Faculty Economics and Business

Optimization of Cold Storage Locations in Madhya Pradesh to Reduce Tomato Losses

Nick Koelemeij

14917327

MSc in Data Science and Business Analytics

Date of final version: 15th July, 2024
Supervisor: Prof. dr. ir. D. den Hertog
Second reader: ir. B. van Veggel
Advisor: MSc. P. Krishnan Krishnakumari

Abstract
Currently up till 30%, or 600.000 tonnes, of the produced tomatoes in Madhya Pradesh, a state of India, goes to waste. Tomatoes are highly perishable and can be kept fresh up till seven days, but with proper storage the shelf life goes up till three weeks. Cold storages could ensure a longer shelf life and therefore reduce waste. An optimization model is to developed to optimize the locations of these cold storages. This optimization model is a Hub Location problem and is solved by applying Mixed-Integer Linear Optimization. Three variations of this model are developed, to gain different insights. The objective of these models is to minimize the total hectare-kilometer and as a result to combat tomato waste.

Statement of originality

This document is written by student [Nick Koelemeij 14917327] who declares to take full responsibility for the contents of this document. I declare that the text and the work presented in this document is original and that no sources other than those mentioned in the text and its references have been used in creating it. I have not used generative AI (such as ChatGPT) to generate or rewrite text. The Faculty of Economics and Business is responsible solely for the supervision of completion of the work, not for the contents.

Contents

Chapter 1

Introduction

1.1 General Introduction

India produces 21,181,000 tonnes of tomatoes annually and ranks number two globally in terms of tomato production according to AtlasBig (2023). The state Madhya Pradesh is the second largest producer of tomatoes in India, with a share of 12% (Board (2020)). Madhya Pradesh is located in the center of India and is able to produce tomatoes all year round. However, Madhya Pradesh experience tomato losses up till 30% annually according to Agarwal et al. (2021). Research has shown that the use of cold storages could reduce the tomato losses, but there is a lack of availability of these storages in Madhya Pradesh (CNBCTV18 (2023)). Currently, it is known that cold storages could be part of a solution to this problem, but the optimal locations of the storages are not determined yet. Therefore, this thesis is focused at the following subject:

Optimization of Cold Storage Locations in Madhya Pradesh to Reduce Tomato Losses

1.2 Objective of This Thesis

The objective of this thesis is to find optimal locations for cold storages in Madhya Pradesh. The optimal locations are found by applying mathematical models, using Mixed-Integer Linear Optimization. Three different models are established, all with a similar objective function but different constraints. The goal of using three different models is to provide varied insights. The goal is to minimize the total hectare-kilometer, while making use of cold storages in the tomato supply chain. The models are developed for the state Madhya Pradesh, however the methodology is applicable to other states with little modification. This thesis aligns with government initiatives such as National Mission on Food Processing, Pradhan Mantri

Fasal Bima Yojana, National Agriculture Market (eNAM), and PM Kisan Sampada Yojana. This thesis ensures synergy to make sure tomato waste is reduced in Madhya Pradesh.

1.3 Structure of This Thesis

For this thesis the following structure holds. Literature review and background information are explained in chapter 2. The literature review is based on existing hub locations problems and previous research regarding tomato supply chains. Mathematical models and methods to obtain the desired results can be found in chapter 3. Followed by chapter 4 where a general overview of the used datasets is provided. Also the preproccesing of these datasets is explained in detail. in chapter 5 a clear overview of the outcomes, both visually and numerically is shown. The thesis ends with chapter 6, where a brief summary about the thesis is provided including the most significant results.

Chapter 2

Background and Literature Review

2.1 Background Information

Tomatoes are one of the most important vegetables in the world, for multiple reasons. Tomatoes are relatively fast growing crops, with a high return which makes it attractive to produce. Besides this advantage, tomatoes are rich of vitamins, minerals and much more (Collins et al. (2022)). Almost all tomatoes in India are sold unprocessed and sold in their own country. In percentages this means that about 1% of all tomatoes are being further processed before. Further processing includes creating ketchup, tomato puree and tomato juice according to van Dam et al. (2005). India exports 1-2% of all tomatoes to other countries (Vala et al. (2021)).

Figure 2.1: Map of India, with the state of Madhya Pradesh being highlighted in blue.

In India there are two broad agricultural seasons called Kharif and Rabi. This season correspondents to the start of the monsoon season, when the seeds are sown. This will happen in the months May till July, where harvesting typically happens around the month of October. The season Kharif is warmer with more rainfall (Madhukar et al. (2022)). Rabi is the more favourable season for producing tomatoes. Tomatoes grow better in slightly cooler temperatures, which occur in the Rabi season.

There is a large difference in the production amounts of farmers between the seasons. Table 2.1 shows that the production in Rabi is almost triple the amount of the season Kharif (Board (2020)). Farmers tend to focus on the Rabi season, since the products most likely will be better which means the tomatoes sell for a higher price.

-	**2018-19**	**2019-20**
Kharif	51.04	49.62
Rabi	139.03	154.86

Table 2.1: Production of tomatoes in Madhya Pradesh (in tonnes)

2.1.1 Current Supply Chain

The current supply chain involves many steps between farmer and customer. After harvesting, the basic steps like sorting and cleaning are performed. Right after these steps the tomatoes are transported from the farmer to a market. At this market an agent will buy tomatoes, which sells it again to a wholesaler. This wholesaler transports the tomatoes to a different region and sells the products to retailers. Finally customers are buying from the retailers (Mohan et al. (2023)). In total tomatoes are loaded and unloaded three times, which leads to a reduced shelf life and physical tomato losses.

2.1.2 Loss of Tomato

Madhya Pradesh struggles with 12.9% tomato losses post harvest on average. One of the reasons has to do with poor sanitation and storaging the crops. There are recommended practices for post-harvest operations, which are often not utilized. These practices are shown in Figure 2.2.

Moreover, there are only uses of short-term storages. On the contrary, the recommended practice is to use either short-term storage or longer-term storage. The longer-term storages are not available in the region. Tomatoes last at maximum one week without the use of cold storages. When using cold storages the shelf life goes all the way up to three weeks according to for Indian Knowledge Systems (2014).

S.No.	Recommended Practice	Actual Practice
A	Harvesting	
1	Harvesting with clippers	Manually (Twisting)
2	Collection in non -sharp edge container	Plastic Crates
3	Harvest time -Early or late hours of day	All day (as per labours working hours)
4	Harvesting stage a. Longer shelf life- Matured green state b. Shorter shelf life- Ripen stage	Producers targeting distant markets harvest their tomatoes in a matured green state For domestic markets ripen tomat o is preferred
B	Precooling, Cleaning & Disinfecting	
1	Dipping fruits in cold water (hydro cooling) mixed with disinfectants such as thiabendazole and sodium hypochlorite to remove field heat and infection	No
2	Sodium hypochlorite solution to reduce fungal infection	No
C	Grading and Sorting	
1	Sorting- Removal of rotten, damaged, or diseased fruits from the healthy and clean ones	Yes
2	Grading- Categorizing on the basis of color, size, stage of maturity, or degree of ripening.	Partially Traditional eye judging method
D	Packaging	
1	Wooden crates, cardboard boxes, plastic crates with shredded paper cushion	Plastic crates with Newspaper base
E	Storage	
1	Short-term storage (up to a week) - ambient conditions	Only short -term storage is done as cold storage facility is not available
2	Longer-term storage - 10–15°C and 85–95% relative humidity	

Figure 2.2: Post-harvest practices according to Solidaridad (2022)

Due to climate change, India had to cope with multiple disease epidemics which is still increasing. This is one of the reasons that India has a relative low production per hectare (Ritchie et al. (2023)). One of the epidemics is the fungus called Alternaria solani, this is spread all accross India Vennila et al. (2020).

2.2 Literature Review

There are various studies performed on the subject of mitigating tomato losses in India. Besides making use of cold storages, polyhouses are also an option to reduce tomato waste. By utilizing polyhouses, the crops are better protected against pest attacks and the cycle is more continuous. Less pesticides are required, which makes exporting tomatoes easier. The tomatoes can be transported easier since the products meet foreign requirements regarding the use of pesticides. Moreover, small factories to further process tomatoes are welcomed. Since, it reduces tomato waste when there is a surplus in production, according to Gulati et al. (2022) The study performed by Sibomana et al. (2016) focused on the mitigation of tomato losses in South-Africa. This study explained there is a lot of margin in the postharvest handling of tomatoes. Firstly, the estimation of the tomato quality should be paid more attention to. Since an accurate estimation of tomato quality would lead to a better estimation to the shelf life. Increasing the accuracy of the estimation would be improved by universal grading conditions and better education regarding tomato farming. The shelf life can be estimated by the temperature and humidity conditions during the whole supply chain. Then storage management can be improved with concepts like First Expired First Out (FEFO) or Least Shelf First Out (LSFO).

Hub locations problems occur often in supply chain management and transportation networks according to Alumur et al. (2021). In fact, the hub location problem is an extension of the facility location problem. The extension consists of the hub locations which connect two destinations (Farahani et al. (2013)). Another study developed a hub location model where the perishability of products is taken into account (Grine et al. (2022)). To include the product quality, a quality function is added to the model. So this model is still a hub location model, with added product quality which makes the model more realistic.

Chapter 3

Methodology

This model is a Mixed-Integer Linear Optimization (MILO) model, since there are continuous and binary variables as can be seen in subsection 3.0.1. This is one of the reasons that MILO offers flexibility in the variables. These variables make it possible to have discrete solutions, like in these models, whether to open a storage or not. MILO is widely used in the supply chain management as can be seen in the book written by Postek et al. (2024).

3.0.1 Notation

As described in section 4.2, three datasets are retrieved. Multiple parameters are used for the models; distance matrices, production, demand and two hyperparameters for the number of existing storages and new storages. Five decision variables are created which all can be found in Table 3.1.

Notation	
Sets	Definition
S	set of storage indexes, where $(s = 1, ..., N)$ refers to all existing storages and indexes $(s = N+1, ..., M)$ corresponds to all potential storage locations
F	index set of farmers
M	index set of markets
Parameters	Definition
c_{fs}	distance from farmer f to storage s, to predetermined market m
d_{fs}	distance from farmer f to storage s
e_{ms}	distance from market m to storage s
p_f	amount of production of farmer f
q_m	demand of market m
N	number of existing storage locations
U	maximum number of new storages planned to be opened
Variables	Definition
y_s	$\begin{cases} 1 & \text{if storage } s \text{ is opened} \\ 0 & \text{else} \end{cases}$
x_{fs}	amount of production of farmer f served by storage s
w_{ms}	amount of demand of market m supplied by storage s
u_{fs}	$\begin{cases} 1 & \text{if farmer } f \text{ is supplied by storage } s \\ 0 & \text{else} \end{cases}$
v_{ms}	$\begin{cases} 1 & \text{if market } m \text{ is supplied by storage } s \\ 0 & \text{else} \end{cases}$

Table 3.1: All sets, parameters and variables including their definitions that are required for the model.

3.1 Models

This section will explain the details of all mathematical models used in this thesis. The objective of every model is to minimize the total weighted distance, expressed in the unit hectare-kilometer. The markets and farmers can utilize an unlimited amount of storages, this model is found in subsection 3.1.1. On the contrary, the second model in subsection 3.1.2 restricts farmers and markets to only use one storage at maximum. The last model uses a totally different approach. The markets which are supplied by farmers are predetermined, shown in subsection 3.1.3.

3.1.1 Basic Hub Location

$$\min \quad \sum_{s\in S}\sum_{f\in F} d_{fs}x_{fs} + \sum_{s\in S}\sum_{m\in M} e_{ms}w_{ms} \tag{3.1}$$

$$\sum_{s=N+1}^{S} y_s \leq U \tag{3.2}$$

$$\sum_{f\in F} x_{fs} \leq \sum_{f\in F} p_f y_s \quad \forall s \in S \tag{3.3}$$

$$\sum_{m\in M} w_{ms} \leq \sum_{m\in M} q_m y_s \quad \forall s \in S \tag{3.4}$$

$$\sum_{s\in S} x_{fs} = p_f \quad \forall f \in F \tag{3.5}$$

$$\sum_{s\in S} w_{ms} = q_m \quad \forall m \in M \tag{3.6}$$

$$\sum_{f\in F} x_{fs} = \sum_{m\in M} w_{ms} \quad \forall s \in S \tag{3.7}$$

$$y_s \in \{0,1\} \quad \forall s \in S \tag{3.8}$$

$$w_{ms}, x_{fs} \geq 0 \quad \forall f \in F, s \in S, m \in M \tag{3.9}$$

This model focuses on the minimization of the total hectare-kilometer between farmer and market, while making use of a cold storage facility. Therefore the objective function is split into two separate parts. The first part consists of the weighted distance between farmer f and storage s, where the weight is the production p_f. The second part focuses on the weighted distance between storage s and market m with weight demand d_m, shown in eq. (3.1)

As explained in section 4.2, there are already existing cold storage locations. Besides these locations, new storages can be opened. These locations match the locations of the markets. Furthermore eq. (3.2) sums over the potential storage locations stored in set S. Since existing storages are already opened, the summation starts at $N+1$, where N are the existing locations. To make sure no more than U locations are opened, an inequality sign is used between U and the summation over y_s.

Crucially is that farmers and markets make use of storage s only when this specific location is opened. For this reason eq. (3.3) and eq. (3.4) are added to the model. When binary variable y_s becomes 0, the left hand side should also become 0 for both equations. In words this means when storage s is not opened, no production from variable x_{fs} can be assigned to this storage. Similar situation holds for assigning demand from variable w_{ms}. When y_s is equal to 1 it means storage s is opened and the left hand side can supply the production to the storage. Similar explanation holds for eq. (3.4).

Farmer f has production p_f and every market m has demand q_m. All farmers and markets should be served to obtain the goal of reducing losses of tomatoes. Therefore eq. (3.5) and eq. (3.6) are added. The left hand side of the constraint represents the total production of farmer f served by storages s. To obtain the total production which is served by all the storages, the summation loops over all storages in set S. To ensure no production is unassigned, the left hand side should equal the total production p_f. Similar explanation holds for eq. (3.6).

Moreover, eq. (3.7) maintains the flow balance. In theory this means that all products that storage s receives from farmers f, should be sent to markets m. This means that storage s is a perfect intermediary since there is no accumulation or shortage at any of the storages.

Furthermore, eq. (3.8) makes sure that variable y_s is binary thus takes value 0 or 1 for every storage s in set S. Finally, variables w_{ms} and x_{fs} should be nonnegative otherwise the production and demand can be negative which is impossible. Table 3.2 presents the size of the model, for the decision variables and the constraints.

Type	**Size**
Decision variables	3375
Constraints	210

Table 3.2: The sizes for both the decisions variables and constraints, for the Basic Hub Location model.

3.1.2 Hub Location (single use storage)

$$\min \quad \sum_{s\in S}\sum_{f\in F} d_{fs}p_f u_{fs} + \sum_{s\in S}\sum_{m\in M} e_{ms}q_m v_{ms} \tag{3.10}$$

$$\sum_{s=N+1}^{S} y_s \leq U \tag{3.11}$$

$$\sum_{f\in F} u_{fs} \leq |F| y_s \quad \forall s \in S \tag{3.12}$$

$$\sum_{m\in M} v_{ms} \leq |M| y_s \quad \forall s \in S \tag{3.13}$$

$$\sum_{s\in S} u_{fs} = 1 \quad \forall f \in F \tag{3.14}$$

$$\sum_{s\in S} v_{ms} = 1 \quad \forall m \in M \tag{3.15}$$

$$y_s, u_{fs}, v_{ms} \in \{0,1\} \quad \forall s \in S, f \in F, m \in M \tag{3.16}$$

The second model restricts farmers and markets to only make use of a single storage. The objective function displayed in eq. (3.10) is similar to the one of the basic hub location model which is found in section 3.1.1. However, since variables u_{fs} and variable v_{ms} do not contain the amount of production and amount of demand, it should be added. So that is why the objective function contains parameters p_f and q_m.

Furthermore, eq. (3.11) restricts the amount of newly opened storages. The two constraints eq. (3.12) and eq. (3.13) make sure that only markets that are opened can be utilized. By using the absolute signs for F and M, the model makes it impossible to use more storages than there are opened.

This largest difference between this model and the model described in subsection 3.1.1 is eq. (3.14) and eq. (3.15). The left hand side of the equations sums over all the used storages. However, the right hand side constraints the model to utilize one storage in total.

Last eq. (3.16) lets the variables y_s, u_{fs} and v_{ms} be binary. Table 3.3 presents the size of the model, for the decision variables and the constraints.

Type	**Size**
Decision variables	3375
Constraints	165

Table 3.3: The sizes for both the decisions variables and constraints, for the Hub Location (single use storage) model.

3.1.3 Hub Location (predetermined markets)

$$\min \quad \sum_{f\in F}\sum_{s\in S} c_{fs} p_f u_{fs} \tag{3.17}$$

$$\sum_{s=N+1}^{S} y_s \leq U \tag{3.18}$$

$$\sum_{f\in F} u_{fs} \leq |F| y_s \quad \forall s \in S \tag{3.19}$$

$$\sum_{s\in S} u_{fs} = 1 \quad \forall f \in F \tag{3.20}$$

$$u_{fs}, y_s \in \{0,1\} \quad \forall s \in S, f \in F \tag{3.21}$$

The third model uses a totally different way of calculating the total distance. The distance matrix for this model contains the all distances between every market and every farmer, using any storage. The objective function shown in eq. (3.17) only exists of one part, since the distances between market and storage are already taken into account. Therefore the distance c_{fs} is multiplied with the production p_f and decision variable u_{fs}.

The constraint in eq. (3.18) makes sure that no more than U new storages are opened. Moreover, eq. (3.19) works similar as the constraint in subsection 3.1.2, where the aim is to only use storages that are opened. Similar situation for eq. (3.20), where only one storage can be used for every farmer f. Last eq. (3.21) makes sure the decision variables y_s and u_{fs} are binary. Table 3.4 presents the size of the model, for the decision variables and the constraints.

Type	**Size**
Decision variables	1710
Constraints	83

Table 3.4: The sizes for both the decisions variables and constraints, for the Hub Location (predetermined markets) model.

Chapter 4

Data

This section will provide explanation about the datasets which are used for the modelling part. section 4.1 shows the supply chain which is used in this thesis. This section explains what players are involved and will provide a detailed overview of each stage. section 4.2 focuses on the introduction of the datasets and the background of the data. All variables contain a description and have a snippet of the dataset. section 4.3 puts more focuses how the data is preprocessed. To obtain a general overview of the data, section 4.4 provides key insights in the data.

4.1 Supply Chain in This Thesis

For this thesis, the following supply chain is displayed in Figure 4.1. One assumption is that possible locations for new storages match the locatios of the markets. Since tomatoes are sold at the markets, the decision has been made to built storages at these locations.

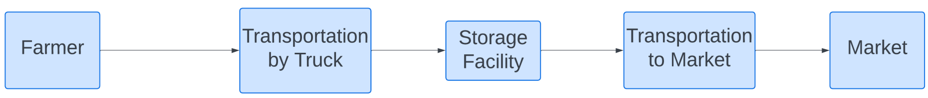


Figure 4.1: Adopted tomato supply chain in Madhya Pradesh for this thesis. The storage facility is either located at an existing location or a newly opened location.

4.2 Data Description

4.2.1 Data Sources

Multiple datasets are used in this research, data regarding the farmers, markets, storages and prices of tomatoes. The source for the data regarding farmers is ICRISAT, which is an District Level Databse (DLD) for Indian agriculture. Data about markets is coming from eNAM (electronic National Agriculture Market), where the exact addresses of all markets are shared. For the cold storages the AgriExchange is used. This is a collaboration between UNCTAT and Ministry of Agriculture.

4.2.2 Datasets

The dataset regarding farmers originates from the year 2017. One important note is that the name of any farmer corresponds to an aggregation of a number of small farmers in the same area. This works for this research, since there is no available data on small farmers specifically. The original dataset did not include any coordinates. So, these coordinates had to be retrieved to obtain an accurate location. How these coordinates are obtained is explained in subsection 4.3.2. A summary of this dataset is displayed in Table 4.1. Where Table 7.1 provides a snippet of the dataset.

Variables	**Description**
Year	Corresponding year of the obtained data
State Name	Name of the state where the farmer is located
Name	Specific name of farmer
Surface (Ha)	Surface in hectare for aggregated farmer
Latitude	Latitude for the farmer
Longitude	Longitude for the farmer

Table 4.1: Variables in the farmers dataset.

For the markets dataset, a similar dataset is available. The state Madhya Pradesh is is separated into multiple districts, and again each district can have multiple markets. Like the dataset of farmers, the coordinates are not provided. In subsection 4.3.2, an explanation about retrieving coordinates will follow.

This dataset also determines the demand per market. How the demand per market is found will be explained in subsection 4.3.3. To determine the demand, four extra columns have been added. A detailed description of the dataset can be found in Table 4.2 with a snippet in Table 7.2.

Variables	Description
Sr. No.	The larger state in India
State	State in Madhya Pradesh
District	District in the state
Name	Name of the market
Latitude	Latitude for the market
Longitude	Longitude for the market
Pop5km	Population in a radius of 5 kilometer
Per5km	Percentage of population in this radius of 5 kilometer relative to the total
Pop10km	Population in a radius of 10 kilometer
Per10km	Percentage of population in this radius of 10 kilometer relative to the total

Table 4.2: Variables in the markets dataset.

Furthermore, a dataset consisting of cold storages is retrieved. As stated in subsection 2.1.2, cold storages are crucial to prevent tomato losses. Table 7.3 shows a snippet of the cold storages dataset, Table 4.3 explains the present variables.

Notable are the different sectors and stored products variables. Furthermore, subsection 4.3.1 explains in detail how these variables are handled in the process.

Variables	Description
Name	Name of the cold storage
Capacity	Storage capacity in metric
Sector	Private, public or cooperative sector
Products stored	Products that can be stored in this cold storage
Latitude	Latitude for the cold storage
Longitude	Longitude for the cold storage

Table 4.3: Variables in the cold storages dataset.

4.3 Data Preprocessing

After receiving all the datasets, the initial task was to start with cleaning the data. Followed by subsection 4.3.1 which explains the first steps of data cleaning that are crucial for further processing the data. Followed by another necessary step where coordinates are being retrieved. Detailed in subsection 4.3.2, how exactly addresses are transformed to accurate coordinates. Finally, the processing of data to obtain demand for each market. Finally, subsection 4.3.3 focuses on this part of the datasets.

4.3.1 Data Cleaning

The first step of data cleaning is removing all the invalid locations. Invalid locations are coordinates which are situated outside the state of Madhya Pradesh. The next step is to only keep exisiting cold storage locations which are available for this thesis. So, cold storage locations should satisfy these requirements. The sector should either be public or cooperative, due to pricing, since private storages are expensive. Furthermore, tomatoes should be able to storage in these cold storages. Most of the available storages are focused on storaging potatoes, which are not useful for this thesis.

4.3.2 Preprocessing

The raw dataset consists of addresses, without specific coordinates. The first step was to retrieve the exact latitude and longitude of these addresses. By using GeoCoding, the most exact locations are retrieved.

Part of data preprocessing is preparing various distance matrices. For every distance matrix, the goal is to obtain the distance for every possible combination. In this thesis there are four different distance matrices required. Distance between every farmer and every storage, distance between every farmer and every market, distance between every storage and every market and finally distance between farmer and predetermined market using every storage. For obtaining all distances, the import googlemaps is used. One important note is that every distance is the driving distance in kilometers.

4.3.3 Demand

There is no data available regarding demand in Madhya Pradesh. For this reason assumptions about demand should be made by using population in the area of the markets. Data from WorldPop is used to extract the population in a certain radius from the provided coordinates Figure 4.2. The population from radii 5 kilometer and 10 kilometer are being retrieved (Bondarenko M. and Tatem (2020)). Next is normalizing the data, so the demand for every market is expressed in a percentage. Then this percentage is multiplied by the total production.

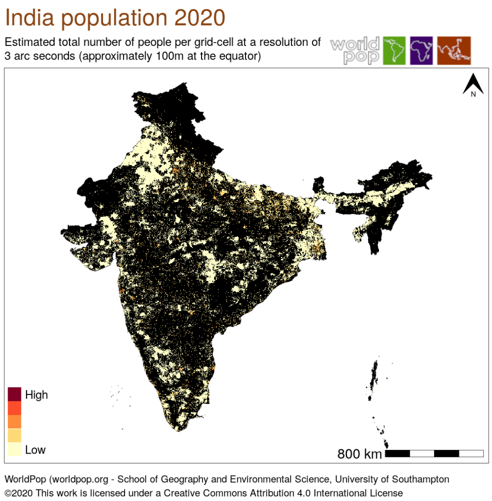
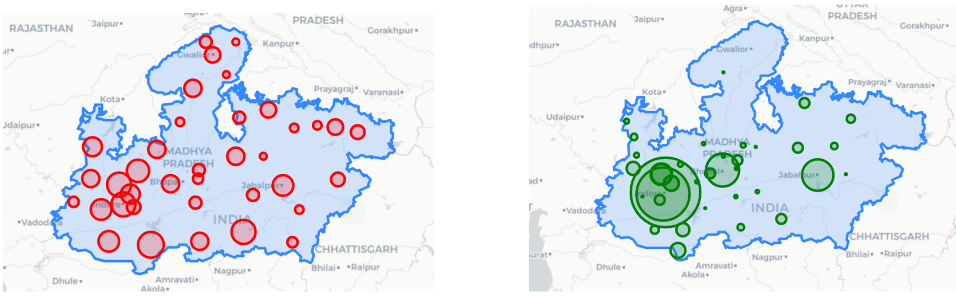


Figure 4.2: Figure from WorldPop which allows to extract population numbers.

4.4 Data Insights

To provide some key insights regarding the data, Table 4.4 summarizes key numbers regarding datasets. Figure 4.3 displays production of the farmers and demand of the markets. Figure 4.3a shows the distribution of farmers over the state of Madhya Pradesh. Moreover, the sizes of the circles indicate the amount of production.

Figure 4.3b displays the distribution of markets. Where a larger portion of the markets is located in the western part of Madhya Pradesh. Also, the percentage of demand varies widely. There exists multiple heavy demanding markets, where others almost have no impact on the demand. Sizes for farmers are more equally distributed.



(a) Farmers with their relative production

(b) Markets with their relative demand

Figure 4.3: Farmers and markets displayed on the state of Madhya Pradesh, where the sizes of the circles indicate relative production/demand

Type	**Size**
Farmers	37
Markets	37
Existing storage locations	8
All potential storage locations	45

Table 4.4: Key numbers regarding datasets.

Chapter 5

Results

5.1 Outcomes Models and Method

5.1.1 Basic Hub Location

The first step was running this model for different values of U, where these values are plotted in Figure 5.1. All objective values can be found in Table 8.1, the corresponding running time is plotted in Figure 8.1. The running time largely increases when U grows. There is a tradeoff between the number of newly opened storages and the objective value. For this model the value three is chosen, since the added value of a new storage reduces. This results in an total of 68201 hectare-kilometer. One important note is the used input for demand. Both demand with radii 5 kilometer and 10 kilometer are tested, where there was no difference in outcome.

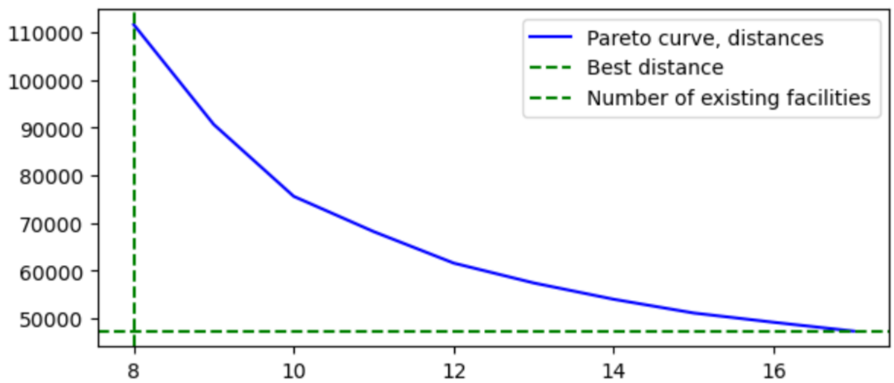
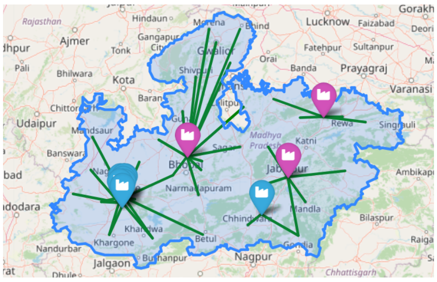
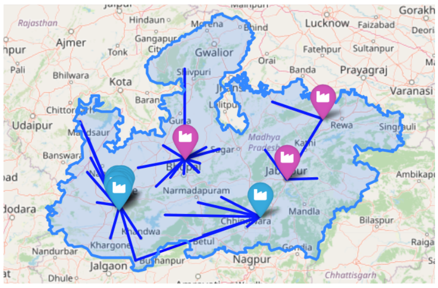


Figure 5.1: Pareto curve for the Basic Hub Location model. The y-axis contains the objective values, which indicates the total hectare-kilometer. The x-axis contain the number of storages (existing + new)

Figure 5.2 consists of two figures which indicate the used storages and all their connections. This model allows farmers and markets to make use of multiple storages, and this situation occurs various times. Figure 8.2 and Figure 8.3 indeed show that farmers occasionally make use of two or three storages. Markets only use one or two storages maximum.



(a) Each green line indicates a connections between a farmer and a storage.



(b) Each blue line indicates a connection between a market and a storage.

Figure 5.2: Two figures displaying all connections from any storage, where $U = 3$, which is the optimal solution for this model.

Table 5.1 displays the percentages of served markets and storages per number of newly opened storages in a specific radius. Again there is a tradeoff between the number of opened storages and the coverage in percentages. When comparing the results for $U = 6$ with $U = 3$, something caught the eye. While adding six additional storages compared to three additional storages, the coverage does not seem to increase significantly. Especially for the radii 150 and 200 kilometer, there is a relatively small increase in coverage. The optimal locations for these storages are found for the values $U = 0$, $U = 1$, $U = 3$ and $U = 6$. These locations are plotted in Figure 5.7. One important note is that the this model is optimizing for the total hectare-kilometer, and not for the coverage. However, the coverage does give key insights in the practical availability of cold storage facilities for markets and farmers.

Radius (km)	U=0	U = 1	U = 3	U = 6
100	20.27%	33.78%	45.95%	60.81%
150	36.49%	55.41%	68.92%	74.32%
200	51.35%	70.27%	79.73%	83.78%

Table 5.1: This table shows the percentages of locations (markets or farmers) that are in a specific radius of a storage location. These percentages are calculated based on the values $U = 0$, $U = 1$, $U = 3$, $U = 6$ and for the radii 100, 150 and 200 kilometer

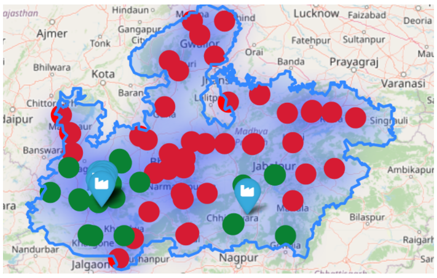


Figure 5.3: Number of new storages equals zero

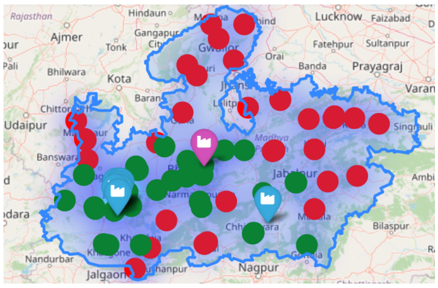


Figure 5.4: Number of new storages equals one

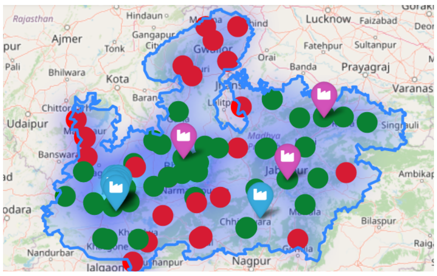


Figure 5.5: Number of new storages equals three

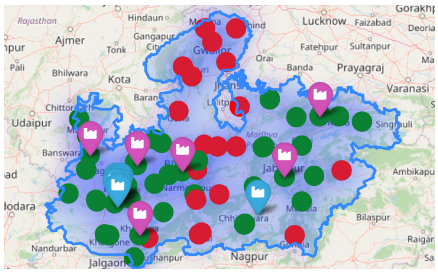


Figure 5.6: Number of new storages equals six

Figure 5.7: These figures plot the optimal locations for cold storages, where the blue markers represent existing storages and the purple markers represent the new storage locations. Locations (market or farmer) that are within a reach of 150 km of a storage location are marked green. Locations not within a reach of 150 km are marked red.

5.1.2 Hub Location (single use storage)

Firstly, this hub location (single use storage) model is executed for multiple values of U. The results are plotted in Figure 5.8, all outcomes can be found in Table 8.2 with corresponding running times in Figure 8.4. The solver used for all the models is GLPK. The chosen value for U is three, since the added value after this reduces significantly. This results in a total of 64058 hectare-kilometer. The opened storages are displayed in Figure 5.9.

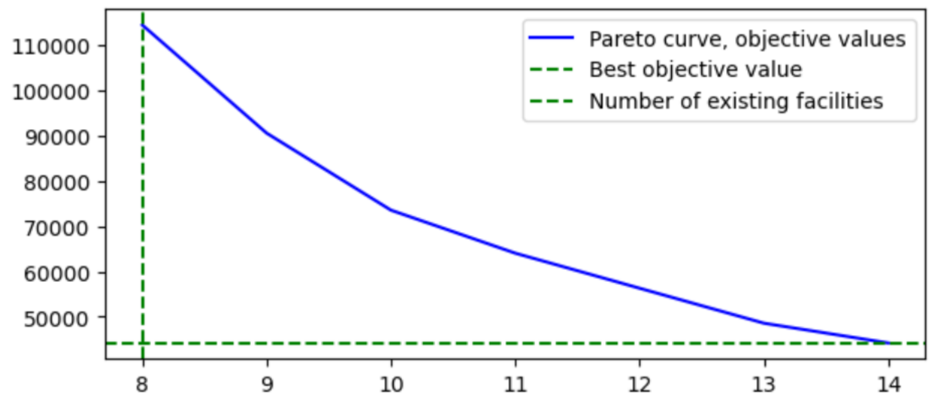
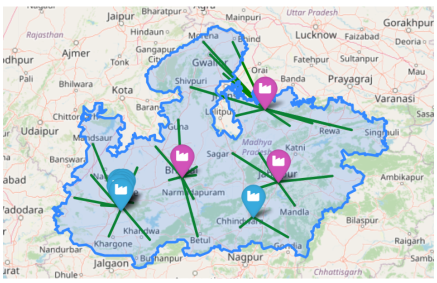
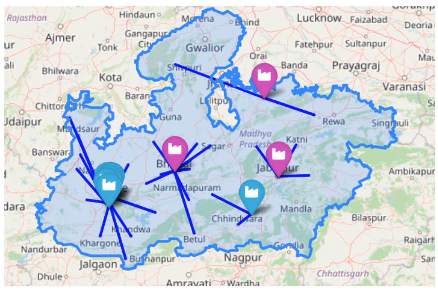


Figure 5.8: Pareto curve for the Hub Location (single use storage) model. The y-axis contains the objective values, which indicates the total hectare-kilometer. The x-axis contain the number of storages (existing + new)



(a) Each green line indicates a connection between a farmer and a storage.



(b) Each blue line indicates a connection between a market and a storage.

Figure 5.9: Two figures displaying all connections from any storage, where $U = 3$, which is the optimal solution for this model.

The boxplot found in Figure 5.10 display the distribution of distances. The distances consist of the distance between farmer & storage, and market & storage. On average, the distance is 100 kilometers. Moreover, there are distances with value zero. These distances are the newly opened storages, since those are placed on the same location as the markets.

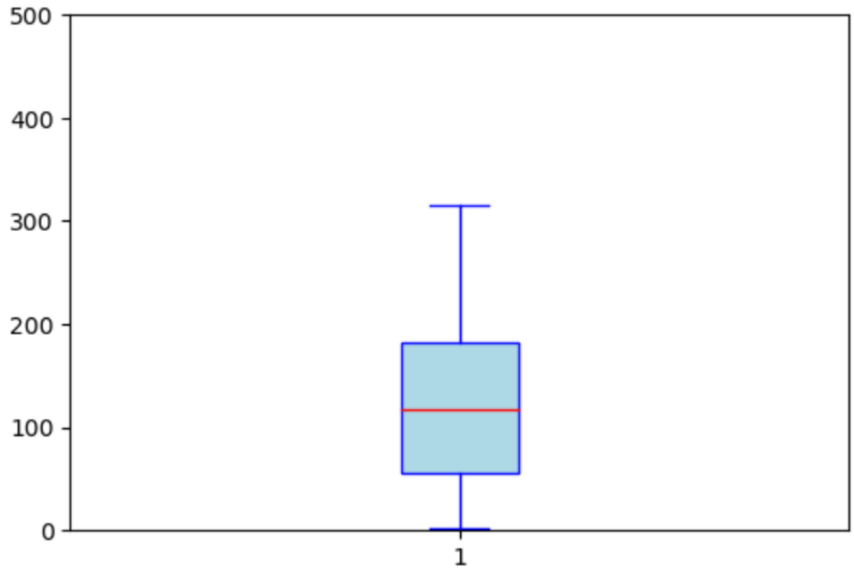


Figure 5.10: Boxplot displaying the distribution of distances, using optimal value $U = 3$.

The coverage of the model is displayed in Figure 5.14 for the values $U = 0$, $U = 1$, $U = 3$ and $U = 6$. The numerical results of the coverage can be found in Table 5.2. Notable is the opening of the storages for $U = 6$. By opening storages in the Northern and Western part of Madhya Pradesh, the coverage goes up to 94.59% with a radius of 200 kilometer.

Radius (km)	U=0	U=1	U = 3	U = 6
100	20.27%	35.14%	47.30%	59.46%
150	36.49%	58.11%	71.62%	85.14%
200	51.35%	71.62%	86.49%	94.59%

Table 5.2: This table shows the percentages of locations (markets or farmers) that are in a specific radius of a storage location. These percentages are calculated based on the values $U = 0$, $U = 1$, $U = 3$, $U = 6$ and for the radii 100, 150 and 200 kilometer

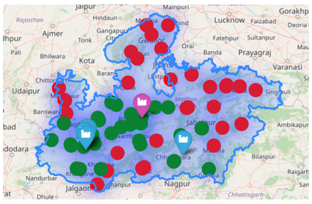


Figure 5.11: Number of new storages equals one

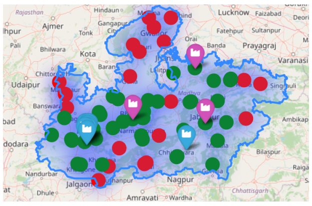


Figure 5.12: Number of new storages equals three

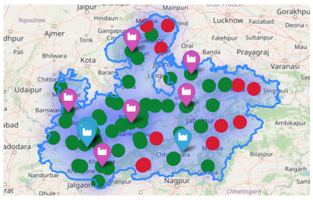


Figure 5.13: Number of new storages equals six

Figure 5.14: These figures plot the optimal locations for cold storages, where the blue markers represent existing storages and the purple markers represent the new storage locations. Locations (market or farmer) that are within a reach of 150 km of a storage location are marked green. Locations not within a reach of 150 km are marked red.

5.1.3 Hub Location (predetermined markets)

The hub location (predetermined markets) model is executed for different values for U, the outcomes are displayed in Figure 5.15. All outcomes can be found in Table 8.3 with corresponding running times in Figure 8.5. The optimal locations are displayed in Figure 5.16. However, this model makes use of predetermined markets. There is no constraint that makes sure that the storage has an equal flow. Table 8.4 shows the balances per used storage. Negative balance means that there is more demand than production at this specific storage. A positive balance means more production came in this storage, than demand that left the storage. Some storages have balance zero, which indicates that these are not used in the optimal model. Figure 8.6 plots these storage locations with as key the ID of each storage.

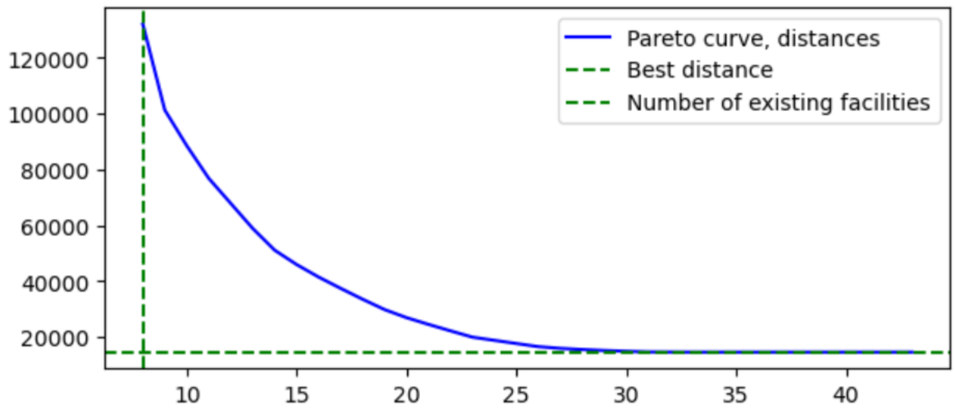
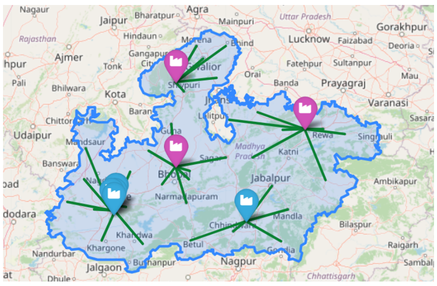
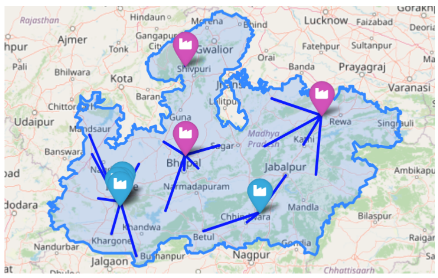


Figure 5.15: Pareto curve for the Hub Location (Single use storage) model. The y-axis contains the objective values, which indicates the total hectare-kilometer. The x-axis contain the number of storages (existing + new)



(a) Each green line indicates a connections between a farmer and a storage.



(b) Each blue line indicates a connection between a market and a storage.

Figure 5.16: Two figures displaying all connections from any storage, where $U = 3$, which is the optimal solution for this model.

Radius (km)	U=0	U = 1	U = 3	U = 6
100	20.97%	25.81%	43.55%	58.06%
150	35.48%	48.39%	70.97%	83.87%
200	50.00%	69.35%	85.48%	93.55%

Table 5.3: This table shows the percentages of locations (markets or farmers) that are in a specific radius of a storage location. These percentages are calculated based on the values $U = 0$, $U = 1$, $U = 3$, $U = 6$ and for the radii 100, 150 and 200 kilometer

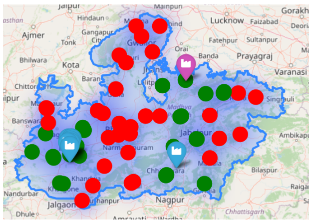


Figure 5.17: Number of new storages equals one

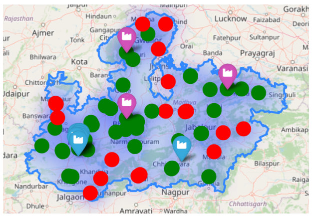


Figure 5.18: Number of new storages equals three

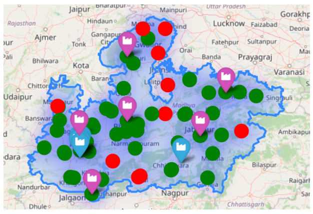


Figure 5.19: Number of new storages equals six

Figure 5.20: These figures plot the optimal locations for cold storages, where the blue markers represent existing storages and the purple markers represent the new storage locations. Locations (market or farmer) that are within a reach of 150 km of a storage location are marked green. Locations not within a reach of 150 km are marked red.

5.1.4 Comparison models

Figure 5.21 displays the objective values for the optimal solutions, where Table 5.4 shows the exact values. It becomes clear that the hub location with single use storage has the lowest total hectare-kilometer, and therefore performs best. Followed by the basic hub location model, and the hub location model with predetermined markets performs worst. The reason that the hub location model with single use storage performs better than the basic hub location model is the absence of the flow balance constraint. This should be taken into account, since there in fact are imbalances at specific storages. Similar situation holds for the hub location model with predetermined markets. However, the outcome for predetermined markets is worst. In practice, it does not seem logical to visit a storage while farmers and markets are located close to each other in specific situations. This might lead to long distances, which translates to a high total hectare-kilometer outcome for this model.

For the value $U = 1$ at the basic hub location model and the hub location model with single use storage, the storage location is similar. This storage location also appears in the hub locations model at $U = 3$, which indicates the significance of this specific location for minimizing the total hectare-kilometer.

Figure 5.22 shows the coverage for values $U = 0$, $U = 1$, $U = 3$ and $U = 6$ with a radius of 150 kilometer. This plot indicates a positive trend for all three models. The coverage for $U = 0$ is approximately similar for the three models. The basic hub location model experiences diminishing returns when U increases, since the coverage does increase with a slower rate for higher values of U. The hub location model with single use storage

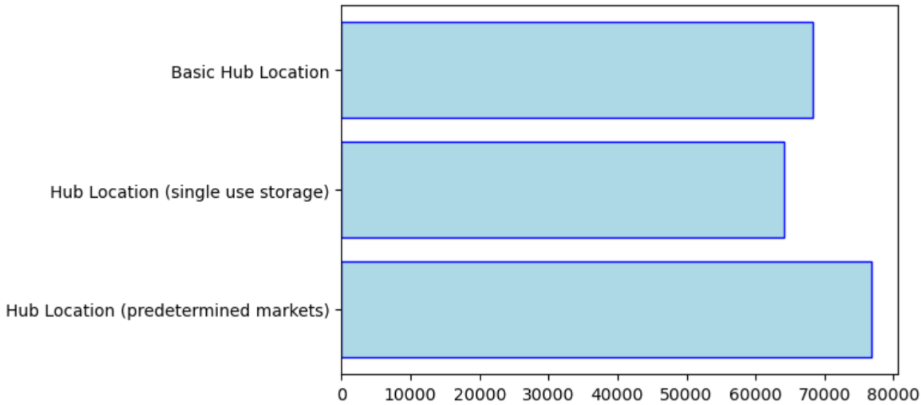


Figure 5.21: This plot displays the objective value of the optimal solution for the three models. The objective value is expressed in hectare-kilometer and can be found on the x-axis.

Model	**Objective Value**
Basic Hub Location	68,201.3064
Hub Location (single Use Storage)	64,058.7225
Hub Location (predetermined Markets)	76,848.04

Table 5.4: Objective values for different models at $U = 3$, expressed in hectare-kilometer.

reaches the highest coverage with 85.14%, with overall a steady increment in coverage. Finally, the hub location model with predetermined markets indicates a different trend. The coverage between $U = 0$ and $U = 1$ does not increase as much as the other two models. However, the largest increment happens between $U = 1$ and $U = 3$.

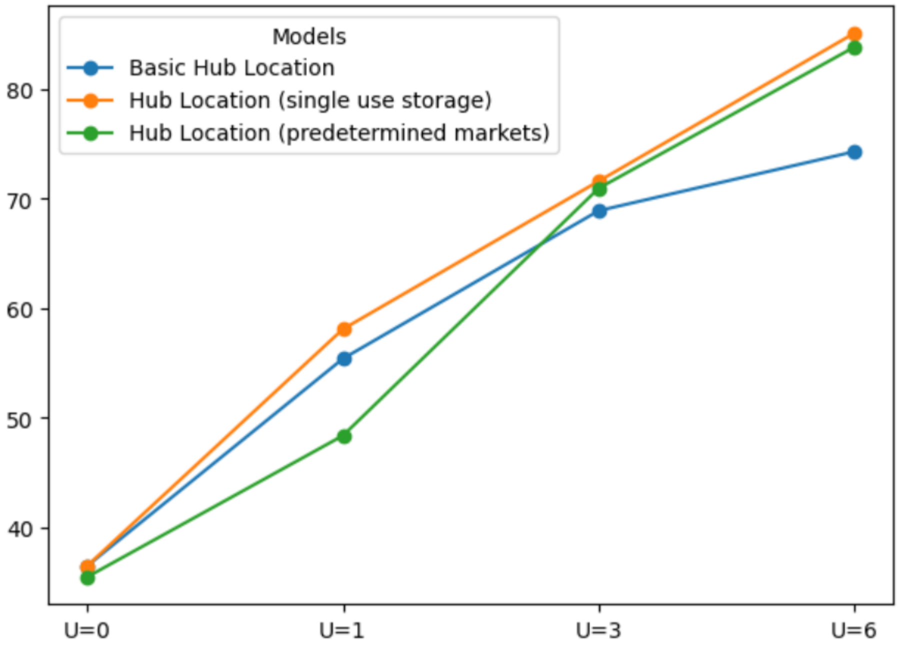


Figure 5.22: Coverage of all models compared in this line plot, for the radius of 150 kilometer. The coverage is shown on the y-axis, for the values $U = 0$, $U = 1$, $U = 3$ and $U = 6$ on the x-axis.

For the running times, the hub location (single use storage) model has the longest running time, followed by the hub location model (predetermined markets) and the basic hub location model has the shortest running time. The exact running times are displayed in Table 5.5. The fact that the hub location (single use storage) has a longer running time compared to the basic hub location model seems counteractive at first since the size of the basic hub location model is larger. The sizes of the basic hub location model and hub location model with single use storage are displayed in Table 3.2 and Table 3.3. There are two reasons for this situation. Firstly, when only one storage can be used in the solution the search complexity increases. On the contrary, the basic hub location model makes it possible to distribute products over different storages and this makes the model less complex and lead to shorter running time. The other reason for this difference in running time is the balance constraint. The absent of the balance constraint leads to more feasible solutions and therefore longer running time.

Model	**Running Time (seconds)**
Basic Hub Location	2.6211
Hub Location (single use storage)	58.9903
Hub Location (predetermined markets)	9.2177

Table 5.5: Running times for different models at $U = 3$.

Chapter 6

Conclusion

6.1 Key Findings

The current tomato supply chain in India leads to a large amount of tomato losses. One of the reasons for these large losses is a lack of availability for cold storages in the state of Madhya Pradesh. Cold storage locations increase the shelf life of tomatoes, which leads to less waste of tomatoes. The goal of this thesis is summarized in the following subject: Optimization of Cold Storage Locations in Madhya Pradesh to Reduce Tomato Losses. The first step of this thesis is retrieving data regarding farmers, storages and markets. Three different models are developed, each model makes use of MILO. It turns out that the hub location model with single use storages has the lowest total hectare-kilometer and therefore performs best. The same model performs best in terms of coverage, the largest percentage of farmers and markets are in the reach of a cold storage in a radius of 150 kilometer. By making the cold storage locations available, less tomatoes will go to waste which.

6.2 Managerial Implications

Three different models are created in this thesis. The basic hub location model provides lots of flexibility, farmers and markets can use as many storages as required to obtain the optimal locations. The hub location model with single use storages makes it easier to trace crops in the tomato supply chain, since each farmer and market are connected to only one storage. The last model makes use of predetermined markets, which might be a more realistic scenario. Therefore, only the connection between farmer and storage had to be found. Another benefit of having cold storage locations in the tomato supply chain, is that it can act as a safety net for farmers. In times of a production surplus, farmers can store tomatoes in the storage for a longer period. These models are developed for Madhya Pradesh, but with little modification can be applied to other states or countries.

6.3 Limitations

There exists a few limitations to this thesis. The first limitation is the demand parameter. In this thesis the demand is based on the population near a market location. However, in practice the demand can be distributed differently. Another limitation is the flow imbalances at the hub location with single use storage and hub location with predetermined markets. In an ideal situation the flow in would equal the flow out, where the storage acts as a perfect intermediary.

This model is applicable to different states or countries. However, there is a limitation in terms of size. While running the basic hub location model and the hub location model with single use storage, the running time grew significantly when U increased. This means these two models are not scalable, and therefore might not be applicable to other states or countries which a larger number of locations.

6.4 Further Research

To extend this research, one can search for different locations to open storages. The current situation is that storages can be opened at the exact same locations as markets. For further research it might be interesting to determine different locations, besides the locations of markets. Also, in this thesis the storage capacities are unlimited. There is no data on the existing locations, and the capacities for the new storages should be determined. By including the capacities, the model becomes more realistic for practical implementation.

Chapter 7

Appendix A : Data

Year	**Name**	**State Name**	**Surface (Ha)**	**Latitude**	**Longitude**
2017	Madhya Pradesh	Jabalpur	20.14	23.1608938	79.9497702
2017	Madhya Pradesh	Balaghat	3.36	21.8133417	80.1916685
...	...	...	...	...	...
2017	Madhya Pradesh	Hoshangabad	4.32	22.7529598	77.7173261

Table 7.1: Snippet of the farmers dataset.

Sr. no.	**State**	**District**	**Name**	**Latitude**	**Longitude**	**Pop5km**	**Per5km**	**Pop10km**	**Per10km**
Madhya Pradesh	Bhopal	Berasia	Bhopal road	23.623873	77.432234	68031	0.49969	129685.0313	0.46151
Madhya Pradesh	Bhopal	Karond	Pandit Laxminarayan Sharma	23.2912759	77.4012875	672312.5	4.93819	1703083.875	6.06078
...	...	...	...	...	...	...	...	...	...
Madhya Pradesh	Vidisha	Vidisha	Krishi Upaj Mandi Samiti	23.5225277	77.798016	188005	1.3809	261665	0.9311

Table 7.2: Snippet of the market dataset.

Name	Sector	Products stored	Latitude	Longitude
HINDUSTAN STORAGE	Private	Potatoes	26.2162438	78.1826037
SAHKARI SHEET LTD.	Cooperative	Multipurpose	22.30163	79.26912
...	...	...	...	...
SHRI. ALU UTPADAK	Cooperative	Potatoes	22.6962244	75.8390046

Table 7.3: Snippet of the final cold storages dataset.

Chapter 8

Appendix B : Results

8.1 Basic Hub Location

U	**Objective Value**	**Running Time (seconds)**
0	111668.5841	0.3054
1	90707.0705	0.4013
2	75606.2475	0.5942
3	68201.3064	2.6211
4	61589.8867	14.5285
5	57424.9686	60.0671
6	53997.6676	169.7894
7	51112.1336	500.1856
8	49162.3528	907.4534
9	47308.8134	1951.5911

Table 8.1: Objective values (hectare-kilometer) and running times for different U values. Retrieved from the Basic Hub Location model.

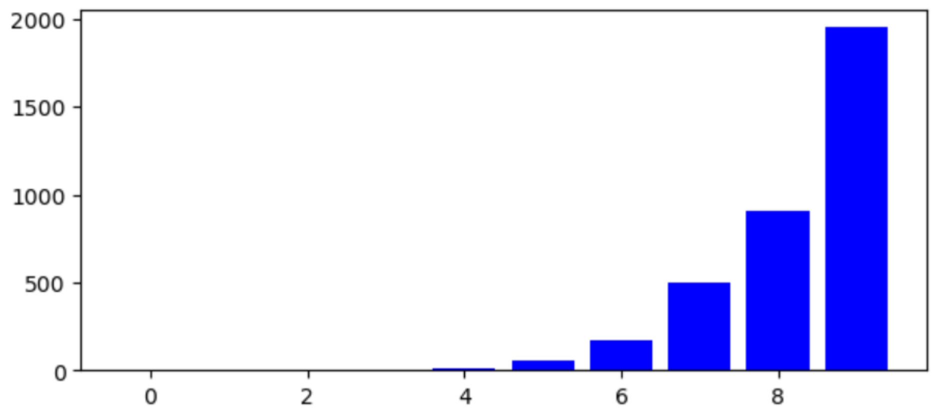


Figure 8.1: This plot displays the running times for every value of U. The running times can be found on the y-axis, the number of new storages on the x-axis.

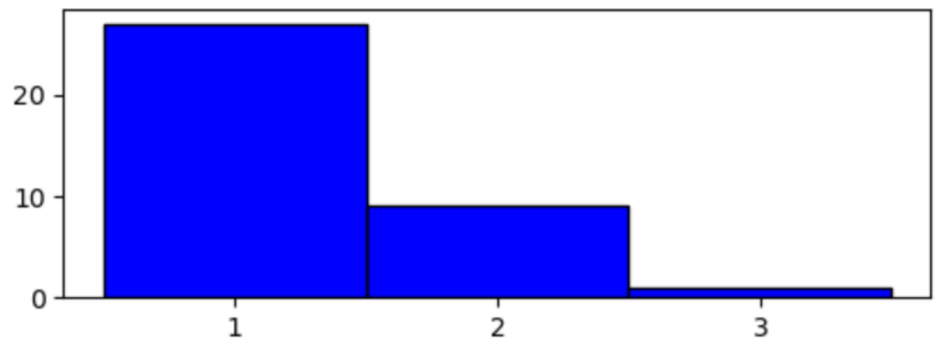


Figure 8.2: The distribution of the number of utilized storages per farmer visualized.

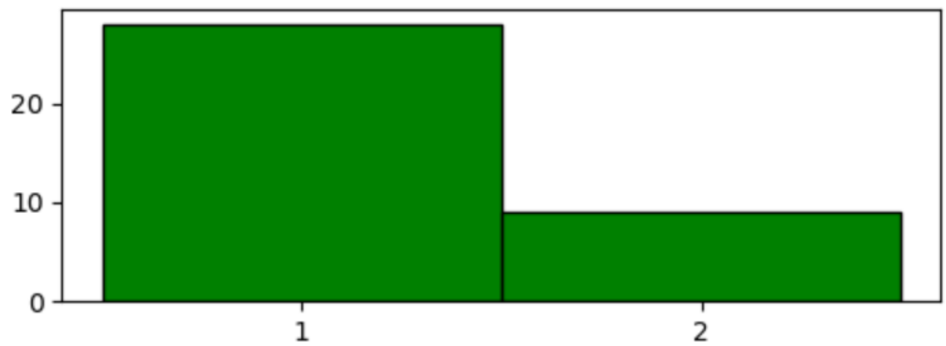


Figure 8.3: The distribution of the number of utilized storages per market visualized.

8.2 Hub Location (single use storage)

U	**Objective Value**	**Running Time (seconds)**
0	114508.1959	0.3421
1	90603.4499	0.9052
2	73571.3829	6.0585
3	64058.7225	58.9903
4	56308.5434	416.9630
5	48580.4152	1594.8992
6	44208.8753	8076.0485

Table 8.2: Objective values (hectare-kilometer) and running times for different U values. Retrieved from the Hub Location (single use storage) model.

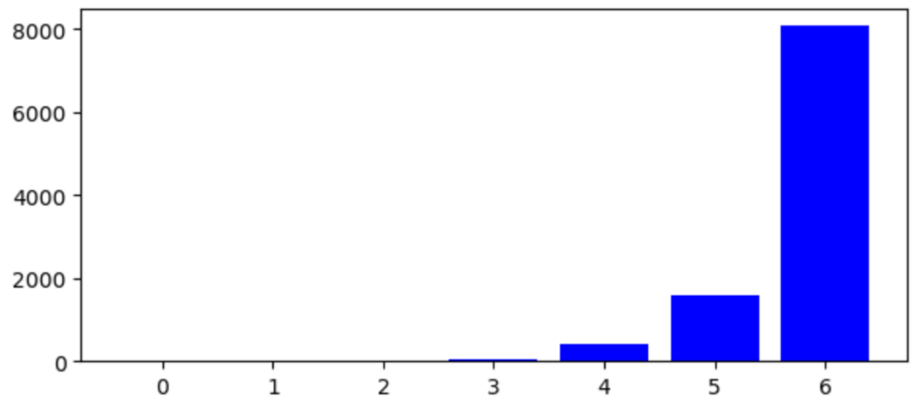


Figure 8.4: This plot displays the running times for every value of U. The running times can be found on the y-axis, the number of new storages on the x-axis.

U	**Objective Value**	**Running Time (seconds)**
0	132073.50	0.599,297
1	101304.27	0.740,074
2	88526.47	2.028,338
3	76848.04	9.217,735
4	67892.71	45.563,055
5	58927.82	109.515,359
6	51128.11	179.215,005
7	45980.98	331.019,313
8	41503.50	538.261,937
9	37470.26	664.687,882
10	33572.54	487.536,709
11	29871.85	224.361,315
12	26976.01	100.485,747
13	24560.90	45.200,856
14	22248.80	14.653,394
15	19992.24	4.605,702
16	18847.60	3.207,055
17	17714.21	2.281,130
18	16593.09	1.518,086
19	15979.65	1.236,635
20	15560.04	1.183,162
21	15211.66	0.899,588
22	14937.31	0.785,503
23	14701.95	0.769,653
24	14666.37	0.657,215
25	14634.43	0.665,841
26	14634.43	0.593,632
27	14634.43	0.653,492
28	14634.43	0.596,080
29	14634.43	0.594,346
30	14634.43	0.578,044
31	14634.43	0.611,993
32	14634.43	0.596,179
33	14634.43	0.671,061
34	14634.43	0.643,337
35	14634.43	0.611,438

Table 8.3: Objective values (hectare-kilometer) and running times for different U values. Retrieved from the Hub Location (predetermined markets) model.

8.3 Hub Location (predetermined markets)

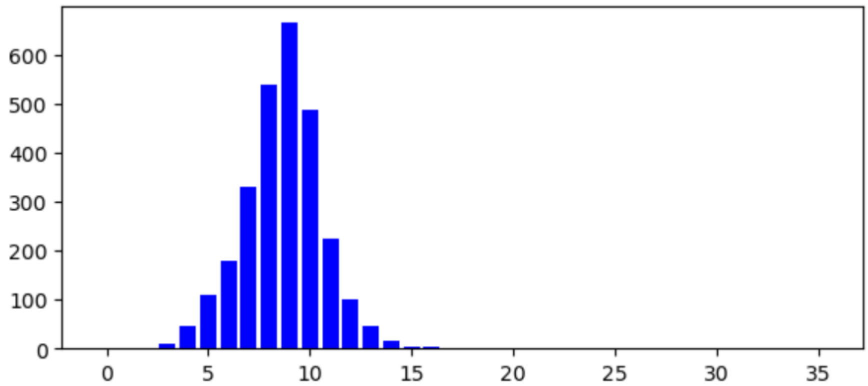
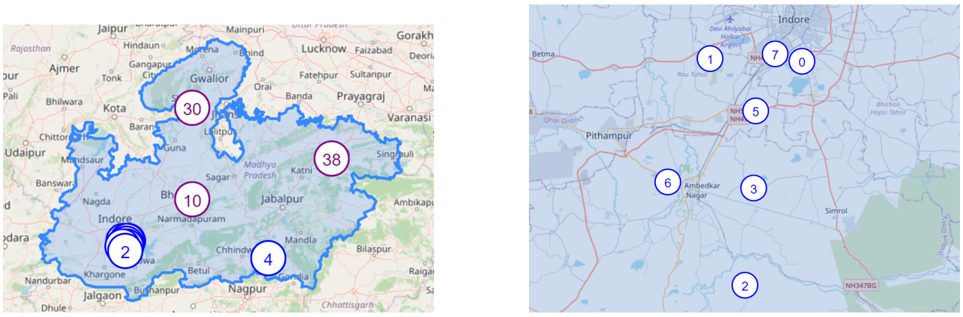


Figure 8.5: This plot displays the running times for every value of U. The running times can be found on the y-axis, the number of new storages on the x-axis.

ID	Storage Name	Balance
0	DEVI SHRI AHILYA CO-OP MKTG. SOCOETY LTD	-36.71
1	GOVT. POULTRY PROJECT, Near MOG Lines, Indore	0.00
2	MALAV SAHAKARI VIPANAN SAMITI MARYADIT	10.09
3	MP STATE COOP MKTG. FED. LTD.	23.07
4	SAHKARI SHEET GRIHA SANSTHA LTD.	28.18
5	SHIV SHAHTI ALU UTPADAK, Vipanan and Prakriya sahkari, Sanstha ltd., T	0.00
6	SHRI AMBIKA ALU UTPADAK, Vipanan and Prakriya Sahakari Ltd, Gawli Pa	32.55
7	SHRI. GANESH ALU UTPADAK, Vipanan Avam Prakriya Sahkari, Sanstha	-5.63
10	Bhopal road, berasia tehsil-berasia dist.-bhopal pin-463106	8.46
30	Krishi upaj mandi samiti, nainagarh road. Morena, m.p.	27.51
38	Krishi upaj mandi, Birla raod, Nayi basti, Satna	6.88

Table 8.4: Balances and ID Numbers of Storages. Negative balance means demand is higher than production for this specific storage. Positive balance indicates a higher production than demand for this storage. Balance with value zero means the storage is not utilized in the optimal solution.



(a) Opened storages, where the ID of each storage is displayed.

(b) This figure displays the storages from the cluster in the figure left.

Figure 8.6: The two figures display the locations of the storages with the corresponding ID which can be linked to Table 8.4.

Bibliography

Monika Agarwal, Sushant Agarwal, Subia Ahmad, Ruchika Singh, and K.M. Jayahari. 2021. Food Loss and Waste in India: The Knowns and the Unknowns. http://www.wri.org/publication/food-loss-and-waste-in-india World Resources Institute India.

Sibel A. Alumur, James F. Campbell, Ivan Contreras, Bahar Y. Kara, Vladimir Marianov, and Morton E. O'Kelly. 2021. Perspectives on modeling hub location problems. *European Journal of Operational Research* 291, 1 (2021), 1–17. https://doi.org/10.1016/j.ejor.2020.09.039

AtlasBig. 2023. World Tomato Production by Country. (2023). https://www.atlasbig.com/en-us/countries-tomato-production

National Horticulture Board. 2020. Monthly Report Tomato. (June 2020). https://nhb.gov.in/statistics/Reports/Monthly%20Report%20on%20Tomato%20for%20June,%202020.pdf

Sorichetta A. Bondarenko M., Kerr D. and A.J. Tatem. 2020. Census/Projection-Disaggregated Gridded Population Datasets, Adjusted to Match the Corresponding UNPD 2020 Estimates, for 183 Countries in 2020 Using Built-Settlement Growth Model (BSGM) Outputs. https://doi.org/10.5258/SOTON/WP00685 dataset.

CNBCTV18. 2023. MP Farmers Bear Losses, Resort to Distress Sales Amid Lack of Cold Storage for Tomatoes, Garlic. https://www.cnbctv18.com/economy/mp-farmer-loss-distress-sales-cold-storage-tomato-garlic-shivraj-singh-chauhan-16561621.htm Accessed: 2024-07-10.

Edward J. Collins, Cressida Bowyer, Audrey Tsouza, and Mridula Chopra. 2022. Tomatoes: An Extensive Review of the Associated Health Impacts of Tomatoes and Factors That Can Affect Their Cultivation. *Biology* 11, 2 (2022). https://doi.org/10.3390/biology11020239

Reza Zanjirani Farahani, Masoud Hekmatfar, Alireza Boloori Arabani, and Ehsan Nikbakhsh. 2013. Hub Location Problems: A Review of Models, Classification, Solution Techniques, and Applications. *Computers Industrial Engineering* 64, 4 (2013), 1096–1109. https://doi.org/10.1016/j.cie.2013.01.012

Centre for Indian Knowledge Systems. 2014. Organic Farming Practices. (2014).

Fatima Zahraa Grine, Oulaid Kamach, and Naoufal Sefiani. 2022. A Multi-objective Hub Location Problem for Perishable Products. In *Advanced Intelligent Systems for Sustainable Development (AI2SD'2020)*, Janusz Kacprzyk, Valentina E. Balas, and Mostafa Ezziyyani (Eds.). Advances in Intelligent Systems and Computing, Vol. 1417. Springer, 785–795. https://doi.org/10.1007/978-3-030-90633-7

Ashok Gulati, Harsh Wardhan, and Pravesh Sharma. 2022. *Tomato, Onion and Potato (TOP) Value Chains*. Springer Nature Singapore.

A. Madhukar, V. Kumar, and K. Dashora. 2022. Spatial Analysis of Temperature Trends During Rabi and Kharif Seasons in India. *Letters in Spatial and Resource Sciences* 15 (2022), 451–467. https://doi.org/10.1007/s12076-022-00304-w

A. Mohan, R. Krishnan, K. Arshinder, J. Vandore, and U. Ramanathan. 2023. Management of Postharvest Losses and Wastages in the Indian Tomato Supply Chain—A Temperature-Controlled Storage Perspective. *Sustainability* 15, 2 (2023). https://doi.org/10.3390/su15021331

Krzysztof Postek, Alessandro Zocca, Joaquim Gromicho, and Jeffrey Kantor. 2024. *Hands-On Optimization with Python*. Online. https://mobook.github.io/MO-book/

Hannah Ritchie, Pablo Rosado, and Max Roser. 2023. Data Page: Tomato Yields. *Agricultural Production* (2023). https://ourworldindata.org/grapher/tomato-yields Data adapted from Food and Agriculture Organization of the United Nations. Retrieved from https://ourworldindata.org/grapher/tomato-yields.

M. S. Sibomana, T. S. Workneh, and K. Audain. 2016. A Review of Postharvest Handling and Losses in the Fresh Tomato Supply Chain: A Focus on Sub-Saharan Africa. *Food Security* 8, 2 (2016), 389–404. https://doi.org/10.1007/s12571-016-0562-1

Solidaridad. 2022. *Study on Post-Harvest Losses in Tomato Supply Chain in Madhya Pradesh*. https://www.agroberichtenbuitenland.nl/documenten/rapporten/2022/05/17/study-tomato-value-chain--mp-india

K. V. Vala, M. T. Kumpavat, and S. Datta. 2021. Long Distance Supply of Tomato: A Case Study of Anand (Gujarat). *IJRAS* 8, 3 (May 2021), 127–131.

B. van Dam, M. de Goffau, J. van Lidt de Jeude, and S. Naika. 2005. *Cultivation of Tomato: Production, Processing and Marketing.* Number 17 in Agrodok-series. Agromisa/CTA, Wageningen, The Netherlands. 69–71 pages. https://hdl.handle.net/10568/52975

S. Vennila, M. Narayana Bhat, D. Anita Kumari, S.K. Yadav, and V.K. Sharma. 2020. *Effect of Climatic Variability and Weather Factors on Development of Tomato Early Blight in a Hot Semi-Arid Region of Southern India.* https://doi.org/10.5958/0974-0112.2020.00046.8