**New York- Restaurant Inspection Score/Grade Prediction**

**Background**

**NY Restaurant Grade:**

This dataset includes NYC restaurant and college cafeteria (hereafter, restaurants) inspection results for up to three years prior to the most recent inspection. The purpose is to provide information on recent inspection results. Restaurants that go out of business are removed. Therefore, this dataset is not appropriate for historical analyses of NYC restaurant inspections that compare previous years of data to the current data. In addition, restaurants can choose to go through the adjudication process, i.e., argue their case at an administrative hearing. Restaurants also have appeal rights, and the entire adjudication process from start to finish can take several months. Scores current as of today may be revised due to adjudication in subsequent weeks or months. The change in scores due to adjudication is another reason why it is not valid to compare current scores to scores from previous years. More Details for Restaurant inspection process and grade are available at <https://www1.nyc.gov/site/doh/services/restaurant-grades.page>

**Forward**

Getting Restaurant Inspection and Grading is a lengthy process and process involves charges and effort, Restaurant management always analyzes and put effort for grading process. Since this involves money, restaurant management want to predict their inspection score prior to the submission,

Objective of this project is to predict the Restaurant (grade)

These studies are intended to the representative of DOHMH for reference, New York restaurant owners/managers and their staff. to use this analysis to predict their restaurant inspection (grade). We are also expected the public to display their favorite restaurant’s grade,

**Executive Summary**

**Problem Statement**

New York is the one of the biggest cities in the east coast of the United States of America. There are thousands of different kinds of food outlets and restaurants are available. Grading of food market, restaurants are very challenging for the department of public health. Different type of cuisines is available in NY, so grading of restaurant are very difficult. Restaurant management need to complete all the pre-requests for the grade evaluation, this tool will help them to predict their restaurant score prior to application of Restaurant Inspection, this prediction will help restaurant management to target better grade

**Data Acquisition and Cleaning**

**Restaurant Inspection Score**

Restaurant Inspection process - <https://www1.nyc.gov/site/doh/services/restaurant-grades.page>

**NY Open Data**

[NY OpenData](https://opendata.cityofnewyork.us/) this source provides all the Open Data about NY;

Open Data is free public data published by New York City agencies and other partners.

**DOHMS**

[Restaurant Score - Graded By DOHMS](https://data.cityofnewyork.us/Health/DOHMH-New-York-City-Restaurant-Inspection-Results/43nn-pn8j) This dataset contains every sustained or not yet adjudicated violation citation from every full or special program inspection conducted up to three years prior to the most recent inspection for restaurants and college cafeterias in an active status on the RECORD DATE (date of the data pull). When an inspection results in more than one violation, values for associated fields are repeated for each additional violation record. Establishments are uniquely identified by their CAMIS (record ID) number. Keep in mind that thousands of restaurants start business and go out of business every year; only restaurants in an active status are included in the dataset.

Records are also included for each restaurant that has applied for a permit but has not yet been inspected and for inspections resulting in no violations. Establishments with inspection date of 1/1/1900 are new establishments that have not yet received an inspection. Restaurants that received no violations are represented by a single row and coded as having no violations using the ACTION field.

Because this dataset is compiled from several large administrative data systems, it contains some illogical values that could be a result of data entry or transfer errors. Data may also be missing. Our data collection and cleaning process will ignore non relevant and missing data,

**Data Cleaning**

In this data set record date is all are identical hence dropping that data

Dates Data type is changing to datetime

Clean text data by removing space and making it lower case to make it easier to manipulate data

Convert datatype of `phone` and `zipcode` to integer

Drop the rows which `INSPECTION DATE` is `1900-01-01`

Drop the rows with null values from our target variable SCORE

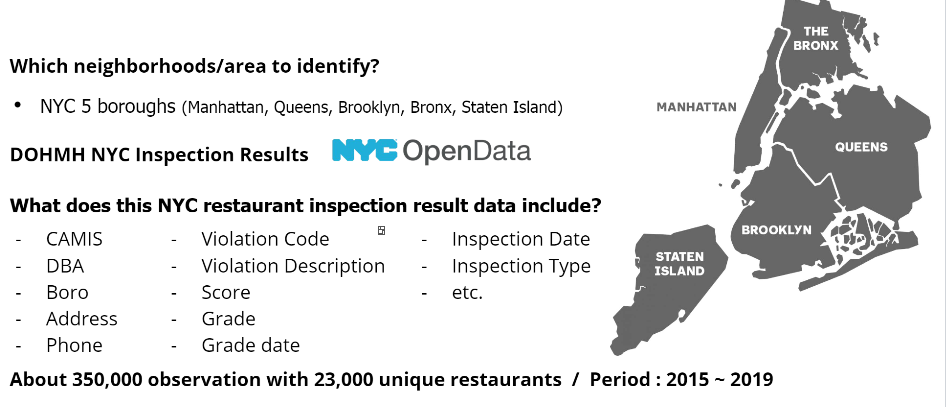
Fill in missing data in BORO columns based on the zip code

Fill n/a for VIOLATION CODE and VIOLATION DESCRIPTION

Save the cleaned data of restaurant inspection data in csv file (restaurant\_inspection\_result\_cleaned.csv)

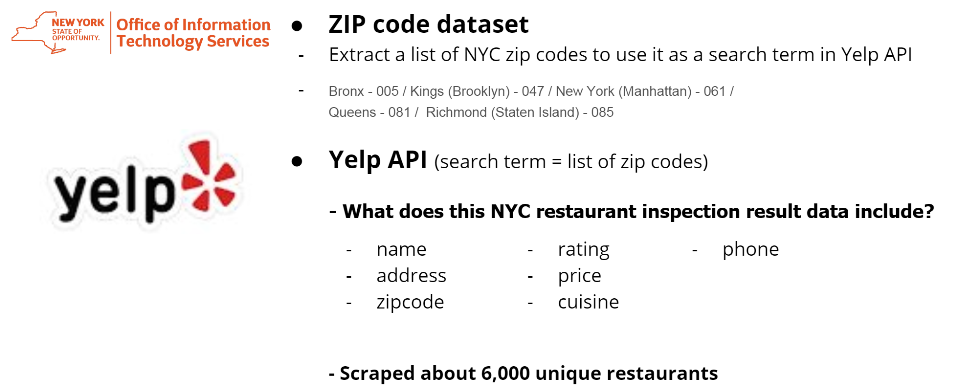
**NY CITY Location**

[NY CITY Locations](https://data.ny.gov/Government-Finance/New-York-State-ZIP-Codes-County-FIPS-Cross-Referen/juva-r6g2) this set of data will provide A listing of NYS counties with accompanying Federal Information Processing System (FIPS) and US Postal Service ZIP codes sourced from the NYS GIS Clearinghouse



**Yelp Data**

**Yelp Data**, with yelp API we are collecting all location details, this will help us to plot the data in NY MAP.



**NY Neighborhood data** -<https://geo.nyu.edu/catalog/nyu_2451_34572>

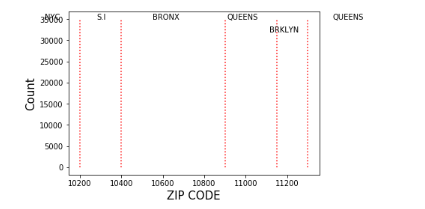
**NY Demographics Data**– <https://en.wikipedia.org/wiki/New_York_City>

**NY population Data** - <https://en.wikipedia.org/wiki/New_York_City>

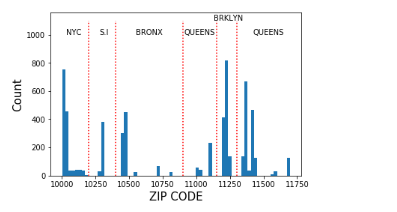
**NY cusine Data** – <https://en.wikipedia.org/wiki/Cuisine_of_New_York_City>

**Exploratory Analysis**

**ZIP Code Data Distribution (original)**



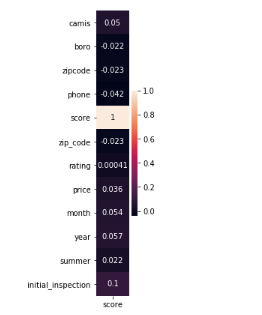
**ZIP Code Data Distribution with Yelp Data**



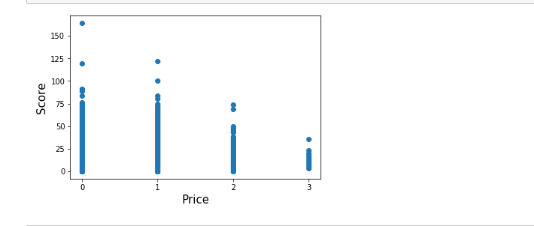
**Cleaned Data with 128275 records and extracted 6002 MY restaurant data**.

data\_collection.ipynb

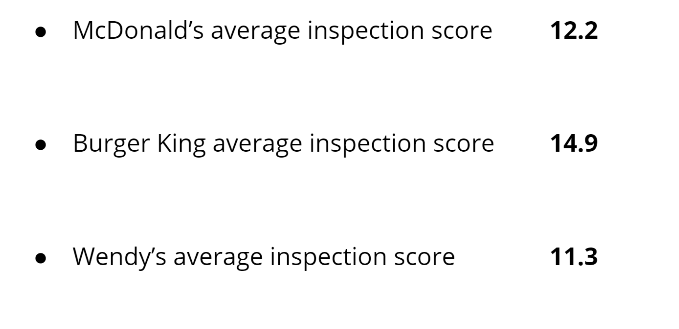
**Heat Map of Restaurant Inspection Score**

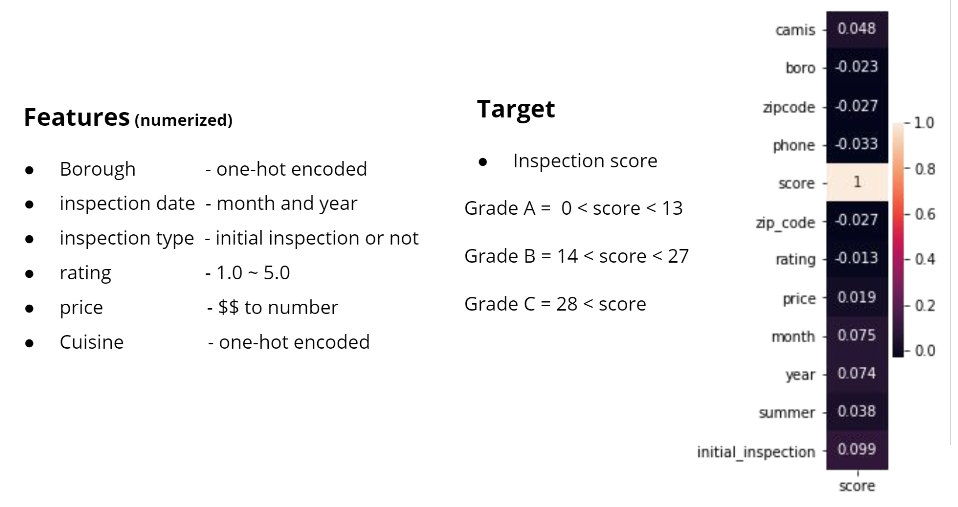
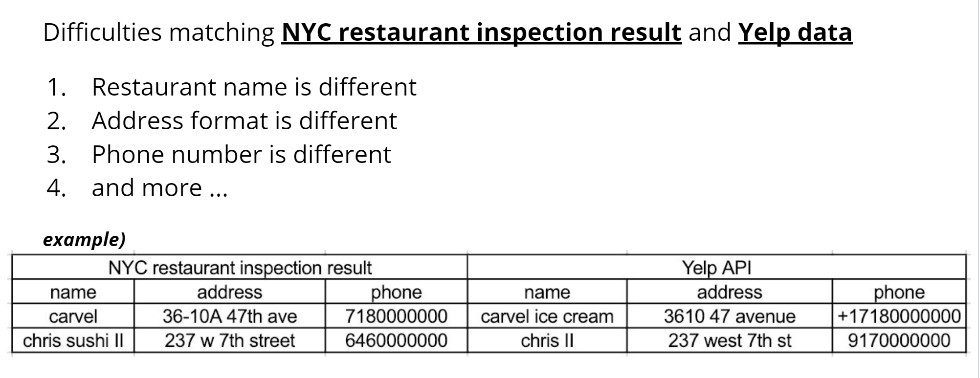


**Score Vs Price Inspection Score**

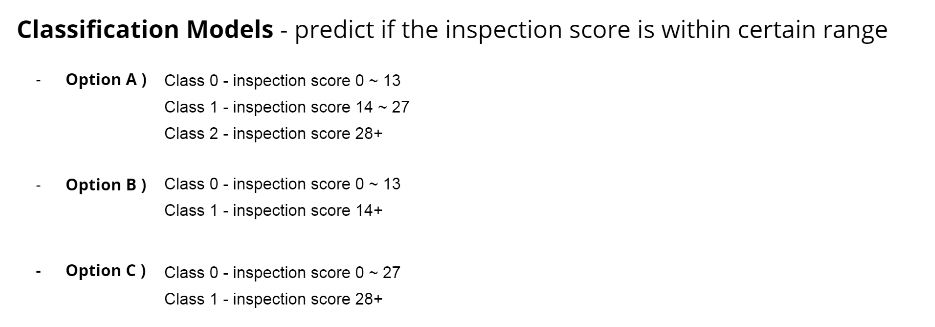


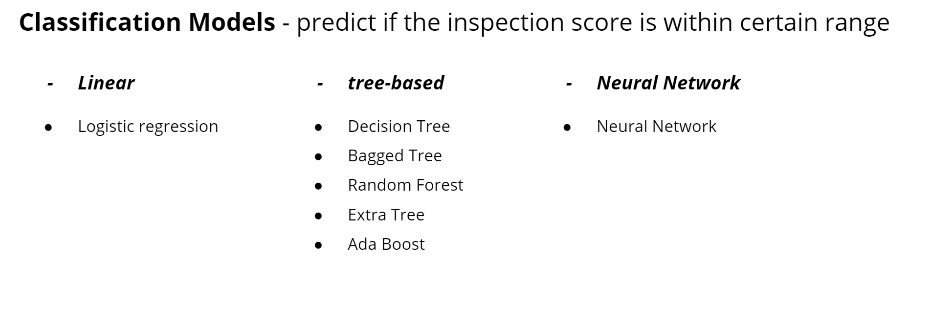
**Interesting fact of Exploratory Data Analysis**

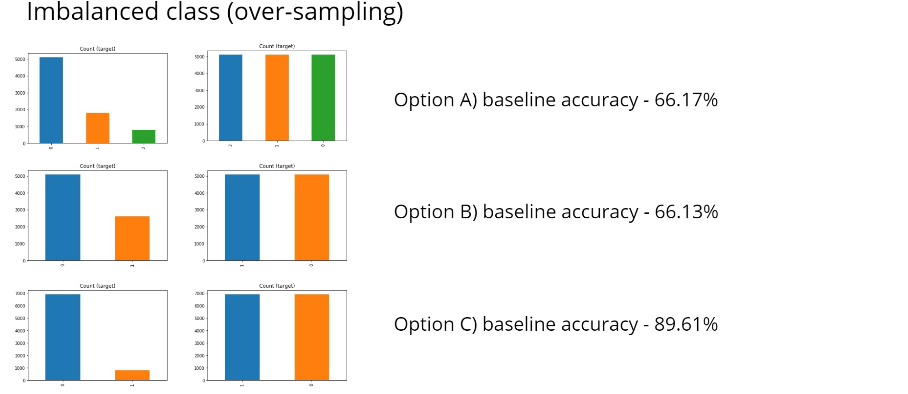


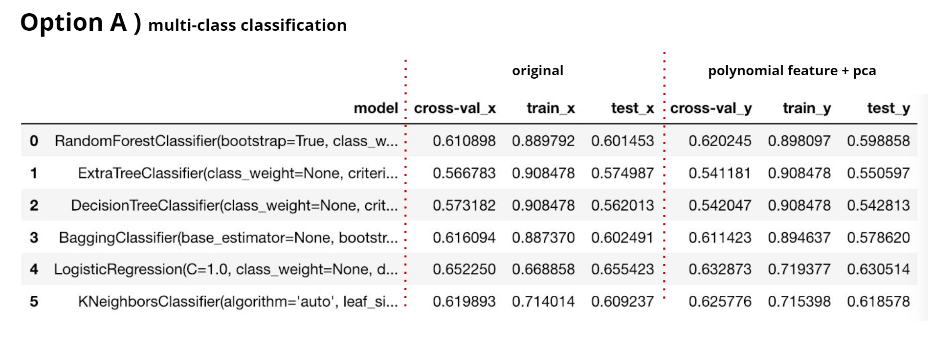
**Pre-Processing**

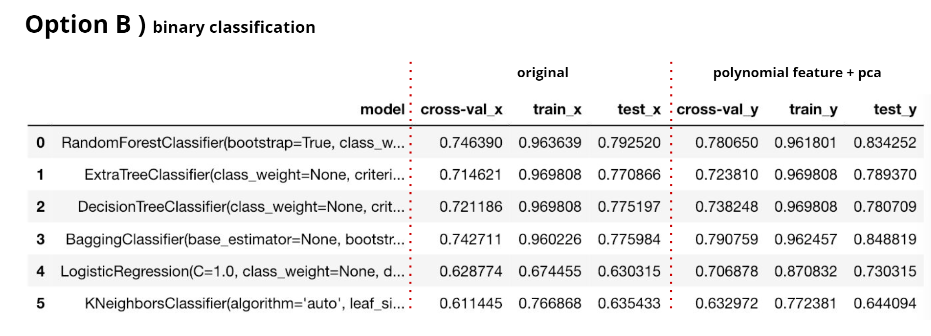
**Modeling and Evaluation**

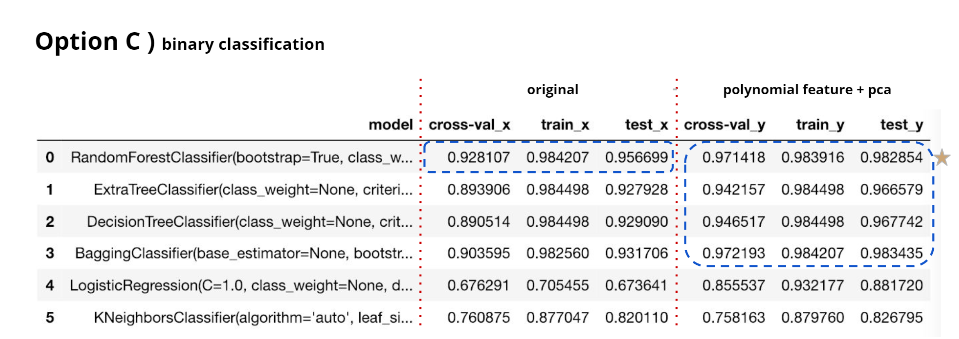


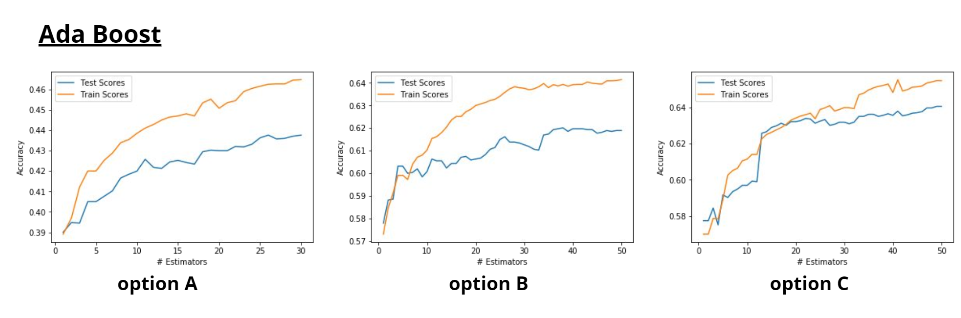


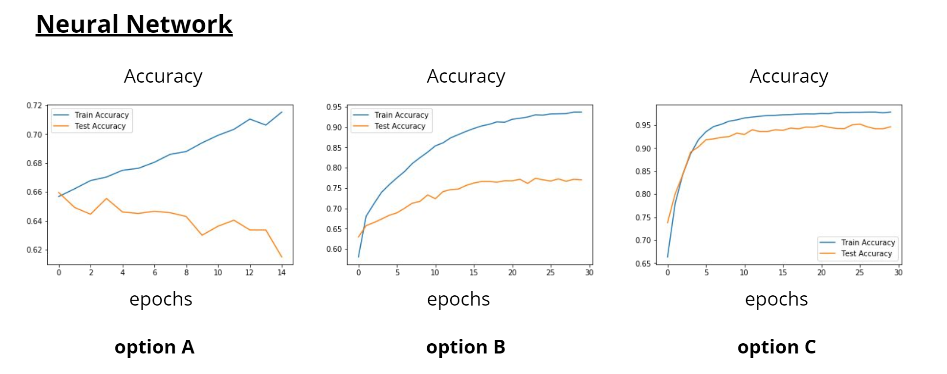


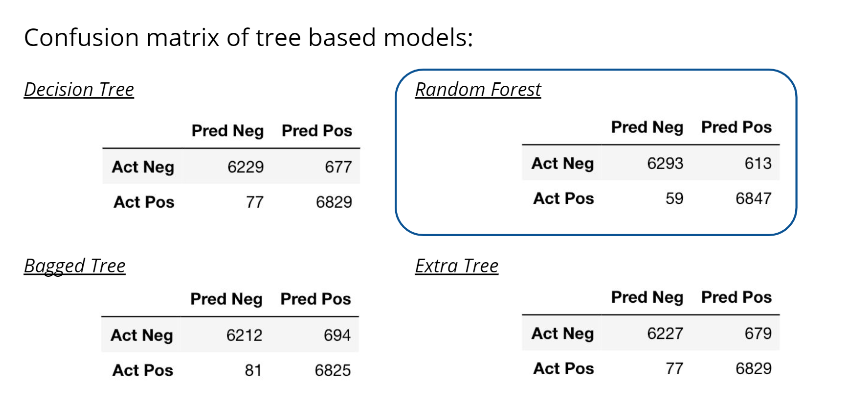




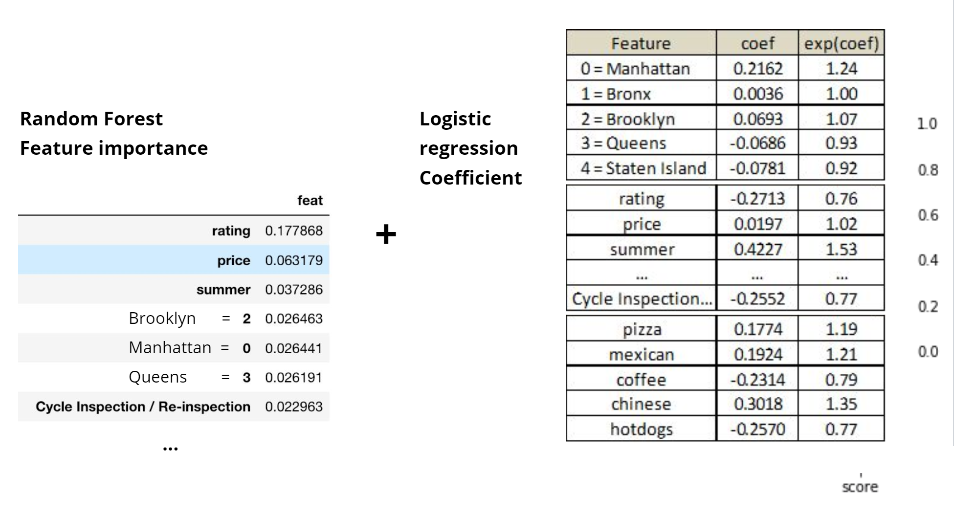








**Recommendation**



**Conclusion and Recommendation**

