IBM Applied Data Science Capstone

Opening a New Mall in Vancouver, British Columbia

By: Andrew Shrestha

Business Problem

- The objective of this Capstone Project will be to utilize data and select the best locations in the city of Vancouver to open a new Shopping Mall.
- Through the use of techniques such as foursquare analysis, clustering, and segmenting, we would like to solve the following business problem:
- “Where would be the optimum location to construct a new mall for property developers/owners in the city of Vancouver?”

Data

- A list of neighborhoods in Vancouver. These will be the locational data that we will confine our analysis to
- Both the Latitude and Longitude coordinates of each of the neighborhood data points so that we may plot it on a map
- Data on Shopping Malls or related so that we may use this in order to perform clustering on the neighborhoods

Data Source

- Obtain a list of 45 neighborhoods from Wikipedia (https://en.wikipedia.org/wiki/List_of_neighbourhoods_in_Vancouver)
- Web scraping techniques to grab the data from Wikipedia (python and Beautifulsoup)
- Geocoder to extract longitude and latitude of the neighborhoods
- Foursquare API to obtain shopping mall venue data of each neighborhood
- Folium for map visualizations

Methodology

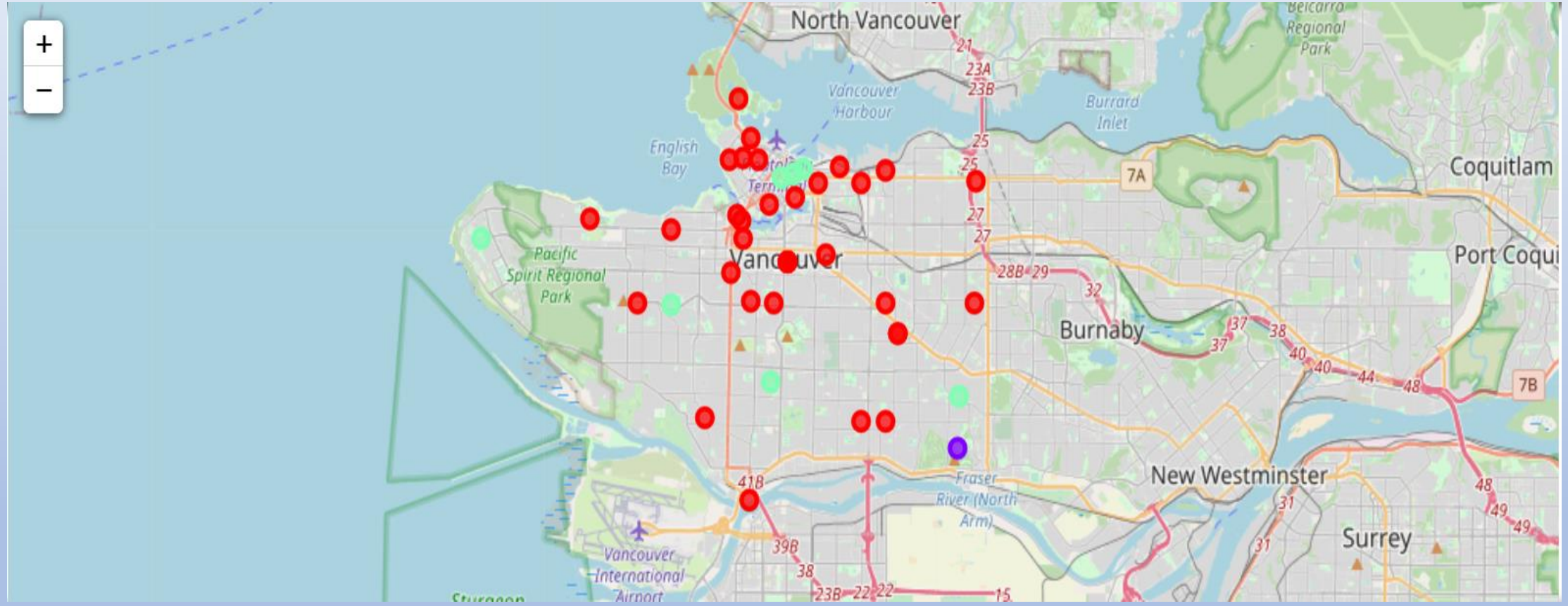
- Extracting Wikipedia Neighborhood data via Web Scraping
- Utilizing Geocoder to obtain the coordinates of Longitude and Latitude
- Venue data via the use of foursquare API
- Grouping and separating neighborhood according to the frequency of “Shopping Malls” present
- Cluster with k-means on the Data
- Create a visual representation of the Data with the use of Folium

Results

- Clustering the Neighborhoods into 3 different clusters where:

1) Cluster 0 (Red) – contained the number of neighborhoods with zero presence of Shopping Malls
2) Cluster 1 (Purple) – contained the number of neighborhoods with the most significant amount of Shopping Mall presence
3) Cluster 2 (Green) – contained the number of neighborhoods with a moderate amount of Shopping Mall Presence

Visual Results



Discussion

- Cluster 0, which was the one with no presence of Shopping Malls was the one with the greatest data points and span the majority of Vancouver.
- Cluster 1, with the highest amount of Shopping Mall presence was only located in one area, while the moderate cluster 2 was mainly located in downtown
- This means that Shopping Malls are only built in highly concentrated areas and neglect quite a wide range of geographical neighborhood locations

Conclusion

- Open a new mall in Cluster 0 where there is an opportunity to meet the demand of more suburban areas of Vancouver
- This will also make it more accessible and reduce the amount of travel time that the majority of neighborhood population has to travel to reach a shopping mall, that might also result in less online purchases and drawn the organic traffic to this new mall
- Avoid building a new Shopping Mall in Cluster 1 or even 2 as these areas already have a moderate to high concentration of existing Malls, which would create competition and will not be optimum for increased profits