Data Science & Business Intelligence Project Proposal

Team 4: Janvi Shah, Ji Yoon Kim, Monica Sarphare, Zack He, Donald Akpovwa

**Project Title:** Predicting Whether a NYC Restaurant will Pass or Fail Inspection

**Data Retrieved from:** Kaggle; NYC Restaurant Inspections: <https://www.kaggle.com/new-york-city/nyc-inspections>

**Potential Data Fields:**

* CAMIS: This is a unique identifier for the restaurant.
* DBA: This field represents the name of the restaurant.
* BORO: Borough in which the restaurant is located. NOTE: There may be discrepancies between zip code and listed boro due to differences in an establishment's mailing address and physical location.
* BUILDING, STREET, ZIPCODE, PHONE: Address information for the restaurant.
* CUISINE DESCRIPTION: This field describes the restaurant cuisine.
* INSPECTION DATE: This field represents the date of inspection. NOTE: Inspection dates of 1/1/1900 mean an establishment has not yet had an inspection.
* ACTION: This field represents the action that is associated with each restaurant inspection.
* VIOLATION CODE: This field represents each violation associated with a restaurant inspection.
* VIOLATION DESCRIPTION: This field describes the violation codes
* CRITICAL FLAG: Critical violations are those most likely to contribute to foodborne illness.
* SCORE: Total score for a particular inspection; updated based on adjudication results.
* GRADE and GRADE DATE: This field represents the grade associated with this inspection. Grades given during a reopening inspection are derived from the previous re-inspection.
* RECORD DATE: The date when the web extract was run to produce this data set
* INSEPECTION TYPE: A combination of the inspection program and the type of inspection performed.

**Business Problem Addressed:** As we emerge from the COVID-19 pandemic, the restaurant industry is likely to experience a boom. Investors will seek restaurants to invest in, and entrepreneurs will look to establish new restaurants. New York City (NYC) is a prime location for this activity. Understanding the historical interaction between restaurants and the Department of Health (DOH) will help individuals decide whether to invest in, or establish, restaurants in NYC. We will prepare a guide to what cuisines, what zip codes, and what types of restaurants have been cited or closed by the DOH.

**Use Scenario of Result:** We will analyze how cuisine type and location, among other factors, affect the DOH’s assessment of restaurants. In addition, we hope to identify if in a particular location, the DOH closes or negatively rates a particular cuisine more than others. We will also identify how chain restaurants fare in restaurant inspections compared to non-chain restaurants.

**Data Instance and Useful Features:** The main data variables we will include in our project are BORO, STREET, ZIPCODE, CUISINE DESCRIPTION, ACTION, VIOLATION CODE, SCORE, GRADE, INSPECTION TYPE, etc.

**Target Variable:** We will use GRADE as the indicator to see how restaurants performed in the inspection.

**Added Business Value:** The added business value is for potential investors and entrepreneurs. They gain more knowledge of factors that potentially affect their business, like whether DOH closes a particular cuisine more frequently in a certain area.