

Introduction and Business Problem

Introduction

The city of New York, NY is a big and multi-cultural city which is packed with restaurants of different cuisines. For people that are new to New York or are visiting it may find it hard to find good quality food.

Business Problem

For this project, I will create a simple guide on where to eat based on Foursquare likes, restaurant category and geographic location data for restaurants in New York. 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.

Data Requirements and Methodology

Data Requirements

For this project, I will be utilizing the Foursquare API to pull the following location data on restaurants in New York, NY:

- Venue Name
- Venue ID
- Venue Location
- Venue Category
- Count of Likes

Data Acquisition Approach

To acquire the data mentioned above, I will need to do the following:

- Get geolocator latitude and longitude coordinates for New York, NY
- Use Foursquare API to get a list of all venues in New York
 - Get venue name, venue ID, location, category, and likes

Methodology

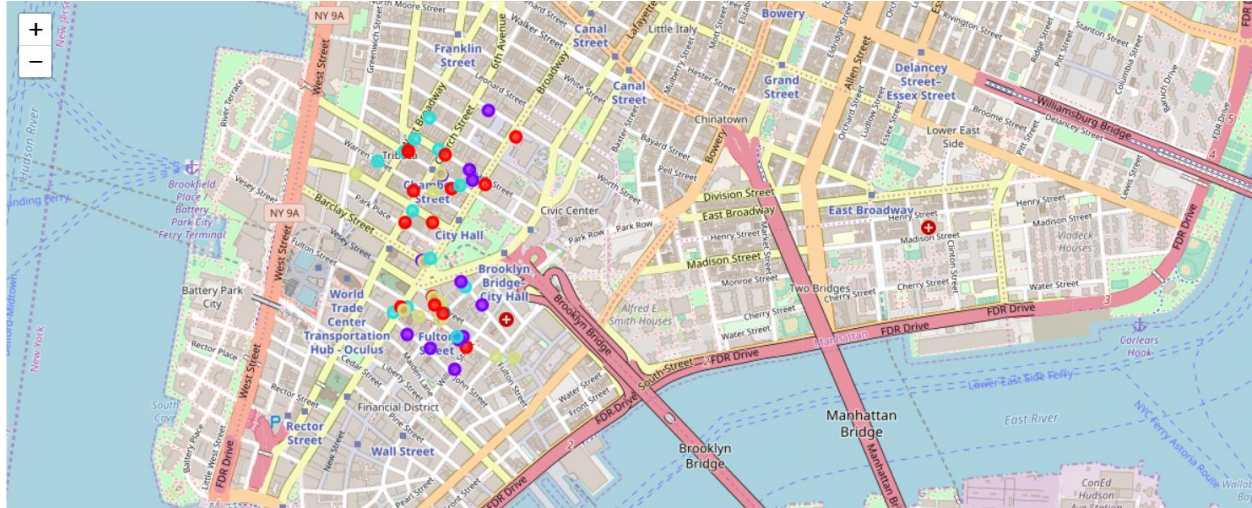
The thought process behind this is that likes are a proxy for quality. The more likes there are, the better the restaurant is. I will then bin this data into a quality categorical variables so we can cluster appropriately.

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 Mexican food or now what type of food might be best to eat in New York if you are new to the area.

Results

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

- Cluster #1: Poor Quality Restaurants
- Cluster #2: Average Quality Restaurants
- Cluster #3: Great Quality Restaurants
- Cluster #4: Below Average Quality Restaurants



Conclusion

There are a variety of restaurants available in New York City. I tried to cluster them according to their likings, but if we increase the cluster size we might able to get rankings according to different cuisines which will help more to our target audience.