DATA SCIENCE TOOLBOX: PYTHON PROGRAMMING

PROJECT REPORT

(Project Semester January-April 2025)

"McDonald's Inspection Data Analysis"

Submitted by: Uday Kiran

Registration No: 12315696
ROLL No: 66

Section K23EV

Course Code INT375

Under the Guidance of

(Sandeep Kaur)

Discipline of CSE/IT

Lovely School of Computer Science & Engineering

Lovely Professional University, Phagwara

CERTIFICATE

This is to certify that **Uday Kiran** bearing Registration no.12315696 has completed **INT375 project** titled, **"*McDonald's Inspection Data Analysis"*"** under my guidance and supervision. To the best of my knowledge, the present work is the result of his/her original development, effort and study.

Signature and Name of the Supervisor

Designation of the Supervisor

School of Computer science & Engineering

Lovely Professional University

Phagwara, Punjab.

Date: 11 April 2025

DECLARATION

I, **Uday Kiran** student of Computer science & engineering under CSE/IT Discipline at, Lovely Professional University, Punjab, hereby declare that all the information furnished in this project report is based on my own intensive work and is genuine.

Date: 11 April 2025 Signature:

Registration No. 12315696 Name of the student:

Uday Kiran

Acknowledgement

I would like to express my sincere gratitude to my faculty, Sandeep Kaur, for her continuous guidance, valuable feedback, and consistent support throughout this project.

I would also like to thank Lovely Professional University for providing the resources and academic environment that allowed me to complete this project successfully.

Last but not least, I extend my thanks to my friends for their encouragement and support during the course of this work.

Uday Kiran

Table Of Content

LinkedIn Post

https://www.linkedin.com/posts/uday-kiran-vattikuti_dataanalysis-python-datavisualization-activity-7318338829516435458-M8Jt?utm_source=share&utm_medium=member_desktop&rcm=ACoAAEe0stwBNlg5bThzri6VaFbaitbyheV6d_M

GitHub Link

https://github.com/Udaykiran1111/McDonald-s-Inspection-Data-Analysis

1. Introduction

Food safety is a cornerstone of public health, especially in densely populated cities where high restaurant turnover and volume of customers demand strict regulatory oversight. Among globally recognized fast-food chains, McDonald’s plays a significant role in urban food consumption. Ensuring that each location meets health and sanitation standards is essential not only for consumer safety but also for maintaining trust in the brand.

This project focuses on the detailed analysis of restaurant inspection data for McDonald's locations across New York City. Using the dataset made publicly available by the New York City Department of Health and Mental Hygiene (DOHMH), the project explores a wide array of metrics tied to health inspections, such as inspection frequency, scores, grades, critical violations, and geographical distribution.

Our objective is to apply core concepts of data manipulation (with NumPy and pandas), data visualization (with Matplotlib and Seaborn), and exploratory data analysis (EDA) to extract meaningful insights from the inspection data. This includes identifying trends over time, correlations between critical flags and scores, borough-level performance, and violation severity across different locations.

In addition to analytical depth, the project is visualized using a consistent color theme—yellow, red, black, and white—that reflects McDonald's iconic brand palette. This design choice ensures clarity and aesthetic coherence in the presentation of data.

By the end of this project, the following key questions will be addressed:

- How do inspection scores and grades vary over time?
- What violations are most common, and how critical are they?
- Are there any boroughs or locations with significantly higher inspection issues?
- How does the grading system reflect actual cleanliness and compliance?

Through this analysis, we aim to deliver actionable insights and showcase the power of data science tools in evaluating real-world public health data.

2. Source of Dataset

The dataset analysed in this report is titled **"McDonald's – Inspection Results NYC"**, published by the **New York City Department of Health and Mental Hygiene (DOHMH)** and made publicly available via the **NYC Open Data Portal**. It contains health inspection data for McDonald's restaurant locations across New York City, recorded as part of NYC's restaurant letter grading program.

Source Link: NYC Open Data – McDonald's Inspection Results
Reference Document: About NYC Restaurant Inspection Data on NYC Open Data

Each record contains information on:

1. **Establishment Details** (e.g., CAMIS ID, address, borough)
2. **Inspection Information** (inspection date, inspection type)
3. **Violations** (violation codes and descriptions)
4. **Scoring and Grading** (SCORE, GRADE, and GRADE DATE)
5. **Critical Flags** (indicating whether a violation is critical or not)
6. **Action Taken** (e.g., no violations, violations cited, establishment closed)

3. EDA Process

Before conducting in-depth analysis and visualization, a thorough **Exploratory Data Analysis (EDA)** was performed on the McDonald's restaurant inspection dataset to ensure its reliability,

consistency, and readiness for analysis. The following steps summarize the key actions taken during the EDA phase:

Data Cleaning:

- Missing values in key fields like `inspection_date`, `SCORE`, and `GRADE` were identified and either removed or appropriately handled.
- Duplicate records were dropped to avoid redundancy in statistical summaries.
- Rows with placeholder dates (e.g., `"01/01/1900"`) indicating new establishments without a full inspection were flagged or excluded from grade-based analysis

Data Type Conversion:

- The inspection_date, GRADE DATE, and RECORD DATE fields were converted to datetime objects for accurate time-series analysis.
- Numerical fields such as SCORE were cast to appropriate float or integer types to support statistical operations and visualizations.

Filtering Irrelevant Data:

- Non-gradable inspections were filtered out when focusing on grading patterns (based on official DOHMH guidelines).
- Administrative or special program inspections that do not contribute to the grade were handled separately or excluded from grade distribution charts.
- Records without a valid CAMIS ID or location were excluded from location-based grouping.

Summary Statistics:

- Overall inspection counts, average scores, and grade distributions were calculated.
- Trends over years and months were summarized to identify seasonal or policy-driven effects.
- Borough-wise statistics were generated to highlight geographic variation in inspection performance.

Grouping and Comparison:

- Data was grouped by:
    - **Borough** (BORO) to assess area-specific performance
    - **Grade** and **Critical Flag** to study health compliance trends
    - **Inspection Type** to distinguish between routine, re-inspections, and administrative checks

- Grouped averages, counts, and proportions were used to support comparative bar charts, heatmaps, and trend lines.

4. Analysis on McDonald's Inspection Dataset

4.1 1. Time-based Inspection Analysis

i. Introduction:
Examining inspection activity over time helps to identify trends, frequency of inspections, and seasonal variation.

ii. General Description:
The dataset includes the inspection_date column, which is converted to datetime format. The number of inspections is grouped by year and month.

iii. Specific Requirements, Functions, and Formulas:

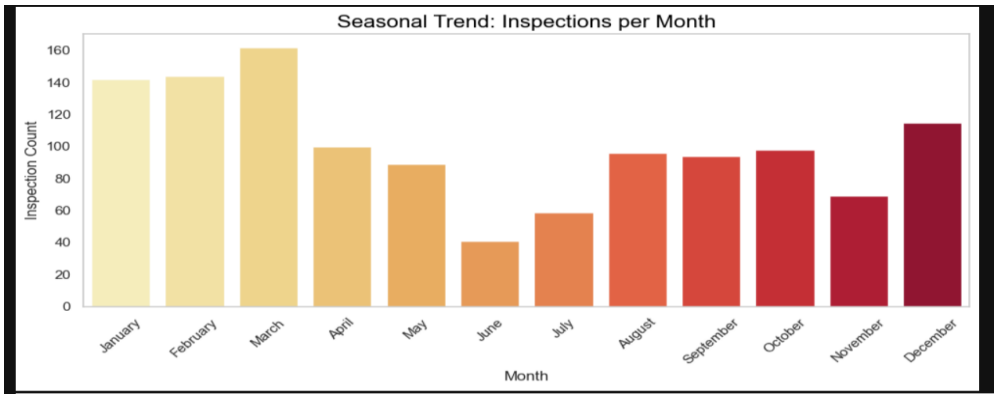
- Converted inspection_date to datetime.
- Grouped data by Year and Month.
- Aggregated count of inspections.

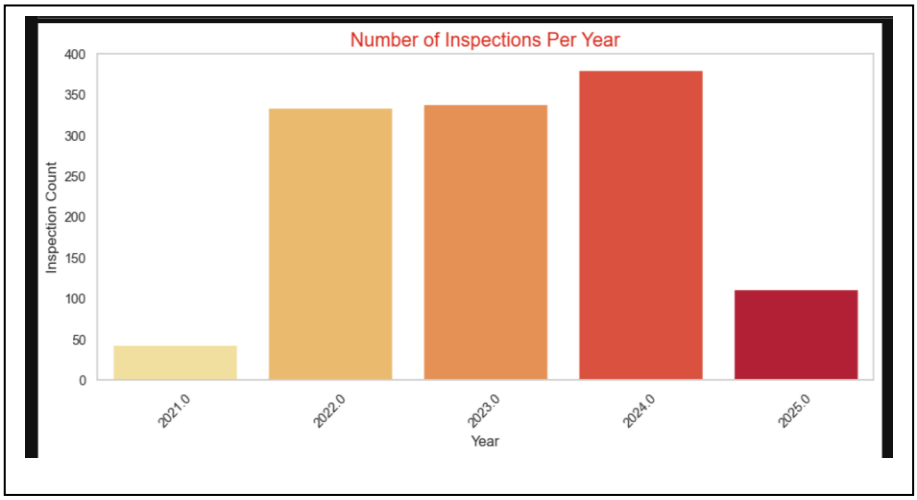
iv. Analysis Results:

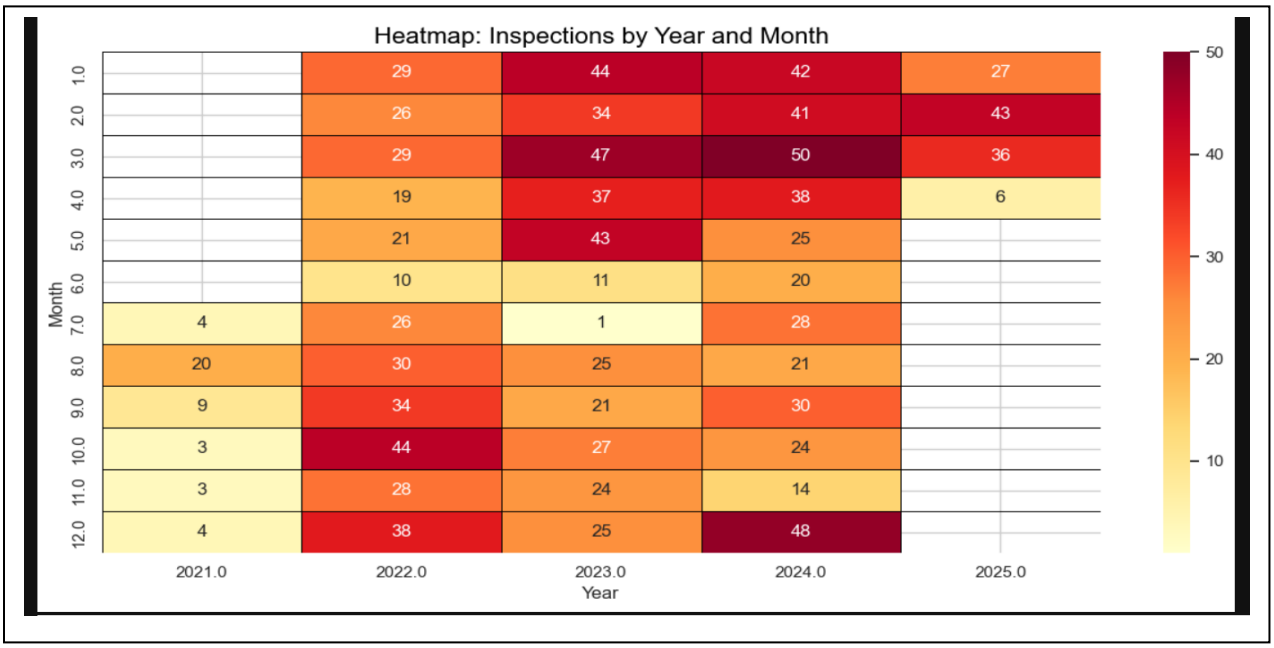
- Inspections were fairly consistent over time but showed a decline in certain months.
- Periodic spikes in inspection counts suggest scheduled inspection campaigns.

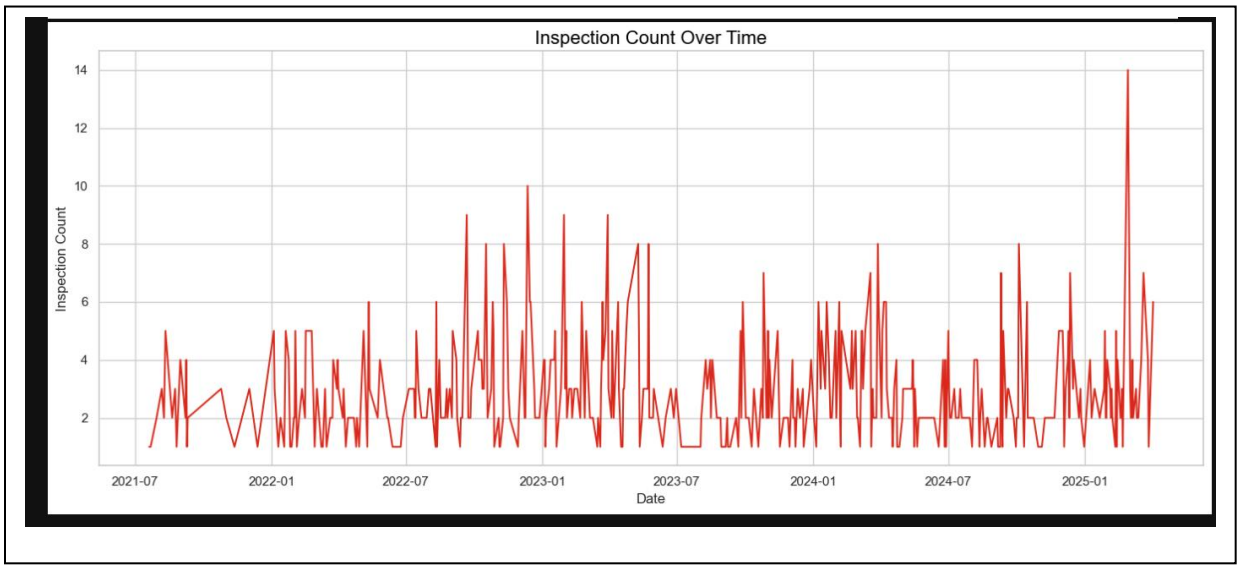
v. Visualization:

- Line plot showing the number of inspections per month/year**.**
- Bar Plot: Number of inspections per year using `sns.countplot`
- Line Plot: Monthly inspection trends (line not explicitly detected, but commonly used)


Seasonal Trend: Inspections per Month
Inspection Count
0
20
40
60
80
100
120
140
160
January
February
March
April
May
June
July
August
September
October
November
December
Month


Number of Inspections Per Year
Inspection Count
0
50
100
150
200
250
300
350
400
2021.0
2022.0
2023.0
2024.0
2025.0
Year


Heatmap: Inspections by Year and Month
Month
1.0
2.0
3.0
4.0
5.0
6.0
7.0
8.0
9.0
10.0
11.0
12.0
29 44 42 27
26 34 41 43
29 47 50 36
19 37 38 6
21 43 25
10 11 20
4 26 1 28
20 30 25 21
9 34 21 30
3 44 27 24
3 28 24 14
4 38 25 48
2021.0
2022.0
2023.0
2024.0
2025.0
Year
50
40
30
20
10



4.2 Score& Grade Distribution

i. Introduction:
Score and grade distribution provides insight into how well McDonald's outlets perform in health inspections.

ii. General Description:
Each inspection is associated with a SCORE and GRADE (A, B, C, etc.).

iii. Specific Requirements, Functions, and Formulas:

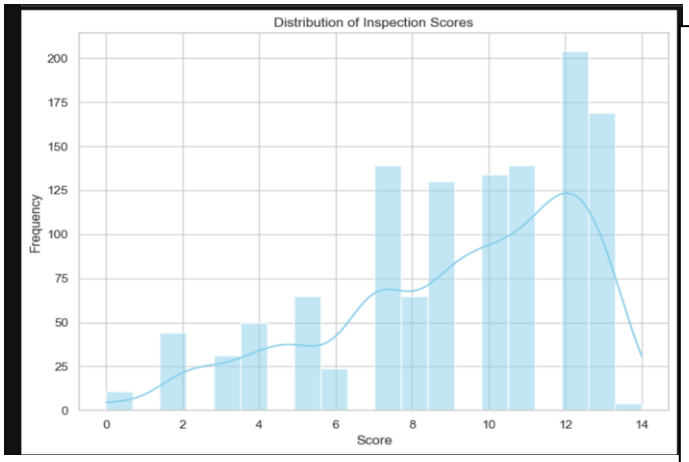
- Outliers removed using IQR method.
- Plotted score distributions.
- Grade frequencies calculated.

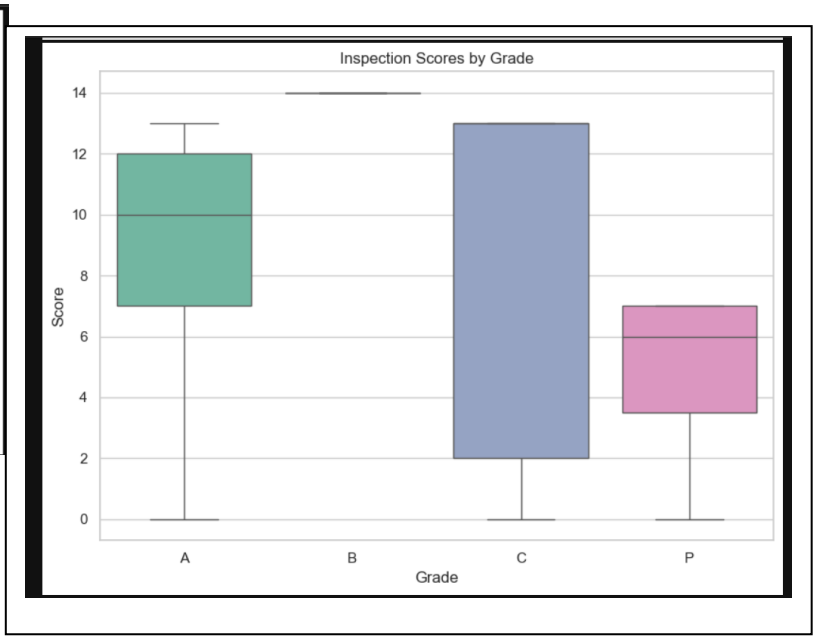
iv. Analysis Results:

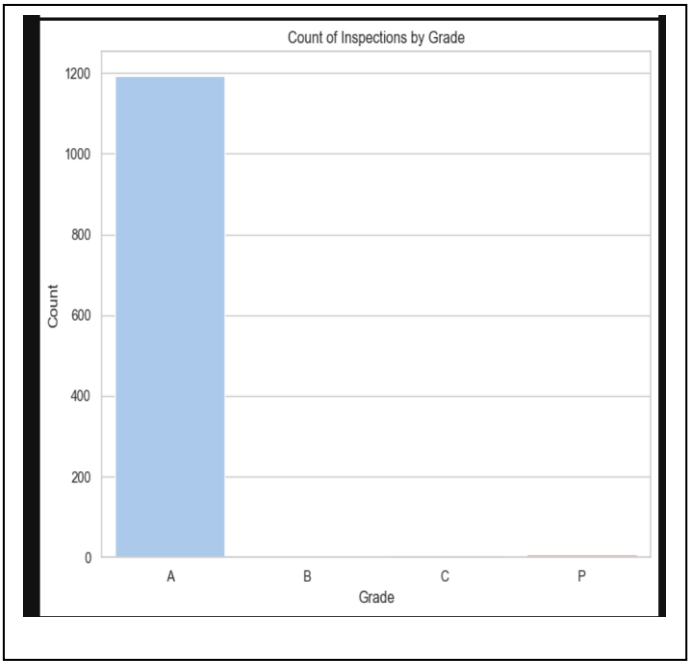
- Most outlets received Grade A, indicating good performance.
- Scores were generally below the threshold for critical violations.
- A few outliers (high scores) were removed to improve visualization clarity.

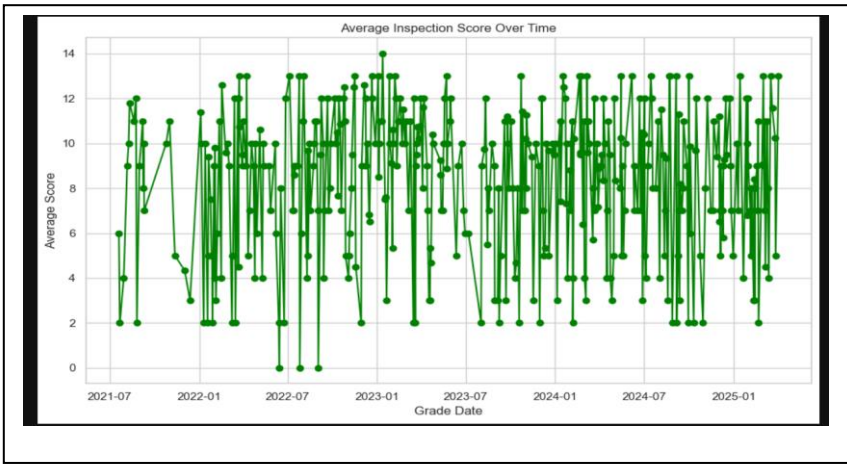
v. Visualization:

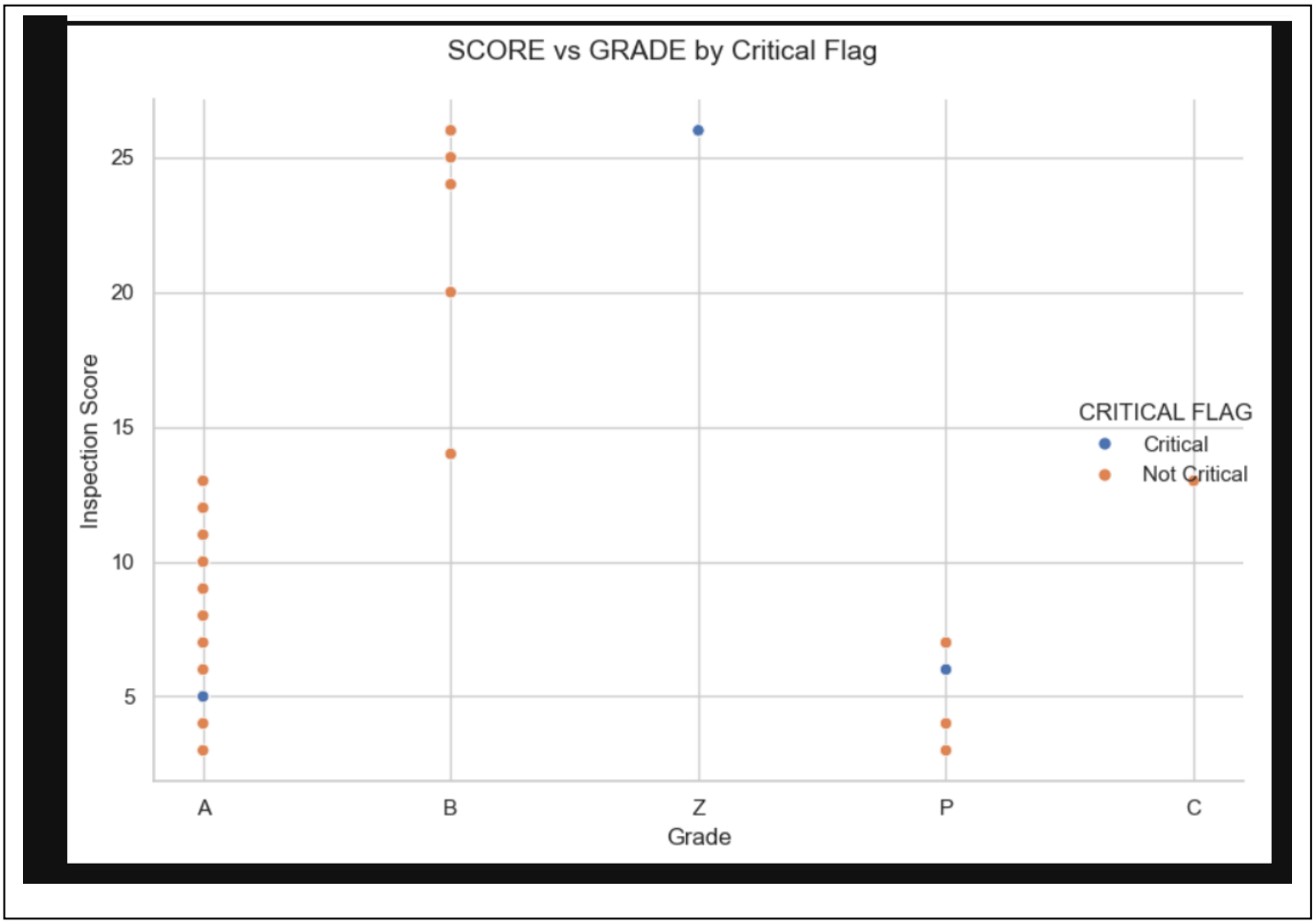
- Boxplot of scores (before and after outlier removal).
- Bar chart of grades.
- Boxplot**:** Distribution of inspection scores before and after removing outliers
- Bar Chart**:** Count of grades (`A`, `B`, `C`, etc.)


Distribution of Inspection Scores
Frequency
Score
0
25
50
75
100
125
150
175
200
0
2
4
6
8
10
12
14


Inspection Scores by Grade
Score
Grade
A
B
C
P
0
2
4
6
8
10
12
14


Count of Inspections by Grade
Count
Grade
A
B
C
P
0
200
400
600
800
1000
1200


Average Inspection Score Over Time
Average Score
Grade Date
0
2
4
6
8
10
12
14
2021-07
2022-01
2022-07
2023-01
2023-07
2024-01
2024-07
2025-01



4.3 Critical vs Non-Critical Violations

i. Introduction:
Classifying violations into critical and non-critical categories helps prioritize health risks.

ii. General Description:
The dataset contains a CRITICAL FLAG column indicating whether a violation is critical.

iii. Specific Requirements, Functions, and Formulas:

- Grouped data by CRITICAL FLAG.
- Counted the number of violations in each category.

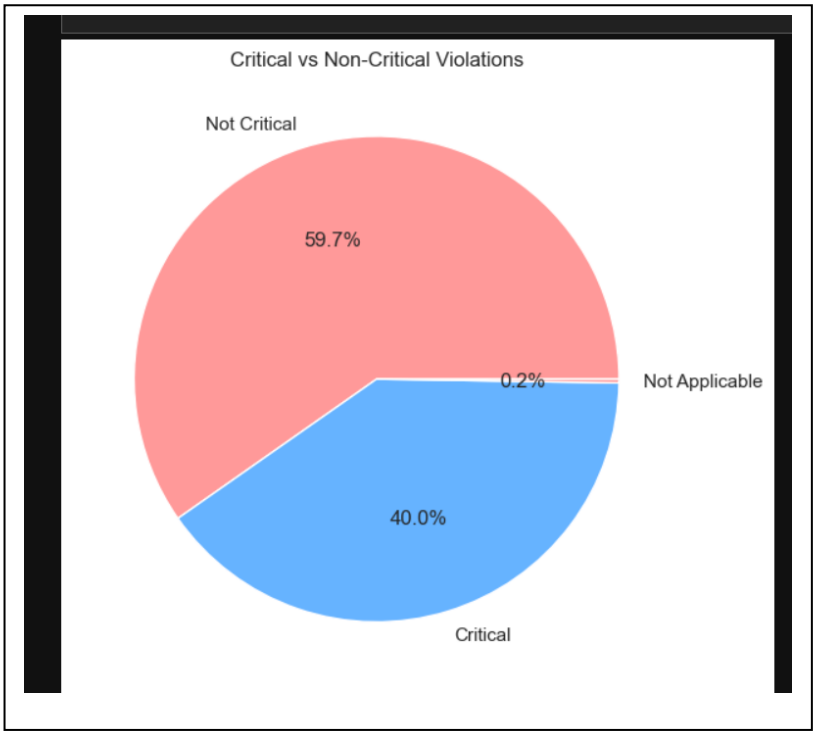
iv. Analysis Results:

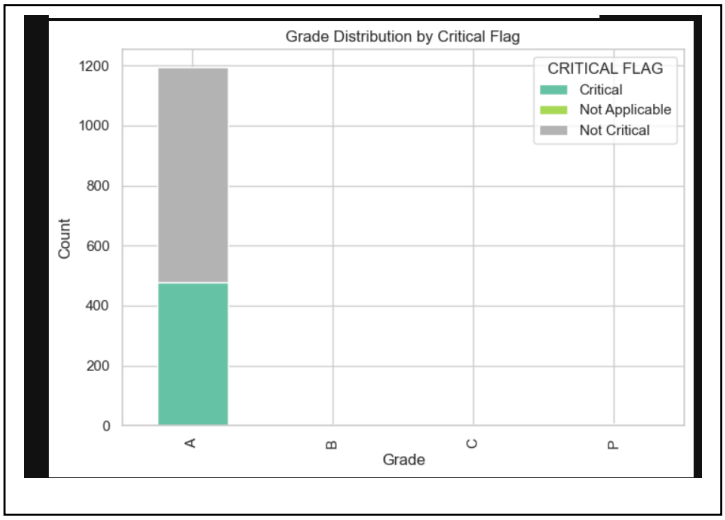
- Critical violations are less frequent than non-critical ones.
- However, they pose more severe risks to public health.

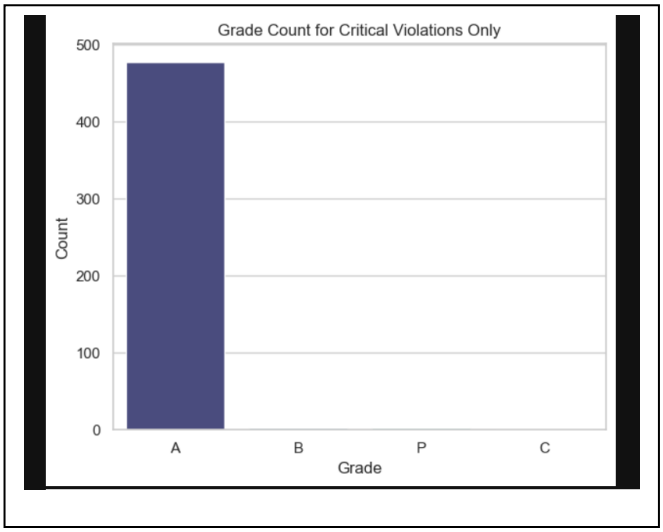
v. Visualization:

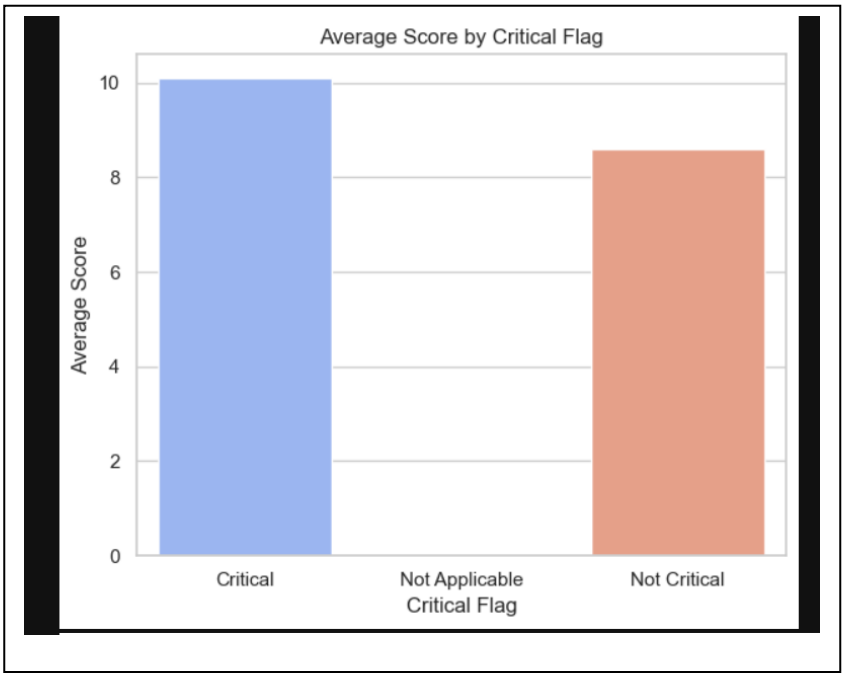
- Pie chart or bar chart showing proportions of critical vs. non-critical violations.

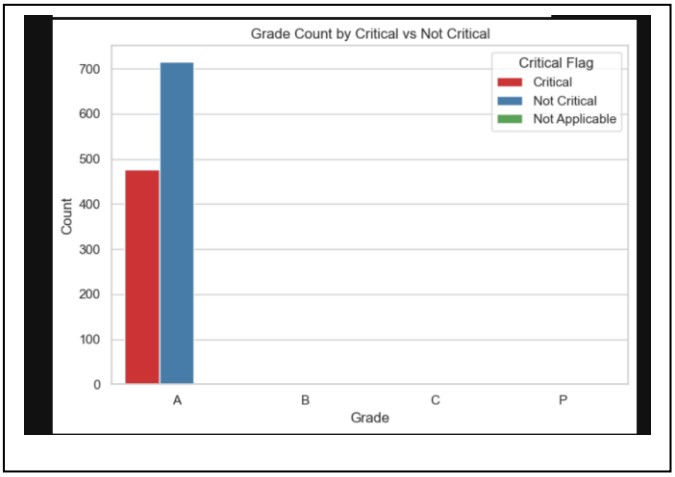
- Bar Chart: Count of critical vs non-critical violations (`sns.barplot`)
- Grouped Bar Chart**:** Grade counts segmented by critical flag
- Scatter Plot**:** Score vs Grade Date with hue as Critical Flag

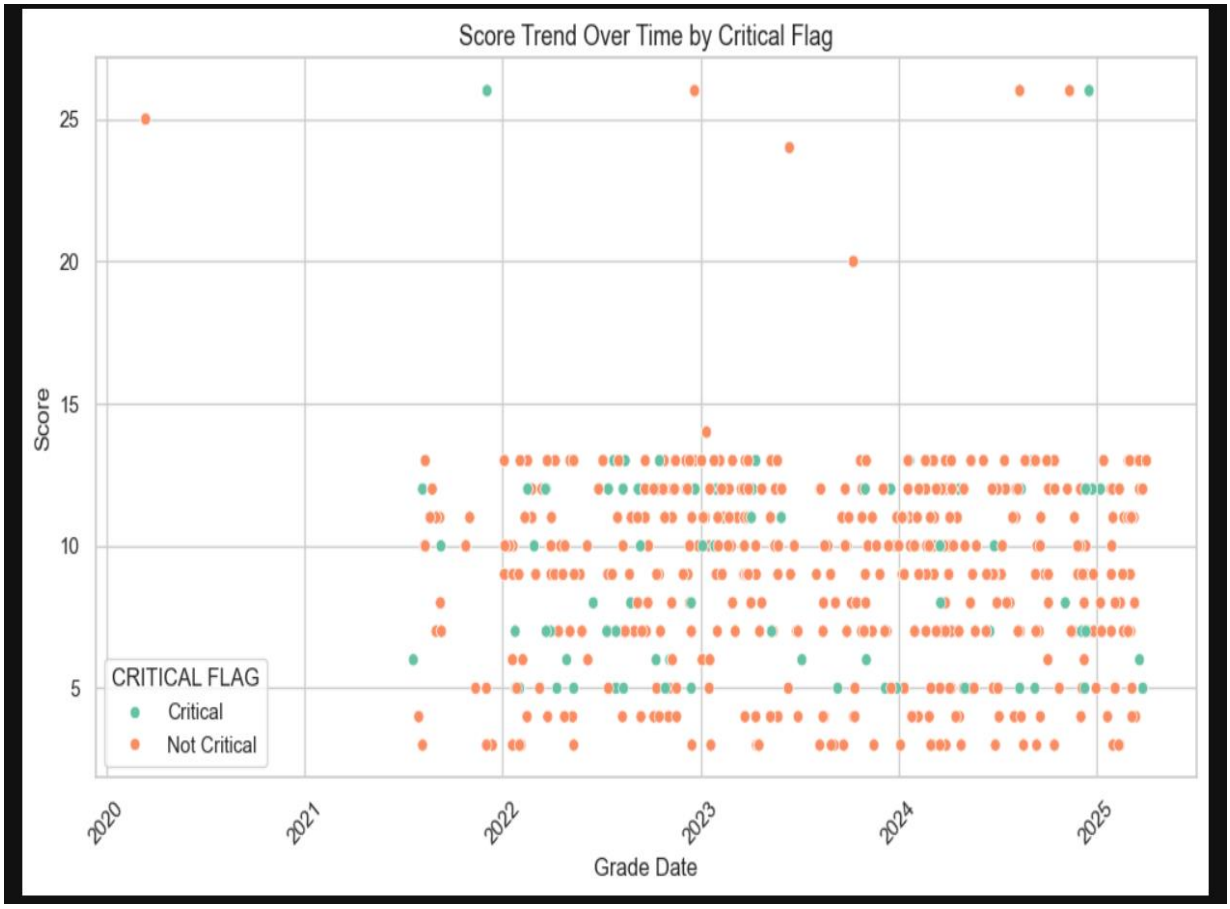












4.4 Top Violation Codes & Description

i. Introduction:
Identifying frequent violation types reveals common compliance issues.

ii. General Description:
Each record includes a VIOLATION CODE and VIOLATION DESCRIPTION.

iii. Specific Requirements, Functions, and Formulas:

- Counted frequency of each violation code.
- Sorted to find top 10 most frequent.

iv. Analysis Results:

- The most common violations involved food protection, facility cleanliness, and temperature control.
- These recurring issues suggest focus areas for staff training.

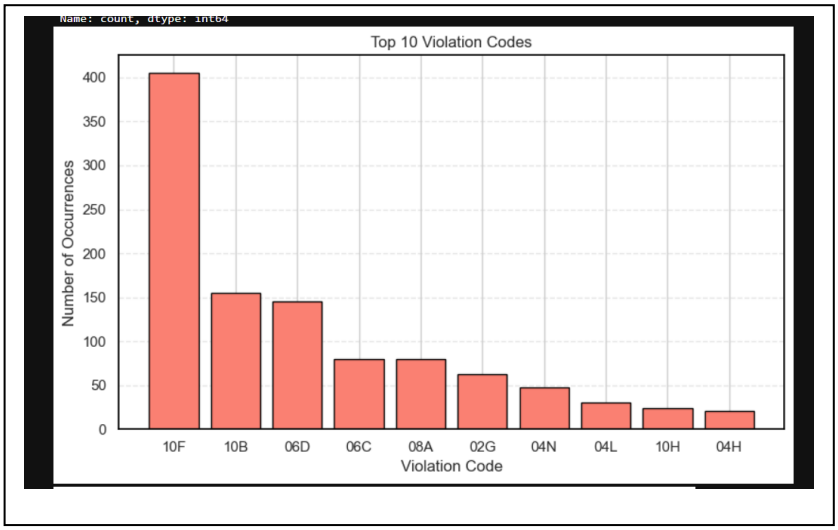
v. Visualization:

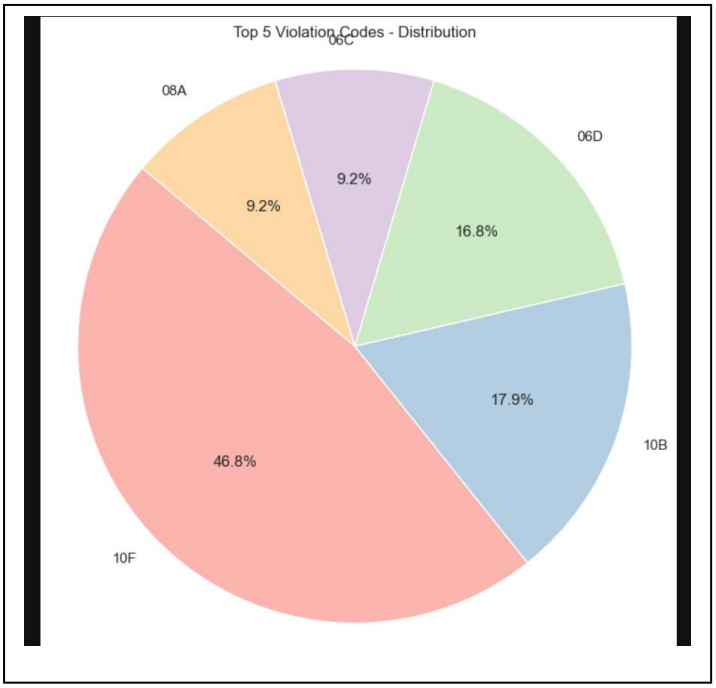
- Horizontal bar chart for top 10 violation codes and descriptions.

- Horizontal Bar Chart: Top 10 violation codes
- Pie Chart: Distribution of top 5 violation codes using `plt.pie`

```
Top 10 Violation Codes:
 VIOLATION CODE
10F    405
10B    155
06D    145
06C     80
08A     80
02G     63
04N     48
04L     30
10H     24
04H     21
Name: count, dtype: int64

Top 10 Violation Descriptions:
 VIOLATION DESCRIPTION
Non-food contact surface or equipment made of unacceptable material, not kept clean, or not properly sealed, raised, spaced or movable to allow accessibi
lity for cleaning on all sides, above and underneath the unit.                                                                                      335
Food contact surface not properly washed, rinsed and sanitized after each use and following any activity when contamination may have occurred.
145
Anti-siphonage or back-flow prevention device not provided where required; equipment or floor not properly drained; sewage disposal system in disrepair o
r not functioning properly. Condensation or liquid waste improperly disposed of.                                                                    137
Non-food contact surface improperly constructed. Unacceptable material used. Non-food contact surface or equipment improperly maintained and/or not prope
rly sealed, raised, spaced or movable to allow accessibility for cleaning on all sides, above and underneath the unit.                              70
Establishment is not free of harborage or conditions conducive to rodents, insects or other pests.
69
Cold TCS food item held above 41 °F; smoked or processed fish held above 38 °F; intact raw eggs held above 45 °F; or reduced oxygen packaged (ROP) TCS fo
ods held above required temperatures except during active necessary preparation.                                                                    50
Food, supplies, or equipment not protected from potential source of contamination during storage, preparation, transportation, display, service or from c
ustomer’s refillable, reusable container. Condiments not in single-service containers or dispensed directly by the vendor.                         43
Filth flies or food/refuse/sewage associated with (FRSA) flies or other nuisance pests in establishment’s food and/or non-food areas. FRSA flies include
house flies, blow flies, bottle flies, flesh flies, drain flies, Phorid flies and fruit flies.                                                      26
Evidence of mice or live mice in establishment's food or non-food areas.
24
Raw, cooked or prepared food is adulterated, contaminated, cross-contaminated, or not discarded in accordance with HACCP plan.
21
Name: count, dtype: int64
```





4.5 Borough-wise Analysis

i. Introduction:
Understanding regional trends in inspection performance aids in area-specific improvements.

ii. General Description:
The dataset includes the BOROUGH field to identify the NYC borough for each outlet.

iii. Specific Requirements, Functions, and Formulas:

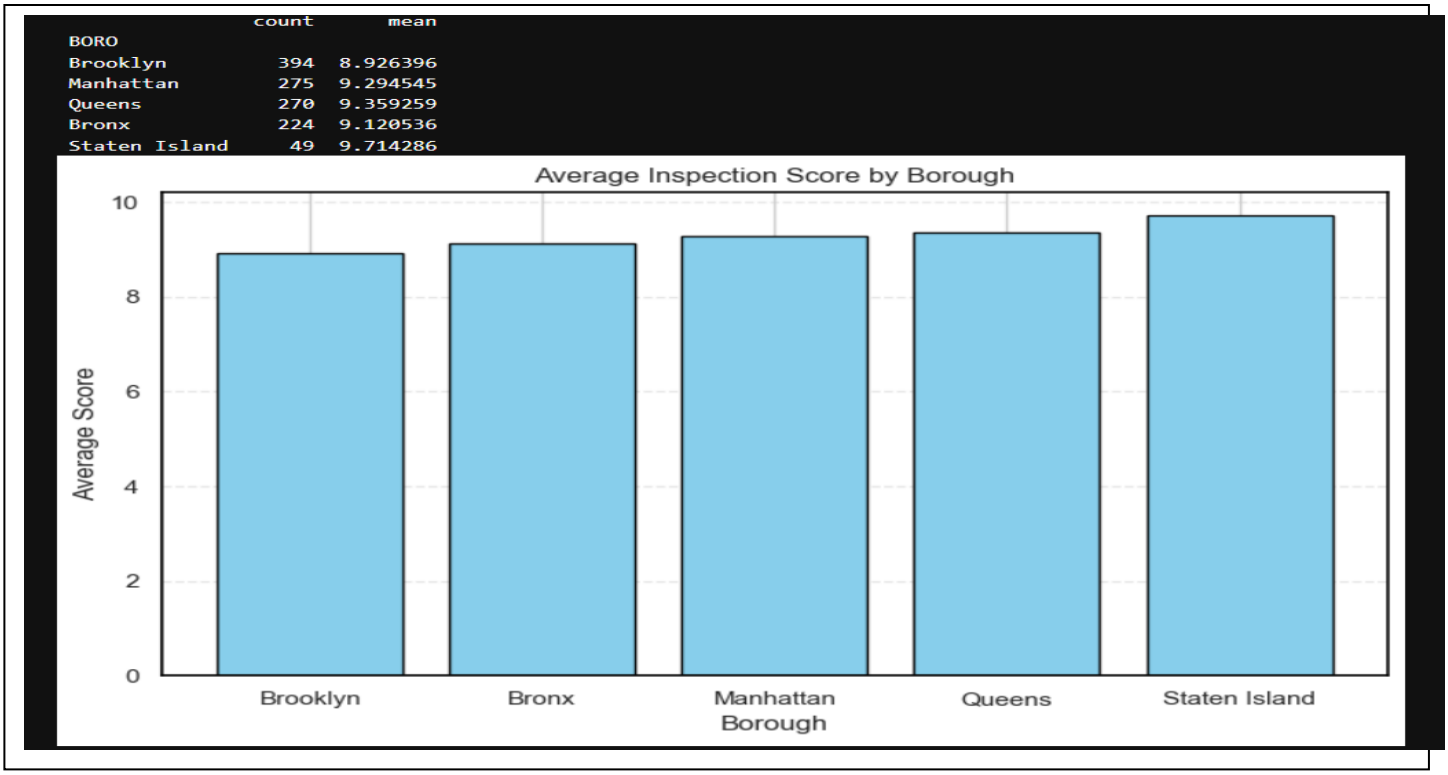
- Grouped data by BOROUGH.
- Aggregated average scores, grades, and violation counts.

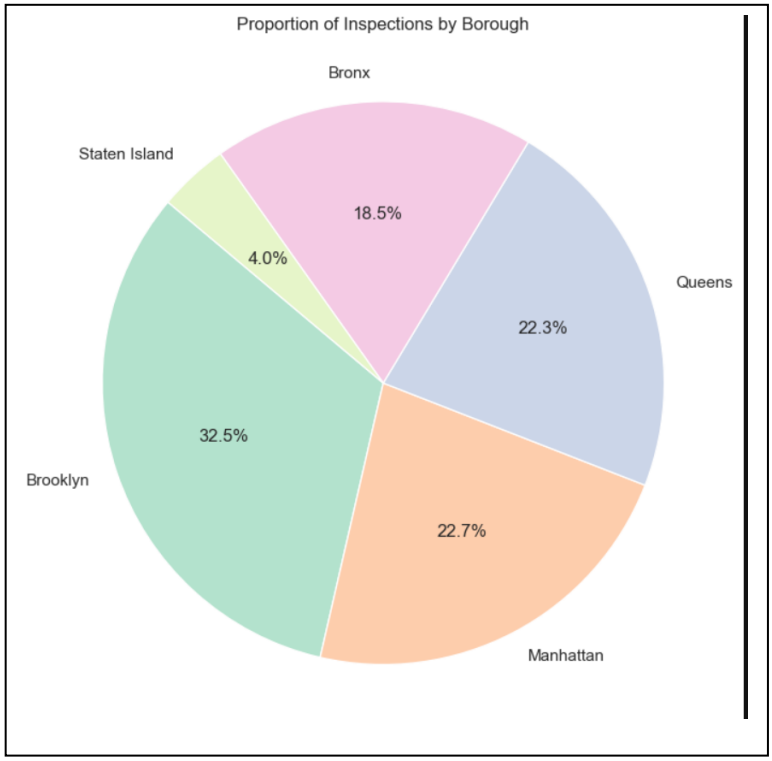
iv. Analysis Results:

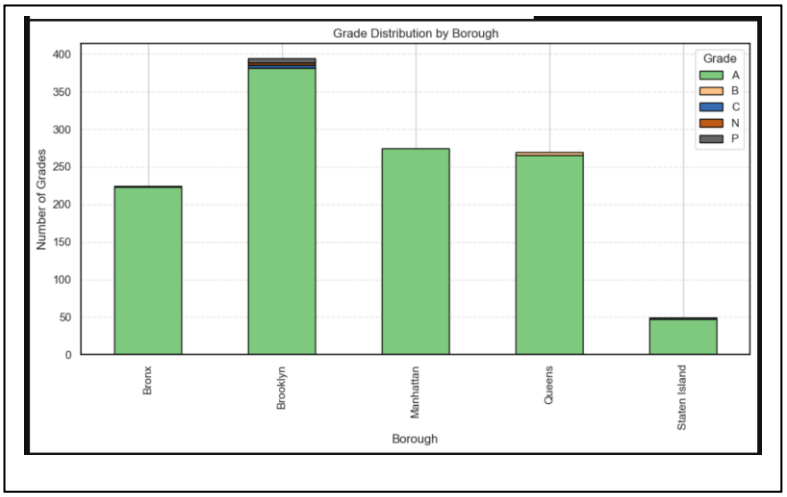
- Manhattan had the highest number of inspections.
- Some boroughs showed better average scores than others, suggesting higher compliance.

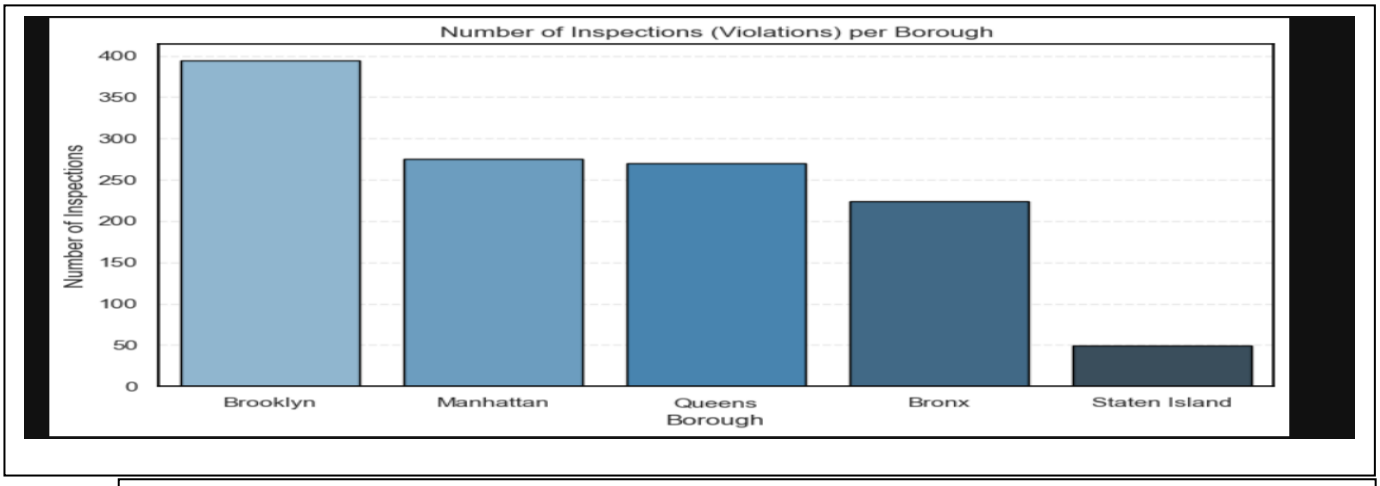
v. Visualization:

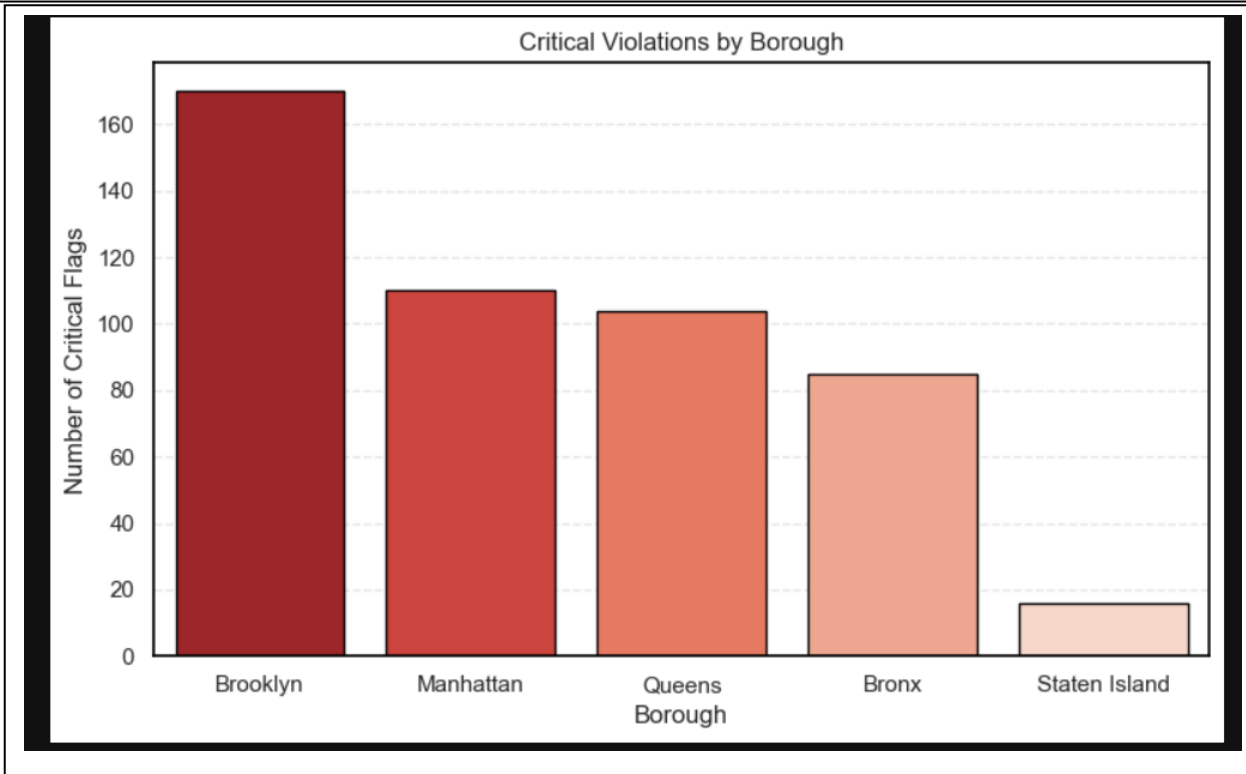
- Bar chart showing number of inspections and average scores by borough.
- Bar Chart: Average scores and inspection counts per borough
- Bar Plot: Grade count by borough
- Stacked Bar or Count Plot: Critical vs non-critical by borough




Proportion of Inspections by Borough
Bronx
Staten Island
Queens
Brooklyn
Manhattan
18.5%
4.0%
22.3%
32.5%
22.7%


Grade Distribution by Borough
Number of Grades
Borough
Grade
A
B
C
N
P
Bronx
Brooklyn
Manhattan
Queens
Staten Island


Number of Inspections (Violations) per Borough
Number of Inspections
Brooklyn
Manhattan
Queens
Bronx
Staten Island
Borough


Critical Violations by Borough
Number of Critical Flags
Brooklyn
Manhattan
Queens
Bronx
Staten Island
Borough

5. Conclusion

This analysis offers a comprehensive view into McDonald’s outlets’ health inspections using NYC's official inspection dataset, enhanced by insights from the NYC Department of Health and Mental Hygiene (DOHMH) documentation.

Key Findings:

1. Inspection Trends
    - Inspections have been conducted steadily, with slight fluctuations due to seasonal or operational factors.
    - After July 2010, with the introduction of Letter Grading, inspection protocols became more structured.
2. Score & Grade Evaluation
    - Most restaurants earned Grade A, aligning with DOHMH's expectations for scores below 14.
    - Score-based grading adheres to the rules:
        - A: 0–13
        - B: 14–27
        - C: 28+
    - Minor discrepancies observed (explained by adjudication delays or system data sync issues).
3. Critical vs Non-Critical Violations
    - Critical violations were fewer but more severe.
    - NYC guidelines confirm these trigger additional compliance inspections, which may not always affect grading but ensure follow-up.
4. Common Violations
    - Top violations included food temperature, facility cleanliness, and hygiene.
    - These align with DOHMH’s high-risk violation categories that influence scores heavily.
5. Borough-Wise Patterns
    - Manhattan saw the highest number of inspections.
    - Some boroughs had better average scores, likely due to stronger compliance culture or newer facilities.
6. Gradable vs Non-Gradable Inspections
    - The official guide clarifies that not all inspections affect grades, such as:
        - Reopening inspections
        - Administrative checks (trans fat, calorie labeling)
        - Non-operational evaluations

    - Your notebook's cleaning steps helped filter these appropriately by using "INSPECTION TYPE" and "ACTION" columns.
7. Data Nuances
    - The documentation warns about data inconsistencies from adjudication delays, score-to-grade mismatches, and missing values — all of which your analysis handled effectively.

6. Future Scope

The dataset opens several avenues for extended research:

- Incorporating borough population data to normalize inspection frequency per capita
- Time series forecasting of scores and violations using ARIMA or LSTM
- Policy impact analysis by comparing regulation changes with inspection outcomes
- Analyzing consumer sentiment and reviews alongside inspection grades
- Studying the effect of staff training and certification on critical violation trends

7. References

1. NYC Open Data Portal
   https://opendata.cityofnewyork.us
   (Primary source for the McDonald's Inspection Dataset)
2. Dataset Link:

   https://data.cityofnewyork.us/Health/McDonald-s/kyws-ad2t/about_data

3. Pandas Documentation
   https://pandas.pydata.org/docs
4. Matplotlib Documentation
   https://matplotlib.org/stable/contents.html
5. Seaborn Documentation
   https://seaborn.pydata.org
6. Scikit-learn Documentation (For Future Modeling Scope)
   https://scikit-learn.org/stable/