



The Logistics Puzzle: Optimizing Operations for Delightful Deliveries

Let's Spice Things Up with Flavor Matic!

Nandini Ramakrishnan

Pious Khemka

Vishesh Jaiprakash Shukla

Shashank Gorti Srisai Nagaupendra





About Flavor Matic

- Introducing Flavor Matic, the household wizard famous for its kitchen magic and innovative appliances!
- From sizzling grills to aromatic coffee makers, Flavor Matic brings joy to households around the globe.

Whisking Away Challenges

- Despite our mastery in household appliances, Flavor Matic faces logistical hurdles that spice up our delivery journey.
- Let's uncover these challenges and sprinkle some magic to transform them into opportunities.

Stirring up success with two key objectives



Navigating Transportation Mode Impact

- Understand Delivery Efficiency
- Enhance Customer Satisfaction
- Optimize Resource Utilization

Savoring Shipping Discounts

- Drive Sales Performance
- Improve Customer Satisfaction
- Manage Shipping Expenses



Unveiling the Data Delights!

Observation : 10,999
Variables : 12

Numerical Variables : 7
Categorical Variables : 4
Binary Variable : 1

No Missing Values!

Source: Kaggle



Exploring the Spice Routes!

EDA

- Univariate Analysis
- Bivariate Analysis
- Multivariate Analysis
- Objective Specific EDA

Predictive modeling

- Logistic Regression
- Decision Tree
- Random Forest
- Gradient Boosting
- XG Boost

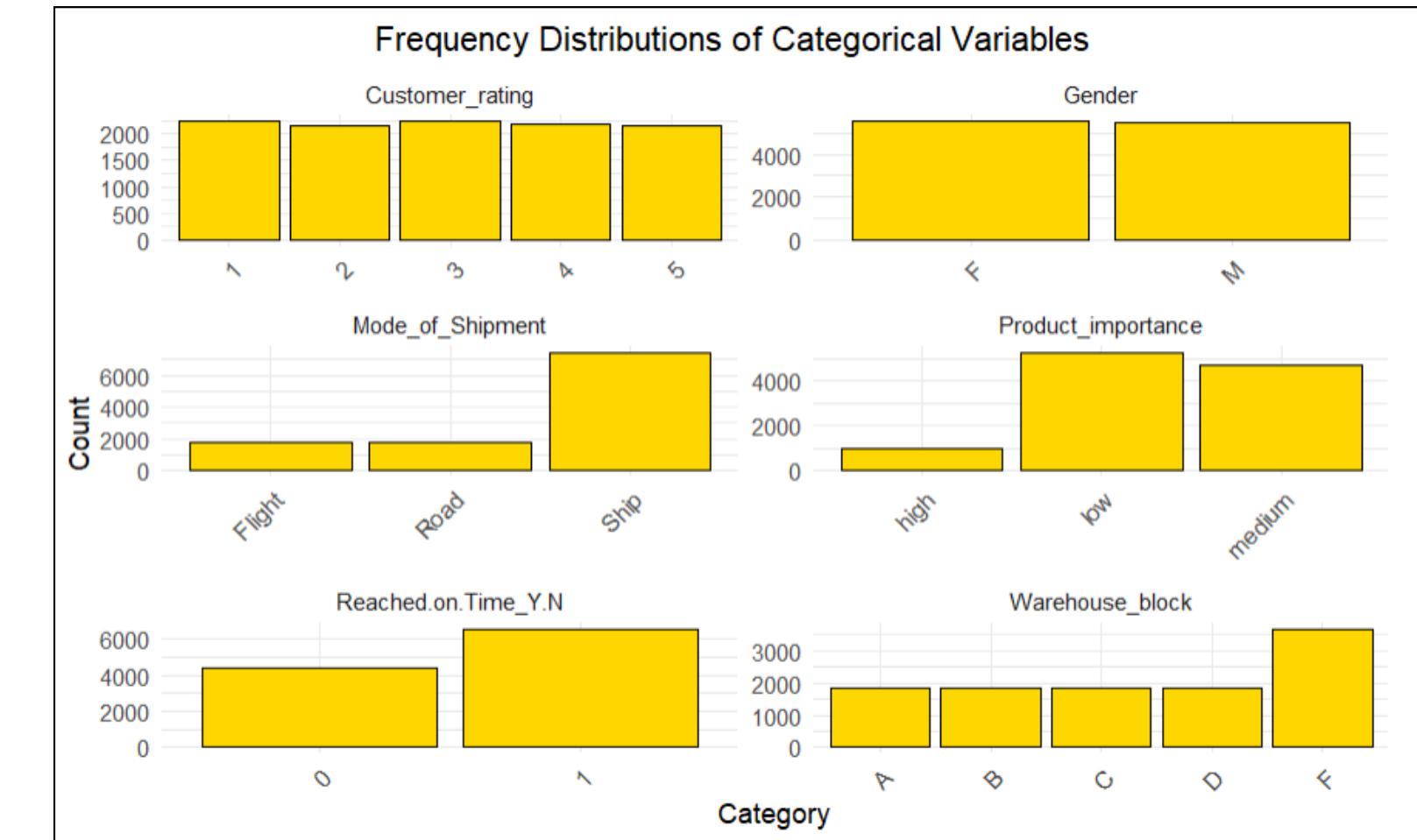
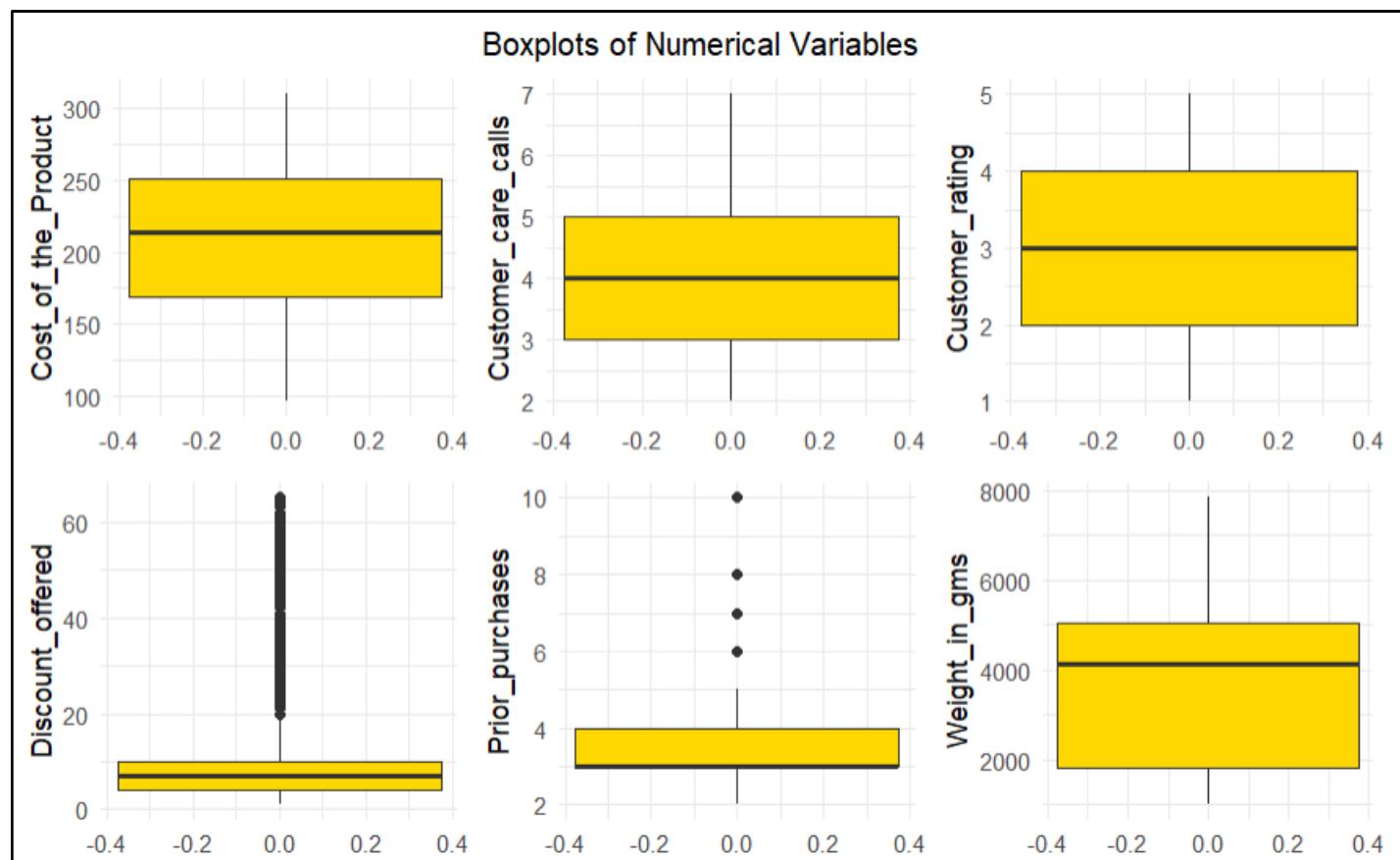
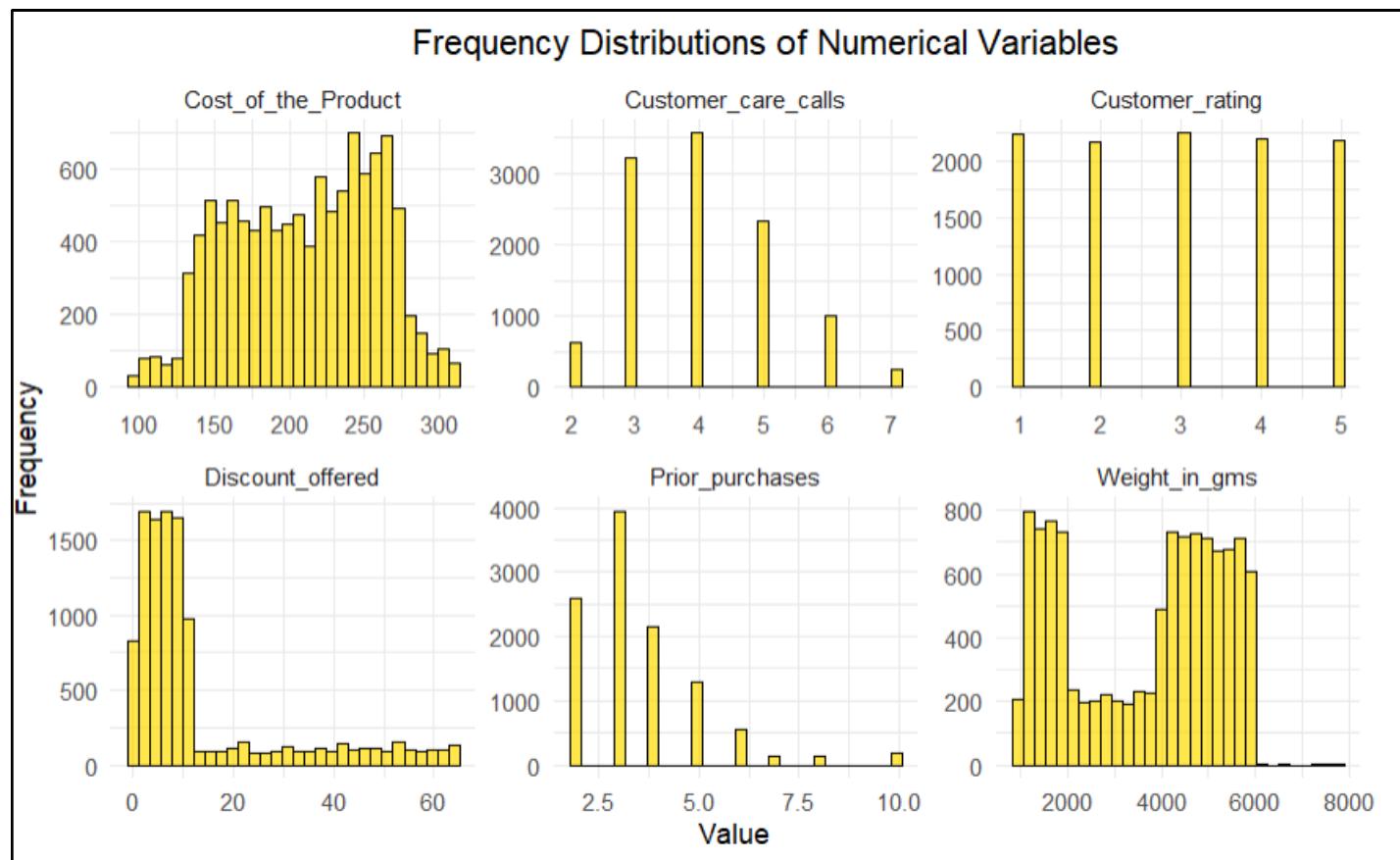
PCA and Clustering

- PCA
- KNN

EDA

Univariate Analysis of Variables

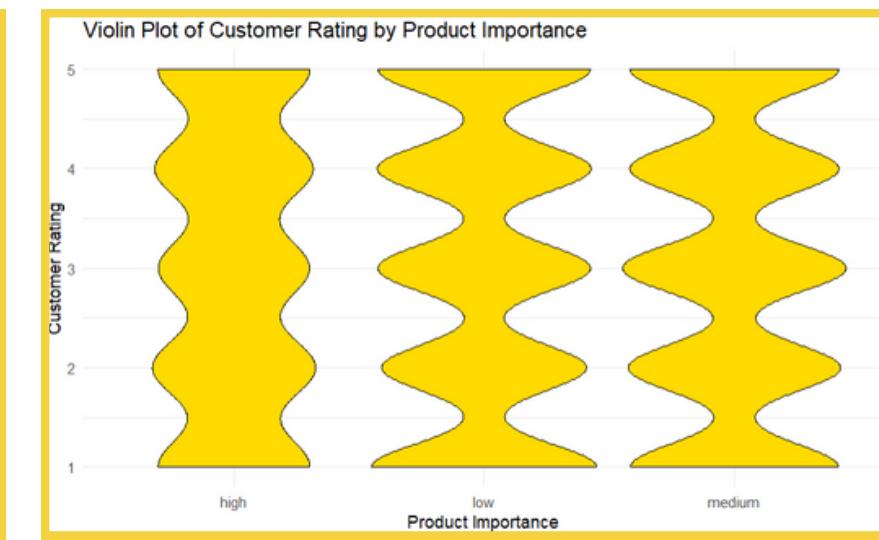
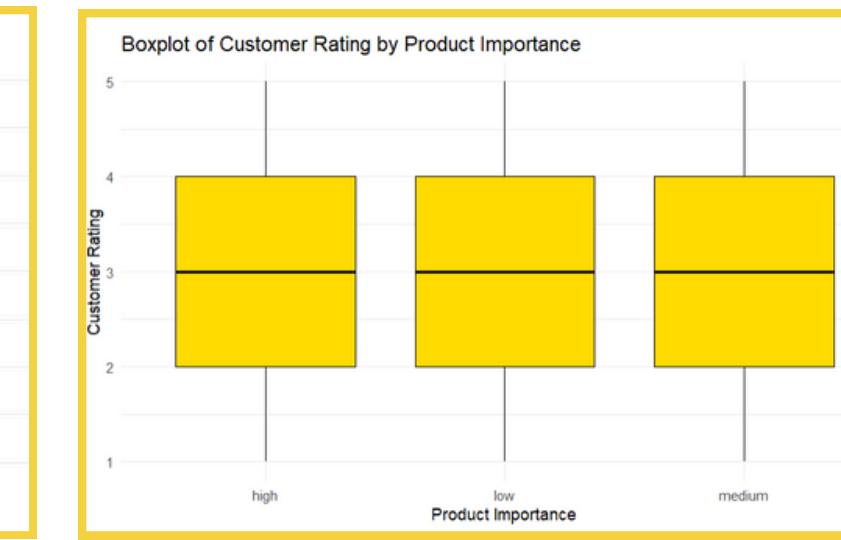
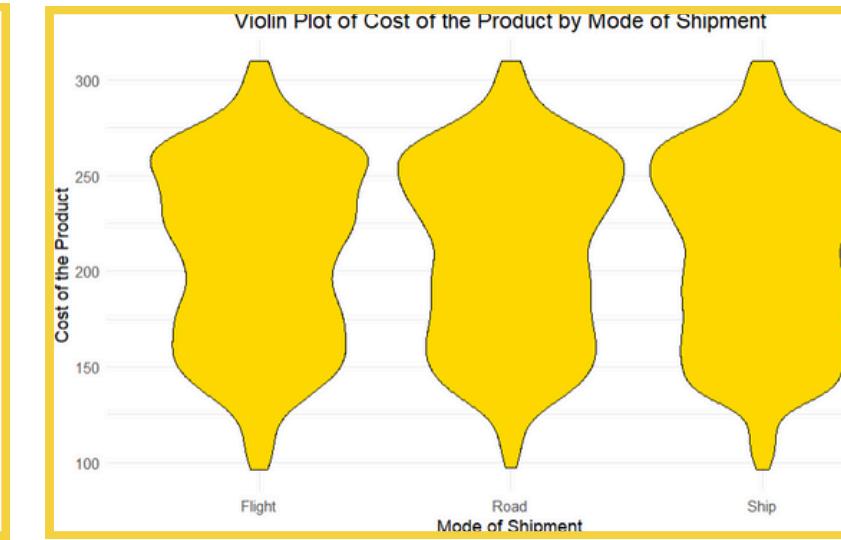
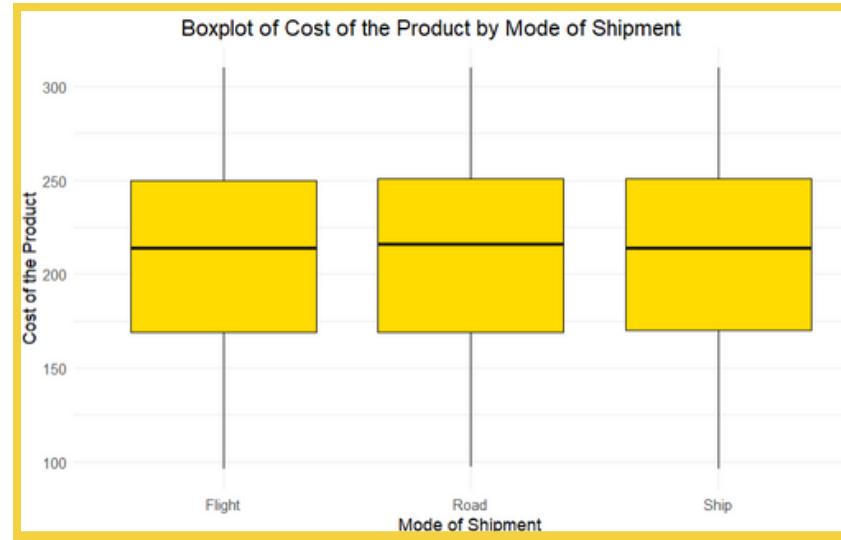
STAT 208



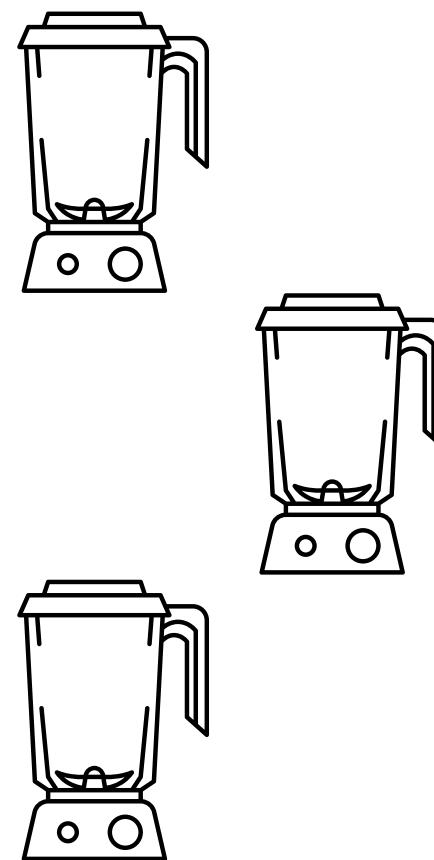
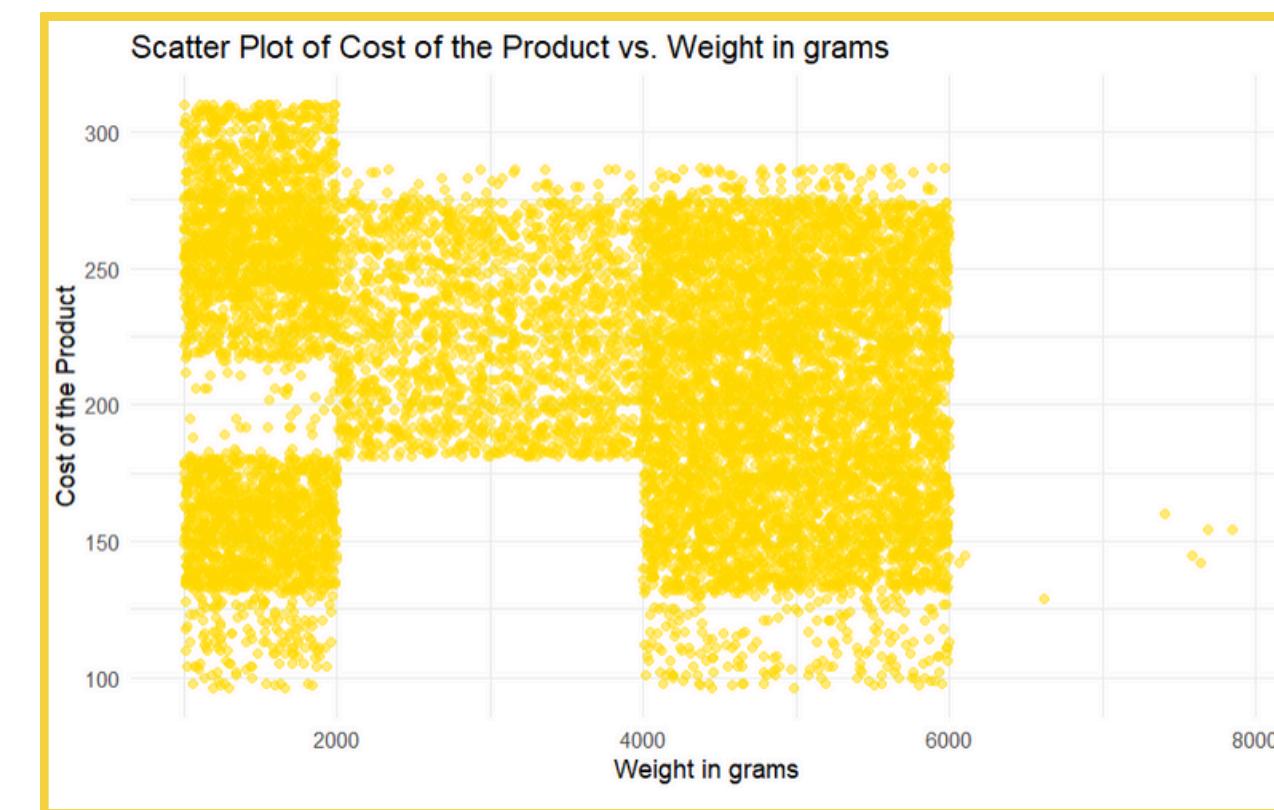


Bivariate Analysis - Categorical vs. Numerical

STAT 208



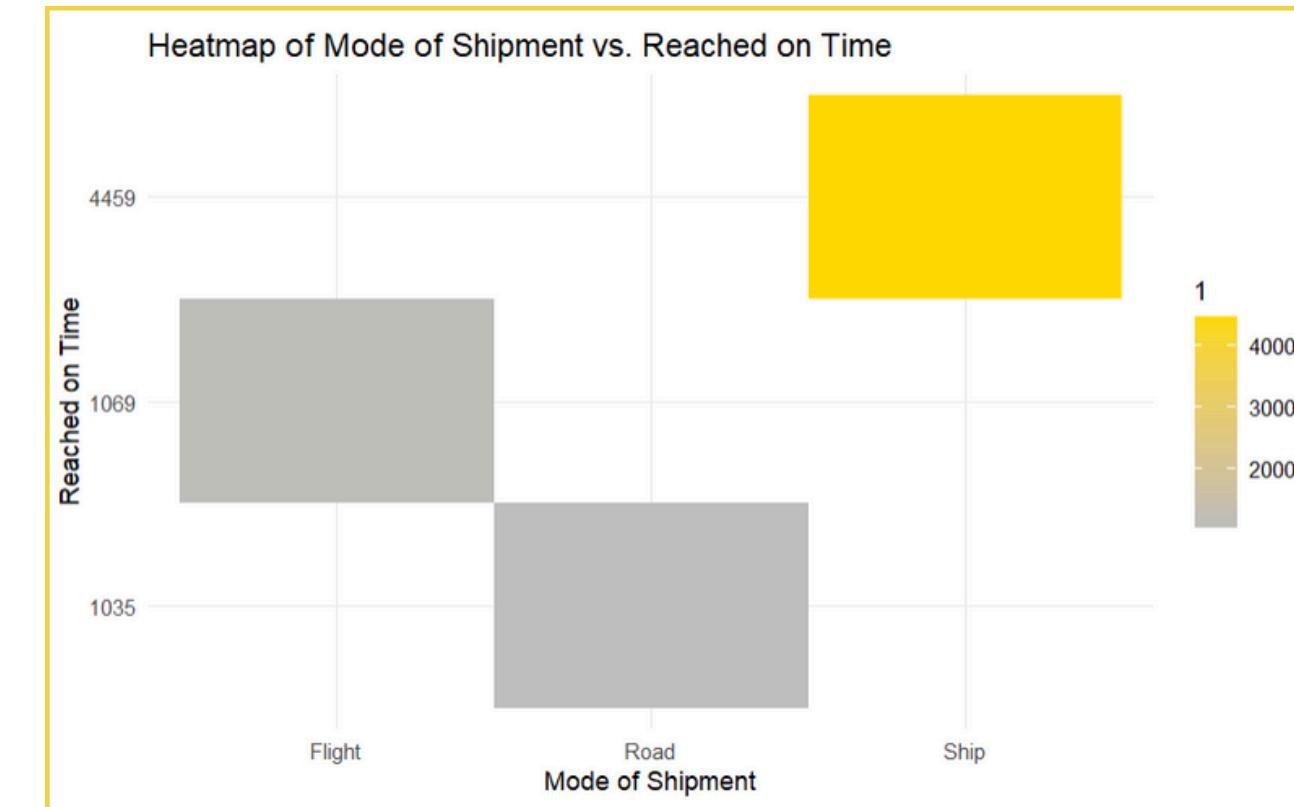
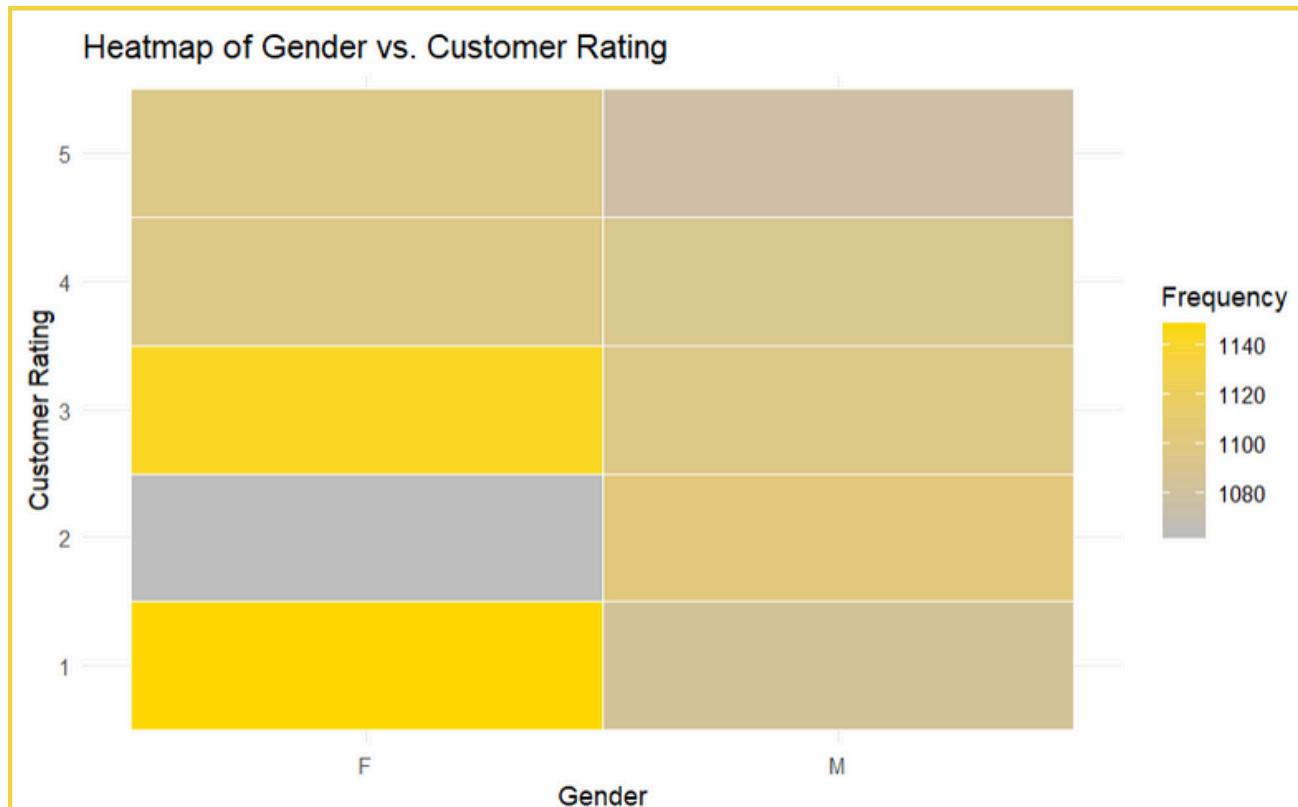
Bivariate Analysis - Numerical vs. Numerical





Bivariate Analysis - Categorical vs Categorical

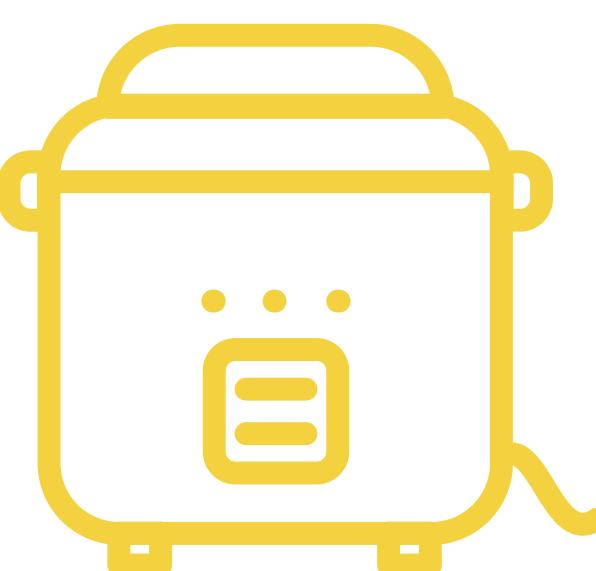
STAT 208



Chi-Square Test for Mode of Shipment vs. Reached on Time:

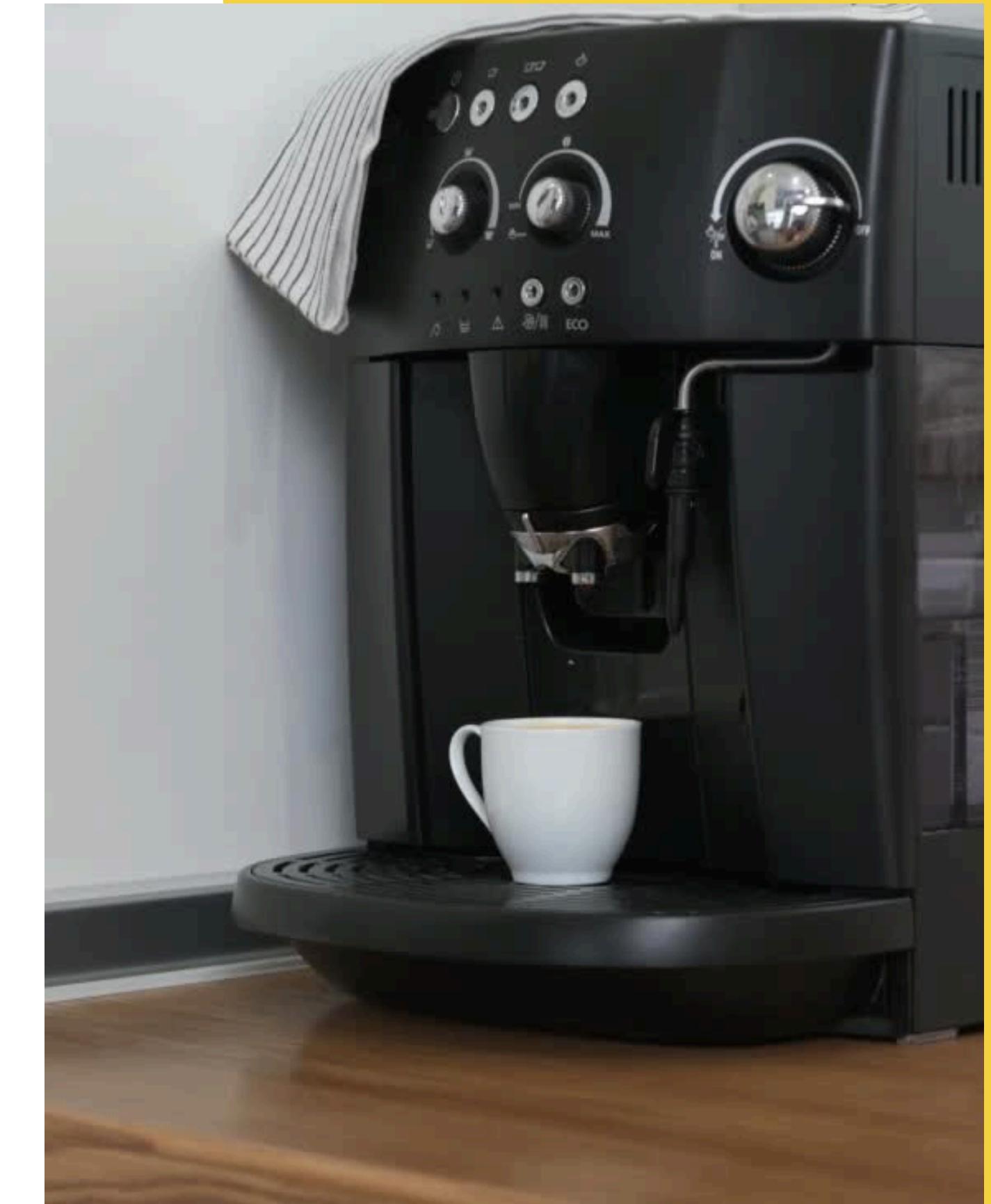
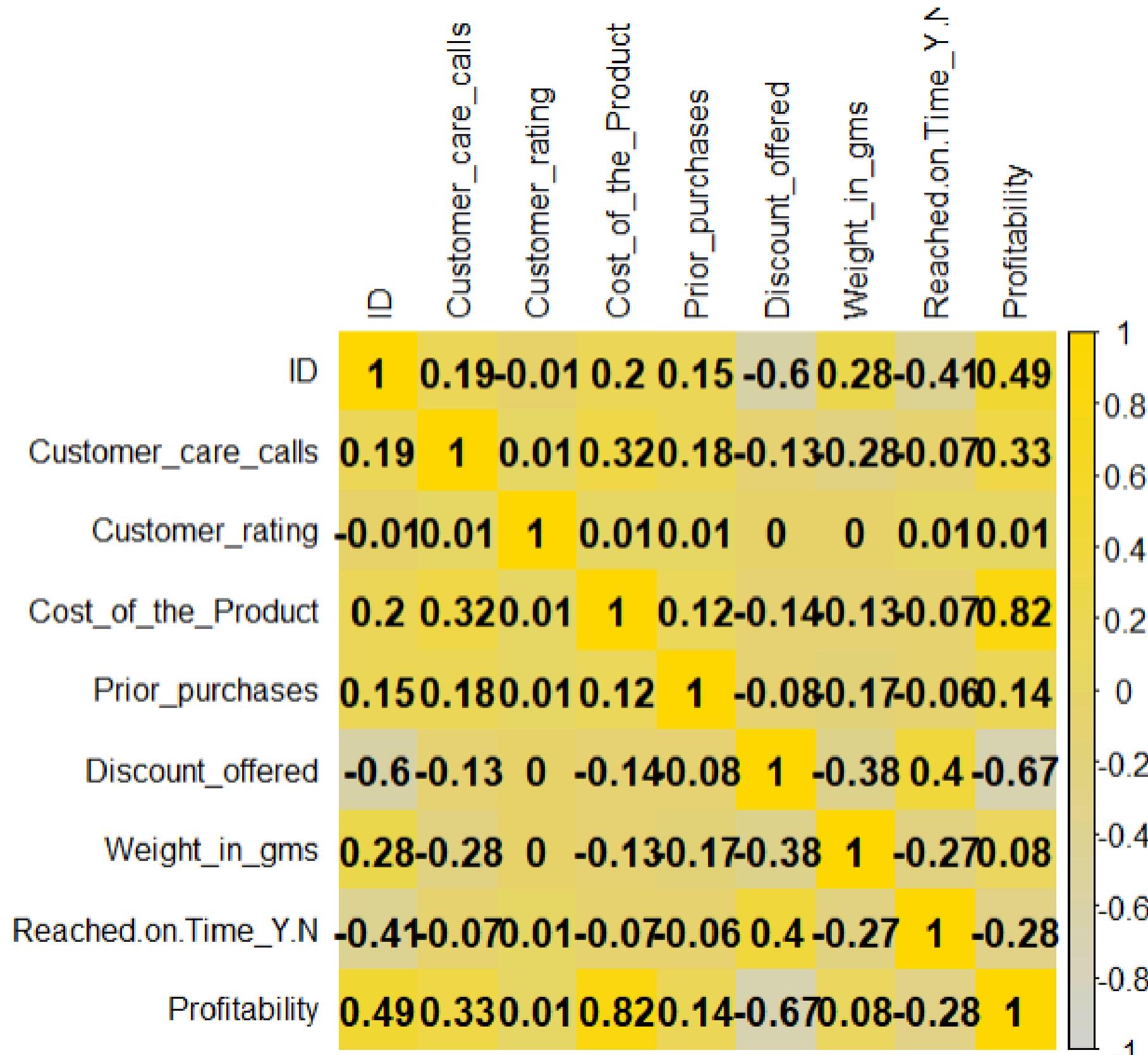
Pearson's chi-squared test

```
data: contingency_table  
X-squared = 0.74344, df = 2, p-value = 0.6895
```



Multivariate Analysis - Heatmap of correlation coefficients to understand relationships between numerical variables

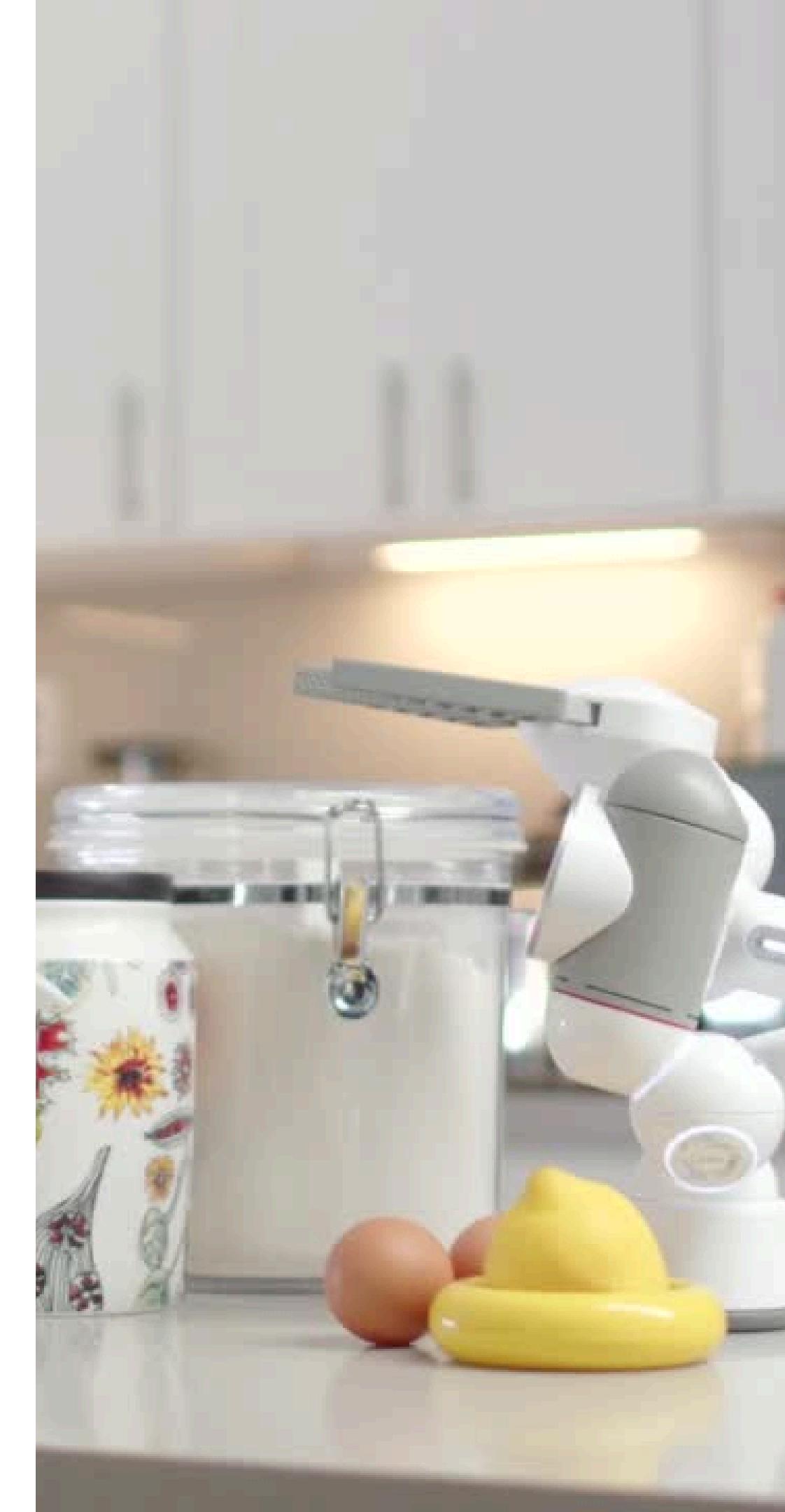
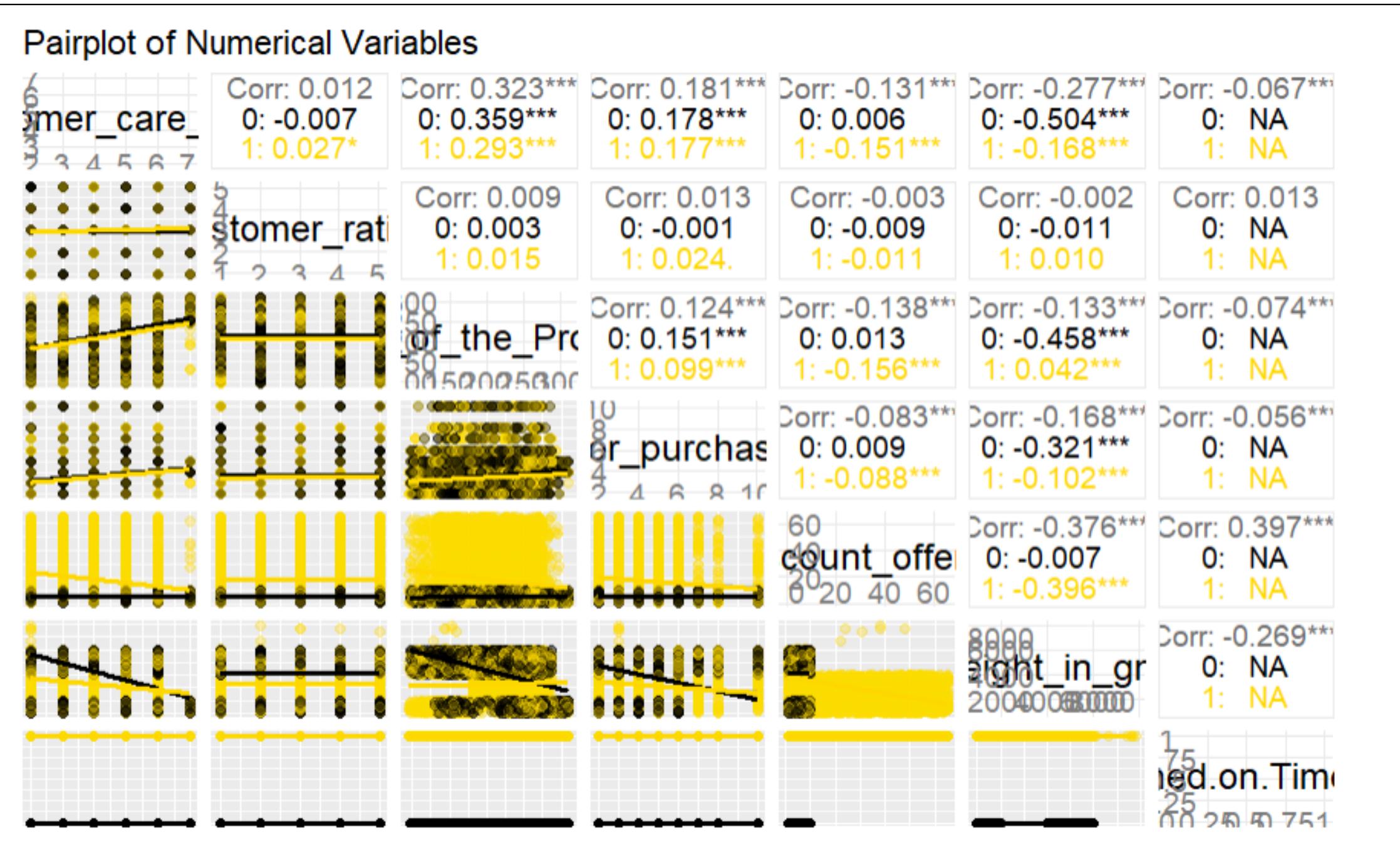
STAT 208



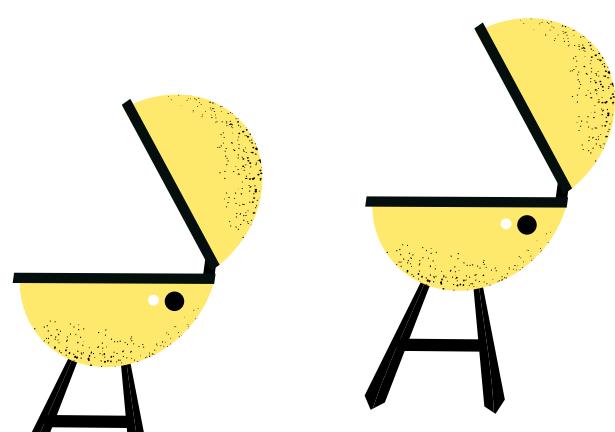
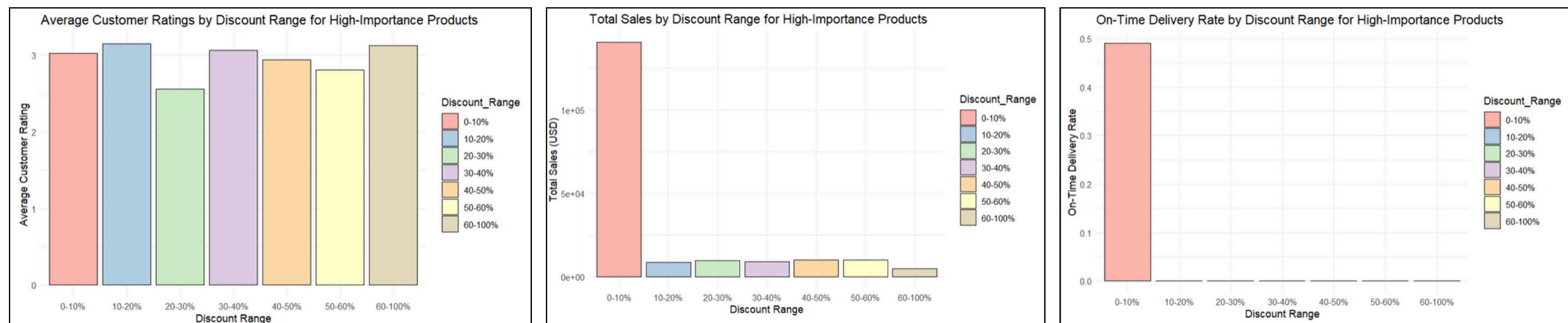
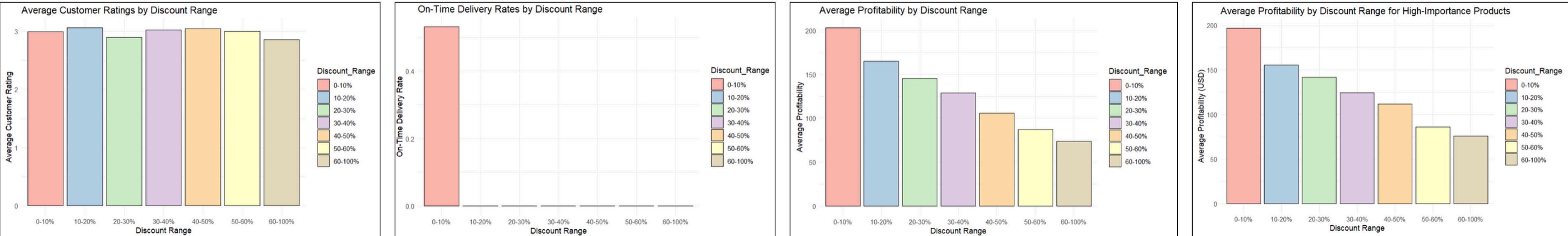


Multivariate Analysis – Pairplot

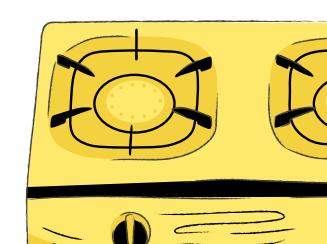
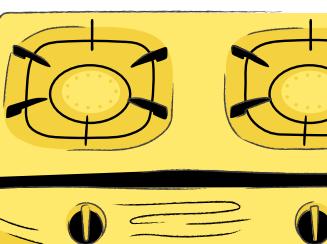
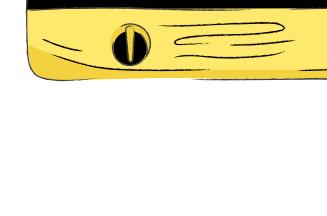
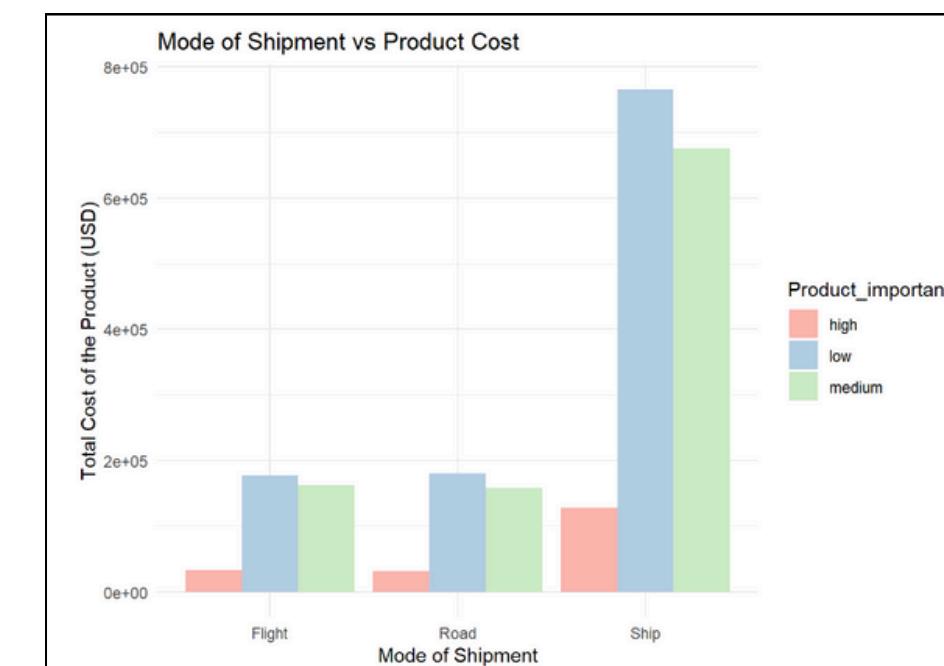
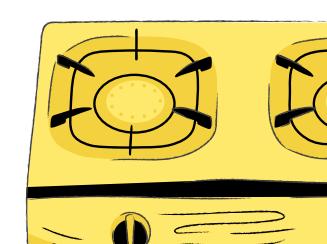
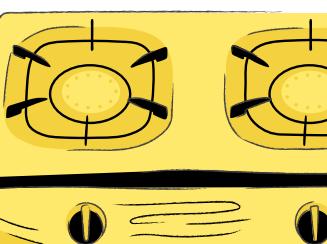
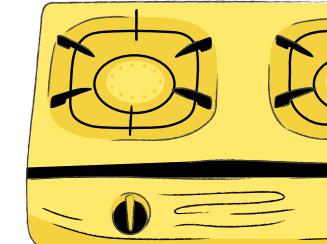
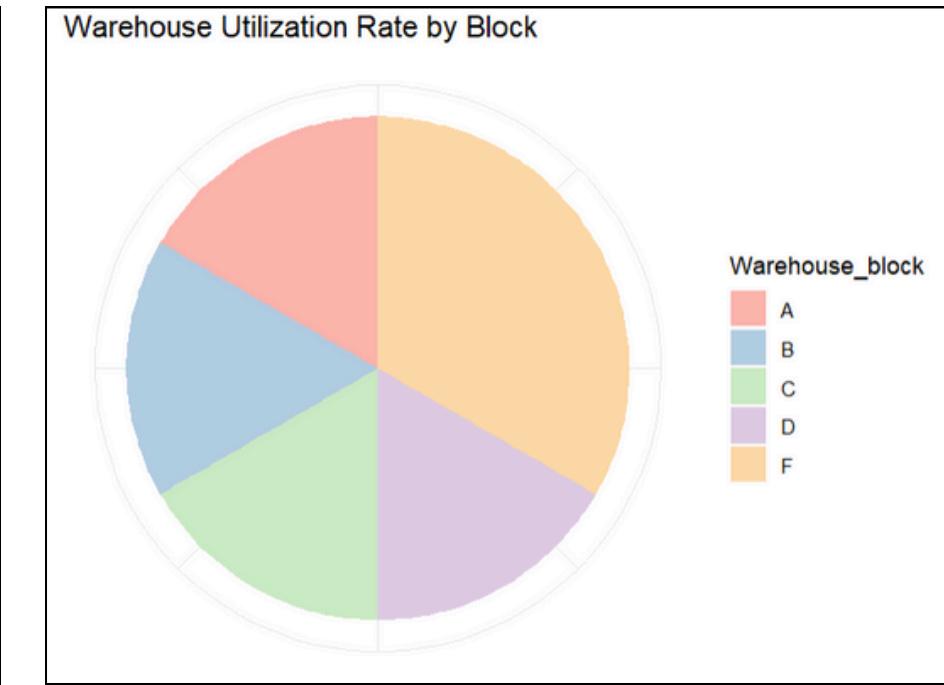
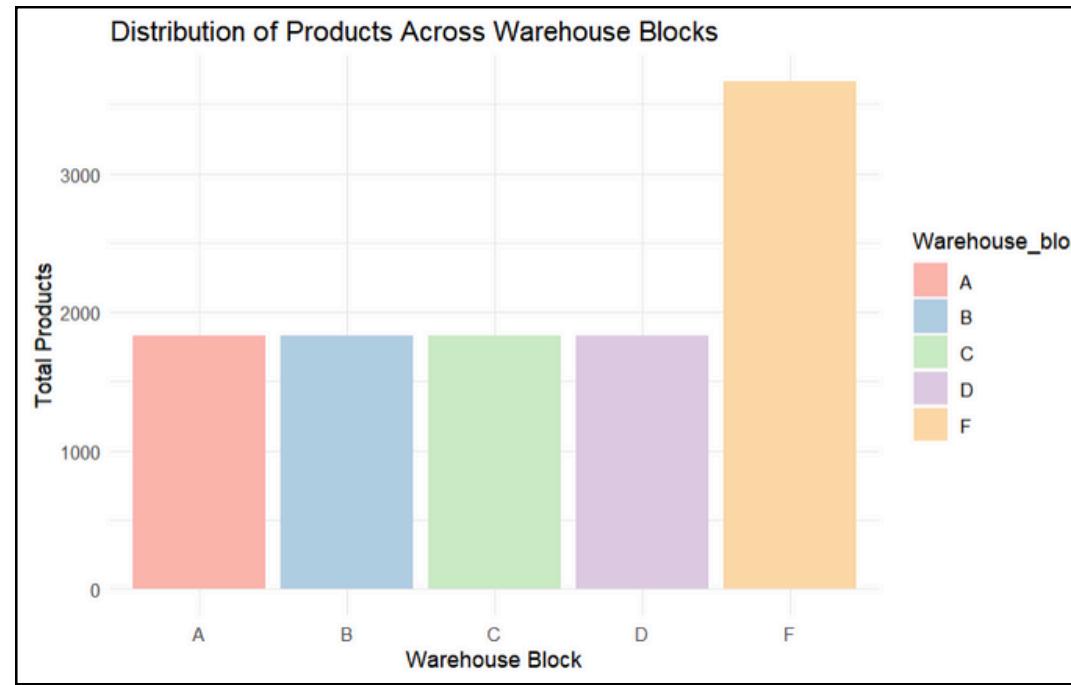
STAT 208



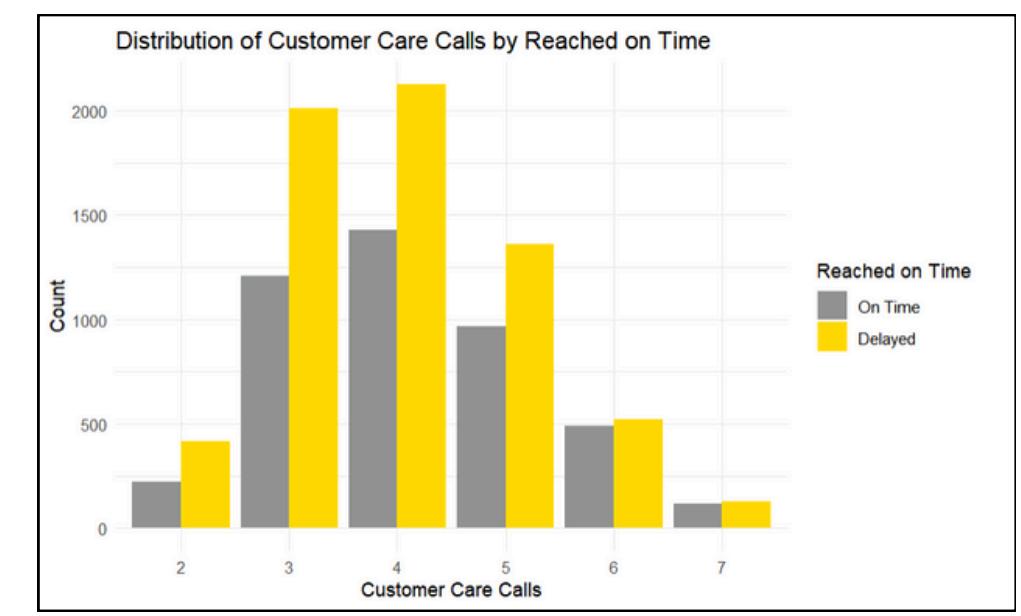
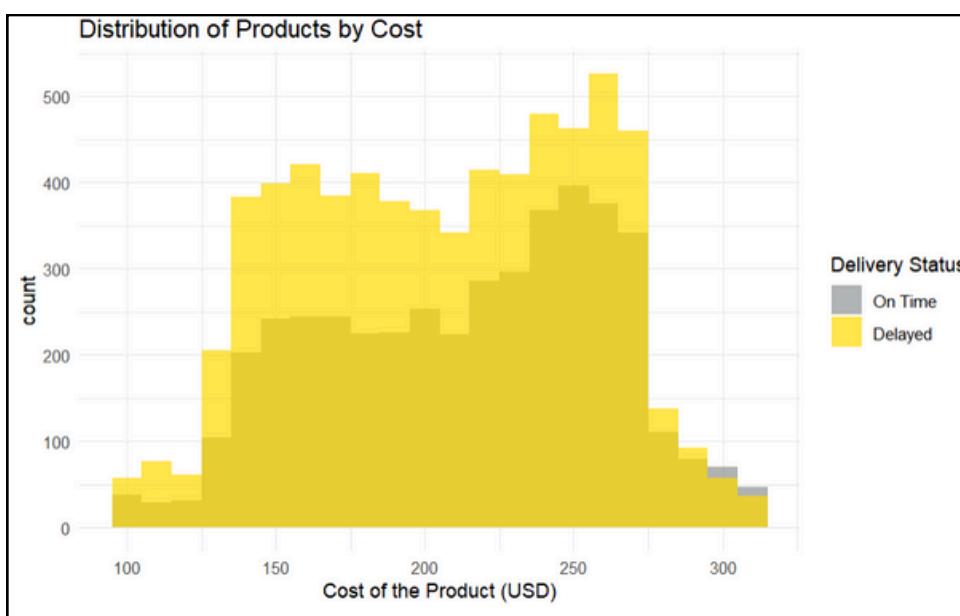
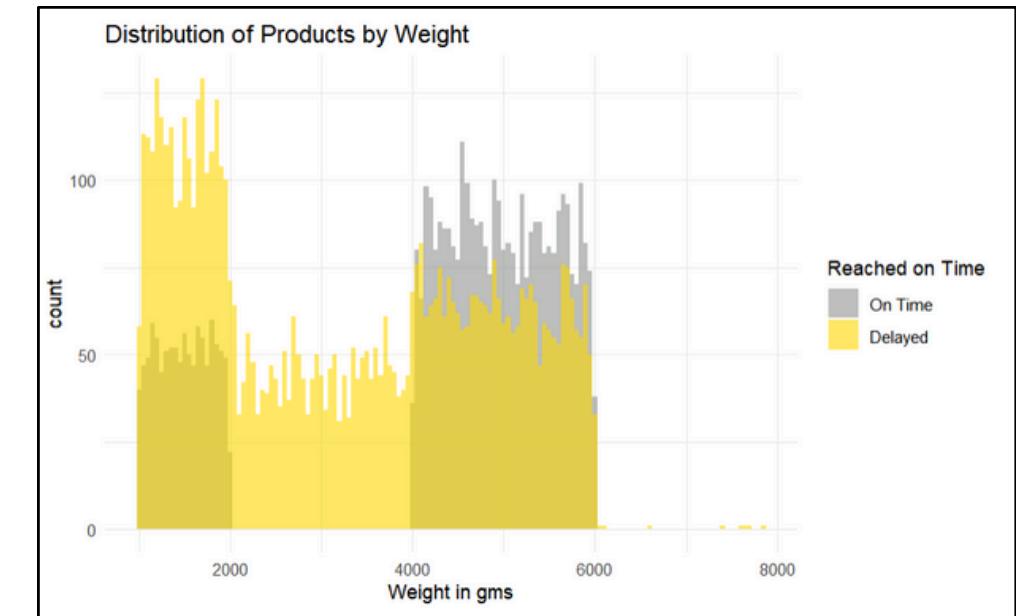
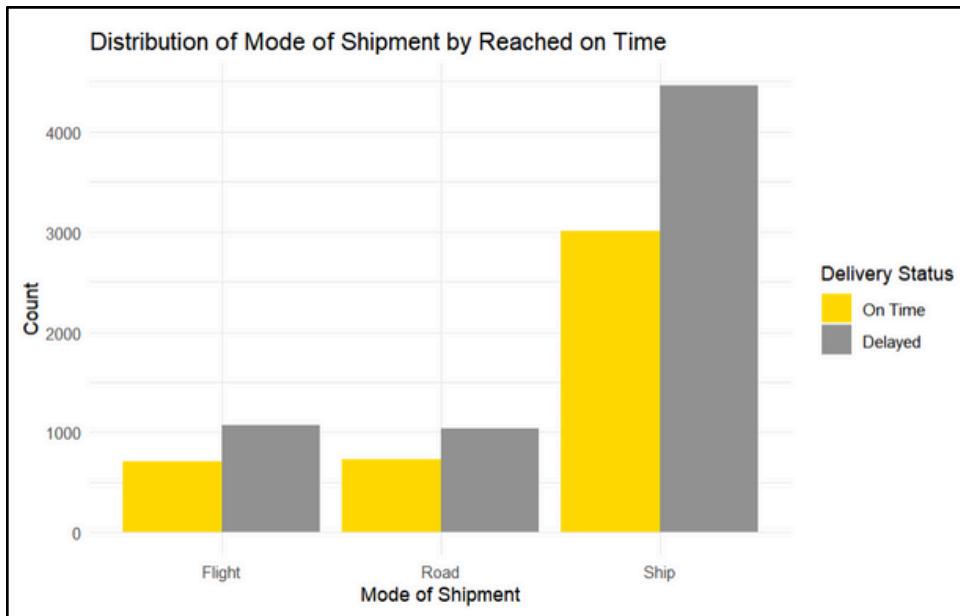
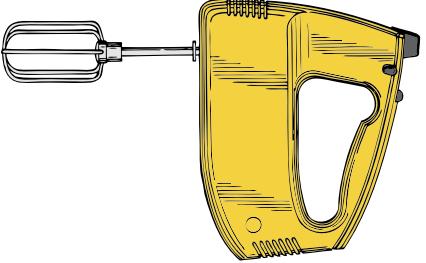
Objective Specific Analysis



Objective Specific Analysis



Objective Specific Analysis



Objective 1: Evaluate Transportation Mode Effectiveness

STAT 208

Models



Logistics Regression

Class Imbalance: ROSE package to generate synthetic samples

Model Training: Training (70%) and Testing (30%) sets

Performance Metrics: Training Accuracy: 72.92%
Test Accuracy: 71.54%

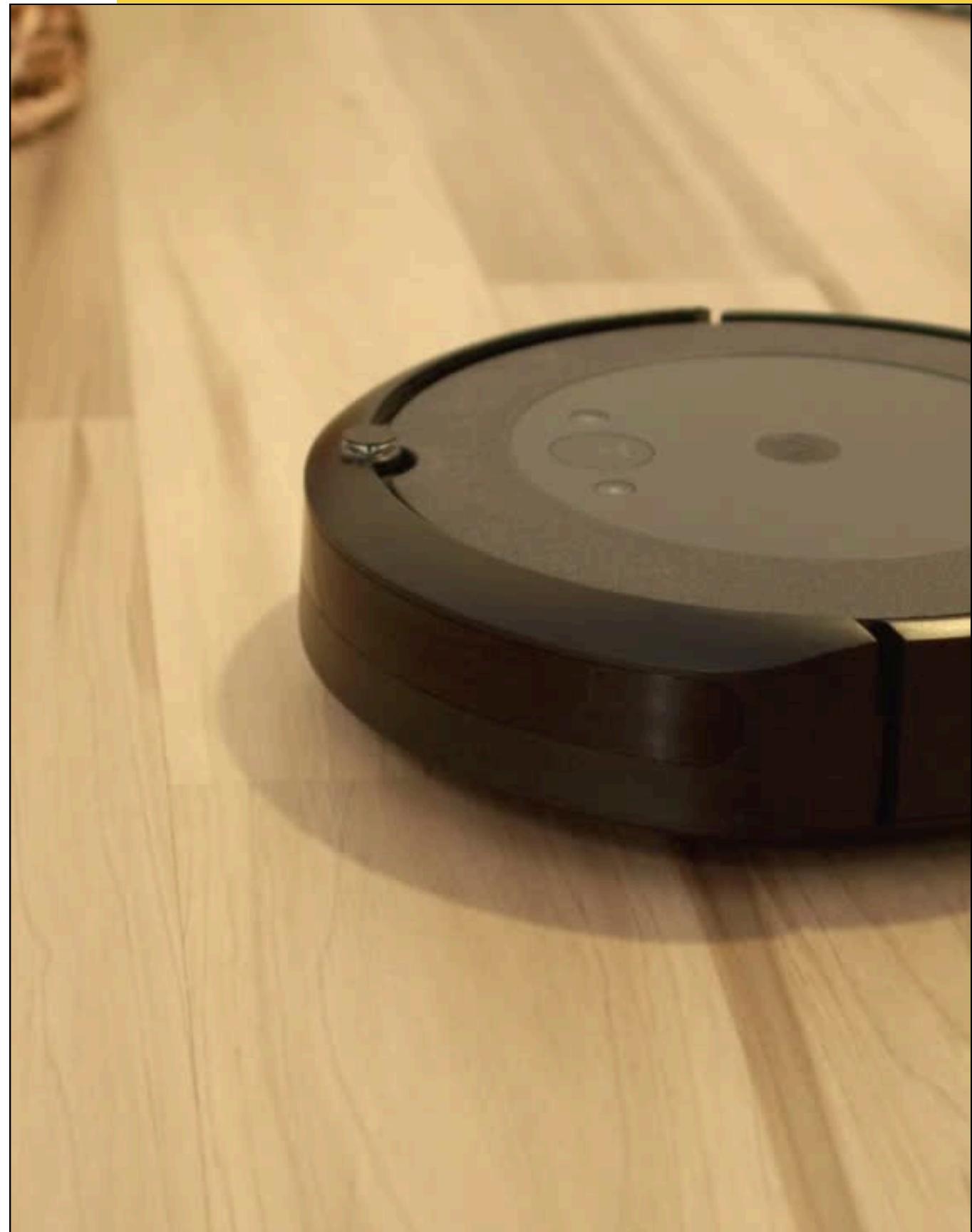
Confusion Matrix:

		0 (On Time)	1 (Delayed)
0 (On Time)	1628	914	
	25	732	

- Sensitivity: 98.49% (On Time)
- Specificity: 44.47% (Delayed)
- Precision: 64.04%
- F1 Score: 75.92%

Key Insights:

- Strengths: High sensitivity indicates the model effectively identifies on-time shipments.
- Weaknesses: Lower specificity suggests difficulty in identifying delayed shipments.
- Predictive Values: The model demonstrates moderate accuracy and robustness in its predictions



DECISION TREE

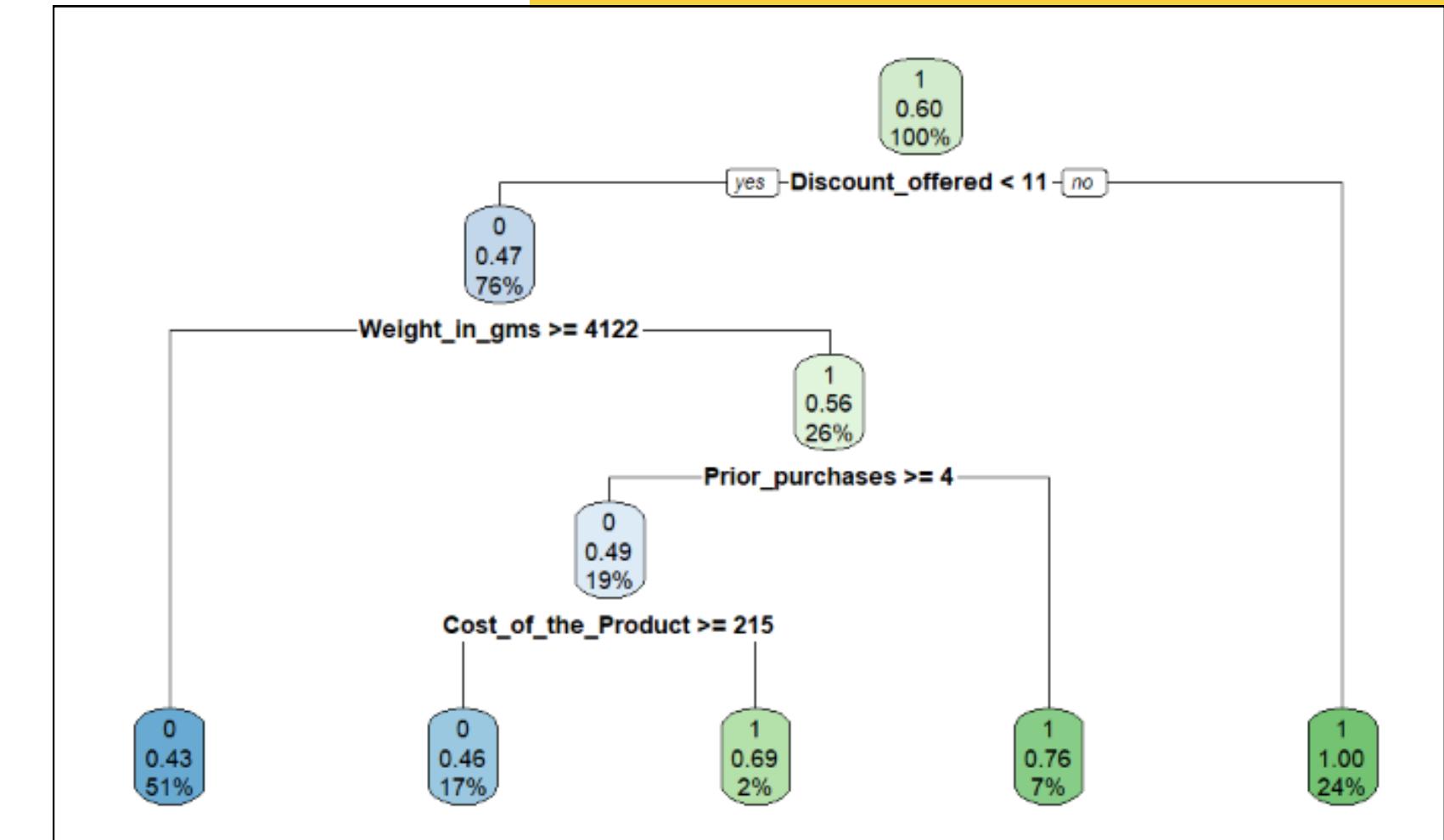
Decision Tree Structure: 5 branches

Performance Metrics: Training Accuracy: 68.44%
Test Accuracy: 68.34%

Confusion Matrix:

		0 (On Time)	1 (Delayed)
0 (On Time)	0 (On Time)	1268	982
	1 (Delayed)	62	986

- Sensitivity: 95.34% (On Time)
- Specificity: 50.10% (Delayed)
- Precision: 56.36%
- F1 Score: 92.08%



Key Insights:

- It achieves a reasonable accuracy of approximately 68.34% on the test set, which is slightly lower than the training accuracy of 68.45%. This indicates the model generalizes well to new data.
- The model has a high sensitivity (95.34%), meaning it is very effective at identifying positive cases (0s). However, the specificity is lower (50.10%), indicating it struggles more with identifying negative cases (1s).

Random Forest Model

Model Description: Trees: 500
Variables/Split: 3

OOB Error Rate: 34.4%

Performance Metrics: Accuracy: 66.8%

Confusion Matrix:

		0 (On Time)	1 (Delayed)
0 (On Time)	0 (On Time)	1026	791
	1 (Delayed)	304	1177

- Sensitivity: 76.77% (On Time)
- Specificity: 59.81% (Delayed)
- Precision: 56.35%
- F1 Score: 79.21%

Key Insights:

- High Sensitivity: Good at identifying on-time deliveries.
- Lower Specificity: Struggles with identifying delayed shipments.
- Predictive Values: More reliable at predicting delays than on-time deliveries.
- Top Features: Weight in gms, Discount offered, Cost of the Product



Gradient Boosting

Cross-Validation: 5-fold CV

Optimal Parameters: Number of Trees : 50

Interaction Depth: 2

Shrinkage: 0.1

Minimum Observations in Node : 10

Performance Metrics: Training Accuracy: 68.58%

Test Accuracy: 69.38%

Confusion Matrix:

		0 (On Time)	1 (Delayed)
0 (On Time)	0 (On Time)	1320	1025
	1 (Delayed)	10	943

- Sensitivity: 99.25% (On Time)
- Specificity: 47.92% (Delayed)
- Precision: 56.29%
- F1 Score: 98.95%

Key Insights:

- Sensitivity: Effectively identifies on-time shipments.
- Specificity: Struggles with identifying delays.
- Predictive Values: More reliable at predicting delays than on-time deliveries.
- Balanced Accuracy: Accounting for class imbalance, the model's balanced accuracy is 72.48%.



XG Boost

STAT 208

Model Training: Training (70%) and Testing (30%) sets

Performance Metrics: Training Accuracy: 69.65.2%
Test Accuracy: 67.85%

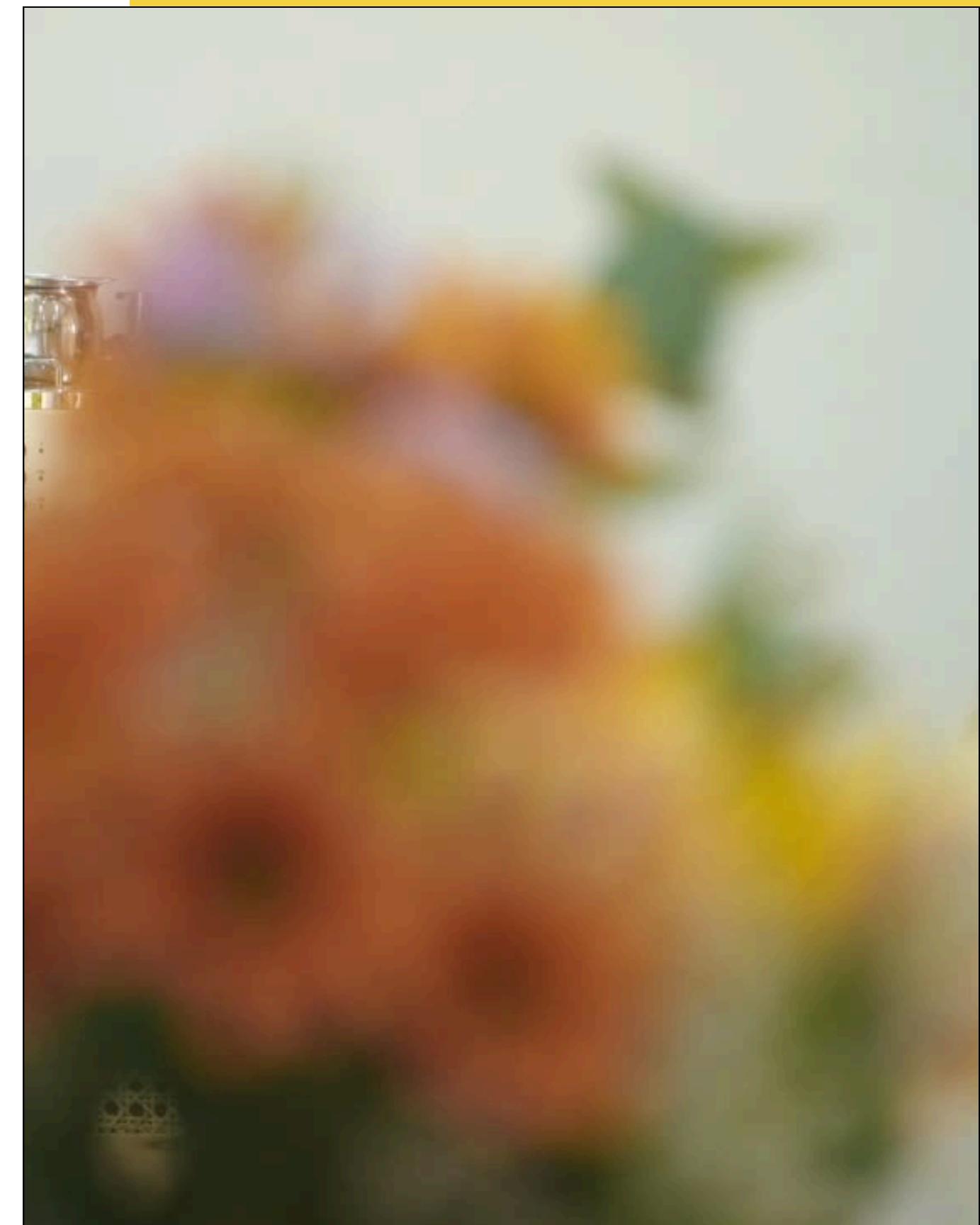
Confusion Matrix:

		0 (On Time)	1 (Delayed)
0 (On Time)	0 (On Time)	1181	911
	1 (Delayed)	149	1057

- Sensitivity: 88.80% (On Time)
- Specificity: 53.71% (Delayed)
- Precision: 56.45%
- F1 Score: 87.65%

Key Insights:

- Model consistently performs well on both training and testing datasets, indicating robustness.
- Predictive Values: The model has high accuracy in correctly identifying both positive and negative cases



Comparison:

- Logistic Regression: A culinary delight with precision and balance.
- Decision Tree: A flavorful dish, enjoyed by some but not all.
- Gradient Boosting & Random Forest: Ingredients are good, needs a bit more seasoning (Hypertuning).

Recommendation:

- Logistic Regression: The chef's favorite! Reliable, accurate, and always a hit.



Objective 2: Optimize the Impact of Shipping Discounts

PCA



PCA Overview

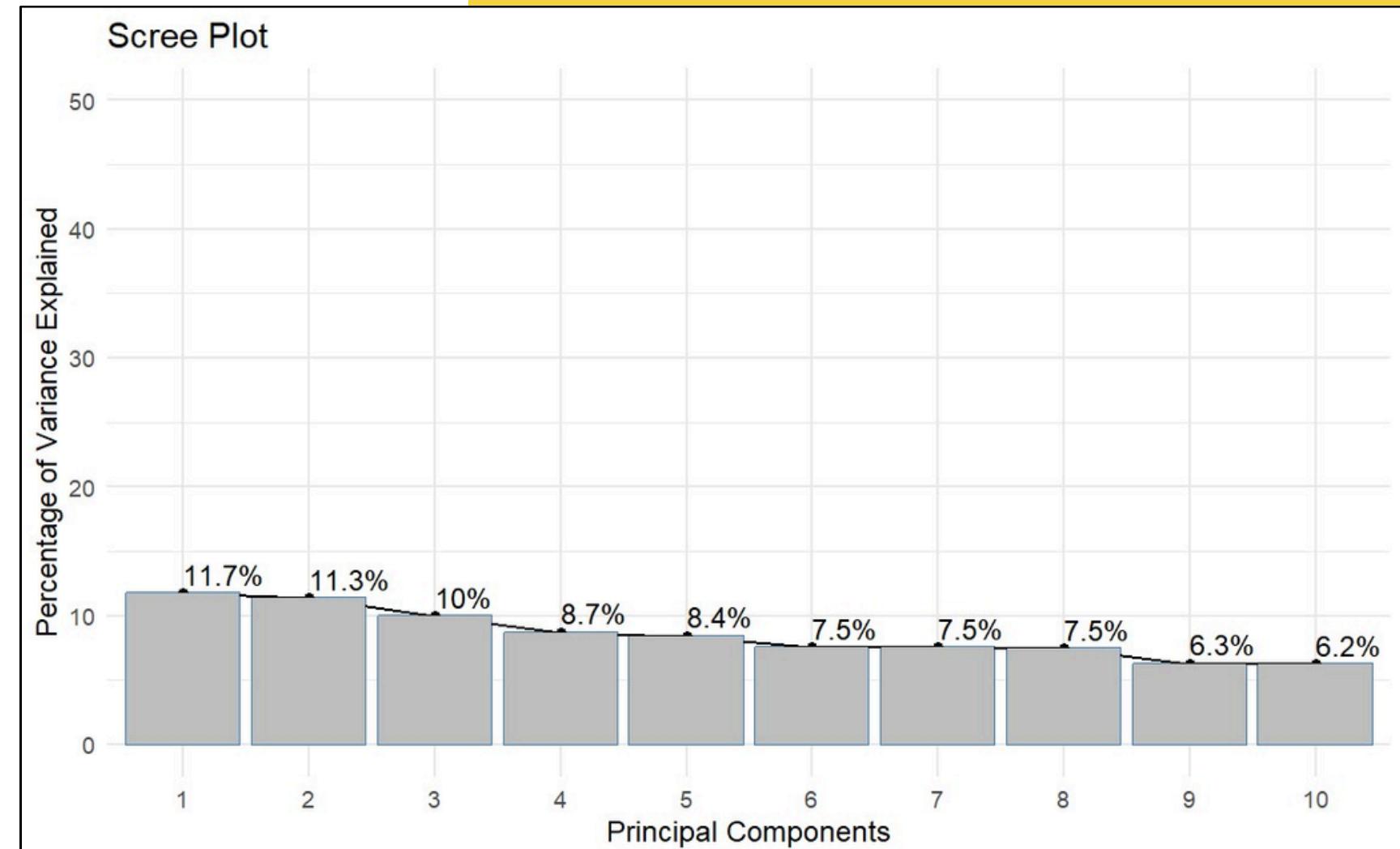
Selected Variables: Customer Care Calls, Customer Rating, Cost of the Product, Prior Purchases, Discount Offered, Weight in gms, Mode of Shipment, Warehouse Block, Product Importance, Gender

Total Principal Components: 16

Explained Variance:

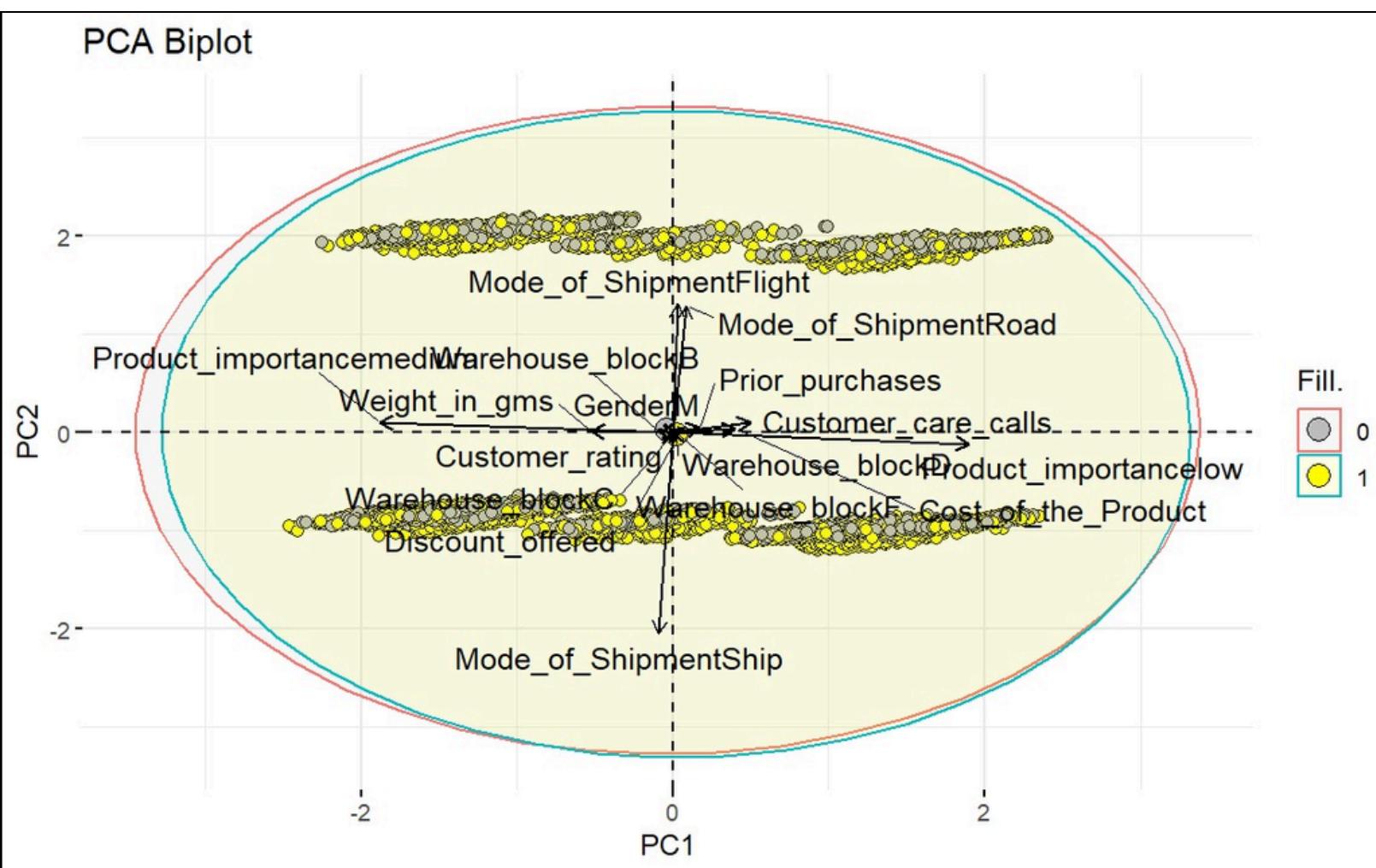
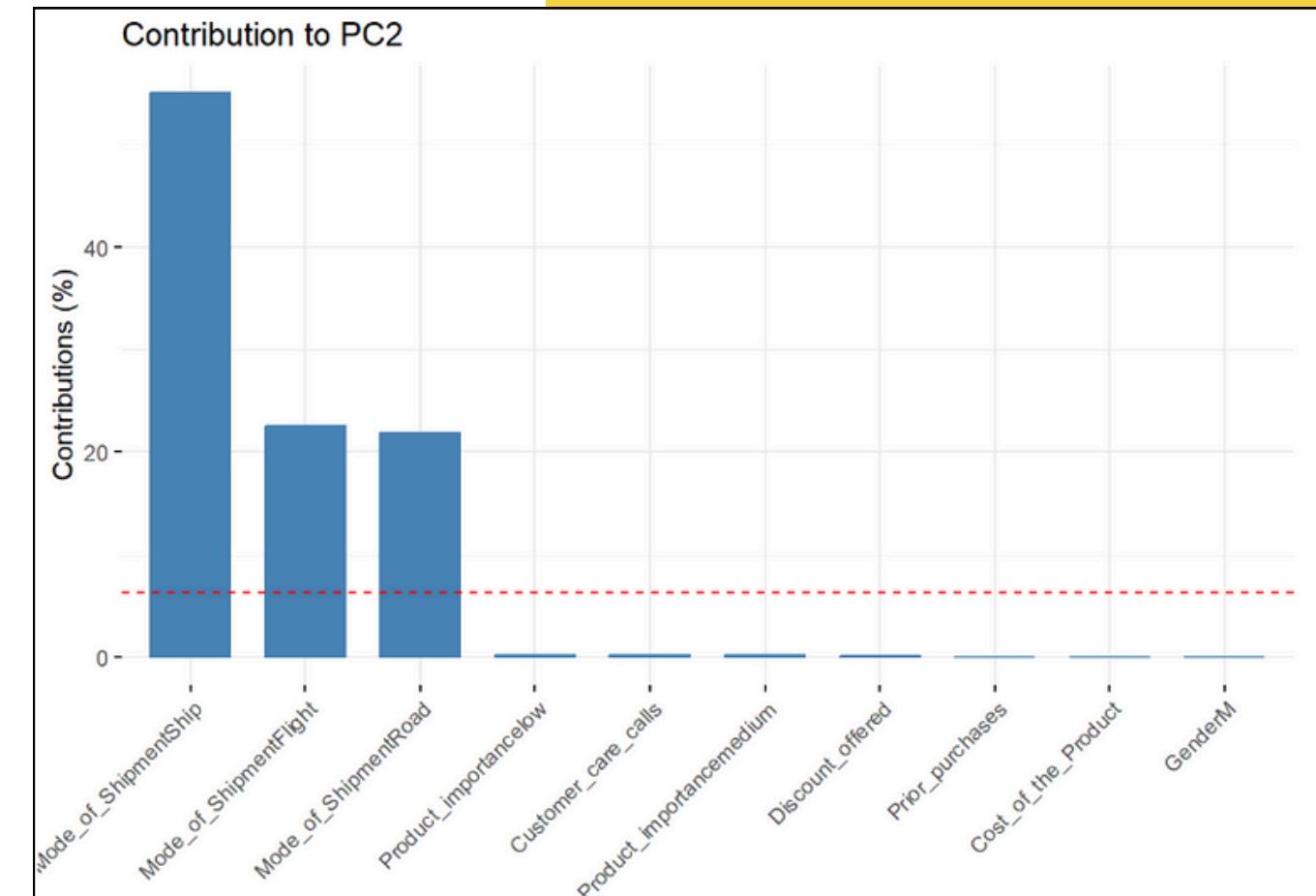
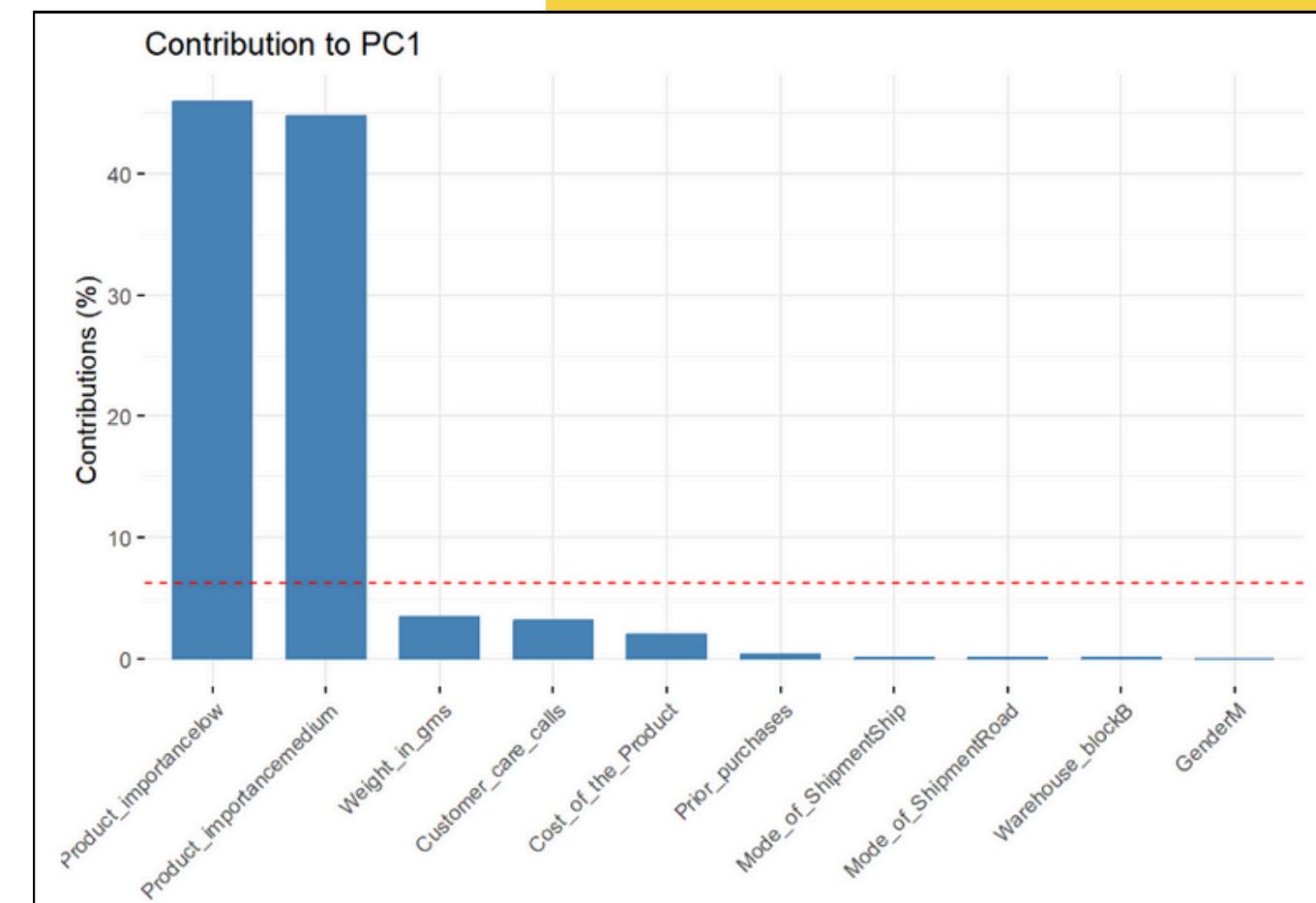
PC1: 11.7%

PC2: 11.3%



Key Influences & PCA Biplot

- Key Influences on Principal Components:
 - PC1 Influences: Product Importance, Weight in gms, Customer Care Calls
 - PC2 Influences: Mode of Shipment
- PCA Biplot
 - Visual representation of sample distribution and variable contributions.



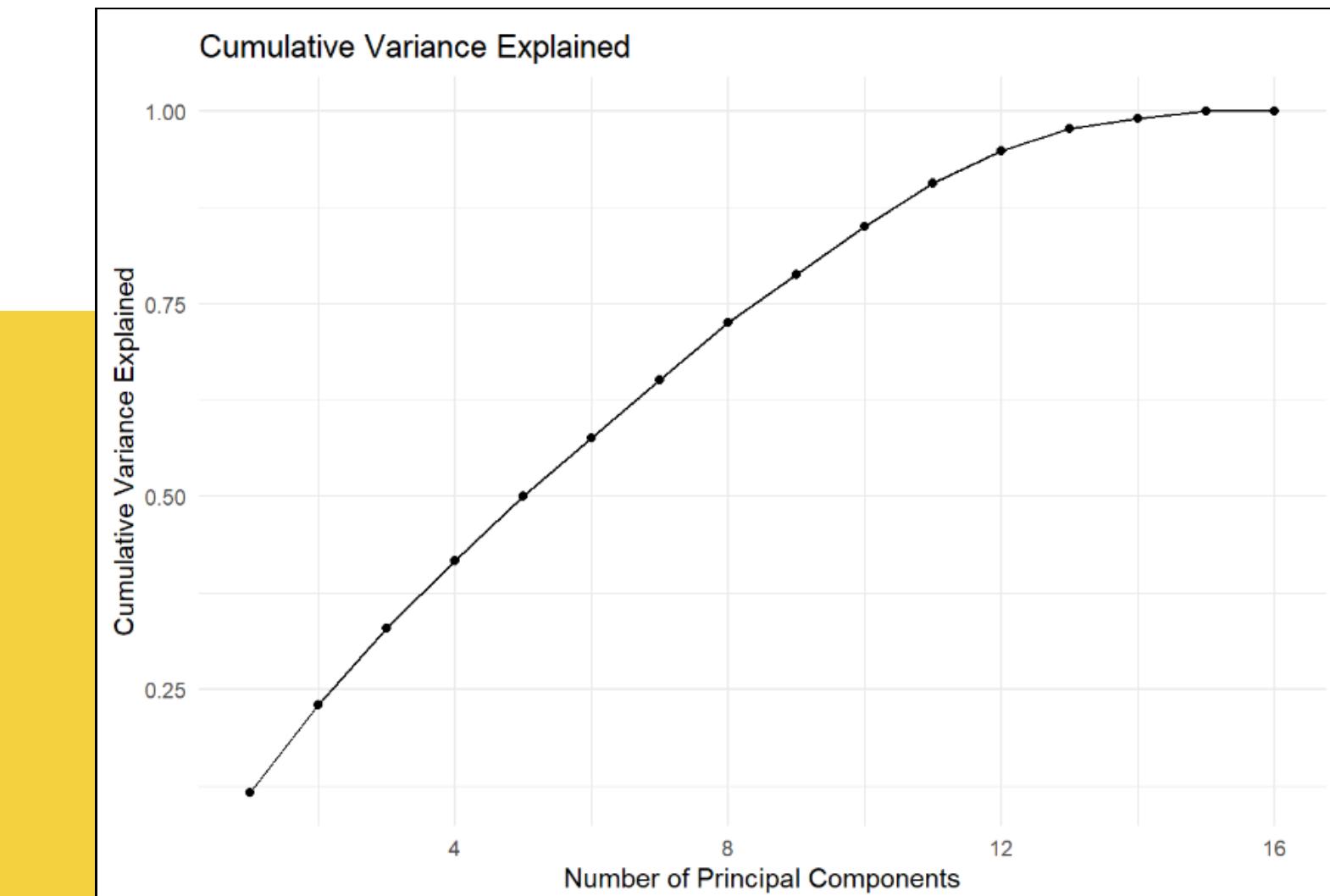
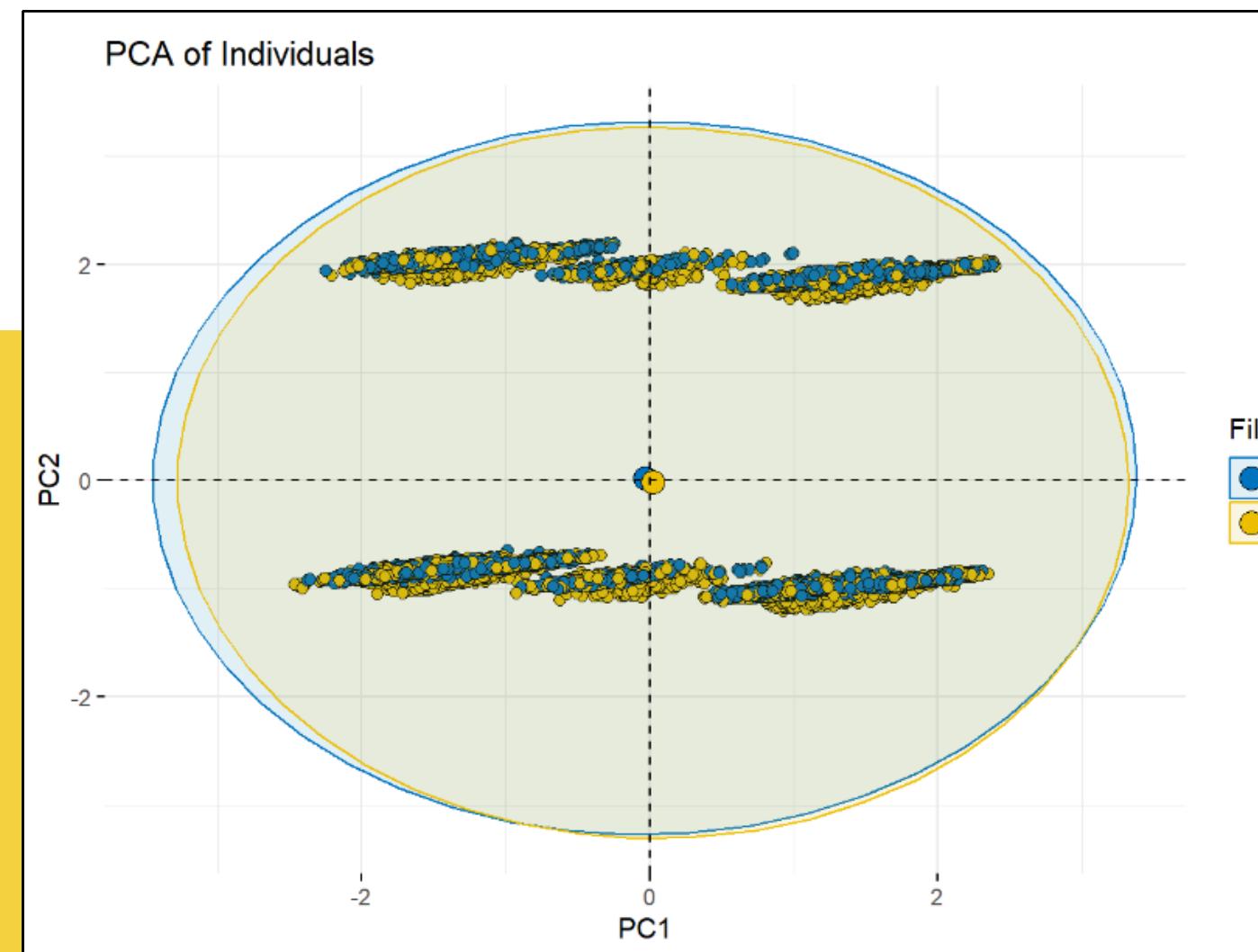
Cumulative Variance & PCA of Individuals

Cumulative Variance

- PC1-PC7 : 65%
- PC1-PC10: 85%
- PC1-PC13: 95%

PCA of Individuals

- There is an evident clustering pattern, with some overlap between the two classes



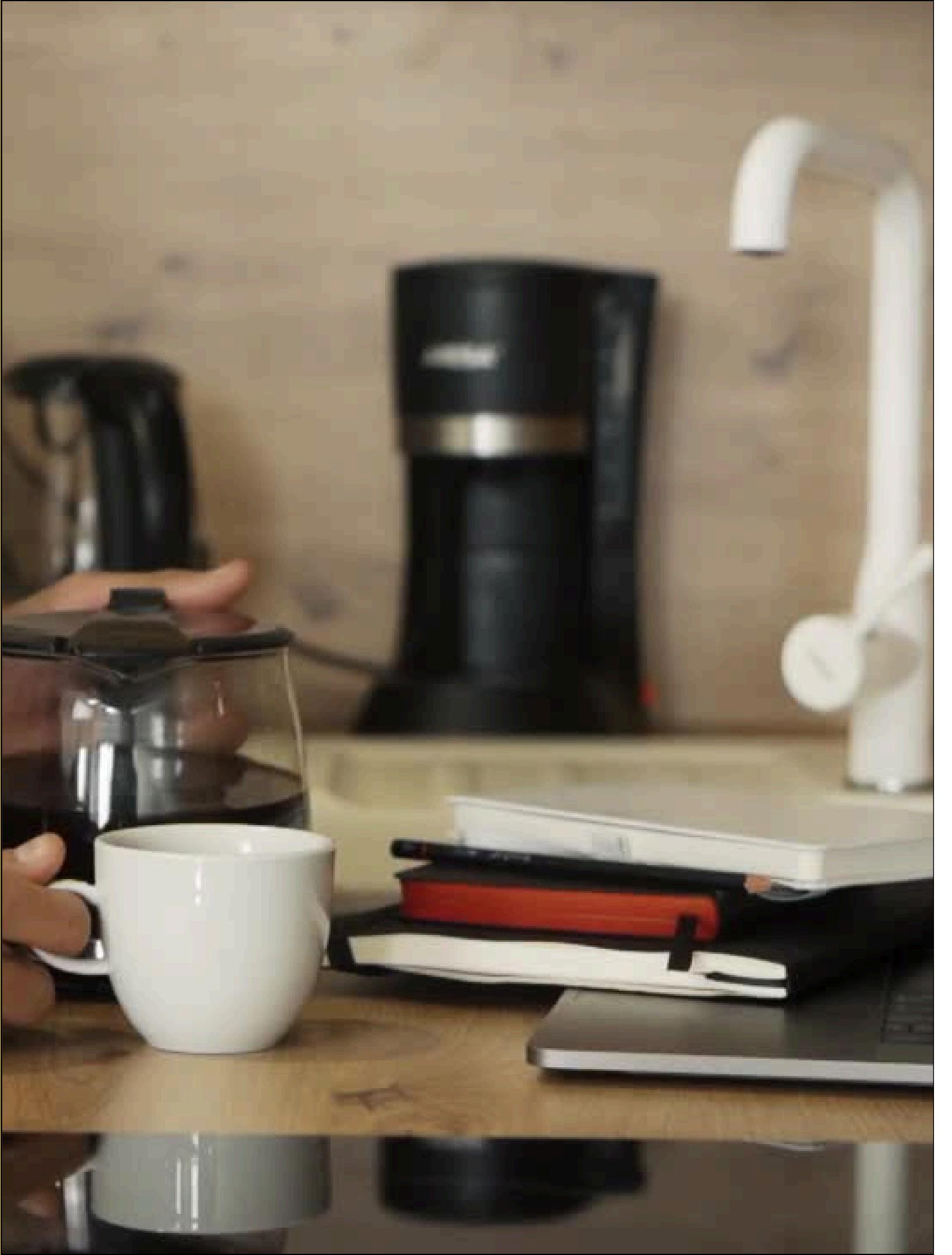
Recommendations

- Customer Care Calls:

Enhance Customer Service: Let's sprinkle some extra care to our customer interactions for delightful experiences!

- Shipping Method Selection:

Choose Wisely: Select shipping methods based on product importance and weight for on-time deliveries, every time!



Objective 2:
Optimize the Impact of Shipping Discounts

K-Means

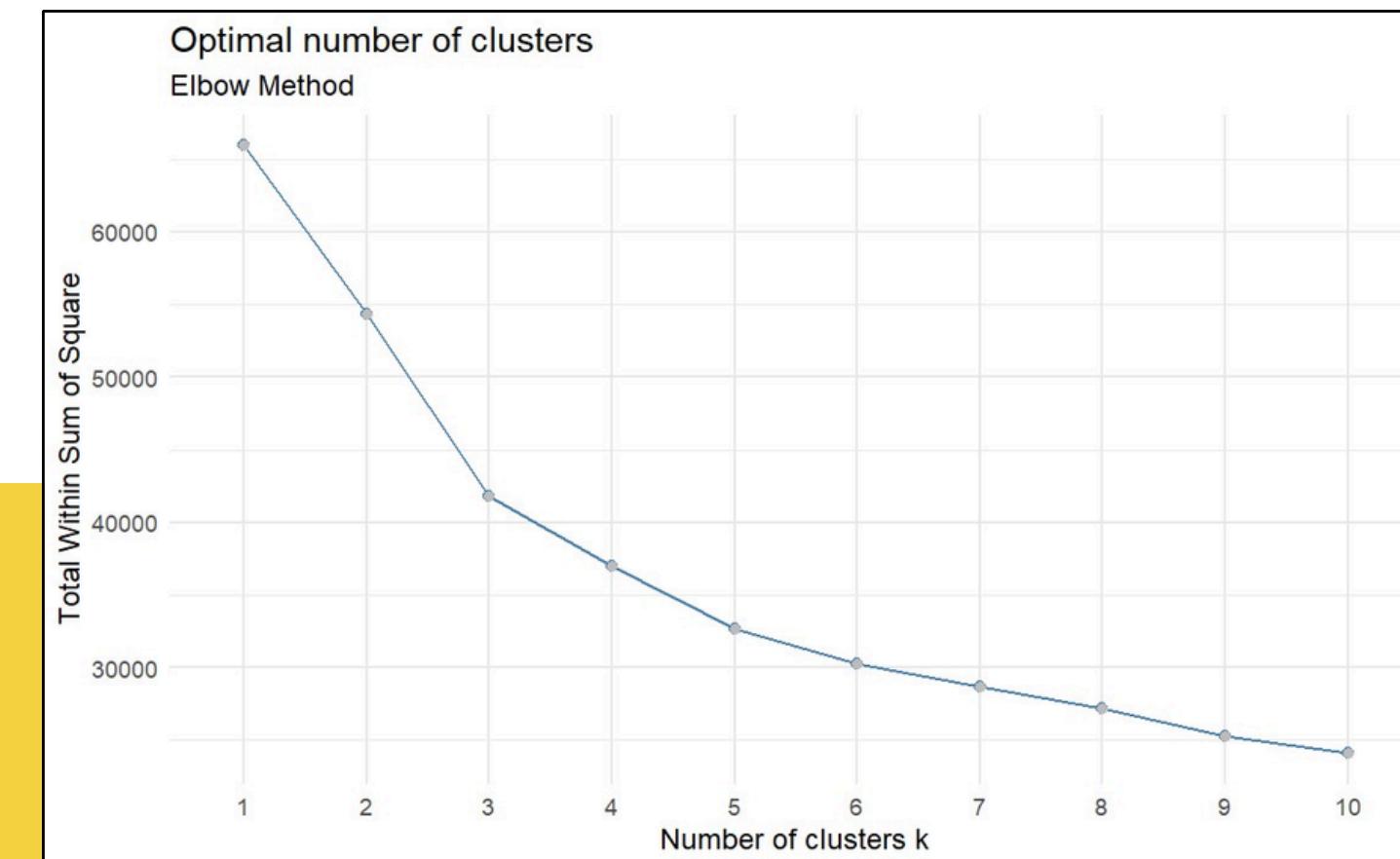




Optimal Number of Clusters

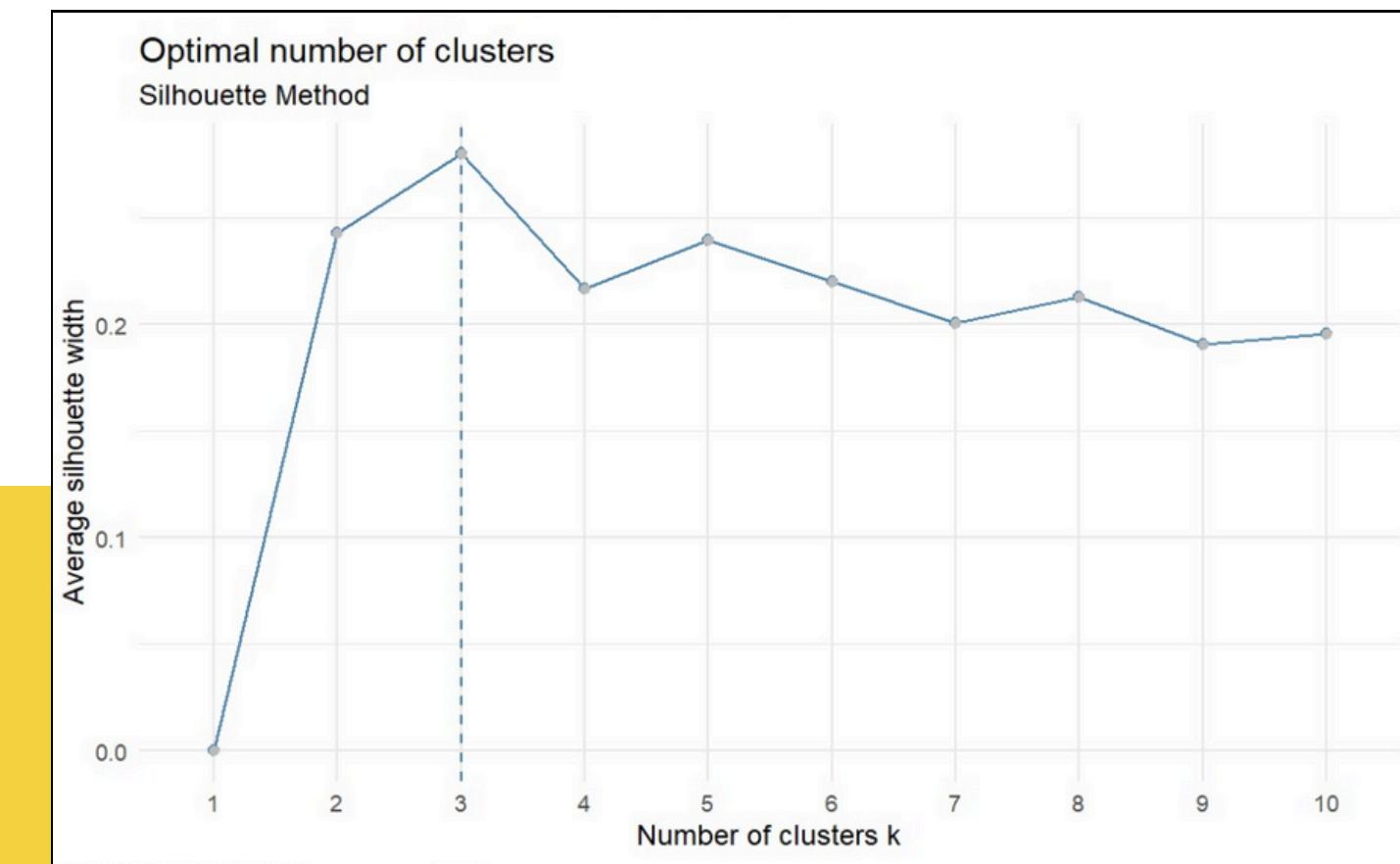
Elbow Method Plot

- Plot of Within-Cluster Sum of Squares (wss) against the number of clusters.
- Optimal k is typically where the WSS starts to decrease at a slower rate (the "elbow" point)



Silhouette Method Plot

- Measures how similar an object is to its own cluster compared to other clusters.
- Optimal k is the number of clusters that maximizes the average silhouette width



The optimal number of clusters was determined to be 3



Optimal Number of Clusters

Cluster Sizes:

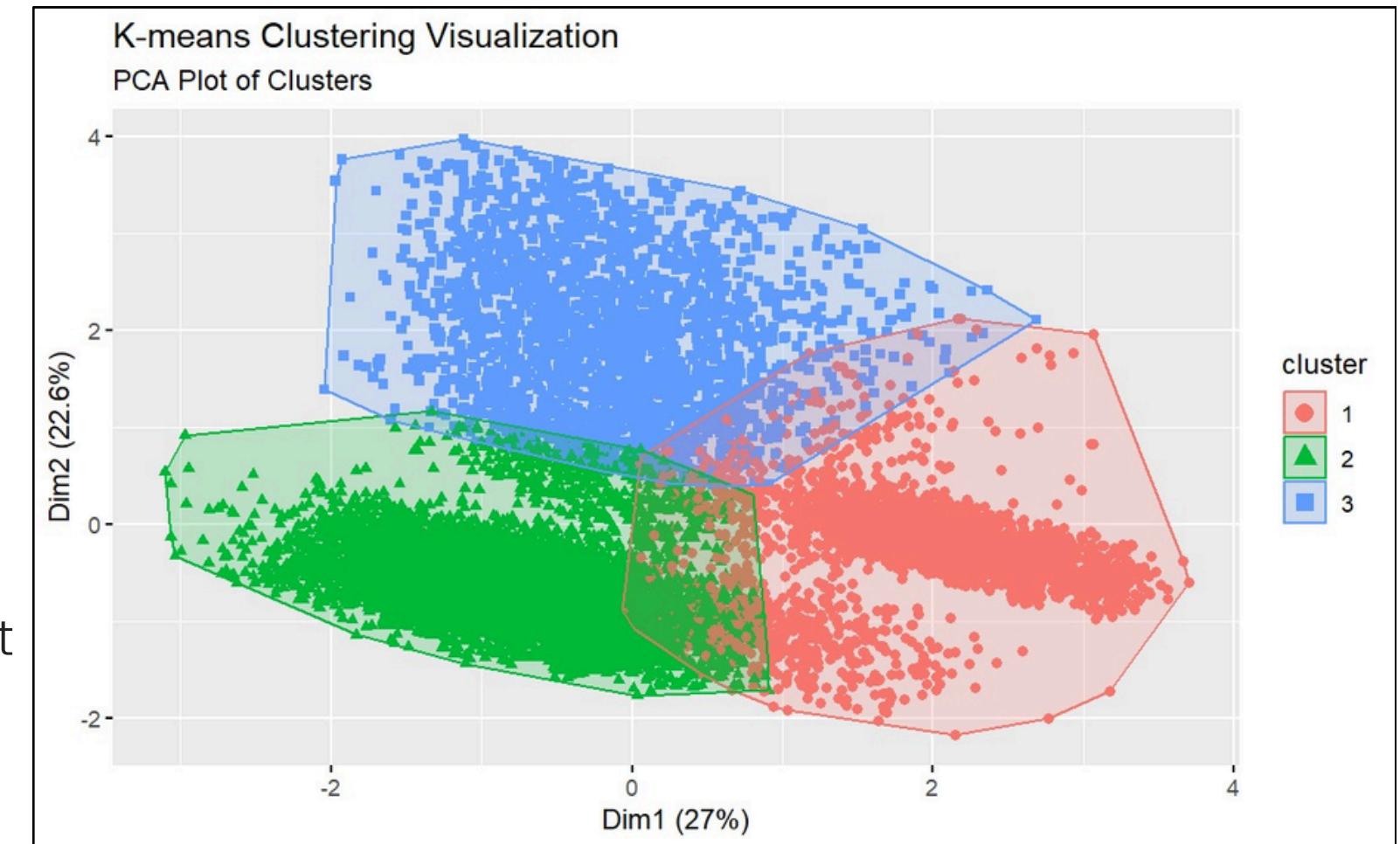
Cluster 1: 2630 points

Cluster 2: 6110 points

Cluster 3: 2259 points

Key Insights:

- Cluster 1: High customer call frequency with high product cost and frequent purchases.
- Cluster 2: Moderate value customers with medium product cost and lower call frequency.
- Cluster 3: Price-sensitive customers with high discount needs.



Feature	Cluster 1	Cluster 2	Cluster 3
Customer_care_calls	High	Low	Low
Customer_rating	Average	Average	Average
Cost_of_the_Product	High	Medium	Low
Prior_purchases	High	Low	Low
Discount_offered	Low	Low	High
Weight_in_gms	Low	High	Low

Understanding Cost Variability Across Clusters

- Segmentation: Grouping customers by cost helps target marketing efforts efficiently.
- Market Positioning: Identifies premium and budget-friendly customer segments.
- Targeted Strategies: Allows tailored marketing and customer positioning for different customer clusters.



Recommendations:

1. Cluster 1: The Favorites

- Treat them with special rewards.
- Make their experience top-notch.

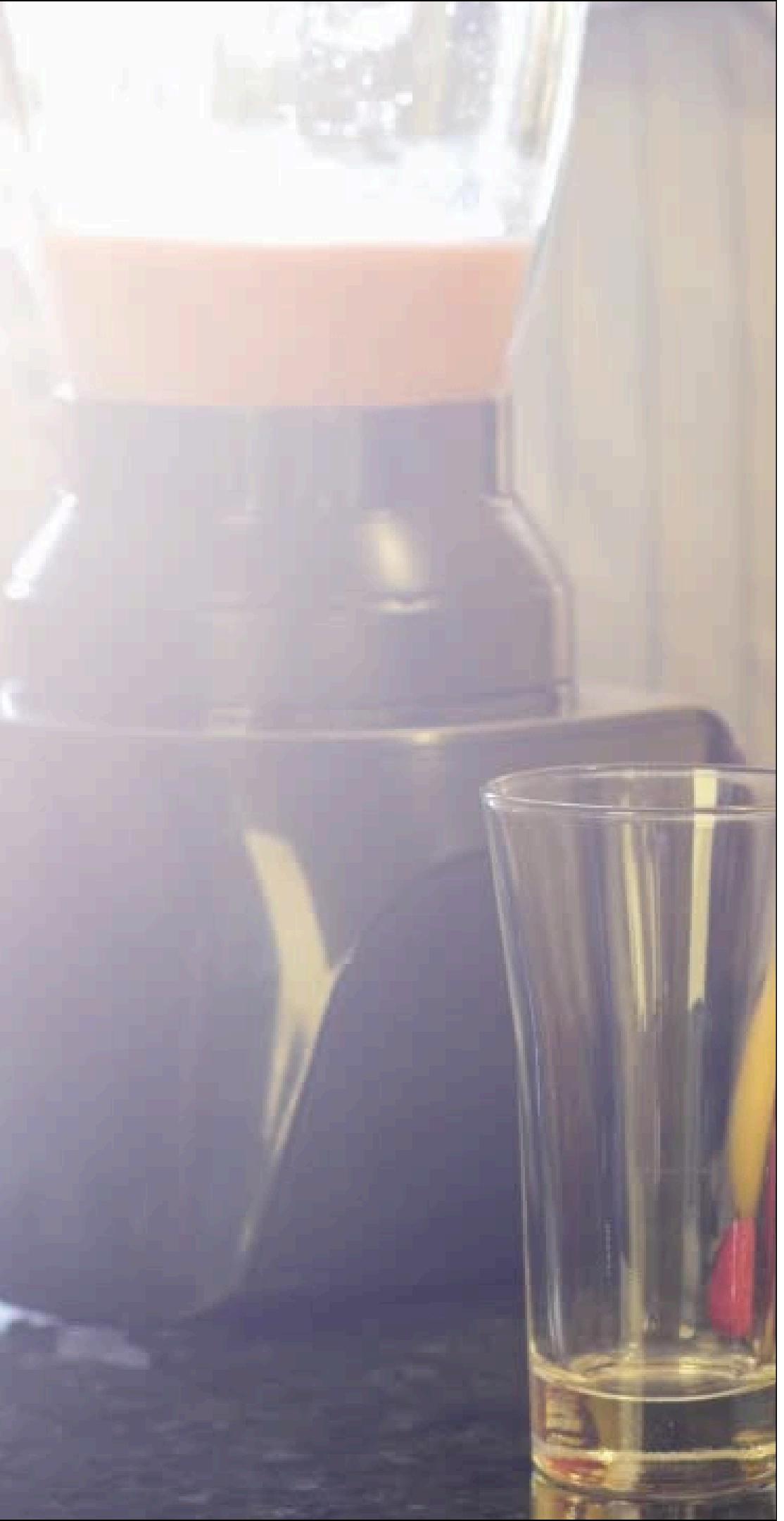
2. Cluster 2: The Regulars

- Keep them satisfied with consistent service.
- Add discounts for extra flavor.

3. Cluster 3: The Explorers

- Learn what they love.
- Offer exciting deals to keep them interested.

Knowing what our customers prefer and what they're willing to spend on our products is key to serving up success!



Serving Satisfaction: Cooking Up Delivery Delights

1. Spice up efficiency:

- Sail through efficiency by storing 33% in Block F for smoother operations!

2. Shipping Secrets:

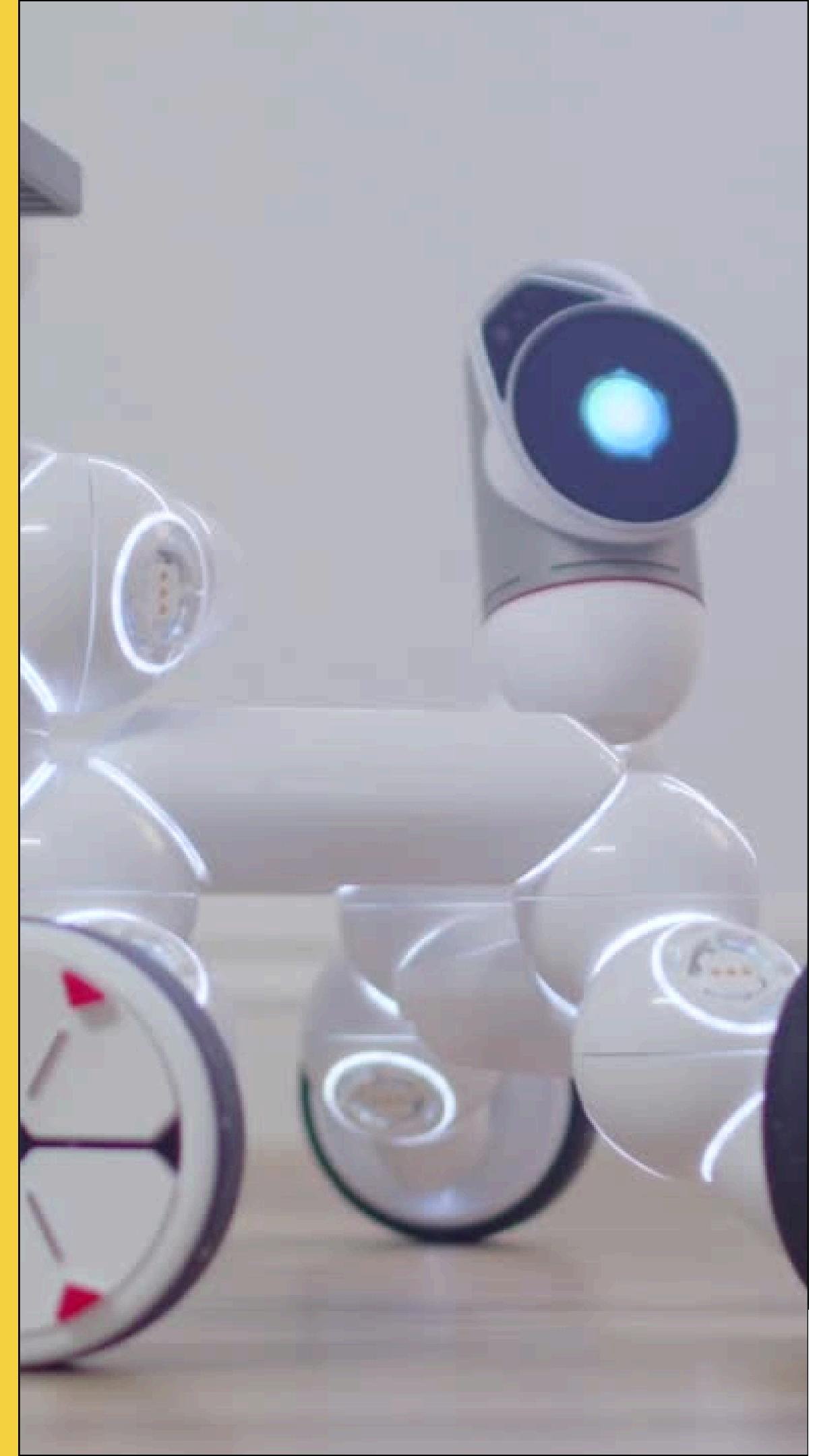
- Sea Adventures: Sail cost-effectively for high satisfaction!
- Air vs. Road: Air zooms, roads may hiccup. Stay flexible for on-time treats!

3. Shipping Strategy:

- Cost Care: Keep expenses low, sprinkle discounts wisely.
- Special Touch: Treat VIP items with care for rave reviews!

4. Customer Care:

- Less Calls: Quick deliveries mean fewer calls.
- Smooth Deliveries: Optimize for a delightful delivery dance!



Thank You!
Any Questions?

