**Introduction and Business Problem**

**Introduction**

NewYork City. A must visit place in USA. City full of adventure. Well great adventure brings great hunger. Many people who are either staying in NewYork or vising the city for the first time may find it tedious to choose a place to eat. So here I am to help finding out best restaurant in NeyYork City.

**Business Problem**

For this project, I am going to create a simple guide on where to eat based on Foursquare likes, restaurant category and geographic location data for restaurants in NewYork. I will then cluster these restaurants based on their similarities so that a user can easily determine what type of restaurants are best to eat at based on Foursquare user feedback.

Sounds delicious right !!

**Data Requirements and Methodology**

**Data Requirements**

For this capstone, I will be using FourSquare API to pull out following data on restaurants in NewYork:

1. Restaurant Name

2. Restaurant ID

3. Restaurant Location

4. Category of the Restaurant ( we'll be deciding this )

5. Total like count.

**Data Acquisition Approach**

To get aforementioned data we need follow below steps:

1. First get latitude and longitude of NewYork City using 'Geopy' package.

2. Using geographical co-ordinates , we can fetch all nearby venues. ( using FourSquare APIs of course )

**Methodology**

The thought process behind this is that likes are a proxy for quality. The more likes there are, the better the restaurant is. This might be incorrect but API call issues (how many I can use for free) holds me back from getting price / rating data. I will then bin this data into a quality categorical variables so we can cluster appropriately.

I am also going to create new categorical variables for the restaurants to better group them based on type of cuisine. This way you can look for good American food or now what type of food might be best to eat in NewYork if you are new to the area.

I will take the gathered data (see above in Data Acquisition Approach and Data Required sections) and will create a k-means clustering algorithm that groups restaurants into 4 clusters so that people looking to eat in NYC can easily see which restaurants are the best to eat at, what cuisine is available and where in the city they can look to eat.

**Results**

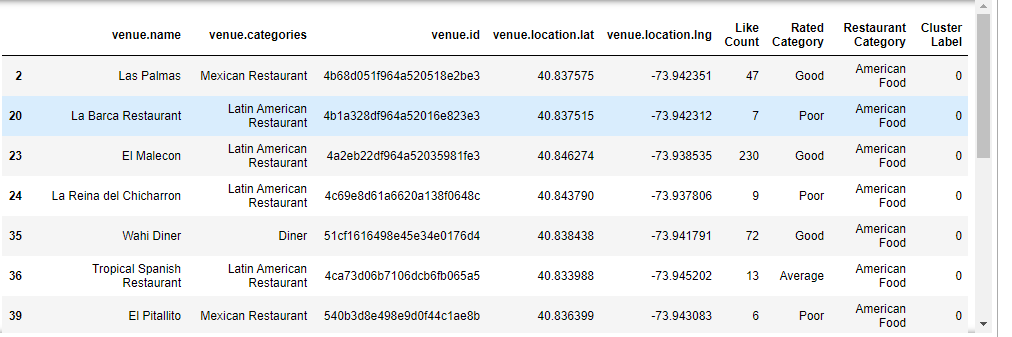
Running my clustering algorithm, I was able to generate four clusters of restaurants. These are as follows:

**Cluster 1 :**

Characteristics:

1. Mostly Eateries.

2. We can get Good food here.

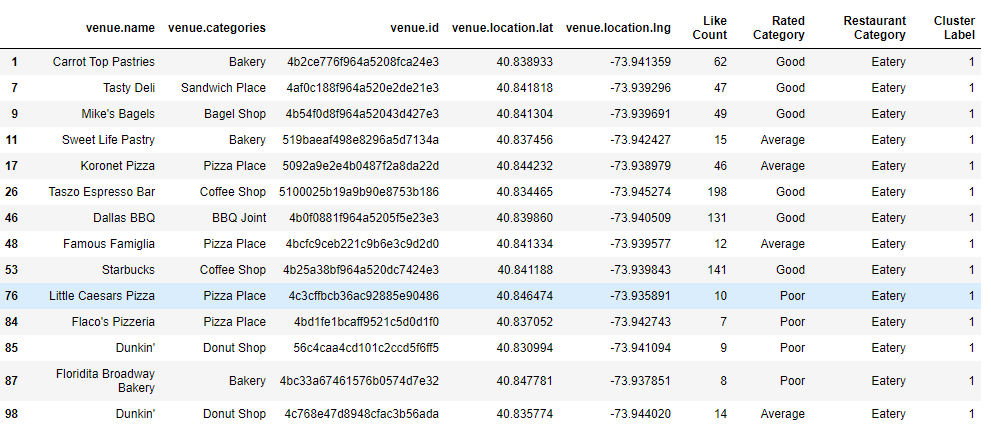


**Cluster 2**

Characteristics:

1. Mostly American

2. Food quality is below average

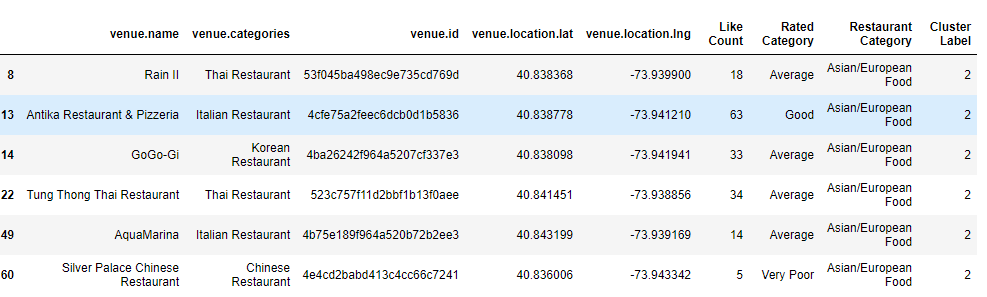


**Cluster 3**

Characteristics:

1. Mostly Asian/European Food.

2. Most of the restaurants have average food quality.

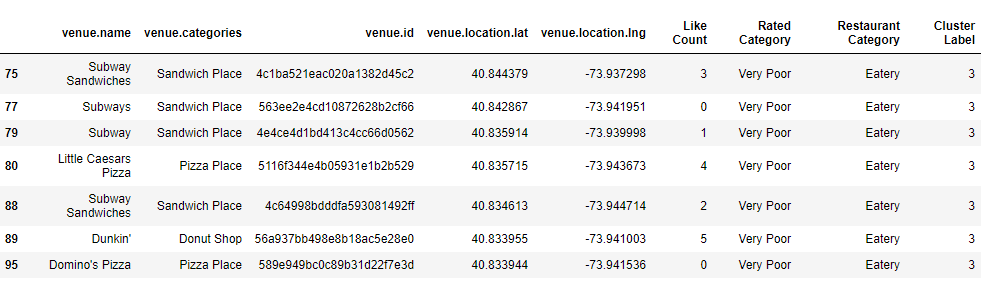


**Cluster 4**

Characteristics:

1. Mostly Eateries.

2. I won't recommend these places as they have poor ratings.



**Map of Clusters for Users**

