**Introduction**

The setting up of a new business establishment is always a headache for new business owners. Conventional thinking is that location would determine the success or failure of a new retail outlet. In this case that we will be looking into, the context would be a budding entrepreneur looking to set up his first cafe in the city of New York.

**Data**

FourSquare API has kindly provided 2 kinds of data which will be useful for our analysis. The first is geographical coordinates in the form of latitude and longitude. We will be using these coordinates to numerically define the addresses of locations of interest which we are analyzing. The second is pricing data which is given in a categorical form. Each establishment is ranked by price tier ranging from 1 to 5. We will be using this as a proxy to measuring the pricing power of cafes.

**Methodology**

For the main machine learning process, as I do not know what to expect, I have opted for the use of unsupervised learning methods. The core method used for this analysis is kMeans Clustering as the datapoints are generally well-distributed rather than differing in density across locations. Hence, kMeans is enough for this analysis. Later, statistical testing in the form of calculation of mean value is used to understand the overall pricing power of cafes within each cluster region. However, to further unveil the effects of location, further scatter plotting of higher price tier cafes is done to show if noise had interfered with the initial prior analysis.

**Results & Discussions**

At first glance, only 1 cluster showed some meaningful increase in pricing power as exhibited by mean pricing power to be at 1.33. However, such a value could potentially be artificially inflated as the sample size of that cluster was smaller compared to other clusters. A few higher tier cafes could skew the mean result. Later, second plotting of only higher tier cafes was done to show distribution of such cafes. It was found that such cafes tended to concentrate themselves in central financial district of New York, which validated the increase in pricing power of clusters formed at the central region. However, the low mean values obtained suggest many other cafes still faced pressures to charge lower. My hypothesis is that due to the oversaturation of cafes attracted by the higher footfall and higher wealth customers at the financial district area, competition has negated much of the pricing power increase that comes with the location. Why some cafes are able to charge higher is still a mystery and further study has to be done.

**Conclusion**

In light of the findings of this analysis, future study could touch on the qualitative factors of the café such as parking lot availability and payment methods.