**New Restaurant Recommendations near Edison, NJ (USA)**

Anuj Sud

Jan 2, 2021

1. **Introduction**
   1. Background

Edison is a township located in Middlesex county, New Jersey, United states and is named after Thomas Edison for inventing first commercial light bulb at a lab situated in this city. Edison has an estimated population of ~100,000 as of 2019, with a population density of 3,339/sq mi. It’s a densely populated suburb with many neighboring cities like Woodbridge, Islin, New Burnswick, North Plainfield, Piscataway etc, given a well-founded schooling system and high school ratings from K-12 grades. According to U.S. News & World Report in 2016, J.P. Stevens ranked 41st within New Jersey and 905th nationally, while Edison H.S. ranked 59th and 2,015th.

Edison is primarily a middle-class community with more than 75 ethnic communities represented. Edison has a large Jewish community next to Highland Park, with multiple synagogues located in Edison. Edison also has a growing Indian community and a number of temples serving the religious needs of the community. Reflecting the number of Edison's residents from India and China, the township has sister city arrangements with Shijiazhuang, China and Baroda, India.

* 1. Problem

Given a vast diversified population around Edison township and over 500+ food and restaurant businesses set up within a radius of 3-5 miles of this city, this project aims at gathering enough data to provide recommendations on type of food/restaurant business that can be newly established in this area and most viable location or city that will suit the business growth needs.

* 1. Audience and Stake holders

Primary audience for project report are entrepreneurs and local business personnel to utilize this information in their business planning. This report can also be used by local muncipal officials to guide business development and revenue generation opportunities.

1. **Data acquisition and cleaning**
   1. Data sources

Idea is to collect a comprehensive set of restaurant / food business data points around Edison, NJ with a radius coverage of 3 miles. This radius coverage will ensure all of the restaurants are covered and can be included for clustering process. Some of the key data points that will be used to derive relationships will be restaurant rating, price, customer review counts, location (Lat, long), and category.

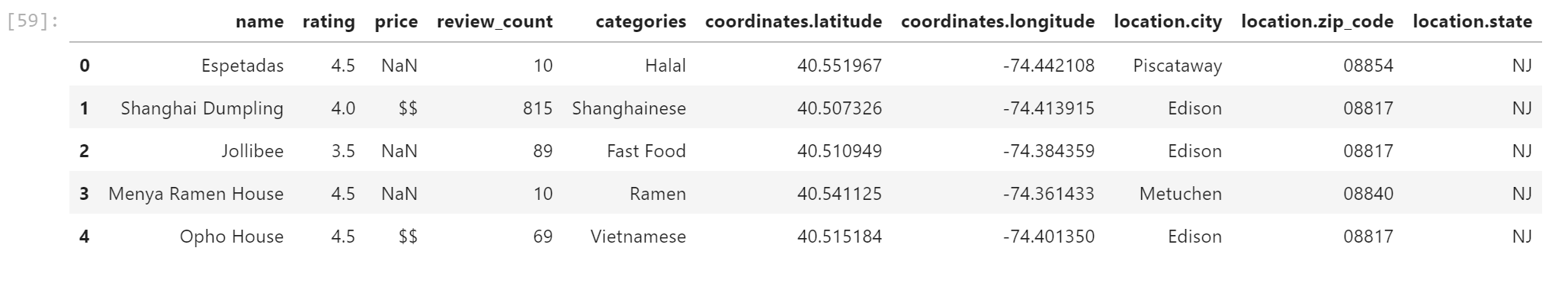
Using FourSquare’s “Explore” API end point, it only provided 100 restaurants around Edison area which didn’t seem to suffice the need of this projects data analysis. In order to obtain more reliable data source for this project, I did some google search and came **across Yelp’s Fusion APIs**. One among them was “[Business Search](https://www.yelp.com/developers/documentation/v3/business_search)” (Hyper liked here), which provided a total of 474 food / restaurant businesses around Edison, NJ location with required data points such as ratings, prices, review counts, locations, city, business category etc.

* 1. Data cleansing

I utilized “Business Search” API from Yelp Fusion platform to get information of 474 restaurants around Edison, NJ. Two data elements that I had to clean and manipulate were: Price and Category. Price provides the pricing of the business in terms of dollar symbols such as $, $$, $$$. I converted these symbols to a numerical value using this mapping: $ =1, $$=2, $$$=3 to represent the price effectiveness of a business.

Another key data element for my analysis was **Category**. I defined a custom function to extract the category value from a string of Json response to populate the desired category name in the dataframe. This categorical column was converted to numerical value for standardization and clustering purposes.

View of consolidated data points from Yelp’s API.



Yelp’s Business Search API had a max result set of 50 rows and had to be run multiple times with offsets to extract all of the available restaurant businesses around the targeted location. Below is the basic structure of calling Yelp’s business search API.

URL = "https://api.yelp.com/v3/businesses/search"

API\_KEY="xyz”

HEADERS = {'Authorization' : 'bearer %s' % API\_KEY}

PARAMETERS = {

"term":"Restaurants",

"latitude":"40.5382375",

"longitude":"-74.39451734508772",

"radius":"5000",

"limit":"50",

"offset" : "100"

}

response = requests.get(url = URL, params=PARAMETERS, headers=HEADERS)

biz\_data=response.json()

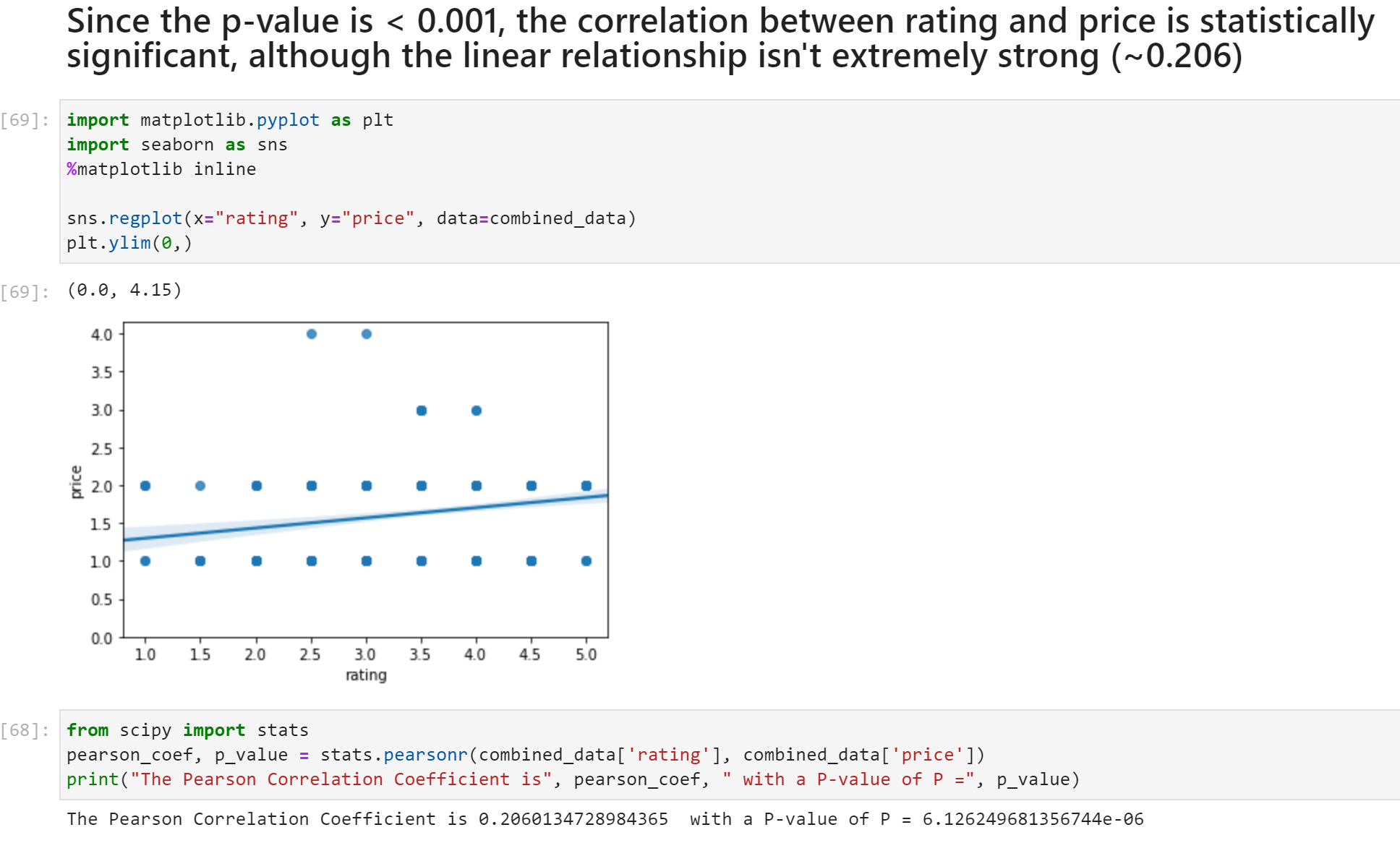
Consolidated result set from calling this API with increments of 50 (offset by 50), along with data cleansing of Price and Category columns looks like this:



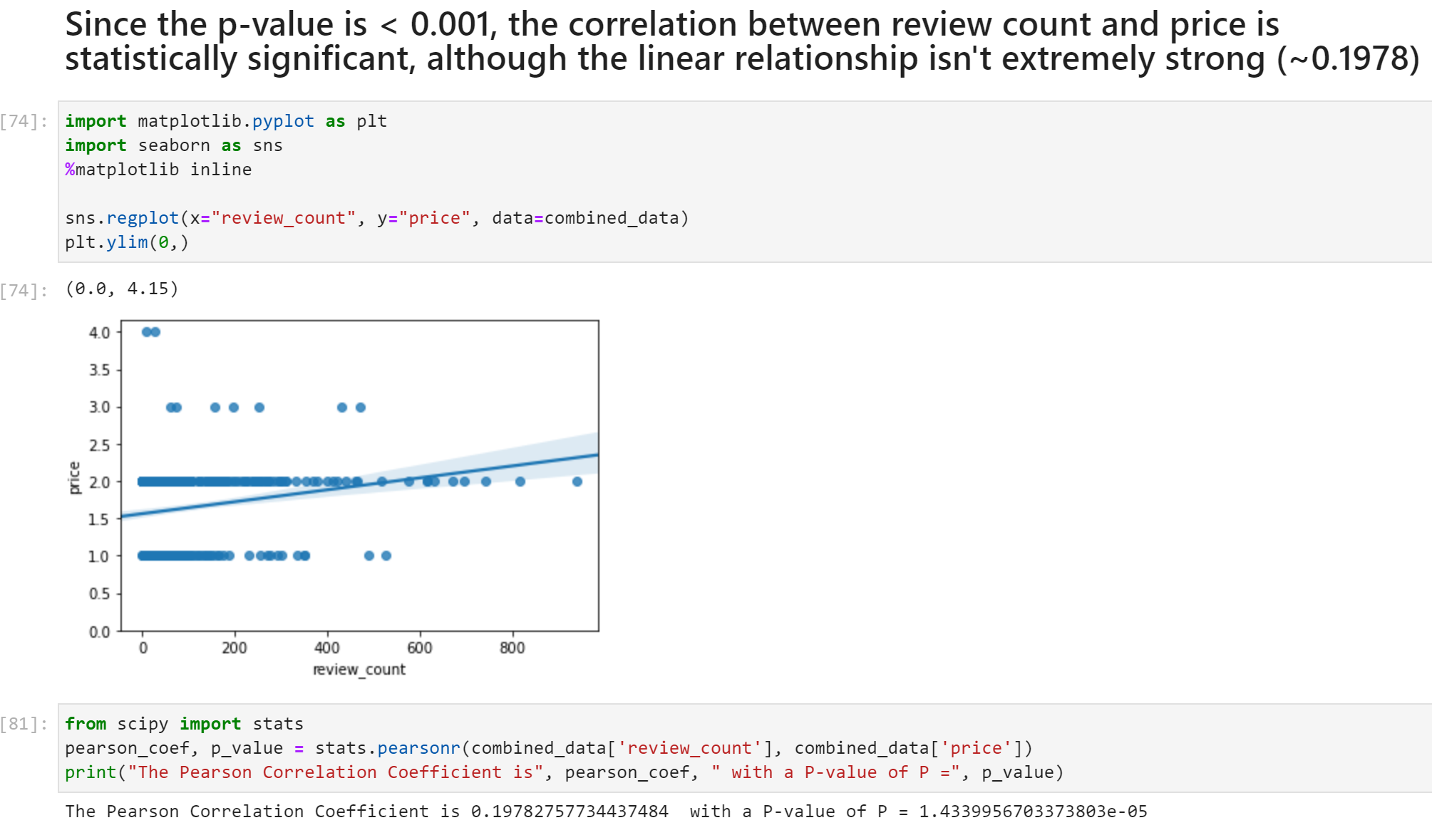
1. **Exploratory Data Analysis**

Looking at the key data elements like Rating, price and review counts, a significant statistical inference was found, even though there wasn’t a strong liner relationship between these variables.

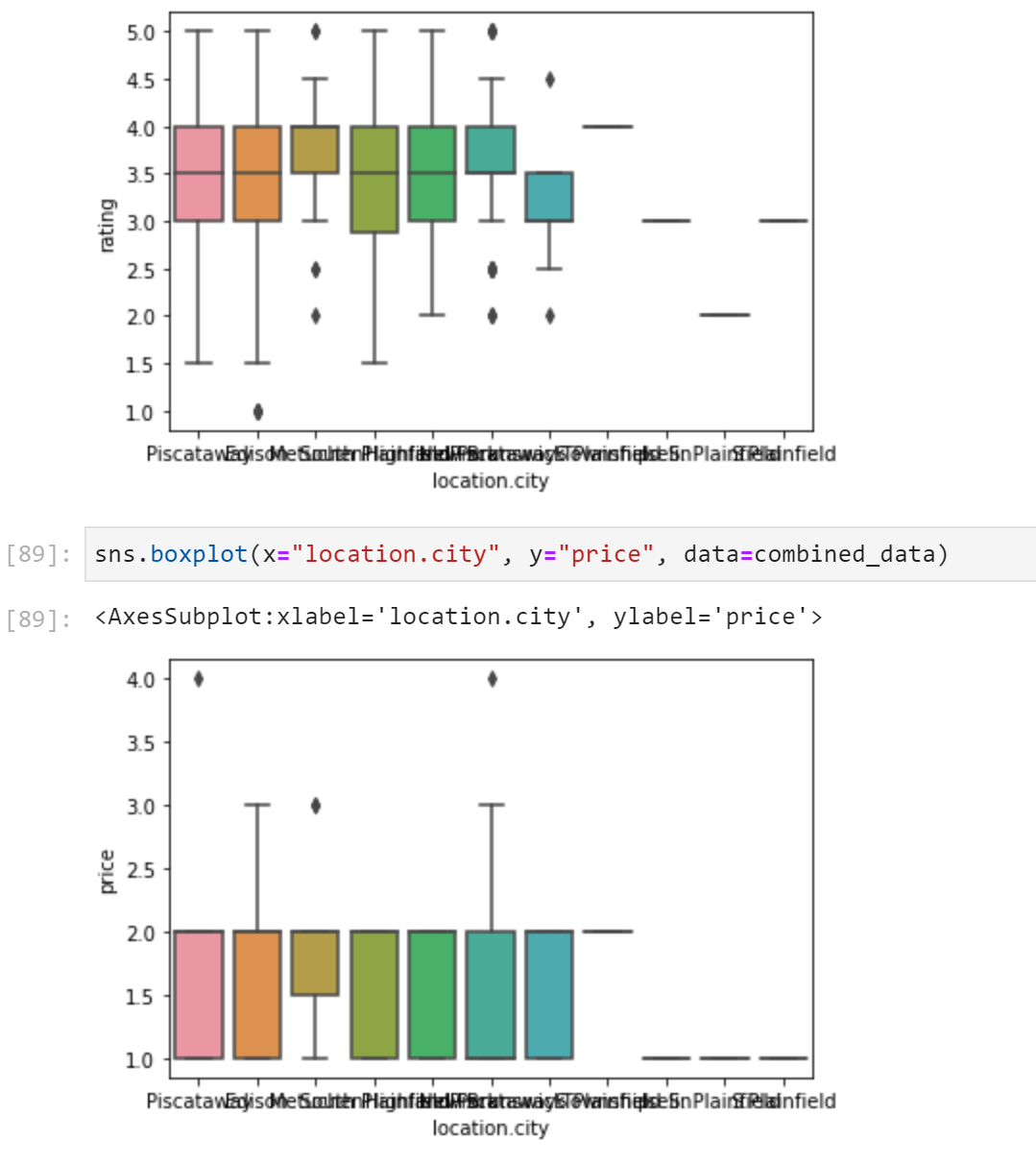
P-value is extremely low, pointing a strong statistical dependence between “Rating” and “Price” values collected from Yelp’s Business Search API.



Similarly, P-value is extremely low, pointing a strong statistical dependence between “review\_count” and “Price” values collected from Yelp’s Business Search API.



Location city variable has too much of overlap with rating and price, hence will not be considered for evaluation.



1. **Predictive Modeling and Clustering**

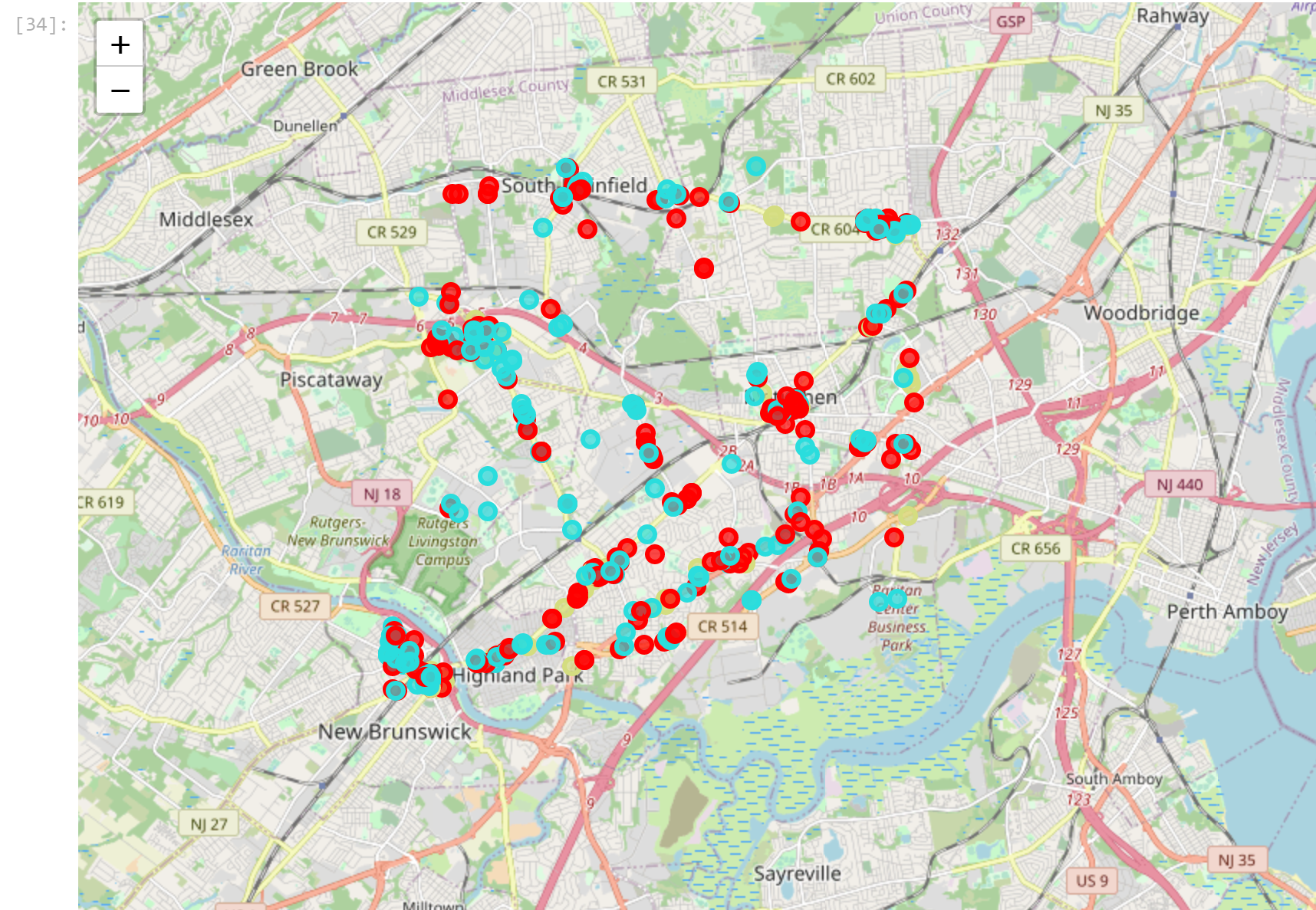
Kmeans clustering model was used to optimally determine best desired clusters and then map them out on the map using Folium Maps. The data was categorized in 4 clusters and the generated labels were mapped back to the original data set for further analysis using map plotting and grouping techniques.

Grouping by 4 (label) clusters and presenting mean values within these clusters for some of the key data points:



**Cluster = 3 comes out to be the most prominent one based on mean values of rating, price, review\_count and categories\_int.**

Mapping of 4 clusters from Kmeans with cluster 0 = 232 businesses, cluster 1 = 180, cluster 2 = 0 business and cluster 3 = 61 businesses



1. **Results**

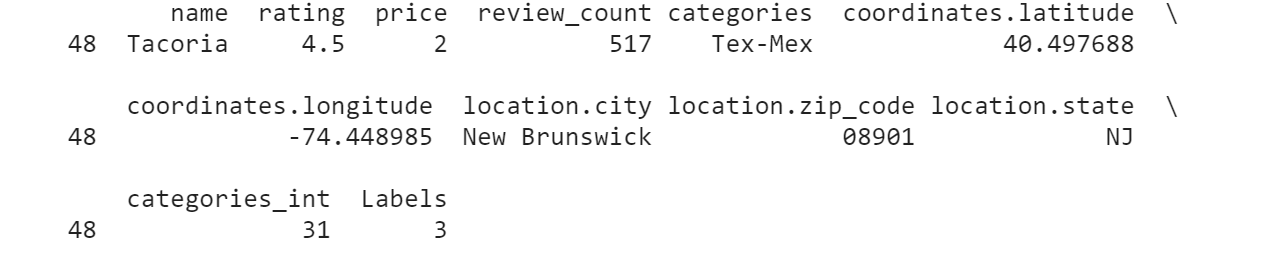
Based on grouping analysis of Cluster 3, it was found that American and Indian restaurants heavily dominated the areas around Edison, NJ.

American restaurant count = 4

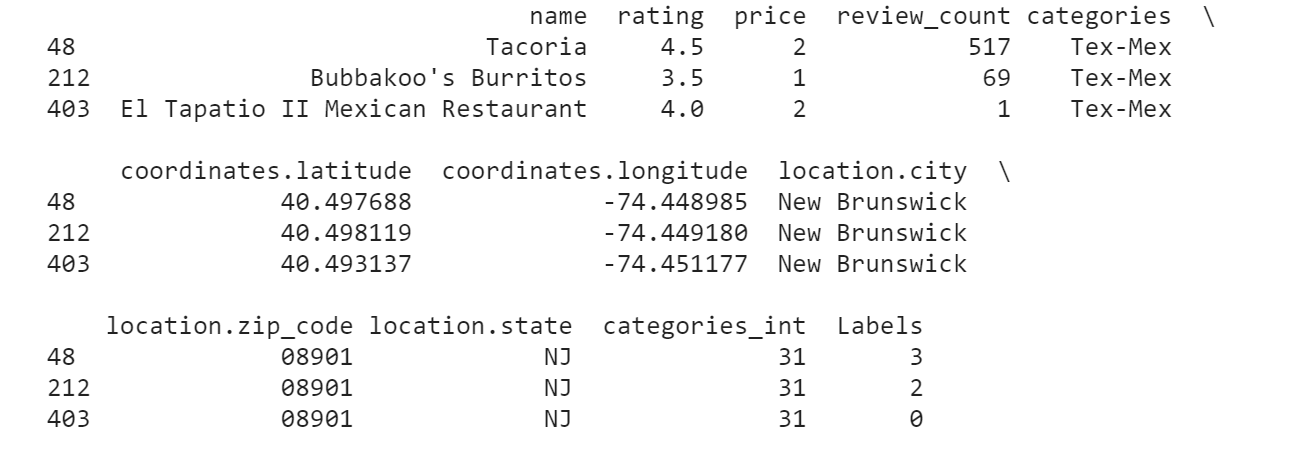
Indian restaurant count = 7

Only 1 Tex-Mex category restaurant was found in this cluster:

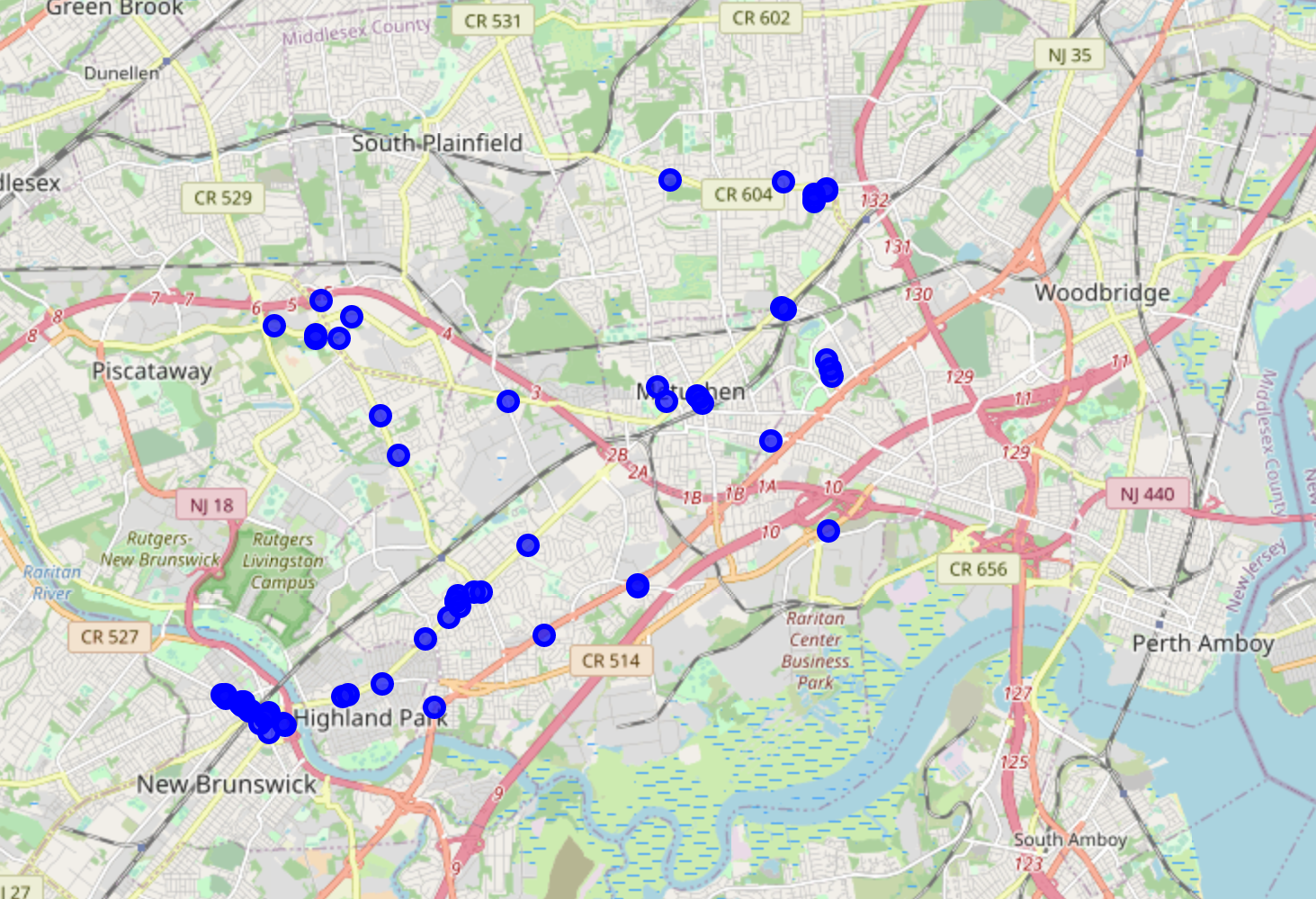
Name= Tacoria, Rating=4.5



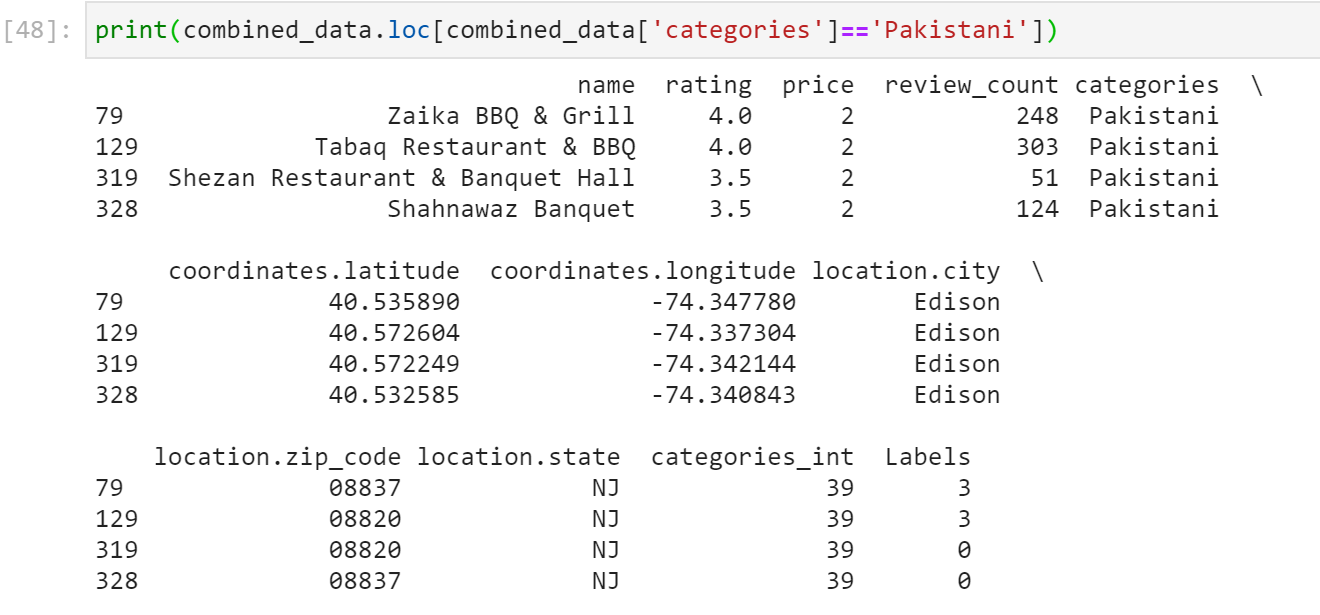
And across the collected data set, only 3 Tex-Mex restaurants were found, with much lower ratings than Tacoria, which leads to conclusions that 3-mile radius around Edison, NJ has only one full-fledged Mexican restaurant. Also, all of these restaurants are located in New Brunswick, NJ.



Cluster 3 mapping of restaurant businesses around Edison, NJ:



Pakistani restaurants in Edison, NJ and their success data points:



1. **Recommendations**

Based on the clustering and further data analysis, my recommendations are as follows for the target audience and stakeholders accessing this report:

1. Launch a Tex-Mex restaurant in Edison, NJ given its success and ratings in New Brunswick area, also currently there no Tex-Mex category of restaurants in Edison, NJ as of now.
2. Launch a Pakistani restaurant in New Brunswick, NJ given there are none in this area and heavy asian population will lead to similar success rates as shown by data points in Edison, NJ area.
3. **Conclusions**

Based on restaurant business data points gathered around Edison, NJ using Yelp’s Business Search API, it provides sufficient information to cluster these restaurants into top performing group based on their Rating, price and reviews. Cluster 3 represents the group of restaurants falling into most successful and best performing based on these data attributes. Further analyzing based on top performing cluster and looking at over all data points, as shown in above sections (5. Results and 6. Recommendations) I can confidently conclude that restaurants under Tex-Mex and Pakistani category will flourish very well in this area.