**Introduction and Business Problem**

**Introduction**

As an international student, schooling at in Huntington, West Virginia has been fun and exciting but there was one problem. I had issues looking for the best restaurants around. It can be difficult finding restaurants that are worth going to and where they are since I am new in the environment.

**Business Problem**

For this project, we will be using Foursquare to create a simple guide on where to eat based on Foursquare categories, likes and geographic location data for the restaurants in Huntington. Using clusters, we will then group similar restaurants based on Foursquare user feedbacks so that students can easily determine which type of restaurant is best for them based on these criteria.

**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 Hoboken, NJ:

* 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 lat and long coordinates for Huntington, WV
* Use Foursquare API to get a list of all venues in Huntington
  + 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. 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 Mexican food or now what type of food might be best to eat in Huntington 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 4-5 clusters so that people looking to eat in Huntington can easily see which restaurants are the best to eat at, what cuisine is available and where in Huntington they can look to eat.

**Results**

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

**Cluster 0**

* Characteristics
  + Above Average



**Cluster 1**

* Characteristics
  + Poor quality food



**Cluster 2**

* Characteristics
  + High quality food

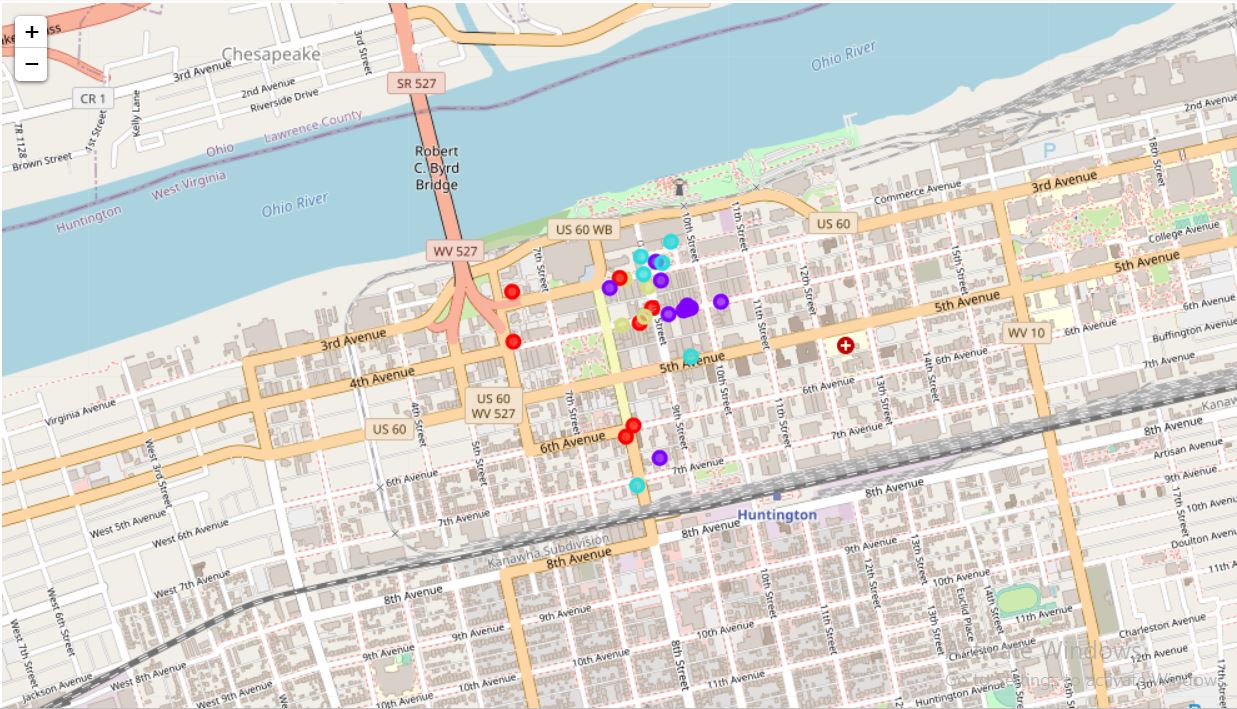


**Cluster 3**

* Characteristics
  + Above average quality food
  + Mostly Bars



**Map of Clusters for Users**



**Discussion**

From these results we see that euro asian indian food make up the most for the above average rating category of restaurants.

Bars make up most of the poor rating category of restaurants.

American food make up most of the great rating category of restaurants.

Bars also make up the most of the below average rating category of restaurants.

**Conclusion**

From the analysis of the data of restaurants in Huntington, I would recommend an American restaurants to a friend who is new in the area. There is a high chance he would enjoy the food.

Thank you.