

Automobile Part Manufacturer Company's Data MRA Project - Milestone 1

By: Pratheek U



Contents:

- Agenda & Executive Summary of the data
 - Problem statement
 - Executive Summary
 - Data Dictionary
 - Summary Stats
 - Assumptions about data
- Exploratory Analysis & Insights
 - Univariate, Bivariate, and multivariate analysis using data visualization
 - Weekly, Monthly, Quarterly, Yearly Trends in Sales
 - Sales Across different Categories of different features in the given data
 - Summarize the inferences from the above analysis
- Customer Segmentation using RFM analysis
 - What is RFM and which tool used
 - What all parameters used and assumptions made
 - Output table head
 - Workflow image to be put when KNIME used
- Inferences from RFM Analysis and identified segments
 - Who are your best customers?
 - Which customers are on the verge of churning?
 - Who are your lost customers?
 - Who are your loyal customers?



Agenda :

01

Agenda &
Executive Summary
of the data

02

Exploratory analysis
and insights

03

RFM analysis for
customer
segmentation

04

Identification of
customers based
on different
parameters

05

Recommendations

01 Agenda & Executive Summary of the data

- Problem statement
- About Data (Info, Shape, Summary Stats, your assumptions about data)

Problem Statement:

An automobile parts manufacturing company has collected data of 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 magical data science skills to provide them with suitable insights about their data and their customers.

Executive Summary:

- Data: past 3 years.
- Objective: identify the underlying buying patterns of the customers and recommend customized marketing strategies for different segments of customers.
- Dataset: 20 columns and 2747 rows,
- Missing values and Duplicate values: None
- Outliers: some columns has few outliers
- The exploratory analysis and insights provide a clear understanding of the data and highlight the key trends and patterns in sales.
- RFM analysis has been performed to segment the customers into four categories based on their buying behavior, and customized marketing strategies have been recommended for each segment.
- The presentation concludes with recommendations for the company to enhance its customer relationships and drive business growth.



Data Dictionary

- ORDERNUMBER : Order Number
- CUSTOMERNAME : customer
- QUANTITYORDERED : Quantity ordered
- PHONE : Phone of the customer
- PRICEEACH : Price of Each item
- ADDRESSLINE1 : Address of customer
- ORDERLINENUMBER : order line
- CITY : City of customer
- SALES : Sales amount
- POSTALCODE : Postal Code of customer

- ORDERDATE : Order Date
- COUNTRY : Country customer
- DAYS_SINCE_LASTORDER : Days_ Since_Lastorder
- CONTACTLASTNAME : Contact person customer
- STATUS : Status of order like Shipped or not
- CONTACTFIRSTNAME : Contact person customer
- PRODUCTLINE : Product line – CATEGORY
- DEALSIZE : Size of the deal based on Quantity and Item Price
- MSRP : Manufacturer's Suggested Retail Price
- PRODUCTCODE : Code of Product

Numeric columns: ORDERNUMBER, QUANTITYORDERED, PRICEEACH, ORDERLINENUMBER SALES, DAYS_SINCE_LASTORDER, MSRP.

Assumptions:

Each row in the data represents a unique transaction made by a customer.

The customer segments may be defined based on the purchasing frequency, amount spent, and recency of purchases (RFM Analysis).

The marketing strategies may vary for each customer segment, and the company may need to personalize their marketing efforts accordingly.

The order date and days since last order columns are accurately calculated.

The sales column is calculated as the product of quantity ordered and price each.

The status column indicates the current status of the order accurately.

The recommendations provided in the presentation are based on the insights gained from the analysis of the transaction data.

Statistical Summary of Numerical Columns

| | ORDERNUMBER | QUANTITYORDERED | PRICEEACH | ORDERLINENUMBER | SALES | DAYS_SINCE_LASTORDER | MSRP |
|-------|--------------|-----------------|-------------|-----------------|--------------|----------------------|-------------|
| count | 2747.000000 | 2747.000000 | 2747.000000 | 2747.000000 | 2747.000000 | 2747.000000 | 2747.000000 |
| mean | 10259.761558 | 35.103021 | 101.098951 | 6.491081 | 3553.047583 | 1757.085912 | 100.691664 |
| std | 91.877521 | 9.762135 | 42.042548 | 4.230544 | 1838.953901 | 819.280576 | 40.114802 |
| min | 10100.000000 | 6.000000 | 26.880000 | 1.000000 | 482.130000 | 42.000000 | 33.000000 |
| 25% | 10181.000000 | 27.000000 | 68.745000 | 3.000000 | 2204.350000 | 1077.000000 | 68.000000 |
| 50% | 10264.000000 | 35.000000 | 95.550000 | 6.000000 | 3184.800000 | 1761.000000 | 99.000000 |
| 75% | 10334.500000 | 43.000000 | 127.100000 | 9.000000 | 4503.095000 | 2436.500000 | 124.000000 |
| max | 10425.000000 | 97.000000 | 252.870000 | 18.000000 | 14082.800000 | 3562.000000 | 214.000000 |

Inference :

- The average number of items ordered per sales order is 35, with a standard deviation of 9.76.
- The average price of each item is 101.09, with a standard deviation of 42.04.
- The average sales amount per order is 3553.05, with a standard deviation of 1838.95.
- The average time since the last order is 1757.09 days, with a standard deviation of 819.28.
- The summary statistics do not indicate any red flags or abnormalities that could potentially indicate issues with the data.

02 Exploratory Analysis & Insights

Univariate, Bivariate, and multivariate analysis using data visualization

- Weekly, Monthly, Quarterly, Yearly Trends in Sales
- Sales Across different Categories of different features in the given data
- Summarize the inferences from the above analysis

Outliers has not been treated.

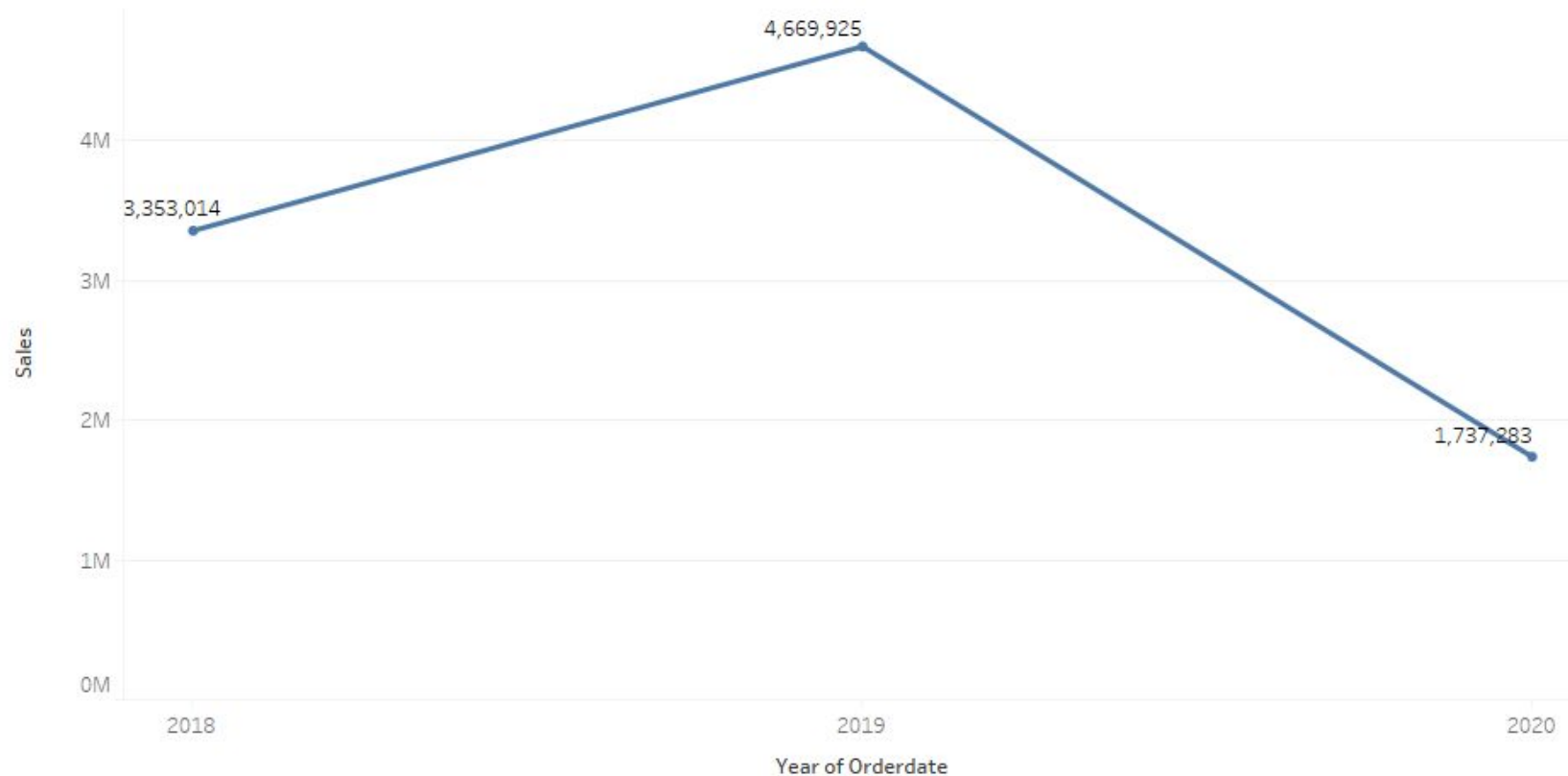


Bivariate Analysis



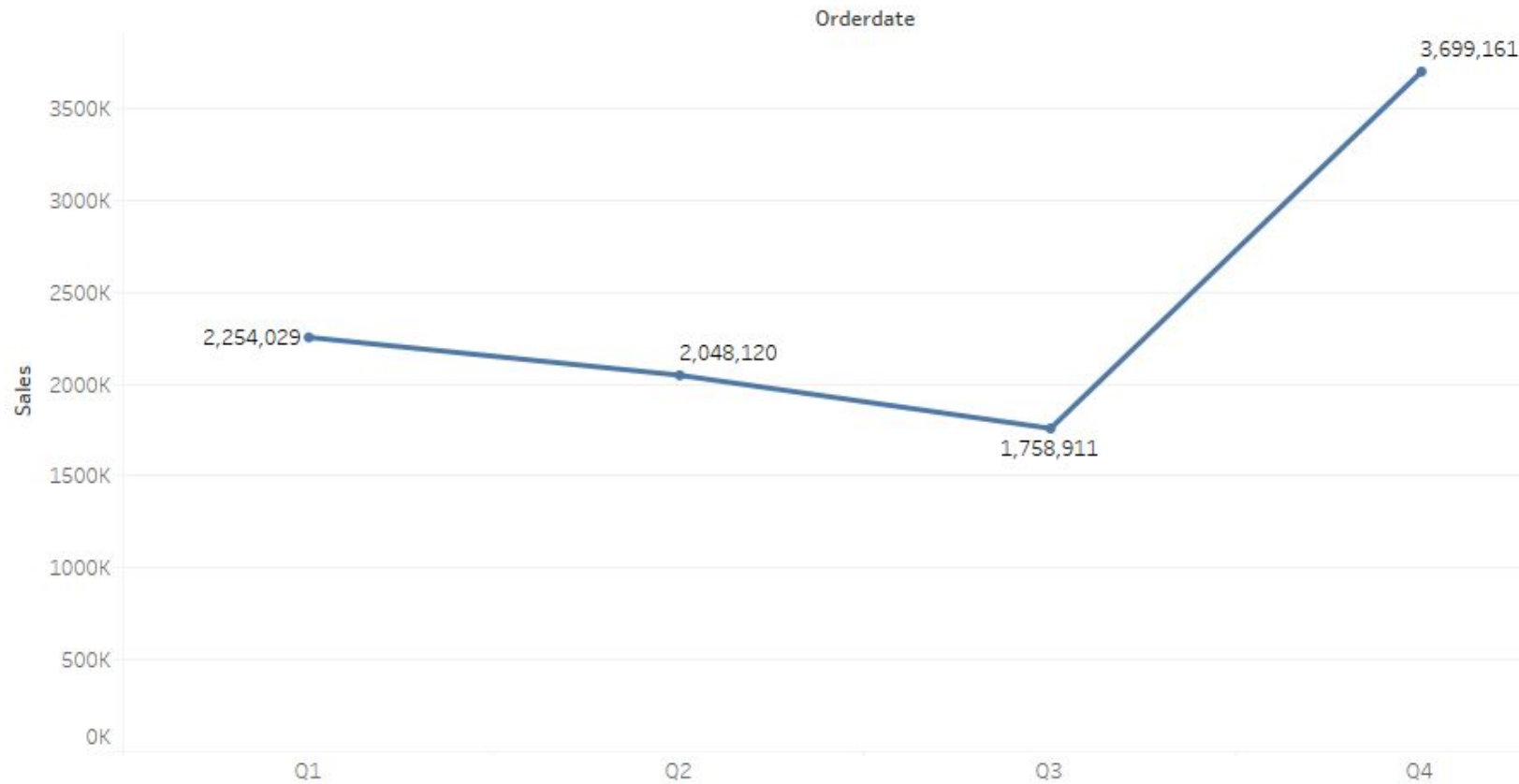
Yearly Sales

- We can see that there is a dip in the yearly sales.
- We need to find more information regarding this as its not a good sign.



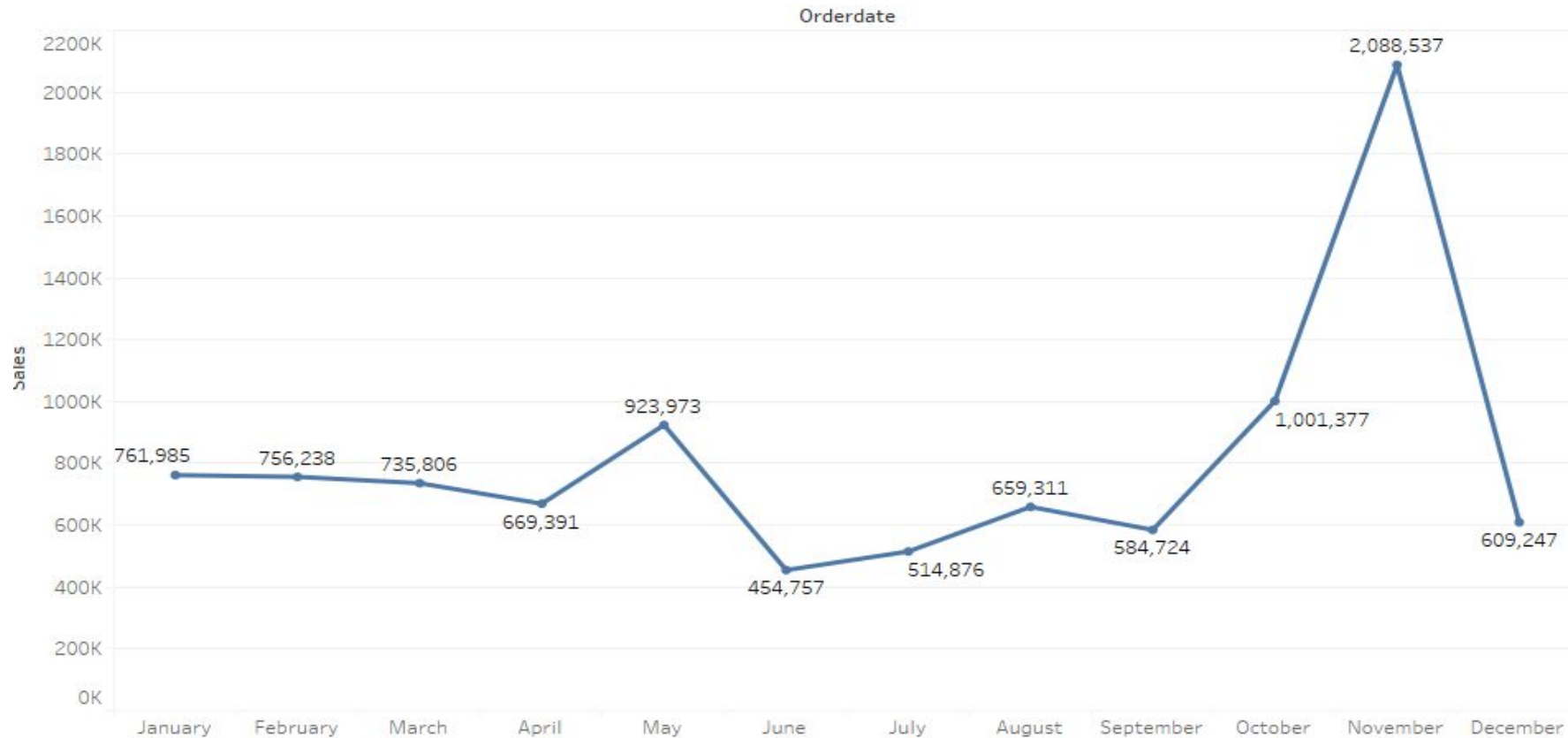
Quarterly Sales

We can see that in the quarter 4 there is high sales as compared to other quarters.



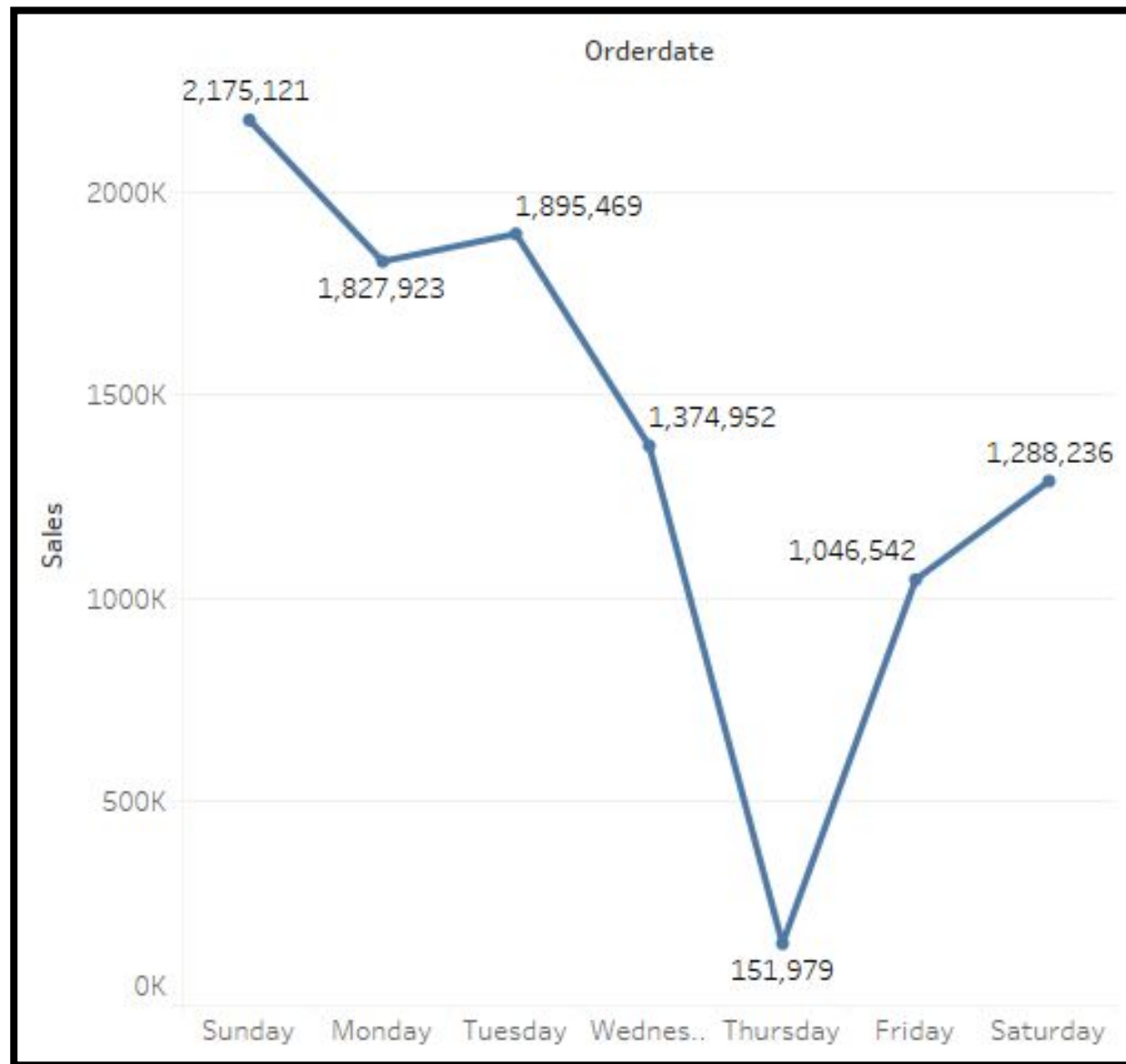
Monthly Sales

- We can see that in the 11th month there is highest sales.
- And in 6th month the sales is the lowest.
- Sales is consistent in the first 4 months of the year.



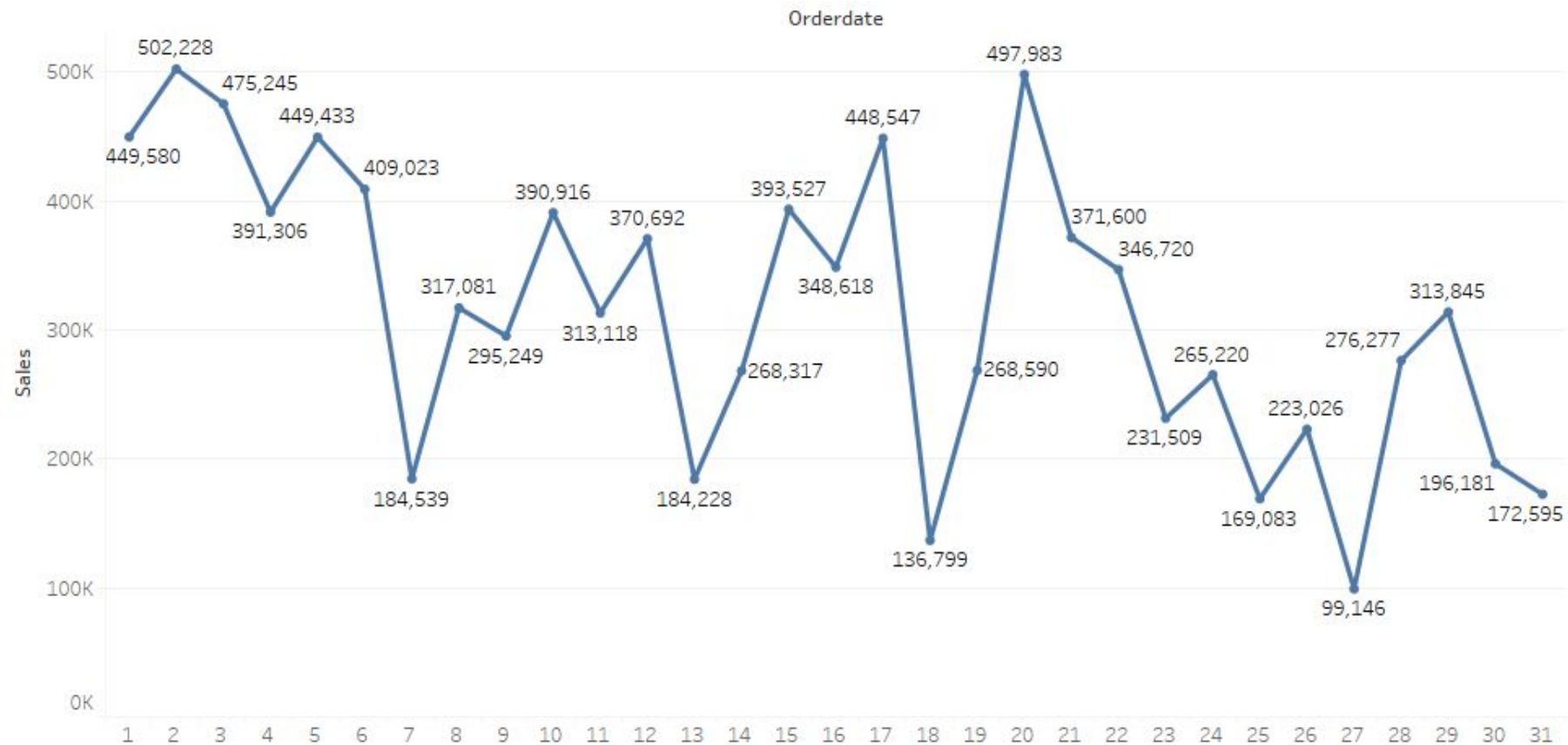
Weekday Sales

- We can see Thursday has the lowest sales and Sunday has the highest sales.
- Sales increases from Friday to Sunday and dips from Monday to Thursday



Day Sales

In the stating days of month there is more sales as compared to the end days.

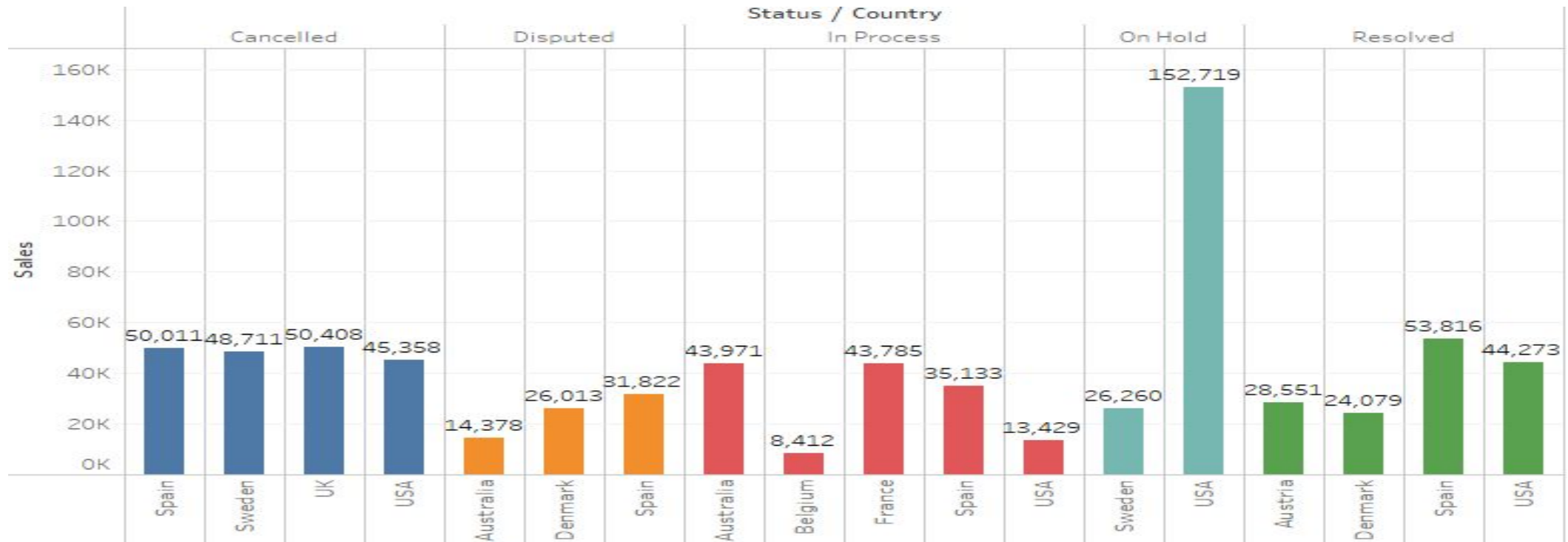




Multi-Variate Analysis

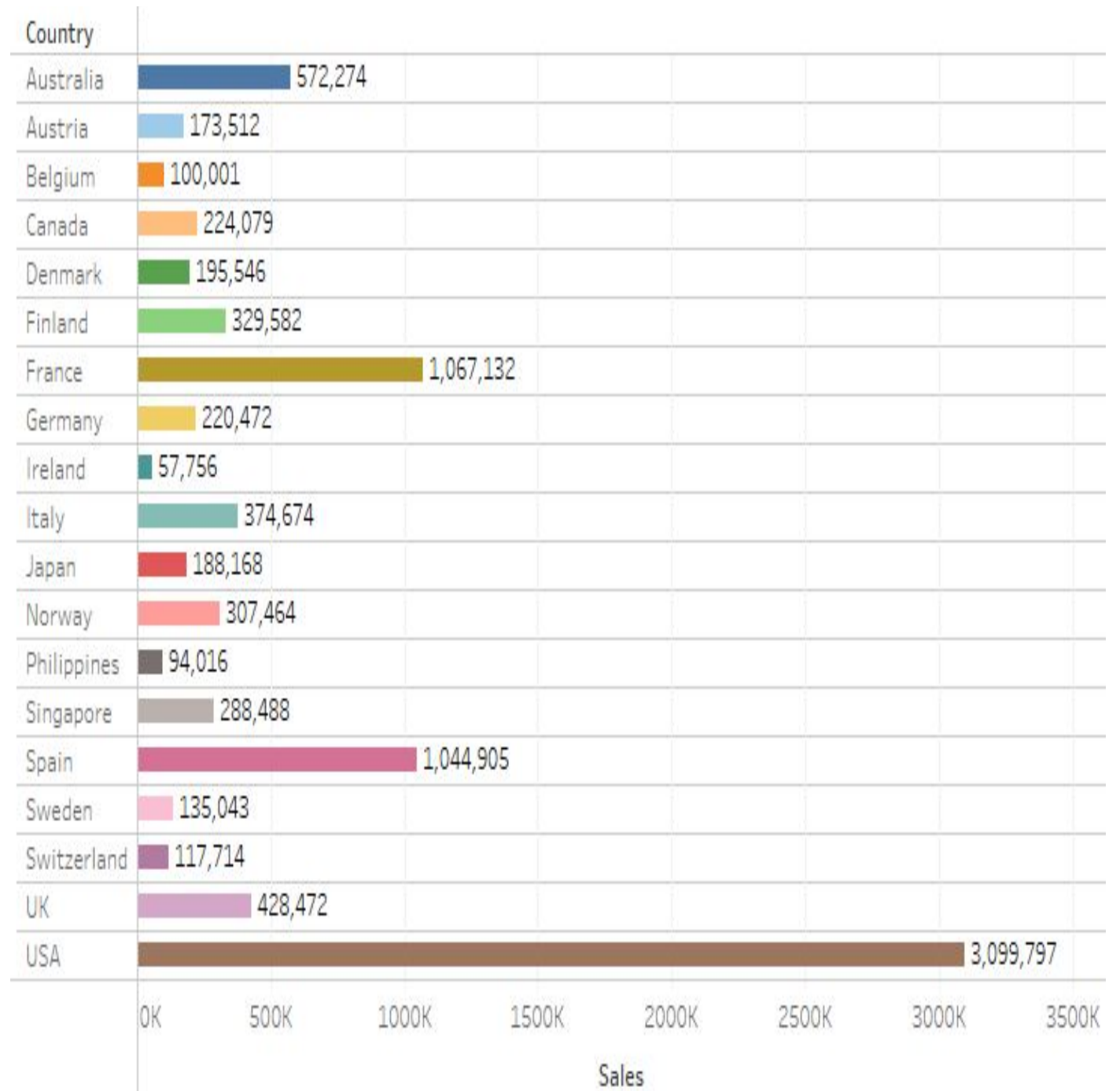
Status, Country & Sales

- We can see most of the orders which are on hold belongs to usa, some orders on hold belong to sweden.
- UK, USA, Spain, Sweden have aprox same number of canceled orders.
- Spain has most disputes and also most disputes are solved.



Order Shipped & Sales

- Usa has most shipped orders.
- Spain and France have aprox same number of orders.
- Ireland have the least number of orders shipped.

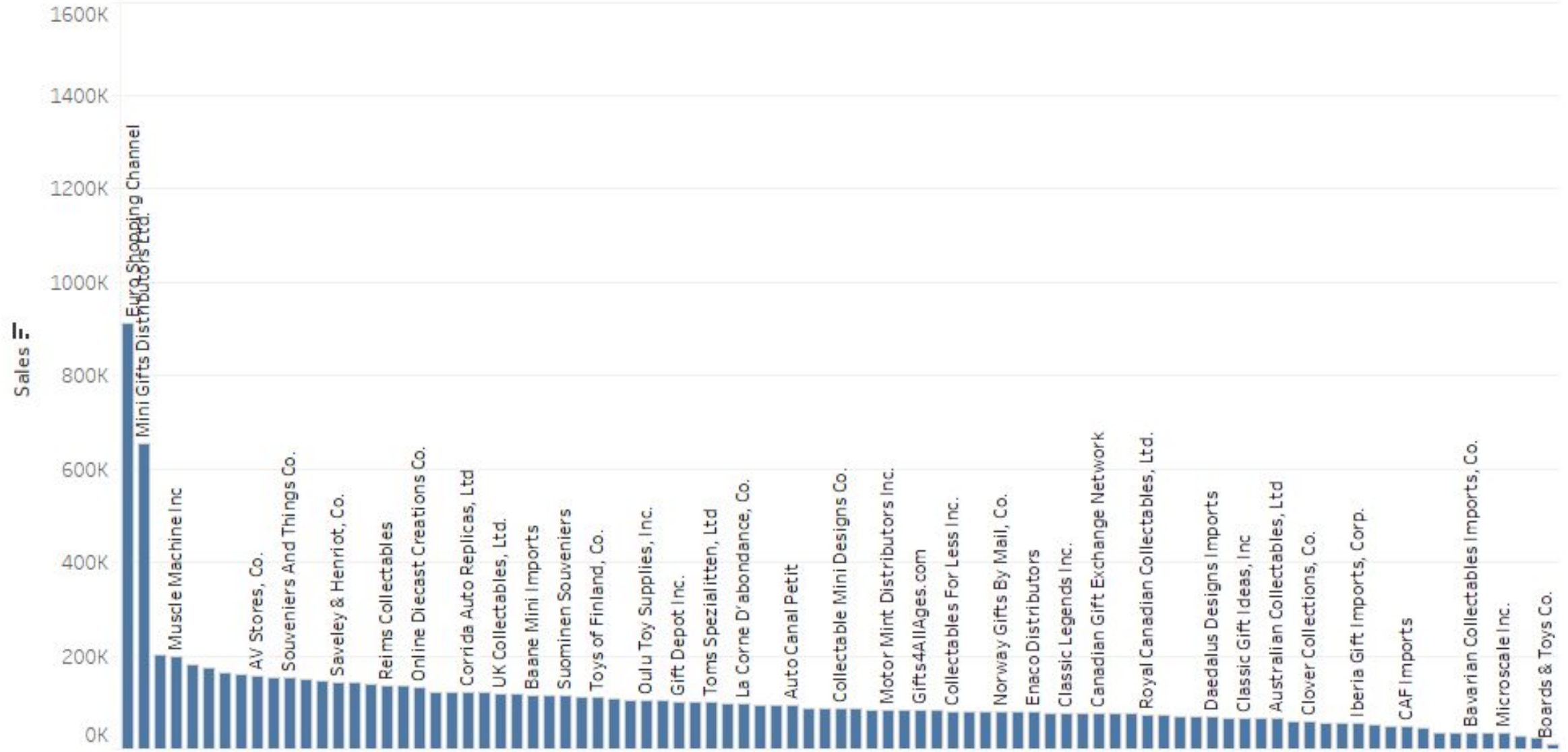


Pivot Table

- Most people deal parts of classic car as it has the higher %.
- Parts of motorcycles are the most Disputed category.
- Least people deal with Train parts.

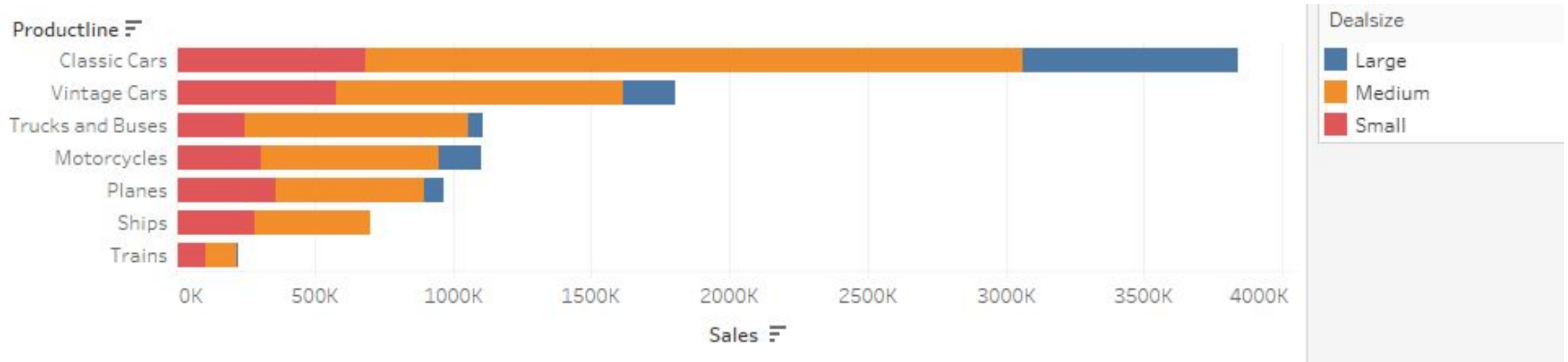
| Productline | Status | | | | | |
|------------------|-----------|----------|------------|---------|----------|---------|
| | Cancelled | Disputed | In Process | On Hold | Resolved | Shipped |
| Classic Cars | 30.46% | 36.02% | 39.91% | 27.40% | 17.12% | 40.19% |
| Motorcycles | | 44.07% | | 2.79% | | 11.83% |
| Planes | 18.22% | 5.32% | | 19.40% | 22.91% | 9.54% |
| Ships | 29.14% | 4.25% | | 13.22% | 26.45% | 6.40% |
| Trains | 2.61% | | | 3.25% | | 2.39% |
| Trucks and Buses | | | 29.73% | 11.28% | 13.58% | 11.40% |
| Vintage Cars | 19.57% | 10.34% | 30.36% | 22.66% | 19.94% | 18.26% |

Sales & Customer names



Sales & Deal Size

- Most sales: Classic cars
- Least sales: Trains
- Most Deal size: medium
- Least Deal size: large
- Sales of truckandbuses and motercycle are aproxx same



Status, Sales and Deal Size

| Deal size | Status | | | | | |
|-----------|-----------|----------|------------|---------|----------|-----------|
| | Cancelled | Disputed | In Process | On Hold | Resolved | Shipped |
| Large | | 42,747 | 27,152 | 42,941 | 8,885 | 1,137,231 |
| Medium | 137,575 | 22,165 | 83,787 | 106,342 | 103,967 | 5,477,396 |
| Small | 56,912 | 7,301 | 33,791 | 29,697 | 37,867 | 2,404,467 |

- No order of large deal type is cancelled it's a good sign.
- Most shipped parts belong to medium deal size type.

PRODUCT LINE AND SALES

Classic Cars
avg Price of Each: 115.20
Sales : 3,842,869

Vintage Cars
avg Price of Each: 90.01
Sales : 1,806,676

Trucks and Buses
avg Price of Each: 104.34
Sales : 1,111,559

Planes
avg Price of Each: 90.52
Sales : 969,323

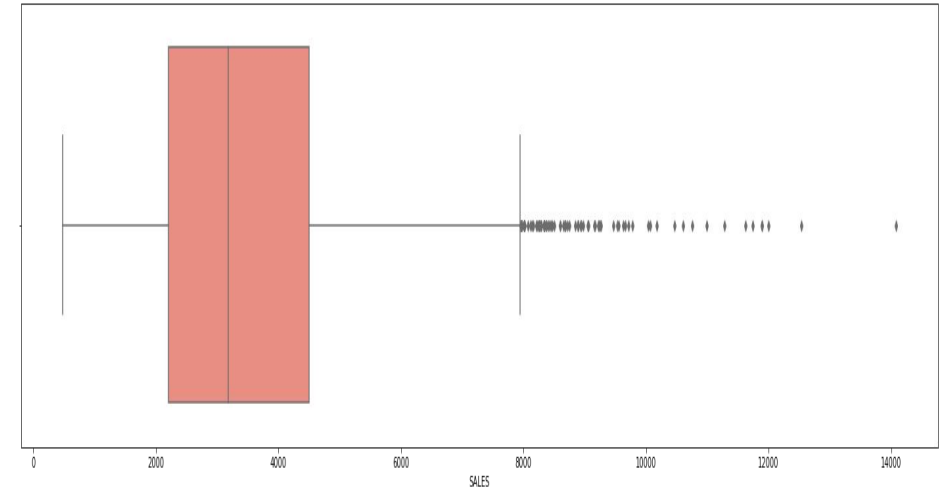
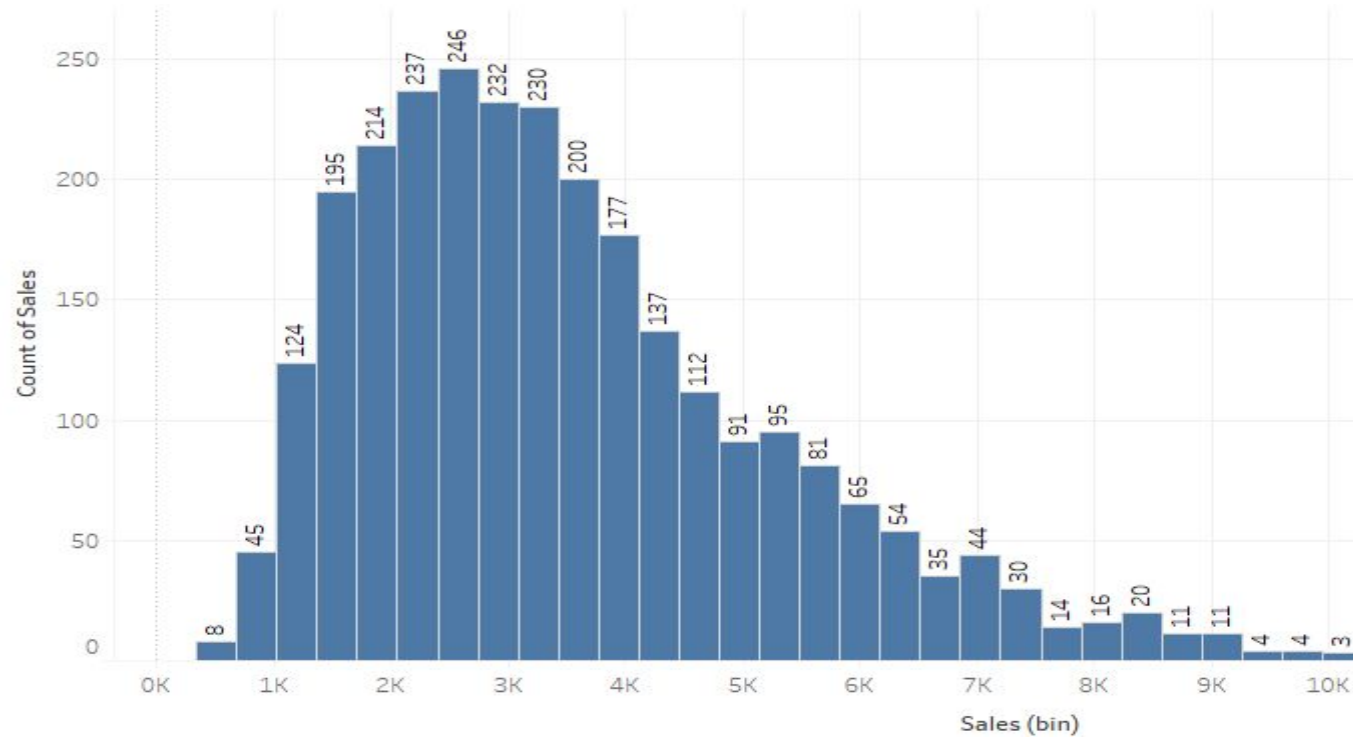
Motorcycles
avg Price of Each: 99.77
Sales : 1,103,512

Ships
avg Price of Each: 88.17
Sales : 700,039

Trains
avg Price of Each: 84.11
Sales : 226,243



Univariate Analysis

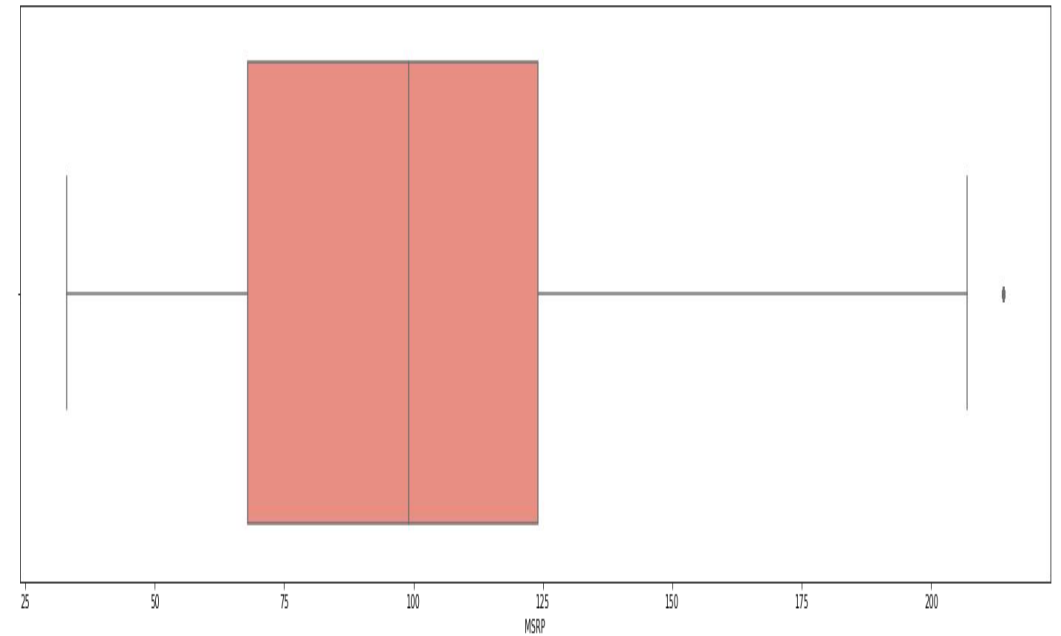
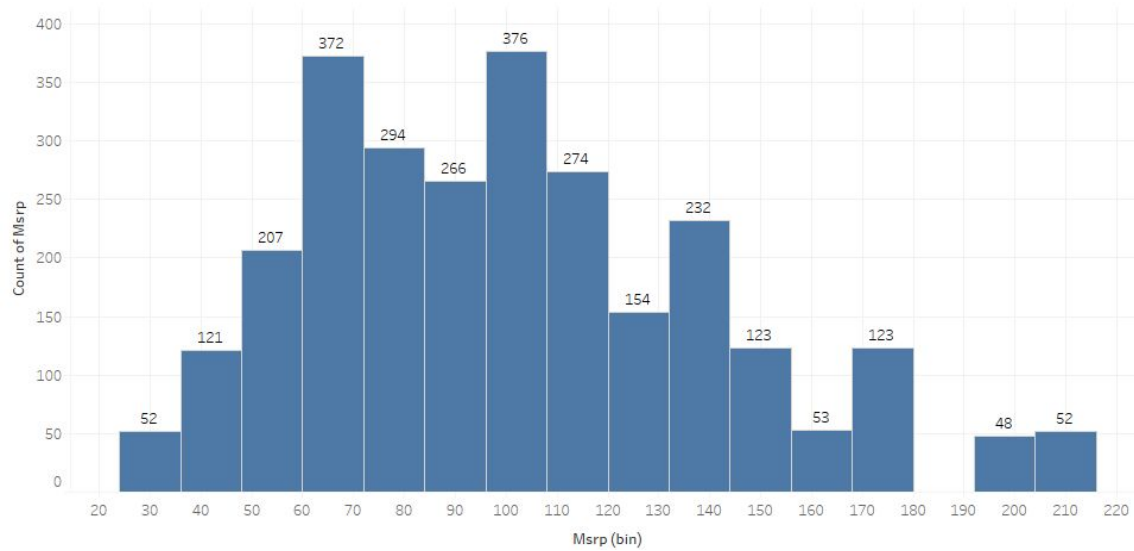


Sales

- We can see that sales data is skewed toward left.
- There are many outliers.
- Most data exist between 1.5k - 3.5k.

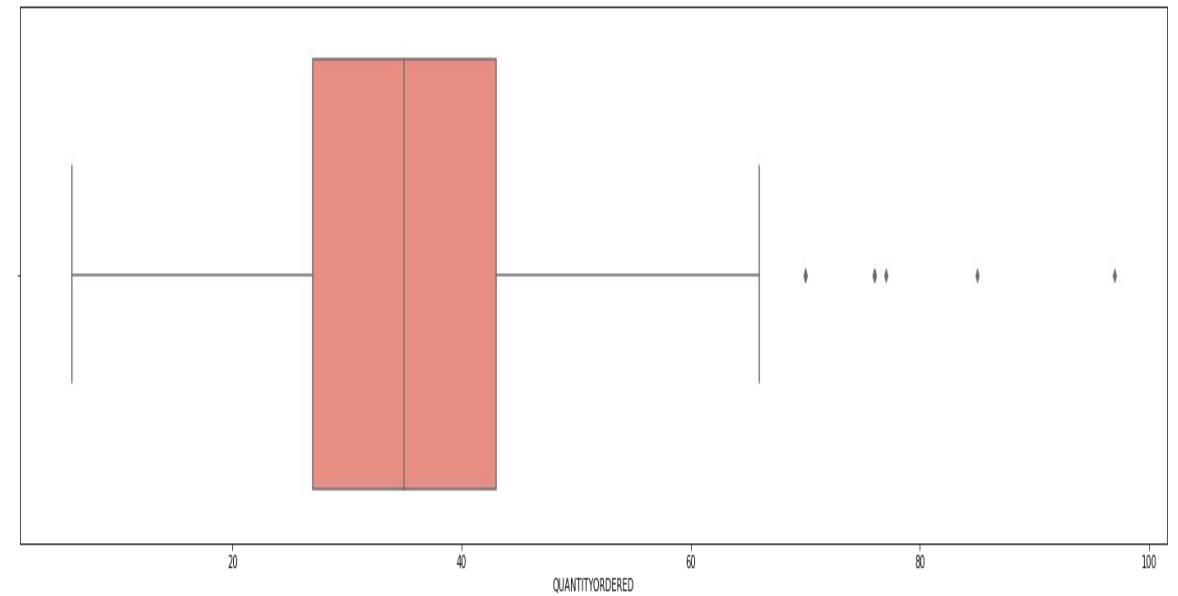
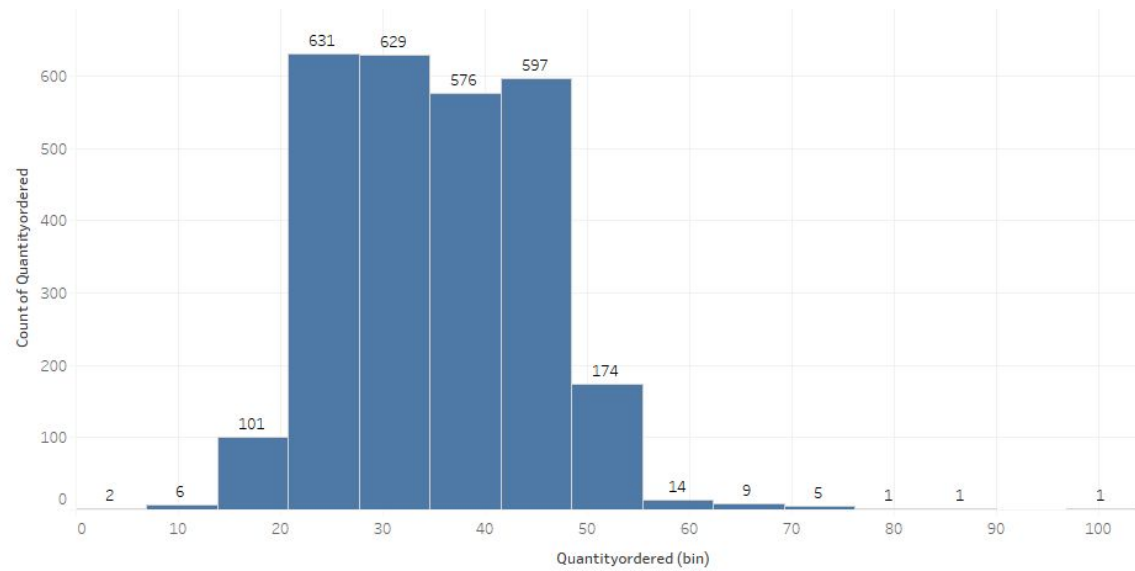
MRSP

- Outliers are present in the MRSP Data
- This data is skewed toward left .



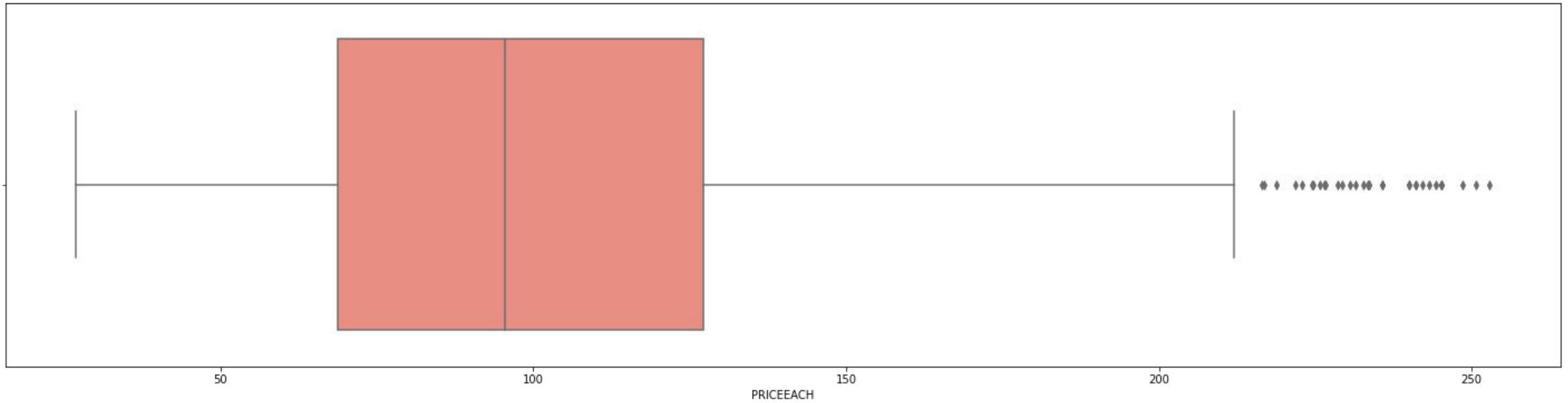
QUANTITY ORDERED

- Outliers are present in the data of Quantity Ordered.



PRICE OF EACH

- Price of Each data has money outliers





Dashboard : Sales

Year Sales



Quartly Sales



Monthly Sales



Productline

(All) ▼

Sales %

| Productline | |
|---------------|--------|
| Classic Cars | 39.37% |
| Motorcycles | 11.31% |
| Planes | 9.93% |
| Ships | 7.17% |
| Trains | 2.32% |
| Trucks and .. | 11.39% |
| Vintage Cars | 18.51% |

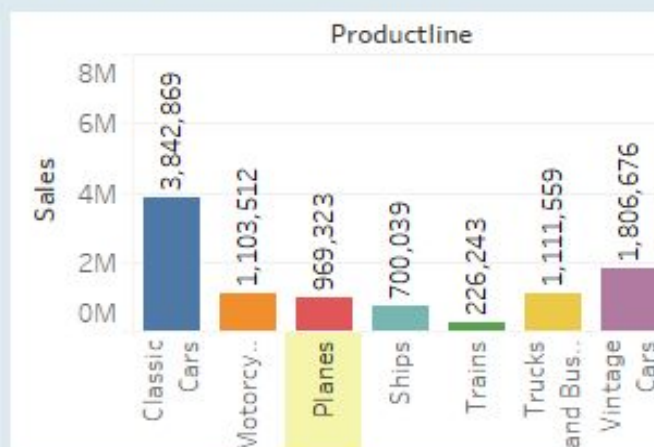
Weekday Sales



Day Sales



Product line and Sales



INFERENCES

The yearly sales have dipped, which is not a good sign.

Quarter 4 has higher sales compared to other quarters.

Sales are consistent in the first 4 months of the year.

Thursday has the lowest sales, and Sunday has the highest sales.

Sales increase from Friday to Sunday and dip from Monday to Thursday.

In the starting days of the month, there are more sales than the end days.

Most orders on hold belong to the USA, and some belong to Sweden.

Spain, USA, UK, and Sweden have approximately the same number of canceled orders.

Spain has the most disputes, and most disputes are solved.

Classic car parts have the highest percentage of sales.

Recommendation

Further investigation is necessary to identify the reasons for the dip in yearly sales.

To capitalize on the high sales in quarter 4, businesses should focus on increasing their inventory during this period.

Businesses should identify the reasons behind the low sales in the 6th month and work towards addressing them.

Businesses should focus their marketing efforts on Thursdays to improve sales on this day.

Strategies such as weekend sales can be employed to increase sales from Friday to Sunday.

Businesses should consider offering discounts or promotions during the start of the month to increase sales.

Efforts should be made to resolve the orders on hold to prevent revenue loss.

Businesses should monitor canceled orders closely and identify any trends to address them.

Steps should be taken to resolve disputes quickly and efficiently to maintain customer satisfaction.

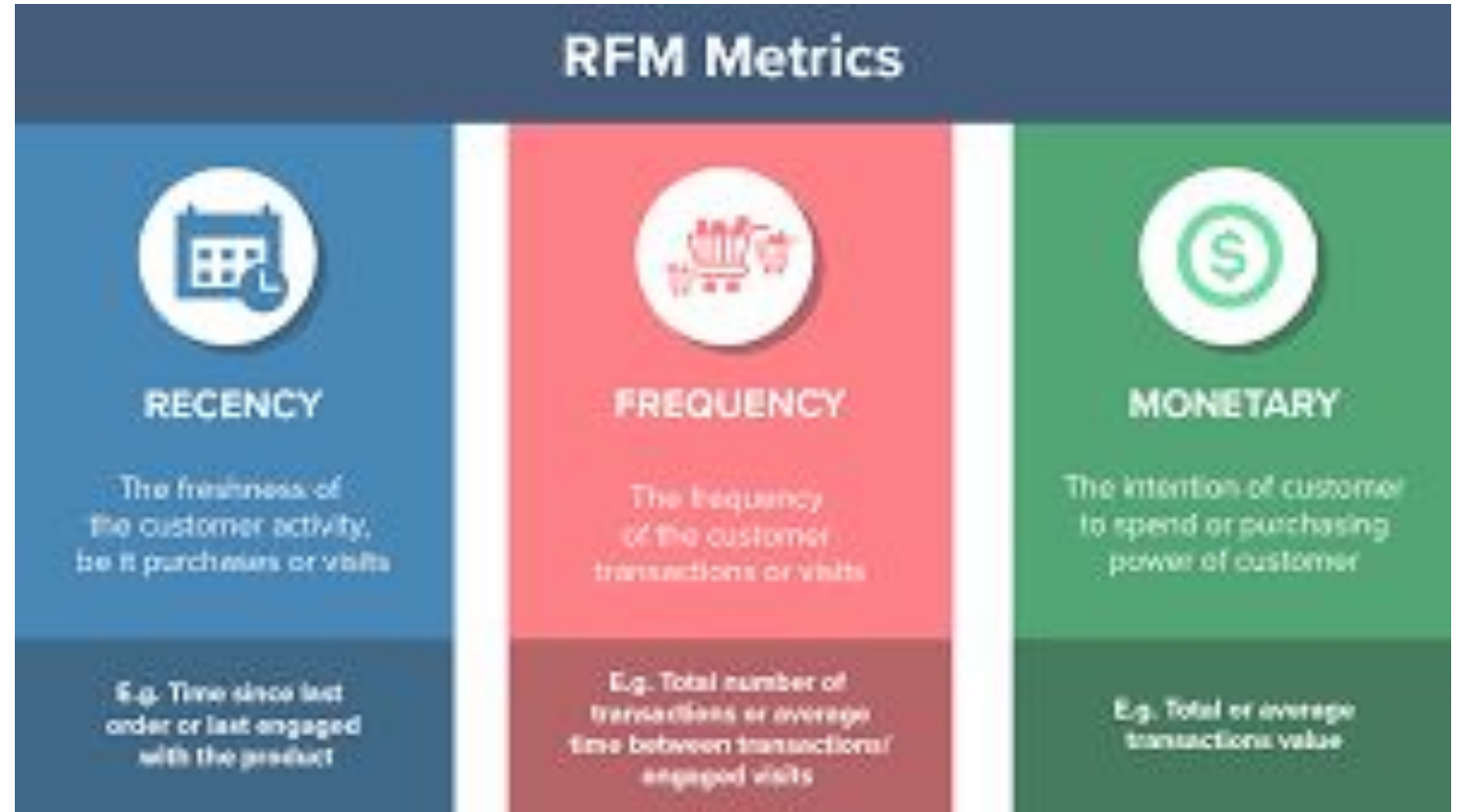
Businesses should consider expanding their inventory of classic car parts, given their high sales percentage.

03 Customer Segmentation using RFM analysis

- What is RFM and which tool used
- What all parameters used and assumptions made
- Output table head
- Workflow image to be put when KNIME used

What is RFM ?

- Recency, frequency, monetary value (RFM) is a marketing analysis tool used to identify a firm's best clients based on the nature of their spending habits.
- An RFM analysis evaluates clients and customers by scoring them in three categories: how recently they've made a purchase, how often they buy, and the size of their purchases.



Tool used : KNIME

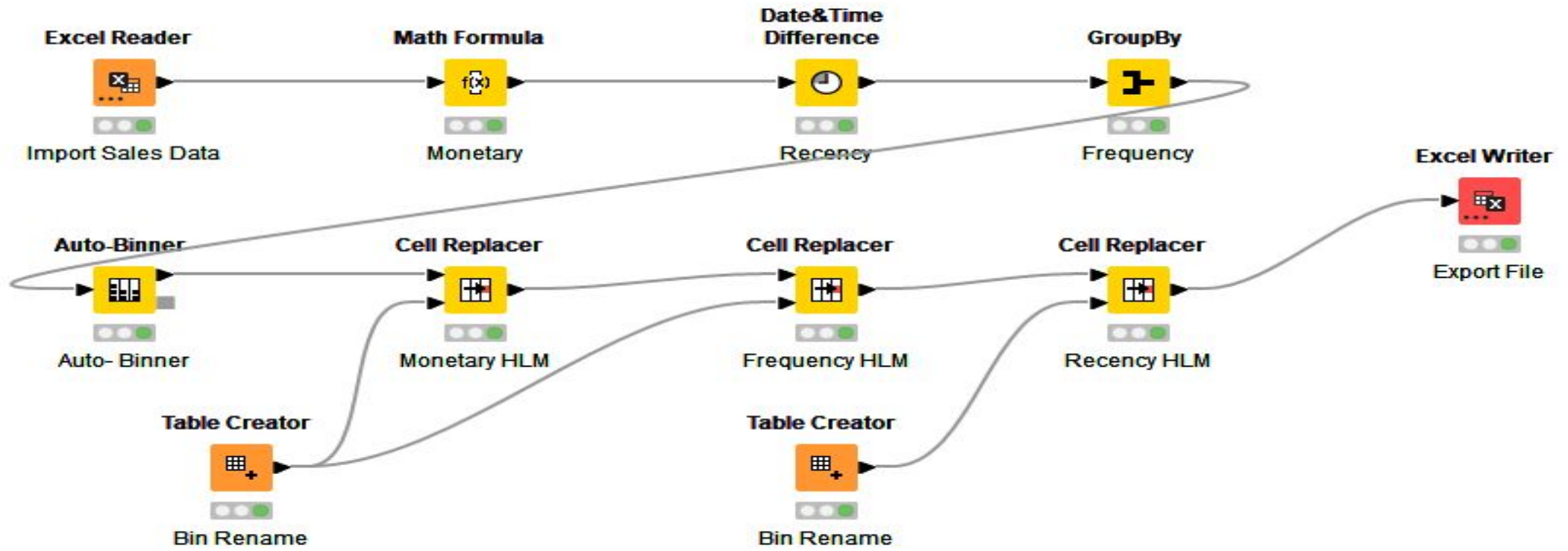
KNIME, the Konstanz Information Miner, is a free and open-source data analytics, reporting and integration platform.



What all parameters used and assumptions made

- As per instructions the column 'Days since last order' is ignored and new column Recency as '[Max(order date)-order date]'
- We have assumed '01-06-2020' as a reference date and created recency column.
- The calculated formula for:
 - Recency :- [min(Recency) customer wise].
 - Frequency:- [count(customer name) customer wise]. We can also take order quantity
 - Monetary:- [sum(unit price + qty ordered) customer wise]. We can also take sales
 - Based on above we have made 3 bins : high , medium , low

KNIME Workflow



Few rows of output

| 1 | CUSTOMER | ORDERNO | QUANTITY | PRICEEACH | ORDERL | SALES | ORDERDATE | DAYS | STATUS | PRODUCT | MSRP | PRODUCT | PHONE | ADDRESSLINE1 | CITY | POSTALCODE | COUNTRY | COMPANY | CONTACT | DEAL | Monetary | Recency | CONTACT | Monetary | Recency | Monetary | Frequency | recency | HML |
|----|-------------|---------|----------|-------------|--------|-----------|------------|------|---------|-----------|------|---------|-------------------|--------------|-----------|------------|---------|---------|---------|-------|-----------|---------|---------|----------|---------|----------|-----------|---------|-----|
| 2 | AV Stores | 51 | 1778 | 91.0845098 | 9.0196 | 157807.81 | 2019-10-14 | 1804 | Shipped | Classic C | 92.8 | S12_110 | (171) 555-1555 | Fauntleroy C | Manche | EC2 5NT | UK | 51 | 51 | Medi | 157807.81 | 197 | Bin 3 | Bin 3 | Bin 2 | H | H | M | |
| 3 | Alpha Cog | 20 | 687 | 101.16 | 4.95 | 70488.44 | 2018-11-08 | 2236 | Shipped | Classic C | 97.2 | S10_475 | 61.77.6555 | 1 rue Alsace | Toulouse | 31000 | France | 20 | 20 | Medi | 70488.44 | 65 | Bin 1 | Bin 1 | Bin 1 | L | L | H | |
| 4 | Amica Mo | 26 | 843 | 110.8526923 | 7.6154 | 94117.26 | 2019-08-17 | 1319 | Shipped | Classic C | 108 | S10_194 | 011-4988555 | Via Monte E | Torino | 10100 | Italy | 26 | 26 | Large | 94117.26 | 266 | Bin 2 | Bin 2 | Bin 3 | M | M | L | |
| 5 | Anna's De | 46 | 1469 | 106.4241304 | 6.4348 | 153996.13 | 2020-03-09 | 1464 | Shipped | Classic C | 105 | S10_194 | 02 9936 8555 | 201 Miller S | North Sy | 2060 | Austral | 46 | 46 | Small | 153996.13 | 84 | Bin 3 | Bin 3 | Bin 2 | H | H | M | |
| 6 | Atelier gri | 7 | 270 | 92.23857143 | 2 | 24179.96 | 2019-09-27 | 1424 | Shipped | Motorcy | 95.6 | S10_201 | 40.32.2555 | 54, rue Roy | Nantes | 44000 | France | 7 | 7 | Medi | 24179.96 | 189 | Bin 1 | Bin 1 | Bin 2 | L | L | M | |
| 7 | Australian | 23 | 705 | 90.04173913 | 6.6957 | 64591.46 | 2018-11-21 | 1690 | Shipped | Vintage | 88.1 | S18_134 | 61-9-3844-6555 | 7 Allen Stre | Glen Wa | 3150 | Austral | 23 | 23 | Medi | 64591.46 | 23 | Bin 2 | Bin 1 | Bin 1 | L | M | H | |
| 8 | Australian | 55 | 1926 | 104.5901818 | 7.0364 | 200995.41 | 2019-02-20 | 1699 | Shipped | Motorcy | 104 | S10_167 | 03 9520 4555 | 636 St Kilda | Melbour | 3004 | Austral | 55 | 55 | Medi | 200995.41 | 185 | Bin 3 | Bin 3 | Bin 2 | H | H | M | |
| 9 | Australian | 15 | 545 | 110.554 | 3.0667 | 59469.12 | 2018-11-06 | 1349 | Shipped | Classic C | 112 | S10_194 | 61-7-3844-6555 | 31 Duncan S | South Br | 4101 | Austral | 15 | 15 | Large | 59469.12 | 120 | Bin 1 | Bin 1 | Bin 2 | L | L | M | |
| 10 | Auto Assc | 18 | 637 | 99.48780142 | 8.5556 | 64834.32 | 2019-10-11 | 1266 | Shipped | Classic C | 100 | S10_194 | 30.59.8555 | 67, avenue | Versaille | 78000 | France | 18 | 18 | Large | 64834.32 | 234 | Bin 1 | Bin 1 | Bin 3 | L | L | L | |
| 11 | Auto Can | 27 | 1001 | 94.25518519 | 6.3333 | 93170.66 | 2019-01-15 | 1763 | Shipped | Motorcy | 94.9 | S10_167 | (1) 47.55.6555 | 25, rue Laur | Paris | 75016 | France | 27 | 27 | Medi | 93170.66 | 55 | Bin 2 | Bin 2 | Bin 1 | M | M | H | |
| 12 | Auto-Mot | 8 | 287 | 92.8 | 2 | 26479.26 | 2018-06-16 | 2282 | Shipped | Ships | 87.4 | S18_302 | 6175558428 | 16780 Pom | Brickhav | 58339 | USA | 8 | 8 | Medi | 26479.26 | 181 | Bin 1 | Bin 1 | Bin 2 | L | L | M | |
| 13 | Baane Mi | 32 | 1082 | 108.57375 | 6.3438 | 116599.19 | 2019-10-15 | 1806 | Shipped | Motorcy | 107 | S10_167 | 07-98 9555 | Erling Skakk | Stavern | 4110 | Norway | 32 | 32 | Medi | 116599.19 | 209 | Bin 2 | Bin 2 | Bin 2 | M | M | M | |
| 14 | Bavarian | 14 | 401 | 84.28928571 | 7.5 | 34993.92 | 2019-09-15 | 2239 | Shipped | Planes | 82.7 | S18_166 | +49 89 61 08 9555 | Hansastr. 15 | Munich | 80686 | Germa | 14 | 14 | Medi | 34993.92 | 260 | Bin 1 | Bin 1 | Bin 3 | L | L | L | |
| 15 | Blauer Se | 22 | 811 | 108.0313636 | 3.8636 | 85171.59 | 2018-10-04 | 1697 | Shipped | Classic C | 106 | S12_109 | +49 69 66 90 2555 | Lyonerstr. 3 | Frankfur | 60528 | Germa | 22 | 22 | Medi | 85171.59 | 209 | Bin 2 | Bin 2 | Bin 2 | M | M | M | |
| 16 | Boards & | 3 | 102 | 89.80666667 | 1.3333 | 9129.35 | 2020-02-08 | 1969 | Shipped | Classic C | 92.3 | S12_338 | 3105552373 | 4097 Dougl | Glendale | 92561 | USA | 3 | 3 | Medi | 9129.35 | 114 | Bin 1 | Bin 1 | Bin 2 | L | L | M | |
| 17 | CAF Impo | 13 | 468 | 104.9630769 | 5.3077 | 49642.05 | 2019-03-19 | 2273 | Shipped | Classic C | 107 | S12_110 | +34 913 728 555 | Merchants | Madrid | 28023 | Spain | 13 | 13 | Large | 49642.05 | 440 | Bin 1 | Bin 1 | Bin 3 | L | L | L | |
| 18 | Cambridg | 11 | 357 | 101.3290909 | 3.2727 | 36163.62 | 2019-03-10 | 1788 | Shipped | Classic C | 97.4 | S10_194 | 6175555555 | 4658 Baden | Cambrid | 51247 | USA | 11 | 11 | Medi | 36163.62 | 390 | Bin 1 | Bin 1 | Bin 3 | L | L | L | |
| 19 | Canadian | 22 | 703 | 105.3409091 | 6 | 75238.92 | 2018-12-05 | 1620 | Shipped | Classic C | 106 | S10_194 | (604) 555-3392 | 1900 Oak St | Vancouv | V3F 2K1 | Canada | 22 | 22 | Large | 75238.92 | 223 | Bin 2 | Bin 2 | Bin 2 | M | M | M | |

Inferences from RFM Analysis and identified segments

Who are your best customers?

Which customers are on the verge of churning?

Who are your lost customers?

Who are your loyal customers?

Top 5 best customers

| CUSTOMERNAME | Monetary HML | Frequency HLM | recency HML |
|------------------------------|--------------|---------------|-------------|
| Euro Shopping Channel | H | H | H |
| La Rochelle Gifts | H | H | H |
| Mini Gifts Distributors Ltd. | H | H | H |
| Souveniers And Things Co. | H | H | H |
| Salzburg Collectables | H | H | H |
| L'ordine Souveniers | H | H | H |

- According to RFM score we have grouped the top customers.
- We have given importance to recency more and ordered the customers accordingly.

Top loyal customers

| CUSTOMERNAME | Monetary HML | Frequency HLM | recency HML |
|----------------------------|--------------|---------------|-------------|
| Australian Collectors, Co. | H | H | M |
| Muscle Machine Inc | H | H | M |
| Dragon Souveniers, Ltd. | H | H | M |
| Land of Toys Inc. | H | H | M |
| AV Stores, Co. | H | H | M |
| Anna's Decorations, Ltd | H | H | M |

- Based on RFM analysis these are the loyal customers
- We have focused on monetary value
- If we focus on these customer we can turn them in to best customers.

Verge on churning Customers

| CUSTOMERNAME | Monetary HML | Frequency HLM | recency HML |
|------------------------------|--------------|---------------|-------------|
| Saveley & Henriot, Co. | H | H | L |
| Vida Sport, Ltd | M | M | L |
| Herkku Gifts | M | M | L |
| Marta's Replicas Co. | M | M | L |
| Amica Models & Co. | M | M | L |
| Collectable Mini Designs Co. | M | M | L |

- As per RFM score we can see that these are the top customers on the verge of churning.
- We should focus on these customers before we lose them.
- We should try some action plan to convert them into regular customers.

Top Lost Customers

| CUSTOMERNAME | Monetary HML | Frequency HLM | recency HML |
|--------------------------------|--------------|---------------|-------------|
| Double Decker Gift Stores, Ltd | L | L | L |
| West Coast Collectables Co. | L | L | L |
| Signal Collectibles Ltd. | L | L | L |
| Daedalus Designs Imports | L | L | L |
| CAF Imports | L | L | L |
| Osaka Souvenirs Co. | L | L | L |

- As per RFM score we can see that these are the customers which we have lost.
- Their recency is very low as well as they have not made purchase frequently.
- We should study them survey them to understand the reasons why we lost them. And take further steps so that we do not lose the customers.

05 Recommendations



Recommendations

- According to RFM analysis, customers can be categorized into four distinct groups: best, loyal, verge of churn, and lost customers. It is important to develop a focused approach for each group in order to optimize customer retention and enhance customer experience.
- **best customers**, it is recommended to provide personalized recognition, exclusive offers, and incentives to ensure that they continue to choose our company over others. By doing so, we can maintain their loyalty and strengthen the long-term relationship.
- **loyal customers**, it is essential to offer periodic discounts and offers to keep them engaged and interested in our products or services. By keeping them engaged, we can turn them into our best customers and improve their satisfaction level with our brand.
- **verge of churn customer**, we need to develop an effective action plan to prevent them from leaving the company. We can conduct surveys, offer incentives, and personalize the communication to identify and address their concerns, and thereby increase their loyalty towards our brand.
- **lost customers**, it is important to analyze their behavior and preferences to identify the reasons for their departure. By identifying the key drivers of customer churn, we can develop targeted strategies to prevent such occurrences in the future and improve our overall retention rate.



Thank you