

Coursera Capstone Project – Where to put a new pharmacy

Introduction:

In light of COVID-19 the most recent and devastating pandemic across the globe, New York has seen record numbers of hospital patients, and due to this has had to bring in a temporary hospital in Central Park and a Hospital ship into the harbour.

Therefore, to help with the medical needs of New York citizens, more pharmacies could be required.

Business Problem:

The objective of this project is to select the optimum location in New York to open a new pharmacy. Therefore, the question is for the government or business owners, using data science and clustering techniques, where would we recommend they build a new pharmacy?

Target Audience:

The audience of this project is not only the government and business owners, to help advise where a new pharmacy should be located, but also the citizens of who would benefit from a new pharmacy in their neighbourhoods.

Data:

Required data:

- List of neighbourhoods in New York.
- Latitude and longitude of these neighbourhoods.
- Data of current hospital locations.

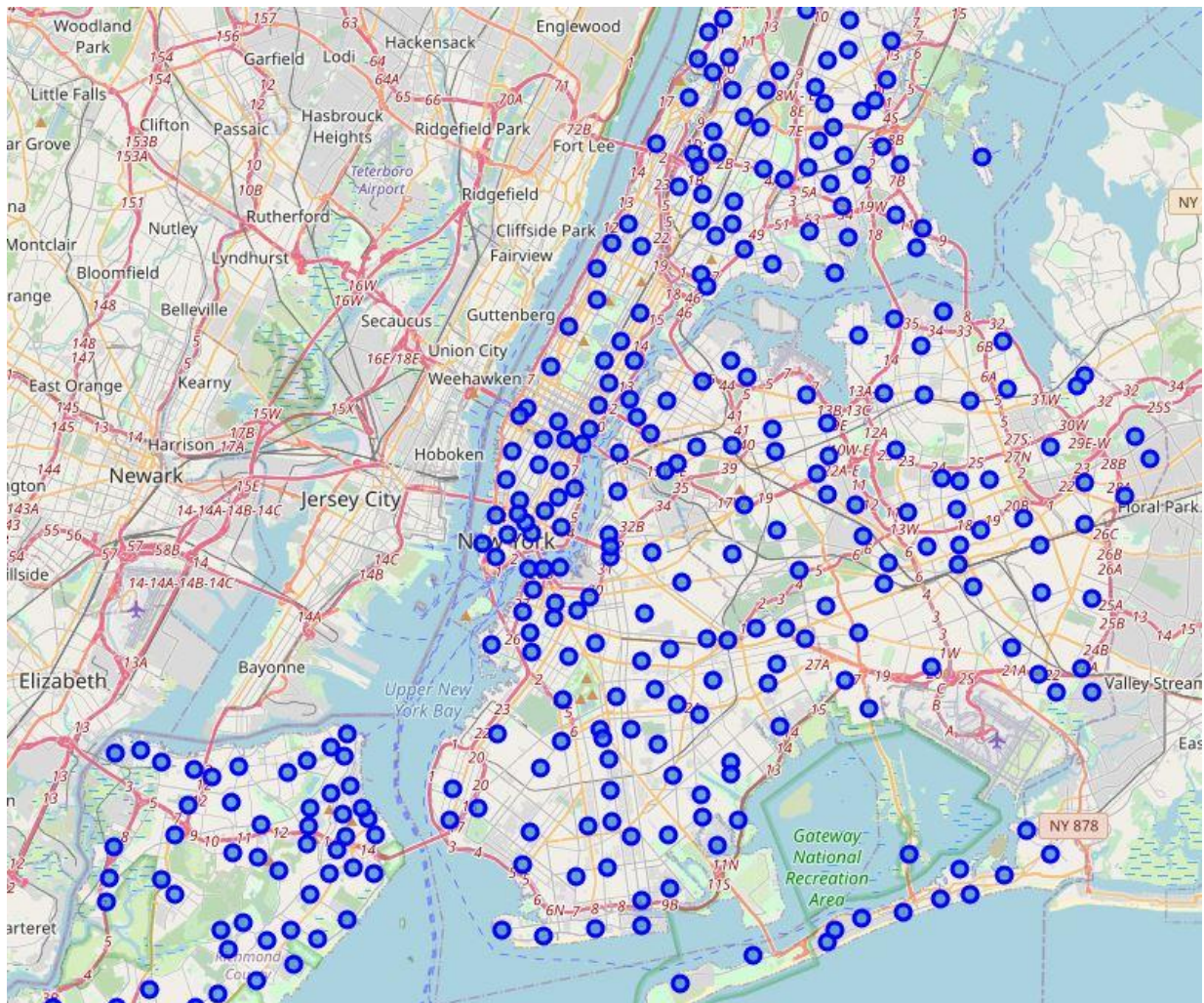
Data Sources:

-https://geo.nyu.edu/catalog/nyu_2451_34572 for New York neighbourhoods data.

-Foursquare API for hospital data

Methodology:

First, we get the list of neighbourhoods in New York from the dataset online. Then we get the geographical coordinates of each neighbourhood using Geocoder, this is then put into a pandas data frame and visualised using Folium.



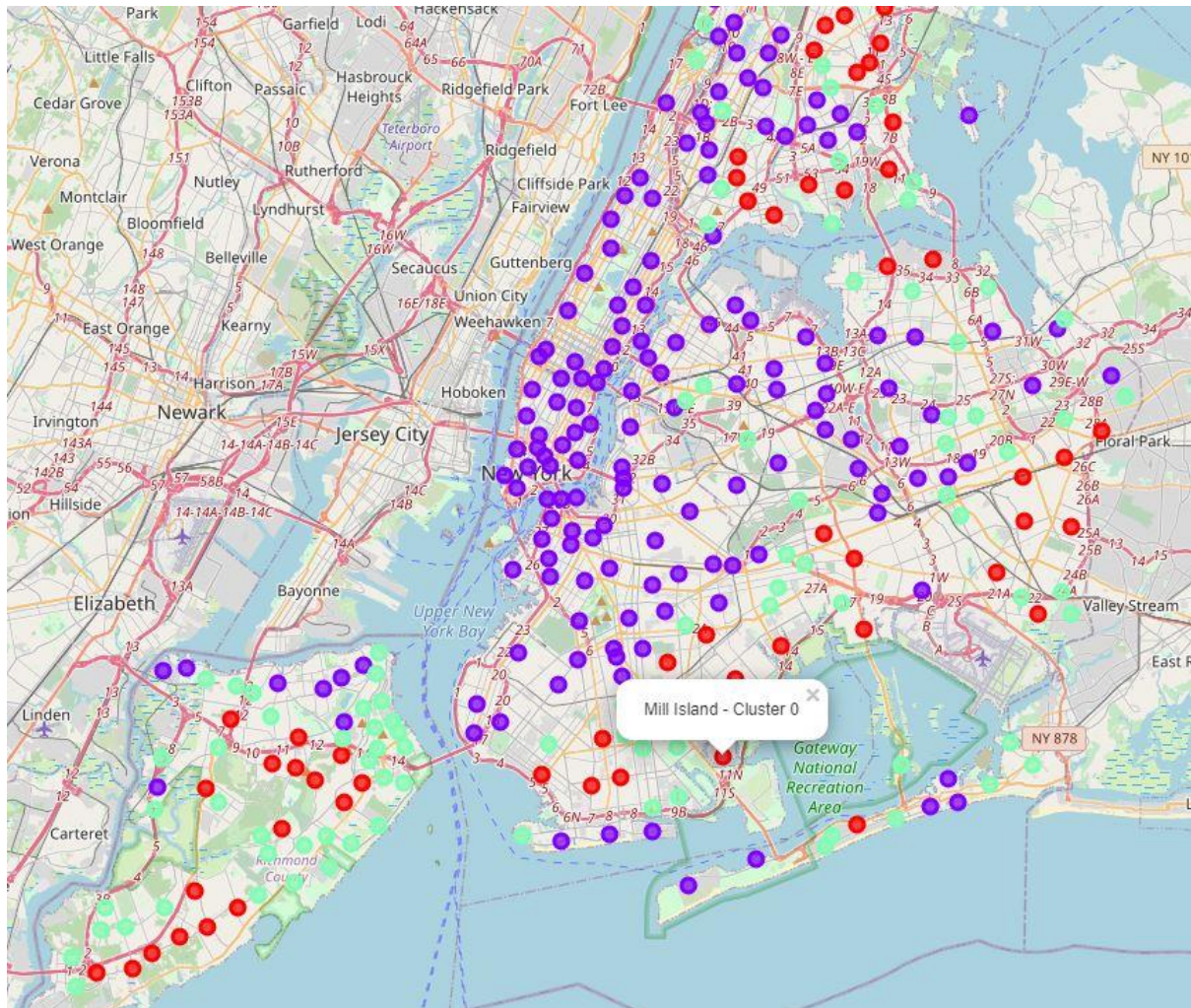
Next, using Foursquare API we get top 100 venues in 2km from each neighbourhood. We then extract this data from the json file we get from foursquare. We can then check that pharmacies are in this data and then we are able to focus on the mean number of pharmacies within each neighbourhood.

Finally, we use K-means clustering (a unsupervised Machine Learning algorithm) to allocate each data point to the nearest cluster. We cluster the data points into 3 distinct clusters, depending on the distance from a pharmacy.

Results:

- Cluster 0: Neighbourhoods with a high number of pharmacies.
- Cluster 1: neighbourhoods with a low number of pharmacies.
- Cluster 2: Neighbourhoods with a medium number of pharmacies.

In the map below this is shown as red for cluster 0, purple for cluster 1 and green for cluster 2.



Discussion:

Observations from the map show us that central New York (Manhattan) has the least number of pharmacies possibly due to its high real estate value and tourist focus. This may be a good opportunity to build pharmacies in this area as there is little competition and a possible large need. Outside the centre we have a mix of medium and high amount of pharmacies, so this is not a good place for prospective business owners to place a new pharmacy.

Furthermore, this model only uses location data and there are other factors that should be considered, such as age demographics.

Conclusion:

We have shown through the method of identifying the problem, accessing the data, performing machine learning clustering on the data and analysis where there is a lack of pharmacies. We have shown that central New York is the optimum place for a new pharmacy and this should help the government, business owners and other stake holders when making a decision on where the best location may be.