RFM Project

Agenda

Problem Statement

Executive Summary

Exploratory Analysis

Customer Segmentation using RFM

Tableau Link for RFM Project -
https://public.tableau.com/views/RFMProject_17281210662530/ProductLineSaleswisedistribution?:language=en-US&:sid=&:redirect=auth&:display_count=n&:origin=viz_share_link

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.

1 - Agenda & Executive Summary of the Data

- Contents of the ppt
- Problem statement
- About Data (Info, Shape, Summary Stats, your assumptions about data)

2 - 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
- Summarise the inferences

3 - Customer Segmentation using RFM analysis (4 segments)

- What is RFM?
- What all parameters used, and assumptions made
- Showcase the KNIME workflow image
- What results are there in the output table head?

4 - Inferences from RFM Analysis and identified segments

- Who are your best customers?
- Which customers are on the verge of churning?
- Who are your lost customers?
- Who are your loyal customers?

1 - Executive Summary of Data

The dataset consists of 2747 rows and 20 columns. There are 1 datetime64[ns], 2 float64, 5 int64, 12 object data types of columns.

The data is of 3 years (2018,2019,2020). We will perform Exploratory Analysis first and then we will perform RFM analysis.

The following are features:

	Column	Non-Null	Count	Data Type
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 Description: - There are no missing values in the data.

	count	mean	min	25%	50%	75%	max	std
ORDERNUMBER	2747.0	10259.761558	10100.0	10181.0	10264.0	10334.5	10425.0	91.877521
QUANTITYORDERED	2747.0	35.103021	6.0	27.0	35.0	43.0	97.0	9.762135
PRICEEACH	2747.0	101.098951	26.88	68.745	95.55	127.1	252.87	42.042548
ORDERLINENUMBER	2747.0	6.491081	1.0	3.0	6.0	9.0	18.0	4.230544
SALES	2747.0	3553.047583	482.13	2204.35	3184.8	4503.095	14082.8	1838.953901
ORDERDATE	2747	2019-05-13 21:56:17.211503360	2018-01-06 00:00:00	2018-11-08 00:00:00	2019-06-24 00:00:00	2019-11-17 00:00:00	2020-05-31 00:00:00	NaN
DAYS_SINCE_LASTORDER	2747.0	1757.085912	42.0	1077.0	1761.0	2436.5	3562.0	819.280576
MSRP	2747.0	100.691664	33.0	68.0	99.0	124.0	214.0	40.114802

Check Null Value

	0
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

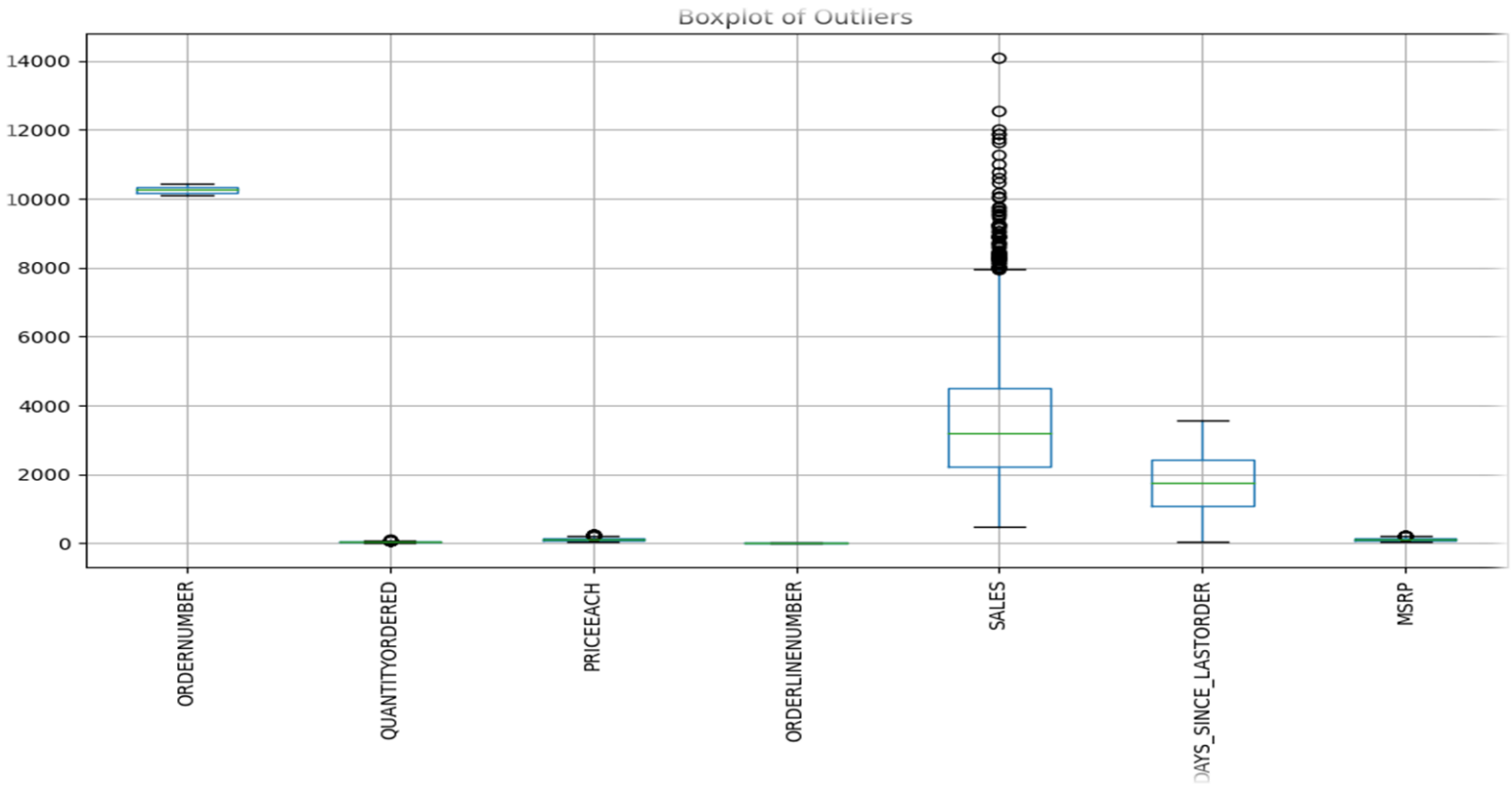
STATUS	0
PRODUCTLINE	0
MSRP	0
PRODUCTCODE	0
CUSTOMERNAME	0
PHONE	0
ADDRESSLINE1	0
CITY	0
POSTALCODE	0
COUNTRY	0
CONTACTLASTNAME	0
CONTACTFIRSTNAME	0
DEALSIZE	0

dtype: int64

RFM/MRA Project

2 - Exploratory Analysis and Inferences –

Outliers Check –



Observations:

• Sales has lot of data points outside the whisker from upper quartile.

• Quantity Ordered, MSRP, and Price Each also have some values outside the whisker from upper quartile.

Unique Values – There are no duplicates values in dataset.

```
STATUS: ['Shipped' 'Disputed' 'In Process' 'Cancelled' 'On Hold' 'Resolved']

PRODUCTLINE: ['Motorcycles' 'Classic Cars' 'Trucks and Buses' 'Vintage Cars' 'Planes'
 'Ships' 'Trains']

PRODUCTCODE: ['S10_1678' 'S10_1949' 'S10_2016' 'S10_4698' 'S10_4757' 'S10_4962'
 'S12_1099' 'S12_1108' 'S12_1666' 'S12_2823' 'S12_3148' 'S12_3380'
 'S12_3891' 'S12_3990' 'S12_4473' 'S12_4675' 'S18_1097' 'S18_1129'
 'S18_1342' 'S18_1367' 'S18_1589' 'S18_1662' 'S18_1749' 'S18_1889'
 'S18_1984' 'S18_2238' 'S18_2248' 'S18_2319' 'S18_2325' 'S18_2432'
 'S18_2581' 'S18_2625' 'S18_2795' 'S18_2870' 'S18_2949' 'S18_2957'
 'S18_3029' 'S18_3136' 'S18_3140' 'S18_3232' 'S18_3259' 'S18_3278'
 'S18_3320' 'S18_3482' 'S18_3685' 'S18_3782' 'S18_3856' 'S18_4027'
 'S18_4409' 'S18_4522' 'S18_4600' 'S18_4668' 'S18_4721' 'S18_4933'
 'S24_1046' 'S24_1444' 'S24_1578' 'S24_1628' 'S24_1785' 'S24_1937'
 'S24_2000' 'S24_2011' 'S24_2022' 'S24_2300' 'S24_2360' 'S24_2766'
 'S24_2840' 'S24_2841' 'S24_2887' 'S24_2972' 'S24_3151' 'S24_3191'
 'S24_3371' 'S24_3420' 'S24_3432' 'S24_3816' 'S24_3856' 'S24_3949'
 'S24_3969' 'S24_4048' 'S24_4258' 'S24_4278' 'S24_4620' 'S32_1268'
 'S32_1374' 'S32_2206' 'S32_2509' 'S32_3207' 'S32_3522' 'S32_4289'
 'S32_4485' 'S50_1341' 'S50_1392' 'S50_1514' 'S50_4713' 'S700_1138'
 'S700_1691' 'S700_1938' 'S700_2047' 'S700_2466' 'S700_2610' 'S700_2824'
 'S700_2834' 'S700_3167' 'S700_3505' 'S700_3962' 'S700_4002' 'S72_1253'
 'S72_3212']

DEALSIZE: ['Small' 'Medium' 'Large']
```

Univariate Analysis –

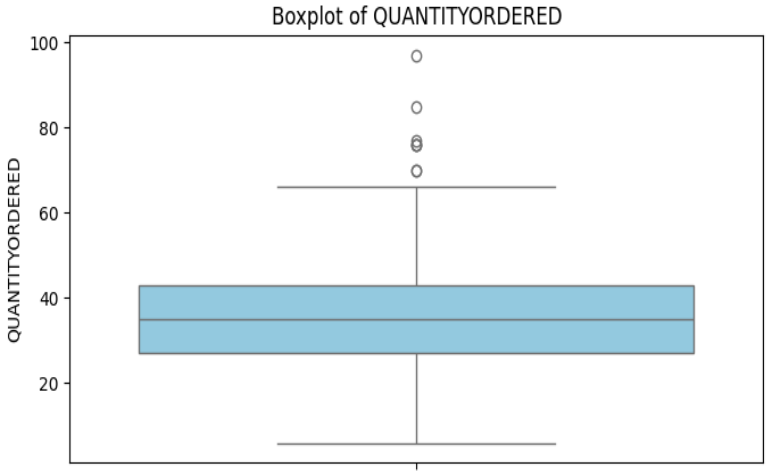
Numeric Data - QUANTITYORDERED Description:

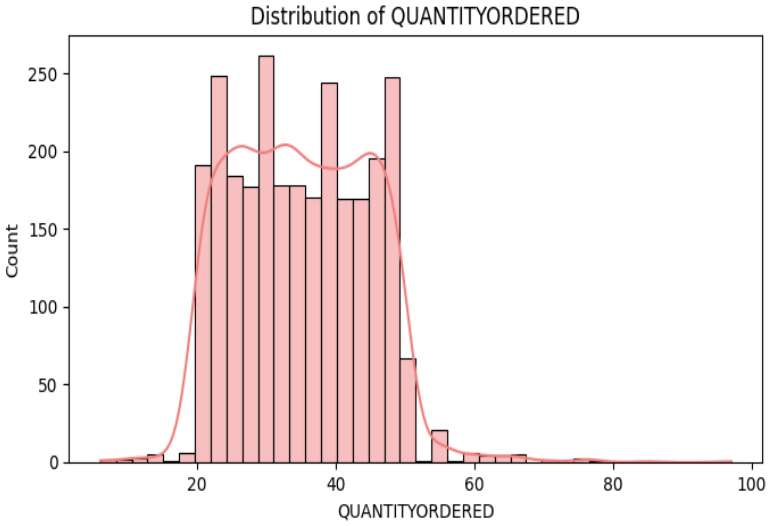
```
count   2747.000000
mean      35.103021
std        9.762135
min        6.000000
25%       27.000000
50%       35.000000
75%       43.000000
max       97.000000
Name: QUANTITYORDERED, dtype: float64
```

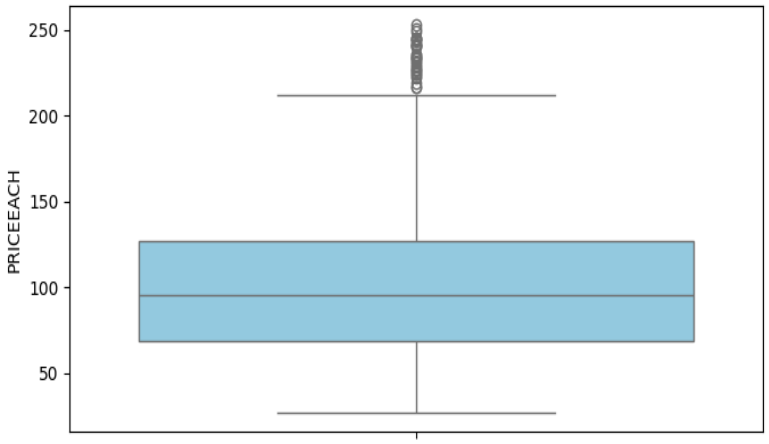
```
PRICEEACH Description:
count   2747.000000
mean     101.098951
std       42.042548
min       26.880000
25%       68.745000
50%       95.550000
75%      127.100000
max      252.870000
Name: PRICEEACH, dtype: float64
```

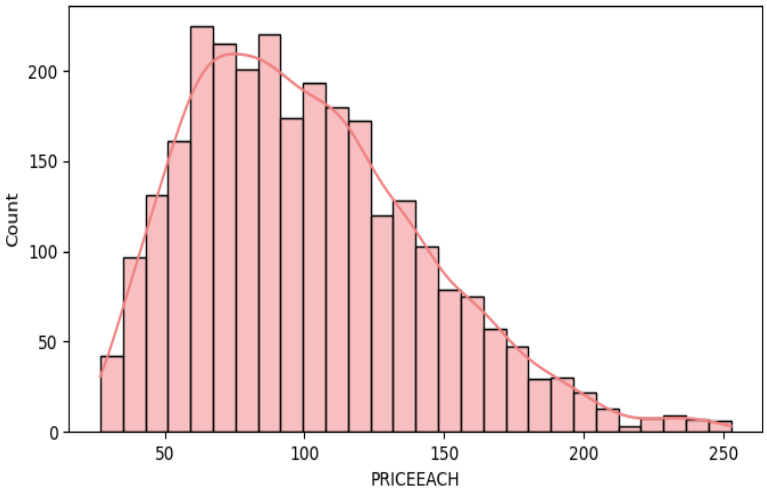
```
SALES Description:
count    2747.000000
mean     3553.047583
std      1838.953901
min       482.130000
25%      2204.350000
50%      3184.800000
75%      4503.095000
max     14082.800000
Name: SALES, dtype: float64
```

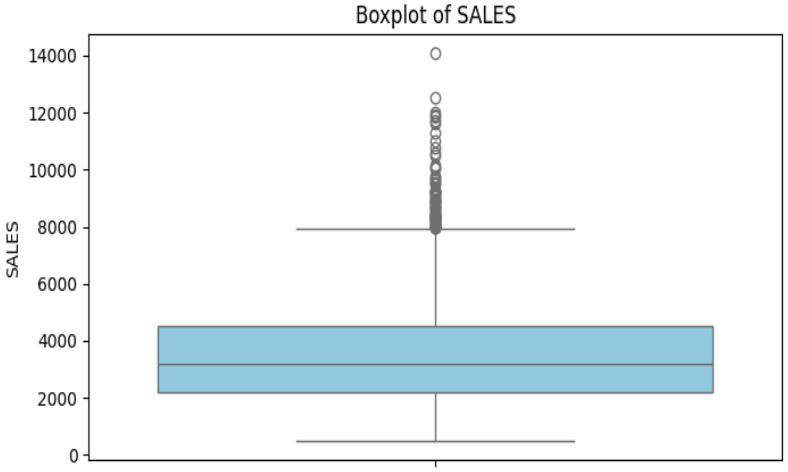
```
MSRP Description:
count   2747.000000
mean     100.691664
std       40.114802
min       33.000000
25%       68.000000
50%       99.000000
75%      124.000000
max      214.000000
Name: MSRP, dtype: float64
```

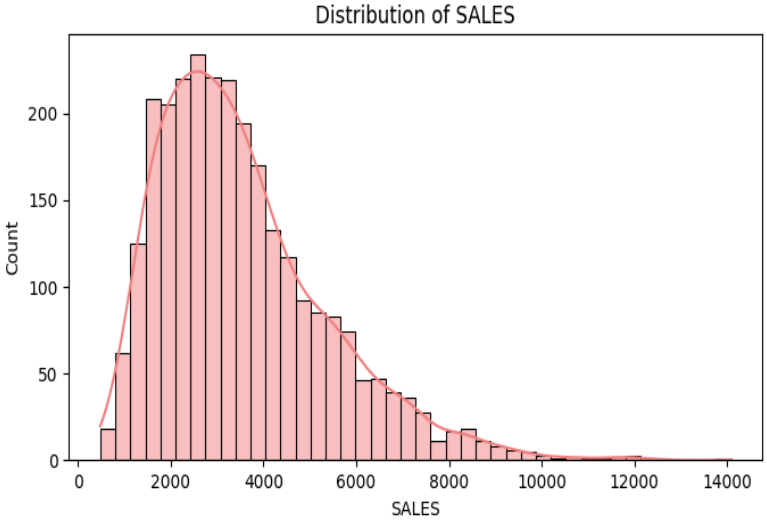

Boxplot of QUANTITYORDERED
QUANTITYORDERED

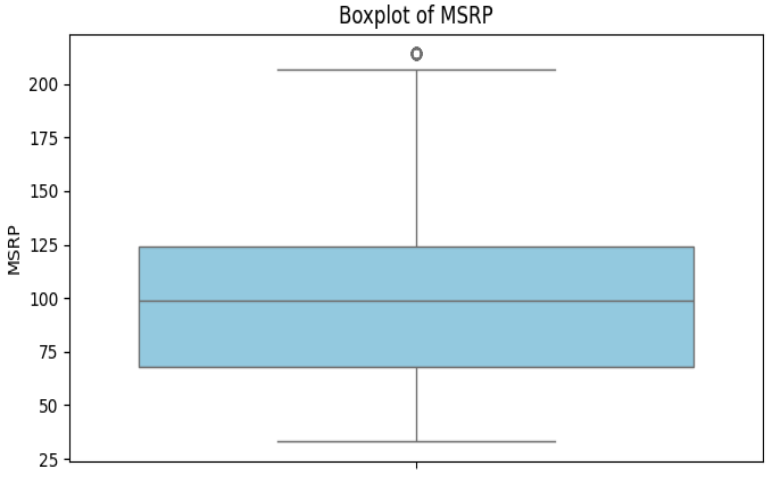

Distribution of QUANTITYORDERED
Count
QUANTITYORDERED

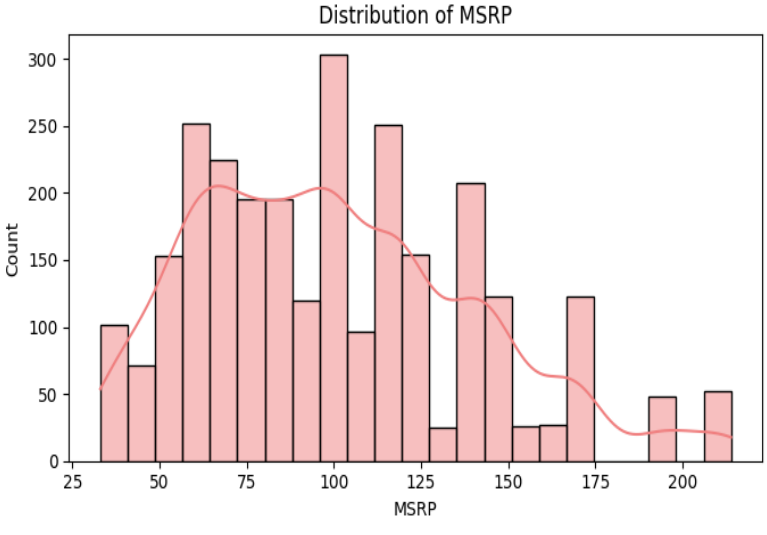

Boxplot of PRICEEACH
PRICEEACH


Distribution of PRICEEACH
Count
PRICEEACH


Boxplot of SALES
SALES


Distribution of SALES
Count
SALES


Boxplot of MSRP
MSRP


Distribution of MSRP
Count
MSRP

Observations –

1. QUANTITYORDERED:

- **Boxplot**: The data has a fairly symmetric distribution, with a few outliers above 60 units. The interquartile range (IQR) suggests most orders fall between 25 and 45 units.
- **Histogram**: The distribution is roughly uniform between 20 and 50 units, with some orders less frequent above 50 units. There is a drop-off after 50 units.

2. PRICEEACH:

- **Boxplot**: The distribution has a large number of outliers above approximately 180. The interquartile range (IQR) suggests the typical price for each product falls between around 70 and 130.
- **Histogram**: The distribution of PRICEEACH is right-skewed (positively skewed), with most prices falling between 50 and 150, and a few higher prices extending up to 250.

3. SALES:

- **Boxplot**: There are numerous outliers on the higher side of the distribution, with values extending beyond 10,000. Most of the sales values are clustered between around 2,000 and 6,000.
- **Histogram**: The SALES data is heavily right-skewed, with most sales concentrated between 0 and 6,000. A long tail of higher sales values stretches out beyond 10,000, indicating occasional large transactions.

4. MSRP (Manufacturer's Suggested Retail Price):

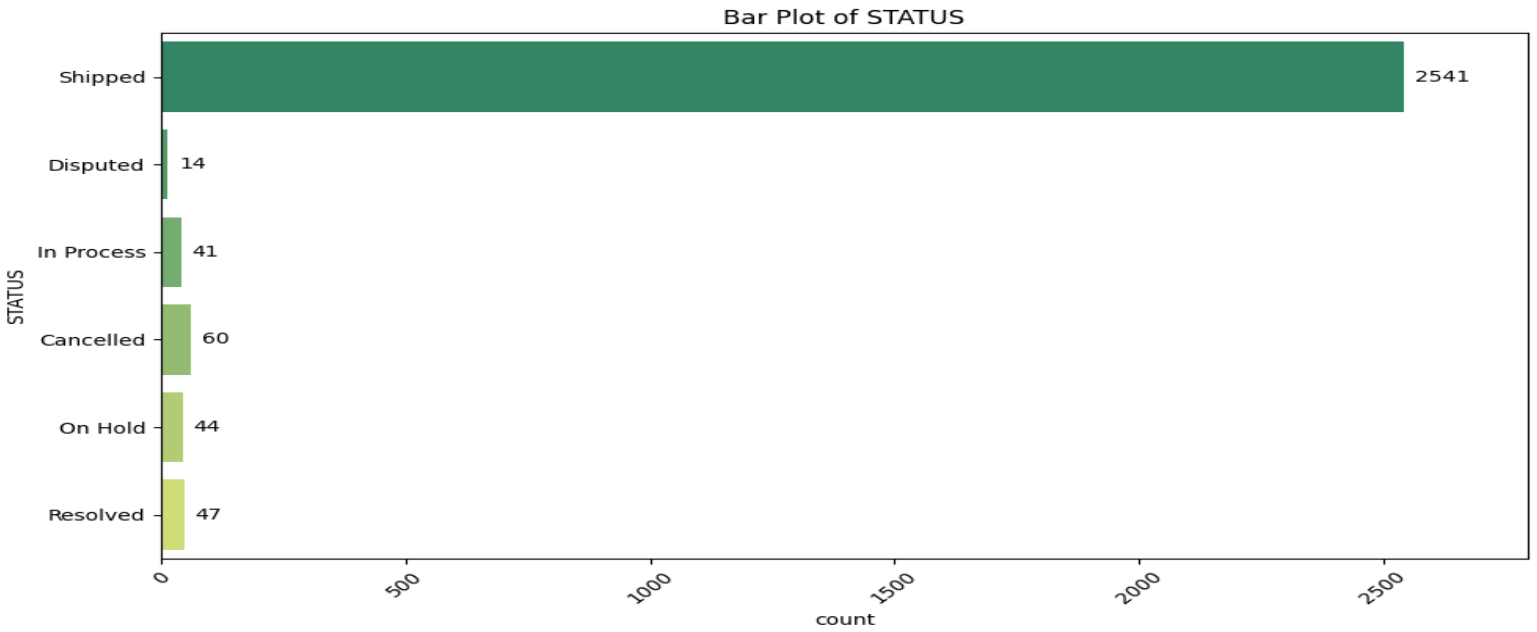
- **Boxplot**: The MSRP data has a more symmetric distribution with fewer outliers. The IQR suggests most MSRP values fall between 75 and 125.
- **Histogram**: The distribution of MSRP is relatively uniform across the range of 50 to 150, with a few higher values around 200. The data has a slightly multi-modal appearance, with peaks around 50, 100, and 150.

Categorical data Analysis –

1 - **Dominance of 'Shipped' Status**: The overwhelming majority of the orders are in the "Shipped" status, with a count of **2541**. This suggests that most of the transactions have been successfully processed and shipped.

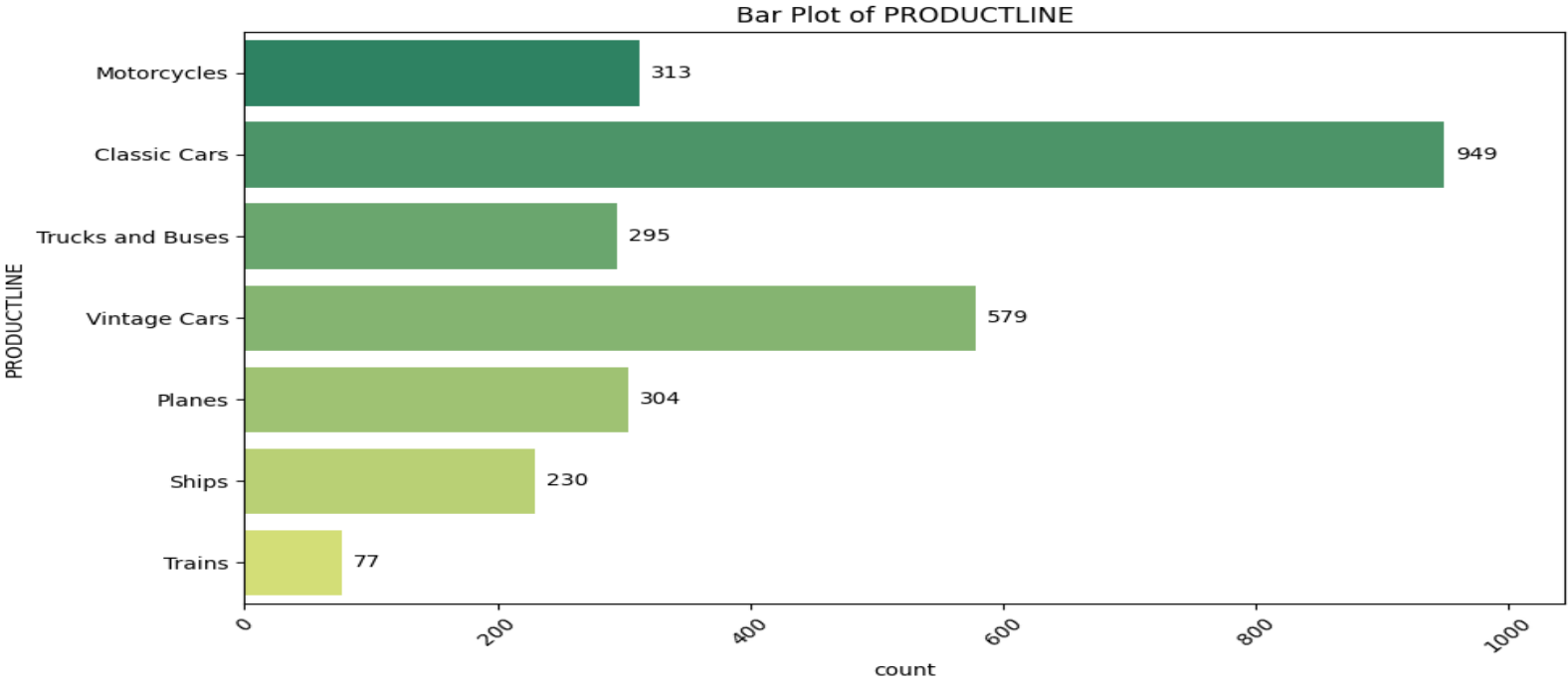
2 - **Smaller Categories**: "Cancelled" has 60 instances.

- "On Hold" has 44 instances.
- "Resolved" has 47 instances.
- "In Process" has 41 instances.
- "Disputed" has the lowest count with just 14 instances.



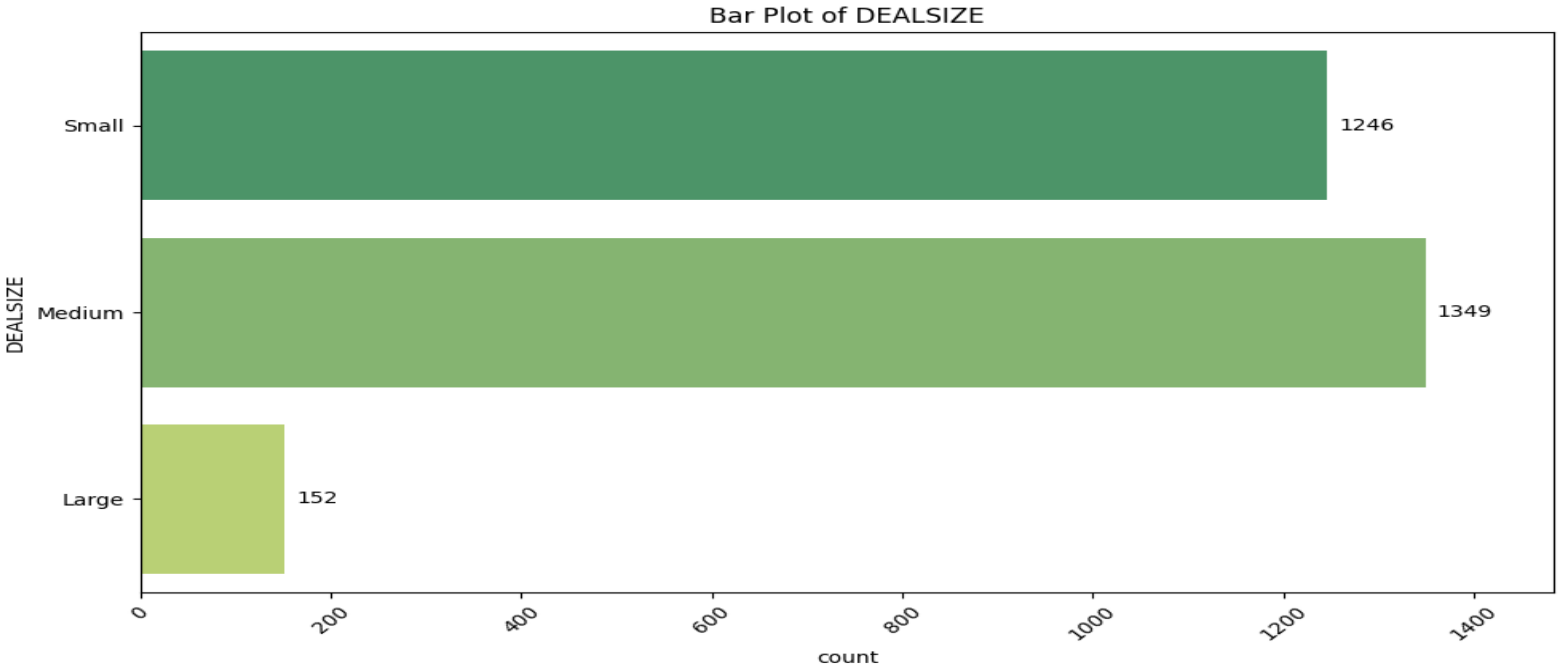
Observation of PRODUCTLINE -

- Classic Cars is the product line with the highest count, followed by Vintage Cars.
- Trucks and Buses and Motorcycles have similar counts.
- Planes, Ships, and Trains have relatively lower counts.



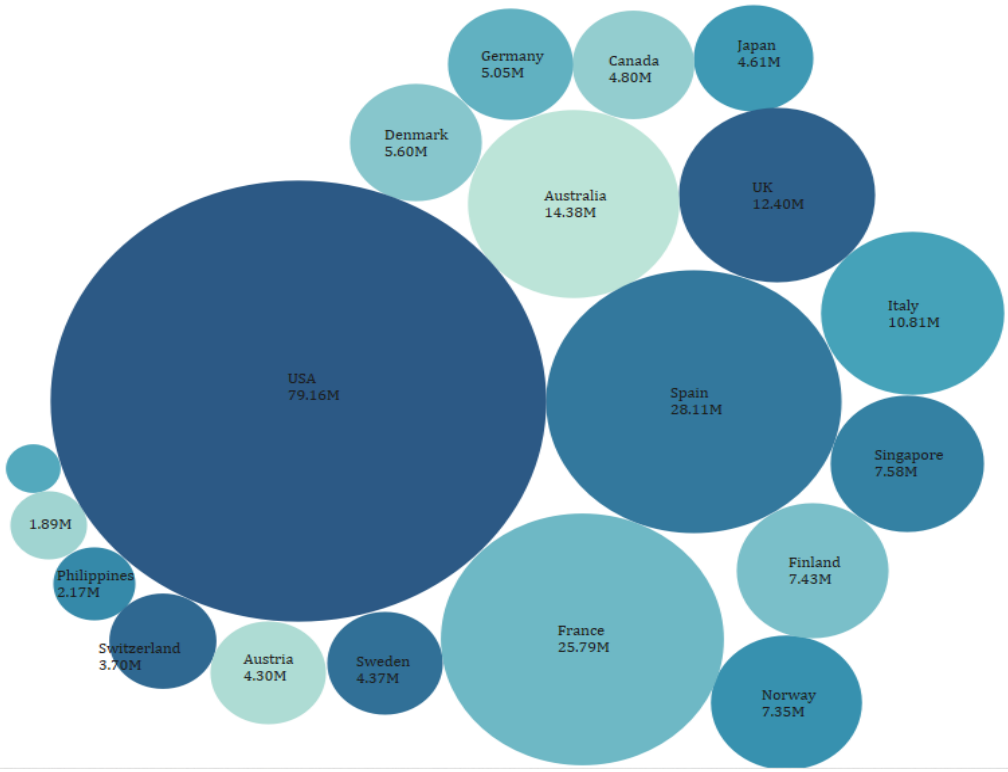
Observation of Deal Size -

- Medium is the deal size with the highest count, followed by Small.
- Large has the lowest count.
- The distribution of the deal sizes is skewed to the right, with a few deal sizes having much higher counts than others.



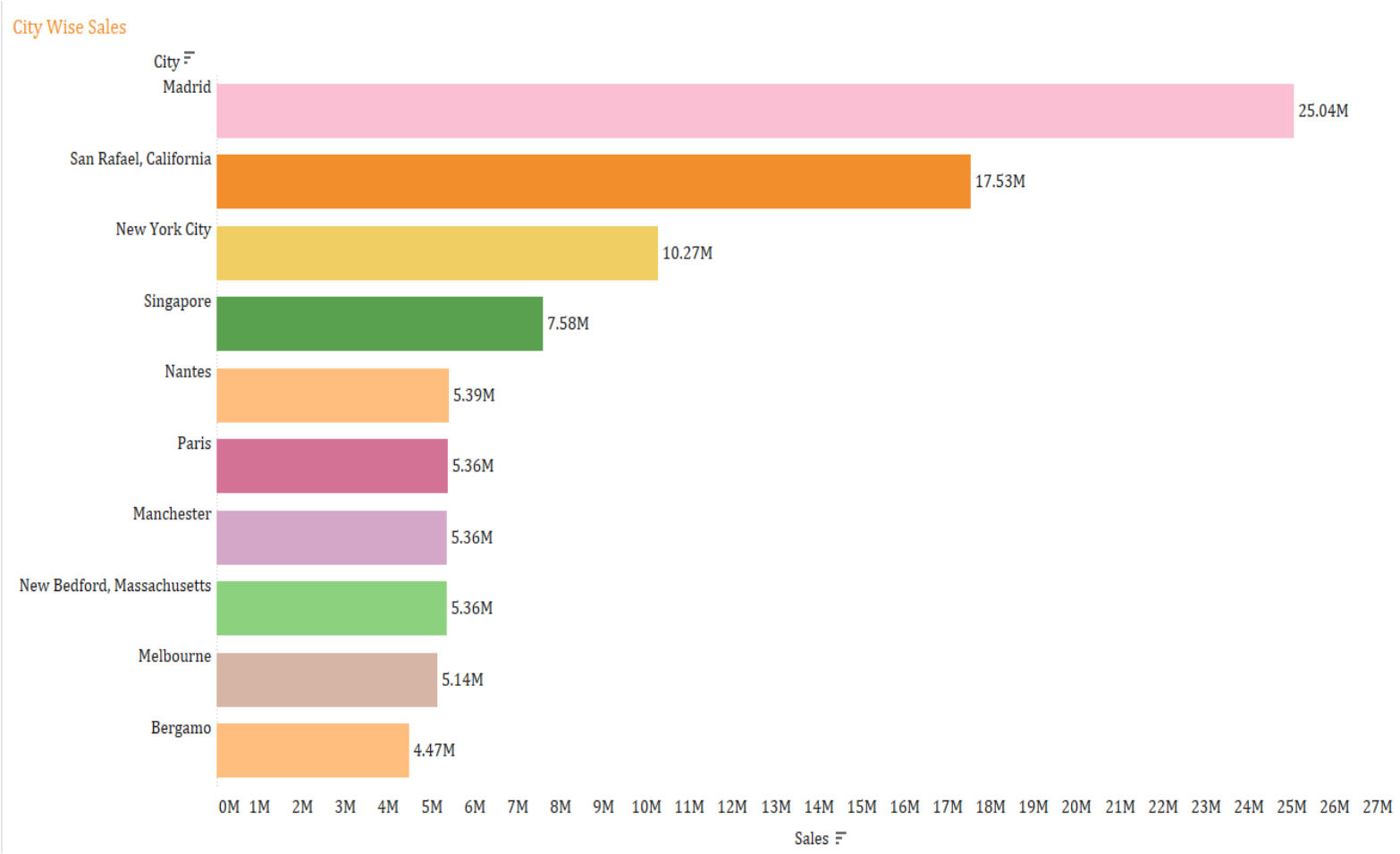
Bivariate Analysis –

Country wise Sales



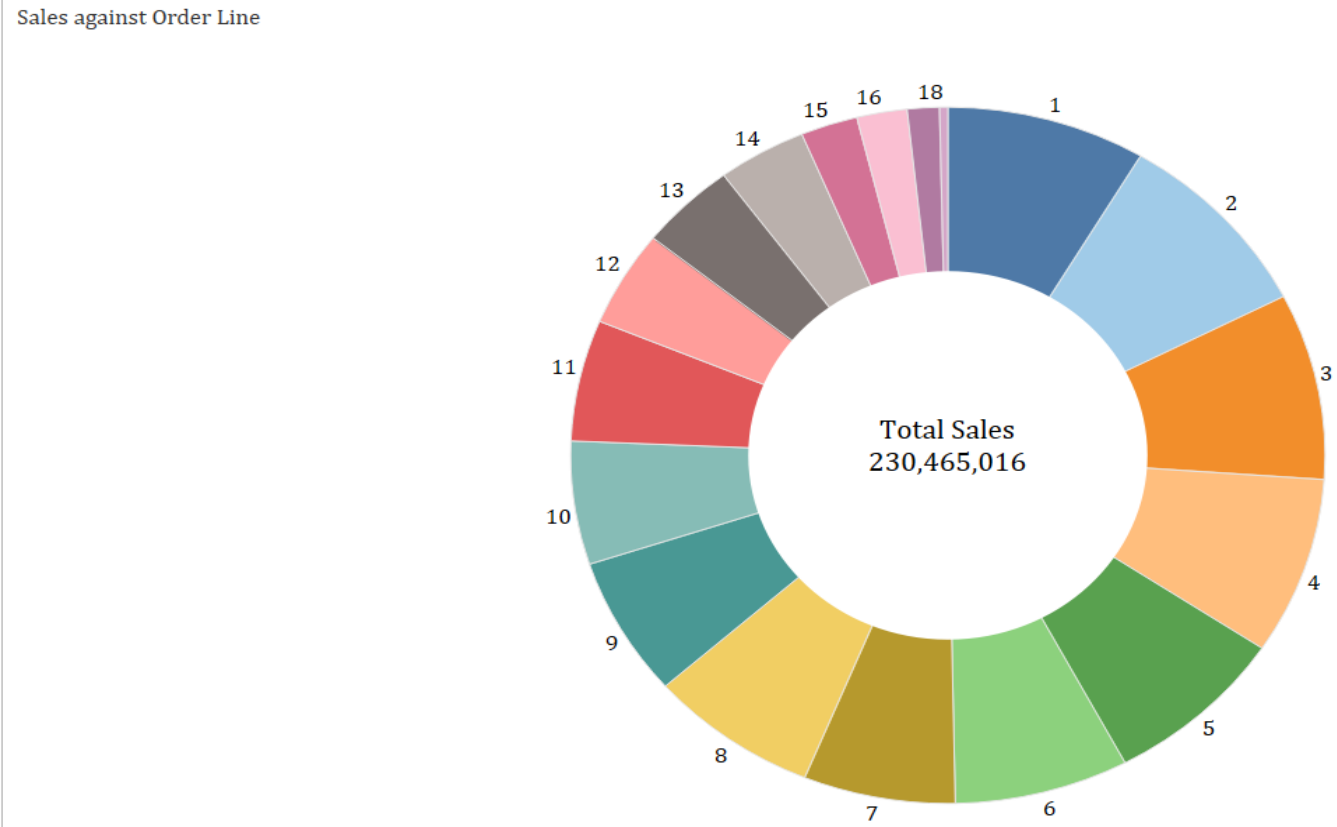
Distribution of Sales across Different Countries –

- USA has the largest bubble, indicating that it has the highest sales volume.
- Spain and Australia also have relatively large bubbles, suggesting that they have significant sales.
- Germany, Canada, Japan, Italy, and Singapore have medium-sized bubbles, indicating moderate sales.



Distribution of Sales across Cities -

- Madrid has the highest sales, followed by San Rafael (California), New York City has significantly lower sales than the top two cities.
- Singapore, Nantes, Paris, Manchester, New Bedford, Massachusetts, and Melbourne all have similar sales, ranging from 5.14 million to 5.39 million.
- Bergamo has the lowest sales 4.447 million.



Observation of Sales for Different Order Lines -

1- Top Performing Order Lines:

Order line 2 leads with the highest total sales at 20,562,469. Order line 3 follows closely with 19,938,939 in total sales.

The top three order lines (2, 3, 1) are all above 19 million in sales, indicating a similarly high level of performance.

2 - Sales Distribution:

The sales figures decrease gradually from the top, with the top 6 order lines all having sales above 17 million.

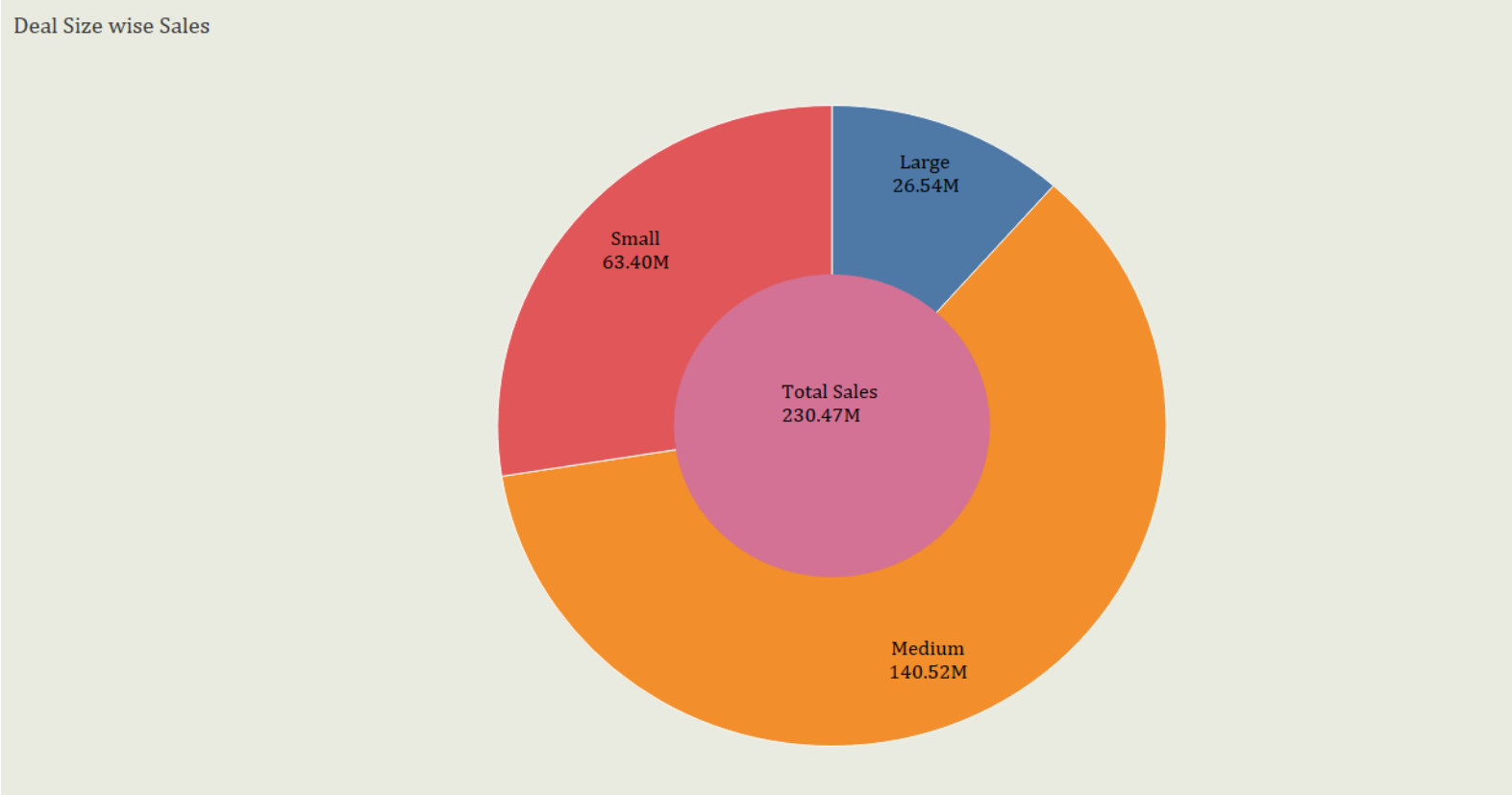
There's a noticeable drop after order line 9, where sales figures drop below 15 million.

3 - Lower Performing Order Lines:

Order line 18 has the lowest sales at 791,765, significantly lower than the rest.

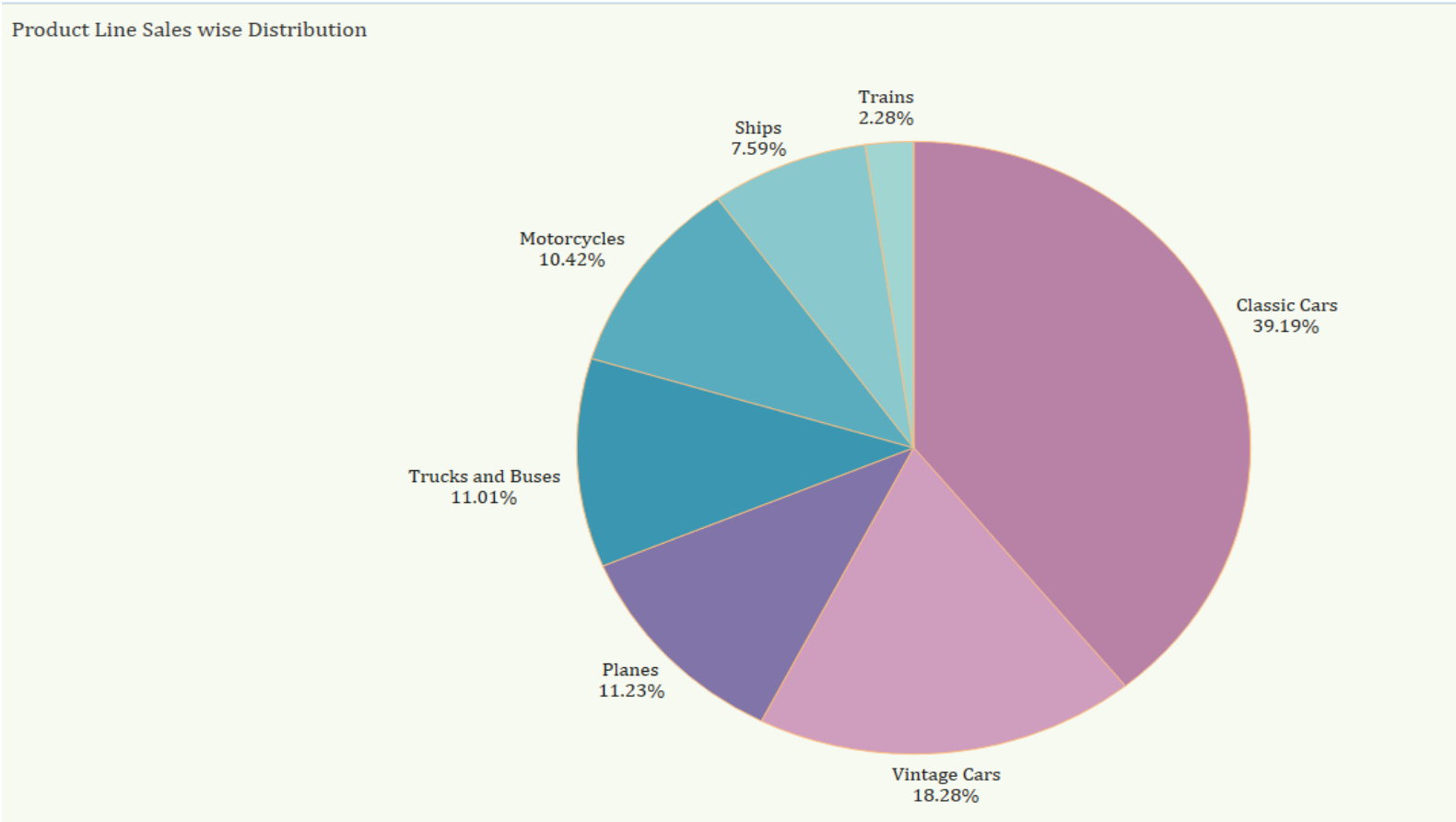
The sales of order lines 17 and 18 are far below others, with order line 17 reaching 3,166,917, much lower than the top order lines.

4 - Overall Range: The range of total sales spans from 791,765 (lowest) to 20,562,469 (highest), showing a substantial variance in performance across the different order lines.



Distribution of Sales Across Different Deal Sizes -

- Medium has the largest slice, indicating that it contributes the most to total sales.
- Small and Large have relatively smaller slices, suggesting that they contribute less to total sales.



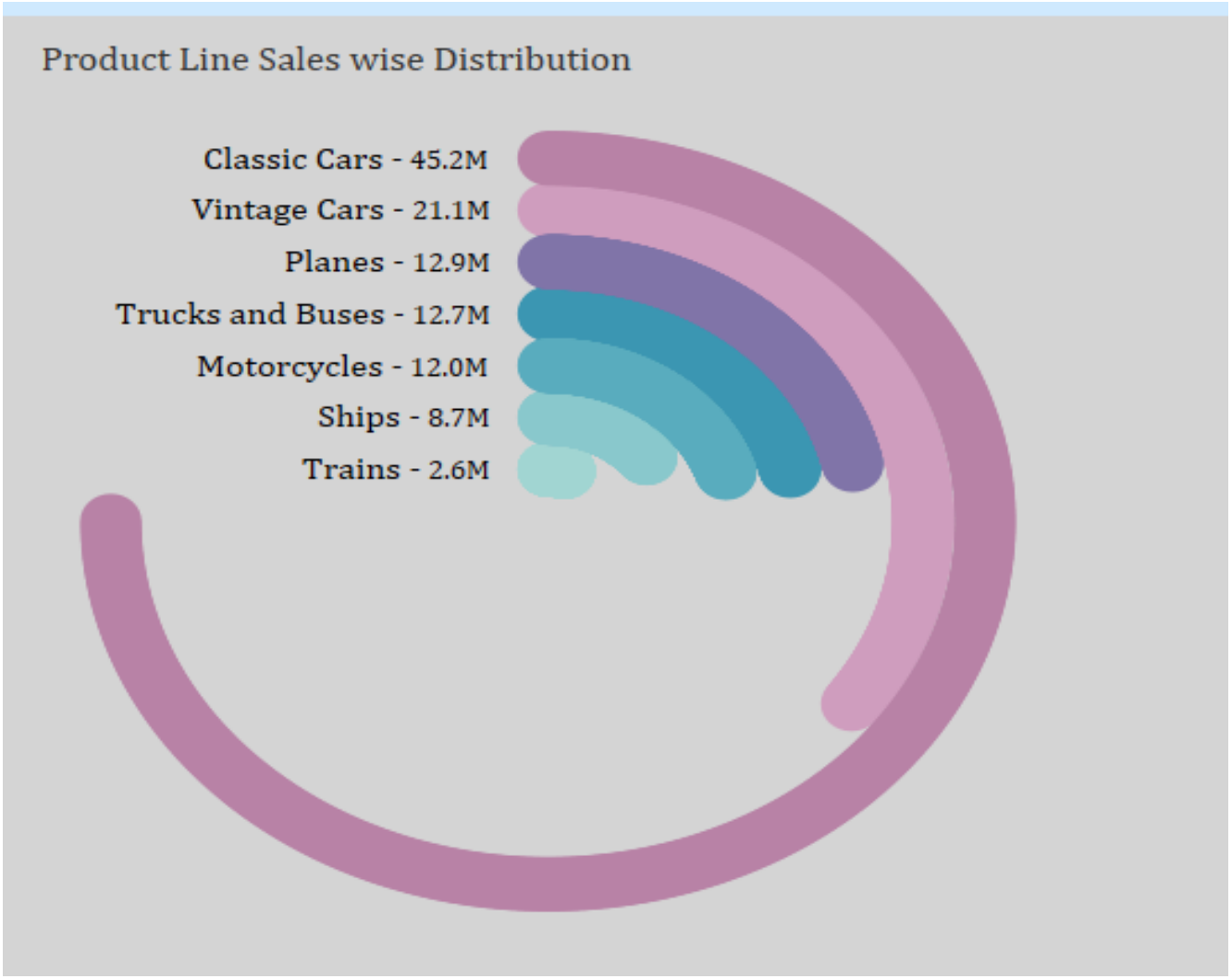
Product Lines Sales Percentage wise -

1 - Dominance of Classic Cars - Classic cars make up the largest portion at 39.19%. This indicates a significant interest or market presence in this category.

2 - Significant Share of Vintage Cars - Vintage cars account for 18.28% of the total. While not as dominant as classic cars, they still represent a substantial segment. This suggests a healthy demand for vintage car parts, maintenance, and collector events.

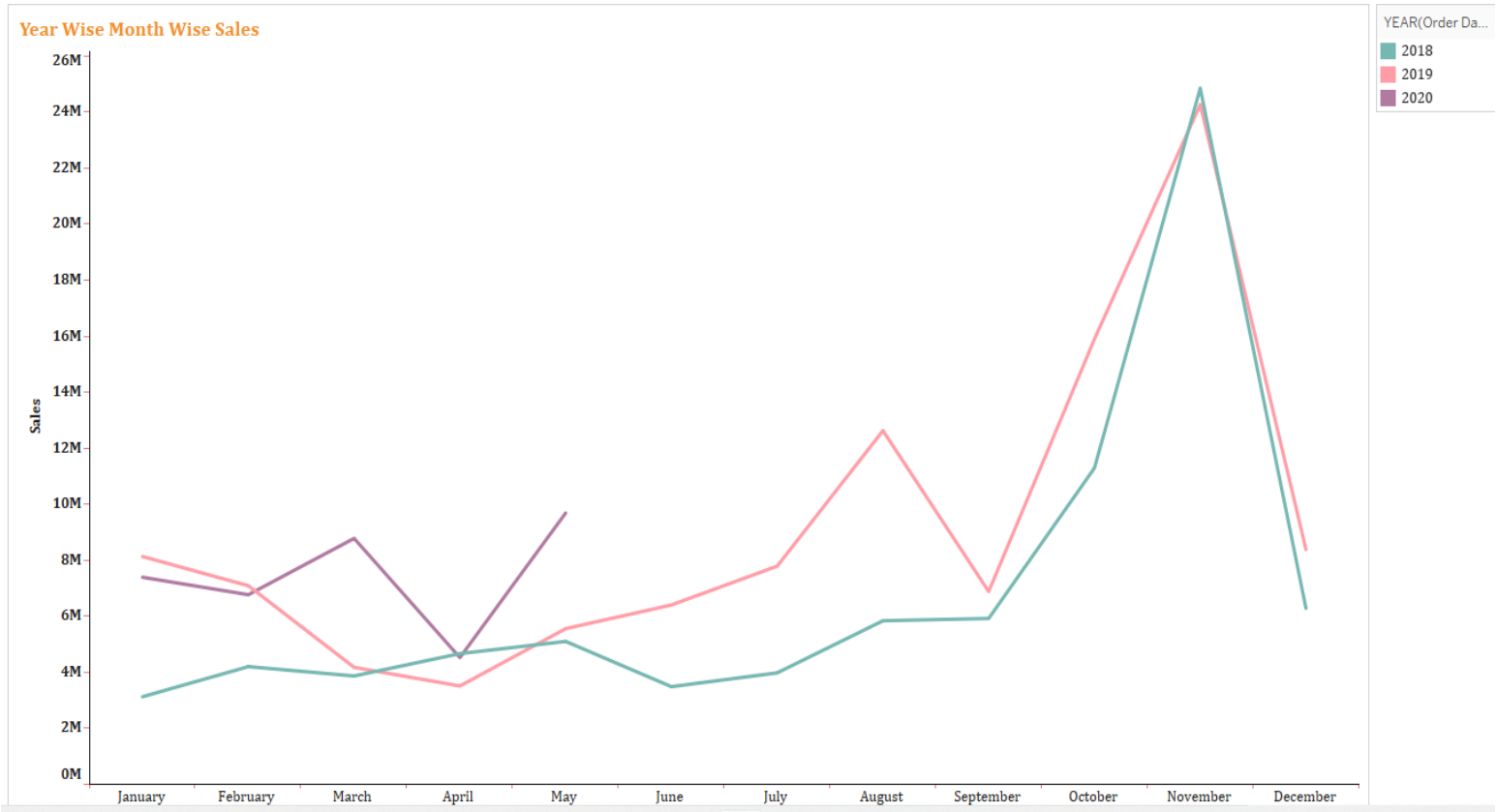
3 - Balanced Distribution Among Other Categories - Planes (11.23%), Trucks and Buses (11.01%), and Motorcycles (10.42%) have relatively similar shares. This balanced distribution indicates diverse interests and opportunities across these categories.

4 - Smaller Segments - Ships and Trains: - Ships (7.59%) and trains (2.28%) have the smallest shares.



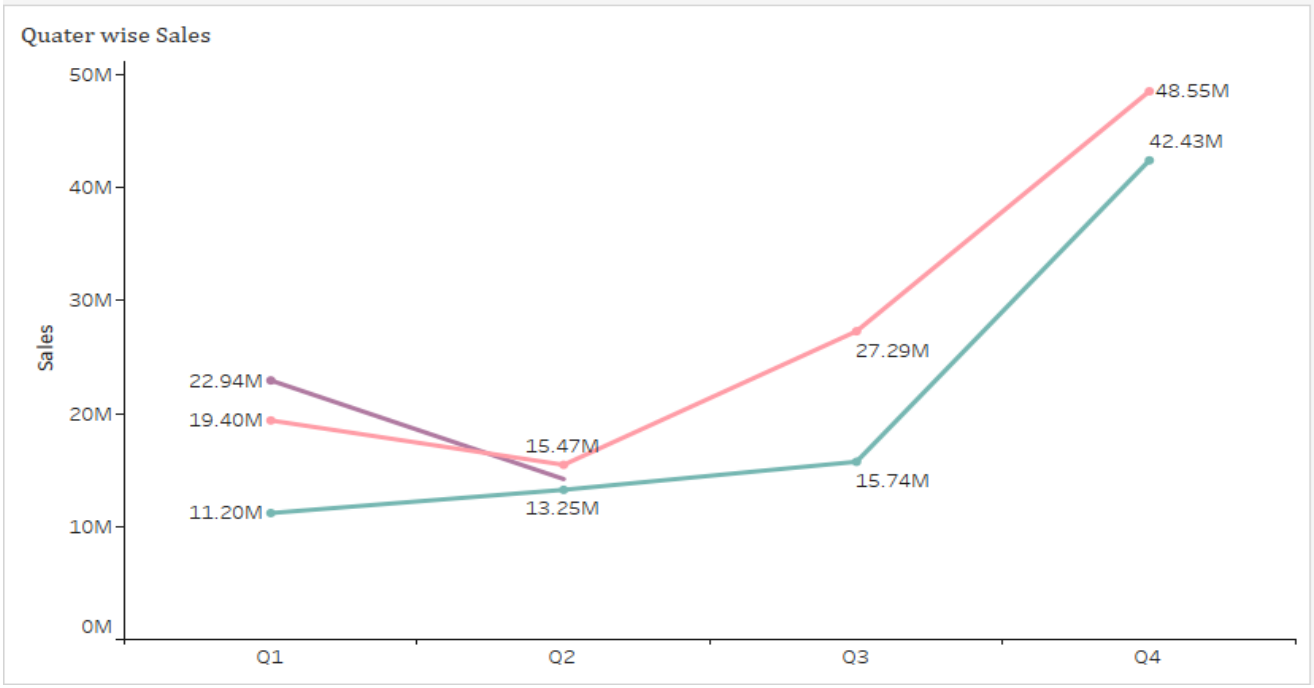
Here are some observations based on the chart:

- **Classic Cars** has the largest slice, indicating that it contributes the most to total sales.
- **Vintage Cars** and **Motorcycles** also have relatively large slices, suggesting that they contribute significantly to total sales.
- **Trucks and Buses**, **Planes**, **Ships**, and **Trains** have smaller slices, suggesting that they contribute less to total sales.



Distribution of Sales Year Wise –

- Revenue generated for last 3 months for both 2018 and 2019 is quite high and November is a peak for both the years
- Overall, sales seem to have increased year over year, indicating growth.



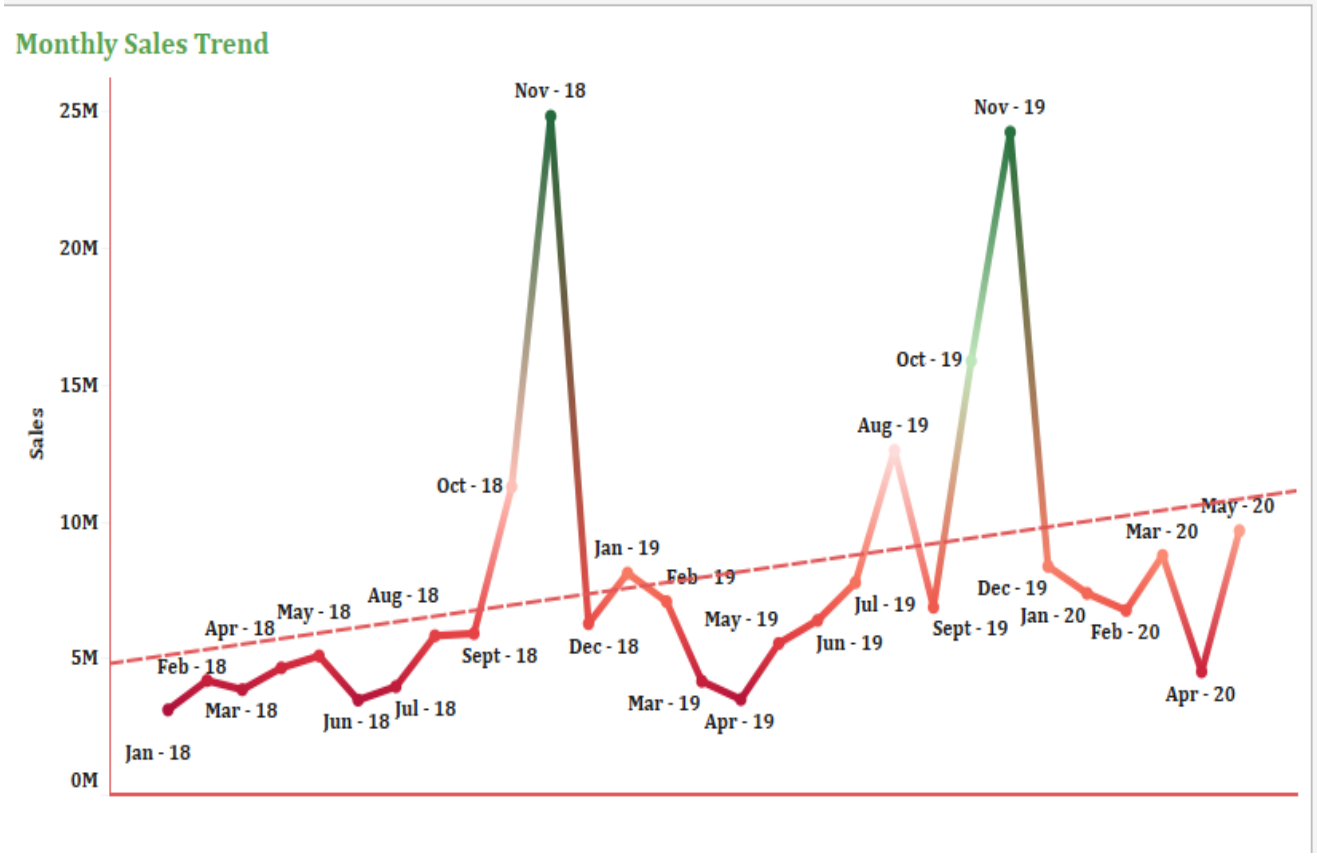
Quarterly Sales Trend -

1 - Year-Over-Year Comparison:

- In 2018, sales increased steadily from Q1 to Q4.
- In 2019, sales decreased from Q1 to Q2, then increased from Q2 to Q4.
- In 2020, sales increased from Q1, then decreased Q2.

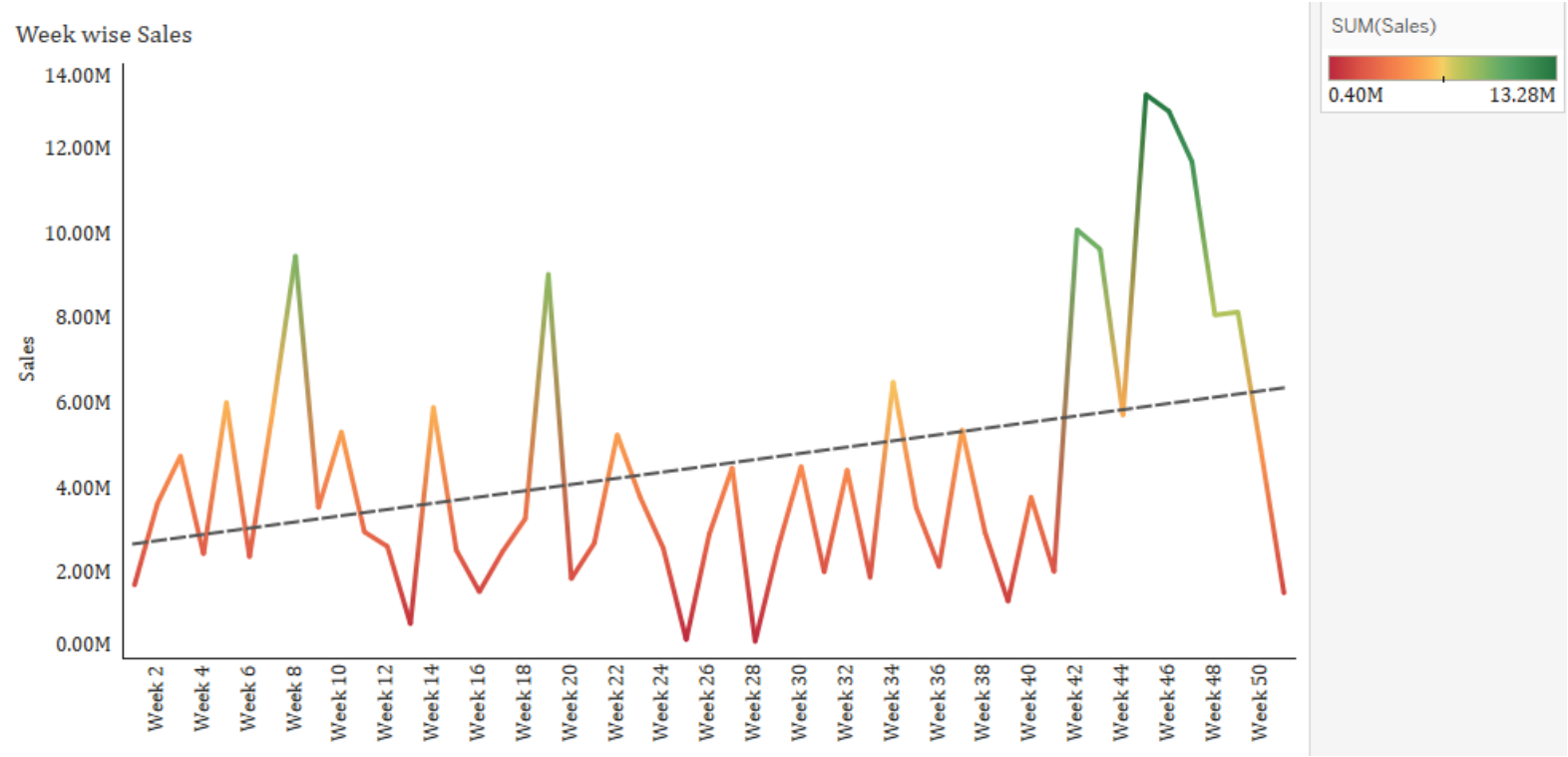
2 - Quarter-wise Comparison:

- Q1 sales have increased each year.
- Q2 sales have fluctuated.
- Q4 sales have been the highest in each year.



Monthly Sales Trend –

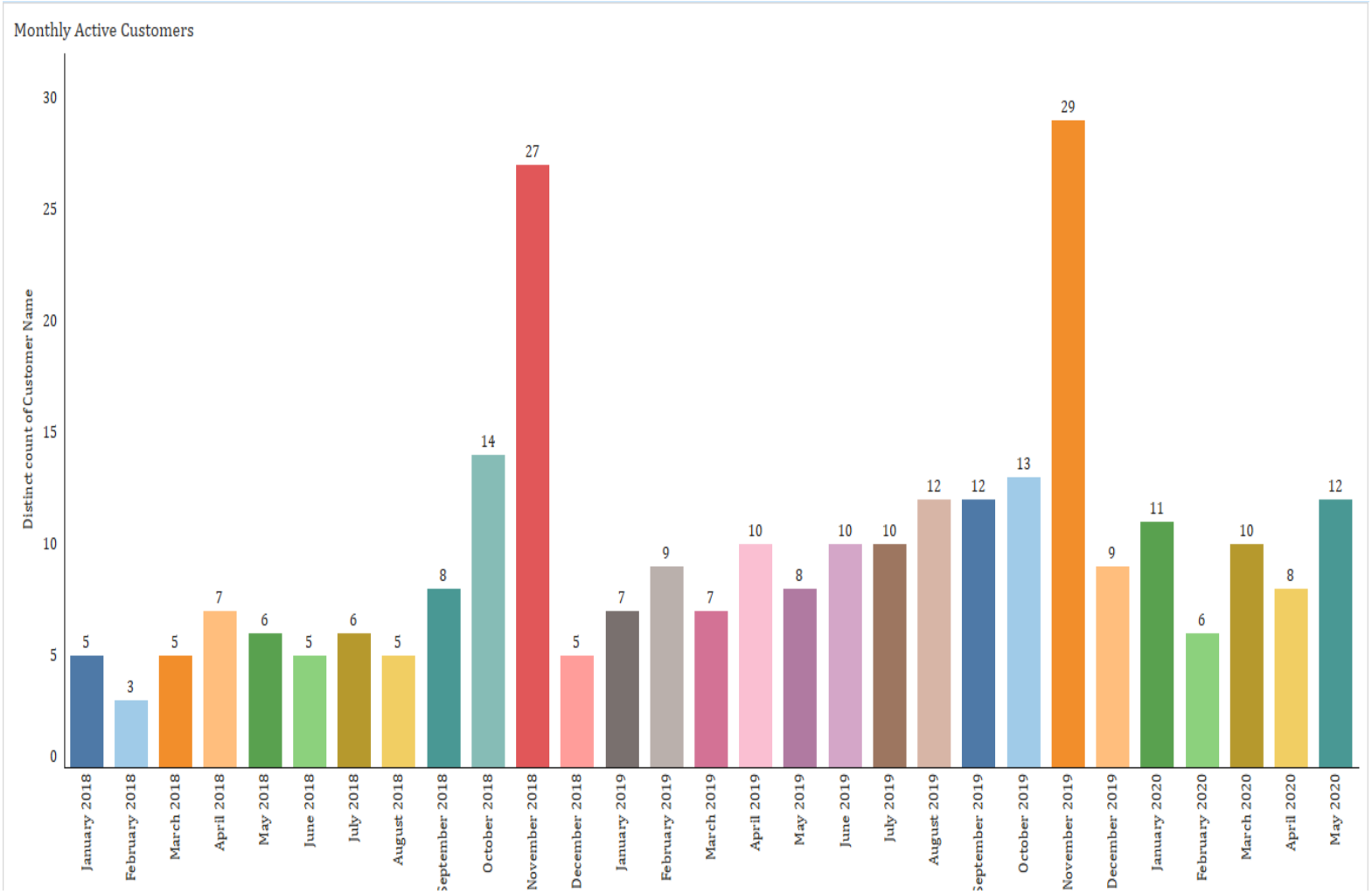
- **Seasonal Peaks**: There are noticeable peaks in sales during November 2018 and November 2019, suggesting a seasonal trend where sales increase significantly during these months.
- **Overall Decline**: Starting from December 2019, there is a significant decline in sales, which continues through to April 2020.
- **Fluctuations**: The sales trend fluctuates throughout the period, with several ups and downs, indicating variability in sales performance.



Weekly Sales Trend –

- Overall sales trend: There is a general upward trend in sales over time, as indicated by the dashed trend line.
- Seasonal patterns: Sales appear to fluctuate seasonally, with peaks around certain weeks and troughs around others.
- Week-over-week comparison: Sales vary significantly from week to week, with some weeks experiencing large increases or decreases in sales.

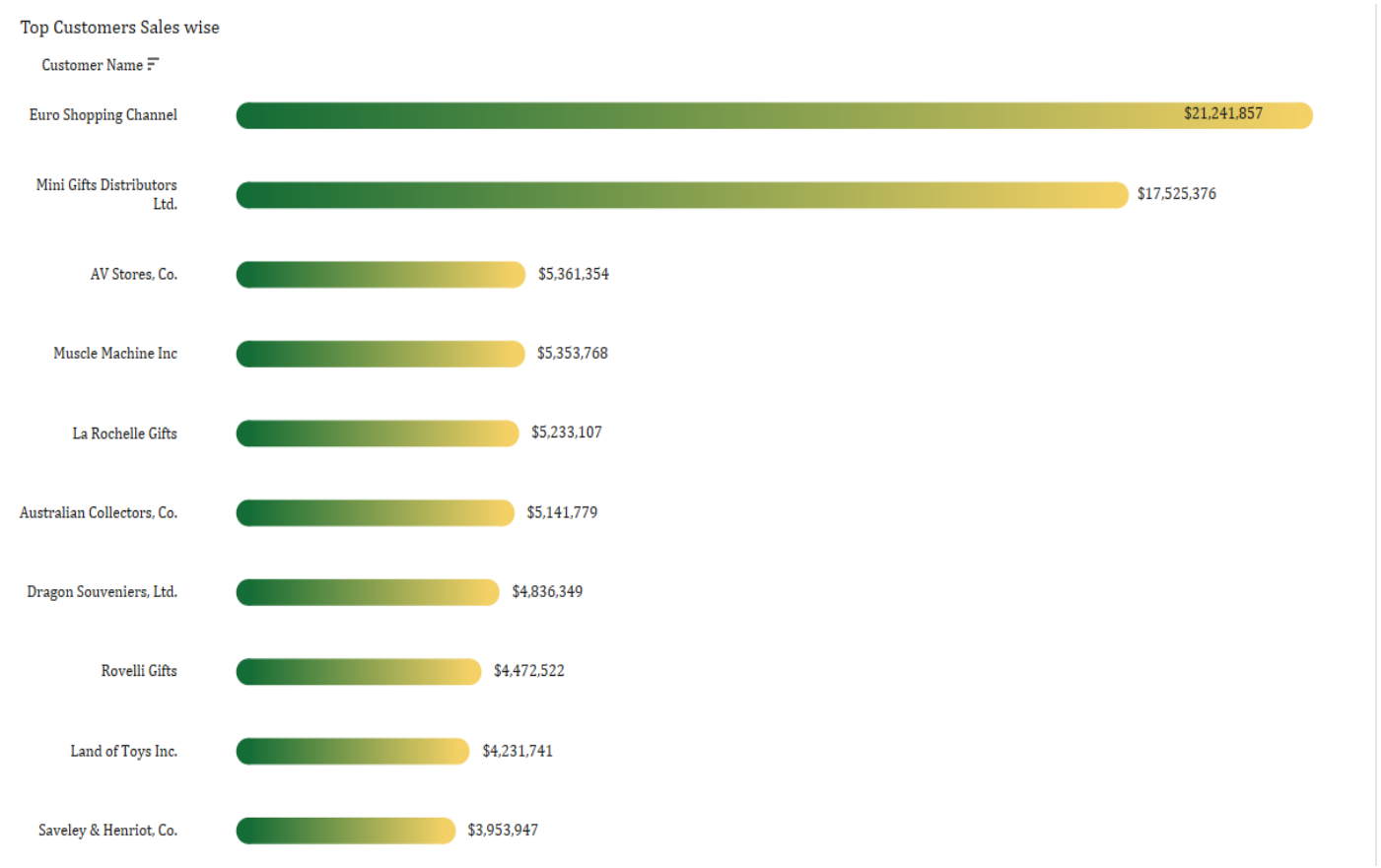
- Correlation with a trend line: The dashed trend line suggests that there is a general upward trend in sales over time, even when accounting for seasonal fluctuations.



Monthly Active Customers -

- Overall trend: The number of active customers has fluctuated over time, with periods of growth and decline.
- Seasonal patterns: There appears to be some seasonality in customer activity, with higher numbers of active customers in certain months compared to others.
- Peak periods: The chart shows two distinct peaks in customer activity: one in November 2018 and another in November 2019.
- Lowest points: The lowest points in customer activity occurred in January 2018, April 2019, and April 2020.

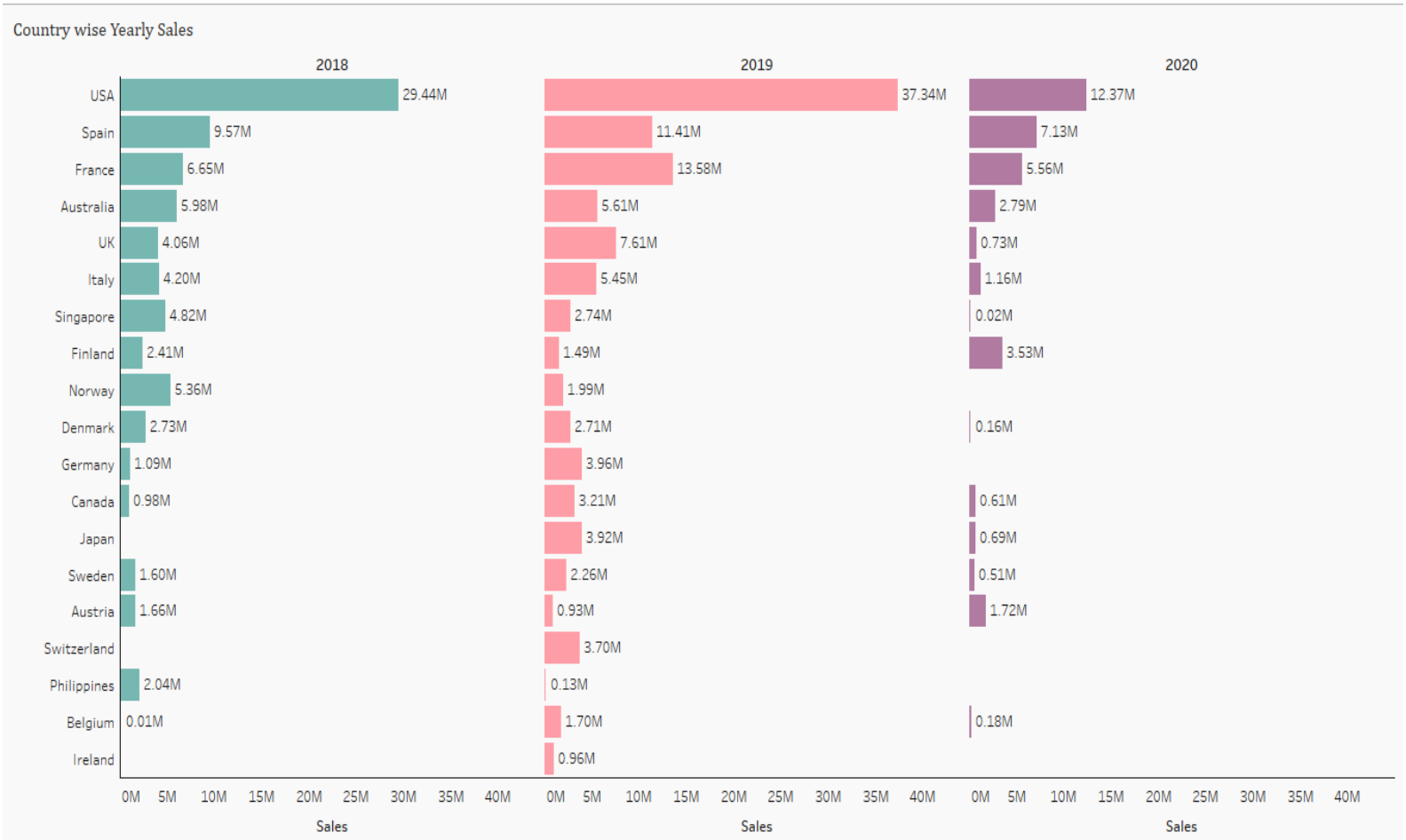
RFM/MRA Project



Customers Sales wise –

- Euro Shopping Channel is the only customer in the prime category, contributing of the total sales with $12.50 million.
- Mini Gifts Distributors Ltd. falls into the best customer category, generating $8.20 million in sales.

Multivariate Analysis –



Yearly Sales by Countries –

1 - USA Dominance: The USA consistently leads the sales in all three years:

- 2018: 29.44M
- 2019: 37.34M (peak)
- 2020: 12.37M

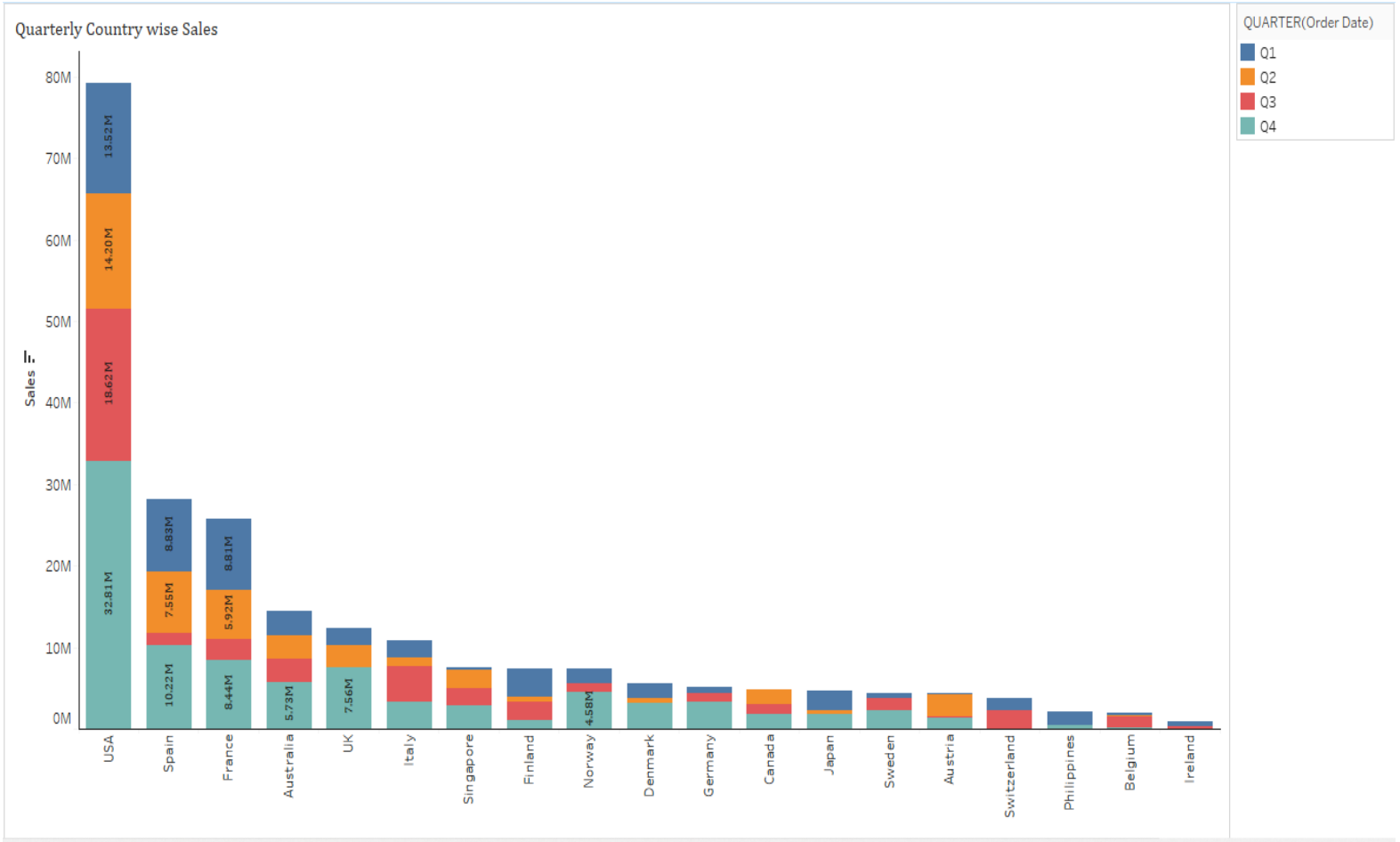
2 - Significant Drop in 2020: Across most countries, sales dropped drastically in 2020, with the USA seeing the largest drop from 37.34M in 2019 to 12.37M.

3 - Top Performers:

- In 2019, countries like France (13.58M), Spain (11.41M), and UK (7.61M) performed well compared to previous years.
- In 2020, despite the drop, France and Spain remained relatively higher compared to other countries but still saw reductions.

4 - Steady Performers: Some countries like Norway and Denmark showed stable sales across the years with slight dips in 2020.

5 - Emerging Markets: In 2020, Germany saw a surprising boost (3.53M) compared to the previous two years, where sales were much lower.



Quarterly Sales by Countries –

1 - USA Dominance:

- The USA has a huge lead over other countries with total sales nearing 80M.
- The highest sales are in Q3 (18.67M), followed by Q4 (32.91M), Q2 (14.20M), and Q1 (13.52M).
- This suggests a strong performance in the last two quarters of the year.

2 - Spain:

- Spain ranks second, with sales totalling around 30M.
- Most of its sales are concentrated in Q4 (10.22M) and Q1 (8.93M), suggesting a strong start and end to the year.

3 - France:

- France shows a similar pattern to Spain, with significant sales in Q4 (8.44M) and Q1 (8.81M).
- Q3 and Q2 contribute relatively lower figures, with Q3 being the weakest.

4 - Other European Markets (UK, Italy, Norway, etc.):

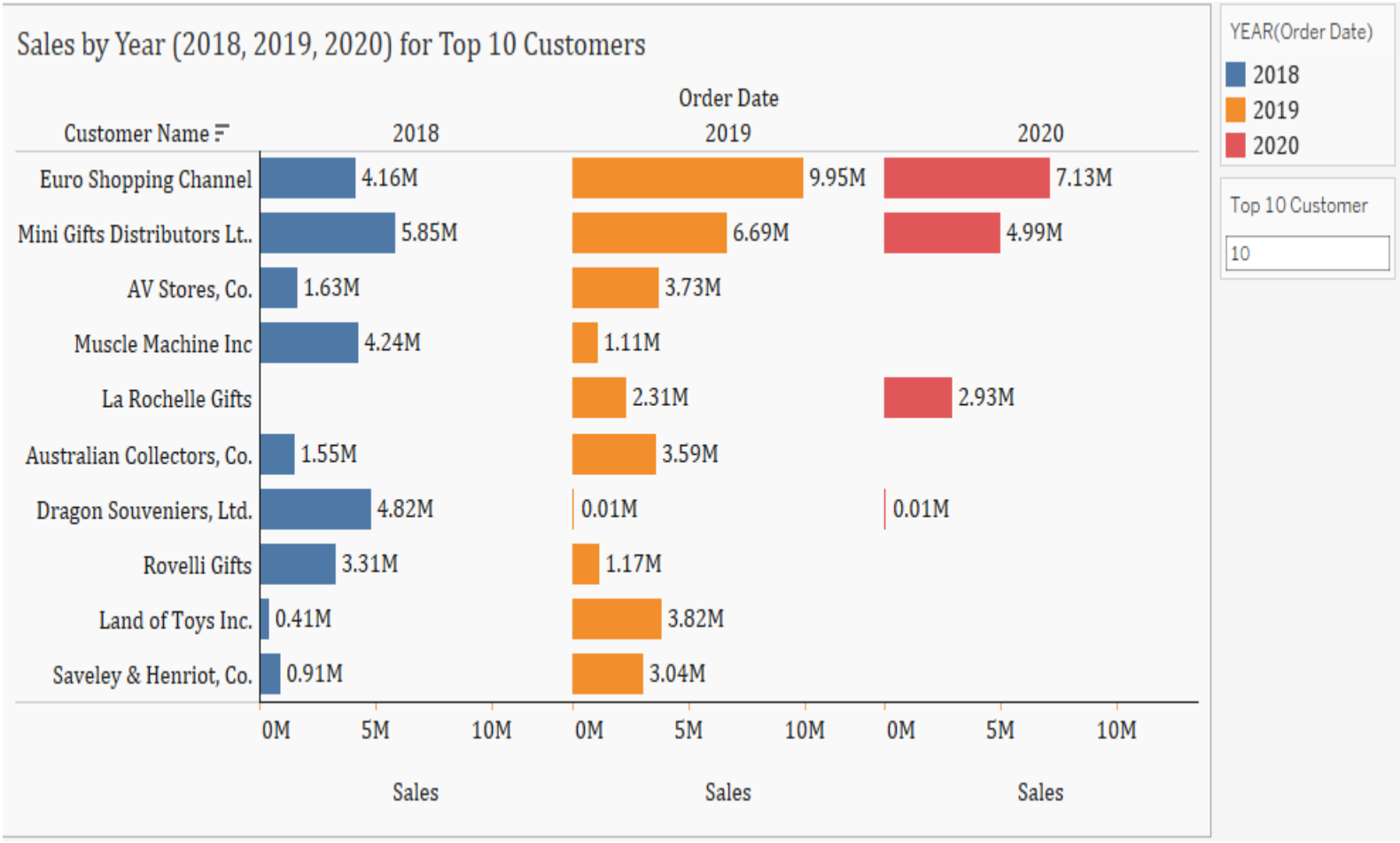
- Countries like the UK, Italy, and Finland exhibit moderate sales (ranging between 5-10M).
- UK has consistent sales across quarters, with Q4 (7.56M) being slightly higher.
- Italy's sales peak in Q4 (5.79M), with other quarters contributing evenly.

5 - Small and Stable Markets:

- Norway, Denmark, and Finland show a steady distribution of sales across the year, without any sharp spikes.
- Other countries like Switzerland, Austria, and Belgium have much lower overall sales with balanced distribution between quarters.

6 - Overall Trends:

- In general, Q4 appears to be the strongest quarter for most countries, likely due to holiday season sales.
- Q1 also sees strong performance in a few countries like the USA, Spain, and France.
- Q3 seems to be a weak quarter for most countries except the USA, where it's one of the stronger periods.



Sales by Year (2018, 2019, 2020) for Top 10 Customers -

Top Customers with Consistent Growth:

- Euro Shopping Channel has shown significant growth each year, peaking in 2020 with sales near 7.13M.
- Mini Gifts Distributors Ltd. also has steady growth, reaching its highest sales in 2020, slightly below Euro Shopping Channel.

Strong Growth in 2019 but drop in 2020:

- AV Stores, Co. and Muscle Machine Inc. had relatively high sales in 2019 but did not maintain the same level in 2020. However, Muscle Machine Inc. saw some sales in 2020, while AV Stores, Co. seems absent for 2020.
- La Rochelle Gifts saw a steady increase from 2018 to 2019 but a dip in 2020.

Irregular Sales Trends:

- Australian Collectors, Co. and Dragon Souvenirs, Ltd. show fluctuating sales. For example, Dragon Souvenirs had strong sales in both 2018 and 2019, but it's absent in 2020.
- Rovelli Gifts shows an increase in 2019 but doesn't appear in 2020.

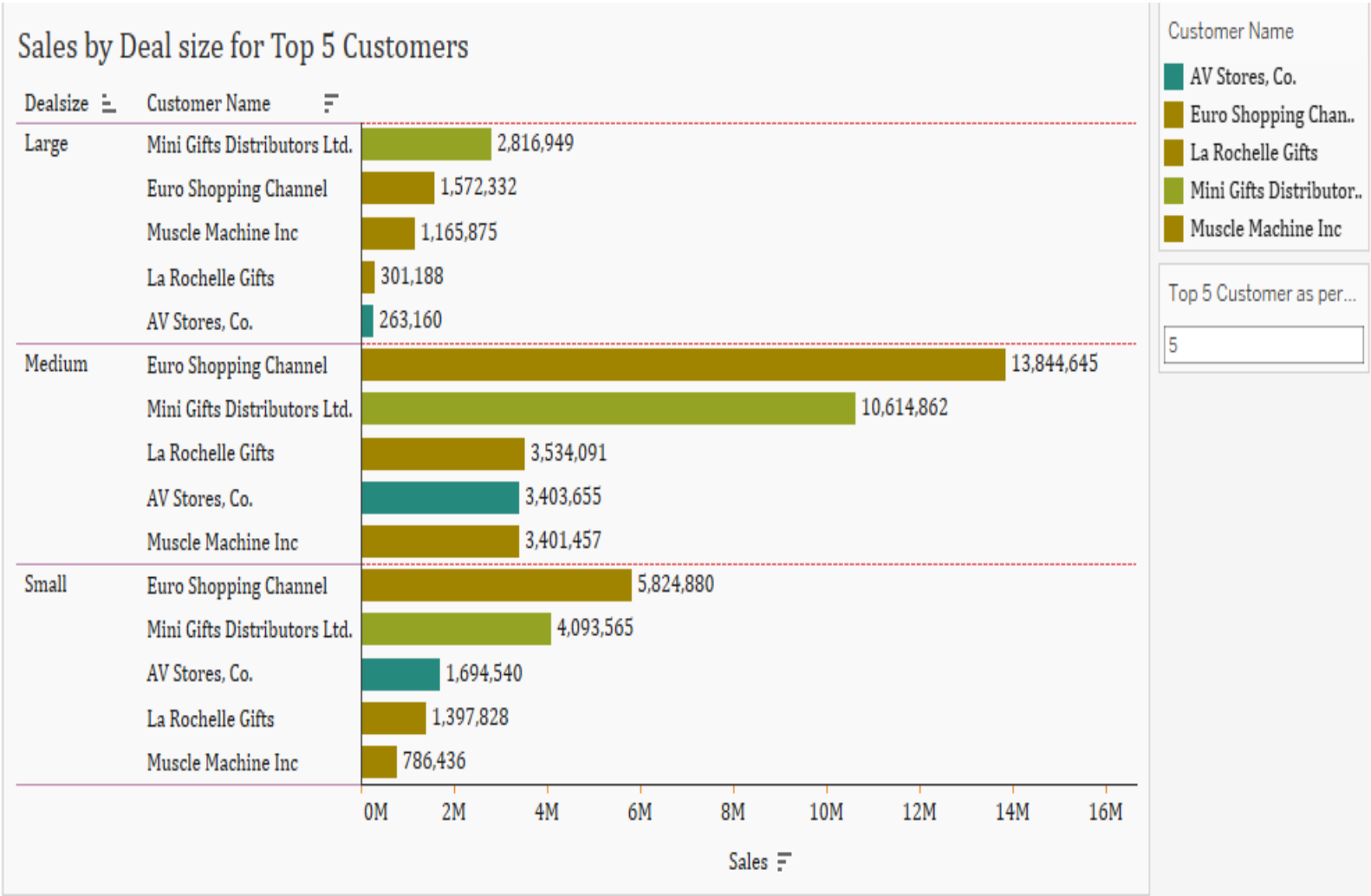
RFM/MRA Project

Absent/Declining Sales:

- Some customers, such as Land of Toys Inc. and Saveley & Henriot, Co., show very limited or no sales in 2020 despite having sales in 2018 or 2019.

General Trends:

- 2020's Top Performers: Euro Shopping Channel and Mini Gifts Distributors were dominant customers in 2020, showing resilience despite a challenging year.
- 2019 Peak for Many: Several customers, like AV Stores, Co. and Dragon Souvenirs, Ltd., reached their highest sales in 2019 but did not continue in 2020.
- Decreased Activity in 2020: A noticeable trend is that many customers either did not make purchases or had reduced activity in 2020, reflecting broader economic challenges.



Sales by Deal Size for the Top 5 Customers across three categories: Large, Medium and Small deals -

Large Deal Size:

- Mini Gifts Distributors Ltd. leads with the highest large deal sales of 2.82M.
- Euro Shopping Channel follows with 1.57M, and Muscle Machine Inc comes third with 1.17M.
- La Rochelle Gifts and AV Stores, Co. have much smaller large deals.

Medium Deal Size:

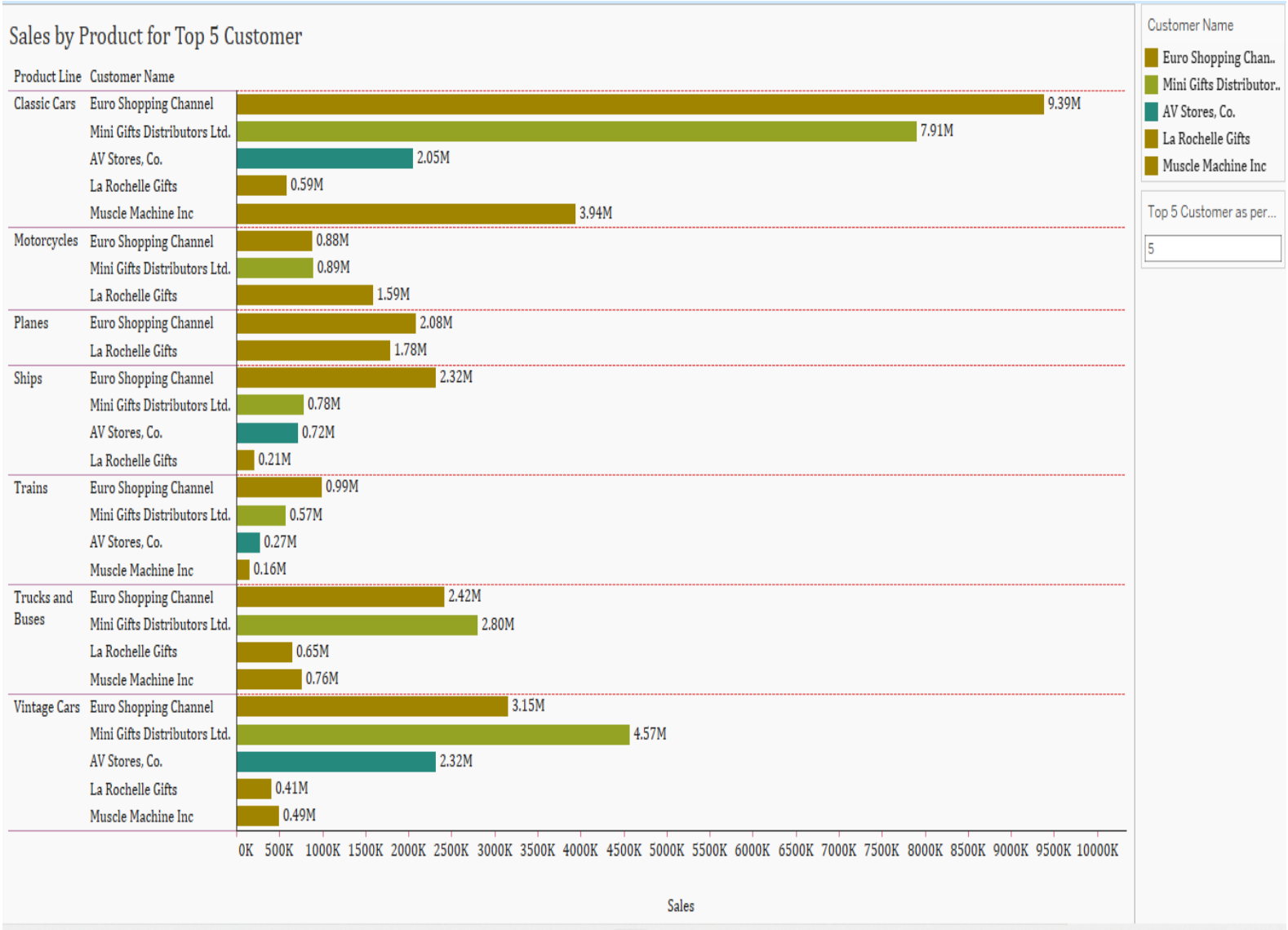
- Euro Shopping Channel has the highest sales in the medium category at 13.84M, indicating strong mid-sized transactions.
- Mini Gifts Distributors Ltd. follows with 10.61M, showing consistent performance in medium deals as well.
- La Rochelle Gifts (3.53M), AV Stores, Co. (3.40M), and Muscle Machine Inc (3.40M) have similar sales in the medium category, contributing moderately.

Small Deal Size:

- Euro Shopping Channel again leads with 5.82M in small deals, showing that it performs well across all deal sizes.
- Mini Gifts Distributors Ltd. follows closely at 4.09M, maintaining a strong presence in small deals.
- AV Stores, Co. (1.69M), La Rochelle Gifts (1.39M), and Muscle Machine Inc (786K) have lower sales in the small category but are still active.

Key Insights:

1. Euro Shopping Channel is the top customer overall, excelling in both medium and small deals, with significant large deals as well.
2. Mini Gifts Distributors Ltd. consistently ranks second across all deal sizes, with particularly strong large and medium deals.
3. Muscle Machine Inc., La Rochelle Gifts, and AV Stores, Co. are more focused on medium and small deals, with smaller contributions from large deals.
4. Medium deals represent the largest chunk of sales for most top customers, especially Euro Shopping Channel and Mini Gifts Distributors Ltd.



Sales by Product Line for the Top 5 Customers –

Classic Cars:

- Euro Shopping Channel leads with sales of 9.39M, followed by Mini Gifts Distributors Ltd. with 7.91M.
- AV Stores, Co. has moderate sales at 2.05M, while La Rochelle Gifts (0.59M) and Muscle Machine Inc (0.59M) contribute much less.

Motorcycles:

- La Rochelle Gifts is the top buyer in this category, with 1.59M in sales.
- Mini Gifts Distributors Ltd. (0.89M) and Euro Shopping Channel (0.88M) show similar contributions.
- Muscle Machine Inc and AV Stores, Co. do not appear to participate in this product line.

RFM/MRA Project

Planes:

- Euro Shopping Channel leads again with 2.08M in sales, followed closely by La Rochelle Gifts with 1.78M.
- Other customers do not seem to buy planes, highlighting the limited market in this product category.

Ships:

- Euro Shopping Channel stands out with 2.32M in sales.
- Mini Gifts Distributors Ltd. (0.78M) and AV Stores, Co. (0.72M) are far behind.
- La Rochelle Gifts has minimal involvement in this product line (0.21M).

Trains:

- Euro Shopping Channel dominates this category with 0.99M.
- Other customers, including Mini Gifts Distributors Ltd. (0.57M) and AV Stores, Co. (0.27M), contribute relatively less.
- Muscle Machine Inc has the least contribution (0.16M) in this category.

Trucks and Buses:

- Mini Gifts Distributors Ltd. is the top buyer in this category with 2.80M in sales, followed by Euro Shopping Channel (2.42M).
- La Rochelle Gifts (0.65M) and Muscle Machine Inc (0.76M) contribute moderately.

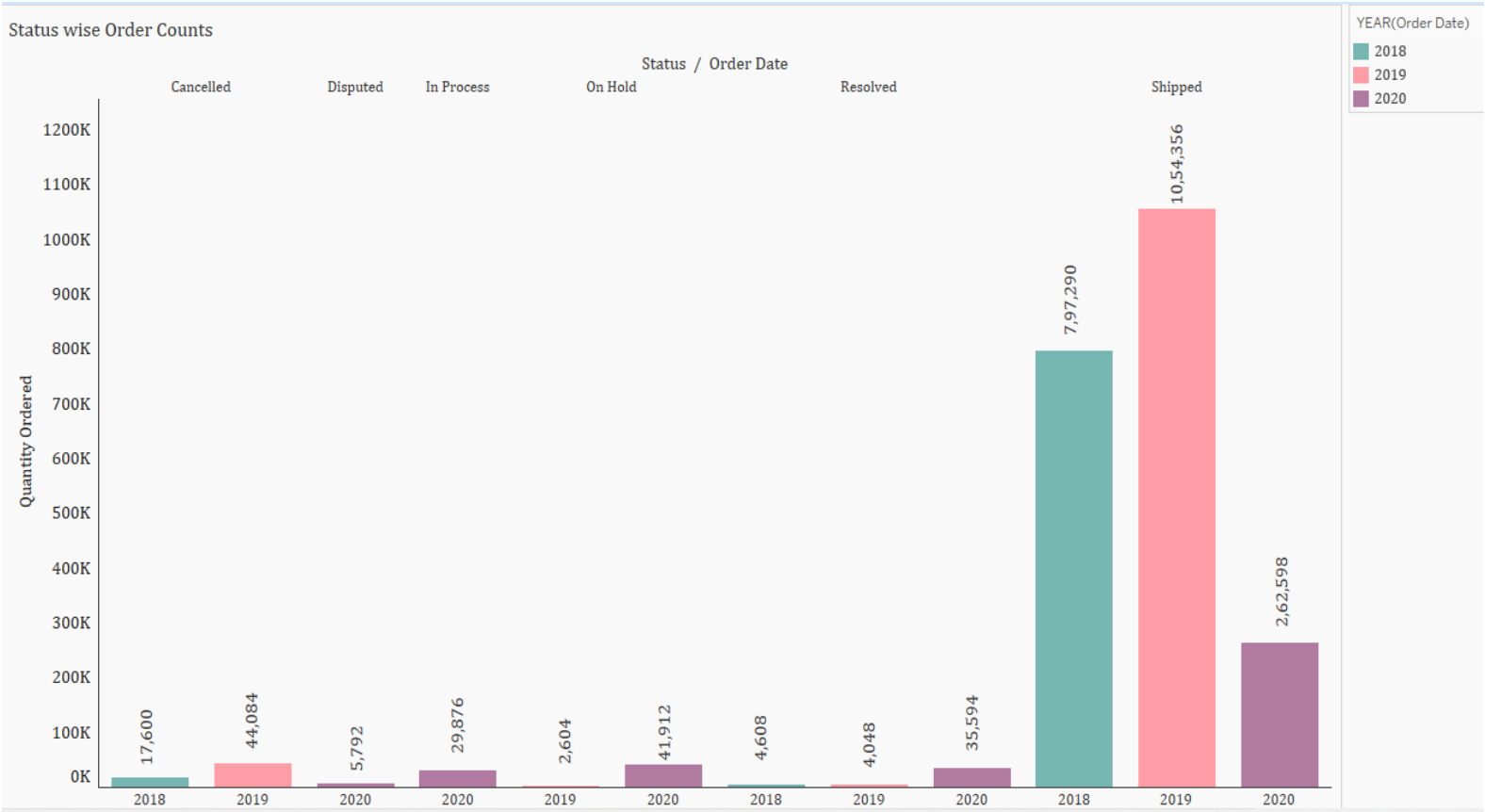
Vintage Cars:

- Mini Gifts Distributors Ltd. leads in vintage car sales with 4.57M.
- Euro Shopping Channel follows with 3.15M.
- AV Stores, Co. also makes a significant contribution at 2.32M, while other customers are less involved.

Key Insights:

1. Euro Shopping Channel is the dominant customer across most categories, especially for Classic Cars, Planes, Ships, and Trains.
2. Mini Gifts Distributors Ltd. shows strong sales across most categories, excelling particularly in Trucks and Buses, and Vintage Cars.

3. La Rochelle Gifts primarily focuses on Motorcycles and Planes, with moderate sales in other categories.
4. AV Stores, Co. focuses more on Classic Cars and Vintage Cars, but shows smaller sales in other product lines.
5. Muscle Machine Inc has smaller contributions overall, with some focus on Motorcycles, Trucks and Buses, and Vintage Cars.



Status-wise Order Counts for the years 2018, 2019, and 2020 –

1 - Shipped Orders:

- The Shipped status has the highest order quantities across all three years.
- 2019 had the highest number of shipped orders (10.54M), followed by 2018 with 7.97M, and 2020 with 2.62M.
- There's a significant drop in the quantity of shipped orders in 2020 compared to the previous years.

2 - Cancelled Orders:

- The number of cancelled orders increases over the years.
- In 2020, there were 5,792 cancelled orders, while 2019 had 44,084, and 2018 had only 17,600.
- The upward trend indicates a growing issue with cancellations over time.

3 - Disputed Orders:

- Disputed orders are only seen in 2019 and 2020, with 29,876 in 2020 and 2,604 in 2019, showing a significant increase in disputes in 2020.

4 - Orders in Process:

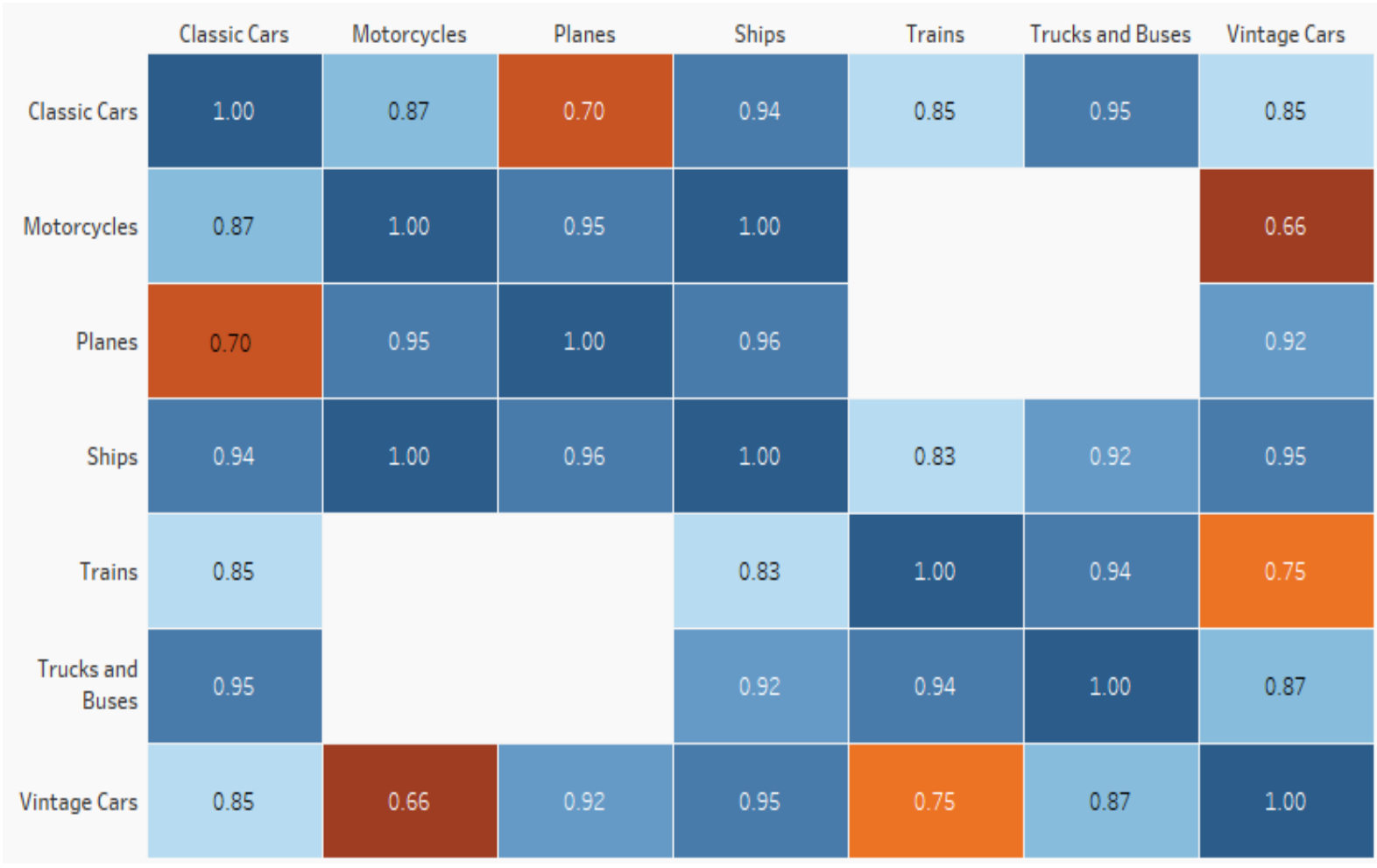
- The number of orders in process slightly increases from 2019 (2,604) to 2020 (41,912), indicating a backlog or delay in processing during 2020.

5 - On Hold Orders:

- On hold orders are only present in 2020 with 35,594, suggesting potential operational issues or delays in that year.

6 - Resolved Orders:

- Resolved orders are present across all three years, with 2020 having the highest count at 35,594, followed by 2019 (4,048) and 2018 (4,608).



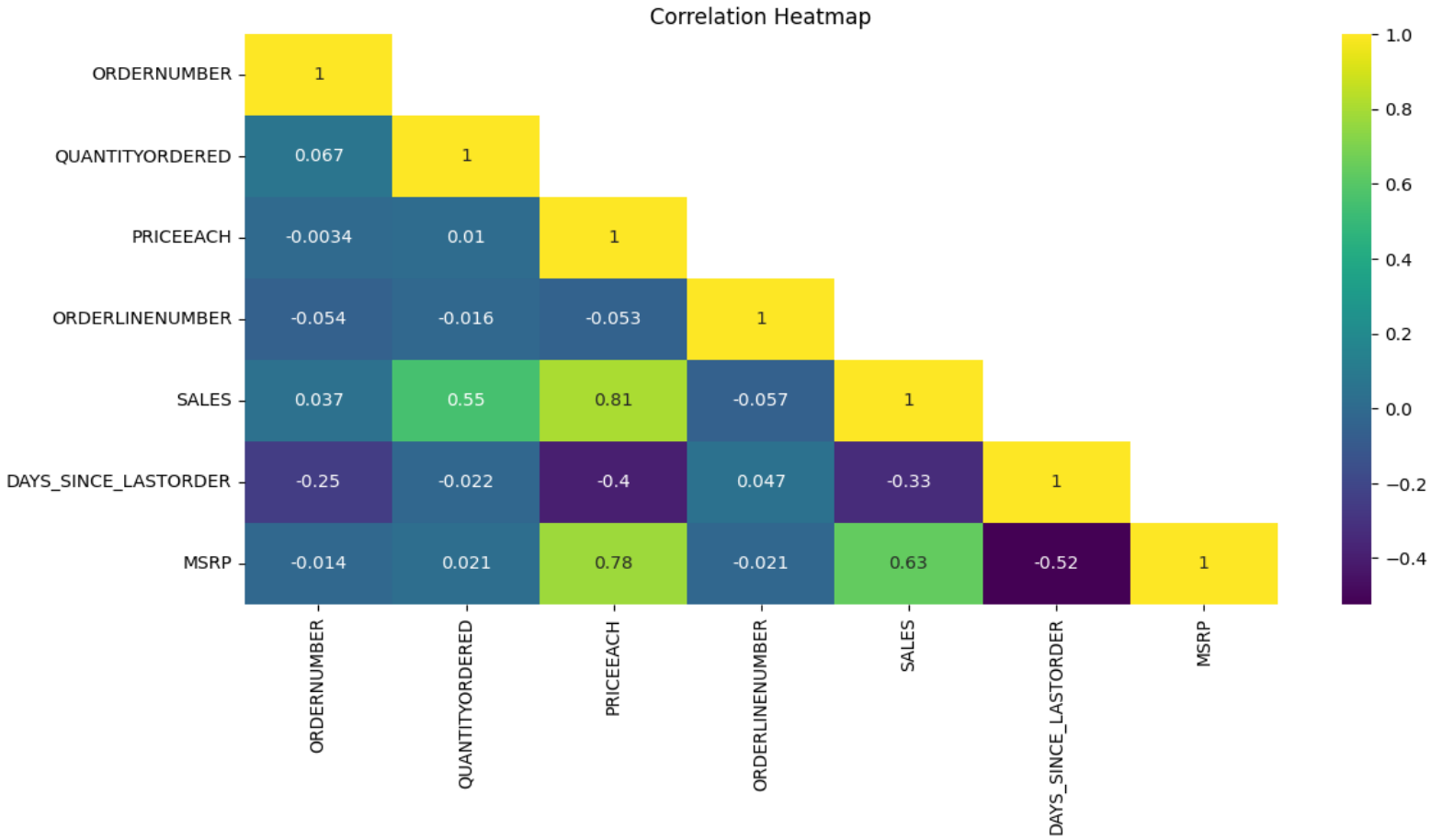
Relationships between different product lines –

1 - Strong positive correlations:

- Classic Cars and Vintage Cars have a strong positive correlation, indicating that they are often purchased together.
- Motorcycles, Ships, and Trucks and Buses also have strong positive correlations with each other, suggesting that they are frequently purchased together.

2 - Weak correlations:

- Planes have relatively weak correlations with the other product lines.
- Trains have slightly stronger correlations with Classic Cars, Trucks and Buses, and Vintage Cars, but still exhibit weak correlations overall.



1- Strong positive correlations:

- SALES and QUANTITYORDERED have a strong positive correlation, which makes sense as the more items ordered, the higher the sales will be.
- SALES and PRICEEACH also have a strong positive correlation, indicating that higher prices lead to higher sales.

- MSRP and PRICEEACH have a strong positive correlation, suggesting that the suggested retail price is closely related to the actual selling price.

2 - Strong negative correlations:

- DAYS_SINCE_LASTORDER and SALES have a strong negative correlation, implying that customers who have ordered more recently are more likely to make future purchases.

3 - Weak correlations:

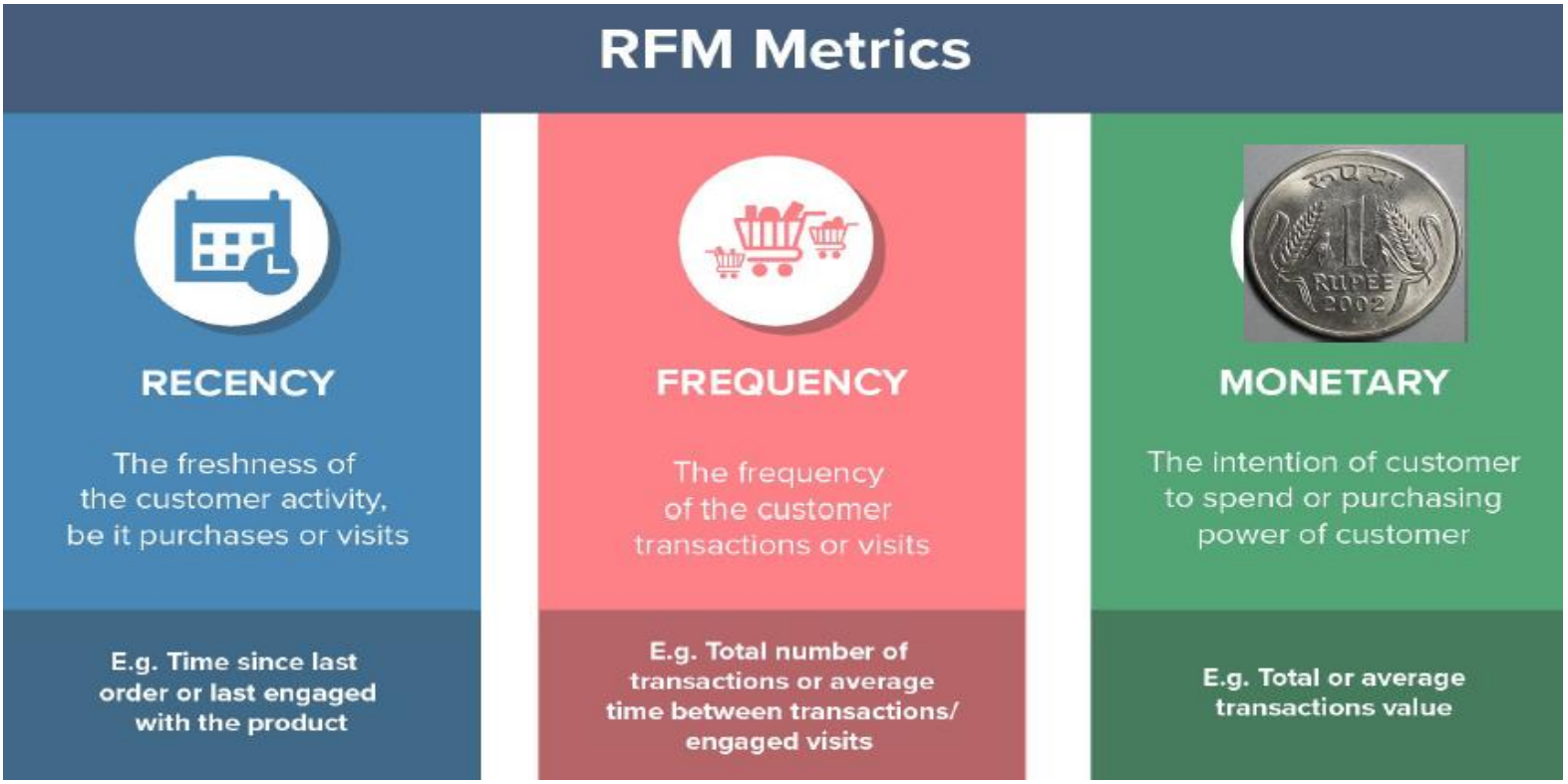
- The remaining variables have weak or no correlations with each other.

3 - Customer Segmentation using RFM Analysis –

What is RFM?

FM analysis is a customer segmentation technique that uses three key metrics to categorize customers:

- Recency: How recently a customer has made a purchase.
- Frequency: How often a customer makes purchases.
- Monetary Value: How much money a customer spends on purchases.



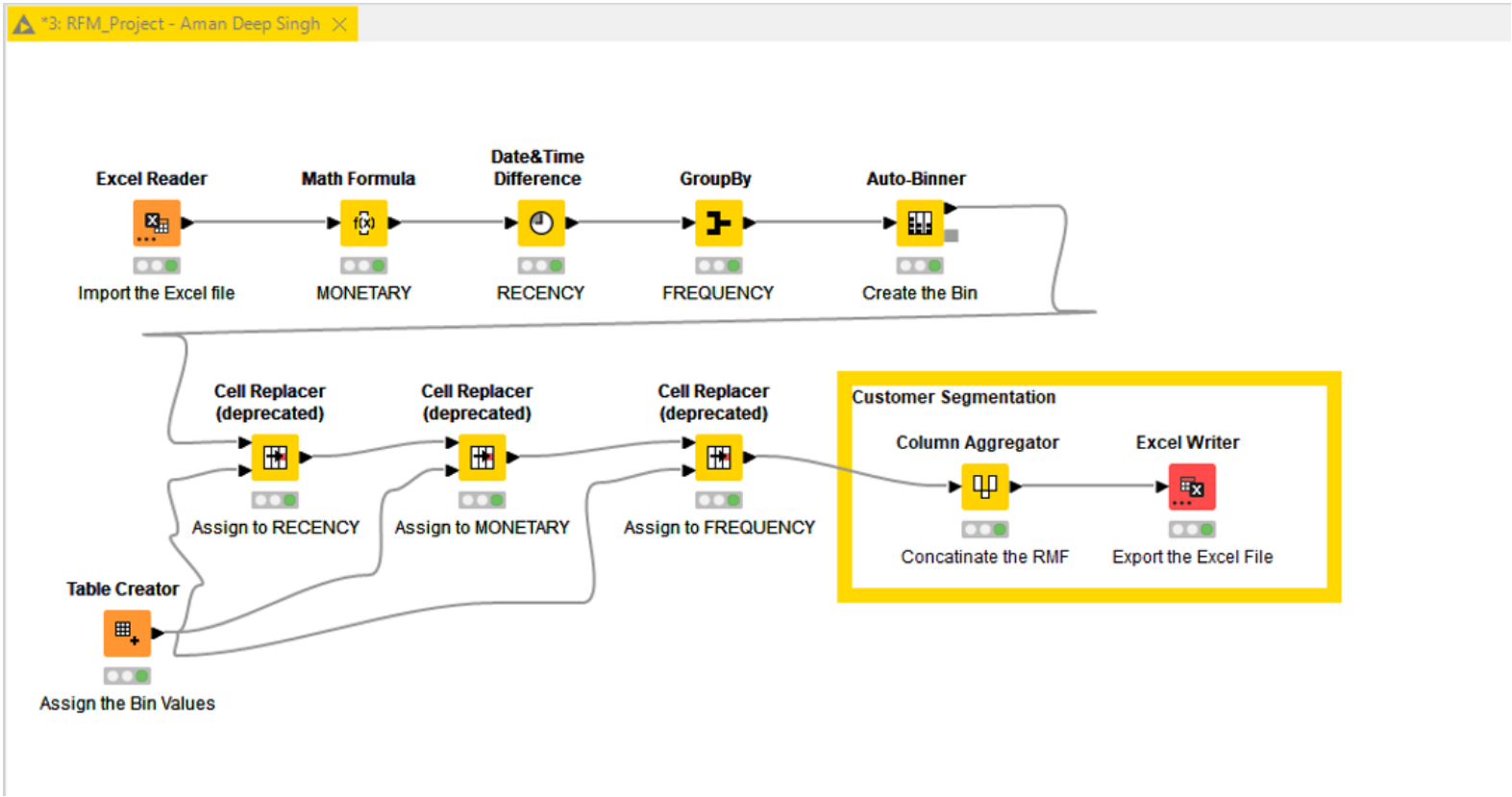
RFM/MRA Project

What all parameters used, and assumptions made?

We have used these parameters (CUSTOMERNAME, QUANTITYORDERED, PRICEEACH, ORDERDATE). Using RFM the customers are divided into 4 segments namely, 4 (High valued), 3 (Mid valued), 2 (At Risk / Potential) and 1(Churn). KNIME workflow was used for the same.

It is assumed that the high and mid valued customers are loyal. The at-risk customers might have some potential left and they could be converted to mid valued by different promotional schemes.

Showcase the KNIME workflow image



What results are there in the output table head?

```
Index(['CUSTOMERNAME', 'QUANTITYORDERED', 'PRICEEACH', 'ORDERLINENUMBER',
       'SALES', 'ORDERDATE', 'DAYS_SINCE_LASTORDER', 'STATUS', 'PRODUCTLINE',
       'MSRP', 'PRODUCTCODE', 'PHONE', 'ADDRESSLINE1', 'CITY', 'POSTALCODE',
       'COUNTRY', 'CONTACTLASTNAME', 'CONTACTFIRSTNAME', 'DEALSIZE',
       'MONETARY', 'RECENCY', 'ORDERNUMBER', 'MONETARY [Binned]',
       'RECENCY [Binned]', 'ORDERNUMBER [Binned]', 'RECENCY_HML',
       'MONETARY_HML', 'FREQUENCY_HML', 'RFMScore'],
      dtype='object')
```

RFM/MRA Project

RFM Results –

	CUSTOMERNAME	RFMScore
0	AV Stores, Co.	242
1	Alpha Cognac	412
2	Amica Models & Co.	131
3	Anna's Decorations, Ltd	344
4	Atelier graphique	212
5	Australian Collectables, Ltd	412
6	Australian Collectors, Co.	344
7	Australian Gift Network, Co	312
8	Auto Assoc. & Cie.	111
9	Auto Canal Petit	432

4 - Inferences from RFM Analysis and identified segments

Who are your best customers?

	CUSTOMERNAME	RFMScore	MONETARY	ORDERNUMBER	RECENCY
32	Euro Shopping Channel	444	912294.110000	26	0
53	Mini Gifts Distributors Ltd.	444	654858.060000	17	2
43	La Rochelle Gifts	444	180124.900000	4	0
78	The Sharp Gifts Warehouse	444	160010.270000	4	39
72	Souveniers And Things Co.	444	151570.980000	4	2

Which customers are on the verge of churning?

	CUSTOMERNAME	RFMScore	MONETARY	ORDERNUMBER	RECENCY
44	Land of Toys Inc.	244	164069.440000	4	198
0	AV Stores, Co.	242	157807.810000	3	196
64	Rovelli Gifts	242	137955.720000	3	201
57	Online Diecast Creations Co.	242	131685.300000	3	209
23	Corrida Auto Replicas, Ltd	242	120615.280000	3	212

Who are your lost customers?

	CUSTOMERNAME	RFMScore	MONETARY	ORDERNUMBER	RECENCY
29	Double Decker Gift Stores, Ltd	111	36019.040000	2	495
87	West Coast Collectables Co.	111	46084.640000	2	488
70	Signal Collectibles Ltd.	111	50218.510000	2	476
25	Daedalus Designs Imports	111	69052.410000	2	465
15	CAF Imports	111	49642.050000	2	439

Who are your loyal customers?

	CUSTOMERNAME	RFMScore	MONETARY	ORDERNUMBER	RECENCY
32	Euro Shopping Channel	444	912294.110000	26	0
53	Mini Gifts Distributors Ltd.	444	654858.060000	17	2
6	Australian Collectors, Co.	344	200995.410000	5	184
55	Muscle Machine Inc	344	197736.940000	4	182
43	La Rochelle Gifts	444	180124.900000	4	0

PROJECT RFM END

RFM/MRA Project

MRA Project

Agenda

Problem Statement

Exploratory Analysis

Market Basket Analysis using KNIME

Inferences & Recommendations

MRA Project Tableau Link –

https://public.tableau.com/views/MRAProject_17282260327100/Year-QuarterProductwiseOrderCounts?:language=en-US&:sid=&:redirect=auth&:display_count=n&:origin=viz_share_link

Problem Statement:

An automobile parts manufacturing company has collected data on transactions for 3 years. They do not have any in-house data science team, thus they have hired you as their consultant. Your job is to use your data science skills to find the underlying buying patterns of the customers, provide the company with suitable insights about their customers, and recommend customized marketing strategies for different segments of customers.

1 - Exploratory Analysis -->

- Exploratory Analysis of data & an executive summary (in PPT) of your top findings, supported by graphs.
- Are there trends across months/years/quarters/days etc. that you are able to notice?

2 - Use of Market Basket Analysis (Association Rules) -->

- Write Something about the association rules and its relevance in this case
- Add KNIME workflow image
- Write about threshold values of Support and Confidence

3 - Associations Identified -->

- Put the associations in a tabular manner
- Explain about support, confidence, & lift values that are calculated.

4 - Suggestion of Possible Combos with Lucrative Offers -->

- Write recommendations
- Make discount offers or combos (or buy two get one free) based on the associations and your experience

Exploratory Analysis -

About Data

There are 20641rows and 3 Columns in Dataset Provided

```
Number of Rows: 20641
Number of Columns: 3
```

There are no missing values in dataset

	0
Date	0
Order_id	0
Product	0

dtype: int64

There are about 4730 duplicate rows which are dropped

```
4730
```

There are 1139 unique Transactions and 37 unique Product

```
37
```

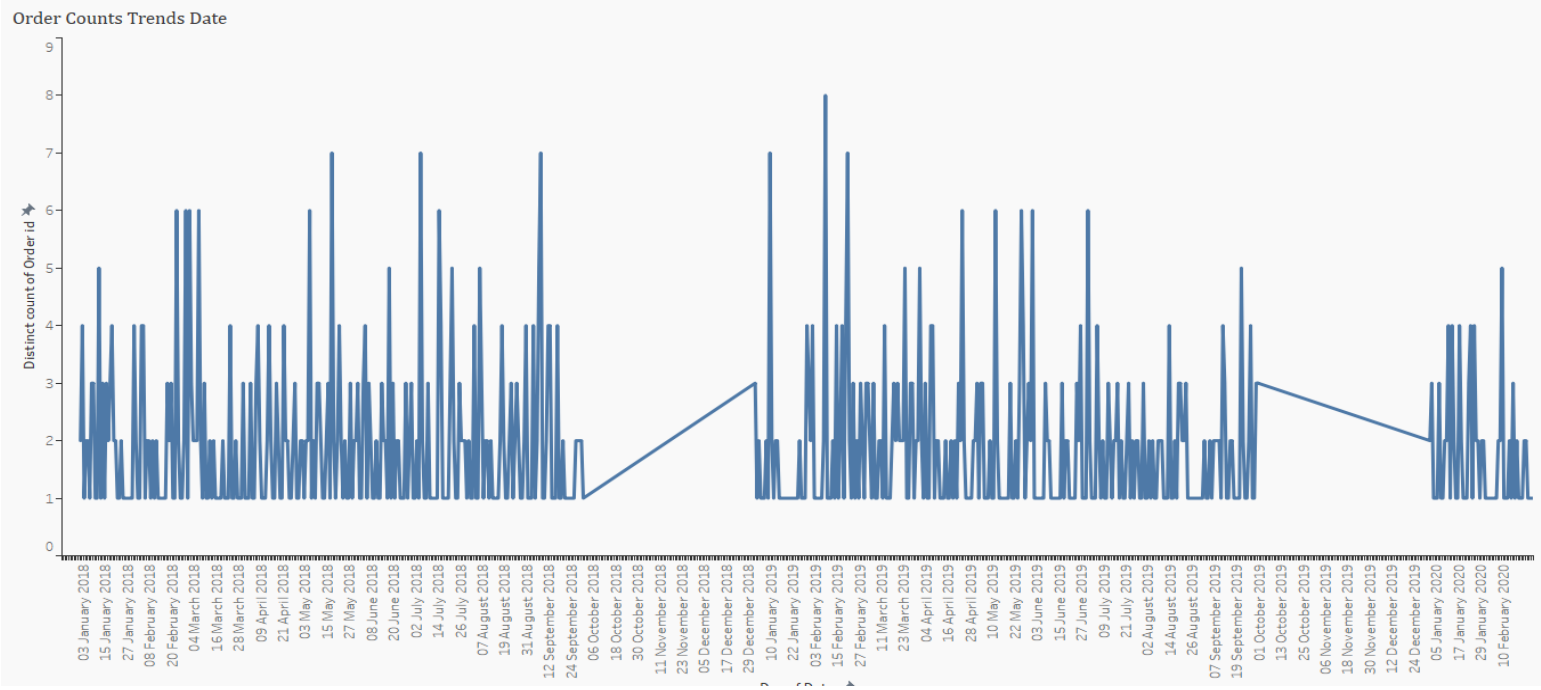
```
data[
```

```
1139
```

The Data is for 3 years 2018, 2019 and 2020

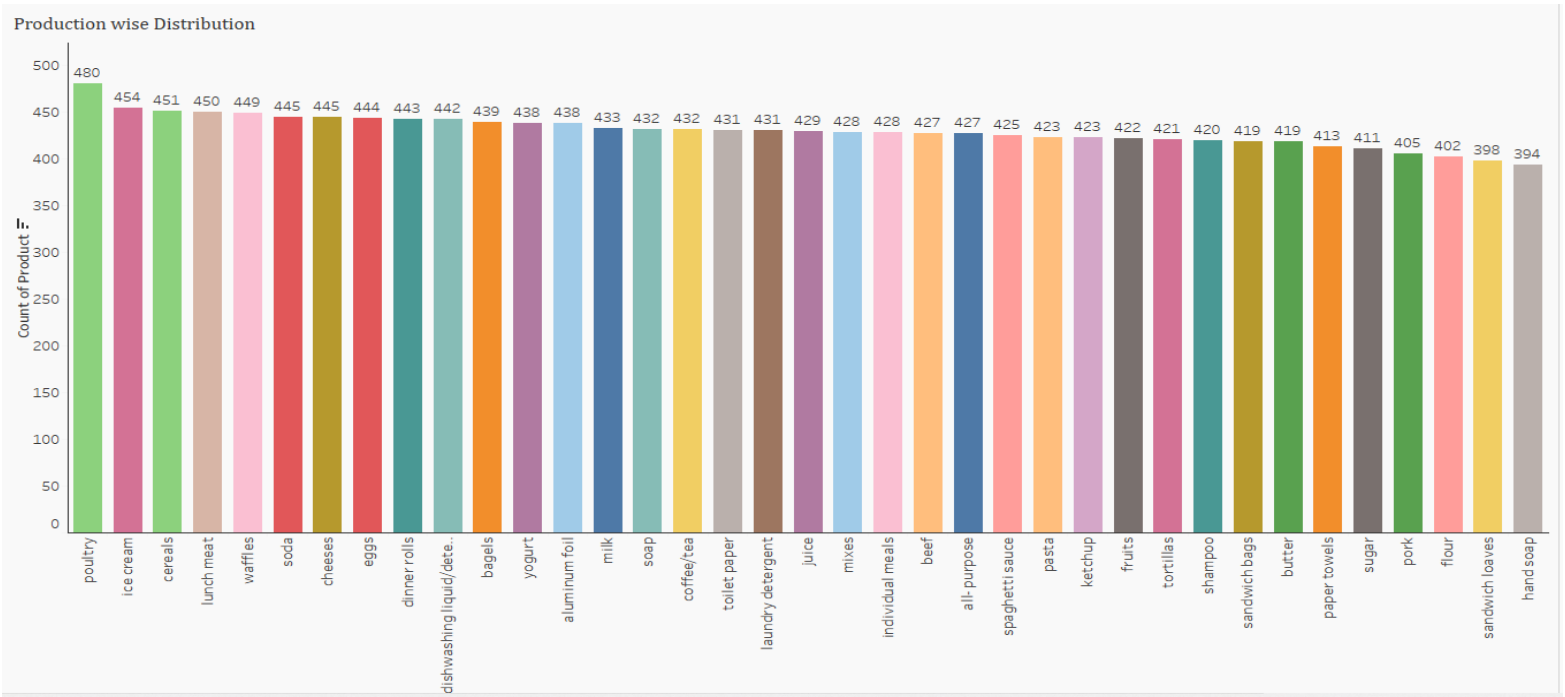
The Data is for 9 months 2018, 9 months of 2019 and 2 months of 2020

The Start Date is 1st Jan 2018 and End Date is 26th Feb 2020.



Order Counts Trend Date

- Data set has seasonality as purchase is high for couple of months.
- No data points from October till December for both 2018 and 2019. Only two months data available for the year 2020.



Production wise Distribution –

RFM/MRA Project

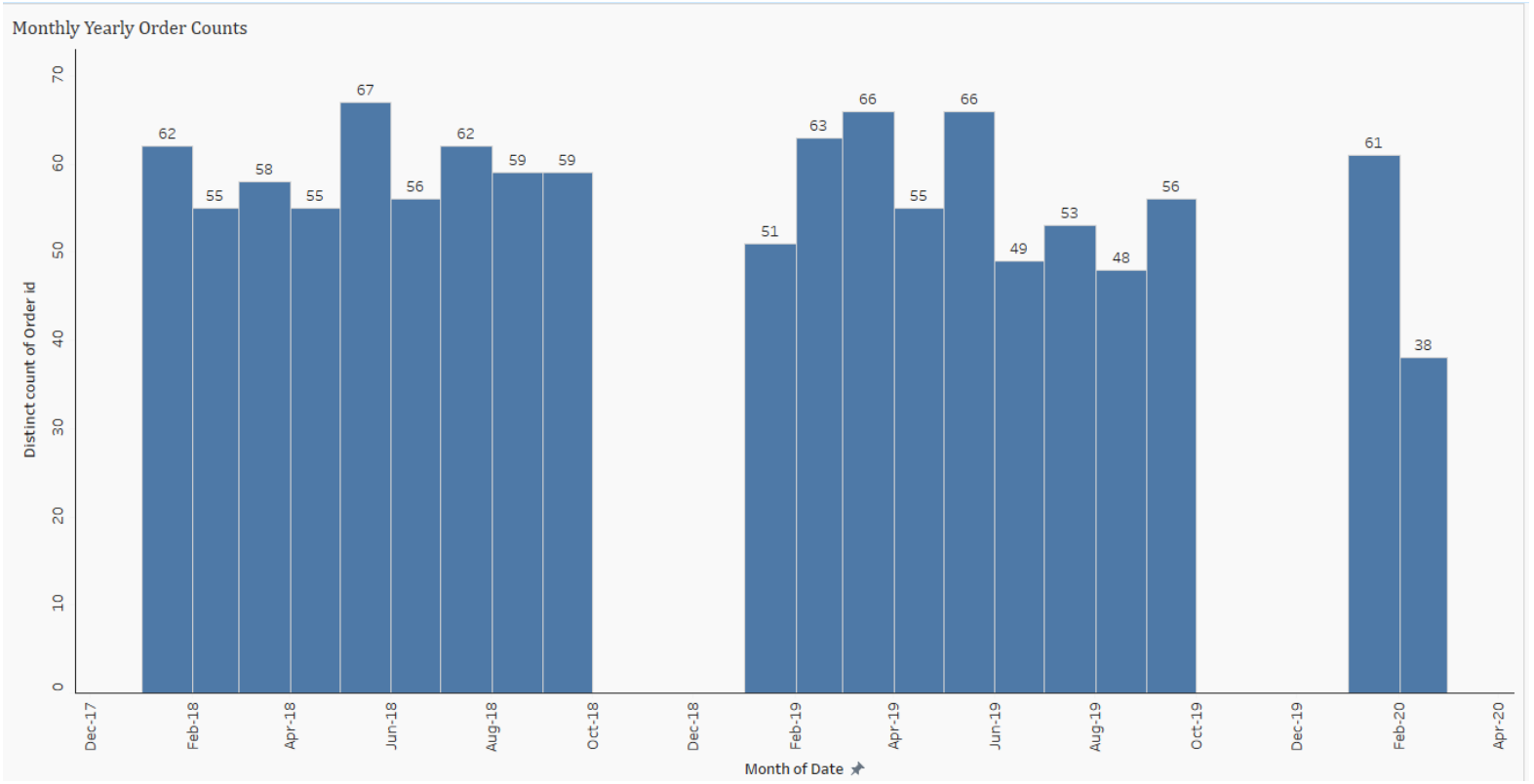
1- **Poultry** has the highest count of products at **454**, indicating it is the most frequently produced product in this dataset.

2 - **Ice cream** and **cereals** follow closely with counts of **451** and **450**, showing they are also very frequently produced items.

3 - **Hand soap** has the lowest count of products, with only **394**, making it the least frequently produced item in the dataset.

4 - The product distribution is fairly consistent, with most products having counts ranging from **400 to 450**, indicating that there isn't a very large variance in production across most product types.

5 - Products like **ketchup**, **pasta**, and **flour** are around the middle in terms of production frequency, suggesting balanced demand or production capacity across these categories.



Monthly Yearly Order Counts -

1 - **Consistent Order Activity**: The number of orders remains relatively stable month-to-month, with most months ranging between 50 to 67 orders.

2 - **Peaks in Orders**:

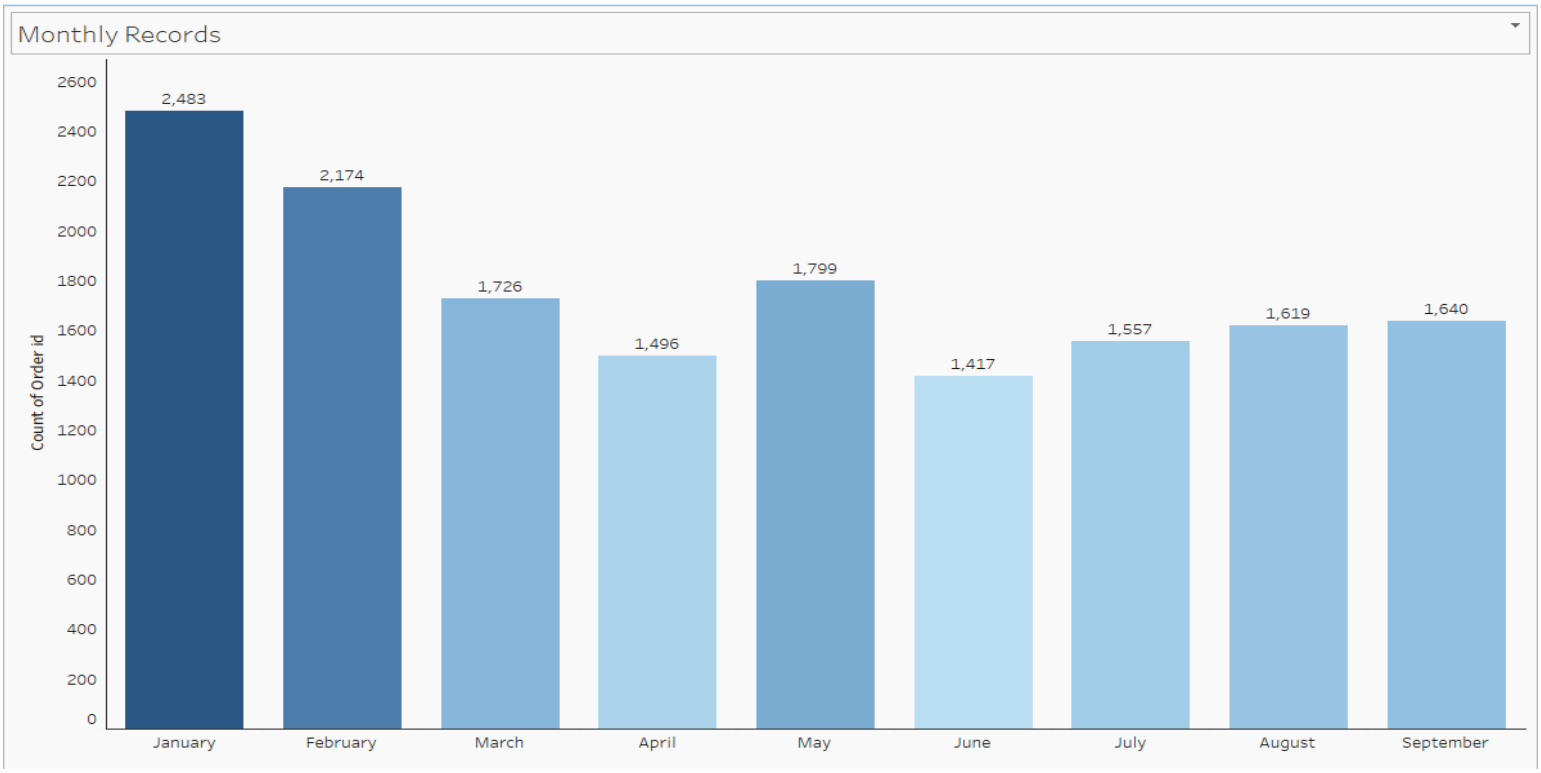
- The highest order count is observed in **July 2018** (67 orders).
- There is also a notable peak in **September 2019** (66 orders).

3 - **Drops in Order Counts**:

- There are drops in **February 2019** (51 orders) and **February 2020** (49 orders), indicating a potential seasonal decline during these months.
- The lowest count is seen in **April 2020** with **38 orders**, which could suggest a significant decline, potentially related to external factors such as market conditions or disruptions.

4 - **Overall Trends**:

- Each year shows a general fluctuation in order activity, but it tends to stabilize towards the end of the year.
- There might be periodic patterns or seasonality in the data, as February in both 2019 and 2020 shows lower activity, while other months like July and September tend to have higher order volumes.



Monthly Records -

1 - **Highest Order Count in January**:

- January leads with **2,483** orders, indicating it is the month with the highest customer activity.

2 - **Consistent Drop After January**:

- A noticeable decrease in orders is seen in February (**2,174**) and continues to decrease steadily through March and April, reaching the lowest point in April with **1,496** orders.

3 - **Recovery in May**:

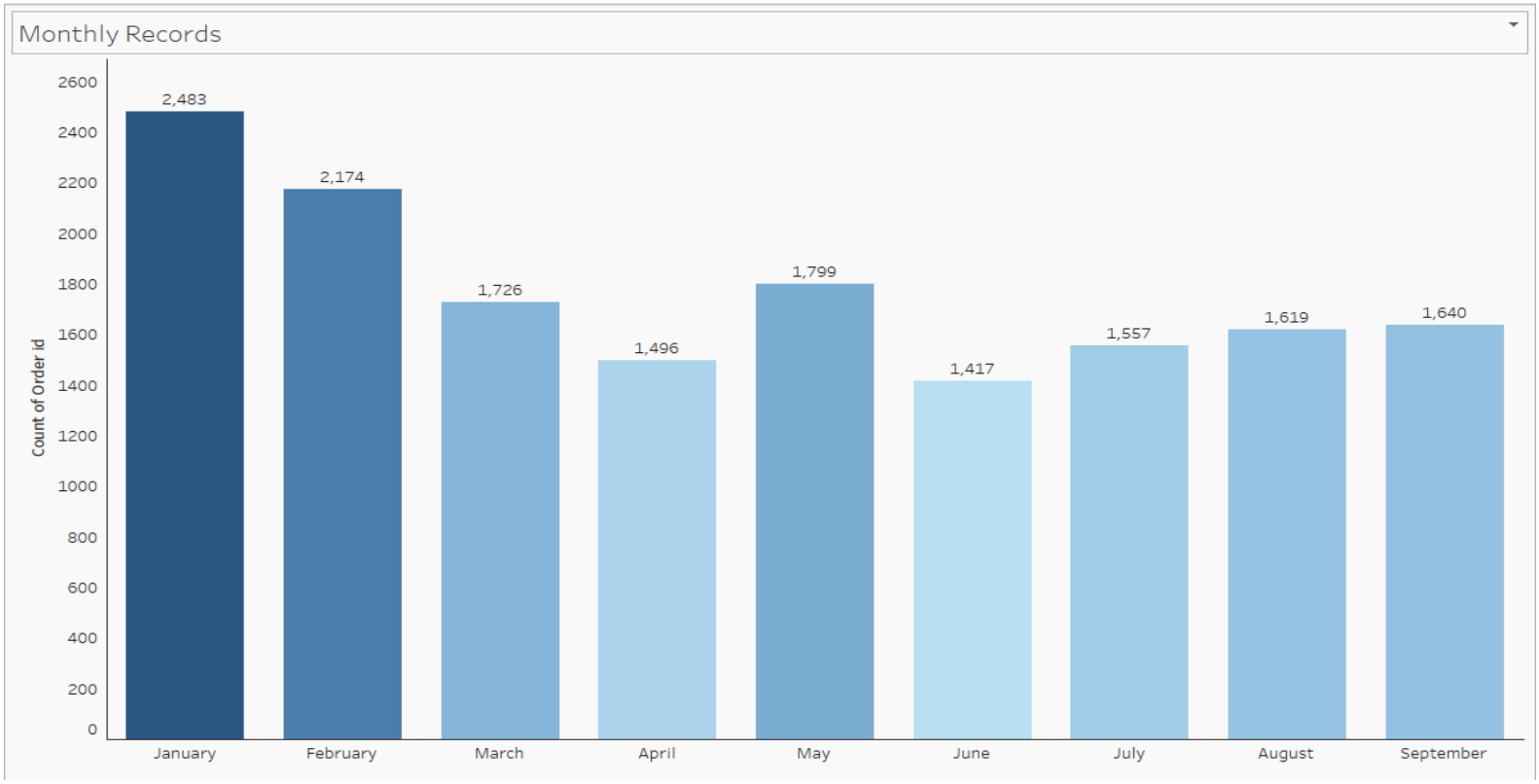
- There is a sharp increase in May (**1,799** orders) after the April drop, indicating a resurgence in order activity.

4 - **Stable Activity in Summer Months**:

- June to September show a relatively stable range of order counts, with small variations between months:
    - **June**: 1,417 orders
    - **July**: 1,557 orders
    - **August**: 1,619 orders
    - **September**: 1,640 orders

5 - **Seasonal Trends**:

- The chart suggests that the beginning of the year (January) is the busiest, while mid-year months experience more stable but lower order volumes.



Monthly wise Counts Records -

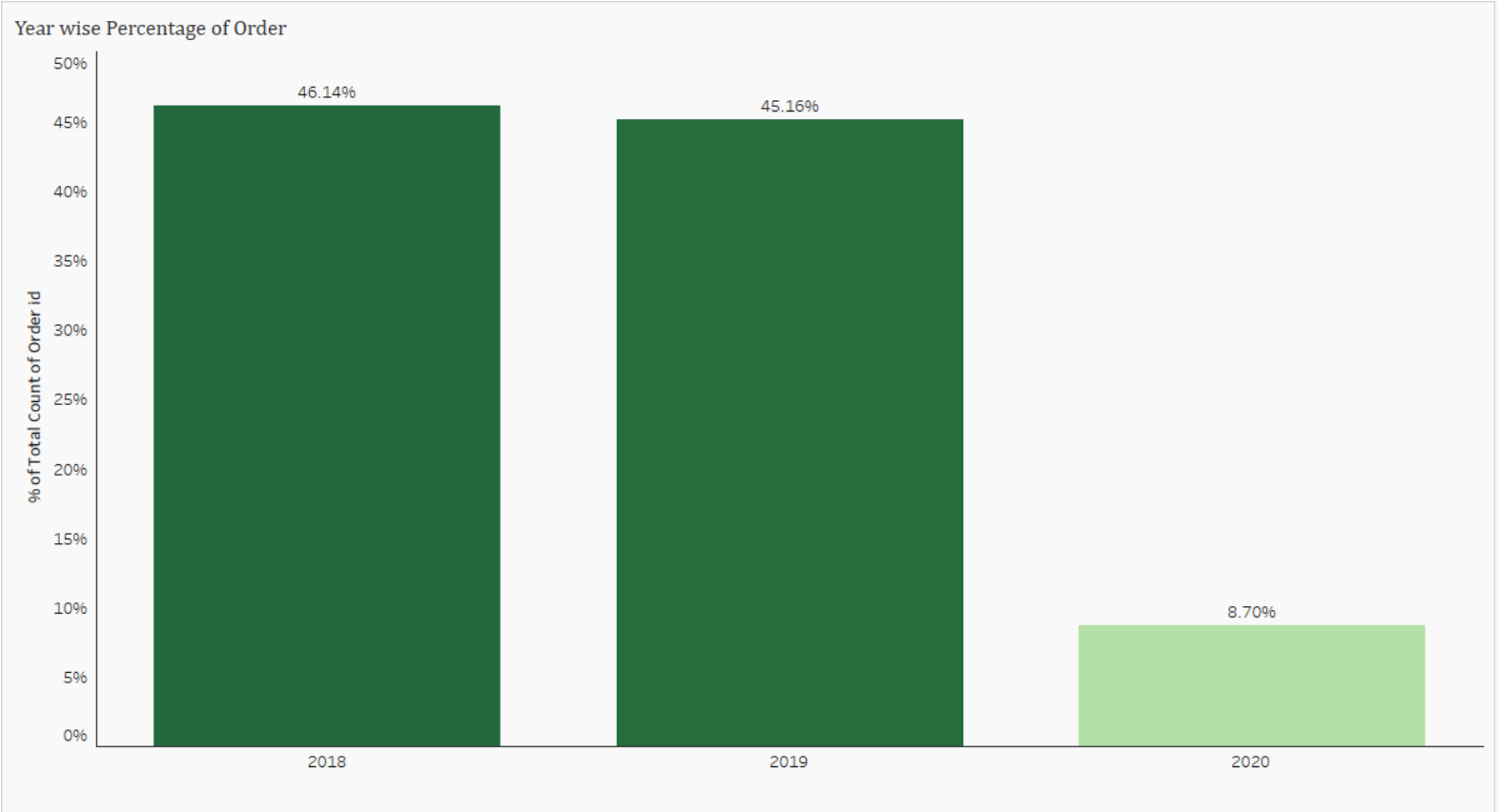
1 - Highest Order Count: January has the highest number of orders at 2,483, followed closely by February with 2,174 orders. This indicates a strong start to the year.

2 - Declining Orders: Following the peak in February, there is a notable decline in order counts for March (1,726) and subsequent months, reaching the lowest point in June with only 1,417 orders.

3 - Gradual Recovery: After June, there is a gradual upward trend in order counts, with July, August, and September showing increases, albeit still below February's peak. September has 1,640 orders, which is a recovery from the low in June.

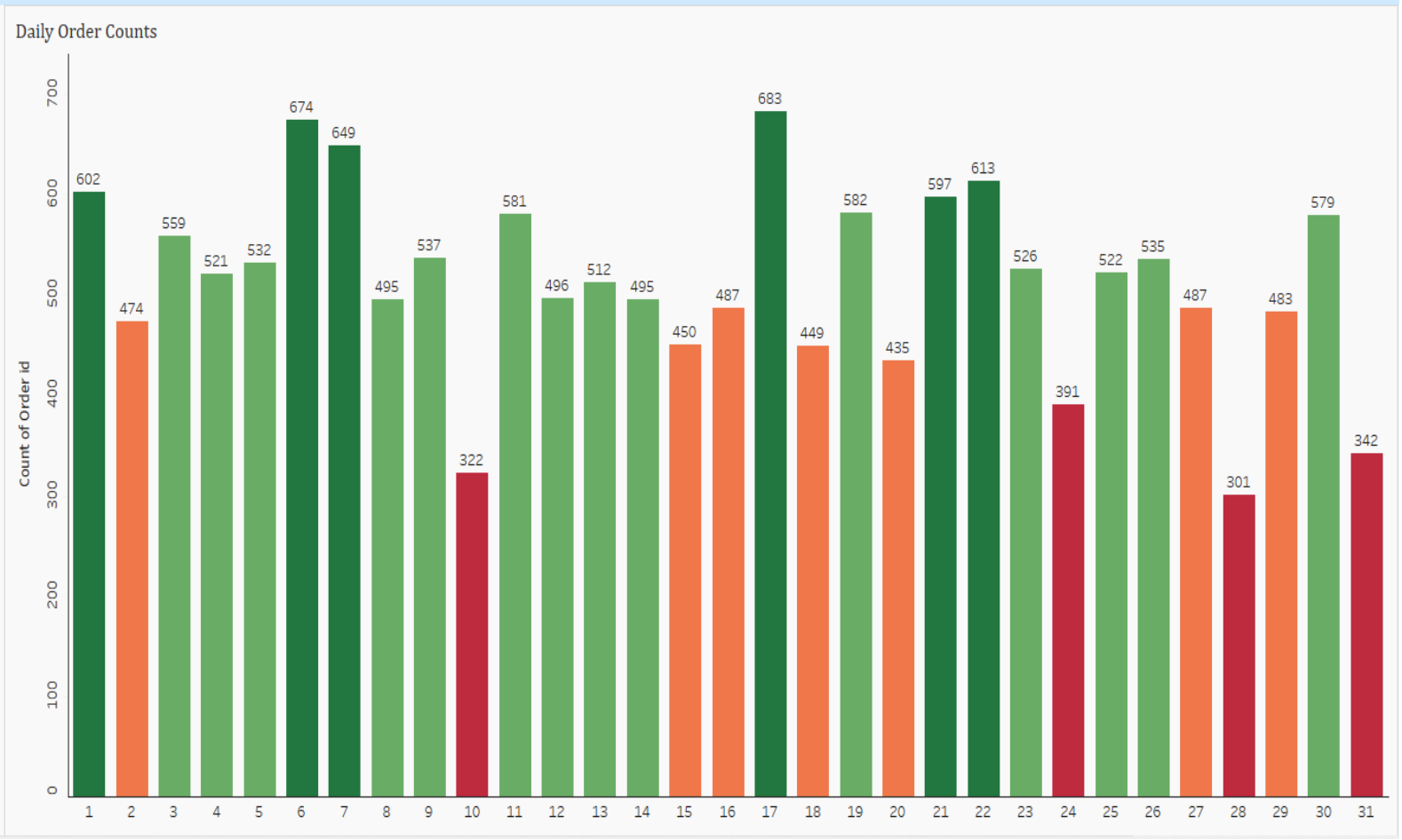
4 - Seasonality Considerations: The initial high numbers could suggest seasonal trends, possibly due to sales or promotions early in the year, while the later months may represent a levelling off as the year progresses.

5 - Volatility in Demand: The chart highlights fluctuations in demand throughout the months, indicating varying consumer behaviour or external factors impacting sales.



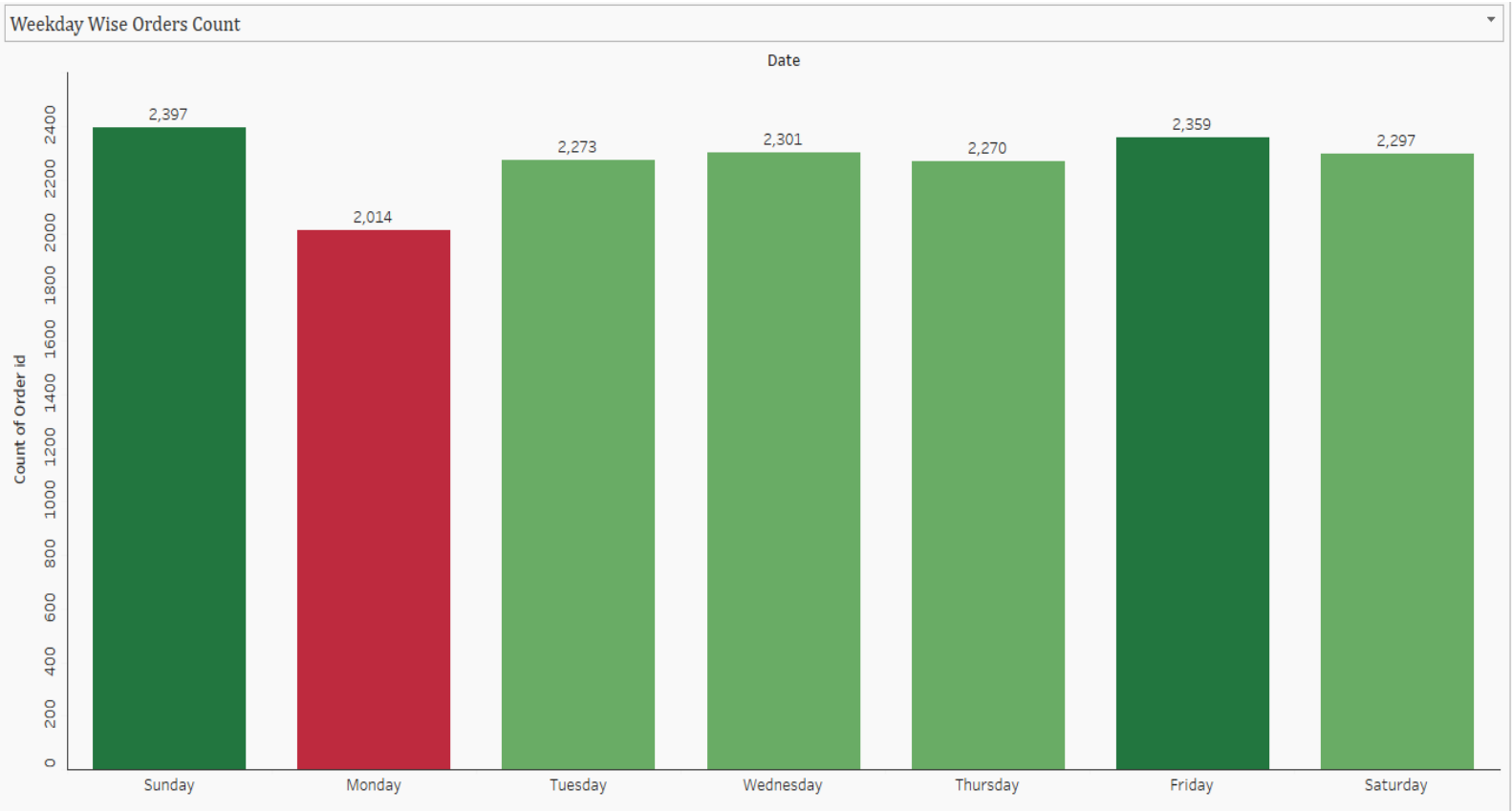
Year-wise Percentage of total Order counts

1. **Stable Performance (2018-2019)**: The percentages for 2018 (46.14%) and 2019 (45.16%) are relatively high and nearly consistent, indicating stable order counts for these years.
2. **Dramatic Decline in 2020**: There is a significant drop in the percentage of total orders in 2020, with only 8.70%. This suggests a drastic decline in orders, possibly due to external factors such as economic downturns, changes in consumer behaviour, or disruptions caused by events such as the COVID-19 pandemic.
3. **Overall Trend**: The data reflects a concerning trend where the count of orders significantly decreased in 2020 compared to the previous two years, highlighting a potential area of concern for the business or industry being analysed



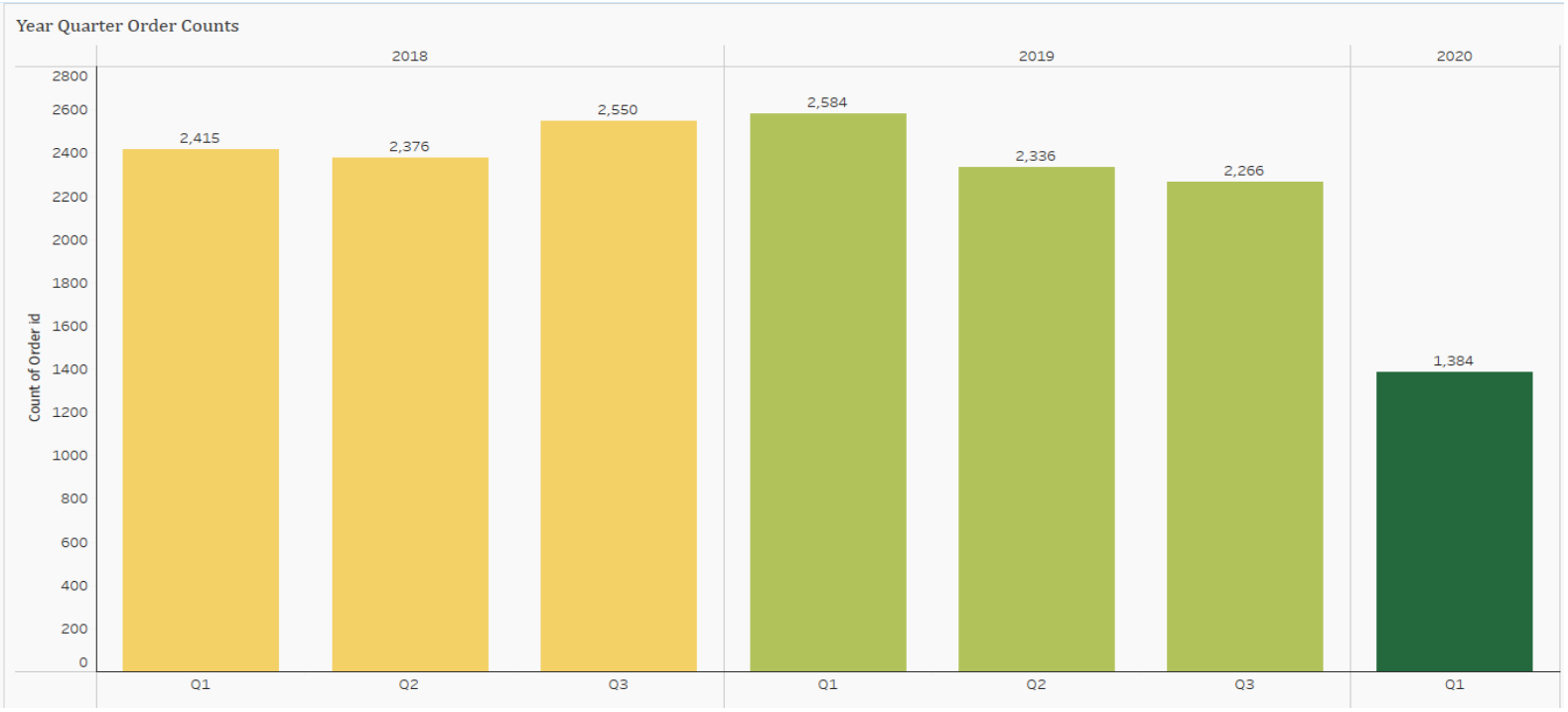
Daily Order Counts –

1. **Peak Days:** The highest order counts are observed on days 8 (674), 9 (649), and 17 (683), indicating these days had significantly more orders compared to others.
2. **Lowest Count**: Day 10 shows a stark drop in orders, with only 322 counted, which is considerably lower than surrounding days. This could be a point of interest for investigation.
3. **Trends:** The order counts fluctuate considerably through the month, with several peaks and troughs. Days 1, 3, and 8 demonstrate strong performance, while days 10, 23, and 28 show lower counts.
4. **General Pattern:** There seems to be a general rise and fall in order counts across the month with certain days consistently performing better, suggesting possible seasonality or promotional events affecting orders.
5. **Average Counts:** Most days hover around 450-600 orders, except notably low days like 10 and 28.
6. **Comparative Analysis:** It may be useful to compare the performance of specific days to assess any correlations with events or marketing strategies.



Weekday Wise Orders Count -

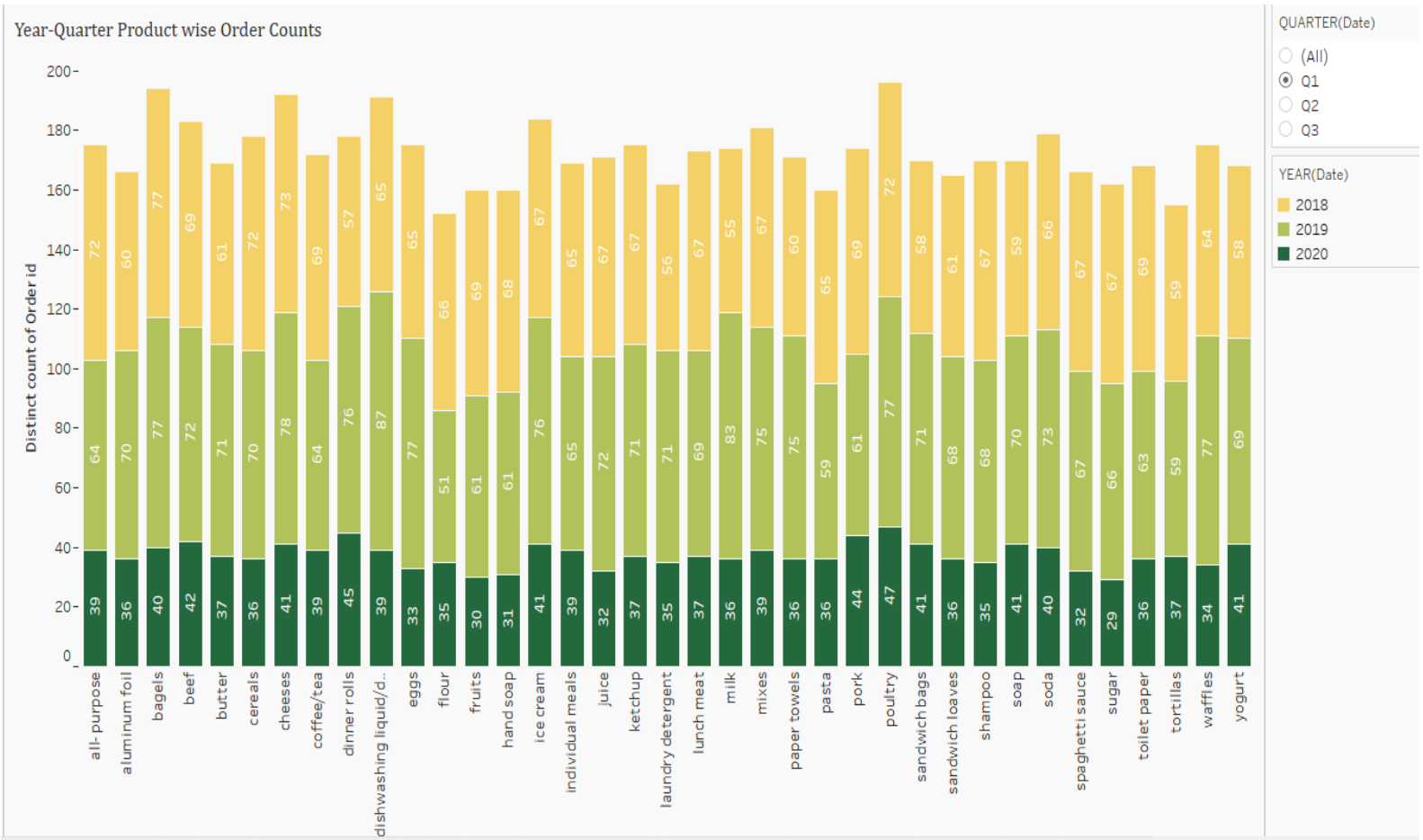
1. **Overall Trends:**
    - Sunday has the highest order count at 2,397.
    - Monday has the lowest order count at 2,014, significantly lower than the other days.
2. **Comparative Analysis:**
    - Tuesday and Wednesday have similar counts, with 2,273 and 2,301, respectively.
    - Thursday also maintains a relatively high order count at 2,270.
    - Friday shows an elevation in orders, reaching 2,359, while Saturday has 2,297, indicating consistent demand throughout the weekend.
3. **Observations on Monday:**
    - The lower count on Monday (highlighted in red) suggests it could be a slower day for orders, which may warrant further investigation to understand customer behaviour or operational efficiencies.
4. **Potential Areas for Focus:**
    - The significant drop in orders from Sunday to Monday signals a potential opportunity for promotions or marketing strategies to boost Monday's sales.



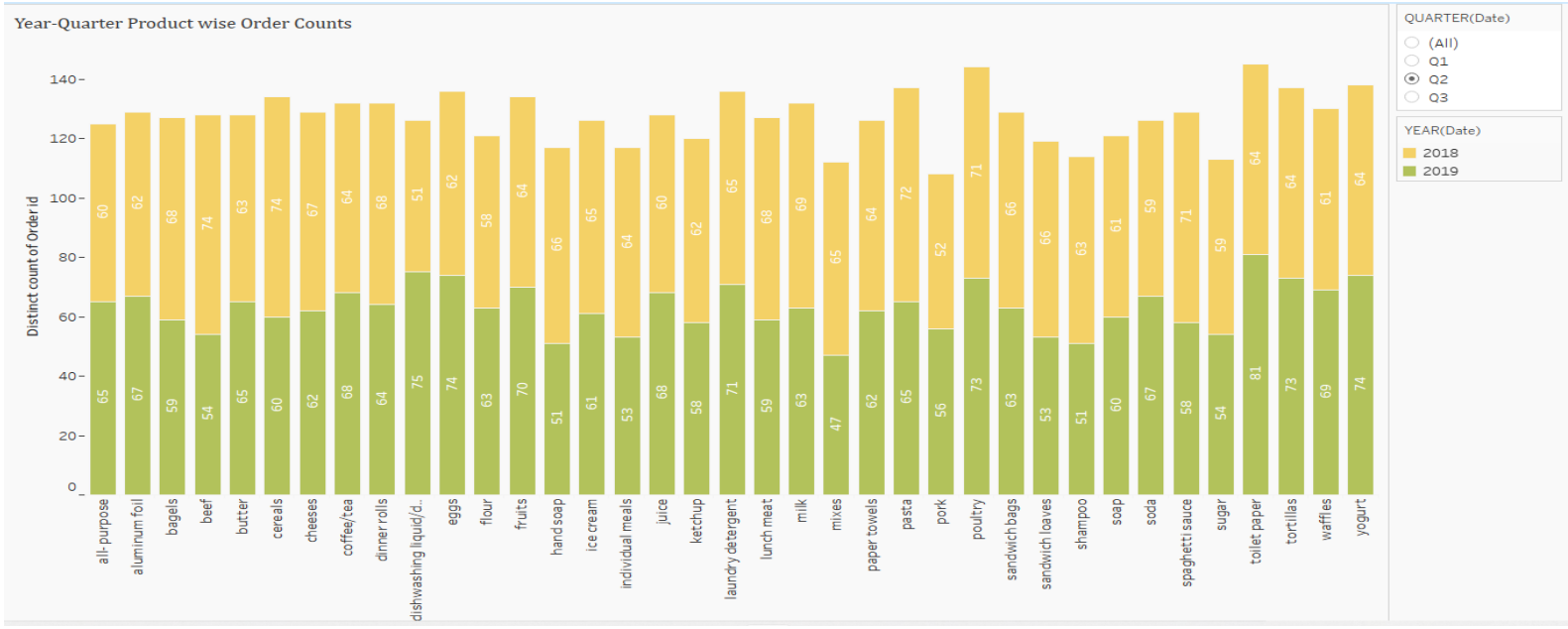
Year Quarter Order Counts -

1. **Consistent Performance (2018 & 2019)**: The order counts in 2018 and 2019 show relatively consistent performance across the first three quarters, with counts hovering around 2,400 to 2,600.
2. **Growth in Orders (2018 to 2019)**: There is a slight increase in order counts from 2018 to 2019. Notably, Q2 and Q3 figures in 2019 are the highest when compared to the previous year.
3. **Significant Drop in 2020**: Q1 of 2020 shows a drastic decline in order counts, dropping to 1,384. This suggests a significant change in demand or market conditions, possibly due to the COVID-19 pandemic's impact starting in early 2020.
4. **Trends Across Quarters**: Each quarter's order counts appear to have a trend of being higher in the earlier quarters (particularly Q3) compared to Q1 of 2020.

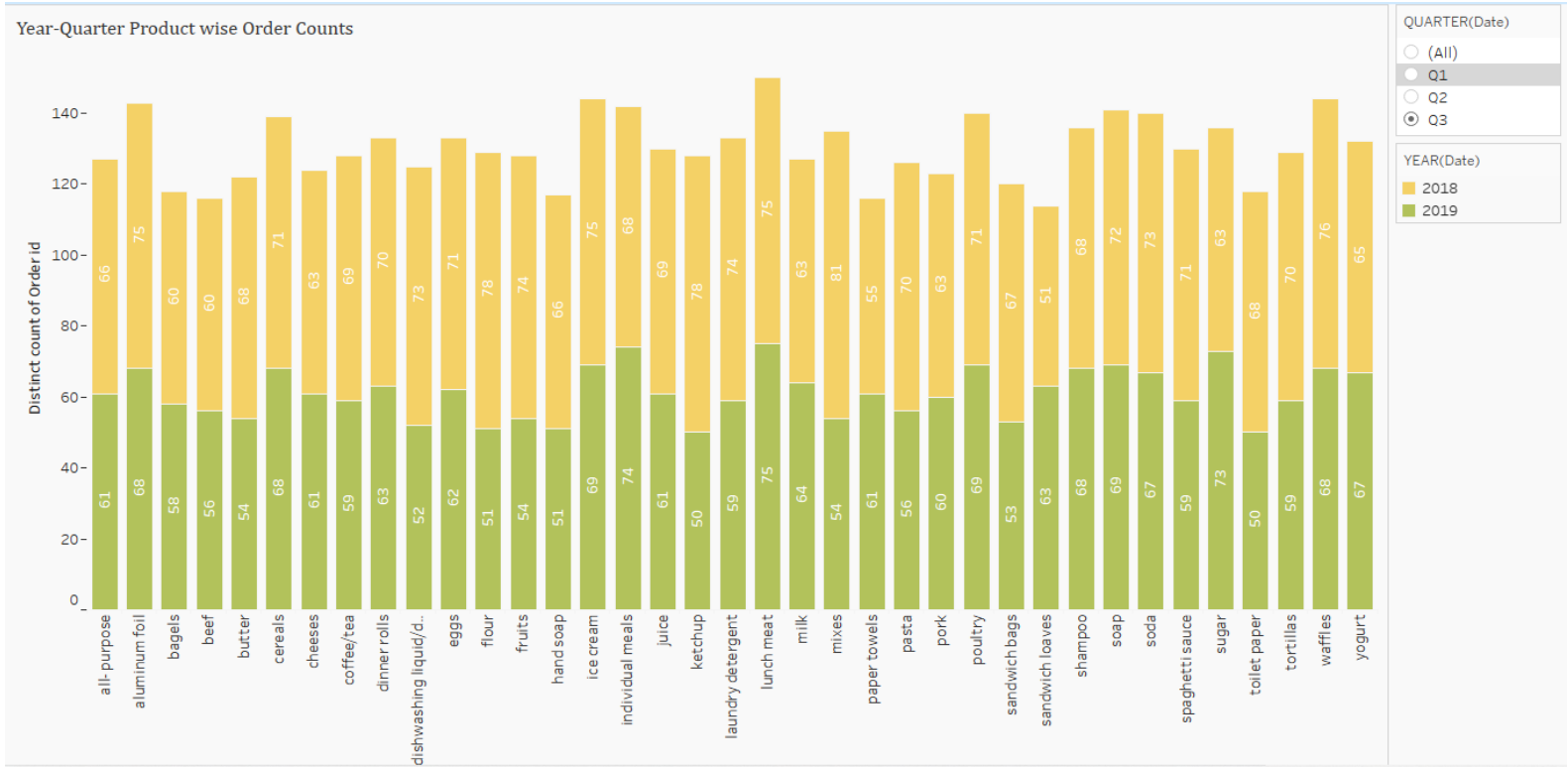
In summary, 2018 and 2019 show stable and growing order counts, while 2020 marks a significant downturn in Q1, reflecting a possible market disruption.



- Polutry and Ice cream had more order in Q1 2019 when compared with Q1 2018.
- Flour Had the maximum from in 2019 when compared with 2018 Q1.



- For Polutry, trend observed between Q2 2018 and Q2 2019 is similar to Q1(more orders in Q2 2019).
- However, for the ice cream order count is higher for Q2 2019.
- Ketchup and mixes had the maximum drop in 2019 when compared to 2018 for Q2.



RFM/MRA Project

- For Poultry trend observed between Q3 2018 and Q3 2019 is different from Q1 and Q2 (more orders in Q3 2018).
- For Ice Cream trend is similar to Q2.
- Ketchup and toilet paper had the maximum drop in 2019 when compared to 2018 for Q2.

Summary –

- Poultry, Soda, Cereal, Ice-cream, Cheese, and Waffles were the top products, indicating high popularity and demand.
- Order activity remained relatively stable from 2018 to 2019, with a slight decrease observed.
- Limited data for 2020 prevents a comprehensive understanding of order trends for that year.
- There was a general downward trend in unique order IDs from Q1 to Q3, suggesting a potential decrease in overall order activity.
- January, February, and May had the highest number of unique orders, while June had the lowest.
- Weekdays showed consistent order activity, with higher counts on Wednesday, Thursday, and Friday, and weekends had the highest overall counts, indicating increased customer engagement
- Over the course of three years, the analysis of customer orders reveals changing preferences and trends. In 2018, cereals, poultry, flour, shampoo, and lunch meat were popular choices, indicating a demand for everyday household items. In 2019, poultry, soda, dishwashing liquid, waffles, and milk took the lead, suggesting a mix of food and cleaning products. However, in 2020, there was a shift towards dinner rolls, poultry, pork, ice cream, and beef, showcasing a desire for more indulgent food options. This suggests a shift in customer preferences between the two years, indicating changing market trends.

Market Basket Analysis -

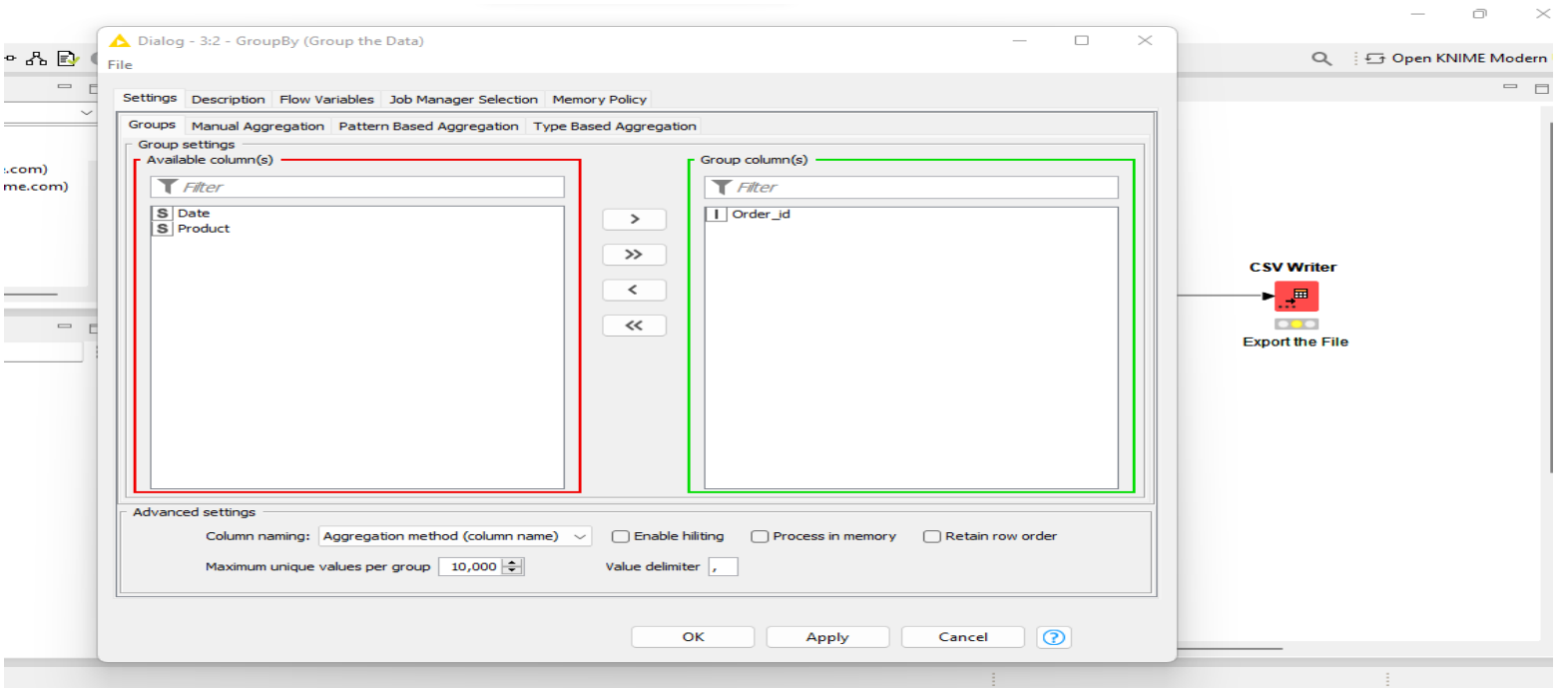
- Market Basket Analysis is a technique used to identify patterns and associations among products that are frequently purchased together in customer transactions. It helps businesses understand customer behaviour and uncover product relationships, which can be used to optimize sales strategies and increase revenue. By analysing transactional data, Market Basket Analysis

generates insights on item co-occurrence and association rules, enabling businesses to make informed decisions on product bundling, cross-selling, and targeted marketing campaigns. This analysis provides valuable insights into customer preferences, allowing businesses to improve customer satisfaction and drive business growth.

Association Rules & its relevance -

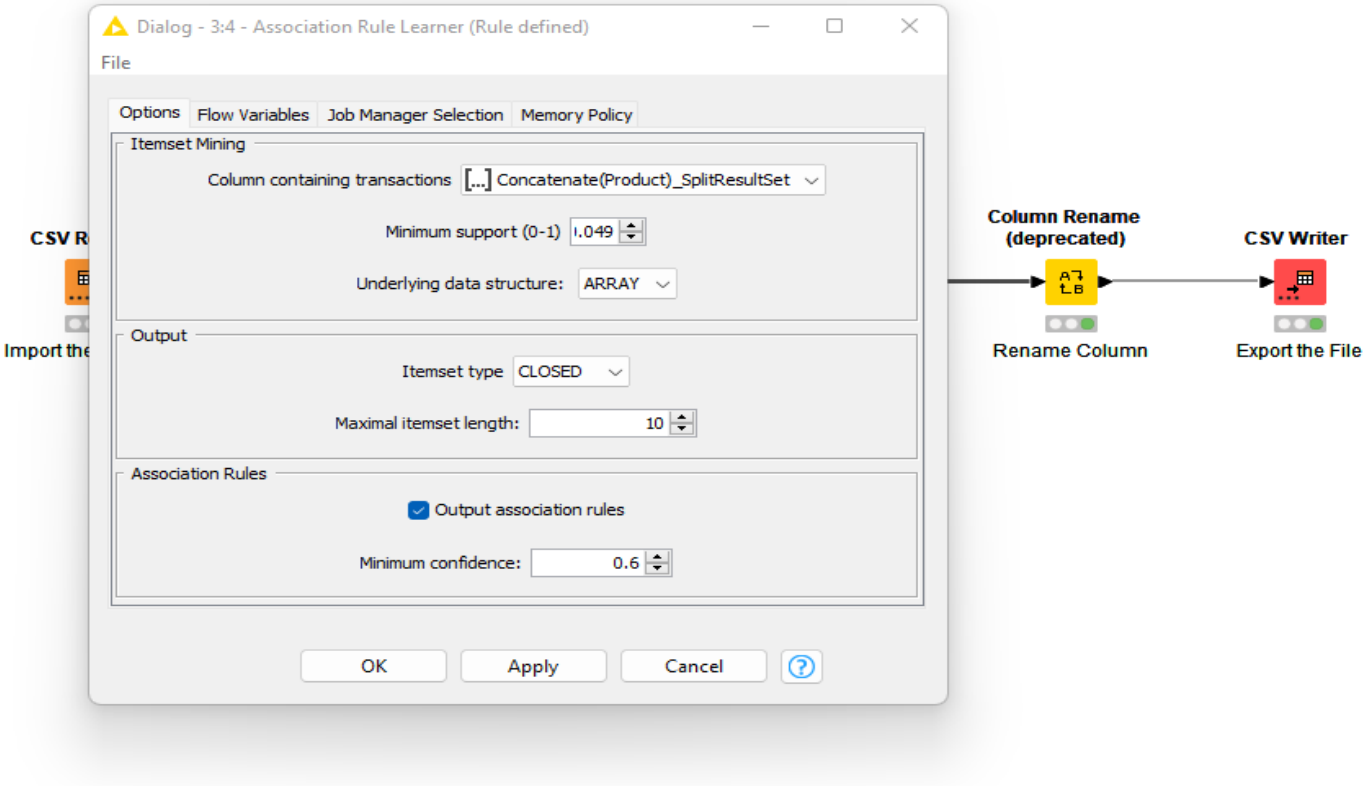
- Association rules in Market Basket Analysis reveal the relationships and co-occurrence patterns between items, providing valuable insights into customer purchasing behaviour and preferences.
- The relevance of association rules lies in their ability to guide businesses in optimizing product placement, creating targeted marketing campaigns, and implementing effective cross-selling and upselling strategies to enhance customer satisfaction and increase revenue.

KNIME workflow image –



Write about threshold values of Support and Confidence -

- Threshold value for Minimum Support is 0.05
- Threshold value for Minimum Confidence is 0.6
- In this analysis, we have defined the threshold values for Support and Confidence as 0.05 and 0.6, respectively. These values help us determine which association rules are significant and trustworthy. By using these thresholds, we can filter out less important rules and focus on the ones that have strong support and confidence levels, ensuring that our analysis provides meaningful insights for decision-making.



Associations Identification –

- Put the associations in a tabular manner
- Explain about support, confidence, & lift values that are calculated

Put the associations in a tabular manner

Row ID	Support	Confidence	Lift	Recommended_item	Recommended_with	Items_list
rule0	0.049	0.629	1.636	yogurt	<---	[eggs, juice, aluminum foil]
rule1	0.049	0.615	1.558	lunch meat	<---	[shampoo, waffles, juice]
rule2	0.049	0.609	1.544	waffles	<---	[shampoo, lunch meat, juice]
rule3	0.049	0.651	1.634	ice cream	<---	[paper towels, eggs, lunch meat]
rule4	0.049	0.615	1.697	paper towels	<---	[eggs, ice cream, lunch meat]
rule5	0.049	0.629	1.614	eggs	<---	[poultry, pasta, soda]
rule6	0.049	0.609	1.561	eggs	<---	[paper towels, pasta, soda]
rule7	0.049	0.622	1.618	aluminum foil	<---	[poultry, beef, juice]
rule8	0.049	0.636	1.51	poultry	<---	[beef, juice, aluminum foil]
rule9	0.049	0.675	1.601	poultry	<---	[beef, toilet paper, sugar]
rule10	0.049	0.622	1.6	dinner rolls	<---	[spaghetti sauce, poultry, sugar]
rule11	0.049	0.659	1.563	poultry	<---	[dinner rolls, spaghetti sauce, sugar]
rule12	0.049	0.644	1.615	ice cream	<---	[cheeses, pasta, waffles]
rule13	0.049	0.622	1.637	milk	<---	[poultry, ice cream, cereals]
rule14	0.05	0.64	1.7	juice	<---	[yogurt, toilet paper, aluminum foil]
rule15	0.05	0.62	1.645	juice	<---	[yogurt, poultry, aluminum foil]
rule16	0.05	0.613	1.616	coffee/tea	<---	[yogurt, cheeses, cereals]
rule17	0.05	0.6	1.424	poultry	<---	[dishwashing liquid/detergent, laundry detergent, mixes]
rule18	0.051	0.63	1.678	mixes	<---	[yogurt, poultry, aluminum foil]
rule19	0.051	0.611	1.66	sandwich bags	<---	[cheeses, bagels, cereals]
rule20	0.051	0.674	1.726	cheeses	<---	[bagels, cereals, sandwich bags]
rule21	0.051	0.617	1.558	cereals	<---	[cheeses, bagels, sandwich bags]
rule22	0.051	0.63	1.621	dinner rolls	<---	[spaghetti sauce, poultry, cereals]
rule23	0.051	0.637	1.512	poultry	<---	[dinner rolls, spaghetti sauce, cereals]
rule24	0.051	0.604	1.589	milk	<---	[poultry, laundry detergent, cereals]
rule25	0.052	0.628	1.61	eggs	<---	[dinner rolls, poultry, soda]
rule26	0.052	0.641	1.649	dinner rolls	<---	[spaghetti sauce, poultry, ice cream]
rule27	0.052	0.686	1.628	poultry	<---	[dinner rolls, spaghetti sauce, ice cream]
rule28	0.052	0.628	1.614	dinner rolls	<---	[spaghetti sauce, poultry, juice]
rule29	0.052	0.602	1.429	poultry	<---	[dinner rolls, spaghetti sauce, juice]
rule30	0.052	0.634	1.627	eggs	<---	[paper towels, dinner rolls, pasta]
rule31	0.052	0.602	1.621	pasta	<---	[paper towels, eggs, dinner rolls]
rule32	0.054	0.642	1.651	dinner rolls	<---	[spaghetti sauce, poultry, laundry detergent]
rule33	0.054	0.656	1.556	poultry	<---	[dinner rolls, spaghetti sauce, laundry detergent]
rule34	0.055	0.624	1.565	ice cream	<---	[paper towels, eggs, pasta]
rule35	0.055	0.63	1.616	eggs	<---	[paper towels, ice cream, pasta]
rule36	0.055	0.643	1.731	pasta	<---	[paper towels, eggs, ice cream]
rule37	0.055	0.649	1.791	paper towels	<---	[eggs, ice cream, pasta]

Explain about support, confidence, & lift values that are calculated

- **Support:** The support value represents the frequency or popularity of an itemset in the dataset. It indicates how often a specific combination of items appears together in customer transactions.
- **Confidence:** Confidence measures the likelihood that a customer who buys one item will also purchase another item. It is calculated as the ratio of the number of transactions where both items are purchased together to the number of transactions where the first item is purchased.

- **Lift:** Lift measures the strength of association between two items in an association rule. It compares the probability of the two items being purchased together to the probability of them being purchased independently. A lift value greater than 1 suggests a positive association, indicating that the items are more likely to be purchased together.

Suggestion of Possible Combos with Lucrative Offers -

Write recommendations –

Make discount offers or combos (or buy two get one free) based on the associations and your experience

Recommendations (discount offers/ combos)

- **Combo Deal:** Offer a special combo deal where customers can buy yogurt, poultry, and aluminium foil along with juice to avail a discounted price or additional item.
- **Buy Two Get One Free:** Introduce a "buy two get one free" offer on dinner rolls, spaghetti sauce, and ice cream to incentivize customers to purchase these items together.
- **Bundle Promotion:** Create a bundle promotion where customers can buy paper towels, eggs, and pasta together at a discounted.
- **Cross-Selling Offer:** Provide a cross-selling offer where customers purchasing cereals can get a discount on cheese, bagels, and sandwich bags.
- **Limited-Time Promotion:** Launch a limited-time promotion where customers buying poultry, laundry detergent, and mixes can receive a percentage savings.
- **Loyalty Program:** Implement a loyalty program where customers who frequently purchase recommended items or participate in the suggested combos can earn rewards or exclusive discounts. This will

incentivize customer loyalty and encourage them to continue shopping with the store, fostering long-term relationships and repeat purchases.

These recommendations are based on the association rules and the occurrence of certain items together, aiming to increase customer satisfaction and encourage them to explore additional products.

Project END