**Introduction/Business Problem**

This project will analyze the neighbored of Boston, MA and potential venues of interest.

As someone who is lives in Greater Boston and works in downtown area, I am often asked for recommendations, things to do, and areas to live from friends, colleagues who either live in different US cities or who are travelling to Boston on a work/pleasure trip. A systematic analysis of neighbored and venues not only benefit them but it will be beneficial for me as well at a personal level to explore areas and places which I have not explored.

To successfully complete our analysis, we will need to explore available dataset, explore neighborhood in Boston, analyze each neighborhood, cluster neighborhood and finally examine the clusters.

**Data**

To perform our analysis, we will need two type of data.

First, we will need location data. For this, we will use foursquare location data which will provide us venues on a given location. We will fetch top 100 venues in nearly 3 mile radius of Boston geographical coordinates.

Second, we would need Boston’s neighborhood data. For this, we will use two sources and merge the data to obtain desired geo coordinates for various neighborhoods. For easy reference, i have downloaded this data in an excel which will be used.

ZIP Code: <http://archive.boston.com/news/local/articles/2007/04/15/sixfigurezipcodes_city/>

GeoData: <https://public.opendatasoft.com/explore/embed/dataset/us-zip-code-latitude-and-longitude/table/?q=boston&refine.state=MA&location=11,42.36681,-71.18952&basemap=jawg.streets>

### Methodology

Our methodology for this analysis will have following five important component

1. Explore available dataset: In this section we will access, clean , reshape location and venue data from sources mentioned in above step.
2. Explore neighborhood in Boston : In this section we will look at different neighborhoods of Boston
3. Analyze each neighborhood : In this section we will start analyzing the neighborhood data
4. Cluster neighborhood : We will use k-means to cluster the neighborhood
5. Examine the clusters: here, we will use the cluster data and examine to understand the top venues.

For our work, we will use different python libraries which are available to us. Some of those are pandas, numpy, json, geopy, matplotlib, folium and sklearn.

**Results**[**¶**](https://render.githubusercontent.com/view/ipynb?commit=5b4c4390df9e5996d747d37f665a44cf6511a7e1&enc_url=68747470733a2f2f7261772e67697468756275736572636f6e74656e742e636f6d2f616d69746a70616e6465792f636f7572736572612d63617073746f6e652f356234633433393064663965353939366437343764333766363635613434636636353131613765312f736372697074732f66696e616c25323070726f6a2f546865253230426174746c652532306f662532304e65696768626f72686f6f6473253230285765656b25323032292e6970796e62&nwo=amitjpandey%2Fcoursera-capstone&path=scripts%2Ffinal+proj%2FThe+Battle+of+Neighborhoods+%28Week+2%29.ipynb&repository_id=281681670&repository_type=Repository#Results)

By examine clusters, some insights are very clear. Such as:

1. Boston’s north end is best for Italian food. It is common knowledge for someone who lives in the area but data proves it that Italian food places are the number one venue in this area. Cluster also shows that inner harbor/east Boston also ranks number one for Italian food.
2. Majority of venues are in cluster 3 which are areas around downtown Boston. It shows variety of venues with quite accurate results. For example, Chinatown’s number one venue is Chinese restaurant, where as people prefer to shop cloths around prudential.

**Discussion**

Even though cluster information is helpful in understanding top venues of Boston, some of the results are not very convincing. For example, Spa seems to be number one venue of many neighborhood which doesn’t sound right. Neighborhood of Roxbury are in two neighborhood. All this indicate poor coverage in location data and also makes me question the foursquare location data. It is advisable, that more data sources are explored to gain confidence

**Conclusion**

In conclusion, the location data and geo data gives us quite a few good insights. We can certainly get an idea which neighborhoods to recommend to our friends and colleagues based on their need. However, in some areas, data coverage doesn’t seems very good but it is recommended that analysis is enhanced with better quality of data or by accessing multiple data sources.