Data Description: Restaurant Placement in Los Angeles

The Framework: For this developmental process, we will make extensive use of Python, utilizing Jupyter Notebooks to process the data and algorithms.

Python Libraries: Python Libraries provide a ready-made assortment of scripts and algorithms to execute the processes we need to perform on the data. The libraries we will make use of will include:

1. json – library to handle json files
2. geopy – converting addresses into latitude and longitude / geocoding
3. folium – for graphing data
4. BeautifulSoup – to scrape web pages
5. csv – to read and write csv files
6. pandas - to read and create data frames
7. numpy –perform data analysis on frames
8. matplotlib – to graph data and colorize plots on a map
9. sklearn – this library provides us with data analytics algorithms including k-means clustering

Wikipedia: The list of neighborhoods in Los Angeles was scraped using BeautifulSoup from this page: https://en.wikipedia.org/wiki/List_of_districts_and_neighborhoods_in_Los_Angeles. This is the first data set from which we determine the clustering of venues – each venue by category is classified by neighborhood.

ArcGIS: ArcGIS will be used to geolocate all the neighborhoods. This will be the starting point from which we locate the relevant categories of venues around each neighborhood.

Foursquare: The Foursquare API for "Places" will be essential in:

1. Locating venues in each neighborhood from the geolocated coordinates from ArcGIS
2. Classifying each venue according to the categorical information in the Foursquare database

Once all venues are plotted and mapped, pivot tables will be used to classify each category of venue by most frequent in that particular neighborhood.

Once the frequency and quantity of venues in the neighborhoods is established, the data will be analyzed by plotting and pivot for vacuums of restaurants in commercial areas.