Marketing and Retail Analytics

By,
Akshaya Nallathambi
16th January 2022

Contents of the Presentation

- Problem Statement
- Aim of the project
- Data Dictionary
- Data Summary
- Inference about the data
- Exploratory Data Analysis (EDA)
- Inference about EDA
- Customer Segmentation using RFM analysis
- Inference about RFM analysis of the 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.

Aim of the project

This project aims to find the underlying buying patterns of the customers of an automobile part manufacturer based on the past 3 years of the Company's transaction data and hence recommend customized marketing strategies for different segments of customers.

Data Dictionary

ORDERNUMBER :	Order Number	PRODUCTCODE :	Code of Product
QUANTITYORDERED :	Quantity ordered	CUSTOMERNAME :	customer
PRICEEACH :	Price of Each item	PHONE :	Phone of the customer
ORDERLINENUMBER :	order line	ADDRESSLINE1 :	Address of customer
SALES :	Sales amount	CITY :	City of customer
ORDERDATE :	Order Date	POSTALCODE :	Postal Code of customer
DAYS_SINCE_LASTORDER :	Days_ Since_Lastorder	COUNTRY :	Country customer
STATUS :	Status of order like Shipped or not	CONTACTLASTNAME :	Contact person customer
PRODUCTLINE :	Product line – CATEGORY	CONTACTFIRSTNAME :	Contact person customer
MSRP :	Manufacturer's Suggested Retail Price	DEALSIZE :	Size of the deal based on Quantity and Item Price

Data Summary

	0	1	2	3	4
ORDERNUMBER	10107	10121	10134	10145	10168
QUANTITYORDERED	30	34	41	45	36
PRICEEACH	95.7	81.35	94.74	83.26	96.66
ORDERLINENUMBER	2	5	2	6	1
SALES	2871.0	2765.9	3884.34	3746.7	3479.76
ORDERDATE	2018-02-24 00:00:00	2018-05-07 00:00:00	2018-07-01 00:00:00	2018-08-25 00:00:00	2018-10-28 00:00:00
DAYS_SINCE_LASTORDER	828	757	703	649	586
STATUS	Shipped	Shipped	Shipped	Shipped	Shipped
PRODUCTLINE	Motorcycles	Motorcycles	Motorcycles	Motorcycles	Motorcycles
MSRP	95	95	95	95	95
PRODUCTCODE	S10_1678	S10_1678	S10_1678	S10_1678	S10_1678
CUSTOMERNAME	Land of Toys Inc.	Reims Collectables	Lyon Souveniers	Toys4GrownUps.com	Technics Stores Inc.
PHONE	2125557818	26.47.1555	+33 1 46 62 7555	6265557265	6505556809
ADDRESSLINE1	897 Long Airport Avenue	59 rue de l'Abbaye	27 rue du Colonel Pierre Avia	78934 Hillside Dr.	9408 Furth Circle
CITY	NYC	Reims	Paris	Pasadena	Burlingame
POSTALCODE	10022	51100	75508	90003	94217
COUNTRY	USA	France	France	USA	USA
CONTACTLASTNAME	Yu	Henriot	Da Cunha	Young	Hirano
CONTACTFIRSTNAME	Kwai	Paul	Daniel	Julie	Juri
DEALSIZE	Small	Small	Medium	Medium	Medium

This table gives the first 5 rows (inverted) of sample data.

```
<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
```

This table gives the basic information of the data set. It is clear that there are variables of type int, float, object and date with 20 columns and 2747 rows. There are no null values. The memory usage is 429.3+ KB.

	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

This table gives the five point summary of the continuous variables in the data set. This tells how the data is distributed.

Inferences:

- The data is about an automobile parts manufacturing company. They have provided over 3 years of data.
- The given data set has variables of type int, float, object and date with 20 columns and 2747 rows. There are no null values. The memory usage is 429.3+ KB.
- This data gives the purchasing behavior of customers in different categories. The company is into manufacturing automobile parts, and they have different product line like Classic cars , Motorcycles, Planes, Ships, Trains, Trucks and Buses, Vintage Cars.
- Each transaction has an order number and for each order number maintained all required information like customer identity details , and product details like price , quantity , product code, and sales for each customer is available. Also it can be identified that one order number has many entries with different product codes.
- Manufacturer's Suggested Retail Price(MSRP) for each product code is decided but it is found that this is not matching with Price of each item.

Exploratory Data Analysis

Exploratory Data Analysis (EDA) is an approach of analyzing data sets to summarize their main characteristics, often using statistical graphics and other data visualization methods.

Box plots help in identifying the outliers in the data

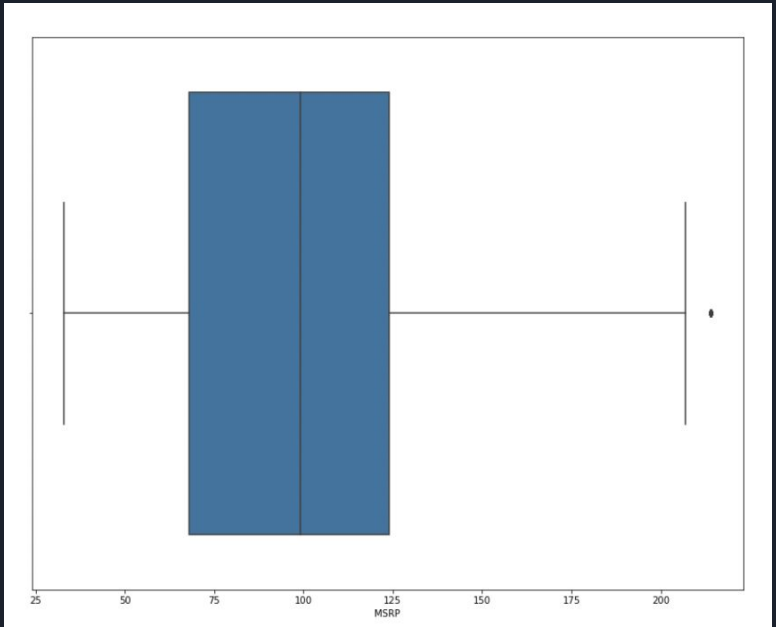
Univariate analysis: It is the simplest form of analyzing data. “Uni” means “one”, so in other words your data has only one variable. The histograms are used for numerical variables and bar plot is used for categorical variable to perform univariate analysis.

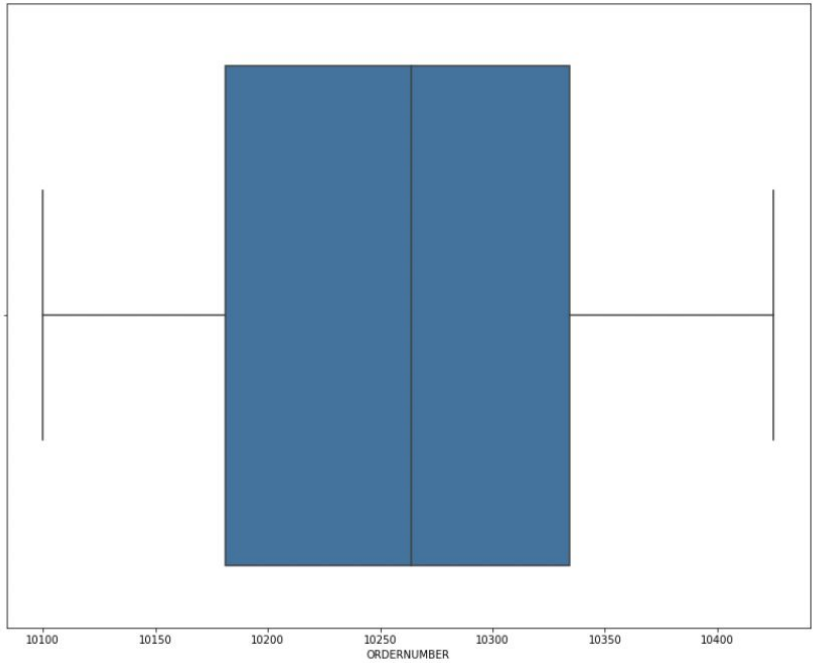
Bivariate analysis: It is the simplest forms of quantitative analysis. It involves the analysis of two variables, for the purpose of determining the empirical relationship between them. The pairplot is generally used for numerical variables to perform bivariate analysis.

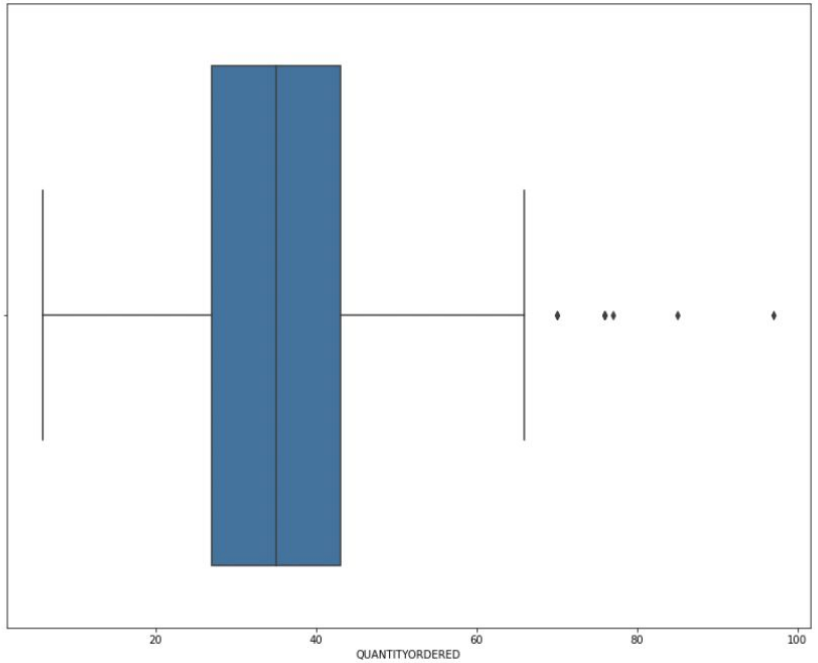
Box Plots

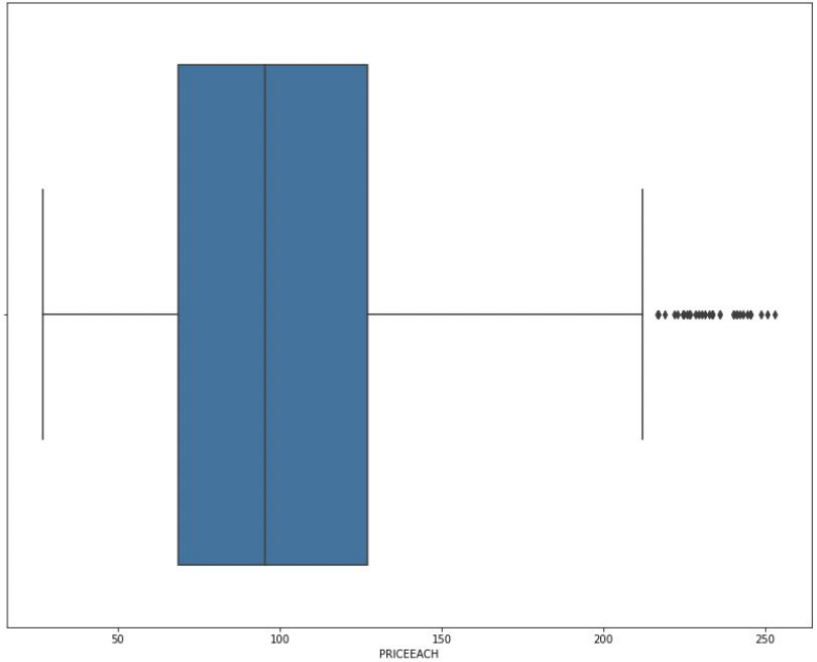
Outliers can be identified using the box plots.

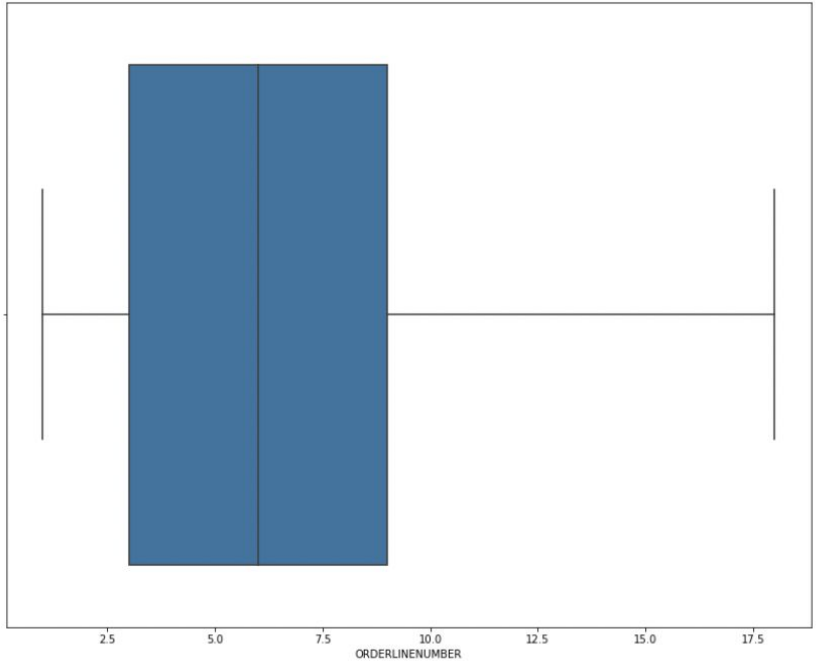
It is clear that the columns 'QUANTITYORDERED', 'PRICEEACH', 'SALES' and 'MSRP' have outliers in them.

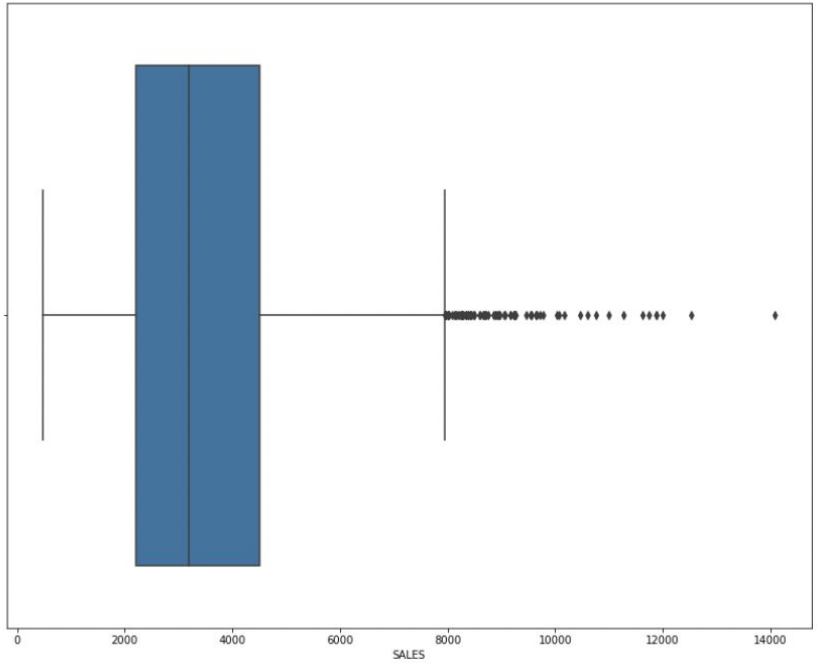


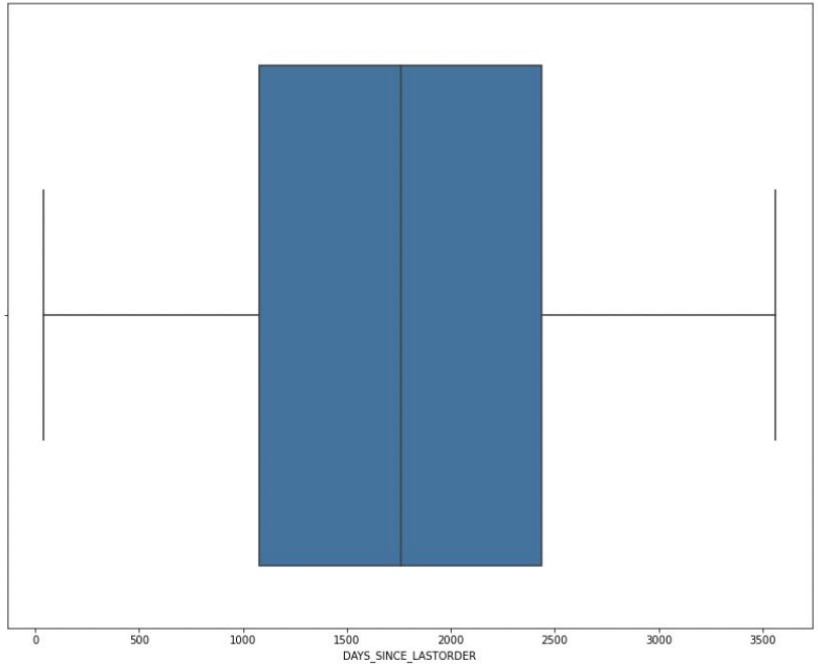

10100
10150
10200
10250
10300
10350
10400
ORDERNUMBER


20
40
60
80
100
QUANTITYORDERED

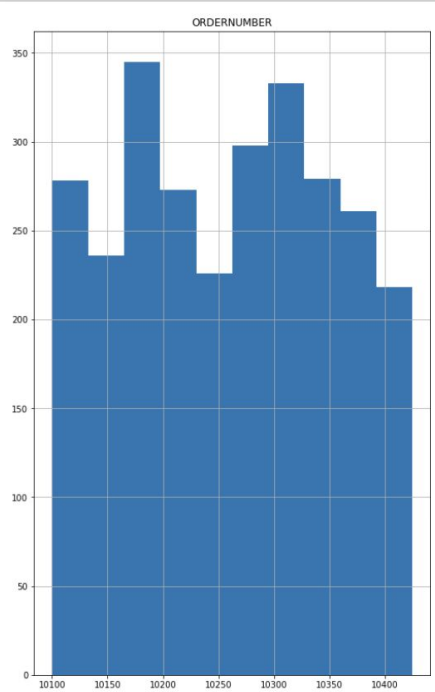

50
100
150
200
250
PRICEEACH

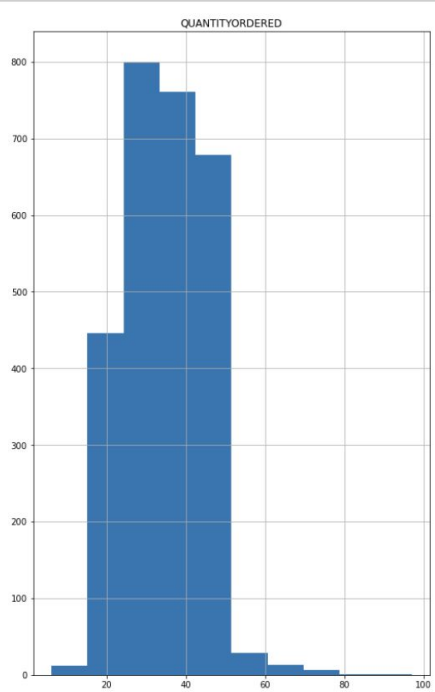

2.5
5.0
7.5
10.0
12.5
15.0
17.5
ORDERLINENUMBER

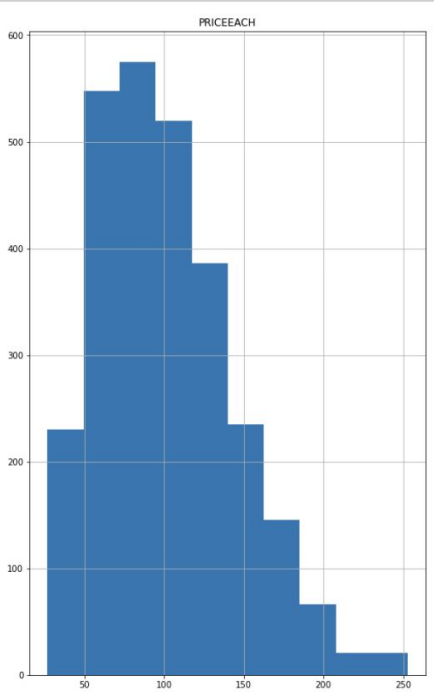

0
2000
4000
6000
8000
10000
12000
14000
SALES

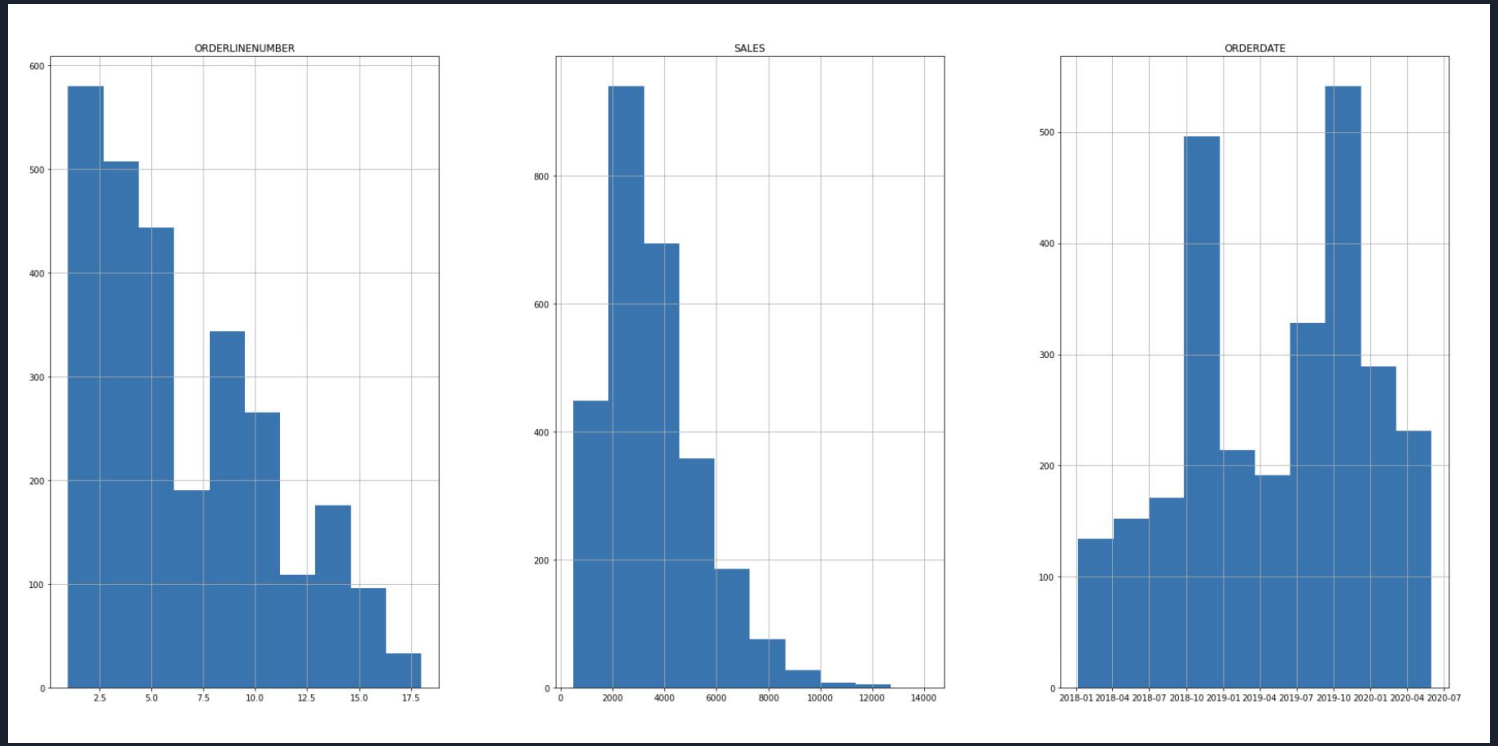

0
500
1000
1500
2000
2500
3000
3500
DAYS_SINCE_LASTORDER

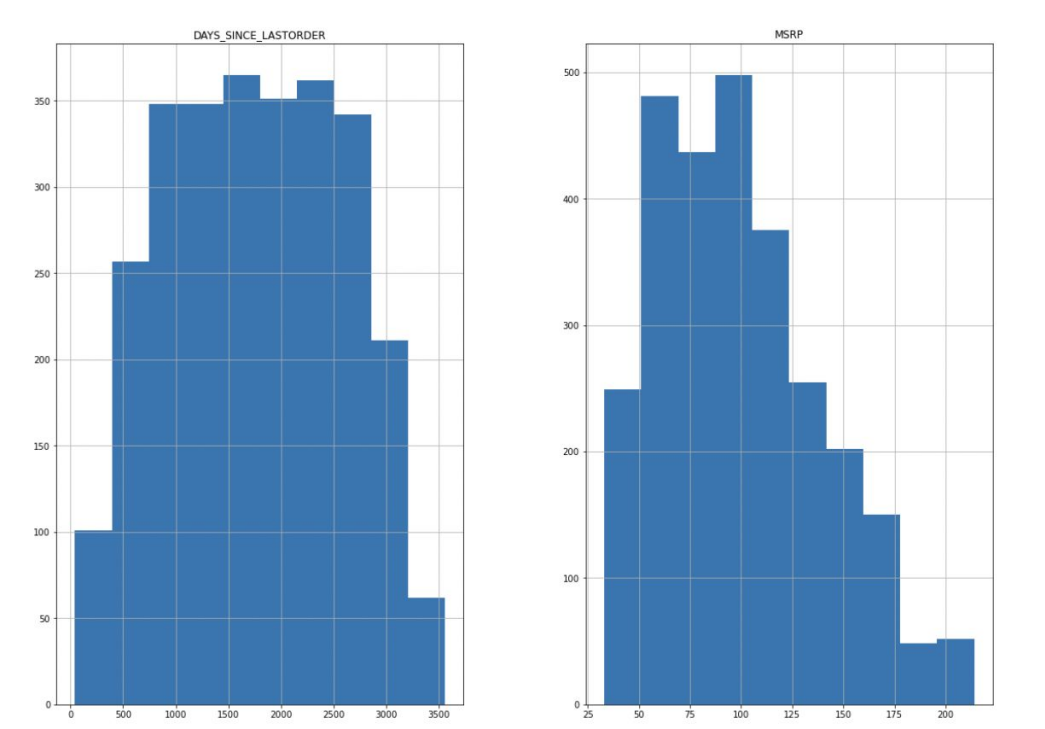
Univariate analysis - Histogram





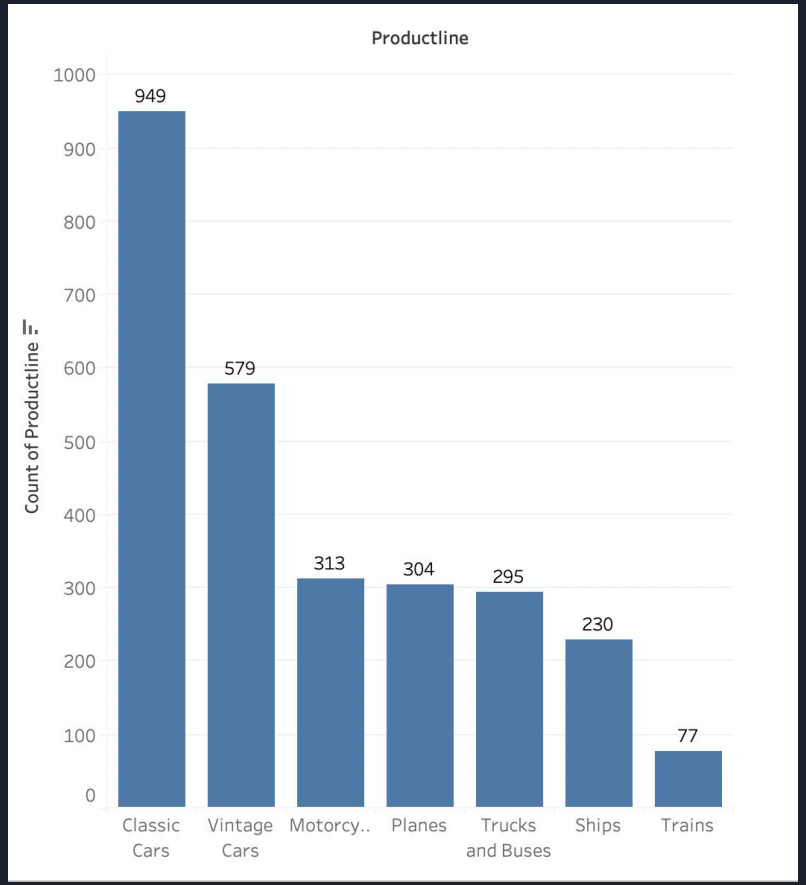



ORDERLINENUMBER
0
100
200
300
400
500
600
2.5
5.0
7.5
10.0
12.5
15.0
17.5
SALES
0
200
400
600
800
0
2000
4000
6000
8000
10000
12000
14000
ORDERDATE
0
100
200
300
400
500
2018-01
2018-04
2018-07
2018-10
2019-01
2019-04
2019-07
2019-10
2020-01
2020-04
2020-07

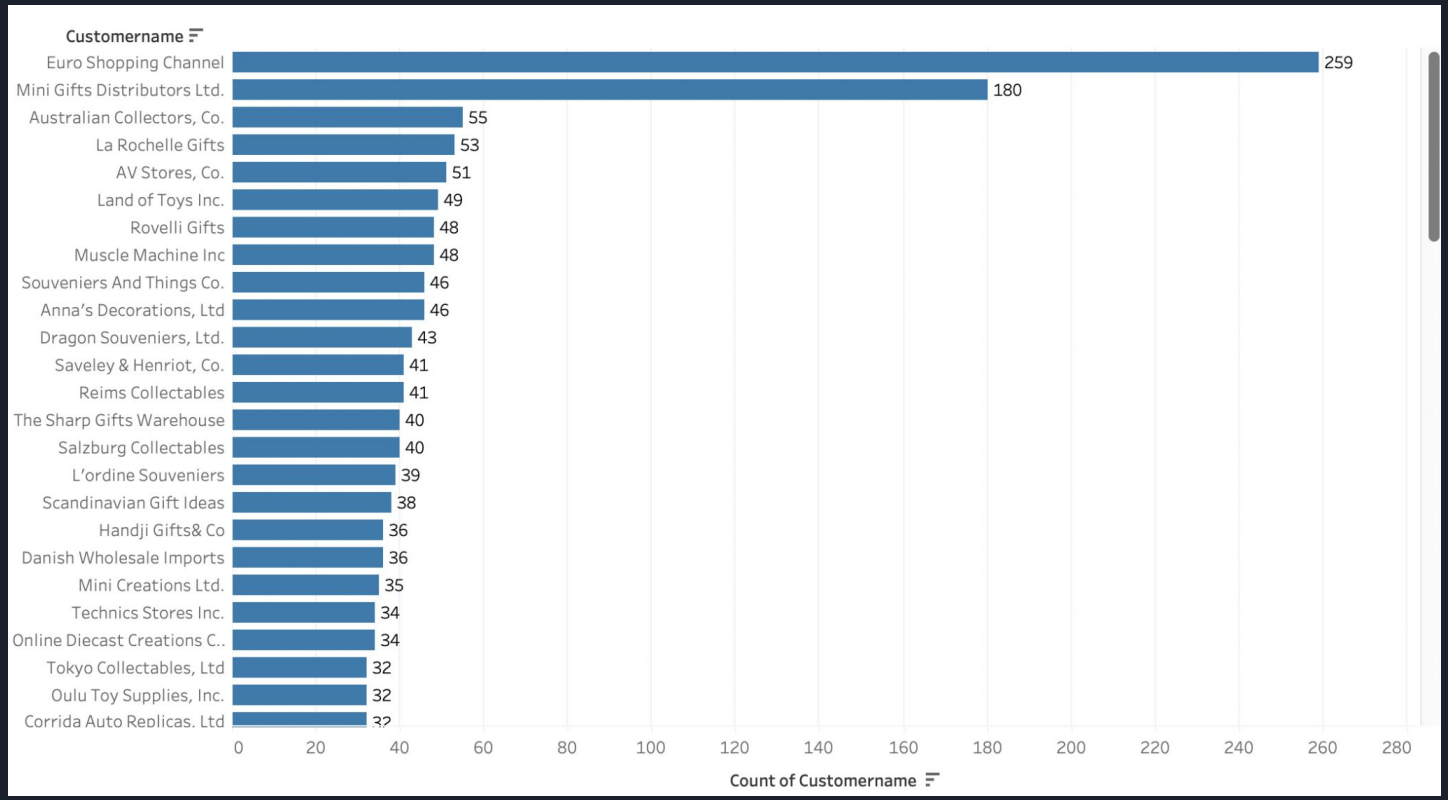


From the histograms, it is clear that the data is less skewed.

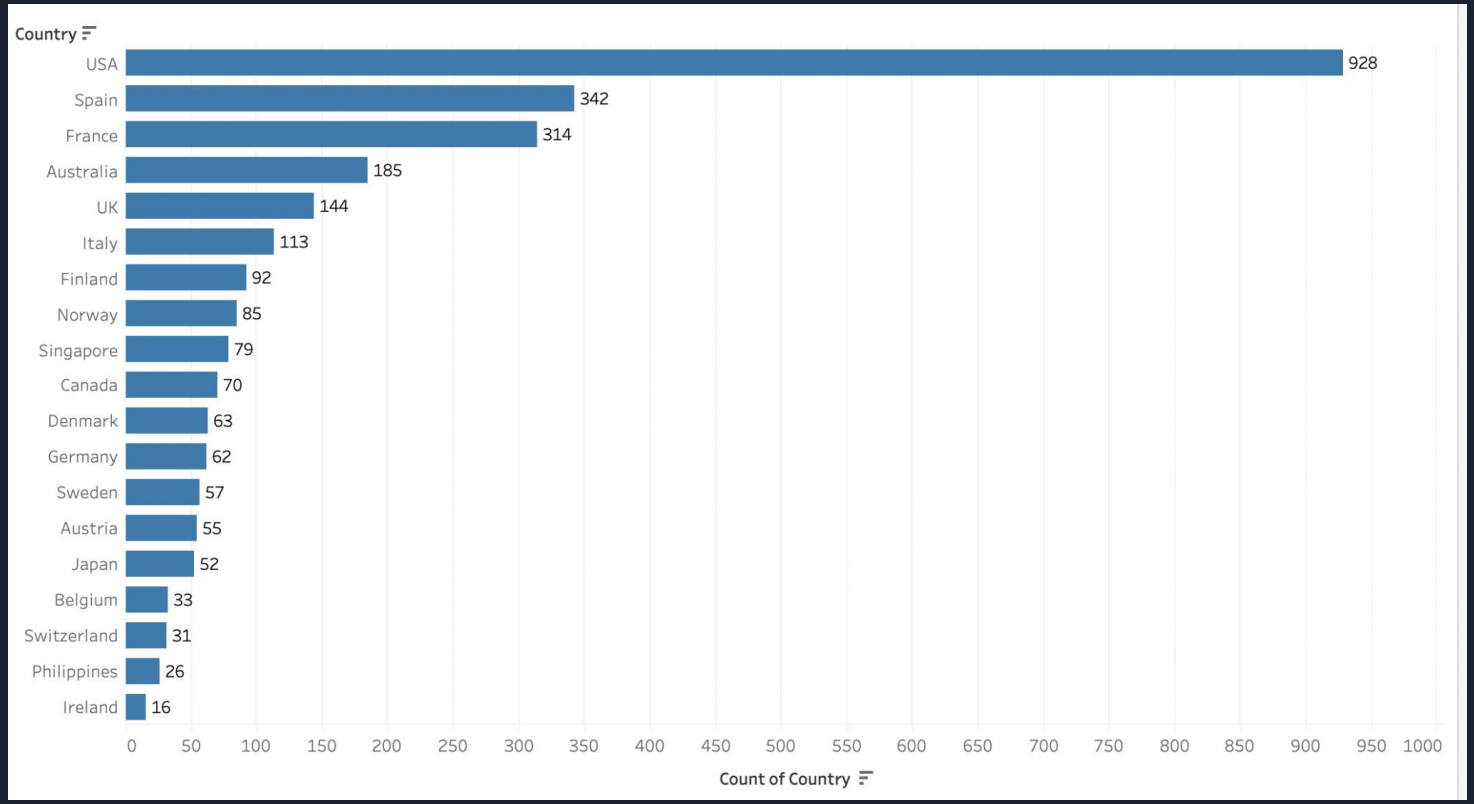
Univariate analysis - Bar Plot



The classic cars are mostly preferred by the customers followed by vintage cars.

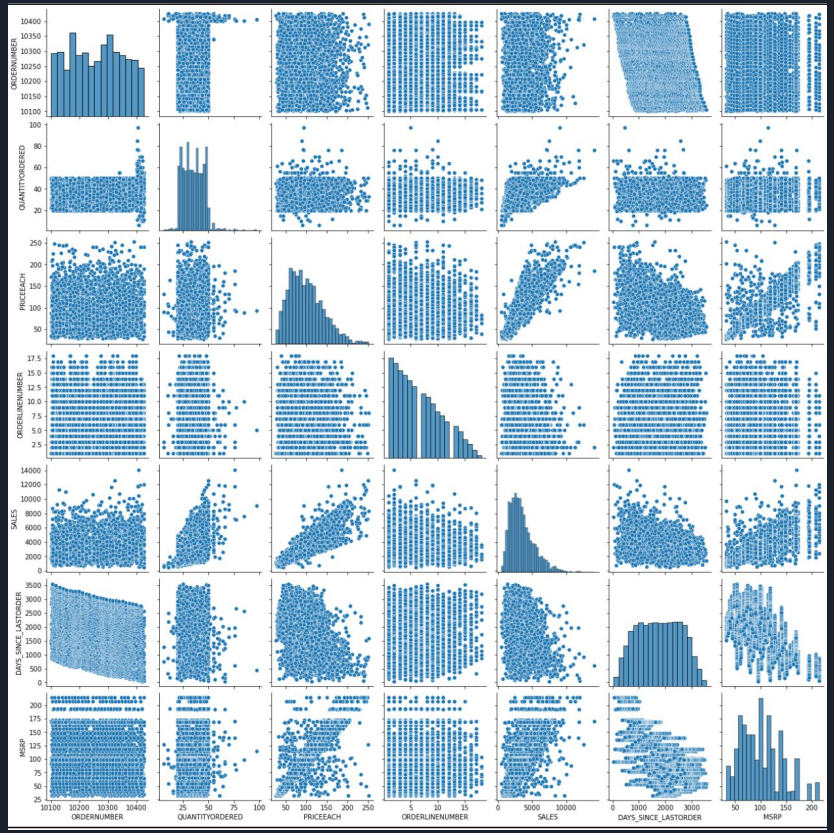


The customer Euro Shopping Channel has the most number of transactions.



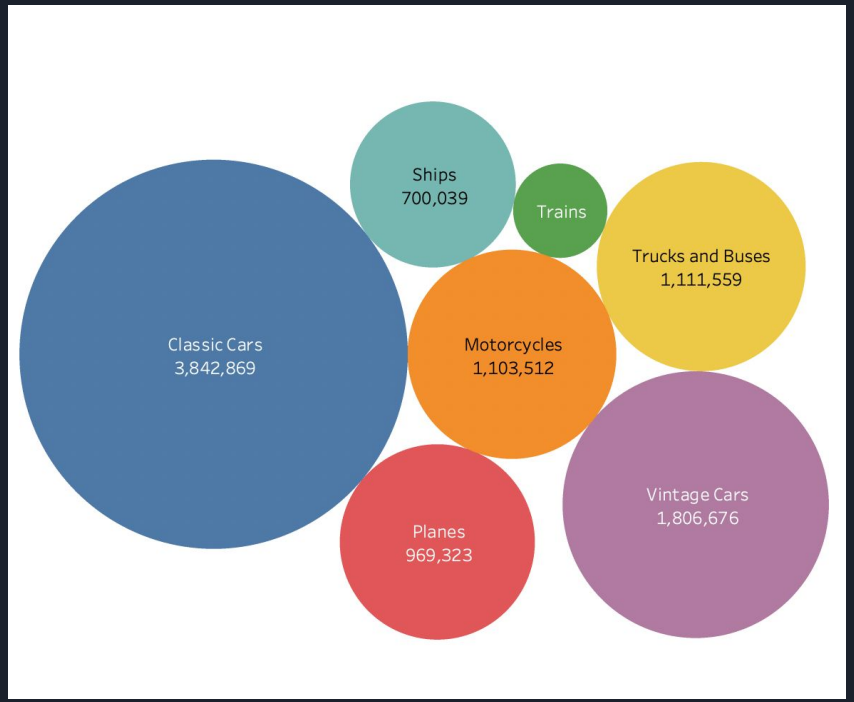
The country USA has the most number of transactions.

Bivariate analysis - Pair Plot



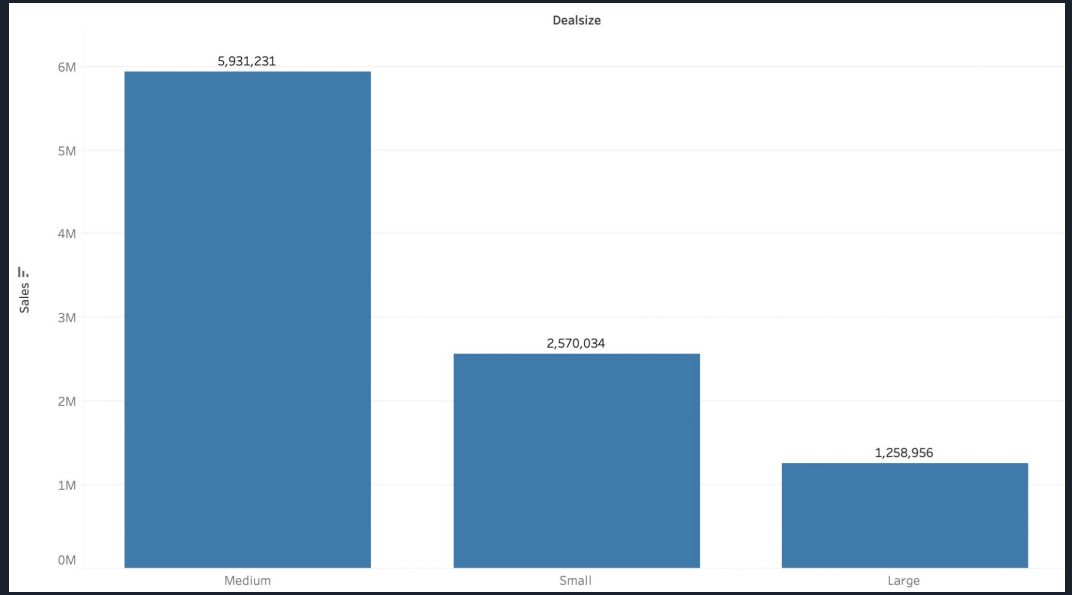
The pairplot gives the relationship between two variables.

Bivariate analysis - Bubble chart



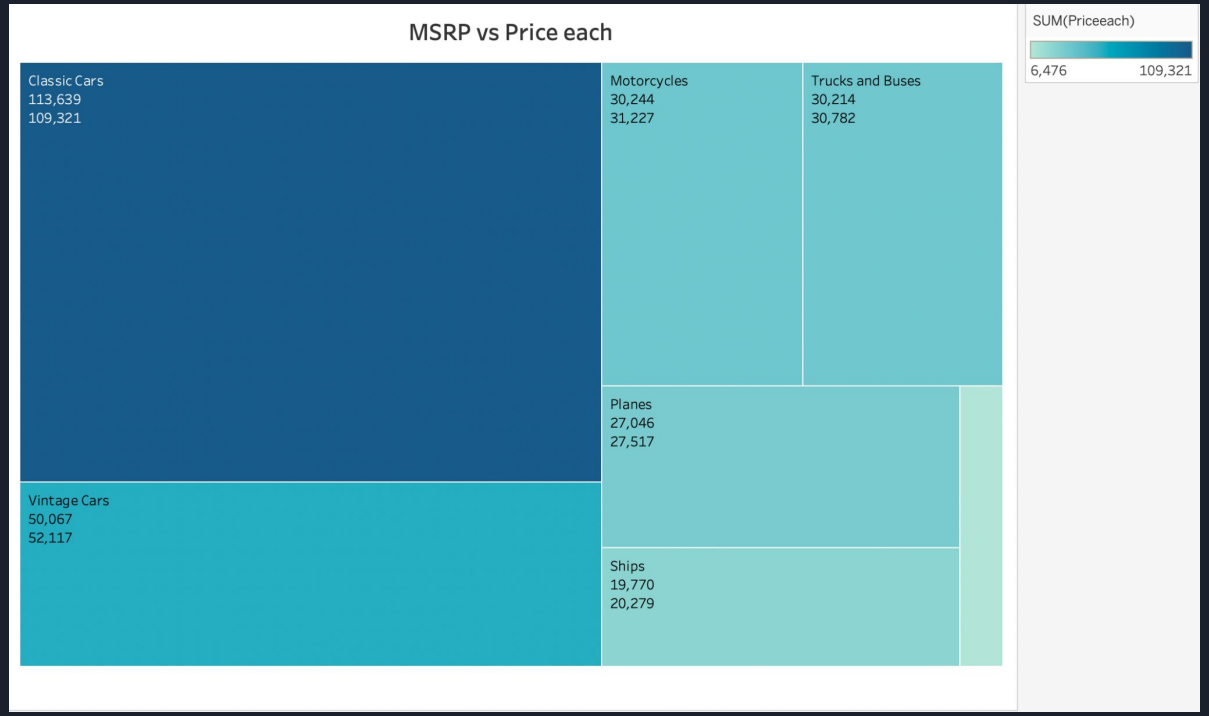
This bubble chart gives the relationship between Product line and Sales.

Bivariate analysis - Bar chart



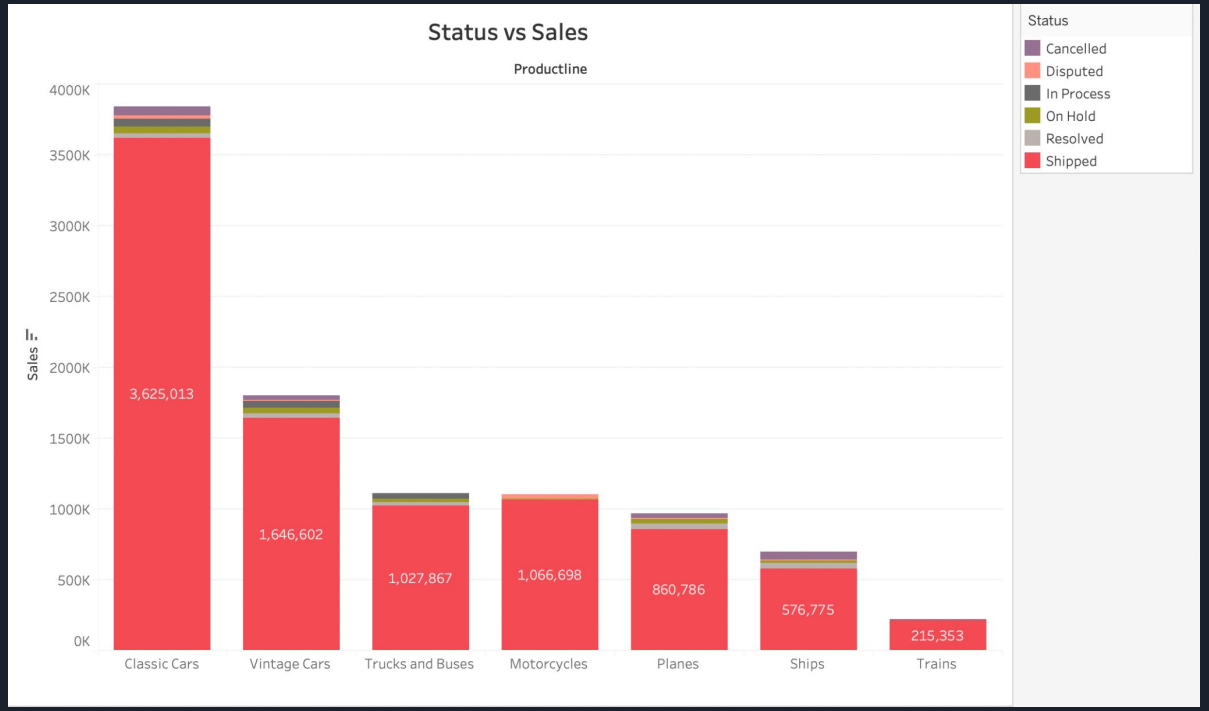
This bar chart gives the relationship between Deal size and Sales.

Multivariate analysis - Tree map



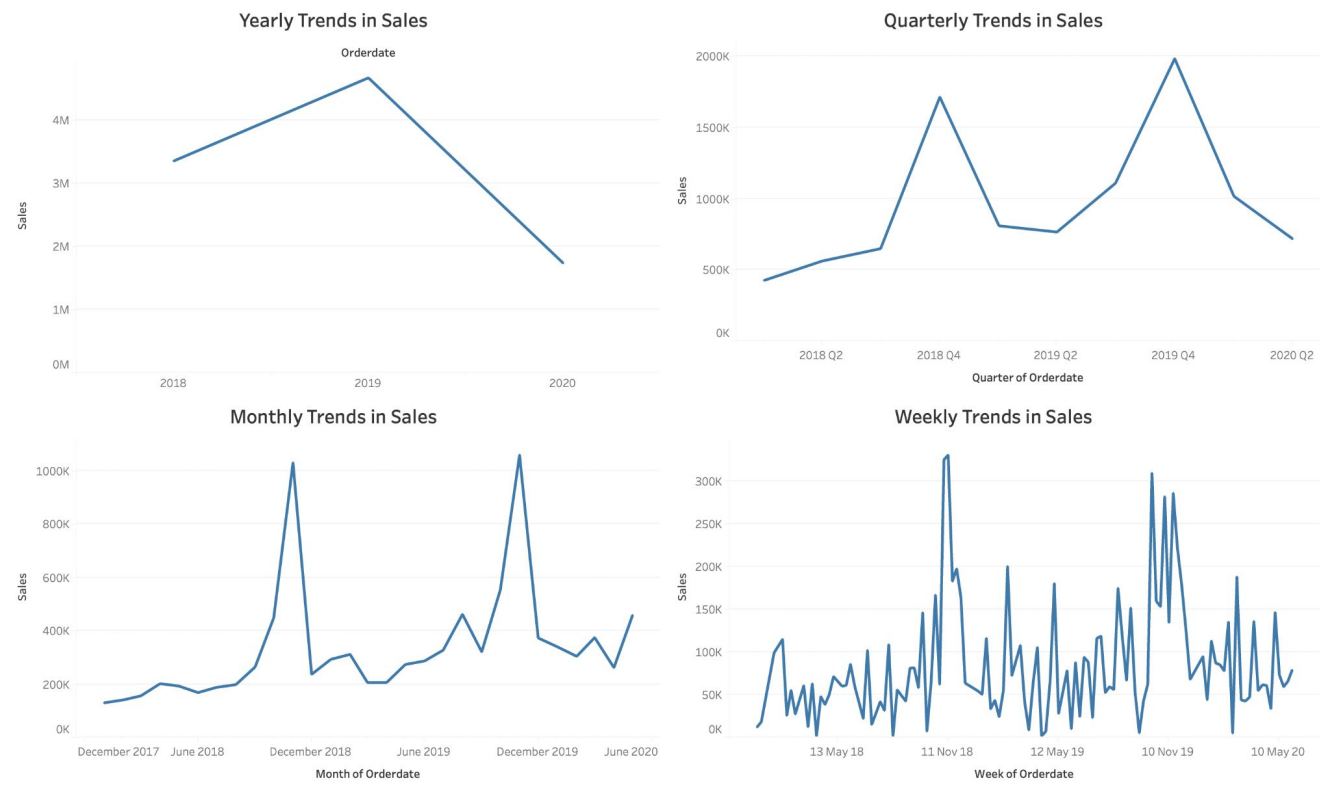
This tree map gives the relationship between Product line, MSRP and price each.

Multivariate analysis - Stacked Bar Chart



This stacked bar chart gives the relationship between Product line, status and sales.

Time series - Line Chart



The line chart gives the yearly, quarterly, monthly and weekly trends in the sales

Inferences:

- The given data has outliers in them. It is clear from the box plot that the columns 'QUANTITYORDERED', 'PRICEEACH', 'SALES' and 'MSRP' have outliers in them.
- The histogram shows that the data is less skewed and almost normally distributed.
- The sales of classic cars are high followed by vintage cars.
- From the bubble chart, it is clear that classic cars has the most sales and the train has the least sales.
- The bar chart between Deal size and Sales shows that medium size deal has the most sales.
- As sales are high for classic cars the company has even sold below MSRP, there might be a chances that the company has given more discounts to its customers. And vice versa for vintage cars were the company has sold above MSRP. Ship, vintage car & train are been sold above the MSRP. By looking at the given data almost all the transactions are been shipped.
- It is observed that in Q4 sales in 2019 is the highest of all. There is a seasonality seen in the graph
- In conclusion, there is a high demand for classic cars followed by vintage cars and the least demand is for trains.

Customer Segmentation using RFM analysis

Which tool used?

- Python and Tableau was used for data summary and EDA. I have used KNIME for the customer segmentation.

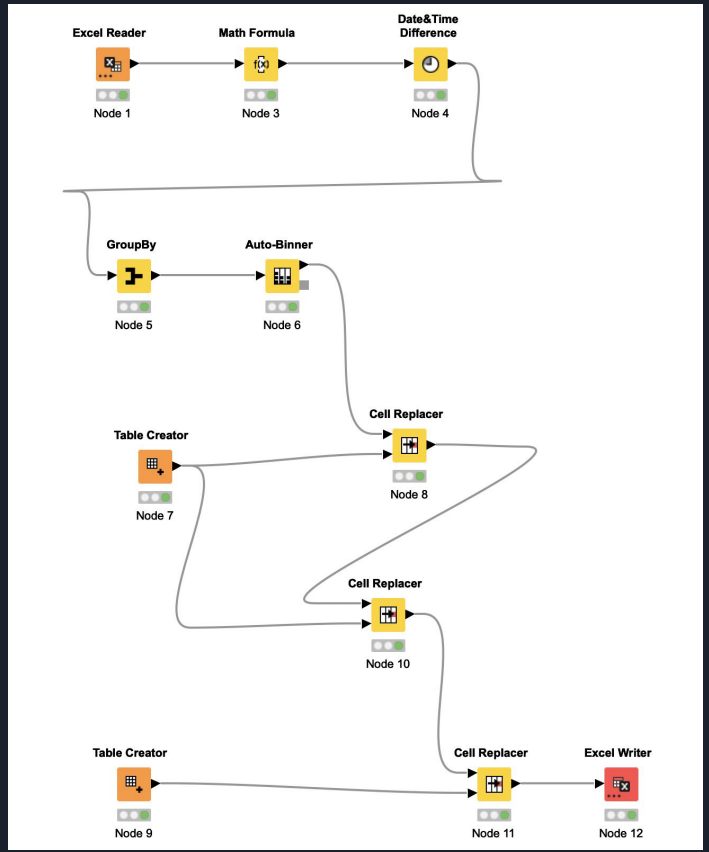
What all parameters used and assumptions made?

- For the **Recency**, I have taken today's date (16th January 2022) into consideration for reference point. Therefore, the column "DAYS_SINCE_LASTORDER" has been ignored.
- For the **Frequency**, the count of "ORDERNUMBER" was taken after grouping it to customer name as there were different order numbers associated with one customer name.
- For the **Monetary**, the sum of the "Sales" was taken post grouping the data.
- There are three bins created H (High), M (Medium) and L (Low) using the percentile range - 0.0, 0.25, 0.75, 1.0.

Sample Data after RFM Analysis

CUSTOMERNAME	PRODUCTCODE	PHONE	ADDRESSLINE1	CITY	POSTALCODE	COUNTRY	CONTACTLASTNAME	CONTACTFIRSTNAME	DEALSIZE	Recency	Frequency	Monetary	Recency Score	Frequency Score	Monetary Score
AV Stores, Co.	S12_1108	(171) 555-1555	Fauntleroy Circus	Manchester	EC2 5NT	UK	Ashworth	Victoria	Medium	791	51	157807.8	M	H	H
Alpha Cognac	S10_4757	61.77.6555	1 rue Alsace-Lorraine	Toulouse	31000	France	Roulet	Annette	Medium	659	20	70488.44	H	L	L
Amica Models & Co.	S10_1949	011-4988555	Via Monte Bianco 34	Torino	10100	Italy	Accorti	Paolo	Large	860	26	94117.26	L	M	M
Anna's Decorations, Ltd	S10_1949	02 9936 8555	201 Miller Street	North Sydney	2060	Australia	O'Hara	Anna	Small	678	46	153996.1	M	H	H
Atelier graphique	S10_2016	40.32.2555	54, rue Royale	Nantes	44000	France	Schmitt	Carine	Medium	783	7	24179.96	M	L	L
Australian Collectables, Ltd	S18_1342	61-9-3844-6555	7 Allen Street	Glen Waverly	3150	Australia	Connery	Sean	Medium	617	23	64591.46	H	M	L
Australian Collectors, Co.	S10_1678	03 9520 4555	636 St Kilda Road	Melbourne	3004	Australia	Ferguson	Peter	Medium	779	55	200995.4	M	H	H
Australian Gift Network, Co	S10_1949	61-7-3844-6555	31 Duncan St. West End	South Brisbane	4101	Australia	Calaghan	Tony	Large	714	15	59469.12	M	L	L
Auto Assoc. & Cie.	S10_1949	30.59.8555	67, avenue de l'Europe	Versailles	78000	France	Tonini	Daniel	Large	828	18	64834.32	L	L	L
Auto Canal Petit	S10_1678	(1) 47.55.6555	25, rue Lauriston	Paris	75016	France	Perrier	Dominique	Medium	649	27	93170.66	H	M	M
Auto-Moto Classics Inc.	S18_3029	6175558428	16780 Pompton St.	Brickhaven	58339	USA	Taylor	Leslie	Medium	775	8	26479.26	M	L	L
Baane Mini Imports	S10_1678	07-98 9555	Erling Skakkes gate 78	Stavern	4110	Norway	Bergulfsen	Jonas	Medium	803	32	116599.2	M	M	M
Bavarian Collectables Imports,	S18_1662	+49 89 61 08 9555	Hansastr. 15	Munich	80686	Germany	Donnermeyer	Michael	Medium	854	14	34993.92	L	L	L
Blauer See Auto, Co.	S12_1099	+49 69 66 90 2555	Lyonerstr. 34	Frankfurt	60528	Germany	Keitel	Roland	Medium	803	22	85171.59	M	M	M
Boards & Toys Co.	S12_3380	3105552373	4097 Douglas Av.	Glendale	92561	USA	Young	Leslie	Medium	708	3	9129.35	M	L	L
CAF Imports	S12_1108	+34 913 728 555	Merchants House, 27-30 Merchant's	Madrid	28023	Spain	Fernandez	Jesus	Large	1034	13	49642.05	L	L	L
Cambridge Collectables Co.	S10_1949	6175555555	4658 Baden Av.	Cambridge	51247	USA	Tseng	Kyung	Medium	984	11	36163.62	L	L	L
Canadian Gift Exchange Netwo	S10_1949	(604) 555-3392	1900 Oak St.	Vancouver	V3F 2K1	Canada	Tannamuri	Yoshi	Large	817	22	75238.92	M	M	M
Classic Gift Ideas, Inc	S10_1949	2155554695	782 First Street	Philadelphia	71270	USA	Cervantes	Francisca	Medium	825	21	67506.97	M	M	L
Classic Legends Inc.	S10_1949	2125558493	5905 Pompton St.	NYC	10022	USA	Hernandez	Maria	Medium	787	20	77795.2	M	L	M
Clover Collections, Co.	S12_1108	+353 1862 1555	25 Maiden Lane	Dublin	2	Ireland	Cassidy	Dean	Large	853	16	57756.43	L	L	L
Collectable Mini Designs Co.	S10_4757	7605558146	361 Furth Circle	San Diego	91217	USA	Thompson	Valarie	Medium	1055	25	87489.23	L	M	M
Collectables For Less Inc.	S10_1949	6175558555	7825 Douglas Av.	Brickhaven	58339	USA	Nelson	Allen	Medium	727	24	81577.98	M	M	M
Corrida Auto Replicas, Ltd	S10_1949	(91) 555 22 82	C/ Araquil, 67	Madrid	28023	Spain	Sommer	Martín	Large	807	32	120615.3	M	M	H
Cruz & Sons Co.	S12_1099	+63 2 555 3587	15 McCallum Street - NatWest Cente	Makati City	1227 MM	Philippines	Cruz	Arnold	Medium	792	26	94015.73	M	M	M
Daedalus Designs Imports	S10_1678	20.16.1555	184, chausse de Tournai	Lille	59000	France	Rance	Martine	Small	1060	20	69052.41	L	L	L

KNIME workflow



Best Customers

CUSTOMERNAME	Recency Score	Frequency Score	Monetary Score
Danish Wholesale Imports	H	H	H
Euro Shopping Channel	H	H	H
L'ordine Souveniers	H	H	H
La Rochelle Gifts	H	H	H
Mini Gifts Distributors Ltd.	H	H	H
Reims Collectables	H	H	H
Salzburg Collectables	H	H	H
Souveniers And Things Co.	H	H	H
The Sharp Gifts Warehouse	H	H	H

There are 10 customers with high Recency, high Frequency and high Monetary Score. Therefore, these customers are considered the best customers.

Customers on the verge of churning

CUSTOMERNAME	Recency Score	Frequency Score	Monetary Score
AV Stores, Co.	M	H	H
Anna's Decorations, Ltd	M	H	H
Australian Collectors, Co.	M	H	H
Dragon Souveniers, Ltd.	M	H	H
Land of Toys Inc.	M	H	H
Muscle Machine Inc	M	H	H
Online Diecast Creations Co.	M	H	H
Rovelli Gifts	M	H	H
Saveley & Henriot, Co.	L	H	H
Scandinavian Gift Ideas	M	H	H
Technics Stores Inc.	M	H	H

These are the customers who purchased often and spent big amounts, but haven't purchased recently. Therefore, they are about to churn and need attention.

Lost Customers

CUSTOMERNAME	Recency Score	Frequency Score	Monetary Score
Auto Assoc. & Cie.	L	L	L
Bavarian Collectables Imports, Co.	L	L	L
CAF Imports	L	L	L
Cambridge Collectables Co.	L	L	L
Clover Collections, Co.	L	L	L
Daedalus Designs Imports	L	L	L
Diecast Collectables	L	L	M
Double Decker Gift Stores, Ltd	L	L	L
Iberia Gift Imports, Corp.	L	L	L
Online Mini Collectables	L	L	L
Osaka Souveniers Co.	L	L	L
Signal Collectibles Ltd.	L	L	L
Super Scale Inc.	L	L	M
West Coast Collectables Co.	L	L	L

These are the customers whose last purchase were long back and had low number of orders. Therefore, they are lost.

Loyal Customers

CUSTOMERNAME	Recency Score	Frequency Score	Monetary Score
Danish Wholesale Imports	H	H	H
Euro Shopping Channel	H	H	H
Handji Gifts& Co	H	H	M
L'ordine Souveniers	H	H	H
La Rochelle Gifts	H	H	H
Mini Gifts Distributors Ltd.	H	H	H
Reims Collectables	H	H	H
Salzburg Collectables	H	H	H
Souveniers And Things Co.	H	H	H
The Sharp Gifts Warehouse	H	H	H

These are the customers who buy on a regular basis. Therefore, the above customers are considered loyal.

Inferences:

- The best customers can become early adopters for new products and will help promote the company's brand. So the company can test the new products with this customer segment.
- The loyal customers should be focussed more as we can easily turn them into best customers. It is noted that classic cars and motorcycles are the product lines preferred by this segment. Offer membership or loyalty programs or recommend related products to upsell them and help them become the best customer.
- We should definitely focus on the churn group before we lose them and try to convert them into our regular customers. Send them personalized reactivation campaigns to reconnect, and offer renewals and helpful products to encourage another purchase.
- For the lost customers, we can take a survey to understand their requirements and try to win them back if this does not incur huge impact in the company budget.

Links:https://public.tableau.com/authoring/MRAproject1_16422437659020/Dashboard1#1

End of Presentation