

Sayyid Ab Latiff, Lindsey Prommer, Temirlan Amir, Madelin Warsito, Matthew A. Lanham

Purdue University Krannert School of Management

sablatif@purdue.edu; lprommer@purdue.edu; tamir@purdue.edu; mwarsito@purdue.edu; lanhamm@purdue.edu

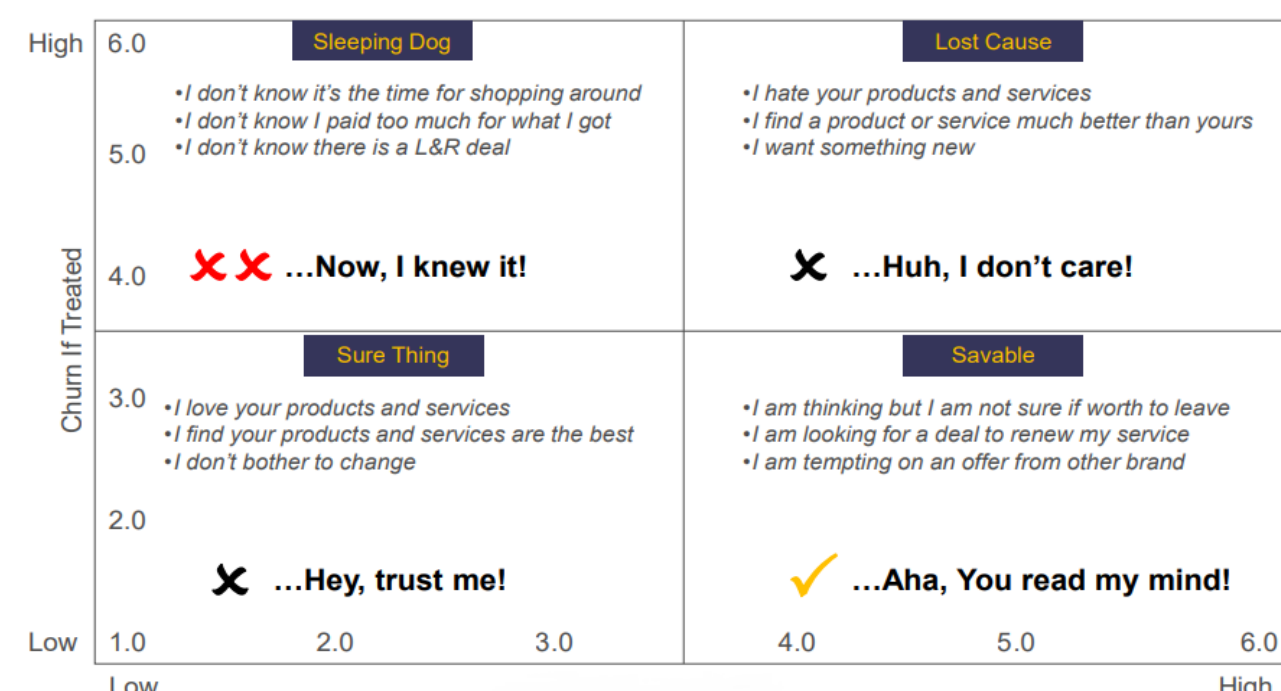


Abstract

This project focuses on understanding how uplift models pertaining to marketing can better predict consumer behaviors, as compared to traditional classification models. Uplift modeling outperforms classification modeling because it allows for testing within a specific product; this cannot be done at a product or customer level using traditional classification models. Using the historical sales data from a liquor retailer, an AB test was conducted in order to determine churn favors that identified people who are receptive to marketing efforts and are likely to become customers. The goal of this study is to build an uplift model in order to show the effect of varying alcohol marketing tactics on customer churning. For our workflow, we first studied the data carefully and performed exploratory data analysis along with feature engineering to clean the data for improved effectiveness of our model and determined input variables that were statistically significant to our model. We then conducted multiple analysis such as classification regressions and clustering. This allows us to identify features about the liquor that will lead to better effects on sales.

Introduction

The big issue in a marketing decision is to allocate marketing spending as effectively as possible in order to gain high return on investment (ROI). Introducing Uplift Modelling which is a two part model which calculate the **magnitude** of the treatment(advertisement) and **identify targets** whether it is individuals or items that will have a higher chance to be converted in a sense to be affected by the treatment (**savable group based on the figure below**).



Research Questions:

- Does promotion (**treatment**) increase the sales of a product (**A/B testing**)?
- Identifying product in the "Savable" groups which are the group that will receive the highest sales (**uplift**) from the promotion (**treatment**)?

Literature Review

Previous studies utilized personal customer data to identify consumer buying patterns while we used a different approach to our data by identifying how the sales of the liquor is affected using logistic regression and decision tree methods.

Study	Decision Trees	Class Transform	Random Forest	Cost-Based	Logistic Regression
(Cao & Xu & Gu, 2017)			✓		
(Soltys & Jaroszewicz & Rzepakowski, 2014)	✓				
(Ahmed & Maheswari, 2019)				✓	
(Karlsson, 2019)		✓	✓		
(Ja'skowski & Jaroszewicz, 2012)					✓
Our Study			✓		✓

Table 1. Literature review summary by methods used

Methodology

Figure 2 outlines the steps that we took to get from our original datasets to our final model.

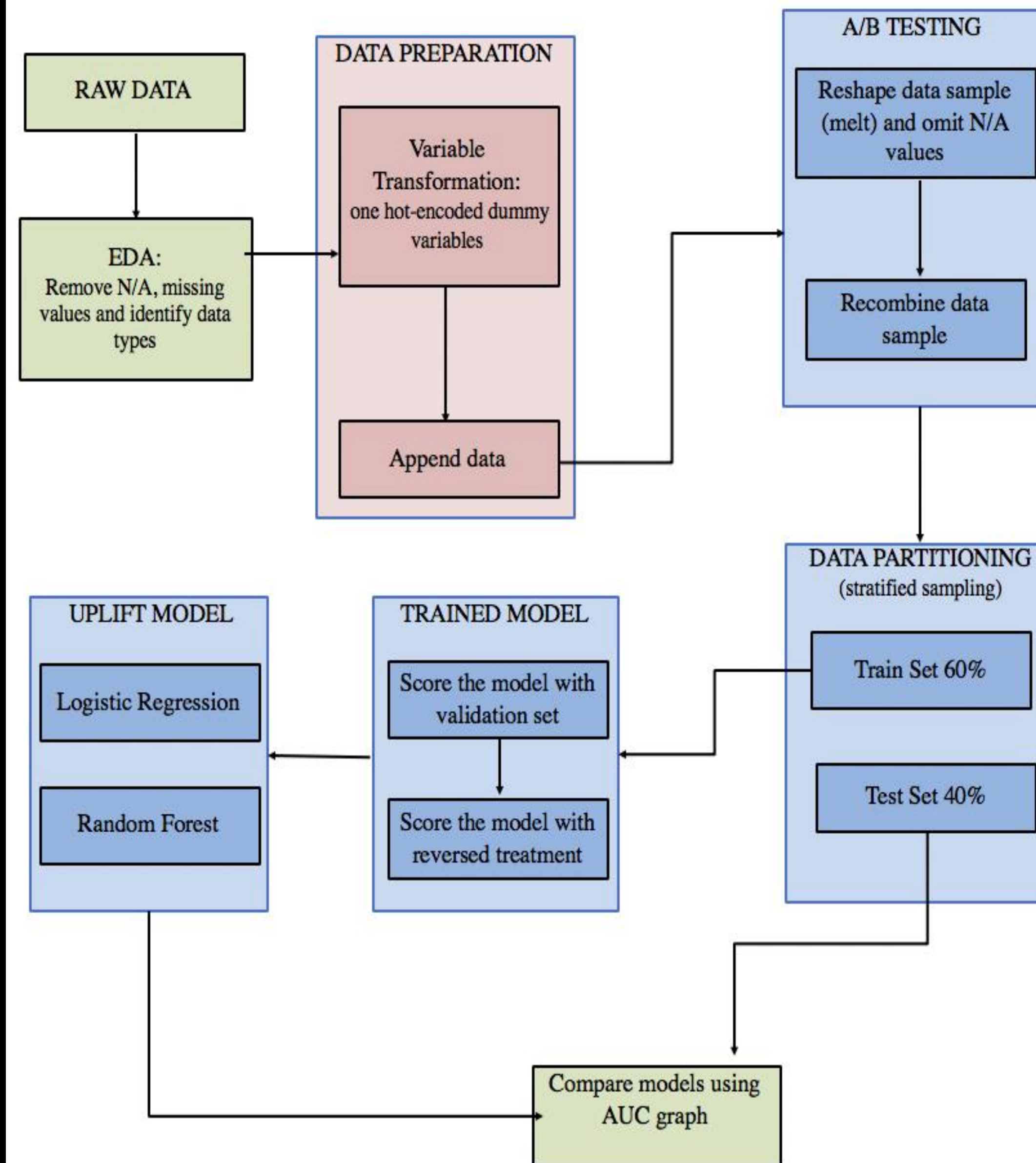


Figure 2. Study Design

Model Evaluation / Statistical & Business Performance Measures

The predictive models were evaluated on overall accuracy and AUC statistical performance measures because they tell us how well the model was at distinguishing between classes. The business performance measure we considered is customer dollars spent (revenue) because this helps us to best compare with the effect of business dollars invested into advertising. Based on the result on the right, we chose Random Forest as our final model as it has a higher AUC than the logistic model. Although lack in easier interpretation, our business problem is focusing on targeting/identify the savable group and thus, a high AUC is needed.

Results

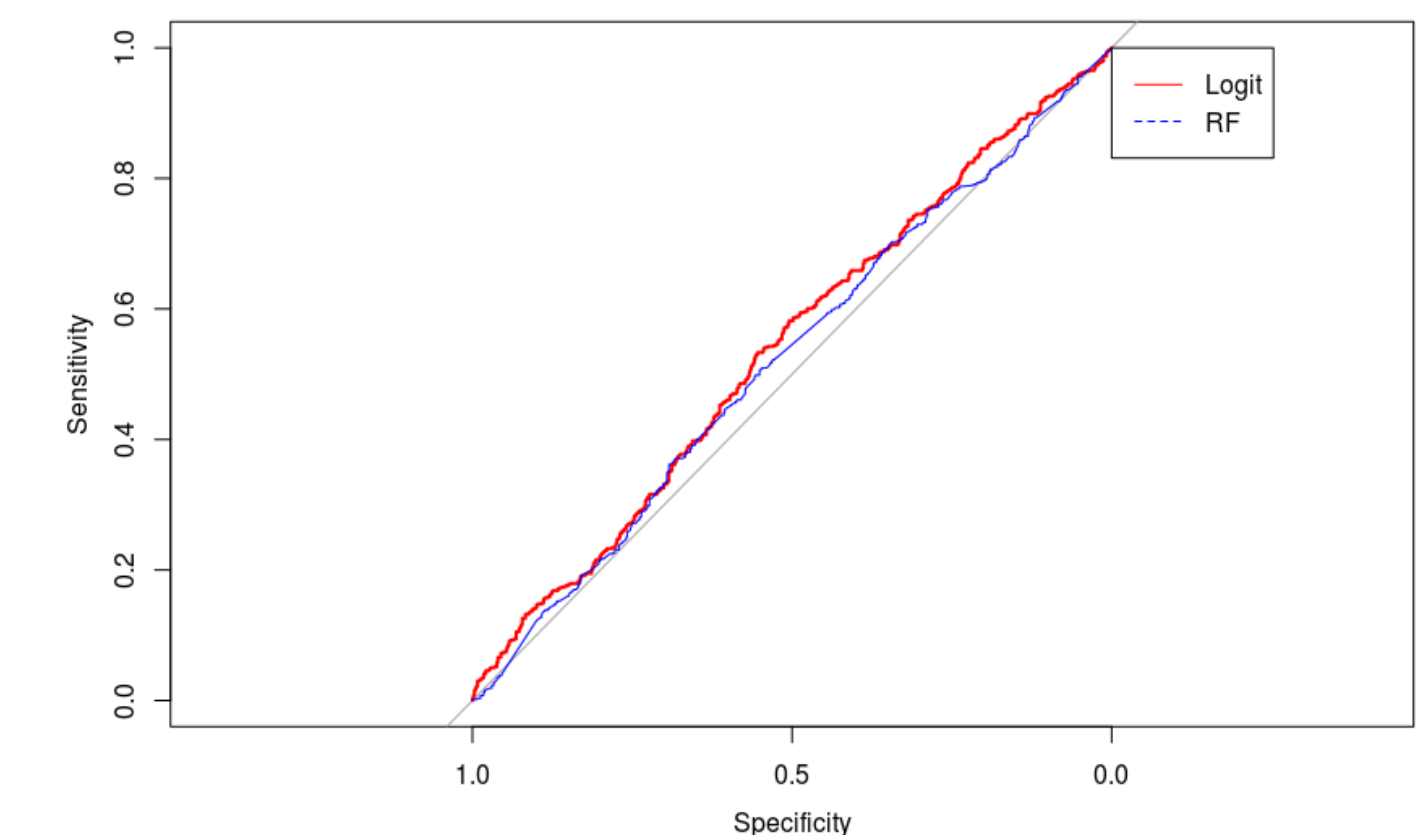


Figure 4. Model Evaluation (AUC)

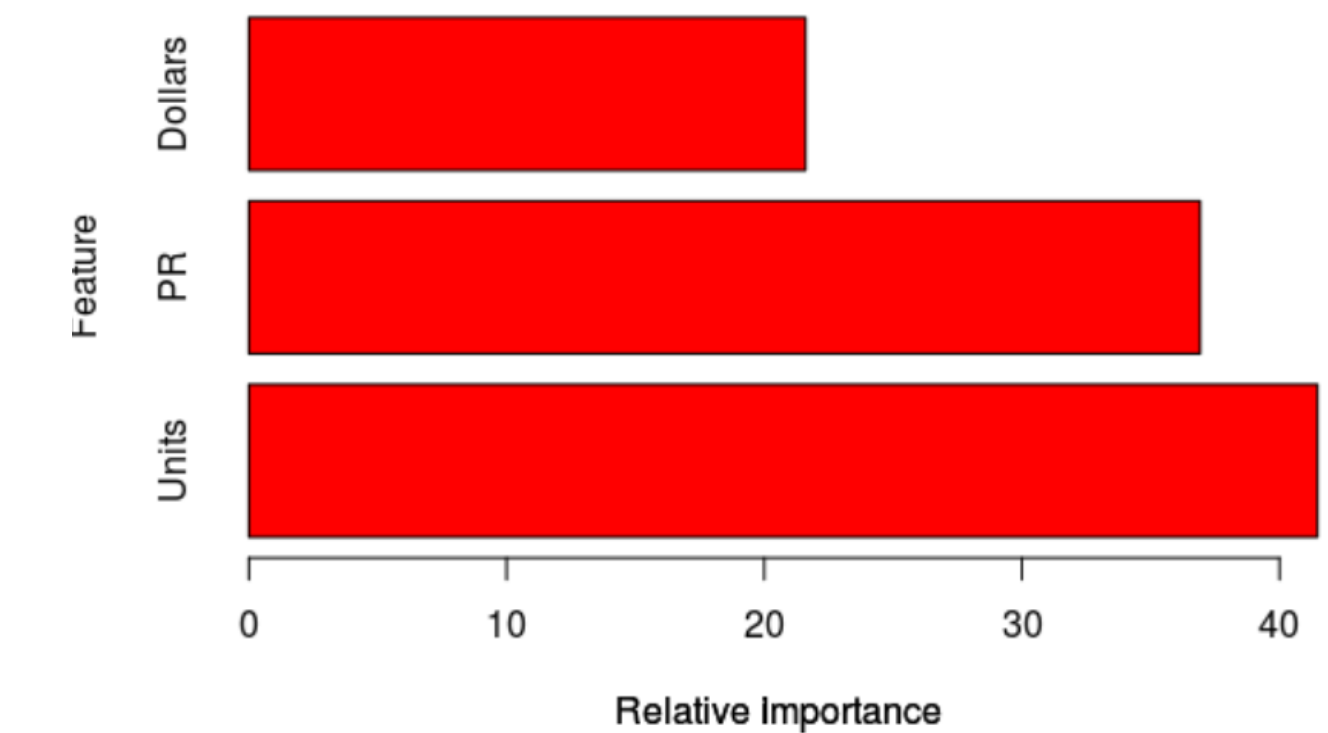


Figure 5. Features with Highest Impact

Received Treatment	Average Spending Before Treatment	Average Spending After Treatment	Change in Dollars	Change in Percent
Yes	\$34.34	\$17.42	\$16.91	97.07%

Figure 6. Average Uplift with Treatment

Conclusions

With our result of 97.07% magnitude of advertisement as well as an uplift value of 21% for liquor X (**Brand is censored**). If the store were to used our uplift model, there are two business decision that can be made which **focuses all marketing effort on liquor X** and suggesting the maximum cost of ad for that liquor.

Assuming liquor X price per unit is \$15. We are able to suggest to the company their maximum cost of advertisement for that liquor to at least break even.

Breakeven (0) = (\$15 * 1.97) * 21% - Cost of Advertisement
Suggested maximum cost of advertisement per liquor X = **\$6.21**

Acknowledgements

We thank Professor Matthew Lanham for constant guidance on this project.