MARKETING AND RETAIL ANALYTICS PROJECT

BY: BENITA MERLIN.E

PGP-DATA SCIENCE AND BUSINESS ANALYTICS.

BATCH: PGP DSBA. O. MAY24.A

PART – A

Analyzing Customer Behavior & Sales Trends

RFM ANALYSIS

CONTENTS

BUSINESS CONTEXT

The automobile parts manufacturing company has been selling to a diverse customer base for the past **3 years**.

They lack in-house expertise to derive insights from transaction data.

By leveraging data-driven insights, the company can **stay ahead of competitors** by offering **personalized marketing, optimized inventory management, and better customer engagement**. This will help in **enhancing customer loyalty** and increasing **repeat purchases**, ultimately driving long-term business growth.

Objective of Analysis:

Identify patterns in customer purchasing behavior.

Segment customers based on their transaction history (RFM Analysis).

Provide **actionable insights** for targeted marketing.

Improve **customer retention & sales growth**.

PROBLEM STATEMENT

Identify hidden patterns in customer purchases.

Segment customers based on purchasing behavior.

Provide actionable insights to optimize marketing strategies.

Recommend personalized approaches to maximize sales and retention.

Data Overview

•Dataset includes **3 years of transactional data**.

Key columns in the dataset:

- Ordernumber, Orderdate, Sales, Quantityordered
- Customername, Productline, Msrp, Dealsize, Status
- Days_since_lastorder

Data Cleaning Steps Taken:

• Checked **duplicate records**
• Checked **missing values**
• Converted **dates to proper format**

Exploratory Data Analysis

STATISTICS SUMMARY

```
Summary Statistics:
         ORDERNUMBER  QUANTITYORDERED    PRICEEACH  ORDERLINENUMBER  \
count   2747.000000      2747.000000  2747.000000      2747.000000
mean   10259.761558        35.103021   101.098951         6.491081
min    10100.000000         6.000000    26.880000         1.000000
25%    10181.000000        27.000000    68.745000         3.000000
50%    10264.000000        35.000000    95.550000         6.000000
75%    10334.500000        43.000000   127.100000         9.000000
max    10425.000000        97.000000   252.870000        18.000000
std       91.877521         9.762135    42.042548         4.230544

              SALES                     ORDERDATE  DAYS_SINCE_LASTORDER  \
count   2747.000000                          2747           2747.000000
mean    3553.047583  2019-05-13 21:56:17.211503360           1757.085912
min      482.130000           2018-01-06 00:00:00             42.000000
25%     2204.350000           2018-11-08 00:00:00           1077.000000
50%     3184.800000           2019-06-24 00:00:00           1761.000000
75%     4503.095000           2019-11-17 00:00:00           2436.500000
max    14082.800000           2020-05-31 00:00:00           3562.000000
std     1838.953901                           NaN            819.280576

              MSRP
count  2747.000000
mean    100.691664
min      33.000000
25%      68.000000
50%      99.000000
75%     124.000000
max     214.000000
std      40.114802
```

Order Volume & Sales:

The **average sales per order** is **$3,553**, with a **maximum of $14,082**.

The **median sales value** is **$3,184.80**, suggesting a right-skewed distribution.

- **Quantity & Pricing:**

The **average quantity ordered** is **35**, with a max of **97**.

Price per unit varies significantly (Min: $26.88, Max: $252.87), indicating different product categories.

- **Order Frequency & Gaps:**

Customers **wait an average of ~1,757 days (~4.8 years)** between orders, with some taking as long as **3,562 days (~9.8 years)**.

The **median time between orders is ~1,761 days**, showing **infrequent repeat purchases**.

- **Trends Over Time:**

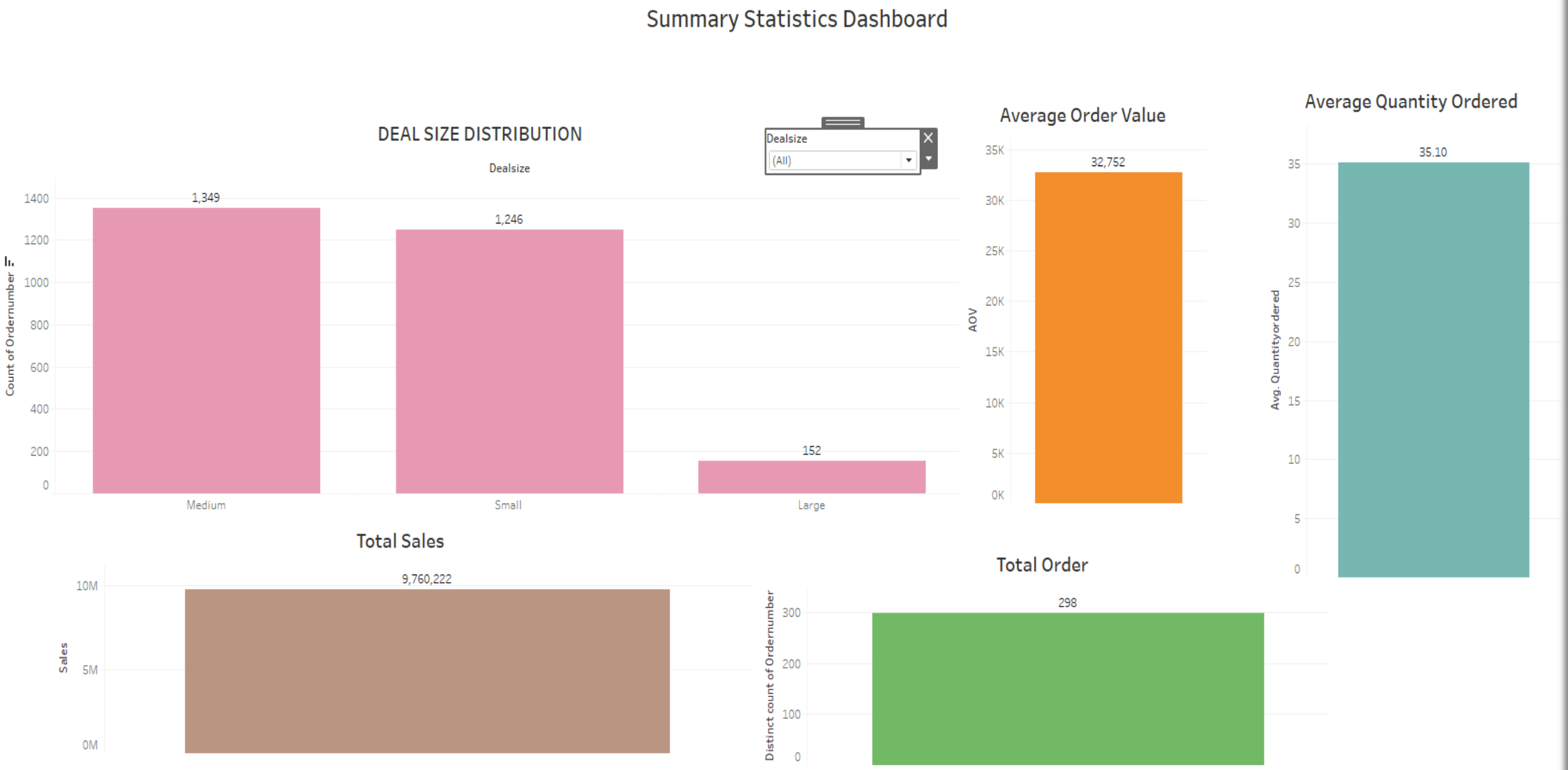
Orders were placed between **January 2018 - May 2020**.

Most sales seem concentrated in later periods (2019-2020).

- **MSRP (Product Pricing):**

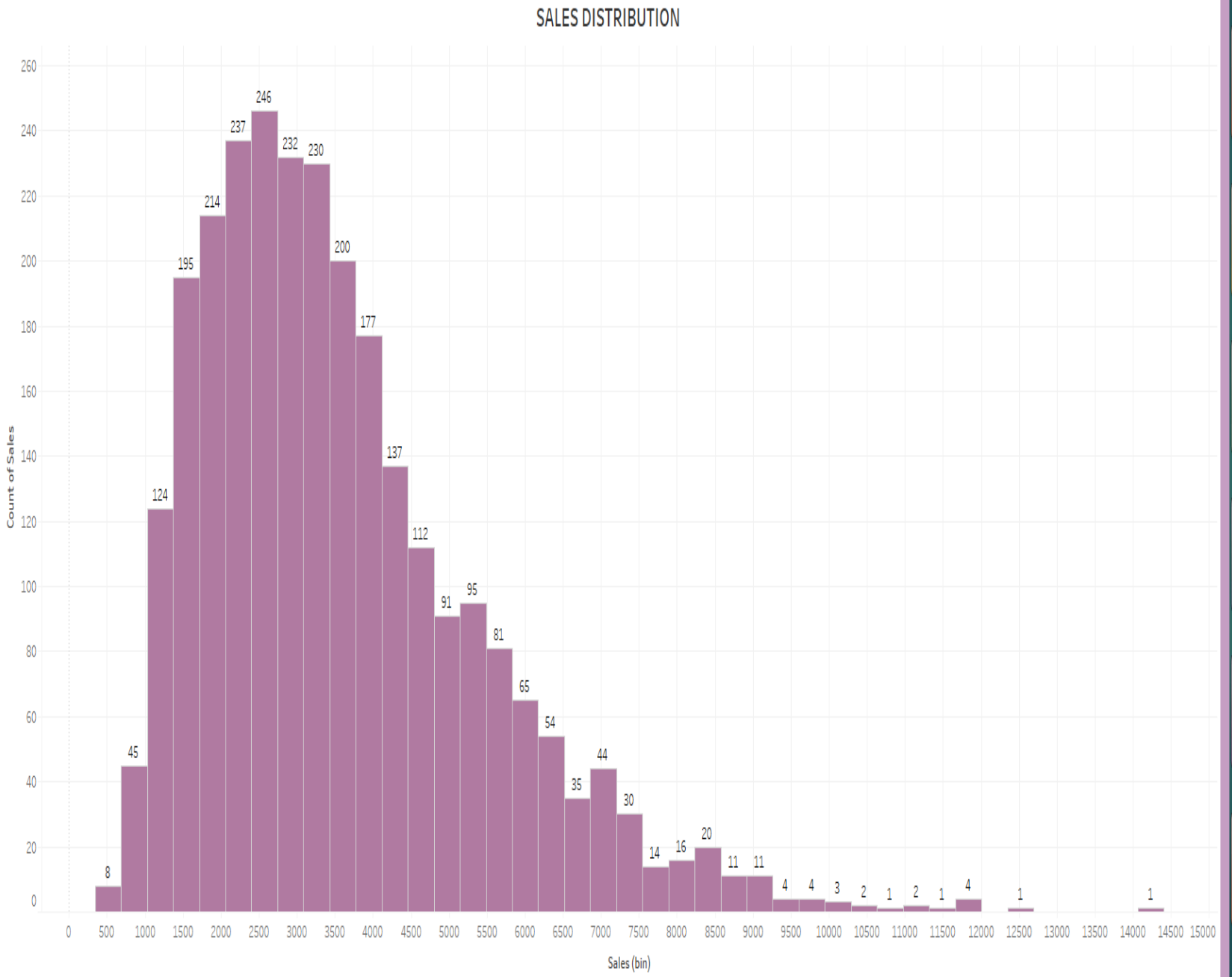
The **average MSRP is $100.69**, but some products reach up to **$214**, reflecting a wide range of product pricing.

STATISTICS SUMMARY



UNIVARIATE ANALYSIS

SALES DISTRIBUTION



Right-Skewed Distribution: Most sales are in the lower range, with fewer high-value sales.

Peak Sales Range: Highest sales occur between **2000–3500**, peaking at **2500–3000 (246 sales)**.

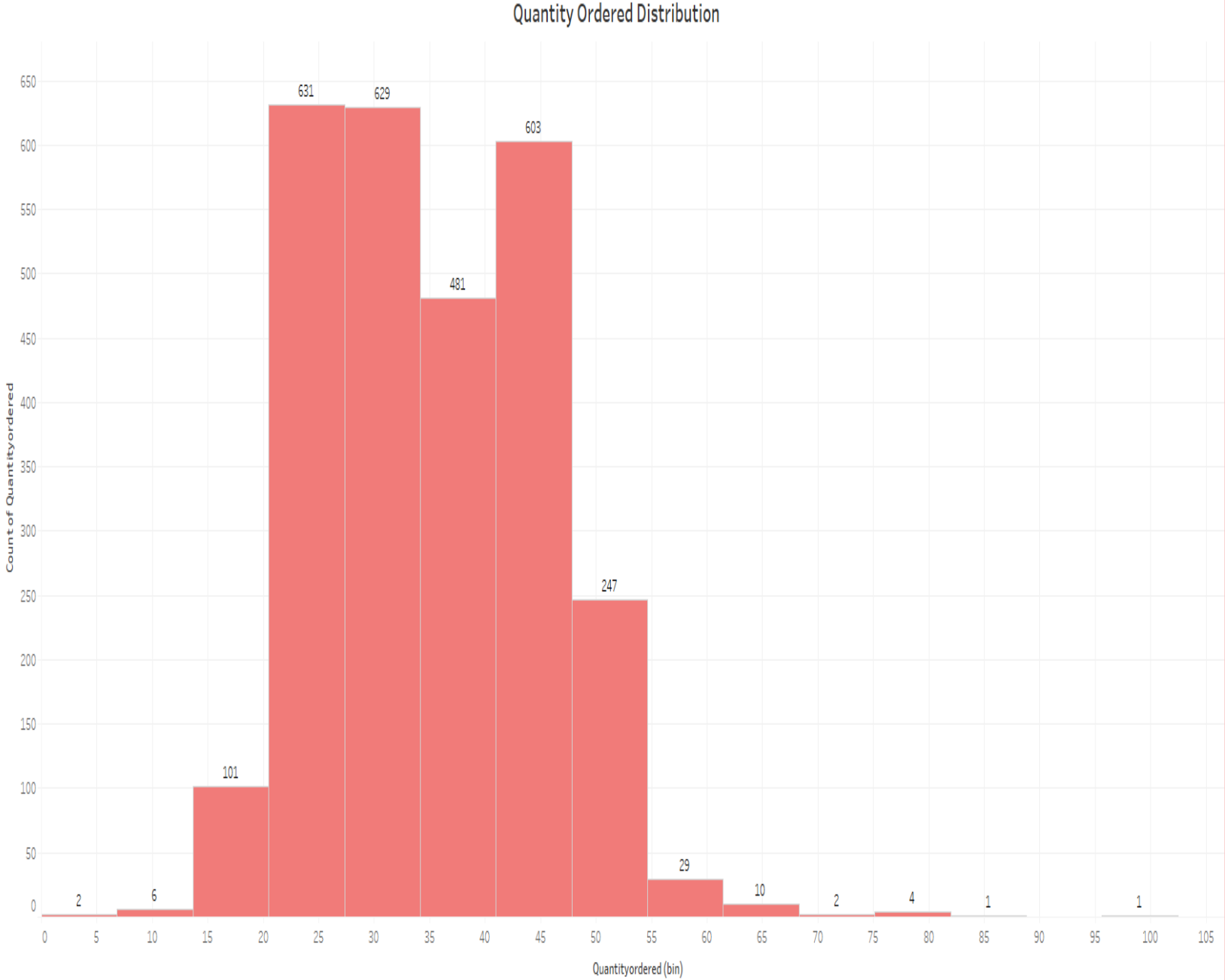
Declining Trend: Sales gradually decrease beyond **3500**.

Low High-Value Sales: Sales above **8000** are rare.

Outliers: A few extreme cases beyond **12000 sales**.

Business Insight: Focus marketing on **1000–4000** sales range for maximum impact.

QUANTITY ORDERED DISTRIBUTION



Bimodal Distribution: Two major peaks around **25–30** and **40–45** quantities ordered.

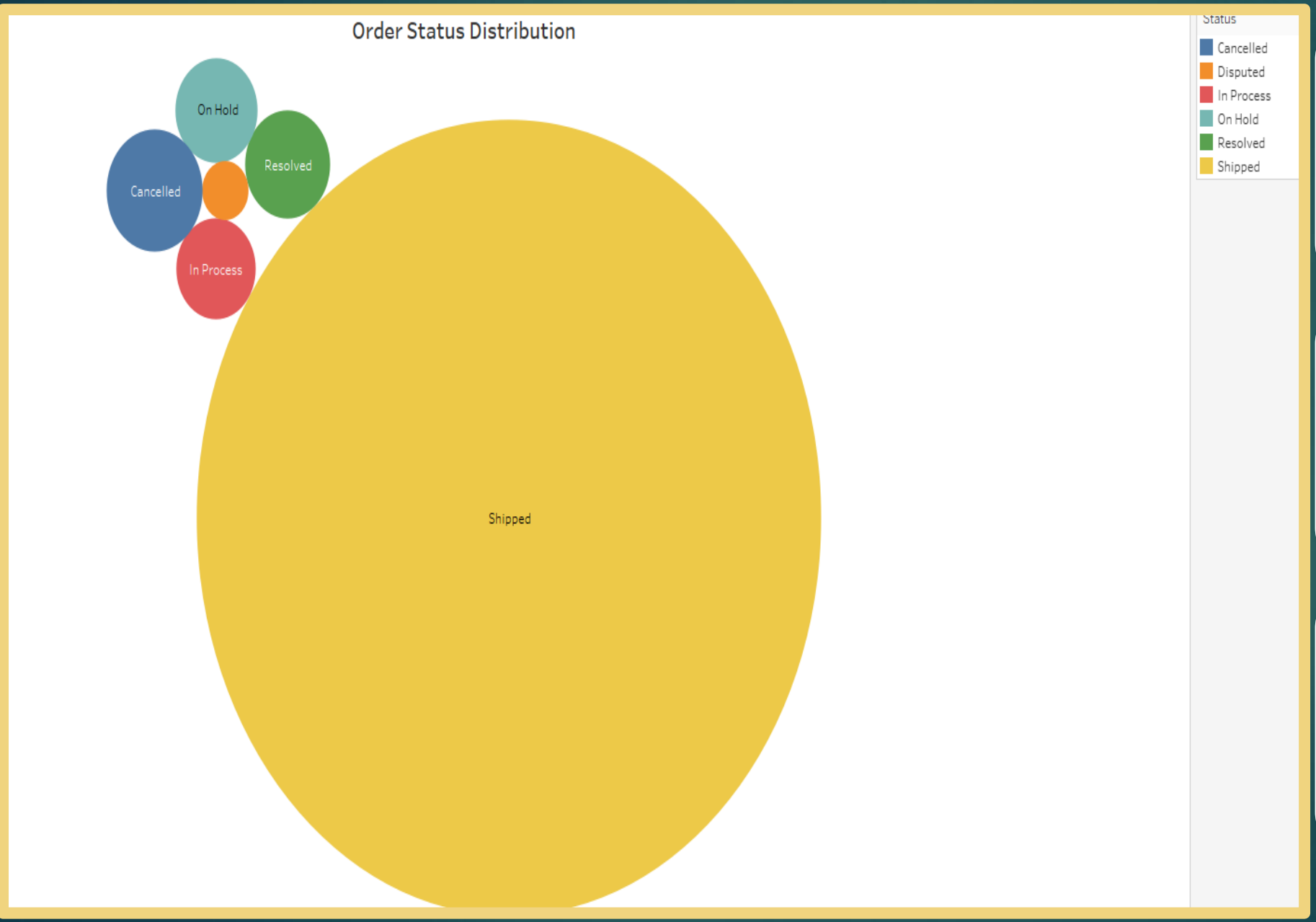
Most Common Orders: Majority of orders are between **20–45**, with highest counts at **25–30 (631 orders)** and **40–45 (603 orders)**.

Lower Order Volumes: Very few transactions below **15** or above **55** quantities.

Rare Large Orders: Quantities exceeding **60** are uncommon, indicating limited bulk purchases.

Business Insight: Focus on optimizing inventory for **20–45 quantity orders**, as they drive most sales

ORDER STATUS DISTRIBUTION



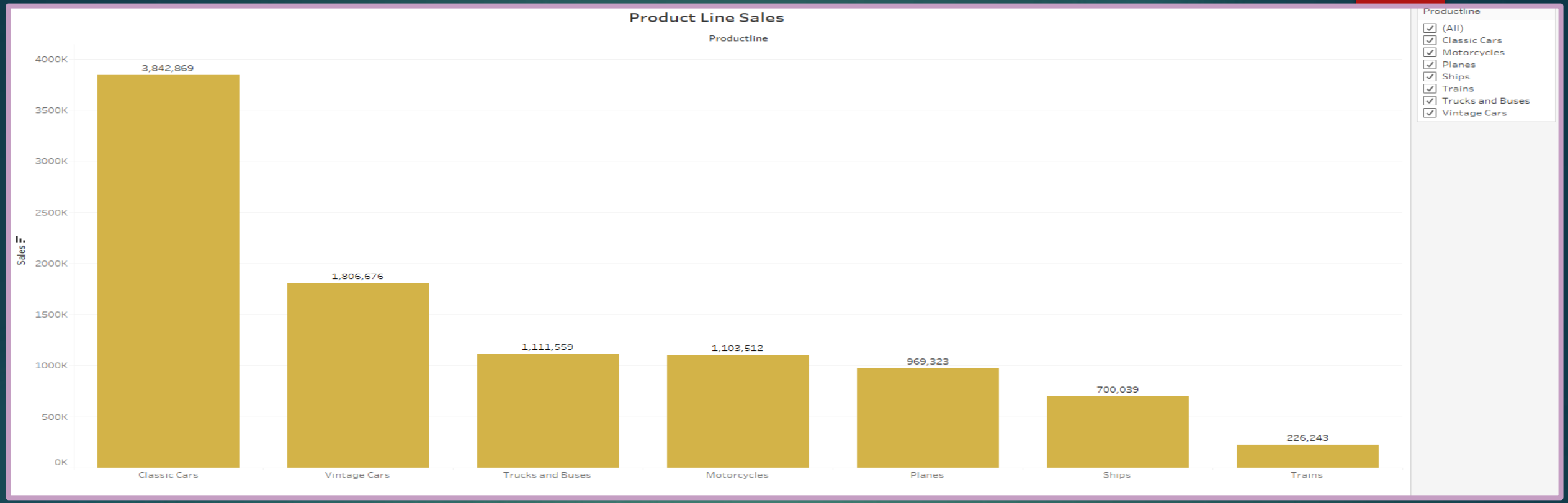
Majority Shipped: Most orders are successfully shipped, indicating efficient logistics.

Cancellations Exist: A notable portion is canceled, requiring analysis to reduce losses.

Minimal Issues: Few orders are **on hold, disputed, or in process**, showing a smooth order flow

Bivariate Analysis & Multivariate Analysis

PRODUCT LINE SALES ANALYSIS



Top Performer: **Classic Cars** lead the sales with **$3.84M**, contributing the most revenue.

Strong Demand: **Vintage Cars** follow with **$1.81M**, showing a high consumer preference.

Mid-Tier Sales: **Trucks & Buses ($1.11M)** and **Motorcycles ($1.10M)** have similar sales volumes.

Lower Sales: **Planes ($969K)** and **Ships ($700K)** have moderate demand.

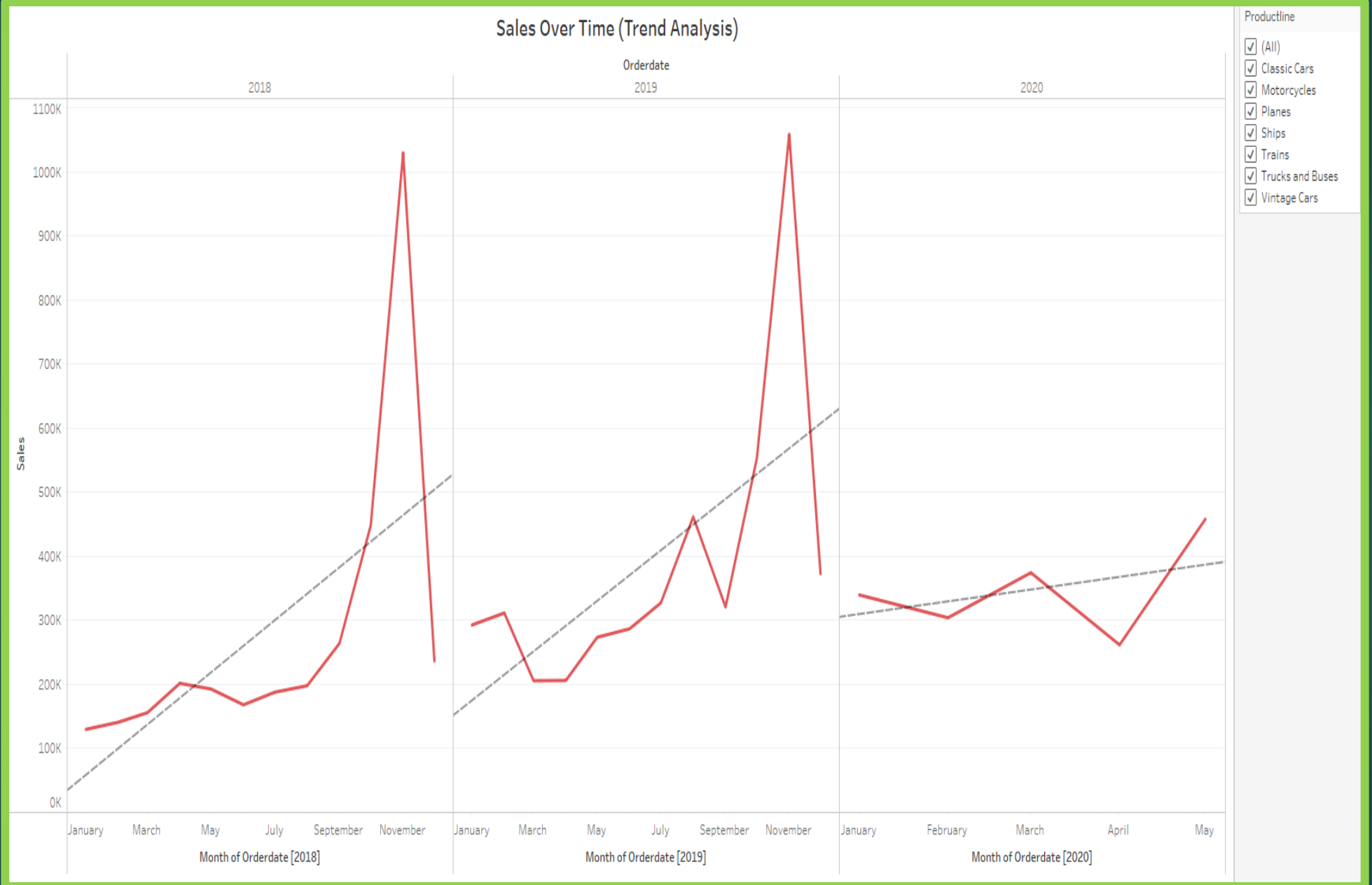
Least Sales: **Trains ($226K)** contribute the least, indicating a niche or underperforming product.

Business Insight:

Focus on expanding inventory and marketing for **Classic and Vintage Cars** to maximize revenue.

Evaluate strategies to boost sales for **Trains and Ships**, such as promotions or product diversification.

SALES TREND (YEAR & MONTH)



Overall Growth

Sales show an increasing trend over time, as indicated by the upward trendline.

Seasonal Peaks

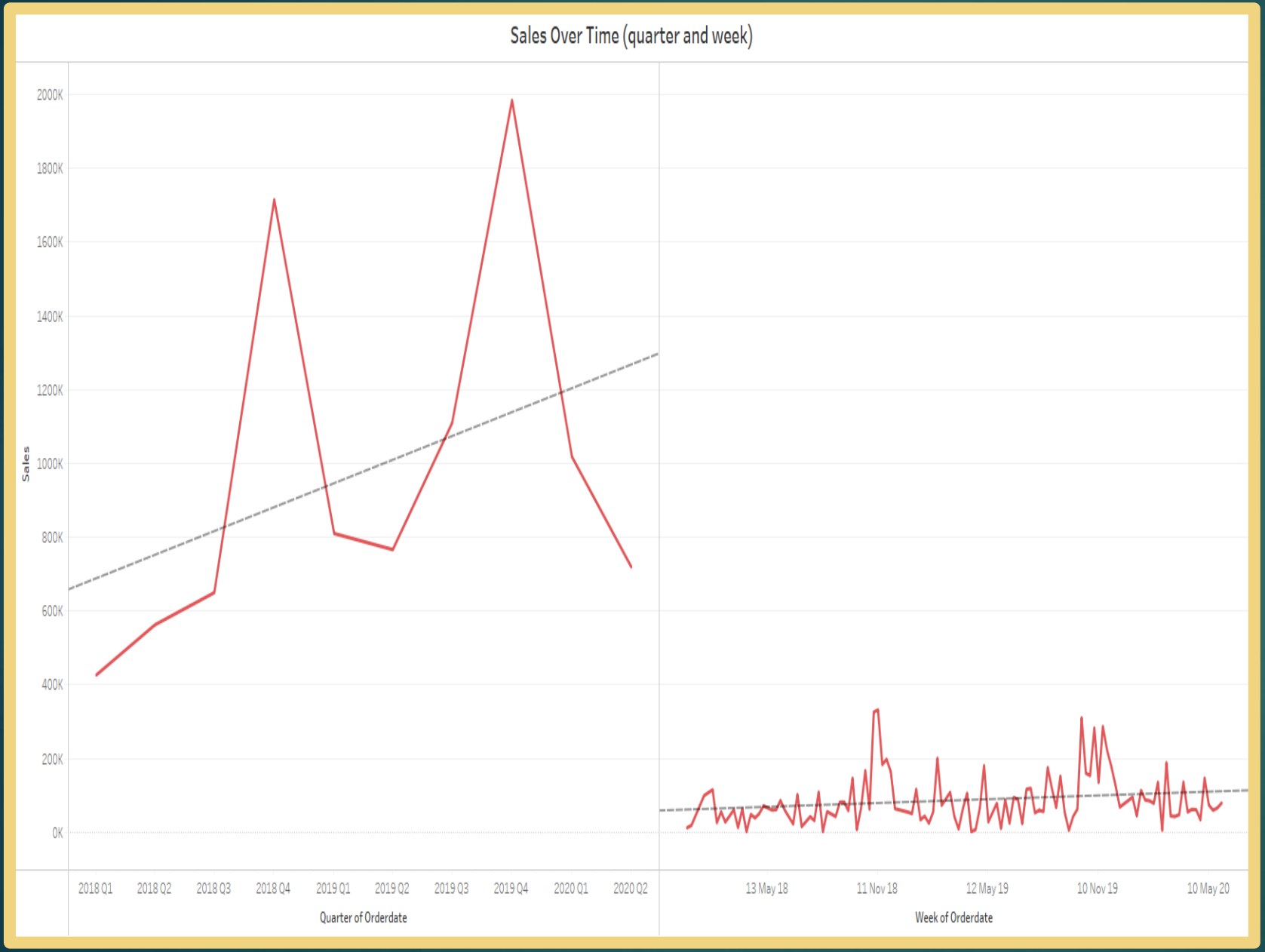
2018 & 2019: Sharp sales spikes around **October–November**, indicating high seasonal demand.

2020: Sales growth is steady but lacks the sharp peaks seen in previous years.

Post-Peak Decline

Noticeable drop in sales after the peak, suggesting cyclical or seasonal patterns.

SALES TREND (QUARTER & WEEK)

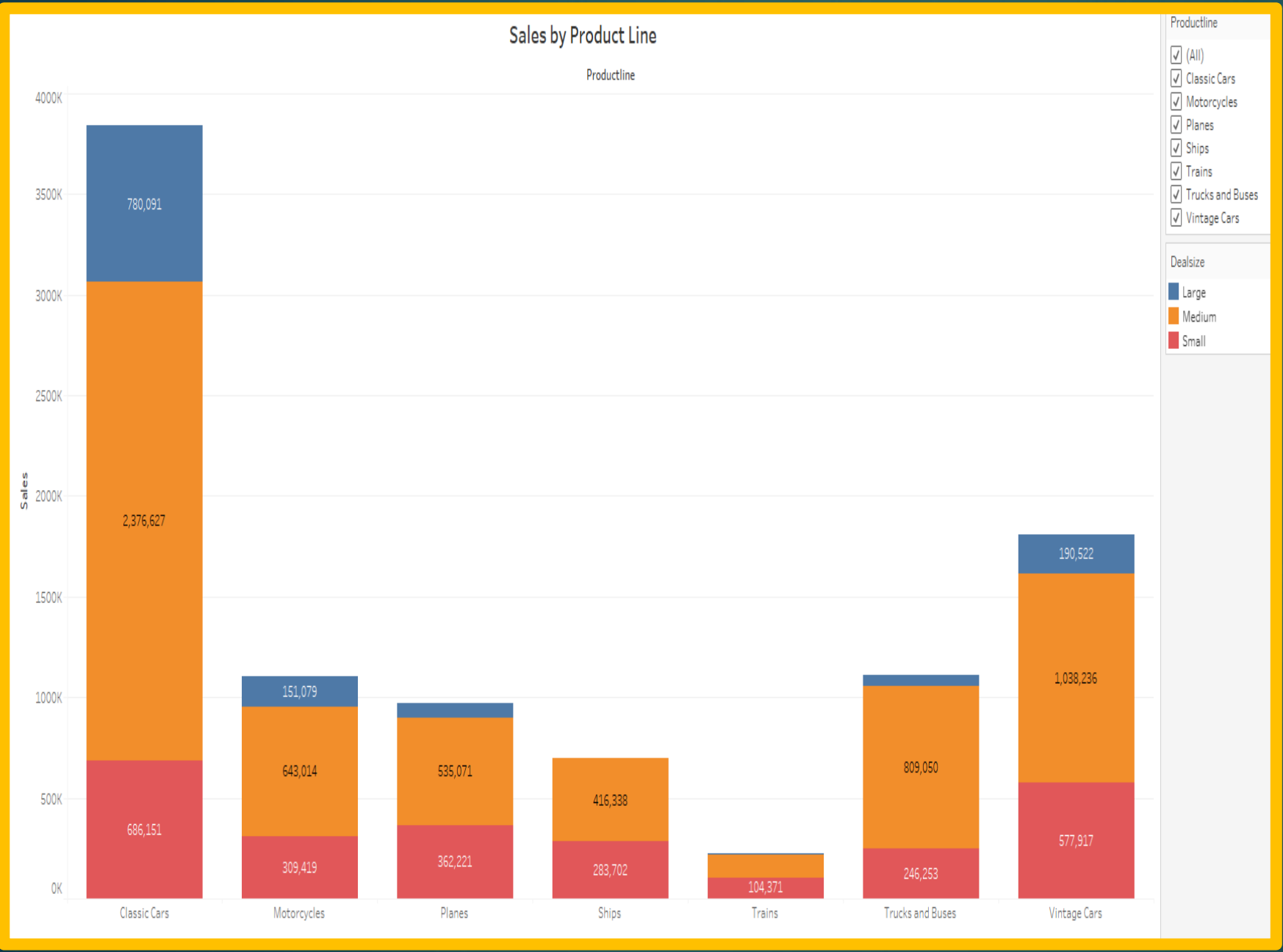


Quarterly: Sales peak in **Q4 (2018 & 2019)**, showing a seasonal demand surge. **Q2 2020 decline** needs further analysis.

Weekly: Sales fluctuate with **occasional spikes**, indicating promotional or seasonal effects.

Overall: Sales show an **upward trend**, but **seasonality plays a key role** in driving performance.

SALES BY PRODUCT LINE



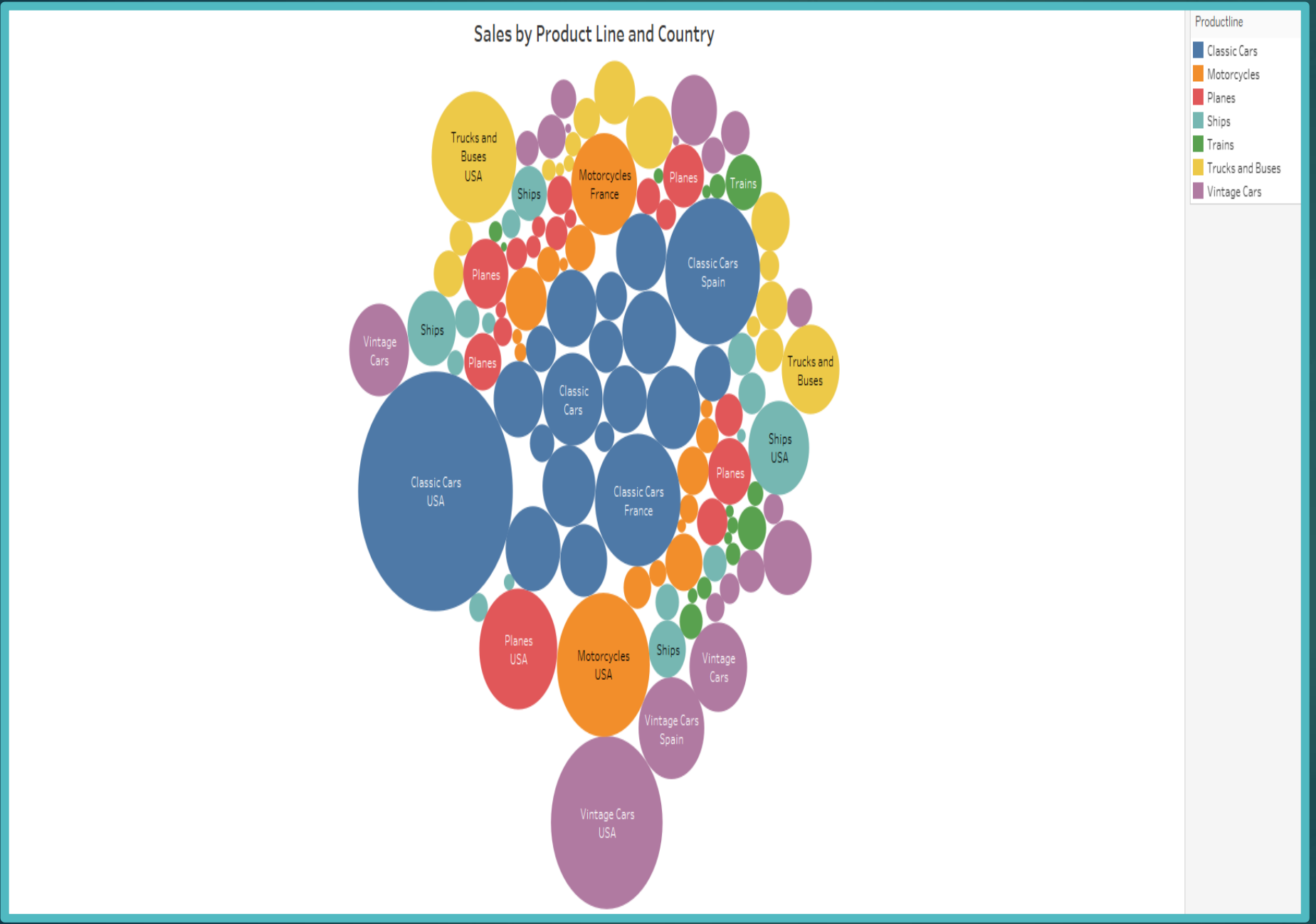
Top Performer: Classic Cars dominate sales with the highest revenue, followed by **Vintage Cars**.

Deal Size Impact: Medium-sized deals contribute the most across all product lines. **Large deals** are significant for Classic and Vintage Cars.

Least Sales: Trains have the lowest sales volume, suggesting lower demand.

Recommendation: Focus on boosting sales for **low-performing categories** while maintaining strong performance in **Classic and Vintage Cars**

SALES BY PRODUCT LINE AND COUNTRY



Top Country: USA dominates sales, especially in **Classic Cars, Vintage Cars, and Motorcycles**.

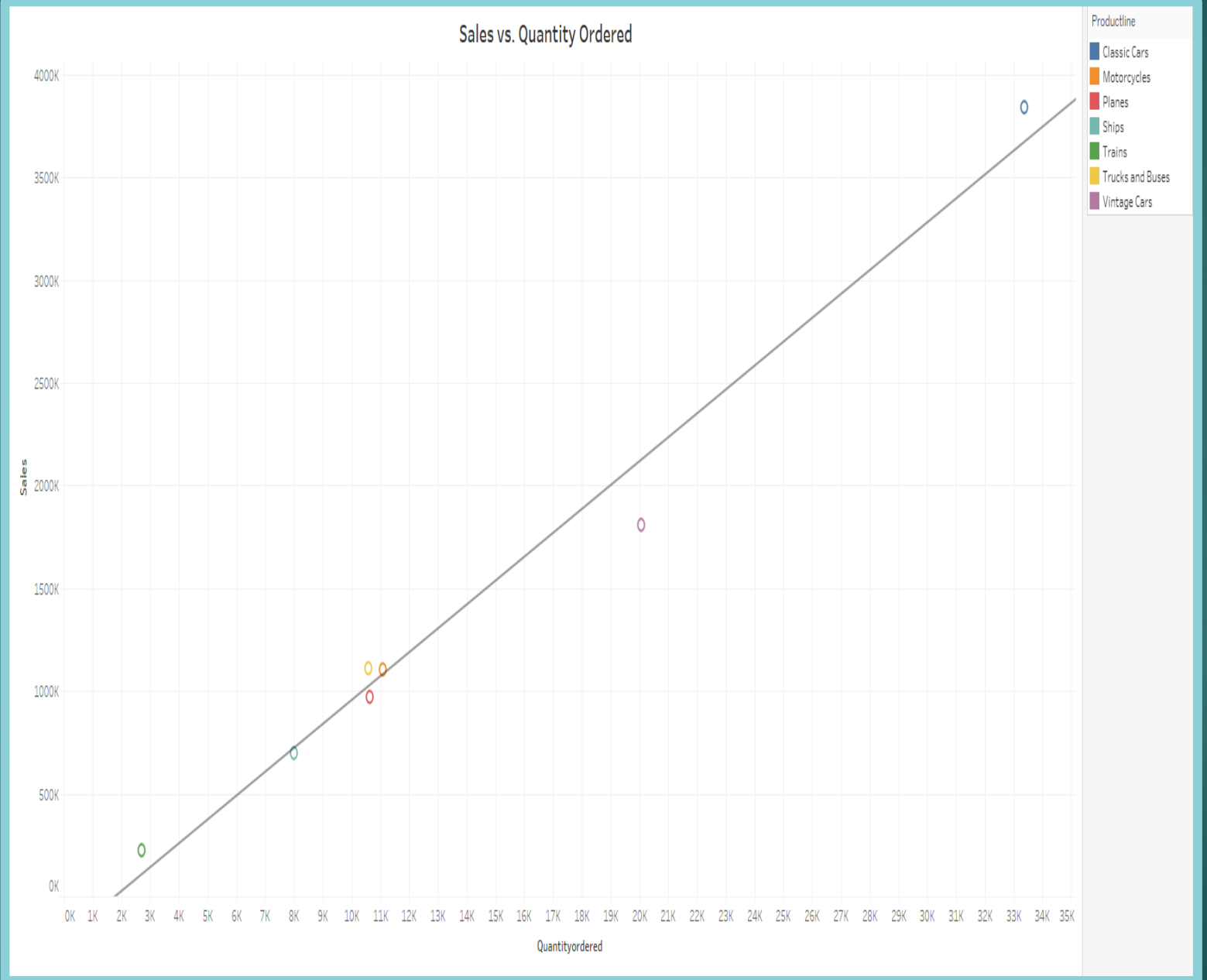
Best-Selling Product Line: Classic Cars lead in multiple countries, including the USA, France, and Spain.

Emerging Markets: Vintage Cars have significant sales in Spain and the USA, indicating a growing interest.

Diverse Product Demand: Trucks and Buses, Planes, and Ships have smaller but distributed sales across different regions.

Recommendation: Strengthen marketing and sales strategies in the USA and expand high-performing product lines in France and Spain.

SALES VS. QUANTITY ORDERED SCATTER PLOT



Strong Positive Correlation:The data points closely follow the trend line, indicating that **higher quantities ordered lead to higher sales**.

This suggests **consistent pricing across product lines**.

Top-Performing Product Line:

Classic Cars (blue point at the top-right) has the **highest quantity ordered and highest sales**, making it the most profitable category.

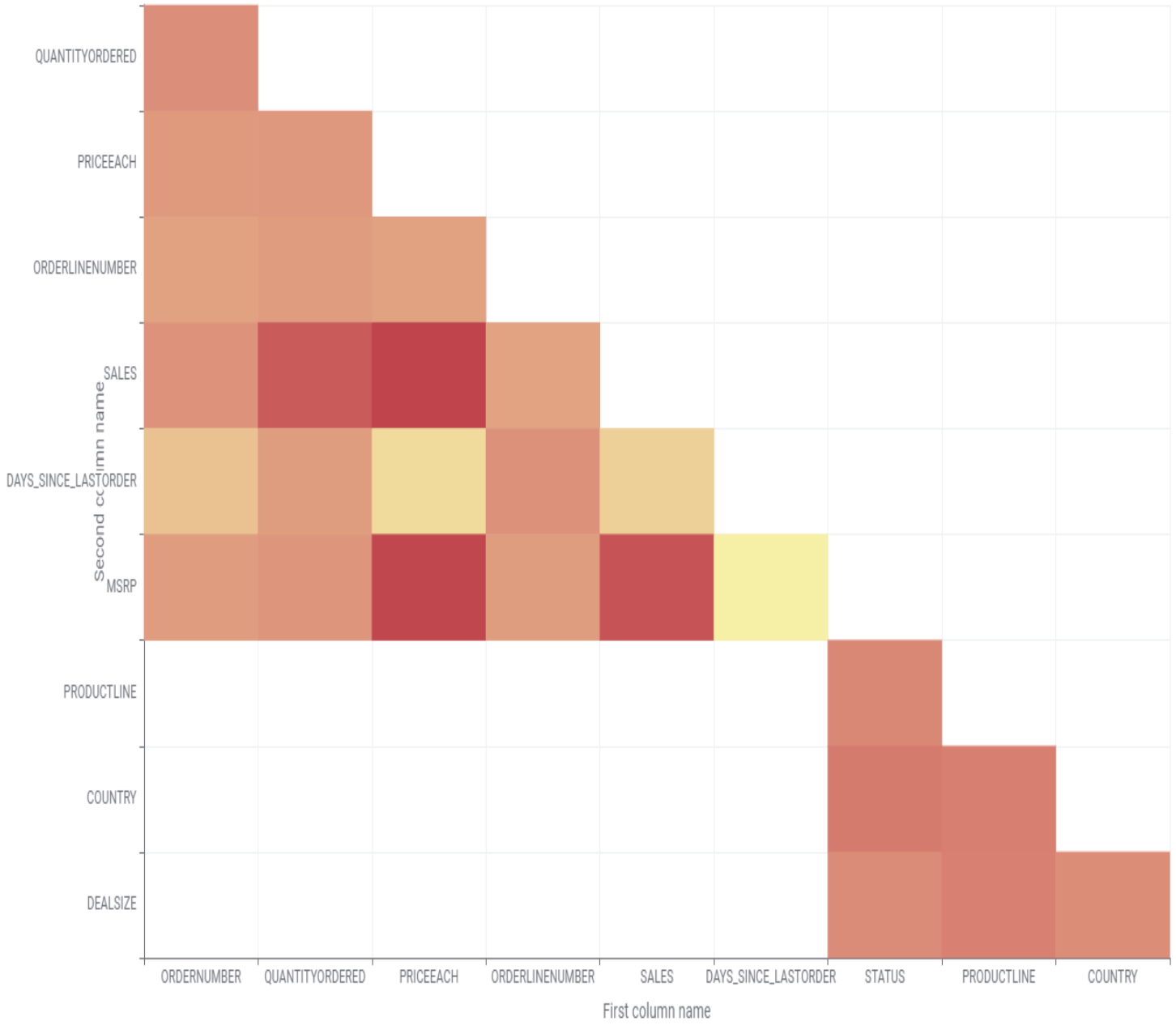
Underperforming Categories:

Trains and Ships are at the lower end of both sales and quantity ordered, indicating **lower demand or higher costs per unit**.

Mid-Tier Product Lines:

Motorcycles, Planes, and Trucks & Buses have moderate sales and quantities, showing **steady performance**.

HEATMAP



Strong Correlations:

Sales & Quantity Ordered: More quantity leads to higher sales.

Sales & MSRP: Higher-priced items drive more revenue.

Weak Correlations:

Days Since Last Order & Sales: No strong impact.

MSRP & Order Frequency: Pricing doesn't affect repeat orders much.

Takeaways:

Focus on **high-performing products** for maximum sales.

Offer **discounts on high-priced items** with low order volume.

Optimize **inventory & pricing based on product line trends**.

CUSTOMER SEGMENTATION USING RFM ANALYSIS

Recency (R): How recently a customer made a purchase.

Frequency (F): How often a customer makes a purchase.

Monetary (M): How much a customer spends.

Customers are scored and segmented based on these three metrics.

UNDERSTANDING RFM ANALYSIS

RFM Metrics Calculation:

Recency (R): Number of days since the last purchase.

Frequency (F): Total number of transactions made by a customer.

Monetary (M): Total amount spent by a customer.

Assumptions:

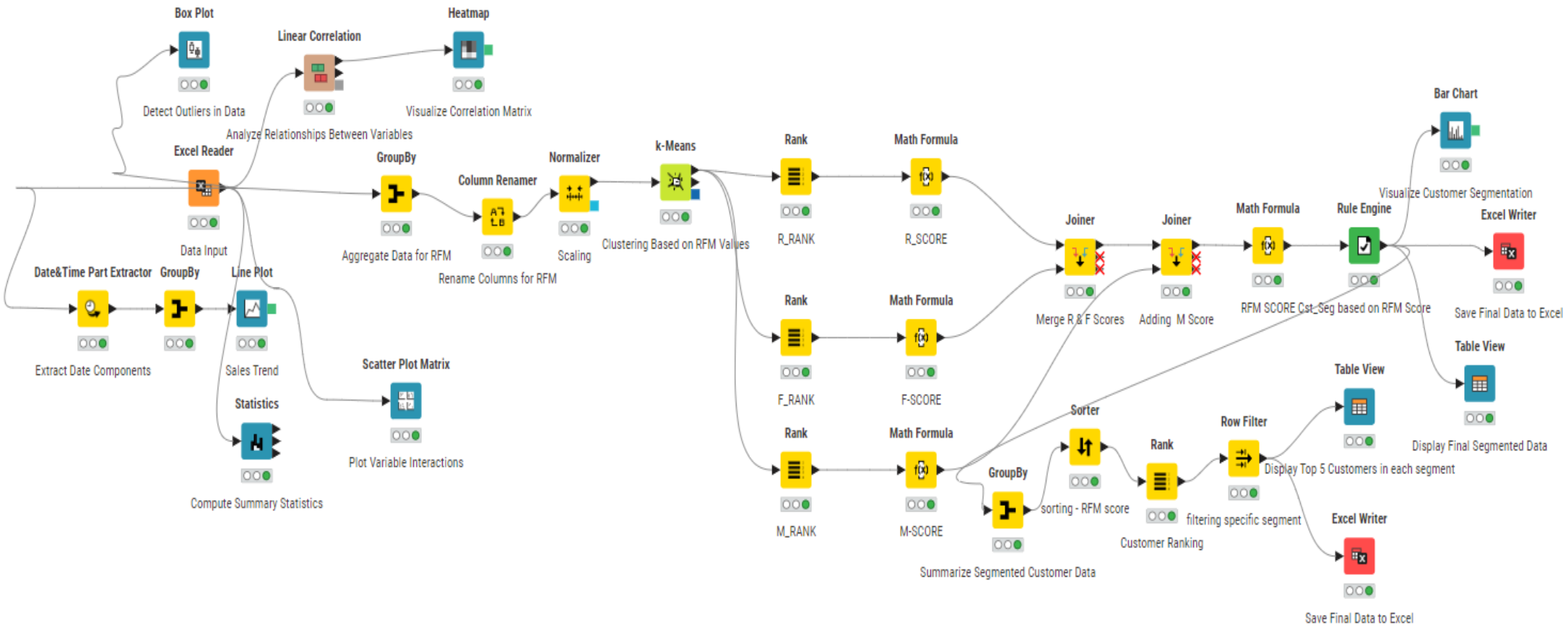
More recent purchases indicate an engaged customer.

Higher purchase frequency signals loyalty.

Higher spending suggests high-value customers.

Customers with low scores in all three categories are at risk of churning.

KNIME WORKFLOW



THREE METRICES: RECENCY, FREQUENCY AND MONETARY

▶ 1: Output Table | Flow Variables

Rows: 89 | Columns: 4

Table | Statistics

#	RowID	CUSTOMERNAME String	MONETARY Number (double)	RECENCY Number (integer)	FREQUENCY Number (integer)
1	Row0	AV Stores, Co.	157,807.81	421	3
2	Row1	Alpha Cognac	70,488.44	675	3
3	Row2	Amica Models & Co.	94,117.26	328	2
4	Row3	Anna's Decorations, Ltd	153,996.13	131	4
5	Row4	Atelier graphique	24,179.96	312	3
6	Row5	Australian Collectables, Ltd	64,591.46	1018	3
7	Row6	Australian Collectors, Co.	200,995.41	229	5
8	Row7	Australian Gift Network, Co	59,469.12	190	3
9	Row8	Auto Assoc. & Cie.	64,834.32	275	2
10	Row9	Auto Canal Petit	93,170.66	127	3
11	Row10	Auto-Moto Classics Inc.	26,479.26	1353	3
12	Row11	Baane Mini Imports	116,599.19	245	4
13	Row12	Bavarian Collectables Imports, Co.	34,993.92	801	1
14	Row13	Blauer See Auto, Co.	85,171.59	705	4
15	Row14	Boards & Toys Co.	9,129.35	410	2
16	Row15	CAF Imports	49,642.05	625	2
17	Row16	Cambridge Collectables Co.	36,163.62	484	2
18	Row17	Canadian Gift Exchange Network	75,238.92	364	2
19	Row18	Classic Gift Ideas, Inc	67,506.97	344	2
20	Row19	Classic Legends Inc.	77,795.2	309	3
21	Row20	Clover Collections, Co.	57,756.43	659	2
22	Row21	Collectable Mini Designs Co.	87,489.23	575	2

PROCESS IN KNIME WORKFLOW

- Calculate Recency, Frequency, and Monetary (RFM) values for each customer.
- Normalize RFM scores
- Apply Clustering Algorithm used K-Means to group customers into 4 clusters.
- Assign rankings for RFM metrics.
- Calculating RFM score using Math formula node.
- Using Rule Engine to *Segment Customers into Best, Loyal,* High Churn Risk *and Lost*
- Visualize customer segmentation (bar charts).

RFM SCORE

▶ 1: Output data Flow Variables

Rows: 89 | Columns: 12 Table | Statistics

#	RowID	CUSTOME... String	MONETARY Number (double)	RECENCY Number (double)	FREQUENCY Number (double)	Cluster String	R_rank Number (integer)	R_SCORE Number (double)	F_rank Number (integer)	F-SCORE Number (double)	M_rank Number (integer)	M_SCORE Number (double)	RFM_SCORE Number (double)
1	Row10	Auto-Moto Classi	0.019	1	0.08	cluster_3	1	1	3	3	3	1	5
2	Row64	Rovelli Gifts	0.143	0.755	0.08	cluster_3	2	1	3	3	74	4	8
3	Row5_	Australian Collec	0.061	0.744	0.08	cluster_3	3	1	3	3	18	1	5
4	Row24	Cruz & Sons Co.	0.094	0.709	0.08	cluster_3	4	1	3	3	49	3	7
5	Row35	Gift Ideas Corp.	0.053	0.69	0.08	cluster_3	5	1	3	3	15	1	5
6	Row40	Iberia Gift Import	0.05	0.658	0.04	cluster_3	6	1	2	2	13	1	4
7	Row70	Signal Collectible	0.045	0.606	0.04	cluster_3	7	1	2	2	11	1	4
8	Row56	Norway Gifts By I	0.078	0.597	0.04	cluster_3	8	1	2	2	35	2	5
9	Row12	Bavarian Collecta	0.029	0.579	0	cluster_3	9	1	1	1	6	1	3
10	Row46	Marseille Mini Au	0.073	0.545	0.08	cluster_3	10	1	3	3	27	2	6
11	Row66	Royale Belge	0.027	0.53	0.12	cluster_3	11	1	4	4	5	1	6
12	Row49	Mini Auto Werke	0.048	0.515	0.08	cluster_3	12	1	3	3	12	1	5
13	Row13	Blauer See Auto,	0.084	0.506	0.12	cluster_3	13	1	4	4	43	3	8
14	Row73	Stylish Desk Dec	0.088	0.503	0.08	cluster_3	14	1	3	3	47	3	7
15	Row1_	Alpha Cognac	0.068	0.483	0.08	cluster_3	15	1	3	3	23	2	6
16	Row28	Diecast Collectab	0.068	0.481	0.04	cluster_3	16	1	2	2	24	2	5
17	Row29	Double Decker Gi	0.03	0.479	0.04	cluster_3	17	1	2	2	7	1	4
18	Row20	Clover Collection	0.054	0.471	0.04	cluster_3	18	1	2	2	16	1	4
19	Row31	Enaco Distributor	0.077	0.471	0.08	cluster_3	18	1	3	3	33	2	6
20	Row71	Signal Gift Stores	0.082	0.469	0.08	cluster_3	19	1	3	3	39	2	6
21	Row30	Dragon Souvenie	0.181	0.463	0.16	cluster_3	20	1	5	5	84	5	11
22	Row82	Toys4GrownUps	0.106	0.463	0.08	cluster_3	20	1	3	3	58	3	7

CUSTOMER SEGMENTATION BASED ON RFM SCORE

View | Flow Variables

Table View

Open in new window

Rows: 89 | Columns: 9

RowID	CUSTOMERNAME (String)	R_rank (Number (integer))	R_SCORE (Number (double))	F_rank (Number (integer))	F-SCORE (Number (double))	M_rank (Number (integer))	M_SCORE (Number (double))	RFM_SCORE (Number (double))	CUST_SEGMENT (String)
Row10	Auto-Moto Classics Inc.	1	1	3	3	3	1	5	Customers on the Verge of Chu
Row64	Rovelli Gifts	2	1	3	3	74	4	8	Customers on the Verge of Chu
Row5_I	Australian Collectables, Ltd	3	1	3	3	18	1	5	Customers on the Verge of Chu
Row24	Cruz & Sons Co.	4	1	3	3	49	3	7	Customers on the Verge of Chu
Row35	Gift Ideas Corp.	5	1	3	3	15	1	5	Customers on the Verge of Chu
Row40	Iberia Gift Imports, Corp.	6	1	2	2	13	1	4	Lost Customers
Row70	Signal Collectibles Ltd.	7	1	2	2	11	1	4	Lost Customers
Row56	Norway Gifts By Mail, Co.	8	1	2	2	35	2	5	Customers on the Verge of Chu
Row12	Bavarian Collectables Imports,	9	1	1	1	6	1	3	Lost Customers
Row46	Marseille Mini Autos	10	1	3	3	27	2	6	Customers on the Verge of Chu
Row66	Royale Belge	11	1	4	4	5	1	6	Customers on the Verge of Chu
Row49	Mini Auto Werke	12	1	3	3	12	1	5	Customers on the Verge of Chu
Row13	Blauer See Auto, Co.	13	1	4	4	43	3	8	Customers on the Verge of Chu
Row73	Stylish Desk Decors, Co.	14	1	3	3	47	3	7	Customers on the Verge of Chu
Row1_I	Alpha Cognac	15	1	3	3	23	2	6	Customers on the Verge of Chu
Row28	Diecast Collectables	16	1	2	2	24	2	5	Customers on the Verge of Chu

RFM Segmentation done using Rule Engine Node

Best Customers: RFM Score **≥ 12**

Loyal Customers: RFM Score **9-11**

Churn Risk Customers: RFM Score **5-8**

Lost Customers: RFM Score **< 5**

CUSTOMER SEGMENTS IDENTIFIED

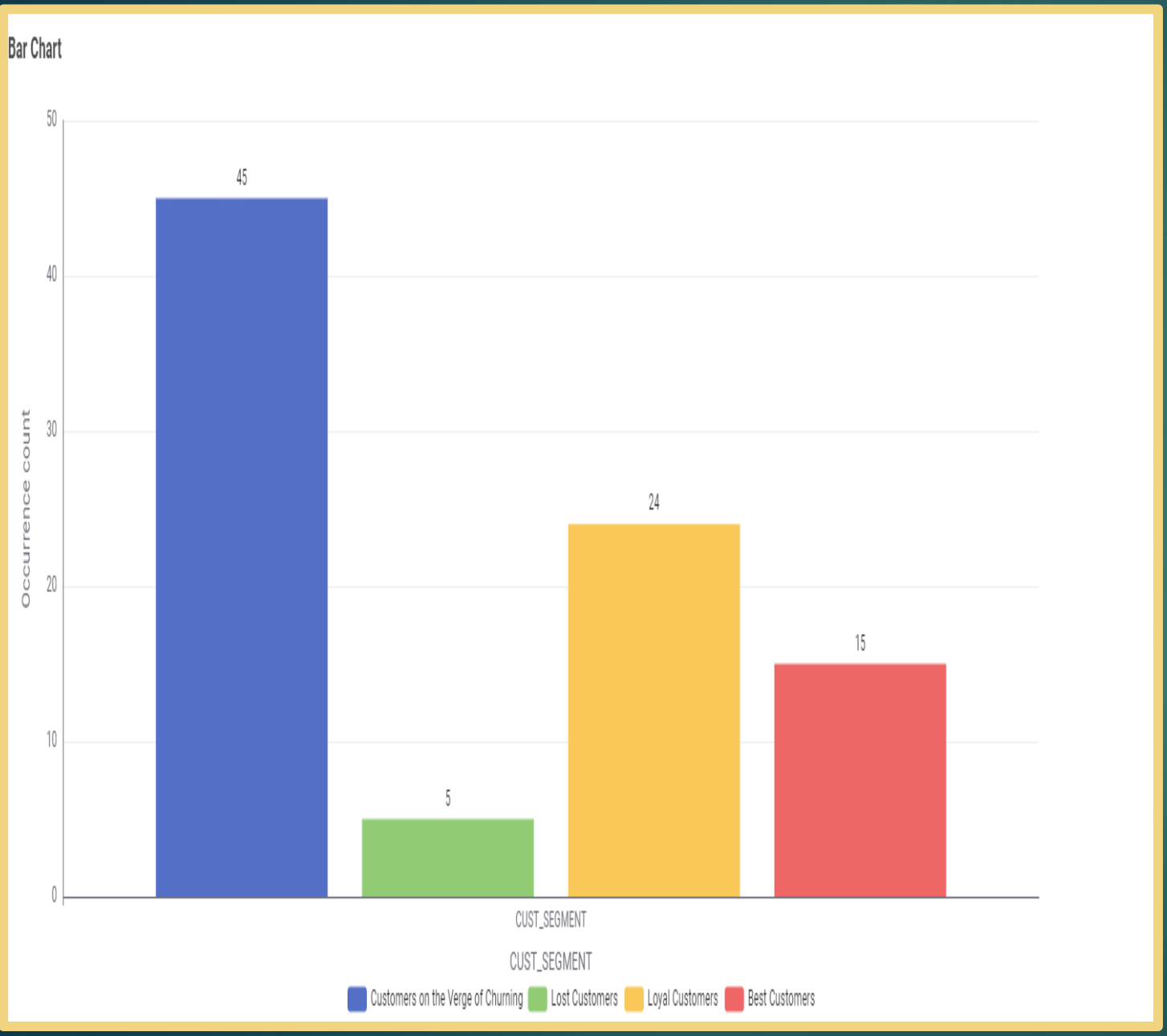
Best Customers: High Recency (R), High Frequency (F), and High Monetary (M) scores.

Loyal Customers: Regular buyers with high Frequency (F) and consistent Monetary (M) scores, even if Recency (R) is slightly lower

Customers on the Verge of Churn: These customers have slightly better scores but are at risk of leaving if not engaged properly.

Lost Customers: These customers have low RFM scores, meaning they haven't purchased recently, buy infrequently, and have low spending.

VISUALIZE CUSTOMER SEGMENTATION



High Churn Risk (45 Customers):

The largest segment consists of customers who are on the verge of churning.

This group needs immediate engagement strategies like personalized offers, re-engagement emails, and loyalty incentives.

Loyal Customers (24 Customers):

This segment includes customers who make frequent purchases but may not always have the highest spending.

They should be nurtured through loyalty programs, exclusive discounts, and appreciation rewards.

Best Customers (15 Customers):

These are the highest-value customers with frequent and high spending.

Retention strategies should focus on VIP services, priority support, and premium experiences to maintain engagement.

Lost Customers (5 Customers):

A small segment of customers who have already stopped engaging.

Win-back strategies such as special comeback offers, targeted ads, and reactivation campaigns can be useful.

CUSTOMER SEGMENTS AND JUSTIFICATION

Segment Name	Count	Justification
Best Customers	15	These customers have high frequency and monetary value, indicating strong brand loyalty and consistent spending.
Loyal Customers	24	Customers who purchase frequently but may not always spend the highest amount. They show consistent engagement.
Customers on the Verge of Churn	45	These customers have lower frequency and monetary value. They are still engaged but need intervention to prevent churn.
Lost Customers	5	Customers with very low recency, frequency, and monetary scores. They have stopped engaging with the business.

RFM ANALYSIS FINAL OUTPUT

View | Flow Variables

Table View

Open in new window

Rows: 89 | Columns: 13

RowID	CUSTOMERNA... String	MONETARY Number (double)	RECENCY Number (double)	FREQUENCY Number (double)	Cluster String	R_rank Number (integer)	R_SCORE Number (double)	F_rank Number (integer)	F-SCORE Number (double)	M_rank Number (integer)	M_SCORE Number (double)	RFM_SCORE Number (double)	CUST_SEGMENT String
Row10	Auto-Moto Classics I	0.019	1	0.08	cluster_3	1	1	3	3	3	1	5	Customers on the Ve
Row64	Rovelli Gifts	0.143	0.755	0.08	cluster_3	2	1	3	3	74	4	8	Customers on the Ve
Row5_I	Australian Collectable	0.061	0.744	0.08	cluster_3	3	1	3	3	18	1	5	Customers on the Ve
Row24	Cruz & Sons Co.	0.094	0.709	0.08	cluster_3	4	1	3	3	49	3	7	Customers on the Ve
Row35	Gift Ideas Corp.	0.053	0.69	0.08	cluster_3	5	1	3	3	15	1	5	Customers on the Ve
Row40	Iberia Gift Imports, C	0.05	0.658	0.04	cluster_3	6	1	2	2	13	1	4	Lost Customers
Row70	Signal Collectibles Lt	0.045	0.606	0.04	cluster_3	7	1	2	2	11	1	4	Lost Customers
Row56	Norway Gifts By Mail,	0.078	0.597	0.04	cluster_3	8	1	2	2	35	2	5	Customers on the Ve
Row12	Bavarian Collectables	0.029	0.579	0	cluster_3	9	1	1	1	6	1	3	Lost Customers
Row46	Marseille Mini Autos	0.073	0.545	0.08	cluster_3	10	1	3	3	27	2	6	Customers on the Ve
Row66	Royale Belge	0.027	0.53	0.12	cluster_3	11	1	4	4	5	1	6	Customers on the Ve
Row49	Mini Auto Werke	0.048	0.515	0.08	cluster_3	12	1	3	3	12	1	5	Customers on the Ve

INFERENCES FROM RFM ANALYSIS

TOP 5 BEST CUSTOMERS

Table View

Open in new window

Rows: 5 | Columns: 6

RowID	MONETARY *Number (double)*	RECENCY *Number (double)*	FREQUENCY *Number (double)*	CUST_SEGMENT *String*	First*(CUSTOMERNAME) *String*	Ranking Segment *Number (integer)*
Ro'	MONETARY	RECENCY	FREQUENCY	1 selected	First*(CUSTOMERNAME)	Ranking Segment
Row63	0.119	0.155	0.12	Best Customers	Baane Mini Imports	1
Row65	0.121	0.026	0.08	Best Customers	UK Collectables, Ltd.	2
Row66	0.123	0.166	0.12	Best Customers	Tokyo Collectables, Ltd	3
Row68	0.124	0.152	0.12	Best Customers	Technics Stores Inc.	4
Row69	0.125	0.142	0.12	Best Customers	Diecast Classics Inc.	5

All the customers in the table belong to the **"Best Customers"** segment, indicating they have high RFM scores.

These customers have frequent transactions, high monetary value, and recent purchases

The highest-ranking customer is **Baane Mini Imports**, followed by **UK Collectables, Ltd.**, **Tokyo Collectables Ltd.**, **Technics Stores Inc.**, and **Diecast Classics Inc.**

These businesses demonstrate strong purchasing behavior and should be prioritized for loyalty programs, exclusive offers, or premium services

TOP 5 LOYAL CUSTOMERS

View | Flow Variables

Table View

Open in new window

Rows: 5 | Columns: 6

RowID	MONETARY *Number (double)*	RECENCY *Number (double)*	FREQUENCY *Number (double)*	CUST_SEGMENT *String*	First*(CUSTOMERNAME) *String*	Ranking Segment *Number (integer)*
Ro'	MONETARY	RECENCY	FREQUENCY	1 selected	First*(CUSTOMERNAME)	Ranking Segment
Row24	0.072	0.028	0.08	Loyal Customers	Quebec Home Shopping Network	1
Row29	0.074	0.226	0.12	Loyal Customers	Volvo Model Replicas, Co	2
Row33	0.077	0.117	0.08	Loyal Customers	Lyon Souveniers	3
Row37	0.08	0.105	0.08	Loyal Customers	Collectables For Less Inc.	4
Row39	0.082	0.081	0.08	Loyal Customers	Gifts4AllAges.com	5

These customers fall under the **"Loyal Customers"** category, meaning they have a moderate-to-high RFM score but do not purchase as frequently as "Best Customers.

The highest-ranking **Loyal Customers** include:

Quebec Home Shopping Network,Volvo Model Replicas, Co,Lyon Souveniers,Collectables For Less Inc.,Gifts4AllAges.com.

These businesses are potential long-term customers who should be nurtured further.

TOP 5 CUSTOMERS ON THE VERGE OF CHURNING CUSTOMERS

Table View

Open in new window

Rows: 5 | Columns: 6

RowID	MONETARY *Number (double)*	RECENCY *Number (double)*	FREQUENCY *Number (double)*	CUST_SEGMENT *String*	First*(CUSTOMERNAME) *String*	Ranking Segment *Number (integer)*
Ro	MONETARY	RECENCY	FREQUENCY	1 selected	First*(CUSTOMERNAME)	Ranking Segment
Row0	0	0.281	0.04	Customers on the Verge of Churning	Boards & Toys Co.	1
Row1	0.017	0.206	0.08	Customers on the Verge of Churning	Atelier graphique	2
Row2	0.019	1	0.08	Customers on the Verge of Churning	Auto-Moto Classics Inc.	3
Row3	0.027	0.314	0.04	Customers on the Verge of Churning	Microscale Inc.	4
Row4	0.027	0.53	0.12	Customers on the Verge of Churning	Royale Belge	5

Customers on the Verge of Churning, high risk of becoming inactive.

recency and frequency scores are low

Top Customers at Risk of Churning: Boards & Toys Co.,Atelier Graphique, Auto-Moto Classics Inc., Microscale Inc., Royale Belge

These customers may leave permanently if no action is taken .

TOP 5 LOST CUSTOMERS

Table View

Open in new window

Rows: 5 | Columns: 6

RowID	MONETARY *Number (double)*	RECENCY *Number (double)*	FREQUENCY *Number (double)*	CUST_SEGMENT *String*	First*(CUSTOMERNAME) *String*	Ranking Segment *Number (integer)*
Ro	MONETARY	RECENCY	FREQUENCY	1 selected	First*(CUSTOMERNAME)	Ranking Segment
Row5	0.029	0.579	0	Lost Customers	Bavarian Collectables Imports, Co.	1
Row6	0.03	0.479	0.04	Lost Customers	Double Decker Gift Stores, Ltd	2
Row10	0.045	0.606	0.04	Lost Customers	Signal Collectibles Ltd.	3
Row12	0.05	0.658	0.04	Lost Customers	Iberia Gift Imports, Corp.	4
Row15	0.054	0.471	0.04	Lost Customers	Clover Collections, Co.	5

These customers are categorized as **"Lost Customers,"** meaning they have **stopped purchasing** and are highly inactive.

These customers are **almost completely lost** unless a **strong win-back strategy** is implemented. Immediate action is needed to **reactivate them and prevent further customer churn**.

Key Insights

High Risk of Churn (50.56%)	More than half of the customer base is at risk of leaving.	Immediate re-engagement strategies are essential to prevent revenue loss.
Loyal Customers (26.97%) – Retention Opportunity	These customers buy frequently but may not spend as much as Best Customers.	They can be converted into Best Customers with targeted promotions and personalized experiences.
Best Customers (16.85%) – High Value, Needs Special Attention	These customers generate the most revenue per transaction.	Ensuring their continued satisfaction through exclusive perks and priority support is vital.
Lost Customers (5.62%) – Reactivation Potential	While small in percentage, re-engaging them can lead to increased retention.	Effective win-back strategies can recover some of these customers.

RECOMMENDATIONS

1. Address High Churn Risk Customers (50.56%)

Implement **personalized re-engagement campaigns** (emails, SMS, social media ads).

Offer **limited-time discounts and loyalty incentives** to encourage repeat purchases.

Conduct **customer feedback surveys** to understand why they are disengaging.

2. Strengthen Loyalty Programs for Loyal Customers (26.97%)

Provide **tiered loyalty rewards** (discounts, early access to sales, birthday offers).

Encourage referrals by offering **bonus points or discounts for bringing in new customers**.

3. Retain Best Customers (16.85%)

Exclusive perks like VIP customer service, premium product access, and personalized recommendations.

Subscription models or memberships to maintain engagement.

Regular engagement through high-value content, appreciation gifts, and personalized outreach.

4. Win Back Lost Customers (5.62%)

Targeted comeback campaigns (e.g., "We Miss You" emails with special offers).

Social media retargeting ads showcasing new arrivals or exclusive deals.

Analyze reasons for churn and address key pain points (pricing, service quality, product variety).

CONCLUSION

The high percentage of customers at risk of churning (50.56%) is a major concern, requiring **immediate retention efforts** to prevent revenue loss.

Loyal Customers (26.97%) have strong potential to become Best Customers, and a **strategic loyalty program** can drive revenue growth.

Best Customers (16.85%) are the most valuable, and **premium engagement strategies** should focus on their retention.

Lost Customers (5.62%) present an opportunity for reactivation, and effective comeback campaigns can recover some of them.

By implementing these strategies, the business can significantly improve retention, boost revenue, and create long-term customer relationships.

PART – B

MARKET BASKET ANALYSIS

CONTENTS

BUSINESS CONTEXT

The grocery retail industry is highly competitive, requiring a deep understanding of customer buying behavior..

Identifying frequently purchased item combinations can optimize sales, improve customer satisfaction, and enhance profitability.

Leveraging POS transaction data allows for targeted marketing strategies, better inventory management, and increased revenue.

Objective of Analysis:

- Analyze POS transactional data to uncover customer buying patterns.
- Identify frequently purchased item combinations.
- Optimize combo offers and discount strategies to increase basket size and sales.
- Enhance customer loyalty through personalized promotions.

PROBLEM STATEMENT

Identify Frequently purchased item combinations from POS data

Understand sales trends (daily, weekly, monthly, quarterly, yearly)

Use Association rule mining insights to recommend combo offers and discounts

Improve **customer retention and sales growth** by targeting the right product groups.

Data Overview

Key columns in the dataset:

- **Date:** Transaction date.
- **Order_ID:** Unique identifier for each customer order.
- **Product:** Name of the purchased item.
- The dataset contains transactional data spanning multiple months/years, allowing for trend analysis and market basket analysis.

Data Cleaning Steps Taken:

- Checked **duplicate records**
- Checked **missing values**
- Converted **dates to proper format**

Exploratory Data Analysis

STATISTICS SUMMARY

	Date	Order_id	Product
count	15911	15911.000000	15911
unique	603	NaN	37
top	08-02-2019	NaN	poultry
freq	138	NaN	480
mean	NaN	574.150462	NaN
std	NaN	328.537425	NaN
min	NaN	1.000000	NaN
25%	NaN	289.500000	NaN
50%	NaN	579.000000	NaN
75%	NaN	859.000000	NaN
max	NaN	1139.000000	NaN

Date 603
Order_id 1139
Product 37
dtype: int64

Total Transactions: 15,911

Unique Order IDs: 1,139

Unique Products: 37

Most Purchased Product: Poultry (480 times)

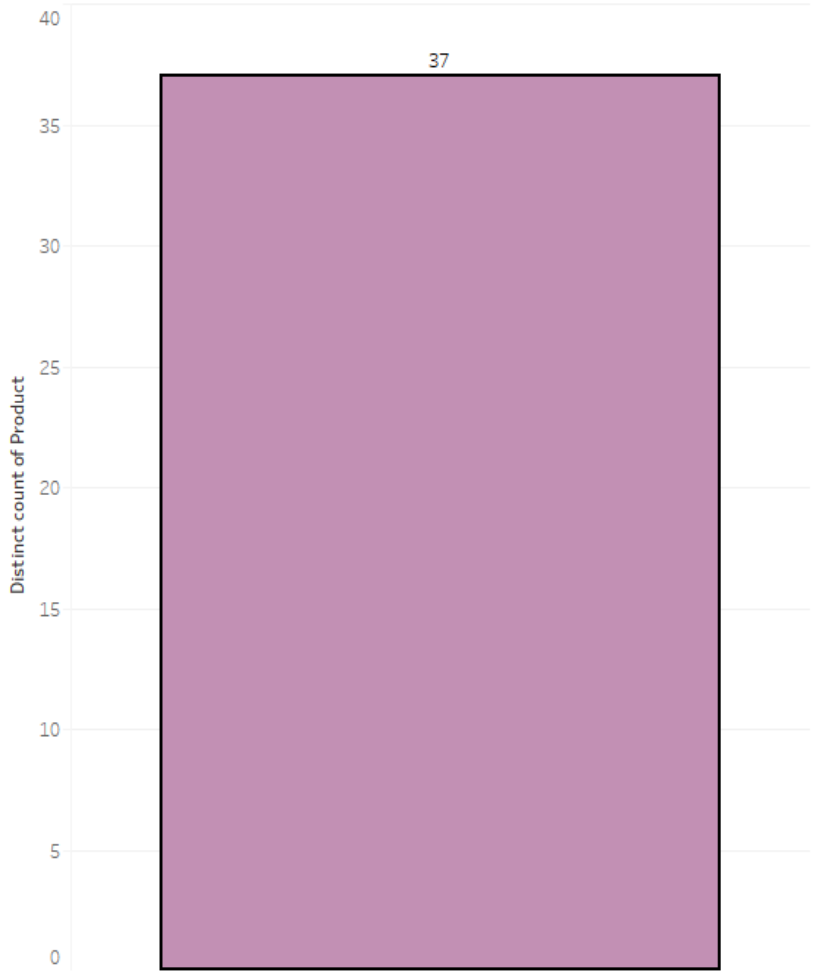
Top Transaction Date:08-02-2019 (138 orders)

Order Distribution: Median Order ID: 579

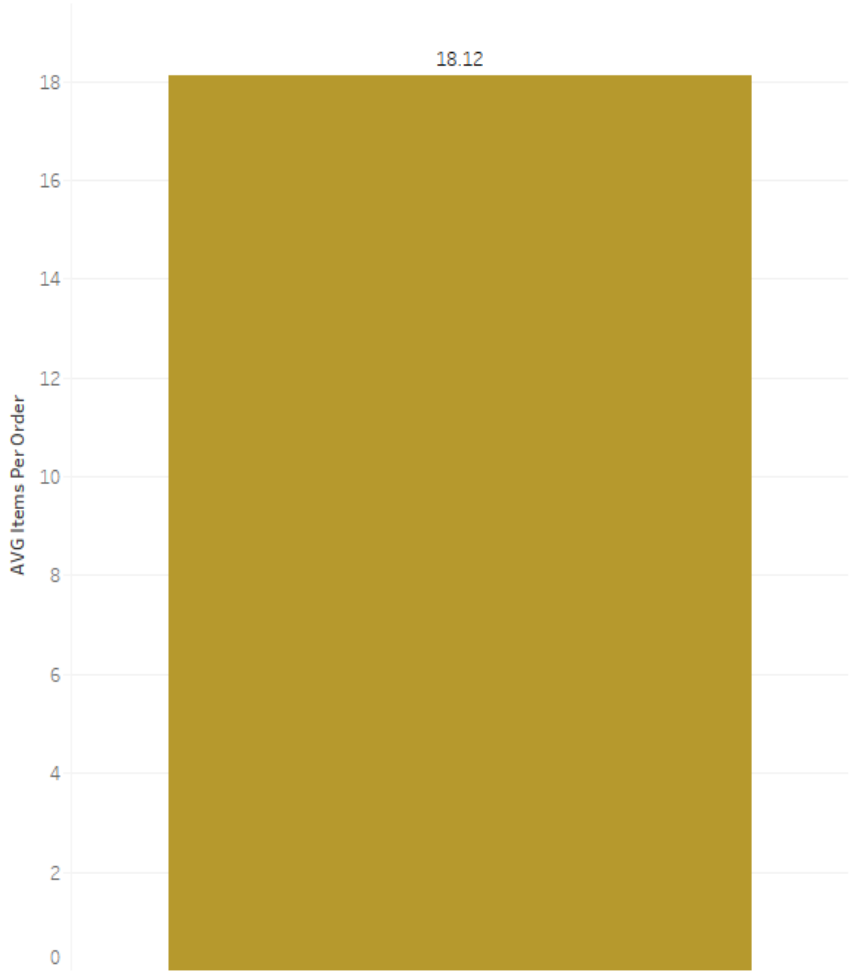
Spread: 25% of orders below **289.5**, 75% below **859**

SUMMARY STATISTICS

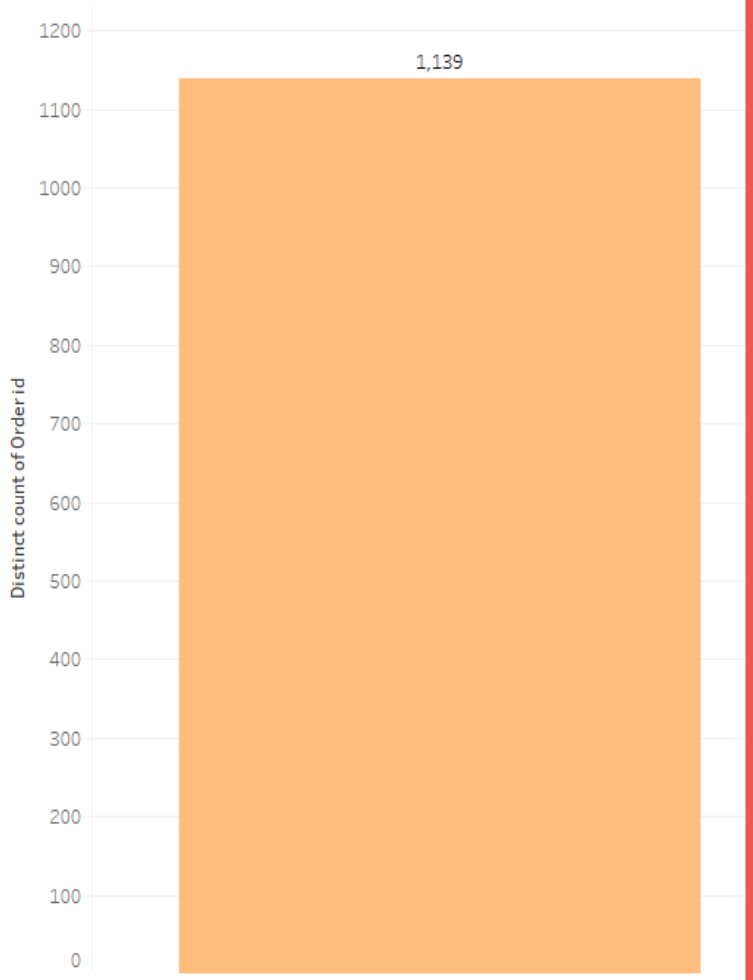
Total unique products sold



Average Items per Order

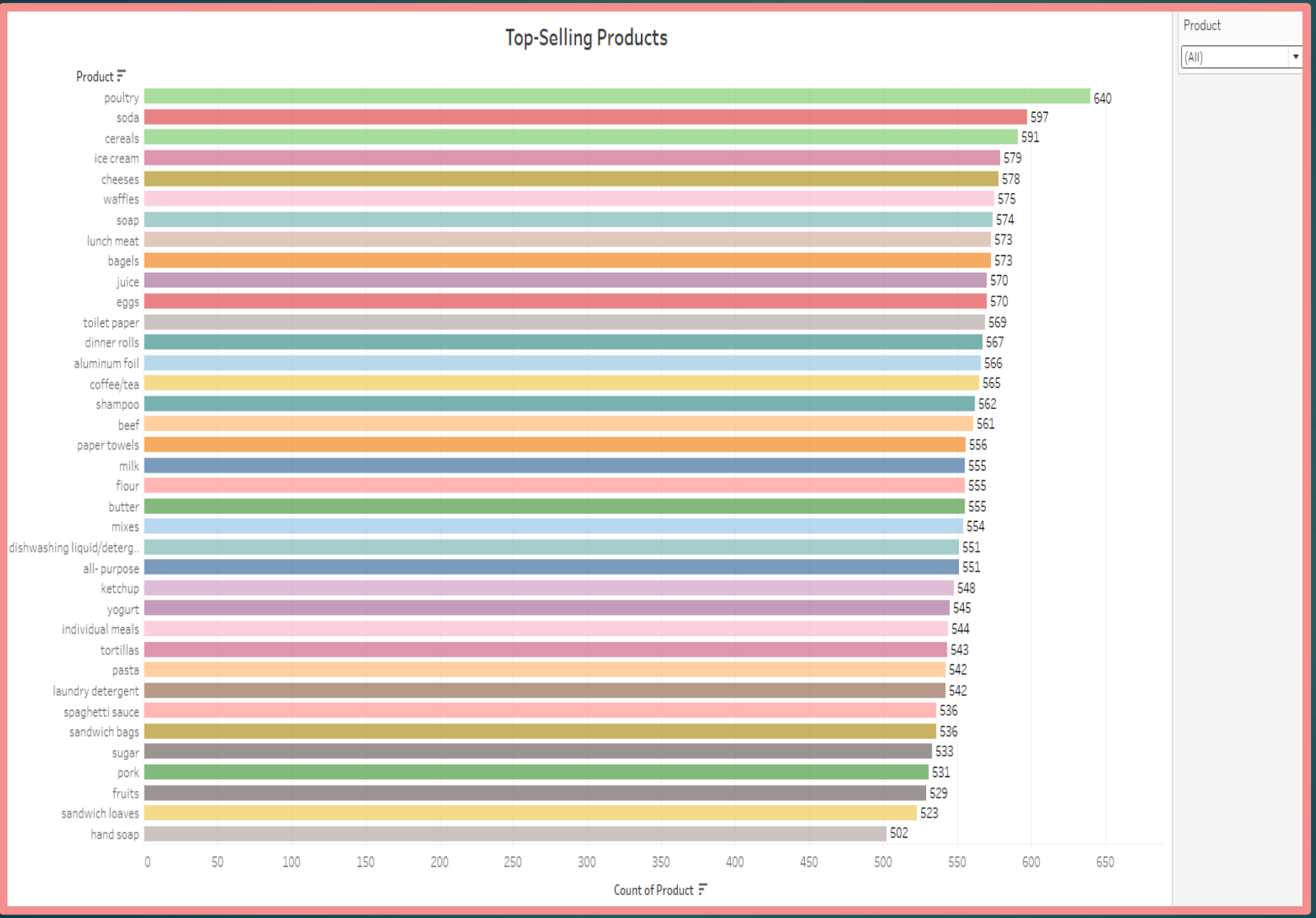


Total number of unique transactions



UNIVARIATE ANALYSIS

TOP SELLING PRODUCTS



Top-Selling Product: Poultry leads with **640** sales, significantly ahead of other products.

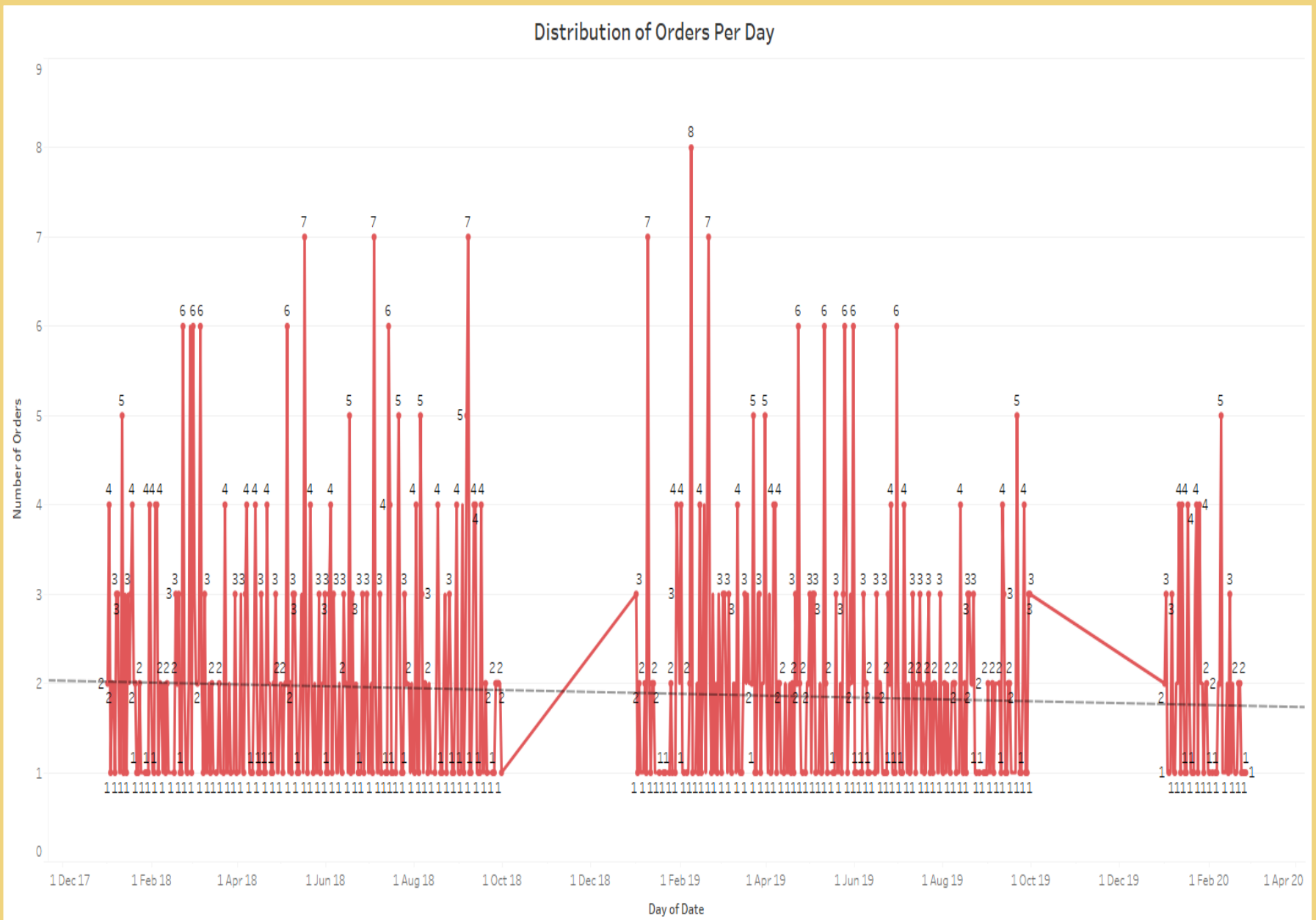
Popular Categories: Soda (**597**) and Cereals (**591**) follow closely, indicating high demand for beverages and breakfast items.

Diverse Consumption: Ice cream, cheeses, and waffles are among the top-selling items, suggesting a mix of essential and indulgence-driven purchases.

Household Essentials: Items like soap, toilet paper, and aluminum foil rank high, reflecting frequent household needs.

Opportunities: Focus on **bundling** high-demand products (e.g., cereals + milk, poultry + spices) to boost sales further.

DISTRIBUTION OF ORDERS PER DAY



Order Frequency: Orders per day vary significantly, ranging from **1 to 8 orders**.

Peak Days: There are multiple spikes where orders reach **6-8 per day**, indicating high-demand periods.

Gaps in Orders: Some periods show **no or minimal orders**, suggesting seasonal trends or operational gaps.

Trend Analysis: The general order volume appears **stable but inconsistent**, with frequent fluctuations.

Bivariate Analysis & Multivariate Analysis

SALES TRENDS OVER TIME

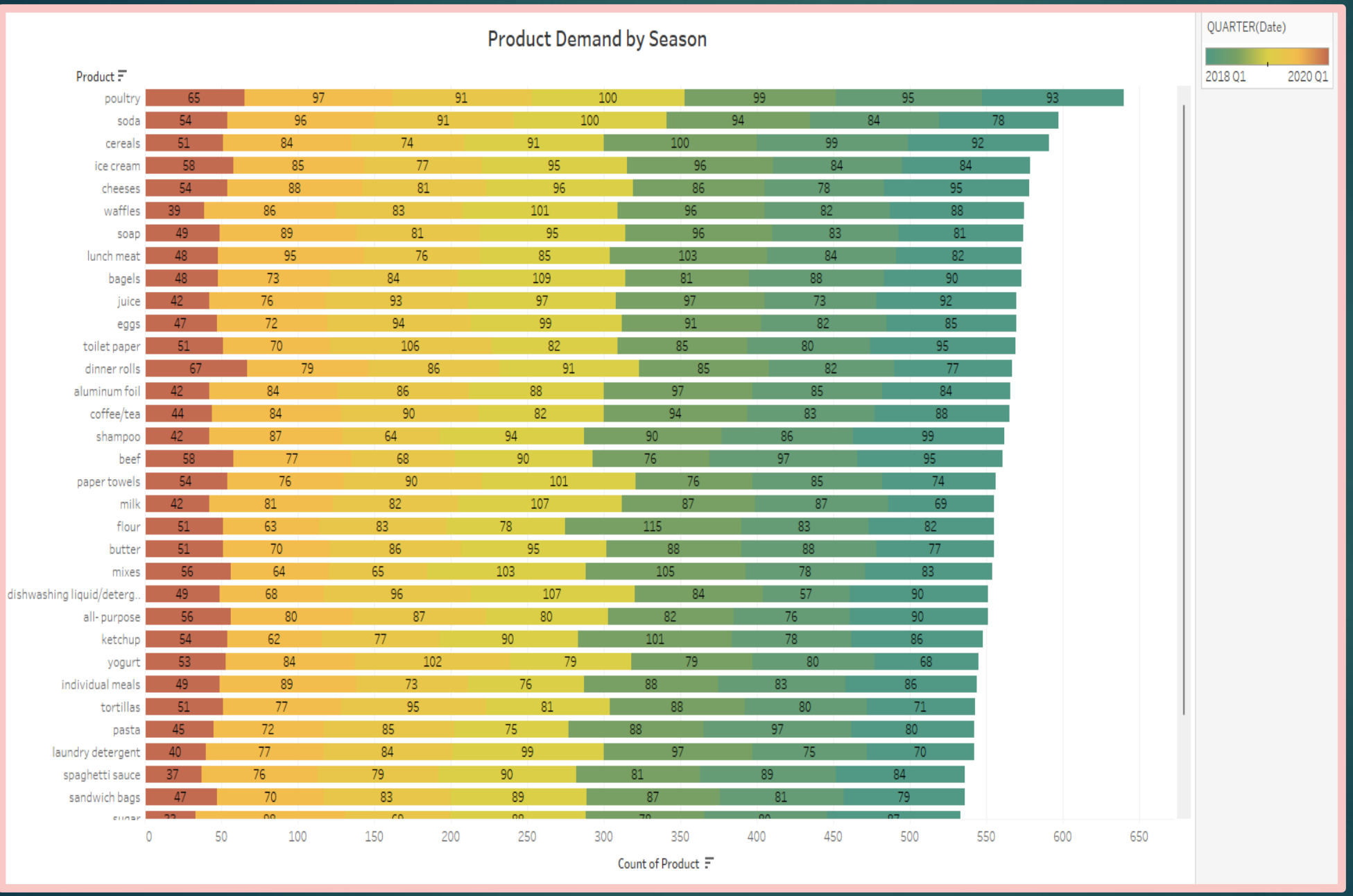
Declining Sales Trend: Orders have steadily **decreased from 533 (2018) to 507 (2019) and sharply to 99 (2020 Q1)**.

Quarterly Performance: Orders peaked in **2018 Q3 & Q4 (~180 orders)** but have been declining since **2019 Q3**.

Monthly Fluctuations: Orders per month vary between **49 and 67**, showing inconsistent demand patterns.

Weekly Orders: High volatility with peaks reaching **22 orders** and dips as low as **4 orders per week**.

PRODUCT DEMAND BY SEASON



Seasonal Trends:

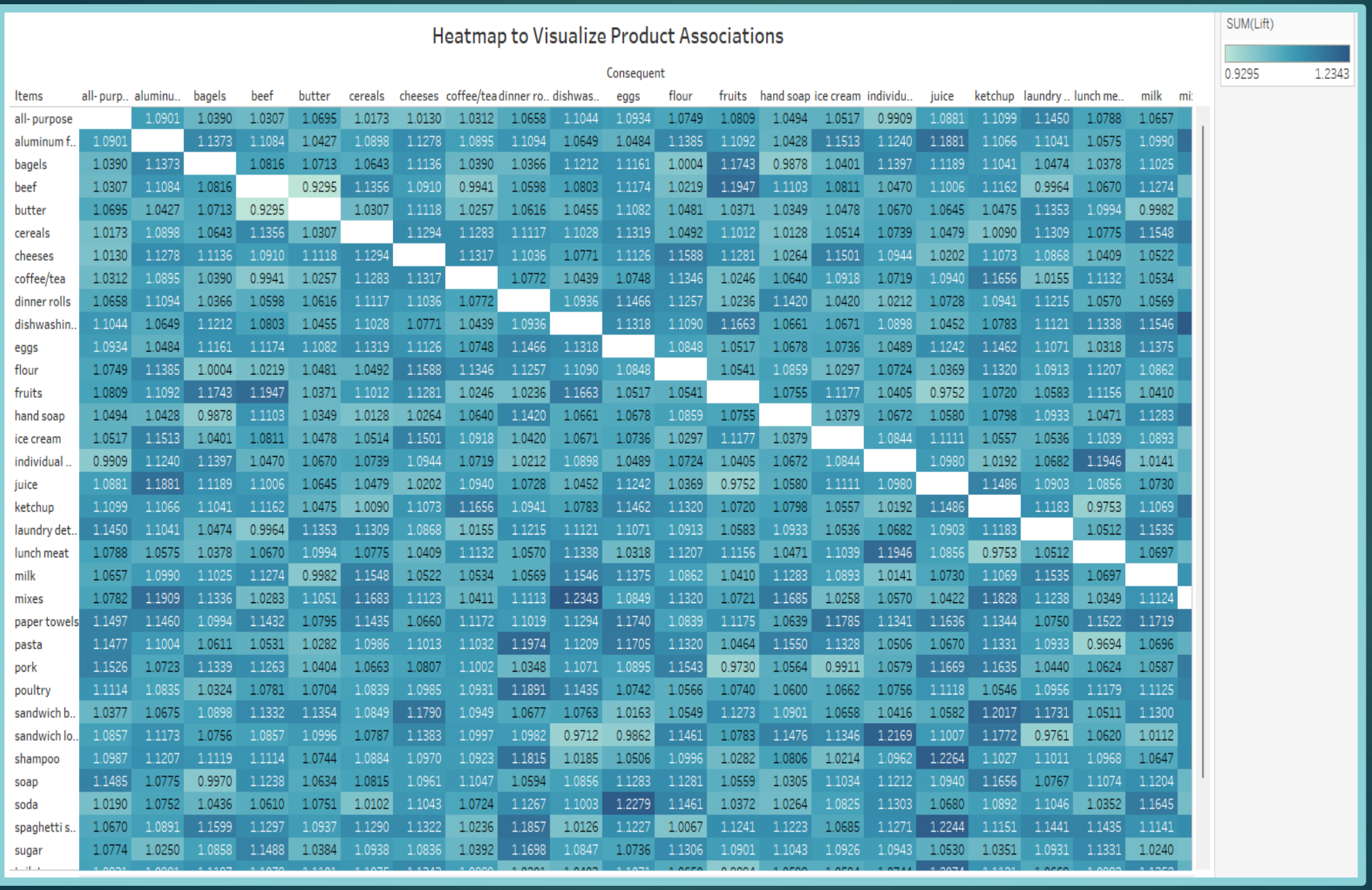
Poultry, cereals, and flour see high demand across multiple quarters.

Ice cream, soda, and juice peak in **summer (Q3)**, indicating seasonal preferences.

Stable Demand Products:

Milk, eggs, bread, and cleaning supplies show consistent demand year-round

HEATMAP TO VISUALIZE FREQUENTLY PURCHASED ITEM COMBINATIONS



Items	all- purp..	aluminu..	bagels	beef	butter	cereals	cheeses	coffee/tea	dinner ro..	dishwas..	eggs	flour	fruits	hand soap	ice cream	individu..	juice	ketchup	laundry ..	lunch me..	milk	mi:
all- purpose		1.0901	1.0390	1.0307	1.0695	1.0173	1.0130	1.0312	1.0658	1.1044	1.0934	1.0749	1.0809	1.0494	1.0517	0.9909	1.0881	1.1099	1.1450	1.0788	1.0657	
aluminum f..	1.0901		1.1373	1.1084	1.0427	1.0898	1.1278	1.0895	1.1094	1.0649	1.0484	1.1385	1.1092	1.0428	1.1513	1.1240	1.1881	1.1066	1.1041	1.0575	1.0990	
bagels	1.0390	1.1373		1.0816	1.0713	1.0643	1.1136	1.0390	1.0366	1.1212	1.1161	1.0004	1.1743	0.9878	1.0401	1.1397	1.1189	1.1041	1.0474	1.0378	1.1025	
beef	1.0307	1.1084	1.0816		0.9295	1.1356	1.0910	0.9941	1.0598	1.0803	1.1174	1.0219	1.1947	1.1103	1.0811	1.0470	1.1006	1.1162	0.9964	1.0670	1.1274	
butter	1.0695	1.0427	1.0713	0.9295		1.0307	1.1118	1.0257	1.0616	1.0455	1.1082	1.0481	1.0371	1.0349	1.0478	1.0670	1.0645	1.0475	1.1353	1.0994	0.9982	
cereals	1.0173	1.0898	1.0643	1.1356	1.0307		1.1294	1.1283	1.1117	1.1028	1.1319	1.0492	1.1012	1.0128	1.0514	1.0739	1.0479	1.0090	1.1309	1.0775	1.1548	
cheeses	1.0130	1.1278	1.1136	1.0910	1.1118	1.1294		1.1317	1.1036	1.0771	1.1126	1.1588	1.1281	1.0264	1.1501	1.0944	1.0202	1.1073	1.0868	1.0409	1.0522	
coffee/tea	1.0312	1.0895	1.0390	0.9941	1.0257	1.1283	1.1317		1.0772	1.0439	1.0748	1.1346	1.0246	1.0640	1.0918	1.0719	1.0940	1.1656	1.0155	1.1132	1.0534	
dinner rolls	1.0658	1.1094	1.0366	1.0598	1.0616	1.1117	1.1036	1.0772		1.0936	1.1466	1.1257	1.0236	1.1420	1.0420	1.0212	1.0728	1.0941	1.1215	1.0570	1.0569	
dishwashin..	1.1044	1.0649	1.1212	1.0803	1.0455	1.1028	1.0771	1.0439	1.0936		1.1318	1.1090	1.1663	1.0661	1.0671	1.0898	1.0452	1.0783	1.1121	1.1338	1.1546	
eggs	1.0934	1.0484	1.1161	1.1174	1.1082	1.1319	1.1126	1.0748	1.1466	1.1318		1.0848	1.0517	1.0678	1.0736	1.0489	1.1242	1.1462	1.1071	1.0318	1.1375	
flour	1.0749	1.1385	1.0004	1.0219	1.0481	1.0492	1.1588	1.1346	1.1257	1.1090	1.0848		1.0541	1.0859	1.0297	1.0724	1.0369	1.1320	1.0913	1.1207	1.0862	
fruits	1.0809	1.1092	1.1743	1.1947	1.0371	1.1012	1.1281	1.0246	1.0236	1.1663	1.0517	1.0541		1.0755	1.1177	1.0405	0.9752	1.0720	1.0583	1.1156	1.0410	
hand soap	1.0494	1.0428	0.9878	1.1103	1.0349	1.0128	1.0264	1.0640	1.1420	1.0661	1.0678	1.0859	1.0755		1.0379	1.0672	1.0580	1.0798	1.0933	1.0471	1.1283	
ice cream	1.0517	1.1513	1.0401	1.0811	1.0478	1.0514	1.1501	1.0918	1.0420	1.0671	1.0736	1.0297	1.1177	1.0379		1.0844	1.1111	1.0557	1.0536	1.1039	1.0893	
individual ..	0.9909	1.1240	1.1397	1.0470	1.0670	1.0739	1.0944	1.0719	1.0212	1.0898	1.0489	1.0724	1.0405	1.0672	1.0844		1.0980	1.0192	1.0682	1.1946	1.0141	
juice	1.0881	1.1881	1.1189	1.1006	1.0645	1.0479	1.0202	1.0940	1.0728	1.0452	1.1242	1.0369	0.9752	1.0580	1.1111	1.0980		1.1486	1.0903	1.0856	1.0730	
ketchup	1.1099	1.1066	1.1041	1.1162	1.0475	1.0090	1.1073	1.1656	1.0941	1.0783	1.1462	1.1320	1.0720	1.0798	1.0557	1.0192	1.1486		1.1183	0.9753	1.1069	
laundry det..	1.1450	1.1041	1.0474	0.9964	1.1353	1.1309	1.0868	1.0155	1.1215	1.1121	1.1071	1.0913	1.0583	1.0933	1.0536	1.0682	1.0903	1.1183		1.0512	1.1535	
lunch meat	1.0788	1.0575	1.0378	1.0670	1.0994	1.0775	1.0409	1.1132	1.0570	1.1338	1.0318	1.1207	1.1156	1.0471	1.1039	1.1946	1.0856	0.9753	1.0512		1.0697	
milk	1.0657	1.0990	1.1025	1.1274	0.9982	1.1548	1.0522	1.0534	1.0569	1.1546	1.1375	1.0862	1.0410	1.1283	1.0893	1.0141	1.0730	1.1069	1.1535	1.0697		
mixes	1.0782	1.1909	1.1336	1.0283	1.1051	1.1683	1.1123	1.0411	1.1113	1.2343	1.0849	1.1320	1.0721	1.1685	1.0258	1.0570	1.0422	1.1828	1.1238	1.0349	1.1124	
paper towels	1.1497	1.1460	1.0994	1.1432	1.0795	1.1435	1.0660	1.1172	1.1019	1.1294	1.1740	1.0839	1.1175	1.0639	1.1785	1.1341	1.1636	1.1344	1.0750	1.1522	1.1719	
pasta	1.1477	1.1004	1.0611	1.0531	1.0282	1.0986	1.1013	1.1032	1.1974	1.1209	1.1705	1.1320	1.0464	1.1550	1.1328	1.0506	1.0670	1.1331	1.0933	0.9694	1.0696	
pork	1.1526	1.0723	1.1339	1.1263	1.0404	1.0663	1.0807	1.1002	1.0348	1.1071	1.0895	1.1543	0.9730	1.0564	0.9911	1.0579	1.1669	1.1635	1.0440	1.0624	1.0587	
poultry	1.1114	1.0835	1.0324	1.0781	1.0704	1.0839	1.0985	1.0931	1.1891	1.1435	1.0742	1.0566	1.0740	1.0600	1.0662	1.0756	1.1118	1.0546	1.0956	1.1179	1.1125	
sandwich b..	1.0377	1.0675	1.0898	1.1332	1.1354	1.0849	1.1790	1.0949	1.0677	1.0763	1.0163	1.0549	1.1273	1.0901	1.0658	1.0416	1.0582	1.2017	1.1731	1.0511	1.1300	
sandwich lo..	1.0857	1.1173	1.0756	1.0857	1.0996	1.0787	1.1383	1.0997	1.0982	0.9712	0.9862	1.1461	1.0783	1.1476	1.1346	1.2169	1.1007	1.1772	0.9761	1.0620	1.0112	
shampoo	1.0987	1.1207	1.1119	1.1114	1.0744	1.0884	1.0970	1.0923	1.1815	1.0185	1.0506	1.0996	1.0282	1.0806	1.0214	1.0962	1.2264	1.1027	1.1011	1.0968	1.0647	
soap	1.1485	1.0775	0.9970	1.1238	1.0634	1.0815	1.0961	1.1047	1.0594	1.0856	1.1283	1.1281	1.0559	1.0305	1.1034	1.1212	1.0940	1.1656	1.0767	1.1074	1.1204	
soda	1.0190	1.0752	1.0436	1.0610	1.0751	1.0102	1.1043	1.0724	1.1267	1.1003	1.2279	1.1461	1.0372	1.0264	1.0825	1.1303	1.0680	1.0892	1.1046	1.0352	1.1645	
spaghetti s..	1.0670	1.0891	1.1599	1.1297	1.0937	1.1290	1.1322	1.0236	1.1857	1.0126	1.1227	1.0067	1.1241	1.1223	1.0685	1.1271	1.2244	1.1151	1.1441	1.1435	1.1141	
sugar	1.0774	1.0250	1.0858	1.1488	1.0384	1.0938	1.0836	1.0392	1.1698	1.0847	1.0736	1.1306	1.0901	1.1043	1.0926	1.0943	1.0530	1.0351	1.0931	1.1331	1.0240	

Strong Product Associations:

Paper towels & mixes (Lift = 1.234) → High likelihood of being purchased together.

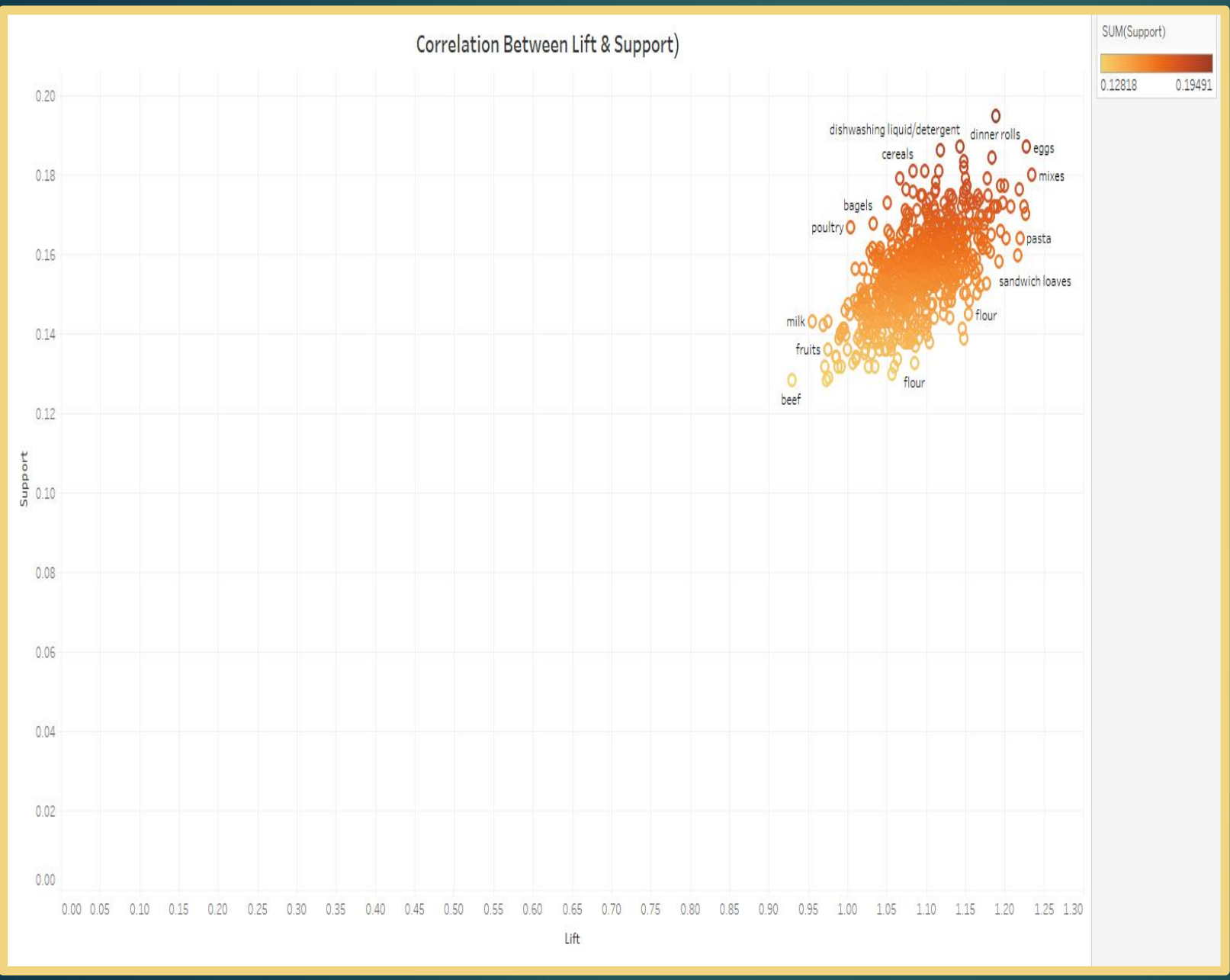
Pasta & sauce show strong correlation, reinforcing natural pairing.

Cross-Selling Opportunities:

Sandwich bags & sandwich loaves → Ideal for bundling offers.

Laundry detergent & soap → Suggests promoting hygiene-related bundles

Lift vs. Support Analysis



High Support & Lift:

Dinner rolls, eggs, mixes, pasta, and sandwich loaves are frequently purchased together and have strong associations.

Dishwashing liquid/detergent and cereals also show high correlation, indicating potential for bundled promotions.

Strategic Product Pairing:

High-support items (e.g., flour, milk, and fruits) are commonly purchased but may not always have a high lift → Opportunities to create new combos.

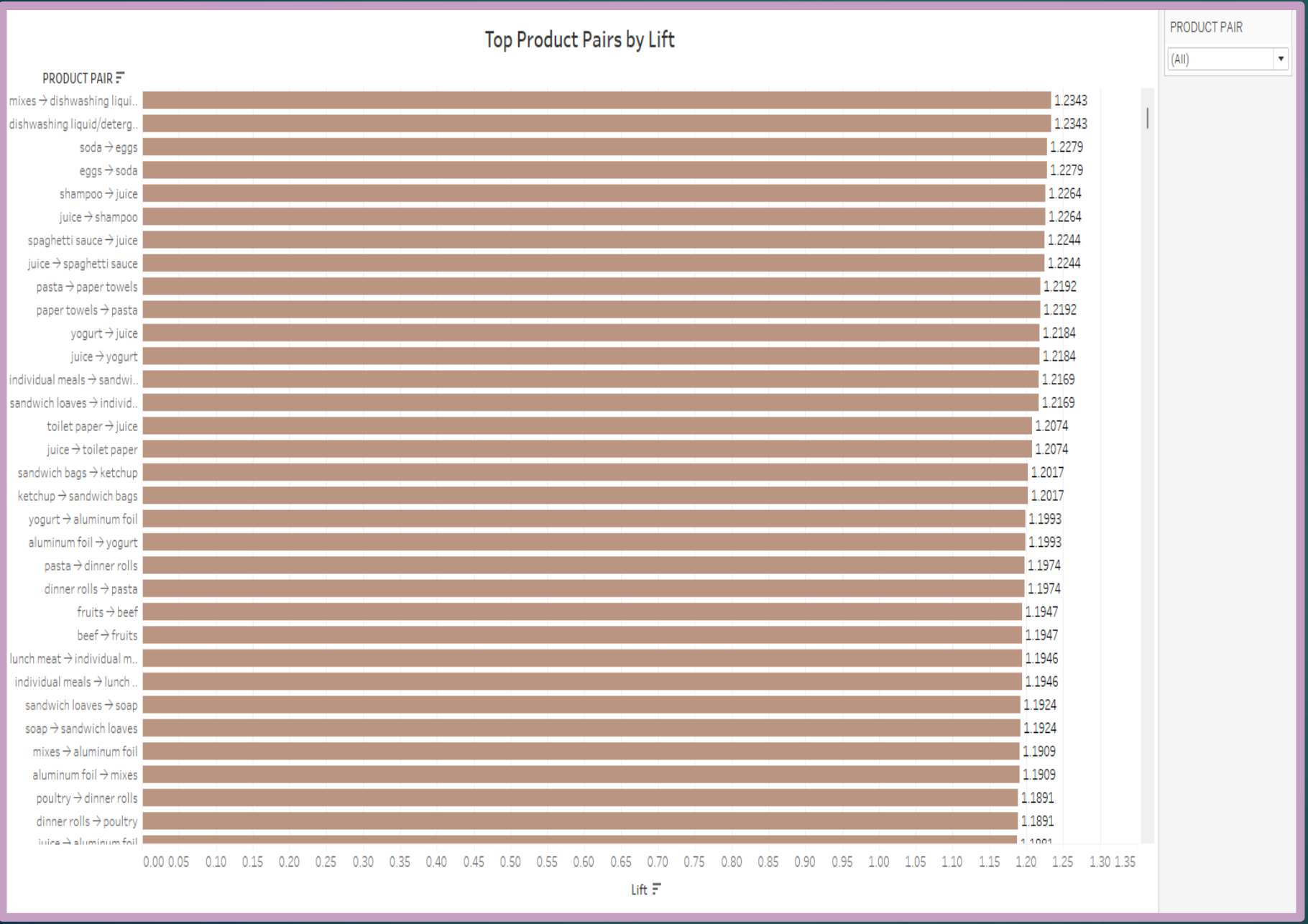
High-lift but lower-support items (e.g., beef and poultry) could benefit from targeted promotions to increase their purchase frequency.

Marketing Implications:

Recommend combo discounts for high-support & high-lift pairs (e.g., eggs & dinner rolls, pasta & sauces).

Encourage bundling of low-support, high-lift items to drive their adoption in customer baskets

TOP PRODUCT PAIRS BY LIFT



Strongest Product Associations:

Mixes & Dishwashing Liquid/Detergent (Lift: 1.2343) have the highest lift, suggesting these items are frequently bought together.

Soda & Eggs (Lift: 1.2279) and **Juice & Shampoo (Lift: 1.2264)** also show strong purchase relationships.

Cross-Category Pairings:

Some top associations involve unrelated categories, such as **shampoo & juice, soda & eggs, and aluminum foil & yogurt**. These might be impulse or habitual purchases.

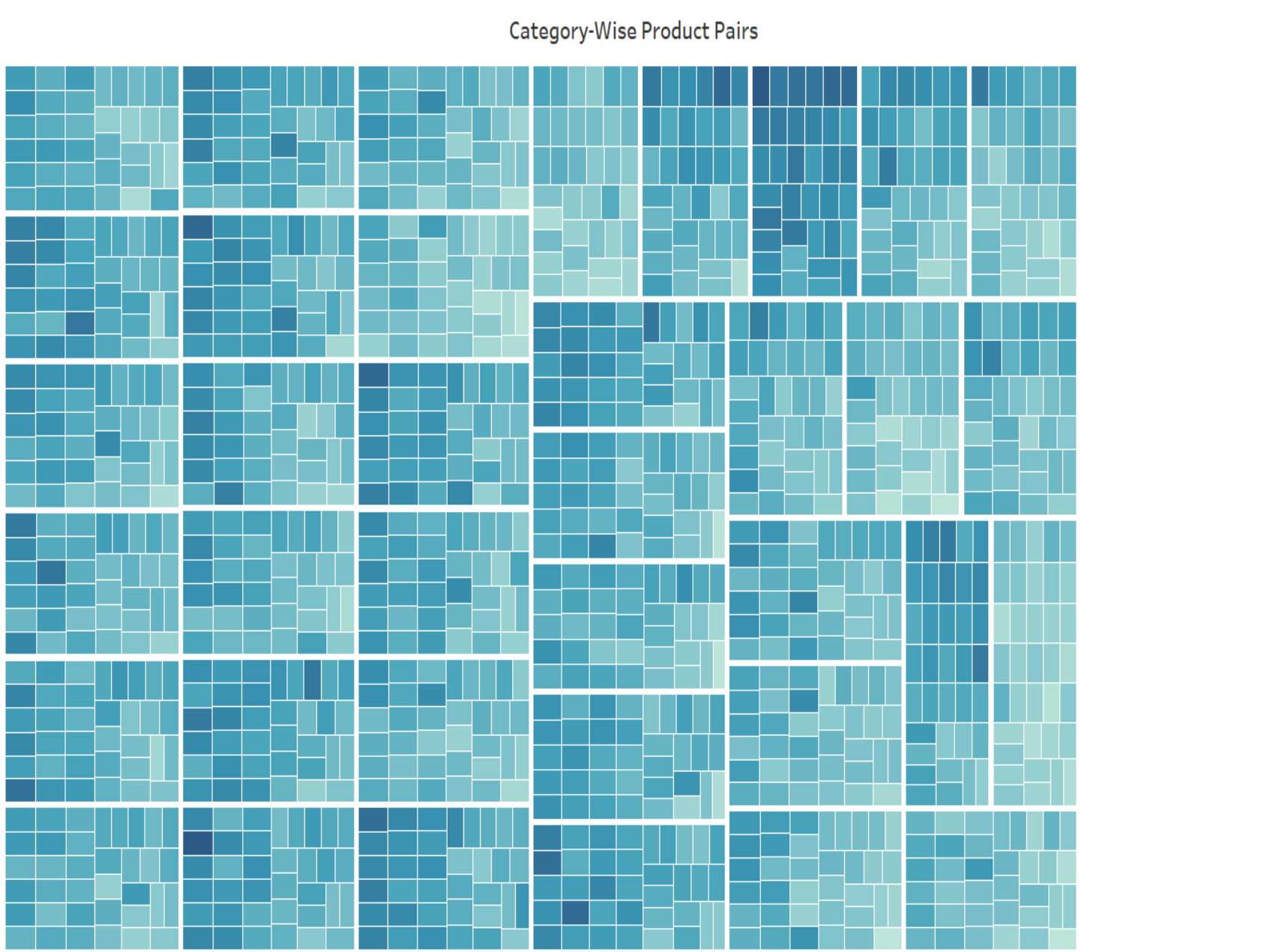
Dinner Rolls & Pasta (Lift: 1.1974) and **Pasta & Paper Towels (Lift: 1.2192)** suggest meal-related bundling opportunities.

Marketing & Promotion Strategies:

Bundle discounts for top pairs (e.g., detergent & mixes, pasta & dinner rolls) to drive more sales.

Cross-category promotions for frequently co-purchased items (e.g., offer a discount on juice with shampoo)

CATEGORY-WISE PRODUCT PAIRS



High-Support Product Categories:

The darker shades indicate **higher support values**, meaning these product categories are frequently bought together.

Some categories dominate in size, suggesting they contain multiple high-support product pairs.

Diverse Purchase Patterns:

The chart is fragmented into multiple sections, indicating a **wide variety of product pairs** across different categories.

Some sections appear denser, implying **higher transaction volume** within specific categories

MARKET BASKET ANALYSIS

Introduction to Market Basket Analysis

Market Basket Analysis helps identify frequently bought item sets.

Using association rules, we can optimize marketing strategies, bundle products, and increase revenue. This report presents findings using KNIME.

UNDERSTANDING SUPPORT, CONFIDENCE, & LIFT

Support:

Proportion of transactions containing both items.

Confidence:

Probability that customers who buy item A also buy item B.

Lift:

Measure of strength of association (>1 means positive correlation).

DEFINITION AND FORMULA FOR KEY METRICS

1. Support

Definition:
Support measures how frequently an item or itemset appears in the dataset. It indicates the popularity of a product combination in transactions.

Formula:
Support(A⇒B) = Transactions containing (A∪B)/Total Transactions

Threshold Value:

- **Common threshold:** 1% to 5% (varies based on dataset size).
- **Higher support:** The itemset appears frequently, making it a strong rule.
- **Lower support:** The itemset is rare and may not be useful for promotions.

2. Confidence

Definition:
Confidence indicates the likelihood that item **B** is purchased when item **A** is purchased. It measures the **strength** of an association rule.

Formula:
Confidence(A⇒B) = Transactions containing (A∪B)/Transactions containing A

Threshold Value:

- **Common threshold:** 50% to 80%.
- **Higher confidence:** Strong dependency between items (e.g., Milk → Bread with 80% confidence means 80% of people who buy milk also buy bread).
- **Lower confidence:** Weak dependency between items.

3. Lift

Definition:

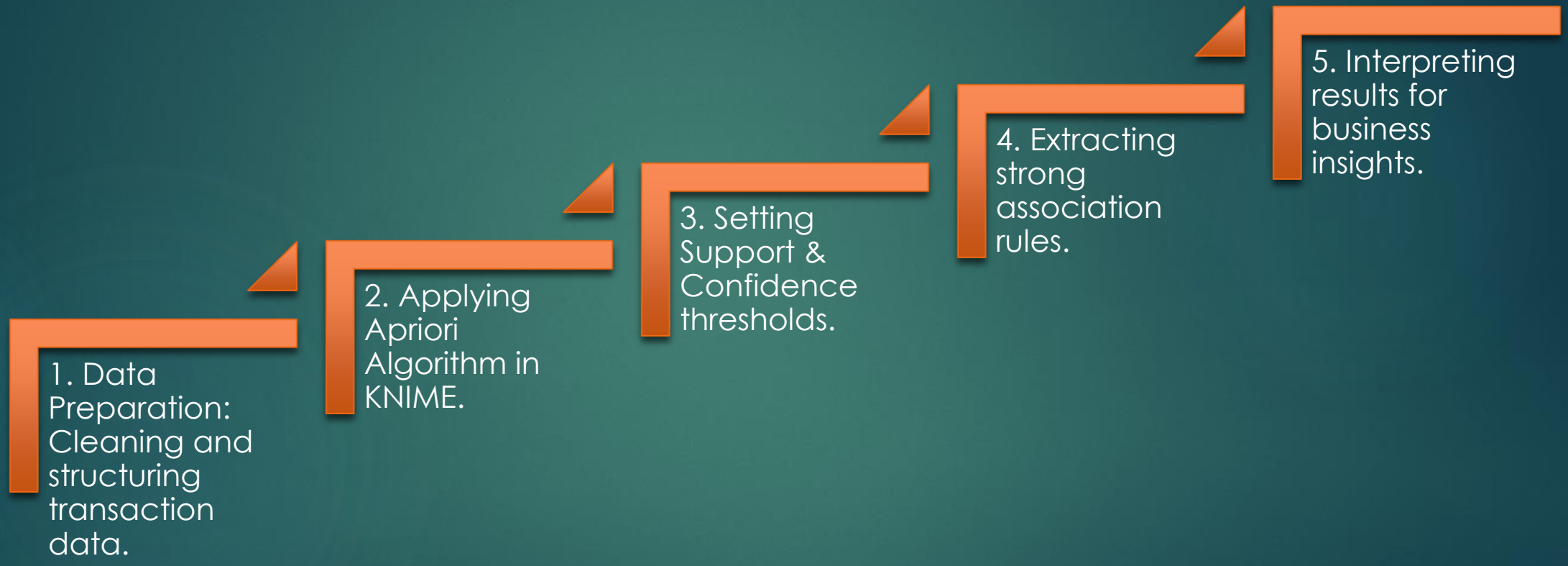
Lift measures how much more likely item **B** is purchased when item **A** is purchased, compared to its normal purchase rate. It helps **identify strong and meaningful associations** beyond random chance.

Formula:
Lift(A⇒B) = Confidence(A⇒B)/Support(B)

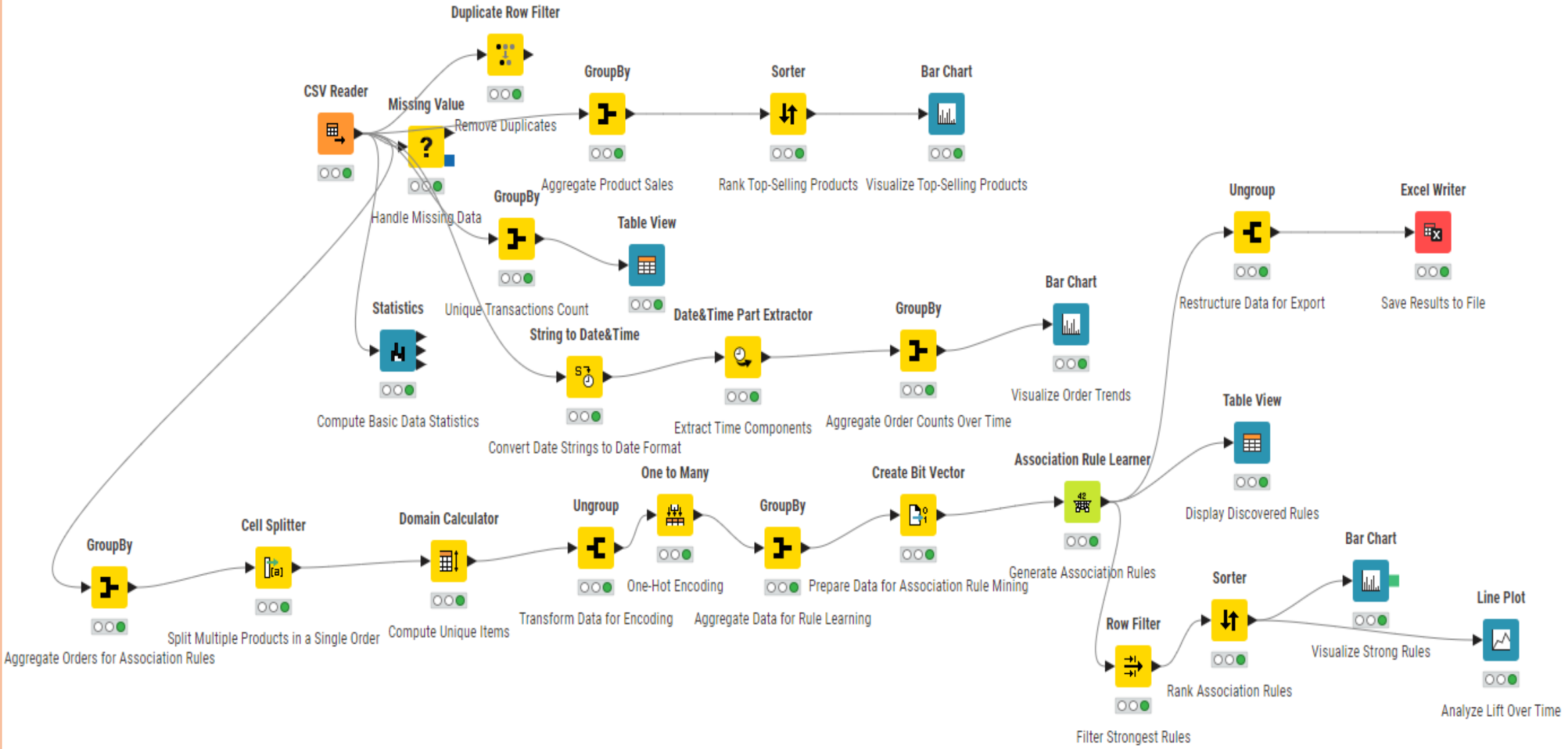
Threshold Value:

- **Lift > 1:** A and B are positively correlated (stronger association).
- **Lift = 1:** No association between A and B.
- **Lift < 1:** A and B are negatively correlated (buying A reduces the chance of buying B).

Process of Association Rule Mining



KNIME WORKFLOW



PROCESS IN KNIME WORKFLOW

Preprocessed data by handling missing values, removing duplicates, and computing basic statistics.

Transformed and aggregated sales and transaction data, converted date formats, and extracted time components.

Prepared data for association rule mining by splitting multiple products, computing unique items, and applying one-hot encoding.

Generated association rules by creating bit vectors and extracting key patterns.

Ranked and filtered rules based on **Support, Confidence, and Lift** to identify the strongest relationships.

Visualized key insights using bar charts and line plots to analyze trends.

Exported final results to an Excel file for business decision-making

THRESHOLD VALUES FOR SUPPORT,LIFT & CONFIDENCE

ASSOCIATION RULES TABLE FINAL OUTPUT

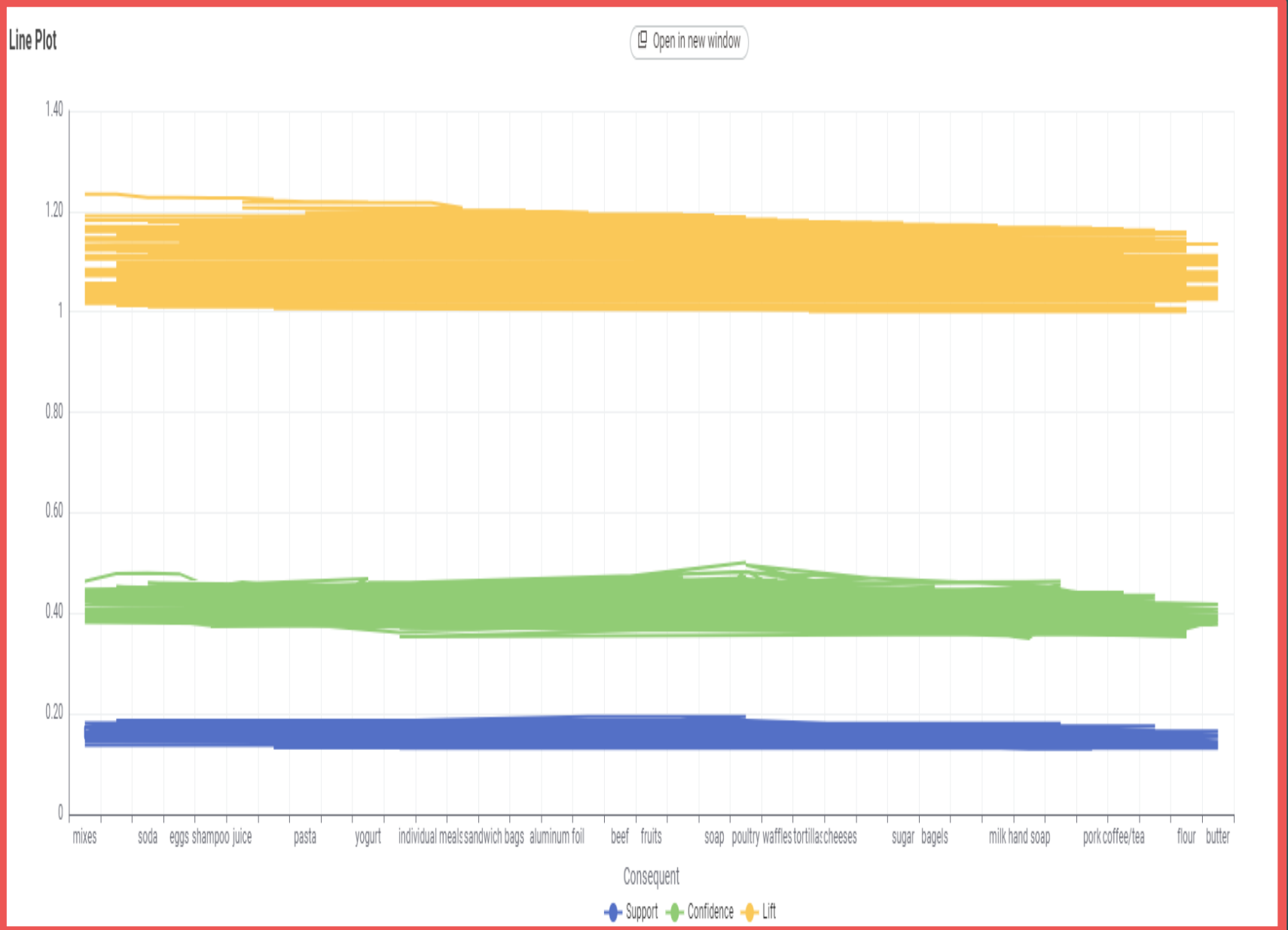
Table View

Open in new window

Rows: 1332 | Columns: 6

RowID	Support *Number (double)*	Confidence *Number (double)*	Lift *Number (double)*	Consequent *String*	implies *String*	Items *Set*
rule0	0.152	0.427	1.111	yogurt	<---	[pork]
rule1	0.152	0.395	1.111	pork	<---	[yogurt]
rule2	0.155	0.422	1.099	yogurt	<---	[sandwich bags]
rule3	0.155	0.404	1.099	sandwich bags	<---	[yogurt]
rule4	0.162	0.409	1.063	yogurt	<---	[lunch meat]
rule5	0.162	0.42	1.063	lunch meat	<---	[yogurt]
rule6	0.168	0.447	1.163	yogurt	<---	[all- purpose]
rule7	0.168	0.436	1.163	all- purpose	<---	[yogurt]
rule8	0.157	0.445	1.158	yogurt	<---	[flour]
rule9	0.157	0.409	1.158	flour	<---	[yogurt]
rule10	0.168	0.429	1.116	yogurt	<---	[soda]
rule11	0.168	0.436	1.116	soda	<---	[yogurt]
rule12	0.14	0.382	0.993	yogurt	<---	[butter]
rule13	0.14	0.365	0.993	butter	<---	[yogurt]
rule14	0.162	0.431	1.121	yogurt	<---	[beef]

VISUALIZES KEY ASSOCIATION RULE METRICS FOR DIFFERENT CONSEQUENT PRODUCTS



Lift Analysis (Orange Line)

Lift values are consistently high (above 1), indicating that the association rules are meaningful.

A lift value above 1 suggests that the presence of the antecedent increases the likelihood of purchasing the consequent.

Higher lift values indicate stronger product relationships.

Confidence Trend (Green Line)

Confidence values remain between **0.3 and 0.5**, implying that when the antecedent items are bought, the consequent products follow with moderate probability.

Certain products like **milk, coffee/tea, and flour** show relatively higher confidence, meaning they frequently appear together in transactions.

Support Levels (Blue Line)

Support values are generally low (<0.2), indicating that while the rules exist, they apply to a smaller portion of total transactions.

Frequently occurring consequents include **soda, eggs, pasta, and sugar**, meaning these are commonly purchased products

Inferences and Recommendations

Strong Item Associations:

Yogurt and Pork

- Support: **15.2%** of transactions include both items.
- Confidence: **42.7%** (If a customer buys pork, there is a **42.7%** chance they also buy yogurt).
- Lift: **1.11** (A slight positive association; customers are **1.11 times** more likely to buy yogurt when buying pork).

Yogurt and Sandwich Bags

- Support: **15.5%** of transactions contain both items.
- Confidence: **42.2%** (If a customer buys sandwich bags, there is a **42.2%** chance they also buy yogurt).
- Lift: **1.10** (Moderate relationship).

Yogurt and Lunch Meat

- Support: **16.2%**, Confidence: **40.9%**, Lift: **1.06**
- Customers buying **lunch meat** are **1.06 times** more likely to buy yogurt.

High Lift Values Indicating Strong Relationships

Milk & Coffee/Tea

- **Lift: 1.35** (A strong association, meaning customers who buy milk are 1.35 times more likely to buy coffee/tea).
- **Recommendation:** Bundle **milk and coffee/tea** together as a morning essentials pack with a **small discount**.

Flour & Butter

- **Lift: 1.29** (Customers buying flour are 1.29 times more likely to buy butter).
- **Recommendation:** Offer a **baking essentials combo** (flour + butter + sugar) with a **discount for bulk purchases**.

Frequently Bought Together Items with Moderate Lift

Eggs & Sugar

- **Lift: 1.22** (Moderate relationship; sugar is often bought with eggs, possibly for baking).
- **Recommendation:** Promote a **"Weekend Baking Pack"** with eggs, sugar, and flour to drive sales.

IMPLICATIONS FOR PROMOTIONS & DISCOUNTS

Create "Breakfast Essentials Combo" including **yogurt, pork, and sandwich bags** (as they often appear together).

"Healthy Snack Combo" with **yogurt and lunch meat** to encourage bundled purchases

Combo Offers:

IMPLICATIONS FOR PROMOTIONS & DISCOUNTS

Milk + Cereal + Fruits (Breakfast Combo)

- **Offer:** Buy any two, get 10% off on the third item.
- **Reason:** Breakfast staples are often purchased together.

Bread + Butter + Jam (Morning Essentials)

- **Offer:** Flat ₹20 off when bought together.
- **Reason:** A common combination, encourages bulk purchase.

Pasta + Sauce + Cheese (Italian Night Combo)

- **Offer:** Buy Pasta and Sauce, get Cheese at 30% off.
- **Reason:** Customers buying pasta usually need sauce and cheese.

Soft Drinks + Chips (Snack Combo)

- **Offer:** Buy 1 large pack of chips, get ₹10 off on a soft drink.
- **Reason:** Frequently purchased for snacking or parties.

IMPLICATIONS FOR PROMOTIONS & DISCOUNTS

Rice + Lentils + Spices (Daily Cooking Essentials)

- **Offer:** Buy 5kg Rice and 2kg Lentils, get 10% off on Spices.
- **Reason:** Encourages larger purchases for regular household needs.

Baby Food + Diapers + Wipes (Parenting Pack)

- **Offer:** Buy any 2 items, get 15% off on the third item.
- **Reason:** Essential products for parents, ensuring repeat purchases.

Shampoo + Conditioner + Body Wash (Personal Care Pack)

- **Offer:** Flat ₹50 off on the total when all three are bought together.
- **Reason:** Customers tend to buy personal care items in sets.

Tea/Coffee + Biscuits + Sugar (Tea Break Combo)

- **Offer:** Buy Tea/Coffee and Biscuits, get Sugar at 20% off.
- **Reason:** Enhances impulse buying for daily consumption.

DISCOUNT STRATEGIES

DISCOUNT STRATEGIES

Loyalty Points for Repeat Purchases

Example: Buy groceries worth ₹2000, get 100 reward points (to encourage return visits).

Late-Night Discounts on Fast-Moving Items

Example: 20% off on bakery items after 8 PM (to clear stock).

Bulk Purchase Discounts

Example: Buy 5kg rice, get ₹50 off on dal/spices (to promote larger cart sizes)

BUSINESS RECOMMENDATIONS

Optimize Product Placement:

- Place high-association items (e.g., bread near butter, milk near cereal) together.
- Cross-category placements for high-value combos.

Targeted Discounts:

- Use loyalty card data to offer personalized discounts on frequently bought items.
- Implement evening discounts on perishable goods to minimize waste.

E-commerce & Digital Strategies:

- Push notifications for personalized combo offers.
- Special discounts for online bulk purchases.

WORKING FILES

1. Jupyter Notebook - MRA PART - A, MRA PART - B

2. Tableau Work Book – PDF AND LINK

Part – A

https://public.tableau.com/app/profile/benita.merlin.e/viz/MRATABLEAUPARTA/SalesvsQuantityOrdered

PART - B

https://public.tableau.com/app/profile/benita.merlin.e/viz/MRATABLEAUPARTB/Totalnumberofuniquetransactions

3.KNIME WORKFLOW - PART-A, PART-B

THANK YOU