IBM Data Science Course Capstone Project

Linh (Lilly) Nguyen

Problems

To boost economic development, NYC city planners want to help current restaurant owners run their business smoothly while attracting more new investors to start a new restaurant business in the city. Two topics of their concern are (1) the **market competition** and (2) the **sanity regulation**.

The project aims to address the following questions:

1. What is the problem with the restaurant having a low sanity rating? What should the restaurant do to maintain a good sanity rating?
2. Do food categories impact the customers' favorability towards a restaurant? If yes, Which food category is the most popular one? Which are the local customers' favorite food categories?
3. Which neighborhood area in New York City is the best for launching a new restaurant business?

Data Description

Two datasets from the following two main sources:

1. **FourSquare API** which provides the surrounding venues of New York City coordinate (40.7128 N, 74.0060 W) and its "likes"

2. The New York City Health Department's Restaurant Inspection Results in 2020 https://data.cityofnewyork.us/Health/DOHMH-New-York-City-Restaurant-Inspection-Results/43nn-pn8j

Methodologies

Data Cleaning

Descriptive Analytics

Data Visualization

Predictive Analytics (Regression Models)

NYC Health Department Sanitary Inspection

Each violation has some point value so a score of 0 means there are no violations. The final grade is based on the sum of all the points.

- 0 to 13 earns an A
- 14 to 27 earns a B
- 28 or more earns a C

Results – The Restaurant Market in NYC

Among the five boroughs, **Manhattan** has the largest number of restaurants (40%), followed by **Brooklyn** (25.5%) and **Queens** (22.9%).

American, Asia-Pacific & Casual dinings is the most popular kind of restaurant in the New York City area.
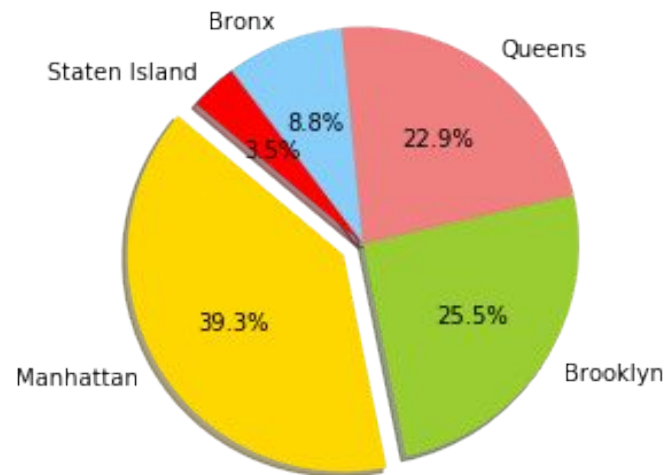
European and **Bar** are the most favorite kinds of restaurants in New York City.
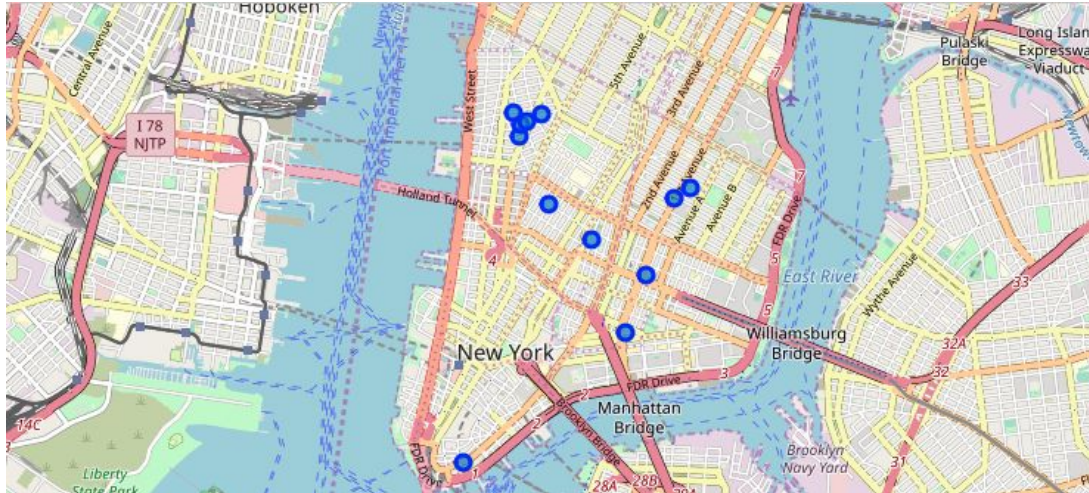
Top most favorite restaurants are located in **West Village, East Village** and **Lower Manhattan.**

Results – The sanitary problem

Although 80% of the restaurants receive grade A rating, and many of them failed to meet the grade A quality in the first sanity inspection.

Restaurants in **Lower Manhattan** and **Northeast Bronx** are the two neighborhoods that have lowest sanity grades in the city.

The common sanity issues are food contact with dirty surfaces or dirty hands.
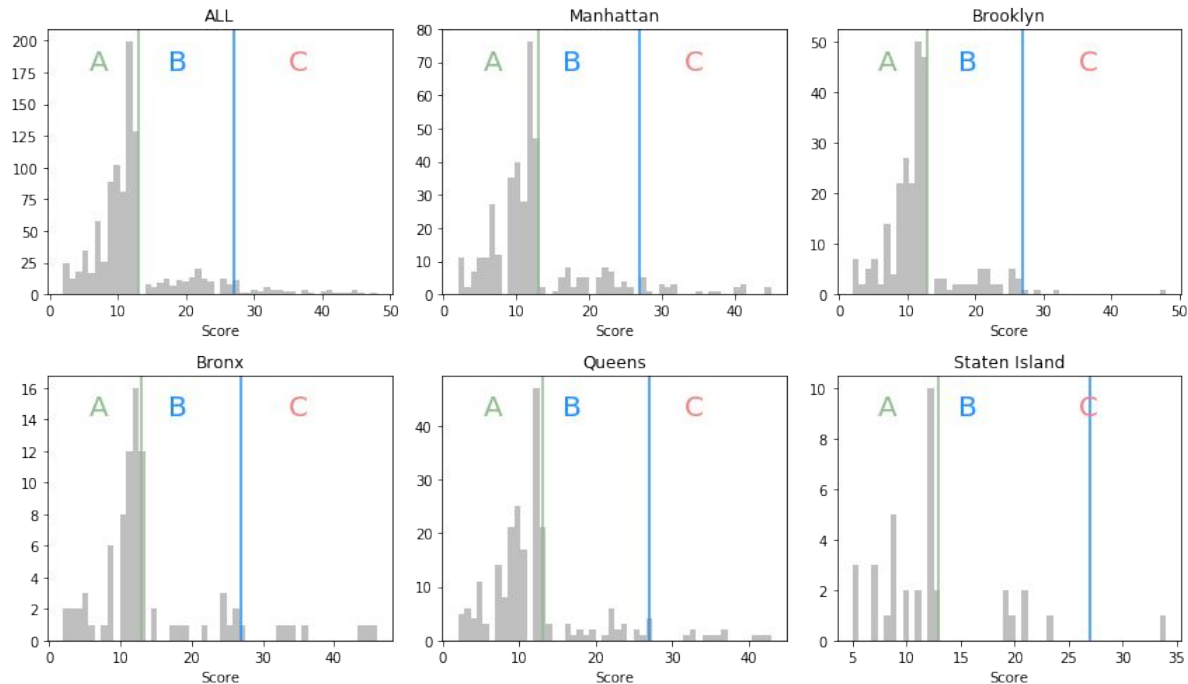
Results – Regression Models

Dataset 1

```
                          OLS Regression Results
================================================================================
Dep. Variable:                    y   R-squared:                      0.673
Model:                          OLS   Adj. R-squared:                 0.671
Method:               Least Squares   F-statistic:                    268.0
Date:              Sat, 28 Mar 2020   Prob (F-statistic):          1.09e-185
Time:                      17:45:53   Log-Likelihood:               -2806.0
No. Observations:               786   AIC:                            5624.
Df Residuals:                   780   BIC:                            5652.
Df Model:                         6
Covariance Type:          nonrobust
================================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
--------------------------------------------------------------------------------
x1            12.2632      0.660     18.589      0.000      10.968      13.558
x2            13.4774      0.693     19.450      0.000      12.117      14.838
x3            11.3988      0.676     16.870      0.000      10.072      12.725
x4            13.1930      0.808     16.329      0.000      11.607      14.779
x5            15.0337      0.914     16.441      0.000      13.239      16.829
x6            14.2174      1.799      7.904      0.000      10.686      17.748
================================================================================
Omnibus:                    554.474   Durbin-Watson:                  1.884
Prob(Omnibus):                0.000   Jarque-Bera (JB):           10259.938
Skew:                         2.963   Prob(JB):                        0.00
Kurtosis:                    19.679   Cond. No.                        2.73
================================================================================
```

Dataset 2

```
                          OLS Regression Results
================================================================================
Dep. Variable:                    y   R-squared:                      0.647
Model:                          OLS   Adj. R-squared:                 0.598
Method:               Least Squares   F-statistic:                    13.42
Date:              Sat, 28 Mar 2020   Prob (F-statistic):          3.41e-05
Time:                      17:52:25   Log-Likelihood:               -191.73
No. Observations:                25   AIC:                            389.5
Df Residuals:                    22   BIC:                            393.1
Df Model:                         3
Covariance Type:          nonrobust
================================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
--------------------------------------------------------------------------------
x1          1158.0000    552.402      2.096      0.048      12.388    2303.612
x2          1239.5000    390.607      3.173      0.004     429.430    2049.570
x3          1254.8000    247.042      5.079      0.000     742.467    1767.133
================================================================================
Omnibus:                      1.820   Durbin-Watson:                  1.629
Prob(Omnibus):                0.403   Jarque-Bera (JB):               0.609
Skew:                        -0.162   Prob(JB):                       0.737
Kurtosis:                     3.693   Cond. No.                       2.24
================================================================================
```

- Customers like food from American restaurants, European restaurants and bars more than other kinds of restaurants.

- One explanation can be that the owners of American, European and bars often have large investment funds that allow them to spend more on improving the sanitary quality.

- Latin restaurants are often the group of restaurants that often violate the sanitary regulation. The owners of this group often do not have much investment fund for their business compared to the Western restaurant owner counterparts

Outcomes & Strategy Recommendations

- Opportunity for investor to open a new restaurant business with an European-style restaurants or bars, especially in the "reputated" restaurant districts in East Village and West Village

- Opportunity to turn Lower Manhattan become the new "restaurant district". However, we need to improve the sanitary quality in this area.

- City planners can provide supporting funds or educational programs to help local restaurants in Lower Manhattan improve sanitary quality, especially providing methods to avoid food contact with dirty surfaces

Sources

Forbes:
https://www.forbes.com/sites/garyocchiogrosso/2019/12/20/the-new-york-city-restaurant-business-is-so-much-more-than-just-the-center-of-the-plate/#138b04a7639c

New York City's Sanitary Inspection:
http://www1.nyc.gov/assets/doh/downloads/pdf/rii/how-we-score-grade.pdf

Data set: https://data.cityofnewyork.us/Health/DOHMH-New-York-City-Restaurant-Inspection-Results/43nn-pn8j

Data Dictionary:
https://data.cityofnewyork.us/api/views/43nn-pn8j/files/ec33d2c8-81f5-499a-a238-0213a38239cd?download=true&filename=RestaurantInspectionDataDictionary_09242018.xlsx