Introduction

New York City, one of the most vibrant financial centers of the United States, is home to thousands of restaurants—each as diverse and unique as the 8.6 million people people who live there. Every business in the city strives to rise above competitors while meeting strict city codes and laws. Such a contentious environments demands that potential restaurant owners thoroughly consider many details before opening their businesses. Considerations include a myriad of things to consider: cuisine, location, atmosphere, and customer-base.

Business Problem

As outlined above, future New York restaurant owners compete in a highly contentious business environment. Therefore, this analysis seeks to capture and present accurate picture of some of the most successful restaurants in New York City.

Business Problem (Continued)

This review will provide an in-depth examination of the following factors to enable the target audience with all the data needed to confidently make an informed decision:
- Demographics of New York
- Population distributions
- Neighborhood statistics
- Avoidance areas, such as those cornered by competitors
- Independent markets, such as open markets and farmer's co-ops
- Local attractions including malls, tourist attractions, theaters, etc.

Target Audience

The target audience for this project primarily represents three types of people, entrepreneurs, professional restaurant staff, and investors. This analysis will use modern data science principles to provide a well-researched recommendation to each of these groups.

Data Sources

Though this review utilizes multiple sources, all data will focus on New York City:
Geographical data comes from GPS-coordinates.org and NYC.gov
    https://gps-coordinates.org/new-york-city-latitude.php
    https://www1.nyc.gov/site/planning/zoning/districts-tools/residence-districts-r1-r10.page
Primarily our base data search began on Wikipedia, at the following sites
    https://en.wikipedia.org/wiki/New_York_City
    https://en.wikipedia.org/wiki/List_of_restaurants_in_New_York_City
    https://en.wikipedia.org/wiki/New_York_City#Cuisine
    https://en.wikipedia.org/wiki/Economy_of_New_York_City
    https://en.wikipedia.org/wiki/New_York_City#Streets_and_highways
Derived data covers over 300 neighborhoods and 5 boroughs
    https://geo.nyu.edu/?f%5Bdc_subject_sm%5D%5B%5D=Neighborhoods
Restaurant is provided by City-data.com
    https://www.city-data.com/city/New-York-New-York.html
Locations of sidewalk cafes and open-air eating establishment
    https://data.cityofnewyork.us/City-Government/Sidewalk-Caf-Regulations-GIS-Shapefile/qsuf-mgjh

Methodology

This analysis uses mapping and table data to depict very detailed data visually. Data was transformed with Pandas dataframe, then looped though the dataframe to render readable results. I then verified that predicted results were matching expeccted data. Trained data and test data matched appropriately. Then, I created a map overlay showing New York Neighborhoods. I used web scraping techniques, via Python/Beautiful Soup, to glean population and demographic data. The project reqired several rounds of web scraping. To clean data, I had to remove white spaces and change the name of some columns to make them usable. When this portion of the project was complete, I saved the dataframe to a .CSV file. Then, I downloaded and analyzed data related to New York cuisine. Next I segmented and clustered neighborhoods with Foursquare API. Then, I used k-means to cluster neighborhoods. Lastly, I configured parameters to visualize the data in this report.

Results – Word Cloud Analysis

Overal Most Favoured Foods in New York City
- Italian   - Purto Rican   - Mexican   - Jewish
- Indian   - Pakistani   - Dominican

Overal Most Favored Foods in Brooklyn
- Italian   - Purto Rican   - Mexican

Overal Most Favored Foods in the Bronx
- Italian   - Purto Rican   - Albanian   - Dominican

Overal Most Favored Foods in Queens
- Italian   - Purto Rican   - Albanian   - Dominican

Overal Most Favored Foods the Staten Island
- Italian   - Mexican   - Chinese   - Indain

Conclusion

This analysis uses numerous tools to locate, gather, clean, and visualize the data from a broad range of data sources. However, this report also does a fair job of predict the success of specific cuisines, based on the patronage of customers visiting those areas.