Analysing Liquor Sales in Iowa for Inventory Optimization

Team 4: Brendan Wilcox, M Zaid , Nimisha Agarwal, Quan Nguyen

Problem Statement

Problem Statement:
New liquor sellers in Iowa struggle to forecast demand and manage inventory

Goals:
Analyzes historical sales data to deliver insights for demand prediction and inventory optimization

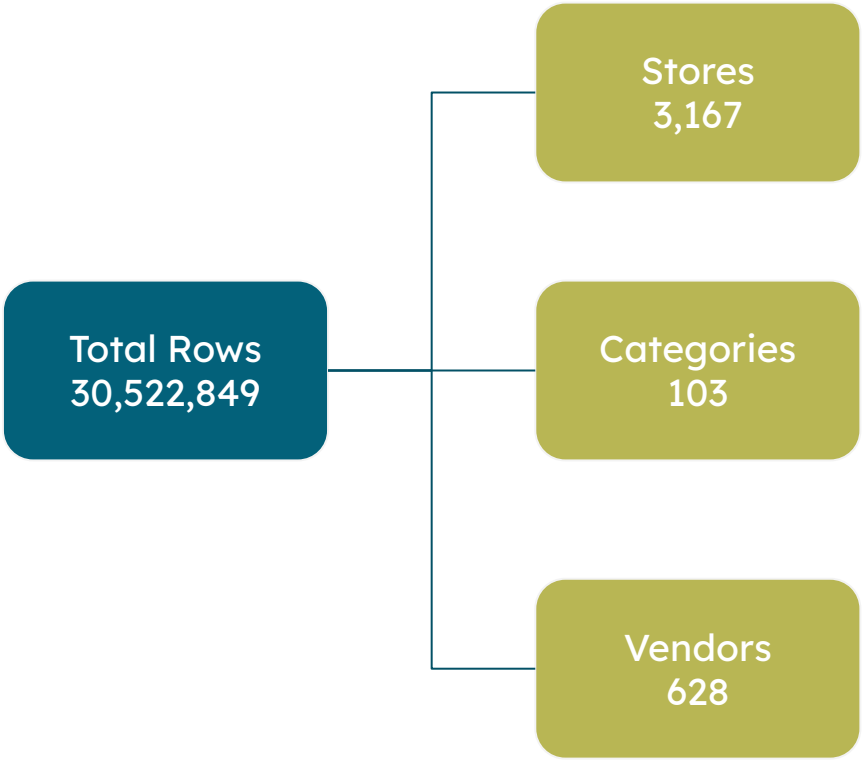
Who Cares and Why?

Sellers: gain market entry insights

Retailers: ensure efficient stock management

Consumers: enjoy consistent supply and good pricing

liquor store



02

Data

Data Source

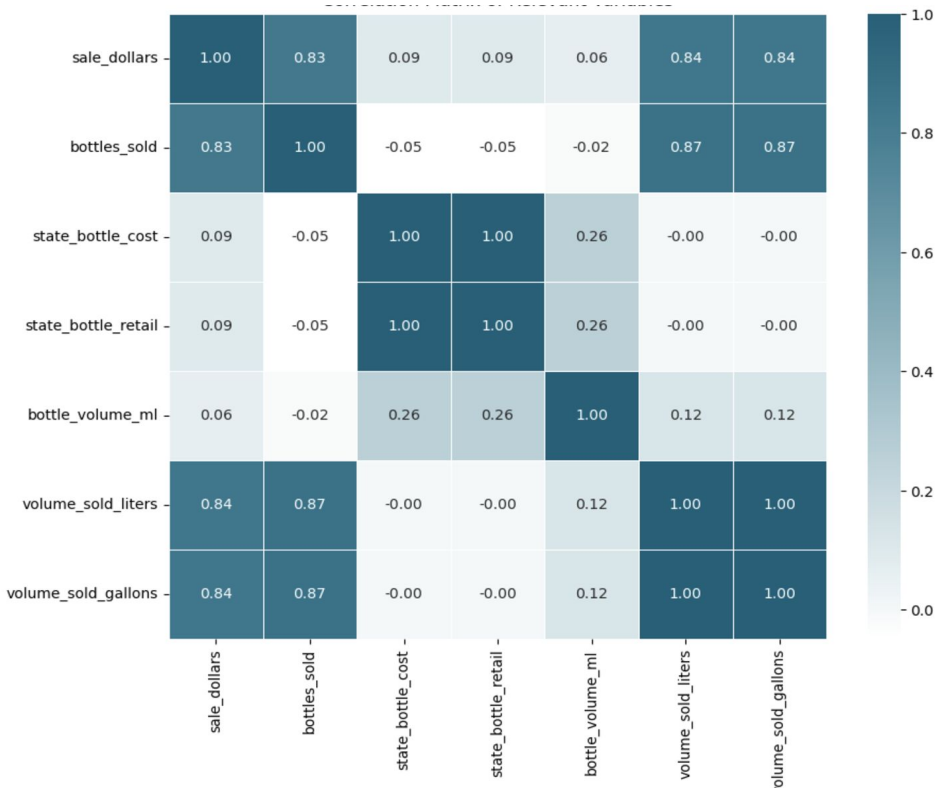
- Iowa Liquor Sales Data
- Provided by Iowa's Alcoholic Beverages Division
- Published on data.iowa.gov
- Also available on BigQuery under bigquery-public-data

About the Dataset

The dataset offers detailed records of liquor sales by Iowa Class “E” license holders.

Timeline	Focused on post covid period, taking the data points from July 2021 to October 2024
Store Information	Fields such as store ID, name, address, city, and county, which will allow us to analyze geographic differences in demand
Product Information	Details such as product ID, description, category, vendor, and bottle volume, enable the categorization for a more targeted analysis
Sales Data	Key metrics like order date, volume / bottles sold, etc. are essential for forecasting sales trends
Data Types	Mixture of text (store and product descriptions), Numeric (sales figures, bottle costs), and Datetime (order dates) variables
Granularity	The dataset provides sales at the product-store-day level

Exploring the Data



Strong Positive Correlation

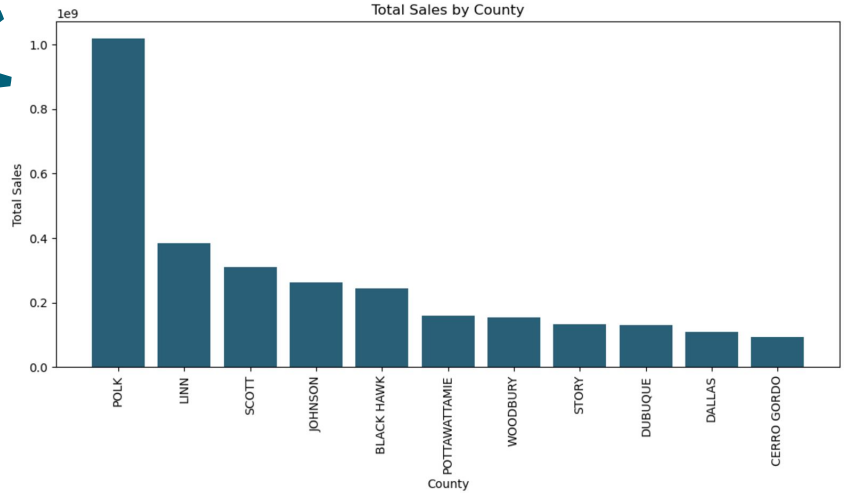
Sales Amount, Bottles Sold & Volume Sold

Low Correlation

Cost and Retail price of Bottles

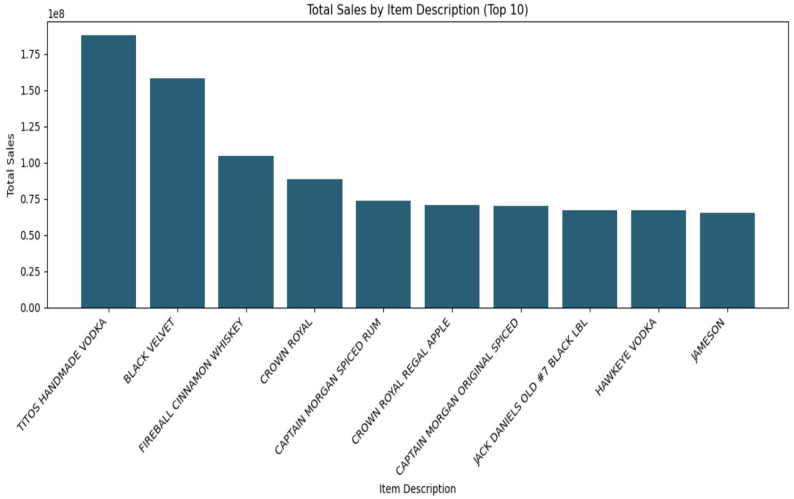
* The correlation matrix helps validate the integrity of the data and performs a sanity check to proceed with feature selection for the data

Exploring the Data



Total Sales by top 10 Counties

- Polk County dominates liquor sales, followed by Linn and Scott, with a steep drop.
- Geography is critical for sales forecasting.



Total Sales by Top 10 Items

- Tito's Handmade Vodka leads total sales, followed by Black Velvet and Fireball.
- Top 10 items reflect customer preferences.

End-to-end ML Process (1)

Cleaning & Preprocessing

- **Rows:** 30M → 16K (GroupBy)
- No imputation
- ~15K encoded columns
- Adj data types (date, objects)
- Standard Preprocessing

Feature Selection

- ***ElasticNetCV***
- Lasso + Ridge combo
- Improved untuned model performance for ⅝ models
- Use reduced feature set ✓

End-to-end ML Process (2)

Hyperparameter Tuning

- **Exhaustive** approach
- Grid, Randomized, Halving, Bayesian👎
- All models tuned: Decision Tree, XGBoost, RF, Ridge, Lasso

Model Selection & Test

- **Lasso,** best RMSE
- **Random Forest,** 3rd best RMSE
- Ridge *not* chosen, 2nd best RMSE (Redundancy)
- Stacking impractical

Main Result

Our **Random Forest** model demonstrated strong performance in predicting Iowa liquor sales, balancing routine accuracy **(low MAE)** and handling outliers **(low RMSE).**

It achieved the lowest Mean Absolute Percentage Error **(MAPE) of 57.17%** and relatively low **RMSE on test data (13,538.83).**

This dual capability ensures reliable inventory management while mitigating risks during high-impact events like holidays or promotions, making it an invaluable tool for supply chain optimization.

Challenges

Computing Power

Limited computing power and the massive dataset slowed processes, required heavy transformations, and reduced predictive efficiency, adding significant delays to our workflow.

Errors

Realizing the need to remove "vendor_name" and "vendor_number" late in the process forced us to restart, significantly extending project time.

Delay in Establishing Objective

We hastily chose "sales" as our prediction target without fully understanding the dataset limitations, leading to late-stage errors and setbacks.

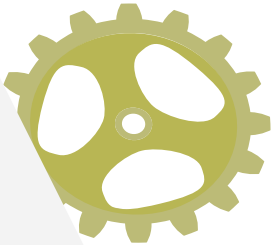
Not Fully Reliable

Despite exploring all feasible approaches, our models remain suboptimal due to the limited predictive power of our data, which was disappointing given our effort.

Key Takeaways

Future Analysis

Thank You



Notebook:
https://colab.research.google.com/drive/1WbTVjyd5M5yKvj-KHM_2goayqLI4CWP4?usp=sharing