MARKET & RETAIL ANALYSIS

M.ABINAYA

AGENDA :

- To find the buying pattern of customers of an automobile spare parts manufacturer based on the past 3 years transaction data and provide the company with suitable insights about their customers, and recommend customized marketing strategies for different segments of customers (Part- A).

- To identify the most commonly occurring sets of items in the customer orders, and provide recommendations through which a grocery store can increase its revenue by popular combo offers & discounts for customers (Part -B).

EXECUTIVE SUMMARY :

- The 3 year automobile transaction data has 2747 entries with 20 variables.
- The grocery store data has 20641 entries with 3 variables.

CONTENTS OF PPT

Automobile Marketing Strategy Analysis

AUTOMOBILE PARTS MANUFACTURING COMPANY

PROBLEM :

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.

DATA DESCRIPTION

- The given dataset has 2747 entries with 20 variables.
- It has 12 Categorical and 27 Numerical and 1 Datetime variable.
- No duplicates and No Null values present in the dataset.
- Outliers are present in Order number, Quantity ordered, Price each, sales , MSRP column.
- We have to treat outliers in sales column only because others wont cause much effect.
- We noticed that one order number has many different entries with different product codes.
- Manufacturer's Suggested Retail Price(MSRP) for each product code is decided but we found that this is not matching with Price of Each item.

DATA DICTIONARY

Column Name	Description
ORDERNUMBER	This column represents the unique identification number assigned to each order.
QUANTITYORDERED	It indicates the number of items ordered in each order.
PRICEEACH	This column specifies the price of each item in the order.
ORDERLINENUMBER	It represents the line number of each item within an order.
SALES	This column denotes the total sales amount for each order, which is calculated by multiplying the quantity ordered by the price of each item.
ORDERDATE	It denotes the date on which the order was placed.
DAYS_SINCE_LASTORDER	This column represents the number of days that have passed since the last order for each customer. It can be used to analyze customer purchasing patterns.
STATUS	It indicates the status of the order, such as "Shipped," "In Process," "Cancelled," "Disputed," "On Hold," or "Resolved"
PRODUCTLINE	This column specifies the product line categories to which each item belongs.
MSRP	It stands for Manufacturer's Suggested Retail Price and represents the suggested selling price for each item.
PRODUCTCODE	This column represents the unique code assigned to each product.
CUSTOMERNAME	It denotes the name of the customer who placed the order.
PHONE	This column contains the contact phone number for the customer.
ADDRESSLINE1	It represents the first line of the customer's address.
CITY	This column specifies the city where the customer is located.
POSTALCODE	It denotes the postal code or ZIP code associated with the customer's address.
COUNTRY	This column indicates the country where the customer is located.
CONTACTLASTNAME	It represents the last name of the contact person associated with the customer.
CONTACTFIRSTNAME	This column denotes the first name of the contact person associated with the customer.
DEALSIZE	It indicates the size of the deal or order, which are the categories "Small," "Medium," or "Large."

DATA INFORMATION

Head

	ORDERNUMBER	QUANTITYORDERED	PRICEEACH	ORDERLINENUMBER	SALES	ORDERDATE	DAYS_SINCE_LASTORDER	STATUS	PRODUCTLINE	MSRP
0	10107	30	95.70	2	2871.00	2018-02-24	828	Shipped	Motorcycles	95
1	10121	34	81.35	5	2765.90	2018-05-07	757	Shipped	Motorcycles	95
2	10134	41	94.74	2	3884.34	2018-07-01	703	Shipped	Motorcycles	95
3	10145	45	83.26	6	3746.70	2018-08-25	649	Shipped	Motorcycles	95
4	10168	36	96.66	1	3479.76	2018-10-28	586	Shipped	Motorcycles	95

Tail

	ORDERNUMBER	QUANTITYORDERED	PRICEEACH	ORDERLINENUMBER	SALES	ORDERDATE	DAYS_SINCE_LASTORDER	STATUS	PRODUCTLINE	MS
2742	10350	20	112.22	15	2244.40	2019-12-02	2924	Shipped	Ships	
2743	10373	29	137.19	1	3978.51	2020-01-31	2865	Shipped	Ships	
2744	10386	43	125.99	4	5417.57	2020-03-01	2836	Resolved	Ships	
2745	10397	34	62.24	1	2116.16	2020-03-28	2810	Shipped	Ships	
2746	10414	47	65.52	9	3079.44	2020-05-06	2772	On Hold	Ships	

Info

```
<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
```

shape

```
(2747, 20)
```

Null values

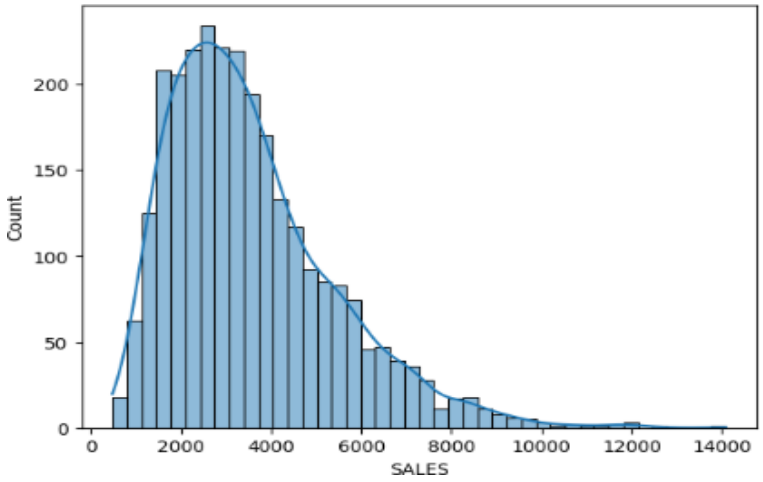
```
ORDERNUMBER           0
QUANTITYORDERED       0
PRICEEACH             0
ORDERLINENUMBER       0
SALES                 0
ORDERDATE             0
DAYS_SINCE_LASTORDER  0
STATUS                0
PRODUCTLINE           0
MSRP                  0
PRODUCTCODE           0
CUSTOMERNAME          0
PHONE   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

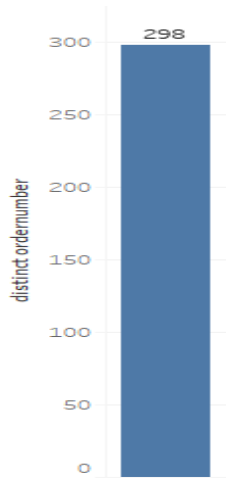
Data summary

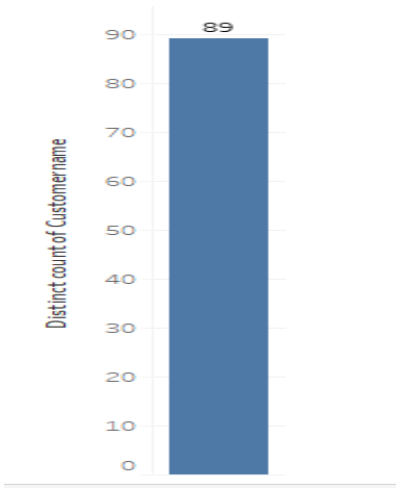
	ORDERNUMBER	QUANTITYORDERED	PRICEEACH	ORDERLINENUMBER	SALES	DAYS_SINCE_LASTORDER	MSRP
count	2747.000000	2747.000000	2747.000000	2747.000000	2747.000000	2747.000000	2747.000000
mean	10259.761558	35.103021	101.098951	6.491081	3553.047583	1757.085912	100.691664
std	91.877521	9.762135	42.042548	4.230544	1838.953901	819.280578	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

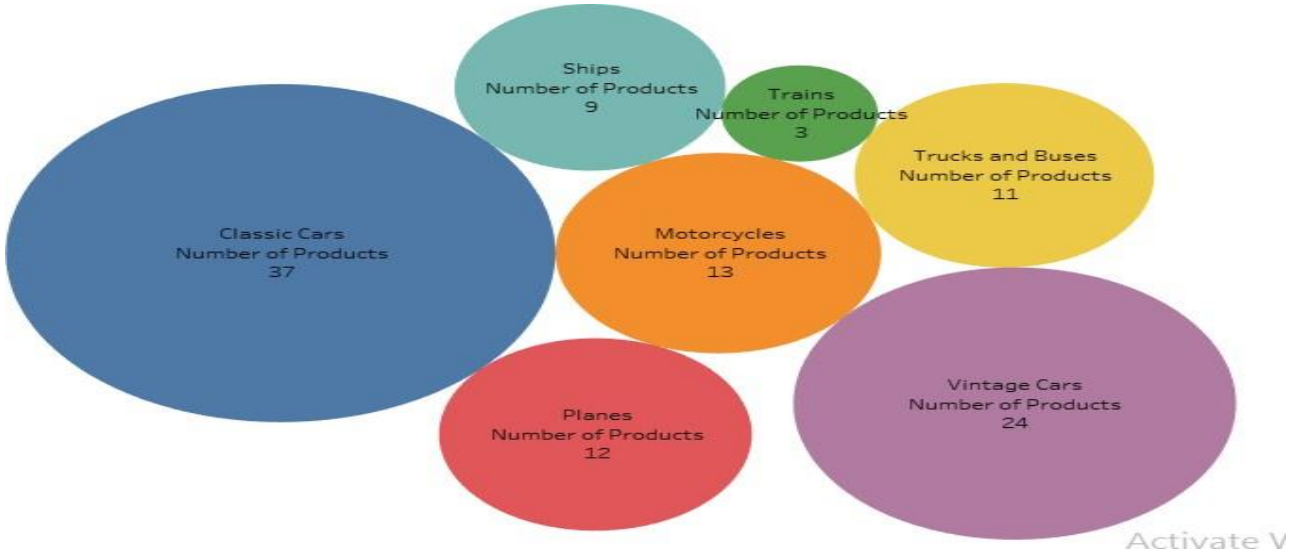
EDA

UNIVARIATE ANALYSIS :

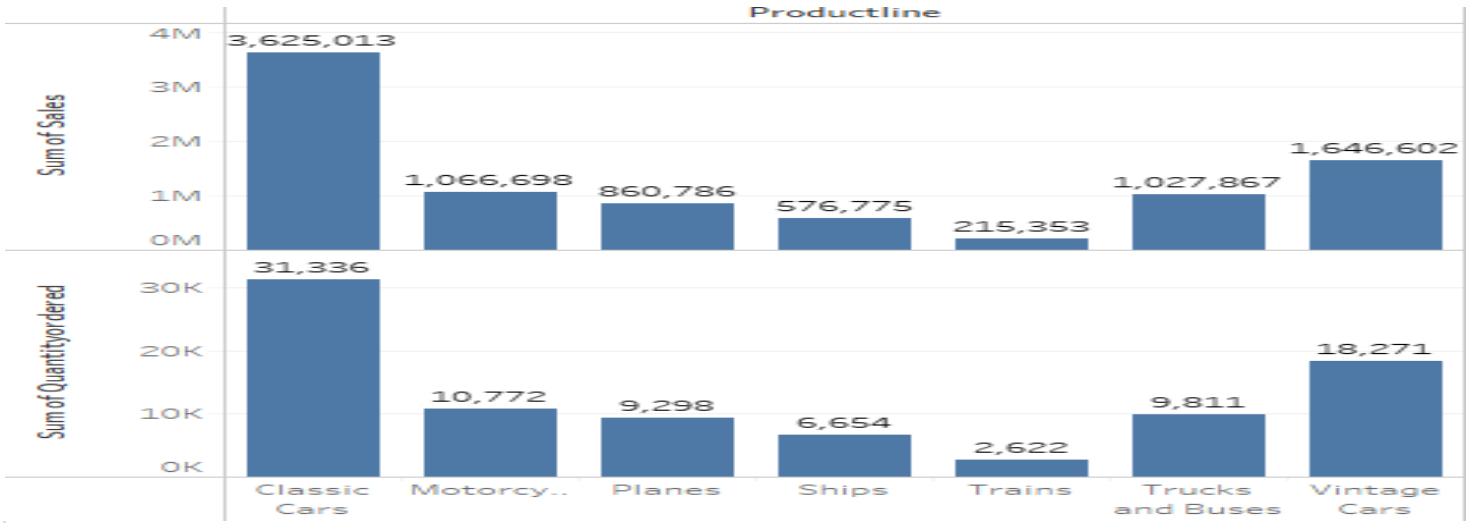


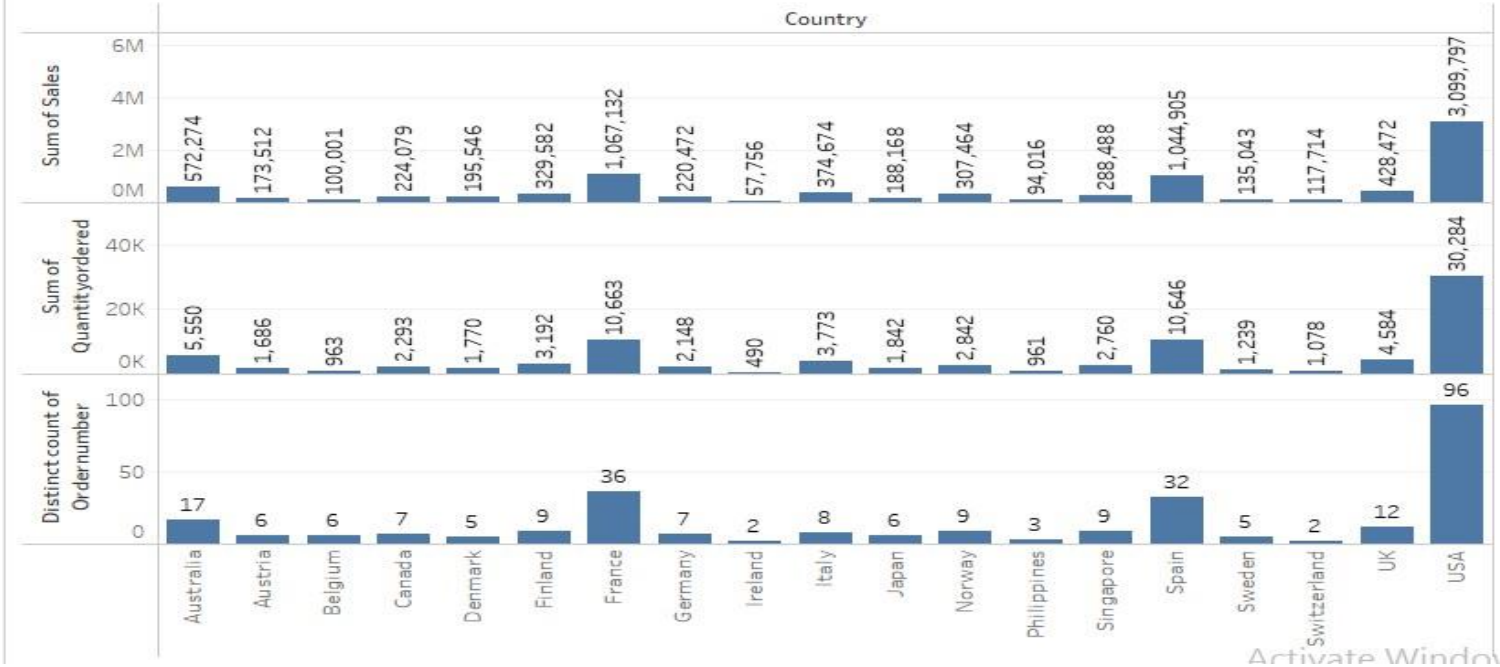




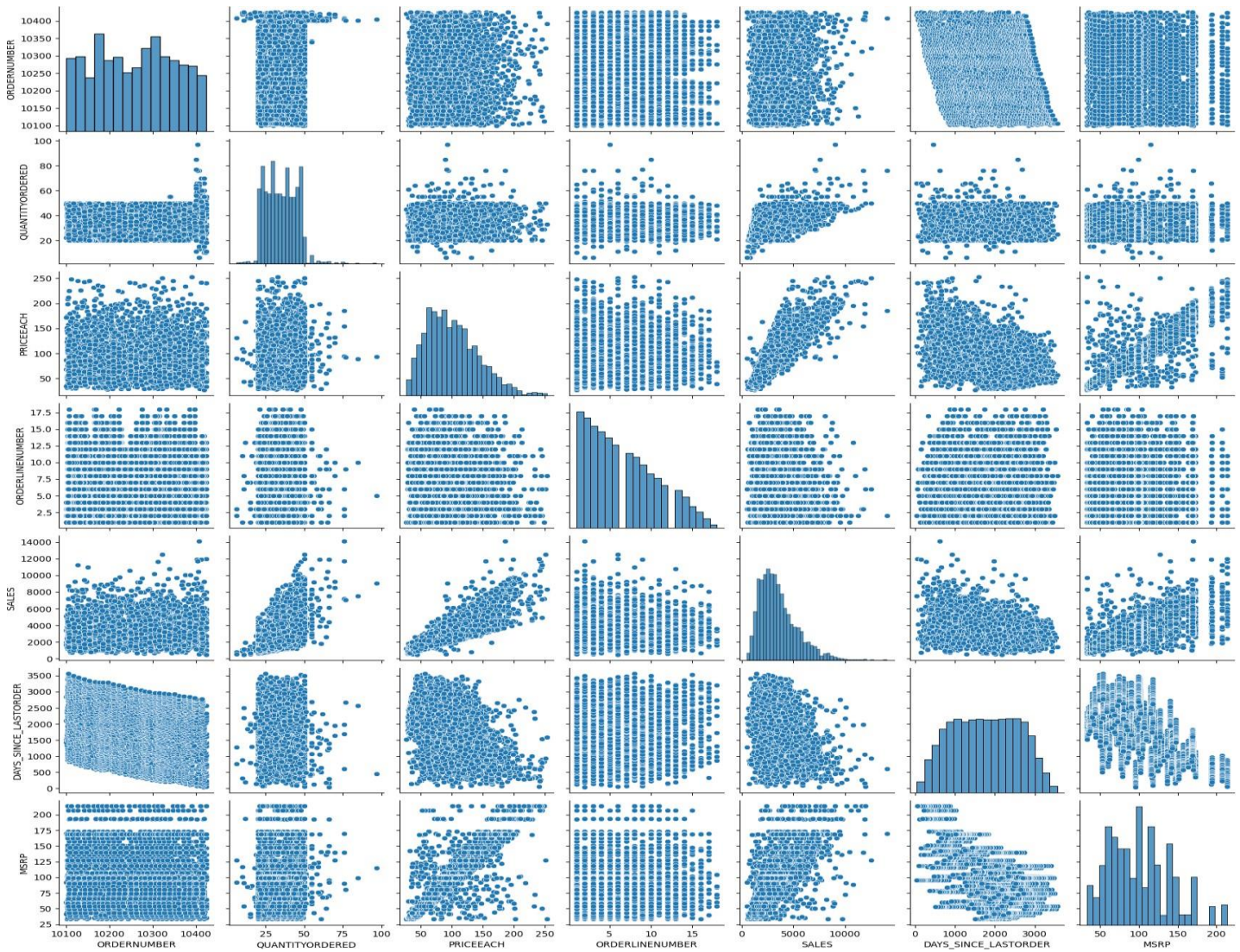


BIVARIATE ANALYSIS :



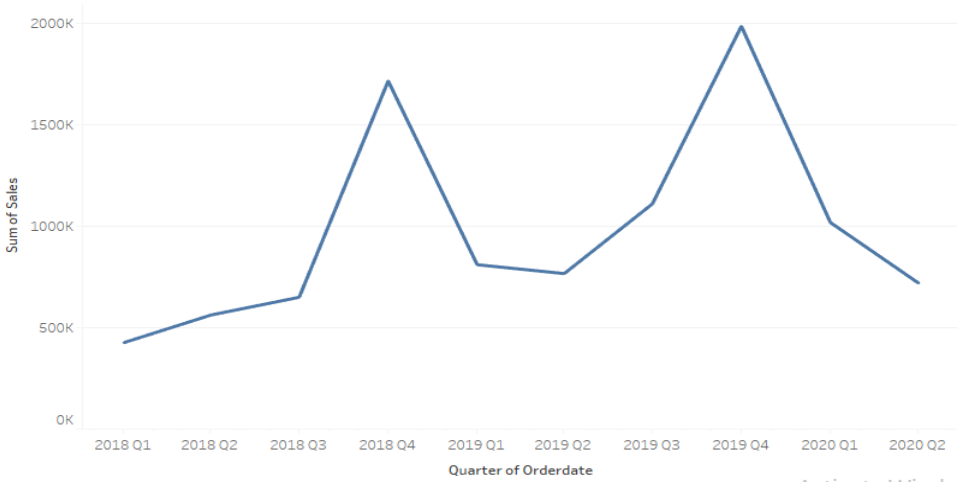


MULTIVARIATE ANALYSIS :

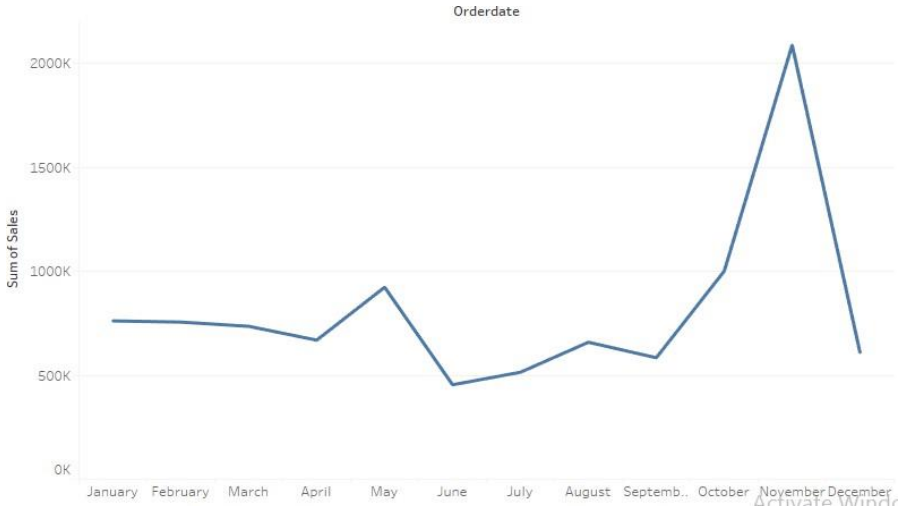


TRENDS IN SALES

- Higher sales occurred in 2018 Q4 and 2019 Q4.So we can say that achieved higher sales in Q4 .

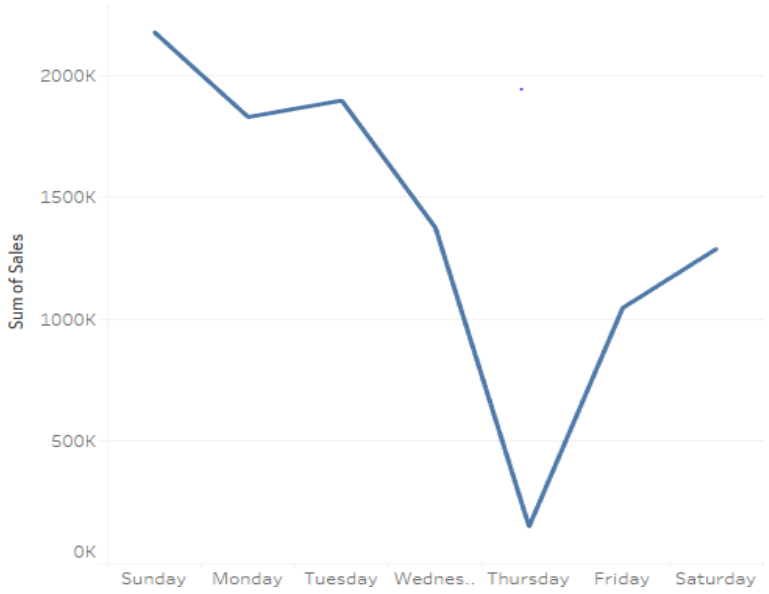


- Generally SALES is higher in the month of NOVEMBER.

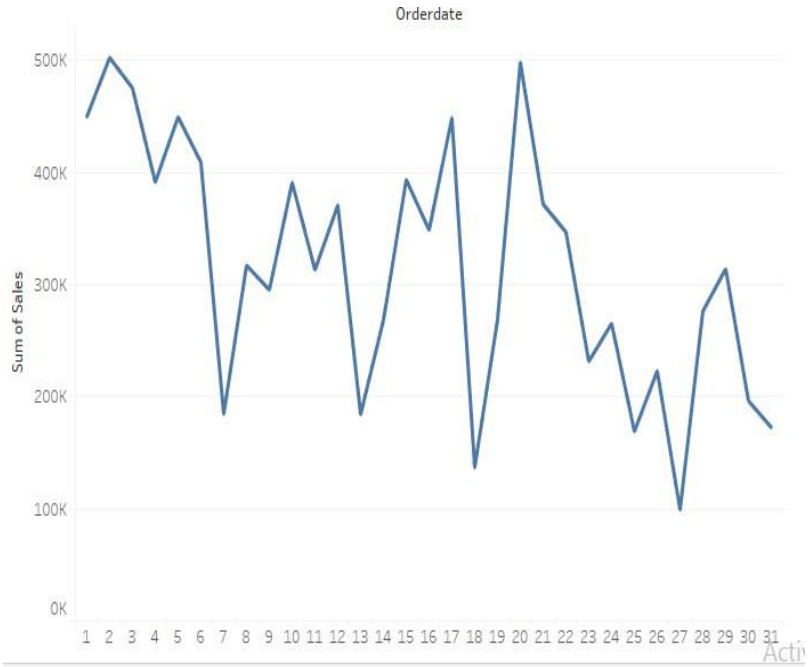


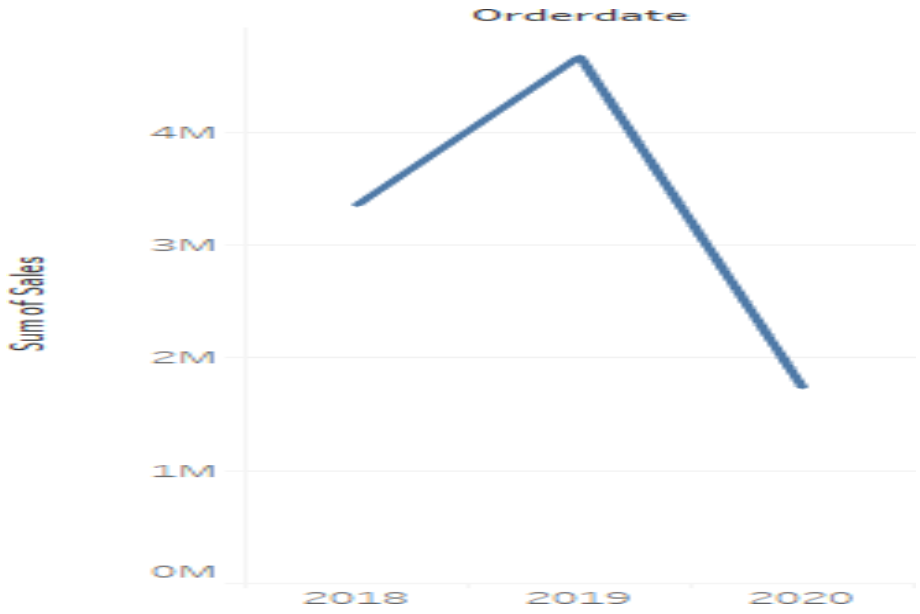
❖ In general ,

Sales is higher in Sunday and lower in Thursday



❖ Generally , each month's higher sales achieved in 2nd and 20th of every month.





- Maximum amount of sales achieved in the year 2019 and then decreases gradually due to covid pandemic in 2020 .

Tableau link :

https://public.tableau.com/app/profile/abinaya.m8348/viz/MRA_16941822183290/Sheet9

INSIGHTS FROM EDA

- Amount of sales is higher among Classic car parts > Vintage cars > Motor cycle > Trucks & Buses > Planes > ships > Trains.
- Size of deal received MEDIUM > SMALL > LARGE.
- Maximum amount of sales done in Madrid city(Spain) i.e) $10,82,551.
- Maximum sales occurred between $2000 to $6000.
- There are 298 distinct orders with 89 customers and more sales occured in Q4 Quarter of every year.
- After the year 2019 , the sales gradually decreases.
- The company has a total of 109 products in different productline.
- Maximum amount of sales occurred in USA then Spain whereas minimum sales in Ireland.
- Maximum number orders received from USA and France whereas minumum number of orders from Ireland and switzerland.

RFM

- RFM Analysis is a marketing technique used to quantitatively rank and group customers based on the recency, frequency, monetary total of their recent transactions to identify the best customers and perform targeted marketing campaigns.
- Recency : The time since last order placed calculated from

  Today's date – Order date
- Frequency : How often the orders placed by a customer is calculated from count of order number.
- Monetary : Total sales . It can be calculated by

  Quantity ordered * price each

PARAMETERS USED

- Here I am using KNIME Tool for RFM Analysis .
- I have grouped the data using CUSTOMER NAME.
- For Monetary , SALES column used
- For Recency , order date column used

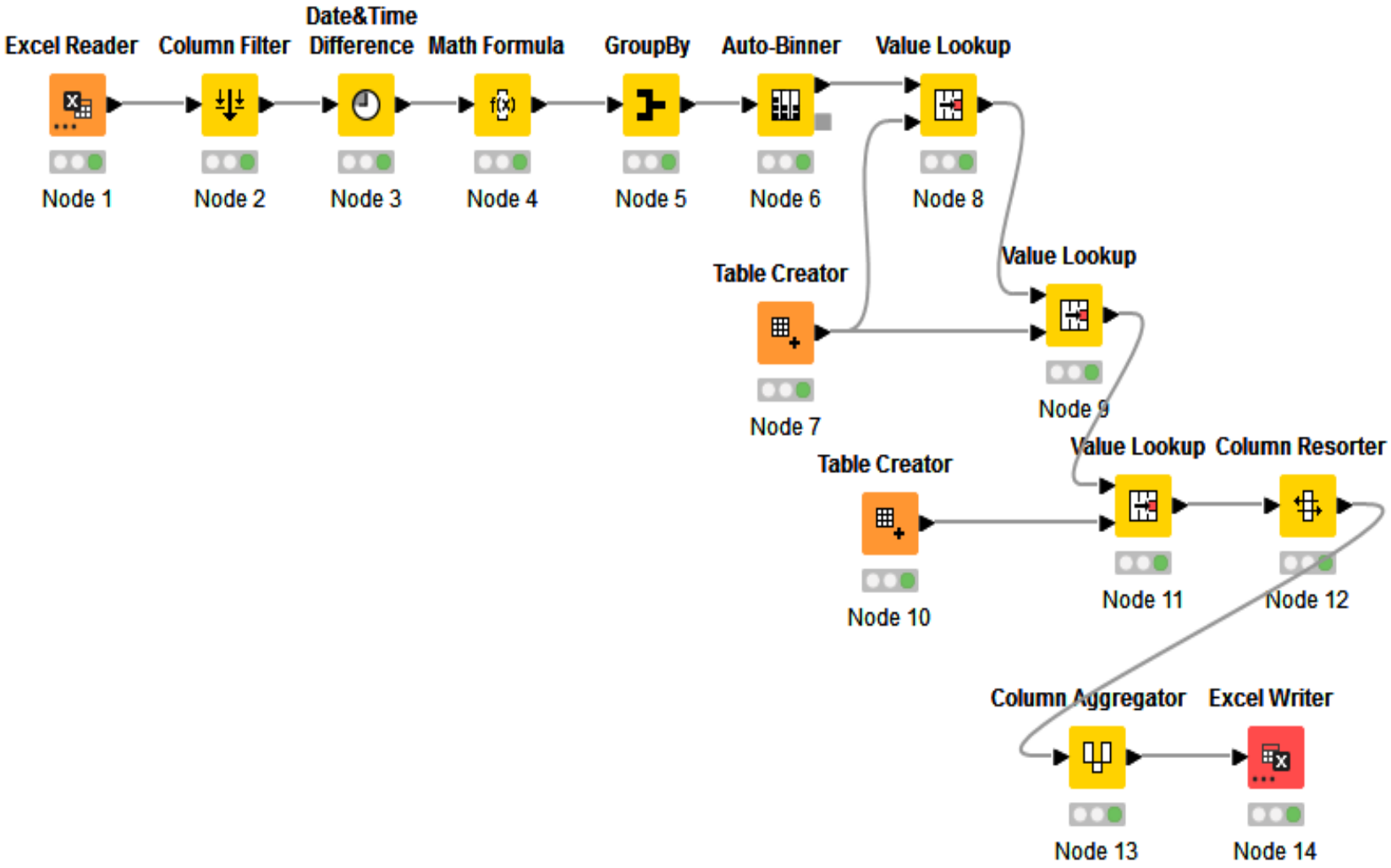
  Today's date – Order date
- For Frequency , Count of Order number is used.
- Creating 4 bins in the range of (0,0.25,0.5,0.75,1) for RFM Analysis then it is assumed to be LOW (L) ,CHURN (C),MEDIUM (M),HIGH (H) respectively.

RFM SEGMENTATION

Customers has been segmented as

- BEST CUSTOMERS are High Recency, High Monitary and High Frequency.
- LOYAL CUSTOMERS are Medium Recency , High and Medium Frequency and High and Medium Monitary .
- Customers are at VERGE OF CHURN are Churn Regency ,Churn and Low Frequency and High Monitary.
- LOST CUSTOMERS are Low Recency, Low Frequency and Churn and Churn Monitary.

KNIME WORKFLOW



FINAL OUTPUT HEAD

Table "default" - Rows: 89 | Spec - Columns: 16 | Properties | Flow Variables

Row ID	CUSTOMERNAME	O...	Q...	PR...	SALES	...	...	L ...	monetary	...	...	...	R...	F...	M...	C...
Row88	giftsbymail.co.uk	26	895	88.979	78,240.84	5	26	1408	78,240.84	Bin 2	Bin 3	Bin 2	C	C	C	CCC
Row87	West Coast Collectables Co.	13	511	88.308	46,084.64	4	13	1684	46,084.64	Bin 1	Bin 4	Bin 1	L	L	L	LLL
Row86	Volvo Model Replicas, Co	19	647	119.289	75,754.88	5	19	1390	75,754.88	Bin 1	Bin 3	Bin 2	C	L	C	CLC
Row85	Vitachrome Inc.	25	787	106.179	88,041.26	4	25	1404	88,041.26	Bin 2	Bin 3	Bin 3	C	C	M	CCM
Row84	Vida Sport, Ltd	31	1078	112.075	117,713.56	1	31	1471	117,713.56	Bin 3	Bin 4	Bin 3	L	M	M	LMM

RFM ANALYSIS

BEST CUSTOMERS :

Row ID	CUSTOMERNAME	O...	Q...	PR...	SALES				monetary	...	..	...	R...	F...	M...	▲ ..
Row26	Danish Wholesale Imports	36	1315	108.038	145,041.6	5	36	1242	145,041.6	Bin 4	Bin 1	Bin 4	H	H	H	HHH
Row32	Euro Shopping Channel	259	9327	97.383	912,294.11	7	259	1196	912,294.11	Bin 4	Bin 1	Bin 4	H	H	H	HHH
Row41	L'ordine Souveniers	39	1280	111.147	142,601.33	6	39	1217	142,601.33	Bin 4	Bin 1	Bin 4	H	H	H	HHH
Row43	La Rochelle Gifts	53	1832	97.046	180,124.9	6	53	1196	180,124.9	Bin 4	Bin 1	Bin 4	H	H	H	HHH
Row53	Mini Gifts Distributors Ltd.	180	6366	102.696	654,858.06	6	180	1198	654,858.06	Bin 4	Bin 1	Bin 4	H	H	H	HHH
Row63	Reims Collectables	41	1433	94.343	135,042.94	7	41	1258	135,042.94	Bin 4	Bin 1	Bin 4	H	H	H	HHH
Row67	Salzburg Collectables	40	1442	101.398	149,798.63	5	40	1210	149,798.63	Bin 4	Bin 1	Bin 4	H	H	H	HHH
Row72	Souveniers And Things Co.	46	1601	95.189	151,570.98	6	46	1198	151,570.98	Bin 4	Bin 1	Bin 4	H	H	H	HHH
Row78	The Sharp Gifts Warehouse	40	1656	93.376	160,010.27	5	40	1235	160,010.27	Bin 4	Bin 1	Bin 4	H	H	H	HHH

- **Euro shopping channel**
- **Mini Gifts Distributors Ltd**
- **La Rochelle Gifts**
- **The Sharp Gifts Warehouse**
- **Souveriers and Things Co**
- **Lordine Souveriers**

LOYAL CUSTOMERS :

CUSTOMERNAME	ORDER	QUANT	PRICEE	SALES	PRODU	PRODU	recency	moneta	ORDER	recency	moneta	RECEN	FREQUE	MONET	Concat
Anna's Decorations, Ltd	46	1469	106.4241	153996.1	4	46	1279	153996.1	Bin 4	Bin 2	Bin 4	M	H	H	MHH
Australian Collectors, Co.	55	1926	104.5902	200995.4	5	55	1380	200995.4	Bin 4	Bin 2	Bin 4	M	H	H	MHH
Dragon Souveniers, Ltd.	43	1524	113.1056	172989.7	6	43	1286	172989.7	Bin 4	Bin 2	Bin 4	M	H	H	MHH
Mini Creations Ltd.	35	1140	95.12914	108951.1	6	35	1341	108951.1	Bin 4	Bin 2	Bin 3	M	H	M	MHM
Muscle Machine Inc	48	1775	111.1508	197736.9	4	48	1378	197736.9	Bin 4	Bin 2	Bin 4	M	H	H	MHH
Oulu Toy Supplies, Inc.	32	1110	95.11875	104370.4	5	32	1317	104370.4	Bin 3	Bin 2	Bin 3	M	M	M	MMM
Scandinavian Gift Ideas	38	1359	97.59737	134259.3	6	38	1285	134259.3	Bin 4	Bin 2	Bin 4	M	H	H	MHH
Suominen Souveniers	30	1031	110.405	113961.2	4	30	1342	113961.2	Bin 3	Bin 2	Bin 3	M	M	M	MMM
Technics Stores Inc.	34	1179	104.9141	120783.1	5	34	1343	120783.1	Bin 4	Bin 2	Bin 4	M	H	H	MHH
Toys of Finland, Co.	30	1051	105.7523	111250.4	5	30	1308	111250.4	Bin 3	Bin 2	Bin 3	M	M	M	MMM
Toys4GrownUps.com	30	1060	97.22467	104562	4	30	1336	104562	Bin 3	Bin 2	Bin 3	M	M	M	MMM

- **Anna’s Decorations Ltd**
- **Australian Collectors Co**
- **Dragon Souveniers Ltd**
- **Scandinavian Gift Ideas**
- **Muscle Machine Inc**
- **Technics Stores Inc**

VERGE OF CHURN CUSTOMERS :

CUSTOMERNAME	ORDER	QUANT	PRICEE	SALES	PRODU	PRODU	recency	monet	ORDER	recency	monet	RECEN	FREQU	MONET	Concat
AV Stores, Co.	51	1778	91.08451	157807.8	4	51	1392	157807.8	Bin 4	Bin 3	Bin 4	C	H	H	CHH
Corrida Auto Replicas, Ltd	32	1163	105.175	120615.3	3	32	1408	120615.3	Bin 3	Bin 3	Bin 4	C	M	H	CMH
Land of Toys Inc.	49	1631	104.1206	164069.4	7	49	1394	164069.4	Bin 4	Bin 3	Bin 4	C	H	H	CHH
Online Diecast Creations Co.	34	1248	108.3021	131685.3	3	34	1405	131685.3	Bin 4	Bin 3	Bin 4	C	H	H	CHH
Rovelli Gifts	48	1650	85.67729	137955.7	4	48	1397	137955.7	Bin 4	Bin 3	Bin 4	C	H	H	CHH

- **AV Stores Co**
- **Corrida Auto Replicas Ltd**
- **Land of Toys Inc**
- **Online Diecast creations co**
- **Rovelli Gifts**

LOST CUSTOMERS :

CUSTOMERNAME	ORDER	QUANT	PRICEE	SALES	PRODU	PRODU	recency	moneta	ORDER	recency	moneta	RECEN	FREQUE	MONET	Concat
Auto Assoc. & Cie.	18	637	99.4878	64834.32	3	18	1429	64834.32	Bin 1	Bin 4	Bin 1	L	L	L	LLL
Bavarian Collectables Imports, Co	14	401	84.28929	34993.92	3	14	1455	34993.92	Bin 1	Bin 4	Bin 1	L	L	L	LLL
CAF Imports	13	468	104.9631	49642.05	4	13	1635	49642.05	Bin 1	Bin 4	Bin 1	L	L	L	LLL
Cambridge Collectables Co.	11	357	101.3291	36163.62	5	11	1585	36163.62	Bin 1	Bin 4	Bin 1	L	L	L	LLL
Clover Collections, Co.	16	490	112.87	57756.43	6	16	1454	57756.43	Bin 1	Bin 4	Bin 1	L	L	L	LLL
Daedalus Designs Imports	20	699	95.474	69052.41	2	20	1661	69052.41	Bin 1	Bin 4	Bin 1	L	L	L	LLL
Double Decker Gift Stores, Ltd	12	357	99.10833	36019.04	4	12	1691	36019.04	Bin 1	Bin 4	Bin 1	L	L	L	LLL
Iberia Gift Imports, Corp.	15	589	93.28267	54723.62	4	15	1434	54723.62	Bin 1	Bin 4	Bin 1	L	L	L	LLL
Online Mini Collectables	15	572	94.68067	57197.96	3	15	1460	57197.96	Bin 1	Bin 4	Bin 1	L	L	L	LLL
Osaka Souveniers Co.	20	692	99.914	67605.07	4	20	1610	67605.07	Bin 1	Bin 4	Bin 1	L	L	L	LLL
Signal Collectibles Ltd.	15	514	95.396	50218.51	3	15	1672	50218.51	Bin 1	Bin 4	Bin 1	L	L	L	LLL
West Coast Collectables Co.	13	511	88.30769	46084.64	4	13	1684	46084.64	Bin 1	Bin 4	Bin 1	L	L	L	LLL

- **Auto Assoc & Cie**
- **Bavarian Collectables Imports Co**
- **CAF Imports**
- **Cambridge Collectables Co**
- **Clover Collections Co**
- **Daedalus Designs Imports**

RFM INFERENCES

- Using Recency, frequency & monetary parameters we have grouped our best , loyal, on the verge of churning and lost customers.
- Customers with good recency has been our best customers whereas customer with low recency are lost customers.
- It is important for the company to convert the customers who are on verge of churning into a regular customer or atleast maintain them.
- The company should focus on Loyal customers to make them best customers so that revenue will be increased.
- From the analysis , we know that maximum sales occured in the 4th Quarter so discounts should be given in other Quarters to increase sales.

- Focus on Large sized deal and promote them by giving offers to increase revenue.
- Focus on sales of Trucks & Bus , Train and Ship category spare parts.
- From the analysis, we know that the sales after 2019 has gradually decreased so focus on to increase revenue by giving discounts and credit points etc.
- Focus on cancelled and disputed orders to increase revenue.
- To increase sales in low performing countries, focus on more customers by giving offers such as Festival sales , low EMI etc.

TOOLS USED

- For Data summary , Python is used
- For EDA , Tableau Tool is used

https://public.tableau.com/app/profile/abinaya.m8348/viz/MRA_16941822183290/Sheet9

- For RFM Analysis , KNIME Tool used.