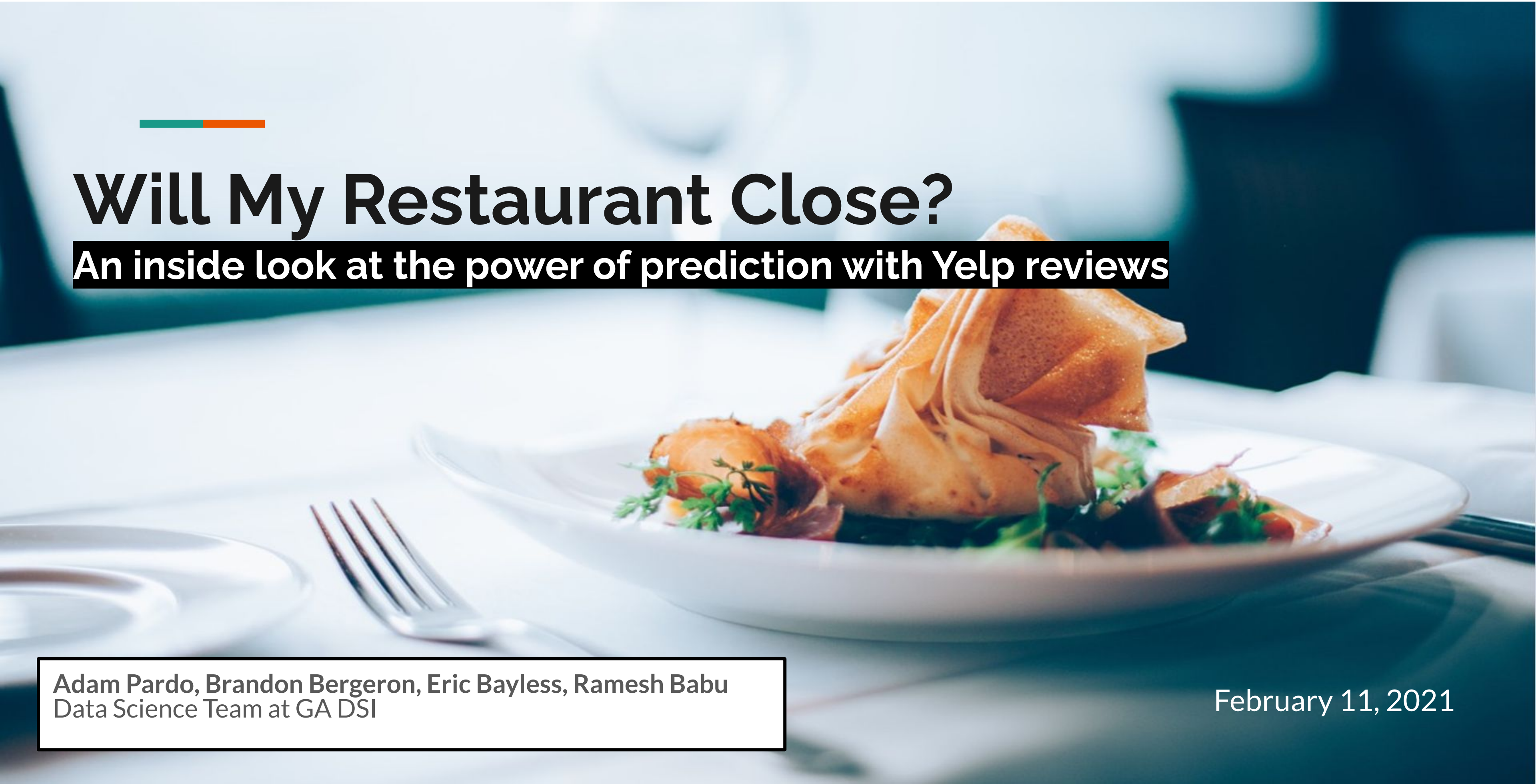


Will My Restaurant Close?

An inside look at the power of prediction with Yelp reviews

Adam Pardo, Brandon Bergeron, Eric Bayless, Ramesh Babu
Data Science Team at GA DSI

February 11, 2021

Problem Statement

The restaurant industry is extremely competitive with 80% of restaurants closing within their first 5 years of operation.

Looking solely at Yelp reviews, does the text have any predictive power into whether or not a restaurant's operating status is open or closed?

Audience

A group of investors, with a semi-technical background, approached our team and asked for help in performing data analysis.

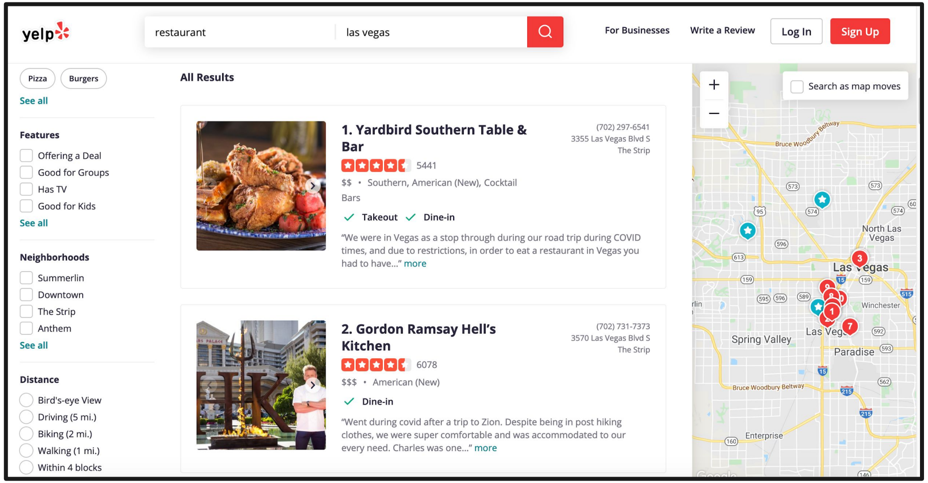
- Provided our team with a Yelp dataset
- Asked us to look into relationship between Yelp reviews and restaurant's operating status.

Agenda

1. Introduction on Yelp
2. Data Collection
3. Exploratory Data Analysis (EDA)
4. Modeling
5. Conclusions / Recommendations
6. Future Areas of Focus
7. Questions

Introduction on Yelp

“**Yelp connects people with great local businesses.** With unmatched local business information, photos and review content, Yelp provides a one-stop local platform for consumers to discover, connect and transact with local businesses of all sizes by making it easy to request a quote, join a waitlist, and make a reservation, appointment or purchase.”



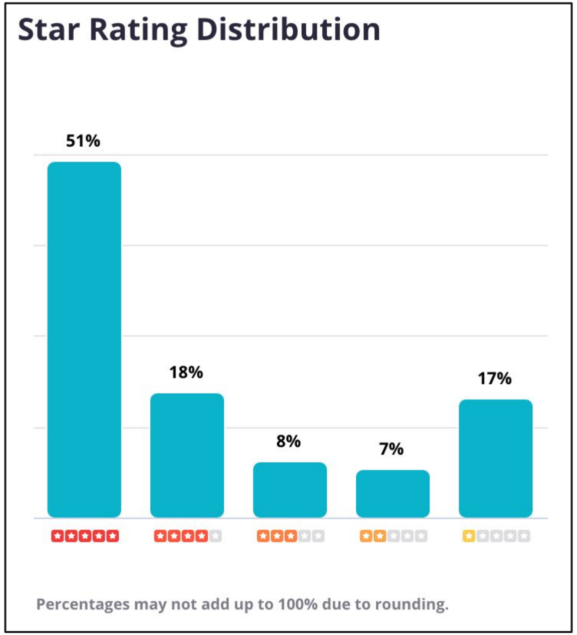
restaurant | las vegas

All Results

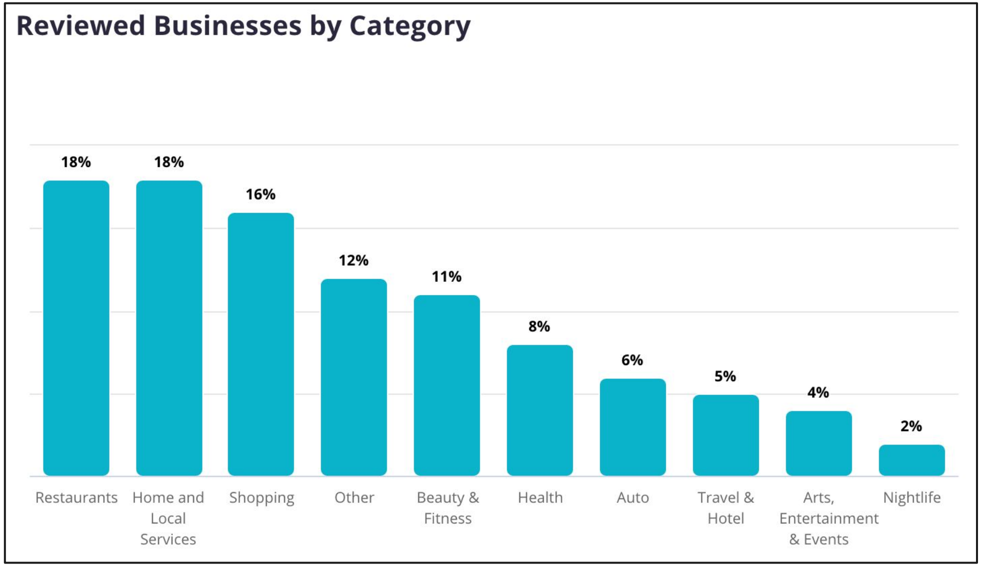
1. Yardbird Southern Table & Bar
5441
$$ • Southern, American (New), Cocktail Bars
✓ Takeout ✓ Dine-in
(702) 297-6541
3355 Las Vegas Blvd S
The Strip
“We were in Vegas as a stop through during our road trip during COVID times, and due to restrictions, in order to eat a restaurant in Vegas you had to have...” more

2. Gordon Ramsay Hell’s Kitchen
6078
$$$ • American (New)
✓ Dine-in
(702) 731-7373
3570 Las Vegas Blvd S
The Strip
“Went during covid after a trip to Zion. Despite being in post hiking clothes, we were super comfortable and was accommodated to our every need. Charles was one...” more

Introduction on Yelp (cont.)

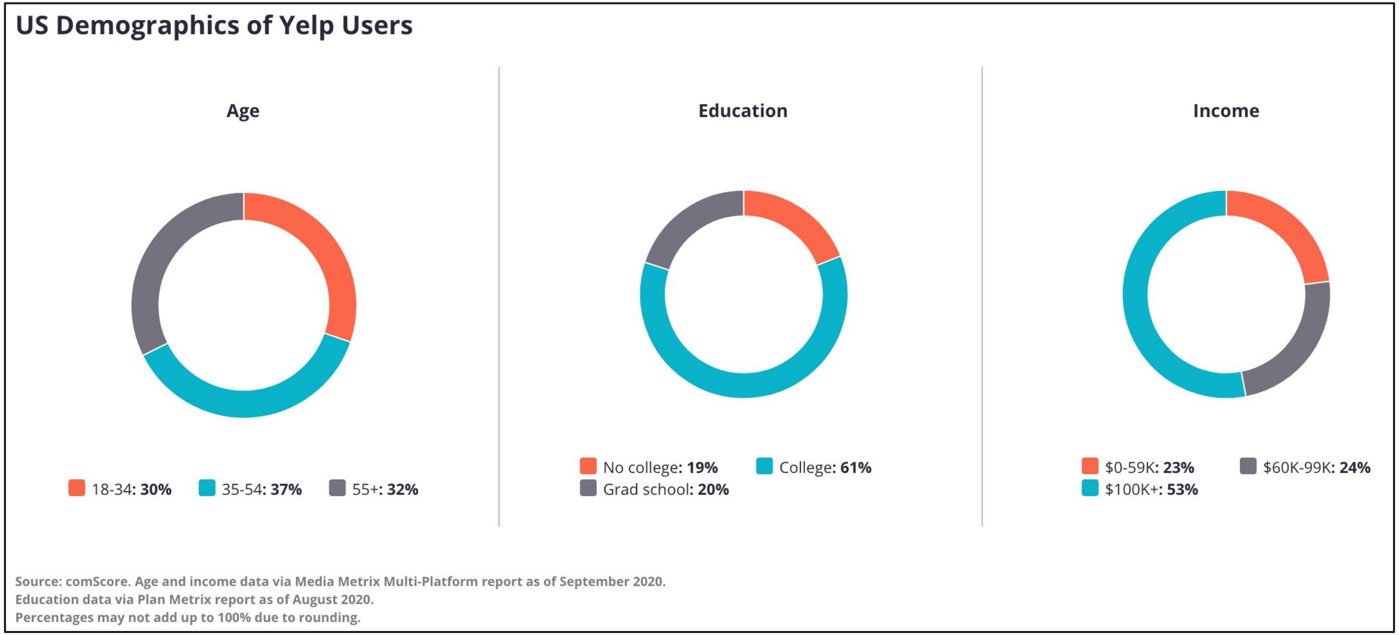

Star Rating Distribution

Percentages may not add up to 100% due to rounding.


Reviewed Businesses by Category

Source: https://www.yelp-press.com/company/fast-facts/default.aspx

Introduction on Yelp (cont.)

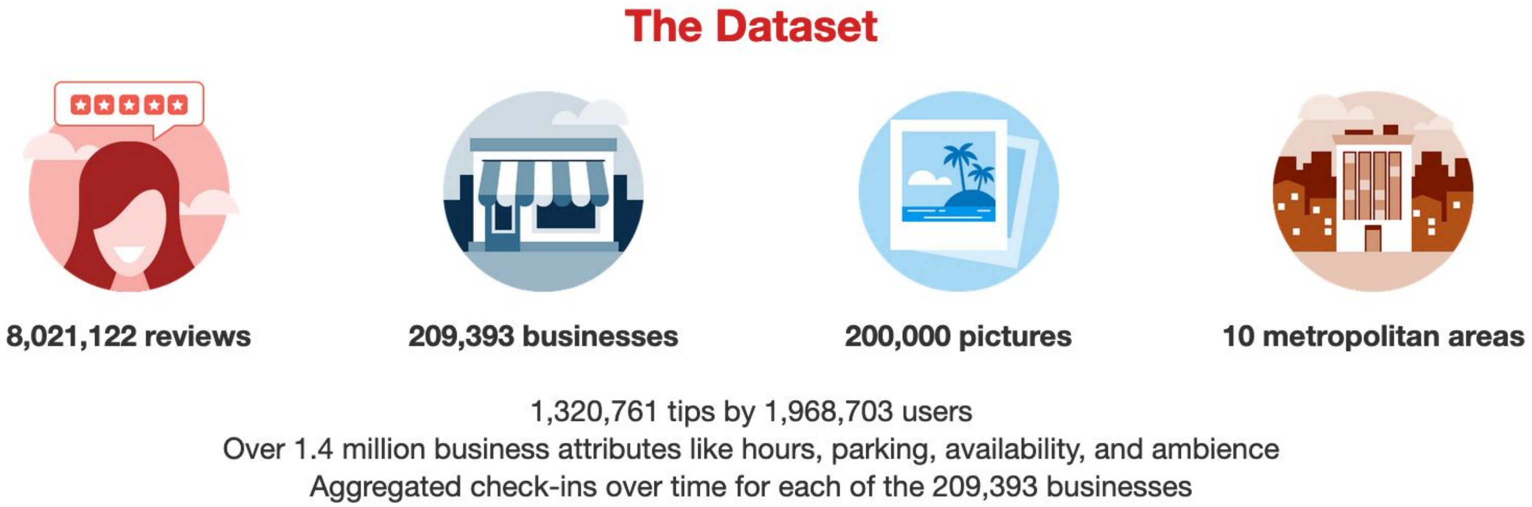


US Demographics of Yelp Users

Source: comScore. Age and income data via Media Metrix Multi-Platform report as of September 2020.
Education data via Plan Metrix report as of August 2020.
Percentages may not add up to 100% due to rounding.

Source: https://www.yelp-press.com/company/fast-facts/default.aspx

Data Collection



The Dataset

8,021,122 reviews

209,393 businesses

200,000 pictures

10 metropolitan areas

1,320,761 tips by 1,968,703 users
Over 1.4 million business attributes like hours, parking, availability, and ambience
Aggregated check-ins over time for each of the 209,393 businesses

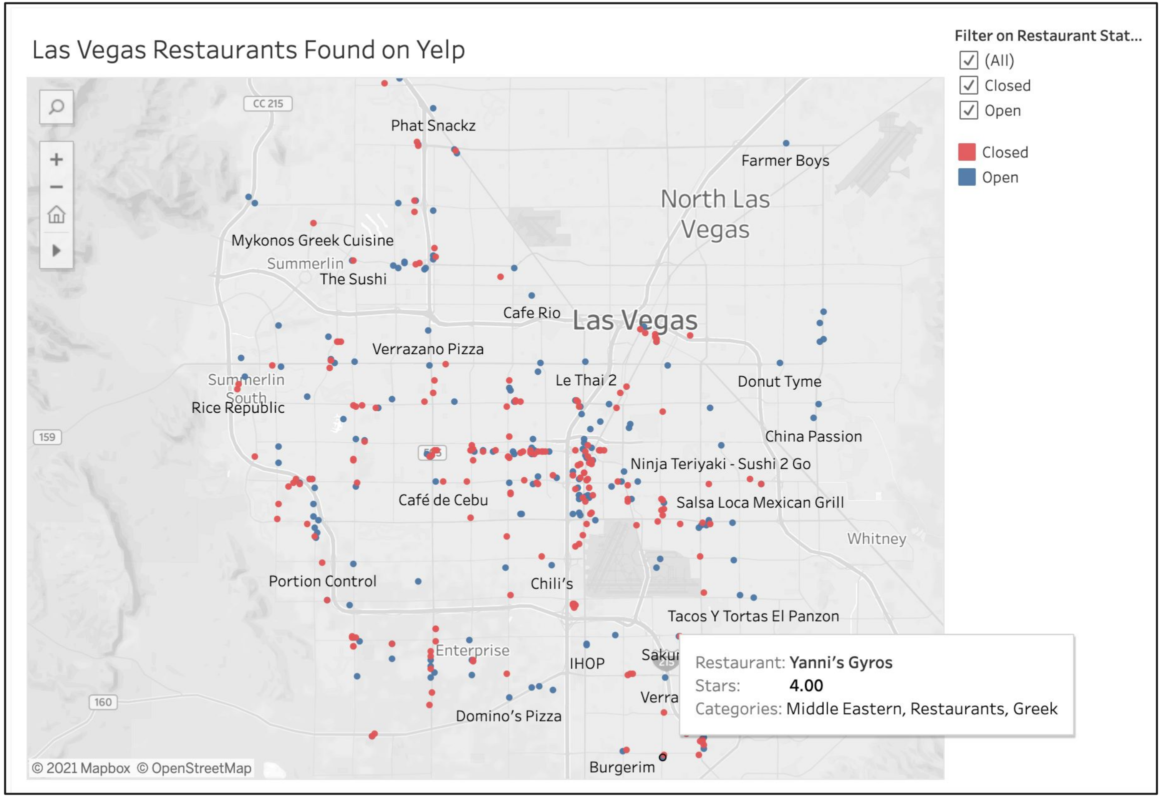
Data Collection (cont.)

- Las Vegas
  - City with the largest amount of reviews in Yelp dataset
  - Dynamic city with high-level of tourism

- Data pull - Las Vegas restaurants that had between 100 and 300 reviews

- Result
  - EDA: 400 restaurants with 100-300 reviews
  - Modeling: all restaurants with 100-300 reviews

Data Collection (cont.)



Las Vegas Restaurants Found on Yelp

Filter on Restaurant Stat...

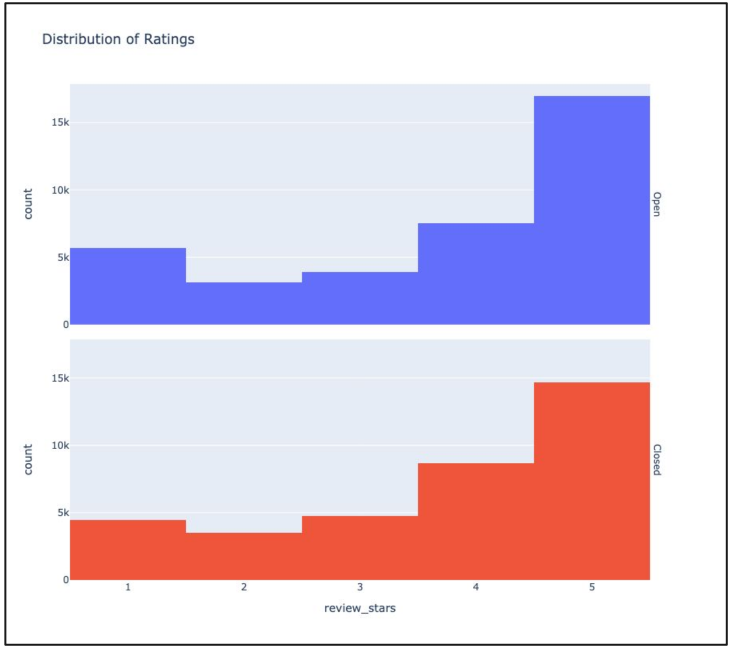
☑ (All)
☑ Closed
☑ Open

🟥 Closed
🟦 Open

Restaurant: Yanni’s Gyros
Stars: 4.00
Categories: Middle Eastern, Restaurants, Greek

EDA

Restaurants in Las Vegas



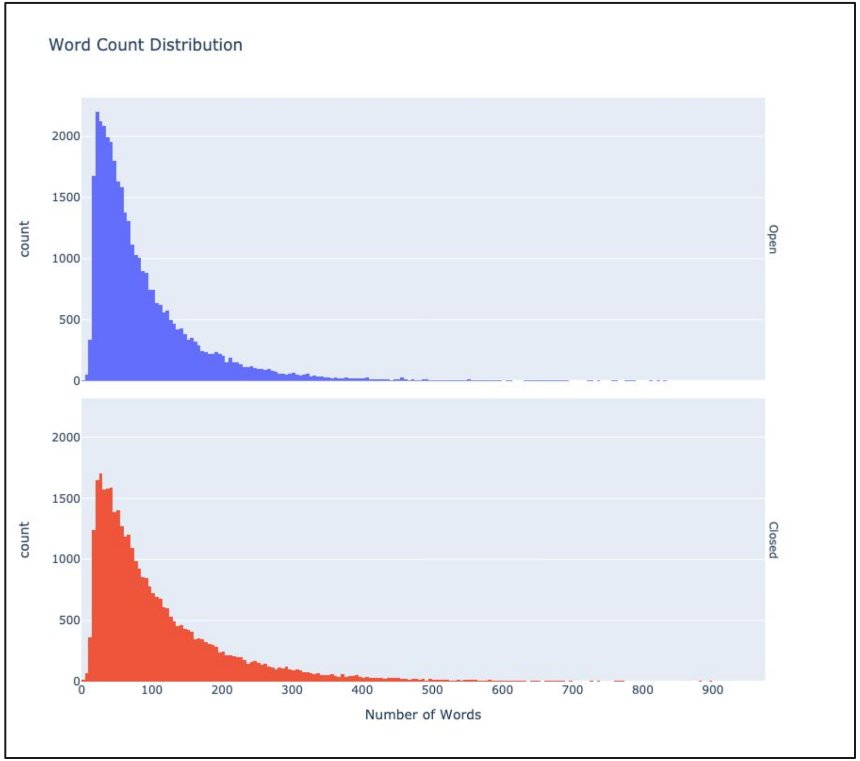
Open Restaurants

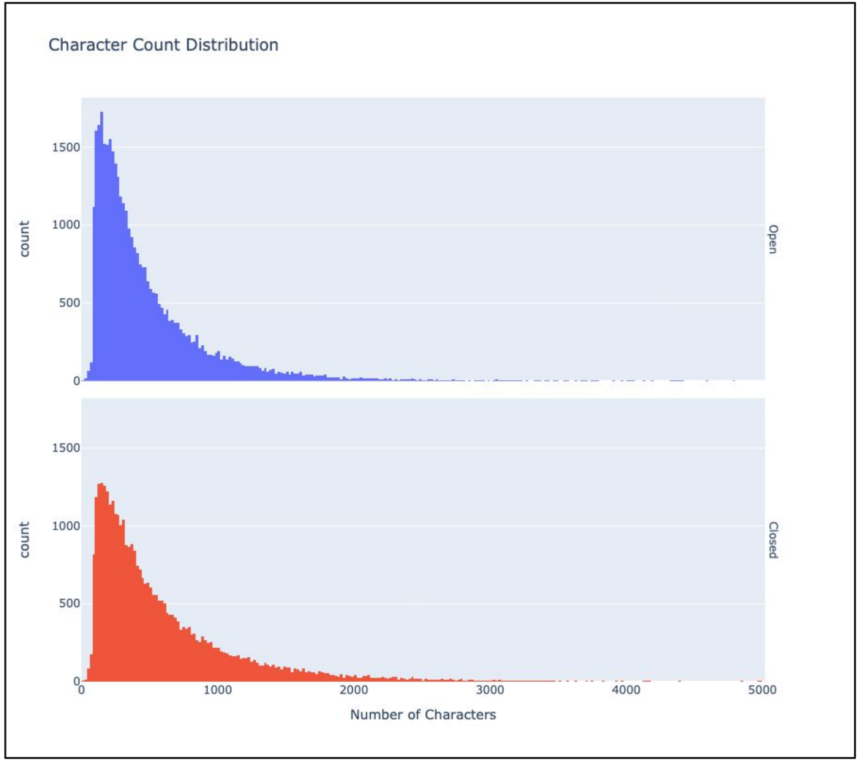
- 200 restaurants
- 37,211 reviews
- Average Rating: 3.7 stars

Closed Restaurants

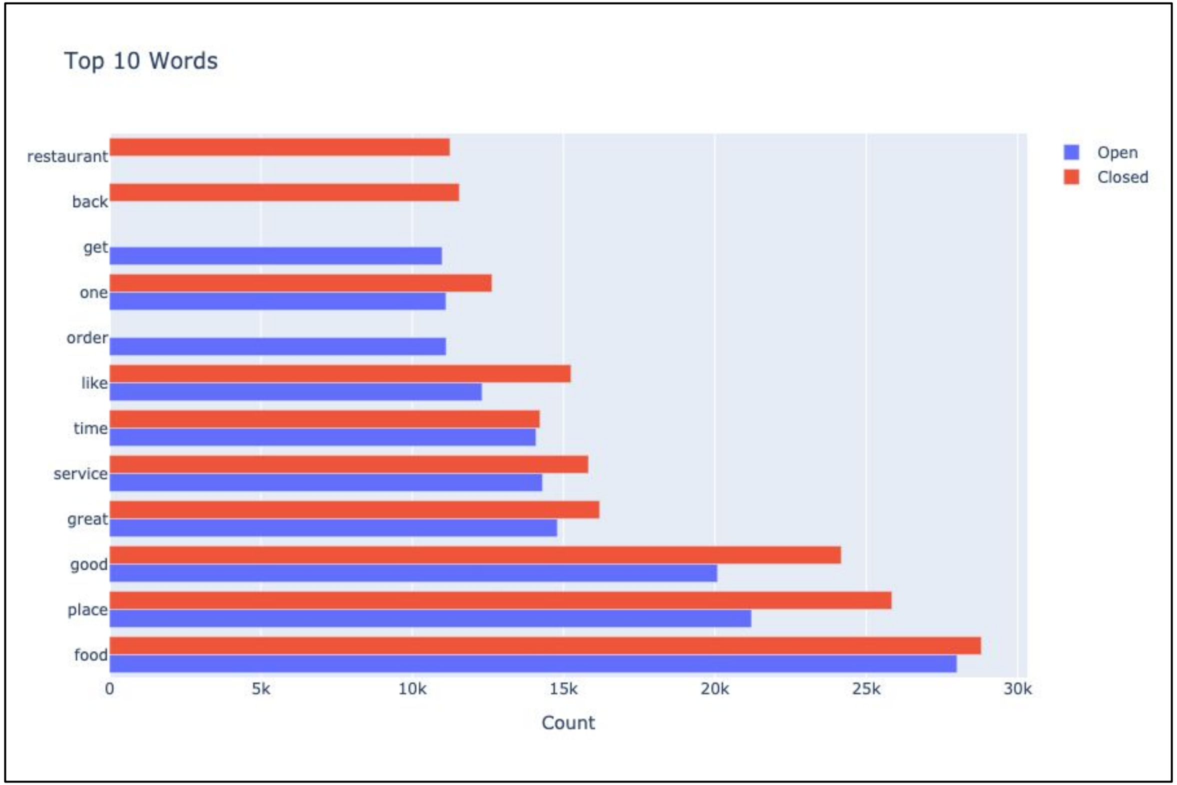
- 200 restaurants
- 36,067 reviews
- Average Rating: 3.7 stars

Restaurants in Las Vegas (cont.)

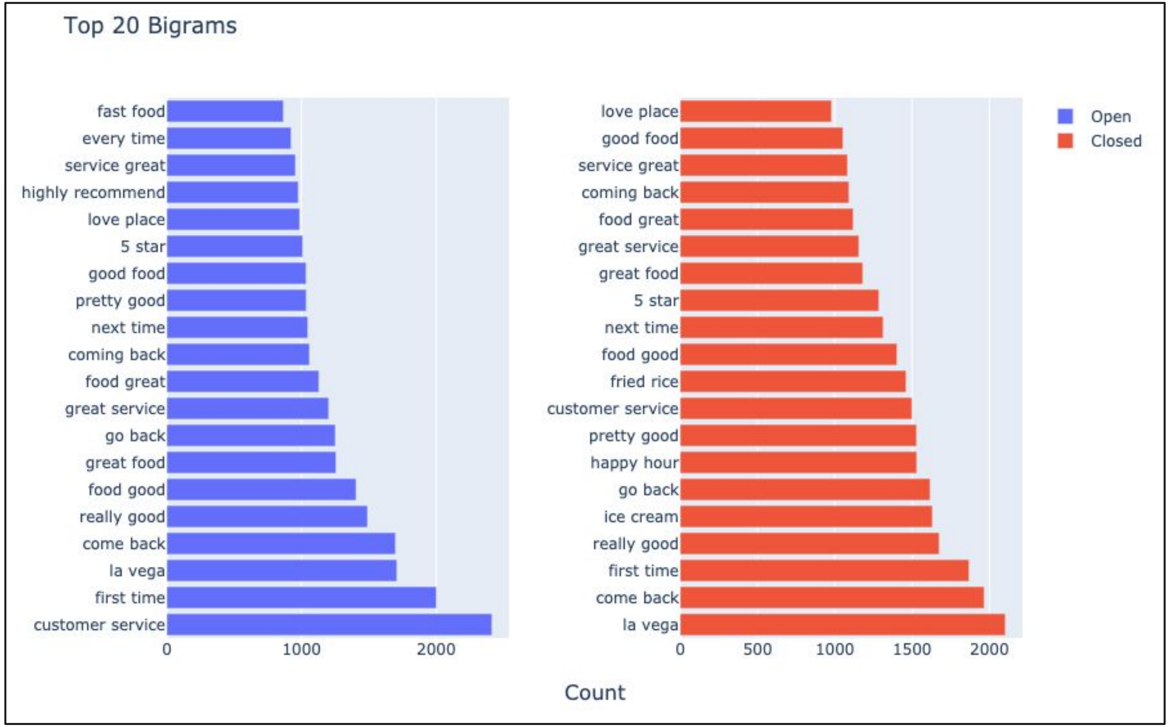




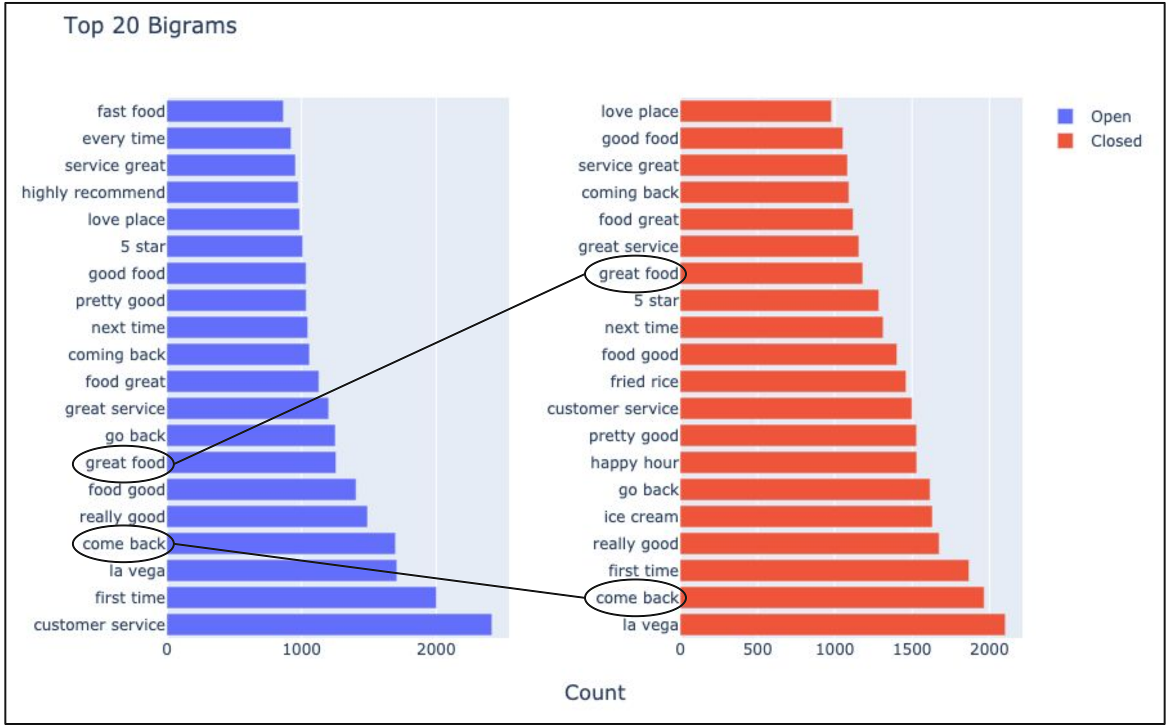
Restaurants in Las Vegas (cont.)



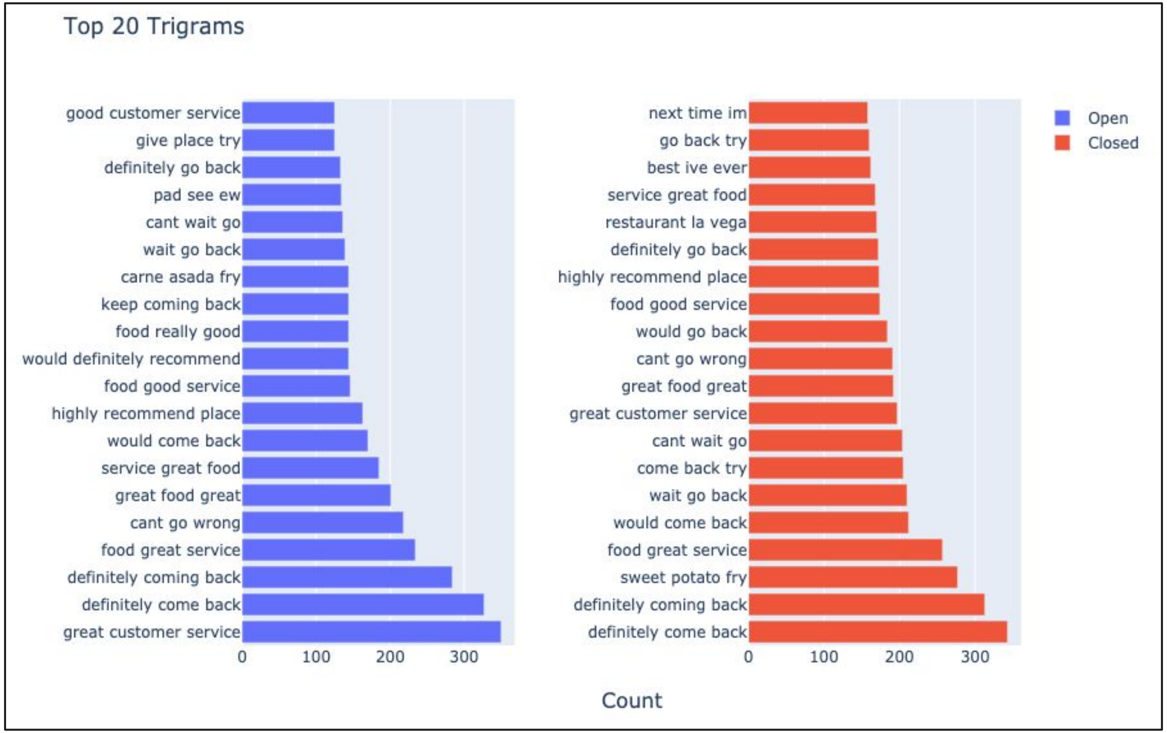
Restaurants in Las Vegas (cont.)



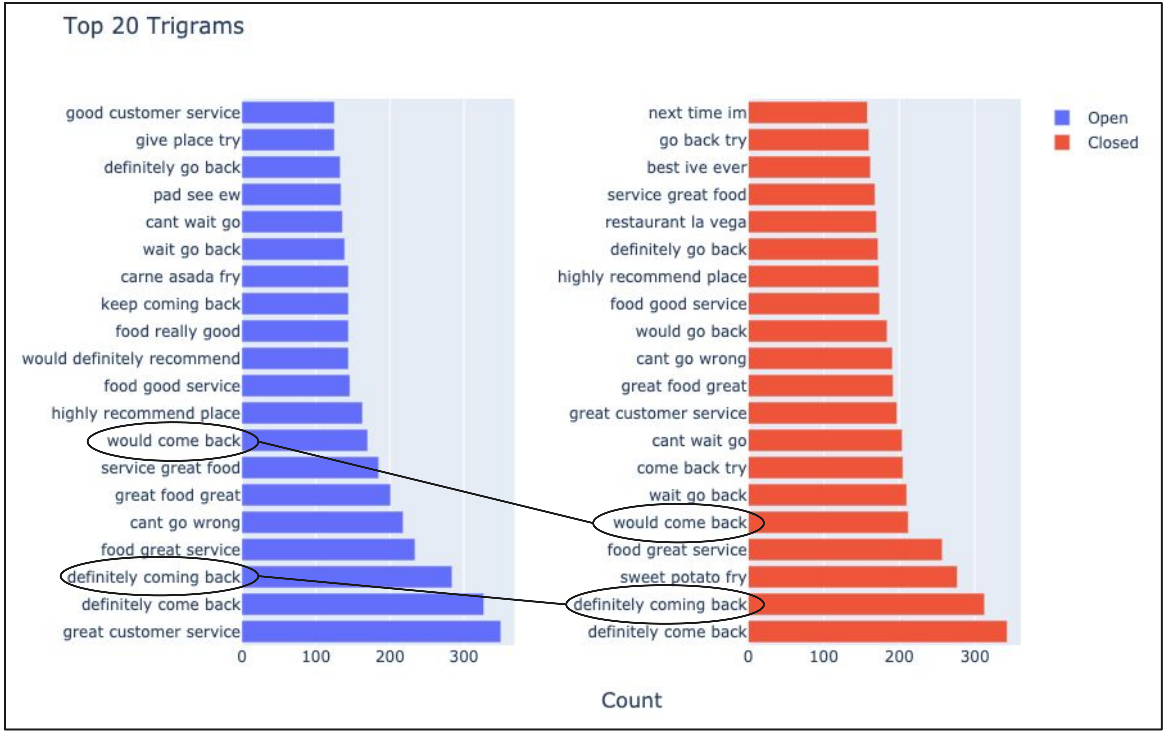
Restaurants in Las Vegas (cont.)



Restaurants in Las Vegas (cont.)



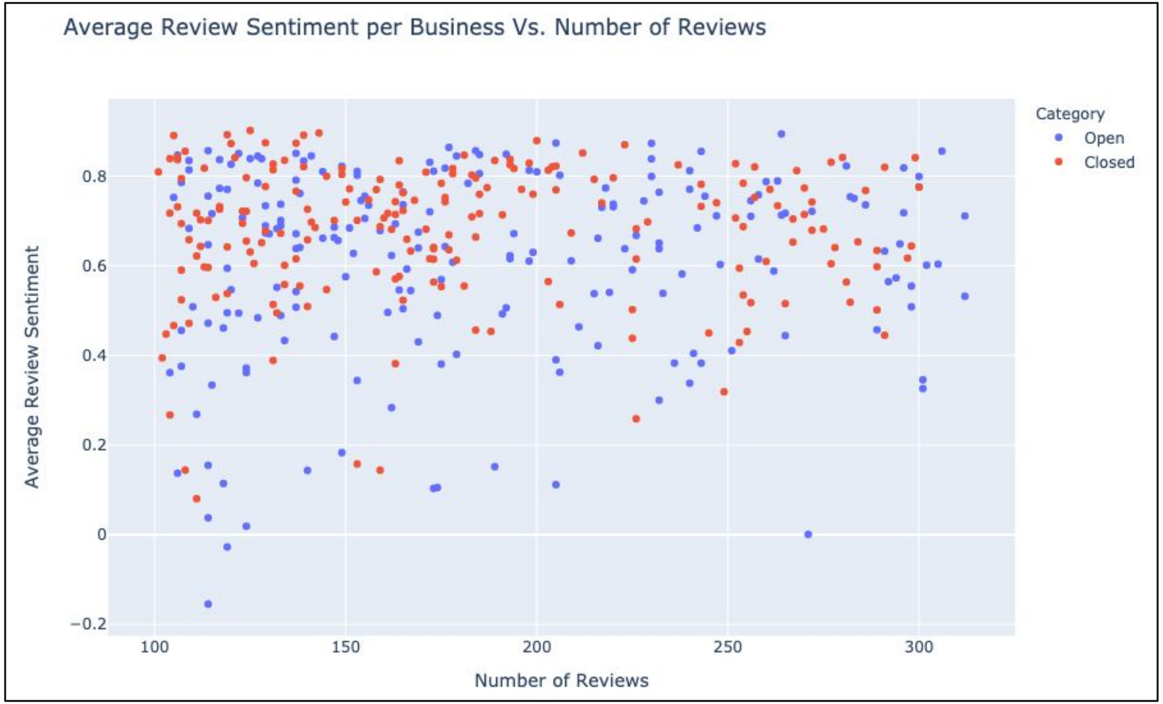
Restaurants in Las Vegas (cont.)



Restaurants in Las Vegas (cont.)

<ins>Sentiment Analysis</ins>

- Evenly distributed between open and closed restaurants
- Wider spread of sentiment for restaurants with lower number of reviews



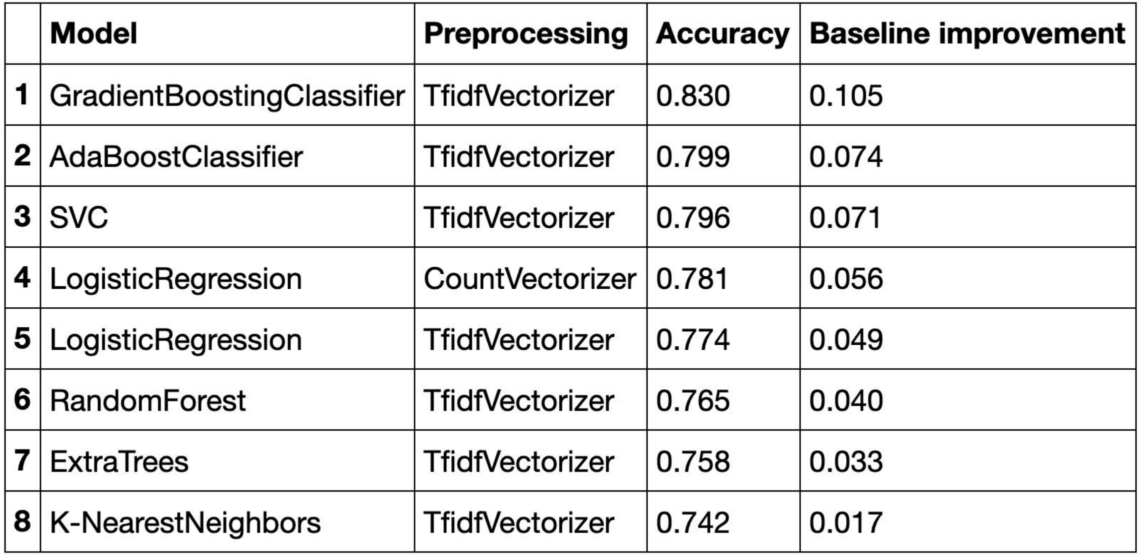
Modeling

Modeling - First Basic Models

The following algorithms were used:

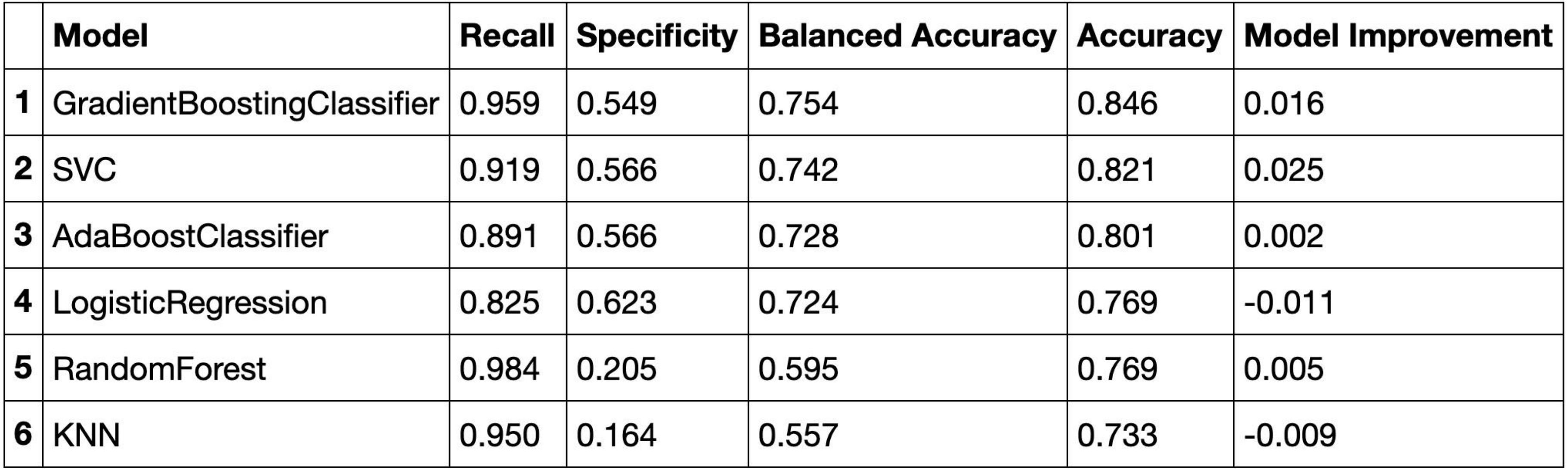
- LogisticRegression
- MultinomialNB
- RandomForest
- ExtraTrees
- K-NearestNeighbors
- SVC
- AdaBoostClassifier
- GradientBoostingClassifier

Baseline accuracy score - 0.72



	Model	Preprocessing	Accuracy	Baseline improvement
1	GradientBoostingClassifier	TfidfVectorizer	0.830	0.105
2	AdaBoostClassifier	TfidfVectorizer	0.799	0.074
3	SVC	TfidfVectorizer	0.796	0.071
4	LogisticRegression	CountVectorizer	0.781	0.056
5	LogisticRegression	TfidfVectorizer	0.774	0.049
6	RandomForest	TfidfVectorizer	0.765	0.040
7	ExtraTrees	TfidfVectorizer	0.758	0.033
8	K-NearestNeighbors	TfidfVectorizer	0.742	0.017

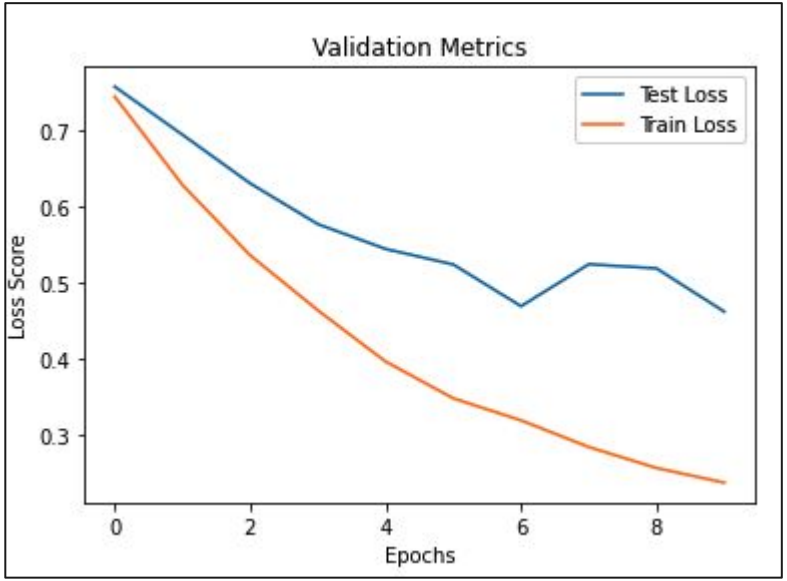
Modeling (cont.)



	Model	Recall	Specificity	Balanced Accuracy	Accuracy	Model Improvement
1	GradientBoostingClassifier	0.959	0.549	0.754	0.846	0.016
2	SVC	0.919	0.566	0.742	0.821	0.025
3	AdaBoostClassifier	0.891	0.566	0.728	0.801	0.002
4	LogisticRegression	0.825	0.623	0.724	0.769	-0.011
5	RandomForest	0.984	0.205	0.595	0.769	0.005
6	KNN	0.950	0.164	0.557	0.733	-0.009

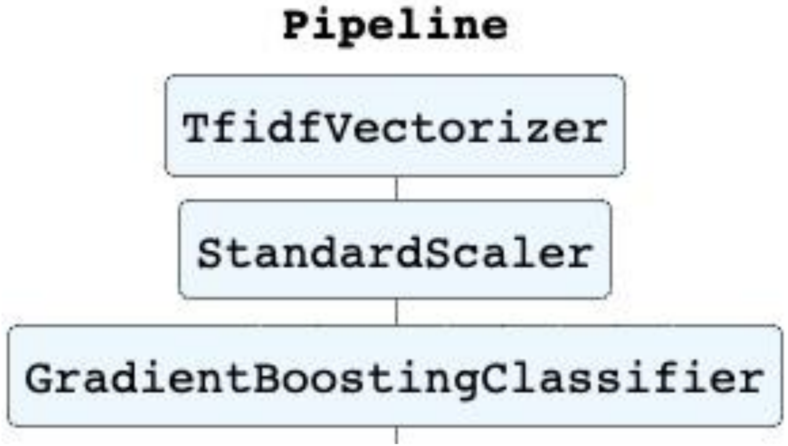
Neural Network Model

- Grid Search - best parameters
  - Keras Classifier
    - Dropout: 0.5
    - Regularization: L2
    - Units: 128
- Scores
  - Accuracy: 0.832
  - Balanced Accuracy: 0.709



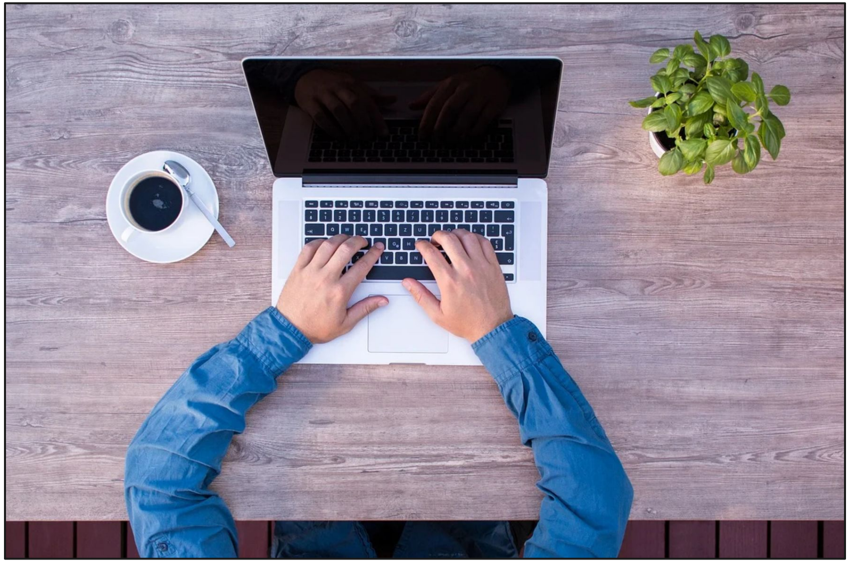
Overall Best Model - GradientBoostingClassifier

- Model parameters:
  - Estimator parameters
    - Max depth = 3
    - Min samples leaf = 20
  - TfidfVectorizer parameters
    - Max Features: 1000
    - Ngram Range: (1, 1)
    - Stop Words: yes, english
- Scores
  - Accuracy: 0.846
  - Balanced Acc: 0.754



Data Limitations / Constraints

- Dataset was large: over 10GB
- Only used restaurants with 100-300 reviews
- Data only current up to the end of 2019
- Reviews pulled from a Yelp provided dataset (unknown missing)



Conclusions / Recommendations

Conclusions / Recommendations

- While a lot can be learned from restaurant reviews, they are not very effective predictors of whether or not a restaurant will close

- Initial modeling efforts led us to believe that customer service played an important role, something to look into

Future Areas of Focus

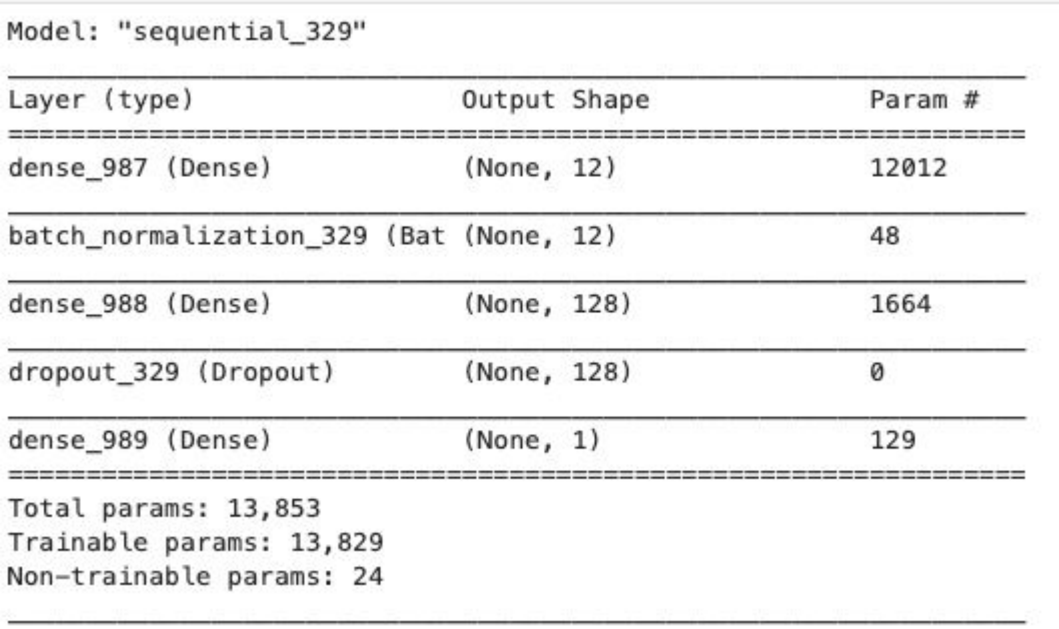
- More advanced NLP methods for better understanding of text
- Model on all restaurants in given city
- Model on other business types
- Expand our modeling efforts to new cities
- Gather recent reviews to observe effect of COVID on our models

Questions?

BACKUP

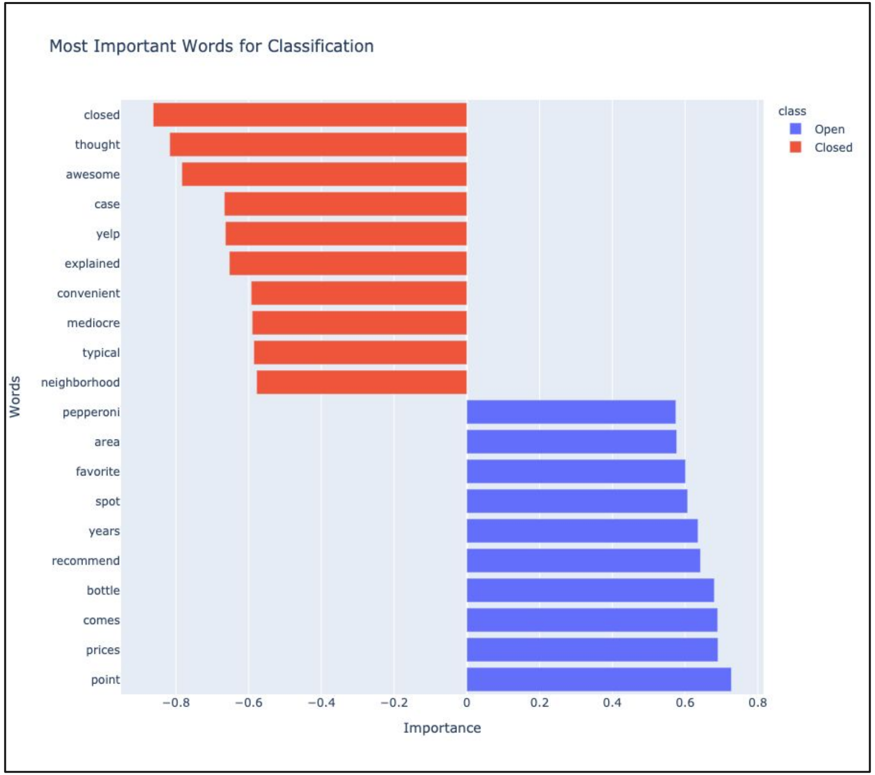
1. Restaurant Statistics
   a. https://www.fsrmagazine.com/expert-takes/restaurant-profitability-and-failure-rates-what-you-need-know
2. Yelp Information
   a. https://www.yelp-press.com/company/fast-facts/default.aspx
3. Yelp Dataset
   a. https://www.yelp.com/dataset
4. Tableau Dashboard
   a. https://public.tableau.com/profile/adam.pardo#!/vizhome/yelp_restaurants/Dashboard1?publish=yes

BACKUP - Neural Network Model



```
Model: "sequential_329"
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
dense_987 (Dense)            (None, 12)                12012
_________________________________________________________________
batch_normalization_329 (Bat (None, 12)                48
_________________________________________________________________
dense_988 (Dense)            (None, 128)               1664
_________________________________________________________________
dropout_329 (Dropout)        (None, 128)               0
_________________________________________________________________
dense_989 (Dense)            (None, 1)                 129
=================================================================
Total params: 13,853
Trainable params: 13,829
Non-trainable params: 24
_________________________________________________________________
```

BACKUP Feature Importance - Open vs Closed Restaurants


Most Important Words for Classification
class
🟦 Open
🟥 Closed
Words
closed
thought
awesome
case
yelp
explained
convenient
mediocre
typical
neighborhood
pepperoni
area
favorite
spot
years
recommend
bottle
comes
prices
point
Importance