**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 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 Hoboken, NJ
* Use Foursquare API to get a list of all venues in Hoboken
  + 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 Hoboken if you are new to the area.

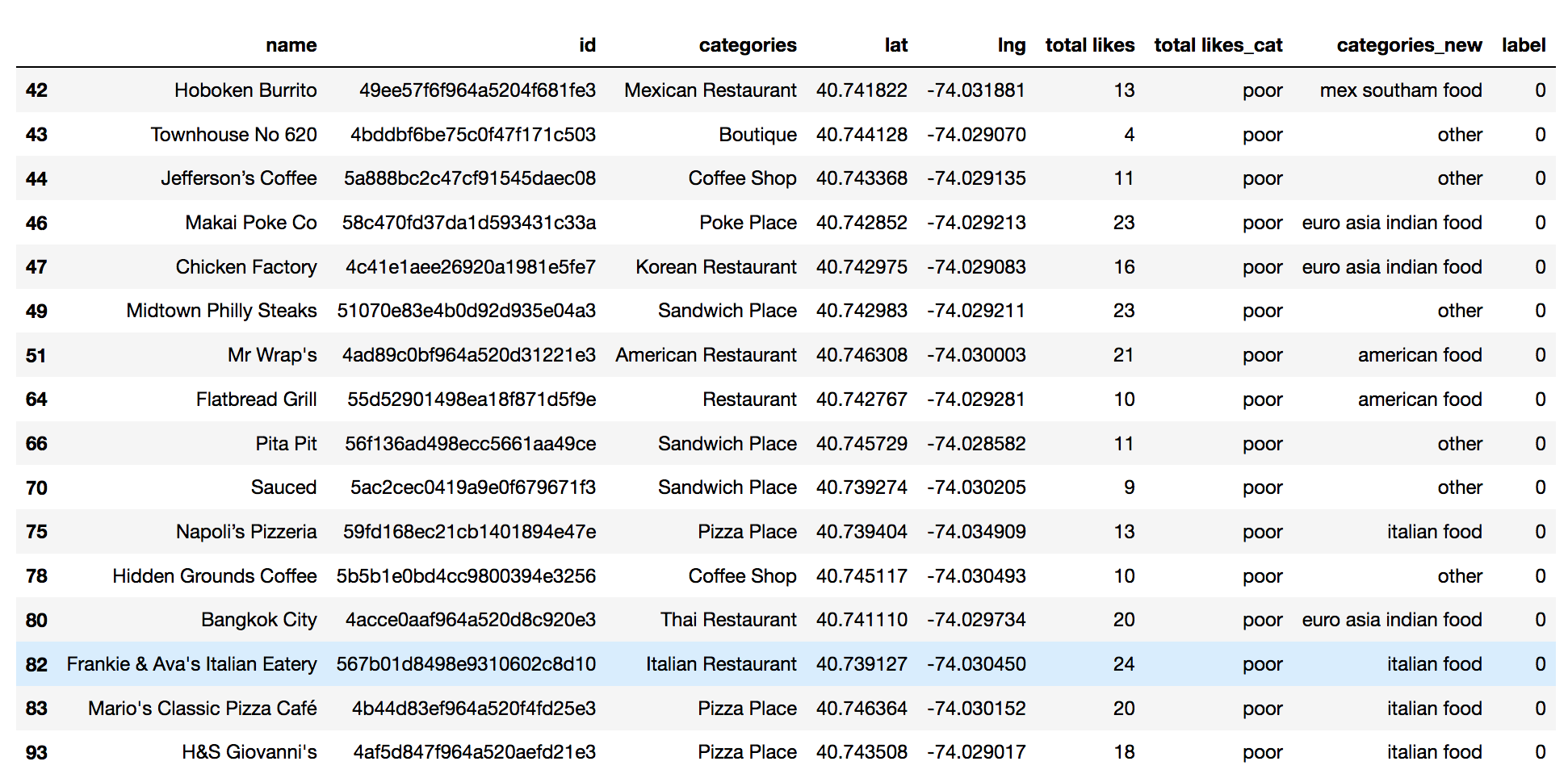
I will take the gathered data (see above in Data Acquisition Approach and Data Required sections) and will create a k-means clustering algorithm that groups restaurants into 4-5 clusters so that people looking to eat in Hoboken can easily see which restaurants are the best to eat at, what cuisine is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

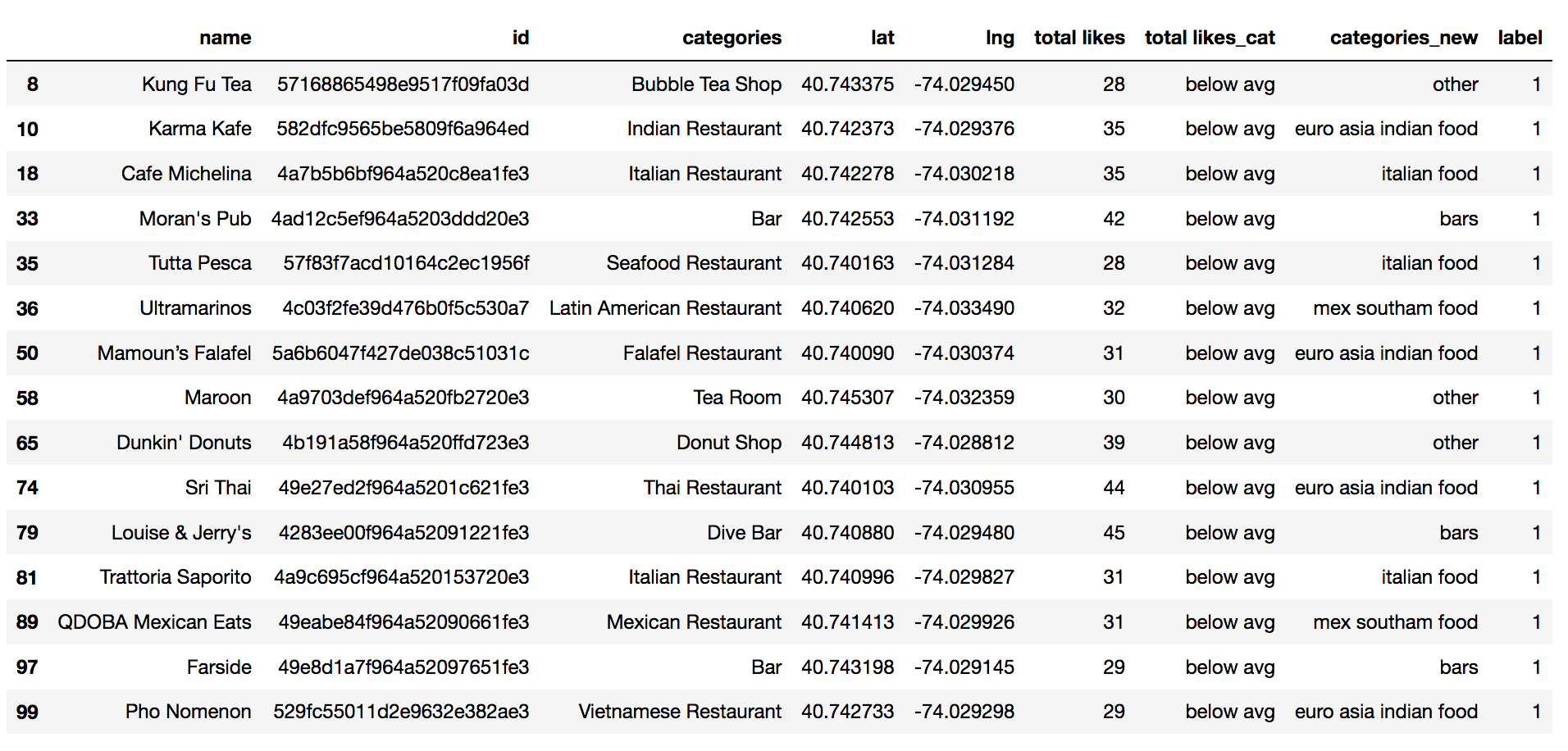
**Cluster 1**

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



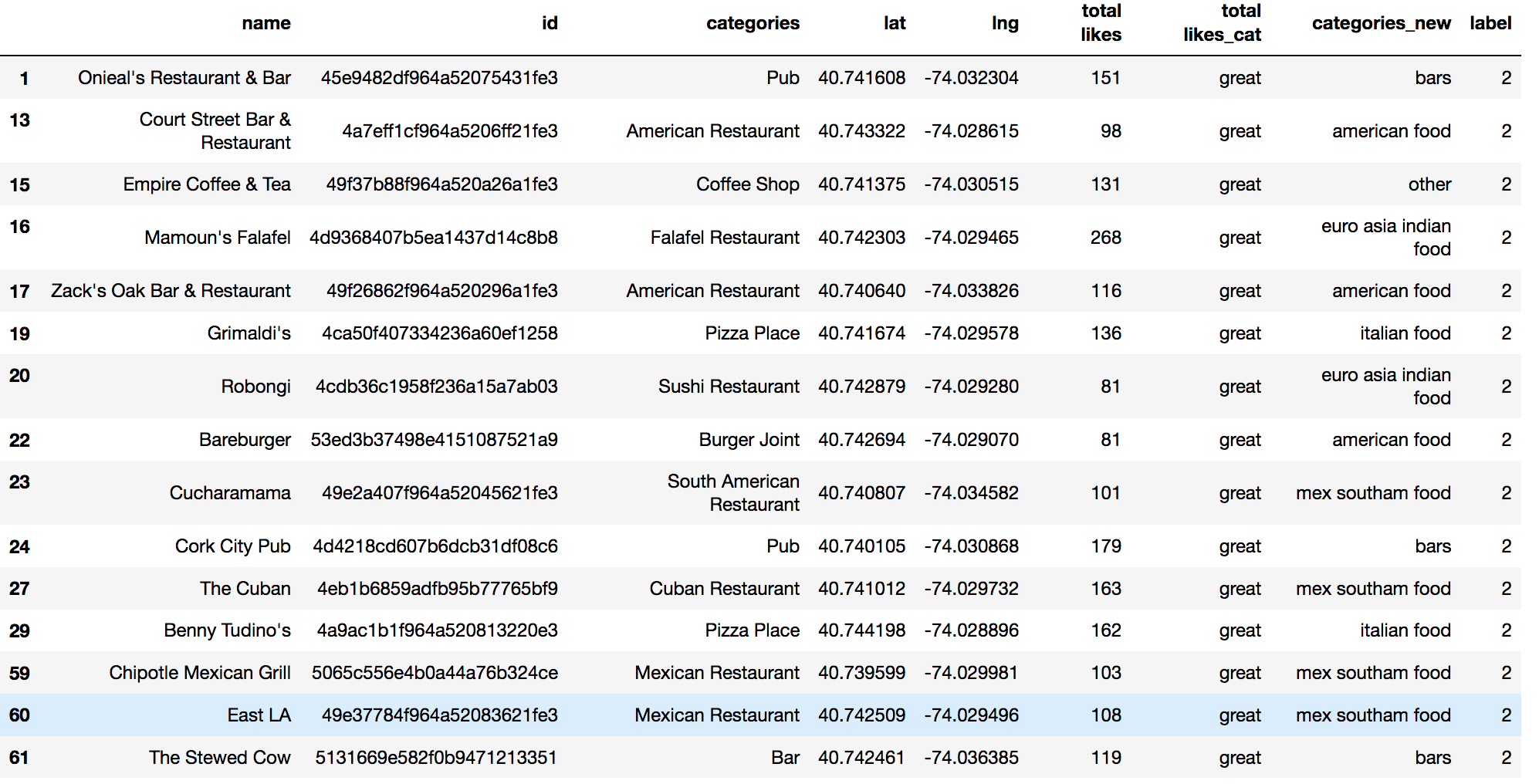
**Cluster 2**

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



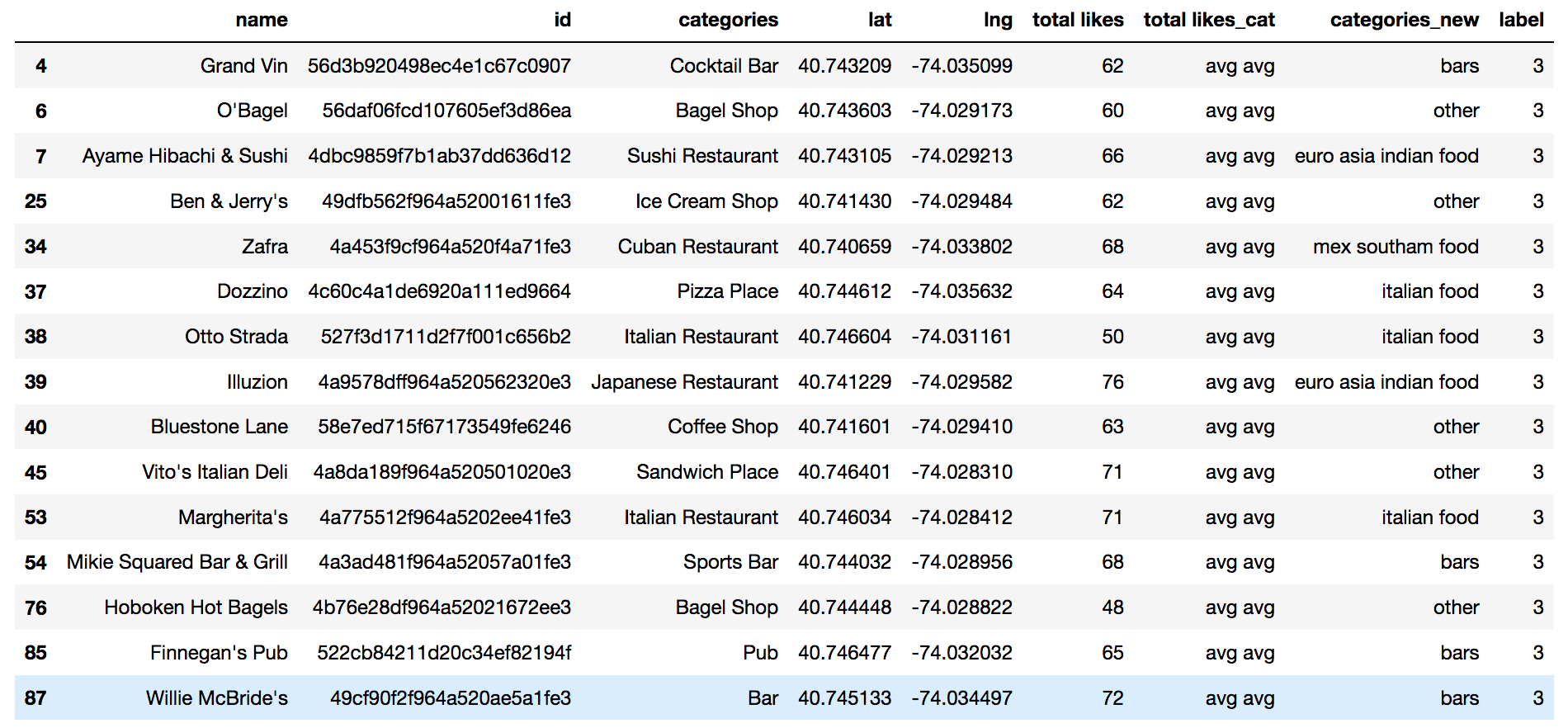
**Cluster 3**

* 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**

