Local Competitive Land Scape for Food Outlet Opening in New York city

IBM CAPSTONE PROJECT

Introduction

- Opening a new food outlet requires a right business strategy to have successful eatery business
- Location and Type of food outlet are crucial aspects of a business strategy
- They affect our ability to draw the customers
- Knowing exactly who we are up against is important before deciding about location and type of food offered
- It helps to determine whether a particular neighborhood is ripe for the picking or is over-saturated with competitors

Objective

- To provide valuable insight on local competitive landscape to those who want to establish new food outlet/eatery in New York city
- Given the location data of various types of food outlets in New York city, predict the common type of food outlets present in different parts of New York city
- It will be quite useful while in deciding about the location and type of new food outlet

Data Acquisition

- Obtain the data that contains the boroughs and the neighborhoods of the New York city
  - •also contains the latitude and longitude of each neighborhood of aborough
  - •are obtained from https://geo.nyu.edu/catalog/nyu_2451_34572 in a '.json' file
  - •5 Boroughs (Manhattan, Brooklyn, Queens, Staten Island, and Bronx) and 306 Neighborhoods are present in New York city
- Using FourSquare API, obtain the data of food outlet/eatery of each borough separately
- obtain a maximum of 25 food outlets present inside 500 meters radius of the neighborhood
- also obtain the type of food outlet/eatery

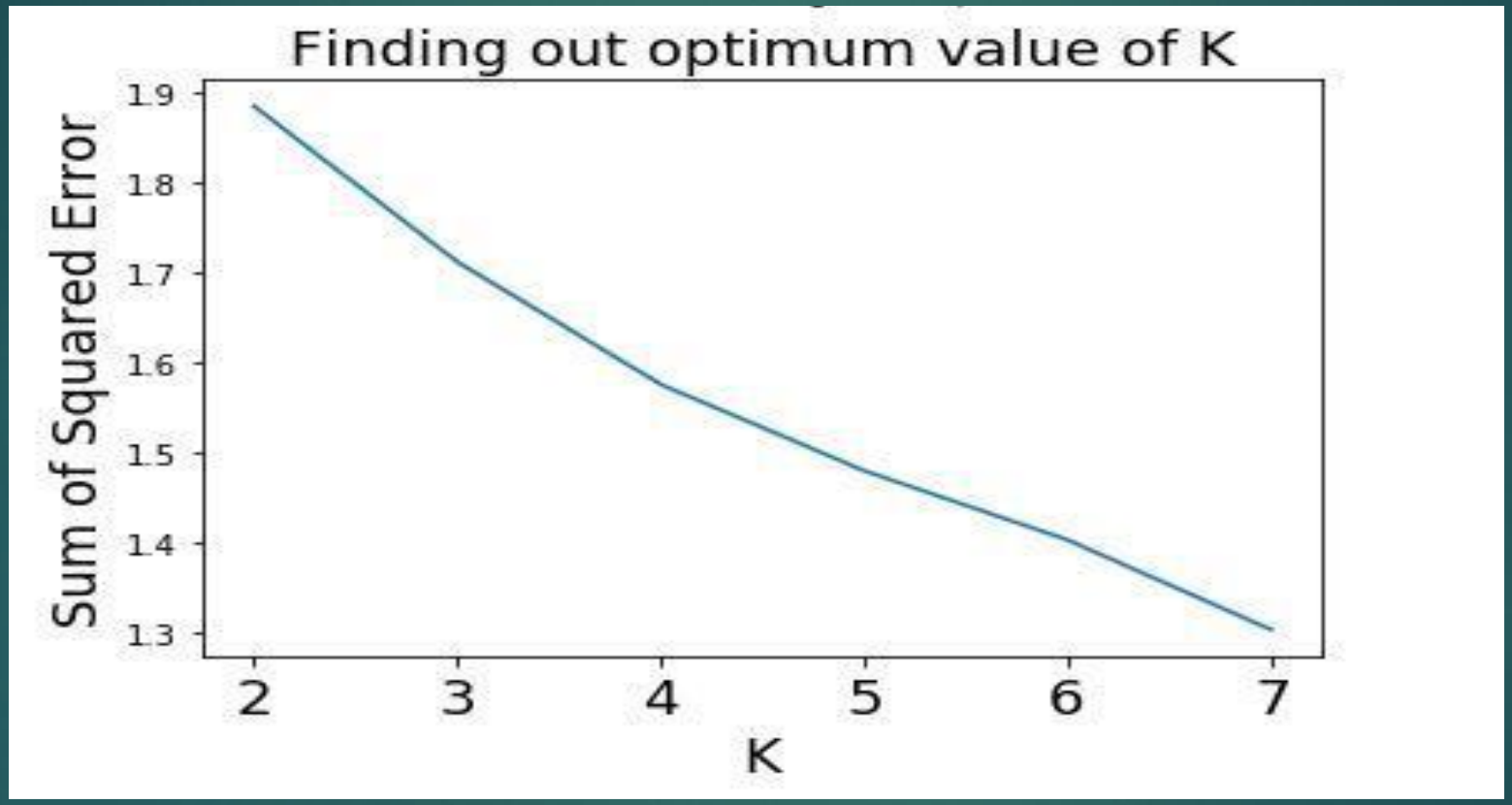
Data Processing

- For each borough of NY city, do the following:
- The food outlet data set of the borough is read into a pandas dataframe
- Perform one-hot coding on this dataframe based on the food outlet type
- Group the rows of this dataframe by neighborhood
- Take the mean of the frequency of occurrence of each type of food outlet
- Create new dataframe containing top 10 most common outlets present in each neighborhood

Data Modeling using Clustering

- Perform K-mean clustering on the dataframe containing top 10 common food outlets
- Perform clustering for different values of K
- Determine the optimum value using elbow technique
- Perform clustering of neighborhoods of a borough using optimum value of K
- Observe the type of food outlets present in each borough of NYcity
- Each cluster will have different types of common food outlets

Clustering of Manhattan

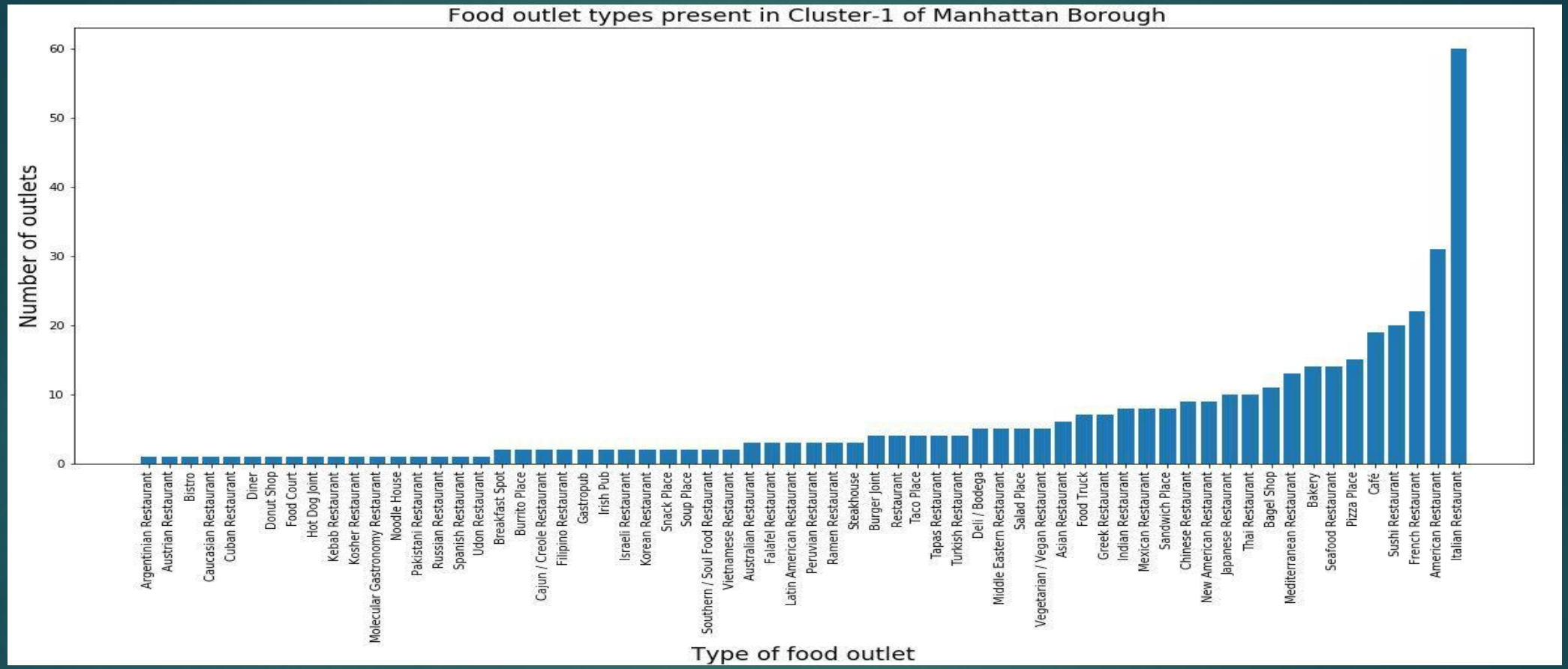


Optimum value of K is 4

Clustering of Manhattan

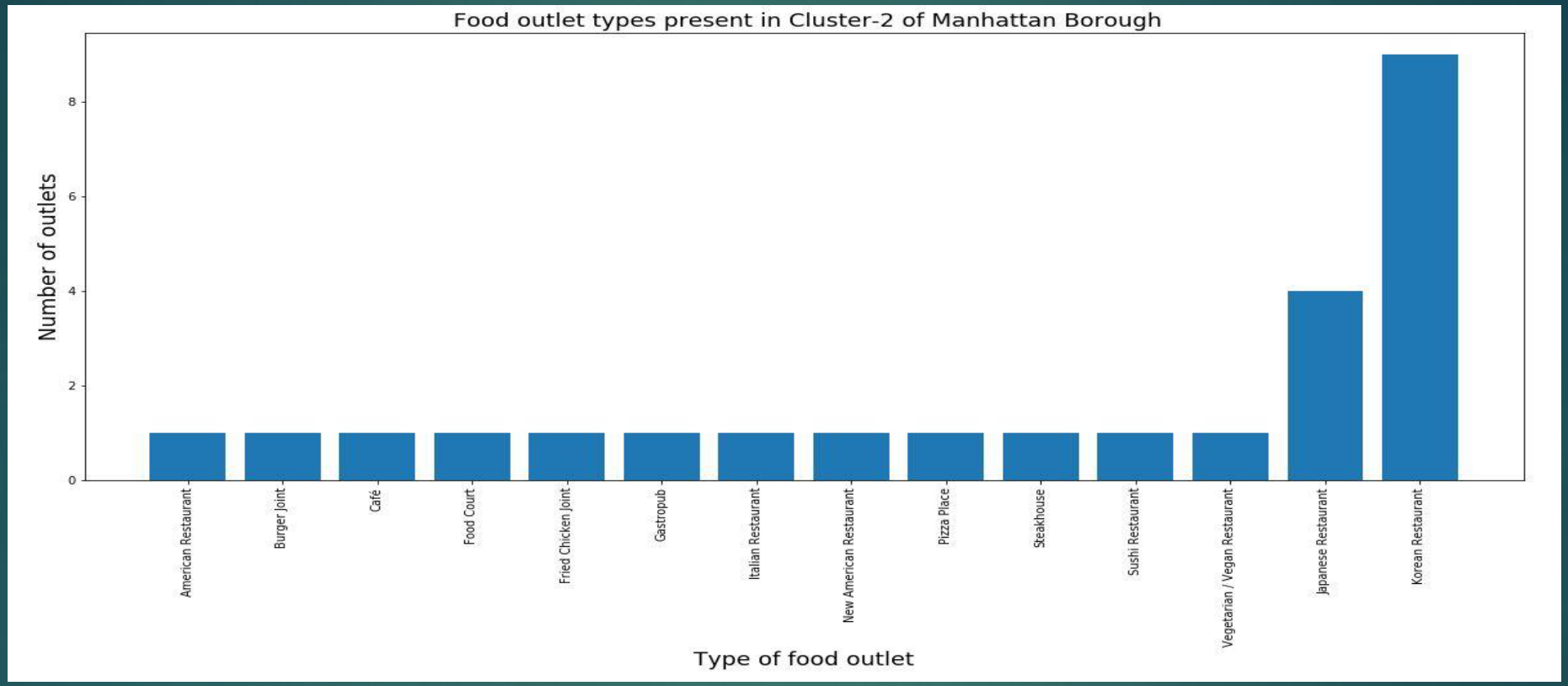
Map showing different clusters of neighborhood present in Manhattan borough

Clustering of Manhattan



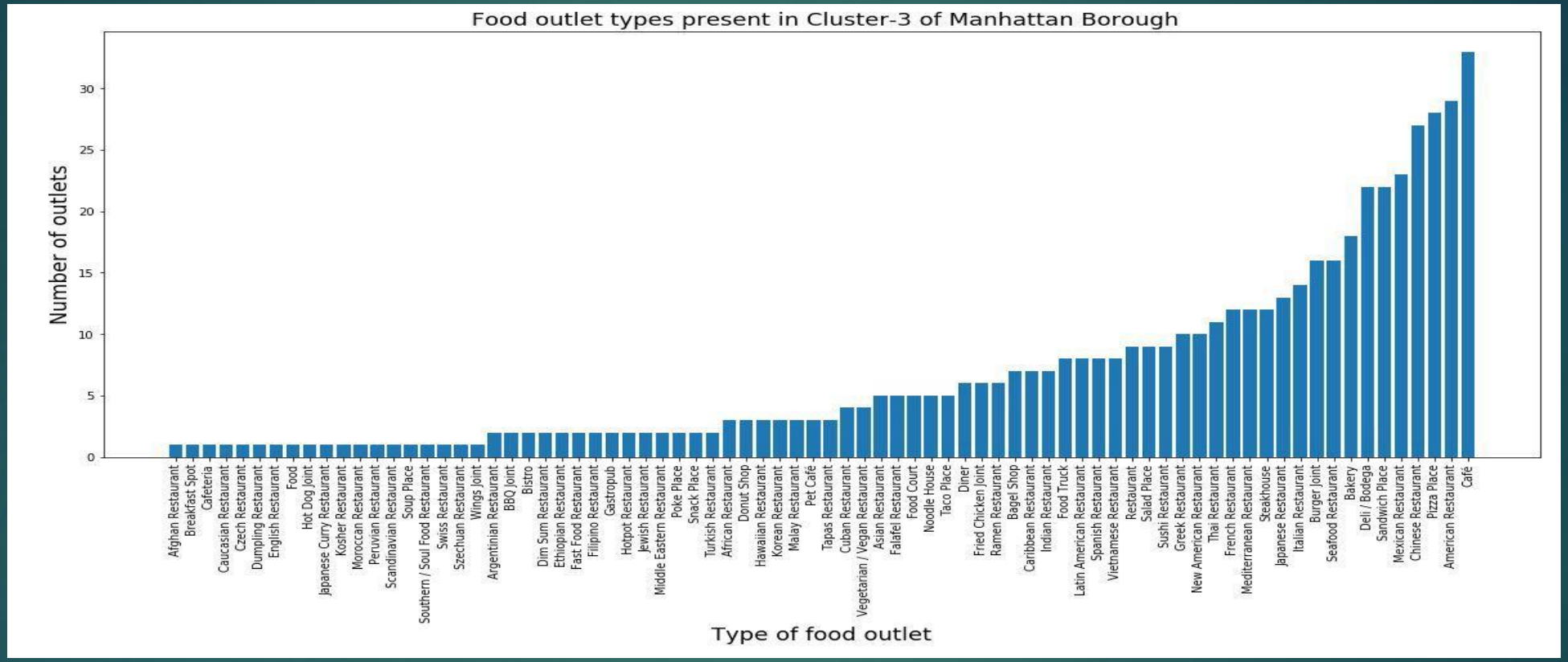
Plot showing Type of food outlets vs Number of outlets in Cluster-1 of Manhattan

Clustering of Manhattan



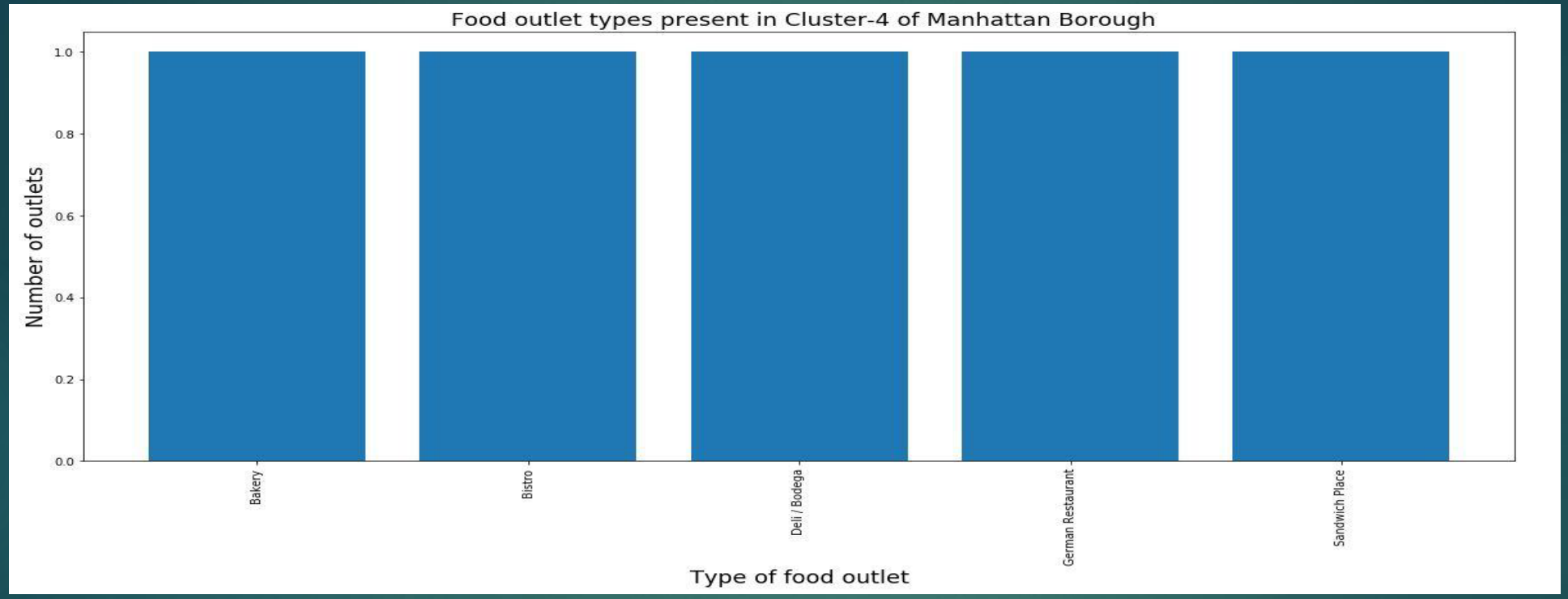
Plot showing Type of food outlets vs Number of outlets in Cluster-2 of Manhattan

Clustering of Manhattan



Plot showing Type of food outlets vs Number of outlets in Cluster-3 of Manhattan

Clustering of Manhattan



Plot showing Type of food outlets vs Number of outlets in Cluster-4 of Manhattan

Conclusions

- Neighborhoods of each borough in NY city has been partitioned into different clusters using K-means clustering algorithm
- Top 10 common food outlets of each neighborhood are used as a data point
- This analysis can be helpful for those who are planning to open a new food outlet in NY city
- It can help them in deciding the type and location of food outlet that provides them the competitive advantage.