Total Wine & More

VODKA UNPLUGGED

INSIGHTS FROM DATA AND MACHINE LEARNING MODELS

Our Super Team

VEDANT KAMAT

SAI MANOHAR BEERAKA

PRAVAH MALUNJKAR

WENDY GUAN

HITARTH SHAH

Table of Contents

Executive Summary

01

Conducted a detailed analysis of the U.S. Vodka market, including a store visit in Laurel to understand operational challenges

02

Cleaned and integrated datasets with external sources, applying feature engineering to uncover key sales drivers

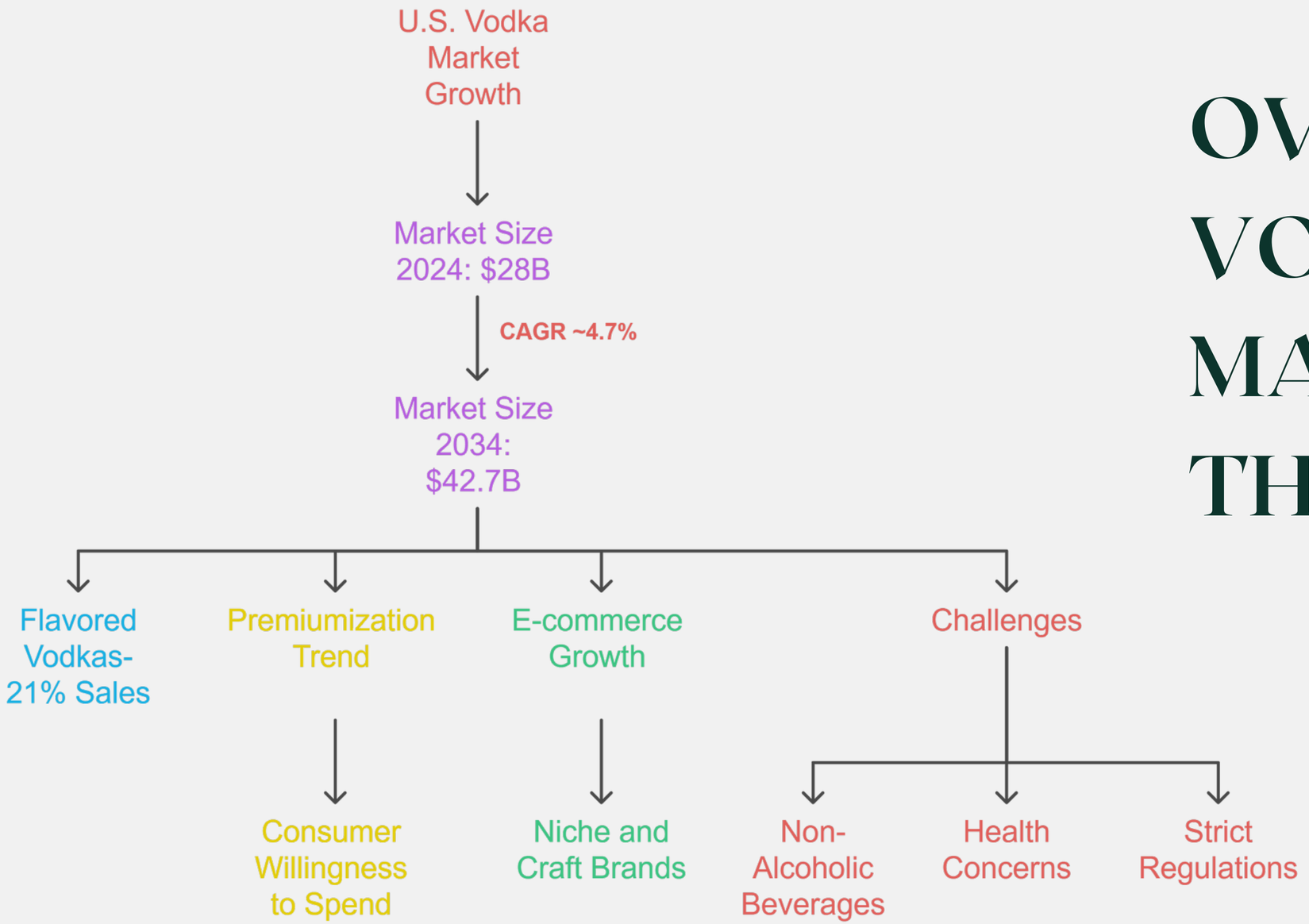
03

Evaluated predictive models, including XGBoost, ANN, and Linear Regression, using metrics like RMSE, MAE, and R-squared.

04

Identified Voting Regressor as the most accurate model, delivering superior performance for vodka sales forecasting with RMSE of 11918

OVERVIEW OF VODKA MARKET IN THE US

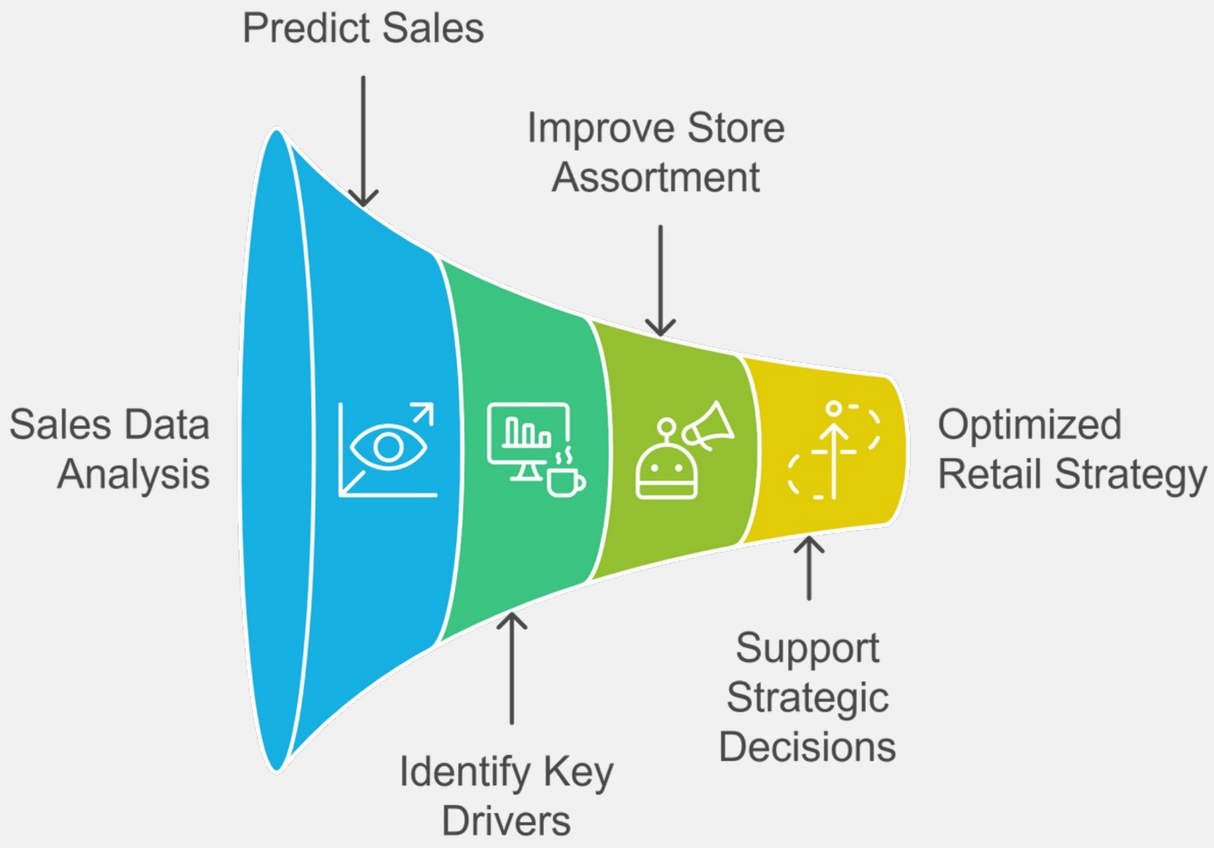


BUSINESS OBJECTIVES

BUSINESS PROBLEMS

- Sales Performance
- Assortment Planning
- Overstocking
- Risk of external factors

PROJECT OBJECTIVES

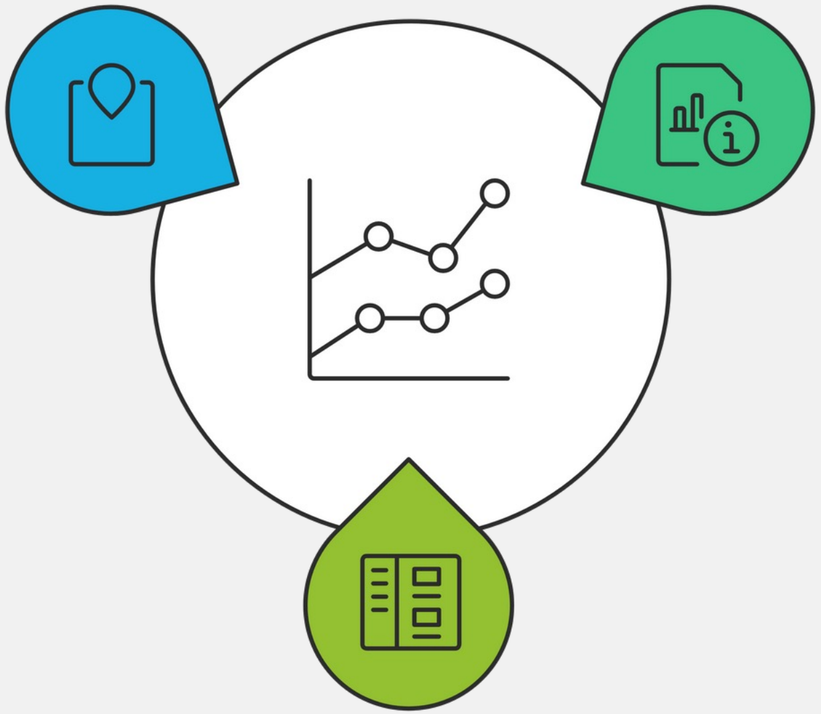


DATA HIGHLIGHTS

INITIAL DATASET OVERVIEW

External Sales Data

Item Info (Code, Name, Package Type, Sales Amount)
Locations with sales of the item



Store Information Data

Vodka Sales Volume
Demographics within 5-mile radius

Internal Sales Data

Item info (Code, Name, Package Type, Retail $ amount)
Normalized Sales $ Amount
Store Info (Number, State)

DATA CLEANING AND MERGING

Retain N/A values

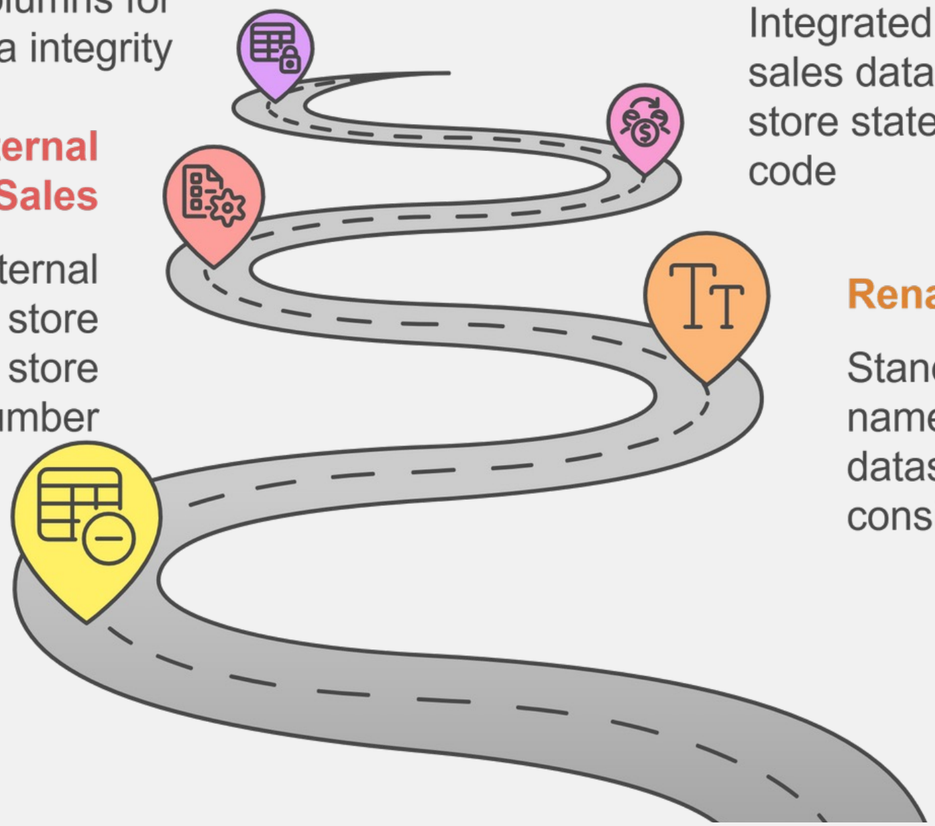
Kept N/A values in specific columns for data integrity

Left Join Internal Sales

Merged internal sales data with store info using store number

Drop Columns

Removed non-essential columns to streamline data



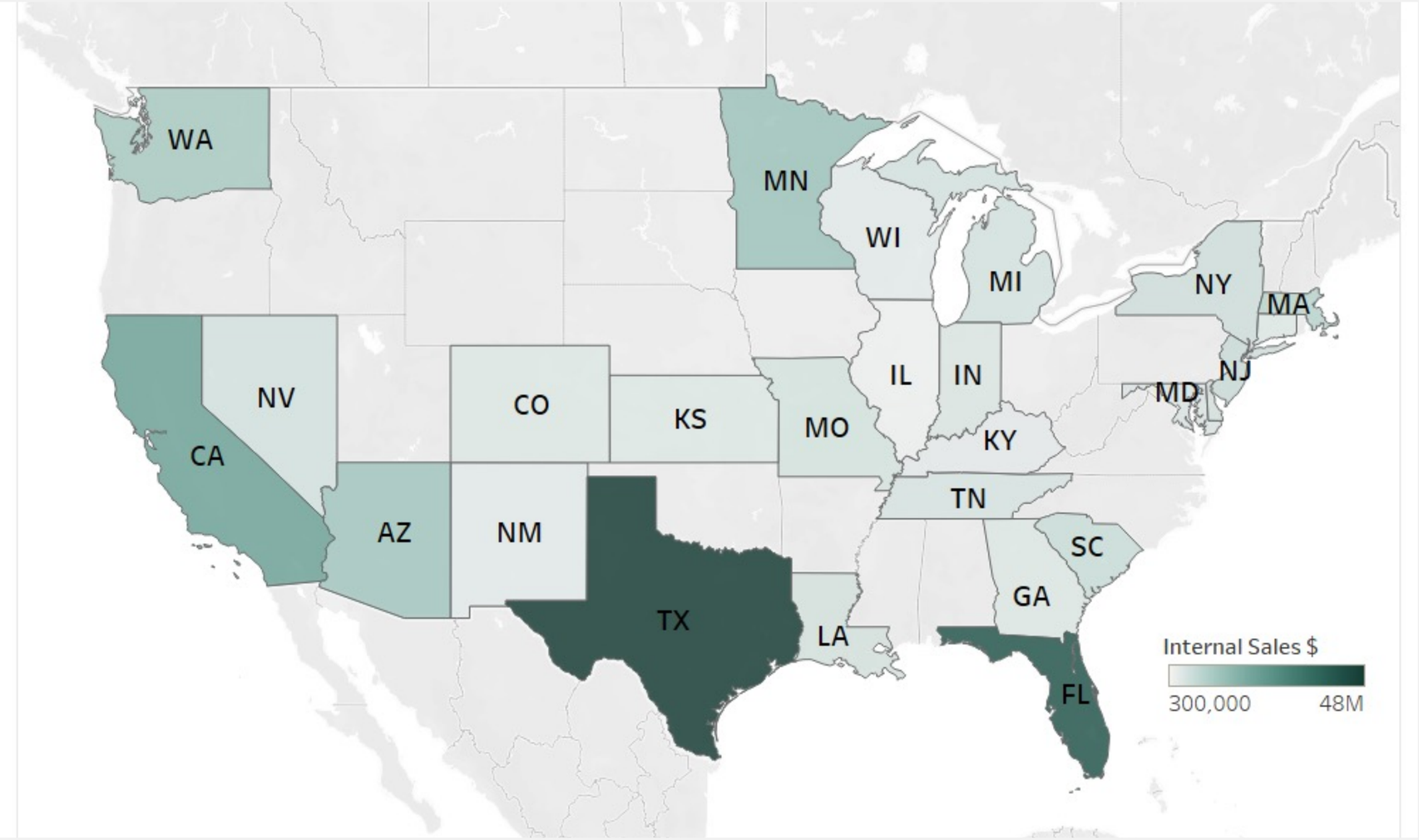
Left Join External Sales

Integrated external sales data using store state and item code

Rename Columns

Standardized column names across all datasets for consistency

TOTAL SALES ACROSS DIFFERENT STATES

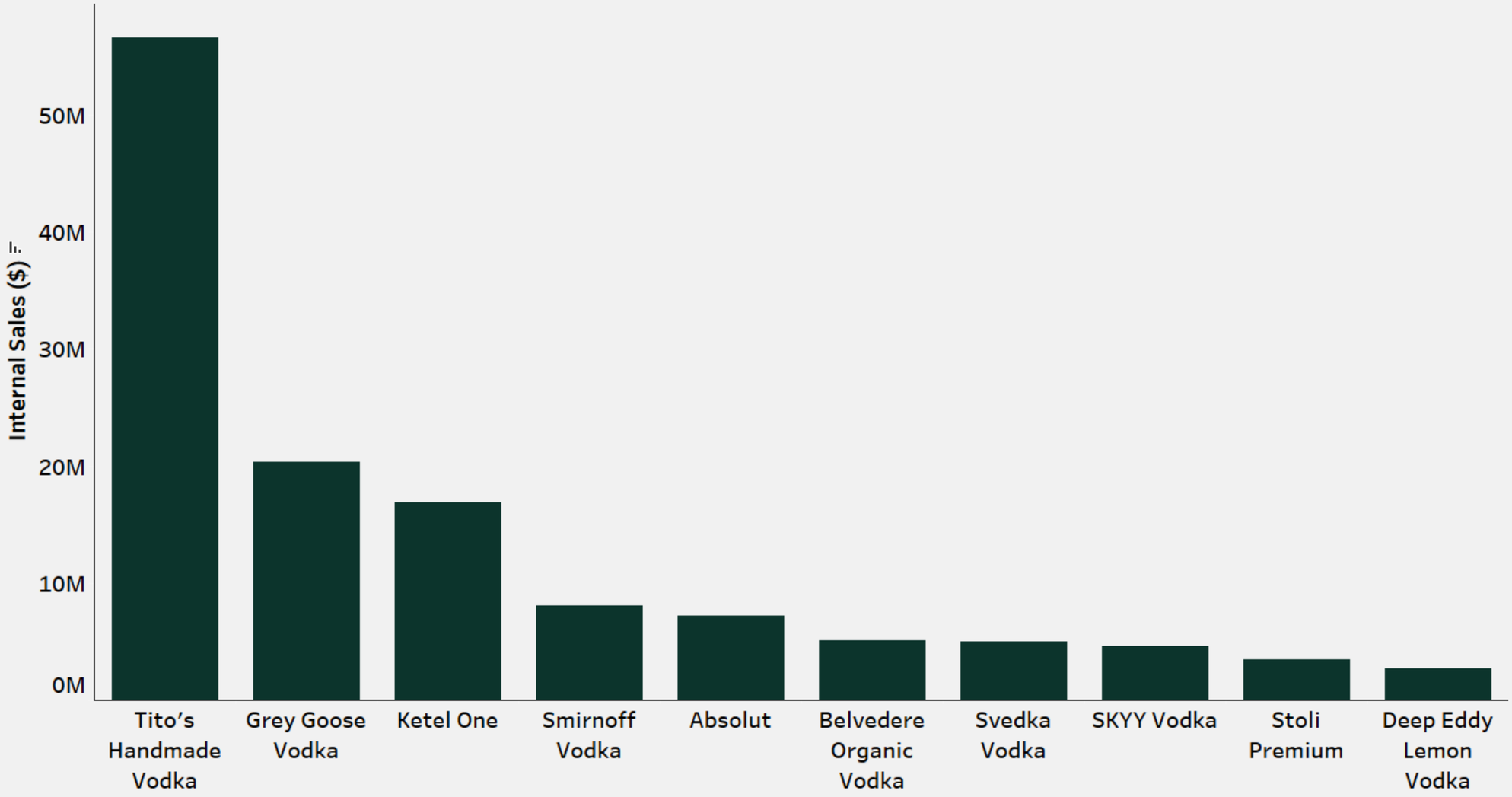


Total Sales : $219 M

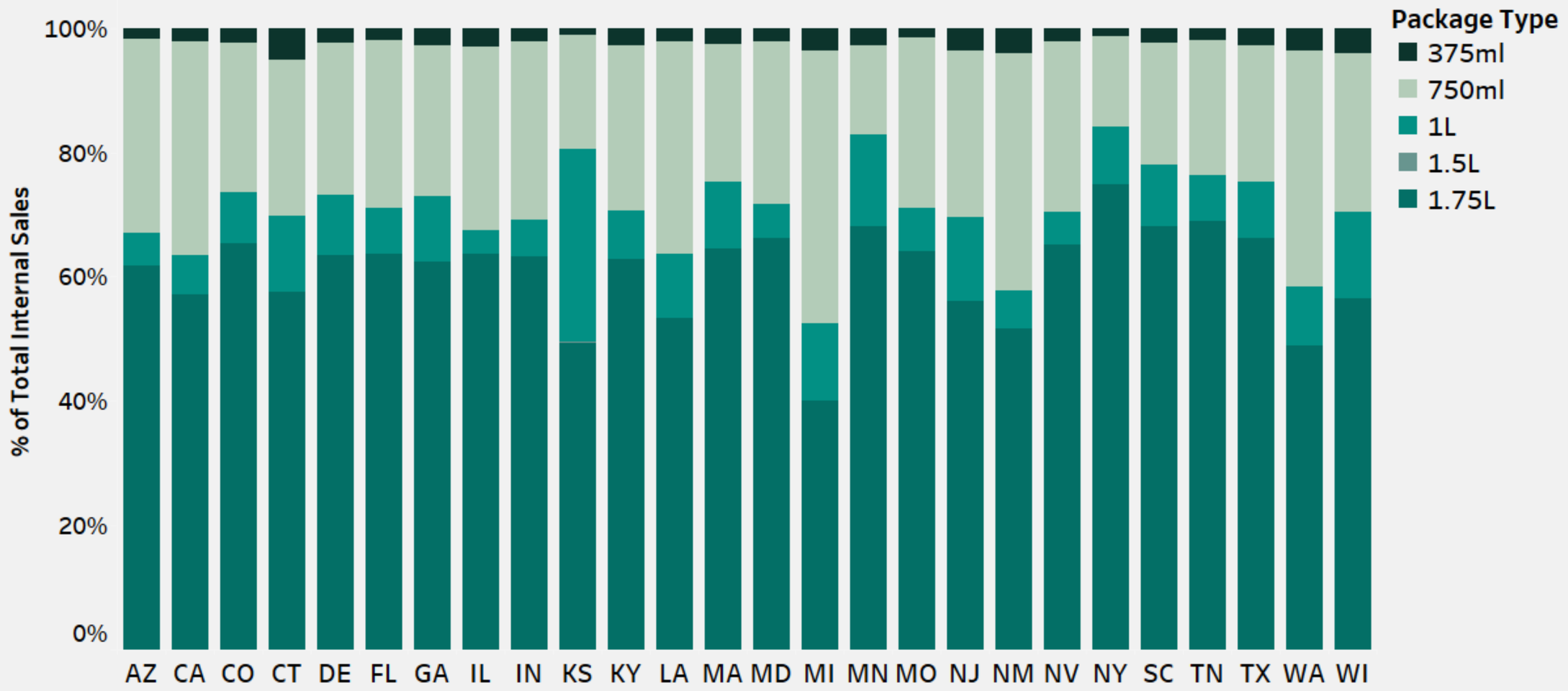
Total Stores: 232

Avg Price: $24.8

TOP SELLING VODKA BRANDS



TOTAL SALES ACROSS DIFFERENT STATES



FEATURE ENGINEERING

Normalized Data

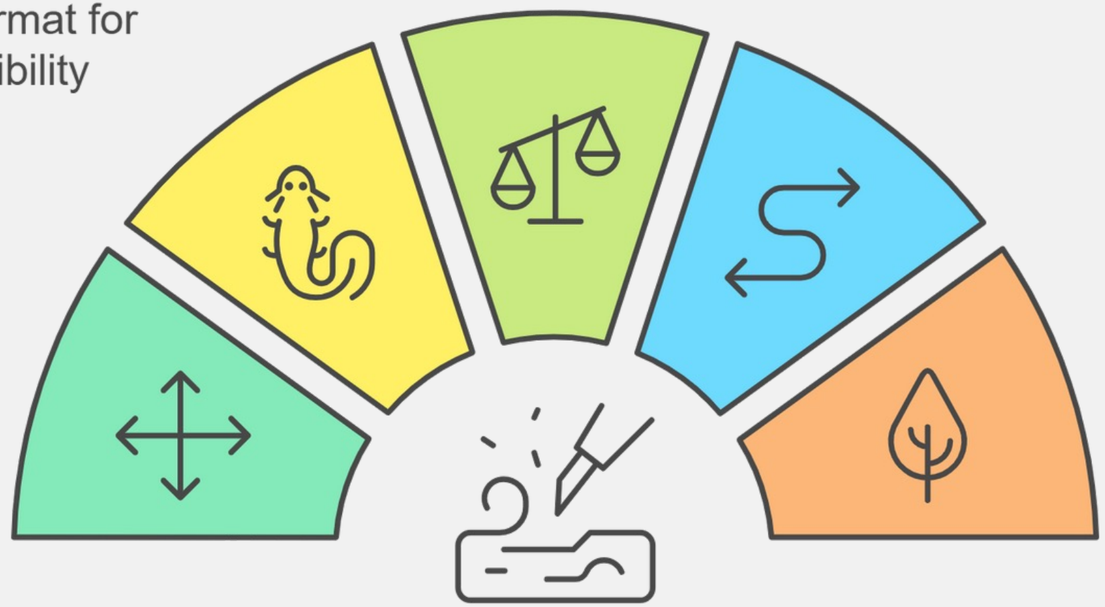
Scaling features such as Count of weeks in stock to a 0-1 range

Dummy Variables

Converting categorical variables like Store Size into numerical format for model compatibility

Log Transformations

Reducing skewness in features like Retail Price



Interaction Terms

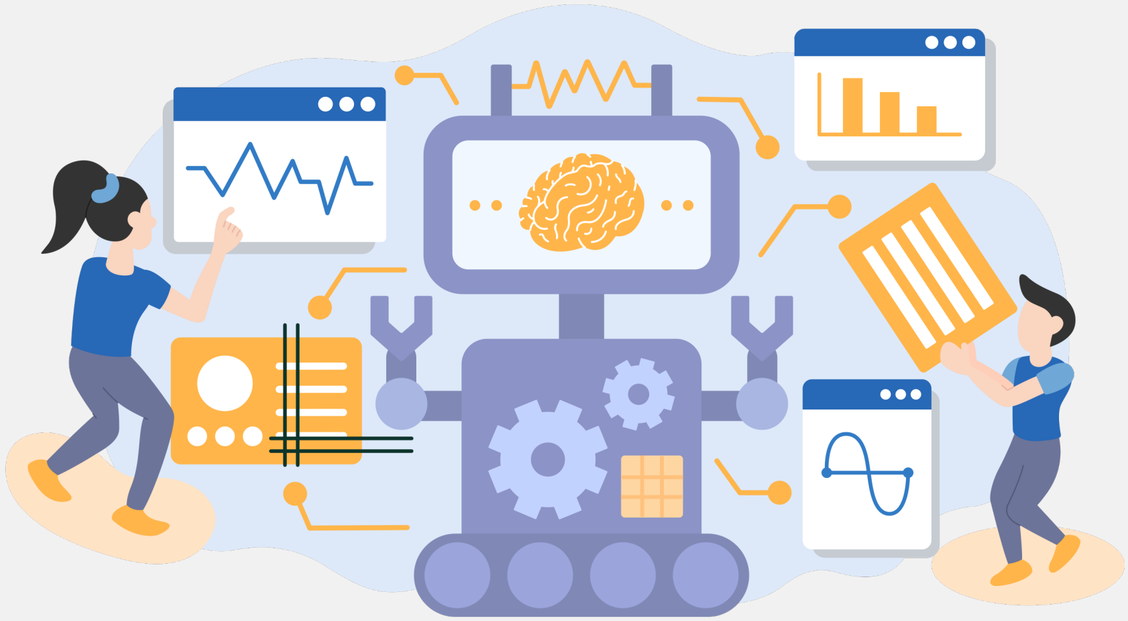
Examining feature relationships with the target variable to enhance model accuracy

New Columns

Adding new features- Spirits Direct, Flavored Vodka etc. to capture additional data insights

MODEL DESCRIPTION

CLUSTERING



STORE-LEVEL:

Features: Vodka sales, retail price, household demographics, store age, average net worth.

DEMOGRAPHICS & DEMAND:

Features: Vodka sales, household demographics, vodka-wine ratio, stock availability.

PRODUCT:

Features: Packaging types (1L, 750ml), vodka-tequila ratios.

SHELF SPACE OPTIMIZATION:

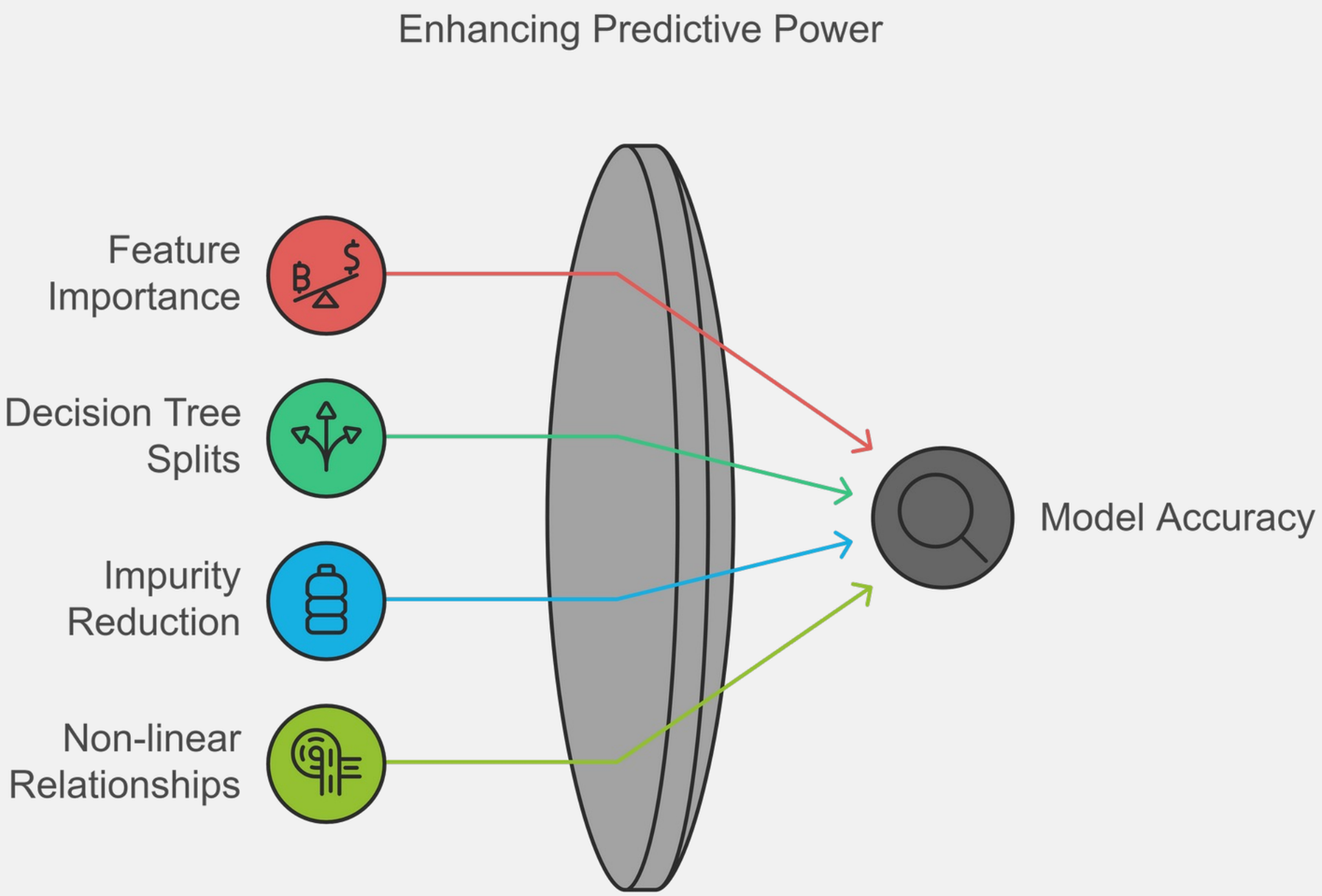
Features: Stock availability (normalized weeks in stock), retail price, vodka-tequila ratio.

STOCKING ANOMALY:

Features: Stock availability, retail price, vodka-wine ratio.

FEATURE SELECTION

USING RANDOM FOREST

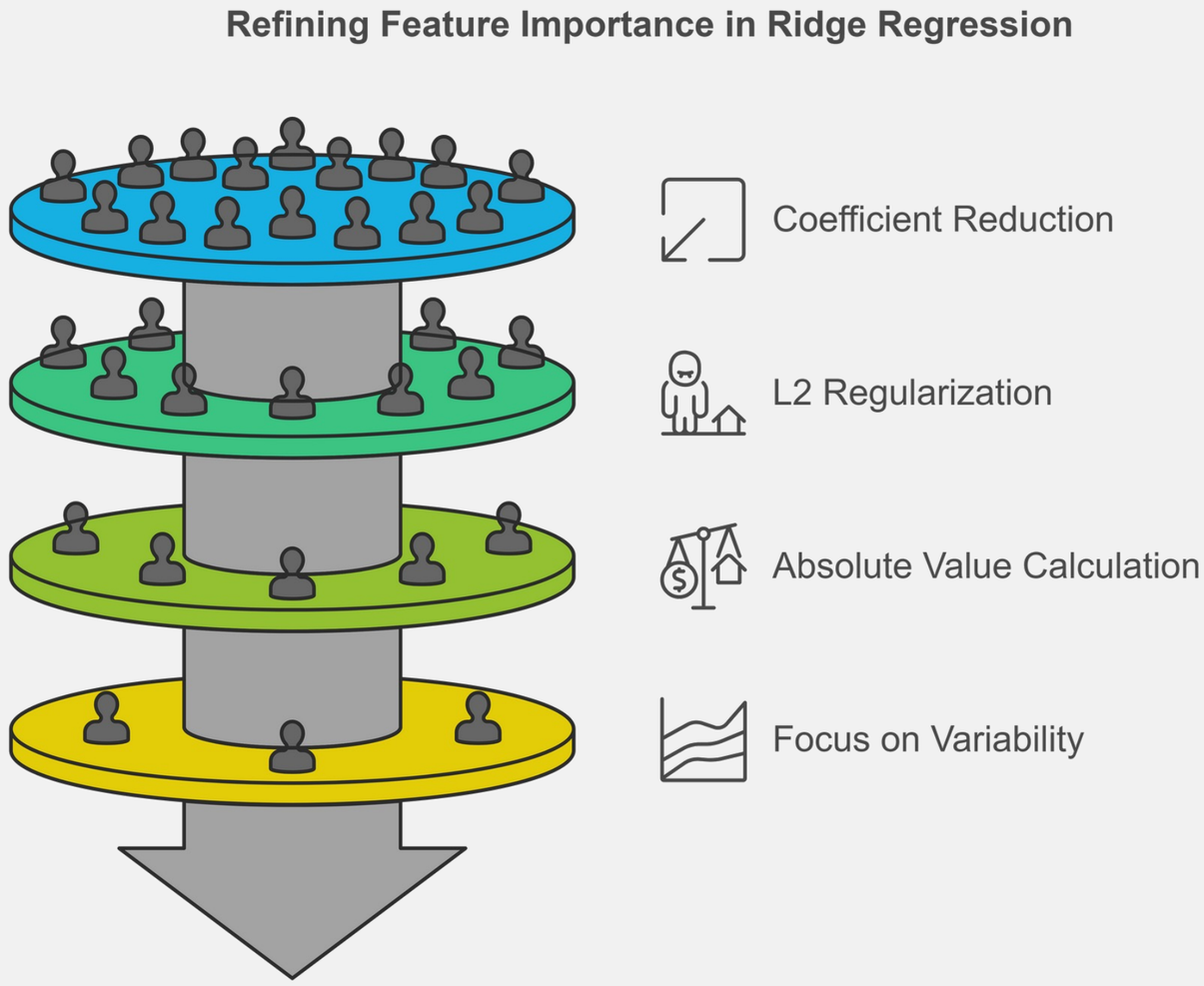


- Ensemble model evaluating feature contributions across decision trees
- Importance reflects the reduction in impurity (Gini/variance)
- Interaction and external terms such as Vodka_Wine_Ratio, Flavored_Vodka, and Spirits Direct showed significance

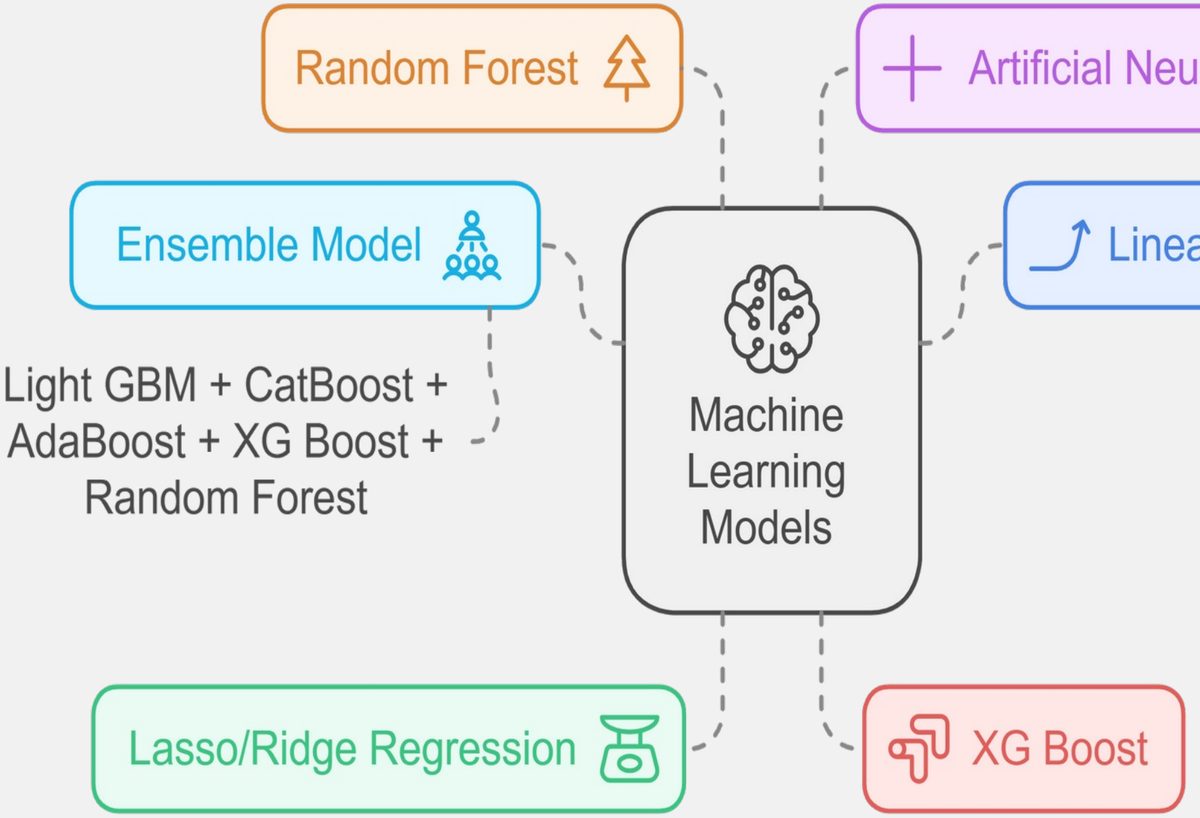
FEATURE SELECTION

USING RIDGE REGRESSOR

- Ridge model reduces overfitting
- Sensitive to scaling; standardized features gain prominence
- Top Features: Store_State_NY, Count_Week_Instock_Normalized, and package-related variables (e.g., Package_Type_750gft)



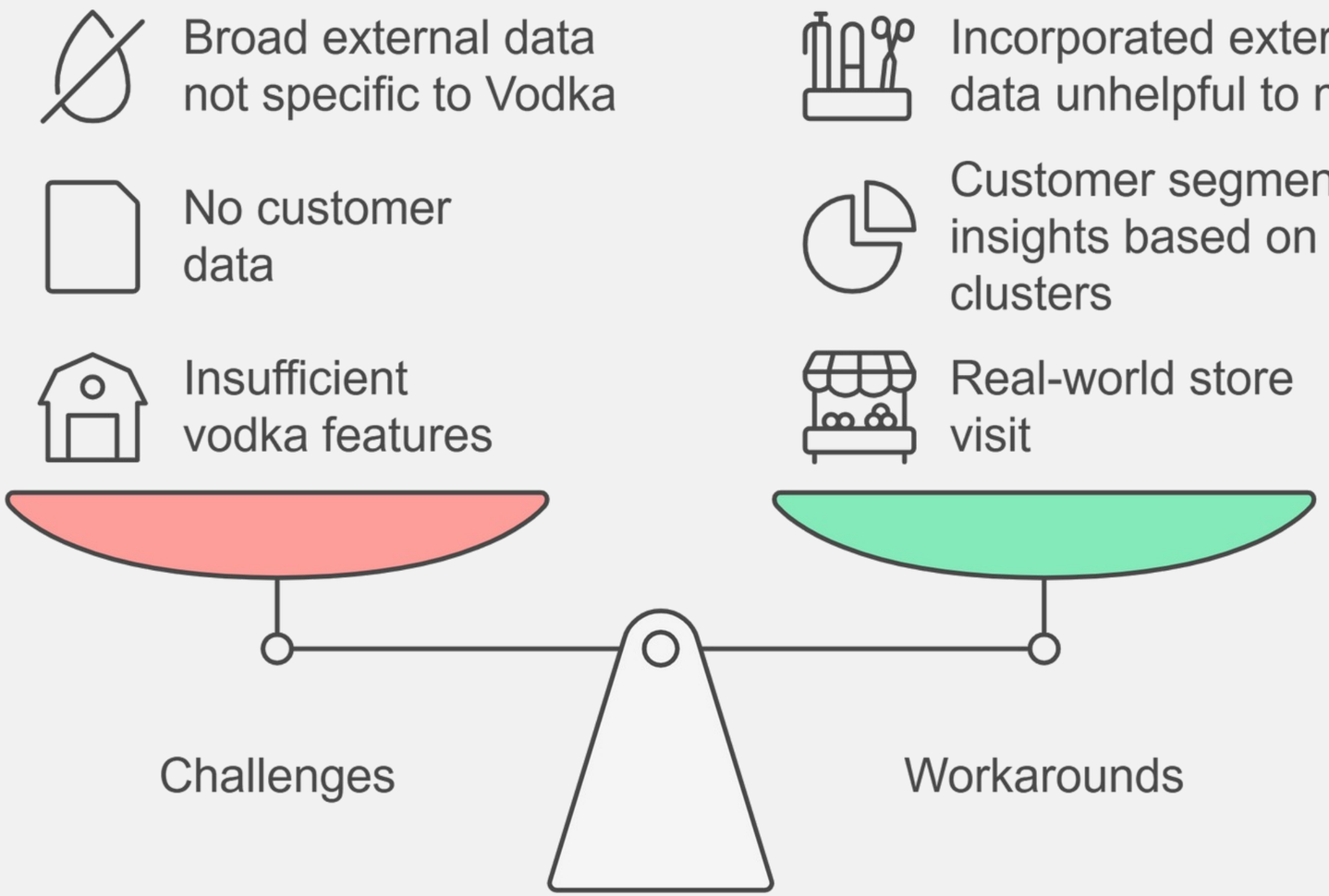
MODELING



	Validation RSME	Validation R2
Ensemble Model	11918	0.5264
Xgboost	14034	0.2848
Random Forest	15635	0.1850

CHALLENGES

CHALLENGES & WORKAROUNDS

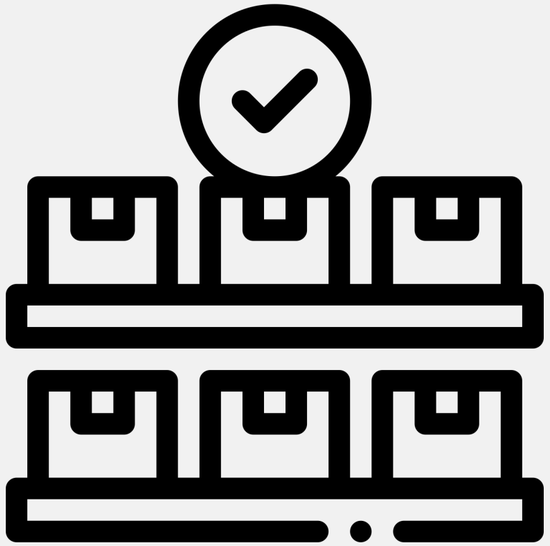


RECOMMENDATIONS

STORE ASSORTMENT

- Predictive models can forecast sales of Vodka brands, aiding in strategic product assortment planning

- Optimizing the assortment boosts overall category sales by maximizing demand and enhances operations by aligning supply with demand



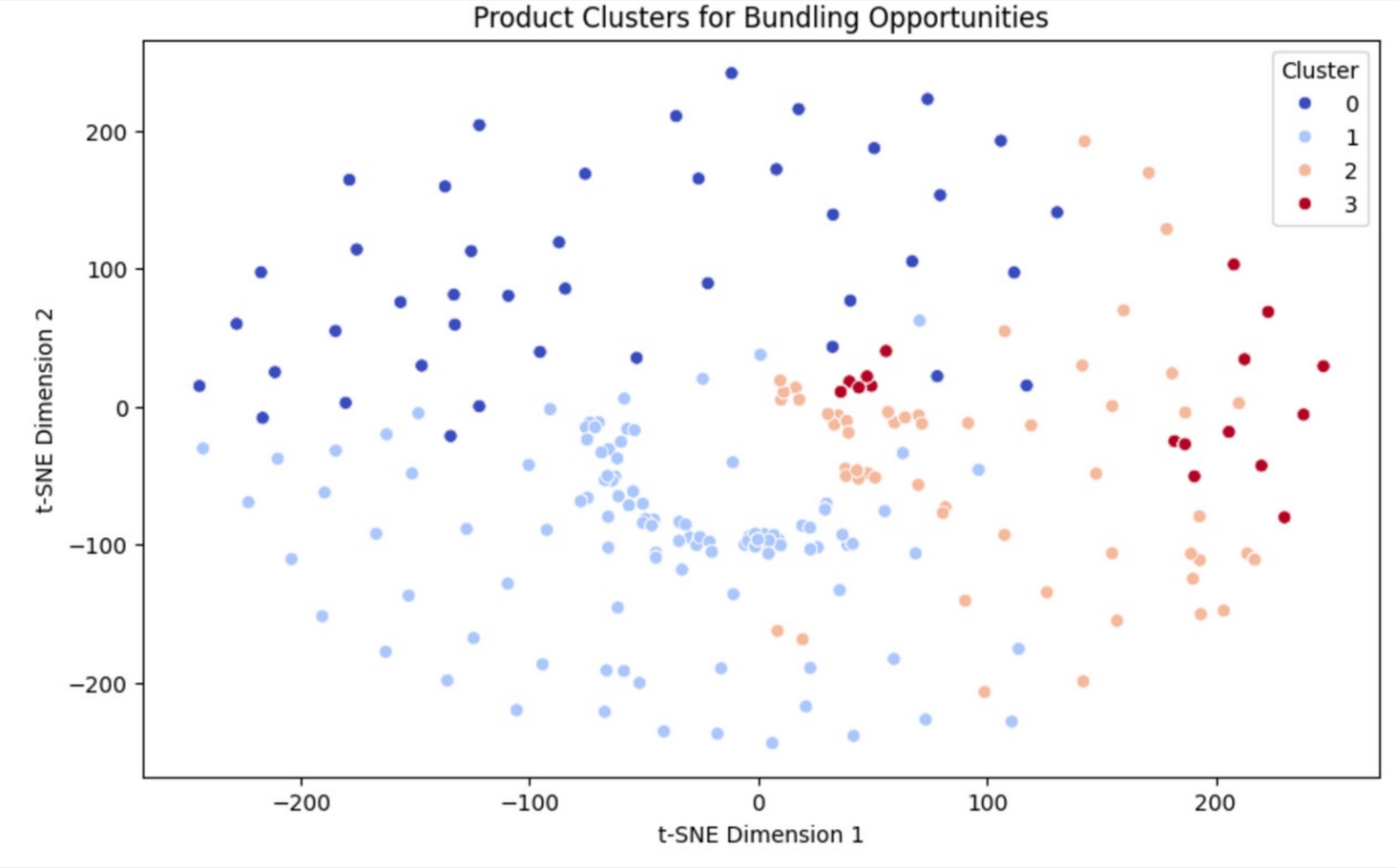
CLUSTERING: CUSTOMERS

Cluster 0: Target niche promotions and unique products

Cluster 1: Stock high-demand, mid-priced brands

Cluster 2: Emphasize premium product selection

CLUSTERING: PRODUCTS



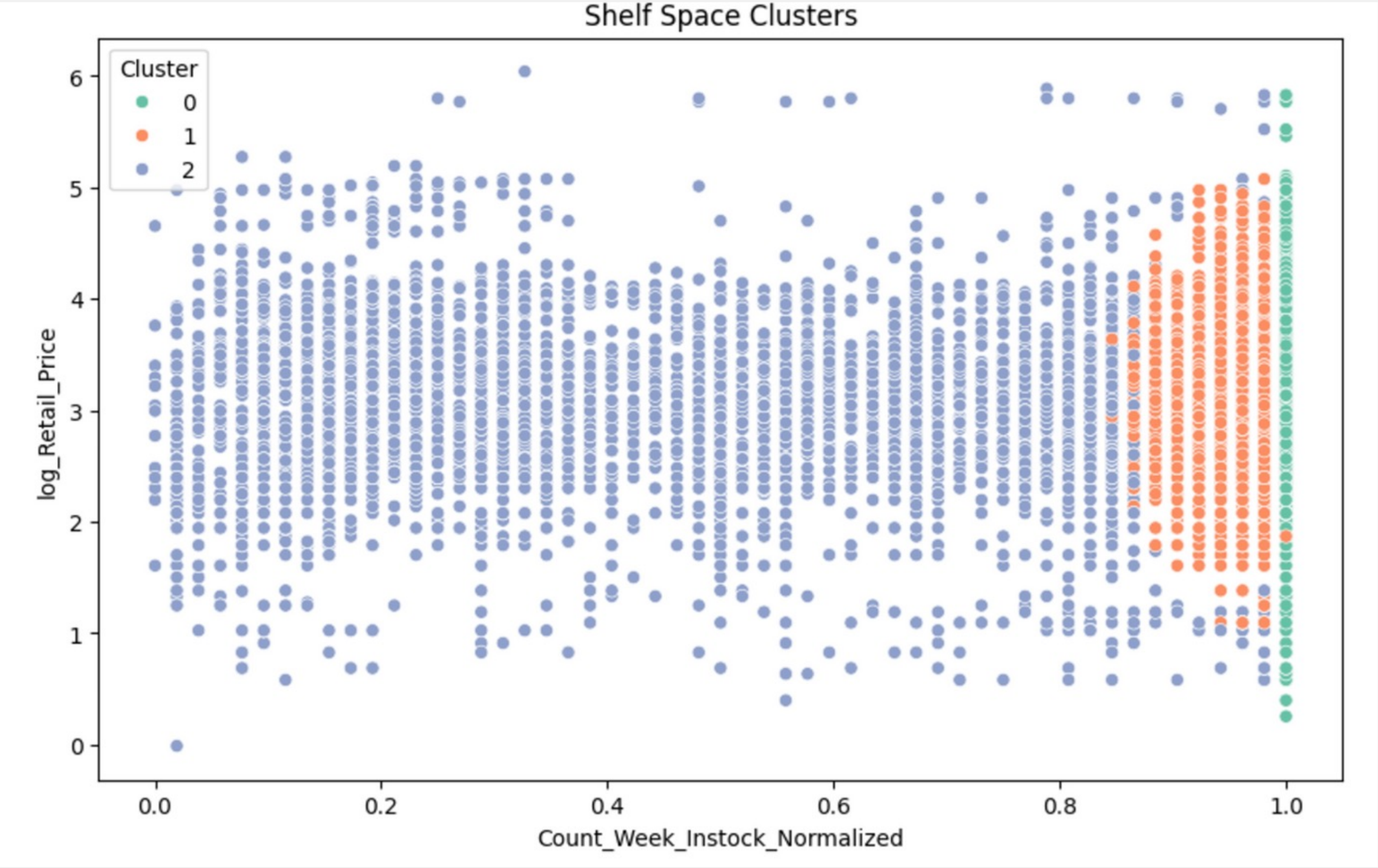
Cluster 0: Broad product range with moderate similarities; suitable for general-purpose bundles targeting a wide audience

Cluster 1: Tightly grouped products with consistent characteristics; ideal for price-based or complementary bundles

Cluster 2: Distinct group of premium/niche products; suitable for high-end or limited-edition bundles

Cluster 3: Small, unique product group; best promoted individually or paired with popular products to boost cross-sales

CLUSTERING: SHELF-SPACE

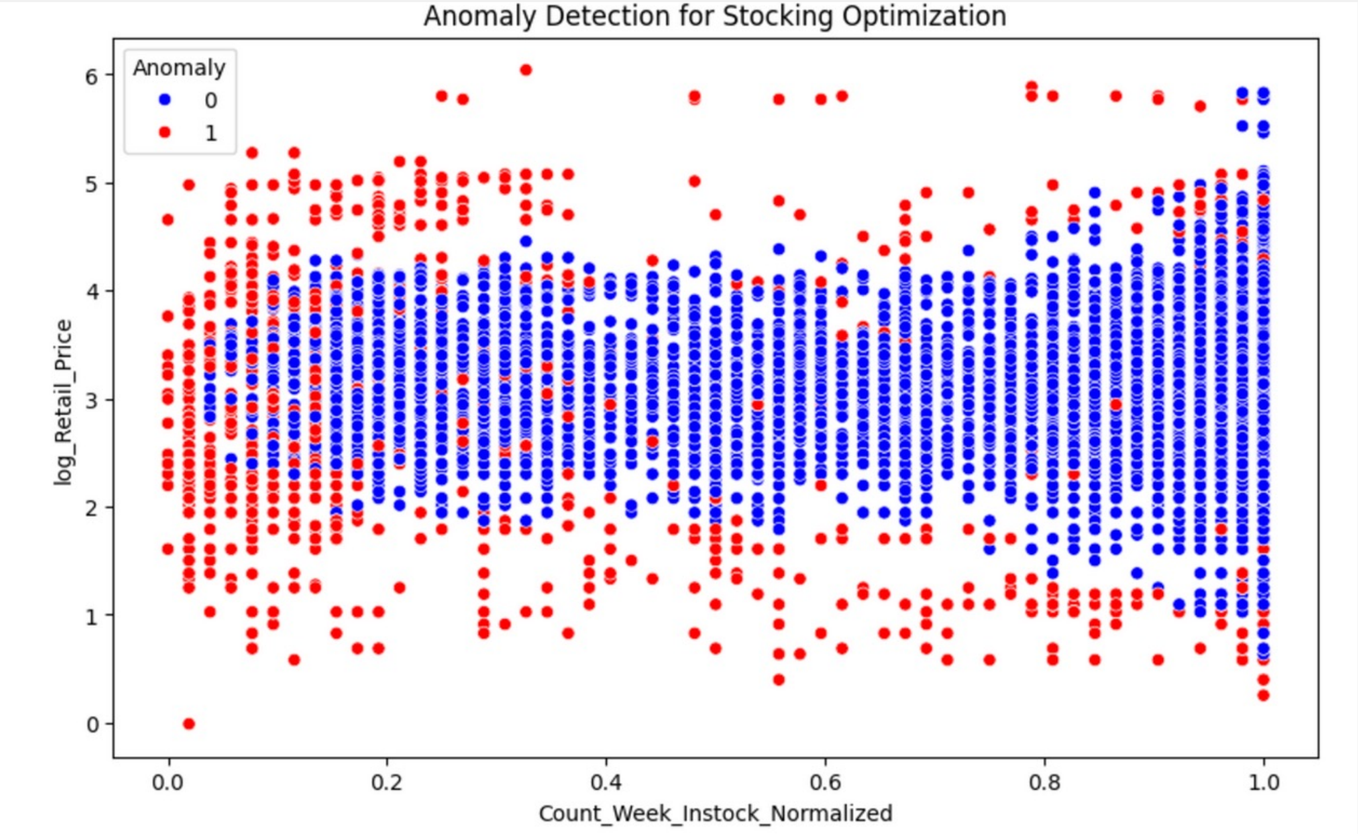


Cluster 0: Prioritize shelf space for easy customer access to high-demand, frequent-purchase items

Cluster 1: Highlight premium products with end caps or creative shelf arrangements to boost cross-sales

Cluster 2: Focus on improving stock consistency and rationalizing shelf space for low-demand items

CLUSTERING: ANOMALY



Normal Products (Blue):
Most products display consistent stocking and pricing behaviors

Anomalous Products (Red):
High Availability & High Price: Premium products with unexpectedly high availability or low demand

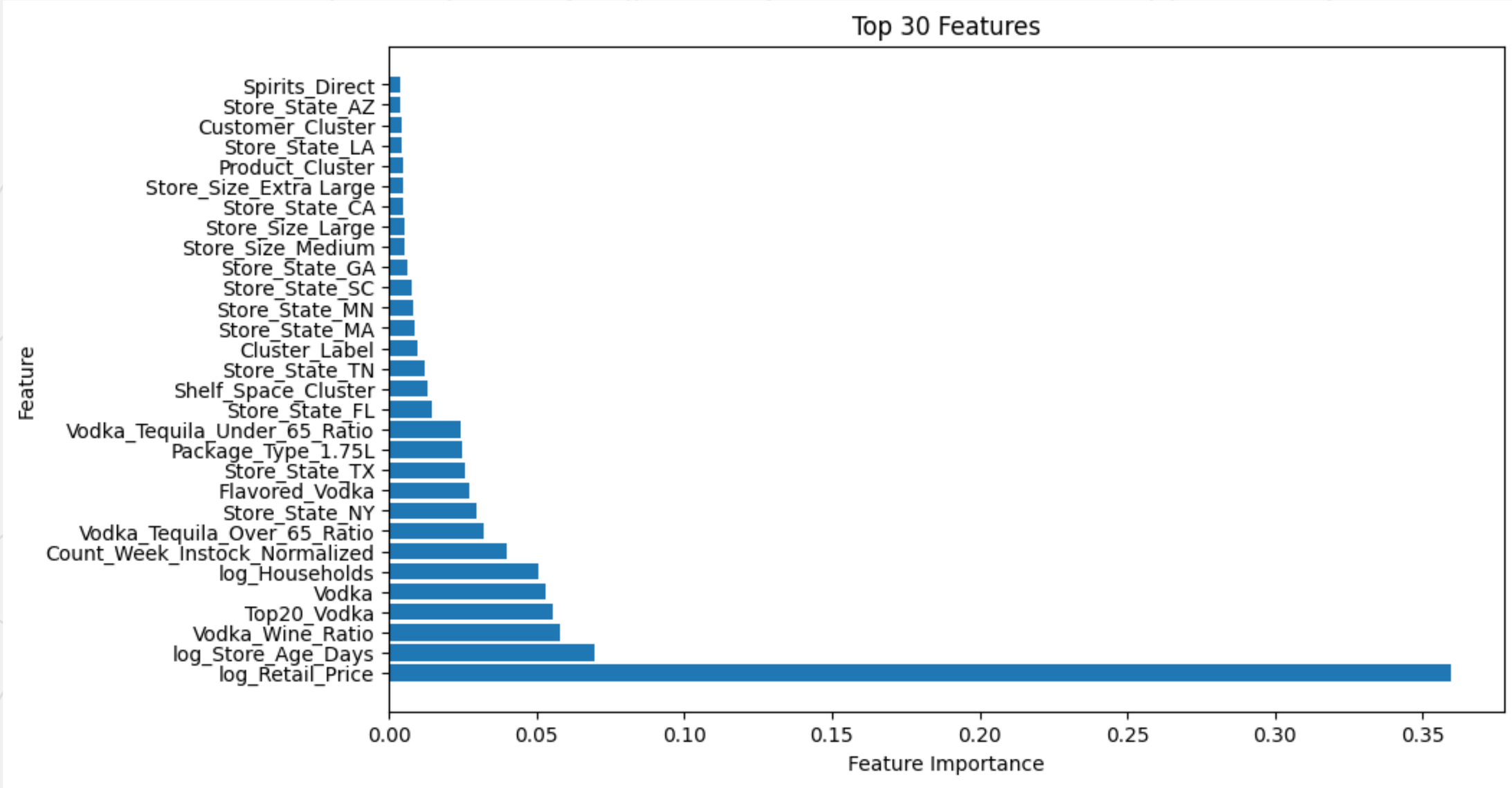
Low Availability & Low Price:
Budget products with inconsistent stocking or low sales

Mid-Price & Low Availability:
Potential missed sales opportunities due to inconsistent stock levels

THANK YOU!

APPENDIX

TOP FEATURES RF



TOP FEATURES RIDGE

	Feature	Importance
25	Store_State_NY	10354.770741
43	Count_Week_Instock_Normalized	5306.913583
41	Package_Type_750gft	4583.502960
32	Package_Type_1.75L	4361.486191
52	Flavored_Vodka	4262.325560
33	Package_Type_1.75Lgft	3914.442949
51	Spirits_Direct	3417.049646
37	Package_Type_200ml	2923.691662
36	Package_Type_200-3gft	2750.759802
34	Package_Type_100ml	2690.280759
38	Package_Type_375ml	2159.482461
39	Package_Type_700ml	2016.070693
53	Top20_Vodka	1823.494208
8	Store_State_CT	1561.321683
6	Store_State_CA	1557.659398

TOP 20 VODKA

The Spirits Business's annual Brand Champions report

- Smirnoff
- Absolut
- Zubrowka
- Magic Moments
- Arkhangelskaya
- Żołądkowa (including Gorzka)
- Grey Goose
- Soplica
- Pyat Ozer
- Nemiroff
- Belenkaya
- Skyy
- Talka
- Ketel One
- Finlandia
- Russian Standard
- Wodka Gorbatschow
- Tsarskaya/Imperial Collection Gold
- Green Mark
- Belaya Bereza Vodka

https://www.google.com/search?q=top+20+vodka+sales&rlz=1C1VDKB_zh-TWUS1074US1076&oq=TOP+20+VOD&gs_lcrp=EgZjaHJvbWUqBggAEEUYOzIGCAAQRRg7MgcIARAAGIAEMgYIAhBFGDkyBwgDEAAYgAQyBwgEEAAYgAQyBwgFEAAYgAQyCAgGEAAYFhgeMggIBxAAGBYYHjIICAgQABgWGB4yCAgJEAAYFhge0gEIMjI1OWowajeoAgCwAgA&sourceid=chrome&ie=UTF-8