**Introduction and Business Problem**

**Introduction**

The city of Hoboken, NJ is relatively small at ~1 square mile but it is packed with restaurants, night life and amazing people. For people that are new to Hoboken, despite its small geographic size, it can be daunting to figure out what restaurants are worth going to and where they are. For people that used to live in Hoboken or are visiting Hoboken, how do you know what the best places are to get something to eat?

**Business Problem**

For this project, I am going to put on my entrepreneur hat and create a simple guide on where to eat based on Foursquare likes, restaurant category and geographic location data for restaurants in Hoboken. 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 Requirement**

**​**

Using the Foursquare API to pull the following location data on restaurants in Hoboken, NJ:

* Venue Name
* Venue ID
* Venue Location
* Venue Category
* Rating
* Price
* Count of Likes
* Phrases

**Data Usage Approach – Clustering**

Using the data acquired from Foursquare create a k-means clustering algorithm that groups restaurants into 4-5 clusters so that people looking to eat in Hoboken can easily identify restaurants will cater to what their needs

**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:

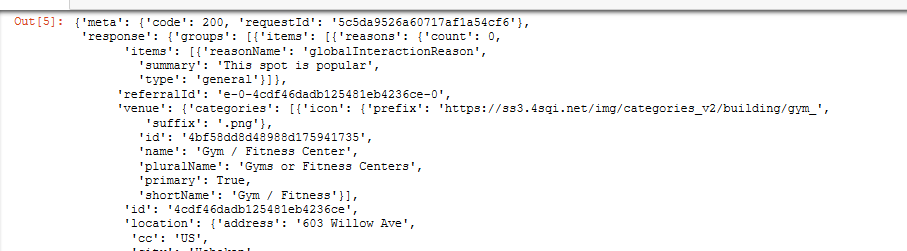
* Venue Name
* Venue ID
* Venue Location
* Venue Category
* Count of Likes

**Data Acquisition and preparation approach**

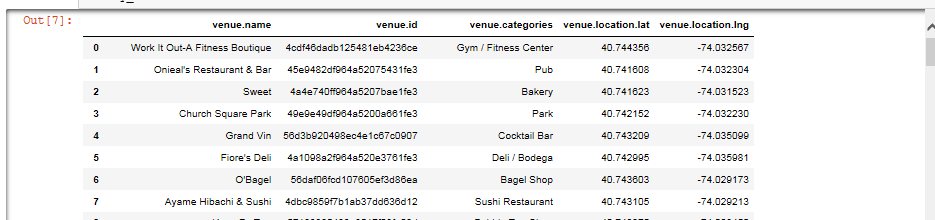
**Data Acquisition Approach**

To acquire the data mentioned above, will follow the below steps

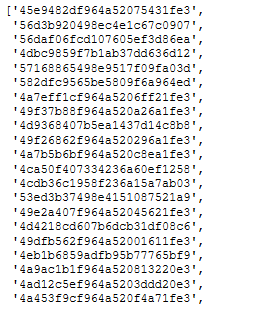
* Get geolocator latitude and long coordinates for Hoboken, NJ
* Use Foursquare API to get a list of all venues in Hoboken
* Creating a URL for all the venues in Hoboken
* Pulling the JSON for the URL of venues.



* Pulling the data from Foursquare into a data frame so we can manipulate and use it.



* Now let's get a list of venue ids so we can pull likes and add to our data frame.



* Get venue name, venue ID, location, category, and likes

**Data preparation approach**

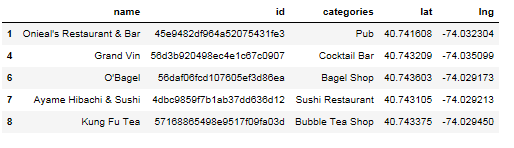
Once the data is acquired from foursquare, will perform the data preparation

The thought process behind this is 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.

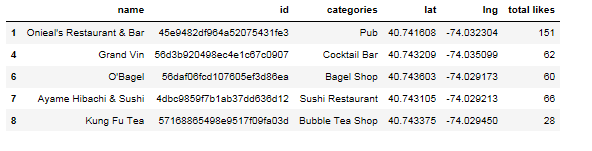
I am also going to create new categorical variables for the restaurants to better group them based on type of cuisine. This way you can look for good Mexican food or now what type of food might be best to eat in Hoboken if you are new to the area

Will perform below steps as part of data preparation

* Prepping our data for clustering. This will include combining data from different lists, creating new categorical data to be used, binning data and then encoding the data for clustering.

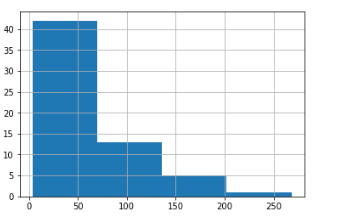


* Combine our list of likes into our data frame



* Let's look at our like data to set bins and based on number of likes set them to appropriate values (see below like histogram)

*Visualize our total likes based on a histogram*



categorize our restaurants based on likes and grouping based on type of restaurants



* Create dummy variables for total likes and categories so we can cluster

**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 Hoboken if you are new to the area.

**Algorithm**

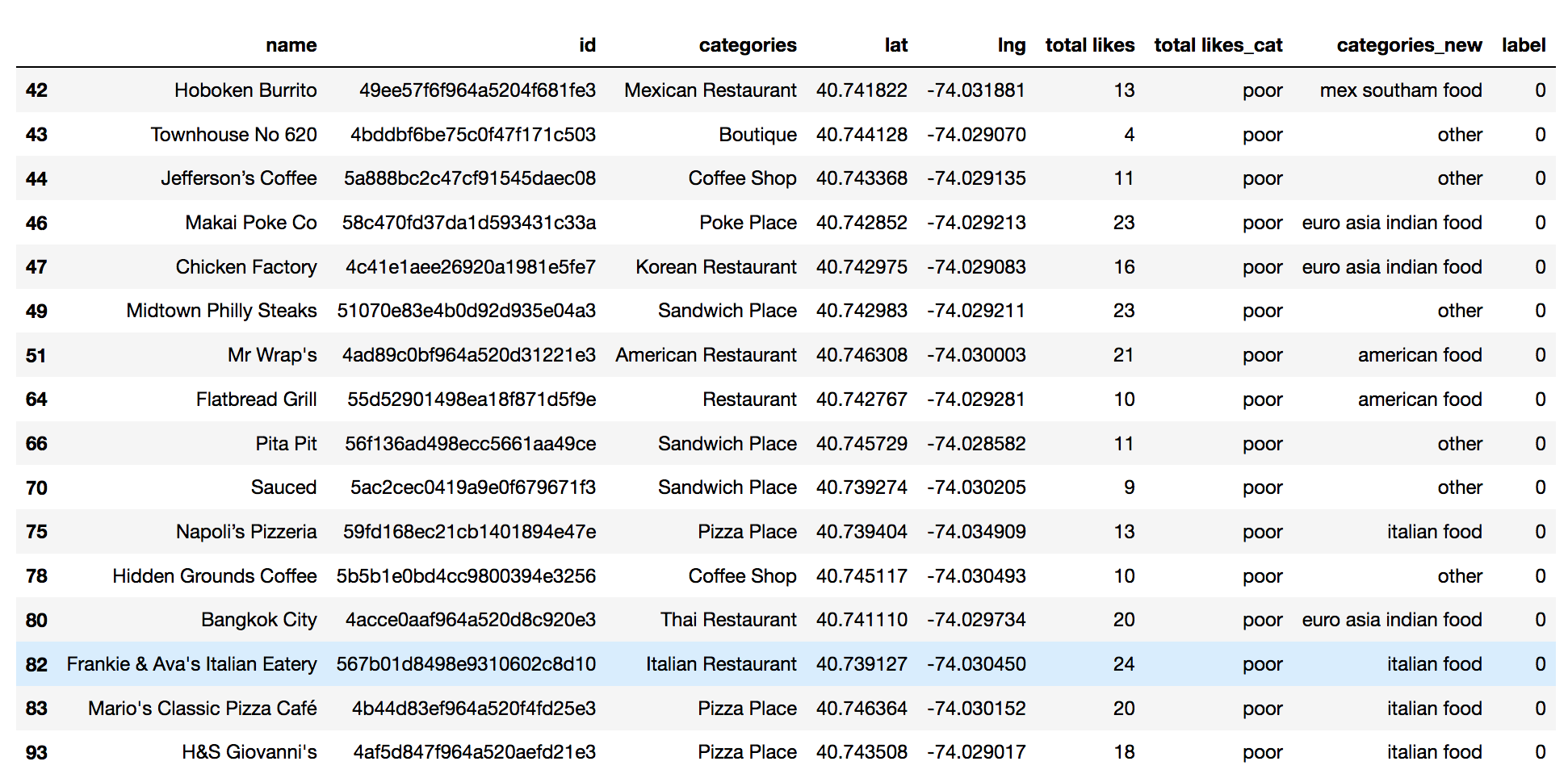
I will take the gathered data (see above in Data Acquisition Approach and Data Required sections) and will create a k-means clustering algorithm that groups restaurants into 4-5 clusters so that people looking to eat in Hoboken can easily see which restaurants are the best to eat at, what cuisine is available and where in Hoboken they can look to eat.

**Results**

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

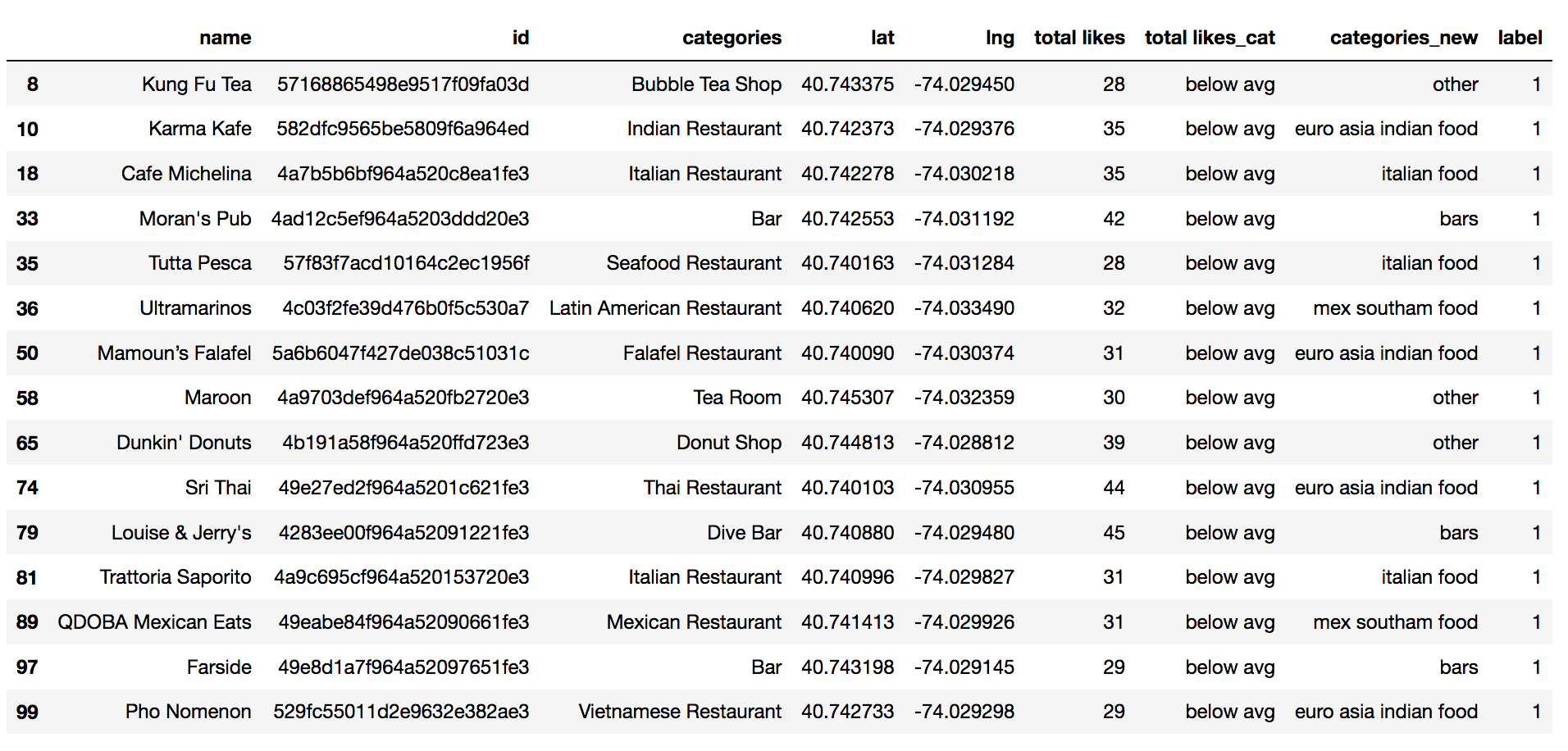
**Cluster 1**

* Characteristics
  + Poor quality food
  + Mostly Italian food or other



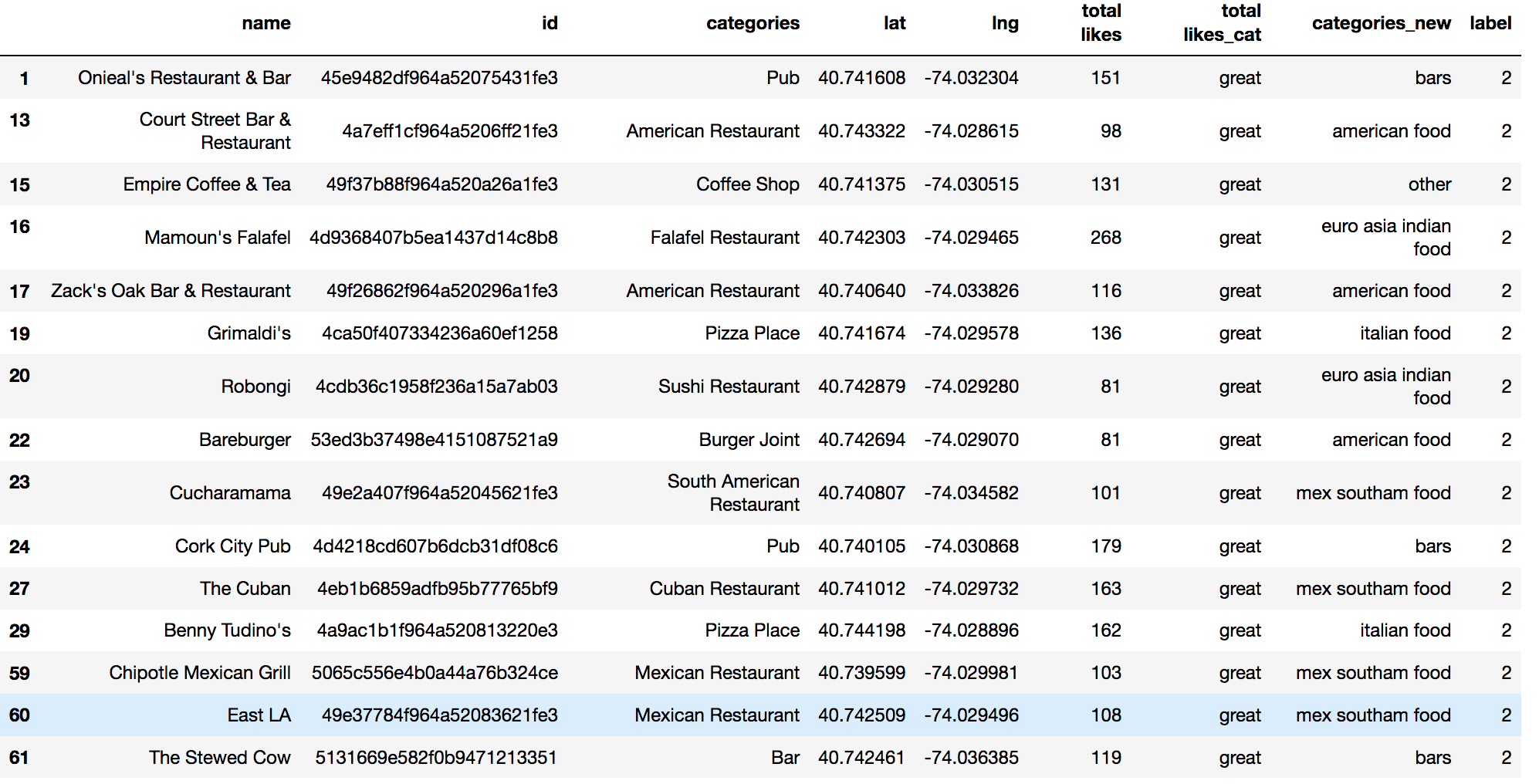
**Cluster 2**

* Characteristics
  + Below average quality food
  + Mostly Europe / Asia inspired food



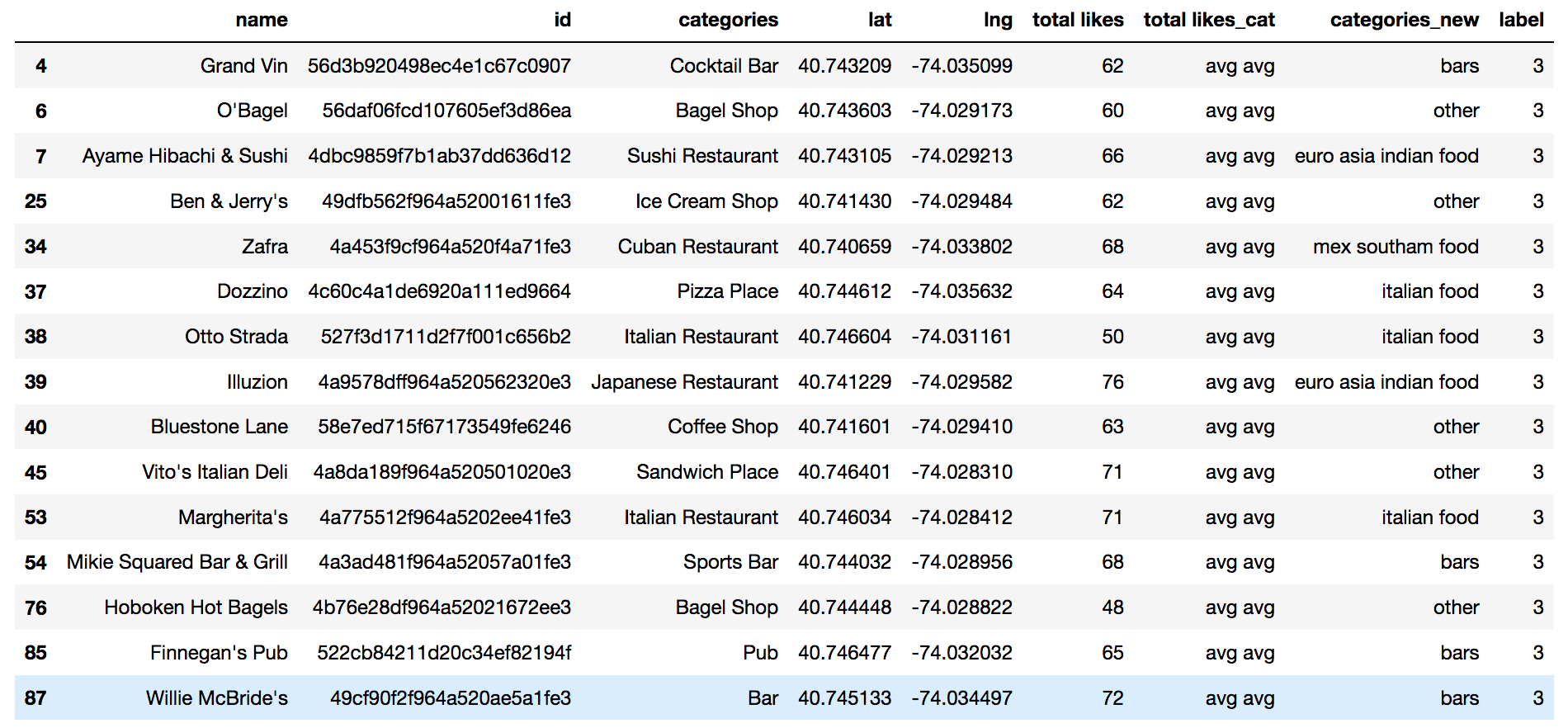
**Cluster 3**

* Characteristics
  + High quality food
  + Mostly Mexican and South American inspired food

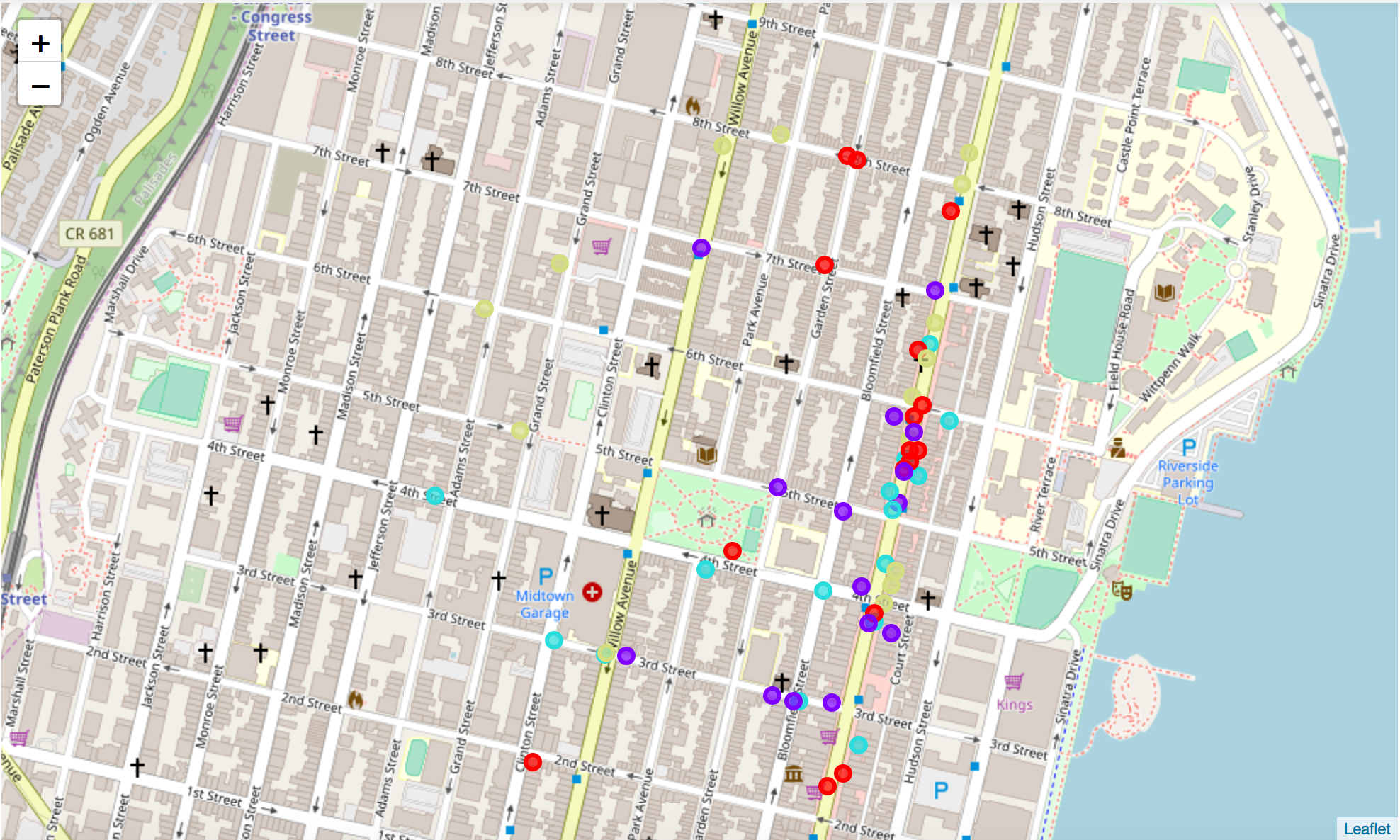


**Cluster 4**

* Characteristics
  + Above average quality food
  + Mostly Bars



**Map of Clusters for Users**



**Conclusion**

I am able to cluster the restaurants into four based on the characteristics, so that people looking to eat in Hoboken can easily see which restaurants are the best to eat at, what cuisine is available and where in Hoboken they can look to eat