

# PUBLIC HEALTH INSPECTION NOTICE

# CLOSED

This establishment was found to have violations which constitute a danger to personal safety or has been found in violation in other provisions of applicable ordinances. As a result, this establishment has been ordered to be closed by the Public Health Department.

# Did diners make a better choice?

Zhiling Hu

Irene Wei

# Data

## NYC Restaurant Inspection Results

**NYC OpenData**

the most recent inspection for restaurants  
and college cafeterias

## Yelp Information Extracted from Yelp API

- Yelp rating
- price levels
- Category
- Text reviews



# Restaurants for clean eating

Of the Violation indicated on Health Inspection Data Set

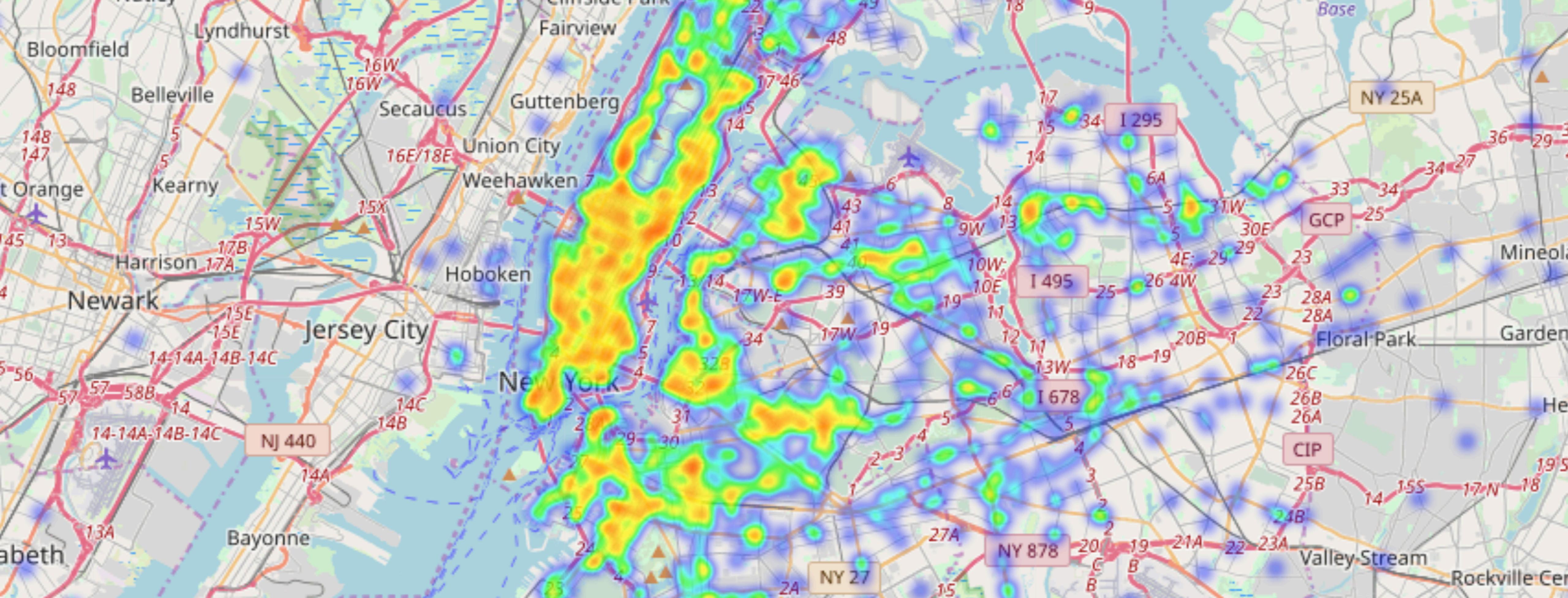
- There is no more than one violation.
- No Critical Violation
- Health inspection grade = A



# Popular Restaurants

Of the popularity indicated on Yelp dataset

➤ Check ins > 500

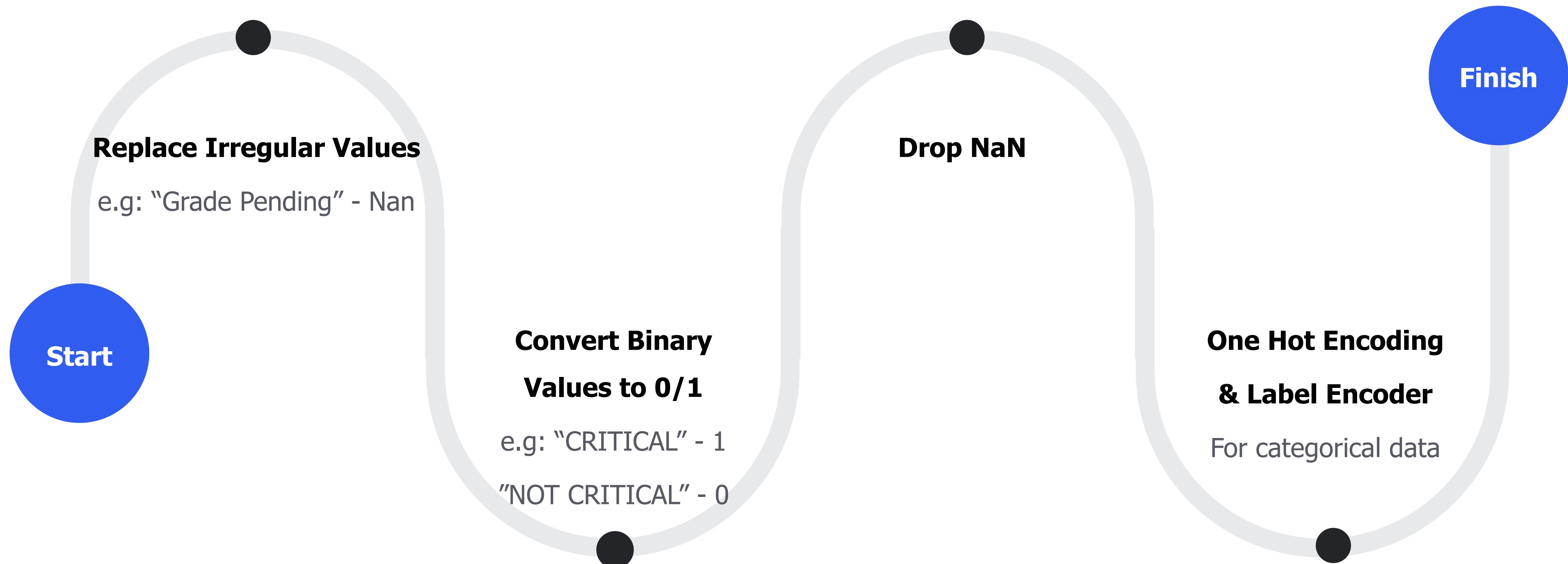


# Top Rated Restaurants

Of the ratings indicated on Yelp dataset

➤ Rating > 4.0

# PREPROCESS



# MERGED TABLE SAMPLE COLUMNS

CAMIS	Rating	Popularity	Price (\$)	BORO	CUISINE DESCRIPTION	GRADE
30075445	4.0	41	2	BRONX	Bakery	A
40356483	3.5	36	3	BROOKLYN	Hamburgers	A
40356731	2.0	99	1	BROOKLYN	Irish	A
40351028	3.0	37	2	MANHATTAN	Ice Cream, Gelato, Yogurt, Ices	A

# STATISTIC MODEL

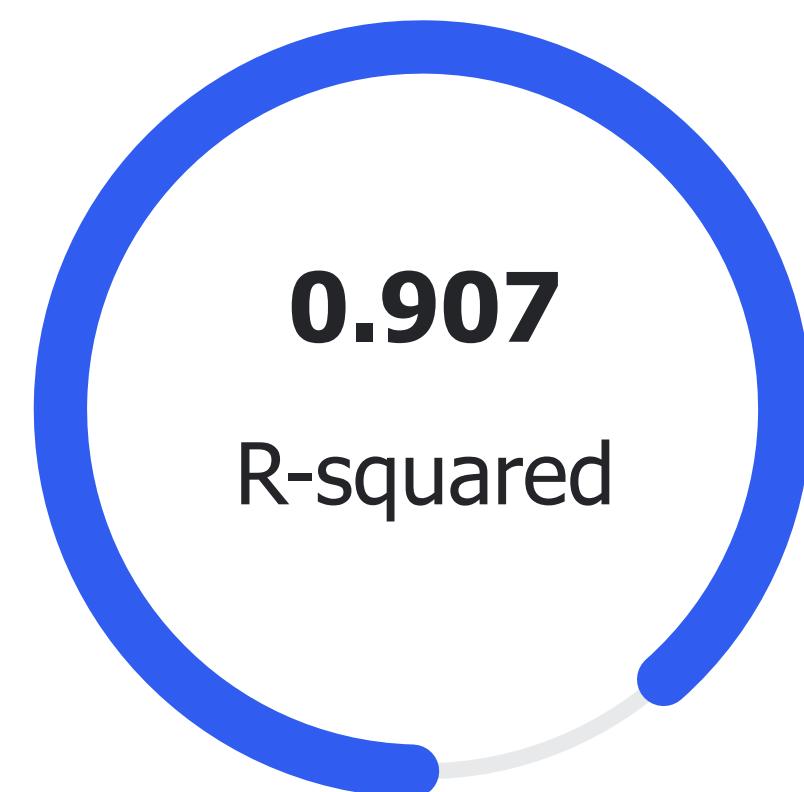
Two independent variables sets

- w/ inspections
- w/o inspections

One dependent variable:

- Rating

# Multiple Linear Regression



w/o Inspection Grade

this model explains 90.7% of  
the variance in our dependent  
variable

		coef	std err	t	P> t	[0.025	0.975]
	<b>BORO</b>	0.4641	0.007	63.696	0.000	0.450	0.478
	<b>Popularity</b>	0.0001	1.15e-05	11.486	0.000	0.000	0.000
	<b>Price</b>	1.1920	0.009	134.050	0.000	1.175	1.209
	<b>CUISINE DESCRIPTION</b>	0.0133	0.000	42.961	0.000	0.013	0.014

# Multiple Linear Regression

**0.919**

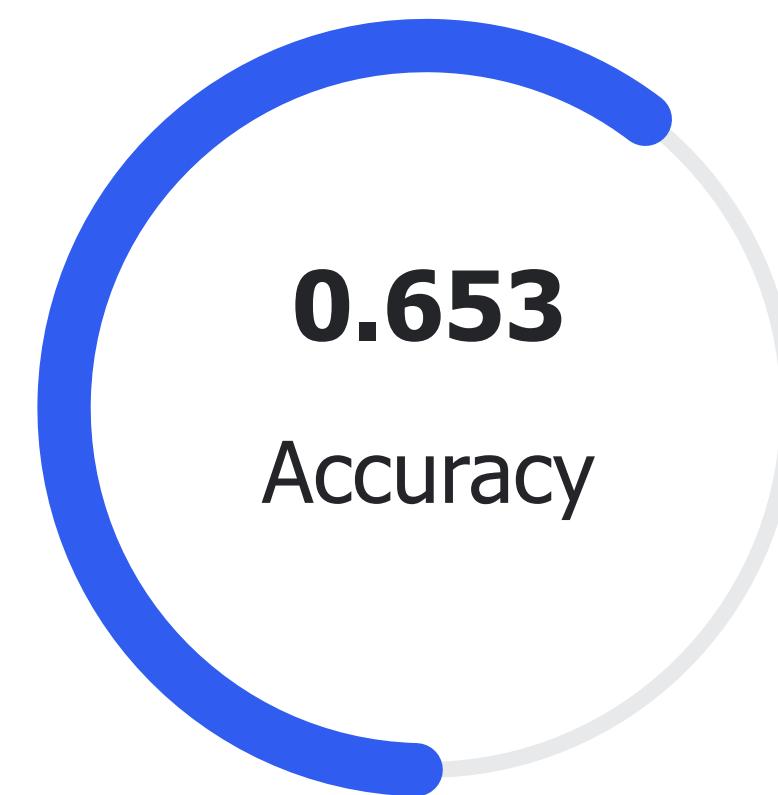
R-squared

w/ Inspection Grade

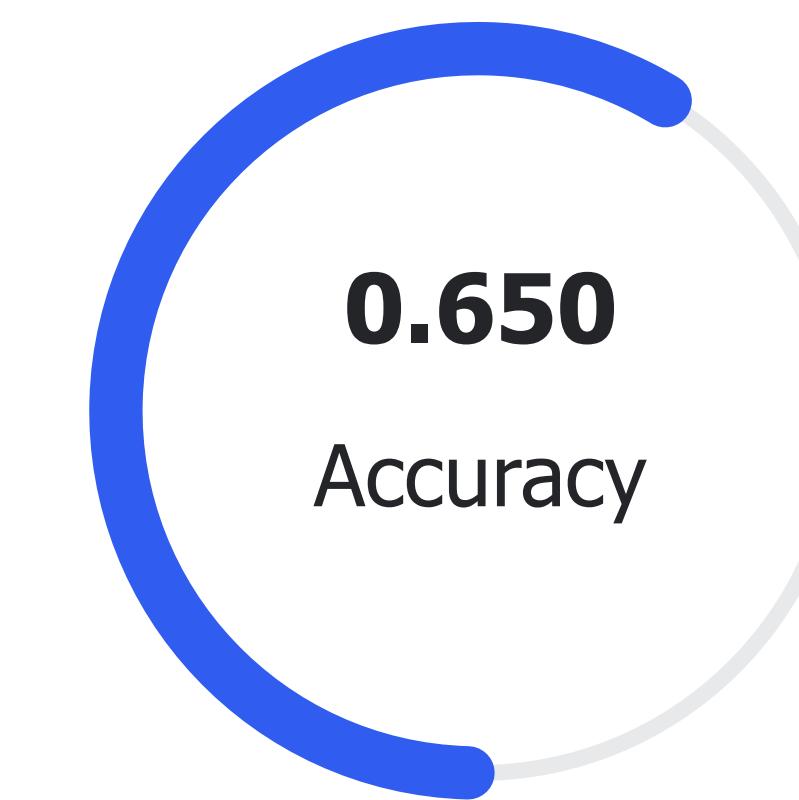
this model explains 91.9% of  
the variance in our dependent  
variable

		coef	std err	t	P> t	[0.025	0.975]
	BORO	0.3731	0.007	53.135	0.000	0.359	0.387
	Popularity	0.0001	1.07e-05	11.688	0.000	0.000	0.000
	SCORE	0.0666	0.001	52.698	0.000	0.064	0.069
	Price	0.9747	0.009	105.083	0.000	0.956	0.993
	CUISINE DESCRIPTION	0.0103	0.000	34.895	0.000	0.010	0.011

# Logistic Regression Model

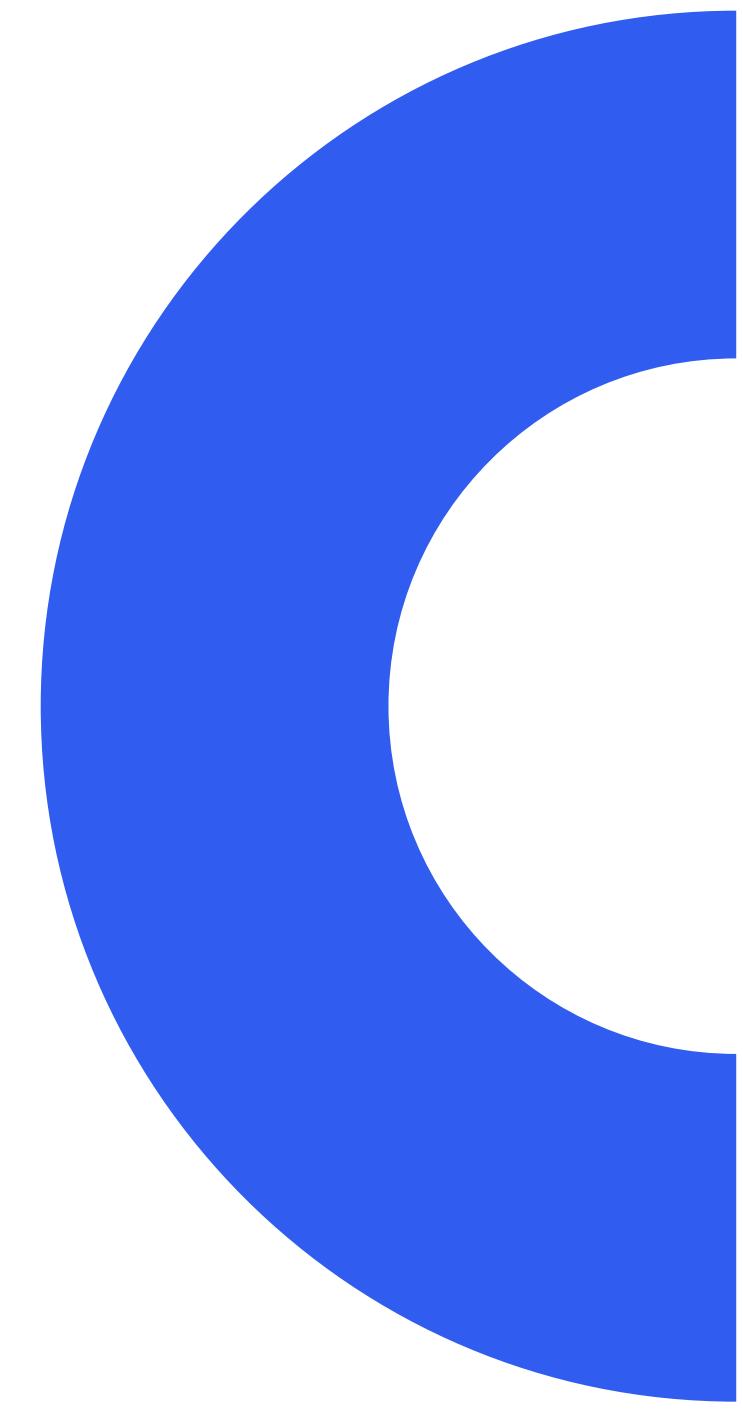
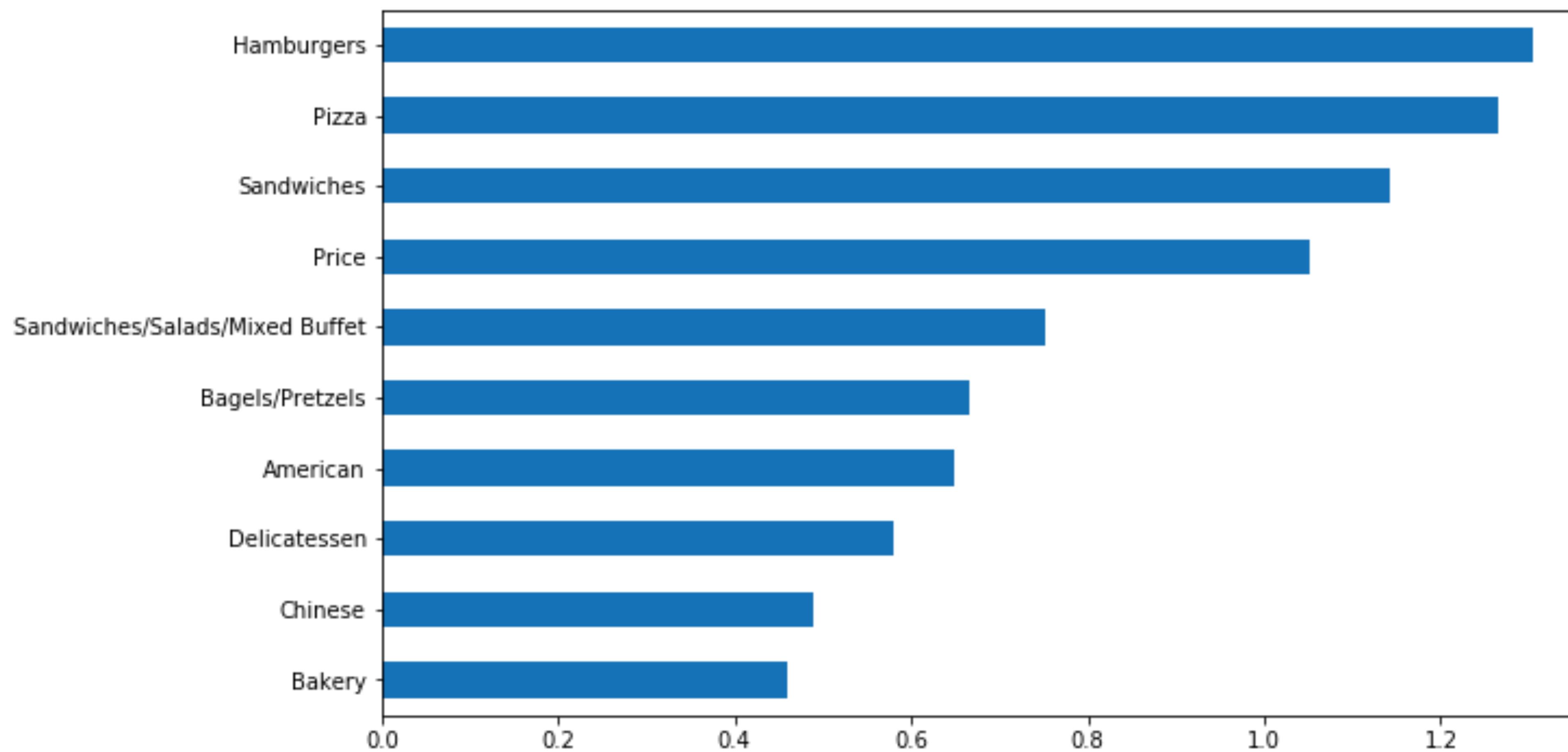


**w/ Inspection Grade**



**w/o Inspection Grade**

# Top 10 Weighted Features w/Grade



# Reflections

Z = Grade Pending

P = Grade Pending issued  
on re-opening following  
an initial inspection that  
resulted in a closure

