

# Customer Segmentation Analysis Report

## MRA Project Part-A

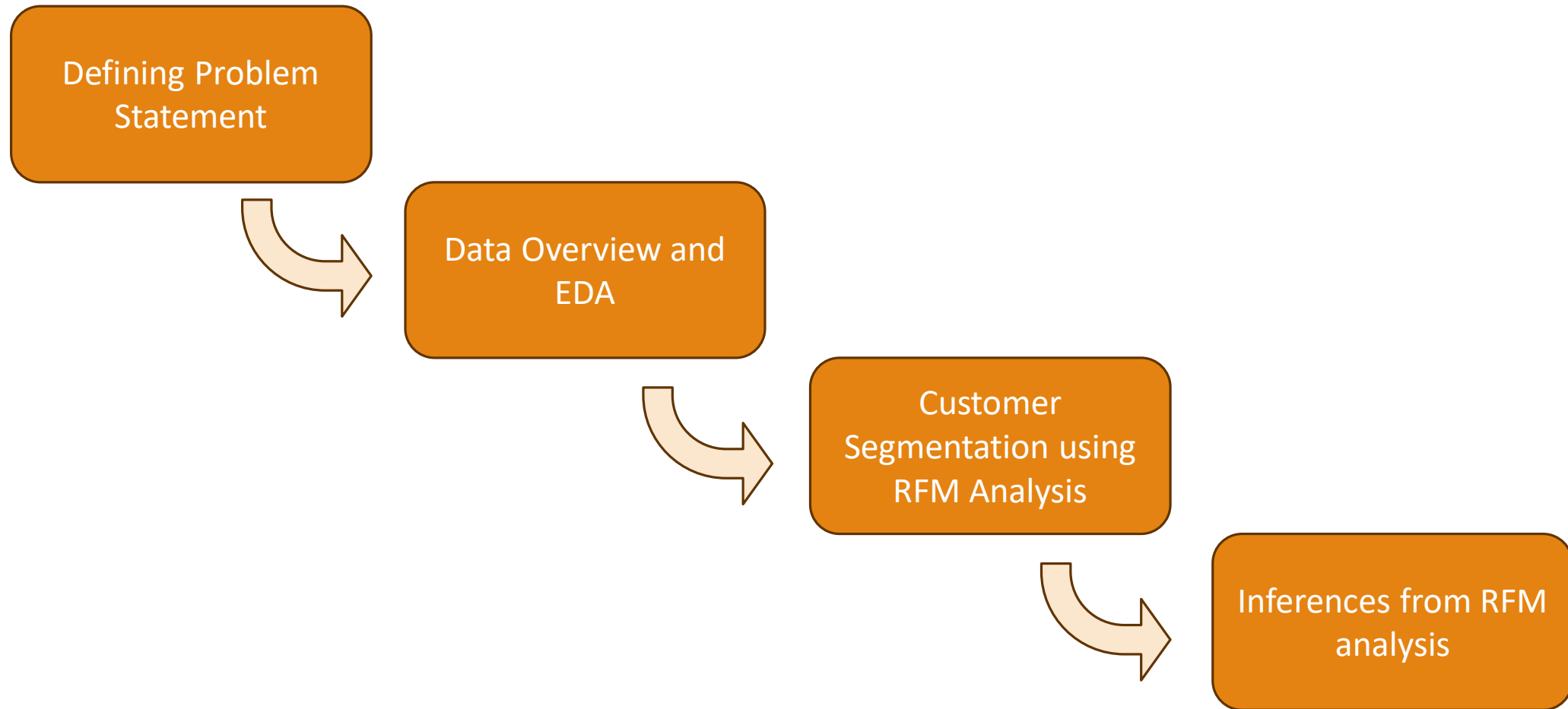
---

Sanjay Rajan J

Batch: PGP-DSBA (PGPDSBA.O.JULY24.A)

sanjaysujana@gmail.com

# AGENDA



# PROBLEM STATEMENT:

## **Business Context:**

An automobile parts manufacturing company has been actively selling products to a diverse range of customers for the past three years. Despite its growth, the company lacks the in-house expertise to derive actionable insights from its transaction data. As a result, they wish to uncover hidden patterns and trends in their customer transactions. By analysing this data, the company aims to better understand customer behaviour, improve customer segmentation, and implement targeted marketing strategies. These insights will help the company not only enhance customer satisfaction but also drive revenue growth by offering more personalized and efficient services.

## **Objective:**

The primary objective of this analysis is to leverage data science techniques to:

1. Identify underlying patterns in customer purchasing behaviour.
2. Segment customers based on their transactional data.
3. Provide actionable insights to optimize the company's marketing efforts.
4. Recommend personalized marketing strategies for each customer segment to maximize sales and customer retention.

# Summary Statistics:

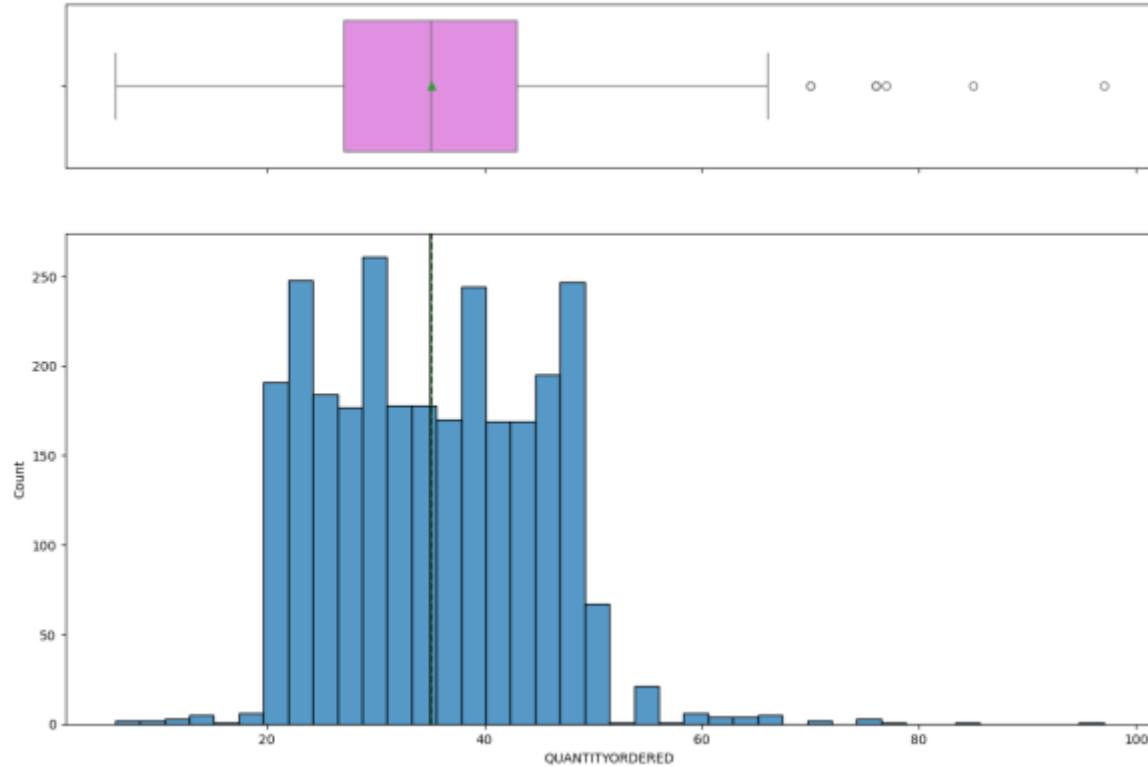
	count	mean	min	25%	50%	75%	max	std
ORDERNUMBER	2747.0	10259.761558	10100.0	10181.0	10264.0	10334.5	10425.0	91.877521
QUANTITYORDERED	2747.0	35.103021	6.0	27.0	35.0	43.0	97.0	9.762135
PRICEEACH	2747.0	101.098951	26.88	68.745	95.55	127.1	252.87	42.042548
ORDERLINENUMBER	2747.0	6.491081	1.0	3.0	6.0	9.0	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

```
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

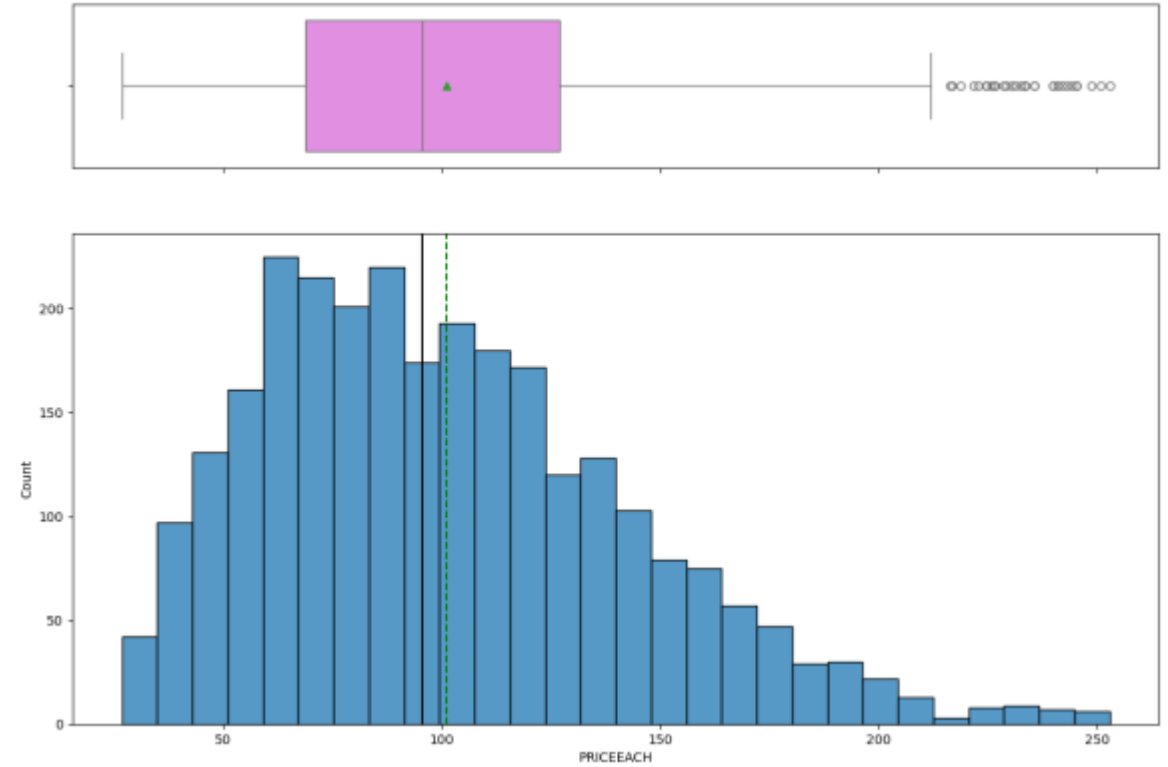
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 2747 entries, 0 to 2746
Data columns (total 20 columns):
#   Column                Non-Null Count  Dtype
---  -
0   ORDERNUMBER           2747 non-null  int64
1   QUANTITYORDERED       2747 non-null  int64
2   PRICEEACH             2747 non-null  float64
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
```

# Univariate Analysis:

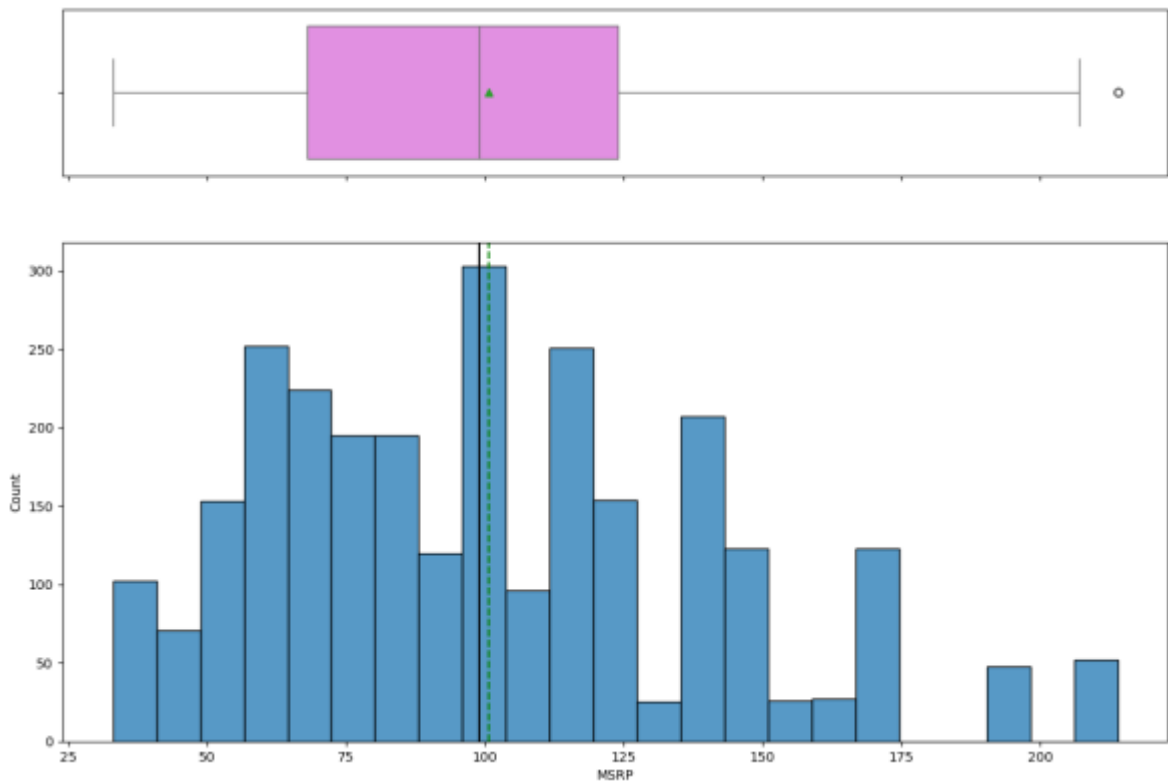
DISTRIBUTION OF QUANTITYORDERED



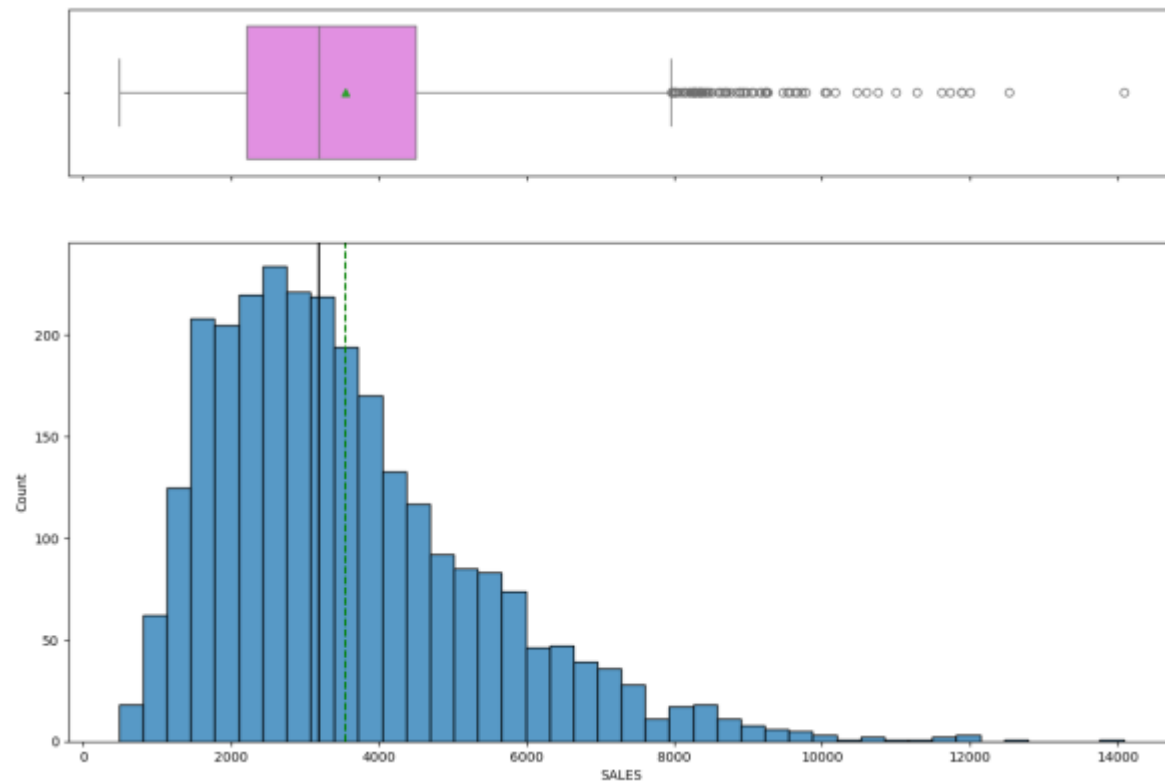
DISTRIBUTION OF PRICEEACH



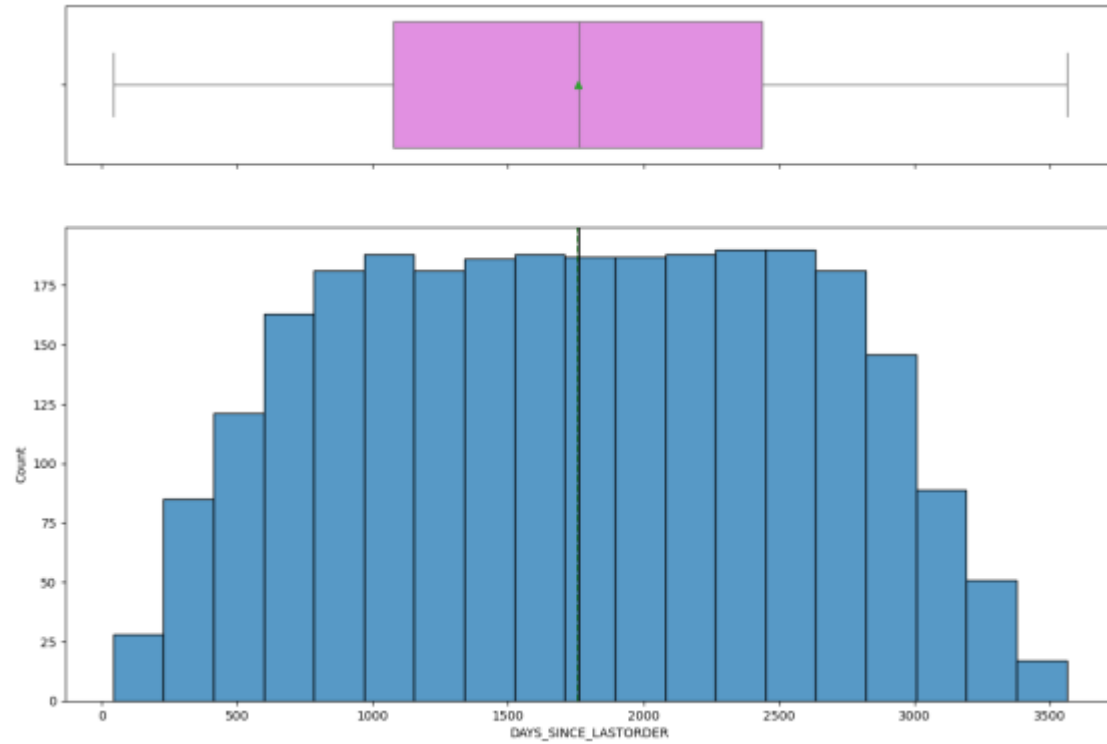
# DISTRIBUTION OF MSRP



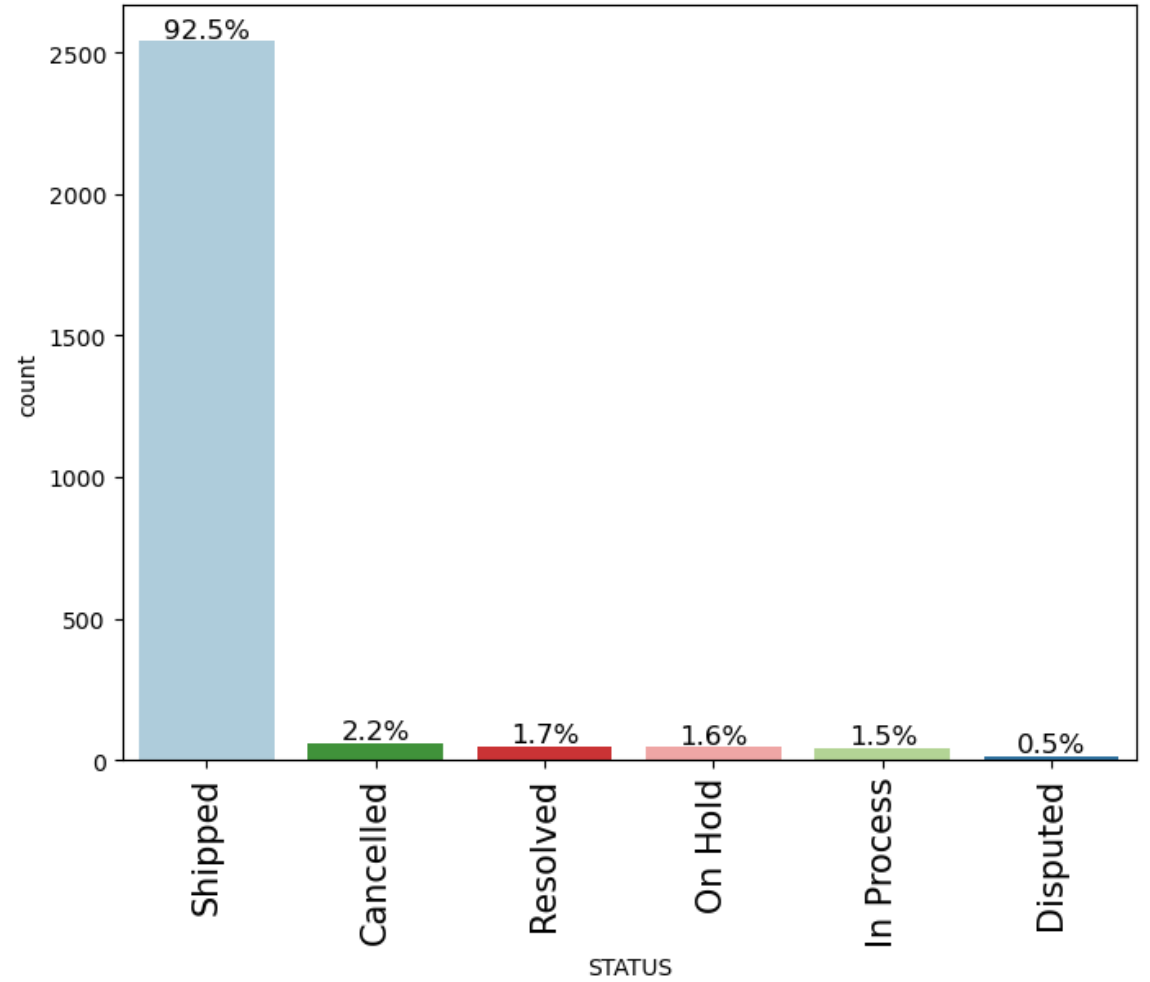
# DISTRIBUTION OF SALES



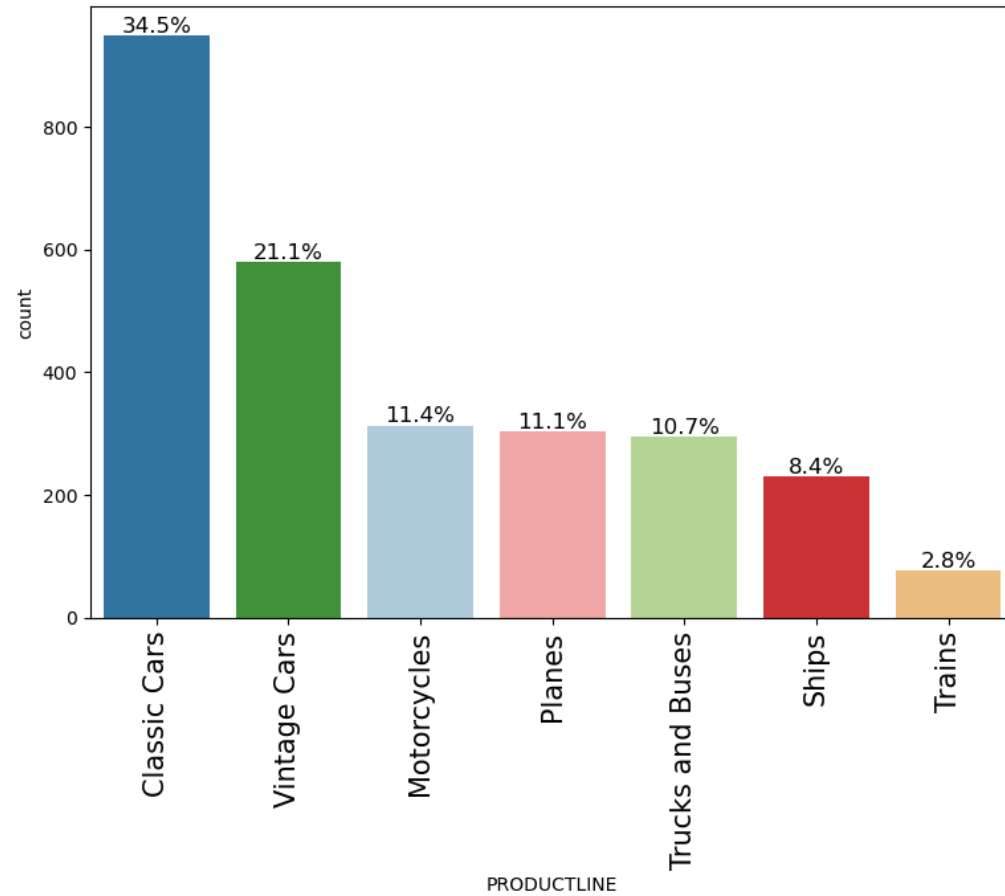
# DISTRIBUTION OF DAYS\_SINCE\_LASTORDER



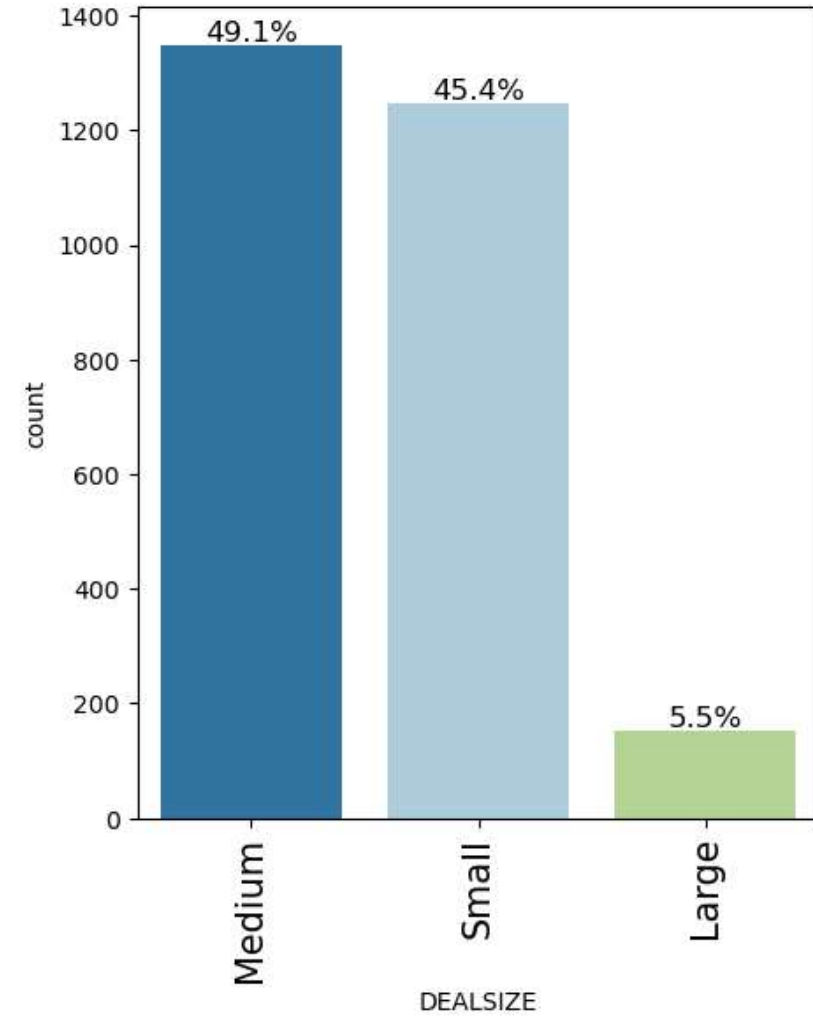
# DISTRIBUTION OF STATUS



DISTRIBUTION OF PRODUCTLINE

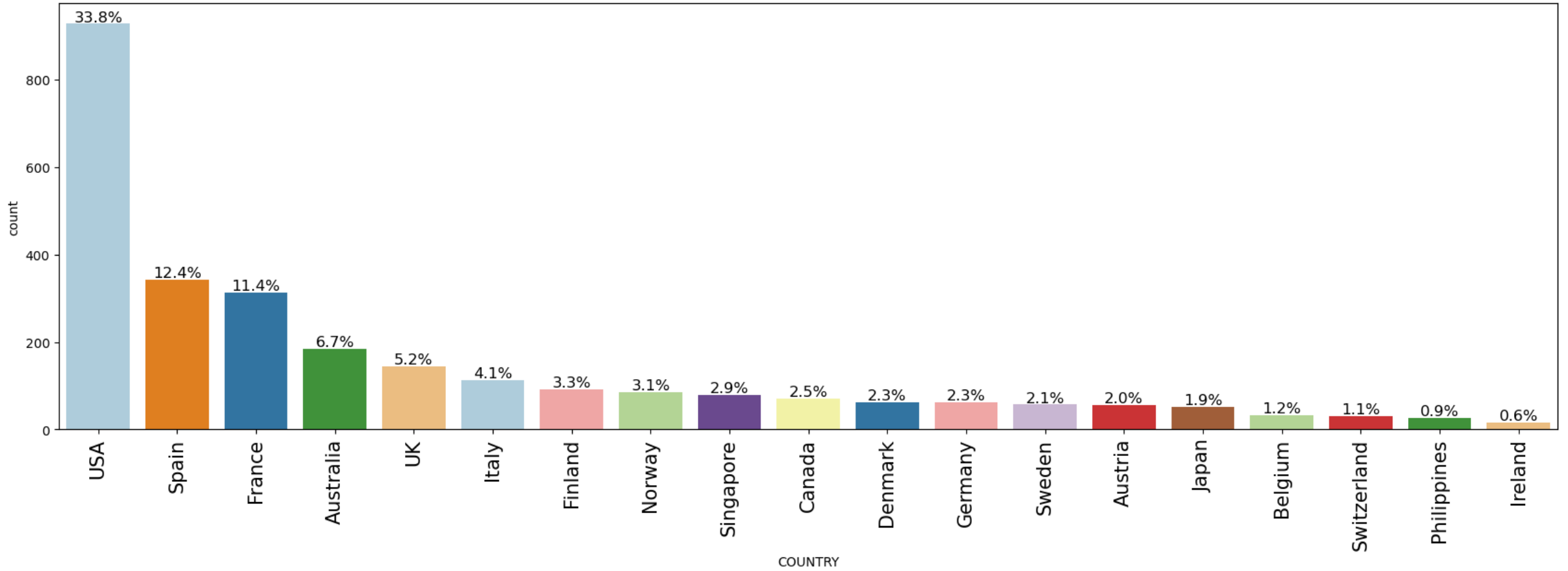


DISTRIBUTION OF DEALSIZE

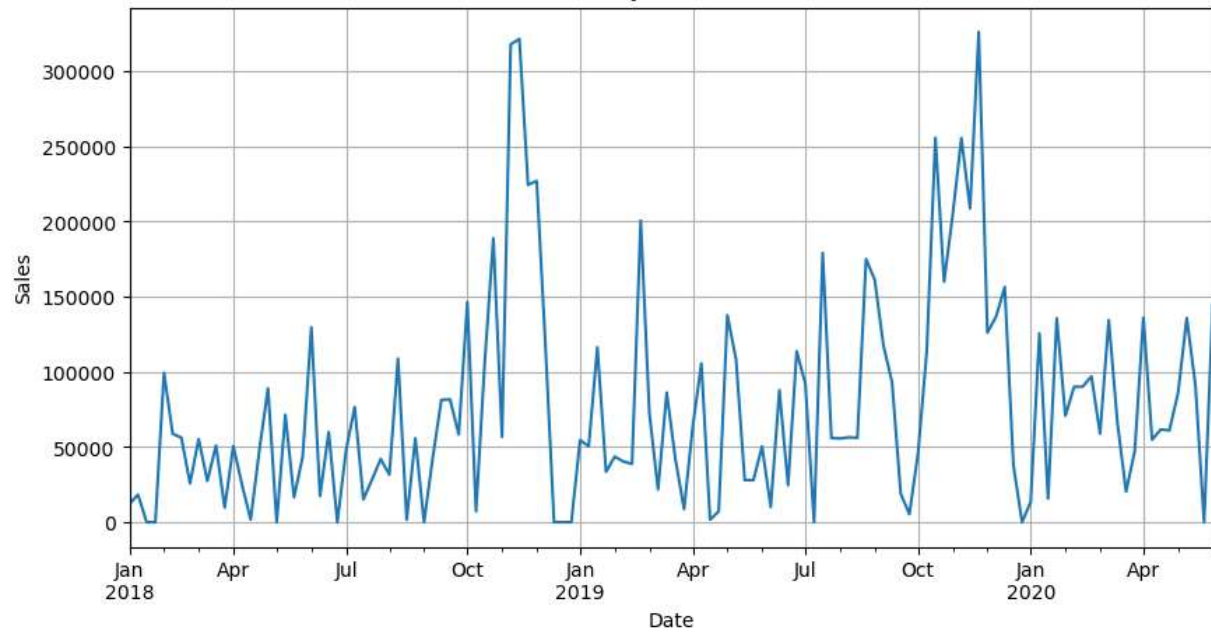




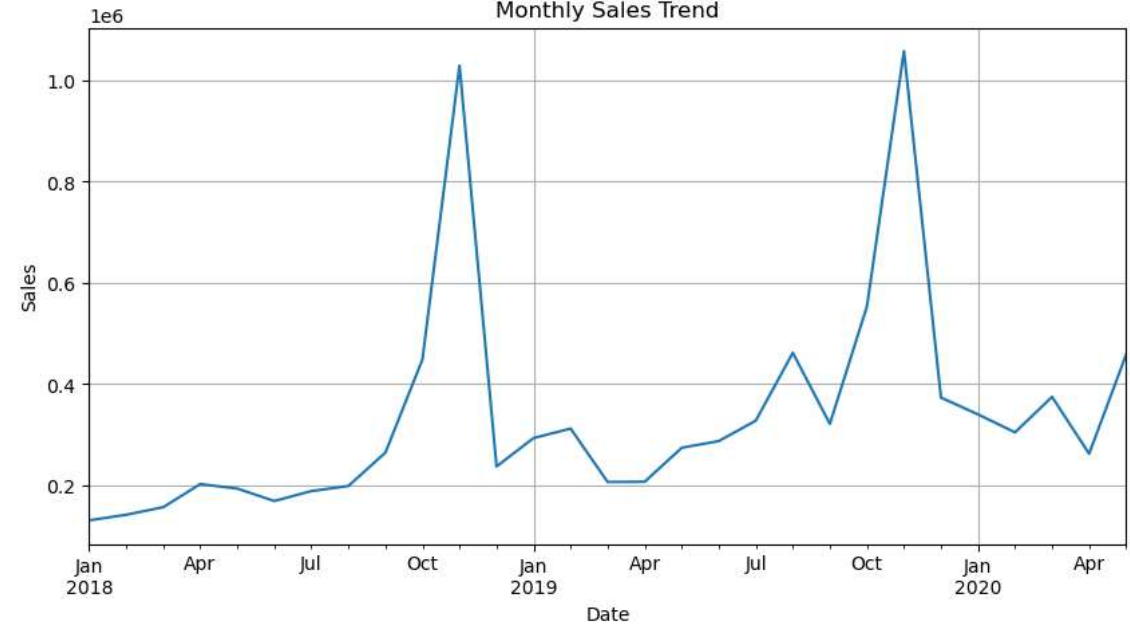
## DISTRIBUTION OF COUNTRY



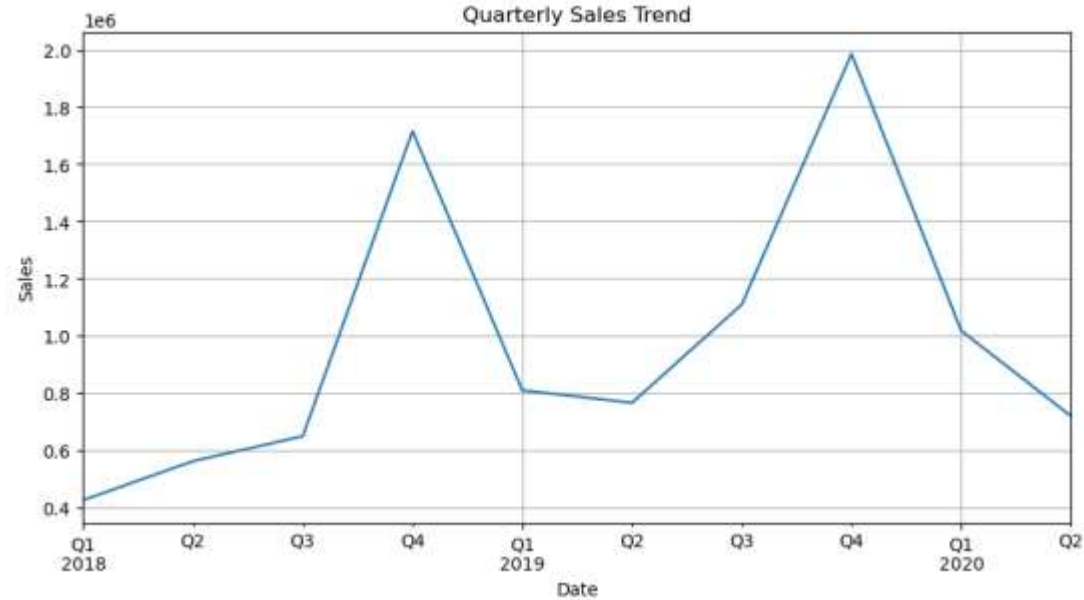
Weekly Sales Trend



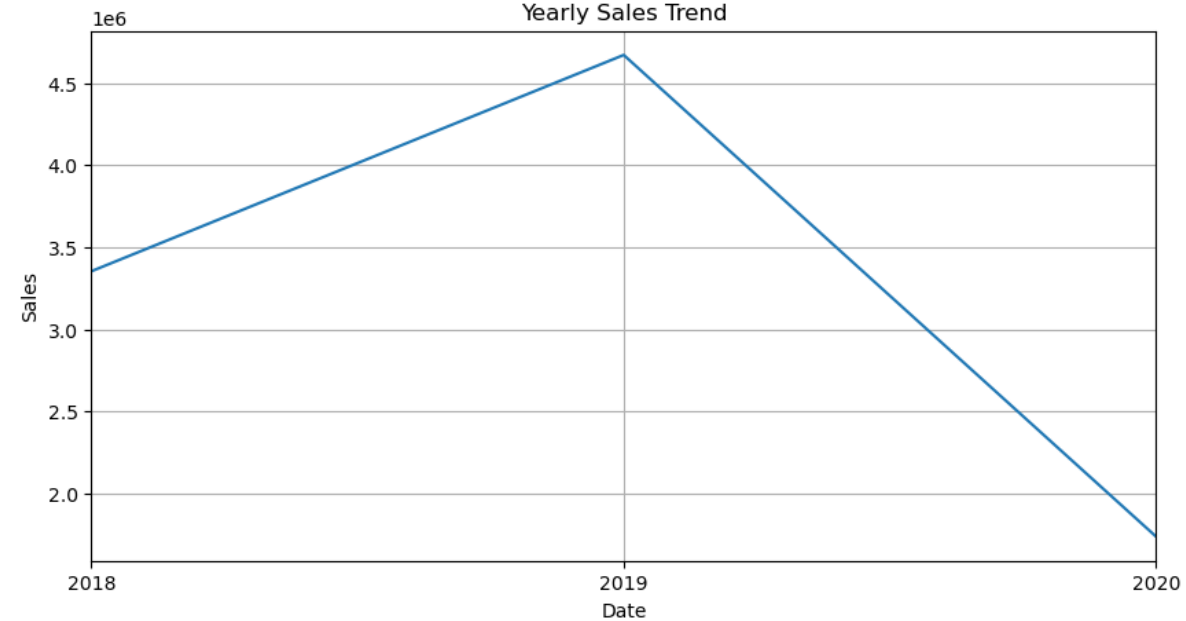
Monthly Sales Trend



Quarterly Sales Trend



Yearly Sales Trend

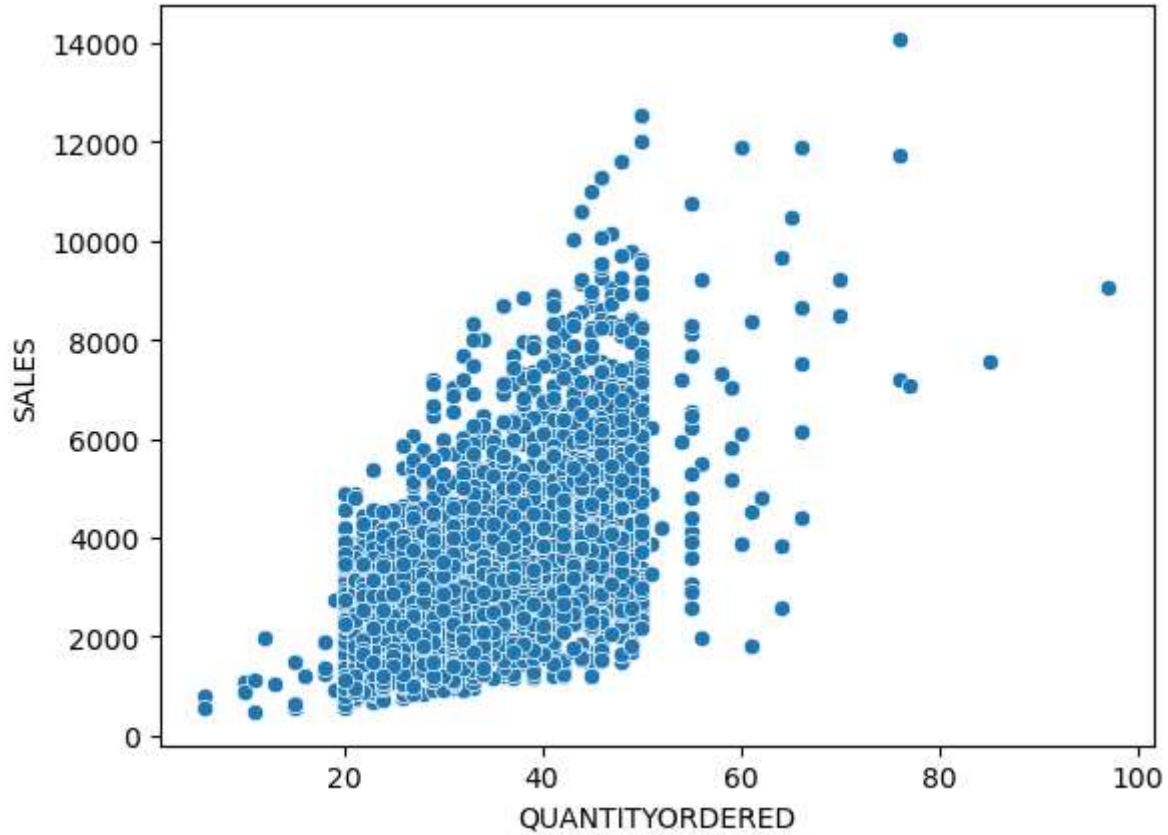


## Inferences:

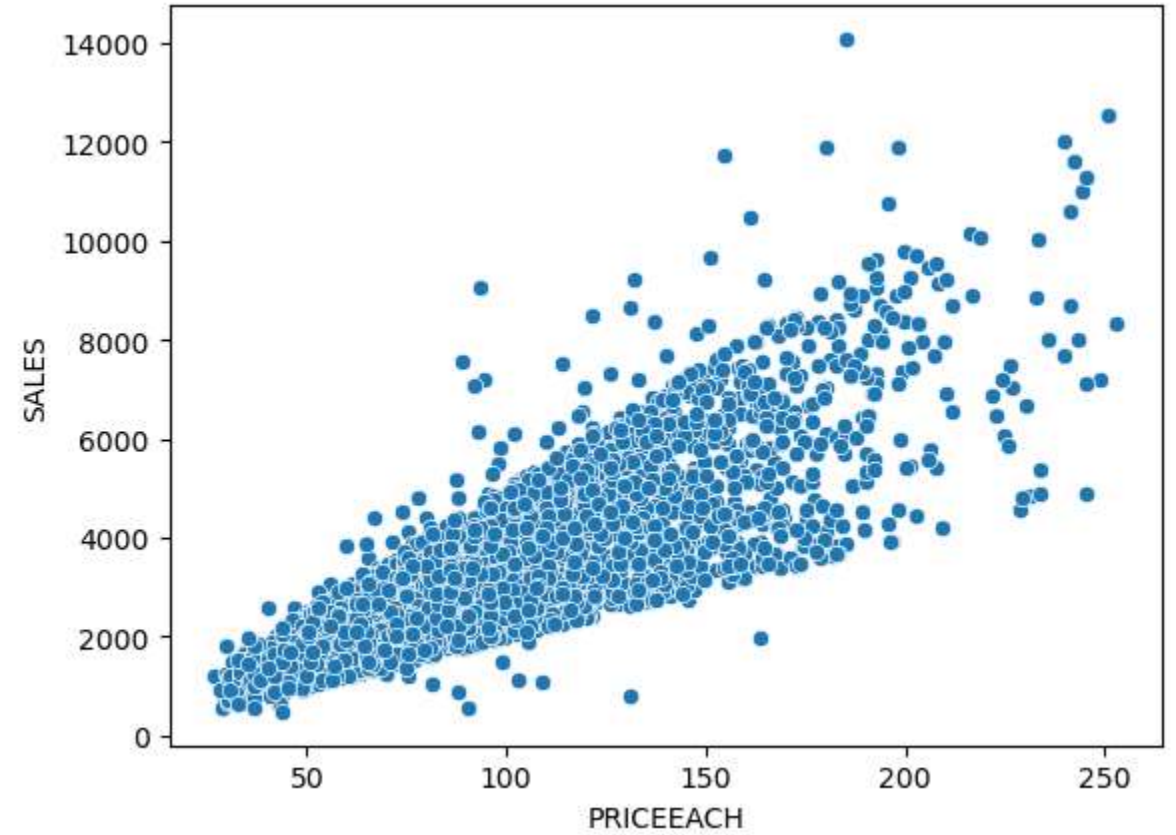
- **“Classic Car type”** sales clearly dominates the market by 34.5% share, when compared to other product line. It is followed by **“Vintage Cars”** with 21.1%
- **Shipped** Status dominates the STATUS field by 92.5%.
- Other Status types are very minimal in comparison.
- **USA** clearly dominates the market share with a huge 33.8%, indicating the largest market.
- Top 5 Countries: USA(33.8%), Spain(12.4%), France(11.4%), Australia(6.7%), UK(5.2%)
- **Q4** significantly shows the highest sales in each year, which indicates the year end performance (Holiday season).
- The year **2019** has the highest sales and it is dropped in the next year.
- Weekly sales trend are highly volatile, indicating sharp ups and downs.
- Monthly and Quarterly Sales trend clearly shows the seasonality, which indicates that the sales are peaked in year end (Q4).

# Bivariate Analysis:

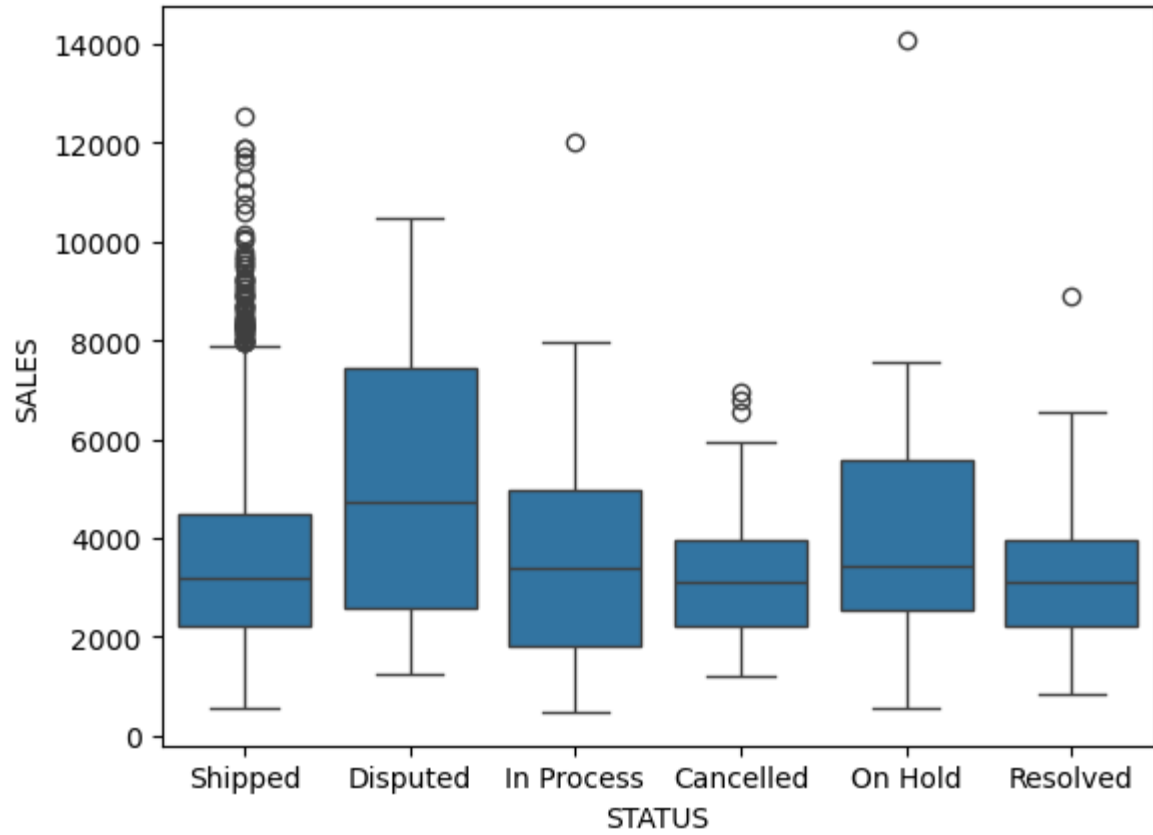
QUANTITYORDERED VS SALES



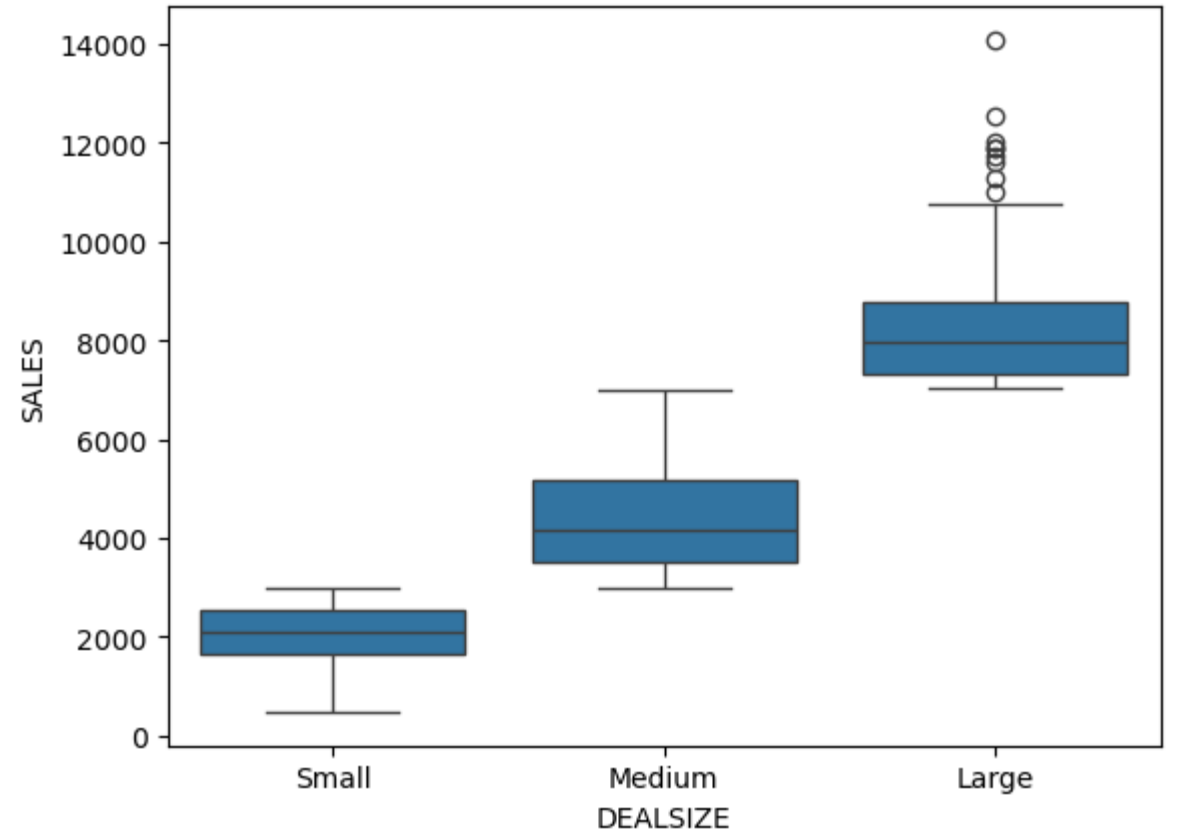
PRICEEACH VS SALES



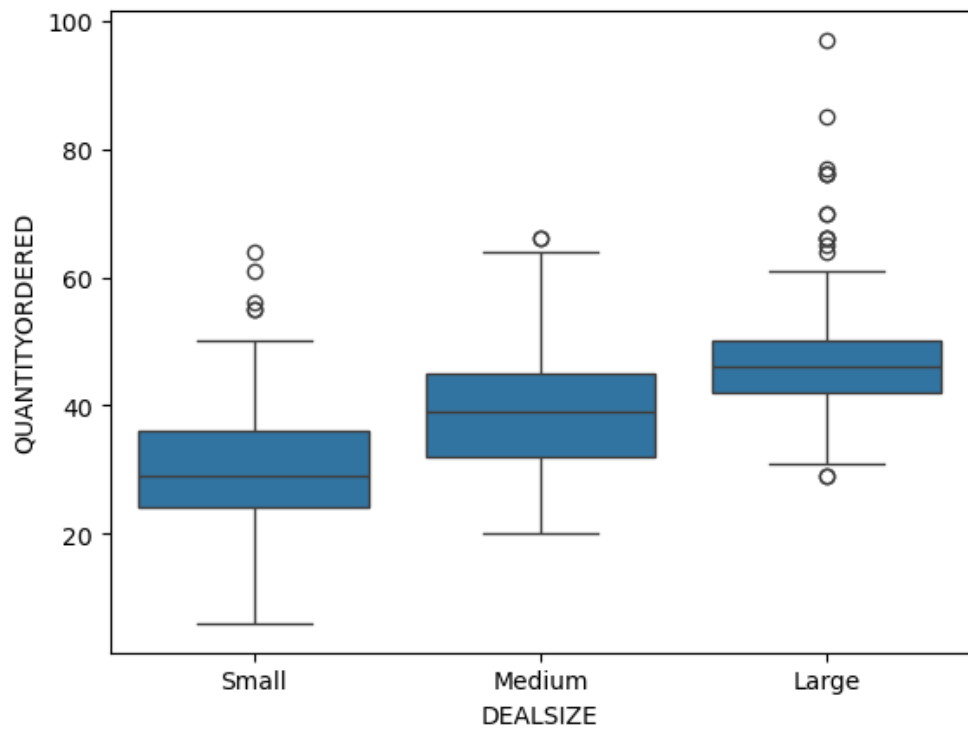
STATUS VS SALES



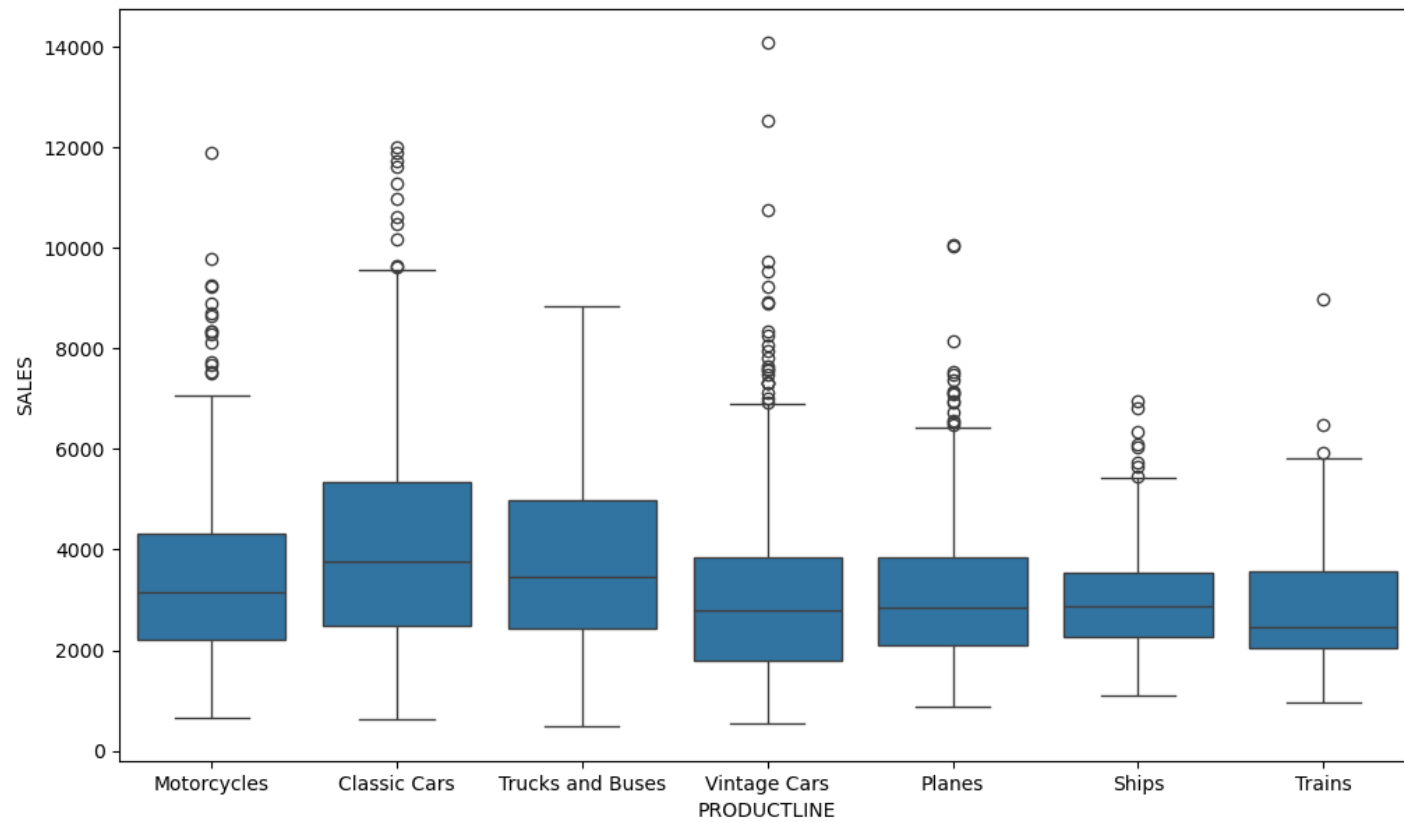
DEALSIZE VS SALES



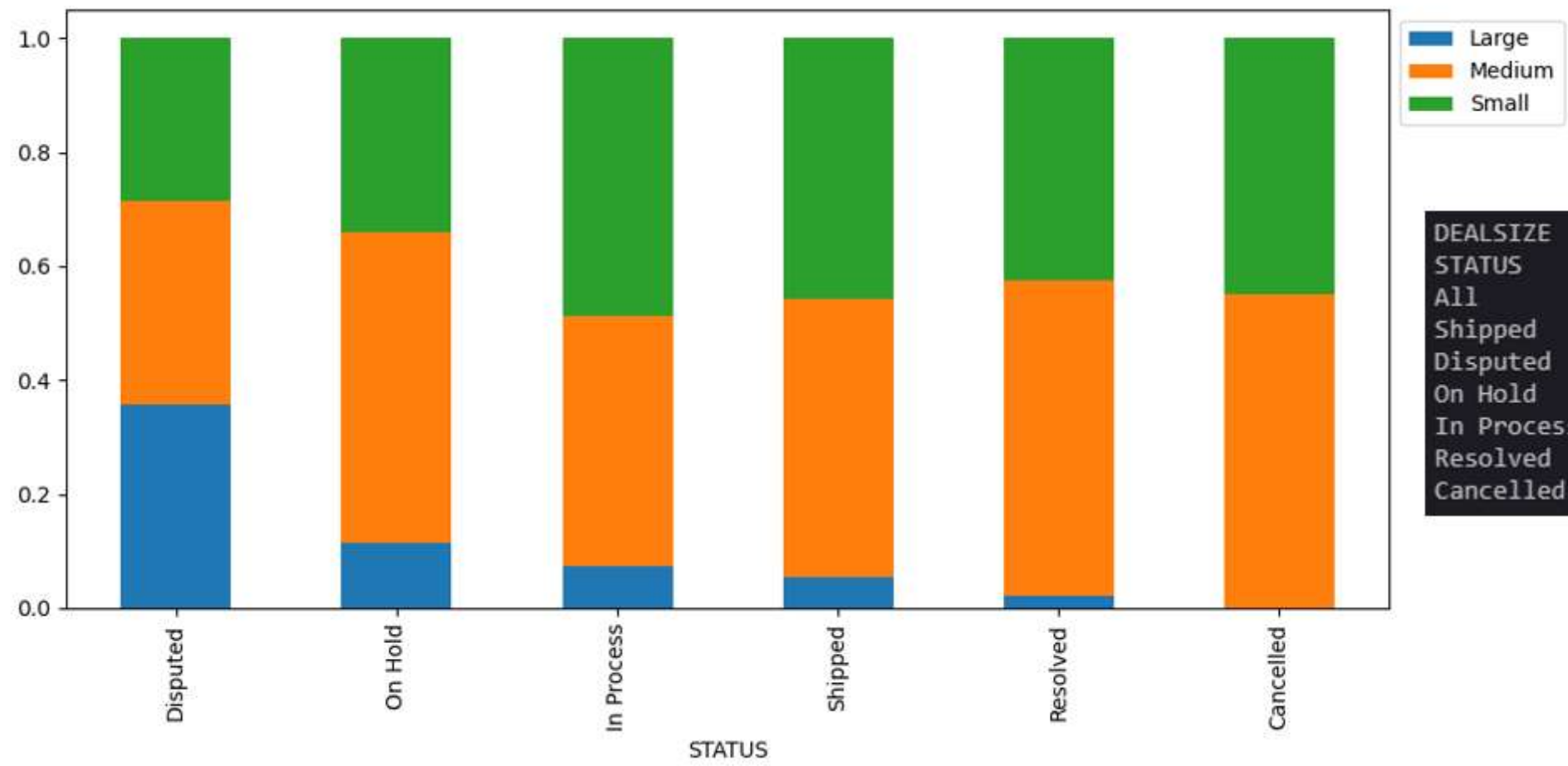
DEALSIZE VS QUANTITYORDERED



PRODUCT\_LINE VS SALES

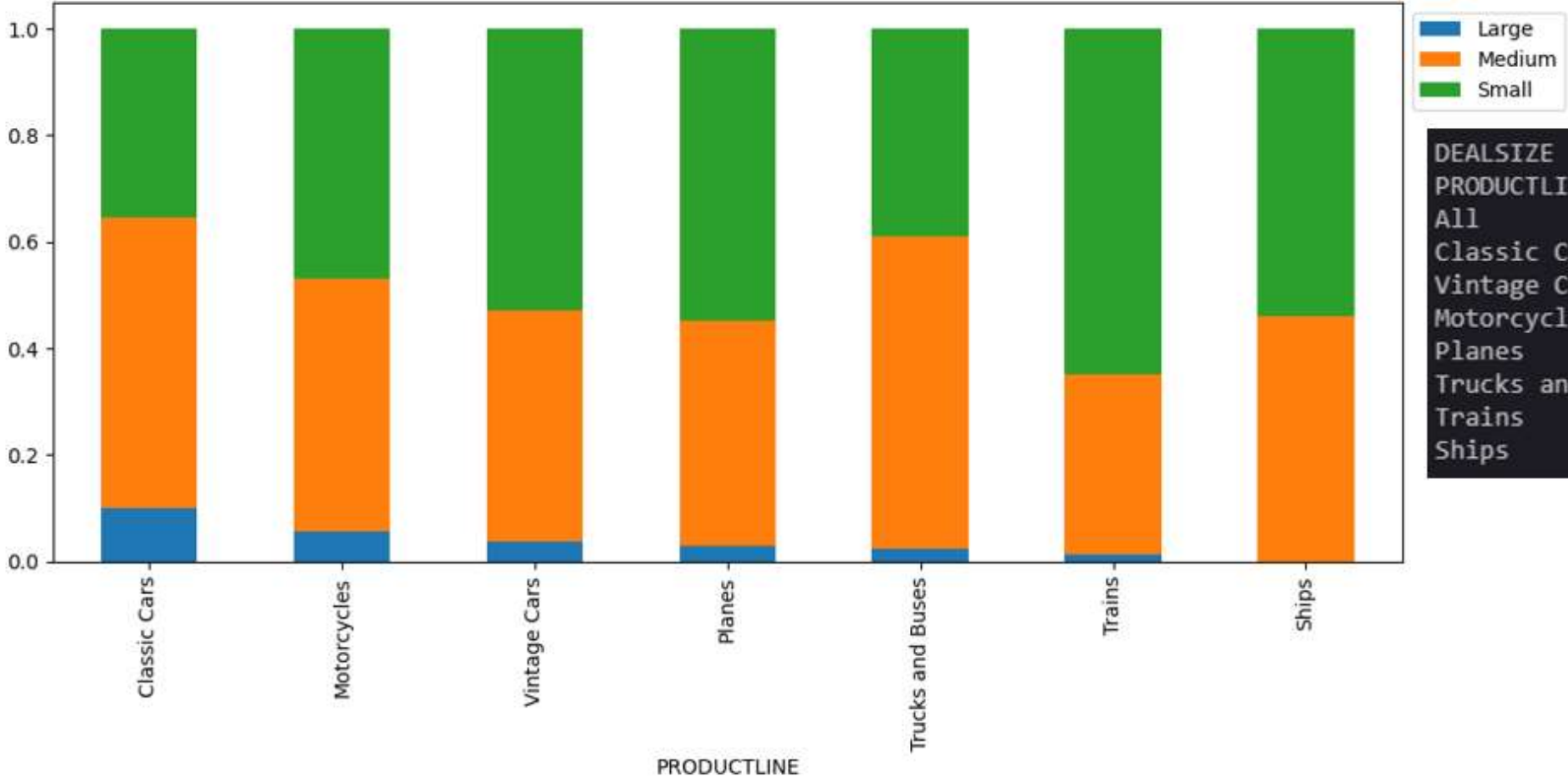


STATUS VS DEALSIZE



DEALSIZE	Large	Medium	Small	All
STATUS				
All	152	1349	1246	2747
Shipped	138	1243	1160	2541
Disputed	5	5	4	14
On Hold	5	24	15	44
In Process	3	18	20	41
Resolved	1	26	20	47
Cancelled	0	33	27	60

# PRODUCTLINE VS DEALSIZE



DEALSIZE	Large	Medium	Small	All
PRODUCTLINE				
All	152	1349	1246	2747
Classic Cars	95	518	336	949
Vintage Cars	22	250	307	579
Motorcycles	18	148	147	313
Planes	9	128	167	304
Trucks and Buses	7	173	115	295
Trains	1	26	50	77
Ships	0	106	124	230



## Inferences:

- A strong positive correlation can be seen between **QUANTITYORDERED** and **SALES**, which indicates that larger orders results in higher sales.
- A strong positive correlation can be seen between **PRICEEACH** and **SALES**, which indicates that highly priced products results in higher revenue.
- Large Deal sizes consistently has higher median sales compared to other deal sizes.
- **Disputed** Orders have the highest median of sales compared to other shipping status.
- Sales increases with the increase of **DEALSIZE**.
- Compared to other product lines, “**Classic cars**” have the highest sales volume and higher median.
- There is a significant distribution of “**On-Hold**” status which needs to be analysed.
- “**Small**” and “**Medium**” deal sizes are the more consistent deal sizes across the different product lines.

# Tool used For RFM Analysis:



- KNIME – Konstanz Information Miner (Open-source) software

# Parameters and Assumptions:

## Monetary:

- Represents the monetary value of each customer.
- Monetary value can be computed by multiplying (**QUANTITYORDERED \* PRICEEACH**) or **SALES** column can be taken.

## Frequency:

- Represents the total no. of purchases of each customer.
- Frequency is calculated by the count of **ORDERNUMBER** and grouped by **CUSTOMERNAME**.

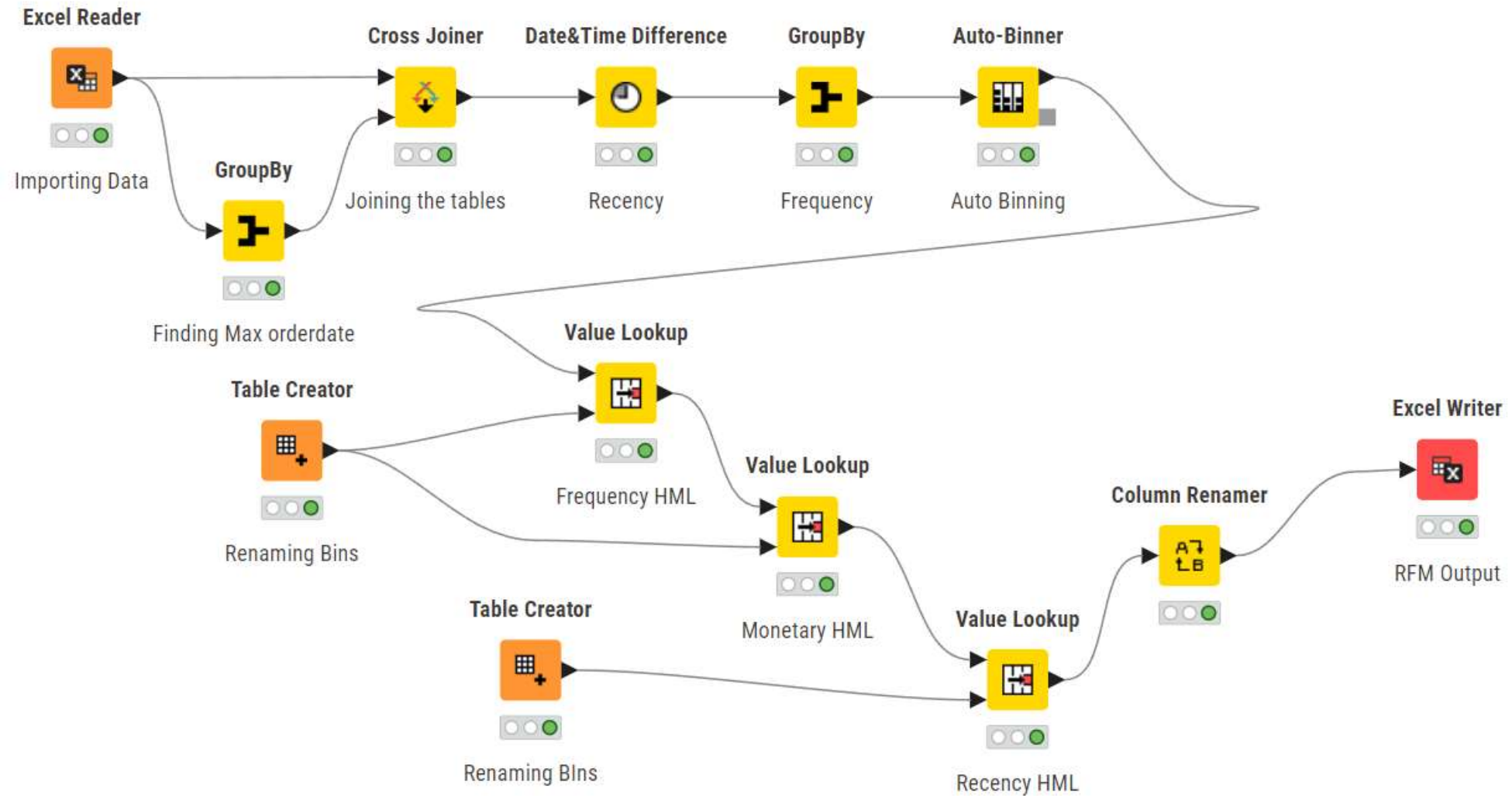
## Recency:

- Represents the recent purchase of the customer.
- Recency value is computed by the difference between **maximum ORDERDATE** and **each customer's ORDERDATE**.

## Segments:

- Based on Recency, Monetary and Frequency, the customer data is segmented into 4 bins.
- 4 BINS: **Very High, High, Medium, Low**
- For Recency, “Very High” being the most recent.
- For Monetary and Frequency, “Very High” being high monetary value or most frequent.

# KNIME Workflow



# Criteria for choosing customers:

	Best Customers	Customers on the verge of Losing	Lost Customers	Loyal Customers
Frequency	Very High	<ul style="list-style-type: none"><li>• Very High</li><li>• High</li><li>• Medium</li></ul>	Low	Very High
Monetary	Very High	<ul style="list-style-type: none"><li>• Very High</li><li>• High</li><li>• Medium</li></ul>	Low	<ul style="list-style-type: none"><li>• Very High</li><li>• High</li></ul>
Recency	Very High	Low	Low	<ul style="list-style-type: none"><li>• Very High</li><li>• High</li></ul>
Sorted By	Monetary	Monetary	Recency	Frequency

# Best Customers (Top 5 )

CUSTOMERNAME	Frequency	Monetary	Recency	Frequency level	Monetary level	Recency level
Euro Shopping Channel	259	912294.11	0	Very High	Very High	Very High
Mini Gifts Distributors Ltd.	180	654858.06	2	Very High	Very High	Very High
La Rochelle Gifts	53	180124.9	0	Very High	Very High	Very High
The Sharp Gifts Warehouse	40	160010.27	39	Very High	Very High	Very High
Souveniers And Things Co.	46	151570.98	2	Very High	Very High	Very High

## Business Recommendations:

- The best customers are those who purchase frequently, purchase recently, and have high monetary value.
- They are very valuable, and active contributors to company income, which are clearly indicated by their high involvement across recency, frequency, and monetary levels.
- These Customers can be rewarded with loyalty and VIP programs.

# Customers on the verge of churning (Top 5)

CUSTOMERNAME	Frequency	Monetary	Recency	Frequency level	Monetary level	Recency level
Saveley & Henriot, Co.	41	142874.25	456	Very High	Very High	Low
Vida Sport, Ltd	31	117713.56	275	High	High	Low
Herkku Gifts	29	111640.28	271	High	High	Low
Marta's Replicas Co.	27	103080.38	231	High	High	Low
Amica Models & Co.	26	94117.26	265	Medium	High	Low

## Business Recommendations:

- These customers are those who have high order numbers and monetary value, but low recency level.
- They are reducing their engagement and are at high risk of being lost.
- Immediate actions like Retention campaigns, targeted approach should be done. Surveys and reviews can be conducted to understand the churning trend.

# Lost Customers (Top 5)

CUSTOMERNAME	Frequency	Monetary	Recency	Frequency level	Monetary level	Recency level
Double Decker Gift Stores, Ltd	12	36019.04	495	Low	Low	Low
West Coast Collectables Co.	13	46084.64	488	Low	Low	Low
Signal Collectibles Ltd.	15	50218.51	476	Low	Low	Low
Daedalus Designs Imports	20	69052.41	465	Low	Low	Low
CAF Imports	13	49642.05	439	Low	Low	Low

## Business Recommendations:

- These lost customers haven't purchased recently, low frequency, and have low monetary value.
- These customers have become inactive and doesn't contribute in company's income.
- Marketing Campaigns in these segments can be reduced or avoided.
- Special Re-entry offers might be provided for those who have high monetary value.



# Loyal Customers (Top 5)

CUSTOMERNAME	Frequency	Monetary	Recency	Frequency level	Monetary level	Recency level
Euro Shopping Channel	259	912294.11	0	Very High	Very High	Very High
Mini Gifts Distributors Ltd.	180	654858.06	2	Very High	Very High	Very High
Australian Collectors, Co.	55	200995.41	184	Very High	Very High	High
La Rochelle Gifts	53	180124.9	0	Very High	Very High	Very High
Muscle Machine Inc	48	197736.94	182	Very High	Very High	High

## Business Recommendations:

- The loyal customers are those who purchase frequently and regularly. They may not have the highest monetary value and recency scores, but they consistently choose our company.
- These customers can be retained by providing personalized offers and discounts.
- They can be rewarded with loyalty and VIP programs.

# Actionable Insights and Recommendations

- Loyal customers can be retained by providing personalized offers and discounts
- Frequent surveys and feedback sessions can be arranged to understand the reasons behind customer churn rate.
- Using Data-driven or ML models can help to make more relevant and engaging campaigns.
- The company should acknowledge and honour consistent purchases even if they don't have the highest monetary value.
- Client behaviour can be monitored regularly to provide engaging offers at the right time.
- Lost customers can be retained with little marketing effort and reactivation programs