

- Customer Segmentation and Retention Strategy using RFM Analysis on B2B Sales Data with KNIME -

Agenda

01 Introduction and Problem Statement

02 About the Dataset

03 Univariate Analysis

04 Bivariate Analysis

Agenda

05 Multivariate Analysis

06 Time-Based Trend Analysis

07 RFM Analysis (Recency, Frequency, Monetary)

08 Inferences from RFM Segmentation

Introduction and Problem Statement

Business Need:

The company has been growing but lacks the internal expertise to extract actionable insights from its transaction data. They are unable to fully understand customer behavior, segment customers effectively, or create targeted marketing strategies.

Our Objective:

To analyze the past three years of customer transaction data in order to:

- **Uncover Buying Patterns:** Understand how customers interact with the products.
- **Segment Customers:** Categorize customers into meaningful segments based on purchasing behavior.
- **Recommend Strategies:** Provide actionable insights for improving customer retention and marketing efforts.

By leveraging data science techniques, the company can enhance customer satisfaction, optimize marketing strategies, and drive revenue growth.

About the Data

Summary Statistics

Dataset Overview:

- **Source:** Internal transaction records of the company
- **Duration Covered:** *January 6, 2018* to *May 31, 2020*
- **Total Orders:** 2,747 transactions
- **Number of Customers:** 89 unique customers
- **Number of Products:** 109 unique products

Key Data Fields:

- **Order Details:** ORDERNUMBER, ORDERDATE, SALES, STATUS
- **Customer Info:** CUSTOMERNAME, COUNTRY, CITY, DEALSIZE
- **Product Info:** PRODUCTCODE, PRODUCTLINE, MSRP, PRICEEACH
- **Behavioral Metrics:** DAYS_SINCE_LASTORDER, QUANTITYORDERED

Assumptions Made:

- Data is clean and representative of actual sales
- Dates are correctly formatted and reflect transaction history
- No major missing values in critical fields (like SALES, ORDERDATE)

Summary Statistics: (Selected Columns):

Metric	Mean	Min	Max	Std Dev
Quantity Ordered	35.1	6	97	9.76
Price Each (USD)	101.1	26.88	252.87	42.04
Sales per Transaction	3,553.05	482.13	14,082.80	1,838.95
Days Since Last Order	1,757.09	42	3,562	819.28

Statistical Summary

Overview of Transaction Data (Jan 2018 – May 2020)

- **Total Orders:** 2,747
- **Total Sales Value:** \$8,707,246
- **Average Order Value:** \$3,169
- **Unique Customers:** 89
- **Product Lines Sold:** 7

Top Performing Product Line

- **Classic Cars** – Contributed the highest share of total sales

Most Frequently Ordered Quantity

- **30 units per order** (most common quantity)

Sales Distribution

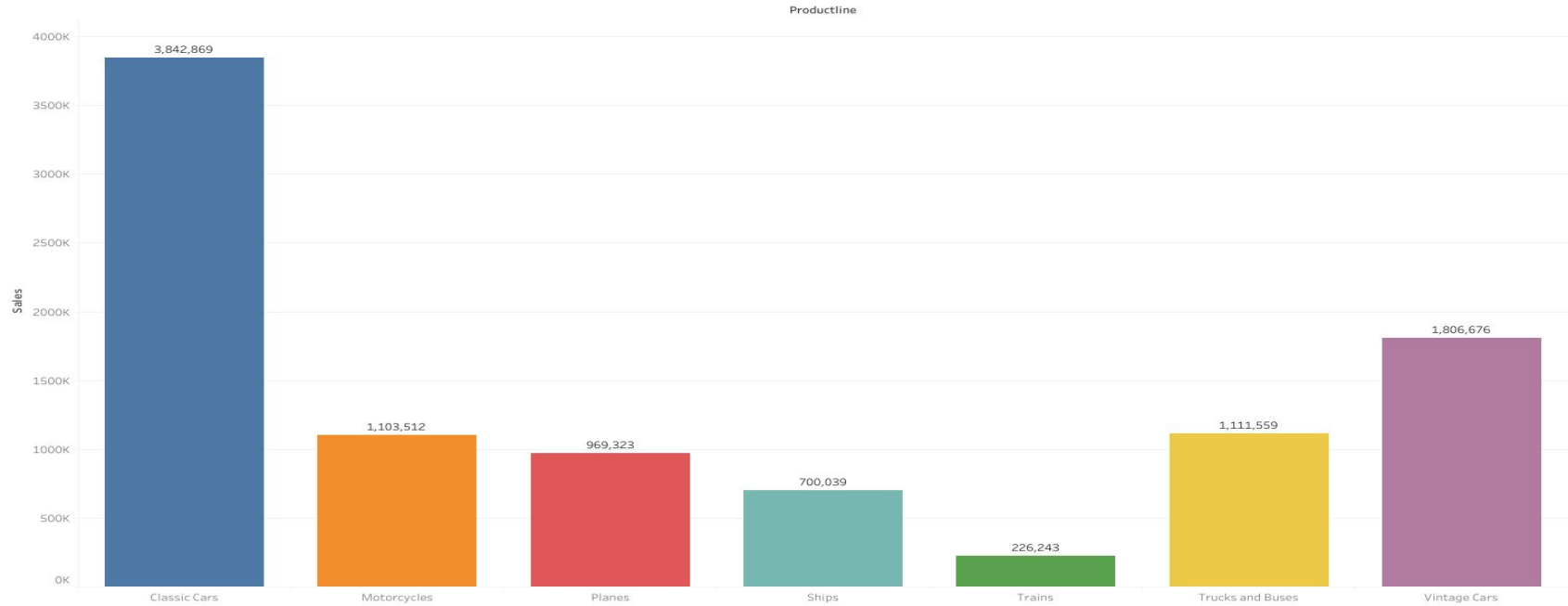
- Right-skewed with majority of orders between \$1,000 and \$5,000
- High-value orders above \$10,000 are rare but impactful

Univariate analysis

univariate analysis

Sales by Product Line

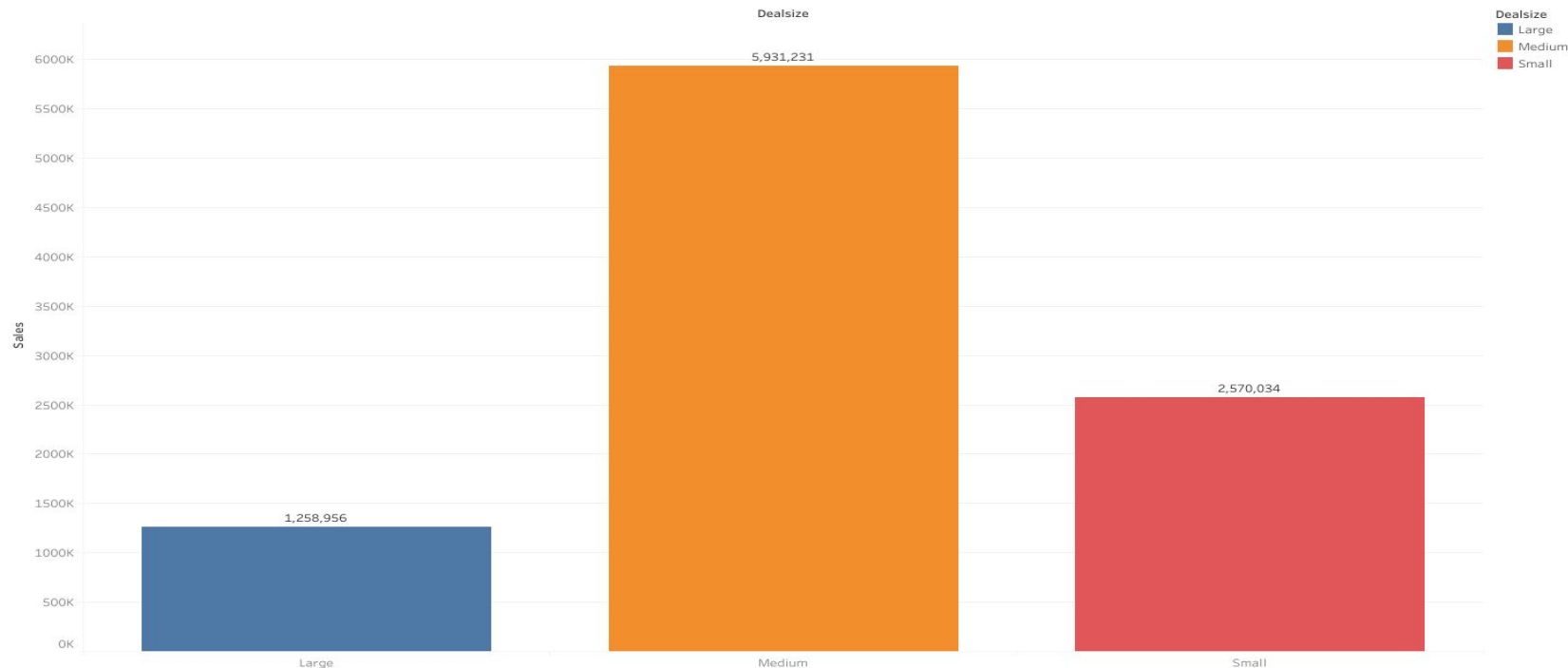
Sales by Product Line



Classic Cars dominate the product portfolio with total sales of **\$3,842,869**, indicating strong customer preference. On the other hand, **Trains** contribute the least with **\$226,243**, suggesting a niche or underperforming category.

Sales by Deal Size

Sales by Deal Size

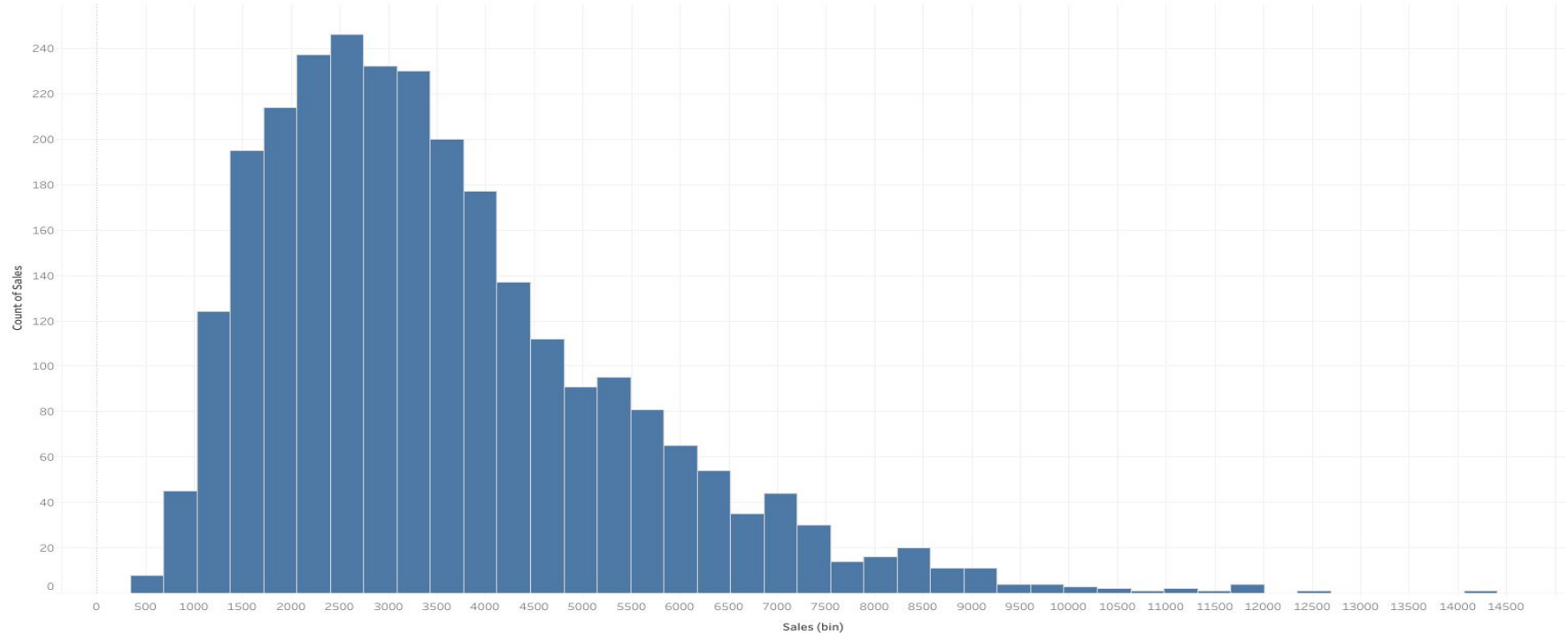


Medium deal sizes generate the highest revenue of **\$5,931,231**, indicating that mid-range products are most popular among customers. **Large deal sizes**, though potentially high in value, contribute only **\$1,258,956**, suggesting lower frequency or uptake.

Sales Distribution



Sales Distribution

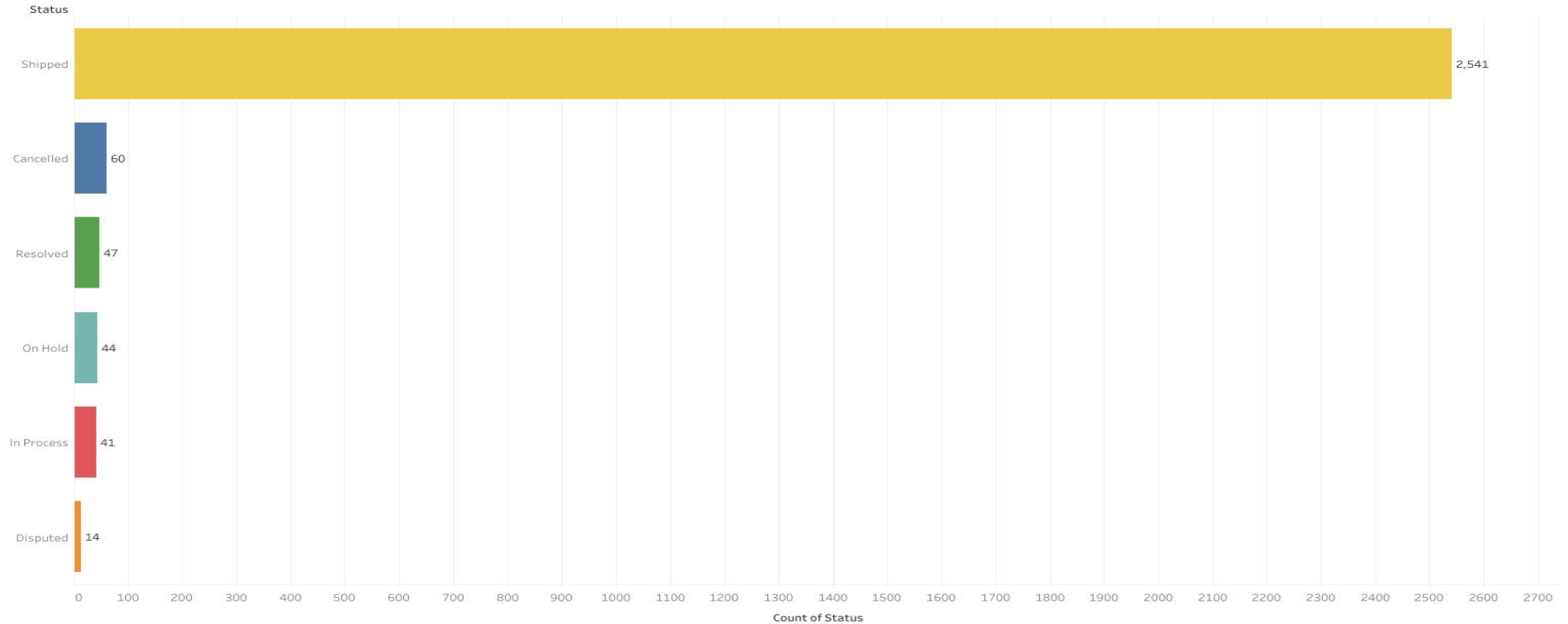


The sales distribution is **right-skewed**, with most orders falling between **\$1,000 and \$5,000**. **High-value transactions above \$10,000 are rare but significant**, highlighting occasional bulk or premium purchases.

Order Count by Status



Order Count by Status



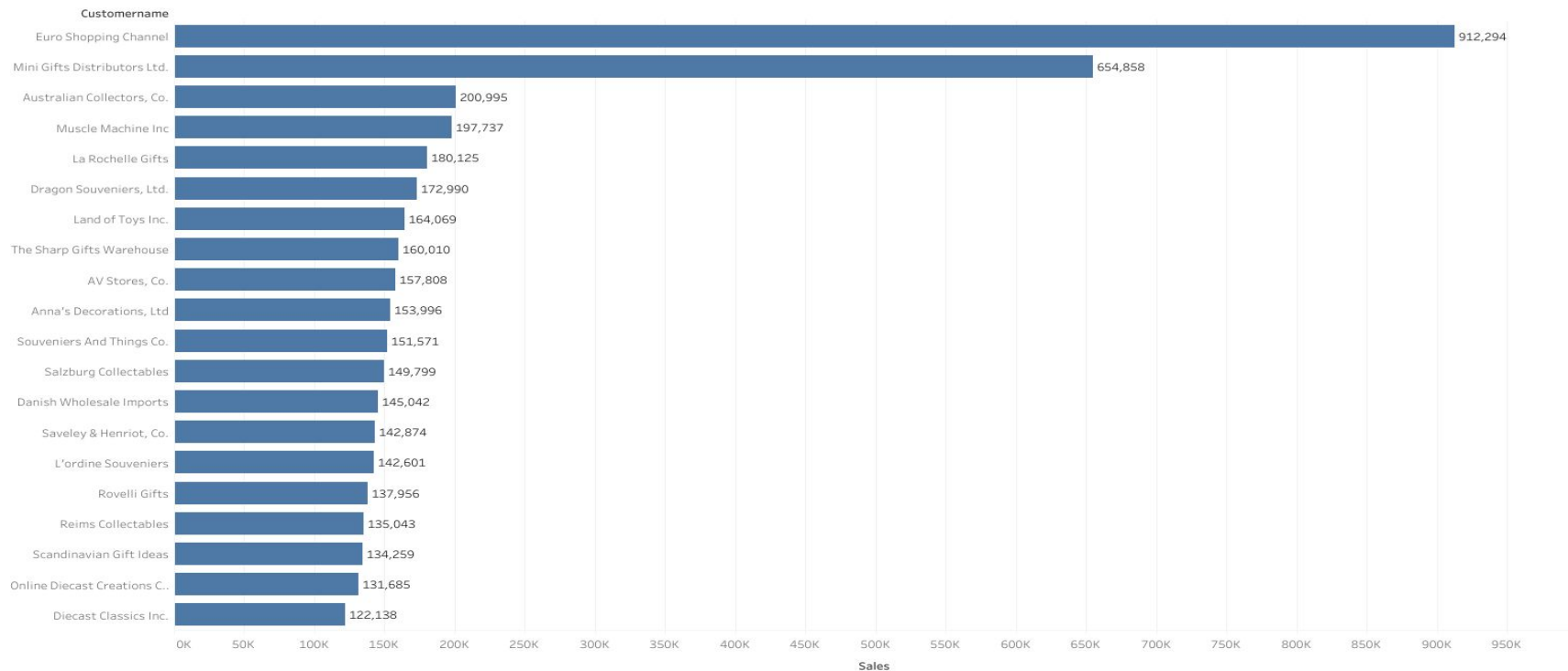
The majority of orders are marked as **Shipped (2,541)**, indicating efficient fulfillment. **Disputed orders (14)** are minimal, reflecting good customer satisfaction or issue resolution processes.

Bivariate Analysis

Top 20 Customers by Sales



Top 20 Customers by Sales

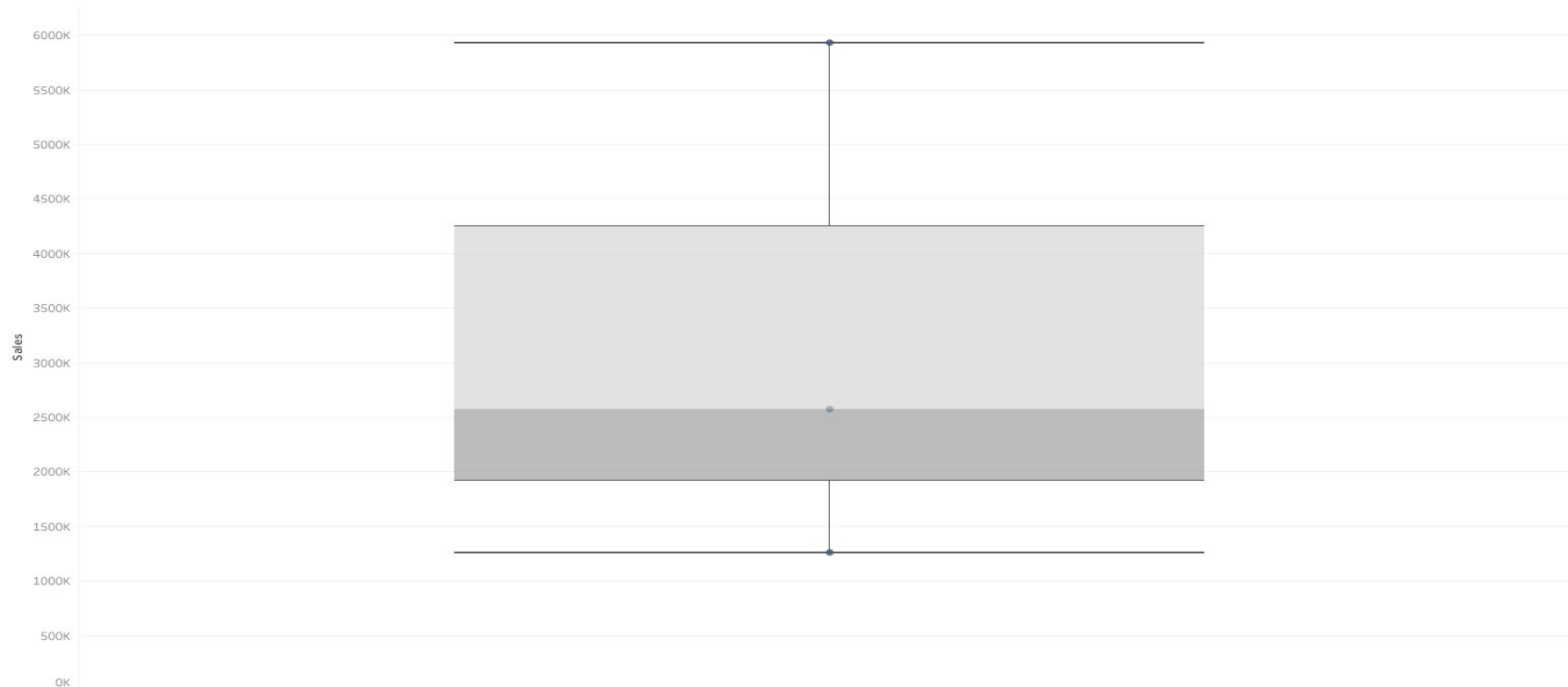


Euro Shopping Channel leads with total sales of **\$912,294**, followed by **Mini Gifts Distributors Ltd.** with **\$654,858**. The **20th customer, Diecast Classics Inc.**, generated **\$122,138**, highlighting the wide gap between top and lower-ranked customers.

Sales Spread by Deal Size



Sales Spread by Deal Size

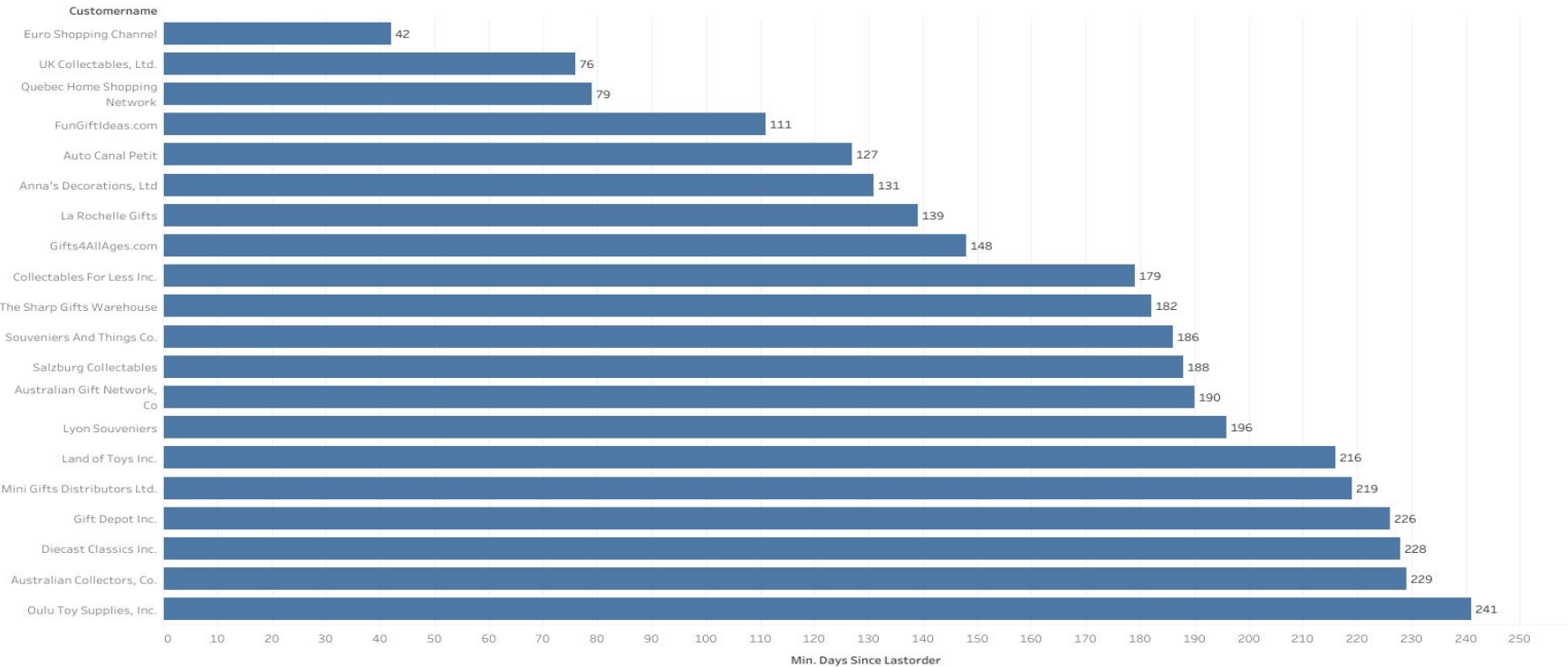


Median value: \$2,570,034. Interquartile range (IQR): \$1,914,495 to \$4,250,633. Upper Whisker: \$5,931,231, Lower Whisker: \$1,258,956. Most sales fall within the IQR, with a few high-value deals extending the upper range.

Top 20 Most Recently Active Customers



Top 20 Most Recently Active Customers



Euro Shopping Channel is the most recently active customer, with their last transaction occurring **42 days ago**. **UK Collectables, Ltd.** and **Quebec Home Shopping Network** followed closely at **76** and **79 days** ago respectively, indicating strong recent engagement from these clients.



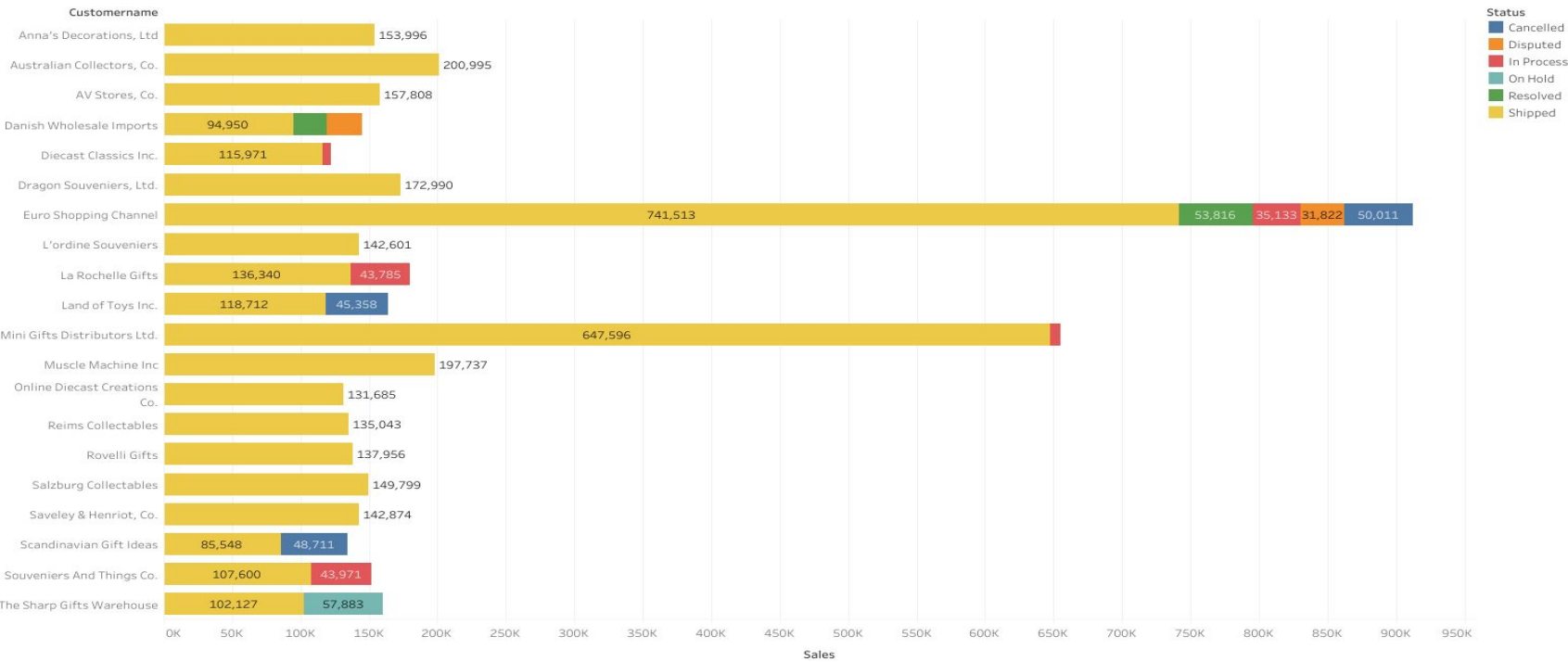
Multivariate analysis

Multivariate analysis

Top 20 Customers – Sales Breakdown by Order Status



Top 20 Customers – Sales Breakdown by Order Status

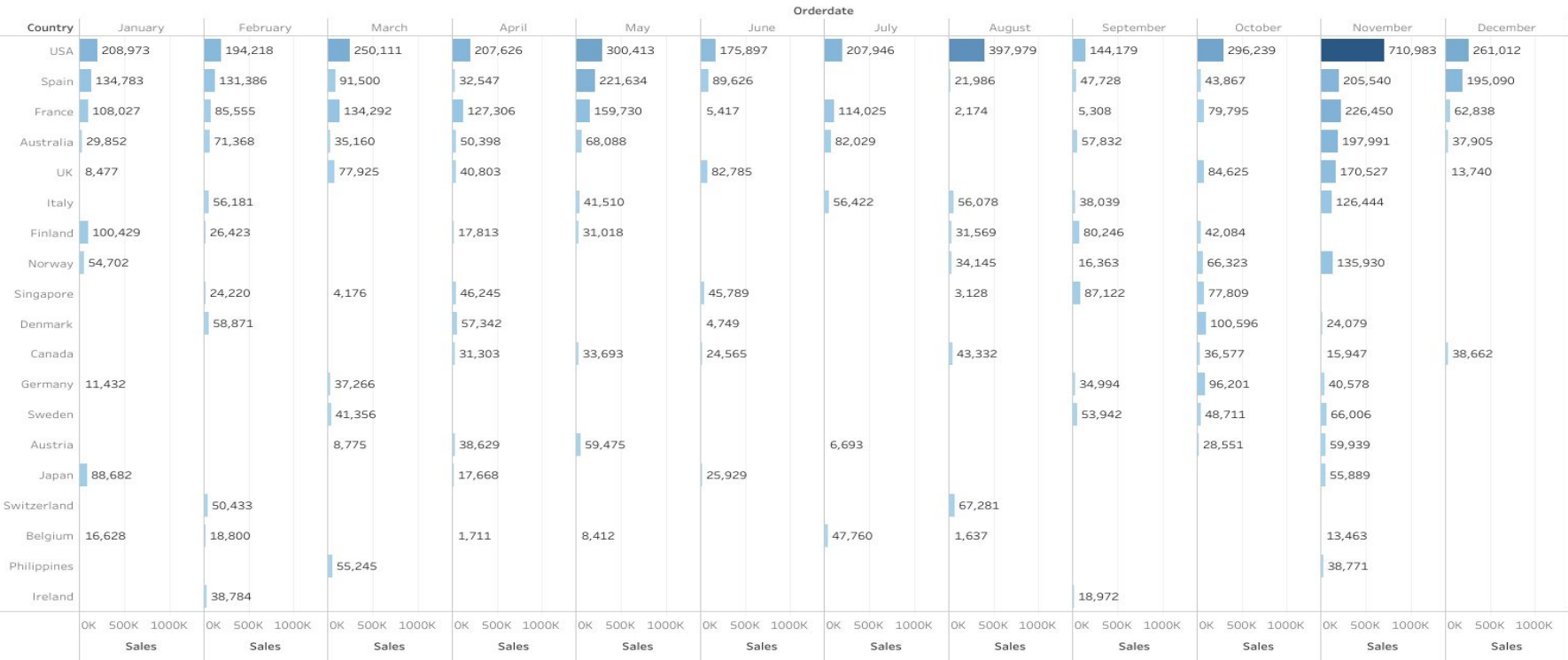


Euro Shopping Channel leads overall with total sales, where: **Shipped orders** account for **\$741,513**. Other statuses include **Resolved: \$53,816**, **In Progress: \$35,133**, **Disputed: \$31,822**, and **Cancelled: \$50,011**. **Mini Gifts Distributors Ltd.** follows with **\$647,596** in **Shipped** and **\$7,262** in **In Progress** status. **Diecast Classics Inc.**, ranked 20th, has shipped orders worth **\$115,971**, and **\$6,167** in **In Progress**

Country vs Sales vs Order Date



Country vs Sales vs Order Date



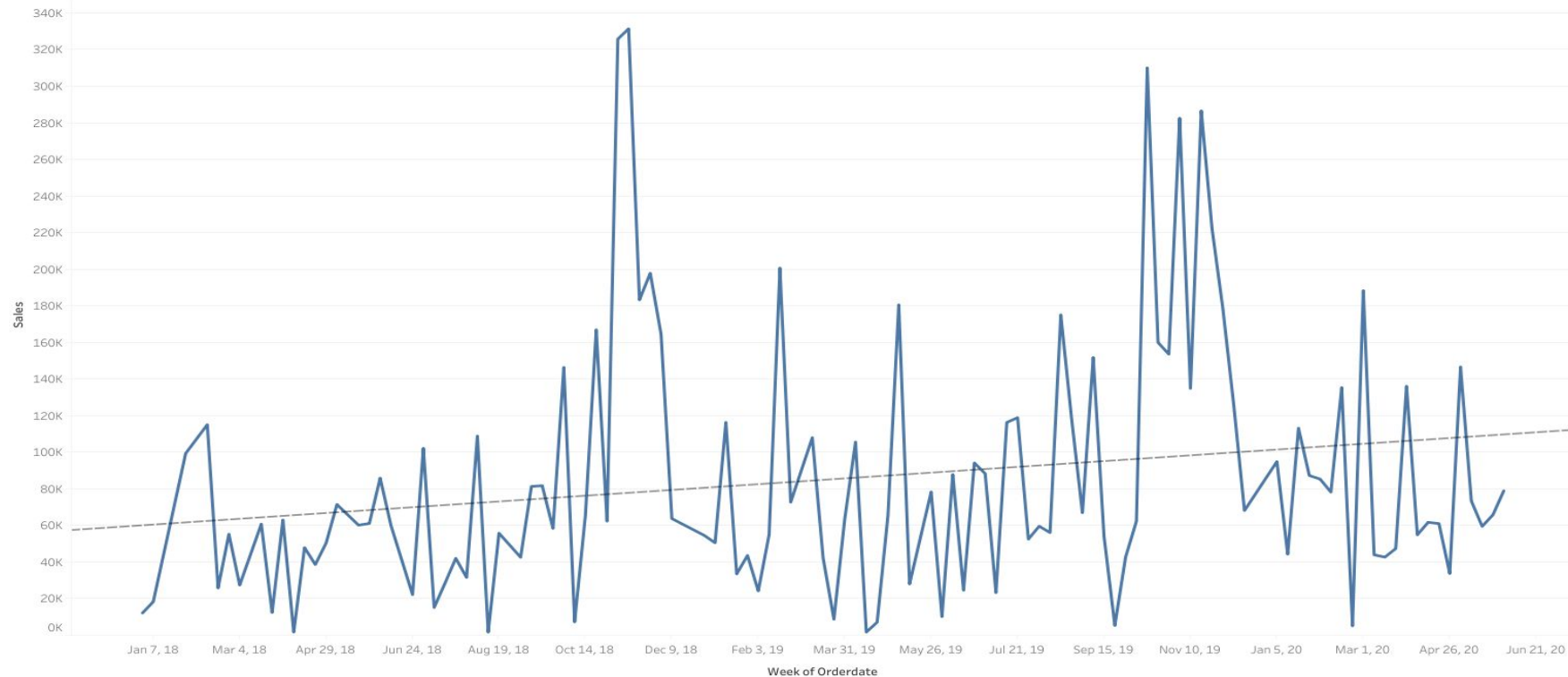
Across the 3-year span, the **USA consistently dominates monthly sales**, maintaining the top position from **January through December**. **Spain** secures the **second-highest sales** during **January, February, May, June, and December**. **France** ranks second in **March, April, July, and November**. **Singapore** rises in **September**, **Switzerland** in **August**, and **Canada** in **October**. This indicates a diverse seasonal performance among countries, with **regional sales peaks varying by month**.

Time-Based Trends analysis

Weekly Sales Pattern



Weekly Sales Pattern

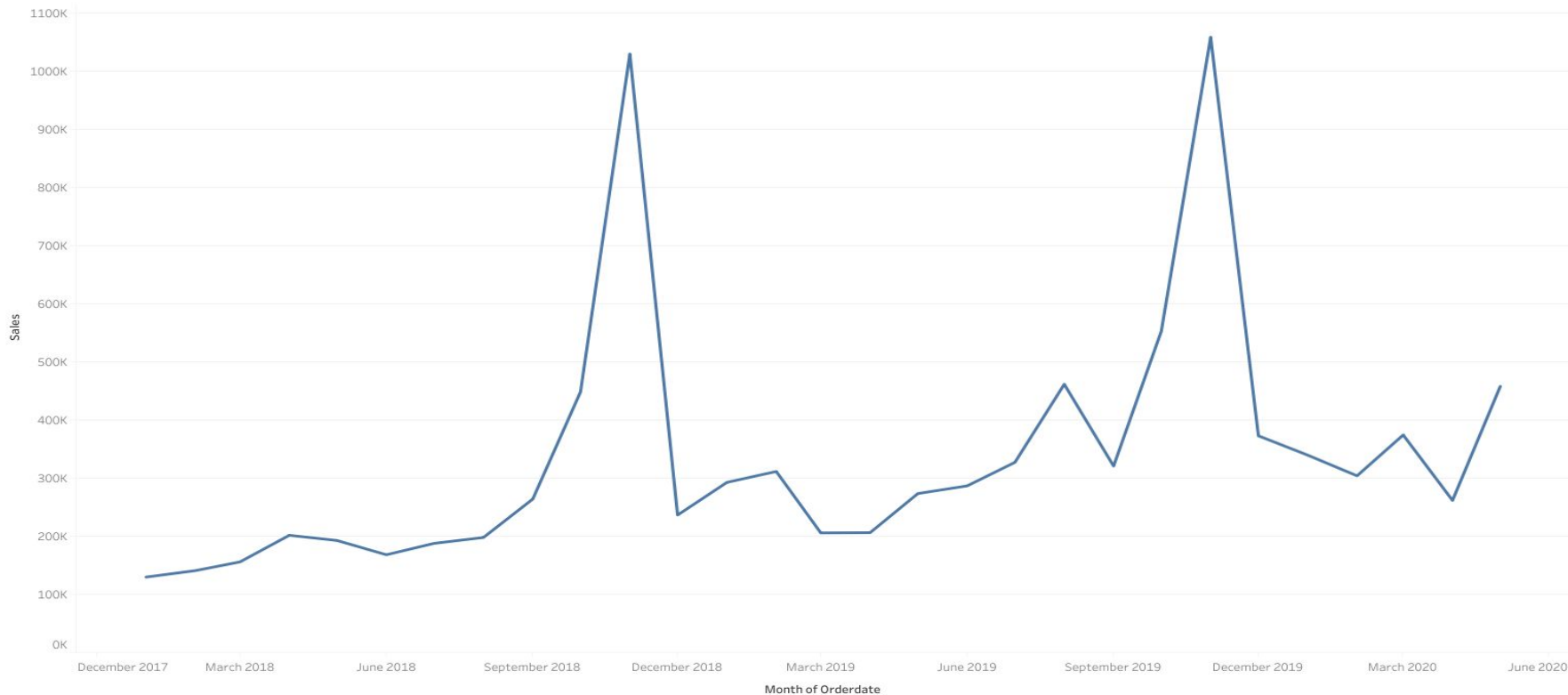


Weekly sales exhibit **high volatility**, marked by **frequent spikes and dips** across the 3-year period. A few weeks show **exceptional peaks**, notably crossing **\$300,000**, reflecting high-value transactions. Despite the fluctuations, the **overall trend shows a gradual upward movement**, indicating **steady growth in sales over time**.

Monthly Sales Trend



Monthly Sales Trend

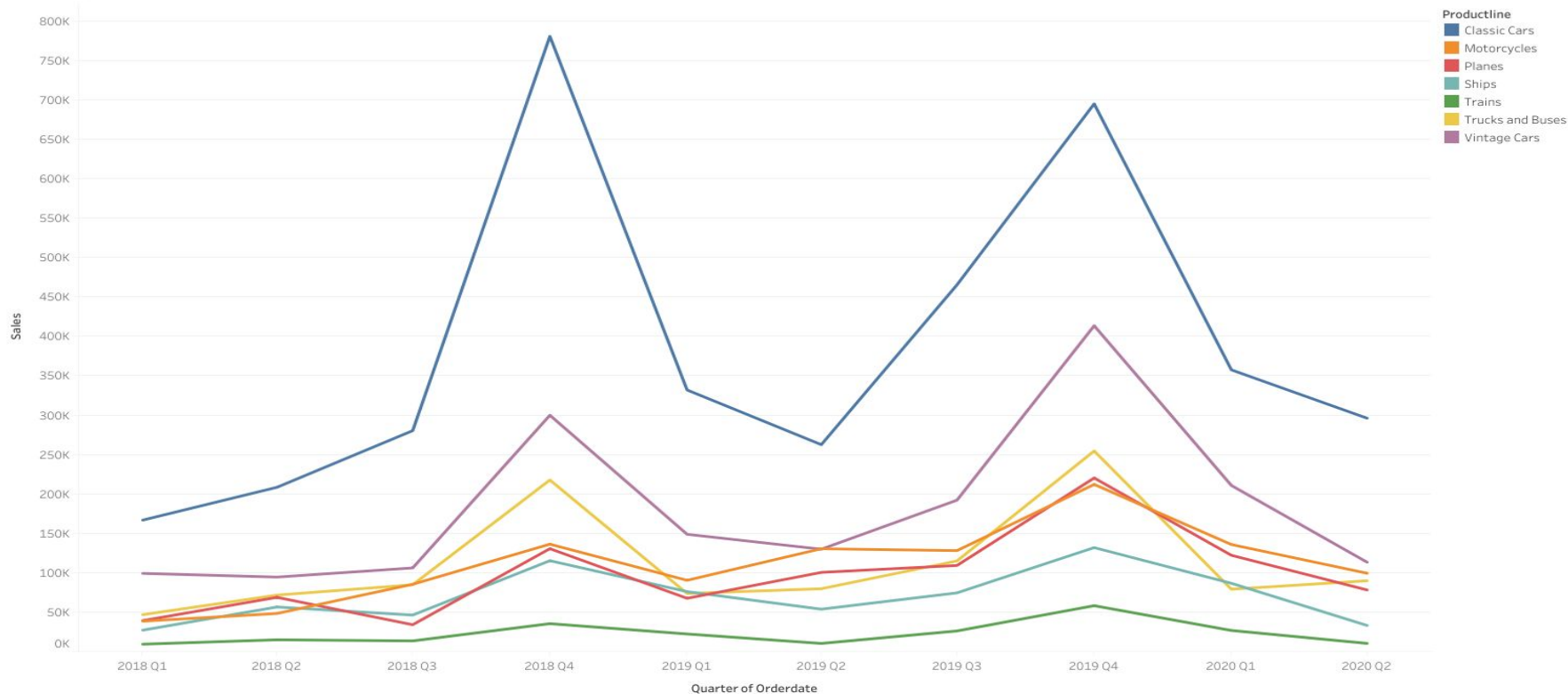


Sales consistently **peak during the months of September to December** across both 2018 and 2019, suggesting a **strong seasonal demand** in the final quarter of each year. This trend highlights the importance of **Q4 for revenue generation**, possibly due to **holiday season promotions or year-end bulk orders**.

Quarterly Product Line Performance



Quarterly Product Line Performance

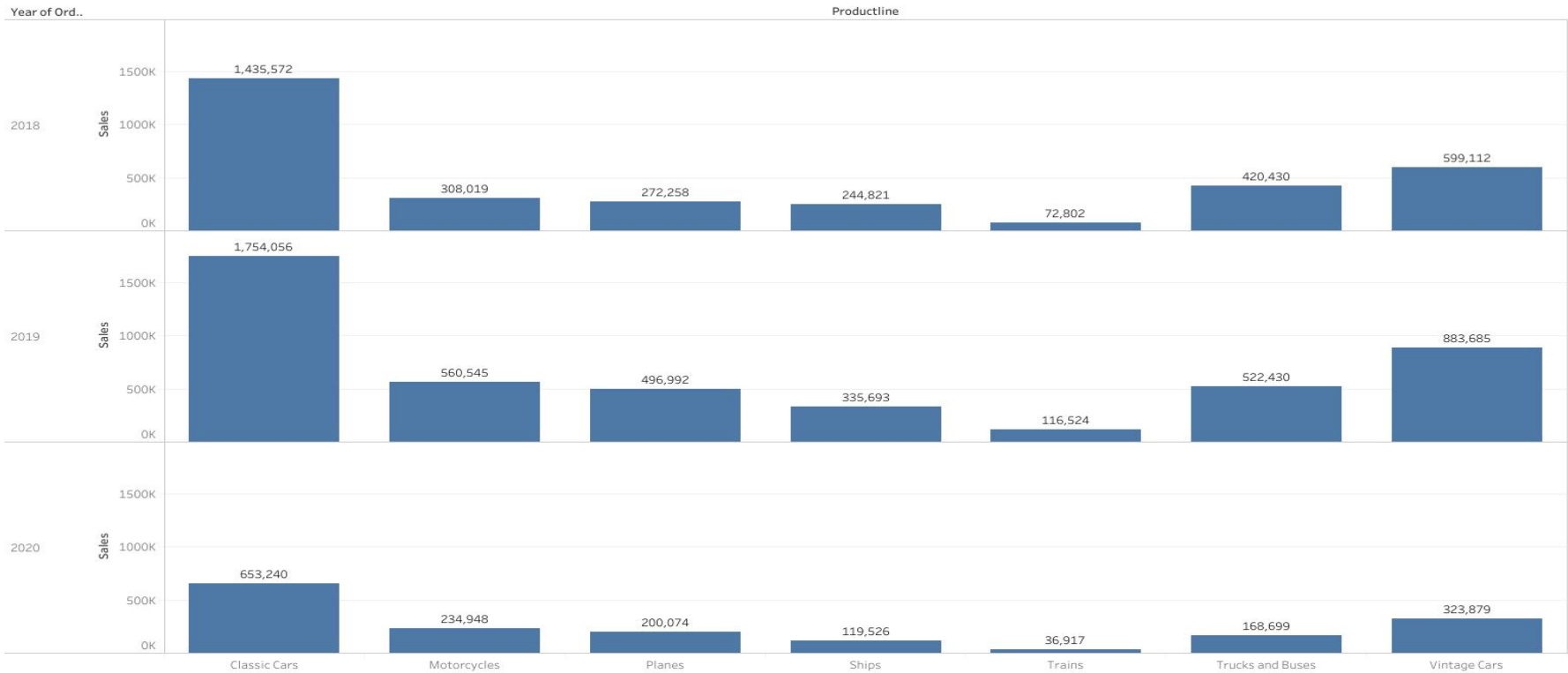


Classic Cars consistently dominate sales across all quarters, peaking in **Q4 of each year (2018 & 2019)**, indicating strong seasonal demand. **Vintage Cars** and **Trucks & Buses** also show significant spikes during Q4, suggesting they contribute heavily to year-end revenue. In contrast, **Trains** and **Planes** show relatively lower and stable performance, indicating niche or low-priority segments.

Yearly Sales Overview



Yearly Sales Overview



Classic Cars remained the top-selling product line across all three years, with a peak in **2019** at over **1.75 million**. Most product lines also saw their highest sales in **2019**, indicating a strong year overall. However, sales **declined sharply in 2020** across all product lines — likely due to external market factors (e.g., economic downturn, pandemic).

Summary of EDA & Time-Based Trends :

♦ Tools Used:

- Tableau (*for Exploratory Data Analysis & Trend Analysis*)



♦ Univariate Analysis:

- **Classic Cars** consistently dominated sales across all years.
- Customer purchases peaked between **September and December**.
- Sales were more concentrated in **specific product lines** such as Classic and Vintage Cars.

♦ Bivariate & Multivariate Analysis:

- **Sales and Product Line** showed a strong correlation, with Classic and Vintage Cars significantly outperforming others.
- The combination of **Product Line and Year** highlighted clear seasonal patterns and growth trends.
- **Geography, Product Type, and Sales** together revealed clustered high-performing categories.

♦ Time-Based Trends:

- **Monthly trends** showed recurring **spikes in Q4** (Sept–Dec) during **2018 and 2019**, indicating seasonal demand.
- **Quarterly analysis** confirmed strong performance in **Q4**, particularly for Classic Cars.
- **2020** showed a drop in sales across all product lines, likely due to **external disruptions**.

RFM Analysis

Objective of RFM Analysis



- To segment customers based on their purchasing behavior
- Helps in identifying best customers, loyal ones, and those at risk of churning

♦ Tool Used

- KNIME (for RFM calculation, binning, and segmentation)



♦ Data Source

- Excel (Transactional data: Jan 2018 – May 2020)

♦ RFM Parameters Used

- **Recency (R):** Days since last order (from snapshot date – 18 April 2025)
- **Frequency (F):** Total number of orders placed
- **Monetary (M):** Total revenue generated by customer

♦ Assumptions Made

- Snapshot date taken as **18 April 2025**
- Each RFM metric binned into 4 categories using quartiles
- Segmentation done based on combined R, F, M bin values



RFM Binning Strategy & Segmentation Logic :





◆ Binning Logic Applied :

Metric	Bin 1	Bin 4	Scoring Logic
Recency	Most Recent	Least Recent	Lower is better
Frequency	Least Often	Most Often	Higher is better
Monetary	Low Spender	High Spender	Higher is better

Assumption: Lower Recency = more recent transaction = more engaged.

◆ Segmentation Logic :



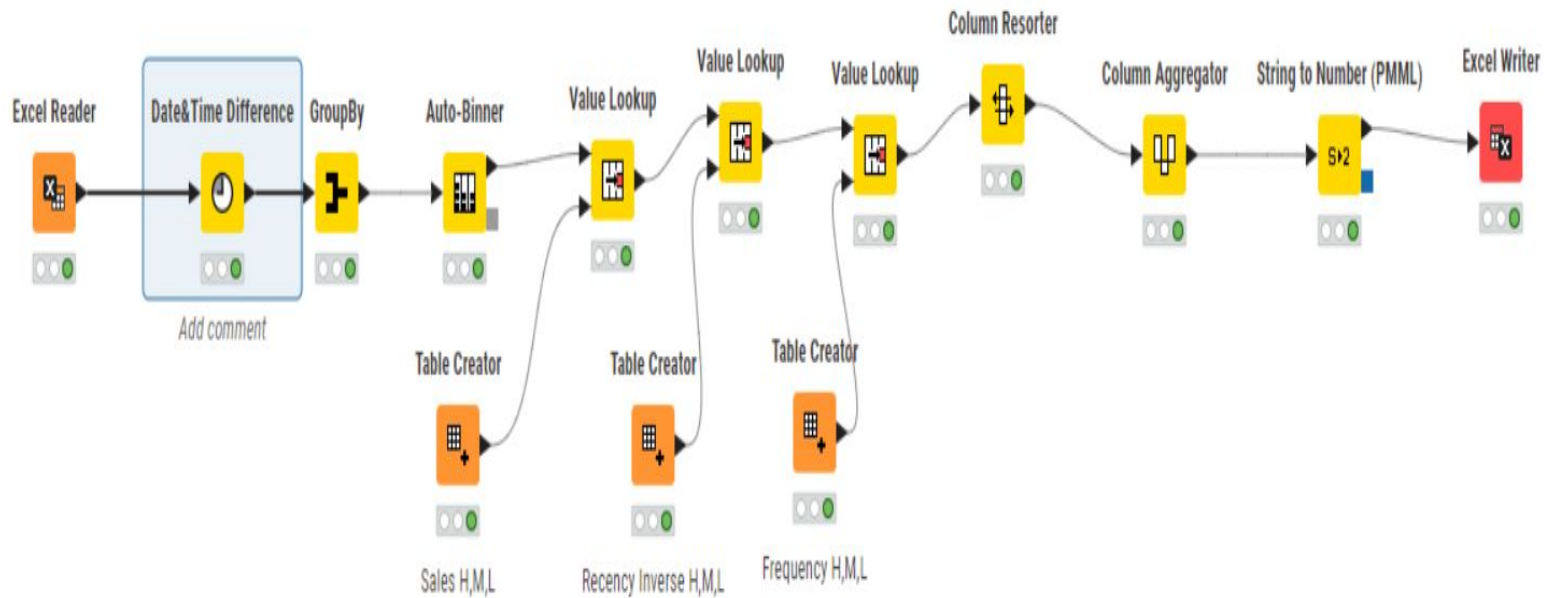
-  **Best Customers:**
 - Recency = 1
 - Frequency = 3 or 4
 - Monetary = 3 or 4
-  **Loyal Customers:**
 - Recency = 1 or 2
 - Frequency = 2 or more
 - Monetary = 2 or more
-  **Customers on the Verge of Churn:**
 - Recency = 2 or 3
 - Frequency = 1 or 2
 - Monetary = 1 or 2
-  **Lost Customers:**
 - Recency = 3 or 4
 - Frequency = 1 or 2
 - Monetary = 1 or 2

Processed RFM Table (Top 10 Rows)

CUSTOMERNAME	Recency	ORDERNUMBER	SALES	ORDERNUMBER [Binned]	SALES [Binned]	Recency [Binned]	Recency H,M,L	Frequency H,L,M	monetary H,L,M	Customers Segmentation	Concatenate
AV Stores, Co.	2181.392157	51	157807.81	Bin 4	Bin 4	Bin 3	2	4	4	Loyal Customers	2, 4, 4
Alpha Cognac	2245.55	20	70488.44	Bin 1	Bin 1	Bin 3	2	1	1	Customers on Verge of Churn	2, 1, 1
Amica Models & Co.	2063.038462	26	94117.26	Bin 2	Bin 3	Bin 1	4	2	3	Lost Customers	4, 2, 3
Anna's Decorations, Ltd	2180.217391	46	153996.13	Bin 4	Bin 4	Bin 2	3	4	4	Customers on Verge of Churn	3, 4, 4
Atelier graphique	2304.428571	7	24179.96	Bin 1	Bin 1	Bin 4	1	1	1	Loyal Customers	1, 1, 1
Australian Collectables, Ltd	2204.304348	23	64591.46	Bin 2	Bin 1	Bin 3	2	2	1	Customers on Verge of Churn	2, 2, 1
Australian Collectors, Co.	2223.072727	55	200995.41	Bin 4	Bin 4	Bin 3	2	4	4	Loyal Customers	2, 4, 4
Australian Gift Network, Co	2185	15	59469.12	Bin 1	Bin 1	Bin 3	2	1	1	Customers on Verge of Churn	2, 1, 1
Auto Assoc. & Cie.	2029.944444	18	64834.32	Bin 1	Bin 1	Bin 1	4	1	1	Lost Customers	4, 1, 1
Auto Canal Petit	2191.555556	27	93170.66	Bin 3	Bin 3	Bin 3	2	3	3	Loyal Customers	2, 3, 3

- **ORDERNUMBER** is used as the **Frequency** metric in the RFM model.
- **SALES** is used as the **Monetary** metric in the RFM model.
- **Recency** is the number of days since the last purchase (already calculated).

◆ THE KNIME WORKFLOW :



Inferences from RFM Analysis

INFERENCES FROM RFM ANALYSIS

Top 5 Best Customers (High R, F, M Scores) :

CUSTOMER NAME	Recency	ORDERNUMBER	SALES	ORDERNUMBER [Binned]	SALES [Binned]	Recency [Binned]	Recency H,M,L	Frequency H,L,M	monetary H,L,M	Customers Segmentation	Concatenate
Muscle Machine Inc	2282.916667	48	197736.94	Bin 4	Bin 4	Bin 4	1	4	4	Best Customers	1, 4, 4
Rovelli Gifts	2345.375	48	137955.72	Bin 4	Bin 4	Bin 4	1	4	4	Best Customers	1, 4, 4
Dragon Souveniers, Ltd.	2412.976744	43	172989.68	Bin 4	Bin 4	Bin 4	1	4	4	Best Customers	1, 4, 4
Saveley & Henriot, Co.	2286.609756	41	142874.25	Bin 4	Bin 4	Bin 4	1	4	4	Best Customers	1, 4, 4
Danish Wholesale Imports	2294.722222	36	145041.6	Bin 4	Bin 4	Bin 4	1	4	4	Best Customers	1, 4, 4

Top 5 Loyal Customers (High R, F, M Scores) :

CUSTOMER NAME	Recency	ORDERNUMBER	SALES	ORDERNUMBER [Binned]	SALES [Binned]	Recency [Binned]	Recency H,M,L	Frequency H,L,M	monetary H,L,M	Customers Segmentation	Concatenate
Australian Collectors, Co.	2223.072727	55	200995.41	Bin 4	Bin 4	Bin 3	2	4	4	Loyal Customers	2, 4, 4
AV Stores, Co.	2181.392157	51	157807.81	Bin 4	Bin 4	Bin 3	2	4	4	Loyal Customers	2, 4, 4
Online Diecast Creations Co.	2234.705882	34	131685.3	Bin 4	Bin 4	Bin 3	2	4	4	Loyal Customers	2, 4, 4
Toys4Grown Ups.com	2206.333333	30	104561.96	Bin 3	Bin 3	Bin 3	2	3	3	Loyal Customers	2, 3, 3
Signal Gift Stores	2256.344828	29	82751.08	Bin 3	Bin 2	Bin 3	2	3	2	Loyal Customers	2, 3, 2

Top 5 Customers on Verge of Churn (High R, F, M Scores) :

CUSTOMER NAME	Recency	ORDERNUMBER	SALES	ORDERNUMBER [Binned]	SALES [Binned]	Recency [Binned]	Recency H,M,L	Frequency H,L,M	monetary H,L,M	Customers Segmentation	Concatenate
Euro Shopping Channel	2090.111969	259	912294.11	Bin 4	Bin 4	Bin 1	4	4	4	Customers on Verge of Churn	4, 4, 4
Mini Gifts Distributors Ltd.	2119.344444	180	654858.06	Bin 4	Bin 4	Bin 2	3	4	4	Customers on Verge of Churn	3, 4, 4
La Rochelle Gifts	1951.09434	53	180124.9	Bin 4	Bin 4	Bin 1	4	4	4	Customers on Verge of Churn	4, 4, 4
Land of Toys Inc.	2155.204082	49	164069.44	Bin 4	Bin 4	Bin 2	3	4	4	Customers on Verge of Churn	3, 4, 4
Anna's Decorations, Ltd	2180.217391	46	153996.13	Bin 4	Bin 4	Bin 2	3	4	4	Customers on Verge of Churn	3, 4, 4

Top 5 Lost Customers (High R, F, M Scores) :

CUSTOMER NAME	Recency	ORDERNUMBER	SALES	ORDERNUMBER [Binned]	SALES [Binned]	Recency [Binned]	Recency H,M,L	Frequency H,L,M	monetary H,L,M	Customers Segmentation	Concatenate
Tokyo Collectables, Ltd	1980.46875	32	120562.74	Bin 3	Bin 3	Bin 1	4	3	3	Lost Customers	4, 3, 3
Marta's Replicas Co.	2036.62963	27	103080.38	Bin 3	Bin 3	Bin 1	4	3	3	Lost Customers	4, 3, 3
Amica Models & Co.	2063.038462	26	94117.26	Bin 2	Bin 3	Bin 1	4	2	3	Lost Customers	4, 2, 3
FunGiftIdeas .com	1998.807692	26	98923.73	Bin 2	Bin 3	Bin 1	4	2	3	Lost Customers	4, 2, 3
Gifts4AllAges.com	1937.692308	26	83209.88	Bin 2	Bin 2	Bin 1	4	2	2	Lost Customers	4, 2, 2