Shopping Mall and Department Store Location Selection in New York

Charles Xi Yan
Applied Data Science: Capstone
Coursera: IBM Data Science

27th March, 2020

Business Problem

- Importance of location for a shopping mall or a department store.
- Objective: analyze the neighbourhoods in New York and select the optimal location.
- Oversupply of shopping malls and department stores.
- Problem Statement:
  - If a real estate developer is planning to open a new shopping mall or a department store in New York, where should be recommended?

Data

- Data Required:
  - List of neighbourhoods in New York.
  - Latitude and longitude coordinates of these neighbourhoods.
  - Venue data, particularly related to shopping malls and department stores.
- Source of data
  - The neighbourhoods data in New York (`https://cocl.us/new_york_dataset`).
  - Geopy package for latitude and longitude coordinates.
  - Foursquare API for venue data.

Methodology

- Load the data (wget command).
- Request latitude and longitude through geopy.
- Venue data through Foursquare API.
- Group data by neighbourhoods: the mean of the frequency of occurrence of each venue category.
- New dataframe: Shopping Mall, Department Store.
- K-mean clustering analysis.
- Visualize the clusters in a map through Folium package.

Results

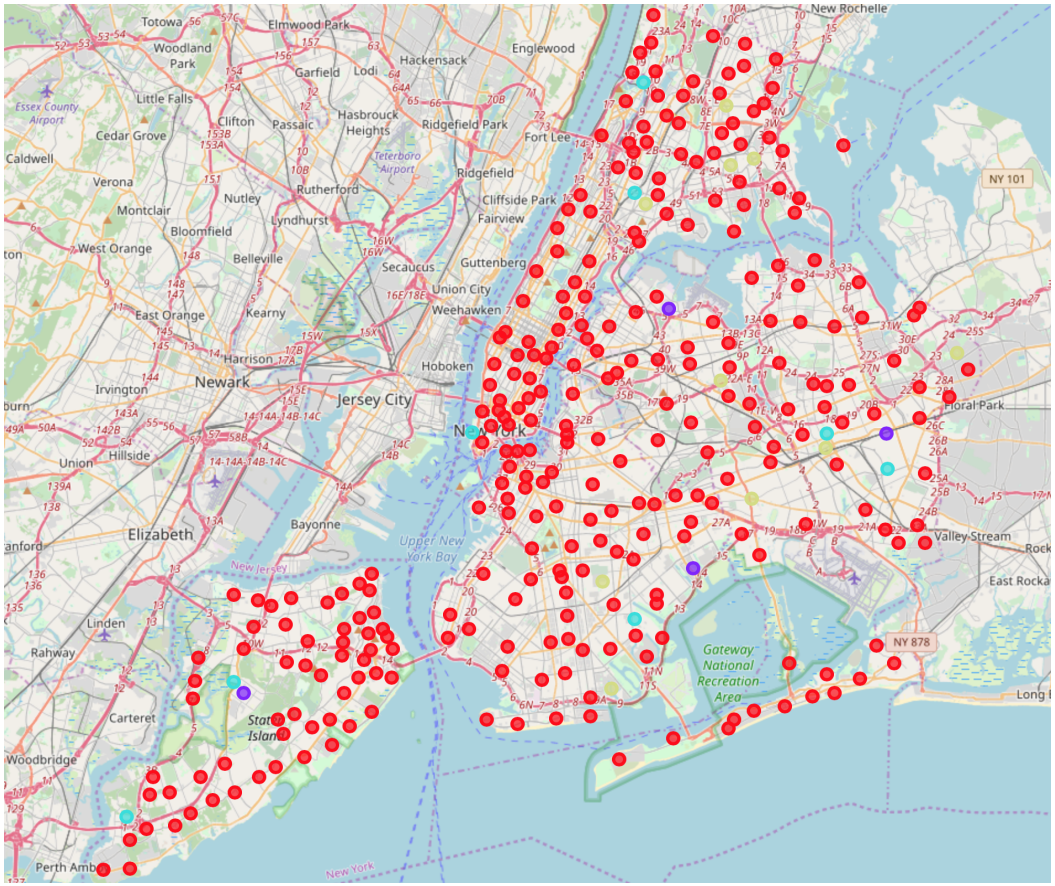

Figure: 4-mean Clustering on Neighbourhoods in New York City.

Results and Discussion

- Cluster 0 (Colour Red): Neighbourhoods with almost no existence of either shopping malls or department stores.
- Cluster 1 (Colour Purple): Neighbourhoods with high concentration of shopping malls but almost no existence of department stores.
- Cluster 2 (Colour Light Blue): Neighbourhoods with moderate number of both shopping malls and department stores.
- Cluster 3 (Colour Gold): Neighbourhoods with high concentration of department stores but almost no existence of shopping malls.

Recommendation and Conclusion

- For the neighbourhoods belonging to Cluster 0, from the perspective of competition due to oversupply and high concentration, both shopping malls and department stores can be considered.
- For the neighbourhoods belonging to Cluster 1, investing on a department store seems to have less pressure from the peer competition.
- For the neighbourhoods belonging to Cluster 3, investing on a shopping mall seems to have less pressure from the peer competition.
- For the neighbourhoods belonging to Cluster 2, the investment condition seems like a intermediate case of Case 2 and Case 3 above.

Limitation and Improvement

- Consider the safety scaling data and the average income of neighbourhood residents.
- Collect more data about shopping malls and departments to avoid serious peer competition.
- More experiments to find the optimal k.

Thanks!