

BAN -A

New York City Restaurant Inspection Results

Final Presentation

Fayzal Olushi, Hanna Blel, Isabella Cahen, Juan Pariente



Foodborne Illness Incident at Kumo Japanese Steakhouse

Stony Brook, New York.

September 9, 2023

28 people reported sick

12 rushed to Hospital for treatment

- Kumo Japanese Steakhouse received 15 health code violations.
- Eight of these violations were related to risk factors for foodborne illness.



“Each year, thousands of New York City residents become sick from consuming foods or drinks that are contaminated with harmful bacteria, viruses or parasites.”

— NYC
HEALTH.GOV

What's the Problem?

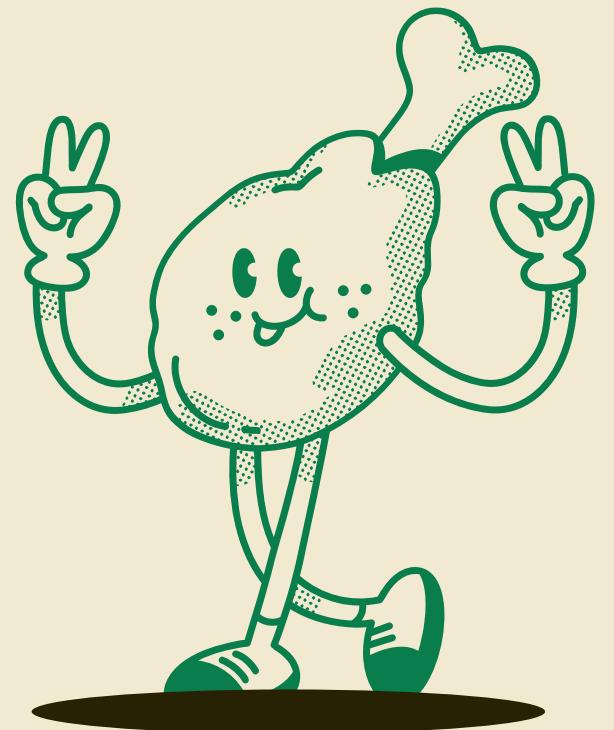
DOHMH

- Reflects poorly (not doing their job)
- Job insecurity
- Wasting Tax Payer´s money

Restaurants

- Endanger public health
- Loose customers
- Risk legal action againts
- Risk Closure
- Subject to fines

(\$100 per violation, maximum of \$2,500 per inspection.)



Our Solution

HeathGrade Solutions (HGS)
aims to...

1. Help track restaurants compliance
2. Aid in preventing faulty restaurant health code compliance in NYC
3. Assist new restaurants in improving their food safety and hygiene scores.
4. Ensure public health and safety



HealthGrade
SOLUTIONS

How we do it

How? Providing the Department of Health and Mental Hygiene (DOHMH) key Insights using predictive modelling.

1) Proactive Intervention:

Faster identification of establishments at risk

2) Improved Efficiency:

Efficient allocation of limited resources

StakeHolders

Consumers

Avoid the risk of food born illnesses,
have peace of mind.

Emerging Restaurants

Enhance their reputation and
performance. Ensure compliance and
avoid the risk of violations.

DOHMH

Increased productivity and efficiency.
No one is slipping through the cracks



Revenue Streams

DOHMH

Pay us to improve their workflow



(SAAS)

Restaurants

Pay us to be consistent with the Health codes

Dataset

NYC Open Data: DOHMH New York City Restaurant Inspection Results

Columns: 27

Rows: 216,080

Contains every sustained **violation** citation
from every full or special **inspection** conducted
up to three years before the most recent
inspection for **restaurants** and **college**
cafeterias in **NYC**

Features

BORO	CUISINE DESCRIPTION	ACTION	VIOLATION CODE	CRITICAL FLAG	SCORE	GRADE
The borough where the restaurant is located (Manhattan, Bronx, Brooklyn, Queens, or Staten Island).	The action taken as a result of the inspection (e.g., Violations were cited, No violations were recorded, Establishment was closed).			Indicates whether the violation is critical to public health (Critical or Not Critical).		The letter grade assigned to the restaurant (A, B, C, or Z)
	The type of cuisine served by the restaurant		The code associated with each specific violation		The inspection score given to the restaurant.	

Features

GRADE = Target

A = 0-13 points

B = 14-27 points

C = 28+ points

N = Not yet graded

P = Grade pending issued on re-opening

Z = Grade pending



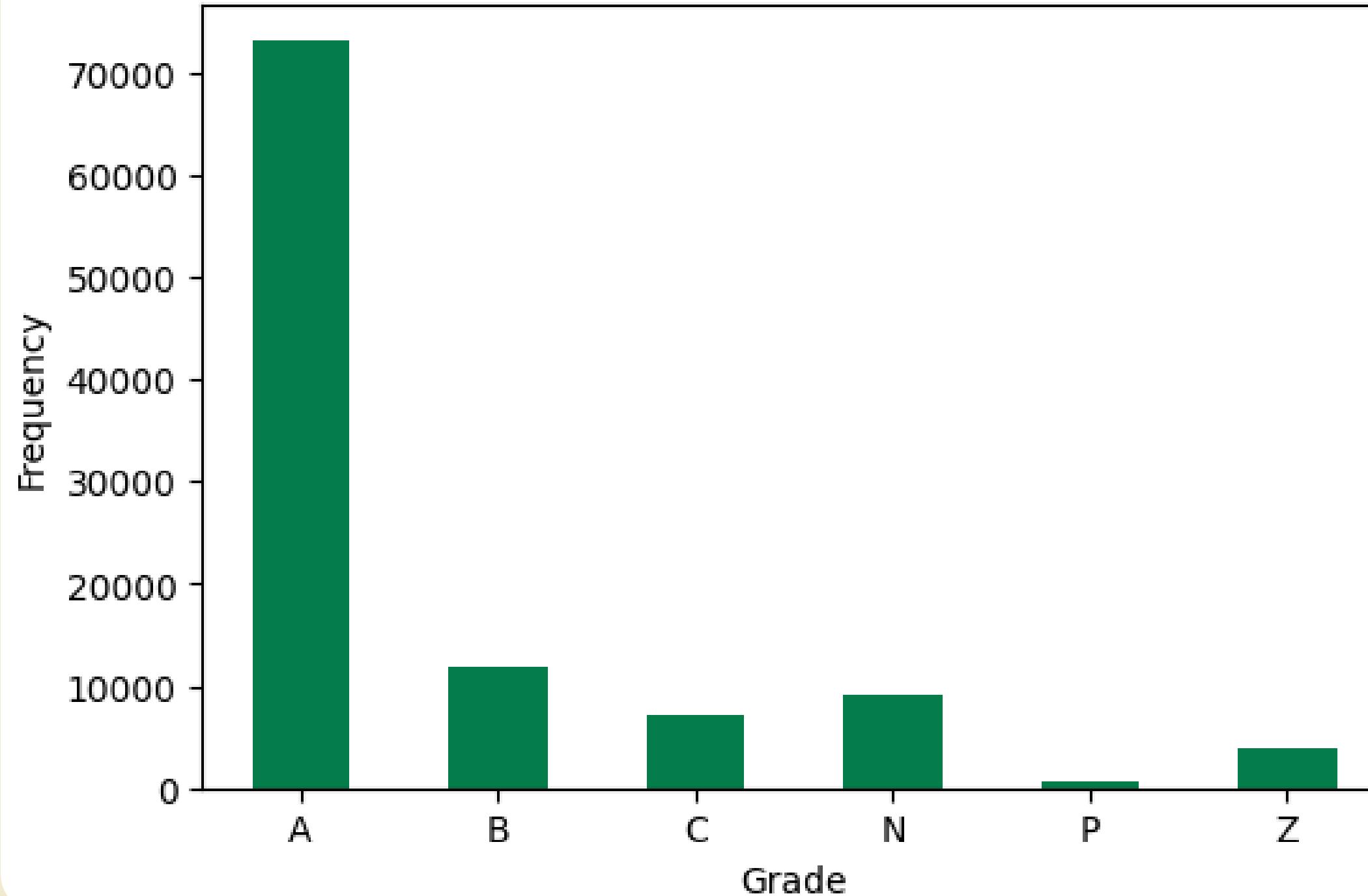
BAN - A

EDA

EDA

Univariate

Bar Plot of Inspection Grade (GRADE)

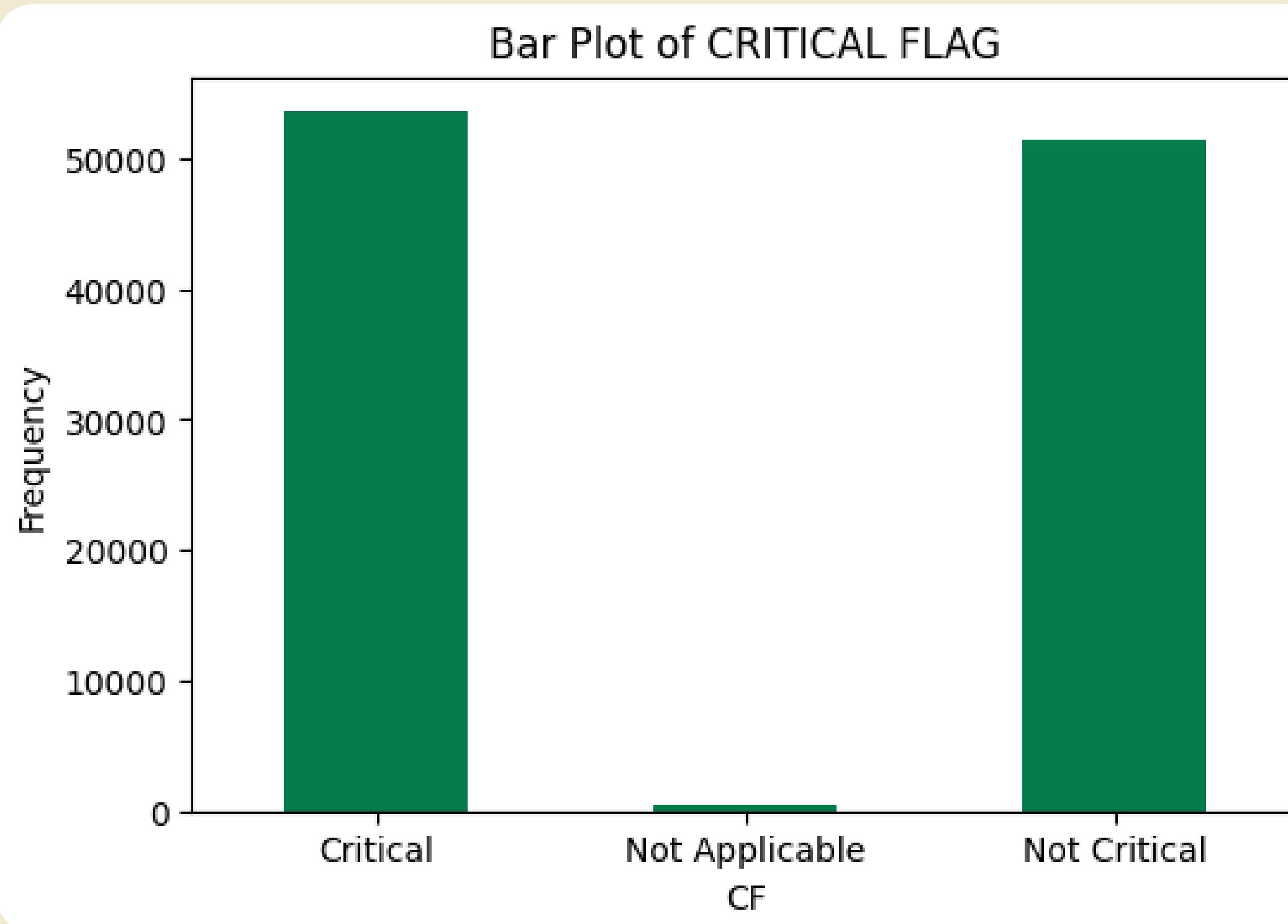


- The plot depicts the **frequency** of different inspection **grades** given to restaurants by the Department of Health and Mental Hygiene (DOHMH) in New York City.
- The **tallest bar** corresponds to **grade A**; this indicates that over 70,000 inspections resulted in this letter grade.
- Grades B and C have significantly lower frequencies.

Overall: Most restaurants received an **“A” grade**, suggesting **high overall compliance** with health standards among NYC restaurants.

EDA

Univariate

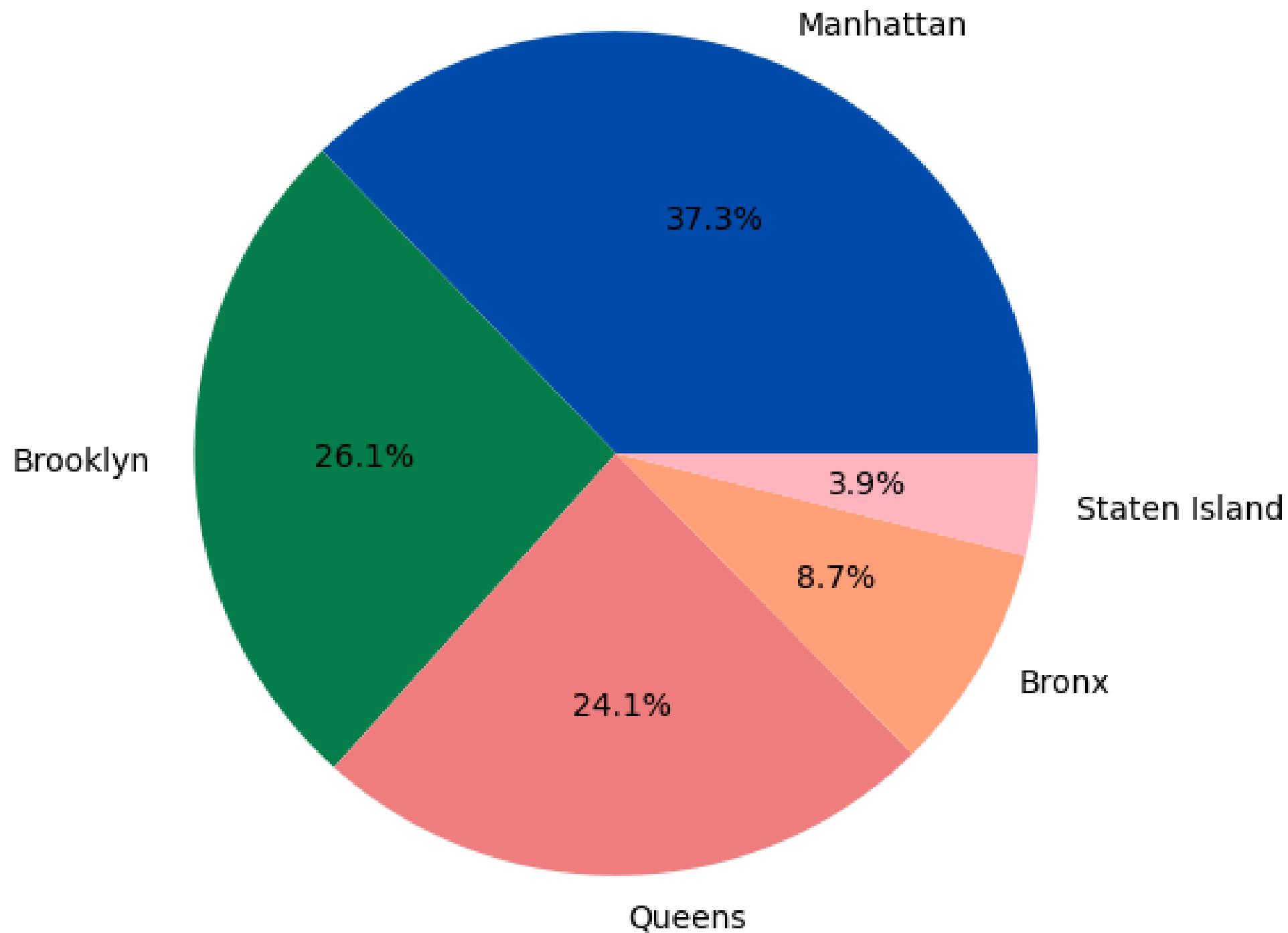


- The bar plot represents the **frequency** of the different categories of the **CRITICAL FLAG** feature.
- Grade Categories: Critical, Not Applicable, and Not Critical.
- The bars for both Critical and Not Critical reach **approximately above 50,000**. However, the **Critical** bar is **slightly higher**.
- In summary, while **most violations are not critical, addressing** all types of violations is **crucial** for maintaining restaurant **hygiene** and **safety**.

EDA

Univariate

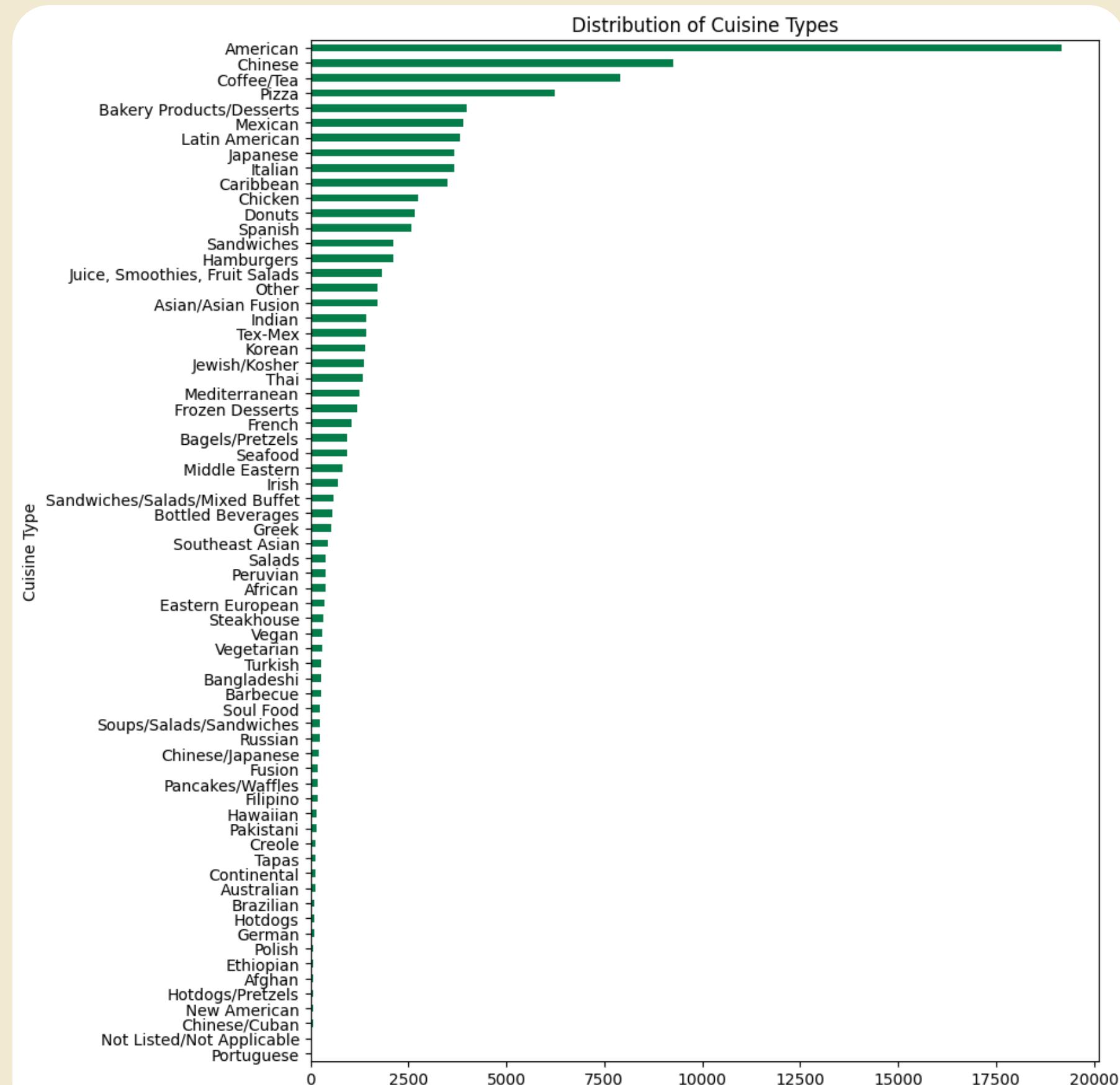
Pie Chart of Borough (BORO)



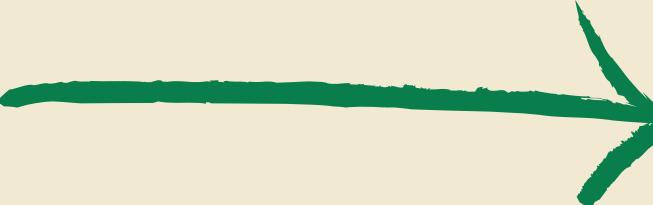
- The pie chart illustrates the distribution of **different boroughs** in the dataset.
- **Manhattan:** Occupies the **largest segment** (37.3%).
- Brooklyn: The second-largest segment (26.1%).
- Queens: Represents 24.1%.
- Bronx: A smaller portion (8.7%).
- Staten Island: The smallest slice (3.9%).
- Manhattan has the highest presence in the dataset, followed by Brooklyn and Queens.
- This distribution helps us **locate** which areas to **target** for our **initial launch**.

EDA

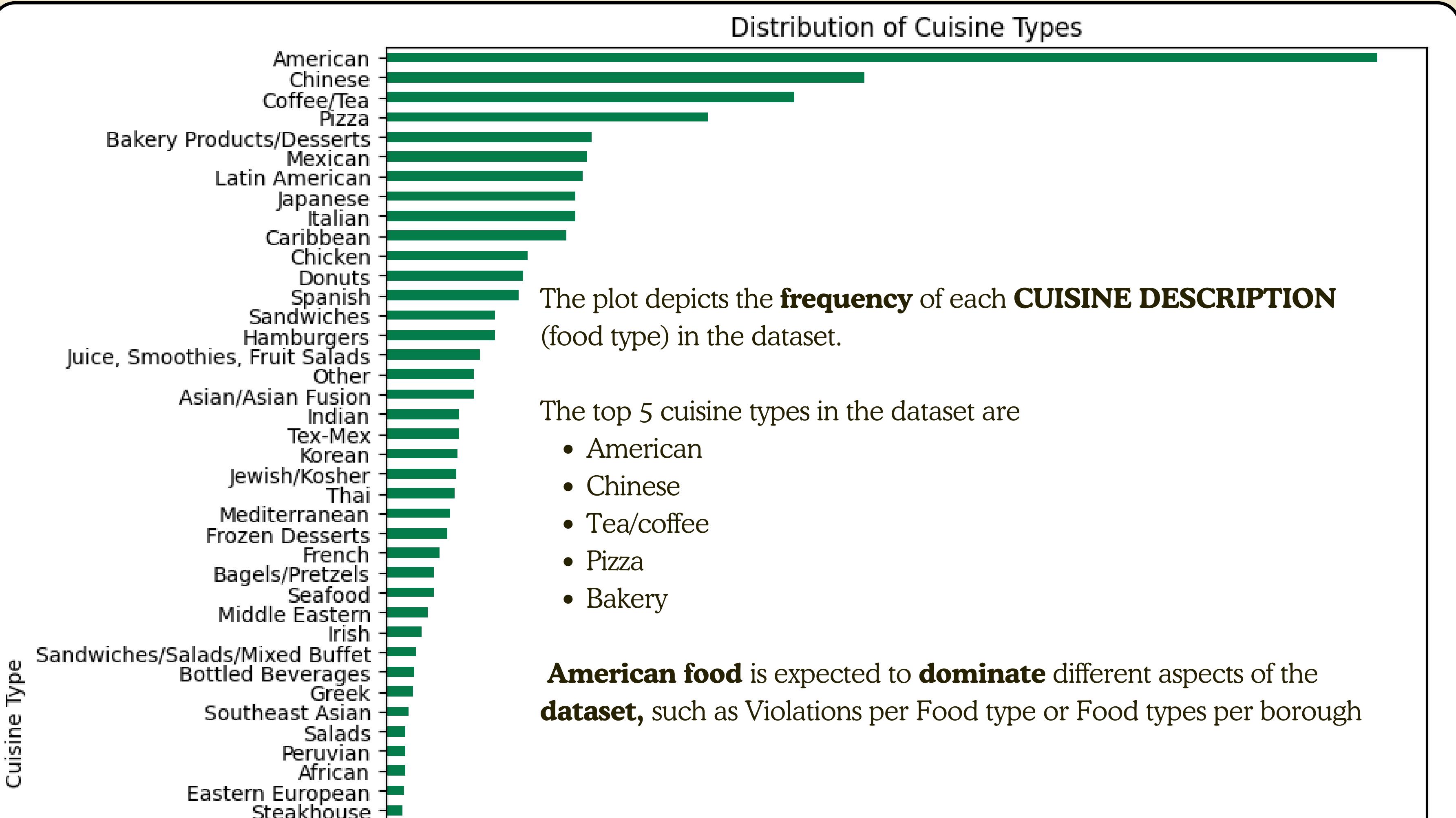
Univariate



Cuisine Distribution



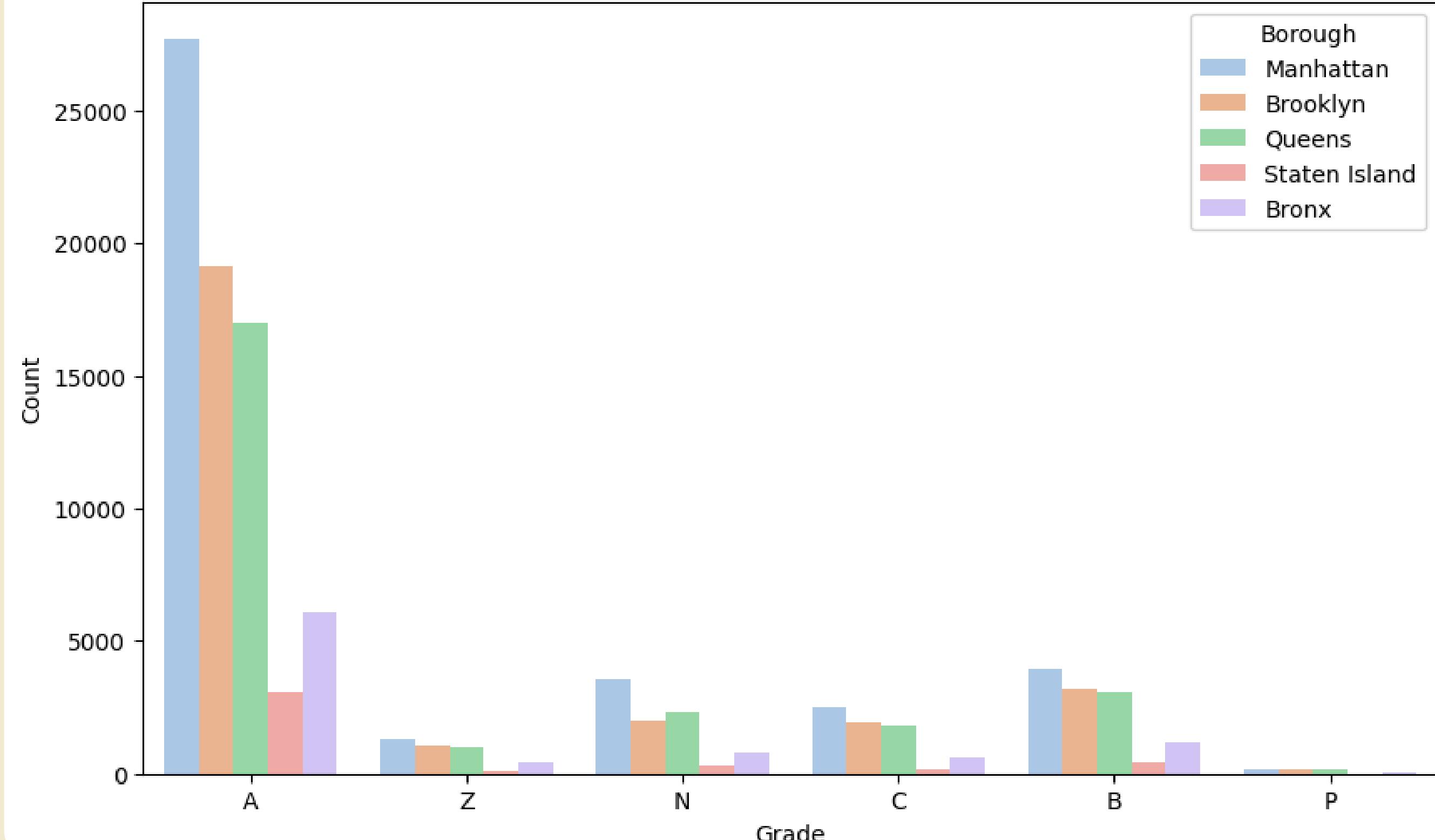
*Next Slide



EDA

Bivariate

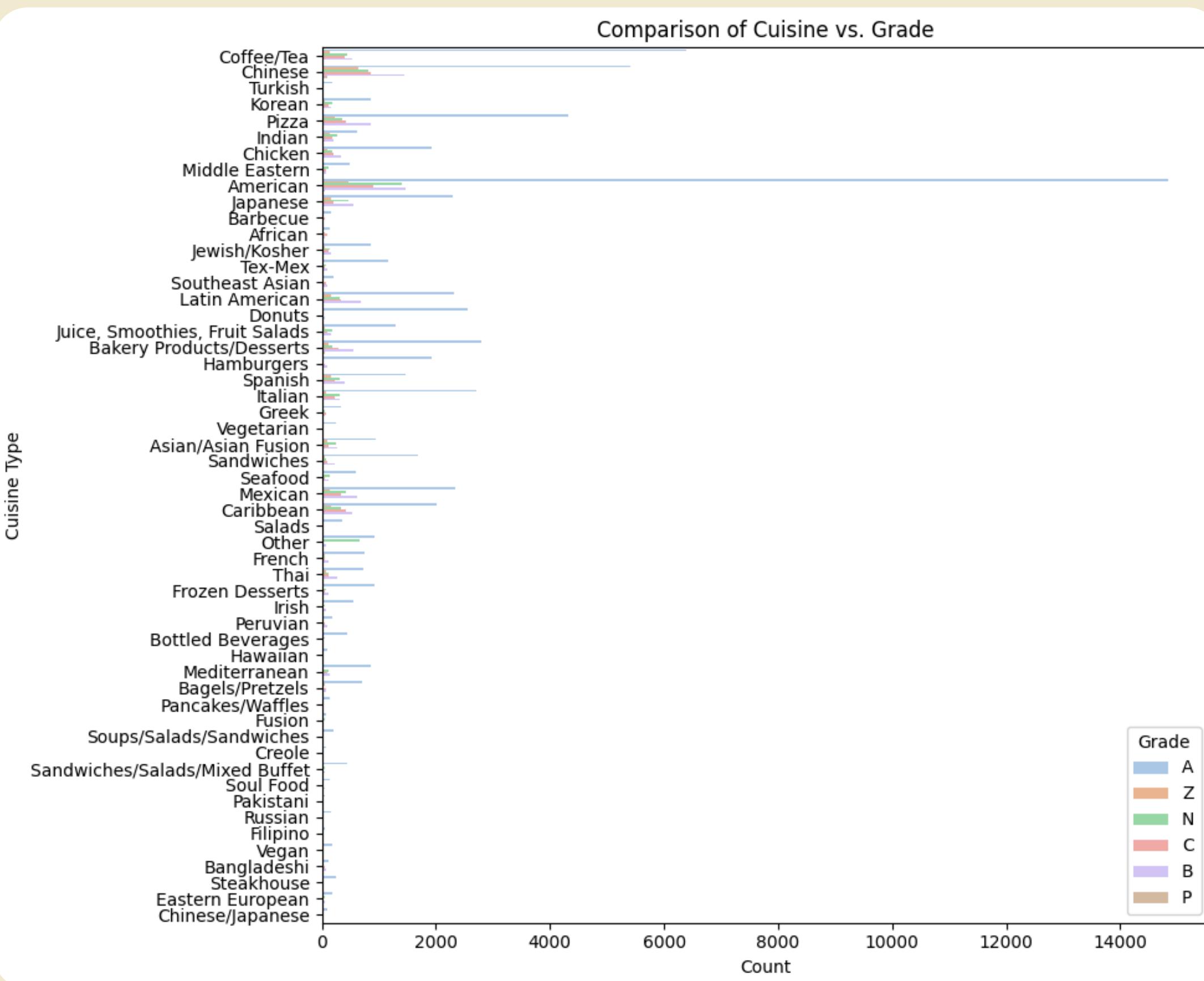
Bar Plot of Grade vs. Borough



- The bar plot illustrates the **distribution** of restaurant **grades** **across** different **boroughs** in New York City.
- **Manhattan has the highest count of restaurants with an “A” grade**, exceeding 20,000.
- **Staten Island has the lowest counts** for all grades.
- Overall Trend: **Grade A is the most common.**

EDA

Bivariate



Cuisine vs Grade



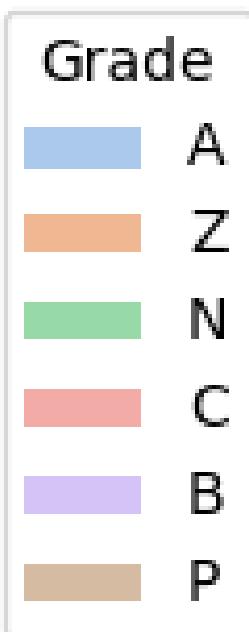
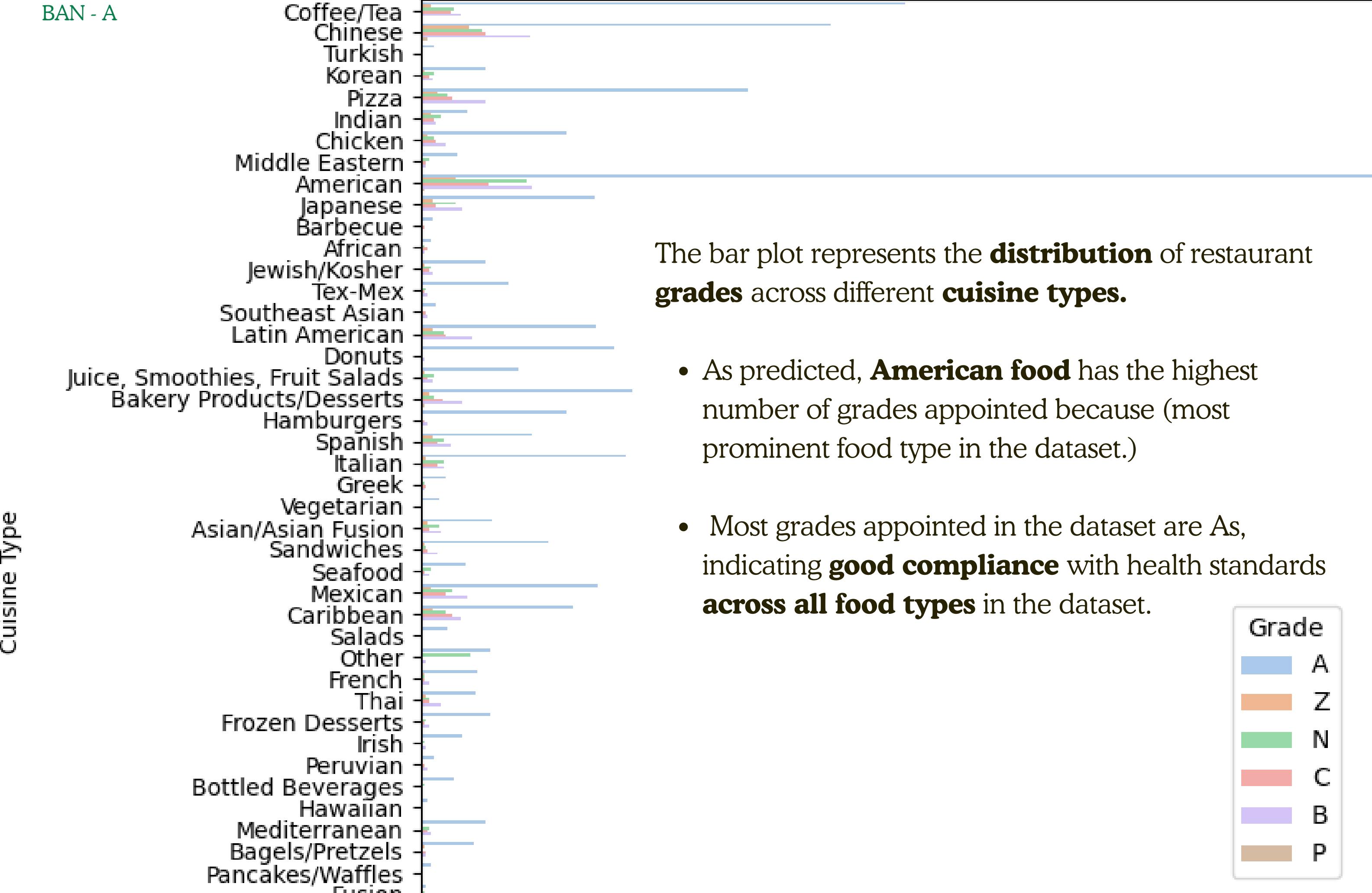
*Next Slide

Comparison of Cuisine vs. Grade

BAN - A

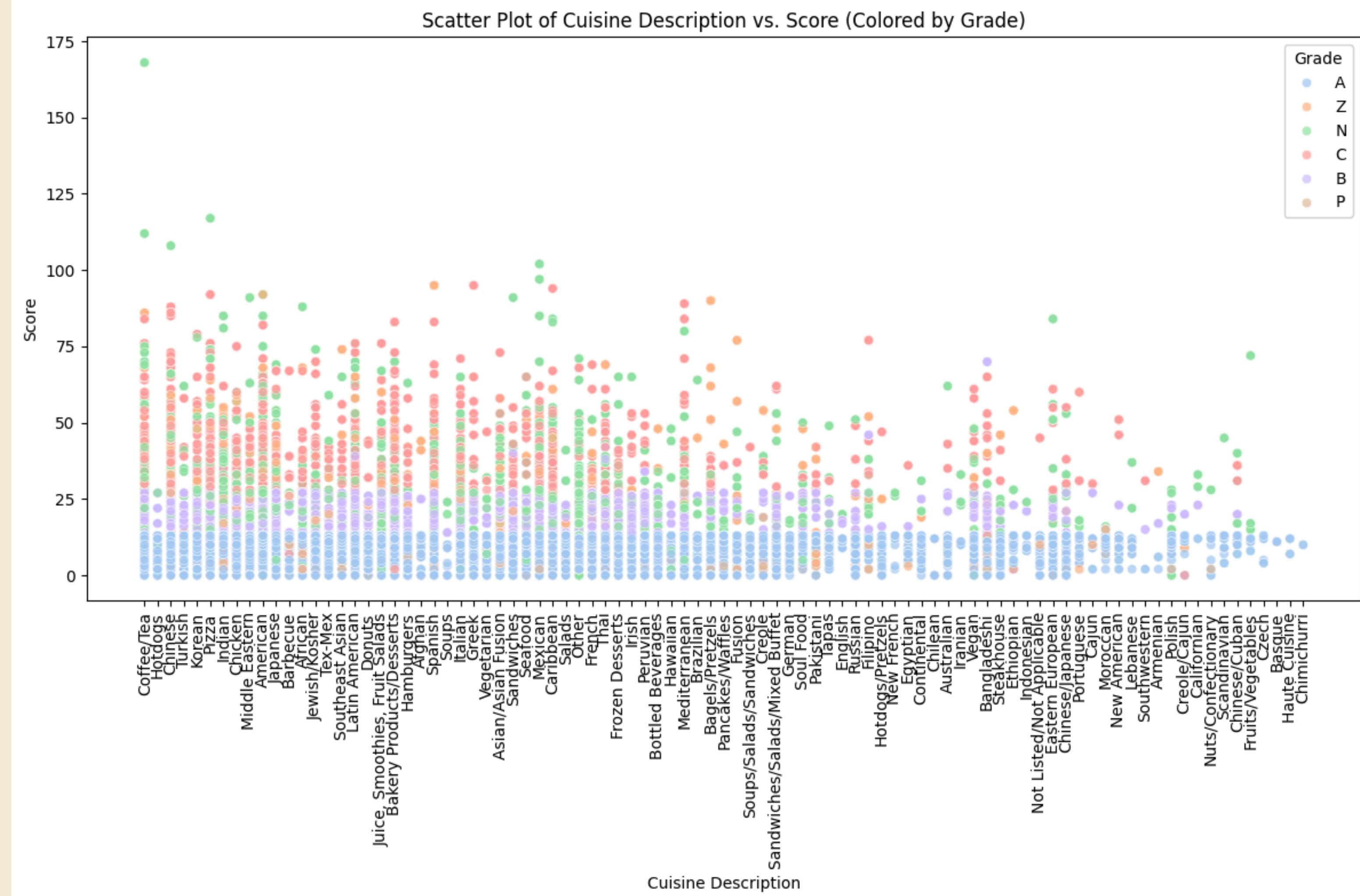
19

EDA
Multivariate



EDA

Multivariate

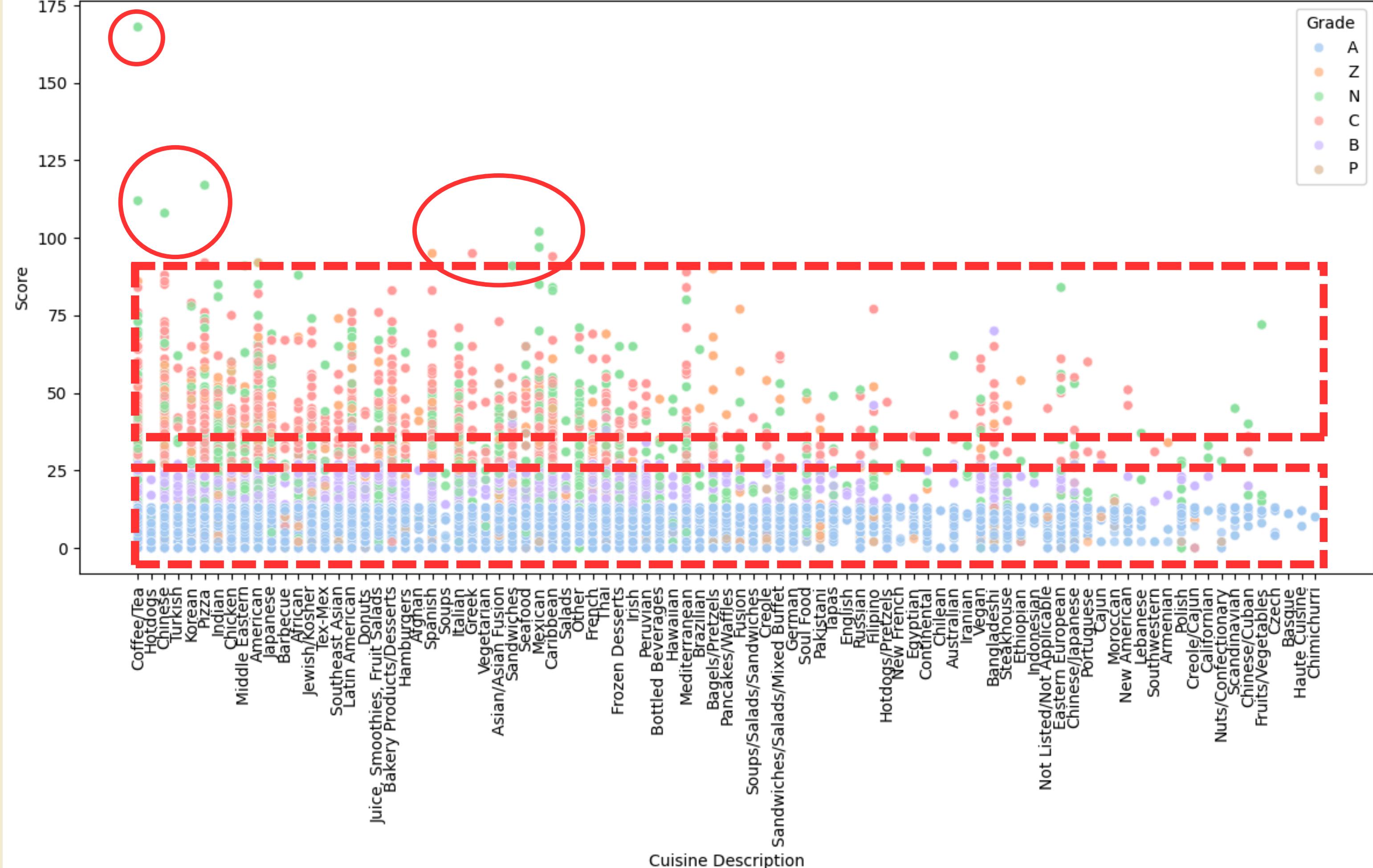


EDA

Multivariate

- The **higher** the score, the **worse** the grade.
- **Density** of plot **decreases** as the **score increases**.

Scatter Plot of Cuisine Description vs. Score (Colored by Grade)



Model Analysis

Top 3 models out of 11 total models

Baseline Model

Baseline: Accuracy: 70%

Target: GRADE

Features:

1. **BORO**
2. **CUISINE DESCRIPTION**
3. **ACTION**
4. **CRITICAL FLAG**
5. **VIOLATION CODE**

The lower the score the better

- A = 0–13 points (score)
- B = 14–27 points (score)
- C = 28+ points (score)
- N = Not yet graded (score)
- P = Grade pending issued on re-opening (score)
- Z = Grade pending (score)

		Confusion Matrix					
		A	B	C	N	P	Z
True Label	Predicted Label	14714	0	0	0	0	2
		2323	0	0	0	0	0
C	A	1382	0	0	3	0	20
		1745	0	0	0	24	11
N	B	0	0	0	17	1	101
		533	0	0	15	0	157
		A	B	C	N	P	Z

Model 1

Kbest 3: Accuracy: 70%

Target: GRADE

Features:

1. CUISINE DESCRIPTION

2. ACTION

- A = 0–13 points (score)
- B = 14–27 points (score)
- C = 28+ points (score)
- N = Not yet graded (score)
- P = Grade pending issued on re-opening (score)
- Z = Grade pending (score)

		Confusion Matrix						
		A	B	C	N	P	Z	
True Label	Predicted Label	A	14672	42	2	0	0	
		B	562	1706	54	1	0	
C	Predicted Label	C	22	256	984	143	0	
		N	5	9	208	1557	0	
P	Predicted Label	P	0	0	1	27	38	
		Z	0	1	2	42	6	
							654	
		A	B	C	N	P	Z	

Model 2

DT classifier Accuracy: 69%

Target: GRADE

Features:

1. **BORO**
2. **CRITICAL FLAG**
3. **ACTION**
4. **VIOLATION CODE**
5. **CUISINE DESCRIPTION**

- A = 0–13 points (score)
- B = 14–27 points (score)
- C = 28+ points (score)
- N = Not yet graded (score)
- P = Grade pending issued on re-opening (score)
- Z = Grade pending (score)

Confusion Matrix

		Predicted Label					
		A	B	C	N	P	Z
True Label	A	21189	381	145	265	2	27
	B	3250	160	43	76	0	10
C	1809	105	74	71	12	22	
N	2333	85	54	197	2	17	
P	0	0	29	4	73	82	
Z	729	35	32	36	80	143	

Model 3

KNN Accuracy: 96%

Target: GRADE

Features:

1. BORO

2. CUISINE DESCRIPTION

3. ACTION

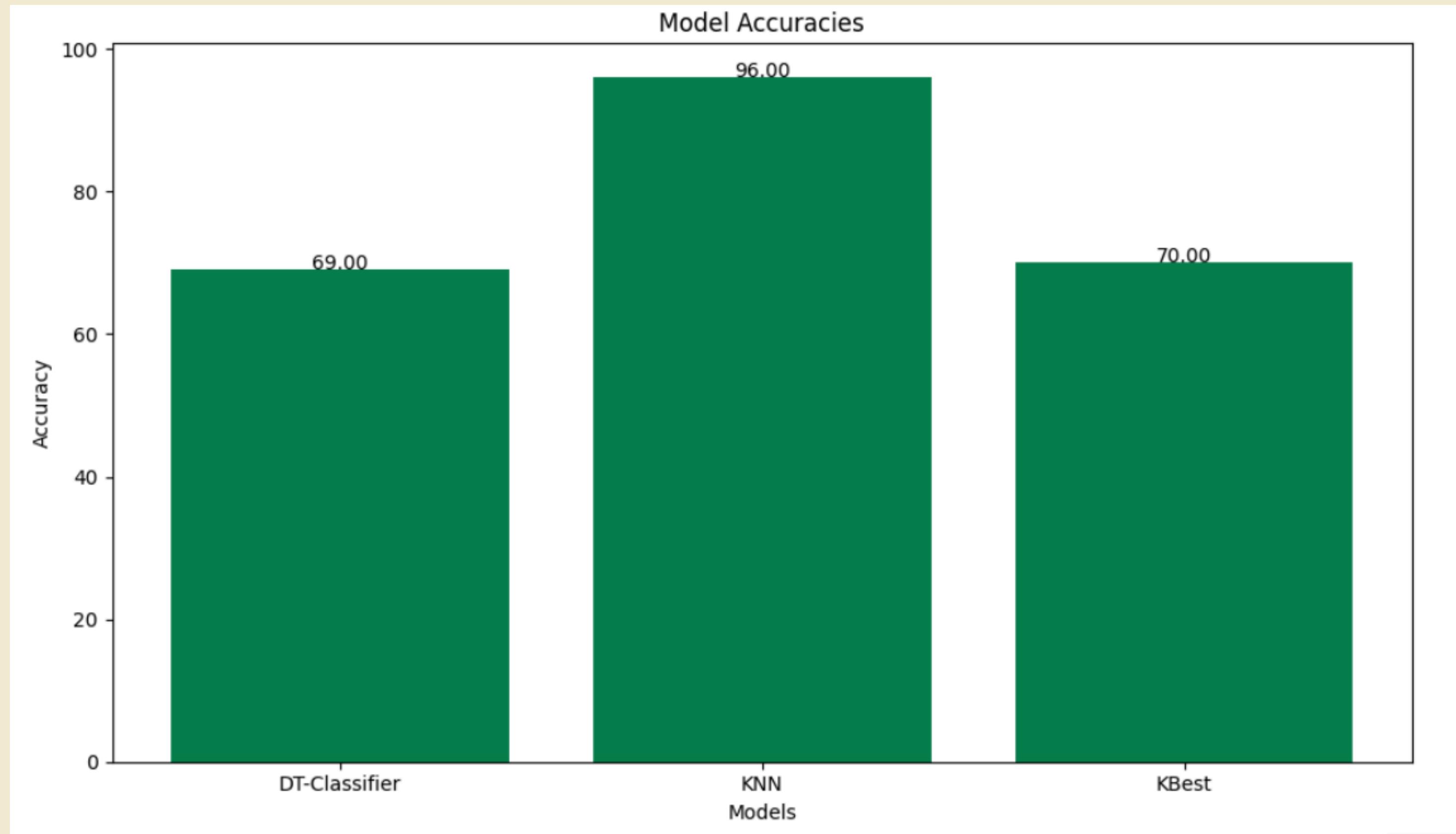
4. CRITICAL FLAG

5. VIOLATION CODE

- A = 0–13 points (score)
- B = 14–27 points (score)
- C = 28+ points (score)
- N = Not yet graded (score)
- P = Grade pending issued on re-opening (score)
- Z = Grade pending (score)

		Confusion Matrix					
		A	B	C	N	P	Z
True Label	Predicted Label	A	14711	3	2	0	0
		B	264	2049	9	1	0
True Label	Predicted Label	C	17	117	1219	52	0
		N	5	5	96	1670	3
True Label	Predicted Label	P	0	0	1	34	71
		Z	0	1	2	36	660

Model 1-3



Model Conclusion



KNN

Accuracy: 96%

Target: GRADE

Features:

1. **BORO**
2. **CUISINE DESCRIPTION**
3. **ACTION**
4. **CRITICAL FLAG**
5. **VIOLATION CODE**

Implementation



Southern Kitchen



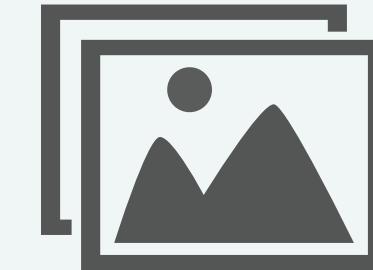
Southern Kitchen 68 Bergen St, Brooklyn, NY 11201

Southern Kitchen's chance of violations and risk percentage

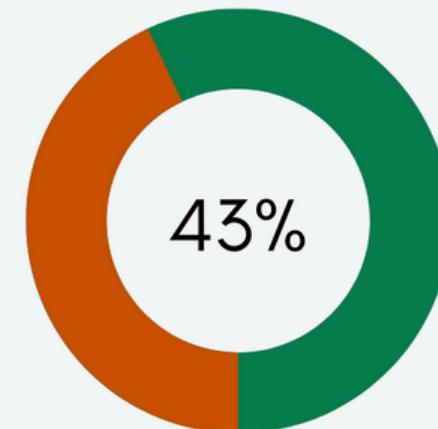
Input Feature 1

Input Feature 2

Run



Not available

**DOHMH
POV**

Percentage risk of committing a violation

C

Predicted Grade

Location ----- 68 Bergen St, Brooklyn

Food Type ----- American

Date of Opening----- Jan 25, 2024

Owner(s)----- Lisa Kudrow

Type of Violations

Similar Establishments



Possible violation chance and type

v1.0

Tell us about your business!

Name of business

What should we call you?

Location/ Adress

Input Street Adress

Food Type

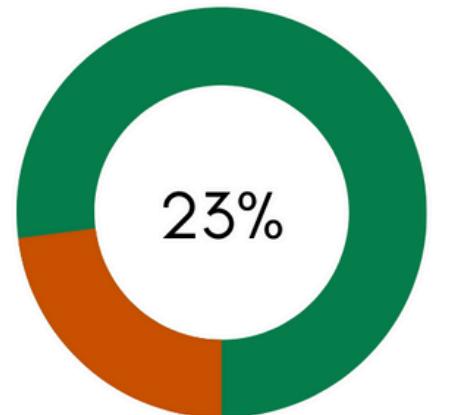
What are you cooking?

Date of Opening

month/ day/ year

Run

*Note, our algorithm predicts the likelihood that your establishment receives a violation based on historical data available to us. Use this as a general guideline to prepare your operations and stay ready for surprise inspections by the DOHMH



A

Predicted Grade

Based on our data, you have a **23%** chance of committing the following violations:

Type of Violations

See how similar restaurants in your community follow the DOHMH regulations!

Similar Establishments

**Restaurant's
POV**

FAQ



What to expect from an inspection?

Complete

BAN - A

END

