Marketing
&
Retail Analytics

Prepared By: Agnes Raja Kumari E

PGPDSBA.O.MAY24.A

Customer Segmentation for an Automobile Parts Manufacturer

Business Context

The company has been selling automobile parts for 3 years but lacks data-driven insights.

Need to identify purchasing patterns, segment customers, and improve marketing strategies.

Expected outcomes: Improved customer satisfaction, targeted promotions, and revenue growth.

Understanding the Dataset

Dataset	Total Sales	Total Number of Orders	Unique Customers
Sales_Data	$ 9760222	298	89

Summary statistics

	ORDERNUMBER	QUANTITYORDERED	PRICEEACH	ORDERLINENUMBER	SALES	ORDERDATE	DAYS_SINCE_LASTORDER	MSRP
count	2747.000000	2747.000000	2747.000000	2747.000000	2747.000000	2747	2747.000000	2747.000000
mean	10259.761558	35.103021	101.098951	6.491081	3553.047583	2019-05-13 21:56:17.211503360	1757.085912	100.691664
min	10100.000000	6.000000	26.880000	1.000000	482.130000	2018-01-06 00:00:00	42.000000	33.000000
25%	10181.000000	27.000000	68.745000	3.000000	2204.350000	2018-11-08 00:00:00	1077.000000	68.000000
50%	10264.000000	35.000000	95.550000	6.000000	3184.800000	2019-06-24 00:00:00	1761.000000	99.000000
75%	10334.500000	43.000000	127.100000	9.000000	4503.095000	2019-11-17 00:00:00	2436.500000	124.000000
max	10425.000000	97.000000	252.870000	18.000000	14082.800000	2020-05-31 00:00:00	3562.000000	214.000000
std	91.877521	9.762135	42.042548	4.230544	1838.953901	NaN	819.280576	40.114802

Average Order Quantity: 35 units per order, with a minimum of 6 and a maximum of 97.

Price per Unit: Ranges from $26.88 to $252.87, with an average of $101.99.

Total Sales: The mean sales value per order is $3,553, but there is high variability (std = $1,838).

Order Line Number: Most orders have around 6 line items, but some go up to 18.

Days Since Last Order: Shows a high standard deviation (819 days), suggesting irregular ordering patterns.

MSRP (Manufacturer's Suggested Retail Price): Ranges from $33 to $214, with a mean of $100.69.

Defining the Problem Statement

Identify distinct customer groups based on purchasing behavior using RFM analysis to optimize marketing efforts.

Detect customers at risk of churning and develop targeted engagement strategies to improve retention.

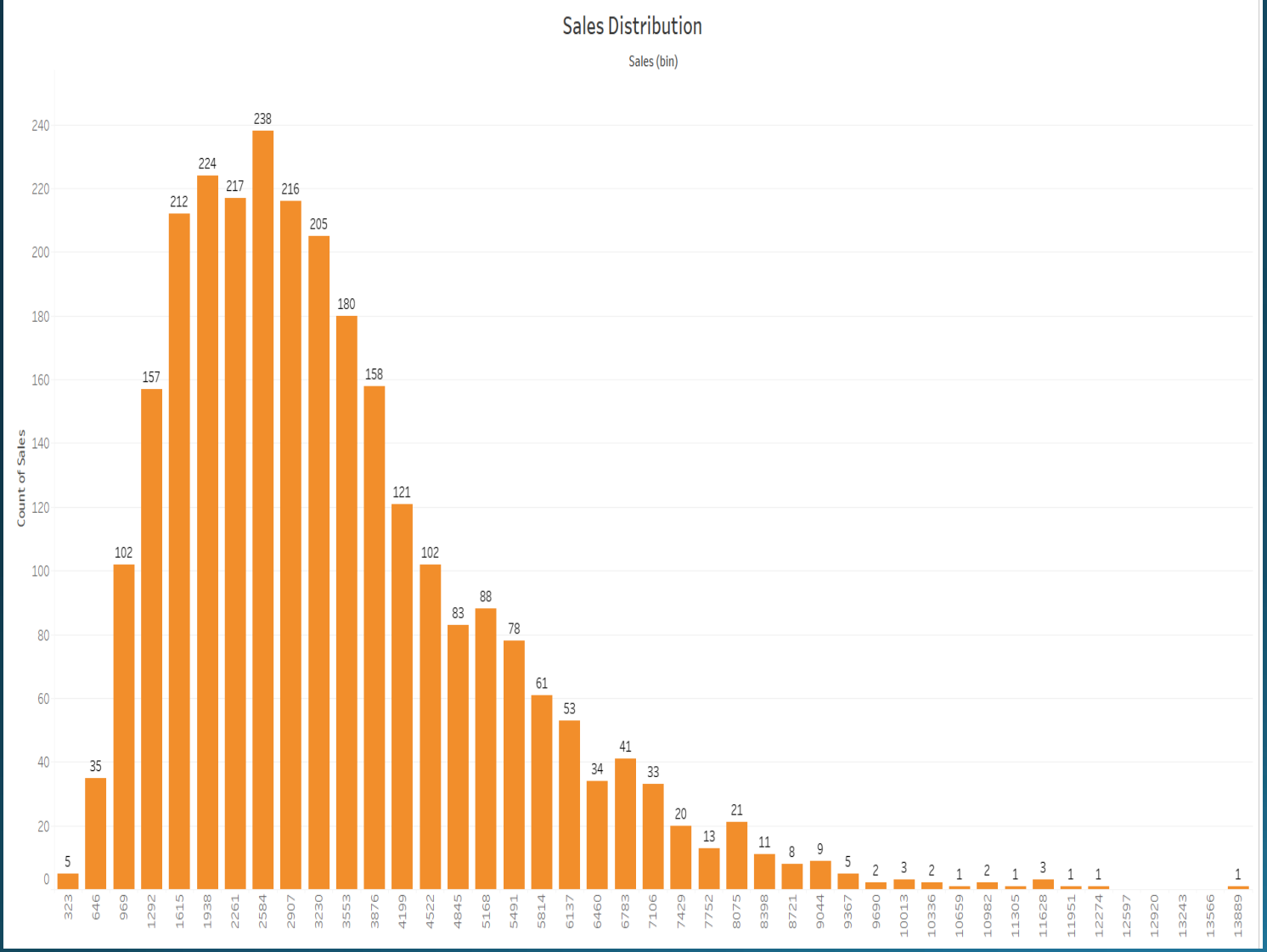
Identify the most valuable customers and recommend personalized strategies to enhance loyalty and maximize revenue.

Analyze sales trends over time (weekly, monthly, quarterly) to uncover seasonality and demand patterns.

Determine which customer segments and products contribute the most to overall profitability.

Leverage transactional data to create targeted marketing campaigns for different customer segments.

Sales Distribution



The peak sales range is around 2,500 - 3,000, where the highest frequency (238 occurrences) is observed.

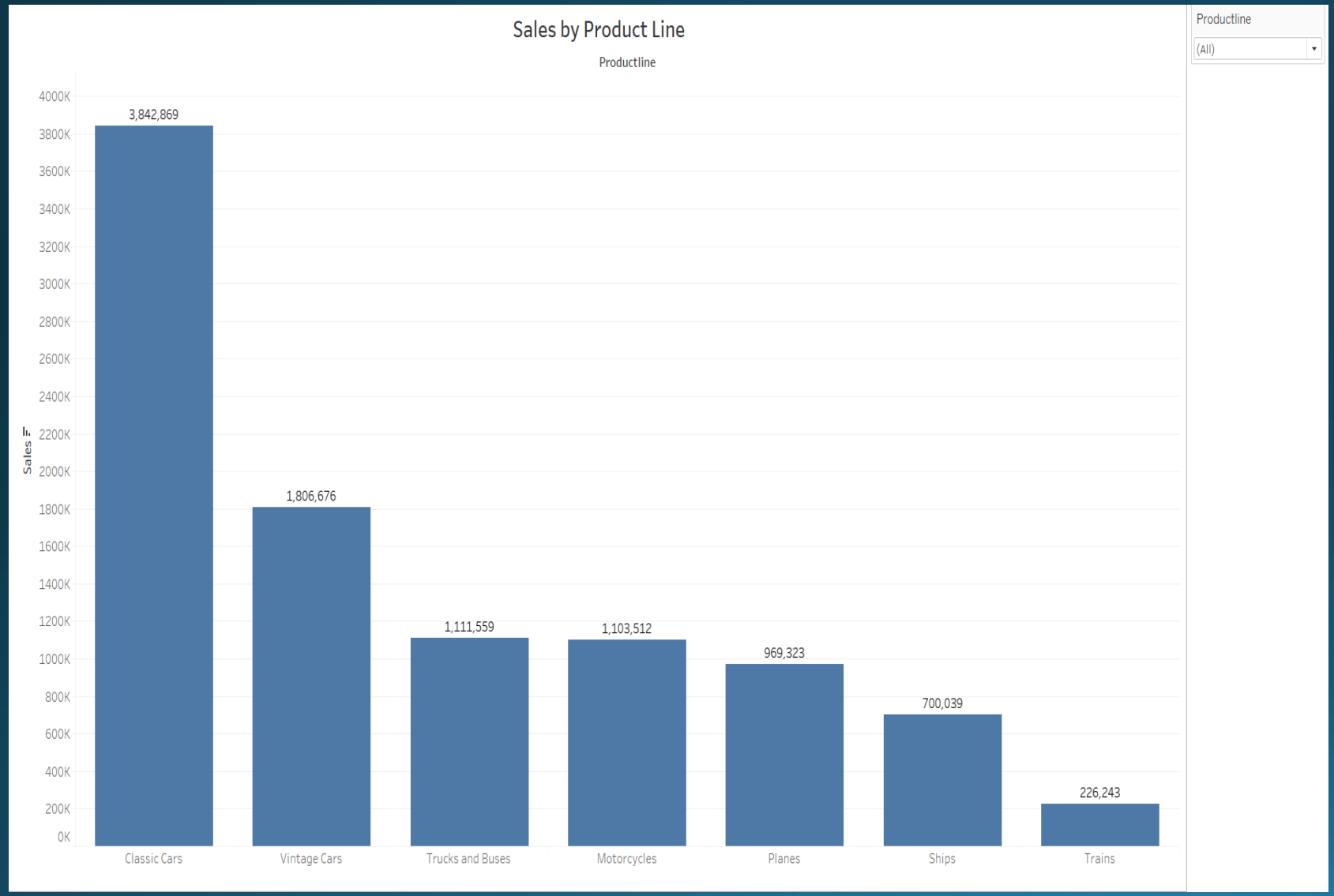
The majority of sales transactions fall within the 0 - 5,000 range, contributing to most of the sales volume.

As sales values increase, the frequency of occurrences decreases significantly, highlighting a long tail effect in sales distribution.

The presence of outliers in the high-sales segment suggests a small number of exceptionally high-value transactions.

Understanding this distribution can help in customer segmentation, inventory management, and targeted marketing strategies to focus on the most frequent sales range.

Sales by Product Line



Classic Cars are the top-performing product line, generating the highest sales revenue of 3.84M, significantly outperforming other categories.

Vintage Cars rank second with 1.81M in sales, indicating strong demand in the collector's market.

Trucks and Buses and Motorcycles have nearly identical sales figures, around 1.1M, suggesting balanced demand in these segments.

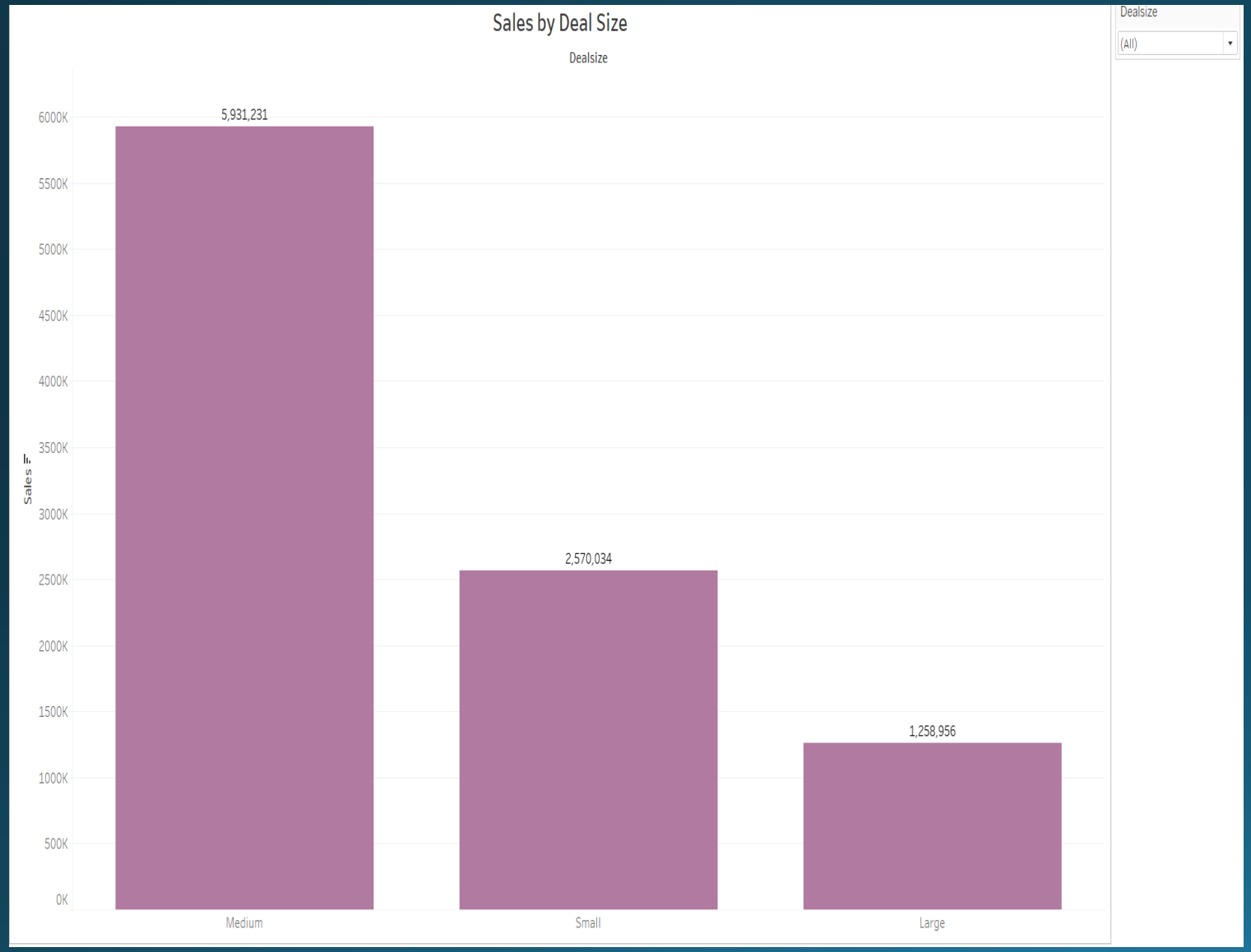
Planes and Ships contribute moderately, with sales of 969K and 700K, respectively.

Trains have the lowest sales at 226K, indicating a niche market or limited product offerings.

The data highlights Classic and Vintage Cars as key revenue drivers, suggesting a focus on expanding marketing and production in these segments.

Diversification strategies may be needed for lower-performing categories like Trains and Ships to improve sales.

Sales by Deal Size



Medium-sized deals contribute the highest sales revenue, totaling 5.93M, indicating that mid-range transactions drive the majority of the business.

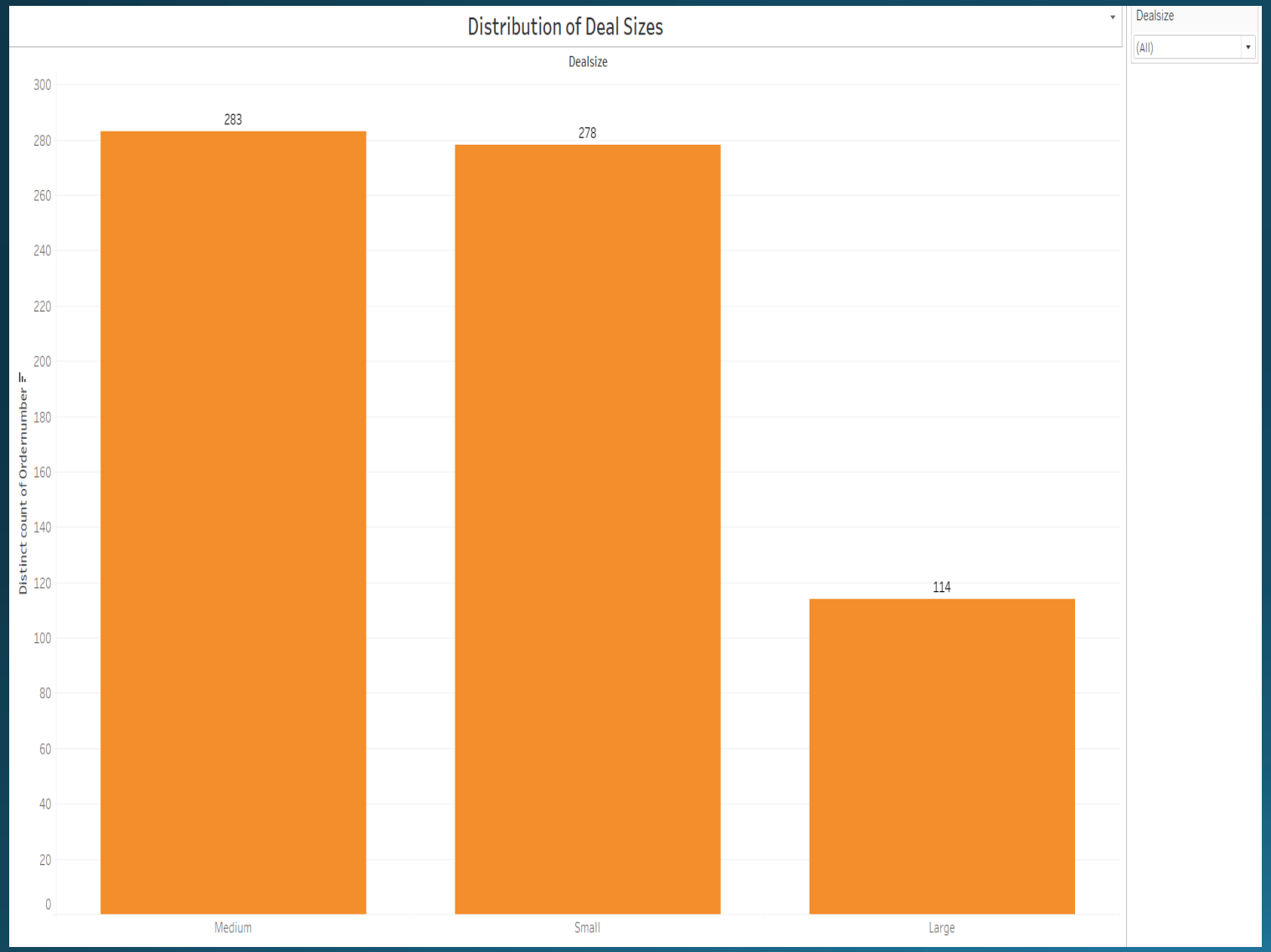
Small-sized deals account for 2.57M in sales, showing a significant contribution but less than half of medium-sized deals.

Large-sized deals generate the lowest revenue at 1.26M, suggesting that fewer high-value transactions occur compared to smaller deals.

The data indicates a strong reliance on medium-sized deals, meaning optimizing strategies for this segment could further enhance revenue.

The relatively low sales in large deals suggest potential challenges, such as high negotiation times, pricing concerns, or limited market demand.

Distribution of Deal Sizes



Medium-sized deals are the most common, with 283 orders, slightly higher than small deals.

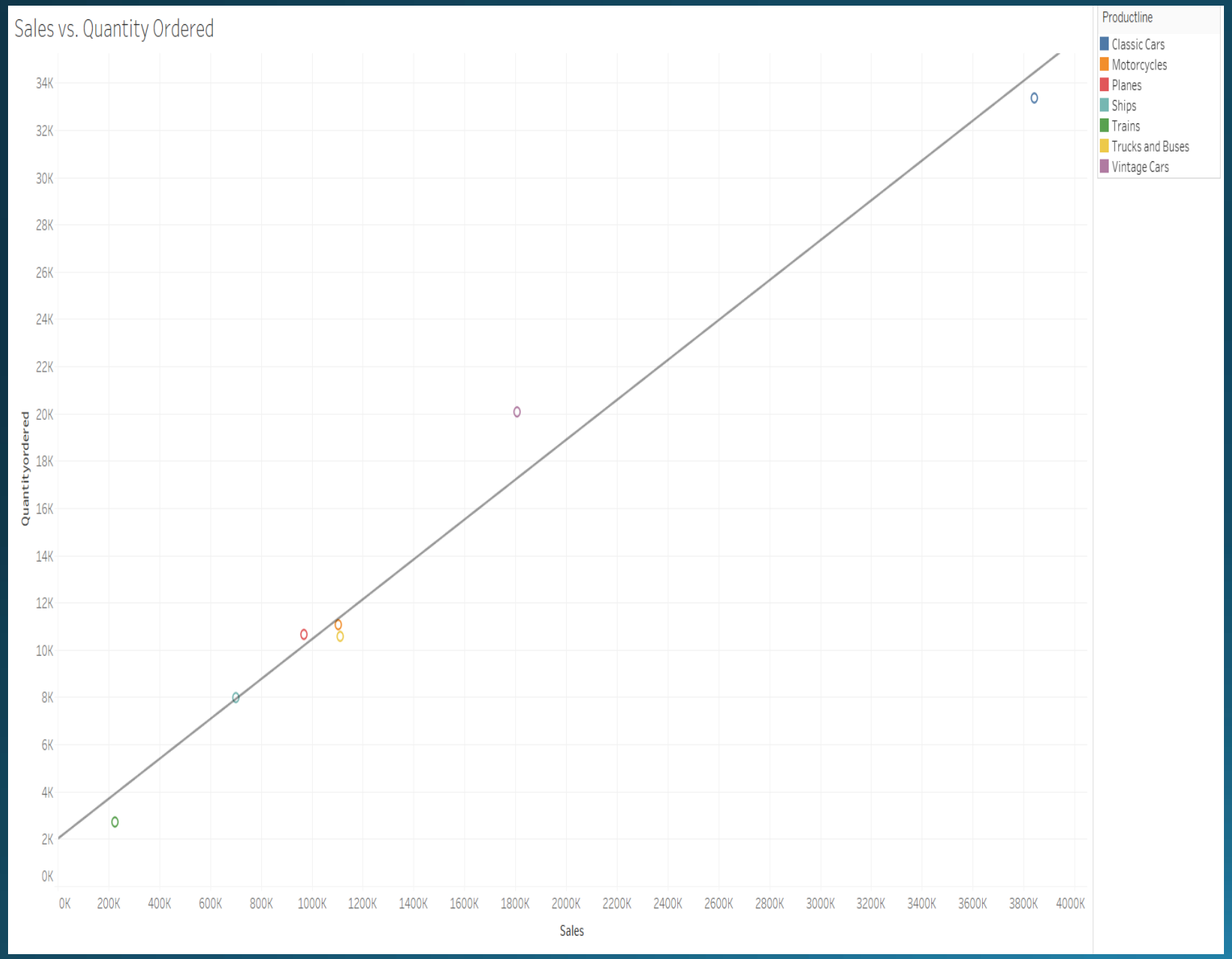
Small-sized deals follow closely behind, with 278 orders, showing a balanced distribution between medium and small deals.

Large-sized deals are significantly fewer, with only 114 orders, indicating that high-value transactions are less frequent.

The similarity in order count between medium and small deals suggests that the company caters primarily to mid-tier and smaller transactions.

The low number of large deals aligns with the lower total revenue observed in the previous sales chart, reinforcing the need to analyze barriers to securing more large deals.

Sales Vs. Quantity Ordered



There is a strong positive correlation between sales and quantity ordered, as indicated by the trend line. This suggests that higher sales are directly associated with a higher number of units ordered.

Classic Cars appear to have the highest sales and quantity ordered, indicating strong demand for this product line.

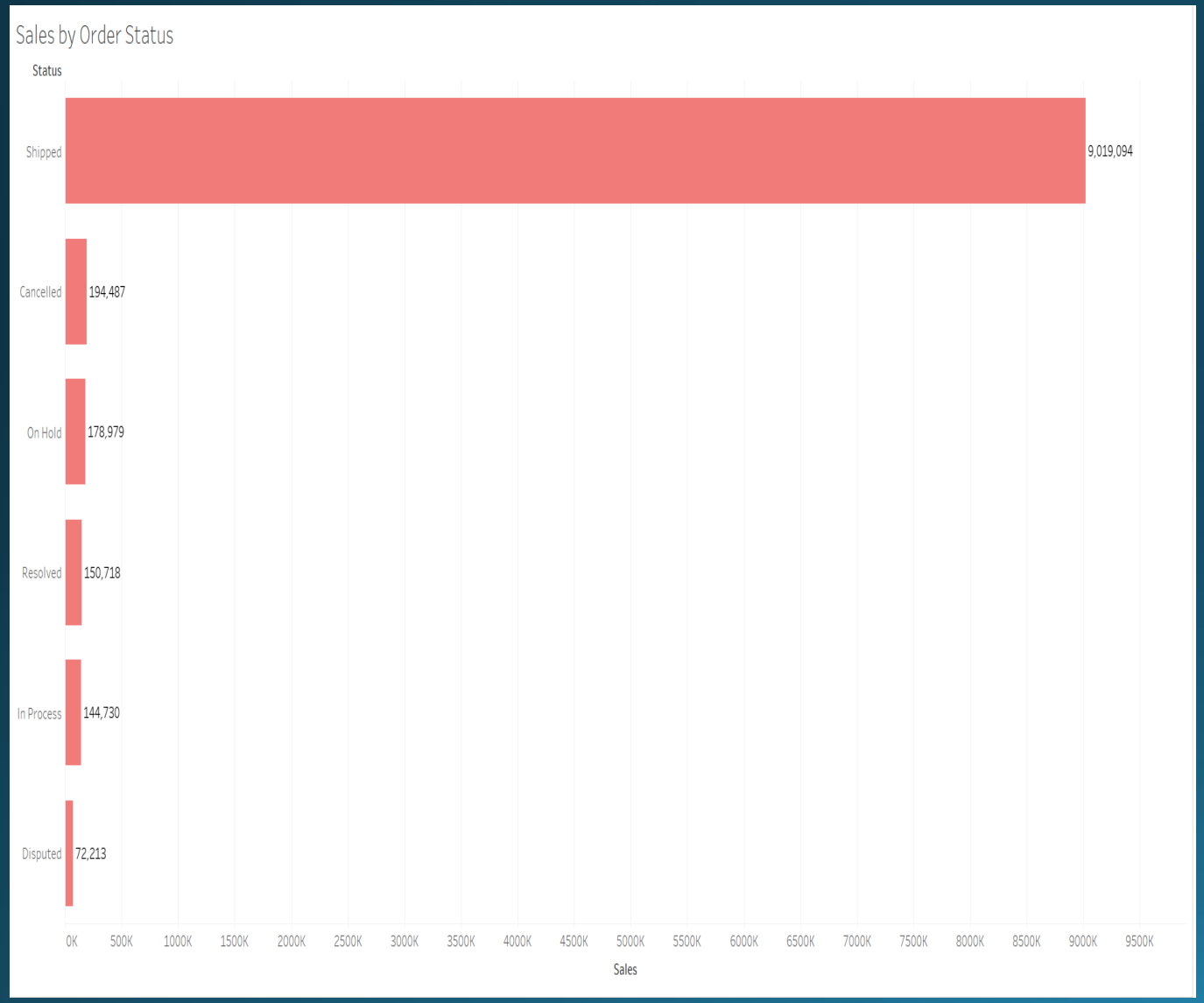
Vintage Cars also have significant sales volume, suggesting they are a key revenue driver.

Motorcycles, Planes, and Trucks/Buses are clustered in a similar sales range, indicating moderate demand.

Ships and Trains have relatively lower sales and quantity ordered, suggesting niche or lower demand for these product lines.

The presence of outliers or deviations from the trend line could indicate products with higher price points but lower quantities sold, or bulk sales at lower prices.

Sales By Order Status



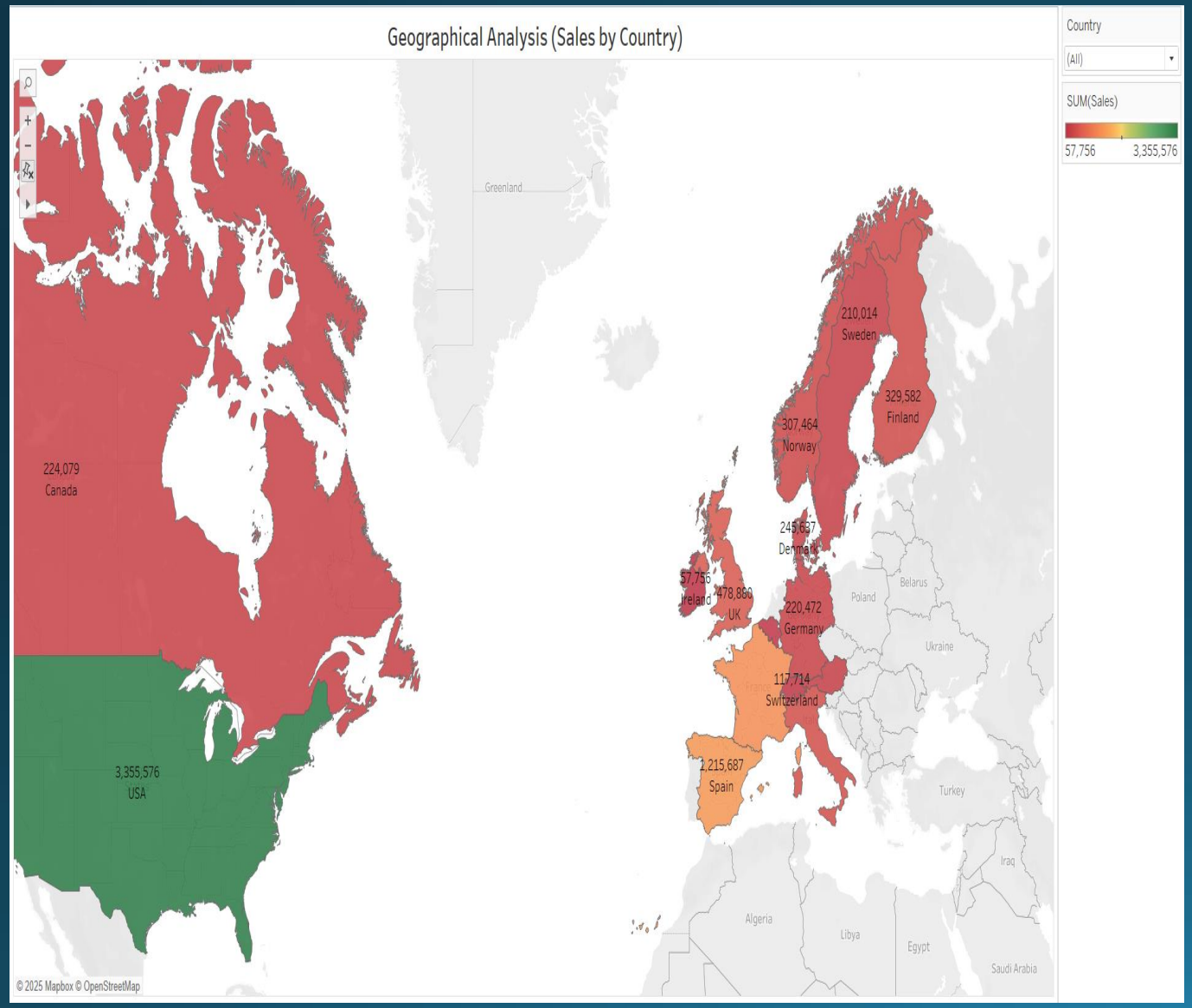
Shipped orders contribute to ~97% of total sales ($9M+), indicating strong fulfillment efficiency.

Cancelled orders ($194K) highlight potential revenue loss—root cause analysis needed to reduce cancellations.

On Hold ($179K) and In Process ($144K) orders indicate pending revenue—faster processing can improve cash flow.

Disputed orders ($72K) suggest customer dissatisfaction—addressing root causes can enhance customer trust.

Sales by Country



USA leads in sales ($3.85M), making it the dominant market.

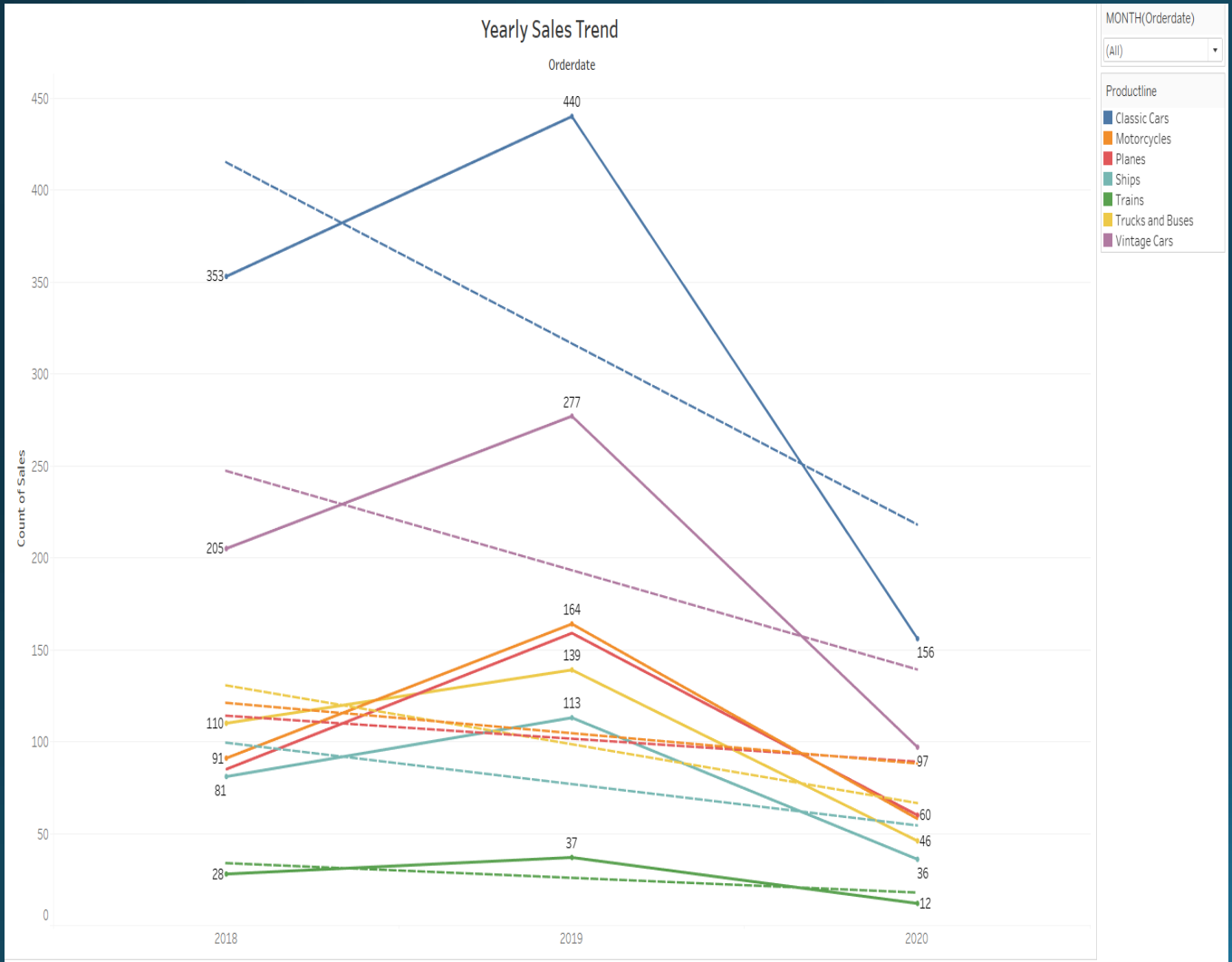
Spain ($2.22M) and the UK ($720K) are strong secondary markets.

Germany ($620K) and Norway ($307K) contribute moderately, indicating potential for growth.

Canada ($224K) and Sweden ($210K) have lower sales—requires further investigation into market barriers.

Focus on high-performing regions for expansion and optimize strategies in low-sales regions.

Yearly Sales Trend



Sales peaked in 2019, with a noticeable decline in 2020 across all product categories.

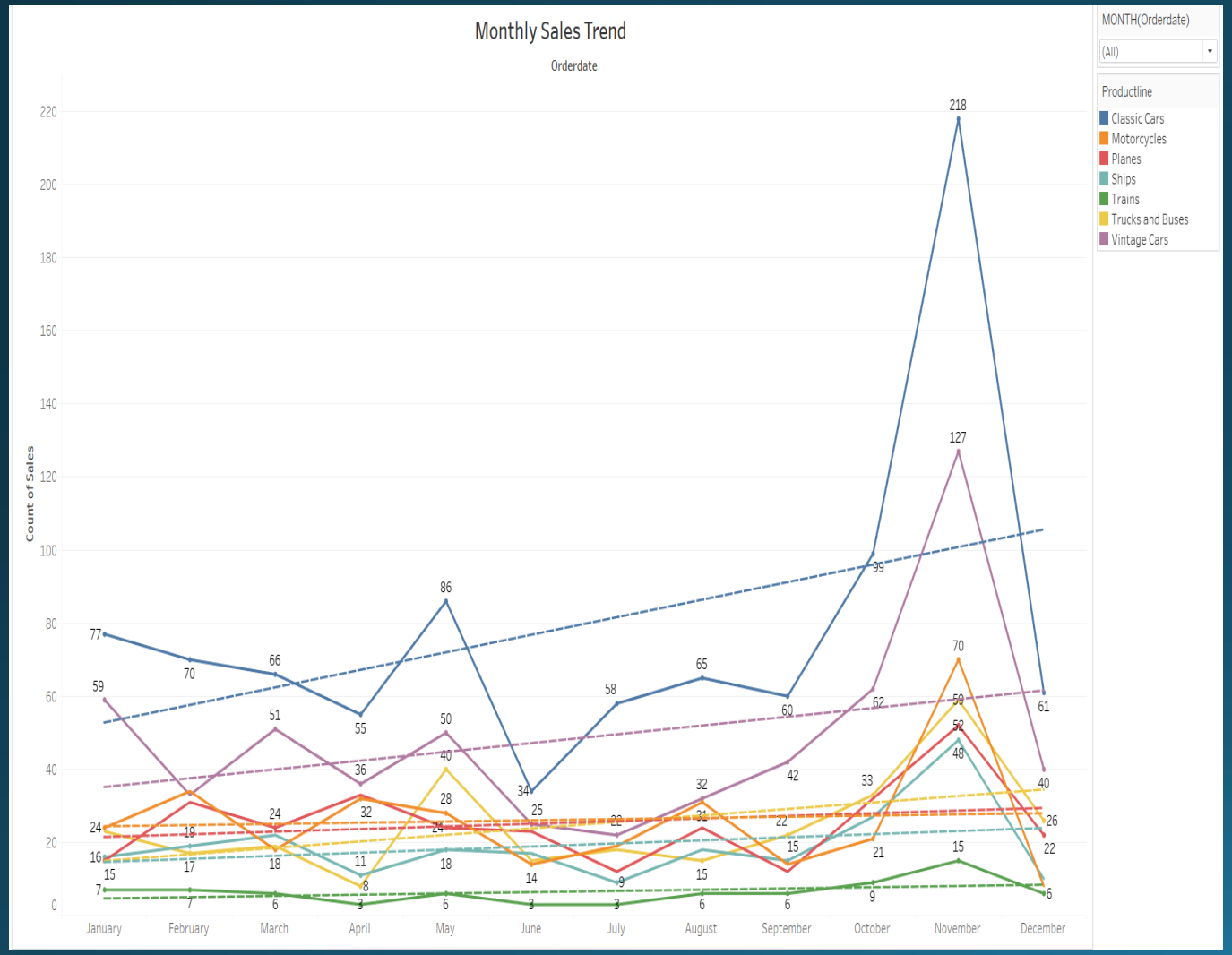
Classic Cars had the highest sales but showed a sharp drop in 2020.

Motorcycles and Ships saw moderate growth in 2019 but declined in 2020.

Market trends suggest a potential impact of external factors in 2020, requiring deeper investigation.

Focus on strategies to recover sales, particularly in high-performing product categories.

Monthly Sales Trend



Seasonal Spike in December

- Classic Cars experienced a huge spike in December (218 sales), significantly higher than other months.
- Vintage Cars also saw a peak in November (127 sales) before dropping in December.

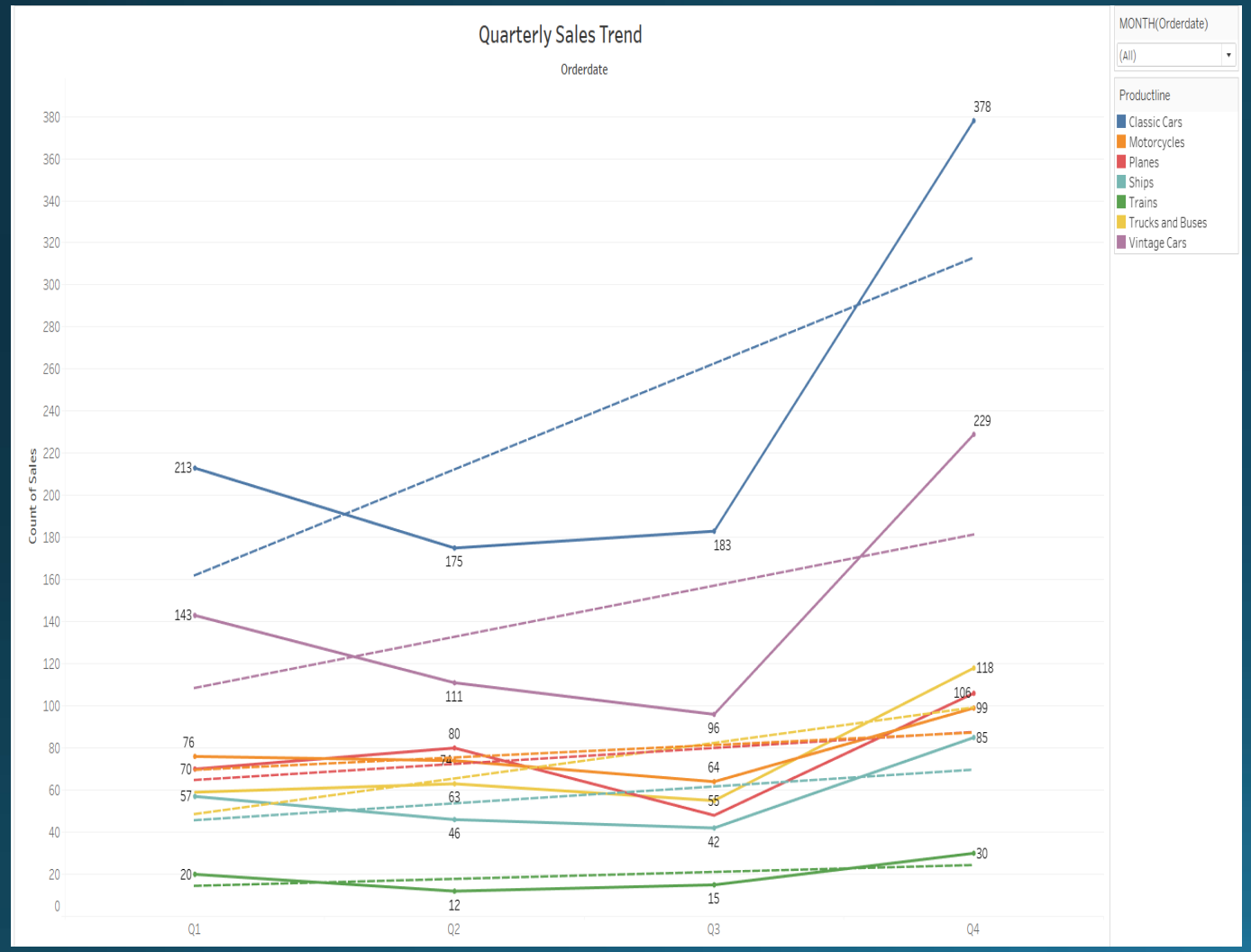
Steady Growth Throughout the Year

- Classic Cars show a gradual increasing trend over the months, with fluctuations.
- Vintage Cars and Motorcycles follow a similar trend but at a lower volume.

Stable but Low Sales for Other Product Lines

- Trucks and Buses, Trains, and Ships show minimal variation in monthly sales.
- Planes and Motorcycles have small peaks in November but remain relatively stable.

Quarterly Sales Trend



Strong Q4 Surge

- Classic Cars saw the highest sales in Q4 (378 sales), highlighting strong year-end demand.
- Vintage Cars also peaked in Q4 (229 sales) after fluctuating in earlier quarters.

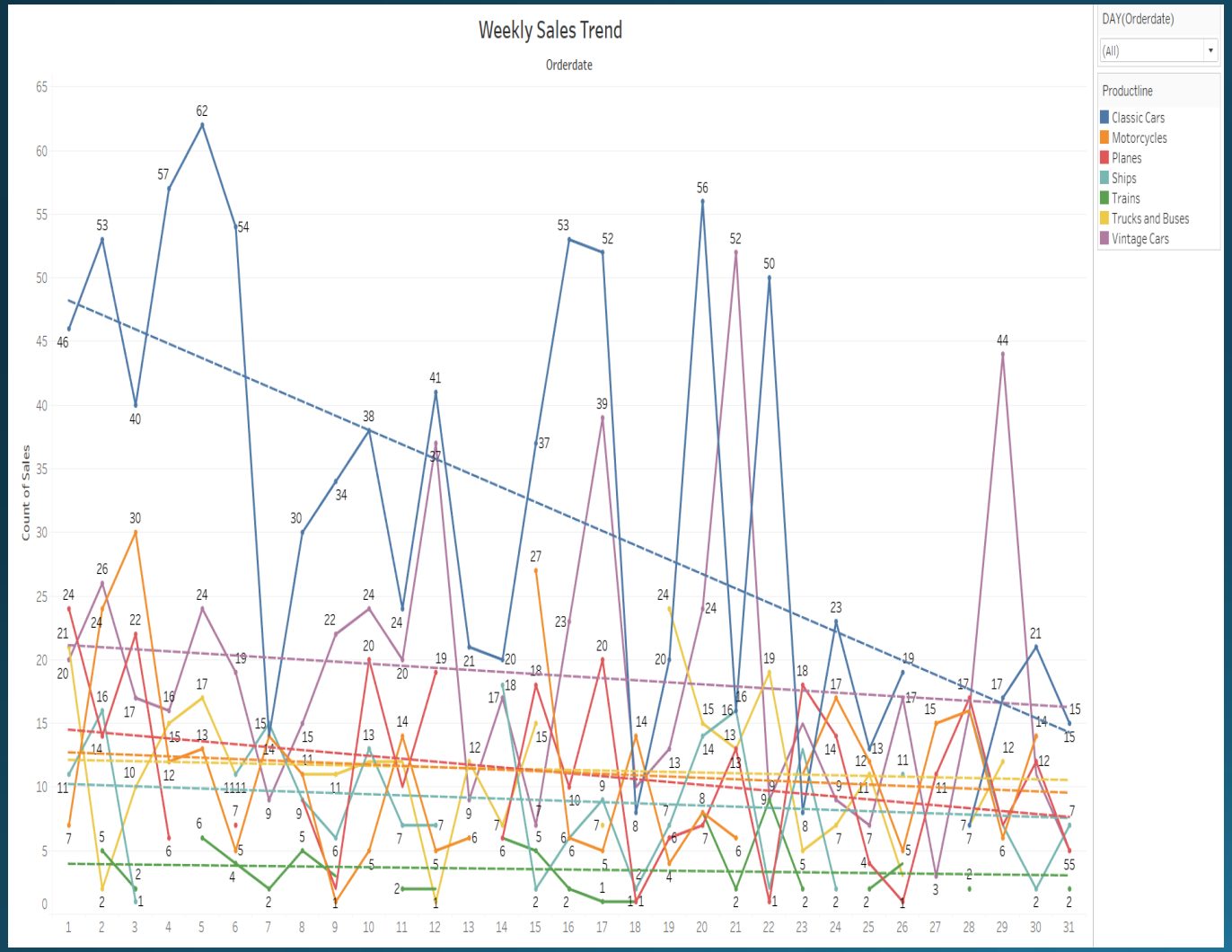
Steady Growth Across Quarters

- Classic Cars had a dip in Q2 but rebounded strongly in Q3 and Q4.
- Motorcycles and Trucks & Buses showed consistent growth, peaking in Q4.

Minimal Variation for Other Categories

- Trains, Ships, and Planes maintained relatively stable but lower sales across all quarters.

Weekly Sales Trend



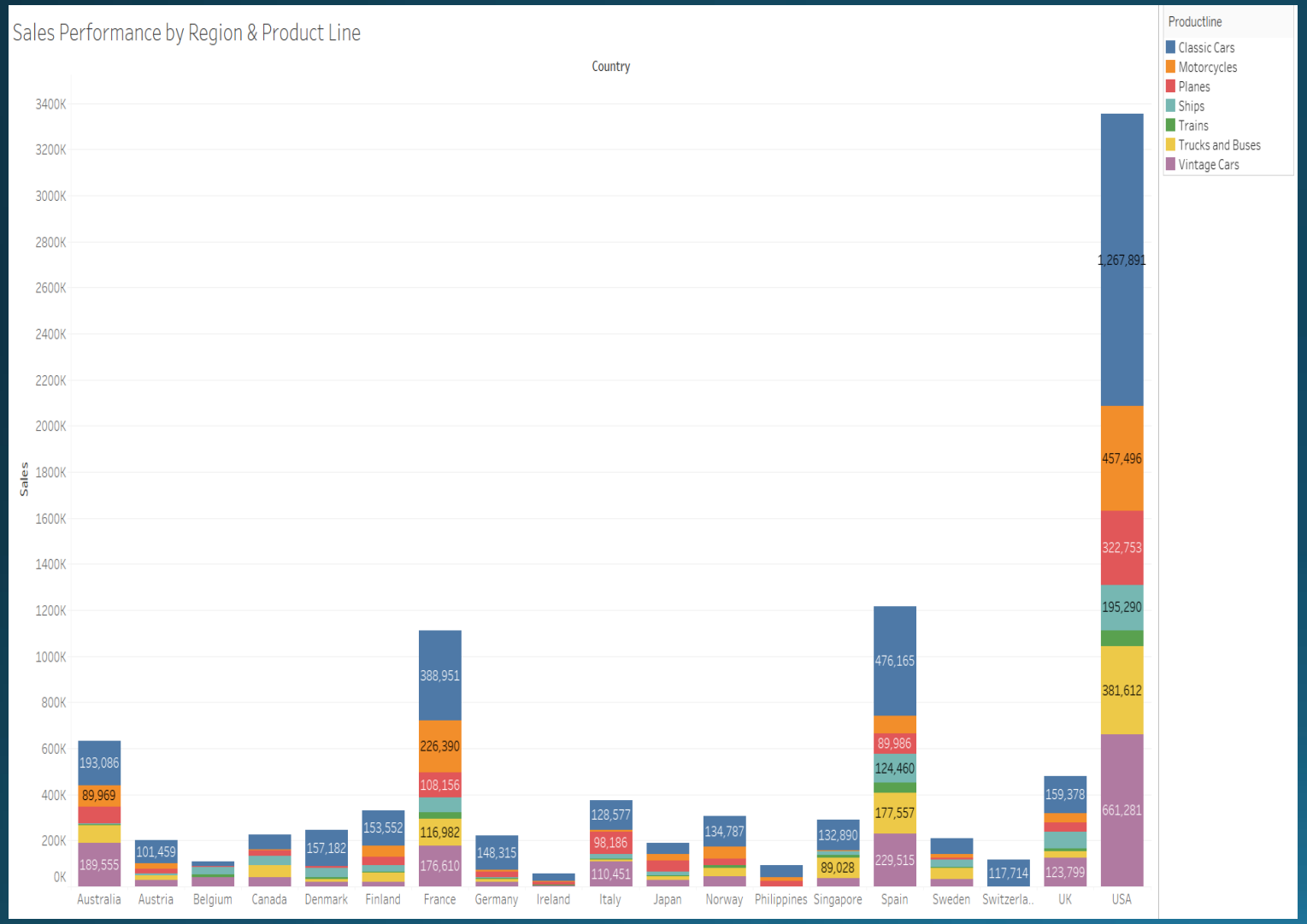
Declining Trend for Classic Cars

- Classic Cars started strong but showed a gradual decline over the weeks, with occasional peaks.
- Vintage Cars exhibited sporadic spikes, indicating demand surges on certain weeks.

High Volatility in Sales

- Motorcycles, Trucks & Buses, and Planes showed fluctuating trends, with no clear upward or downward movement.
- Trains and Ships maintained consistently low sales across all weeks.

Sales Performance by Region & Product Line



Dominance of the USA in Sales

The USA leads in total sales, significantly higher than other regions.

Classic Cars contribute the most, followed by Motorcycles and Vintage Cars.

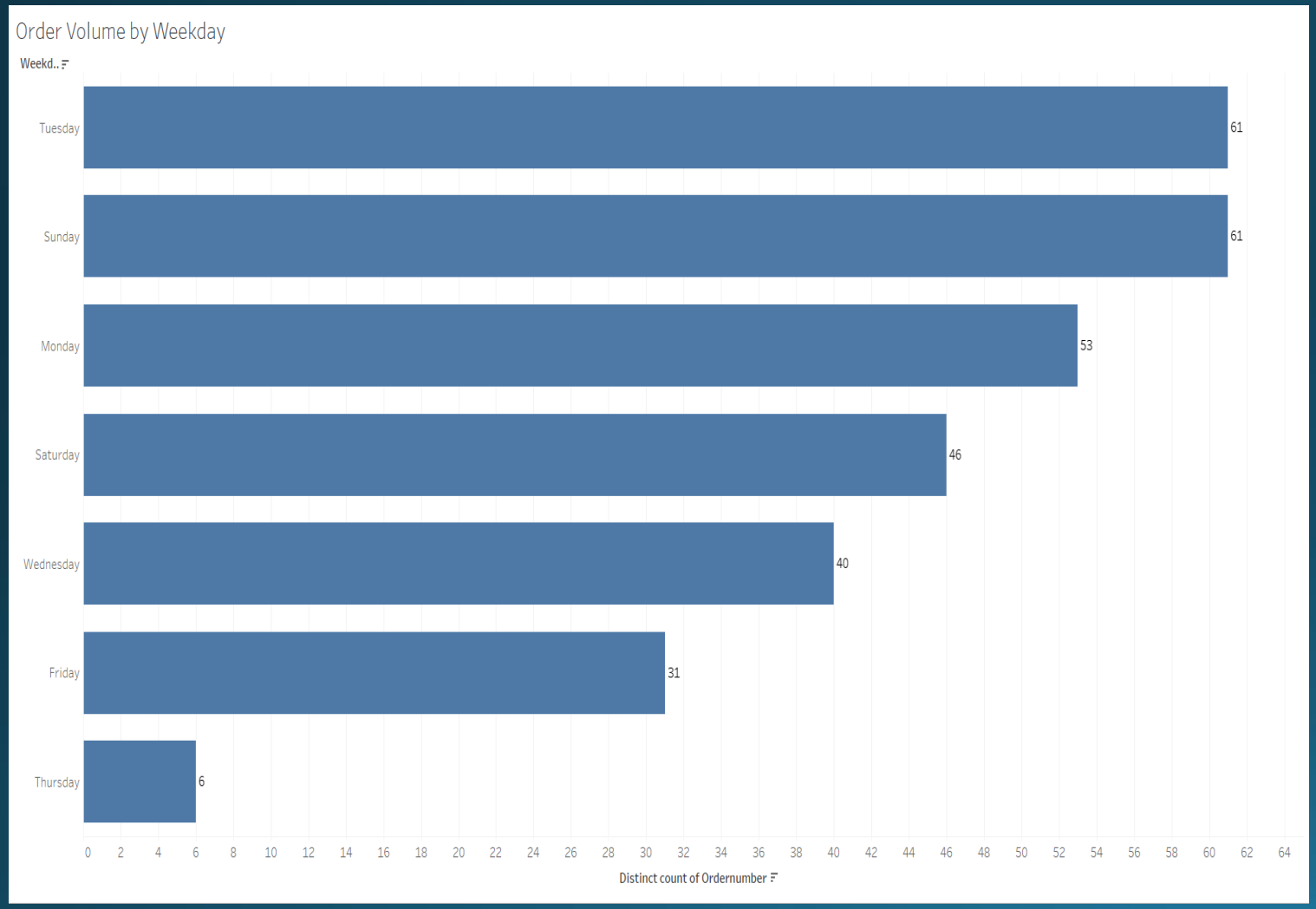
Strong Sales in France & Spain

France and Spain exhibit high sales volumes, with Classic Cars and Motorcycles as key drivers.

Lower Sales in Smaller Markets

Countries like Belgium, Ireland, and the Philippines show comparatively lower sales across all product lines.

Order Volume by Weekday



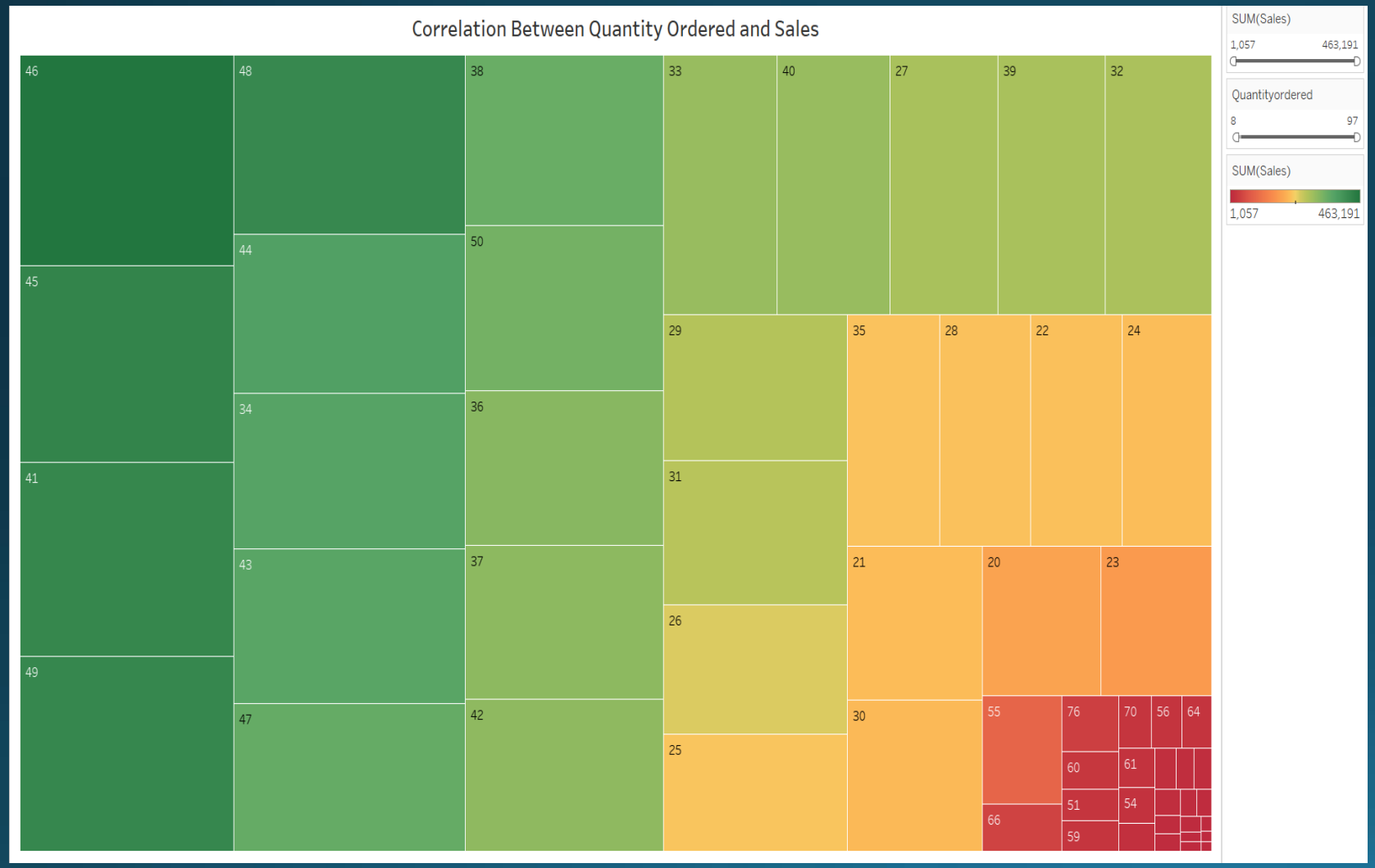
Tuesday and Sunday have the highest order volume (61 orders each).

Monday follows with 53 orders, showing strong activity at the beginning of the week.

Thursday has the lowest order volume (only 6 orders), indicating a significant drop.

Friday and Wednesday have moderate order counts, but still lower than peak days.

Correlation Between Quantity Ordered And Sales



Darker green shades represent higher sales, while red and orange shades indicate lower sales.

Higher quantity ordered (larger blocks) is generally associated with higher sales, but some smaller blocks in red show exceptions where high quantity doesn't translate to high sales.

Lower quantity orders (smaller blocks) are more prevalent in red/orange shades, indicating lower sales contributions.

What is RFM Analysis?

A customer segmentation technique based on **Recency, Frequency, and Monetary (RFM) scores**.

Helps identify **best customers, inactive customers, and at-risk customers**

Enhances **targeted marketing strategies**

RFM Metrics

RECENCY

The freshness of the customer activity, be it purchases or visits

E.g. Time since last order or last engaged with the product

FREQUENCY

The frequency of the customer transactions or visits

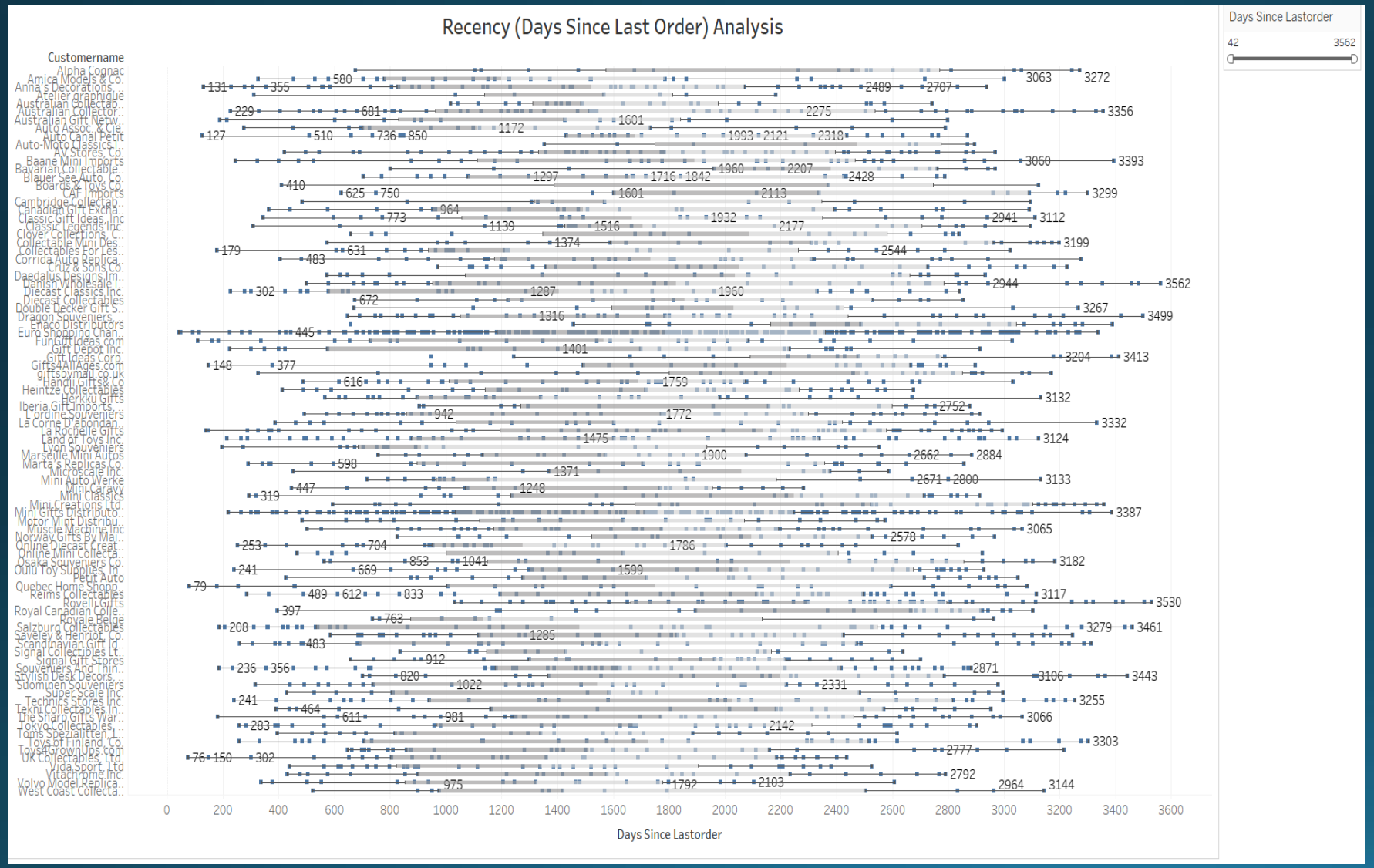
E.g. Total number of transactions or average time between transactions/ engaged visits

MONETARY

The intention of customer to spend or purchasing power of customer

E.g. Total or average transactions value

Recency Analysis



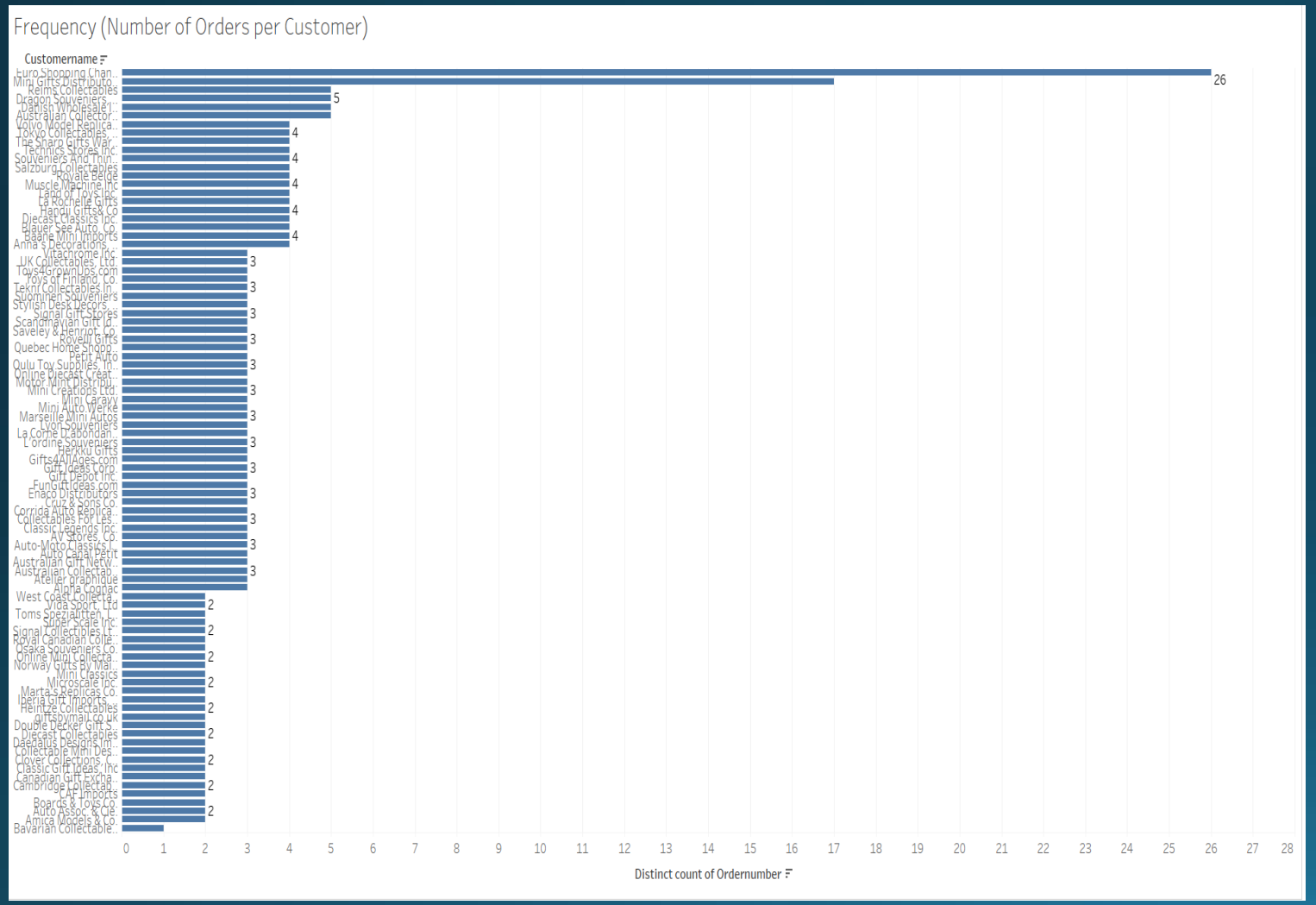
The visualization shows the distribution of days since the last order for different customers.

Some customers have recent transactions (low days since last order), while others have not ordered in a very long time.

The spread of days is wide, indicating a diverse purchasing pattern among customers.

Some customers have extremely high recency values (e.g., above 3000 days), suggesting they may have churned or stopped engaging with the business.

Frequency Analysis

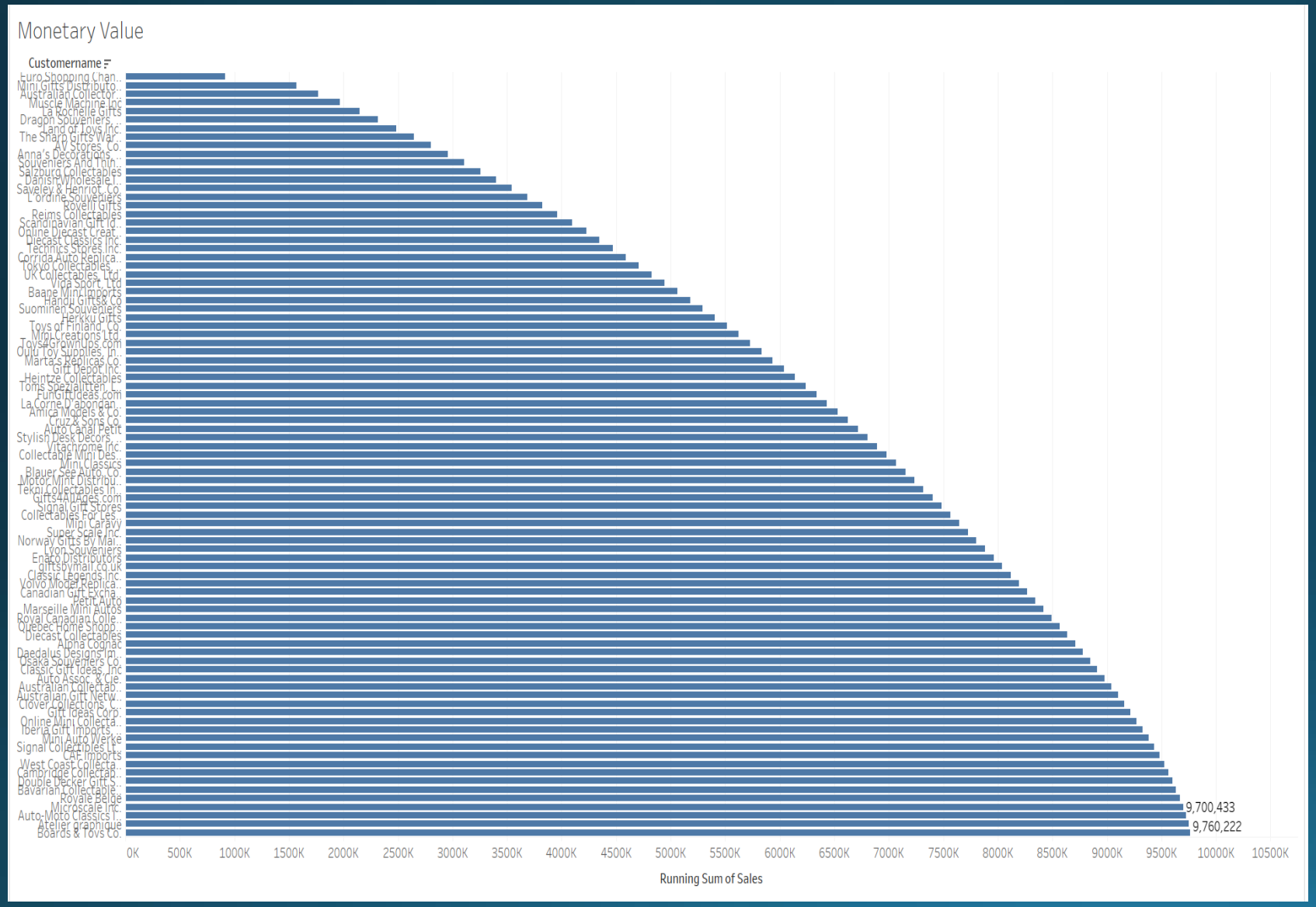


The majority of customers have made only a few purchases (2-4 orders).

One customer (Euro Shopping) has placed 26 orders, which is significantly higher than the rest.

A few other customers have placed 5 or 4 orders, but the distribution is heavily skewed towards lower order frequencies.

Monetary Value Analysis

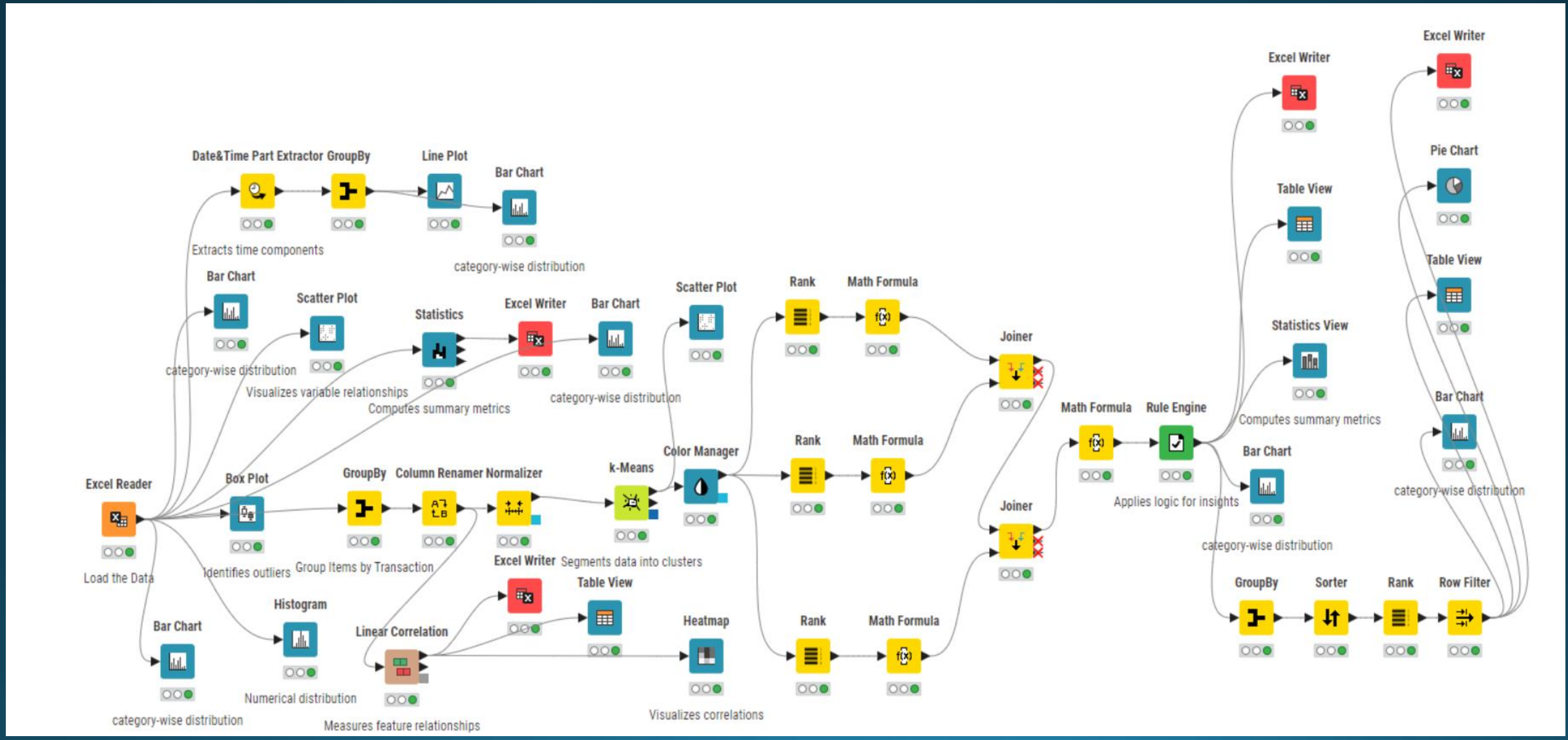


The distribution of monetary value is heavily skewed, with a few customers contributing significantly to total sales.

The highest-value customer accounts for approximately $9.76M in sales, followed closely by another with $9.70M.

A long tail of customers contributes smaller sales amounts, reinforcing the Pareto principle (80/20 rule)—a small percentage of customers drive most of the revenue.

KNIME Workflow For Customer Segmentation



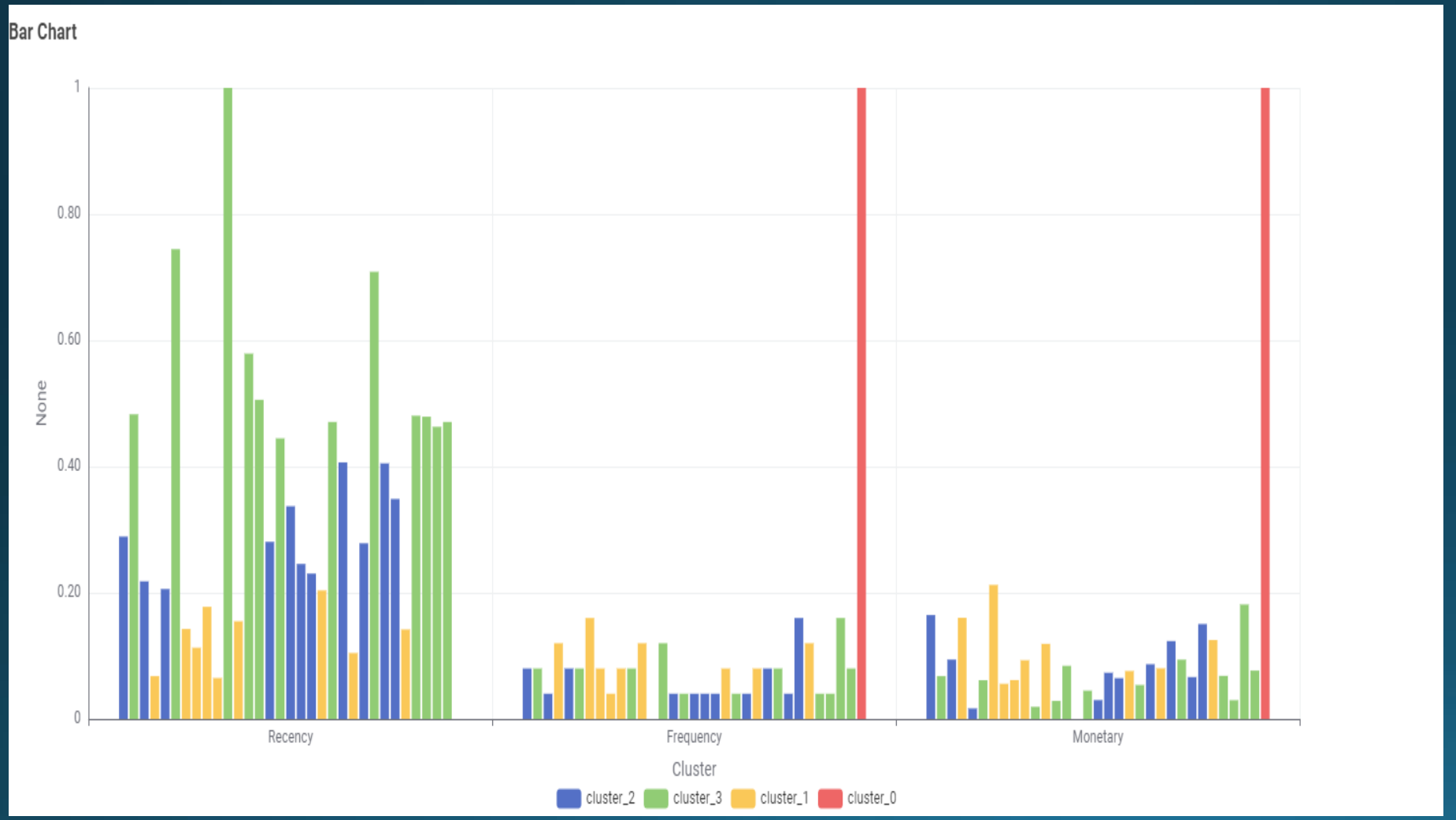
KNIME Workflow Process

RFM Segmentation Process

Based on RFM analysis, customers were segmented into four distinct groups to optimize targeted marketing and retention strategies.

Segment Name	RFM Score Criteria	Business Justification
Best customers	R: 4-5, F: 4-5, M: 4-5	Most valuable customers with high engagement and spending. Need loyalty programs and exclusive offers.
Loyal Customers	R: 3-5, F: 3-5, M: 3-4	Regular buyers who consistently engage with the brand. Need personalized engagement and retention efforts.
Customers on the verge of churning	R: 1-2, F: 3-5, M: 3-5	Previously high-value customers who haven't purchased recently. Require reactivation campaigns and special discounts.
Lost customers	R: 1-2, F: 1-2, M: 1-2	Least engaged customers with minimal activity. May need aggressive win-back offers or lower-priority marketing focus.

The Number Of Customer Segments



Cluster 0 (Red) are high-value customers who spend significantly more.

Cluster 3 (Green) consists of inactive customers who need re-engagement strategies.

Cluster 2 (Blue) and Cluster 1 (Yellow) show moderate engagement and spending patterns.

Customer Segments & Justification

Based on natural clustering of **RFM scores**

Each segment represents a **distinct behavior pattern**

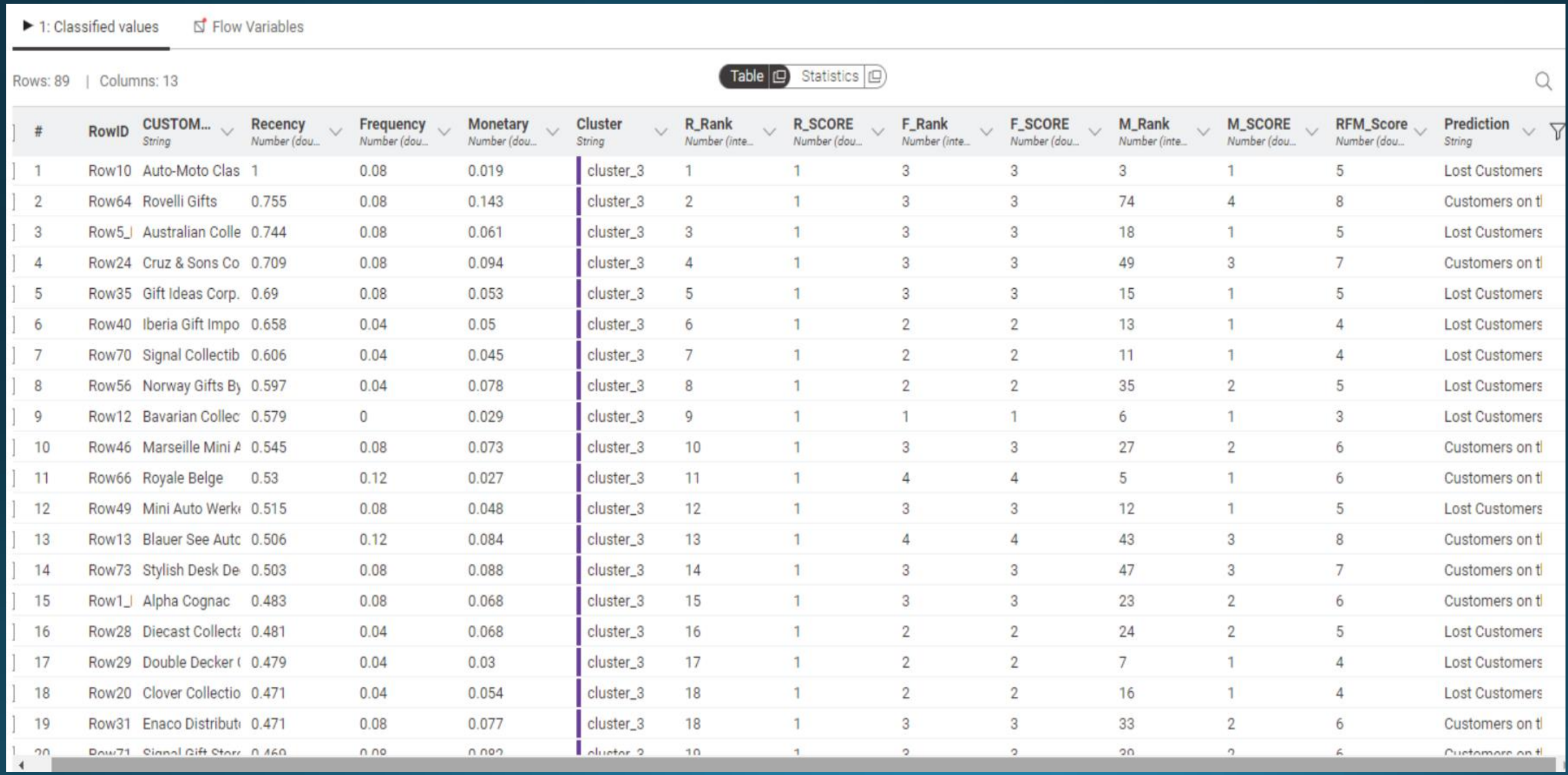
Helps businesses apply **targeted marketing strategies**

Based on RFM analysis, we segmented customers into four groups: High-Value, Loyal, At-Risk, and Lost. These segments reflect distinct purchasing behaviors. High-Value customers are top revenue drivers, requiring personalized engagement.

Loyal customers maintain steady spending, while At-Risk customers need reactivation efforts.

Lost customers show minimal engagement and may need win-back strategies.

Customer Segmentation using RFM analysis



▶ 1: Classified values | Flow Variables

Rows: 89 | Columns: 13 — Table | Statistics

#	RowID	CUSTOM... String	Recency Number (dou...	Frequency Number (dou...	Monetary Number (dou...	Cluster String	R_Rank Number (inte...	R_SCORE Number (dou...	F_Rank Number (inte...	F_SCORE Number (dou...	M_Rank Number (inte...	M_SCORE Number (dou...	RFM_Score Number (dou...	Prediction String
1	Row10	Auto-Moto Clas	1	0.08	0.019	cluster_3	1	1	3	3	3	1	5	Lost Customers
2	Row64	Rovelli Gifts	0.755	0.08	0.143	cluster_3	2	1	3	3	74	4	8	Customers on tl
3	Row5_I	Australian Colle	0.744	0.08	0.061	cluster_3	3	1	3	3	18	1	5	Lost Customers
4	Row24	Cruz & Sons Co	0.709	0.08	0.094	cluster_3	4	1	3	3	49	3	7	Customers on tl
5	Row35	Gift Ideas Corp.	0.69	0.08	0.053	cluster_3	5	1	3	3	15	1	5	Lost Customers
6	Row40	Iberia Gift Impo	0.658	0.04	0.05	cluster_3	6	1	2	2	13	1	4	Lost Customers
7	Row70	Signal Collectib	0.606	0.04	0.045	cluster_3	7	1	2	2	11	1	4	Lost Customers
8	Row56	Norway Gifts By	0.597	0.04	0.078	cluster_3	8	1	2	2	35	2	5	Lost Customers
9	Row12	Bavarian Collec	0.579	0	0.029	cluster_3	9	1	1	1	6	1	3	Lost Customers
10	Row46	Marseille Mini A	0.545	0.08	0.073	cluster_3	10	1	3	3	27	2	6	Customers on tl
11	Row66	Royale Belge	0.53	0.12	0.027	cluster_3	11	1	4	4	5	1	6	Customers on tl
12	Row49	Mini Auto Werk	0.515	0.08	0.048	cluster_3	12	1	3	3	12	1	5	Lost Customers
13	Row13	Blauer See Autc	0.506	0.12	0.084	cluster_3	13	1	4	4	43	3	8	Customers on tl
14	Row73	Stylish Desk De	0.503	0.08	0.088	cluster_3	14	1	3	3	47	3	7	Customers on tl
15	Row1_I	Alpha Cognac	0.483	0.08	0.068	cluster_3	15	1	3	3	23	2	6	Customers on tl
16	Row28	Diecast Collecta	0.481	0.04	0.068	cluster_3	16	1	2	2	24	2	5	Lost Customers
17	Row29	Double Decker (	0.479	0.04	0.03	cluster_3	17	1	2	2	7	1	4	Lost Customers
18	Row20	Clover Collectio	0.471	0.04	0.054	cluster_3	18	1	2	2	16	1	4	Lost Customers
19	Row31	Enaco Distribut	0.471	0.08	0.077	cluster_3	18	1	3	3	33	2	6	Customers on tl
20	Row71	Signal Gift Stor	0.469	0.08	0.082	cluster_3	19	1	3	3	30	2	6	Customers on tl

Table Representation of Segments

Category	Top 5 Customers (ID)	Justification	Business Strategy
Best Customers	R21,R1,R24,R20,R17	High RFM scores (5-5-5), frequent purchases, high spending	VIP treatment, exclusive rewards, loyalty benefits
Customers on the Verge of Churning	R43,R31,R78,R12,R26	Decreasing frequency, reduced recency	Win-back campaigns, personalized re-engagement offers
Lost Customers	R88,R51,R80,R72,R53	Very low RFM scores (1-1-1), inactive for a long time	Last-attempt reactivation offers or deprioritization
Loyal Customers	R2,R35,R13,R8,R7	Consistently high frequency but moderate monetary value	Reward loyalty, encourage upselling, personalized offers

Top 5 Best Customer

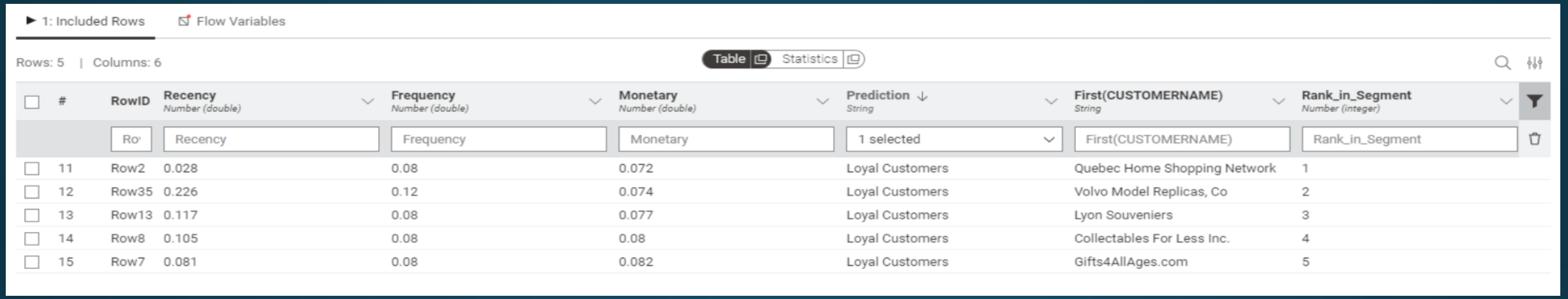
Rows: 5 | Columns: 6

Table | Statistics

#	RowID	Recency *Number (double)*	Frequency *Number (double)*	Monetary *Number (double)*	Prediction *String*	First(CUSTOMERNAME) *String*	Rank_in_Segment *Number (integer)*
	Ro	Recency	Frequency	Monetary	1 selected	First(CUSTOMERNAME)	Rank_in_Segment
16	Row21	0.155	0.12	0.119	Best Customers	Baane Mini Imports	1
17	Row1	0.026	0.08	0.121	Best Customers	UK Collectables, Ltd.	2
18	Row24	0.166	0.12	0.123	Best Customers	Tokyo Collectables, Ltd	3
19	Row20	0.152	0.12	0.124	Best Customers	Technics Stores Inc.	4
20	Row17	0.142	0.12	0.125	Best Customers	Diecast Classics Inc.	5

Baane Mini Imports - Rank #1 → UK Collectables, Ltd. - Rank #2 → Tokyo Collectables, Ltd. - Rank #3 → Technics Stores Inc. - Rank #4 → Diecast Classics Inc. - Rank #5

Top 5 Loyal Customer



1: Included Rows | Flow Variables

Rows: 5 | Columns: 6 — Table | Statistics

#	RowID	Recency (Number (double))	Frequency (Number (double))	Monetary (Number (double))	Prediction (String)	First(CUSTOMERNAME) (String)	Rank_in_Segment (Number (integer))
	Ro	Recency	Frequency	Monetary	1 selected	First(CUSTOMERNAME)	Rank_in_Segment
11	Row2	0.028	0.08	0.072	Loyal Customers	Quebec Home Shopping Network	1
12	Row35	0.226	0.12	0.074	Loyal Customers	Volvo Model Replicas, Co	2
13	Row13	0.117	0.08	0.077	Loyal Customers	Lyon Souveniers	3
14	Row8	0.105	0.08	0.08	Loyal Customers	Collectables For Less Inc.	4
15	Row7	0.081	0.08	0.082	Loyal Customers	Gifts4AllAges.com	5

Quebec Home Shopping Network – Rank #1

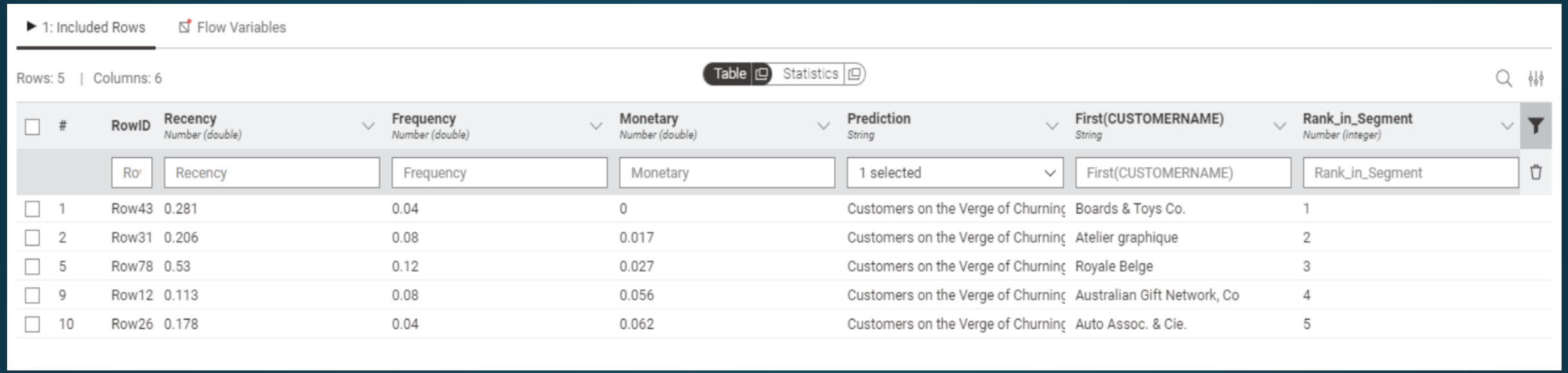
Volvo Model Replicas, Co – Rank #2

Lyon Souvenirs – Rank #3

Collectables For Less Inc. – Rank #4

Gifts4AllAges.com – Rank #5

Top 5 Customers on the verge of churning Customer



▶ 1: Included Rows | Flow Variables

Rows: 5 | Columns: 6 — Table | Statistics

#	RowID	Recency *Number (double)*	Frequency *Number (double)*	Monetary *Number (double)*	Prediction *String*	First(CUSTOMERNAME) *String*	Rank_in_Segment *Number (integer)*
	Ro'	Recency	Frequency	Monetary	1 selected	First(CUSTOMERNAME)	Rank_in_Segment
1	Row43	0.281	0.04	0	Customers on the Verge of Churning	Boards & Toys Co.	1
2	Row31	0.206	0.08	0.017	Customers on the Verge of Churning	Atelier graphique	2
5	Row78	0.53	0.12	0.027	Customers on the Verge of Churning	Royale Belge	3
9	Row12	0.113	0.08	0.056	Customers on the Verge of Churning	Australian Gift Network, Co	4
10	Row26	0.178	0.04	0.062	Customers on the Verge of Churning	Auto Assoc. & Cie.	5

Boards & Toys Co. – Rank #1

Atelier graphique – Rank #2

Royale Belge – Rank #3

Australian Gift Network, Co – Rank #4

Auto Assoc. & Cie. – Rank #5

Top 5 Lost customers Customer

▶ 1: Included Rows | Flow Variables

Rows: 5 | Columns: 6 — Table | Statistics

#	RowID	Recency *Number (double)*	Frequency *Number (double)*	Monetary *Number (double)*	Prediction *String*	First(CUSTOMERNAME) *String*	Rank_in_Segment *Number (integer)*
	Ro'	Recency	Frequency	Monetary	1 selected	First(CUSTOMERNAME)	Rank_in_Segment
3	Row88	1	0.08	0.019	Lost Customers	Auto-Moto Classics Inc.	1
4	Row51	0.314	0.04	0.027	Lost Customers	Microscale Inc.	2
6	Row80	0.579	0	0.029	Lost Customers	Bavarian Collectables Imports, Co.	3
7	Row72	0.479	0.04	0.03	Lost Customers	Double Decker Gift Stores, Ltd	4
8	Row53	0.337	0.04	0.03	Lost Customers	Cambridge Collectables Co.	5

Auto-Moto Classics Inc. – Rank #1

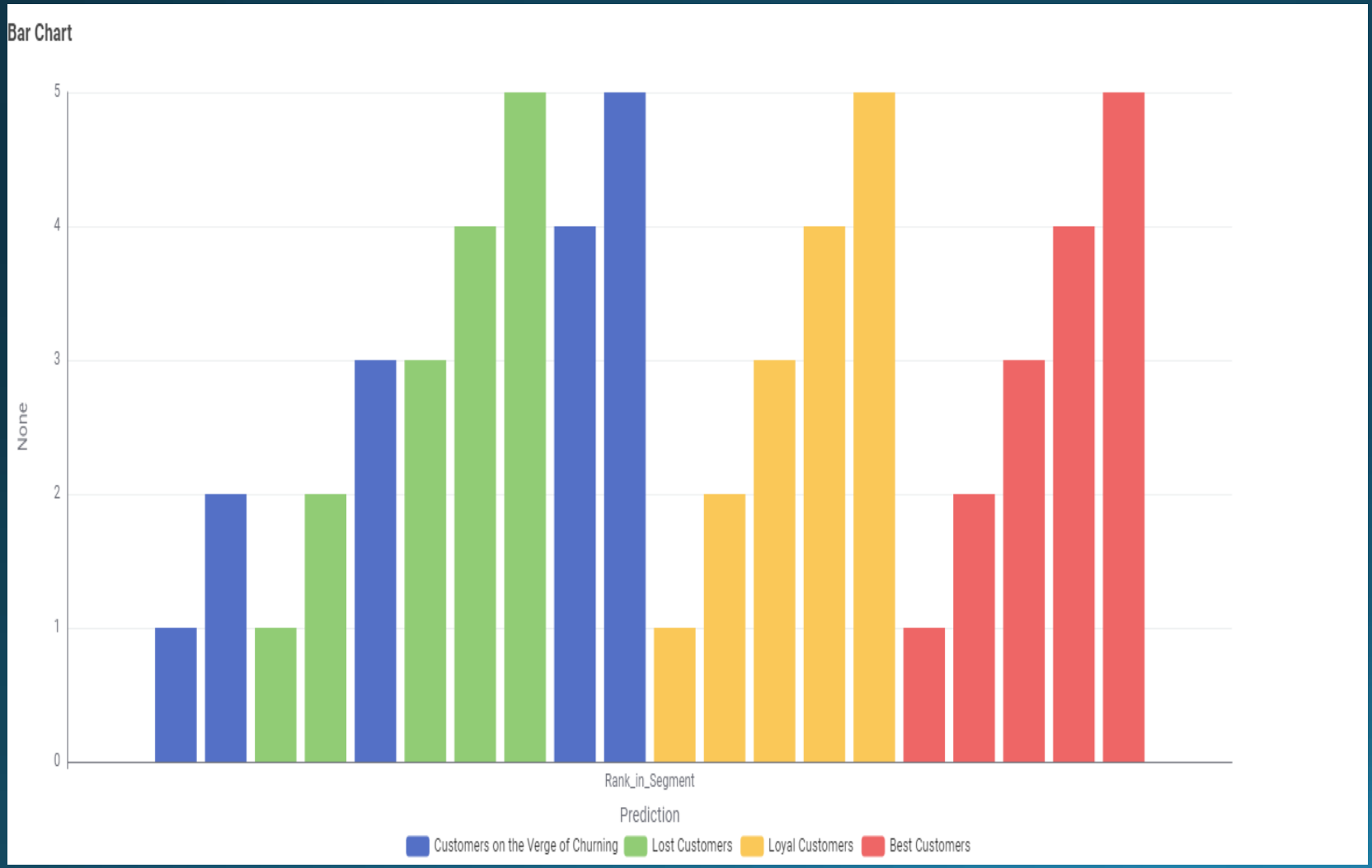
Microscale Inc. – Rank #2

Bavarian Collectables Imports, Co. – Rank #3

Double Decker Gift Stores, Ltd – Rank #4

Cambridge Collectables Co. – Rank #5

Comparing average RFM scores per category



Best Customers: Highest RFM scores; need continued engagement to maintain loyalty.

Customers on the Verge of Churning: At risk of dropping off; require targeted promotions.

Lost Customers: Low engagement; final reactivation attempts before deprioritization.

Loyal Customers: Consistently engaged; focus on retention and upselling strategies.

Strategic Recommendations

Reward Best Customers – Offer loyalty programs, exclusive perks, and personalized discounts to retain high-value customers.

Engage Loyal Customers – Encourage upselling, cross-selling, and referrals through targeted offers and premium memberships.

Win Back At-Risk Customers – Use reactivation campaigns, special discounts, and personalized outreach to prevent churn.

Re-target Lost Customers – Deploy retargeting ads, feedback surveys, and incentives to regain inactive buyers.

Optimize Marketing & Inventory – Align promotions with customer segments and focus on high-demand products for better ROI.

Final Insights & Future Actions

Key Takeaways from RFM Analysis – Identified best, loyal, at-risk, and lost customers for targeted strategies.

Customer Retention & Marketing Impact – Helps personalize promotions, optimize engagement, and improve retention rates.

Data-Driven Decision Making – Enables better resource allocation and higher ROI on marketing efforts.

Next Steps for Business Growth – Implement A/B testing, automate segmentation, and refine customer outreach strategies.

Long-Term Strategy – Continuously monitor RFM segments and adapt campaigns based on customer behavior trends.

Market Basket Analysis for Grocery Store

Understanding Customer Buying Behavior & Business Objective

Identifying frequently purchased item combinations helps optimize promotions, inventory, and customer retention.

Insights from POS data enable personalized marketing, combo offers, and discount strategies to increase revenue and loyalty.

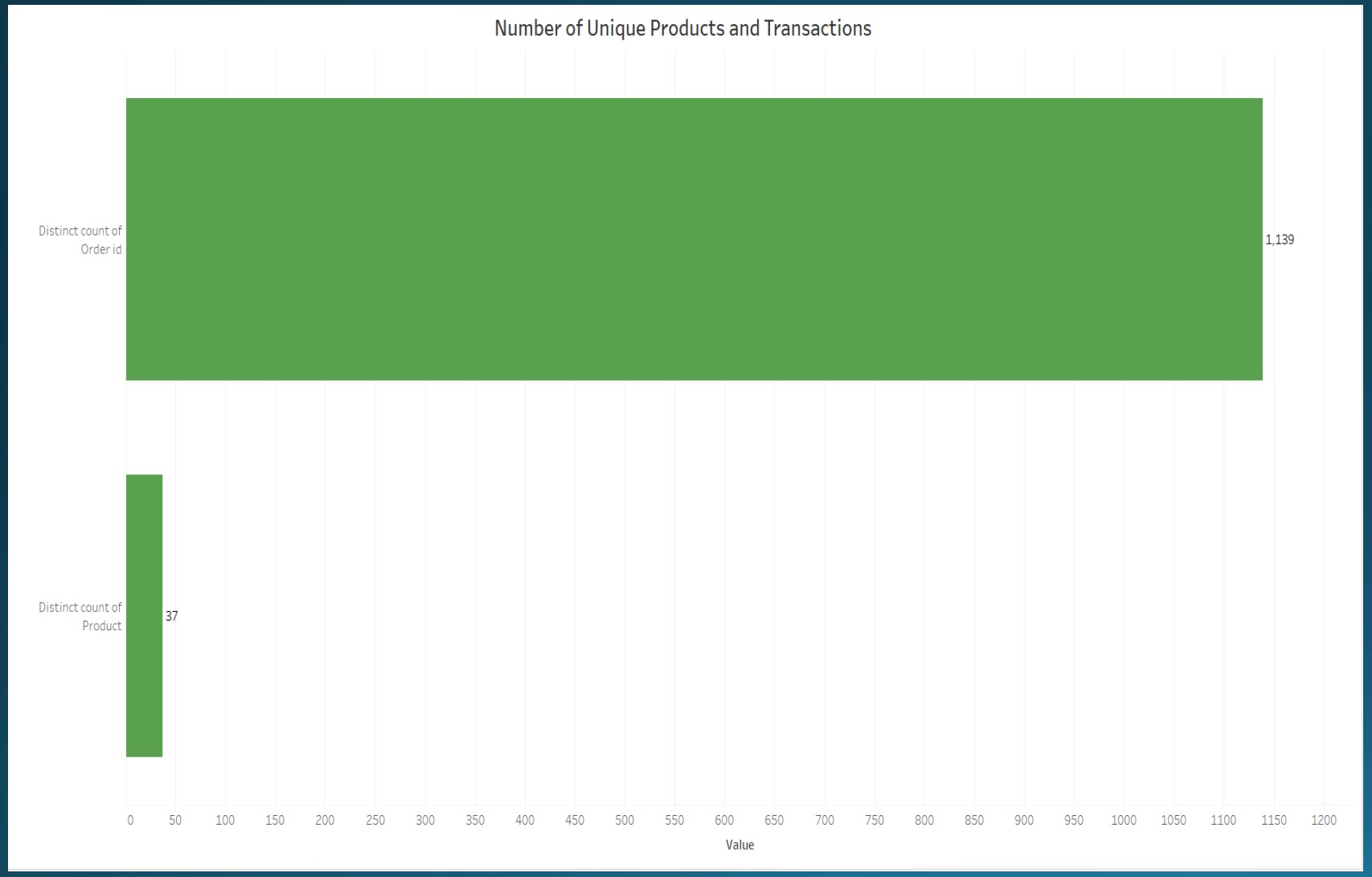
Use association rule mining to uncover meaningful purchasing patterns and drive higher sales.

Leverage insights to create data-driven marketing campaigns and efficient inventory management for sustained business growth.

Data Overview

Dataset	Columns	Time Period	Total Transactions
dataset_group.CSV	• Date, • Order_ID • Product	2018-01-01 to 2020-02-26	1139

Understanding the Dataset



Total Unique Transactions: 1,139 distinct orders.

Total Unique Products: 37 distinct products.

Key Observation: High number of transactions compared to a limited product range, indicating strong sales concentration within a small set of products.

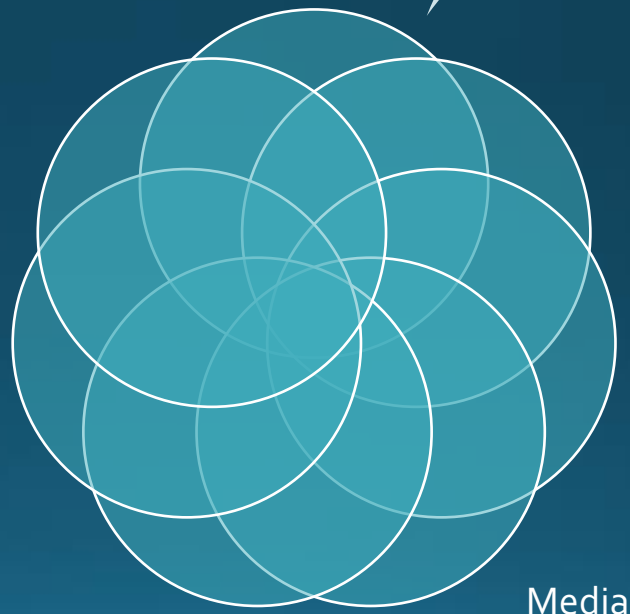
Summary statistics

	Order_id
count	20641.000000
mean	575.986289
std	328.557078
min	1.000000
25%	292.000000
50%	581.000000
75%	862.000000
max	1139.000000

Total Orders Analyzed: 20,641

75% of orders are below 862

Order ID Range: 1 to 1,139



50% of orders are below 581

Mean Order ID: 576, with a standard deviation of 328, indicating a wide distribution of orders.

25% of orders have an ID below 292

Median Order ID (50th percentile): 581, close to the mean, suggesting a relatively symmetric distribution

Defining the Problem Statement

Grocery retail is highly competitive, requiring data-driven strategies to boost sales and customer retention.

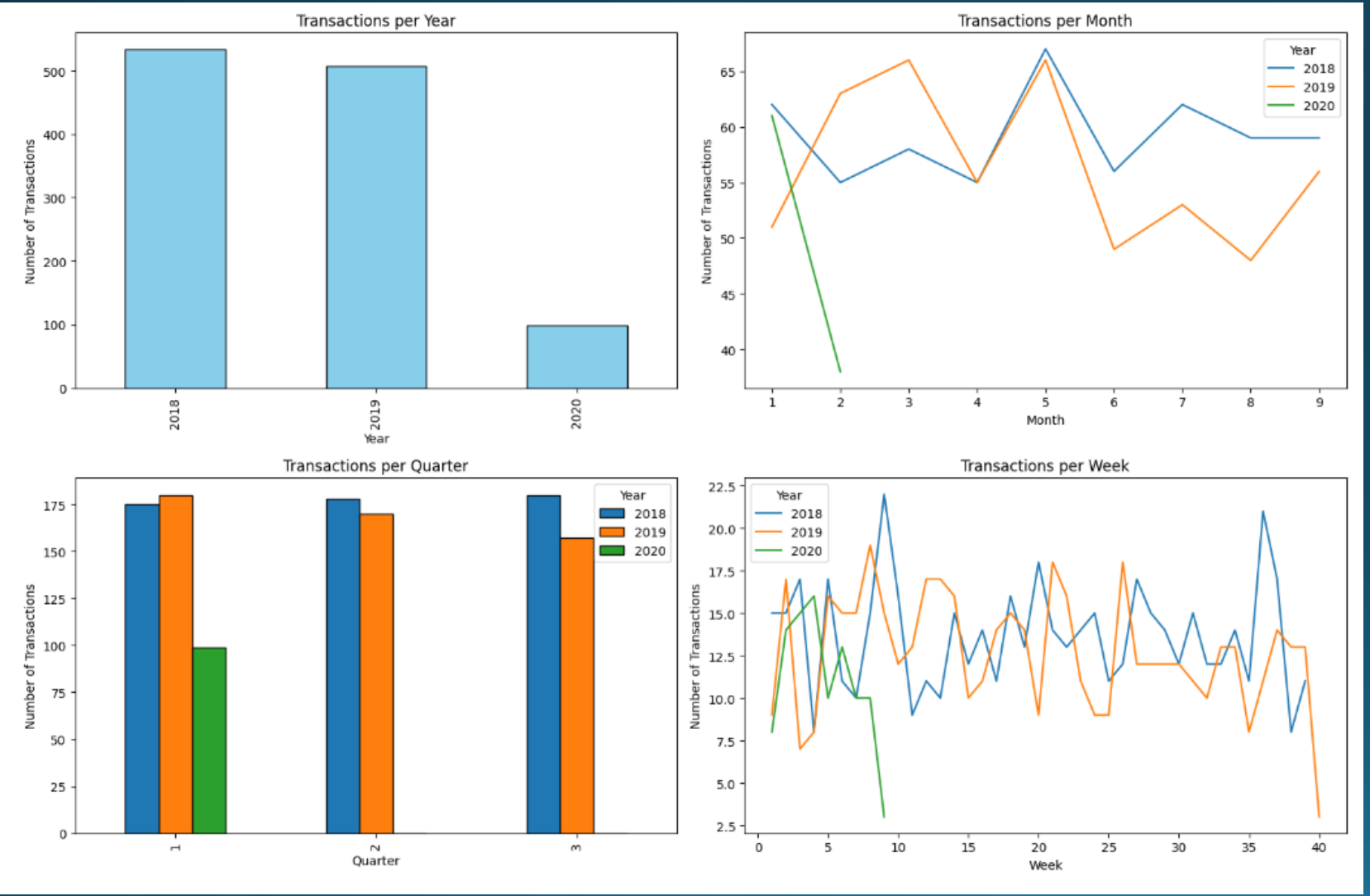
Understanding customer buying patterns can help optimize promotions, inventory management, and marketing efforts.

The challenge is to analyze POS transactional data to identify frequently purchased item combinations.

Using association rule mining, uncover purchase patterns that enable the creation of targeted combo offers and discounts.

Insights from the analysis will help increase basket size, maximize revenue, and enhance customer satisfaction.

Transactions Over Time



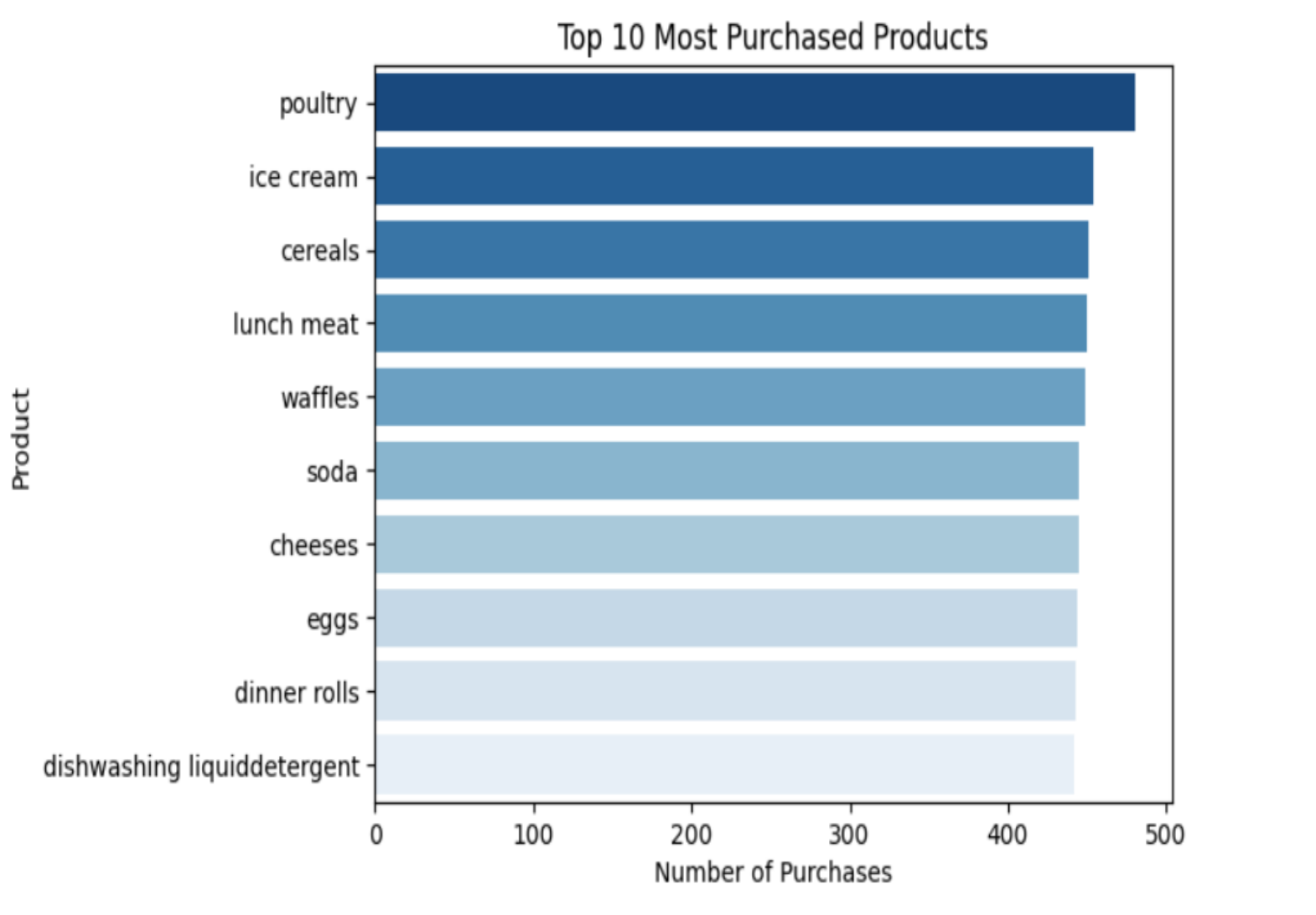
Yearly trend: Transactions peaked in 2018 and 2019, with a significant drop in 2020

Monthly variation: Some months show a spike in transactions (e.g., May),indicating possible seasonal trends

Quarterly insights: Consistent transaction volume in 2018 and 2019, but lower in 2020 Q1.

Weekly fluctuations: Regular variations indicate changing customer demand week over week.

Top Selling Products

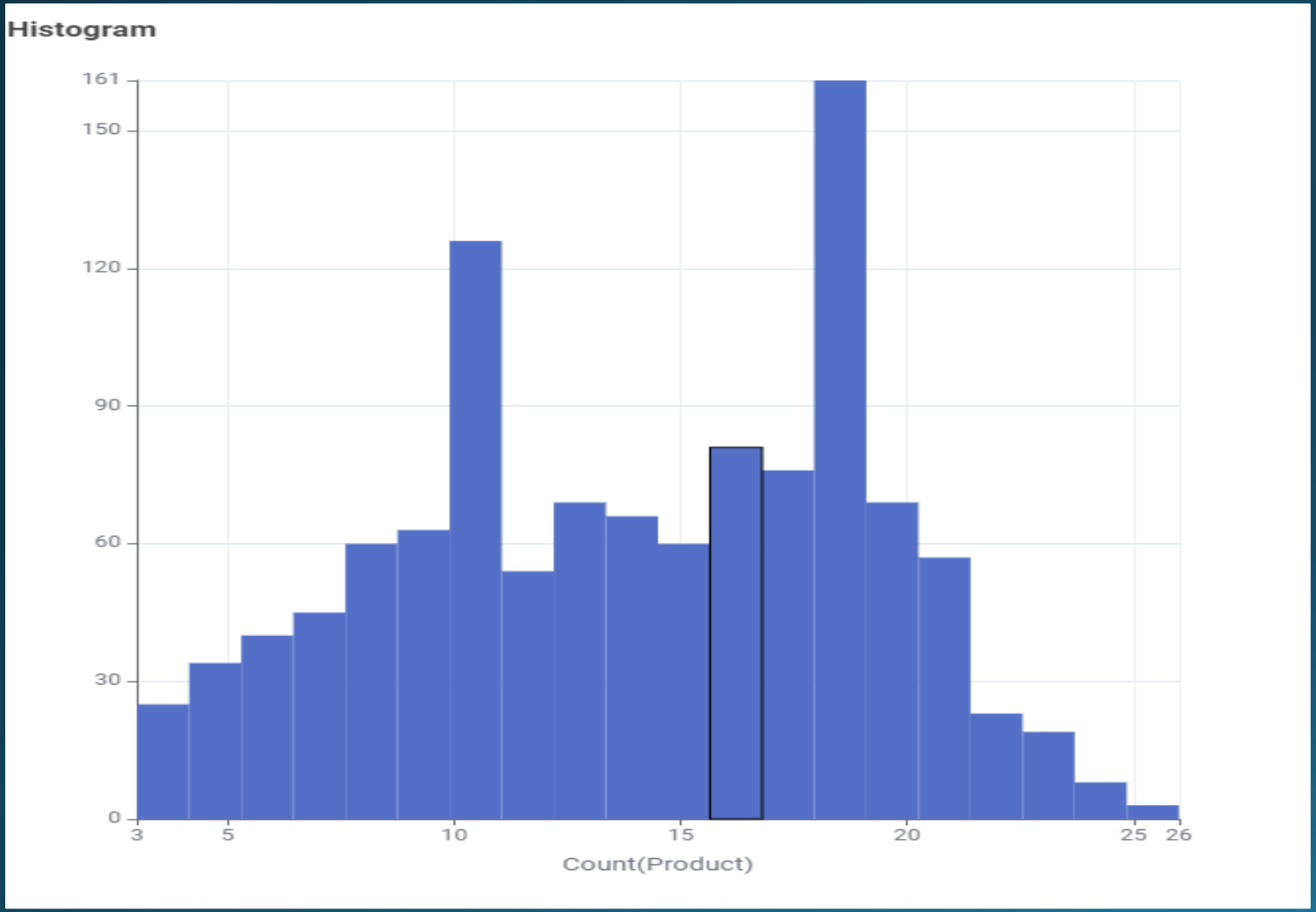


The top-selling product is poultry, followed by ice cream and cereals.

The dominance of food products suggests they are highly popular among customers.

Non-food items like dishwashing liquid detergent are among the top 10, indicating cross-category purchasing behavior.

Basket Size Distribution



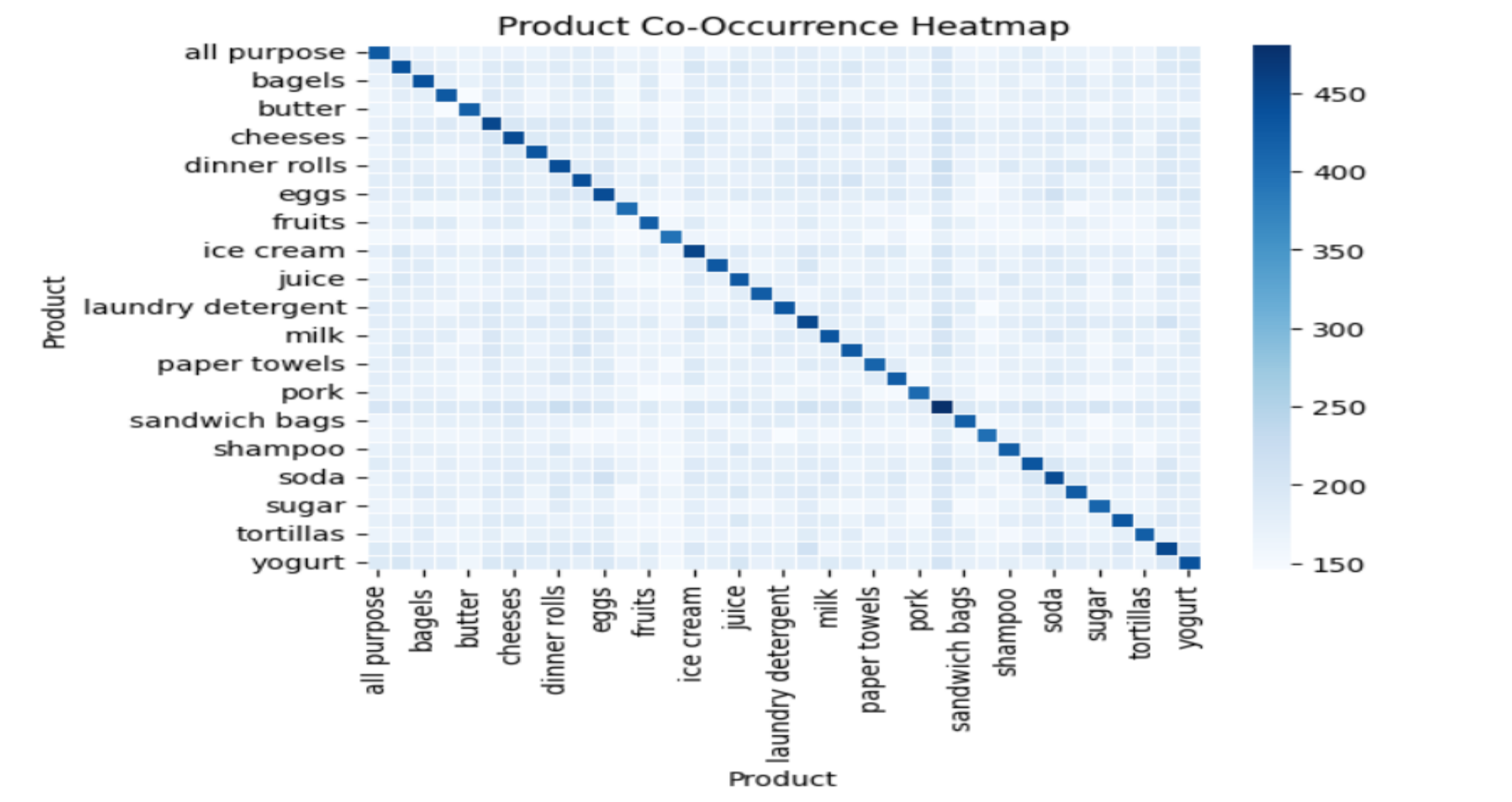
Basket sizes range from 3 to 25 items per order, with peaks around 10, 15 and 20 items.

This indicates two key shopping behaviors: small basket shoppers (~10 items) and bulk shoppers (~20 items).

Fewer transactions occur in the 12-14 item range, suggesting a gap that can be leveraged.

Strategic upselling and promotions can encourage small-basket customers to add more items.

Product Co-Occurrence



Key Insights

Frequent Co-Purchased Items

- The heatmap shows **product co-occurrence** trends, indicating which items are frequently bought together.
- **Darker shades** suggest **stronger co-purchase relationships** among certain products.

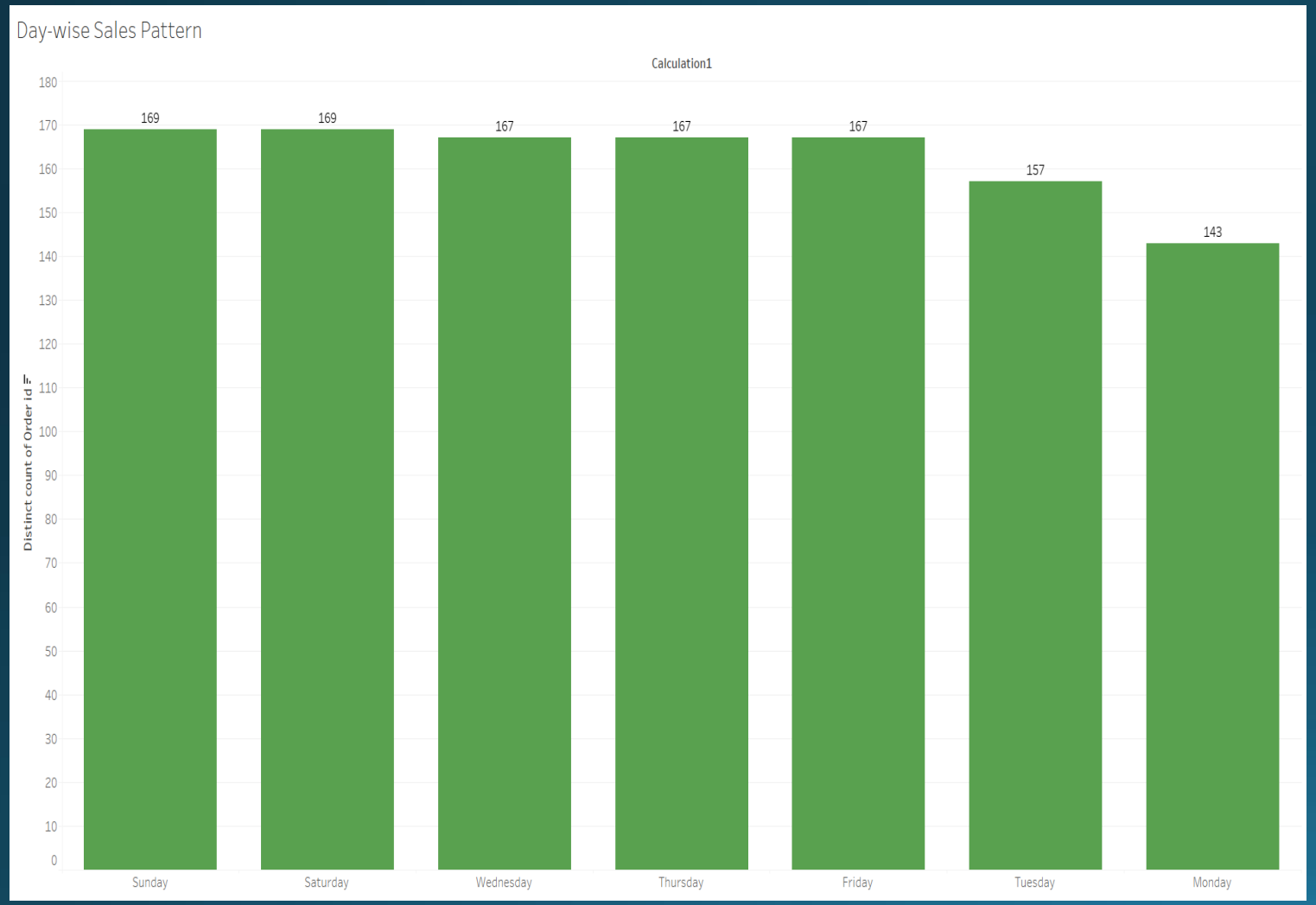
Key Product Associations

- **Milk & Eggs, Butter & Cheese, Bread & Spread (Butter/Jam) → High Co-Occurrence**
- **Laundry Detergent & Paper Towels, Shampoo & Soap → Household Essentials Grouped**
- **Juice & Fruits, Soda & Snacks → Beverage & Snack Pairing**

Gaps in Co-Occurrence

- Some products have **weak associations**, meaning cross-selling potential exists.
- Opportunity to encourage **bundling strategies** for low-association products.

Day-wise Sales Pattern

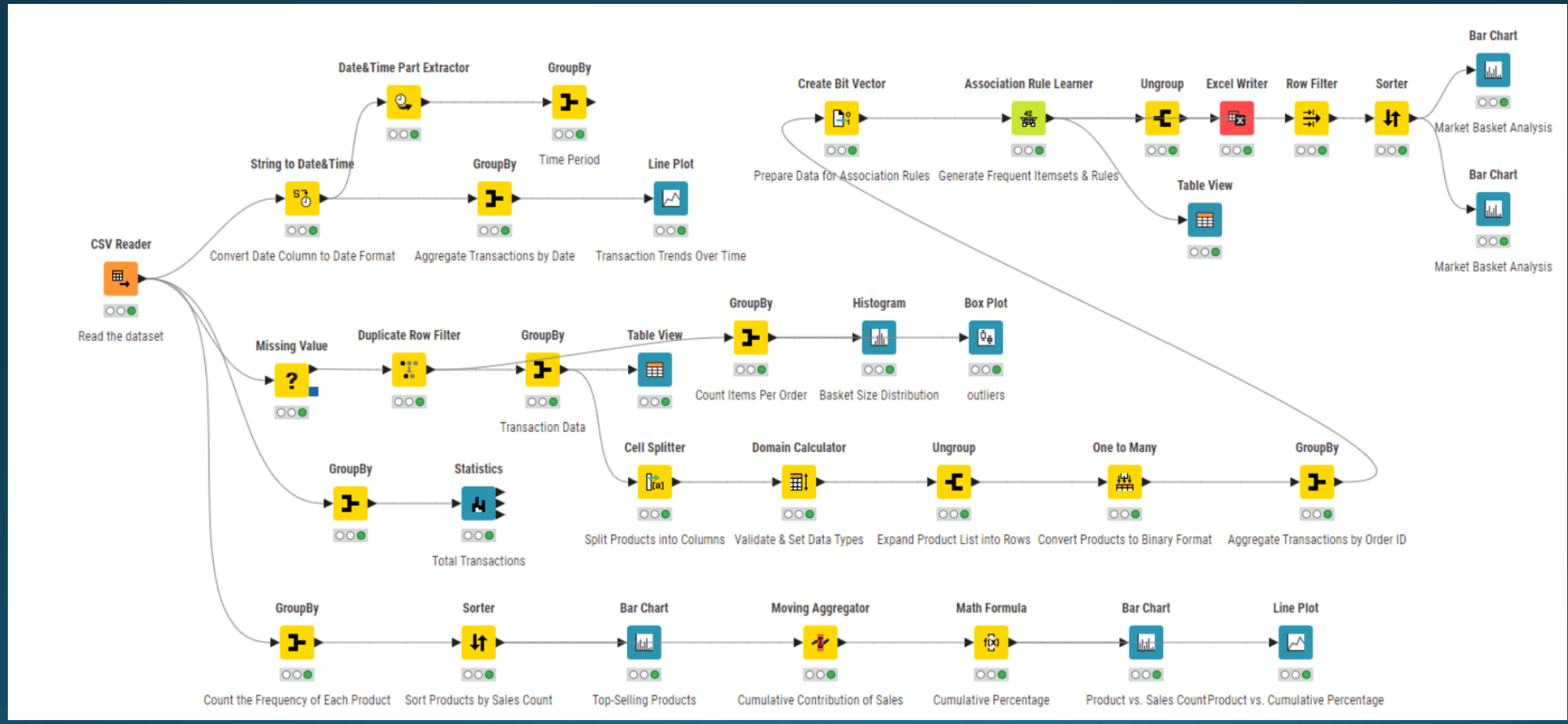


Technique to analyze customer purchasing behavior.

Identifies relationships between items bought together.

Purpose: Improve cross-selling, recommend products, optimize marketing strategies

KNIME Workflow for Market Basket Analysis



KNIME Workflow Process

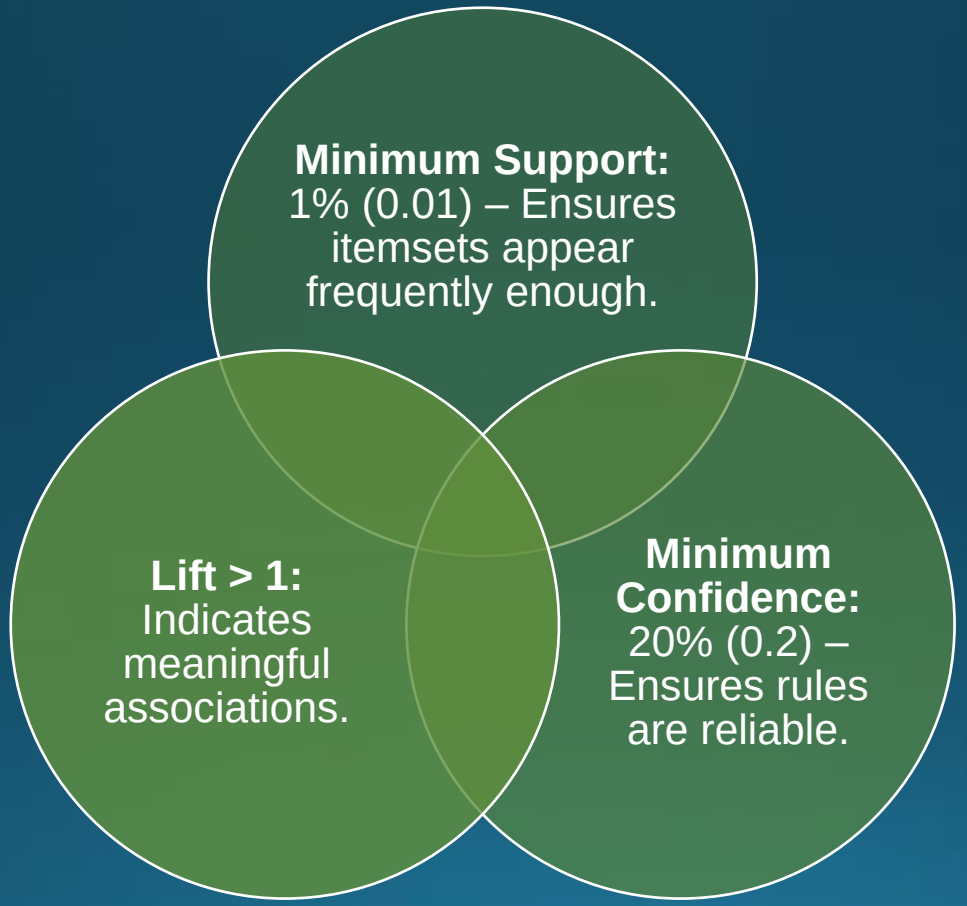
Apriori Algorithm Overview

Definition	A rule-based technique for frequent itemset mining and association rule learning.
Key Concept	If an itemset is frequent, then its subsets must also be frequent.
Technique Used	Identifying strong associations between items in transactional data.

Market Basket Analysis - Methodology

Support : Frequency of an itemset in transactions.	Confidence: Probability that Consequent occurs given Antecedent.	Lift: Strength of association compared to random chance.
(Transactions containing itemset) / (Total transactions)	Support(Itemset) / Support(Antecedent)	Confidence / Expected Confidence

Association Rule Mining & Thresholds



Association Rules Summary

#	RowID	Support *Number (double)*	Confidence *Number (double)*	Lift *Number (double)*	Consequent *String*	implies *String*	Items *Set*
1	rule665	0.195	0.501	1.189	poultry	<---	[dinner rolls]
2	rule127	0.179	0.496	1.178	poultry	<---	[sugar]
3	rule809	0.183	0.484	1.148	poultry	<---	[soap]
4	rule763	0.182	0.484	1.148	poultry	<---	[mixes]
5	rule110	0.187	0.482	1.144	poultry	<---	[dishwashing liquid/detergent]
6	rule448	0.187	0.48	1.228	soda	<---	[eggs]
7	rule749	0.18	0.479	1.234	dishwashing liquid/detergent	<---	[mixes]
8	rule449	0.187	0.479	1.228	eggs	<---	[soda]
9	rule246	0.177	0.472	1.195	lunch meat	<---	[individual meals]
10	rule259	0.186	0.471	1.118	poultry	<---	[lunch meat]
11	rule55	0.181	0.47	1.116	poultry	<---	[yogurt]
12	rule904	0.17	0.47	1.178	ice cream	<---	[paper towels]
13	rule107	0.178	0.469	1.112	poultry	<---	[milk]
14	rule60	0.176	0.469	1.218	yogurt	<---	[juice]
15	rule126	0.176	0.469	1.112	poultry	<---	[juice]
16	rule323	0.176	0.468	1.111	poultry	<---	[all- purpose]
17	rule193	0.172	0.468	1.11	poultry	<---	[sandwich bags]
18	rule238	0.184	0.468	1.184	lunch meat	<---	[waffles]
19	rule124	0.163	0.467	1.109	poultry	<---	[sandwich loaves]
20	rule239	0.184	0.467	1.184	waffles	<---	[lunch meat]

Support (0.163 - 0.195): Indicates how often item pairs appear in transactions.

Confidence (0.467 - 0.501): Measures the probability that the consequent item is purchased when the antecedent is bought.

Lift (1.109 - 1.234):

A Lift > 1 means the items are positively correlated, meaning they are more likely to be purchased together.

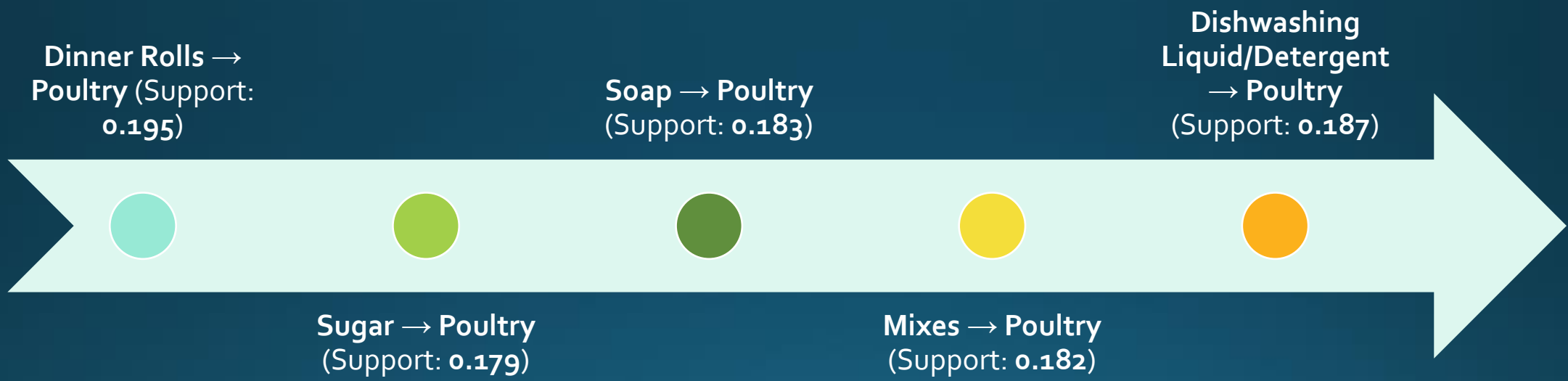
The strongest rule (Lift = 1.234) is between soda and dishwashing liquid/detergent, showing a significant relationship

Top Associations:

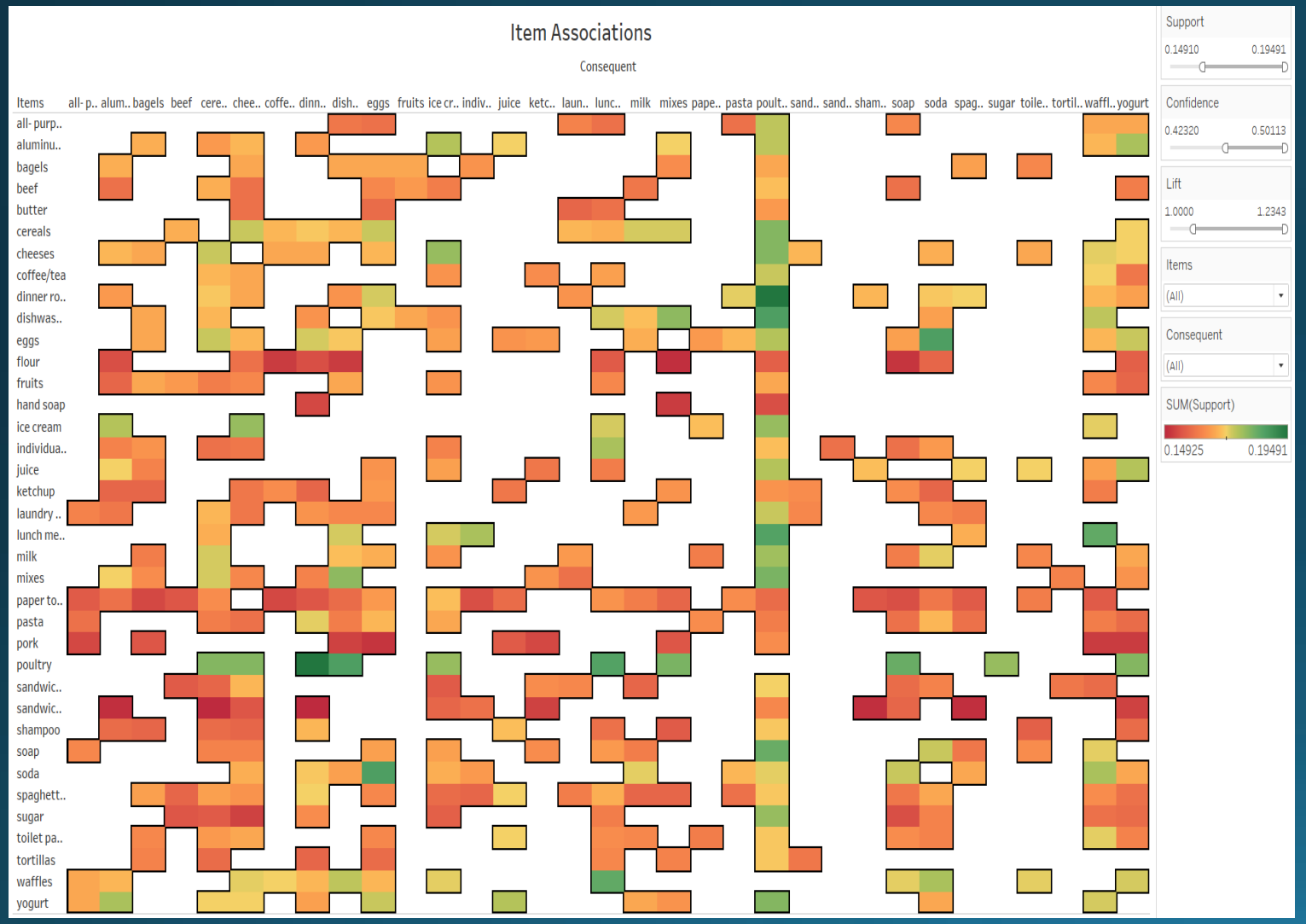
Poultry is frequently associated with dinner rolls, sugar, soap, and mixes.

Lunch meat is commonly bought with individual meals and sandwich loaves.

Key Insights from Market Basket Analysis



Item Associations



Green cells indicate strong associations (higher support), while red/orange cells indicate weaker associations.

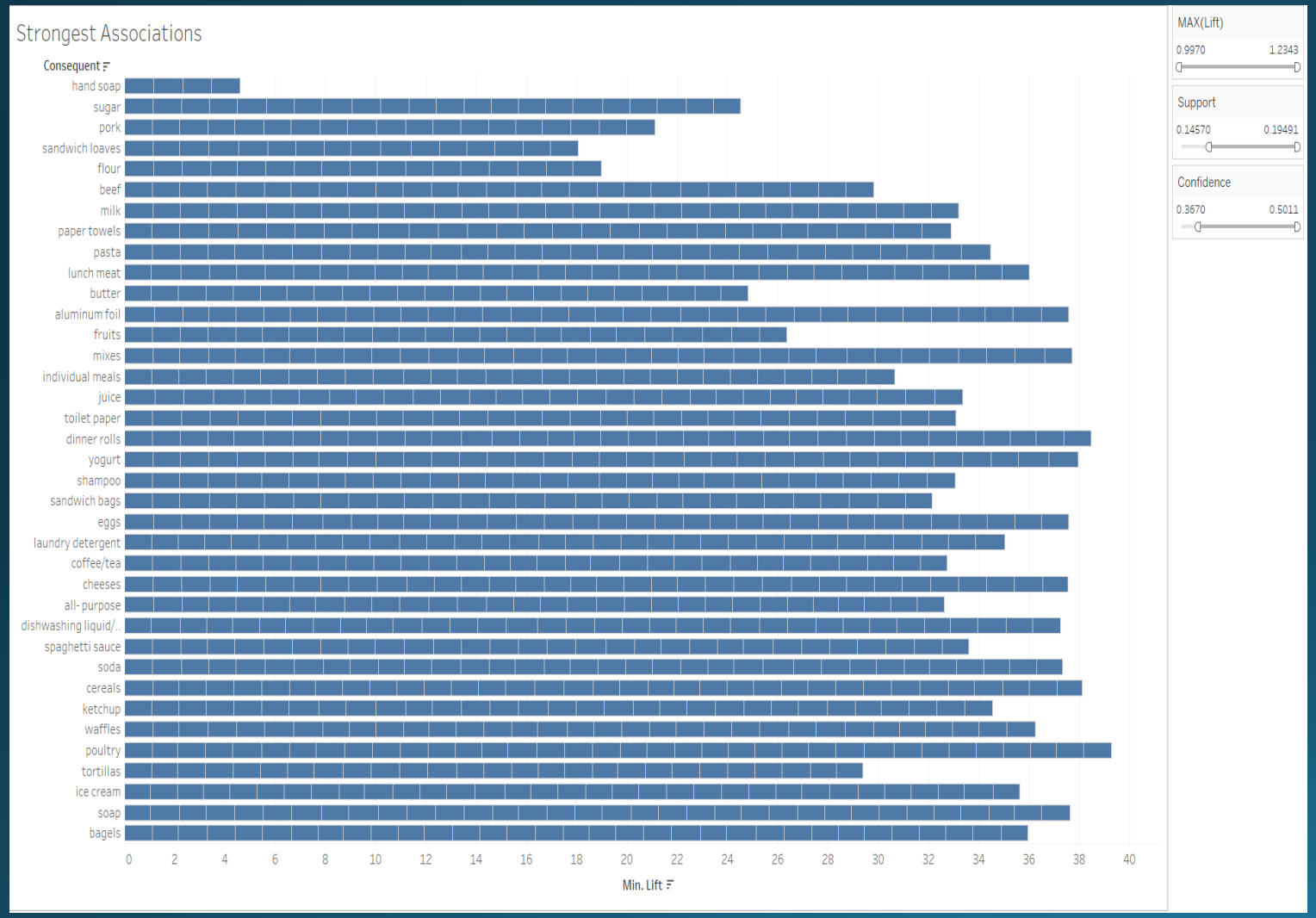
Items such as pasta, flour, and milk show strong associations with multiple products.

Support (0.149 - 0.1949): Frequency of these item pairs appearing together in transactions.

Confidence (0.423 - 0.501): Probability that a customer buys one item given they bought another.

Lift (1.000 - 1.234): Strength of association beyond random chance.

Strong Associations



The bar chart displays frequent item associations, with Min. Lift on the x-axis representing the strength of association.

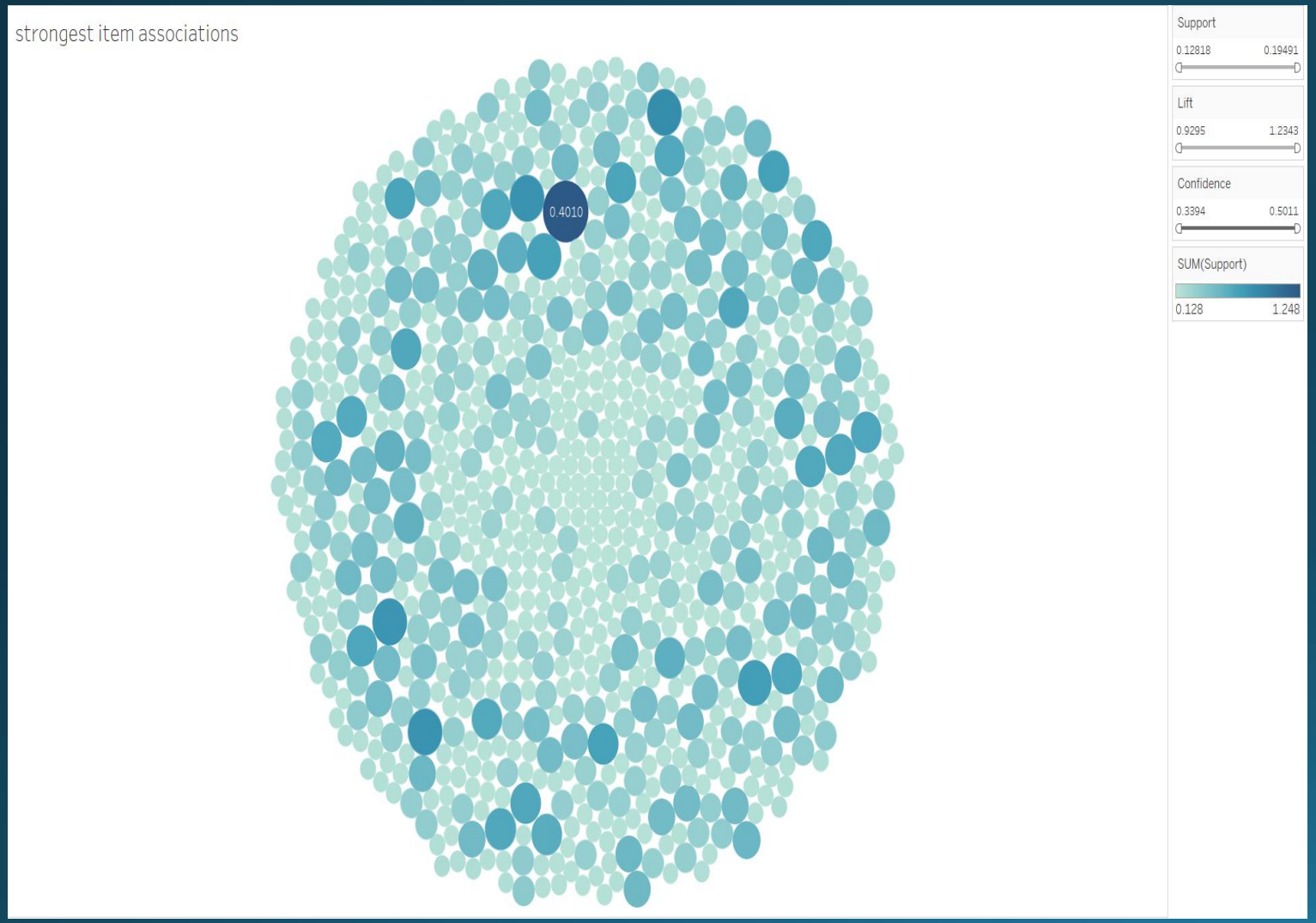
The top associated items include hand soap, sugar, pork, sandwich loaves, and flour, among others.

Lift Value (0.997 - 1.234) suggests how much more likely two products are bought together compared to random chance.

Support (0.1457 - 0.1949) indicates how often these associations appear in transactions.

Confidence (0.367 - 0.501) reflects the probability of purchasing an item when another is bought.

Association Rule Strength



- The bubble chart represents item associations, where larger bubbles indicate stronger relationships between items.
- The size of each bubble corresponds to the strength of the association, with higher support values being more significant.
- The side panel metrics (Support, Lift, and Confidence) help in understanding:

Support: Frequency of item pairs appearing together.

Lift: The strength of an item association beyond random chance.

Confidence: The likelihood that an item is bought given the presence of another.

Association Rule Metrics (Support, Confidence, Lift) Analysis

Strong Associations

Poultry (16.55), Cereals (15.49), and Waffles (15.48) frequently appear in transactions together.

Strong Dependency

Butter (40.40), Yogurt (39.95), and Cheeses (39.57) have the highest lift values, indicating strong cross-selling potential.

Frequent Purchases

Poultry (6.22%), Cereals (5.84%), and Waffles (5.84%) are among the most frequently purchased items.

Lower Association

Hand Soap, Sandwich Loaves, and Pork show weaker associations.

Strongest Associations (High Lift Values)

Dishwashing Liquid/Detergent → Poultry (Lift: **1.144**)

- Poultry is **1.144 times** more likely to be purchased when Dishwashing Liquid/Detergent is bought.

Soda → Dishwashing Liquid/Detergent (Lift: **1.234**)

- This means customers who buy soda also tend to buy dishwashing detergent **1.234 times more than expected**.

Mixes → Dishwashing Liquid/Detergent (Lift: **1.228**)

- Customers who buy mixes have a strong likelihood of buying detergent.

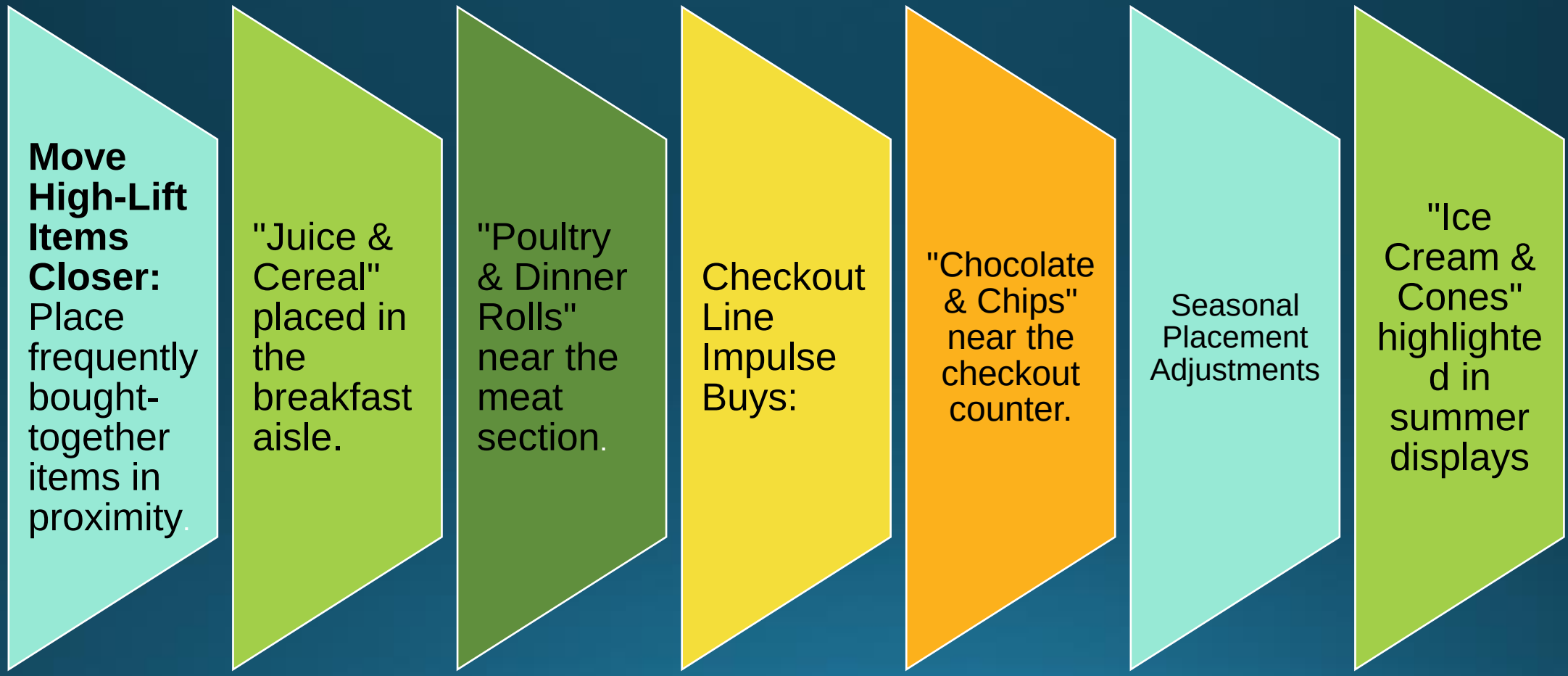
Lunch Meat → Poultry (Lift: **1.195**)

- Customers who buy Lunch Meat are **1.195 times more likely** to buy Poultry.

Juice → Poultry (Lift: **1.184**)

- Juice and Poultry are frequently purchased together.

Store Layout Optimization



Business Recommendations

Place **Poultry near Dinner Rolls, Sugar, and Soap** in stores.

Offer bundle discounts like **"Buy Dinner Rolls, Get a Discount on Poultry."**

Use **personalized online recommendations** (E.g., "Customers who bought Poultry also bought Dinner Rolls").

Position Soda and Mixes near Dishwashing Detergent for cross-selling.

Create promotions such as **"Buy Soda, Get a Discount on Detergent."**

BOGO & Discount Strategy

Combo Offers (Bundling)

Pre-Packaged Meal Deals:

Breakfast Bundle→ Cereal + Milk + Juice at a **discounted price**.

Movie Night Combo → Popcorn + Soda + Nachos at **10% off**.

Cross-Selling Promotions:

Shampoo & Conditioner offered as a **discounted bundle.**

Festival or Seasonal Combos:

Holiday Baking Kit → Flour + Sugar + Butter + Vanilla Essence at a **special price**.

Expected Business Impact

Conclusion


Optimized store layout enhances customer experience.
BOGO & discounts drive repeat purchases.
Bundling increases transaction value.

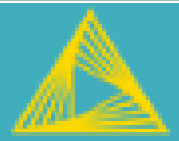
Working Files

Jupyter Note Book


RFM.ipynb


MBA.ipynb

Knime Work Book


RFM.knwf

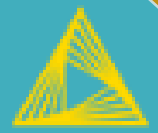
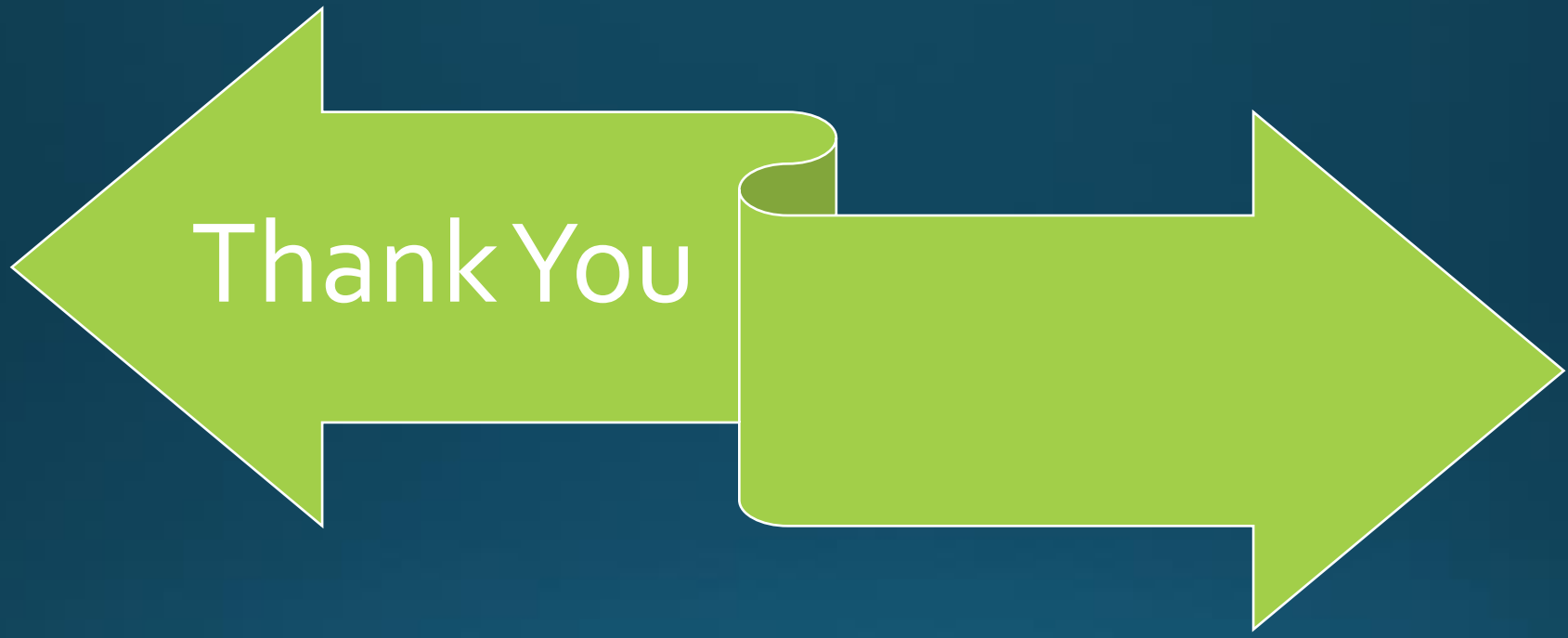

MBA.knwf

Tableau Work Book

RFM | Tableau Public

MBA | Tableau Public


Thank You