MARKETING & RETAIL ANALYTICS MILESTONE 1

PROJECT BY :- AKASH KAMBLE, JUNE - A BATCH 2022

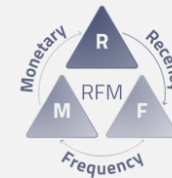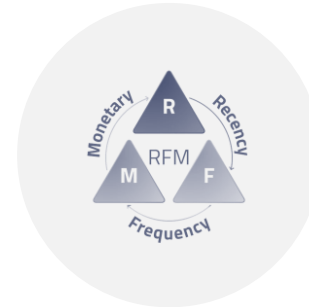# CONTENTS

CONTENTS

Agenda & Executive Summary of the Data

Exploratory Analysis and Inferences

Customer Segmentation using RFM Analysis

Inferences from RFM Analysis and Identified Segments

PROBLEM STATEMENT

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

EXECUTIVE SUMMARY & INTRODUCTION

❖ **Executive Summary :-**

- The intend of the study is to find the underlying buying patterns of the customers of an automobile part manufacturers based on the past 3 years of the Company's transaction data and hence recommend customized marketing strategies for different segments of customers using RFM analysis.

❖ **Introduction :-**

- In this study, we will be using different tools for the analysis such as Python, Tableau and KNIME.

- We will go-through EDA, sales trends, customer segmentation using RFM analysis and eventually drawing inferences based upon the analysis performed.
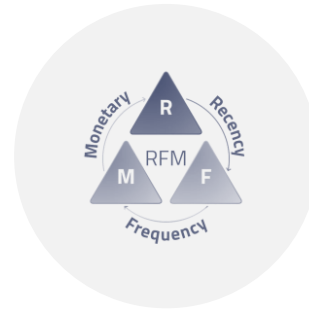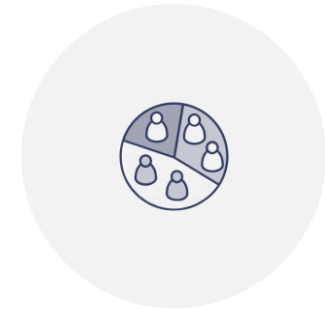
DATA SUMMARY

- **Data Description :-**

Data	Description
PRICEEACH	Price of each item
ORDERLINENUMBER	Order Line
SALES	Sales amount
DAYS_SINCE_LASTORDER	No. of Days Since Last Order
STATUS	Status of order like Shipped or not
PRODUCTLINE	Product line – CATEGORY
MSRP	Manufacturer's Suggested Retail Price
DEALSIZE	Size of the deal based on Quantity and Item Price

- **Data Info :-**

1. There are total 2747 number of rows and 20 number of columns in the dataset.

2. Data consists of datetime64, float64, int64 and object datatypes.

3. There are no missing values and no duplicate rows in the data.

```
 #   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
```

Akash_Kamble_MRA_1.ipynb

STATISTICAL SUMMARY & ASSUMPTIONS

- **Data Description :-**

	count	unique	top	freq	first	last	mean	std	min	25%	50%	75%	max
ORDERNUMBER	2747.0	NaN	NaN	NaN	NaT	NaT	10259.761558	91.877521	10100.0	10181.0	10264.0	10334.5	10425.0
QUANTITYORDERED	2747.0	NaN	NaN	NaN	NaT	NaT	35.103021	9.762135	6.0	27.0	35.0	43.0	97.0
PRICEEACH	2747.0	NaN	NaN	NaN	NaT	NaT	101.098951	42.042548	26.88	68.745	95.55	127.1	252.87
ORDERLINENUMBER	2747.0	NaN	NaN	NaN	NaT	NaT	6.491081	4.230544	1.0	3.0	6.0	9.0	18.0
SALES	2747.0	NaN	NaN	NaN	NaT	NaT	3553.047583	1838.953901	482.13	2204.35	3184.8	4503.095	14082.8
ORDERDATE	2747	246	2018-11-14 00:00:00	38	2018-01-06	2020-05-31	NaN	NaN	NaN	NaN	NaN	NaN	NaN
DAYS_SINCE_LASTORDER	2747.0	NaN	NaN	NaN	NaT	NaT	1757.085912	819.280576	42.0	1077.0	1761.0	2436.5	3562.0
STATUS	2747	6	Shipped	2541	NaT	NaT	NaN	NaN	NaN	NaN	NaN	NaN	NaN
PRODUCTLINE	2747	7	Classic Cars	949	NaT	NaT	NaN	NaN	NaN	NaN	NaN	NaN	NaN
MSRP	2747.0	NaN	NaN	NaN	NaT	NaT	100.691664	40.114802	33.0	68.0	99.0	124.0	214.0
PRODUCTCODE	2747	109	S18_3232	51	NaT	NaT	NaN	NaN	NaN	NaN	NaN	NaN	NaN
CUSTOMERNAME	2747	89	Euro Shopping Channel	259	NaT	NaT	NaN	NaN	NaN	NaN	NaN	NaN	NaN
PHONE	2747	88	(91) 555 94 44	259	NaT	NaT	NaN	NaN	NaN	NaN	NaN	NaN	NaN
ADDRESSLINE1	2747	89	C/ Moralzarzal, 86	259	NaT	NaT	NaN	NaN	NaN	NaN	NaN	NaN	NaN
CITY	2747	71	Madrid	304	NaT	NaT	NaN	NaN	NaN	NaN	NaN	NaN	NaN
POSTALCODE	2747	73	28034	259	NaT	NaT	NaN	NaN	NaN	NaN	NaN	NaN	NaN
COUNTRY	2747	19	USA	928	NaT	NaT	NaN	NaN	NaN	NaN	NaN	NaN	NaN
CONTACTLASTNAME	2747	76	Freyre	259	NaT	NaT	NaN	NaN	NaN	NaN	NaN	NaN	NaN
CONTACTFIRSTNAME	2747	72	Diego	259	NaT	NaT	NaN	NaN	NaN	NaN	NaN	NaN	NaN
DEALSIZE	2747	3	Medium	1349	NaT	NaT	NaN	NaN	NaN	NaN	NaN	NaN	NaN

- **Assumptions :-**

1. Currency considered for the dataset id USD.

Exploratory Analysis & Inferences

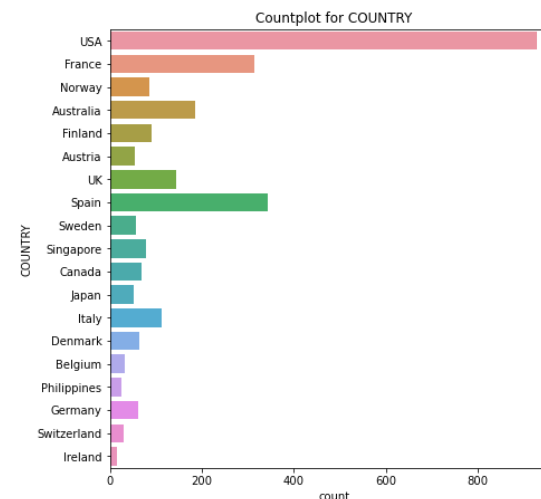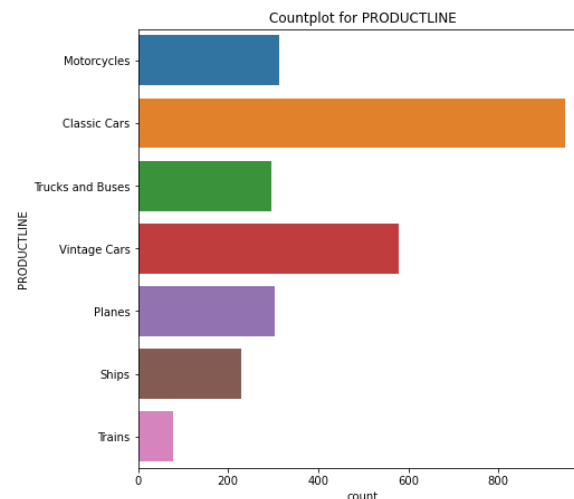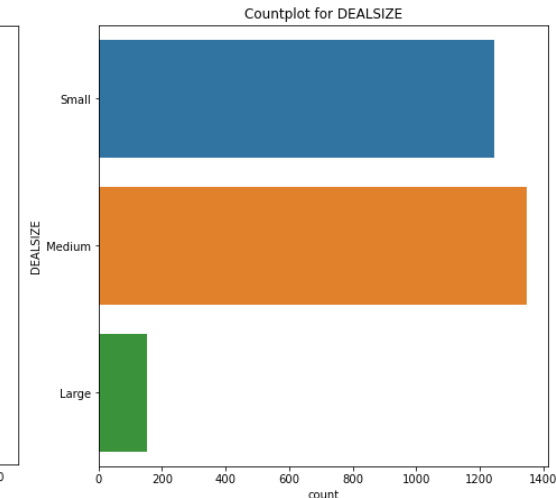- **Univariate Analysis :-**
 - Numerical Variables
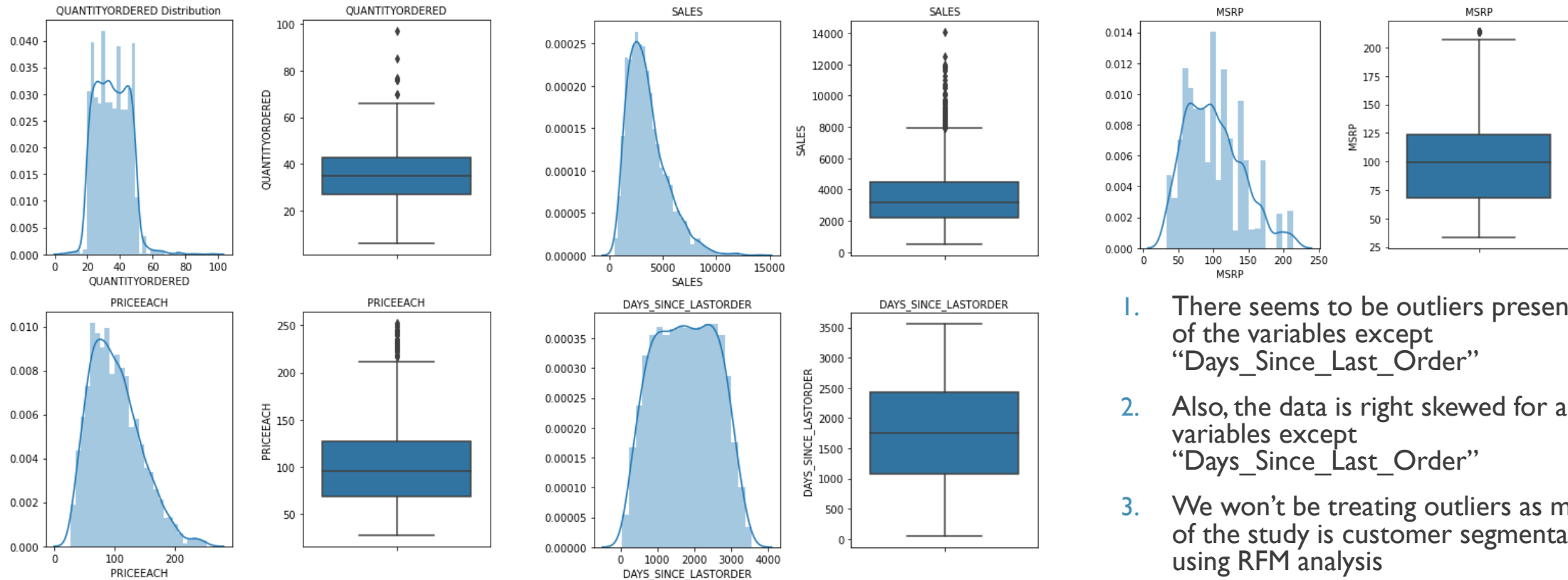
1. There seems to be outliers present in all of the variables except "Days_Since_Last_Order"

2. Also, the data is right skewed for all variables except "Days_Since_Last_Order"

3. We won't be treating outliers as motive of the study is customer segmentation using RFM analysis

Exploratory Analysis & Inferences

- **Univariate Analysis :-**
 o Major Categorical Variables

1. Maximum observations belongs to "Orderline numbe 1".

2. Most of the order status is "shipped".

3. "Large" deal size is the least whereas "Medium" is the large.

4. Most of the auto component products belongs to category "Classic Cars".

5. Maximum number of the customer comes from USA.
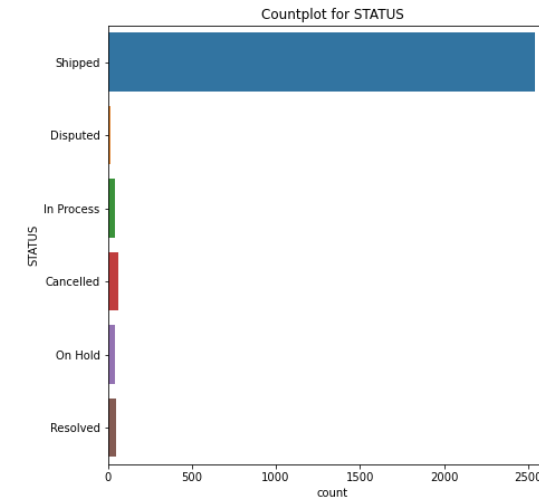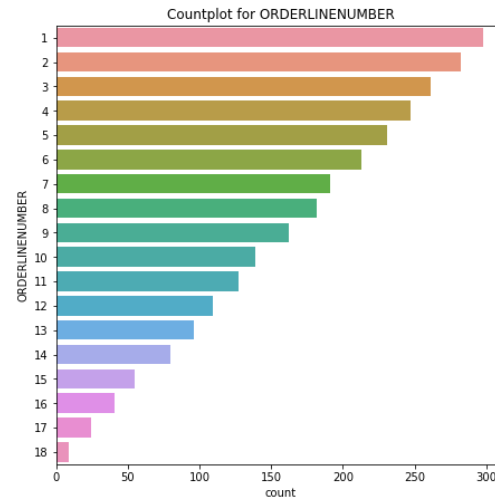
Exploratory Analysis & Inferences

- **Bivariate Analysis :-**

1. "Classic Cars" contribute the maximum sales among the all product line.

2. Similarly, customers belong to the country USA contribute the highest to the sales.

3. Among all product lines and countries, "Medium" Size deals yields highest revenue in comparison with other deals.
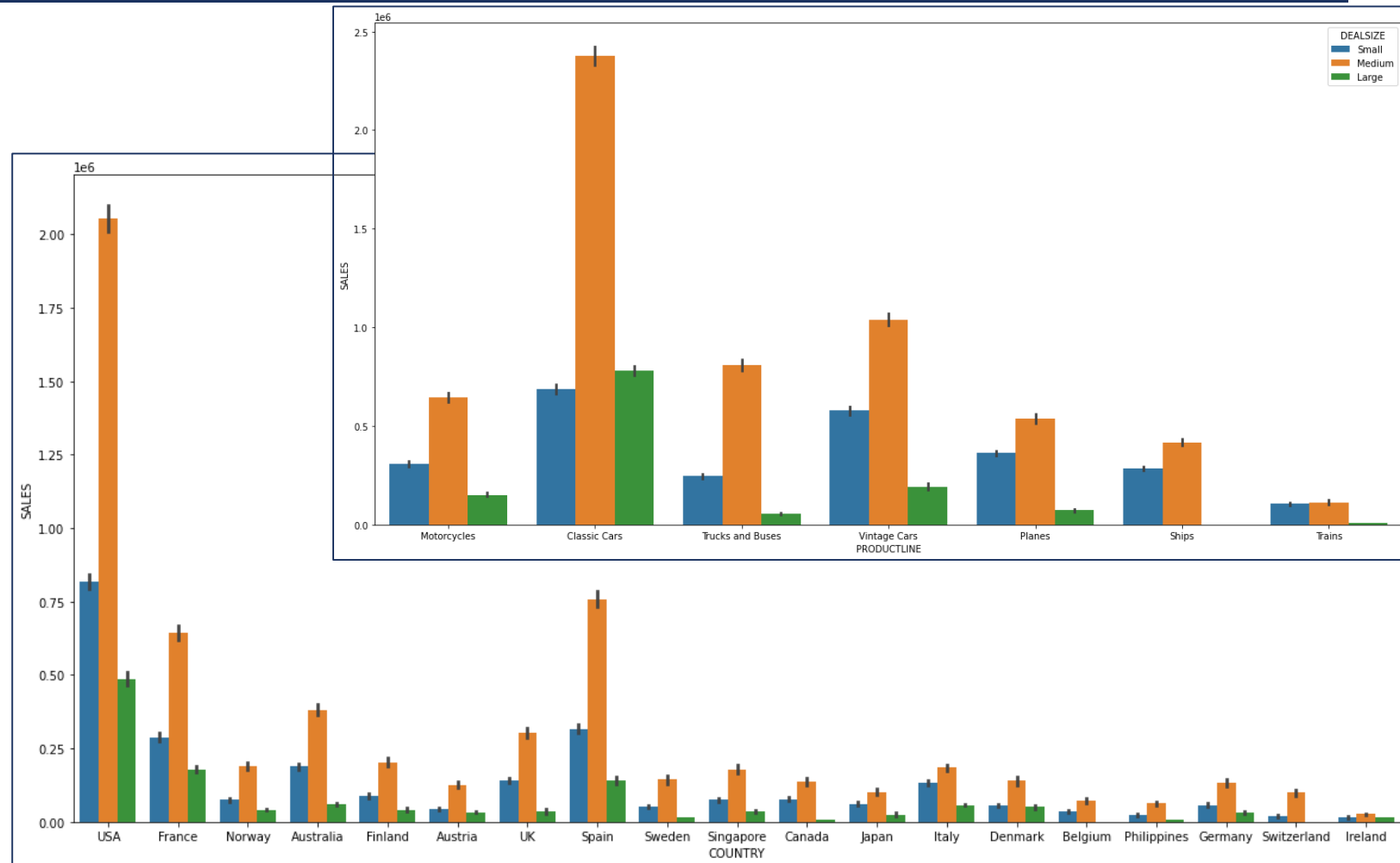
Exploratory Analysis & Inferences

- **Multi-variate Analysis :-**

1. There is correlation between sales and quantity ordered & price each, which is obvious.

2. Also, MSRP shows correlation with sales and price each and negative correlation with days since last order.

Heatmap
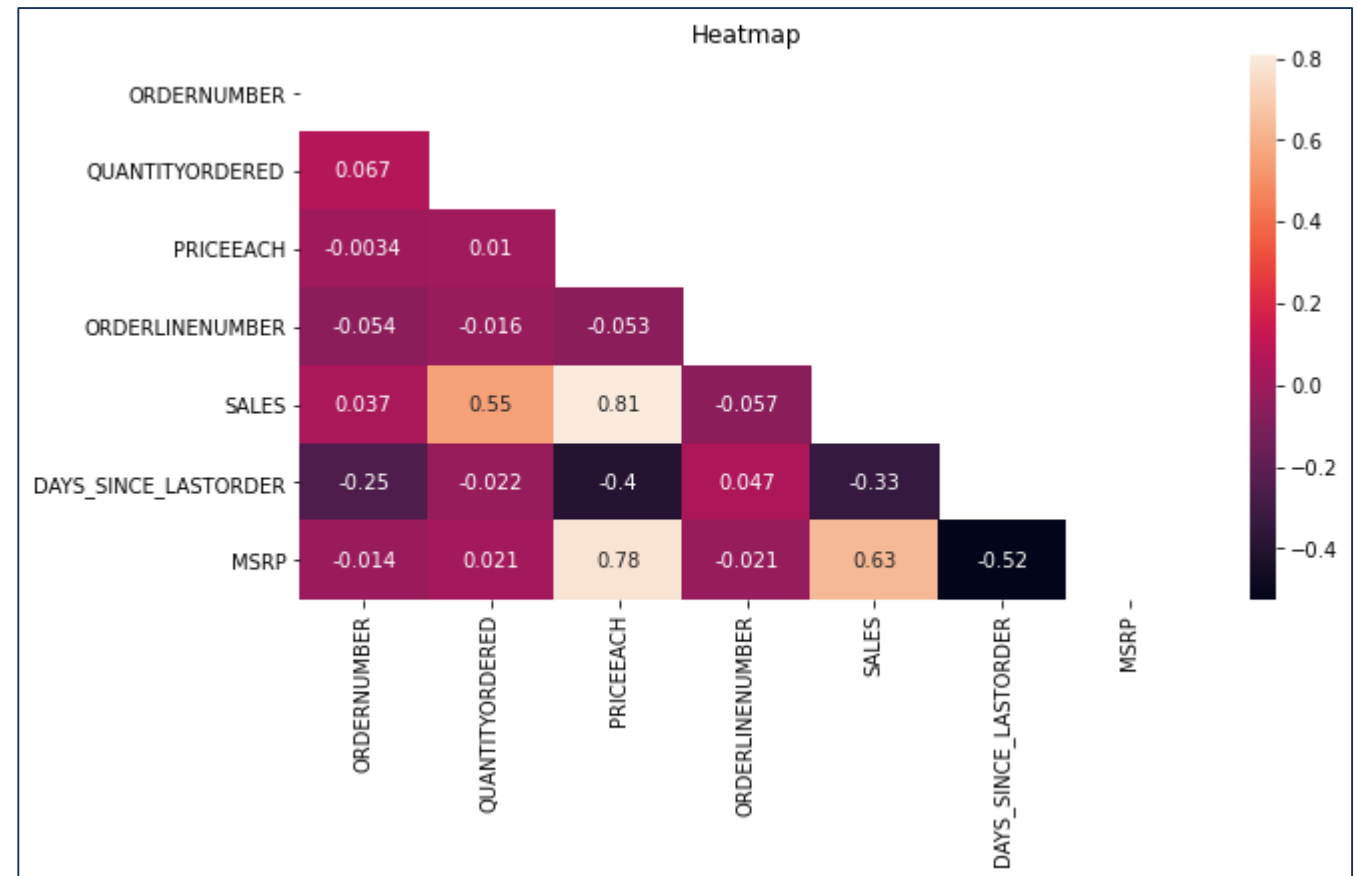
SALES TRENDS

- **Yearly Sales Trend:-**

Year Sales Trend

- **Quarterly Sales Trend:-**

Quarterly Sales Trend
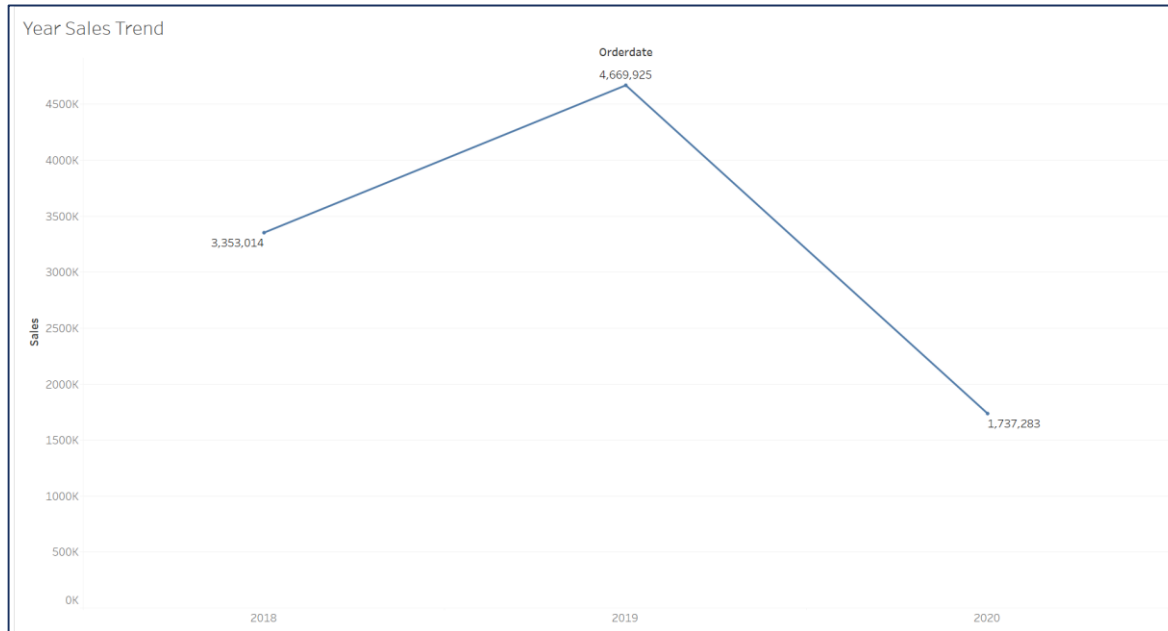
1. There is significant sales increase in 2019 over 2018.

2. In 2020, data is available for 5 months only. This is the reason we see a drop in sales in 2020 over 2019.
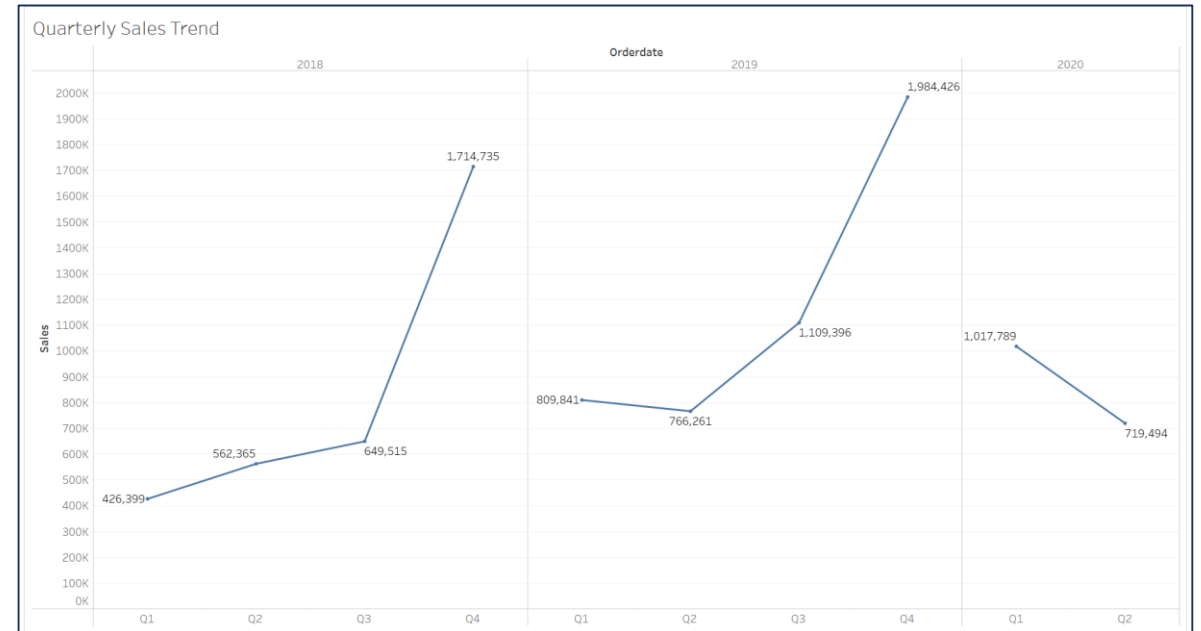
1. There seems to be a upward sales trend in each of the quarter except the Q1 & Q2 of 2020.

2. Q2 of 2020 shows a drop in sales as only 2 months are observed in Q2'2020.

3. Q1'2020 seems to be performing better than Q1'2018 and Q1'2019.

RFM_EDA_SalesTrends_MRA_1.twbx

SALES TRENDS

- **Monthly Sales Trend:-**
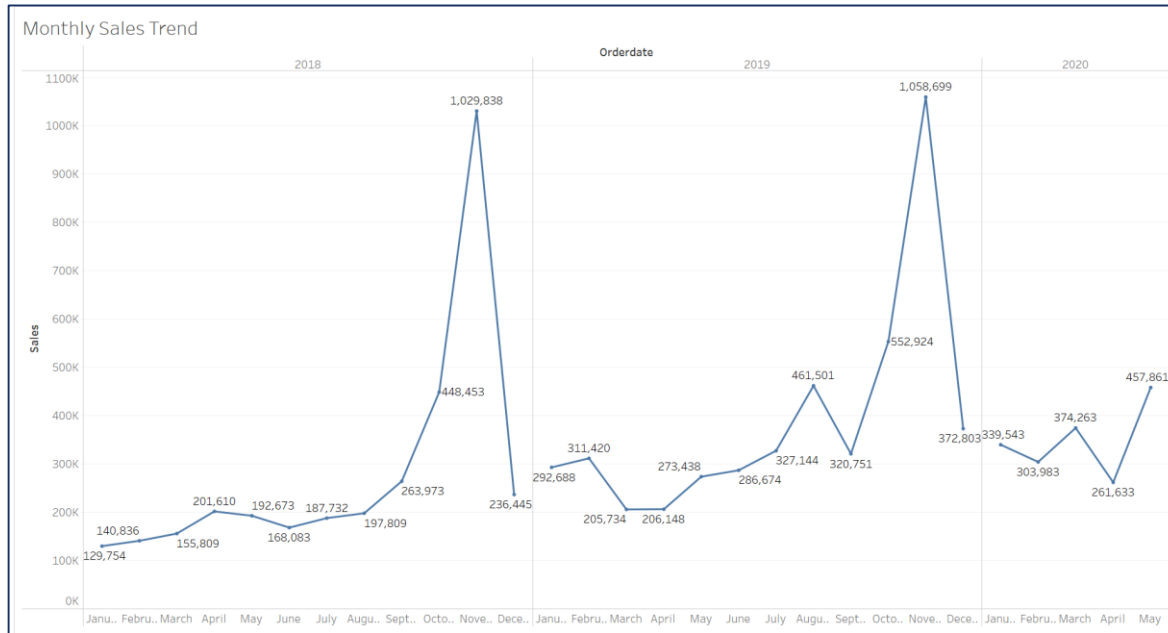
1. The monthly trend is clearly upward throughout the year.

2. However, December show a drastic fall in sales for 2018 and 2019. This could be due to the fact that the OEMs manufacture lower volumes of vehicle as customer doesn't prefer to buy vehicle in the last month as it have low resale value due to Model Year. If the end consumer of vehicle waits for one month and gets vehicle model of next year, its resale value improves.

- **Weekly Sales Trend:-**
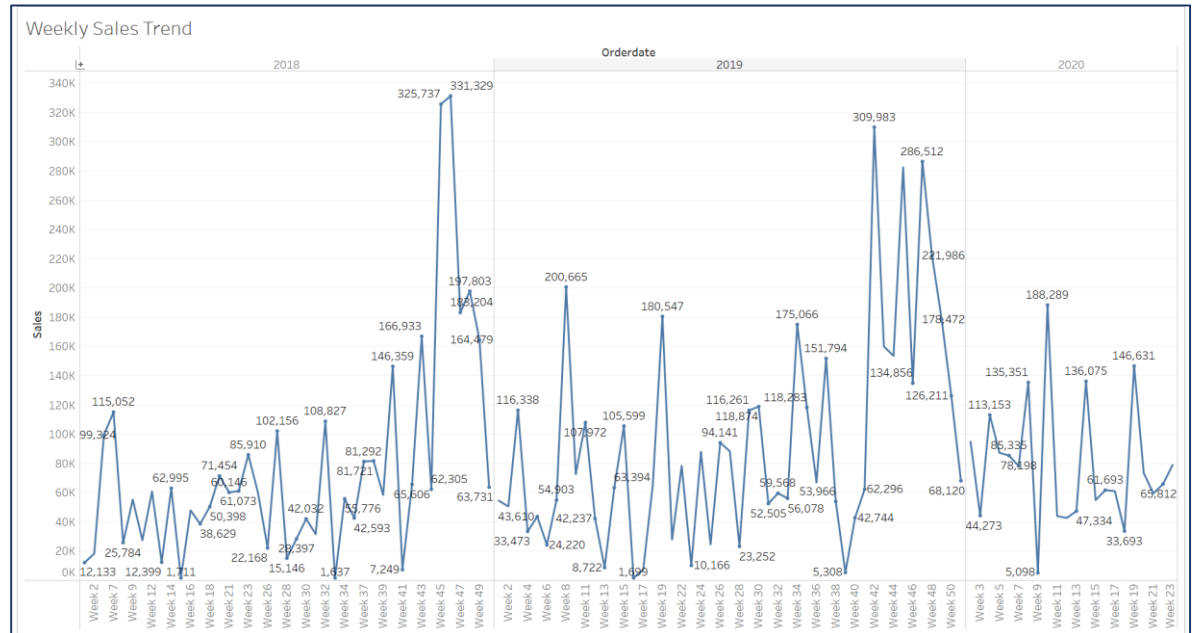
1. Weekly trend is noisy, still we could observe a uptrend in the sales.

2. There are many spikes visible in the weekly sales which show demand fluctuations.

SALES ACROSS CATEGORIES

- **Sales across Product Lines:-**

Sales across Product Lines

Productline	Sales
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
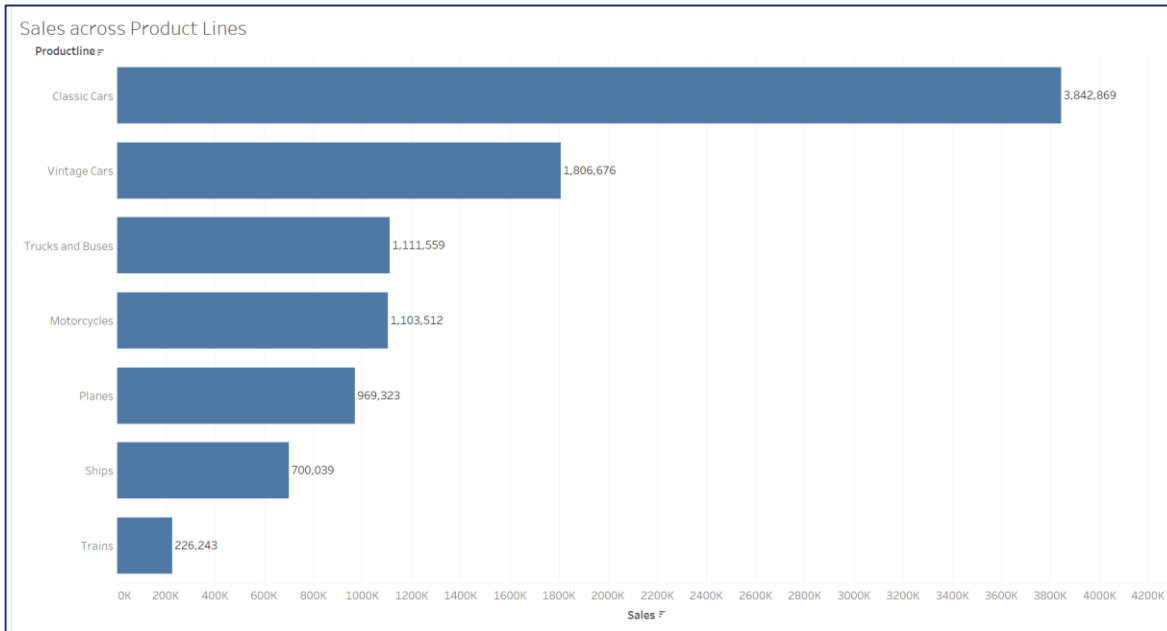
1. Sales contribution by the product line category is highest for **"Classic Cars"**.

2. "Classic Cars" and "Vintage Cars" together constitutes around **60% of the total sales**.

- **Sales across Deal Sizes:-**

Sales across Deal Sizes

Deal	Sales
Medium	5,931,231
Small	2,570,034
Large	1,258,956
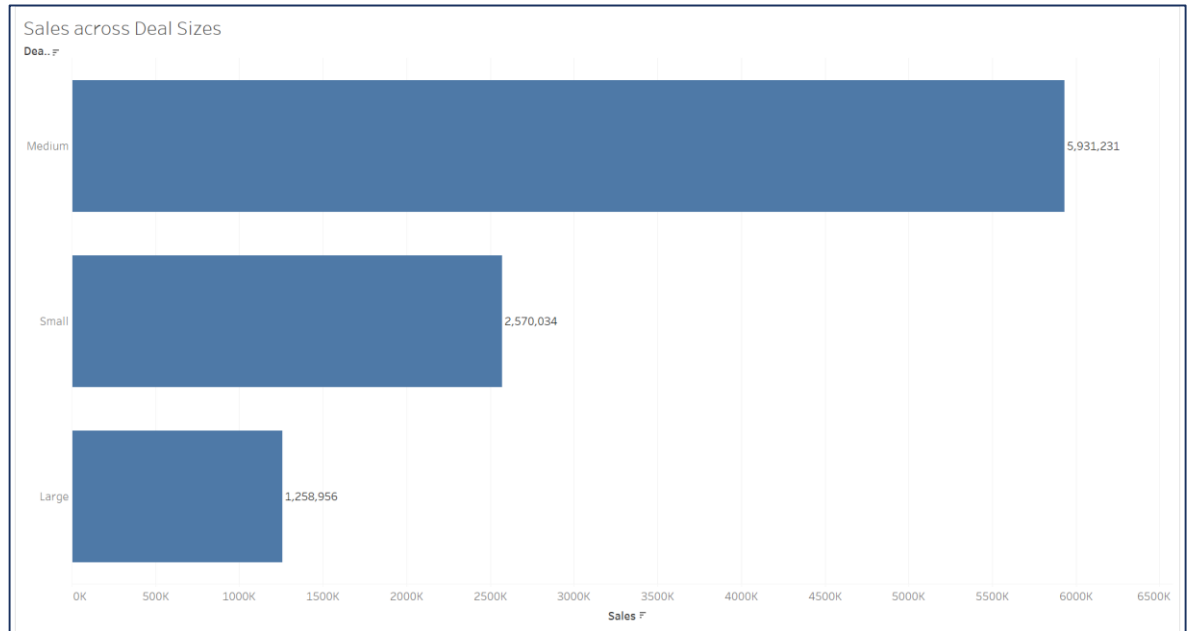
1. **"Medium" deal size** have highest sales in the category which is around 50% of the total sales.

2. "Large" deal size expected to have higher sales in comparison with other categories, but it is the least.

SALES ACROSS CATEGORIES

- **Monthly Sales Trend:-**

Sales across Countries

Country	Sales
USA	3,355,576
Spain	1,215,687
France	1,110,917
Australia	630,623
UK	478,880
Italy	374,674
Finland	329,582
Norway	307,464
Singapore	288,488
Denmark	245,637
Canada	224,079
Germany	220,472
Sweden	210,014
Austria	202,063
Japan	188,168
Switzerland	117,714
Belgium	108,413
Philippines	94,016
Ireland	57,756

- **Weekly Sales Trend:-**

Sales across Cities

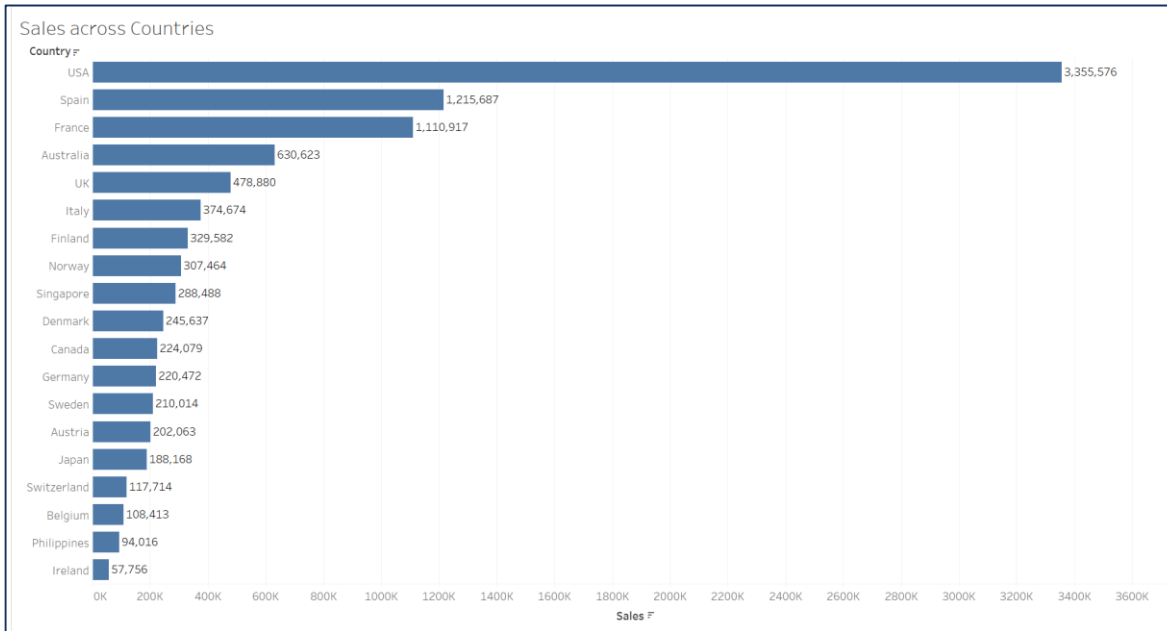| City	Sales
Madrid	1,082,551
San Rafael	654,858
NYC	560,788
Singapore	288,488
Paris	
New Bedford	207,875
Nantes	200,995
Melbourne	
Brickhaven	
San Jose	160,010
Manchester	
Boston	154,070
North Sydney	
Chatswood	151,571
Philadelphia	
Salzburg	149,799
Kobenhavn	
Lyon	142,874
Reggio Emilia	
Cambridge	139,244
Bergamo	
Reims	135,043
Boras	
Nashua	131,685
London	
Allentown	122,138
Burlingame	
Minato-ku	120,563
Liverpool	
Gensve	117,714
Stavern	
Espoo	113,961
Bergen	
Helsinki	111,250
Pasadena	
Oulu	104,370
Bridgewater	
Aaarhus	100,596
Koln	
Torino	94,117
Makati City	
San Diego	87,489
White Plains	

1. Maximum contribution of sales belong to **USA**, followed by Spain and France.

2. These 3 countries constitutes maximum revenue.
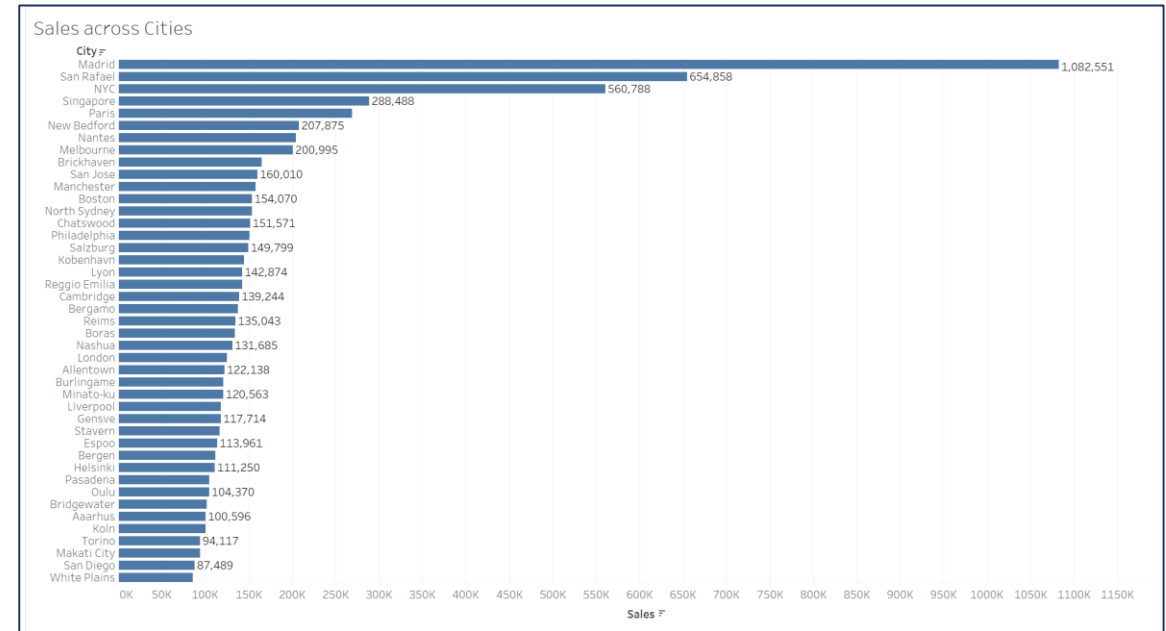
1. Similarly, maximum contribution of sales belongs to **Madrid**, followed by San Rafeal and New York.

2. These are the cities of maximum revenue generation.

SALES ACROSS CATEGORIES

- **Sales by Customers**

1. **Euro Shopping Channel** is on top of Customer Sales list, followed by Mini Gifts Distributors Ltd and Australian Collectors, Co.
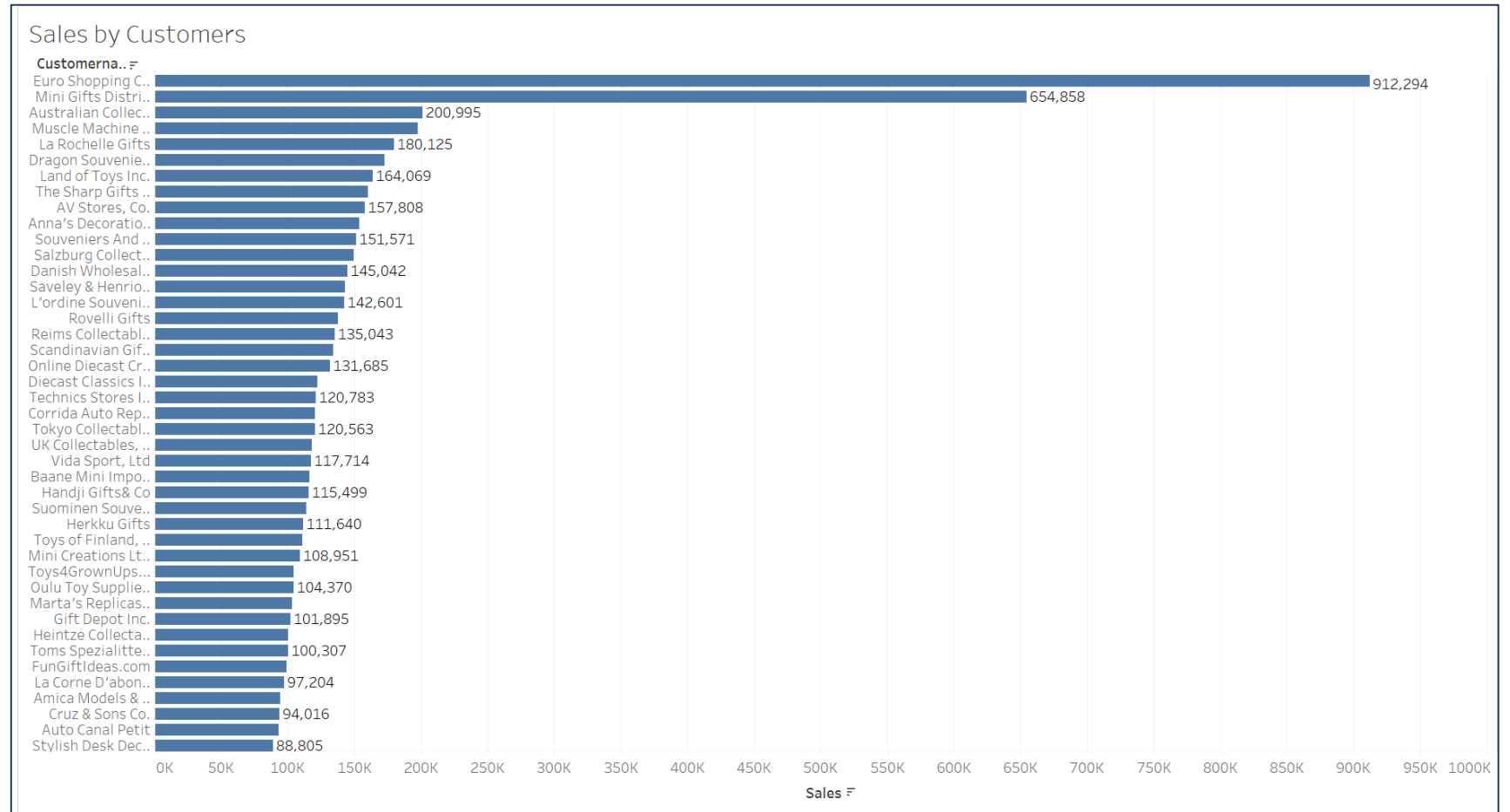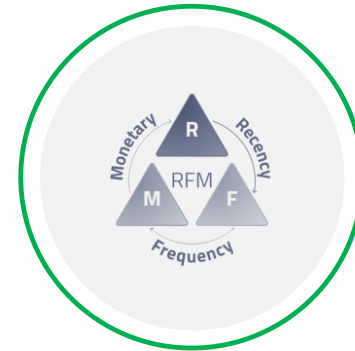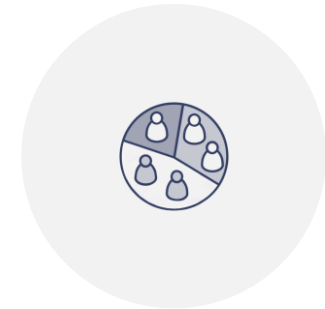

Sales by Customers

CONTENTS

Agenda & Executive Summary of the Data

Exploratory Analysis and Inferences

Customer Segmentation using RFM Analysis

Inferences from RFM Analysis and Identified Segments

CUSTOMER SEGMENTATION USING RFM ANALYSIS

❑ **What is RFM?**

- RFM stands for recency, frequency and monetary.

- RFM segmentation allows marketers to target specific clusters of customers with communications that are much more relevant for their particular behaviour and thus generate much higher rates of response, plus increased loyalty and customer lifetime value. Like other segmentation methods, RFM segmentation is a powerful way to identify groups of customers for special treatment.

- RFM is used to identify,

  1. Best Customers
  2. Big Spenders
  3. Loyal Customers
  4. Faithful Customers
  5. At-risk Customers
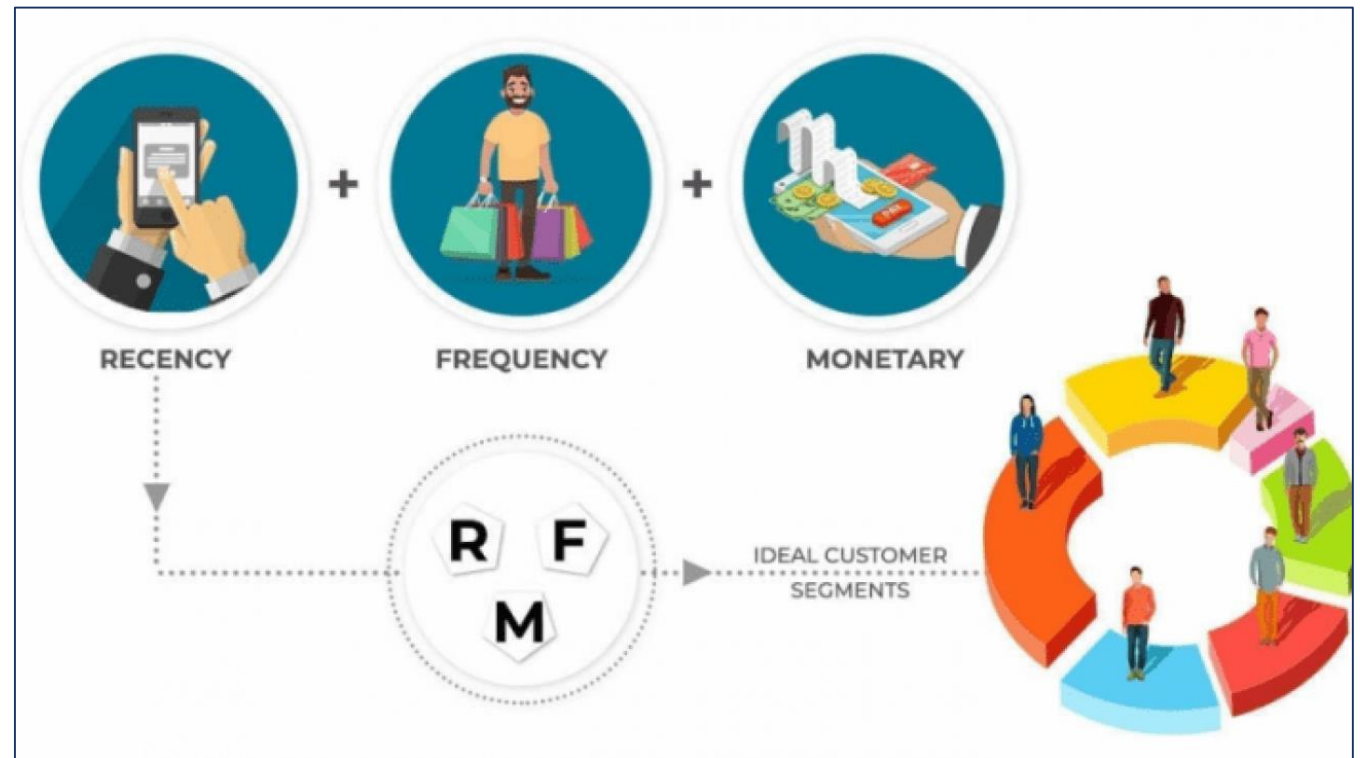
CUSTOMER SEGMENTATION USING RFM ANALYSIS

❑ **What all parameters used?**

• Underlying the RFM segmentation technique is the idea that marketers can gain an extensive understanding of their customers by analyzing three quantifiable factors. These are:

• Recency: How much time has elapsed since a customer's last activity or transaction with the brand? In most cases, the more recently a customer has interacted or transacted with a brand, the more likely that customer will be responsive to communications from the brand.

• Frequency: How often has a customer transacted or interacted with the brand during a particular period of time? Clearly, customers with frequent activities are more engaged, and probably more loyal, than customers who rarely do so.

• Monetary: How much a customer has spent with the brand during a particular period of time. Big spenders should usually be treated differently than customers who spend little.

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 are the assumptions made?**

• Customers with more purchases are more likely to buy products than customers with fewer purchases.

CUSTOMER SEGMENTATION USING RFM ANALYSIS

- **Analysis, KNIME Workflow :-**

CUSTOMER SEGMENTATION USING RFM ANALYSIS

KNIME_output_ex
cel

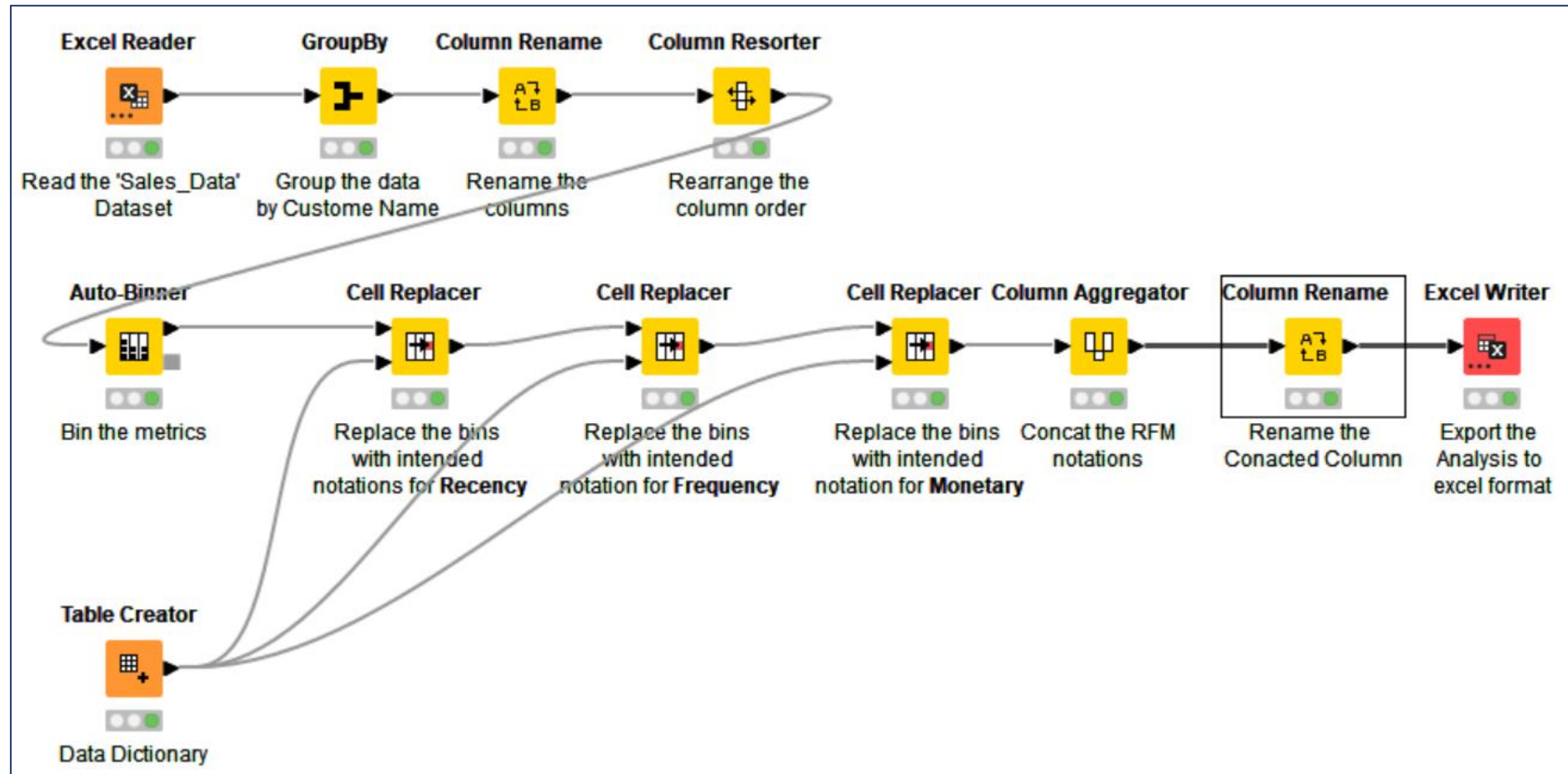- **Output Table :-**

Row ID	S CUSTO...	D QUANT...	D PRICEE...	D MSRP	S STATUS	S CITY	S COUNTRY	S DEALSIZE	I Recency	I Freque...	D Monetary	S Recenc...	S Freque...	S Moneta...	S RFM
Row0	AV Stores, Co.	34.863	91.085	92.843	Shipped	Manchester	UK	Medium	421	51	157,807.81	2	3	3	233
Row1	Alpha Cognac	34.35	101.16	97.15	Shipped	Toulouse	France	Medium	675	20	70,488.44	1	1	1	111
Row2	Amica Model...	32.423	110.853	107.654	Shipped	Torino	Italy	Large	328	26	94,117.26	2	2	2	222
Row3	Anna's Decor...	31.935	106.424	104.717	Shipped	North Sydney	Australia	Small	131	46	153,996.13	3	3	3	333
Row4	Atelier graph...	38.571	92.239	95.571	Shipped	Nantes	France	Medium	312	7	24,179.96	2	1	1	211
Row5	Australian C...	30.652	90.042	88.13	Shipped	Glen Waverly	Australia	Medium	1018	23	64,591.46	1	2	1	121
Row6	Australian C...	35.018	104.59	103.527	Shipped	Melbourne	Australia	Medium	229	55	200,995.41	3	3	3	333
Row7	Australian Gi...	36.333	110.554	111.533	Shipped	South Brisbane	Australia	Large	190	15	59,469.12	3	1	1	311
Row8	Auto Assoc. ...	35.389	99.488	100.389	Shipped	Versailles	France	Large	275	18	64,834.32	2	1	1	211
Row9	Auto Canal P...	37.074	94.255	94.852	Shipped	Paris	France	Medium	127	27	93,170.66	3	2	2	322
Row10	Auto-Moto Cl...	35.875	92.8	87.375	Shipped	Brickhaven	USA	Medium	1353	8	26,479.26	1	1	1	111
Row11	Baane Mini I...	33.813	108.574	107.469	Shipped	Stavern	Norway	Medium	245	32	116,599.19	3	2	2	322
Row12	Bavarian Coll...	28.643	84.289	82.714	Shipped	Munich	Germany	Medium	801	14	34,993.92	1	1	1	111
Row13	Blauer See A...	36.864	108.031	105.818	Shipped	Frankfurt	Germany	Medium	705	22	85,171.59	1	2	2	122
Row14	Boards & To...	34	89.807	92.333	Shipped	Glendale	USA	Medium	410	3	9,129.35	2	1	1	211
Row15	CAF Imports	36	104.963	106.923	Shipped	Madrid	Spain	Large	625	13	49,642.05	2	1	1	211
Row16	Cambridge C...	32.455	101.329	97.364	Shipped	Cambridge	USA	Medium	484	11	36,163.62	2	1	1	211
Row17	Canadian Gif...	31.955	105.341	106.409	Shipped	Vancouver	Canada	Large	364	22	75,238.92	2	2	2	222
Row18	Classic Gift I...	31.81	103.32	102.476	Shipped	Philadelphia	USA	Medium	344	21	67,506.97	2	2	1	221
Row19	Classic Lege...	36	109.803	106.65	Shipped	NYC	USA	Medium	309	20	77,795.2	2	1	2	212
Row20	Clover Collec...	30.625	112.87	106.875	Shipped	Dublin	Ireland	Large	659	16	57,756.43	1	1	1	111
Row21	Collectable M...	38.16	91.535	93.12	Shipped	San Diego	USA	Medium	575	25	87,489.23	2	2	2	222
Row22	Collectables ...	33.125	97.237	99.458	Shipped	Brickhaven	USA	Medium	179	24	81,577.98	3	2	2	322
Row23	Corrida Auto...	36.344	105.175	102.625	Shipped	Madrid	Spain	Large	407	32	120,615.28	2	2	3	223
Row24	Cruz & Sons ...	36.962	96.08	97.962	Shipped	Makati City	Philippines	Medium	971	26	94,015.73	1	2	2	122
Row25	Daedalus De...	34.95	95.474	94.5	Shipped	Lille	France	Small	573	20	69,052.41	2	1	1	211
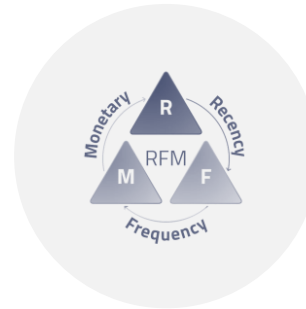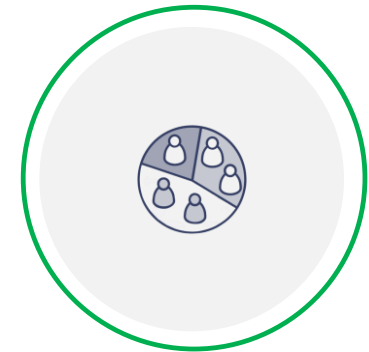
INFERENCES FROM RFM ANALYSIS AND IDENTIFIED SEGMENTS

- **Inferences :-**

1. Euro Shopping Channel have <u>high monetary</u> value and <u>high frequency</u> as well.
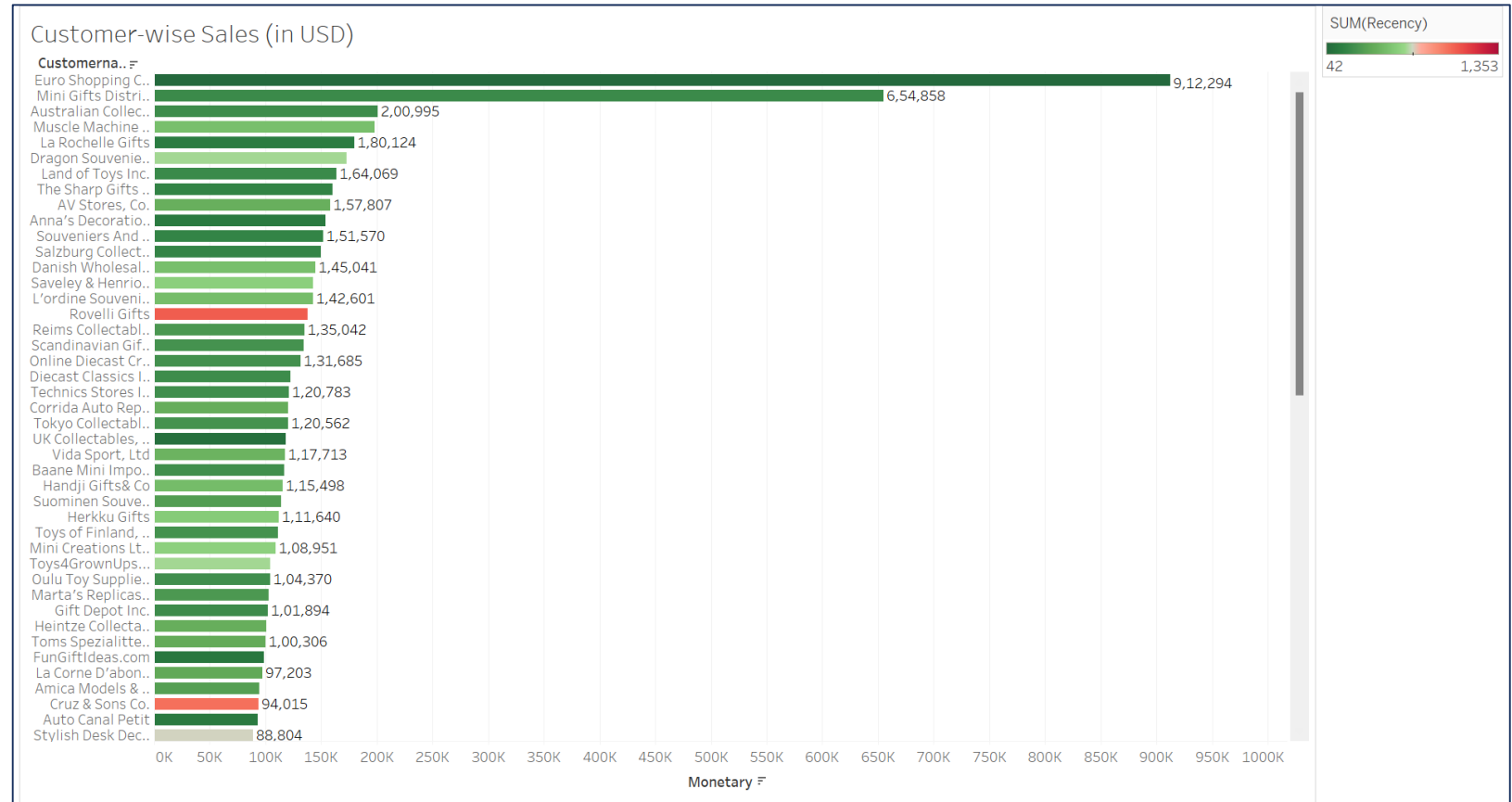
2. Followed by Mini Gifts Distributors.

Customer-wise Sales (in USD)
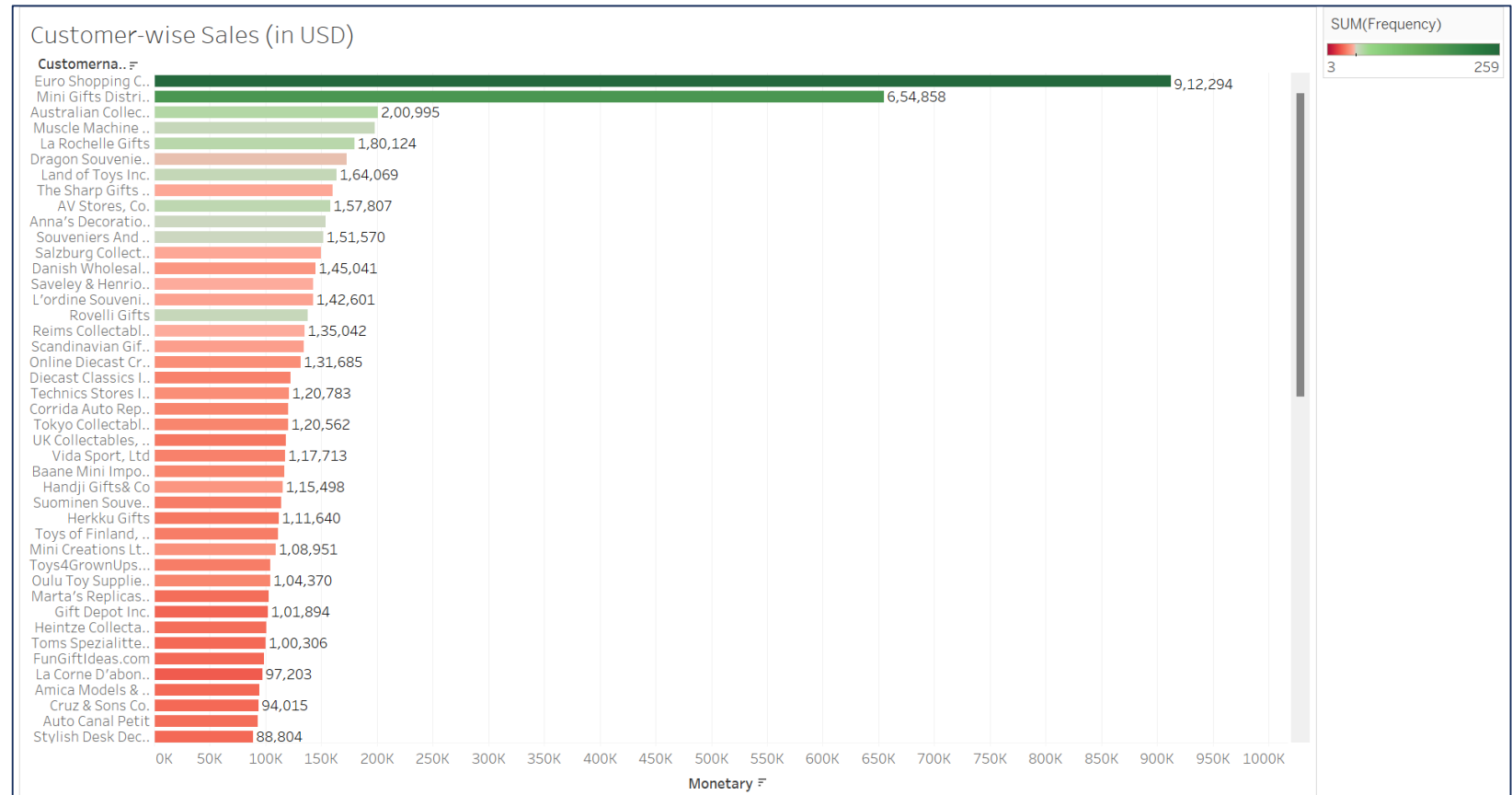
Knime_RFM_analysis

INFERENCES FROM RFM ANALYSIS AND IDENTIFIED SEGMENTS

- **Inferences :-**

1. Similarly, Euro Shopping Channel have <u>high monetary</u> value with <u>low recency value</u> which is good indicator.

2. Followed by Mini Gifts Distributors.

INFERENCES FROM RFM ANALYSIS AND IDENTIFIED SEGMENTS

- **Customer Segmentation:**

Best Customers : **11 nos.**

1. Euro Shopping Channel
2. Anna's Decorations, Ltd
3. La Rochelle Gifts
4. The Sharp Gifts Warehouse
5. Souveniers and Things Co.

Customers on verge of churning : **7 nos.**

1. Scandinavian Gift Ideas
2. Reims Collectables
3. AV Stores, Co.
4. L'ordine Souveniers
5. Danish Wholesale Imports

Lost Customers : **8+1+1+10 = 20 nos.**

1. Signal Gift Stores
2. Cruz & Sons Co.
3. Australian Collectables, Ltd
4. Clover Collections, Co.
5. Enaco Distributors

Loyal Customer : **11+3+2 = 16 nos.**

1. Danish Wholesale Imports
2. Muscle Machine Inc
3. Handji Gifts & Co
4. Mini Creations
5. Salzburg Collectables

RFM_Analysis

Recency	Frequency	Monetary (Binned)		
		3	2	1
3	3	11		
	2	1	9	
	1		1	1
2	3	7	2	
	2	1	19	1
	1		4	10
1	3	2		
	2		8	1
	1		1	10

THANK YOU