The Battle of Neighborhoods of two cities

Coursera Capstone

IBM Data Science Professional Certificate

By: Rajeev Kavety

Problem

- The top 10 most common venues misses the niche market, such as, food courts.

- Identifying food courts can be useful for the entrepreneurs who want to setup a business.

- The challenges of overlaying the top most common venues having food courts on the maps of Toronto and Manhattan, can address both preferences of the public.

Data

- Used scrapping techniques demonstrated by [beautiful soup4](#), and using the python results package

- The neighborhood data was extracted from the [Wikipedia link](#).

- For Toronto map coordinates, the information in a CVS file was downloaded from the link http://cocl.us/Geospatial_data . For the Manhattan neighborhoods dataset, the link is in the json format. https://geo.nyu.edu/catalog/nyu_2451_34572.

Foursquare API

Foursquare API is a location data provider that uses RESTful API calls to retrieve data about most common venue categories in each neighborhood within a radius of 500m.

Methodology

Exploratory Data Analysis

- For Toronto city, Neighborhoods of North Toronto, Central Toronto, West Toronto, East Toronto, Downtown Toronto, North York, East York, York, and Etobicoke are taken.

- Another is Manhattan neighborhoods of New York.

- To focus on the scope of this project, food courts along with the venues was combined to the top 10 venues list.

Unsupervised Machine Learning

Using the K-means clustering algorithm- Elbow method, the cluster number is estimated
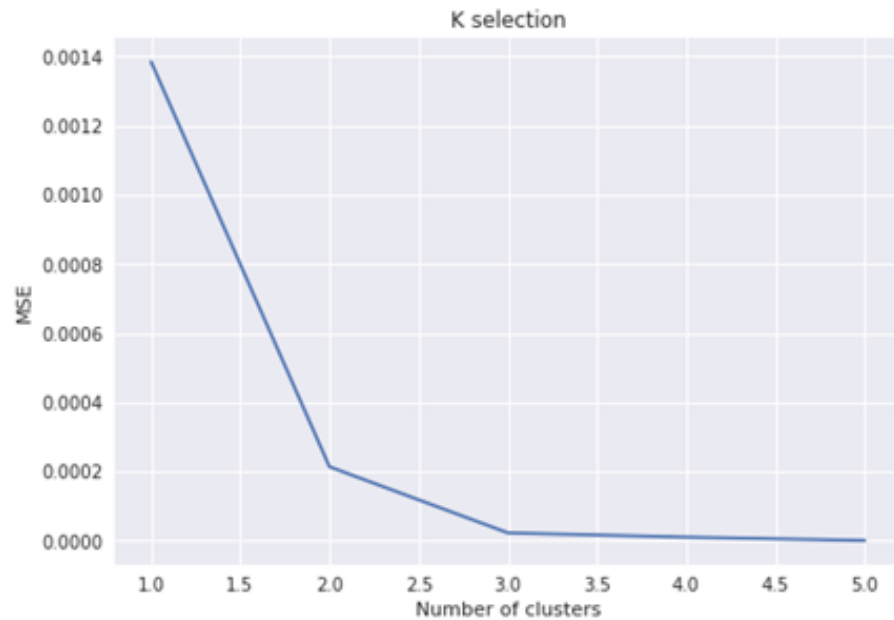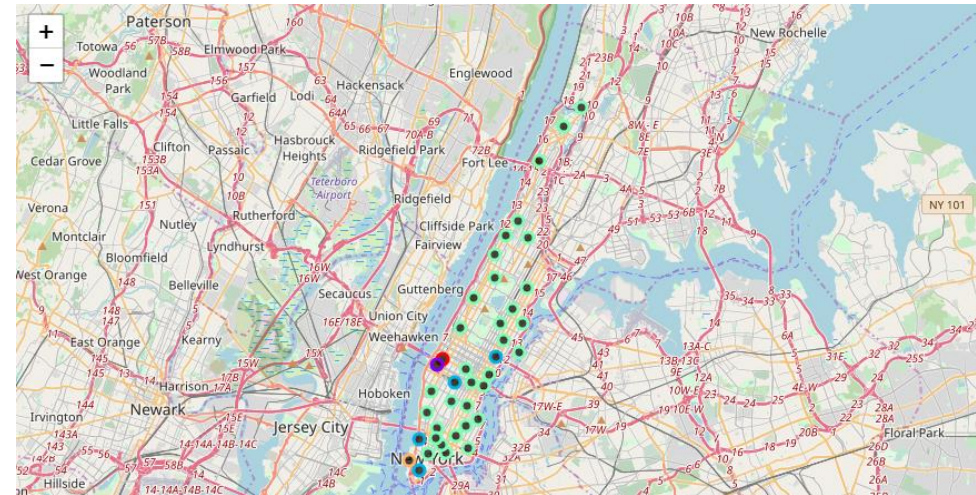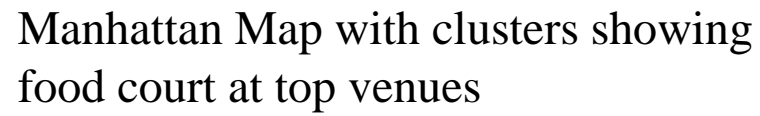

The number of Kcluster set: 5
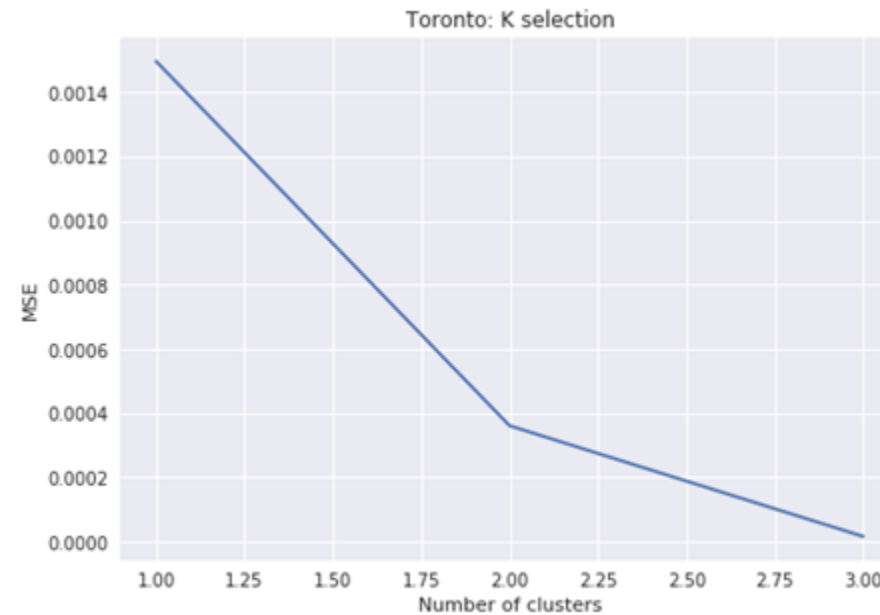
The number of Kclusters set: 3

Results

Toronto Map with clusters showing food court at top venues

Manhattan Map with clusters showing food court at top venues

Conclusions

- The precedence of food courts in some of the top 10 venues, gives an insight for the niche market of food courts.

- Visitors prefer to visit the location that has food courts as well as other top most venues.

- Overlaying food courts with top most venues gives information for the start-up entrepreneurs of a business opportunity.

- Including population density would have been helpful for in-depth insight to the entrepreneurs.

Thank you