AUTOMOBIL COMPANY

MRA PROJECT MILESTONE - 1

Pooja Kabadi

PGP-DSBA Online Batch- A4

20-02-2022

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. Your job is to use your magical data science skills to provide them with suitable insights about their data and their customers.

Auto Sales Data: Sales_Data.xlsx

- **Agenda & Executive Summary of the data**
  - Contents of the presentation
  - Problem statement
  - About Data (Info, Shape, Summary Stats, your assumptions about data)
- **Exploratory Analysis and Inferences**
  - Univariate, Bivariate, and multivariate analysis using data visualization
  - Weekly, Monthly, Quarterly, Yearly Trends in Sales
  - Sales Across different Categories of different features in the given data
  - Summarize the inferences from the above analysis
- **Customer Segmentation using RFM analysis (make 4 segments)**
  - What is RFM?
  - What all parameters used and assumptions made
  - Output table head
  - When KNIME used, Workflow image to be put
- **Inferences from RFM Analysis and identified segments**
  - Who are your best customers? (give at least 5)
  - Which customers are on the verge of churning? (give at least 5)
  - Who are your lost customers? (give at least 5)
  - Who are your loyal customers? (give at least 5)

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

Executive Summary:

- This project attempts to discover the underlying buying patterns of an automobile part manufacturer's customers analyzing transaction data from the last 3 years, and then recommend customized marketing strategies for various client segments.
- Customer analysis and segmentation based on RFM
- Possibilities for improvement
- To drive business solutions, identify patterns and other analytical inferences.

About Data:

- The data provided is from a Automobile parts manufacturing company of transactions for 3 years.
- The data has 20 attributes and 2747 records.
- This data reflects the purchasing behavior of customers in several categories.
- The company is into automobile parts and has a variety of product lines, including classic cars, motorcycles, planes, trains, ships, buses, trucks, and vintage cars.
- The company is dealing with 109 unique products across 7 different product line.
- Currently, the company has 89 customers from 19 different countries.
- The data is clean and has neither null values nor duplicate records

Data Information:

```
<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
```

Data information:

- The data is about an automobile parts manufacturing company of transactions for 3 years.
- The dataset has 20 variables and 2747 records.
- The data has 1 datetime64 , 2 float64, 5 int64, and 12 Object data types.
- There are no null values in any column of the dataset.
- There are no duplicates in the dataset.

Data Description:

	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	101.098951	42.042548	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

Summary of Numeric variables:

- The mean sales of the automobile parts manufacturing company ranges from 482 to 14082 with average sales of 3553 and standard deviation of 1838.
- From above summary we can infer that the price of each automobile parts is more than the MSRP [Manufacturer's Suggested Retail Price]
- The average price of MSRP is 100.69 where price of each is 101.098.
- The Quantity of ordered ranges from 9 to 97 with average of 35 orders.

Data Description:

	count	unique	top	freq
STATUS	2747	6	Shipped	2541
PRODUCTLINE	2747	7	Classic Cars	949
PRODUCTCODE	2747	109	S18_3232	51
CUSTOMERNAME	2747	89	Euro Shopping Channel	259
PHONE	2747	88	(91) 555 94 44	259
ADDRESSLINE1	2747	89	C/ Moralzarzal, 86	259
CITY	2747	71	Madrid	304
POSTALCODE	2747	73	28034	259
COUNTRY	2747	19	USA	928
CONTACTLASTNAME	2747	76	Freyre	259
CONTACTFIRSTNAME	2747	72	Diego	259
DEALSIZE	2747	3	Medium	1349

Summary of Categorical variables:

- **Status :** There are 6 unique categories to define the order status out of which most of the orders are 'shipped'.
- **Customer Name :** There are 89 different customers out of which 'Euro Shopping Channel' is the one with maximum orders.
- **City :** Orders are placed in 71 different cities with maximum orders from 'Madrid'
- **Country :** Orders are placed in 19 different countries with maximum orders from 'USA'
- **Deal Size :** There are 3 different categories of deal size. Most of the orders have 'Medium' deal size..

Exploratory Analysis

General Summary:

- Shape of data: (2747, 20)
- Continuous variables: 7
- Categorical variables: 12
- Date-Time variables: 1
- Null values: 0
- Duplicate records: 0

Tools & Libraries used:

PYTHON :

EDA for the dataset is performed using Python libraries:

- Pandas
- Numpy
- Seaborn
- Matplotlib

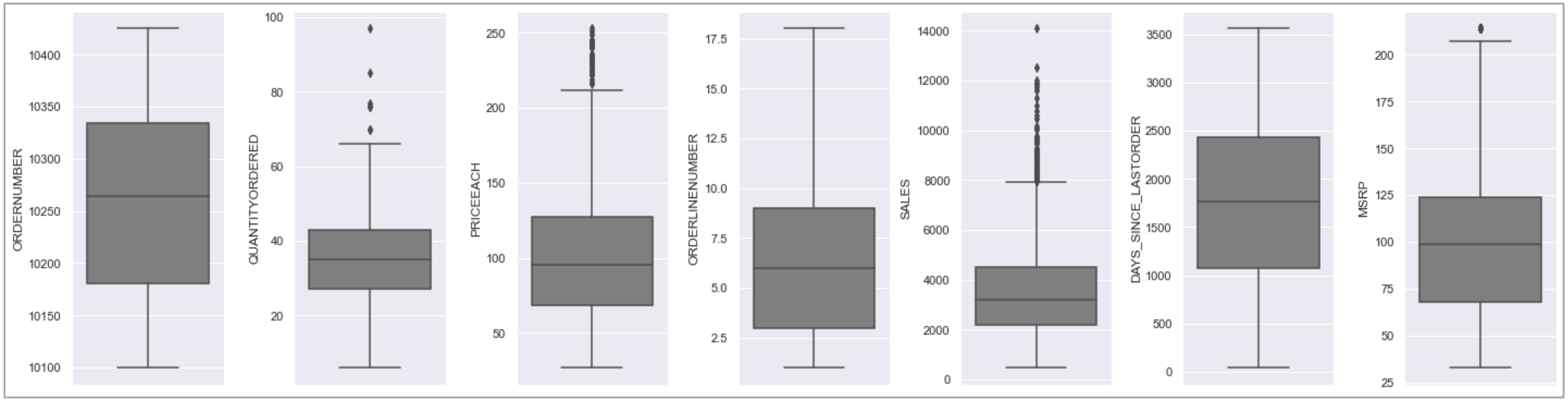
TABLEAU :

Data Visualization [Bivariate, Multivariate] is performed using tableau public.

KNIME :

RFM Analysis is performed using KNIME.

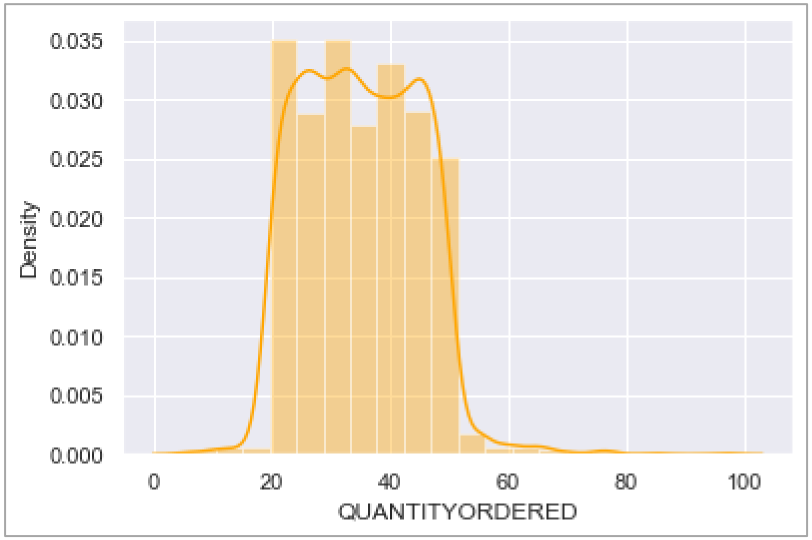
Outlier detection:

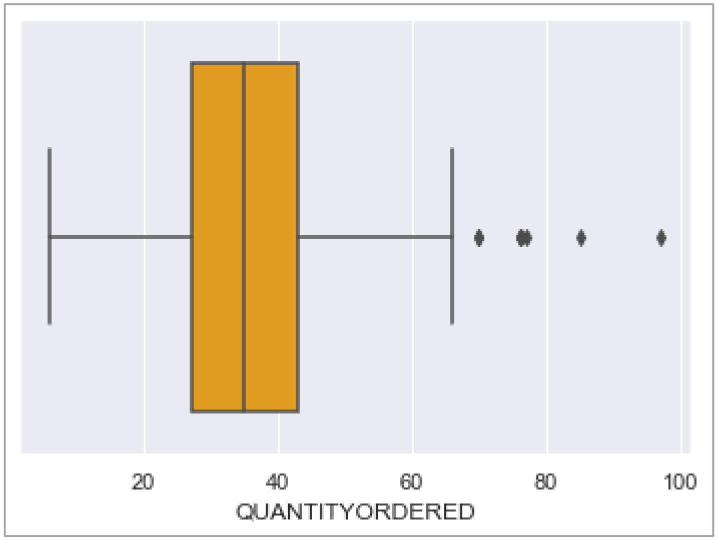


Inferences:

- The data shows varied range of entries across each variables
- We can see that, Sales and Price of each part of automobile have more number of outliers

Univariate Analysis : Quantity Ordered

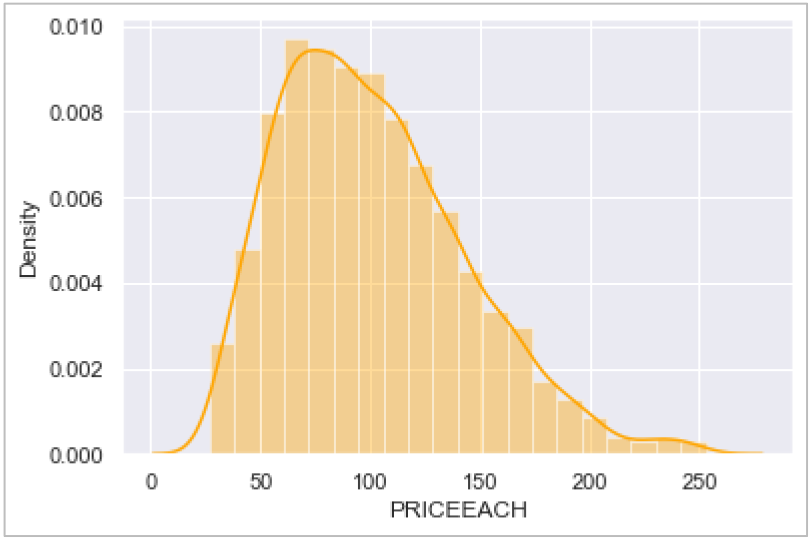


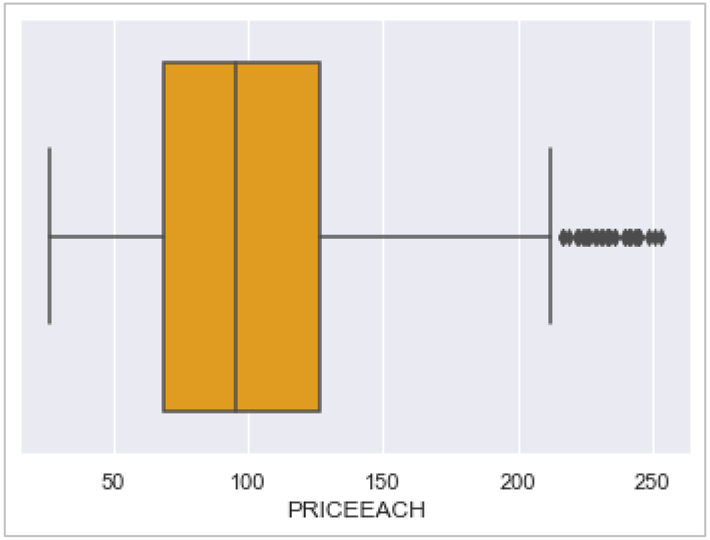


Inferences:

- The distribution of ‘Quantity ordered' is slightly right skewed with skewness value of 0.37.
- The distribution is almost normally distributed.
- The distplot shows the distribution of most of data from 20 to 65.
- The box plot of the ‘Quantity ordered' variable shows few outliers.

Univariate Analysis : Price of each item

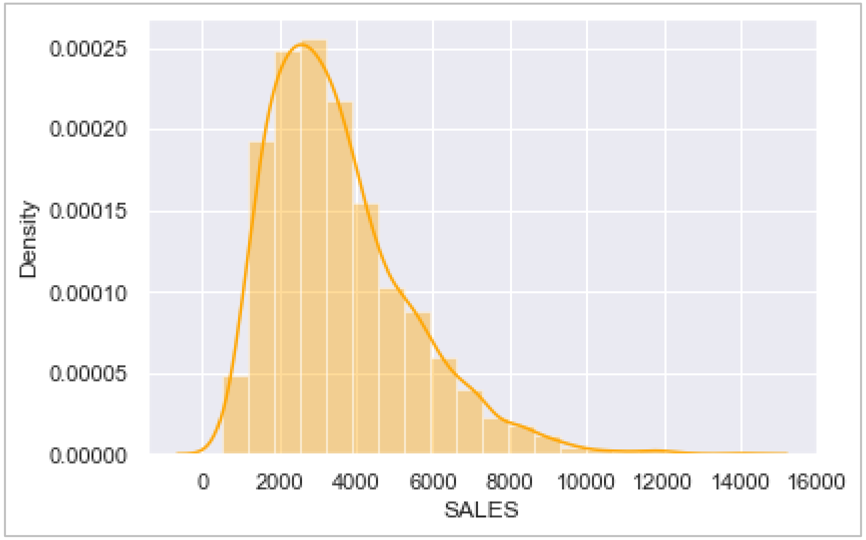


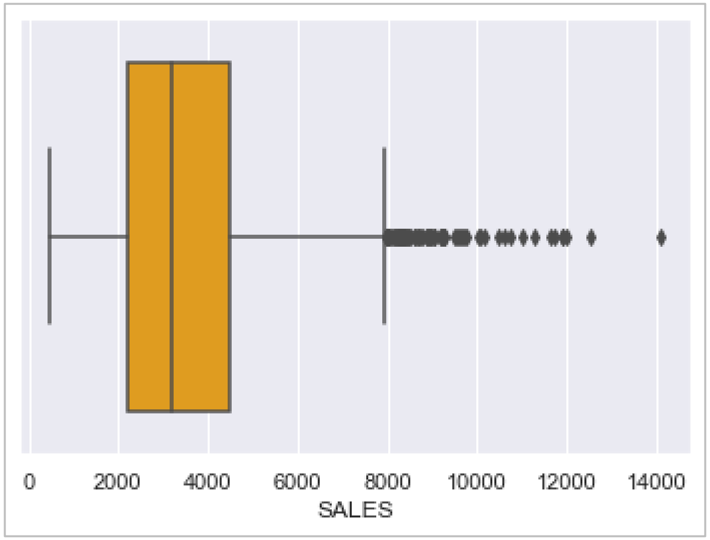


<u>Inferences:</u>

- The distribution of ‘Price each' is right skewed with skewness value of 0.7.
- The distplot shows the distribution of most of data from 20 to 220.
- The box plot of the ‘Price each' variable shows presence of outliers.

Univariate Analysis : Sales

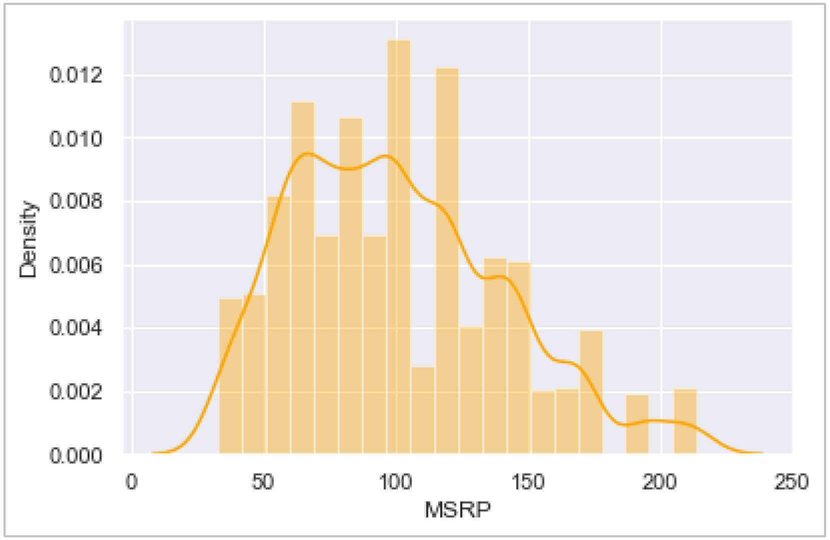


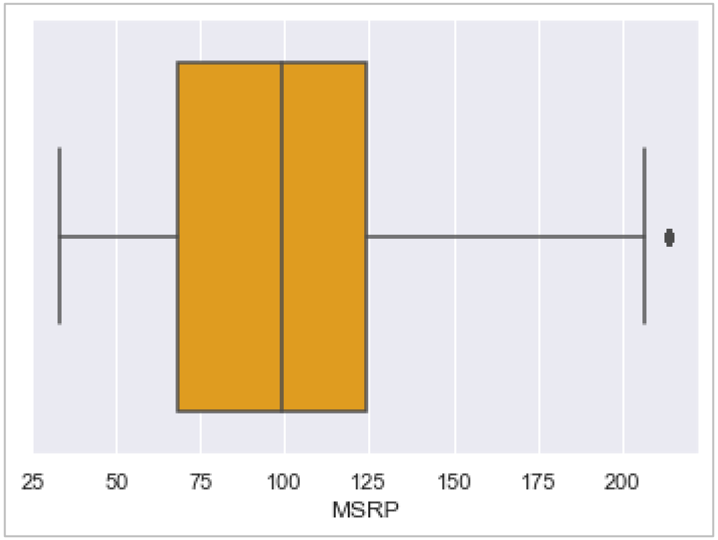


Inferences:

- The distribution of ‘Sales' is right skewed with skewness value of 1.16.
- The distplot shows most of the sales are in range of 500 to 8000.
- The box plot of the ‘Sales' variable shows presence of many outliers.

Univariate Analysis : MSRP

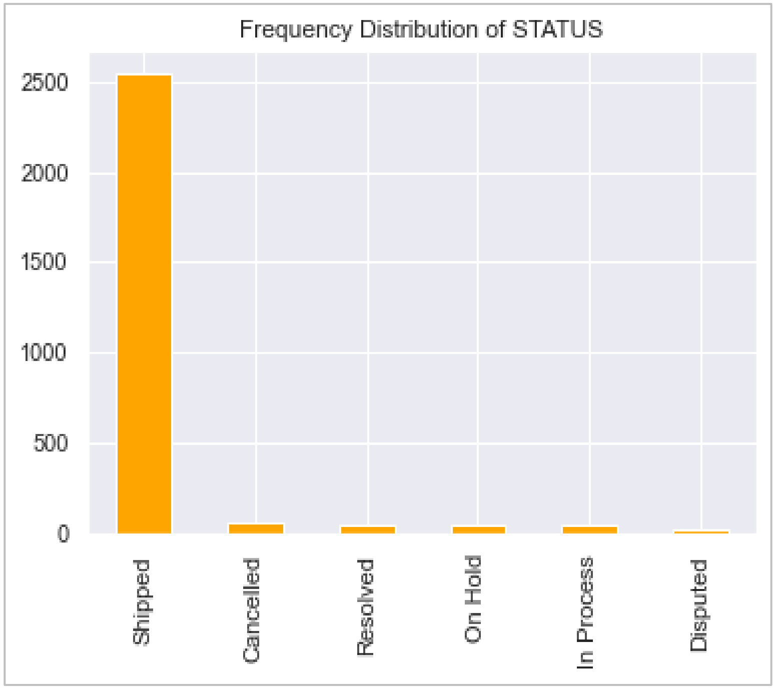


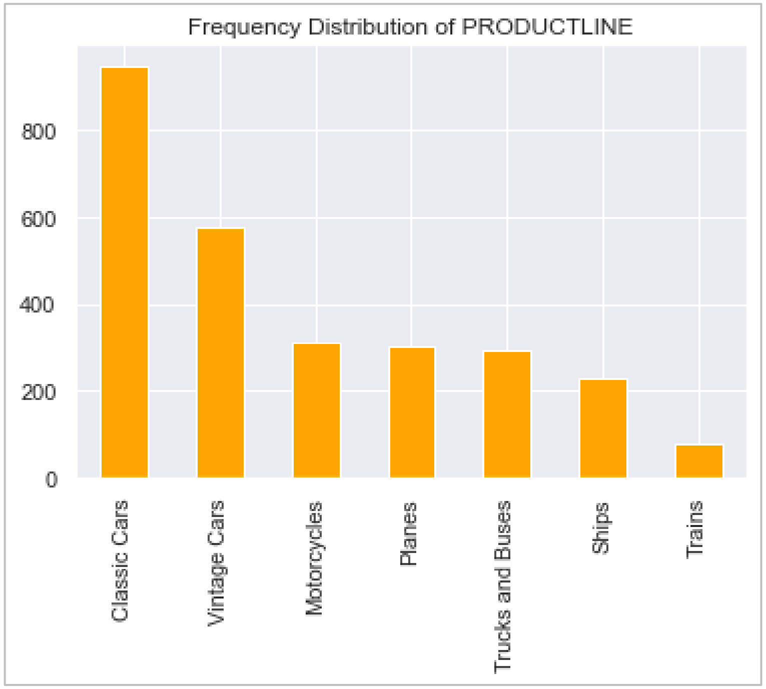


<u>Inferences:</u>

- The distribution of ‘MSRP’ is slightly right skewed with skewness value of 0.58.
- The distribution is almost normally distributed.
- The distplot shows the range of MSRP from 35 to 200 of each item .
- The box plot of the ‘MSRP' variable shows presence of one outliers.

Univariate Analysis : Status , Product line

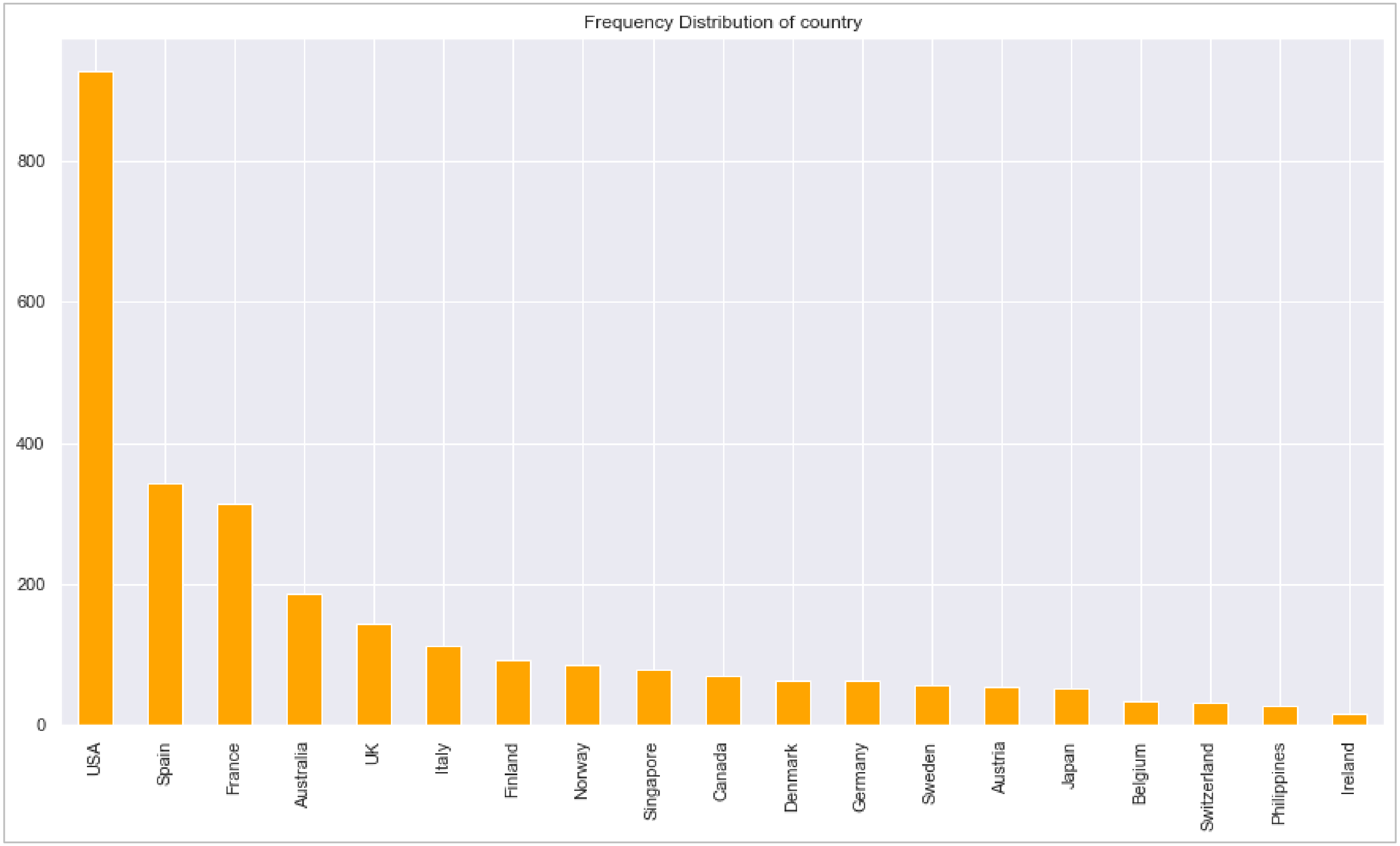




<u>Inferences:</u>

- Status is categories into 6 different types, most of the items are shipped.
- There are 7 different Product lines. Classic cars parts have the maximum orders followed vintage cars.
- The train parts are having the least orders.

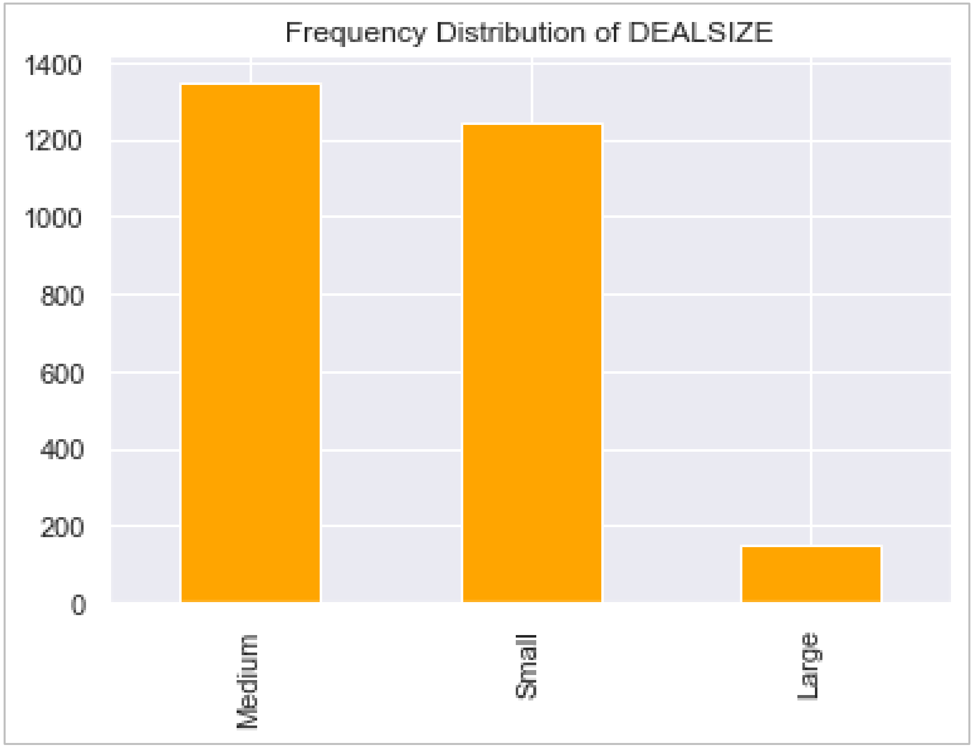
Univariate Analysis : Country



Inferences:

- USA has the highest number of orders followed by Spain and France.
- Ireland has minimum number of orders followed by Philippines and Switzerland.

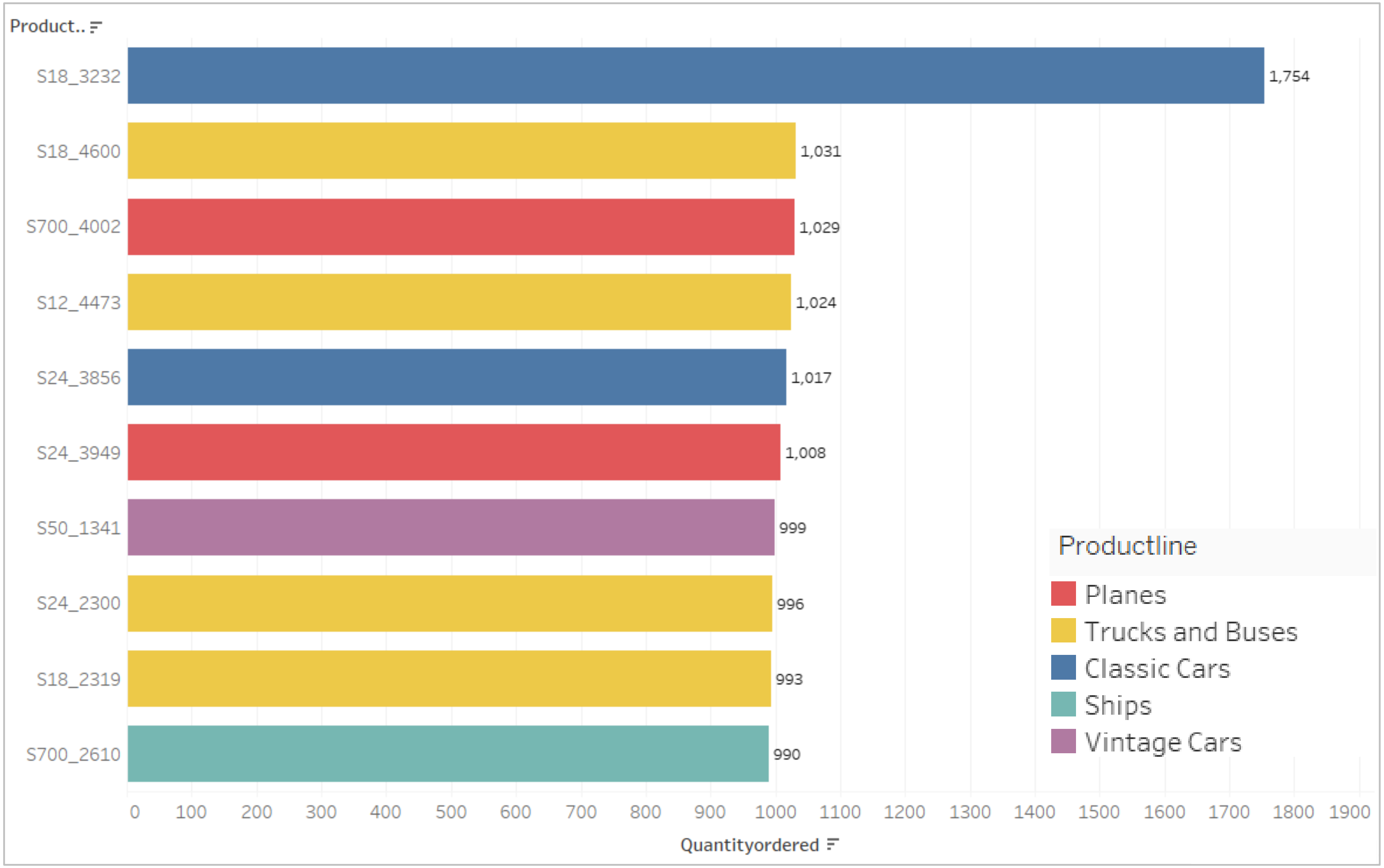
Univariate Analysis : Deal size



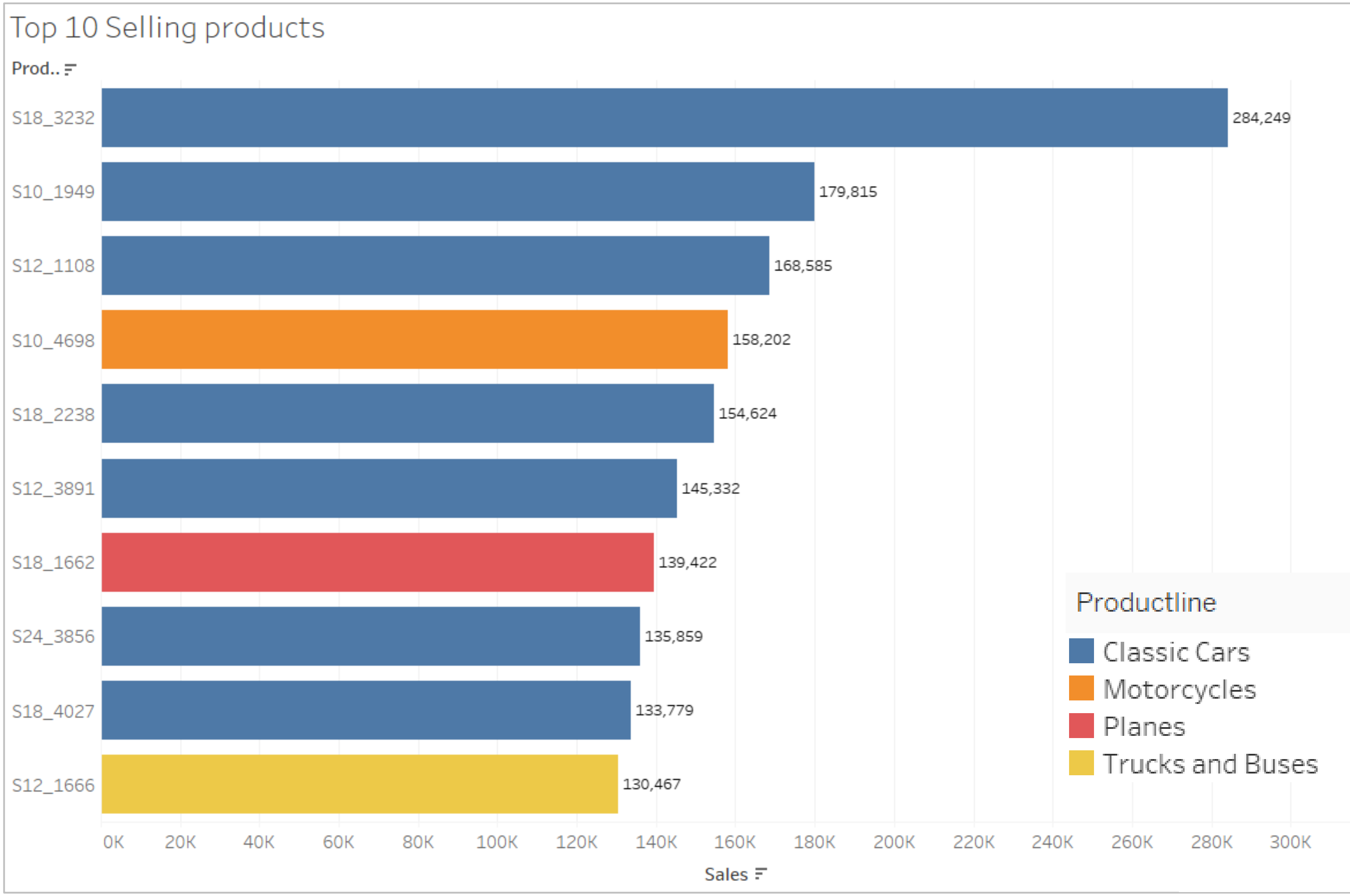
Inferences:

- The deal size of orders are mostly median size followed small and the least are the large.

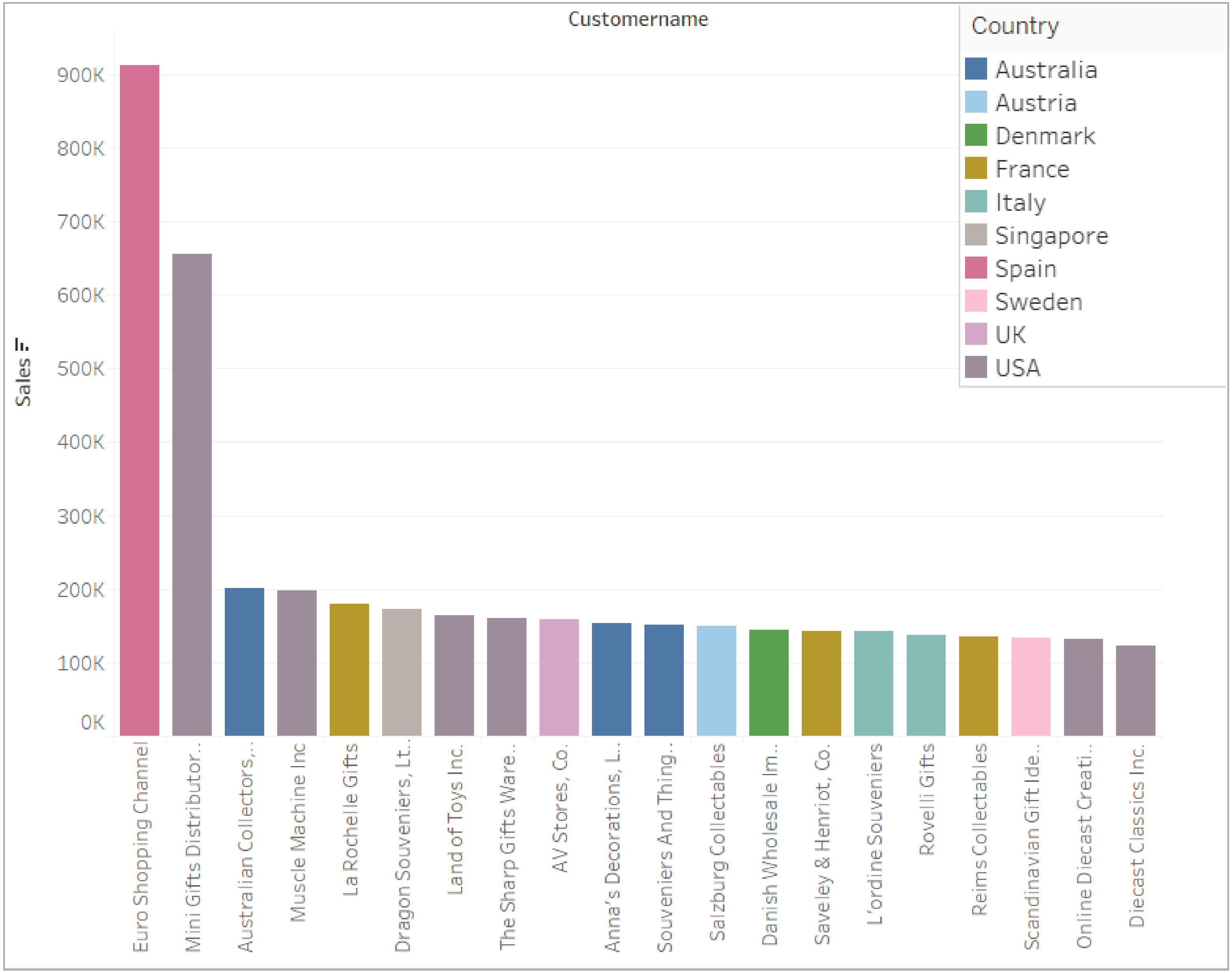
Top 10 purchased products



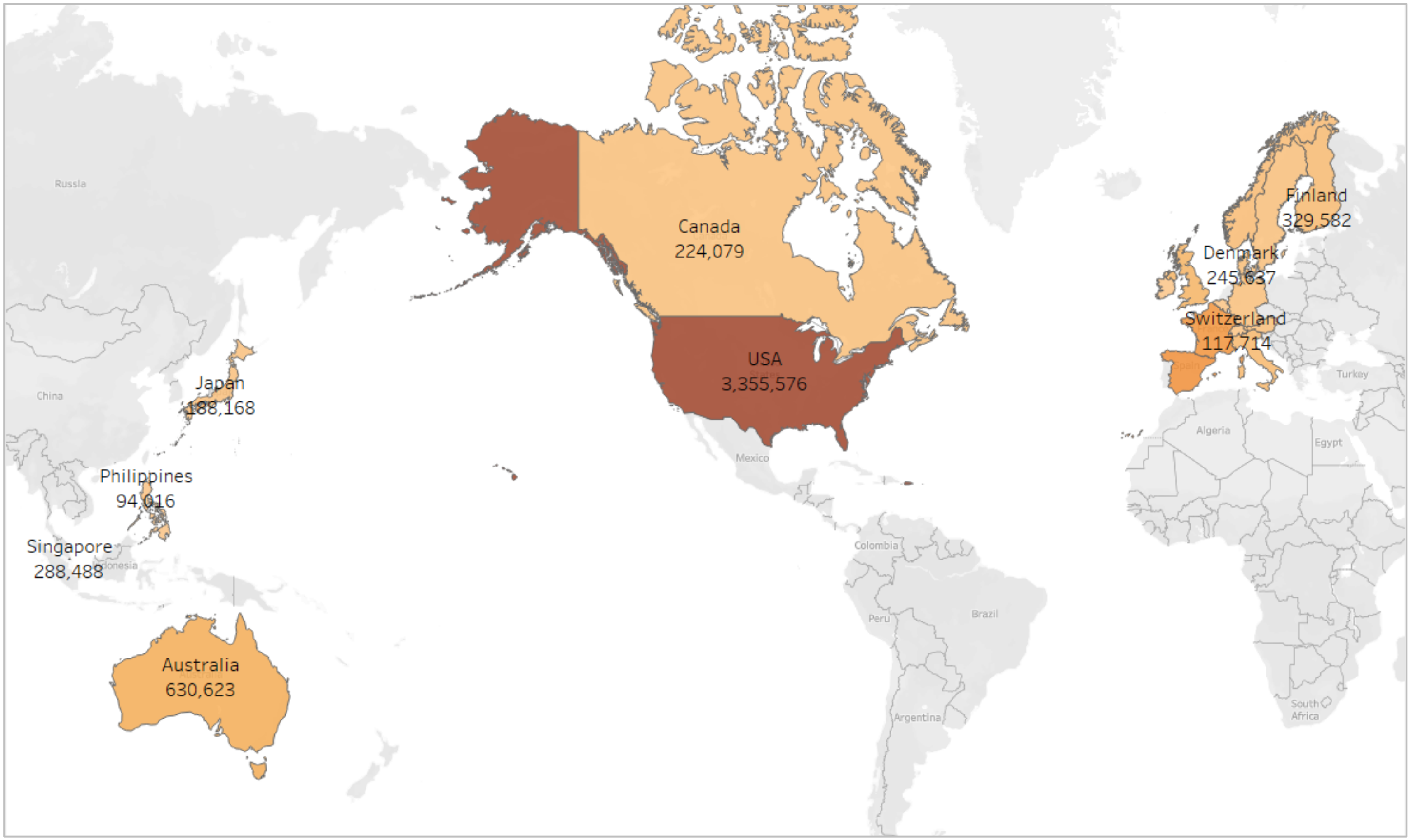
Top 10 selling products



Top 20 Customers across Globe



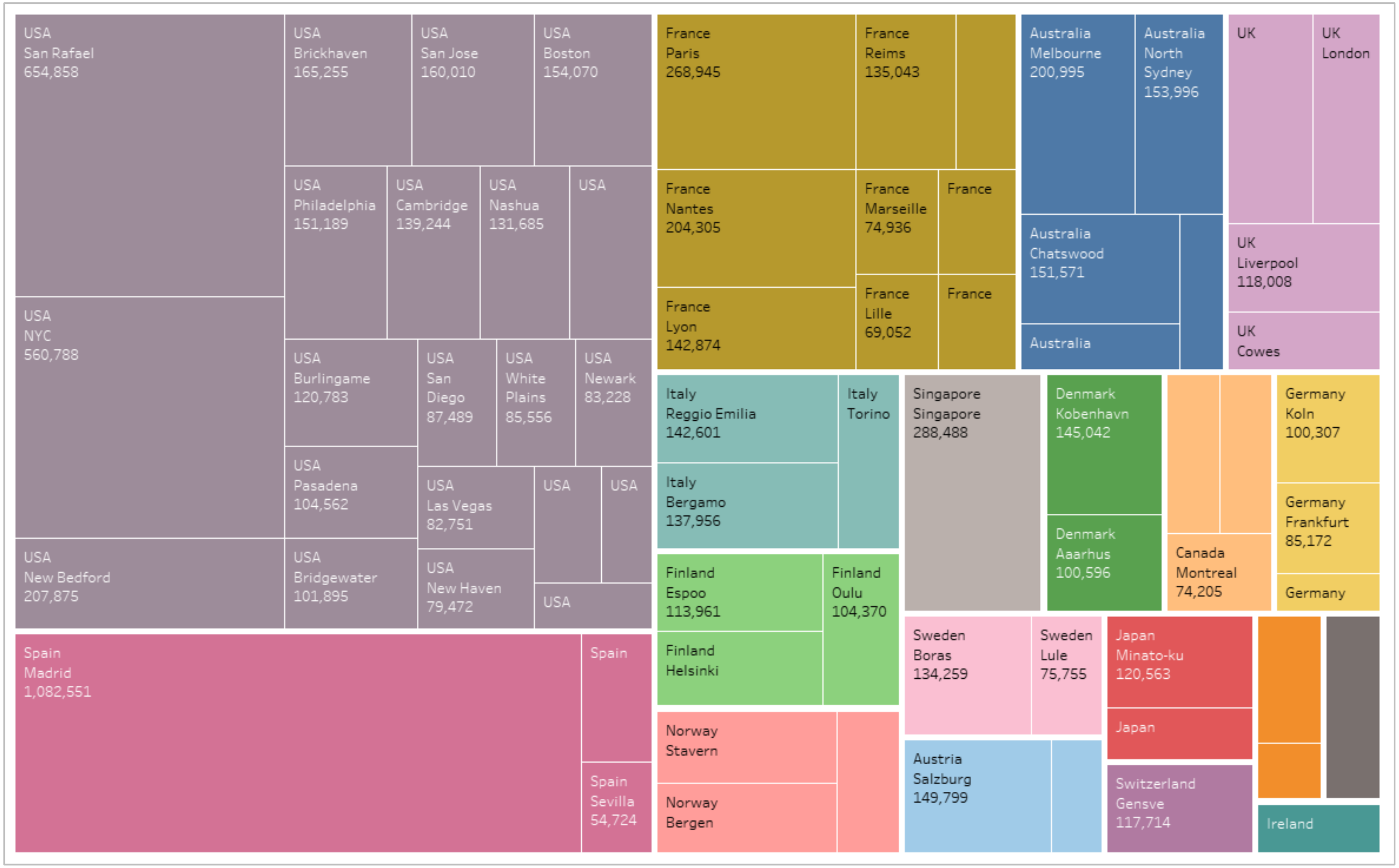
Bivariate Analysis : Sales across Countries



Inferences:

- This graph shows the sales across different countries. The company is having a viable market across 19 countries.
- USA is the primary market of the company contributing maximum turn-over Some European countries like Spain and Switzerland shows good amount of sales following USA.

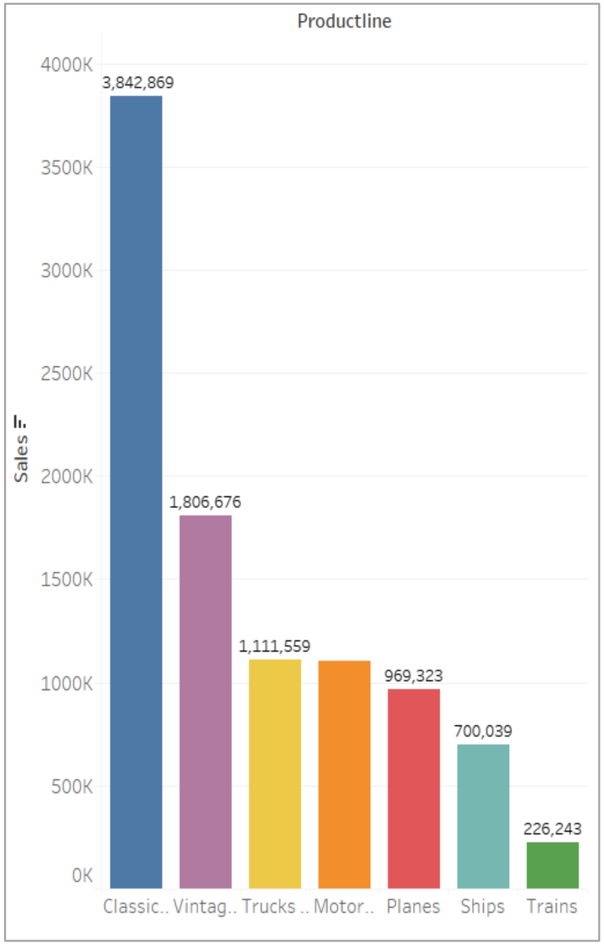
Sales across different cities around the globe

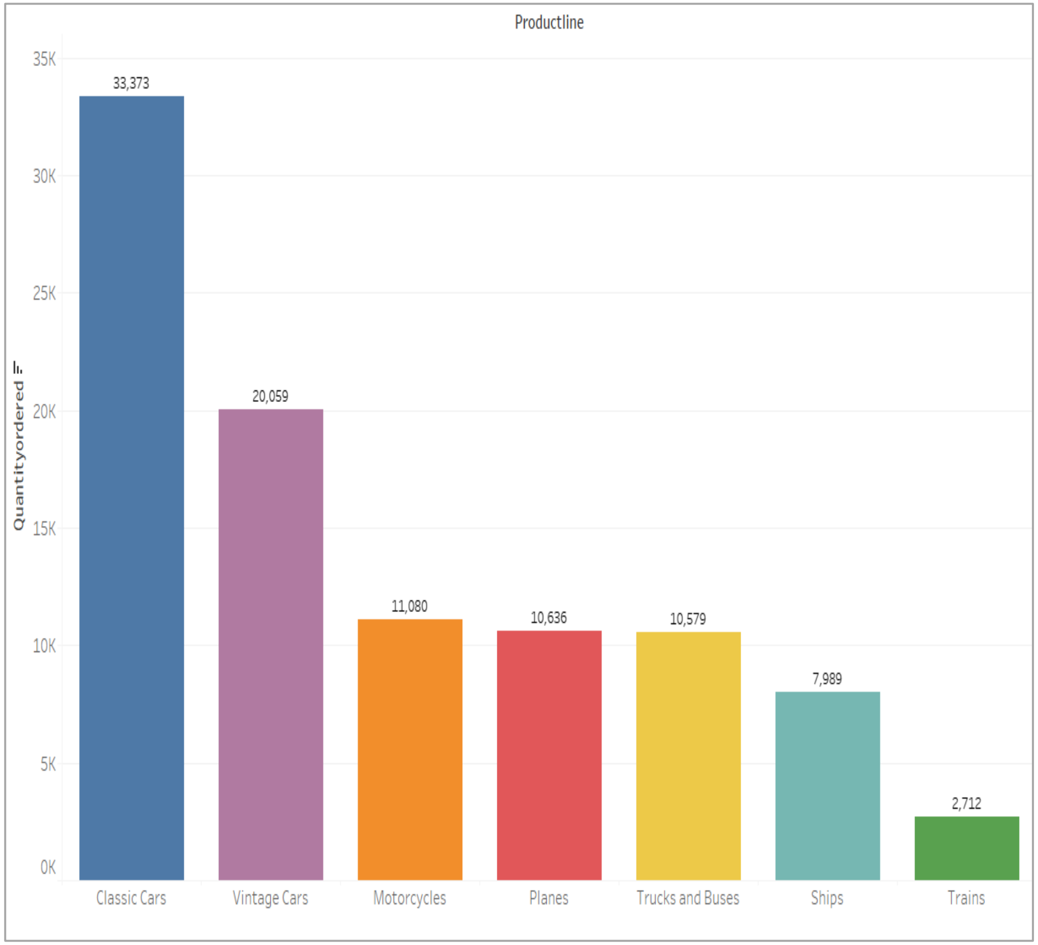


Inferences:

- In USA, San Rafael and NYC are having major sales.
- Madrid makes major turn over entire Globe.

Sales across different Product line & Quantity of orders across different Product line

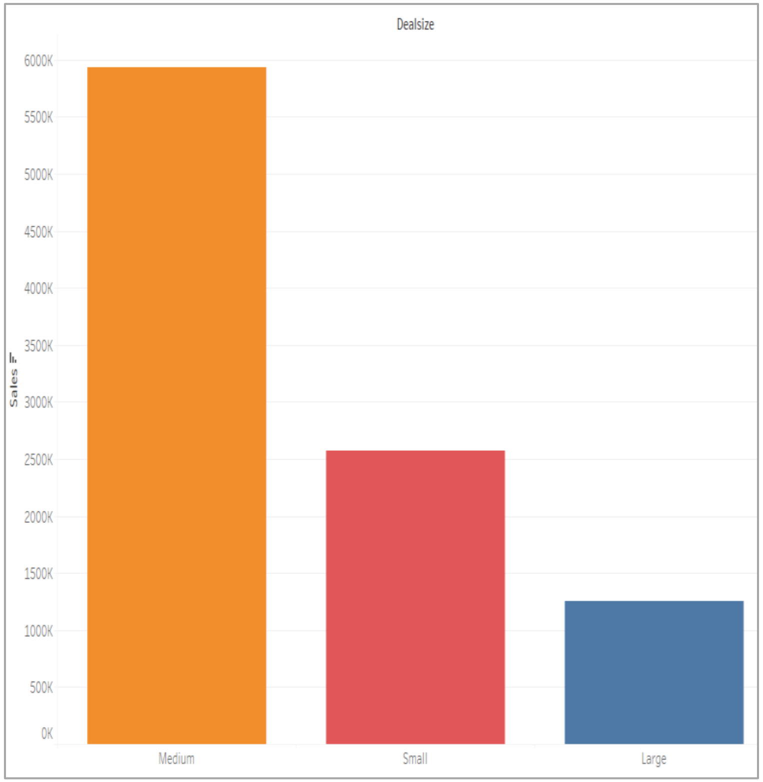


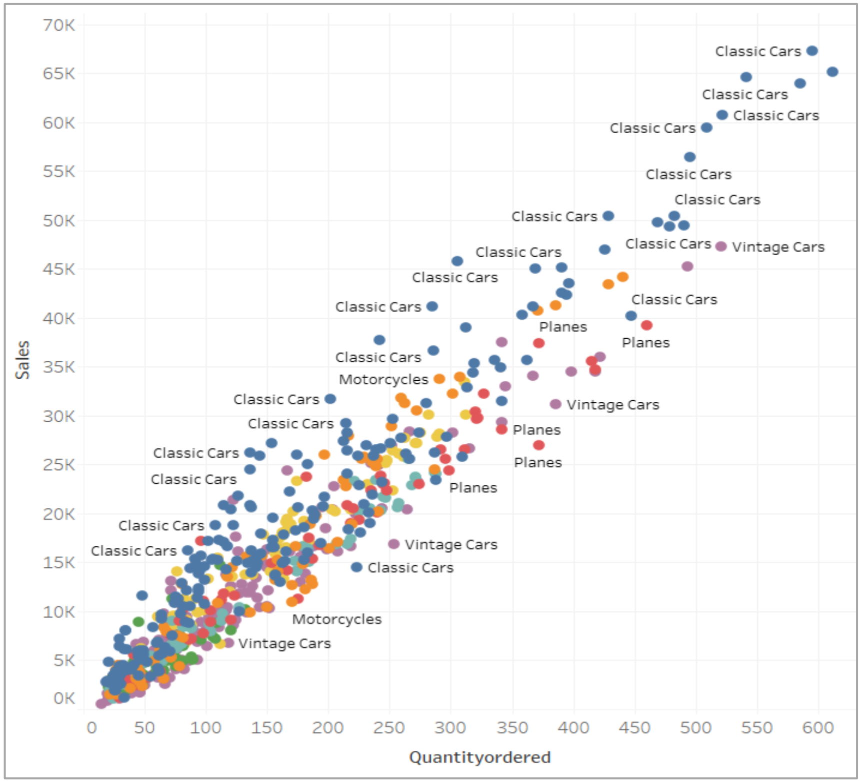


Inferences:

- The Sales and the Quantity of orders are highest on the Classic car products followed by the vintage cars, where the sales of Classic car parts is almost double the sales of vintage car parts.
- In both the cases the Train parts are contributing the least followed by the products of ship.

Sales vs Deal size & Sales, Quantity vs Product line

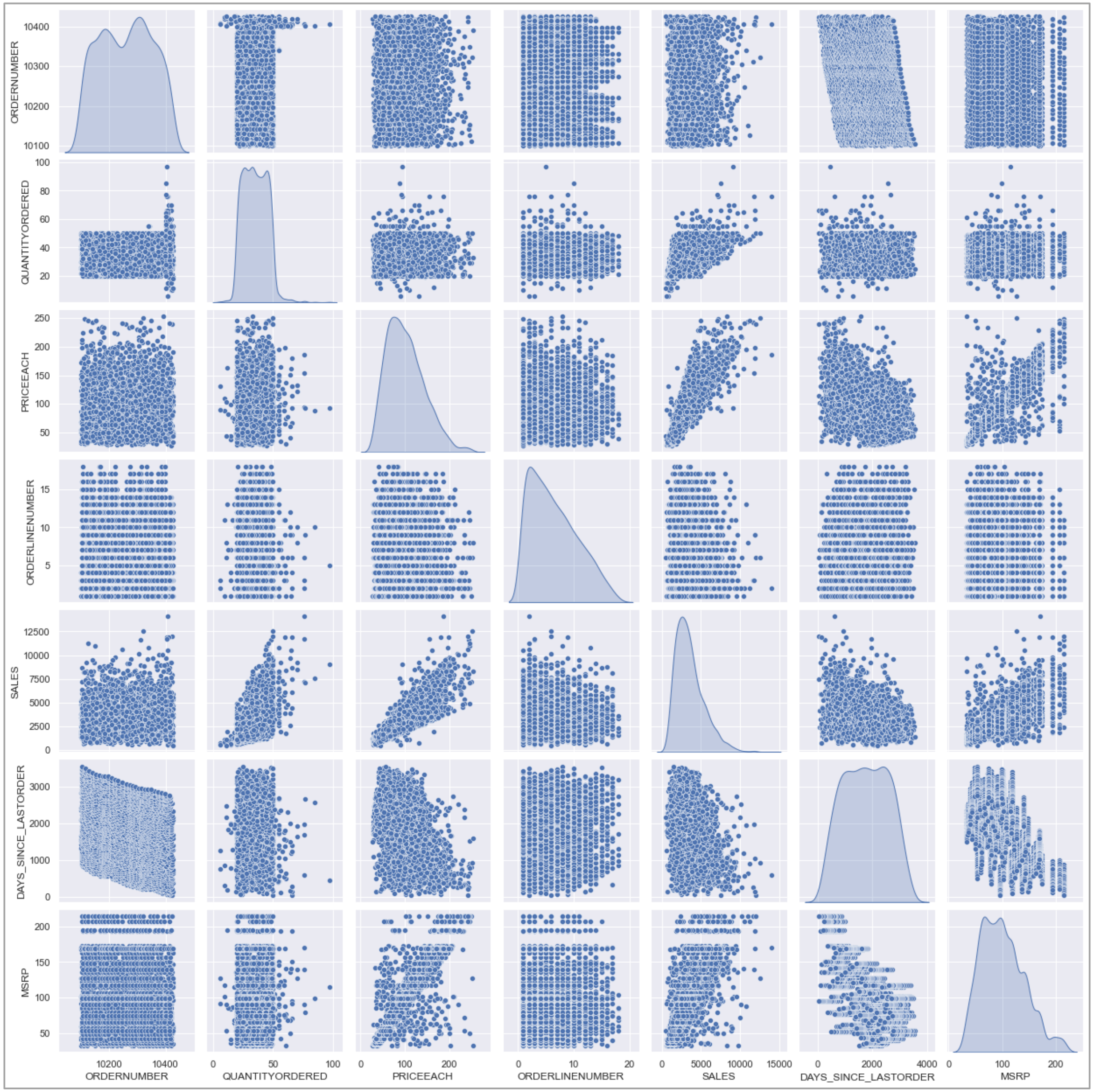




Inferences:

- Clearly we can see that medium seal size are having the maximum sale of around approximately 600K.
- The scatter plot represents the distribution of various product lines with respect to Quantity ordered and Sales price 2. Classic cars products are visibly the most selling product line followed by Vintage cars

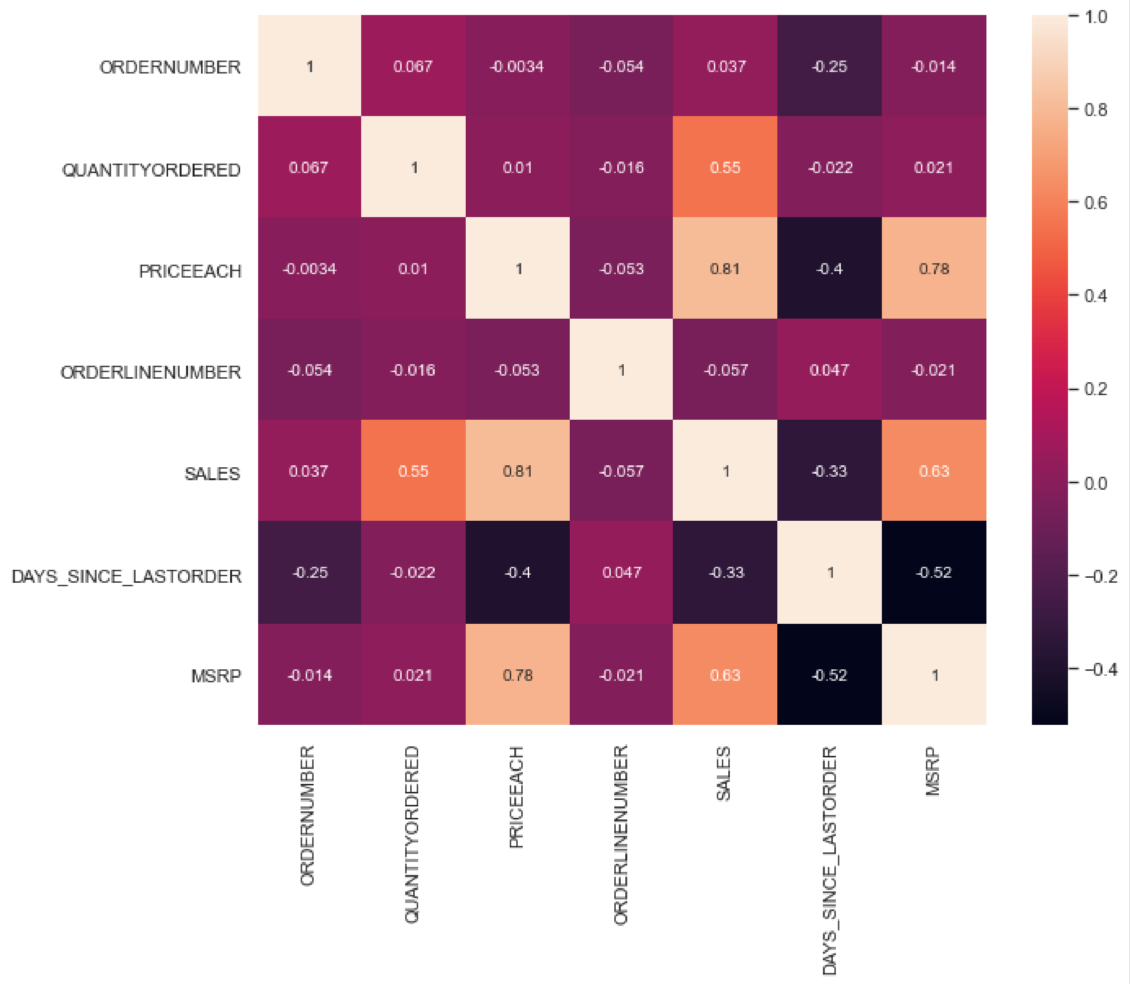
Multivariate Analysis : Pair plot



Inferences:

- The pair plot shows all variables in data are normally distributed or slightly right skewed.
- There is visible correlation between sales, MSRP and price each.
- Other variables don't show any signs of correlation and have cloud like distribution.

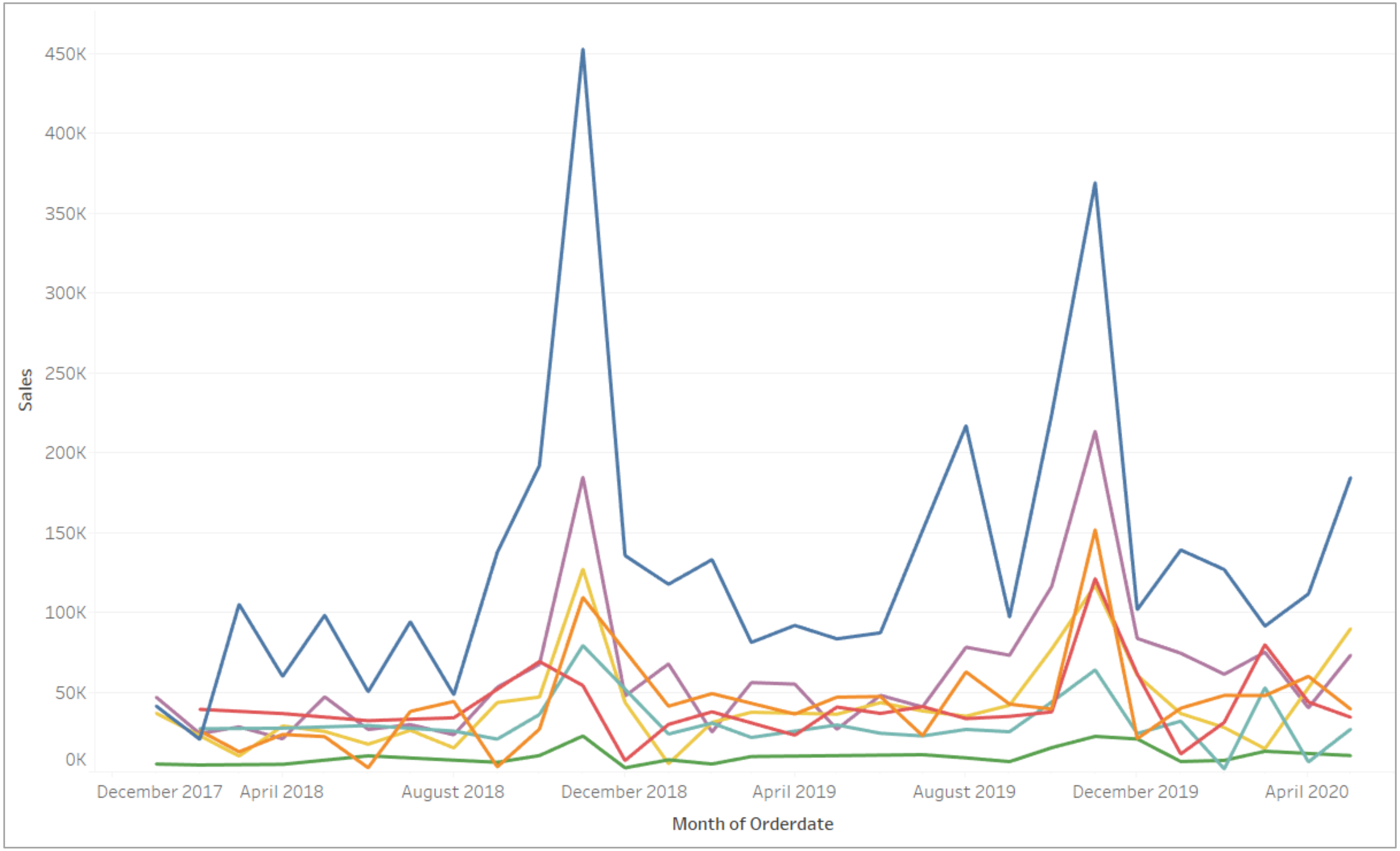
Multivariate Analysis : Heatmap



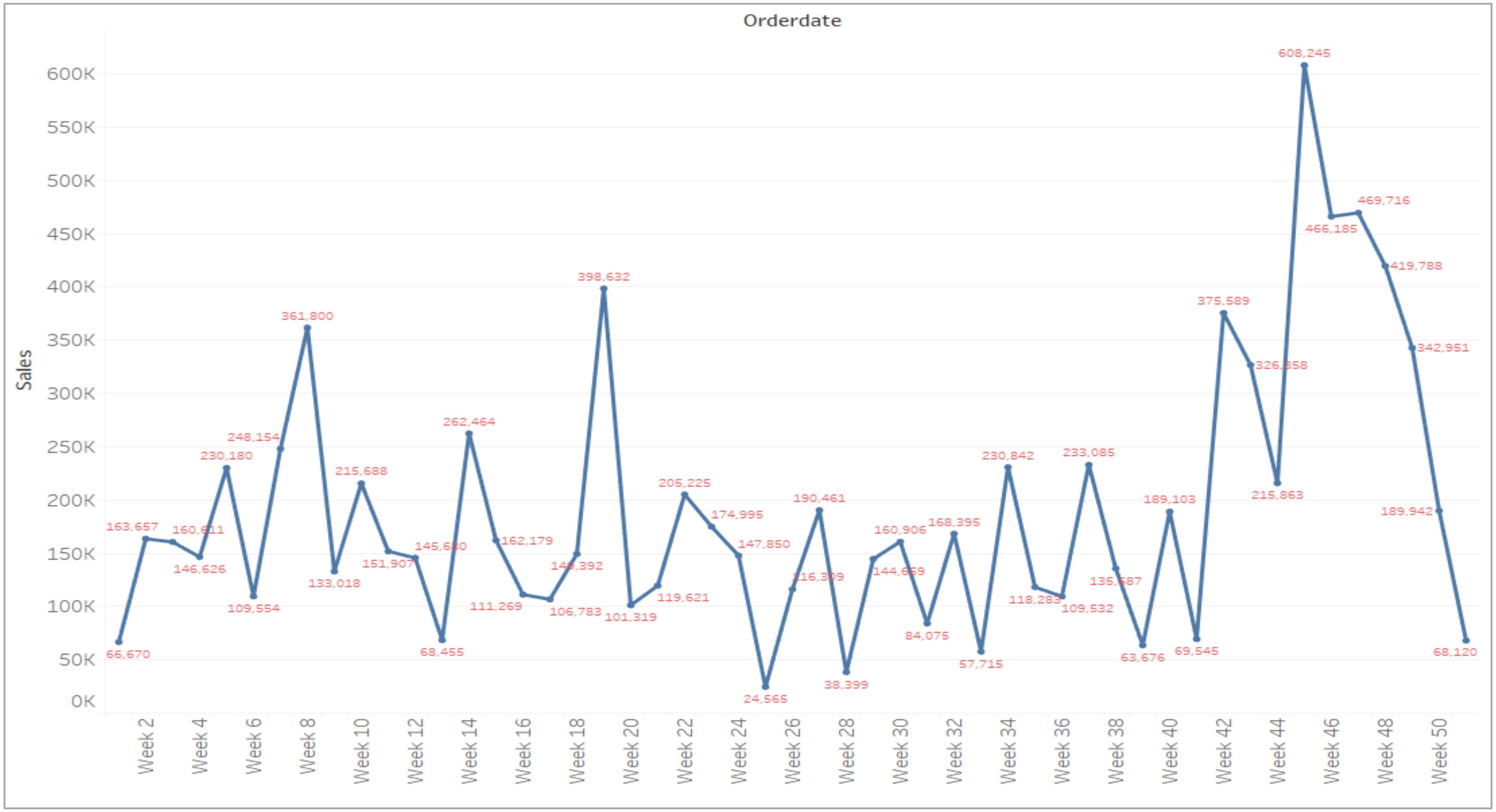
Inferences:

- The heatmap provides asseveration to our observation in pair plot.
- Sales variables is highly correlating with MSRP and Price each
- There is also a partial correlation observed between Quantity of order and Sales.

Sales over Time based on Product line



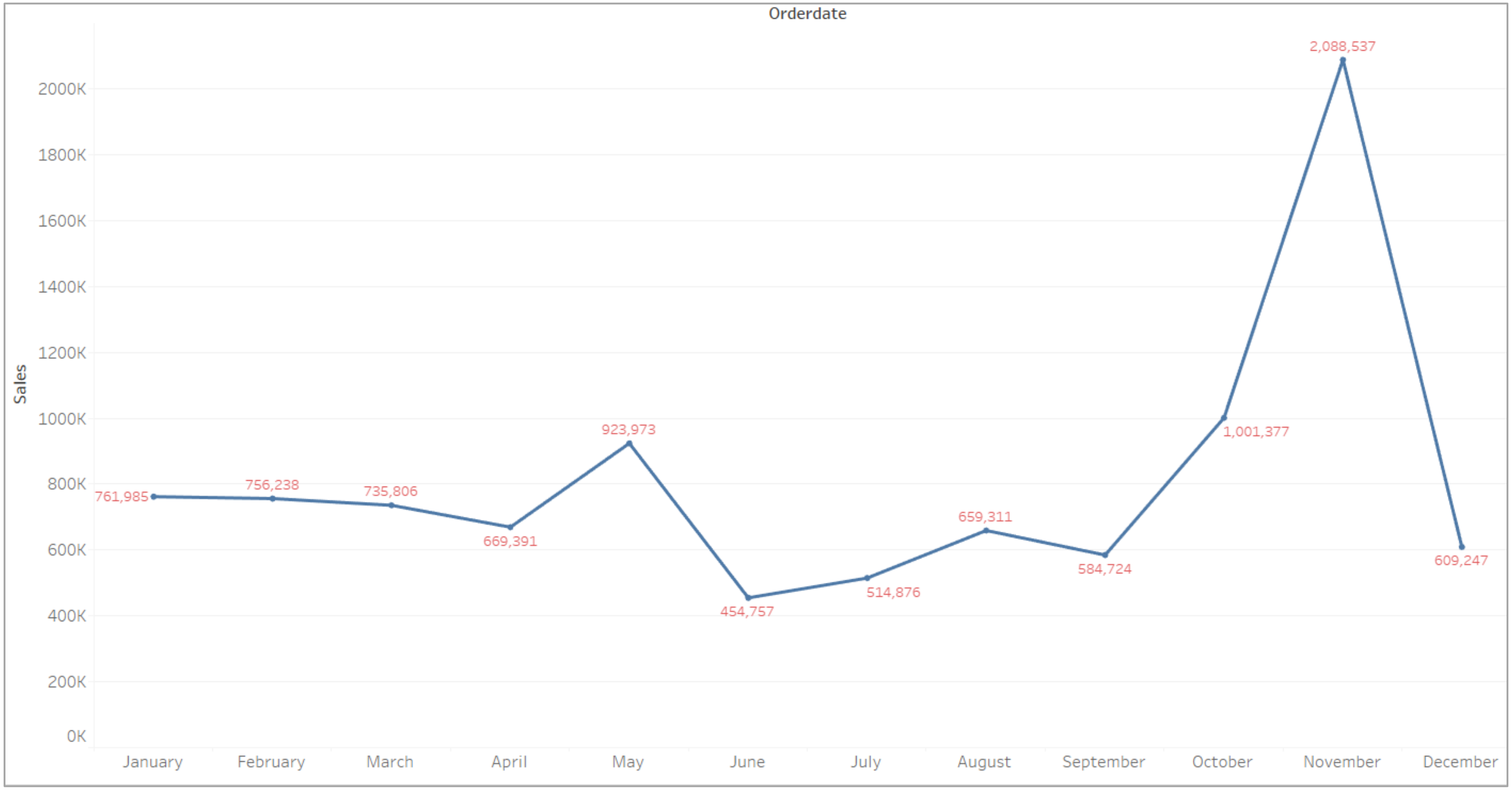
Weekly Trend of Sales



Inferences:

- Sales are highest in the weeks commencing from 42 to 50 and partially low in the mid weeks of 20th to 40th week.
- Sales of 2018 and 2019 are having almost same trend and 2020 is moving with same trend as well.

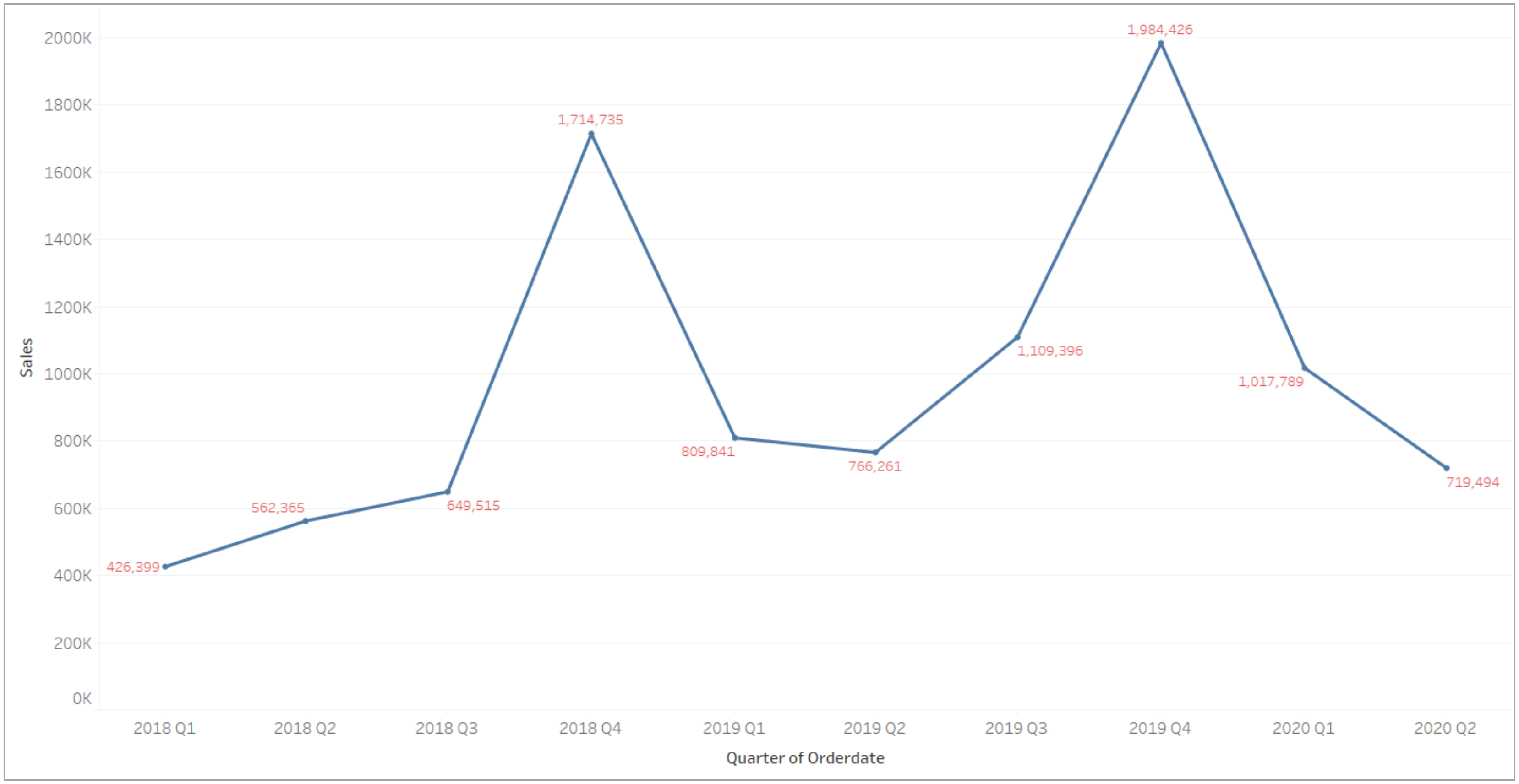
Monthly Trend of Sales



Inferences:

- Highest sales are in the month of November.
- Looks like the sales trend is exalting towards the end of the year i.e; from September to November.

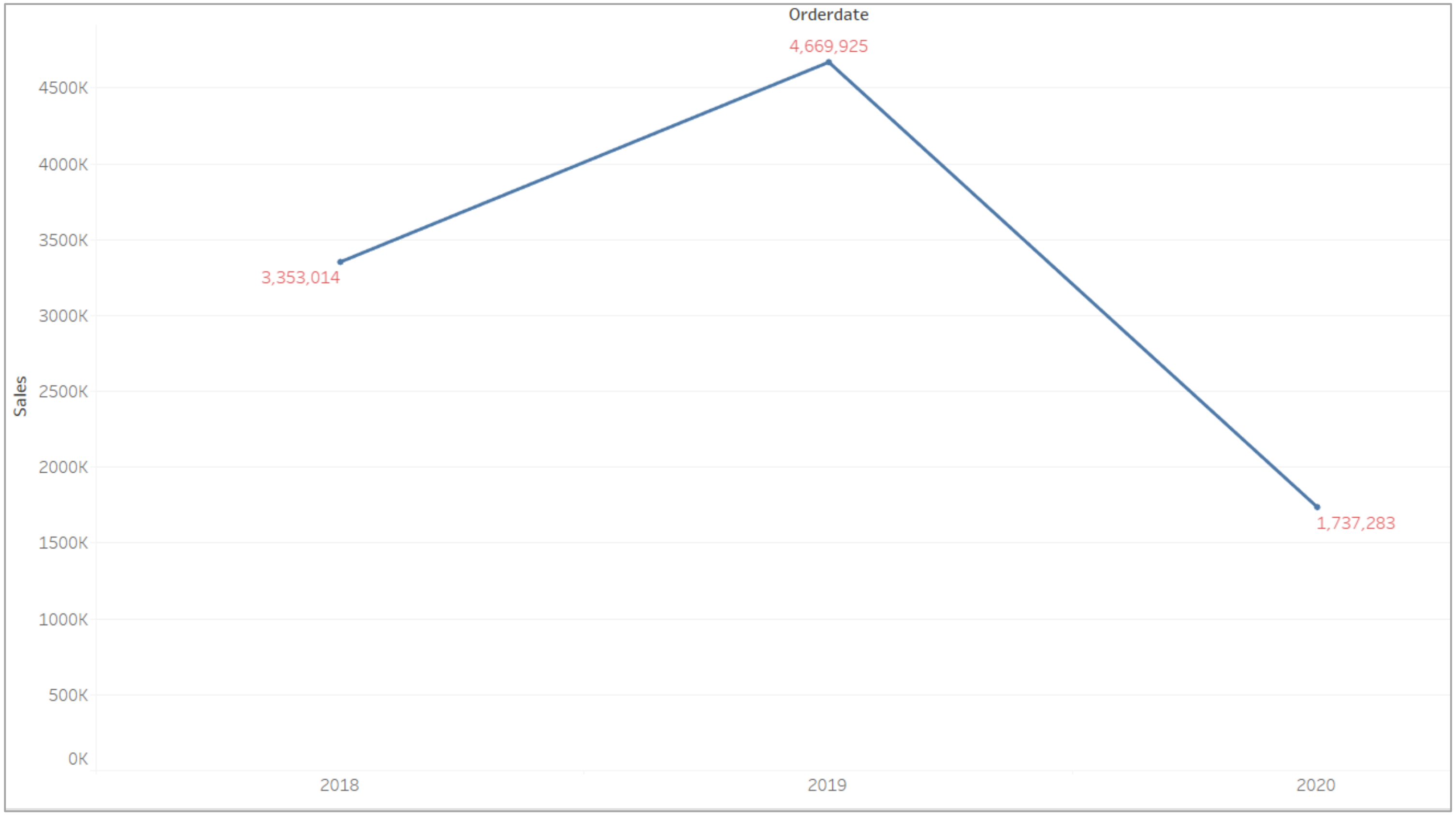
Quarterly Trend of Sales from 2018-20



Inferences:

- Clearly, we can see that the sales are maximum at every 4th quarter affirming from the previous monthly trend.
- The sales trend is least at the 1st quarter and has increasing trend at 4th and shows the seasonality pattern.

Yearly Trend of Sales from 2018-20



<u>Inferences:</u>

- The sales has increased from 2018 to 2019 by approximately more than 30%
- After 2019. the sales have been drastically decreasing.
- Till 2019, the sales were to peak.

EDA Inferences

- USA is the primary market of the company contributing maximum turn-over Some European countries like Spain and Switzerland shows good amount of sales following USA.
- Ireland, Philippines and Belgium are performing very low in terms of sales
- Madrid in Spain, Sand Rafael in USA are the cities which account for maximum sales
- Most gold customers are present in USA market
- Classic cars and Vintage cars product producing good sales. The company can focus on Trains and Ships more to generate more sales across all products.
- The deal size of orders are mostly median size followed small and the least are the large. The company has highest sales in 4^{th} quarter of the year
- The sales trend is least at the 1^{st} quarter and has increasing trend at 4^{th} and shows the seasonality pattern.
- Euro shopping channel present in Spain is the most loyal customer

Customer Segmentation using RFM Analysis

- **What is RFM?**

RFM analysis stands for recency, frequency and monetary value. RFM analysis is a marketing approach that ranks and groups clients statistically based on the recency, frequency, and monetary amount of their recent transactions in order to find the best customers and conduct focused marketing campaigns.

- KNIME tool to perform the RFM analysis on our data

- **What all parameters used and assumptions made?**

Customer Name, Order Date, Price of each item, Quantity ordered and Product code are the major parameters which are used for this RFM analysis.

- **Assumptions made**

❑ For **Recency value**, we take **Max(Order Date) – Order Date**. In our case, Max(Order Date) is equal to 30st May 2020 as mentioned in FAQ.

❑ For **Monetary value,** we take **Price of each item X Quantity ordered**

❑ For **Frequency value,** we group the data set by **Customer Name** column to get Recency, Frequency and Monetary value of customers.

❑ For Recency column binning, lowest value of recency gets best score of 4 while highest value of recency gets least score of 1.

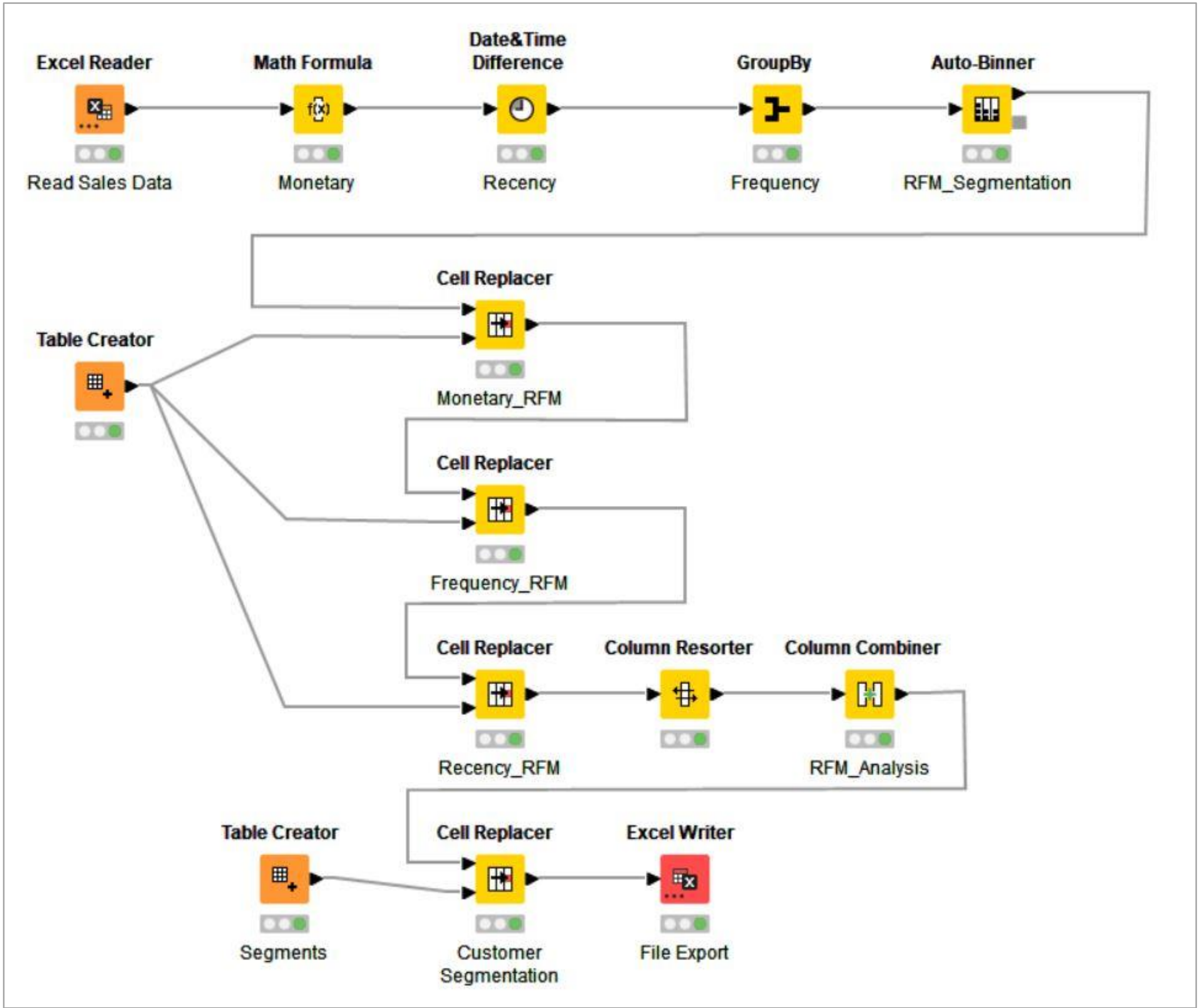
❑ For Frequency and Monetary value binning, highest values get best score of 4 while lowest scores get worst score of 1.

Output Table from KNIME

Row ID	S POSTA...	S COUNTRY	S CONTA...	S CONTA...	I DEALSIZE	S PRODU...	S RECEN...	S FREQU...	S MONET...	D MONET...	L RECENCY	S MONET...	S RECEN...	S RFM_S...	S CUSTOM...
Row0	EC2 5NT	UK	Ashworth	Victoria	51	Bin 4	2	4	4	157,807.81	195	Bin 4	Bin 3	244	Loyal Customer
Row1	31000	France	Roulet	Annette	20	Bin 1	4	1	1	70,488.44	63	Bin 1	Bin 1	411	Lost Customer
Row2	10100	Italy	Accorti	Paolo	26	Bin 2	1	2	3	94,117.26	264	Bin 3	Bin 4	123	Lost Customer
Row3	2060	Australia	O'Hara	Anna	46	Bin 4	3	4	4	153,996.13	82	Bin 4	Bin 2	344	Best Customer
Row4	44000	France	Schmitt	Carine	7	Bin 1	2	1	1	24,179.96	187	Bin 1	Bin 3	211	Lost Customer
Row5	3150	Australia	Connery	Sean	23	Bin 2	4	2	1	64,591.46	21	Bin 1	Bin 1	421	Loyal Customer
Row6	3004	Australia	Ferguson	Peter	55	Bin 4	3	4	4	200,995.41	183	Bin 4	Bin 2	344	Best Customer
Row7	e 4101	Australia	Calaghan	Tony	15	Bin 1	3	1	1	59,469.12	118	Bin 1	Bin 2	311	Lost Customer
Row8	78000	France	Tonini	Daniel	18	Bin 1	1	1	1	64,834.32	232	Bin 1	Bin 4	111	Lost Customer
Row9	75016	France	Perrier	Dominique	27	Bin 3	4	3	3	93,170.66	53	Bin 3	Bin 1	433	Best Customer
Row10	58339	USA	Taylor	Leslie	8	Bin 1	3	1	1	26,479.26	179	Bin 1	Bin 2	311	Lost Customer
Row11	4110	Norway	Bergulfsen	Jonas	32	Bin 3	2	3	3	116,599.19	207	Bin 3	Bin 3	233	Customer on ...
Row12	80686	Germany	Donnermeyer	Michael	14	Bin 1	1	1	1	34,993.92	258	Bin 1	Bin 4	111	Lost Customer
Row13	60528	Germany	Keitel	Roland	22	Bin 2	2	2	2	85,171.59	207	Bin 2	Bin 3	222	Customer on ...
Row14	92561	USA	Young	Leslie	3	Bin 1	3	1	1	9,129.35	112	Bin 1	Bin 2	311	Lost Customer
Row15	28023	Spain	Fernandez	Jesus	13	Bin 1	1	1	1	49,642.05	438	Bin 1	Bin 4	111	Lost Customer
Row16	51247	USA	Tseng	Kyung	11	Bin 1	1	1	1	36,163.62	388	Bin 1	Bin 4	111	Lost Customer
Row17	V3F 2K1	Canada	Tannamuri	Yoshi	22	Bin 2	2	2	2	75,238.92	221	Bin 2	Bin 3	222	Customer on ...
Row18	71270	USA	Cervantes	Francisca	21	Bin 2	2	2	1	67,506.97	229	Bin 1	Bin 3	221	Lost Customer
Row19	10022	USA	Hernandez	Maria	20	Bin 1	2	1	2	77,795.2	191	Bin 2	Bin 3	212	Lost Customer
Row20	2	Ireland	Cassidy	Dean	16	Bin 1	1	1	1	57,756.43	257	Bin 1	Bin 4	111	Lost Customer
Row21	91217	USA	Thompson	Valarie	25	Bin 2	1	2	2	87,489.23	459	Bin 2	Bin 4	122	Lost Customer
Row22	58339	USA	Nelson	Allen	24	Bin 2	3	2	2	81,577.98	131	Bin 2	Bin 2	322	Customer on ...
Row23	28023	Spain	Sommer	Martin	32	Bin 3	2	3	4	120,615.28	211	Bin 4	Bin 3	234	Customer on ...
Row24	1227 MM	Philippines	Cruz	Arnold	26	Bin 2	2	2	3	94,015.73	196	Bin 3	Bin 3	223	Customer on ...
Row25	59000	France	Rance	Martine	20	Bin 1	1	1	1	69,052.41	464	Bin 1	Bin 4	111	Lost Customer
Row26	1734	Denmark	Petersen	Jytte	36	Bin 4	4	4	4	145,041.6	45	Bin 4	Bin 1	444	Best Customer
Row27	70267	USA	Yu	Kyung	31	Bin 3	4	3	4	122,138.14	0	Bin 4	Bin 1	434	Best Customer
Row28	51003	USA	Franco	Valarie	18	Bin 1	1	1	2	70,859.78	400	Bin 2	Bin 4	112	Lost Customer
Row29	WA1 1DP	UK	Hardy	Thomas	12	Bin 1	1	1	1	36,019.04	494	Bin 1	Bin 4	111	Lost Customer
Row30	79903	Singapore	Natividad	Eric	43	Bin 4	3	4	4	172,989.68	89	Bin 4	Bin 2	344	Best Customer
Row31	8022	Spain	Saavedra	Eduardo	23	Bin 2	2	2	2	78,411.86	188	Bin 2	Bin 3	222	Customer on ...
Row32	28034	Spain	Freyre	Diego	259	Bin 4	4	4	4	912,294.11	-1	Bin 4	Bin 1	444	Best Customer
Row33	50553	USA	Benitez	Violeta	26	Bin 2	3	2	3	98,923.73	88	Bin 3	Bin 2	323	Loyal Customer
Row34	97562	USA	King	Julie	25	Bin 2	4	2	3	101,894.79	25	Bin 3	Bin 1	423	Best Customer
Row35	97561	USA	Lewis	Dan	19	Bin 1	3	1	1	57,294.42	178	Bin 1	Bin 2	311	Lost Customer
Row36	51003	USA	Yoshido	Juri	26	Bin 2	4	2	2	83,209.88	24	Bin 2	Bin 1	422	Loyal Customer

- The RFM Analysis is done and customers are segmented into 4 bins from 1 to 4, 1 being the low and 4 the highest.
- After the bins are generated from 1 to 4 based on there values, the RFM score is generated by combing the bind of Recency, Frequency and Monetary in the same sequence.

Work flow image from KNIME



KNIME Workflow Explanation

1 - Excel Reader : Reading all the entries from the excel data file.

2 - Math Formula : Math formula node is used to calculate the Monetary column by take product of price of each item and Quantity ordered.

3 - Date & Time difference : This node is used to calculate the Recency value by considering the fixed date difference. i.e; formula (Max(order date) – order date) where Max(order date) is 30st May 2020 as mentioned in FAQ's.

4 - Groupby : Group by node is used to calculate the frequency of customers orders by grouping by customer name and keeping count of orders and products.

5 - Auto-Binner: Customer segmentation is done by binning the Recency, Frequency and Monetary values into 4 quartiles ranging from 0 - 0.25, 0.25 - 0.5, 0.5 - 0.75 and 0.75 - 1, into Bin 1, Bin 2, Bin 3 and Bin 4 respectively.

6 - Table Creator: To rename the bins, initially a manual table is created, where for monetary and frequency the higher scores has higher values and lower scores has lower values and for Recency high recency value has lower scores while low recency value has higher scores.

7 - Cell replacer nodes : Three cell replacer nodes are used to Recency, Frequency and Monetary to rename the scores of bins and the column of each are appended.

8 - Column Resorter : This node is used to sort the columns of Recency, Frequency and Monetary columns in sequence of RFM.

9 - Column Combiner : This node is used to combine all the values of Recency, Frequency and Monetary into one column of RFM score.

10 - Table Creator : Another table creator is used to assign the 4 categories of customer segmentations based on the RFM score.

11 - Cell replacer : This columns renames the customer segmentations into Best customers, Loyal customers, Customers on the verge of churn and Lost customers based on the previous table creator node.

12 - Excel writer : Finally exporting the RFM analysis into excel.

Inferences from RFM Analysis and identified segments

- According to the RFM analysis, customers are classified into 4 segments, ranging from 1 to 4; 1 being the lowest and 4 being the highest.
- Recency – High recency value has lower scores while low recency value has higher scores
- Monetary and Frequency – High monetary and frequency values have higher scores while low values have lower scores.
- These scores are then combined in a set order of RFM to finally get RFM scores. These scores range from 111 to 444 where 111 is the lowest score while 444 is the highest score.
- Based on these scores, customers are segmented into 4 different categories:
  - **Best Customers**
  - **Loyal Customers**
  - **Customers on the verge of Churning**
  - **Lost Customers**

Recency	**Frequency**	**Monetary**			
		4	**3**	**2**	**1**
4	**4**	Best	Best		
	3	Best	Best		
	2		Best	Loyal	Loyal
	1		Loyal		Lost
3	**4**	Best	Best		
	3		Loyal	Loyal	
	2		Loyal	Churn	
	1				Lost
2	**4**	Loyal			
	3	Churn	Churn		
	2		Churn	Churn	Lost
	1			Lost	Lost
1	**4**	Churn			
	3		Churn		
	2		Lost	Lost	
	1			Lost	Lost

Best Customers [Based on RFM score]

Below mentioned are the customers which fall into Best Customers category based on RFM score.

Best Customers

Customername	Customer Segmentation
Euro Shopping Channel	Best Customer
Mini Gifts Distributors Ltd.	Best Customer
Australian Collectors, Co.	Best Customer
Muscle Machine Inc	Best Customer
La Rochelle Gifts	Best Customer
Dragon Souveniers, Ltd.	Best Customer
The Sharp Gifts Warehouse	Best Customer
Anna's Decorations, Ltd	Best Customer
Souveniers And Things Co.	Best Customer
Salzburg Collectables	Best Customer
Danish Wholesale Imports	Best Customer
L'ordine Souveniers	Best Customer
Reims Collectables	Best Customer
Scandinavian Gift Ideas	Best Customer
Diecast Classics Inc.	Best Customer
Technics Stores Inc.	Best Customer
Tokyo Collectables, Ltd	Best Customer
UK Collectables, Ltd.	Best Customer
Handji Gifts& Co	Best Customer
Mini Creations Ltd.	Best Customer
Gift Depot Inc.	Best Customer
Auto Canal Petit	Best Customer

Customers on the verge of churning
[Based on RFM score]

Below mentioned are the customers which fall into Customers on the verge of Churn category based on RFM score.

Customers on verge of Churn

Customername	Customer Segmentation
Saveley & Henriot, Co.	Customer on the Verge of Churn
Corrida Auto Replicas, Ltd	Customer on the Verge of Churn
Vida Sport, Ltd	Customer on the Verge of Churn
Baane Mini Imports	Customer on the Verge of Churn
Herkku Gifts	Customer on the Verge of Churn
Marta's Replicas Co.	Customer on the Verge of Churn
Heintze Collectables	Customer on the Verge of Churn
Toms Spezialitten, Ltd	Customer on the Verge of Churn
La Corne D'abondance, Co.	Customer on the Verge of Churn
Cruz & Sons Co.	Customer on the Verge of Churn
Vitachrome Inc.	Customer on the Verge of Churn
Mini Classics	Customer on the Verge of Churn
Blauer See Auto, Co.	Customer on the Verge of Churn
Motor Mint Distributors Inc.	Customer on the Verge of Churn
Collectables For Less Inc.	Customer on the Verge of Churn
Enaco Distributors	Customer on the Verge of Churn
giftsbymail.co.uk	Customer on the Verge of Churn
Canadian Gift Exchange Network	Customer on the Verge of Churn
Marseille Mini Autos	Customer on the Verge of Churn

Lost Customers [Based on RFM score]

Below mentioned are the customers which fall into Lost Customers category based on RFM score.

Lost Customers

Customername	Customer Segmentation
Auto Assoc. & Cie.	Lost Customer
Bavarian Collectables Imports, Co.	Lost Customer
CAF Imports	Lost Customer
Cambridge Collectables Co.	Lost Customer
Clover Collections, Co.	Lost Customer
Daedalus Designs Imports	Lost Customer
Double Decker Gift Stores, Ltd	Lost Customer
Iberia Gift Imports, Corp.	Lost Customer
Online Mini Collectables	Lost Customer
Osaka Souveniers Co.	Lost Customer
Signal Collectibles Ltd.	Lost Customer
West Coast Collectables Co.	Lost Customer
Diecast Collectables	Lost Customer
Super Scale Inc.	Lost Customer
Collectable Mini Designs Co.	Lost Customer
Norway Gifts By Mail, Co.	Lost Customer
Royal Canadian Collectables, Ltd.	Lost Customer
Amica Models & Co.	Lost Customer
Atelier graphique	Lost Customer
Microscale Inc.	Lost Customer
Classic Legends Inc.	Lost Customer
Volvo Model Replicas, Co	Lost Customer
Classic Gift Ideas, Inc	Lost Customer
Australian Gift Network, Co	Lost Customer
Auto-Moto Classics Inc.	Lost Customer
Boards & Toys Co.	Lost Customer
Gift Ideas Corp.	Lost Customer
Mini Auto Werke	Lost Customer
Royale Belge	Lost Customer
Alpha Cognac	Lost Customer

Loyal Customers [Based on RFM score]

Below mentioned are the customers which fall into Loyal Customers category based on RFM score.

Loyal Customers

Customername	Customer Segmentation
Land of Toys Inc.	Loyal Customer
AV Stores, Co.	Loyal Customer
Rovelli Gifts	Loyal Customer
Online Diecast Creations Co.	Loyal Customer
Suominen Souveniers	Loyal Customer
Toys of Finland, Co.	Loyal Customer
Toys4GrownUps.com	Loyal Customer
Oulu Toy Supplies, Inc.	Loyal Customer
FunGiftIdeas.com	Loyal Customer
Stylish Desk Decors, Co.	Loyal Customer
Tekni Collectables Inc.	Loyal Customer
Gifts4AllAges.com	Loyal Customer
Signal Gift Stores	Loyal Customer
Mini Caravy	Loyal Customer
Lyon Souveniers	Loyal Customer
Petit Auto	Loyal Customer
Quebec Home Shopping Network	Loyal Customer
Australian Collectables, Ltd	Loyal Customer

Recommendations:

The best customers and Loyal customers are top priority. Focus can be given on the customers who are on the verge of churning.

The company can initiate loyalty programs and campaigns for customer relationship.

This model will help the company to maintain its sales and customers and can focus on criteria's for many lost customers & can take actions.

In addition to US market, the company can invest and do more business in Spain, France and Australia, as they have promising sales performance.

In order to venture into new markets, Classic cars and Vintage cars product line will be most effective since most of the orders and sales are on it.

Reference Links:

1- EDA Data visualisation: https://public.tableau.com/app/profile/pooja.kabadi8245/viz/MRA-Milwstone1/SalesDistribution?publish=yes

2- RFM Customer segmentation : https://public.tableau.com/app/profile/pooja.kabadi8245/viz/RFMCustomersegmentation/BestCustomers?publish=yes

3 - Python file: Attached along with presentation pdf

4 – **KNIME :** Attached along with presentation pdf

Thank you