IBM Data Science Capstone

Restaurant Recommendation System in Pittsburgh

April 2020

**Introduction**:

Pittsburgh is one of the largest city in the state of Pennsylvania and is also one of the important food joints in the state. The city has diverse options for food with different types of restaurants offering different cuisines.

**Problem description**

A business tycoon is planning to open a restaurant in the Pittsburgh. With Pittsburgh having already lot of restaurants, he wants to analyze the locality and provide him the report which would give him suggestions about the location for profitable business

**Data Description**

To recommend the user about the location for opening a restaurant, we need data about the entire locality.

We need

* The Neighborhoods available in the city considered, say Pittsburgh here
* Geographical coordinates for each neighborhood and each venue
* Number of the venues available in each neighborhood
* Number of Restaurants available in each neighborhood
* Type of restaurants available in each neighborhood

**Data Collection**

Collecting Neighborhood data:

I eventually ended up in collecting list of Neighborhoods in Pittsburgh through Wikipedia. I manually created a csv file with the list of neighborhoods in it.

Later I collected latitudes and longitudes of the Neighborhoods in Pittsburgh programmatically.

Once the required data is available, I programmatically collected different venues and categorized them. Later I used K Means to create clusters and from there, I derived observations for the best place to open a restaurant.

Methodology:

Initially the list of neighborhoods are collected manually and then programmatically, I collected latitudes and longitudes.

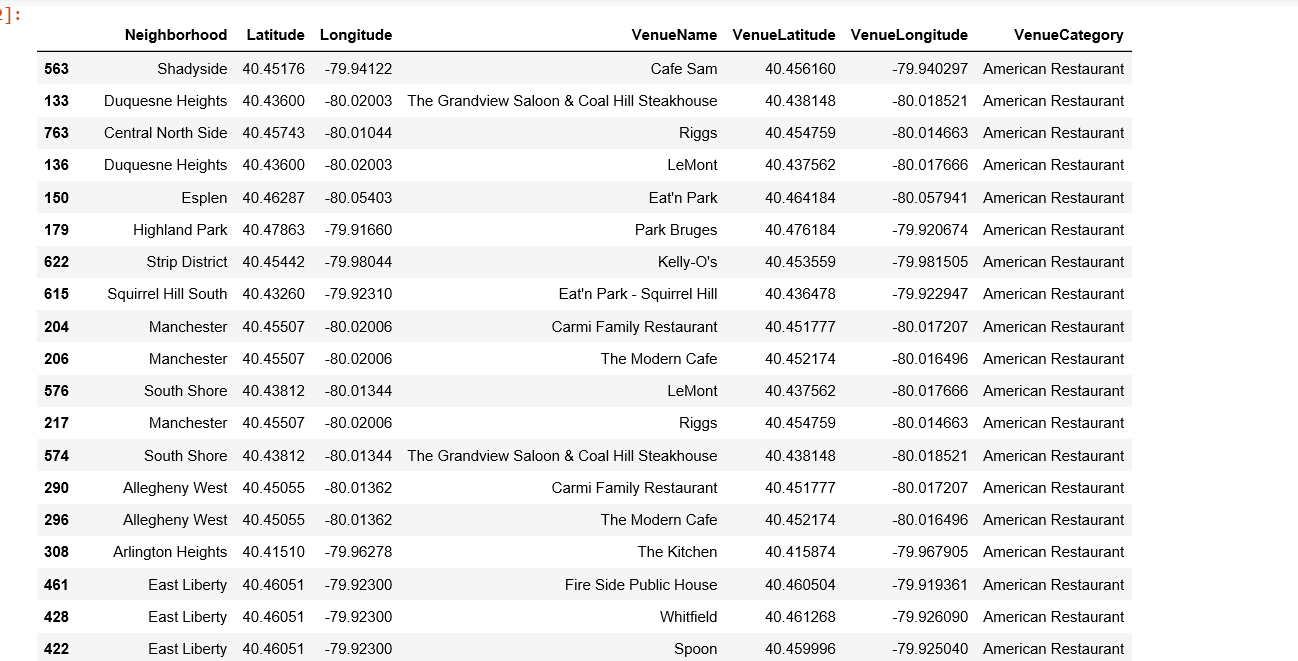
The cleaned data is shown as below



Once the geographical coordinates are available for the neighborhoods, I used Foursquare API to explore the venues in the neighborhood.

After exploring the venues, the data is cleaned up using the venue category which contains the word “Restaurant”

The data now contains top 100 restaurant’s within 500 m radius in each neighborhood



Later, I used K Means algorithm to analyze the data as the data is unsupervised data. The algorithms create clusters. The clusters are shown as below

Cluster 1



Cluster 2:



**Results:**

The below observations are made from the analysis

Pittsburgh has many American, Italian, Seafood and Thai restaurant. There are more restaurants in Downtown and in strip district area. There are only two seafood restaurants in East Hills. Any restaurant other seafood has good scope in East Hills

The best bet for opening restaurant is summer’s Hill where you have just only one American Restaurant

It is better to avoid areas in Cluster for opening any new restaurants as they already have many restaurant’s offering different type of cuisines

The locations in the cluster 4 has very few restaurants and they are of the same type - Thai.