Project on Marketing & Retail Analytics

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.

Contents:

- Basic Information of the data
- Results of Exploratory Data Analysis
- Results of RFM Analysis
- Results of Segmentation of data

Summary:

Exploratory Data Analysis (EDA) was done on the dataset to understand whether data had any discrepancies. EDA helped in drawing useful insights & inferences from the data to understand business & customer patterns. Recency, Frequency & Monetary analysis was used to create bins based on quantile values and understand customer distribution in much better fashion. This helped in segmenting the customers, by which customers can be analyzed easily and new strategies for further business growth can also be devised.

Basic Information of the dataset:

Basic Information of the dataset:

```
<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
```

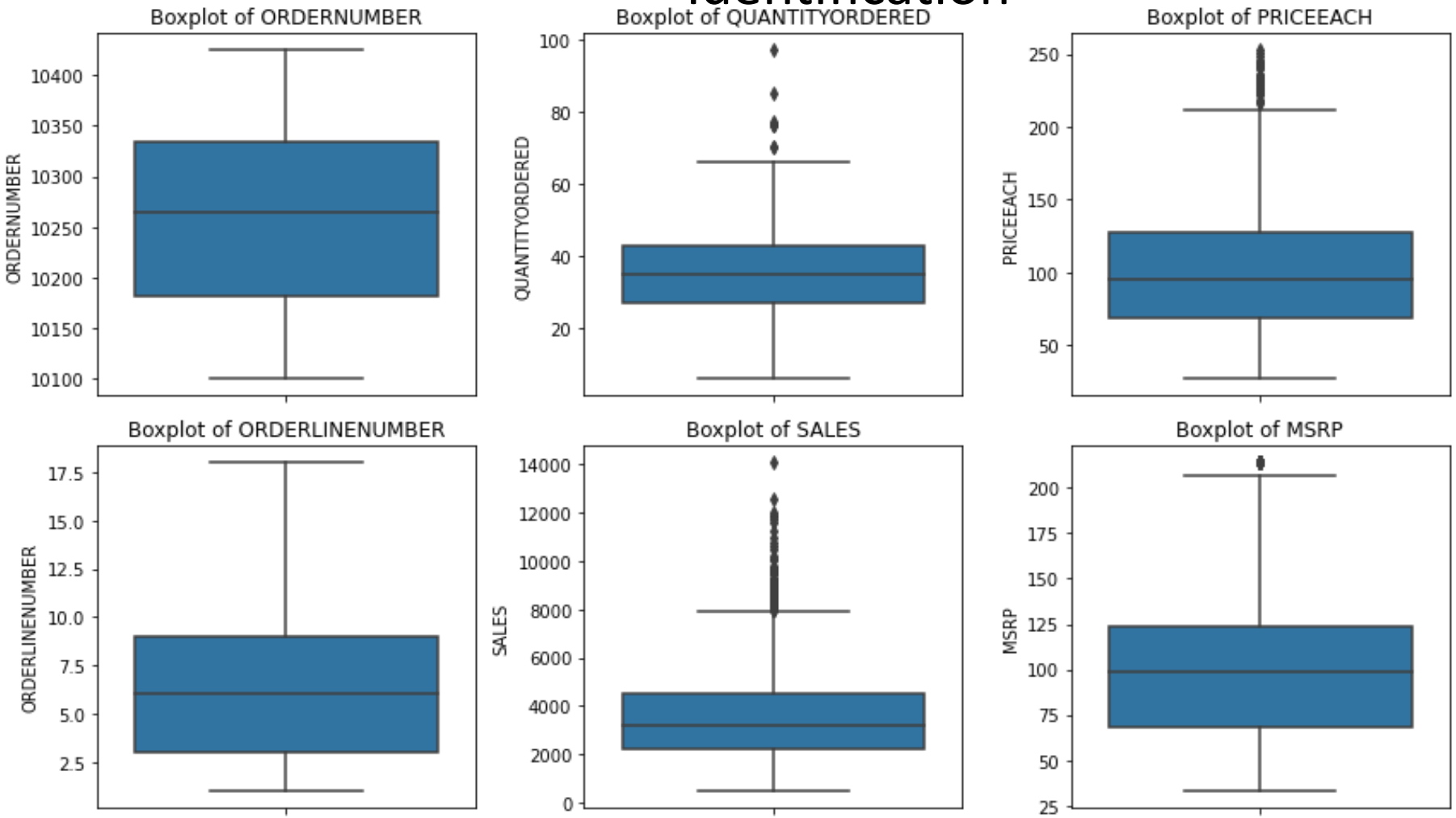
Insights:

- There are no null & duplicate values in the dataset
- There are 12 object, 7 numerical and 1 date-time datatype variables.
- Summary table includes minimum, maximum, quantile values, count, mean and std. deviation values of all numerical variables.
- Certain redundant variables/columns were removed for ease of working.

Summary of the dataset:

	ORDERNUMBER	QUANTITYORDERED	PRICEEACH	ORDERLINENUMBER	SALES	MSRP
count	2747.000000	2747.000000	2747.000000	2747.000000	2747.000000	2747.000000
mean	10259.761558	35.103021	101.098951	6.491081	3553.047583	100.691664
std	91.877521	9.762135	42.042548	4.230544	1838.953901	40.114802
min	10100.000000	6.000000	26.880000	1.000000	482.130000	33.000000
25%	10181.000000	27.000000	68.745000	3.000000	2204.350000	68.000000
50%	10264.000000	35.000000	95.550000	6.000000	3184.800000	99.000000
75%	10334.500000	43.000000	127.100000	9.000000	4503.095000	124.000000
max	10425.000000	97.000000	252.870000	18.000000	14082.800000	214.000000

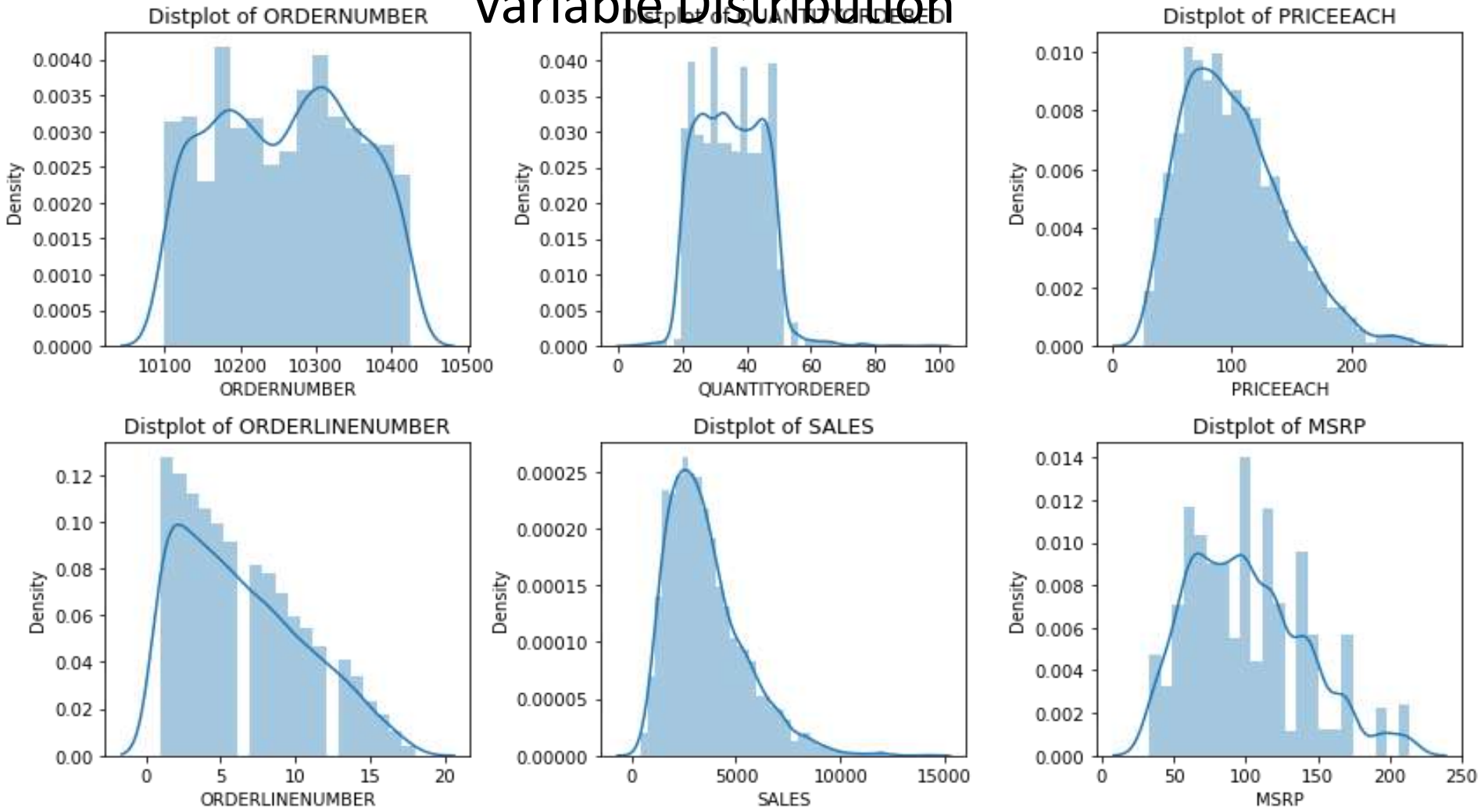
Univariate Analysis: Boxplot for Outlier Identification



Insights:

From the plot above, we can say that, variables 'QUANTITYORDERED, 'PRICEEACH', 'SALES' and 'MSRP' are the only ones which have outliers and since, they are less in numbers, they can be ignored.

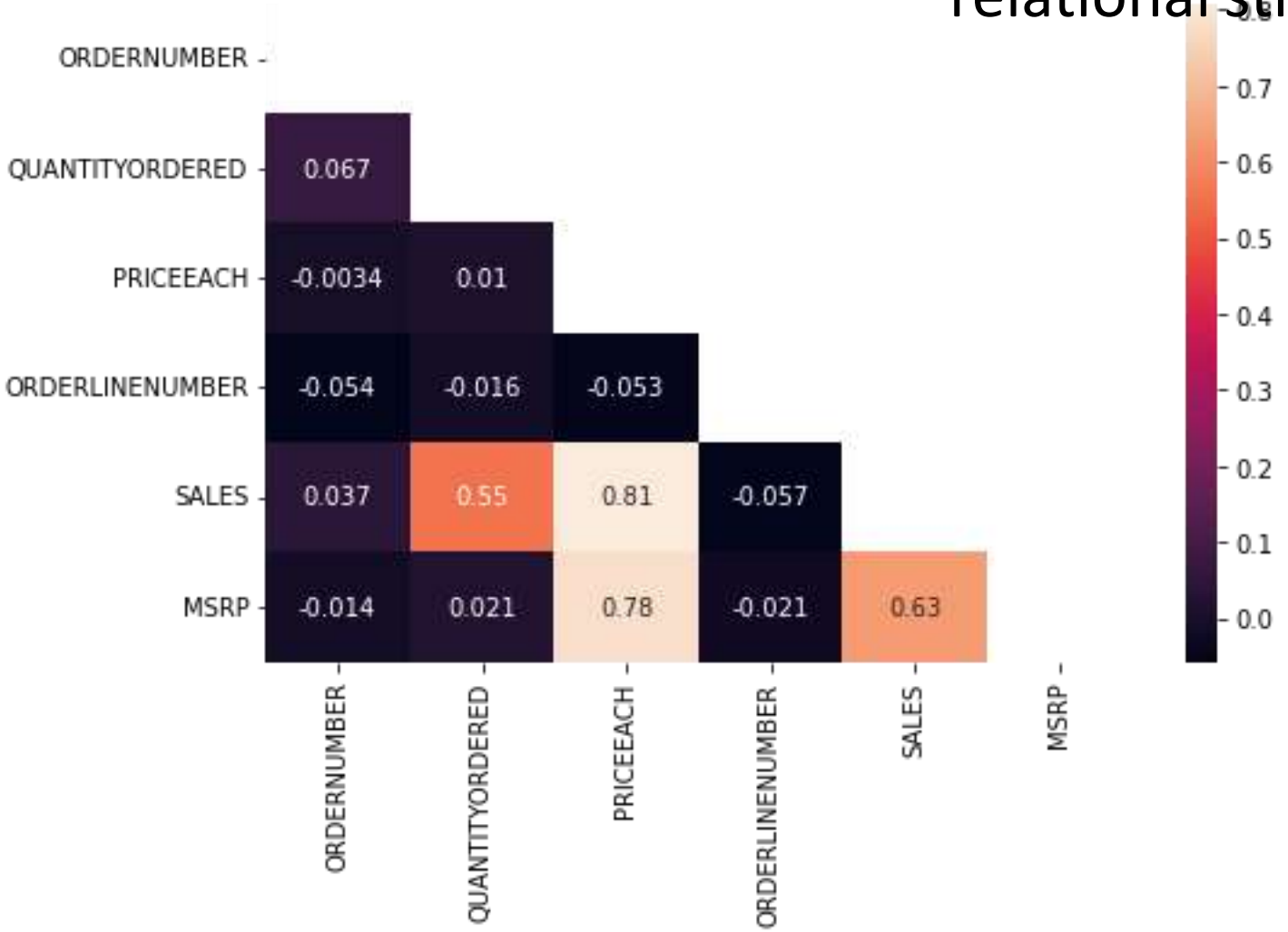
Univariate Analysis: Distplot for Variable Distribution



Insights:

From the plot above, we can say that, except for the variables ‘SALES’ and ‘ORDERLINENUMBER’ all the other variables show symmetric distribution.

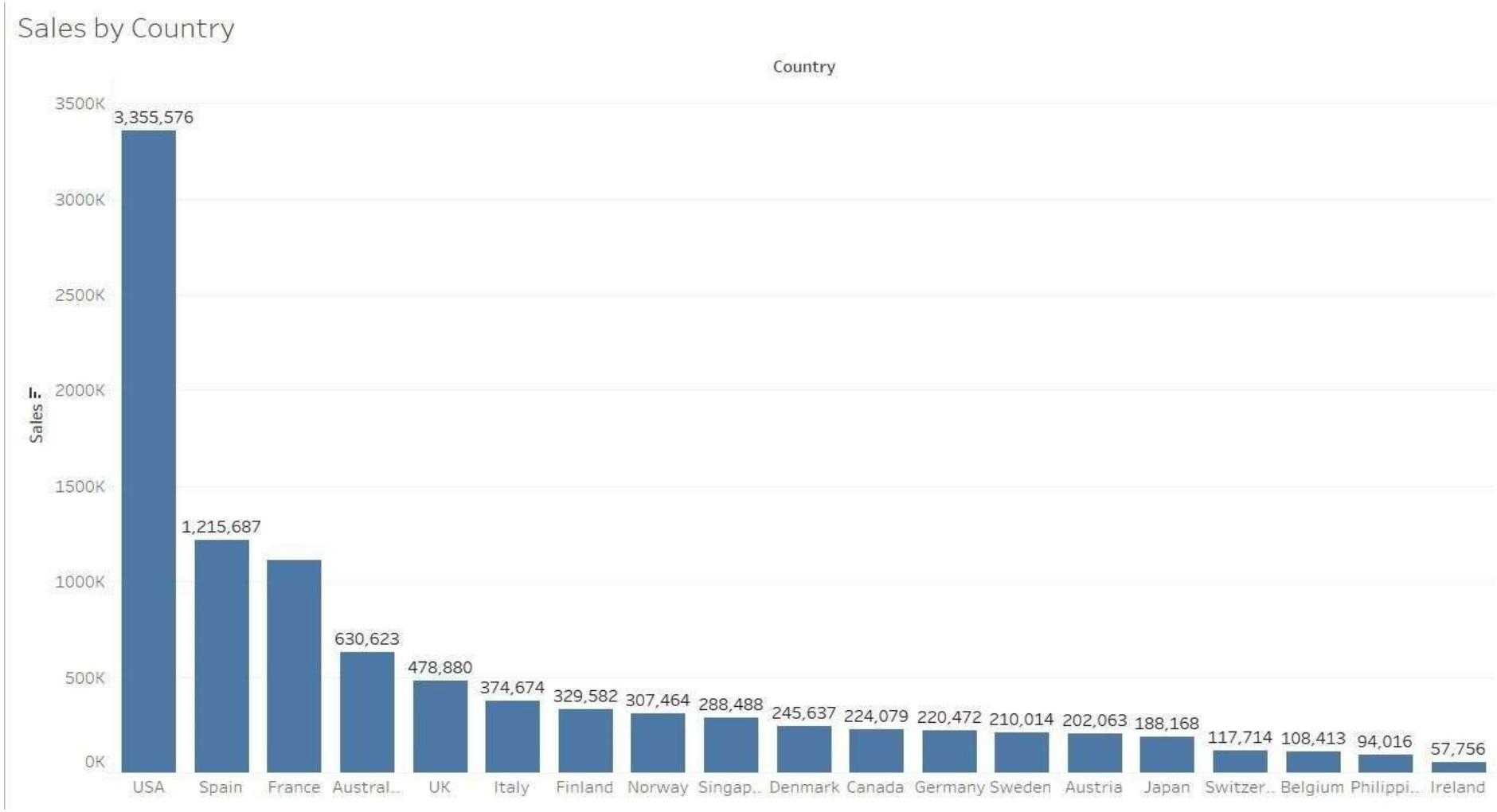
Multivariate Analysis: Correlation Heatmap for variables' relational strength



Insights:

From the plot above, we can say that, variables 'SALES' & 'MSRP' are strongly correlated to variable 'PRICEEACH' and variable 'SALES' is moderately correlated to variables 'QUANTITYORDERED' & 'MSRP'.

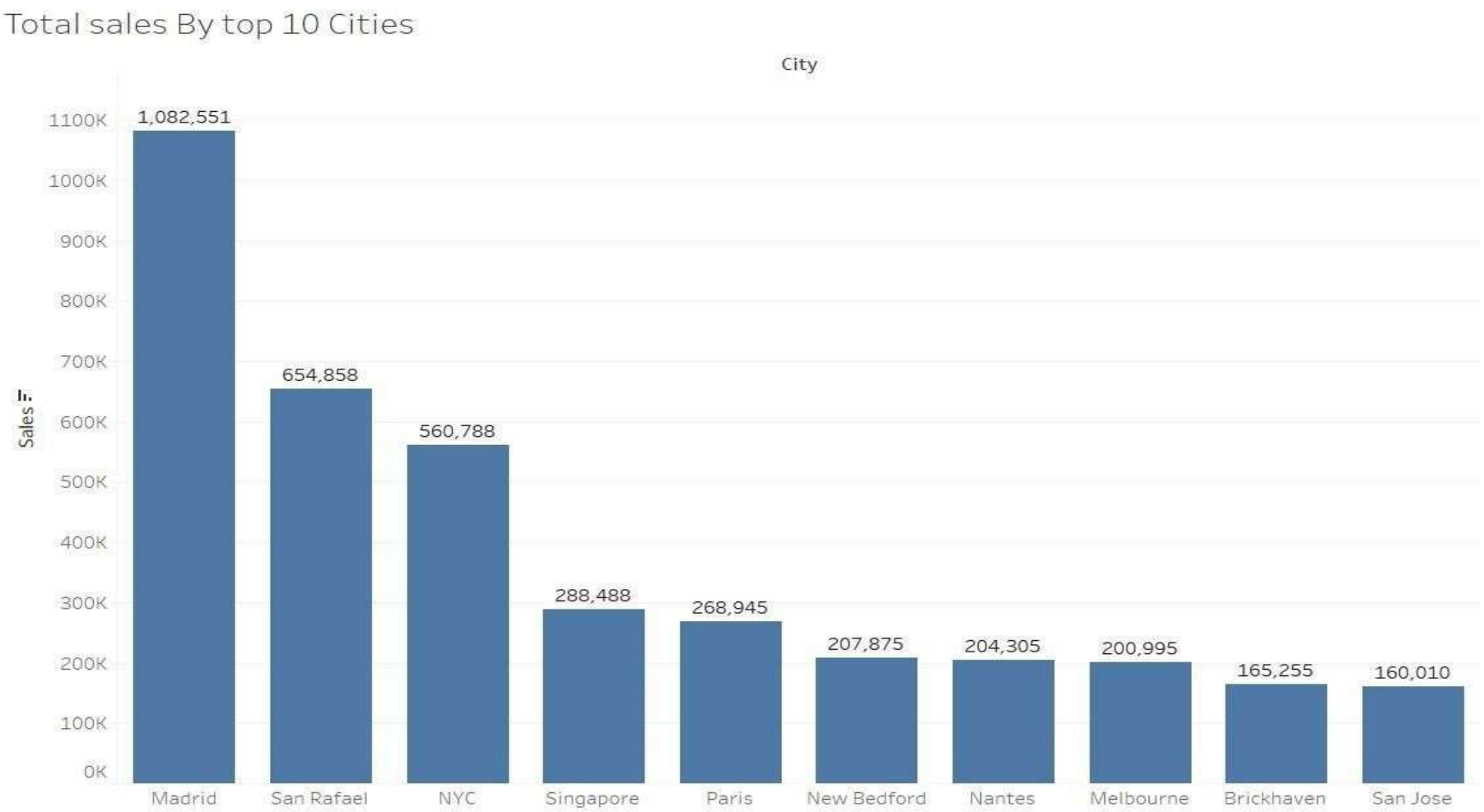
Multivariate Analysis: Total Sales for each Country



Insights:

From the plot above, we can say that, highest sales is in ‘USA’ whereas the lowest is in ‘Ireland’.

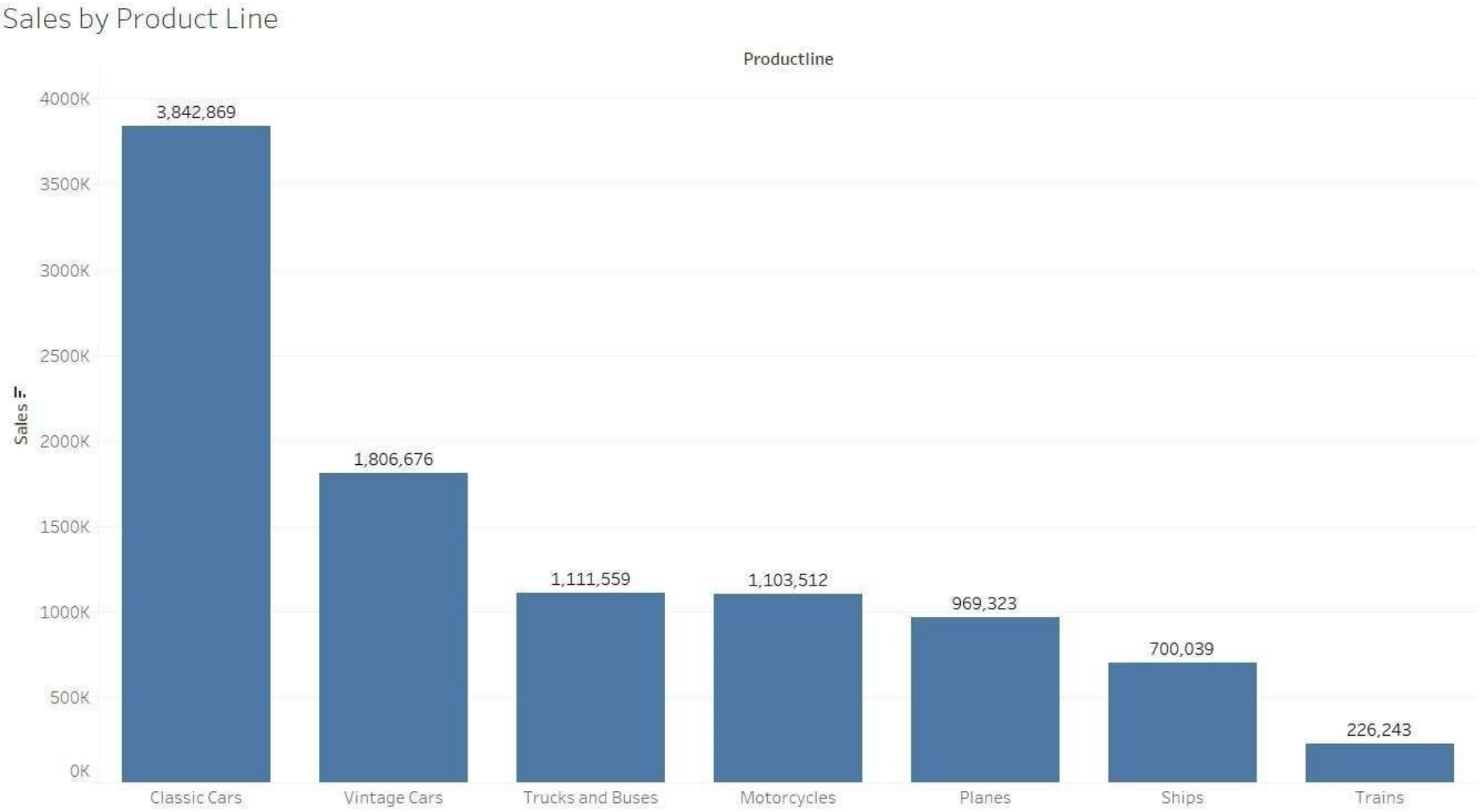
Multivariate Analysis: Total Sales for top 10 Cities



Insights:

From the plot above, we can say that, highest sales is in ‘Madrid’ followed by ‘San Rafael’ and ‘NYC’.

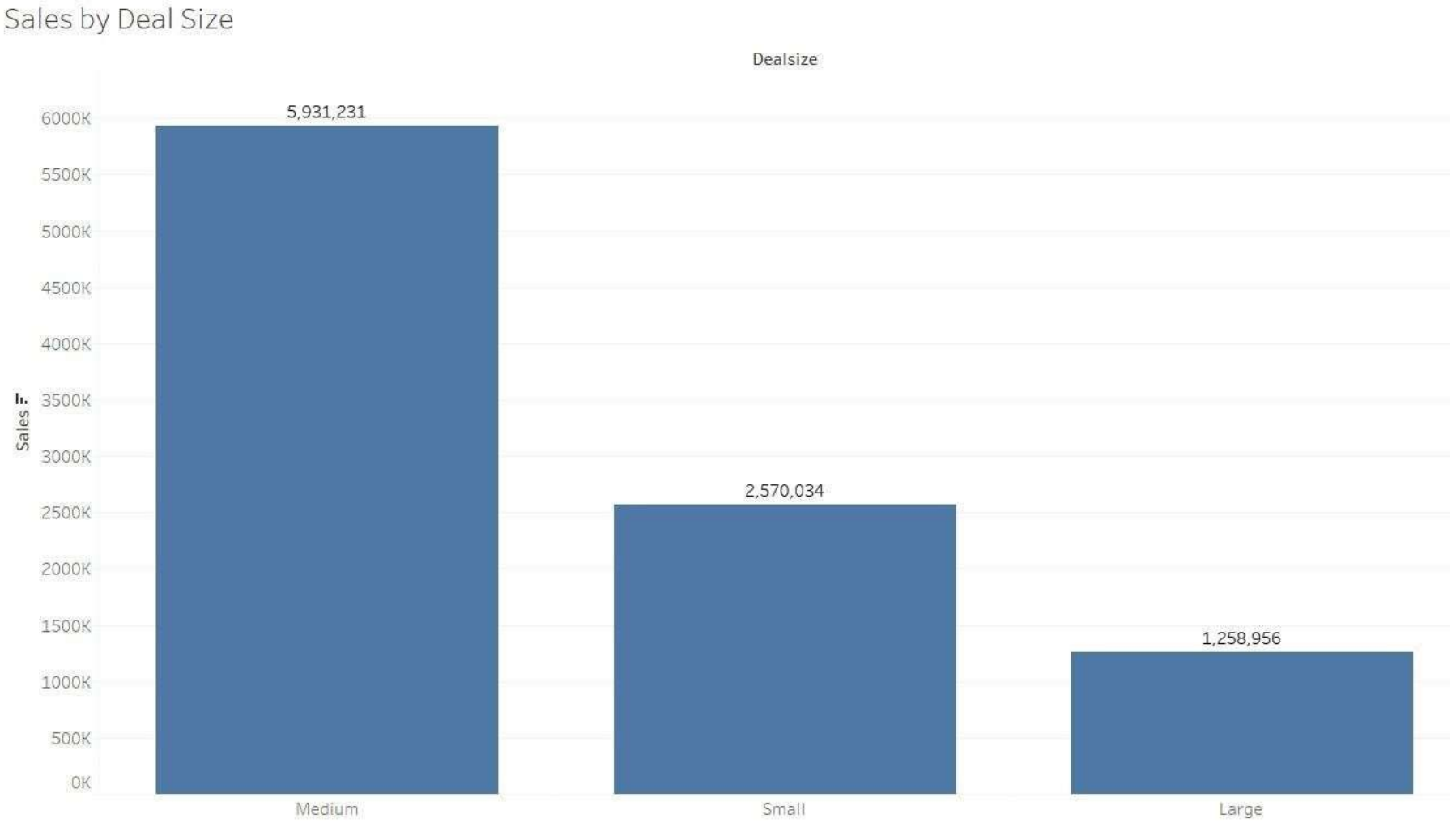
Multivariate Analysis: Total Sales for Product Line



Insights:

From the plot above, we can say that, highest sales is for ‘Classic’ Cars followed by ‘Vintage’ cars, whereas, ‘Trains’ has the lowest sales.

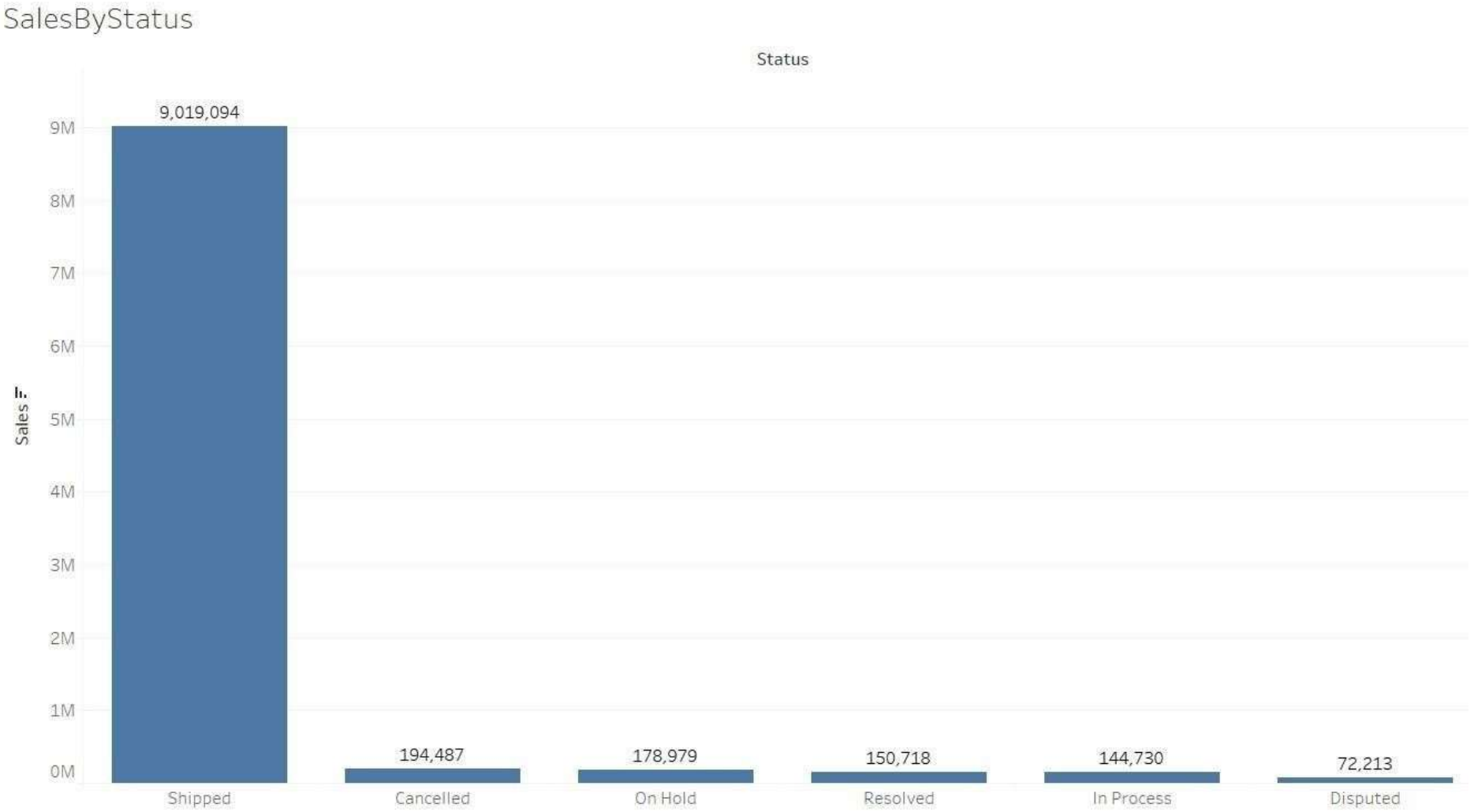
Multivariate Analysis: Total Sales for Deal Size



Insights:

From the plot above, we can say that, highest sales is for ‘Medium’ deal size whereas it is lowest for ‘Large’ deal size.

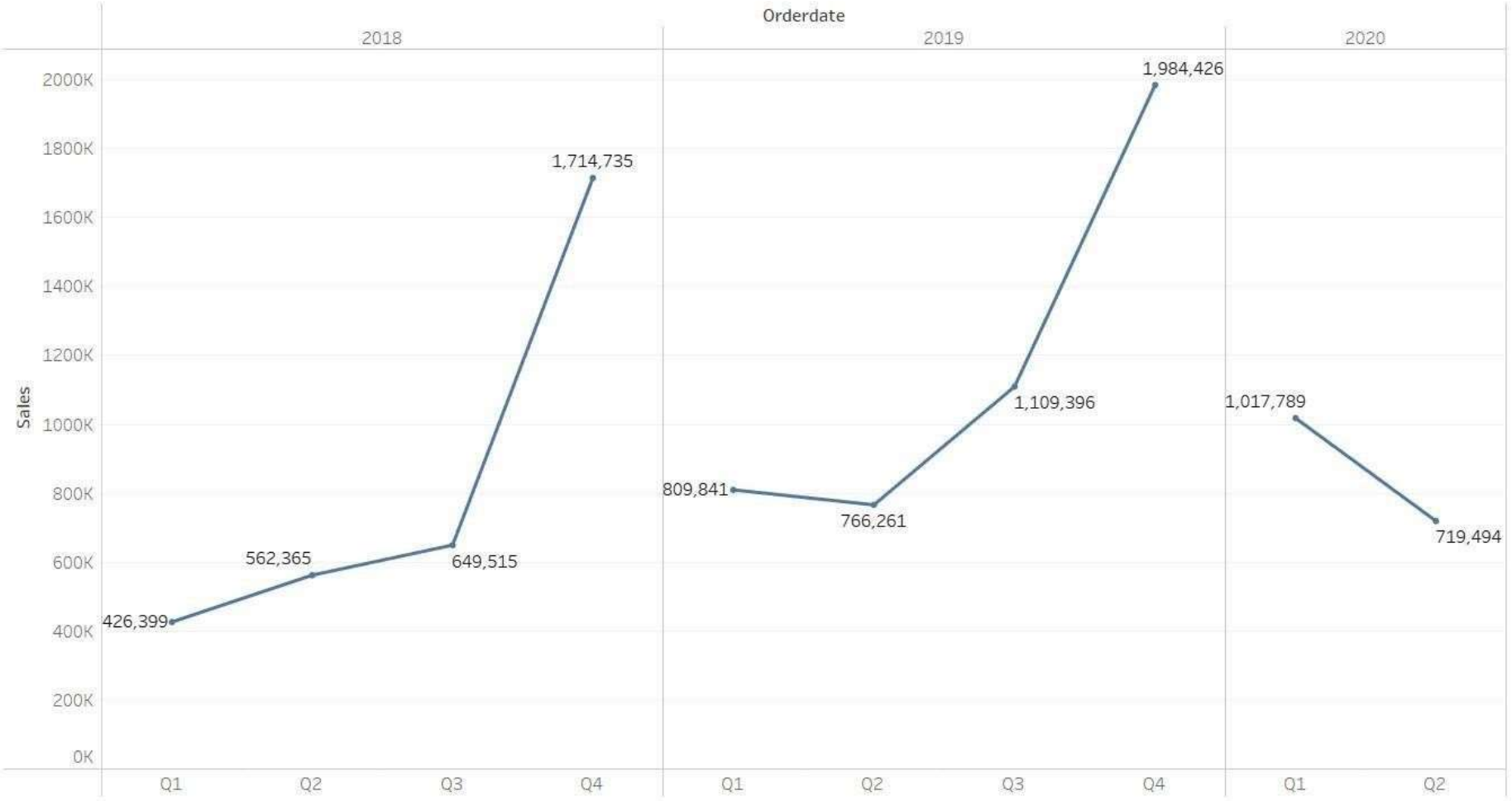
Multivariate Analysis: Total Sales for Status



Insights:

From the plot above, we can say that, highest sales is for ‘Shipped’ status, the rest all are significantly low.

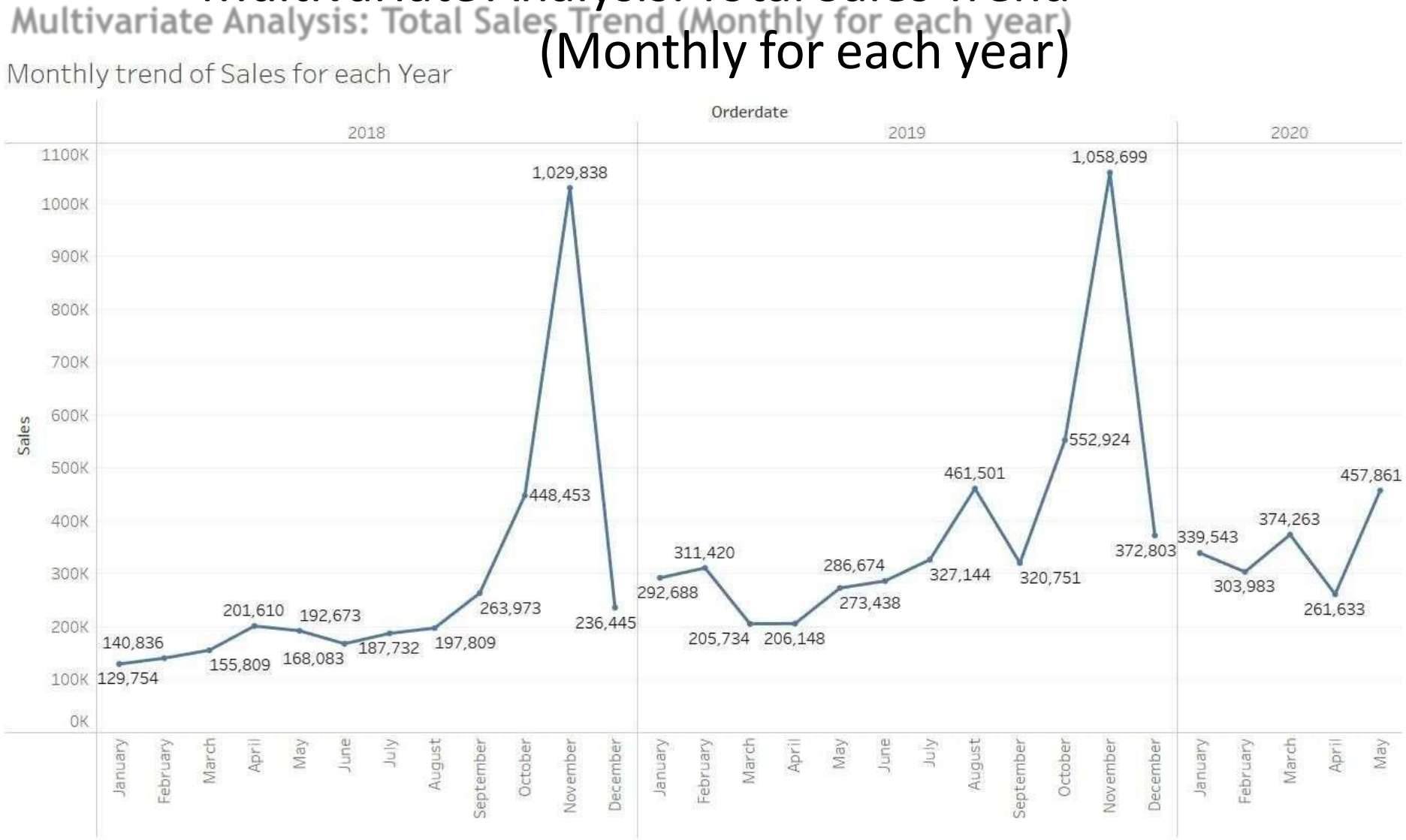
Multivariate Analysis: Total Sales Trend (Quarterly for each year)



Insights:

From the plot above, we can say that, 4th Quarter has higher sales compared to other quarters in the year 2018 & 2019. Whereas, in the year 2020 1st Quarter had more compared to 2nd Quarter.

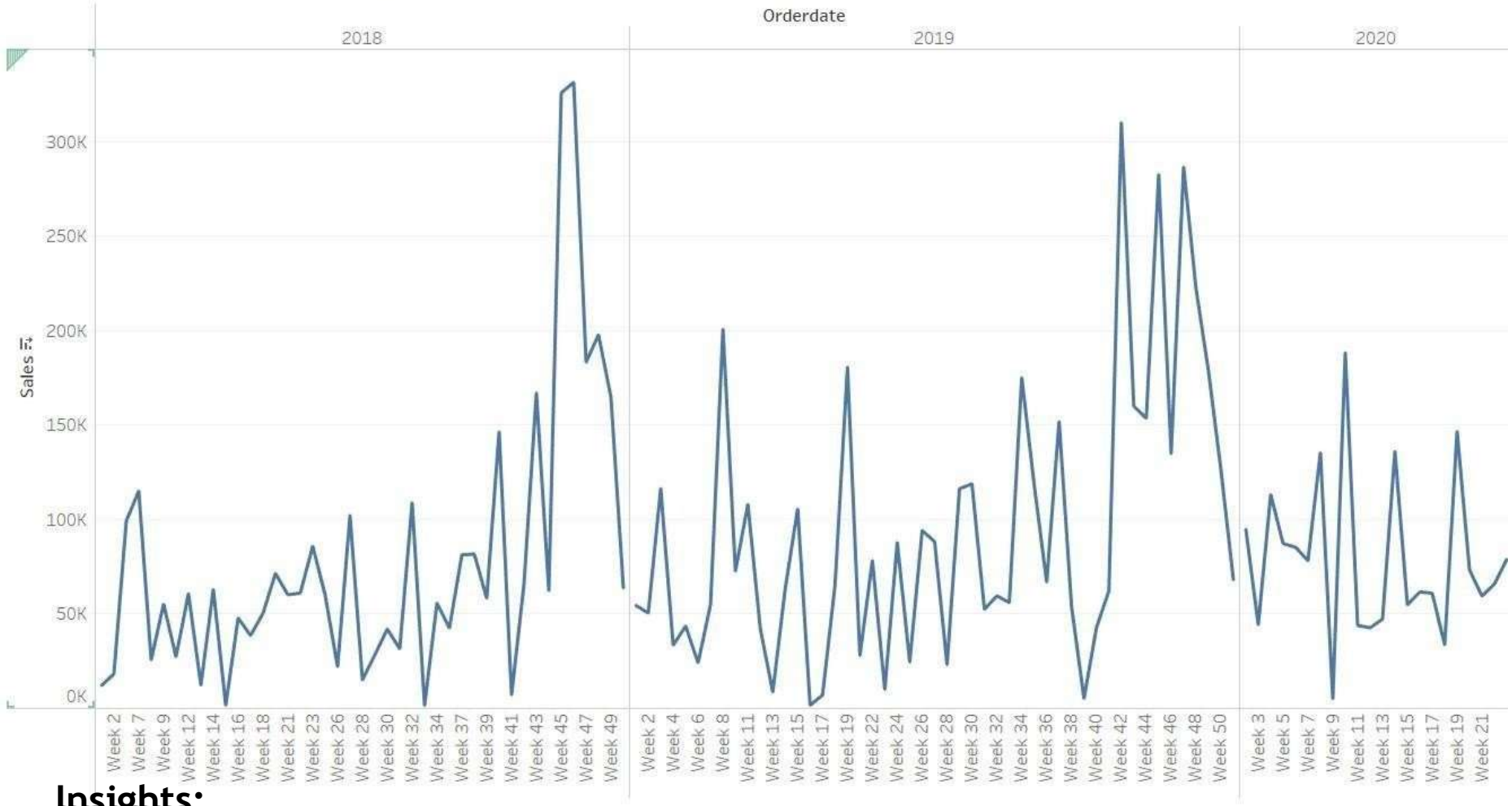
Multivariate Analysis: Total Sales Trend (Monthly for each year)



Insights:

From the plot above, we can say that, the month of November has significantly more sales compared to any other month in 2018 & 2019. Whereas in 2020, its highest in the month of May.

Multivariate Analysis: Total Sales Trend (Weekly for each year)



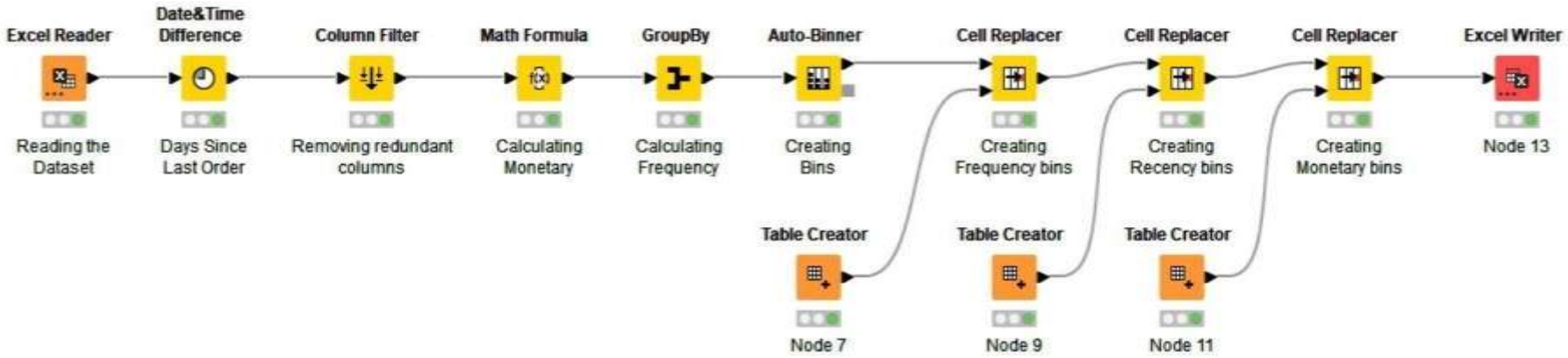
Insights:

From the plot above, we can say that, the last 7 weeks of the year 2018 & 2019 have the highest sales compared to rest of the weeks. Whereas, in the year 2020 Week 9 has the highest sales.

What is RFM?

RFM analysis is a marketing technique used to quantitatively rank and group customers based on the recency, frequency and monetary total of their recent transactions to identify the best customers and perform targeted marketing campaigns. The system assigns each customer numerical scores based on these factors to provide an objective analysis.

KNIME Workflow:



Parameters used / Assumptions made:

- Firstly, after reading the dataset, calculated the Recency metric. Then removed the redundant columns for ease of work.
- Then, created the Monetary metric (Quantity * Cost/piece). Then, created the Frequency metric by keeping Customer name as group by key.
- Created bins for quantile values(0-25, 25-50, 50-75, 75-100). Assigned values to those bins (4 bins created). (1-Very High, 2-High, 3-Medium, 4-Low) for Monetary & Frequency. (4-Very High, 3-High, 2-Medium, 1-Low) for Recency.

Pivot Table of customer count

		Monetary Bin / Frequency Bin										
		High			Low		Medium			Very High		
Recency Bin	Frequency Bin	High	Medium	Very High	Low	Medium	High	Low	Medium	High	Very High
High	High	142					100			21	
	Low				182						
	Medium					32			59		
	Very High			39							17
Low	High	94					25			23	
	Low				734			36			
	Medium					31			88		
	Very High			39							20
Medium	High	49					52				
	Low				264			35			
	Medium		29						122		
	Very High										68
Very High	High	75					26				
	Low				41						
	Medium								87		
	Very High			35							182

Customer Segmentation:

From the bins created before customers were segregated for combinations as listed below.

- Best Customers (Very High Recency, Frequency & Monetary)
- Loyal Customers (High & Medium Recency, Frequency & Monetary)
- On the verge of Churning Customers (Medium & Low Recency, Frequency & Monetary)
- Lost Customers (Low Recency, Frequency & Monetary)

Best Customers

Customername	Recency ..	Frequenc..	Monetar..	
Daedalus Designs I..	Very High	Very High	Very High	20
Osaka Souveniers Co.	Very High	Very High	Very High	20
Auto Assoc. & Cie.	Very High	Very High	Very High	18
Clover Collections, C..	Very High	Very High	Very High	16
Iberia Gift Imports, ..	Very High	Very High	Very High	15
Online Mini Collecta..	Very High	Very High	Very High	15
Signal Collectibles L..	Very High	Very High	Very High	15
Bavarian Collectable..	Very High	Very High	Very High	14
CAF Imports	Very High	Very High	Very High	13
West Coast Collecta..	Very High	Very High	Very High	13
Double Decker Gift S..	Very High	Very High	Very High	12
Cambridge Collectab..	Very High	Very High	Very High	11

Insights:

Customer Segmentation:

The Companies Mentioned in the image along, are the best customers as they have Very High bin of all 3 Recency, Frequency & Monetary.

Namely:

1) Daedalus Designs Imports
2) Osaka Souveniers Co.
3) Auto Assoc. & Cie.
4) Clover Collections Co.
5) Iberia Gift Imports, Corp.

Loyal Customers

Customername	Recency ..	Frequenc..	Monetar..	
Baane Mini Imports	High	Medium	Medium	32
Oulu Toy Supplies, I..	Medium	Medium	Medium	32
Suominen Souveniers	Medium	Medium	Medium	30
Toys of Finland, Co.	Medium	Medium	Medium	30
Toys4GrownUps.com	Medium	Medium	Medium	30
Signal Gift Stores	Medium	Medium	High	29
Heintze Collectables	High	Medium	Medium	27
Cruz & Sons Co.	High	High	Medium	26
FunGiftIdeas.com	Medium	High	Medium	26
giftsbymail.co.uk	High	High	High	26
Mini Classics	High	High	High	26
Stylish Desk Decors,..	Medium	High	Medium	26
Toms Spezialitten, L..	High	High	Medium	26
Marseille Mini Autos	Medium	High	High	25
Vitachrome Inc.	High	High	Medium	25
Collectables For Les..	Medium	High	High	24
Enaco Distributors	High	High	High	23
La Corne D'abondan..	High	High	Medium	23
Motor Mint Distribu..	High	High	High	23
Blauer See Auto, Co.	High	High	High	22
Canadian Gift Excha..	High	High	High	22

Insights:

The Companies Mentioned in the image along, are the loyal customers as they have a combination of High & Medium bins of all 3 Recency, Frequency & Monetary.

Namely:

1) Baane Mini Imports
2) Oulu Toy Supplies
3) Suominen Souveniers
4) Toys of Finland
5) Toys4GrownUps.com

Churn Customers

Customername	Recency ..	Frequenc..	Monetar..	
Euro Shopping Chan..	Low	Low	Low	259
Mini Gifts Distributo..	Low	Low	Low	180
Australian Collector..	Medium	Low	Low	55
Muscle Machine Inc	Medium	Low	Low	48
La Rochelle Gifts	Low	Low	Low	53
Dragon Souveniers, ..	Medium	Low	Low	43
The Sharp Gifts War..	Low	Low	Low	40
Anna's Decorations, ..	Medium	Low	Low	46
Souveniers And Thin..	Low	Low	Low	46
Salzburg Collectables	Low	Low	Low	40
Danish Wholesale I..	Low	Low	Low	36
L'ordine Souveniers	Low	Low	Low	39
Reims Collectables	Low	Low	Low	41
Scandinavian Gift Id..	Medium	Low	Low	38
Diecast Classics Inc.	Low	Medium	Low	31
Technics Stores Inc.	Medium	Low	Low	34
Tokyo Collectables, ..	Low	Medium	Medium	32
UK Collectables, Ltd.	Low	Medium	Medium	29
Handji Gifts& Co	Low	Low	Medium	36
Suominen Souveniers	Medium	Medium	Medium	30
Toys of Finland, Co.	Medium	Medium	Medium	30
Mini Creations Ltd.	Medium	Low	Medium	35
Toys4GrownUps.com	Medium	Medium	Medium	30
Oulu Toy Supplies, I..	Medium	Medium	Medium	32
Auto Canal Petit	Low	Medium	Medium	27

Insights:

The Companies Mentioned in the image above, are on the verge of churning customers as they have a combination of Low & Medium bins of all 3 Recency, Frequency & Monetary.

Namely:

1) Dragon Souveniers
2) Anna's Decorations.
3) Australian Collector Co.
4) Muscle Machine Inc.
5) UK Collectables Ltd.

Lost Customers

Customername	Recency ..	Frequenc..	Monetar..	
Euro Shopping Chan..	Low	Low	Low	259
Mini Gifts Distributo..	Low	Low	Low	180
La Rochelle Gifts	Low	Low	Low	53
The Sharp Gifts War..	Low	Low	Low	40
Souveniers And Thin..	Low	Low	Low	46
Salzburg Collectables	Low	Low	Low	40
Danish Wholesale I..	Low	Low	Low	36
L'ordine Souveniers	Low	Low	Low	39
Reims Collectables	Low	Low	Low	41

Insights:

The Companies Mentioned in the image above, are the lost customers as they have Low bin of all 3 Recency, Frequency & Monetary.

Namely:

1) Euro Shopping Channel
2) Mini Gifts Distribution Ltd.
3) Reims Collectibles
4) L'ordine Souveniers
5) La Rochelle Gifts

Business Recommendations:

- As we have seen, the company should focus on Best & Loyal Customers as they are more likely to purchase from the company again.
- We should also understand the average market sales of our leading competitors' sales to understand our performance in the market much better.
- We should devise new dynamic pricing system to determine best prices for products which would maintain balance between company's profits & customer satisfaction. This would lead to increase of company's goodwill amongst the customers.
- There should be some incentive given to loyal customers as well customers who have very low purchases over the course of time. This will help retain our existing customers against our competitors.
- Company should not invest much more time & money on low customers as they returns might no be that substantial.

THE END