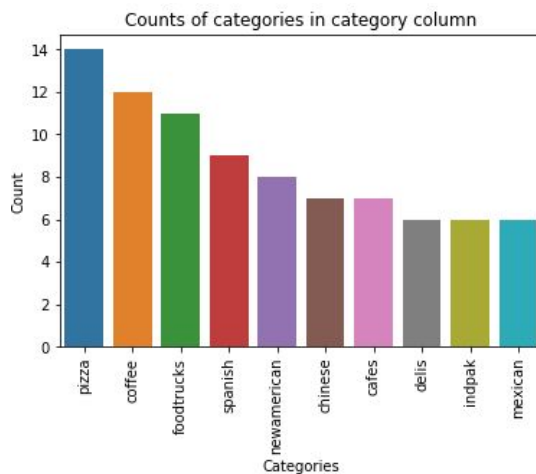


Data

The data that was collected for this report was allocated from the Yelp API's businesses search. The columns in the data correspond to important characteristics of each restaurant. The columns used are, Name, Reviews, Rating, Latitude, Longitude, and the category alias or Type. The type of restaurant was one hot encoded for purposes of segmentation. Descriptive statistics of each of the columns are listed below.

	reviews	rating	long	lat
count	198.000000	198.000000	198.000000	198.000000
mean	77.348485	3.760101	-74.179943	40.735841
std	112.100502	0.773465	0.591560	0.051162
min	1.000000	1.500000	-82.401843	40.058863
25%	10.000000	3.500000	-74.172499	40.732373
50%	29.000000	4.000000	-74.169795	40.735776
75%	81.750000	4.500000	-74.163111	40.740788
max	737.000000	5.000000	-73.922892	40.867447

Countplot provided for different types of establishments in Newark NJ.



The categorical data will be one-hot-encoded into numerical features and fed into a Kmeans clustering algorithm using sklearn. The algorithm will provide clusters that will give insight into the highest rated restaurants as well as restaurants that have foot traffic (a high number of reviews)