

Preventing Bar Closure

Group-15: Max Schleck, Shravan Kaul, Jiayang Wang

November 17, 2022

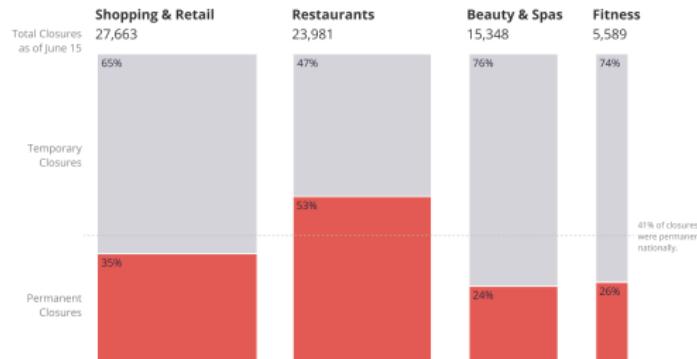
Motivation

Restaurants specifically are Closing at rates that resemble the Great Depression

- Bars and Food trucks are most vulnerable

Restaurants and Retail have been Hit Hardest

Number of businesses marked temporarily or permanently closed*
on Yelp that were open on March 1



*Among U.S. businesses that were open on March 1. A business was counted as closed if it was marked as permanently or temporarily closed, changed its hours to mark itself as closed for entire days, or included the word "closed" in the COVID-19 business updates section of its Yelp page. Closures updated through June 15. Data may lag actual closures because of time taken to report and verify closures.

Source: Yelp
Chart: The DataFace

Reliability of the Closed Flag and Model Choice

Type of Closure	Number of Bars	Proportion
Permanent Closure*	256	.85
No Information/Assumed Closure	40	.13
Reopened	2	.006
Never Closed*	2	.006

Final Closure Numbers

Closed	Open*
298	2

Logistic Vs. Survival Regression?

Summary of Data Cleaning

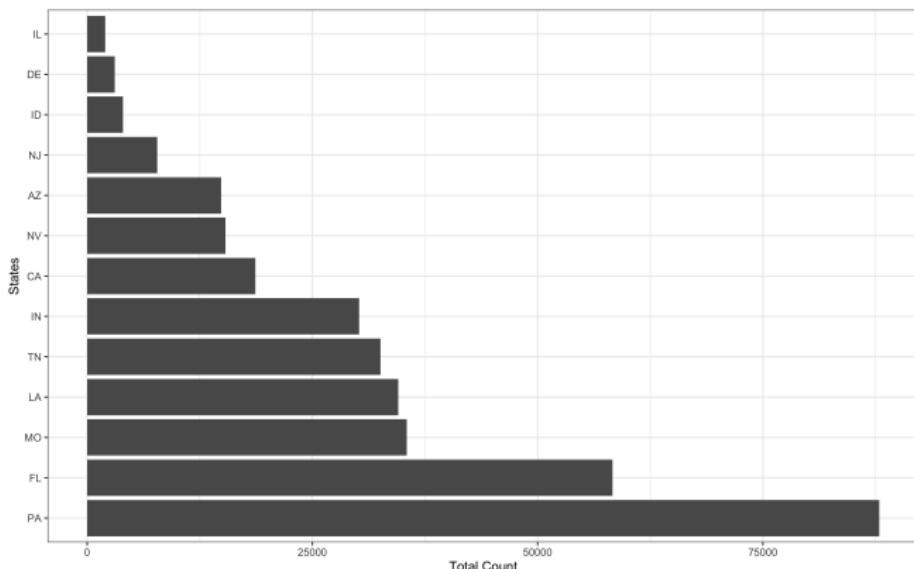
- ▶ We used the "is open" indicator to find all the businesses that were closed.
- ▶ We decided to go with categories bars/nightlife as they had the highest count after category tags food and restaurant with a good amount of reviews.

Category	Count
Restaurants	52268
Food	27781
Shopping	24395
Home Services	14356
Beauty & Spas	14292
Nightlife	12281
Health & Medical	11889
Local Services	11198
Bars	11065
Total	18772

- ▶ Then we merged the dataframes created by reviews, business and tips json files and compiled it into a csv.

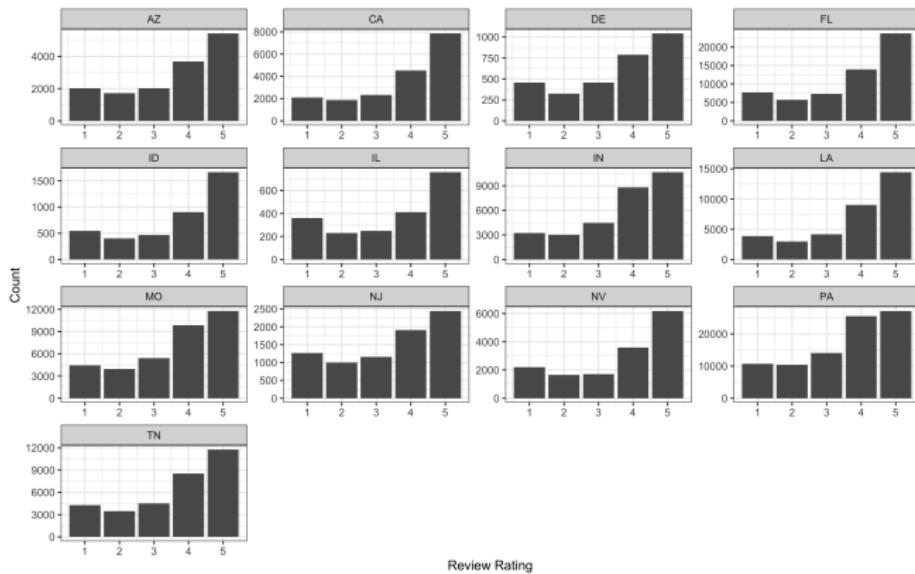
Exploratory Data Analysis

- ▶ The data that it is not standardized as there are difference in review counts within each restaurant and region.



Exploratory Data Analysis

- The distribution of the ratings is surprisingly very similar in each of the regions where the bars closed down.



Words Cloud

- ▶ We divided the dataset into three groups according to the ratings from customers – less than 3 stars, equal to 3 stars and higher than 3 stars For each group, we graphed a wordcloud to visualize the keywords.
- ▶ Our aim is to identify the words that impact the star rating in a review. Also, It helps us to find the predictors we can classify into categories for our survival model.

Word Clouds

Ratings of Customers are Less than Three Stars



It is interesting to note that the bad ratings are being associated with server and bartender in terms of service.

Words Cloud

Ratings of Customers are Equal to Three Stars



Here we can note the impact of food and drinks on neutral reviews.

Words Cloud

Ratings of Customers are Higher than Three Stars



We see that happy hour has a big impact on high ratings.

Natural Language Processing

Three basic aspects of Bars: Food, Drinks and Service

In each category, we considered some features related to it and made target clusters for text classifications using gen-sim(NLP).

- ▶ Food: 'food','dish','snacks','appetizers','starters'
- ▶ Drinks: 'drinks','cocktails','alcohol','liquor','shots','whiskey','vodka','jager','rum'
- ▶ Service: 'bartender','server','waiter','waitress','service','mixologist'

Natural Language Processing

Text Cleaning steps performed

- ▶ Tokenization, Stemming, Lemmatization ...

Tokenization

- ▶ Tokenization is the process of breaking down the given text in natural language processing into the smallest unit in a sentence called a token.

Stemming

- ▶ Stemming is the process of finding the root of words.

Lemmatization

- ▶ Lemmatization is the process of finding the form of the related word in the dictionary.

Future Work in Pipeline

- ▶ Working on text classification of reviews into categories like food,drinks, service and happy hours using NLP models like BERT to create word embeddings and then further classifying to target clusters based on categories via cosine similarity.
- ▶ Standardization of reviews.
- ▶ Feature selection,engineering and parameter tuning for survival model.