frequent class) is used as a baseline. These baseline scores are then compared with the F1 scores obtained from models trained with different sets of features. For both pita and pizza restaurants, F1 scores that are slightly better than the baseline F1 scores are obtained. Interestingly, the best result is obtained by solely using the statistical features for both, which were expected to be less effective than using only the menu content or combining them together. An F1 score of 0.49 is obtained with pizza restaurants compared to a baseline score of 0.446 while an F1 score of 0.339 is obtained with pita restaurants compared to a baseline score of 0.312. These models' F1 scores along with the baseline are shown in Table 6.

From the best models used to predict each restaurant, average weights and median ranks are calculated for the used features. While statistical features are found to be most useful for both pita and pizza restaurant, the importance of individual statistical features vary as shown in Table 7. Feature importances are calculated using Mean Decrease Accuracy (permutation-based) and Mean Decrease in Impurity (impurity-based) obtained from the training and validation sets. It can be seen that the average description length and the percentage of photographed items are more important for pizza restaurants while the response percentages, order cancellation percentages, and the number of comments are more important for pita restaurants. Price medians and score geometric means are useful for both, but more prominent for pita restaurants.

¹ This chain removal is not applied to pita restaurants because there are no pita restaurants that have their own categories, the number of branches are relatively low for pita chains, and determining the pita chains is not as easy due to the fact that pita restaurants seemingly try to imitate each other's names instead of differentiating themselves. For example, many pita restaurants in Ankara use the name "Aspava" that is originally used by a restaurant upon which there is still not an agreement.

Table 6: Classification scores for pizza and pita restaurants using different types of features

Food	Feature set	Train-Validation		Test
		Mean F1 score	F1 score std.	F1 score
Pizza	Menu Content	0.843	0.011	0.457
	Statistics	0.791	0.012	0.490
	Menu Content + Statistics	0.864	0.016	0.460
	Baseline	-	-	0.446
Pita	Menu Content	0.640	0.010	0.327
	Statistics	0.850	0.008	0.339
	Menu Content + Statistics	0.842	0.012	0.330
	Baseline	-	-	0.312

Baseline: Predicting all data points to be a member of the most common class (ZeroR).

While being slightly over the baseline score, this approach is not reliable to distinguish restaurants either. Confusion matrices for the best models for pizza and pita restaurants are given in Table 8 and Table 9. They indicate a bias towards Çankaya. Most of confusions also happen between adjacent and closer districts, which may be suggesting an intermingling relationship between districts as expected. However, these districts are mostly the closest ones to the city center anyway, which increases the chance of confusions between adjacent districts. Working with higher resolutions (such as the exact coordinates instead of districts) may improve the analyses and the prediction success.

Table 7: Feature importances for pizza and pita restaurants, ranked by permutation-based feature importances

Food	Feature	Permutation-based				Impurity-based			
		Mean	Std.	Min.	Max.	Mean	Std.	Min.	Max.
Pizza	ITEM_AVG_DESC	0.101	0.010	0.071	0.124	0.134	0.004	0.123	0.145
	SCORE_GEO_MEAN	0.075	0.009	0.053	0.098	0.126	0.004	0.118	0.136
	RESPONSE_PERCENTAGE	0.074	0.010	0.049	0.096	0.102	0.005	0.090	0.112
	PRICE_MEDIAN	0.069	0.010	0.045	0.092	0.105	0.004	0.096	0.116
	PHOTO_PERCENTAGE	0.055	0.008	0.035	0.072	0.075	0.005	0.058	0.086
	DESC_WORDS	0.045	0.013	0.010	0.073	0.082	0.004	0.072	0.091
	LATENCY_GEO_MEAN	0.044	0.012	0.017	0.071	0.061	0.003	0.052	0.069
	N_ITEM	0.044	0.010	0.024	0.069	0.069	0.003	0.063	0.076
	N_COMMENT	0.042	0.013	0.001	0.071	0.082	0.005	0.071	0.093
	CANCEL_PERCENTAGE	0.036	0.009	0.009	0.056	0.065	0.004	0.055	0.074
	N_CANCEL	0.022	0.008	0.003	0.040	0.043	0.003	0.036	0.05
	N_CANCEL_RESTAURANT	0.020	0.008	-0.001	0.042	0.041	0.003	0.035	0.047
	N_CANCEL_CUSTOMER	0.003	0.004	-0.003	0.016	0.016	0.002	0.010	0.023
Pita	PRICE_MEDIAN	0.189	0.007	0.171	0.208	0.160	0.003	0.152	0.168
	RESPONSE_PERCENTAGE	0.154	0.008	0.132	0.179	0.109	0.003	0.101	0.115
	SCORE_GEO_MEAN	0.152	0.009	0.128	0.175	0.121	0.003	0.112	0.128
	CANCEL_PERCENTAGE	0.095	0.009	0.069	0.128	0.085	0.003	0.078	0.096
	N_COMMENT	0.089	0.009	0.059	0.117	0.103	0.003	0.094	0.112
	N_ITEM	0.077	0.008	0.053	0.105	0.075	0.003	0.067	0.082
	ITEM_AVG_DESC	0.076	0.008	0.051	0.099	0.086	0.003	0.079	0.093
	DESC_WORDS	0.073	0.009	0.046	0.096	0.081	0.002	0.075	0.088
	LATENCY_GEO_MEAN	0.051	0.007	0.027	0.071	0.064	0.002	0.058	0.070
	N_CANCEL	0.026	0.006	0.014	0.043	0.042	0.002	0.037	0.048
	N_CANCEL_RESTAURANT	0.024	0.006	0.010	0.041	0.034	0.002	0.030	0.039
	PHOTO_PERCENTAGE	0.013	0.004	0.001	0.023	0.015	0.001	0.012	0.019
	N_CANCEL_CUSTOMER	0.012	0.004	0.001	0.023	0.026	0.002	0.022	0.031

ITEM_AVG_DESC	Average number of description words per item
SCORE_GEO_MEAN	Score geometric mean
RESPONSE_PERCENTAGE	Percentage of comments that are responded by the restaurant
PRICE_MEDIAN	Menu item price median
PHOTO_PERCENTAGE	Percentage of menu items that have a photograph
DESC_WORDS	Total number of description words
LATENCY_GEO_MEAN	Order cancellation latency geometric mean
N_ITEM	Number of menu items
N_COMMENT	Number of comments
CANCEL_PERCENTAGE	Order cancellation percentage compared to the number of comments
N_CANCEL	Number of order cancellations
N_CANCEL_RESTAURANT	Number of order cancellation by the restaurant
N_CANCEL_CUSTOMER	Number of order cancellation by the customer

Table 8: Confusion matrix for pizza restaurants using the best model

Predicted \ Truth	Çankaya	Etimesgut	Gölbaşı	Keçiören	Mamak	Yenimahalle
Çankaya	83	11	6	16	7	11
Etimesgut	0	0	0	0	0	0
Gölbaşı	0	0	0	0	0	0
Keçiören	0	1	0	1	0	0
Mamak	0	0	0	0	1	0
Yenimahalle	0	1	0	0	0	2

Table 9: Confusion matrix for pita restaurants using the best model

Predicted \ Truth	Altındağ	Çankaya	Etimesgut	Gölbaşı	Keçiören	Mamak	Sincan	Yenimahalle
Altındağ	0	0	0	0	0	0	0	0
Çankaya	9	159	35	15	24	18	6	32
Etimesgut	0	3	2	0	2	1	1	4
Gölbaşı	0	0	1	2	0	0	0	1
Keçiören	1	4	0	0	2	1	2	1
Mamak	1	0	0	0	2	0	2	0
Sincan	0	0	2	0	1	2	4	1
Yenimahalle	1	4	3	0	0	0	2	3

3.2.4 Socioeconomic Status Prediction with User Reviews, Menu Contents, and Menu/Restaurant Statistics

RQ3 Is it possible to predict a restaurant’s surrounding area’s socioeconomic status based on its user reviews?

It is hypothesized that areas with residents from higher socioeconomic layers may have a different language or different expectations/emphasis on restaurants’ features/problems that can be extracted from their reviews. It is known and shown in district-level education dataset (ED) (see Table 23 in Appendix B) that Ankara’s districts have varying levels of education. It has been also explained in Section 3.2.2 that education is a good proxy of socioeconomic status. However, as the previous analyses have shown, using district-level data may not be suitable to make neighborhood-level predictions due to various reasons:

1. Administrative (district) borders do not necessarily apply on how people settle. A neighborhood may show characteristics that are more similar to a neighbor district.
2. Districts can be relatively big meanwhile restaurants typically only accept orders from closer neighborhoods. Therefore, a restaurant’s customers might not represent the district or vice versa due to possible special cases creating microclimate-like settlements in districts.
3. Some central and crowded districts such as Çankaya can have neighborhoods that significantly vary.

To overcome these difficulties, it is decided to work with finer-level datasets for this research question. However, neighborhood-level education data are not publicly accessible. Instead, as a socioeconomic status indicator, it is decided to use real estate listings instead to approximate an area’s wealth as it has been found to be relevant [23] and useful [22]. In a previous study [22], online real estate listings were scraped from Hürriyet Emlak [62]. A dataset that is collected later the same way (RE) is used. The dataset was collected in June-July 2019 and pre-processed to remove listings that suspiciously overlap and point to the coordinates of real estate agents’ offices. Due to real estate properties being an investment tool and the fact that the socioeconomic status of the landlords do not necessarily reflect the socioeconomic status of the tenants who live and review those restaurants, only the rental listings are used. Listings that are not classified as residential are removed. To reduce the noise, top and bottom 20% of the listings for each neighborhood from the dataset are removed. The remaining 4017 out of 10965 listings in total are used with restaurants that are pinpointed.

Since restaurants only accept orders from people who are located at closer neighborhoods, and people can only review restaurants through their previous orders on Yemeksepeti, it is assumed to be relatively safe to associate the comments to the local residents. Through manual considerations of some restaurants’ delivery regions on the map, it is decided that a 3-5km of radius is a fair approximation for a restaurant’s delivery area. For all the restaurants with known coordinates, rental listings along with their price per square meter values within a 5km ellipsoid distance are aggregated. To minimize the outliers’ effect, geometric mean is used and restaurants that did not have at least three listings nearby were discarded. After the removal, out of 526 restaurants with coordinates, 499 restaurants (198 serving pizza, 270 serving pita, and 31 serving both) with enough nearby realty listings are obtained in total. Restaurants that serve only pizza or pita are shown with their districts and their square

meter price geometric means in Figure 4 and Figure 5. Number of nearby realty listings within 5km radius for these restaurants are shown in Figure 6 and 7. In terms of both square meter prices and number of listings, pita restaurants seemingly have more variance compared to pizza restaurants. This is especially clear with Altındağ, Etimesgut, Gölbaşı, and Sincan districts.

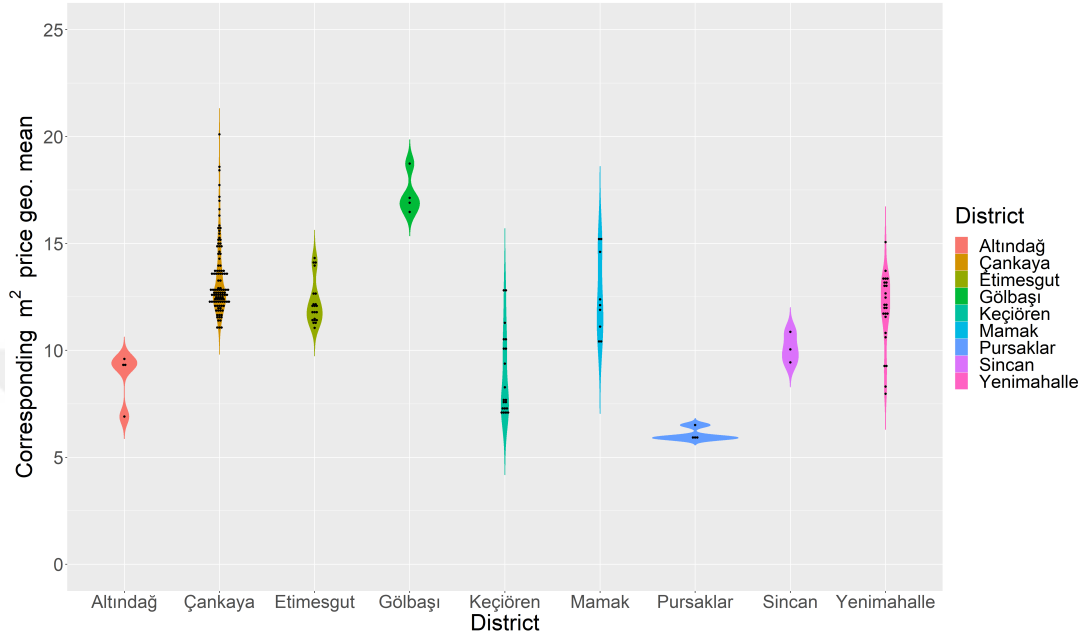


Figure 4: District and square meter price distributions of pizza restaurants

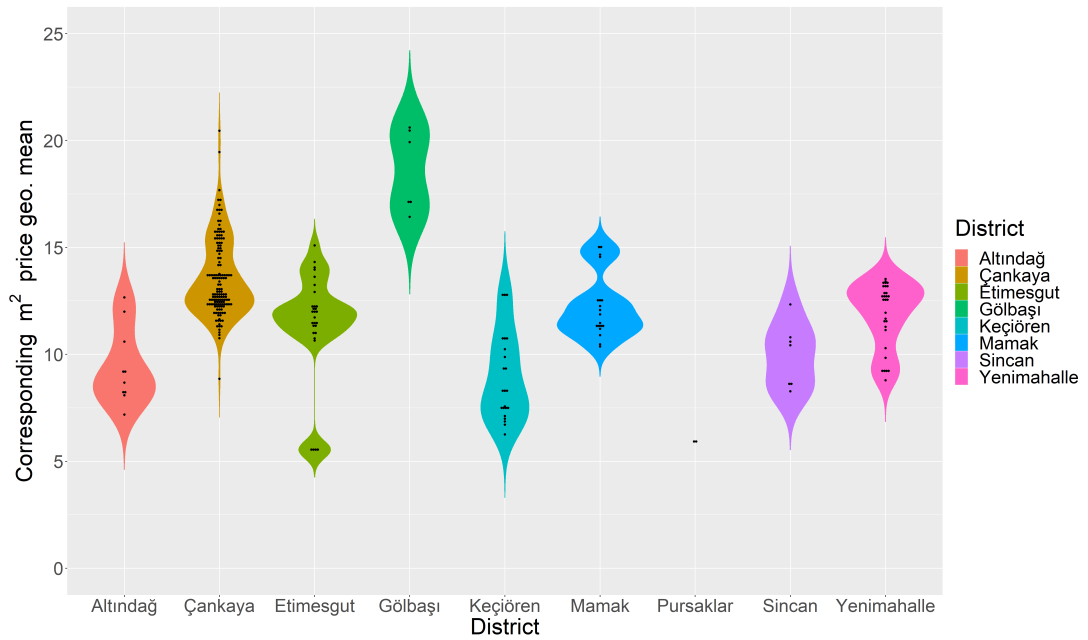


Figure 5: District and square meter price distributions of pita restaurants

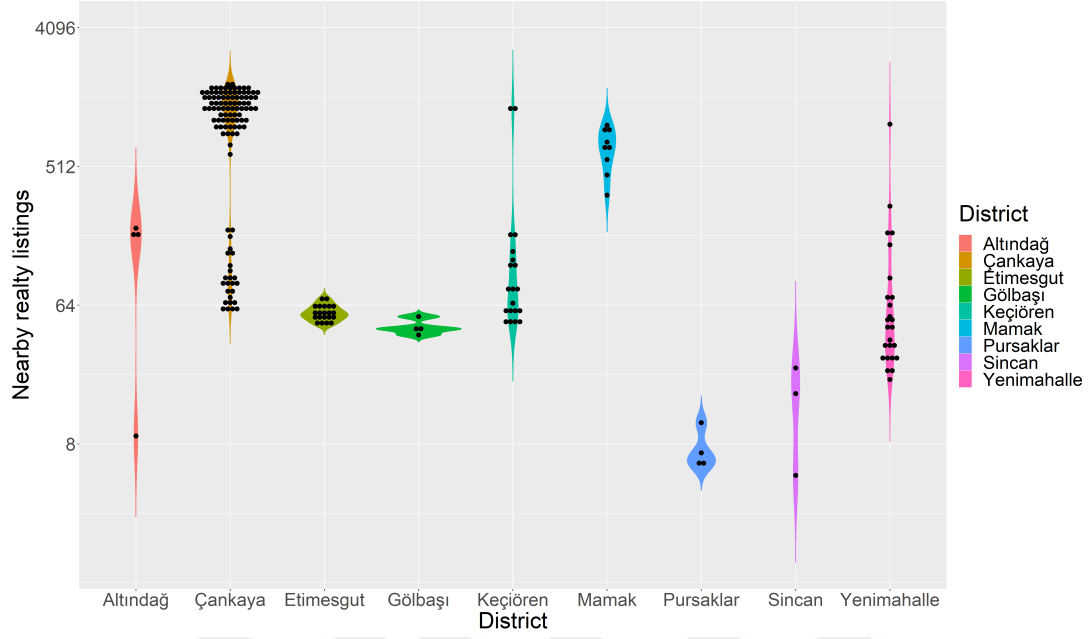


Figure 6: District and realty listings distributions of pizza restaurants

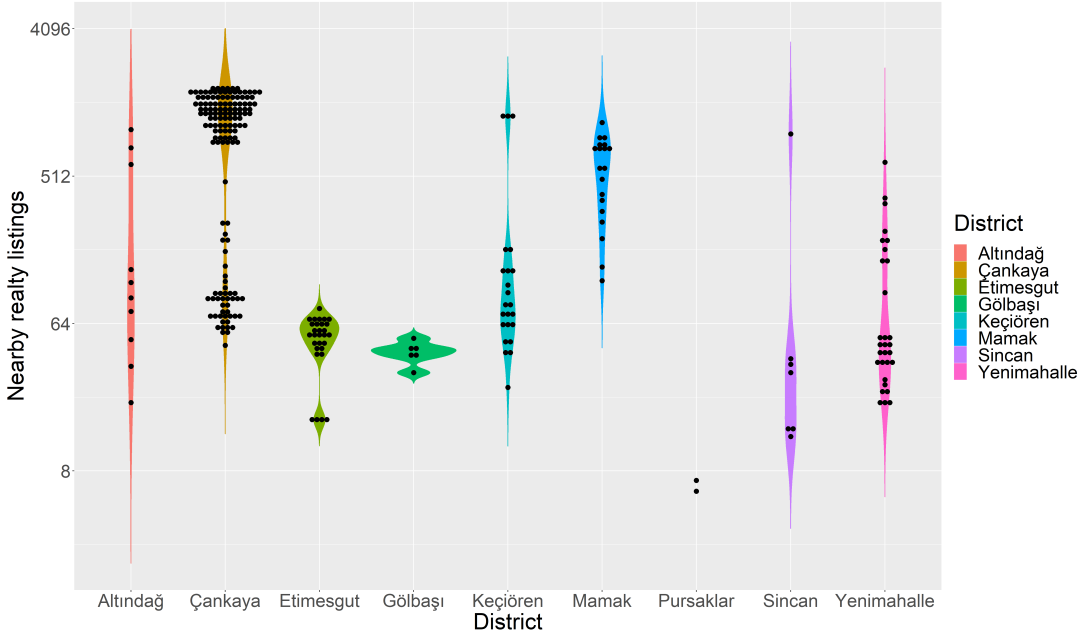


Figure 7: District and realty listings distributions of pita restaurants

As explained in Section 3.1.4.2, user comments for pizza and pita restaurants are separately modeled using BTM to obtain a topic likelihood vector for each comment. Since it is attempted to find about which topics customers talk independent from their sentiment, review rates or sentiments are not used. Restaurants that serve both pizza and pita are not used for modeling in order to prevent more noise in user comments. Comment-level topic likelihood vectors are then aggregated based on the restaurant

and geometric means for each topic likelihood are obtained, producing a restaurant-level topic likelihood that generalizes their customers' comment focus. A total of 100 comment topics (named T1 to T100) are obtained for pita and pizza restaurants, separately.

To predict restaurants' corresponding square meter price geometric means, these restaurant-level vectors are used to train random forest regressors, with similar reasons explained in the previous section. Restaurants that belong to pizza chains are removed at this point to reduce noise and interference to districts. Apart from this, since restaurants are required to either serve pizza or pita, have coordinates, have comments, have enough realty listings nearby, and have menu items collected (assuming menu data would be used to boost the model success), only 67 pizza and 143 pita restaurants were usable at this point. Pizza restaurants and pita restaurants are trained separately since their comments would largely differ from each other. Due to the small number of datasets, leave-one-out cross-validation is used. Model parameters are optimized through grid search.

Other than the topic likelihoods; menu vectors, one-hot encoded districts, number of ingredients and side dishes, number of unique ingredients and side dishes, food-specific menu item price medians, score geometric means calculated from category-level restaurant scores, average number of description words per item, and restaurant response percentages to reviews are used as well. Different combinations of different types of feature sets are used to see how each one affects the success of the model. While random forests are considered to be able to handle many features, the curse of dimensionality affects their success nonetheless, especially for this regression task. For each used feature set combination, a semi-exhaustive feature selection is applied through successive iterations of feature subsetting and grid search that make use of feature importances and hyperparameter optimization. This lead the models to have few features that perform better compared to models that have many features. While both type of restaurants have 100 comment topic features and 13 statistical features, the number of menu content features differ for pizza restaurants (134) and pita restaurants (97).

Model results are provided in Table 10 for pizza restaurants and in Table 11 for pita restaurants. Feature explanations and importances for each model are given in Table 12 for pizza and Table 13 for pita restaurants. The main difference between permutation-based importances and impurity-based importances is the impurity-based approach is biased towards high cardinality features [68], causing binary features (menu or district features in this case) being likely to have a lower importance score. Baseline models that predict solely based on the arithmetic or geometric means of square meter price values for each district are provided as well. However, these models are not comparable with the models that do not use district information at all. Also, these baseline scores being rather high is possibly caused by the cut offs applied to the data. It is found that, without using any district information, menu content and statistics yield the best results for pizza restaurants while comment topics replace menu content for pita restaurants, yielding the best results with the help of statistical features. The existence of "roka" (rocket) ingredient, the percentage of user reviews getting responded by the restaurant, and the existence of "köfte" (meatball) ingredient yield the best result (an R^2 of 0.205) for pizza restaurants while three different comment topics along with the score geometric means and menu item price medians yield the best result (an R^2 of 0.181) for pita restaurants. While the response percentage of restaurants to user reviews is an important feature for pizza restaurants, its relationship with the square meter prices is not that simple or linear. A faceted scatter plot that shows their relationship for each district is given in Figure 8. While there is a clear positive correlation for districts like Çankaya and Yenimahalle, Keçiören defies the trend and all districts together produce a highly scattered figure. It mostly leads to higher square meter prices on trees, but it can occasionally be seen to lead both ways in the same tree.

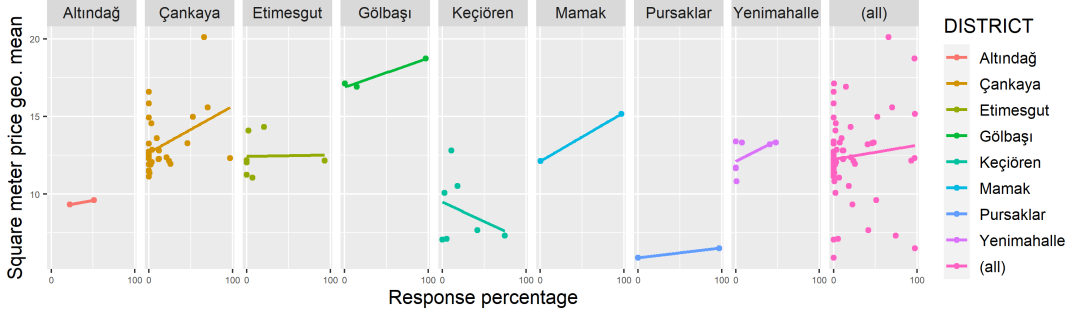


Figure 8: Scatter plots of response percentages and square meter prices for each district

Admittedly, while these models may be giving an insight, their predictive success is rather limited. To see the difference having the district information makes, restaurant districts are later one-hot encoded and combined with the feature sets. Using comment topic, menu, and district features together increases the R^2 to 0.242 for pizza restaurants while using district features with comment topics increase the R^2 to 0.258 for pita restaurants.

Table 10: Model results for pizza restaurants with different feature sets

Feature set	Train-Validation		Test	
	Mean R^2	R^2 std.	R^2	MSE
Comment Topics	0.375	0.012	0.104	5.90
Menu	0.222	0.010	0.140	5.66
Menu + Statistics	0.426	0.016	0.205	5.24
Comment Topics + Menu + Statistics	0.531	0.013	0.171	5.45
Comment Topics + District + Menu	0.520	0.012	0.242	4.99
District + Menu	0.317	0.013	0.208	5.21
District	0.200	0.016	0.145	5.63
Baseline 1	-	-	0.629	2.44
Baseline 2	-	-	0.627	2.46

Baseline 1: Predictions solely based on the arithmetic mean of each district’s square meter prices.

Baseline 2: Predictions solely based on the geometric mean of each district’s square meter prices.

To understand how these selected topic features are helpful, BTM topics are visualized using LDavis [69] and terms’ relevance scores are calculated. According to Sievert and Shirley [69], relevance score for a given term and topic is shown in Equation 1:

$$r(w, k|\lambda) = \lambda \log(\phi_{kw}) + (1 - \lambda) \log\left(\frac{\phi_{kw}}{p_w}\right) \quad (1)$$

where ϕ_{kw} denotes the probability of term w in topic k , p_w denotes the probability of term w in the corpus, and therefore λ denotes the weight given to the term’s topic-specific probability relative to the weight given to the term’s lift. In this sense, relevance is a weighted average between log values of a term’s topic-specific probability and its topic exclusivity (similar to "distinctness" measure proposed

Table 11: Model results for pita restaurants with different feature sets

Feature set	Train-Validation		Test	
	Mean R^2	R^2 std.	R^2	MSE
Comment Topics	0.577	0.007	0.147	5.06
Comment Topics + Statistics	0.639	0.005	0.181	4.86
Comment Topics + Menu + Statistics	0.637	0.005	0.158	4.99
Comment Topics + District	0.470	0.005	0.258	4.40
District	0.239	0.006	0.200	4.75
Baseline 1	-	-	0.451	3.26
Baseline 2	-	-	0.445	3.29

Baseline 1: Predictions solely based on the arithmetic mean of each district’s square meter prices.

Baseline 2: Predictions solely based on the geometric mean of each district’s square meter prices.

in [70]) to the topic with a given weight λ that is between 0 and 1. λ directly determines the weight of topic-specific probability while inversely determining the weight of the lift ($1 - \lambda$). A weight of 0.6 is suggested and it is stated that relevance-wise top terms tell a cohesive story and help the topic distinguish itself [69]. Another research also suggests a weight of 0.6 works fine with a different dataset as well [71]. Therefore, a weight of 0.6 is used in this study to interpret and visualize the topics. However, it is also observed that a higher weight such as 1 does not significantly change the overall narratives of the topics visualized below, which is different from what is observed in [69]. All 100 topics along with their top 10 relevant terms in descending order (calculated with $\lambda = 0.6$) are provided in Table 28 and Table 29. A minimum threshold to filter the terms is not applied.

Table 12: Feature importances for square meter price prediction with pizza restaurants

Feature set	Feature	Permutation-based importance			Impurity-based importance				
		Mean	Std.	Min.	Max.	Mean	Std.	Min.	Max.
Comment Topics	T39	0.302	0.033	0.162	0.371	0.441	0.039	0.280	0.528
	T9	0.175	0.023	0.108	0.241	0.325	0.037	0.219	0.472
	T15	0.103	0.018	0.066	0.169	0.234	0.035	0.156	0.349
Menu	ROKA	0.361	0.024	0.293	0.417	0.651	0.039	0.549	0.748
	KÖFTE	0.131	0.009	0.101	0.154	0.349	0.039	0.252	0.451
Menu + Statistics	ROKA	0.536	0.058	0.381	0.661	0.250	0.024	0.183	0.305
	RESPONSE_PERCENTAGE	0.314	0.016	0.263	0.343	0.588	0.024	0.534	0.694
	KÖFTE	0.110	0.018	0.039	0.169	0.162	0.029	0.053	0.245
Comment Topics + Menu + Statistics	ROKA	0.280	0.041	0.184	0.402	0.156	0.019	0.112	0.220
	T39	0.231	0.037	0.128	0.312	0.291	0.040	0.193	0.384
	T9	0.229	0.025	0.154	0.309	0.316	0.032	0.226	0.423
	RESPONSE_PERCENTAGE	0.142	0.016	0.090	0.177	0.237	0.026	0.159	0.309
Comment Topics + District + Menu	DISTRICT_KEÇİÖREN	0.246	0.028	0.163	0.305	0.293	0.032	0.187	0.376
	T39	0.183	0.024	0.055	0.219	0.273	0.031	0.138	0.328
	T15	0.177	0.015	0.137	0.205	0.224	0.019	0.175	0.283
	T9	0.074	0.012	0.039	0.100	0.146	0.019	0.092	0.181
	ROKA	0.052	0.012	0.023	0.094	0.063	0.013	0.036	0.098
District + Menu	DISTRICT_KEÇİÖREN	0.208	0.027	0.130	0.264	0.425	0.049	0.288	0.547
	ROKA	0.197	0.021	0.137	0.234	0.349	0.035	0.259	0.444
	KÖFTE	0.092	0.010	0.049	0.117	0.226	0.035	0.111	0.340
District	DISTRICT_KEÇİÖREN	1	0	1	1	1	0	1	1

T39 Comment topic that includes terms with positive sentiments on the speed, hotness of the food, and the courier

T9 Comment topic that includes terms with positive sentiments on the taste, speed, hotness of the food, and the abundance of the ingredients

T15 Comment topic that includes terms with positive sentiments on the taste, speed, and service, featuring some compliments

ROKA Existence of roka (rocket) in the restaurant's menu

KÖFTE Existence of köfte (meatball) in the restaurant's menu

RESPONSE_PERCENTAGE Percentage of reviews getting a response from the restaurant

DISTRICT_KEÇİÖREN Boolean Keçiören district indicator

Table 13: Feature importances for square meter price prediction with pita restaurants

Feature set	Feature	Permutation-based importance				Impurity-based importance			
		Mean	Std.	Min.	Max.	Mean	Std.	Min.	Max.
Comment Topics	T14	0.519	0.019	0.455	0.592	0.393	0.012	0.354	0.420
	T70	0.347	0.010	0.309	0.369	0.302	0.011	0.279	0.334
	T88	0.295	0.012	0.253	0.333	0.305	0.013	0.273	0.332
Comment Topics + Statistics	T14	0.341	0.020	0.293	0.433	0.261	0.010	0.234	0.294
	T70	0.206	0.010	0.166	0.229	0.184	0.010	0.147	0.213
	T88	0.205	0.013	0.165	0.250	0.201	0.009	0.169	0.227
	SCORE_GEO_MEAN	0.167	0.011	0.124	0.197	0.187	0.011	0.153	0.216
	PRICE_MEDIAN	0.162	0.011	0.133	0.194	0.166	0.009	0.141	0.195
Comment Topics + Menu + Statistics	T14	0.322	0.018	0.280	0.414	0.256	0.010	0.230	0.291
	T70	0.201	0.010	0.157	0.220	0.179	0.010	0.144	0.203
	T88	0.198	0.012	0.160	0.238	0.196	0.009	0.163	0.219
	SCORE_GEO_MEAN	0.158	0.011	0.118	0.186	0.180	0.010	0.143	0.207
	PRICE_MEDIAN	0.151	0.011	0.123	0.182	0.160	0.009	0.136	0.185
	MANTAR	0.018	0.003	0.010	0.026	0.028	0.003	0.020	0.042
Comment District	DISTRICT_ÇANKAYA	0.264	0.021	0.194	0.312	0.273	0.014	0.229	0.333
	T88	0.163	0.020	0.111	0.210	0.217	0.016	0.162	0.253
	T14	0.113	0.018	0.074	0.157	0.192	0.029	0.126	0.257
	T70	0.091	0.006	0.076	0.110	0.132	0.008	0.111	0.164
	DISTRICT_KEÇİÖREN	0.057	0.010	0.031	0.091	0.187	0.026	0.109	0.263
District	DISTRICT_ÇANKAYA	0.187	0.010	0.155	0.219	0.568	0.029	0.472	0.667
	DISTRICT_KEÇİÖREN	0.128	0.008	0.103	0.156	0.432	0.029	0.333	0.528

T14 Comment topic that includes terms with positive sentiments on side dishes, desserts, and free offerings

T88 Comment topic that mentions "ıçlı köfte" and "çığ köfte" (side dishes)

T70 Comment topic that mentions ordering spicy lahmacun (a pita-like dish with minced meat) and not being satisfied with it

SCORE_GEO_MEAN Geometric mean of the restaurant's scores in speed, service, and taste

PRICE_MEDIAN Food-specific price median in menu

MANTAR Existence of mantar (mushroom) in the restaurant's menu

DISTRICT_ÇANKAYA Boolean Çankaya district indicator

DISTRICT_KEÇİÖREN Boolean Keçiören district indicator

These topics' top 10 relevant term visualizations are shown in Figure 9, 10, 11, 12, 14, and 13. As also explained in [69], in these figures, the blue bars indicate overall frequency of a term in the corpus while the red bars indicate the topic-specific frequency, which helps understanding a term's commonality and exclusivity. English translations for the terms are provided right below the words, italicized and between parentheses. While Topic 39 for pizza restaurants emphasizes the courier and how hot the order was, it seems like the most distinguishing topics for pizza restaurants are somewhat similar to each other, having similar positive sentiments with some different nuances such as courier or the abundance of ingredients. At the first glance, it looks like Topic 88 for pita restaurants is about meatballs' ("köfte") insides ("iç") being raw ("çiğ"). However, after scrutinizing relevant reviews, it is understood that those reviews actually mention "çiğ köfte" (a side dish that got its name from actually being made of raw meat before it was banned to do so) and "içli köfte" (a side dish that is made of bulgur wheat paste and filled with spiced ground meat). While there are two other meatball-defining terms ("hamburger" and "kasap") along with some miscellaneous words, due to having prominently low topic-specific or overall frequencies, it can be said that this topic is about these two side dishes. Another deceiving term from pita restaurant reviews is "ezmek" (to crush) as it actually refers to "ezme" (a spicy side dish made from mashed tomatoes along with some other vegetables and herbs) which comes from the verb "ezmek" (to crush/smash). The other useful topics for pita restaurants are about spicy lahmacun and positive sentiments on side dishes, which indicate lower and higher square meter prices respectively. Some minor lemmatization or user errors can be spotted in topics. For example, Zemberek has a tendency to lemmatize "salata" (salad) to "salât" while the term "herşey" is a common misspelling of "her şey" (everything). In Topic 70, "belirmek" (to appear) might be also a lemmatization error on a word that could be about "belirtmek" (to state). Topic 14 for pita restaurants is clearly about positive remarks on side dishes, salad, desserts, free offerings, and them being fresh.

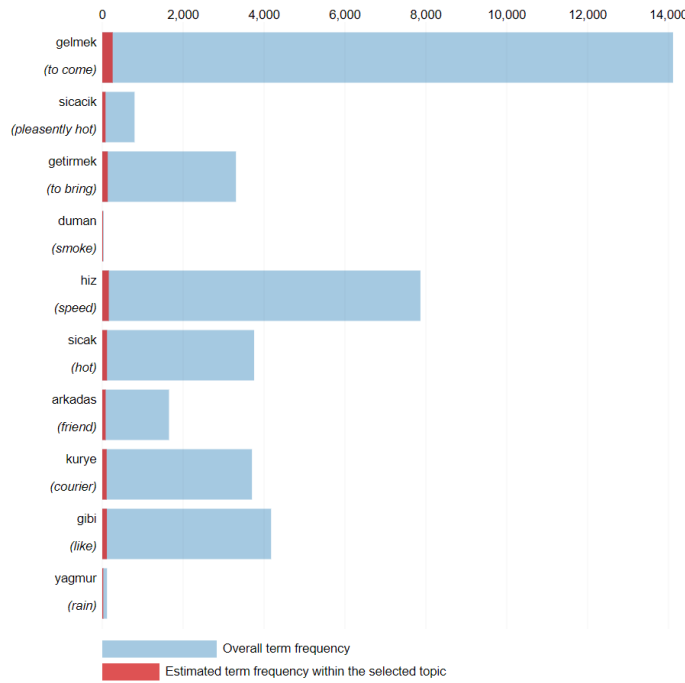


Figure 9: Top 10 most relevant terms of Topic 39 (from pizza restaurant reviews)

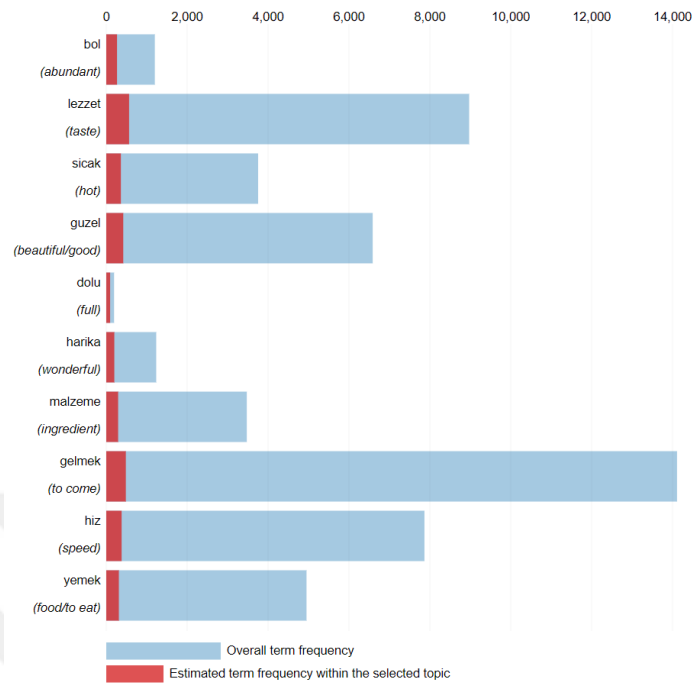


Figure 10: Top 10 most relevant terms of Topic 9 (from pizza restaurant reviews)

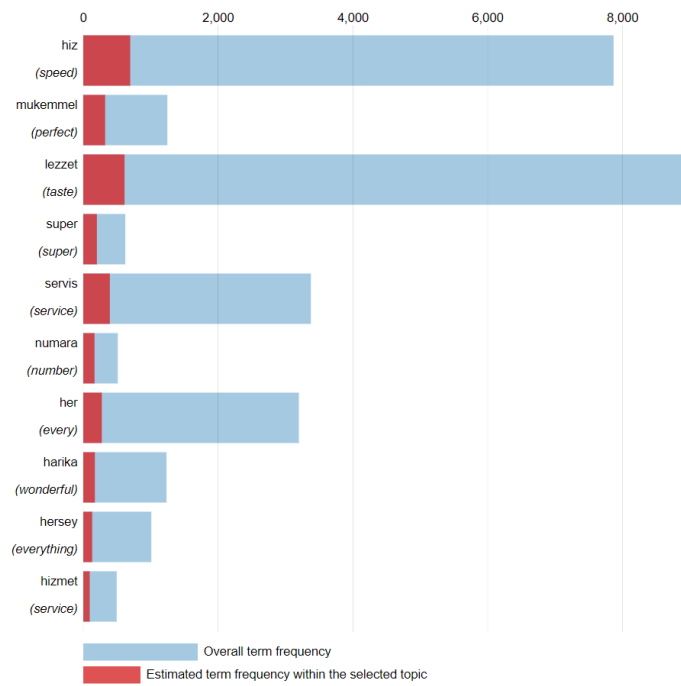


Figure 11: Top 10 most relevant terms of Topic 15 (from pizza restaurant reviews)

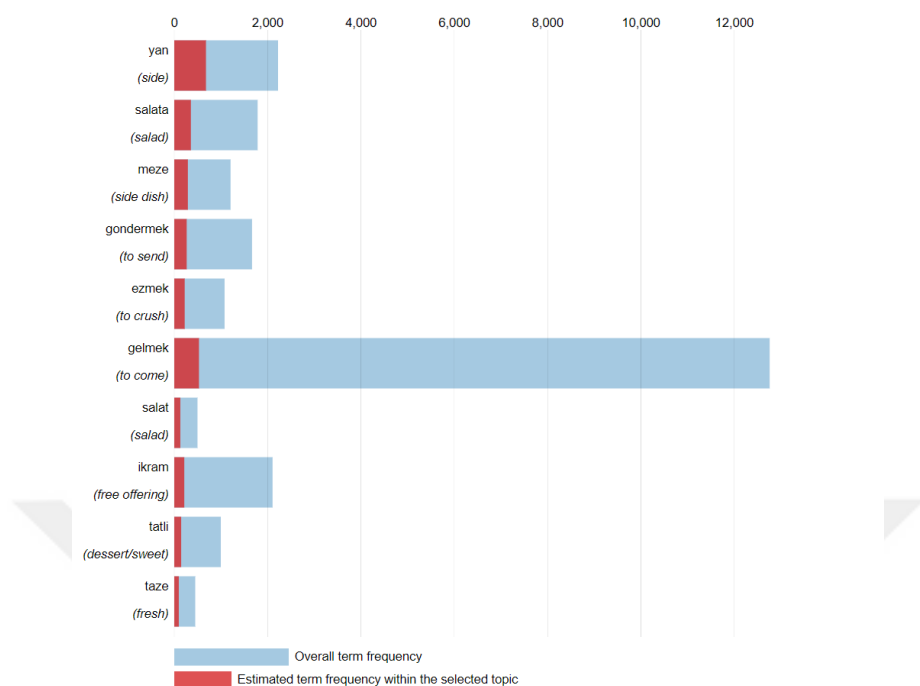


Figure 12: Top 10 most relevant terms of Topic 14 (from pita restaurant reviews)

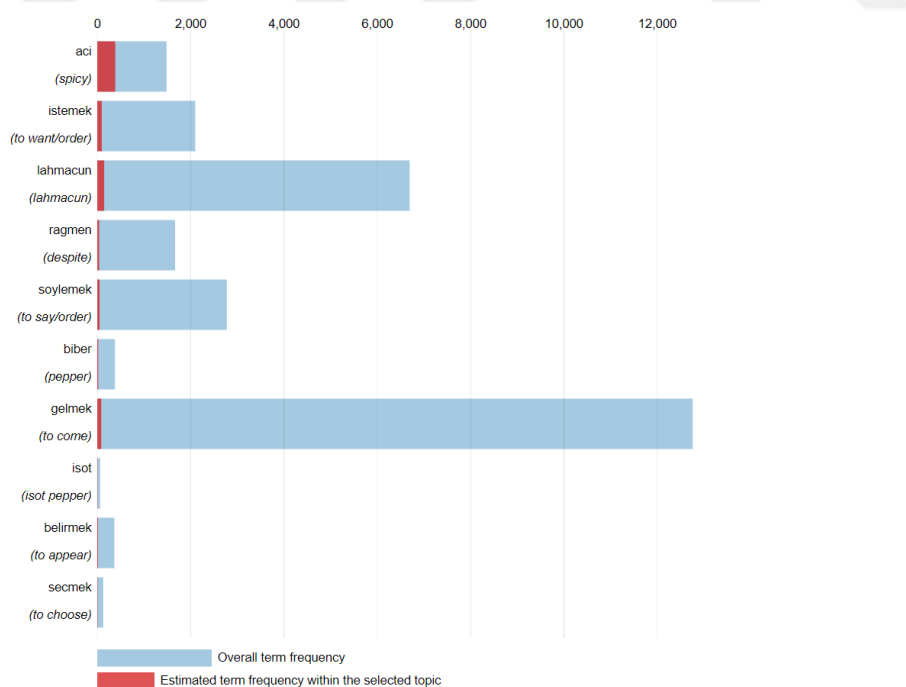


Figure 13: Top 10 most relevant terms of Topic 70 (from pita restaurant reviews)

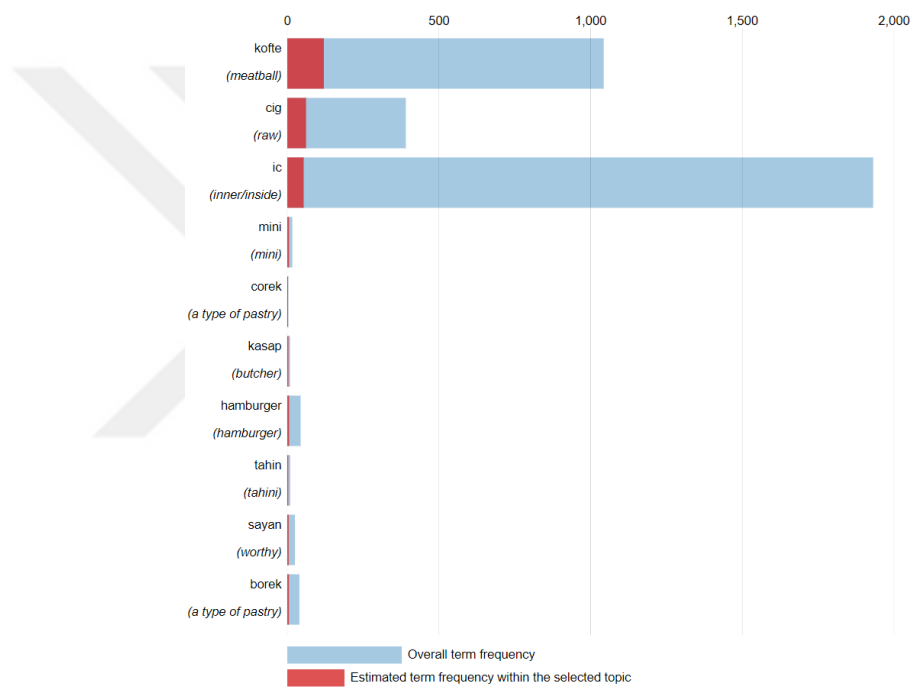


Figure 14: Top 10 most relevant terms of Topic 88 (from pita restaurant reviews)

To see if ingredient or side dish existences correlate with the square meter prices, the point-biserial correlation [72] which is suitable for correlations between binary and continuous variables is used. While being parametric and requiring normal distribution, the Central Limit Theorem suggests that the sample group approximates to normal distribution with the increase of the sample size, especially with sample sizes larger than 30 [73], which suggests that it can be used with pizza (N=67) and pita (N=143) datasets. It is found that only pizza restaurants' menu contents have significant moderate correlations, shown in Table 14, which is inline with the other findings. While the square meter price increases with the existence of "roka" (rocket), "soya sosu" (soy sauce), and "nacho" (nacho chips), they decrease with the existence of "köfte" (meatball). The reason nacho chips and soy sauce are not seen as important as meatball or rocket by random forest models is possibly their relative low frequency even at the city level, which may also render their point-biserial analysis less reliable as well.

Table 14: Point-biserial correlations between menu contents and square meter prices for pizza restaurants (N=67).

Menu content	R	P
ROKA (rocket)	0.395	<.001
SOYA SOSU (soy sauce)	0.329	.007
NACHO (nacho chips)	0.304	.012
KÖFTE (meatball)	-0.305	.012

With these restaurants, correlations of the district and the corresponding education rates for that district with the square meter prices are analyzed as well. Point-biserial correlation is used for one-hot encoded binary district features while Spearman correlation is used for education rates. Significant moderate correlations that are found are shown in Table 15, 16, 17, and 18. It can be seen that, for both pizza and pita restaurants, square meter prices increase with higher education rates or Gölbaşı district while they decrease with lower education rates or Keçiören district. These correlations not only justify the use of realty prices as an indicator of socioeconomic status but also support the previous findings and indicates the possibility of using district-level data with this dataset to predict the socioeconomic status. However, instead of directly using districts, this time it is decided to cluster them based on their education rates and check if the features used with the random forest models significantly differ between education clusters.

To cluster districts, district education rates shown in Table 23 are rescaled and clustered into two groups using k-means clustering. The ratio of people with unknown educations are excluded. The number of clusters is determined by looking at average silhouette widths with different cluster counts. As shown in Figure 15, having two clusters produces the most cohesive groups with an average silhouette width that is clearly closer to 1. Education rates are not grouped before clustering as they were grouped in Section 3.2.2 for correlation analysis, because of a slight silhouette width difference. The cluster centers are shown in Table 19. These clusters suggest that the breaking point is secondary education for Ankara districts, which is in line with the fact that education up until secondary education (high school) has been compulsory in Turkey. Therefore, Cluster 1 districts (Çankaya, Etimesgut, Gölbaşı, and Yenimahalle) are the ones that go well beyond the compulsory education while Cluster 2 districts (Akyurt, Altındağ, Ayaş, Bala, Beypazarı, Çamlıdere, Çubuk, Elmadağ, Evren, Güdül, Haymana, Kahramankazan, Kalecik, Keçiören, Kızılcahamam, Mamak, Nallıhan, Polatlı, Pursaklar, Şereflikoçhisar, and Sincan) lag behind with more of their residents having lower levels of education.

Table 15: Spearman correlations between education rates and square meter prices for pizza restaurants (N=67). Education rates correspond to the highest education obtained. See Table 23 for more information.

Education rate	R	P
PHD	0.391	.001
MASTERS	0.336	.005
UNDERGRAD	0.336	.005
ELEMENTARY	-0.335	.006
PRIMARY	-0.340	.005
MIDDLE	-0.345	.004
LITERATE	-0.346	.004
ILLITERATE	-0.348	.004

Table 16: Spearman correlations between education rates and square meter prices for pita restaurants (N=143). Education rates correspond to the highest education obtained. See Table 23 for more information.

Education rate	R	P
PHD	0.491	<.001
MASTERS	0.463	<.001
UNDERGRAD	0.463	<.001
PRIMARY	-0.459	<.001
MIDDLE	-0.463	<.001
ELEMENTARY	-0.463	<.001
LITERATE	-0.464	<.001
ILLITERATE	-0.473	<.001

Table 17: Point-biserial correlations between one-hot encoded binary district features and square meter prices for pizza restaurants (N=67).

District feature	R	P
DISTRICT_GÖLBAŞI	0.438	<.001
DISTRICT_PURSAKLAR	-0.424	<.001
DISTRICT_KEÇİÖREN	-0.461	<.001

These clusters are visualized in Figure 16. It can be seen that districts that have higher education rates are located in the central area.

The Kruskal-Wallis test is used to compare comment topic feature likelihoods between education clusters. The Bonferroni correction is applied to prevent false positives. While no significance is found for pizza restaurants, likelihoods of Topic 14 and Topic 2 of pita restaurants are found to significantly differ between clusters, which are shown in Table 20. All comparisons are provided in Table 24 and Table 25 for pizza and pita restaurants respectively. Top 10 most relevant terms of Topic 2 for pita

Table 18: Point-biserial correlations between one-hot encoded binary district features and square meter prices for pizza restaurants (N=143).

District feature	R	P
DISTRICT_ÇANKAYA	0.409	<.001
DISTRICT_GÖLBAŞI	0.337	<.001
DISTRICT_KEÇİÖREN	-0.385	<.001

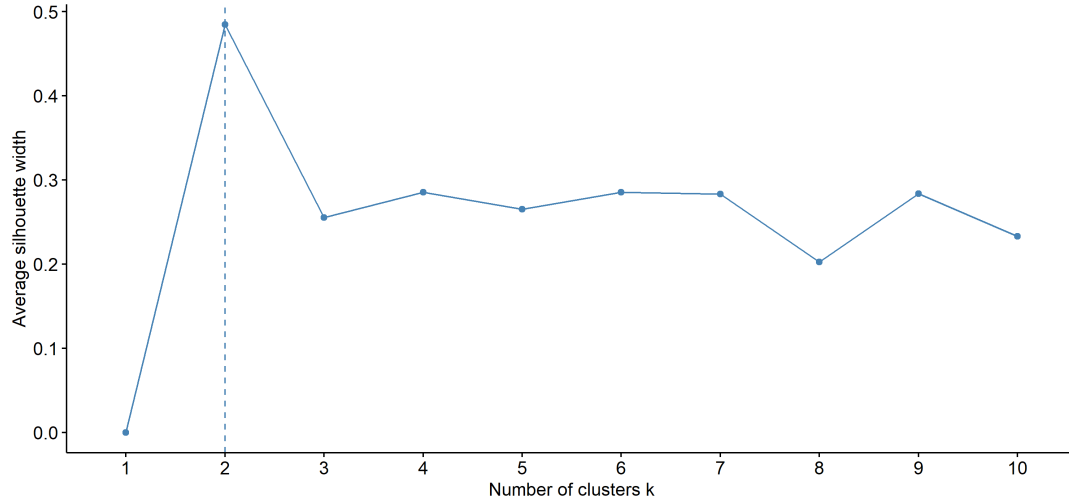


Figure 15: Average silhouette width with different numbers of clusters

Table 19: District clusters centers with their scaled education rates

Cluster	ILLITERATE	LITERATE	ELEMENTARY	MIDDLE	PRIMARY
1	-0.93	-0.93	-1.52	-1.42	-1.54
2	0.18	0.19	0.29	0.27	0.29
Cluster	SECONDARY	UNDERGRAD	MASTERS	PHD	
1	1.10	1.96	1.94	1.80	
2	-0.21	-0.37	-0.37	-0.34	

restaurants are shown in Figure 17. Topic 2, seemingly not having significantly distinctive terms, includes terms with positive sentiments on the delivery speed and taste. According to the cluster means for these topics' likelihoods, restaurants from districts with higher education rates (Cluster 1) are more likely to have positive comments on side dishes or positive comments in general.

For binary features that indicate the existence of a certain ingredient/side dish for a restaurant, contingency tables are created for existence/non-existence and two education clusters. Due to previously mentioned filtering processes, some ingredients and side dishes only had restaurants that unanimously feature or not feature them, so it was not possible to make comparisons for them and they are removed. Fisher's exact test is preferred over chi-square test due to having fewer observations and its suitability

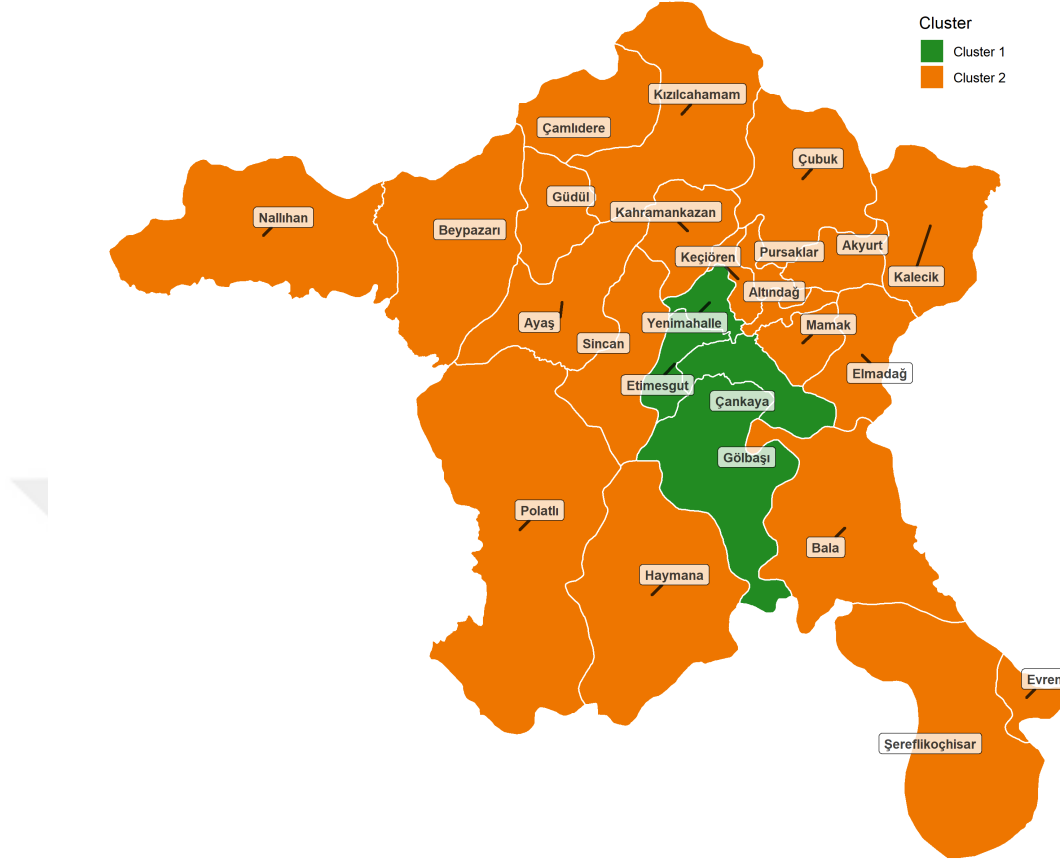


Figure 16: Districts clustered by their education rates

Table 20: Pita restaurants' comment topic likelihoods that significantly differ between two education clusters, the Kruskal-Wallis test results with Bonferroni-corrected p-values

Comment topic	H	P	Cluster 1 mean $\pm$ std.	Cluster 2 mean $\pm$ std.
T14	7.09	.016	0.004 $\pm$ 0.007	0.002 $\pm$ 0.002
T2	5.71	.034	0.019 $\pm$ 0.016	0.011 $\pm$ 0.007

for these conditions [74, 75]. Ingredients or side dishes that significantly differ between education clusters are shown in Table 21 while all comparison results are given in Table 26 for pizza restaurants and Table 27 for pita restaurants. For pizza restaurants, it is found that the existence of "köfte" (meatball), "roka" (rocket), and "kaburga" (rib, mostly referring to short ribs) significantly differ between two education clusters. Odds ratios indicate that meatball is more popular in less educated districts while it is the opposite for rocket and ribs. For pita restaurants, French fries seem to be more common in more educated districts, which is not surprising.

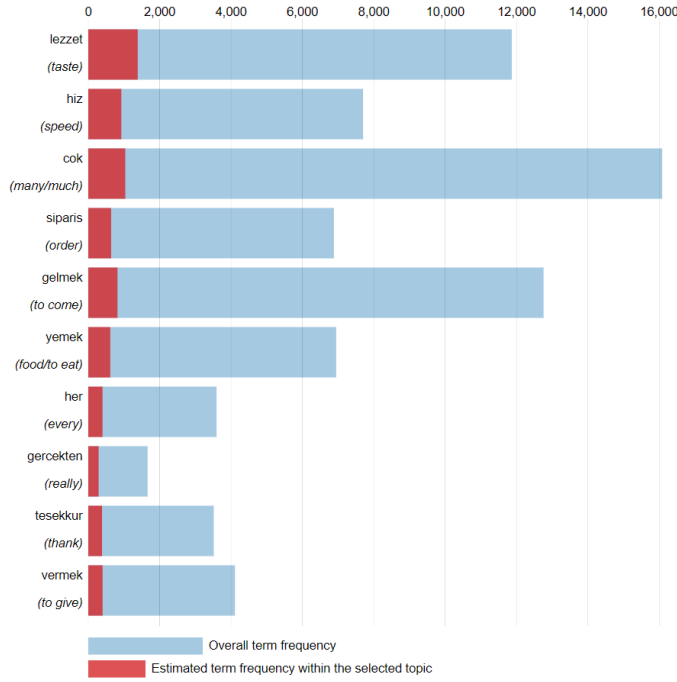


Figure 17: Top 10 most relevant terms of Topic 2 (from pita restaurant reviews)

For both types of tests, it should be noted that the Bonferroni correction is a harsh countermeasure which can increase the number of false negatives [76].

Table 21: Pizza and pita restaurants' menu term existences compared with two education clusters using Fisher's exact test. P-values are Bonferroni-corrected. An odds ratio higher than 1 indicates that the item is more common for the less educated districts' cluster while an odds ratio lower than 1 indicates that the item is more common for the more educated districts' cluster.

Food	Item	Odds ratio	P
Pizza	KÖFTE (meatball)	13.64	.001
	ROKA (rocket)	0.07	.008
	KABURGA (rib)	0	.028
Pita	PATATES KIZARTMASI (French fries)	0	.048

CHAPTER 4

DISCUSSION

In this chapter, firstly, methodologies and findings of the previous chapter are summarized and discussed further. Later, possible uses of these methodologies and findings are mentioned.

4.1 Methodologies and Results

4.1.1 Menu Statistics and Socioeconomic Status

RQ1 Is there a relationship between a restaurant's menu curation characteristics and its district's demographic characteristics?

It is hypothesized that districts with higher socioeconomic status would have more diverse and elaborate menus or vice versa, reflecting the districts' residents and their expectations. Through applying Pearson correlation, a moderate positive correlation between food-specific menu-item price medians and university graduate rates is found for pita restaurants and pizza restaurants that are not big chain branches. However, no correlation between menu elaboration/diversity and education is found. This suggests that either different districts do not have significantly distinguishable format-related expectations from menus or restaurants do not care. However, average description length is later found to be useful for pizza restaurants to predict the district itself.

4.1.2 District Prediction

RQ2 Is it possible to predict a restaurant's district by looking at its menu contents and/or statistics?

It is hypothesized that, due to cultural or economic reasons, restaurants in certain districts might typically have certain ingredients or side dishes that may distinguish them from other districts. Random forest classifiers are trained with menu and statistical features. Random forest models' scores suggest that it is possible to predict pizza and pita restaurants' districts to a limited extent. The confusion matrices suggest there is a visible bias towards Çankaya, the district that dominates Yemeksepeti. It is found that statistical features are more important than content-based features to predict a restaurant's district. While the price median is one of the most prominent features for both type of restaurants, it is less so for pizza restaurants, although pizza price medians tend to vary more. This can be interpreted in different ways. One possible explanation is that pita has a higher value for money, and pizza is closer

to be a pleasure product, which explains its higher price. The fact that pizza restaurants serve different sizes may have an effect on this as well, but it would be assumed that the number of different types of pizzas per size would be roughly the same for a given pizza restaurant. It may be related to the fact that "medium size" is not standardized, which can affect the usability of price medians as well.

Looking at the menu content does not provide much benefit over the statistical features for both pizza and pita restaurants. While this is expected for pita restaurants, this is surprising for pizza restaurants since pizza menus are naturally rich and diverse compared to pita restaurants. Besides, there are certain ingredients found to only exist in a single district. One of them is pork meat, only served in a few restaurants in Çankaya. Pork meat, as an imported product, being more expensive than beef and not being eaten by Muslims, is could be a good indicator of high education rates and socioeconomic status. The existence of pork meat also coincides with the foreign populations since international company offices, embassies, and universities (such as Bilkent University) that have many foreign workers and students dominantly reside in Çankaya. The reason such ingredients are not helpful is possibly because they exist in only a few restaurants that makes it hard for them to give enough information. This suggests that while pizza restaurants have more variety compared to pita restaurants, they still mostly use the same ingredients, which is not as helpful to predict a restaurant's district. The fact that existence vectors are preferred over occurrence vectors may also have an effect on this, because the ingredient difference may be more about the frequency rather than existence.

4.1.3 Socioeconomic Prediction

RQ3 Is it possible to predict a restaurants' surrounding area's socioeconomic status by looking at a restaurant's user reviews?

It is hypothesized that areas with residents from higher socioeconomic layers may have a different language or different expectations/emphasis on restaurants' features/problems that can be extracted from their reviews. Random forest regressors are trained with comment topic, menu, and statistical features. For pizza restaurants, it is found that user reviews are not that as helpful as they are for pita restaurants. Comment topics that are most helpful are mostly about positive remarks on the speed, service, hotness of the food, courier, and the abundance of ingredients in the pizza. The best results are obtained with menu and statistical data combined. Interestingly, the most helpful feature in regression is the percentage of user reviews responded by the restaurants. While at first it can be thought to be about higher status neighborhoods having restaurants that provide a better service and care more about the customers (which can still be true), smaller districts (in terms of both population and restaurants) having lower square meter prices but high response percentages complicates its interpretation. Perhaps, this non-monotonic behavior of response percentages can be explained with lower sales volumes and/or lower percentages of writing a review in smaller districts, which may be influencing those restaurants' owners to care more about what their customers say. In comparison, a restaurant from Çankaya with high sales volumes seems to be less likely to find time for all those reviews or suffer from the lack of interaction. Another issue is Keçiören as it features a negative correlation. As shown in Table 2, being the second most populated district yet having an education level that is right in the middle of the developed districts and the less developed districts, Keçiören indeed features unique characteristics. This is also in line with the fact that Boolean Keçiören district indicator is deemed helpful with both restaurant types. Sales volumes may be behind of this as well. While Keçiören is slightly less crowded than Çankaya, most of its population is densely populated in a relatively small area meanwhile Çankaya

is spread over the lower half of the inner city, shown in Figure 18. This may be causing restaurants in Keçiören to serve more households in a given radius, which can inversely affect their responsiveness. Perhaps, looking at the number of comments (as a proxy of number of orders) along with these values can provide more insight. Even when we only remove Keçiören, it can be generalized that higher status neighborhoods have higher response percentages within the district.

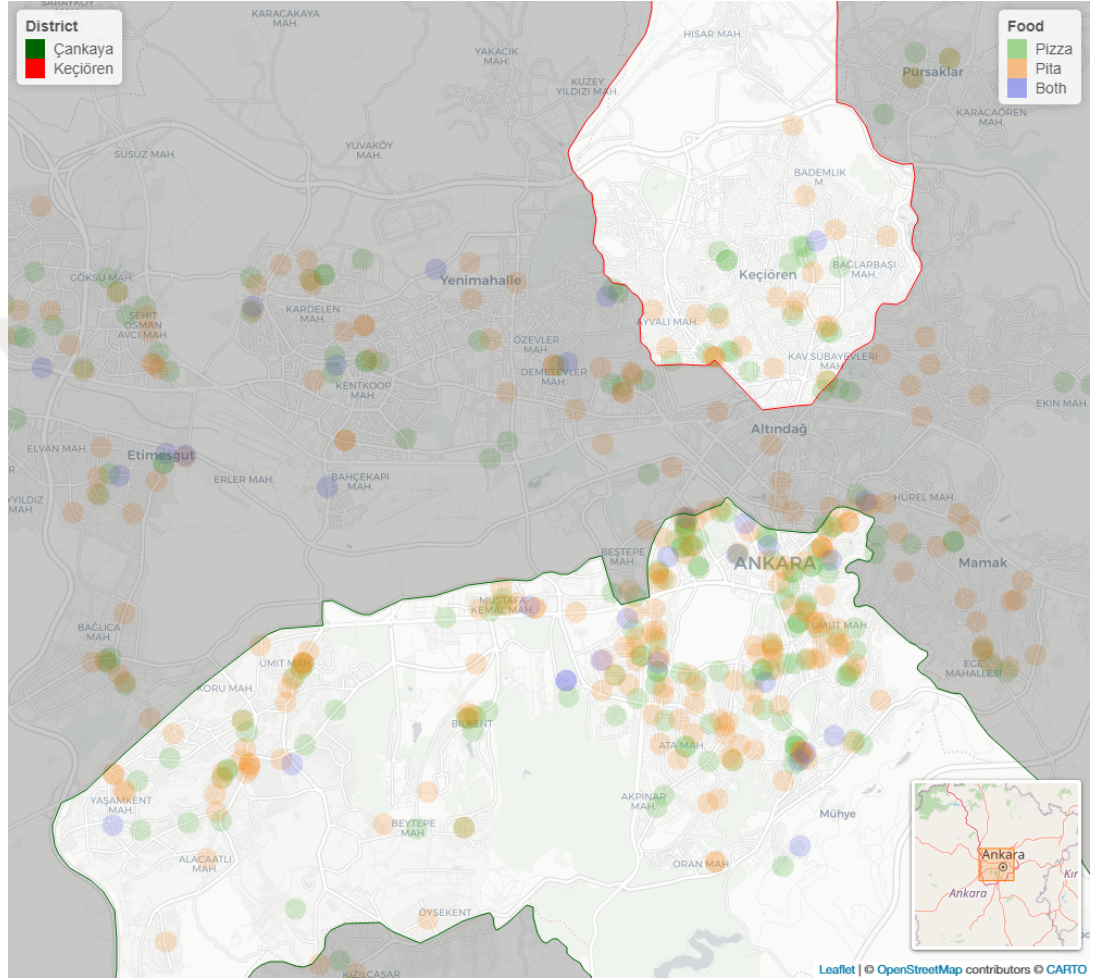


Figure 18: Keçiören's (with a red border) population is denser and restaurants are fewer compared Çankaya (with a green border)

According to the random forest models, the existence of "köfte" (meatball) indicates lower square meter prices. Since ground meat can be mixed with cheaper animal parts, it may be a cheaper way to eat meat for lower-income households, which may explain this relationship. Interestingly, while "roka" (rocket) is a relatively cheap ingredient and a t-test shows 204 pizza items that have rocket are significantly cheaper than 4470 pizza items that do not have rocket, random forest trees indicate that the existence of it leads to higher square meter prices. This might be related to the fact that rocket is also commonly used in pizzas that have beef steak, fillet steak, and ribs. Since what it accompanies varies, perhaps the models fail to notice these actually more expensive ingredients. Therefore, rocket might be just a reflection of more expensive ingredients in this context. Later through Fisher's exact test, it is found that the existence of "kaburga" (rib, short ribs), along with rocket, statistically differs

between two district clusters that are formed according to their education rates, which is in line with this possibility. However, this might be related to local preferences or other causes as well, so further analyses that scrutinize the relationship between rocket and ribs (or other relatively expensive ingredients) are required. Just like rocket, meatball is also found to be useful with these district clusters as well. Similar to how it indicates lower square meter prices, the existence of meatball also indicates lower education rates.

For pita restaurants, user comment topics are actually useful along with statistical features, to some extent. The topics that are most helpful are also more varied and interesting compared to pizza restaurants. Restaurants that have higher likelihoods for having positive comments about side dishes, desserts, and free offerings (Topic 14) seemingly have higher priced realty listings nearby. It should be noted that one of the relevant terms in Topic 14 is "taze" (fresh), which suggests these restaurants serve fresher side dishes, which is indeed a telltale sign of restaurant quality in pita restaurants. Perhaps, neighborhoods with lower income levels or their restaurants are more focused on the main dish while there is more focus on side dishes and serving them fresh in neighborhoods with higher income levels.

Another useful topic (Topic 70) is about people who ordered spicy lahmacun (a pita-like dish) and did not like something about it. It seems like neighborhoods with lower level income levels are not as proficient when it comes to serve spicy lahmacun. However, it may also indicate that neighborhoods with lower income levels prefer their lahmacuns to be spicy, leading to more negative comments about lahmacun in general. As shown in Figure 13, while lahmacun has a high overall frequency in the corpus, its spiciness context only exists in this topic. Either way, the existence of "acı" (spicy), "biber" (pepper), and "isot" (isot pepper) suggests this topic is about spiciness. The fact that the Arabic culture especially has an effect on Southeastern Turkey and it has the spiciest cuisine in the country [77] is also in line with this possibility since these districts' residents have education demographics that are more similar with Southeastern Turkey [78]. Households with lower income levels may prefer lahmacun due to its relative cheap price while their traditional and perhaps eastern side prefers lahmacun spicy, and therefore they may specifically comment on spicy lahmacuns.

The hardest topic to interpret is Topic 88, the one that has comments on "çiğ köfte" and "ıçlı köfte" which are side dishes as well. Interestingly, while the topic does not have an apparent negative sentiment, it has a negative correlation with the square meter prices. Considering Topic 88 is the only topic that mentions "çiğ köfte" and "ıçlı köfte" in its top 10 terms and the other terms have low probabilities, it can be said that this topic represents comments that are about these side dishes in general. Therefore, it brings the question if households with lower income levels consume more of these side dishes. The biggest difference of these side dishes from the ones mentioned in the other side dish topic is that these side dishes are mostly deliberately purchased in a meal while the others usually come with the main dish whether the customers want them or not. Another difference is that these two side dishes require more time and effort compared to the other ones. The fact that these two side dishes or their slightly different versions (such as "kibbeh" and "kibbeh nayyeh") also exist in other eastern cultures is in line with the idea of pita restaurants from lower income neighborhoods being more traditional and having eastern ties. This may be also an indicator of internal migrations. For example, these lower income neighborhoods may have more residents that are originally from Southeastern Turkey. Perhaps, more traditional restaurants have a dedicated labor force for preparing these side dishes while the others either do not serve them or their customers do not comment on these side dishes. This topic requires further analyses to understand why these side dishes are more commonly reviewed in lower-income neighborhoods.

Statistical features that are most helpful for pita restaurants are different than pizza restaurants. For pita restaurants, the important statistics are score geometric mean (geometric mean of a restaurant's scores in speed, service, and taste) and food-specific menu item price median. This suggests that restaurants from higher income level neighborhoods are more expensive yet more successful. While using menu items does not yield better square meter price predictions with pita restaurants, it is found that the existence of "patates kızartması" (French fries) in the restaurant menu significantly differs between two educational district clusters. According to the Fisher's exact test result, French fries are more commonly served as a side dish in districts that have higher education rates. Since districts that have higher education rates are expected to have more multiculturalism and more variety, this is not surprising. French fries can be considered non-traditional and therefore possibly not as popular in more traditional districts that also happen to be less educated. Just like it was prominent in the random forest models, the Kruskal-Wallis test and cluster means show that Topic 14 (a topic that includes terms with positive sentiments on side dishes) also significantly differs between two district clusters with a higher likelihood in more educated districts. These similarities strengthen the visible relationship between realty prices and education levels.

For both pizza and pita restaurants, Boolean Keçiören district indicator is found to be useful to predict the socioeconomic level. While Çankaya is found to be even more useful for pita restaurants, it does not give as much information for pizza restaurants. Correlations and feature importances show that, for some districts, knowing the district helps with socioeconomic prediction.

In overall, it can be argued that the neighborhood-level prediction using restaurant data is not sufficiently successful. However, according to a study that analyzes realty prices in Ankara, there are many unrelated factors such as the number of bathrooms, number of floors, building age, and heating system that affect the square meter prices [79]. Therefore, obtaining R^2 values of 0.205 and 0.181 by simply using restaurant data is still notable. After all, it is found that even district-level education rates are at most moderately correlated with the square meter prices. It is also found that once districts are clustered according to their education rates, it is possible to find certain features that show a significantly different distribution among these clusters. Statistical tests with district clusters also show that menu features are more useful for pizza restaurants while comment topics are more useful for pita restaurants, which echoes a previous remark about how pizza and pita restaurants differ among each other. In general, findings of different tests and approaches throughout the study have similar or at least non-contradictory indications.

4.1.4 Pita versus Pizza

Unsurprisingly, pita restaurants outnumber pizza restaurants. However, the fact that pizza restaurants have penetrated more districts on Yemeksepeti might indicate that pizza restaurants are less traditional and more open to online food ordering systems, at least in rural areas. For Pursaklar, the only district that has more pizza restaurants than it has pita restaurants, Google Maps indicates that it actually has much more pita restaurants, which is in line with this possible explanation. Nearby realty listing frequency and price distributions in each district for pizza and pita restaurants are interesting as well. It is possible to see that pita restaurants are not only more common but also located in more varying neighborhoods.

Pizza restaurants mostly featuring more variance in their statistics such as food-specific menu item price medians or ingredient counts indicates a more varying and less standardized experience with

pizza restaurants. Price medians not having the same importance for pizza restaurants for both district prediction and socioeconomic status prediction also suggests that pizza may be more about pleasure and less about price-performance ratio. It seems like pita restaurants not only provide a more monotonous service but they also do not try to differentiate themselves based on their menu item descriptions and use of photos (or lack thereof). Considering that many of the pita restaurants have "Aspava" in their name (a name originally being used by a single restaurant and refers to a dining culture rich with side dishes [80]) and pita restaurants come from Middle Eastern traditions that favor collectivism over individualism, them not having as much brand identity is not surprising. However, their unmotivatedness to present photographs or more detailed explanations might be partially explained with the fact that pitas have significantly less number of ingredients (mostly one or two) and the ingredients are already given in the product's name. Therefore, the ingredients and the end products are mostly exactly the same for a specific pita dish between two different restaurants, which makes them require less descriptions and less photos. This difference is also echoed in socioeconomic prediction. Pizza restaurants are distinguished with their contents among each other while looking at pita restaurants' user reviews gives more information than looking at their menus. It looks like side dishes that come with the main dish in pita restaurants are a big deal for their customers, but it can be only seen by looking at their reviews, possibly because most restaurants technically provide the same side dishes.

There are certain similarities between pizza and pita restaurants. Both of them seemingly set their prices according to the socioeconomic status of their districts. Both of them also have discriminant statistical features for both district and socioeconomic prediction. While pita restaurants from higher income level neighborhoods have higher prices and scores, pizza restaurants from higher income level neighborhoods responding to more user reviews suggests there is more than just the product quality for pizza customers.

4.1.5 Predicting Place versus Attributes

As explained in Section 3.2.4, administrative (district) borders do not necessarily apply on how people settle. Analyses and models that work with district clusters or neighborhood-level data work better than working with district classes. Districts can be too big to generalize and too similar to differentiate, at least in Ankara. Therefore, at the cost of prediction granularity, working with attributes or attribute-based clusters rather than administrative entities is suggested.

4.1.6 Related Work

The relationships found between ingredient-square meter prices and ingredient-education rates, along with the correlations between district-square meter prices and education rates-square meter prices, are in line with the results of previous studies in terms of the relationship between rental prices, income, education, and socioeconomic status [20–23].

It is found that districts in which French fries (a non-traditional side dish) are served with pitas more also have higher education levels. Similarly, although it has a relatively low frequency in overall, soy sauce on pizza is found to be more frequent in neighborhoods with higher realty prices. Considering pizza's history, soy sauce being linked to Asian cuisine [32], and a celebrity's tweet about combining

soy sauce and pizza caused a public debate in 2017 [81], soy sauce can be considered rather non-traditional for pizzas as well. Openness personality trait is linked to having a less traditional diet in [6] while another study states that Openness is linked to higher education levels [82]. The findings of this study are in line with these previous studies as well.

In [30], it is stated that more complex terminologies used in menus can cause an increase in perceived quality and pricing expectations. While price medians are found to correlate with education rates, descriptions lengths are not found to have a significant correlation with them. However, this study does not attempt to directly adapt this study, and term complexity does not necessarily require lengthiness. Setting and cuisine may have an effect on this as well.

4.2 Implications of Use

Apart from its research contributions, this study and especially its methodologies have multiple implications of use for both businesses and customers.

On Yemeksepeti, users can decide on trying a new restaurant by looking at category-specific (speed, service, and taste) rates or other users' comments. Rates can only give an overall look and learning about certain dishes' quality at best requires some digging through the user reviews. Yemeksepeti could feature an aspect-based highlights for restaurants or it could recommend restaurants based on the user's expectations (extracted from previous reviews) and restaurants' specific service details (extracted from other users' reviews). For example, if a pita customer does not like the side dishes of a certain restaurant and decides to write about it, next time Yemeksepeti could recommend restaurants that seemingly perform well with side dishes according to user reviews. While an increased informedness might cause certain restaurants to lose some of their income and leave the system, it can be also used to promote certain dishes of restaurants that have low ratings overall. That way, restaurants can reach their niche audience and the system can alleviate its possibly harmful effects, at least for restaurants that have something that is marketable. This can also affect restaurant to optimize their menus, specialize, and cut their business costs.

Apart from restaurant and dish recommendations, there is also a remarkable analytics potential for online food ordering systems. They can not only use these methods to gain insight but also they can inform restaurants in order to increase their performance that could benefit the system as well. For example, restaurants can be notified of which topics their customers talk about lately. Through adding a sentiment or rating layer to BTM, restaurants can also see if there is a recently rising problem in their service without the need of reading each individual comment. Restaurants can be also notified of the strengths and weaknesses of nearby competitors based on customer reviews, which can influence them to improve their services without stalking each competitor's user reviews. They can be also recommended to add certain menu items or ingredients based on their nearby competitors' menu contents, supported with the sales data, or provide a baseline menu to help new businesses with their menu design. These potential analytics features can be also put behind a paywall to help online food ordering systems directly drive even more revenue from their partner restaurants.



CHAPTER 5

CONCLUSION

This study attempts to explain how pita and pizza restaurants are different from each other, how their menus and reviews differ based on location, and how these differences can be used to distinguish restaurants' locations and socioeconomic status. To do so, restaurant data on Yemeksepeti are scraped and datasets are enriched with other collected or appropriated datasets. NLP techniques are applied to restaurant menus and user reviews. Restaurants menus are vectorized and topic modeling is applied to restaurant reviews to be used with random forest models. Education rates and realty listing prices are used as proxies for socioeconomic status.

Menu curation of restaurants are found not to be related to the restaurants' district education rates while average description length is found to be useful for predicting a pizza restaurant's district. The findings may indicate a relationship between a restaurant's prices and the neighborhood's income level. Restaurants are found to be useful to predict a neighborhood's realty listing prices, but only to a limited extent. It is found that menu contents and restaurants' responses to reviews are more useful for pizza restaurants while menu prices, restaurant scores, and user reviews are more useful for pita restaurants to predict a restaurant's location or location characteristics. The Kruskal-Wallis and Fisher's exact tests with district education clusters also yield results that are in line with these findings, indicating significant differences among different education clusters with some ingredients for pizza restaurants and with comments related to side dishes for pita restaurants. The findings suggest that pita restaurants from higher-income neighborhoods get more positive comments about their side dishes, desserts, and free offerings that they serve with the main dish. Meanwhile, lower-income neighborhoods seemingly make more comments on "çiğ köfte" (kibbeh nayyeh) and "içli köfte" (kibbeh) or ordering something spicy. The findings also suggest that, compared to pizza restaurants, pita restaurants are menu-wise less but neighborhood-wise more diverse. However, it is also indicated that serving French fries with pita is significantly more common in more educated districts. Pizza restaurants have more diverse menus and the existence of certain ingredients such as rocket and meatball is significantly different between education-based district clusters, respectively indicating higher and lower education rates. It is also indicated that, within the same district, pizza restaurants that are located in more expensive locations generally respond more to user reviews.

While predictive models proposed in this study do not yield impressive results, this study is significant, because there has been no other study that scrutinizes restaurant menus and reviews in the context of cuisine, location, and socioeconomics. This study's contributions are given in Section 1.2, the main one being proposing of a set of methods to enrich an online food ordering system's data with socioeconomic data and do an analysis. It also provides Turkish ingredient and side dish lexicons specifically

for pizza and pita restaurants, which can be used for restaurant/recipe analysis. Furthermore, this study's findings also have some valuable implications and further research questions for future studies.

5.1 Limitations and Assumptions

Limitations and some related assumptions of this study are explained in this section.

5.1.1 Data Collection

It is not possible to find a reliable number of restaurants for some (mostly peripheral) districts of Ankara. For some, it is not even possible to find any restaurants.

Yemeksepeti has no public API or a structurally simple website. This requires using a WebDriver-based scraping approach, simulating user interactions, such as scrolling and clicking, and pausing between requests to avoid being temporarily denied of service. Combined with the previously explained not-so-predictable availability of the restaurants, it causes some restaurants to not be scraped. However, given the amount of trials for each restaurant, it can be assumed that those restaurants are not active anyway, and including them could also increase the noise.

Reviewers' names are not public or even uniquely identifying, which could otherwise have been used for further analysis. Reviews have relative dates (such as "X days/months ago"), and they reflect the moment the reviews are written, not the moment orders are placed. Only the reviews that are written in the last six months can be seen. It is also assumed that all of the reviewers live in an area close to the corresponding restaurant. However, it is possible that a considerable amount of them were written by people who rather work in an area close to the corresponding restaurant. This may have caused a significant noise especially for districts like Çankaya and Yenimahalle. Still, it is believed that people who work in an area interact, affect, and be affected by the area just like its residents. So, it is assumed that it does not interfere the study enough to be a significant problem.

How the restaurant scores are calculated or manipulated are not known. For example, it is not known if the given scores are normalized based on the scoring habits of individual users or they are calculated only by aggregating the last six months' reviews.

In general, starting with limiting the study to Yemeksepeti, varying levels of bias are introduced at each step. Yemeksepeti may not correctly represent restaurants or customers in Ankara. It is assumed that, due to its popularity and market domination, limiting the study to data provided by Yemeksepeti does not introduce a significant bias. It is stated without much further explanation that the age of the most active users on Yemeksepeti is 27 [83] while the average age of Ankara calculated using the explicitly given neighborhoods and the age group lower boundaries from PD is 31. Assuming the distribution is similar to the actual population distribution and the userbase of Ankara reflects the general userbase of Yemeksepeti, this expectedly suggests that Yemeksepeti has a slightly younger userbase than the actual population. The exact age distribution of Yemeksepeti users is unknown.

Çankaya district, having at least 11 university campuses [84], likely has a significant amount of students that are officially not Çankaya residents, which can skew Çankaya's Yemeksepeti userbase. However, since the dataset is collected in October 2019, it contains the last six months' reviews which

include summer. This may have alleviated this bias to a certain degree. Some other unforeseen biases might also exist in the review dataset based on its time frame. Having full knowledge of the users, their addresses, and order details would remarkably increase the potential and success of analyses concerning user reviews. It is assumed that a person reviewing a restaurant in a location shares some common characteristics with the location at which they happen to be at that specific time.

The user review dataset is inherently biased because they may not correctly represent the users' views or demographics. The gamification and point system on Yemeksepeti is assumed to help capture the views of the userbase by encouraging the users to review their order. Sampling the review dataset might reduce the fairness even more, but a filtering is not applied to them other than removing the short ones. It is assumed that the reviews represent the userbase. Creating promotions can also affect the amount of comments or their sentiments. Comment topics mentioning promotions (limited-time "joker" promotions) are also observed, but they did not show a significance as shown in Table 24 and Table 25.

5.1.2 Data Analysis and Predictions

There are some accidental noises introduced in MD. As explained, menu categories are filtered to remove duplicate sizes of each menu item. However, it is later found out that some restaurants did not follow the expected size order and put the medium-size category first. While this only affected three pizza restaurants that are used in analyses, pizza prices are also slightly skewed due to certain restaurants not serving every menu item in every possible size and causing an item to be retrieved from a larger size category if it does not exist in the small-size category. Using different datasets for price analysis and menu analysis would be better to address this issue. It is also found out that some pita restaurants (two for the last research question and three for the rest) had duplicate menu items. This did not affect menu vectorization since the existence vectors are used, but it slightly affected statistics such as number of items or price median.

Menu item descriptions, provided by restaurants themselves, do not have a standard. Some simply leave it empty, some give the bare minimum amount of information, some give detailed information and even use it as a medium for further communication/branding, and some even provide the English translation. Fortunately, bilingual descriptions do not affect the study since common ingredient names are almost always paired with a language-specific word which helps to distinguish them using an n-gram solution or Turkish suffixes (an example would be "mozzarella cheese/mozzarella peyniri" or "cheddar/cheddarlı"). However, there is a slight duplication of words related to primary ingredients. For example, a menu item having the word "pineapple" in both the item name and item description can cause the word to be counted two times that can affect the analysis if the analysis uses raw frequencies. It is assumed that the menu item descriptions are correct. However, based on some user reviews, some of the menu items' side dishes can slightly differ from their descriptions. While lemmatization is applied to menu description words, they are not morphologically analyzed. Part of speech tagging could help to identify when a word is actually an ingredient and when it is not, which can be used to detect more ingredients with more precision.

Each considerably different cuisine requires many specific words to be added to the lexicon. Meanwhile, each word in the lexicon can have a drastically different importance based on the cuisine. An important word for a cuisine can improve the performance for that cuisine while decreasing the performance for another cuisine. This context-wise fragility creates an adaptability/scalability issue. More

sophisticated approaches such as analyzing inter-dependencies between words are not implemented as it was determined to be out-of-scope for this study. Also, it is later detected that some residual terms are created due to how ingredients/side dishes are collected using the lexicons. For example, barbecue sauce is searched as "barbekü" (barbecue) instead of "barbekü sos" (barbecue sauce) since there were occurrences of it without "sos" (sauce). However, due to not having an alternative version as "barbekü sos" to merge them later, this caused occurrences of "barbekü sos" (barbecue sauce) to be separately counted as barbecue and generic sauce. Fortunately, this mistake is relatively rare and rather unimportant since it does not affect any relatively important features and its effect is almost non-existent since existence vectors are preferred over occurrence vectors. These detected errors are later fixed in the lexicon for future work. While this lexicon can provide a starting point for future work, it does not include ingredients or side dishes that do not exist in the dataset specific for this study. Even if used for similar food, an effort to obtain case-specific ingredients or side dishes to extend the lexicons for future work is suggested.

While it is possible to normalize text data to some extent, user reviews can be quite colloquial, syntax/grammar-wise delinquent, or context-specific. No additional lexicon is used while processing the user reviews, but review-specific lexicons can help reduce the noise and catch certain potentially missed patterns. For example, using ingredient or side dish lexicons with user reviews can improve the use of collocations (like "çiğ köfte") and lemmatization as some minor lemmatization mistakes (such as lemmatizing "salata" to "salat" as if "a" is a suffix) are observed.

It is assumed that the age distributions of districts do not significantly affect their education rates. Therefore, education data is used as is. However, normalizing each district using their population that is for example at least 30 years old could improve the educatedness accuracy.

Since most restaurants' locations are strategically designated, restaurants form clusters in shopping malls or certain regions, so there are blind spots about which we cannot make rental price predictions. More importantly, each restaurant is treated as an independent point, but restaurants are very likely to affect other restaurants in close proximity at least in terms of pricing decisions, which can cause skewness in certain locations. This can both increase or decrease the prices. It is also possible that this spatial correlation affects restaurants' menus/ingredients, working hours, or other qualities. It may be also harder to detect it as not all of the restaurants are on Yemeksepeti. This spatial correlation can be attempted to be tackled in the future with further analyses that take into account this complication.

Certain variables in the study are not perfectly optimized. Feature selection and hyperparameter optimization in random forests for socioeconomic prediction are handled in a semi-automatic way due to the excessive computational complexity, which means it might be possible to increase the models' success. Feature selection is not applied for district prediction, and its application may increase its success. Realty listing search radius for restaurants is made with an average radius that is obtained by manual observation. It might be more accurate to use restaurant-specific search distances through pinpointing the delivery regions for each restaurant. Biterm search window is set to be 4, which is not optimized. The comments used with BTM have a mean sentence length of 7.6 and a median sentence length of 6 while losing 20% of their terms in average after the filtering process. This indicates that the window size is slightly shorter than the average sentence length. Optimizing the window size can yield better results. While [57] mentions using "a small, fixed-size window" to obtain co-occurrences, it also suggests that short texts can be the window themselves. However, considering the negations are not handled and the comments can have multiple aspects, this may also cause the topics to get convoluted. Biterms are extracted after the terms are pre-processed and recombined, which probably

helped with the sparsity while causing some words that would be outside the window's range to be paired. However, this should not have caused a significant problem as similar studies have worked with documents that are formed by whole sentences or even comments. BTM topic count are not optimized as well. While the obtained comment topics are highly explainable, optimizing the number of topics could increase the success of the features and predictions at the cost having more generalized topics.

5.2 Future Work

Considering the unused parts of the data, other existing methodologies, and the new questions brought up in the previous chapter, this study can be extended. Scaling up the scope in terms of cuisine and region is likely to provide more insight. For example, as mentioned in Chapter 2, it seems like the relationship between a healthier diet and higher education levels is debatable. To explore location-based healthy dietary habits and opportunities, restaurants from many different cuisines can be analyzed. Instead of districts of the same city, comparing two different cities or even countries can lead to other interesting findings as well.

Due to the amount of different data collected, it is possible to find numerous new ways to analyze it. With the most methodologies, pizza chain branches are simply removed from pizza restaurants. However, comparing chain branches with non-chain restaurants can also produce interesting results for pizza restaurants. Ingredients are handled at the restaurant level, but looking at how ingredients are paired can also provide additional insight about pizza restaurants and perhaps regions. Instead of directly comparing locations with all of their restaurants, comparing restaurants with other restaurants at the same caliber (expensiveness) can also provide insight. This way, how similar restaurants from different locations differ from each other can be studied as well. Since major university campuses that restaurants serve are explicitly provided by Yemeksepeti and collected in the dataset, a university-oriented approach could be also adopted, which may help with controlling the neighborhood-level population variations. While prominent comment topics do not suggest a significant and explicit difference in language and politeness, using a politeness lexicon and isolating politeness from the customer satisfaction/sentiment may still reveal some differences. Statistics such as sentence and comment length can be also scrutinized as well. Topic modeling is only used with user reviews. However, restaurant responses are collected and processed just like user reviews. Modeling restaurant responses can also produce interesting results especially since restaurant response percentage is deemed useful to distinguish pizza restaurants' neighborhood characteristics. Certain restaurant characteristics such as working hours, delivery locations, and minimum order prices are collected but not analyzed. Especially working hours can give some information about the neighborhood and its functional characteristics. While clustering is used with statistical tests, they are not used with predictions. Predictions can be made with location clusters instead of directly working with locations.

It is also possible to improve certain methods used in this study, some of which are mentioned in the previous section. Firstly, some menu filtering mistakes made during the data manipulation phase shall be fixed in future studies. Multiple and more sophisticated lexicons can be used for restaurant menus. For example, while there is a general "peynir" (cheese) term, it only reflects cheese that are not specified (mozzarella, feta, etc.). Generalized and overlapping ingredient or side dish terms that count the total occurrences for certain similar terms can reduce the sparsity and be useful along with the detailed ones. After all, sometimes it can be more important to see if any kind of an ingredient is

used instead of looking at its sub-classes. A saliency threshold can be applied after obtaining the most relevant terms for each topic to improve topic interpretability, which would filter out the insignificant terms and possibly insignificant topics [70].

While PCA is also briefly tried before training models without success, dimension reduction methods can be still useful with certain types of features. However, a significant portion of the data is not continuous. A more sophisticated topic modeling approach (like the ones mentioned in Chapter 2) can be used. For example, while topics frequently and inherently have a sentiment, adding a sentiment or rating layer and aspect-based topic modeling can be useful. It is likely to be useful to merge synonym terms in order to reduce noise and obtain even better topics. For example, in Topic 15 for pizza restaurants, both "servis" and "hizmet" terms refer to the same thing (service), which creates redundancy. While adding a user layer is not possible due to reviews' anonymity, adding certain restaurant-specific layers can have interesting outcomes as well.

Finally, Chapter 4 explores possible socioeconomic explanations for this study's findings and asks new questions. For example, why do lower-income neighborhoods make more comments on certain side dishes? Do lower-income neighborhoods prefer more spicy foods? Why do higher-income neighborhoods have pizza restaurants that are more responsive to the reviews and why is Keçiören an exception? In the future, these questions can be tackled to better understand district/neighborhood/cuisine-specific characteristics and differences.

REFERENCES

- [1] J. Washington, "The magical power of cannibalism," *Crossroads: An Interdisciplinary Journal for the Study of History, Philosophy, Religion and Classics*, vol. 6, no. 1, pp. 46–57, 2012.
- [2] M. Fabinyi, "Historical, cultural and social perspectives on luxury seafood consumption in China," *Environmental Conservation*, vol. 39, no. 1, p. 83–92, 2012.
- [3] L. M. Long, *Culinary Tourism*. Springer, 2013.
- [4] L. R. Goldberg and L. A. Strycker, "Personality traits and eating habits: The assessment of food preferences in a large community sample," *Personality and Individual Differences*, vol. 32, no. 1, pp. 49–65, 2002.
- [5] R. Möttus, A. Realo, J. Allik, I. Deary, T. Esko, and A. Metspalu, "Personality traits and eating habits in a large sample of Estonians," *Health Psychology: Official Journal of the Division of Health Psychology, American Psychological Association*, vol. 31 6, pp. 806–14, 2012.
- [6] C. S. Kessler, S. Holler, S. Joy, A. Dhruva, A. Michalsen, G. Dobos, and H. Cramer, "Personality profiles, values and empathy: Differences between lacto-ovo-vegetarians and vegans," *Complementary Medicine Research*, vol. 23, no. 2, pp. 95–102, 2016.
- [7] M. Tasviri, S. A. H. Golpayegani, and H. Ghavamipoor, "Presenting a model based on social network analysis in order to offer a diet to users proper to their mood," in *2017 3th International Conference on Web Research (ICWR)*, pp. 133–139, IEEE, 2017.
- [8] C. Hirschberg, A. Rajko, T. Schumacher, and M. Wrulich, "The changing market for food delivery," 2016. [Online]. Available: <http://dln.jaipuria.ac.in:8080/jspui/bitstream/123456789/2874/1/The-changing-market-for-food-delivery.pdf>. [Accessed: Sep. 9, 2020].
- [9] A. Ray, A. Dhir, P. K. Bala, and P. Kaur, "Why do people use food delivery apps (FDA)? A uses and gratification theory perspective," *Journal of Retailing and Consumer Services*, vol. 51, pp. 221 – 230, 2019.
- [10] Yemek Sepeti, "Yemek Sepeti - online yemek siparişi ve paket servis." [Online]. Available: <https://www.yemeksepeti.com>. [Accessed: Aug. 17, 2020].
- [11] Novinite, "Turkey's online food ordering portal Yemeksepeti to expand to Bulgaria, Romania." [Online]. <https://www.novinite.com/articles/168647/Turkey%E2%80%99s+Online+Food+Ordering+Portal+Yemeksepeti+to+Expand+to+Bulgaria>. [Accessed: Aug. 17, 2020].
- [12] Yemeksepeti, "2019 bizim için böyle geçti!." [Online]. Available: <https://twitter.com/yemeksepeti/status/1211943885712310273>. [Accessed: Aug. 17, 2020].

- [13] H. Ögütçü, "Yemeksepeti, 2019 yılında verilen 340 milyon porsiyon siparişin analizini paylaştı," *egirisim*, December 2019. [Online]. Available: <https://egirisim.com/2019/12/27/yemeksepeti-2019-yilinda-verilen-340-milyon-porsiyon-siparisin-analizini-paylasti>. [Accessed: Aug. 17, 2020].
- [14] Just Eat, "Investor presentation." [Online]. Available: https://www.just-eat.com/download_file/831/197. [Accessed: Aug. 17, 2020].
- [15] S. Kemp, "Digital 2020: Global digital yearbook," *DataReportal*, January 2020. [Online]. Available: <https://datareportal.com/reports/digital-2020-global-digital-yearbook>. [Accessed: Aug. 17, 2020].
- [16] S. Sakarya and N. Soyer, "Cultural differences in online shopping behavior: Turkey and the United Kingdom," *International Journal of Electronic Commerce Studies*, vol. 4, no. 2, pp. 213–238, 2014.
- [17] F. S. Chapin, "A quantitative scale for rating the home and social environment of middle class families in an urban community: A first approximation to the measurement of socio-economic status," *Journal of Educational Psychology*, vol. 19, no. 2, p. 99, 1928.
- [18] D. Donovan, R. Huyser, P. Kearney, J. Olsen, and D. E. Schooley, "Levels of educational performance and related factors in Michigan," *The Fifth Report of the 1970-71 Michigan Educational Assessment Program*, 1972.
- [19] S. Acar, M. C. Meydan, L. Bilen Kazancık, and M. Işık, *İllerin ve Bölgelerin Sosyo-ekonomik Gelişmişlik Sıralaması Araştırması: SEGE-2017*. Sanayi ve Teknoloji Bakanlığı, Kalkınma Ajansları Genel Müdürlüğü, Dec 2019. [Online]. Available: https://www.bebka.org.tr/admin/datas/sayfas/89/sege-2017_1581687211.pdf. [Accessed: Sep. 23, 2020].
- [20] S. Kalaycıoğlu, K. Çelik, Ü. Çelen, and S. Türkyılmaz, "Temsili bir örnekleme sosyo-ekonomik statü (SES) ölçüm aracı geliştirilmesi: Ankara kent merkezi örneği," *Sosyoloji Araştırmaları Dergisi*, vol. 13, no. 1, pp. 182–220, 2010.
- [21] O. Işık and E. Ataç, "Yoksulluğa dair: Bildiklerimiz, az bildiklerimiz, bilmediklerimiz," *Birikim*, 268, vol. 269, pp. 66–86, 2011.
- [22] Ö. O. Kılıç, M. A. Akyol, O. Işık, B. Günel Kılıç, A. U. Aydınöğlu, E. Surer, H. Ş. Düzgün, S. Kalaycıoğlu, and T. Taşkaya-Temizel, *The Use of Big Mobile Data to Gain Multilayered Insights for Syrian Refugee Crisis*, pp. 347–379. Cham: Springer International Publishing, 2019.
- [23] H. Li, Y. D. Wei, and Y. Wu, "Analyzing the private rental housing market in Shanghai with open data," *Land Use Policy*, vol. 85, pp. 271–284, 2019.
- [24] J. P. Block, R. A. Scribner, and K. B. DeSalvo, "Fast food, race/ethnicity, and income: A geographic analysis," *American Journal of Preventive Medicine*, vol. 27, no. 3, pp. 211 – 217, 2004.
- [25] E. Çeltekin and M. Bozdoğan, "Turizm işletmelerinde e-ticaret: Yemeksepeti.com'da satış yapan yiyecek-içecek işletmelerinin incelenmesi," *Gaziantep University Journal of Social Sciences*, vol. 12, no. 3, 2013.
- [26] M. Çuhadar and B. Aşıroğlu, "Zincir fast-food işletmelerine yönelik çevrimiçi değerlendirmelerin analizi: Eskişehir örneği," in *20. Ulusal - 4. Uluslararası Turizm Kongresi Bildiriler Kitabı*, p. 415, 2019.

- [27] E. Armağan and Y. Eskici, “Tüketicilerin online yemek servislerine karşı tutum, davranış ve satın alma niyetleri,” *EKEV Akademi Dergisi*, p. 39–75, Mar 2019.
- [28] K. Annaraud, “Restaurant menu analysis,” *Journal of Foodservice Business Research*, vol. 10, no. 4, pp. 25–37, 2007.
- [29] K. Fakih, G. Assaker, A. G. Assaf, and R. Hallak, “Does restaurant menu information affect customer attitudes and behavioral intentions? A cross-segment empirical analysis using PLS-SEM,” *International Journal of Hospitality Management*, vol. 57, pp. 71–83, 2016.
- [30] M. McCall and A. Lynn, “The effects of restaurant menu item descriptions on perceptions of quality, price, and purchase intention,” *Journal of Foodservice Business Research*, vol. 11, no. 4, pp. 439–445, 2008.
- [31] E. L. Serrano and V. B. Jedda, “Comparison of fast-food and non-fast-food children’s menu items,” *Journal of Nutrition Education and Behavior*, vol. 41, no. 2, pp. 132–137, 2009.
- [32] W. Min, B. Bao, S. Mei, Y. Zhu, Y. Rui, and S. Jiang, “You are what you eat: Exploring rich recipe information for cross-region food analysis,” *IEEE Transactions on Multimedia*, vol. 20, no. 4, pp. 950–964, 2018.
- [33] C.-Y. Teng, Y.-R. Lin, and L. A. Adamic, “Recipe recommendation using ingredient networks,” in *Proceedings of the 4th Annual ACM Web Science Conference*, pp. 298–307, 2012.
- [34] S. Dahmann and S. Anger, “The impact of education on personality: Evidence from a German high school reform,” *IZA Discussion Paper*, 8139, 2014.
- [35] A. W. Dynesen, J. Haraldsdóttir, L. Holm, and A. Astrup, “Sociodemographic differences in dietary habits described by food frequency questions—results from Denmark,” *European journal of clinical nutrition*, vol. 57, no. 12, pp. 1586–1597, 2003.
- [36] C. Akbay, G. Y. Tiryaki, and A. Gul, “Consumer characteristics influencing fast food consumption in Turkey,” *Food Control*, vol. 18, no. 8, pp. 904–913, 2007.
- [37] R. M. Nayga Jr and O. Capps Jr, “Impact of socio-economic and demographic factors on food away from home consumption: Number of meals and type of facility,” *Journal of Restaurant & Foodservice Marketing*, vol. 1, no. 2, pp. 45–69, 1994.
- [38] G. Mihalopoulos Vasilis and P. Demoussis Michael, “Greek household consumption of food away from home (fafh): A microeconomic approach. v: 71st eaae seminar-the food consumer in the early 21 st century,” *Zaragoza*, pp. 19–20, 2001.
- [39] A. M. Angulo, J. M. Gil Roig, and J. Mur, “Spanish demand for food away from home: A panel data approach,” *Journal of agricultural economics*, pp. 289–307, 2007.
- [40] K. Glanz, M. Basil, E. Maibach, J. Goldberg, and D. Snyder, “Why Americans eat what they do: Taste, nutrition, cost, convenience, and weight control concerns as influences on food consumption,” *Journal of the American Dietetic Association*, vol. 98, no. 10, pp. 1118 – 1126, 1998.
- [41] S. Abbar, Y. Mejova, and I. Weber, “You tweet what you eat: Studying food consumption through Twitter,” in *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, CHI ’15, (New York, NY, USA), p. 3197–3206, Association for Computing Machinery, 2015.

- [42] T. H. Silva, P. O. Vaz de Melo, J. M. Almeida, M. Musolesi, and A. A. Loureiro, “A large-scale study of cultural differences using urban data about eating and drinking preferences,” *Information Systems*, vol. 72, pp. 95 – 116, 2017.
- [43] S. Blair-Goldensohn, K. Hannan, R. McDonald, T. Neylon, G. Reis, and J. Reynar, “Building a sentiment summarizer for local service reviews,” in *Proceedings of WWW-2008 Workshop on NLP in the Information Explosion Era*, 2008.
- [44] M. Hu and B. Liu, “Mining and summarizing customer reviews,” in *Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '04, (New York, NY, USA), p. 168–177, Association for Computing Machinery, 2004.
- [45] G. A. Miller, R. Beckwith, C. Fellbaum, D. Gross, and K. J. Miller, “Introduction to WordNet: An on-line lexical database,” *International Journal of Lexicography*, vol. 3, no. 4, pp. 235–244, 1990.
- [46] H. Kang, S. J. Yoo, and D. Han, “Senti-lexicon and improved naïve bayes algorithms for sentiment analysis of restaurant reviews,” *Expert Systems with Applications*, vol. 39, no. 5, pp. 6000 – 6010, 2012.
- [47] A. Esuli and F. Sebastiani, “SENTIWORDNET: A publicly available lexical resource for opinion mining,” in *Proceedings of the Fifth International Conference on Language Resources and Evaluation (LREC'06)*, (Genoa, Italy), European Language Resources Association (ELRA), May 2006.
- [48] S. Kiritchenko, X. Zhu, C. Cherry, and S. Mohammad, “NRC-Canada-2014: Detecting aspects and sentiment in customer reviews,” in *Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014)*, pp. 437–442, 2014.
- [49] D. M. Blei, A. Y. Ng, and M. I. Jordan, “Latent Dirichlet allocation,” *Journal of Machine Learning Research*, vol. 3, no. Jan, pp. 993–1022, 2003.
- [50] I. Titov and R. McDonald, “A joint model of text and aspect ratings for sentiment summarization,” in *Proceedings of ACL-08: HLT*, pp. 308–316, 2008.
- [51] C. Lin and Y. He, “Joint sentiment/topic model for sentiment analysis,” in *Proceedings of the 18th ACM Conference on Information and Knowledge Management*, pp. 375–384, 2009.
- [52] C. Lin, Y. He, R. Everson, and S. Ruger, “Weakly supervised joint sentiment-topic detection from text,” *IEEE Transactions on Knowledge and Data Engineering*, vol. 24, no. 6, pp. 1134–1145, 2011.
- [53] Z. Hai, G. Cong, K. Chang, P. Cheng, and C. Miao, “Analyzing sentiments in one go: A supervised joint topic modeling approach,” *IEEE Transactions on Knowledge and Data Engineering*, vol. 29, no. 6, pp. 1172–1185, 2017.
- [54] X. Pu, G. Wu, and C. Yuan, “User-aware topic modeling of online reviews,” *Multimedia Systems*, vol. 25, no. 1, pp. 59–69, 2019.
- [55] D. Andrzejewski, X. Zhu, and M. Craven, “Incorporating domain knowledge into topic modeling via Dirichlet forest priors,” in *Proceedings of the 26th Annual International Conference on Machine Learning*, pp. 25–32, 2009.

- [56] Z. Chen, A. Mukherjee, B. Liu, M. Hsu, M. Castellanos, and R. Ghosh, "Exploiting domain knowledge in aspect extraction," in *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing*, pp. 1655–1667, 2013.
- [57] X. Cheng, X. Yan, Y. Lan, and J. Guo, "BTM: Topic modeling over short texts," *IEEE Transactions on Knowledge and Data Engineering*, vol. 26, no. 12, pp. 2928–2941, 2014.
- [58] N. Li, C.-Y. Chow, and J.-D. Zhang, "Seeded-BTM: Enabling biterm topic model with seeds for product aspect mining," in *2019 IEEE 21st International Conference on High Performance Computing and Communications; IEEE 17th International Conference on Smart City; IEEE 5th International Conference on Data Science and Systems (HPCC/SmartCity/DSS)*, pp. 2751–2758, IEEE, 2019.
- [59] H. H. Do, P. Prasad, A. Maag, and A. Alsadoon, "Deep learning for aspect-based sentiment analysis: A comparative review," *Expert Systems with Applications*, vol. 118, pp. 272–299, 2019.
- [60] J. M. Digman, "Personality structure: Emergence of the five-factor model," *Annual Review of Psychology*, vol. 41, no. 1, pp. 417–440, 1990.
- [61] L. R. Goldberg, "An alternative" description of personality": The big-five factor structure.,," *Journal of Personality and Social Psychology*, vol. 59, no. 6, p. 1216, 1990.
- [62] Hürriyet Emlak, "Hürriyet Emlak | Türkiye'nin emlak sitesi & emlak ilanları." [Online]. Available: <https://www.hurriyetemlak.com>. [Accessed: Aug. 17, 2020].
- [63] Yemek Sepeti, "User agreement - Yemek Sepeti." [Online]. Available: <https://www.yemeksepeti.com/en/user-agreement>. [Accessed: Aug. 17, 2020].
- [64] V. I. Levenshtein, "Binary codes capable of correcting deletions, insertions, and reversals," *Soviet Physics Doklady*, vol. 10, no. 8, pp. 707–710, 1966.
- [65] A. A. Akın and M. D. Akın, "Zemberek, an open source NLP framework for Turkic languages," *Structure*, vol. 10, pp. 1–5, 2007.
- [66] J. L. Fleiss, "Measuring nominal scale agreement among many raters," *Psychological Bulletin*, vol. 76, no. 5, p. 378, 1971.
- [67] R. Bayar, "CBS yardımıyla modern alışveriş merkezleri için uygun yer seçimi: Ankara Örneği (location choice for shopping mall centers using GIS: Case study of Ankara)," *Coğrafi Bilimler Dergisi/Turkish Journal Geographical Sciences*, vol. 3, no. 2, pp. 19–38, 2005.
- [68] "Permutation importance vs random forest feature importance (MDI)." [Online]. Available: https://scikit-learn.org/stable/auto_examples/inspection/plot_permutation_importance.html. [Accessed: Sep. 28, 2020].
- [69] C. Sievert and K. Shirley, "LDAvis: A method for visualizing and interpreting topics," in *Proceedings of the Workshop on Interactive Language Learning, Visualization, and Interfaces*, pp. 63–70, 2014.
- [70] J. Chuang, C. D. Manning, and J. Heer, "Termite: Visualization techniques for assessing textual topic models," in *Proceedings of the International Working Conference on Advanced Visual Interfaces*, pp. 74–77, 2012.

- [71] M. Tran and M. Truong, "Clustering short text messages using unsupervised machine learning," *LU-CS-EX 2019-20*, 2019.
- [72] R. F. Tate, "Correlation between a discrete and a continuous variable. point-biserial correlation," *The Annals of Mathematical Statistics*, vol. 25, no. 3, pp. 603–607, 1954.
- [73] S. G. Kwak and J. H. Kim, "Central limit theorem: the cornerstone of modern statistics," *Korean Journal of Anesthesiology*, vol. 70, no. 2, p. 144, 2017.
- [74] H.-Y. Kim, "Statistical notes for clinical researchers: Chi-squared test and fisher's exact test," *Restorative Dentistry & Endodontics*, vol. 42, no. 2, pp. 152–155, 2017.
- [75] M. L. McHugh, "The chi-square test of independence," *Biochemia Medica*, vol. 23, no. 2, pp. 143–149, 2013.
- [76] R. Simas, F. Maestri, and D. Normando, "Controlling false positive rates in research and its clinical implications," *Dental Press Journal of Orthodontics*, vol. 19, no. 3, pp. 24–25, 2014.
- [77] S. Güler, "Türk mutfak kültürü ve yeme içme alışkanlıkları," *Dumlupınar Üniversitesi Sosyal Bilimler Dergisi*, vol. 26, no. S 1, pp. 24–30, 2010.
- [78] E. Tomul, "Türkiye'de eğitime katılım üzerinde gelirin etkisi," *Pamukkale Üniversitesi Eğitim Fakültesi Dergisi*, vol. 22, no. 22, pp. 122–131, 2007.
- [79] L. Alkan Gökler, "Ankara'da konut fiyatları farklılaşmasının hedonik analiz yardımıyla incelenmesi," *Megaron*, 2017.
- [80] E. Aybars, "Ankara Aspava kültürü," Mar 2016. [Online]. Available: <http://www.radikal.com.tr/yazarlar/evren-aybars/ankara-aspava-kulturu-1522747>. [Accessed: Sep. 2, 2020].
- [81] "Perfect pairing: Pizza and soy sauce a thing now?," Nov 2017. [Online]. Available: <https://torontosun.com/life/food/perfect-pairing-pizza-and-soy-sauce-a-thing-now>. [Accessed: Sep. 9, 2020].
- [82] I. C. McManus and A. Furnham, "Aesthetic activities and aesthetic attitudes: Influences of education, background and personality on interest and involvement in the arts," *British Journal of Psychology*, vol. 97, no. 4, pp. 555–587, 2006.
- [83] G. Ulukan, "Günde 400 bin sipariş gönderen yemeksepeti'nin güncel verileri," Jun 2019. [Online]. Available: <https://webrazzi.com/2019/06/27/400-bin-siparis-yemeksepeti-guncel-veriler>. [Accessed: Sep. 23, 2020].
- [84] "YÖK Üniversiteler." [Online]. Available: <https://www.yok.gov.tr/universiteler/universitelerimiz>. [Accessed: Sep. 23, 2020].

APPENDIX A

MENU LEXICONS AND ANNOTATIONS

A.1 Ingredient Lexicon

- ahtapot (octopus)
- ananas (pineapple)
- anason (aniseed)
- ançuez (anchovies)
- antrikot (rib steak)
- avokado (avocado)
- baharat (spice)
- bal (honey)
- balsamik sos, balzamik sos (balsamic glaze)
- barbekü sos, barbekü (barbecue sauce)
- beyaz peynir (feta cheese)
- bezelye (green pea)
- biber (pepper)
- biftek (beef steak)
- bolonez sos, bolonez (bolognese sauce)
- bonfile (fillet steak)
- brokoli (broccoli)
- ceviz (walnut)
- chili sos (chili sauce)
- ciğer (liver)
- çarliston biber (banana pepper)
- çedar, cheddarlı, çedar peyniri, cheddar peyniri (cheddar cheese)
- çemen (a paste made of fenugreek seeds, cumin, and other spices)
- çilek (strawberry)
- çörekotu, çörek otu (black seed)
- dana (calf)
- dil peyniri (string cheese)
- domates (tomato)
- domates sos, pizza sos (tomato sauce, pizza sauce)
- domuz (pork)
- döner (doner kebab)
- edam peyniri (Edam cheese)
- enginar (artichoke)
- et (meat, mostly refers to red meat by itself)
- ezine peyniri (Ezine cheese, a type of feta cheese)
- fasulye (bean)
- fesleğen (sweet basil)
- findık (hazelnut)
- fontina peyniri (Fontina cheese)
- füme hindi, hindi füme (smoked turkey)
- gorgonzola peyniri (Gorgonzola cheese)
- gouda peyniri (Gouda cheese)
- gravyer peyniri (Gruyère cheese)
- hamur (dough)
- havuç (carrot)
- hellim peyniri (halloumi cheese)
- hindi (turkey)
- ıspanak (spinach)
- isot (dried isot pepper spread)
- istiridye mantarı (oyster mushroom)
- jalapeno (jalapeño pepper)
- jambon (calf ham)
- kabak (zucchini)
- kaburga (rib)
- kalamar (squid)
- kapyra (kapyra pepper)
- karabiber (black pepper)

- karides (shrimp)
- karnabahar (cauliflower)
- kaşar, kaşar peyniri (kasseri cheese)
- kavurma (fried meat)
- kekik (thyme)
- kepek (whole wheat)
- ketçap (ketçap)
- kırmızısoğan, kırmızı soğan (red onion)
- kıyma (ground meat)
- kokoreç (grilled lamb intestines)
- köfte (meatball)
- köri sos (curry sauce)
- köz biber, közde biber, közlenmiş biber (roasted pepper)
- köz patlıcan, közde patlıcan, közlenmiş patlıcan, közlenmiş patlıcan ezmesi, yoğurtlu patlıcan, yoğurtlu patlıcan salatası (roasted eggplant)
- köz soğan, közde soğan, közlenmiş soğan (roasted onion)
- krema (cream)
- kuşkonmaz (asparagus)
- kuyruk yağı (tail fat)
- kuzu (lamb)
- labne, labne peyniri (cream cheese)
- levrek (sea bass)
- limon (lemon)
- lor peynir (curd cheese)
- mantar (mushroom)
- marinara (marinara sauce)
- marul (lettuce)
- maydanoz, maydanoz (parsley)
- mayonez (mayonnaise)
- mercimek (lentil)
- mısır (corn)
- mozzarellalı, mozzarella peyniri, mozzarella peyniri, mozzarella peyniri (mozzarella cheese)
- muz (banana)
- nacho (nacho chips)
- nane (mint)
- nutella, çikolata (nutella, chocolate)
- otlu peynir (herby cheese)
- paprika biber, kırmızıbiber, kırmızı biber (red pepper)
- parmesan peyniri (parmesan cheese)
- pastırma (cured beef similar to pastrami, covered by çemen and air-dried)
- pastrami (pastrami)
- patates (potato)
- patlıcan (eggplant)
- pepperoni (pepperoni)
- pesto sos (pesto sauce)
- peynir (cheese)
- pırasa (leek)
- porçini mantarı (porcini mushrooms)
- pul biber (chili pepper)
- ricotta peyniri (ricotta cheese)
- roka (rocket)
- rokfor, rokfor peyniri (roquefort cheese)
- salam (salami)
- salatalık (cucumber)
- salatalık turşusu, kornişon turşu (pickled cucumber)
- salça (tomate and/or pepper paste)
- sarımsak (garlic)
- sebze (vegetables)
- siyah zeytin, siyahzeytin (black olive)
- soğan (onion)
- somon (salmon)
- sos (sauce)
- sosis (sausage)
- soya sosu, soya sos (soy sauce)
- sucuk (fermented sausage)
- sumak (sumac)
- susam (sesame seed)
- tahıl (whole grain)
- tahin (tahini)
- tavuk (chicken)
- tere yağ, tereyağı (butter)
- teriyaki (teriyaki)
- ton balık, ton balığı (tuna fish)
- trüf (truffle)
- turşu (pickle)
- tuz (salt)
- yağ (oil)
- yerfıstığı, yer fıstığı (peanut)
- yeşil zeytin, yeşilzeytin (green olive)
- yeşilbiber, yeşil biber (green pepper)
- yoğurt (yogurt)
- yumurta (egg)
- zahter (za'atar)
- zeytin (olive)
- zeytinyağı (olive oil)

A.2 Side-dish Lexicon

- barbekü sos, barbekü (barbecue sauce)
- biber (pepper)
- cacık (a side dish made of yogurt, water, cucumber, and mint)
- cips (chips)
- çarliston biber (banana pepper)
- çiğ köfte (a dish made of fine bulgur, used to be made of raw meat)
- çorba (soup)
- domates (tomato)
- ezme (a spicy side dish made from mashed tomatoes along with some other vegetables and herbs)
- havuç (carrot)
- humus (hummus)
- irmik helvası (semolina helva)
- kadayıf (a traditional dessert made of pastry and syrup)
- karalahana (red cabbage)
- kırmızısoğan, kırmızı soğan (red onion)
- köz biber, közde biber, közlenmiş biber (roasted pepper)
- köz patlıcan, közde patlıcan, közlenmiş patlıcan, közlenmiş patlıcan ezmesi, yoğurtlu patlıcan, yoğurtlu patlıcan salatası (roasted eggplant)
- köz soğan, közde soğan, közlenmiş soğan (roasted onion)
- limon (lemon)
- marul (lettuce)
- maydanoz, maydonoz (parsley)
- meze (appetizer/snack/side dish)
- nane (mint)
- patates (potato)
- patates kızartma (fried potatoes)
- pilav (cooked rice)
- roka (rocket)
- salata, mevsim yeşillikleri (salad, salad greens)
- salatalık (cucumber)
- sebze (vegetables)
- semizotu (purslane)
- soğan (onion)
- sumaklı soğan (onions with sumac)
- tarator (a side dish mostly made of yogurt, garlic, and other ingredients)
- turşu (pickle)
- yeşilbiber, yeşil biber (green pepper)
- yoğurt (yogurt)

A.3 Annotator Votes

Table 22: Annotator votes for each lexicon item

Item	Is ingredient?			Is side dish?		
	Vote 1	Vote 2	Vote 3	Vote 1	Vote 2	Vote 3
ahtapot (octopus)	Yes	Yes	Yes	No	No	No
ananas (pineapple)	Yes	Yes	Yes	No	No	No
anason (aniseed)	Yes	Yes	Yes	No	No	No
ançuez (anchovies)	Yes	Yes	Yes	No	No	No
antrikot (rib steak)	Yes	Yes	Yes	No	No	No
avokado (avocado)	Yes	Yes	Yes	No	No	No
baharat (spice)	Yes	Yes	Yes	No	No	No
bal (honey)	Yes	Indecisive	Yes	No	Indecisive	No
balsamik sos, balzamik sos (balsamic glaze)	Yes	Yes	Yes	No	No	No
barbekü sos, barbekü (barbecue sauce)	Yes	Yes	Yes	Yes	Yes	Yes
beyaz peynir (feta cheese)	Yes	Yes	Yes	No	No	No
bezelye (green pea)	Yes	Yes	Yes	No	No	No
biber (pepper)	Yes	Yes	Yes	Yes	Yes	Yes
biftek (beef steak)	Yes	Yes	Yes	No	No	No
bolonez sos, bolonez (bolognese sauce)	Yes	Yes	Yes	No	No	No
bonfile (fillet steak)	Yes	Yes	Yes	No	No	No
brokoli (broccoli)	Yes	Yes	Yes	No	No	No
cacık (a side dish made of yogurt, water, cucumber, and mint)	No	No	No	Yes	Yes	Yes
ceviz (walnut)	Yes	Yes	Yes	No	No	No
chili sos (chili sauce)	Yes	Yes	Yes	No	No	No
ciğer (liver)	Yes	Yes	Yes	No	No	No
cips (chips)	No	No	No	Yes	Yes	Yes
çarliston biber (banana pepper)	Yes	Yes	Yes	Yes	Yes	Yes
çedar, cheddarlı, çedar peyniri, cheddar peyniri (cheddar cheese)	Yes	Yes	Yes	No	No	No
çemen (a paste made of fenugreek seeds, cumin, and other spices)	Yes	Yes	Yes	No	No	No
çiğ köfte (a dish made of fine bulgur, used to be made of raw meat)	No	No	No	Yes	Yes	Yes
çilek (strawberry)	Yes	Yes	Yes	No	No	No
çorba (soup)	No	No	No	Yes	Yes	Yes
çörekotu, çörek otu (black seed)	Yes	Yes	Yes	No	No	No
dana (calf)	Yes	Yes	Yes	No	No	No
dil peyniri (string cheese)	Yes	Yes	Yes	No	No	No
domates (tomato)	Yes	Yes	Yes	Yes	Yes	Yes
domates sos, pizza sos (tomato sauce, pizza sauce)	Yes	Yes	Yes	No	No	No
domuz (pork)	Yes	Yes	Yes	No	No	No
⋮	⋮	⋮	⋮	⋮	⋮	⋮

Table 22: (continued)

Item	Is ingredient?			Is side dish?		
	Vote 1	Vote 2	Vote 3	Vote 1	Vote 2	Vote 3
⋮	⋮	⋮	⋮	⋮	⋮	⋮
döner (doner kebab)	Yes	Yes	Yes	No	No	No
edam peyniri (Edam cheese)	Yes	Yes	Yes	No	No	No
enginar (artichoke)	Yes	Yes	Yes	No	No	No
et (meat)	Yes	Yes	Yes	No	No	No
ezine peyniri (Ezine cheese, a type of feta cheese)	Yes	Yes	Yes	No	No	No
ezme (a spicy side dish made from mashed tomatoes along with some other vegetables and herbs)	No	No	No	Yes	Yes	Yes
fasulye (bean)	Yes	Yes	Yes	No	No	No
fesleğen (sweet basil)	Yes	Yes	Yes	No	No	No
findık (hazelnut)	Yes	Yes	Yes	No	No	No
fontina peyniri (Fontina cheese)	Yes	Yes	Yes	No	No	No
füme hindi, hindi füme (smoked turkey)	Yes	Yes	Yes	No	No	No
gorgonzola peyniri (Gorgonzola cheese)	Yes	Yes	Yes	No	No	No
gouda peyniri (Gouda cheese)	Yes	Yes	Yes	No	No	No
gravyer peyniri (Gruyère cheese)	Yes	Yes	Yes	No	No	No
hamur (dough)	Yes	Yes	Yes	No	No	No
havuç (carrot)	Yes	Yes	Yes	Yes	Yes	Yes
hellim peyniri (halloumi cheese)	Yes	Yes	Yes	No	No	No
hindi (turkey)	Yes	Yes	Yes	No	No	No
humus (hummus)	No	No	No	Yes	Yes	Yes
ıspanak (spinach)	Yes	Yes	Yes	No	No	No
irmik helvası (semolina helva)	No	No	No	Yes	Yes	Yes
isot (dried isot pepper spread)	Yes	Yes	Yes	No	No	No
istiridy mantarı (oyster mushroom)	Yes	Yes	Yes	No	No	No
jalapeno (jalapeño pepper)	Yes	Yes	Yes	No	No	No
jambon (calf ham)	Yes	Yes	Yes	No	No	No
kabak (zucchini)	Yes	Yes	Yes	No	No	No
kaburga (rib)	Yes	Yes	Yes	No	No	No
kadayıf (a traditional dessert made of pastry and syrup)	No	No	No	Yes	Yes	Yes
kalamar (squid)	Yes	Yes	Yes	No	No	No
kapya (kappa pepper)	Yes	Yes	Yes	No	No	No
karabiber (black pepper)	Yes	Yes	Yes	No	No	No
karalahana (red cabbage)	No	No	No	Yes	Yes	Yes
karides (shrimp)	Yes	Yes	Yes	No	No	No
karnabahar (cauliflower)	Yes	Yes	Yes	No	No	No
kaşar, kaşar peyniri (kasseri cheese)	Yes	Yes	Yes	No	No	No
kavurma (fried meat)	Yes	Yes	Yes	No	No	No
kekik (thyme)	Yes	Yes	Yes	No	No	No
kepek (whole wheat)	Yes	No	Yes	No	No	No
ketçap (ketchup)	Yes	Yes	Yes	No	No	No
kırmızısoğan, kırmızı soğan (red onion)	Yes	Yes	Yes	Yes	Yes	Yes
kıyma (ground meat)	Yes	Yes	Yes	No	No	No
kokoreç (grilled lamb intestines)	Yes	Yes	Yes	No	No	No
köfte (meatball)	Yes	Yes	Yes	No	No	No
⋮	⋮	⋮	⋮	⋮	⋮	⋮

Table 22: (continued)

Item	Is ingredient?			Is side dish?		
	Vote 1	Vote 2	Vote 3	Vote 1	Vote 2	Vote 3
⋮	⋮	⋮	⋮	⋮	⋮	⋮
köri sos (curry sauce)	Yes	Yes	Yes	No	No	No
köz biber, közde biber, közlenmiş biber (roasted pepper)	Yes	Yes	Yes	Yes	Yes	Yes
köz patlıcan, közde patlıcan, közlenmiş patlıcan, közlenmiş patlıcan ezmesi, yoğurtlu patlıcan, yoğurtlu patlıcan salatası (roastted eggplant)	Yes	Yes	Yes	Yes	Yes	Yes
köz soğan, közde soğan, közlenmiş soğan (roasted onion)	Yes	Yes	Yes	Yes	Yes	Yes
krema (cream)	Yes	Yes	Yes	No	No	No
kuşkonmaz (asparagus)	Yes	Yes	Yes	No	No	No
kuyruk yağı (tail fat)	Yes	Yes	Yes	No	No	No
kuzu (lamb)	Yes	Yes	Yes	No	No	No
labne, labne peyniri (cream cheese)	Yes	Yes	Yes	No	No	No
levrek (sea bass)	Yes	Yes	Yes	No	No	No
limon (lemon)	Yes	No	Yes	Yes	Yes	Yes
lor peynir (curd cheese)	Yes	Yes	Yes	No	No	No
mantar (mushroom)	Yes	Yes	Yes	No	No	No
marinara (marinara sauce)	Yes	Yes	Yes	No	No	No
marul (lettuce)	Yes	No	Yes	Yes	Yes	Yes
maydanoz, maydonoz (parsley)	Yes	Yes	Yes	Yes	Yes	Yes
mayonez (mayonnaise)	Yes	Yes	Yes	No	No	No
mercimek (lentil)	Yes	Yes	Yes	No	No	No
meze (appetizer/snack/side dish)	No	No	No	Yes	Yes	Yes
mısır (corn)	Yes	Yes	Yes	No	No	No
mozzarellalı, mozzarella peyniri, mozzaralle peyniri, mozzarella peyniri (mozzarella cheese)	Yes	Yes	Yes	No	No	No
muz (banana)	Yes	Yes	Yes	No	No	No
nacho (nacho chips)	Yes	Yes	Yes	No	No	No
nane (mint)	Yes	Yes	Yes	Yes	Yes	Yes
nutella, çikolata (nutella, chocolate)	Yes	Yes	Yes	No	No	No
otlu peynir (herby cheese)	Yes	Yes	Yes	No	No	No
paprika biber, kırmızıbiber, kırmızı biber (red pepper)	Yes	Yes	Yes	No	No	No
parmesan peyniri (Parmesan cheese)	Yes	Yes	Yes	No	No	No
pastırma (cured beef similar to pastrami, covered by çemen and air-dried)	Yes	Yes	Yes	No	No	No
pastrami (pastrami)	Yes	Yes	Yes	No	No	No
patates (potato)	Yes	Yes	Yes	Yes	Yes	Yes
patates kızartması, patates kızartma (French fries)	No	No	No	Yes	Yes	Yes
patlıcan (eggplant)	Yes	Yes	Yes	No	No	No
pepperoni (pepperoni)	Yes	Yes	Yes	No	No	No
pesto sos (pesto sauce)	Yes	Yes	Yes	No	No	No
peynir (cheese)	Yes	Yes	Yes	No	No	No
pırasa (leek)	Yes	Yes	Yes	No	No	No
pilav (cooked rice)	No	No	No	Yes	Yes	Yes
⋮	⋮	⋮	⋮	⋮	⋮	⋮

Table 22: (continued)

Item	Is ingredient?			Is side dish?		
	Vote 1	Vote 2	Vote 3	Vote 1	Vote 2	Vote 3
⋮	⋮	⋮	⋮	⋮	⋮	⋮
porcini mantarı (porcini mushrooms)	Yes	Yes	Yes	No	No	No
pul biber (chili pepper)	Yes	Yes	Yes	No	No	No
ricotta peyniri (ricotta cheese)	Yes	Yes	Yes	No	No	No
roka (rocket)	Yes	Yes	Yes	Yes	Yes	Yes
rokfor, rokfor peyniri (roquefort cheese)	Yes	Yes	Yes	No	No	No
salam (salami)	Yes	Yes	Yes	No	No	No
salata, mevsim yeşillikleri (salad, salad greens)	No	No	No	Yes	Yes	Yes
salatalık (cucumber)	Yes	No	Yes	Yes	Yes	Yes
salatalık turşusu, kornişon turşu (pickled cucumber)	Yes	Yes	Yes	No	No	No
salça (tomate and/or pepper paste)	Yes	Yes	Yes	No	No	No
sarımsak (garlic)	Yes	Yes	Yes	No	No	No
sebze (vegetables)	Yes	Yes	Yes	Yes	Yes	Yes
semizotu (purslane)	No	No	No	Yes	Yes	Yes
siyah zeytin, siyahzeytin (black olive)	Yes	Yes	Yes	No	No	No
soğan (onion)	Yes	Yes	Yes	Yes	Yes	Yes
somon (salmon)	Yes	Yes	Yes	No	No	No
sos (sauce)	Yes	Yes	Yes	No	No	No
sisos (sausage)	Yes	Yes	Yes	No	No	No
soya sosu, soya sos (soy sauce)	Yes	Yes	Yes	No	No	No
sucuk (fermented sausage)	Yes	Yes	Yes	No	No	No
sumak (sumac)	Yes	Yes	Yes	No	No	No
sumaklı soğan (onions with sumac)	No	No	No	Yes	Yes	Yes
susam (sesame seed)	Yes	Yes	Yes	No	No	No
tahıl (whole grain)	Yes	No	Yes	No	No	No
tahin (tahini)	Yes	Yes	Yes	No	No	No
tarator (a side dish mostly made of yogurt, garlic, and other ingredients)	No	No	No	Yes	Yes	Yes
tavuk (chicken)	Yes	Yes	Yes	No	No	No
tere yağ, tereyağı (butter)	Yes	Yes	Yes	No	No	No
teriyaki (teriyaki)	Yes	Yes	Yes	No	No	No
ton balık, ton balığı (tuna fish)	Yes	Yes	Yes	No	No	No
trüf (truffle)	Yes	Yes	Yes	No	No	No
turşu (pickle)	Yes	Yes	Yes	Yes	Yes	Yes
tuz (salt)	Yes	Yes	Yes	No	No	No
yağ (oil)	Yes	Yes	Yes	No	No	No
yerfıstığı, yer fıstığı (peanut)	Yes	Yes	Yes	No	No	No
yeşil zeytin, yeşilzeytin (green olive)	Yes	Yes	Yes	No	No	No
yeşilbiber, yeşil biber (green pepper)	Yes	Yes	Yes	Yes	Yes	Yes
yoğurt (yogurt)	Yes	Yes	Yes	Yes	Yes	Yes
yumurta (egg)	Yes	Yes	Yes	No	No	No
zahter (za'atar)	Yes	Yes	Yes	No	No	No
zeytin (olive)	Yes	Yes	Yes	No	No	No
zeytinyağı (olive oil)	Yes	Yes	Yes	No	No	No

APPENDIX B

EXTRA MATERIAL

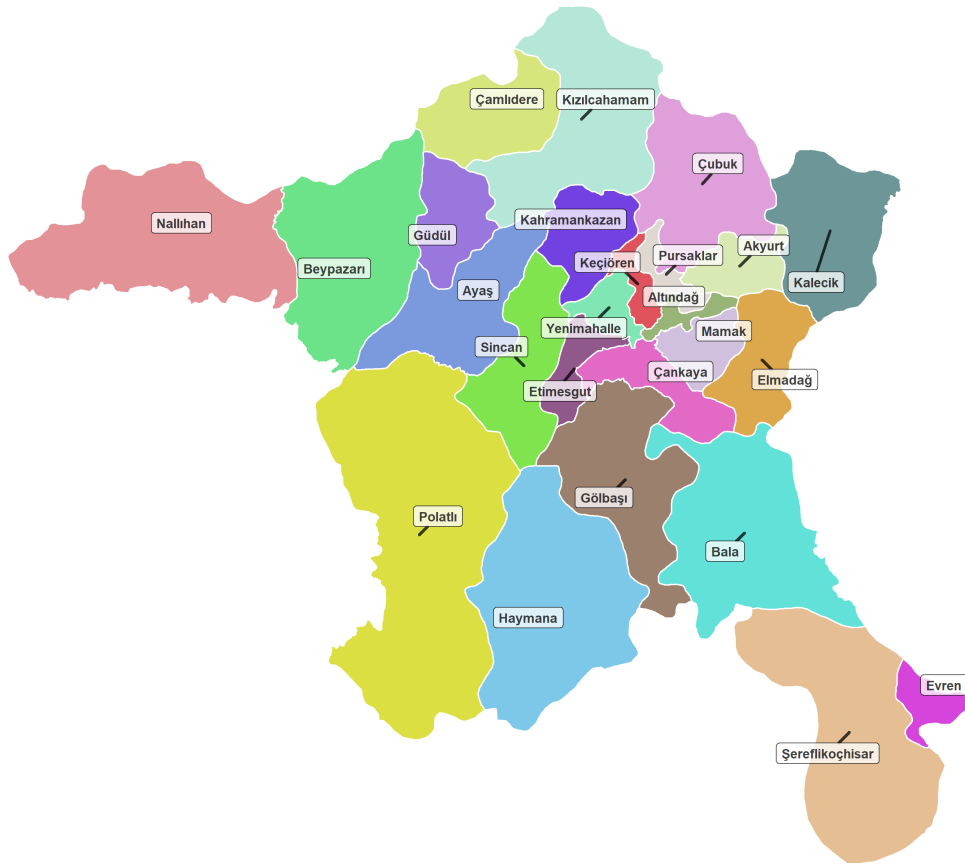


Figure 19: Districts of Ankara

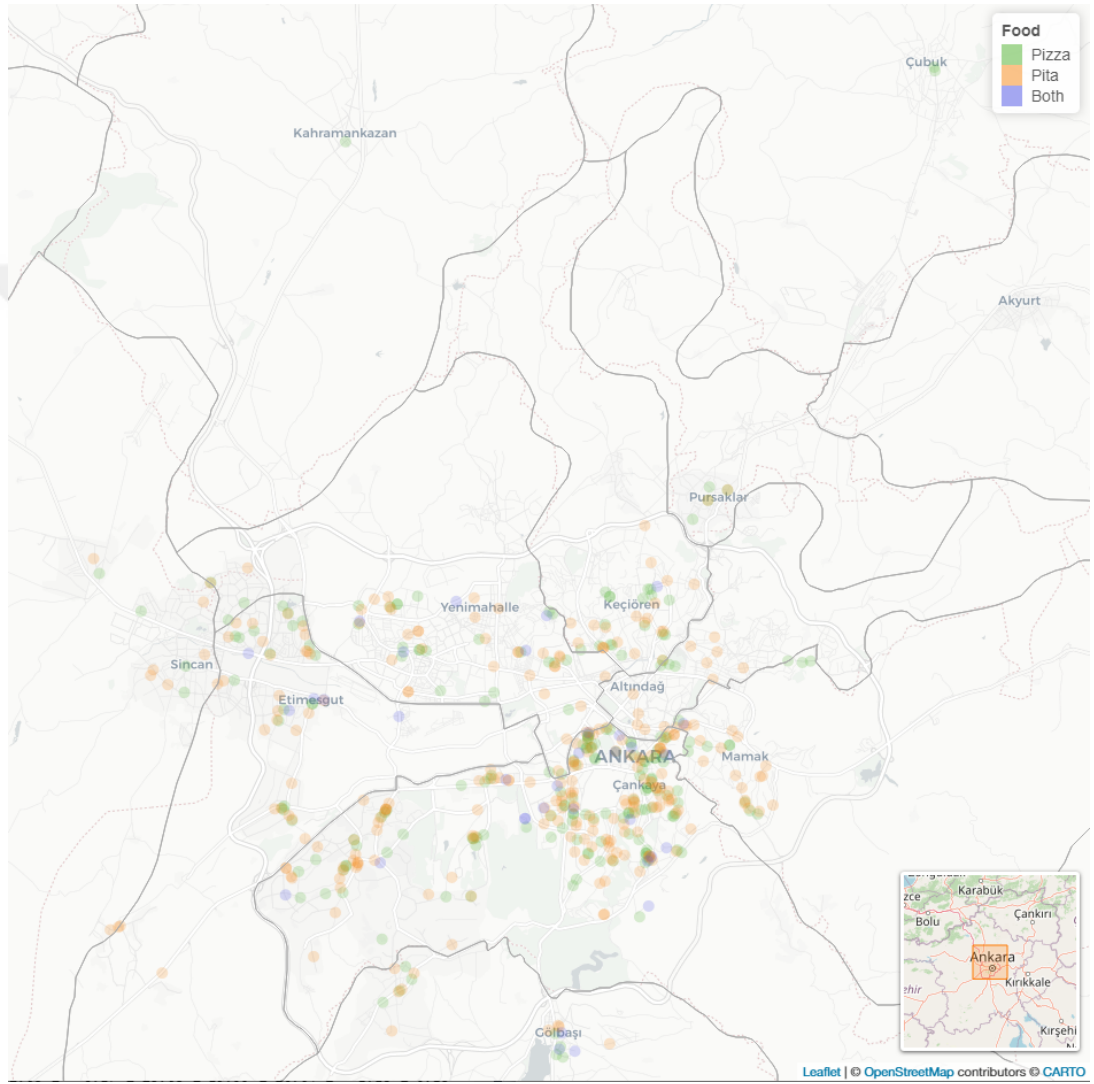


Figure 20: Locations of restaurants that serve pizza, pita, or both on map along with the district borders

Table 23: Ankara's education level percentages by district, rounded (Everyone is represented only under the highest education they have)

District	Unknown	Illiterate	Literate	Elementary school	Middle school	Primary school	Secondary school	Undergraduate degree	Master's degree	PhD degree
Akyurt	0.35	1.53	9.5	29.46	14.87	14.57	20.58	8.56	0.58	0
Altındağ	1.05	2.71	9.37	26.94	12.58	14.17	18.02	13.62	1.34	0.19
Ayaş	0	0.26	6.31	35.36	8.84	15.85	22.44	10.81	0.14	0
Bala	0.18	3.84	8.41	31.22	13.73	14.63	21.24	6.75	0	0
Beypazarı	0.35	1.7	8.72	35.92	8.9	12.22	19.64	11.74	0.82	0
Çamlıdere	0	6.1	7.69	33.12	10.96	14.06	21.47	6.61	0	0
Çankaya	1	0.86	5.14	10.95	2.97	7.95	24.98	33.52	10.17	2.45
Çubuk	0.53	3.1	8.77	29.36	11.74	14.79	21.54	9.38	0.76	0.03
Elmadag	0.21	2.85	8.53	22.86	9.73	14.61	26.56	13.9	0.74	0
Etimesgut	0.66	1.03	7.65	16.39	7.3	11.41	24.81	25.31	4.82	0.61
Evren	0.6	4.82	14.5	33.22	9.34	14.1	18.66	4.76	0	0
Gölbaşı	0.64	1.36	7.71	19.15	7.57	11.77	25.05	21.94	3.89	0.92
Güdül	0	2.48	6.74	44.66	7.35	12.86	18.72	6.95	0.24	0
Haymana	0.43	3.74	10.85	33.91	15.41	14.29	16.74	4.53	0.11	0
Kahramankazan	0.41	1.52	9.34	27.11	12.87	14.62	21.26	11.83	0.96	0.08
Kalecik	0	2.63	8.34	43.03	8.3	11.55	19.05	7.1	0	0
Keçiören	0.77	1.88	8.16	21.63	9.27	13.63	25.55	16.64	2.23	0.23
Kızılcahamam	0.43	3.38	7.18	33.97	7.61	11.88	24.52	10.28	0.75	0
Mamak	0.72	2.52	8.23	22.76	11.73	14.23	24.49	13.78	1.46	0.09
Nallıhan	0	1.53	9.1	37.48	9.88	12	19.91	9.43	0.66	0
Polatlı	0.9	2.1	8.59	29.89	10.35	12.14	22.57	12.29	1.14	0.05
Pursaklar	0.6	1.5	9.44	23.1	12.06	15.21	23.97	12.7	1.3	0.12
Şereflikoçhisar	2.73	4.61	10.35	28.8	8.69	13.37	21.16	9.69	0.61	0
Sincan	0.54	1.86	8.98	23.61	13.52	15.46	24.01	11.05	0.93	0.05
Yenimahalle	0.73	1.44	7	17.7	5.62	11.08	25.73	25.24	4.81	0.66

While there are no overlaps between categories since everyone is counted under the highest level of education they completed, some of the categories might be harder to interpret. Due to changes on the education system, the first eight years of formal education (primary school) is now divided into two four-year parts (elementary school and middle school). Therefore, someone who completed the first eight years before the change is counted under "primary school" while someone who completed the first eight years after the change is counted under "middle school." Associate's degrees are counted with bachelor's degrees under the undergraduate degree category. This is one of the reasons education data is grouped before use as explained in Section 3.2.2.

Lexicon evaluation - Ingredients and side dishes

Ingredient: A substance used to prepare a specific dish.

Side dish: A food item that accompanies the main dish.

For my thesis, I need to analyze restaurant menus, specifically pizza and pita (pide) dishes from 'Yemek Sepeti'. To do so, I need custom lexicons for ingredients and side dishes.

Please look at the items listed below and decide if they are ingredients and/or side dishes. Be careful that some items below can be considered as both. For example, pepper (biber) can be put on a pizza as an ingredient, but it can be a side dish that is served with a pita. *Check all that apply.* If you think neither of them is suitable, leave both of them unchecked. If you are not sure, please check "Not sure." A food that would be considered as a main dish can be an ingredient as well. For example, a pita can have döner on it, and a pizza can have kokoreç on it. Since my thesis focuses on pizza and pita, it might help to consider these items from pizza and pita perspectives.

If you would like to see some examples for a specific item, you can click on the ⓘ icon right next to it. At most five related menu item names and descriptions will be shown to you. However, not all the items have examples. If you have no idea what that item is, you can click on the 🍷 icon right next to it to make a Google search.

karabiber ⓘ

Kıymalı Bafra Pidesi

kavurulmuş dana kıyma, soğan, karabiber

Kaşarlı Karadeniz Pide

kıyma, soğan, kaşar peyniri, yeşillikbiber, karabiber, salata ve acılı ezme ile

Karadeniz Pide

kıyma, soğan, yeşillikbiber, karabiber, salata ve acılı ezme ile

Romesco Pizza

pizza sos, mozzarella peyniri, kavurma, kız patlıcan, kız biber, karabiber

Quattro Formaggi Pizza

mozzarella peyniri, provolone peyniri, gorgonzola peyniri, gouda peyniri, parmesan peyniri, cherry domates, krema, karabiber

☐ Ingredient ☐ Side dish ☐ Not sure

salça ⓘ 🍷

☐ Ingredient ☐ Side dish ☐ Not sure

kekik ⓘ 🍷

☐ Ingredient ☐ Side dish ☐ Not sure

Figure 21: A screenshot of the custom web-based tool used to label lexicon items

73

Table 24: Pizza restaurants' comment topic likelihoods compared with two education clusters using the Kruskal-Wallis test. P-values are Bonferroni-corrected.

Comment topic	H	P	Comment topic	H	P
V74	3.32	.137	V81	0.29	1.180
V9	3.1	.157	V63	0.26	1.224
V59	2.56	.219	V73	0.24	1.246
V68	2.41	.240	V1	0.24	1.246
V35	2.13	.289	V27	0.24	1.246
V3	2.08	.298	V30	0.24	1.246
V48	2.08	.298	V56	0.21	1.291
V11	1.99	.316	V85	0.21	1.291
V18	1.99	.316	V43	0.2	1.314
V37	1.95	.326	V20	0.18	1.337
V5	1.82	.356	V54	0.18	1.337
V17	1.82	.356	V10	0.17	1.360
V70	1.57	.421	V69	0.17	1.360
V87	1.49	.444	V96	0.17	1.360
V90	1.41	.469	V16	0.16	1.384
V50	1.27	.521	V95	0.16	1.384
V42	1.2	.548	V8	0.14	1.407
V64	1.16	.562	V36	0.14	1.407
V45	1.06	.605	V22	0.13	1.431
V6	1.03	.620	V46	0.13	1.431
V28	1	.636	V7	0.12	1.454
V24	0.88	.699	V34	0.12	1.454
V77	0.85	.716	V61	0.12	1.454
V21	0.79	.749	V86	0.12	1.454
V52	0.76	.766	V65	0.11	1.478
V78	0.73	.784	V75	0.11	1.478
V83	0.73	.784	V80	0.1	1.502
V4	0.71	.801	V23	0.09	1.526
V62	0.71	.801	V29	0.08	1.551
V94	0.65	.837	V32	0.08	1.551
V49	0.63	.856	V51	0.08	1.551
V71	0.63	.856	V76	0.06	1.624
V44	0.6	.874	V92	0.06	1.624
V98	0.6	.874	V100	0.06	1.624
V33	0.58	.893	V84	0.05	1.649
V79	0.58	.893	V99	0.04	1.673
V82	0.51	.951	V38	0.04	1.698
V39	0.49	.971	V15	0.03	1.723
V57	0.46	.991	V2	0.01	1.823
V58	0.46	.991	V19	0.01	1.823
V40	0.44	1.011	V88	0.01	1.848
V67	0.44	1.011	V41	0.01	1.874
V53	0.42	1.031	V60	0	1.899
V93	0.42	1.031	V26	0	1.924
V13	0.4	1.052	V31	0	1.924
V89	0.4	1.052	V91	0	1.924
V55	0.38	1.073	V66	0	1.949
V25	0.36	1.094	V12	0	1.975
V47	0.36	1.094	V72	0	1.975
V97	0.34	1.115	V14	0	2.000

Table 25: Pita restaurants' comment topic likelihoods compared with two education clusters using the Kruskal-Wallis test. P-values are Bonferroni-corrected.

Comment topic	H	P	Comment topic	H	P
V14	7.09	.016	V53	0.23	1.261
V2	5.71	.034	V48	0.22	1.282
V63	4.09	.086	V60	0.21	1.289
V75	3.33	.136	V6	0.19	1.318
V44	3.27	.141	V18	0.19	1.325
V73	3.01	.165	V82	0.15	1.391
V70	2.98	.169	V79	0.15	1.398
V9	2.89	.178	V85	0.15	1.398
V47	2.55	.221	V99	0.13	1.442
V66	2.55	.221	V4	0.12	1.457
V92	2.35	.251	V27	0.12	1.465
V42	2.15	.284	V11	0.11	1.472
V96	1.93	.330	V57	0.11	1.487
V21	1.82	.355	V72	0.11	1.487
V40	1.61	.409	V31	0.1	1.502
V81	1.56	.423	V78	0.1	1.502
V97	1.36	.488	V55	0.09	1.525
V8	1.3	.508	V58	0.09	1.525
V52	1.29	.512	V86	0.09	1.525
V59	1.28	.517	V13	0.09	1.540
V23	1.27	.521	V5	0.08	1.563
V26	1.2	.546	V54	0.08	1.563
V98	1.17	.560	V51	0.07	1.585
V84	1.09	.591	V41	0.05	1.631
V50	1.06	.605	V67	0.05	1.639
V62	1	.633	V88	0.05	1.647
V32	0.94	.662	V17	0.05	1.655
V36	0.93	.672	V90	0.05	1.655
V61	0.82	.728	V56	0.04	1.693
V38	0.79	.750	V68	0.04	1.693
V65	0.67	.827	V29	0.03	1.724
V28	0.64	.849	V93	0.03	1.740
V24	0.62	.861	V25	0.03	1.748
V49	0.62	.861	V76	0.03	1.748
V10	0.61	.867	V15	0.02	1.756
V64	0.61	.873	V83	0.02	1.756
V1	0.58	.890	V39	0.02	1.771
V7	0.58	.890	V45	0.02	1.771
V43	0.57	.902	V30	0.02	1.787
V46	0.57	.902	V91	0.01	1.818
V71	0.53	.932	V3	0.01	1.850
V19	0.51	.951	V37	0.01	1.850
V20	0.5	.957	V77	0.01	1.858
V95	0.43	1.019	V16	0	1.929
V34	0.4	1.058	V69	0	1.945
V89	0.34	1.124	V74	0	1.945
V100	0.33	1.137	V35	0	1.976
V12	0.27	1.205	V80	0	1.976
V22	0.26	1.219	V87	0	1.976
V94	0.26	1.219	V33	0	2.000

Table 26: Pizza restaurants' menu term existences compared with two education clusters using Fisher's exact test. P-values are Bonferroni-corrected. An odds ratio higher than 1 indicates that the item is more common for the less educated districts' cluster while an odds ratio lower than 1 indicates that the item is more common for the more educated districts' cluster.

Item	Odds ratio	P
KÖFTE	13.64	.001
ROKA	0.07	.008
KABURGA	0	.028
FASULYE	0	.056
SİYAH ZEYTİN	0.25	.076
KEKİK	0.25	.094
SEBZE	0.14	.099
ROKFOR PEYNİRİ	0	.115
BİFTEK	0	.207
ZEYTİNYAĞI	0	.210
JAMBON	0.25	.219
SARIMSAK	0.2	.317
ANANAS	0.22	.330
PATLICAN	0.22	.330
PESTO SOSU	0	.376
RİCOTTA PEYNİRİ	0	.376
PUL BİBER	0	.379
BAHARAT	0	.379
PARMESAN PEYNİRİ	0.31	.387
AHTAPOT	Inf	.388
DEREOTU	Inf	.388
PASTIRMA	0.4	.394
JALAPENO	0.4	.394
ET	2.3	.418
FESLEĞEN	0.38	.429
KIRMIZI BİBER	2.3	.443
TON BALIĞI	4.13	.537
CEVİZ	0.27	.544
PEPPERONİ	0.27	.544
SOS	0.46	.579
ENGİNAR	0	.676
KABAK	0	.676
SOĞAN	2.27	.684
PEYNİR	2.27	.684
ÇİKOLATA	4.28	.706
CHİLİ SOS	4.28	.706
DÖNER	1.99	.788
LOR PEYNİRİ	0.33	.870
YEŞİL ZEYTİN	1.71	.870
BARBEKÜ	0.3	.872
MAYDANOZ	0.3	.872
BONFİLE	1.72	.933
İSTİRİDYE MANTARI	2.14	.964
⋮	⋮	⋮

Table 26: (continued)

Item	Odds ratio	P
⋮	⋮	⋮
KIRMIZI SOĞAN	0.56	1.042
KIYMA	1.57	1.061
KAVURMA	1.66	1.076
MOZZARELLA PEYNİRİ	Inf	1.149
HİNDİ	0	1.149
DİL PEYNİRİ	0	1.149
KAPYA	0	1.149
BOLONEZ	0	1.149
DOMATES	Inf	1.149
LİMON	0	1.159
GORGONZOLA PEYNİRİ	0	1.159
TERİYAKİ	0	1.159
MARİNARA SOS	0	1.159
KREMA	0	1.159
KOKOREÇ	0	1.159
ANASON	0	1.159
MARUL	0	1.176
HAVUÇ	0	1.176
SOMON	0	1.176
BİBER	0.57	1.228
LABNE PEYNİRİ	0.42	1.345
BROKOLİ	0.42	1.345
SUSAM	0.37	1.354
KÖZLENMİŞ PATLICAN	0.58	1.433
TAVUK	1.73	1.433
KARİDES	0.58	1.433
ZEYTİN	1.35	1.520
NANE	1.41	2.000
HELLİM PEYNİRİ	1.15	2.000
KALAMAR	1.41	2.000
DANA	1.16	2.000
ÇEDAR PEYNİRİ	0.93	2.000
SALAM	1.1	2.000
YOĞURT	0	2.000
KARNABAHAAR	1.04	2.000
ÇÖREK OTU	1.04	2.000
NACHO	0	2.000
KÖZLENMİŞ BİBER	1.17	2.000
DOMATES SOSU	1.49	2.000
TURŞU	0.82	2.000
ÇEMEN	0	2.000
KÖRİ SOSU	0.56	2.000
BEZELYE	0	2.000
GOUDA PEYNİRİ	0	2.000
ISPANAK	0.95	2.000
⋮	⋮	⋮

Table 26: (continued)

Item	Odds ratio	P
⋮	⋮	⋮
PASTRAMİ	0	2.000
SOSİS	1.22	2.000
YEŞİL BİBER	1.12	2.000
KARABİBER	0	2.000
PATATES	0.8	2.000
FÜME HİNDİ	1.04	2.000
KAŞAR PEYNİRİ	0.8	2.000
YUMURTA	0.56	2.000
MISIR	Inf	2.000
MAYONEZ	0	2.000
MUZ	0	2.000
HAMUR	1.04	2.000
ANTRİKOT	0	2.000
KETÇAP	0	2.000
BEYAZ PEYNİR	1.02	2.000
GRAVYER PEYNİRİ	0	2.000
KUŞKONMAZ	0	2.000
EZİNE PEYNİRİ	0	2.000
ÇARLİSTON BİBER	0	2.000
YERFISTIĞI	0	2.000
PORÇİNİ MANTARI	0	2.000
SALÇA	0	2.000
LEVREK	0	2.000
FONTİNA PEYNİRİ	0	2.000
SOYA SOSU	0	2.000
FINDIK	0	2.000
TAHİN	0	2.000
TAHİL	0	2.000
DOMUZ	0	2.000
ZAHTER	0	2.000

Table 27: Pita restaurants' menu term existences compared with two education clusters using Fisher's exact test. P-values are Bonferroni-corrected. An odds ratio higher than 1 indicates that the item is more common for the less educated districts' cluster while an odds ratio lower than 1 indicates that the item is more common for the more educated districts' cluster.

Item	Odds ratio	P
PATATES KIZARTMASI	0	.048
KÖZLENMİŞ BİBER	12.15	.058
CACIK	0	.083
BEYAZ PEYNİR	0.37	.107
SOĞAN	0.17	.150
PATATES	0.32	.170
KIYMA	3.55	.210
SALATA	0.54	.277
MANTAR	0.55	.307
KÖFTE	0	.408
ÇİĞ KÖFTE	0	.408
CEVİZ	Inf	.420
SEBZE	2.01	.554
MAYDANOZ	0.29	.603
LİMON	0	.687
DÖNER	0.52	.817
DOMATES	0.63	.973
PEYNİR	1.58	.985
EZME	1.4	1.011
SOSİS	0	1.159
MEZE	0	1.159
YAĞ	0	1.168
TURŞU	0	1.168
TARATOR	1.94	1.211
SUCUK	1.35	1.299
KAVURMA	1.27	1.348
TAVUK	0.79	1.356
YEŞİL BİBER	0.4	1.377
DANA	1.28	1.431
ET	0.8	1.618
YUMURTA	0.88	1.674
ISPANAK	0.83	2.000
KARABİBER	0	2.000
NANE	0	2.000
KAŞAR PEYNİRİ	0.79	2.000
SUMAK	0	2.000
ANTRİKOT	0	2.000
ZEYTİN	0	2.000
KABURGA	0	2.000
İRMİK HELVASI	0.79	2.000
⋮	⋮	⋮

Table 27: (continued)

Item	Odds ratio	P
∴	∴	∴
SUMAKLI SOĞAN	0	2.000
DEREOTU	0	2.000
SALAM	0	2.000
KIRMIZI SOĞAN	0	2.000
YOĞURT	0.75	2.000
SOS	1.26	2.000
KUZU	0	2.000
PATLICAN	0.75	2.000
KEKİK	0	2.000
TAHİL	0	2.000
PASTIRMA	0.87	2.000
ÇÖREK OTU	0	2.000
OTLU PEYNİR	0	2.000
FESLEĞEN	0	2.000
SUSAM	0	2.000
BAHARAT	0	2.000
İSOT	0	2.000
HELLİM PEYNİRİ	0	2.000
KIRMIZI BİBER	0	2.000
HAMUR	0	2.000
MARUL	0	2.000
KETÇAP	0	2.000
TEREYAĞI	0.94	2.000
KOKOREÇ	0	2.000
KAPYA	0	2.000
BONFİLE	0	2.000
ÇEDAR PEYNİRİ	0	2.000
BİBER	0.88	2.000
KADAYIF	0	2.000

APPENDIX C

COMMENT TOPICS

Top 10 most relevant topic terms are provided for pizza and pita restaurants below. To order the terms based on their relevance, the relevance score mentioned in Section 3.2.4 is used with $\lambda = 0.6$. Each topic has its terms listed in descending order.

Table 28: Top 10 most relevant terms for each topic obtained from pizza restaurant reviews

Topic	Term	Topic	Term	Topic	Term	Topic	Term
Topic 1	gelmek yok hamur gibi peynir ust soguk tat uzeri ekmek	Topic 2	cok ben icin lezzet bu daha guzel vermek siparis iyi	Topic 3	cok guzel iyi gercekten lezzet begenmek pizza hersey tat gayet	Topic 4	hiz sicak gelmek lezzet cok tesekkur gayet sekil siparis servis
Topic 5	kurye getirmek kibar cok bey arkadas abi tesekkur guleryuz nazik	Topic 6	demek gelmek getirmek para aramak siparis saat kabul ben soylemek	Topic 7	daha biraz fazla iyi guzel dis gelmek cok dikkat patates	Topic 8	malzeme peynir koymak konmak ekstra hamur soylemek istemek ust secmek
Topic 9	bol lezzet sicak guzel dolu harika malzeme gelmek hiz yemek	Topic 10	bu kadar kotu yemek musteri daha para boyle iyi siz	Topic 11	patates kucuk menu kadar boy tane almak buyuk yazmak kisilik	Topic 12	gayet lezzet hiz guzel basari servis doyurmak iyi gelmek fiyat
Topic 13	gelmek gec soguk gecmek saat biraz cok yavas sogumak sicak	Topic 14	yemek bura hep favori artik sevmek yer en tercih yapmak	Topic 15	hiz mukemmel lezzet super servis numara her harika hersey hizmet	Topic 16	saat dakika gelmek sonra siparis tam sure buz yaklasik yarim
Topic 17	yag kotu tat cok tuz patates koku yemek fazla asiri	Topic 18	tesekkur icin hiz ilgi dolayi bey guleryuz kurye abiye nezaket	Topic 19	lezzet kalite fiyat yok bu urun malzeme zaman tat ayni	Topic 20	yemek icin catal soylemek baska kalmak kasik urun daha yer
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮

Table 28: (continued)

Topic	Term	Topic	Term	Topic	Term	Topic	Term
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
Topic 21	siparis vermek ilk once daha bura kez defa son yer	Topic 22	saglik el emek cok hersey guzel herkes tesekkur yine usta	Topic 23	mi bilmek diye bu degismek yapmak anlamak yok boyle mu	Topic 24	bu kadar sefer sicak gelmek hiz nasil yemek sekil kenar
Topic 25	yasamak aramak sube sorun bilgi siparis bu telefon sikinti vermek	Topic 26	daha iyi gore cok kotu degil once beklemek gelmek asla	Topic 27	her zaman gibi sey guzel mukemmel cok harika sefer ayni	Topic 28	boy orta kucuk buyuk boyut porsiyon kuculmek soylemek kisi doymak
Topic 29	malzeme az miktar bol cok azmak azalmak biraz daha gore	Topic 30	tat damak ben pek begenmek peynir zevk sos agir almak	Topic 31	salata ic domates et marul biraz tavuk daha koymak peynir	Topic 32	ketcap mayonez istemek baharat ragmen koymak yan unutmak ekstra gondermek
Topic 33	kurye insan calismak guleryuz arkadas tavir saygi kibar is davranmak	Topic 34	ketcap mayonez mendil pecete islak catal yan koymak yok bicak	Topic 35	en yemek hayat kotu iyi kadar sepet azi son uzun	Topic 36	fiyat performans gore uygun yuksek pahali artmak porsiyon biraz kalite
Topic 37	malzeme bol kalite kullanmak peynir az lezzet iyi hamur cok	Topic 38	yemek diye siparis sonra kadar beklemek vermek cikmak demek bu	Topic 39	gelmek sicacik getirmek duman hiz sicak arkadas kurye gibi yagmur	Topic 40	patates kizarmak halka sogan kizartmak elma tavuk yan citir yag
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮

Table 28: (continued)

Topic	Term	Topic	Term	Topic	Term	Topic	Term
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
Topic 41	boyle devam bozmak ummak hep kalite lutfen insallah hic cizgi	Topic 42	yanlis adres eksik gelmek siparis yazmak dogru ragmen unutmak telafi	Topic 43	yok sey demek hicbir baska zaten soz ic laf dis	Topic 44	getirmek gitmek kendi cikmak paket girmek arkadas alan arka demek
Topic 45	biber zeytin mantar yesil istemek ragmen belirmek koymak ic aci	Topic 46	servis hiz konu lezzet iyi eleman cok paket sikinti gayet	Topic 47	aramak diye ben sormak bakmak demek kapi yol goz kurye	Topic 48	sos yan barbeku sarimsak aci yogurt koymak domates gondermek yerine
Topic 49	kutu yapismak ic dagilmak birbiri folyo dokmek karton kapak aluminyum	Topic 50	yanmak kenar pismek hamur sert kurumak gibi yanik kupkuru kitir	Topic 51	ikram tatli hediye icin gondermek tesekkur kurabiye yan cok ayrica	Topic 52	not dikkat almak okumak yazmak alinmak nota dusmek lutfen yorum
Topic 53	yemek salca almak diye daha ekmek market bile ne mi	Topic 54	hamur ince kalin secmek istemek klasik ragmen normal kenar gelmek	Topic 55	joker indirim diye siparis soylemek ragmen vermek mi almak icin	Topic 56	tavuk citr katir top barbeku kanat parca sos izgara salata
Topic 57	iftar gece saat siparis yogun vermek vakit aksam kala ragmen	Topic 58	hak is bu calismak yorum vermek emek gerçekten yapmak tavsiye	Topic 59	ozur aramak canli yardim dilemek iletisim sormak gecmek telefon gecikmek	Topic 60	zincir diger marka firma kat fast food sube cogu gore
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮

Table 28: (continued)

Topic	Term	Topic	Term	Topic	Term	Topic	Term
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
Topic 61	ekmek tost peynir kasar kavurmak bazlama ara sarimsak sucuk sandvic	Topic 62	ozen paketlemek hazirlamak gostermek servis urun dikkat yapmak temiz bas	Topic 63	kenar sarimsak odemek ucret istemek ekstra normal ragmen para susam	Topic 64	once daha sefer gore gecen ayni fark kafa sene tat
Topic 65	kalmak memnun zor bu sefer cok siparis hic gec vermek	Topic 66	yapmak is bu ne siz sevmek adam bilmek zam insan	Topic 67	pasta cilek fistik cikolata mozaik muz baklava krema meyve frambuaz	Topic 68	yorum bakmak gerek yok kusur hic sapmak okumak yapmak sacma
Topic 69	para vermek almak odemek kurus karsilik hak ust bozuk degmek	Topic 70	denemek tavsiye kesin deney mutlaka deneme onermek fark pisman daha	Topic 71	kart nakit cihaz odemek pos kredi odem post makine ariza	Topic 72	isitmak firin mikrodalga cikmak gibi sanki yeni zor hazir isimak
Topic 73	baska sube bulmak yer adres yonlenmek gerek konum bulunmak ora	Topic 74	kesmek dilim dilimlemek koparmak duzgun bolmek ayirmak esit zor malzeme	Topic 75	puan kirmak dusuk vermek hatir yukse sebep neden haketmek anlamak	Topic 76	degil fena pek hos hic yenmek iyi sorun gibi yumak
Topic 77	eski kuculmek eser yok gittikce gibi azalmak artik gitmek tat	Topic 78	mesafe ev yakın uzak ragmen yurumek dakika oturmak saat gelmek	Topic 79	buz gibi tas herzaman gelmek dalga gecer kayis gorunmek yenmek	Topic 80	sure kisa uzun hava ragmen yagmur ortalama teslim dakika yazmak
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮

Table 28: (continued)

Topic	Term	Topic	Term	Topic	Term	Topic	Term
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
Topic 81	kola icecek kol seker istemek seftali icmek fanta gelmek ayran	Topic 82	dilim tane parca adet sucuk elma patates var koymak gram	Topic 83	ilk defa kez denemek hayat boyle bu karsilasmak yemek uye	Topic 84	ne var mi yazik acaba demek kadar bilmek anlamak mu
Topic 85	sucuk sosis salam karisik jambon pastirma mangal istemek sandvic kumpir	Topic 86	tek kelime mukemmel kisilik harika ayiklamak sikinti ile muhtesem adres	Topic 87	gun tarih son gecmis gecen dogum donem ay ertesi hafta	Topic 88	kral selam milliyet adam abi sen surat isinlamak agabey abiye
Topic 89	kofte hamburger ekmek et burger kantin market pinar misket gram	Topic 90	sufle cikolata dondurma irmik helva pudra kek uzeri kasik tuy	Topic 91	vicik akmak kagit poset yag su dokmek tasimak isi kese	Topic 92	calmak zil kapi basmak uyumak car bebek uyanmak telefon bayan
Topic 93	cop atmak zor kalmak direk yemek gitmek yari ayiklamak kosmak	Topic 94	doymak mide rahat gonul karin kisi yemek agri bulanmak gak	Topic 95	musteri memnuniyet sayan odak kaybetmek takdir anlayis yaklasim sergilemek daimi	Topic 96	borek kiymak kilo peynir gozleme sigara yarim ceviz kabuk lor
Topic 97	fume kaburga et roka kervansaray ancuez kebab kapari kikirdak dana	Topic 98	hayal kirik ugramak yasamak piril yaratmak mahcup buyuk sekte sefer	Topic 99	acmak vurmak takir beri tabir ters ihtiyac ses midye girc	Topic 100	ton balik belgelemek tun yilan bosalmak huzunlenmek salata gorunur yuklemek

Table 29: Top 10 most relevant terms for each topic obtained from pita restaurant reviews

Topic	Term	Topic	Term	Topic	Term	Topic	Term
Topic 1	cok guzel lezzet hiz gayet tesekkur servis gelmek sicak hersey	Topic 2	lezzet hiz cok siparis gelmek yemek her gercekten tesekkur vermek	Topic 3	gayet lezzet doyurmak guzel iyi lahmacun hiz servis porsiyon cok	Topic 4	yag az ic cok tuz et biraz kiyma pismek pide
Topic 5	lahmacun tavsiye harika lezzet efsane yemek mukemmel kesin cok guzel	Topic 6	hiz sicak gelmek lezzet cok servis gayet sicacik sekil siparis	Topic 7	cok lahmacun guzel gelmek begenmek gercekten lezzet yan tat iyi	Topic 8	biraz ezmek guzel aci dis lezzet tat salata gayet daha
Topic 9	bu gelmek siparis ragmen sefer ayni kadar yazmak once soylemek	Topic 10	aramak siparis telefon getirmek iptal icin sonra gondermek gelmek yanlis	Topic 11	bu bura ben yer favori kesin restoran siz icin islemek	Topic 12	kadar bu yemek lahmacun kotu yapmak soylemek guzel anlamak simdi
Topic 13	gelmek istemek normal menu yok yazmak var siparis kola aci	Topic 14	yan salata meze gondermek ezmek gelmek salat ikram tatli taze	Topic 15	saglik el usta hersey cok emek guzel tesekkur kol numara	Topic 16	demek ben bakmak bilmek akil once siparis aramak diye yok
Topic 17	degil fena hic gibi tat kotu eski soguk gelmek iyi	Topic 18	salata kotu pilav yan gelmek helva gibi yok bulgur ezmek	Topic 19	maydanoz limon koymak yan lahmacun sogan salata daha yesil gondermek	Topic 20	poset dokmek akmak ic kutu birbiri yapismak paket kagit acmak
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮

Table 29: (continued)

Topic	Term	Topic	Term	Topic	Term	Topic	Term
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
Topic 21	cok iyi hiz servis lezzet guzel kotu daha her sey	Topic 22	biraz guzel dis tavuk gelmek hersey fazla ekmek sadece sey	Topic 23	getirmek kurye arkadas abi kibar guleryuz cok paket nazik kardes	Topic 24	siparis vermek surekli bura once son sevmek yermek yer ilk
Topic 25	gelmek gec soguk biraz saat gecmek sogumak yogun sicak yavas	Topic 26	musteri islemek memnuniyet onem saygi bu boyle ben vermek siz	Topic 27	daha biraz fazla iyi once yag beklemek pismek pisirmek az	Topic 28	her zaman gibi sey mukemmel guzel ayni numara harika cok
Topic 29	servis hiz numara lezzet mukemmel yildiz super harika hizmet kalite	Topic 30	parca ic cikmak et kikirdak yanmak pismek yanik kemik cig	Topic 31	yemek yer bura baska gitmek sonra soylemek lahmacun sevmek guzel	Topic 32	saat iftar ragmen siparis gelmek tam vakit mesafe yakin dakika
Topic 33	gondermek mi koymak bu ayip atmak gormek oyle kokmak kadar	Topic 34	doner durum et tavuk sos ekmek lavas ic cok iskender	Topic 35	fiyat performans gore uygun pahali iyi yukse oran artmak bu	Topic 36	kasar pide kiymak kusbasi kapali soylemek yumurta mantar kavurmak istemek
Topic 37	diye mi bura lahmacun artik demek soylemek yok siz biz	Topic 38	devam ummak boyle hep bozmak insallah cizgi dilek dilemek hic	Topic 39	porsiyon kucuk buyuk gore fiyat az kuculmek boyut doyurmak biraz	Topic 40	not dikkat almak okumak yazmak icin alinmak tesekkur kisim yorum
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮

Table 29: (continued)

Topic	Term	Topic	Term	Topic	Term	Topic	Term
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
Topic 41	tesekkur ilgi icin bey alaka hizmet guleryuz kardes dolayi kurye	Topic 42	sure saat dakika gelmek kisa siparis uzun surmek sonra yaklasik	Topic 43	diger yer var restoran fark gore uzak kadar kalite urun	Topic 44	kibar kurye ozur guleryuz nazik arkadas tavir getirmek sahip dilemek
Topic 45	tavuk sis kanat kuzu pirzola ciger soylemek izgara ses pismek	Topic 46	vermek siparis tekrar daha asla dusunmek gonul rahat kesin tereddut	Topic 47	damak tat tuz ben begenmek zevk uymak pek baharat hos	Topic 48	citir hamur ince kalin lahmacun yumusak lavas pide kenar gibi
Topic 49	catal pecete mendil islak bicak kasik plastik tuz ne vs	Topic 50	kalite eski bozmak dusmek ummak degismek zaman lutfen ayni malzeme	Topic 51	icin ikram tatli tesekkur hediye ayrica gondermek puding ayni jest	Topic 52	istemek sogan belirmek ragmen not yazmak yumurta hal ekstra diye
Topic 53	gonul doymak rahat karin insan acmak) kisi goz ogrenci	Topic 54	domates biber salata salatalik marul kozlemek sogan curuk maydanoz yesil	Topic 55	hayat en yemek kotu sepet kadar uzun guzel gormek berbat	Topic 56	ilk defa kez siparis vermek denemek boyle fa bu memnun
Topic 57	irmik helva tatli ikram guzel kasik ezmek gondermek patates sekerpare	Topic 58	yok gerek demek soz hic sey laf alaka zaten baska	Topic 59	bol malzeme dolu ikram ic meze bolmek gelmek salata kismak	Topic 60	koku tat var kokmak garip et tuhaf agir tarih eksi
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮

Table 29: (continued)

Topic	Term	Topic	Term	Topic	Term	Topic	Term
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
Topic 61	ayran kola unutmak icecek fanta menu istemek salgam icmek yerine	Topic 62	tane menu adet soylemek kisi kisilik lahmacun ayran ragmen yazma	Topic 63	yol dakika yurumek gelmek kadar cikmak gelir beklemek kapi saat	Topic 64	yasamak sikinti sorun konu yok ufak tefek problem hicbir ummak
Topic 65	en iyi bolge civar azi cevre biri acik yemek lahmacun	Topic 66	ekmek parca kesmek koca dilim koymak dilimlemek tirnak yarim kucuk	Topic 67	puan dusuk kirmak yuksek vermek neden haketmek sebeup anlamak bu	Topic 68	kalmak memnun zor cok geri sinif gec damak berber memnu
Topic 69	is yapmak hak adam sevgi bu bilmek siz layikiyla zam	Topic 70	aci istemek lahmacun ragmen soylemek biber gelmek isot belirmek secmek	Topic 71	joker indirim ragmen soylemek diye faydalanmak ile siparis vermek almak	Topic 72	kunefe serbet kaymak kadayif fistik peynir vicik hazir market seker
Topic 73	mide yag agri bulanmak bulanti agiz gun agrimak sonra kotu	Topic 74	para hak helal vermek haketmek karsilik almak kurus ucret son	Topic 75	karton kutu kap tabak aluminyum kopuk kagit koymak plastik sarmak	Topic 76	gibi buz herzaman kayis sanki isitmak tas yemis gelmek mikrodalga
Topic 77	tek kelime mukemmel harika ile adres ote muhtesem kisilik elestirim	Topic 78	sefer bu sonra kadar ummak siz siparis nazar lutfen gecen	Topic 79	ozen hazirlamak gostermek paketlemek yapmak paket daha servis sagmak sekil	Topic 80	emek hayir herkes kazanc gecen is tavsiye gecmek dilemek bol
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮

Table 29: (continued)

Topic	Term	Topic	Term	Topic	Term	Topic	Term
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
Topic 81	cop atmak gitmek zor yemek yari yazik gunah direk para	Topic 82	corba paca mercimek kelle icmek iskembe tuzlamak su sirke sarimsak	Topic 83	karisik izgara kebap patlican kofte kiremit soylemek vali sigara menu	Topic 84	kart cihaz pos nakit odemek kredi post cekmek secmek makine
Topic 85	pilav bulgur pirinc fasulye lapa diri ispir ust kuru sut	Topic 86	tam pismek kivam anlam sutlac firin gibi fiyasko fevk karar	Topic 87	yorum bakmak yapmak aldanmak okusur kusur cevap puan olumsuz olumlu	Topic 88	kofte cig ic mini corek kasap hamburger tahin sayan borek
Topic 89	mi bilmek hangi acaba mu ne nasil marka degismek anlamak	Topic 90	tutar minimum limit fici ucet gum kampus gonderim inmek hey	Topic 91	goz lazim ramazan acilen acil ay bazen elbette sifa art	Topic 92	calmak zil basmak kapi bangir uyumak telefon canli aramak acmak
Topic 93	ust alt ortalama karman corman orta beklenti okur duman kismen	Topic 94	yeni firin cikmak bayram tasinmak siki kosmak tasimak gun cicar	Topic 95	patates kizarmak kizartmak ketcap mayonez gozleme cacik pure sos haslamak	Topic 96	ne var yazik tak coluk piskin cildirmek kaldirmek duldul uyumak
Topic 97	ayri yogurt kesik manti kebap kap huzur civik paragraf gelmic	Topic 98	kirik hayal ugramak topak torpulemek tava husran yaratmak paramparca ben	Topic 99	ara bicak dag fark aman acik ekmek sira kapsam nisan	Topic 100	pudra kut girmek hatir seker bile acmak turku ayip merkez