LEVERAGING LARGE-SCALE DATA FOR SUPPLY CHAIN NETWORK DESIGN: A LOCATION-ALLOCATION MODEL FOR RWANDA

A THESIS SUBMITTED TO

THE GRADUATE SCHOOL OF ENGINEERING AND SCIENCE

OF BILKENT UNIVERSITY

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR

THE DEGREE OF

MASTER OF SCIENCE

IN

INDUSTRIAL ENGINEERING

By

Zeynep Göze Gürkan

August 2024

LEVERAGING LARGE-SCALE DATA FOR SUPPLY CHAIN NETWORK DESIGN: A LOCATION-ALLOCATION MODEL FOR RWANDA

By Zeynep Göze Gürkan

August 2024

We certify that we have read this thesis and that in our opinion it is fully adequate, in scope and in quality, as a thesis for the degree of Master of Science.

Ayşe Selin Kocaman(Advisor)

Pablo Duenas Martinez(Co-Advisor)

Firdevs Ulus

Saltuk Buğra Selçuklu

Approved for the Graduate School of Engineering and Science:

Orhan Arıkan
Director of the Graduate School

ABSTRACT

LEVERAGING LARGE-SCALE DATA FOR SUPPLY CHAIN NETWORK DESIGN: A LOCATION-ALLOCATION MODEL FOR RWANDA

Zeynep Göze Gürkan
M.S. in Industrial Engineering
Advisor: Ayşe Selin Kocaman
Co-Advisor: Pablo Duenas Martinez
August 2024

Clean cooking strategies are significant contributors to the enhancement of development and sustainability. Regarding the lower emission levels compared to biomass usage, we consider Liquefied Petroleum Gas (LPG) a clean cooking strategy to promote, especially in developing countries. Hence, we design large-scale supply chain operations for the LPG distribution in Rwanda. This involves addressing the location-allocation problem of facilities by utilizing a large dataset on the location and LPG demand of each rooftop by formulating a Mixed-Integer Linear Programming (MILP) model. In order to decrease the size of the problem, we propose three methods. First of all, we design the system independent of time index. Next, we use the agglomerative hierarchical clustering-based heuristic approach to cluster the rooftops and locate retailers on the distance-constrained geomedian point of each cluster. Finally, we propose to decompose the formulated MILP model to get adequate solutions in less time. For computational analysis, we compare the system configurations with different retailer locations obtained by the village centroid approach and agglomerative hierarchical clustering based-heuristic approach. In addition, we investigate whether the existing system configuration can be extended when the projected increase in yearly LPG demand is introduced. Moreover, we conduct a sensitivity analysis to show the trade-off between the infrastructure and transportation costs due to the volatility in diesel fuel prices. Finally, we compare the results and performances of the main model and the decomposed model.

Keywords: Energy access, clean cooking, SDG7, location-allocation, agglomerative clustering.

ÖZET

BÜYÜK ÖLÇEKLİ VERİ KULLANARAK TEDARİK ZİNCİRİ AĞI TASARIMI: RUANDA İÇİN BİR YERLEŞİM-ATAMA MODELİ

Zeynep Göze Gürkan
Endüstri Mühendisliği, Yüksek Lisans
Tez Danışmanı: Ayşe Selin Kocaman
İkinci Tez Danışmanı: Pablo Duenas Martinez
Ağustos 2024

Temiz pişirme stratejileri, kalkınma ve sürdürülebilirliğin iyileştirilmesine katkı sağlayan elementlerdir. Biyokütle kullanımına kıyasla daha düşük emisyon seviyesine sahip olması nedeniyle Sıvılaştırılmış Petrol Gazı (LPG) gelişmeke olan ülkelerde teşvik edilmesi gereken bir temiz pişirme stratejisi olarak değerlendirilmektedir. Bu nedenle, bu çalışmada, Ruanda'da LPG dağıtımı için büyük ölçekli tedarik zinciri operasyonu tasarlanmıştır. Bu kapsamda evlerin konum ve LPG talebini içeren büyük ölçekli bir veri seti kullanarak tesislerin yerleşim-atama problemi bir Karma-Tamsayılı Doğrusal Programlama (MILP) modeli formüle ederek çözülmüştür. Problemin boyutunu azaltmak için üç yöntem önerilmiştir. İlki, sistemi zaman indeksinden bağımsız tasarlamaktır. İkinci olarak, çatıları kümelemek ve mesafe kısıtlı jeomedyan noktasına perakendeciler yerleştirmek için yığışımlı hiyerarşik kümeleme tabanlı sezgisel yaklaşımı kullanılmıştır. Son olarak, az zamanda uygun çözümler elde etmek için MILP modelini ayrıştırma yöntemi önerilmiştir. Analiz için, köy merkezi yaklaşımı ve yığışımlı hiyerarşik kümeleme tabanlı sezgisel yaklaşımla kurulan sistem yapılandırmaları karşılaştırılmıştır. Ayrıca, yıllık LPG talebindeki projeksiyon artışı dikkate alındığında mevcut sistem yapılandırmasının genişletilip genişletilemeyeceği araştırılmıştır. Aynı zamada, dizel yakıt fiyatlarındaki dalgalanma nedeniyle altyapı ve ulaşım maliyetleri arasındaki dengeyi göstermek için bir duyarlılık analizi yapılmıştır. Son olarak, ana model ile ayrıştırılmış modelin sonuçlarını ve performansları karşılaştırılmıştır.

Anahtar sözcükler: Enerji erişimi, temiz pişirme, SDG7, yerleşim-atama, yığışımlı kümeleme.

Acknowledgement

I would like to express my deepest gratitude to my advisor, Assoc. Prof. Ayşe Selin Kocaman, for her unwavering guidance and support throughout this journey. Her insightful advice, constant encouragement, and profound knowledge have been invaluable in shaping this thesis. It has been a remarkable journey from NYC to Rwanda hocam, I am grateful.

I am also thankful to my co-advisor, Research Scientist Pablo Duenas Martinez, for his expertise and continuous support. His constructive feedback and detailed insights have greatly enhanced the quality of this work.

I extend my sincere thanks to the members of the jury, Assoc. Prof. Firdevs Ulus and Asst. Prof. Saltuk Buğra Selçuklu, for taking the time to review my thesis and for providing invaluable feedback.

I consider myself the luckiest to have Yunus Emre Çakır by my side anytime I need support. I truly cherish every moment we have spent together and would not trade a single minute. Also, I am grateful to Ali Eren Demir. Your friendship means a lot to me, and it holds immense value in any sense.

I am fortunate to have the support of wonderful friends, Mustafa, Deniz Akkaya, Osman, İrem, Doğuş, Deniz Şahin, Rana, Gülin, Sena, Ali, Gizem, Sarp, Batuhan, and of course, Sedo and Lodi who have been by my side throughout this journey. Thank you for all the fun times. You made these two years memorable.

Last but not least, I owe the deepest thanks to my family, Nuriye Acar, Mehmet Gürkan, and Azra Çağlasu Gürkan. You have provided endless love, encouragement, and understanding during both the highs and lows of this journey. I owe everything to you. Sincerest thanks to Gökhan Yetişir and Ceyda Acar for their support and help. And, of course, to Mabel, for always being present just next to my side.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

As a part of Agenda 2030, in 2015, the Sustainable Development Goals (SDGs) were approved by the United Nations to enhance global sustainability by 2030. SDGs comprise 17 measurable objectives that address all facets of sustainability and social domains [1]. One of these objectives, SDG7, aims to ensure "access to affordable, reliable, sustainable and modern energy for all" [2]. This goal highlights the critical role of energy in achieving broader sustainability and development outcomes by increasing the usage of renewable energy resources, improving energy efficiency, expanding energy infrastructure plans, and investing in clean energy technologies [3, 4, 5, 6]. Considering the current inadequate energy access, particularly in developing countries, SGD7 is of significant importance in addressing global inequalities [7, 8, 9, 10].

Clean cooking strategies are significant contributors to sustainability. They address the need for sustainable, safe, and efficient cooking methods, particularly in developing countries where traditional practices are still prevalent. Commonly adopted clean cooking options include biogas, biofuels, ethanol, liquefied petroleum gas (LPG), and electricity. It is important to note that while not all of these resources are renewable, they are considered clean due to their low emission levels. These strategies are integral to achieving SDG7, which focuses on ensuring access to affordable, reliable, sustainable, and modern energy for all

[11, 12, 13, 14].

In many developing countries, biomass is commonly used for heating and cooking, leading to negative impacts on human health and the environment because of emissions [15]. The World Health Organization predicts that approximately 2.3 billion people worldwide cook by burning biomass, kerosene, and coal, causing 3.2 million deaths per year in 2020 due to household air pollution [16]. If current trends in cooking strategies continue, it is estimated that 23% of the world's population will still not be able to adopt clean cooking techniques by 2030, where the sub-Saharan African population constitutes a big portion of this ratio [17].

Rwanda is a pertinent example of these challenges. As a developing and landlocked country in East Africa, Rwanda's population of 13.777 million in 2023 faces significant difficulties related to energy consumption and its impacts [18]. In the country, cooking energy composes a large portion of the household energy consumption, and it heavily relies on fossil fuels [19]. Namely, at least 83% of energy consumption was based on biomass (mostly firewood and charcoal) in 2020, causing significant environmental degradation and health problems [20]. More specifically, between 2001 and 2019, 6.9% of Rwanda's total tree coverage area was destroyed to provide wood for domestic energy [19]. In addition, the Rwanda Ministry of Health reports that each year, over three million Rwandans develop respiratory problems, with air pollution accounting for 13% of these cases [20].

Given the negative impacts of biomass utilization, switching to cleaner alternatives such as LPG has been a foremost objective in Rwanda. In a study conducted to compare the emission levels of different kinds of stoves, it is recorded that LPG stoves have 90% lower mean particulate matter and carbon monoxide emission factors than biomass stoves [21]. In line with this, several studies have been carried out to promote the usage of LPG in Rwanda since it is a clean and available resource. With this motivation in mind, one of the biggest steps taken is to prepare the "Rwanda National LPG Master Plan, Feasibility, and Investment Report" to evaluate the current system and the future outcomes. The master plan asserts that Rwanda's National Strategy for Transformation (NST-1) of 2017 sets the target to reduce the percentage of the Rwandan population relying on biomass as their primary cooking fuel from 83% in 2017 to 42% by 2024 [22]. Promoting the usage of LPG is one of the key elements in fulfilling this goal. According to the report, if the projected LPG usage occurs, 25.6 to 40.1 million metric tonnes of CO2eq emission will be deterred, and 2600 to 7700 premature deaths could be averted cumulatively from 2021 to 2030 [22].

While increasing LPG usage across the country is important, it's important to recognize that distributing LPG to households requires a carefully managed infrastructural plan. Therefore, this thesis is motivated by the need to design a large-scale supply chain for LPG distribution, aiming to promote clean cooking strategies in developing countries. With this motivation in mind, we are addressing a data-driven hierarchical location-allocation problem to model the Rwandan LPG Supply Chain System. Although various studies in the literature focus on the hierarchical location-allocation problem, to the best of our knowledge, using a large-scale dataset to solve this problem within the context of an LPG supply chain system represents a novel approach.

In this thesis, we begin by analyzing the dataset, which contains information on the coordinates of each rooftop in Rwanda and their annual LPG demands. We identify candidate locations for depots, bottling stations, and warehouses based on insights from the Rwanda National LPG Master Plan. To obtain accurate distance measurements between facilities, we use the OSRM API. OSRM API is a routing engine that finds the shortest path between two coordinates. To determine the locations of facilities and their allocations, we formulate a Mixed-Integer Linear Programming (MILP) model to solve our hierarchical location-allocation problem, considering annual operations within the supply chain.

We explore various methods to solve the model efficiently, primarily focusing on reducing the problem's scale. Assuming no seasonality in LPG demand, one approach is to model the problem on a monthly basis rather than annually. Another method we propose involves minimizing the number of retailers—the final points in our supply chain—while adhering to specific conditions. A common approach to determine retailer locations is by selecting the centroid point of each village, but this approach results in a large number of retailers. To address this, we propose an agglomerative hierarchical clustering-based heuristic approach to reduce the number of retailers while maintaining fairness among rooftops. With this heuristic, we cluster rooftops and locate retailers at the distance-constrained geomedian point of each cluster, ensuring the distance between a rooftop and its assigned retailer is less than 1 mile. Lastly, we develop another approach that involves decomposing the MILP model. We first solve the location-allocation problem for the warehouse-retailer phase, and using the solutions from this phase, we then solve the bottling station-warehouse phase of the problem

Finally, we discuss the outputs for different scenarios. We compare the results for the two approaches used for locating the retailers since the number of retailers affects the overall design of the supply chain system. In addition, we compare the outputs of the models solved using different dissimilarity measures used in the agglomerative hierarchical clustering based-heuristic approach. We also analyze the results obtained from various demand scenarios, considering the increasing trend of LPG demand in Rwanda. Lastly, we conduct a sensitivity analysis on the number of infrastructural units and discuss the results of the decomposed model.

Overall, this thesis is organized as follows. In Chapter 2, we present the relevant literature for clean cooking strategies, location-allocation problems within the scope of the LPG supply chain system, and location-allocation problems with large-scale data. In Chapter 3, we provide a data analysis and explain the agglomerative hierarchical clustering based-heuristic approach. In Chapter 4, we introduce the parameters used in the mathematical model and the assumptions made to simplify the problem. In Chapter 5, we provide our MILP problem formulation and model decomposition. In Chapter 6, we present the overall results. Finally, in Chapter 7, we conclude this thesis, including suggestions for future work.

Chapter 2

Literature Review

In this chapter, we present recent works related to the supply chain optimization of various clean cooking strategies. We also discuss the literature that focuses on supply chain problems within the context of LPG.

The infrastructural plan of clean cooking methods can be managed efficiently via supply chain optimization of the products. There are various studies within supply chain optimization in the literature that mainly consider biodiesel and biofuel systems, which are among the resources classified as clean cooking strategies. Akgul et al. investigated the trade-off between the economic and environmental performance of the biofuel supply chain. They formulated a static multi-objective, mixed integer linear programming model using ϵ-constraint method, where they minimized the total system cost and environmental impact [23]. Leduc et al. focused on developing an MILP for deciding on the optimal number and location of the biodiesel plants while minimizing the total cost of the supply chain system in terms of biomass production, and transportation costs [24]. Moreover, Awudu and Zhang suggested a stochastic production planning model for biofuel supply chain optimization regarding the fluctuations in demand and price of end products. With this model, they decided on the amount of raw material to purchase and consume, as well as the level of production, while maximizing the expected profit [25]. From another perspective, Zhang and Jiang worked on the waste cooking oil-for-biodiesel conversion supply chain optimization. Decisions on the number, sizes and locations of the bio-refineries, the amount of oil collected from each site, and the transportation plan of waste cooking oil and biodiesel are made in order to construct the system optimally. Regarding the uncertainty in biodiesel fuel prices, they used a robust model [26].

Other than biodiesel and biofuel, LPG is considered another clean cooking strategy that can be promoted. Various studies address the LPG supply chain optimization problem in the literature. Van Roy solved a multi-level production and distribution network optimization problem for LPG products to minimize the cost of the system. The location of the facilities, the production and stock levels in the facilities, the transportation volume between the facilities, the number of vehicles necessary to operate the system, and the customer assignments are determined, which are the set of decisions included in our problem, too. Additionally, he worked on the transportation shift systems and schedules. He also worked on the problem by introducing different scenarios such as seasonal variations, varying levels of of change in facility location and customer allocation, and changes in the size of the fleet and shift system. His system consists of 2 entities, 2 refineries, 10 candidate bottling plant locations, 40 potential depot location, and 200 customers. This is one of the papers that is closely in parallel with this thesis regarding the structure of the supply chain system, the objective function of the problem, and the set of decisions to be made [27]. However, this problem's scale low compared to ours, as the number of entities, potential locations, and customers is significantly smaller, and the customer locations are pre-determined.

Fölz et al. worked on the distribution of gas cylinders by developing a single-echelon location-allocation problem that is composed of two stages. In the first stage, they determined the location of the facilities and the assignment of demand points to the facilities in a way that the total delivery cost and operational cost of facilities are minimized. In the second stage, the daily routing plans of the vehicles travelling from the facilities to the demand points are found while maximizing a utility function based on the customer priority and minimizing the total routing transportation cost. Their system contains 7 filling stations and 1700 sales points [28].

Masudin modelled the multi-echelon LPG distribution system of Indonesia by opening new facilities considering the already existing ones. He formulated a capacitated fixed-charged facility location-allocation problem to find the optimum location and allocation of the facilities and to examine the relation between inventory costs, transportation cost, and location selections. He also conducted a sensitivity analysis to show the trade-off between facility locations and transportation and inventory costs. He is using a dataset constructed from the Indonesian LPG distribution system with 2 warehouses, 52 filling stations, 52 agents. In total, his formulation has 5,566 variables with 263 constraints [29].

Sankaran and Raghavan investigated the LPG supply chain system in India with the aim of minimizing the total cost of the system. They decided on the location of the facilities and their capacities by formulating a two-echelon facility location problem [30].

Starting with the motivation of energy conversion from kerosene to LPG in Indonesia, Santoso et al. worked on a distribution network design of LPG cylinders. The structure of the network is similar to ours as they considered a hierarchical system, with a difference of the sales points that can only be served from a single distributor while a distributor can serve multiple sales points. In addition, they assumed that each distributor has a specific number of vehicles with a predetermined number of empty cylinders that need to be filled in the filling station. Under these assumptions, they decided on the allocation of LPG cylinders between facilities while minimizing the total system cost. Their system consists of 2 filling stations, 4 distributors, and 77 retailers [31].

Blanc et al. investigated the reverse logistics network design system of LPG tanks in the Netherlands. They formulated a mixed integer model for the vehicle routing problem to optimize the number of LPG depots and allocation of end-of-life vehicle-dismantlers to depots. The objective of the study is to minimize the total system cost. In order to capture the uncertainty in the system, they conducted a sensitivity analysis for elements such as projected costs and collection periods. They worked with 29 candidate locations for depots and distances and driving times of vehicles provided by an outer company [32].

Although the hierarchical location-allocation problem is a well-known problem in the literature, discussing the problem with a data-driven approach is an area that has not yet been extensively studied. There are several papers that address supply chain optimization problems using large-scale dataset in different contexts. Gumte et al. [33], Sarkar et al. [34], Yang et al. [35], Yücel et al. [36], and Panahi et al. [37] utilized large-scale datasets while working on supply chain optimization problems in various contexts from electric vehicle charging station location problem to patient allocation problem. However, to the best of our knowledge, supply chain optimization problems have not been studied specifically for LPG systems using large-scale datasets. Hence, this thesis aims to fill that gap.

Chapter 3

Data and System Description

This thesis investigates the LPG supply chain system in Rwanda to propose a decision-making framework. The flow chart illustrating the work packages in this thesis can be found in Figure 3.1. Therefore, in this chapter, insights on Rwanda, the structure of the supply chain system, analysis on the data that we use, retailer location selection methods, and the distance calculations will be provided.

3.1 Location Specifications and Dataset

Rwanda is a landlocked country in East Africa, with a population of over 13.7 million as of 2023. It is bordered by the Democratic Republic of Congo, Tanzania, Uganda, and Burundi. The country features a hilly terrain, predominantly composed of mountainous areas and swamps [38]. It consists of 5 provinces, namely Northern Province, Southern Province, Western Province, Eastern Province, and the City of Kigali. Each province is further subdivided into districts, with a total of 30 districts in the country. Similarly, each district is divided into 14,185 villages.

The population density varies significantly across the country. Since the City

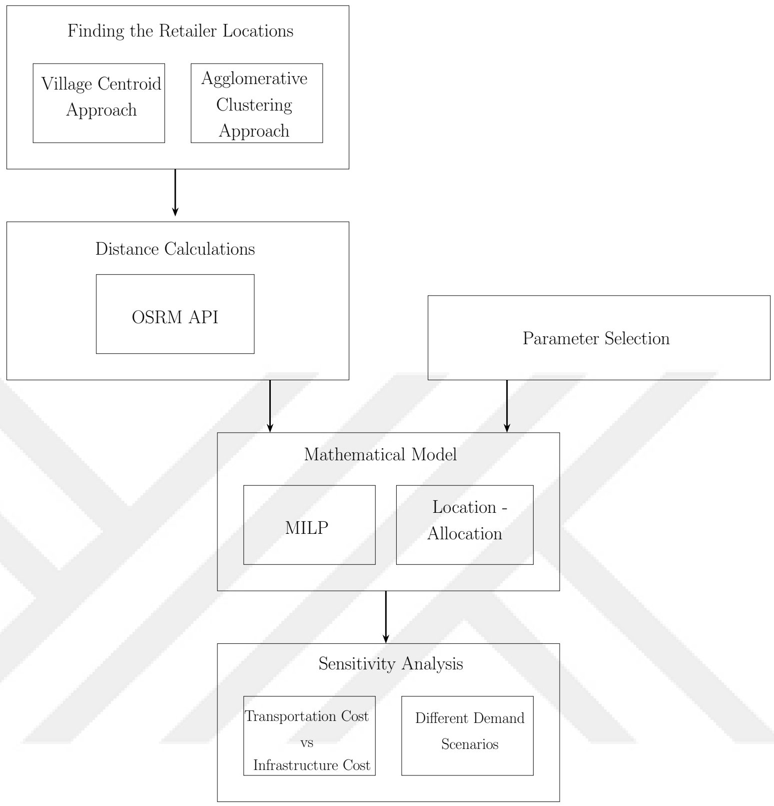


Figure 3.1: Work Packages

of Kigali is the capital and largest city of Rwanda, it contains densely populated districts. On the contrary, the population density of other provinces is substantially lower than the City of Kigali. This implies that rural settlements and agricultural activities are more common in these areas [38]. This disparity in population densities indicates that logistics dynamics may differ between urban and rural areas, with distinct challenges and opportunities in each region.

We have data files containing information about districts, villages and rooftops within these villages. Geospatial datafiles about district and village polygons were obtained from [39]. In addition, we curated building footprint information from [40] and [41] to obtain a complete database including 3,341,549 building footprints in Rwanda. These data files are integrated into QGIS, a powerful geographic information system tool where we can visualize the precise locations of each rooftop across the country. Additionally, we can see the distribution of rooftops within each district, enabling us to observe the spatial patterns.

3.2 System Description

LPG supply chain network consists of several stages, starting from importation to getting the products to the final demand point. The stages that are examined in this thesis are importation, storage in the depot, bottling operations and storage in the bottling stations, storage in the warehouses, and transportation of the products between the facilities.

The system starts with importing LPG into the country. Since Rwanda does not have any oil refineries, it has to import LPG overseas from Dar es Salaam (Tanzania) and Mombasa (Kenya). In Rwanda, transportation infrastructure primarily builds upon road transportation since Rwanda lacks connections to regional railways and Inland Water Transport [42]. Consequently, according to the Rwanda National LPG Master Plan, the LPG marketers in Rwanda would rather import from Tanzania with their own bulk trucks, considering logistic operations and regulatory requirements [22].

The imported LPG is stored in a LPG depot. The depot acts as a hub storing large bulk quantities of LPG. From the depot, LPG is periodically transported to the bottling station using tankers. Tankers, with their large capacities, carry the LPG in bulk form. The transfer of LPG from the depot to the bottling stations is referred to as bulk transportation.

Bottling stations are one of the major components of the LPG supply chain system. After the tankers carry LPG to the bottling stations, at these facilities, the cylinders are filled with LPG. In addition, bottling stations have the capacity to store the filled cylinders and bulk LPG quantities for a period of time. It should be noted that cylinders are considered the end product distributed to the end customers in this thesis. Although there may be customers requiring LPG in bulk form, such as hospitals, in this thesis, we solely focus on the customers who demand LPG in cylinders for the sake of simplification.

After bottling operations in the bottling stations, trucks carry the cylinders to the warehouses, which are named as primary transportation. At this point, the warehouses function as temporary storage units. As the final step, vans transfer cylinders to the retailers, which are considered to be the final demand point. In this thesis, we assume that each household satisfies its LPG cylinder demand from the retailer that is allocated to serve that specific household. Additionally, as we do not consider the last mile logistic part of the system, individuals in households have to reach their designated retailer on their own. An illustration of the described system can be seen in Figure 3.2.

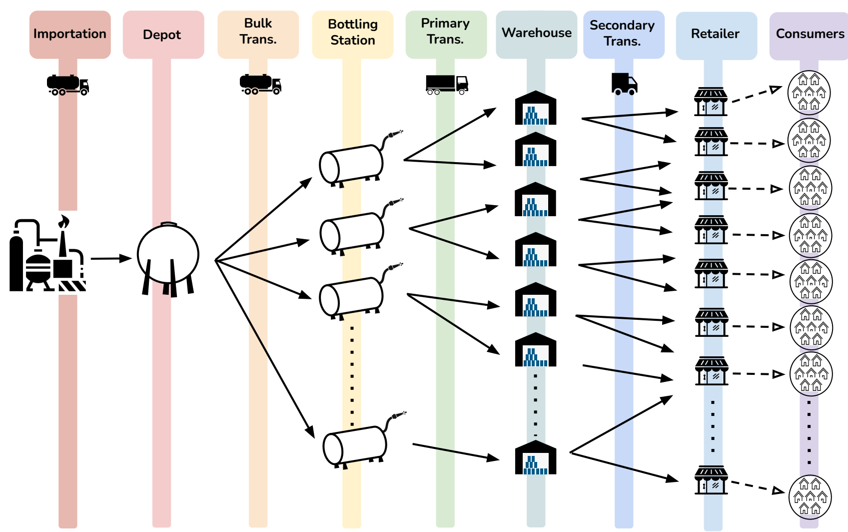


Figure 3.2: LPG supply chain system illustration

3.3 Finding the Locations of the Retailers

In our problem, the individual rooftops are not considered as the final demand point in the system. Instead, we assume that each rooftop satisfies its LPG cylinder demand from the assigned retailer and retailers are the final demand point in the system. Thus, we neglected the last mile logistics. At this point, selecting the locations of the retailers in the system wisely is crucial for our problem.

Generally, most of the studies in the literature work with pre-determined final demand locations within the scope of the location-allocation problem. However, in this thesis, we also determine the locations of retailer points (final demand point in the system). In order to decide on the location of the retailers, we follow two different methods: 1) finding each village's centroid, and 2) using a heuristic based on agglomerative hierarchical clustering algorithm.

3.3.1 Finding Village Centroids

In this approach, we locate a retailer at each village's centroid. We use QGIS to find the centroid of each village. The centroid is calculated in such a way that the centroid minimizes the sum of squared Euclidean distances between itself and each vertex point of the polygon. Then, the coordinates of each centroids are extracted to be used as the retailer locations. Consequently, the total number of retailers is equal to the number of villages, which is 14,185. A sample depiction of the villages with their centroids can be seen in Figure 3.3.

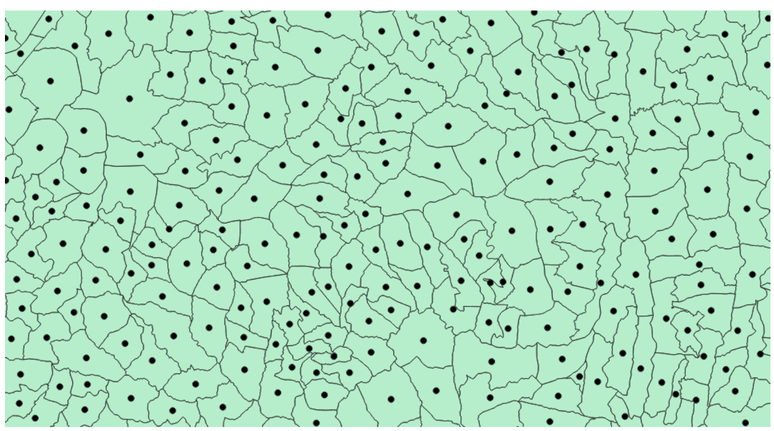


Figure 3.3: Sample village centroids

However, it should be noted that locating the retailers in the centroids of the villages may not be the most fair approach. This is due to the significant variation in the surface areas of the villages across the country. Some villages cover a considerably larger area than others. In particular, it is observed that the villages in the center of the country generally have a smaller surface area, while those near the borders of the country are spread over a larger region. Consequently, the distance between a rooftop and its retailer may not be similar for all rooftops.

For rooftops located near the borders of the villages that have a larger surface area, the distance travelled to reach the centroid of the village is greater compared to rooftops around the borders of villages with a small surface area. As illustrated in Figure 3.4, the distance travelled in the village colored in blue is less than the distance travelled in the village colored in yellow due to the varying surface areas.

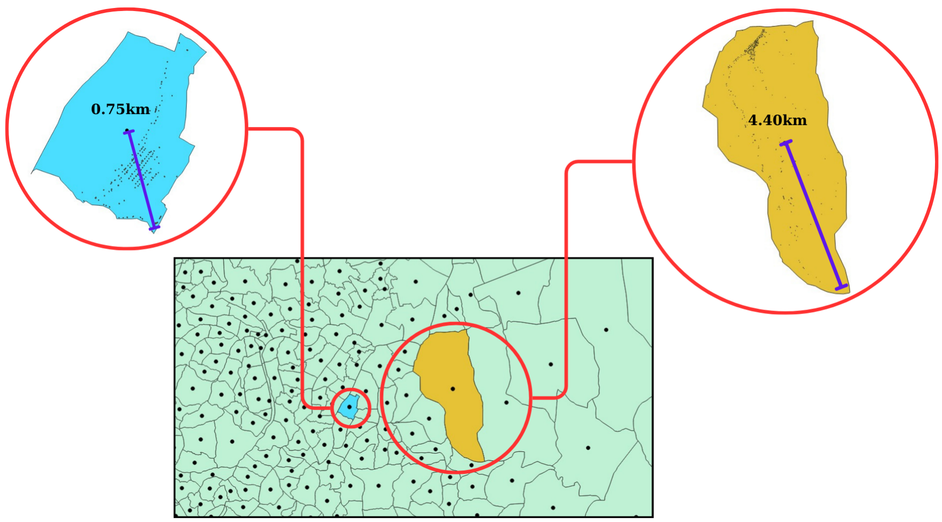


Figure 3.4: Villages with small and large surface area

Hence, in order to ensure fairness among rooftops, we introduce agglomerative hierarchical clustering-based heuristic approach to decide on the locations of the retailers.

3.3.2 Agglomerative Hierarchical Clustering-Based Heuristic Approach

The other method used to find the location of retailers is based on the agglomerative hierarchical clustering algorithm. Agglomerative hierarchical clustering is a widely used clustering approach where the nodes are clustered according to their proximity [43]. It is a bottom-up approach where each point acts as a cluster on its own at the beginning and they are successively merged into larger clusters according to a dissimilarity measure. The algorithm continues to merge clusters until all points are connected and there is one single cluster formed if there is no stopping criterion.

As a pre-processing step, before running the agglomerative hierarchical clustering-based heuristic, we divide each district into grids of 5 kilometers due to the sheer volume of data, i.e., 3,341,549 rooftops. Due to the variety in population density over the country, some districts contain considerably higher numbers of rooftops. In this case, to avoid memory issues, we divide those districts into grids of 2 kilometers. More specifically, Gasabo, Kicukiro, Musanze, Nyarugenge, and Rubavu are the districts that are divided into grids of 2 kilometers. A sample illustration of a district with its grid lines can be seen in Figure 3.5.

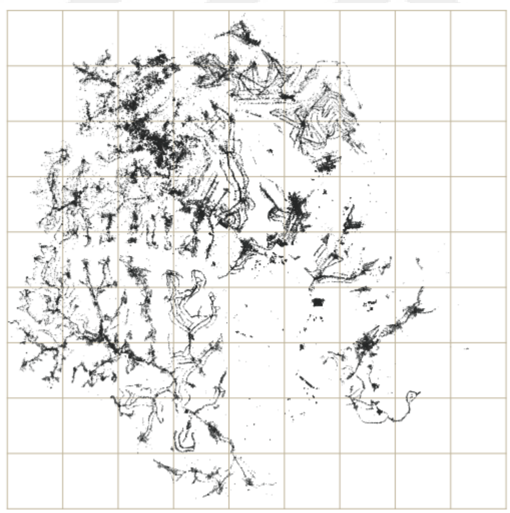


Figure 3.5: Bugesera with grid lines (5km)

To find the locations of the retailers, we use the agglomerative clustering-based heuristic approach proposed by [44]. In order to do that, for the agglomeration process of the rooftops into clusters, we utilize a distance threshold as the stopping criterion of the clustering algorithm. The aggregation continues step-by-step until the most similar clusters cannot be merged into a cluster due to the distance limit. Distance limit ($limDist$) is determined as 1 mile in this thesis in order to ensure each rooftop is within a walkable distance to its allocated retailer. The distance limit prevents the formation of clusters from sparsely distributed rooftops and enables us to make sure that the maximum distance between a rooftop and a retailer can be set at 1 mile in the following steps.

At each iteration, the pair of clusters that has the least dissimilarity measure are selected and merged. As dissimilarity measures used in the agglomerative clustering algorithm, we use commonly preferred measures in the literature, namely, Complete [45], Average [46], Centroid, and Ward's minimum variance [47]. In this thesis, these dissimilarity measures are referred to as linkage parameters used in the agglomerative hierarchical clustering algorithm. Since each linkage parameter performs differently and their performances vary depending on the dataset used [48], we preferred to try each of them and select the one that yield the best result for our problem to be used in future analysis.

In order to explain the linkage methods, let us define j and g as two clusters and N_j and N_g as the set of nodes in clusters j and g respectively. Also, let (x_i, y_i) be a point in cluster j ($i \in N_j$) and (x_k, y_k) be a point in cluster g ($k \in N_g$).

First, Johnson specified that Complete linkage method calculates the dissimilarity between two clusters, j and g, in a way that it will yield the maximum distance between rooftop $i \in N_j$ and $k \in N_g$ [45] as in Equation (3.1):

$$d(j, g) = \max_{i \in N_j,\, k \in N_g} \|(x_i, y_i) - (x_k, y_k)\| \tag{3.1}$$

Sneath and Sokal presented the Average linkage method, and they named it as unweighted pair group method with arithmetic mean. The Average linkage method determines the dissimilarity between two clusters, j and g, as the average of the pairwise Euclidean distance between each rooftop in these clusters [46]. It is represented in Equation (3.2):

$$d(j, g) = \frac{1}{|N_j||N_g|} \sum_{i \in N_j} \sum_{k \in N_g} \|(x_i, y_i) - (x_k, y_k)\| \tag{3.2}$$

In Centroid linkage method, the distance between two clusters are computed using Equation (3.3). In this method, the centroid point of each cluster is determined and the distance between the centroid points are specified as the dissimilarity measure [48]:

$$d(j,g) = \left\| \frac{1}{|N_j|} \sum_{i \in N_j} (x_i, y_i) - \frac{1}{|N_g|} \sum_{k \in N_g} (x_k, y_k) \right\| \tag{3.3}$$

As distinct from the above-explained linkage methods, in Ward linkage method, Ward considered the information loss associated with each agglomeration. He quantified the amount of loss due to agglomeration using Error Sum of Squares (ESS) criterion and he proposed to merge clusters yielding the minimum increase in information loss, in other words with minimum ESS value [47]. Hence, in Ward linkage method, the dissimilarity is associated with the increase in the total within-cluster variance as a result of agglomeration. As expected, in this method, the goal is to merge the pair of clusters that results in the smallest possible increase in the total within-cluster variance. Consequently, we compute the Ward dissimilarity measure as the sum of squared distances between the centroid of each cluster and each rooftop in the cluster using Equation (3.4):

$$d(j,g) = \sqrt{\frac{2|N_j||N_g|}{|N_j| + |N_g|}} \left\| \frac{1}{|N_j|} \sum_{i \in N_j} (x_i, y_i) - \frac{1}{|N_g|} \sum_{k \in N_g} (x_k, y_k) \right\| \tag{3.4}$$

As the clusters are merged based on the above-explained linkage methods, in each agglomeration, the distance-constrained geometric median point (geomedian) of each cluster is determined. If the distance between the geomedian point and the rooftops in the cluster is less than the predetermined distance limit, $limDist$, then the algorithm merges this pair of clusters and stores the cluster configuration and total cost of the related configuration to obtain the cluster system yielding the minimum cost. Otherwise, the agglomeration does not occur.

Geomedian point of a cluster is defined as the point that minimizes the sum of distances to the nodes in the cluster. It should be noted that the geomedian point of a cluster does not necessarily need to be one of the nodes in the cluster, but it can be any point in the planar space. Let us follow the notation: Given that L denotes the set of all clusters, N_j indicates the set of all nodes in cluster $j \in L$. The nodes in cluster j is represented as (x_i, y_i) where $i \in N_j$ and $j \in L$.

Moreover, (z_t^j, w_t^j) denotes the geomedian point found in iteration t for cluster $j \in L$ and (z_a^j, w_a^j) indicates the newly found suggested geomedian point.

Until now, there is no direct formulation to calculate the exact location of the geomedian point in the literature as Weber showed [49]. Hence, iterative algorithms are preferred to find the geomedian point. A well-known method to find the distance-constrained geometric median point is using the Weiszfeld algorithm. The Weiszfeld algorithm converges to geomedian point if the suggested point is not one of the nodes in the cluster.

The Weiszfeld algorithm starts with an initial guess for the geomedian point, which can be any point in the Euclidean space. Then, in each iteration, it computes the next approximation of geomedian point using the following formula:

$$(z_a^j, w_a^j) = \frac{\sum_{i \in N_j} \frac{(x_i, y_i)}{\|(x_i, y_i) - (z_t^j, w_t^j)\|}}{\sum_{i \in N_j} \frac{(x_i, y_i)}{\|(x_i, y_i) - (z_t^j, w_t^j)\|}} \quad , \tag{3.5}$$

where $\|(x_i, y_i) - (z_t^j, w_t^j)\|$ is the Euclidean distance between (x_i, y_i) and (z_t^j, w_t^j) [37].

The algorithm iteratively finds new approximations of the geomedian point until the change between successive approximations is below a certain threshold, indicating convergence. When it converges, the Weiszfeld algorithm finds the geomedian point. However, when the geomedian point found in an iteration is computed to be the exact location of one of the nodes in the cluster, the denominator yields zero in the formula. In order to prevent this, in this thesis, the distance-constrained geometric median points are found using Vardi and Zhang's modified Weiszfeld algorithm [50].

The modified Weiszfeld algorithm iteratively recommends a suggested geomedian point until no further changes are made in the suggested point. At the beginning of the algorithm, as an initialization step, we select the suggested geomedian point as the center point of the smallest circle (z_0^j, w_0^j), which includes all the nodes in cluster $j \in L$. If the radius of this smallest circle exceeds the defined distance limit, $limDist$, the algorithm terminates. Because, in this case, it is not possible to set the maximum distance between a rooftop and a retailer to 1 mile. On the other hand, if the radius of the smallest circle is less than $limDist$ and the geomedian point of cluster $j \in L$ at iteration t (z_t^j, w_t^j) is not located exactly on the same location of a rooftop (x_i, y_i) of cluster $j \in L$, where $i \in N_j$, then the location of the new suggested geomedian point is calculated using the following formula:

$$(z_a^j, w_a^j) = \frac{\sum\limits_{i \in N_j, d_i^t > 0} (x_i, y_i) \frac{1}{\|(x_i, y_i) - (z_t^j, w_t^j)\|}}{\sum\limits_{i \in N_j, d_i^t > 0} \frac{1}{\|(x_i, y_i) - (z_t^j, w_t^j)\|}} = \frac{\sum\limits_{i \in N_j, d_i^t > 0} (x_i, y_i) \frac{1}{d_i^t}}{\sum\limits_{i \in N_j, d_i^t > 0} \frac{1}{d_i^t}} \quad , \tag{3.6}$$

where d_i^t denotes the distance between rooftop $i \in N_j$, $j \in L$ and the geomedian location found in iteration t (z_t^j, w_t^j).

If (z_t^j, w_t^j) coincides with one of the rooftop locations (x_i, y_i) where $i \in N_j$ and $j \in L$, then the total attraction $\sigma(z_t^j, w_t^j)$ between (z_t^j, w_t^j) and the rest of the rooftops in N_j and the resistance to movement η are computed. The attraction indicator σ is calculated using the following formula:

$$\sigma(z_t^j, w_t^j) = \left\| \sum_{i \in N_j, d_i^t > 0} \frac{(x_i, y_i) - (z_t^j, w_t^j)}{\|(x_i, y_i) - (z_t^j, w_t^j)\|} \right\| \tag{3.7}$$

Next, the resistance of movement is specified in the following manner:

$$\eta(z_t^j, w_t^j) = \begin{cases} \sigma & \text{if } (z_t^j, w_t^j) \in N_j \\ 0 & \text{otherwise} \end{cases} \tag{3.8}$$

The attraction indicator σ and the resistance of movement η are identified in order to decide whether relocating the suggested geomedian point will create a significant impact. In case the attraction indicator σ is smaller than the resistance to movement η, the current location of the suggested geomedian point is not altered. On the other hand, if the attraction indicator σ is higher than the resistance to movement η, the new location of the suggested geomedian point is determined as follows:

$$(z_a^j, w_a^j) = \left(1 - \frac{\eta(z_t^j, w_t^j)}{\sigma(z_t^j, w_t^j)}\right)(z_a^j, w_a^j) + \frac{\eta(z_t^j, w_t^j)}{\sigma(z_t^j, w_t^j)}(z_t^j, w_t^j) \tag{3.9}$$

Note that a distance limit is not regarded for convergence in the modified Weiszfeld algorithm. Consequently, the distance between the determined geomedian point and any rooftop in the cluster may be more than *limDist*, which is 1 mile. Hence, in order to ensure fairness among rooftops, after the algorithm converges, we shift the determined geomedian point to a feasible point. In this context, a feasible point refers to the location where the geomedian point is within 1-mile distance from each rooftop in the cluster. In order to shift the suggested geomedian point to a feasible point, [44] utilize a projection method developed by [51].

According to the method proposed by [51], if the suggested geomedian point is within the distance limit, then the suggested geomedian point (z_a^j, w_a^j) is set as the geomedian location for the next iteration. Hence, $(z_{t+1}^j, w_{t+1}^j) = (z_a^j, w_a^j)$. In Figure 3.6, the feasible region can be seen as the shaded area. The feasible region is depicted as the intersection of circles, each with a center of a rooftop and a radius of *limDist*. Hence, in order to ensure that a suggested geomedian point is within the *limDist* distance to each rooftop in the cluster, the suggested geomedian point should fall to the shaded region.

On the other hand, if the suggested geomedian point does not obey the distance limit and falls to a region outside the feasible area, then a line segment between the suggested geomedian point (z_a^j, w_a^j) and the current geomedian point (z_t^j, w_t^j) is created. Suppose set D represents the set of rooftops (m_r, n_r) that are more than *limDist* to the suggested geomedian point (z_a^j, w_a^j). For each rooftop $(m_r, n_r) \in$

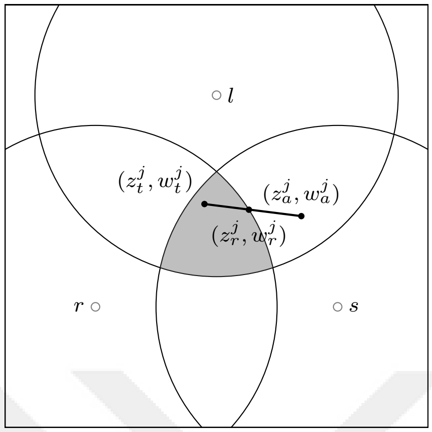


Figure 3.6: Depiction of (z_r^j, w_r^j)

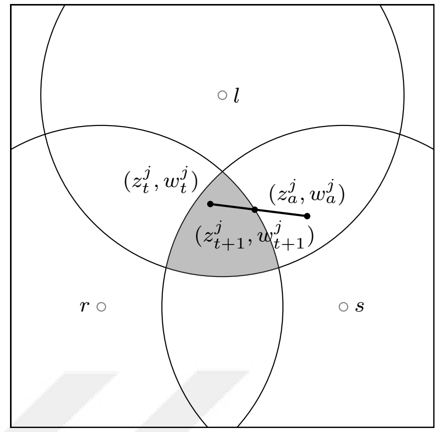


Figure 3.7: Projection to find (z_{t+1}^j, w_{t+1}^j)

D, a point that is *limDist* away and on the line segment between (z_a^j, w_a^j) and (z_t^j, w_t^j) is determined. In Figure 3.6, one of these points are labeled as (z_r^j, w_r^j) for rooftop r.

Each point on line segment between points (z_a^j, w_a^j) and (z_t^j, w_t^j) can be expressed by the following equation while $\theta \in [0, 1]$:

$$(z_t^j, w_t^j) + \theta(z_a^j - z_t^j, w_a^j - w_t^j) \tag{3.10}$$

With this equation, we are able to find a feasible point (z_r^j, w_r^j) on the line segment, that is precisely *limDist* away from the rooftop (m_r, n_r). This feasible point is represented as follows:

$$(z_r^j, w_r^j) = (z_t^j, w_t^j) + \theta_r(z_a^j - z_t^j, w_a^j - w_t^j) \tag{3.11}$$

Hence, in order to find the associated $\theta_r \in [0, 1]$ for the rooftop (z_r^j, w_r^j), we solve the following second order polynomial equation for all rooftops $r \in D$:

$$(z_t^j + \theta_r(z_a^j - z_t^j) - m_r)^2 + (w_t^j + \theta_r(w_a^j - w_t^j) - n_r)^2 = limDist^2 \tag{3.12}$$

For rooftops that are at most $limDist$ away from (z_a^j, w_a^j), i.e. $N_j \setminus D$, all points on the line segment are within $limDist$ threshold. On the contrary, for rooftops $r \in D$, the points on the line segment with $\theta \in [0, \theta_r]$ are the only ones that are within $limDist$ threshold. Consequently, we select $\theta_{min} = \min_{r \in D} \theta_r$. Then, we project the suggested infeasible geomedian point (z_a^j, w_a^j) to the feasible region for the next iteration using the following equation:

$$(z_{t+1}^j, w_{t+1}^j) = (z_t^j, w_t^j) + \theta_{min}(z_a^j - z_t^j, w_a^j - w_t^j) \tag{3.13}$$

Consequently, (z_{t+1}^j, w_{t+1}^j) corresponds to the intersection point of the line segment and the boundary of the feasible region as it can be seen in Figure 3.7. This projected location,(z_{t+1}^j, w_{t+1}^j), is named as distance-constrained geometric median point of cluster j. As this projection is performed in every iteration, it is ensured that the feasibility is sustained.

In the end, the configuration with the least cost is given as the final output with retailers' locations. Overall, at the end of this algorithm, we find retailer locations as many as the number of clusters formed by the agglomerative hierarchical clustering-based heuristic, which are located at the distance-constrained geometric median point of each cluster.

3.4 Locations of Facilities and Distance Calculations

We determine the locations of the candidate bottling stations considering the insights from the Rwanda National LPG Master Plan. Bottling stations should be situated in a way that they are as close as possible to the sales areas without compromising any risk to the public. Hence, the strategy while choosing the locations of the bottling stations is minimizing filling cost while maximizing the safety level [22]. Regarding these, Rwanda's LPG Master Plan suggests 7 locations for the bottling stations. The candidate bottling station locations can be found in Appendix A.

Rwanda's LPG Master Plan asserts that Jabana is known for a fuel depot location. Thus, the depot is located in Jabana, specifically at (-1.8735, 30.0811) [22]. In addition, candidate warehouse locations are selected to be in the center of each district. Hence, there are 30 candidate warehouse locations in total. Candidate warehouse locations are given in Appendix B.

As the locations of the depot, candidate bottling stations, candidate warehouses, and retailers are determined, we proceed with calculating the distance between the facilities. To calculate the distance between two coordinates, we use the Open Source Routing Machine (OSRM) routing engine. It uses OpenStreetMap (OSM) data to find the shortest path between two coordinates for various types of transportation modes [52]. In this thesis, we use "route" as the service parameter to find the fastest route between coordinates, "v1" as the version parameter, and "driving" as the profile parameter to specify the type of transportation. A depiction of the driving route is provided in Figure 3.8. Using OSRM, we calculate the distance between the depot and each bottling station, the distance between every bottling station and every warehouse, and the distance between every warehouse and retailer. These distances will later be used in the mathematical model.

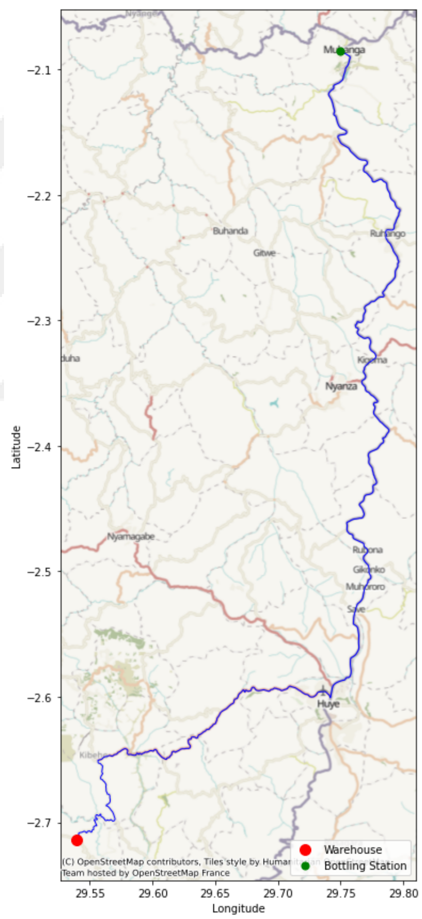


Figure 3.8: Driving route between bottling station 5 and warehouse 29

Chapter 4

Parameter Selection and Assumptions

4.1 Parameter Selection

Parameters used in the mathematical model are selected from the Rwanda National LPG Master Plan [22] and [38]. The values selected for the parameters used in the mathematical model can be found in Table 4.1, Table 4.2, and Table 4.3. The values of discount rate, speed of vehicles, labor cost, annual salary, and daily working hours are taken as assumptions. While the weight of the cylinder is taken from the Rwanda National LPG Master Plan [22], the rest of the parameter values are retrieved from [38]. The necessary currency exchanges are conducted according to the current values as of July 2, 2024, with the exchange rate of one USD ($) being equal to 0.93 Euros (€), and one Rwandan franc (RWF) being equal to 0.00076 USD ($).

4.1.1 Importation Parameters

A price is incurred per kilogram for importing LPG. Currently, Rwanda imports LPG overseas from Dar es Salaam (Tanzania) and Mombasa (Kenya), with importation costs of approximately 765$ per metric ton from Kenya and 807$ per metric ton from Tanzania. Hence, the estimated importation price is taken as an average of these two costs, 786$ per metric ton [38].

The importation license is a mandatory regulatory requirement to import LPG into the country, which secures safe and efficient importation. It should be renewed every 5 years. The importation license is taken as 2,600,000 RWF [38].

4.1.2 Transportation Parameters

Diesel fuel is used in vans, trucks, and tankers. The diesel fuel price is calculated based on the current diesel price in Rwanda and the diesel price cap determined by the Rwandan government. As of July 2, 2024, the current diesel price and the diesel cap price are 1652 RWF per liter [53] - [54].

The fuel consumption of a standard vehicle used in bulk transportation is estimated to be 34.5 liters per 100 kilometers. Accordingly, the cost incurred per kilometer due to the fuel consumption of tankers is found to be 0.435$ per kilometer. The fuel consumption of a standard vehicle utilized in primary transportation is approximated to be 27 liters per 100 kilometers. Regarding this, 0.340$ is incurred per kilometer due to the fuel consumption of the trucks. Lastly, the fuel consumption of a standard vehicle used in secondary transportation is estimated to be 19 liters per 100 kilometers. Hence, the cost incurred per kilometer due to the fuel consumption of vans is taken as 0.239$ per kilometer [54].

A license is issued for each type of transportation by The Rwanda Utilities Regulatory Authority (RURA) to ensure the safe and efficient transportation of the products. The license cost is 550,000 RWF for each type of transportation. It should be noted that the license is valid for 5 years for all three types of transportation [38].

The Overnight costs (ON) of vehicles are referred to as the cost of acquisition in the model. It is estimated that the costs of acquisition of a tanker, truck, and van are estimated to be 183,835.59€, 170,200\$, and 53,771.28€, respectively. In addition, each vehicle should be renewed when their lifetimes are completed. The lifetimes of tankers, trucks, and vans are specified as 35 years, 15 years, and 10 years respectively [38].

The Variable operation and maintenance (VOM) costs of vehicles are accounted for per kilometer vehicles travel in the model. This includes the cost of tires, maintenance, and repair costs. Additionally, the cost of fuel consumption is also added to the VOM costs of vehicles. It is estimated that the VOM costs of tankers, trucks, and vans are 0.186€/km, 0.186€/km, and 0.184€/km respectively [38].

In order to estimate the fixed operation and maintenance costs of vehicles (FOM) transportation prices of Africa's international corridors are taken into consideration. It is stated that the fixed costs compose 56% of the yearly total costs while the variable costs account for the remaining 44% of the yearly total costs. Thus, the fixed costs are calculated as 127% of the total variable costs in the model [38].

Vehicle capacities are regarded in the model. The common range of tanker capacities is between 20 to 28 metric tons. Hence, for simplification, the capacity of the tankers is taken as 24 metric tons. It is stated that the trucks can carry a maximum of 833 12kg-equivalent cylinders. Thus, the capacity of the trucks is set to be 9996kg. Although the Rwanda National LPG Master Plan indicates that there are different types of vehicles with various capacities for secondary transportation, in this thesis, we focus on only one type of secondary transportation vehicle, which is a van. The capacity of a van is estimated to be 270 12kg-equivalent cylinders as a mean capacity value of different vehicles functioning for secondary transportation. Therefore, the capacity of a van is determined to be 3240kg [38].

4.1.3 Infrastructural Parameters

A license, known as the License for LPG Plant or Bulk Storage Facility Installation, is issued by RURA to make sure that LPG is stored and handled in the depot carefully. The license fee is specified as 1,000,000 RWF. Another license, named the License for LPG Plant or Bulk Storage Facility Operations, is a necessary regulation that ensures safe operation and management activities in bulk facilities. The license cost is 5,200,000 RWF, and it is valid for 3 years. Additionally, the lifetime of the depot is indicated as 50 years [38].

As in the LPG depot, installation and operation licenses are issued by RURA. The License for LPG Plant Installation costs 1,000,000 RWF. The cost of License for LPG Plant Operations is 5,200,000 RWF. This license is also valid for 3 years [38].

The overnight cost (ON) of a bottling station refers to the expenses related to the opening of a bottling station. More specifically, the ON includes expenses related to site preparation, construction, and acquisition of equipment necessary for bottling, handling, and storing LPG. The ON cost of a bottling station, with a capacity of 35,000 tons per year, is estimated to be 7,000,000$ [38].

The fixed cost of operation and maintenance (FOM) of a bottling station includes elements such as staff cost, facility maintenance cost, utilities cost, insurance, and administrative expenses. Overall, the FOM of a bottling station, of the above capacity, is approximated to 384,400$ [38].

The variable cost of operation and maintenance (VOM) of a bottling station is incurred for each kilogram of LPG bottled. The variable cost is approximated to be 35$ per ton. Furthermore, the approximate maximum bottling capacity of an individual bottling station is determined to be 35,000 tons per year. It should be taken into account that the lifetime of a bottling station is 15 years, according to industry standards [38].

An LPG wholesale license is required for warehouse operations with a cost of

1,100,000 RWF. The license is valid for 5 years [38].

The overnight cost (ON) of warehouses includes expenses to open a warehouse, such as the acquisition of land, properties, and equipment to operate. It is estimated that the ON of a warehouse is 71,105,252.32 RWF [38].

Expenses such as staff costs, rent, utilities, insurance, and maintenance costs constitute the fixed cost (FOM) for warehouses. It is approximated that the FOM cost of a warehouse is 154,800$ [38]. In addition, following [38], the Variable cost (VOM) of warehouses is assumed to be 0$.

4.1.4 Cylinder Parameters

Although in the Rwanda National LPG Master Plan, different cylinder sizes are considered (6kg, 12kg, 35kg), since 12kg cylinders are the predominant choice in the market, this thesis only considers the 12kg cylinder demands [22]. In other words, the total LPG demand, which is available in kilogram form, is distributed to 12kg cylinders, and our MILP model is constructed based on this assumption. Moreover, note that the lifetime of any cylinder type is approximately 20 years [38].

In order to calculate the empty cylinder acquisition price of 12kg cylinders, the offers from different suppliers are considered by [38]. Following the lead of [38], the offers from the suppliers are recorded as 10$, 12.8$, and 13.5$. Hence, the empty cylinder acquisition cost is found to be 12.10$ as of their average value.

LPG cylinder requalification involves evaluating and confirming the safety and reliability of LPG cylinders after a specific period or usage. This process ensures the continued suitability and security of the cylinders by checking the existence of any leaks, damages, and flaws. The requalification cost is set at 7.17$ per cylinder. In addition, cylinders must be requalified every 10 years [38].

Table 4.1: Parameter values - part 1

Parameter	Abbreviation	Value
Discount Rate	r	0.1
Weight of cylinder (kg)	w	12
Capacity of a van (kg)	$CapV$	3,240
Capacity of a truck (kg)	$CapTr$	9,996
Capacity of a tanker (kg)	$CapTa$	24,000
Speed of a van (km/h)	Sv	15
Speed of a truck (km/h)	Str	30
Speed of a tanker (km/h)	Sta	20
Lifetime of a van (year)	Lv	10
Lifetime of a truck (year)	Ltr	15
Lifetime of a tanker (year)	Lta	35
Lifetime of a cylinder (year)	Lc	20
Van license validity (year)	$LicV$	5
Truck license validity (year)	$LicTr$	5
Tanker license validity (year)	$LicTa$	5
Bottling capacity of a bottling station (kg)	$BotCap$	35,000,000
Capacity of a warehouse (kg)	$Capw$	100,000,000
Bulk inventory capacity of a bottling station(kg)	$BulkCap$	100,000,000
Cylinder inventory capacity of a bottling station (cylinder)	$CylCap$	100,000,000
Fixed cost of a warehouse ($/year)	$FOMw$	154,800
Overnight cost of a warehouse ($)	ONw	54,039.992
Fixed cost of a bottling station ($/year)	$FOMb$	348,400
Overnight cost of a bottling station ($)	ONb	7,000,000
Variable cost of a bottling station ($/kg)	$VOMb$	0.035

Table 4.2: Parameter values - part 2

Parameter	Abbreviation	Value
Lifetime of a warehouse (year)	Lw	15
Lifetime of a bottling station (year)	Lbs	15
Lifetime of the depot (year)	Ld	50
Fixed cost of a van (\$/year)	$FOMv$	127% of $VOMv$
Overnight cost of a van (\$)	ONv	57,818.581
Variable cost of a van (\$/km)	$VOMv$	0.437
Fixed cost of a truck (\$/year)	$FOMtr$	127% of $VOMtr$
Overnight cost of a truck (\$)	$ONtr$	170,200
Variable cost of a truck (\$/km)	$VOMtr$	0.540
Fixed cost of a tanker (\$/year)	$FOMta$	127% of $VOMta$
Overnight cost of a tanker (\$)	$ONta$	197,672.677
Variable cost of a tanker (\$/km)	$VOMta$	0.635
Cost of a 12-kg cylinder i (\$)	$cylCost$	12.10
Labor cost (%)	$laborC$	1.18
Annual salary (\$)	$salary$	2,604
Daily working hours (h)	$dayWh$	8
Van license cost (\$)	$licenseCv$	418
Truck license cost (\$)	$licenseCtr$	418
Tanker license cost (\$)	$licenseCta$	418
Warehouse wholesale license cost (\$)	$Wsal$	836
Cylinder requalification cost (\$/cylinder)	$Creq$	7.17
Bottling station installation fee (\$)	$InsB$	760
Depot installation fee (\$)	$InsD$	760
Bottling station operating license cost (\$)	$OpLb$	3,952
Depot operating license cost (\$)	$OpLd$	3,952
Importation license cost (\$)	ImL	1,976
LPG importation price (\$/kg)	$PriceI$	0.786

Note that discount factors are used to calculate the present value of future cash flows in the system. Discount factors are calculated using the following formula:

Table 4.3: Parameter values - part 3

Parameter	Abbreviation	Value
Import license validity (year)	$LicI$	5
Warehouse wholesale license validity (year)	$LicW$	5
Bottling station operating license validity (year)	$LicBs$	3
Depot operating license validity (year)	$LicD$	3
Cylinder requalification validity (year)	$LicC$	10
Discount factor of a warehouse	dFw	0.131
Discount factor of a bottling station	dFb	0.131
Discount factor of the depot	dFd	0.101
Discount factor of a van	dFv	0.163
Discount factor of a truck	$dFtr$	0.131
Discount factor of a tanker	$dFta$	0.104
Discount factor of a cylinder	dFc	0.117
Discount factor of a van license	$dFvl$	0.264
Discount factor of a truck license	$dFtrl$	0.263
Discount factor of a tanker license	$dFtal$	0.264
Discount factor of a warehouse wholese license	$dFwsal$	0.264
Discount factor of a cylinder requalification	$dFcreq$	0.163
Discount factor of a bottling station operating license	$dFbop$	0.402
Discount factor of a depot operating license	$dFdop$	0.402
Discount factor of import license	$dFil$	0.264

$$dF = \frac{r}{1-(1+r)^{-L}} \tag{4.1}$$

where r denotes the discount rate (yearly) and L denotes the lifetime (years).

4.2 Assumptions

In order to simplify the supply chain structure and the mathematical model, we make some assumptions. While these assumptions do not alter the essence of the model, they are helpful in creating a practical setting. In this section, these assumptions are explained and justified. The assumption that we use are as follows:

- **Disregarding the geographical availability of retailer locations:** As explained before, to find the locations of the retailers, either the centroid approach or agglomerative hierarchical clustering-based heuristic approach is used. However, these methods do not consider the geographical structure of the topology. More specifically, the retailers may be situated in places that are not easily reachable by road transportation, such as backlands, mountains, bodies of water, etc. Nonetheless, this is a necessary assumption to incorporate into the problem, as integrating the details of geographical structures of the topology may greatly augment the complexity of the clustering algorithms. Considering the scale of the data, this is a crucial simplification to carry out quicker and more efficient analysis.

- **Constant demand assumption:** We only consider the yearly demand information from the Rwanda National LPG Master Plan. Hence, we assume that there is no seasonality in demand and that demand is constant throughout the year. This is one of the key steps to model problem that is independent of time index t.

  Yearly LPG demand values are taken from the Rwanda National LPG Master Plan. The projected demand values for the Business-as-usual (BAU) scenario are used. BAU scenario assumes that if the current trends in LPG usage persist, the percent of the total Rwandan population using LPG will reach 16.8% in 2024 and 30.4% in 2030 [22]. Note that these ratios include both household and institutional demand. In Table 4.4 and Figure 4.1, the yearly projected LPG demand (in tons) across all market segments is shown.

  Table 4.4: Yearly BAU LPG demand (in tons)

Scenario	**2025**	**2026**	**2027**	**2028**	**2029**	**2030**
BAU	81,935	92,069	103,481	116,248	130,379	145,770

- **Bulk and cylinder demand assumption:** There are three market

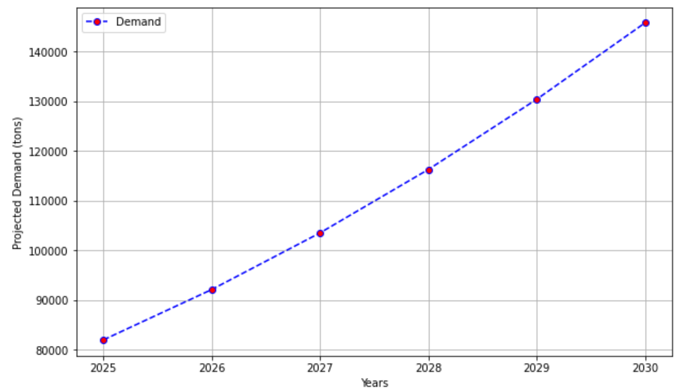


Figure 4.1: Projected yearly LPG demand for BAU scenario (in tons)

segments in Rwanda for LPG consumption: households, public non-commercial institutions, and commercial and industrial institutions. Households represent citizens' houses; public non-commercial institutions correspond to schools, hospitals, prisons, health facilities, and refugee camps; and commercial and industrial segment stands for hotels, restaurants, and agribusinesses. In 2020, it is known that the household segment consumed 73% of LPG volume. This volume is anticipated to grow to at least 80% by 2030 [22]. While cylinders are used to satisfy household demand, tankers are used for the other segments, where LPG is provided in bulk form. In this thesis, since we only consider the households with cylinder demand, we work with 80% of yearly projected LPG demand provided in Table 4.4, considering the projected cylinder demand ratio in 2030.

- **Uniform fuel prices:** We assume that fuel prices are uniform over time. In addition, we simplify the problem by assuming that fuel prices are constant across different regions. Although, in reality, fuel prices may fluctuate over time and region, focusing on fuel prices and incorporating this into the model would make the problem complicated. Since the primary objective of this thesis does not rely on fuel prices, this is a useful assumption to make.

- **Excluding last mile logistics:** As mentioned before, this thesis considers down to the retail point. Although some consumers might be far from a retailer, we ignore the possibility of house shipping. While optimizing last mile logistic systems is effective in increasing the accessibility and usage of the product, this thesis focuses on decisions at the higher planning level. Thus, any cost factors related to last mile logistics are excluded.

- **Excluding routing decisions of vehicles:** In this thesis, we optimize the number of vehicles used in each facility and their allocations to the higher-level facilities on the hierarchy. We do not optimize vehicle routing decisions. We assume that a vehicle chooses the shortest path to its destination facility without visiting any other facility as we use OSRM API. At the same time, we assume that the vehicle chooses the same path while returning back to its origin facility.

- **Identical road conditions:** In this thesis, we acknowledge that the road conditions - its slope, surface quality, etc. - may affect the fuel consumption level of the vehicles. Nevertheless, regardless of their actual state, we assume that the road conditions are identical across the country. This simplification enables us to focus on the impact of distance on overall transportation costs, which is one of the primary investigation points in this thesis.

- **Null residual value of vehicles and infrastructure:** In reality, when the lifetimes of vehicles and infrastructure are completed, their residual value may not be zero. However, in order to simplify the cost calculations, the residual values of these properties are considered to be zero at the end of their useful lives.

- **Daily working hours assumption:** Qualified personnel are required to operate the vehicles. Their daily working hours are set at 8 hours. This assumption can be considered reasonable when the average trip hours of the vehicles are investigated. In the majority of cases, the trip durations stay within the 8-hour threshold when trip distances are analyzed. In this context, a trip is defined as going to the destination facility and returning back to the origin facility.

  Though the value of daily working hour is set at 8 hours, yearly working days may vary slightly from year to year. The values of yearly working days are provided in Table 4.5.

Table 4.5: Yearly Working Days

Year	Working Days
2025	248
2026	249
2027	249
2028	248
2029	249
2030	249

- **Losses through the system:** Some portion of LPG may be lost during the transportation, storage, and distribution of the product. These losses may occur due to factors such as leakage and evaporation. We assume that there is no loss at any stage of the supply chain system. However, the decision-maker can easily incorporate loss to the system by multiplying the loss factor with flow variables in the inventory constraints in the mathematical model formulated in Section 5.1. In addition, the amount of volume bottled, the amount of demand, and the amount of LPG imported should be multiplied with the loss factor too. In this case, the model designs the system regarding this loss to ensure that the demand is satisfied sufficiently at the final point.

- **Total Cylinder Number Assumption:** We assume that there are as many empty cylinders in the system as many as the full cylinders. Therefore, the total number of cylinders in the system is expressed as two times the number of full cylinders.

Chapter 5

Mathematical Model

5.1 Main Model

Regarding the above mentioned assumptions and system description, we formulate the following MILP model to design a hierarchical supply chain system based on the monthly operations. List of index sets, parameters and decision variables are given in Table 5.1, Table 5.2, Table 5.3, and Table 5.4. Index t is the time index and it represents a month.

Table 5.1: Mathematical model notations - part1

	Index sets
B	Bottling Stations
W	Warehouses
R	Retailers
T	Time
I	Cylinder

Table 5.2: Mathematical model notations - part2

	Parameters
r	Discount rate
w_i	Weight of cylinder i (kg)
$CapV$	Capacity of a van (kg)
$CapTr$	Capacity of a truck (kg)
$CapTa$	Capacity of a tanker (kg)
Sv	Speed of a van (km/h)
Str	Speed of a truck (km/h)
Sta	Speed of a tanker (km/h)
$BotCap$	Bottling capacity of a bottling station (kg)
$Capw$	Capacity of a warehouse (kg)
$BulkCap$	Bulk inventory capacity of a bottling station (kg)
$CylCap$	Cylinder inventory capacity of a bottling station (cylinder)
$FOMw$	Fixed cost of a warehouse ($/year)
ONw	Overnight cost of a warehouse ($)
$FOMb$	Fixed cost of a bottling station ($/year)
ONb	Overnight cost of a bottling station ($)
$VOMb$	Variable cost of a bottling station ($/kg)
$FOMv$	Fixed cost of a van ($/year)
ONv	Overnight cost of a van ($)
$VOMv$	Variable cost of a van ($/km)
$FOMtr$	Fixed cost of a truck ($/year)
$ONtr$	Overnight cost of a truck ($)
$VOMtr$	Variable cost of a truck ($/km)
$FOMta$	Fixed cost of a tanker ($/year)
$ONta$	Overnight cost of a tanker ($)
$VOMta$	Variable cost of a tanker ($/km)
$cylCost_i$	Cost of cylinder i ($)
$dem_{i,r,t}$	Cylinder i demand of retailer r at time t (kg)
$dist_{w,r}$	Distance between warehouse w and retailer r (km)
$dist_{b,w}$	Distance between bottling station b and warehouse w (km)
$dist_b$	Distance between the depot and bottling station b (km)

Table 5.3: Mathematical model notations - part3

Parameters	
dFw	Discount factor of a warehouse
dFb	Discount factor of a bottling station
dFv	Discount factor of a van
$dFtr$	Discount factor of a truck
$dFta$	Discount factor of a tanker
dFc	Discount factor of a cylinder
$dFvl$	Discount factor of a van license
$dFtrl$	Discount factor of a truck licence
$dFtal$	Discount factor of a tanker license
$dFwsal$	Discount factor of a warehouse wholesale license
$dFcreq$	Discount factor of cylinder requalification cost
$dFbop$	Discount factor of a bottling station operating license cost
$dFdop$	Discount factor of a depot operating license cost
$dFil$	Discount factor of import license cost
$laborC$	Labor cost (%)
$salary$	Annual salary ($)
$yearWh$	Yearly working hours (hour)
$monthWd$	Monthly working days (day)
$dayWh$	Daily working hours (hour)
$licenseCtr$	Truck license cost ($)
$licenseCv$	Van license cost ($)
$licenseCta$	Tanker license cost ($)
$Wsal$	Warehouse wholesale license cost ($)
$Creq$	Cylinder requalification cost ($/cylinder)
$InsB$	Bottling station installation fee ($)
$InsD$	Depot installation fee ($)
$OpLb$	Bottling station operating license cost ($)
$OpLd$	Depot operating license cost ($)
ImL	Importation license cost ($)
$PriceI$	LPG importation price ($/kg)

Table 5.4: Mathematical model notation - part4

Decision Variables	
I_w	$= \begin{cases} 1, & \text{if warehouse } w \in W \text{ is opened} \\ 0, & \text{otherwise} \end{cases}$
I_b	Number of bottling stations opened in candidate bottling station location $b \in B$
$Capd$	Capacity of the depot (kg)
$numV_w$	Number of vans in warehouse w
$numTr_b$	Number of trucks in bottling station b
$numTa$	Number of tankers in the depot
$vanTrip_{w,r,t}$	Number of trips a van is performing between warehouse w and retailer r at time t
$truckTrip_{b,w,t}$	Number of trips a truck is performing between bottling station b and warehouse w at time t
$tankerTrip_{b,t}$	Number of trips a tanker is performing between the depot and bottling station b at time t
$volbol_{b,t}$	Volume bottled in bottling station b at time t
$bulkInv_{b,t}$	Amount of bulk inventory at bottling station b at time t
$Inv_{i,b,t}$	Inventory level of cylinder i in bottling station b at time t
$Inv_{i,w,t}$	Inventory level of cylinder i in warehouse w at time t
$f_{i,w,r,t}$	Amount of flow of cylinder i from warehouse w to retailer r at time t (cylinder)
$f_{i,b,w,t}$	Amount of flow of cylinder i from bottling station b to warehouse w at time t (cylinder)
$f_{b,t}$	Amount of flow from the depot to bottling station b at time t (kg)
$numCbot_{i,b,t}$	Number of cylinder i bottled in bottling station b at time t (cylinder)
$totNumFullCyl_i$	Total number full cylinders i in the system (cylinder)

5.1.1 Objective Function

The objective of the model is to minimize the total system cost. Hence, each cost component of the system is taken into consideration in the objective function.

The contribution of fixed costs and overnight costs of warehouses and bottling stations are represented in (5.1). The overnight costs are discounted using the discount factor. In addition, the variable cost of bottling stations is included via (5.2). A cost is incurred for each m^3 LPG bottled in a bottling station.

$$\mathrm{Z1} = \sum_{w \in W} (FOMw + ONw \cdot dFw) \cdot I_w + \sum_{b \in B} (FOMb + ONb \cdot dFb) \cdot I_b \tag{5.1}$$

$$\mathrm{Z2} = \sum_{b \in B} \sum_{t \in T} (VOMb \cdot volbol_{b,t}) \tag{5.2}$$

The fixed costs and overnight costs of vans, trucks, and tankers are represented in (5.3). The costs are added according to the number of vehicles operating in the system in a year and discounted accordingly.

$$\begin{aligned} \mathrm{Z3} = & \sum_{w \in W} (FOMv + ONv \cdot dFv) \cdot numV_w \\ & + \sum_{b \in B} (FOMtr + ONtr \cdot dFtr) \cdot numTr_b \\ & + (FOMta + ONta \cdot dFta) \cdot numTa \end{aligned} \tag{5.3}$$

Variable costs are incurred for each kilometer the vehicles travel in a month. Since the fixed costs are considered to be 127% of the variable costs, the variable costs are multiplied by 2.27 in (5.4). In addition, considering the return of the vehicles to the facilities, distances are multiplied by 2.

$$
\begin{aligned}
\text{Z4} = 2 \cdot & \sum_{w \in W} \sum_{r \in R} \sum_{t \in T} (VOMv \cdot dist_{w,r} \cdot vanTrip_{w,r,t} \cdot 2.27) \\
& + 2 \cdot \sum_{b \in B} \sum_{w \in W} \sum_{t \in T} (VOMtr \cdot dist_{b,w} \cdot truckTrip_{b,w,t} \cdot 2.27) \\
& + 2 \cdot \sum_{b \in B} \sum_{t \in T} (VOMta \cdot dist_{b} \cdot tankerTrip_{b,t} \cdot 2.27)
\end{aligned} \tag{5.4}
$$

The discounted cost of cylinders and the requalification cost of cylinders are expressed in (5.5). The total number of full cylinders in the system is multiplied by 2 since it is assumed that there must be as many empty cylinders in the system as there are full cylinders.

$$
\begin{aligned}
\text{Z5} = & \sum_{i \in I} (totNumFullCyl \cdot cylCost \cdot dFc \cdot 2) \\
& + \sum_{i \in I} (totNumFullCyl \cdot Creq \cdot dFcreq \cdot 2)
\end{aligned} \tag{5.5}
$$

Labor costs are incurred on an hourly basis for the workers who are responsible for driving the vehicles. Hence, the labor costs are calculated according to the total vehicle operating hours for a month, considering the total number of trips made by the vehicles. Labor costs are expressed in (5.6) for van, truck, and tanker trips, respectively.

$$
\begin{aligned}
\text{Z6} = & \sum_{w \in W} \sum_{r \in R} \sum_{t \in T} \frac{laborC \cdot salary}{yearWh} \cdot vanTrip_{w,r,t} \cdot \frac{dist_{w,r} \cdot 2}{Sv} \\
& + \sum_{b \in B} \sum_{w \in W} \sum_{t \in T} \frac{laborC \cdot salary}{yearWh} \cdot truckTrip_{b,w,t} \cdot \frac{dist_{b,w} \cdot 2}{Str} \\
& + \sum_{b \in B} \sum_{t \in T} \frac{laborC \cdot salary}{yearWh} \cdot tankerTrip_{b,t} \cdot \frac{dist_{b} \cdot 2}{Sta}
\end{aligned} \tag{5.6}
$$

Importation cost is included via (5.7).

$$Z7 = \sum_{b \in B} \sum_{t \in T} f_{b,t} \cdot priceI \tag{5.7}$$

Finally, vehicle and facility license costs are added via (5.8).

$$\begin{aligned} Z8 = &(licenseCv \cdot dFvl) + (licenceCtr \cdot dFtrl) + (licenseCta \cdot dFtal) \\ &+ (Wsal \cdot dFwsal) + (InsB \cdot dFb) + (OpLb \cdot dFbop) + (InsD \cdot dFd) \\ &+ (OpLd \cdot dFdop) + (ImL \cdot dFil) \end{aligned} \tag{5.8}$$

5.1.2 Constraints

5.1.2.1 Depot - Related Constraints

$$\sum_{b \in B} f_{b,t} \leq Capd \quad \forall t \in T \tag{5.9}$$

$$\frac{f_{b,t}}{CapTa} \leq tankerTrip_{b,t} \quad \forall b \in B, t \in T \tag{5.10}$$

$$\sum_{b \in B} tankerTrip_{b,t} \cdot \frac{dist_b \cdot 2}{Sta} \leq numTa \cdot dayWh \cdot monthWd \quad \forall t \in T \tag{5.11}$$

$$numTa, tankerTrip_{b,t} \in \mathbb{Z}^+ \cup \{0\} \quad \forall b \in B, t \in T \tag{5.12}$$

$$Capd, f_{b,t} \geq 0 \quad \forall b \in B, t \in T \tag{5.13}$$

Constraint (5.9) ensures that the LPG stored in the depot does not exceed the capacity of the depot. Constraint (5.10) specifies the number of trips each tanker makes in a month. Also, it ensures that the flow carried from the depot to the bottling station does not exceed the capacity of the tanker. Constraint (5.11) determines the number of tankers necessary to run the monthly operations. Constraints (5.12) - (5.13) defines the decision variables related to depot operations.

5.1.2.2 Bottling Station - Related Constraints

$$\sum_{i\in I}\sum_{w\in W} f_{i,b,w,t} \leq I_b \cdot BigM \quad \forall b \in B, t \in T \tag{5.14}$$

$$\sum_{t\in T} volbol_{b,t} \leq BotCap \cdot I_b \quad \forall b \in B \tag{5.15}$$

$$volBottled_{b,t} = \sum_{i\in I} numCBottled_{i,b,t} \cdot cylinder\_kg \quad \forall b \in B, t \in T \tag{5.16}$$

$$bulkInv_{b,t} = bulkInv_{b,t-1} + f_{b,t} - volbol_{b,t} \quad \forall b \in B, t \in \{2, ..., T\} \tag{5.17}$$

$$bulkInv_{b,t} \leq BulkCap \quad \forall b \in B, t \in T \tag{5.18}$$

$$Inv_{i,b,t} = Inv_{i,b,t-1} + numCbot_{i,b,t} - \sum_{w\in W} f_{i,b,w,t} \quad \forall i \in I, b \in B, t \in \{2, ..., T\} \tag{5.19}$$

$$Inv_{i,b,t} \leq CylCap \quad \forall i \in I, b \in B, t \in T \tag{5.20}$$

$$\sum_{w\in W} truckTrip_{b,w,t} \cdot \frac{dist_{b,w} \cdot 2}{Str} \leq numTr_b \cdot dayWh \cdot monthWd \quad \forall b \in B, t \in T \tag{5.21}$$

$$\sum_{i\in I} \frac{f_{i,b,w,t} \cdot w_i}{CapTr} \leq truckTrip_{b,w,t} \quad \forall b \in B, w \in W, t \in T \tag{5.22}$$

$$I_b, numTr_b, truckTrip_{b,w,t}, f_{i,b,w,t} \in \mathbb{Z}^+ \cup \{0\} \quad \forall i \in I, b \in B, w \in W, t \in T \tag{5.23}$$

$$Cap_b, volbol_{b,t}, numCbot_{b,t}, totNumFullCyl \geq 0 \quad \forall i \in I, b \in B, t \in T \tag{5.24}$$

$$Inv_{i,b,t}, bulkInv_{b,t} \geq 0 \quad \forall i \in I, b \in B, t \in T \tag{5.25}$$

Constraint (5.14) states that if there is a flow from a bottling station to a warehouse, then that bottling station must be open. Constraint (5.15) ensures that the volume bottled in each bottling station does not exceed the bottling station's bottling capacity. Constraint (5.16) sets the number of cylinders bottled in the bottling station. Constraint (5.17) sets the inventory level for LPG stored in bulk in bottling stations. Additionally, Constraint (5.18) indicates that the bulk inventory level cannot exceed the inventory capacity of a bottling station. Similarly, Constraint (5.19) sets the inventory level for LPG stored in cylinders in bottling stations, and Constraint (5.20) states that cylinder inventory level cannot exceed the inventory capacity of a bottling station. Constraint (5.21) determines the number of trucks needed to run the monthly deliveries from bottling stations to warehouses, considering the monthly working hours. Finally, Constraint (5.22) ensures that the weight of the cylinders carried from the bottling stations to the warehouses does not exceed the capacity of the trucks. Constraints (5.23) - (5.25) defines the decision variables related to bottling station operations.

5.1.2.3 Warehouse - Related Constraints

$$\sum_{r\in R}\sum_{i\in I} f_{i,w,r,t} \leq I_w \cdot BigM \quad \forall w \in W, t \in T \tag{5.26}$$

$$Inv_{i,w,t} = Inv_{i,w,t-1} + \sum_{b\in B} f_{i,b,w,t} - \sum_{r\in R} f_{i,w,r,t} \quad \forall i \in I, w \in W, t \in \{2,..,T\} \tag{5.27}$$

$$\sum_{i\in I} Inv_{i,w,t} \cdot cylinder\_kg \leq Capw \quad \forall w \in W, t \in T \tag{5.28}$$

$$\sum_{r\in R} vanTrip_{w,r,t} \cdot \frac{dist_{w,r} \cdot 2}{Svan} \leq numV_w \cdot dayWh \cdot monthWd \quad \forall w \in W, t \in T \tag{5.29}$$

$$\sum_{i\in I} \frac{f_{i,w,r,t} \cdot w_i}{CapV} \leq vanTrip_{w,r,t} \quad \forall w \in W, r \in R, t \in T \tag{5.30}$$

$$I_w \in \{0,1\} \quad \forall w \in W \tag{5.31}$$

$$numV_w, vanTrip_{w,r,t}, f_{i,w,r,t} \in \mathbb{Z}^+ \cup \{0\} \quad \forall i \in I, w \in W, r \in R, t \in T \tag{5.32}$$

$$Cap_w, Inv_{i,w,r,t} \geq 0 \quad \forall i \in I, w \in W, r \in R, t \in T \tag{5.33}$$

Constraint (5.26) indicates that if there is a flow from a warehouse to a retailer, then that warehouse must be open. Constraint (5.27) sets the inventory level in each warehouse. Constraint (5.28) ensures that the number of cylinders stored in a warehouse does not exceed the capacity of a warehouse. Constraint (5.29) determines the number of vans needed to carry the cylinders from warehouses to retailers in a month, considering the monthly working hours. Constraint (5.30) ensures that the weight of the cylinders carried from the warehouses to the retailers does not exceed the capacity of the vans. Constraints (5.31) - (5.33) defines the decision variables related to warehouse operations.

5.1.2.4 Retailer - Related Constraints

$$\sum_{w \in W} f_{i,w,r,t} \geq \frac{dem_{i,r,t}}{w_i} \quad \forall i \in I, r \in R, t \in T \tag{5.34}$$

$$\begin{aligned} totNumFullCyl_i = & \sum_{w \in W} \sum_{r \in R} f_{i,w,r,t} + \sum_{w \in W} \sum_{b \in B} f_{i,b,w,t} \\ & + \sum_{w \in W} Inv_{i,w,t} + \sum_{b \in B} Inv_{i,b,t} \quad \forall i \in I, t \in T \end{aligned} \tag{5.35}$$

Constraint (5.34) ensures that each retailer's demand is satisfied. Lastly, Constraint (5.35) specifies the number of cylinders present in the entire system.

Entire model can be found in Appendix C.

After solving the main model, we report the objection function value, which is the summation of Z_i values explained in Section 5.1.1. We also calculate the cost incurred per kilogram of LPG using $\frac{Z^*}{\text{total demand}}$ where Z^* indicates the objective function value of the main model.

5.2 Model Decomposition

Since this is a large-scale problem, solving the model to optimality takes a considerably high amount of time. Hence, to obtain solutions in a shorter time, we propose to decompose the location-allocation problem formulated in Section 5.1.

While decomposing the model, we follow a bottom-up approach. In other words, we divide the main location allocation model into two phases. While the first phase represents the warehouse-retailer part of the problem, the second phase models the depot-bottling station-warehouse part of the problem.

The idea behind using a bottom-up approach is to first decide on the locations of the warehouses and the total flow originating from each warehouse. Later, the sum of flows sent from each warehouse is used as the LPG demand of the corresponding warehouse. Consequently, we treat the system as if the final demand points are warehouses instead of retailers in the second phase model. Given the location of demand points and their demands, the second phase model decides on the locations of the bottling stations and other operational decisions related to the depot-bottling station - warehouse system.

5.2.1 First Phase Model

For the first phase model, we provide the candidate warehouse locations, retailer locations, and LPG demand of each retailer as parameters. In return, we determine the locations of the warehouses, the number of vans used in each warehouse, and the flow sent from warehouses to retailers. Regarding these, the first phase model is formulated below. Note that while we use $totNumFullCyl_i$ in the main model, since it is also associated with decision variables related to the bottling station, a new decision variable, $totNumFullCyl_i^{wr}$, is defined for the first phase model. $totNumFullCyl_i^{wr}$ represents the total number of full cylinders in the warehouse-retailer system.

$$
\begin{aligned}
\min \quad Z9 = & \sum_{w \in W} (FOMw + ONw \cdot dFw) \cdot I_w \\
& + \sum_{w \in W} (FOMv + ONv \cdot dFv) \cdot numV_w \\
& + 2 \cdot \sum_{w \in W} \sum_{r \in R} \sum_{t \in T} (VOMv \cdot dist_{w,r} \cdot vanTrip_{w,r,t} \cdot 2.27) \\
& + \sum_{i \in I} (totNumFullCyl_i^{wr} \cdot cylCost \cdot dFc \cdot 2) \\
& + \sum_{i \in I} (totNumFullCyl_i^{wr} \cdot Creq \cdot dFcreq \cdot 2) \\
& + \sum_{w \in W} \sum_{r \in R} \sum_{t \in T} \frac{laborC \cdot salary}{yearWh} \cdot vanTrip_{w,r,t} \cdot \frac{dist_{w,r} \cdot 2}{Sv} \\
& + (licenseCv \cdot dFvl) + (Wsal \cdot dFwsal)
\end{aligned}
$$

s.t.

$$(5.26), (5.29), (5.30), (5.31), (5.32), (5.33), (5.34)$$

$$totNumFullCyl_i^{wr} = \sum_{w \in W} \sum_{r \in R} f_{i,w,r,t} \quad \forall i \in I, t \in T \tag{5.36}$$

Note that the constraints related to $Inv_{i,w,t}$ are not included in the first phase model ((5.27) and (5.28)). Since in the current system, there are only warehouses and retailers; warehouses do not receive any flow $f_{i,b,w,t}$ in Constraint (5.27). Hence, without any replenishment, the flow going out from the warehouses $f_{i,w,r,t}$ makes the model infeasible eventually, depending on the starting inventory level. Therefore, in the first phase model, it is assumed that warehouses do not keep any inventory.

By solving the first phase model, we determine the flow sent from each warehouse. $\sum_{r \in R} f_{i,w,r,t} \quad \forall i \in I, w \in W, t \in T$ denotes the total flow sent from each warehouse which will be used in the second phase model as the demand of each warehouse.

5.2.2 Second Phase Model

For the second phase model, we provide the warehouse locations and demand of each warehouse. The demand of warehouses is represented by $dem_{i,w,t}$ which is equal to $\sum_{r \in R} f_{i,w,r,t} \quad \forall i \in I, w \in W, t \in T$. As a result of the model, we retrieve the bottling station locations, number of trucks used in each bottling station, truck trip numbers, number of cylinders bottled in bottling stations, number of tankers used, and tanker trip numbers. In the second phase model, to represent the total number of full cylinders in the system, we use $totNumFullCyl_i^{bw}$ decision variable. The second phase model is formulated as follows:

$$
\begin{aligned}
\min \quad \mathrm{Z10} = & \sum_{b \in B} (FOMb + ONb \cdot dFb) \cdot I_b \\
& + \sum_{b \in B} \sum_{t \in T} (VOMb \cdot volbol_{b,t}) \\
& + \sum_{b \in B} (FOMtr + ONtr \cdot dFtr) \cdot numTr_b \\
& + (FOMta + ONta \cdot dFta) \cdot numTa \\
& + 2 \cdot \sum_{b \in B} \sum_{w \in W} \sum_{t \in T} (VOMtr \cdot dist_{b,w} \cdot truckTrip_{b,w,t} \cdot 2.27) \\
& + 2 \cdot \sum_{b \in B} \sum_{t \in T} (VOMta \cdot dist_b \cdot tankerTrip_{b,t} \cdot 2.27) \\
& + \sum_{i \in I} (totNumFullCyl_i^{bw} \cdot cylCost \cdot dFc \cdot 2) \\
& + \sum_{i \in I} (totNumFullCyl_i^{bw} \cdot Creq \cdot dFcreq \cdot 2) \\
& + \sum_{b \in B} \sum_{w \in W} \sum_{t \in T} \frac{laborC \cdot salary}{yearWh} \cdot truckTrip_{b,w,t} \cdot \frac{dist_{b,w} \cdot 2}{Str} \\
& + \sum_{b \in B} \sum_{t \in T} \frac{laborC \cdot salary}{yearWh} \cdot tankerTrip_{b,t} \cdot \frac{dist_b \cdot 2}{Sta} \\
& + \sum_{b \in B} \sum_{t \in T} f_{b,t} \cdot priceI
\end{aligned}
$$

$$+ (licenceCtr \cdot dFtrl) + (licenseCta \cdot dFtal) + (InsB \cdot dFb)$$

$$+ (OpLb \cdot dFbop) + (InsD \cdot dFd) + (OpLd \cdot dFdop) + (ImL \cdot dFil)$$

s.t.

$$(5.9) - (5.25)$$

$$totNumFullCyl_i^{bw} = \sum_{b \in B} \sum_{w \in W} f_{i,b,w,t} + \sum_{b \in B} Inv_{i,b,t} \quad \forall i \in I, t \in T \tag{5.37}$$

$$\sum_{b \in B} f_{i,b,w,t} \geq dem_{i,w,t} \quad \forall i \in I, w \in W, t \in T \tag{5.38}$$

After solving the second phase model, we have decided on all the tactical and strategic actions that are present in the main model. As a result, we report the total cost of the system as the sum of the objective function values of first and second phase models: Z9 + Z10. In addition, we express the total unit cost of the system as $\frac{Z9+Z10}{\text{total demand}}$.

Chapter 6

Computational Analysis

Since the simple plant facility location problem is an NP-hard problem, the main and the decomposed models are NP-hard problems. Hence, while it is ideal to solve both the main and decomposed model to optimality, it takes a very long time to do that. Thus, due to the scale of the problem, we run the models either with optimality gap thresholds or time limits. The threshold used will be specified while presenting the solution of each configuration. All the models are coded in Python and solved using Gurobi.

Since there is no seasonality in demand, we have the agility to run the models independent of time index t. By omitting index t on the decision variables and parameters, we are able to decrease the size of the model presented in Section 5.1. While other planning horizons may also be selected, we prefer to specify $t = 1$ where t indicates a month. The yearly demand values are allocated equally to 12 months regarding the absence of demand seasonality. Consequently, the objective function value of the model where $t = 1$ is 12 times smaller than the objective function value of the model where $t \in \{1, ..., 12\}$. In the system, the vehicles and cylinders are purchased at the beginning of the first planning horizon. Hence, when the acquisition costs of these entities are discounted to present value using monthly discount factors and the monthly operational cost of the system is calculated, we will obtain a model independent of index t that is equivalent to a model including index t.

In this chapter, first, we will discuss the outputs of models in which the retailer locations are selected based on the agglomerative hierarchical clustering-based heuristic approach with different linkage parameters, as explained in Section 3.3.2. Then, we will compare the result of the model yielding the least total cost among these with the result of the model that uses retailer locations as the centroid point of each village. Next, we will discuss the impact of different demand scenarios on the system configuration. In addition, we will investigate whether there is a trade-off between the infrastructure cost and transportation cost by a sensitivity analysis. Finally, we will provide the outputs of the decomposed model that is provided in Section 5.2.

6.1 Linkage Parameter Comparison

The agglomerative hierarchical clustering-based heuristic is run with Complete, Average, Centroid, and Ward's minimum variance linkage parameters to determine the retailer locations. As a result, the algorithm yields 5969, 6734, 6856, and 6920 retailer locations for Complete, Average, Centroid, and Ward's minimum variance linkage parameters, respectively.

The main model formulated in Section 5.1 is run for each linkage parameter for the 2025 BAU demand scenario. These models are run for 24 hours, and their outputs are reported in Table 6.1. In the table, The Total Cost corresponds to the objective function value of the model. The Importation Cost is associated with the cost of importing LPG into the country to satisfy demand. Since the 2025 BAU demand scenario is used in each model, the Importation Costs are expected to be similar in the four models. Hence, in order to clearly see the impact of retailer numbers on the system design, the Total Cost-Importation Cost is also reported. The Total Infrastructure Cost contains each cost component related to the depot, bottling stations, and warehouses. Total Transportation Cost includes each cost component related to the tankers, trucks, and vans used in the system.

In addition, the Total Cylinder cost is reported, which represents the acquisition cost of the total number of cylinders in the system. For each model, the Cost Per Kilogram of LPG is recorded. Finally, since the models' execution time is specified to be 24 hours, the optimality gaps at the end of 24 hours are listed.

Table 6.1: Outputs of different linkage parameters

	Complete	**Average**	**Centroid**	**Ward**
# Retailers	5,969	6,734	6,856	6,920
# Warehouse	17	20	21	19
# Bottling Station	2	2	2	2
# Van	132	139	138	145
# Truck	16	16	16	16
# Tanker	4	4	4	4
# Van Trip	6,215	6,986	7,115	7,143
# Truck Trip	553	553	553	554
# Tanker Trip	230	230	230	230
Total Cost	5,961,626.403	6,028,864.969	6,042,454.261	6,038,678.242
Importation Cost	4,321,710.383	4,325,379.431	4,326,020.807	4,325,983.079
Total Cost-Importation Cost	1,639,916.021	1,703,485.538	1,716,433.455	1,712,695.163
Total Cost of Infrastructure	811,775.569	852,566.465	866,113.545	839,126.169
Total Cost of Transportation	611,070.242	633,664.572	633,033.194	656,284.174
Total Cost of Cylinder	217,070.210	217,254.500	217,286.716	217,284.821
Cost Per Kg	1.0914	1.1037	1.1062	1.1055
Optimality Gap	0.785%	0.829%	1.021%	0.812%

Since the total demand is the same, all four models yield the same number of trucks and tankers, as well as the number of truck trips and tanker trips. However, it is observed that as the number of retailers increases, the number of van trips also increases. As trip decision variables are specified to be integers, even if there is a small amount of flow that needs to be sent to a retailer, the trip number increases by at least 1. Hence, the increase in the retailer numbers directly corresponds to the increase in the van trip number. Consequently, the total transportation cost increases as the van numbers and van trip numbers increase.

Among four different linkage parameters, it is seen that the model using the retailer locations obtained by the Complete linkage parameter yields the minimum total system cost. Hence, the cost per kg is also the lowest for the Complete model. Thus, in the following analysis, the Complete model's outputs will be compared with the outputs of the model using the centroid of each village as retailer locations.

6.2 Retailer Location Comparison

As explained in Section 3.3.1, the other approach used to decide on the locations of retailers is selecting the centroid point of each village, as an uninformed decision-maker might do in reality. Since there are 14,815 villages in Rwanda, there are 14,815 retailers found with this approach. Using these locations, the main model is run for the 2025 BAU demand scenario to compare its outputs with the least cost model in Section 6.1. The Complete model is run for 24 hours, and the optimality gap is recorded as 0.785 %. However, it is seen that after 24 hours, the optimality gap is still high for the 14K model. Thus, the 14K model is run until the optimality gap drops to 1 % with an execution time of 53.1 hours. The complete model has 185,435 constraints, 67 continuous and 358,642 integer variables (30 binary). The 14K model has 459,661 constraints, 67 continuous and 889,402 integer variables (30 binary). The outputs of these two models can be found in Table 6.2.

Table 6.2: Retailer location comparison model outputs

	Complete	**14K**
# Retailers	5,969	14,815
# Warehouse	17	25
# Bottling Station	2	2
# Van	132	207
# Truck	16	18
# Tanker	4	4
# Van Trip	6,215	13,709
# Truck Trip	553	560
# Tanker Trip	230	232
Total Cost	5,961,626.403	6,383,343.268
Importation Cost	4,321,710.383	4,354,741.247
Total Cost-Importation Cost	1,639,916.021	2,028,602.021
Total Cost of Infrastructure	811,775.569	922,545.113
Total Cost of Transportation	611,070.242	887,327.611
Total Cost of Cylinder	217,070.210	218,729.296
Cost Per Kg	1.0914	1.1686
Optimality Gap	0.785%	0.996 %

While the model consistently prefers to open two bottling stations in both scenarios due to the same amount of total demand, the higher number of retailers in the 14K case results in more warehouses being opened to minimize transportation costs. With the number of bottling stations remaining constant, the 14K model opens more warehouses, leading to increased infrastructure costs. As previously explained, the increase in retailer locations leads to a higher number of van trips and an increased number of vans. Consequently, the total transportation cost is higher for the 14K model.

Although total demand is the same in both cases, since the flows carried by trucks and vans are specified to be integer (cylinder flow), depending on the retailer's demand, the total number of cylinders delivered to the retailer may vary.

To be more precise, the total number of cylinders is 916,384 in the Complete model and 923,388 in the 14K model. The number of cylinders used in the two cases explains the slight difference in the total cost of cylinder. This is the reason for the slight difference between the importation costs of the two cases. However, it is seen that regardless of the importation cost, the Complete model yields a lower cost solution compared to the 14K model, where the total cost of the Complete model is 1.6M$, and the total cost of the 14K model is 2.0M$, excluding the importation cost. When the cost per kilogram values are evaluated, it is also seen that the Complete model yields lower costs.

In addition, the motivation to propose another approach in selecting the retailer locations is to ensure fairness among rooftops, as explained in Section 3.3.2. By using the agglomerative hierarchical clusterin-based heuristic approach, we make sure that each rooftop is within a 1-mile radius of its retailer, unlike the case in the centroid approach. In-field conversations informed us that 1 mile is considered as the minimum distance to a retailer when designing a logistic chain in this context. By examining the distance statistics between the retailers and rooftops of the Complete model and the 14K model, we can see this distinction clearly. The distance statistics are provided in Table 6.3 for two selected districts: Nyarugenge and Kayonza. While Nyarugenge is an urban region with the highest population density among all districts, Kayonza is a rural region with the lowest population density in the country.

Table 6.3: Complete and 14K distance statistics (km)

	Complete - Urban	**Complete - Rural**	**14K - Urban**	**14K - Rural**
Average	0.671	0.677	0.231	0.628
Median	0.674	0.648	0.177	0.532
Min	$1.3e^{-06}$	0	0.0006	0.0015
Max	1.609	1.609	1.234	9.861
Sum	73,006.509	77,846.686	25,328.569	72,551.875
S.D.	0.301	0.364	0.174	0.533

As targeted, in the Complete model, the maximum distance between a rooftop and its retailer is 1.6090 kilometers (1 mile), regardless of whether the district is an urban or rural region. It is also seen that the minimum distance is 0 km in the Complete model, indicating at least one retailer is placed exactly on the same location of a rooftop. While this scenario may be hard to implement in reality, within the context of this thesis, it is a feasible scenario. Furthermore, it is recorded that the average distance between a rooftop and its retailer is almost the same for the urban region (0.6708 km) and rural region (0.6770 km), indicating fairness among rooftops in the Complete model.

On the other hand, in the 14K model, it is seen that the maximum distance between a rooftop and its retailer is 1.2338 km in the urban region but 9.8612 km in the rural region. Considering the varying surface areas of the villages, this is an expected difference to encounter. Hence, in the 14K model, even though it is possible to keep the maximum distance between a rooftop and its retailer within a 1-mile threshold in urban regions, this is not the case in rural regions. Consequently, using the village centroid approach may not perform well in rural regions. Additionally, the difference between the average distances for the urban and rural regions indicates rooftops in the urban regions are closer to their retailers. The biggest difference is seen in the sum of the distances travelled by each rooftop to reach their retailers. The sum of distances is significantly higher in the rural region, highlighting the unfair distribution of retailers between urban and rural regions.

As a result, it is observed that the Complete model dominates the 14K model since it provides a lower-cost solution in less time with a lower optimality gap. Furthermore, it is also seen that the Complete model manages to ensure fairness among rooftops, regardless of the location of the rooftops. Hence, we will continue with the Complete model for the next analysis.

6.3 Different Demand Scenarios

In order to observe the impact of the projected increase in demand on the system design, the Complete model is run for each demand scenario, provided in Table 4.4. These models are executed for 24 hours, and at the end, the optimality gaps are recorded. The Complete model outputs for projected demand scenarios from 2025 to 2030 can be found in Table 6.4.

The aim of this analysis is to see the impact of projected demand increase on the system configuration. It is observed that starting from 2026, the bottling capacities of 2 bottling stations are not sufficient to fulfill the yearly LPG demand. Hence, over the years, as demand increases, the number of bottling stations increases. While this is an expected result, we decided to investigate further the locations of the opened bottling stations.

	2025	**2026**	**2027**	**2028**	**2029**	**2030**
Bottling Station Index	2-5	2-5-6	2-5-6	2-2-5	2-2-5	1-2-4-5

Table 6.5: Demands and Bottling Station Indices

In Table 6.5, the opened bottling station indices can be found. Bottling Station 2 is in the same location as the LPG depot. Hence, in order to minimize the transportation cost by reducing the number of tankers and tanker trips, the model prefers opening Bottling Station 2 in all cases. When Bottling Station 2 is open, it is assumed that LPG is not carried with tankers as the depot and Bottling Station 2 is in the same location.

Except for the case in 2030, it is observed that the model prefers to open Bottling Stations 2,5 and 6 with different combinations. In 2026, it is seen that the model prefers to keep the already existing bottling stations from 2025 (Bottling Stations 2 and 5) and adds Bottling Station 6 to satisfy the demand. As the demand in 2027 can also be satisfied by 3 bottling stations, the model keeps the current selection of bottling stations from 2026.

Since the decision variable I_b is specified to be an integer variable, the model has the agility to open more than one bottling station in the same location. As can be seen in 2028 and 2029, the model prefers to open 2 bottling stations on the location of Bottling Station 2. Since the demand in 2029 can be satisfied by 3 bottling stations, the model sustains the system configuration created in 2028.

In 2030, at least 4 bottling stations need to be open to fulfill the yearly LPG demand. Different than the other years, in 2030, the model introduces Bottling Stations 1 and 4. While Bottling Stations 1, 2, and 5 lay relatively close to each other on the map (Figure 6.1b), Bottling Station 4 stands separately (Figure 6.1d). The reason for choosing Bottling Station 4 is interpreted in a way that minimizes the primary transportation cost. With an increase in demand, the amount of flow carried between bottling stations and warehouses inclines. Hence, the location of bottling stations and warehouses should be selected in a way that minimizes the primary transportation cost.

Since in 2029, Bottling Stations 2,2,5 are open, we decide to examine the change in the system in case the existing configuration is kept by introducing only Bottling Station 1 in 2030. An increase in the number of truck trips indicates that the model prefers to open Bottling Station 4 in 2030 to decrease transportation costs due to increased demand. Hence, the configurations with Bottling Stations 1, 2, 2, 5 and Bottling Stations 1, 2, 4, 5 are compared. These outputs related to these two configurations are compared in Table 6.6. It is observed that when the model is forced to open Bottling Stations 1, 2, 2, 5 the truck number increases while the trip numbers are almost the same. This is because the distances travelled by trucks are 67052.896 km and 50759.291 km when Bottling Stations 1, 2, 2, 5 and Bottling Stations 1, 2, 4, 5 are opened, respectively. Therefore, when Bottling Stations 1, 2, 4, 5 are opened, the primary transportation cost is lower. Thus, it can be stated that Bottling Station 4 is introduced to the system as a new bottling station in order to minimize the transportation cost with an increase in demand. On the contrary, it is seen that the van numbers decrease due to a higher number of warehouses when Bottling Stations 1, 2, 2, 5 are opened. However, since the number of warehouses increases, the total infrastructure cost is higher in the configuration with Bottling Stations 1, 2, 2, 5. This is the major component that increases the total system cost for the configuration with Bottling Station 1, 2, 2, 5. Finally, it is observed that the tanker number for the Bottling Station 1, 2, 2, 5 configuration is half of the Bottling Station 1, 2, 4, 5 as the flow is not carried by tankers between the depot and Bottling Station 2. With a higher number of vans and tankers, the total cost of transportation for the Bottling Station 1, 2, 4, 5 seems to be higher.

Table 6.6: Model outputs when Bottling Stations 1,2,4,5 and 1,2,2,5 are opened

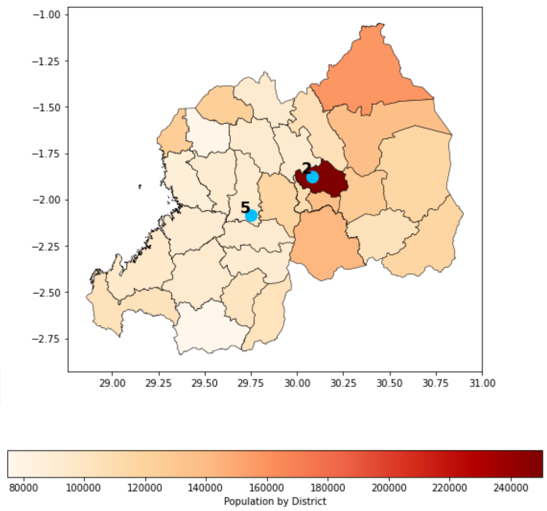
	Bs 1,2,4,5	**Bs 1,2,2,5**
# Warehouse	18	26
# Bottling Station	4	4
# Van	140	122
# Truck	21	25
# Tanker	10	5
# Van Trip	6,758	6,758
# Truck Trip	977	980
# Tanker Trip	480	407
Total Cost	10,097,006.634	10,160,991.709
Importation Cost	7,666,844.639	7,666,844.639
Total Cost-Importation Cost	2,430,161.995	2,494,147.070
Total Cost of Infrastructure	1,323,625.044	1,431,561.606
Total Cost of Transportation	721,446.158	677,494.672
Total Cost of Cylinder	385,090.793	385,090.793
Cost Per Kg	1.0390	1.0456
Optimality Gap	0.879%	0.999 %

Since in 2026 and 2027, Bottling Stations 2,5, and 6 are opened, we investigate the scenario when Bottling Station 6 is opened instead of Bottling Station 1 in 2030. Depending on the change in the total system cost, the decision-maker may prefer to keep the existing Bottling Stations 2,5, and 6, and add Bottling Station 1 in 2030 since Bottling Station 1 and Bottling Station 6 lay close to each other on the map. The system cost of configuration with Bottling Stations 1,2,4,5 and Bottling Stations 2,4,5,6 can be found in Table 6.7. It is observed that when Bottling Station 6 is opened instead of Bottling Station 1 in 2030, the total cost of the system (excluding the importation cost) increases by 1.894 % with a change of 46,905.572 \$. The configuration with Bottling Station 6 opens more warehouses, and the total bulk transportation cost is higher. However, considering the cost difference between the two configurations, the decision-maker may choose to open Bottling Station 6 instead of 1 to keep the already existing bottling stations in the system instead of opening completely new ones.

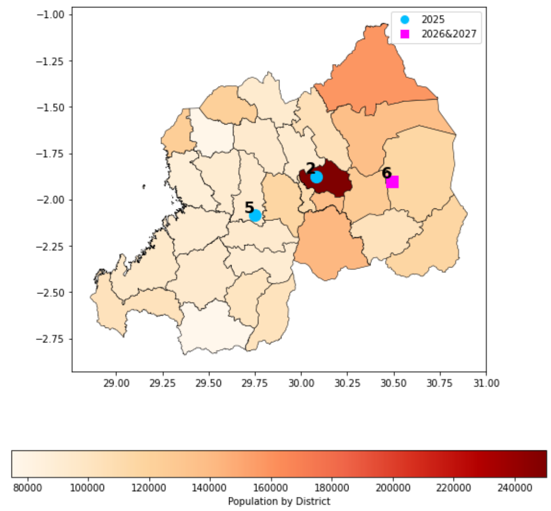
Table 6.7: Model outputs when Bottling Stations 1,2,4,5 and 2,4,5,6 are opened

	Bs 1,2,4,5	**Bs 2,4,5,6**
# Warehouse	18	25
# Bottling Station	4	4
# Van	140	124
# Truck	21	20
# Tanker	10	14
# Van Trip	6,758	6,760
# Truck Trip	977	980
# Tanker Trip	480	407
Total Cost	10,097,006.634	10,143,912.206
Importation Cost	7,666,844.639	7,666,844.639
Total Cost-Importation Cost	2,430,161.995	2,477,067.567
Total Cost of Infrastructure	1,323,625.044	1,418,069.535
Total Cost of Transportation	721,446.158	673,907.239
Total Cost of Cylinder	385,090.793	385,090.793
Cost Per Kg	1.0390	1.0438
Optimality Gap	0.879%	0.998 %

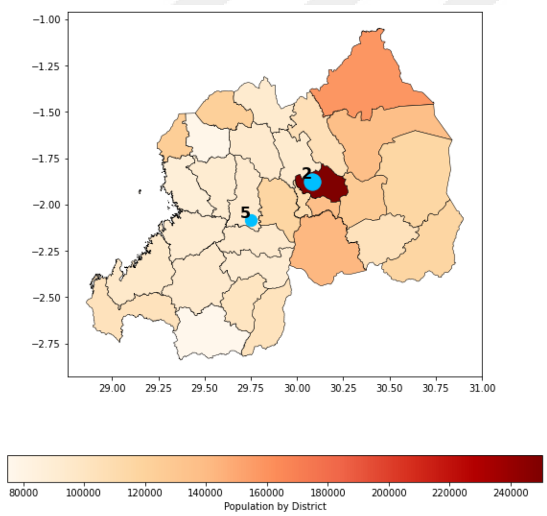
Overall, it is seen that, except for the configuration in 2030, the model prefers to keep the existing bottling stations and open new ones in case of an increase in demand. Hence, we can assert that we can expand the system and build upon the already existing ones.



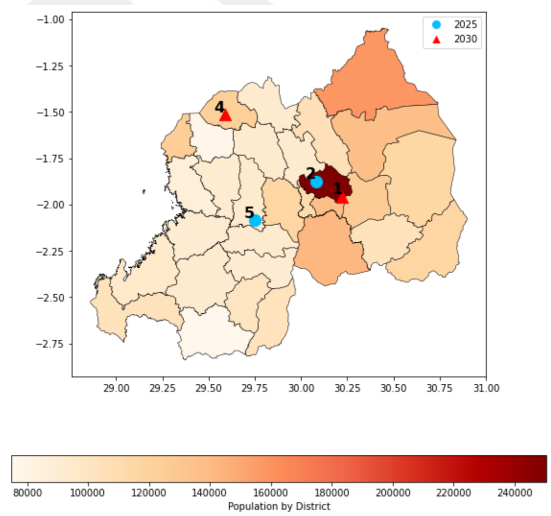
(a) Bottling stations opened in 2025



(b) Bottling stations opened in 2026-2027



(c) Bottling stations opened in 2028-2029



(d) Bottling stations opened in 2030

Figure 6.1: Locations of bottling stations opened for each year: This figure illustrates the distribution of bottling stations over different time periods, highlighting the expansion of facilities to meet increasing LPG demands. Bottling Stations 2 and 5 are opened in 2025 (Figure 6.1a). In 2026 and 2027, Bottling Station 6 (denoted with pink square) is added to the existing configuration (denoted with blue circles) (Figure 6.1b). In 2028 and 2029, Bottling Stations 2,2,5 are opened. The increased size of the label for Bottling Station 2 indicates the existence of two bottling stations in the location of Bottling Station 2 (Figure 6.1c). Finally, in 2030, Bottling Stations 1 and 4 are introduced (denoted with red triangles) on top of the existence configuration of Bottling Stations 2 and 5 (denoted with blue circles) (Figure 6.1d). The maps are colored according to the population density.

6.4 Sensitivity Analysis

In Rwanda, diesel fuel prices may fluctuate. Hence, in case of an increase in the diesel fuel price, the total supply chain system cost may be impacted heavily due to the incline in the transportation cost. In this section, we aim to show the trade-off between transportation costs and infrastructure costs through a sensitivity analysis. A rational decision-maker, having information about the possibility of fluctuations in diesel fuel prices, may prefer to invest either in transportation units or infrastructure.

In Section 6.3, it is seen that the Complete model opens 2 bottling stations and 17 warehouses. Hence, to show the trade-off, we report the solutions of different scenarios with combinations of 2 and 4 bottling stations; and 10, 17, and 25 warehouses. These models are run with a 48-hour limit and a 1% optimality gap. The model is terminated when one of these criteria is satisfied. The results can be seen in Table 6.8 and Table 6.9.

Table 6.8: Fixed number of warehouses and bottling stations (2 Bs)

	2bs_10w	**2bs_17w**	**2bs_25w**
# Warehouse	10	17	25
# Bottling Station	2	2	2
# Van	175	135	123
# Truck	18	20	18
# Tanker	0	0	4
# Van Trip	6,217	6,215	6,216
# Truck Trip	551	552	556
# Tanker Trip	230	230	230
Total Cost-Importation Cost	1,693,040.22	1,662,695.218	1,706,688.371
Total Cost of Infrastructure	717,331.08	811,775.569	919,712.130
Total Cost of Transportation	758,638.93	633,849.439	569,906.031
Cost Per Kg	1.1011	1.0956	1.1036
Optimality Gap	1.012 %	1.001 %	0.991 %

Table 6.9: Fixed number of warehouses and bottling stations (4 Bs)

	4bs_10w	**4bs_17w**	**4bs_25w**
# Warehouse	10	17	25
# Bottling Station	4	4	4
# Van	209	141	120
# Truck	13	12	12
# Tanker	5	6	6
# Van Trip	6,215	6,216	6,216
# Truck Trip	553	553	555
# Tanker Trip	230	232	230
Total Cost-Importation Cost	2,008,979.108	1,860,217.652	1,890,287.972
Total Cost of Infrastructure	928,783.818	1,023,228.309	1,131,164.870
Total Cost of Transportation	863,125.080	619,919.133	542,052.891
Cost Per Kg	1.1590	1.1317	1.1372
Optimality Gap	2.621 %	1.000%	0.938%

In Table 6.8 and Figure 6.2, it can be seen that when the model is forced to open 2 bottling stations and 10 warehouses, the total cost of transportation is more than the total cost of infrastructure by 5.445 %. This is the scenario in which the number of infrastructural units is the smallest among all the configurations tested. Since the number of bottling stations and warehouses is low, the total infrastructure cost is low. Nevertheless, opening fewer facilities directly results in an incline in the distance travelled by vehicles between the facilities. Considering the 8-hour daily working limit, as the distance between the facilities increases, the total number of vehicles used increases. Therefore, the operational cost of vehicles and their acquisition costs augment the total transportation cost.

On the other hand, when the rest of the scenarios presented in Table 6.8 and Table 6.9 are evaluated, it is seen that the total cost of infrastructure dominates the total cost of transportation. While the number of bottling stations is 2, and the number of warehouses has increased from 10 to 17, the total infrastructure cost does not increase drastically, as the Overnight cost of a warehouse is relatively

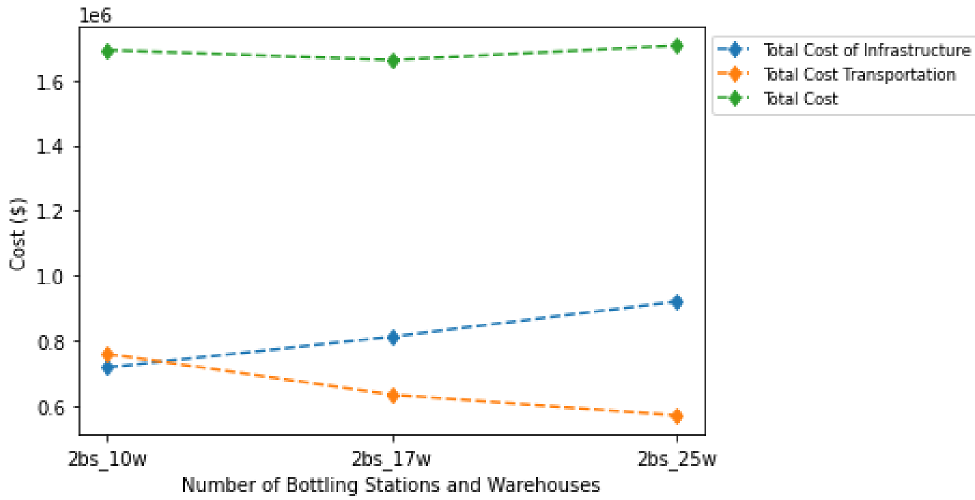


Figure 6.2: Infrastructure, transportation, and total cost of systems with 2 bottling stations: This graph represents the infrastructure, transportation and total cost of each scenario. Total cost line represents the total system cost, excluding the importation cost. Labels on the x-axis represent each scenario. For example, 2bs_10w indicates the scenario with 2 bottling stations and 10 warehouses.

low. However, the increase in the number of warehouses leads to a decrease in the number of vans as the distance between warehouses and retailers diminishes. Note that the number of tankers in this scenario is recorded as 0 even though there is a flow between the depot and bottling stations. The reason for this is the model prefers to open both of the bottling stations at the location of Bottling Station 2. In case of a different selection in the locations of bottling stations, the acquisition cost of tankers would add up to the total cost of transportation. This change would help to close the gap between the total cost of infrastructure and the total cost of transportation.

In Figure 6.3, it is seen that the difference between the total cost of infrastructure and the total cost of transportation starts to increase more. Since the Overnight cost of a bottling station is high, the total infrastructure cost is always higher than the total transportation cost when the model is forced to open 4 bottling stations. However, it should be highlighted that even if there are 4 bottling stations in scenario 4bs_10w, due to the high number of van numbers with the incline in the travelling distances, the total cost of infrastructure is only 7.069 % higher than the total cost of transportation.

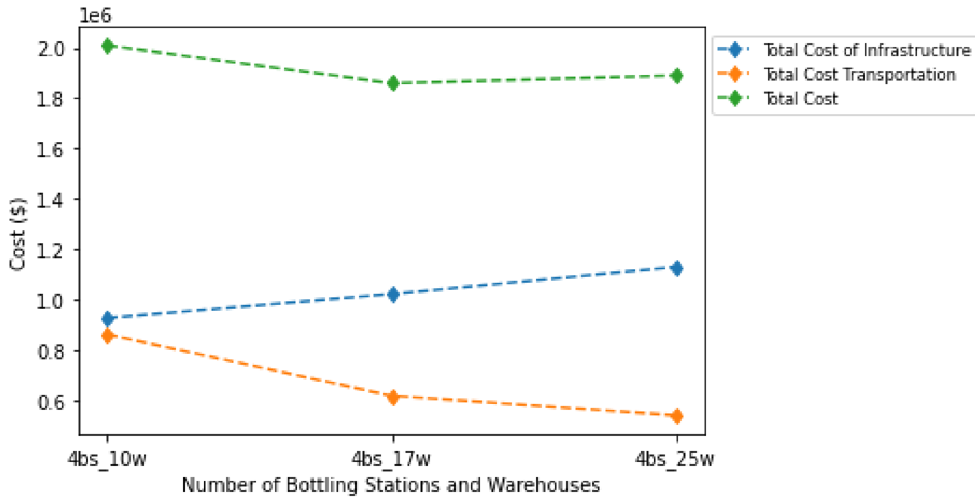


Figure 6.3: Infrastructure, transportation, and total cost of systems with 4 bottling stations: This graph represents the infrastructure, transportation and total cost of each scenario. Total cost line represents the total system cost, excluding the importation cost. Labels on the x-axis represent each scenario. For example, 4bs_10w indicates the scenario with 4 bottling stations and 10 warehouses.

By evaluating the results, it can be seen that there is indeed a trade-off between the total cost of infrastructure and the total cost of transportation. With an increase in the price of diesel fuel, the total transportation cost may worsen the total system cost. Regarding this trade-off, the decision-maker should decide on whether to invest in infrastructure or transportation units.

6.5 Main Model and Decomposed Model Comparison

In this section, we compare the outputs of the main model (Section 5.1) and the decomposed model (Section 5.2), including their performances. As in Section 6.3, the decomposed model is run for each projected demand scenario presented in Table 4.4. The outputs of the decomposed model can be found in Table 6.10. Both the first phase model (WR) and the second phase model (BW) are executed until the optimality gaps drop to 1%.

While it is ideal to solve the first and second phase models to optimality, it is observed that the optimality gaps drop to 1% in a very short time. However, beyond this point, it takes a considerably higher amount of time to solve the models to optimality. As an applied work, this thesis also aims to provide a practical decision-making framework for policymakers. Thus, it is important to obtain adequate solutions with a 1% optimality gap in a relatively short period. This allows policymakers to make timely and effective decisions without the need to wait for full optimization.

When Table 6.4 and Table 6.10 are compared, it is seen that the difference between total cost (excluding the importation cost) is not significantly high. More specifically, the total cost differences between the two models are 1.707 %, 0.168 %, 0.770 %, 0.693 %, 0.130 %, and 0.496 % from 2025 to 2030, respectively. However, it should be noted that in each case, the total cost of the decomposed model is higher than the total cost of the main model.

While the values of cost components do not alter drastically between the two models, the locations of facilities differ. In Table 6.12, the indices of bottling stations opened in each model are reported. While "M" denotes the main model, "D" stands for the decomposed model in Table 6.12. Although there are some similarities, the selection of bottling stations mostly varies. While both of the models prefer to open Bottling Station 2 in all cases (due to its location), the difference in other bottling stations indicates that the main and decomposed models are not completely equivalent. Hence, the difference between the cost components is investigated further. The cost components for the 2025 case are provided in Table 6.11.

Table 6.11: Cost components for the main and decomposed model in 2025

	M	**D**
Total Cost of Secondary Transportation	457,503.173	438,310.839
Total Cost of Primary Transportation	128,372.697	172,483.078
Total Cost of Bulk Transportation	25,194.371	15,254.439
Total Cost of Warehouse	229,383.571	242,875.641
Total Cost of Bottling Station	582,253.181	582,253.181
Total Cost of Depot	138.818	138.818
Total Cost of Cylinder	217,070.210	217,070.210
Total Cost of Importation	4,321,710.383	4,321,710.383
Total Cost-Importation Cost	1,639,916.02	1,668,386.206

It is seen that the cost of bulk transportation is higher in the main model. Since the location of the depot and the total amount of flow carried between the depot and bottling stations are the same, we can state that selecting Bottling Station 1 instead of Bottling Station 5, results fewer tanker numbers in the decomposed model. This is because the distance travelled between the depot and the bottling stations directly affects the number of tankers used in the system. Consequently, the total bulk transportation cost is lower in the decomposed model.

However, it is also observed that the total primary transportation cost of the decomposed model is higher than the total primary transportation cost of the main model. We know that the total primary transportation cost is directly affected by the locations of the bottling stations and warehouses. Hence, we can assert that when the locations of the warehouses are fixed, the decomposed model needs to increase the truck numbers and truck trip numbers in order to reach the existing warehouses from the newly opened bottling stations. At this point, the main model dominates the decomposed model as this is the main cost component that deteriorates the total system cost of the decomposed model.

Table 6.12: Bottling station indices for the main and decomposed model

2025		**2026**		**2027**		**2028**		**2029**		**2030**	
M	**D**	**M**	**D**	**M**	**D**	**M**	**D**	**M**	**D**	**M**	**D**
2	1	2	2	2	2	2	2	2	2	1	1
5	2	5	4	5	2	2	5	2	2	2	2
		6	5	6	5	5	6	5	5	4	2
										5	5

In Table 6.13, the warehouse indices that are not common in the two models for each year are presented. Overall, the decomposed model opens more warehouses than the main model. Hence, the total warehouse cost of the decomposed model is higher. Given that the warehouses do not hold any inventory, the motivation to open more warehouses is to minimize the total secondary transportation cost as much as possible. With increased number of warehouses, fewer number of vans are required due to the decline in the total travelling distance between warehouses and retailers. When the total secondary transportation costs of the two models are compared, it is indeed seen that the total secondary transportation cost of the decomposed model is lower.

Overall, we record that the total system cost of the decomposed model is higher than the total system cost of the main model, as expected. Hence, in order to obtain the configuration yielding the lowest cost, the decision-maker should prefer the main model. Nonetheless, it should also be kept in mind that the execution times of the two models vary significantly. Thus, within a limited time interval, the decision-maker may also prefer to use the decomposed model to get a sufficient solution in a small time interval.

Table 6.13: Warehouse indices for the main and decomposed model

2025		2026		2027		2028		2029		2030	
M	**D**	**M**	**D**	**M**	**D**	**M**	**D**	**M**	**D**	**M**	**D**
10	9	18	20	4	5	26	2	5	2	8	2
18	14	26	25		29		25	18	14	18	25
	20							26	20	26	
								30	25		
									29		

Table 6.4: Different demand scenarios main model outputs

	2025	**2026**	**2027**	**2028**	**2029**	**2030**
Demand	5,462,333.33	6,137,933.33	6,898,733.33	7,749,866.67	8,691,933.33	9,718,000.00
# Retailers	5,969	5,969	5,969	5,969	5,969	5,969
# Warehouse	17	18	17	17	17	18
# Bottling Station	2	3	3	3	3	4
# Van	132	131	133	135	141	140
# Truck	16	15	17	22	24	21
# Tanker	4	6	7	4	5	10
# Van Trip	6,215	6,280	6,353	6,471	6,598	6,758
# Truck Trip	553	620	699	781	876	977
# Tanker Trip	230	258	289	325	365	408
Total Cost	5,961,626.403	6,682,873.39	7,382,556.24	8,174,773.02	9,057,876.14	10,097,006.63
Importation Cost	4,321,710.383	4,852,835.73	5,450,494.41	6,119,911.75	6,860,484.09	7,666,844.64
Total Cost - Importation Cost	1,639,916.021	1,830,037.65	1,932,061.82	2,054,861.27	2,197,392.05	2,430,161.99
Total Cost of Infrastructure	811,775.569	976,547.44	1,014,315.21	1,071,729.64	1,135,246.87	1,323,625.04
Total Cost of Transportation	611,070.242	609,742.45	643,979.43	675,740.73	717,556.55	721,446.16
Total Cost of Cylinder	217,070.21	243,747.76	273,767.17	307,390.90	344,588.63	385,090.79
Cost Per Kg	1.0914	1.0888	1.0701	1.0548	1.0421	1.0390
Optimality Gap	0.785 %	1.765 %	1.242 %	0.902 %	0.503 %	0.878 %

Table 6.10: Decomposed model outputs

	2025	**2026**	**2027**	**2028**	**2029**	**2030**
Demand	5,462,333.333	6,137,933.333	6,898,733.333	7,749,866.677	8,691,933.333	9,718,000.000
# Warehouse	18	18	18	18	18	18
# Bottling Station	2	3	3	3	3	4
# Van	126	128	129	132	131	133
# Truck	22	16	22	21	28	30
# Tanker	3	6	4	8	5	6
# Van Trip	6,215	6,280	6,353	6,470	6,597	6,758
# Truck Trip	559	627	704	790	884	987
# Tanker Trip	230	258	290	325	364	408
Total Cost	5,990,096.588	6,685,958.762	7,397,542.143	8,189,104.406	9,060,733.011	10,114,709.032
Importation Cost	4,321,710.383	4,852,835.735	5,450,494.415	6,119,911.751	6,860,484.095	7,666,844.639
Total Cost-Importation Cost	1,668,386.206	1,833,123.027	1,947,047.729	2,069,192.656	2,200,248.917	2,447,864.393
Total Cost of Infrastructure	825,267.639	976,547.440	1,027,807.285	1,085,221.709	1,148,738.938	1,323,625.044
Total Cost of Transportation	626,048.356	612,827.826	645,473.272	676,580.044	706,921.347	739,148.556
Total Cost of Cylinder	217,070.210	243,747.762	273,767.172	307,390.902	344,588.631	385,090.793
Cost Per Kg	1.0966	1.0893	1.0723	1.0567	1.0424	1.0408
Execution Time (WR) (s)	157.436	285.512	268.886	259.831	369.885	336.276
Execution Time (BW) (s)	159.052	288.135	271.570	262.198	372.280	338.637
Optimality GAP (WR)	0.369 %	0.516 %	0.930 %	0.888 %	0.743 %	0.909 %
Optimality Gap (BW)	0.757 %	0.976 %	0.999 %	0.480 %	0.167 %	0.850 %
Main Model % Difference	1.707 %	0.168 %	0.770 %	0.693 %	0.130 %	0.496 %

Chapter 7

Conclusion and Future Work

In the literature, various studies have been conducted on the location-allocation problem. Different than the existing ones, in this thesis, we aim to solve a large-scale location-allocation problem formulated as a MILP, using a real-life dataset on LPG demands of rooftops in Rwanda. The main motivation behind this thesis is to design a supply chain system for LPG distribution in Rwanda, aiming to promote clean cooking strategies in developing countries.

Given the location of each rooftop and yearly LPG demand, we first analyze the data and make necessary assumptions. Then, in order to decrease the size of the problem, we propose three different methods. The first method is to run the formulated MILP model independent of time index t due to the absence of demand seasonality. The second method is to use the agglomerative hierarchical clustering-based heuristic approach to decide on the locations of retailers to be used in the index t independent mathematical model. It is seen that this approach yields fewer retailers than the village centroid approach. Finally, we propose a decomposition of the main MILP model formulated in order to decrease the execution time.

Furthermore, we compare the results of the models that use different retailer locations and different demand scenarios. In this analysis, we make comments on the selection of retailer locations and whether the existing system configuration can be expanded with the projected demand increase. In addition, considering the volatile diesel fuel prices, we conduct a sensitivity analysis to show the trade-off between the transportation cost and the infrastructure costs. Finally, we compare the results and performances of the main model and the decomposed model.

In future work, incorporating demand seasonality and uncertainty could be key to enhancing the system, particularly in ensuring the functionality of warehouses. Moreover, integrating the routing decisions of the vehicles could lead to a decrease in the investment in transportation units. Furthermore, retailer locations may be selected based on alternative criteria that account for the real life applicability, such as the proximity of the retailer's location to roads. Additionally, given the difficulties in solving large-scale optimization models due to the size and complexity of the dataset, further decomposition methods may be introduced, such as Benders Decomposition, to enhance computational efficiency. Finally, evolutionary algorithms, which are capable of exploring large and complex solution spaces through mechanisms like selection, mutation, and crossover, may also be employed to provide approximate solutions.

In conclusion, this thesis provides a supply chain design decision-making framework for LPG distribution in Rwanda, with the aim of promoting clean cooking techniques in developing countries using a real-life, large-scale dataset.

Bibliography

[1] A. Fleming, R. M. Wise, H. Hansen, and L. Sams, "The sustainable development goals: A case study," *Marine Policy*, vol. 86, pp. 94–103, 2017.

[2] D. McCollum, L. Gomez Echeverri, K. Riahi, and S. Parkinson, "Sdg7: Ensure access to affordable, reliable, sustainable and modern energy for all," 2017.

[3] B. Akbas, A. S. Kocaman, D. Nock, and P. A. Trotter, "Rural electrification: An overview of optimization methods," *Renewable and Sustainable Energy Reviews*, vol. 156, p. 111935, 2022.

[4] J. E. Adkins, V. Modi, S. Sherpa, R. Han, A. S. Kocaman, N. Zhao, C. Natali, and J. Carbajal, "A geospatial framework for electrification planning in developing countries," in *2017 IEEE Global Humanitarian Technology Conference (GHTC)*, pp. 1–10, IEEE, 2017.

[5] A. S. Kocaman, W. T. Huh, and V. Modi, "Initial layout of power distribution systems for rural electrification: A heuristic algorithm for multilevel network design," *Applied energy*, vol. 96, pp. 302–315, 2012.

[6] A. Zvoleff, A. S. Kocaman, W. T. Huh, and V. Modi, "The impact of geography on energy infrastructure costs," *Energy Policy*, vol. 37, no. 10, pp. 4066–4078, 2009.

[7] J. E. Adkins, V. Modi, S. Sherpa, R. Han, A. S. Kocaman, N. Zhao, C. Natali, and J. Carbajal, "A geospatial framework for electrification planning in developing countries," in *2017 IEEE Global Humanitarian Technology Conference (GHTC)*, pp. 1–10, 2017.

[8] S. Fobi, A. S. Kocaman, J. Taneja, and V. Modi, "A scalable framework to measure the impact of spatial heterogeneity on electrification," *Energy for Sustainable Development*, vol. 60, pp. 67–81, 2021.

[9] Ö. Karsu and A. S. Kocaman, "Towards the sustainable development goals: A bi-objective framework for electricity access," *Energy*, vol. 216, p. 119305, 2021.

[10] A. S. Kocaman, E. Ozyoruk, S. Taneja, and V. Modi, "A stochastic framework to evaluate the impact of agricultural load flexibility on the sizing of renewable energy systems," *Renewable Energy*, vol. 152, pp. 1067–1078, 2020.

[11] A. S. Kocaman and V. Modi, "Value of pumped hydro storage in a hybrid energy generation and allocation system," *Applied energy*, vol. 205, pp. 1202–1215, 2017.

[12] M. Peker, A. S. Kocaman, and B. Y. Kara, "Benefits of transmission switching and energy storage in power systems with high renewable energy penetration," *Applied Energy*, vol. 228, pp. 1182–1197, 2018.

[13] A. S. Kocaman, C. Abad, T. J. Troy, W. T. Huh, and V. Modi, "A stochastic model for a macroscale hybrid renewable energy system," *Renewable and Sustainable Energy Reviews*, vol. 54, pp. 688–703, 2016.

[14] P. Toufani, E. C. Karakoyun, E. Nadar, O. B. Fosso, and A. S. Kocaman, "Optimization of pumped hydro energy storage systems under uncertainty: A review," *Journal of Energy Storage*, vol. 73, p. 109306, 2023.

[15] G. Bolukbasi and A. S. Kocaman, "A prize collecting steiner tree approach to least cost evaluation of grid and off-grid electrification systems," *Energy*, vol. 160, pp. 536–543, 2018.

[16] World Health Organization, "Household air pollution and health." `https://www.who.int/news-room/fact-sheets/detail/household-air-pollution-and-health`, 2024. Accessed: 2024-07-02.

[17] United Nations, "The sustainable development goals report 2023." https://unstats.un.org/sdgs/report/2023/The-Sustainable-Development-Goals-Report-2023.pdf, 2023. Accessed: 2024-07-02.

[18] United Nations, "World statistics pocketbook 2023." https://unstats.un.org/unsd/publications/pocketbook/files/world-stats-pocketbook-2023.pdf, 2023. Accessed: 2024-07-02.

[19] I. Čukić, C. Kypridemos, A. W. Evans, D. Pope, and E. Puzzolo, "Towards sustainable development goal 7 "universal access to clean modern energy": National strategy in rwanda to scale clean cooking with bottled gas," *Energies*, vol. 14, no. 15, p. 4582, 2021.

[20] E. Hakizimana, U. Wali, D. Sandoval, and K. Venant, "Environmental impacts of biomass energy sources in rwanda," *Energy Environ. Eng*, vol. 7, no. 3, pp. 62–71, 2020.

[21] M. M. Islam, R. Wathore, H. Zerriffi, J. D. Marshall, R. Bailis, and A. P. Grieshop, "In-use emissions from biomass and lpg stoves measured during a large, multi-year cookstove intervention study in rural india," *Science of The Total Environment*, vol. 758, p. 143698, 2021.

[22] Rwanda Ministry of Infrastructure, "Rwanda lpg master plan, feasibility and investment report (2021)." https://static1.squarespace.com/static/5633c4c2e4b05a5c7831fbb5/t/6199d38cbd361e425e811bcf/1637471271053/Rwanda+LPG+Master+Plan%2C+Feasibility+and+Investment+Report+%282021%29.pdf, 2021. Accessed: 2024-07-02.

[23] O. Akgul, N. Shah, and L. G. Papageorgiou, "An optimisation framework for a hybrid first/second generation bioethanol supply chain," *Computers & Chemical Engineering*, vol. 42, pp. 101–114, 2012.

[24] S. Leduc, K. Natarajan, E. Dotzauer, I. McCallum, and M. Obersteiner, "Optimizing biodiesel production in india," *Applied Energy*, vol. 86, pp. S125–S131, 2009.

[25] I. Awudu and J. Zhang, "Stochastic production planning for a biofuel supply chain under demand and price uncertainties," *Applied Energy*, vol. 103, pp. 189–196, 2013.

[26] Y. Zhang and Y. Jiang, "Robust optimization on sustainable biodiesel supply chain produced from waste cooking oil under price uncertainty," *Waste Management*, vol. 60, pp. 329–339, 2017.

[27] T. J. Van Roy, "Multi-level production and distribution planning with transportation fleet optimization," *Management Science*, vol. 35, no. 12, pp. 1443–1453, 1989.

[28] F. Fölsz, C. Mészáros, and T. Rapcsák, "Distribution of gas cylinders," *European journal of operational research*, vol. 87, no. 3, pp. 613–623, 1995.

[29] I. Masudin, "An investigation of the relationship between facility location decisions, service level and distribution costs: A proposed model for indonesian lpg supply chain," *International Journal of Business and Society*, vol. 16, no. 1, 2015.

[30] J. K. Sankaran and N. S. Raghavan, "Locating and sizing plants for bottling propane in south india," *Interfaces*, vol. 27, no. 6, pp. 1–15, 1997.

[31] A. Santoso, D. N. Prayogo, and J. Parung, "Integrated supply chain network model for allocating lpg in a closed distribution system," *Automation, Control and Intelligent Systems*, vol. 3, no. 5, pp. 95–99, 2015.

[32] H. Le Blanc, H. A. Fleuren, and H. Krikke, "Redesign of a recycling system for lpg-tanks," *OR spectrum*, vol. 26, pp. 283–304, 2004.

[33] K. M. Gumte, P. D. Pantula, S. S. Miriyala, and K. Mitra, "Data driven robust optimization for handling uncertainty in supply chain planning models," *Chemical Engineering Science*, vol. 246, p. 116889, 2021.

[34] S. Sarkar, A. Pramanik, J. Maiti, and G. Reniers, "Covid-19 outbreak: A data-driven optimization model for allocation of patients," *Computers & Industrial Engineering*, vol. 161, p. 107675, 2021.

[35] J. Yang, J. Dong, and L. Hu, "A data-driven optimization-based approach for siting and sizing of electric taxi charging stations," *Transportation Research Part C: Emerging Technologies*, vol. 77, pp. 462–477, 2017.

[36] E. Yücel, F. S. Salman, B. Bozkaya, and C. Gökalp, "A data-driven optimization framework for routing mobile medical facilities," *Annals of Operations Research*, vol. 291, pp. 1077–1102, 2020.

[37] H. Panahi, F. Sabouhi, A. Bozorgi-Amiri, and S. Ghaderi, "A data-driven optimization model for renewable electricity supply chain design," *Renewable and Sustainable Energy Reviews*, vol. 202, p. 114618, 2024.

[38] J. Serra Llavona, "Mit-iit universal access lab: optimization and simulation models for clean-cooking planning," 2023.

[39] World Bank, "Rwanda - admin boundaries and villages." `https://datacatalog.worldbank.org/search/dataset/0041453/Rwanda-Admin-Boundaries-and-Villages`, 2021. Accessed: 2024-08-13.

[40] "Openstreetmap - browse map." `https://www.openstreetmap.org/\#map=4/47.18/57.48`. Accessed: 2024-08-14.

[41] "Microsoft buildings dataset." `https://planetarycomputer.microsoft.com/dataset/ms-buildings`. Accessed: 2024-08-14.

[42] J. M. V. Ntamwiza, "Econometric analysis of transport sector on economic growth in rwanda (1999-2018)," *Journal of Transportation Technologies*, vol. 10, no. 4, pp. 380–391, 2020.

[43] A. S. Kocaman, *Essays on infrastructure design and planning for clean energy systems*. Columbia University, 2014.

[44] B. Akbas and A. S. Kocaman, "A planar facility location–allocation problem with fixed and/or variable cost structures for rural electrification," *Computers & Operations Research*, vol. 154, p. 106202, 2023.

[45] S. C. Johnson, "Hierarchical clustering schemes," *Psychometrika*, vol. 32, no. 3, pp. 241–254, 1967.

[46] P. H. A. Sneath and R. R. Sokal, *Numerical Taxonomy: The Principles and Practice of Numerical Classification.* W. H. Freeman and Co., 1973.

[47] J. H. Ward Jr, "Hierarchical grouping to optimize an objective function," *Journal of the American statistical association*, vol. 58, no. 301, pp. 236–244, 1963.

[48] A. M. Jarman, "Hierarchical cluster analysis: Comparison of single linkage, complete linkage, average linkage and centroid linkage method," *Georgia Southern University*, vol. 29, 2020.

[49] W. Lexis, "V. alfred weber, ueber den standort der industrien," *Jahrbücher für Nationalökonomie und Statistik*, vol. 94, no. 1, pp. 818–826, 1910.

[50] Y. Vardi and C.-H. Zhang, "A modified weiszfeld algorithm for the fermat-weber location problem," *Mathematical Programming*, vol. 90, pp. 559–566, 2001.

[51] K. Gokbayrak and A. S. Kocaman, "A distance-limited continuous location-allocation problem for spatial planning of decentralized systems," *Computers & Operations Research*, vol. 88, pp. 15–29, 2017.

[52] S. Huber and C. Rust, "Calculate travel time and distance with openstreetmap data using the open source routing machine (osrm)," *The Stata Journal*, vol. 16, no. 2, pp. 416–423, 2016.

[53] Global Petrol Prices, "Rwanda gasoline prices." `https://www.globalpetrolprices.com/Rwanda/`, 2024. Accessed: 2024-07-02.

[54] Rwanda Utilities Regulatory Authority, "Announcement for fuel pump prices from june 2024." `https://www.rura.rw/fileadmin/publication/Announcement_for_Fuel_pump_prices_from_June_2024.pdf`, 2024. Accessed: 2024-07-02.

Appendix A

Candidate Bottling Station Locations

Table A.1: Candidate bottling station locations

Province	**District**	**Town/Area**	**Coordinate**
Kigali City	Gasabo	Kabuga/Rusororo	(-1.9569, 30.2239)
		Jabana	(-1.8735, 30.0811)
	Kicukiro	Gahanga	(-2.0380, 30.0986)
Northern	Musanze	Musanze	(-1.5148, 29.5881)
Southern	Muhanga	Muhanga	(-2.0849, 29.7495)
Eastern	Kayonza	Kayonza	(-1.9010, 30.4934)
Western	Rusizi	Kamembe	(-2.4823, 28.8966)

Appendix B

Candidate Warehouse Locations

Table B.1: Candidate warehouse locations

District	Coordinate	District	Coordinate
Bugesera	(-2.2330, 30.1589)	Karongi	(-2.1713, 29.4413)
Gatsibo	(-1.5893, 30.3979)	Ngororero	(-1.8547, 29.6325)
Kayonza	(-1.9050, 30.5102)	Nyabihu	(-1.6410, 29.5181)
Kirehe	(-2.2613, 30.6285)	Nyamasheke	(-2.3580, 29.1546)
Ngoma	(-2.1663, 30.5392)	Rubavu	(-1.6936, 29.2599)
Nyagatare	(-1.2986, 30.3295)	Rusizi	(-2.4888, 28.8958)
Rwamagana	(-1.9512, 30.4349)	Rutsiro	(-1.8954, 29.2997)
Gasabo	(-1.9523, 30.1705)	Gisagara	(-2.6069, 29.8467)
Kicukiro	(-1.9854, 30.1036)	Huye	(-2.6053, 29.7402)
Nyarugenge	(-1.9509, 30.0589)	Kamonyi	(-2.0279, 29.9021)
Burera	(-1.4609, 29.8012)	Muhanga	(-2.0431, 29.7180)
Gakenke	(-1.6718, 29.7859)	Nyamagabe	(-2.3993, 29.4811)
Gicumbi	(-1.5762, 30.0690)	Nyanza	(-2.3431, 29.7702)
Musanze	(-1.4745, 29.5693)	Nyaruguru	(-2.7136, 29.5395)
Rulindo	(-1.7515, 29.9976)	Ruhango	(-1.8568, 29.3844)

Appendix C

Main Model

$$
\begin{aligned}
\min \quad & \sum_{w \in W} (FOMw + ONw \cdot dFw) \cdot I_w \\
& + \sum_{b \in B} (FOMb + ONb \cdot dFb) \cdot I_b \\
& + \sum_{b \in B} \sum_{t \in T} (VOMb \cdot volbol_{b,t}) \\
& + \sum_{w \in W} (FOMv + ONv \cdot dFv) \cdot numV_w \\
& + \sum_{b \in B} (FOMtr + ONtr \cdot dFtr) \cdot numTr_b \\
& + (FOMta + ONta \cdot dFta) \cdot numTa \\
& + 2 \cdot \sum_{w \in W} \sum_{r \in R} \sum_{t \in T} (VOMv \cdot dist_{w,r} \cdot vanTrip_{w,r,t} \cdot 2.27) \\
& + 2 \cdot \sum_{b \in B} \sum_{w \in W} \sum_{t \in T} (VOMtr \cdot dist_{b,w} \cdot truckTrip_{b,w,t} \cdot 2.27) \\
& + 2 \cdot \sum_{b \in B} \sum_{t \in T} (VOMta \cdot dist_b \cdot tankerTrip_{b,t} \cdot 2.27) \\
& + \sum_{i \in I} (totNumFullCyl_i \cdot cylCost \cdot dFc \cdot 2) \\
& + \sum_{i \in I} (totNumFullCyl_i \cdot Creiq \cdot dFcreq \cdot 2) \\
& + \sum_{w \in W} \sum_{r \in R} \sum_{t \in T} \frac{laborC \cdot salary}{yearWh} \cdot vanTrip_{w,r,t} \cdot \frac{dist_{w,r} \cdot 2}{Sv} \\
& + \sum_{b \in B} \sum_{w \in W} \sum_{t \in T} \frac{laborC \cdot salary}{yearWh} \cdot truckTrip_{b,w,t} \cdot \frac{dist_{b,w} \cdot 2}{Str} \\
& + \sum_{b \in B} \sum_{t \in T} \frac{laborC \cdot salary}{yearWh} \cdot tankerTrip_{b,t} \cdot \frac{dist_b \cdot 2}{Sta} \\
& + \sum_{b \in B} \sum_{t \in T} f_{b,t} \cdot priceI \\
& + (licenseCv \cdot dFvl) + (licenceCtr \cdot dFtrl) + (licenseCta \cdot dFtal) \\
& + (Wsal \cdot dFwsal) + (InsB \cdot dFb) + (OpLb \cdot dFbop) + (InsD \cdot dFd) \\
& + (OpLd \cdot dFdop) + (ImL \cdot dFil)
\end{aligned}
$$

$$
\begin{aligned}
\text{s.t.} \quad & \sum_{b \in B} f_{b,t} \leq Capd \quad \forall t \in T \\
& \frac{f_{b,t}}{CapTa} \leq tankerTrip_{b,t} \quad \forall b \in B, t \in T \\
& \sum_{b \in B} tankerTrip_{b,t} \cdot \frac{dist_b \cdot 2}{Sta} \leq numTa \cdot dayWh \cdot monthWd \quad \forall t \in T \\
& \sum_{i \in I} \sum_{w \in W} f_{i,b,w,t} \leq I_b \cdot BigM \quad \forall b \in B, t \in T \\
& \sum_{t \in T} volbol_{b,t} \leq BotCap \cdot I_b \quad \forall b \in B \\
& volBottled_{b,t} = \sum_{i \in I} numCBottled_{i,b,t} \cdot cylinder\_kg \quad \forall b \in B, t \in T \\
& bulkInv_{b,t} = bulkInv_{b,t-1} + f_{b,t} - volbol_{b,t} \quad \forall b \in B, t \in \{2, ..., T\} \\
& bulkInv_{b,t} \leq BulkCap \quad \forall b \in B, t \in T \\
& Inv_{i,b,t} = Inv_{i,b,t-1} + numCbot_{i,b,t} - \sum_{w \in W} f_{i,b,w,t} \quad \forall i \in I, b \in B, t \in \{2, .., T\} \\
& Inv_{i,b,t} \leq CylCap \quad \forall i \in I, b \in B, t \in T \\
& \sum_{w \in W} truckTrip_{b,w,t} \cdot \frac{dist_{b,w} \cdot 2}{Str} \leq numTr_b \cdot dayWh \cdot monthWd \quad \forall b \in B, t \in T \\
& \sum_{i \in I} \frac{f_{i,b,w,t} \cdot w_i}{CapTr} \leq truckTrip_{b,w,t} \quad \forall b \in B, w \in W, t \in T \\
& \sum_{r \in R} \sum_{i \in I} f_{i,w,r,t} \leq I_w \cdot BigM \quad \forall w \in W, t \in T \\
& Inv_{i,w,t} = Inv_{i,w,t-1} + \sum_{b \in B} f_{i,b,w,t} - \sum_{r \in R} f_{i,w,r,t} \quad \forall i \in I, w \in W, t \in \{2, .., T\} \\
& \sum_{i \in I} Inv_{i,w,t} \cdot cylinder\_kg \leq Cap_w \quad \forall w \in W, t \in T \\
& \sum_{r \in R} vanTrip_{w,r,t} \cdot \frac{dist_{w,r} \cdot 2}{Svan} \leq numV_w \cdot dayWh \cdot monthWd \quad \forall w \in W, t \in T \\
& \sum_{i \in I} \frac{f_{i,w,r,t} \cdot w_i}{CapV} \leq vanTrip_{w,r,t} \quad \forall w \in W, r \in R, t \in T
\end{aligned}
$$

$$\sum_{w \in W} f_{i,w,r,t} \geq \frac{dem_{i,r,t}}{w_i} \quad \forall i \in I, r \in R, t \in T$$

$$totNumFullCyl_i = \sum_{w \in W} \sum_{r \in R} f_{i,w,r,t} + \sum_{w \in W} \sum_{b \in B} f_{i,b,w,t}$$

$$+ \sum_{w \in W} Inv_{i,w,t} + \sum_{b \in B} Inv_{i,b,t} \quad \forall i \in I, b \in B, w \in W, t \in T$$

$$numTa, tankerTrip_{b,t} \in \mathbb{Z}^+ \cup \{0\} \quad \forall b \in B, t \in T$$

$$Capd, f_{b,t} \geq 0 \quad \forall b \in B, t \in T$$

$$I_b, numTr_b, truckTrip_{b,w,t}, f_{i,b,w,t} \in \mathbb{Z}^+ \cup \{0\} \quad \forall i \in I, b \in B, w \in W, t \in T$$

$$Cap_b, volbol_{b,t}, numCbot_{b,t}, totNumFullCyl_i, Inv_{i,b,t}, bulkInv_{b,t} \geq 0 \quad \forall i \in I, b \in B, t \in T$$

$$I_w \in \{0,1\} \quad \forall w \in W$$

$$numV_w, vanTrip_{w,r,t}, f_{i,w,r,t} \in \mathbb{Z}^+ \cup \{0\} \quad \forall i \in I, w \in W, r \in R, t \in T$$

$$Cap_w, Inv_{i,w,r,t} \geq 0 \quad \forall i \in I, w \in W, r \in R, t \in T$$