PREDICTING THE CATERING LOAD ON INFLIGHT RETAIL PLATFORMS

(UÇAK İÇİ SATIŞ PLATFORMLARINDA İKRAM TAHMİNLEME)

by

Salih Enver YURTER, B.S.

Thesis

Submitted in Partial Fulfillment

of the Requirements

for the Degree of

MASTER OF SCIENCE

in

COMPUTER ENGINEERING

in the

GRADUATE SCHOOL OF SCIENCE AND ENGINEERING

of

GALATASARAY UNIVERSITY

Oct 2024

ACKNOWLEDGEMENTS

First and foremost, I would like to express my deepest gratitude to my supervisor, Asst. Prof. Reis Burak Arslan, for his unwavering guidance, encouragement, invaluable insights throughout the course of this research and belief in me. His deep commitment to academic excellence and meticulous attention to detail have significantly shaped this dissertation.

I extend my heartfelt thanks to the members of my thesis committee, Asst. Prof. Atay ÖZGÖVDE and Assoc. Prof. Gülfem ALPTEKİN for dedicating their time and effort to evaluate my work. Your thoughtful questions, constructive feedback, and professional perspectives inspired me to think critically and refine my approach. I deeply appreciate your commitment and contributions to the academic rigor of this thesis.

I am deeply grateful to my family and friends for their encouragement and support all through my studies.

Oct 2024
Salih Enver YURTER

TABLE OF CONTENTS

LIST OF SYMBOLS

AI **:** Artificial Intelligence
ANN **:** Artificial Neural Network
App **:** Appendix
FSC **:** Full-Service Carriers
IATA **:** International Air Transport Association
ICAO **:** International Civil Aviation Organization
LCC **:** Low Cost Carriers
BoB **:** Buy on Board
XGBoost **:** Extreme Gradient Boost
RMSE **:** Root Mean Square Error
Inflight **:** During the flight
R^2 : R Square value
PAX : Passenger
DEP : Departure
ARR : Arrival
ReLU : Rectified Linear Unit
SGD : Stochastic Gradient Descent
Adam : Adaptive Moment Estimation
Orig : Origin
Dest : Destination
FLT_LINE : Flight Line Domestic/International
Top-up : Fill exhausted sources
Uplift : Load full catering
DT : Decision Tree
MAPE : Mean Absolute Percentage Error
SLR : Simple Linear Regression
MLR : Multiple Linear Regression

LIST OF FIGURES

LIST OF TABLES

ABSTRACT

In the competitive landscape of commercial aviation, efficient resource management is pivotal for both environmental sustainability and economic performance. This thesis introduces a novel application of AI models in predicting food and beverage consumption for buy-on-board services across different flight legs. By analyzing historical sales data, the AI model forecasts total consumption, enabling airlines to optimize provisioning and reduce waste. This is critical, as excess inventory not only leads to increased wastage of perishable goods but also contributes to higher fuel consumption due to additional weight. This research not only advances the application of AI in aviation but also serves as a blueprint for enhancing onboard service logistics and sustainability practices.

This research focuses on fresh food prediction in BoB system. Fresh Food sales data is used for training and development purposes. Historical sales data is used from a commercial airline operating in Turkiye. Two models are selected for this approach; one is XGBoost and other one is neural networks. These models were selected for their differing strengths in handling complex, nonlinear data relationships. The comparative analysis focused on evaluating their predictive accuracy, robustness, and efficiency in operational environments. Results indicate that while both models enhance forecasting precision over traditional methods, each offers unique advantages that can be leveraged depending on specific operational needs.

To ensure the relevance and accuracy of the predictions, the research analyzes data from the last two years only, a period following the COVID-19 pandemic when BoB services were gradually reinstated. This decision was guided by the significant impact of the pandemic on airline services, where BoB was either forbidden or severely restricted, and subsequent consumer hesitance to purchase from BoB systems. This specific focus helps isolate the effects of normalizing consumer behavior and operational changes post-

pandemic, ensuring that the AI models' predictions are grounded in the current operational realities of airlines.

The research shows that after training and evaluating both models, XGBoost and the deep learning neural network are suitable for predicting fresh food consumption with considerable accuracy. Both models demonstrate robust performance, with implications for real-world application in optimizing airline food services. Additionally, the predictive models are designed to improve as new sales data becomes available, suggesting that ongoing data collection and model retraining will further enhance the system's efficiency and reliability, aligning with dynamic market conditions and consumer preferences.

RÉSUMÉ

Dans le paysage compétitif de l'aviation commerciale, la gestion efficace des ressources est cruciale tant pour la durabilité environnementale que pour la performance économique. Cet état de fait est l'occasion d'une nouvelle application de modèles d'IA dans la prévision de la consommation de nourriture et de boissons pour les services à bord sur différents segments de vol. En analysant les données historiques de ventes, le modèle d'IA prévoit la consommation totale, permettant aux compagnies aériennes d'optimiser l'approvisionnement et de réduire les déchets. C'est crucial, car un excès d'inventaire conduit non seulement à un gaspillage accru de produits périssables mais contribue également à une consommation de carburant plus élevée en raison du poids supplémentaire. Cet état de fait non seulement fait avancer l'application de l'IA dans l'aviation mais sert également de modèle pour améliorer la logistique des services à bord et les pratiques de durabilité.

Cet état de fait se concentre sur la prévision de nourriture fraîche dans le système BoB. Les données de ventes de nourriture fraîche sont utilisées à des fins de formation et de développement. Les données de ventes historiques proviennent d'une compagnie aérienne commerciale opérant en Turquie. Deux modèles ont été sélectionnés pour cette approche ; le premier repose sur XGBoost tandis que le second s'appuie sur les réseaux neuronaux. Ces modèles ont été choisis pour leurs forces différentes dans la gestion des relations de données complexes et non linéaires. L'analyse comparative s'est concentrée sur l'évaluation de leur précision prédictive, de leur robustesse et de leur efficacité dans les environnements opérationnels. Les résultats indiquent que chacun offre des avantages uniques qui peuvent être exploités en fonction des avantages spécifiques bien que les deux modèles améliorent la précision des prévisions par rapport aux méthodes traditionnelles.

Pour garantir la pertinence et la précision des prédictions, cet état de fait analyse les données des deux dernières années uniquement, une période suivant la pandémie de

COVID-19 lorsque les services BoB ont été progressivement rétablis. Cette décision a été guidée par l'impact significatif de la pandémie sur les services aériens, où le BoB était soit interdit soit sévèrement limité, et par l'hésitation subséquente des consommateurs à acheter dans les systèmes BoB. Cet accent spécifique aide à isoler les effets de la normalisation du comportement des consommateurs et des changements opérationnels post-pandémie, garantissant que les prédictions des modèles d'IA sont ancrées dans les réalités opérationnelles actuelles des compagnies aériennes.

Cet état de fait montre qu'après la formation et l'évaluation des deux modèles, tant XGBoost que le réseau neuronal de deep learning conviennent pour prédire la consommation de nourriture fraîche avec une précision considérable. Les deux modèles démontrent une performance robuste, avec des implications pour l'application dans le monde réel dans l'optimisation des services alimentaires des compagnies aériennes. De plus, les modèles prédictifs sont conçus pour s'améliorer au fur et à mesure que de nouvelles données de ventes deviennent disponibles, suggérant que la collecte de données continue et la réentrainement des modèles amélioreront encore l'efficacité et la fiabilité du système, s'alignant sur les conditions de marché dynamiques et les préférences des consommateurs.

ÖZET

Ticari havacılıkta rekabetçi bir ortamda, etkin kaynak yönetimi hem çevresel sürdürülebilirlik hem de ekonomik performans için kritik öneme sahiptir. Bu tez, gerçekleşen her bir uçuş seferi kapsamında satın alma hizmetleri için yiyecek ve içecek tüketiminin tahmin edilmesinde yapay zekâ modellerini kullanılmasını hedeflemektedir. Tarihsel satış verilerini analiz ederek, yapay zekâ modeli toplam tüketimi tahmin etmekte, böylece havayollarının tedarikini optimize etmesine ve atığı azaltmasına olanak tanımaktadır. Uçağa fazla yüklenen yiyecek ve içecekler sadece çabuk bozulabilen malların israfını arttırmakla kalmayıp aynı zamanda ekstra ağırlık nedeniyle yakıt tüketimini artırmasına da sebep olmaktadır. Bu çalışma, havacılıkta yapay zekâ uygulamalarının kullanımına bir örnek oluşturmakta ve uçuş içi hizmet lojistiği ve sürdürülebilirlik konularına da katkıda bulunmaktadır.

Bu çalışma, uçak içi satış sisteminde taze yiyecek tahmini üzerine odaklanmaktadır. Geçmişe dönük taze yiyecek satış verileri, eğitim ve geliştirme amaçları için kullanılmıştır. Tarihsel satış verileri, Türkiye'de faaliyet gösteren ticari bir havayolu şirketinden alınmıştır. Bu yaklaşım için iki model seçilmiştir; biri XGBoost ve diğeri ise yapay sinir ağlarıdır. Bu modeller, karmaşık, doğrusal olmayan veri ilişkilerini yönetme yeteneklerindeki farklı güçleri nedeniyle seçilmiştir. Karşılaştırmalı analiz, onların tahmin doğruluğunu, sağlamlığını ve operasyonel ortamlardaki verimliliklerini değerlendirmeye odaklanmıştır. Sonuçlar, her iki modelin de geleneksel yöntemlere kıyasla tahmin hassasiyetini artırdığını, her birinin özgün avantajlar sunduğunu ve bu avantajların belirli operasyonel ihtiyaçlara bağlı olarak değerlendirilebileceğini göstermektedir.

Tahminlerin ilgili ve doğru olmasını sağlamak için, çalışma sadece son iki yılın verileri kullanılmıştır. Buradaki temel yaklaşım Covid Pandemi sürecinin uça içi satış performansına etkisinden kaçınmaktır. Bu kararla, uçak içi satışın ya yasaklandığı ya da

ciddi şekilde kısıtlandığı, ardından tüketicilerin uçak içi satış sistemlerinden satın alma konusunda tereddütlü davrandıkları dönemin verileri kapsam dışında bırakılmıştır. Bu yaklaşım sayesinde, pandemi sonrası tüketici davranışlarının ve operasyonel değişikliklerin etkileri izole edilmiş, böylece yapay zekâ modellerinin tahminlerinin havayollarının mevcut operasyonel gerçeklikleriyle uyumlu olması hedeflenmiştir.

Çalışma, her iki modelin de eğitim ve değerlendirme sonrasında taze yiyecek tüketimini önemli bir doğrulukla tahmin etmek için uygun olduğunu göstermektedir. Her iki model de etkin bir performans sergilemekte olup, havayolu yiyecek hizmetlerini optimize etme konusunda gerçek dünya uygulamaları için tutarlı sonuçlar sunmaktadır. Ayrıca, tahmin modelleri, yeni satış verileri mevcut hale geldikçe iyileşecek şekilde tasarlanmıştır, bu da sürekli veri toplama ve model yeniden eğitiminin, dinamik pazar koşulları ve tüketici tercihleri ile uyumlu olarak sistemin verimliliğini ve güvenilirliğini daha da artıracağını önermektedir.

1. INTRODUCTION

1.1 Motivation

The airline industry is constantly seeking innovative methods to enhance customer satisfaction and maximize revenue streams. Ancillary revenues become critical each year, and airline companies put a significant effort to increase the ancillary revenue percentage on total revenues. Table 1.1 show the percentage of ancillary revenue in total revenues for top 10 airlines. Some of the airlines generates their revenue from ancillary more than scheduled services (tickets). The table shows that most of the airlines increased the total percentage of ancillary revenue from 2022. This trend is not a temporary situation for airlines, but it keeps increasing globally. Figure 1.1 show the global trend for airline industry between years 2013 and 2022. Even in post Covid period (2020-2021) total percentage of ancillary revenue among total revenue is increased. Airline companies will continue to improve their ancillary services to maximize their revenues.

Table 1.1: Top 10 Airlines - Ancillary Revenue as % of Total Revenue (IdeaWorksCompany, 2024)

Rank	Airlines	2023 Result	2022 Result	Change from 2022
1	Spirit	56.4%	51.5	+4.9 points
2	Frontier	56.2%	50.8	+5.4 points
3	Breeze	51.3%	38.9	+12.4 points
4	Allegiant	49.8%	48.9	+0.9 points
5	Volaris	48.7%	41.4	+7.3 points
6	Viva Aerobus	45.5%	44.5	+1.0 points
7	Wizz Air	44.7%	48.0	-3.3 points
8	Flair Airlines	40.0%	No data	n/a
9	Azul	37.4%	26.5	+10.9 points
10	easyJet	36.1%	33.9	+2.2 points

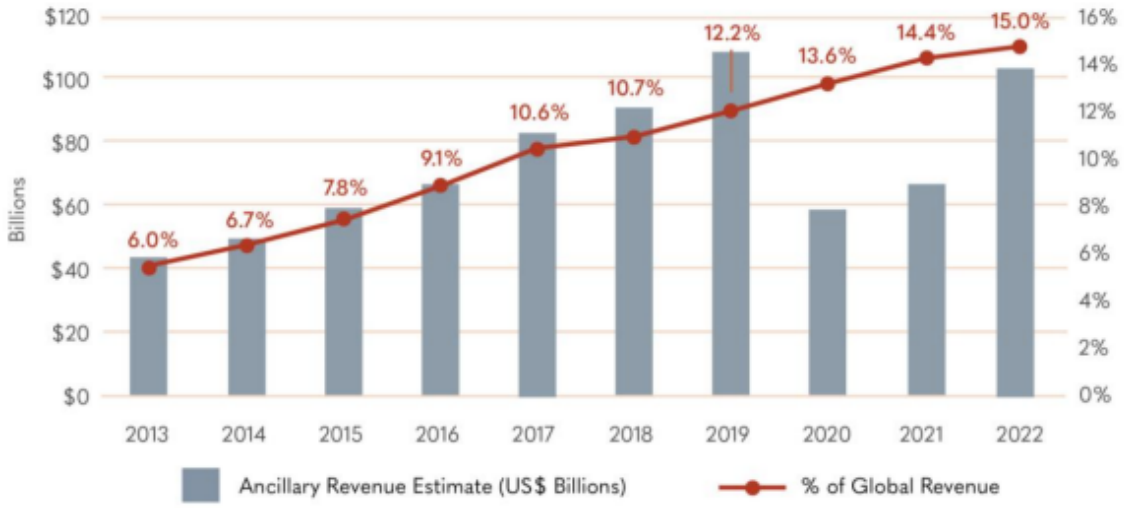


Figure 1.1: Worldwide Estimate of Ancillary Revenue (IdeaWorksCompany, 2022)

Inflight retail is the biggest subset of ancillary services for airline companies. Inflight retail plays a crucial role in increasing the total ancillary revenue by providing passengers with a wide array of products and services during their journey. Efficiently managing inflight retail requires accurate predictions of the load that can be purchased during the flight. There are many factors affect the customer demand during the year. Catering providers help airline companies to uplift and collect the products to the aircrafts. Figure 1.2 shows the process for catering provider to uplift all items to aircraft and then return. Most of the items are refundable during those processes and companies are well organized to manage those services. However, some items are exceptional cases because of external factors or nature of the product.

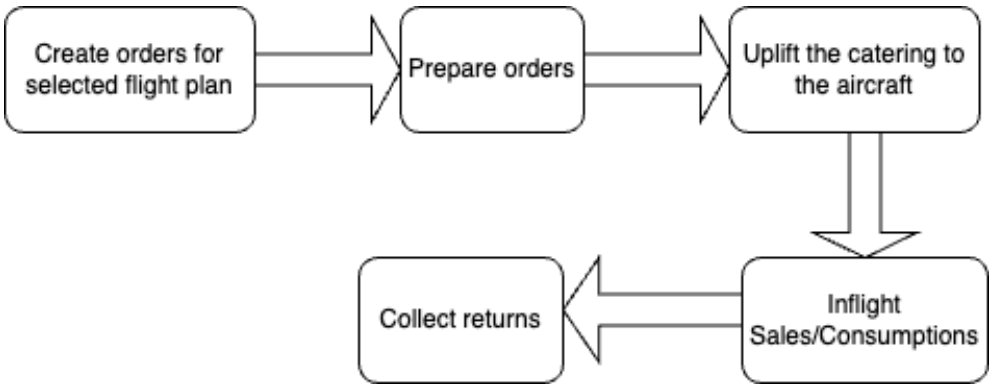


Figure 1.2: Airline Catering Process

Fresh food items are one of the exceptional key items for airline catering operations. Loading the correct number of Fresh Food items are important for two factors. One of them is nature of the product. Fresh food products are consumable in limited time range depending on storage conditions (dry ice etc.). Refund of return items is not an option since they will not be consumable after flight. Arline company pays all Fresh Food products whether they are sold or not. Another important factor is CO2 emissions. A lot of regulations are in progress to reduce carbon footprint of airline companies. IATA member airlines are committed for zero carbon by 2050 (Mason, 2005). So, reducing the wastage is another target for airline companies.

In recent years, advancements in artificial intelligence (AI) and machine learning techniques have provided new avenues for accurate and data-driven predictions. AI models, such as regression, random forest, gradient boosting, and neural networks, offer powerful tools to analyze complex datasets and identify patterns. These models, when combined with historical inflight retail data, passenger demographics, and weather information, can yield precise load predictions.

The objective of this research is to develop a predictive model that forecast the load of inflight retail purchases for Low-Cost Carriers. By leveraging historical data from an airline's inflight retail transactions and weather databases, this research aims to create an accurate and robust model capable of anticipating passenger behavior patterns. The proposed model will employ a combination of statistical methods and AI techniques, such as gradient boosting or neural networks, to capture the complex relationships between inflight retail load, and other relevant features.

The dataset utilized in this research comprises sales records of fresh foods, categorized specifically by flight route, capturing the nuances of consumer purchasing behaviors in different geographic and operational contexts. The data spans the most recent two-year period, a deliberate selection aimed at avoiding the anomalous conditions influenced by the COVID-19 pandemic. During the early stages of this period, buy-on-board (BoB) services were completely prohibited due to health and safety regulations, which significantly impacted the availability of onboard food services. As restrictions eased,

BoB operations resumed but were initially limited to a reduced range of products, reflecting cautious adjustments to new operational guidelines and consumer hesitance.

The dataset is structured around four main categories of fresh foods. These categories are consistently maintained across different seasons, although the specific types of fresh foods available vary seasonally. This categorization helps in analyzing trends and patterns in food consumption without the fluctuations that individual product variations might introduce. Each category aggregates different items, ensuring that despite seasonal changes in the food offerings, the analysis remains structured and focused on broader consumption patterns.

This focus on recent data is essential as it more accurately reflects the current consumer behavior and operational realities post-pandemic. Customer purchasing habits have undergone significant changes during the pandemic, and it took some time after the resumption of normal services for patterns to stabilize near pre-pandemic levels. The analysis of this dataset will provide insights into how consumer behaviors have adapted and stabilized in the new normal, allowing for more accurate predictions and strategic adjustments in BoB services.

Passenger preferences are critical on demand fluctuations, dietary actions, food habits etc. all have significant effect (You et al., 2020). Airline companies try to increase customer satisfaction by applying over catering strategy. Any fluctuations will increase the existing over estimated items dramatically and this results with wastage and fuel expenses (Garcia-Garcia, 2015). Catering services are not present for all destinations that airlines operate. Each airline company plans catering uplifts according to their flight routes and predetermined uplift bases. In this flight pair last legs are sensitive to pax changes since there is no option to top-up or cancel the catering that is already uplifted in the beginning. Optimizing the uplift locations significantly reduces costs and wastage (Yilmaz and Yücel, 2021).

1.2 Structure of the Thesis

The rest of this thesis is organized as follows:

- The problem statement is defined in Chapter 2, and a review of related studies is presented to provide context and background for the study.
- The solution methodology is described in Chapter 3, covering the detailed explanations of the techniques and approaches applied through the study.
- The dataset analysis is explored in the next chapter including the characteristics and attributes of the data used in the study.
- The results from selected models are evaluated according to the selected evaluation metrics in the Chapter 5.
- The conclusions based on the results presented in the previous chapter discussed and recommendations for future work are provided in Chapter 6.

2. LITERATURE REVIEW

2.1 Problem Statement

Predicting the load plays important role on maximizing the revenue for an airline. The first effect of prediction is maximizing the profit by loading enough products to supply customer demand and reducing the waste. In 1994, waste from operations at Heathrow Airport approximately 35.000 tons and 70% of all waste generated through catering offices (Coggins, 1994). Also annually, approximately 1.3 billion tons of edible food is produced and one third of them end up as waste (Garcia-Garcia, 2015).

The indirect effect of load prediction is it is effect on fuel consumption. According to IATA reports fuel expenses is 30% of all operating expenses (Garcia-Garcia, 2015). To load retail products, airline companies use trolleys. All products must be placed inside the trolleys. So, loading excessive amount of retail items increases weight with both own weights and trolley weights. Reducing the waste becomes important target for airlines.

According to business models, Carriers are divided into two categories: Low-Cost Carriers (LCC) and Full-Service Carriers (FSC) (InterVistas, 2016). Developing prediction model strictly depends on airline business model. While working with FSCs, product substitution becomes an important factor. Since FSCs offer meals for each customer, customer tend to take the meal even they do not need at that time. So FSCs need to load excessive safety stock. To minimize this load FSCs considers product substitution while developing prediction models (Walt and Bean, 2002).

For LCCs, customers tend to purchase the exact same product they want and not to make purchases otherwise. In their research A. Van der Walt and W.L. Bean finds out that substitute products increase post-consumer food waste in the form of left-over food items (Walt and Bean, 2002). Another result is substitute product usage is inconclusive since

the decision about the menu items are depend on catering or airline company subjective factors. They found out that substitute products increase the performance in cost of decreasing the reliability. In this research we will develop a model for LCCs and we do not consider substitute products for the model.

Customer travel purpose in LCCs has significant effect on their purchase behaviors. Leisure traveler expectations start with minimum expenses especially in the beginning of departure flight. Business customers are less sensitive to price, and this is a common finding. They are more likely spend inflight and their habits are more stable. Leisure travelers are more likely to spend their money on beverages whereas business travelers are more focused on special services. According to IATA only 15% of the airlines' capacity is allocated to Business Class but it generates 28% of revenue (Mason, 2005).

Traditional approaches primarily consider factors such as flight duration, destination, and passenger demographics. However, some countries and their airlines have more than traditional factors. Especially passenger demographics may change each season and year. Passengers can select an airline for business trips during winter and same airline for leisure during the summer. Some sort of passengers choose the airline for leisure and nothing else. All seasonal changes promote different set of products for inflight retail platforms. Airline companies not only offer beverages but also, they offer sunscreen, beach towel etc. There are lots of recurring events that also affect passenger demographics so the set of products. Weather conditions also have significant impact on customer behavior, influencing purchasing decisions and preferences. Passenger behavior is another effect in food waste, dietary habits, food ordering and situational factors have significant effect (Your et al., 2020). Another research show that flight characteristics affects the consumption details. Especially for the long duration flights, optimizing the consumption increases the revenue and service quality with similar effects (Goto et al., 2004).

Meal demand fluctuations in the airline catering industry significantly effects the wastage. The meal demand change in last 24 led catering company to increase the production or cancel prepared meals. Both actions create extra wastage, and the cost is distributed between the catering company and airline (Megodawickrama, 2018). Megodawickrama

emphasizes that managing last 24 hours demand by some rules or arrangements significantly reduces the wastage.

2.2 Related Research

In their research E. Samunderu and M. Farrugia show that customer segmentation can be predictable by using Machine Learning techniques and this prediction can be done in real time (Samunderu and Farrugia, 2022). For all business models, passenger preorder also has effect on waste optimization and is a significant factor on prediction. If passengers buy preorder meals, this guarantees the consumption and reduces waste (Ernits et al., 2022). Their approach based on pattern recognitions on while purchasing the tickets. Since customer travel purpose is significant feature on inflight demand, we can build a prediction model on the catering load. Similar research took place in Australia, airline data is used for predicting the domestic passenger demand and resulted with high success rates with ANNs (Srisaeng, 2015).

Catering operations are not available for all airports that airlines operate. Each airline makes agreements with airport catering operators according to their flight plans and uplift the ordered caterings from predetermined airports. The uplifted catering covers multiple flight legs until the aircraft reaches another predetermined airport for top-up or uplift. Airlines operational changes on flight plans and last-minute passenger number increases due to operational or standard factors lead to over catering situations. In their research Yılmaz and Yücel shows that increasing or adjusting the snack loading locations by correct optimization strategy reduces the airline costs (Yılmaz and Yücel, 2021). This study helps decreasing the wastage, but the targeted airline was not LCC. Their future work suggestion is to extend the study to cover LCCs.

Selecting the suitable prediction model is key factor in the study for meal demand. Bozkir and Sezer study Decision Tree models to predict the food demand in food courts. In their study they used Classification and Regression Tree (CART), Chi Squared Automatic Interaction Detection (CHAID) and Microsoft Decision Trees (MSDT) models with two years historic data from a university food court. Study results demonstrates that DT technology is suitable for food consumption prediction (Bozkir and Sezer, 2010).

Levis and Papageorgiou evaluates the SVRs on customer demand forecasting in their study. Their prediction model based on historical sales data. Their proposed algorithm shows adaptive and flexible regression function that extracts the customer demand patterns and captures the customer behavior with a prediction accuracy of 93% in all cases (Levis and Papageorgiou 2005).

In their study Geem and Roper proposed an ANN model to estimate the South Korea's energy demand with leveraging features like GDP, population, imports and exports. The ANN model outperformed traditional methods like linear regression and exponential models. RMSE is used as evaluation metric. The ANN model was successful even predicting the spikes in the demand during the year without showing the signs of overfitting (Geem and Roper, 2009).

Another approach is to predict the sales rather than demand. Silva and Cordoso studies the influence of external environment on a supermarket chain store's sale performance. A regression tree is used to predict the store sales performance. They found out that predictions are influenced by external factors such as lack of visibility, heavy road traffic. They demonstrated that proposed tree model is used successfully in predicting the performance of new stores (Silva and Cardoso, 2005).

In similar study, Jain, Menon and Chandra compare XGBoost predictor with traditional regression algorithms Linear Regression and Random Forest Regression. Their study focuses on forecasting potential sales for retail outlets of a major European Pharmacy. They use prior sales data, promotions, school and state holidays, location and the time of the year. They concluded that XGBoost predictor clearly outperforms other traditional approaches (Jain et al., 2015).

Chi-Jie Lu prosed a hybrid model combines variable selection model and Support Vector Regression for computer product sales forecasting. The study shows that selecting the relevant predictor variables is important for improving the forecast accuracy. Weekly sales data for five products Notebook, LCD, MB, HD and DC is used for training. Study shows that proposed hybrid model outperforms traditional methods and provides insights into variable importance (Levis and Papageorgiou 2005).

Addressing challenges in sales forecasting for short time-series data, Croda, Romero and Morales studied on the applicability of neural networks, specifically a multilayer perceptron, to predict sales accurately. The study highlights the independence of neural networks from statistical assumptions, making them better candidate then the traditional time-series methods. Their results, with a validation error of less than five percent, show that neural networks can provide accurate predictions even with limited data (Croda et al., 2017).

Emirates uses AI models developed by a third company that uses intelligent cameras, smart scales and meters to analyze most wasted food items. Emirates successfully applied the technology, and they managed to reduce the food waste by %35 in all their operations (Caswell, 2020).

3. SOLUTION METHODOLOGY

The solution methodology for predicting fresh food consumption in airlines is structured to ensure a comprehensive, step-by-step approach that maximizes predictive accuracy and operational relevance. This section outlines the entire process, from initial data preparation to the final model training and evaluation, providing a clear roadmap for implementing the proposed solution. The methodology is built around a multi-stage framework designed to systematically address the challenges inherent in demand forecasting. The main phases of the solution include Data Preparation, Dataset Analysis, and Model Training, each contributing to the robustness of the final predictive models.

Data Preparation phase involves gathering, cleaning, and transforming the raw data into a format suitable for analysis. Techniques such as handling missing values, data normalization, and feature encoding are employed to ensure that the dataset is ready for accurate model training. Data Preparation ensures that raw data is transformed into a model-ready format through techniques such as feature scaling and encoding, which are detailed in the Data Preparation subsection.

Before training the models, a broad analysis of the dataset is conducted to understand its underlying structure in the Dataset Analyses Section. This includes statistical summaries, distribution checks, and correlation analysis. The goal is to identify relationships between features, assess potential multicollinearity, and uncover any hidden patterns that may inform feature selection or engineering. Dataset Analysis builds on this by performing exploratory data analysis (EDA) to extract valuable insights. This step involves plotting distributions, visualizing feature correlations, and conducting tests for feature importance.

Model Training stage involves training and evaluating predictive models. Model Training utilizes the findings from the analysis phase to construct and optimize models.

The training processes for both models are clearly explained, including hyperparameter tuning and the strategies used to avoid overfitting and improve generalization.

3.1 Data Preparation

In the research airline inflight sales transaction is used. In order to properly analyze and use the data we need data processing.

3.1.1 Data Cleaning

The data cleaning process began with the removal of records with missing flight information. This step was essential to ensure that the analysis would be based on complete and accurate data, as the flight details are critical for contextualizing sales data within specific operational parameters. Table 3.1 shows all attributes from the original dataset with example records.

Following the removal of incomplete entries, the dataset was further scrutinized for duplicates. Duplicate records were identified and removed to prevent any bias or redundancy in the analysis, ensuring that each data point represented a unique instance of food sales.

Some missing values are collected from other records especially from the same flight. Most common missing value was PAX and Duration information. This information may be missing for a product but may be available for another product from the same flight.

An essential aspect of data cleaning involves identifying and handling outliers to improve the overall quality and reliability of the data used in predictive modeling. Outliers can distort the results of the analysis, leading to inaccurate forecasts and potentially misleading conclusions.

In this study, outliers were defined based on the sale quantities of fresh food items. Specifically, any record where the item sale quantity exceeded 20 was considered an outlier. This threshold was determined through an analysis of the distribution of sales quantities across all items and flight routes. Records showing such high sales figures

were rare and significantly higher than typical sales, indicating atypical purchasing behavior or possible data entry errors.

Table 3.1: Data attributes for BoB sales

Feature Name	Description	Example
FLT_LINE	International or Domestic flag for the flight, International flights are important since customs and passport control takes time during preflight period more than domestic flights	DOM, INT
ORIG	Starting airport of the flight	IST
DEST	Arrival airport of the flight	FRA
FLT_NO	Assigned from the carrier company it may or may not be unique	508
ITEM_NAME	Name of the fresh food	Tavuklu Sandviç
TRANSACTION_TYPE	Sale, complimantary. For operational reasons the crew can give the item without charging the customer.	SALE
UNIT_PRICE	Unit price of the item	35
SALE_TOTAL	Total basket amount of the transaction that includes this item	120
PROMOTION_AMOUNT	If there are any promotions, total amount of the promotion	5
DISCOUNT_AMOUNT	Total discount amount including promotion and other discounts	8
QUANTITY	Number of units for this transaction on selected item	2
MEAL_IBS	Total number of meals ordered to this flight from reservation system, this information is not specific to transaction but the flight itself	3
STD	Scheduled time of departure	3/13/23 5:50
ARR_DT	Arrival time	3/13/23 7:10
DURATION	Total duration of the flight	0 01:25:00.000000
PAX_MALE	Number of male passengers on the flight	90
PAX_FEMALE	Number of female passengers on the flight	80
PAX_CHILDREN	Number of child passengers on the flight	15
PAX_INFANT	0-24 months old passengers	3

The decision to remove these outliers was informed by the need to model typical consumer behavior accurately. High variance caused by extreme values could skew the performance of the predictive models, especially in cases where the goal is to understand and predict normal consumption patterns. Therefore, records with item sale quantities more than 20 were excluded from the dataset prior to any further processing.

The removal of these outliers helps to stabilize the variance in the data, providing a more homogeneous dataset that better represents the general trends and patterns in consumer behavior on different flight routes. This step ensures that the predictive models are trained on data that reflect typical sales scenarios, thereby enhancing the accuracy and reliability of the consumption forecasts.

3.1.2 Data Preprocessing

After cleaning the data, the next step was to prepare it for effective usage in predictive modeling. This involved several preprocessing techniques aimed at transforming raw data into a format more suitable for analysis.

3.1.2.1 Encoding Categorical Data

One of the most critical data in the data set is departure and arrival airport which are defined as ORIG and DEST in the dataset. Another significant attribute is type of the flight domestic or international which is defined as FLT_LINE. Those attributes are categorical data and needs to be converted to suitable format.

Time of day value which is extracted from flight date is also categorized into 4 categories. Those categories are Morning, Afternoon, Evening and Night. Those categories are introduced as new attribute to existing dataset. This data also needs encoding to use in AI models.

One Hot Encoding is one of the suitable techniques to represent categorical data for AI models. In this technique all possible values per attribute is extracted and added a new attribute and marked with 0 or 1 according to state of the record. For example, all possible departure airports are listed per record as ORIG_IST, ORIG_JFK etc. The biggest disadvantage of this technique is increased dimensionality. This encoding technique is used for all categorical attributes.

3.1.2.2 Handling Cyclic Data

The date of flight is another critical attribute for models. In this research month of date value is used to train the models. The nature of this data is cyclic and passenger information is known to be seasonal.

To address this, sine and cosine transformations were applied to the month of flight data as shown in formula 3-1 and 3-2. This method, often referred to as circular or periodic encoding, preserves the cyclical continuity of months. For example, December and January are only one month apart, but numerically they are at the extreme ends if encoded traditionally from 1 to 12. Using sine and cosine ensures that the cyclic nature of the data is maintained, thus allowing the models to more accurately understand and predict patterns based on seasonal trends.

$$\sin_{month} = sin\frac{2\pi M}{12} \tag{3-1}$$

$$\cos_{month} = cos\frac{2\pi M}{12} \tag{3-2}$$

These sine and cosine values are then used as two separate features in the dataset, replacing the original month feature. This approach not only captures the cyclical trend but also avoids introducing arbitrary distances between the end and the start of the year.

3.1.2.3 Other applications

The dataset included a variable for the promotion amount associated with each sale, which reflects the discount or special offer applied at the time of purchase.

Given the objectives of this research, it was determined that the presence of a promotion has impact on purchasing behavior. The amount does not show significant effect. To better capture this effect, a new binary variable called has_promotion was introduced.

The has_promotion variable was seamlessly integrated into the dataset by applying a conditional transformation to the existing PROMOTION_AMOUNT column. This method ensured that all information regarding promotions was retained but restructured in a way that aligns more closely with the analytical needs of this study.

The impact of Turkish national holidays on fresh food sales was identified as potentially significant, warranting inclusion in the dataset for a more accurate analysis. To integrate this aspect into the predictive models, a binary variable called is_holiday was created. This variable indicates whether a sale occurred on a Turkish national holiday.

This transformation was achieved by cross-referencing the date of each sale with a list of official Turkish national holiday dates for the corresponding years. If a sale's date matched any date on the holiday list, the is_holiday attribute for that record was set to 1. Including the is_holiday attribute allows the models to account for the potential increase in sales volume or changes in purchasing patterns during national holidays.

The day of year information, that is extracted from date of flight, is another effective attribute. Due to its larger numerical range, normalization was applied. This transformation adjusts the scale of the day of year data from a range of 0-365 to a more uniform scale of 0-1. The normalization was performed using the following formula:

$$normalized\_day\_of\_year\_ = \frac{day_{of_{year}}}{365} \tag{3-3}$$

The normalized "day of year" variable was integrated back into the dataset, replacing the original unnormalized values. This normalized attribute now serves as a continuous feature, providing a consistent input scale across all numerical data within the model.

Pax information shows number of passengers according to different categories as gender and infant types. However total pax count has significant effect so other pax count types are removed from the model training data set.

Final state of processed attributes is shown in Table 3.2.

Table 3.2: All Modified Attiributes in Dataset

Feature Name	Description	Example
FLT_LINE	Categorical data is converted into encoding	FLT_LINE_DOM=0,1
ORIG	Categorical data is converted into encoding	ORIG_IST=0,1
DEST	Categorical data is converted into encoding	DEST_FRA=0,1
flight_id	Composite key from transaction data, yyyymmdd_ORIG_DEST	20230315_IST_KZR
total_pax	Total number of passengers	190
has_promotion	shows whether promotion is available or not	TRUE/FALSE
is_holiday	show whether date is national holiday	TRUE/FALSE
duration_in_mins	total duration of the flight in minutes	280
month_sin	sine transformation	0.67
month_cos	cosine transformation	0.56
day_of_year_normalized	all values are normalized between 0,1	0.23
time_of_day	Categorical data is converted into encoding; Morning, Afternoon, Evening, Night	time_of_day_night=0,1

3.2 Dataset Analysis Techniques

3.2.1 K-Means Clustering

K-Means Clustering is an unsupervised machine learning algorithm used to partition the dataset into given number of distinct non-overlapping (k) clusters where each data point belongs only one group. Clustering group data points such that those within the same cluster are more similar to each other than those in other clusters (Lloyd, 1982). The algorithm checks all the data point with one cluster and assigns each of them to one of k clusters by minimizing the variance. This is an iterative process to minimize the sum of distances between the data points and their cluster centroids. K-Means uses distance-based algorithm to determine the similarity between the data points, usually Euclidean distance is selected. Distance formula is shown as formula 3-4.

$$Distance = \sqrt{(x2 - x1)^2 + (y2 - y1)^2} \quad (3\text{-}4)$$

K-Means algorithm works as:

1. Specify number of clusters K
2. Initialize centroids by randomly selecting K data points as the centroids
3. Categorize each item to its closest centroid and compute the means for each cluster then update the cluster centroids
4. Repeat the step 3 until the centroids does not change or maximum number of iterations have been reached

3.2.2 Elbow Method

The Elbow Method is used to compare the number of clusters for best fit, by plotting the sum of squared distances (intertia) between the data points and the associated cluster centroids. The elbow in the plot refers that the point where adding more clusters does not change the slope of the elbow line (Humaira and Rasyidah, 2020). Figure 3.1 shows an example Elbow Point for a typical clustering scenario.

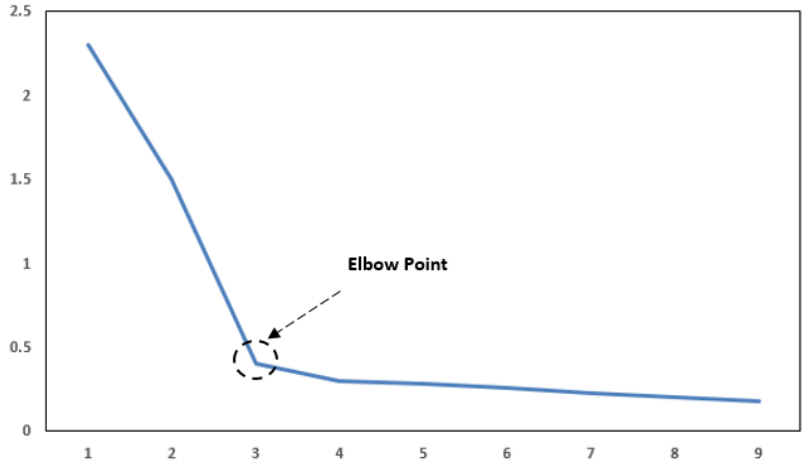


Figure 3.1: Elbow Point Example

3.2.3 Silhouette Score

Silhouette Score is another technique for assessing the quality of centroids by calculating Silhouette Score for clustering algorithms (Rousseeuw, 1987). The score ranges between -1 to 1 whereas the nearest data point +1 is the best match among the alternatives. 0 means that data point lies on or near the boundary between clusters. Values around -1 score shows the values assigned to wrong cluster (Shahapure, 2020).

$$Silhouette\ Score = \frac{(b-a)}{\max(a,b)} \tag{3-5}$$

Silhouette score is calculated using above formula where:

- For each data point, the average distance to other data points in the same cluster calculated as a.
- For each data point, the average distance to all other cluster it does not belong to is calculated as b.

3.2.4 Prophet Forecasting Model

Prophet is a powerful tool developed by Facebook, designed for handling time-series data with strong seasonal components. It is particularly effective for capturing daily, weekly, and yearly patterns and making it easier to visualize and quantify these recurring trends (Zhao et al., 2018). Prophet uses three main model components such as trend, seasonality and holidays.

$$y(t) = g(t) + s(t) + h(t) + \epsilon_t \tag{3-6}$$

Function shows the prophet model where (Taylor and Letham 2017):

- g(t) is the trend function which models non-periodic changes in the value of the time series
- s(t) represents periodic changes (weekly, yearly seasonality)

- h(t) represents the effects of holidays
- ϵ_t the error term represents any idiosyncratic changes that are not accommodated by the model

Prophet model evaluation can be utilized with RMSE (Chai and Draxler, 2014), MAPE (Makridakis and Hibon, 2000) values that are already included in the python package for prophet.

3.3 Machine Learning Algorithms

Gradient Boosting model is a part of decision tree family and widely used for capturing non-linear interactions and highly affective handling weakly correlated features. These properties make this model a good candidate for predicting the fresh food counts.

ANN excel at detecting complex, non-linear relationships by learning intricate patterns within the data. This is particularly useful with the limited linear correlation between features and the target. ANN is another possible candidate for this research.

3.3.1 Gradient Boosting

Gradient Boosting is a machine learning technique used for both classification and regression tasks. It builds models in a sequential manner by combining multiple weak learners, most commonly decision trees. In this technique, the aim is to reduce the residual errors of the previous ensemble after adding a new tree so the model's accuracy is boosted step by step. The underlying principle of gradient boosting is to optimize a loss function by iteratively adding weak learners in a manner like gradient descent, which allows for efficient error correction (Friedman, 2001).

The weak learners used in gradient boosting are usually shallow decision trees, which are well-suited for capturing simple relationships in the data. These shallow trees, also known as "stumps," have limited depth and thus contribute only a small portion to the model's overall prediction power. The idea is to continuously add these simple models to minimize the loss, thus collectively forming a strong model.

Decision trees are good at modeling non-linear relationships, making them especially useful for capturing the complex interactions between features. In the context of predicting fresh food consumption for airlines, such features could include passenger demographics, flight schedules, seasonal variations, and other influencing factors. By aggregating multiple decision trees, gradient boosting forms an ensemble that can handle complex data structures and deliver accurate predictions.

Gradient boosting algorithms, such as XGBoost, LightGBM, and AdaBoost, have been widely adopted in various domains due to their excellent predictive performance and flexibility. They can handle both numerical and categorical features, capture nonlinear relationships, and effectively handle large datasets.

XGBoost(eXtreme Gradient Boosting) builds on the concept of gradient boosting by incorporating several enhancements that improve both the efficiency and accuracy of the model (Chen and Guestrin, 2016). The model maintains the foundational principles of gradient boosting but incorporates additional system optimizations, making it one of the most effective implementations of the algorithm. Figure 3.2 shows the general architecture for XGBoost model.

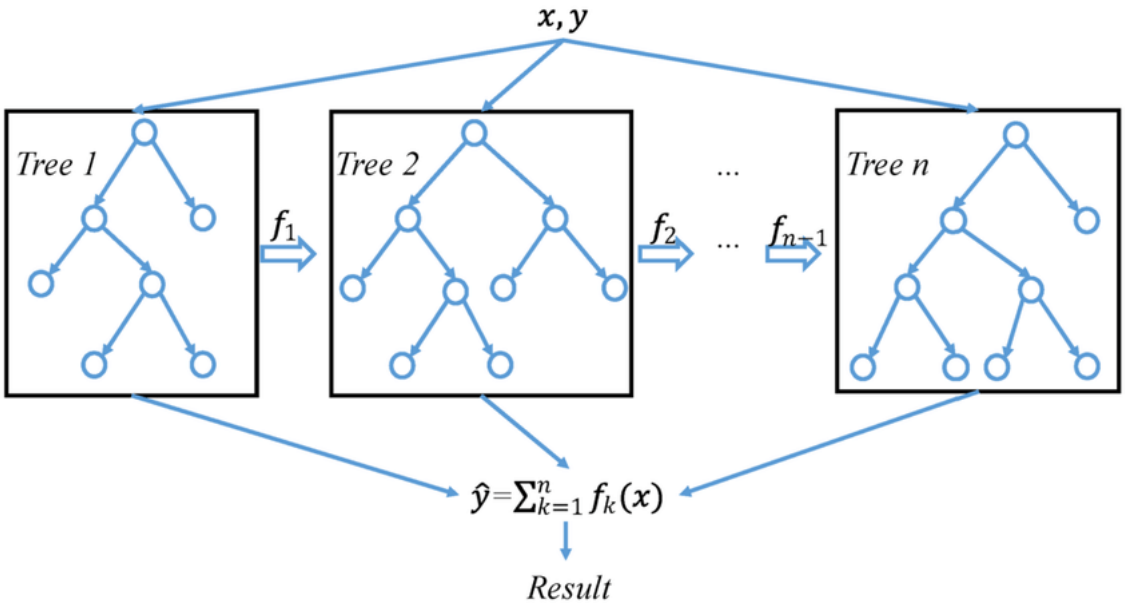


Figure 3.2: XGBoost General Architecture

Major enhancements of the algorithm:

- XGBoost introduces both L1 (Lasso) and L2 (Ridge) regularization, adding a penalty to the loss function that helps prevent overfitting.
- XGBoost introduces shrinkage, which is referred as the learning rate, that scales the contribution of each tree added to the model. This helps better generalization to the new data and prevents it from overfitting the training data quickly.
- XGBoost uses a "depth-first" approach to grow trees and employs a loss-guided pruning strategy. This ensures that the trees are grown only as much as needed which makes the model remain efficient.
- XGBoost is particularly adept at handling missing data. It has an in-built mechanism that automatically finds the best path for missing values during tree construction, making it highly suitable for real-world scenarios where data is often incomplete.
- Traditional gradient boosting is inherently a sequential process, which limits scalability. XGBoost, however, utilizes advanced parallel processing techniques, allowing it to grow tree nodes simultaneously. This drastically reduces training time and makes XGBoost highly efficient even when dealing with large datasets.

Given the intricate and dynamic nature of fresh food consumption for airlines, where numerous factors influence passenger preferences and consumption patterns, XGBoost's tree-based ensemble approach offers a powerful solution for regression-based problem (Zhang et al., 2019). The combination of decision tree learning with gradient boosting allows XGBoost to capture both the linear and non-linear relationships that exist within the data.

XGBoost model offers some important parameters to for tuning:

- Learning Rate (eta); controls how much the model adjusts weights with respect to the gradient step.
- Number of Boosting Rounds (n_estimators); the number of trees built during mode training.

- Max Depth (max_depth); the maximum depth of each tree, controlling the model's complexity.
- Min Child Weight (min_child_weight); the min sum of instance weight needed in a child node.
- Subsample; the fraction of training data used to build each tree.
- Gamma (min_split_loss), the minimum loss reduction required to make split.

3.3.2 Neural Networks

Artificial neural networks are a class of machine learning models inspired by the structure and functioning of biological neural networks in the human brain. McCulloch and Pitts proved that by assembling sufficient neurons and assembling appropriately the synaptic links, any computational problem can be solved (McCulloch and Pitts 1943).

Neural networks are powerful tools for solving complex tasks, including pattern recognition, classification, regression, and sequence generation. Neural networks are highly flexible and can be adapted to various problem domains (Mitchell, 1997). They require careful selection of architecture, hyperparameters, and training techniques. Additionally, they often benefit from large amounts of labeled training data for optimal performance.

ANN is composed of layers of interconnected nodes(may be referred as neurons). Each node in separate layer can process the data from another node on previous layer. Those layers are grouped into three categories. First category is called Input layer, which takes the input data. Then some multiple layers are grouped into hidden layers section. Weight based calculations are handled at this layer through connections between nodes. The hidden layers are core of the neural network. Each neuron in a hidden layer applies a mathematical transformation to the inputs it receives. These hidden layers allow ANNs to learn and model complex non-linear relationships by capturing intricate patterns within the data. A typical ANN can consist of multiple hidden layers, which is the reason why they are called deep neural networks when the number of hidden layers is big. At the end output layer produces final output from the network. Figure 3.3 shows basic ANN structure.

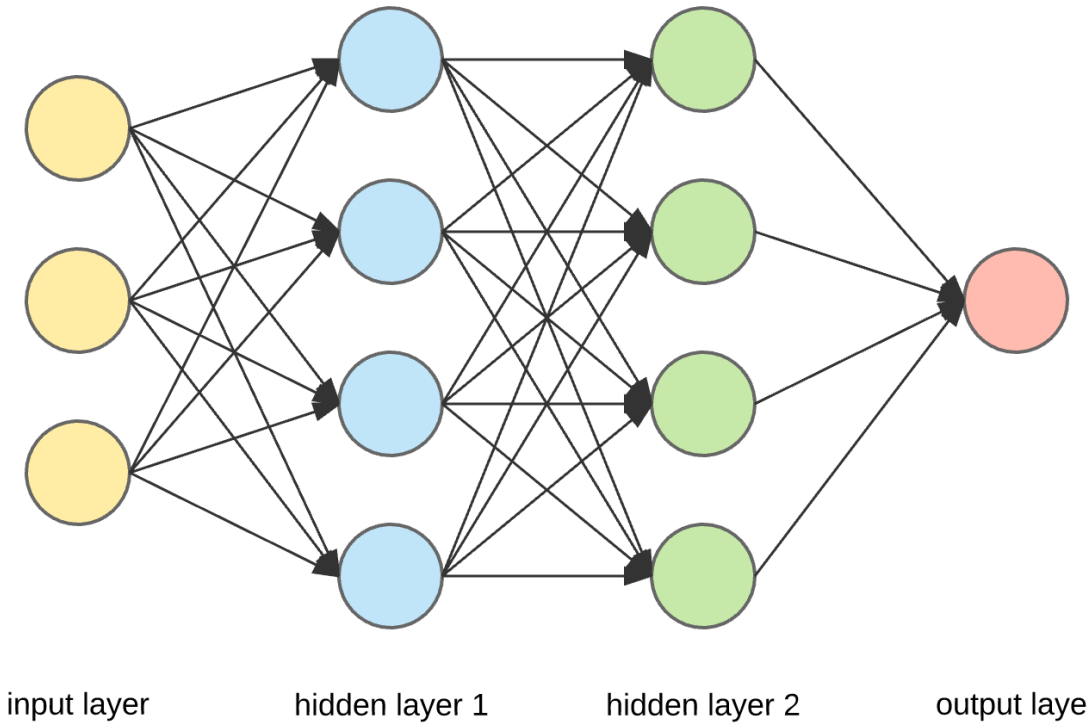


Figure 3.3: ANN Structure

Major components and advantages:

- Activation functions are used to introduce non-linearity into the model, enabling it to learn complex patterns. Most common activation functions are ReLU (Rectified Linear Unit), Sigmoid, and Tanh (Nwankpa et al., 2018).
- ANNs are trained using backpropagation, which adjusts the weights of the connections between neurons to minimize prediction errors (Rumelhart et al., 1986). During training, an optimization algorithm such as Stochastic Gradient Descent (SGD) or Adam is used to minimize the loss function (Kingma and Ba, 2014).
- One of the major strengths of ANNs is their ability to learn from data with complex and non-linear relationships. In the context of fresh food consumption, factors such as changing passenger preferences, seasonal trends can interact in highly complex ways. ANNs can capture these relationships more effectively.
- ANNs, when designed and tuned correctly, can generalize well to new data. This is crucial for airline industry that generates new data rapidly for each flight.

Predicting fresh food consumption for airlines involves complex patterns that can be influenced by a wide range of features: passenger demographics (age, class, preferences), flight duration, season, and even external factors such as cultural preferences during events. ANNs can model these relationships efficiently by processing these diverse input features through multiple hidden layers to capture deep correlations and generate an accurate forecast. This advantage also makes ANN a better candidate for regression-based problems than linear regression.

The flexibility of ANNs makes them suitable for handling non-linear data and dynamic environments like airline operations, where customer preferences can change frequently. Additionally, ANNs can be enhanced with techniques like dropout to prevent overfitting (Srivastava, 2014), which is particularly helpful in ensuring that the model doesn't become too tailored to past flights, thus generalizing better to new scenarios.

3.4 Performance Metrics for Models

Evaluating the effectiveness of predictive models is crucial to ensure their accuracy and reliability. In this study, Root Mean Square Error (RMSE) and Coefficient of Determination (R^2) are employed as performance metrics to evaluate the models. These metrics provide complementary insights: RMSE measures the average magnitude of prediction errors, while R^2 evaluates the proportion of variance in the target variable explained by the model. Together, they offer a comprehensive measurement of model performance and predictive accuracy.

This combination of metrics has been utilized in recent years for comparing machine learning models such as XGBoost and ANN in similar studies. In their research Abdikan, Sekertekin, Narin and Sanli used RMSE and R^2 in assessing the crop height estimation performance of SLR, MLR, ANN, XGBoost and CNN (Abdikan et al., 2023). In another study Chakraborty and Elzarka used this combination of metrics to make comparative analysis between XGBoost and ANN models on energy consumption prediction (Chakraborty and Elzarka 2017).

3.4.1 Root Mean Squared Error

Root Mean Squared Error (RMSE) is the square root of the mean squared differences between predicted and actual values (Chai and Draxler, 2014). It provides an error metric in the same units as the target variable, making it easy to interpret (Stevanović et al., 2024). RMSE is selected because of its ability to penalize large prediction errors more than small ones, which is important when the cost of incorrect forecasting can lead to significant operational waste or shortages. Lower RMSE values indicate that the model is making predictions closer to the actual values. The lower the RMSE, the better the model is at minimizing large deviations from true consumption, thereby providing a reliable forecast for fresh food needs. The accuracy of the models can be evaluated with RMSE for this research.

Root mean square error can be expressed as

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n} \|y(i) - y\hat{}(i)\|^2} \quad (3\text{-}7)$$

- n is the number of data points
- y(i) is the i-th measurement
- y^(i) is its corresponding prediction

3.4.2 Coefficient of Determination (R^2)

R^2 Score measures the proportion of the variance in the dependent variable that is predictable from the independent variables. It ranges from 0 to 1, where a higher R^2 indicates a better fit of the model to the data. R^2 is considered a good evaluation metric for nonlinear regression algorithms (Chicco et al., 2021). R^2 was used to assess how well each model explains the variance in fresh food consumption across different flights. A higher R^2 score suggests that the model effectively captures the relationships within the features of data, which is critical for understanding the patterns that influence customer demand. R^2 score evaluates how well the model covers the seasonality, passenger

demographics, etc. R^2 value close to 1 indicates that the model can explain most of the variability in food consumption.

Residual sum of squares expressed as:

$$SS_{res} = \sum_i (y_i - f_i)^2 \tag{3-8}$$

Total sum of squares expressed as:

$$SS_{tot} = \sum_i (y_i - y\hat{})^2 \tag{3-9}$$

General definition of the coefficient of determination is:

$$R^2 = 1 - \frac{SS_{res}}{SS_{tot}} \tag{3-10}$$

- y is the observed point
- y^ is the mean of observed data
- f is the corresponding predicted value

4. MATERIAL METHOD

4.1 Exploratory Data Analysis

Initial data is collected from operational data from airline company. Dataset covers two years period between 2022 – 2024. Dataset consists of schedule information, transaction information and passenger statistics. Dataset consists of 450k records in 25 columns CSV format. Table 4.1 shows the initial set of attributes and sample data for single origin and destination.

Initial data set consists of similar date time attributes that comes from different data sources. FLT_DATE, FLT_DATE_YMD attributes exactly same data with different formats.

In airline operations, date/time data is managed using two main time zones: UTC (Coordinated Universal Time) and the Local Time Zone of the departure country. This dual time zone usage is essential for accurate scheduling, operational management, and post-flight data analysis. Departure and arrival timestamps are further categorized based on their application, specifically into scheduled time, estimated time, and actual time.

- Scheduled Time represents the planned timing for flight operations and is primarily used for scheduling and tracking flight legs. Notably, scheduled time remains static, meaning it does not change even if the flight status is updated (e.g., due to delays or cancellations).
- Estimated Time utilized during daily operations, estimated time helps monitor and track ongoing flights. It serves operational purposes by providing a dynamic forecast of expected departure and arrival times. However, this metric is not used for data mapping or in the final analysis since it is inherently variable and subject to change during flight operations.

- Actual Time records the real-time departure or arrival of a flight and is only available post-flight completion. For instance, the actual departure time is only available if the aircraft has physically departed.

Table 4.1: Sample Records from Dataset

FLT_LINE	DOM	DOM	INT
ORIG	AYT	AYT	ESB
DEST	ESB	ESB	FRA
FLT_DATE	44987	44988	45115
FLT_DATE_YMD	20230302	20230303	20230708
FLT_NO	1000	1002	2015
FLT_SFX	-	-	-
ITEM_NAME	Group-I	Group-II	Group-I
ITEM_CATEGORY	Fresh Food	Fresh Food	Fresh Food
TRANSACTION_TYPE	SALE	SALE	SALE
UNIT_PRICE	35	45	6
SALE_TOTAL	102	135	11
PROMOTION	3	3	1
DISCOUNT_AMOUNT	3	3	1
QUANTITY	3	3	2
PRE_ORDER_COUNT	4	2	2
STD	3.02.2023 05:50:00	3.03.2023 05:50:00	7.08.2023 13:35:00
STD_LT	3.02.2023 08:50:00	3.03.2023 08:50:00	7.08.2023 16:35:00
DEP_DT	3.02.2023 05:42:00	3.03.2023 05:38:00	7.08.2023 13:58:00
DEP_DT_LT	3.02.2023 08:42:00	3.03.2023 08:38:00	7.08.2023 16:58:00
DURATION	0 01:30:00.000000	0 01:37:00.000000	0 03:40:00.000000
PAX_MALE	106	78	59
PAX_FEMALE	67	54	68
PAX_CHILDREN	12	4	32
PAX_INFANT	2	3	5

In airline operations data usage, up to six types of date-time attributes may be associated with each flight. The key attributes used for analysis in this study are:

- STD (Scheduled Time of Departure) and STD_LT (Scheduled Time of Departure in Local Time): These attributes provide the scheduled departure time in UTC and in the local time zone of the departure country, respectively.
- DEP_DT (Actual Departure Time) and DEP_DT_LT (Actual Departure Time in Local Time): These attributes indicate the actual departure time in both UTC and the local time zone.

DURATION information is another critical information that is calculated according to actual departure and arrival times. Dataset has attribute for this data so that no further calculations are required.

Table 4.2: Statistical Results for Numeric Attributes

	Unit Price	**Sale Total**	**Promotion Amount**	**Discount Amount**	**Quantity**	**Preorder Count**	**Flight Duration**
Mean	196,40	149,46	66,85	63,53	4,32	5,90	198,41
Std	52,85	491,21	81,83	78,68	4,33	5,81	44,45
Min	30	0	0	0	1	0	19
Max	272	8806	850	714	71	190	411

Table 4.3: Statistical Results for Pax Numeric Attributes

	Male Pax	**Female Pax**	**Children Pax**	**Infant Pax**	**Total Pax**
Mean	68,45	73,38	18,07	3,41	163,32
Std	17,93	19,23	13,08	2,65	37,97
Min	0	0	0	0	0
Max	185	145	90	23	208

Initial statistical analysis was conducted on the numerical attributes of the dataset to understand their basic characteristics and presented in Table 4.2 and Table 4.3. Key summary statistics, including mean, standard deviation (std), minimum (min), and maximum (max) values, were calculated for each attribute to provide a foundational understanding of the data's distribution.

It is observed that most numerical attributes in the dataset have a minimum value of zero. The presence of these zero values is significant and covers careful interpretation, as their source and distribution vary across different attributes. Therefore, each attribute should be assessed individually to understand its implications and potential impact on the overall analysis.

The flight line information is a common attribute across all records in the dataset and is represented as categorical data. To draw meaningful insights from this attribute, it is essential to analyze the distribution of sales across different flight lines. This analysis allows for more accurate interpretations of the dataset and informs subsequent modeling and feature engineering decisions.

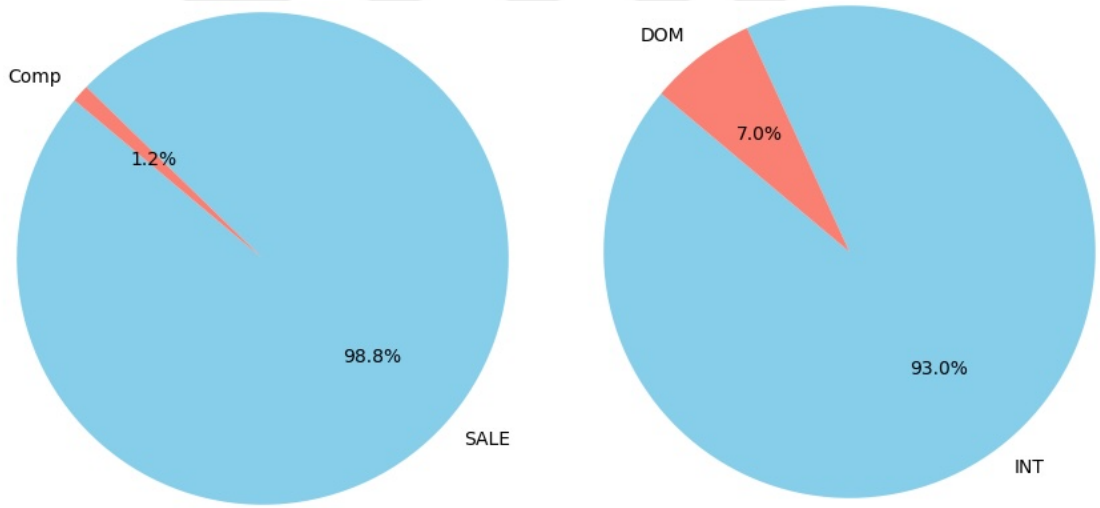


Figure 4.1: Distribution of transactions for Sale and Route Types

The dataset encompasses two primary types of sales: standard paid sales and complimentary sales. This distinction is important for understanding the presence of zero values in numeric attributes and the implications for data analysis.

- Standard Paid Sales represent typical transactions where passengers purchase items, contributing directly to revenue. The data from these sales include non-zero values for attributes such as price and quantity.
- In Complimentary Sales transactions, passengers are not charged for their purchases, resulting in zero values for numeric attributes related to pricing and payment. While complimentary transactions constitute only 1.2% of the total sales, they are still indicative of demand and play a crucial role in understanding passenger service patterns.

The distribution of complimentary versus standard transactions is shown in Figure 4.1 left plot. Although the proportion of complimentary sales is small, their presence highlights an essential aspect of operational management. The primary drivers for providing complimentary sales are operational disruptions, like:

- Preorder Uplift Failures where preordered items are not uplifted as planned.
- Extended Waiting Times in the plane or at the gate that ground operation cannot handle.
- Logistic irregularities such as damaged item package or rotten food.

In these cases, offering complimentary items helps maintain passenger satisfaction and reflects the airline's commitment to service quality. Understanding the context and frequency of these transactions is critical for interpreting sales data accurately and identifying patterns related to demand that may not be apparent from paid transactions alone. These disruptions are not uniform across all operations; their frequency distribution may exhibit local effects. Factors such as airport infrastructure, staffing levels, or regional regulations can influence the likelihood and frequency of disruptions.

As illustrated in Figure 4.1 right plot, approximately 93% of total sales are attributed to international flights. This significant proportion underscores the importance of differentiating between domestic and international flight operations when interpreting data trends.

A key distinction between domestic and international flights lies in the pricing structure and the variety of products offered:

- Domestic Flights have typically lower prices and a limited selection, with only two categories of fresh food available at the same time during the season.
- Internal Flights have higher prices and offer a broader range of products, encompassing four categories of fresh food.

Understanding these differences is crucial for analyzing customer purchasing behavior and demand forecasting. The variance in pricing and product availability between domestic and international flights impacts not only sales distribution but also consumer preferences and operational logistics.

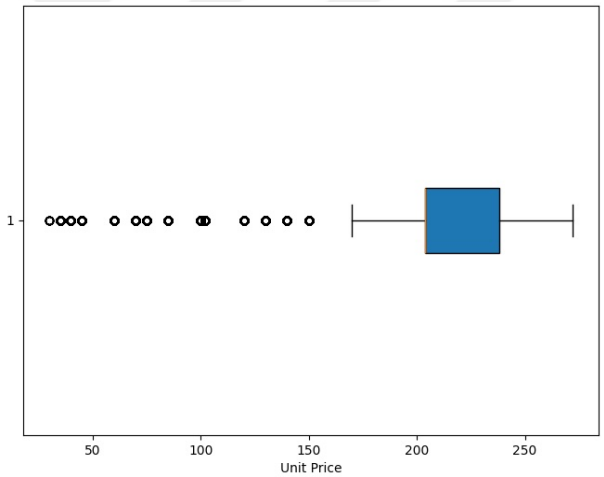


Figure 4.2: Box Plot for Unit Price of Products

Figure 4.2 presents a box plot illustrating the distribution of unit prices within the dataset. The plot highlights that most significant unit price values are above 200, with the distribution skewed towards higher prices. The visualization suggests that unit prices below 150 falls outside the main data distribution and may be considered outliers.

While removing these lower-priced values could potentially enhance the model's focus on more consistent data, it is essential to consider the origin of these prices. Analysis reveals that unit prices below 150 primarily come from domestic flights, which account

for 7% of all transactions, as shown in Figure 4.1. If prices below 150 are excluded as outliers, all domestic transactions would be eliminated from the dataset.

Removing these data points could skew the dataset and lead to the loss of valuable information regarding domestic flight sales patterns. This action would not only reduce the comprehensiveness of the dataset but could also negatively impact the model's ability to predict demand for the entire range of flights. Consequently, the decision to classify and remove lower unit prices as outliers must be carefully weighed against the importance of maintaining a representative sample that reflects both domestic and international transactions.

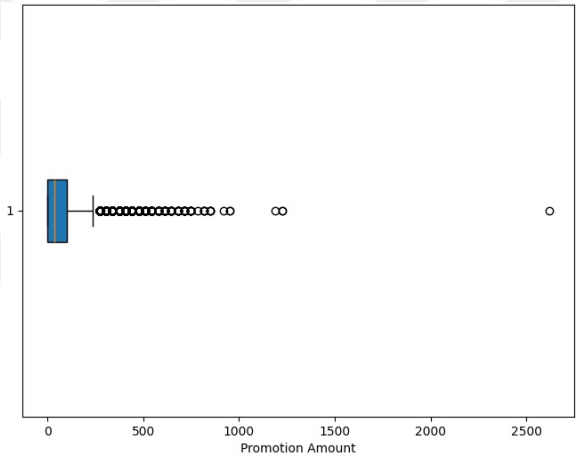


Figure 4.3: Box Plot for Promotion Amount

The dataset includes attributes for promotion amount and discount amount, which exhibit similar behavior. This is because the discount amount is calculated as the total sum of promotion amounts applied to a given transaction across all items. For this analysis, the promotion amount is examined in detail, as it is the primary source of this data.

Figure 4.3 presents a box plot of the promotion amount distribution. The plot reveals a wide range of values, but a significant portion falls into the outlier category. This indicates that while high promotion amounts are present, they occur less frequently and may distort the overall interpretation of the data.

A noteworthy observation is that complimentary product sale transactions often include a recorded promotion amount that matches the total sale amount. This reflects the operational practice of compensating passengers during disruptions or as a gesture of goodwill, resulting in a recorded transaction that includes a full promotional adjustment. While this practice may be financially justifiable and operationally routine for the airline, it requires careful consideration for research purposes. Therefore, when interpreting and modeling the dataset, the role and frequency of promotional amounts in complimentary sales should be accounted for to avoid bias and ensure more accurate demand forecasting and predictive modeling outcomes.

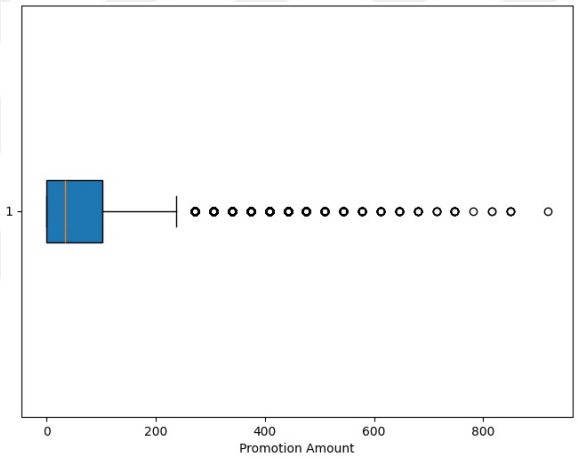


Figure 4.4: Box Plot for Promotion Amount without Complementary Transactions

Figure 4.4 illustrates the results after updating the PROMOTION_AMOUNT attribute for complimentary transactions to zero. This adjustment was made to better reflect the nature of these transactions, where no actual monetary exchange occurs. By setting the promotion amount to zero for complimentary sales, the dataset provides a more accurate representation of true promotional activities linked to paid transactions.

The new box plot shows a narrower range for PROMOTION_AMOUNT, indicating a more consolidated distribution of data. However, despite this adjustment, the plot suggests that there may still be outliers within the range. These outliers warrant further

investigation to understand their source and potential impact on the analysis. While setting promotion amounts to zero for complimentary transactions helps align the data more closely with financial reality, it is important to note that the presence of remaining outliers could still affect model training and interpretation.

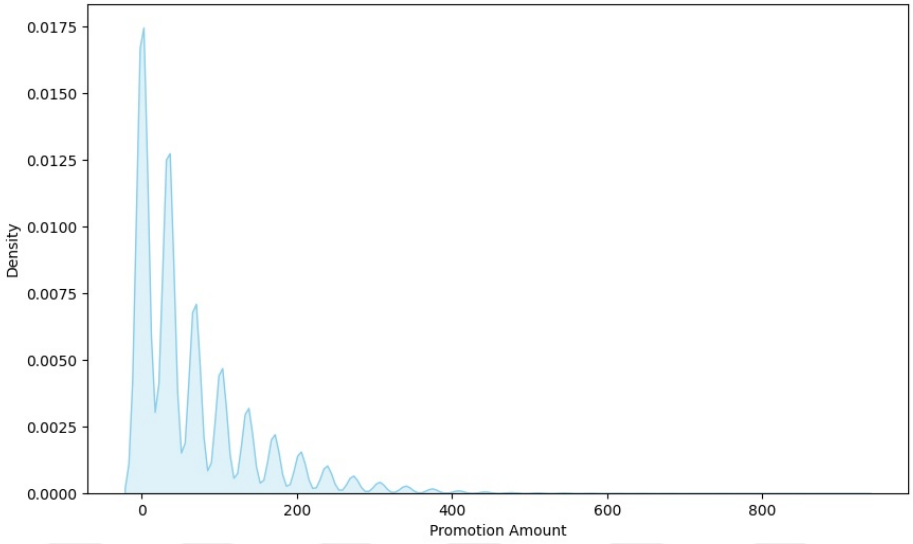


Figure 4.5: Density Plot for Promotion Amount without Complementary Transactions

Figure 4.5 shows the density plot for the adjusted PROMOTION_AMOUNT attribute. The plot reveals that density drops significantly for values above 400, suggesting that higher promotion amounts are rare within the dataset. Upon manual inspection of the data, it was observed that while the quantities associated with these transactions are valid, the corresponding promotion amounts are challenging to justify or interpret. This anomaly may point to potential data inconsistencies or corruptions.

In the daily airline operations, numerous edge cases can occur, which may contribute to these unexplained promotion amounts. Such cases could involve complex scenarios during flight operations, such as adjustments made for operational disruptions, manual overrides, or data entry errors. Due to the variety and complexity of these edge cases, it is difficult to pinpoint a singular explanation for these high promotion values or classify them definitively.

Given that the density of outlier values is low and the corresponding quantities are valid, removing these outliers may not be the most effective approach. Excluding them could result in an incomplete representation of the dataset, potentially skewing the analysis and impacting the predictive models' ability to generalize. Therefore, while acknowledging these outliers, it is recommended to retain them to maintain the dataset's integrity and reflect the true variability present in flight operations.

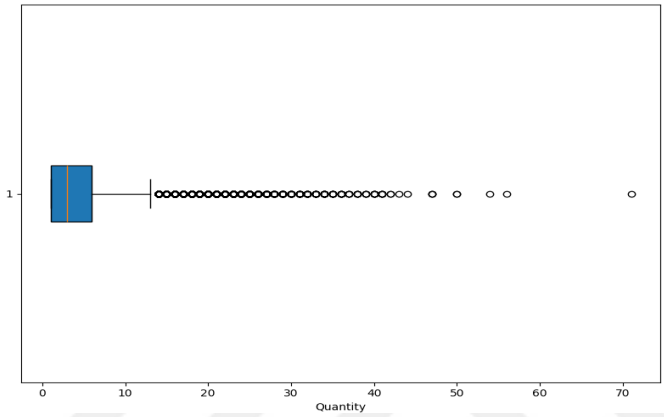


Figure 4.6: Box Plot for Quantity

The Quantity attribute represents the total number of different products purchased during a flight. Figure 4.6 provides a visualization of the distribution of item quantities, revealing that most of the transactions fall within the 0 to 15 range. The mean value for the Quantity attribute is 4.32, with a standard deviation of 4.33. These statistical values suggest that, on average, approximately four different items are purchased per flight per meal group, with some variation that extends up to approximately four items above or below this mean. The values align with the corresponding box plot, confirming that the dataset's distribution of quantities is consistent and does not exhibit significant deviations or anomalies.

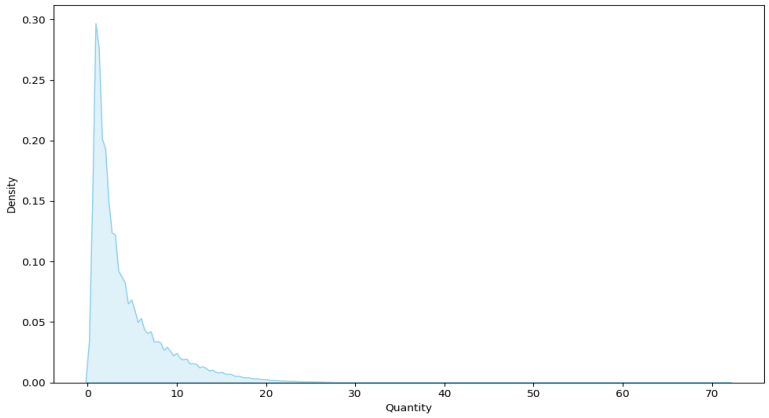


Figure 4.7: Density Plot for Quantity

Figure 4.7 presents the density graph for the Quantity attribute, which illustrates the frequency distribution of the number of different products purchased during flights. The graph indicates that outlier values, particularly those above 40, have a very low density, suggesting potential data irregularities. These low-density values could be the result of data corruption or, more plausibly, operational exceptions within the airline context.

Upon reviewing the dataset, it is found that nearly all quantities exceeding 40 occur only once. This pattern confirms that these are not data corruptions but rather signs of operational anomalies. Such anomalies might include unusual events or specific cases that lead to atypical purchase quantities, such as emergency provisions or special catering requirements during flight disruptions. The current approach focuses on retaining these operational exceptions rather than removing them. Removing these outlier values could simplify the dataset and potentially improve model training but it would also lead to the loss of significant information that reflects the operational realities of airline service.

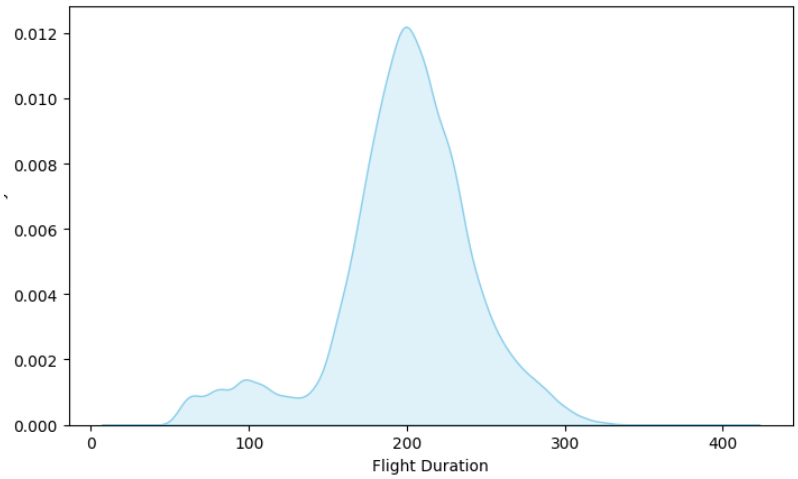


Figure 4.8: Density Plot for Flight Durations

Figure 4.8 presents the density plot of flight durations across the dataset, calculated as the interval between actual departure and arrival times. While there may be outliers within the dataset, these outliers originate from operational factors, such as unexpected delays, extended taxiing times, or other real-time flight disruptions. Given that these anomalies are rooted in operational realities, their presence provides valuable context for understanding potential variability in flight operations.

The Table 4.3 summarizes the statistical analysis of passenger counts, which is influenced by various factors, including the layout and configuration of the aircraft. The total number of available seats on a plane depends significantly on the airline's chosen configuration, which may vary based on factors such as leg distance, the presence of a business class section, and overall seat layout. Low-cost airlines, for example, typically prioritize maximizing the number of economy seats, often opting out of installing business class sections. This results in configurations with approximately 200 economy seats per plane. The dataset reveals a maximum observed passenger count of 208. This slightly exceeds typical commercial seating capacity and could be attributed to airline staff occupying non-commercial seats during a flight.

4.2 Correlation Analysis

A heatmap visualizes data in grids where individual values are represented by colors. Correlation Heatmaps used to display relationship between features of the data. Values

inside the heatmap varies between -1 and 1. Color density shows the strength of the relation between features (Vacanti, 2019).

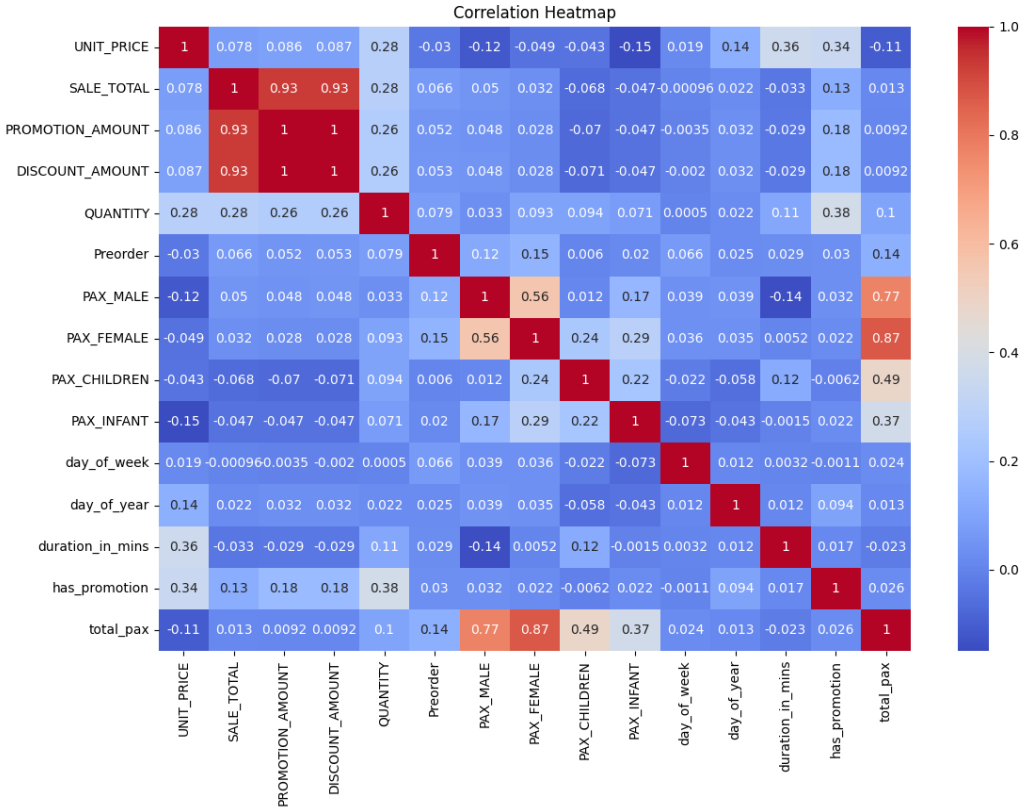


Figure 4.9: Heatmap for numeric attributes

In this research, the primary objective is to predict the quantity for each fresh food category. A comprehensive correlation analysis was conducted to identify relationships between key attributes in the dataset. The heatmap visualization indicates a strong correlation between the total transaction amount (SALE_TOTAL) and the promotion-related attributes.

The dataset has three main attributes related to promotions; PROMOTION_AMOUNT represents the discount applied to the unit price, DISCOUNT_AMOUNT denotes the total discount applied to the entire transaction and has_promotion a derived attribute during the research to evaluate whether the presence of the promotion or amount has more effect.

Initial findings show that the DISCOUNT_AMOUNT has a notably high correlation (approximately 0.93) with the SALE_TOTAL, indicating its significant impact on transaction values. Conversely, the has_promotion attribute exhibits a much lower correlation (~0.13), suggesting that while promotions are present, their mere existence does not substantially influence the total transaction amount. This finding underscores that the actual amount of the discount has a far more considerable effect on customer spending than simply offering a promotion.

Another conclusion from the correlation analysis is that discount amounts tend to be higher for more expensive products. This outcome aligns with industry practices where discounts are often applied as a percentage of the product price, leading to larger absolute discount amounts for higher-priced items. Additionally, it appears that customers are more inclined to make purchases when the total discount amount increases, potentially due to the perceived value of savings on more substantial transactions.

Overall, the results highlight that while the presence of a promotion may influence customer decisions to some extent, it is the magnitude of the discount amount that has a more significant and quantifiable impact on the total transaction value.

The analysis of correlations between promotion attributes and the quantity of products purchased reveals underlying patterns. Both the PROMOTION_AMOUNT and DISCOUNT_AMOUNT exhibit a relatively weak correlation of 0.26 with the quantity of products. This suggests that while discounts may influence the overall sales value, they have limited direct impact on the number of items purchased per transaction.

On the other hand, the has_promotion attribute, which indicates the presence of a promotion irrespective of its amount, shows a slightly higher correlation of 0.38 with product quantity. Although this is still considered a weak correlation, it suggests that the mere existence of a promotion may have a more notable effect on the number of products purchased compared to the actual discount amounts. These findings highlight that the psychological aspect of promotions, such as customer perception of getting a deal, may drive higher purchase volumes, even when the discount amount is not substantial.

These observations are critical for understanding how promotional strategies affect consumer purchasing patterns. While the correlation data suggest that promotion amounts have limited direct impact on product quantities, the existence of a promotion may be a more influential factor in driving customer purchases.

The total transaction amount (SALE_TOTAL) shows a weak correlation of 0.28 with the quantity of products purchased. This value suggests that while there is a relationship between the total amount spent and the number of products purchased, it is not particularly strong. This correlation is expected, as the total transaction amount naturally tends to increase when more products are purchased.

The unit price of products exhibits a weak correlation of 0.28 with the quantity of items purchased. This suggests that, while there is some relationship between the price of individual products and the number of items purchased, it is not particularly strong. One potential explanation for this correlation lies in passenger behavior. Certain passengers may tend to purchase higher-priced items, driven by preferences for premium products or a perception of quality.

Passenger counts are an essential factor influencing total sales in the context of airline catering. However, an analysis of the heatmap indicates that the correlation between total passenger count and the quantity of products purchased is relatively weak, with a correlation coefficient of 0.1. The correlation between total passenger count and the total amount of sales is even lower, at -0.11, suggesting that, within the dataset, the number of passengers on a flight has limited direct impact on the transaction values and product quantities.

The dataset also includes correlations for passenger subcategories—male, female, children, and infants. Female and children passenger counts each have a correlation of 0.092 with product quantity, indicating a slightly stronger (though still weak) association compared to the overall passenger count. Infant passenger count shows a similarly weak correlation of 0.071. Male passenger count exhibits the lowest correlation among the categories, at 0.033. These weak correlations suggest that, while the number of

passengers on a flight is naturally expected to influence total sales, the relationship is not pronounced when analyzing data at the transaction or item level. A more comprehensive analysis that aggregates sales by total number of items sold per route may yield stronger correlations and provide deeper insights into how passenger demographics influence overall sales.

Flight duration is an essential characteristic of flights and could be expected to influence passenger behavior and purchasing patterns. However, an analysis of the dataset reveals that flight duration has a weak correlation of 0.11 with the quantity of products purchased. This suggests that, contrary to potential expectations, the length of a flight on its own does not significantly impact the number of items purchased by passengers.

The preorder amount represents the total value of preorders placed before the flight. The nature of the data suggests that the existence of preorders may negatively affect in-flight fresh food sales, as passengers who have preordered items are less likely to make additional purchases during the flight. In the correlation analysis, the preorder amount shows a weak correlation of 0.079 with the quantity of products purchased. This indicates that while there is some relationship between the presence of preorders and the number of fresh food items bought on board, it is not particularly strong. The weak negative impact suggests that preordering may slightly reduce in-flight sales volumes, as expected, but it is not a dominant factor.

The correlation analysis indicates strong relationships between total passenger count and the individual passenger subcategories, specifically male and female passengers. The correlation coefficient for male passengers with the total passenger count is 0.77, while female passengers exhibit an even stronger correlation of 0.87. These high correlation values are expected, as the total passenger count is inherently calculated as the sum of all subcategories, including male, female, children, and infants.

An examination of the correlation analysis involving attributes related to the transaction date reveals that these factors generally show weak correlations with other numerical values in the dataset. Specifically, attributes such as day of the year and day of the week exhibit correlation coefficients below 0.1. This indicates that, based on numerical

correlation analysis alone, these attributes do not appear to have a significant direct relationship with product quantity or transaction value. The weak correlations are somewhat unexpected, as transaction date attributes are often anticipated to play an important role in influencing sales patterns due to factors such as seasonality, weekend effects, and holiday travel surges.

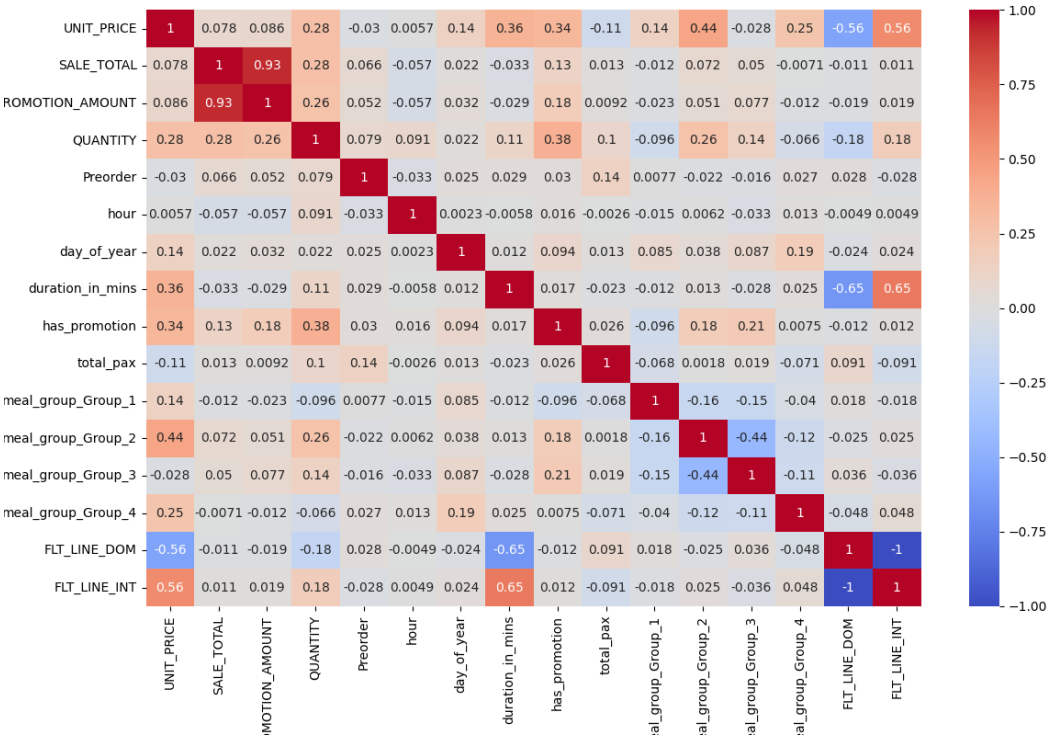


Figure 4.10: Heatmap with encoded categorical attributes

Figure 4.10 illustrates the correlation map that includes both numerical and categorical features, providing an overview of relationships within the dataset. This analysis extends beyond purely numeric attributes to incorporate relevant categorical variables, offering a more comprehensive understanding of the dataset's structure and potential influences on sales and product demand.

One of the key categorical features in the dataset is the meal group classification. Throughout the season, four distinct meal groups are offered to customers, although the specific items within each group may vary due to supplier constraints or requirements.

This variability adds complexity to the analysis, as the product mix and availability can influence purchasing patterns and customer preferences.

Another significant categorical attribute is the flight line information, which indicates whether a transaction is related to a domestic or international flight. In the dataset, 93% of transactions are associated with international flights, while 7% to domestic flights. This attribute has been added to the heatmap analysis to explore its influence on other variables, such as total sales, product quantities, and promotion effectiveness.

Another important categorical feature in the dataset is the departure and arrival airport information. This airline operates flights to more than 80 destinations, representing a wide geographical range and diverse operational conditions. Due to the extensive number of unique values for these attributes, it is not feasible to include all destinations in the heatmap or encode them in a way that allows comprehensive visualization within a single plot.

When comparing the current set of attributes to those in the previous heatmap, it is evident that this group does not contribute significant additional strong correlations, as illustrated in Figure 4.9. The analysis shows that meal groups, a key categorical feature representing the different types of meals offered, exhibit only weak correlations with unit prices and minimal correlation between the groups themselves. This finding suggests that while meal groups are essential for understanding product categorization and availability, their influence on other quantitative metrics such as unit price or sales quantity is limited.

The analysis indicates weak but significant correlations between the flight line attribute (domestic versus international) and other key attributes, such as unit prices and flight duration. While these correlations are not strong, they are meaningful and align with the operational characteristics of the airline. One notable observation is that certain meal groups are not offered on domestic flights throughout the year. This limited availability could contribute to differences in unit prices and purchasing behavior between domestic and international flights. The variation in meal offerings suggests that domestic flights, which typically have a more streamlined product selection, may drive lower unit prices

compared to international routes that offer a broader range of meal options and higher-priced items.

In the heat map departure time attribute is added as hour. Strong correlation was expected between the departure time and the meal group selections. However, heatmap suggest strong relations between the departure time and almost all other attributes. This should be inspected carefully.

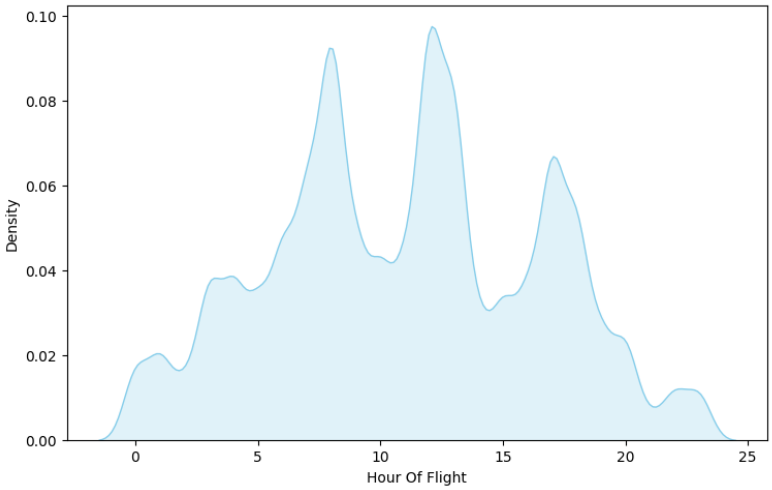


Figure 4.11: Density plot for departure times

Figure shows the density of the departure hours for 24 hours day. Density graph does not suggest any outliers and flight schedule of the airline is consistent with the distribution. Intervals of the day may lead potential differences on passenger behavior.

4.3 Clustering Analysis

Density graph shows that departure times can be grouped into subgroups which may suggest more accurate results. To find out the ideal number of subgroups elbow method may be used.

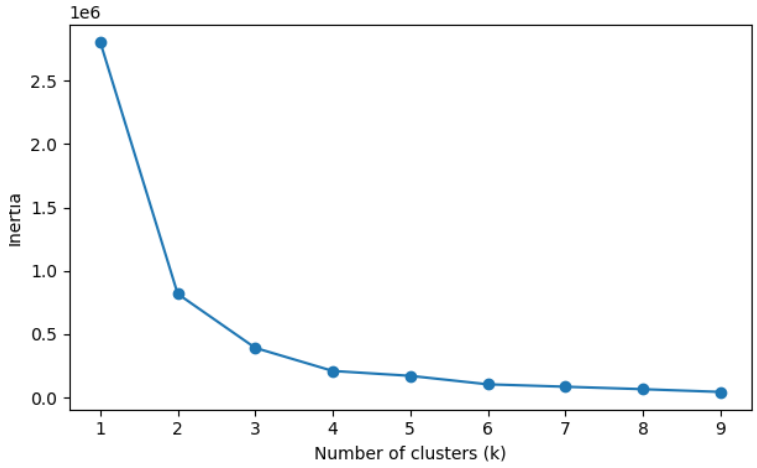


Figure 4.12: Elbow Method Plot

Figure 4.12 shows the elbow method results for departure hours. Elbow plot shows 3 and 4 clusters as possible candidates.

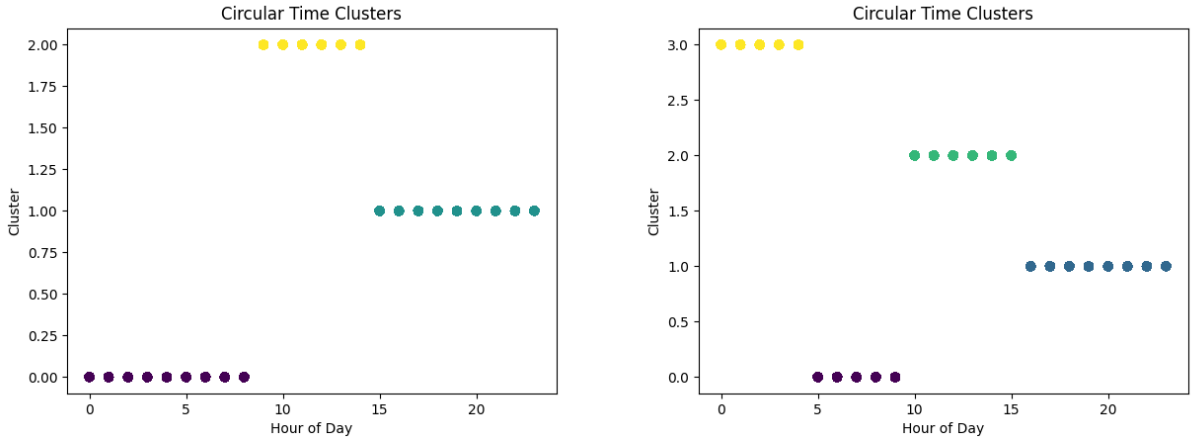


Figure 4.13: K-Means Clustering results for departure hours for 3&4

Figure 3.13 shows K-Means clustering results for departure time for two clusters. Clustering completed with 3 and 4 clusters.

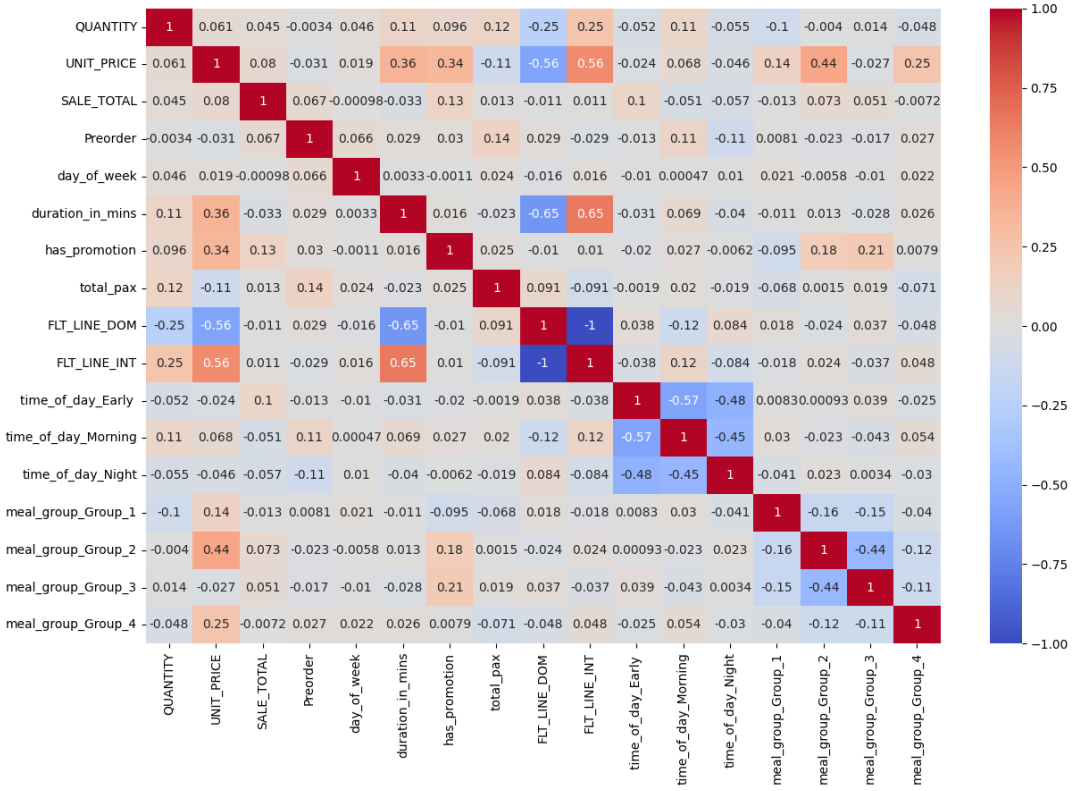


Figure 4.14: Heatmap for all attributes with three-hour clusters

Figure 4.14 presents heatmap for three clusters for hour intervals, those are:

- Early: 00:00 – 08:00 (UTC)
- Morning: 08:00 – 14:00 (UTC)
- Night: 14:00 – 23:59 (UTC)

Time intervals are defined according to airlines timetables and clustering results are used as new attributes in the dataset. However, these attributes does not show any kind of strong correlation with existing attributes and does not show potential use cases.

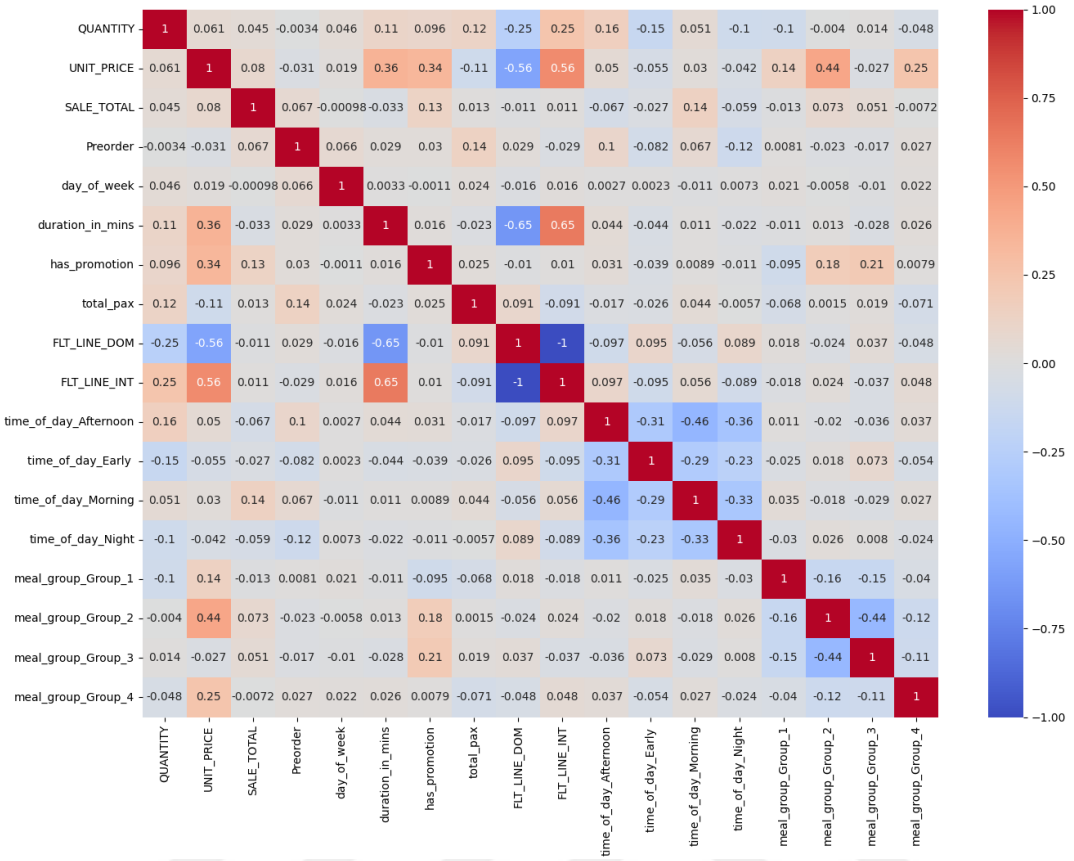


Figure 4.15: Heatmap for all attributes with four-hour clusters

Figure 4.15 presents heatmap incorporating attributes based four-time interval groupings.

- Early: 00:00 - 05:00 (UTC)
- Morning: 05:00 – 10:00 (UTC)
- Afternoon: 10:00 – 15:00 (UTC)
- Night: 15:00 – 23:59 (UTC)

While defining time intervals that align with the airline company's timetable and adding these as attributes to the dataset was explored, this approach did not reveal any strong correlations or significantly improve the model.

For both Figure 4.14 and Figure 4.15 only weak correlations were observed between certain time intervals and attributes such as total sale amount and promotion amount. These results may indicate potential inconsistencies in the distribution of consumption data across routes and flights.

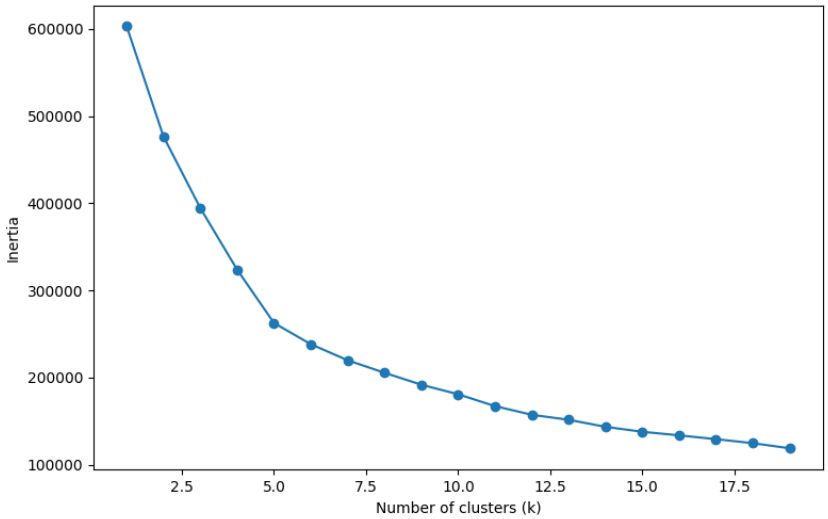


Figure 4.16: Elbow Analyze for Feature Set

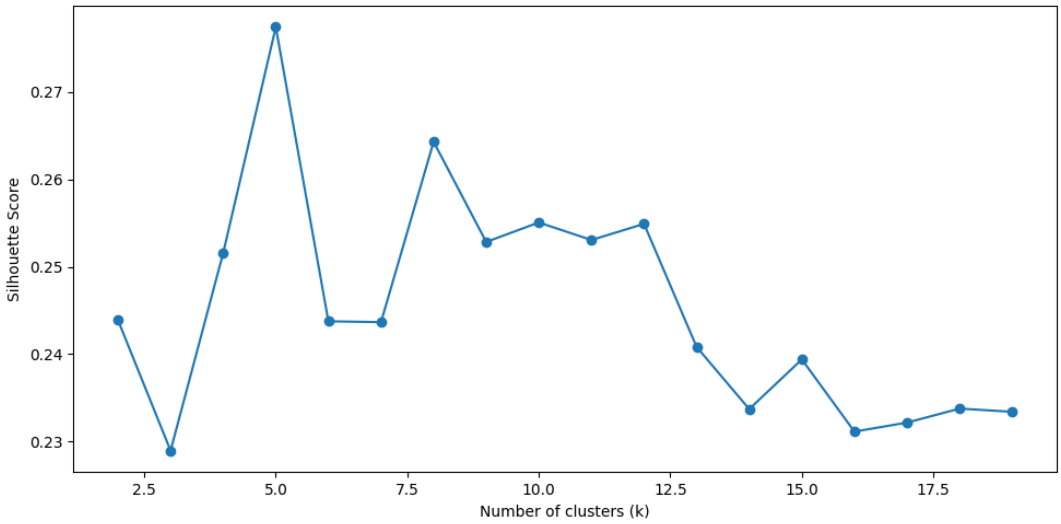


Figure 4.17: Silhouette Score for Feature Set

Figure 4.16 illustrates the Elbow analysis performed on the feature set, which includes departure hour, flight duration, quantity, and meal groups. To assess the quality of the clusters, Silhouette scores were also calculated and illustrated on Figure 4.17. The highest Silhouette score achieved is 0.2774, which, despite being the best score obtained, indicates that the clusters are not well-defined or distinctly separated, as a score close to 1 would represent more clearly separated clusters. Meal groups and flight duration has overlapping nature, and this may lead to the poor cluster separability. Quantity and departure hour is expected to be significant contributors; however, their contributions are

enough for clear separation of clusters. Seasonality and route specific effects may also introduce noise that affects the final quality of the clustering.

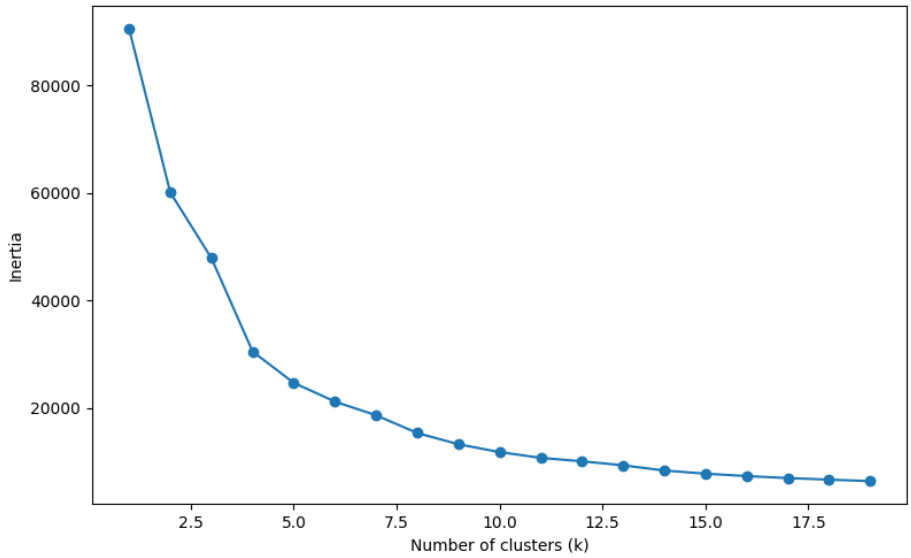


Figure 4.18: Elbow Analyze for Duration, Quantity, Promotion

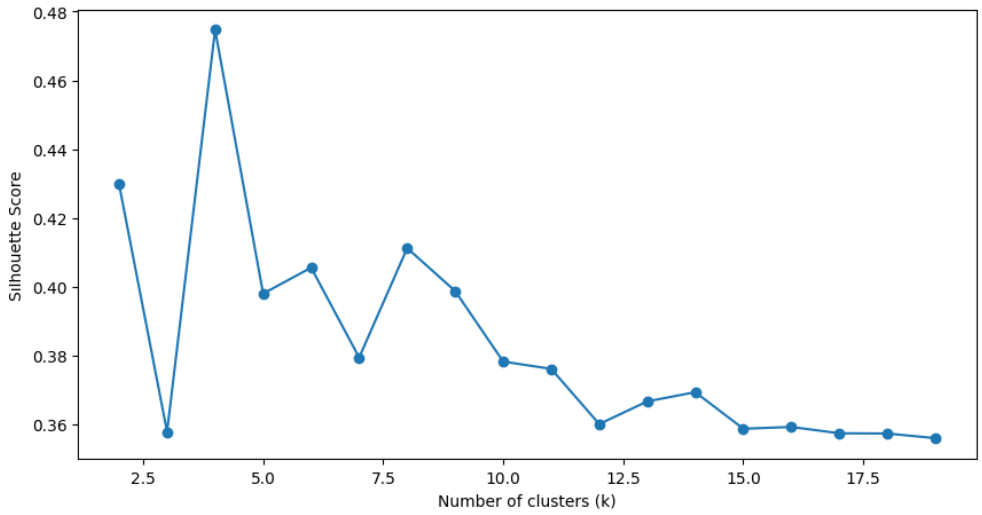


Figure 4.19: Silhouette Score for Duration, Quantity, Promotion

Figure 4.18 illustrates the Elbow Analyzes for features Duration, Quantity and Promotion presence. To assess the quality of the clusters, Silhouette scores were also calculated and illustrated on Figure 4.19. The highest Silhouette Score achieved is 0.4746 for k=4

clusters. This score suggests a moderate level of clustering quality, indicating that the clusters are somewhat distinct but not strongly separated. This feature set defines clearly better clusters than the previous one with four features. Although there is improvement on silhouette scores, this clustering can be considered as moderate and has some level overlaps. These features produce better clusters then previous set. Especially presence of the promotion and the quantity of the purchased products offers natural correlation.

In the research three different approaches may be used. First approach is directly using the departure time hour. The other one is grouping flight times into similar intervals and using those intervals as attributes. The last approach is to merge all data for each destination per daily and make daily predictions rather than per flight bases.

4.4 Seasonality Analysis

An integral part of the dataset analysis involved understanding the seasonality of fresh food consumption, as it plays a crucial role in demand forecasting for airlines. Seasonality analysis helps uncover recurring patterns over specified periods, which can significantly influence model performance by informing feature engineering and model training processes.

To analyze seasonality, the Prophet library was employed. Figure 4.20 shows the seasonality components plots for a selected route and selected meal group between March 2023 – 2024. This prophet model also covers the national holidays for both countries on the route. Dataset shows signs of seasonality on weekly and monthly basis. Passengers are less likely to travel in weekdays. Weekly trend reflects the travel behavior. Weekends and some holidays are high demand intervals. However, holidays have strange affects. The beginning and the end dates of the holidays are most demanding times. On the other hand, the interval between the beginning and end time of long holidays are the less demanding times and have negative effects.

At this point it is important If time series analysis model is sufficient for predictions or research should move advanced machine learning techniques?

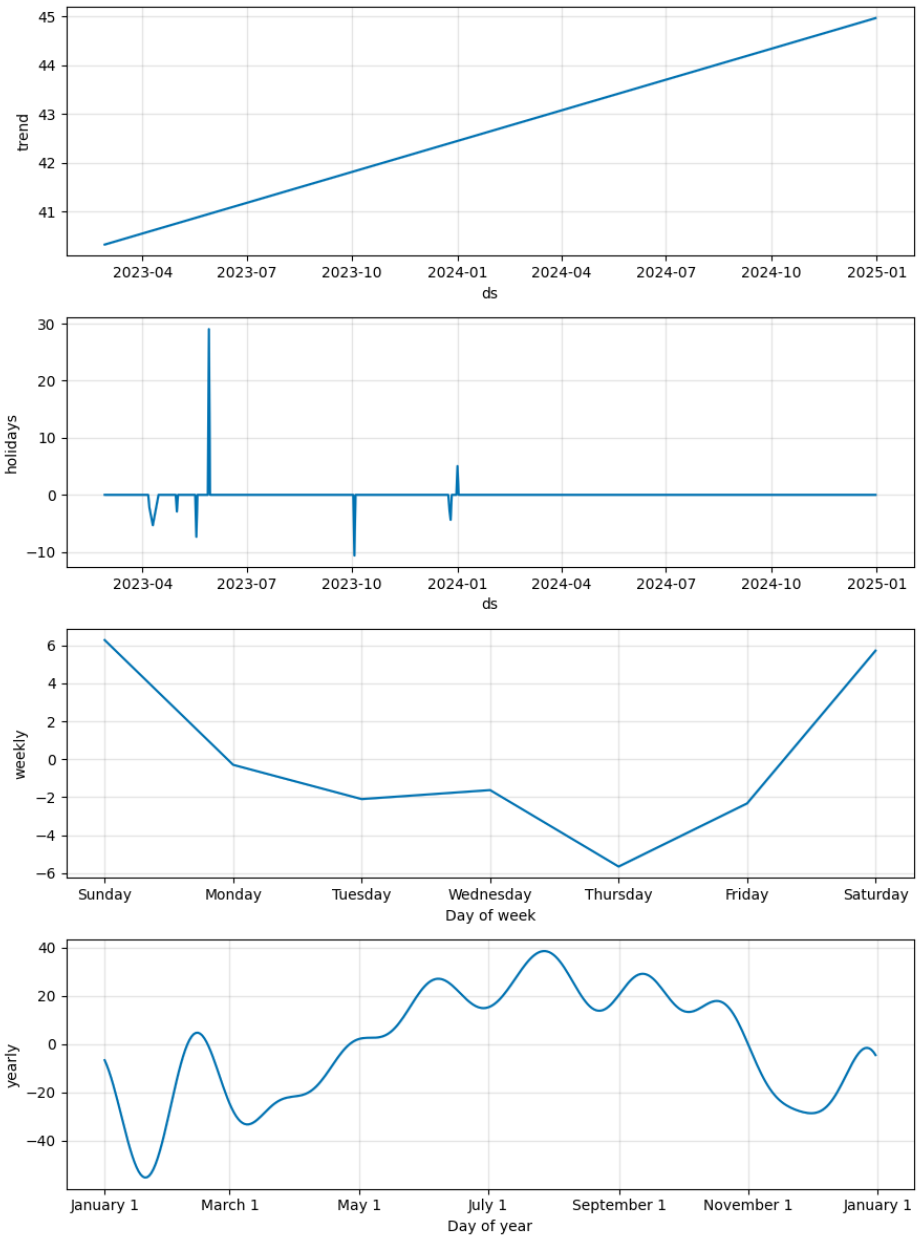


Figure 4.20: Seasonality Components

Figure 4.21 shows the residuals for prophet model introduced above for selected route and meal group. Residual shows large fluctuations and not randomly distributed. The non-randomness in the plot shows signs of more complex relationships in the data. Time series models like Prophet are limited in modelling non-linear relationships.

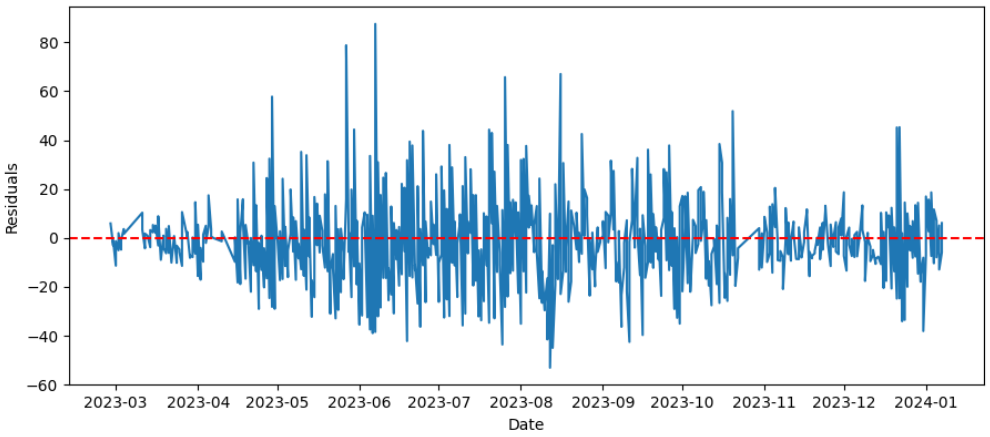


Figure 4.21: Residual Plot

Overall performance for the prophet model may be measured aggregating individual performance metrics like average RMSE (Chai and Draxler, 2014) and average MAPE (Makridakis and Hibon, 2000) for all routes and meal groups.

Table 4.4: Prophet Model Parameters

Parameter Name	**Default**	**Value**
changepoint_prior_scale	0.05	0.5
seasonality_prior_scale	10.0	8.0
holidays_prior_scale	10.0	10.0
seasonality_mode	additive	additive
changepoint_range	0.8	0.8
yearly_seasonality	True	True
weekly_seasonality	False	True
daily_seasonality	False	False

Prophet model parameters are presented Table 4.4, model predictions are processed with these parameters. The results are Overall RMSE is 103.3088 and Overall MAPE 8.1722. The results should be compared with the results for advanced models.

All features are interpreted in the data analysis section. Heatmap values are distributed between 0 to -+1. Positive correlations between two features show both variable changes

in same direction and the values close to the 1 suggest strong correlations. Negative correlations between two features show variable changes happen in opposite directions and the values close to -1 show strong negative correlation. Color density shows how strong is the existing correlation. Red color shows positive correlation, and blue shows the negative correlation. The value and total distribution of the features shows the accurate outcomes.

Figure 4.10, Figure 4.14 and Figure 4.15 shows heatmap for different features derived from same attributes. The initial results offer limited number of considerable correlations. There are limited number of correlation values above 0.5 which may be considered as moderate strong correlation. Most of the features are distributed between 0 - 0.1.

Correlation results show that the model should capture weak direct correlations and non-linear patterns. One possible candidate model may be linear regression. However, linear regression model relies heavily on linear relationships between feature and target variable. With low correlations linear regression would likely to underfit the data.

5. RESULTS

This section presents the outcomes of applying the predictive models—XGBoost and Artificial Neural Networks (ANNs)—to forecast fresh food consumption for airlines. The objective is to evaluate each model's performance in terms of prediction accuracy and its ability to capture underlying data patterns, using Root Mean Squared Error (RMSE) and the R^2 Score as evaluation metrics.

The results are structured to provide a comprehensive comparison between the two models, highlighting both their strengths and weaknesses. RMSE is used to quantify the average magnitude of errors, with particular emphasis on the penalty for larger deviations, which is critical for minimizing operational inefficiencies in airline catering. R^2, on the other hand, is used to assess how well the models capture the variance in fresh food demand, indicating the model's capability in understanding broader consumption patterns influenced by features such as passenger demographics, flight duration, and seasonality.

Both models are tested in the Standard Macbook Pro with 16 GB of Ram with the Apple Silicon M2 processor. Python version 3.9.6, scikit-learn, and tensor flow libraries are used for models. The dataset was split into training and test sets with an 80:20 ratio, and each model was assessed based on its ability to predict food consumption accurately.

In the data analysis section, three approaches offered:

- First approach is using the actual departure time as the feature.
- Second approach is using time intervals defined in Figure 3.13.
- Third approach is merging the daily total transactions per route and predicting the total consumption instead of per flight predictions. This approach needs aggregate sale quantities per meal group per day.

To decide the final approach, initial model runs are completed for both models and approaches.

Table 5.1: Comparison of approaches with two candidate models

	XGBoost		**ANN**	
	R^2	RMSE	R^2	RMSE
First Approach	0.4901	23.91	0.4632	27.21
Second Approach	0.4815	25.83	0.4311	29.17
Third Approach	0.6112	13.55	0.6149	13.49

All three approaches are tested with initial parameters and original standard features from the dataset. Table 5.1 shows the results for all candidate approaches. The third approach clearly outperforms other approaches in both performance metrics. The third approach aggregates the dataset size 130k rows.

The next step is adding all features and testing the models with all associated tuning parameters.

5.1 XGBoost Results

XGBoost's important tuning parameters are learning_rate, max_depth, min_child_weight, n_estimators, subsample, colsample_bytree and gamma. All parameters are explained in the 3.3.1 section. For performance analysis, the most significant candidate values are selected for each parameter, and results are processed for each parameter combination. Other significant parameters are the objective which is the only alternative for the regression type reg:squarederror and early_stopping_rounds which is the input for the training function.

Table 5.2 shows the candidate tuning parameter alternatives for the XGBoost model. The GridSearchCV technique is a widely used hyperparameter tuning method that evaluates all possible combinations for XGBoost models in similar studies (Darmawan et al., 2022).

All possible combinations (8.6k) are processed during the research. Each combination was tested through separate runs, and the corresponding results were recorded.

Table 5.2: XGBoost Candidate Tuning Parameters

Parameter Name	Default Value	Parameter Set
learning_rate	0.3	0.01, 0.0.5, 0.1, 0.3, 0.5
n_estimators	100	100, 200, 500
max_depth	6	3,6,8,10
min_child_weight	1	1,5
subsample	1	0.8, 1.0
colsample_bytree	1	0.8, 1.0
gamma	0	0,1
early_stopping_rounds	10	10, 30, 50

Table 5.3: XGBoost Parameter Set Sample Results

	Parameter Set	R^2	RMSE
1	{'colsample_bytree': 1.0, 'gamma': 1, 'learning_rate': 0.3, 'max_depth': 8, 'min_child_weight': 1, 'n_estimators': 500, 'subsample': 0.8, 'early_stopping_rounds': 10}	0.9199	6.4303
2	{'colsample_bytree': 0.8, 'gamma': 0, 'learning_rate': 0.3, 'max_depth': 8, 'min_child_weight': 1, 'n_estimators': 500, 'subsample': 0.8, 'early_stopping_rounds': 10}	0.9177	6.5161
3	{'colsample_bytree': 1.0, 'gamma': 1, 'learning_rate': 0.1, 'max_depth': 8, 'min_child_weight': 1, 'n_estimators': 500, 'subsample': 0.8, 'early_stopping_rounds': 10}	0.8930	7.4301
4	{'colsample_bytree': 1.0, 'gamma': 1, 'learning_rate': 0.1, 'max_depth': 8, 'min_child_weight': 1, 'n_estimators': 100, 'subsample': 1.0, 'early_stopping_rounds': 10}	0.8586	8.6403
5	{'colsample_bytree': 0.8, 'gamma': 1, 'learning_rate': 0.5, 'max_depth': 6, 'min_child_weight': 5, 'n_estimators': 200, 'subsample': 1.0, 'early_stopping_rounds': 10}	0.8576	8.5732
6	{'colsample_bytree': 1.0, 'gamma': 1, 'learning_rate': 0.3, 'max_depth': 6, 'min_child_weight': 1, 'n_estimators': 100, 'subsample': 1.0, 'early_stopping_rounds': 10}	0.8140	9.7669
7	{'colsample_bytree': 1.0, 'gamma': 0, 'learning_rate': 0.05, 'max_depth': 6, 'min_child_weight': 1, 'n_estimators': 500, 'subsample': 0.8, 'early_stopping_rounds': 10}	0.8120	9.8501
8	{'colsample_bytree': 1.0, 'gamma': 0, 'learning_rate': 0.01, 'max_depth': 6, 'min_child_weight': 5, 'n_estimators': 500, 'subsample': 1.0, 'early_stopping_rounds': 10}	0.6828	12.7952
9	{'colsample_bytree': 1.0, 'gamma': 1, 'learning_rate': 0.01, 'max_depth': 8, 'min_child_weight': 1, 'n_estimators': 100, 'subsample': 1.0, 'early_stopping_rounds': 10}	0.5137	15.8433
10	{'colsample_bytree': 1.0, 'gamma': 1, 'learning_rate': 0.3, 'max_depth': 6, 'min_child_weight': 1, 'n_estimators': 100, 'subsample': 1.0, 'early_stopping_rounds': 30}	0.8140	9.7669
11	{'colsample_bytree': 1.0, 'gamma': 1, 'learning_rate': 0.1, 'max_depth': 8, 'min_child_weight': 1, 'n_estimators': 100, 'subsample': 1.0, 'early_stopping_rounds': 30}	0.8586	8.6403

Table 5.3 presents a subset of significant results from the extensive model tuning process, showcasing selected parameter combinations from a total of approximately 8.6k tested sets. The table highlights only the most relevant configurations, chosen to illustrate the impact of parameter adjustments on model performance.

In this research, an R^2 score above 0.85 and an RMSE below 9 are considered acceptable thresholds for model performance. While some results in the sample set fall below these criteria, they have been included to demonstrate the effects of parameter tuning and to highlight the trade-offs involved in optimizing the model.

To finalize the selection of optimal parameters, it is essential to carefully evaluate overfitting and underfitting tendencies. This requires examining both validation and training losses to ensure that the chosen model configuration generalizes well to unseen data. Only after assessing these factors should the final parameter set be determined, ensuring a balance between model accuracy and robustness.

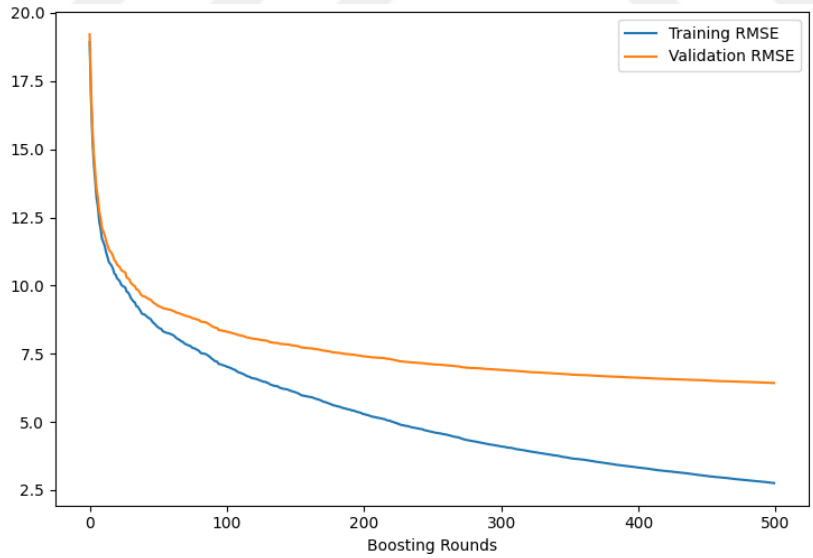


Figure 5.1: Loss Plot for XGBoost Result 1

Figure 5.1 shows the test and validation loss plot for parameters from the first row of the Table 5.3. The gap between the training and the validation lines starts to increasing early and validation line start to become flat.

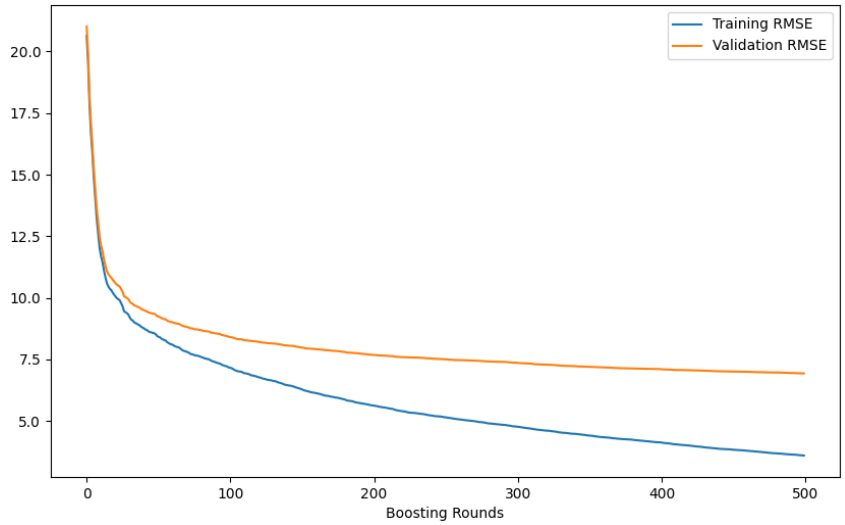


Figure 5.2: Loss Plot for XGBoost Result 3

Figure 5.2 shows the test and validation loss plot for parameters from Table 5.3 third row. The gap between the training and the validation lines starts increasing early, and for this parameter, the validation line starts to become more flattened.

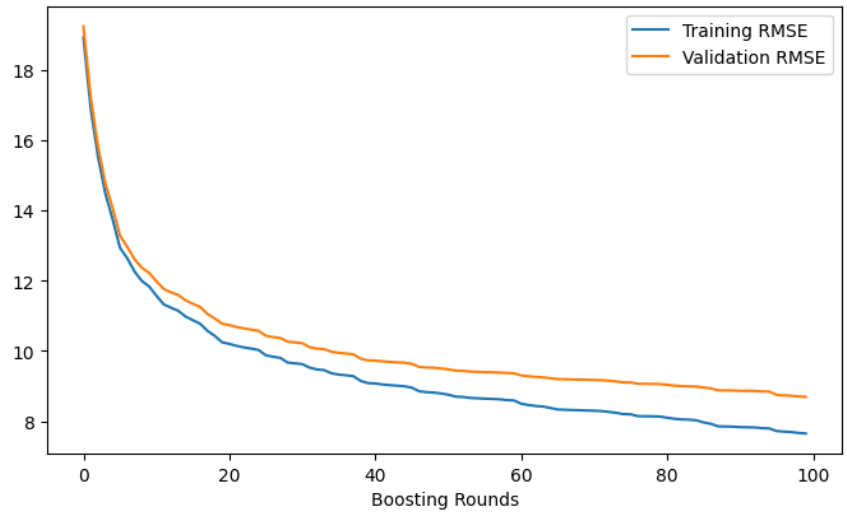


Figure 5.3: Loss Plot for XGBoost Result 4

Figure 5.3 shows the test and validation loss plot for parameters from Table 5.3, 4th row. The gap between the training and the validation lines starts increasing smoothly and

reflects the changes on both lines. The Validation line does not become flattened early and not at all. Figure 5.1 and Figure 5.2 show the signs of overfitting. R^2 and RMSE parameters are both in the acceptable range, and the loss plot does not show any sign of overfitting. Among the candidate parameter set, Table 5.4 shows the values that offer the most applicable results.

Table 5.4: XGBoost Best Parameters

Parameter Name	Value
learning_rate	0.3
n_estimators	100
max_depth	8
min_child_weight	1
subsample	1.0
colsample_bytree	1.0
gamma	1
early_stopping_rounds	10

Top 5 feature importances are listed in Table 5.5 below. Feature importances are based on gain importance type among three options. Gain is selected because the most reliable indicator of feature importance is this type. Other types are weight and cover.

Table 5.5: XGBoost Feature Importances

	Feature	Importance
1	ORIG_FRA	0.095240
2	ORIG_BSL	0.07455891
3	ORIG_ZRH	0.057896443
4	DEST_DUS	0.04633553
5	ORIG_DUS	0.03982096

Feature importances suggest that destination and origin features are more important and significant for model performance on this dataset. Other important features not listed here but available in the top 100 are duration, day of the year, and total passenger count.

5.2 ANN Results

In this study, ANNs are used for regression-based predictions. The Tensorflow library is used with the Python programming language. The RelU activation function is used for hidden layers since regression uses a linear activation function. Sequential model used 3 layers:

- Input Layer: 64 Units with ReLU activation
- Hidden Layer: 32 Units with ReLU activation
- Output Layer: 1 unit with Linear activation

Table 5.6: ANN Candidate Tuning Parameters

Parameter Name	**Default Value**	**Parameter Set**
learning_rate	0.001	0.01, 0.001
epochs	50	50, 100
batch_size	32	32, 64
optimizer	Adam	Adam, RMSprop

Table 5.6 shows the candidate tuning parameters for the ANN model. Hyper parameter tuning is accomplished through the RandomizedSearchCV technique that is applied in similar studies (Sharma et al., 2023). Using this technique, different random combinations for selected parameters are selected, and separate runs are processed.

Table 5.7 displays sample results from the model's tuning process, featuring various combinations of parameters. Each parameter configuration demonstrates a significant impact on model performance. While only one R^2 score exceeds the threshold of 0.85, other strong candidates fall within the 0.80 to 0.85 range. These configurations approach

the acceptable performance criteria but may require further refinement to meet the optimal standards.

Table 5.7: ANN Parameter Set Sample Results

	Parameter Set	R^2	RMSE
1	{ 'learning_rate': 0.001, 'batch_size': 32, 'epochs': 100, 'optimizer': 'Adam'}	0.7739	10.9331
2	{ 'learning_rate': 0.1, 'batch_size': 32, 'epochs': 100, 'optimizer': 'Adam'}	0.6241	14.0981
3	{ 'learning_rate': 0.001, 'batch_size': 32, 'epochs': 50, 'optimizer': 'Adam'}	0.8161	9.8570
4	{ 'learning_rate': 0.01, 'batch_size': 32, 'epochs': 50, 'optimizer': 'Adam'}	0.8012	10.2532
5	{ 'learning_rate': 0.001, 'batch_size': 64, 'epochs': 50, 'optimizer': 'Adam'}	0.7981	10.7201
6	{ 'learning_rate': 0.001, 'batch_size': 64, 'epochs': 100, 'optimizer': 'Adam'}	0.8208	9.7341
7	{ 'learning_rate': 0.01, 'batch_size': 64, 'epochs': 100 , 'optimizer': 'Adam'}	0.8302	9.472
8	{ 'learning_rate': 0.001, 'batch_size': 32, 'epochs': 100, 'optimizer': 'RMSprop'}	0.7358	11.8132
9	{ 'learning_rate': 0.001, 'batch_size': 32, 'epochs': 100, 'optimizer': 'RMSprop'}	0.6702	13.2049
10	{ 'learning_rate': 0.001, 'batch_size': 64, 'epochs': 100, 'optimizer': 'RMSprop'}	0.7333	12.0960
11	{ 'learning_rate': 0.01, 'batch_size': 32, 'epochs': 50, 'optimizer': 'RMSprop'}	0.8586	8.6403

Additionally, several results with lower R^2 values are included to illustrate the effects of different parameter choices, even when they do not meet the target criteria. This approach provides insights into how specific parameter adjustments influence model accuracy, highlighting the sensitivity of the model to tuning decisions.

To gain a comprehensive understanding of model performance, it is also necessary to analyze the loss plots for each parameter set. Interpreting these training and validation loss trends can help identify overfitting or underfitting issues, leading to a more informed selection of final parameters that balance accuracy with generalization.

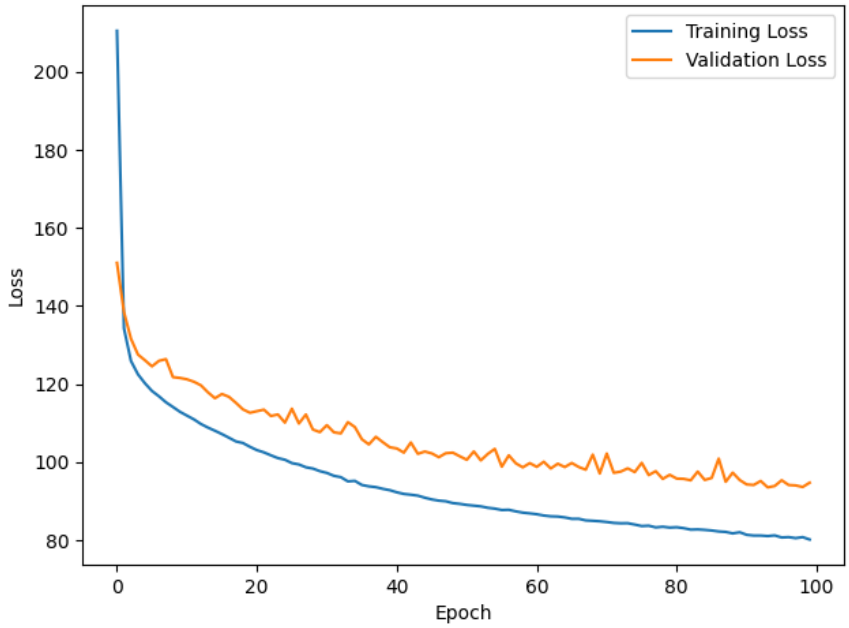


Figure 5.4: Loss Plot for ANN Result 3

Figure 5.4 shows the test and validation loss plot for parameters from Table 5.7, third row. The validation loss line has some simple fluctuations. This may be interpreted as a sign of insufficient validation data.

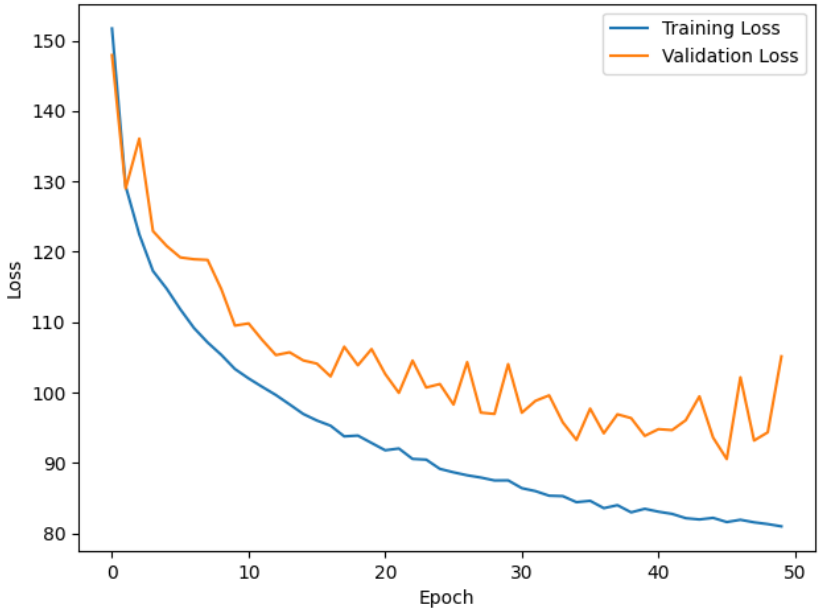


Figure 5.5: Loss Plot for ANN Result 4

Figure 5.5 shows the test and validation loss plot for parameters from Table 5.7, fourth row. The validation loss line has a lot of significant fluctuations. And the last portion of the validation loss line trend is rising. This may be interpreted as a clear sign of overfitting. Fluctuations may be the sign of insufficient data or a high learning rate.

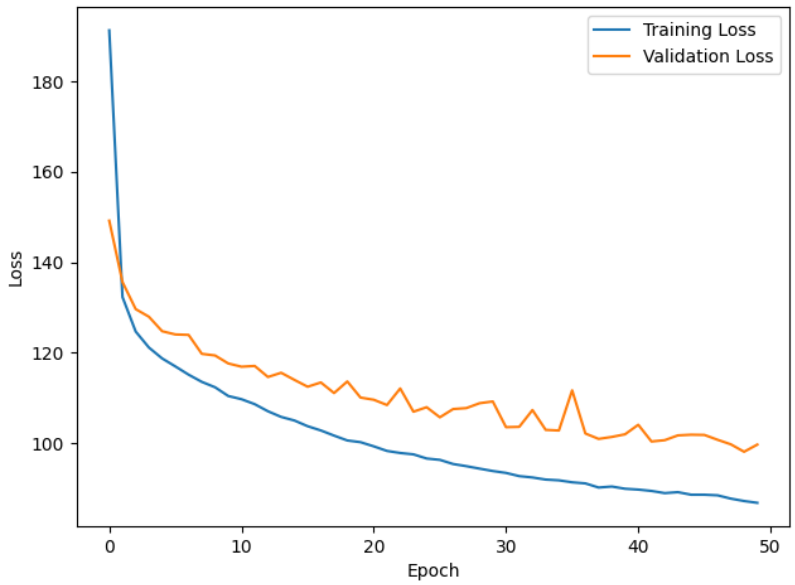


Figure 5.6: Loss Plot for ANN Result 6

Parameter set 6 loss plot shows similarities with the parameter set 3 on Figure 5.6. Fluctuations are more significant but not too sharp. The gap between the loss lines is increasing but not too much. There is no clear sign of overfitting, but fluctuations may suggest insufficient validation data.

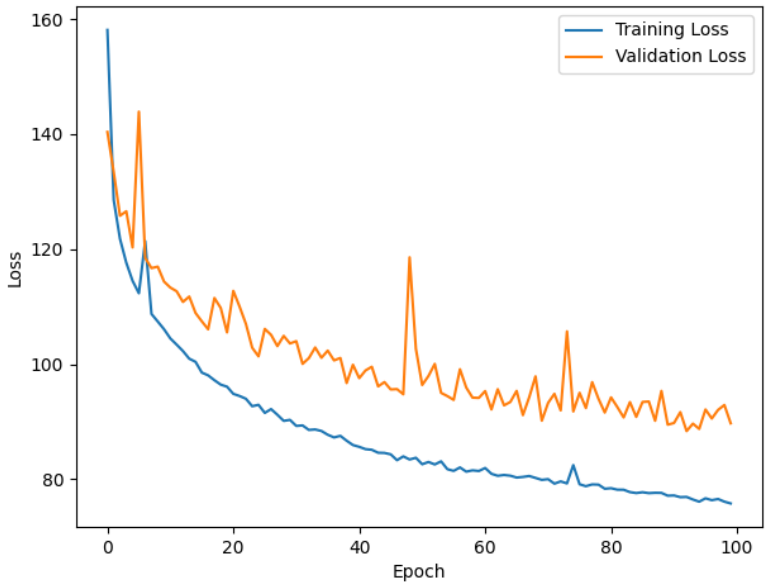


Figure 5.7: Loss Plot for ANN Result 7

Parameter set 7 shows sharp fluctuations on validation loss lines in Figure 5.7. Also, the training line has some simple but significant fluctuations. The learning rate is the main reason behind this result. Fluctuations on both lines are not fully compatible; this shows poor generalization. Increasing the learning rate does not perform well on the data set.

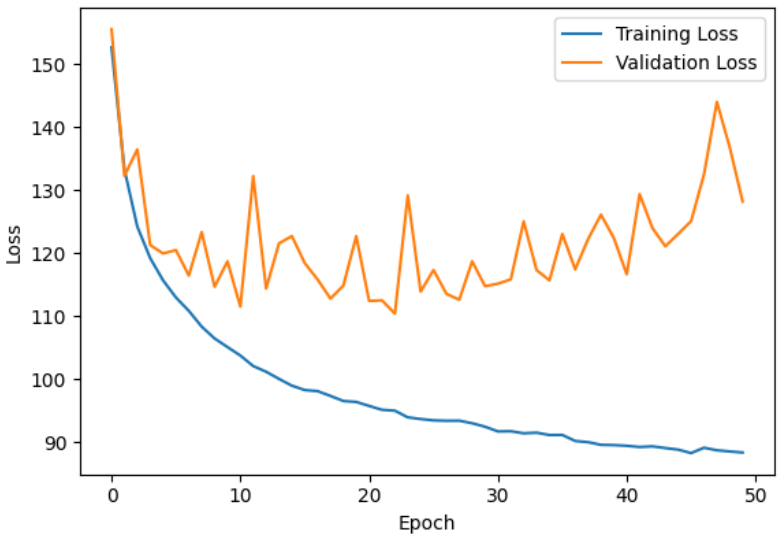


Figure 5.8: Loss Plot for ANN Result 11

Figure 5.8 shows the loss plot for the dataset with the best R^2 and RMSE values. Validation loss has a lot of sharp fluctuations, which may be the sign of insufficient validation data. The validation line trend keeps increasing after some point; this may be the sign of overfit. This parameter set has an RMSprop optimizer function that is different from other possible candidates.

Interpretation of Figure 5.4 to Figure 5.8 shows fluctuations in validation losses. This may be considered as the insufficient data for the ANN model. The Adam optimizer generates more proper results. The parameter candidate is the third parameter set.

Table 5.8: ANN Best Parameters

Parameter Name	**Parameter Set**
learning_rate	0.001
epochs	50
batch_size	32
optimizer	Adam

Top 5 feature weights are listed in Table 5.9. Weights is the sign for the significance of the feature on the final prediction. Weight indicates how much a given input feature contributes to the output of the mode.

Table 5.9: Top 5 Feature Weights ANN

	Feature	**Importance**
1	day_of_year	9.352919
2	day_of_week	6.6551905
3	month	6.582795
4	total_pax	5.51459
5	is_weekend	4.936882

ANN top feature weights show that the ANN model is more sensitive to date-related features. Duration and total passenger count are also other significant features in the top 10. Destination and origin features are distributed among the lists.

5.3 Model Comparison Analysis

In order to interpret the results between two models, R^2 scores and RMSE values are considered. Table 5.10 shows the R^2 score and RMSE for both models.

Table 5.10: XGBoost and ANN Model Results

Feature Name	R^2	RMSE
XGBoost	0.8586	8.64
ANN	0.8161	9.85

The R^2 score indicates how well each model captures the variance in the target variable. The ANN performed slightly better than XGBoost, explaining more variance (85.8% vs. 81.6%). This suggests that the ANN may be better at learning the overall relationships between the input features and the target variable, especially when dealing with complex, non-linear patterns. This could be due to the ANN's ability to model non-linear relationships more comprehensively through its multiple layers of neurons.

The RMSE measures the average magnitude of prediction errors, with lower values indicating better accuracy. Here, XGBoost outperformed ANN, with an RMSE of 8.64 compared to 9.85 for the ANN. The lower RMSE of XGBoost suggests that it produces more accurate predictions for individual instances. In practical terms, this means that XGBoost is likely to make fewer extreme errors, which is critical in ensuring operational efficiency in airline catering.

6. CONCLUSION

This research has successfully demonstrated the application of XGBoost and Artificial Neural Networks (ANNs), in predicting fresh food consumption for buy-on-board services in the aviation industry. The results obtained from both models are promising and provide a solid foundation for enhancing inventory management and reducing waste in airline catering services.

The XGBoost model demonstrated strong predictive performance with an R^2 value of 0.8586 (which is better than the ANN model) and an RMSE of 8.64. This indicates that the model was able to capture the majority of the variance in the data while maintaining precision in individual predictions. The lower RMSE makes XGBoost particularly suitable for operational deployment, where accurate predictions of individual flights' food needs are essential to prevent waste and ensure adequate provisioning.

The ANN model achieved lower R^2 value of 0.8161, demonstrating its capability is less likely to capture more complex, non-linear relationships in the dataset. Also, the RMSE of 9.85 indicates that the ANN model struggled with prediction accuracy on an individual level compared to XGBoost. The higher variability in its predictions could lead to inconsistencies that may negatively impact inventory planning if deployed without further refinement. There are signs of insufficient data in the loss plots. Expanding the time interval for the data set may increase the performance of the ANN model.

The application of XGBoost and ANN to predict fresh food demand provides a data-driven methodology that can significantly enhance the accuracy of food supply forecasting, reducing both excess and insufficient provisioning. Deploying XGBoost for real-time prediction of fresh food needs can lead to significant cost savings by minimizing

food waste for each flight. This contributes to both economic efficiency and environmental sustainability, which are the key priorities for the airline industry.

The absence of some historical data due to operational disruptions during the COVID-19 pandemic has likely affected the models' learning. Although there is little that can be done about lost historical data, this limitation highlights the importance of robust data collection strategies. Especially the ANN model is more sensitive to the amount of data. Rather than playing with the model parameters, dataset modifications may generate better results.

The ANN model is more sensitive to date features and the total number of passengers, which is considered suitable with the nature of the dataset for initial impressions. However, XGBoost importances are focused on destinations, and results are more accurate. The dataset suggests more complex relations than the nature of the data reflects.

In addition to the Artificial Neural Network (ANN) model used in this research, a sample Long Short-Term Memory (LSTM) model with default parameters was also explored as a preliminary approach to assess its potential. LSTM networks are well-suited for capturing sequential dependencies, which can be advantageous for demand forecasting involving temporal data. However, initial results from the sample LSTM model were not substantially different from those achieved by the ANN, reflecting the limitations of the current dataset in fully leveraging LSTM's capabilities. The relatively small dataset used in this study posed challenges for both ANN and LSTM models, highlighting the need for a larger and more diverse dataset to effectively utilize the temporal strengths of LSTM. Although LSTMs could theoretically enhance model performance by capturing longer-term dependencies, further exploration would require an expanded dataset to allow the model to achieve optimal predictive accuracy in comparison to ANN. Details of the LSTM study are added to the appendix.

For future work this study can be extended to cover some critical external factors. Incorporating loyalty data for known passengers could refine predictions further, allowing for more personalized forecasting that considers individual buying habits. Adding new features to reflect seasonal or local events, public holidays (both the base for

airline country and all destinations). Another approach may be combining multiple methods to improve predictions.

The use of AI in forecasting fresh food consumption for airlines provides a powerful tool to address the operational challenges posed by fluctuating demand. By implementing XGBoost and ANN, this study has demonstrated the feasibility of reducing forecasting errors, thereby contributing to reduced food waste and enhanced customer satisfaction. As airlines continue to navigate challenges related to operational efficiency and sustainability, data-driven approaches such as those presented here can play an instrumental role in optimizing resource usage and enhancing overall service quality.

REFERENCES

Abdikan, S., Sekertekin, A., Narin, O.G., Delen, A., Sanli, F.B. (2023). A comparative analysis of SLR, MLR, ANN, XGBoost and CNN for crop height estimation of sunflower using Sentinel-1 and Sentinel-2, *Advances in Space Research* 71(7): 3045–3059.

Bozkir, A.S. and Sezer, E.A. (2010). Predicting food demand in food courts by decision tree approaches. *Procedia Computer Science* 3: 759–763.

Caswell, M. (2020). Emirates Flight Catering to use AI to reduce food waste by 35 percent URL: *https://www.businesstraveller.com/business-travel/2020/10/18/emirates-flight-catering-to-use-ai-to-reduce-food-waste-by-35-per-cent*

Chai, T. and Draxler, R. R. (2014): Root mean square error (RMSE) or mean absolute error (MAE)? – Arguments against avoiding RMSE in the literature, Geosci. *Model Dev*., 7: 1247–1250.

Chakraborty, D. and Elzarka, H. (2017). Advanced machine learning techniques for building performance simulation: a comparative analysis, *Journal of Building Performance Simulation* 12(2): 193–207.

Chen, T., and Guestrin, C. (2016). XGBoost: A scalable tree boosting system. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining.

Chicco D, Warrens MJ, Jurman G. (2021). The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation. *PeerJ Computer Science* 7(3): e623.

Coggins, P. (1994). Reduce and recycle at home, at work, and at play: perspectives on waste management and travel and tourism. *J. Waste Manag. Resour. Recovery* 1: 1-10.

Croda, R.M.C., Romero, D.E.G., Morales, S.O.C., (2017). Sales Prediction through Neural Networks for a Small Dataset. *International Journal of Interactive Multimedia and Artificial Intelligence*,5: 35–41.

Darmawan, H., Yuliana M., Hadi, M.Z.S. (2022). GRU and XGBoost Performance with Hyperparameter Tuning Using GridSearchCV and Bayesian Optimization on an IoT-Based Weather Prediction System. *International Journal on Advanced Science Engineering Information Technology* 13(3): 851-862.

Ernits, R. M., Reiß, M., Bauer, M., Becker A. (2022). Individualisation of Inflight Catering Meals—An Automation Concept for Integrating Pre-Ordered Meals,. *Aerospace*,9: 736.

Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. *Annals of Statistics*, 29(5): 1189-1232.

Garcia-Garcia, G., Woolley, E., Rahimifard, S. (2015). A Framework for a More Efficient Approach to Food Waste Management. *International Journal of Food Engineering* 1(1): 65-72.

Geem, Z.W. and Roper, W.E. (2009). Energy demand estimation of South Korea using artificial neural network. *Energy Policy* 37(10): 4049–4054.

Goto J. H., Lewis, M.E., Puterman M.L. (2004). Coffee, tea, or?: A Markov decision process model for airline meal provisioning. Transportation Science, 38: 107–118.

Humaira H. and Rasyidah R. (2020). Determining The Appropiate Cluster Number Using Elbow Method for K-Means Algorithm. *Proceedings of the WMA-2*, Padang, Indonesia.

IdeaWorksCompany (2022). Airline Ancillary Revenue Nears Pre-Pandemic Level with a 56% Increase to $102.8 Billion for 2022, IdeaWorksCompany Press Release URL: *https://ideaworkscompany.com/wp-content/uploads/2022/11/Press-Release-169-Global-Estimate-2022.pdf*

IdeaWorksCompany (2024), CarTrawler Yearbook of Ancillary Revenue (2024) URL: *https://ideaworkscompany.com/2024-cartrawler-yearbook-of-ancillary-revenue-report/*

InterVistas (2016). Istanbul Technical University, Airline Business Models URL: *http://aviation.itu.edu.tr/%5Cimg%5Caviation%5Cdatafiles/Lecture%20Notes/Aviation%20Economics%20and%20Financial%20Analysis%2020152016/Readings/Module%2007/Airline%20Business%20Models.pdf*

Jain, A., Menon, M.N., Chandra, S. (2015). Sales Forecasting for Retail Chains URL: *https://cseweb.ucsd.edu/classes/wi15/cse255-a/reports/fa15/004.pdf*

Kingma, D. P. and Ba, J. (2014). Adam: A method for stochastic optimization, *International Conference on Learning Representations*.

Levis, A.A. and Papageorgiou, L.G. (2005). Customer Demand Forecasting via Support Vector Regression Analysis. *Chemical Engineering Research and Design*, 83: 1009–1018.

Lloyd, S. (1982). Least squares quantization in PCM. *IEEE Transactions on Information Theory*, 28(2): 129–137.

Makridakis, S. and Hibon, M. (2000) The M3-Competition: results, conclusions and implications, *International Journal of Forecasting* 16: 451–476.

Mason, K. J. (2005). Observations of fundamental changes in the demand for aviation ser-vices. *Journal of Air Transport Management* 11: 19–25.

McCulloch, W.S. and Pitts, W. (1943). A logical calculus of the ideas immanent in nervous activity. *Bulletin of Mathematical Biophysics* 5: 115–133.

Megodawickrama, P.L. (2018). Significance of Meal Forecasting in Airline Catering on Food Waste Minimization, *15th International Conference on Business Management, ICBM 2018*, Colombo, Sri Lanka, pp. 1131-1146.

Mitchell T. (1997). Machine learning. McGraw-Hill series in artificial intelligence. New York: McGraw-Hill

Nwankpa, C., Ijomah, W., Gachagan, A., & Marshall, S. (2018). Activation Functions: Comparison of trends in practice and research for deep learning. *2nd International Conference on Computational Sciences and Technologies*, Jamshoro 2020, pp. 124-133.

Rousseeuw, P.J. (1987). Silhouettes: a graphical aid to the interpretation and validation of cluster analysis, Journal of Computational and Applied Mathematics, 20: 53–65.

Rumelhart, D. E., Hinton, G. E., Williams, R. J. (1986). Learning representations by back-propagating errors. Nature, 323(6088): 533-536.

Samunderu, E. and Farrugia M. (2022). Predicting customer purpose of travel in a low-cost travel environment—A Machine Learning Approach, *Machine Learning with Applications: Volume* 9.

Shahapure, R.K. and Nicholas C. (2020) Cluster Quality Analysis Using Silhoutte Score, 2020 IEEE 7th International Conference on Data Science and Advanced Analytics (DSAA)

Sharma N., Malviya L., Jadhav A., Lalwani P., (2023). A hybrid deep neural net learning model for predicting Coronary Heart Disease using Randomized Search Cross-Validation Optimization, Decision Analytics Journal 9.

Silva, A. and Cardoso, M. (2005). Predicting supermarket sales: The use of regression trees. J Target Meas Anal Mark 13, 239–249.

Srisaeng, P., Baxter, G., Wild, G. (2015). Forecasting demand for Low Cost Carriers in Australia using an artificial neural network approach, *Aviation* 19(2): 90–103.

Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. (2014). Dropout: A simple way to prevent neural networks from overfitting. Journal of Machine Learning Research, 15(1): 1929-1958.

Stevanović S, Dashti H, Milošević M, Al-Yakoob S, Stevanović D (2024) Comparison of ANN and XGBoost surrogate models trained on small numbers of building energy simulations. PLoS ONE 19(10): e0312573.

Taylor, S.J. and Letham, B., (2017). Forecasting at Scale. Peerj Preprints 5:e3190v2.

Vacanti, N.M. (2019). The Fundamentals of Constructing and Interpreting Heat Maps. In: Fendt, SM., Lunt, S. (eds) Metabolic Signaling. Methods in Molecular Biology, vol 1862. Humana Press, New York, NY.

Walt, A. and Bean, W.L. (2022). Inventory management for the in-flight catering industry: A case of uncertain demand and product substitutability, *Computers & Industrial Engineering: Volume* 165.

Yılmaz, S.B. and Yücel, E. (2021). Optimizing onboard catering loading locations and plans for airlines. Omega [online]. Volume 99 - Article 102301 URL: https://www.sciencedirect.com/science/article/pii/S0305048320306551 [accessed September 5, 2024]

You, F., Bhamra, T., Lilley, D. (2020). Why is airline food always dreadful? analysis of factors influencing passengers' food wasting behaviour. Sustainability, 12, 8571.

Zhang, X., Yan, C., Gao, C. et al.(2019) Predicting Missing Values in Medical Data Via XGBoost Regression. J Healthc Inform Res 4: 383–394.

Zhao, N., Liu, Y., Vanos, J.K., Cao, G. (2018) Day-of-week and seasonal patterns of PM2.5 concentrations over the United States: Time-series analyses using the Prophet procedure, Atmospheric Environment 192 pp. 116–127.

APPENDIX

In this study LSTM is used for regression-based predictions. Tensorflow library is used with python programming language. Sequential model used 3 layers and dropouts between each layer with 0.2 rate:

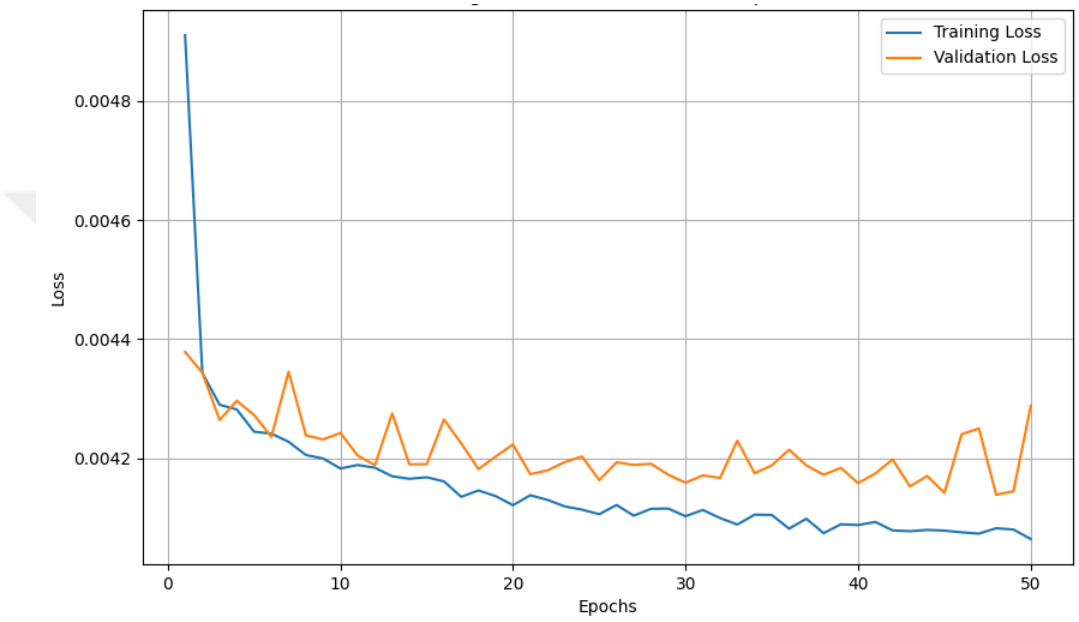
- Input Layer: LSTM 50 Units, return_sequences=True, input_shape(tims_steps)
- Hidden Layer: LSTM 50 Units, return_sequences=False
- Output Layer: 1 unit Dense for regression

Parameter	**Value**	**Description**
Units (Neurons)	50	Number of neurons in the LSTM layer
Dropout Rate	0.2	Fraction of units to drop for regularization.
Batch Size	32	Number of samples per gradient update.
Epochs	50	Total number of training iterations over the entire dataset.
Learning Rate	0.001	Default learning rate
Optimizer	Adam	Adaptive Moment Estimation, a common optimizer for LSTM models.
Loss Function	Mean Squarred Error (MSE)	Standard loss function for regression tasks, measuring prediction error.

Initial Results:

- R^2 Score 0.6496
- RMSE 13.5056

The initial results indicate an R^2 score of 0.6496 and an RMSE of 13.5056, suggesting that the LSTM model achieves a moderate level of accuracy but does not surpass the performance of the ANN model in this context. This outcome reflects the constraints posed by the dataset size, which may limit the LSTM's capacity to leverage its strength in capturing sequential dependencies effectively.



The mismatch between training and validation loss trends could indicate overfitting or a high variance problem, where the model learns patterns specific to the training data but struggles to generalize to new data. The sharp fluctuations in validation loss may also suggest that the dataset size is insufficient for the LSTM's complexity, leading to inconsistencies in validation performance. More data would help the model generalize better and reduce the variance seen in validation performance.

BIOGRAPHICAL SKETCH

Salih Enver Yurter is currently pursuing a master's degree in computer engineering at Galatasaray University. He is applying advanced machine learning and predictive modeling techniques to solve operational challenges in the airline industry, specifically focusing on forecasting fresh food consumption to optimize airline catering operations. This work reflects his interest in using technology to create data-driven, impactful solutions in dynamic environments.

Salih Enver Yurter is graduated with a Bachelor of Science (B.S.) in Computer Engineering from Bilkent University in 2009. Since then, he has amassed substantial industry experience across diverse sectors, including defense, airline, finance, and e-commerce industries. His career began as a Software Developer, and through dedication and skill, he progressed to Team Leader, Software Architect, Manager, Director, and currently holds the position of Chief Technology Officer (CTO). His experience spans a range of responsibilities, from technical implementation and software architecture to leading teams and managing large-scale projects.