ABS STORE PERFORMANCE ANALYSIS

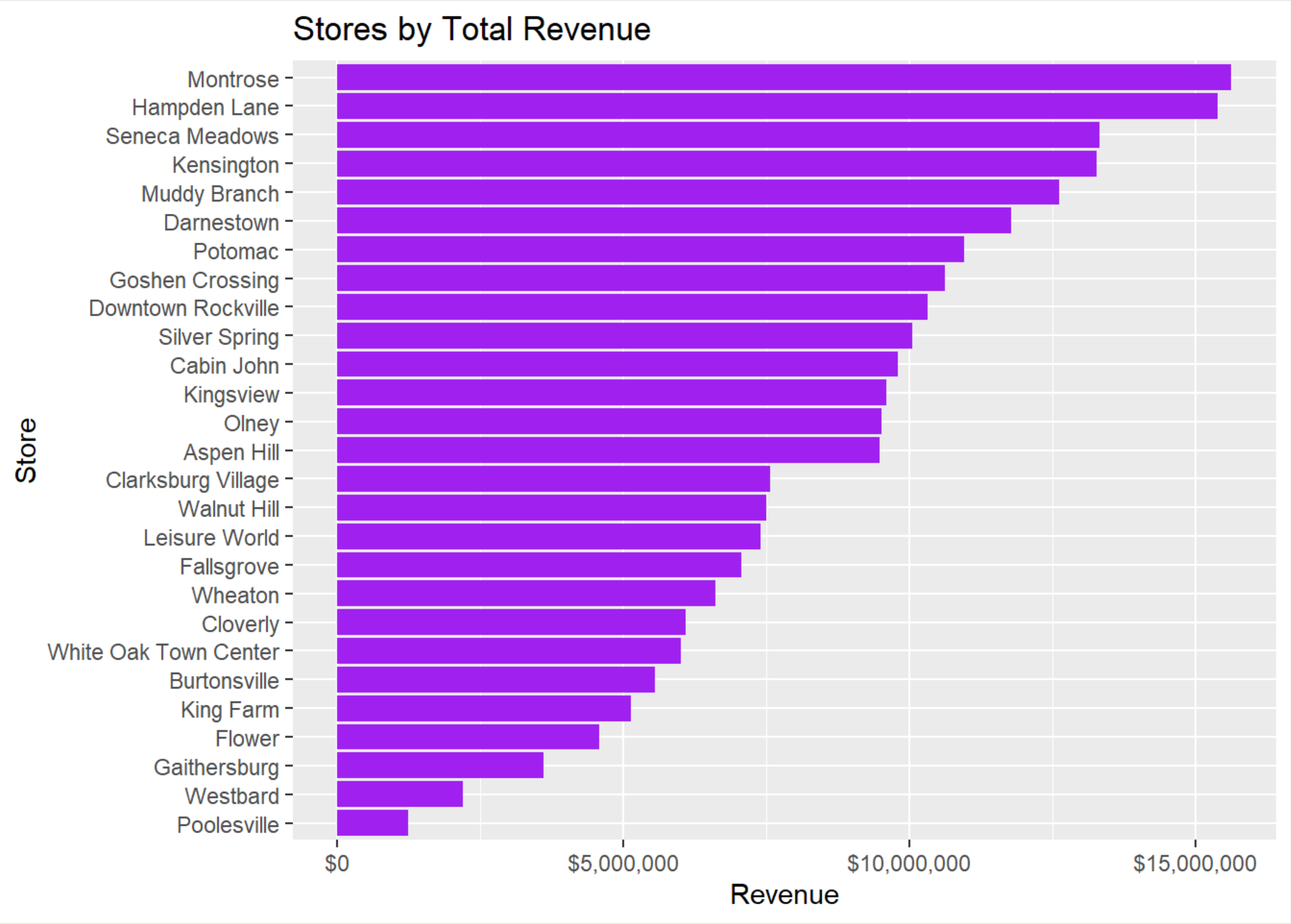
Arash Nateghian

Data 205

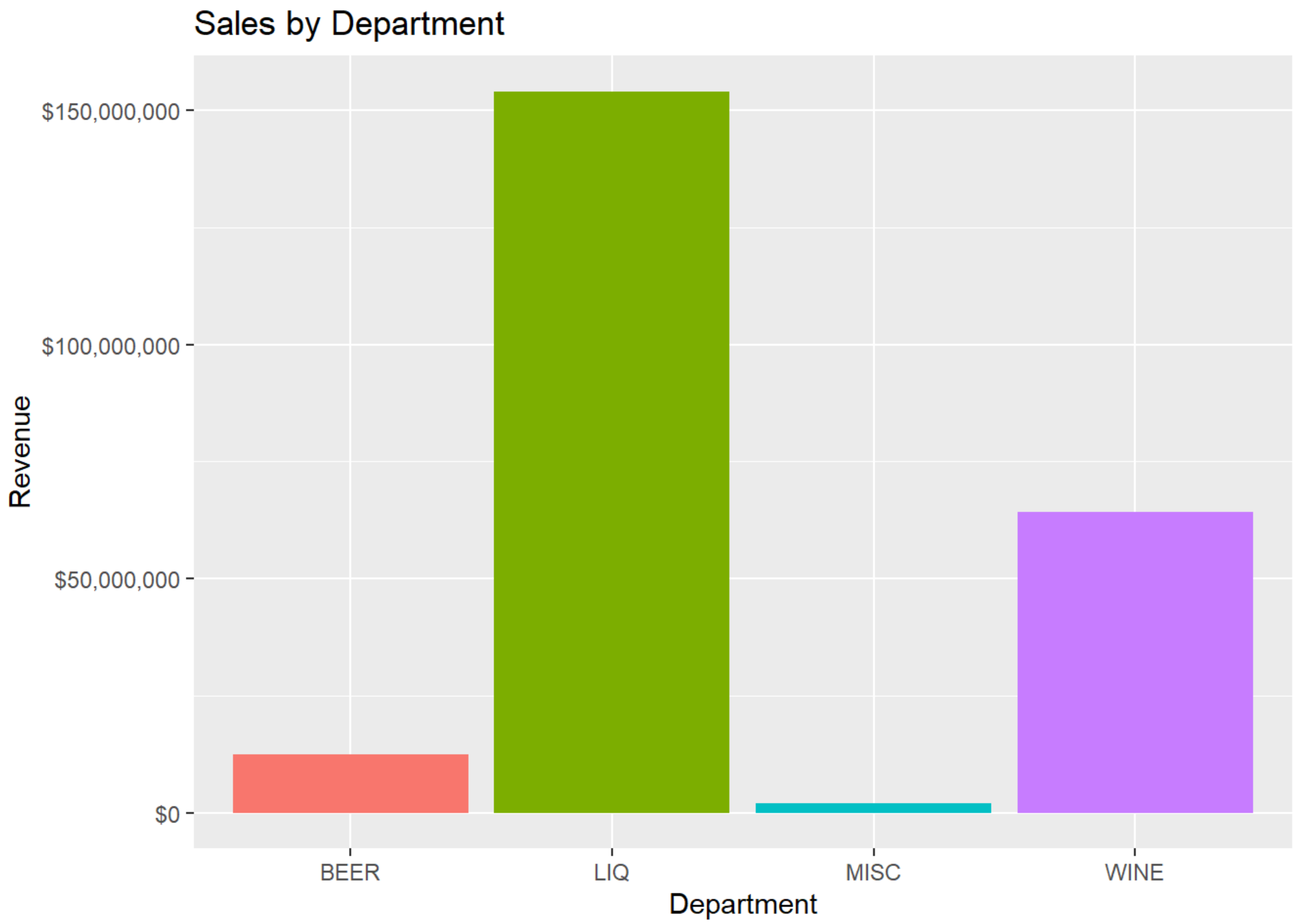
Prof. Perine

DATA OVERVIEW & CLEANING STRATEGY

- 3 datasets: transactions, designation, demographics
- Major cleaning: normalize store names, pack sizes, remove noise
- Result: analytics-ready dataset



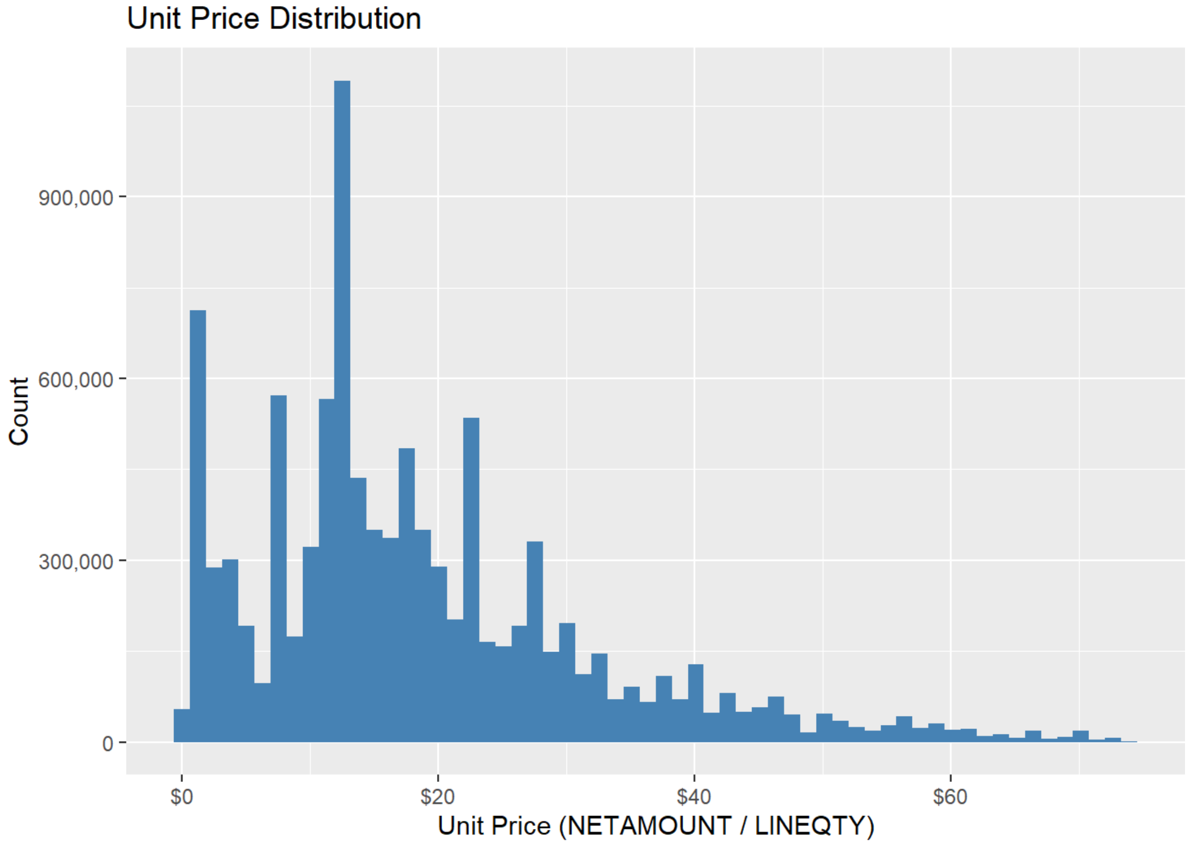
- High-revenue ≠ high efficiency
- Montrose, Hampden Lane

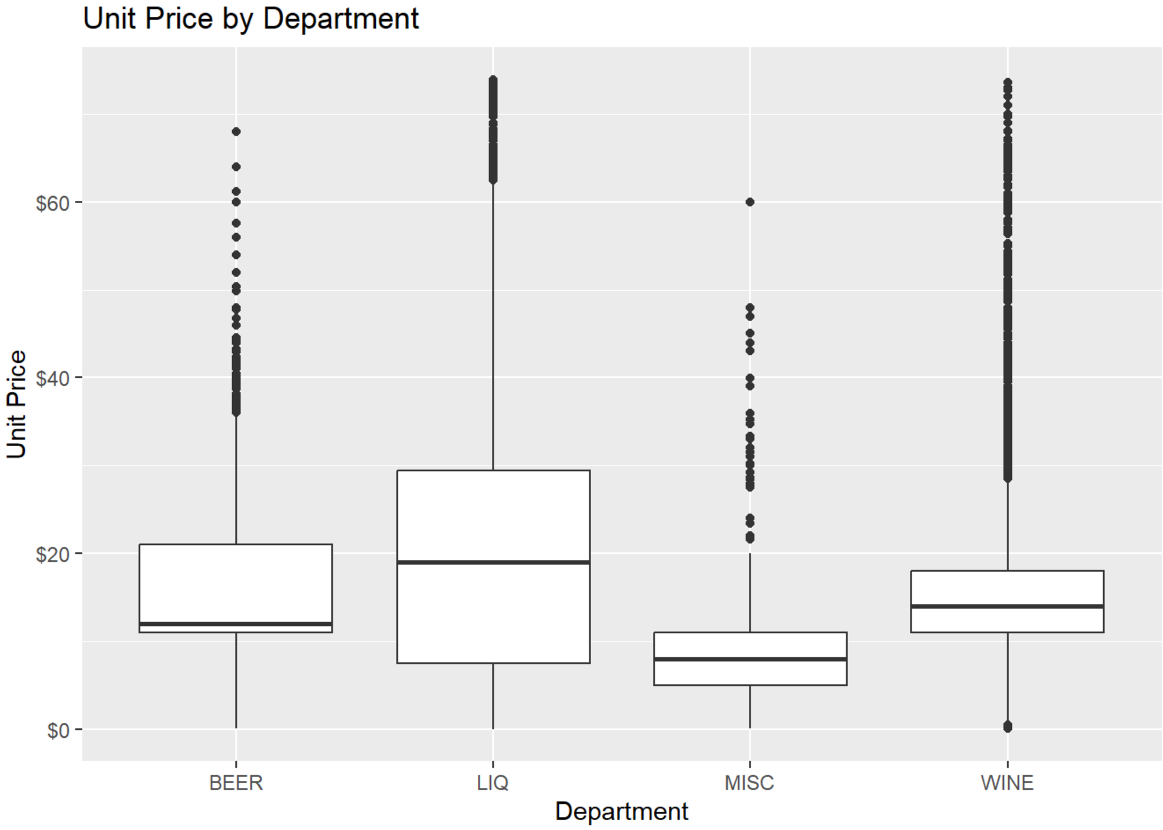


Liquor drives revenue

Beer = volume machine

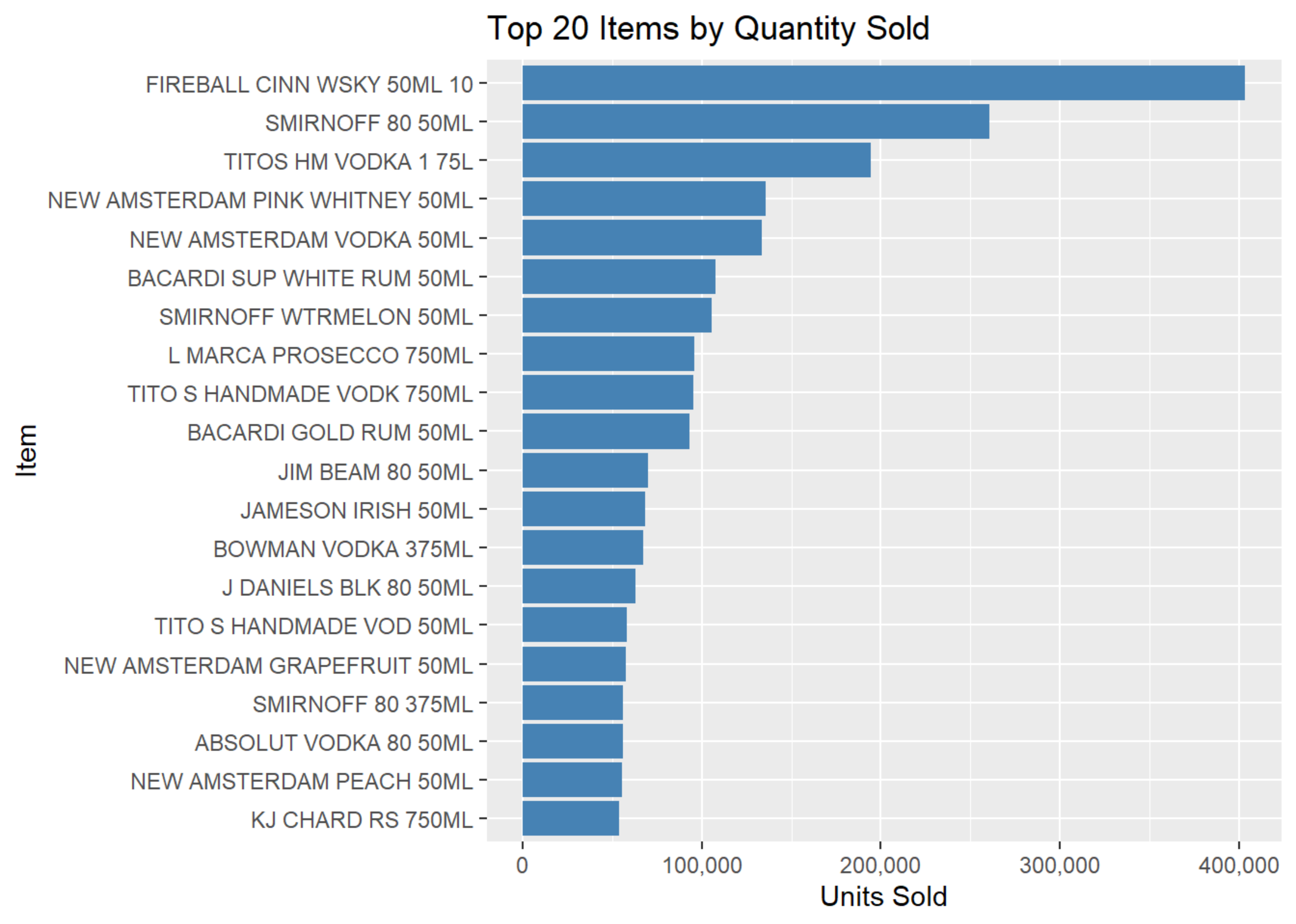
Wine moderate

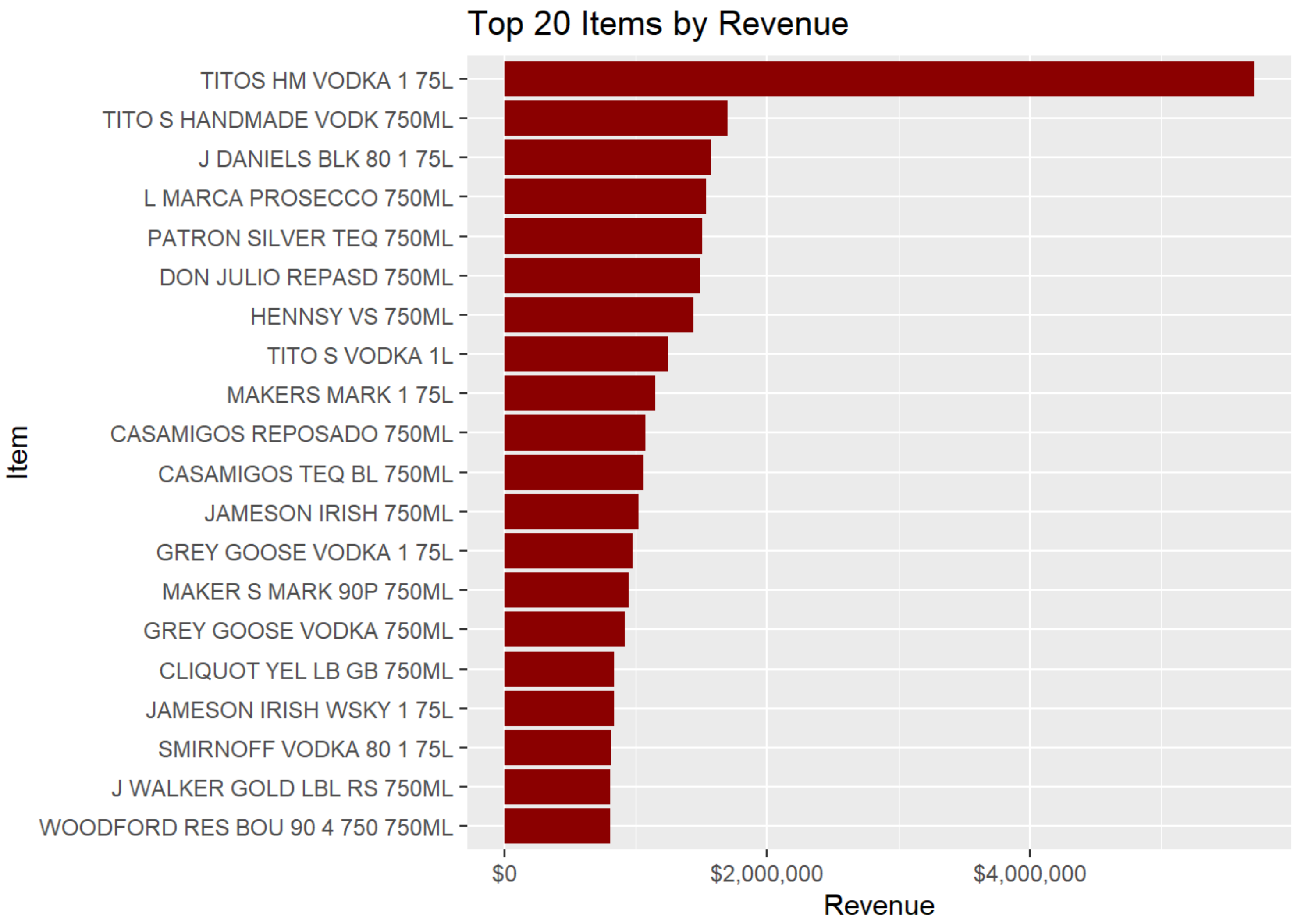




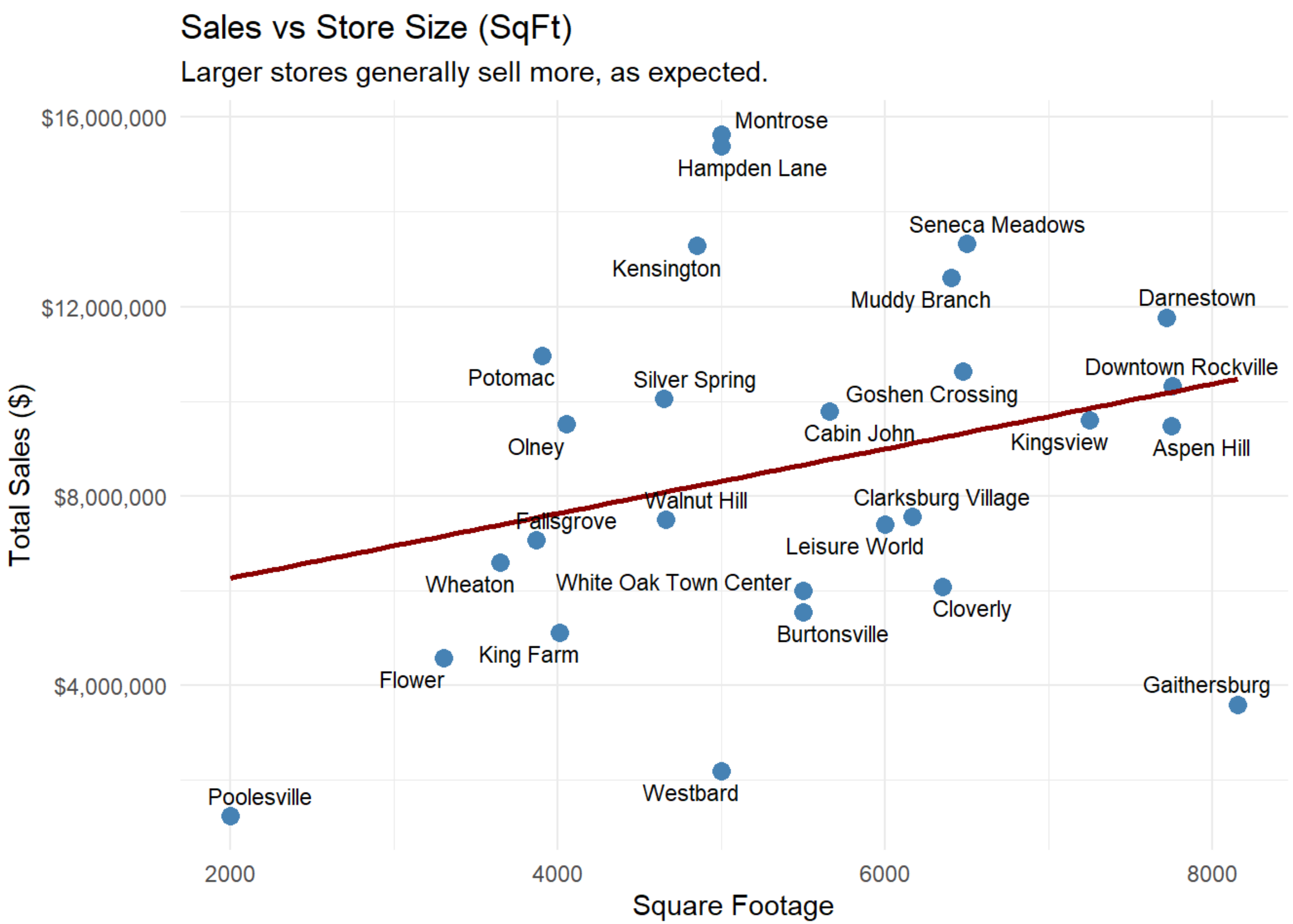
- Strong right-skew with premium items
- Liquor & wine have widest price ranges
- Beer & misc cluster at lower prices
- Premium pricing concentrated in spirits/wine

- Small bottles, beer packs
- Drives foot traffic
- Low individual revenue





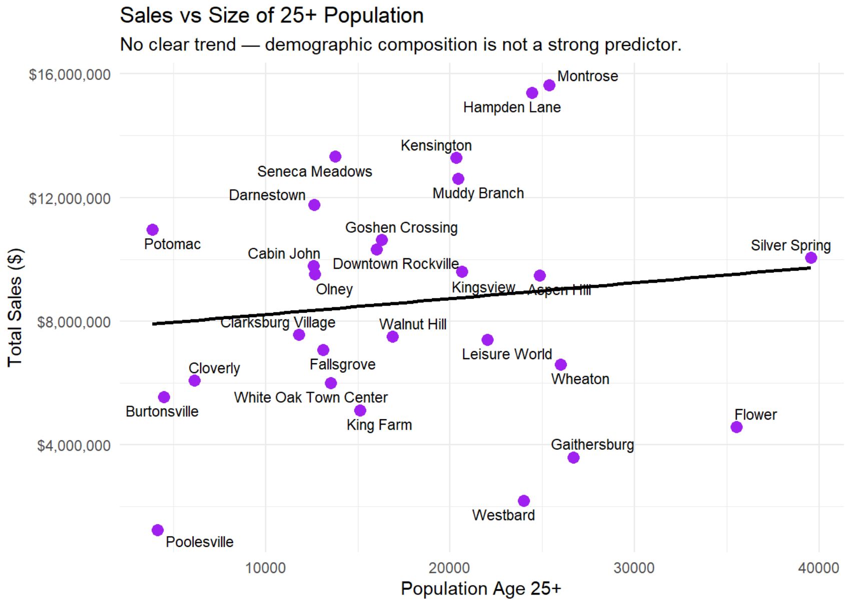
- Premium spirits dominate
- Big impact on revenue
- Fewer SKUs drive most dollars



- Weak correlation
- Bigger store ≠ better store
- Correlation = 0.2813958

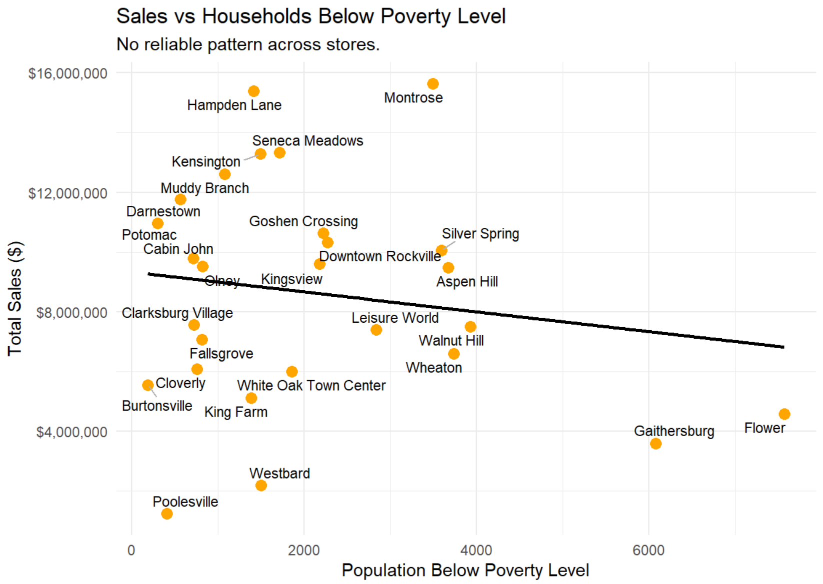
DEMOGRAPHICS VS SALES

Correlation = 0.0928748

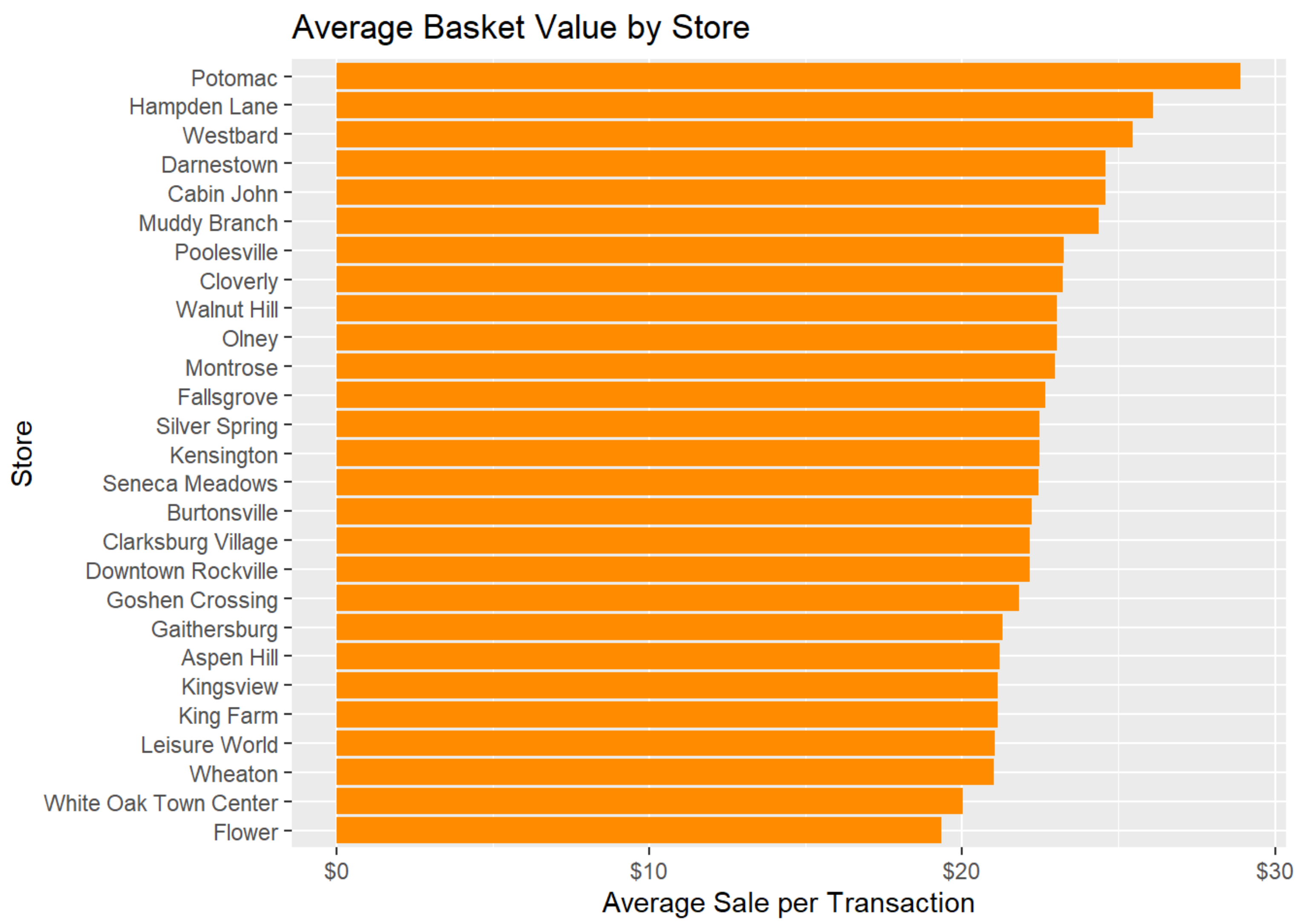


Correlation = 0.1209889

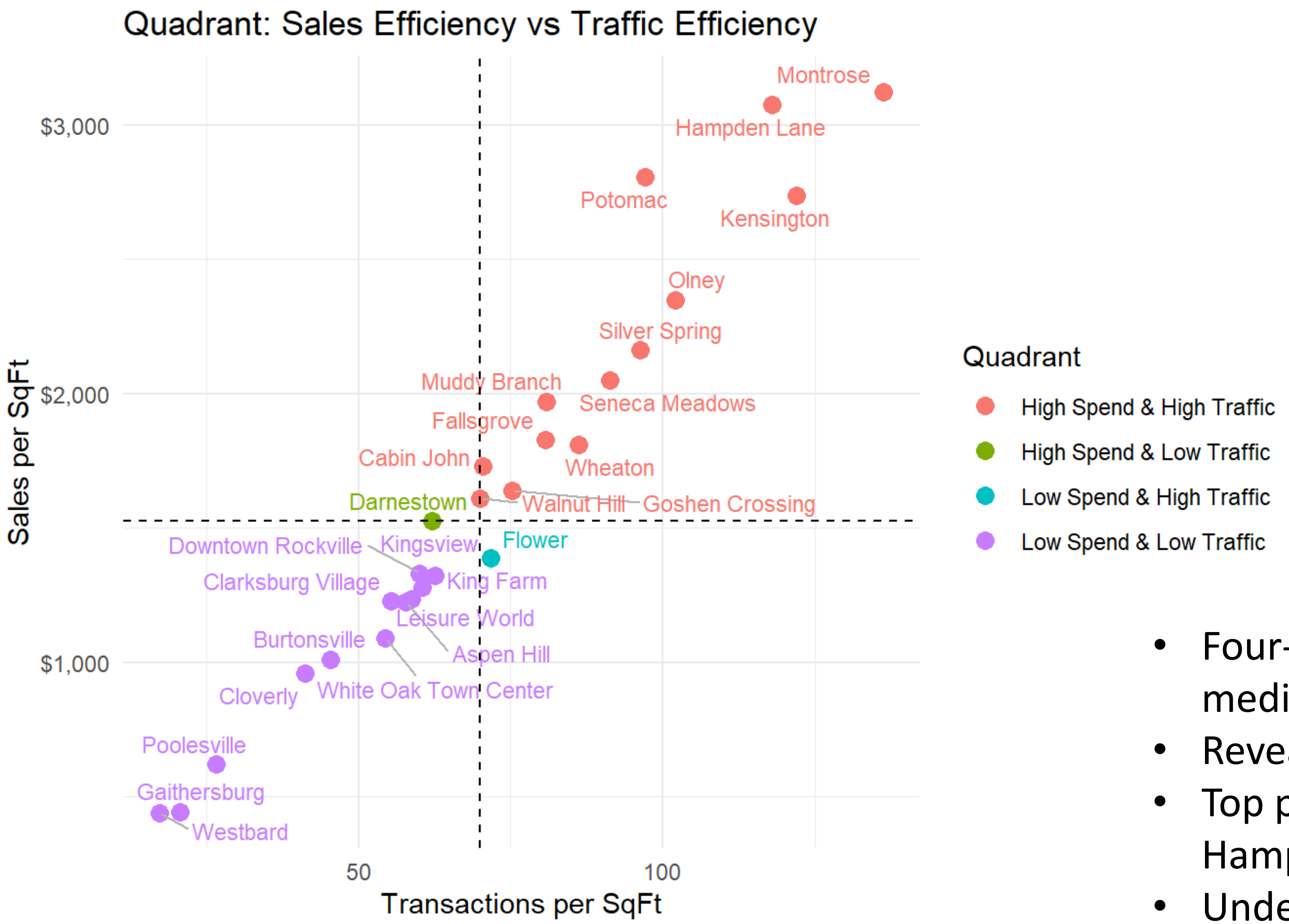
- Demographics show weak predictive power
- Performance differences arise from in-store behavior, not population
- Reinforces the need for operational and assortment strategies



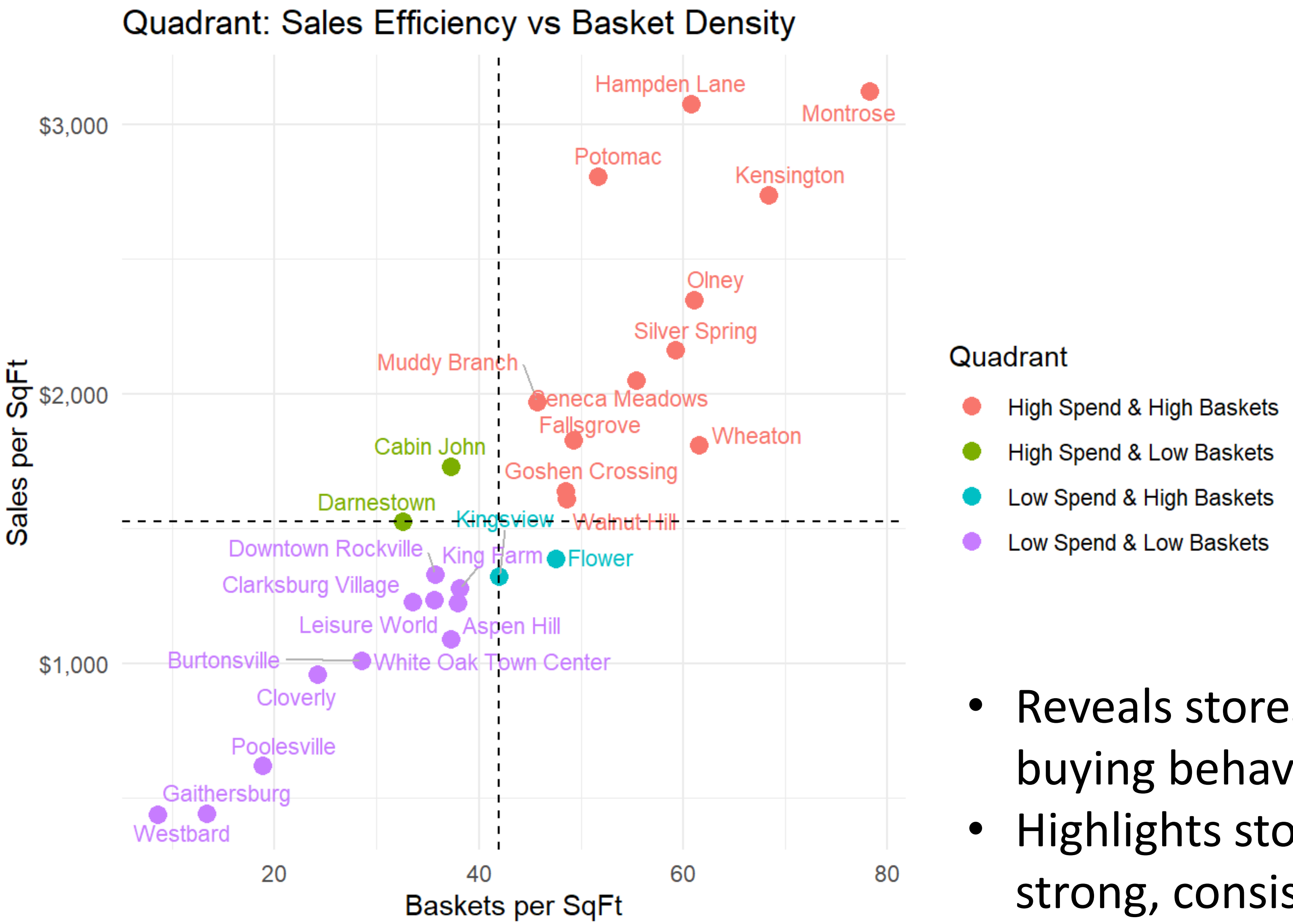
Correlation = -0.1579369



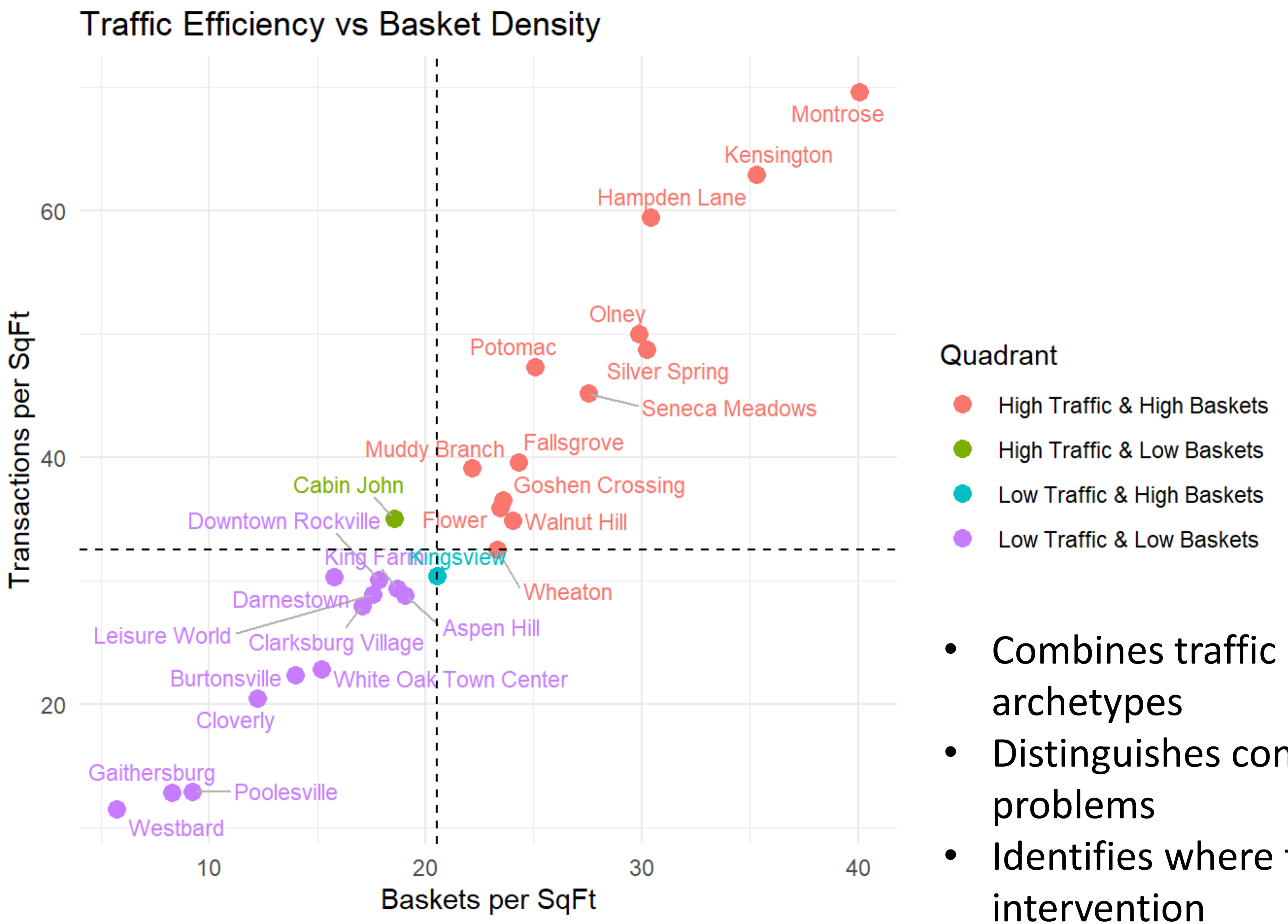
- Very wide variation
- Customer behavior differs by location



- Four-quadrant segmentation based on medians
- Reveals structural positions across stores
- Top performers: Montrose, Kensington, Hampden Lane
- Underperformers: Westbard, Poolesville, Gaithersburg
- Traffic is influential, but not sufficient → operational gaps remain



- Reveals stores with high traffic but weak buying behavior
- Highlights stores with low traffic but strong, consistent baskets
- Shows basket density as a critical driver of sales productivity



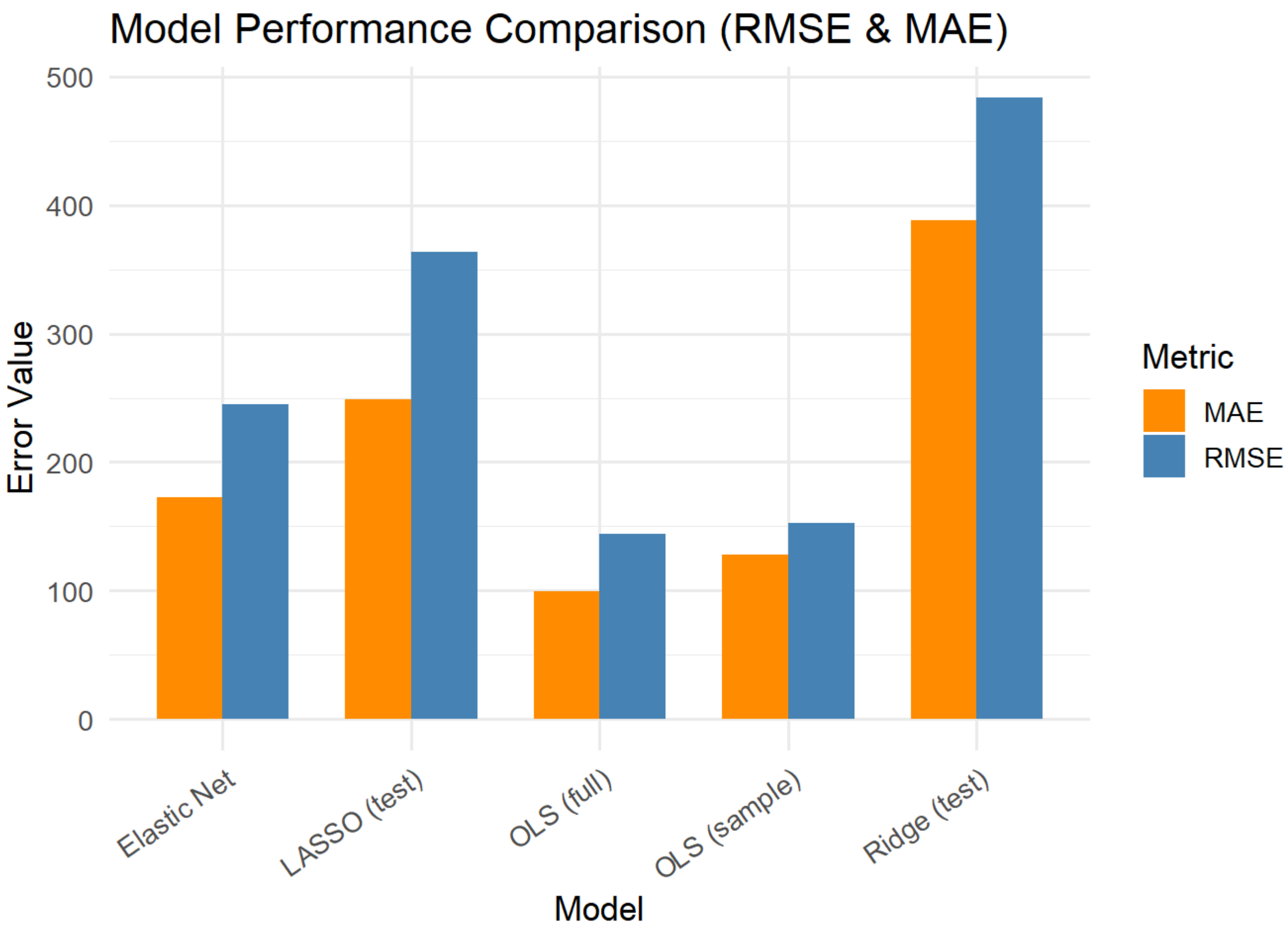
- Combines traffic and basket density into four store archetypes
- Distinguishes conversion problems from demand problems
- Identifies where traffic, basket value, or both need intervention
- Provides the framework for targeted operational strategy

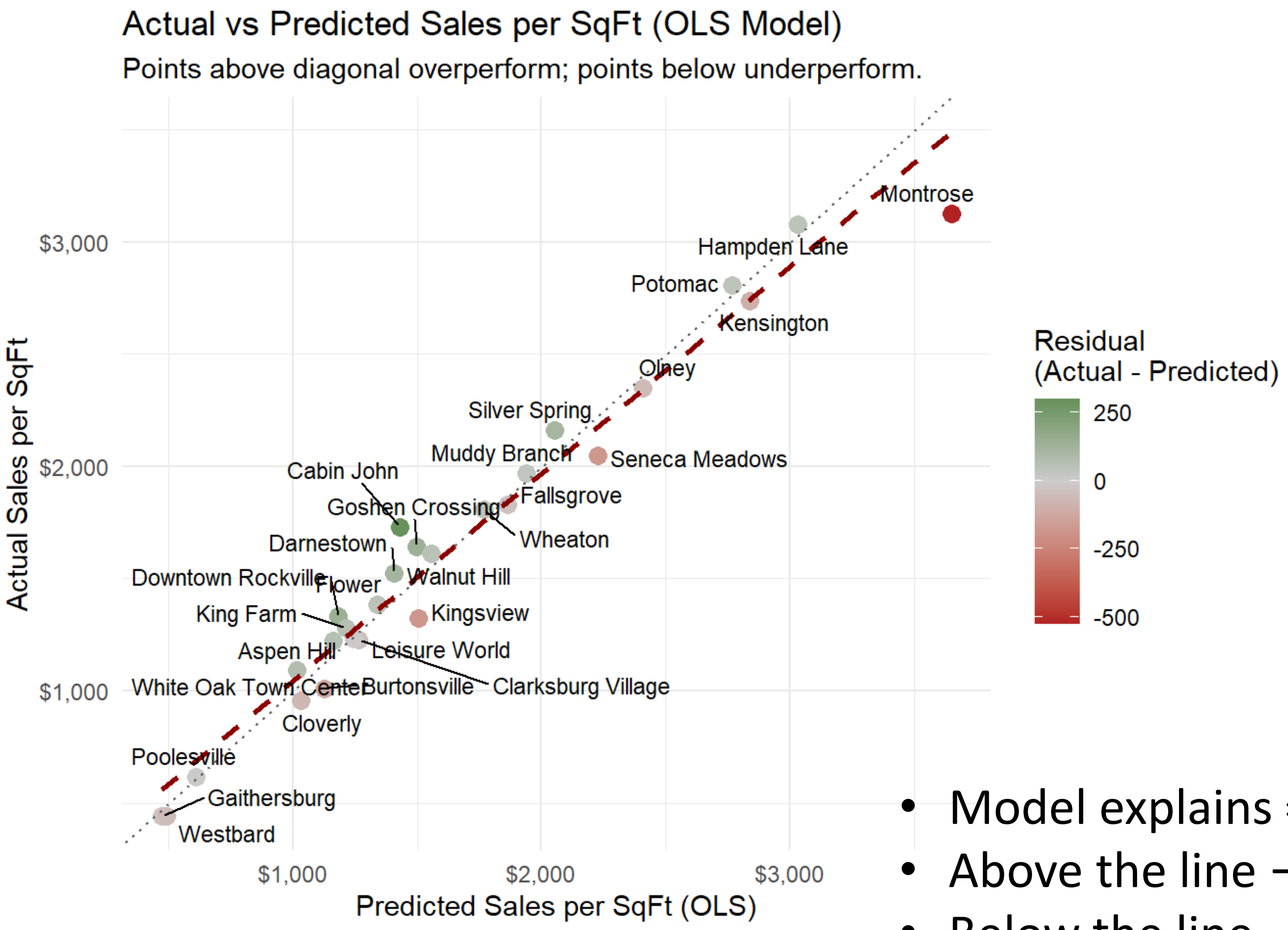
OLS Coefficient Significance Table (α = 0.05)

Predictor	Estimate	Std. Error	t-value	p-value	Significance
TotalTransactionsPerSqFt	0.01607	0.000915	17.565	6.2e-11	**Strong***
Share_BEER	56.82	19.14	2.97	0.0102	**Significant**
Share_LIQ	55.69	18.45	3.02	0.0092	**Significant**
Share_WINE	55.66	18.08	3.08	0.0082	**Significant**
Share_Weekday	−76.74	35.65	−2.15	0.0493	**Significant**
Share_Weekend	−80.51	36.00	−2.24	0.0421	**Significant**
Avg_BasketValue	0.0204	0.0101	2.02	0.0634	Borderline
AvgItemsPerBasket	−0.2195	0.2440	−0.90	0.3835	NS
SquareFootage	0.0000163	0.0000199	0.82	0.4250	NS
TotalPopulation	0.0000241	0.0000300	0.80	0.4358	NS
Older25Years	−0.000031	0.0000392	−0.79	0.4421	NS
PovertyLevel	−0.000016	0.0000316	−0.51	0.6163	NS

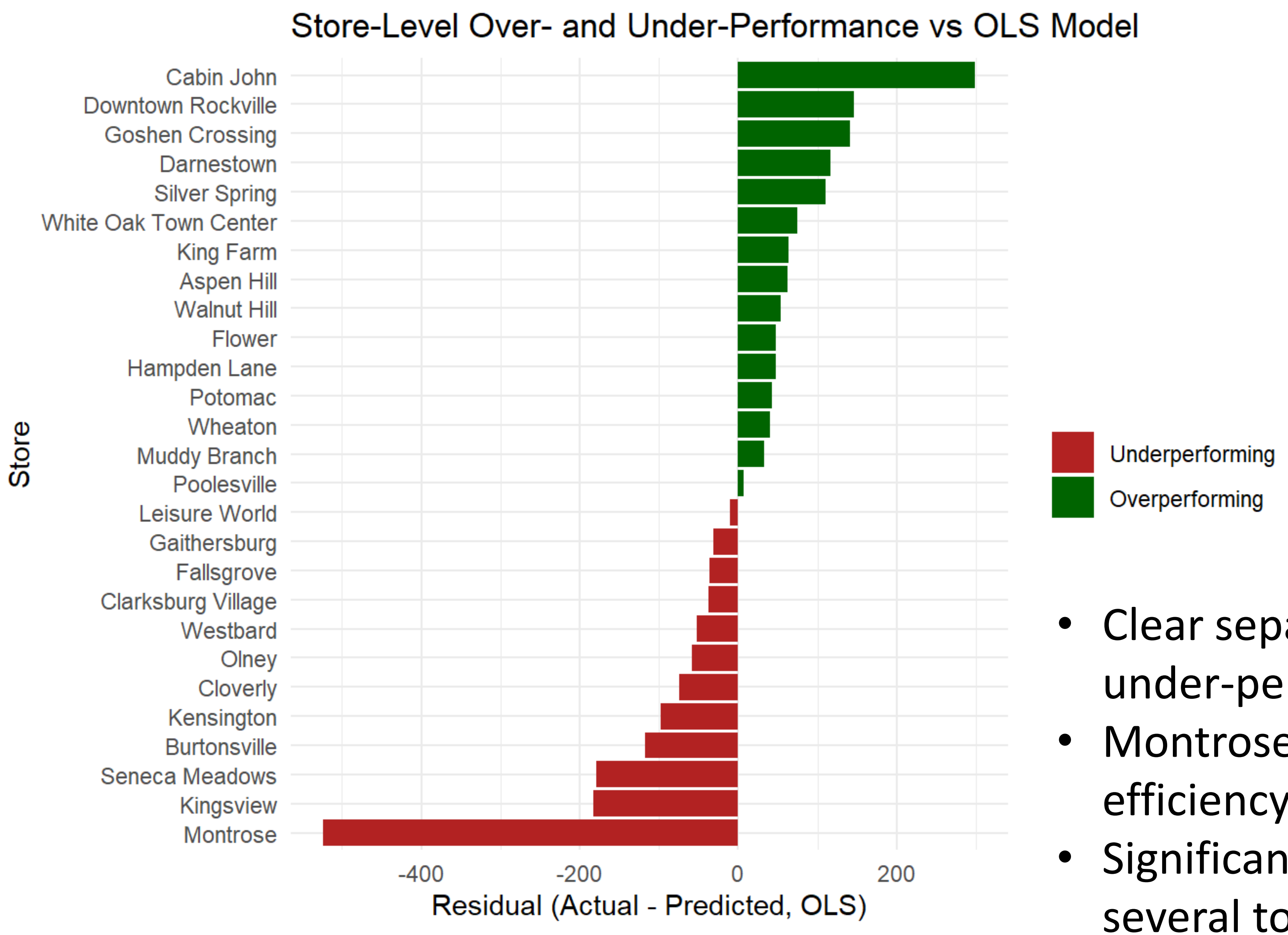
Model Comparison Table (Fit Metrics)

Model	RMSE	MAE
OLS (Full Model)	**144.11**	**99.46**
OLS (Sampled Train/Test)	**152.52**	**127.68**
Elastic Net	**245.04**	**172.54**
Ridge (Test)	**483.30**	**388.38**
LASSO (Test)	**363.41**	**248.77**

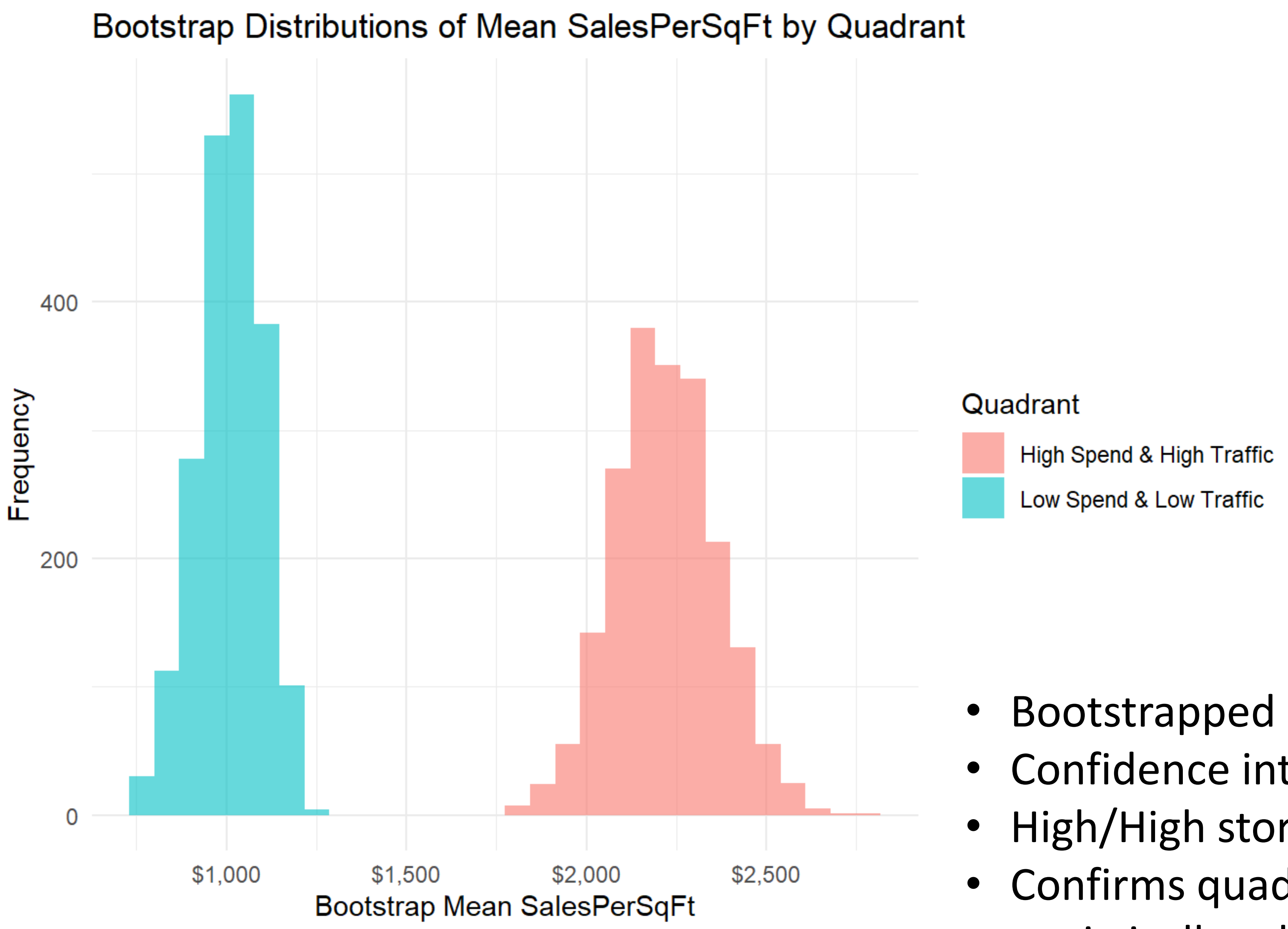




- Model explains ≈97% of variation in sales per SqFt
- Above the line → outperforming expectations
- Below the line → underperforming expectations
- Establishes an objective performance benchmark



- Clear separation of true over- and under-performance
- Montrose: high revenue ≠ high efficiency
- Significant unrealized value in several top stores



- Bootstrapped means: **~$2,200 vs ~$900**
- Confidence intervals do not overlap
- High/High stores are genuinely stronger
- Confirms quadrant segmentation is statistically robust

KEY INSIGHTS

- Traffic intensity is the strongest driver of performance
- Basket density + category mix determine revenue efficiency
- Demographics provide little explanatory power
- Quadrants reveal stable behavioral patterns across stores
- OLS residuals expose hidden underperformance
- Bootstrapping confirms differences are statistically significant
- Foundation for targeted, store-specific operational improvements

RECOMMENDATION

Boost traffic where it is weak; raise basket value where it underperforms

Use quadrant positions to target the highest-impact lever at each store

Tighten replenishment and align ordering with observed buying patterns

Reduce long-tail SKUs to increase efficiency and improve shopper navigation

Tailor promotions to quadrant type: *premium pushes* for high-traffic stores, *traffic builders* for low-traffic stores

Use the KPI monitoring framework for continuous, data-driven oversight

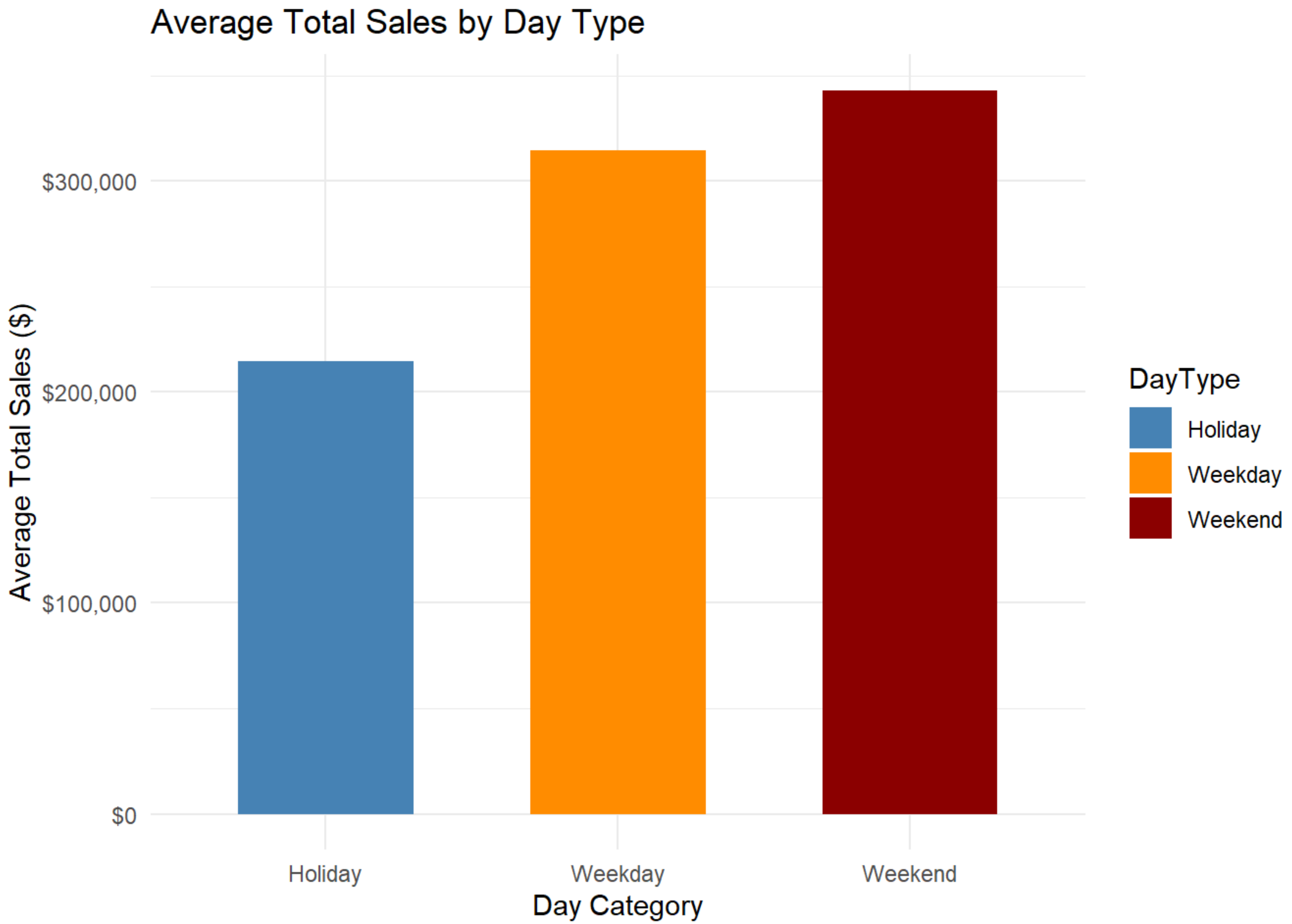
ACKNOWLEDGMENT

- **The ABS team** for meeting with me weekly, helping me understand the dataset, and making sure I stayed on the right analytical path.
- **The Data Montgomery team** for coordinating meetings, supporting communication, and helping identify supplementary datasets—even the ones we ultimately didn’t need.
- **Prof. Lori Perine** for her guidance on how to advance and refine my modeling approach.
- **My teammate Andre**, whose thoughtful input consistently sharpened my reasoning and guided my analytical decisions.

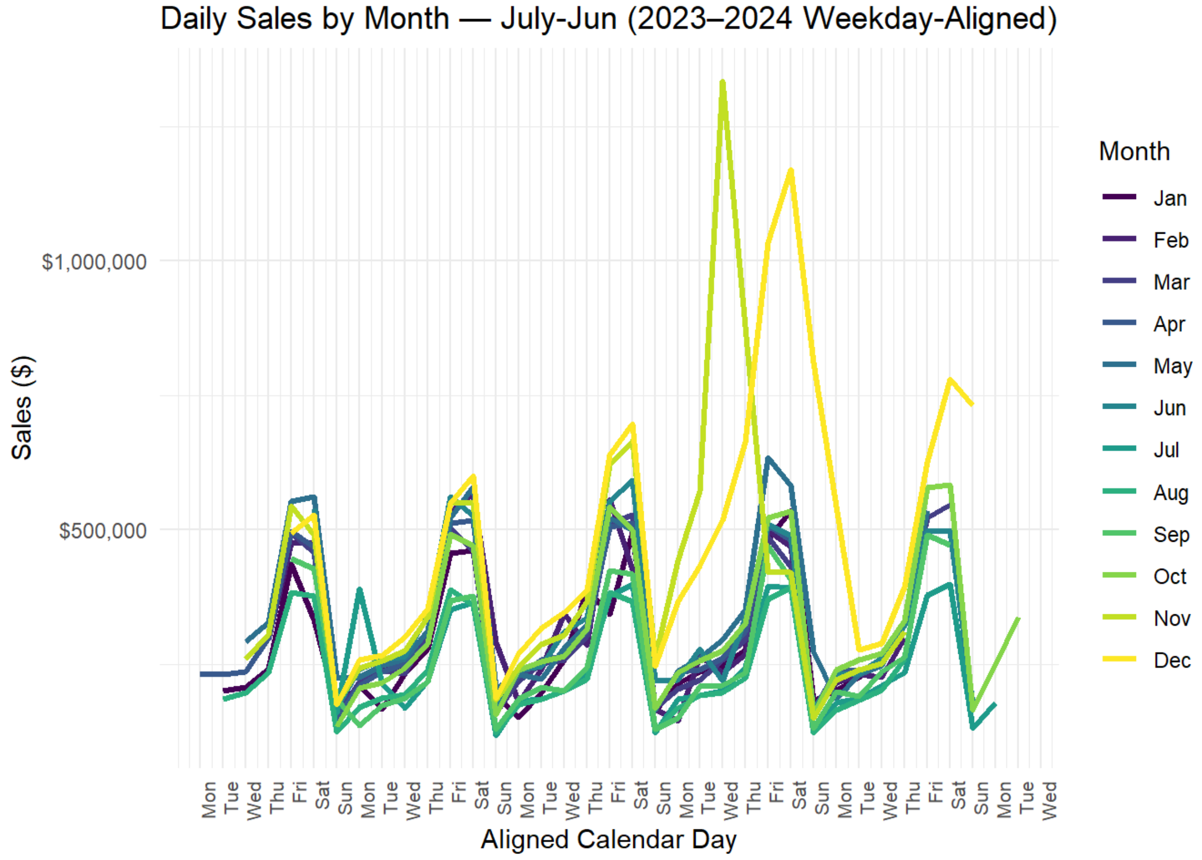
THANK YOU

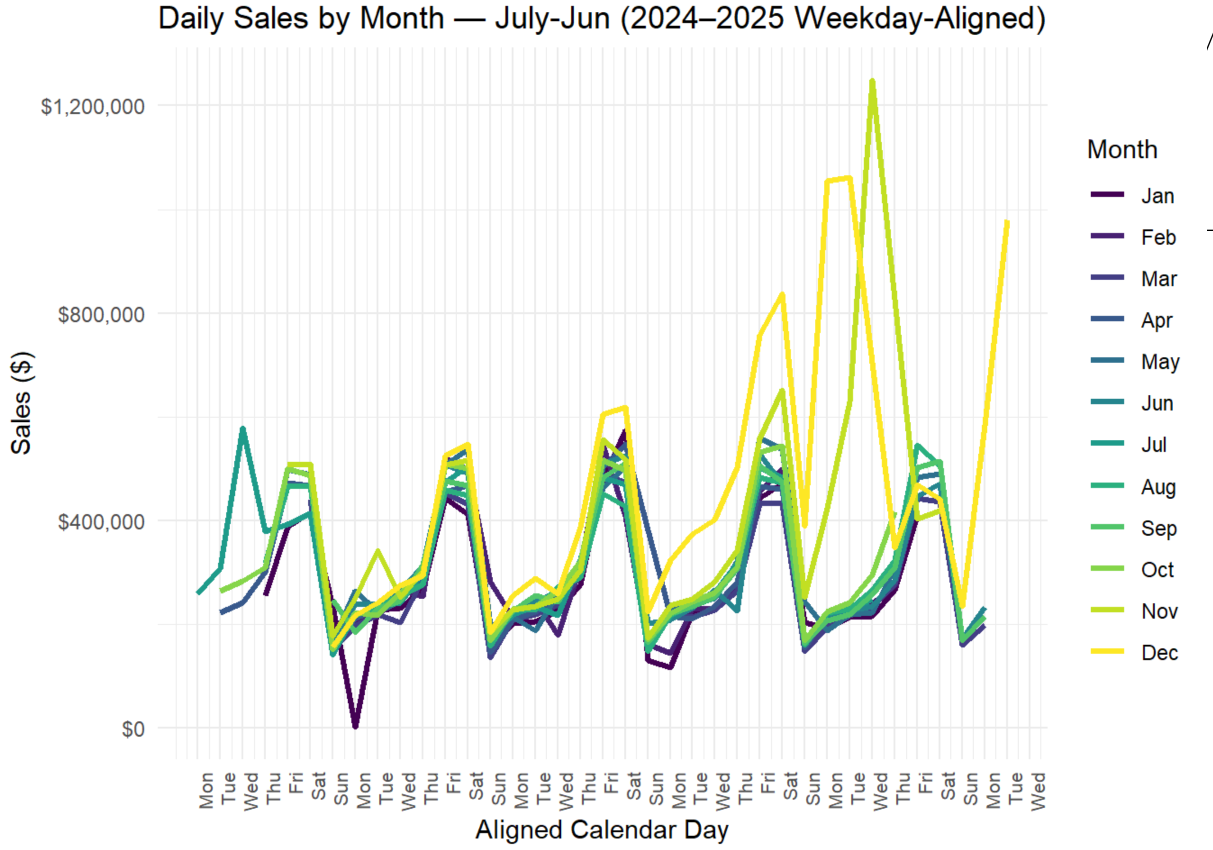
ABS MISSION & WHY THIS ANALYSIS MATTERS

- Dual mandate: public safety + profit
- Store variation influences county revenue
- Better decisions require accurate metrics

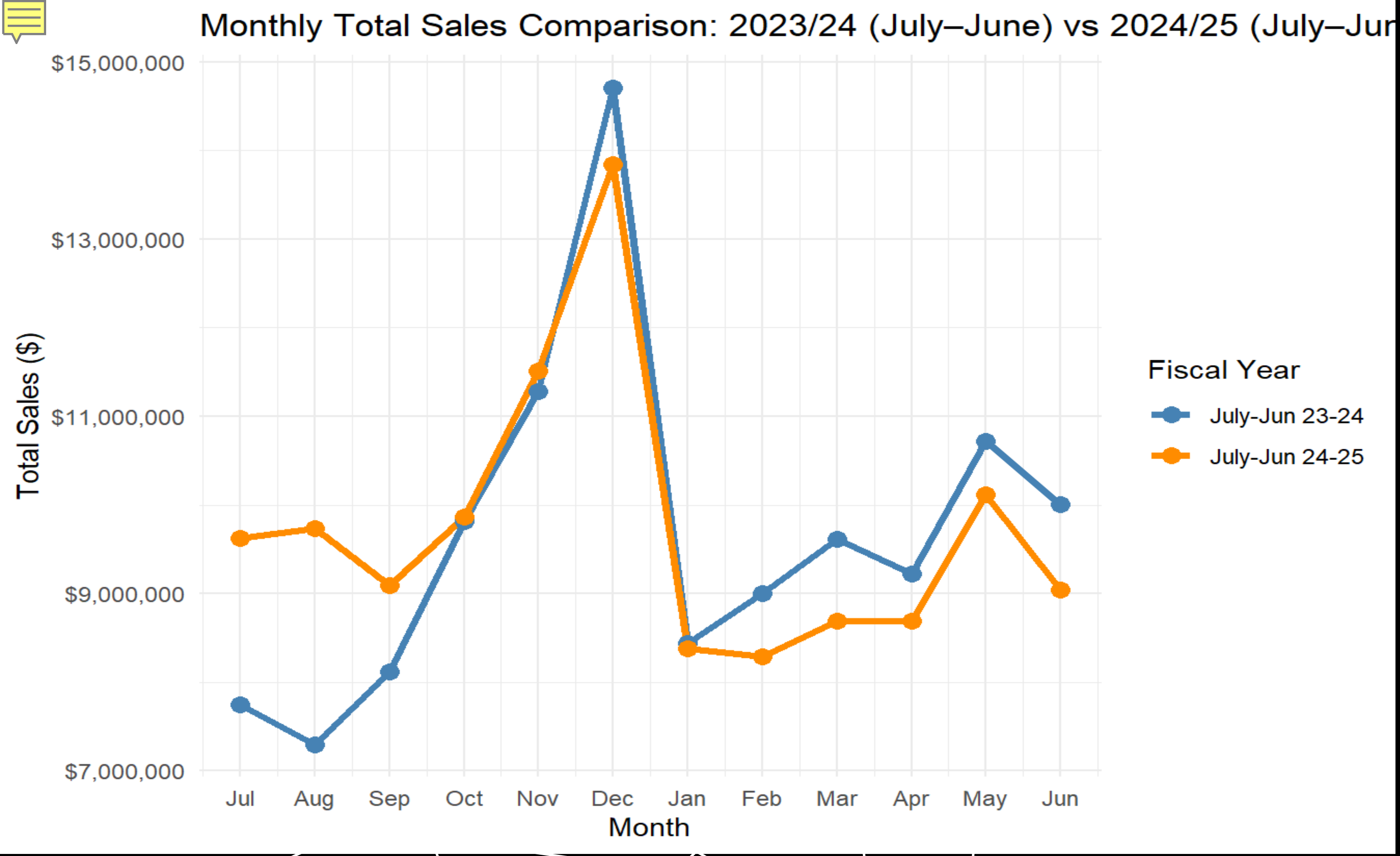


- Weekends generate the highest total daily sales
- Holidays also produce elevated revenue levels
- Weekdays remain steady but lower than peak periods

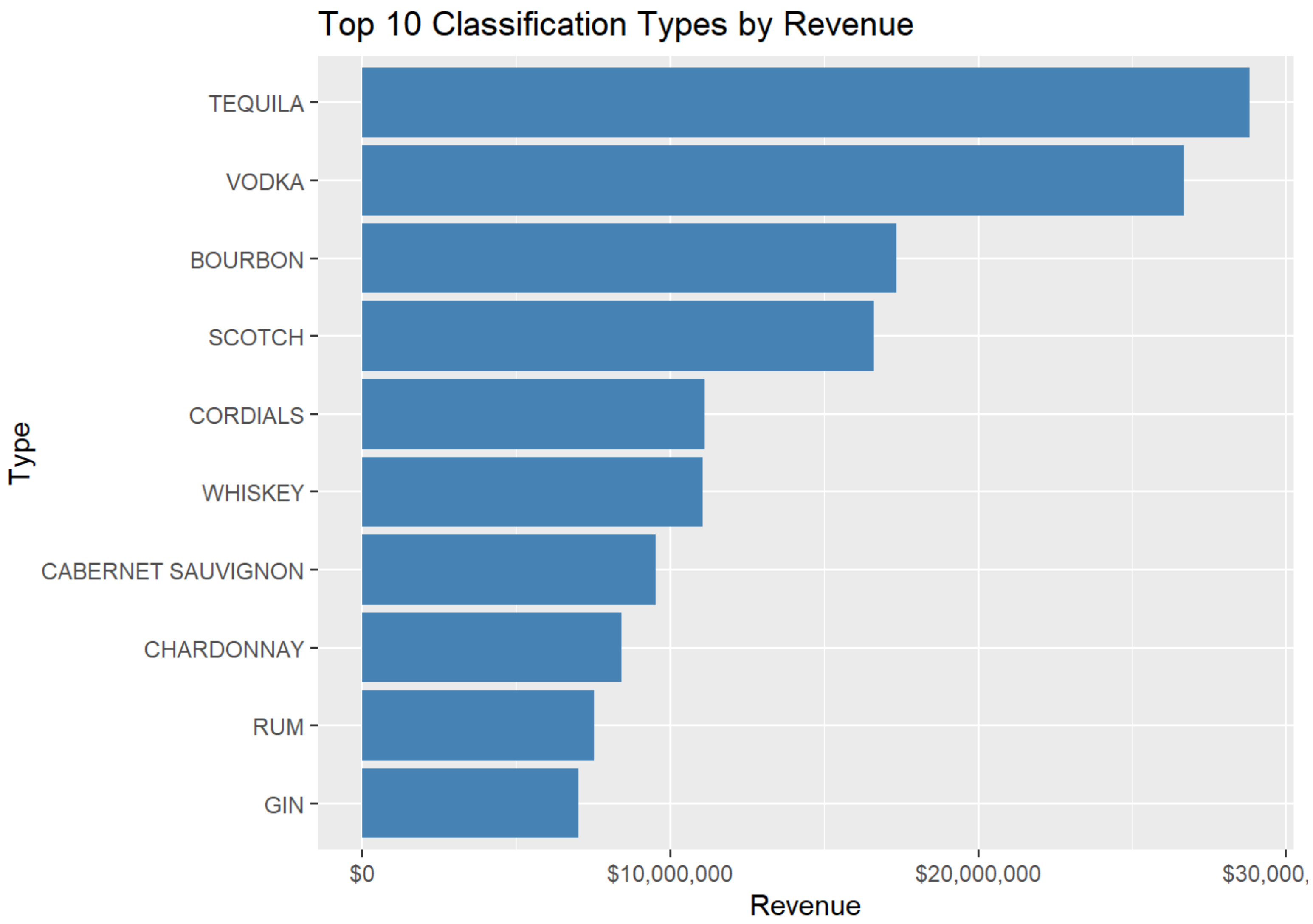




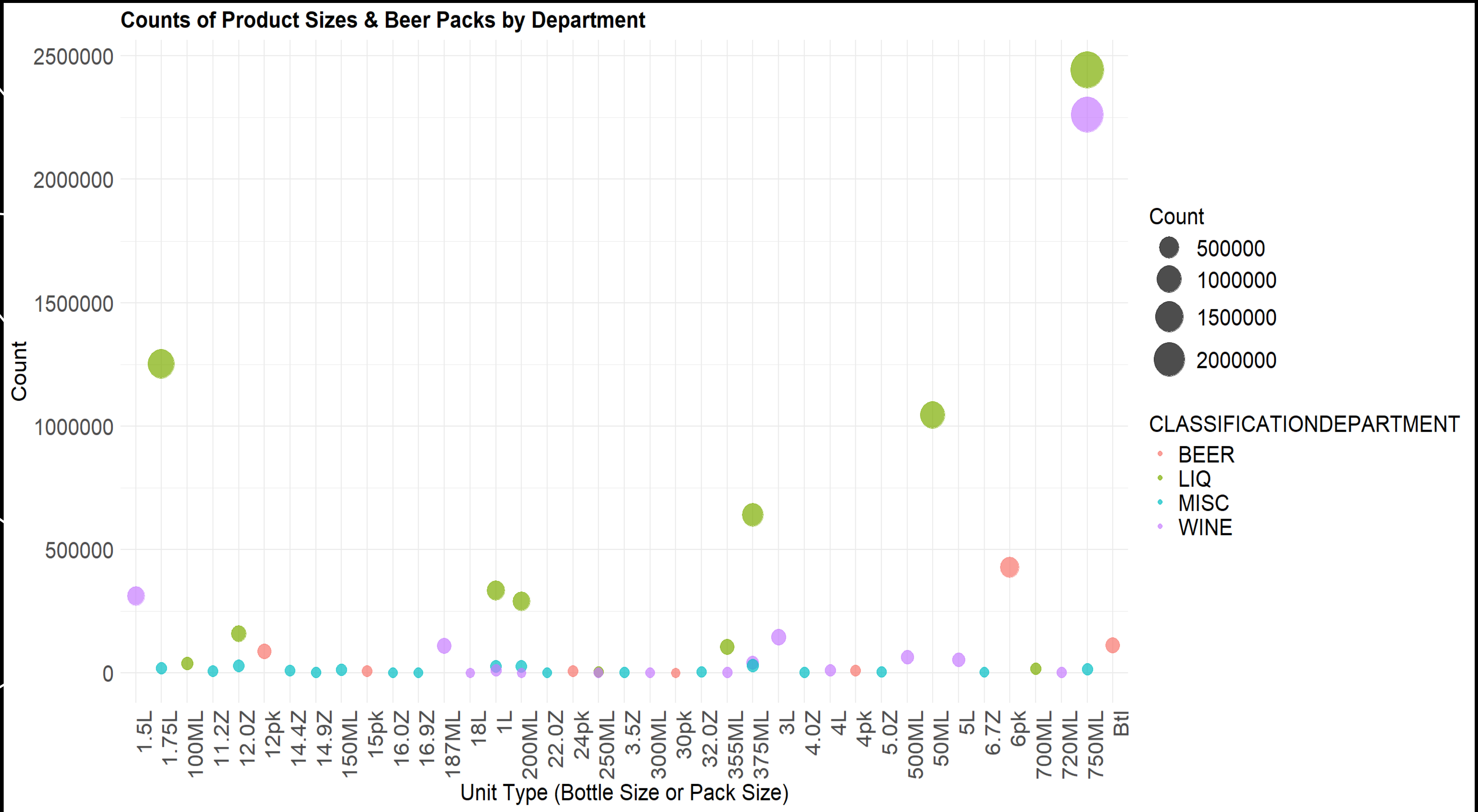
- Consistent Friday–Saturday peaks
- Stable weekly rhythm
- Clendar alignment removes date bias



DECEMBER IS CONSISTENTLY THE PEAK MONTH
JANUARY DROPS SHARPLY
SPRING RECOVERY FOLLOWS THE SAME PATTERN

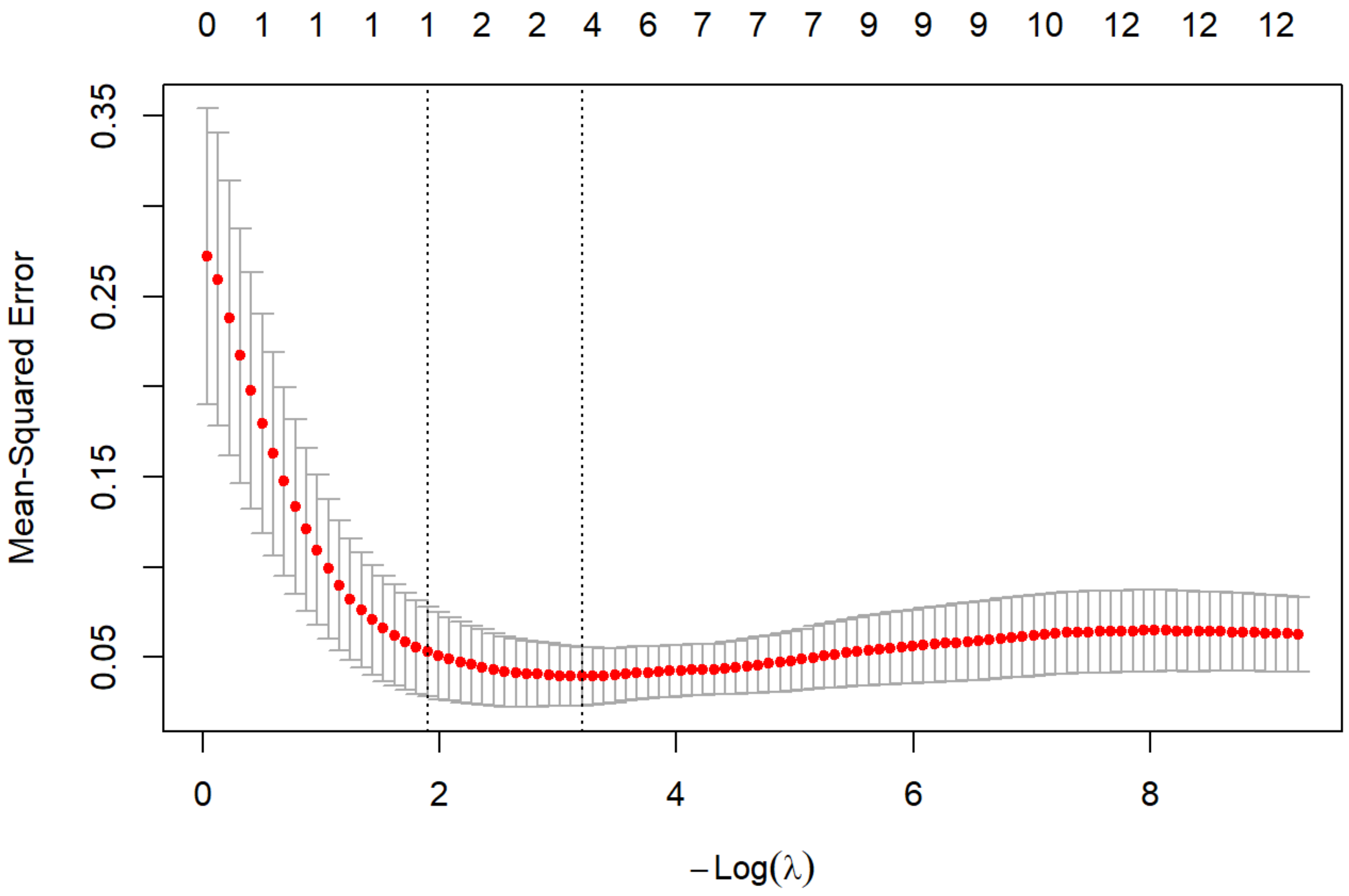


- Tequila, vodka, bourbon high impact
- Key categories for assortment and forecasting

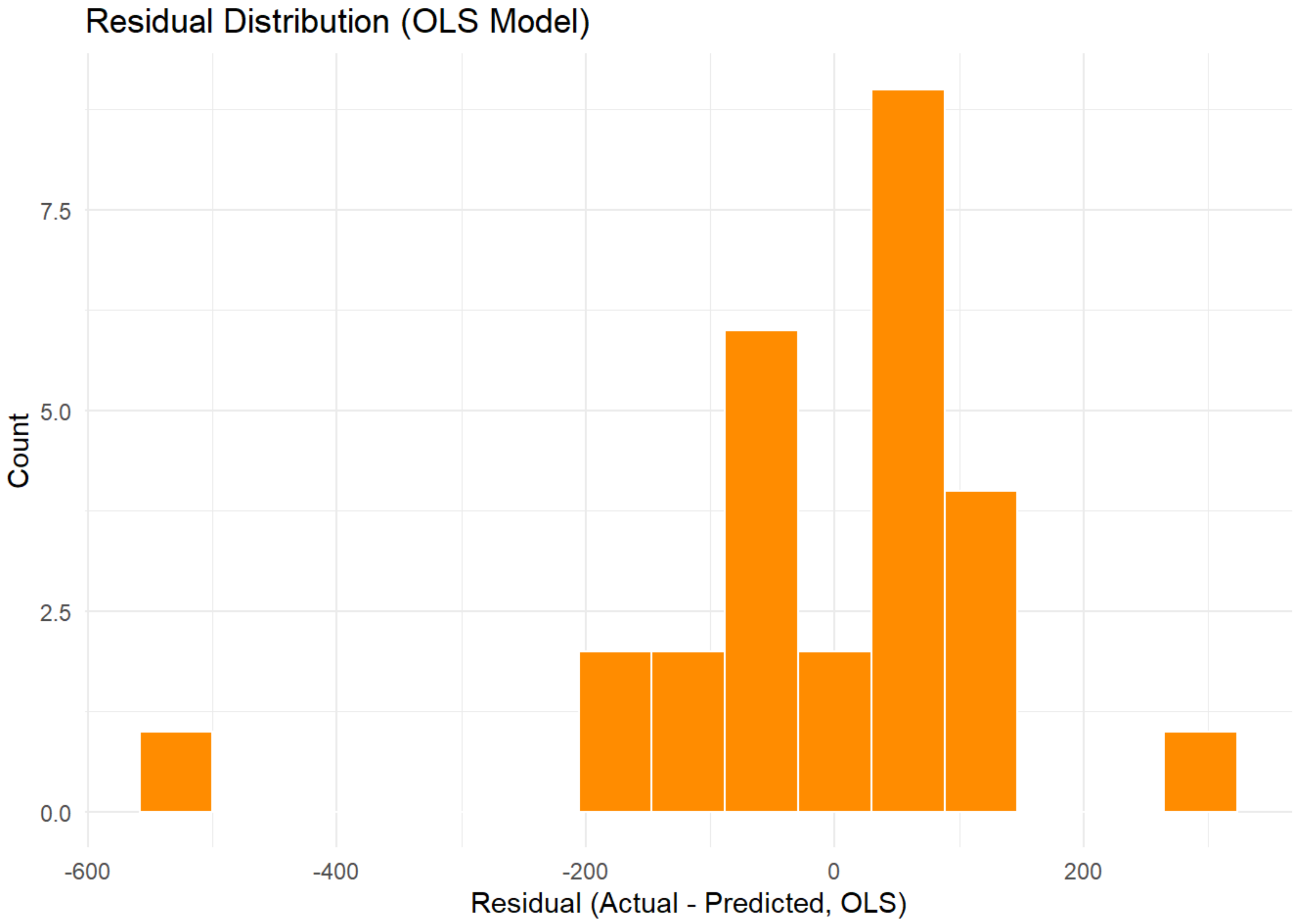


- Core sizes dominate
- Fringe formats low impact

ELASTIC NET CROSS-VALIDATION

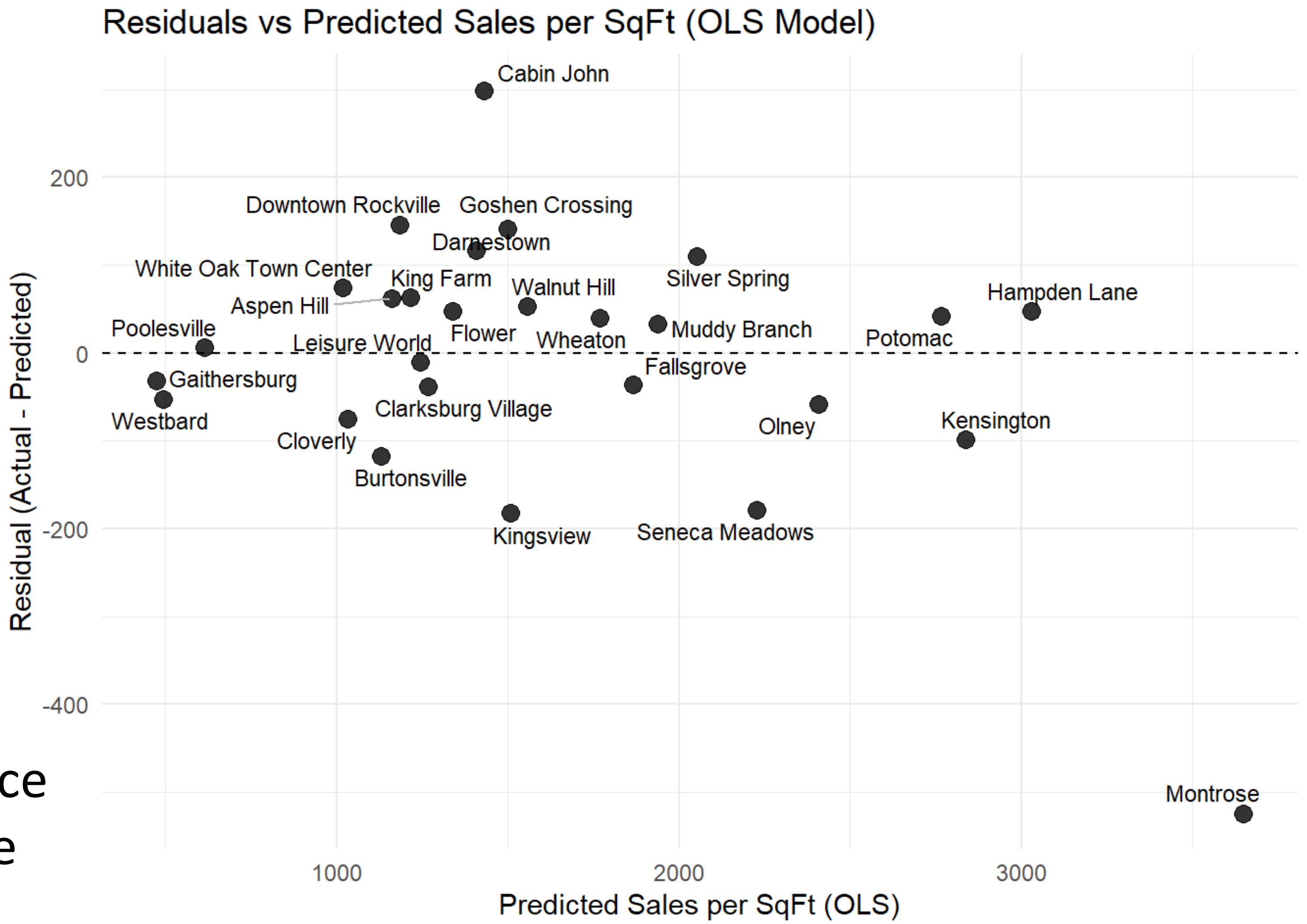


- Elastic Net used to test stability under correlated predictors
- Strong penalty quickly removes weaker variables
- Traffic intensity is the only predictor that consistently survives
- Basket metrics and category mix shrink to zero under regularization
- ENET confirms traffic's dominance but fits worse than OLS
- Best used for variable screening, not for final prediction



OLS fits on the log scale:

log (SalesPerSqFt) ~ sim SquareFootage + TotalPopulation + Older25Years + PovertyLevel + Avg_BasketValue + AvgItemsPerBasket + TotalTransactionsPerSqFt + Share_BEER + Share_LIQ + Share_WINE + Share_Weekday + Share_Weekend



- Below the line → underperformance
- Above the line → overperformance
- Underperformance appears at all predicted sales levels
- Reveals operational gaps hidden by revenue totals