MRA PROJECT - 1

Harikrishnan M
PGP DSBA

AGENDA

Problem Statement

Data Analysis

Exploratory Data Analysis

Customer Segmentation using RFM Analysis

RFM Inferences and Identified Segments

Problem Statement

An automobile parts manufacturing company has collected data on transactions for 3 years. They do not have any in-house data science team, thus they have hired you as their consultant. Your job is to use your data science skills to find the underlying buying patterns of the customers, provide the company with suitable insights about their customers, and recommend customized marketing strategies for different segments of customers.

DATA ANALYSIS

Data Analysis

INFO

1. Dataset is having 20 variables out of which 12 are categorical, 7 Numerical and one date field
2. There is no missing values and duplicate values found.
3. Total Number of observations are 2747

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 2747 entries, 0 to 2746
Data columns (total 20 columns):
 #   Column                Non-Null Count  Dtype
---  ------                --------------  -----
 0   ORDERNUMBER           2747 non-null   int64
 1   QUANTITYORDERED       2747 non-null   int64
 2   PRICEEACH             2747 non-null   float64
 3   ORDERLINENUMBER       2747 non-null   int64
 4   SALES                 2747 non-null   float64
 5   ORDERDATE             2747 non-null   datetime64[ns]
 6   DAYS_SINCE_LASTORDER  2747 non-null   int64
 7   STATUS                2747 non-null   object
 8   PRODUCTLINE           2747 non-null   object
 9   MSRP                  2747 non-null   int64
 10  PRODUCTCODE           2747 non-null   object
 11  CUSTOMERNAME          2747 non-null   object
 12  PHONE                 2747 non-null   object
 13  ADDRESSLINE1          2747 non-null   object
 14  CITY                  2747 non-null   object
 15  POSTALCODE            2747 non-null   object
 16  COUNTRY               2747 non-null   object
 17  CONTACTLASTNAME       2747 non-null   object
 18  CONTACTFIRSTNAME      2747 non-null   object
 19  DEALSIZE              2747 non-null   object
dtypes: datetime64[ns](1), float64(2), int64(5), object(12)
memory usage: 429.3+ KB
```

Data Analysis

Data Shape

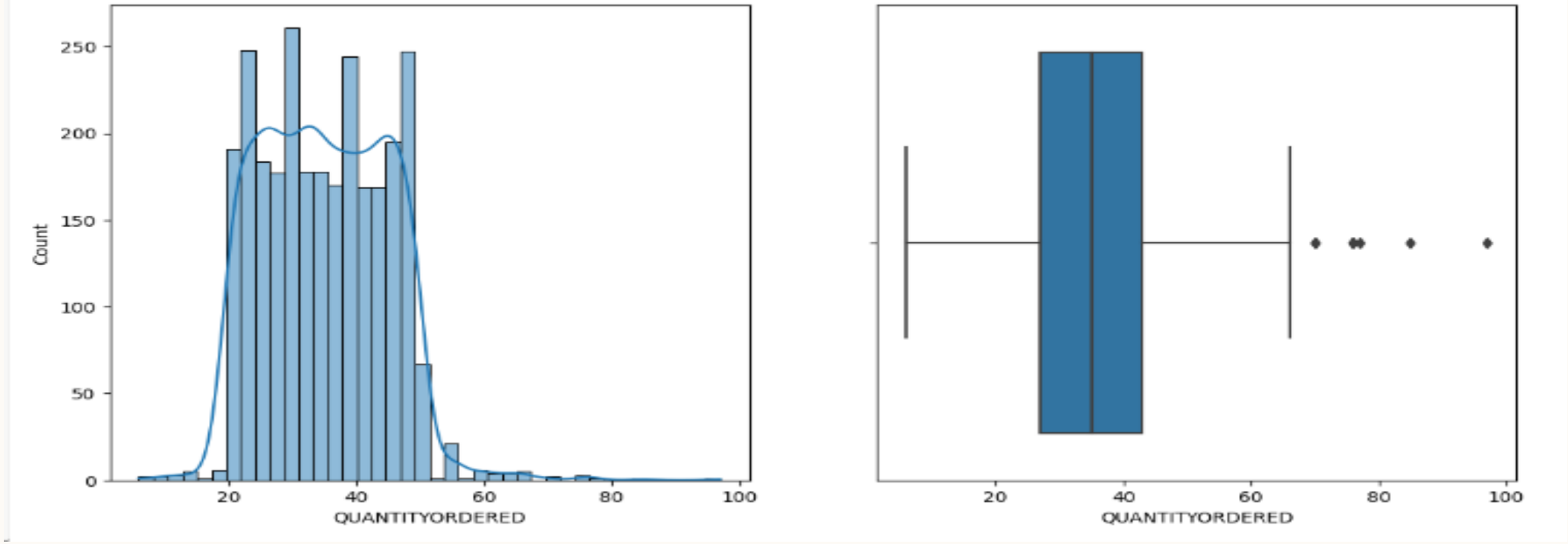
```
(2747, 20)
```

Data Summary

	count	mean	min	25%	50%	75%	max	std
ORDERNUMBER	2747.0	10259.761558	10100.0	10181.0	10264.0	10334.5	10425.0	91.877521
QUANTITYORDERED	2747.0	35.103021	6.0	27.0	35.0	43.0	97.0	9.762135
PRICEEACH	2747.0	101.098951	26.88	68.745	95.55	127.1	252.87	42.042548
ORDERLINENUMBER	2747.0	6.491081	1.0	3.0	6.0	9.0	18.0	4.230544
SALES	2747.0	3553.047583	482.13	2204.35	3184.8	4503.095	14082.8	1838.953901
ORDERDATE	2747	2019-05-13 21:56:17.211503360	2018-01-06 00:00:00	2018-11-08 00:00:00	2019-06-24 00:00:00	2019-11-17 00:00:00	2020-05-31 00:00:00	NaN
DAYS_SINCE_LASTORDER	2747.0	1757.085912	42.0	1077.0	1761.0	2436.5	3562.0	819.280576
MSRP	2747.0	100.691664	33.0	68.0	99.0	124.0	214.0	40.114802

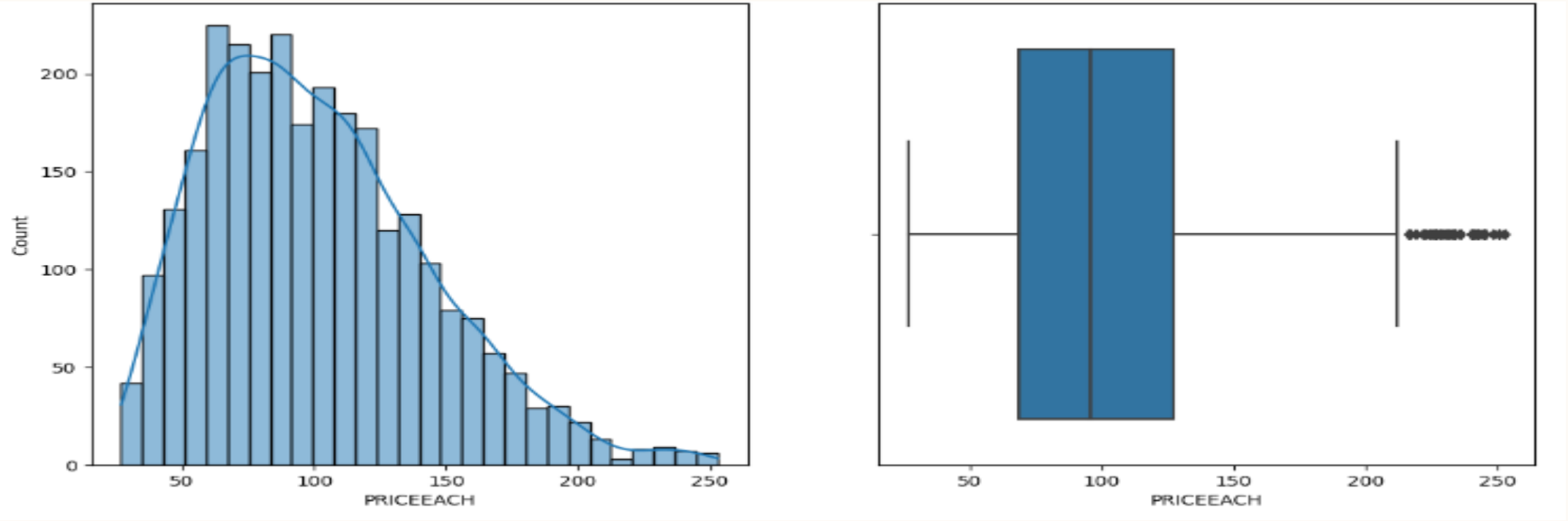
Order Quantity

After Exploration of the data it is found that there are Outliers present in this variable and data not perfect normally distributed.



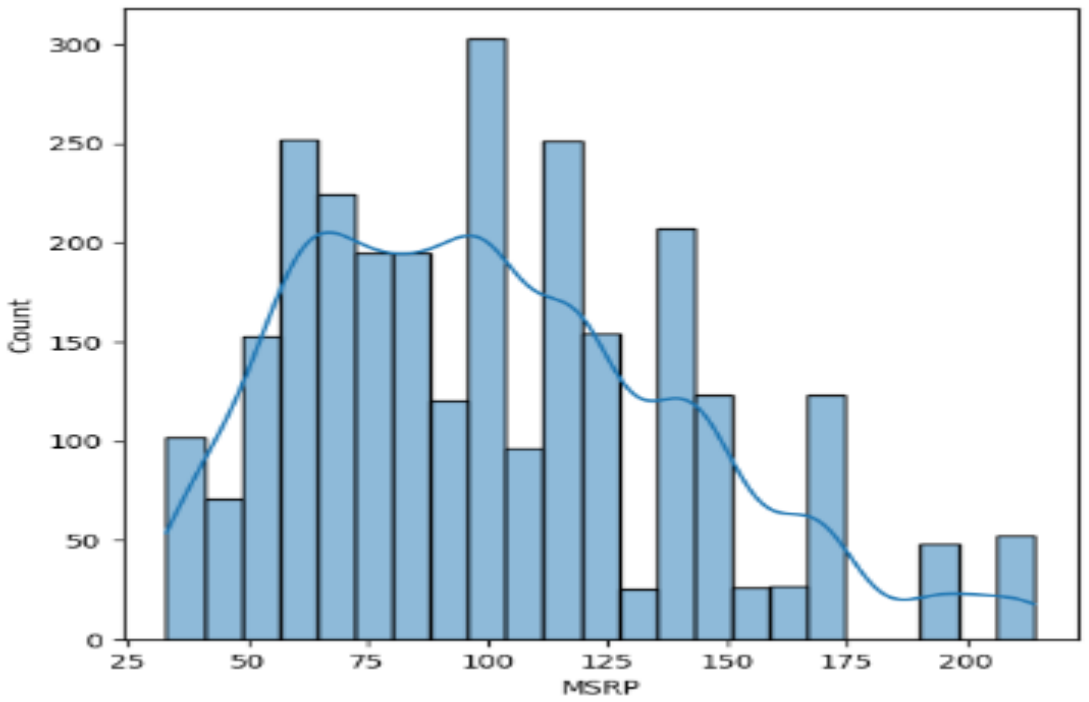
Price of Each Item

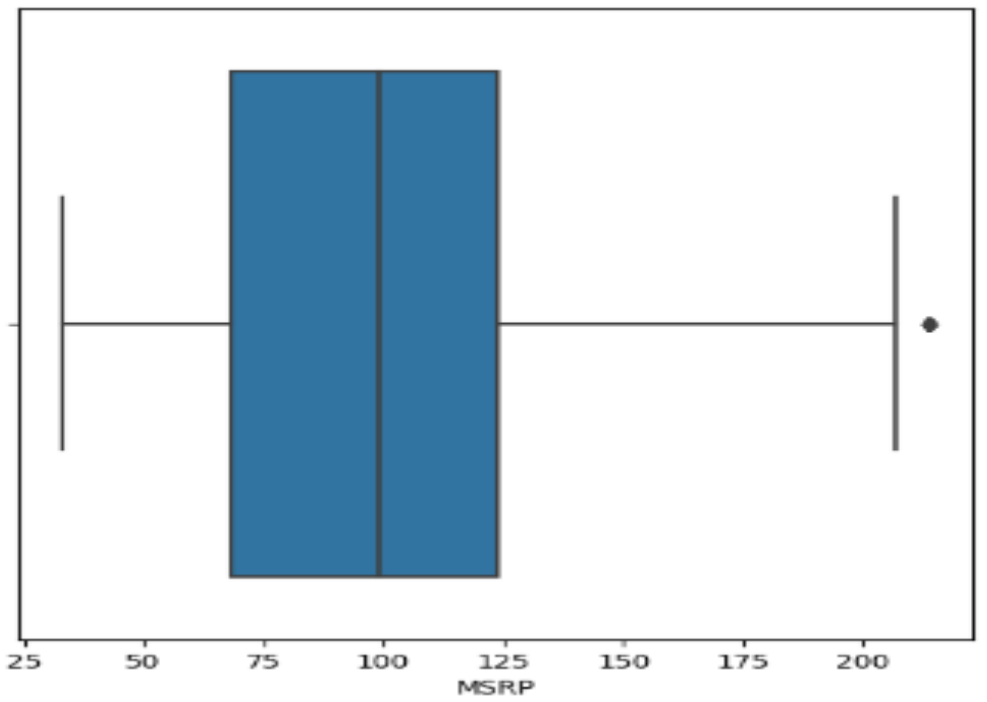
This variable is having approximate normal distribution (slight right skewed) but having many outliers.



Manufacturer's Suggested Retail Price

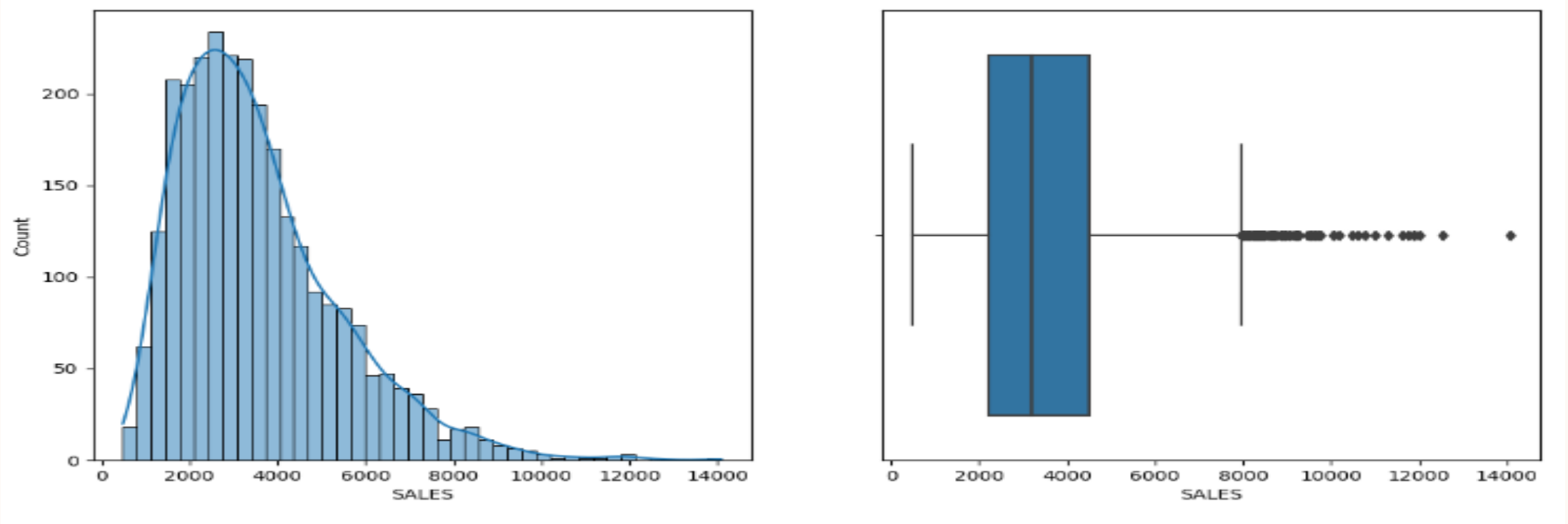
This variable is having very few outliers





Sales Amount

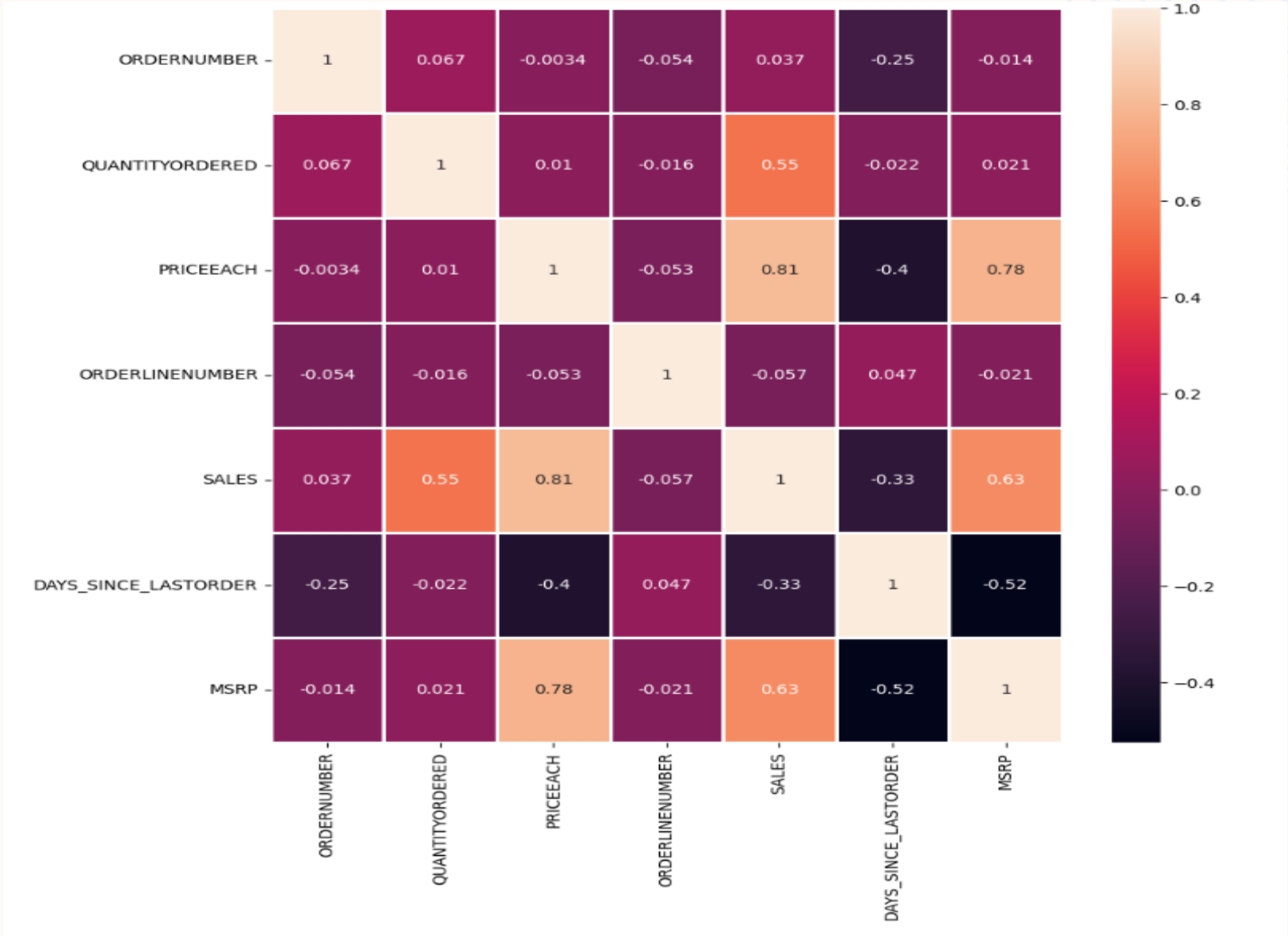
This variable is right skewed with a lot of outliers.



Data Analysis

Correlation

Highly correlated variables are
SALES and PRICEEACH,
MSRP and PRICEEACH,
SALES and QUANTITYORDERED,
SALES and MSRP



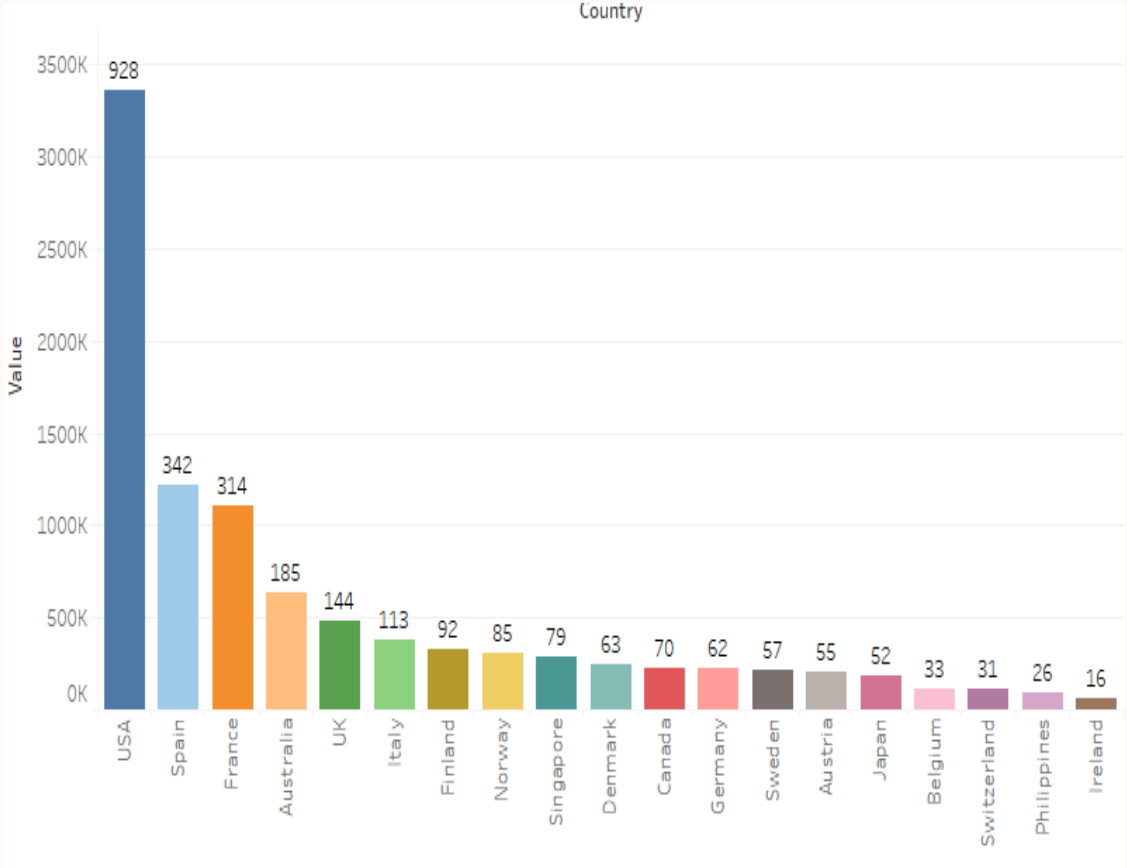
EXPLORATORY DATA ANALYSIS

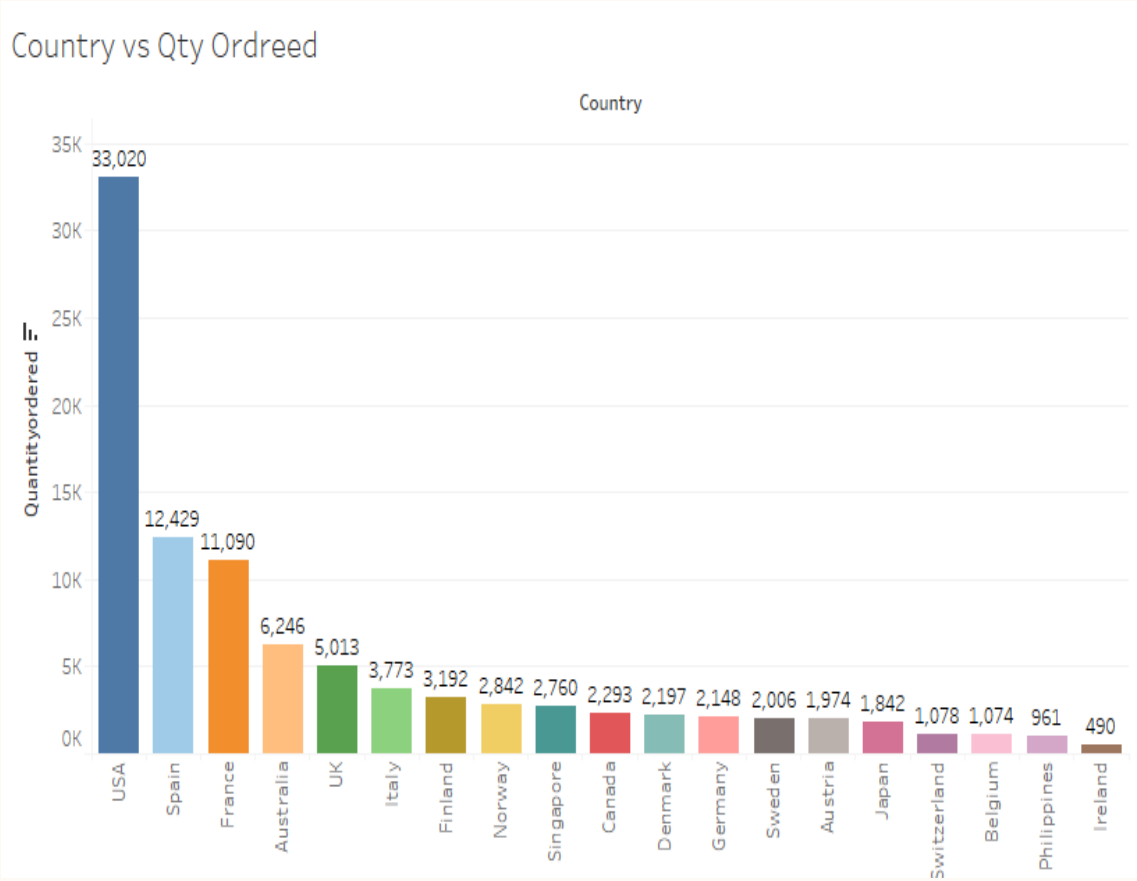
1. UNIVARIATE ANALYSIS
2. BIVARIATE ANALYSIS
3. MULTIVARIATE ANALYSIS

TABLEAU LINK: RFM_Harikrishnan | Tableau Public

Sales across countries

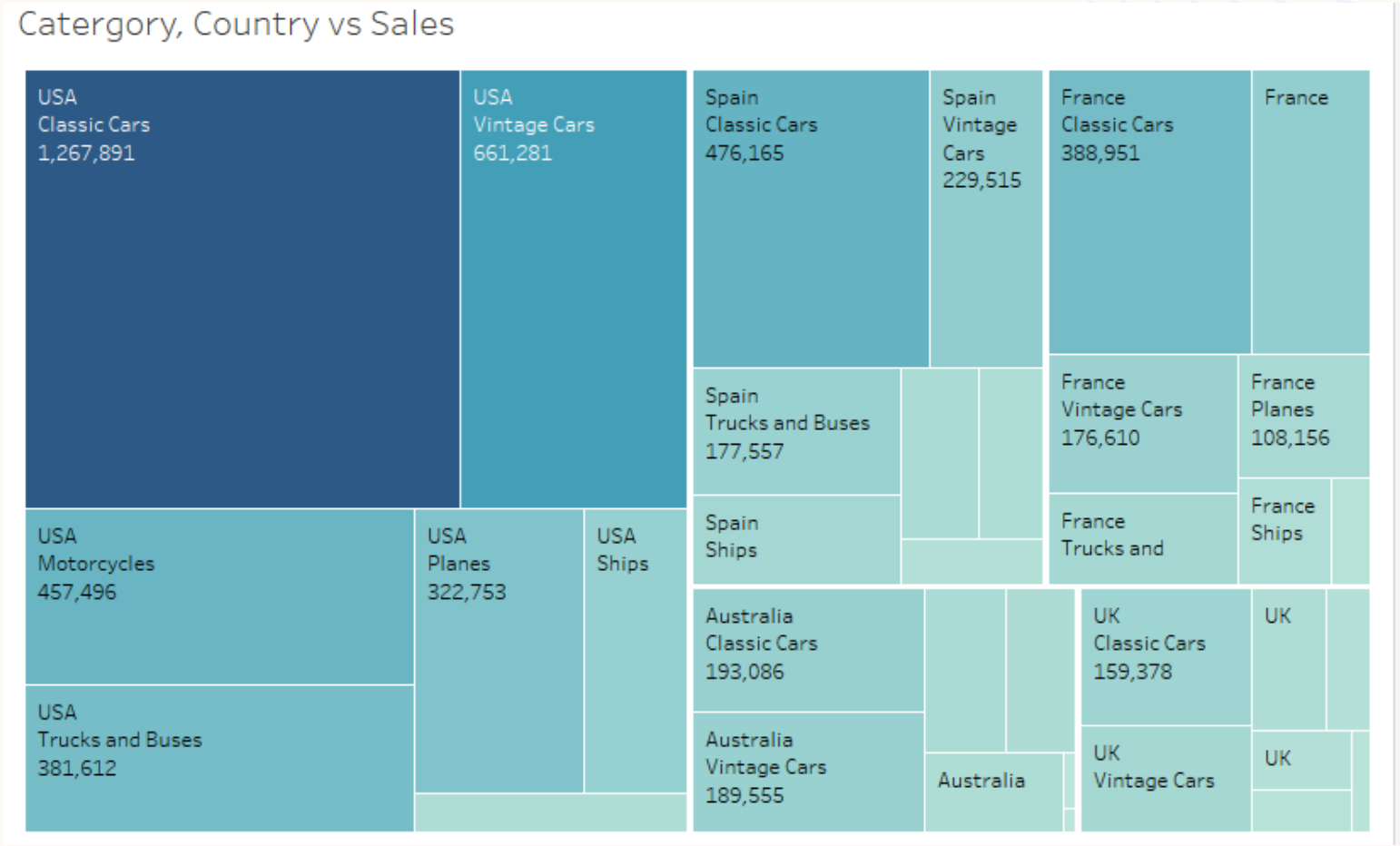
USA is top most seller in the among countries.





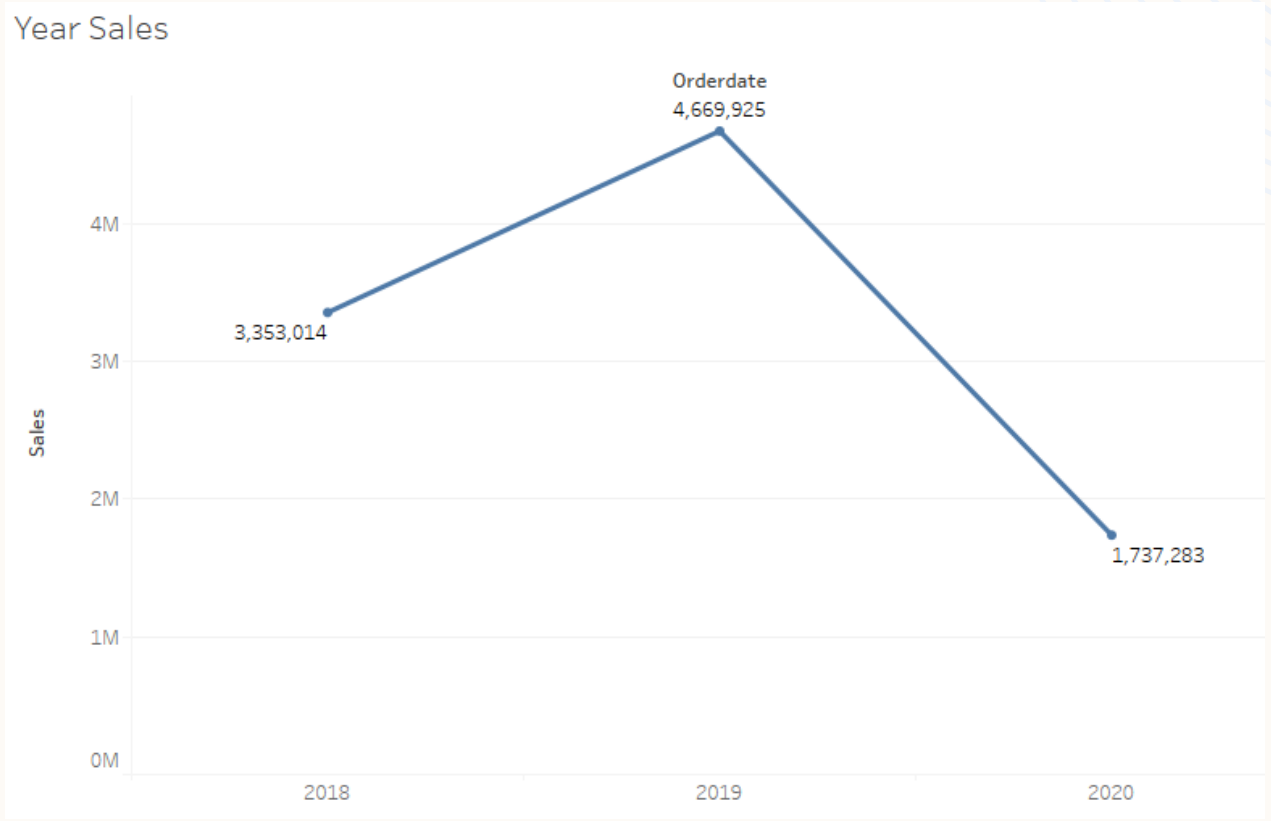
Category, Country Vs Sales

Classic cars have the highest sales, while trains have the lowest.



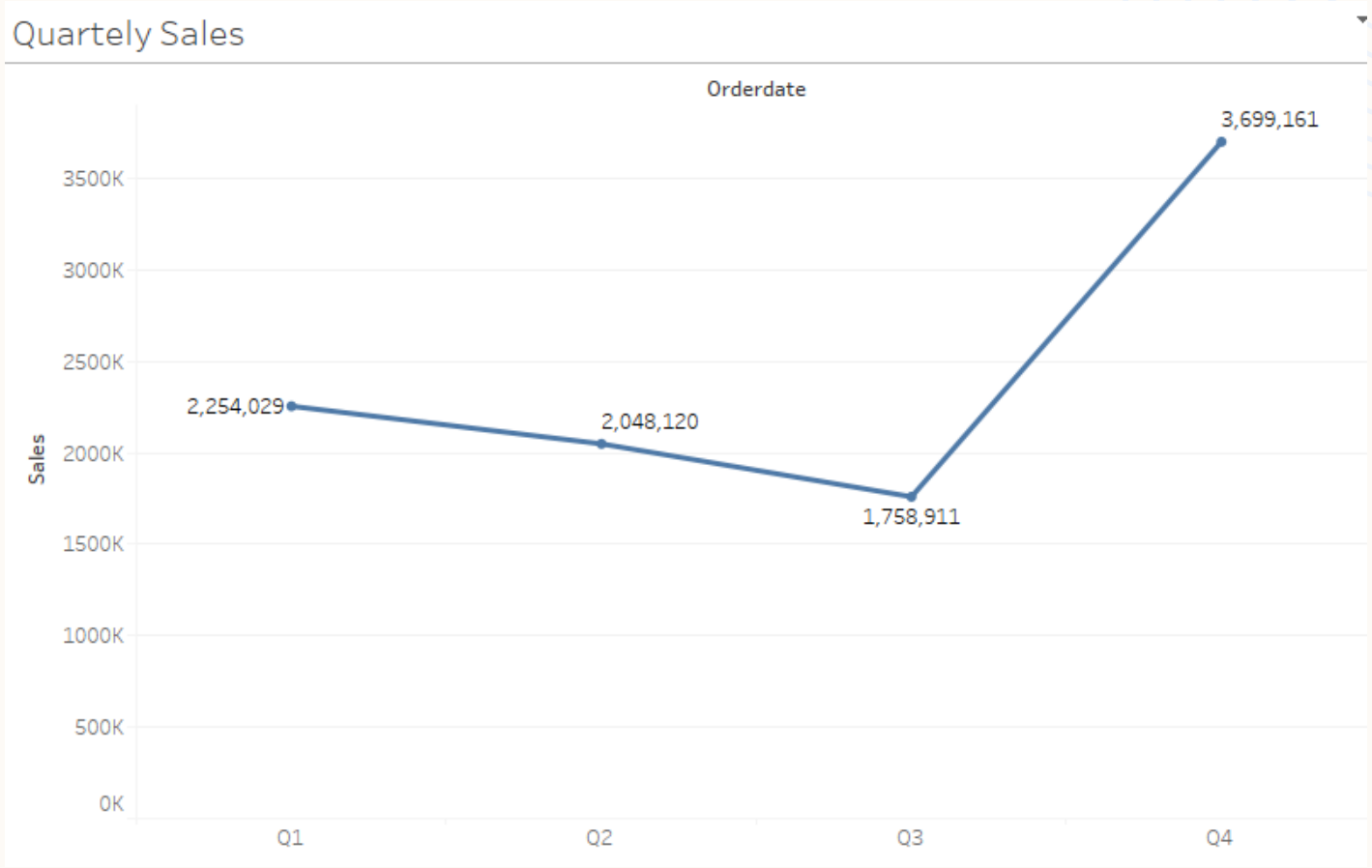
Yearly Sales

2019 appears to be the best year for sales.
The year 2020 has the lowest sales.



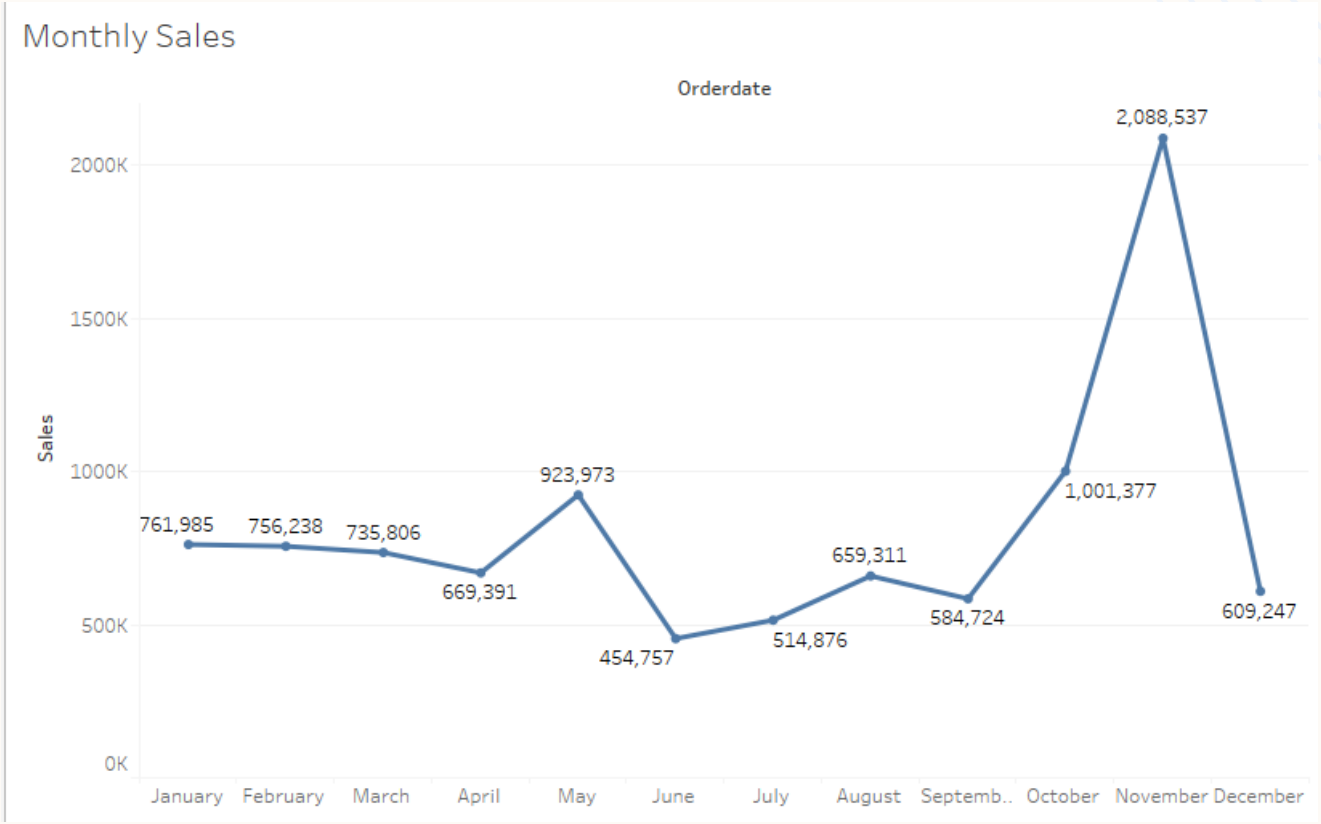
Quarterly Sales

Sales decreased from Q1 to Q3, but increased dramatically in Q4.



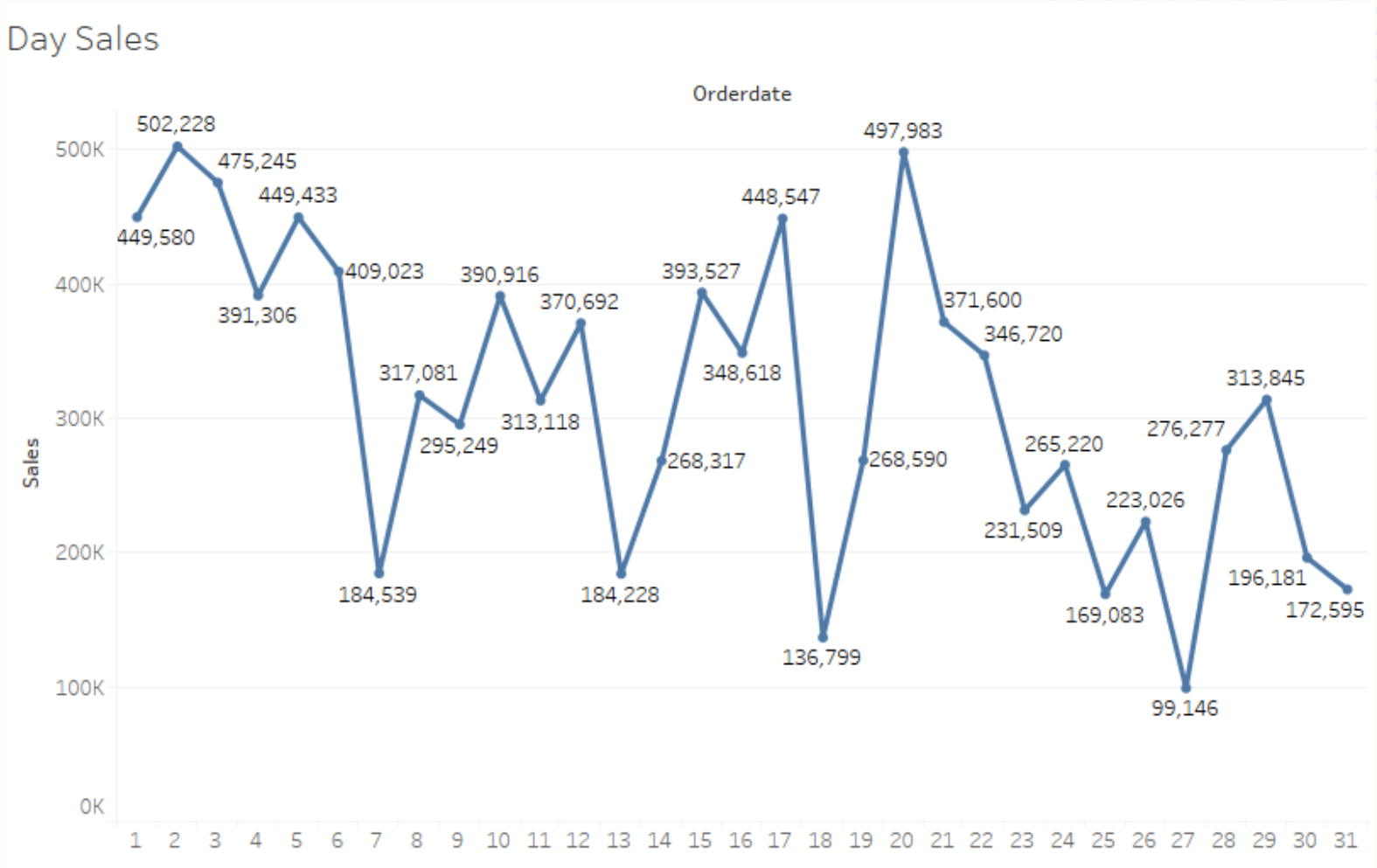
Monthly Sales

The month of November had the biggest sales.
June is a month with poor sales.



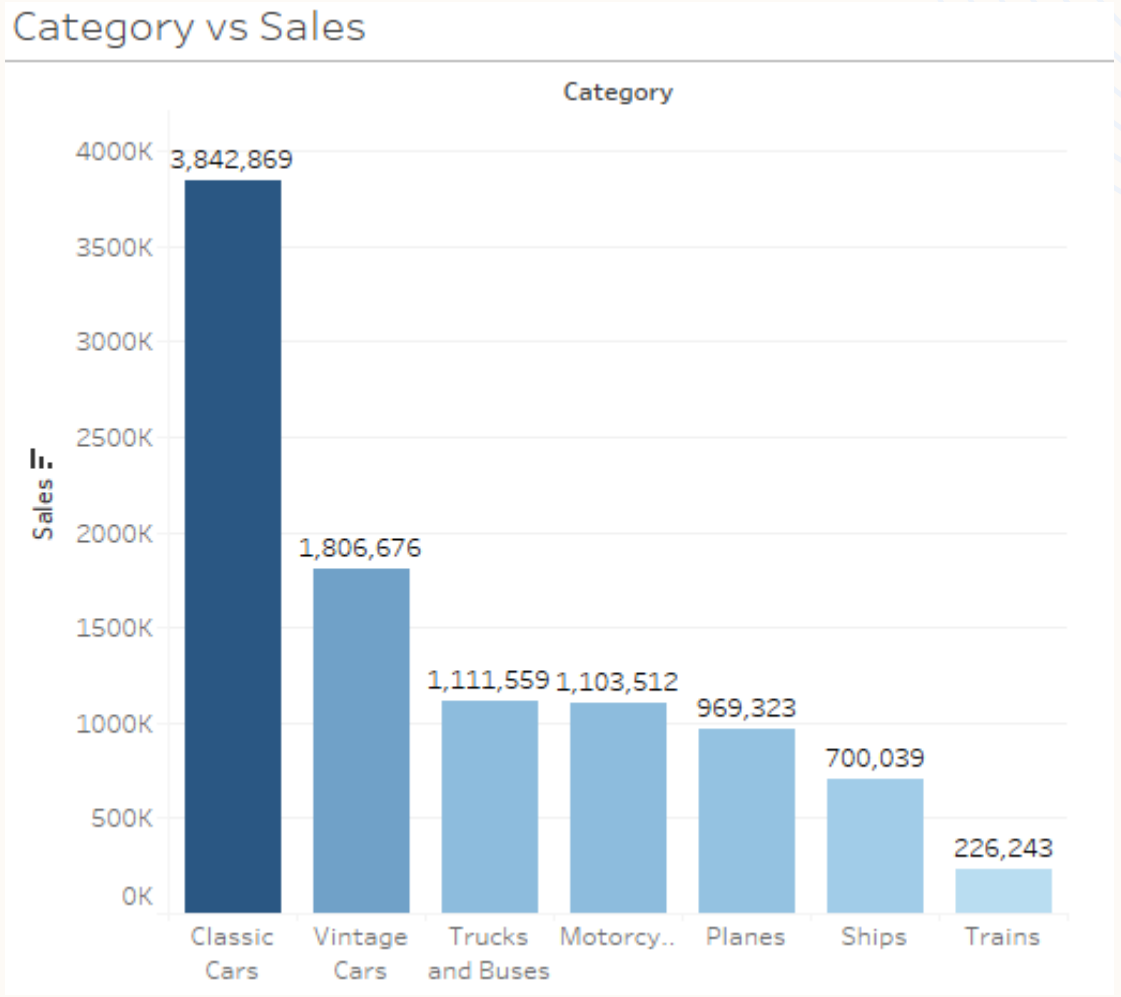
Day Sales

There is a pattern in day sales. People want to order in the first few days of the month. People ordered less at the end of the month.



Bivariate Analysis – Category vs Sales

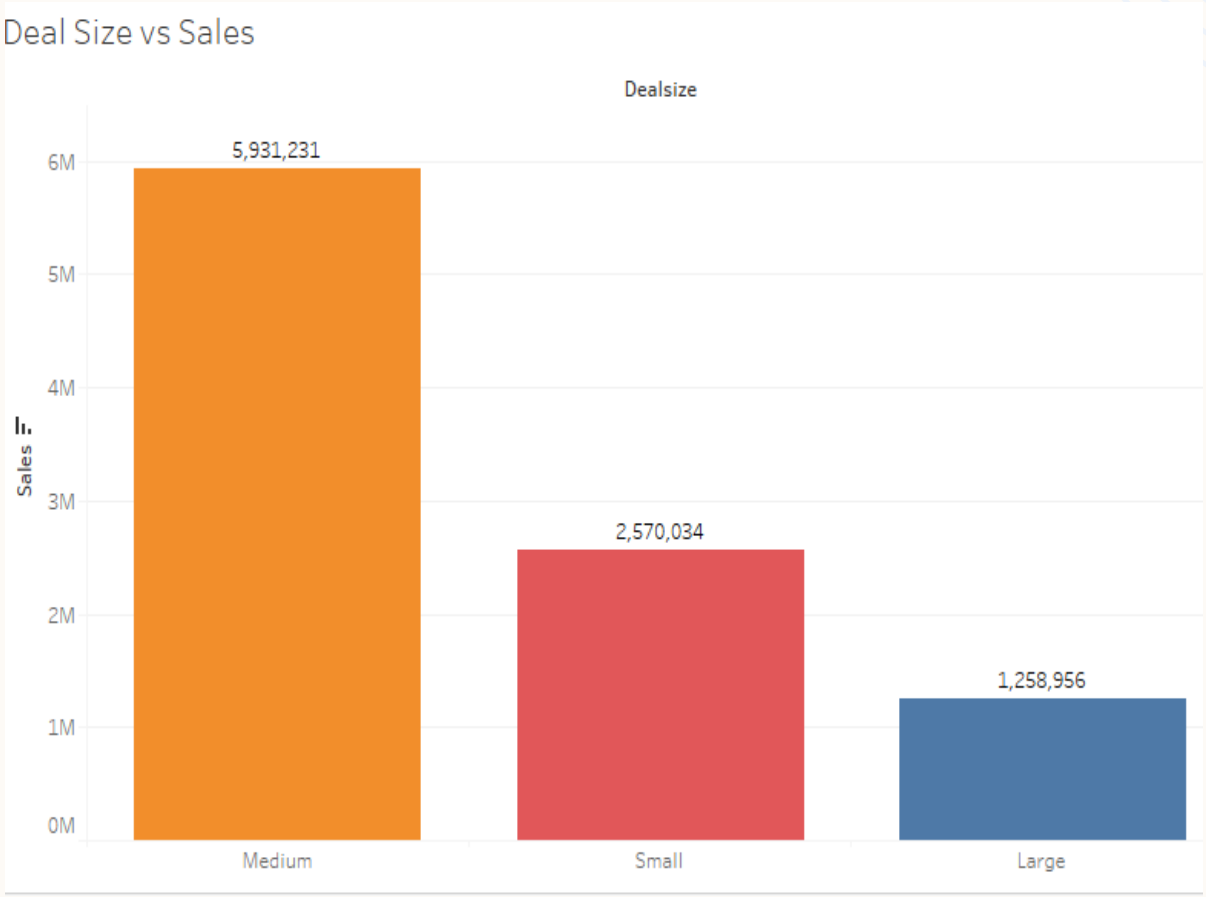
Classic cars have the highest sales, while trains have the lowest.



Bivariate Analysis – Deal size vs Sales

1. When compared to small and large deals, medium-sized deals produce much larger sales.

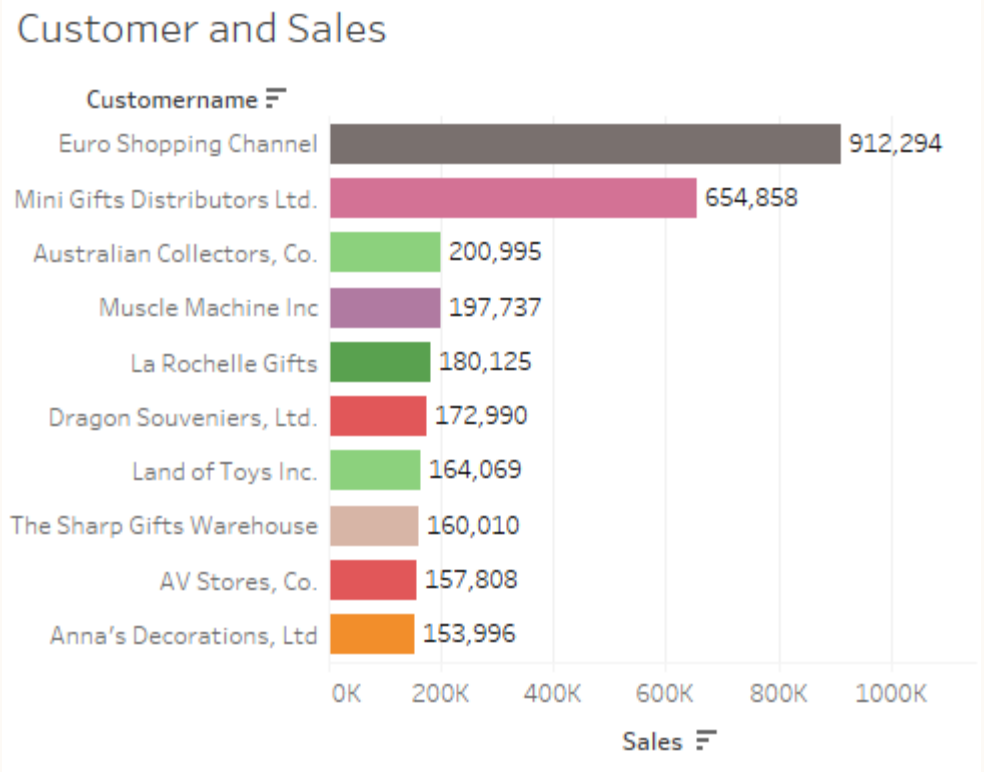
2. Large size deal sales have almost remained flat throughout the years, and it may be assumed that the corporation should focus on acquiring large size chunk projects.



Bivariate Analysis – Customer vs Sales

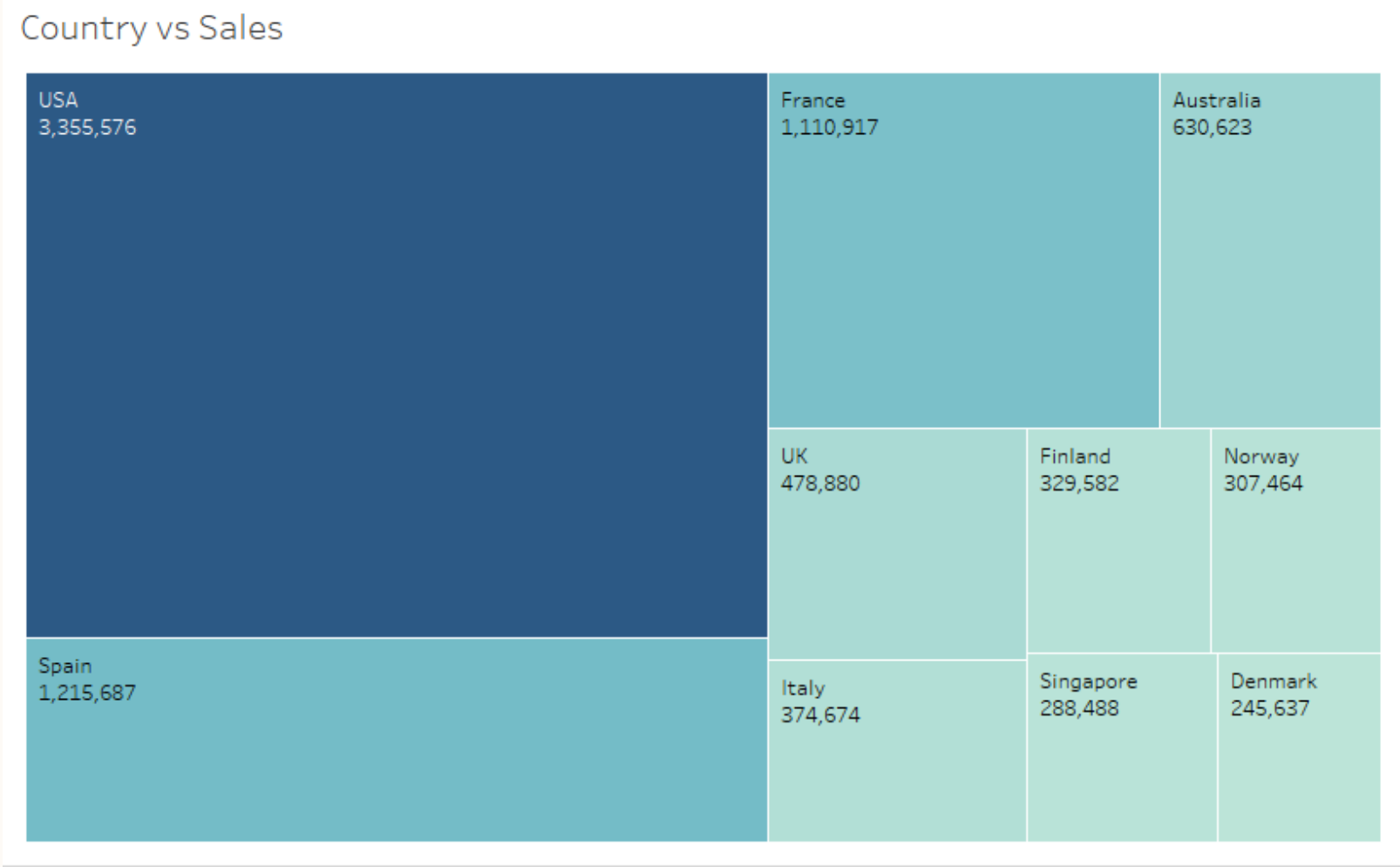
1. In terms of sales volume, the Euro shopping channel appears to be the best customer.

2. The company is a customer driver because 4-5 customers account for the majority of their sales. As a result, the corporation should spend more on customer scouting in a strategic manner, because client attrition will have a significant impact on the company's sales.



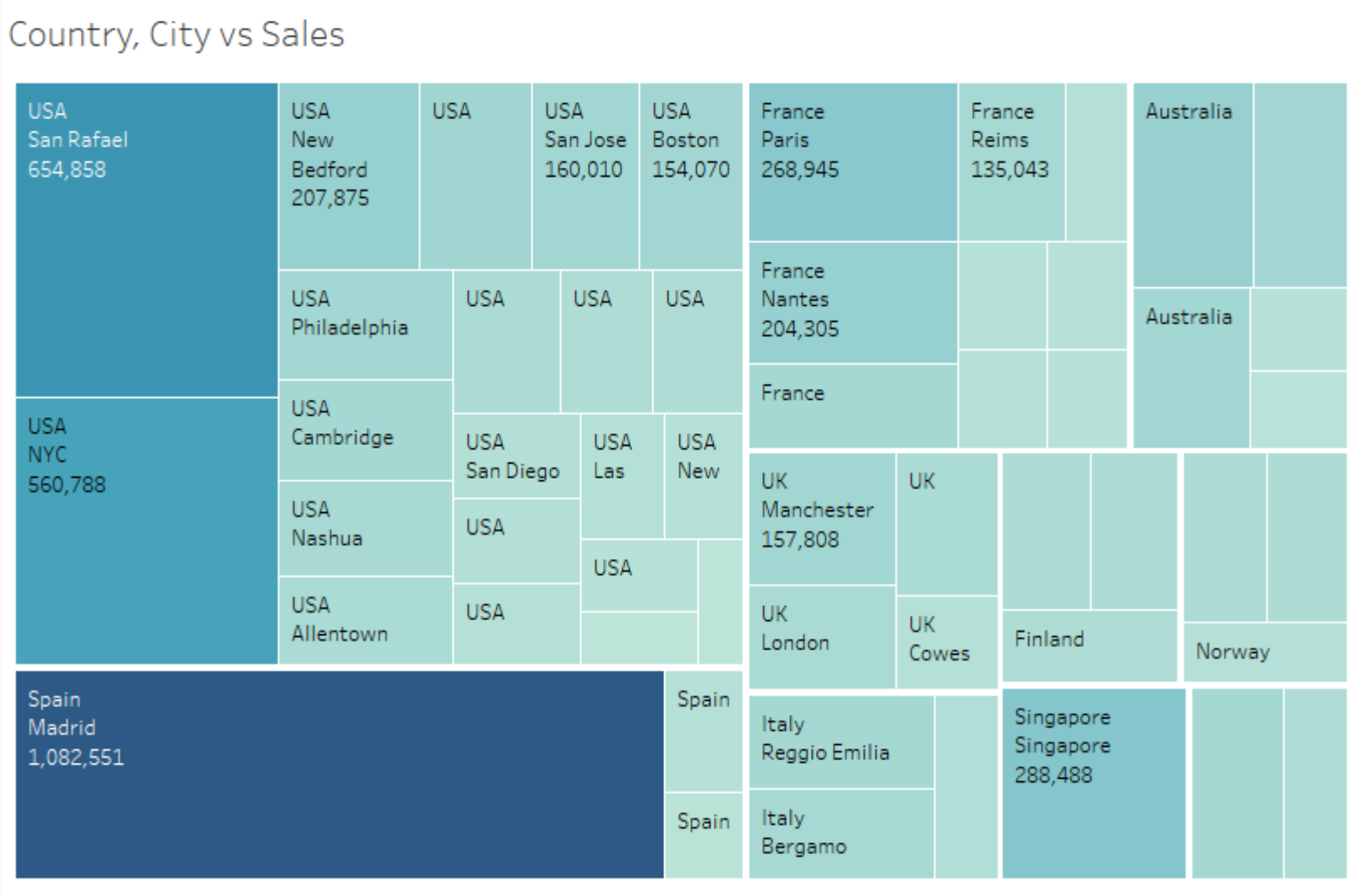
Bivariate Analysis – Country vs Sales

USA has the highest sales as compared to other countries



Bivariate Analysis – Country, City vs Sales

City Madrid in country Spain is delivering the highest sales as compared to other countries



SUMMARY ADDON

- Classic cars are the most popular and generate the most income.
- Planes, trucks, and ships have the longest order lines.
- The top two selling countries are the United States and Spain.
- San Rafael and New York City have the highest sales in the United States, while Madrid is the largest market in Spain
- Euro Shopping Channel and Mini Gifts Distributor are two of the most important consumers.
- Classic cars, as well as Euro shopping and Mini presents, are at the top of the list.
- Annual sales are down, but when broken down, they show a little upward tendency.

RFM

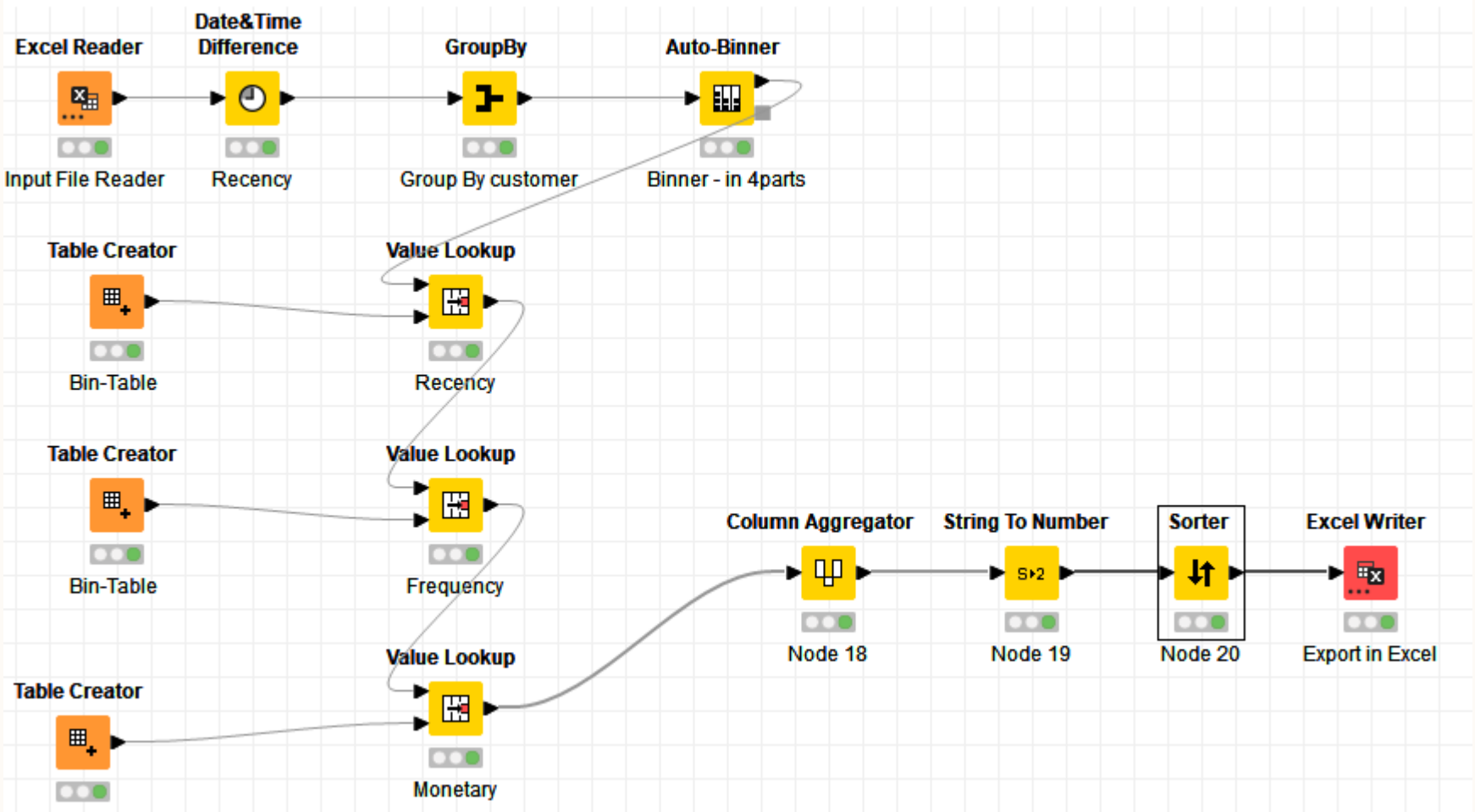
- RFM, also known as RFM analysis, is a type of customer segmentation and behavioral targeting used to help businesses rank and segment customers based on the recency, frequency, and monetary value of a transaction.
- RFM marketing can help marketers and small business owners determine their target audience to use their budget most effectively
- This method gives customers scores based on 3 factors: recency, frequency, and monetary
- Recency refers to how recent a customer's last purchase was. Customers who have made a recent purchase, typically within the last few weeks, still have the product and brand on their minds and are most likely to make a repeat purchase.
- Frequency is how often the customer makes purchases, which can help you identify repeat customers. For example, many clients make frequent repeat purchases within a set timeframe
- Monetary value refers to how much a customer spends within a given period. It's always important to consider because it can tell you a few things about consumer behavior

CUSTOMER SEGMENTATION USING RFM ANALYSIS

- KNIME tool is used for RFM analysis and customer segmentation
- Customer name, Quantity ordered, Order date, order quantity and Sales are the used parameters.
- Predictions
  1. Recency calculated using Max(ORDERDATE)-ORDERDATE
  2. Frequency calculated using count of order for a particular customer.
  3. Monetary is calculated using sum of sales through particular customer.

TABLEAU LINK: RFM_Harikrishnan | Tableau Public [https://public.tableau.com/app/profile/harikrishnan.m4921/viz/RFM_Harikrishnan/Story1?publish=yes]

KNIME Workflow



RFM Analysis

Output Table Head

Row ID	CUSTOMERNAME	Count*...	Sum(Q...	Mean(P...	Sum(SA...	Count(...	Count(...	Count(...	Min*(R...	Count(...	Recency	Freque...	Monetary	RFM Sc...
Row32	Euro Shopping Channel	259	9327	97.383	912,294.11	259	259	259	0	259	4	4	4	444
Row53	Mini Gifts Distributors...	180	6366	102.696	654,858.06	180	180	180	2	180	4	4	4	444
Row43	La Rochelle Gifts	53	1832	97.046	180,124.9	53	53	53	0	53	4	4	4	444
Row78	The Sharp Gifts Ware...	40	1656	93.376	160,010.27	40	40	40	39	40	4	4	4	444
Row72	Souveniers And Thin...	46	1601	95.189	151,570.98	46	46	46	2	46	4	4	4	444
Row67	Salzburg Collectables	40	1442	101.398	149,798.63	40	40	40	14	40	4	4	4	444
Row26	Danish Wholesale Im...	36	1315	108.038	145,041.6	36	36	36	46	36	4	4	4	444
Row41	L'ordine Souveniers	39	1280	111.147	142,601.33	39	39	39	21	39	4	4	4	444
Row63	Reims Collectables	41	1433	94.343	135,042.94	41	41	41	62	41	4	4	4	444
Row37	Handji Gifts& Co	36	1236	95.593	115,498.73	36	36	36	38	36	4	4	3	443
Row27	Diecast Classics Inc.	31	1111	108.566	122,138.14	31	31	31	1	31	4	3	4	434
Row79	Tokyo Collectables, Ltd	32	1150	101.183	120,562.74	32	32	32	39	32	4	3	3	433
Row83	UK Collectables, Ltd.	29	1046	108.536	118,008.27	29	29	29	53	29	4	3	3	433
Row9	Auto Canal Petit	27	1001	94.255	93,170.66	27	27	27	54	27	4	3	3	433
Row34	Gift Depot Inc.	25	903	108.932	101,894.79	25	25	25	26	25	4	2	3	423
Row77	Tekni Collectables Inc.	21	906	93.571	83,228.19	21	21	21	58	21	4	2	2	422
Row36	Gifts4AllAges.com	26	933	91.564	83,209.88	26	26	26	25	26	4	2	2	422
Row61	Petit Auto	25	796	93.494	74,972.52	25	25	25	1	25	4	2	2	422
Row62	Quebec Home Shoppi...	22	717	104.515	74,204.79	22	22	22	30	22	4	2	2	422

INFERENCE FROM RFM ANALYSIS

BEST CUSTOMERS	CUSTOMERS ON THE VERGE OF CHURNING	LOST CUSTOMERS	LOYAL CUSTOMERS

BEST CUSTOMERS

EURO SHOPPING CHANNEL	MINI GIFTS DISTRIBUTORS	LA ROCHELLE GIFTS	THE SHARP GIFTS WAREHOUSE	SOUVENIERS AND THINGS CO.
Total Sales 912294 with RFM 444	Total Sales 654858 with RFM 444	Total Sales 180124 with RFM 444	Total Sales 160010 with RFM 444	Total Sales 151570 with RFM 444

CUSTOMERS ON THE VERGE OF CHURNING

SAVELEY & HENRIOT, CO.	LAND OF TOYS INC.	AV STORES, CO	ROVELLI GIFTS	ONLINE DIECAST CREATIONS CO.
Total Sales 142874 with RFM 144	Total Sales 164069 with RFM 244	Total Sales 157807 with RFM 244	Total Sales 137955 with RFM 244	Total Sales 131685 with RFM 244

LOST CUSTOMERS

DAEDALUS DESIGNS IMPORTS	OSAKA SOUVENIERS CO.	AUTO ASSOC. & CIE.	CLOVER COLLECTIONS, CO.	ONLINE MINI COLLECTABLES
Total Sales 69052 with RFM 111	Total Sales 67605 with RFM 111	Total Sales 64834 with RFM 111	Total Sales 57756 with RFM 111	Total Sales 57197 with RFM 111

LOYAL CUSTOMERS

EURO SHOPPING CHANNEL	MINI GIFTS DISTRIBUTORS	HANDJI GIFTS& CO	DIECAST CLASSICS INC	TOKYO COLLECTABLES.
Total Sales 912294 with RFM 444	Total Sales 654858 with RFM 444	Total Sales 115498 with RFM 443	Total Sales 122138 with RFM 434	Total Sales 120562 with RFM 433

RFM Analysis

SUMMARY

- We classified our top, loyal, on the point of churning, and lost customers using recency, frequency, and monetary parameters. Customers with a recent purchase have been our top customers, as we have lost customer lists.
- Customers on the edge of leaving can be salvaged and transformed into good buyers.
- The RFM model is used to determine customer kinds such as loyal, top or best, on the verge of churning, and lost consumers.
- Recency, frequency, and monetary criteria were commonly utilized to categorize clients.
- This methodology can assist a firm maintain its sales and customers by focusing on how the company lost customers and taking various actions to attract them back.
- It is critical for the company to convert or keep clients who are on the point of leaving into regular customers.
- It is also possible to identify ways to boost the sales ratio.

THANK YOU

Harikrishnan M