
Predicting NYC Restaurant Closures in the COVID-19 Pandemic

Eric Roberts & Nick Subic
November 13, 2020





Introduction

- The COVID-19 pandemic has dramatically affected the Restaurant Industry
 - 1 in 6 restaurants have closed nationwide
 - Approx. 3 million employees are out of work
 - Est. \$240 billion in sales lost by end of 2020
- Population density and global travel have increased the rate at which new infections (re)emerge



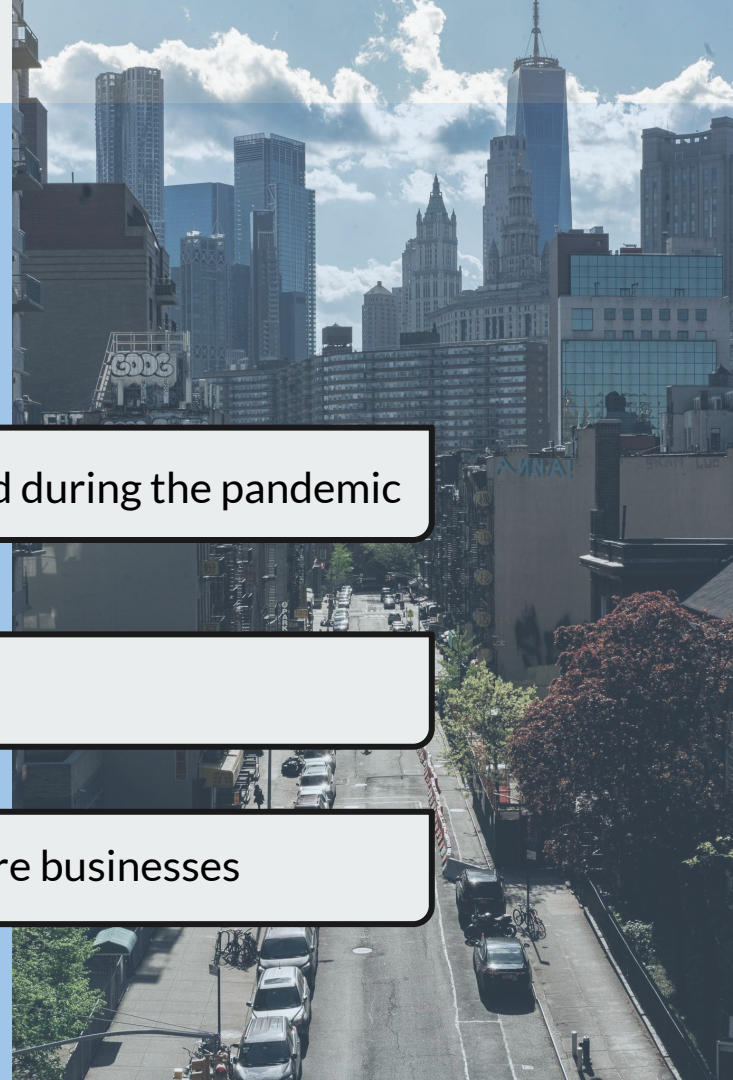


Project Goals

Identify common characteristics of restaurants closed during the pandemic

Find common attributes of surviving businesses

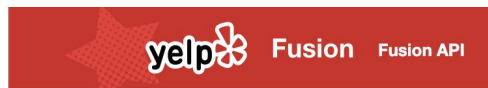
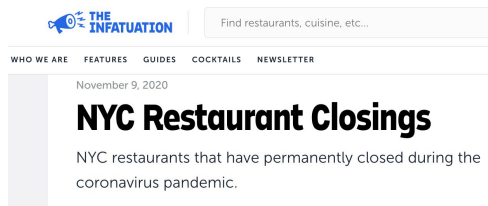
Make predictions and recommend strategies for future businesses





Data Sources

- COVID-19 related restaurant closures were compiled from lifestyle websites
- Characteristics of open and closed restaurants were gathered from Yelp's business search API
- COVID-19 safety precautions were scraped from Yelp's website
- Population density by zip code came from US Census data
- Restaurant density came from NYC Health restaurant inspection data



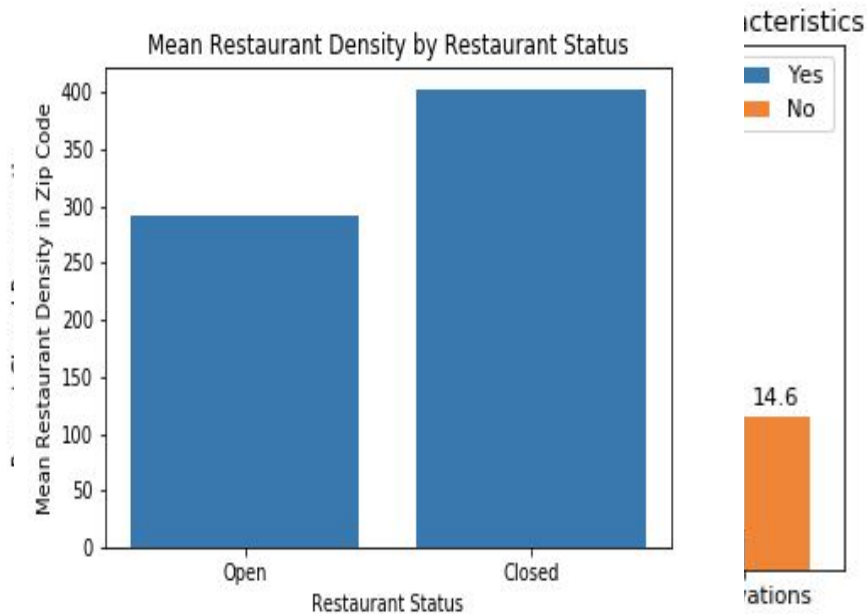
NYC OpenData



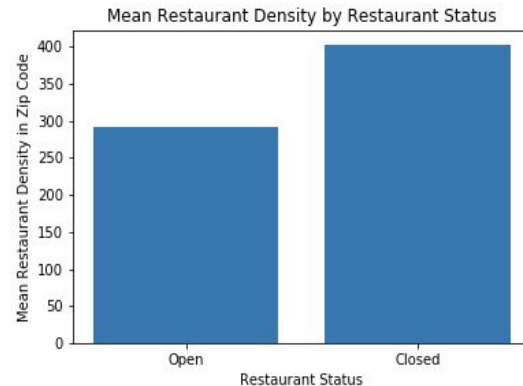
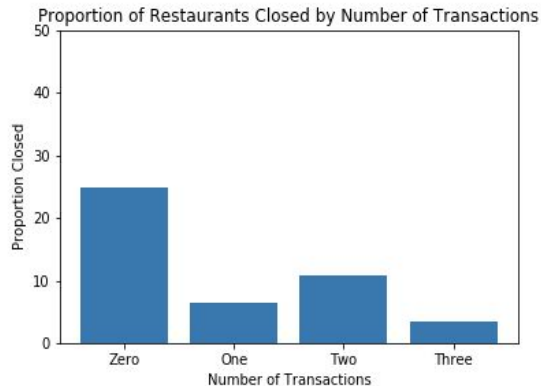
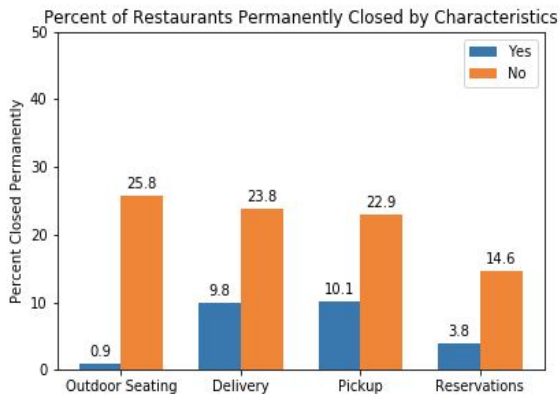
Sample and Analyses

- 187 restaurants closed since March 2020
- 1173 restaurants currently open
 - API calls for 35 restaurants by zip code were requested
 - Restricted to Manhattan, Brooklyn or Queens, with 16 or more reviews
- Univariable and bivariable analyses
- Modeled with Logistic Regression and Decision Trees
- Predicted closures with over 80% accuracy

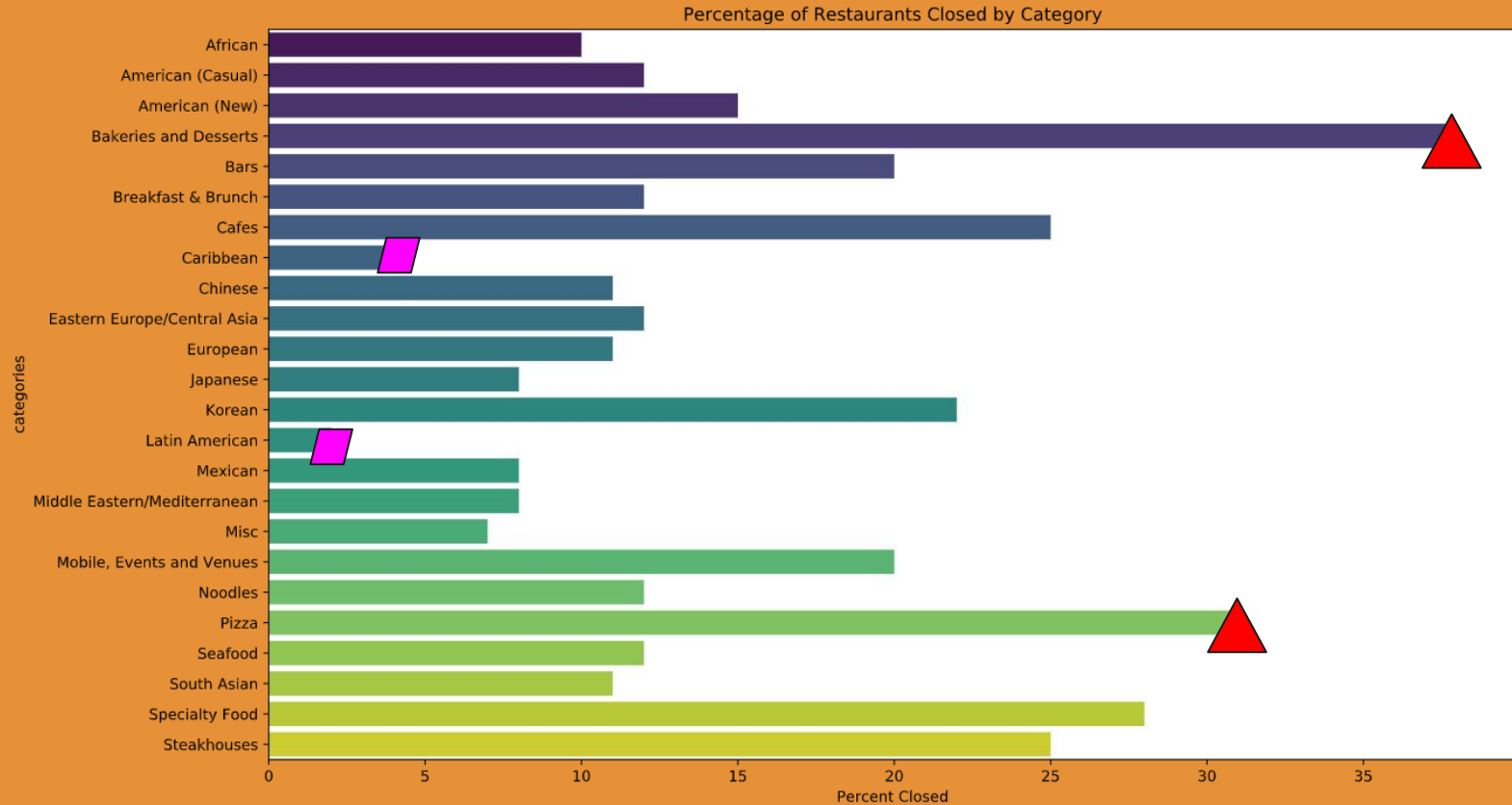
Exploratory Data Analyses



Exploratory Data Analyses



Characteristics associated with closure - no outdoor seating, no delivery, no takeout, no reservations, offering none of the three transaction types, and being in a zip code with a higher mean restaurant density.



Closures were most common with Bakeries and Pizzerias

Caribbean and Latin American restaurants had the fewest closures



Models

Logistic Regression and Decision Trees

Balanced Class Weights

Recall and Accuracy evaluation metrics

Achieved a recall score of .91 and accuracy of .8



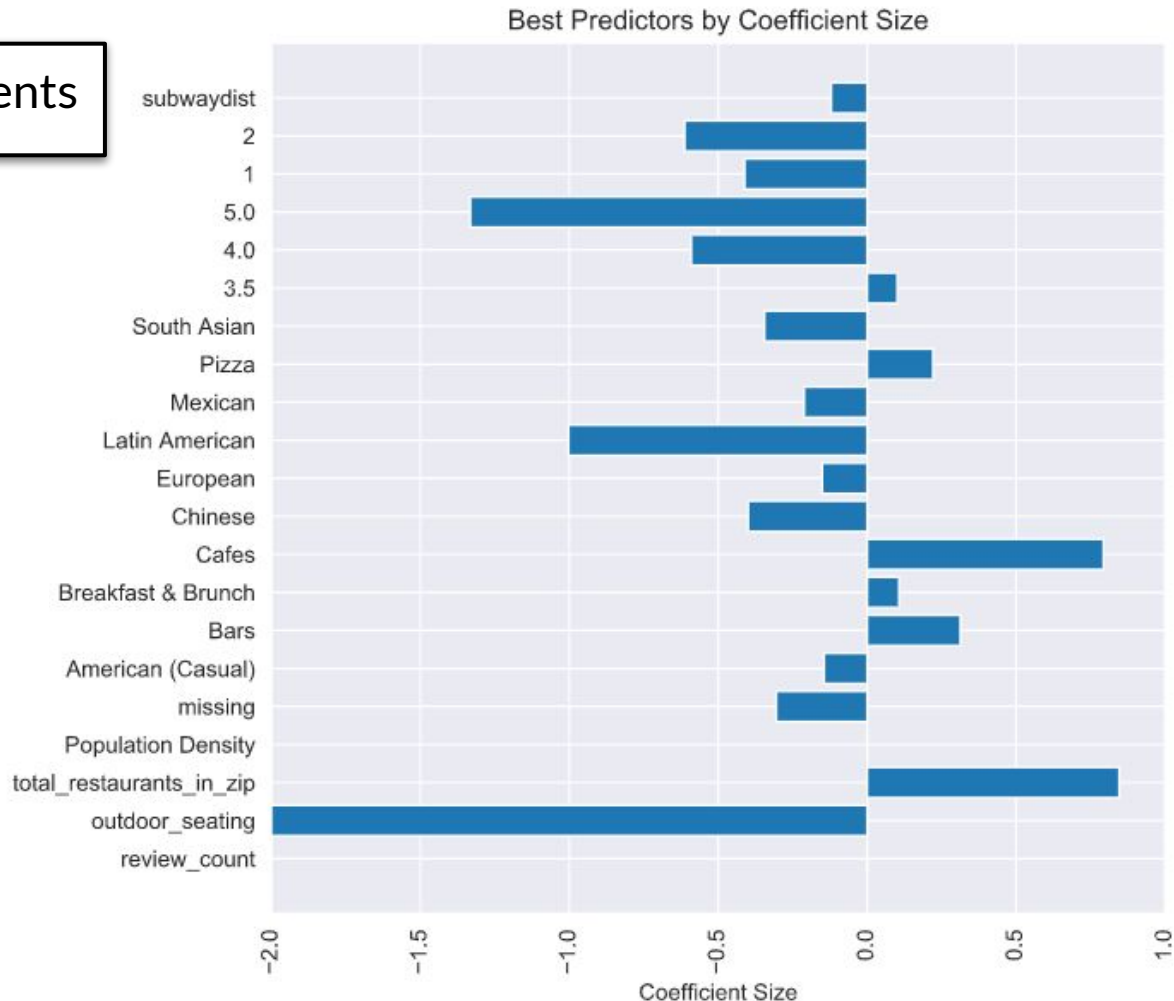
Evaluating the Model Coefficients

Main Survival Factors:

- Outdoor Seating
- 5.0 Yelp rating
- Latin American Cuisine
- 2 or More Transaction Types

Main Closure Attributes:

- Cafe
- Bars
- Pizza
- High number of restaurants in zip code





Our Findings:

Outdoor seating is a make or break factor for most businesses

Higher Yelp ratings improve odds of survival

A combination of delivery, takeout and reservations is ideal



Next Steps

Working with data in seasons with no outdoor seating

Gathering info on indoor modifications

Finding more data to generate a sample with less bias

Incorporating features for restaurants opened during pandemic



Any Questions?