Capstone Project

Selecting the best option to open a Sushi Bar in Manhattan, New York.

The Battle of Neighborhoods’

## Part 2

# Data Section

Data Selection

* To identify the characteristics of our competitors' venues in Manhattan, we would first need to find out the number of Sushi bars in Manhattan currently and their location.
* We then used Google Map API to find their geographic coordinates based on their postal code addresses.
* In Manhattan, there are 1763 Sushi bars are currently operating.
* <https://geo.nyu.edu/catalog/nyu_2451_34572> (1763, 7)
* Next, we also used Google Map API to find their geographic coordinates of the 5 locations shortlisted for our Sushi bar;

A screenshot of a cell phone

Description automatically generated

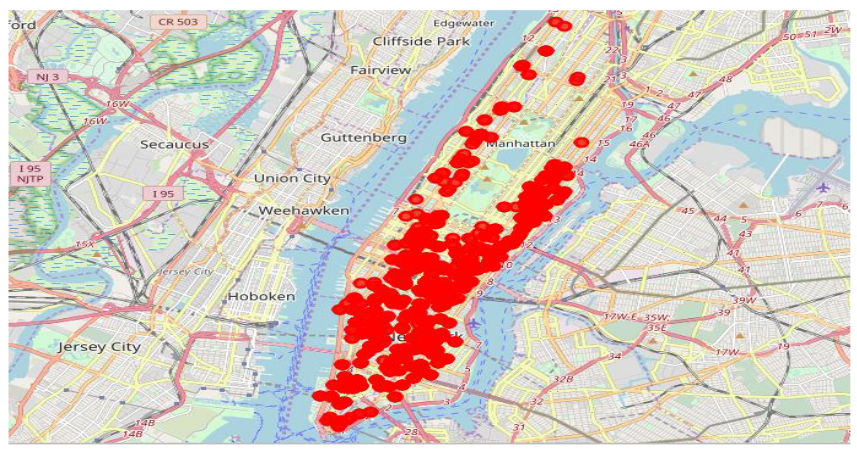
Methodology

* Addresses are converted into their equivalent latitude and longitude values.
* Foursquare API is used to explore neighbourhoods' in Manhattan, New York.
* A screenshot of a cell phone

  Description automatically generatedAfter that, explore function to get Sushi bar categories in each neighbourhood.

Table 3: Data frame containing counts of neighborhood.

\* Sushi bars in Manhattan



* Then using this feature to group the neighbourhoods into clusters K-means clustering algorithm will be used to complete this task. And also, the Folium library to visualize the neighbourhoods’ in Manhattan and its emerging clusters.

A screenshot of a computer

Description automatically generated

Results

* Using K-mean to clustering data area with less number of Sushi bars;

A picture containing sitting, computer, holding, street

Description automatically generated

Cluster 1

A screenshot of a cell phone

Description automatically generated

Cluster 2

A close up of a piece of paper

Description automatically generated

Cluster 3

A screenshot of a cell phone

Description automatically generated

A screenshot of a social media post

Description automatically generatedCluster 4

A screenshot of a cell phone

Description automatically generated

A close up of a map

Description automatically generated

Discussion

* This analysis is performed on limited data, results may vary. But if good amount of data is available there is opportunity to get better results.
* There is high competition in Midtown and Sohoso it is very risky to open business in these areas.
* Central Harlem has also potential where closes to Morning side Heights area.
* More detailed analysis is suggested by adding other factors such as transportation, demographics of inhabitants, etc.

Conclusion

* Although all of the goals of this project are met, there is definitely opportunity for further improvement.
* Apart, goals of the project, this can easily be extended into a fully functional application that could support the entrepreneurs’ in opening a new business in an unknown location.