

# MRA Project Milestone 1

Presented By: Tejas Padekar  
Batch: PGP DSBA – G1  
Email: [tnpadekar@gmail.com](mailto:tnpadekar@gmail.com)

---

# AGENDA:

- About data (Continuous and Categorical)
  - EDA
  - Univariate, Bivariate and Multivariate Analysis
  - Inferences
  - RFM Analysis
  - Top Customers across different segments
-

# 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.
- Below is the data description:

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

# ABOUT DATA

- There are 2747 rows and 20 columns.
- There are 1 datetime, 2 float, 5 Integer and 12 object data types.
- There are no null values in the dataset.
- There are no duplicate rows in the dataset.

```
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)
```

# ABOUT DATA (CONTINUOUS VARIABLES)

- Average of quantity ordered is 35 with min 6 and max 97.
- Average for price each is 101 with min 26 and max 252
- Average for Sales is 3553 with min 482 and max 14082.
- There seems to be a few outliers present in the dataset.

	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

# ABOUT DATA (CATEGORICAL VARIABLES)

- There are 6 unique Sales Statuses.
- There are 7 different product lines with 109 product codes.
- There are total of 89 customers across 19 countries.
- Sale Size is divided into 3 deal size categories.

	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/ Moralarzal, 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

# EXPLORATORY DATA ANALYSIS

We have used Python and Tableau for performing EDA

---

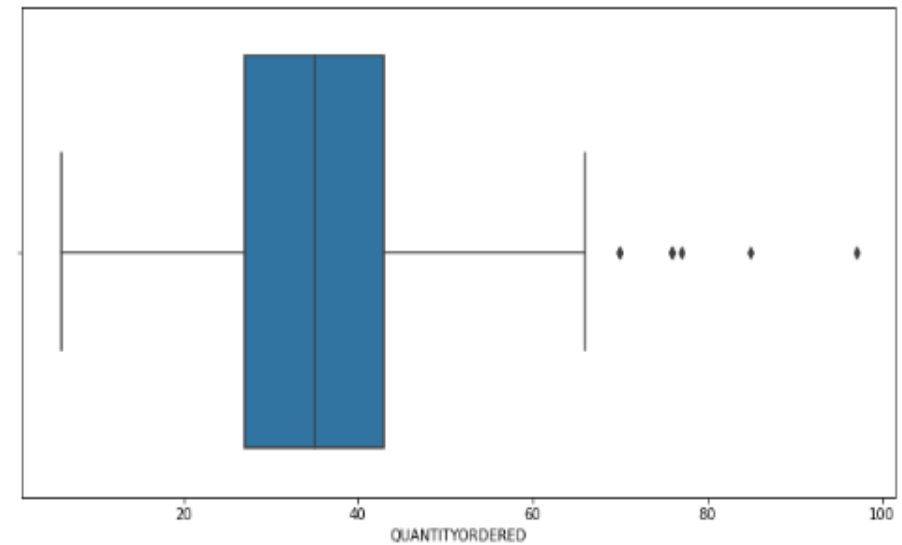
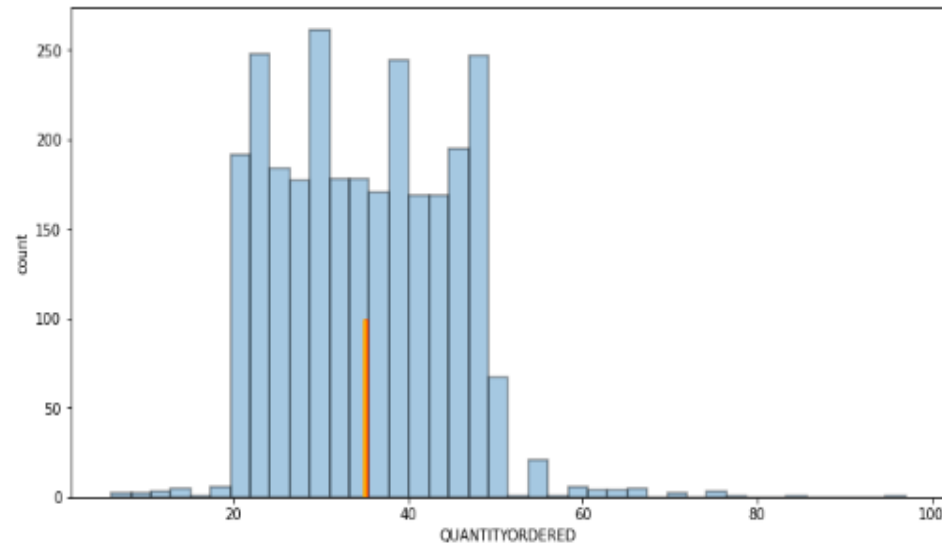
# UNIVARIATE ANALYSIS

## DATA DISTRIBUTION & OUTLIERS

- There are outliers and high skewness in QUANTITYORDERED.

QUANTITYORDERED

Skew: 0.37

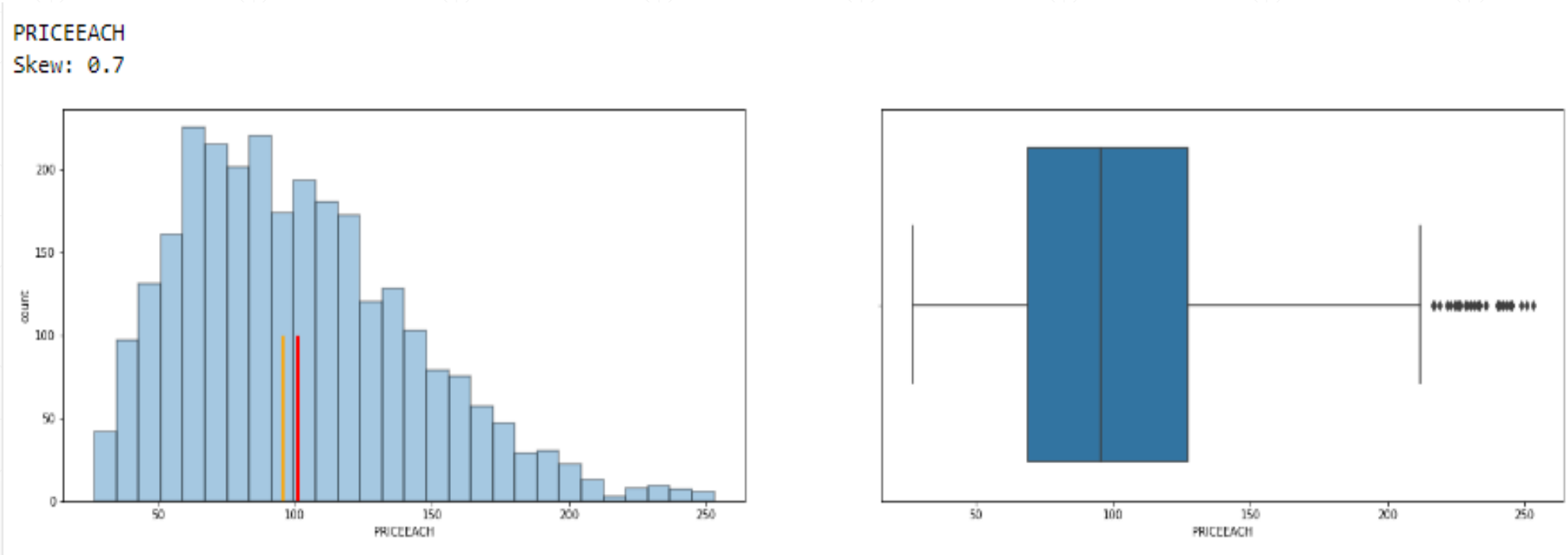




# UNIVARIATE ANALYSIS

## DATA DISTRIBUTION & OUTLIERS

- There are outliers and high skewness in PRICEEACH.



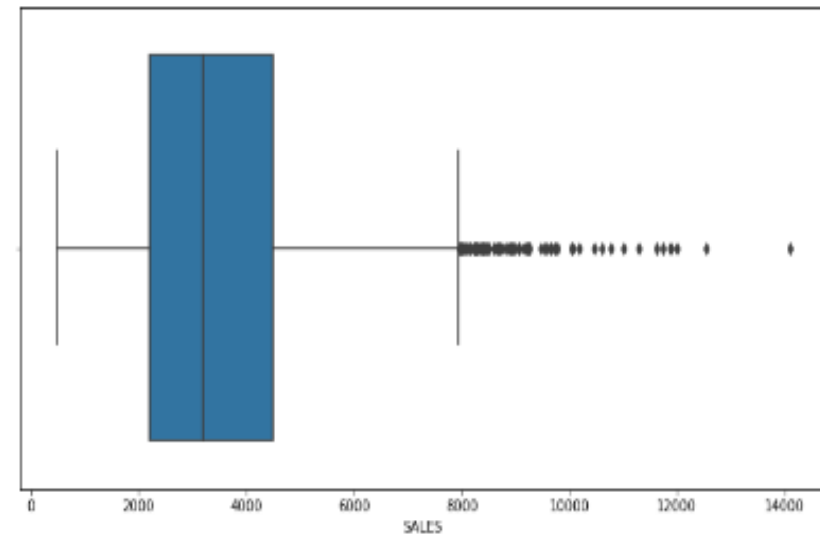
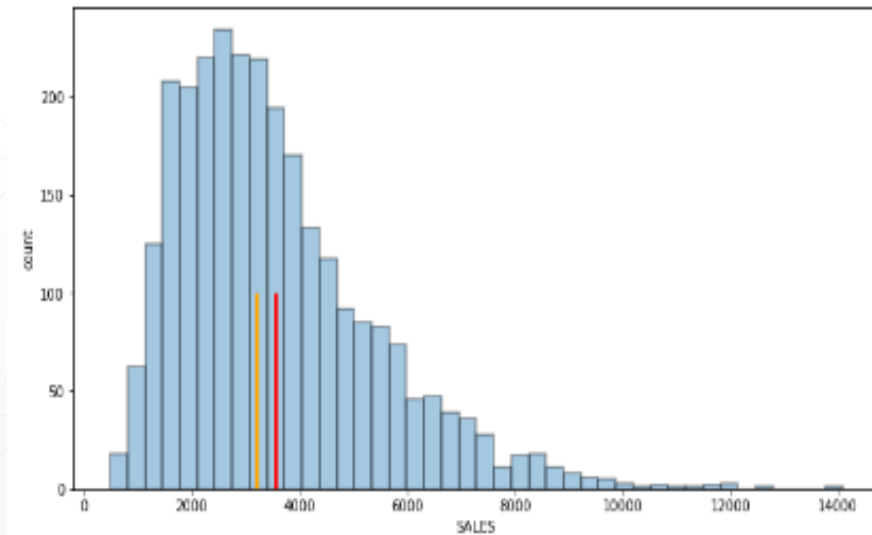
# UNIVARIATE ANALYSIS

## DATA DISTRIBUTION & OUTLIERS

- There are outliers and high skewness in SALES.

SALES

Skew: 1.16



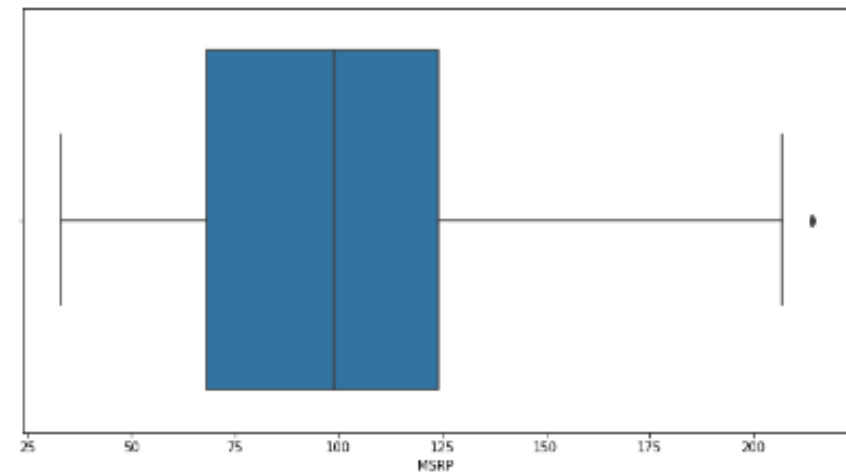
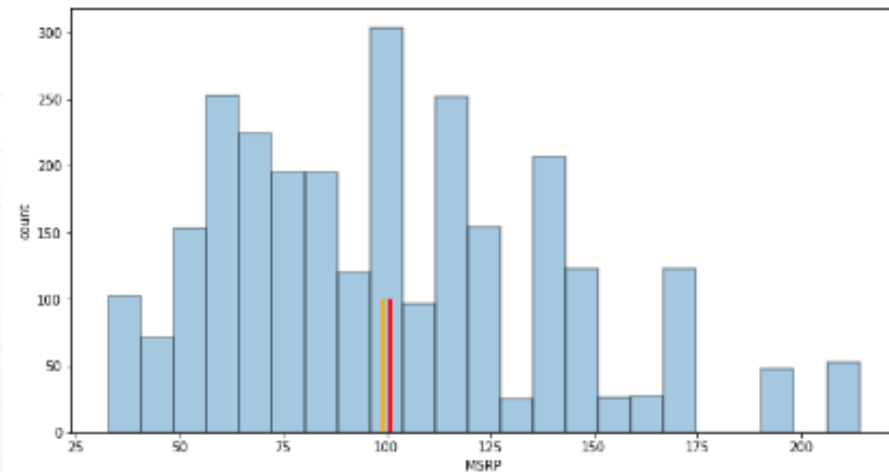
# UNIVARIATE ANALYSIS

## DATA DISTRIBUTION & OUTLIERS

- There are outliers and high skewness in MSRP.

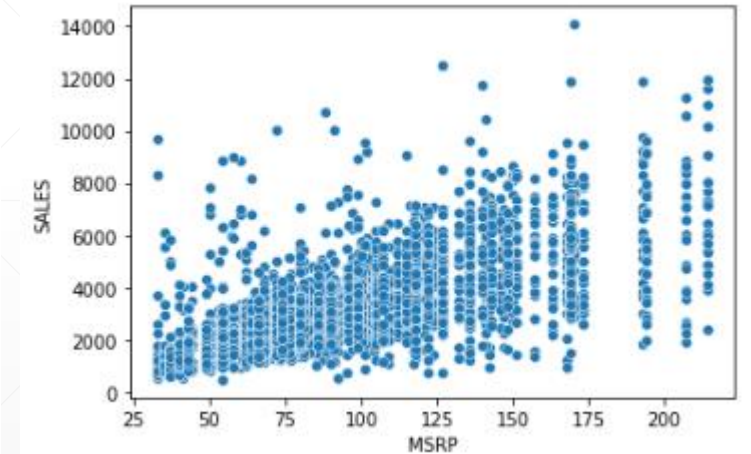
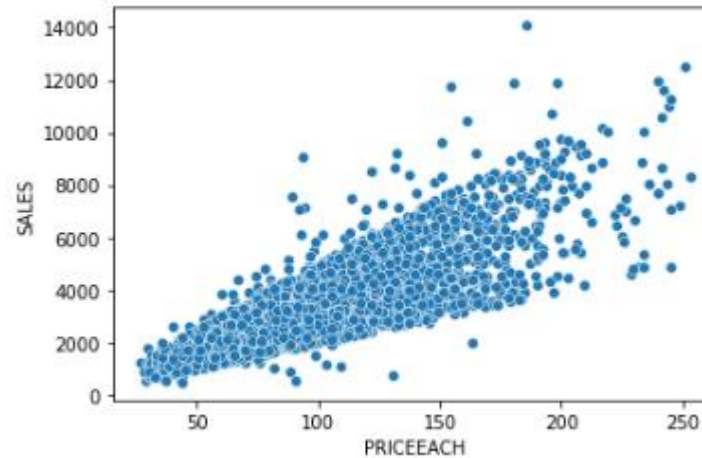
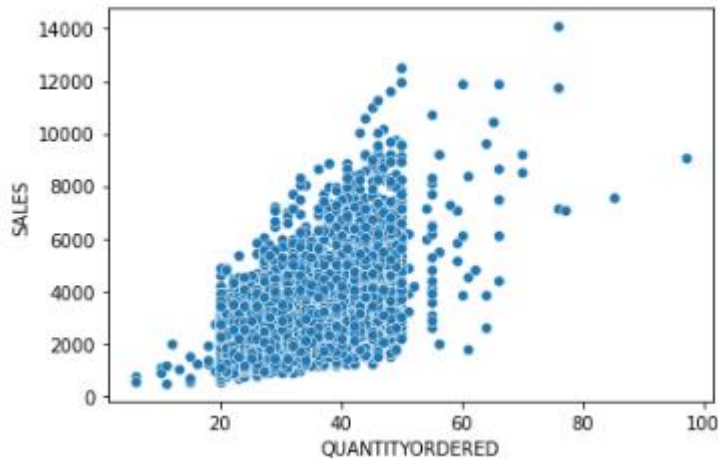
MSRP

Skew: 0.58



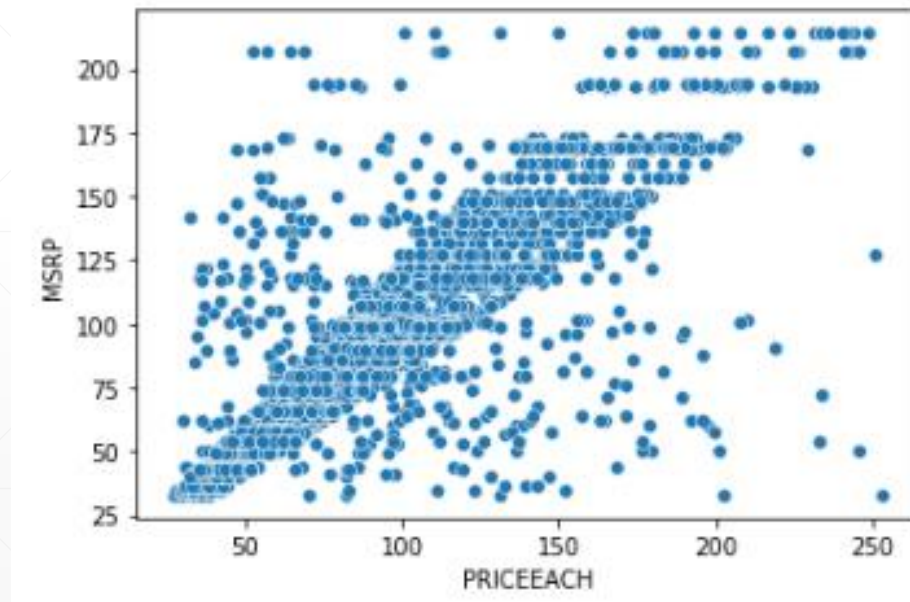
# BIVARIATE ANALYSIS

- We can observe below that SALES has high correlation with QUANTITYORDERED, PRICEEACH and MSRP as well.



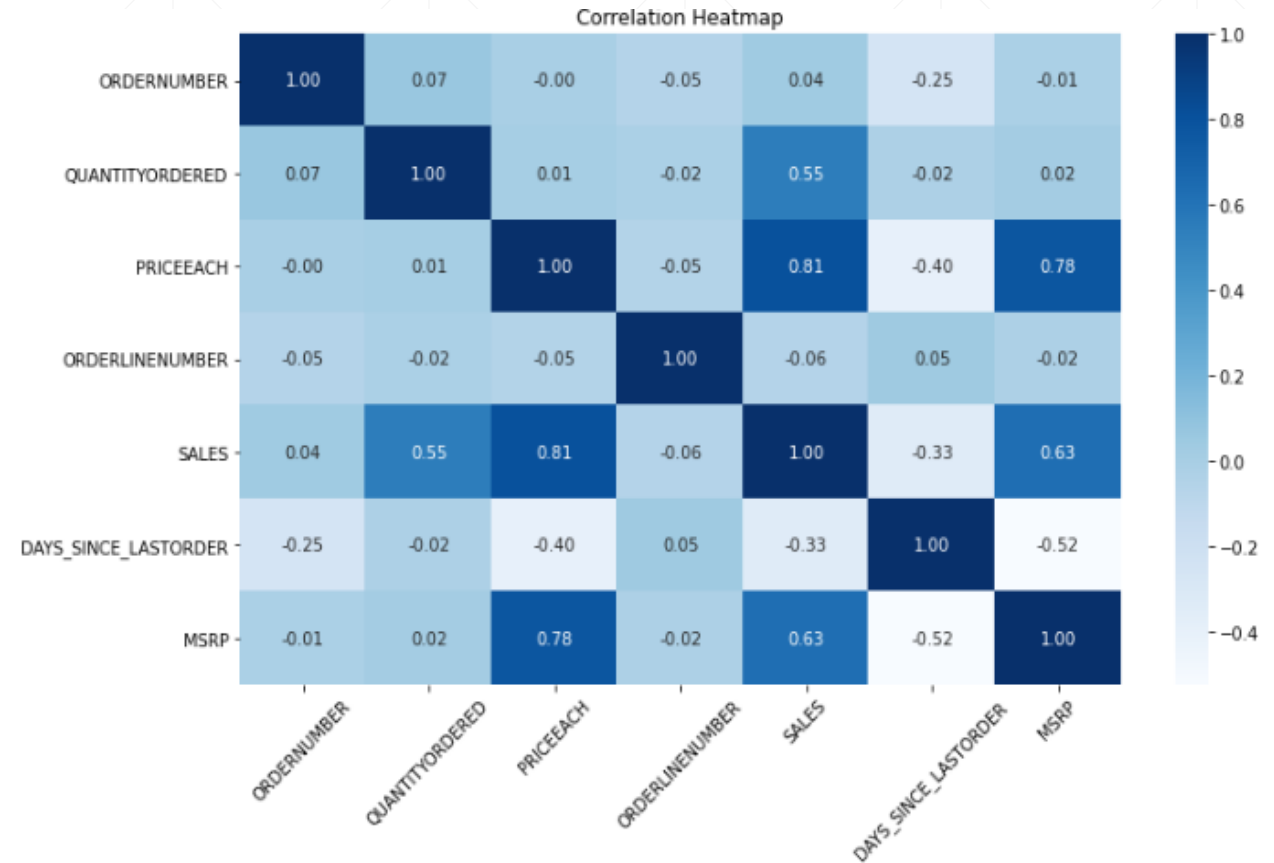
# BIVARIATE ANALYSIS

- We can observe below that MSRP has high correlation with PRICEEACH.

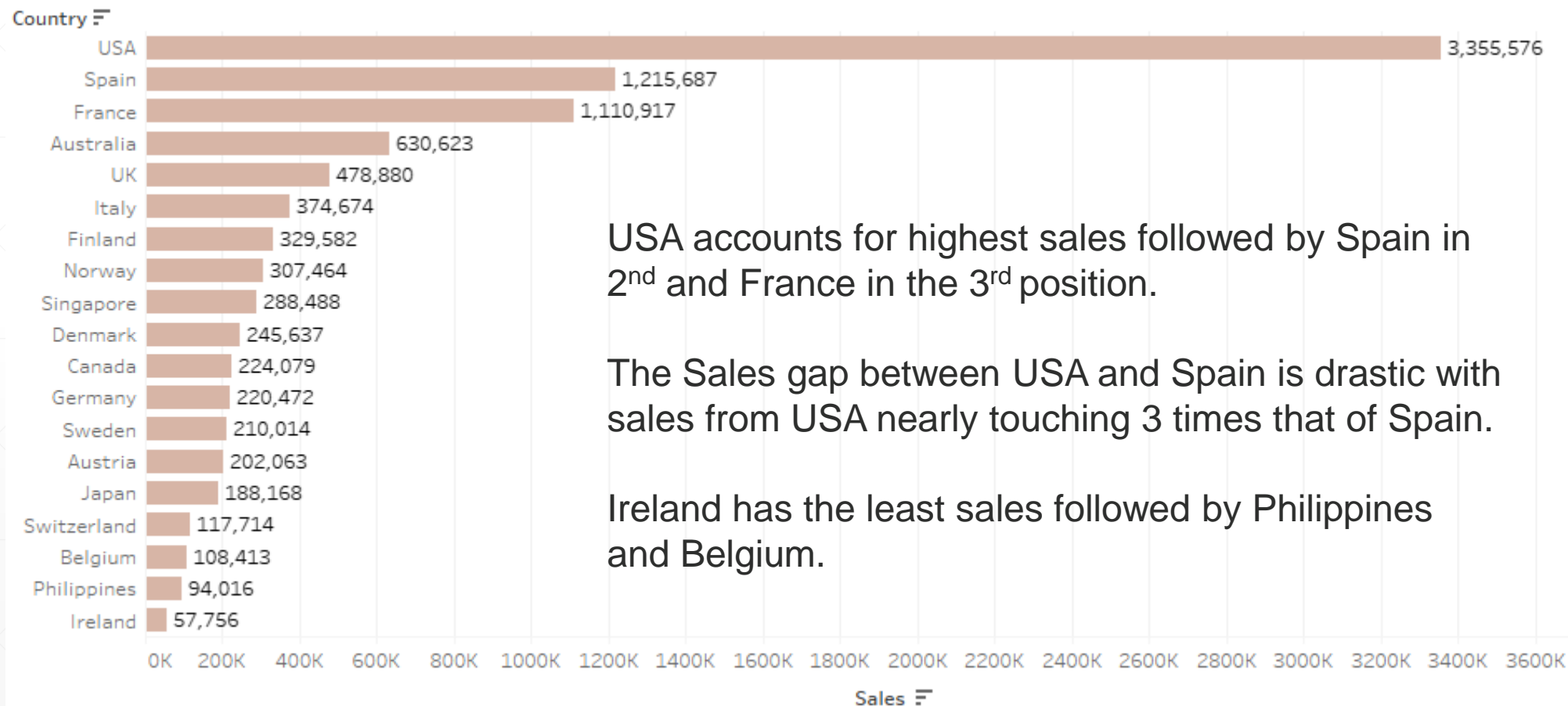


# MULTIVARIATE ANALYSIS

- As stated earlier, we can observe strong correlation between SALES and QUANTITYORDERED, PRICEEACH and MSRP
- PRICEEACH and MSRP also have a high correlation
- There is nearly no correlation between QUANTITYORDERED and PRICEEACH.



# TOTAL SALES ACROSS ALL COUNTRIES



# TOTAL SALES ACROSS CITIES

City	Country		
Madrid	Spain	1,082,551	Madrid leads the cities with the highest sales followed by San Rafael and NYC
San Rafael	USA	654,858	
NYC	USA	560,788	
Singapore	Singapore	288,488	5 out of the top 10 cities belong to USA which contribute to maximum sales.
Paris	France	268,945	
New Bedford	USA	207,875	
Nantes	France	204,305	
Melbourne	Australia	200,995	
Brickhaven	USA	165,255	
San Jose	USA	160,010	



# TOP 10 CUSTOMERS BY SALES

- Below we see that the top 10 customers with highest sales are based across multiple countries with USA having maximum with 4 followed by Australia with 2.
- Spain, France, Singapore, UK each comprise of 1.
- Euro Shopping Channel from Spain is the top customer by overall sales.

Customername	Country	City	
Euro Shopping Channel	Spain	Madrid	912,294
Mini Gifts Distributors Ltd.	USA	San Rafael	654,858
Australian Collectors, Co.	Australia	Melbourne	200,995
Muscle Machine Inc	USA	NYC	197,737
La Rochelle Gifts	France	Nantes	180,125
Dragon Souvenirs, Ltd.	Singapore	Singapore	172,990
Land of Toys Inc.	USA	NYC	164,069
The Sharp Gifts Warehouse	USA	San Jose	160,010
AV Stores, Co.	UK	Manchester	157,808
Anna's Decorations, Ltd	Australia	North Sydney	153,996

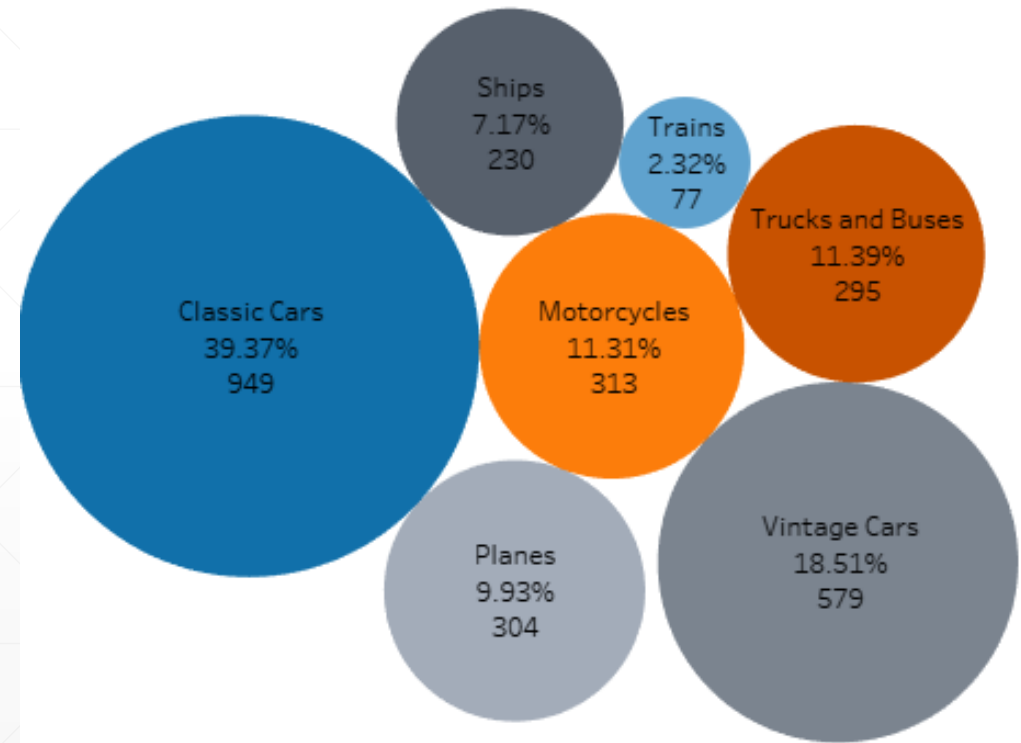
# TOP 10 CUSTOMERS BY ORDERS

- Top 10 customers by orders are also across multiple countries with USA having maximum with 3. Rest of the other countries have only client with high orders.
- The top 2 customers, Euro Shopping Channel with 26 and Mini Gifts Distribution Ltd with 17 are the only clients with double digit orders.
- Euro Shopping Channel from Spain is again the top customer by overall orders placed.

Customername	Country	City	
Euro Shopping Channel	Spain	Madrid	26
Mini Gifts Distributors Ltd.	USA	San Rafael	17
Reims Collectables	France	Reims	5
Dragon Souvenirs, Ltd.	Singapore	Singapore	5
Danish Wholesale Imports	Denmark	Kobenhavn	5
Australian Collectors, Co.	Australia	Melbourne	5
Volvo Model Replicas, Co	Sweden	Lule	4
Tokyo Collectables, Ltd	Japan	Minato-ku	4
The Sharp Gifts Warehouse	USA	San Jose	4
Technics Stores Inc.	USA	Burlingame	4

# TOTAL SALES & ORDERS BY PRODUCTLINE

- Classic Cars has the lions share of overall sales with 39.37 % followed by Vintage Cars at 18.51 % and Trucks and Buses at 11.39 %
- Least sales are seen in Trains at 2.32% followed by Ships at 7.17%
- So far 949 units have been sold for Classic Cars followed by 579 for Vintage Cars and 313 for Motorcycles
- Train are the least ordered at 77 followed by ships at 230 units

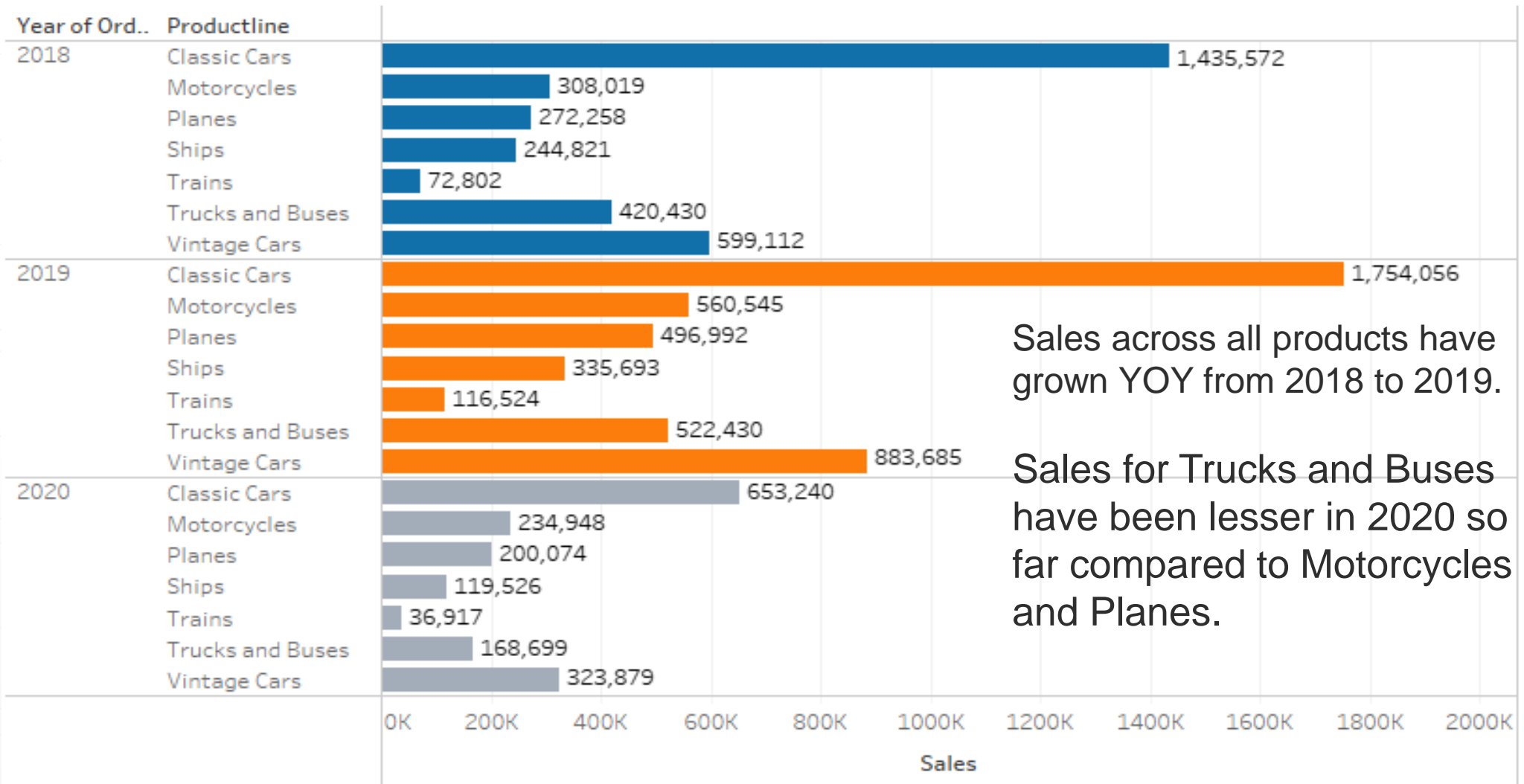


# CURRENT STATUS OF PRODUCTLINE

- Below is the current status as per each of the seven product lines.
- 2541 products are shipped, 47 are ready to be shipped, 41 are in process to be delivered.
- However, 44 are on hold, 14 are under dispute and 60 have been cancelled.

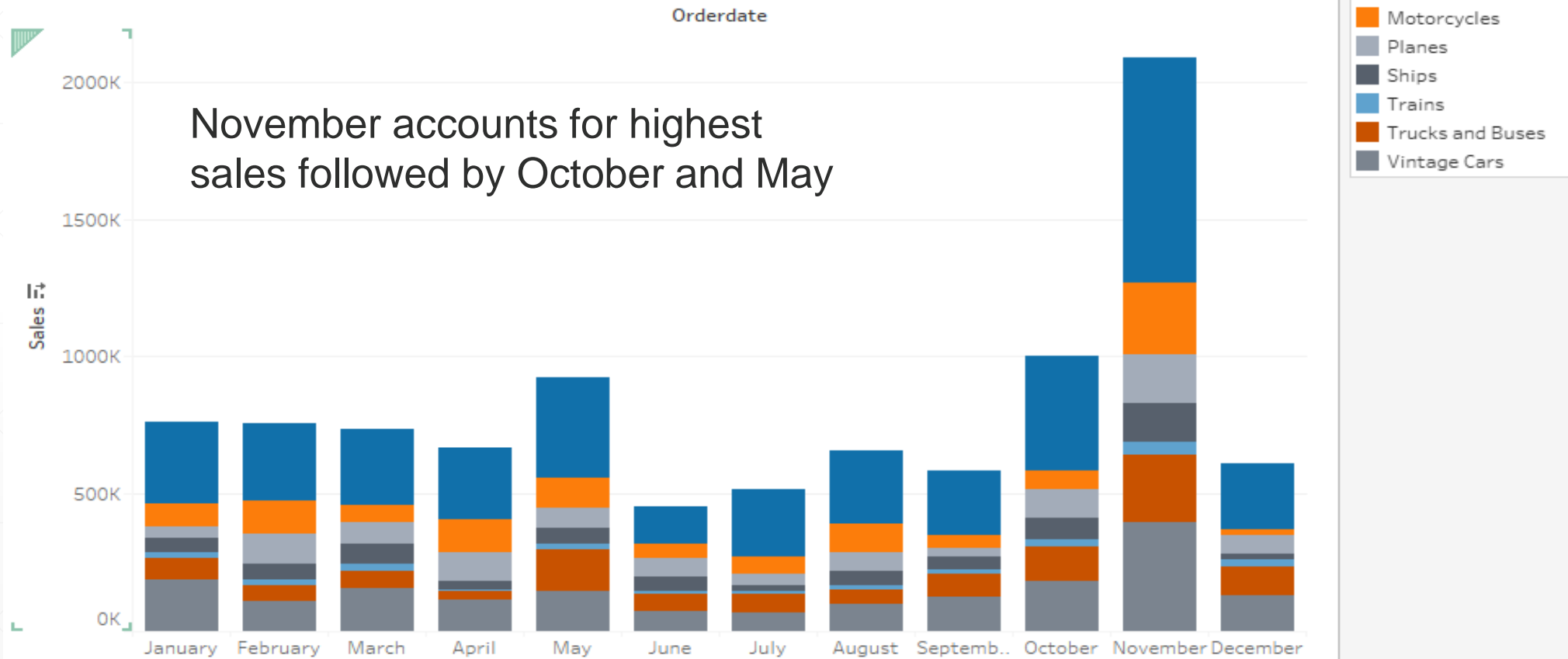
Productline	Status						Grand To..
	Cancelled	Disputed	In Process	On Hold	Resolved	Shipped	
Classic Cars	16	3	14	12	8	896	949
Vintage Cars	13	2	16	9	10	529	579
Motorcycles		6		1		306	313
Planes	12	2		9	12	269	304
Trucks and Buses			11	4	5	275	295
Ships	18	1		8	12	191	230
Trains	1			1		75	77
Grand Total	60	14	41	44	47	2,541	2,747

# YEARLY SALES BY PRODUCTLINE



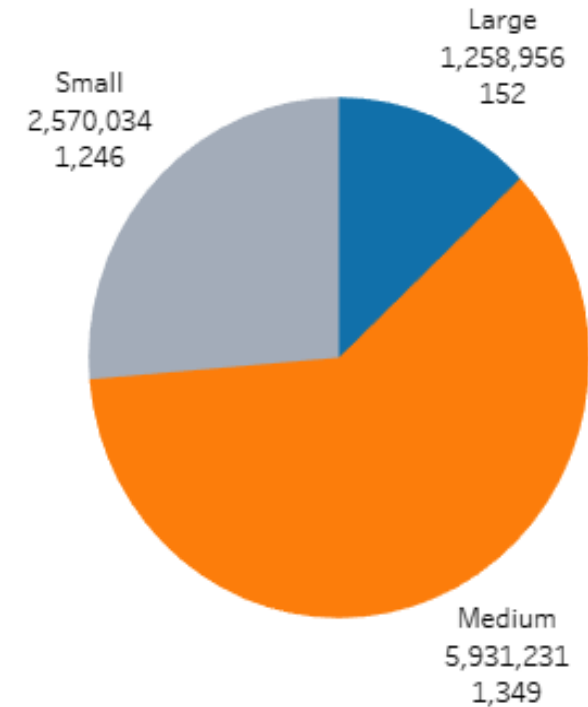
# MONTHWISE SALES FOR PRODUCTINE

Monthly Sales for Productline

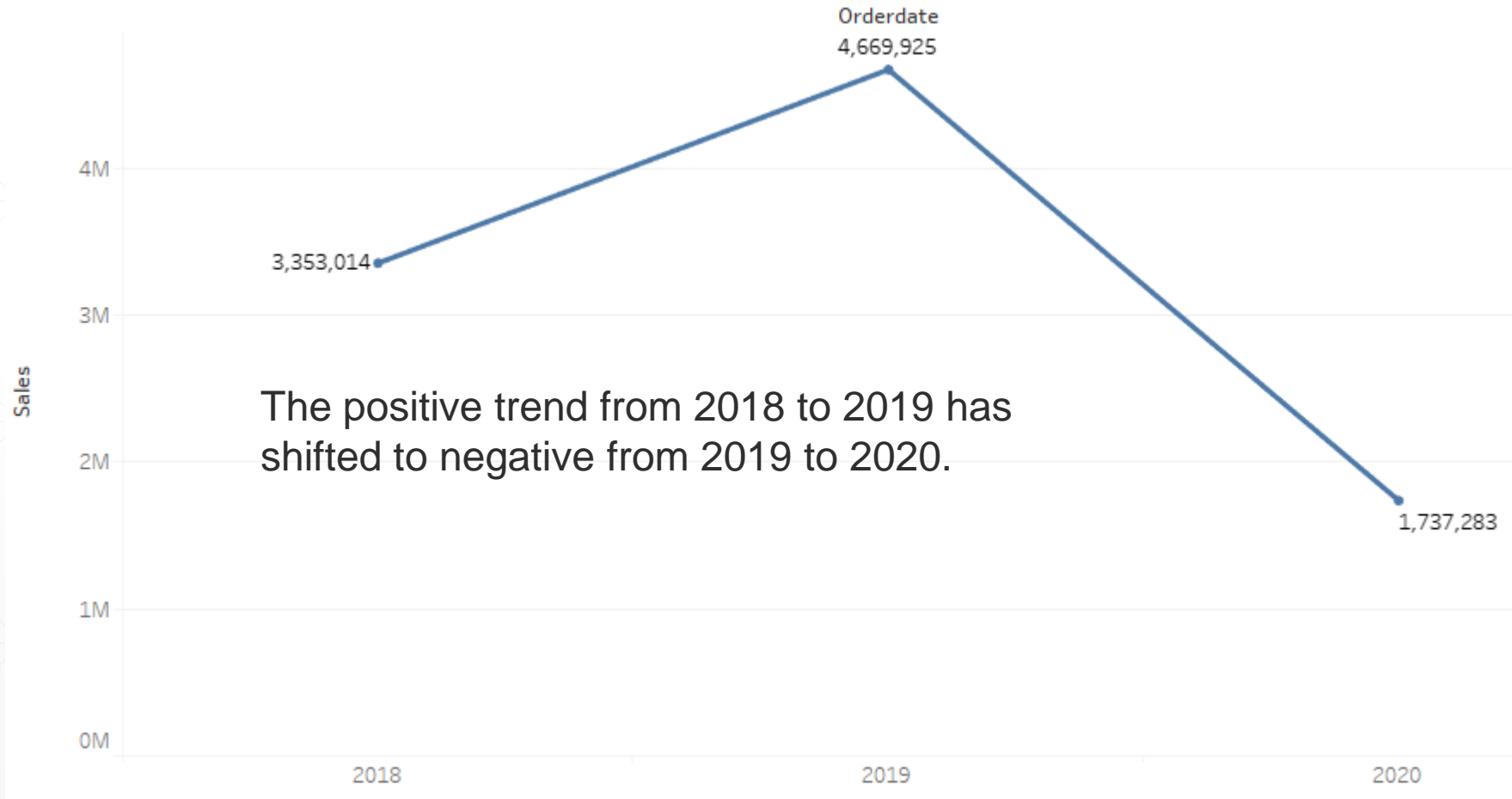


# DEALSIZE BY SALES AND TOTAL CUSTOMERS

- Medium deal size generates the highest revenue followed by Small.
- Large deal sizes are very few in no of orders and revenue both
- Medium (1349) and Small (1246) have very similar no of orders. However, the difference in revenue is more than double for medium comparatively.

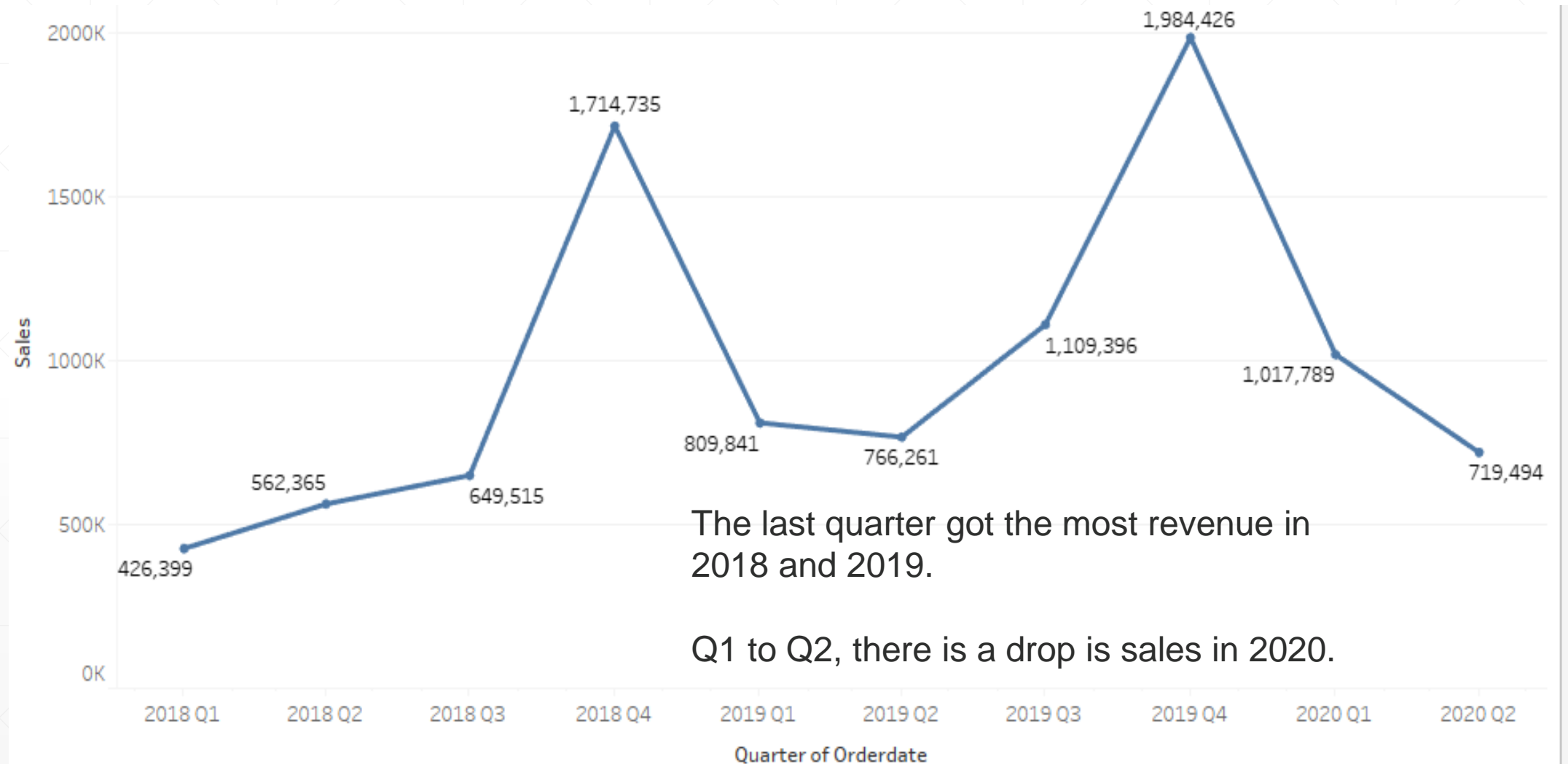


# YEARLY SALES TREND

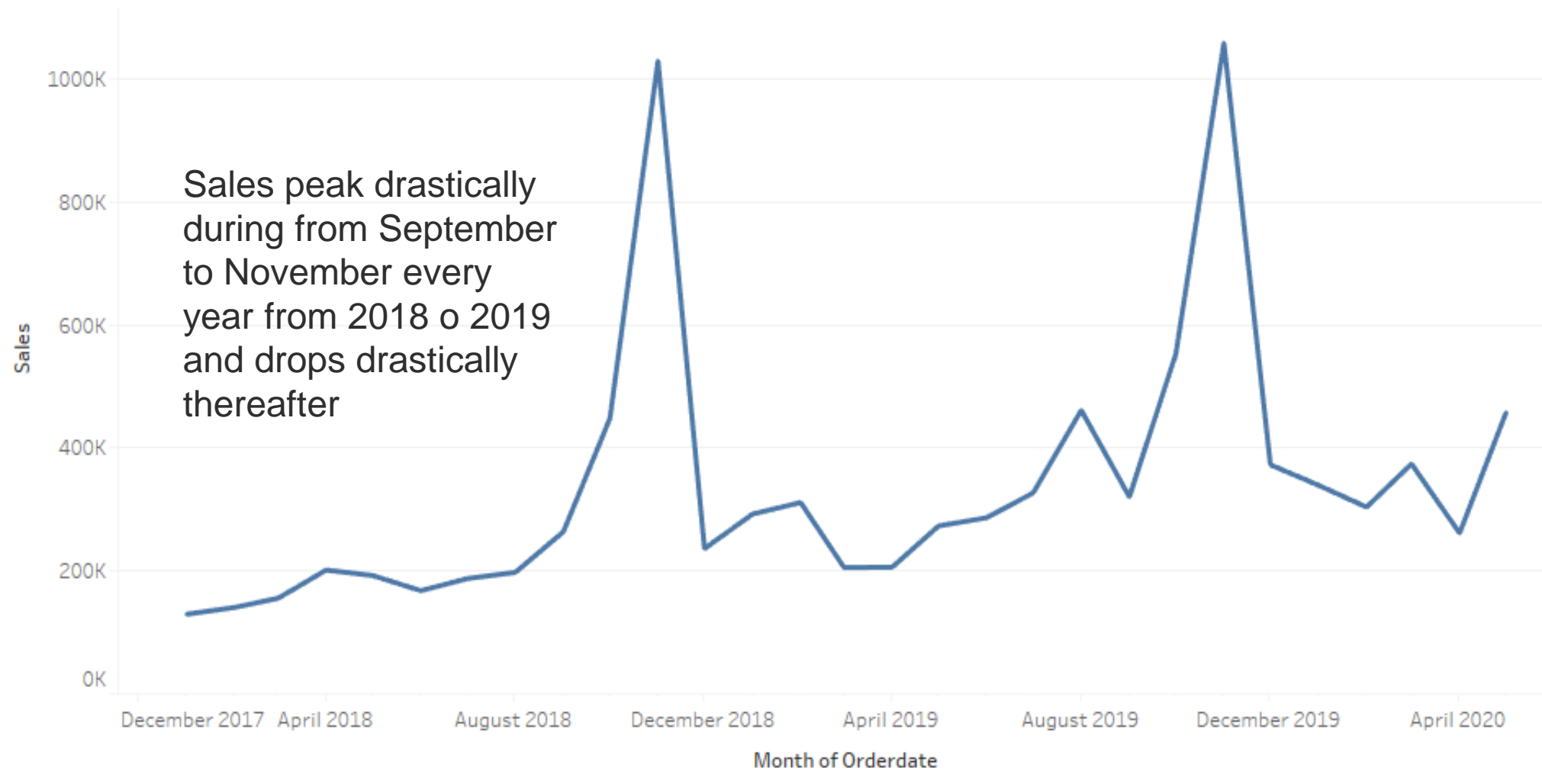




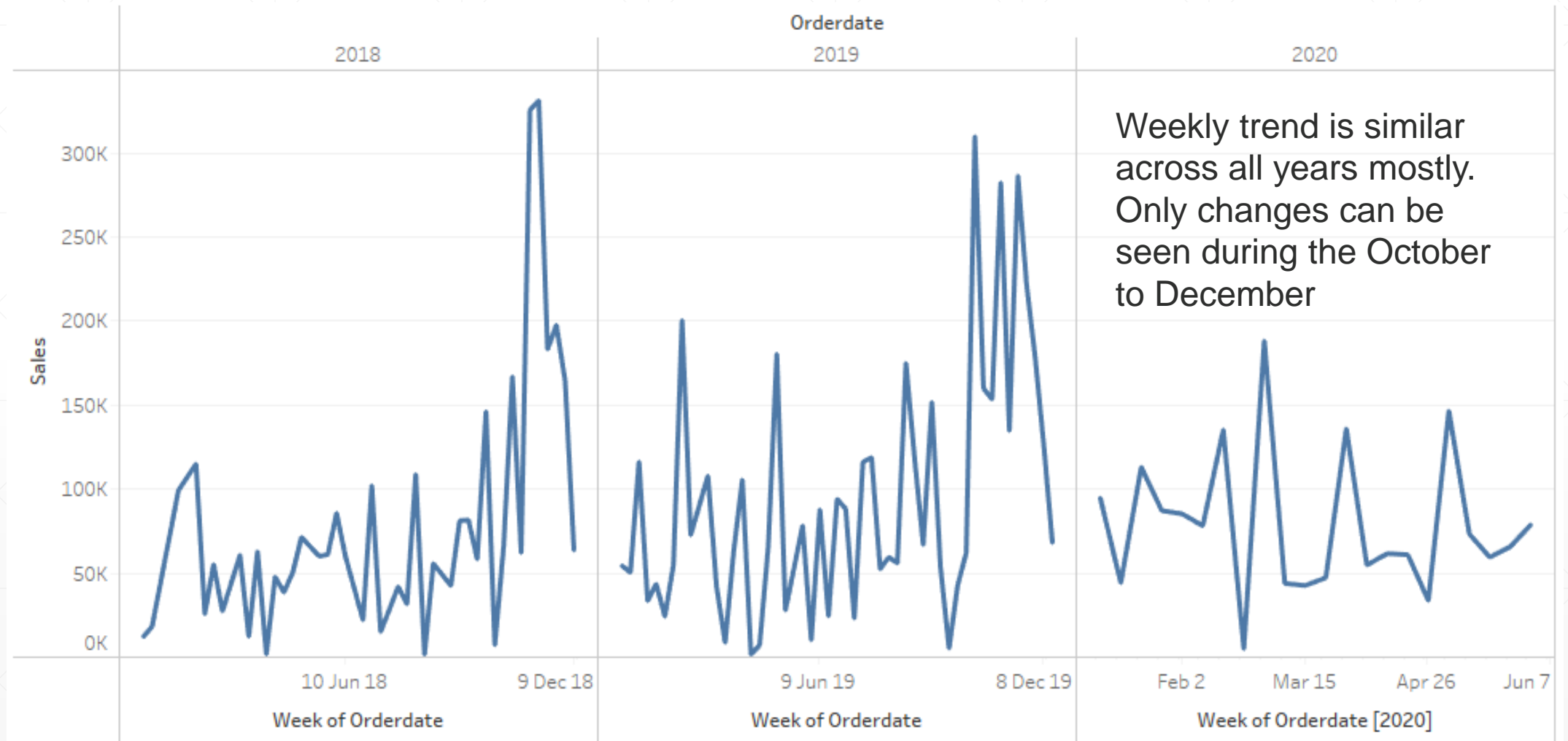
# QUARTERLY SALES TREND



# MONTHLY SALES TREND



# WEEKLY SALES TREND



# INFERENCES:

- Average of quantity ordered is 35 with min 6 and max 97.
  - Average for price each is 101 with min 26 and max 252
  - Average for Sales is 3553 with min 482 and max 14082.
  - Only the top 2 customers account for double digit orders.
  - Majority of the deals are medium sized.
  - Medium (1349) and Small (1246) have very similar no of orders. However, the difference in revenue is more than double for medium comparatively.
  - Classic Cars has the lions share of overall sales with 39.37 % followed by Vintage Cars at 18.51 % and Trucks and Buses at 11.39 %
  - Out of 2747, 2541 products are shipped, 47 are ready to be shipped, 41 are in process to be delivered. However, 44 are on hold, 14 are under dispute and 60 have been cancelled.
  - Top 10 customers by orders are also across multiple countries with USA having maximum with 3. Rest of the other countries have only client with high orders.
-

# INFERENCES:

- USA accounts for highest sales followed by Spain in 2<sup>nd</sup> and France in the 3<sup>rd</sup> position. The Sales gap between USA and Spain is drastic with sales from USA nearly touching 3 times that of Spain. Hence, efforts can be made to increase sales in Spain and France and the other top 10 countries.
  - Ireland has the least sales followed by Philippines and Belgium. We can try to cut operational costs in these countries and use it to try and increase customer base.
  - 5 out of the top 10 cities belong to USA which contribute to maximum sales. So, there is a need to try and increase sales and market share in other regions as well.
  - Trains have the least impact on revenue. So, can think of discontinuing these products or offer them at lower discounted prices
  - Sales from September to November are at the peak. Hence, need to focus more on marketing and advertising for products month or quarter before to generate interest for new customer acquisitions.
  - Promotional offers can be run during the periods of fall in sales, for eg; February, June and December especially.
-

# RFM ANALYSIS

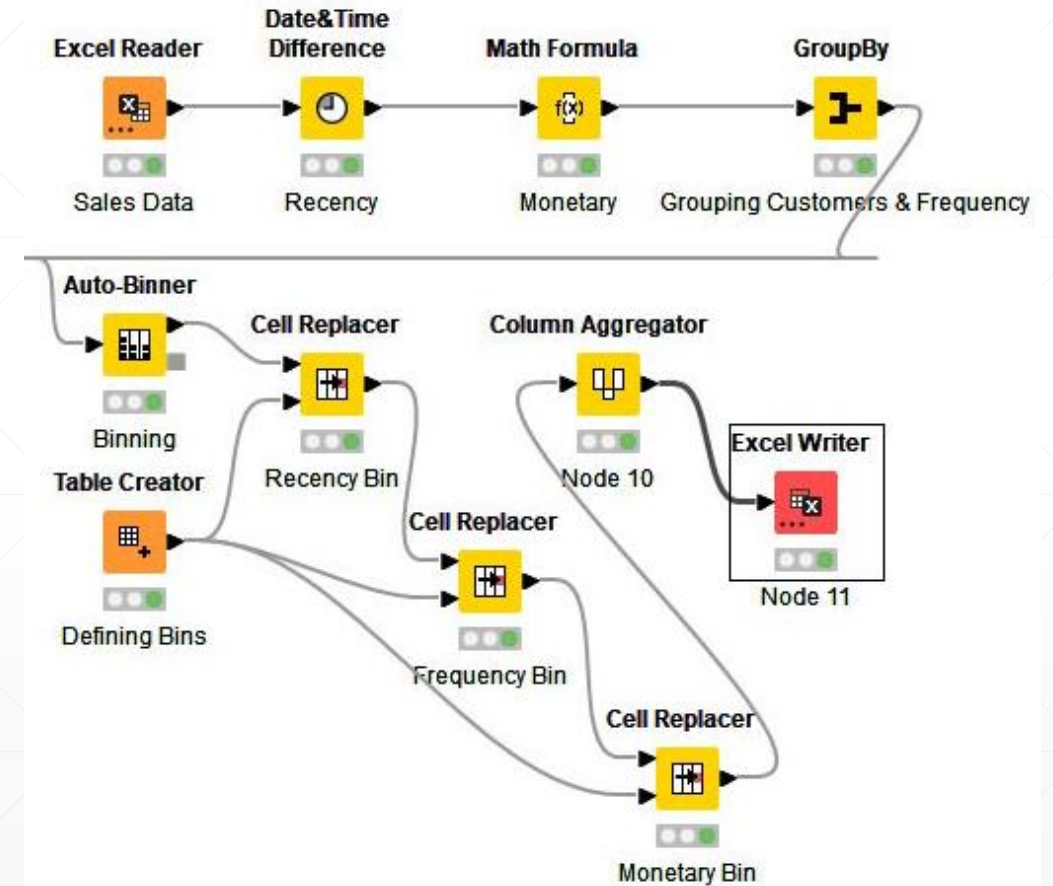


# What is RFM?

- RFM analysis is customer segmentation technique used to measure a worthiness of customer based on three measures as below.
  - 1. Recency (How recently has the customer made a purchase)
  - 2. Frequency (How frequently does the customer purchase)
  - 3. Monetary (What is the revenue that we generate from the customer)
-

# KNIME RFM Analysis

- Data is Grouped by CUSTOMERNAME
- Customers are divided into 4 segments
- We have made the following assumptions:
  - A) PRICEEACH x QUANTITYORDERED for Monetary
  - B) Unique Count of ORDERNUMBER for Frequency
  - C) 31/05/2020 – ORDERDATE for Recency





## Top 5 Best Customer

- These are customers having an RFM Score of 444 which is the best/highest.

	CUSTOMERNAME	QUANTITYORDERED	SALES	ORDERNUMBER	Recency	Monetary	Recency_Score	Frequency_Score	Monetary_Score	RFM
32	Euro Shopping Channel	9327	3522.371081	26	0	912294.11	4	4	4	444
53	Mini Gifts Distributors Ltd.	6366	3638.100333	17	2	654858.06	4	4	4	444
43	La Rochelle Gifts	1832	3398.583019	4	0	180124.90	4	4	4	444
78	The Sharp Gifts Warehouse	1656	4000.256750	4	39	160010.27	4	4	4	444
72	Souvenirs And Things Co.	1601	3295.021304	4	2	151570.98	4	4	4	444

## Top 5 Customers on verge of churning

	CUSTOMERNAME	QUANTITYORDERED	SALES	ORDERNUMBER	Recency	Monetary	Recency_Score	Frequency_Score	Monetary_Score	RFM
48	Microscale Inc.	381	3314.493000	2	210	33144.93	2	1	1	211
18	Classic Gift Ideas, Inc	668	3214.617619	2	230	67506.97	2	1	1	211
88	giftsbymail.co.uk	895	3009.263077	2	212	78240.84	2	1	2	212
51	Mini Classics	929	3290.615000	2	229	85555.99	2	1	2	212
17	Canadian Gift Exchange Network	703	3419.950909	2	222	75238.92	2	1	2	212

## Top 5 Lost Customer

- These are customers having an RFM Score of 111 which is the least.

	CUSTOMERNAME	QUANTITYORDERED	SALES	ORDERNUMBER	Recency	Monetary	Recency_Score	Frequency_Score	Monetary_Score	RFM
29	Double Decker Gift Stores, Ltd	357	3001.586667	2	495	36019.04	1	1	1	111
87	West Coast Collectables Co.	511	3544.972308	2	488	46084.64	1	1	1	111
70	Signal Collectibles Ltd.	514	3347.900667	2	476	50218.51	1	1	1	111
25	Daedalus Designs Imports	699	3452.620500	2	465	69052.41	1	1	1	111
15	CAF Imports	468	3818.619231	2	439	49642.05	1	1	1	111

# Top 5 Loyal Customers

	CUSTOMERNAME	QUANTITYORDERED	SALES	ORDERNUMBER	Recency	Monetary	Recency_Score	Frequency_Score	Monetary_Score	RFM
3	Anna's Decorations, Ltd	1469	3347.741957	4	83	153996.13	3	4	4	344
53	Mini Gifts Distributors Ltd.	6366	3638.100333	17	2	654858.06	4	4	4	444
79	Tokyo Collectables, Ltd	1150	3767.585625	4	39	120562.74	4	4	3	443
78	The Sharp Gifts Warehouse	1656	4000.256750	4	39	160010.27	4	4	4	444
76	Technics Stores Inc.	1179	3552.443235	4	147	120783.07	3	4	4	344

**THANK YOU**

