

# AUTO SALES DATA

## RFM Analysis and Segmentation

VIDYA V  
PGPDSBA.O.2023 B

# AGENDA



Analyze the sales data to derive insights and identify patterns

Perform RFM Analysis and Segment Customers

Develop Specific Strategies for each segment

# EXECUTIVE SUMMARY

**2747**

Orders



**19**  
Countries



**71**

Cities

**89**

Customers



**7**  
Product Lines

**96,428**  
Total quantity



**\$9.76 M**  
Total Sales



**92.5%**  
Shipped Orders

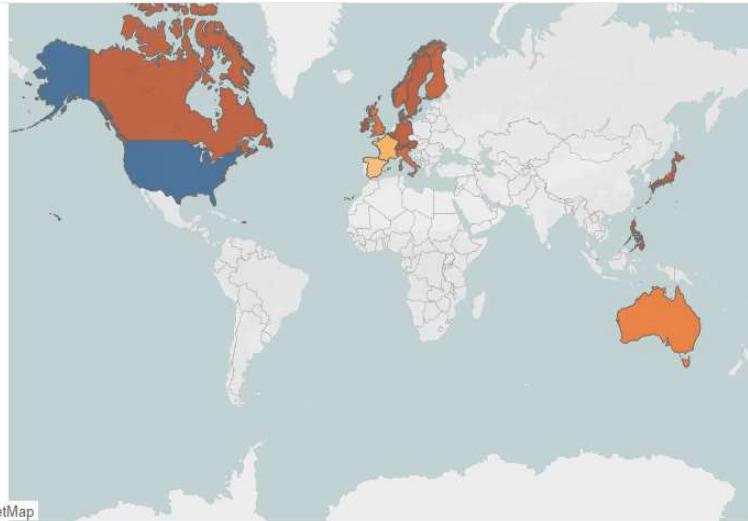




# UNIVARIATE ANALYSIS

Category	Value
1	10
2	12
3	15
4	18
5	20
6	22
7	25
8	28
9	30
10	32
11	35
12	38
13	40
14	42
15	45
16	48
17	50

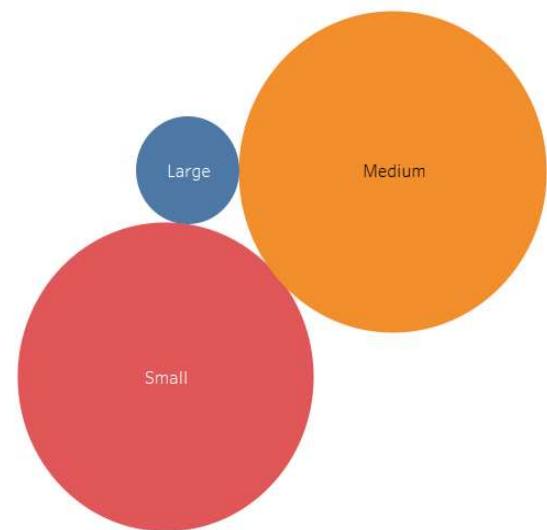
Country count



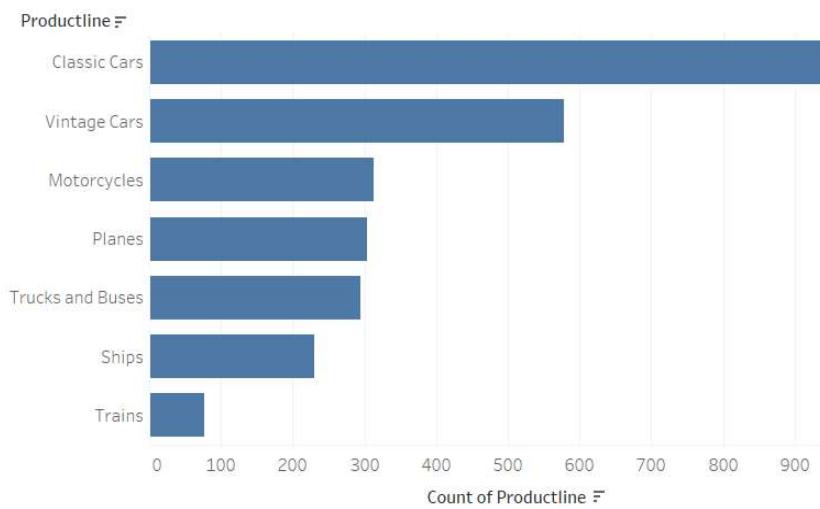
City count



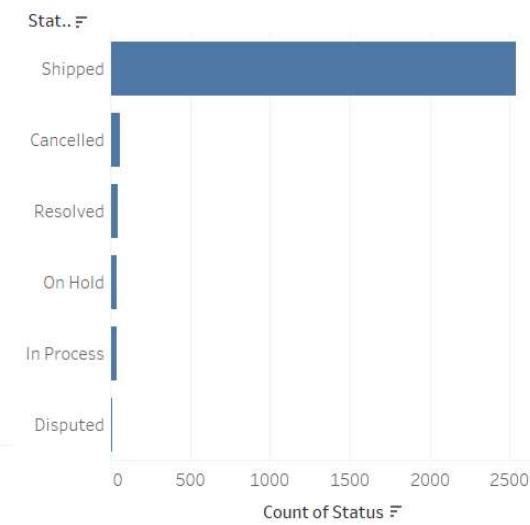
Deal Count



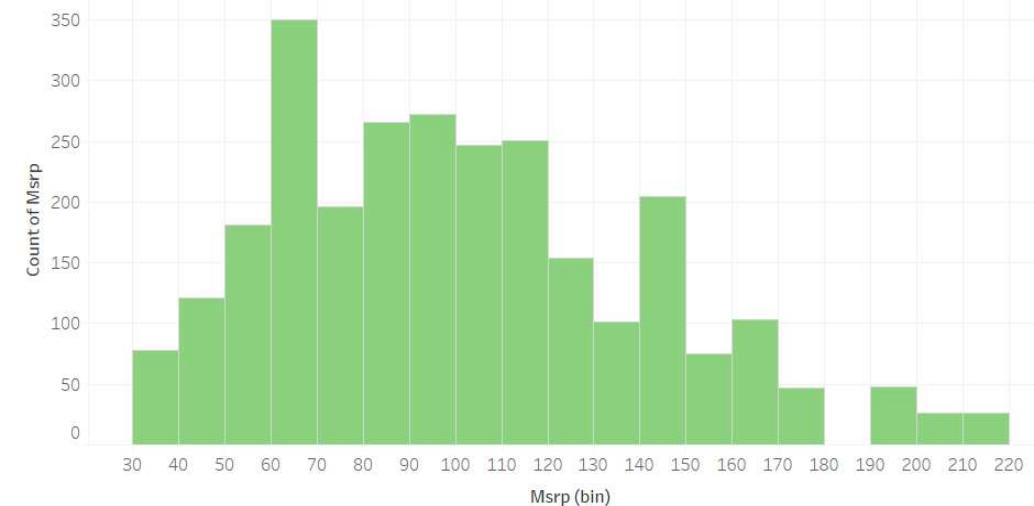
Product count



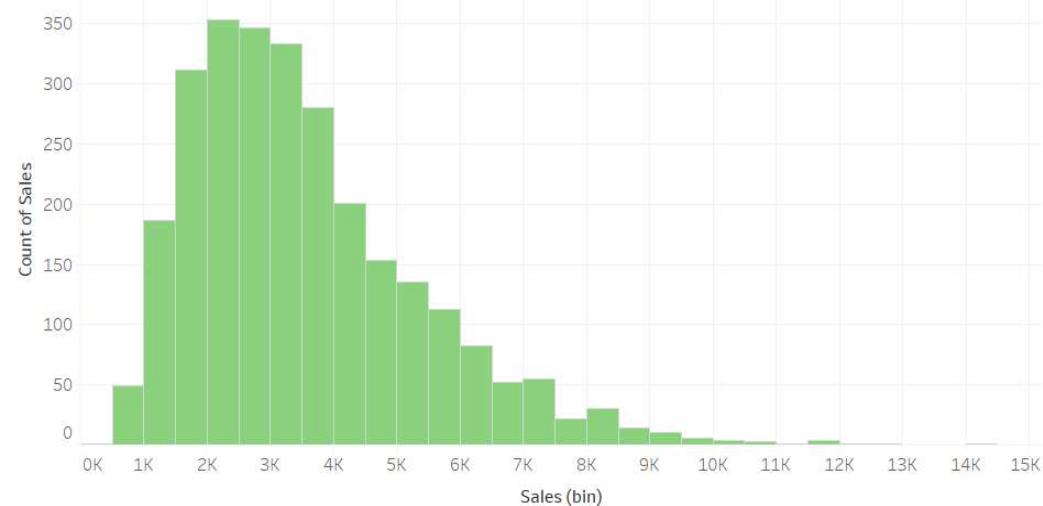
Status Count



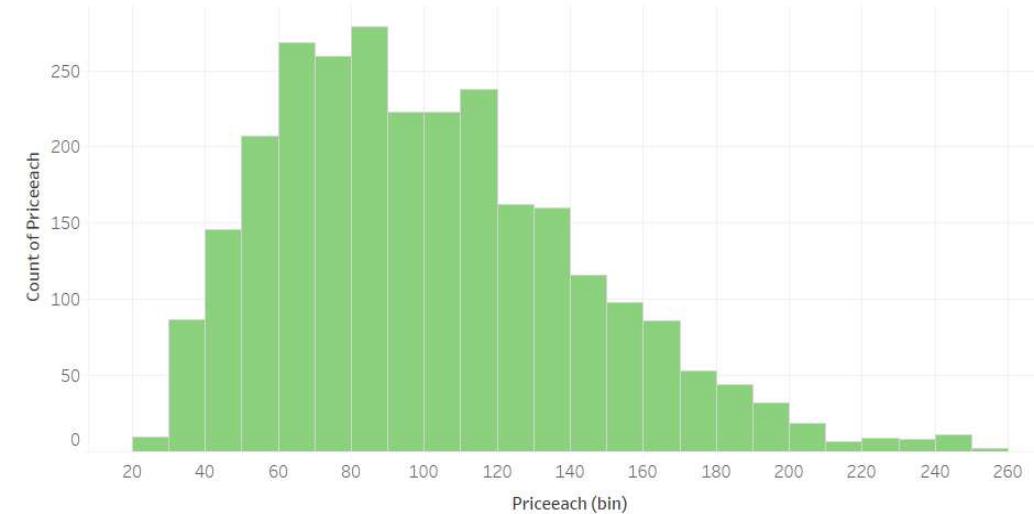
MSRP Distribution



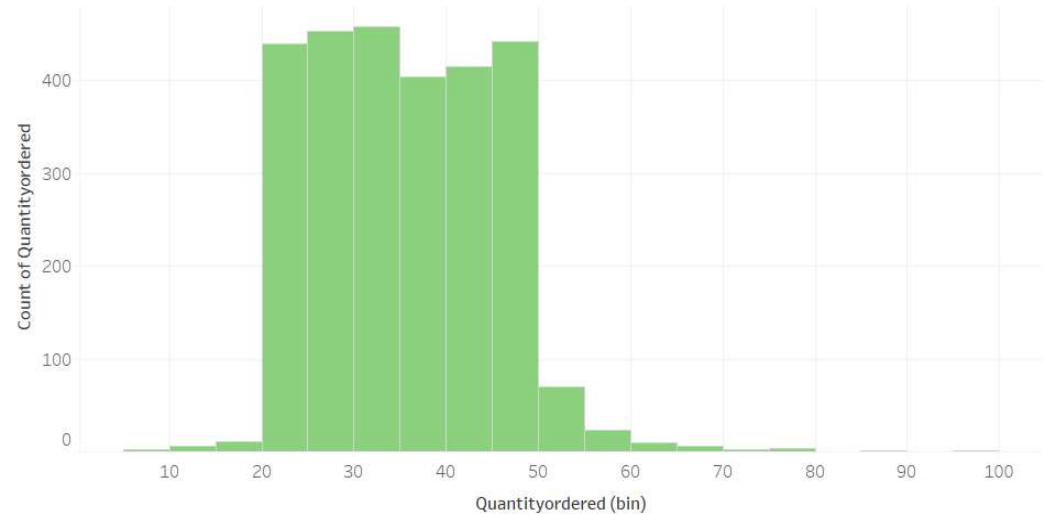
Sales Distribution



Unit Price Distribution



Quantity Distribution

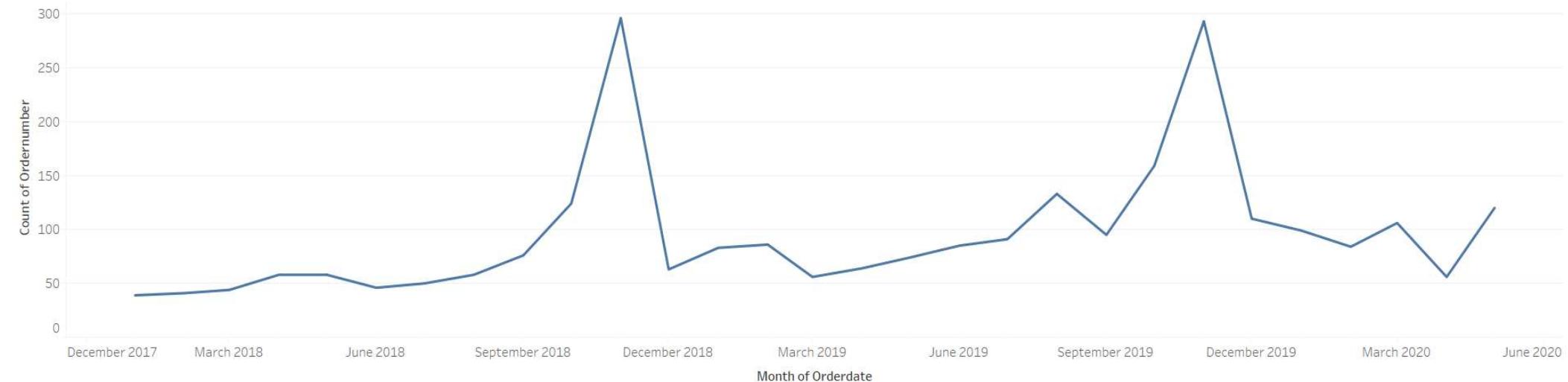




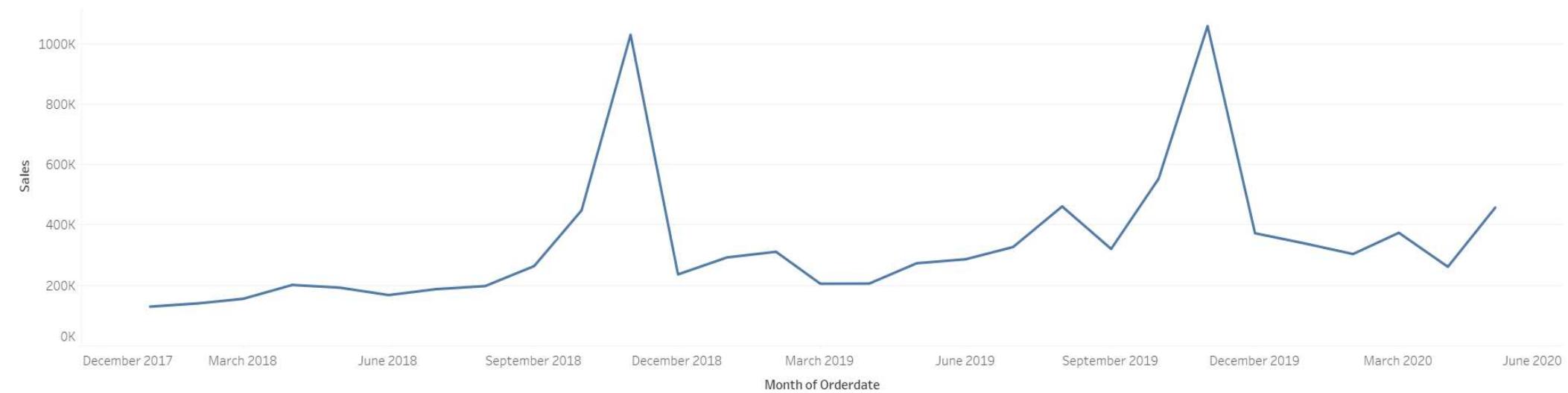
## BIVARIATE ANALYSIS

50  
40  
30  
20  
10

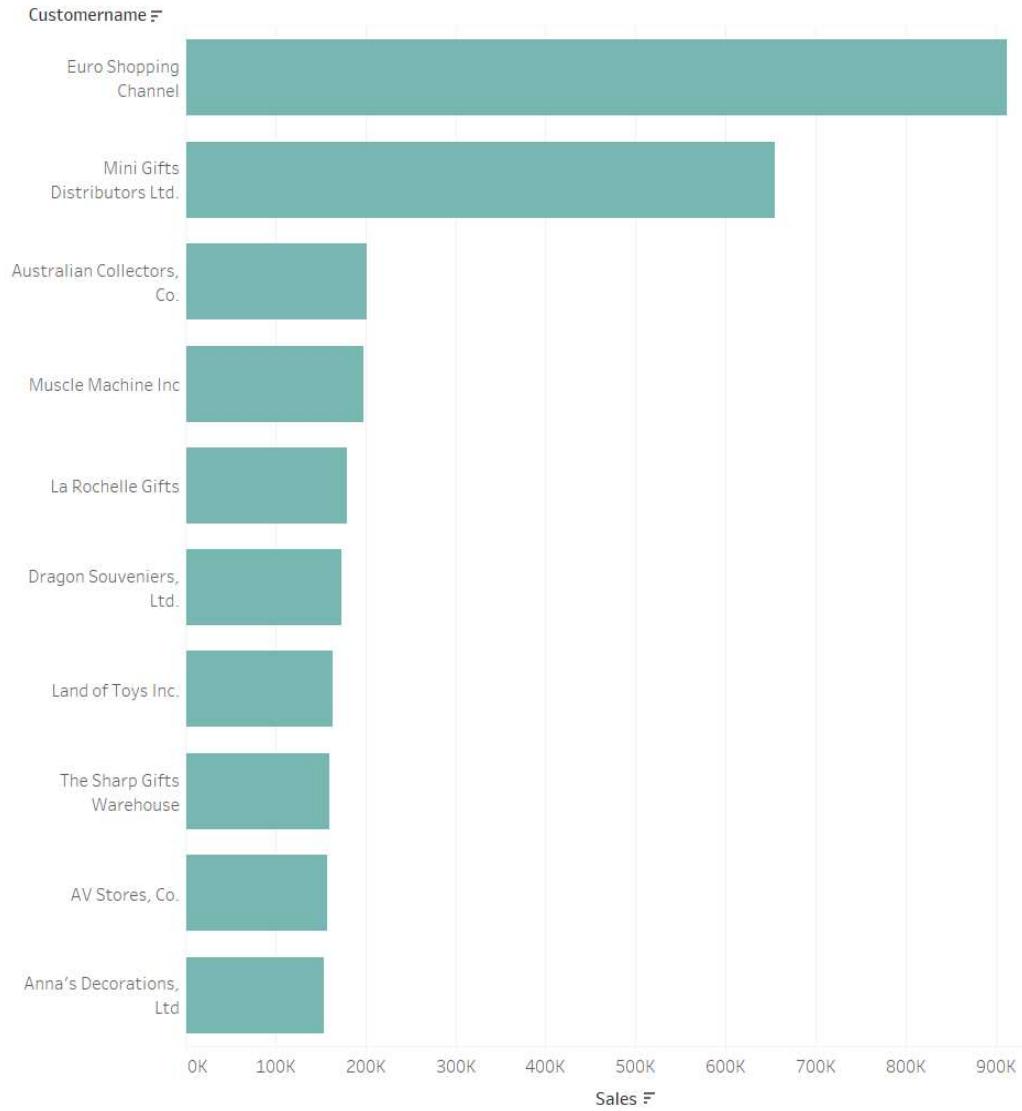
### Order count over time



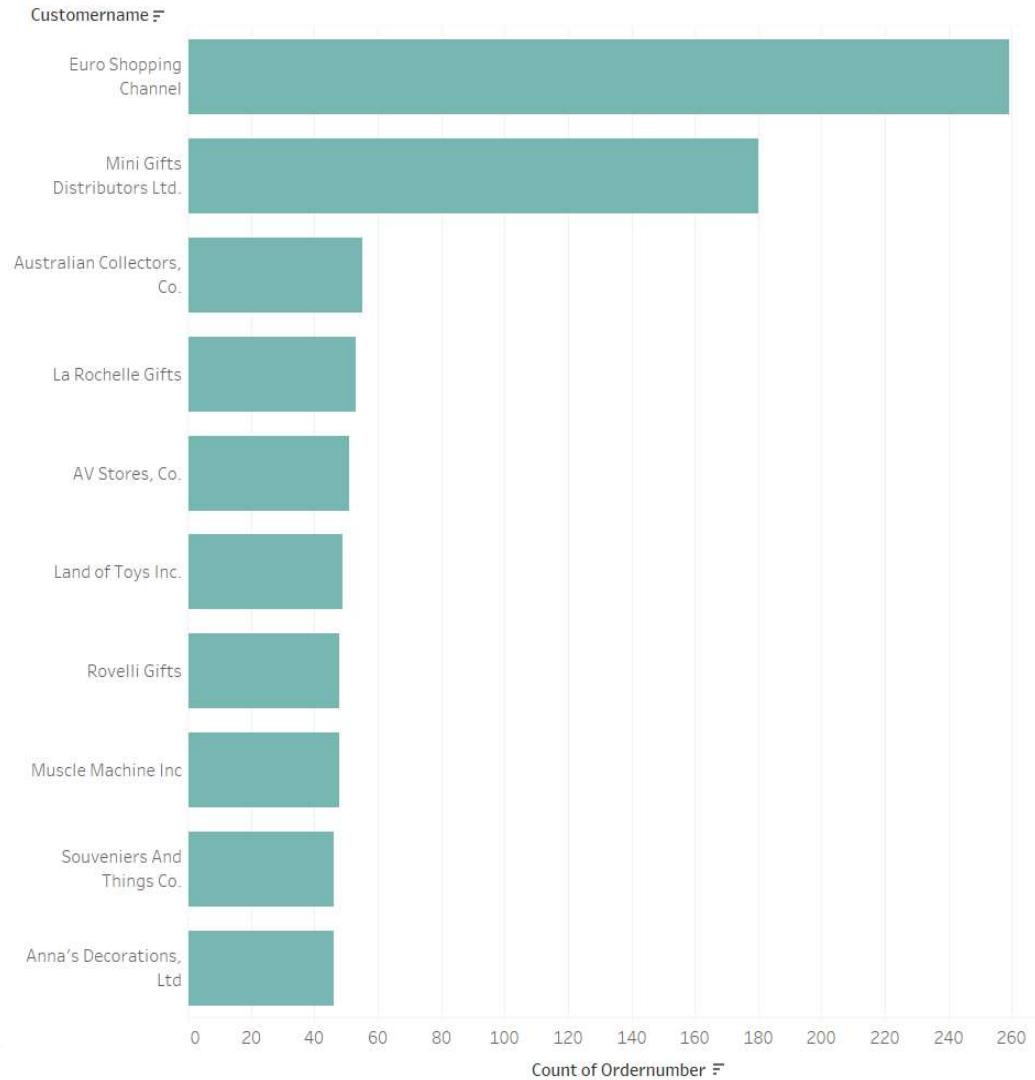
### Sales over time



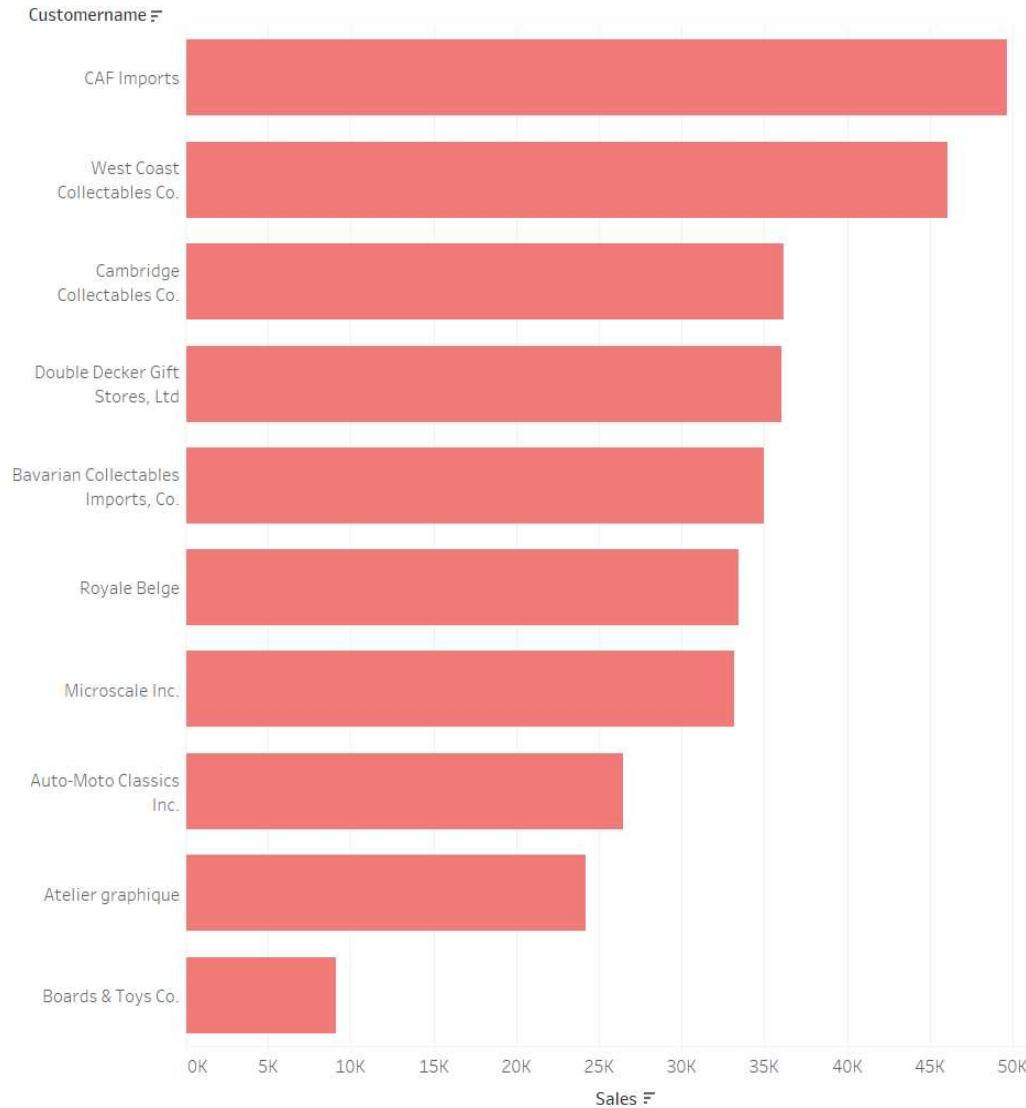
## Top 10 Customers by Sales



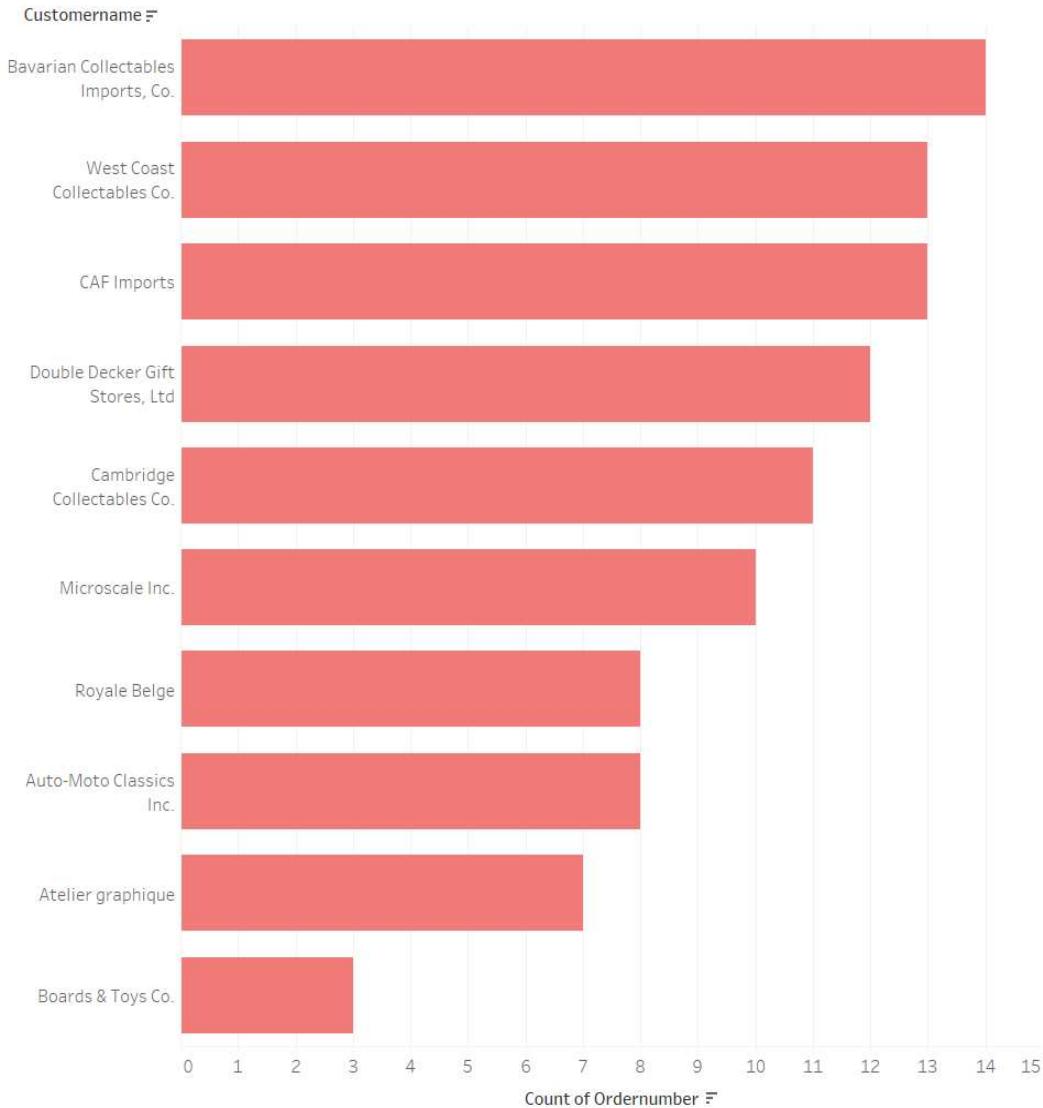
## Top 10 Customers by Orders



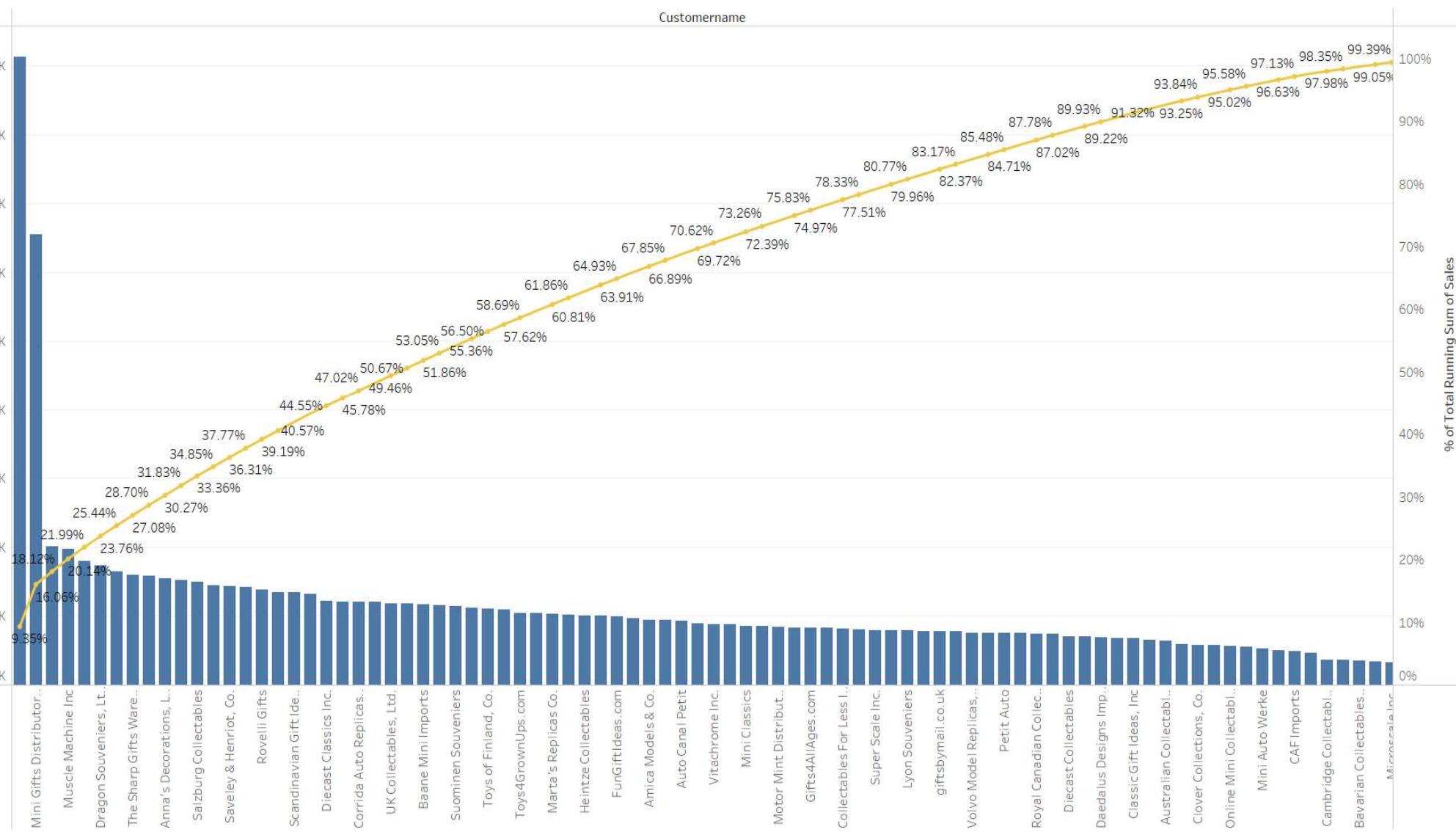
### Bottom 10 Customers by Sales



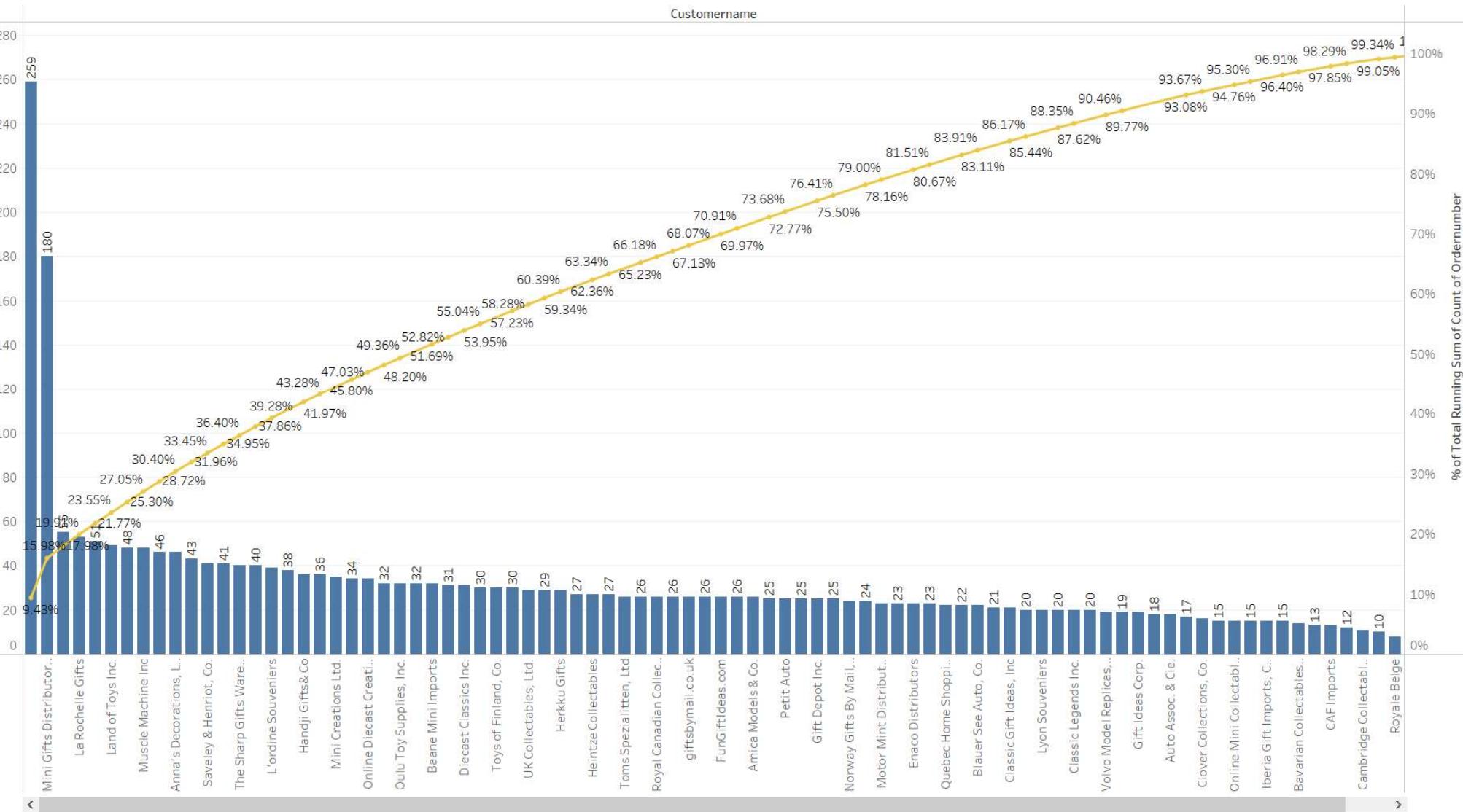
### Bottom 10 Customers by Orders



# pareto-sales



## pareto-orders

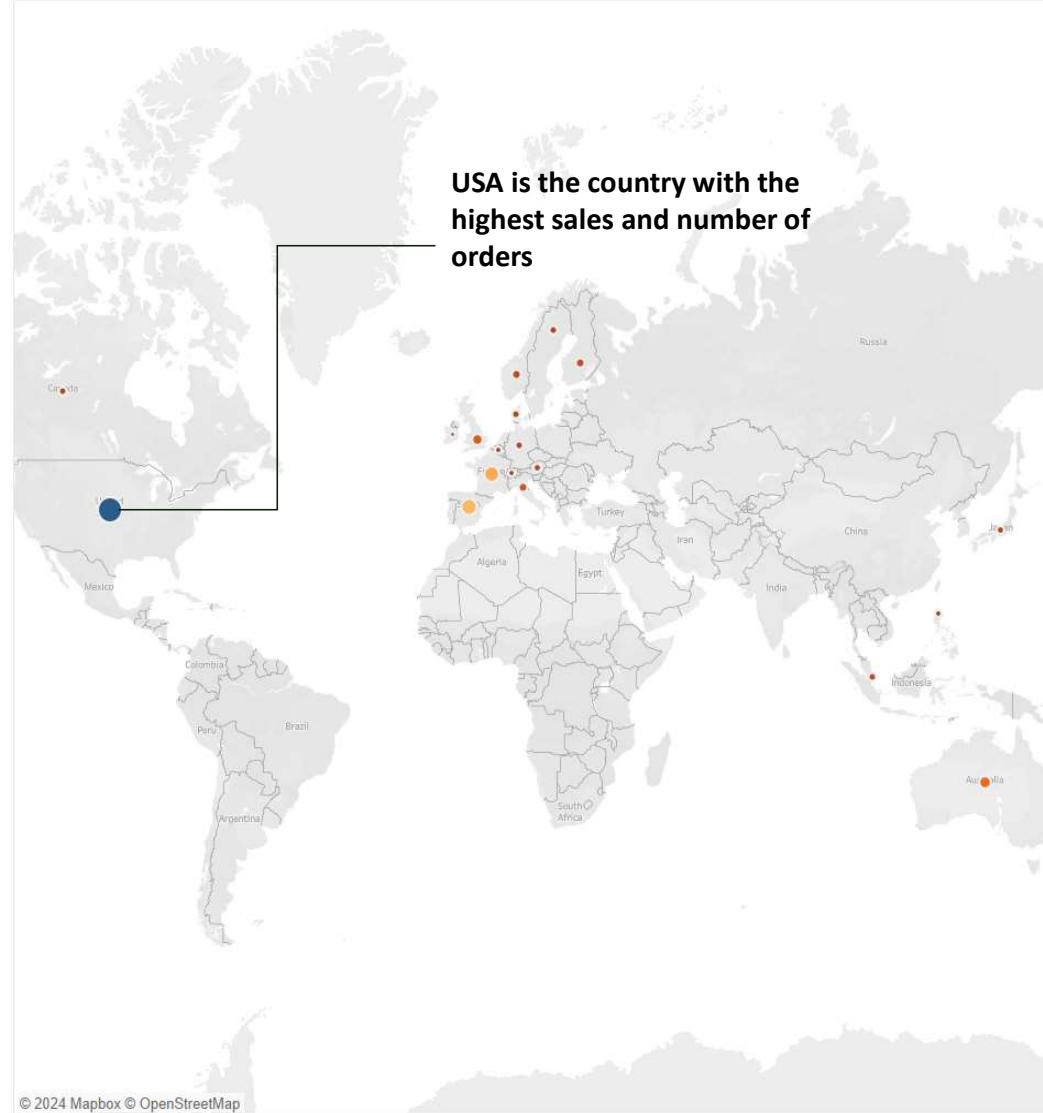




# MULTIVARIATE ANALYSIS

50  
40  
30  
20  
10

Country wise sales and orders

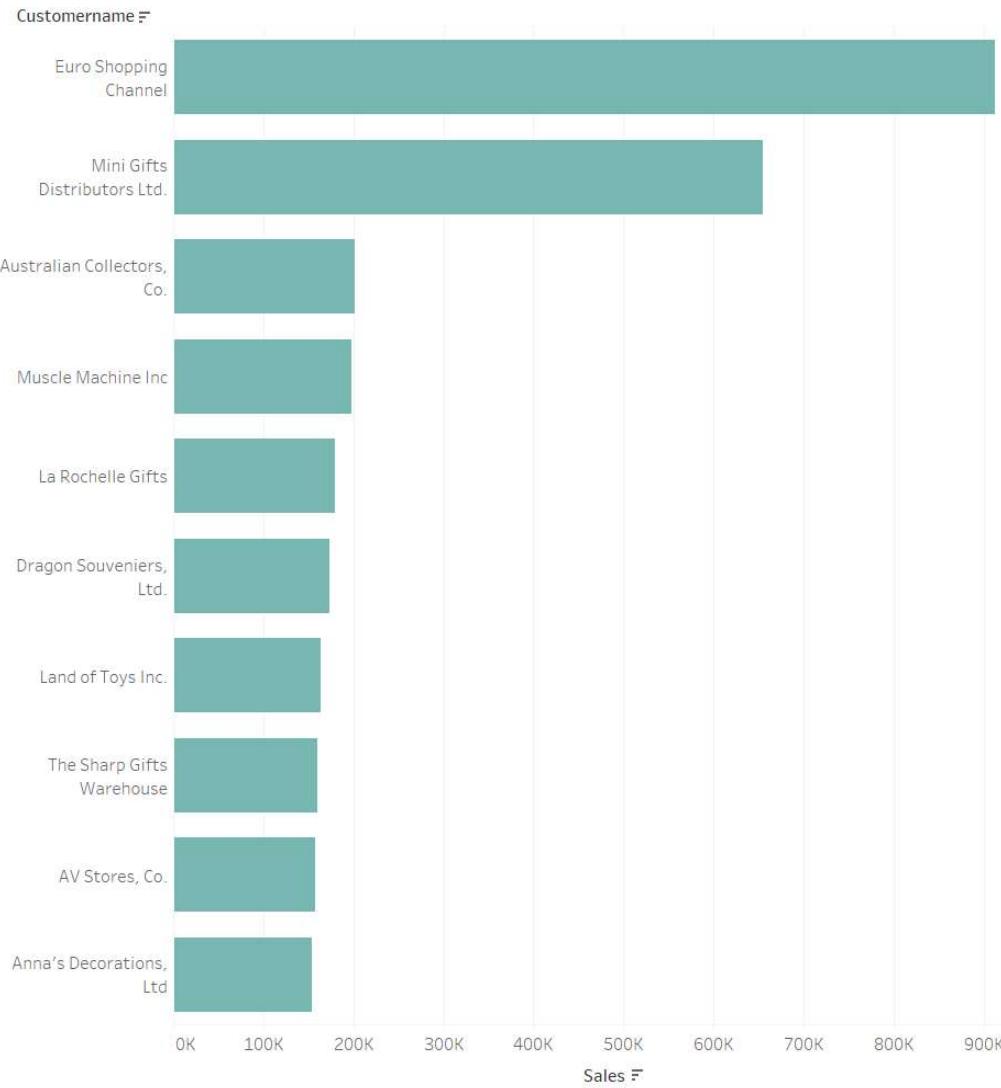


Citywise sales and orders

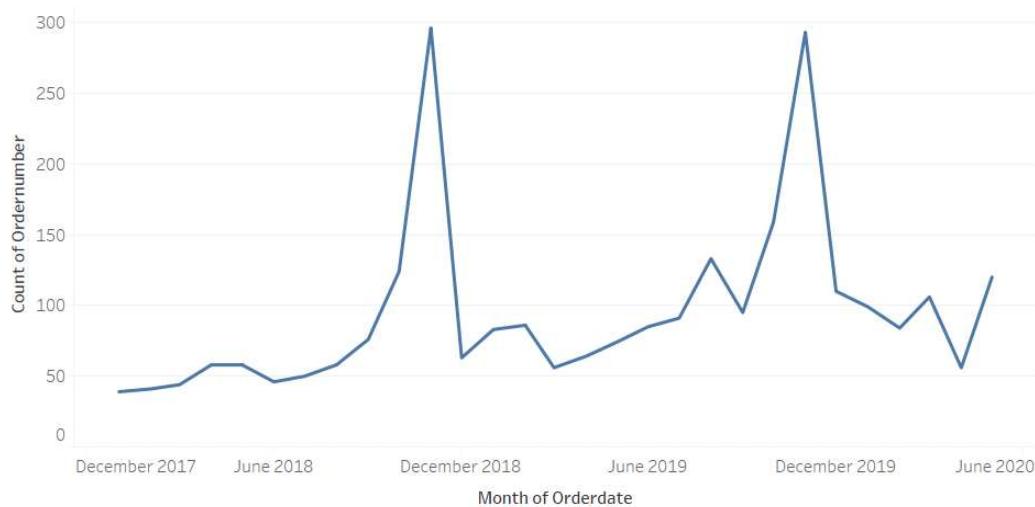


A1 Author, 1/12/2024

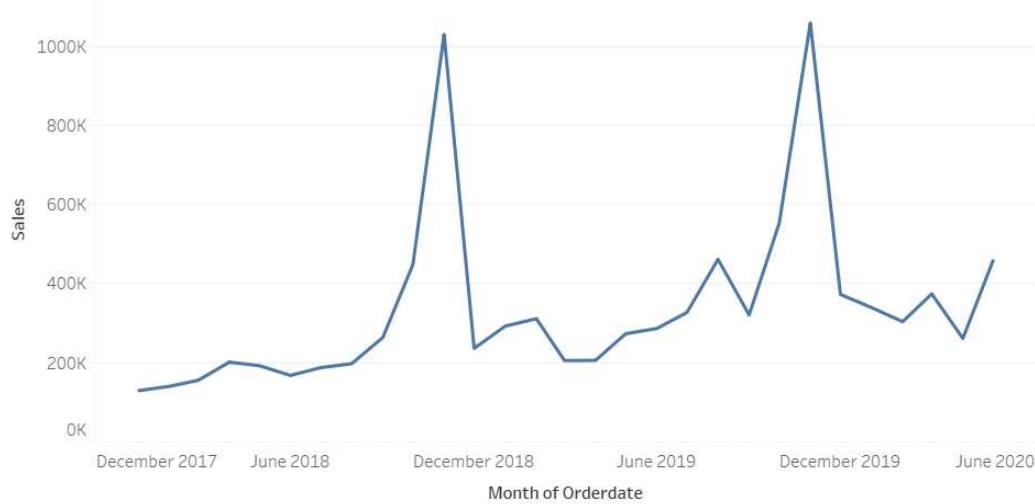
## Top 10 Customers by Sales



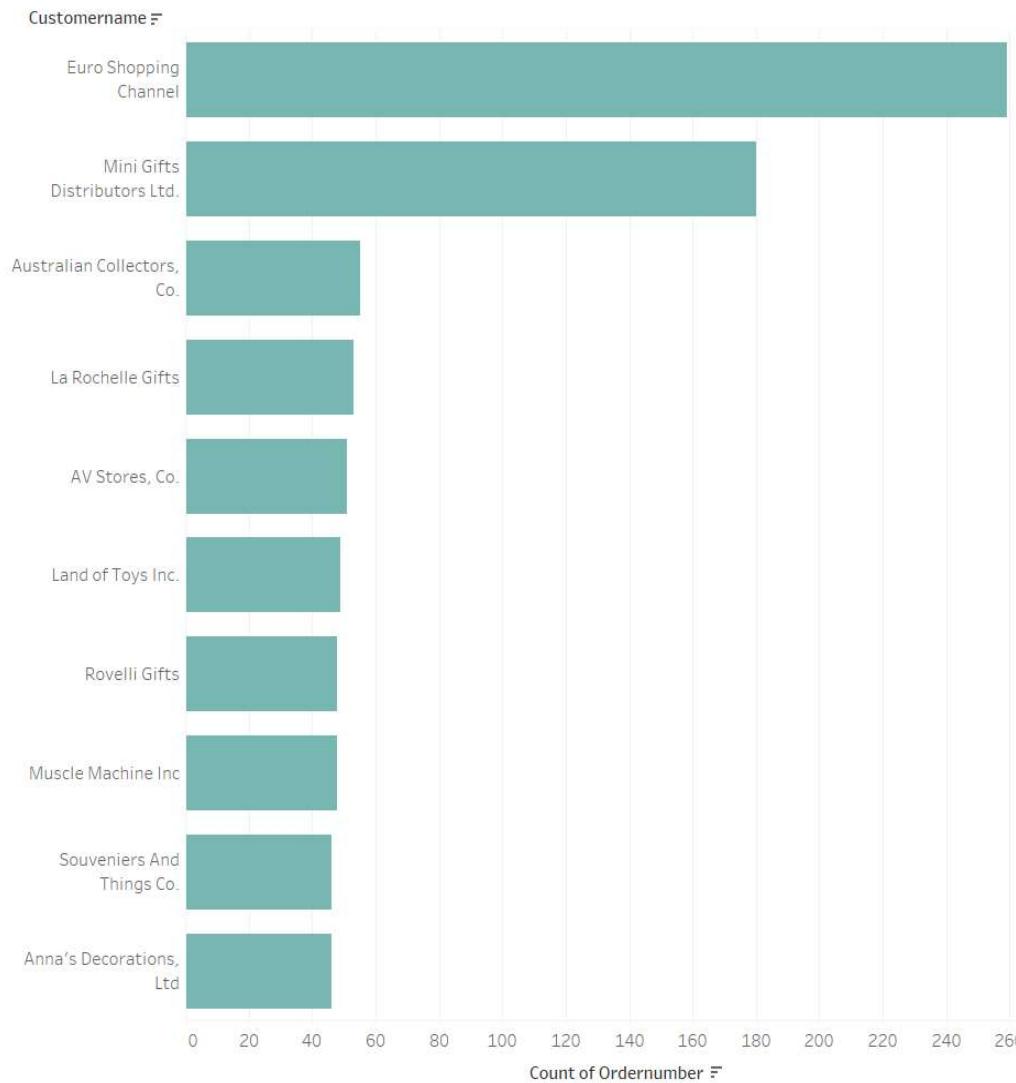
## Order count over time



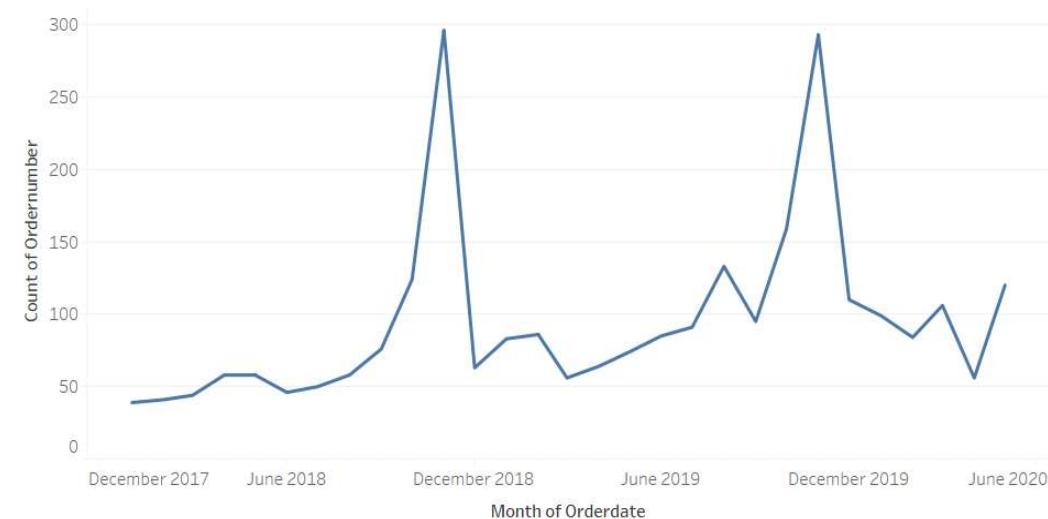
## Sales over time



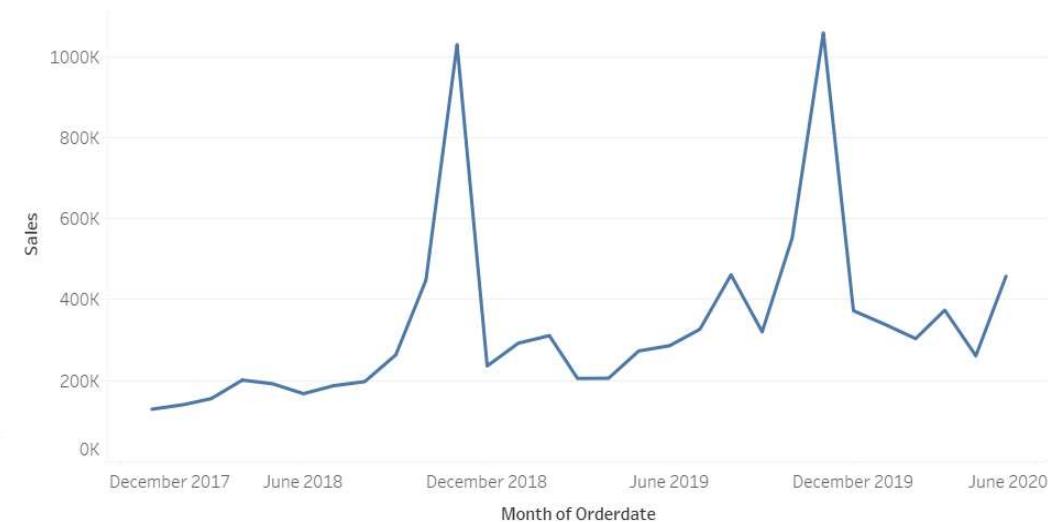
## Top 10 Customers by Orders



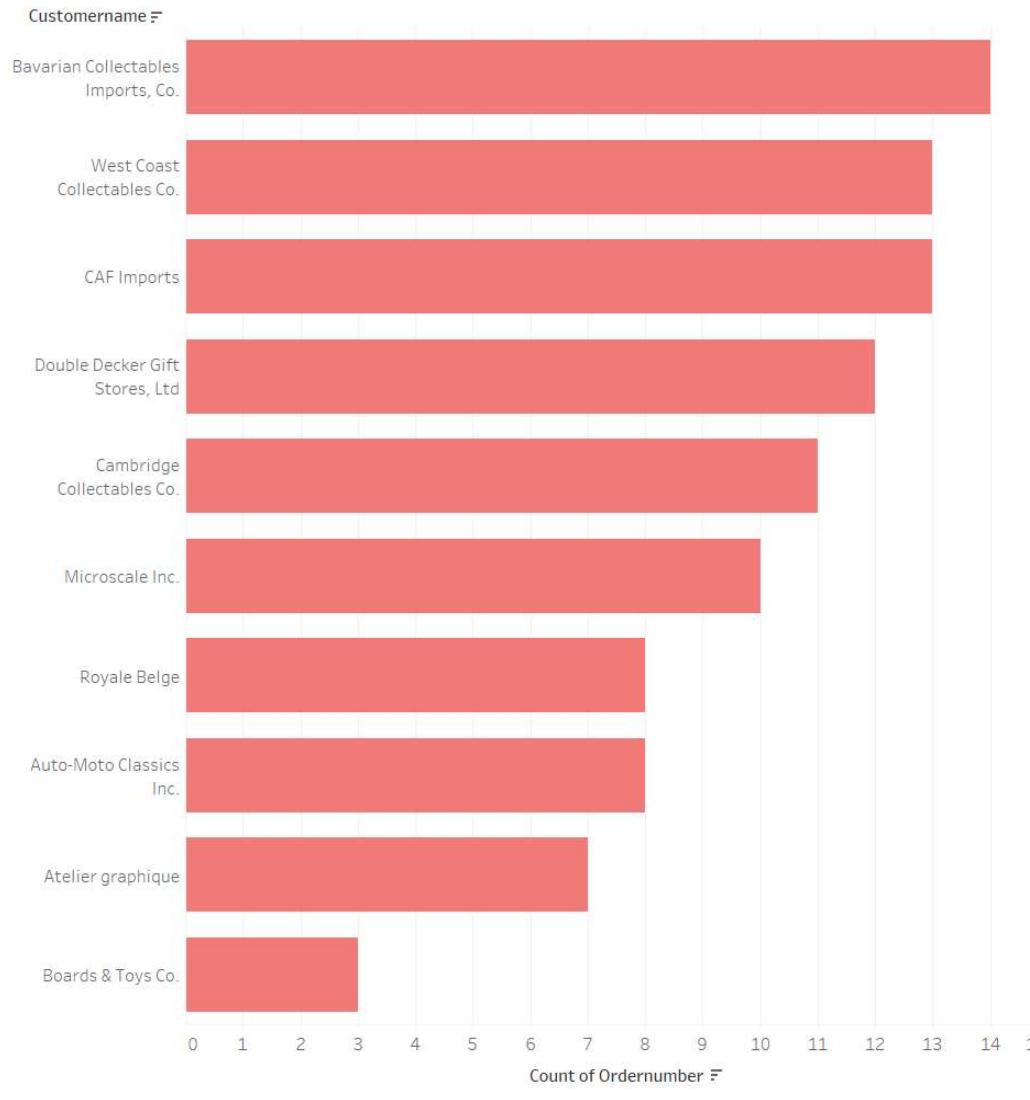
## Order count over time



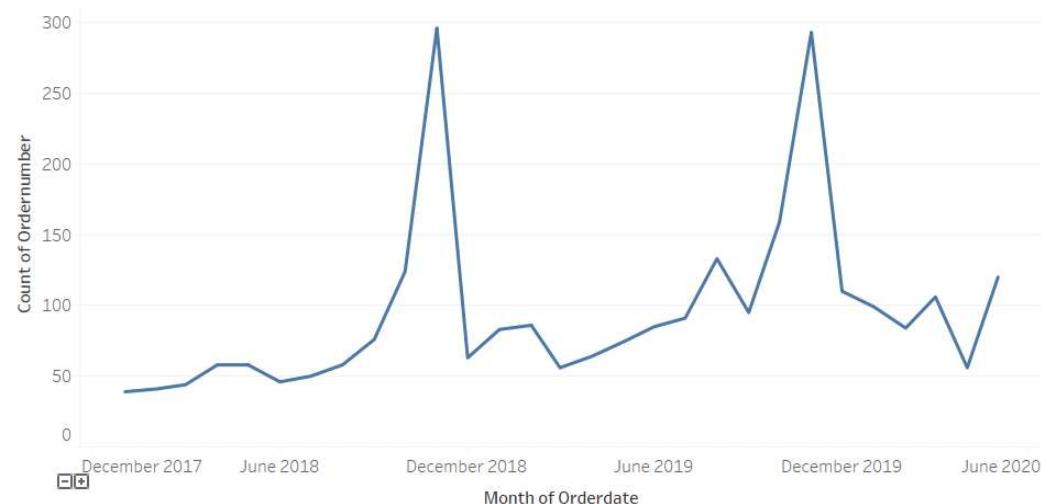
## Sales over time



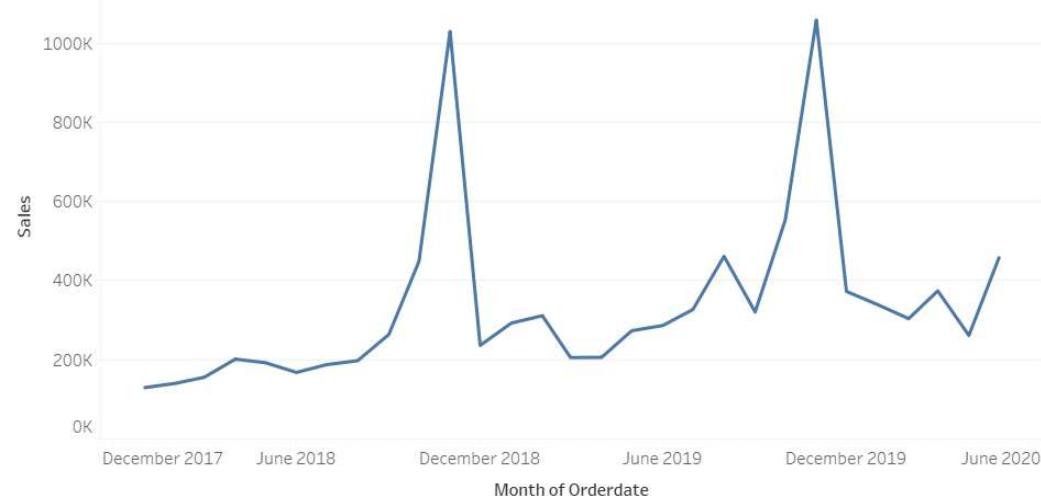
## Bottom 10 Customers by Orders



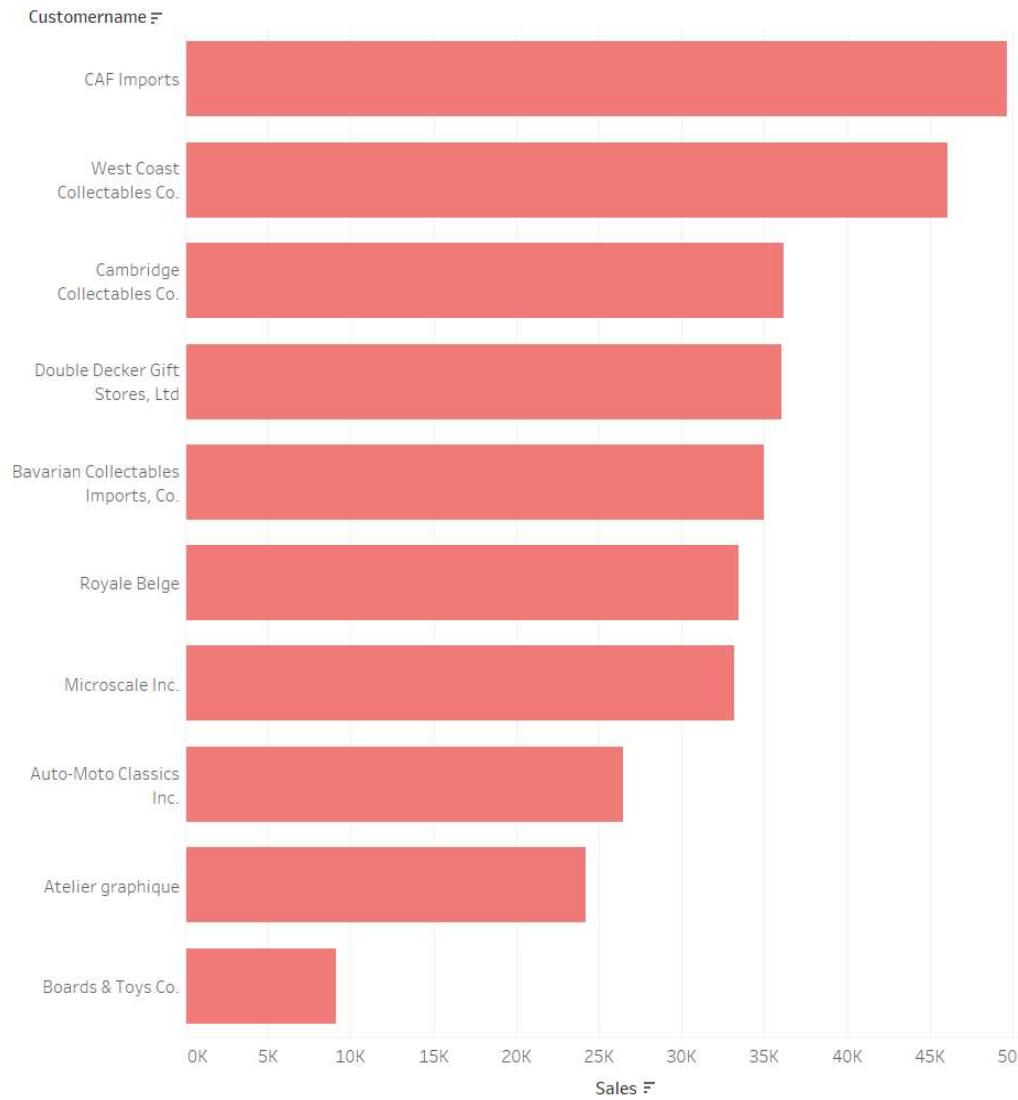
## Order count over time



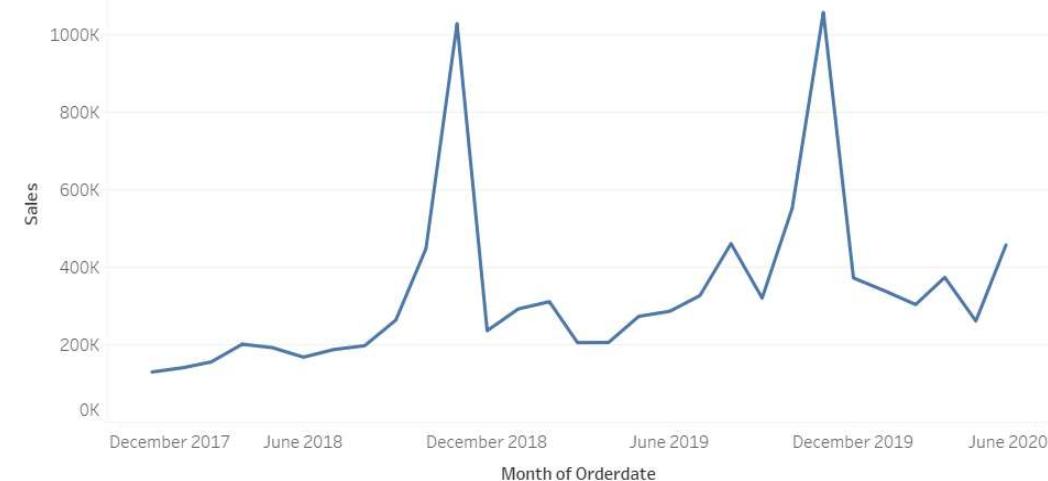
## Sales over time



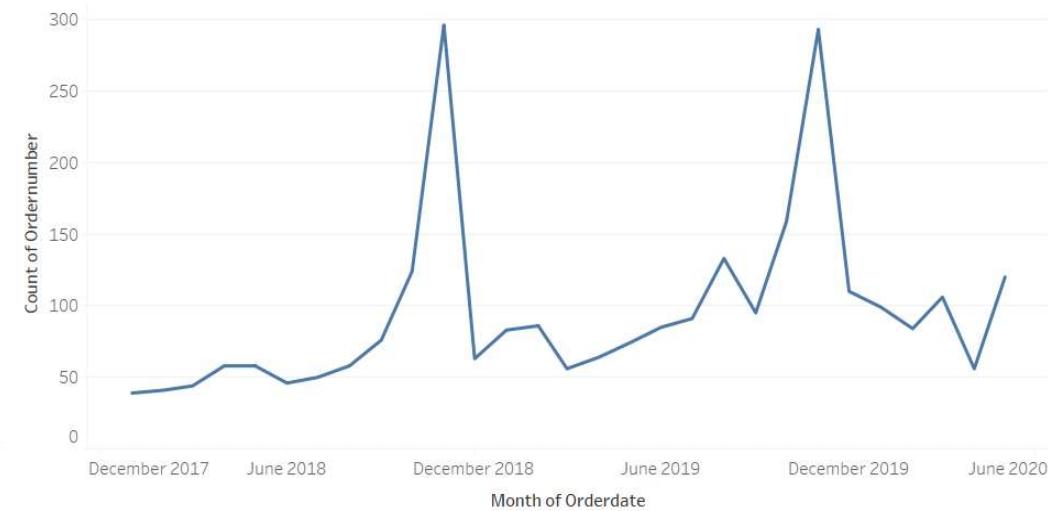
### Bottom 10 Customers by Sales



### Sales over time



### Order count over time



# EDA OBSERVATIONS

- USA is the country where most of the orders come from and is also the country with the highest sales
- Madrid is the city with most number of orders and highest sales
- There is no observable trend in the number of orders or the sales over the given period of 29 months
- Euro Shopping Channel, Mini Gifts Distributors Ltd are two best customers with the highest number of orders and highest total sales
- The top 10 customers contribute to approximately 30% of total orders and total sales

# RFM ANALYSIS

RFM

Frequency

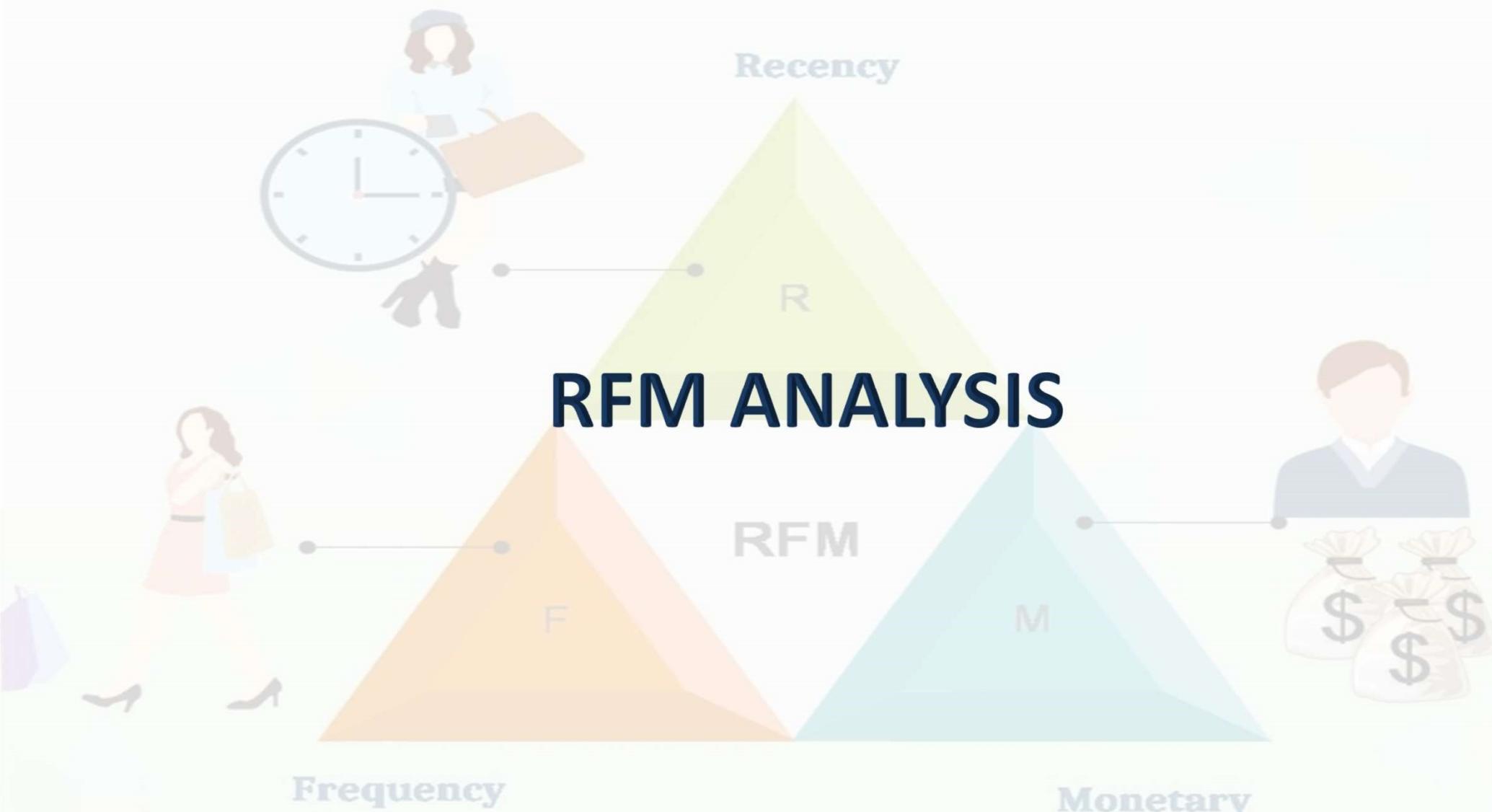
Monetary

Recency

R

M

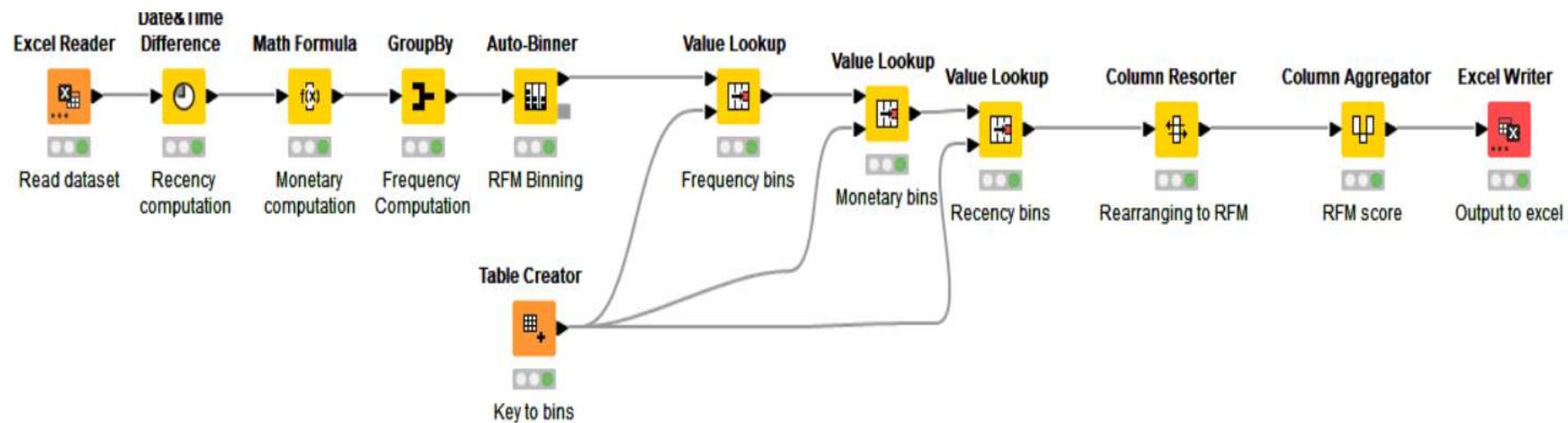
F



# RFM ANALYSIS

- Technique used to identify three metrics- Recency, Frequency and Monetary values of each customer using their transactions
- Objective: Customer Segmentation and development of specific marketing strategies to the target segments
- Recency: Measure of how recently the purchase was made. Helps determine the engagement and responsiveness of customers
- Frequency: Measure of how frequent the purchases are. Helps determine the loyalty of customers
- Monetary: Measure of the sales bought in by the customers. Helps determine the monetary value of customers

# KNIME WORKFLOW



# METHOD OF DETERMINING RFM

- Recency:
  - Determine the date difference of each transaction from a reference date- 01-07-2020
  - Find the minimum date difference after grouping by customer name
- Frequency:
  - Group by customer name and count the number of orders.
  - This is the total number of orders by the customer in 29 months
- Monetary:
  - Group by customer name and compute the sum of sales for all the orders
  - This is the total sales for the company from the customer

# SEGMENTATION AFTER RFM

- After computing the RFM metrics, a scoring system was developed using the auto binner node
- Each customer was assigned a bin for recency, frequency and monetary columns
- The key for the assignment of bins for frequency and monetary columns is as follows:
  - Bin 1- 0-25<sup>th</sup> percentile of column values
  - Bin 2- 25<sup>th</sup> -50<sup>th</sup> percentile of column values
  - Bin 3- 50<sup>th</sup>-75<sup>th</sup> percentile of column values
  - Bin 4- 75-100<sup>th</sup> percentile of column values
- The key for the assignment of bins for recency column is as follows:
  - Bin 4- 0-25<sup>th</sup> percentile of column values
  - Bin 3- 25<sup>th</sup> -50<sup>th</sup> percentile of column values
  - Bin 2- 50<sup>th</sup>-75<sup>th</sup> percentile of column values
  - Bin 1- 75-100<sup>th</sup> percentile of column values

# SEGMENTATION AFTER RFM

- That is for all these columns, the worst values would be given the lowest bin and the best values would be assigned the highest bin
- After this, an RFM score column was computed by concatenating the bins assigned
- This score had values 111,112,113,114,121,122,123,124,.....444
- Based on these scores, the customers were segmented into four categories:
  - Best Customers- RFM scores => [444]
  - Loyal Customers- RFM scores => [144,244,343,344,443,444]
  - Customers on the verge of churning- RFM scores =>[112,122,123,133,144]
  - Lost Customers- RFM scores =>[111]
- That is,
  - Best customers are those who have ordered recently, frequently and have contributed highly to the sales volume

# SEGMENTATION AFTER RFM

- That is,
  - Best customers are those who have ordered recently, frequently and have contributed highly to the sales volume
  - Loyal customers have high frequency of orders, even if it does not have a high monetary score
  - Customers who have the lowest recency score are on the verge of churning, even if they have high frequency and monetary scores
  - Lost customers have the highest days since last order, have a low frequency and monetary scores

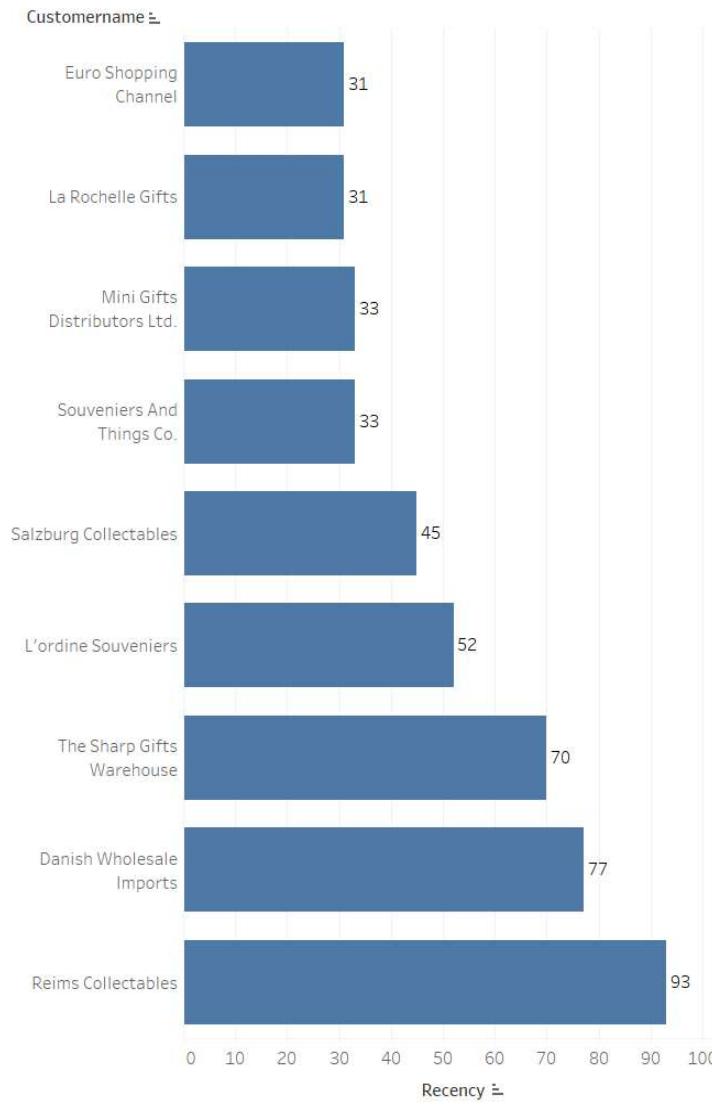
Rows: 29 | Columns: 14

Name	Type	# Missing values	# Unique values	Minimum	Maximum	25% Quantile	50% Quantile (Median)	75% Quantile	Mean	Mean Absolute Deviation	Standard Deviation	Sum	10 most common values
CUSTOMERNAME	String	0	89	?	?	?	?	?	?	?	?	?	AV Stores, Co. ...
ORDERNUMBER	Number (integer)	0	41	3	259	19.5	26	32	30.865	13.883	31.37	2,747	26 (9; 10.11%), ...
QUANTITYORDERED	Number (integer)	0	84	102	9,327	685.5	903	1,155	1,083.461	493.461	1,124.113	96,428	357 (2; 2.25%), ...
PRICEEACH	Number (double)	0	89	84.289	128.452	94.299	100.548	108.652	101.123	7.443	8.812	8,999.967	84.289 (1; 1.12...)
ORDERLINEINDEX	Number (integer)	0	41	3	259	19.5	26	32	30.865	13.883	31.37	2,747	26 (9; 10.11%), ...
SALES	Number (double)	0	89	9,129.35	912,294.11	69,770.425	87,489.23	121,839.975	109,665.412	49,989.241	111,837.42	9,760,221.71	9,129.35 (1; 1...)
ORDERDATE	Local Date	0	84	?	?	?	?	?	?	?	?	?	2018-11-14 (3; ...)
DAYS_SINCE_LAST_ORDER	Number (integer)	0	85	1,767	3,530	2,772.5	2,915	3,135.5	2,920.337	235.217	318.84	259,910	2,780 (2; 2.25...)
STATUS	String	0	3	?	?	?	?	?	?	?	?	?	Shipped (86; 9...)
PRODUCTLINE	String	0	6	?	?	?	?	?	?	?	?	?	Classic Cars (4...)
MSRP	Number (double)	0	89	82.714	123.471	94.83	100.533	106.987	100.692	6.699	8.183	8,961.617	82.714 (1; 1.12...)
PRODUCTCODE	Number (integer)	0	36	3	106	17.5	24	31	25.809	8.913	13.704	2,297	26 (7; 7.87%), ...
PHONE	String	0	88	?	?	?	?	?	?	?	?	?	6175558555 (...)
ADDRESSLINE1	String	0	89	?	?	?	?	?	?	?	?	?	1 rue Alsace-L... ...
CITY	String	0	71	?	?	?	?	?	?	?	?	?	NYC (5; 5.62%)
POSTALCODE	String	0	73	?	?	?	?	?	?	?	?	?	10022 (5; 5.62...)
COUNTRY	String	0	19	?	?	?	?	?	?	?	?	?	USA (32; 35.96...)
CONTACTLASTNAME	String	0	76	?	?	?	?	?	?	?	?	?	Young (4; 4.49...)
CONTACTFIRSTNAME	String	0	72	?	?	?	?	?	?	?	?	?	Valarie (4; 4.49...)
DEALSIZE	Number (integer)	0	41	3	259	19.5	26	32	30.865	13.883	31.37	2,747	26 (9; 10.11%), ...
Recency	Number (long)	0	76	31	526	100.5	215	261.5	210.607	98.658	128.894	18,744	239 (3; 3.37%), ...
Monetary	Number (double)	0	89	9,129.35	912,294.11	69,770.425	87,489.23	121,839.975	109,665.412	49,989.241	111,837.42	9,760,221.71	9,129.35 (1; 1...)
ORDERNUMBER_BINNED	String	0	4	?	?	?	?	?	?	?	?	?	Bin 1 (27; 30.3...)
Recency_BINNED	String	0	4	?	?	?	?	?	?	?	?	?	Bin 1 (23; 25.8...)
Monetary_BINNED	String	0	4	?	?	?	?	?	?	?	?	?	Bin 1 (23; 25.8...)
Recency (#1)	String	0	4	?	?	?	?	?	?	?	?	?	4 (23; 25.84%), ...
Frequency	String	0	4	?	?	?	?	?	?	?	?	?	1 (27; 30.34%), ...
Monetary (#1)	String	0	4	?	?	?	?	?	?	?	?	?	1 (23; 25.84%), ...
Concatenate	String	0	30	?	?	?	?	?	?	?	?	?	111 (12; 13.48...)

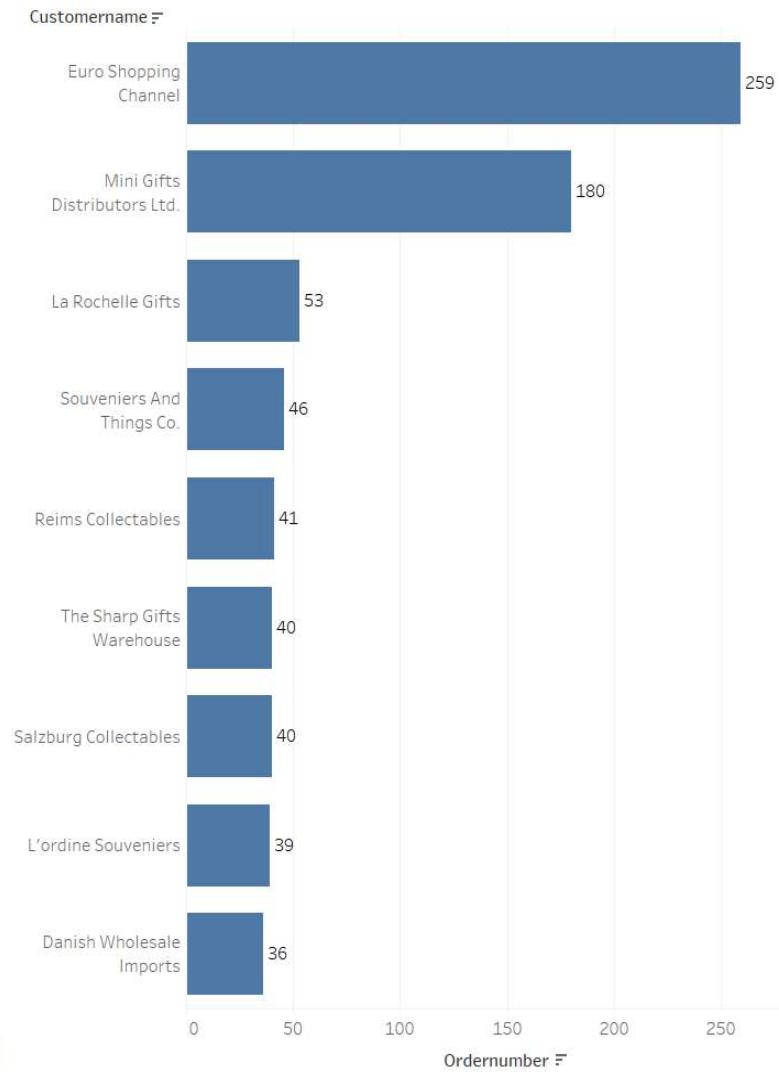


**BEST CUSTOMERS**

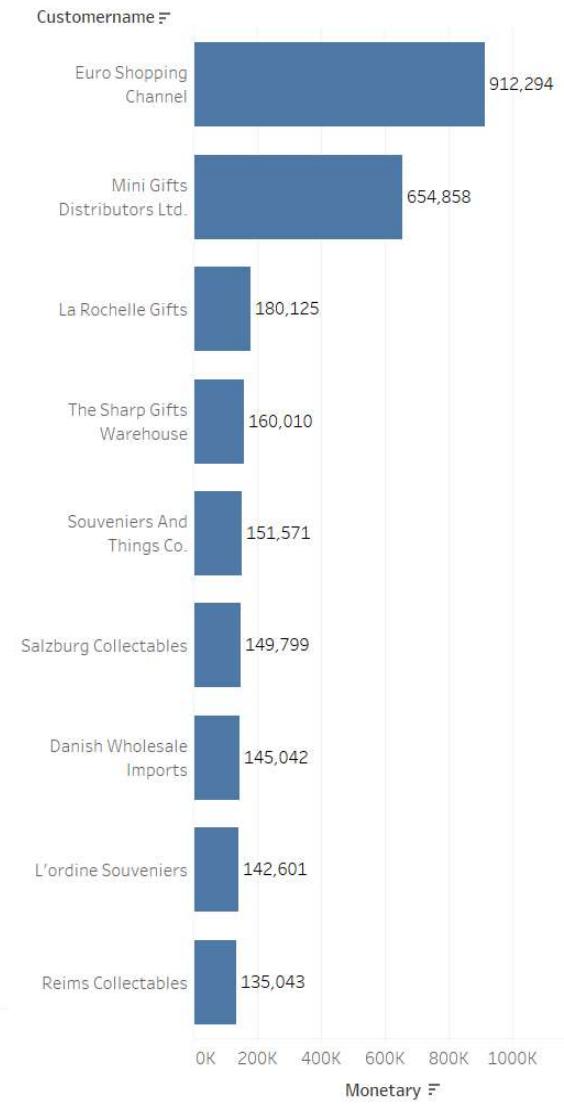
Best-Recency



Best-Frequency

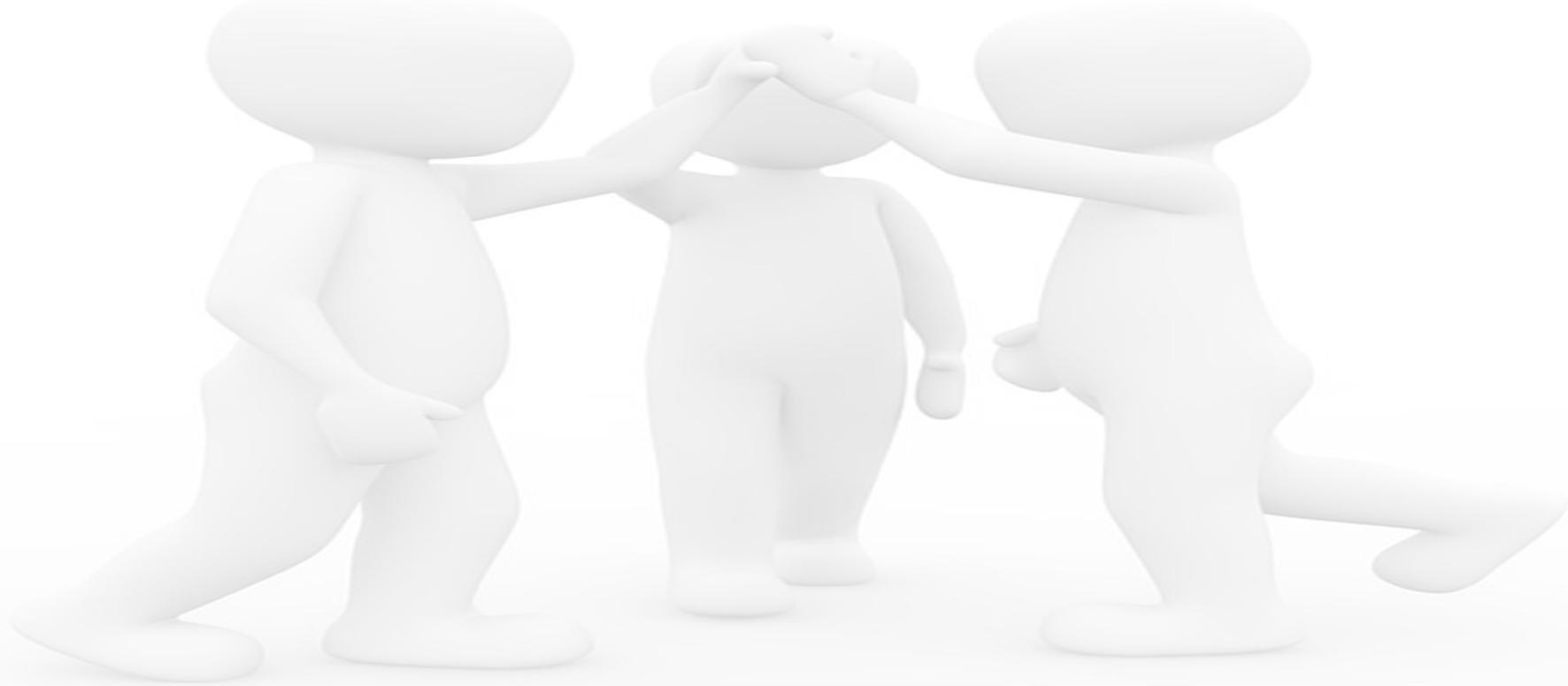


Best-Monetary



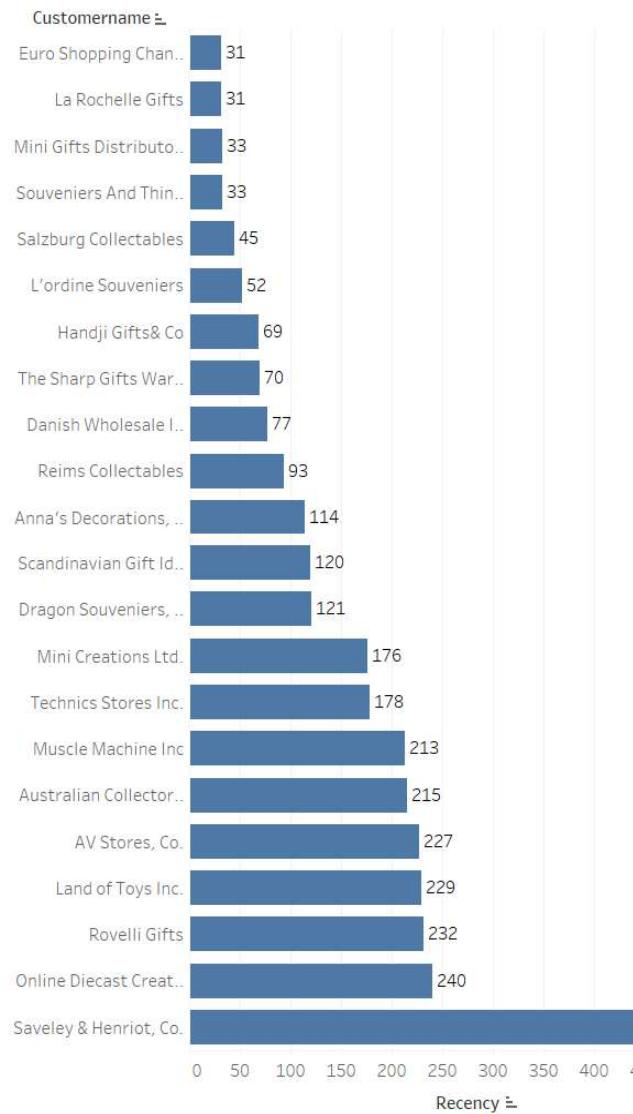
# INSIGHTS & RECOMMENDATIONS

- To retain these, any amount of offers could be given, some of which include:
  - Lowering of MSRP
  - Offering Big Deals
  - Expediting shipments

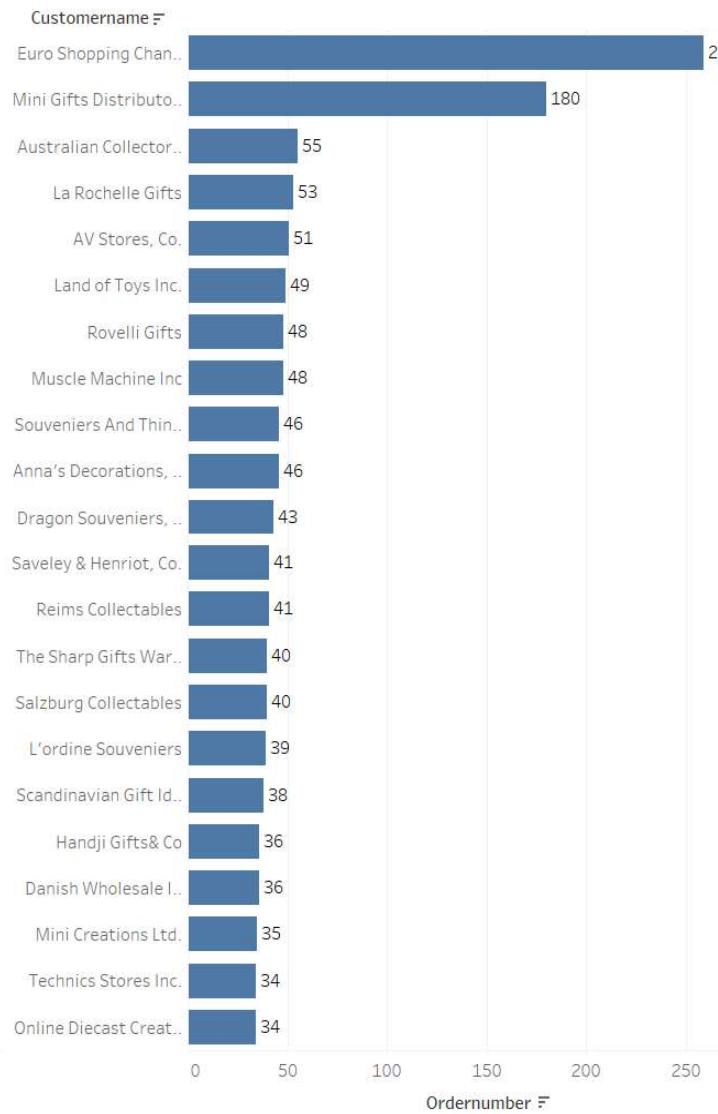


**LOYAL CUSTOMERS**

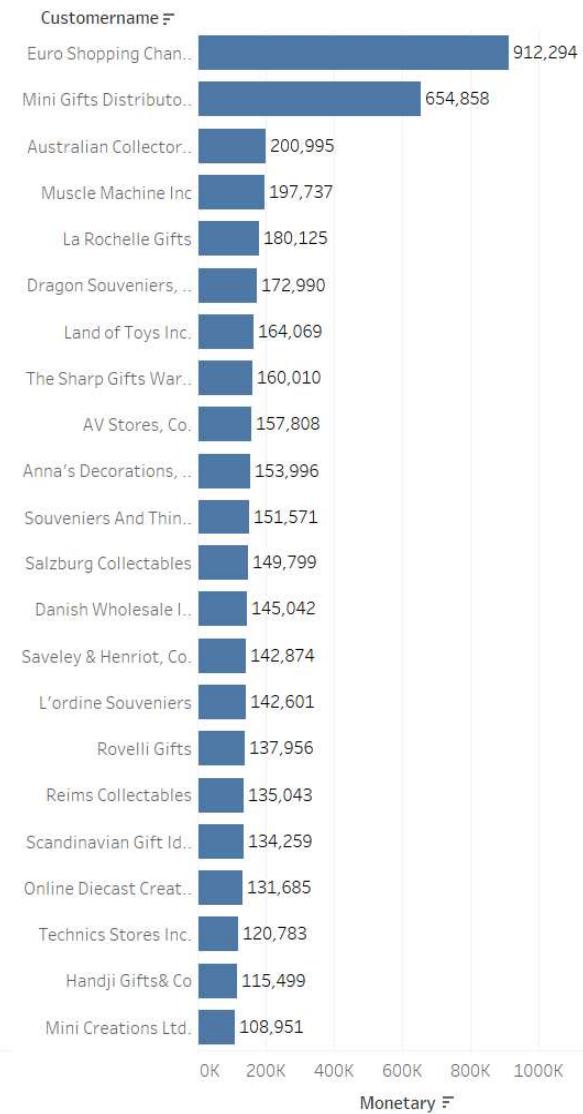
## Loyal-Recency



## Loyal-Frequency



## Loyal-Monetary



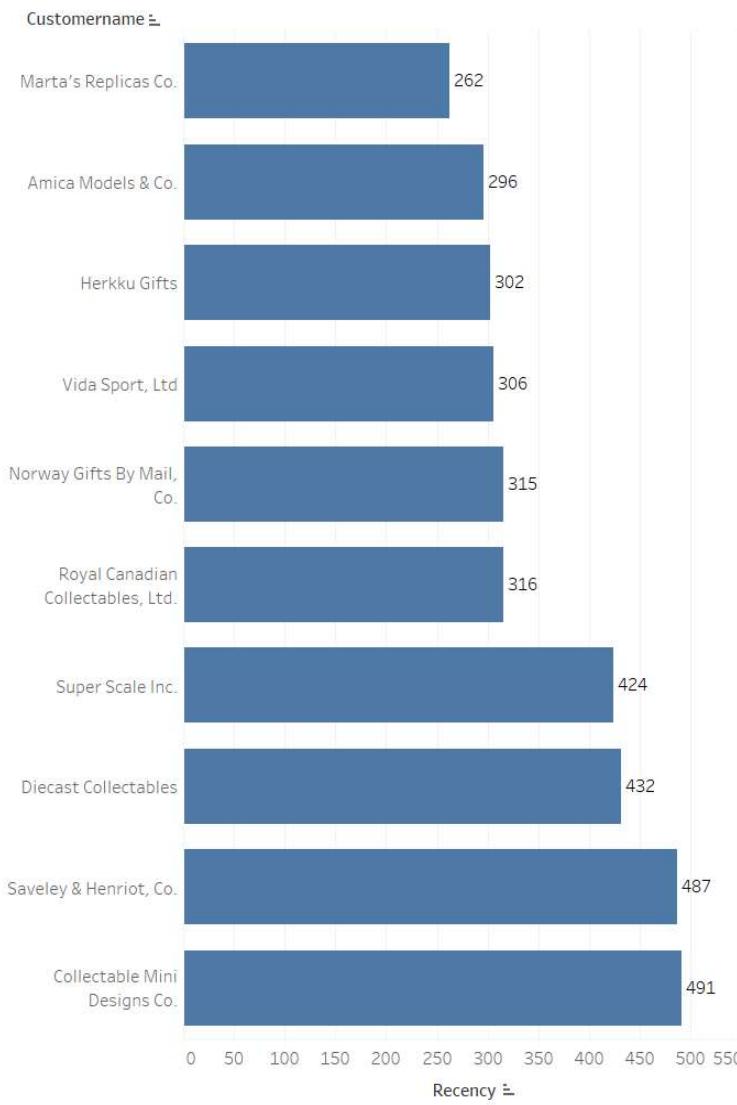
# INSIGHTS & RECOMMENDATIONS

- Though these customers have a high frequency, their monetary contributions need to improve.
- This could be achieved by :
  - Offering better deals
  - Upselling product lines with higher MSRP
  - Offering package deals to encourage buying more quantity
- Care should be taken to improve the recency of purchase for those customers, whose recency is low

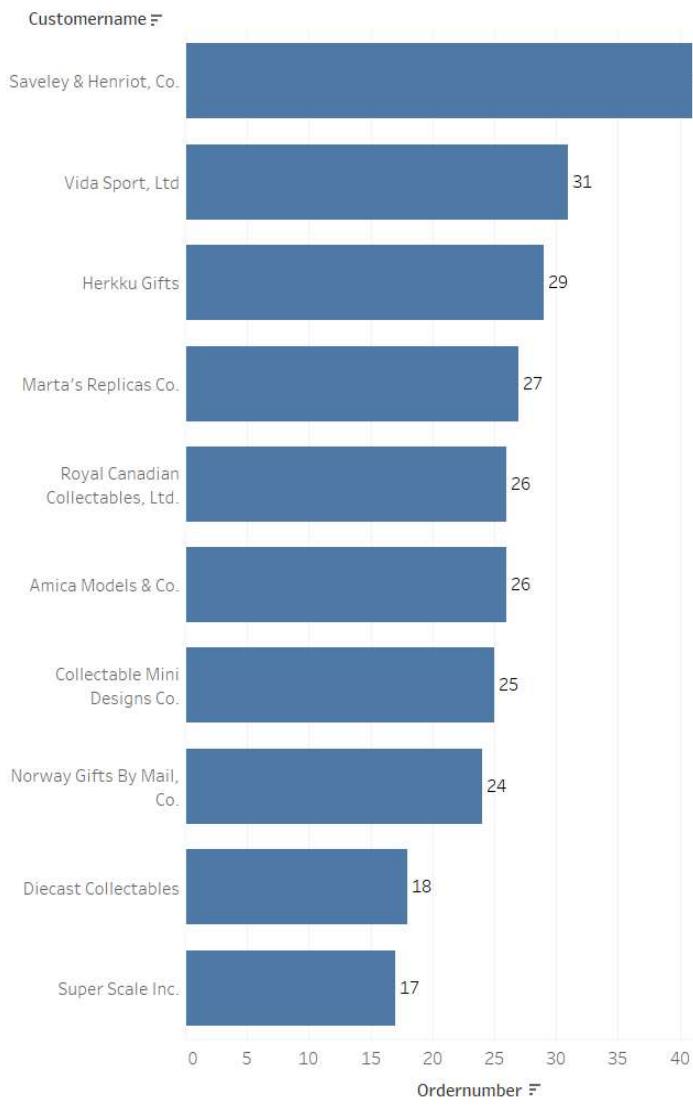


**CUSTOMERS ABOUT TO CHURN**

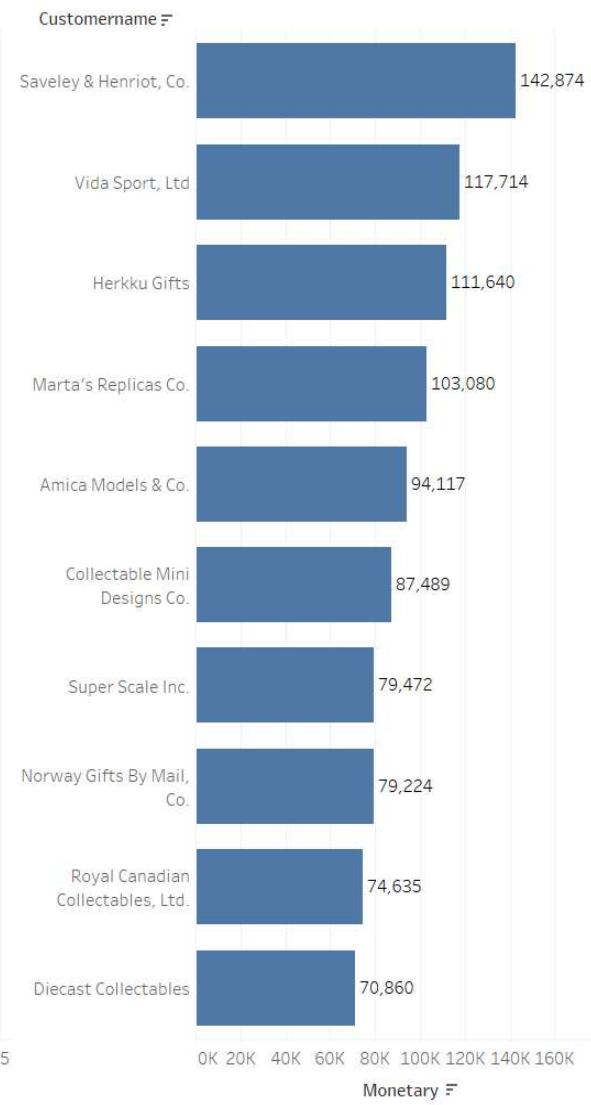
### Churn-Recency



### Churn-Frequency



### Churn-Monetary



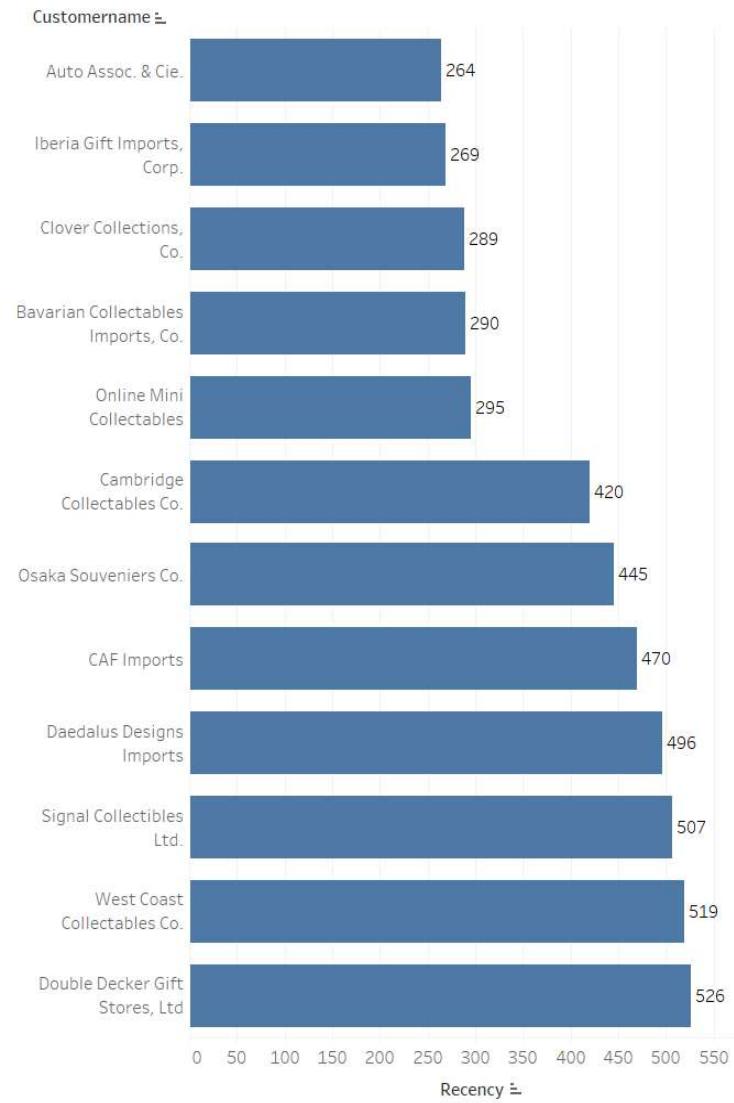
# INSIGHTS & RECOMMENDATIONS

- The reasons for the increase in the gap of purchase of these customers need to be investigated
- If no action is taken, these customers could churn, leading to a loss of about \$35k per year
- Particular care should be taken in the case of Saveley Henriot & Co, which also falls under the loyal customer category, and also have a good Monetary score
- Efforts to be taken to contact the customer and determine the reason for gap in purchases
- If it is something at the company's end, it has to be immediately addressed

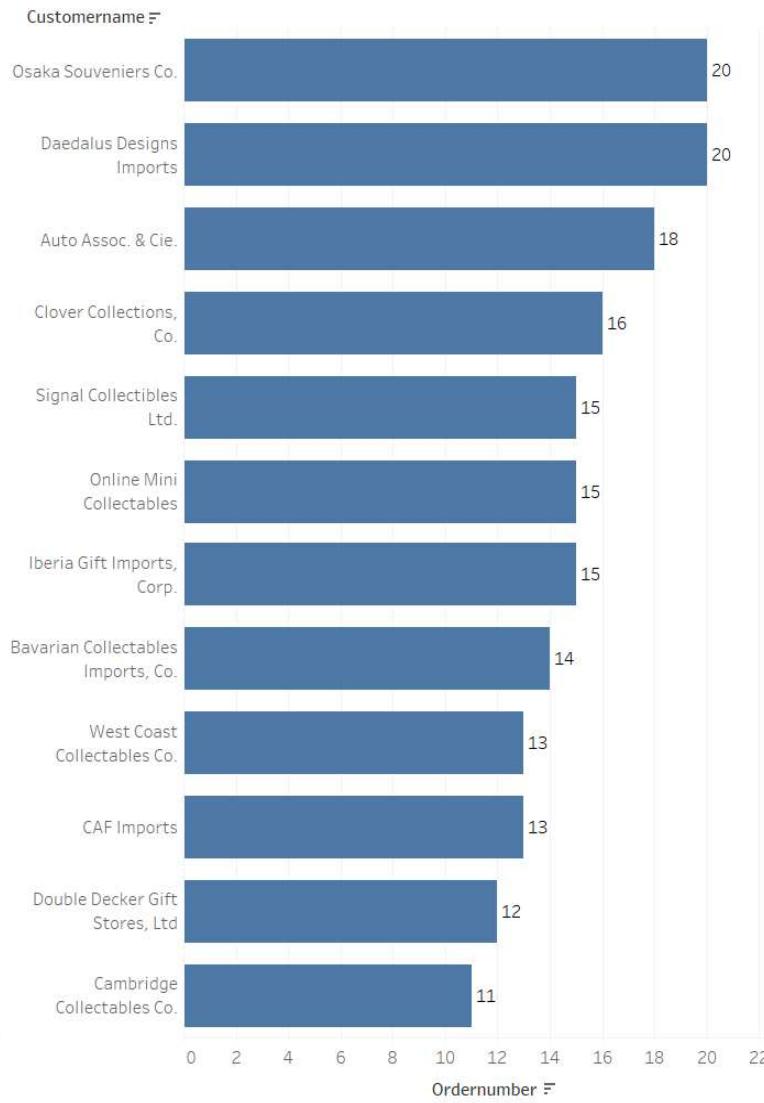
# **LOST CUSTOMERS**



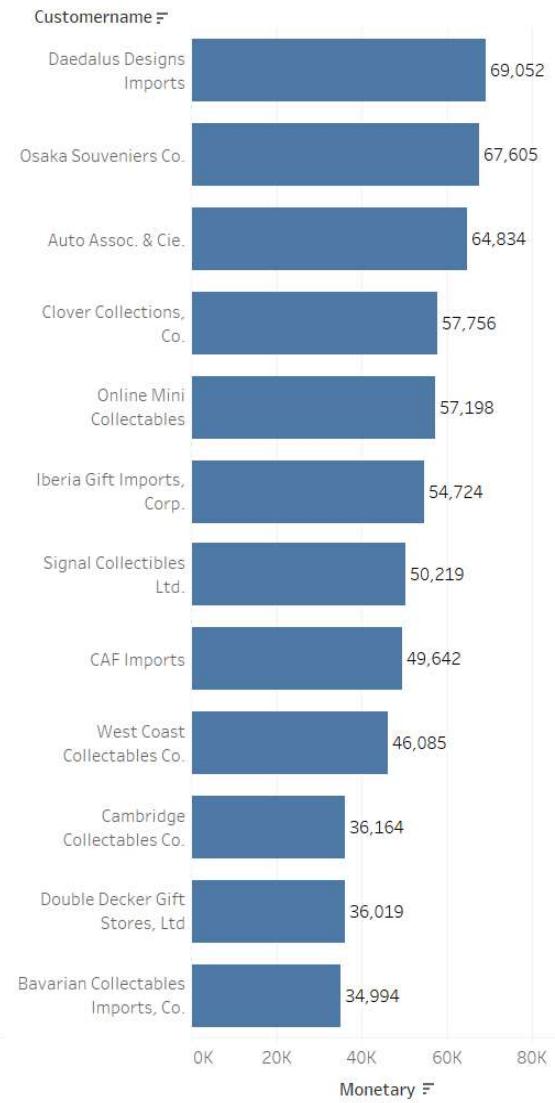
### Lost-Recency



### Lost-Frequency



### Lost-Monetary



# INSIGHTS & RECOMMENDATIONS

- Lost customers might prove hard to earn back
- A practical actionable deed could be to investigate the cause of the loss and improve the bottlenecks so that no more customers face the need to leave the company

**THANK YOU**