**Introduction/Business Problem**

**Analyze Steakhouse vs Vegetarian / Vegan Restaurant Concentration in New York (Brooklyn, Manhattan)**

* Business problem: Decide which New York neighborhoods to target for a PETA (People for the Ethical Treatment of Animals) animal awareness campaign
* As a proxy, we will compare the # of Steakhouses vs Vegetarian / Vegan Restaurant in various NY neighborhoods, trying to see if there are obvious geographical patterns for fellow animal supporters

**Data Section**

* For NYC geographical data, we will first use the NYU library's "2014 New York City Neighborhood Names". It contains the name each NYC neighborhood
* Next, we will find the coordintes of the neighborhoods via geopy.geocoders
* We will the load NYC neighborhood data into a Pandas dataframe
* For contextual data (steakhouse vs veg), we will use the Foursquare API, filtering via category codes
* For this study, we will focus on the boroughs of Manhattan and Brooklyn

**Visualization setup**

* Get coordinates of Manhattan / Brooklyn
* Populate raw Folium map

**Methodology - Data Analysis**

* We need to set a threshold for minimum number of results, as some neighborhoods doesn't have many steakhouses or veg restaurants. Looking at the sorted # of results, it seems 10 is a reasonable cutoff.
* If a neighborhood has less than 5 Vegetarian / Vegan Restaurants and less than 5 Steakhouses, we consider the neighborhood "undecided" and drop it from the analysis. This faciliates one-hot processing / normalization, as it gives equal-weight to each neighboorhood (e.g., 0 veg + 1 steak has same one-hot output as 0 veg + 100 steak).

**Methodology: Model Selection**

* We apply a KMeans(n=3) Clustering analysis
* We tried various valuse (n=3..7) and noticed that there are really just three big groups, namely...
* Group 1 = "More steakhouse than veg";
* Group 2 = "Mostly veg";
* Group 3...n = "More veg than steakhouses"
* With the above observation, it seems n=3 is a good choice

**Results: Visualization & Conclusion**

* We found that there are many areas with predominately veg places with little to no steakhouses (group 1). This makes up 23 of the 61 nighborhoods.
* We didn't find the reverse - namely, areas with mostly steakhouses and little veg places except for two areas (Manattan Terrace, Homecrest)
* As noted in the previous section, the map shows a relatively contiguous clustering...
* Group 1 (red) = "More steakhouse than veg";
* Group 2 (blue) = "Mostly veg";
* Group 3 (green) = "More veg than steakhouses"
* The clustering shows a very clear pattern
* Group 1 concentrates in just two areas: midtown Manhattan, the Financial District of Manhattan and southern Brooklyn. The two Manhattan neighborhoods happens to be where financial companies are heavily concentrated, and are thur frequent "corproate dinner" destinations.
* The rest of Manhattan is composed of Group 3. Given that generally there are in general more veg places than steakhouses, this is likely a neutral composition and simply reflects a balance of preferences. Also keep in mind that steakhouses are generally far larger establishments than veg spots, so the # of patrons served is closer to 40/60 than 30/70 as indicated in the previous section.
* Downtown Brooklyn (immediate southern area over the Manhattan bridge) also comprises of group 3. This likely reflects the relatively higher income of the area, as that region serves as homes for many Manhattan commuters.
* Group 2 is composed of the rest of Brooklyn, including young and trendy neighborhoods like Williamsburg, Greenpoint, and Bushwick. More surprising is that this trend extends to the rest of Brooklyn. This likely reflects not so much a high concentration of veg places than a lack of steakhouses.
* With the data in mind, our suggestion for the PETA animal awareness campaign is to target business addresses in Manhattan Midtown + Finacial District, as well as residential areas of downtown Brooklyn, and possibly the (Russian) ethnic neighborhood in southern Brooklyn