Marketing and Retail Analytics Project

AUTO SALES

Agenda and Executive Summary of the data

Contents of the presentation

- Problem statement

- Data Analysis

  - Info, Shape, Data type, Null value count, Summary Stats

- 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

  - Summary of the inferences from the above analysis

  - Measures to improve sales

- Customer Segmentation using RFM analysis (make 4 segments)

  - What is RFM?

  - What all parameters used and assumptions made

  - Output table head

  - KNIME Workflow

- Inferences from RFM Analysis and identified segments

  - Who are your best customers? (give at least 5)

  - Which customers are on the verge of churning? (give at least 5)

  - Who are your lost customers? (give at least 5)

  - Who are your loyal customers? (give at least 5)

Problem statement

An automobile parts manufacturing company has collected data of transactions for 3 years. They do not have any in-house data science team, thus they have hired you as their consultant. Your job is to use your magical data science skills to provide them with suitable insights about their data and their customers.

Auto Sales Data: Sales_Data.xlsx

Data Analysis – Info

```
<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
```

The dataset has 20 columns and 2747 records as seen in the information of the dataset on the left.

There are 2747 non-null records in all the columns meaning there are no missing records based on the initial analysis that was done.

Data Analysis – Shape

(2747, 20)

The no. of rows: 2747
The no. of columns: 20

The shape of the data is (2747, 20) meaning the dataset has 2747 rows and 20 columns as shown on the left.

Data Analysis – Data type

```
ORDERNUMBER                      int64
QUANTITYORDERED                  int64
PRICEEACH                      float64
ORDERLINENUMBER                  int64
SALES                          float64
ORDERDATE               datetime64[ns]
DAYS_SINCE_LASTORDER             int64
STATUS                          object
PRODUCTLINE                     object
MSRP                             int64
PRODUCTCODE                     object
CUSTOMERNAME                    object
PHONE                           object
ADDRESSLINE1                    object
CITY                            object
POSTALCODE                      object
COUNTRY                         object
CONTACTLASTNAME                 object
CONTACTFIRSTNAME                object
DEALSIZE                        object
dtype: object
```

The dataset has 20 variables out of which there are:

- 12 categorical variables,

- 7 numerical variables and

- 1 date time field.

Data Analysis – Null Value count

```
ORDERNUMBER            0
QUANTITYORDERED        0
PRICEEACH              0
ORDERLINENUMBER        0
SALES                  0
ORDERDATE              0
DAYS_SINCE_LASTORDER   0
STATUS                 0
PRODUCTLINE            0
MSRP                   0
PRODUCTCODE            0
CUSTOMERNAME           0
PHONE                  0
ADDRESSLINE1           0
CITY                   0
POSTALCODE             0
COUNTRY                0
CONTACTLASTNAME        0
CONTACTFIRSTNAME       0
DEALSIZE               0
dtype: int64
```

There are no null values or missing values in the dataset.

Data Analysis – Summary stats

	ORDERNUMBER	QUANTITYORDERED	PRICEEACH	ORDERLINENUMBER	SALES	DAYS_SINCE_LASTORDER	MSRP
count	2747.000000	2747.000000	2747.000000	2747.000000	2747.000000	2747.000000	2747.000000
mean	10259.761558	35.103021	101.098951	6.491081	3553.047583	1757.085912	100.691664
std	91.877521	9.762135	42.042548	4.230544	1838.953901	819.280576	40.114802
min	10100.000000	6.000000	26.880000	1.000000	482.130000	42.000000	33.000000
25%	10181.000000	27.000000	68.745000	3.000000	2204.350000	1077.000000	68.000000
50%	10264.000000	35.000000	95.550000	6.000000	3184.800000	1761.000000	99.000000
75%	10334.500000	43.000000	127.100000	9.000000	4503.095000	2436.500000	124.000000
max	10425.000000	97.000000	252.870000	18.000000	14082.800000	3562.000000	214.000000

- The descriptive statistics of the numerical columns is shown in the table.

- All the columns in the dataframe have 2747 values. There are no missing values.

- By looking at the table we can see that SALES has the highest mean value and ORDERLINENUMBER has the lowest mean value. The same is true in the case of standard deviation too.

- This is probably because SALES and ORDERLINENUMBER are two different measurements.

- The mean and median values are approximately equal in all variables.

Exploratory Analysis and Inferences

Univariate, Bivariate and multivariate analysis using data visualization

Skewness value of numerical variables

```
SALES                  1.155940
PRICEEACH              0.697222
MSRP                   0.575646
ORDERLINENUMBER        0.575327
QUANTITYORDERED        0.369286
DAYS_SINCE_LASTORDER  -0.002983
ORDERNUMBER           -0.006995
dtype: float64
```
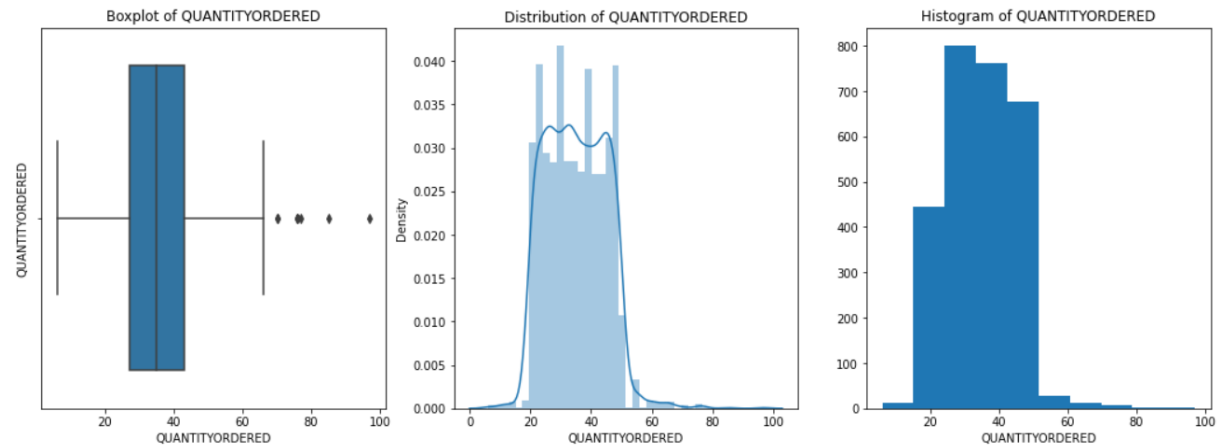
Univariate Analysis – Quantity Ordered

- The boxplot of 'QUANTITYORDERED' variable has outliers.

- The distribution of data is moderately right skewed.

- The skewness value of the variable is 0.369286.

- The mean of the data is 35.10 meaning the quantity of items ordered is 35 on average.

```
Description of QUANTITYORDERED
----------------------------------------
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
```
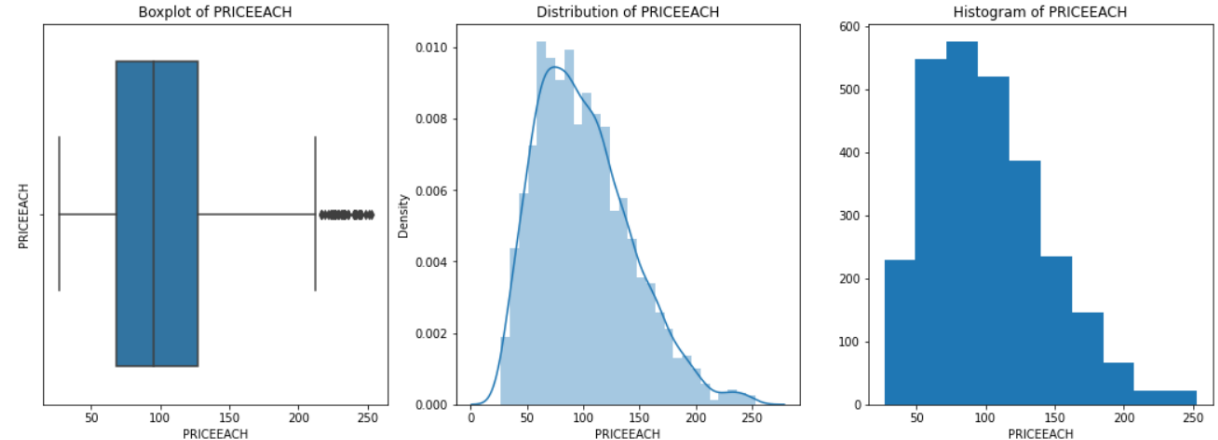
Univariate Analysis – Price Each

- The boxplot of 'PRICEEACH' variable has outliers.

- The distribution of data is moderately right skewed.

- The skewness value of the variable is 0.697222.

- The mean of the data is 101.09 meaning the price of each item is 101.09 on average.

```
Description of PRICEEACH
--------------------------------
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
```
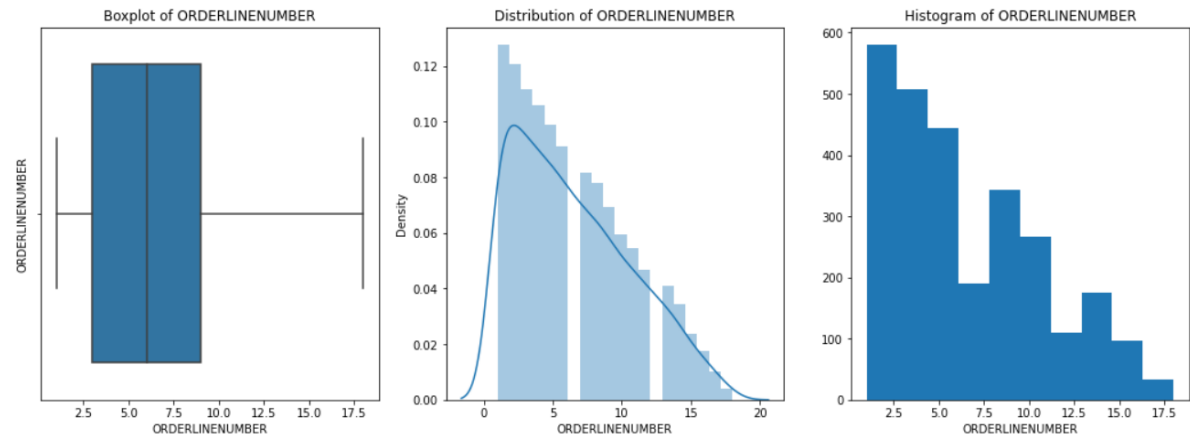
Univariate Analysis – Order Line Number

- The boxplot of 'ORDERLINENUMBER' variable has no outliers.

- The distribution of data is moderately right skewed.

- The skewness value of the variable is 0.575327.

```
Description of ORDERLINENUMBER
-----------------------------------------
count    2747.000000
mean        6.491081
std         4.230544
min         1.000000
25%         3.000000
50%         6.000000
75%         9.000000
max        18.000000
Name: ORDERLINENUMBER, dtype: float64
```
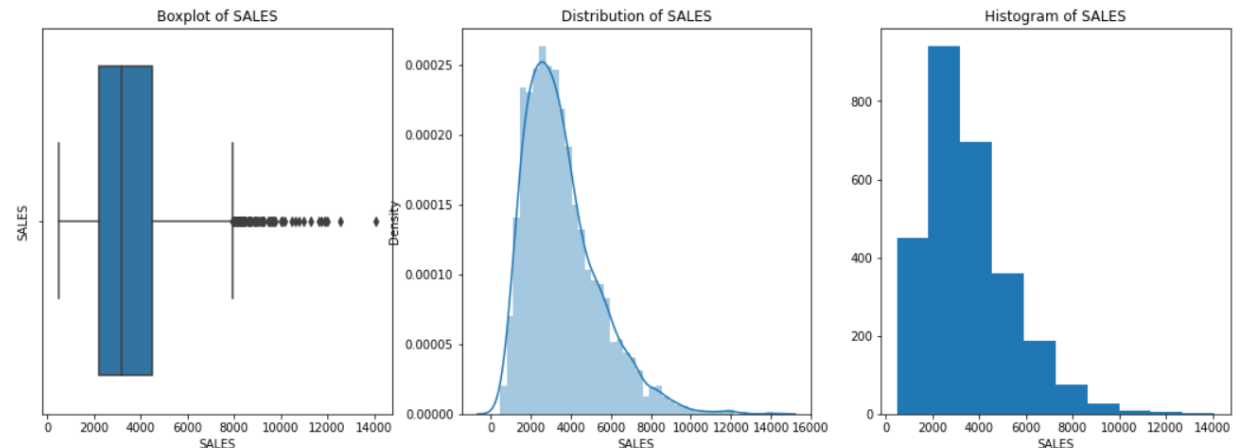
Univariate Analysis – Sales

- The boxplot of 'SALES' variable has outliers.

- The distribution of data is moderately right skewed.

- The skewness value of the variable is 1.155940.

- The mean of the data is 3553.05 meaning the amount of sales is 3553.05 on average.

```
Description of SALES
------------------------------
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
```

Univariate Analysis – MSRP

```
Description of MSRP
----------------------------
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
```
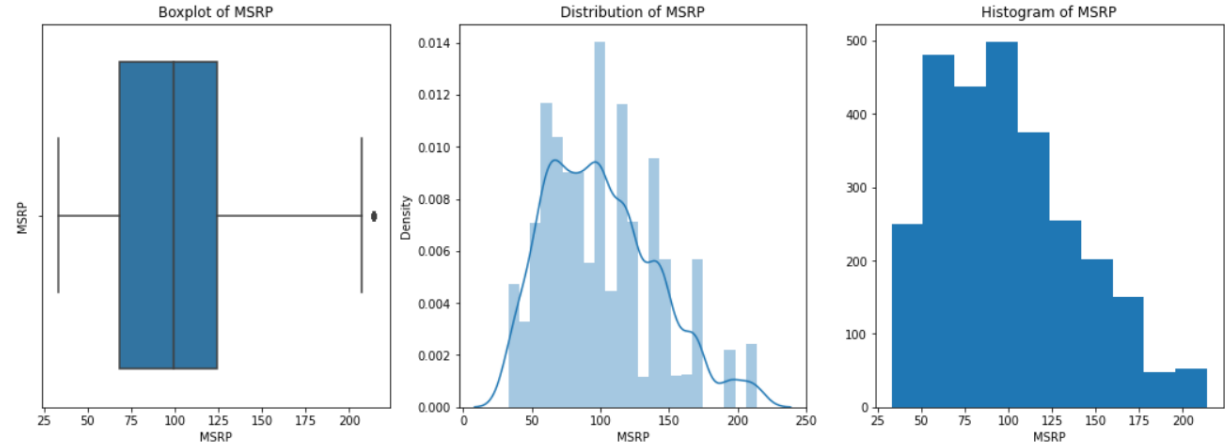
- MSRP is Manufacturer's Suggested Retail Price.

- The boxplot of 'MSRP' variable has outliers.

- The distribution of data is moderately right skewed.

- The skewness value of the variable is 0.575646.

- The mean of the data is 100.69 meaning the Manufacturer's Suggested Retail Price is 100.69 on average.

Bivariate Analysis
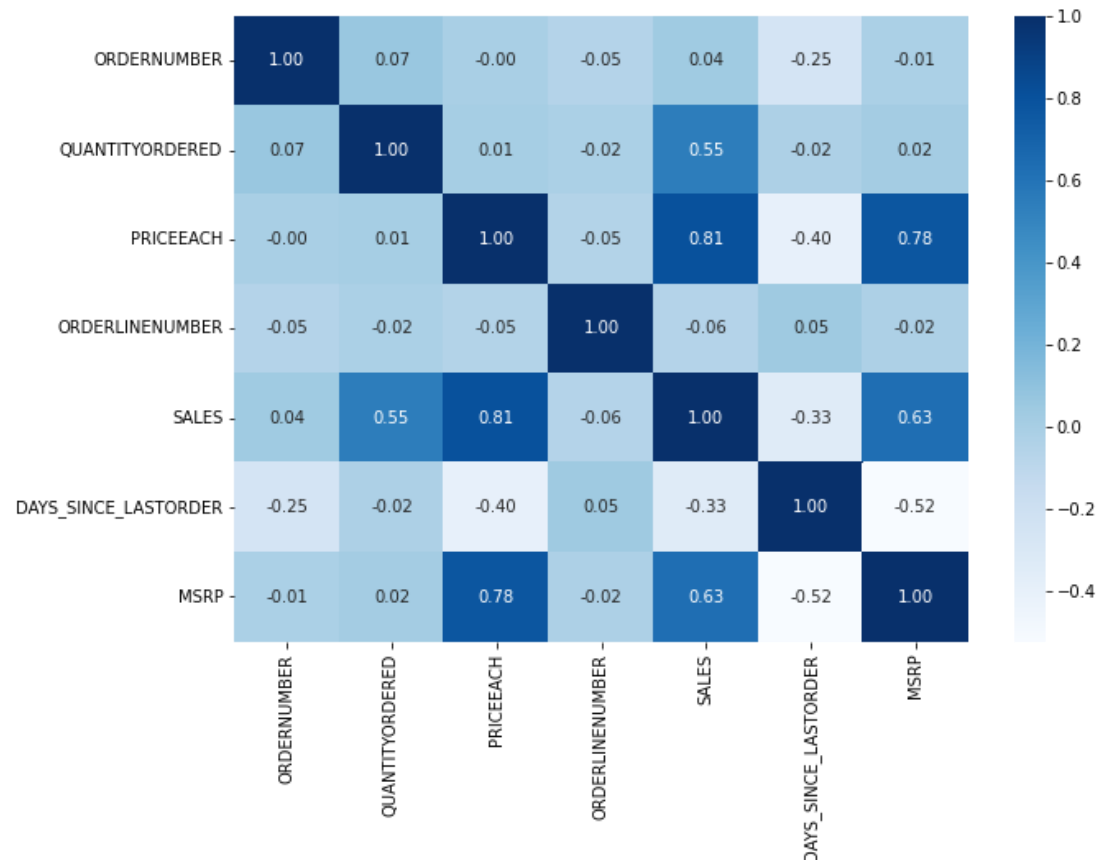
- Bivariate Analysis is done with the help of a heatmap.

- The highest correlation is between SALES and PRICEEACH (81%).

- There is a strong correlation between MSRP & PRICEEACH, MSRP & SALES and SALES & QUANTITYORDERED.

	ORDERNUMBER	QUANTITYORDERED	PRICEEACH	ORDERLINENUMBER	SALES	DAYS_SINCE_LASTORDER	MSRP
ORDERNUMBER	1.000000	0.067110	-0.003369	-0.054300	0.037289	-0.251476	-0.013910
QUANTITYORDERED	0.067110	1.000000	0.010161	-0.016295	0.553359	-0.021923	0.020551
PRICEEACH	-0.003369	0.010161	1.000000	-0.052646	0.808287	-0.397092	0.778393
ORDERLINENUMBER	-0.054300	-0.016295	-0.052646	1.000000	-0.057414	0.046615	-0.020956
SALES	0.037289	0.553359	0.808287	-0.057414	1.000000	-0.334274	0.634849
DAYS_SINCE_LASTORDER	-0.251476	-0.021923	-0.397092	0.046615	-0.334274	1.000000	-0.524285
MSRP	-0.013910	0.020551	0.778393	-0.020956	0.634849	-0.524285	1.000000

Multivariate Analysis

- Multivariate Analysis is done with the help of a pair plot to understand the relationship between all the numerical values in the dataset.

- There is a high positive correlation between SALES and PRICEEACH.
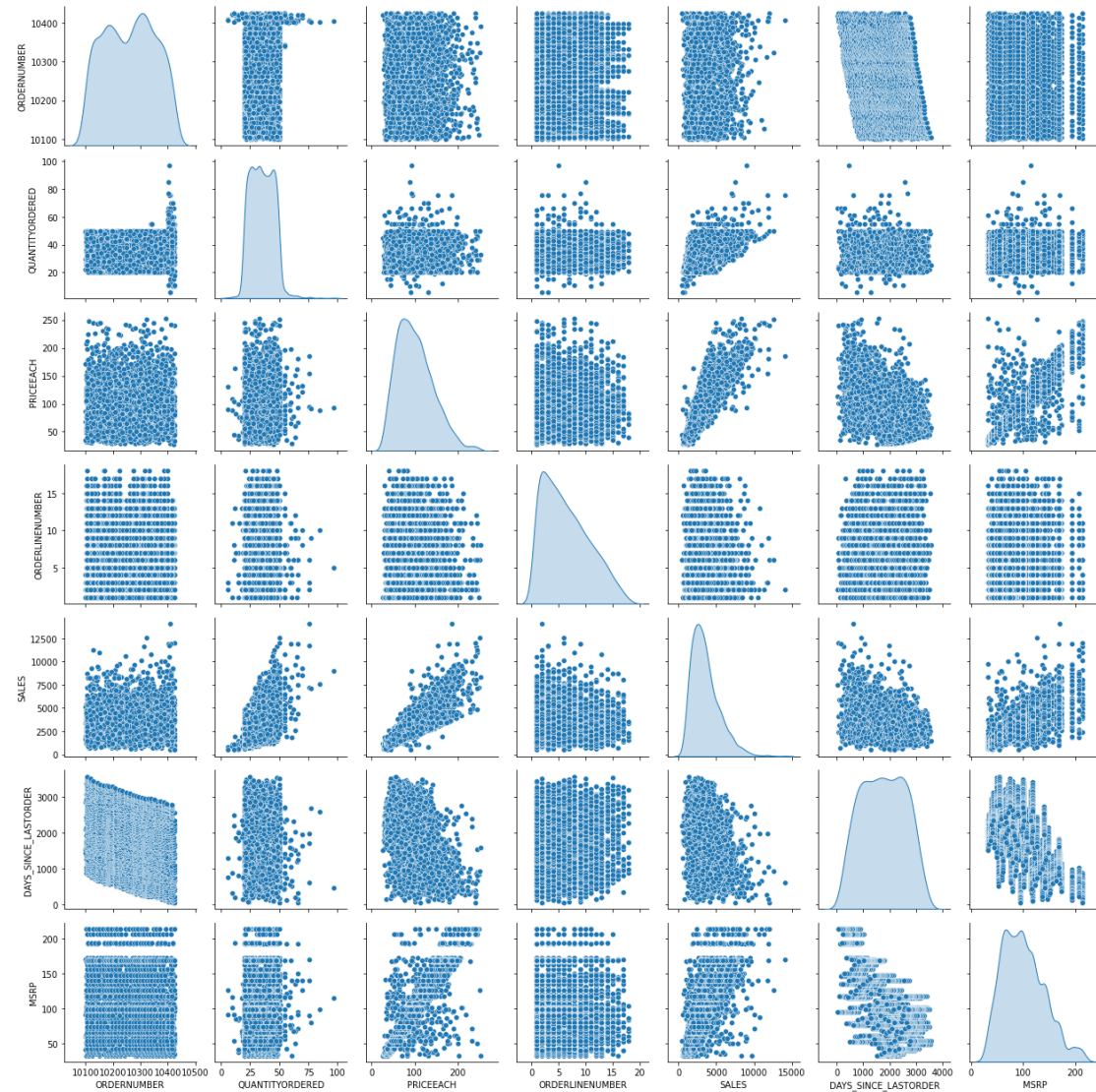
Weekly, Monthly, Quarterly, Yearly Trends in Sales

Trend in Sales

- In 2018 and 2019, sales is highest in the 4th quarter compared to the other 3 quarters of the year.

- In 2020, sales is highest in quarter 2 compared to quarter 1.

SUM(Sales)

129,754   1,058,699

Yearly, quarterly and monthly trend in sales
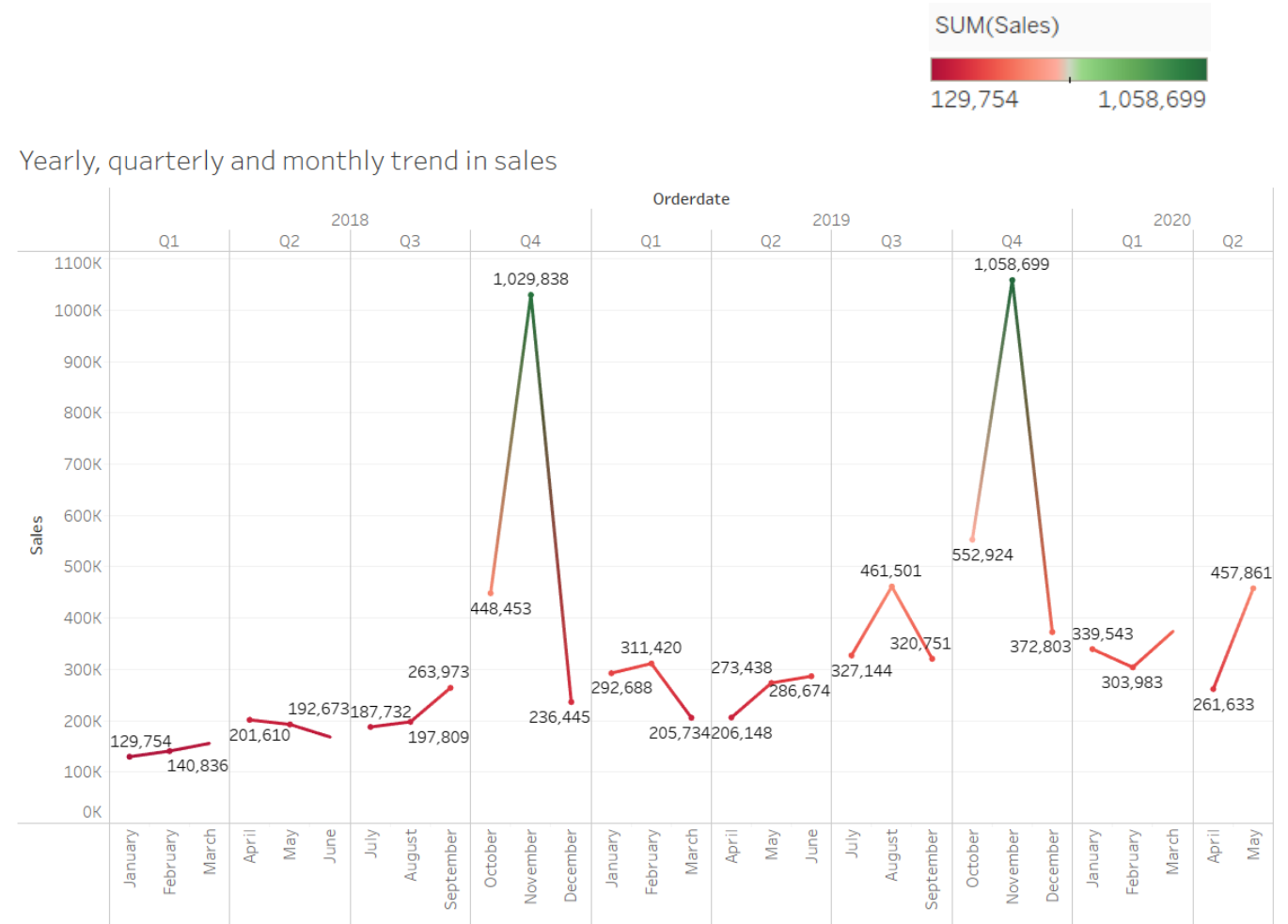
Daily Trend in Sales

- Thursday has the lowest sales compared to all other days

- Sunday has the highest sales followed by Tuesday.

SUM(Sales)

151,979   2,175,121

Daily trend in sales

Orderdate

2,175,121
1,827,923
1,895,469
1,374,952
1,288,236
1,046,542
151,979

Sunday   Monday   Tuesday   Wednesday   Thursday   Friday   Saturday
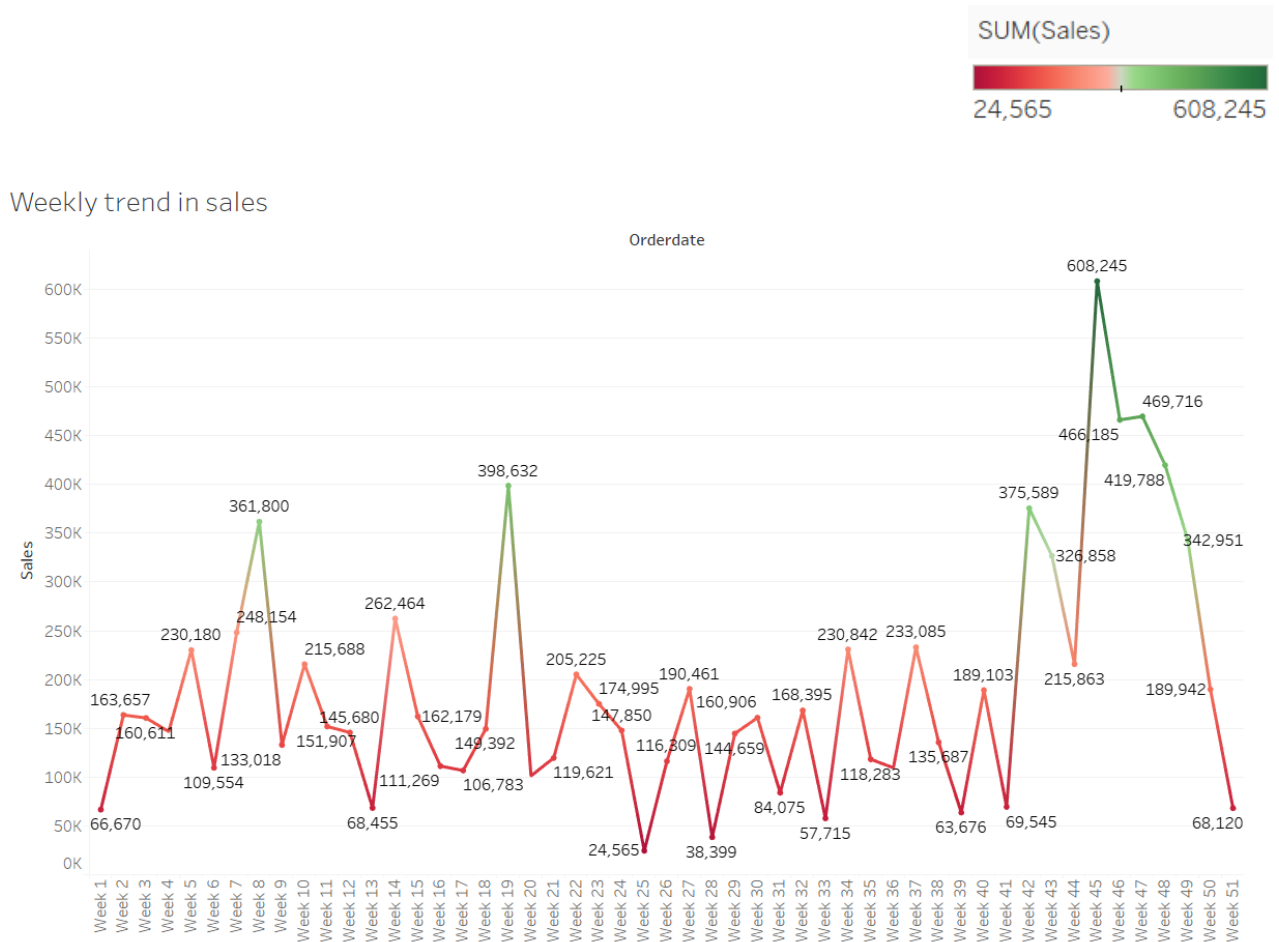
Sales

Monthly Trend in Sales

- There is a downward trend in sales from January to April after which there is a sudden increase in May.

- From May to September, there is a fluctuation in the sales.

- September to November has an increasing trend after which the sales drops again in December.

- November month seems to have the highest sales which validates the weekly trend in sales as Week 45 falls in November.

SUM(Sales)

454,757   2,088,537

Monthly trend in sales

Orderdate
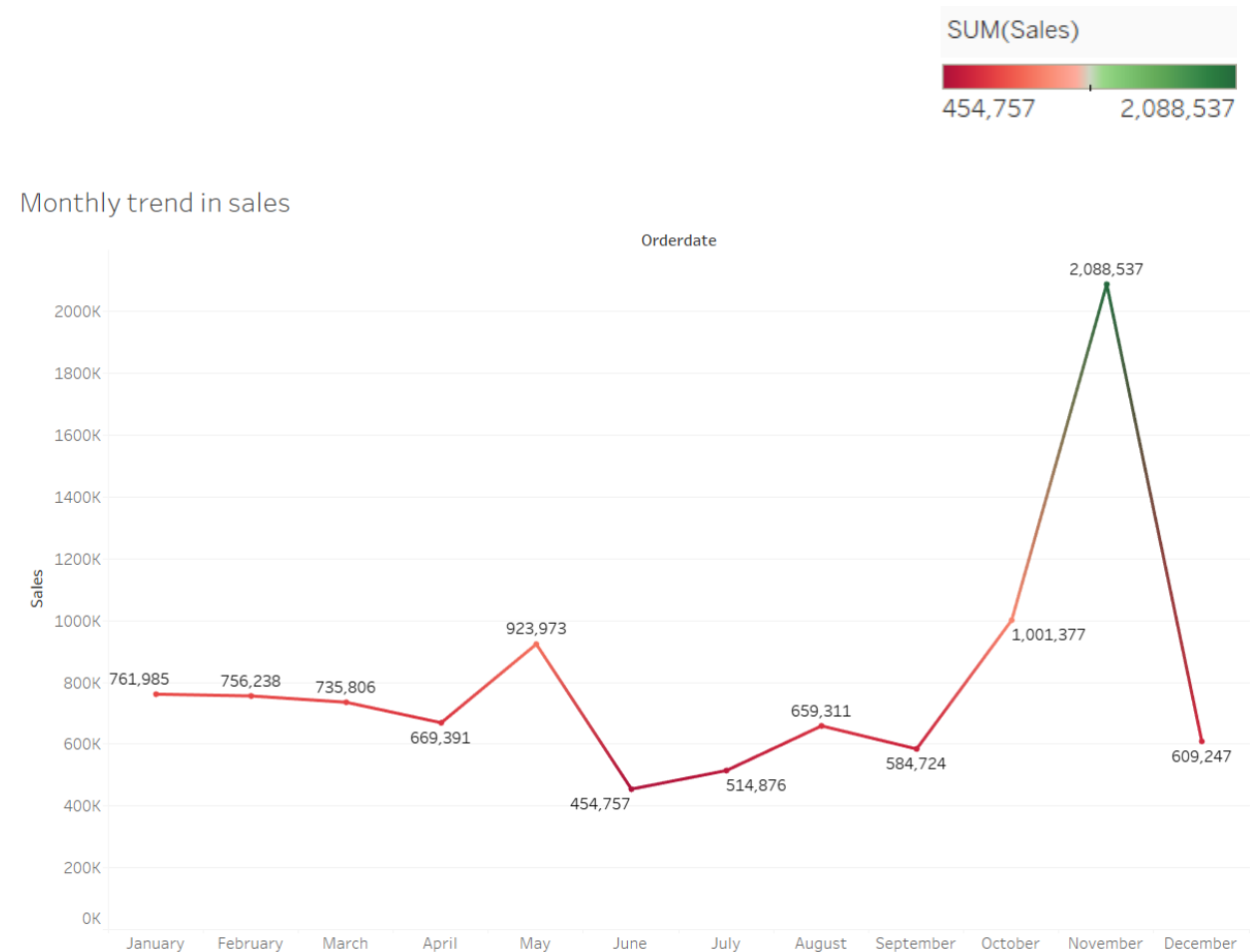
Quarterly Trend in Sales

- Sales is highest in the 4th quarter.

- There is a downward or decreasing trend from the 1st to 3rd quarter.

- From the 3rd to the 4th quarter there is an increasing trend.

SUM(Sales)

1,758,911   3,699,161

Quarterly trend in sales

Orderdate

3,699,161

2,254,029

2,048,120

1,758,911

Sales

Q1   Q2   Q3   Q4
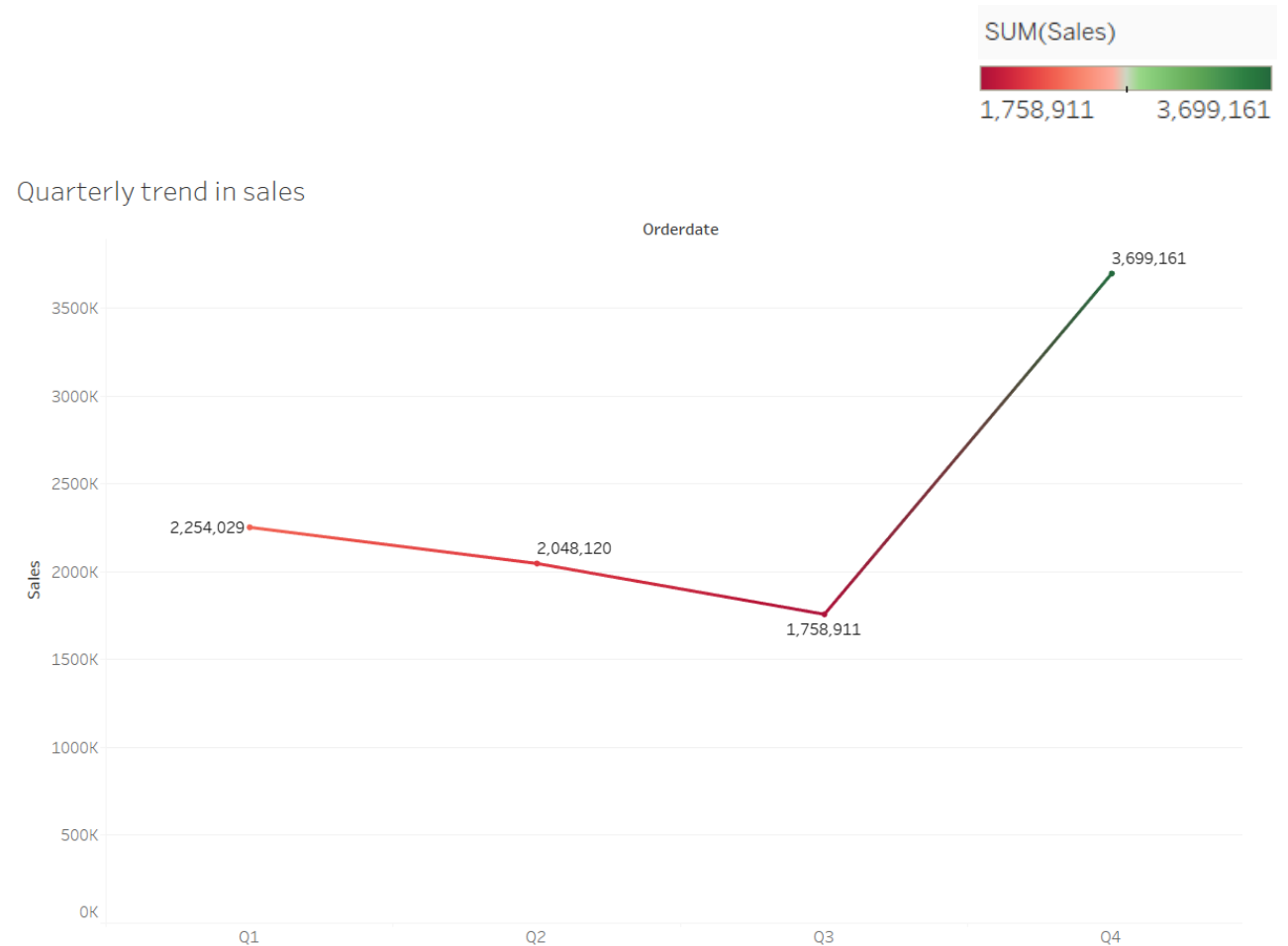
Yearly Trend in Sales

- 2019 has the highest sales.

SUM(Sales)

1,737,283   4,669,925

Yearly trend in sales

Orderdate
4,669,925

3,353,014

1,737,283

2018   2019   2020
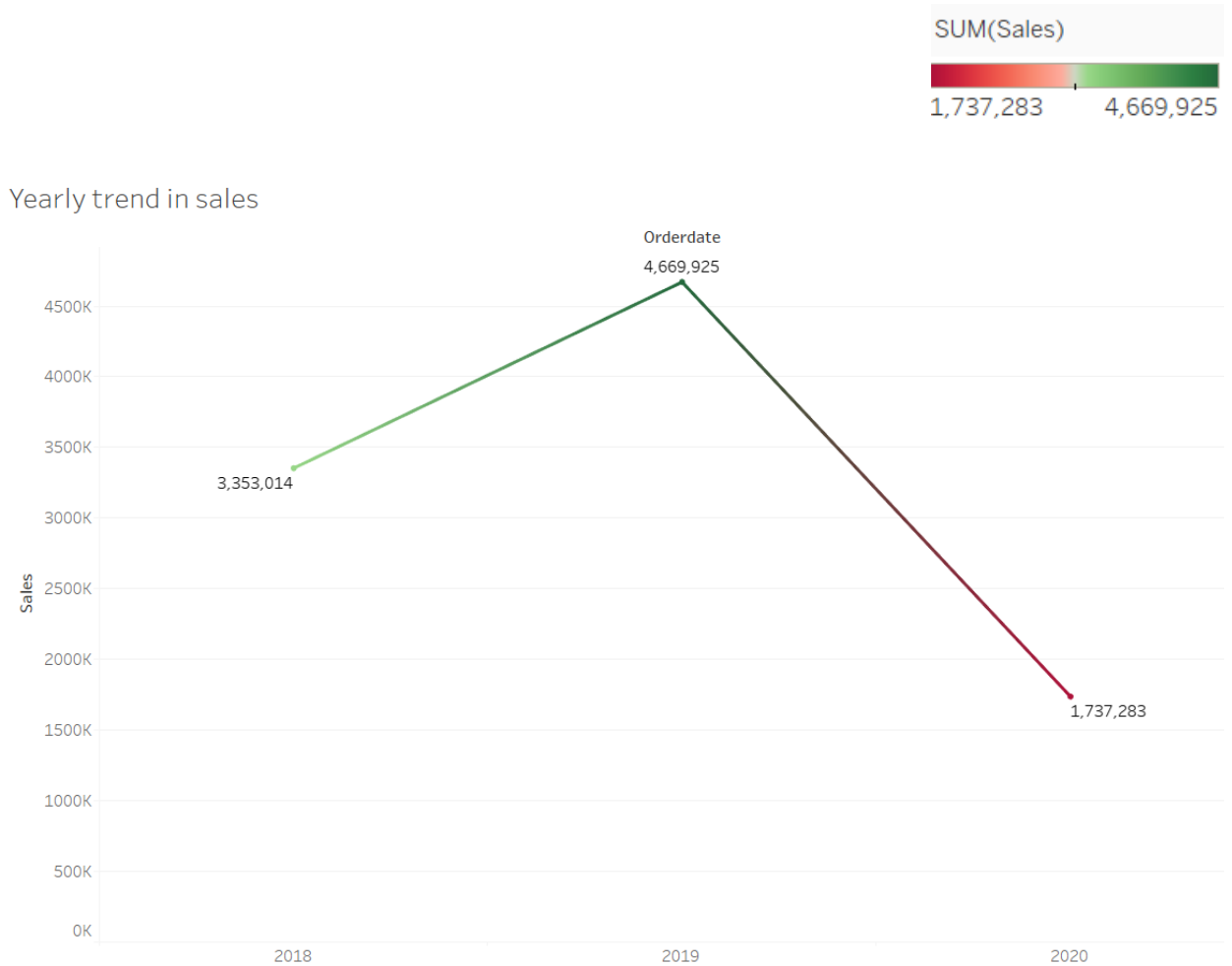
Sales Across different Categories of different features in the given data

Map of Sales and Country

- The map shows the countries that are listed in the dataset and the colour coding shows the sum of sales of each country.

- It is seen that USA has the highest sales.

SUM(Sales)

57,756   3,355,576

Map - Sales and Country

Finland
329,582

Denmark
245,637

Switzerland
117,714

Spain
1,215,687

Canada
224,079

USA
3,355,576

Japan
188,168

Philippines
94,616

Singapore
288,488

Australia
630,623

© 2022 Mapbox © OpenStreetMap
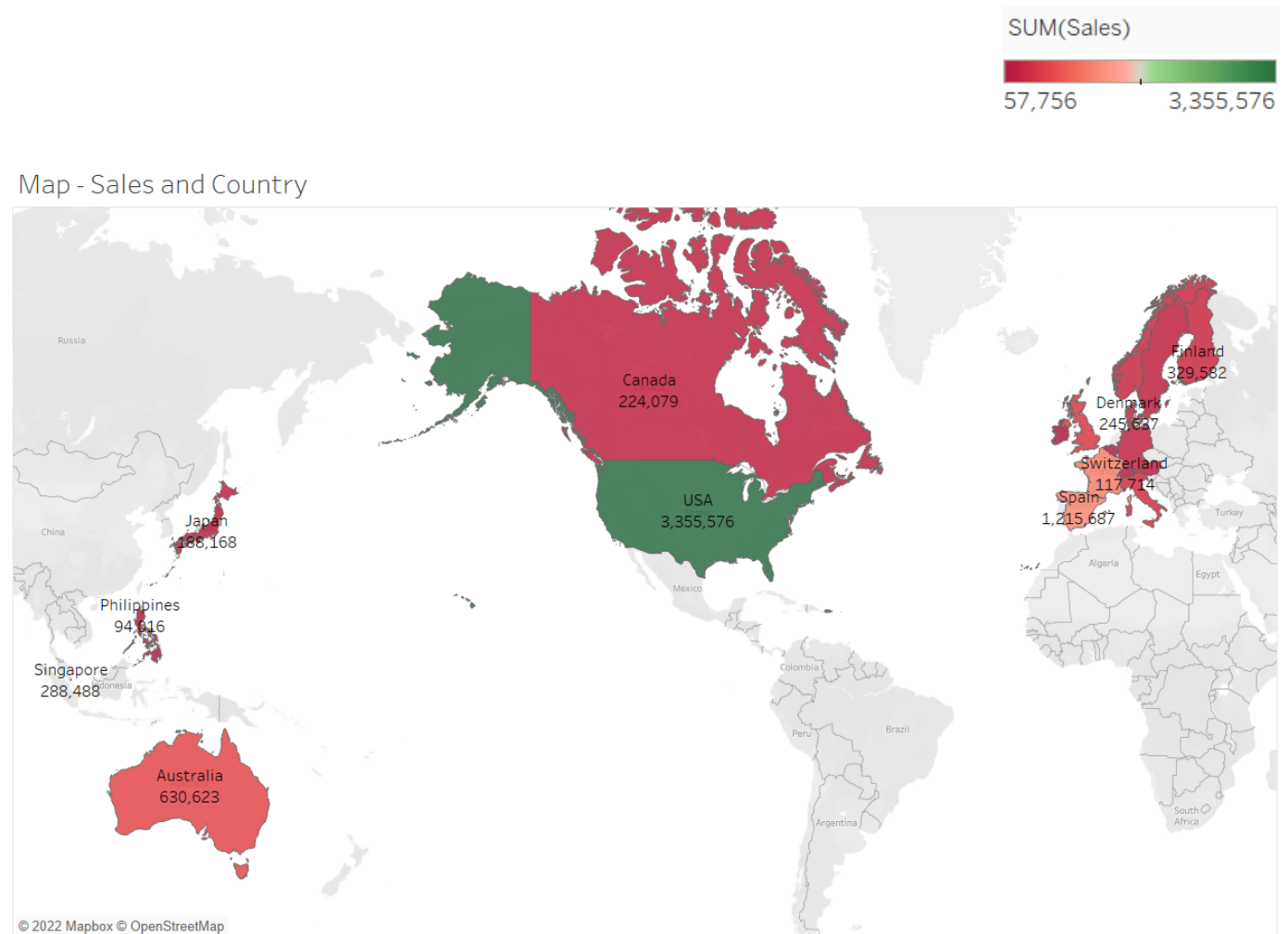
Bubble Chart of Country and Sales

- The bubble chart of sales across country shows that USA has the maximum or the highest sales with the sum of sales of 3,355,576.

- This is followed by Spain having sum of sales as 1,215,687.

Bubble chart - Country and Sales

Country

- Australia
- Austria
- Belgium
- Canada
- Denmark
- Finland
- France
- Germany
- Ireland
- Italy
- Japan
- Norway
- Philippines
- Singapore
- Spain
- Sweden
- Switzerland
- UK
- USA

UK
478,880

Singapore
288,488

Spain
1,215,687

Norway
307,464

Canada

Denmark

Sweden

Japan

Australia
630,623

Austria

Finland
329,582

Italy
374,674

Germany

France
1,110,917

USA
3,355,576
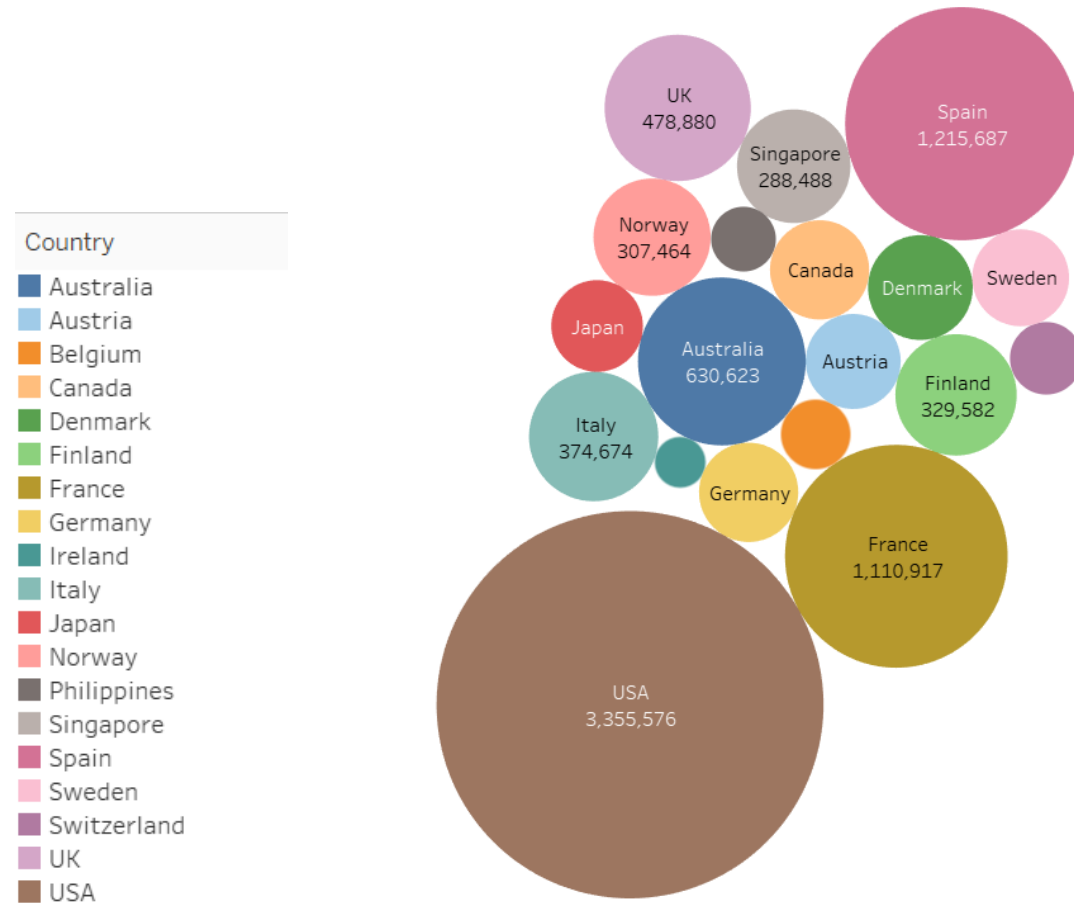
Country and City wise Sales

- The country with the lowest or the least sales is Ireland.

- The treemap also confirms that the highest sales is in USA followed by Spain.

- However, by looking at the sales city wise, it is seen that the city Madrid in Spain is delivering the highest sales compared to other cities.

SUM(Sales)

33,440   1,082,551

Treemap - Country and City wise Sales
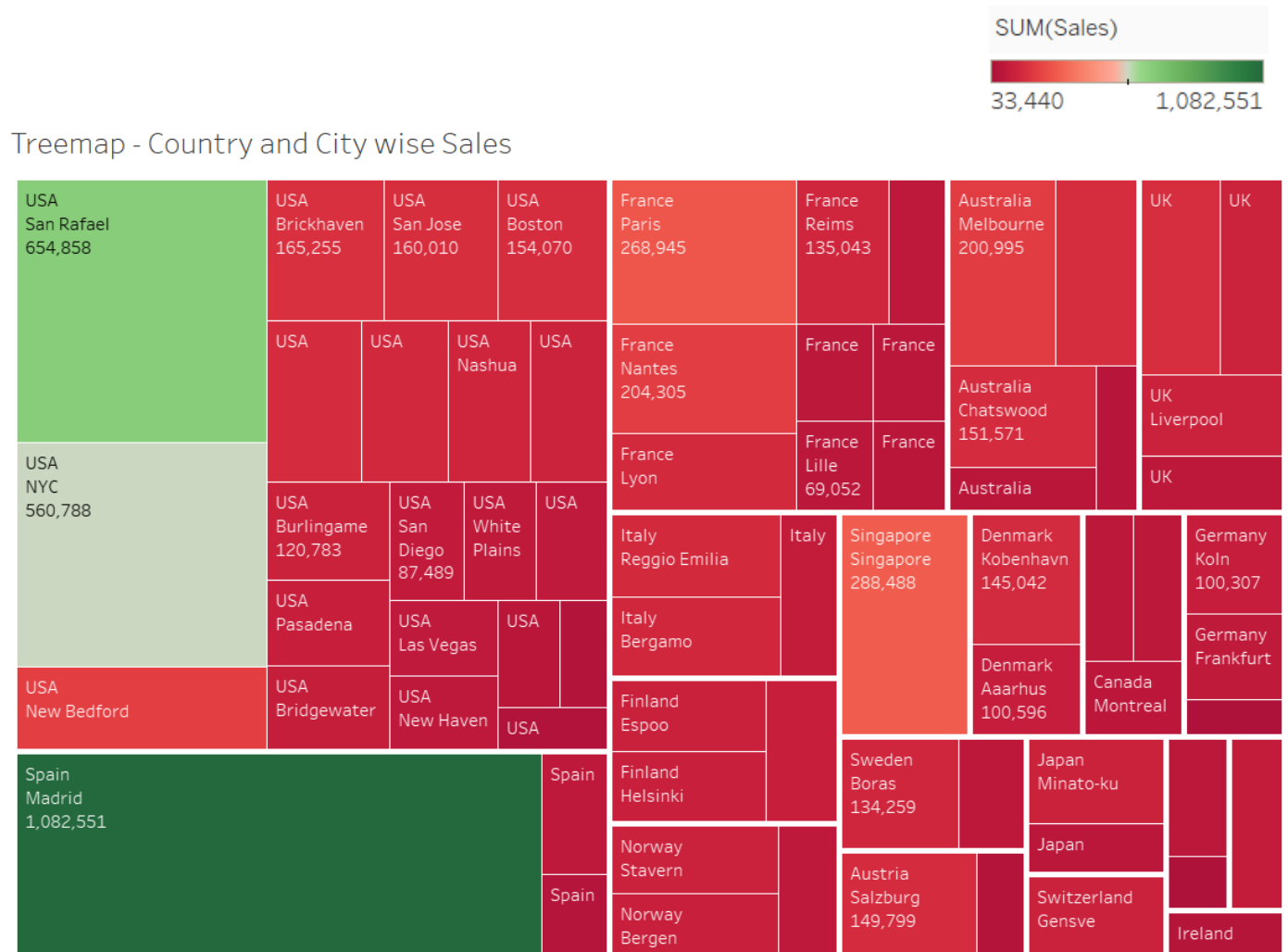
Sales and Category

- From the graph, we can see that 'Classic Cars' have the highest sales.

- 'Trains' have the lowest sales.

Category vs Sales

Productline

Productline

Classic Cars	3,842,869
Vintage Cars	1,806,676
Trucks and Buses	1,111,559
Motorcycles	1,103,512
Planes	969,323
Ships	700,039
Trains	226,243
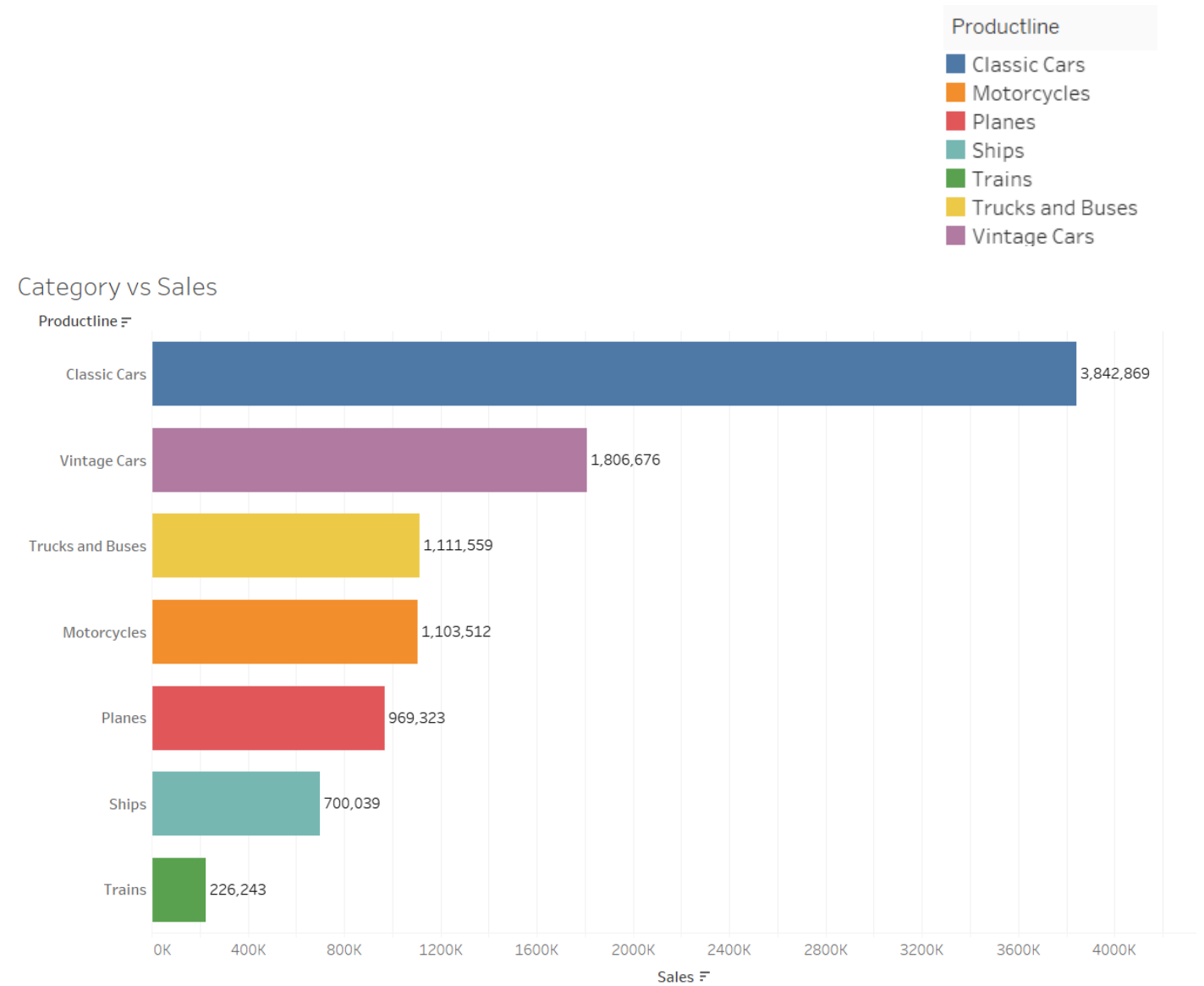
Sales

Quantity Ordered and Category

- Classic cars is the most ordered product followed by Vintage cars.

- Trains are the least ordered product.

Category vs Quantity Ordered

Productline ⌐

Classic Cars	33,373
Vintage Cars	20,059
Motorcycles	11,080
Planes	10,636
Trucks and Buses	10,579
Ships	7,989
Trains	2,712

Quantityordered ⌐

Productline
- Classic Cars
- Motorcycles
- Planes
- Ships
- Trains
- Trucks and Buses
- Vintage Cars
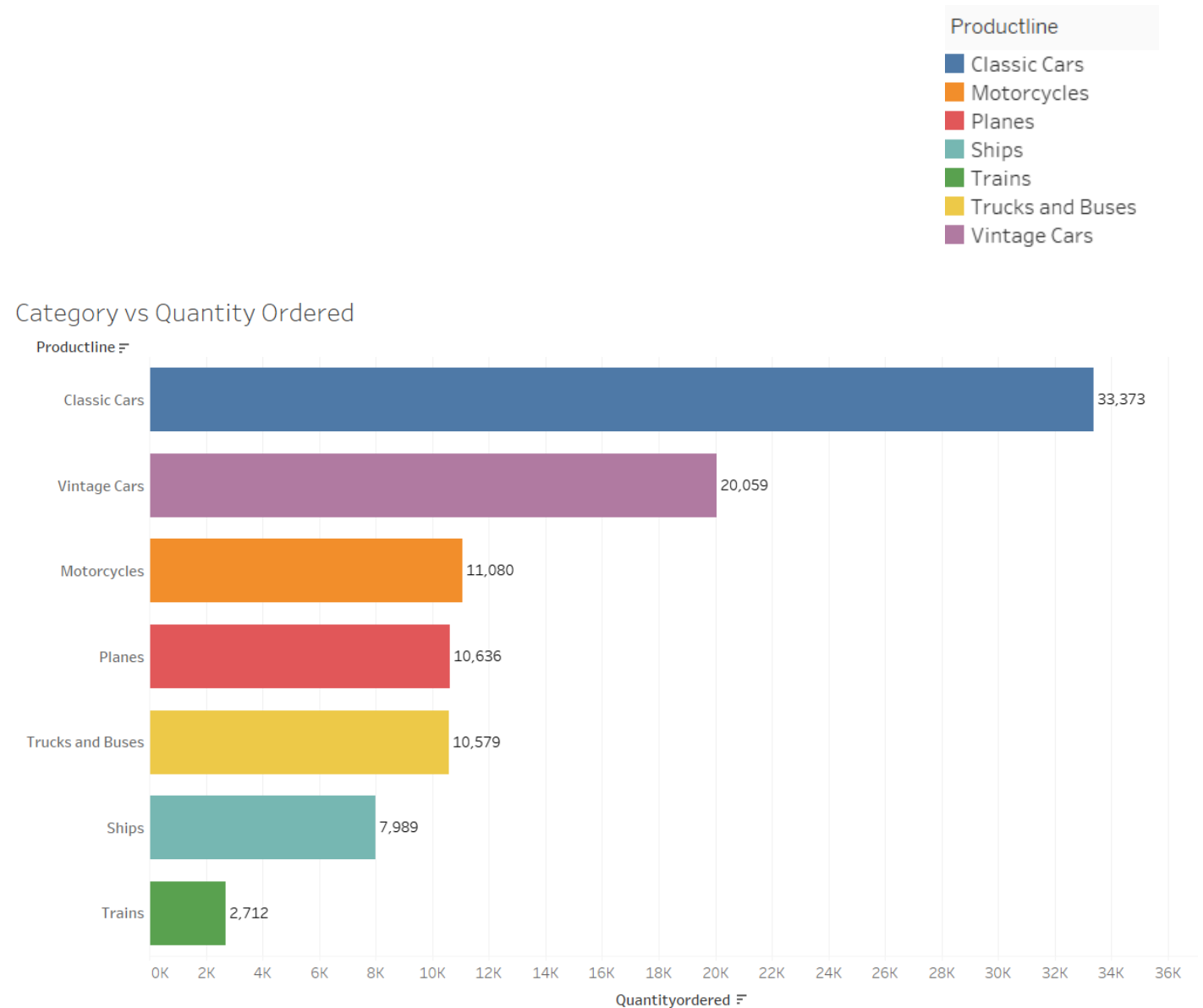
Sales across deal size

- Sales is the highest in medium deal size compared to small and large deal size.

- Medium deal size deliver the highest sales.

- The least orders and sales are of large deal size.

Pie chart - Deal size and Sales

Large 1,258,956
Small 2,570,034
Medium 5,931,231

Dealsize
- Large
- Medium
- Small

Productline
- Classic Cars
- Motorcycles
- Planes
- Ships
- Trains
- Trucks and Buses
- Vintage Cars
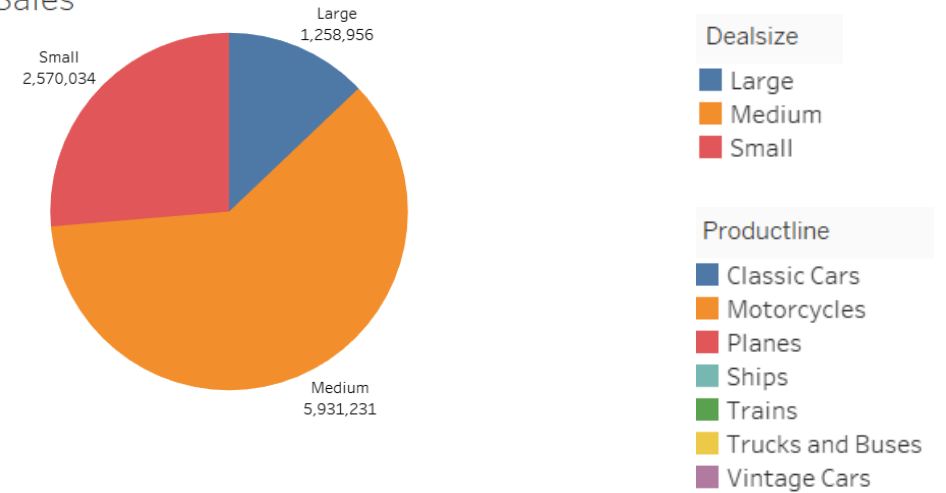
Sales vs Category and Deal size

Dealsize / Productline

Medium
- Classic Cars: 2,376,627
- Vintage Cars: 1,038,236
- Trucks and Buses: 809,050
- Motorcycles: 643,014
- Planes: 535,071
- Ships: 416,338
- Trains: 112,895

Small
- Classic Cars: 686,151
- Vintage Cars: 577,917
- Planes: 362,221
- Motorcycles: 309,419
- Ships: 283,702
- Trucks and Buses: 246,253
- Trains: 104,371

Large
- Classic Cars: 780,091
- Vintage Cars: 190,522
- Motorcycles: 151,079
- Planes: 72,032
- Trucks and Buses: 56,256
- Trains: 8,977

Sales across Customer

- Euro Shopping Channel is the top most customer in terms of sales.

- The sum of sales from Euro Shopping Channel customer alone is 912,294.

SUM(Sales)
9,129   912,294

Top Customer and Sales

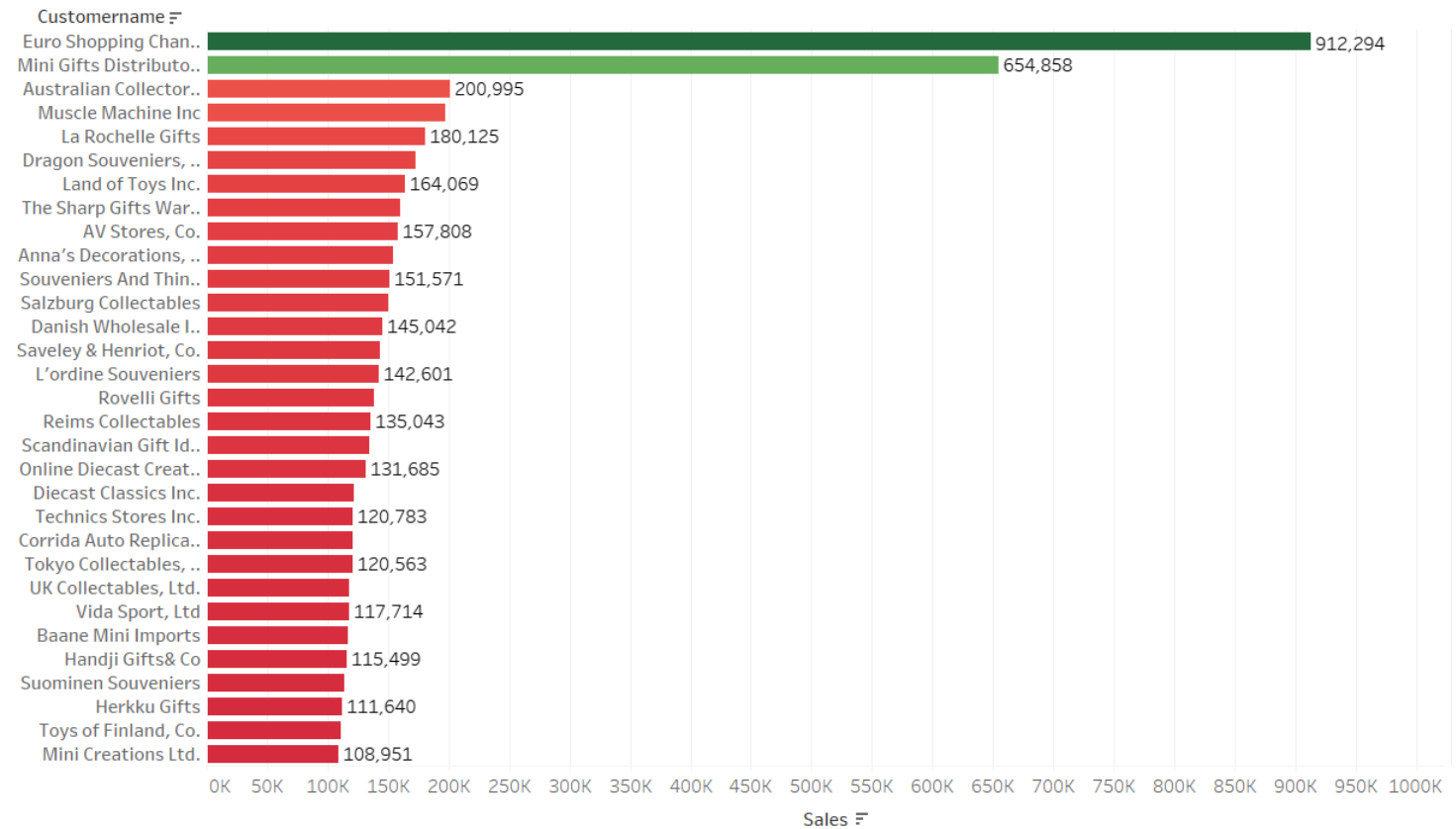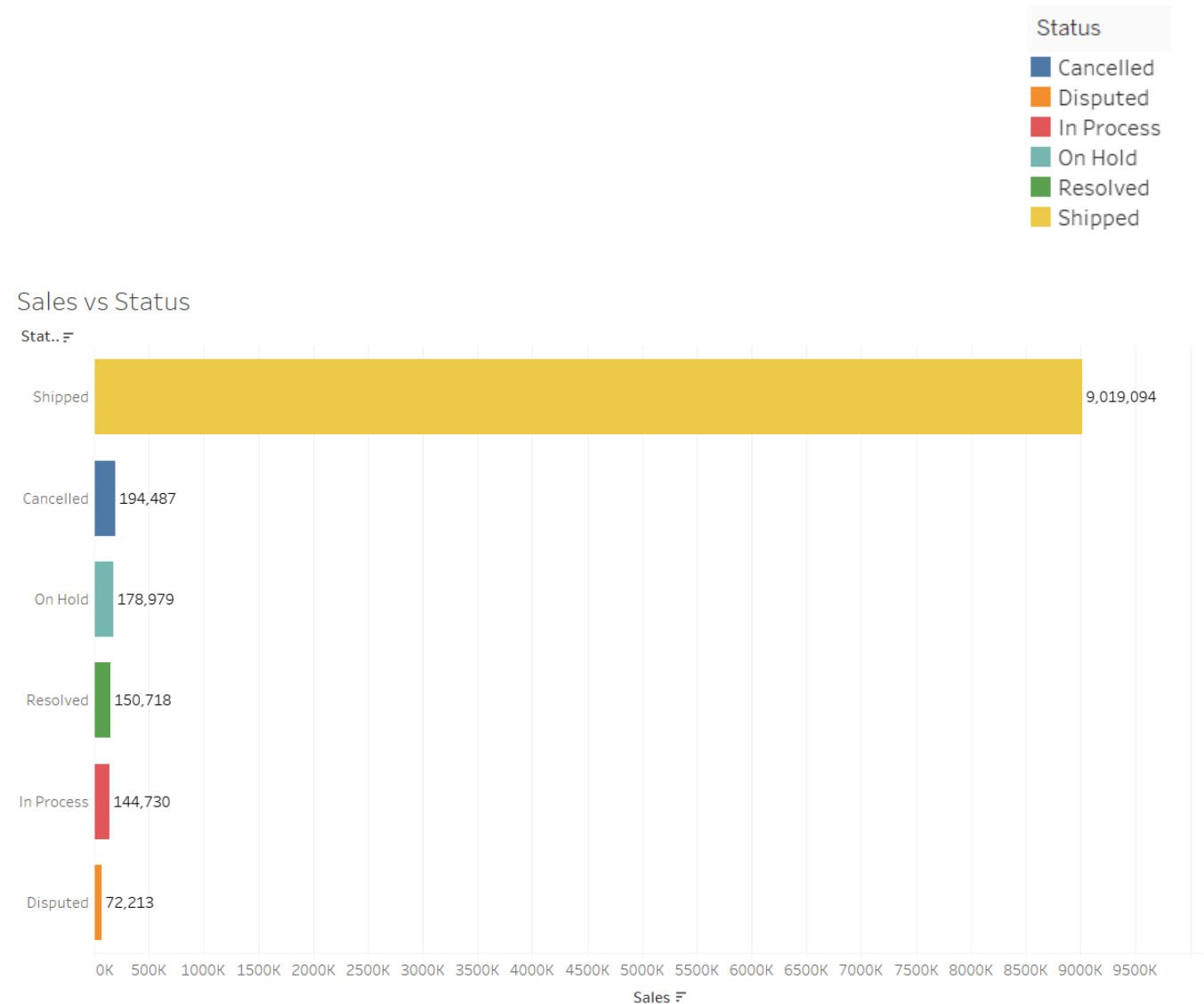| Customername	Sales
Euro Shopping Chan..	912,294
Mini Gifts Distributo..	654,858
Australian Collector..	200,995
Muscle Machine Inc	
La Rochelle Gifts	180,125
Dragon Souveniers, ..	
Land of Toys Inc.	164,069
The Sharp Gifts War..	
AV Stores, Co.	157,808
Anna's Decorations, ..	
Souveniers And Thin..	151,571
Salzburg Collectables	
Danish Wholesale I..	145,042
Saveley & Henriot, Co.	
L'ordine Souveniers	142,601
Rovelli Gifts	
Reims Collectables	135,043
Scandinavian Gift Id..	
Online Diecast Creat..	131,685
Diecast Classics Inc.	
Technics Stores Inc.	120,783
Corrida Auto Replica..	
Tokyo Collectables, ..	120,563
UK Collectables, Ltd.	
Vida Sport, Ltd	117,714
Baane Mini Imports	
Handji Gifts& Co	115,499
Suominen Souveniers	
Herkku Gifts	111,640
Toys of Finland, Co.	
Mini Creations Ltd.	108,951

34

Pareto Chart of Customer and Sales

- From the pareto chart, it is seen that a major part of the sales for the company comes from 4-5 customers.

- The top 5 customers contribute to 21.99% of the total sales of the company.

Pareto chart - Customer and Sales
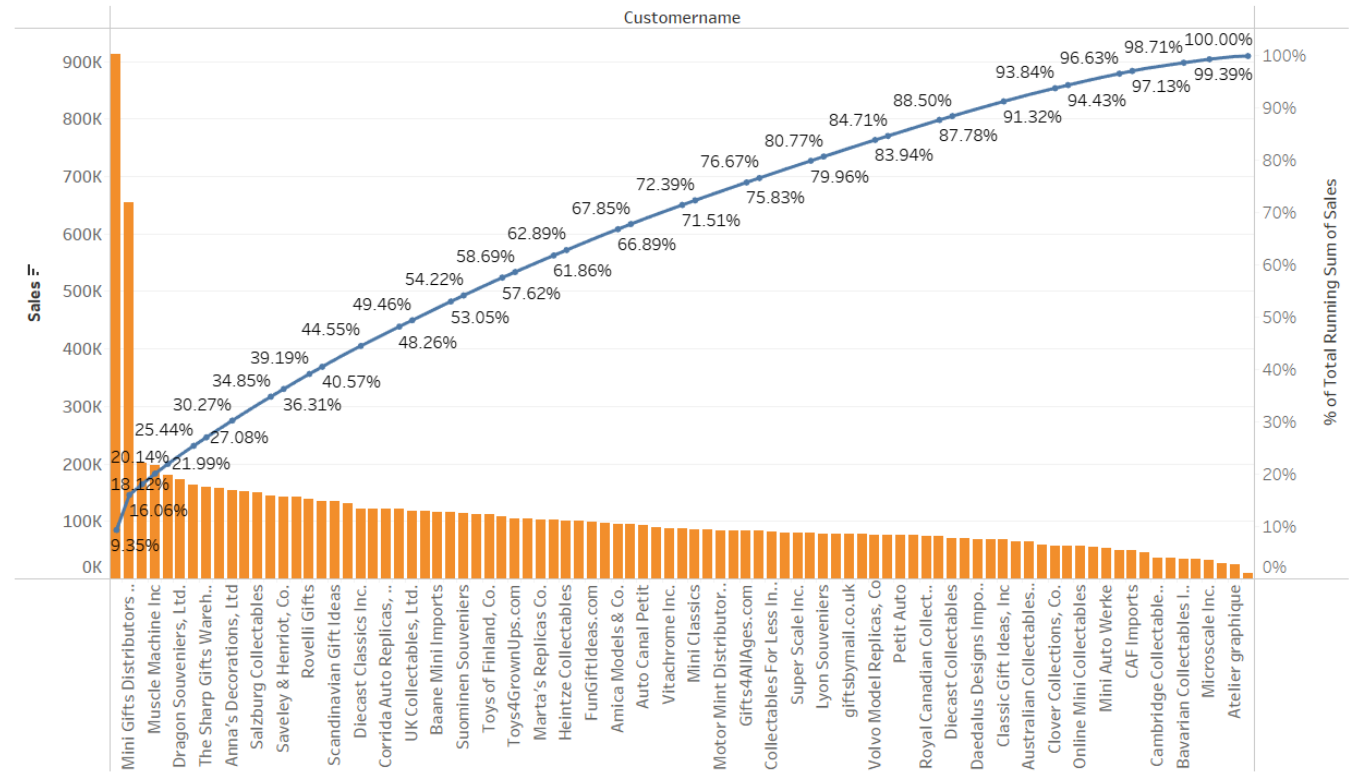
Sales across Status

- Majority of the orders are shipped.

- There are orders that have been cancelled. Attention should be given to check the reason behind the cancellation and rectify it.

- Some orders have been resolved but there are a few 'in process' and 'disputed' status.

- Disputed orders should be resolved at the earliest so that there is no negative feedback from the customer.

Status
- Cancelled
- Disputed
- In Process
- On Hold
- Resolved
- Shipped

Sales vs Status

Status	Sales
Shipped	9,019,094
Cancelled	194,487
On Hold	178,979
Resolved	150,718
In Process	144,730
Disputed	72,213

Summary of the inferences from the above analysis

Trends in sales show that sales in the maximum in the 4th quarter.

USA has the highest sales country wise.

Madrid has the highest sales city wise.

Classic cars has the highest sales and it is also the highest ordered product among all other products. Trains have the lowest sales. The sales of trucks and buses, motorcycles and planes can be increased.

Medium deal size deliver much higher sales compared to small and large deal size.

Euro Shopping Channel is the best customer in terms of sales contributing 9.35% of the total sales of the company.

Majority of the orders are shipped.

Measures to improve sales

The 4th quarter and especially November month has the highest sales. This can be taken to our advantage and new models can be introduced during this season to increase the sales.

To increase sales across all quarters, discounts and offers can be given for every purchase with a valid time period. This way we can improve sales and the customers will also be content with the discounted price.

Warranty can have a great impact on the sales. If warranty is offered, this will encourage the customers to buy even if it is expensive as the damage will be covered in the warranty.

Discounts and promotions can be given to increase the sales in countries with low sales. Low EMI options can promote sales in these countries.

The sales of large deal size remain low. Therefore the company should take efforts to promote and attract customers to buy orders of large size.

Attention should be given to cancelled orders and rectify the reason for cancellation. This can increase the sales. Disputed orders should be resolved at the earliest. This will give a positive feel for the customer.

Marketing and promoting the products can play a major role to increase sales. Paid partnerships and collaborations with social media influencers can have a great impact and also have a good reach to the people. This is a way of advertising the brand which can increase sales.

Customer Segmentation using RFM Analysis (4 segments)

What is RFM?

- The "RFM" in RFM analysis stands for recency, frequency and monetary value.

- It is a way to use data based on existing customer behavior to predict how a new customer is likely to act in the future.

- The analysis is based on the marketing adage that "80% of your business comes from 20% of your customers."

- The analysis enables marketers to identify the best customers and increase revenue by performing targeted marketing campaigns (offers, messages). This leads to increased response rates, customer retention, customer satisfaction, and customer lifetime value (CLV).

RFM Metrics

RECENCY
The freshness of the customer activity, be it purchases or visits

E.g. Time since last order or last engaged with the product

FREQUENCY
The frequency of the customer transactions or visits

E.g. Total number of transactions or average time between transactions/ engaged visits

MONETARY
The intention of customer to spend or purchasing power of customer

E.g. Total or average transactions value

What all parameters used and assumptions made

KNIME is used to perform the RFM Analysis.

Parameters used for RFM Analysis: Customer name, Order number, Quantity ordered, Price each, Sales, Order date and Days since last order

Assumptions made:

1. Recency is calculated by subtracting order date from the maximum order date

2. Frequency is calculated by taking the count of quantity ordered for a particular customer

3. Monetary is calculated by taking the sum of sales made by a particular customer

Recency = [Max (order date) – order date]

Frequency = Count (Quantity ordered)

Monetary = Sum (Sales)

Output table head

Table "default" - Rows: 89 | Spec - Columns: 15 | Properties | Flow Variables

Row ID	CUSTOMERNAME	First*(...	Count*...	Mean(P...	Sum(SA...	Max(O...	First*(...	First(O...	Recency	Freque...	Monetary	Recenc...	Freque...	Moneta...	RFM Sc...
Row32	Euro Shopping Channel	10417	259	97.383	912,294.11	2020-05-31	42	2020-05-13	18	259	912,294.11	4	4	4	444
Row3	Anna's Decorations, Ltd	10391	46	106.424	153,996.13	2020-03-09	131	2020-03-09	0	46	153,996.13	4	4	4	444
Row57	Online Diecast Creati...	10322	34	108.302	131,685.3	2019-11-04	253	2019-11-04	0	34	131,685.3	4	4	4	444
Row83	UK Collectables, Ltd.	10403	29	108.536	118,008.27	2020-04-08	76	2020-04-08	0	29	118,008.27	4	3	3	433
Row74	Suominen Souveniers	10363	30	110.405	113,961.15	2020-01-06	317	2020-01-06	0	30	113,961.15	4	3	3	433
Row60	Oulu Toy Supplies, Inc.	10373	32	95.119	104,370.38	2020-01-31	241	2020-01-31	0	32	104,370.38	4	3	3	433
Row38	Heintze Collectables	10314	27	112.268	100,595.55	2019-10-22	414	2019-10-22	0	27	100,595.55	4	3	3	433
Row33	FunGiftIdeas.com	10388	26	109.587	98,923.73	2020-03-03	111	2020-03-03	0	26	98,923.73	4	2	3	423
Row21	Collectable Mini Desig...	10222	25	91.535	87,489.23	2019-02-26	575	2019-02-19	7	25	87,489.23	4	2	2	422
Row51	Mini Classics	10308	26	95.03	85,555.99	2019-10-15	295	2019-10-15	0	26	85,555.99	4	2	2	422
Row22	Collectables For Less ...	10369	24	97.237	81,577.98	2020-01-20	179	2020-01-20	0	24	81,577.98	4	2	2	422
Row88	giftsbymail.co.uk	10316	26	88.979	78,240.84	2019-11-01	327	2019-11-01	0	26	78,240.84	4	2	2	422
Row65	Royal Canadian Colle...	10283	26	86.449	74,634.85	2019-08-20	397	2019-08-20	0	26	74,634.85	4	2	2	422
Row62	Quebec Home Shoppi...	10411	22	104.515	74,204.79	2020-05-01	79	2020-05-01	0	22	74,204.79	4	2	2	422
Row50	Mini Caravy	10405	19	99.27	80,438.48	2020-04-14	447	2020-04-14	0	19	80,438.48	4	1	2	412
Row75	Super Scale Inc.	10245	17	128.452	79,472.07	2019-05-04	430	2019-05-04	0	17	79,472.07	4	1	2	412
Row8	Auto Assoc. & Cie.	10304	18	99.488	64,834.32	2019-10-11	275	2019-10-11	0	18	64,834.32	4	1	1	411
Row70	Signal Collectibles Ltd.	10219	15	95.396	50,218.51	2019-02-10	836	2019-02-10	0	15	50,218.51	4	1	1	411
Row15	CAF Imports	10231	13	104.963	49,642.05	2019-03-19	625	2019-03-19	0	13	49,642.05	4	1	1	411
Row87	West Coast Collectab...	10215	13	88.308	46,084.64	2019-01-29	523	2019-01-29	0	13	46,084.64	4	1	1	411
Row12	Bavarian Collectables...	10296	14	84.289	34,993.92	2019-09-15	801	2019-09-15	0	14	34,993.92	4	1	1	411
Row48	Microscale Inc.	10319	10	88.494	33,144.93	2019-11-03	453	2019-11-03	0	10	33,144.93	4	1	1	411
Row14	Boards & Toys Co.	10376	3	89.807	9,129.35	2020-02-08	410	2020-02-08	0	3	9,129.35	4	1	1	411
Row78	The Sharp Gifts Ware...	10400	40	93.376	160,010.27	2020-04-22	182	2020-04-01	21	40	160,010.27	3	4	4	344
Row0	AV Stores, Co.	10306	51	91.085	157,807.81	2019-11-17	421	2019-10-14	34	51	157,807.81	3	4	4	344
Row72	Souveniers And Thin...	10361	46	95.189	151,570.98	2020-05-29	186	2019-12-17	164	46	151,570.98	3	4	4	344
Row67	Salzburg Collectables	10341	40	101.398	149,798.63	2020-05-17	208	2019-11-24	175	40	149,798.63	3	4	4	344
Row68	Saveley & Henriot, Co.	10194	41	100.548	142,874.25	2019-03-02	586	2018-11-25	97	41	142,874.25	3	4	4	344
Row69	Scandinavian Gift Ideas	10291	38	97.597	134,259.33	2020-03-03	307	2019-09-08	177	38	134,259.33	3	4	4	344
Row27	Diecast Classics Inc.	10318	31	108.566	122,138.14	2020-05-30	228	2019-11-02	210	31	122,138.14	3	3	3	334
Row84	Vida Sport, Ltd	10225	31	112.075	117,713.56	2019-08-30	624	2019-02-22	189	31	117,713.56	3	3	3	333
Row11	Baane Mini Imports	10309	32	108.574	116,599.19	2019-11-05	245	2019-10-15	21	32	116,599.19	3	3	3	333
Row81	Toys of Finland, Co.	10299	30	105.752	111,250.38	2020-02-09	259	2019-09-30	132	30	111,250.38	3	3	3	333
Row47	Marta's Replicas Co.	10285	27	107.072	103,080.38	2019-10-13	292	2019-08-27	47	27	103,080.38	3	3	3	333
Row71	Signal Gift Stores	10278	29	91.429	82,751.08	2019-11-29	741	2019-08-06	115	29	82,751.08	3	3	2	332
Row2	Amica Models & Co.	10280	26	110.853	94,117.26	2019-09-09	328	2019-08-17	23	26	94,117.26	3	2	3	323
Row85	Vitachrome Inc.	10237	25	106.179	88,041.26	2019-11-05	432	2019-04-05	214	25	88,041.26	3	2	3	323
Row28	Diecast Collectables	10207	18	101.783	70,859.78	2019-04-26	672	2018-12-09	138	18	70,859.78	3	1	2	312

KNIME workflow

KNIME is used to perform the RFM Analysis and the workflow diagram is shown on the right.
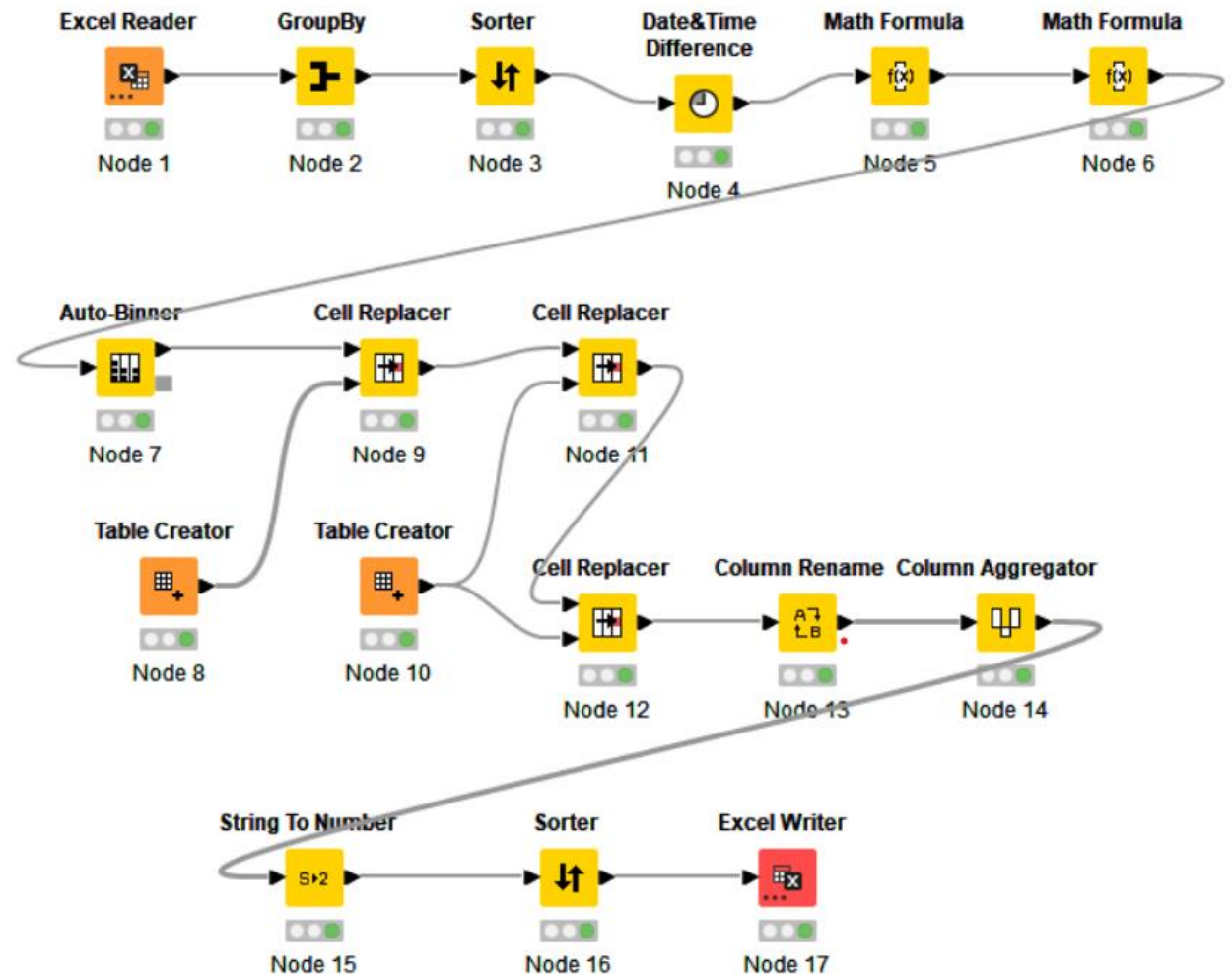
Inferences from RFM Analysis and identified segments

Who are your best customers?

- Best customers are those who are extremely recent, very active/frequent with a high monetary value.

- The company should not loose these customers.

- The best customers will have a RFM score of:

✓ Extremely recent – therefore recency score of 4, 3

✓ Very frequent – therefore a frequency value of 4

✓ High monetary value – therefore a monetary value of 4

	CUSTOMERNAME	RFM Score
0	Euro Shopping Channel	444.0
1	Anna's Decorations, Ltd	444.0
2	Online Diecast Creations Co.	444.0
23	The Sharp Gifts Warehouse	344.0
24	AV Stores, Co.	344.0
25	Souveniers And Things Co.	344.0
26	Salzburg Collectables	344.0

Which customers are on the verge of churning?

- Customers with moderate recency, more or less frequent (not very active) and inconsistent monetary value are those customers on the verge of churning.

- These customers are the risk group. If nothing is done to improve the sales of this group, the company is at the risk of loosing them. Sales can be improved in this group by giving offers, promotions, discounts etc.,

- Customers on the verge of churning will have a RFM score of:

✓ Moderate recency – therefore recency score of 3, 2

✓ More or less frequent – therefore a frequency value of 3, 2

✓ Inconsistent monetary value – therefore a monetary value of 3, 2

	CUSTOMERNAME	RFM Score
30	Vida Sport, Ltd	333.0
31	Baane Mini Imports	333.0
32	Toys of Finland, Co.	333.0
33	Marta's Replicas Co.	333.0
35	Amica Models & Co.	323.0
36	Vitachrome Inc.	323.0
51	Tokyo Collectables, Ltd	233.0

Who are your lost customers?

- Lost customers are inactive customers. They are customers with low recency, low frequency and low monetary value.

- It is okay to let go off these customers as there is no use in investing in them.

- The lost customers will have a RFM score of:

✓ Very low recency – therefore recency score of 1

✓ Not frequent – therefore a frequency value of 2, 1

✓ Very low monetary value – therefore a monetary value of 2, 1

	CUSTOMERNAME	RFM Score
80	Marseille Mini Autos	122.0
81	Australian Collectables, Ltd	121.0
82	Lyon Souveniers	112.0
83	Volvo Model Replicas, Co	112.0
84	Alpha Cognac	111.0
85	Australian Gift Network, Co	111.0
86	Gift Ideas Corp.	111.0

Who are your loyal customers?

- Loyal customers are those who are extremely recent, moderate frequent with a moderate monetary value.

- They are valuable customers. The company can invest in them by providing offers.

- The loyal customers will have a RFM score of:

✓ Extremely recent – therefore recency score of 4, 3

✓ Moderate frequency – therefore a frequency value of 4, 3

✓ Moderate monetary value – therefore a monetary value of 4, 3

	CUSTOMERNAME	RFM Score
0	Euro Shopping Channel	444.0
1	Anna's Decorations, Ltd	444.0
2	Online Diecast Creations Co.	444.0
3	UK Collectables, Ltd.	433.0
4	Suominen Souveniers	433.0
5	Oulu Toy Supplies, Inc.	433.0
6	Heintze Collectables	433.0

Inferences

- Best customers are those who are extremely recent, very active/frequent with a high monetary value. The company should not loose these customers. They are the gold customers.

- Customers with moderate recency, more or less frequent (not very active) and inconsistent monetary value are those customers on the verge of churning. These customers are the risk group. If nothing is done to improve the sales of this group, the company is at the risk of loosing them. Sales can be improved in this group by giving offers, promotions, discounts etc.,

- Lost customers are inactive customers. They are customers with low recency, low frequency and low monetary value. It is okay to let go off these customers as there is no use in investing in them.

- Loyal customers are those who are extremely recent, moderate frequent with a moderate monetary value. They are valuable customers. The company can invest in them by providing offers.

TABLEAU Link

https://public.tableau.com/app/profile/athulya4213/viz/MRAProjectMilestone1_16527395828840/Salesvs Status

Thank you