



# **AUTOMOBILE PART MANUFACTURE COMPANY'S DATA MRA PROJECT - MILESTONE 1**

BY - DURGESH BHARGAVA

PGPDSBA.O.JULY24.A



# CONTENTS

## Agenda & Executive Summary

### 1. Project Overview & Objective

1. Problem Statement
2. Executive Summary

### 2. Data Overview

1. Data Dictionary
2. Summary Statistics
3. Assumptions Made About the Data

### 3. Exploratory Data Analysis & Insights

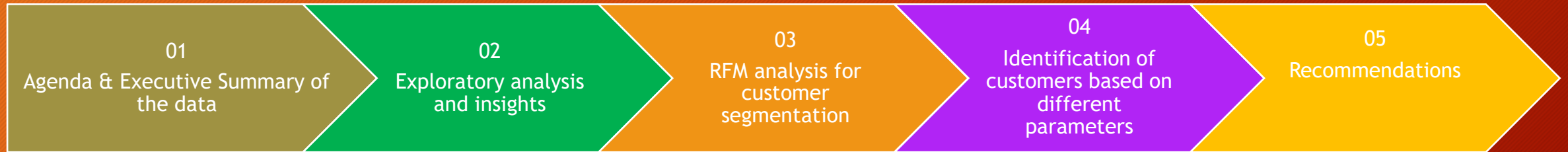
1. Univariate, Bivariate, and Multivariate Analysis (with Visualizations)
2. Sales Trends: Weekly, Monthly, Quarterly, and Yearly
3. Sales Distribution Across Categories and Features
4. Key Inferences Drawn from the Analysis

### 4. Customer Segmentation Using RFM Analysis

1. Introduction to RFM (Recency, Frequency, Monetary)
2. Tools Used for Analysis (e.g., KNIME)
3. Parameters Considered and Assumptions Made
4. Output Snapshot (Table Head)
5. Workflow Diagram (Insert KNIME Workflow Image)
6. Segment Insights:
  1. Identification of Top Customers
  2. Customers at Risk of Churning
  3. Lost Customers



# AGENDA:-





# Agenda & Executive Summary of the data

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

# Problem statement

An automobile parts manufacturing company has accumulated transaction data over a span of three years. Lacking an internal data science team, they have brought you on board as a consultant. Your role is to leverage your data science expertise to uncover meaningful insights about their data and customer behavior.

# Executive Summary:

## Project Summary

- **Data:** Last 3 years of customer transactions
- **Dataset Size:** 20 columns, 2,747 rows
- **Data Quality:** No missing or duplicate values; a few outliers in some columns

## Objective

- Identify customer buying patterns
- Recommend personalized marketing strategies for different customer segments

## Approach

- Performed **Exploratory Data Analysis (EDA)** to find key sales trends and patterns
- Conducted **RFM Analysis** to group customers into 4 segments based on their purchase behavior
- Suggested custom marketing strategies for each customer group



# Data Dictionary

- ORDERNUMBER : Order Number
- CUSTOMERNAME : customer
- QUANTITYORDERED : Quantity ordered
- PHONE : Phone of the customer
- PRICEEACH : Price of Each item
- ADDRESSLINE1 : Address of customer
- ORDERLINENUMBER : order line
- CITY : City of customer
- SALES : Sales amount
- POSTALCODE : Postal Code of customer
- ORDERDATE : Order Date
- COUNTRY : Country customer
- DAYS\_SINCE\_LASTORDER : Days\_ Since\_Lastorder
- CONTACTLASTNAME : Contact person customer
- STATUS : Status of order like Shipped or not
- CONTACTFIRSTNAME : Contact person customer
- PRODUCTLINE : Product line - CATEGORY
- DEALSIZE : Size of the deal based on Quantity and Item Price
- MSRP : Manufacturer's Suggested Retail Price
- PRODUCTCODE : Code of Product

Numeric columns: ORDERNUMBER, QUANTITYORDERED, PRICEEACH, ORDERLINENUMBER  
SALES, DAYS\_SINCE\_LASTORDER, MSRP.

# ASSUMPTIONS

Each row in the data represents a unique transaction made by a customer.

The customer segments may be defined based on the purchasing frequency, amount spent, and recency of purchases (RFM Analysis).

The marketing strategies may vary for each customer segment, and the company may need to personalize their marketing efforts accordingly.

The order date and days since last order columns are accurately calculated.

The sales column is calculated as the product of quantity ordered and price each.

The status column indicates the current status of the order accurately.

The recommendations provided in the presentation are based on the insights gained from the analysis of the transaction data



# Statistical Summary

	ORDERNUMBER	QUANTITYORDERED	PRICEEACH	ORDERLINENUMBER	SALES	ORDERDATE	DAYS_SINCE_LASTORDER	MSRP
count	2747.000000	2747.000000	2747.000000	2747.000000	2747.000000	2747	2747.000000	2747.000000
mean	10259.761558	35.103021	101.098951	6.491081	3553.047583	2019-05-13 21:56:17.211503360	1757.085912	100.691664
min	10100.000000	6.000000	26.880000	1.000000	482.130000	2018-01-06 00:00:00	42.000000	33.000000
25%	10181.000000	27.000000	68.745000	3.000000	2204.350000	2018-11-08 00:00:00	1077.000000	68.000000
50%	10264.000000	35.000000	95.550000	6.000000	3184.800000	2019-06-24 00:00:00	1761.000000	99.000000
75%	10334.500000	43.000000	127.100000	9.000000	4503.095000	2019-11-17 00:00:00	2436.500000	124.000000
max	10425.000000	97.000000	252.870000	18.000000	14082.800000	2020-05-31 00:00:00	3562.000000	214.000000
std	91.877521	9.762135	42.042548	4.230544	1838.953901	NaN	819.280576	40.114802

## Inference :

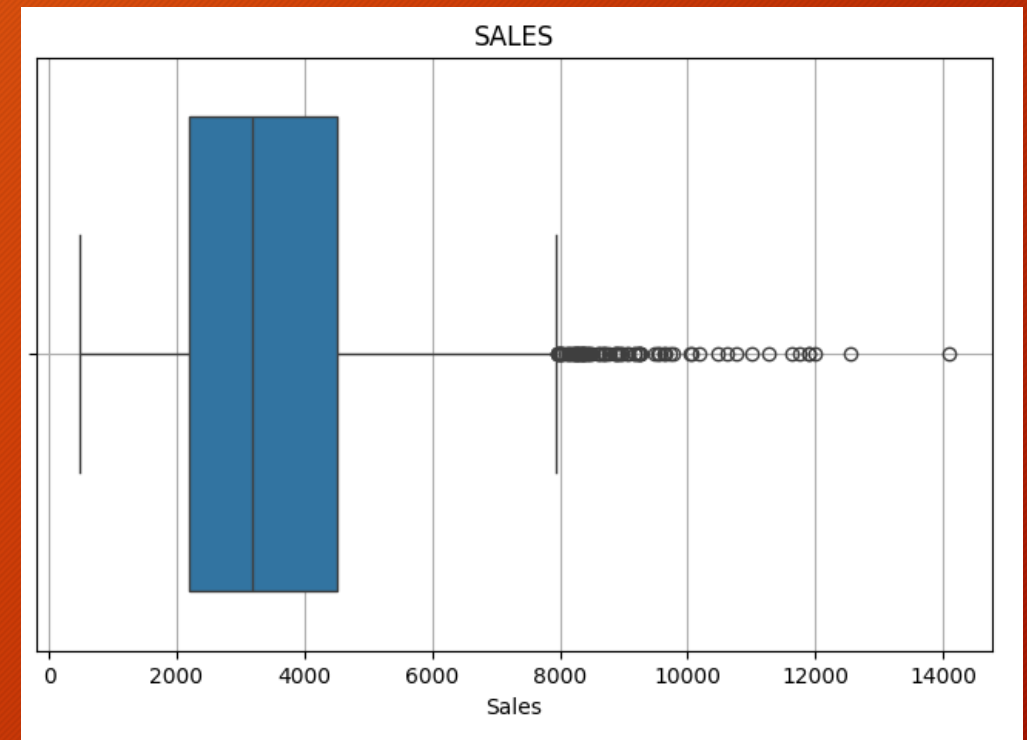
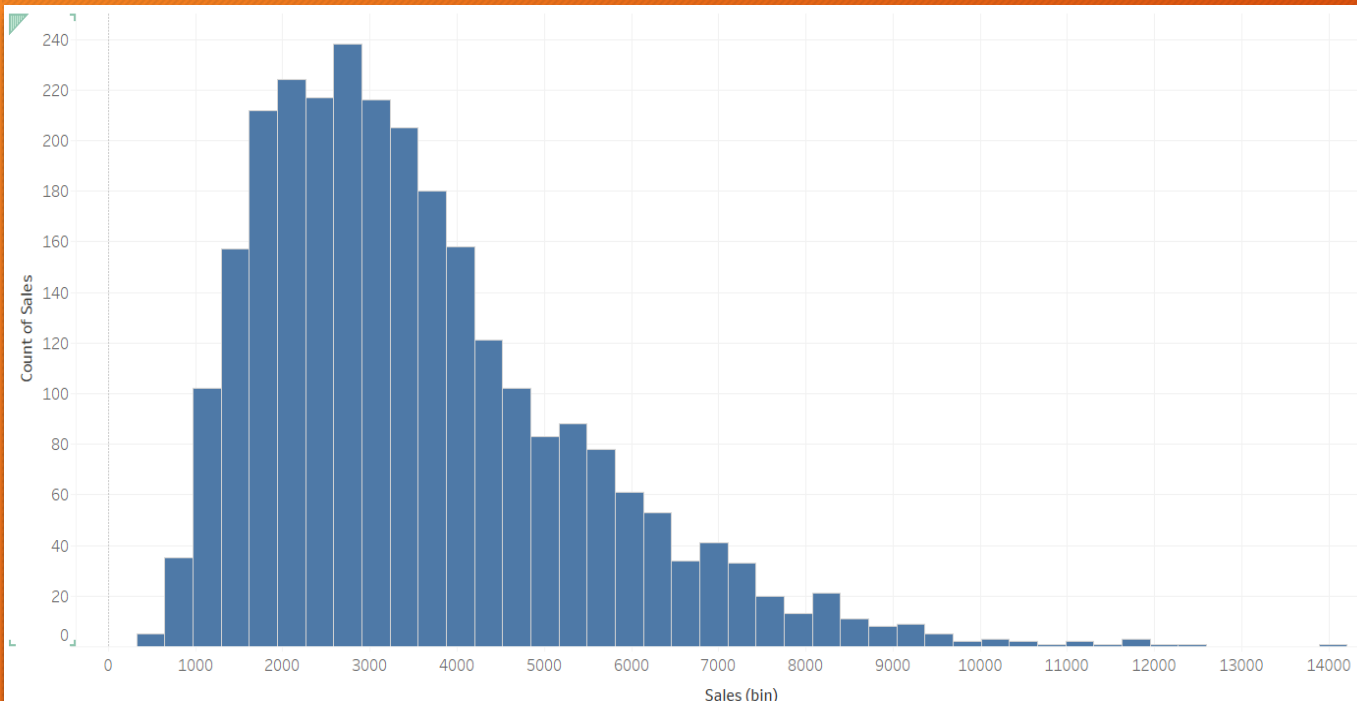
- The average number of items ordered per sales order is 35, with a standard deviation of 9.76.
- The average price of each item is 101.09, with a standard deviation of 42.04.
- The average sales amount per order is 3553.05, with a standard deviation of 1838.95.
- The average time since the last order is 1757.09 days, with a standard deviation of 819.28.
- The summary statistics do not indicate any red flags or abnormalities that could potentially indicate issues with the data.

# Exploratory Analysis & Insights

- 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
- Summarize the inferences from the above analysis



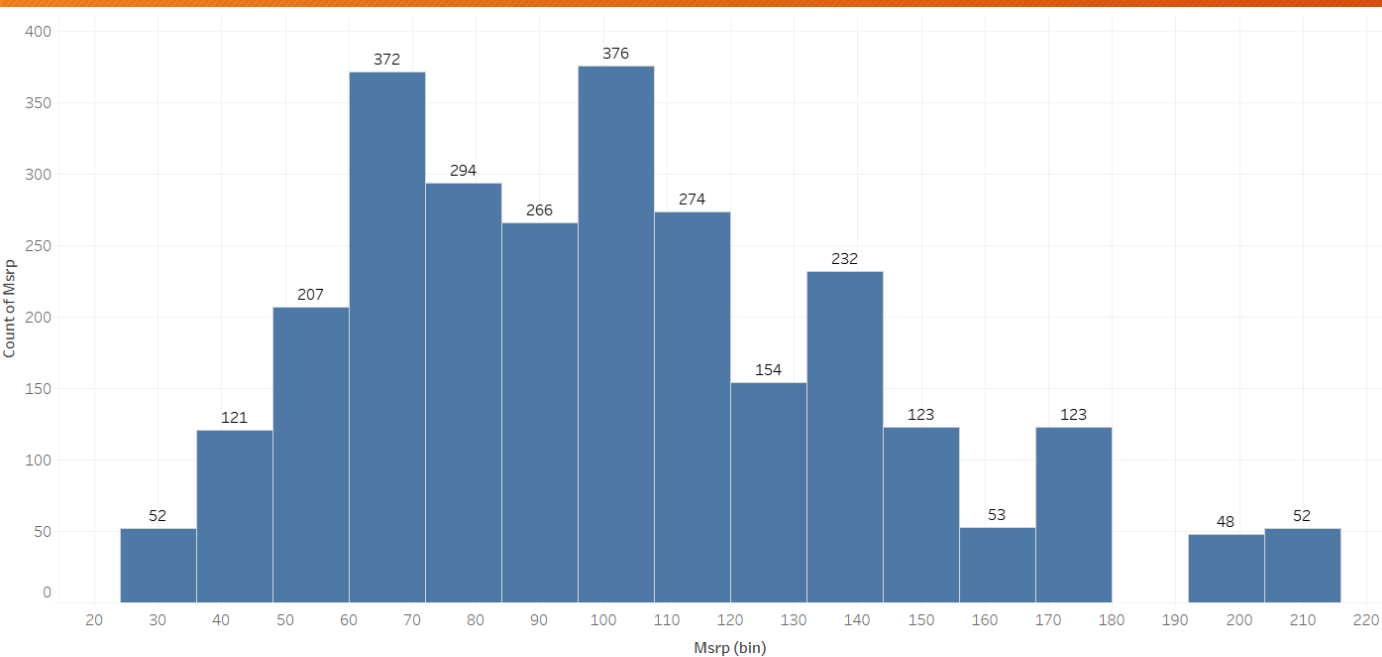
# UNIVARIATE ANALYSIS



- We can see that sales data is skewed toward left.
- There are many outliers.
- Most data exist between 1.5k - 3.5k.

SALES

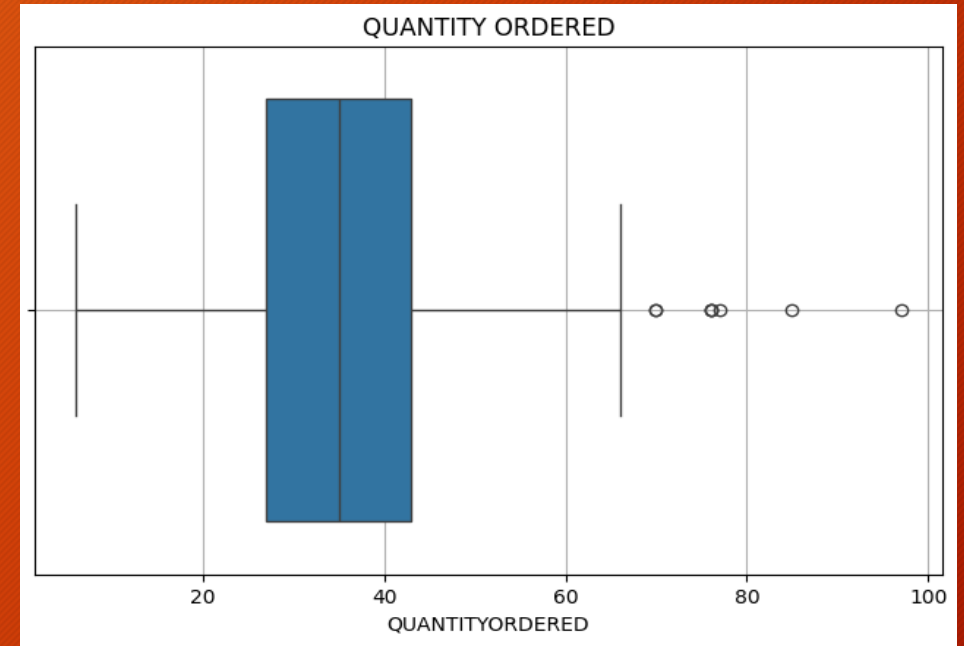
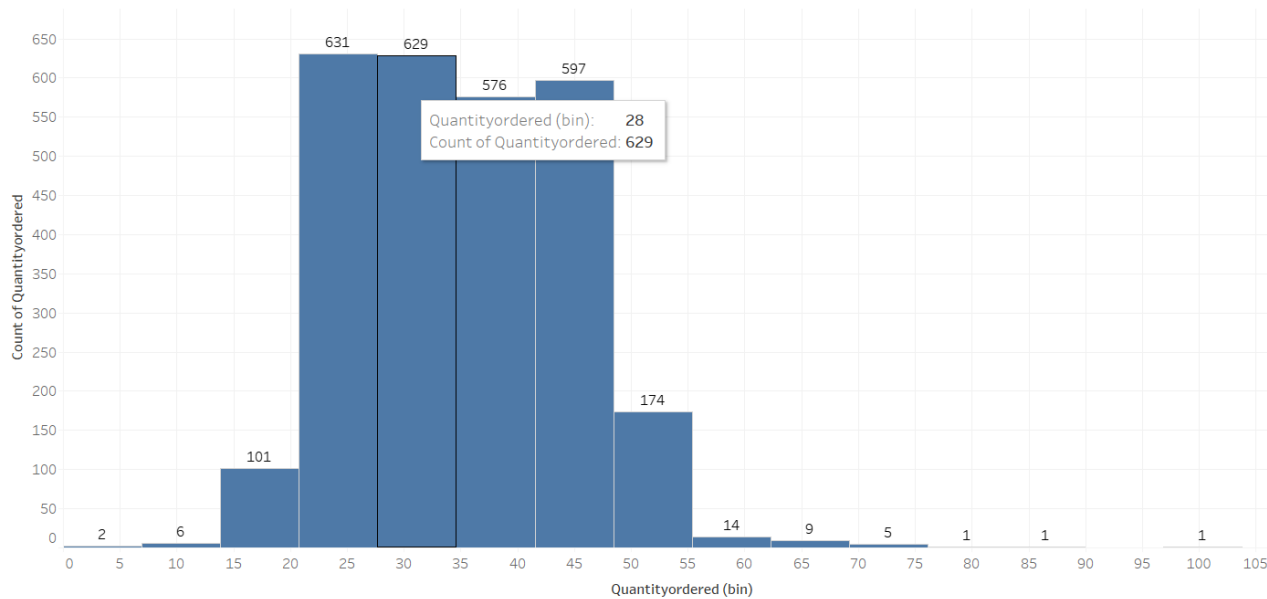




- Outlier are present in the MRSP Data
- This data is shewed toward left

# MRSP

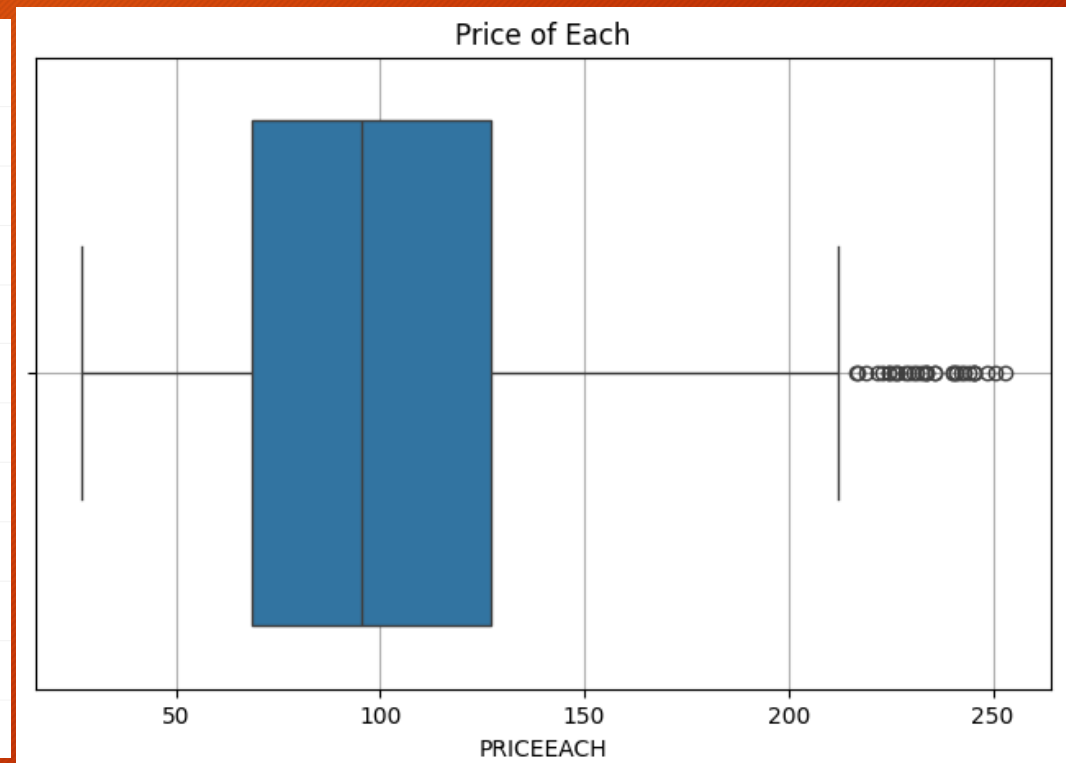
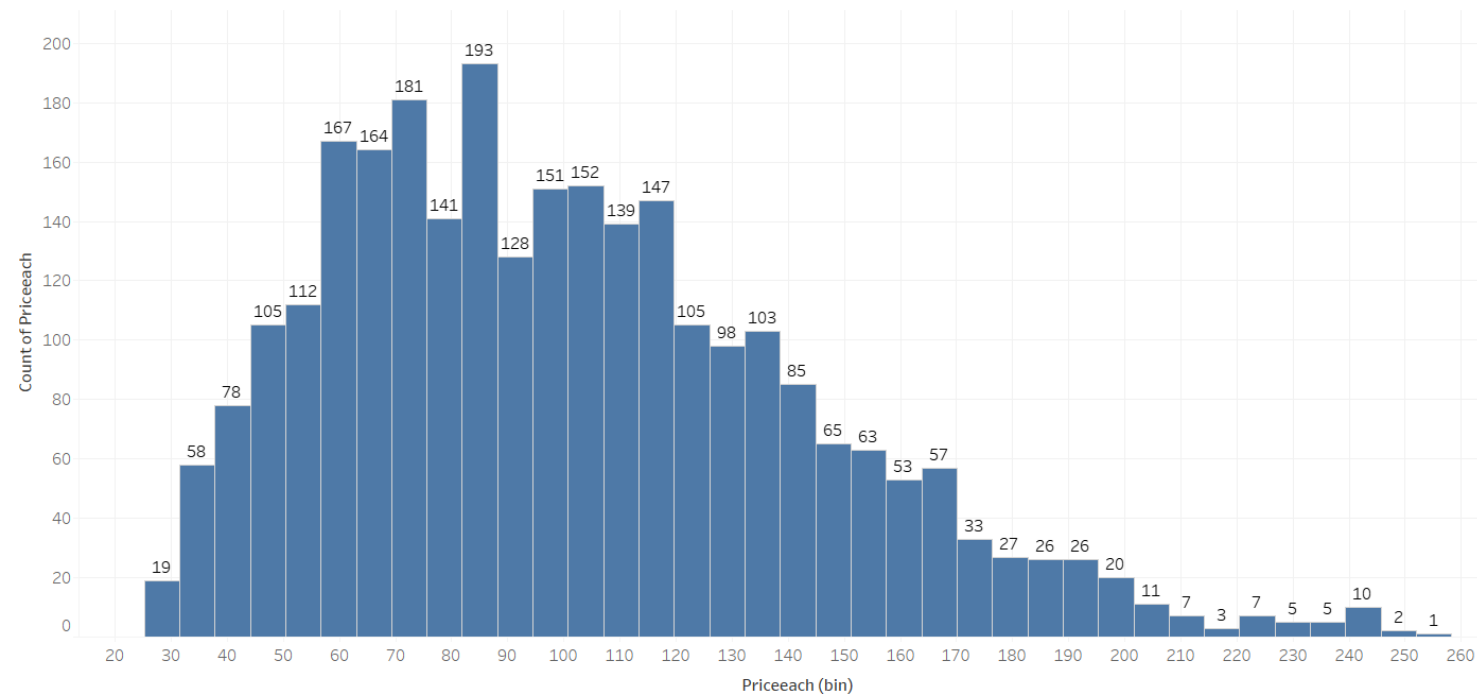
Quantity Ordered



Outliers are present in the data of Quantity Ordered.

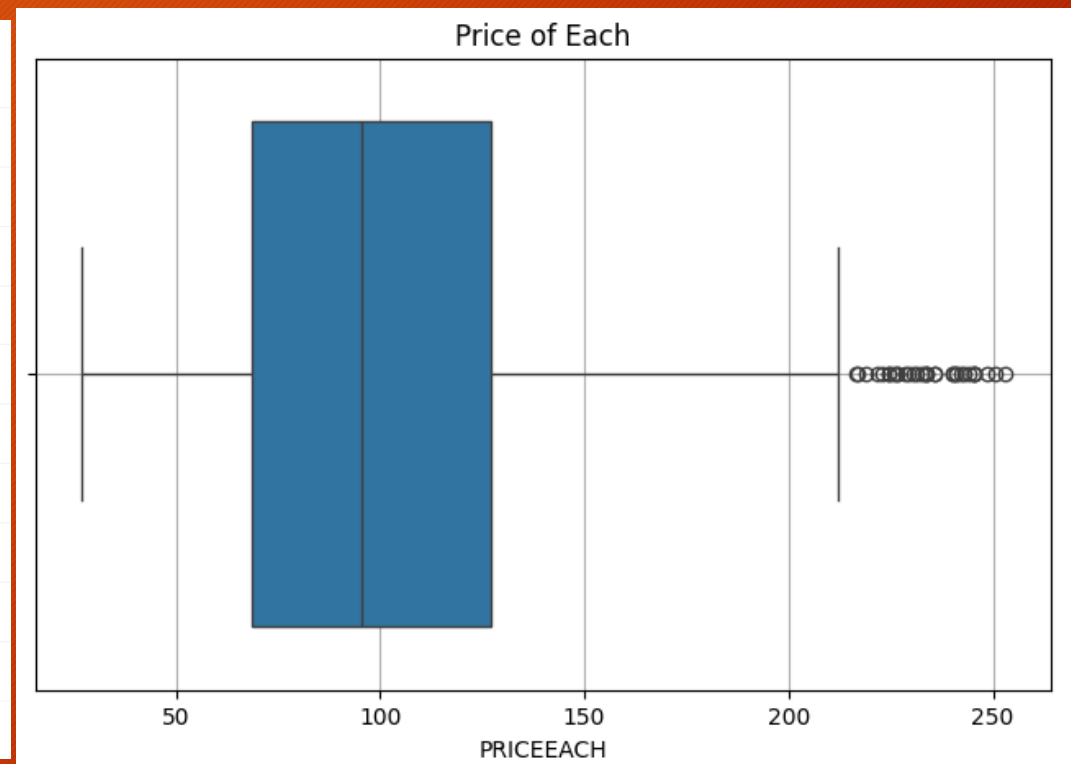
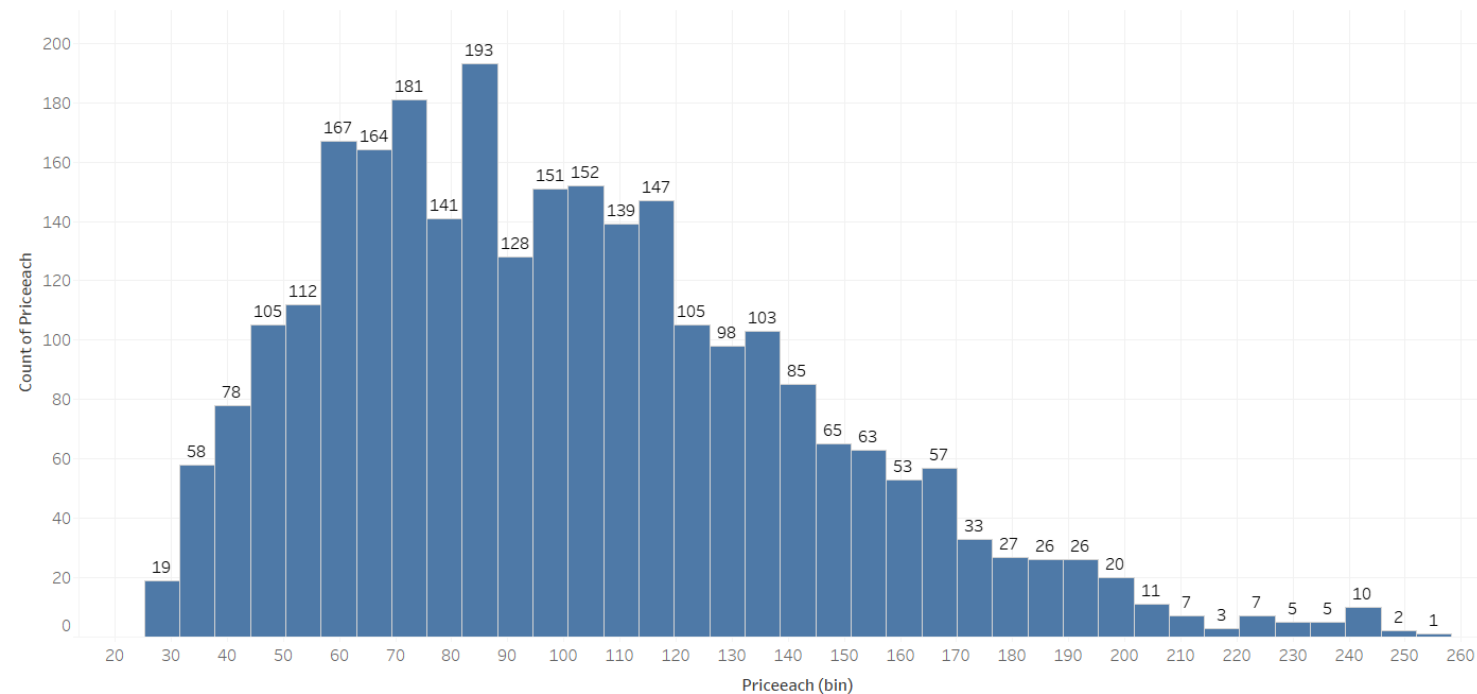
Quantity  
Ordered





Slightly right skewed but with many outliers

Price of  
each item



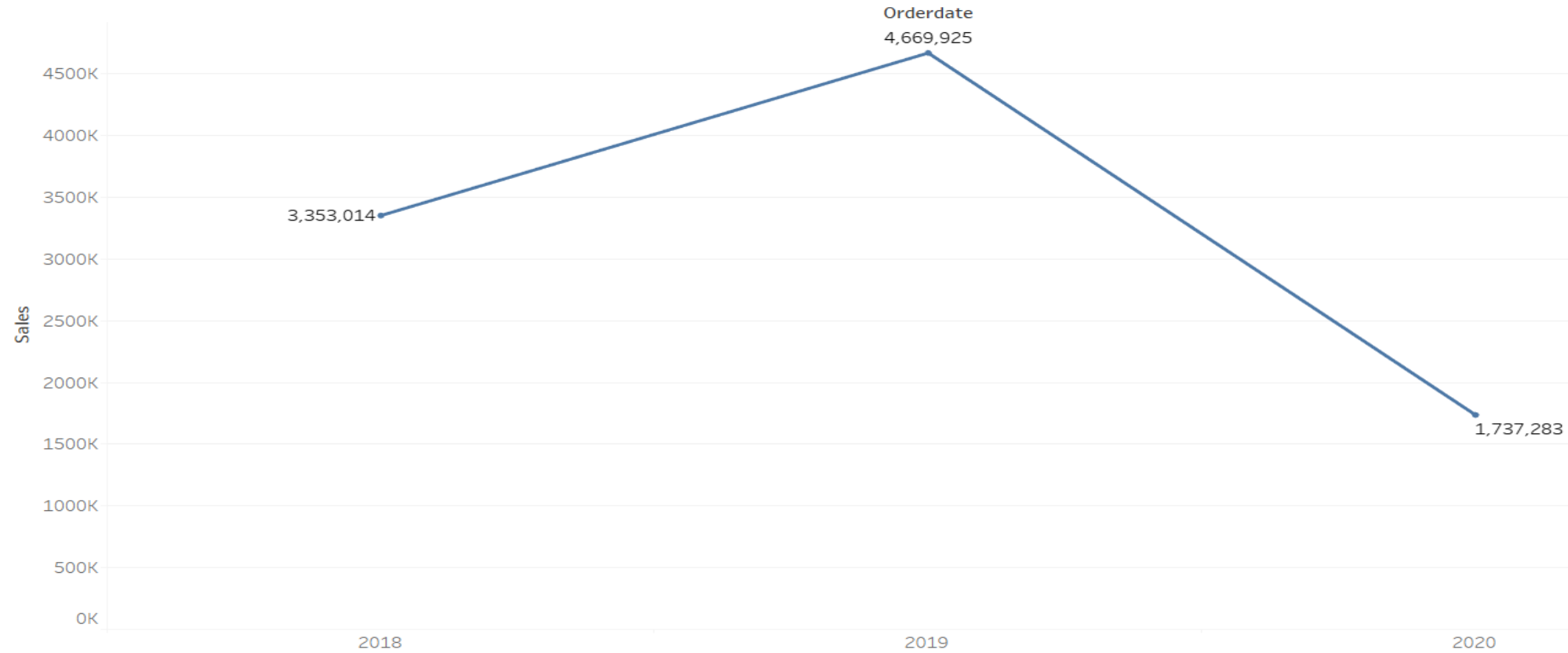
Slightly right skewed but with many outliers

Price of  
each item



# BIVARIATE ANALYSIS

## Year Sales

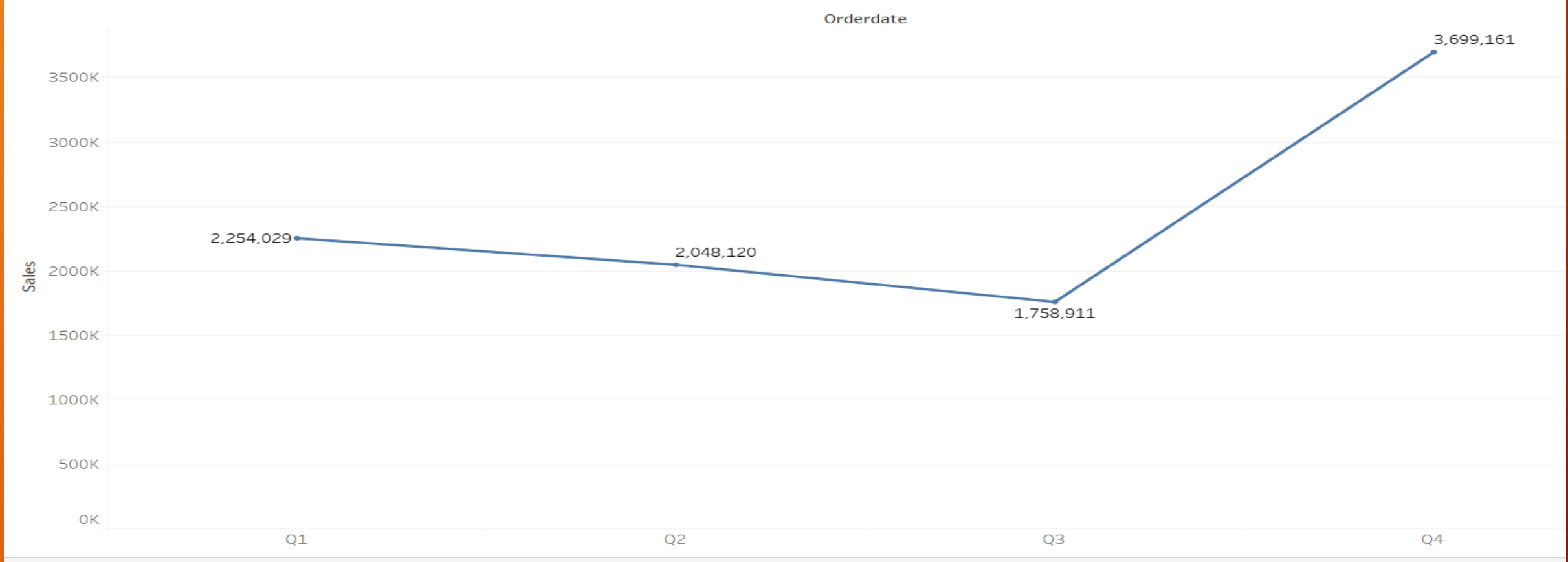


- We can see that there is a dip in the yearly sales
- We need to find more information regarding this as its not a good sign.

Yearly Sales



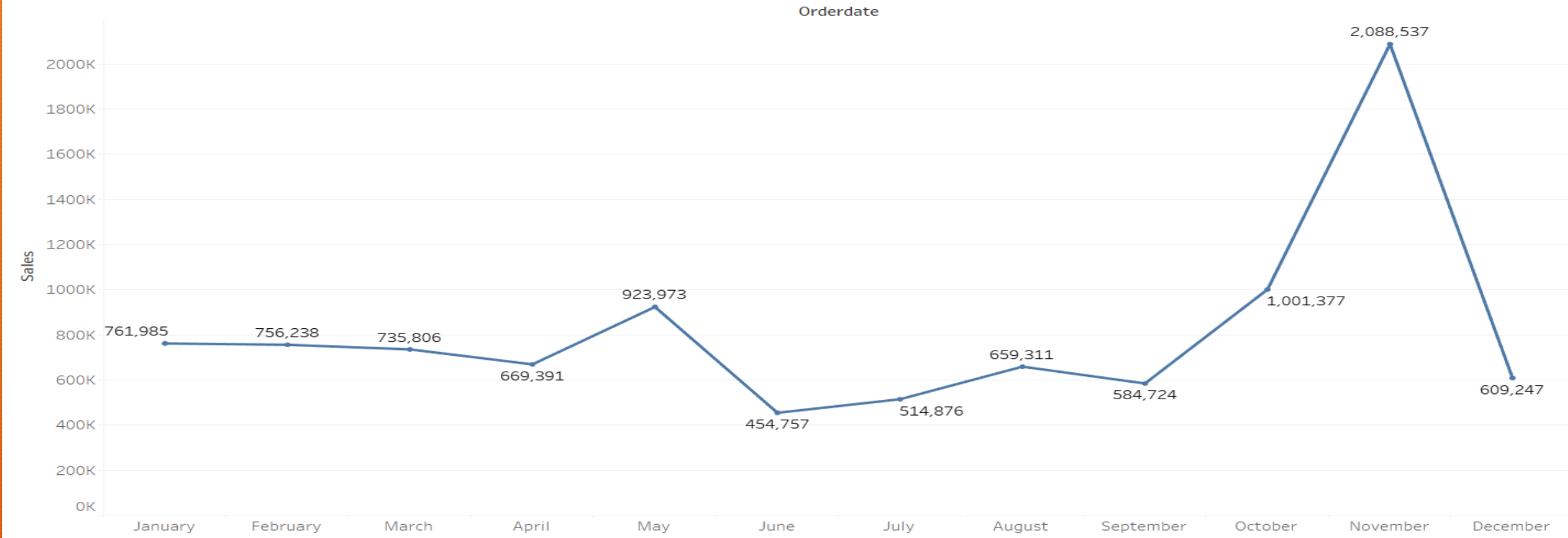
## Quarter Sales



- We can see that in the quarter 4 there is high sales as compared to other quarters.

Quarterly  
Sales

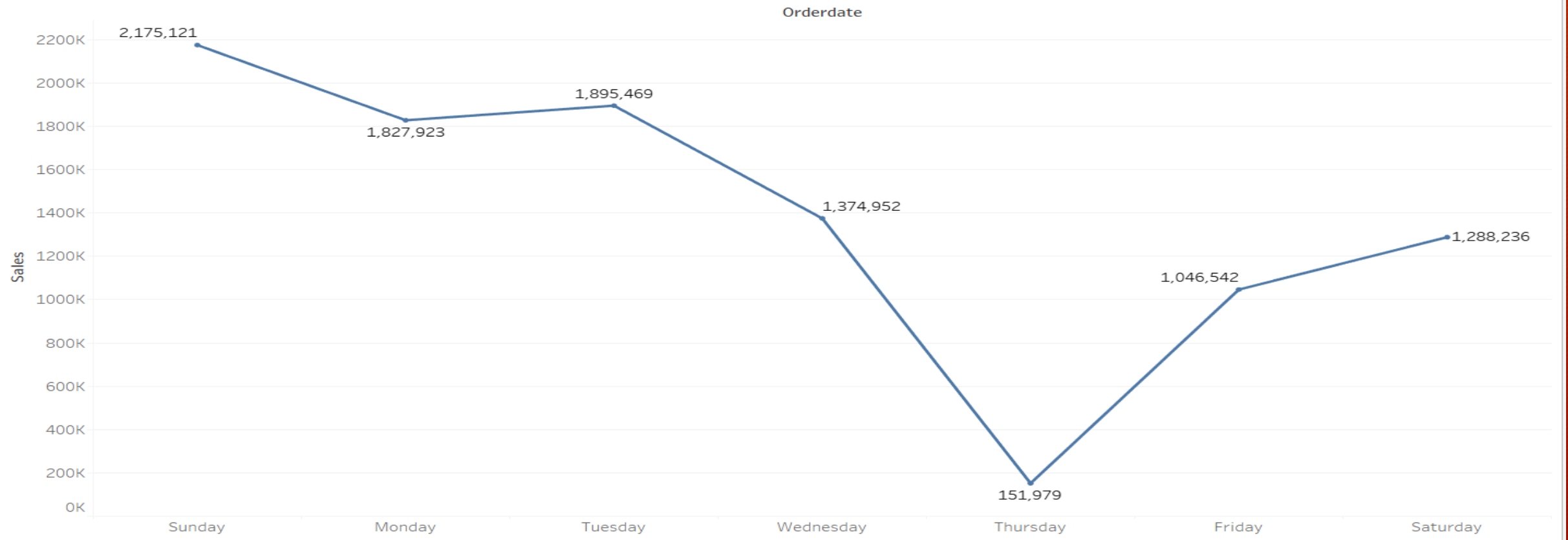
## Monthly Sales



- We can see that in the 11th month there is highest sales.
- And in 6th month the sales is the lowest.
- Sales is consistent in the first 4 months of the year

Monthly  
Sales

## Weekday Sales

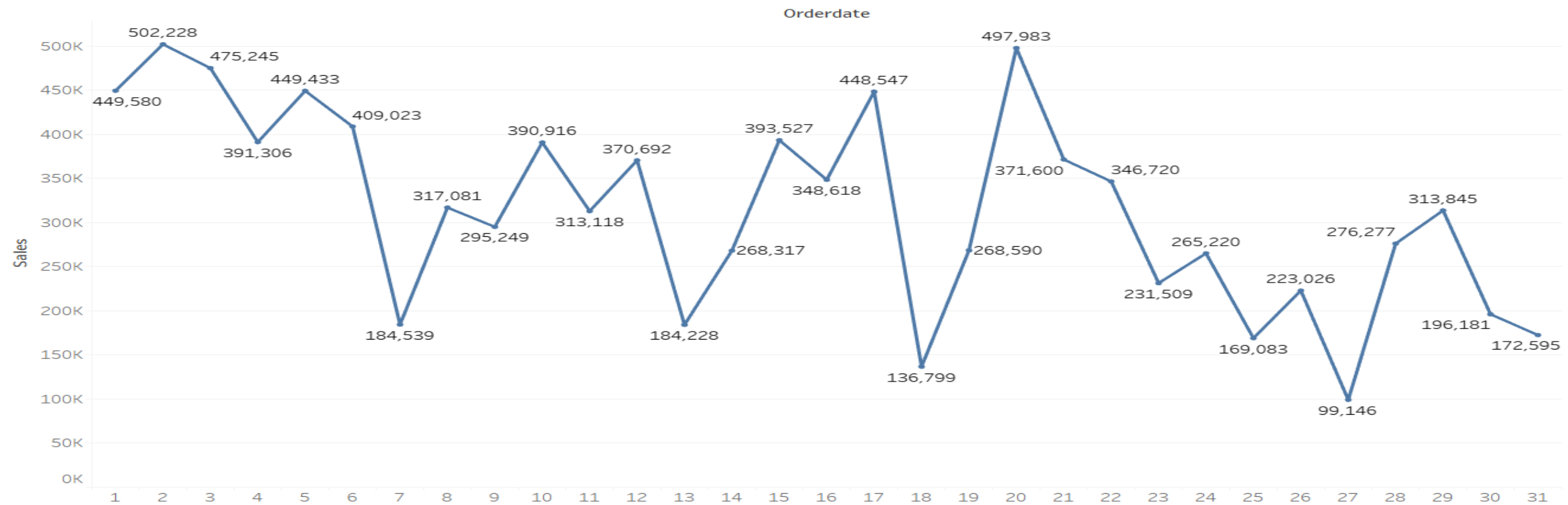


- We can see Thursday has the lowest sales and Sunday has the highest sales.
- Sales increases from Friday to Sunday and dips from Monday to Thursday

Weekdays  
Sales



## Day Sales

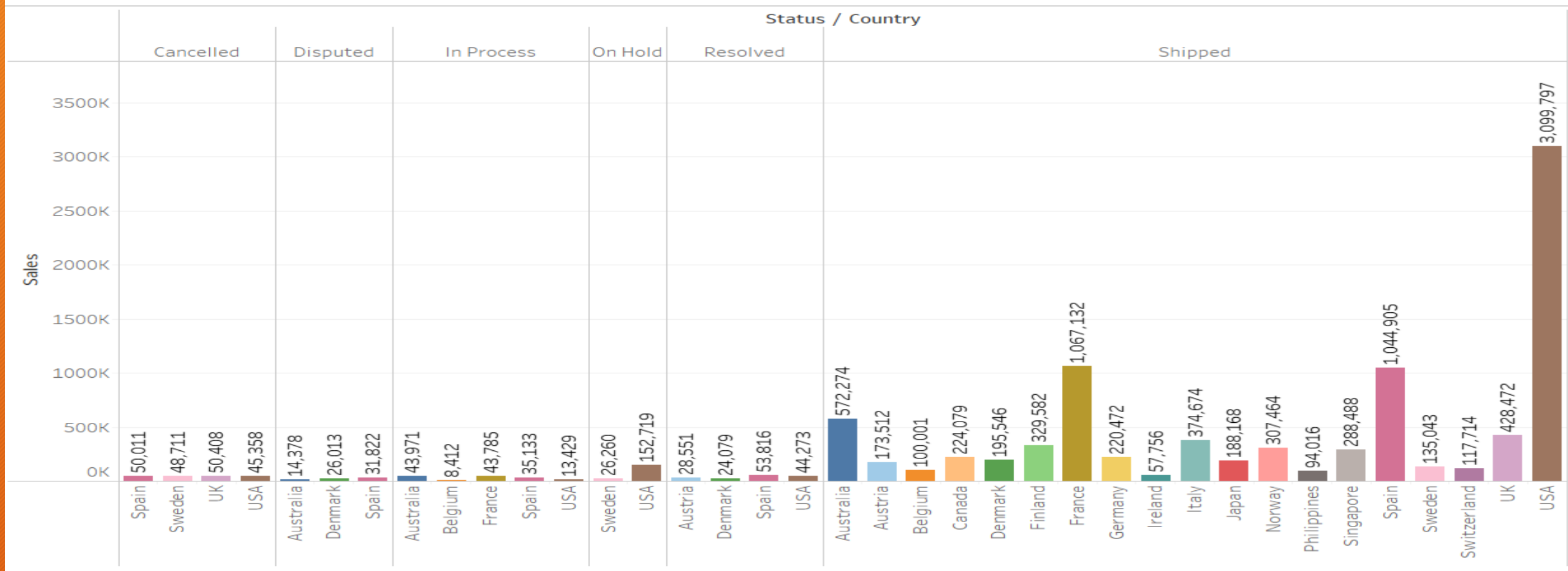


- In the stating days of month there is more sales as compared to the end days.

Day Sales

# MULTIVARIATE ANALYSIS

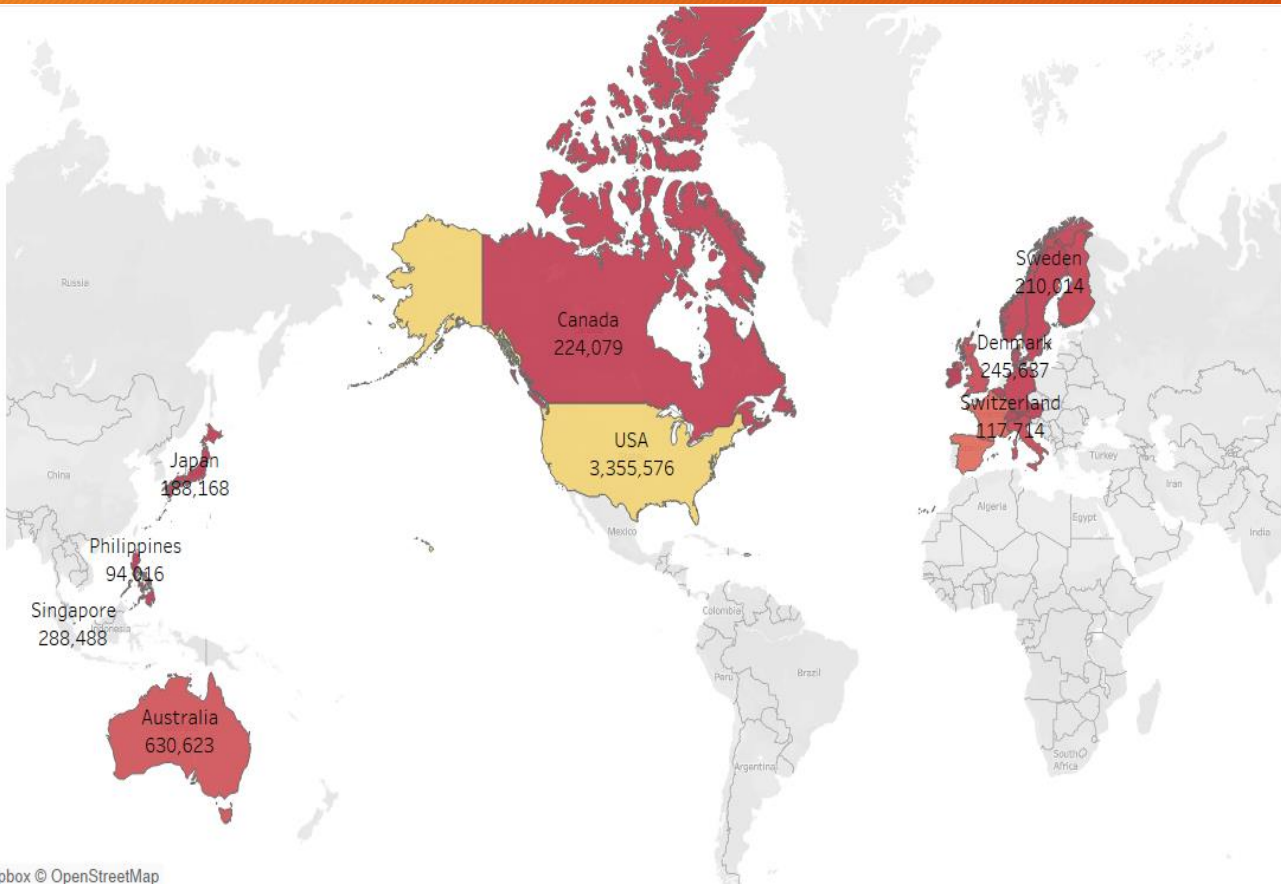
## Production-Shipped Countries



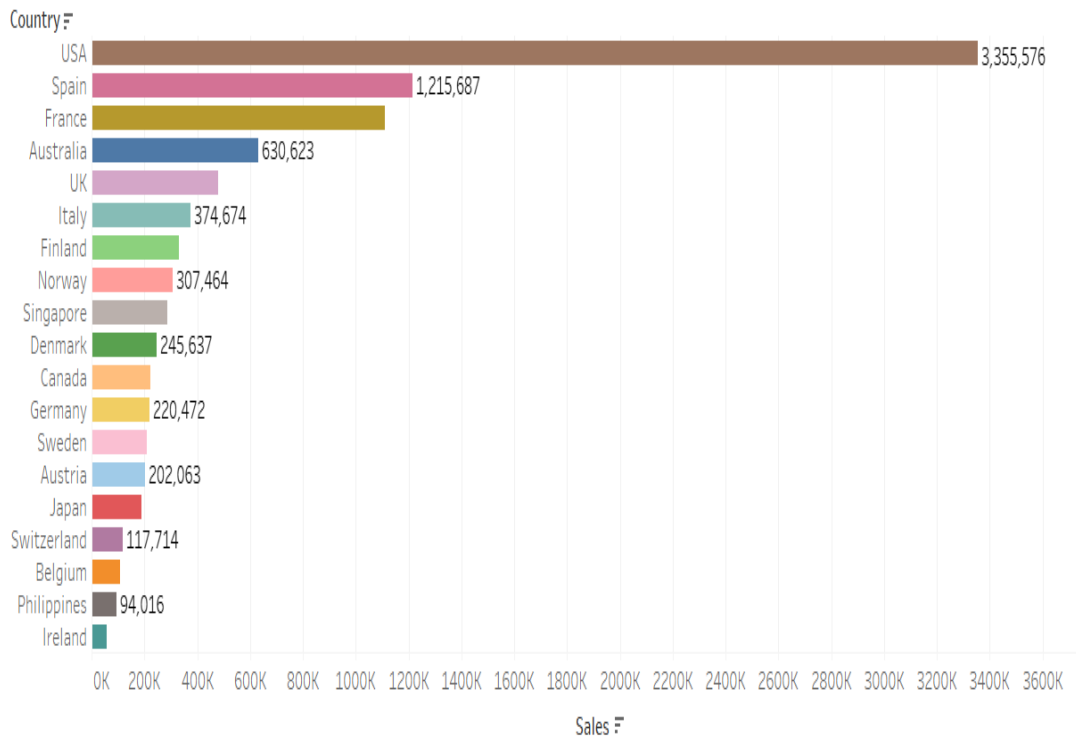
- We can see most of the orders which are on hold belongs to USA, some orders on hold belong to Sweden.
- UK, USA, Spain, Sweden have approx. same number of canceled orders.
- Spain has most disputes and also most disputes are solved.

Status,  
Country &  
Sales





### Country Sales



- Usa has most shipped orders.
- Spain and France have approx same number of orders.
- Ireland have the least number of orders shipped

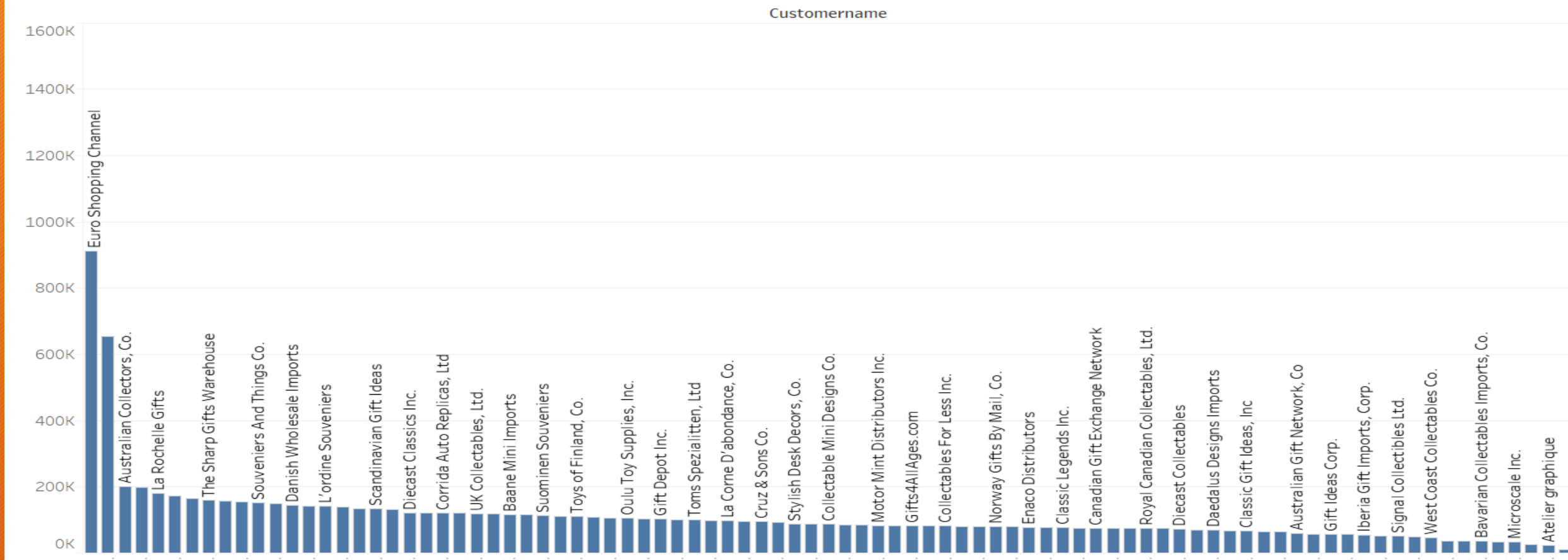
Order  
Shipped &  
Sales

Status	Productline						
	Classic Ca..	Motorcyc..	Planes	Ships	Trains	Trucks an..	Vintage C..
Cancelled	30.46%		18.22%	29.14%	2.61%		19.57%
Disputed	36.02%	44.07%	5.32%	4.25%			10.34%
In Process	39.91%					29.73%	30.36%
On Hold	27.40%	2.79%	19.40%	13.22%	3.25%	11.28%	22.66%
Resolved	17.12%		22.91%	26.45%		13.58%	19.94%
Shipped	40.19%	11.83%	9.54%	6.40%	2.39%	11.40%	18.26%

- Most people deal parts of classic car as it has the higher %.
- Parts of motorcycles are the most Disputed category.
- Least people deal with Train parts.

Pivot Table

## Customer name -Sales

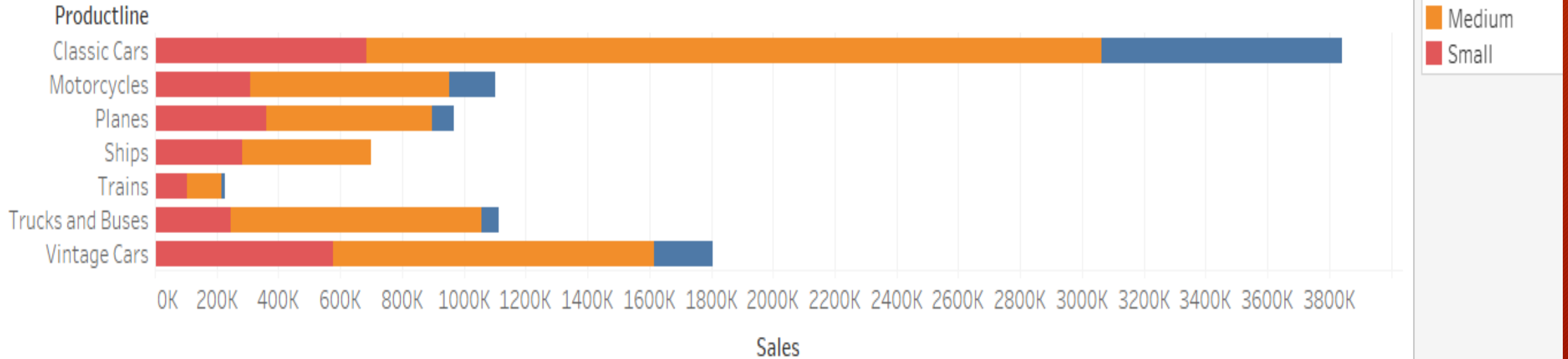


- Euro Shopping Channel has Highest sales.

Sales &  
Customer names



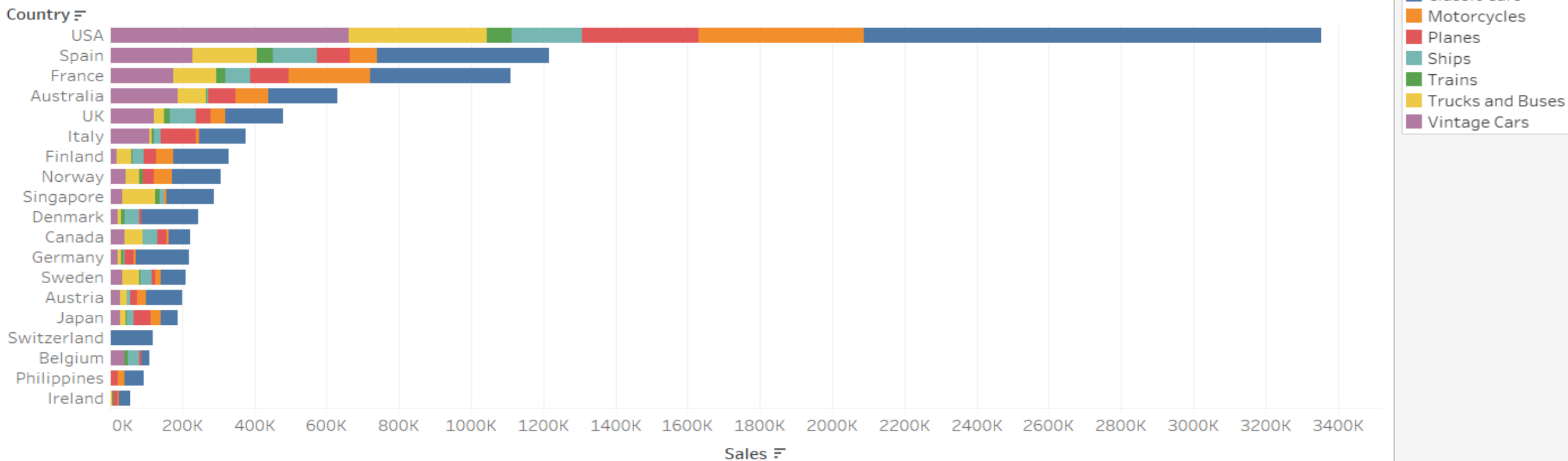
## Production line-Deal Size



- Most sales: Classic cars
- Least sales: Trains
- Most Deal size: medium
- Least Deal size: large
- Sales of truck and buses and motorcycle are approx. same

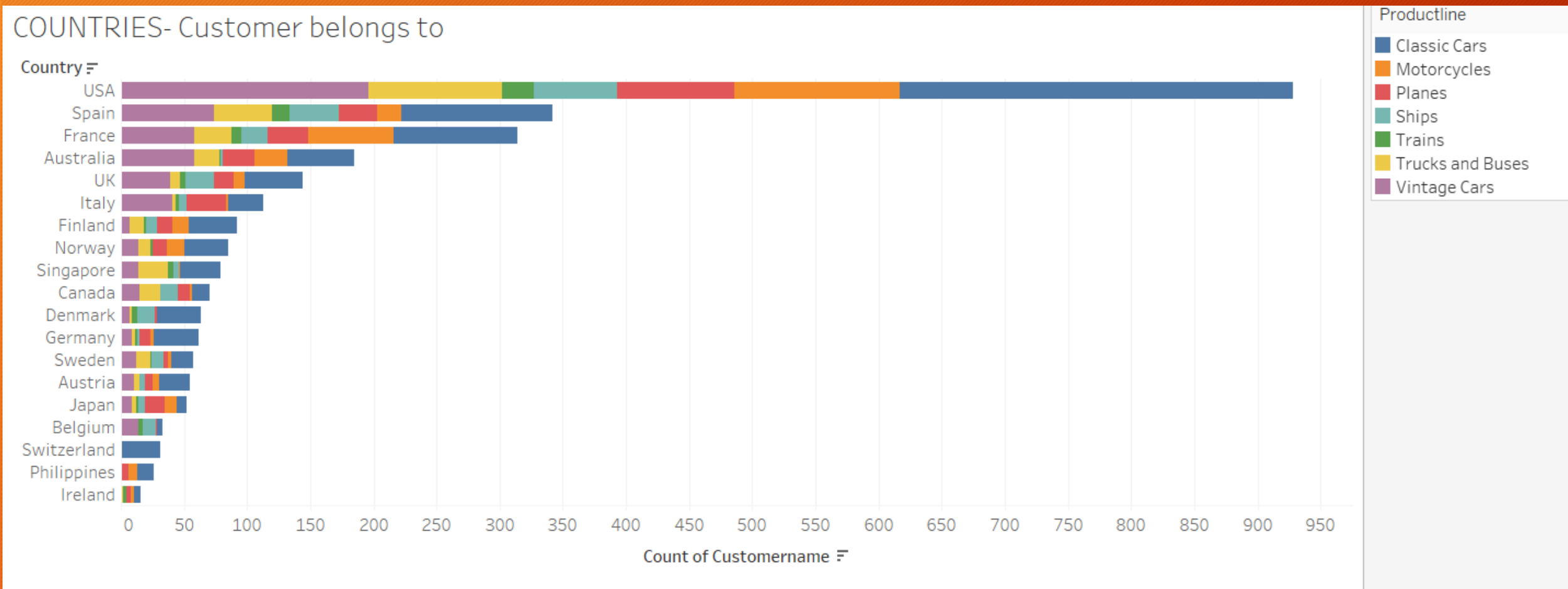
Sales & Deal  
Size

## Product Line -Country sales



- The USA dominates sales, particularly in the "Classic Cars" category, while other countries show varying sales performance across different product lines.
- European markets like Spain and France have strong sales, though lower than the US, and several countries show considerably lower overall sales.

## Product Line Sales Insights



The USA has the highest count of customers across most product lines, especially Classic Cars, while other countries have fewer customers and a different distribution of product line interest.

Customer & their country

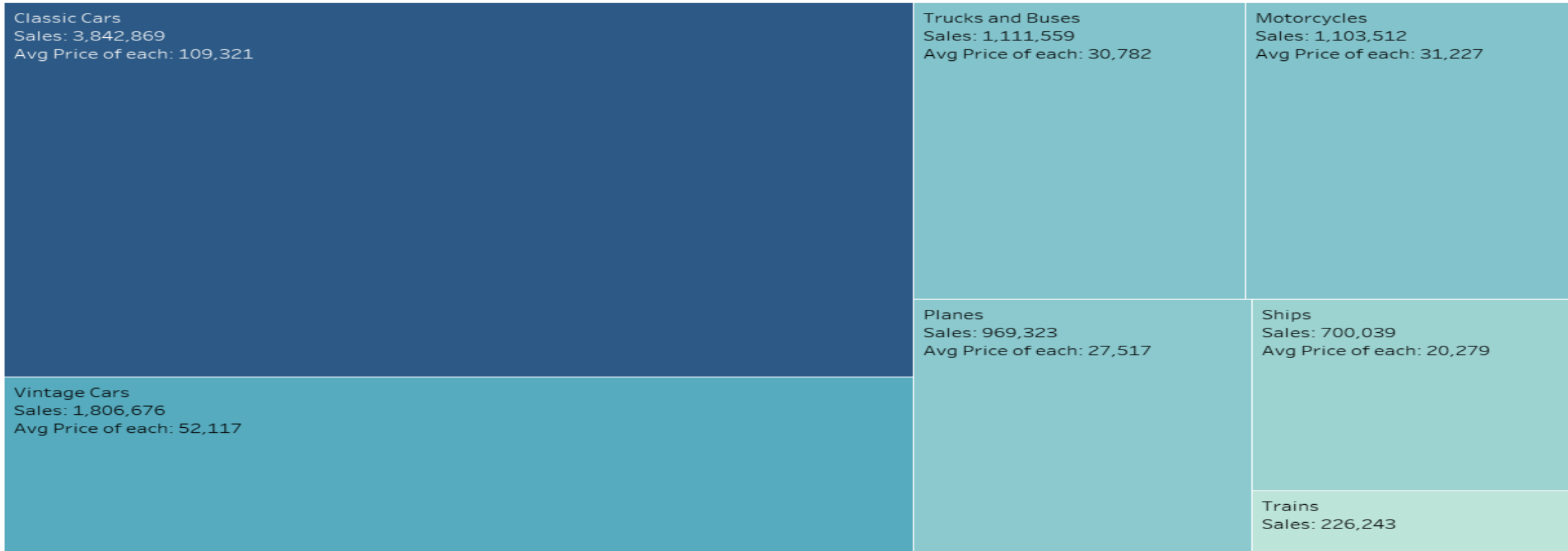


Dealsize	Status					
	Cancelled	Disputed	In Process	On Hold	Resolved	Shipped
Large		42,747	27,152	42,941	8,885	1,137,231
Medium	137,575	22,165	83,787	106,342	103,967	5,477,396
Small	56,912	7,301	33,791	29,697	37,867	2,404,467

- No order of large deal type is cancelled it's a good sign.
- Most shipped parts belong to medium deal size type.

Status, Sales  
and Deal Size

## Prodcution Line data



Classic Cars generate the highest sales and have the highest average price, while Trains have the lowest sales.

PRODUCT  
LINE AND  
SALES



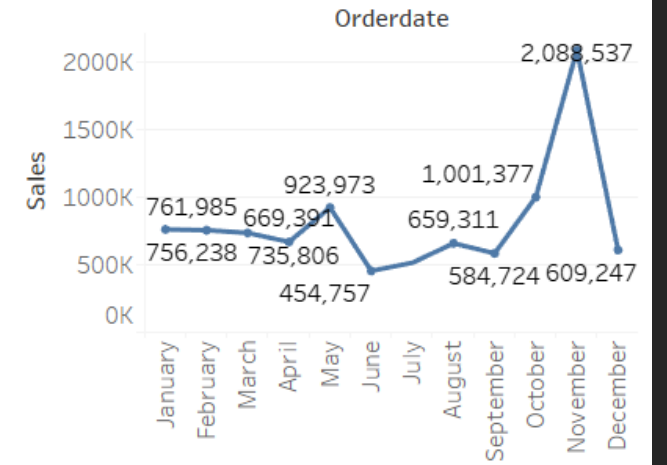
# SALES-DASHBOARD



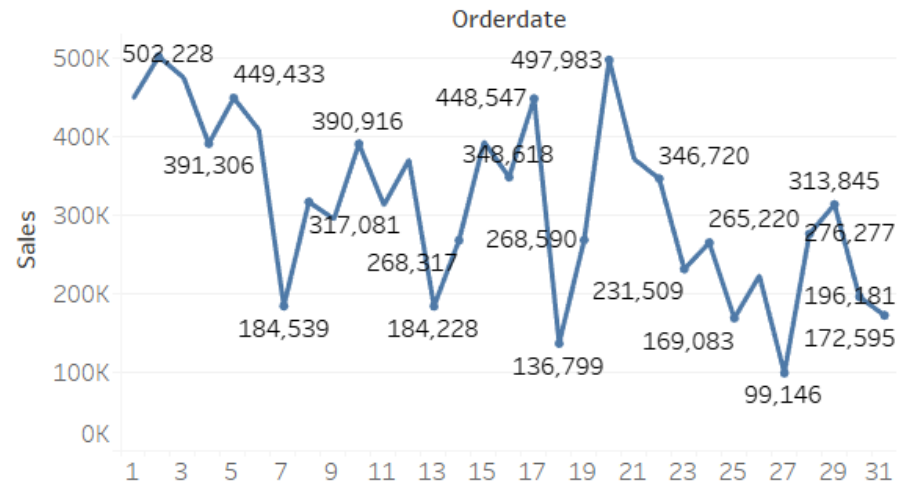
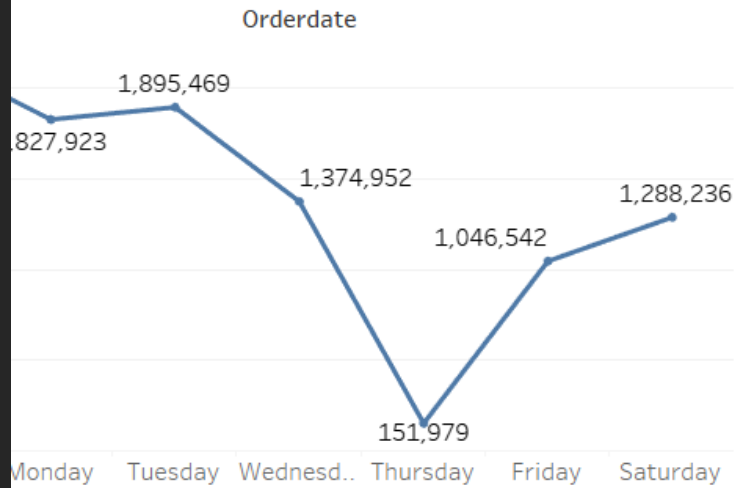
### Quarter Sales



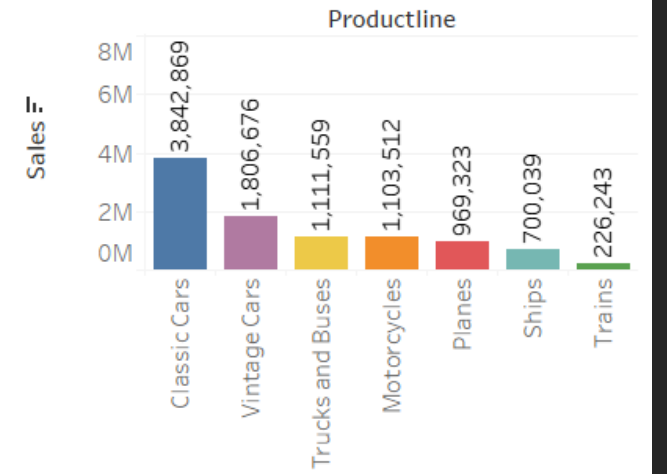
### Monthly Sales



### Day Sales



### Production line Sales





# INFERENCES

The yearly sales have dipped, which is not a good sign.

Quarter 4 has higher sales compared to other quarters.

Sales are consistent in the first 4 months of the year.

Thursday has the lowest sales, and Sunday has the highest sales.

Sales increase from Friday to Sunday and dip from Monday to Thursday.

In the starting days of the month, there are more sales than the end days.

Most orders on hold belong to the USA, and some belong to Sweden.

Spain, USA, UK, and Sweden have approximately the same number of canceled orders.

Spain has the most disputes, and most disputes are solved.

Classic car parts have the highest percentage of sales.

# RECOMMENDATIONS

Further investigation is necessary to identify the reasons for the dip in yearly sales.

To capitalize on the high sales in quarter 4, businesses should focus on increasing their inventory during this period.

Businesses should identify the reasons behind the low sales in the 6th month and work towards addressing them.

Businesses should focus their marketing efforts on Thursdays to improve sales on this day.

Strategies such as weekend sales can be employed to increase sales from Friday to Sunday.

Businesses should consider offering discounts or promotions during the start of the month to increase sales.

Efforts should be made to resolve the orders on hold to prevent revenue loss.

Businesses should monitor canceled orders closely and identify any trends to address them.

Steps should be taken to resolve disputes quickly and efficiently to maintain customer satisfaction.

Businesses should consider expanding their inventory of classic car parts, given their high sales percentage.

# Customer Segmentation using RFM analysis

- What is RFM and which tool used
- What all parameters used and assumptions made
- Output table head
- Workflow image to be put when KNIME used



# What is RFM ?

## RFM (Recency, Frequency, Monetary) Analysis

RFM is a marketing technique used to identify the most valuable customers based on their purchasing behavior.

It scores customers in three key areas:

- **Recency** - How recently a customer made a purchase
- **Frequency** - How often they make purchases
- **Monetary** - How much they spend in total

## RFM Metrics



### RECENCY

The frequency of the customer transactions 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



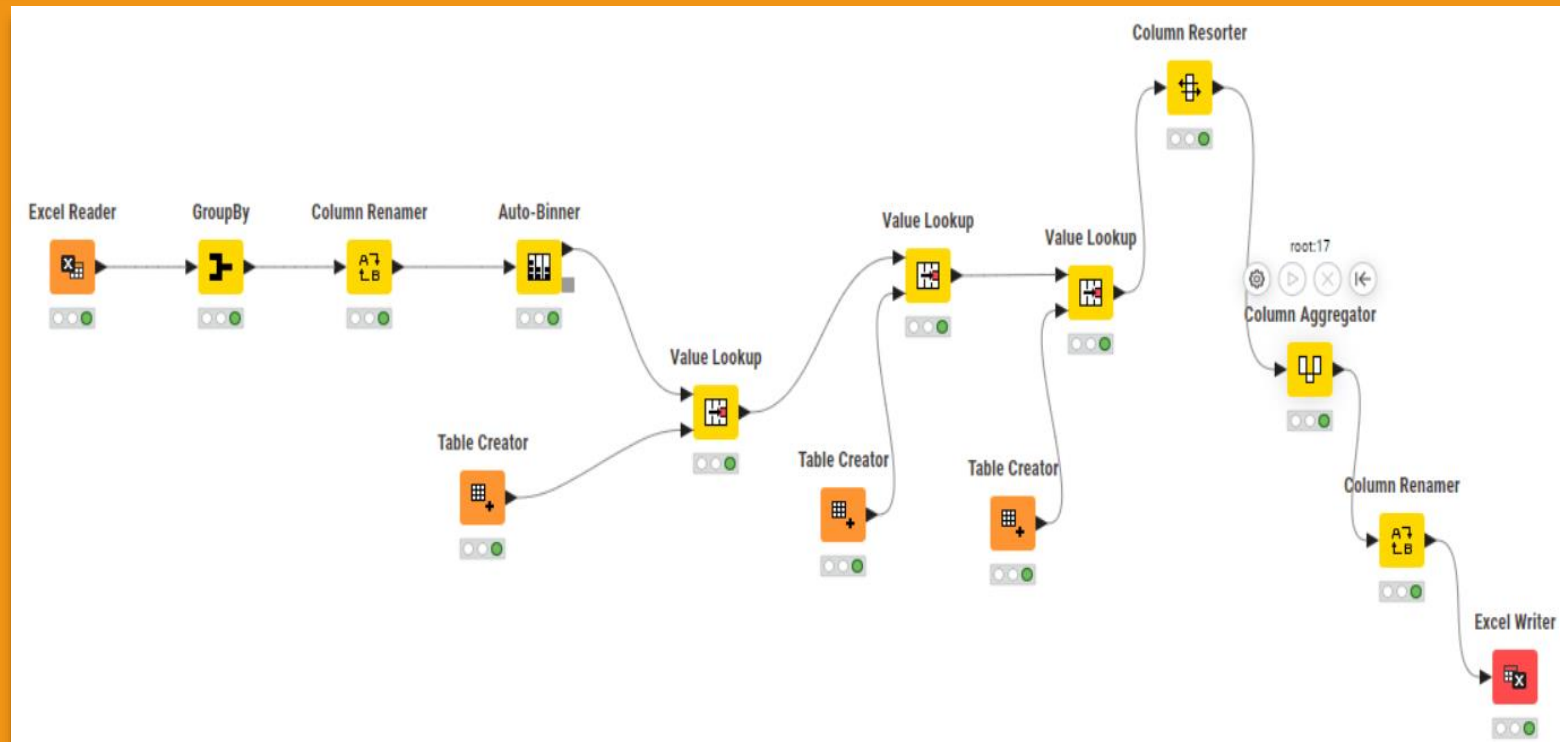
# Tool used : KNIME

- KNIME, the Konstanz Information Miner, is a free and open-source data analytics, reporting and integration platform.



# WORKFLOWS

This workflow calculates RFM (Recency, Frequency, Monetary) values to analyze customer behavior. It reads transaction data, groups it by customer to determine Recency (using the most recent order date), Frequency (using order counts), and Monetary value (using total sales). These RFM values are then binned, scored, and combined to produce a final RFM score for each customer, enabling customer segmentation.



# Few rows of output

CUSTOMERNAME	Recency	Frequency	Monetary	Recency[Binned]	Frequency[Binned]	Monetary[Binned]	R	F	M	RFMValue
AV Stores, Co.	421	51	157807.81	Bin 3	Bin 4	Bin 4	3	4	4	344
Alpha Cognac	675	20	70488.44	Bin 4	Bin 1	Bin 1	2	1	1	211
Amica Models & Co.	328	26	94117.26	Bin 2	Bin 2	Bin 3	4	2	3	423
Anna's Decorations, Ltd	131	46	153996.13	Bin 1	Bin 4	Bin 4	5	4	4	544
Atelier graphique	312	7	24179.96	Bin 2	Bin 1	Bin 1	4	1	1	411
Australian Collectables, Ltd	1018	23	64591.46	Bin 4	Bin 2	Bin 1	2	2	1	221
Australian Collectors, Co.	229	55	200995.41	Bin 1	Bin 4	Bin 4	5	4	4	544
Australian Gift Network, Co	190	15	59469.12	Bin 1	Bin 1	Bin 1	5	1	1	511
Auto Assoc. & Cie.	275	18	64834.32	Bin 2	Bin 1	Bin 1	4	1	1	411
Auto Canal Petit	127	27	93170.66	Bin 1	Bin 3	Bin 3	5	3	3	533
Auto-Moto Classics Inc.	1353	8	26479.26	Bin 4	Bin 1	Bin 1	2	1	1	211
Baane Mini Imports	245	32	116599.19	Bin 1	Bin 3	Bin 3	5	3	3	533
Bavarian Collectables Imports, Co.	801	14	34993.92	Bin 4	Bin 1	Bin 1	2	1	1	211
Blauer See Auto, Co.	705	22	85171.59	Bin 4	Bin 2	Bin 2	2	2	2	222
Boards & Toys Co.	410	3	9129.35	Bin 2	Bin 1	Bin 1	4	1	1	411
CAF Imports	625	13	49642.05	Bin 3	Bin 1	Bin 1	3	1	1	311
Cambridge Collectables Co.	484	11	36163.62	Bin 3	Bin 1	Bin 1	3	1	1	311
Canadian Gift Exchange Network	364	22	75238.92	Bin 2	Bin 2	Bin 2	4	2	2	422
Classic Gift Ideas, Inc	344	21	67506.97	Bin 2	Bin 2	Bin 1	4	2	1	421
Classic Legends Inc.	309	20	77795.2	Bin 2	Bin 1	Bin 2	4	1	2	412



# 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?



# Top 5 best customers

CUSTOMERNAME	Recency	Frequency	Monetary	Recency[Binned]	Frequency[Binned]	Monetary[Binned]	R	F	M	RFM Value
Anna's Decorations, Ltd	131	46	153996.13	Bin 1	Bin 4	Bin 4	5	4	4	544
Australian Collectors, Co.	229	55	200995.41	Bin 1	Bin 4	Bin 4	5	4	4	544
Euro Shopping Channel	42	259	912294.11	Bin 1	Bin 4	Bin 4	5	4	4	544
La Rochelle Gifts	139	53	180124.9	Bin 1	Bin 4	Bin 4	5	4	4	544
Land of Toys Inc.	216	49	164069.44	Bin 1	Bin 4	Bin 4	5	4	4	544

- According to RFM score we have grouped the top customers.
- We have given importance to recency more and ordered the customers accordingly.

# Top loyal customers

CUSTOMERNAME	Recency	Frequency	Monetary	Recency[Binned]	Frequency[Binned]	Monetary[Binned]	R	F	M	RFM Value
Euro Shopping Channel	42	259	912294.11	Bin 1	Bin 4	Bin 4	5	4	4	544
Mini Gifts Distributors Ltd.	219	180	654858.06	Bin 1	Bin 4	Bin 4	5	4	4	544
Australian Collectors, Co.	229	55	200995.41	Bin 1	Bin 4	Bin 4	5	4	4	544
La Rochelle Gifts	139	53	180124.9	Bin 1	Bin 4	Bin 4	5	4	4	544
Land of Toys Inc.	216	49	164069.44	Bin 1	Bin 4	Bin 4	5	4	4	544

- Based on RFM analysis these are the loyal customers
- We have focused on monetary value
- If we focus on these customer's we can turn them into best customers.

# Verge on churning Customers

CUSTOMERNAME	Recency	Frequency	Monetary	Recency[Binned]	Frequency[Binned]	Monetary[Binned]	R	F	M	RFM Value
Auto-Moto Classics Inc.	1353	8	26479.26	Bin 4	Bin 1	Bin 1	2	1	1	211
Rovelli Gifts	1032	48	137955.72	Bin 4	Bin 4	Bin 4	2	4	4	244
Australian Collectables, Ltd	1018	23	64591.46	Bin 4	Bin 2	Bin 1	2	2	1	221
Cruz & Sons Co.	971	26	94015.73	Bin 4	Bin 2	Bin 3	2	2	3	223
Gift Ideas Corp.	947	19	57294.42	Bin 4	Bin 1	Bin 1	2	1	1	211

- As per RFM score we can see that these are the top customers on the verge of churning.
- We should focus on these customers before we lose them.
- We should try some action plan to convert them into regular customers.



# Top Lost Customers

CUSTOMERNAME	Recency	Frequency	Monetary	Recency[Binned]	Frequency[Binned]	Monetary[Binned]	R	F	M	RFMValue
Alpha Cognac	675	20	70488.44	Bin 4	Bin 1	Bin 1	2	1	1	211
Auto-Moto Classics Inc.	1353	8	26479.26	Bin 4	Bin 1	Bin 1	2	1	1	211
Bavarian Collectables Imports, Co.	801	14	34993.92	Bin 4	Bin 1	Bin 1	2	1	1	211
Clover Collections, Co.	659	16	57756.43	Bin 4	Bin 1	Bin 1	2	1	1	211
Double Decker Gift Stores, Ltd	670	12	36019.04	Bin 4	Bin 1	Bin 1	2	1	1	211

- As per RFM score we can see that these are the customers which we have lost.
- Their recency is very low as well as they have not made purchases frequently.
- We should study them, survey them to understand the reasons why we lost them. And take further steps so that we do not lose the customers.



# RECOMMENDATIONS

# Recommendations

Based on the RFM analysis, customers can be grouped into four categories: **Best**, **Loyal**, **At Risk (Verge of Churn)**, and **Lost**. Each group requires a tailored strategy to enhance engagement and improve retention.

- **Best Customers:**  
Offer personalized recognition, exclusive deals, and loyalty rewards to maintain their interest and long-term relationship with the brand.
- **Loyal Customers:**  
Keep them engaged with regular offers and discounts. With the right incentives, they can be nurtured into best customers.
- **At Risk Customers:**  
These customers show signs of leaving. Use feedback surveys, targeted offers, and personalized outreach to address their issues and rebuild loyalty.
- **Lost Customers:**  
Analyze their past behavior to understand why they left. Use these insights to build strategies that prevent future churn and enhance customer satisfaction.

THANKYOU