Bar Closure Executive Summary Report

Max Schleck, Shravan Kaul, Jiayang Wang

Introduction:

It was no shock that, after a 2 year long pandemic which emphasized isolation, Yelp's quarterly economic status report labeled the restaurants as one of the top two failing service industries. Within the restaurant category, bars continually rank as the restaurant type with the highest rate of closure consistently in this post Covid landscape. With this as motivation our group chose to research reasons why bars are continuing to fail at a higher rate. To assist in this goal we chose to limit our search type further to a specific state, Pennsylvania, for both availability of closed restaurant data and group familiarity with the region. Using this review data, our group focused on creating a tool to provide bar owners insight into how at-risk their bar is, along with areas for improvement.

Data Cleaning:

After the initial preprocessing of the open and closed restaurants from loading the business information along with the review data we converged towards choosing the state (PA) with the most number of businesses with an equal amount of reviews for each of the two categories.

During the data cleaning, we found many reviews which were not related to the bar at all and would cause our model to be skewed, so we curated the stopword list to cater towards this issue. We also have considered the reviews that were written in English for better contextual sentiment analysis and text classification model accuracies. Within the text cleaning, we also added certain profanity filters words to the stopword list for added interpretability at the modeling stages.

We limited the scope for bars, by removing the closed bars that had reviews less than 25 as minimal reviews would skew the results and also would maximize the chances of understanding the reason behind closure of bars. Similarly, for open bars we decided to pick a higher upper bound as 700 as they naturally had more reviews and could be used to support the survival model.

Our proportional hazards model requires a time variable to represent the "age" of these bars at the time of closure. Unfortunately perfectly accurate data on open and close dates does not exist. Instead we chose to substitute this with the difference, in weeks, between the most recent review and the oldest available review for a rough estimate of age of each bar.

Finally, we additionally chose to categorize our scores (explained below) into the typical 5 star rating. This categorization naturally lends itself to a survival model and significantly increases our predictive power from the model.

Language Processing and Sentiment Analysis:

Our idea was to capture the entirety of the aspects conveyed in the reviews by using a combination of word embeddings, text classification modeling and sentiment analysis.

Bar Closure Executive Summary Report

After performing EDA, we found six major aspects that govern the survival of the bar derived from the reviews which would further be used as predictors for the proportional hazard model. The aspects chosen were 'food', 'drink', 'price', 'service', 'ambience' along with the sentiment score for the reviews.

The first step before we get into text classification would be to perform text cleaning and embed the aspect labels we have chosen. Since the labels are textual they can be projected into an embedded vector space. Therefore each review can be classified based on the distance from its word vector centroid with respect to each label. This is a form of zero-shot learning for NLP [1].

We experimented with a few embedding techniques like ELMO, glove and BERT embeddings but it had issues handling reviews with tokens more than 512 which is prevalent in most transformer based embeddings. We then decided to go with SpaCy word2vec embeddings as it was both fast and accurate to create the vector embeddings for the reviews as well as the aspects. For finding the distances from the label vectors to the centroid we used NearestNeighbors from scikit-learn package.

Finally, for the sentiment analysis we decided to test out a few of the techniques available like NLTK's Vader, TextBlob Sentiment and Flair architecture based DistilBERT, and decided to use the distilled version of Bi-directional Encoder Representations from Transformers model (DistilBERT),. We chose it since it aptly recognized the contextual sentiment from the reviews and had the highest accuracy and precision out of the 3.

Combining the centroid distances and averaging the sentiment scores for each of the aspects we made the predictors for the survival model.

Model:

We chose to use a Cox Proportional Hazards (CPH) model to estimate failure percentage of bars given their scores described above. However, immediately there were some concerns with time dependence of our predictors. Initially we planned to utilize all 5 sentiment scores we created. However, in this model we found a strong correlation between service and time. This naturally makes sense as service is the predictor that would change the most with time as people who work at these places will ebb and flow. With this in mind we chose to take advantage of stratification. Stratifying the service variable will dice up our data set into strata depending on their service score. This allows us to still take into consideration the service information without violating the assumption of proportional hazards, which will be explored later . However, we will not be making any claims about the impact service has on your hazard. Thus our final model, in laymans, is:

$$Surv(time, is\_open) \sim Drink + Ambience + Price + strata(Service)$$

Where surv(time,is_open) is our survival function and each of the predictors is their respective categories sentiment score. The significance of a CPH model is typically tested by

Bar Closure Executive Summary Report

looking at the significance of three tests, the Likelihood ratio test, Walk test and logrank test which are all equivalent with large sample sizes. Our sample is on the smaller end so we specifically looked to optimize the Likelihood ratio test significance as this test performs the best on small data sets. The significance of each test is shown below in the p value, where we consider any p<.05 to be significant. Clearly our model is significant over all tests, with an even better performance on our key test. Specific conclusions from the model will be discussed in the next section.

```
Likelihood ratio test= 40.85  on 3 df,   p=7e-09
Wald test            = 41.42  on 3 df,   p=5e-09
Score (logrank) test = 41.61  on 3 df,   p=5e-09
```

Assumption Checking:

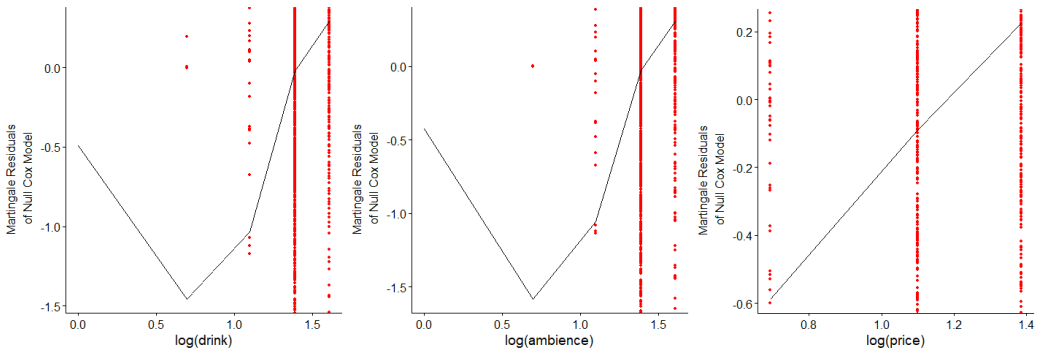
Proportional Hazards:

The CPH model requires the relative hazard (or risk of closure) to be constant over time and as our other predictors vary. This assumption in the real world is extremely hard to justify as essentially everything varies with time. This assumption is checked by referencing the scaled Schoenfeld residuals correlation with time as these two conditions are equivalent. Generating these correlation coefficients (p values in the table to the right) shows that none of our predictors are correlated at the 10% significance level, double the industry standard of 5%. And at the global level, taking all predictors into account, we still see no correlation with time. With the addition of a stratified service score this assumption was met.

	chisq	df	p
drink	0.01649	1	0.90
ambience	0.31946	1	0.57
price	0.00559	1	0.94
GLOBAL	0.41103	3	0.94

Linearity of log(Hazard) and Predictors:

The CPH model additionally requires there to be a roughly linear relationship between the log of our hazard and our predictors. The best way to check this model assumption is by looking at the marginal residuals plotted against the log of each of our predictors. Graphically we expect to see a roughly linear trend between residuals and our log function. Below are all 4 of our predictors plotted in this fashion:

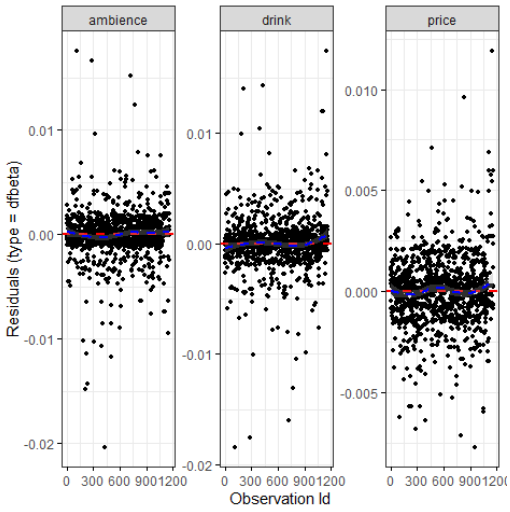


Bar Closure Executive Summary Report

From this we see a very clear linear relationship in our price graph. The other three also indicate a linear relationship, just a bit less clear. With a lack of 1 star reviews for drinks, food, and ambience it leads to the upward tailing at the beginning of the graph, thus this is not an indication of non-linearity rather just displaying that the predictors are skewed right. We don't see this in the price graph because price had a significant amount of 1 star reviews to work with, giving us this very clear linear trend.

Influential Points:

While not a direct assumption of the model, additionally we looked to remove potential outliers impacting our model. This is checked visually by reviewing a dfbeta plot. This plot represents the changes to the regression coefficients that occur when said point is removed. We will consider any bars with a dfBeta Residual over .02 as an outlier with grounds for removal. Looking at the dfBeta plots to the right we see that all points are within this threshold meaning we do not need to do any outlier removal at this point with our data.



Results:

When interpreting CPH results we look at the hazard ratio for each predictor. The hazard ratio in our case represents the increase in the bar's likelihood to survive if a 1 unit change was made to said predictor while all others remain unchanged and is calculated by taking exp(regression coefficients). This is not the typical interpretation of the rations, rather the opposite because of the inverted nature of our is_open column. We can find the traditional hazard ratios changing our hazard ratio calculation exp(- regression coefficients). From this we find our most influential predictor to be ambience with a hazard ratio of 1.24 (or traditional hazard ratio of .80) this ratio means that with an increase in one rating of a bar's drink score, it is 24% less likely to close. Drinks and Price have similar ratios of 1.19 and 1.13 respectively. Thus creating a model with strong predictive power, along with significant recommendations on specific fields to work on for efficient review improvement.

Conclusions:

After a holistic review we found that price, ambience and price are the three driving factors around bar closure in Pennsylvania when stratifying based on service. While changes in food do not matter. But more interestingly we found that atmosphere is the most important factor to predicting bar closure, meaning any bar owner looking to turn their establishment around should first focus on creating a positive customer experience the moment they enter the bar, rather than focusing on improving the menu.

Bar Closure Executive Summary Report

Contributions:

All ideas and models were discussed before coming to a conclusion.

Max Schleck - Wrote code for creating time columns and categorizing our score values. Wrote model code and assumption checking code, wrote Intro, Model, Results and Conclusions section. Created the Shiny App.

JW wrote the diagnostic and data cleaning part of the summary, worked on slides 1 and 7 to 9. JW also created code related to model diagnostics, Figure 2 and Figure 4. JW is ultimately responsible for the model diagnostic portion of the code.

Shravan wrote the code for natural language processing, word embedding and sentiment analysis. Additionally worked on the initial data cleaning, text cleaning and wrote the data cleaning and Language Processing section of the report.