Population-Based Covid-19 Distribution Centers

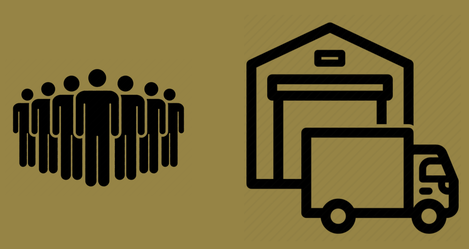
Obaida Albaroudi, Meghana Manne, Wilmer Tejada, Adam Aalto, and Vikas Krishna Manne

Problem Definition & Motivation

Although the vaccine has been essential in the fight against COVID-19, it is lacking in certain key aspects. The most notable of which involves its efficient distribution methods or rather, its lack thereof. As of now, the methods being utilized to distribute the vaccine are notorious for being ill-conceived and haphazard, a fact further perpetuated by the significant number of doses that have been left to expire. A large part of this inadequacy and wastefulness can be attributed to the less than optimal locations of the distribution hubs, specifically the location of the 18 FedEx and UPS centers that are used to distribute the vaccine in the United States. In order for individuals to be vaccinated as quickly as time allows, distribution centers should be positioned in places that are in close proximity to heavily populated areas in the US which would in return contribute to quick and methodical transportation.

How it works

Inputs:



Population

Existing Centers

Model:

- *Weighted k-means model->Adjusted and created our own weight system based off population sizes*
- *Looked at the population size of different cities along with their latitude and longitude*
- *Final model output returns cluster centers*

User Interface:

- *Used Shiny, an R package that allows us to easily host web applications*
- *Through a slider, the interface allows the user to determine the number of k-means clusters that should be used to train the underlying model with*
- *An elbow graph acts as a recommendation tool to select an optimal number of cluster centers*
- *User can compare locations of currently existing distribution centers with model based cluster centers*

Data

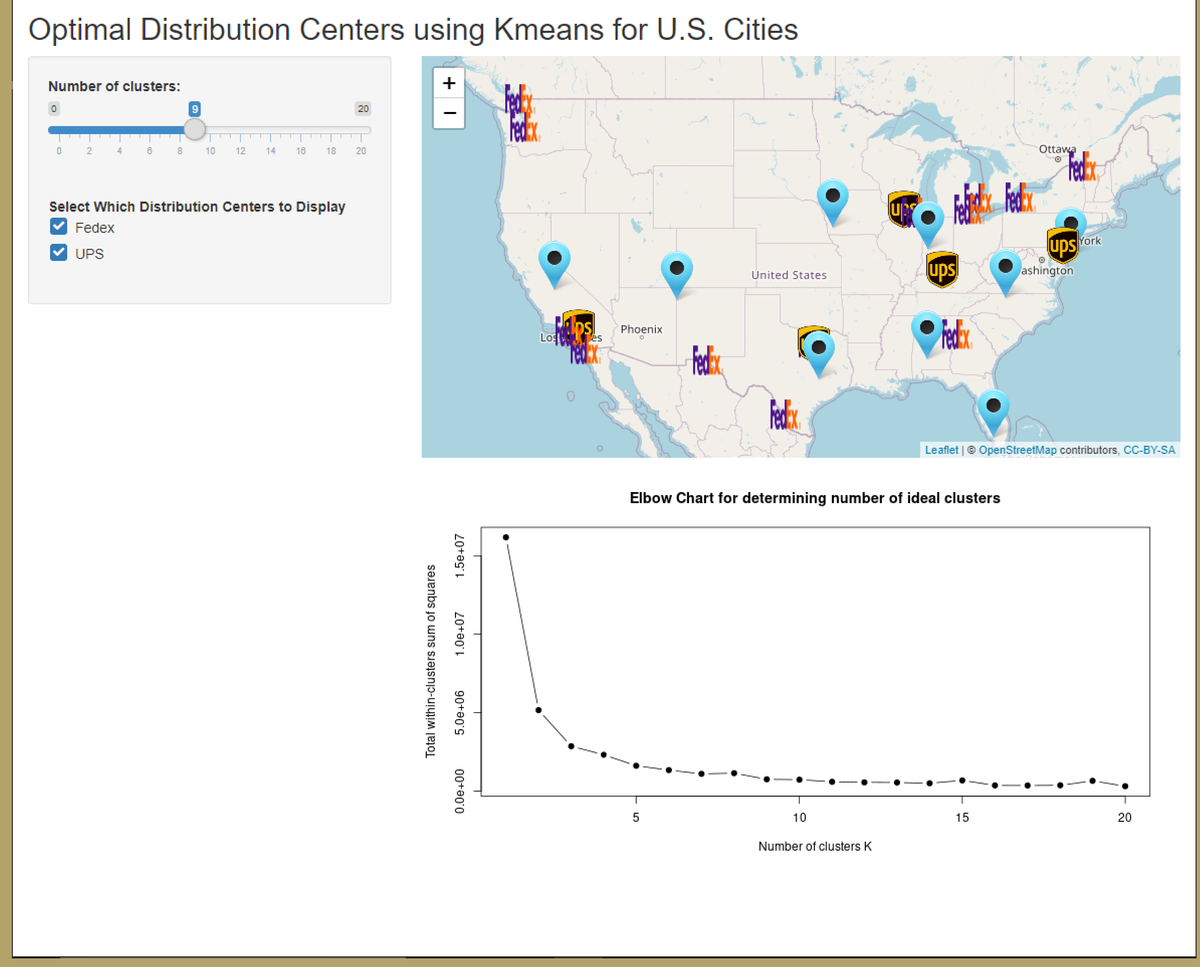
- The data was provided by the US Census Bureau and was obtained through an online database that houses several types of geographical datasets.
- The link for this dataset is as follows: https://simplemaps.com/data/us-cities
- The data consists of 19,501 rows which tell us the City, State, Population, and several other metadata values regarding that specific city.

Methods

- Weighted k-means clustering was chosen as it is a simple but also popular unsupervised machine learning model. Additionally, It is the go to method for the density-based clustering we were hoping to achieve.
- We cleaned up the data by scaling the population size per city.
- In our model, we only focus on the contiguous US (all states excluding Hawaii and Alaska) as opposed to all fifty states to ensure every cluster is located on land as opposed to a body of water.
- We overlaid the cluster centers from our model with the actual locations of the COVID-19 distribution centers to allow us to compare between the two and determine if there are any differences
- The interface also permits users to hover over the points on the map to display a tooltip that specifies the latitude and longitude points of these different centers to get more precise details on the location of each.

User Interface

https://wilmer-tejada.shinyapps.io/Kmeans_USCities/



Results

- Our final output was able to address the following: What is the optimal number of distribution centers? In which regions will the distribution centers be located? Will some regions have more or less distribution centers than others? How close are the existing distribution centers to the ones from our model?
- Based off our elbow chart, there should be 9 centers as that is considered the most optimal value
- Distribution centers tend to be scattered along the middle of the United States with just one or two centers polarized to the south
- There are a greater number of centers located towards the east coast than the west coast
- The existing distribution centers are primarily all scattered along the coast whereas the opposite stands true for the model-induced centers.
- Primarily, the centers obtained from the model do not line up with existing distribution centers. However, anywhere between 1-4 are in close proxmity depending on the number of centers chosen.

Conclusion

- **Based on the locations of the existing distribution centers, we can resonably conclude one of our initial assumptions: Existing hubs were chosen from a number of peripheral factors, likely including foreign imports and shipping channels.**
- **Our results show that if looking to vaccinate individuals in a quick and effective manner then these existing considerations are not ideal for choosing the locations of vaccine distribution hubs**
- **We believe that if utilized, our model which adheres to a population-based algorithm, can maximize impact and vaccinate as many individuals as rapidly as possible.**