**PROJECT TITLE**

**FINDING THE OPTIMUM WAREHOUSE LOCATIONS**

**SUBMITTED BY**

**Mr. Sujeet Virendra Gupta**

**M.Sc. (Applied Statistics) PRN:22060641006**



**THE ACADEMIC YEAR 2022 - 23**

**Email Id:**[**sujeetgupta662000@**](mailto:sujeetgupta662000@)**gamil.com**

**Mobile: 8788721910**

# AIMS & OBJECTIVES

The objective of the company is to make two key decisions regarding the placement of

warehouses and the allocation of customers to those warehouses. The aim is to establish the minimum number of warehouses required while ensuring that all customers are served

efficiently.

To achieve this objective, the company needs to take into account various constraints, such as the maximum distance of 1000km between the warehouse and the customer demand centers. The

company must analyze the geographic locations of all demand centers and determine the optimal placement of warehouses based on factors such as demand density, transportation costs, and

accessibility.

Once the warehouse locations have been identified, the company must assign each customer to the appropriate warehouse based on factors such as proximity, demand volume, and shipping

costs. By making these decisions carefully, the company can minimize the number of

warehouses needed while ensuring that all customers receive timely and cost-effective service.

# METHODOLOGY

The dataset we analyzed includes a matrix that displays the distances between one city and all the other cities in the USA, as well as the corresponding demand for each city.

## DATA PREPROCESSING

The dataset used for clustering analysis did not contain any missing values or irrelevant features. However, two outliers were identified during the analysis, which were the cities from Puerto

Rico. These outliers were removed from the dataset as they were geographically distant from the rest of the cities in the USA (the minimum distance from them was more than 1500 km) and therefore could not be considered for clustering purposes. Additionally, since the data consisted only of distances, there was no need for data normalization or transformation.

## DISTANCE METRIC SELECTION

For clustering analysis, k-means clustering was used, which is a distance-based algorithm. The Euclidean distance metric was selected to calculate the distance between the data points, as it is the most commonly used distance metric and suitable for numerical data.

## CLUSTERING ALGORITHM SELECTION

The clustering algorithm chosen for the analysis was k-means clustering, which is a popular and widely used unsupervised machine learning algorithm. As K-means clustering is known for its

simplicity and efficiency in clustering large datasets.

Also, Another reason why k-means clustering was chosen for the analysis is that it was required to cluster the cities based on their geographical distance from each other. Specifically, the goal

was to form clusters where the cities within each cluster were at a maximum distance of 1000 km from the cluster centroid. K-means clustering is well-suited for this task as it aims to minimize the sum of squared distances between data points and their assigned cluster centroid. This

ensures that the resulting clusters are compact and have a small within-cluster variance, which aligns with the desired outcome of clustering the cities based on their distance from each other.

## DETERMINATION OF THE OPTIMAL NUMBER OF CLUSTERS

In order to determine the optimal number of clusters to form, an elbow plot was generated. The elbow plot is a graphical method that displays the relationship between the number of clusters and the within-cluster sum of squares. The point of inflection in the elbow plot indicates the optimal number of clusters, which was selected based on the trade-off between the decrease in the within-cluster sum of squares and the increase in the number of clusters.

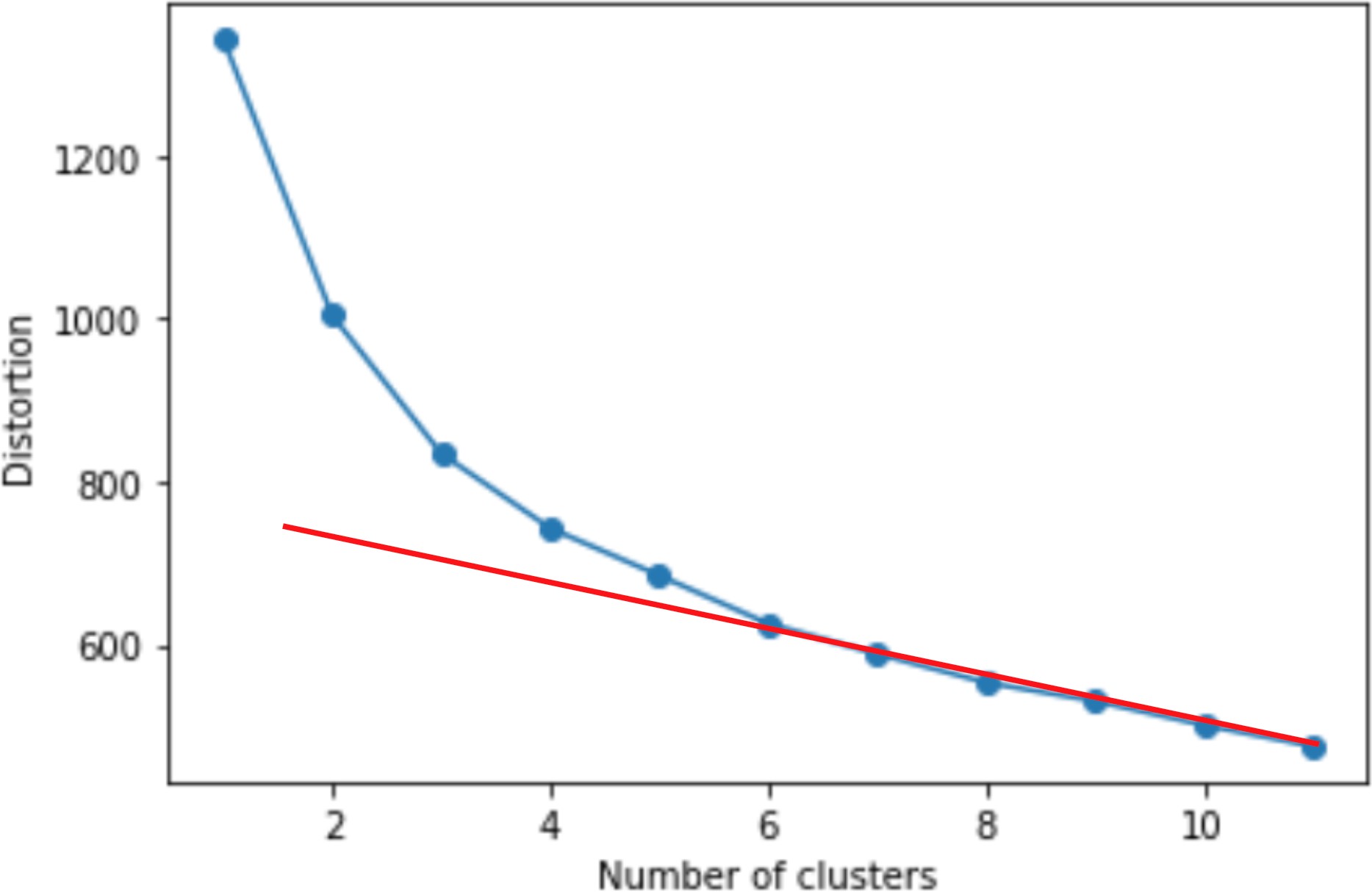


Fig.1 Elbow plot for the optimal number of clusters

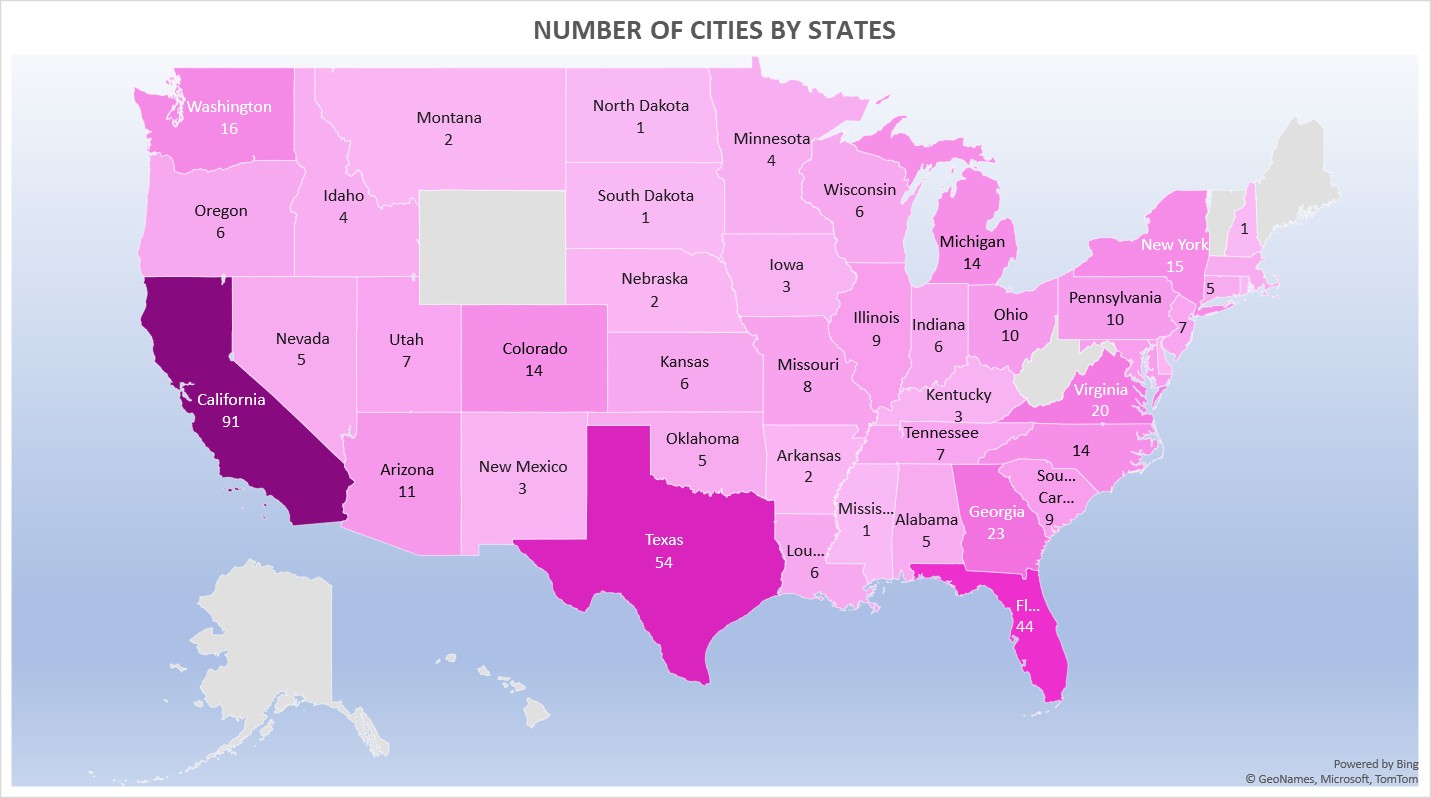
Hence, the optimal number of clusters chosen for clustering was 7.

## CLUSTER ANALYSIS AND VALIDATION

After forming the clusters, the next step was to validate and analyze them to ensure they were meaningful and relevant to the main objective. The clusters were examined to confirm that they consisted of cities that were within the desired maximum distance of 1000 km from their respective cluster centroids. Based on this analysis, it was concluded that the clusters formed

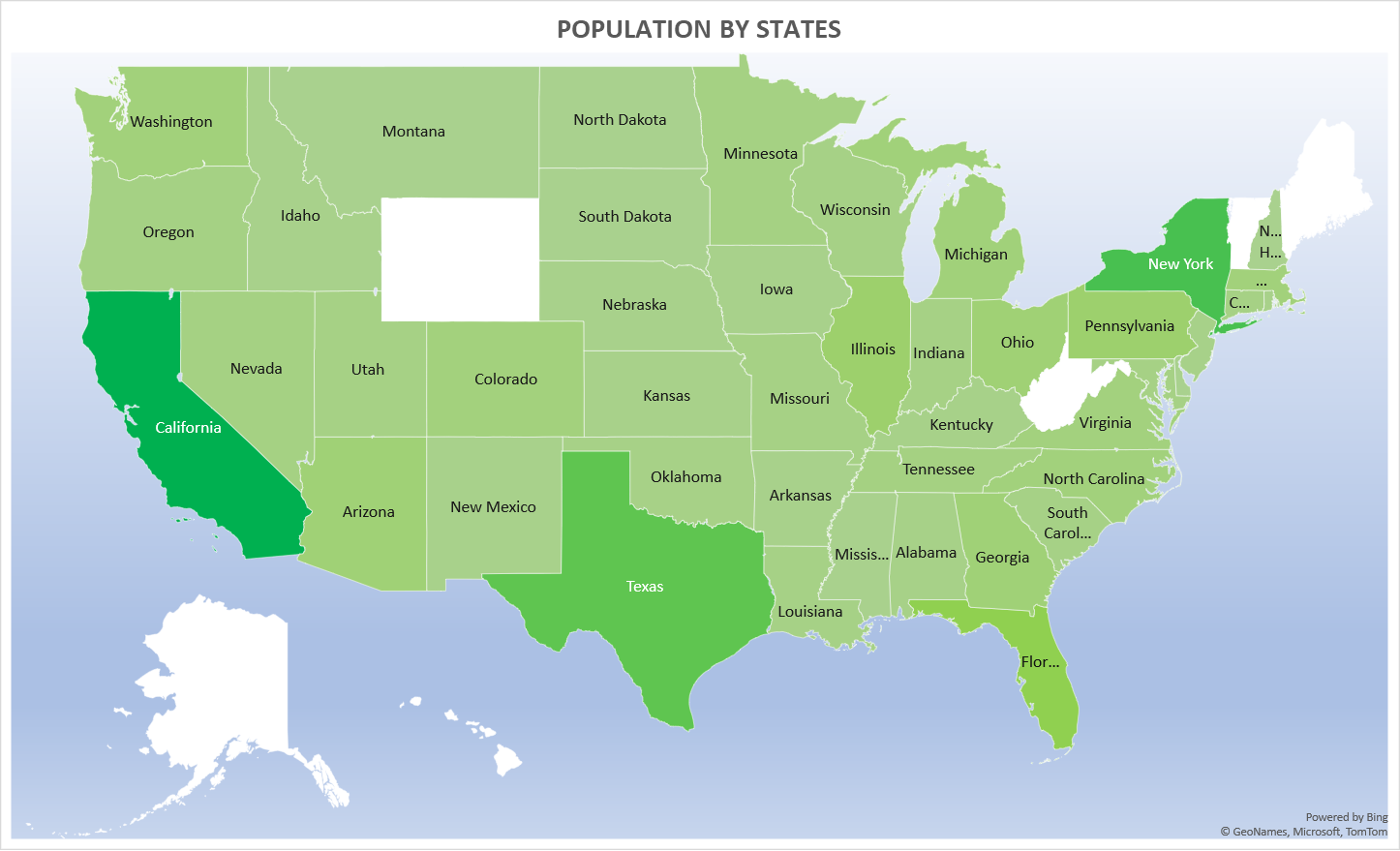
were optimal and satisfied the main objective of the clustering analysis. Therefore, the cluster analysis and validation were successful in meeting the project's objectives.

## RESULTS

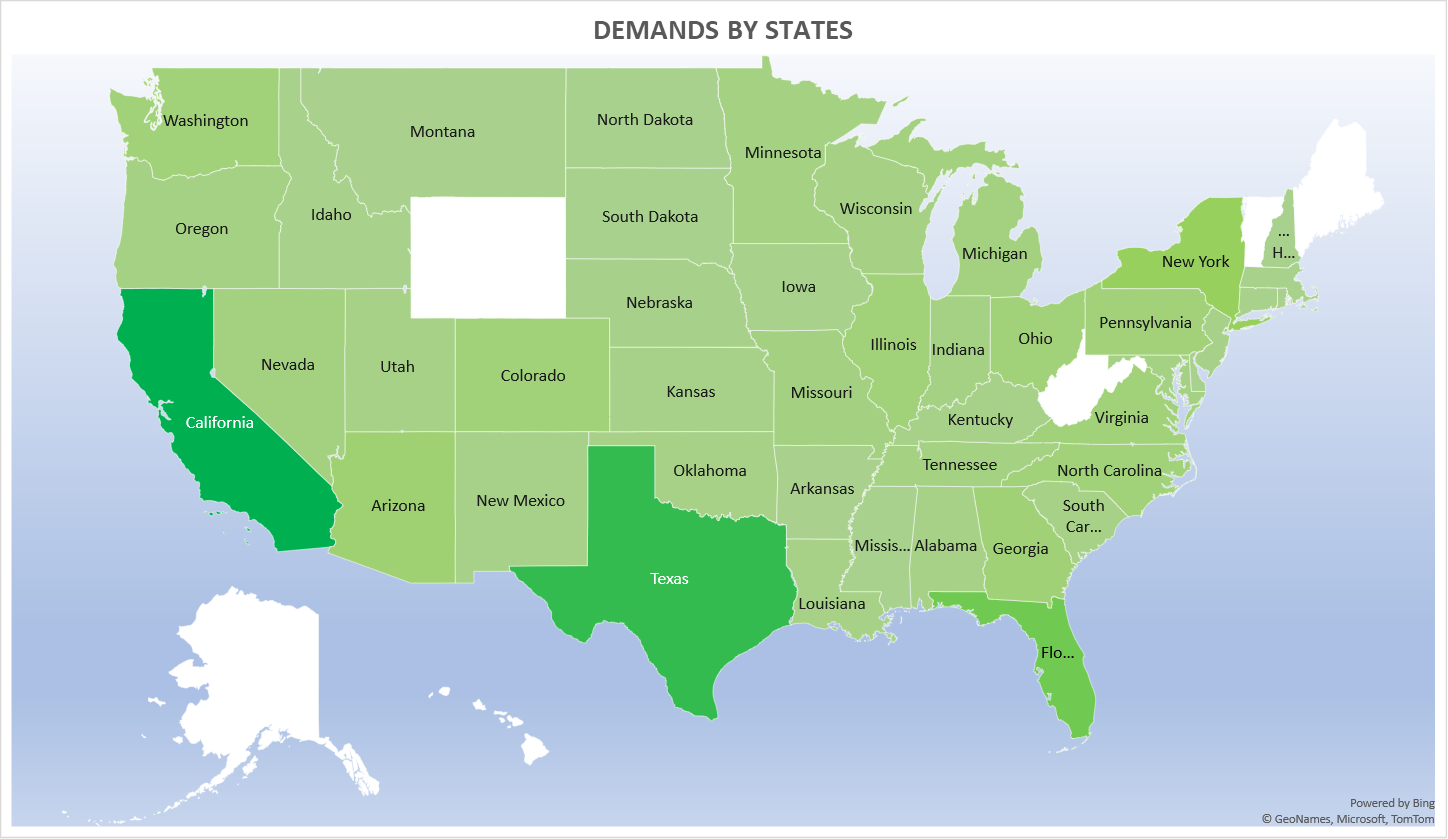


**Fig.1**

From Fig.1 we can observe that the states with a higher number of cities are represented by darker regions on the map. Specifically, California, Texas, Florida, Georgia, Virginia, New York, Michigan, Washington, and North Carolina are the states that have more than 15 cities to meet the demand.



**Fig.2**



**Fig.3**

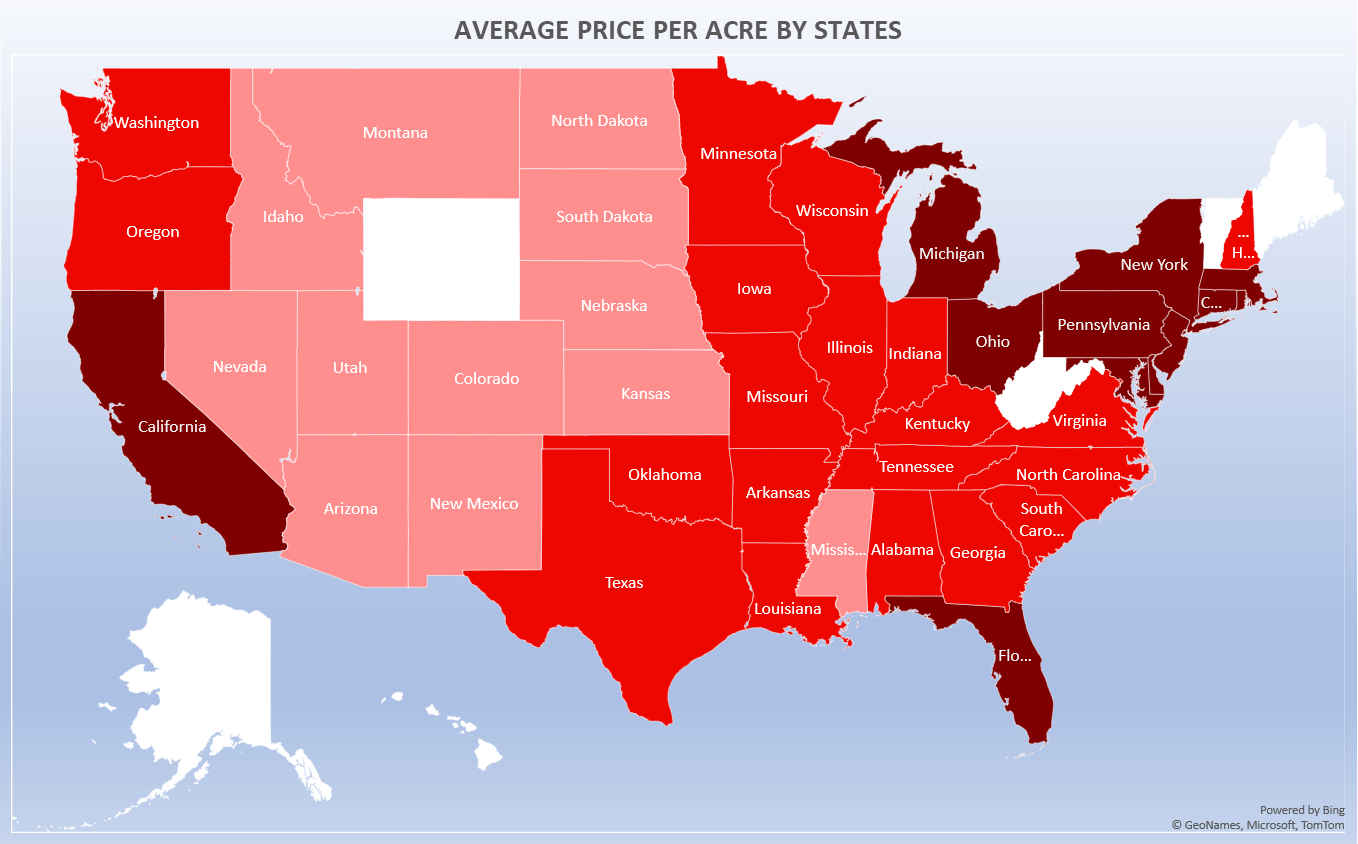
From Fig.2 we can observe that the states with higher number of population are represented by darker regions on the map.

From Fig.3 we can observe that the states with a higher number of demands are represented by darker regions on the map.

By observing both the Fig.2 and Fig.3 we can conclude that the regions having higher populations are the regions having higher demands and vice versa.

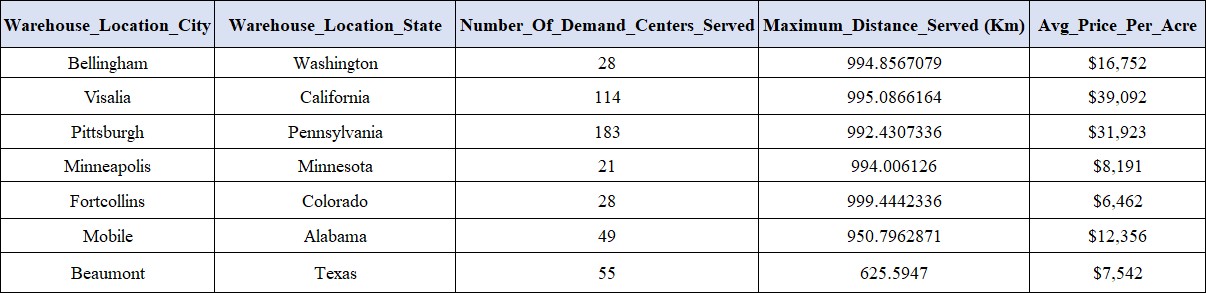
Also, the correlation between population and demand comes out to be 0.94.

Therefore, we can infer that there is a high positive correlation between population and demand in these regions.



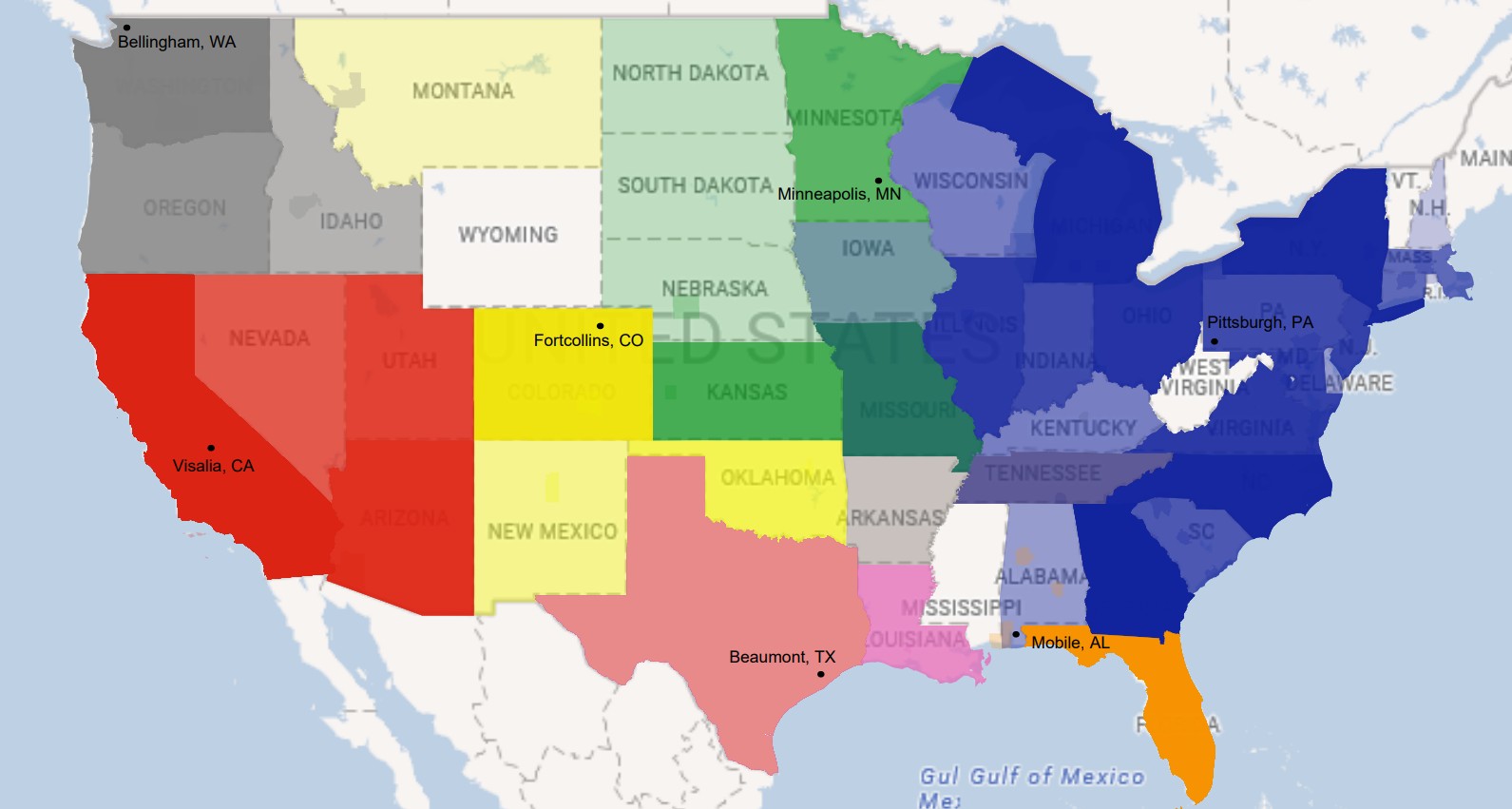
**Fig.4**

Figure 4 displays that darker regions on the map correspond to states with high prices per acre of land, indicating a variation in land prices across different states. Specifically, California, Florida, New York, Michigan, Pennsylvania, New Jersey, Rhode Island, and Massachusetts are the states that have particularly high land prices throughout the state.



**Table.1**

The Table.1 above provides information regarding the locations of the warehouses to be constructed, along with the number of cities each warehouse can serve in meeting the demands. Additionally, it indicates the maximum distance between the warehouse and the farthest city that requires fulfillment of demand. Also, the average price per acre of land in each state where the warehouse is to be constructed is been provided.



**Fig.5**

Fig.5 provides a visual representation of the clusters, which includes the locations of the warehouses and the regions that will have their demands fulfilled by these warehouses.

Additionally, if any city falls within a maximum distance of 1000 km from the warehouse location, it may also be served as a secondary distribution center, even if it is not part of the same cluster.

## DISCUSSION AND CONCLUSION

In conclusion, based on the information presented in the table, it can be concluded that setting up warehouses in seven specific locations, namely Bellingham-Washington, Visalia-California, Pittsburgh-Pennsylvania, Minneapolis-Minnesota, Fortcollins-Colorado, Mobile-Alabama, and Beaumont-Texas, would be sufficient to fulfill the demands of all 480 cities. These seven locations would serve as primary distribution centers, catering to a total of 478 cities, with the number of cities served by each location being 28, 114, 183, 21, 28, 49, and 55, respectively.

Importantly, all the cities served by these distribution centers are located within a maximum distance of 1000km.

## RECOMMENDATION & LIMITATIONS

The warehouses that were identified as optimal were chosen solely based on their distance,

without considering other factors such as maintenance, costs, and complexity. Therefore, while they may be useful in fulfilling a particular aspect of the business, they may not be beneficial overall. As a result, there are two approaches to consider: having multiple smaller warehouses or fewer, more advanced larger warehouses. Both approaches have their advantages and disadvantages, and the selection of the approach depends on the specific needs of the business

and organization.

In general, having multiple smaller warehouses can offer advantages such as:

1. **Faster Delivery Times:** With more distribution centers, it becomes easier to reach

customers in different locations quickly, which can help to improve customer satisfaction and loyalty.

1. **Lower Transportation Costs:** By having more distribution centers, companies can often reduce transportation costs by shipping products shorter distances and using smaller vehicles.
2. **Reduced Dependence on a Single Location:** Having more distribution centers can help to reduce the risk of disruptions, such as natural disasters or supply chain disruptions, by spreading inventory across multiple locations.

However, having multiple smaller warehouses can also come with some tradeoffs, such as:

1. **Higher Inventory Carrying Costs:** Maintaining inventory across multiple distribution centers can result in higher inventory carrying costs, as companies may need to stock more inventory to meet customer demand in different locations.
2. **Increased Complexity:** Managing multiple distribution centers can be more complex and require more sophisticated logistics and inventory management systems to ensure that inventory is distributed efficiently and effectively.
3. **Higher Upfront Costs:** Establishing additional distribution centers can require a

significant upfront investment in infrastructure, such as building and equipping the facilities, installing technology systems, and hiring and training staff.

In contrast, having fewer, more advanced larger warehouses can offer advantages such as:

1. **Lower Inventory Carrying Costs:** By consolidating inventory in fewer distribution centers, companies can often reduce inventory carrying costs by stocking fewer items.
2. **Reduced Complexity:** Managing fewer distribution centers can be simpler and require less sophisticated logistics and inventory management systems.
3. **Lower Upfront Costs:** Establishing fewer distribution centers can require a lower upfront investment in infrastructure and staffing.

However, having fewer, more advanced larger warehouses can also come with tradeoffs, such as:

1. **Longer Delivery Times:** With fewer distribution centers, it may take longer to reach customers in different locations, which can impact customer satisfaction and loyalty.
2. **Higher Transportation Costs:** By having fewer distribution centers, companies may need to transport products longer distances and use larger vehicles, which can result in higher transportation costs.

Ultimately, the optimal number of distribution centers will depend on the specific needs and characteristics of a company's supply chain, such as the volume of sales, customer demand patterns, transportation costs, inventory carrying costs, and overall profitability goals. A thorough analysis of these factors can help companies determine the most profitable and

effective distribution strategy for their business.

## REFERENCES

* <https://phoenix.tech/>
* <https://www.mapsofworld.com/lat_long/usa-lat-long.html>
* <https://www.landsearch.com/price>
* <https://towardsdatascience.com/k-means-clustering-for-beginners-2dc7b2994a4>
* [https://www.gep.com/blog/technology/4-steps-in-successful-supply-chain-inventory-man](https://www.gep.com/blog/technology/4-steps-in-successful-supply-chain-inventory-management) [agement](https://www.gep.com/blog/technology/4-steps-in-successful-supply-chain-inventory-management)