Booze ‘R’ Us Sales Analysis

James Rounthwaite, Jamie Luna, Lana Huynh,
Rachel Roggenkemper, Sophia Chung

Overview

- Our goal was to create a model that would project sales in the coming year given a list of your stores' data, ultimately so you can decide whether you will have enough income to expand your operations
  - We prioritized a model that would make accurate predictions
- Since the data provided does not enclose the liquor stores' exact sales, in place we are utilizing the total cost of a liquor order, which is the number of bottles multiplied by the state bottle retail

Data Cleaning

- We collapsed the categories of alcohol type into broader categories (amaretto, brandy, gin, liqueur, rum, schnapps, tequila, vodka, and whiskey) to have better predictions
  - Sacrificing a little precision for higher accuracy
- Dropped columns that were not relevant to this analysis
- Split the date column to be able to use month, day, year as predictors
- Standardized columns
  - Consistent model interpretation

Methodology

- We chose to use the data from one store at a time to test our practices
  - This method be applied to any store that Booze R Us owns
- Aggregated bottle counts by category

Model Selection Process

- When creating our models, we cross referenced our results with pre-built packages like scikit-learn to verify our results
- The different models we tested included fitting our model through ordinary least squares regression, ridge regression, and with Log-Cosh loss using gradient descent, testing with both cross validation and time series cross validation

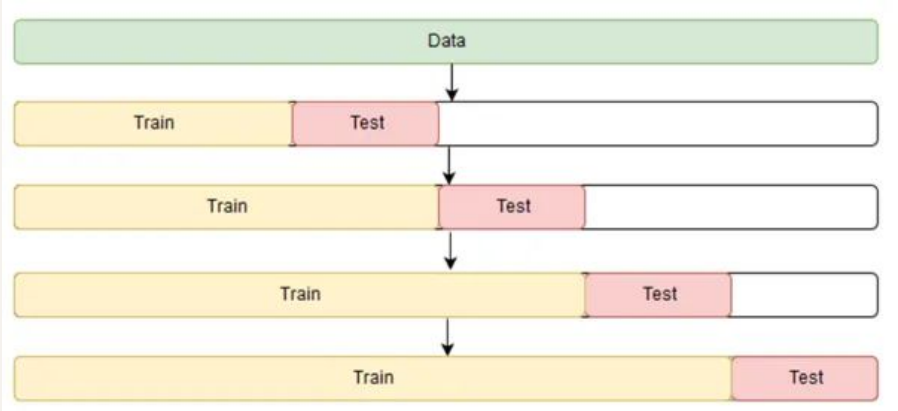
Model Performance

Sales = Year + Month + Amaretto + Brandy + Gin + Liqueur + Rum + Schnapps + Tequila + Vodka + Whiskey + Other

- Ordinary Least Squares Regression
- R-squared =
  - Time series: 0.556
  - Regular series: 0.775
- Year, month, and the broad categories of alcohol predicts 55.6%/77.5% of the variability in sales.

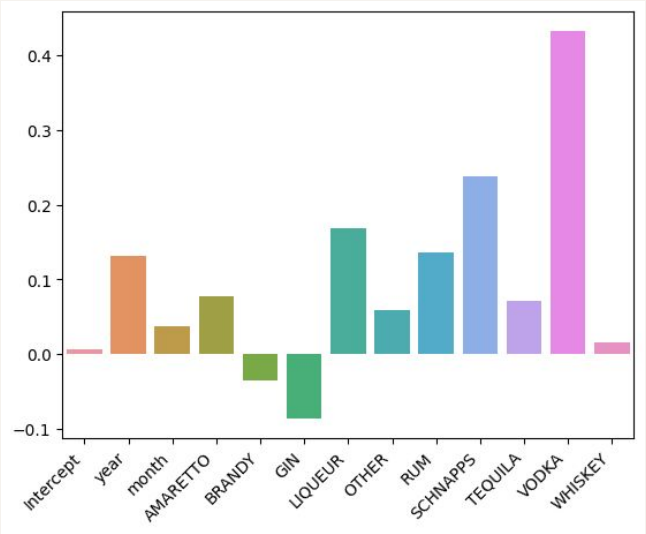
Cross-validation with Time Series Data

- Our data was in time-ordered data, so we wanted to validate our data on a rolling basis
- Would only make sense to use previous years to predict future years



Overall Takeaways and Conclusions

- Our model works very well for this particular store
- We can use this framework and apply it to other stores that Booze ‘R’ Us may own



Thank you!