Capstone Project Report

on

Solving Spatial Demand Optimization Problems Using Automated Workflows in Alteryx, Tableau, and SQL

(Submitted for the partial fulfilment of the requirements for the award of the Executive Master of Business Administration in Data Science and Analytics)

Submitted By:

Name: Shaleen Sharma **Roll No: 23/UEMBA/10**

Under the guidance of:

Dr. Kusum Lata, Assistant Professor

UNIVERSITY SCHOOL OF MANAGEMENT & ENTREPRENEURSHIP

Delhi Technological University

East Delhi Campus, Vivek Vihar, Phase 2, Delhi-110095

July 2025

CERTIFICATE FROM GUIDE

This is to certify that the project titled " Solving Spatial Demand Optimization Problems Using Automated Workflows in Alteryx, Tableau, and SQL" submitted in partial fulfilment of the requirements for the Executive Master of Business Administration in Data Science and Analytics by Shaleen Sharma at the University School of Management & Entrepreneurship, Delhi Technological University is a record of original research work carried out by him.

____

Shaleen Sharma

Roll No. 23/UEMBA/10

Executive MBA in Data Science and Analytics

Batch 2023-25

CERTIFICATE OF ORIGINALITY

This is to certify that the above-mentioned project titled "Solving Spatial Demand Optimization Problems Using Automated Workflows in Alteryx, Tableau, and SQL" submitted by Shaleen Sharma, Roll No. 23/UEMBA/10, has been carried out under my supervision.

Project Guide

____

Dr. Kusum Lata

Delhi Technological University

ACKNOWLEDGEMENT

I would like to express my gratitude to my project guide, Dr. Kusum Lata, for the invaluable opportunity to work under his expert guidance. Her dedication and support throughout the project have been instrumental in my learning journey.

I am also deeply appreciative of the faculty members at Delhi Technical University for their unwavering support and assistance. Additionally, I extend my thanks to the Prof. Manoj Kumar Sharma & staff at University School of Management and Entrepreneurship, Delhi Technological University for always supporting and guiding me. Your collective efforts have greatly enriched my experience, and I am truly grateful.

NAME and SIGNATURE OF STUDNET

Declaration

I hereby declare that the project work entitled "Solving Spatial Demand Optimization Problems Using Automated Workflows in Alteryx, Tableau, and SQL" submitted to the USME, DTU, is a record of an original work done by me under the guidance of Dr. Kusum Lata, Professor & Program Coordinator, DTU and this project work is submitted in the partial fulfilment of the requirements for the award of-the degree of Master of Business Administration in Data Science And Analytics. This thesis has not been submitted to any other University or Institute for the award of any degree or diploma.

Abstract

In today's data-driven business environment, organizations are increasingly adopting AI and automation to enhance operational efficiency and customer satisfaction. As customer expectations evolve toward faster and more convenient services, there is a growing need for intelligent systems that can optimize operations based on spatial and demand-related data. This project addresses the challenge of solving spatial demand optimization problems using automated workflows built with low-code tools.

The study focuses on developing a scalable, automated solution using Alteryx for data transformation and logic, Tableau for clustering and geospatial visualization, and SQL for data storage and integration. Using only minimal inputs store-level latitude, longitude, and demand values the workflow identifies high-demand zones and strategically allocates resources based on clustering analysis and weighted spatial logic.

Despite limitations in available data, the solution successfully enables decision-making for spatial demand optimization, offering insights into how low-code platforms can be leveraged for strategic planning. The outcome demonstrates that even with limited datasets, meaningful business insights can be achieved through the right combination of tools and analytical methods.

Keywords: Alteryx, Tableau, High-Demand Zones, Low-Code Platforms, Business Insights,warehouse, Spatial Demand Optimization

TABLE OF CONTENTS

Contents

LIST OF TABLES

LIST OF FIGURES

LIST OF ABREVIATIONS

Abbreviation	Full Form
AI	Artificial Intelligence
SQL	Structured Query Language
GIS	Geographic Information System
BI	Business Intelligence
SME	Small and Medium-sized Enterprises
KPI	Key Performance Indicator
API	Application Programming Interface
UX	User Experience
IT	Information Technology
ML	Machine Learning
GPS	Global Positioning System
ETL	Extract, Transform, Load
CSV	Comma-Separated Values
UI	User Interface
EDA	Exploratory Data Analysis
ERP	Enterprise Resource Planning
DB	Database
RDBMS	Relational Database Management System
K-means	K-means Clustering Algorithm

CHAPTER 1 - INTRODUCTION

In the recent decades the retail environment and the way people approach it have changed dramatically. Shopping was a local and easy task in the past. The consumers used to go to small vendors or Kirana stores located near their place to buy groceries and other household items. These small shops were widely engaged in neighborhoods and were generally run by the people who people knew of, and transactions were based on trust and personal connections. Taking into consideration the concept of shopping, it was not strategized, but an everyday activity done out of convenience.

With the evolution of urbanization, retail experience started to change. Supermarkets and departmental stores became bigger and located in other areas of the city and the assortment of goods under a roof was increased. This change became the precursor of formal retailing formats. Shoppers started arriving more frequently every week or every month to these stores where they found it convenient with the larger choice of goods and other systematic display of goods. This model was further enhanced by shopping plazas and hypermarkets such as; Big Bazaar and Reliance Fresh. These retail stores were strategially placed in places like metro, commercial centre and they had everything, food, electronics and so on in one stop.

Figure 1 – Importance of Warehouse Location

With the emergence of the e-commerce platform, the retail sector witnessed a more significant disruption. The online shopping has brought about a different dimension in consumer behavior. When the entry of such platforms as Amazon and Flipkart occurred, consumers started establishing orders online and getting deliveries a few days later. This shift was not just a matter of convenience but itself reshaped the expectations. Customers would be able to shop and compare prices, read reviews, follow delivery and experience home delivery without going out. E-commerce introduced a scale, range, and visibility in purchasing.

Figure 2 - Evolution of Retail Shopping

The development of the e-commerce environment promoted thousands of new start-ups and specialized platforms to appear in the market. Digital-first companies started competing based on elements such as: speed of deliveries, product quality, the level of user experience, guarantee conditions in the case of returns, whether it was fashion, electronics, groceries, or lifestyle products. This is a race that spurred on fast advances in logistics, warehousing, customer services and older technology. Promises of delivery time that used to take five to seven days were brought down to two days, then one day and finally the creation of same a day deliveries in some areas.

The norm of convenience has been even further advanced today with the development of the ultra-fast delivery services. Companies such as Blinkit, Zepto, Swiggy Instamart and Amazon Fresh offer to deliver groceries within 15 minutes. Speed is no longer a luxury in the market, but it is a minimum requirement. The consumers of present days do not only require the possibility of shopping at any place and at any time, but also to have their commodities delivered to them immediately. Any delay, however, minor, can be the cause of dissatisfaction. Responsiveness, precision and convenience have assumed positions as main pillars of customer satisfaction.

To meet this need, companies are making huge spends on destination optimization, decision engine and automation of processes. Among the most important features of this change is location intelligence, i.e., a strategic location of service nodes such as warehouses, dark stores, micro-fulfillment centers. Locating these facilities near the areas of high demand to save the delivery time and the operational cost to a considerable degree. Fast delivery services rely on the ability of a company to devise their network of supply points.

But where to locate services is not a simple thing to locate. Ideally, decision-makers can access smaller details that include population density, trends of delivery patterns in the past, coverage of the road network, and real-time traffic conditions. The truth is that those solid datasets are not always provided off-the-shelf (especially when it comes to smaller businesses or new markets). A lot of companies have to make some critical plans using fewer inputs which could even be just some simple geographic coordinates and combined demands.

This project deals with how optimization of the spatial demand can still be possible despite such restraints. To be precise, it explores the possibility of using the data with very modest details in analytics, like store latitude, store longitude, and aggregate demand, to help in the effective planning of service zones using low-code solutions. It is concentrated on providing data-light, automation-supported, business-friendly solution so that a company could take action even without tremendous IT or data science structure.

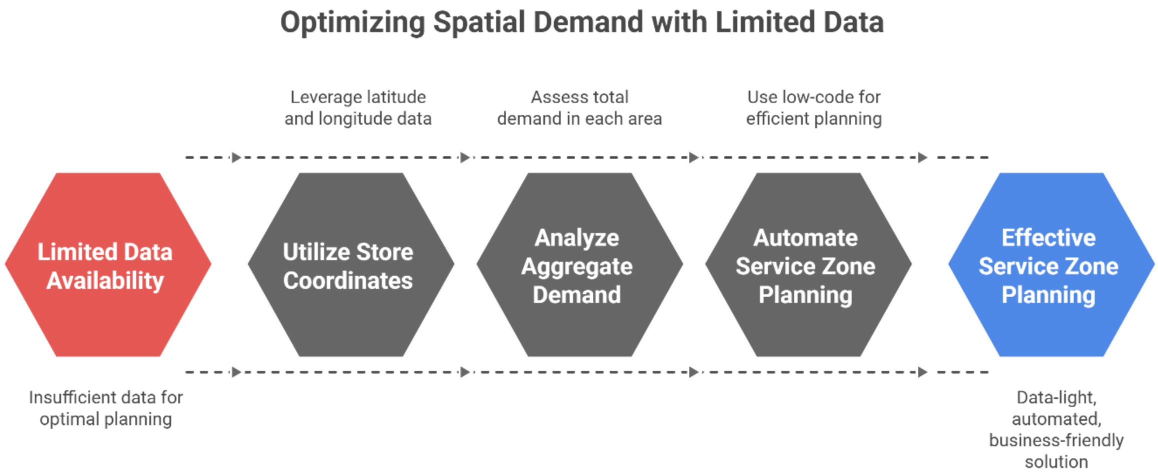


Figure 3 - Optimizing Spatial Demand with Limited Data

It starts with the data preparation of the dataset visualised with a low-code data transformation tool, Alteryx. The dataset normally comprises of a list of stores and the list of geographical locations and demand. Because demand is one of the most important factors that influence service needs, it is applied as weight in further calculations. Data is then filtered to represent the best 80 percent of demand generating stores in accordance with Pareto principle, so as to locate the most influential areas.

Then Tableau will be employed to perform K-means clustering according to geographic co-ordinates of these stores with high demands. The option of clustering groups contains stores that are in the same space. This technique is used during the process of determining natural clusters of the demand centers that can efficiently be served using an easily reached location. The individual stores are given cluster IDs and the clusters are regarded as possible service areas.

After this the clustered data is imported back into Alteryx, where weighted centroids are generated. This entails temperature calculation of average latitude and longitude per cluster, depending on demand of each store. The bigger the demand of the stores, the greater is the influence on the centroid position. These centroid points are the most desirable points to establish the service centers as to cut down the travel that also causes maximum responsiveness.

Not every cluster can be viable. Thus, such spatial validation tools as Create Point and Spatial Match in Alteryx are used. Clusters which are beyond serviceable distances, e.g. within a city boundary or within a specified delivery zone are eliminated in the analysis. This will make sure that only feasible and strategically interesting clusters are maintained.

Following the verification of the zones positions, the activity dwells on the completion of the network. The stores that did not fall in the 80 percent-demand first group and/or falling outside a cluster will be reassigned according to proximity. The distance between each store and the final zone centroid is computed with the help of the Haversine formula, which comprises the geographic distance between the two points. The nearest centroid is then used to assign every store, and the whole network was covered. The approach is both cost-effective and positionally legitimate.

When the analysis is complete, the augmented dataset (which will have zone assignments, centroids and distance measures) is exported to a SQL database. This move makes it easier to scale and integrate. One can report, dashboard or feed other supply chain systems using the structured database. The most notable thing is that it is a reference point of choice to planners and decision-makers.

The last thing in the project is to develop an interactive Tableau dashboard.

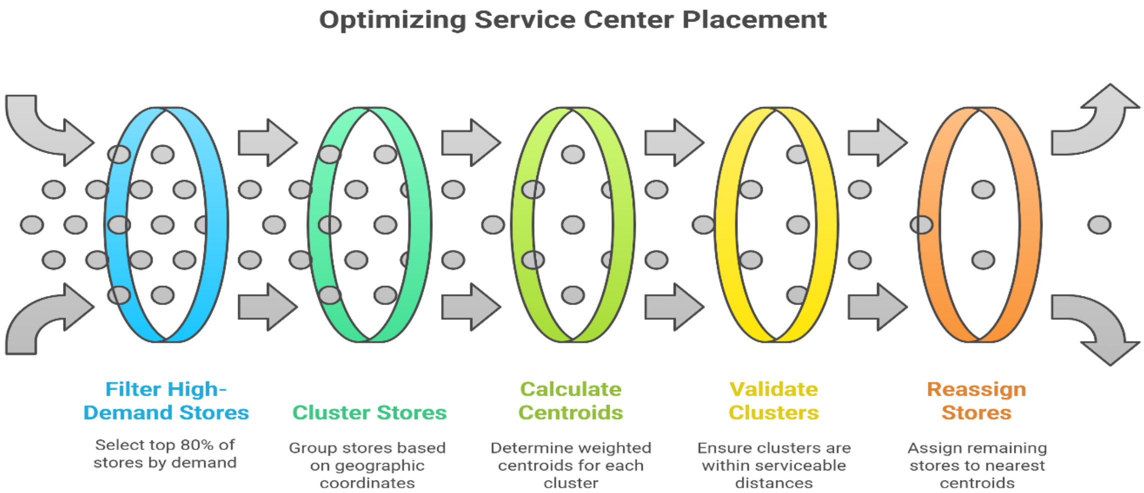


Figure 4 - Optimizing Service Center Placement

This visualization has a connection to the SQL database and loads a map displaying all store locations, reserves dedicated to each of the stores, and lines between stores. Dynamic spatial visuals are generated by use of calculated fields like MAKEPOINT and MAKELINE. Other charts and filters provide an opportunity to focus on analyzing zone-wise performance, the coverage of demand, and service efficiency. To indicate what zones get the greatest demand or serve the biggest amount of stores, non-standard funnel charts are applied.

The ultimate outcome is the fully-computerized, repeatable, and scalable workflow that aids in optimizing the spatial logistics with only three data fields. This project demonstrates the potential of having little data and using the available tools and sense to draw business results. It is even eliminating huge infrastructure investments or high level of technical knowhow and thus is open to small scale business use.

Another potential area in this project is the increasing contribution of low-code analytics to the solution of complicated business problems. With the desire of the companies to be fast in responding to market needs and what their customers want, flexibility and speed are essential in modern day decision making, which tools such as Alteryx and Tableau can offer. Data scientists have lost the sole right to exploiting the ability to experiment on various inputs, test hypotheses, and come up with real-time insights. Business users, analysts, and operations managers should also be in a position to assume ownership of analytical processes and come up with innovation within their groups.

At the end, possibility to optimize service areas and delivery networks is turning into paramount competitive edge. The consumers anticipate speedier service more than ever before, and the companies need to invest in intelligent planning solutions that are cost-effective, efficient, and satisfying customers. The results of the project indicate that a business can take empowered, strategic decisions informed using automated workflows and simple visual analytics inputs even when the size of the dataset is small. It can relate well to other organizations that aim to close the gap between the complexity and expectations of the operations in the current highly demanding digital market.

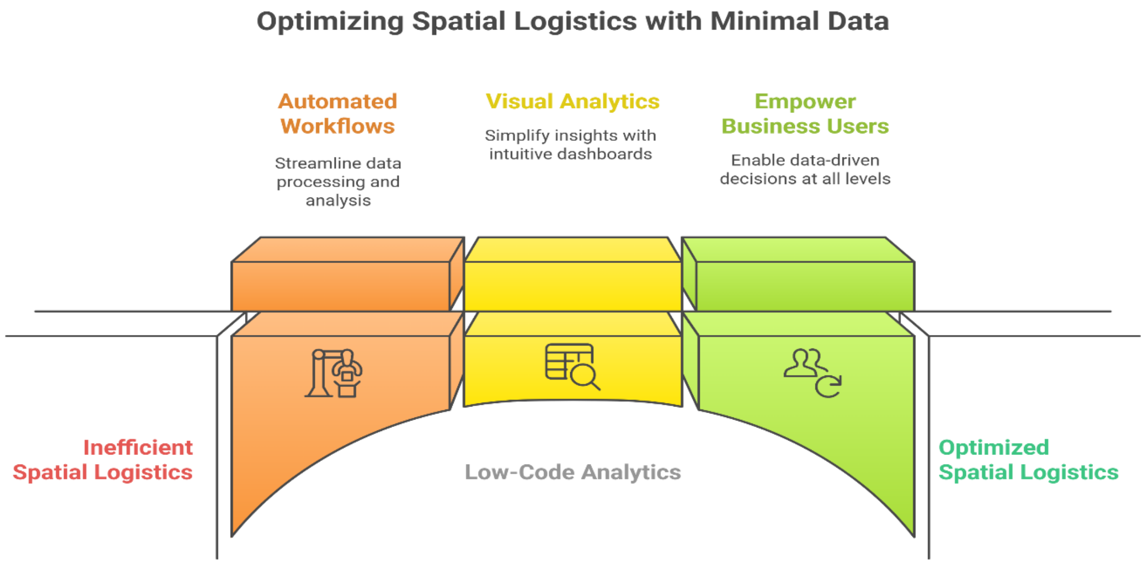


Figure 5 - Optimizing Spatial Logistics with Minimal Data

1.1 Problem Statement

One of the main challenges in this project was the limited data provided. The dataset contained only three columns:

- Latitude of the store
- Longitude of the store
- Total Demand in each store

Key factors such as population density, road connectivity and number of customers were not provided. Since two out of the three columns were purely dimensional, it was challenging task.

1.2 Research Objective

The main objective of this study is to develop a strategic location for our store.

Specifically, this study aims to:

- Identify optimum location for warehouses to effectively meet the store demand
- Reduce delivery time to improve customer convenience.
- To prioritise key business stores that are important for profitability

CHAPTER 2

CHAPTER 2 - LITERATURE REVIEW

In the literature review, some main themes associated with spatial demand optimization and the changing importance of the low-code tools in data analytics are described. It starts by looking into the importance of spatial planning in supply chain efficiency especially when it comes to locating the best service or warehouse locations. Since the proximate demand points-between the warehouses and the customers-can also be clustered using clustering methods such as K-means as shown in the studies, this aspect will also help in improving the responsiveness of the delivery in various ways.

Besides, the review shows that the use of geospatial data is on the rise, particularly in a setting where there is scarcity of data, i.e., latitude, longitude, and demand. There is a significant emphasis on introduction of low-code tools such as Alteryx and Tableau that streamline data processing and visualization efforts and allow advanced analytics to be used by non-experts.

Notwithstanding the current progress that has been made, areas of needs in terms of achieving scalability, automation, and data-light means are evident within the review. Most conventional models require close data input and coding skills and are restricted to practical use. It necessitates the requirement of applications which are flexible, repeatable, and user friendly, and work with low resource levels. The literature provides the basis on which this project will be done, i.e., the low-code will be leveraged to construct scalable and automate workflows in spatial optimisation.

2.1 Spatial Optimization in Supply Chain Planning

Sharma et al. (2024) use the available body of research on the hub location problem (HLP) to develop a comprehensive overview of the main advancements, methods, and trends of the current research to find the future research directions. The proposed method will be the systematic literature review of the scholarly sources on HLPs, with the approaches classified according to problem variant, modeling, solution approaches, and application domains. The results give away striking advancements in modeling complex and realistic problems hub location, such as single and multiple allocation problems, capacitated hubs, and stochastic aspects, as well as the utilization of exact algorithms, heuristics, and metaheuristics as the means of a solution. It is within the same study that the study further identifies that, despite the notable progress in theory and computational development, there still exists gaps in the applications of introducing real-time information, dynamic and multi-period environments, and accounting of sustainability perspectives into the HLPs. Future studies, therefore, need to change the emphasis on these unexplored dimensions that can support the practical challenges and make them more applicable to the developing transportation and logistics networks, as mentioned in the paper.

Singh et al. (2018) to determine the best location of the warehouses to improve efficiency and to minimize dealing with global supply chains. The purpose of the research will be used is a case study approach, where data analysis, decision making models, geographical information systems will be used to analyze various location variables to include transportation costs, closeness to markets, and infrastructure. Results indicate that strategic location of warehouses has added values in real sense regarding better delivery performance and cost containment with reference to regional economic dynamics. Nevertheless, the gap that exists in the study is the inadequate use of dynamic factors such as political instability and the changing trade policies, thus giving indication of the need of incorporating real time risk assessment in future.

As Ghadge et al. (2016) state, optimal locations of facilities that handle both forward and the reverse logistic and manage them efficiently in a closed-loop supply chain are determined. The approach focuses on modeling of Math with the approach of metaheuristic algorithms to solve the multi-objective problem of minimizing costs and maximizing sustainability. It was found that the hybrid strategy approaches are more effective than the classical strategies as it balances the economic and environmental goals, enhances resource recovery and minimizes the wastes. Nevertheless, the limitation that was identified is diminishing this real-world validation and scalability of the model meaning that there will be a necessity of future research to test the approach in various industrial environments and at larger-scale networks.

Zadeh and Garay-Rondero (2023) discuss the possibility of decreasing the carbon footprint by using innovative strategies of supply chain in a smart urban setting. The methodology involves the use of a combination of case studies, data analytics and sustainability assessment frameworks in investigating the effect of integration of smart technology, optimization of logistics as well as green procurement on urban carbon emissions. The results have shown that the application of improved supply chains reduces the emission levels and maximizes the utilization of resources in smart cities. The missing link is, however, the fact that social acceptance and policy implications are under-analyzed with a call over having wider research between disciplines to help in transformation of urban supply chains towards sustainability.

The article by Modgil et al. (2021) analyzes the ways through which AI technologies can be used to increase the robustness and adaptability of the supply chains in the context of a crisis such as the pandemic. To conduct them, the methodology implies the qualitative analysis of the case studies, literature search, and assessment of the AI applications efforts like predictive analytics, demand forecasting, and risk management that have been realized in context of the Covid-19 disruptions. The results have shown that AI increases the visibility of the supply chain, the speed of decision making, and the mitigation of risk boosting resilience. Nevertheless, the identified gap is the absence of common structures of the AI implementation and the scarce research of the ethics and data privacy issues, which prompts that future research must integrate the challenges to achieve its widespread use.

Cataldo et.al (2022) reveal the research gaps in the area of sustainability in construction supply chains and study the existing trends. The research methodology includes a systemic literature review, which analyzes articles and industry reports using academic, industry, and government reports on the topic of sustainability construction SCM in relation to environmental, social, and economic aspects. Results indicate an increasing focus on green materials, minimization of wastes and cooperation, but present mixed implementation and evaluation of sustainability procedures. Their absence is in empirical research, particularly that relating Digital technology integration and stakeholder engagement with the main need in future research to come up with practical frameworks and evaluate long-term effects of digital technology integration on construction supply chain sustainability.

The purpose of the proposed research study is to design a successful decision-making model to determine the best destination of a cross-dock terminal to advance the effectiveness of supply chain. The approach has a party multi-step method that combines CI-DEA in efficiency measurement, IDOCRIW when dealing with uncertainty in weighting the criteria, and the MABAC in the final ranking through a real life case study. Results indicate that such combined model offers a strong, precise, and adaptable complex location tool. The identified gap however is that not much exploration is done on dynamic market conditions and environment hence a need to explore adaptive models in different evolving supply chain contexts in the future.

Castillo-Villar (2014) study use of metaheuristic algorithms to optimize the different problems in the bioenergy supply chain. This is done through methodology of an extensive literature review that examines the nature of application of program such as genetic algorithms, particle swarm optimization, and ant colony optimization to solve problems in the field of facility location, logistics and resource allocation. The results show that such algorithms are practical to improve the quality of solutions and computational performance, but such issues as problem complexity and parameterization are perceived. The shortcoming is due to a lack of connections with real-time suggestions and sustainability indices, and future studies should focus on the formation of dynamic and large-scale bioenergy supply networks by creating adaptive, hybrid metaheuristic optimisation models.

2.2 Clustering Techniques in Location Analytics

Khalid and Herbert-Hansen (2018) targets the deployment of k-means clustering in the high-quality international location decision-making process. The methodology implies searching most geographic, economic and market-related factors through clustering of the potential locations according to their similarities and allows making a clearer comparison and selection of the strategic locations. The results have revealed the effectiveness of the tool of k-means in grouping different locations, so that decision-makers could know which locations have similar features and which locations to select first based on the business goals. The limitation is however in the sensitivity of the method to absolute numbers of clusters and non-integration with the qualitative side meaning that future studies should engage in integration of clustering alongside multi-criteria decision analysis in order to produce a more measured international location strategy.

The aim of the study by Miraftabzadeh et. al (2023) examination of K-Means and other Clustering applications in the modern power system is to assess and compare clustering algorithms to an effective data analysis of the system. This methodology will involve the use of K-Means and other clustering algorithms like hierarchical and density based algorithms to power system data in order to find some pattern, anomalies, and groupings in the operating behaviors. It has been found that the use of K-Means though computationally economical and most common also has its drawback of not capturing the non-linear and complex data structure of current power system operations to the full. There is however still a gap in coming up with adaptive clustering techniques that would effectively fit into dynamically and real-time power system conditions in a robust yet more accurate manner.

Liao and Guo (2008) examined the Capacitated Facility Location Problem (CFLP) through a clustering-based approach because efficient solutions are sought to determine an optimal location of a facility with regard to its capacity. This is carried out by breaking down the problem by employing clustering techniques to randomly cluster the customer demands together firstly in space and the location and allocation problem is then solved which makes the problem less complex. Results indicate that the clustering-based algorithm enhances the quality of the solution as well as the level of computational performance in comparison with conventional optimization procedures and continuously balanced facility capacities and customer assigning. The missing part is on how to handle very dynamic or large-scale data sets where the clustering accuracy and the scalability may need to be improved further in how they may need to manipulate real world variability and complexity of decision making on facility site locations.

Abo-Elyousr et al. (2022) minimise the joint optimisation of distributed generation (DG) and electric vehicle (EV) parking lots combined with demand response programmes to increase grid efficiency. The approach is a fusion between K-means clustering to group load profiles and a Particle Swarm Optimization (PSO) algorithm that is self-adjusted to identifying optimal scheduling solutions to balance the supply and demand. Results indicate enhanced energy regulation, lower cost of operations and grid stability due to flexible scheduling. The thing is that it lacks handling real-time uncertainties and dynamic behavior of the EV users and renewable generation, which suggests that more adaptation in real-time optimization frameworks, is needed to adapt to the turbulences in the natural settings.

To improve the choice of demand analysis and forecasting, Azad et.al (2014) define common electricity consumptions patterns based on the K-means clustering technique. The algorithm procedure is to treat data containing grouped electricity usage using K-means, cluster consumers or time slots by characteristics into groups that represent commonality in the load pattern. The results indicate that K-means can also be considered as a good method to extract the representative load profiles that can help understand the consumption behavior in a better way and in the effective improvement of demand side management strategies. Nevertheless, that is where the discrepancy is founded since K-means have limitations that include being sensitive to the initial cluster centers and challenging to process non-spherical or overlapping groups without clarifying the use of more pronounced or mixed clustering methodologies that serve better and meaningful recognition of a load profile.

Chicco (2012) give a survey and benchmark of different clustering algorithms to cluster electrical load patterns so as to enhance load forecasts and demand management. The approach includes a systematic application and comparison of various clustering algorithms, such as K-means, hierarchical and density based to electrical load data so as to assess their capability to form patterns in an effort to identify the most appropriate and suitability. The evidence shows that although K-means is simple and fast, there is a possibility that other methods could be used to detect complex load behaviors and outliers. The allocation of the gap lies in the fact that there is no clustering strategy optimal in every situation, and this area requires adequate attention toward the adaptive, hybrid strategies that would fit a wide range of various load properties and therefore improve clustering quality, due to its application in practice.

2.3 Importance of Geospatial Data in Business Intelligence

Andrienko et al. (2007) set a research plan to transform geovisual analytics as a system of spatial decision support to improve the interpretation and analysis of a spatial data. It consists of the overview of the existing geovisual analytics methods, determination of their main challenges and the development of future directions of research involving visual analytics and spatial reasoning and interactive technologies. Results indicate the promise of geovisual analytics to enhance decision making through exploratory data analysis through dynamic visualizations combined with real-time data processing as well as the deficiencies concerning the complex spatial-temporal data processing, user-focused design, and scaling. The research recommends the need to employ the use of interdisciplinary approaches to realize these challenges and come up with better spatial decision support systems.

The study by Erskine et al. (2013) comes up with a detailed research structure within which business can utilize geospatial information to make better decisions. The strategy is to have a literature search, to study the available methods, instruments, and geospatial data applications in the business environment. Results demonstrate that geospatial data enhances the quality of decisions to a great extent through spatial information on market analysis, logistics, and customer segmentation. Nevertheless, the research finds that there is a gap in geospatial analytics integration with classical business intelligence and that the studies of issues related to bad data and data confidentiality, on-demand processing and high-tech frameworks, which address these concerns and integrate the geospatial data into business decision-making, are more focused.

Daly (2006) set principles that can be used in determining suitability of the spatial climate data sets, in relation to different environmental and climate related uses. The methodology includes specific challenging aspects achieved through the analysis of various spatial climate data sets on the background of the spatial resolution, the time period, the accuracy of data, and its applicability on the background of certain research or operational demands. The results indicate that the appropriateness of data differs greatly between context of application, and there is a trade-off between coverage and detail, and a comprehensive evaluation based on standardized guidelines will enhance selection of dataset. The identified gap is missing application-specific assessment protocols that have universal agreement, and there is a clear need to develop more application-specific transparency frameworks that would help users choose the right data to use in various applications in the field of spatial climate data.

To improve support of spatial decision-making, Bdard and Marchand (2001) also specify the key features of the Spatial On-Line Analytical Processing (SOLAP). It is a methodological approach wherein existing SOLAP system and the needs of spatial data will be examined so that the main aspects that could facilitate easy processing of multidimensional spatial data and visualization are determined. The results indicate the promise of SOLAP in enhancing spatial data exploration and decision support with rapid query response and more easy to understand interfaces, as well as shows that there are difficulties integrating complex spatial data elements into the more conventional OLAP models. The shortcoming is that there is a need of having some standardized SOLAP architectures and scalable solutions that can be much more effective in terms of maximising the benefits of the large scale heterogeneous spatial data used in making decision.

Providing a spatial assistance on the development of aquaculture and its management, Nath et al. (2000) investigate their opportunities that are based on Geographical Information Systems (GIS). The methodology consists of the review of case studies and available GIS facilities that can be studied in relation to the space data regarding the site selection, environmental surveillance, and management of resources in aquaculture. The result shows that GIS can strengthen the decision-making process by combining various spatial data, a better evaluation of the suitability of sites and promote sustainable aquaculture. But a gap has been observed as not much real time data is integrated with advanced modeling in GIS to make the dynamic aquaculture environment functioning, and this means the need to have more flexible and forecast spatial decision support system in relation to aquaculture.

Fotheringham et al. (2000) discuss sophisticated ways to compute spatial data that aims at accomplishing a more in-depth knowledge of the geographic patterns and processes. The methodology mostly consists of statistical methods, spatial modeling and computer algorithms to handle and analyze massive spatial data. Results demonstrate the efficiency of such quantitative methods in the realization of complicated spatial associations and decision-making domain in estate control, environmental study, and social studies. Nevertheless, there is a wide vacuum with respect to the interdisciplinarity between the qualitative and quantitative spatial analysis and hence, there is a need to incorporate the qualitative insights with quantitative analysis in the future so as to holistically comprehend the socio-spatial phenomena.

2.4 Role of Low-Code Tools in Analytical Workflows

The emergence of low-code platform has changed business relationship with data analysis, as well as data automation. The tools allow their users to develop, run, and control data flows with limited-to-no insights or expertise in programming, which not only increases the accessibility of analytics, but also enables the use of technology in more industries. Previously, companies could only rely on IT departments or other versed developers to make their data clean, run models, and produce reports. However, some tools, such as Alteryx and Tableau, are closing that divide today because they enable non-technical people to access data, automate workflows and deliver insights on their own.

Alteryx is a popular tool to develop automated data workflow. It provides input, transformation, blending, filtering, summarizing, and exporting of data with a wide range of tools and drag-and-drop interface. When such can be created once, they could be saved and then utilized again and again and changed without any coding. Alteryx does this better when handling structured data that must be cleansed, enriched, and processed on a regular basis. Users have the options of creating repeatable workflow patterns that can automatically process large sets of data and take lesser time and effort that would be needed in manual processing.

Alteryx has one of the most powerful features, such as the capability of using/managing spatial data. It has included features to develop geographic points using latitude and longitude, compute the distance between sites, spatially join, and establish customized geographic boundaries. It is extremely valuable to projects involving the location-based analysis, e.g. store network optimization, warehouse location planning, delivery area planning etc. Rather than being contingent upon externally developed GIS tools or libraries, or relying upon code development, users can complete more advanced geospatial analysis in its entirety in Alteryx.

Alteryx also assists logic-based tasks by formulae that could be simply implemented using visual based interface. As an example, users could place a weighting on fields, use conditional rules, or provide ranking of entries using business designated parameters. The characteristics contribute to making sure that business logic has been incorporated into data preparation processes. It [the platform] also interacts well with different databases, networks, APIs, and external tools, so it makes the transfer of data and connectivity easier and smoother.

By contrast, Tableau model is most renowned due to its effectiveness of transforming processed information into interactive representations. It offers an extremely user-friendly user interface that allows the creation of charts, graphs, maps and dashboards through the drag and drop of data fields. Tableau has a rich chart types and customization capabilities where a user can use to adjust visuals to their particular requirements in business. Its dashboard features enable the stakeholders to view trends, filter information and view Shots in real-time without prior specialist knowledge.

One option of Tableau that can be handy is the geospatial visualization feature. Customers are allowed to create maps of views with spatial aspects of the data, including store locations, customer density, or delivery areas. Tableau, when used together with calculated fields, makes spatial analysis powerful. It also provides inbuilt clustering capabilities like K-means clustering whereby clustering of data can be done in regard to neighborhood or similarity. These groupings may then be presented pictorially either on maps or other charts and provide on the spot analysis of the regional trends.

The advantage of Tableau is that it allows direct connection to live or data sources. Automatic visualization in Tableau dashboards can be used when connected to workflow/data processing and transformation platforms such as Alteryx. This allows to put together entirely automated analytical pipelines where new data will be ingested, analyzed and visualized without any human interaction. These types of automation minimize the chances of making human errors and boost the effectiveness of decision-making procedures.

Alteryx and Tableau combined provide an ecosystem where one can process and visualize data. Alteryx (or any other backend tool) performs the backend activities; Data Cleaning, applying logic, clustering, and preparing. And Tableau presents those findings to the world in an easy-to-understand dashboard and a graphical picture. The mixture is highly practical in applications that require spatial demand optimization, as both transformation and visualizing knowledge are the keys to success.

Among the greater benefits of low-code tools, one can distinguish the fact that they give non-programmers a chance to contribute to analytics. The business users who have profound knowledge of the field but do not possess technical expertise can now construct and apply analytics tools where knowledge on coding was essential. This makes data more democratic and organizational-wide by creating a culture of data. Rather than the work truly being associated with data teams, either marketing managers, operations planners, or supply chain analysts can discover data and derive solutions independently through a low-code platform.

The other strength of such tools is the possibility to automate routine procedures. The same tasks that are repeated often such as weekly reports, KPI monitoring, or the switching of zones may have to be updated manually, by scripting or that the IT department will need to support. One can automate these tasks with Alteryx and Tableau. Workflows can be scheduled to repeat at fixed interval once created and dashboards can be refreshed automatically with data. This is particularly useful in changing landscapes such as retail and logistics, where the insights should be timely.

Low-code tools are more accommodating of the utilization of limited datasets. Structuring and detailed data cannot be accessed by many organizations and yet their operations require some important decisions to be made. Alteryx and Tableau allow analyzing such datasets (which usually have coordinates and volume data as their sole input) through the usage of clustering and filtering and spatial reasoning. As an example, business can possess only location of the stores and total demand, and still, it can shine high-priority areas and allocate resources efficiently.

The other advantage of low-code platforms is scalability. With increased business requirement, the workflows may be scaled in order to accommodate a bigger dataset or to compute the new parameters. The solution does not have to be redesigned completely. Also, teams work together more unified as working processes and dashboards are visual and self-contained. This increases transparency and wasted time ushering in new people on board or training new members of the staff.

To conclude, the contribution of low-code such as Alteryx and Tableau is playing and becoming more significant in the analytics sphere. These frameworks are scalable, performant, and user-friendly data transformation and visualization solutions. They limit the consumption of technical resources, decrease time to insight and enable high-level analytical work to be introduced to a broader audience. Their functions are especially useful in the environment of spatial demand optimization. Even using little data, businesses can conduct meaningful analysis, generate automated, repeatable workflows that will enable them to plan better and make smarter decisions.

2.5 Research Gap

Though significant improvements in spatial demand optimization have been achieved, it is possible to determine certain important research gaps, especially regarding accessibility, automation of the research and data availability. A lot of the existing models and frameworks in the scholarly literature are based on complex algorithms, high-quality data and large computing power. The solutions are sometimes effective and right in theory but most of the time, they do not work out to be scalable and/or a reality to some organizations based on the lack of technical knowledge or resources. Practically, not every business has access to sophisticated infrastructure and elaborate data sets and consequently lacks the capability of executing such solutions.

Among the most severe gaps, excessive preference to data of high resolution or granularity, including traffic reports on a real-time data feed, customer behavioral activities, or infrastructure mapping, can be cited. Whereas these factors, undoubtedly, increase the precision of the location planning and delivery modeling, they are not always realistic to gather, particularly with regard to small and medium-sized enterprises (SMEs) or those companies that enter new markets. What is left to most of the organizations is usually just mere basic information like store coordinates or past demand data. Sadly, there is little help on how to make operations optimal when there are minimal pieces of information available. This brings out the grain of necessity stressing on simplistic, data-lite models that are implementable and cost-friendly yet applicable to aid in generating insight.

An example of such gaps is that data processing and visualization tools are not integrated in most of the academic studies. It is common that traditional methods associate data transformation, analysis and reporting as unique activities. This break in connection does not only delay decision-making but also creates the possibilities of error and misunderstanding as well. Industry requirements, on the other side, require integrated solutions capable of ingesting, processing and visualizing data on a continuous stream. There is increased demand in frameworks which aggregate these into a unified system- ideally in a low code or no code system to help drive adoption.

Another issue is the scrimpy attention toward repeatability and automation. Most of the prevalent models are also static, i.e., they can only be used once at a time or have to be updated manually at high cost. That is why they are not as applicable in dynamic environments, such as retailing or logistics where data is changed frequently and knowledge bases need to be updated on a regular basis. The scalable solution must be easily replicated, adapted, and scheduled, thus the decisions would not become obsolete. The unspecificity of the focus on repeatability of the current research leaves the room of solutions capable of changing according to the business requirements.

Moreover, most of the relevant literature is focused on high-tech technologies and approaches based on programming, which are defined as machine-learning, geographic information systems (GIS), or Python/R- models. As much as they are powerful, they are specialized and steep in their learning. Due to this, they may not be readily available to the frontline managers, operations planners or those who are in close contact with the decision making process, who would be supply chain coordinators. It would require friendlier tools and methods that would enable non-technical individuals to be more active in analytics. The difference between the technical complexity and easy usability should be filled with solutions that are scalable.

Besides, an issue related to low-code analytical platforms lacks attention in academia either. To some extent, the tools used in the industry such as Alteryx or Tableau are poorly presented in the literature. They have immense opportunities as far as flexibility, speed and user activity are concerned. They allow us to build un-scary automated, scalable workflows that can resolve real world problems with limited data, and simple rules. The introduction of these tools into the research discussion can create new opportunities to come up with solutions that can be really effective as well as accessible.

It represents an opportunity because there is no practical, modular, and easy-to-deploy solutions available. Businesses need models which are- fast-generated, simple to up-keep, and able to produce actionable knowledge other than making significant technical investments. It is so especially in case of spatial demand optimization when the decisions made should be able to respond to shifts in demand, pressures of delivery, and customer expectations. The increased need of fast service delivery, transport time and high customer satisfaction will necessitate the creation of solutions that not only grow vertically (data volume) but also are able to grow horizontally (between business units or geographies).

The proposed project tries filling these gaps by proposing a fully automated, end-to end, workflow that reduces spatial demand optimization problems to three variables, including latitude, longitude, and the demand at stores. It discards the need to have elaborate customer correspondence or traffic information and uses clever logic and clumping methodologies instead. The solution is modular, repeatable and scaleable since Data transformation would be done using Alteryx and visualization using Tableau. The output is a model that may be utilized by any given organization with diverse sizes that may or may not be technologically mature and consist of sufficient data.

To sum it up, scalable, accessible, and practical solutions in the area of spatial optimization are demanded with a high level of urgency and can be considered a true challenge. The current body of knowledge frequently ignores the facts of constraint on data, user restrictions as well as the joins of systems. Future study and implementation projects aiming at working with low-code technology and automation will be capable of closing this gap and providing valuable benefits to businesses operating within the environment of rapid changes and increased competition.

On the other hand, our project is going to employ Tableau and Alteryx, two software whose use we are curious to know how to perform clustering in Tableau, and make calculations in Alteryx another benefit that utilizing automated workflow of Alteryx will give us the opportunity to inject new input and output verification. These tools were used to conduct clustering, spatial boundary analysis and calculation of distance type centroid using Haversine formula. Such a low-code methodology allows realistic and repeatable modelling of warehouse locations without using machine learning or code. As far as we know, such combination of tools was not used before to tackle the warehouse location problem in an academic setting, and, as a result, our project is quite innovative and approachable.

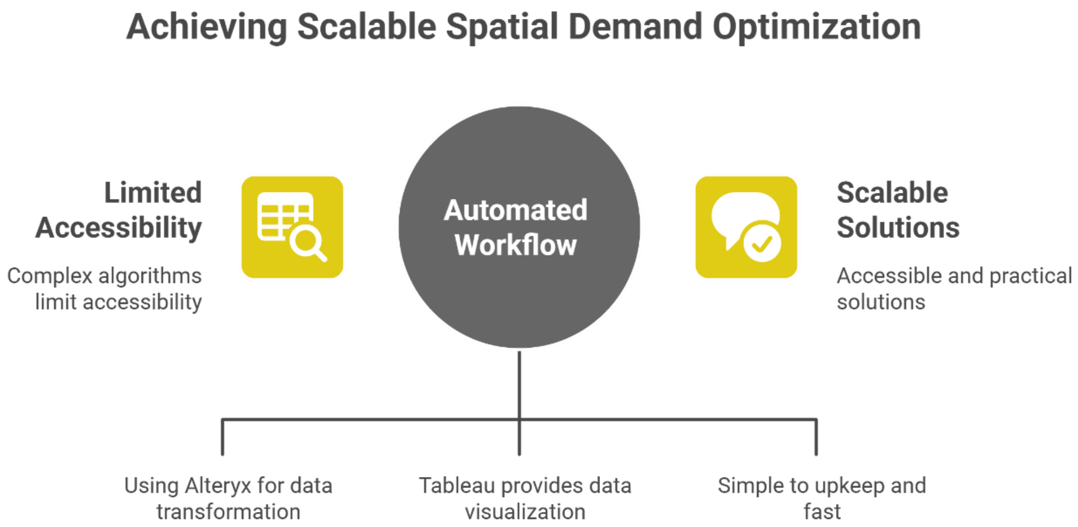


Figure 6 - Achieving Scalable Spatial Demand Optimization

CHAPTER 3

CHAPTER 3 - METHODOLOGY

This study follows a quantitative and analytical approach to solve the business problem of warehouse location optimization. The methodology consists of data preparation, demand-based filtering, clustering analysis, and final output generation using a combination of industry-standard tools.

Figure 7 - Optimizing warehouse Logistics

- **Data Collection and Preparation**

  The dataset used in this study contained three columns: latitude, longitude, and total demand of each store. No additional data such as road connectivity, customer demographics, or service radius was available. Therefore, the analysis was conducted solely on spatial and demand-related inputs. All data was cleaned, structured, and imported into Alteryx for further processing.

- **Demand-Based Filtering**

  To focus on the most impactful areas, we applied a Pareto-based approach by selecting the top 80% of demand-contributing stores. This helped us prioritize high-demand locations and eliminate low-impact outliers from the clustering process.

- **Clustering Using Tableau**

  K-means clustering was performed in Tableau using latitude and longitude fields. This method grouped geographically close stores together into clusters. After testing multiple cluster counts, we finalized the optimal number of clusters based on visual spread, geographic relevance, and internal validation.

- **Data Transformation in Alteryx**

  All data transformations, formula logic, ranking, and demand-based calculations were performed in Alteryx. Tools like **Summarize, Join, create point, formula, Filter, transform, Auto field, Sort, crosstab, browse, Data Cleansing, multiple Join, input tool, output and spatial match** were used to structure the data before and after clustering. Final output included warehouse cluster IDs, centroid points, and demand summaries.

- **Data Output and Storage**

  The final processed data, including the assigned clusters and suggested warehouse points, was written to a SQL database. This step ensures scalability and future use for other internal dashboards or business operations.

- **Dashboard**

  After saving output in Data Base then we connect Tableau with database and use some formulas in calculative field like generating origin point, destination point and line between these two points by using this field we made a map chart which shows connection of warehouse with its store.

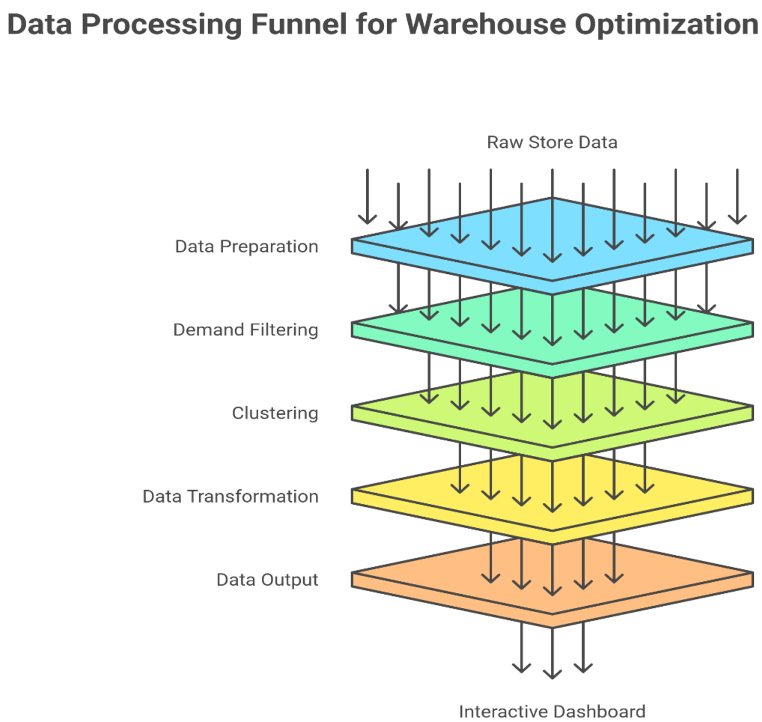


Figure 8 – Data Processing Funnel for Warehouse Optimization

CHAPTER 4

CHAPTER 4 - DATA ANALYSIS

4.1 Data Understanding and Preparation

Preparation of the raw dataset and its understanding was the initial part of the analysis. The data had three fields vital namely Destination_Latitude, Destination_Longitude, and Total_Demand, that corresponds to location and volume of business at the store level. The Input tool was used in ensuring this data set was imported to Alteryx with ease so that we could connect our .csv file very easily. It was important to clean and organize data before undertaking any form of analysis. First of all, we examined the types of data and made sure that longitude and latitude were numbers. The missing values or inconsistency were dealt with at this level. Auto Field and Data Cleansing feature through Alteryx have allowed us to change data type automatically and delete whitespace or null values. After cleaning the data we ranked the stores ranked in descending order according to Total_Demand-using the sort tool. This ranking concluded in detecting which stores gave the biggest contribution to the total demand. It was essential to comprehend this distribution, as we planned to put our optimization efforts on the areas, where our effect would be greatest. The thorough preparation of the data guaranteed that the following processes, clustering, zone establishment, and spatial calculations would be correct and significant. Its simplicity allowed testing of optimization using automation, and the clarity of the format was advanced to be input into Tableau and other spatial tools easily. It prepared the ground for a strong analysis and reduced the number of errors in the further stages of the process.

Table 1 - Store Demand Data

Destination_Latitude	Destination_Longitude	Total_Demand
28.7041	77.1025	1231231221
28.5006	77.2772	1231231220
28.6829	77.289	1231231219
28.6417	77.2197	1231231218
28.6842	77.2066	1231231217
28.5582	77.2066	1231231216
28.5362	77.2684	1231231215
28.5672	77.21	1231231214
28.6139	77.209	1231231213

28.6519	77.2315	1231231213

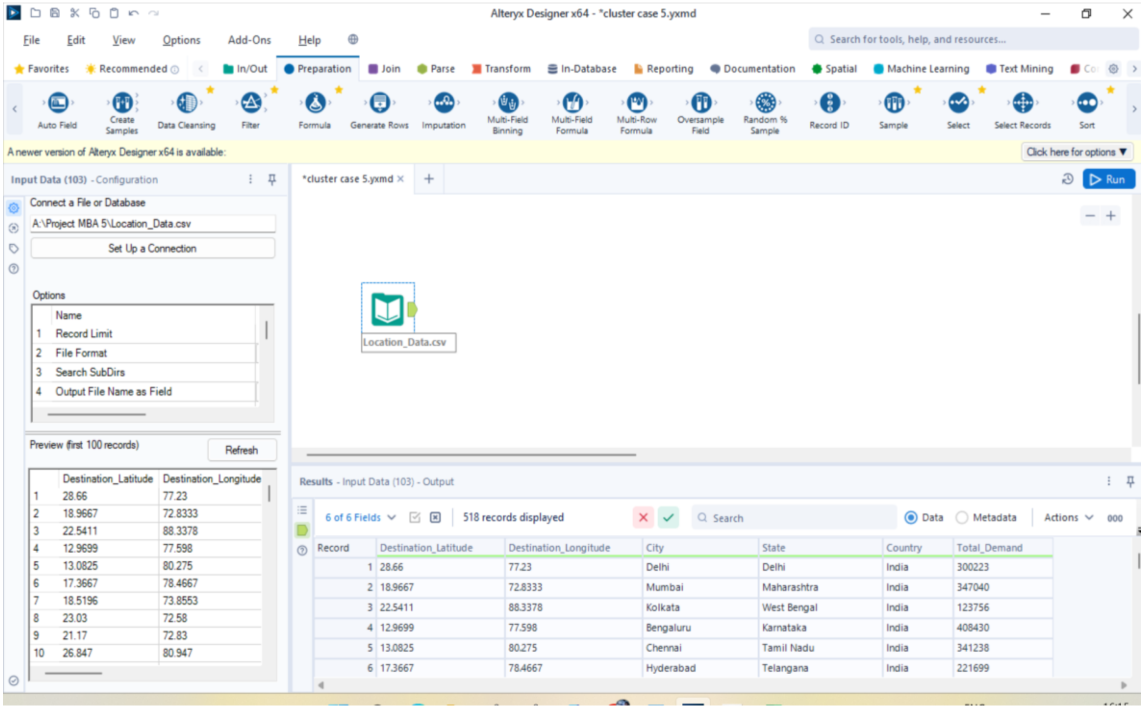


Figure 9 - Alteryx Data Import

4.2 Demand-Based Filtering and Clustering in Tableau

Once we prepared the data, we concentrated on bringing its scope to the most relevant entries. Through the Sample tool in Alteryx, we sampled the 80 percent of the records using cumulative Total_Demand as the sampling criterion. This is in line with the Pareto rule of thumb that most of the demands are usually realized by a small percentage of stores. After separate area stores that should have been at the tip of the demand, we sent data along and brought the data to Tableau. We managed to plot Latitude and Longitude in Tableau on the map view that provided us with a visualized perception of how the stores were distributed throughout the region. The tableau was then clustered, through the inbuilt K-means in tableau. This was the grouping of stores by their proximity, because of which clusters emerged, which resembled the service areas. Tableau automatically assigned colors to every cluster and in the result it is easy to see the picture and interpret the results. This visual aid assisted in deciding the manner in which stores issued naturally grouped because of position, even before any infrastructure or customer facts existed. The stores that did not combine well in clusters were identified as being under the collection of stores that were not clustered which we went over this later. Clustering allowed us to go beyond raw data by geographic area and group data into meaningful zones preparing us to do more analyses. The clustering logic involved only geospatial data however, it was successful in identifying demand centers. At the end of this step we placed a cluster ID on each store so that we could treat each group separately in the following steps. This section of the analysis was incredibly productive and informative because Tableau is a rather user-friendly tool with visual responses.

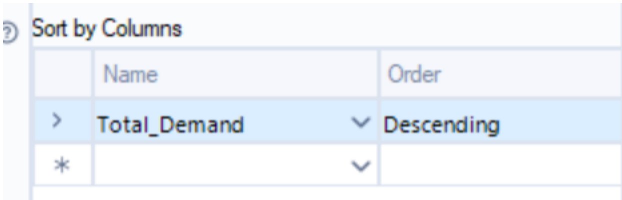


Figure 10 - Demand Sorting View

4.3 Weighted Centroid Calculation and Spatial Filtering in Alteryx

Having assigned clusters to each of the stores, we went back to Alteryx to determine central points (or, in other words, centroids) of each cluster. These major ones would be optimized service points. We import the clustered data then the Appendix Field made the optimization of data types through auto field tool. After that, we used Formula tool to evaluate weighted coordinates. We established two additional fields Weight_Latitude (Latitude * Total_Demand) and Weight_Longitude (Longitude * Total_Demand). These sectors made sure that stores that acted in higher demand affected the centroid more than the lower ones. The data was classified by cluster ID and the total demand, the number of stores, and weight values were summed in the table with the help of the Summarize tool. Then applying the Formula tool one more time helped us to calculate centroid of each cluster as dividing the sum of the weights by the sum of demand gave us the accurate and demand-sensitive location of the centroid. These points we called as the services potential areas. We did the validation by Create Point and Spatial Match tools before finishing them to ensure the correctness of the location. Such tools enabled us to define geographic parameters - borders of cities or regions - and assure that the centroids were in acceptable service area. In case a centroid was off-boundaries or was a part of a cluster with a short amount of stores (less than 15 stores or low demand) we filtered it. On the completion of this step we locked in three good zones which satisfied both spatial and business rules. The center point of each zone was weighted and validated and formed the core point of reference in the determination of store connection and delivery routes in the following stages.

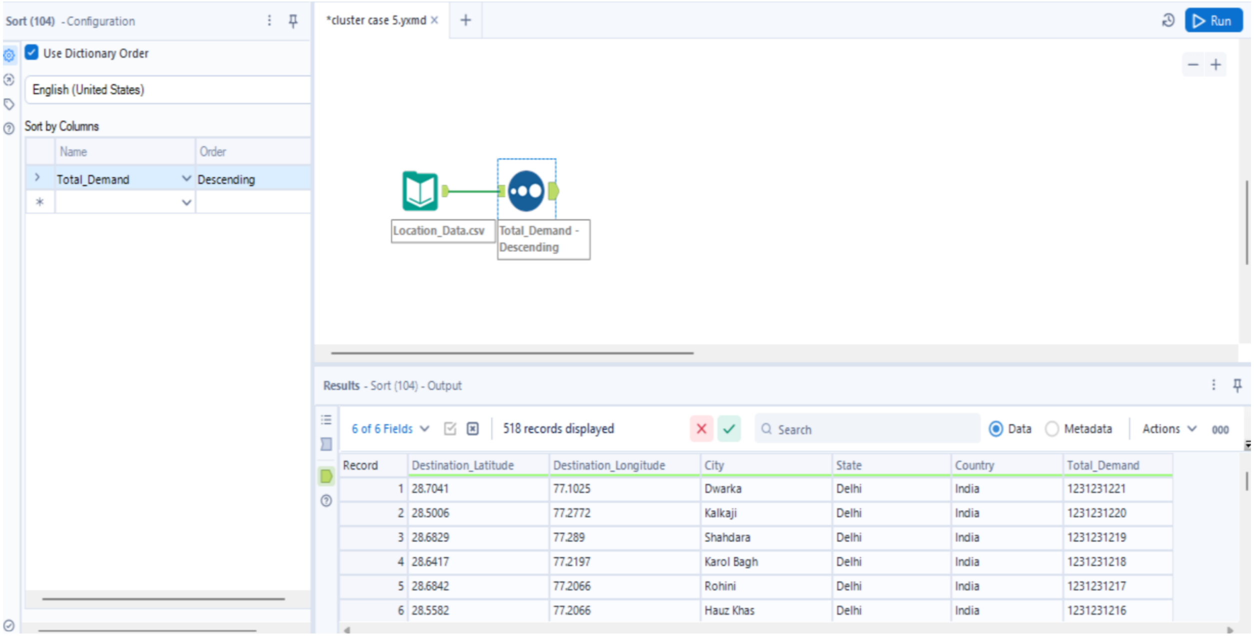


Figure 11 - Demand Ranking Output

4.4 Store-Zone Assignment Using Distance Calculation

Once, we established relevant zone centers, we shifted to allocating stores to their close zone. Of special importance was this to two kinds of stores: the bottom 20 per cent (omitted previously because of low demand) and stores in filtered-out clusters. The first thing we done was to make a reference table of the three completed zones and their coordinates. Then we estimated the distance between every store and the three points of the zones by the formula Haversine in Alteryx Formula tool. The Haversine formula aids in calculation of the shortest path between two points on a sphere and they are determined in latitude and longitude. The store had the value of three distances respectively in every zone. Then we applied the other Formula tool to determine the shortest distance between the three and the closest zone was assigned to the store. On this logic, the Assigned_Zone You were assigned to field was created. When each of the stores was assigned a zone, we merged these records with the stores that had been clustered earlier using the Join tool. We made sure that we assigned no store that was unassigned and the unified data had all the details regarding the store IDs, coordinates, demand, and assigned zones. The last mapping made it possible to cover the network fully. A store that was not initially clustered previously now had a logically nearest area that is based on location and demand. This was an important step to be thorough and efficient in the end product. The resultant product was a completely filled data set which is ready to be visualized and strategically used.

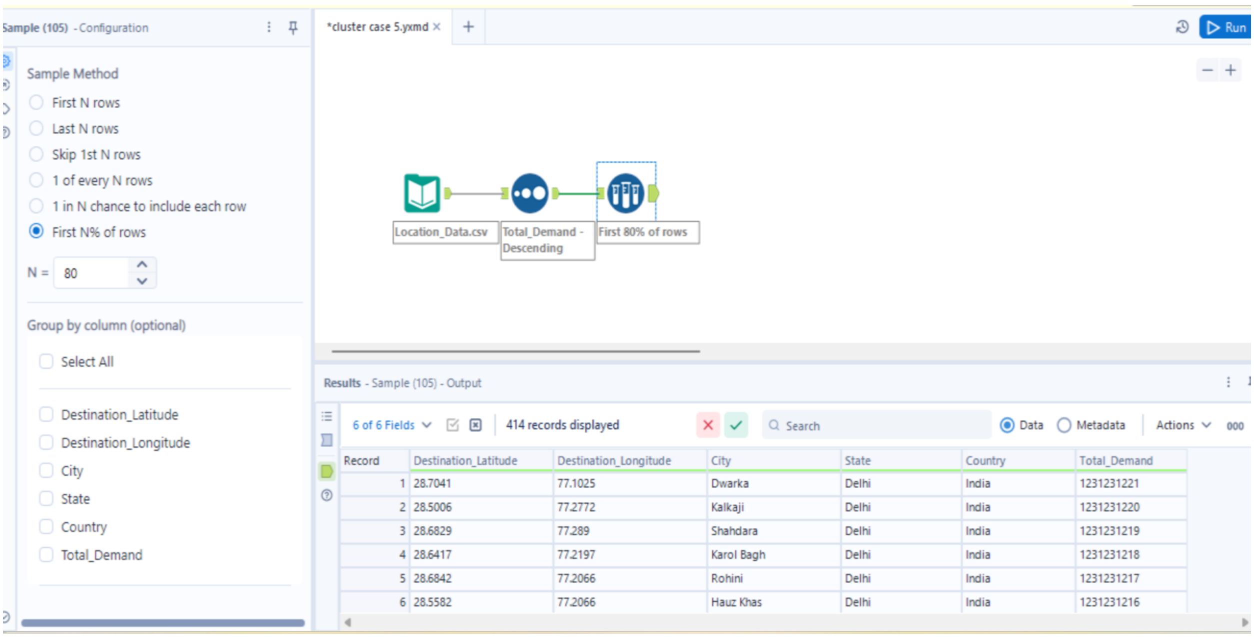


Figure 12 - Top Demand Filter

4.5 Output Storage and Visualization in Tableau

Finally, through the Output Data tool we exported the entire dataset on Alteryx to a SQL database. This made the results to be stored safely and they could be retrieved to make reports or dashboards. The data set stored in it contained such fields as store IDs, assigned zones, latitudes, longitudes, demand values, and computed distances. Then we linked this SQL database with Tableau to provide a visual, interactive dashboard. Then we created visualization of how each store was connected to the zone it was assigned using calculated fields including MAKEPOINT and MAKELINE. Distribution of stores and zones was mapped by MAKEPOINT, whereas MAKELINE mapped the direction of links between stores and zones displaying clearly the paths of service areas. This image allowed to easily notice whether stores got incorrect assignment, and they were given too far away. We have also resorted to color coding and filtering used in comparing the volumes of demands across various zones. Non-standard funnel chart was also added to make the comparison between total store counts and cumulative demands per zone. This assisted in explaining the network distribution of business. Viewer interaction in Tableau zoomed in, filtered, and checked single zones or stores. Both SQL-supported data and Tableau representations combined allowed making the analysis not only effective but also ready to be presented in the business environment and user-friendly. The workflow demonstrated that valuable location-based planning can be performed with the utilization of low-code tools, even if the data available is limited, and preconditioned future automation or 24/7 tracking.

4.6 Data Type Optimization and Zone Renaming

After assigning all the stores to zones, we used the optimization of data types to be used in the analysis process and to be clear. In Alteryx, we have a tool called Select which we use to check and change the data types of the new created fields. As an example Wh_Latitude and Wh_Longitude (zone centroid coordinates) fields were changed to the type double to guarantee exactness of the following calculations. In the same way, other field types like Zone_ID and Store_Name were typed correctly so as to have uniformity. Having done this cleanup we proceeded to supplying each zone with a recognizable and identifier name. We decided to rename zones as something meaningful rather than calling them Cluster 1, 2 or 3 (as created by Tableau clustering in Tableau) using Formula tool. Columns such as Zone_Name were introduced and filled with their zone labels e.g. Zone A, Zone B and Zone C in an order or sequence of the design zone or contribution of demand. This made the ultimate information easy to read and present. It also enables ordinary business users who are not so technologically oriented to understand the output intuitively. This was done in order to clean up the data model, scale and prepare the data to be served in presentations. Such improvements concerning organization and naming ensured that the outcomes would be (a) applicable in different dashboards or (b) exportable into a report without additional reformatting. In general, this is a trivial but needed step of the analysis in order to guarantee consistency of the data, accuracy of formatting, and improved conveying results to the decision-makers.

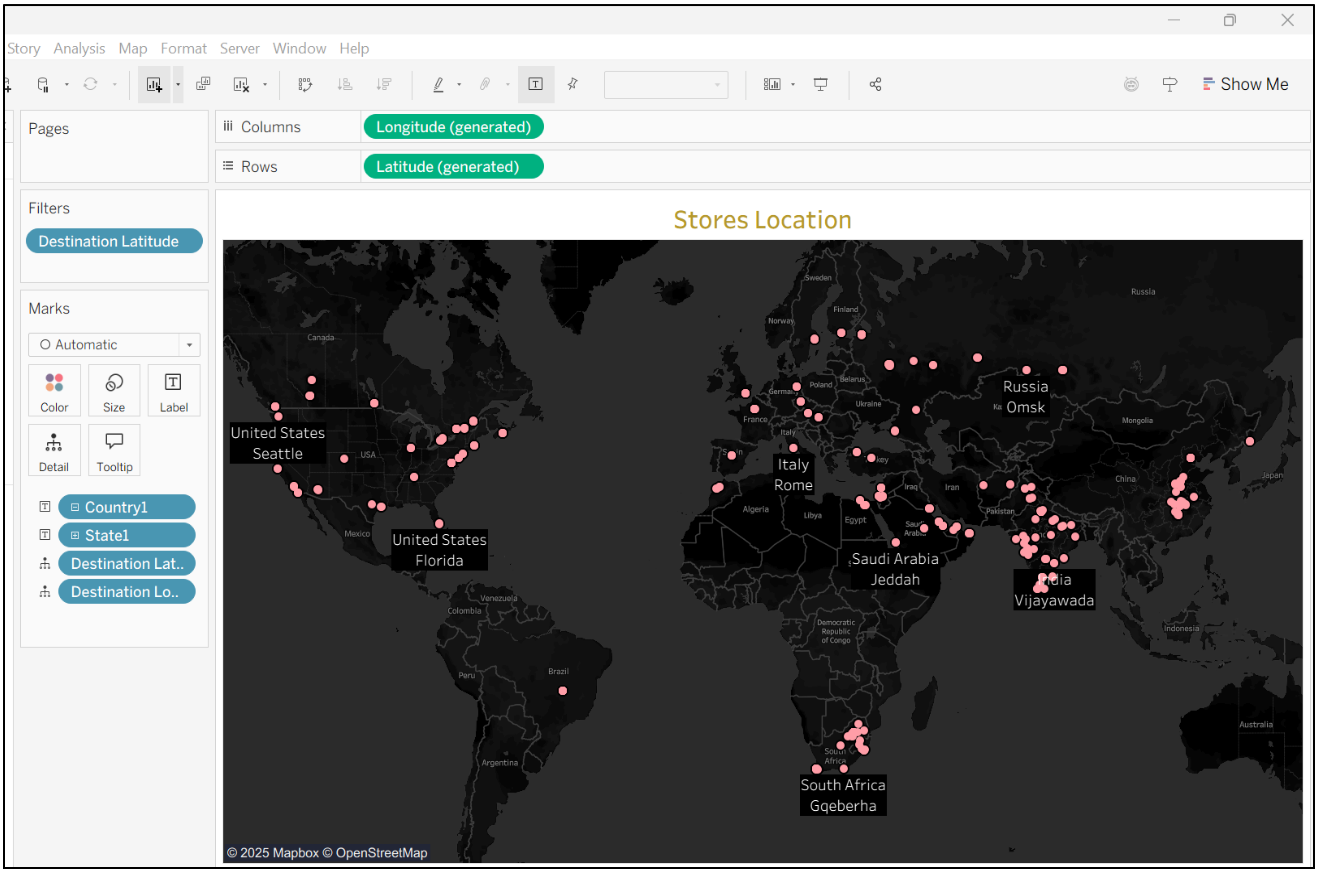


Figure 13 - Store Map Plot

4.7 Combining Remaining Stores and Filtered Records

The step was dedicated to the combination of all the relevant records in a single dataset. Only the 80 per cent of the highest-in-demand stores were initially clustered. Once we made production zones optimized, we had to consider the rest of the 20% stores that were not available at first in clusters, and others stores in the ignored clusters. With the help of Transpose tool in Alteryx we reshaped the dataset to make it ready to be appended to the original dataset. Subsequently, using the Crosstab and Formula tools, we recreated the records at the stores level in an identical form. We made all the stores of the same structure: ID, coordinates, demand and zone assignments so to be aggregate without any issues. Data Cleansing tool has been used to remove nulls and normalize values. Having got it ready we then appended (Union and Join) the original Stores that were Processed and the new clustered stores that had been processed hence creating a single dataset. This also guaranteed that all the stores, not just those in the high demand category, now had a record and were matched with a service zone. This was crucial in the assurance of completeness of data and coverage of business. There was no shop that was abandoned. It also enabled the use of downstream processes like visualizations or distance optimization be conducted on the whole network without special processing. With this merge, the dataset turned into an actual single source of truth a well-edited and supplemented table that contained all the needed fields and reached the full range of business locations.

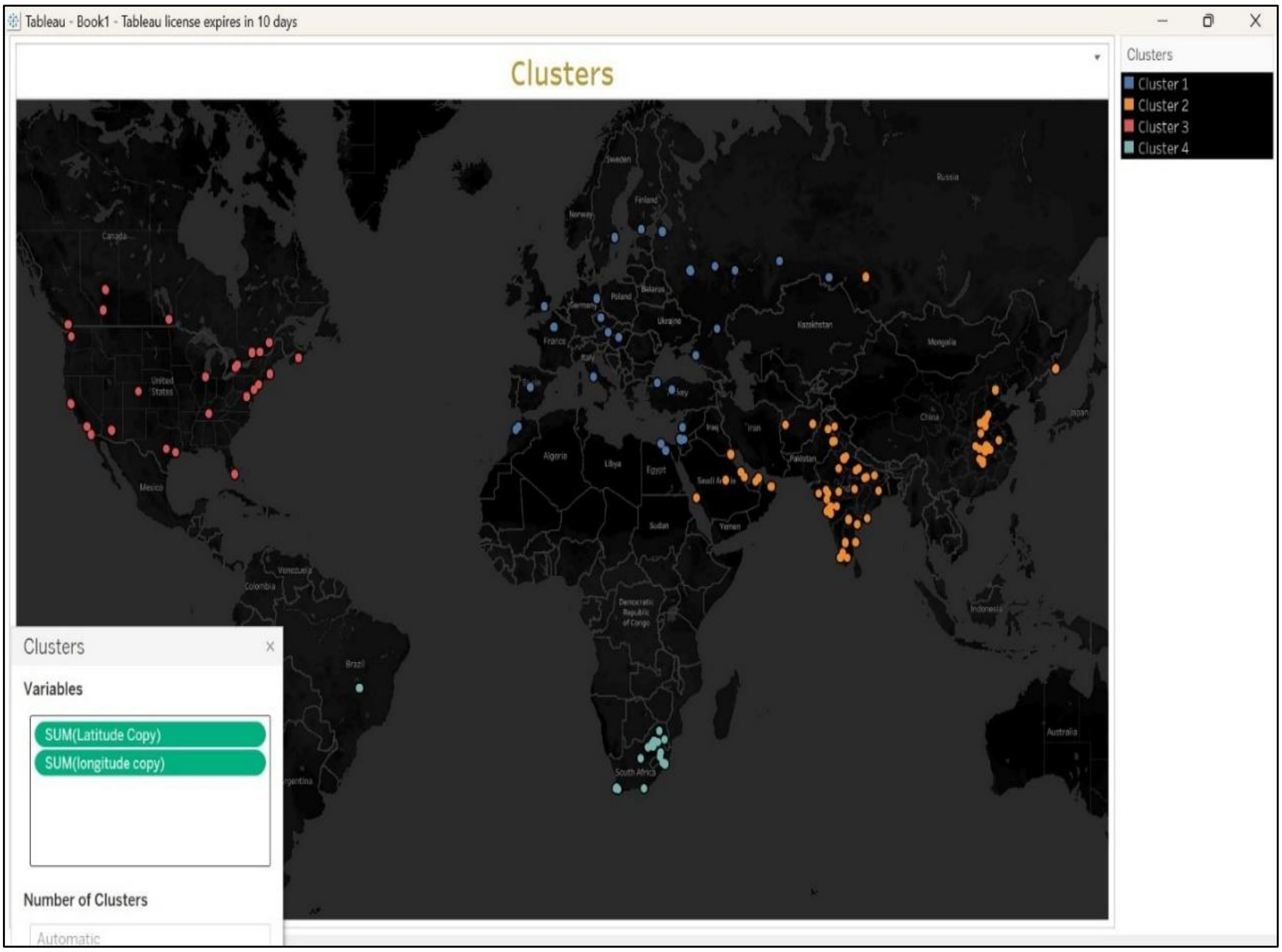


Figure 14 - Store Merge View

4.8 Final Distance Optimization Using Haversine Formula

As all the stores are now connected to the service zone, we measured the distance between every store and the location center of a zone using the formula of Haversine. It computes minimum distance between two points on the surface of the earth taking into account the curvature of the planet- this is a best method to make computer calculations in the real geographic world. In Alteryx, this logic has been implemented through Formula tool. Three distance fields: Dist1, Dist2 and Dist3 was created to represent how distance each store is to the three zones available centroids. Thereafter, the lowest of the three distances was determined among the stores using another Formula tool. Referring to the result, we have assigned the closest zone (The closest zone is set using a conditional logic formula: say MinDistance = Dist1, then assign Zone A; MinDistance = Dist2, then Zone B and so on.). This was done such that every store was associated not only with the clustered area, but also the most effective area to make services and deliveries. Once the zone assignment was done we joined this result to earlier zone data and tidied up discrepancies using the Join tool. Such an optimization step enhanced zone planning and logistics on a significant basis. It also enabled an improvement of visual as well as resource attainment in the dashboard. Once this last spatial distribution had been done, we had the logical mapping of all stores with the most apt zone.

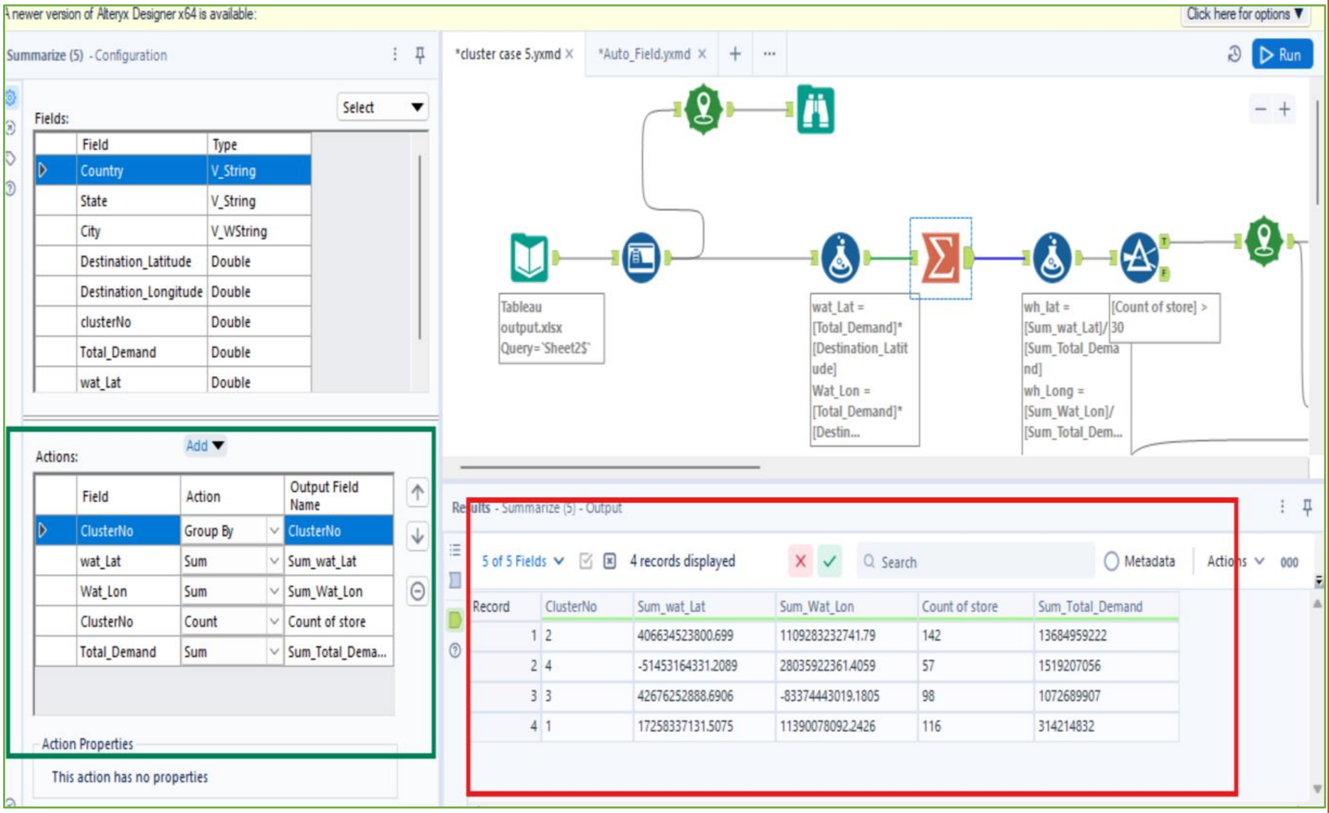


Figure 15 - Cluster Summary Stats

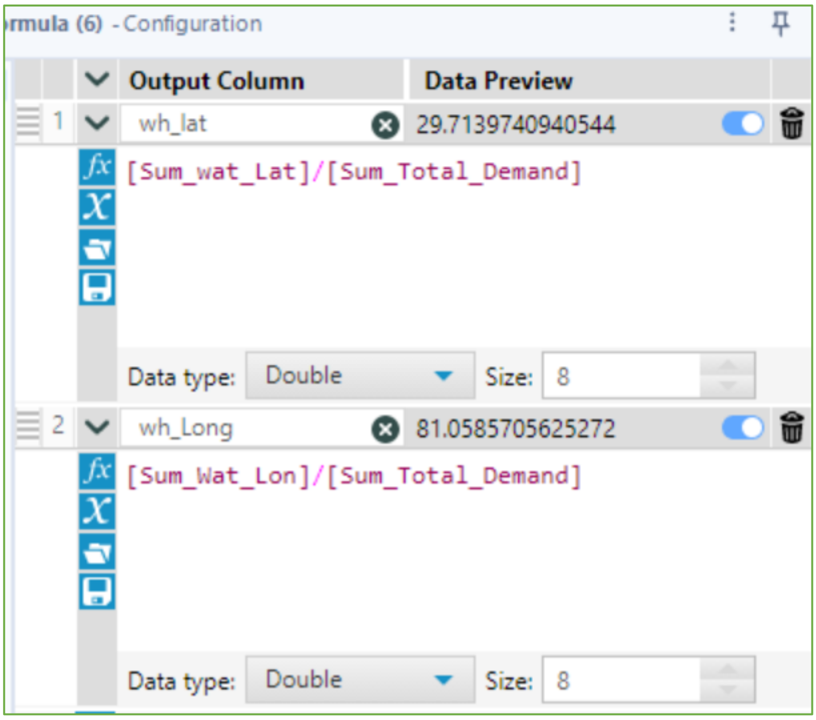


Figure 16 - Centroid Formula Calculation

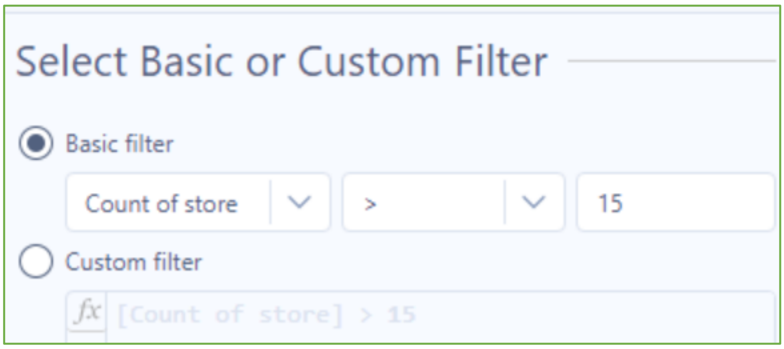
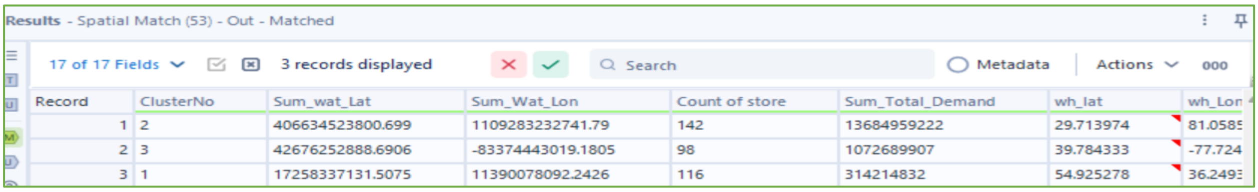


Figure 17 - Low-demand Zone Filter



Record	ClusterNo	Sum_wat_Lat	Sum_Wat_Lon	Count of store	Sum_Total_Demand	wh_lat	wh_Lon
1	2	406634523800.699	1109283232741.79	142	13684959222	29.713974	81.0585
2	3	42676252888.6906	-83374443019.1805	98	1072689907	39.784333	-77.724
3	1	17258337131.5075	11390078092.2426	116	314214832	54.925278	36.2493

Figure 18 - Spatial Match Tool

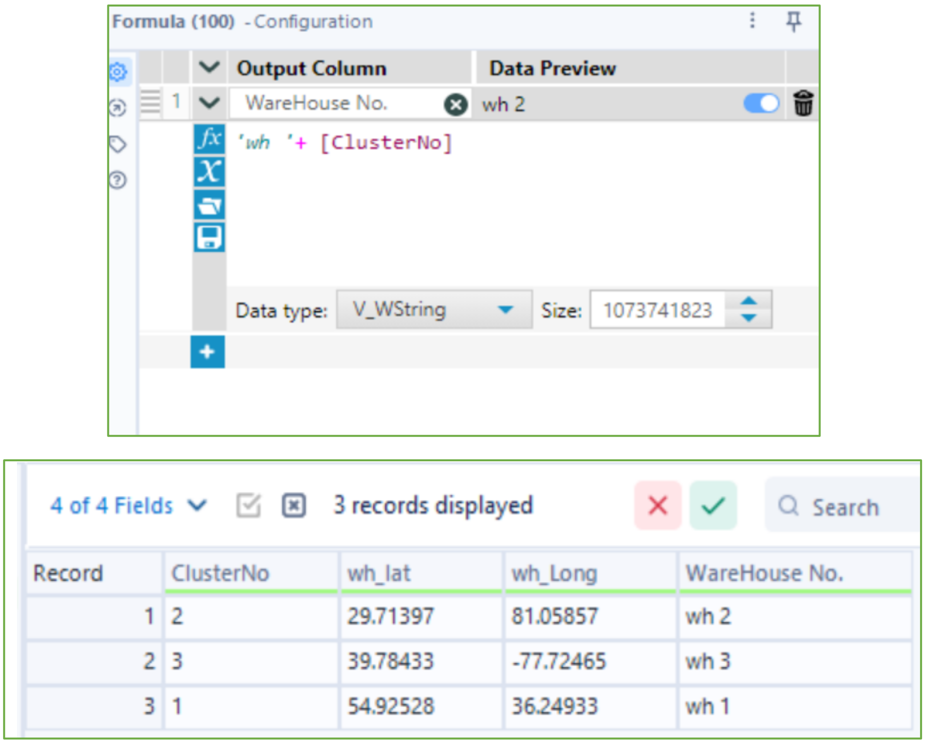


Figure 19 - Zone Naming Formula

4.9 Database Export and Integration Setup

Once the entire dataset had been finalized, having zone assignments, store coordinates as well as the distance calculations, we exported the results into a SQL database through Alteryx Output Data tool. The purposes of this database were a central storage place where it could be then used on the BI dashboards and as an operating system. We reconfirmed all the field names, data types and records before exporting them, so that they could fit the requirements of SQL schema. We also made sure that it was well indexed particularly in such fields as Store_ID and Zone_ID, to be able to use the information efficiently in on-the-spot applications. After the export had been done, we checked the number of records and sample outputs in the database to be sure that it did not get lost or packed in wrong format. The well-organized SQL results enabled us to join Tableau or any other reporting software easily with the most recent data. This integration assisted in the generation of live dashboards and made it possible to filter, drill-down and monitor performance in zones. Also, centralization of this data means that any future upgrade such as real-time updates, automation triggers or real-time updates can be deployed in a simple manner. This completed the backend component of the project and resulted in the availability of the results to the stakeholders with a safe, sound, and scale-able storage. It also brought forth the possibility of automated data update and re-evaluation of the zones after some period of demand inputs being entered.

3 of 3 Fields — 6 records displayed

Record	ClusterNo	Name	Value
1	2	wh_lat	29.71397
2	2	wh_Long	81.05857
3	3	wh_lat	39.78433
4	3	wh_Long	-77.72465
5	1	wh_lat	54.92528
6	1	wh_Long	36.24933

Figure 20 - Transpose Tool Output

Results - Cross Tab (66) - Output

6 of 6 Fields — 1 record displayed

Record	1_wh_lat	1_wh_Long	2_wh_lat	2_wh_Long	3_wh_lat	3_wh_Long
1	54.925278	36.249332	29.713974	81.058571	39.784332	-77.724648

Figure 21 - Crosstab Tool Output

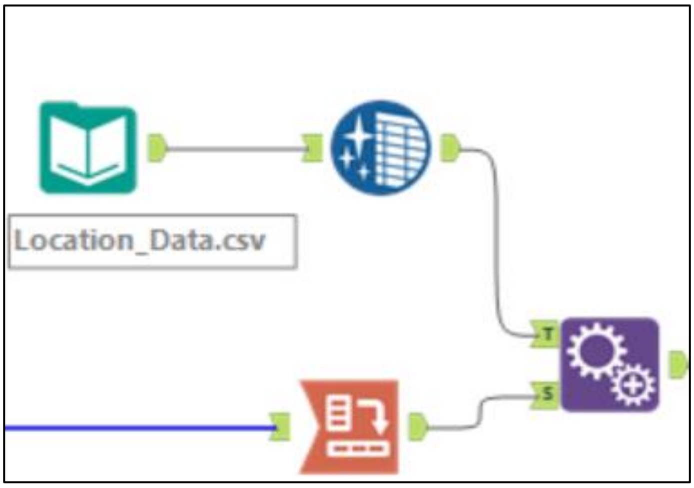
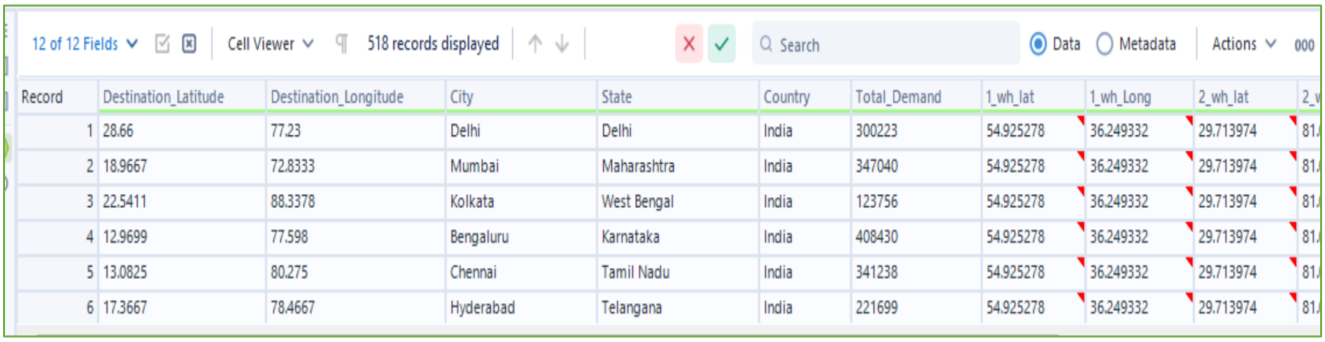


Figure 22 - File Append Process



Record	Destination_Latitude	Destination_Longitude	City	State	Country	Total_Demand	1_wh_lat	1_wh_Long	2_wh_lat	2_w
1	28.66	77.23	Delhi	Delhi	India	300223	54.925278	36.249332	29.713974	81.
2	18.9667	72.8333	Mumbai	Maharashtra	India	347040	54.925278	36.249332	29.713974	81.
3	22.5411	88.3378	Kolkata	West Bengal	India	123756	54.925278	36.249332	29.713974	81.
4	12.9699	77.598	Bengaluru	Karnataka	India	408430	54.925278	36.249332	29.713974	81.
5	13.0825	80.275	Chennai	Tamil Nadu	India	341238	54.925278	36.249332	29.713974	81.
6	17.3667	78.4667	Hyderabad	Telangana	India	221699	54.925278	36.249332	29.713974	81.

Figure 23 - Combined Store Records

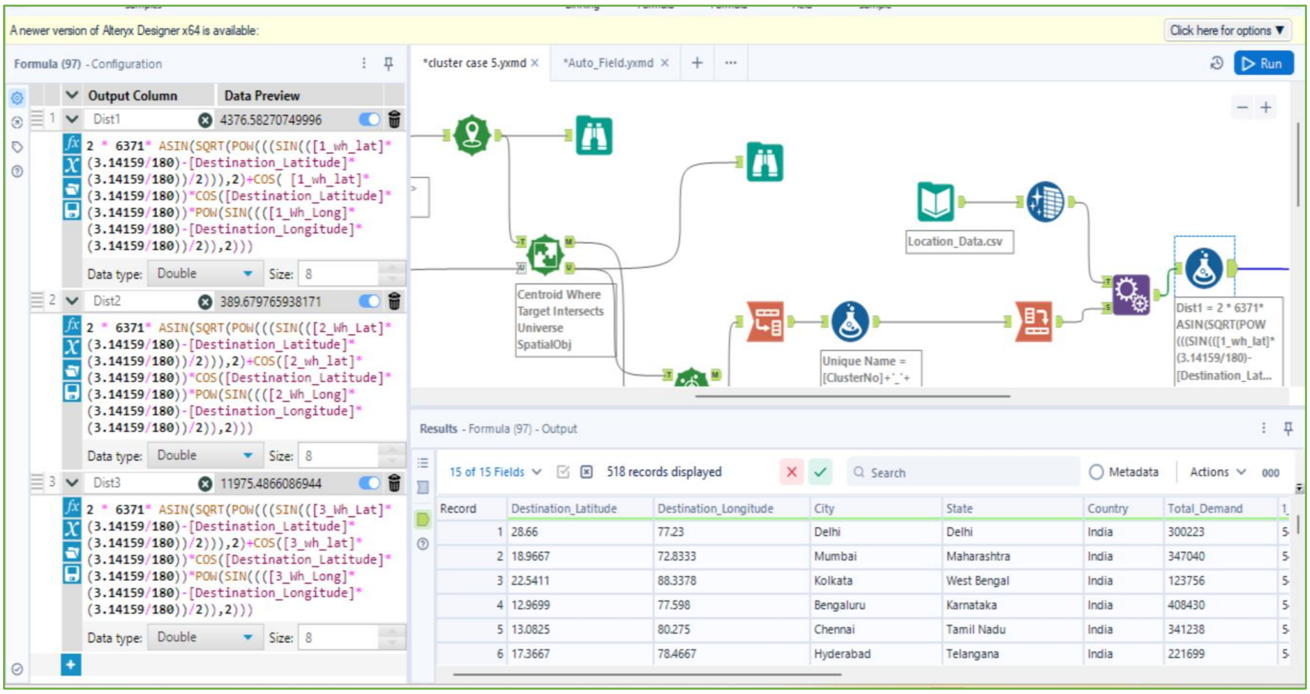


Figure 24 - Cluster Mapping Snapshot

Next step to find which distance is nearest to the warehouse for this we use formula tool bar and make two columns.

Minimum Distance= Min([Dist1],[Dist2],[Dist3])

Warehouse No.=

IF [min dist] =[Dist1] THEN 'wh 1' ELSEif

[min dist] = [Dist2] THEN 'wh 2' else 'wh 3' endif

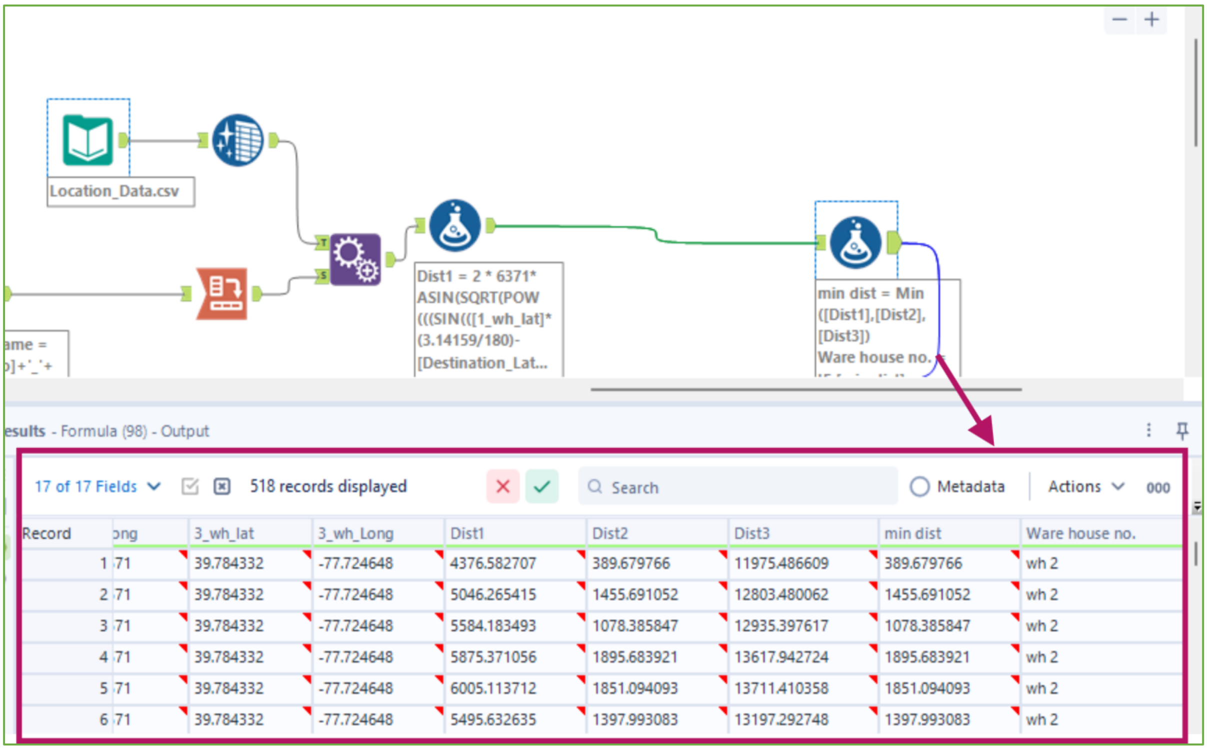


Figure 25 – Minimum Distance Logic

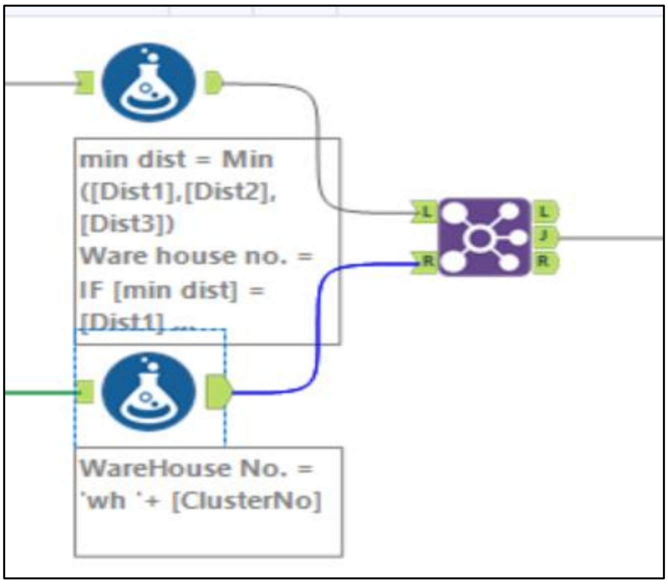
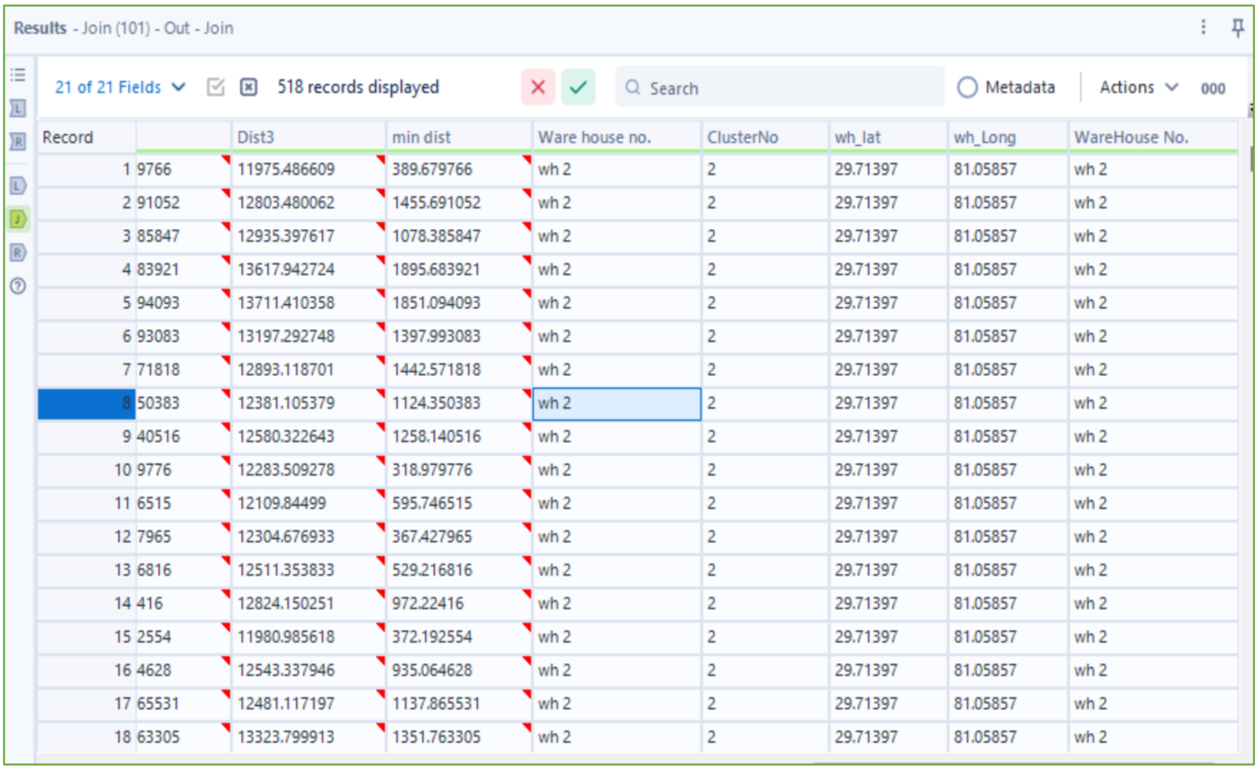


Figure 26 - Final Store Join



Results - Join (101) - Out - Join

21 of 21 Fields | 518 records displayed | Search | Metadata | Actions

Record		Dist3	min dist	Ware house no.	ClusterNo	wh_lat	wh_Long	WareHouse No.
1	9766	11975.486609	389.679766	wh 2	2	29.71397	81.05857	wh 2
2	91052	12803.480062	1455.691052	wh 2	2	29.71397	81.05857	wh 2
3	85847	12935.397617	1078.385847	wh 2	2	29.71397	81.05857	wh 2
4	83921	13617.942724	1895.683921	wh 2	2	29.71397	81.05857	wh 2
5	94093	13711.410358	1851.094093	wh 2	2	29.71397	81.05857	wh 2
6	93083	13197.292748	1397.993083	wh 2	2	29.71397	81.05857	wh 2
7	71818	12893.118701	1442.571818	wh 2	2	29.71397	81.05857	wh 2
8	50383	12381.105379	1124.350383	wh 2	2	29.71397	81.05857	wh 2
9	40516	12580.322643	1258.140516	wh 2	2	29.71397	81.05857	wh 2
10	9776	12283.509278	318.979776	wh 2	2	29.71397	81.05857	wh 2
11	6515	12109.84499	595.746515	wh 2	2	29.71397	81.05857	wh 2
12	7965	12304.676933	367.427965	wh 2	2	29.71397	81.05857	wh 2
13	6816	12511.353833	529.216816	wh 2	2	29.71397	81.05857	wh 2
14	416	12824.150251	972.22416	wh 2	2	29.71397	81.05857	wh 2
15	2554	11980.985618	372.192554	wh 2	2	29.71397	81.05857	wh 2
16	4628	12543.337946	935.064628	wh 2	2	29.71397	81.05857	wh 2
17	65531	12481.117197	1137.865531	wh 2	2	29.71397	81.05857	wh 2
18	63305	13323.799913	1351.763305	wh 2	2	29.71397	81.05857	wh 2

Figure 27 - Join Tool Results

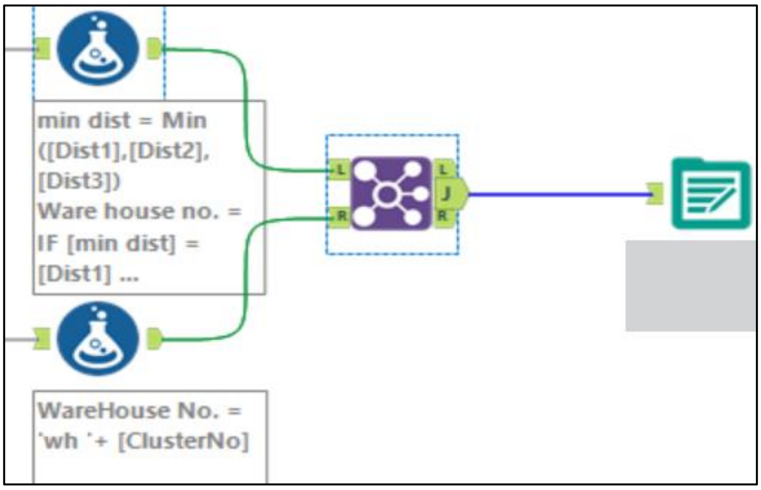


Figure 28 - Database Export Process

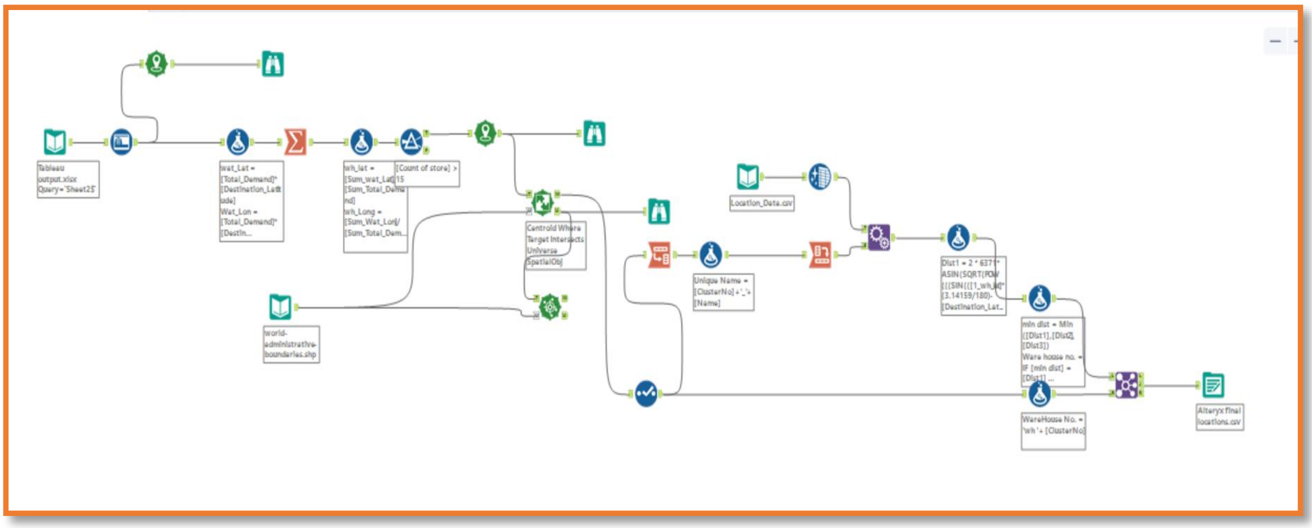


Figure 29 - Alteryx Workflow Overview

4.10 Visualization and Dashboard Development in Tableau

Visualization of the optimized data using Tableau was the last stage in the process of data analysis. Observing the SQL database with Tableau has allowed creating a dashboard on a map, where the stores are shown as a point, and each store has its line directed towards the centre of the corresponding zone. This has been done with the help of calculated fields such as MAKEPOINT([Latitude], [Longitude]) to calculate where stores and zones are located, and a MAKELINE([StorePoint], [ZonePoint]) to create a visual connection of both points. These graphical associations were used to enable the users to find the location of stores in different service areas and also the grouping in different areas within a short time. Filters were added according to zone, demand range and region to facilitate interactive analysis. The number of stores and cumulative demand that every zone handled was presented in a non-standard funnel chart. Color coding made an area distinction clearer, and tooltips gave an immediate access to a corresponding measurement of every point in the map. It helped simplify the process of evaluating the performance of the spatial optimization and find the congruence between visualization and business objectives. It allowed evident detection of trends and differences, i.e., demand concentrations or possible inefficiencies of zone assignments. The visualization of the data played a critical role in confirming the effectiveness of the optimization process since it transformed complicated statistics and raw data into business-intelligent feedbacks. Finally, Tableau did succeed in its core mission of converting backend logics into a user-friendly, professional, and real-time decision-making platform which in turn facilitated strategic plans in a clear and accurate way.

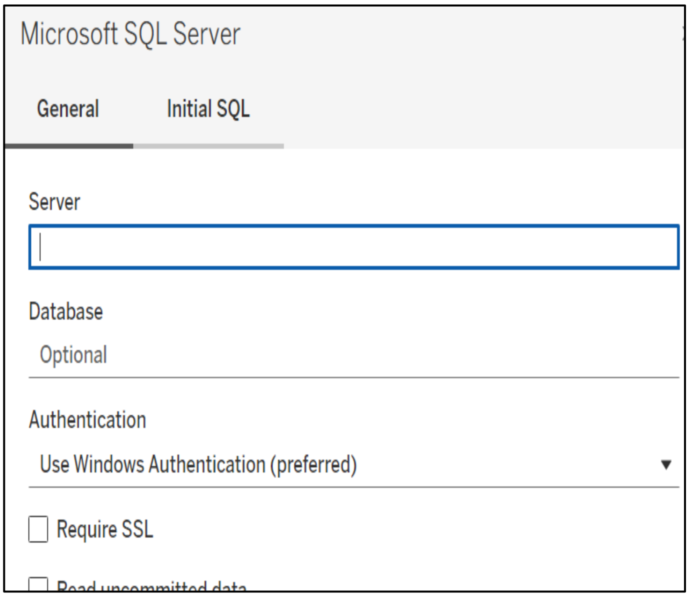


Figure 30 - Tableau SQL Connection

Figure 31 - Zone Location Map

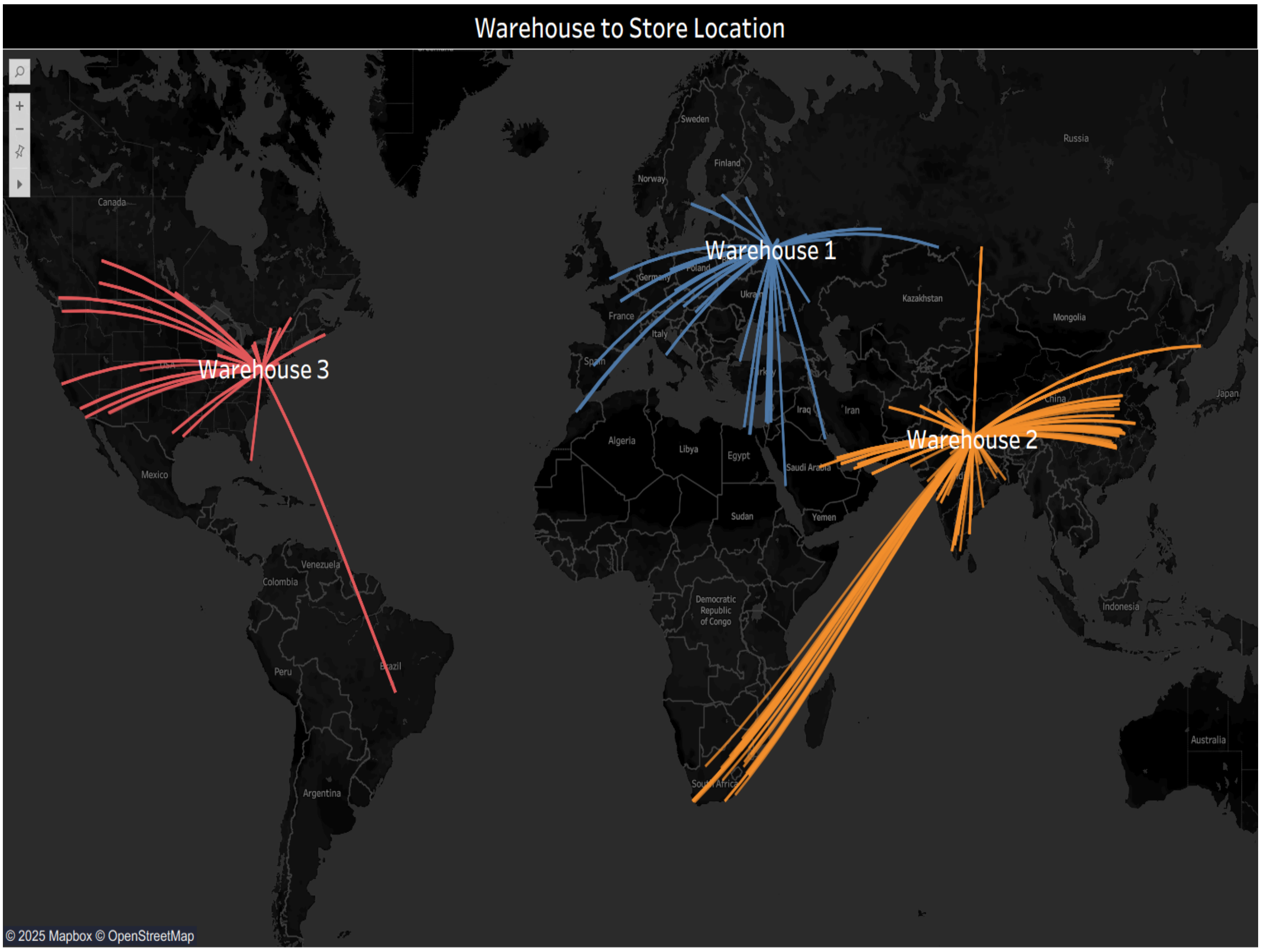


Figure 32 - Zone Store Lines

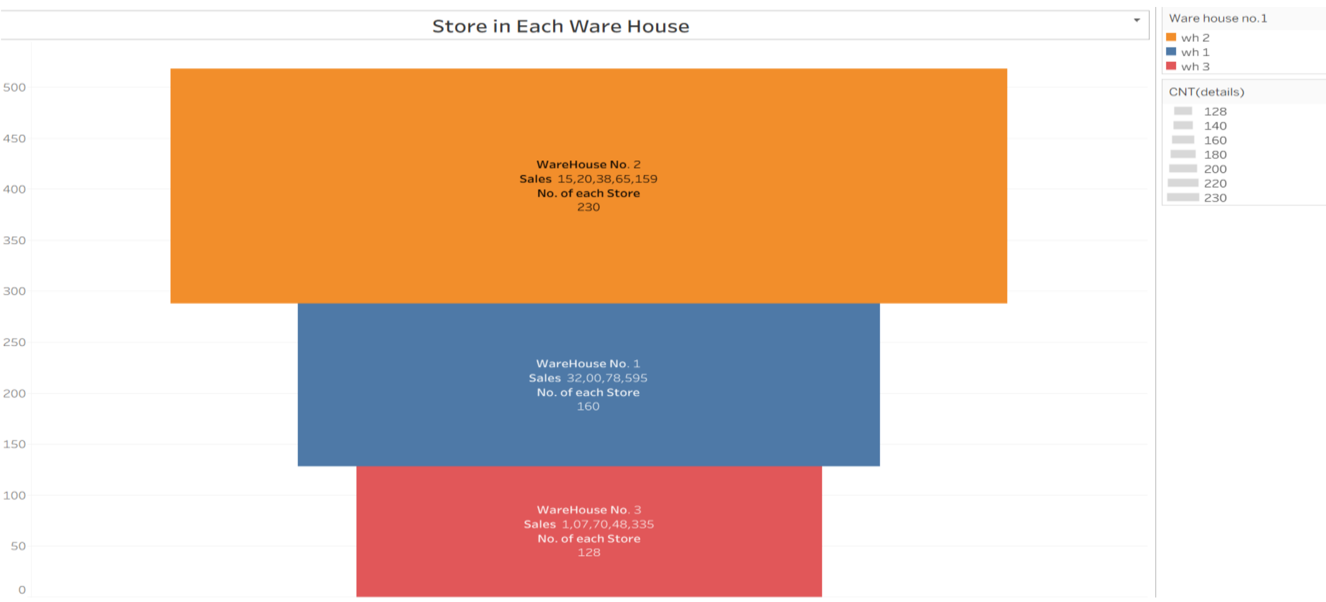
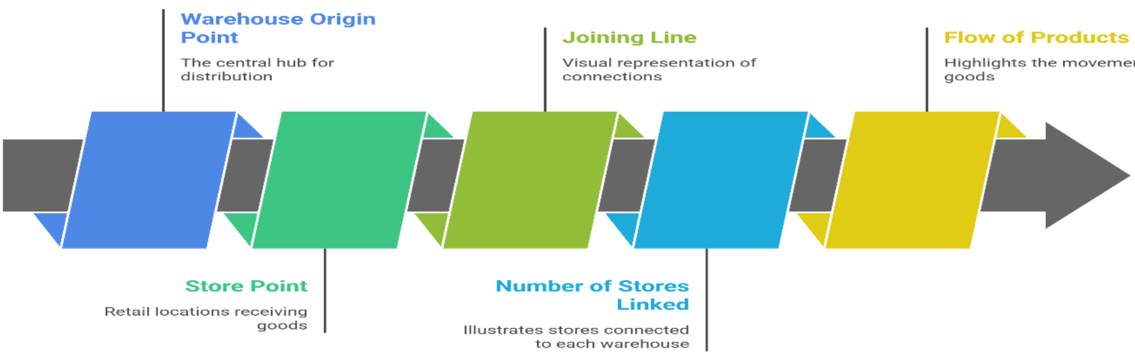


Figure 33 - Zone Funnel Chart

Figure 34 - Warehouse Demand and Store Coverage Dashboard



CHAPTER 5

CHAPTER 5 - RESULT

5.1 Cluster-wise Store and Sales Coverage (Final Output)

The final clustering and analysis resulted in the formation of three distinct warehouse zones, each responsible for a specific set of stores. These zones were derived using location-based clustering logic and aligned well with demand distribution across the dataset. Below is a summary of each warehouse's store count and corresponding sales performance:

Warehouse No. 1

Total Sales: ₹32,00,78,595

Number of Stores: 160

Warehouse 1 covers a medium-sized region, serving 160 stores with a solid sales contribution. Although not the largest in store count, it plays a vital role in fulfilling demand in its area. Its coverage is relatively balanced in terms of both volume and reach.

Warehouse No. 2

Total Sales: ₹15,20,38,65,159

Number of Stores: 230

This is the largest cluster in terms of store count, handling 230 stores. It also contributes the highest overall sales, indicating that it likely covers high-density regions or commercially stronger zones. Strategically, this warehouse is the core node in the supply chain and demands high operational efficiency.

Warehouse No. 3

Total Sales: ₹1,07,70,46,335

Number of Stores: 128

Warehouse 3 manages the fewest stores and the lowest total sales among the three. It likely represents a more scattered or less urban cluster. Despite lower numbers, this warehouse is still essential for servicing its zone, ensuring full geographic coverage.

5.2 Strategic Interpretation

- Warehouse 2 is the most critical in terms of both coverage and profitability. It should be prioritized for high-capacity infrastructure, fast-moving inventory, and service-level optimization.
- Warehouse 1 provides a steady balance between store count and sales, making it ideal for standard operations.
- Warehouse 3, while smaller in scale, is strategically important to ensure complete market reach and reduce dependency on distant hubs.

This output supports the decision to establish three warehouse locations, each positioned to maximize service efficiency, reduce delivery time, and align with business demand patterns. The visual representation of store-to-warehouse connections and the clear segmentation also make it easy to implement in real-world logistics planning.

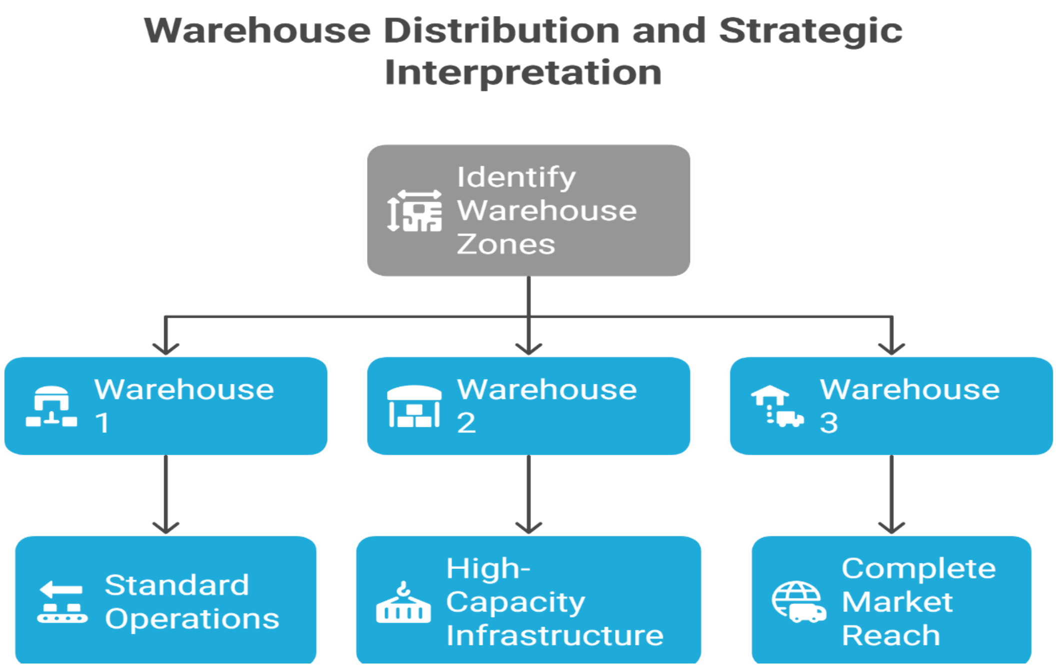


Figure 35 - Warehouse Distribution and Strategic Interpretation

CHAPTER 6

CHAPTER 6 - CONCLUSION

The purpose of this project was to identify the most suitable warehouse locations based on store demand and geographical distribution, with the goal of improving delivery efficiency and reducing operational costs. Despite working with a limited dataset containing only latitude, longitude, and total store demand the analysis was able to deliver practical and strategic results.

Using clustering techniques, we divided the store network into three optimized groups, each linked to a centrally placed warehouse. These clusters were not only logical in terms of location but also aligned well with demand patterns. Among them, Warehouse No. 2 emerged as the most critical, handling the highest number of stores and the highest overall sales. The other two warehouses also showed balanced performance, supporting efficient service in their respective zones.

This outcome confirms that even with minimal data, meaningful business insights can be drawn using the right tools and approach. The combination of Tableau for clustering and visualization, Alteryx for data transformation, and SQL for storage made the entire process efficient, scalable, and ready for future business use.

The findings support key supply chain goals such as faster delivery, lower logistics costs, and higher customer satisfaction and can help companies make informed decisions when planning warehouse infrastructure.

REFRENCES

Sharma, A., Jakhar, S. K., Vlachos, I., & Kumar, S. (2024). Advances in hub location problems: a literature review and research agenda. *International Journal of Productivity and Performance Management*. https://doi.org/10.1108/ijppm-09-2023-0514

Singh, R. K., Chaudhary, N., & Saxena, N. (2018). Selection of warehouse location for a global supply chain: A case study. *IIMB Management Review*, *30*(4), 343–356. https://doi.org/10.1016/j.iimb.2018.08.009

Ghadge, A., Yang, Q., Caldwell, N., König, C., & Tiwari, M. K. (2016). Facility location for a closed-loop distribution network: a hybrid approach. *International Journal of Retail & Distribution Management*, *44*(9), 884–902. https://doi.org/10.1108/ijrdm-07-2015-0094

Zadeh, S. B. I., & Garay-Rondero, C. L. (2023). Enhancing Urban Sustainability: Unravelling Carbon Footprint Reduction in Smart Cities through Modern Supply-Chain Measures. *Smart Cities*, *6*(6), 3225–3250. https://doi.org/10.3390/smartcities6060143

Modgil, S., Singh, R. K., & Hannibal, C. (2021). Artificial intelligence for supply chain resilience: learning from Covid-19. *The International Journal of Logistics Management*, *33*(4), 1246–1268. https://doi.org/10.1108/ijlm-02-2021-0094

Cataldo, Ieva, Banaitis, Audrius, Samadhiya, Ashutosh, Banaitienė, Nerija, Kumar, Anil and Luthra, Sunil (2022) *Sustainable supply chain management in construction: an exploratory review for future research.* Journal of Civil Engineering and Management, 28 (7). pp. 536-553. ISSN 1822-3605

Wang, J., Wen, J., Pajić, V., & Andrejić, M. (2024). Optimizing Cross-Dock Terminal Location Selection: A Multi-Step Approach based on CI-DEA–IDOCRIW–MABAC for Enhanced Supply Chain Efficiency—A Case Study. *Mathematics*, *12*(5), 736. https://doi.org/10.3390/math12050736

Castillo-Villar, K. (2014). Metaheuristic Algorithms Applied to Bioenergy Supply Chain Problems: theory, review, challenges, and future. *Energies*, *7*(11), 7640–7672. https://doi.org/10.3390/en7117640

Khalid, W., & Herbert-Hansen, Z. N. L. (2018). Using k-means clustering in international location decision. *Journal of Global Operations and Strategic Sourcing*, *11*(3), 274–300. https://doi.org/10.1108/jgoss-11-2017-0056

S. M. Miraftabzadeh, C. G. Colombo, M. Longo and F. Foiadelli, "K-Means and Alternative Clustering Methods in Modern Power Systems," in *IEEE Access*, vol. 11, pp. 119596-119633, 2023, doi: 10.1109/ACCESS.2023.3327640.

Liao, K., & Guo, D. (2008). A Clustering-Based Approach to the Capacitated Facility Location problem1. *Transactions in GIS*, *12*(3), 323–339. https://doi.org/10.1111/j.1467-9671.2008.01105.x

Abo-Elyousr, F. K., Sharaf, A. M., Darwish, M. M. F., Lehtonen, M., & Mahmoud, K. (2022). Optimal scheduling of DG and EV parking lots simultaneously with demand response based on self-adjusted PSO and K-means clustering. *Energy Science & Engineering*, *10*(10), 4025–4043. https://doi.org/10.1002/ese3.1264

S. A. Azad, A. B. M. S. Ali and P. Wolfs, "Identification of typical load profiles using K-means clustering algorithm," *Asia-Pacific World Congress on Computer Science and Engineering*, Nadi, Fiji, 2014, pp. 1-6, doi: 10.1109/APWCCSE.2014.7053855.

Chicco, G. (2012). Overview and performance assessment of the clustering methods for electrical load pattern grouping. *Energy*, *42*(1), 68–80. https://doi.org/10.1016/j.energy.2011.12.031

Andrienko, G., Andrienko, N., Jankowski, P., Keim, D., Kraak, M., MacEachren, A., & Wrobel, S. (2007). Geovisual analytics for spatial decision support: Setting the research agenda. *International Journal of Geographical Information Science*, *21*(8), 839–857. https://doi.org/10.1080/13658810701349011

Erskine, M., Gregg, D., Karimi, J., & Scott, J. (2013). Business Decision-Making Using Geospatial Data: A research Framework and Literature review. *Axioms*, *3*(1), 10–30. https://doi.org/10.3390/axioms3010010

Daly, C. (2006). Guidelines for assessing the suitability of spatial climate data sets. *International Journal of Climatology*, *26*(6), 707–721. https://doi.org/10.1002/joc.1322

Bédard, Y., & Marchand, P. (n.d.). *Toward better support for spatial decision making: defining the characteristics of spatial on-line analytical processing (solap)*. GEOMATICA. https://cdnsciencepub.com/doi/abs/10.5623/geomat-2001-0061

Nath, S. S., Bolte, J. P., Ross, L. G., & Aguilar-Manjarrez, J. (2000). Applications of geographical information systems (GIS) for spatial decision support in aquaculture. *Aquacultural Engineering*, *23*(1–3), 233–278. https://doi.org/10.1016/s0144-8609(00)00051-0

Fotheringham, A. S., Charlton, M., & Brunsdon, C. (2000). *Quantitative Geography : Perspectives on Spatial Data analysis*. Fotheringham, a Stewart - Charlton, Martin - Brunsdon, Chris - Sage - Torrossa. https://www.torrossa.com/en/resources/an/4913049