MRA Project - Milestone 1

https://okcredit.in/blog/how-to-start-automobile-parts-manufacturing-business/

By Debsmita Chakraborty
Batch- July 'C'

Table of Contents

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. Our job is to use our magical data science skills to provide them with suitable insights about their data and their customers.
The data given to us is of 2747 rows and 21 columns.

Data Dictionary:

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

- The data has 2747 rows and 20 columns with the below headings. The data can be seen having 2 float type categories , 5 integers and 13 object type 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   object
 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: float64(2), int64(5), object(13)
memory usage: 429.3+ KB
```

Data information

- The data has no duplicates and each column has unique data.

```
Number of duplicate rows = 0
```

- There are no null values in the data as well.

```
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
```

Null Values

Exploratory Data Analysis

	count	mean	std	min	25%	50%	75%	max
ORDERNUMBER	2747.0	10259.761558	91.877521	10100.00	10181.000	10264.00	10334.500	10425.00
QUANTITYORDERED	2747.0	35.103021	9.762135	6.00	27.000	35.00	43.000	97.00
PRICEEACH	2747.0	101.098952	42.042549	26.88	68.745	95.55	127.100	252.87
ORDERLINENUMBER	2747.0	6.491081	4.230544	1.00	3.000	6.00	9.000	18.00
SALES	2747.0	3553.047583	1838.953901	482.13	2204.350	3184.80	4503.095	14082.80
DAYS_SINCE_LASTORDER	2747.0	1757.085912	819.280576	42.00	1077.000	1761.00	2436.500	3562.00
MSRP	2747.0	100.691664	40.114802	33.00	68.000	99.00	124.000	214.00

Data Description

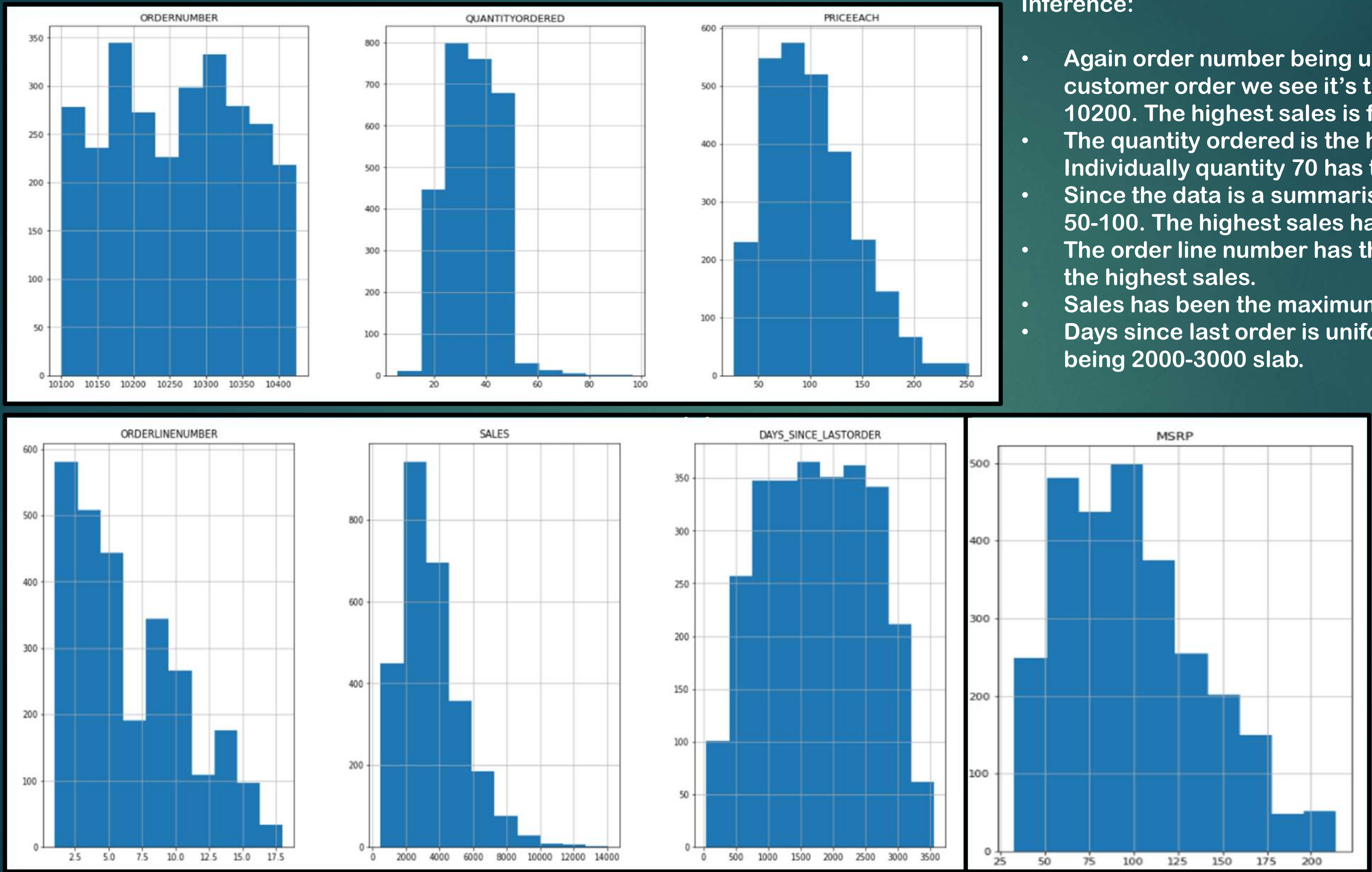
Inferences:

- The data has count of 2747 which confirms no missing values.
- The highest mean is for order number which is a unique field to differentiate orders placed. Considering sales our target we see it has a mean of 3553.05 and minimum value of 482.13 which indicates we need to check outliers. The difference is high and we need to investigate the same.
- Quantity ordered seems to be having good numbers with minimum order of 6 and maximum 97. Price and days since last order also seem to be inline without much difference with less difference in minimum and maximum values.
- MSRP(Manufactures suggested retail price) also seem to be having a healthy difference between mean and maximum values.

Product Line	Count of Products sold
Classic Cars	949
Vintage Cars	579
Motorcycles	313
Planes	304
Trucks and Buses	295
Ships	230
Trains	77
Grand Total	**2747**

- The maximum cars sold is classic cars and the least is Trains.

Univariate Analysis



Histogram

Inference:

- Again order number being unique identification number for customer order we see it's the highest in the range if 10150-10200. The highest sales is for order#10407 .
- The quantity ordered is the highest in the range of 20-30. Individually quantity 70 has the highest sales.
- Since the data is a summarised representation maximum price is 50-100. The highest sales has been for Rs. 185.3.
- The order line number has the highest range from 2.5-5. 2 having the highest sales.
- Sales has been the maximum in the range 2000-4000.
- Days since last order is uniformly distributed dated. Maximum lag being 2000-3000 slab.

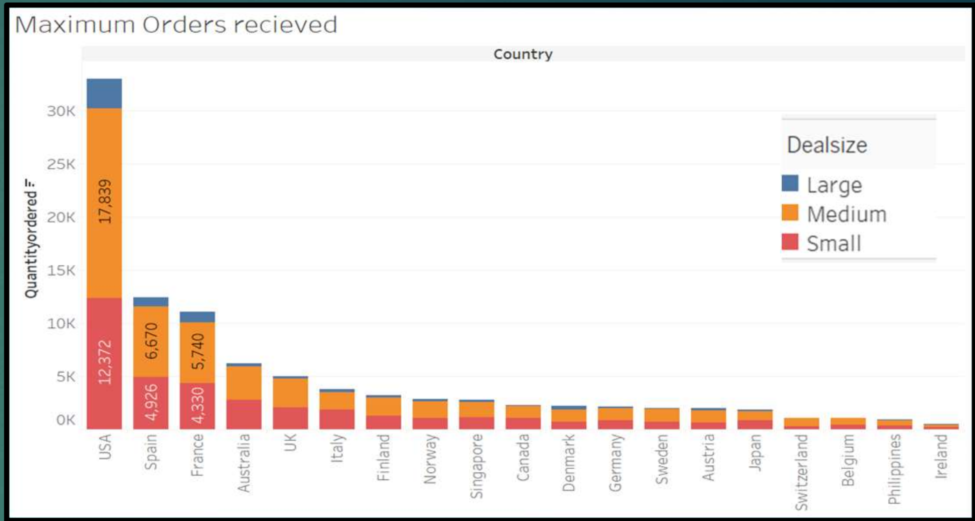
- The MSRP is maximum at 100. The lowest price set by the retailer is at 175-200.

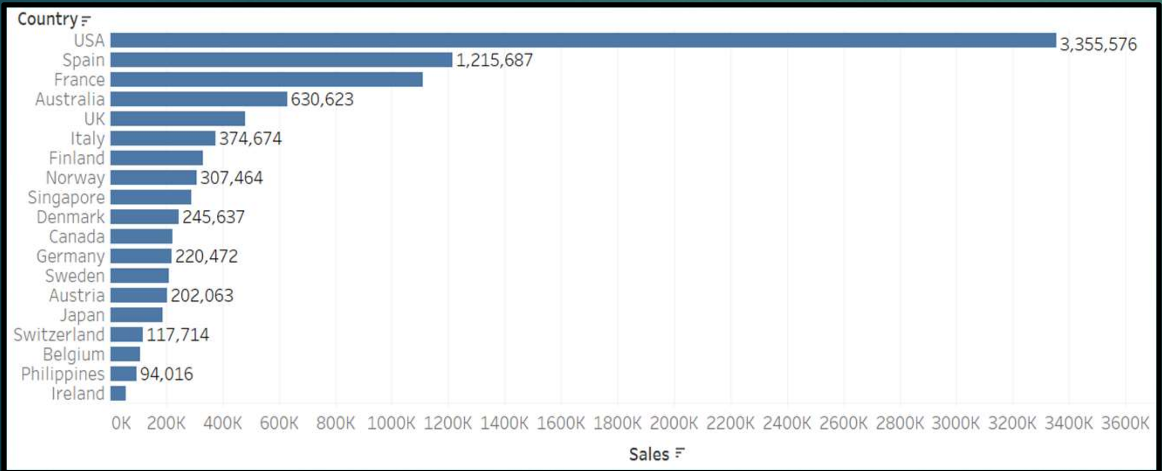
Bivariate and Multivariate Analysis

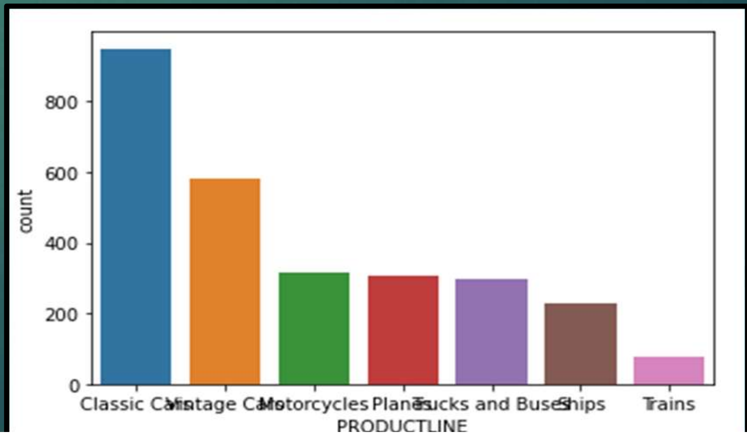
Inference:

- Country wise product line we see sales is the highest for USA and lowest for Ireland. Maximum orders is received from USA and least from Ireland.
- The maximum sales is of Classic Cars , second being vintage cars and the minimum is of Trains.

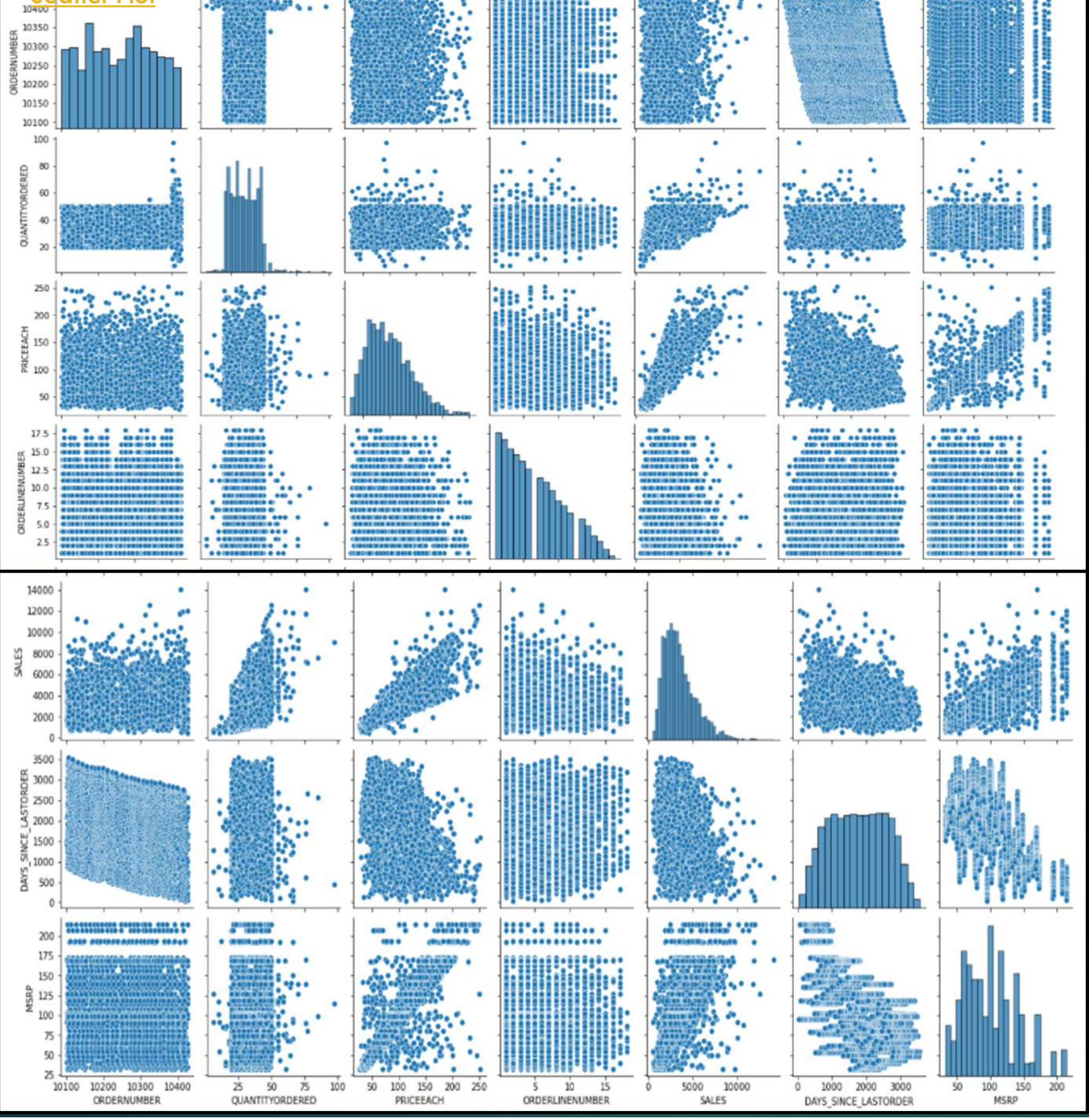
	Productline						
Country	Classic Cars	Motorcycles	Planes	Ships	Trains	Trucks and Bu..	Vintage Cars
USA	1,267,891	457,496	322,753	195,290	69,254	381,612	661,281
Spain	476,165	74,635	89,986	124,460	43,370	177,557	229,515
France	388,951	226,390	108,156	66,487	27,341	116,982	176,610
Australia	193,086	89,969	74,854	4,160	1,681	77,319	189,555
UK	159,378	40,803	41,164	72,959	12,636	28,143	123,799
Italy	128,577	7,568	98,186	17,704	6,275	5,915	110,451
Finland	153,552	47,867	34,375	29,808	5,117	40,479	18,383
Norway	134,787	51,769	29,501		11,310	37,076	43,021
Singapore	132,890	4,176		14,156	13,279	89,028	34,960
Denmark	157,182		7,586	38,697	11,476	9,589	21,106
Canada	61,623	4,177	25,510	40,309		51,946	40,513
Germany	148,315	7,498	23,001	5,501	5,043	10,178	20,936
Sweden	69,088	15,567	8,900	30,916	3,808	47,931	33,804
Austria	101,459	26,048	17,860	9,025		20,473	27,197
Japan	47,271	26,536	49,177	18,860	3,524	13,349	29,450
Switzerland	117,714						
Belgium	20,137		5,625	31,708	9,017		41,926
Philippines	53,112	18,062	20,907				1,935
Ireland	31,689	4,953	11,784		3,113	3,983	2,234







Scatter Plot



	ORDERNUMBER	QUANTITYORDERED	PRICEEACH	ORDERLINENUMBER	SALES	DAYS_SINCE_LASTORDER	MSRP	TotalPrice
ORDERNUMBER	1.000000	0.067110	-0.003369	-0.054300	0.037289	-0.251476	-0.013910	0.037290
QUANTITYORDERED	0.067110	1.000000	0.010161	-0.016295	0.553359	-0.021923	0.020551	0.553359
PRICEEACH	-0.003369	0.010161	1.000000	-0.052646	0.808287	-0.397091	0.778393	0.808287
ORDERLINENUMBER	-0.054300	-0.016295	-0.052646	1.000000	-0.057414	0.046615	-0.020956	-0.057414
SALES	0.037289	0.553359	0.808287	-0.057414	1.000000	-0.334274	0.634849	1.000000
DAYS_SINCE_LASTORDER	-0.251476	-0.021923	-0.397091	0.046615	-0.334274	1.000000	-0.524285	-0.334274
MSRP	-0.013910	0.020551	0.778393	-0.020956	0.634849	-0.524285	1.000000	0.634849
TotalPrice	0.037290	0.553359	0.808287	-0.057414	1.000000	-0.334274	0.634849	1.000000

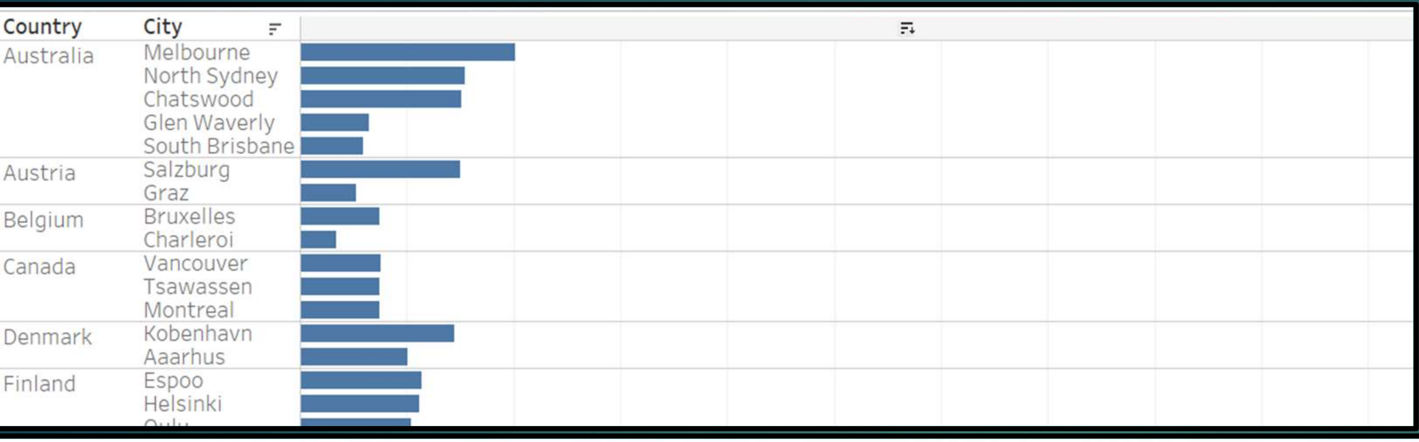
Correlation

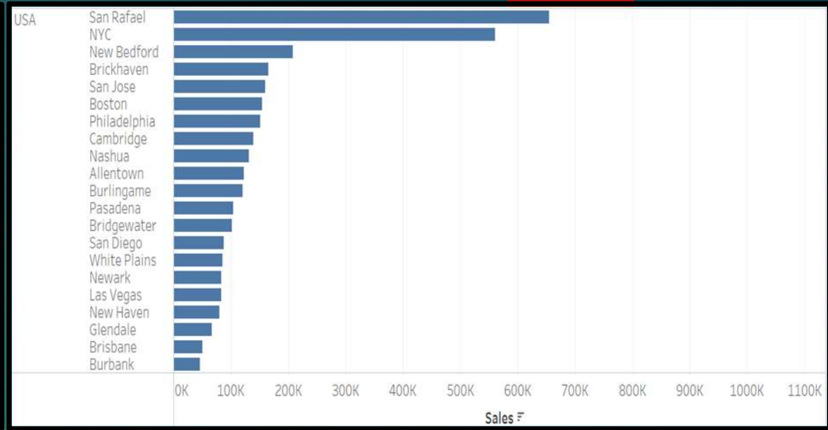
Inference:

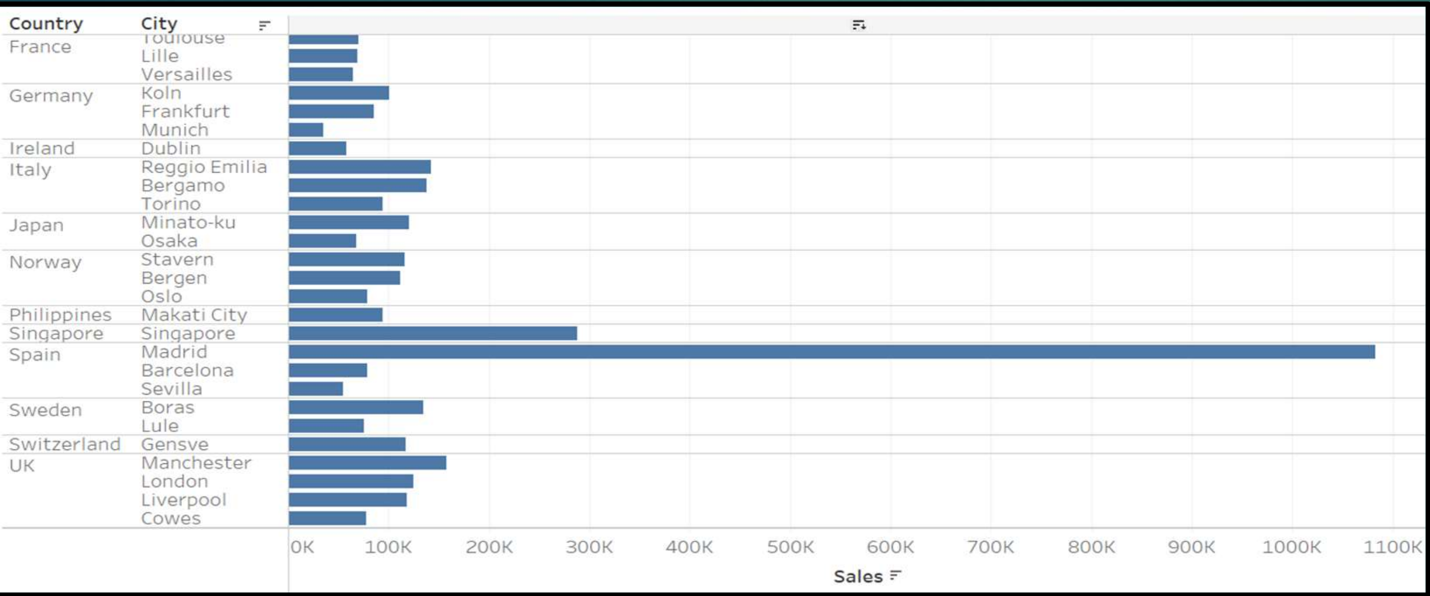
- The data is corelated to each other with normal distribution.
- A good correlation score is -1 to 1 which is valid in the data we have.
- The highest correlation is between price of each order and total price.
- The result also confirms Sales is not hindered by most of the data and they are uniformly distributed.

Sales Analysis with different parameters:

- ***Country to City wise sales.***





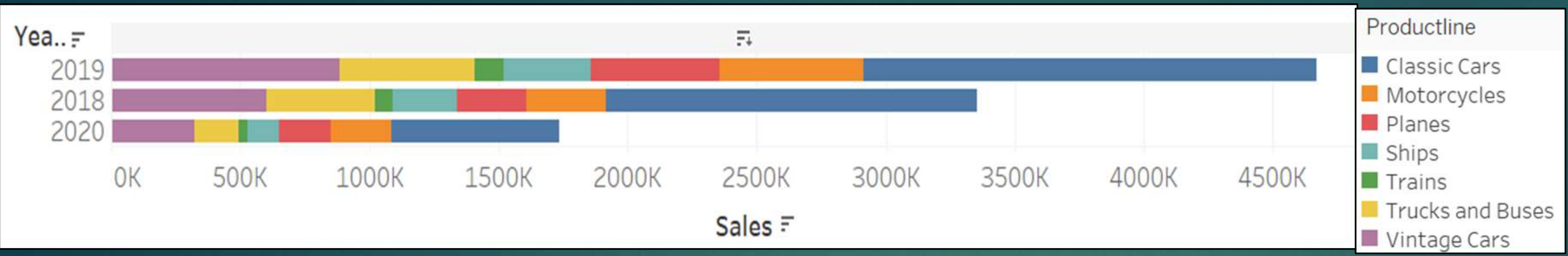


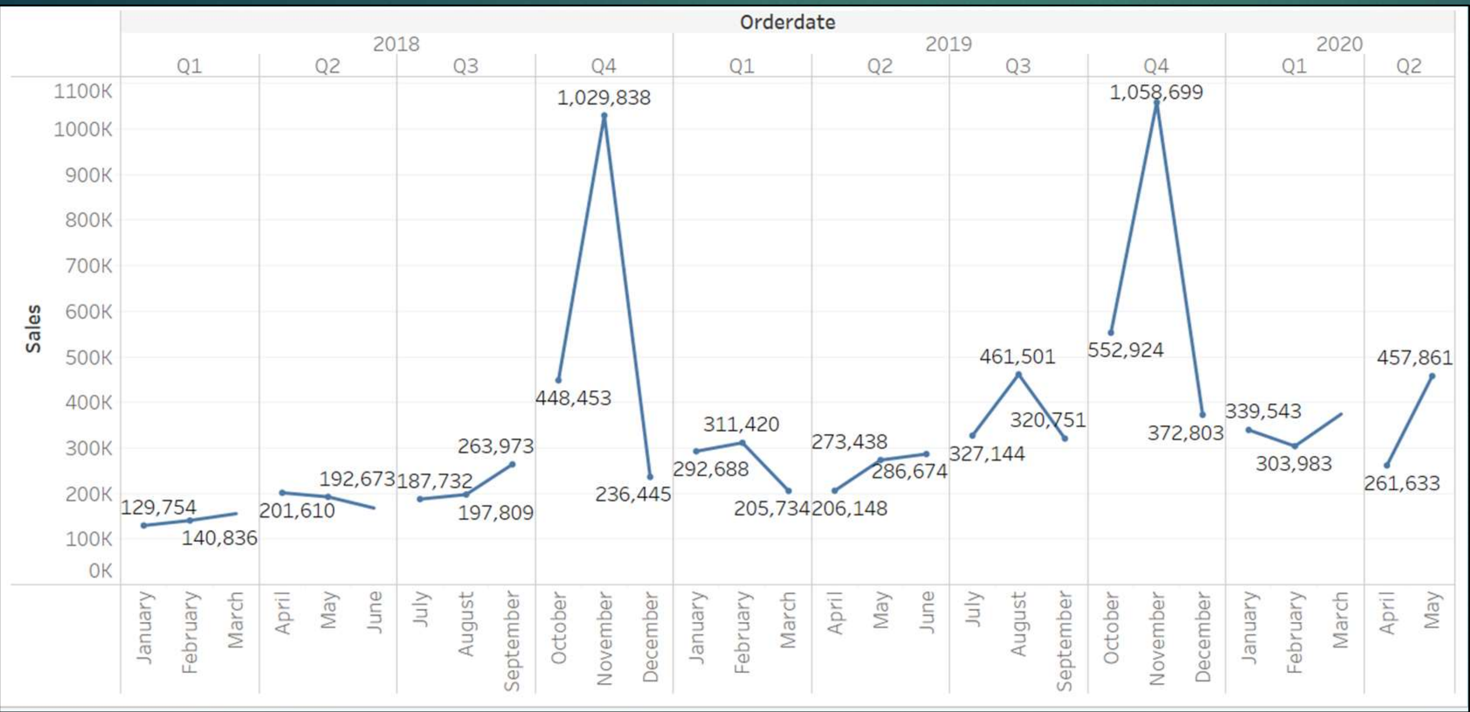
Inference:

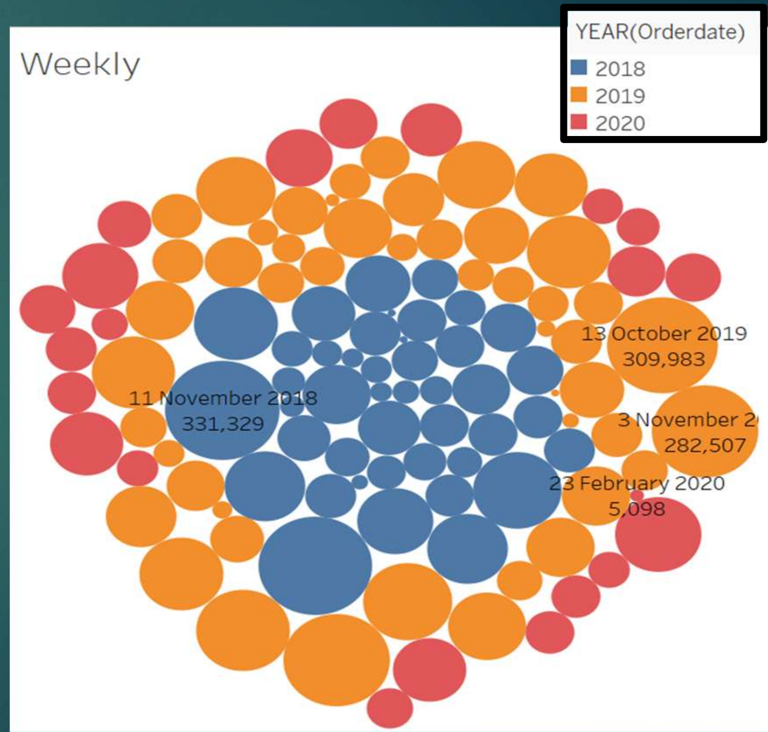
- Highest sales is in the USA in the city San Rafael.
- Lowest is in France in the city Nantes.
- There is a total of 71 cities in the data.
- Dimensions
- City has 71 members on this sheet
  Members: Aaarhus; Cambridge; Liverpool; Osaka; San Jose; ...
  Country has 19 members on this sheet
  Members: Denmark; France; Japan; UK; USA; ...
- Measures

Sum of Sales ranges from 33,440 to 1,082,551 on this sheet.

- ***Yearly-Quarterly-Monthly-Weekly wise sales.***







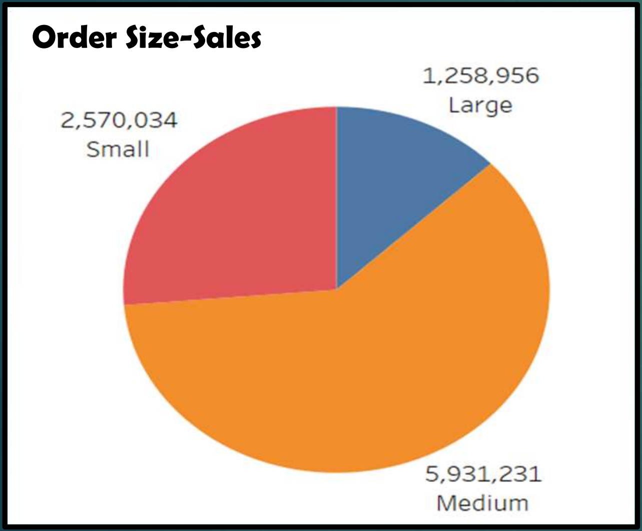
Inference:

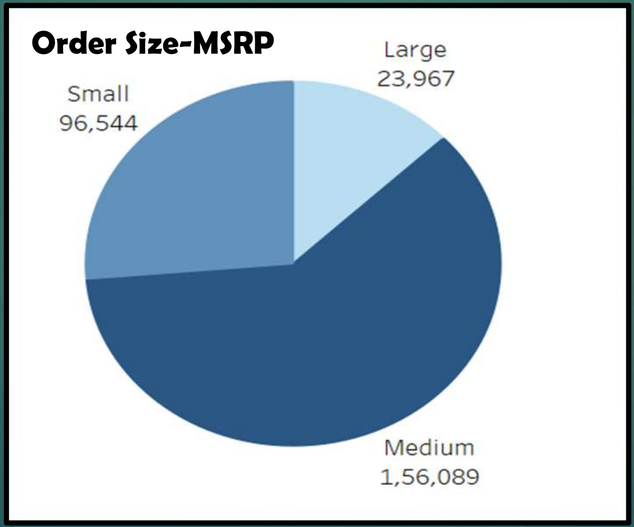
- Sales is the maximum in the year 2019 for classic cars.
- Data represents and confirms classic cars being more in demand and least being trains.
- Quarterly sales is highest in Q4 which is year 2019.
- The data shows month wise October and November 2019 has the highest sales and lowest is January 2018.
- Weekly we see 13th October 2019 having highest sales and lowest being 2nd June 2019.

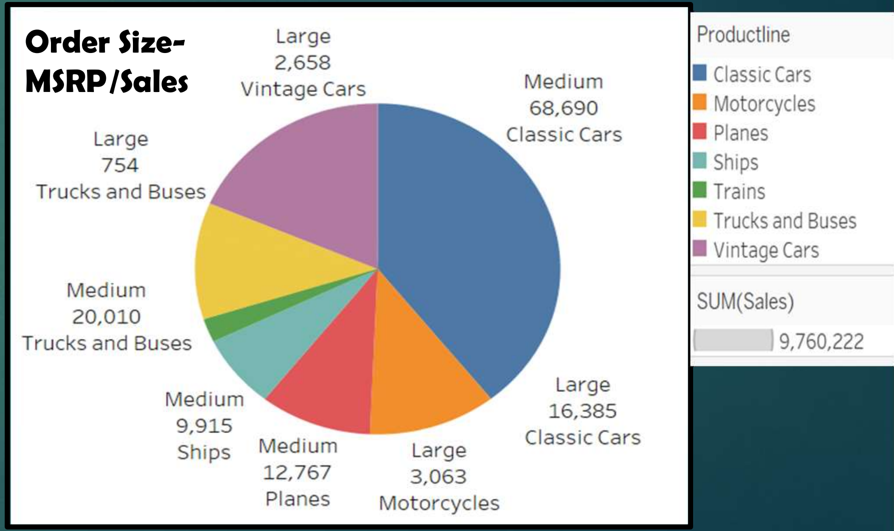
- ***Sales with deal size and manufactures suggested retail price.***

Inference:

- We see that deal size medium has maximum sales and the price that manufactures decided is also the maximum for medium size deal.
- We in the third figure that the classic cars had the maximum retail price as well as the maximum deal size. Least being deal size large for trucks and buses.



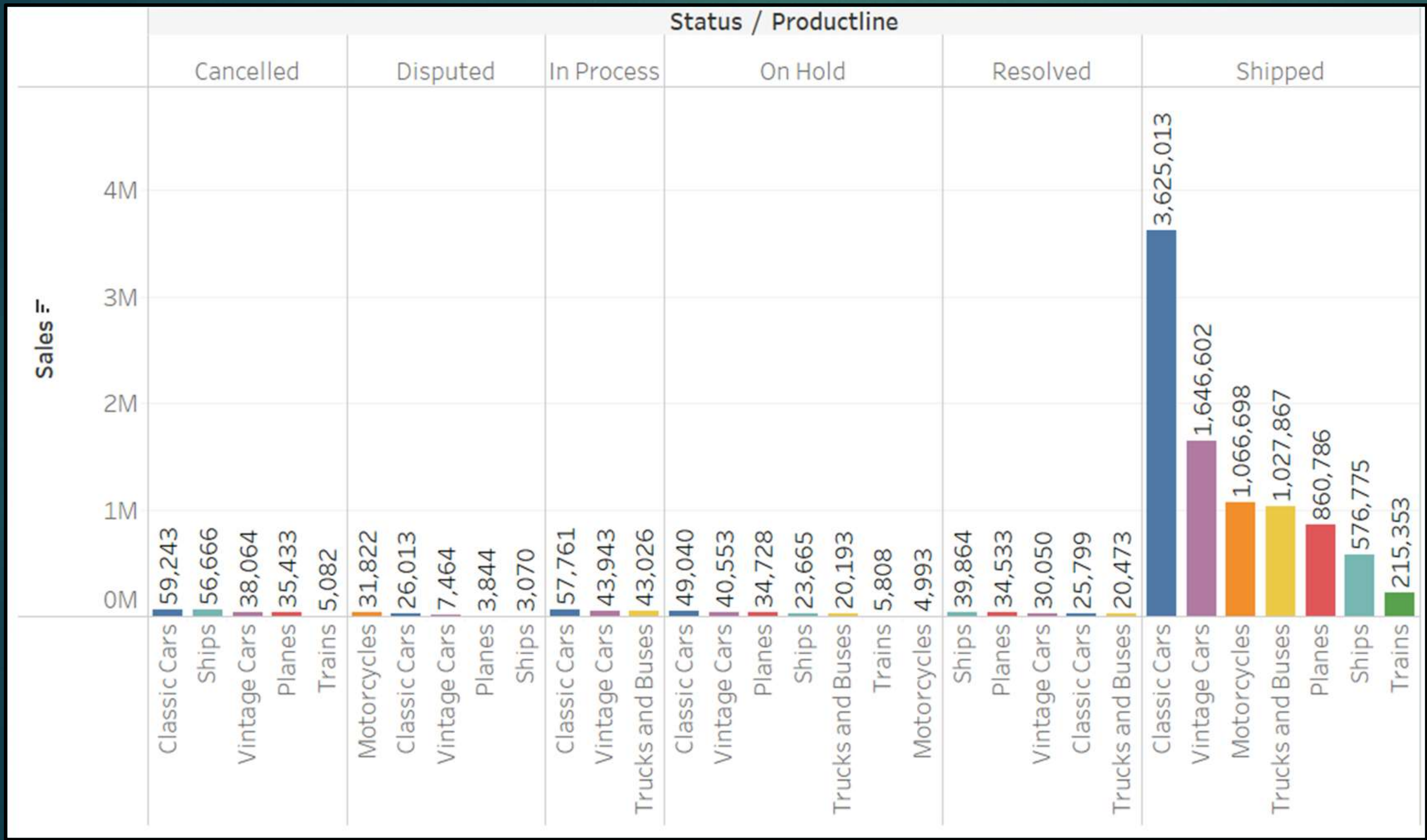


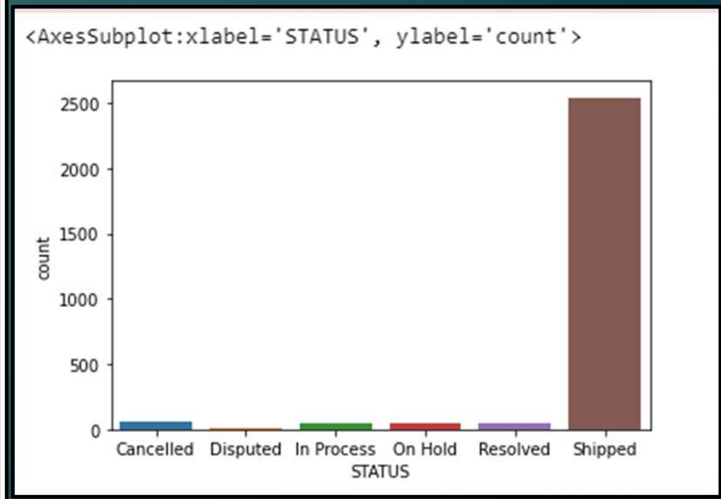


- ***Sales with product status and productline***

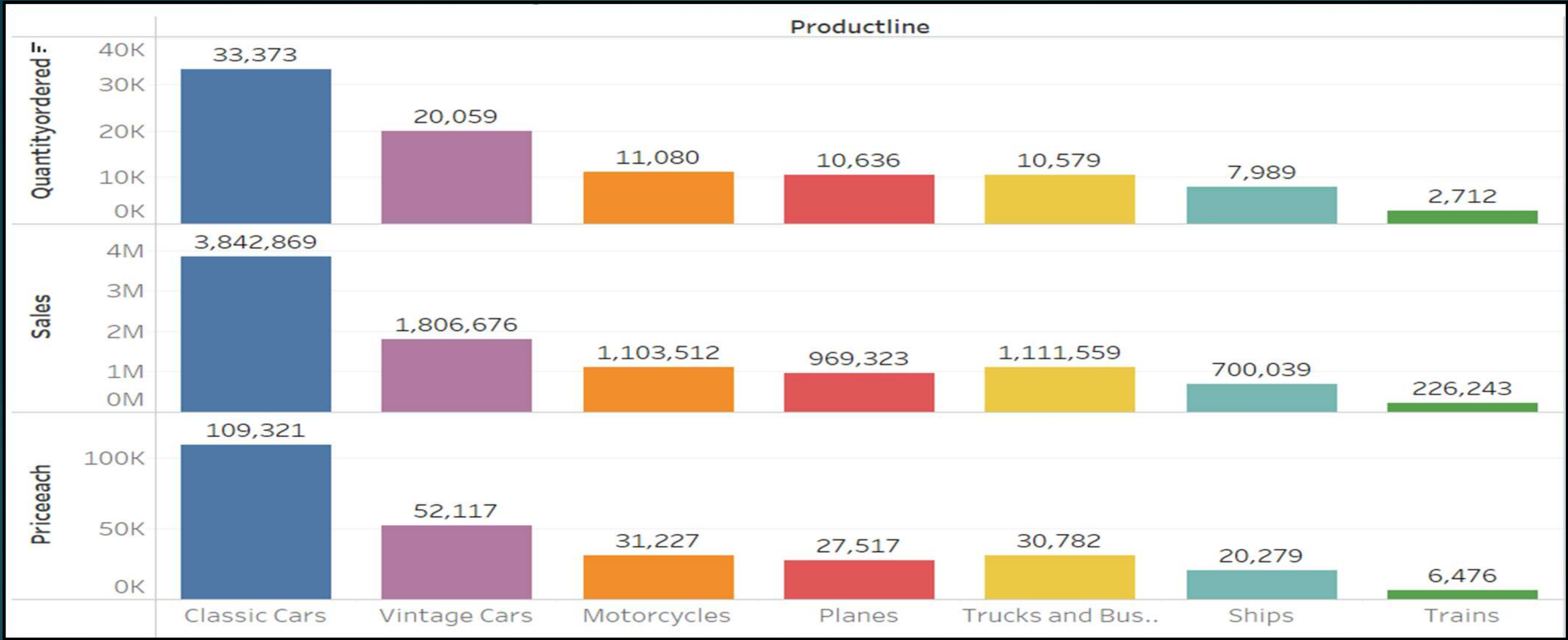
Inference:

- **We see that maximum status of the product is shipped. Resolved is of a lesser value than cancelled and disputed which will be analyzed further.**





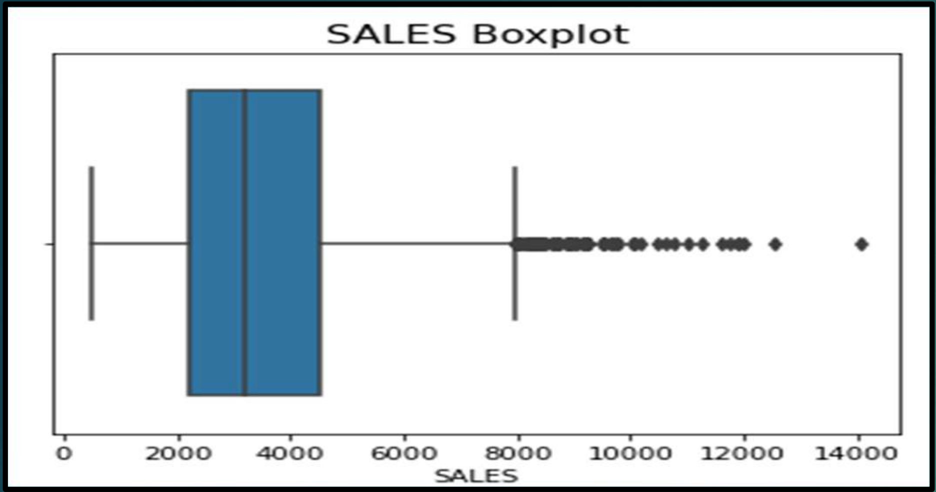
- ***Price of the product and quantity ordered compared to sales.***



Inference:

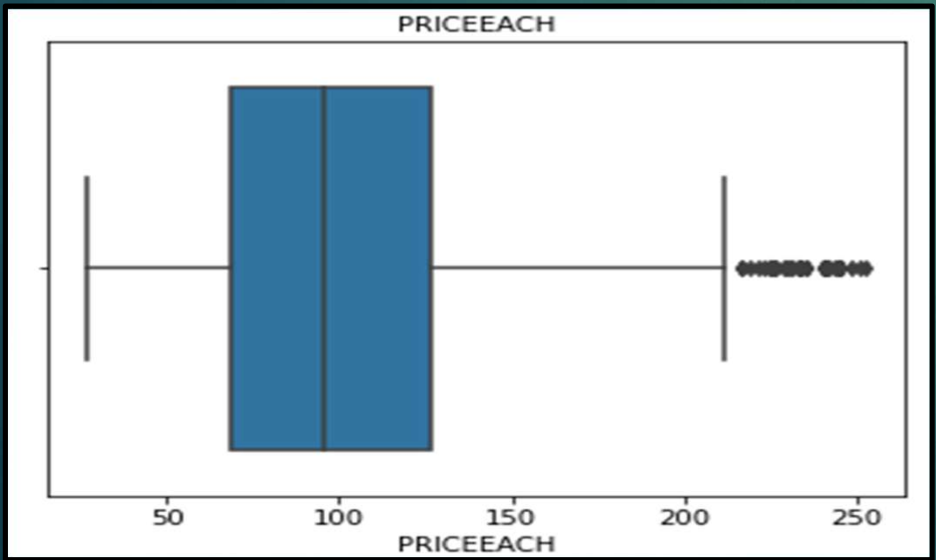
- We can see that when quantity was 33,373 we did a sale of 3,842,869 wherein the price was 109321. Here the profit is too high.
- The products seem to have profits when compared to the price and orders placed.
- Trains and ships needs to be analyzed as to why the orders are less.

- ***Checking outliers***



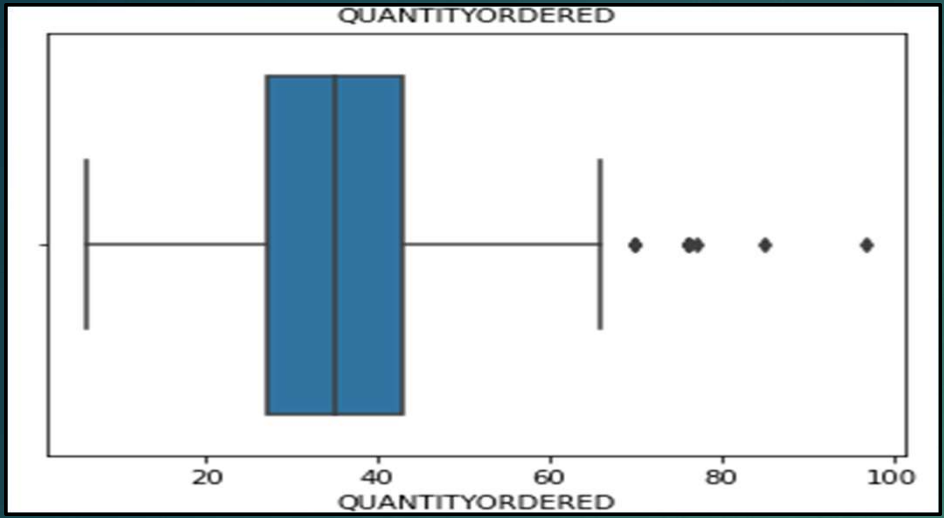
Inference:

- There are outliers in the data for sales.
- The data seems to have a mean sales of 2000-4000 and then spread across maximum to 14000

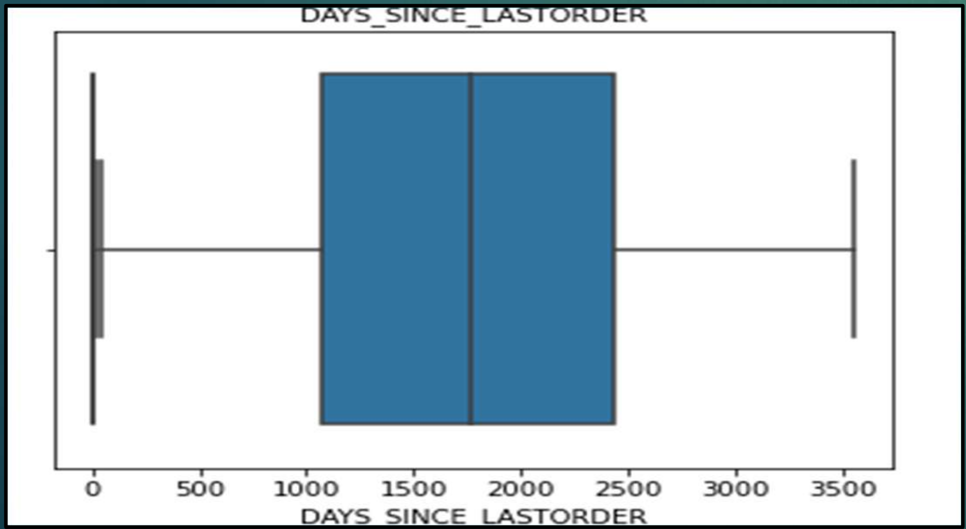


Inference:

- There are outliers in the data for price of product.
- The data seems to have a mean price at 100 and then spread across maximum to 250.

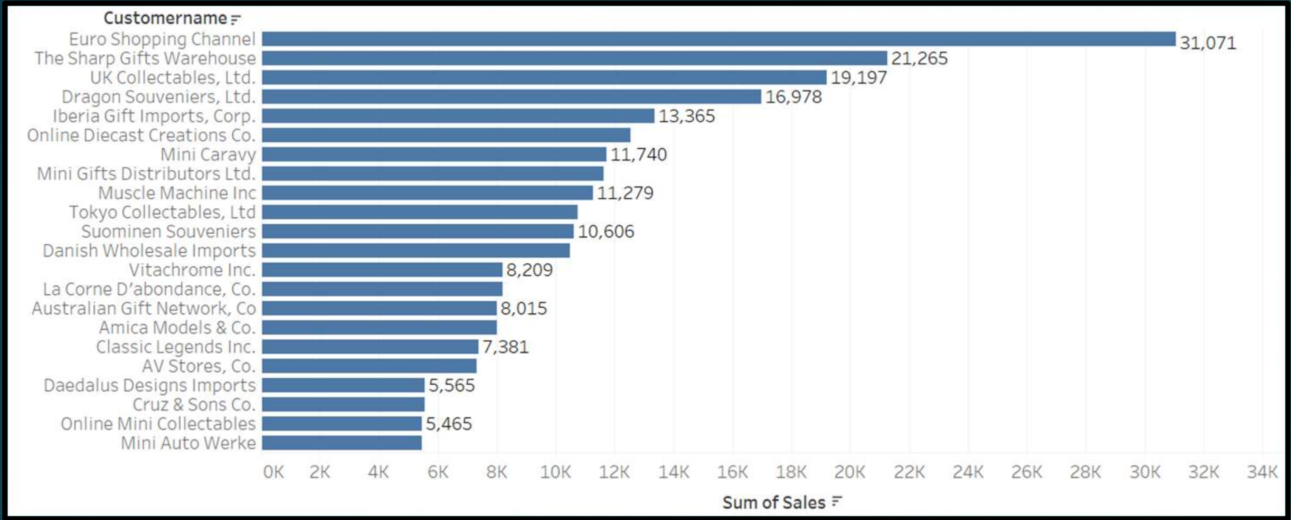


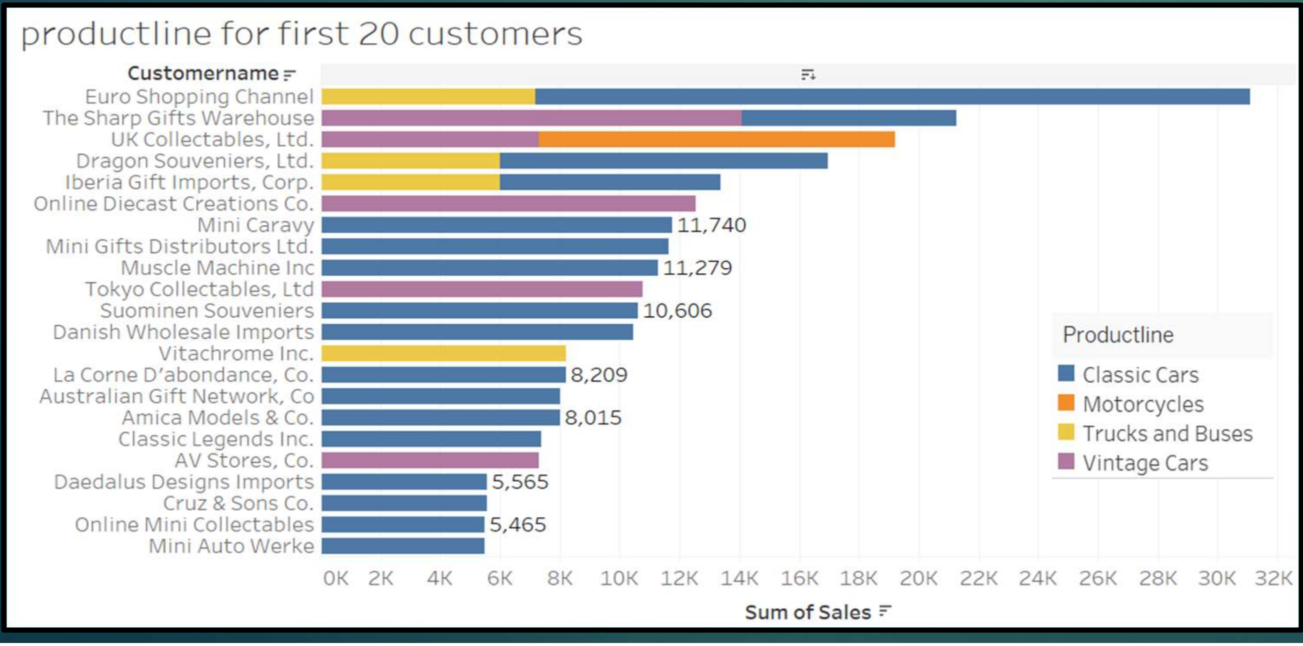
Inference:
- There are outliers in the data for quantity ordered of product.
- The data seems to have a minimum quantity at 20-40 and the maximum at 100.



Inference:
- The data for days since last order seems to be perfect as it is normally distributed and seems sales is not impacted for gap between orders.
- No outliers are present in the data.

• Top 20 Customers

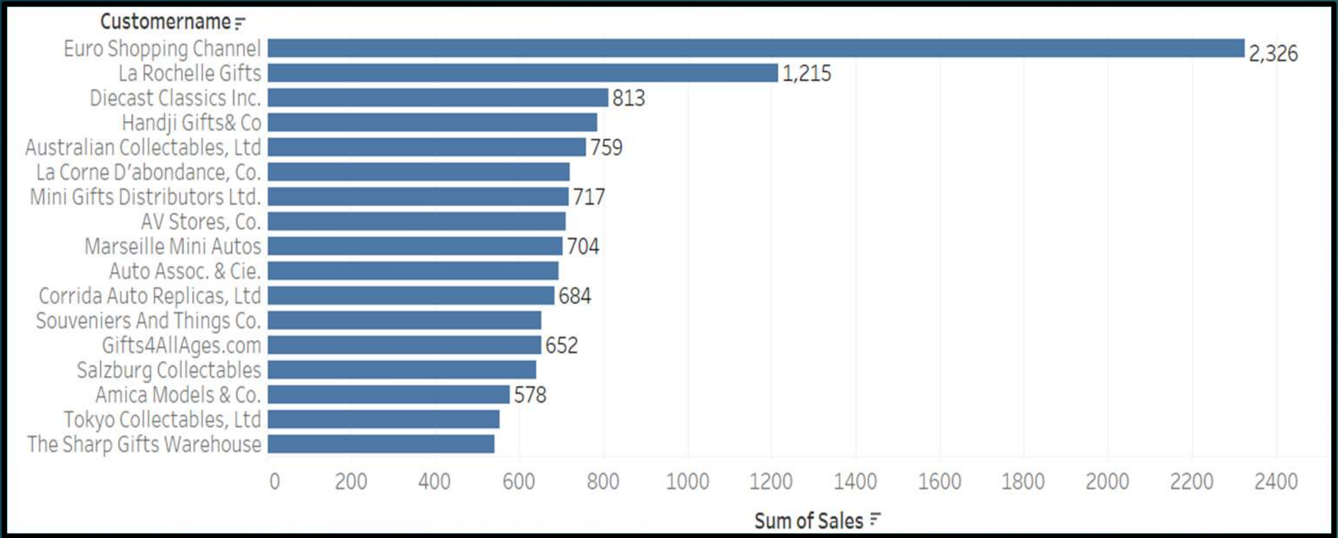


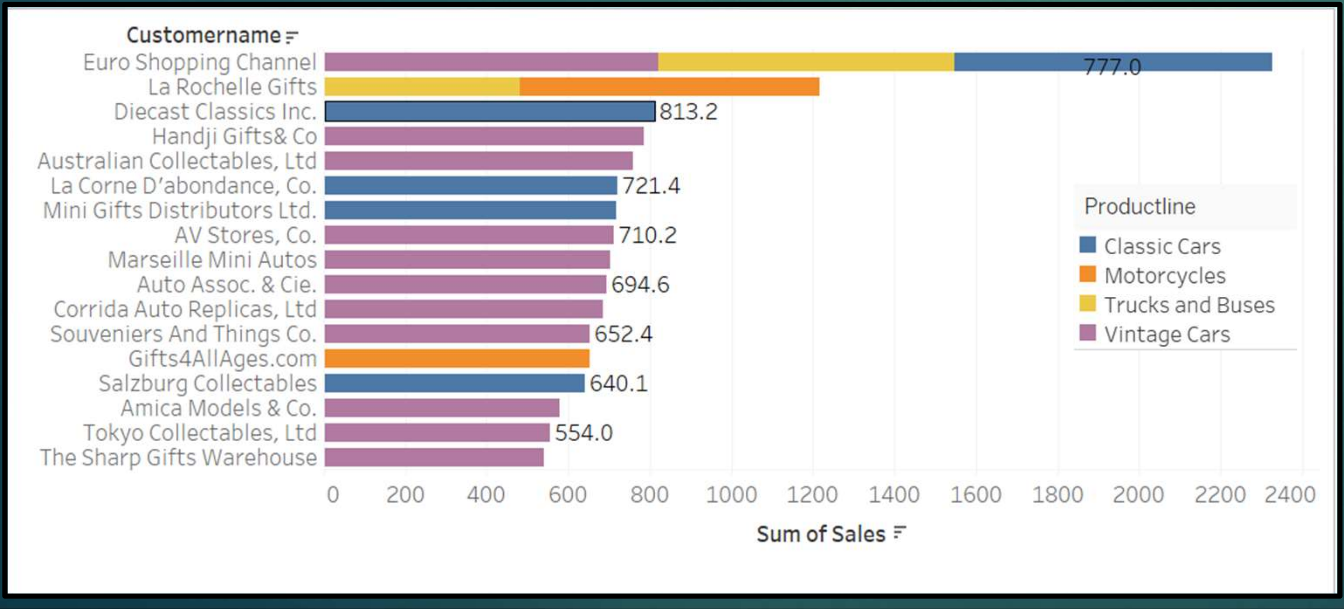


Inference:

- The highest customer is Euro Shopping Channel with a sale of 31,071.
- The lowest sale in top 20 customers is nearly 5,000.
- We see from both the outputs that our top customers is more into classic cars only.
- UK collectables is mostly into Motorcycles and vintage cars only.
- Summing up we have top 20 only into classic cars and vintage cars. Motorcycle and trucks buses being only for some.

- ***Last 20 Customers***



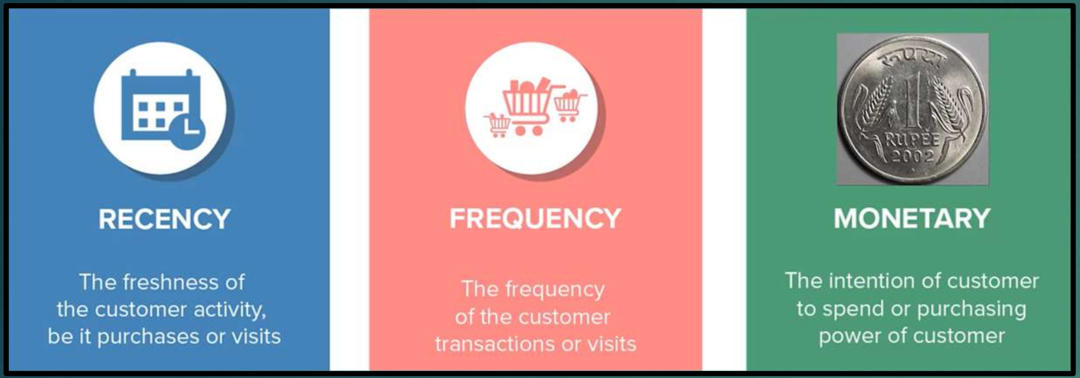


Inference:

- The highest customer is Euro Shopping Channel with a sale of 777.
- The lowest sale in last 20 customers is nearly 500.
- We see from both the outputs that our last customers is more into vintage cars only.
- Diecast Classics is mostly into classic cars only.
- The last customers are mostly into Vintage cars and we can say here is a mixture of all the productline.

Customer Segmentation using RFM analysis

RFM stands for recency, frequency and monetary. This method is used mostly to segment customers or products.



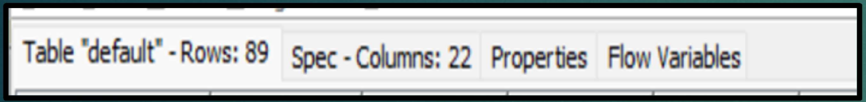
This method is used mostly to segment customers or products. How recently a customer visits, how frequently they came and purchased and what is the monetary value they added. A way to segment customers to know whom we can churn and whom we can concentrate on to get more results.

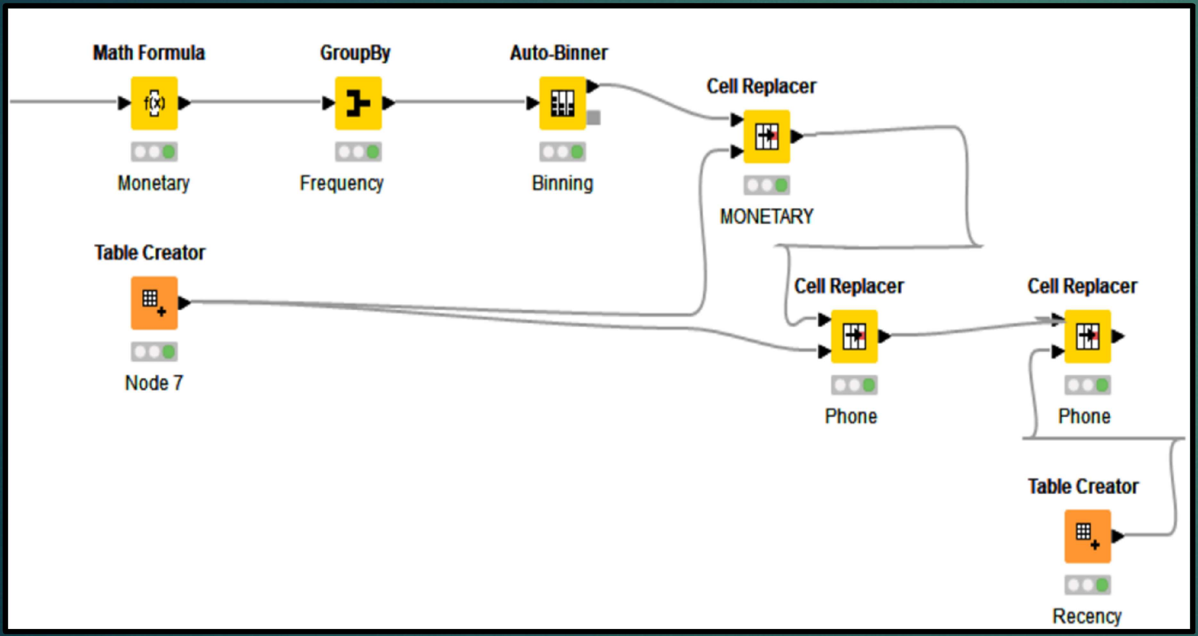
We have used KNIME to do the RFM analysis.

Formulas:

- **MONETARY = Price Each *Quantity ordered**
- **FREQUENCY= Count of Orders**
- **RECENCY = column "**DAYS_SINCE_LASTORDER**" has been taken**
- **To make data in recency reliable we have used phone numbers as phone number for all the customers will be same with the same customer name.**

- When we started in KNIME which RFM we had 2747 rows and 21 columns.
- Applying the Math formula to know the monetary aspect we got the customer “The sharp gifts warehouse” having the highest monetary contribution with a demand for Vintage cars. Least being the customer “La Rochelle Gifts” with a demand for Trucks and Buses.
- We have taken DAYS_SINCE_LASTORDER as our recency data.
- Since the data was huge and to focus on any customer was not clear we grouped the data by Customer Name. All the fields were assigned as per their nature as count or summation or means in order to avoid any default value reading. The data now stands as:





- Post grouping we have created bins on the basis of High, Medium and Low.
- Recency works the other way than monetary and frequency. In recency low means a good score because customer visited recently and in the other two high is a good score.
- Since Customer is a categorical field , we used a numerical field Phone number to detect bins for Frequency.
- OUTPUT-

KNIME_Debsmita_Sales Data.knwf

Customer Segmentation

The assumptions made are:

- We have used KNIME .
- We will consider price of each item as that will determine how many orders we will receive.
- Sales will depend on City as well as it will decide how frequently can the customer visit or order from us.
- Product line has the most impact which is what exactly the customer needs.
- Deal size will impact sales as the more the deal size the higher the sales.
- MSRP will also impact as the price will determine the deal size and the quantity ordered.
- Days since last order to know how frequently a customer is ordering.
- It gives us the means. Each cluster shows different combinations and the respective means.

Table "default" - Rows: 10 | Spec - Columns: 10 | Properties | Flow Variables

Row ID	D QUANT...	D PRICEE...	D SALES	D DAYS_...	D PRODU...	D MSRP	D PHONE	D POSTA...	D DEALSIZE	D Monetary
cluster_0	36.05	98.415	53,850.072	15.111	15.111	1,493.667	15.111	15.111	15.111	53,850.072
cluster_1	34.018	95.408	29,193.773	9.125	9.125	845.875	9.125	9.125	9.125	29,193.773
cluster_2	34.337	98.74	67,848.35	20	20	1,959.857	20	20	20	67,848.35
cluster_3	35.689	100.04	783,576.085	219.5	219.5	21,789	219.5	219.5	219.5	783,576.085
cluster_4	35.501	106.473	187,961.732	49.75	49.75	5,217.75	49.75	49.75	49.75	187,961.732
cluster_5	35.255	99.788	146,670.287	42.231	42.231	4,180.077	42.231	42.231	42.231	146,670.287
cluster_6	34.776	105.788	116,476.827	31.75	31.75	3,323.417	31.75	31.75	31.75	116,476.827
cluster_7	35.42	98.498	84,951.186	24.7	24.7	2,425.6	24.7	24.7	24.7	84,951.186
cluster_8	34.642	102.133	77,068.855	22.154	22.154	2,237	22.154	22.154	22.154	77,068.855
cluster_9	35.318	104.946	99,294.609	26.818	26.818	2,787.545	26.818	26.818	26.818	99,294.609

- Labelled input has original dataset intact only the rows have been shrinked.

Table "default" - Rows: 89 | Spec - Columns: 28 | Properties | Flow Variables

Row ID	S CUSTO...	I ORDER...	D QUANT...	D PRICEE...	I ORDER...	D SALES	I ORDER...	I DAYS_...	I STATUS	I PRODU...	I MSRP	I PRODU...	I PHONE	I ADDRE...	I CITY
Row0	AV Stores, Co.	51	34.863	91.085	51	157,807.81	51	51	51	51	4735	51	51	51	51
Row1	Alpha Cognac	20	34.35	101.16	20	70,488.44	20	20	20	20	1943	20	20	20	20
Row2	Amica Model	26	32.423	110.853	26	94,117.26	26	26	26	26	2799	26	26	26	26

- **The best cluster is cluster zero as we have lesser difference in visiting since they last ordered. The average sales earned is the highest from them. Also the Average of MSRP is high so loss incurred is lesser comparatively to other clusters.**

Row Labels	Average of QUANTITYORDERED	Count of CUSTOMERNAME	Average of DAYS_SINCE_LASTORDER	Average of SALES	Average of MSRP
cluster_0	48.58	85	992.88	₹ 9,059.83	₹ 158.18
cluster_3	43.51	276	1229.14	₹ 6,492.83	₹ 141.43
cluster_2	37.89	354	1090.77	₹ 4,812.61	₹ 130.84
cluster_5	40.37	344	2476.42	₹ 4,144.26	₹ 97.54
cluster_1	32.54	526	1109.71	₹ 3,141.15	₹ 110.65
cluster_4	33.11	601	2650.12	₹ 2,528.41	₹ 79.17
cluster_6	28.48	561	1762.26	₹ 1,598.95	₹ 68.56
Grand Total	35.10302148	2747	1757.085912	3553.047583	100.6916636

- **The top 10 cluster_0 customers are:**

Top 10 Customers in Cluster_0	**Sales**
Mini Gifts Distributors Ltd.	71136.69
Euro Shopping Channel	57865.36
The Sharp Gifts Warehouse	31904.54
Mini Caravy	29076.91
Dragon Souveniers, Ltd.	28342.71
Danish Wholesale Imports	27533.95
L'ordine Souveniers	25411.55
Gift Depot Inc.	25251.49
Salzburg Collectables	25096.48
Amica Models & Co.	24679.38

- We have binned the data as well and have divided it into Bin1, Bin 2 and Bin 3.
- Post grouping we have created bins on the basis of High, Medium and Low.
- Recency works the other way than monetary and frequency. In recency low means a good score because customer visited recently and in the other two high is a good score.

Row ID	...	PHONE	ADDRE...	CITY	POSTA...	COUNTRY	CONTA...	CONTA...	DEALSIZE	Monetary	DAYS_...	PHONE ...	Moneta...	Moneta...	Custom...	Recenc...
Row0		51	51	51	51	51	51	51	51	157,807.81	Bin 3	Bin 3	Bin 3	H	H	L
Row1		20	20	20	20	20	20	20	20	70,488.44	Bin 1	Bin 1	Bin 1	L	L	H
Row2		26	26	26	26	26	26	26	26	94,117.26	Bin 2	Bin 2	Bin 2	M	M	M
Row3		46	46	46	46	46	46	46	46	153,996.13	Bin 3	Bin 3	Bin 3	H	H	L
Row4		7	7	7	7	7	7	7	7	24,179.96	Bin 1	Bin 1	Bin 1	L	L	H
Row5		23	23	23	23	23	23	23	23	64,591.46	Bin 2	Bin 2	Bin 1	L	M	M
Row6		55	55	55	55	55	55	55	55	200,995.41	Bin 3	Bin 3	Bin 3	H	H	L
Row7		15	15	15	15	15	15	15	15	59,469.12	Bin 1	Bin 1	Bin 1	L	L	H
Row8		18	18	18	18	18	18	18	18	64,834.32	Bin 1	Bin 1	Bin 1	L	L	H
Row9		27	27	27	27	27	27	27	27	93,170.66	Bin 2	Bin 2	Bin 2	M	M	M
Row10		8	8	8	8	8	8	8	8	26,479.26	Bin 1	Bin 1	Bin 1	L	L	H
Row11		32	32	32	32	32	32	32	32	116,599.19	Bin 2	Bin 2	Bin 2	M	M	M
Row12		14	14	14	14	14	14	14	14	34,993.92	Bin 1	Bin 1	Bin 1	L	L	H
Row13		22	22	22	22	22	22	22	22	85,171.59	Bin 2	Bin 2	Bin 2	M	M	M
Row14		3	3	3	3	3	3	3	3	9,129.35	Bin 1	Bin 1	Bin 1	L	L	H
Row15		13	13	13	13	13	13	13	13	49,642.05	Bin 1	Bin 1	Bin 1	L	L	H
Row16		11	11	11	11	11	11	11	11	36,163.62	Bin 1	Bin 1	Bin 1	L	L	H

Inferences from RFM Analysis and Identified segments. Good, Loyal, Lost and churn customers.

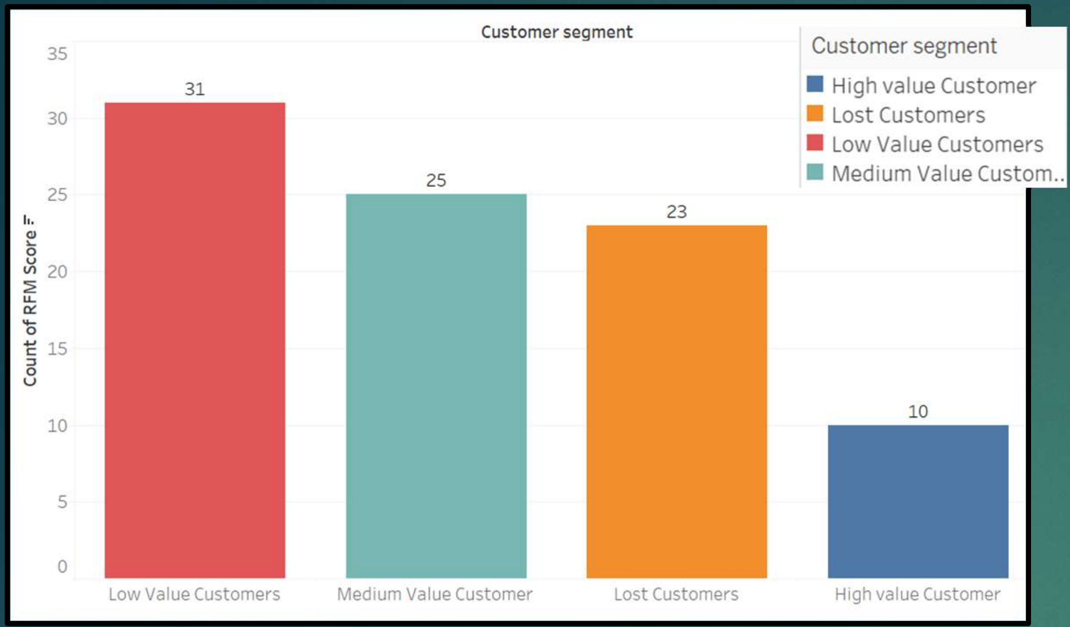
The RFM analysis post dividing the data into Recency, Frequency and Monetary has been given ranks under the segments in JUPYTER(Python):

- **High Value,**
- **Lost Customers,**
- **Low Value Customers,**
- **Medium Value Customers**

	CUSTOMERNAME	Recency	Frequency	Monetary	R_rank_norm	F_rank_norm	M_rank_norm	RFM_Score	Customer_segment
0	AV Stores, Co.	2969	51	5676559.99	55.617978	95.505618	95.505618	4.48	High value Customer
1	Alpha Cognac	3272	20	2603716.92	20.224719	28.089888	28.089888	1.35	Lost Customers
2	Amica Models & Co.	3003	26	3345240.88	50.561798	53.932584	53.932584	2.67	Low Value Customers
3	Anna's Decorations, Ltd	2939	46	5113397.35	60.674157	90.449438	90.449438	4.30	High value Customer
4	Atelier graphique	2183	7	943849.26	100.000000	2.247191	2.247191	0.85	Lost Customers
5	Australian Collectables, Ltd	2743	23	2270716.64	86.516854	38.764045	38.764045	2.30	Low Value Customers
6	Australian Collectors, Co.	3356	55	7583338.75	12.359551	97.752809	97.752809	4.25	High value Customer
7	Australian Gift Network, Co	2799	15	2261737.10	82.022472	14.606742	14.606742	1.24	Lost Customers
8	Auto Assoc. & Cie.	2790	18	2490636.72	84.269663	20.786517	20.786517	1.52	Lost Customers
9	Auto Canal Petit	2870	27	3641885.74	75.280899	60.674157	60.674157	3.14	Medium Value Customer

Inference Rating Customer based upon the RFM score

- **rfm score >4.5 : Top Customer**
- **4.5 > rfm score > 4 : High Value Customer**
- **4>rfm score >3 : Medium value customer**
- **3>rfm score>1.6 : Low-value customer**
- **rfm score<1.6 :Lost Customer**



INFERENCE:

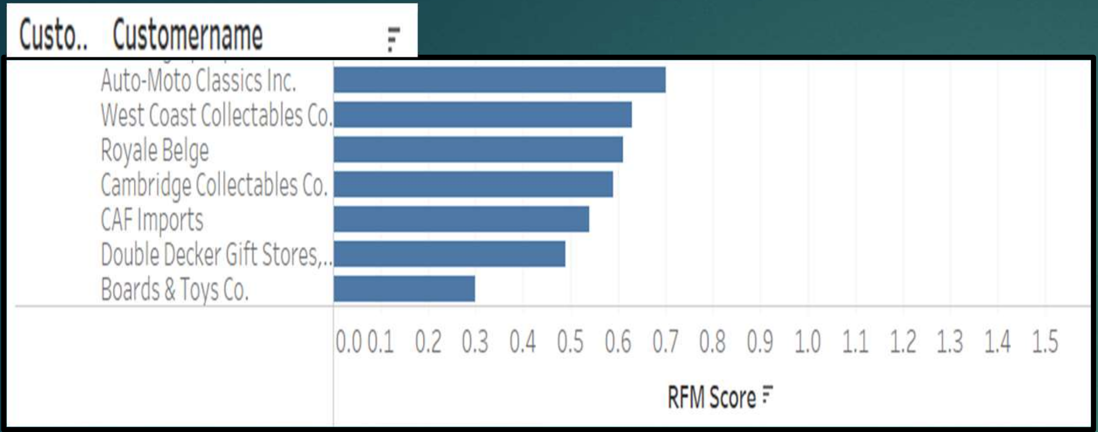
- The RFM score has maximum count of Customers in Low value customers and the least that is a count of 10 in high value customers.
- The Medium Value customer is the one we need to focus more on and target Low value customers for retention or conversions through marketing strategies.

Top 5 customers

INFERENCE:

- The RFM score has been sorted by maximum score wise which is giving us the top five customers as in the image.
- The highest score of La Rochelle Gifts being 4.5.

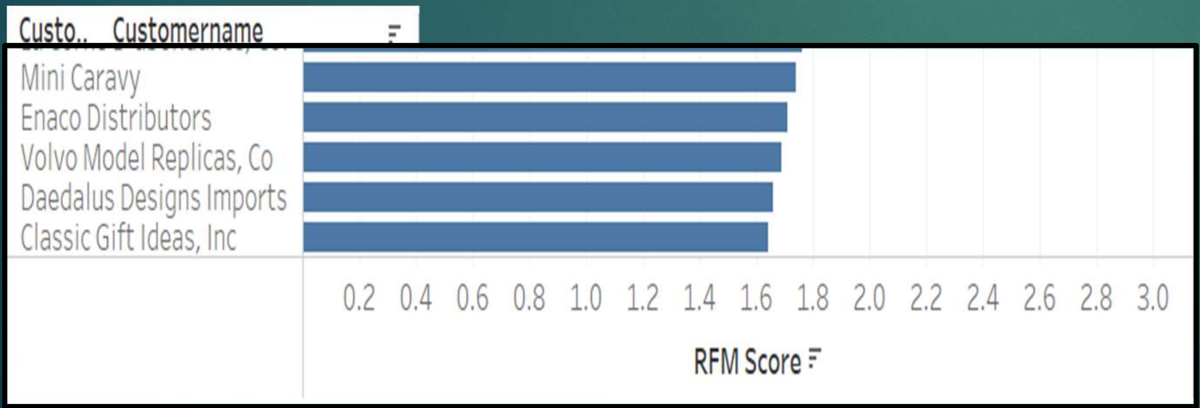
Lost Customers



INFERENCE:

- A RFM score below 1.6 means lost customer.
- The above customers have been lost and the lowest amongst them is Board & Toys Company.
- These customers will be removed and will not be looked into to retain.

Verge of Churning



INFERENCE:

- The last 5 low value customers who are with the least RFM score.
- They are at the verge of churning.
- The least score being of Classic Gift Ideas Inc which is 1.6 exactly.
- These customers are at the verge because RFM below 1.6 means lost.

Loyal Customers

INFERENCE:

- **Although the scores are not above 4.5 but these customers are stagnant and contribute well throughout.**
- **They maintain a decent RFM score hence these should be the main focus for us.**

Recommendation:

- Maximum lost customers had ordered Vintage cars which were in shipped status. We need to focus on why its taking so long to process any order. Make our delivery and order management department strong.
- Trains have few orders of all. We should analyse on where the company is lacking is it the quality , meeting the demand or failing to provide as good as competitors.
- Our loyal customer is the medium value customers who have again highest volume in shipped which should be analysed since these customers contribute more to our sales.
- Classic Cars are getting cancelled the most. The company needs to study is it better price and product some other competitor is providing or is it company driven default.
- A customer survey form on every product line should be made. One section should consist of the service satisfaction and the other should have customer requirement analysis form. The customer requirement analysis form can take basic details of the customer like industry, need they have from our product line and on why they cancel orders.
- In 2019 we had the highest sales, a study of that years data can be done further in detail to know what went well.

Tools Used:

- Python
- Tableau
- KNIME

THE END