**Machine Learning Capstone**

**Data**

**Miami Neighborhood Data**

Neighborhood data will be scraped from [**https://en.wikipedia.org/wiki/List\_of\_neighborhoods\_in\_Miami**](https://en.wikipedia.org/wiki/List_of_neighborhoods_in_Miami)**.**

Only neighborhood names and coordinates will be used from this.

Here is a partial screenshot:

****

**Preprocessing**

After scraping, the data will be fed into a pandas dataframe with the irrelevant columns removed.

The neighborhood containing Chicago Boys Town will be appended to this dataframe because the objective is to find a neighborhood like a gay neighborhood, so I’m interested in which Miami neighborhoods cluster with a gay neighborhood in Chicgo.

**Foursquare**

Foursquare is used to ascertain various venues for each neighborhood.

The venue information is processed to return the top 10 venues by count for each neighborhood. This is saved in a dataframe and the neighborhoods are then clustered using kmeans clustering.

See the attached notebook for details.

  
Figure 1: Sample of Neighborhoods with Counted Venues

  
Figure 2: Sample of Cluster which contains KitKat