HARLEM, NYC: WHAT CUISINE SHOULD YOU INVEST IN?

IBM Capstone Project

July, 2020

RESTAURANT SATURATION BY CUISINE

- The restaurant and food culture in New York City is one of the best in the world. At it's core, the food scene in NYC is diverse and difficult to succeed in.

- Harlem, one of the neighborhoods in Manhattan's Upper end, is a hub of that food culture.

- An article recently described Harlem as "New York's Best Food Neighborhood."

- When restaurant investors, or small business owners, are evaluating starting a new restaurant, they consider the number of existing, similar restaurants they will compete with.

- In Harlem, the diversity of cuisines means you have a lot of competition. It would be useful to know which cuisines are more saturated with existing restaurants, and which are under represented.

- By considering the current state of restaurant saturation, by type of cuisine, investors and new owners can make more informed decisions about what types of restaurants will be most successful.

DATA ACQUISITION AND CLEANING

Sources
- Longitude and Latitude for US Zip Codes sourced from public.opendatasoft.com
- New York City, NY Zip Code and Neighborhood data sourced from www.health.ny.gov
- Popularity Ratings of Food Cuisines from a Google Study.
- Harlem, NY local business data sourced from Foursquare

Cleaning and Feature Selection

- Joined the US Zip Code data with the corresponding Manhattan neighborhoods, using the zip code feature as the key to combine both data sets; resulted in a dataframe containing the features: Longitude, Latitude, Neighborhood, Borough, and State

- Reduced the dataframe by selecting for only neighborhoods corresponding to 'Harlem'

- Queried for local businesses using Foursquare API, targeting businesses within 500 meters of the zip codes correlating to Harlem

- Initially resulted in a dataset containing hundreds of rows, which I reduced to only contain businesses serving food using unique Venue Categories; resulting in a cleaned dataframe of 170 rows and 7 features: Neighborhood, Latitude, Longitude, Venue, Venue Latitude, Venue Longitude, and Venue Category

# RESTAURANTS' STAKE IN LOCAL BUSINESS

Calculating the number of Restaurants in Harlem:

- The number of restaurants, compared to the number of businesses overall, was staggering.

- Restaurants and food services currently account for 56.3% of all businesses operating.

Restaurants vs. Other Businesses

Restaurants 56.3%

Other Businesses 43.7%

# DISTRIBUTION OF SERVICE TYPES
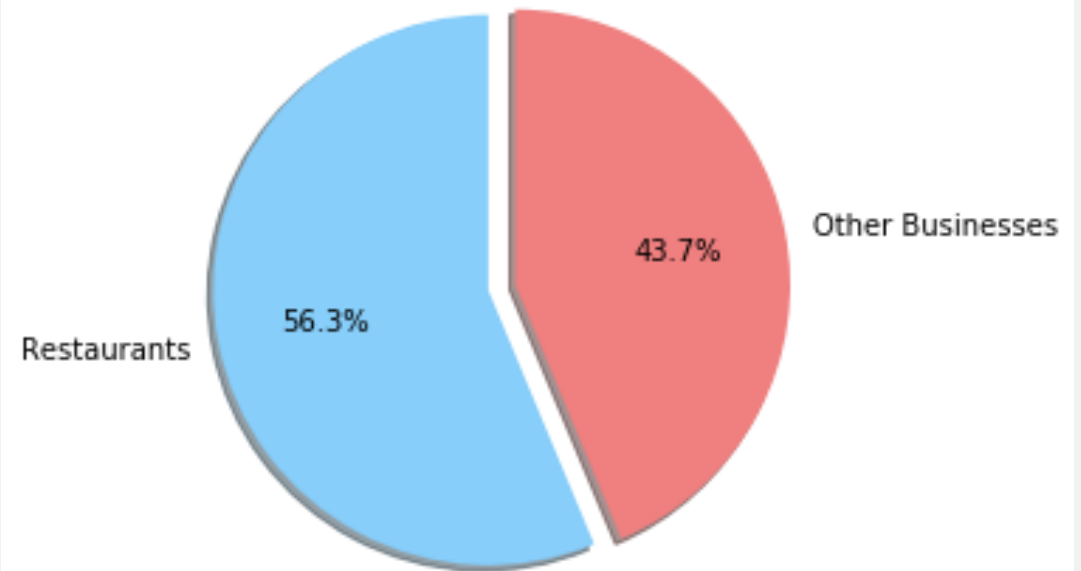
Distribution of Service Types
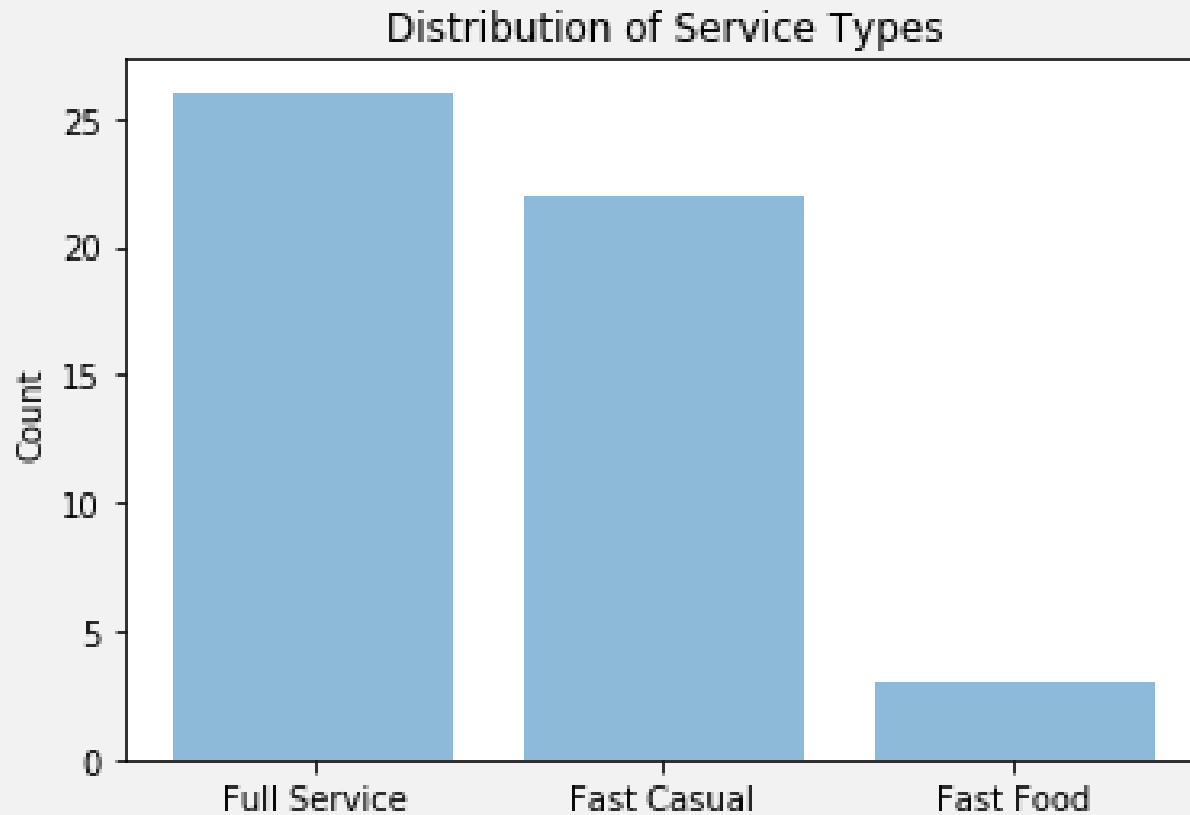
- The type of service style for restaurants is depicted in this graph.
- The two most common in Harlem are Full Service and Fast Casual dining.
- As a potential investor, you can conclude that by and large Fast Food restaurants are not as successful

MODELING: CLUSTERING BASED ON POPULARITY, SERVICE TYPE, AND CUISINE SATURATION
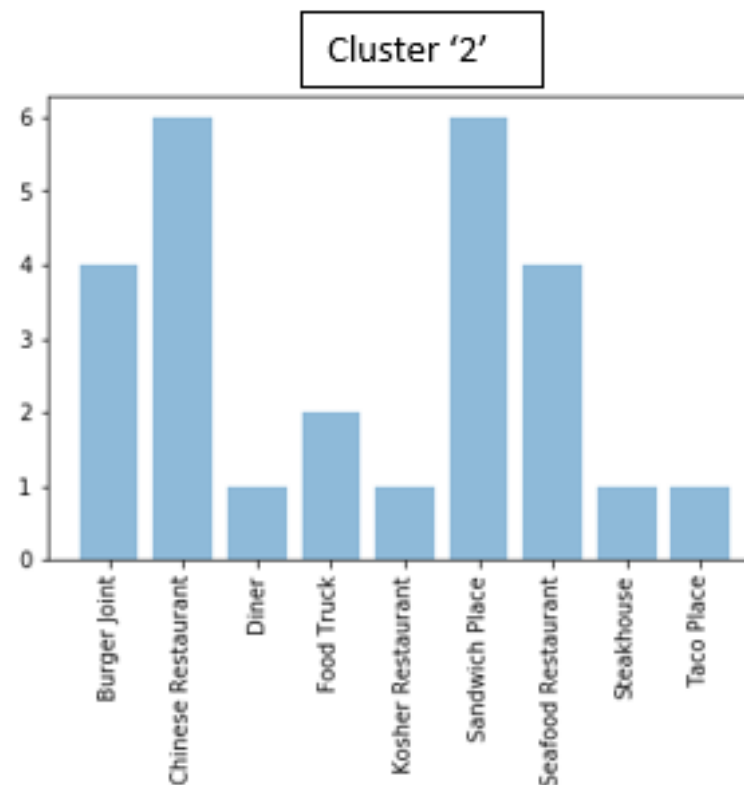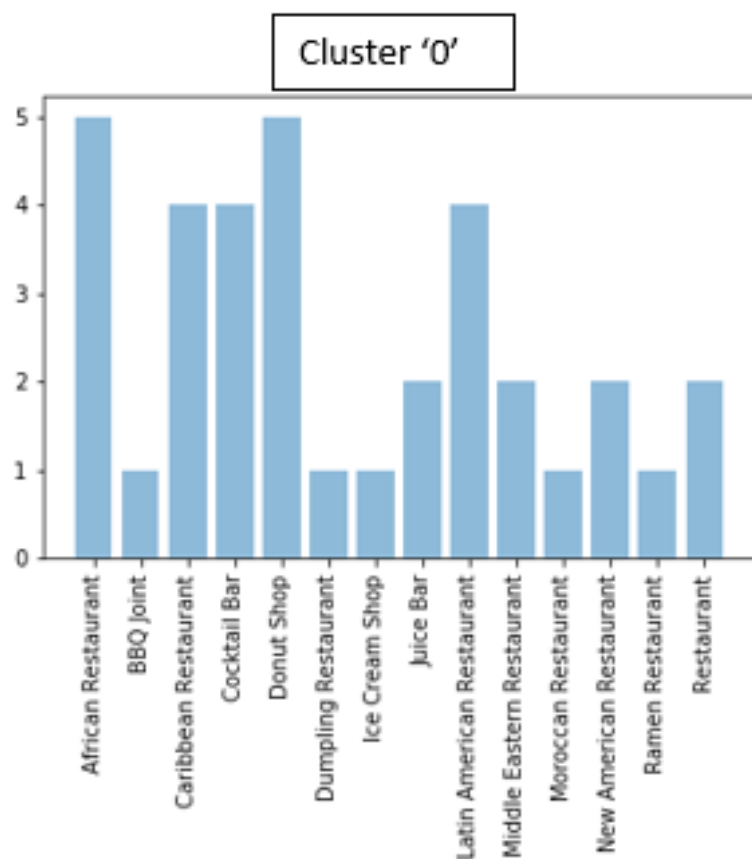
- The Kmeans algorithm produced 5 clusters. Two appear to have the highest popularity score and number of restaurants.
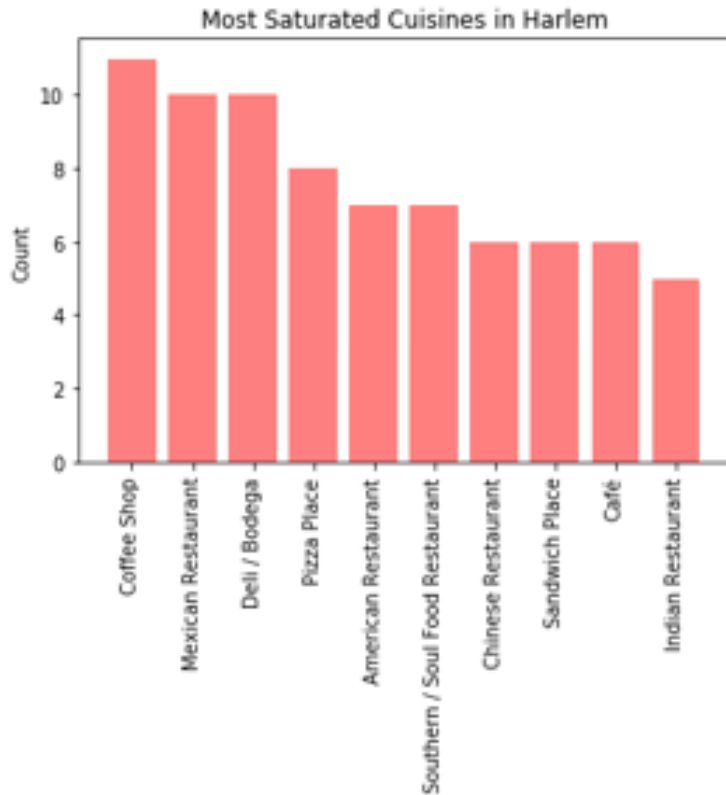- Clusters '0' and '2' had the highest aggregate values for each feature

CLUSTERS: POPULARITY SCORE DISTRIBUTION

- Clusters represented by their distribution of popularity scores

- Clusters 0,2 and 4 have the highest average scores
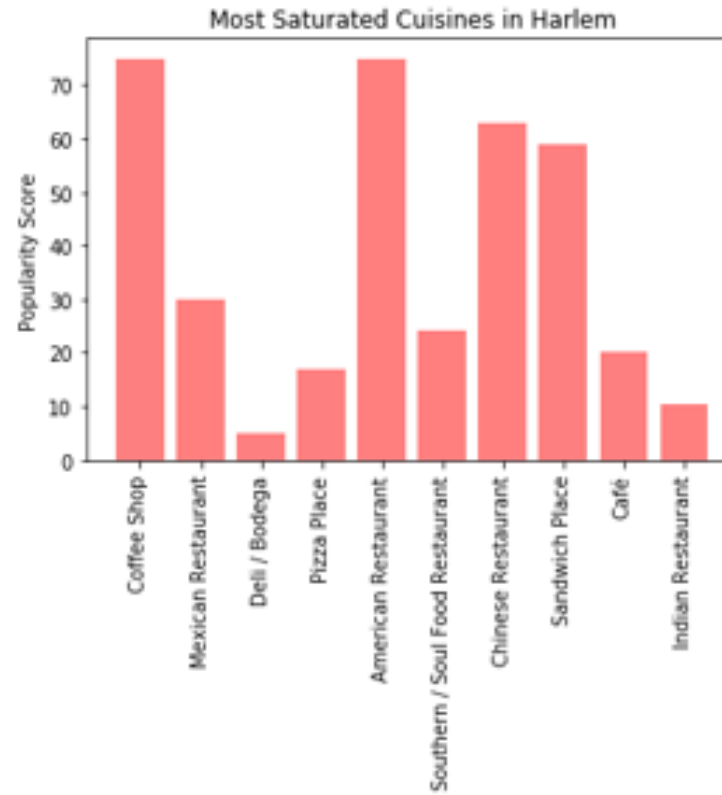
- Cluster 3 has the lowest
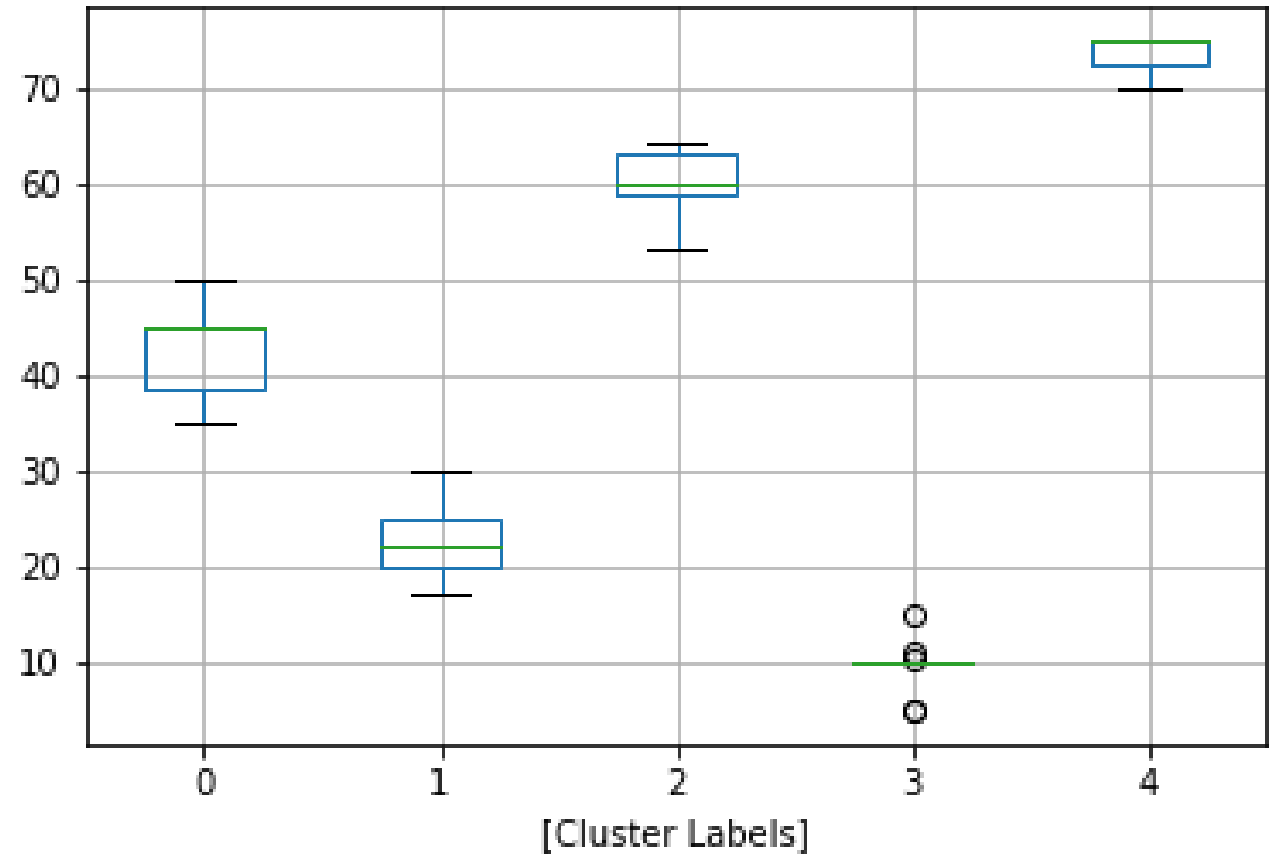
SATURATION OF CUISINES: COUNT AND POPULARITY SCORE

Overall Rating

Overall Rating

- Clusters '0' and '2' had the highest aggregate values for each feature
- Cluster 0, when looking more closely, is too overly populated to be wise for investment, when considering the overall rating

CONCLUSIONS

Cluster 2: Best Popularity Scores and Lowest Saturation

	Cluster Labels	Cuisine Type	Count	Popularity Score	Type
7	2	Burger Joint	4	60.0	2
10	2	Chinese Restaurant	6	63.0	1
16	2	Diner	1	63.0	2
23	2	Food Truck	2	60.0	3
31	2	Kosher Restaurant	1	59.0	1
40	2	Sandwich Place	6	59.0	2
41	2	Seafood Restaurant	4	64.0	1
45	2	Steakhouse	1	64.0	1
46	2	Taco Place	1	53.0	2

- Based on this study, restaurant investors should consider making their investments into restaurant types represented in Cluster 2.