

MRA PROJECT

RFM Analysis



By Suneel Kumar Pentapalli



- ❑ Problem Statement
- ❑ What's in our data
- ❑ Exploratory Data Analysis
- ❑ Customer Segmentation using RFM Model
- ❑ Inferences

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.

What's in our data



- The dataset comprises 2747 entries and 20 columns, representing 89 unique customers. None of the columns have missing values

| # | 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 |

What's in our data(Contd.)



- Given data has no duplicates.
- Below is the five point summary of numeric columns.

```
df.shape
```

```
(2747, 20)
```

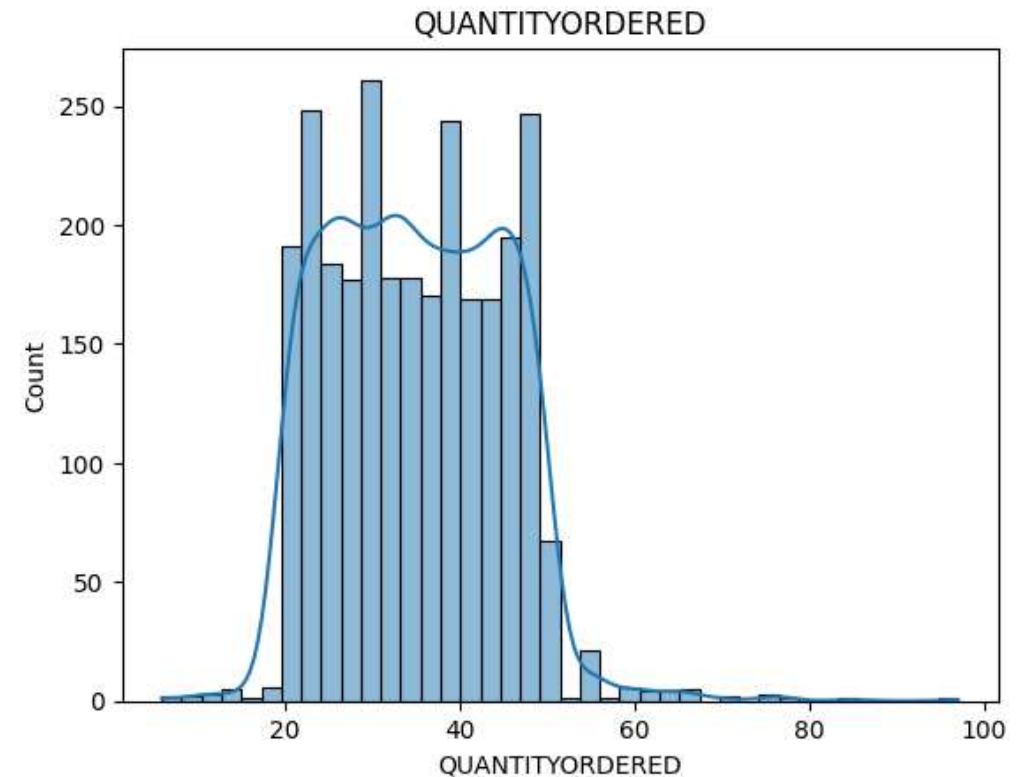
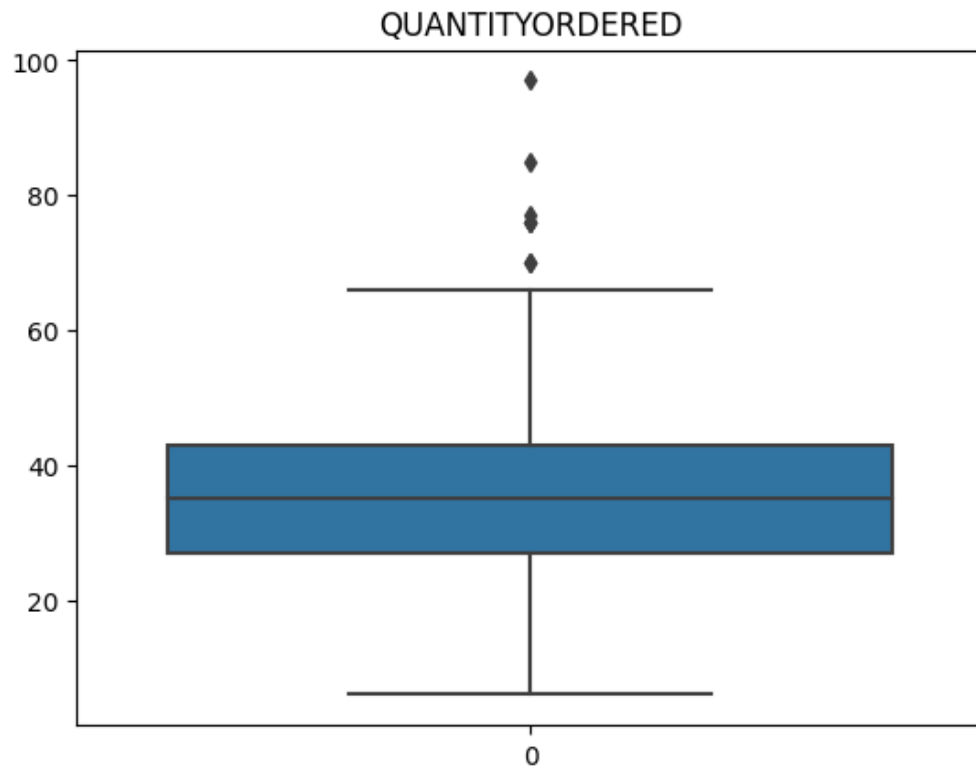
| | count | mean | std | min | 25% | 50% | 75% | max |
|----------------------|--------|----------|---------|----------|----------|----------|----------|----------|
| ORDERNUMBER | 2747.0 | 10259.76 | 91.88 | 10100.00 | 10181.00 | 10264.00 | 10334.50 | 10425.00 |
| QUANTITYORDERED | 2747.0 | 35.10 | 9.76 | 6.00 | 27.00 | 35.00 | 43.00 | 97.00 |
| PRICEEACH | 2747.0 | 101.10 | 42.04 | 26.88 | 68.74 | 95.55 | 127.10 | 252.87 |
| ORDERLINENUMBER | 2747.0 | 6.49 | 4.23 | 1.00 | 3.00 | 6.00 | 9.00 | 18.00 |
| SALES | 2747.0 | 3553.05 | 1838.95 | 482.13 | 2204.35 | 3184.80 | 4503.09 | 14082.80 |
| DAYS_SINCE_LASTORDER | 2747.0 | 1757.09 | 819.28 | 42.00 | 1077.00 | 1761.00 | 2436.50 | 3562.00 |
| MSRP | 2747.0 | 100.69 | 40.11 | 33.00 | 68.00 | 99.00 | 124.00 | 214.00 |

Exploratory Data Analysis



Univariate Analysis(QUANTITYORDERED):

- On an average, 30 items have been ordered.
- Distribution of orders is almost uniform with outliers due to very few high order quantities.

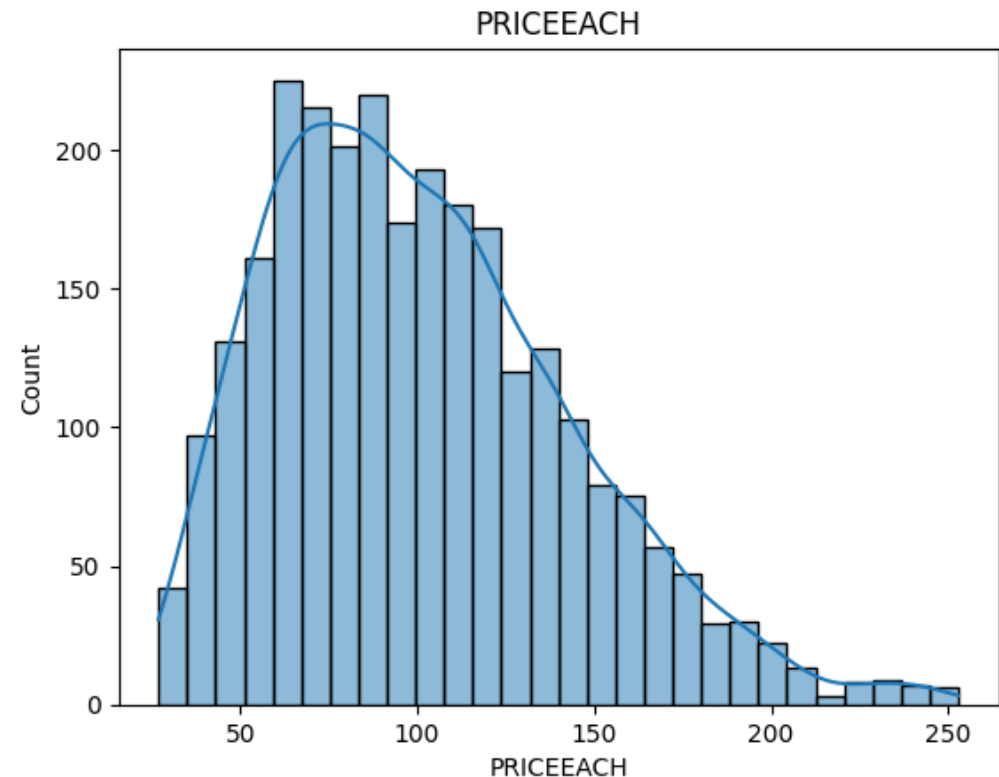
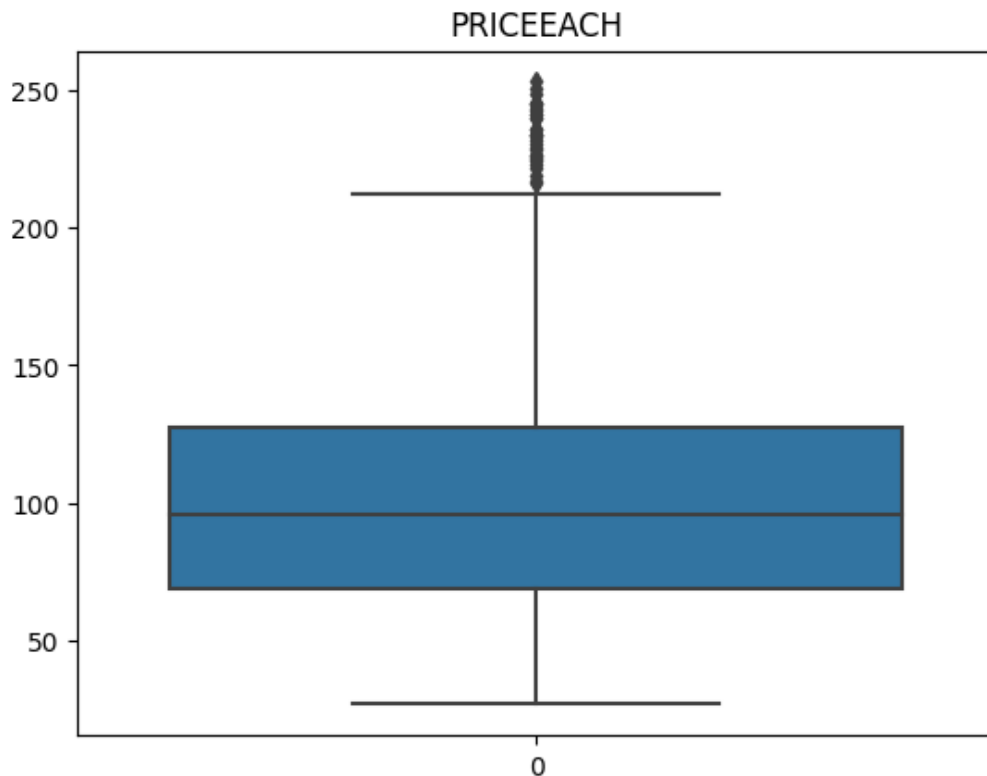


Exploratory Data Analysis(Contd.)



Univariate Analysis(PRICEEACH):

- Average price of all of our products is around 101.
- Distribution of prices is almost normal with outliers due to some premium priced products.

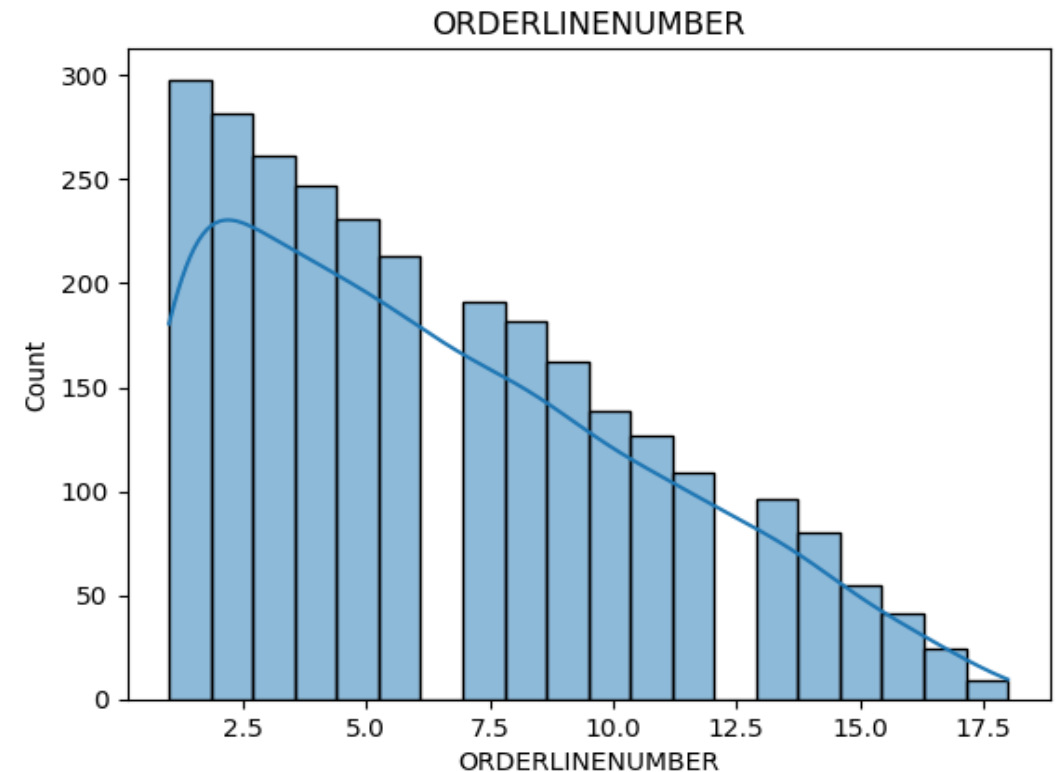
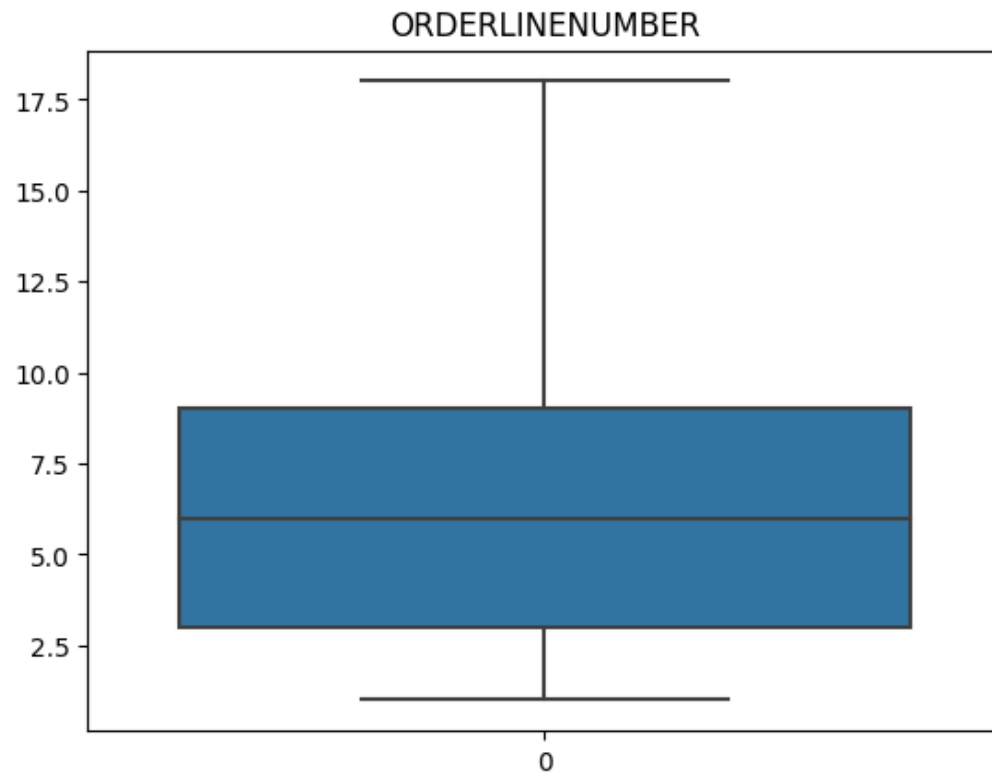


Exploratory Data Analysis(Contd.)



Univariate Analysis(ORDERLINENUMBER):

- It represents the line number of each item within an order.

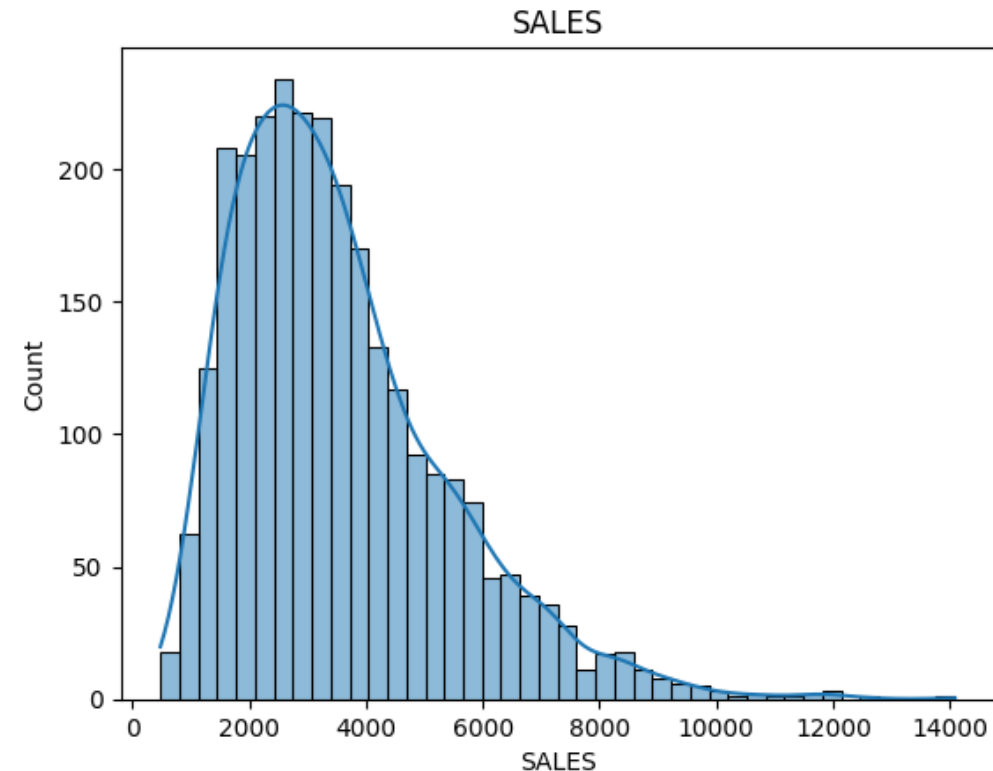
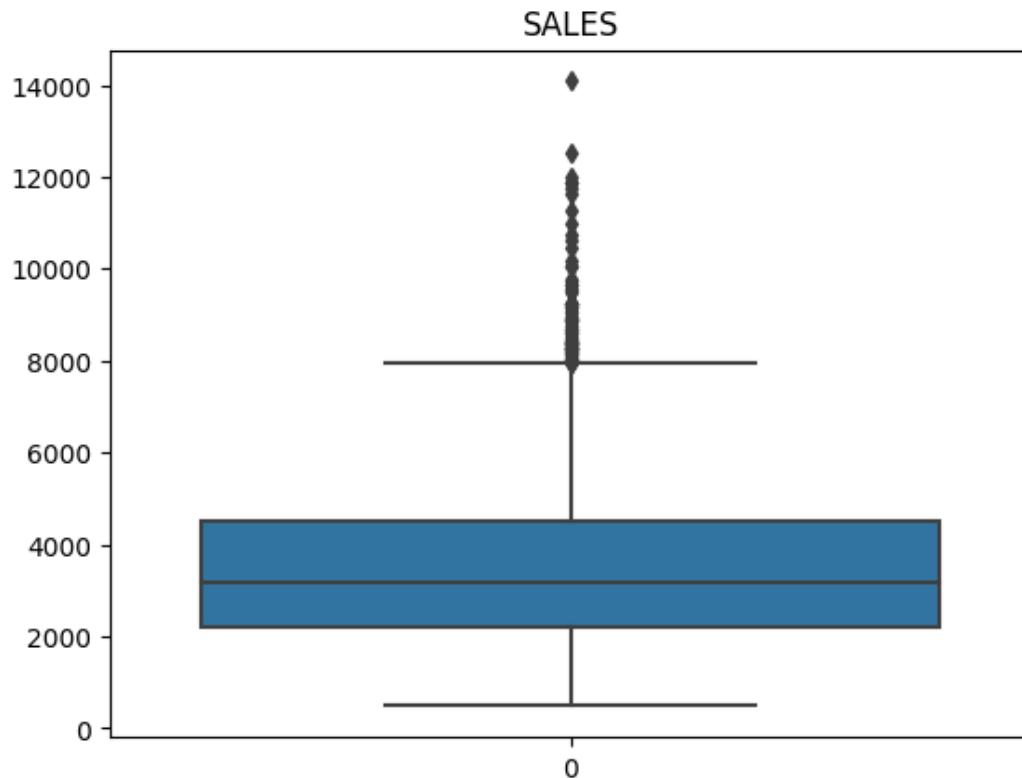


Exploratory Data Analysis(Contd.)



Univariate Analysis(SALES):

- Majority of our sales amount lies between 2000 and 7000 currency units.
- Yet there are very few sales crossing 8000 currency units. Every sale that's crossing 8000 is observed to be a Large Order.

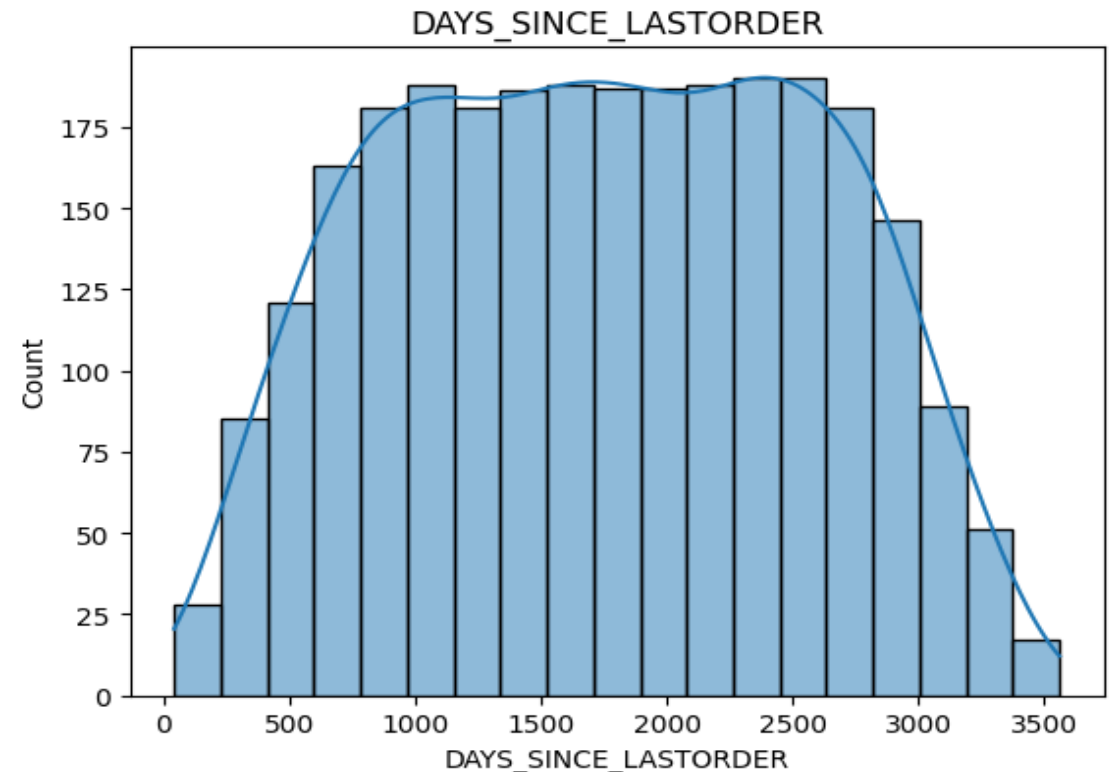
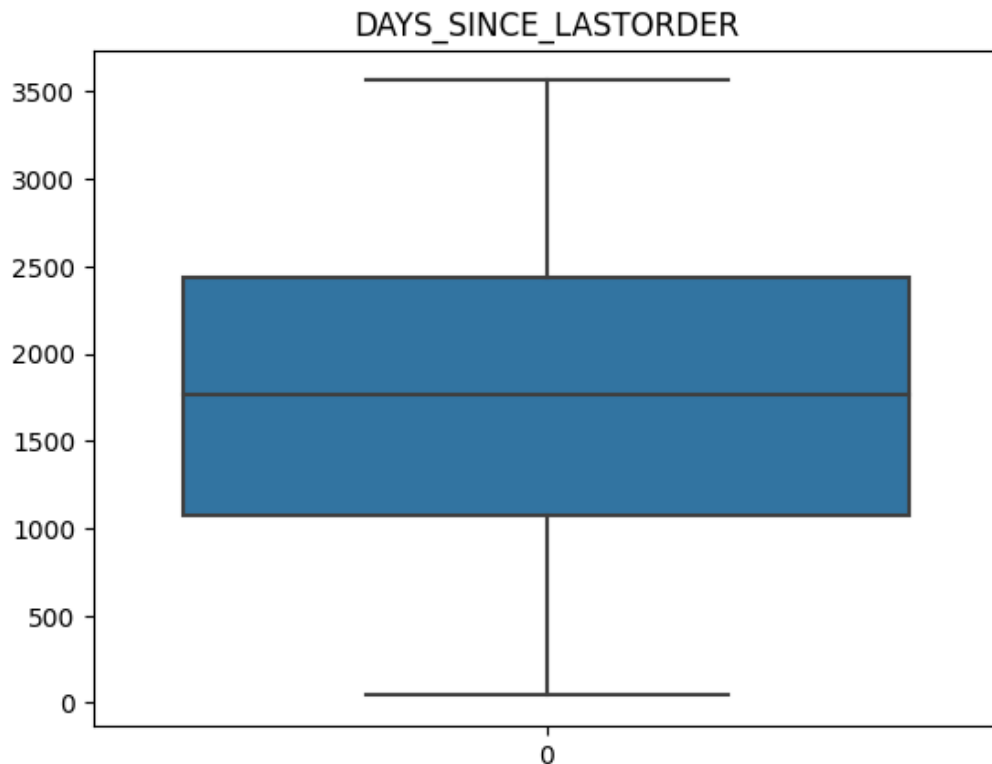


Exploratory Data Analysis(Contd.)



Univariate Analysis(DAYS_SINCE_LASTORDER):

- On an average, customer returns to the store to buy again after almost 1757 days, which is very concerning.
- There are very few customers that shows up before 500 days.

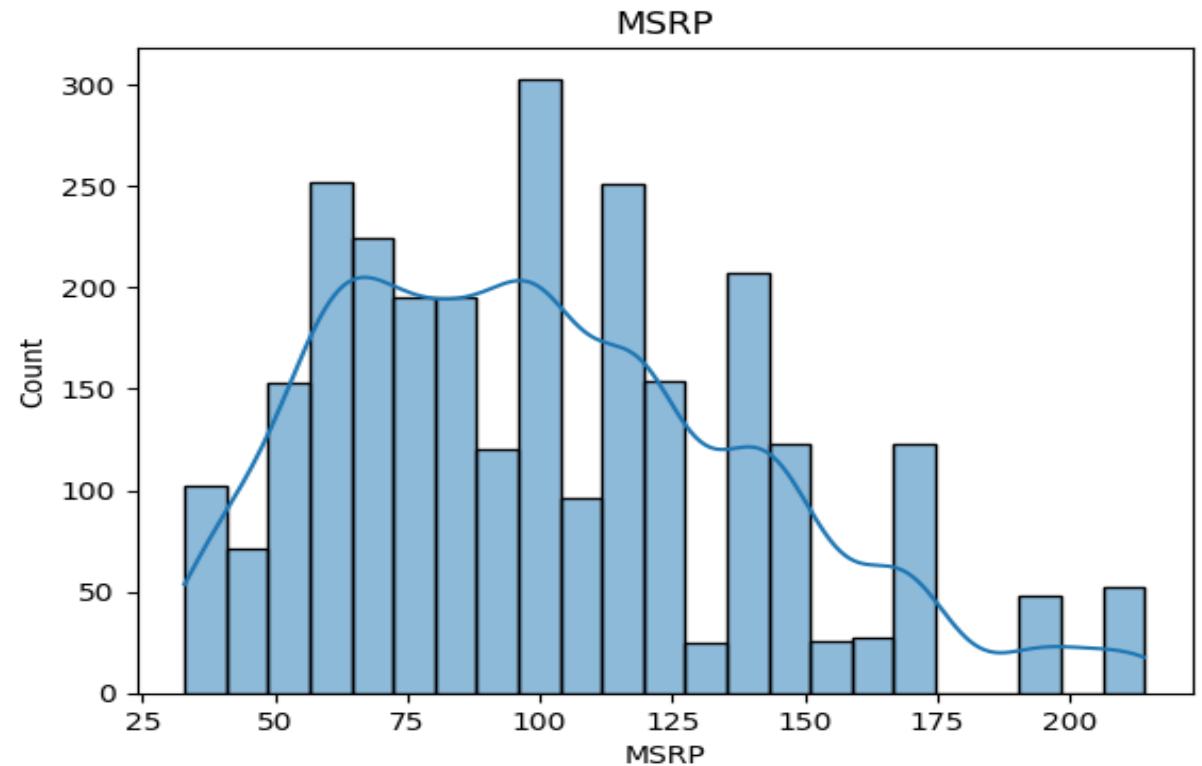
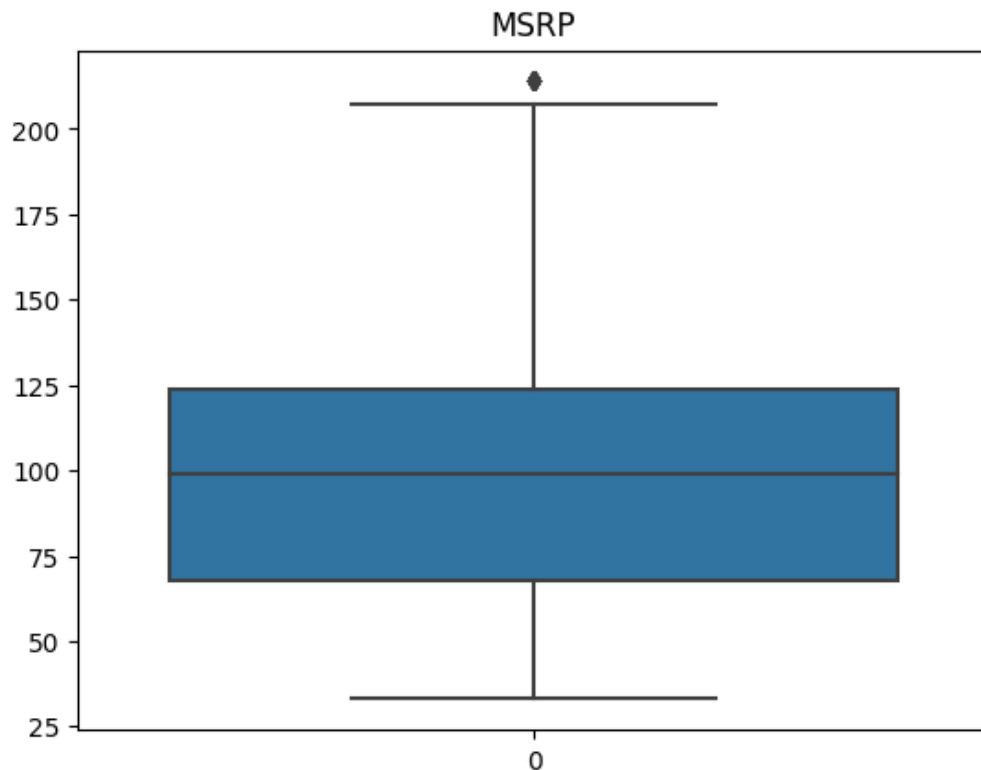


Exploratory Data Analysis(Contd.)



Univariate Analysis(MSRP):

- It stands for Manufacturer's Suggested Retail Price and represents the suggested selling price for each item.

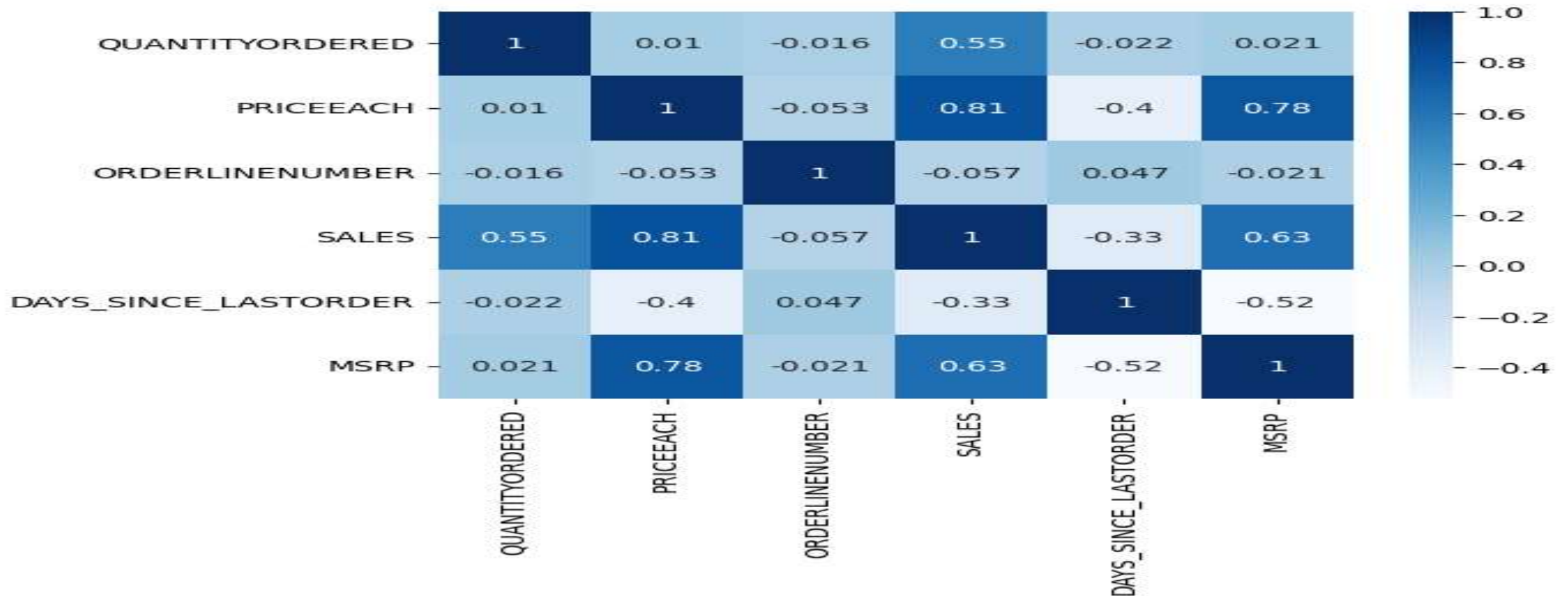


Exploratory Data Analysis(Contd.)



Correlation between above discussed variables:

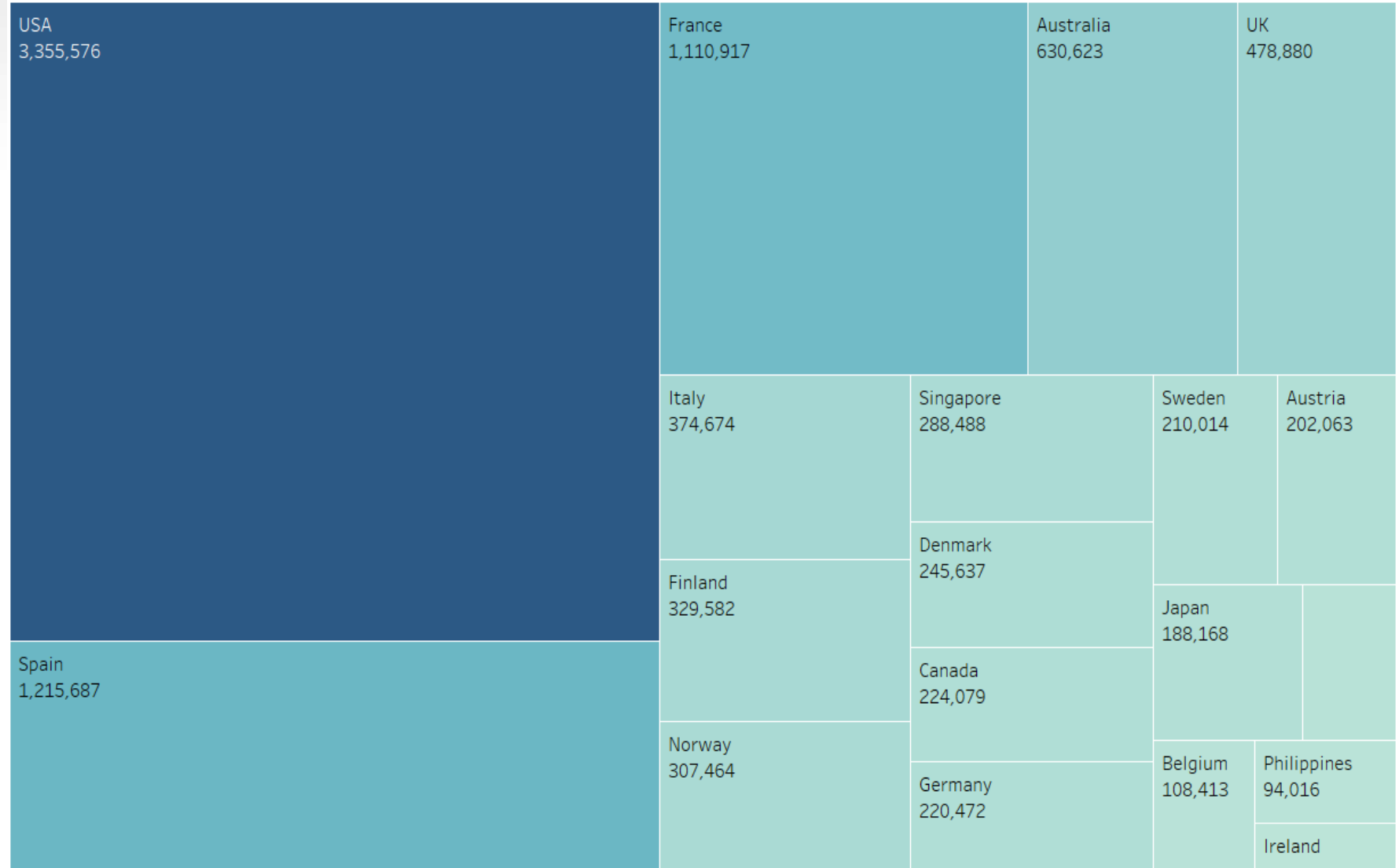
- There is high amount of positive correlation between sales and item's price. None of the other attributes has any useful correlations.



Sales across Countries



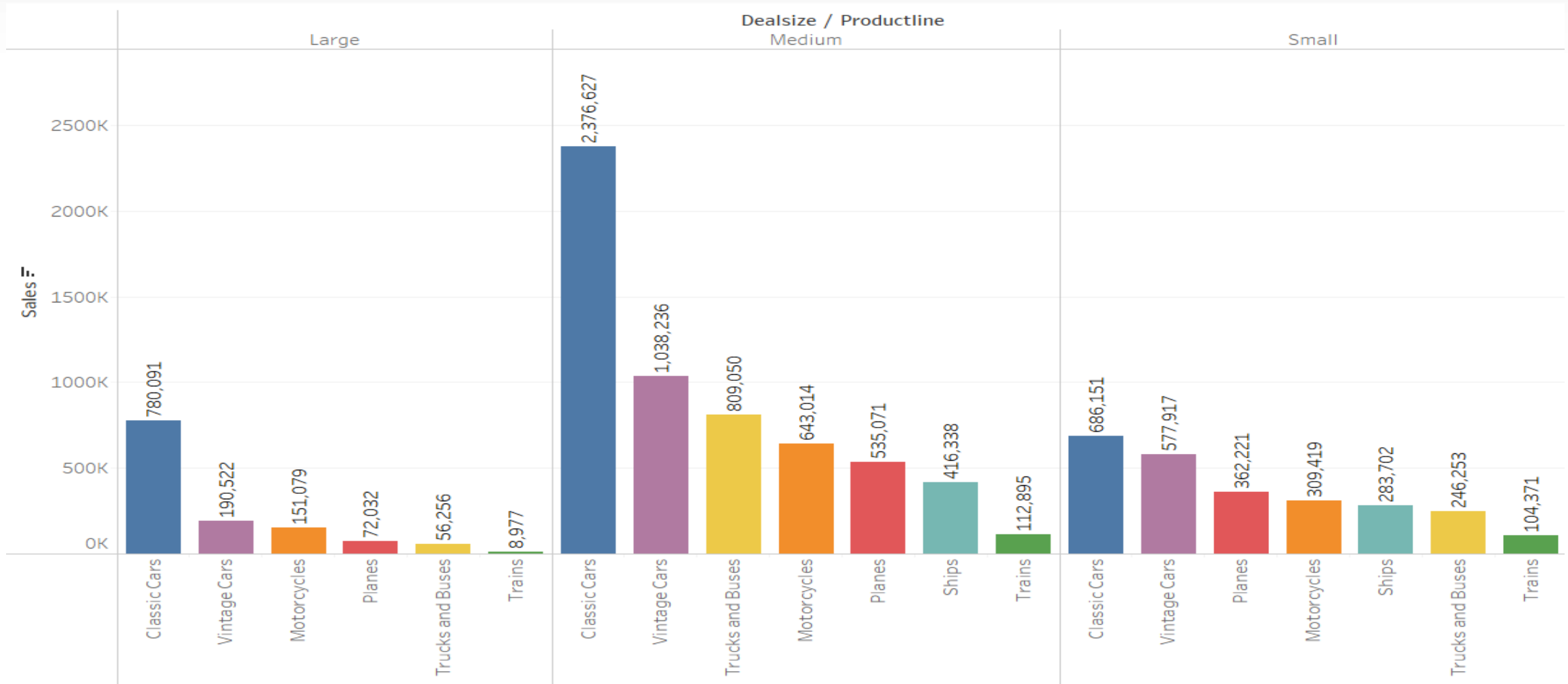
- USA contributed the highest in terms of sales with total of 33,55,576 currency units.
- Ireland contributed to the least amount of sales with 57,756 currency units.



Sales vs Productlines across each deal size.



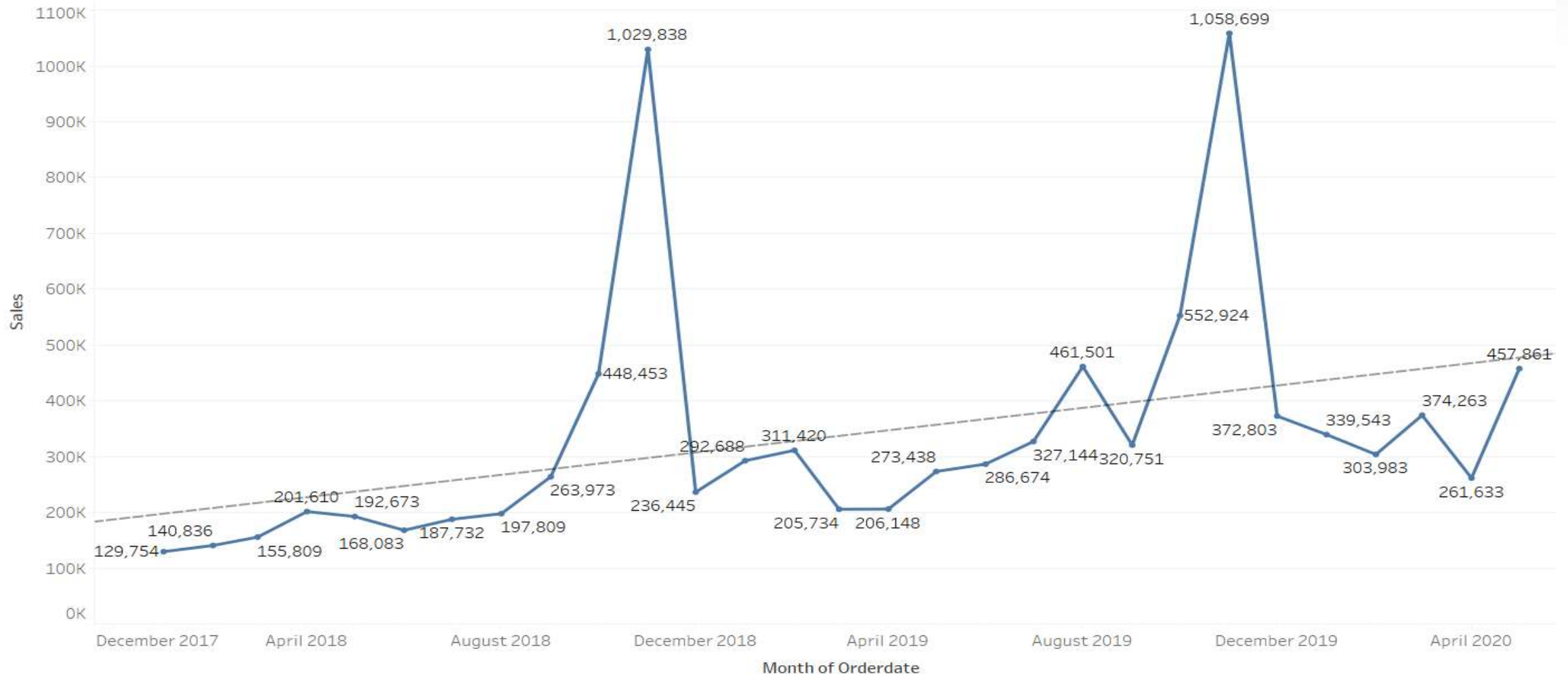
- Across all the deal sizes, Classic Cars are the highest sold and Trains are the least sold.



Sales Trend/Seasonality



There is a clear seasonality across years with peak sales in November, trending slightly upwards.

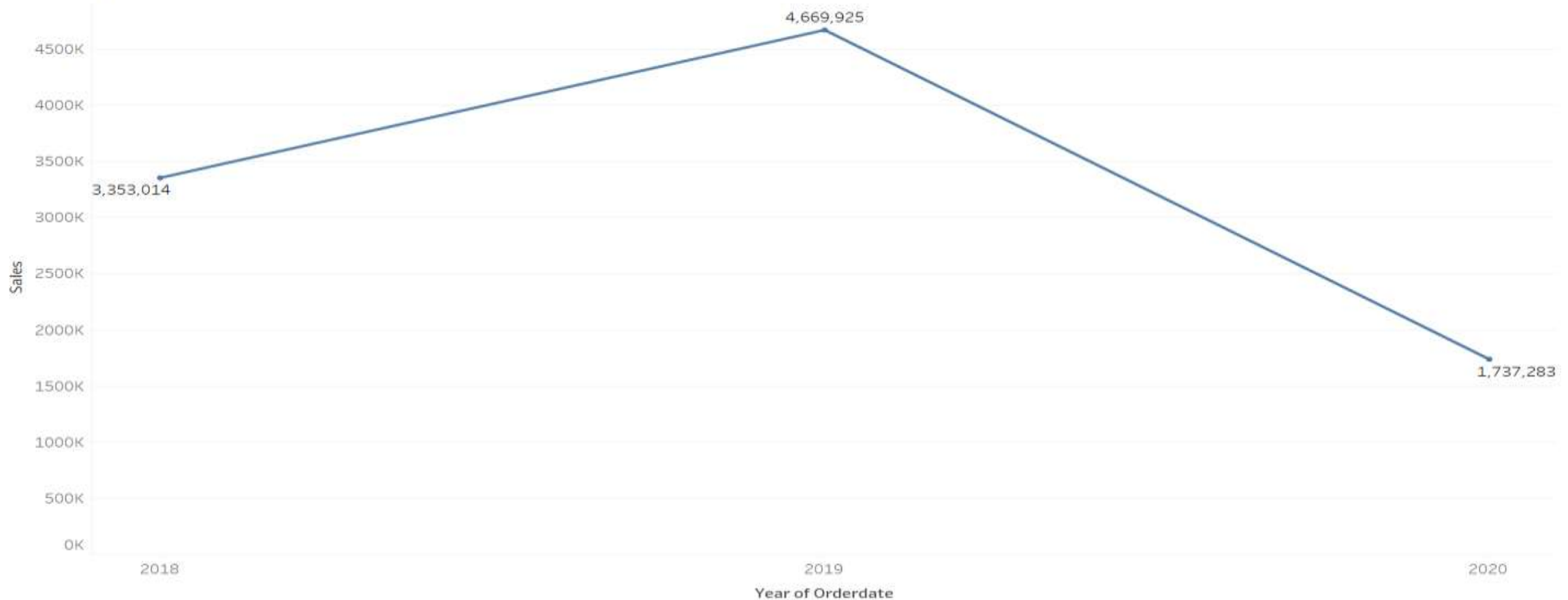


Yearly Sales



There is considerable dip in sales in year 2020, which should be given attention to. 2019 is the year with more sales out of all the three years.

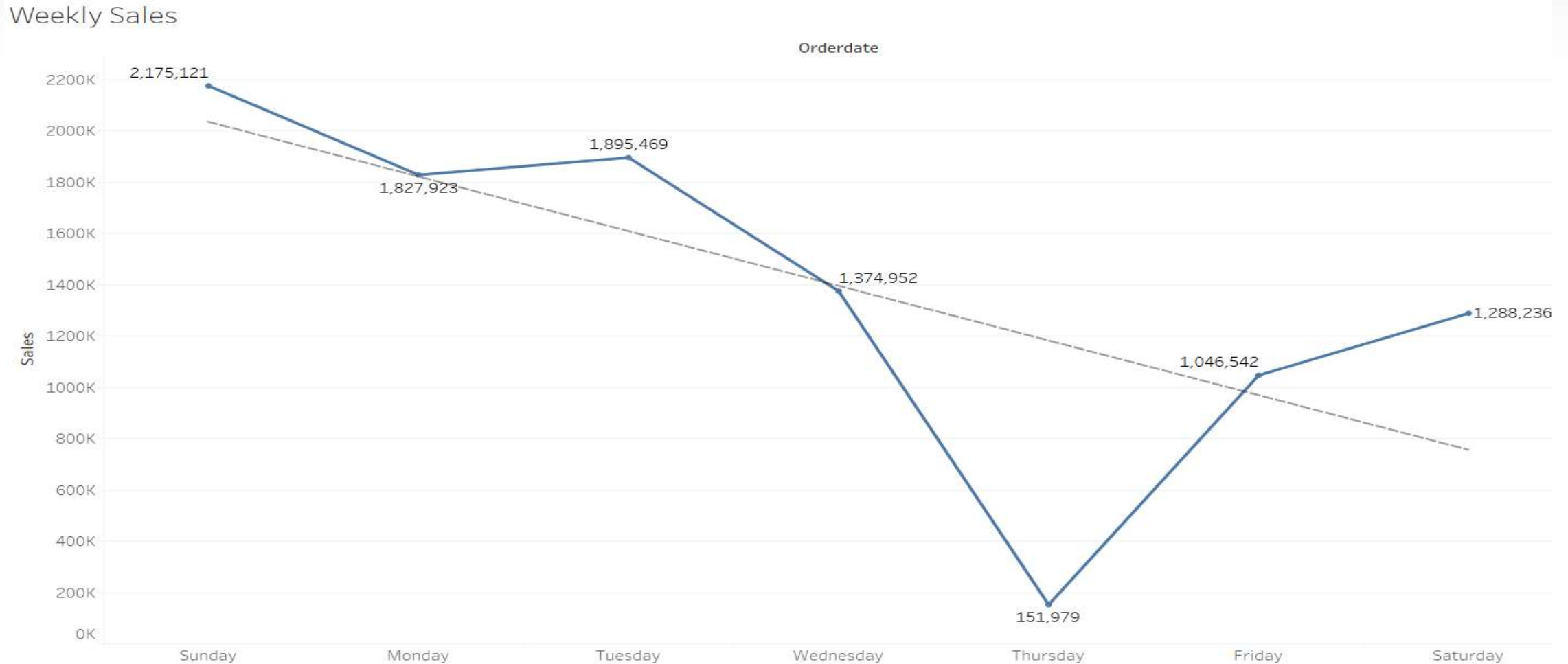
Yearly Sales



Weekly Sales



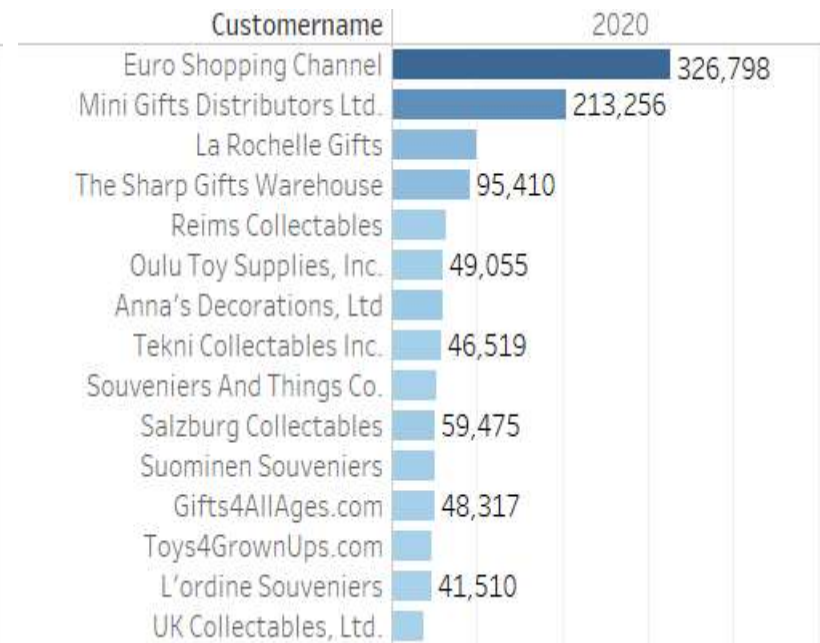
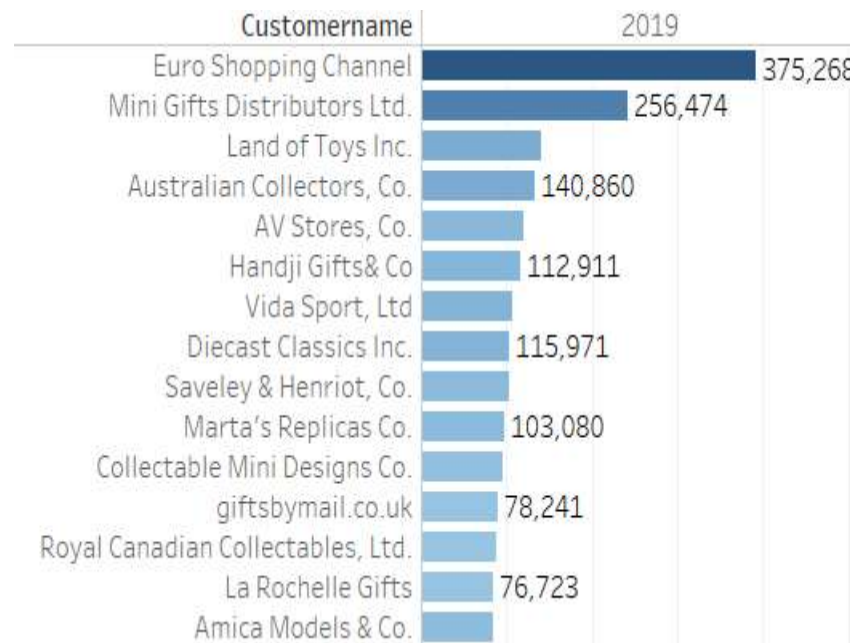
On an average, Sunday seems to have highest number of sales and Thursday seems to have the lowest sum of sales.



Top 15 Customers across years



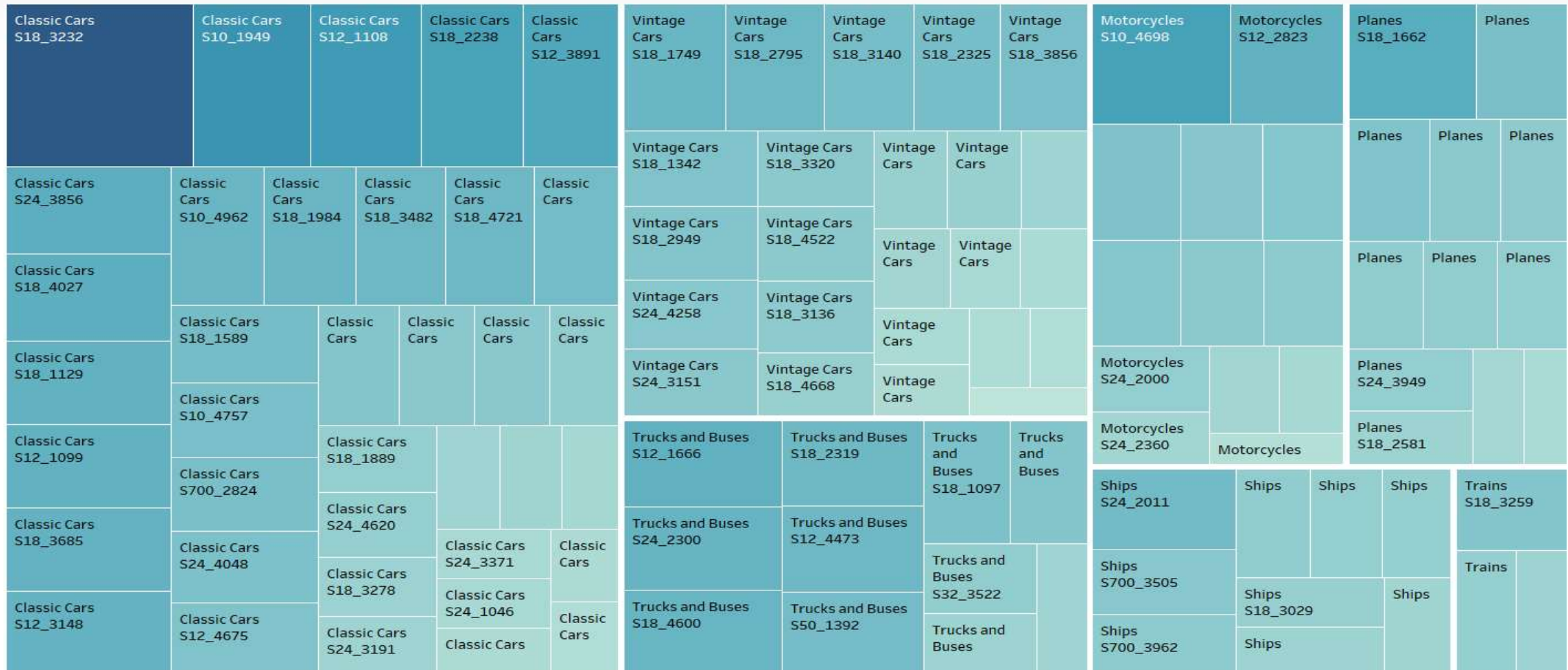
Euro Shipping and Mini Gifts are the most contributing customer across all three years. Other customer among top 15 across years are as follows.



Highest selling product code



Below is the view of highest selling product code across each product line.

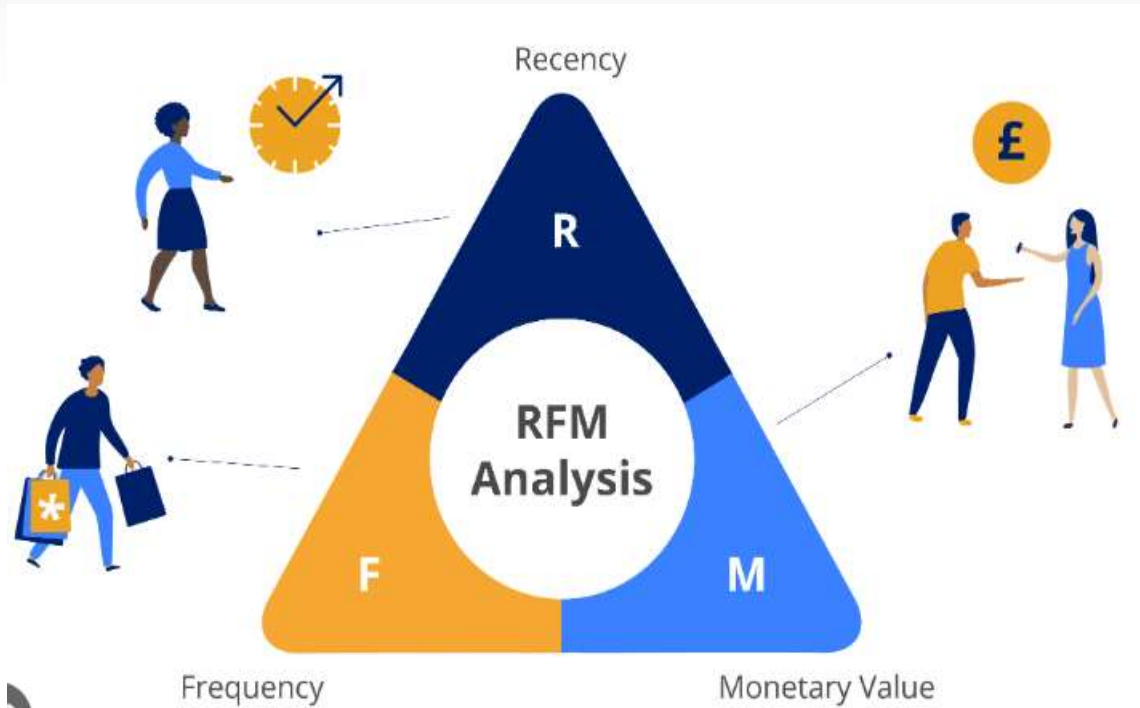


Customer Segmentation using RFM Model



- RFM stands for
 - Recency(days since last ordered/bought)
 - Frequency(number of times same customer is contributing towards sales.)
 - Monetary(Quantity of each item*Price of each item)
- RFM analysis is a great technique to identify set of customers for distinctive treatments.
- RFM segmentation allows marketers to target specific groups of customers.
- With promotions that are more significant for their actual behaviour, we can better help our individual customer.

Customer Segmentation using RFM Model



- Customer segmentation has been done into four categories(High Value, Medium Value, Low Value and Very Low Value)
- Below aggregations were used while segmenting customers.
 - ✓ Recency - Days since last ordered.
 - ✓ Frequency - Unique count of OrderID.
 - ✓ Monetary - Total Sum of sales for each customer.

Inferences from RFM(Golden Customers)



Below are our best customers with high monetary values, highly frequent buying and more active(recent) buying style.

| CUSTOMERNAME | RFM_Score |
|------------------------------|-----------|
| The Sharp Gifts Warehouse | 444 |
| Technics Stores Inc. | 444 |
| Souvenirs And Things Co. | 444 |
| Salzburg Collectables | 444 |
| Mini Gifts Distributors Ltd. | 444 |
| Land of Toys Inc. | 444 |
| La Rochelle Gifts | 444 |
| Euro Shopping Channel | 444 |
| Diecast Classics Inc. | 444 |
| Australian Collectors, Co. | 444 |
| Anna's Decorations, Ltd | 444 |

Inferences from RFM(Medium Value Customers)



Below are our customers with average recency, average frequent and medium monetary value. We can entice these customers to convert into golden customers.

| CUSTOMERNAME | RFM_Score |
|------------------------------|-----------|
| Baane Mini Imports | 443 |
| Online Diecast Creations Co. | 424 |
| UK Collectables, Ltd. | 423 |
| Oulu Toy Supplies, Inc. | 423 |
| Gift Depot Inc. | 423 |
| FunGiftIdeas.com | 423 |
| Auto Canal Petit | 423 |
| Quebec Home Shopping Network | 422 |
| Lyon Souveniers | 422 |
| Gifts4AllAges.com | 422 |

Inferences from RFM(Low Value Customers)



Below are our customers that are at the verge of churning.

| CUSTOMERNAME | RFM_Score |
|----------------------------|-----------|
| Amica Models & Co. | 313 |
| Atelier graphique | 321 |
| Tekni Collectables Inc. | 322 |
| Classic Legends Inc. | 322 |
| Toys of Finland, Co. | 323 |
| Suominen Souvenirs | 323 |
| La Corne D'abondance, Co. | 323 |
| Scandinavian Gift Ideas | 324 |
| Corrida Auto Replicas, Ltd | 324 |
| Volvo Model Replicas, Co | 342 |
| Tokyo Collectables, Ltd | 343 |
| Reims Collectables | 344 |

Inferences from RFM(Very low value Customers)

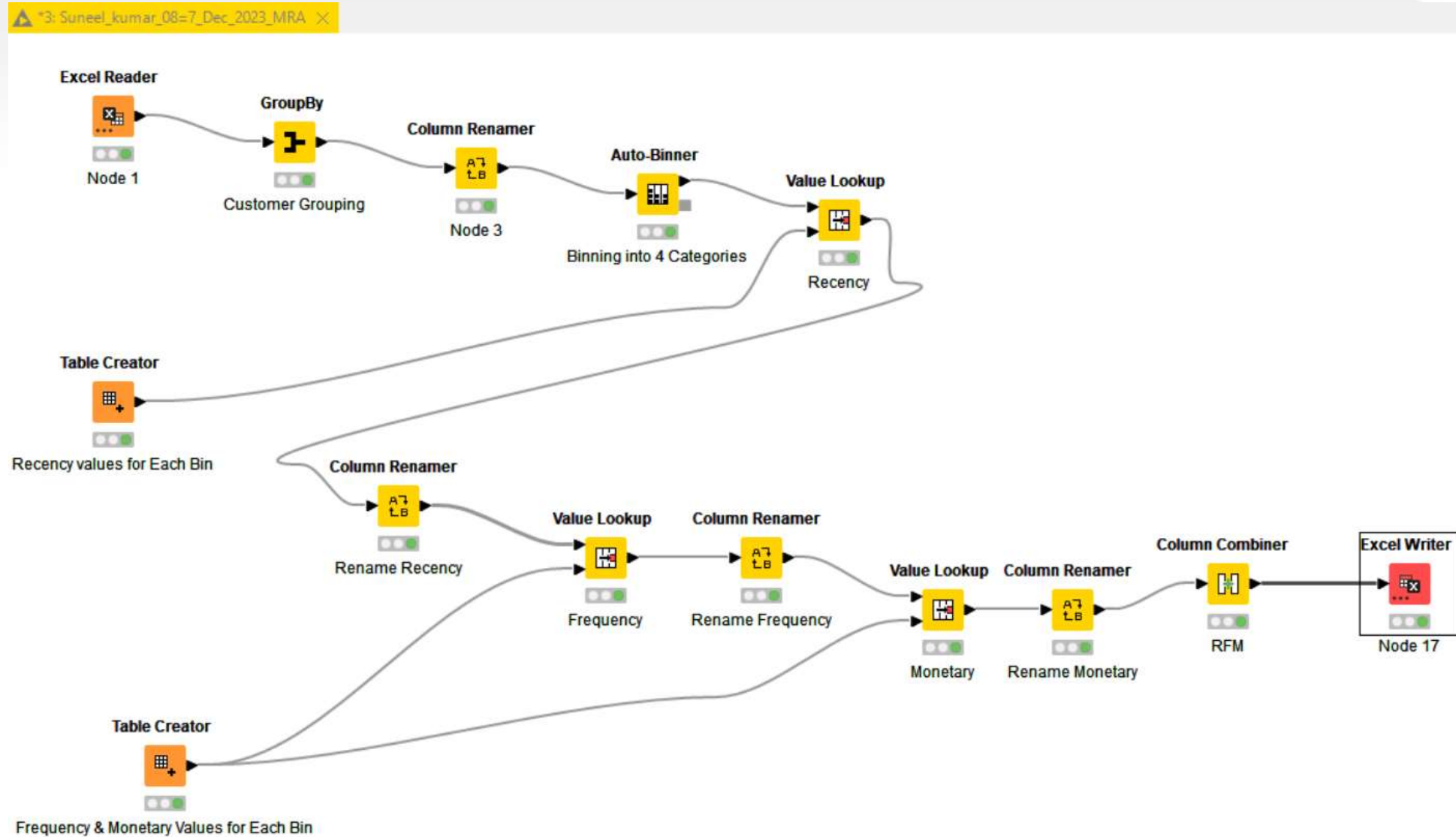


These are customers that can be considered lost as they have very less frequent and less monetary values and not very active customers.

| CUSTOMERNAME | RFM_Score |
|------------------------------------|-----------|
| Signal Collectibles Ltd. | 111 |
| Iberia Gift Imports, Corp. | 111 |
| Double Decker Gift Stores, Ltd | 111 |
| Clover Collections, Co. | 111 |
| Bavarian Collectables Imports, Co. | 111 |
| Norway Gifts By Mail, Co. | 112 |
| Diecast Collectables | 112 |
| Mini Auto Werke | 121 |
| Gift Ideas Corp. | 121 |
| Auto-Moto Classics Inc. | 121 |

Note: Customers mentioned in each category are just top 10 each category. Please refer attached excel for total list of customers.

Knime Workflow





Recommendations:

- Golden customers are the ones because of whom, the company is surviving. We should retain them at any cost.
- Medium customers are the ones who can be enticed with offers and services and make them our golden customers.
- Low value customers are very hard to retain but we can reduce the conversion to Lost customers by giving basic offers only after planned expenditure is made on golden and Medium value customers.
- Very Low value customers are lost customers and we don't need to do anything to retain them.



Thank You