

Data Analysis Report

Generated on: 2025-01-09 17:32:28

Key Findings:

- Dataset contains 1952 records and 25 columns
- Strong upward trend detected in Customer ID
 - Consider investigating outliers in Row ID

Detailed Insights:

- Dataset contains 1952 records and 25 columns
- Found 54 outliers in Row ID
- Found 2 outliers in Discount
- Found 21 outliers in Unit Price
- Found 45 outliers in Shipping Cost
- Found 40 outliers in Profit
- Found 39 outliers in Quantity ordered new
- Found 32 outliers in Sales
- Found 85 outliers in Order ID
- Most common Order Priority: Low (398 occurrences)
- Most common Customer Name: Jenny Gold (9 occurrences)
- Most common Ship Mode: Regular Air (1437 occurrences)
- Most common Customer Segment: Corporate (684 occurrences)
- Most common Product Category: Office Supplies (1071 occurrences)
- Most common Product Sub-Category: Paper (284 occurrences)
- Most common Product Container: Small Box (991 occurrences)
- Most common Product Name: Bevis 36 x 72 Conference Tables (9 occurrences)
- Most common Country: United States (1952 occurrences)
- Most common Region: Central (566 occurrences)
- Most common State or Province: California (214 occurrences)
- Most common City: New York City (47 occurrences)
- Most common Order Date: 4/9/2015 (28 occurrences)
- Most common Ship Date: 1/8/2015 (23 occurrences)
- Strong correlation (0.92) between Row ID and Order ID
- Missing values in Product Base Margin: 16

Recommendations:

- Consider investigating outliers in Row ID
- Consider investigating outliers in Discount
- Consider investigating outliers in Unit Price
- Consider investigating outliers in Shipping Cost
- Consider investigating outliers in Profit
- Consider investigating outliers in Quantity ordered new
- Consider investigating outliers in Sales
- Consider investigating outliers in Order ID
- Consider grouping less frequent categories in Product Sub-Category
- Consider grouping less frequent categories in State or Province
- Consider grouping less frequent categories in City
- Consider grouping less frequent categories in Order Date
- Consider grouping less frequent categories in Ship Date
- Consider handling missing values in Product Base Margin

Trends and Predictions:

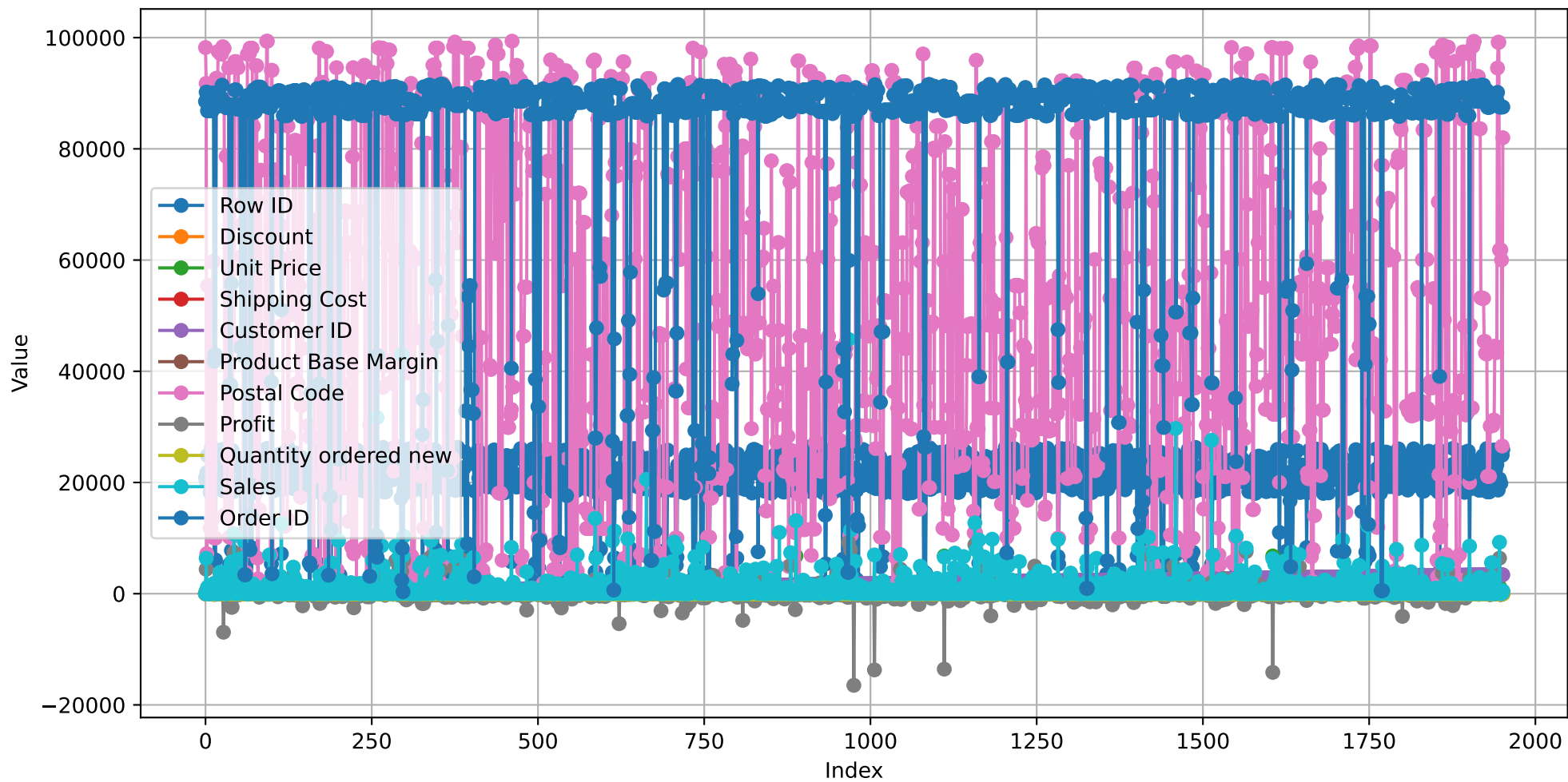
Trends:

- Strong upward trend detected in Customer ID

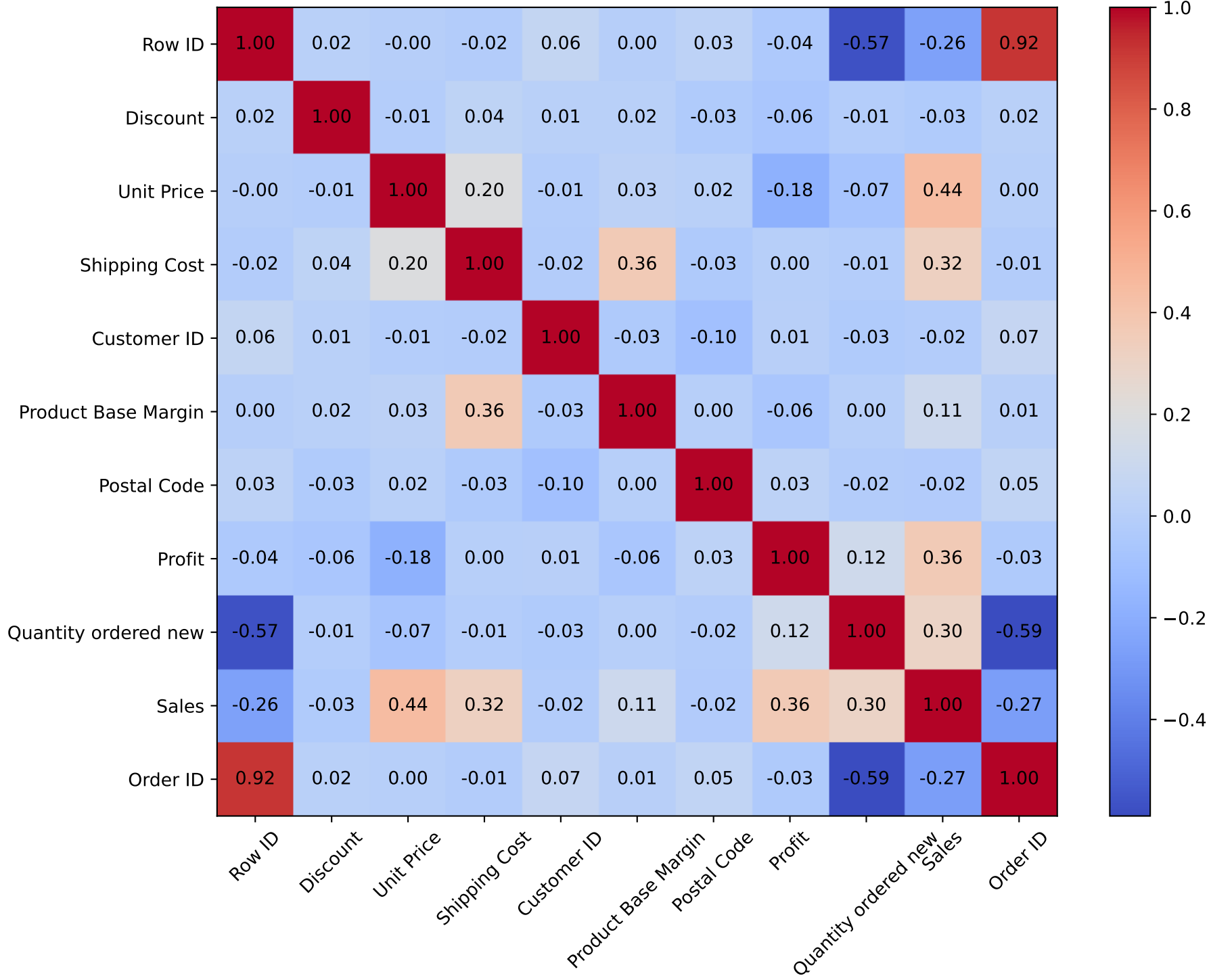
Predictions:

- Predicted next 5 values for Row ID: [20563.53, 20564.2, 20564.86, 20565.52, 20566.18]
- Predicted next 5 values for Discount: [0.05, 0.05, 0.05, 0.05, 0.05]
- Predicted next 5 values for Unit Price: [102.38, 102.38, 102.37, 102.36, 102.36]
- Predicted next 5 values for Shipping Cost: [12.42, 12.42, 12.42, 12.42, 12.42]
- Predicted next 5 values for Customer ID: [3451.97, 3453.73, 3455.49, 3457.25, 3459.0]
- Predicted next 5 values for Postal Code: [46725.64, 46720.71, 46715.79, 46710.86, 46705.94]
- Predicted next 5 values for Profit: [132.79, 132.8, 132.82, 132.84, 132.86]
- Predicted next 5 values for Quantity ordered new: [12.32, 12.31, 12.31, 12.31, 12.31]
- Predicted next 5 values for Sales: [914.95, 914.88, 914.81, 914.74, 914.66]
- Predicted next 5 values for Order ID: [84606.91, 84609.2, 84611.5, 84613.79, 84616.09]

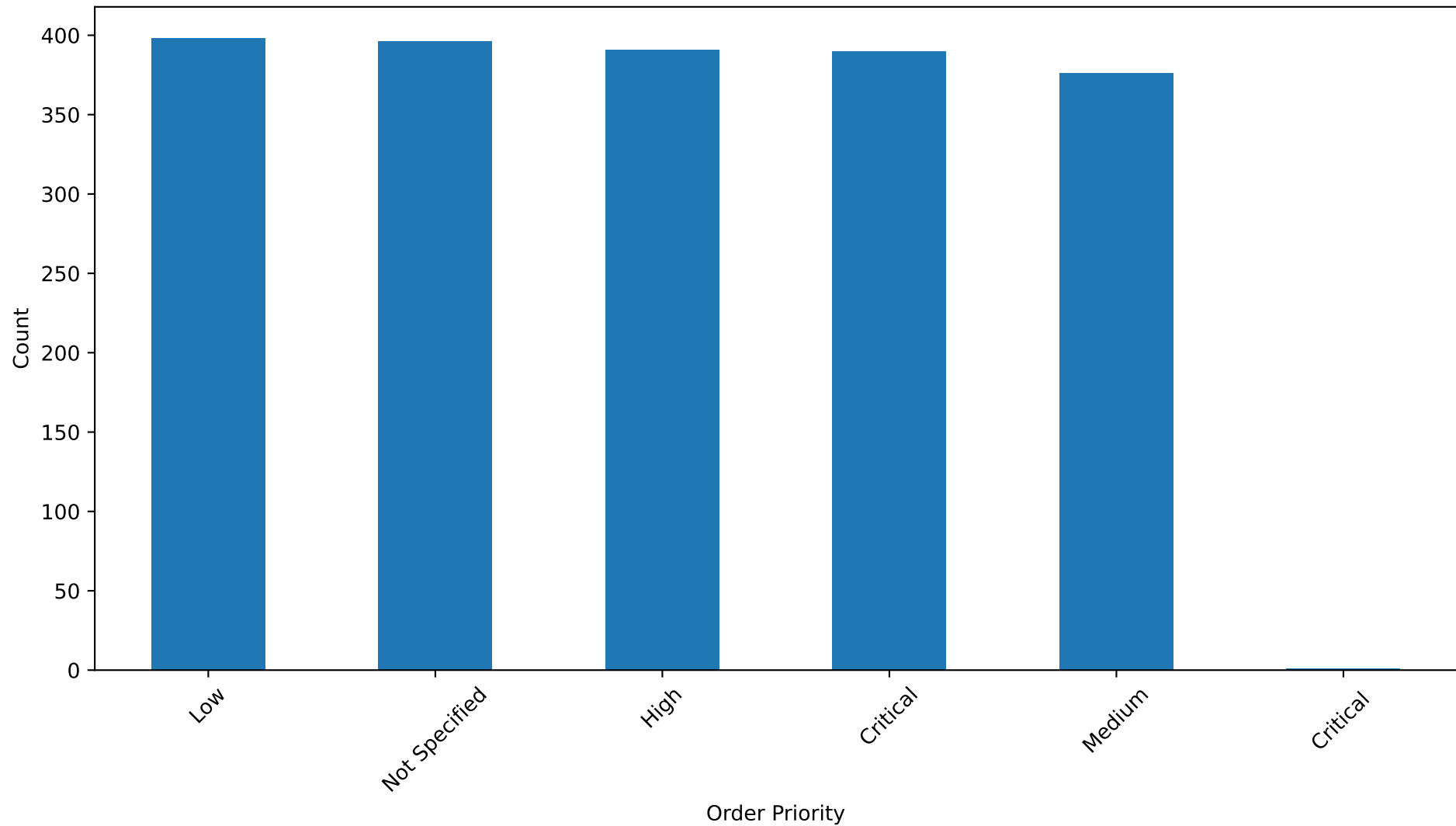
Numeric Data Trends



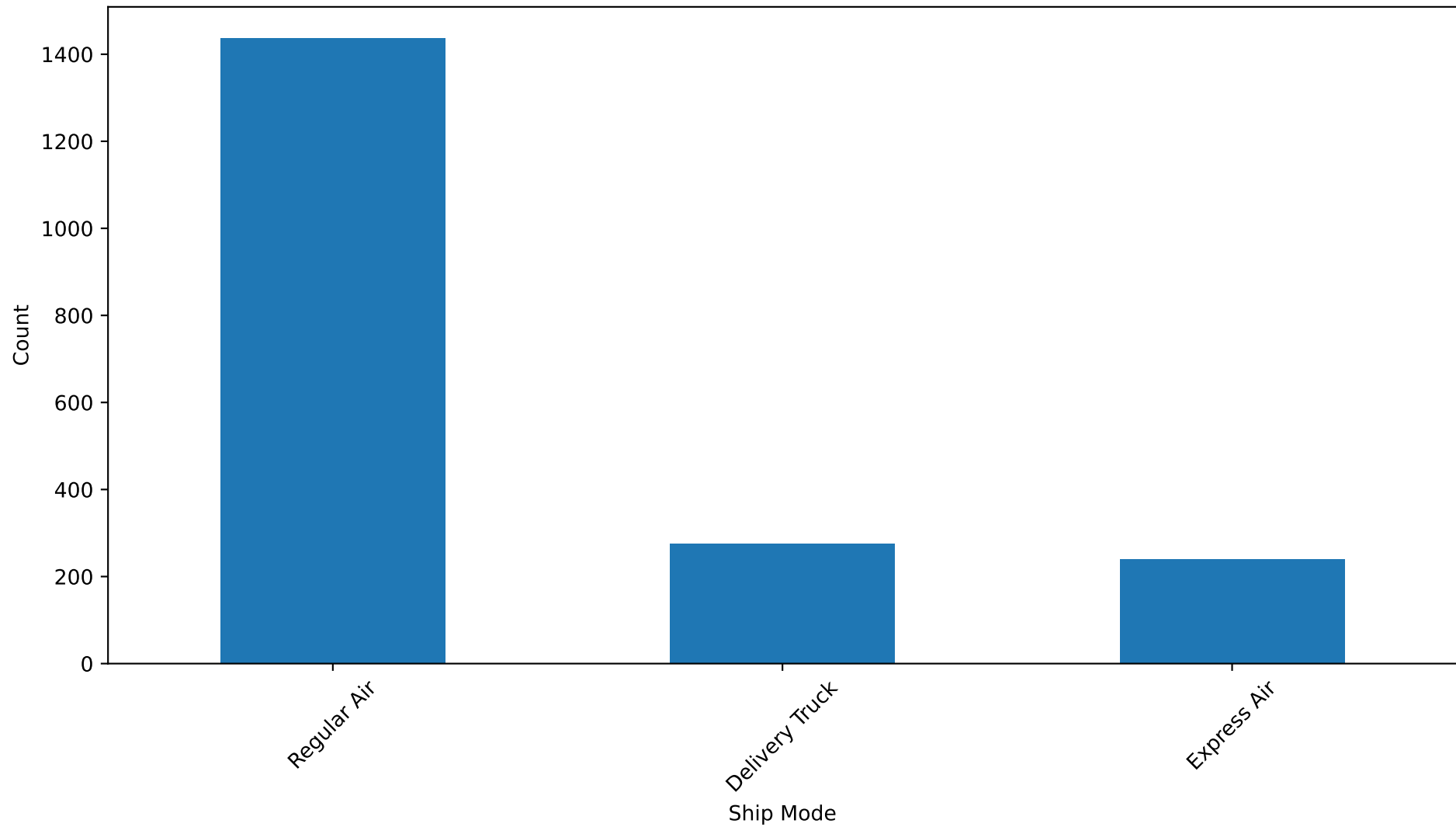
Correlation Matrix



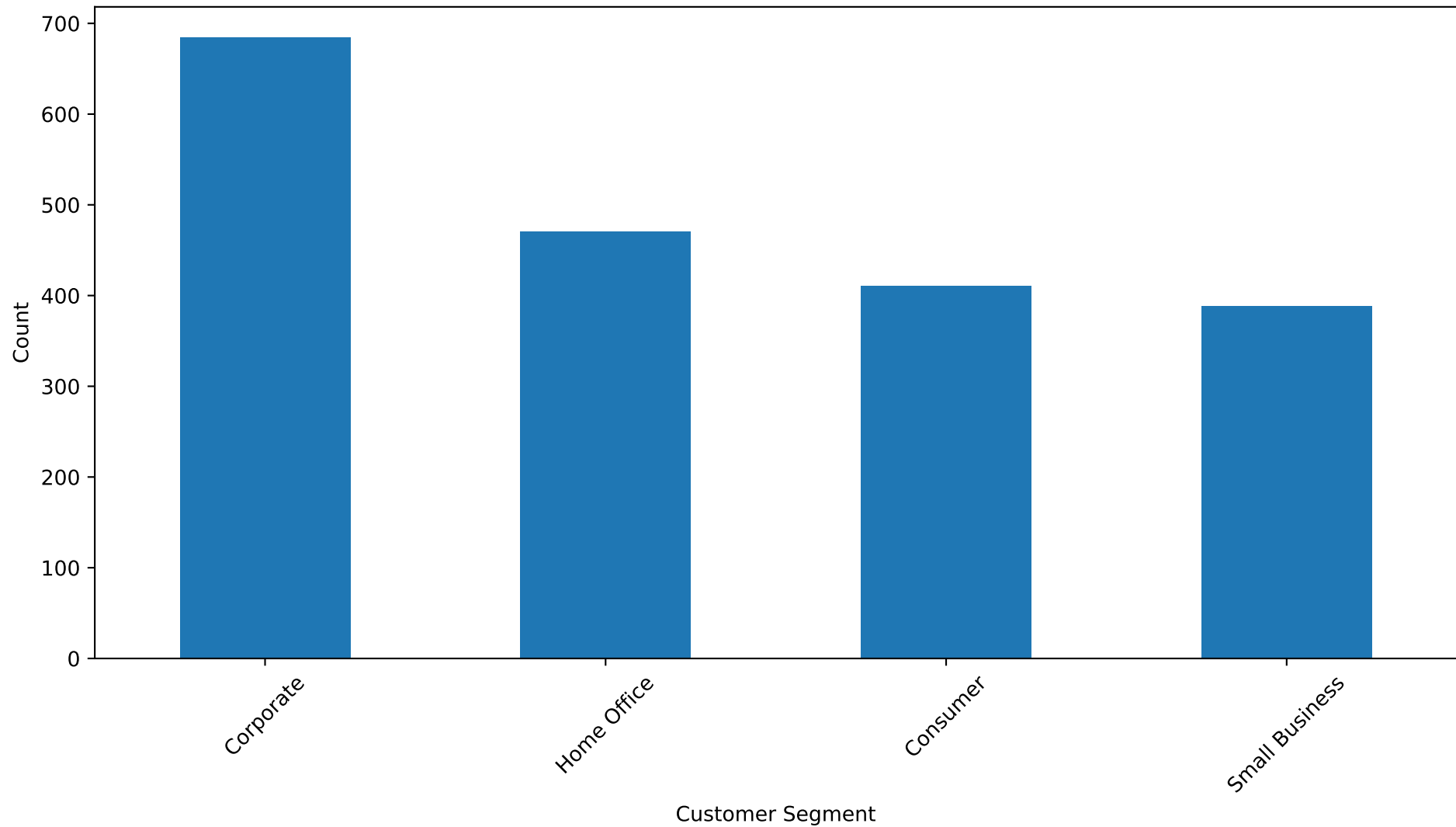
Distribution of Order Priority



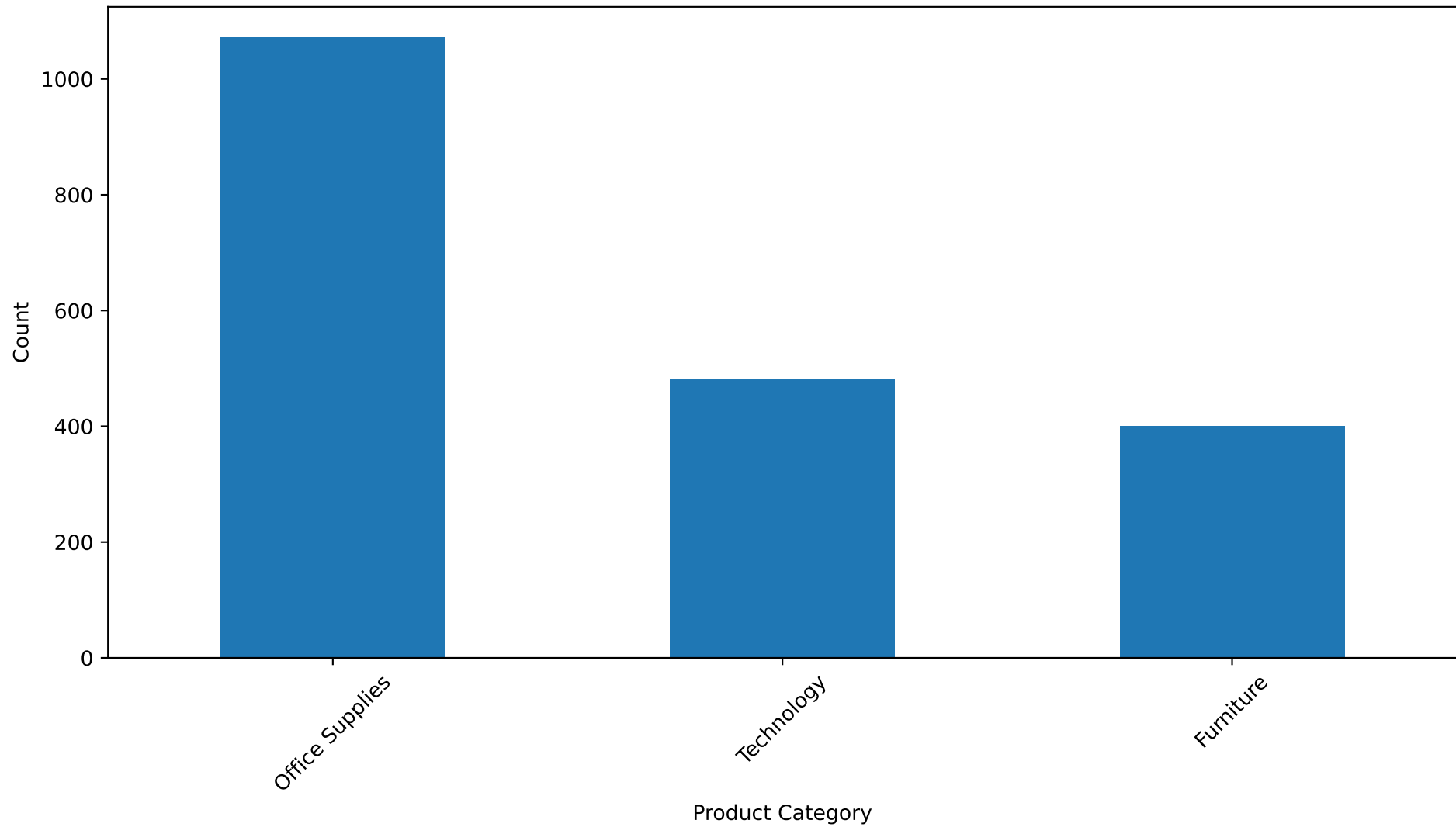
Distribution of Ship Mode



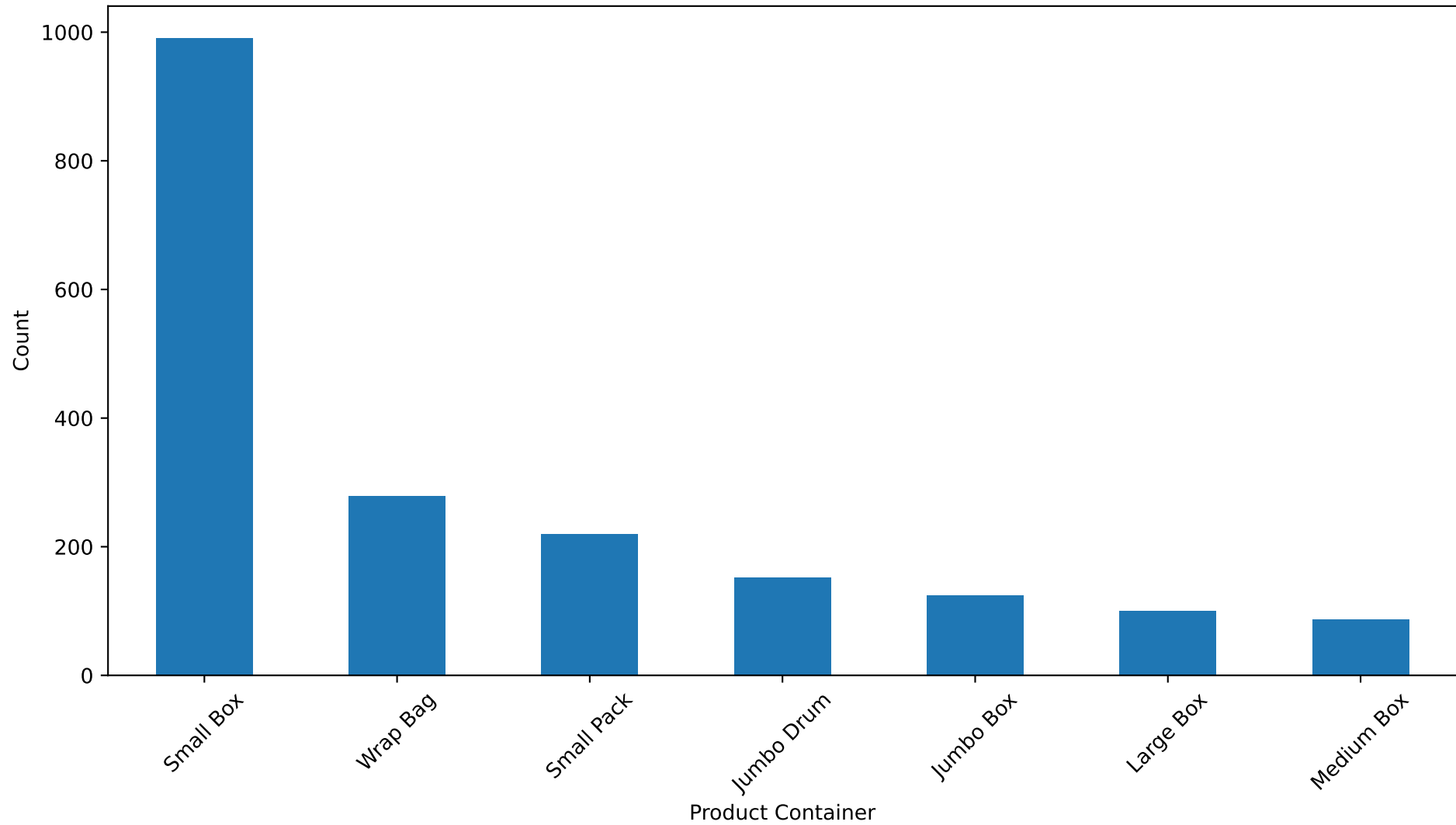
Distribution of Customer Segment



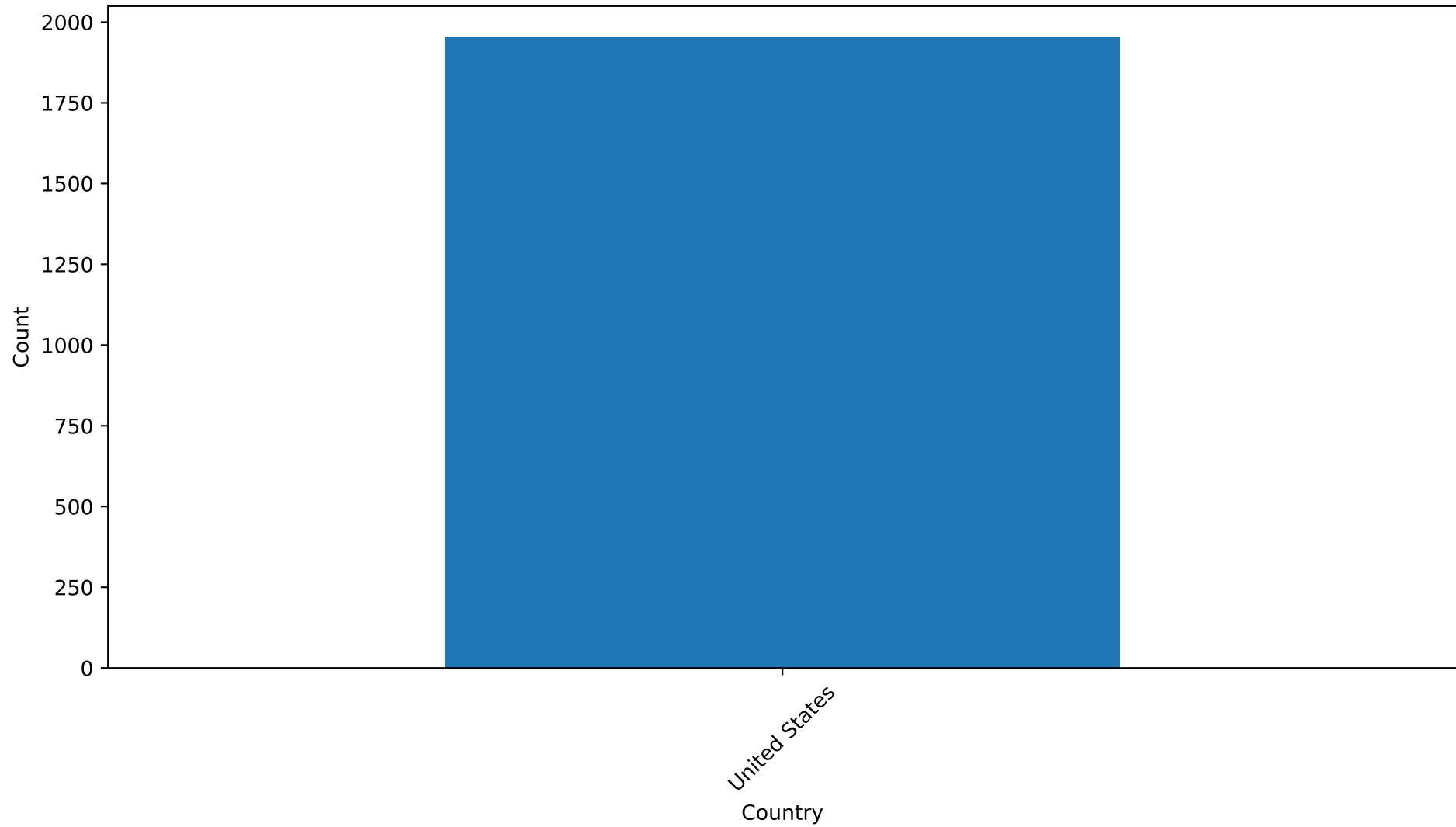
Distribution of Product Category



Distribution of Product Container



Distribution of Country



Distribution of Region

