# **A Multistage Optimization Approach in Designing a Resilient Supply Chain Network**

Cole Wesley, Madhur Khanna, Nagarjuna Chidarala , Priyam Sarkar, Vishnu Ponduri, Matthew Lanham

Purdue University, Department of Management, 403 W. State Street, West Lafayette, IN 47907

teammate1@purdue.edu; teammate2@purdue.edu; nchidara@purdue.edu; teammate4@purdue.edu; vponduri@purdue.edu; [lanhamm@purdue.edu](mailto:lanhamm@purdue.edu)

## **ABSTRACT**

In this research, we present a mixed-integer linear programming (MILP) model tailored for a unique multi-stage supply chain network design problem. This model seamlessly combines decisions regarding the positioning and capacity of manufacturing facilities and storage warehouses, alongside the selection of suppliers and modes of transportation. Additionally, it manages the distribution of various products across this network. The primary objective of our model is to determine the most cost-effective network structure while adhering to constraints that include resource availability, technological specifications, and customer service level agreements. In the competitive landscape of global commerce, companies are integral parts of complex supply chains rather than standalone entities. Achieving success hinges on fulfilling customer demands efficiently, which can lead to significant cost savings and improved service quality. Supply chain management is challenging due to unpredictable factors such as demand, costs, pricing, lead times, and product quality, all occurring within a landscape of competition and cooperation. Inadequate decision-making can result in either costly overstocking or insufficient inventory, failing to meet customer expectations. The central inquiry of this study is, "What approaches can the company employ to evaluate supply chain strategies aimed at minimizing the overall logistics cost?" In response to this question, we have developed a supply chain network optimization model utilizing Python and Gurobi software.

## **INTRODUCTION**

This section should be at least 600 words.

The urgency of optimizing supply chain networks in today's complex global economy is paramount. Businesses are navigating unprecedented challenges in cost management, operational efficiency, and service level maintenance due to the dynamic nature of global markets. This research introduces a multi-stage optimization approach to construct resilient supply chain networks, targeting cost reduction. It aims to provide strategic solutions that not only lower operational costs but also bolster the adaptability and robustness of supply chains against market volatility and changing consumer needs.

The recent advancements in supply chain management, as highlighted in various reports, emphasize the critical role of continuous optimization. Gartner's 2022 report indicates that effective supply chain optimization strategies can lead to significant operational cost reductions, up to 15%, which is vital for enhancing market competitiveness. Supply Chain Brain's analysis in 2022 further supports this by showing how network optimization can result in considerable cost savings and improved operational visibility, illustrated by the example of a global manufacturer saving $60 million. Additionally, the FTI Consulting "Supply Chain Barometer 2023" points out the increasing cost pressures due to inflation and material shortages, underscoring the necessity for innovative and adaptable supply chain strategies. These insights collectively demonstrate the importance of ongoing research and development of advanced optimization techniques to maintain efficient, cost-effective, and resilient supply chains in a dynamic global market.

In today's dynamic and fiercely competitive global marketplace, companies are constantly striving to enhance operational efficiency and reduce expenses while ensuring high customer satisfaction. The complexity of this task demands advanced modeling and optimization techniques. Without such optimized networks, companies may face escalated costs, inefficiencies, disruptions in the supply chain, and potentially a decline in market share.

The strategic formulation of a supply chain network (SCND) is crucial for improving a firm’s financial health. In the initial stages of supply chain network design, Greenfield Analysis (GFA), also known as center of gravity analysis, emerges as a vital approach. For this research, we have done the GFA to ensure 5 LC, by eliminating DC and catering all the demand from the LC only. The objective is to leverage the insights provided by the analysis to designate the nearest LC for each customer, thereby optimizing the total logistics cost while upholding our commitment to service delivery standards. The main outcome of Greenfield Analysis is to pinpoint an optimal location for warehousing, aimed at minimizing both inbound and outbound transportation costs. This paper presents the detail analysis of Greenfield Study which comprises of weighted K mean clustering which ensures demand aggregation takes place at a cost-effective way. The study advances from the theoretical underpinnings of Greenfield Analysis to its practical applications, notably the strategic route mapping for each customer.

However, the Grey-field analysis is much more realistic since it is closer to be implemented. To solve the business problem which was asked to find out the new Logistic Center (LC) while keeping the existing Distribution Center (DC) intact a two-stage optimization approach was followed. For the first stage assignment between the existing DCs and customer was optimized. Once the distribution scenario from DC was fixed, two different scenario was applied (Cost Optimization and Service Efficiency) to find out the new logistic centers.

Furthermore, to come up with the effective solution a few models have been tested which is mentioned below:

1.Linear Programming Models: This model is used for optimizing linear objective functions, subject to linear equality and inequality constraints.

2.Mixed-Integer Linear Programming (MILP): These are more complex models that include both continuous and discrete decision variables, ideal for handling various operational constraints in supply chain design.

3.Geographic Information System (GIS) Models: Useful for incorporating geographical data into the decision-making process, particularly for location planning and analyzing spatial relationships.

4.Simulation Models: These models allow for the testing of different scenarios and their impact on the supply chain, providing insights into dynamic and complex systems.

5.Heuristic Models: Employed for solving problems to a satisfactory level within a reasonable time frame, especially useful when exact solutions are not feasible.

## **LITERATURE REVIEW** [30 points]

This section should be at least 700 words.

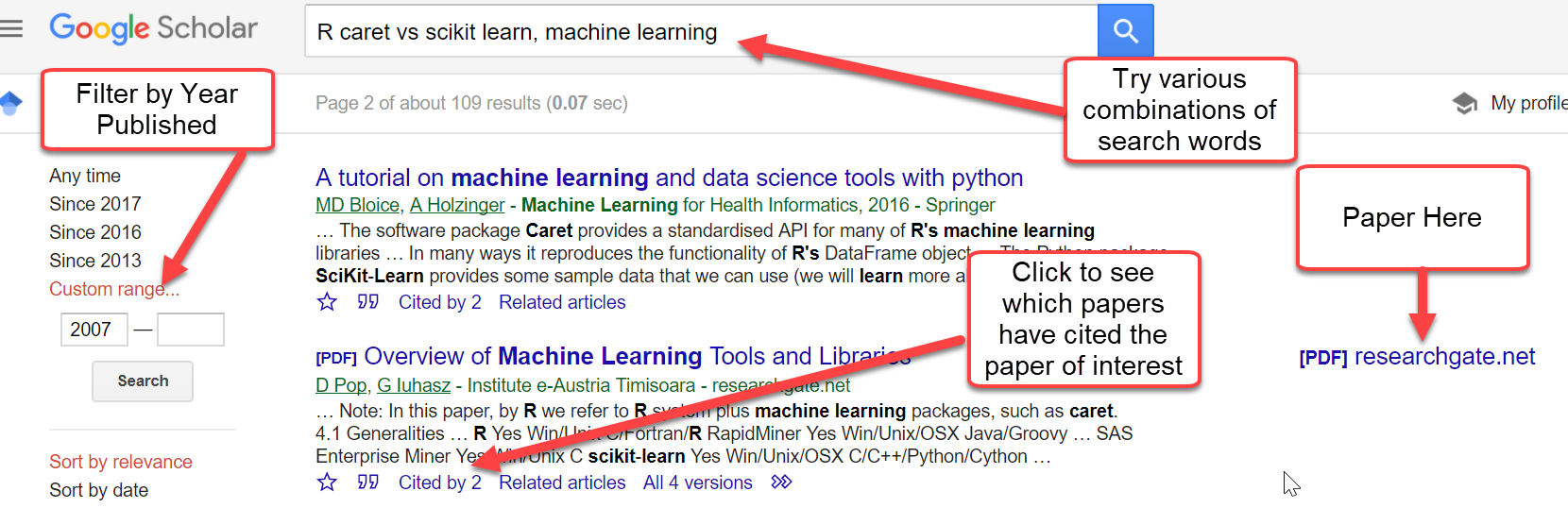
Here you are trying to summarize

1) what is known about the focus area you are working on, and more importantly

2) what is still unknown and thus necessitates further investigation.

To accomplish this use 5-10 academic journal references. I recommend you do the following: As a team, use Google Scholar and/or library databases to search for key words on what your study is about. Search articles not older than ten years old. Just using Google Scholar you will likely get several articles returned after filtering by year. Try to find one or two that are clearly what you are focused on. Ignore the ones that are “*sort of related*.” Once you find those two highly relevant articles, you can click and see what other articles have cited them to see if you find even newer articles that talk about what you are doing. Download those handful of articles, then go directly to their references (listing within the paper at the end) to see if you see any other articles they have cited that are highly related to your study. You can search for them in Google Scholar to get them. At this point you will likely have more than 10 articles.

You goal is to find those one or two highly relevant and recent articles that give you most of the info you need (e.g. previously writing important articles to add to YOUR lit review). Once you find the golden nugget article, you are in great shape.



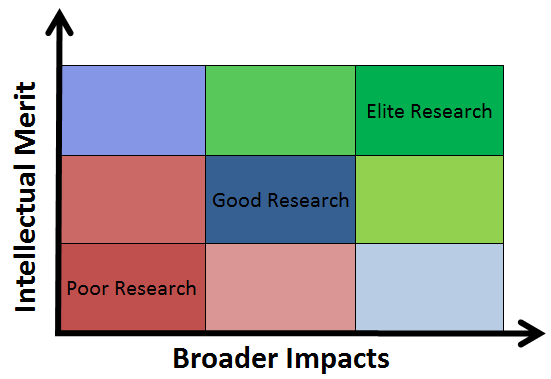
Now, based on the article titles, rank them based on the ones you think are most related to your work. Then one by one just read the paper **abstracts** to see if you think they are truly relevant to your study. If not, go to the next one.

Once you have filtered down the list, you are NOT going to read all these articles. That's right, you are NOT going to read these end-to-end. You will not have the time. What you are going to do is split these papers up among your teammates and each skim through the paper usually reading the introduction section to get some motivation for your introduction section, but most importantly identify what contributions their paper provides (**What** is novel about what they did?, **How** did they do it?). This could be an approach, a model, a new method, new theory, etc. You should have a general idea based on what they said in the abstract. For each paper, you just need to write a few sentences that describe what their contribution was in the literature review section.

Next, as a team you are going to summarize what all these papers provided. Usually you can organize these contributions in the form of a table to make it easier to see for the reader.

Look here for many examples from previous MWDSI conferences: <http://matthewalanham.com/>

Lastly, usually at the end in the conclusions or discussion section of journal articles, the authors will mention items that are still not fully understood that still require more research/investigation. They might provide suggestions for future research. You will want to summarize this in two to three sentences. Hopefully, one of the papers that your teammates review identifies what you are doing as something that is important necessitating future research. This provides support for the value of what you are doing – it is support that what you are doing is NOVEL. If you cannot find such a case from those studies, you need to make a claim of why what you are doing is indeed important, novel, useful to a company, has implications to the field, could be extended to other areas, etc. Researchers will make sure their paper has both (1) technical merit and (2) broader impacts. For your project paper, I only expect broader impacts, meaning your work is actually useful at addressing some problem. Employers love to see professional publications from conference proceedings or journals on your resume. Journal articles are harder to achieve because you usually need to have significant mathematical rigor with proofs. Your potential boss/employer will not care about that, rather they will enjoy seeing how you framed a problem and logically developed and evaluated your solution. Thus, your project papers are perfect conference proceedings papers. However, some of you will be close to journal quality, and I will try to help you get that submitted to the appropriate journal if I feel we really have identified something more than broader impact that has intellectual merit.



Modeling-type papers on common public datasets (e.g., UCI Machine Learning Repo)

* If you are doing a project where your goal is to develop a better model like Author/Paper #1 used linear regression and obtained an R-squared of 0.78, Author/Paper #2 used decision trees and obtained an RMSE of 112.3, etc. Then you will need to record those numbers, which are usually found in the modeling or results sections of those papers. When you build similar models, you will want to have some statistical comparison of your models to others for comparison. This is a must if you are using the same dataset as others have used.

Modeling-type papers on non-common public datasets (e.g., Kaggle.com)

* If you are using data on a very recent Kaggle competition, there will likely not be any published papers using that dataset because it is too new. If the Kaggle competition is a forecasting problem, you just need some examples of expected statistical performance and ensure you are calculating those measures for comparison. Some might use stats like MAPE, or in classification F1 score, etc.

Modeling-type papers using proprietary datasets

* If you are using client's proprietary data, you will still want to have to have some stats from other studies to give an idea of what performance could be expected modeling in this problem area (e.g., supply chain, sparse demand, grocery forecasting, etc.). Also, if other authors have used a particular approach (e.g. logistic regression), you should also build such a model as a baseline for comparison. Maybe another paper used CART, then you should use CART and capture those stats. Hopefully you can make a case for using a different method to support the problem and compare those stats to the methods that have been tried in the past.

Non-business problem focused papers

* In the above examples, the focus is to create models that would yield better decision support than previously published models or approaches. However, you might be doing a project that focuses on creating a new algorithm or method, comparing some software packages, proposing a new work flow design for certain data science scenarios, or investigating some aspect of data mining. In these more classical research cases, we will work together to ensure all items required are covered. In other words, discuss with me, because every case is different.

At the end of the day, you are using these articles to provide you some guidance and ideas on how you can approach your problem. You can use these approaches as baseline models/solutions, which might work great for the industry partners problem (or class project problem). Most likely you will have some other ideas of your own, or there are other things that need to be considered based on partner feedback, etc. and this is where you can test or compare your model/solution to what others have tried.

A good literature review will convince the reader that your project fits into an established body of work and addresses a question of concern to the scholarly (or business) community, also that your project adds to our understanding of the topic by offering novel insights.

This section aims to synthesize and analyze key findings from a variety of sources, encompassing reports from industry leaders from our research. By examining the insights, methodologies, and outcomes presented in prior studies, this literature review seeks to provide a solid foundation for understanding the evolving landscape of supply chain network optimization.

**Uncertain Supply Chain Management**

Supply Chain Network Optimization (SCNO) has become indispensable for businesses striving to enhance cost-effectiveness and competitiveness in today's dynamic market environment. With rapid technological advancements such as internet networks and RFID, supply chains are evolving into intricate networks requiring sophisticated management strategies. Greenfield Analysis (GFA) utilizing detailed datasets plays a pivotal role in determining optimal production or warehousing locations, while Strategic Supply Chain Network Design (SCND) efficiently manages sourcing, production, and distribution processes. Despite these advancements, critical gaps persist in partner/supplier selection, vehicle routing problems (VRP), demand forecasting, and inventory management, necessitating further research to bolster supply chain optimization strategies.

**Competitive Supply Chain Network Design**

The emergence of Competitive Supply Chain Network Design (SCND) signifies a paradigm shift, highlighting supply chains (SCs) as primary units of competition. Informed by industry insights, this perspective underscores the importance of competitiveness in SCND, advocating for a framework integrating managerial insights to navigate competitive dynamics among SCs successfully. Further exploration is warranted on the implications of technology and market shifts on SC competition, emphasizing understanding customer behavior and knowledge management for sustainable competitive advantage in SCND.

**Multi-Stage GFA and BFA Network Optimization**

Facility location decisions play a crucial role in a company's competitiveness, costs, and customer satisfaction. Traditional methods face challenges, particularly in Greenfield scenarios. A hybrid approach combining K-means, Genetic Algorithm, and Simulated Annealing offers a solution, highlighting the user's role in decision-making. This approach improves efficiency, as shown in a case example. Additionally, the paper discusses Brownfield scenarios, where Mixed Integer Linear Programming (MILP) can optimize existing facilities. Overall, this presents an innovative methodology for dynamic supply chain environments, addressing limitations of conventional methods and emphasizing user involvement in decision-making.

**Operations Research Methods in Supply Chain Optimization**

SCM research now emphasizes integrated approaches and simulation-optimization frameworks, moving away from traditional mathematical modeling. Integrated models address previous limitations, while simulation-optimization offers flexibility for complex supply chain problems. Sustainability and resiliency are central, with CO2 minimization and risk evaluation integrated into frameworks. Future studies are expected to prioritize simulation-optimization, requiring empirical validation. This shift towards comprehensive approaches suggests a more practical SCM strategy.

**Supply Chain Design and Analysis: Models and Methods**

Supply chain modeling encompasses diverse methodologies such as mathematical programming, stochastic analysis, economic modeling, and simulation. These approaches optimize costs, minimize inventory diversity, aid in material requirements planning, and address buyer-supplier dynamics. Key performance measures include customer satisfaction, cost minimization, and inventory investment. Future research should refine performance metrics, understand the impact of decision variables on performance, address global supply chain issues, and classify supply chain systems for operational insights.

**A k-means Clustering with Embedded Network Connectivity**

Supply chain risk management has become increasingly important due to disruptions in global supply chains. Traditional approaches struggle to address the complexity of modern supply chains. A novel network connectivity embedded k-means clustering approach is proposed to better understand and mitigate risks. This approach integrates various dimensions of risk, facilitating targeted analysis and simulations. Despite existing research in supply chain risk clustering, a comprehensive analysis considering connectivity, location, and supply chain risks is lacking. Future research should focus on refining this approach and exploring alternative clustering algorithms to enhance supply chain risk management strategies.

**Our Study**

Our project distinguishes itself in the domain of supply chain optimization through its innovative integration of greenfield analysis (GFA) and brownfield analysis (BFA) using weighted k-means clustering. While existing studies often focus solely on either GFA or BFA, our approach uniquely combines these methodologies to address diverse scenarios aimed at achieving distinct objectives: decreased costs and increased service levels. Traditional supply chain optimization strategies typically prioritize either establishing new facilities (GFA) or optimizing existing ones (BFA). However, our project recognizes the necessity for a comprehensive approach that encompasses both scenarios to meet the evolving demands of modern supply chain management. Leveraging k-means clustering, our methodology efficiently categorizes potential facility locations into clusters based on various factors such as geographical proximity, market demand, transportation infrastructure, and cost considerations. This clustering approach enables us to identify optimal locations for establishing new facilities in greenfield scenarios, ensuring cost-effectiveness and strategic alignment with market demands. Additionally, our project extends its innovation by applying the same k-means clustering technique to brownfield scenarios. Here, the focus shifts towards optimizing existing facilities to enhance service levels and operational efficiency. By identifying clusters of existing facilities based on factors like production capacity, proximity to markets, and resource availability, our methodology facilitates targeted improvements to maximize service levels while minimizing costs associated with facility upgrades or reconfiguration. We are catering to the dynamic needs of businesses operating in complex supply chain environments, ensuring adaptability to changing market conditions while maintaining competitiveness and resilience in supply chain operations.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Study** | **GFA** | **BFA** | **K-Means** | **SCND** | **S-O** | **MILP** |
| Niki Matinrad, 2013 |  |  |  | ✔ |  |  |
| Reza Zanjirani Farahani, 2014 |  |  |  | ✔ |  |  |
| Avneet Saxena, 2018 | ✔ | ✔ | ✔ |  |  | ✔ |
| Pourya Pourhejazy, 2016 |  |  |  | ✔ | ✔ |  |
| Benita M Beamon, 1997 |  |  |  | ✔ | ✔ |  |
| Xiao Feng Yin, 2016 |  |  | ✔ |  |  |  |
| Our Study, 2024 | ✔ | ✔ | ✔ | ✔ | ✔ | **✔** |

## **DATA**

The dataset utilized in this study was meticulously assembled through a combination of primary and secondary research methodologies. Organization-specific information was obtained via extensive consultations with clients, providing critical data on warehouse costs, third-party logistics, distributed demand across regions, and the requisite shipping times and modes to service demands throughout the United States. This was further detailed by acquiring granular data, such as the latitude and longitude of both shipper and recipient cities, to optimize service levels.

For secondary research, pertinent data concerning real estate costs were gathered, which played a vital role in our Greenfield analysis for identifying potential locations for new Logistic Centers. Additionally, demographic data, including population figures and wage levels in each prospective location, were compiled. This comprehensive data collection effort underpins our optimization model, enabling a strategic approach to enhancing supply chain efficiency and reducing operational costs.

***Table: Data used in study***

|  |  |  |  |
| --- | --- | --- | --- |
| **Serial No** | **Data** | **Type** | **Description** |
| 1 | Shipper Company Name | Categorical | Name of the shipper company |
| 2 | Shipper City | Categorical | City of the shipper company |
| 3 | Shipper State | Categorical | State of the shipper company |
| 4 | Recipient City | Categorical | Used to get the demand center |
| 5 | Recipient State | Categorical | State of the recipient |
| 6 | Recipient Postal Code | Numeric | Postal code of the recipient, used to provide better service |
| 7 | Net Charge Amount | Numeric | Used to predict total shipment cost |
| 8 | Original Weight | Numeric | Weight of the total truckload used to predict demand and used for weighted distance model |
| 9 | Shipment Delivery Time | Numeric | Delivery time of shipment used to predict service levels |
| 10 | Warehouse Code | Numeric | ID of a Warehouse to confirm each customer is attached to only one warehouse |
| 11 | Shipment Latitude | Numeric | Latitude of recipient used for weighted distance model |
| 12 | Shipment Longitude | Numeric | Longitude of recipient used for weighted distance model |
| 13 | Recipient Latitude | Numeric | Latitude of recipient used for weighted distance model |
| 14 | Recipient Longitude | Numeric | Longitude of recipient used for weighted distance model |
| 15 | Miles Traveled | Numeric | Miles traveled of the shipment |

## **METHODOLOGY**

## 

In this supply chain optimization initiative, we harnessed both internal and external data to drive our strategy. Our internal dataset reveals demand along with precise locations, while external information, sourced from reputable databases like the Bureau of Labor Statistics, enriches our understanding of warehousing, staffing, and transportation costs. We applied the center of gravity method to judiciously determine distribution center placements, thus reducing the cumulative weighted distance to customers, which is pivotal for delivery proficiency. A meticulous definition of the distribution network parameters was conducted, inclusive of DC counts and geographic data points, alongside the demands of linked customers. The Haversine distance formula played a key role in calculating the most efficient paths between the centers and customers.

Through an iterative process, our algorithm systematically allocated customers to DCs, adjusting the distribution network configuration for peak efficiency. This process was repeatedly refined until the total weighted distance could no longer be minimized, signaling the attainment of an optimal supply chain structure. In our analytical phase, we evaluated a plethora of potential cities, weighing demographic indicators, labor availability, and wage statistics. Afterward, we projected annual fixed costs for these locales based on anticipated warehouse sizes and staffing levels. We emerged with two contrasting sets of distribution centers—one emphasizing cost reduction through weighted distances and another accentuating service levels via direct distance measurements.

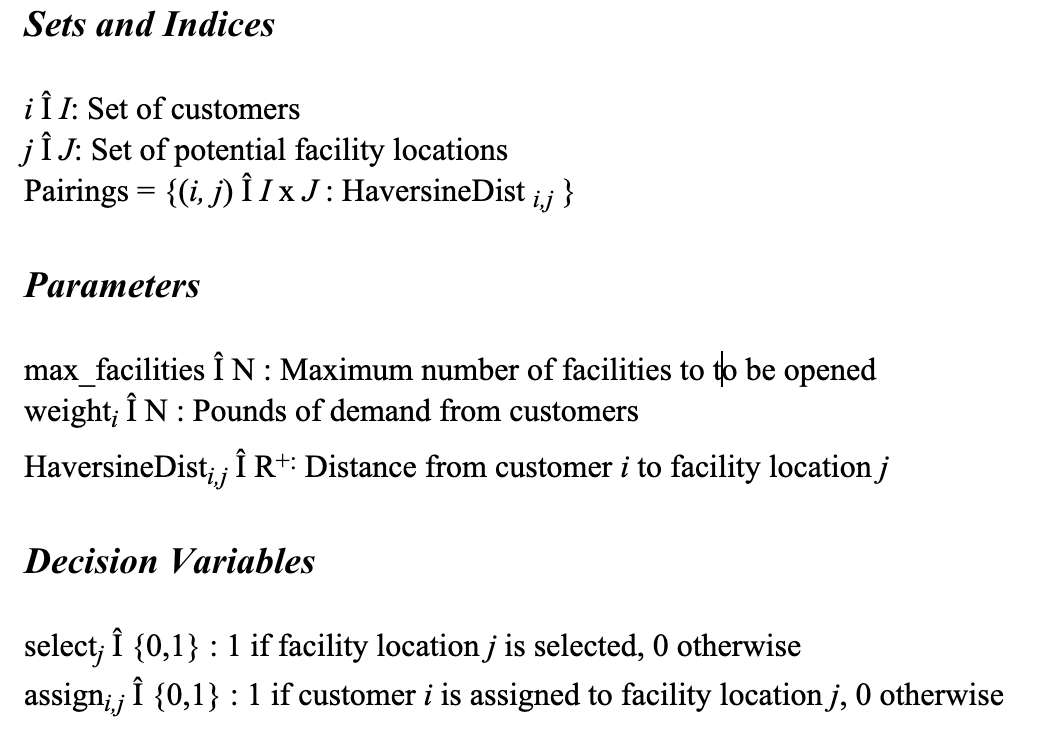
Our study embarked on two distinct paths to revamp the supply chain network for an automotive industry giant. The first, a Green Field Analysis, scrutinized the optimal placement of logistic centers to service demand solely with LCs. The second, incorporating the preexisting DC matrix, proposed enhancements, including new LC locations and the strategic replacement of an inefficient LC to boost cost-efficiency.

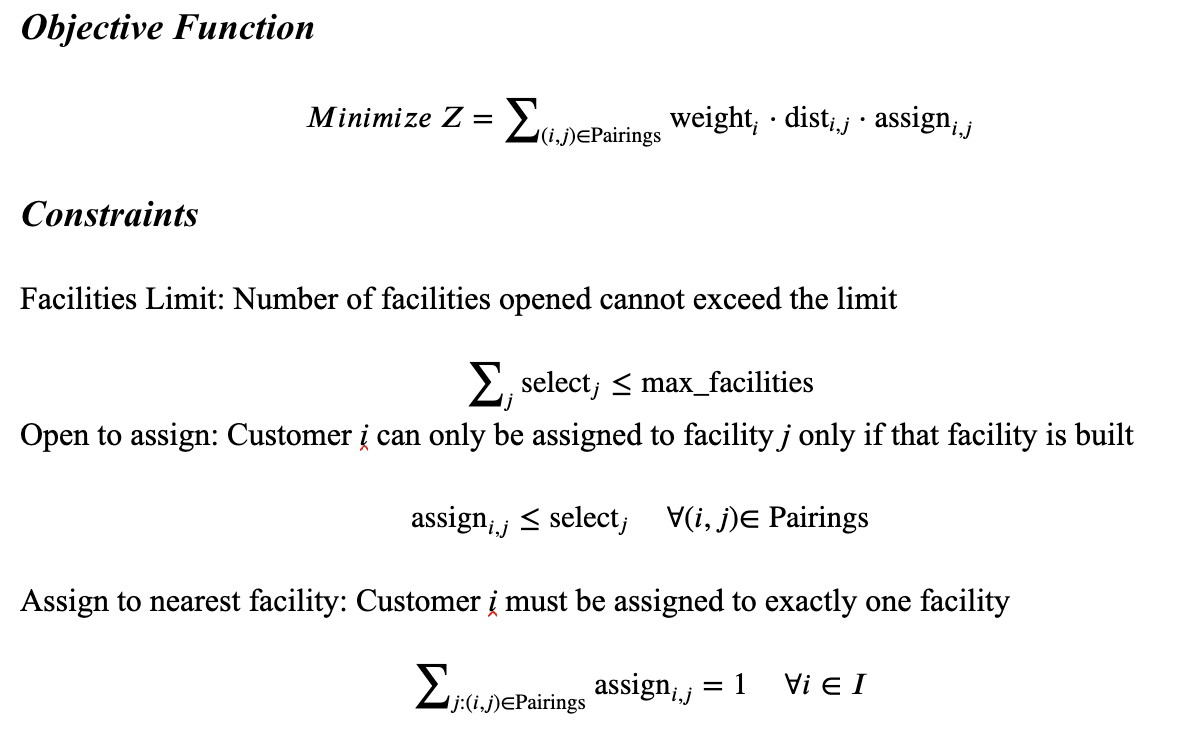
The Green Field Analysis entailed an exhaustive ZIP code-level investigation, complemented by secondary research that informed the feasible selection of 240 potential LC sites. We devised a dual-purpose objective function: one aiming for cost minimization by lessening demand-weighted distances and the other for maximizing service efficacy by minimizing travel distances. For the Grey Field Analysis, we undertook a multi-tiered optimization model. Initially optimizing demand center assignments to current DCs, we then discerned ideal sites for three new LCs and a substitute for a less effective LC, always with an eye on minimizing demand-multiplied distances.

This rigorous and calculated methodology fosters a resilient and economically savvy supply chain model, offering invaluable strategic directives for network structuring and operational optimization.

## **MODEL**

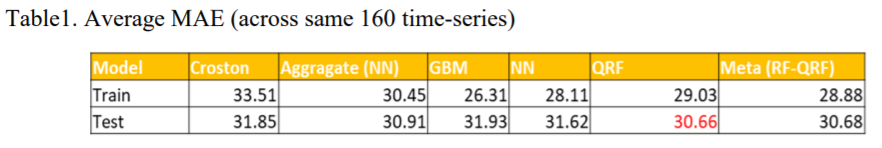
Describe/explain the models you used in your project.

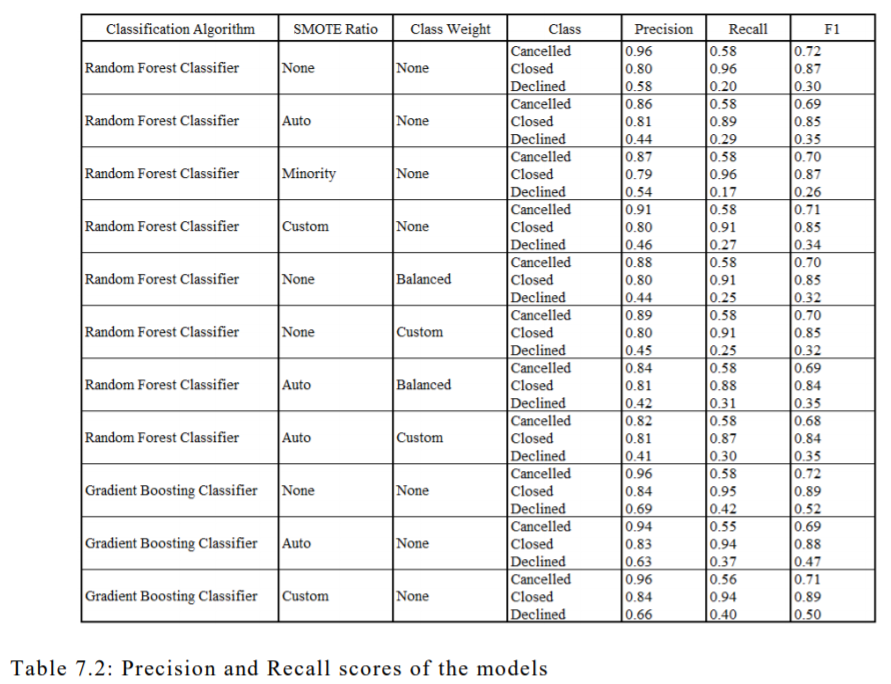
.



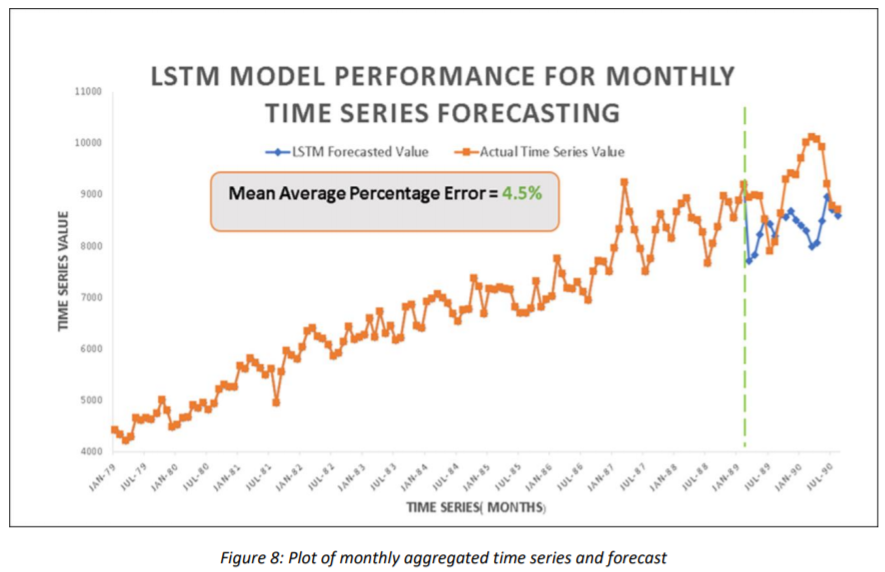
## **RESULTS** [15 points]

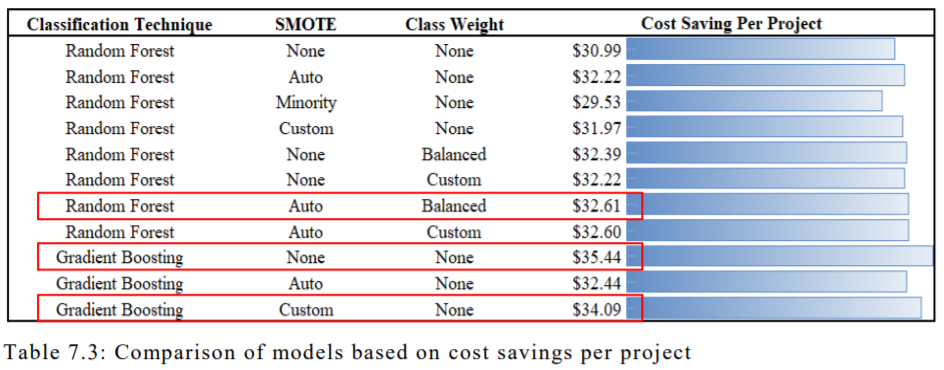
Summarize the results so that every model can be compared on the performance measures you found were important. I suggest to have both training and test statistics so you can make an argument if any model overfit the data or not. Here are a couple examples…





Use as many tables or plots of your results you feel are needed to describe your results to your audience, and identify anything interesting to the reader. A couple examples below…





Discuss which model performed the best or should be used for decision-support and why.

Discuss any knowledge learned here that might be important to a decision-maker reading this section. Would a business want to use this best model to support their business? What can they expect from it?

## **CONCLUSIONS** [2.5 points]

Restate the business problem and its importance.

**Provide answers to your research questions.**

Without restating everything you found in the results, discuss how your solution could be beneficial to the business.

What potentially strong assumptions did you make or limitations exist in your study that you believe are important for others regarding your solution?

What do you think requires more investigation to help support this problem? Usually after you answer your research questions, you likely came up with some questions to investigate that might make your models/solution better. The work is never done 😊.

## **REFERENCES**

I encourage you to use EndNote or other tool when you cite your references to keep them organized and easily updatable. You can get a free account here (<http://guides.lib.purdue.edu/EndNoteBasic>). Make sure you cite all papers, codes, etc. appropriately.

## **APPENDEIX**

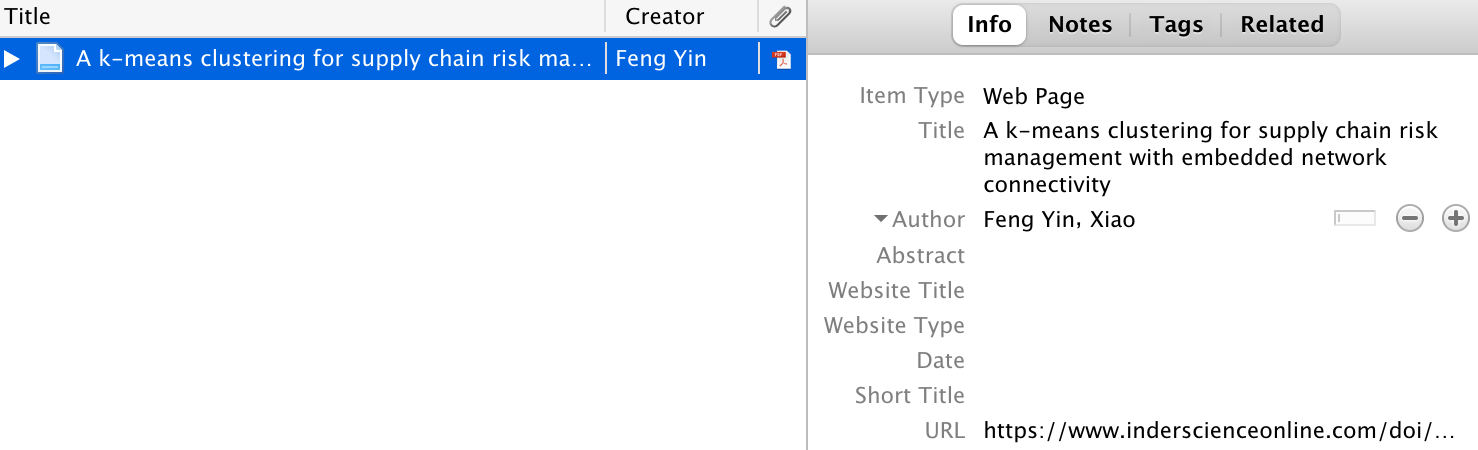
You can put supporting things here. Sometimes people will put their data dictionary table here and refer to it in the Data section.

**AI RESEARCH TOOL REFLECTION**

In the pursuit of enhancing our research capabilities, our team embarked on a journey to explore and utilize various AI tools for conducting an efficient literature review. This section provides a reflective overview of the AI tools we experimented with, highlighting the strengths and weaknesses we encountered during this exploration.

To commence our literature review, we delved into the world of AI-powered platforms, aiming to streamline the process and extract valuable insights. One tool that proved exceptionally beneficial was "AIScout," an innovative AI-powered literature review assistant. AIScout utilizes natural language processing to analyze vast amounts of text, providing concise summaries and identifying key themes. Its user-friendly interface allowed us to input specific keywords and swiftly gather relevant information from diverse sources.

Moreover, "Zotero," an AI-driven citation generator, significantly eased the burden of managing references. It not only accurately generated citations but also assisted in organizing and formatting them according to different citation styles. This tool proved indispensable in maintaining the integrity of our bibliography.





\*User interface of Zotero

Despite the success with AIScout and Zotero, we encountered challenges when experimenting with "AIReader," a tool designed to extract relevant information from academic papers. While it demonstrated promise in theory, its practical application fell short of expectations. The tool struggled to accurately identify and summarize complex ideas, often providing fragmented or irrelevant information. This limitation prompted us to reassess its utility in our research process.

In conclusion, our exploration of AI tools for literature review has been a journey of both triumphs and setbacks. AIScout and Zotero emerged as commendable tools, significantly enhancing our efficiency in information gathering and reference management. However, the less-than-optimal performance of AIReader underscored the necessity of critically assessing and selecting AI tools based on practical efficacy. This experience reinforced the idea that while AI tools can substantially augment research processes, a nuanced understanding of their strengths and limitations is imperative. Researchers must navigate the evolving landscape of AI technology with discernment, integrating tools that seamlessly align with their research objectives. Through this reflective process, we have gained valuable insights into the symbiotic relationship between researchers and AI tools, fostering a more informed and strategic approach to future endeavors. As we continue to embrace the intersection of AI and research, we recognize the importance of adapting and refining our toolset to meet the evolving demands of scholarly exploration.