Battle of Neighborhoods
Determining a suitable business location in Raleigh with the power of Data and Science

Battle of Neighborhoods

Determining a suitable business location in Raleigh with the power of Data and Science

Problem Description

- Wake county located in North Carolina is one of the fastest growing counties in the USA
- Large influx of immigrants Its major city Raleigh
- Highly educated young workforce
- Tremendous possibilities for new businesses to thrive
- North Carolina is also famous for its barbecue and bacon, but other cuisines are starting to become popular
- Develop a recommender system to open a new restaurant that can be successful in the area.

Datasets

The following datasets will be used for our assessment

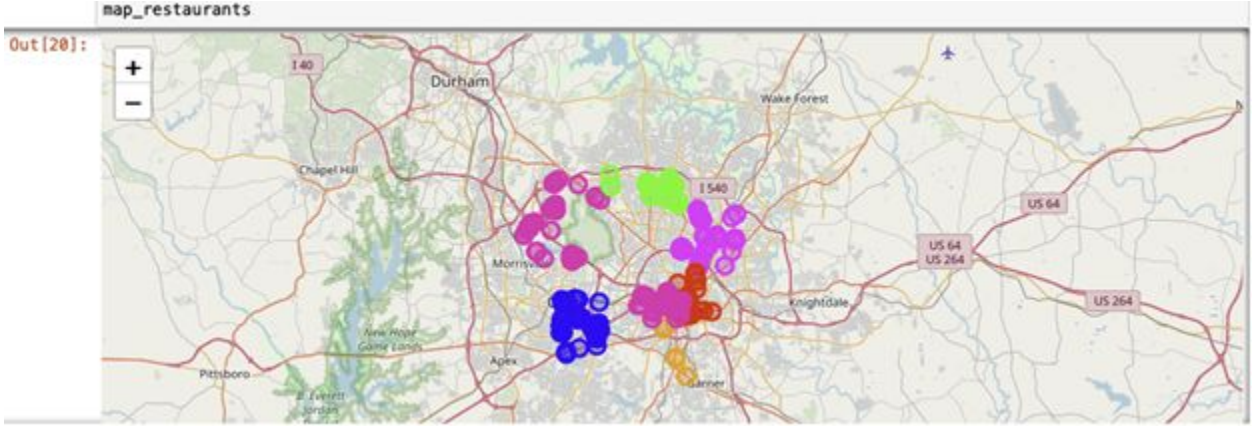
- 1. *https://en.wikipedia.org/wiki/Raleigh,_North_Carolina_neighborhoods*
- *2. https://en.wikipedia.org/wiki/Category:Neighborhoods_in_Raleigh,_North_Carolina*
- *3. Restaurant density data from Foursquare API's*
- *4. Classification of types of restaurants based on cuisines*

Exploration,Data cleanup and Machine Learning tools

- The data will be read via ***pandas*** dataframe and clean up to classify Restaurants to Neighborhoods.
- ***Geopyclient*** to obtain location information
- ***Foursquare API's*** to explore restaurant data in the neighborhood
- ***Folium*** for visualizations
- Then we will be able to cluster the restaurants in the area and as well as correlate the quality of restaurants again based on cuisine to the neighborhood which would potentially provide us insights on the type of restaurant that can be setup.
- We will then perform data clustering using **KNN** on the restaurant data and generate visualization plots and graphs that will aid in the final decision

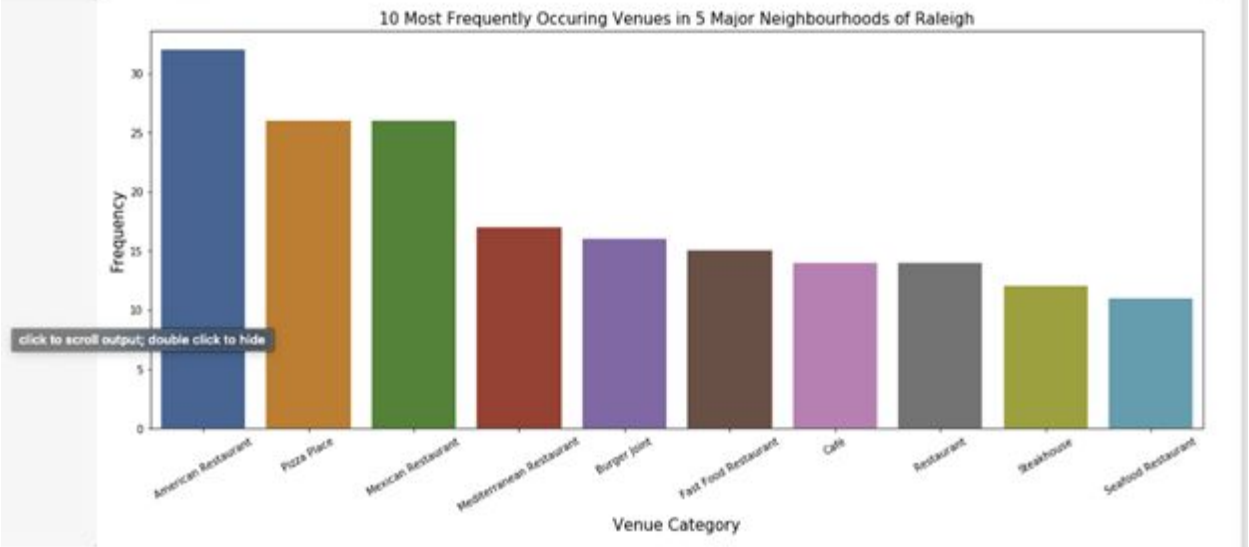
Data Visualization

Restaurant concentration in the various Raleigh neighborhood



Ten most common venues across each Neighborhood

In [23]:
```
import seaborn as sns
fig = plt.figure(figsize=(18,7))
s=sns.barplot(x="Venue_Category", y="frequency", data=final_Ral_clusters_top10)
s.set_xticklabels(s.get_xticklabels(), rotation=30)
plt.title('10 Most Frequently Occuring Venues in 5 Major Neighbourhoods of Raleigh', fontsize=15)
plt.xlabel("Venue Category", fontsize=15)
plt.ylabel ("Frequency", fontsize=15)
plt.savefig("Most_Freq_Venues.png", dpi=300)
plt.show()
```



- 54 unique categories
- Top three common restaurant across Neighborhood
  - American
  - Pizza Place
  - Mexican

%%%%%East Raleigh,Raleigh,NC%%%%%

	Venue	Freq
0	American Restaurant	0.08
1	Bakery	0.08
2	Southern / Soul Food Restaurant	0.06
3	Café	0.06
4	Pizza Place	0.06

%%%%%North Raleigh,Raleigh,NC%%%%%

	Venue	Freq
0	Pizza Place	0.20
1	Mexican Restaurant	0.12
2	Café	0.08
3	Restaurant	0.04
4	Italian Restaurant	0.04

%%%%%Northeast Raleigh,Raleigh,NC%%%%%

	Venue	Freq
0	American Restaurant	0.14
1	Fast Food Restaurant	0.08
2	Burger Joint	0.08
3	Restaurant	0.08
4	Mexican Restaurant	0.06

%%%%%Northwest Raleigh,Raleigh,NC%%%%%

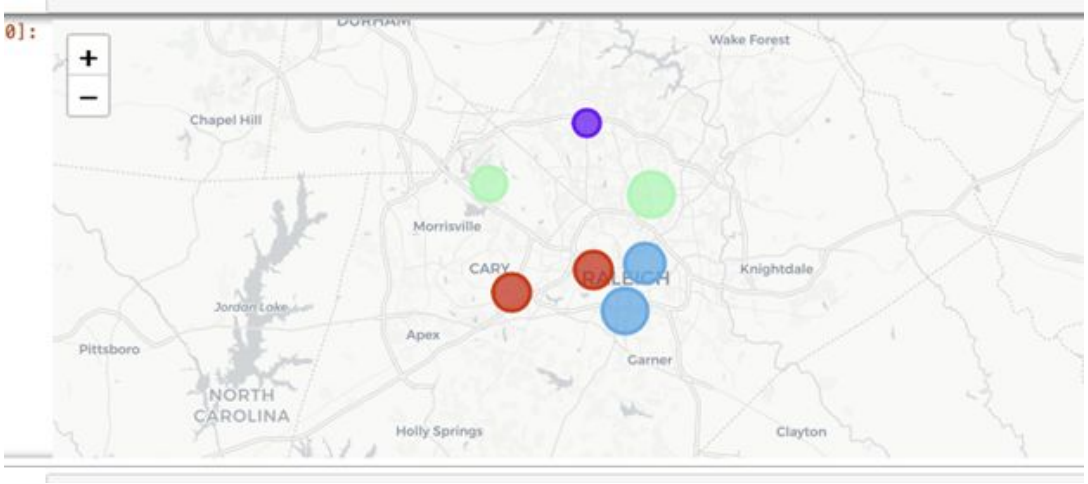
	Venue	Freq
0	American Restaurant	0.12
1	Steakhouse	0.08
2	Pizza Place	0.08
3	Mexican Restaurant	0.08
4	Fast Food Restaurant	0.06

%%%%%Southeast Raleigh,Raleigh,NC%%%%%

	Venue	Freq
0	American Restaurant	0.14
1	Seafood Restaurant	0.08
2	Italian Restaurant	0.06
3	Asian Restaurant	0.06
4	Burger Joint	0.06

Results

- From the various exploratory data analysis and clustering operations performed on our dataset,
- Cluster 1(North Raleigh) has the lowest number(smallest radius) of neighborhood
- This could mean a potential location to open a restaurant in this area.
- Southern Raleigh is oversaturated with restaurants
- Our advice to new potential business owners is to avoid this area if you plan to open American or Mexican restaurants.



Conclusions

This however is a small study based on the learnings from the course. In the future we can expand our analysis by getting trending venues, restaurant ratings information to dissect the data further to provide any new businesses revealing insights on the type (breakfast, sit-down, fast food), cuisine (American, Indian,Mexican..),the ideal price point that will determine if it can become a successful venture