

MARKET RESEARCH ANALYSIS OF CUSTOMERS PROJECT

CONTENTS

- PROBLEM STATEMENT
- EDA
- RFM ANALYSIS
- RFM ANALYSIS
INFERENCES
- RECOMMENDATIONS

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. The aim is to provide them with suitable insights about their data and their customers.

Information of the Dataset

| # | 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 | | | |

The dataset has 2747 rows and 20 columns

ORDERDATE Column is a time series object, whereas the rest of the datatypes of the columns are either int or object data types.

Check Null Values or Duplicate Rows.

```
ORDERNUMBER          0  
QUANTITYORDERED    0  
PRICEEACH           0  
ORDERLINENUMBER     0  
SALES               0  
ORDERDATE            0  
DAYS_SINCE_LASTORDER 0  
STATUS               0  
PRODUCTLINE          0  
MSRP                 0  
PRODUCTCODE          0  
CUSTOMERNAME         0  
PHONE                0  
ADDRESSLINE1          0  
CITY                  0  
POSTALCODE            0  
COUNTRY               0  
CONTACTLASTNAME       0  
CONTACTFIRSTNAME      0  
DEALSIZE              0  
dtype: int64
```

Number of duplicate rows = 0

There are no null values present in the dataset and number of duplicate rows is none.

Descriptive Statistics summary

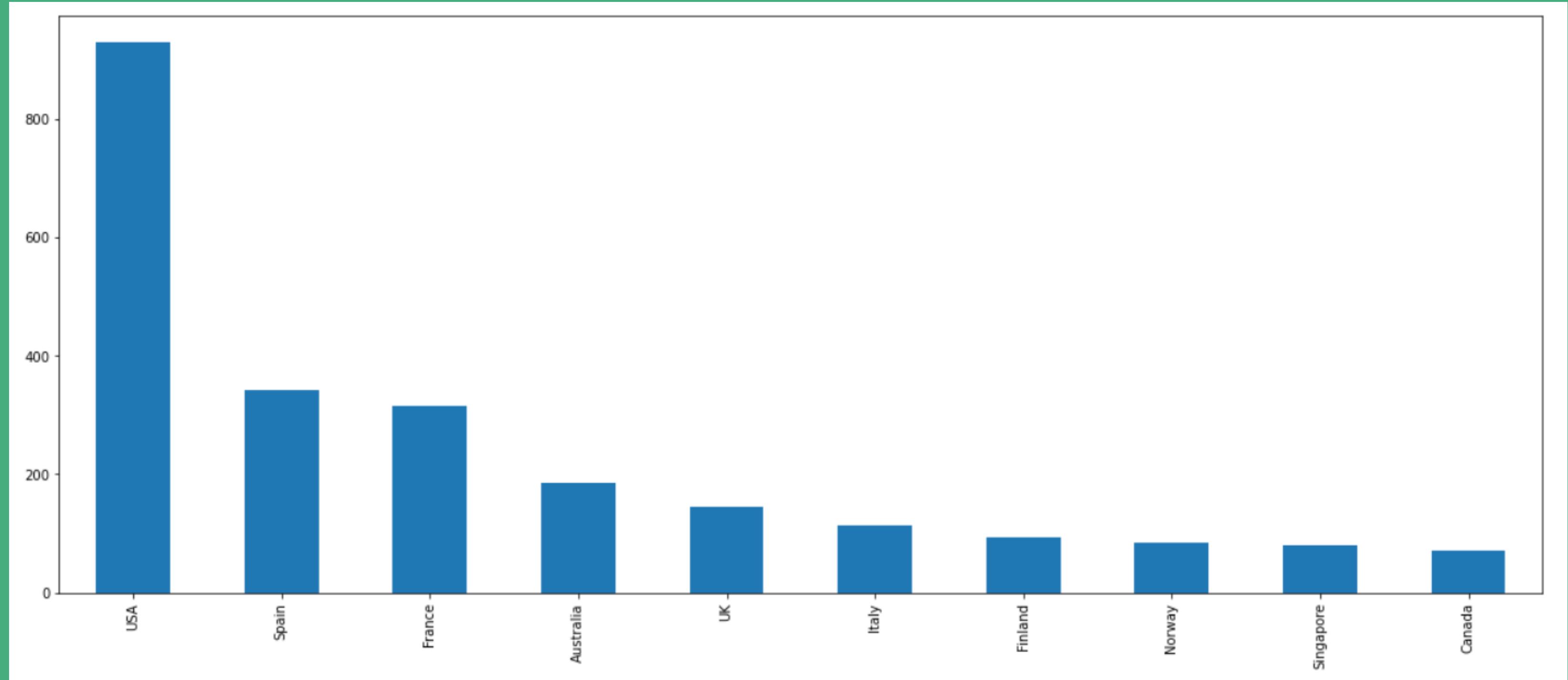
| | ORDERNUMBER | QUANTITYORDERED | PRICEEACH | ORDERLINENUMBER | SALES | DAYS_SINCE_LASTORDER | MSRP |
|--------------|--------------|-----------------|-------------|-----------------|--------------|----------------------|-------------|
| count | 2747.000000 | 2747.000000 | 2747.000000 | 2747.000000 | 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 |

■ The minimum and the maximum number of sales amounts are 482 and 14082 respectively. The average number of quantities ordered from the shop by a customer are 35 items.

■ The lowest and highest price of the items in the market are 26 and 97 respectively. As observed, the Market Retail price is higher than the sold items.

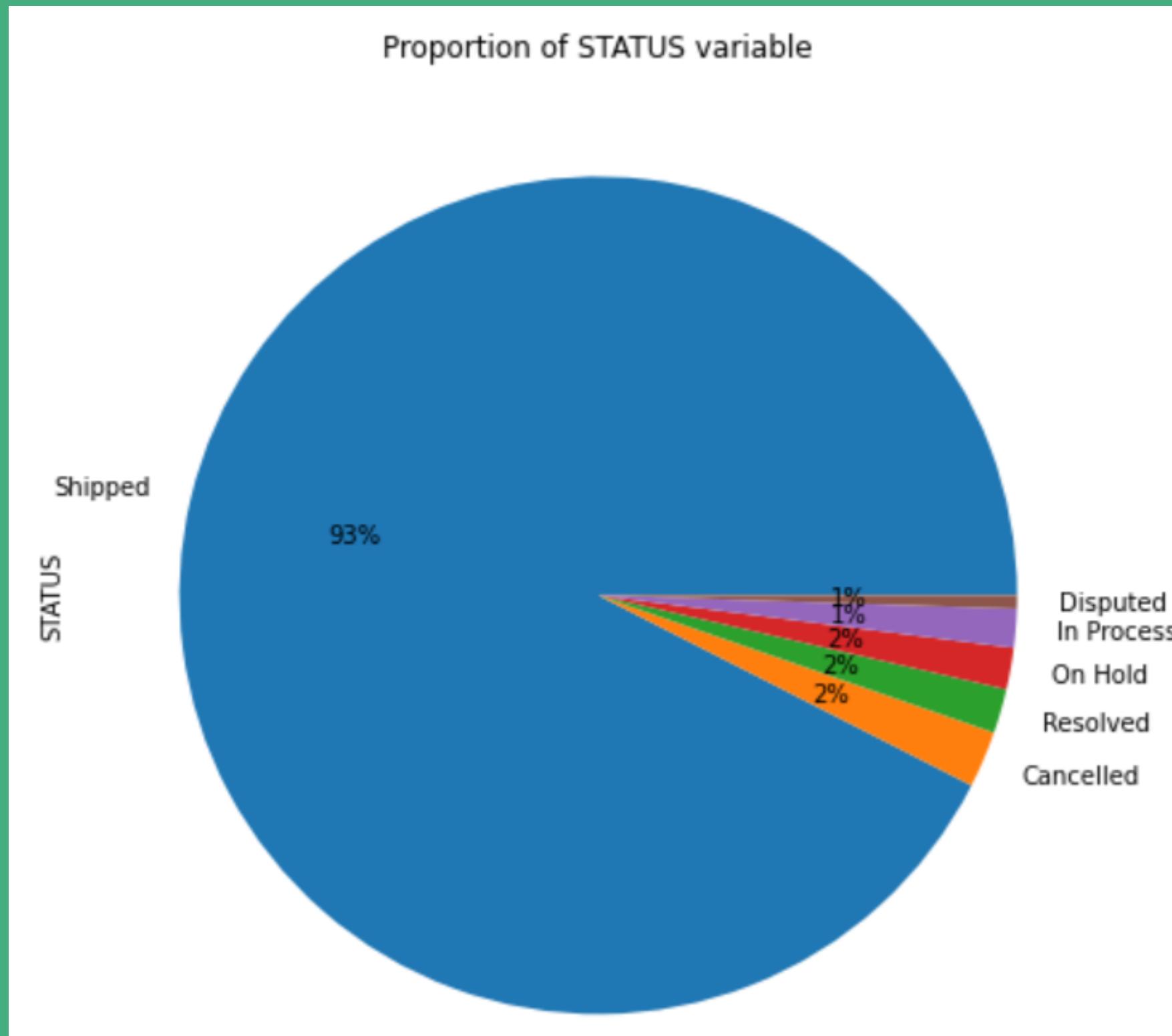
EXPLORATORY DATA ANALYSIS





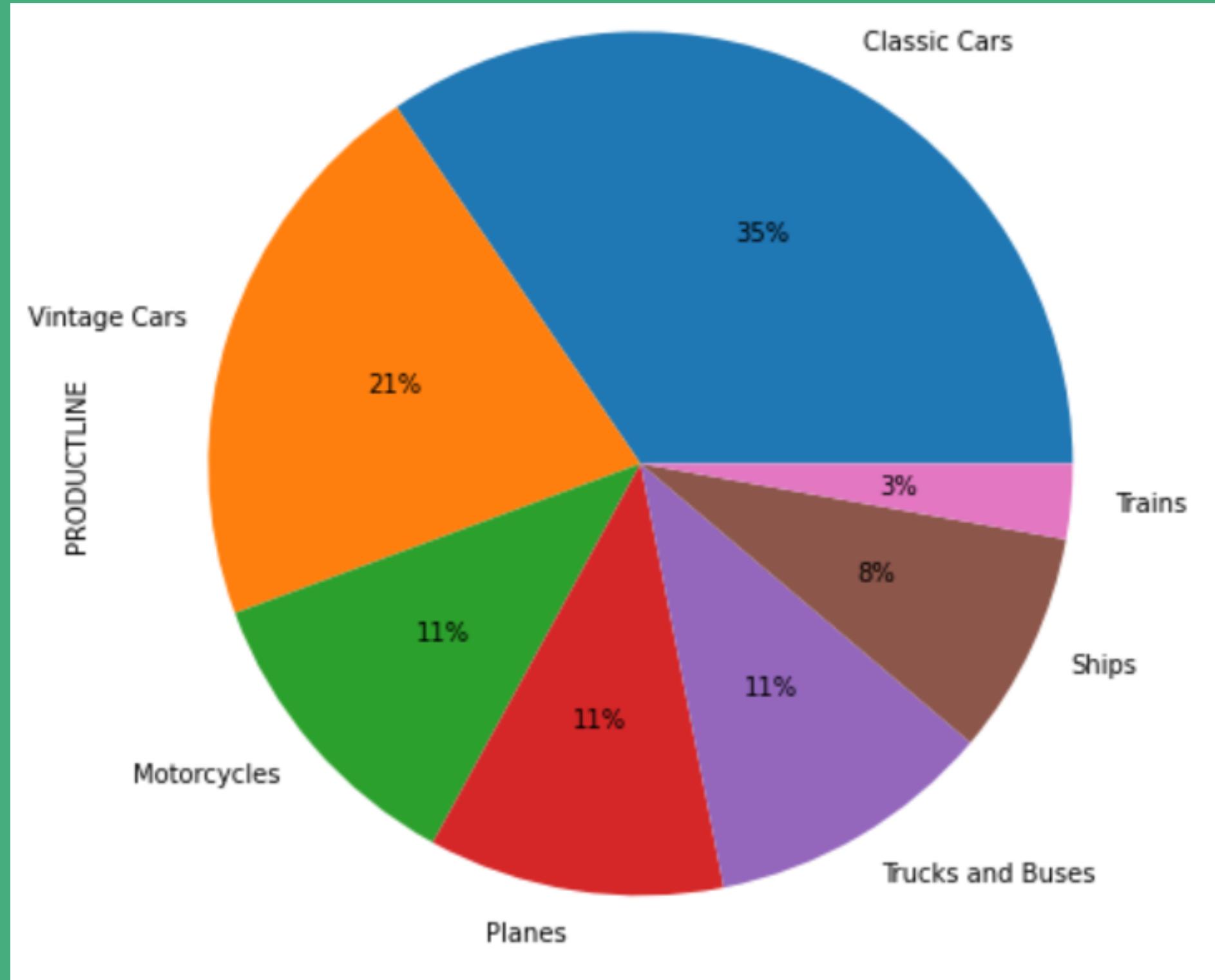
Large Number of customers are from USA and there are only few customers from Canada

ANALYSIS OF STATUS VARIABLE



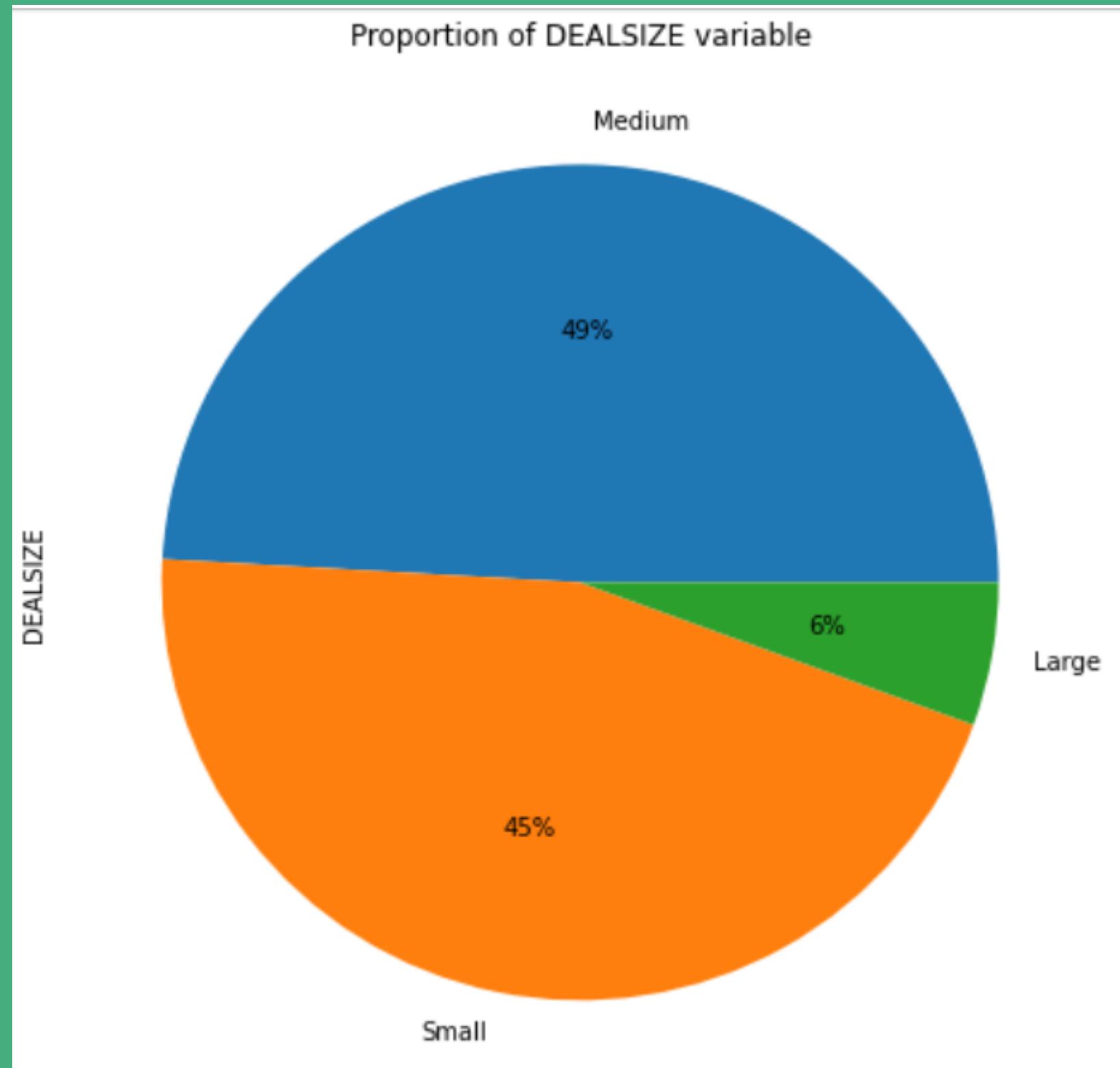
The total percentage of the shipped status of the orders is 93% which counts for a high percentage among other statuses of the orders.

ANALYSIS OF CATEGORIES VARIABLES



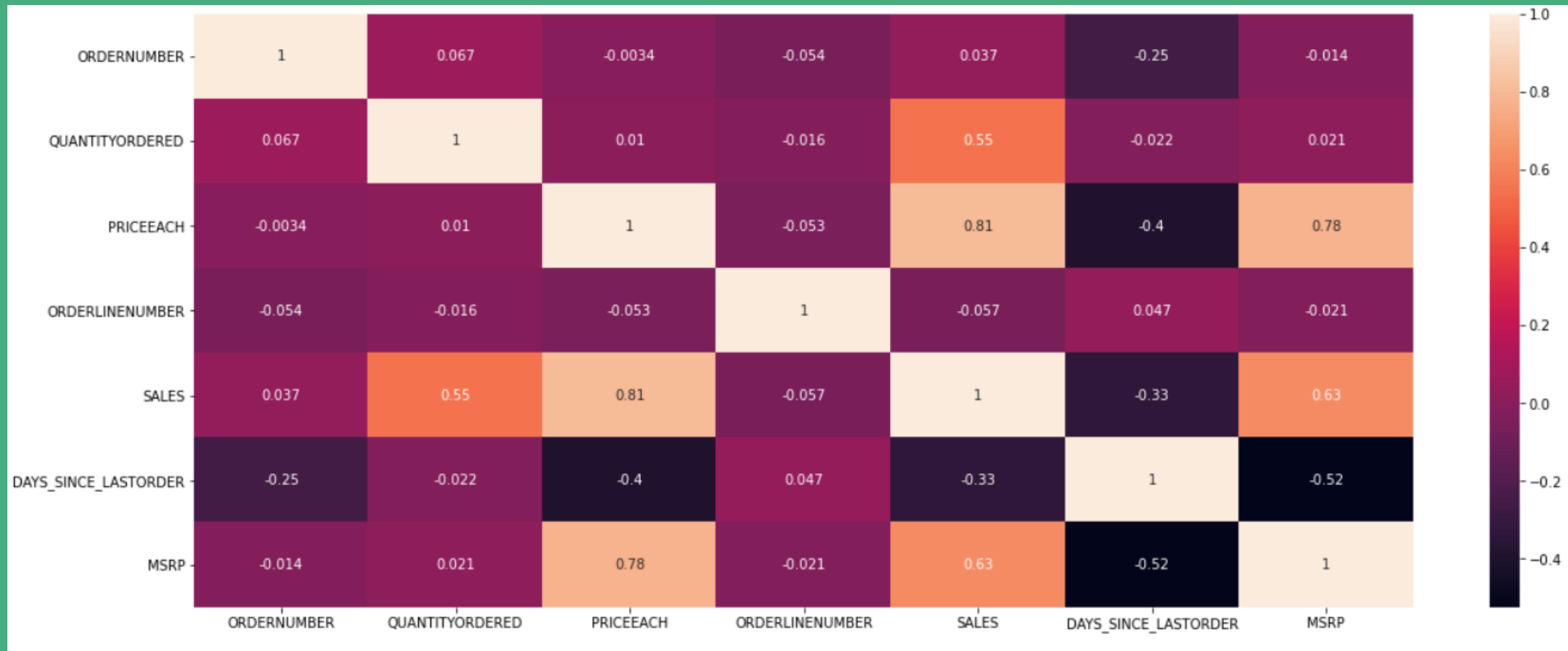
The most number of customers that buy automobile parts are for classic cars followed by vintage cars. The category that has least number of orders is ships

ANALYSIS OF CATEGORIES VARIABLES



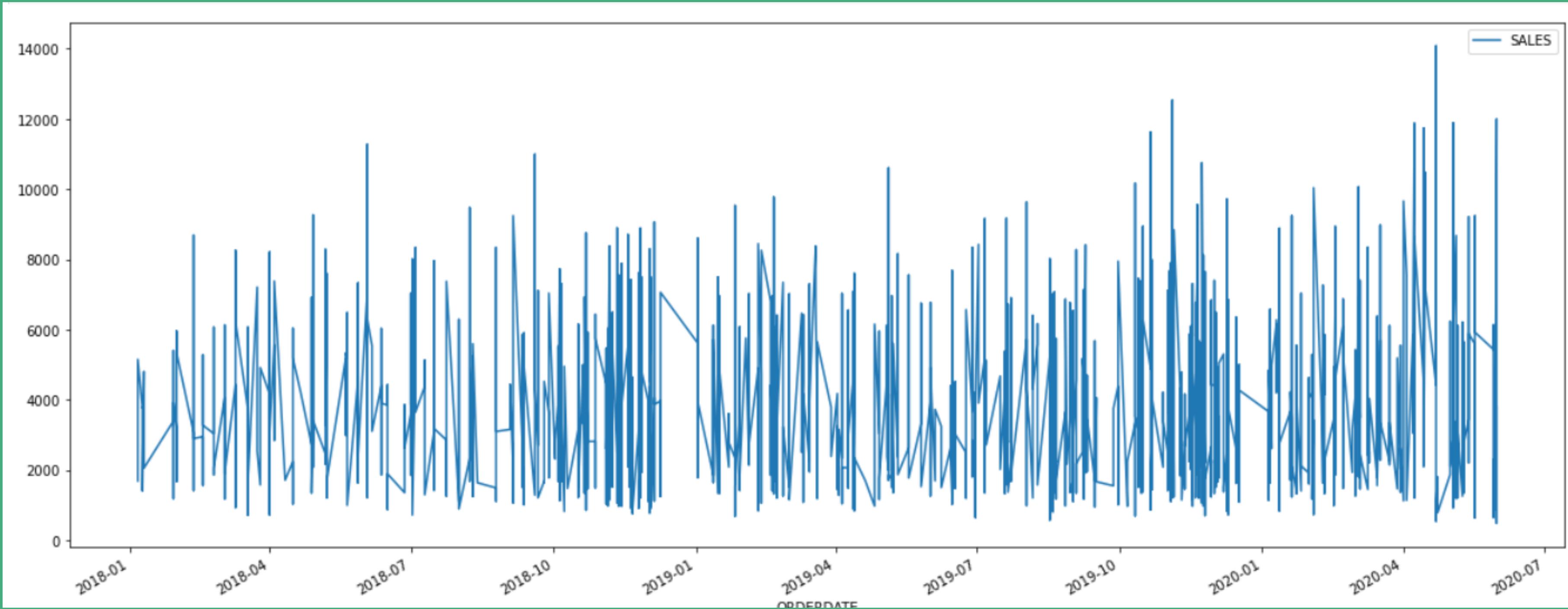
The size of the most deals are medium based on the Quantities and Item prices.

HEATMAP OF THE DATASET



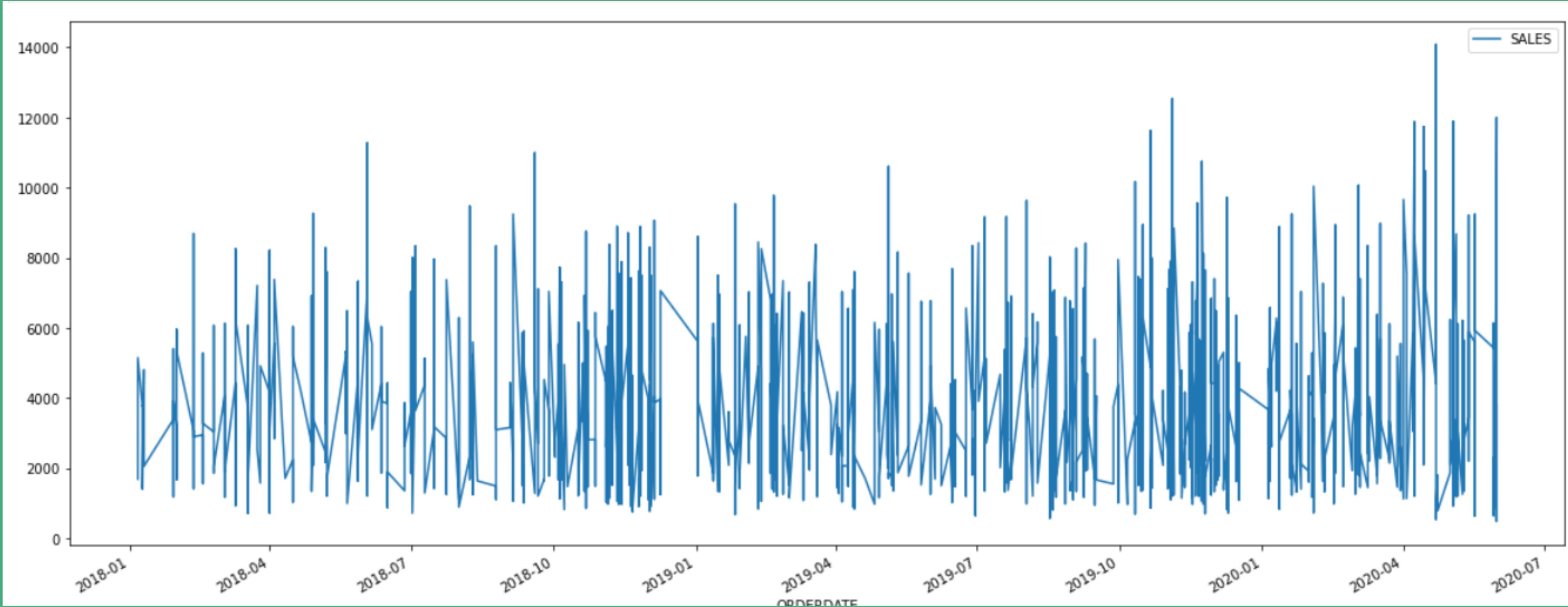
There is a high correlation between the price of each item and the sales amount.

QUARTERLY TRENDS IN SALES



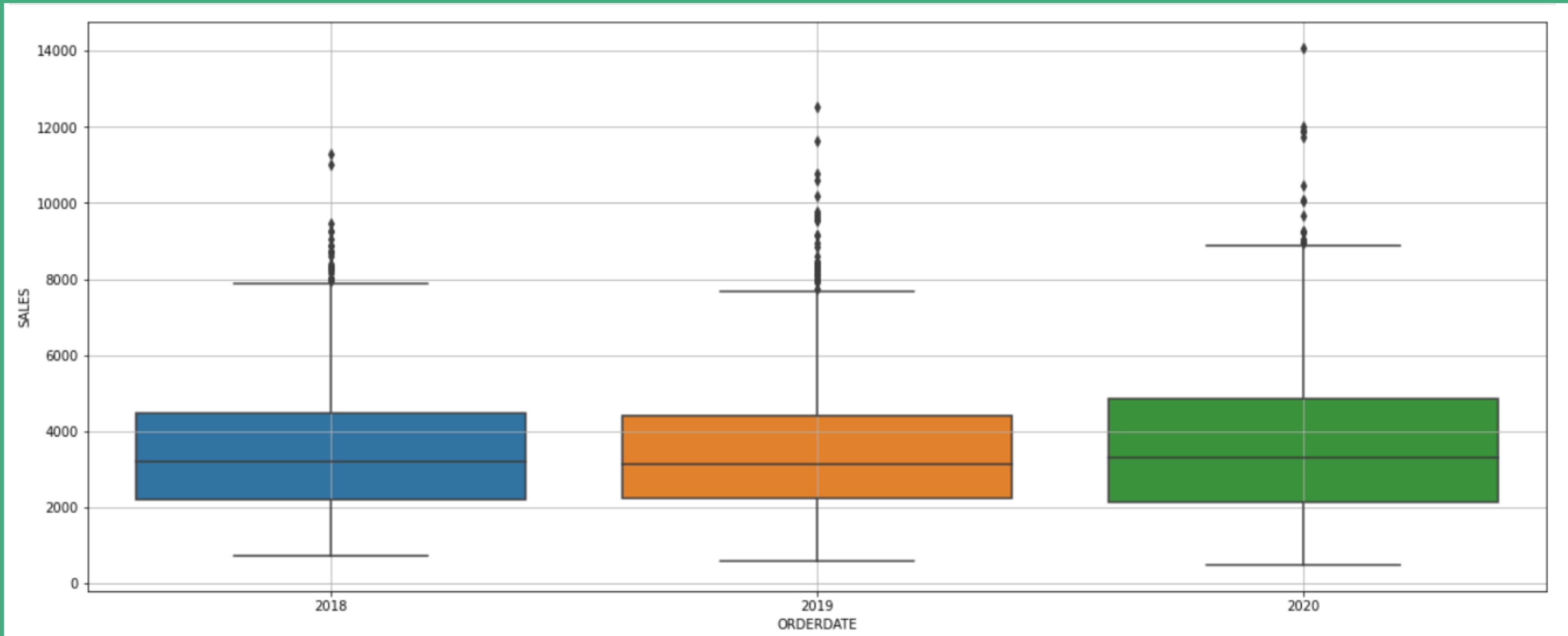
There is a high correlation between the price of each item and the sales amount.

QUARTERLY TRENDS IN SALES



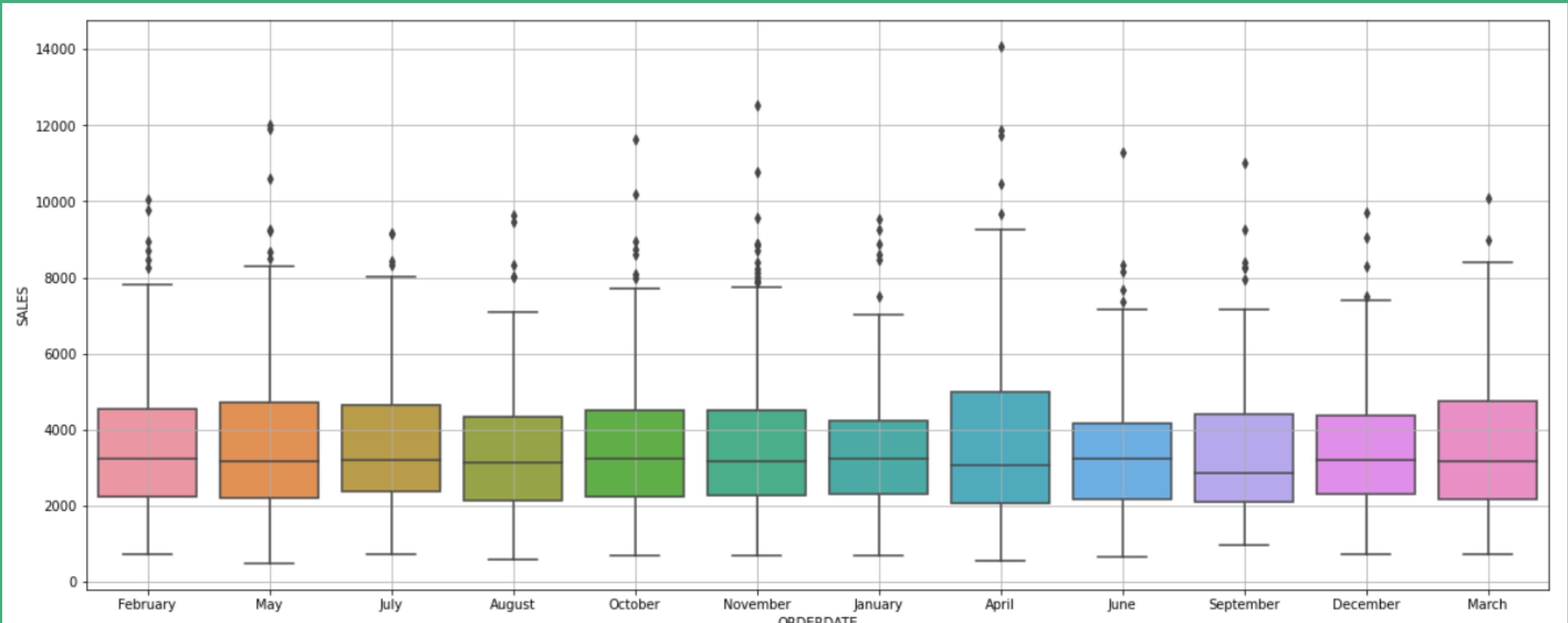
There is a high correlation between the price of each item and the sales amount.

Box plot of yearly trends in sales.



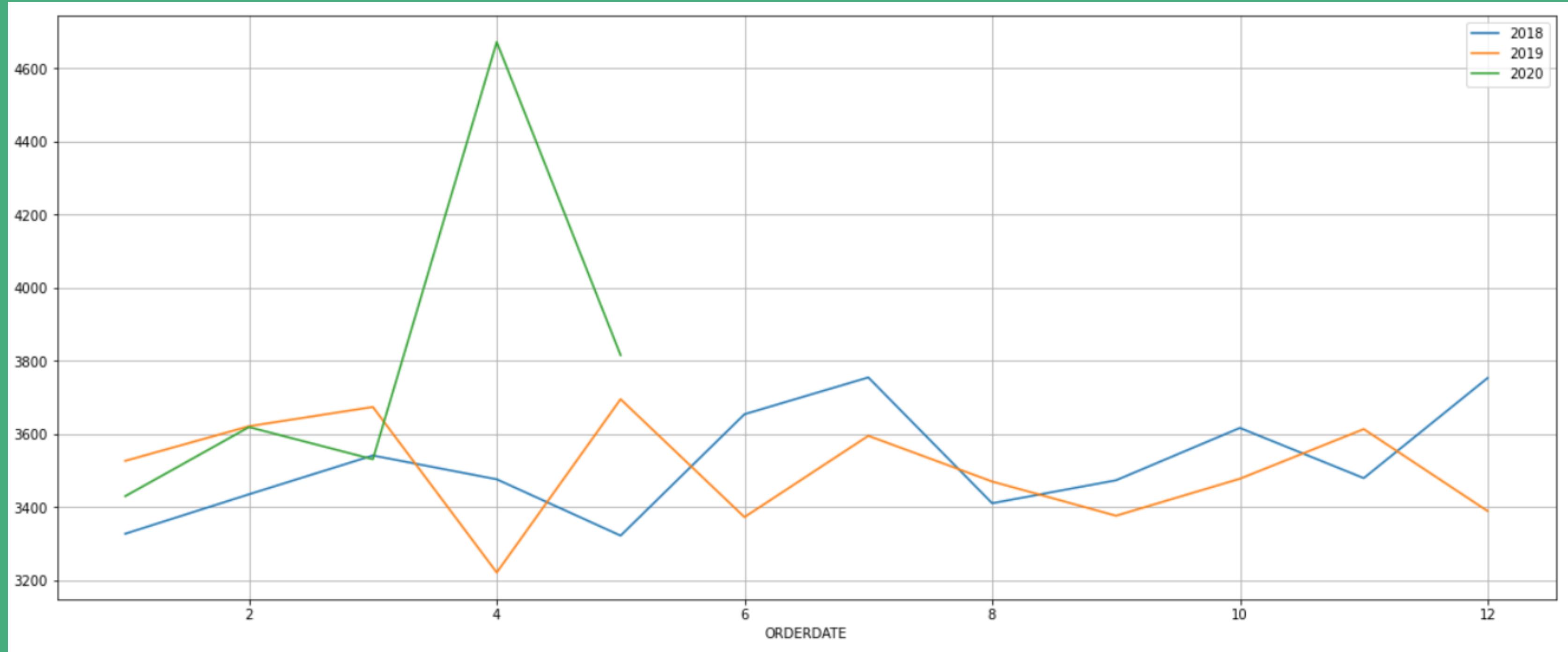
There are higher sales in the year 2020 and the least number of sales in the previous year.

Box plot of monthly trends in sales.

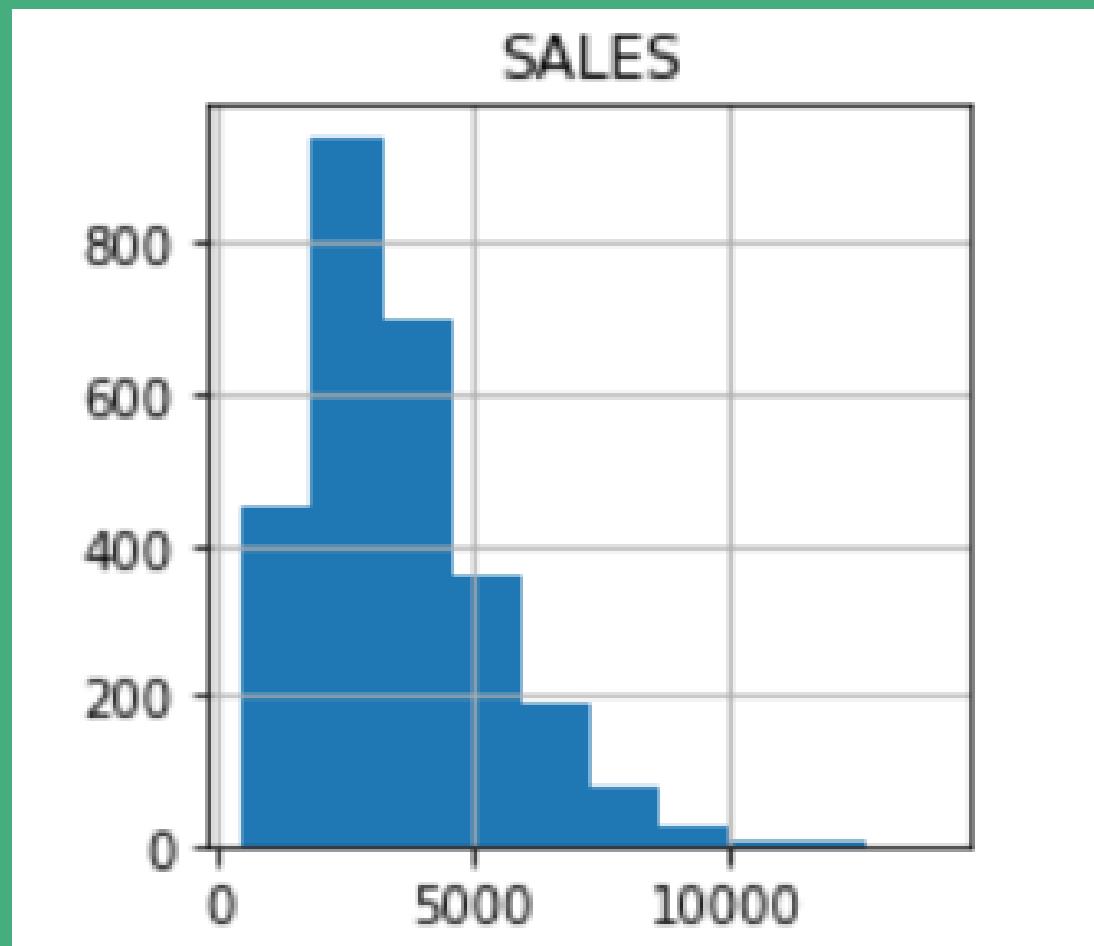
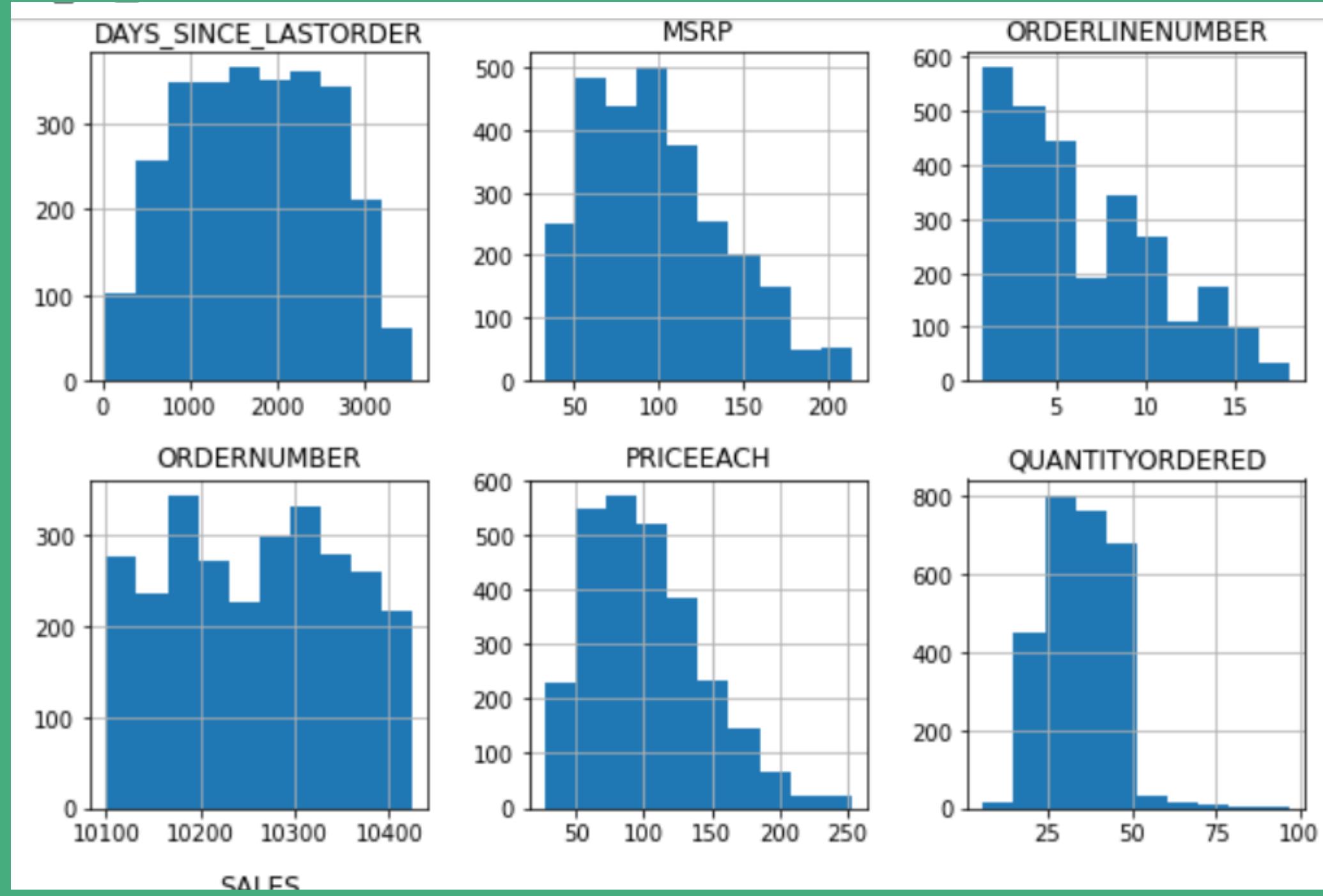


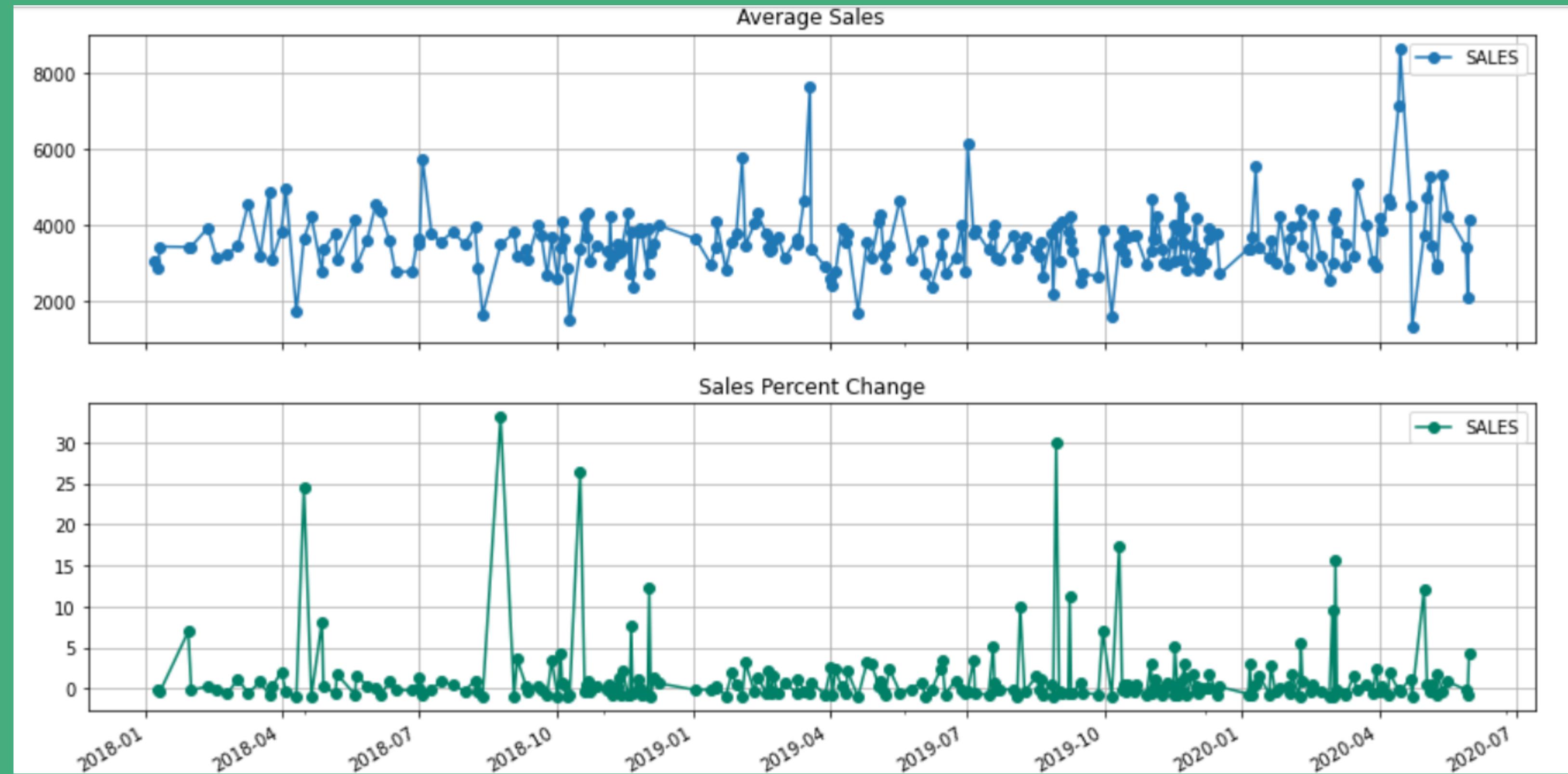
The highest number of sales amount in all the three years are in the month of April followed by March. January and August recorded the lowest sales amount.

Monthly sales across years



The monthly sales of 2018 and 2019 has been observed to fluctuate and there has been a peak sales amount in 2020.





CUSTOMER SEGMENTATION

USING RFM ANALYSIS

TOOL USED

RFM ANALYSIS

WHAT IS IT?

It combines and groups them into different customer segments for easy recall and campaign targeting. It is useful in understanding the responsiveness of your customers and for segmentation-driven database marketing.

VARIABLES USED

'CUSTOMERNAME' , 'ORDERNUMBER' , 'ORDERDATE' AND
'SALES'

OVERVIEW

It groups customers based on their transaction history. Recency: How recently did the customer purchase, Frequency: How often do they purchase? , Monetary: How much do they spend? Then Create RFM Table with respect to customer names. We have used the last order date as the "Now" date to calculate Recency. Then make four equal parts based on the quantile values to calculate RFM score.

RFM TABLE

| | CUSTOMERNAME | ORDERNUMBER | ORDERDATE | SALES |
|---|----------------------|-------------|------------|---------|
| 0 | Land of Toys Inc. | 10107 | 2018-02-24 | 2871.00 |
| 1 | Reims Collectables | 10121 | 2018-05-07 | 2765.90 |
| 2 | Lyon Souveniers | 10134 | 2018-07-01 | 3884.34 |
| 3 | Toys4GrownUps.com | 10145 | 2018-08-25 | 3746.70 |
| 4 | Technics Stores Inc. | 10168 | 2018-10-28 | 3479.76 |

Recency, Frequency and Monetary

1

Recency

Deduct the most recent purchase date from today to calculate the recency value.

2

Frequency

How many times has the customer purchased from our store?

3

Monetary

How much amount has this customer spent? Total up the money from all transactions to get the M value.

Calculation of RFM Score

Customer segments from RFM Model can be created by using Quartiles. Assign a score from 1 to 4 to Recency, Frequency, and Monetary. Four is the best/highest value, and one is the lowest/worst value. A final RFM score is calculated simply by combining individual RFM score numbers.

| CUSTOMERNAME | Recency | Frequency | Monetary | R_Quartile | F_Quartile | M_Quartile | RFMScore |
|-------------------------|---------|-----------|-----------|------------|------------|------------|----------|
| AV Stores, Co. | 196 | 51 | 157807.81 | 2 | 4 | 4 | 244 |
| Alpha Cognac | 64 | 20 | 70488.44 | 4 | 1 | 1 | 411 |
| Amica Models & Co. | 265 | 26 | 94117.26 | 1 | 2 | 3 | 123 |
| Anna's Decorations, Ltd | 83 | 46 | 153996.13 | 3 | 4 | 4 | 344 |
| Atelier graphique | 188 | 7 | 24179.96 | 2 | 1 | 1 | 211 |

INFERENCES FROM RFM ANALYSIS

and Identified Segments.

Q1 . Who are the best Customers ?

| CUSTOMERNAME | Recency | Frequency | Monetary | R_Quartile | F_Quartile | M_Quartile | RFMScore |
|------------------------------|---------|-----------|-----------|------------|------------|------------|----------|
| Euro Shopping Channel | 0 | 259 | 912294.11 | 4 | 4 | 4 | 444 |
| Mini Gifts Distributors Ltd. | 2 | 180 | 654858.06 | 4 | 4 | 4 | 444 |
| La Rochelle Gifts | 0 | 53 | 180124.90 | 4 | 4 | 4 | 444 |
| The Sharp Gifts Warehouse | 39 | 40 | 160010.27 | 4 | 4 | 4 | 444 |
| Souveniers And Things Co. | 2 | 46 | 151570.98 | 4 | 4 | 4 | 444 |

Condition applied: When "RFM Score = 444" . The above output screenshot shows the top 6 the best customers

Q2. Which customers are on the verge of churning?

| | | Recency | Frequency | Monetary | R_Quartile | F_Quartile | M_Quartile | RFMScore |
|--------------|-------------------------------------|---------|-----------|-----------|------------|------------|------------|----------|
| CUSTOMERNAME | | | | | | | | |
| | Land of Toys Inc. | 198 | 49 | 164069.44 | 2 | 4 | 4 | 244 |
| | AV Stores, Co. | 196 | 51 | 157807.81 | 2 | 4 | 4 | 244 |
| | Saveley & Henriot, Co. | 456 | 41 | 142874.25 | 1 | 4 | 4 | 144 |
| | Rovelli Gifts | 201 | 48 | 137955.72 | 2 | 4 | 4 | 244 |
| | Online Diecast Creations Co. | 209 | 34 | 131685.30 | 2 | 4 | 4 | 244 |
| | Corrida Auto Replicas, Ltd | 212 | 32 | 120615.28 | 2 | 3 | 4 | 234 |

Condition applied: When "Recency value is low >2". The above output screenshot shows the top 6 customers that are on the verge of churning.

Q3. Who are the lost Customers ?

| CUSTOMERNAME | Recency | Frequency | Monetary | R_Quartile | F_Quartile | M_Quartile | RFMScore |
|--------------------------------|---------|-----------|----------|------------|------------|------------|----------|
| Double Decker Gift Stores, Ltd | 495 | 12 | 36019.04 | 1 | 1 | 1 | 111 |
| West Coast Collectables Co. | 488 | 13 | 46084.64 | 1 | 1 | 1 | 111 |
| Signal Collectibles Ltd. | 476 | 15 | 50218.51 | 1 | 1 | 1 | 111 |
| Daedalus Designs Imports | 465 | 20 | 69052.41 | 1 | 1 | 1 | 111 |
| CAF Imports | 439 | 13 | 49642.05 | 1 | 1 | 1 | 111 |
| Osaka Souveniers Co. | 414 | 20 | 67605.07 | 1 | 1 | 1 | 111 |

Condition applied: When "RFM Score = 111". The above output screenshot shows the top 6 lost customers.

Q4. Who are your loyal Customers ?

| CUSTOMERNAME | Recency | Frequency | Monetary | R_Quartile | F_Quartile | M_Quartile | RFMScore |
|------------------------------|---------|-----------|-----------|------------|------------|------------|----------|
| Euro Shopping Channel | 0 | 259 | 912294.11 | 4 | 4 | 4 | 444 |
| Mini Gifts Distributors Ltd. | 2 | 180 | 654858.06 | 4 | 4 | 4 | 444 |
| La Rochelle Gifts | 0 | 53 | 180124.90 | 4 | 4 | 4 | 444 |
| The Sharp Gifts Warehouse | 39 | 40 | 160010.27 | 4 | 4 | 4 | 444 |
| Souveniers And Things Co. | 2 | 46 | 151570.98 | 4 | 4 | 4 | 444 |

Condition applied: When "Frequency value is more than 3".
The above output screenshot shows the top 6 loyal customers.

Additional Insights

| CUSTOMERNAME | Recency | Frequency | Monetary | R_Quartile | F_Quartile | M_Quartile | RFMScore |
|------------------------------|---------|-----------|-----------|------------|------------|------------|----------|
| Euro Shopping Channel | 0 | 259 | 912294.11 | 4 | 4 | 4 | 444 |
| Mini Gifts Distributors Ltd. | 2 | 180 | 654858.06 | 4 | 4 | 4 | 444 |
| Australian Collectors, Co. | 184 | 55 | 200995.41 | 3 | 4 | 4 | 344 |
| Muscle Machine Inc | 182 | 48 | 197736.94 | 3 | 4 | 4 | 344 |
| La Rochelle Gifts | 0 | 53 | 180124.90 | 4 | 4 | 4 | 444 |

Condition applied: When "Monetary quartile value is equal to 4". The above output screenshot shows the top 6 customers who spend more money.

RECOMMENDATIONS

Best customers are most likely to buy new products, so as the best customers of the company are already identified, you can reward them and may be start a "Refer a friend" program.

Customers who are on the verge of churning can be approached by sending personalized emails to shop and sending emails regarding latest prices and products.