RESTAURANT INSPECTION
EDA PROJET

Name : Navaneetha krishnan

Course: DA/DS

Date : 22/09/2025

INTRODUCTION

What is the dataset?

The dataset I am using is the Restaurant Inspection, which contains information about restaurants Inspection

Why did you choose it?

I chose this dataset because it is a large dataset that provides detailed information about restaurant inspections

Objective of your analysis

The main objectives of this EDA project To understand a Data, Data Cleaning, Statistical Tests, Analysis using Plots

DATASET OVERVIEW

- Dataset : Restaurant inspections
- Number of Records: Large dataset with 290023 Rows
- Total Number of Columns 22
- Data Types:Numerical categorical and alphabetical categorical date columns
- Contains some missing and null values

OVERVIEW

	CAMIS	DBA	BORO	BUILDING	STREET	ZIPCODE	PHONE	CUISINE DESCRIPTION	INSPECTION DATE	ACTION	...	INSPECTION TYPE	Latitude	Longitude	Community Board	Council District	Census Tract	BIN	BBL	NTA	Location Point1
0	50114537	MAMA'S CUPBOARD	Manhattan	42	WEST 42 STREET	10036.0	6469191465	NaN	1/1/1900	NaN	...	NaN	40.754068	-73.982321	105.0	4.0	8400.0	1085593.0	1.012570e+09	MN17	NaN
1	50164148	TIENDA CHICOJ	Queens	10840	ROOSEVELT AVE	11368.0	7189166374	NaN	1/1/1900	NaN	...	NaN	40.750845	-73.858280	404.0	21.0	40100.0	4307718.0	4.019960e+09	QN26	NaN
2	50169870	TEX'S CHICKEN & BURGERS	Queens	21712	MERRICK BLVD	11413.0	9178850048	NaN	1/1/1900	NaN	...	NaN	40.679915	-73.752087	413.0	31.0	35800.0	4279942.0	4.130110e+09	QN66	NaN
3	50132087	FANTASTIC BEASTS	Queens	36-10	UNION STREET	11354.0	2538807717	NaN	1/1/1900	NaN	...	NaN	40.763482	-73.828056	407.0	20.0	86900.0	4112354.0	4.049770e+09	QN22	NaN
4	50172040	BROAD STREET DELI	Manhattan	270	PARK AVENUE	10017.0	7043286780	NaN	1/1/1900	NaN	...	NaN	40.755634	-73.975109	105.0	4.0	9400.0	1035421.0	1.012830e+09	MN17	NaN

5 rows × 27 columns

DATA CLEANING STEPS

Handling Missing Values

- Checked for null or missing values in all columns.
- Dropped rows containing null values in both numerical (e.g., score) and categorical columns to ensure data completeness.
- Ensured dataset consistency by keeping only rows with valid values, avoiding potential bias from imputed data.

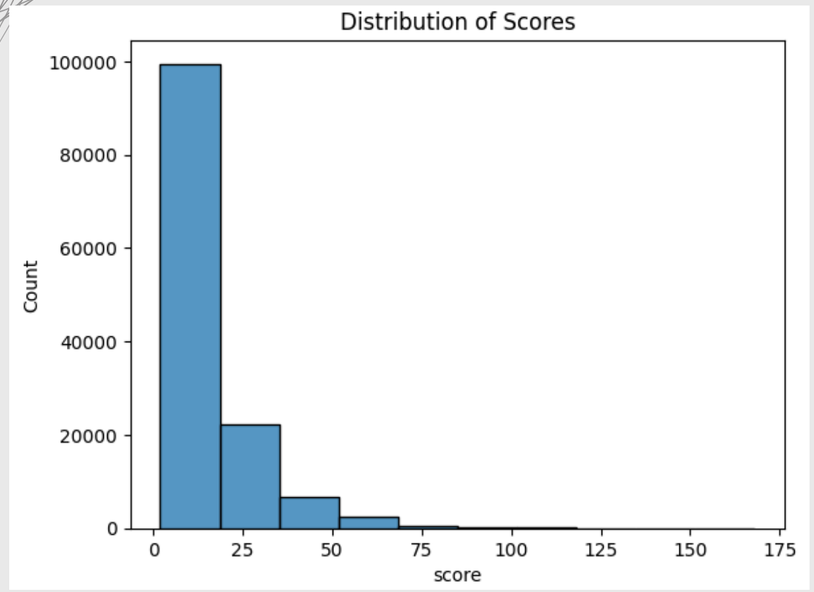
Removing Duplicates

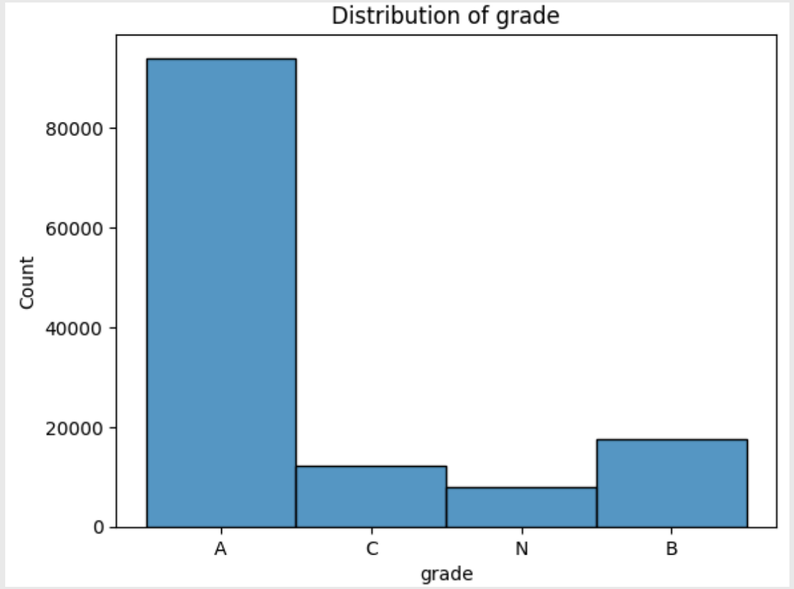
- Identified and removed any duplicate rows to ensure data integrity.

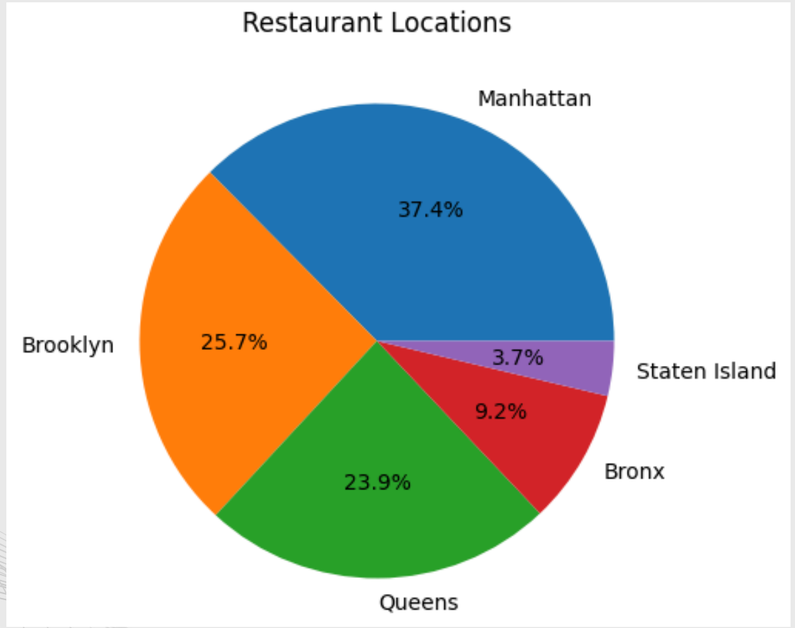
Filtering Unnecessary Data

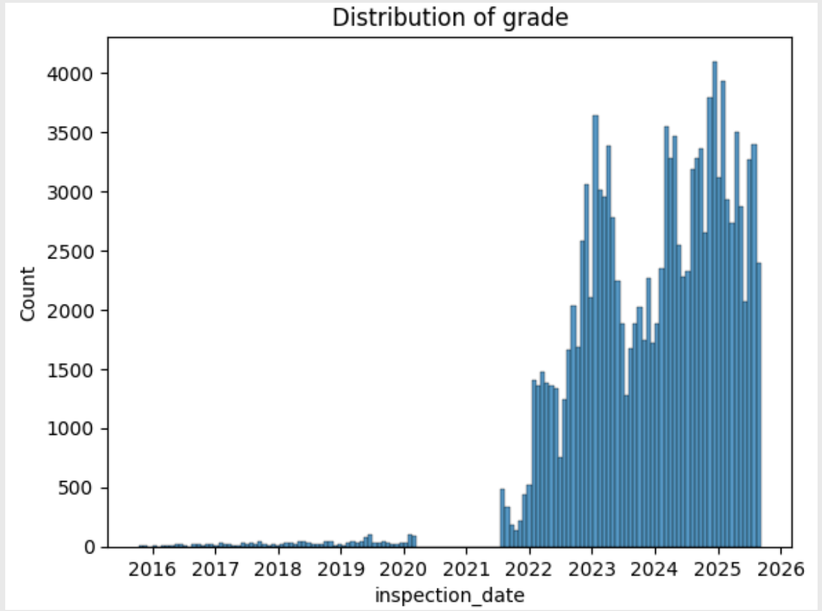
- Removed irrelevant columns e.g records with unavailable inspection info.

UNIVARIATE ANALYSIS

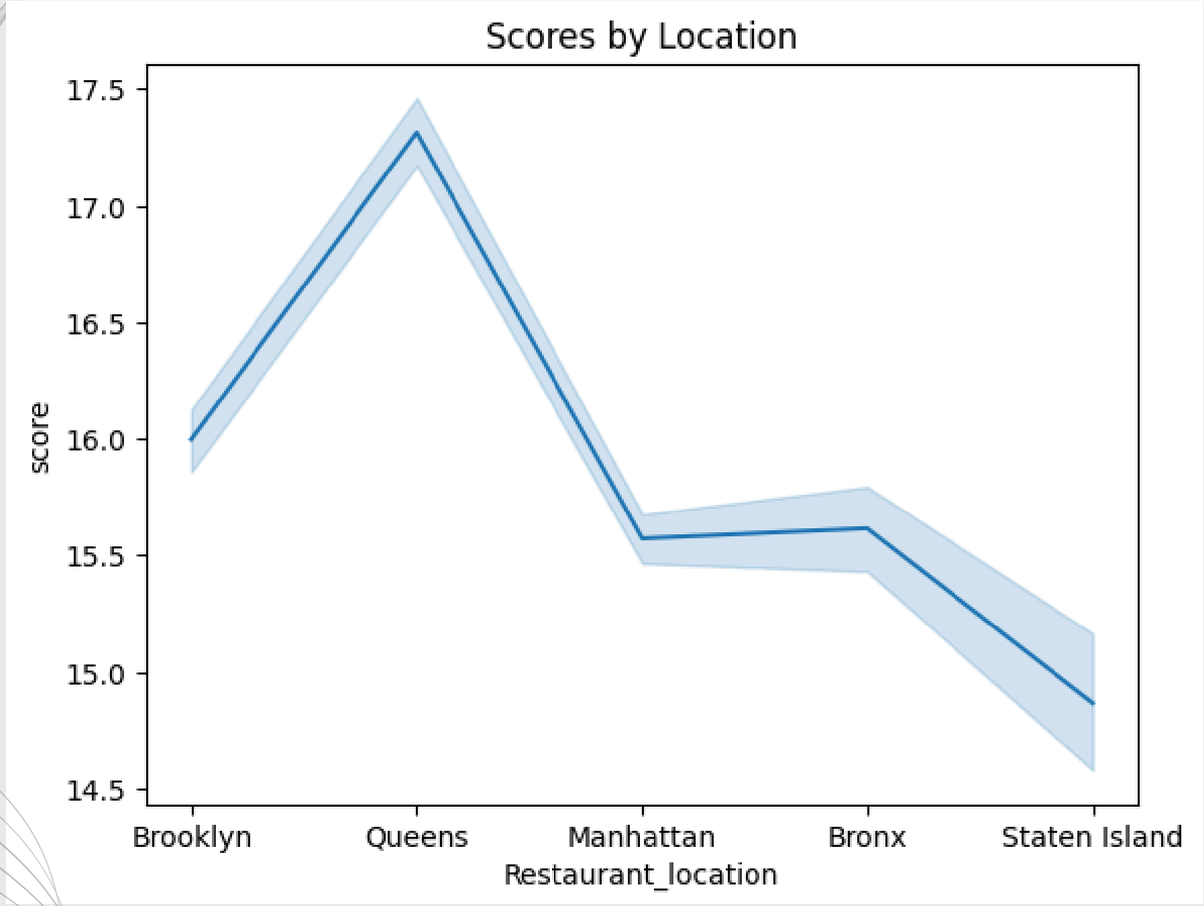


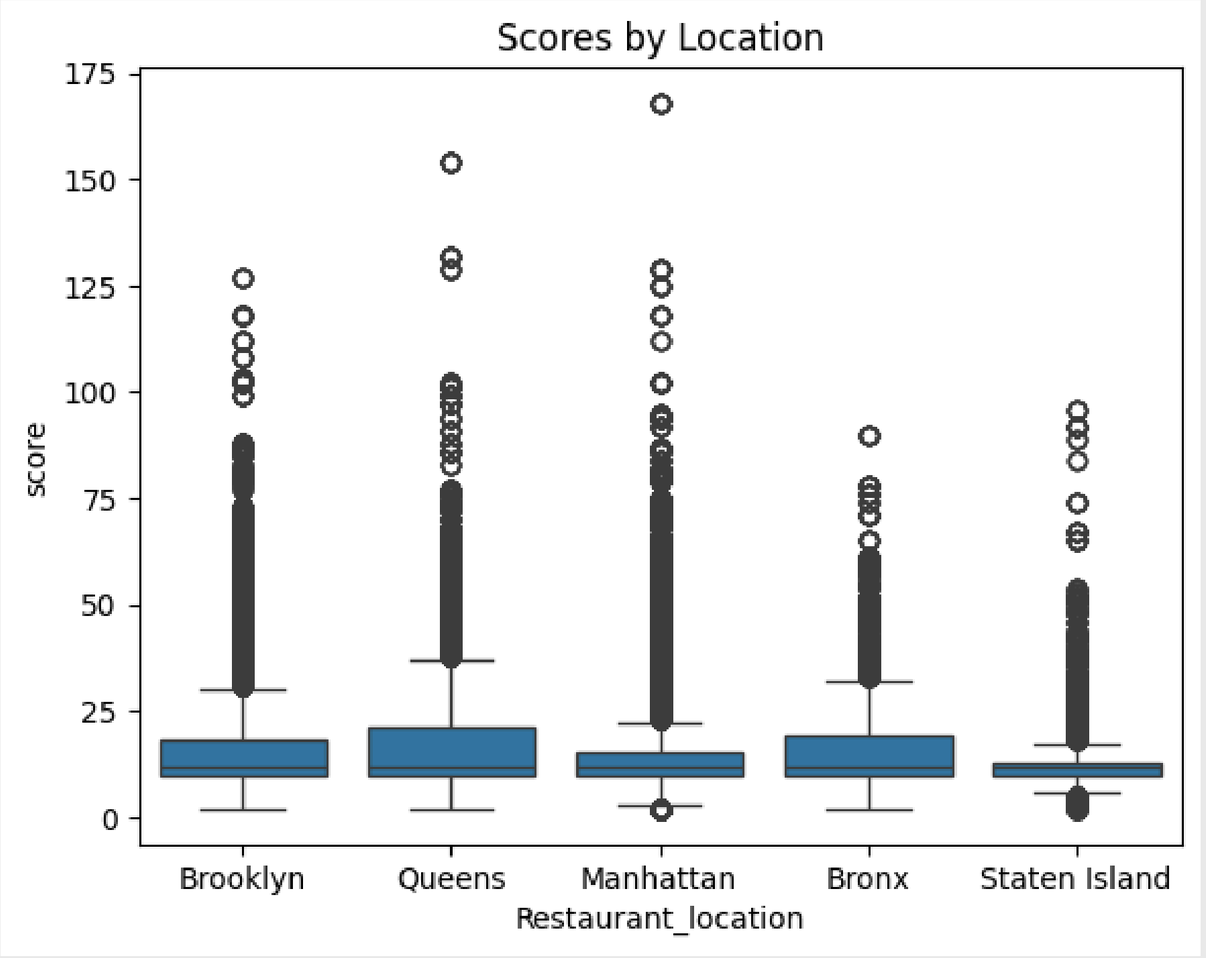




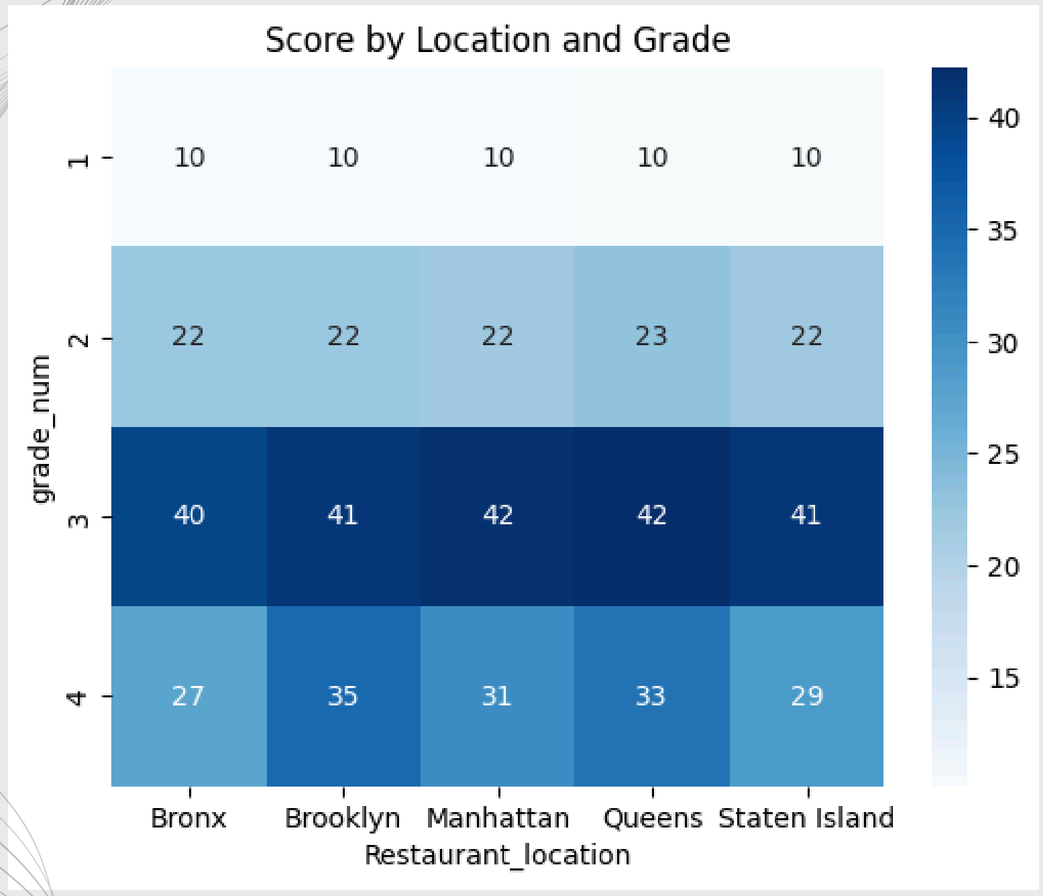


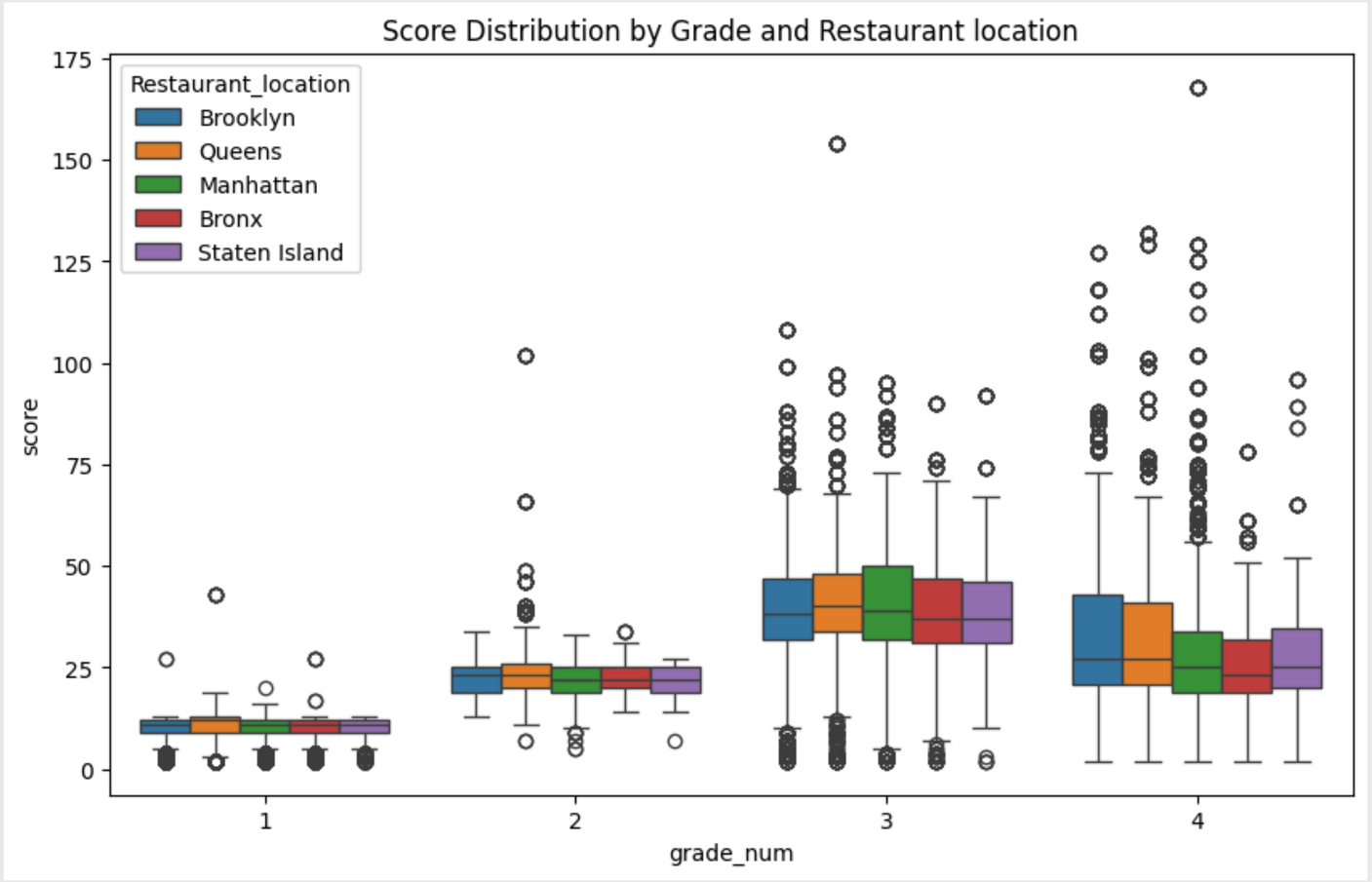
BIVARIATE ANALYSIS





MULTIVARIATE ANALYSIS





INSIGHTS

- GRADE DISTRIBUTION: MOST RESTAURANTS RECEIVED AN A GRADE, WITH FEWER B AND C GRADES, INDICATING OVERALL GOOD COMPLIANCE.
- BOROUGH TRENDS: MANHATTAN AND BROOKLYN HAVE A HIGHER NUMBER OF INSPECTIONS, WHILE VIOLATION PATTERNS VARY ACROSS BOROUGHS.
- COMMON VIOLATIONS: FOOD HANDLING, SANITATION, AND CLEANLINESS-RELATED VIOLATIONS WERE MOST FREQUENT.
- CUISINE PATTERNS: CERTAIN CUISINE TYPES SHOWED SLIGHTLY HIGHER VIOLATION RATES, REFLECTING OPERATIONAL CHALLENGES.
- RE-INSPECTIONS: MANY RESTAURANTS IMPROVED THEIR GRADES AFTER RE-INSPECTION, HIGHLIGHTING THE EFFECTIVENESS OF THE GRADING PROGRAM.
- GEOGRAPHIC INSIGHTS: MAPPING REVEALED CLUSTERS OF RESTAURANTS WITH RECURRING VIOLATIONS, WHICH CAN INFORM TARGETED INSPECTIONS AND PUBLIC HEALTH INTERVENTIONS.
- THESE INSIGHTS HELP UNDERSTAND COMPLIANCE TRENDS, IDENTIFY HIGH-RISK AREAS, AND SUPPORT DECISION-MAKING FOR BOTH REGULATORS AND RESTAURANT OWNERS.

STATISTICAL TESTS

CHI-SQUARE TEST OF INDEPENDENCE

- TO CHECK IF THERE IS A RELATIONSHIP BETWEEN TWO CATEGORICAL VARIABLES RELATIONSHIP **BETWEEN RESTAURANT_LOCATION VS GRADE.**

ONE-SAMPLE T-TEST

- TO CHECK IF THE MEAN OF A SAMPLE IS SIGNIFICANTLY DIFFERENT FROM THE POPULATION MEAN.

INDEPENDENT T-TEST

- TO COMPARE THE MEANS OF TWO INDEPENDENT GROUPS TO SEE IF THEY ARE SIGNIFICANTLY DIFFERENT.
- IN YOUR CASE: INSPECTION SCORES IN MANHATTAN VS BROOKLYN.

ANOVA TEST

- TO CHECK IF MEAN SCORES DIFFER ACROSS MORE THAN TWO GROUPS.
- IN YOUR CASE: COMPARING INSPECTION SCORES ACROSS ALL RESTAURANT LOCATIONS.

CONCLUSION

- MOST RESTAURANTS MAINTAIN GOOD HYGIENE, WITH GRADE **A** .
- INSPECTION SCORES VARY ACROSS LOCATIONS SOME AREAS AND RESTAURANTS NEED IMPROVEMENT.
- STATISTICAL TESTS SHOW SIGNIFICANT RELATIONSHIPS BETWEEN LOCATION AND GRADES, AND DIFFERENCES IN MEAN SCORES.
- OUTLIERS AND MISSING DATA WERE HANDLED CAREFULLY TO ENSURE ACCURATE INSIGHTS.
- OVERALL, THE DATASET PROVIDES VALUABLE INFORMATION TO MONITOR AND IMPROVE RESTAURANT STANDARDS.

RECOMMENDATIONS

- IMPROVE DATA COLLECTION: ENSURE ALL INSPECTIONS HAVE COMPLETE SCORES AND GRADES TO REDUCE MISSING VALUES.
- REGULAR MONITORING: USE INSIGHTS TO MONITOR RESTAURANTS WITH LOW SCORES OR CRITICAL VIOLATIONS REGULARLY.
- LOCATION-BASED ANALYSIS: FOCUS ON AREAS WITH CONSISTENTLY LOWER SCORES TO IMPLEMENT TARGETED INTERVENTIONS.
- PREDICTIVE MODELING: BUILD MODELS TO PREDICT INSPECTION OUTCOMES BASED ON RESTAURANT FEATURES.
- TREND ANALYSIS: ANALYZE SEASONAL OR YEARLY TRENDS TO IDENTIFY PATTERNS IN RESTAURANT PERFORMANCE.

THANK YOU