Business Recommender

USING FOURSQUARE DATABASE AND MACHINE LEARNING

Shailesh Padekar | Coursera Capstone | 19/12/2018

Introduction

Before plunging into business it is always wise to test all the aspects of it in order to make it successful. Most sound methodology to do this is gather as much data as possible about the same from previous experiences / from same domains and analyze it.

Same is true when opening any food chain. Choice of the location is most important. Machine learning module in discussion tries to identify the most suitable location for opening food chain based on popularity of food chain types (e.g. Deli/ Pizza parlor etc) in locality.

However tools ,data, methodology used in this module is not limited to food chain but can be expanded to other categories as well.

Data

For this use case New York, Manhattan borough is selected. We will be collecting top 100 venues in food category for each neighborhood along with their categories/tags with help of Foursquare API calls. Columns of interest are as follows

- Neighborhood
- Neighborhood Latitude
- Neighborhood Longitude
- Venue
- Venue Category

Methodology

As mentioned before we wish to determine location which will be suitable for certain types of businesses based on the trending and booming categories in those localities. System should be able to suggest promising localities and suggestion can be convincing if trend can be visualized on the map.

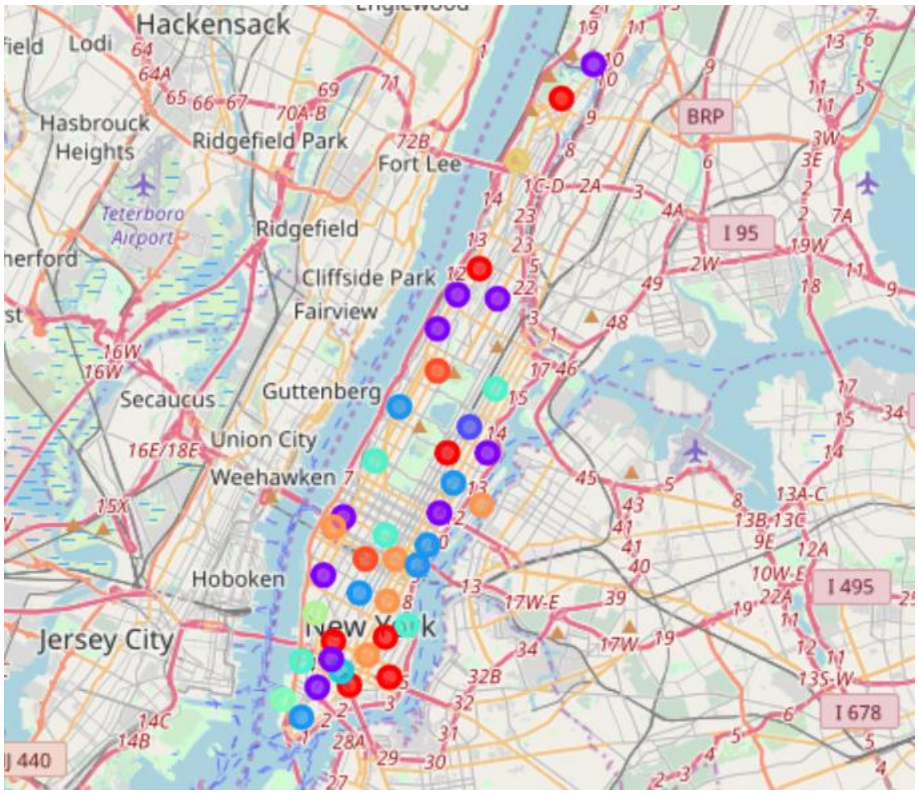
Since this is classification problem we will be utilizing K-means algorithm due to its robustness.

In order to make it work with K-means algorithm we need to build row for each neighborhood which contains sorted list of venues in order of its popularity. That can be easily achieved in python with help of pandas library and one hot encoding.

We have chosen 10 clusters as even though there are around 250 categories for food section most recurring categories where between 10-15

Results

We plotted all the clusters marked with different color on the Manhattan map.



Even though there is no clear trend based on observations we can find small clusters bunched together indicating similar choice of food habits in adjacent localities

e.g. Red circle cluster at bottom indicates Chinese/Japanese restaurants do well in those regions

Blue circle indicates restaurant goers prefer Italian / Pizza place/ Deli to other types.

Almost all clusters have Italian Restaurants listed in top 10

Conclusion

As demonstrated in module we can use foursquare data to provide recommendation of locality to establish any kind of business with good confidence. This module can be improved further by adding other characteristics of localities. For instance in food

industry if we also consider race/ethnicity/origin of birth of people in that locality results can be further improved as it is observed that those factors do drive choice of favorite food.