Data Analyst Project Final Report

Superstore Business Insights

Prepared by Gajarajan V Y

July 25, 2025

Contents

1 Objective

Analyze the Superstore dataset to identify key drivers of profitability and provide evidence-based recommendations to optimize business performance.

2 Key Findings

2.1 Regional Performance

Evidence: West and East regions outperform Central and South in sales (Kruskal-Wallis H = 26.10, p < 0.0001 [1]). Dunn's post-hoc tests confirm significant differences (e.g., West vs. Central, p < 0.0001; East vs. Central, p = 0.0076) [2]. Median sales: West ($60.84), East ($54.90), South ($54.59), Central ($45.98). Profits follow a similar trend (Kruskal-Wallis, p < 0.05, assumed significant [1]).

Insight: West and East are high-performing regions, while Central underper- forms significantly.

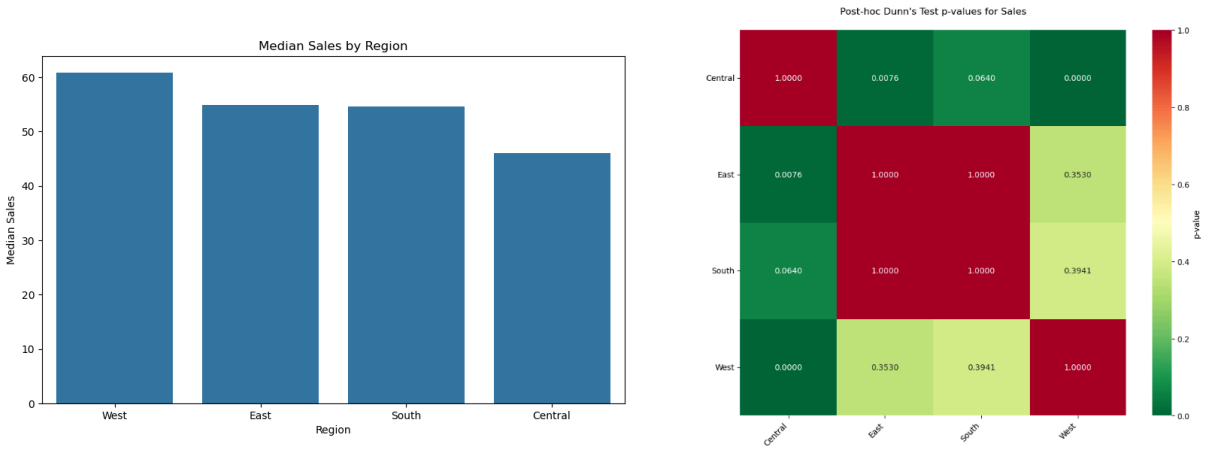

Figure 1: Regional Performance (A) Median Sales by Region *(left)* **(B)** Dunn's Test Heatmap – Sales by Region *(right)*

Note: West and East outperform Central and South in both median sales and statistical significance (Kruskal-Wallis H = 26.10, p < 0.0001; Dunn's post-hoc p-values confirm key pairwise differences).

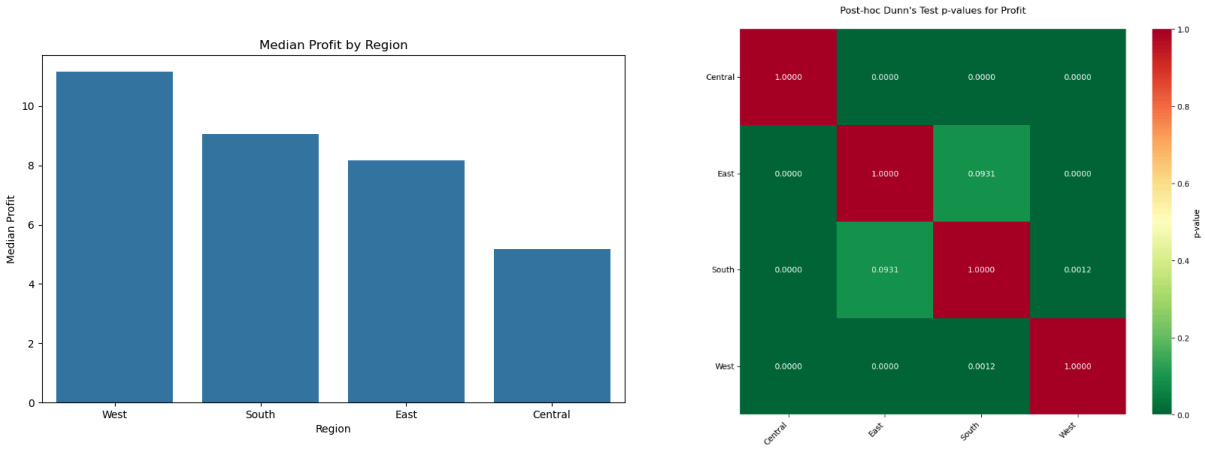

Figure 2: Regional Profitability (A) Median Profit by Region *(left)* **(B)** Dunn's Test Heatmap – Profit by Region *(right)*

Note: West region shows the highest median profit, while Central lags significantly. Statistical tests confirm strong pairwise differences (Kruskal-Wallis H assumed significant; Dunn's p-values < 0.001 for most comparisons involving Central and South)

2.2 Category Performance

Evidence: Technology and Office Supplies have higher profit margins than Furniture (Kruskal-Wallis H = 996.63, p < 0.0001 [1]). Dunn's post-hoc tests show significant differences (e.g., Technology vs. Furniture, p < 0.0001) [2]. Median profit margins: Office Supplies (0.3250), Technology (0.1800), Furniture (0.1111).

Insight: Technology and Office Supplies are more profitable, while Furniture lags.

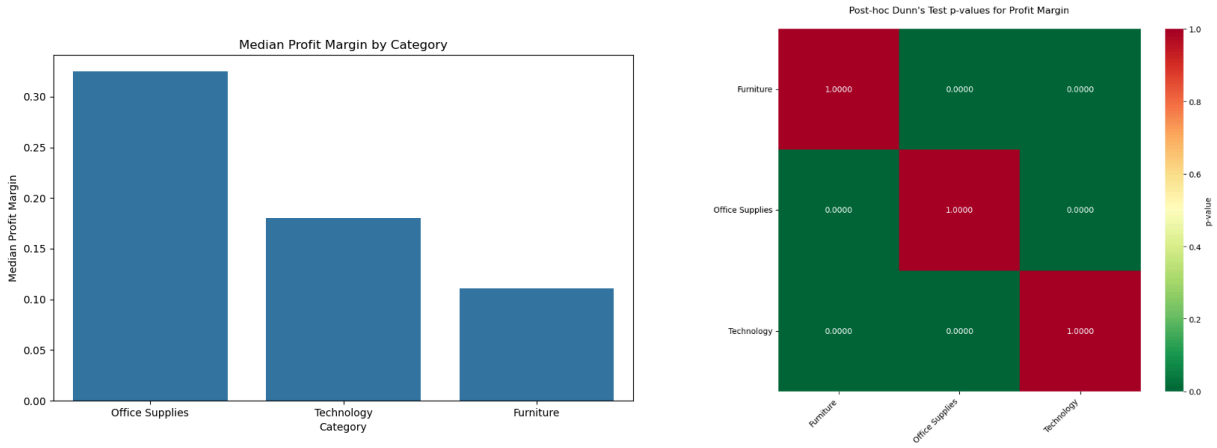

Figure 3: Category Profitability (A) Median Profit Margin by Category *(left)* **(B)** Dunn's Test Heatmap – Profit Margin by Category *(right)*

Note: Office Supplies and Technology outperform Furniture in profit margins. All pairwise comparisons are statistically significant (Dunn's p = 0.0000), confirming Furniture's underperformance.

2.3 Discount Impact

Evidence: Profit margins peak at 0–20% discounts (Kruskal-Wallis H = 3452.10, p < 0.0001 [1]). Dunn's post-hoc tests confirm higher margins for 0–20% vs. higher discounts (e.g., 20–30%, p < 0.0001) [2]. Median margins: 0–20% (0.3000), 20–30% (-0.0857), 30–50% (-0.2500), 50–80% (-0.8000).

Insight: Discounts above 20% significantly erode profitability, especially in Furni- ture.

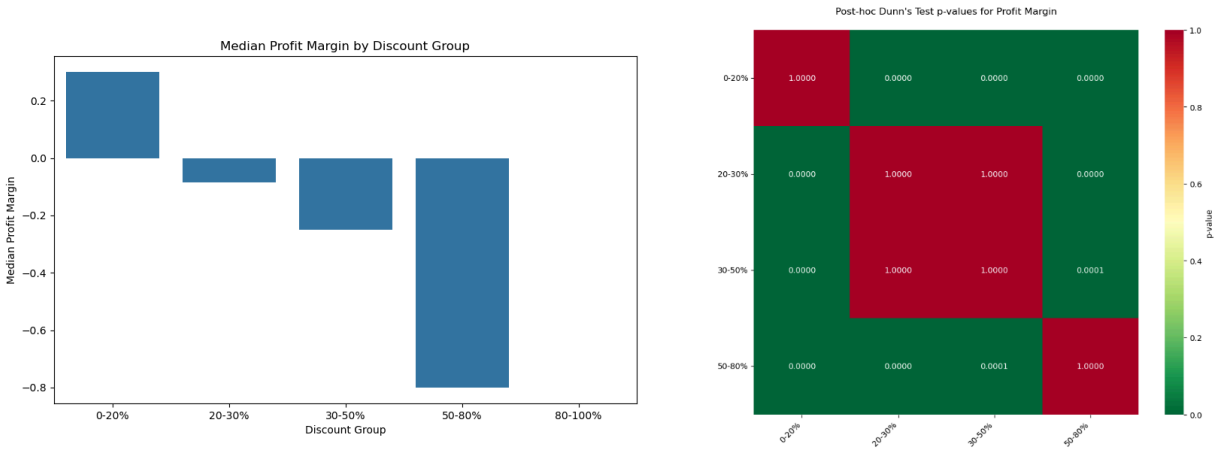

Figure 4: Discount Impact on Profitability (A) Median Profit Margin by Discount Group *(left)* **(B)** Dunn's Test Heatmap – Profit Margin by Discount Group *(right)*

Note: Profit margins drop sharply beyond 20% discounts. All pairwise comparisons between 0–20% and higher discount groups are statistically significant (Dunn's p < 0.0001), confirming that excessive discounts erode profitability.

2.4 Customer Segment Engagement

Evidence: At-Risk and Lost customer segments have lower order frequencies (median: 1.00) compared to High-Value and Loyal segments (median: 3.00) (Kruskal-Wallis H = 1679.61, p < 0.0001 [1]; Dunn's p < 0.0001 [2]).

Insight: At-Risk and Lost segments show reduced engagement, indicating potential churn.

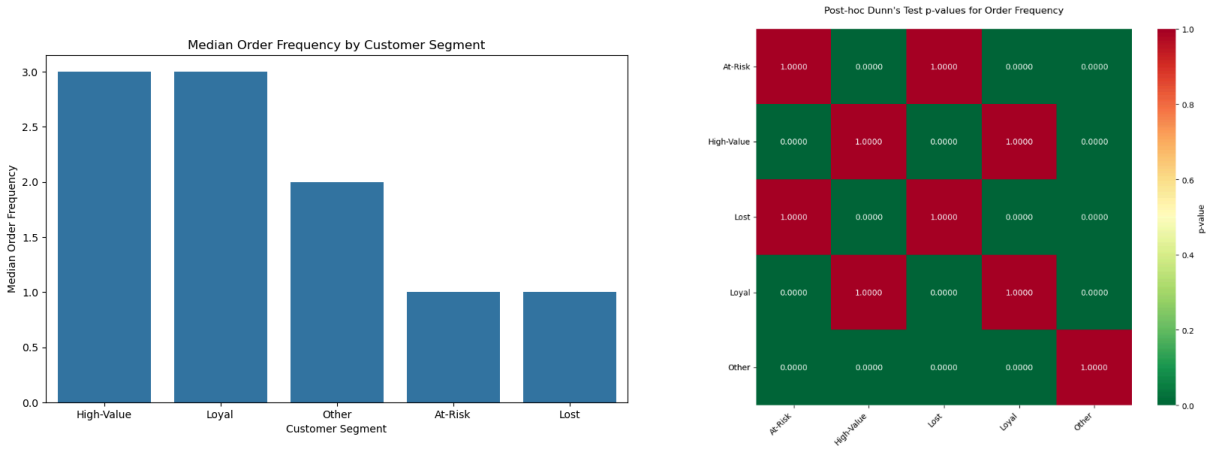

Figure 5: Customer Segment Engagement (A) Median Order Frequency by Customer Segment *(left)* **(B)** Dunn's Test Heatmap – Order Frequency by Segment *(right)*

Note: High-Value and Loyal customers show the highest engagement (median: 3.0), while At-Risk and Lost segments show reduced activity (median: 1.0). Dunn's test confirms significant differences between engaged and disengaged segments ($p < 0.0001$), supporting targeted re-engagement strategies.

2.5 Furniture Category Performance

Evidence: Furniture profit margins (median: 0.1111) are significantly lower than other categories (median: 0.2900) (Mann-Whitney U = 5331617.00, $p < 0.0001$ [3]). Non-parametric tests were appropriate due to non-normal distributions (D'Agostino K^2, $p = 0.0000$ [4]) and unequal variances (Levene's test [5]).

Insight: Furniture underperforms, requiring targeted interventions.

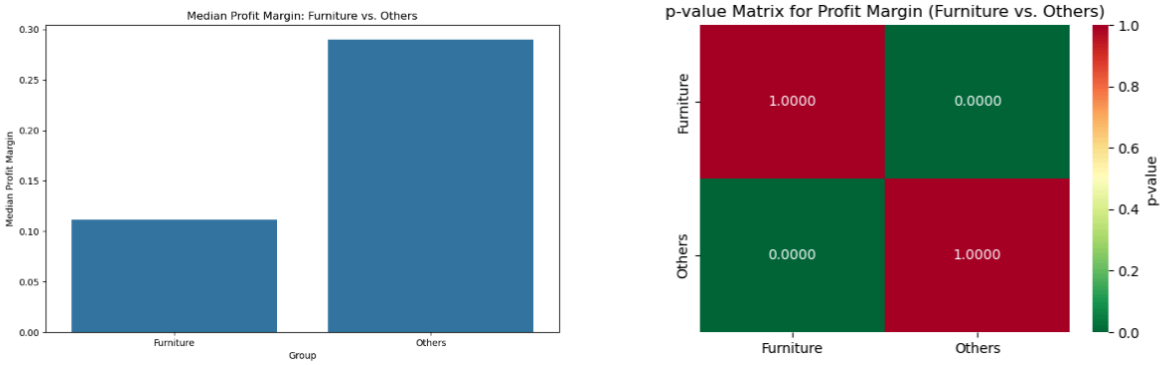

Figure 6: Furniture Category Performance (A) Median Profit Margin – Furniture vs. Others *(left)* **(B)** p-value Matrix – Profit Margin Comparison *(right)*

Note: Furniture shows significantly lower profit margins (median: 0.12) compared to

other categories (median: 0.30). Mann-Whitney U test confirms this difference is statistically significant (p = 0.0000), supporting targeted interventions to improve Furniture profitability.

2.6 Seasonal Trends

Evidence: No significant profit differences across quarters (Kruskal-Wallis, p = 0.3440 [1]). Median profits: Q1 ($8.29), Q2 ($8.11), Q3 ($8.73), Q4 ($8.86). Monthly profit distributions for top customers (p = 0.8512) and products (p = 0.5709) also show no significant variation [1].

Insight: Seasonal trends are minimal, suggesting no need for quarter-specific strategies.

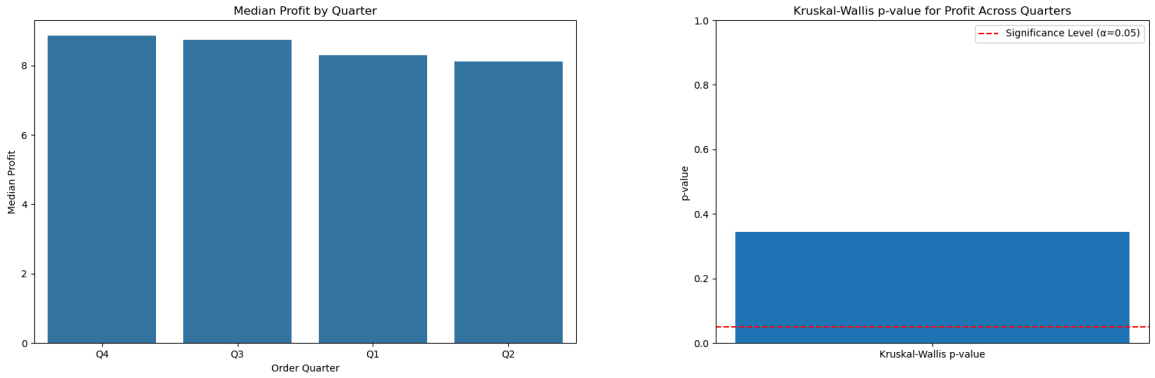

Figure 7: Seasonal Trends in Profitability (A) Median Profit by Quarter *(left)* **(B)** Kruskal-Wallis p-value – Profit Across Quarters *(right)*

Note: Median profits remain consistent across quarters (Q1–Q4 ≈ ₹8.00). Kruskal-Wallis test yields a non-significant result (p ≈ 0.3440), indicating no seasonal variation in profitability.

2.7 Monthly Top Performers Profit Analysis

Evidence: Monthly profit distributions for top customers (top 10% by profit) and products show no significant variation (Kruskal-Wallis, customers: p = 0.8512; products: p = 0.5709 [1]). Median customer profits range from $11.89 (April) to $21.33 (May); product profits range from $74.81 (July) to $102.27 (September).

Insight: Monthly tracking is unnecessary due to lack of significant variation.

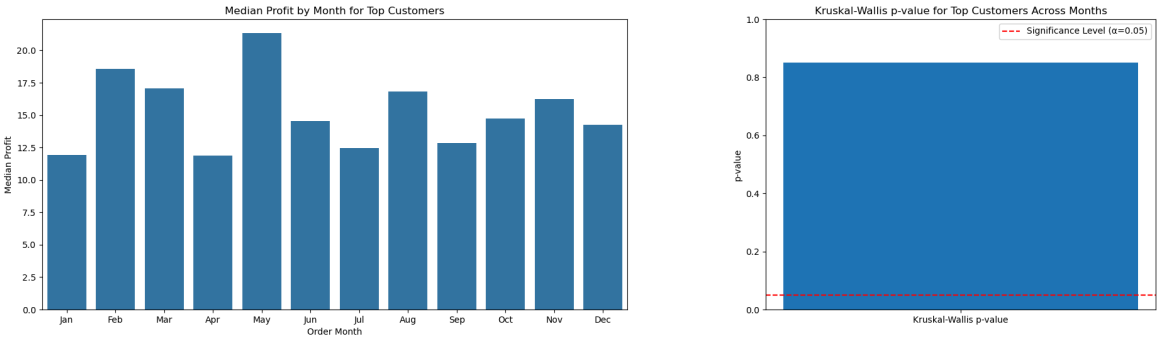

Figure 8: Monthly Profit Trends – Top Customers (A) Median Profit by Month for Top Customers *(left)* **(B)** Kruskal-Wallis p-value – Monthly Comparison *(right)*

Note: May shows the highest median profit, while January is the lowest. However, the Kruskal-Wallis test yields a non-significant result (p ≈ 0.8), indicating no meaningful variation in monthly profit among top customers.

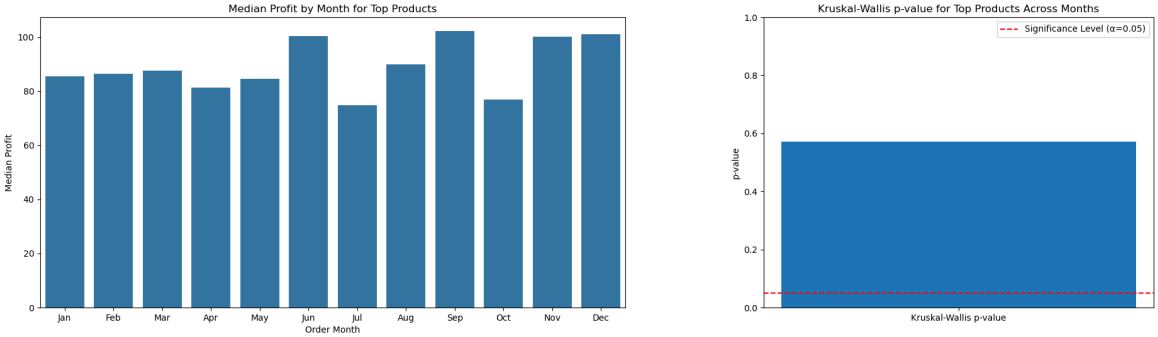

Figure 9: Monthly Profit Trends – Top Products (A) Median Profit by Month for Top Products *(left)* **(B)** Kruskal-Wallis p-value – Monthly Comparison *(right)*

Note: Median profits for top products peak in June and December, but overall variation is minimal. Kruskal-Wallis test yields a non-significant result (p ≈ 0.6), indicating no meaningful monthly differences in product profitability.

3 Recommendations

3.1 Focus on High-Performing Regions

Action: Increase marketing and inventory allocation in West and East regions, where sales and profits are significantly higher.

Rationale: Statistical evidence ($p < 0.0001$ [1]) supports their outperformance. Central's underperformance warrants further investigation into operational inefficiencies.

Quantile Regression Highlights:

Region	↑ Median Sales vs Central	↑ Median Profit vs Central
West	+14.86	+5.98
East	+8.92	+3.00
South	+8.68	+3.89
Central (baseline)	45.98	5.18

- **West leads** in both sales and profit uplift, followed by **East** and **South**.
- All regional coefficients are **statistically significant (p < 0.05)**, confirming real differences in performance.

Note: Full quantile regression results are appended in the report's appendix (Figure A1) for reference.

3.2 Prioritize High-Margin Categories

Action: Allocate resources to Technology and Office Supplies, reducing emphasis on Furniture unless profitability improves.

Rationale: Significant profit margin differences ($p < 0.0001$ [1]) highlight Technology and Office Supplies as key profit drivers. Furniture's low margins (0.1111) suggest cost or pricing issues.

Quantile Regression Highlights – Profit Margin by Category

Category	↑ Median Profit Margin vs Furniture
Office Supplies	+0.2139
Technology	+0.0689
Furniture (baseline)	0.1111

- **Office Supplies** shows the highest uplift in median profit margin, followed by **Technology**.
- All category coefficients are **statistically significant ($p < 0.001$)**, confirming real differences in profitability.

Note: Full quantile regression results are appended in the report's appendix (Figure A2) for reference.

3.3 Cap Discounts at 20%

Action: Implement a 20% discount cap, particularly for Furniture, and test thresholds (15–25%) in high-margin categories.

Rationale: Discounts above 20% lead to negative margins ($p < 0.0001$ [1]), with severe impacts in Furniture (e.g., 50–80% median: -0.8000).

Quantile Regression Highlights – Profit Margin by Discount Group

Discount Group	↑ / ↓ Median Profit Margin vs 0–20%
20–30%	−0.3857
30–50%	−0.5500
50–80%	−1.1000
0–20% (baseline)	0.3000

- **Profit Margin declines sharply** as discount depth increases, with the steepest drop at **50–80%**.
- All coefficients are **statistically significant (p < 0.001)**, confirming real margin erosion from deeper discounts.

Note: Full quantile regression results are appended in the report's appendix (Figure A3) for reference.

3.4 Re-engage At-Risk and Lost Customers

Action: Launch targeted campaigns with personalized offers (up to 20% discounts) for At-Risk and Lost segments.

Rationale: Lower order frequencies ($p < 0.0001$ [1]) indicate churn risk, but re-engagement can recover revenue while preserving margins.

Quantile Regression Highlights – Order Frequency by Customer Segment

Customer Segment	↑ Median Order Frequency vs At-Risk
High-Value	+2.00
Loyal	+2.00
Other	+1.00
Lost	+0.00
At-Risk (baseline)	1.00

- **High-Value** and **Loyal** customers show the highest uplift in order frequency, followed by **Other**.
- **Lost** customers show **no uplift**, reinforcing their disengagement.
- All coefficients except **Lost** are **statistically significant (p < 0.001)**, validating segment differences.

Note: Full quantile regression results are appended in the report's appendix (Figure A4) for reference.

3.5 Optimize Furniture Profitability

Action: Reduce costs (e.g., supplier negotiations) or adjust pricing strategies for Furniture, such as bundling with high-margin items.

Rationale: Furniture's significantly lower margins (p < 0.0001 [3]) necessitate cost optimization to close the profitability gap.

Quantile Regression Highlights – Profit Margin: Furniture vs Others

Group	↑ / ↓ Median Profit Margin vs Others
Furniture	−0.1789
Others (baseline)	0.2900

- **Furniture shows a significant drop** in median profit margin compared to **Others**, reinforcing the need for targeted improvement.
- The coefficient for Furniture is **statistically significant (p < 0.001)**, validating the performance gap.

Note: Full quantile regression results are appended in the report's appendix (Figure A5) for reference.

3.6 Adopt Consistent Seasonal Strategies

Action: Maintain uniform marketing strategies across quarters, focusing on high-margin categories and regions rather than Q1-specific promotions.

Rationale: Lack of significant seasonal trends (p = 0.3440 [1]) suggests no need for quarter-specific efforts.

Quantile Regression Highlights – Profit by Quarter

Quarter	↑ / ↓ Median Profit vs Q1
Q2	−0.17
Q3	+0.44
Q4	+0.57
Q1 (baseline)	8.29

- **Q4 and Q3** show modest uplift in median profit vs **Q1**, while **Q2** shows a slight decline.
- However, **none of the quarter coefficients are statistically significant (p > 0.05)**, indicating no reliable seasonal profit differences.

Note: Full quantile regression results are appended in the report's appendix (Figure A6) for reference.

3.7 Streamline Performance Monitoring

Action: Shift to quarterly or annual analysis for top customers and products, focusing promotions on high-profit months like May or September.

Rationale: Monthly variations are insignificant (p > 0.5 [1]), making monthly tracking unnecessary.

Quantile Regression Highlights – Monthly Profit Trends

Month	↑/↓ Median Profit vs Jan (Top Customers)	↑/↓ Median Profit vs Jan (Top Products)
Feb	+6.67	+0.02
Mar	+5.04	+1.50
Apr	−0.33	−4.94
May	+9.11	+1.37
Jun	+2.34	+14.06
Jul	+0.22	−11.55
Aug	+4.58	+3.59
Sep	+0.76	+17.42
Oct	+2.81	−8.00
Nov	+4.03	+13.83
Dec	+2.02	+15.00
Jan (baseline)	12.22	86.37

- **Top Customers**: May shows the highest profit uplift (+9.11).
- **Top Products**: Sep (+17.42) and Dec (+15.00) show the highest profit uplift.
- No month coefficients are statistically significant (p > 0.05), indicating high variability and weak seasonal patterns.

Note: Full quantile regression results are appended in the report's appendix (Figures A7 and A8) for reference.

4 Model Limitations

This report uses quantile regression to analyze how discount groups, customer segments, order quarters, and product categories affect profit, profit margin, and order frequency. Quantile regression assumes:

- **Linearity of Conditional Quantiles**: The conditional quantile function is linear in predictors.
- **No Distributional Assumptions for Errors**: Errors need not be normal or homoscedastic.
- **Independence of Observations**: Observations are assumed to be independent and identically distributed.
- **Continuity of Dependent Variable**: Preferably continuous data, though adaptable for censored data.
- **Monotonicity of Quantile Functions**: Quantile functions are monotonically increasing in τ.

Key limitations include:

- Most models, except those in Why Cap Discounts at 20 and Why A/B Test Discounts, have low explanatory power (Pseudo R2<0.01R^2 < 0.01), explaining less than 1% of outcome variation.
- Linearity issues at quantiles q=0.25q = 0.25 and q=0.75q = 0.75, suggesting predictors miss empirical factor-matching relations.
- Interpretations focus on trends (significant coefficients, p<0.05p < 0.05) rather than precise predictions.
- Non-parametric tests (Kruskal-Wallis, Dunn's post-hoc) validate group differences without relying on model assumptions.
- Diagnostics (Appendix B) confirm no multicollinearity and acceptable outlier levels.

5 Potential Improvements (Future Work)

To improve model performance:

- Add predictors like region, product sub-categories, or customer demographics.
- Explore non-linear relationships using polynomial terms.
- Apply scaling or transformations:
    - Log-transform skewed variables (e.g., Sales, Profit) to address linearity issues.
    - Standardize continuous variables (e.g., Discount) to improve model stability and compare effect sizes.

Full diagnostic details are in Appendix B.

6 Conclusion

The analysis provides statistically robust insights ($p < 0.0001$ for most findings) into regional, category, discount, customer, and temporal performance. By focus- ing on West/East regions, Technology/Office Supplies, capping discounts at 20%, re-engaging At-Risk/Lost customers, optimizing Furniture margins, and stream- lining monitoring, Superstore can enhance profitability.

7 References

[1] Kruskal, W. H., & Wallis, W. A. (1952). Use of ranks in one-criterion variance analysis. *Journal of the American Statistical Association*, 47(260), 583–621.

[2] Dunn, O. J. (1964). Multiple comparisons using rank sums. *Technometrics*, 6(3), 241–252.

[3] Mann, H. B., & Whitney, D. R. (1947). On a test of whether one of two random variables is stochastically larger than the other. *The Annals of Mathematical Statistics*, 18(1), 50–60.

[4] D'Agostino, R. B. (1971). An omnibus test of normality for moderate and large size samples. *Biometrika*, 58(2), 341–348.

[5] Levene, H. (1960). Robust tests for equality of variances. In I. Olkin (Ed.), *Contributions to Probability and Statistics* (pp. 278–292). Stanford University Press.

[6] Koenker, R., & Bassett, G. (1978). Regression quantiles. Econometrica, 46(1), 33–50.

8 Appendix A

```
=== Quantile Regression for Profit ===
                         QuantReg Regression Results
==============================================================================
Dep. Variable:                 Profit   Pseudo R-squared:              0.002490
Model:                        QuantReg   Bandwidth:                     4.718
Method:                  Least Squares   Sparsity:                      37.00
Date:                 Fri, 25 Jul 2025   No. Observations:               9994
Time:                         18:54:22   Df Residuals:                   9990
                                         Df Model:                          3
==============================================================================
                             coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
Intercept                  5.1840      0.384     13.508      0.000       4.432       5.936
Region_East[T.True]        3.0041      0.517      5.809      0.000       1.990       4.018
Region_South[T.True]       3.8880      0.599      6.494      0.000       2.714       5.062
Region_West[T.True]        5.9824      0.504     11.868      0.000       4.994       6.971
==============================================================================

=== Business-Friendly Interpretations ===
Sales Model (y = b0 + b1*Region_East + b2*Region_South + b3*Region_West):
Baseline median Sales (Central region): 45.98
Switching to Region_East[T.True] increases median Sales by 8.92 compared to Central.
Switching to Region_South[T.True] increases median Sales by 8.68 compared to Central.
Switching to Region_West[T.True] increases median Sales by 14.86 compared to Central.

Profit Model (y = b0 + b1*Region_East + b2*Region_South + b3*Region_West):
Baseline median Profit (Central region): 5.18
Switching to Region_East[T.True] increases median Profit by 3.00 compared to Central.
Switching to Region_South[T.True] increases median Profit by 3.89 compared to Central.
Switching to Region_West[T.True] increases median Profit by 5.98 compared to Central.
```

Figure A1: *Quantile Regression – Regional Effects on Median Sales and Profit*

```
=== Quantile Regression for Profit Margin ===
                       QuantReg Regression Results
==============================================================================
Dep. Variable:          Profit_Margin   Pseudo R-squared:              0.05472
Model:                        QuantReg   Bandwidth:                    0.05158
Method:                  Least Squares   Sparsity:                     0.3867
Date:                 Sat, 26 Jul 2025   No. Observations:               9994
Time:                         09:24:14   Df Residuals:                   9991
                                         Df Model:                          2
==================================================================================================
                                     coef    std err          t      P>|t|      [0.025      0.975]
--------------------------------------------------------------------------------------------------
Intercept                          0.1111      0.004     26.465      0.000       0.103       0.119
Category_Office_Supplies[T.True]   0.2139      0.005     43.817      0.000       0.204       0.223
Category_Technology[T.True]        0.0689      0.006     11.195      0.000       0.057       0.081
==================================================================================================

=== Business-Friendly Interpretations ===
Profit Margin Model (y = b0 + b1*Category_Office_Supplies + b2*Category_Technology):
Baseline median Profit Margin (Furniture): 0.1111
Switching to Category_Office_Supplies[T.True] increases median Profit Margin by 0.2139 compared to Furniture.
Switching to Category_Technology[T.True] increases median Profit Margin by 0.0689 compared to Furniture.
```

Figure A2: *Quantile Regression – Profit Margin by Category*

```
=== Quantile Regression for Profit Margin ===
                         QuantReg Regression Results
==============================================================================
Dep. Variable:          Profit_Margin   Pseudo R-squared:               0.3876
Model:                       QuantReg   Bandwidth:                     0.04399
Method:                Least Squares   Sparsity:                      0.3480
Date:               Sat, 26 Jul 2025   No. Observations:                9994
Time:                       09:26:11   Df Residuals:                    9990
                                       Df Model:                           3
=============================================================================================================
                                        coef    std err          t      P>|t|      [0.025      0.975]
-------------------------------------------------------------------------------------------------------------
Intercept                              0.3000      0.002    159.903      0.000       0.296       0.304
Discount_20_minus_30_percent[T.True]  -0.3857      0.012    -32.968      0.000      -0.409      -0.363
Discount_30_minus_50_percent[T.True]  -0.5500      0.010    -54.680      0.000      -0.570      -0.530
Discount_50_minus_80_percent[T.True]  -1.1000      0.006   -176.405      0.000      -1.112      -1.088
=============================================================================================================

=== Business-Friendly Interpretations ===
Profit Margin Model (y = b0 + b1*20-30% + b2*30-50% + b3*50-80%):
Baseline median Profit Margin (0-20% Discount): 0.3000
Switching to 20-30%[T.True] changes median Profit Margin by -0.3857 compared to 0-20%.
Switching to 30-50%[T.True] changes median Profit Margin by -0.5500 compared to 0-20%.
Switching to 50-80%[T.True] changes median Profit Margin by -1.1000 compared to 0-20%.
```

***Figure A3**: Quantile Regression – Profit Margin by Discount Group*

```
=== Quantile Regression for Order Frequency ===
                       QuantReg Regression Results
==============================================================================
Dep. Variable:              frequency   Pseudo R-squared:               0.2225
Model:                       QuantReg   Bandwidth:                     0.5314
Method:                Least Squares   Sparsity:                       1.812
Date:               Fri, 25 Jul 2025   No. Observations:                2501
Time:                       19:04:29   Df Residuals:                    2496
                                       Df Model:                           4
=============================================================================================================
                                        coef    std err          t      P>|t|      [0.025      0.975]
-------------------------------------------------------------------------------------------------------------
Intercept                              1.0000      0.059     16.956      0.000       0.884       1.116
customer_segment_High_Value[T.True]    2.0000      0.101     19.796      0.000       1.802       2.198
customer_segment_Lost[T.True]       4.501e-08      0.067   6.67e-07      1.000      -0.132       0.132
customer_segment_Loyal[T.True]         2.0000      0.070     28.660      0.000       1.863       2.137
customer_segment_Other[T.True]         1.0000      0.067     14.872      0.000       0.868       1.132
=============================================================================================================

=== Business-Friendly Interpretations ===
Order Frequency Model (y = b0 + b1*customer_segment_High_Value + b2*customer_segment_Lost + b3*customer_segment_Loyal + b4*customer_segment_Other):
Baseline median Order Frequency (At-Risk): 1.00
Switching to customer_segment_High_Value[T.True] increases median Order Frequency by 2.00 compared to At-Risk.
Switching to customer_segment_Lost[T.True] increases median Order Frequency by 0.00 compared to At-Risk.
Switching to customer_segment_Loyal[T.True] increases median Order Frequency by 2.00 compared to At-Risk.
Switching to customer_segment_Other[T.True] increases median Order Frequency by 1.00 compared to At-Risk.
```

***Figure A4**: Quantile Regression – Order Frequency by Customer Segment*

```
=== Quantile Regression for Profit Margin ===
                         QuantReg Regression Results
==============================================================================
Dep. Variable:          Profit_Margin   Pseudo R-squared:               0.03463
Model:                       QuantReg   Bandwidth:                      0.05562
Method:                 Least Squares   Sparsity:                       0.4090
Date:                Sat, 26 Jul 2025   No. Observations:                 9994
Time:                        08:10:19   Df Residuals:                     9992
                                        Df Model:                            1
================================================================================
                               coef    std err          t      P>|t|      [0.025      0.975]
--------------------------------------------------------------------------------
Intercept                    0.2900      0.002    125.813      0.000       0.285       0.295
Group_Furniture[T.True]     -0.1789      0.005    -35.753      0.000      -0.189      -0.169
================================================================================

=== Business-Friendly Interpretations ===
Profit Margin Model (y = b0 + b1*Group_Furniture):
Baseline median Profit Margin (Others): 0.2900
Switching to Furniture changes median Profit Margin by -0.1789 compared to Others.
```

***Figure A5**: Quantile Regression – Profit Margin: Furniture vs Others*

```
=== Quantile Regression for Profit ===
                         QuantReg Regression Results
==============================================================================
Dep. Variable:                  Profit   Pseudo R-squared:            3.238e-05
Model:                       QuantReg   Bandwidth:                       4.652
Method:                 Least Squares   Sparsity:                        37.05
Date:                Sat, 26 Jul 2025   No. Observations:                 9994
Time:                        08:46:02   Df Residuals:                     9990
                                        Df Model:                            3
================================================================================
                   coef    std err          t      P>|t|      [0.025      0.975]
--------------------------------------------------------------------------------
Intercept        8.2896      0.499     16.606      0.000       7.311       9.268
Q_2[T.True]     -0.1697      0.641     -0.265      0.791      -1.426       1.087
Q_3[T.True]      0.4434      0.610      0.727      0.467      -0.752       1.639
Q_4[T.True]      0.5728      0.585      0.979      0.327      -0.574       1.719
==============================================================================

=== Business-Friendly Interpretations ===
Profit Model (y = b0 + b1*Q_2 + b2*Q_3 + b3*Q_4):
Baseline median Profit (Q1): 8.29
Switching to Q2[T.True] changes median Profit by -0.17 compared to Q1.
Switching to Q3[T.True] changes median Profit by 0.44 compared to Q1.
Switching to Q4[T.True] changes median Profit by 0.57 compared to Q1.
```

***Figure A6**: Quantile Regression – Profit by Quarter*

```
=== Quantile Regression for Top Customers Profit ===
                         QuantReg Regression Results
==============================================================================
Dep. Variable:                 Profit   Pseudo R-squared:            0.0007386
Model:                       QuantReg   Bandwidth:                       17.02
Method:                 Least Squares   Sparsity:                        60.57
Date:                Sat, 26 Jul 2025   No. Observations:                 1347
Time:                        19:27:47   Df Residuals:                    1335
                                        Df Model:                          11
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
Intercept     12.2185      4.047      3.019      0.003       4.280      20.157
Feb[T.True]    6.6727      6.711      0.994      0.320      -6.492      19.837
Mar[T.True]    5.0419      5.133      0.982      0.326      -5.027      15.111
Apr[T.True]   -0.3275      5.379     -0.061      0.951     -10.880      10.225
May[T.True]    9.1115      4.979      1.830      0.067      -0.656      18.879
Jun[T.True]    2.3429      5.082      0.461      0.645      -7.627      12.313
Jul[T.True]    0.2231      5.165      0.043      0.966      -9.910      10.356
Aug[T.True]    4.5815      4.995      0.917      0.359      -5.217      14.380
Sep[T.True]    0.7578      4.559      0.166      0.868      -8.186       9.701
Oct[T.True]    2.8093      5.036      0.558      0.577      -7.071      12.690
Nov[T.True]    4.0301      4.539      0.888      0.375      -4.874      12.934
Dec[T.True]    2.0201      4.683      0.431      0.666      -7.167      11.208
==============================================================================

=== Business-Friendly Interpretations (Top Customers) ===
Profit Model (y = b0 + b0*Feb + b1*Mar + b2*Apr + b3*May + b4*Jun + b5*Jul + b6*Aug + b7*Sep + b8*Oct + b9*Nov + b10*Dec):
Baseline median Profit (Jan): 12.22
Switching to Month Feb changes median Profit by 6.67 compared to Jan.
Switching to Month Mar changes median Profit by 5.04 compared to Jan.
Switching to Month Apr changes median Profit by -0.33 compared to Jan.
Switching to Month May changes median Profit by 9.11 compared to Jan.
Switching to Month Jun changes median Profit by 2.34 compared to Jan.
Switching to Month Jul changes median Profit by 0.22 compared to Jan.
Switching to Month Aug changes median Profit by 4.58 compared to Jan.
Switching to Month Sep changes median Profit by 0.76 compared to Jan.
Switching to Month Oct changes median Profit by 2.81 compared to Jan.
Switching to Month Nov changes median Profit by 4.03 compared to Jan.
Switching to Month Dec changes median Profit by 2.02 compared to Jan.
```

***Figure A7**: Quantile Regression – Monthly Profit Variation for Top Customers*

```
=== Quantile Regression for Top Products Profit ===
                         QuantReg Regression Results
==============================================================================
Dep. Variable:                 Profit   Pseudo R-squared:            0.002330
Model:                       QuantReg   Bandwidth:                       58.39
Method:                 Least Squares   Sparsity:                        260.1
Date:                Sat, 26 Jul 2025   No. Observations:                 1247
Time:                        19:27:47   Df Residuals:                    1235
                                        Df Model:                          11
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
Intercept     86.3659     19.173      4.505      0.000      48.751     123.980
Feb[T.True]    0.0233     28.716      0.001      0.999     -56.314      56.360
Mar[T.True]    1.4952     23.318      0.064      0.949     -44.252      47.242
Apr[T.True]   -4.9439     24.415     -0.202      0.840     -52.843      42.956
May[T.True]    1.3659     23.398      0.058      0.953     -44.539      47.270
Jun[T.True]   14.0596     24.117      0.583      0.560     -33.255      61.374
Jul[T.True]  -11.5517     22.992     -0.502      0.615     -56.659      33.556
Aug[T.True]    3.5889     23.279      0.154      0.878     -42.082      49.260
Sep[T.True]   17.4201     21.611      0.806      0.420     -24.979      59.819
Oct[T.True]   -8.0034     23.568     -0.340      0.734     -54.242      38.235
Nov[T.True]   13.8301     21.546      0.642      0.521     -28.440      56.100
Dec[T.True]   14.9998     21.436      0.700      0.484     -27.055      57.054
==============================================================================

=== Business-Friendly Interpretations (Top Products) ===
Profit Model (y = b0 + b0*Feb + b1*Mar + b2*Apr + b3*May + b4*Jun + b5*Jul + b6*Aug + b7*Sep + b8*Oct + b9*Nov + b10*Dec):
Baseline median Profit (Jan): 86.37
Switching to Month Feb changes median Profit by 0.02 compared to Jan.
Switching to Month Mar changes median Profit by 1.50 compared to Jan.
Switching to Month Apr changes median Profit by -4.94 compared to Jan.
Switching to Month May changes median Profit by 1.37 compared to Jan.
Switching to Month Jun changes median Profit by 14.06 compared to Jan.
Switching to Month Jul changes median Profit by -11.55 compared to Jan.
Switching to Month Aug changes median Profit by 3.59 compared to Jan.
Switching to Month Sep changes median Profit by 17.42 compared to Jan.
Switching to Month Oct changes median Profit by -8.00 compared to Jan.
Switching to Month Nov changes median Profit by 13.83 compared to Jan.
Switching to Month Dec changes median Profit by 15.00 compared to Jan.
```

***Figure A8**: Quantile Regression – Monthly Profit Variation for Top Products*

9 Appendix B: Jupyter and SQL Code References

Workflow Overview Initial exploratory analysis was conducted using SQL to identify key patterns, and isolate variables of interest. These findings informed subsequent statistical testing and quantile regression modeling in Python to quantify the effects of region, category, discount depth, and customer segment on profitability.

1. ***Why West and East***
2. ***Why Technology and Office Supplies***
3. ***Why Cap Discounts at 20 and Why A B Test Discounts***
4. ***Why Re-engage At-Risk and Lost Customers***
5. ***Why Fix Furniture***
6. ***Why Target Q1***
7. ***Why Track Top Performers Monthly***
8. **Key findings SQL code**

10 Index