Grand Opening!

Joseph Ostroman

Giving options of what restaurant to open in Newark, NJ

Introduction

Approached by an investor we were tasked with determining what placement and type of restaurant should be opened in Newark, NJ. According to ‘thebalancesmb.com’ a store should have a parking lot, accessibility, visibility, and a population base. An understanding of the people and current layout of Newark would be important for any decision being made for this situation and could generate good insight into the accessibility and population base that is needed to make an educated decision on where the new restaurant should be located.

Data Acquisition

Data will be used from:

Yelp Fusion API: For businesses/ratings in the township of Newark

Foursquare API: For comparison of business/ratings in the township of Newark

Average Annual Daily Traffic from the state of NJ website : For traffic common spaces.

Population statistics of Newark NJ: For information on People that live in Newark NJ

Cleaning

No missing data was present in the API data.

The data from the “Category” variable was one hot encoded into 69 columns.

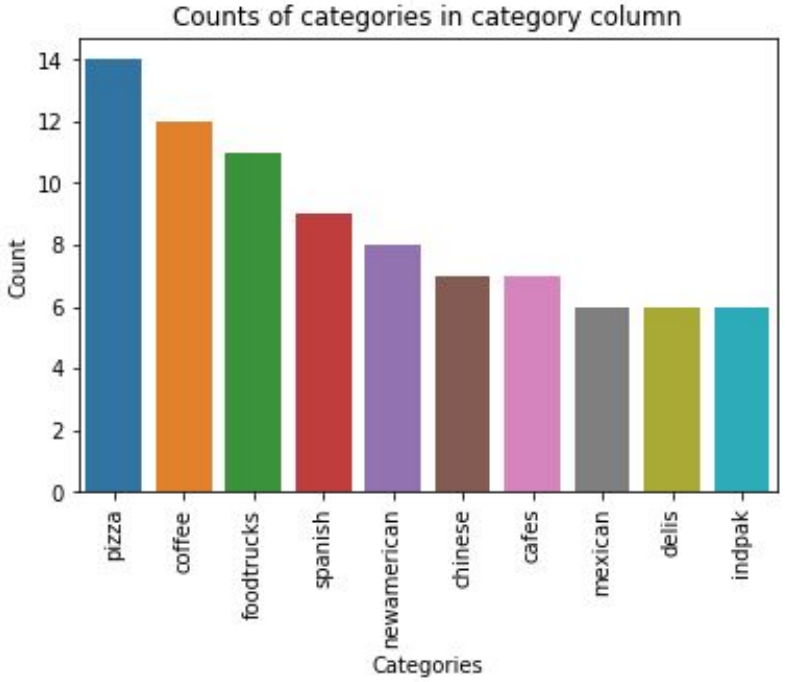
The data also contained some restaurants that were outside the area of Newark, NJ so the data had to be filtered for a boxed section of Newark, NJ.

Variable Descriptive Statistics

	reviews	rating	long	lat
count	163.000000	163.000000	163.000000	163.000000
mean	70.055215	3.628834	-74.169887	40.735573
std	103.997819	0.752750	0.004388	0.004281
min	1.000000	1.500000	-74.179360	40.726841
25%	10.000000	3.000000	-74.172988	40.732056
50%	26.000000	3.500000	-74.170701	40.735347
75%	75.000000	4.000000	-74.166612	40.738695
max	737.000000	5.000000	-74.160670	40.744943

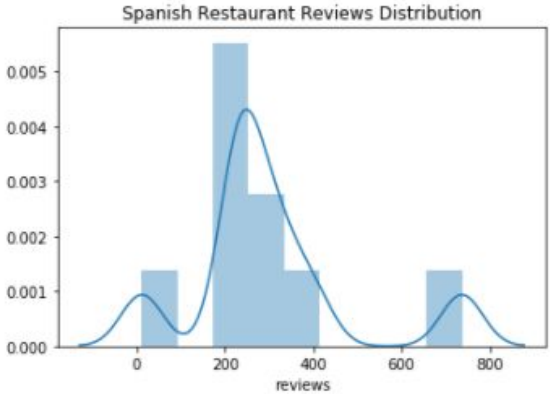
Restaurant Counts in our data.

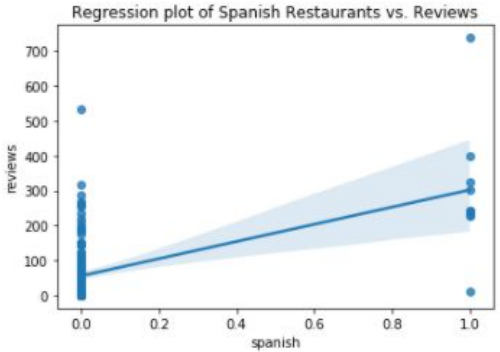
There are many pizza and coffee restaurants in Newark, NJ.

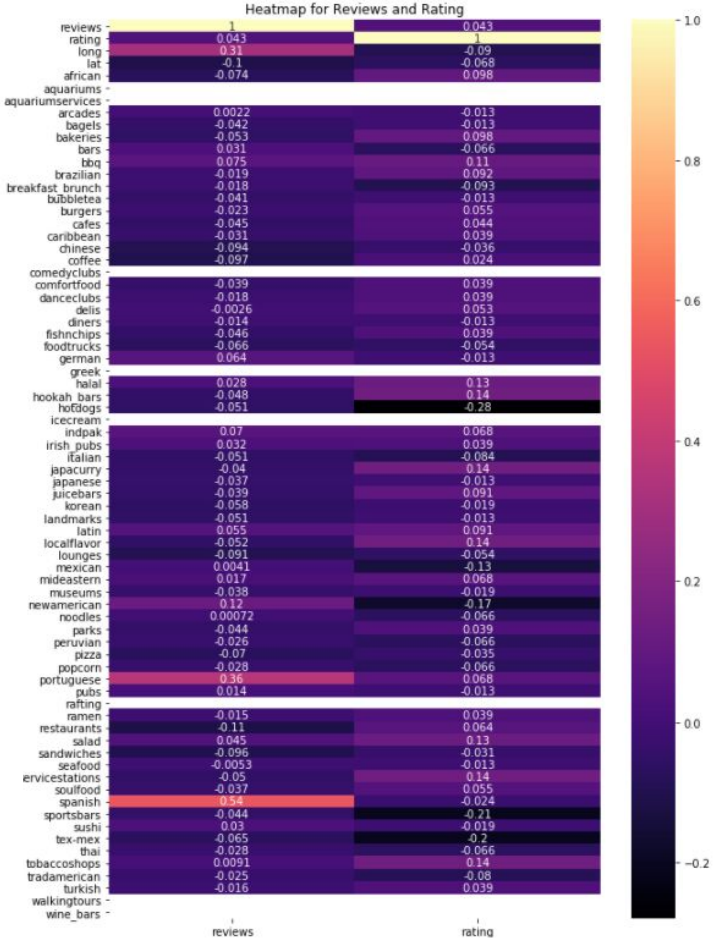


Correlations with “Spanish”

This heatmap shows that there is a correlation between spanish restaurants and the amount of reviews. A distribution plot shows that there are many restaurants above 200 reviews for spanish restaurants.

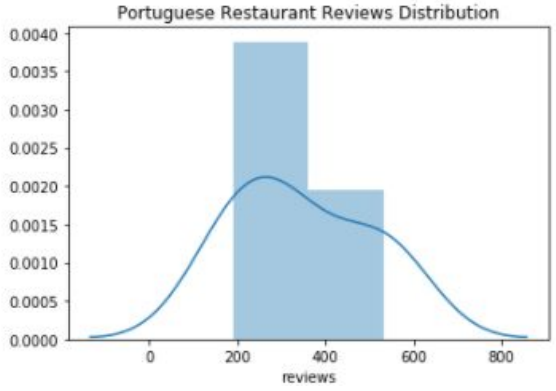


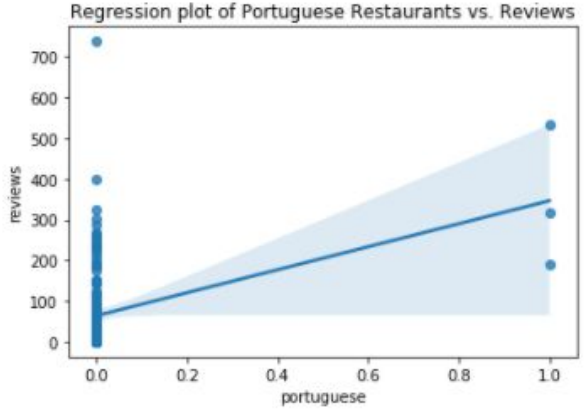


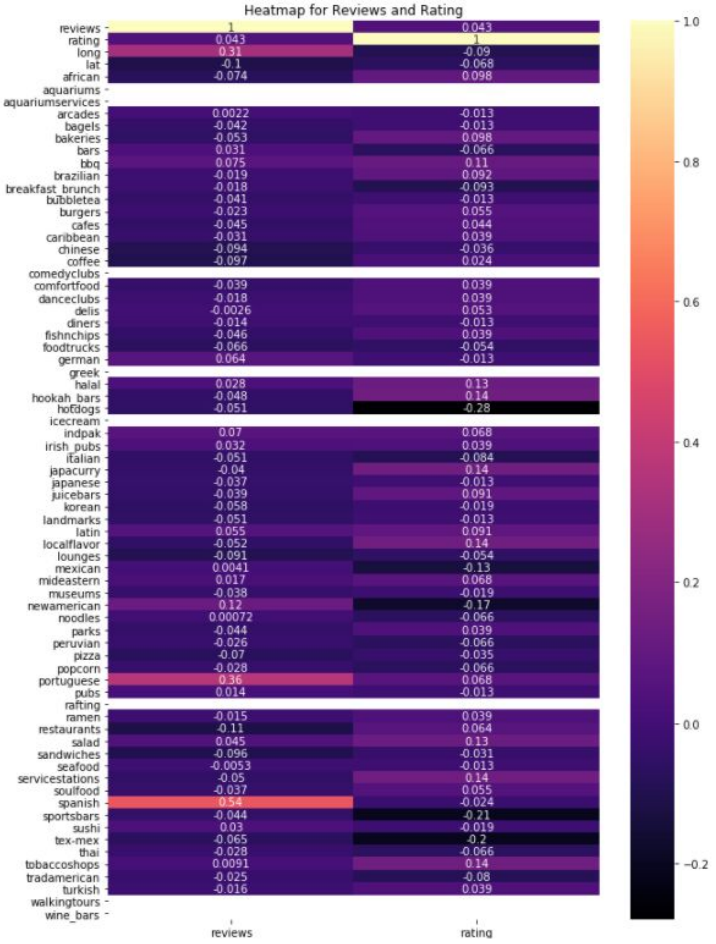


Correlations with “Portuguese”

This heatmap shows that there is a correlation between portuguese restaurants and the amount of reviews. A distribution plot shows that there all the restaurants are above 200 reviews for portuguese restaurants.

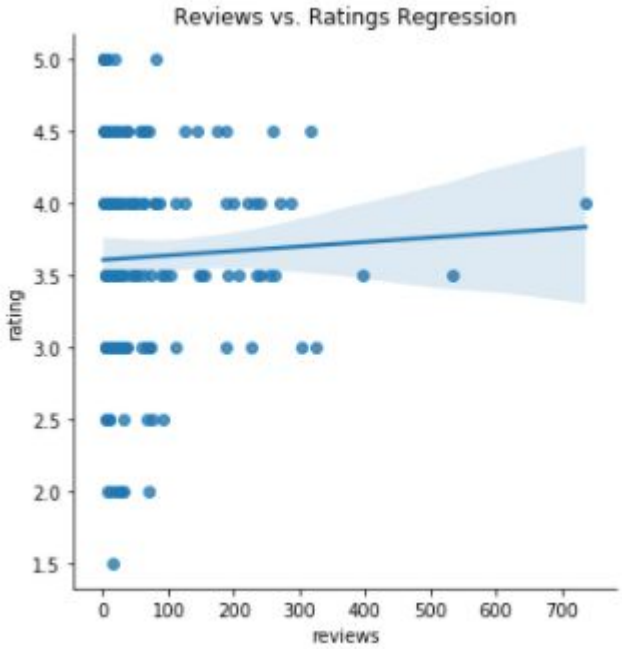






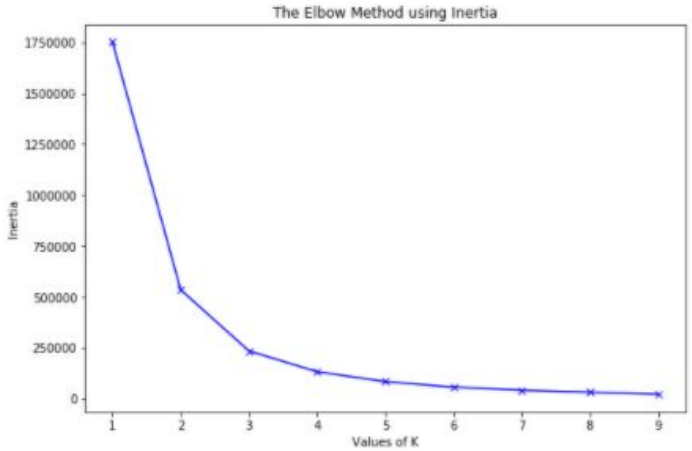
Ratings and Reviews

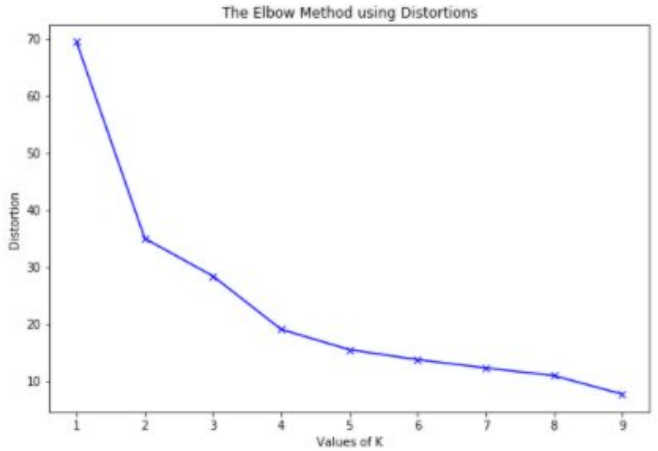
Ratings and Reviews have little to no correlation meaning that higher review restaurants do not have higher ratings.



Elbow Method for Clustering Models

Using the elbow method with both inertia and distortions a K of 2 was chosen to model the dataset.





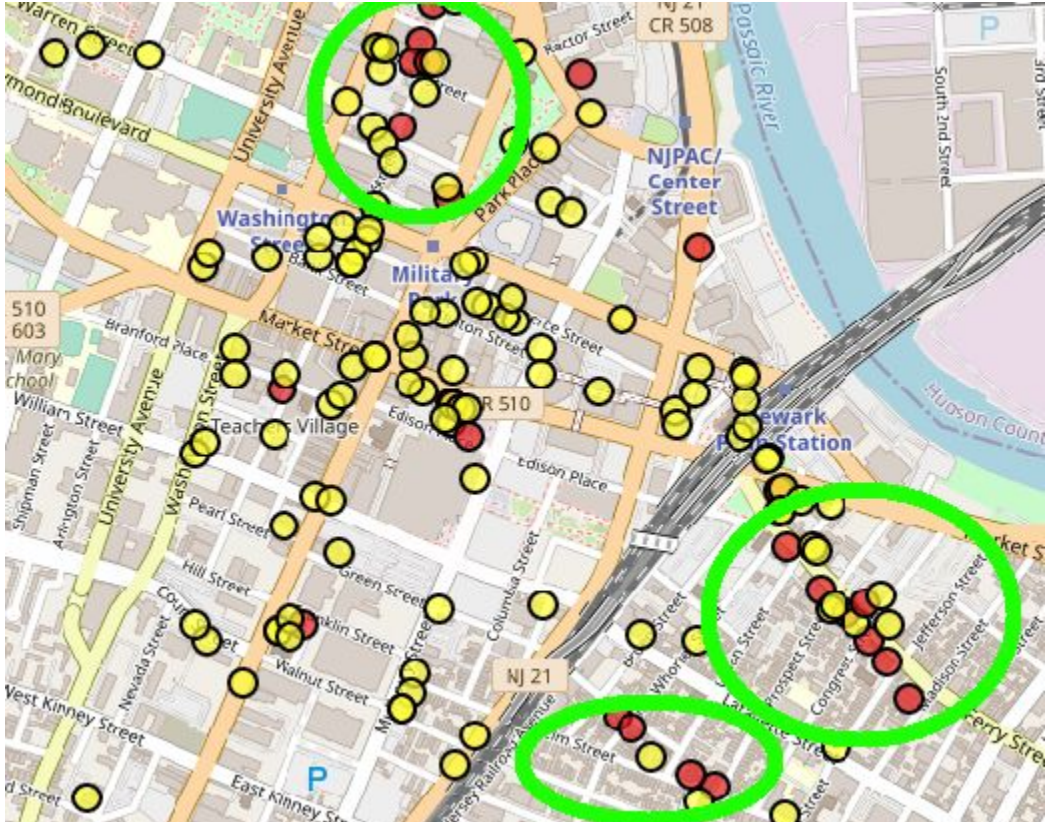
K-Means Clustering

There are three distinct areas where clusters of value one were found. This cluster has reviews that were on average in the 200's with a rating that is slightly higher on average than the others in cluster zero.

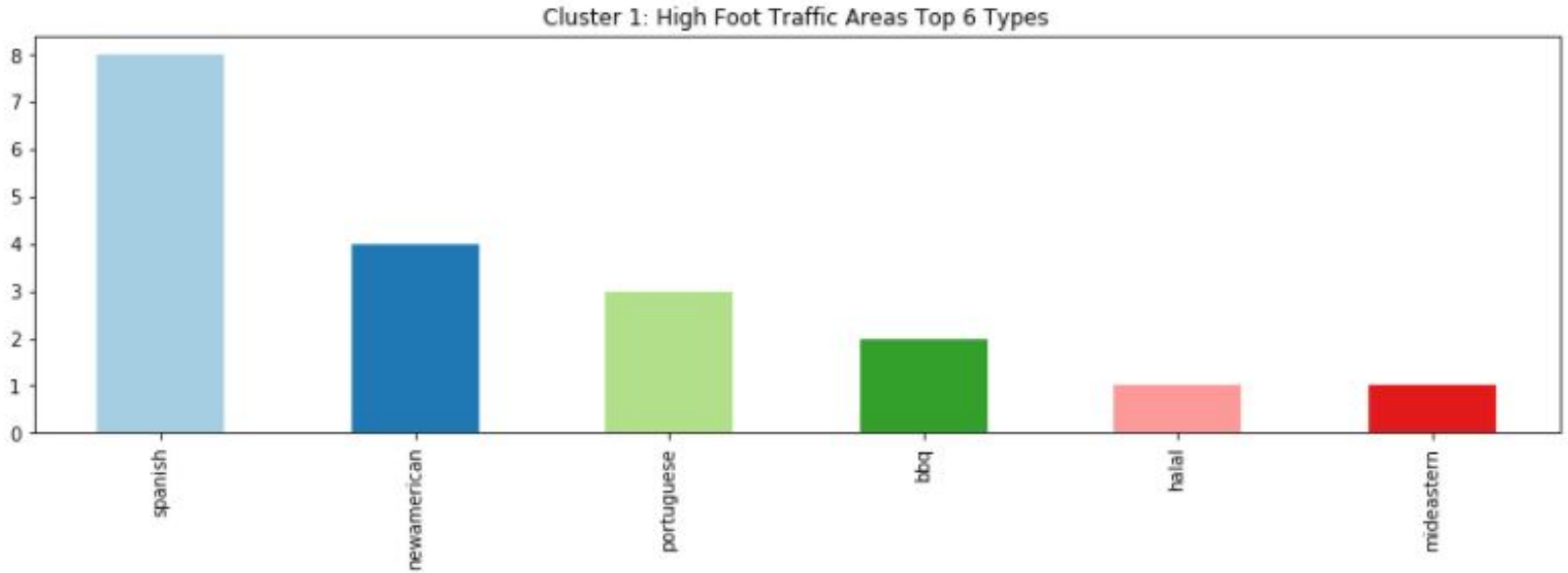
cluster	reviews	rating
0	34.136691	3.607914
1	278.083333	3.750000

Yellow

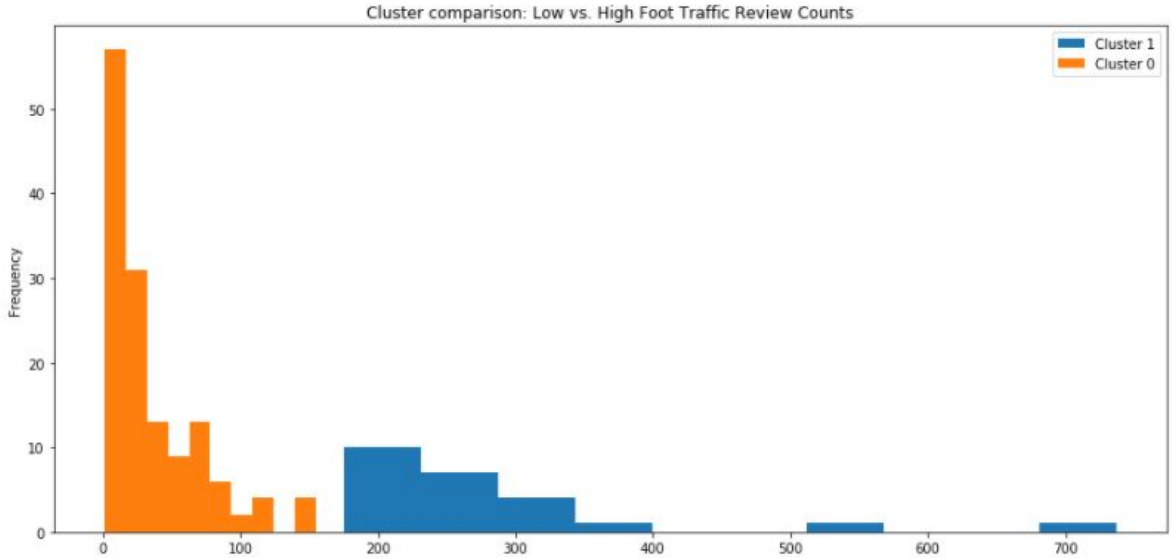
Red



Cluster one top 6 restaurant types



Review counts for each cluster



Conclusion and Future Direction

- K means clustering algorithm created for unsupervised learning using API from both Foursquare and Yelp.
- The model found a cluster of restaurants that show high review counts where more people visit and leave reviews.
- More data could be used like crime rate and traffic in different parts of the city.