



# Marketing and Retail Analysis (MRA) Project

By:- Harsh Patel

Date:- 27 October 2024



# **PART-A: AutoMobile Parts Manufacturing Company Sales**

## *Problem Statement:*

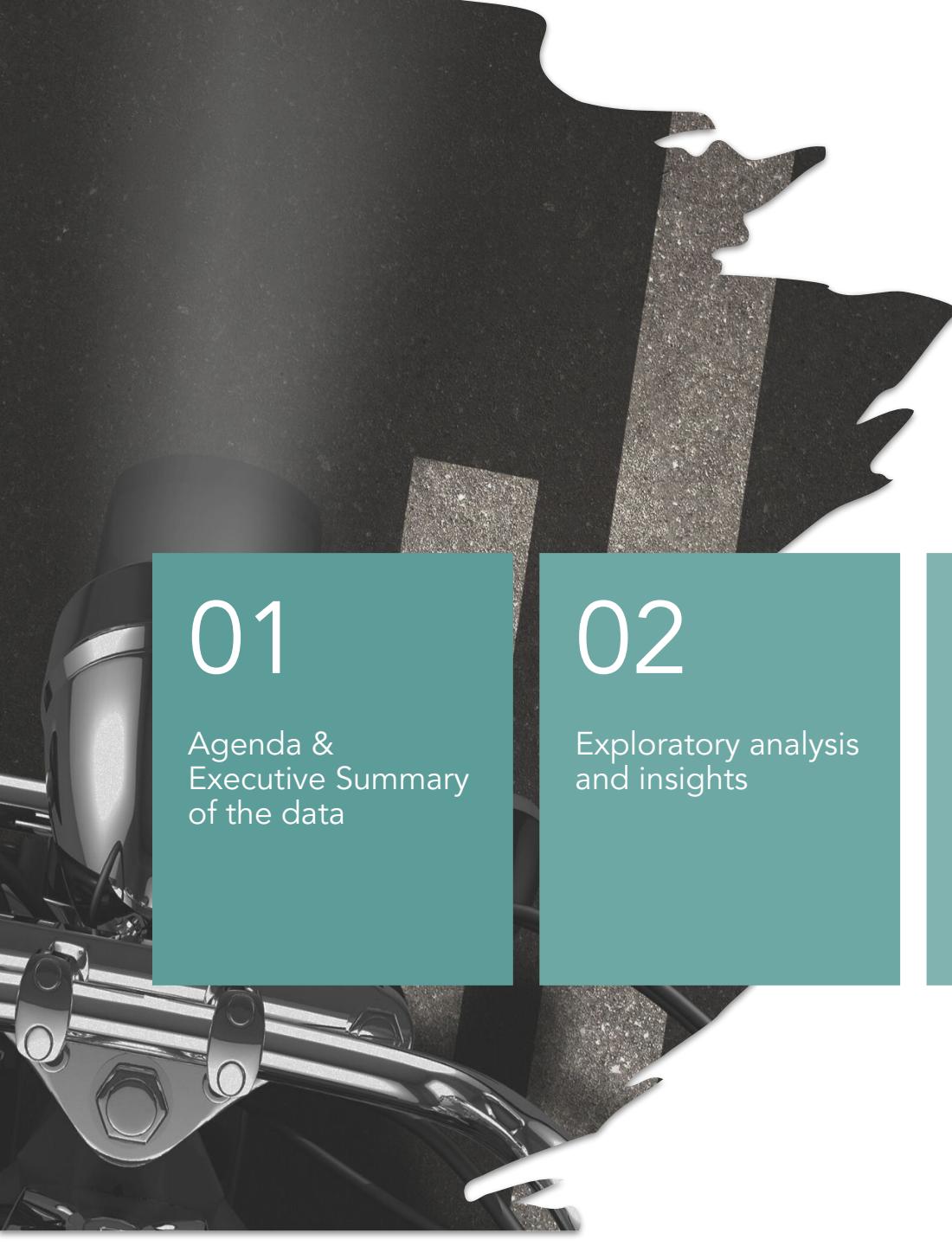
An automobile parts manufacturing company has collected data on transactions for 3 years. They do not have any in-house data science team; thus, they have hired you as their consultant. Your job is to use your data science skills to find the underlying buying patterns of the customers, provide the company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



# CONTENT

1. Agenda & Executive Summary of the data
  - Problem statement
  - Executive Summary
  - Data Dictionary
  - Summary Stats
  - Assumptions about data
2. 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
  - 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
  - Inferences from RFM Analysis and identified segments
  - Who are your best customers?
  - Which customers are on the verge of churning?
  - Who are your lost customers?
  - Who are your loyal customers?





# 1. Agenda :

---

01

Agenda &  
Executive Summary  
of the data

02

Exploratory analysis  
and insights

03

RFM analysis for  
customer  
segmentation

04

Identification of  
customers based  
on different  
parameters

05

Recommendations



## Executive Summary

**Objective:** This analysis aims to uncover the underlying buying patterns of customers, offering valuable insights into customer behavior and recommending tailored marketing strategies for various customer segments.

**Dataset Overview:** The analysis is based on 2,747 transactions from an automobile parts manufacturing company. The dataset comprises 20 columns, including 2 float, 6 integer, and 12 object data types, providing a comprehensive view of customer and transaction data. Importantly, there are no duplicate entries.

### Key Findings:

- **Order Insights:** On average, customers placed orders for 35 items per transaction, with a range from a minimum of 6 to a maximum of 97 items.
- **Customer Activity:** The average interval since the last order is approximately 1,757 days, with a minimum of 42 days and a maximum of 3,562 days.
- **Pricing Information:** The average Manufacturer's Suggested Retail Price (MSRP) of the products is \$100.69, with a standard deviation of \$40.11.

**Analytical Approach:** We employed RFM (Recency, Frequency, Monetary) analysis to evaluate customer behavior based on their transaction history. By assessing how recently a customer made a purchase, how often they buy, and the total monetary value of their transactions, RFM analysis enables the identification and segmentation of customers. This segmentation allows for the development of targeted marketing strategies and enhanced customer retention efforts.

# 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.

# Statistical Summary

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

Observation :

- **Average number of items per sales order:** 35 (Standard Deviation: 9.76)
- **Average price per item:** 101.09 (Standard Deviation: 42.04)
- **Average sales amount per order:** 3553.05 (Standard Deviation: 1838.95)
- **Average time since the last order:** 1757.09 days (Standard Deviation: 819.28)

# Assumptions:



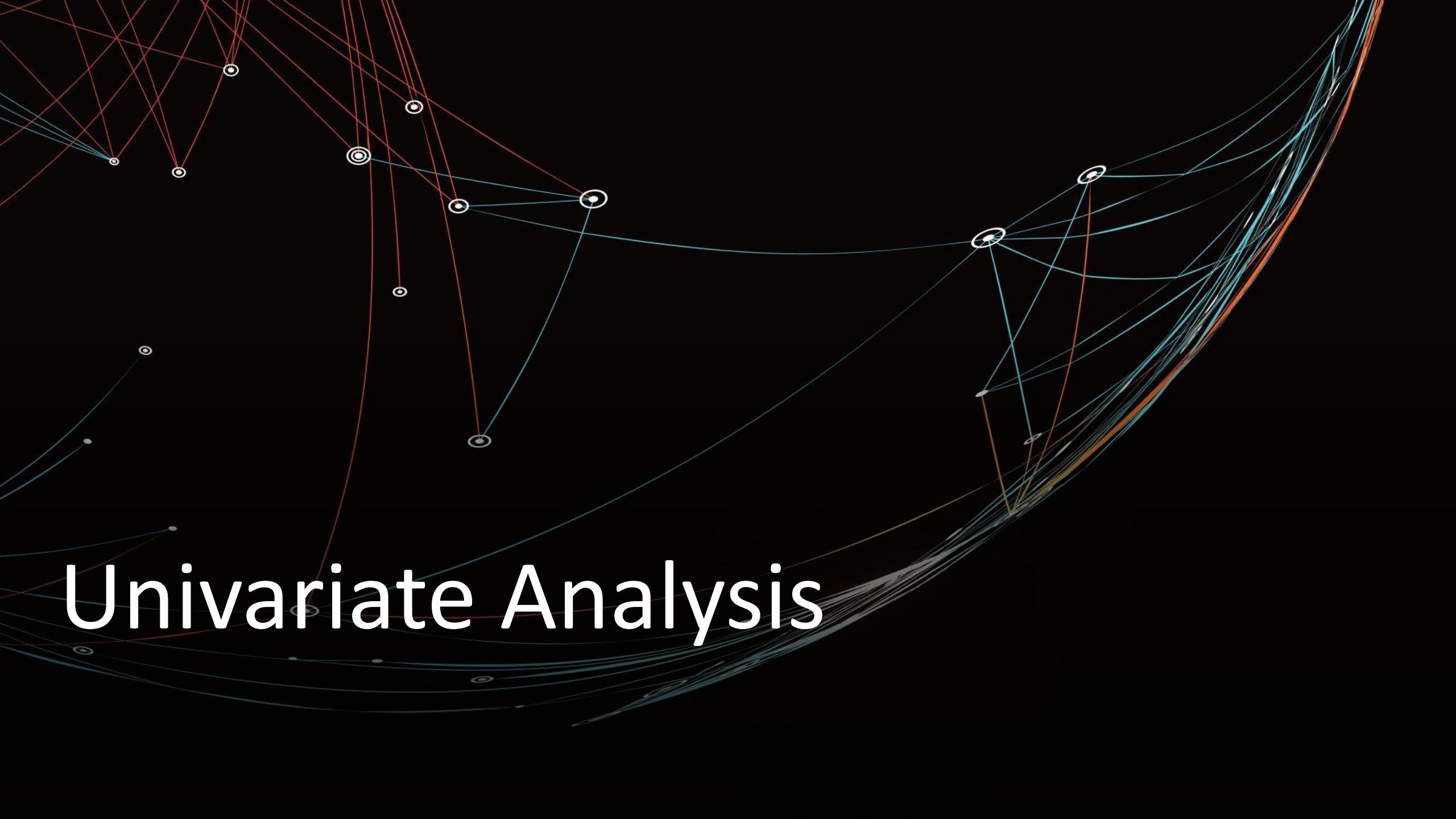
- The dataset contains detailed information about customer orders, with each order uniquely identified by an ORDERNUMBER. It includes transactional data related to customer purchases, such as the number of items bought, the days since the last order, and the manufacturer's suggested retail price (MSRP) for each product. Importantly, there are no duplicate entries, meaning each transaction record is distinct, with no repetitions for any single customer.
- Each row in the dataset represents a unique transaction made by a customer. Customer segments can be defined using RFM (Recency, Frequency, Monetary) analysis, which considers purchasing frequency, total amount spent, and recency of purchases. Marketing strategies may vary by segment, necessitating personalized marketing efforts tailored to different customer groups.
- The columns for order date and days since the last order are accurately calculated, and the sales amount is determined by multiplying the quantity ordered by the price per item. The status column reflects the current status of each order accurately. Recommendations provided in the presentation are based on insights derived from the analysis of this transactional data.

## 2. Exploratory Analysis of data

- 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

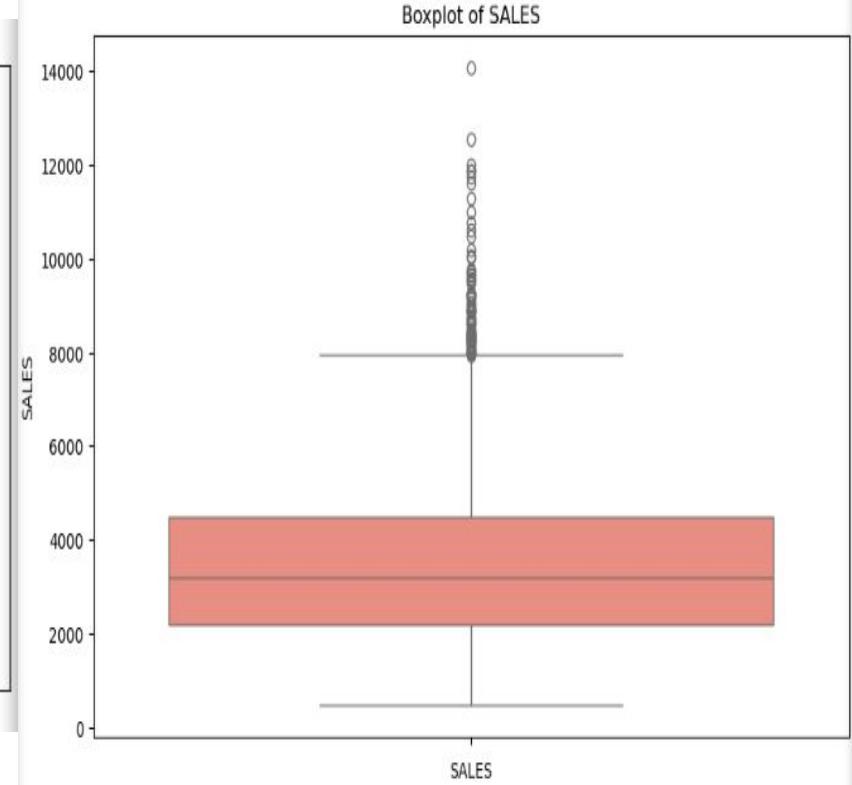
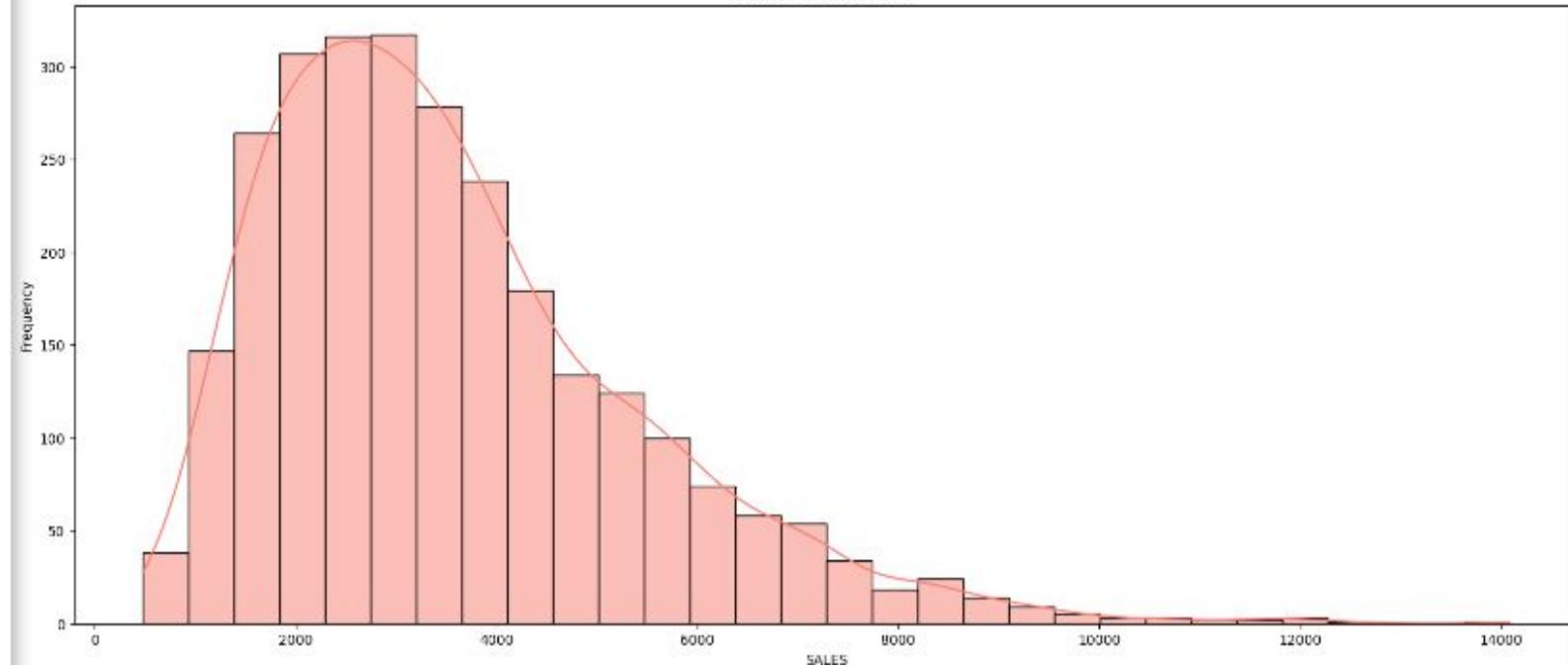


# Univariate Analysis



# Sales

Distribution of SALES



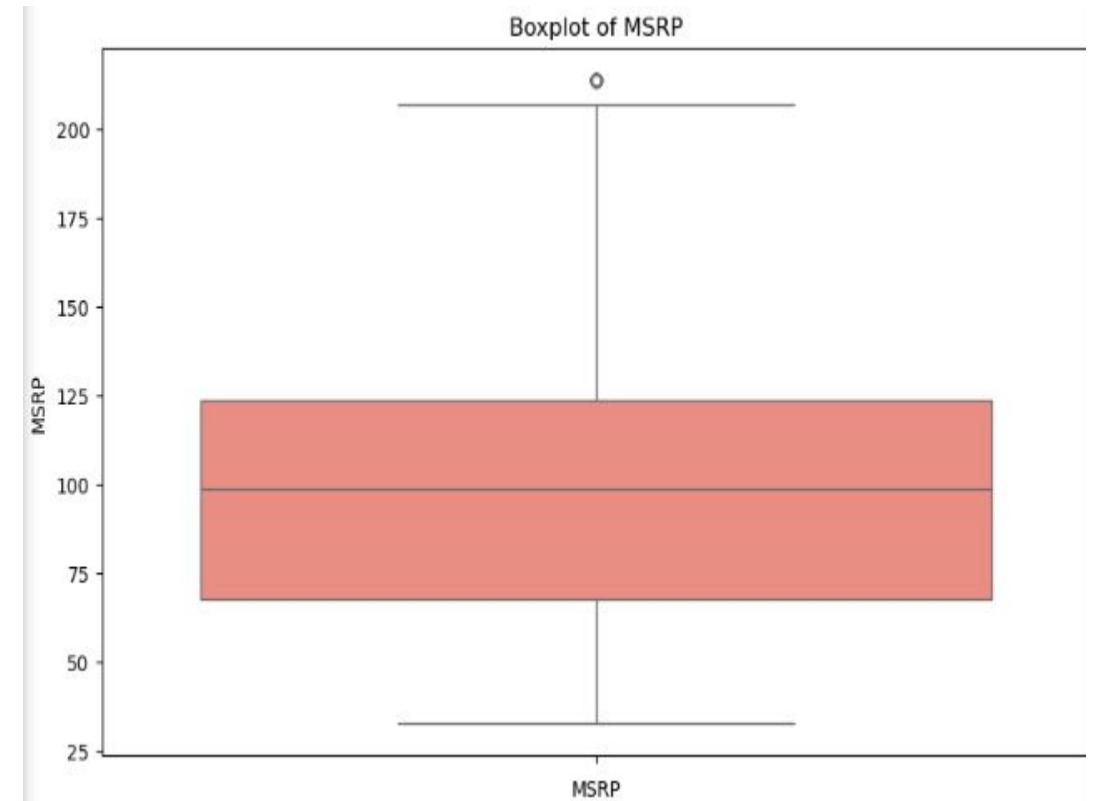
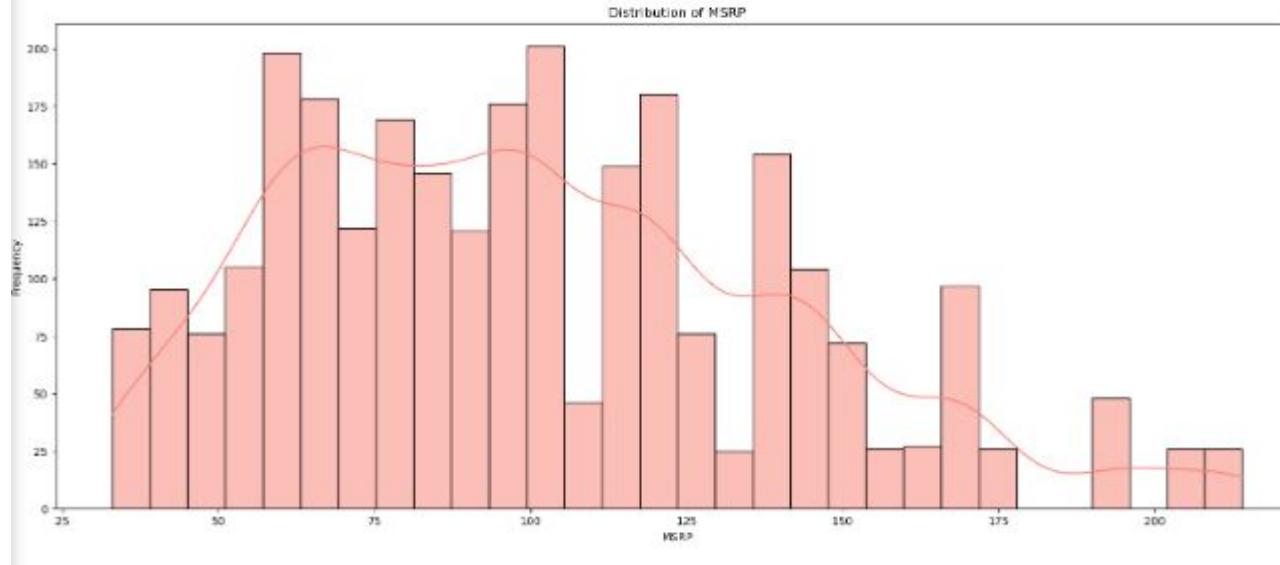
## Observation:

The sales data exhibits a left skew, indicating that most values cluster towards the higher end of the range. Additionally, there are several outliers present in the dataset. The majority of the sales figures fall within the range of 1.5k to 3.5k, suggesting a concentration of transactions around these values.

# MRSP

## Observation:

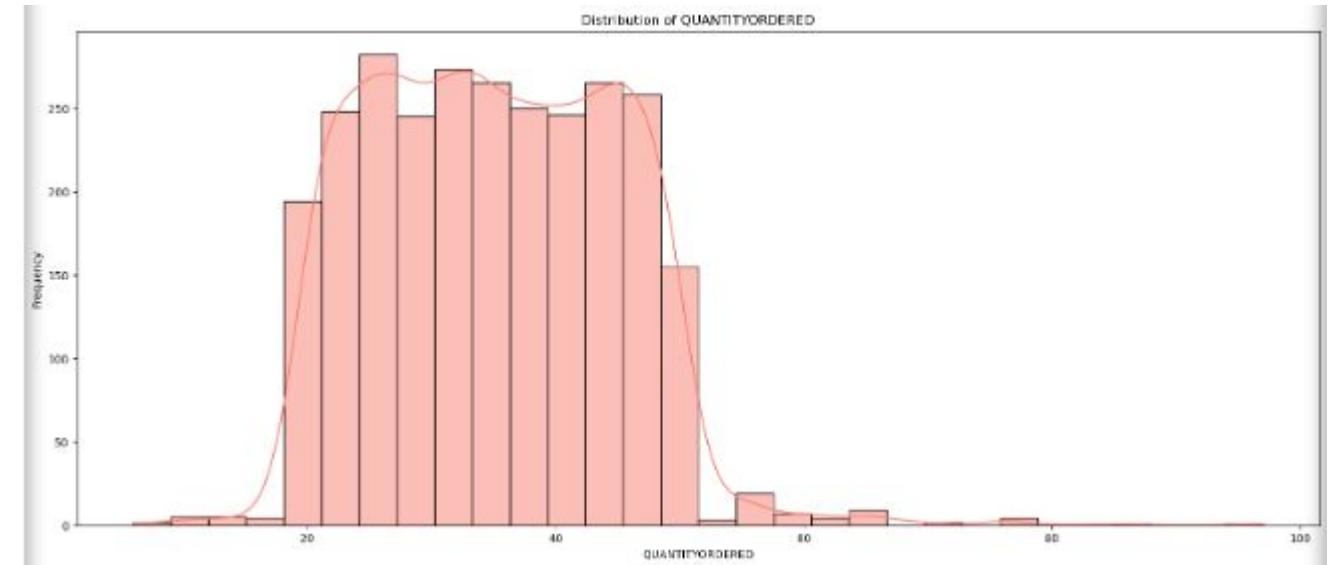
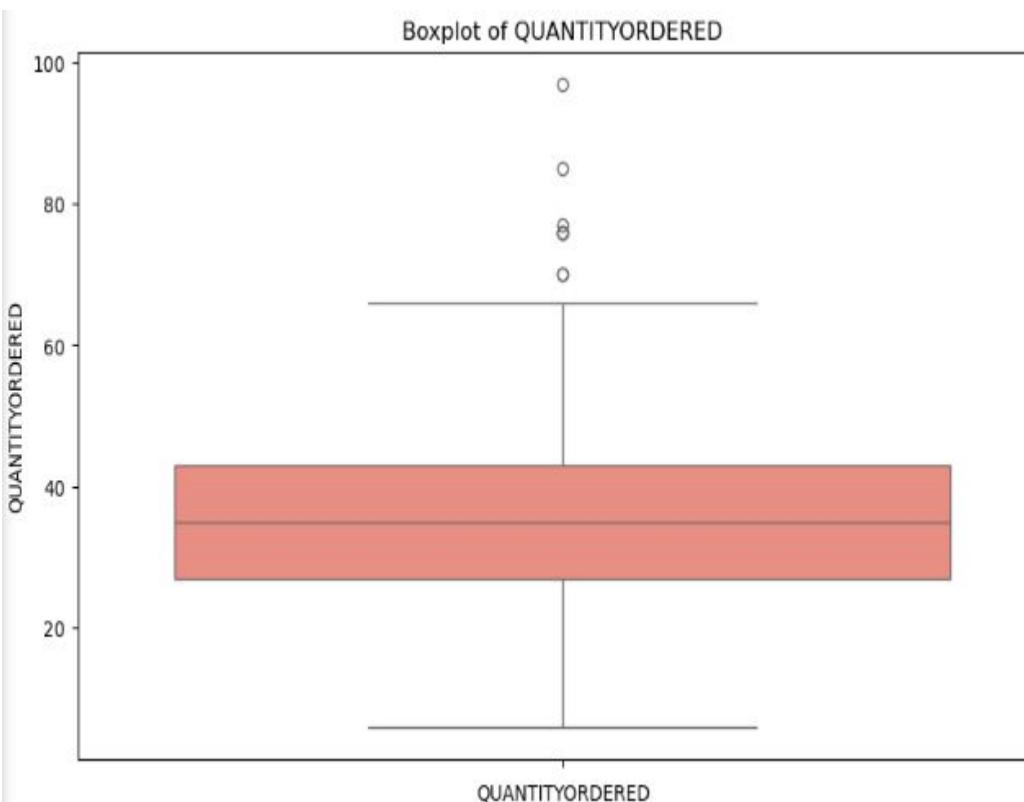
The MSRP data contains outliers and is left-skewed, indicating that most values are concentrated at the higher end of the distribution. This skewness suggests that there are relatively few low-priced items compared to a greater number of higher-priced ones.



# QUANTITY ORDERED

## Observation:

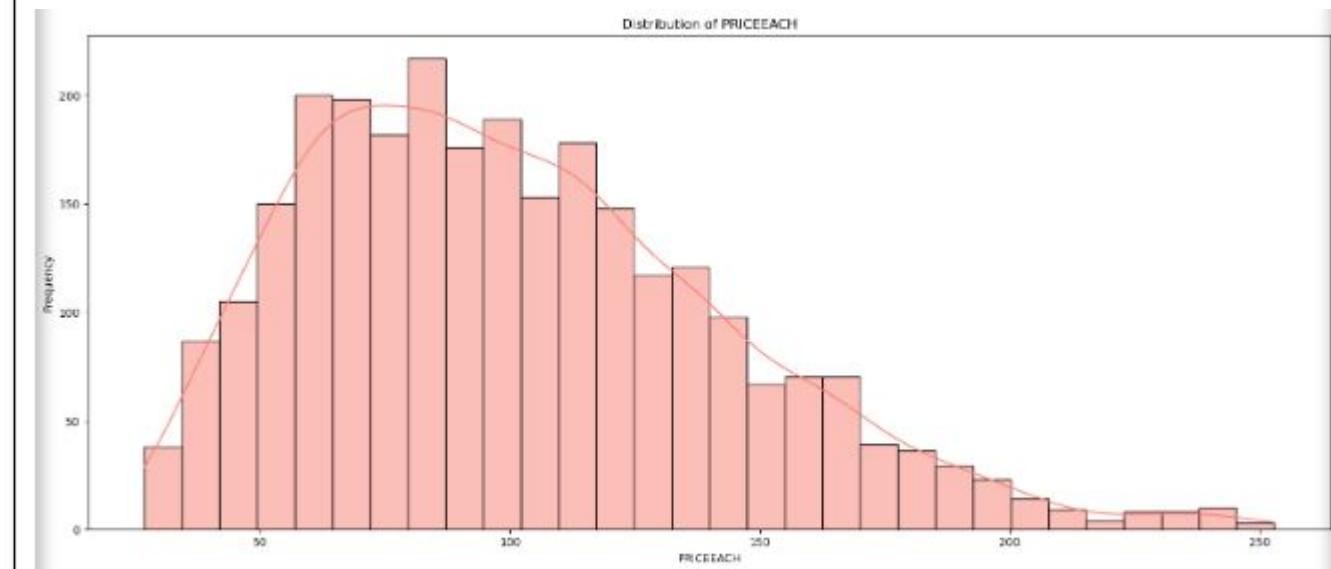
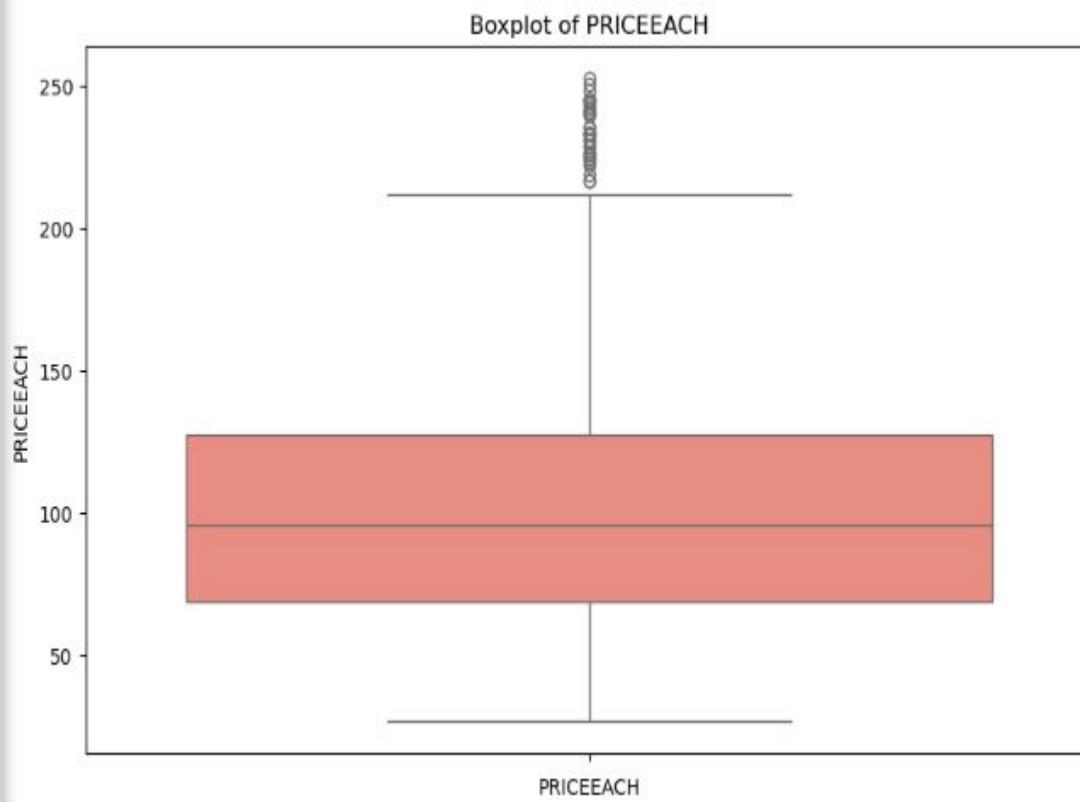
The Quantity Ordered data also contains outliers, indicating that there are some unusually high or low values that deviate significantly from the typical range of quantities purchased. These outliers may impact overall analysis and should be examined further to understand their implications.



# PRICE OF EACH

## Observation:

The Price of Each data shows the presence of monetary outliers, which are values that deviate significantly from the typical pricing range. These outliers may influence the overall analysis.

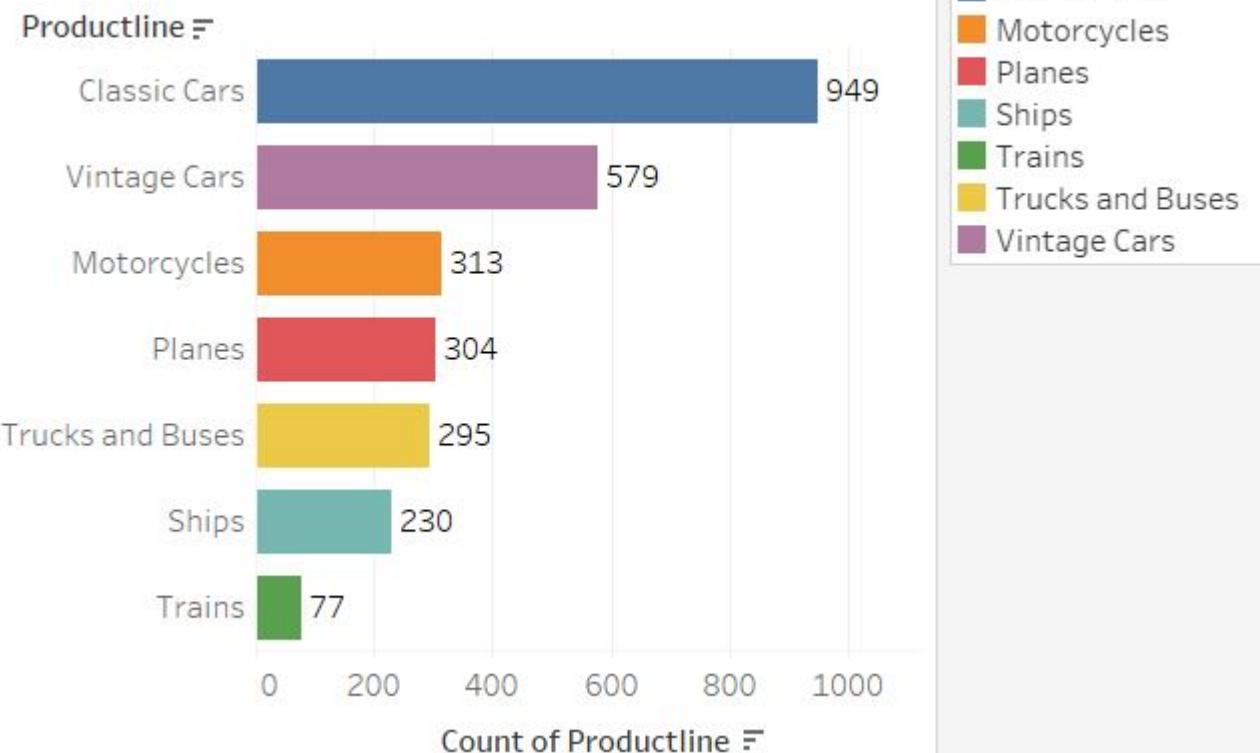


## Product Line Analysis

### Observation:

Classic cars and vintage cars are the most popular product lines, demonstrating a strong demand for these vehicle models. In contrast, ships, trucks, buses, and trains have relatively lower counts, indicating a more niche market or a specialized customer base for these categories.

### Univariate - Product Line Analysis

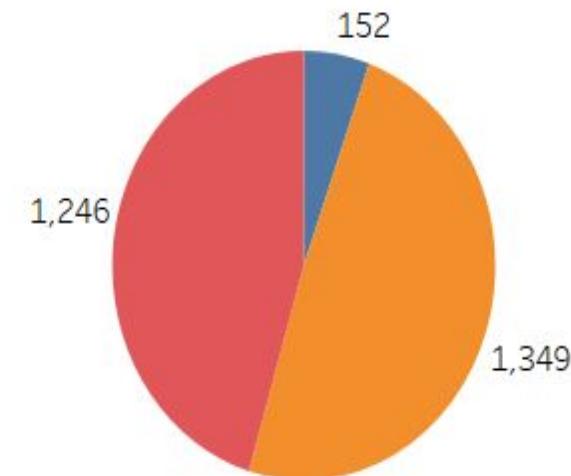


## Deal Size Analysis

### Observation :

The majority of deals are categorized as small and medium, with counts of 1,246 and 1,349, respectively. This suggests that the company primarily engages in transactions involving moderate to smaller deal sizes.

### Univariate - Deal Size Analysis

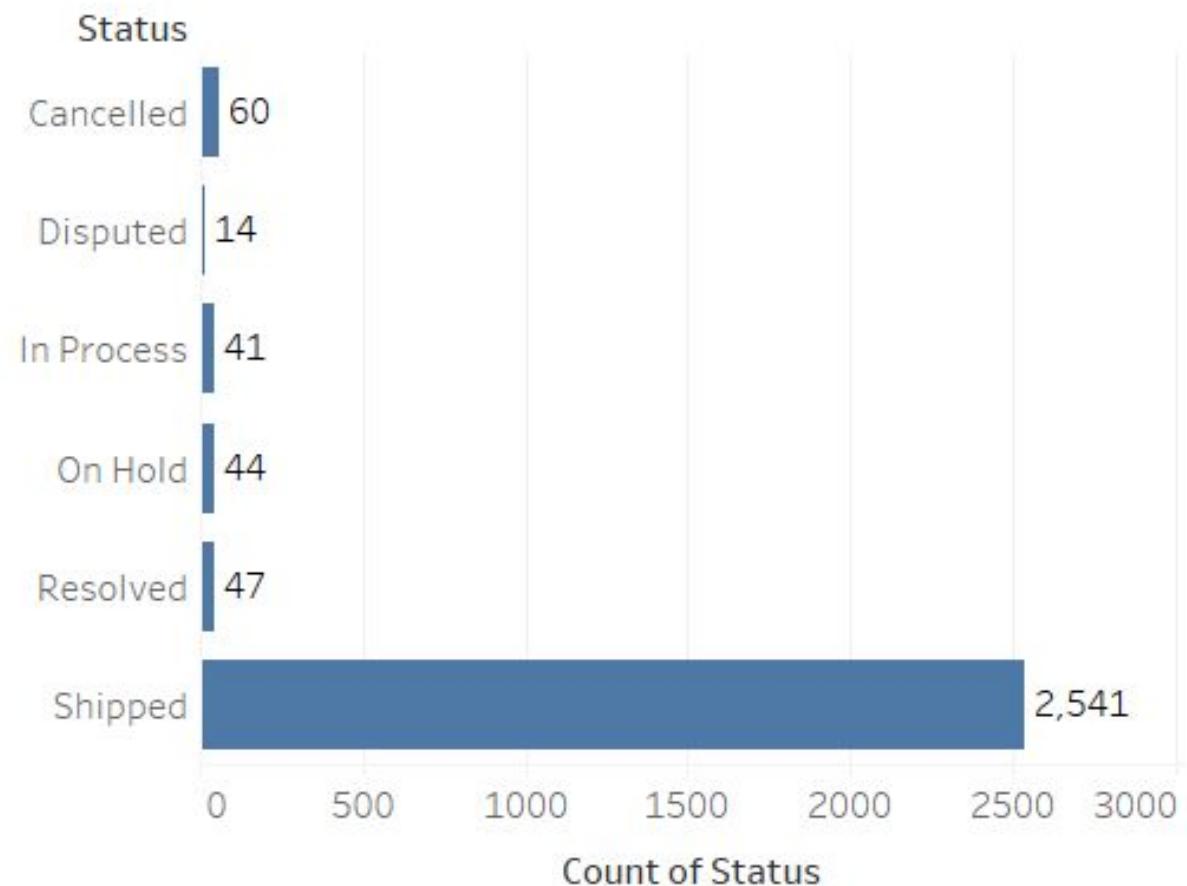


## Order Status Analysis

### Observation :

The substantial count of 2,541 successfully shipped orders highlights the company's strong order fulfillment capabilities and its commitment to delivering orders in a timely manner.

Univariate - Order Status Analysis

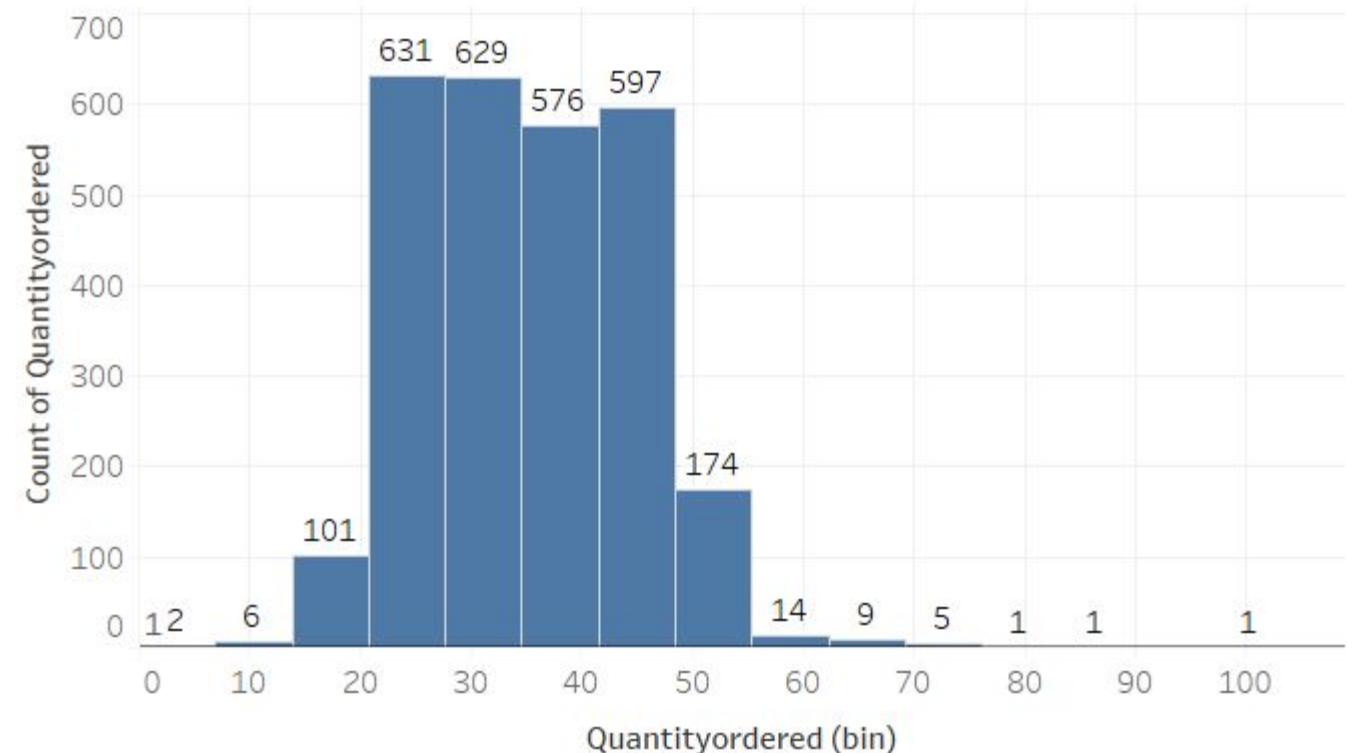


## Quantity Ordered Analysis

### Inferences :

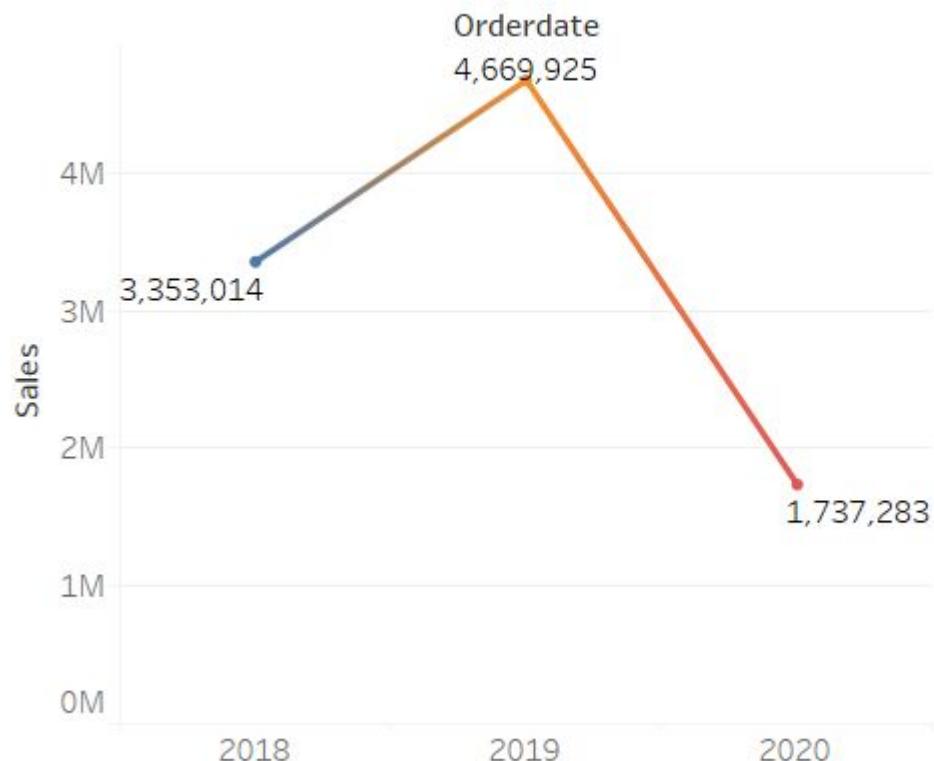
The bins for quality orders of 21, 28, 35, and 42 have the highest counts, suggesting their popularity among customers. With counts ranging from 576 to 631, these quantities effectively meet the needs of a significant portion of the customer base.

Univariate - Quantity Ordered Analysis



# Bivariate Analysis

## Bivariate - Sales over Years



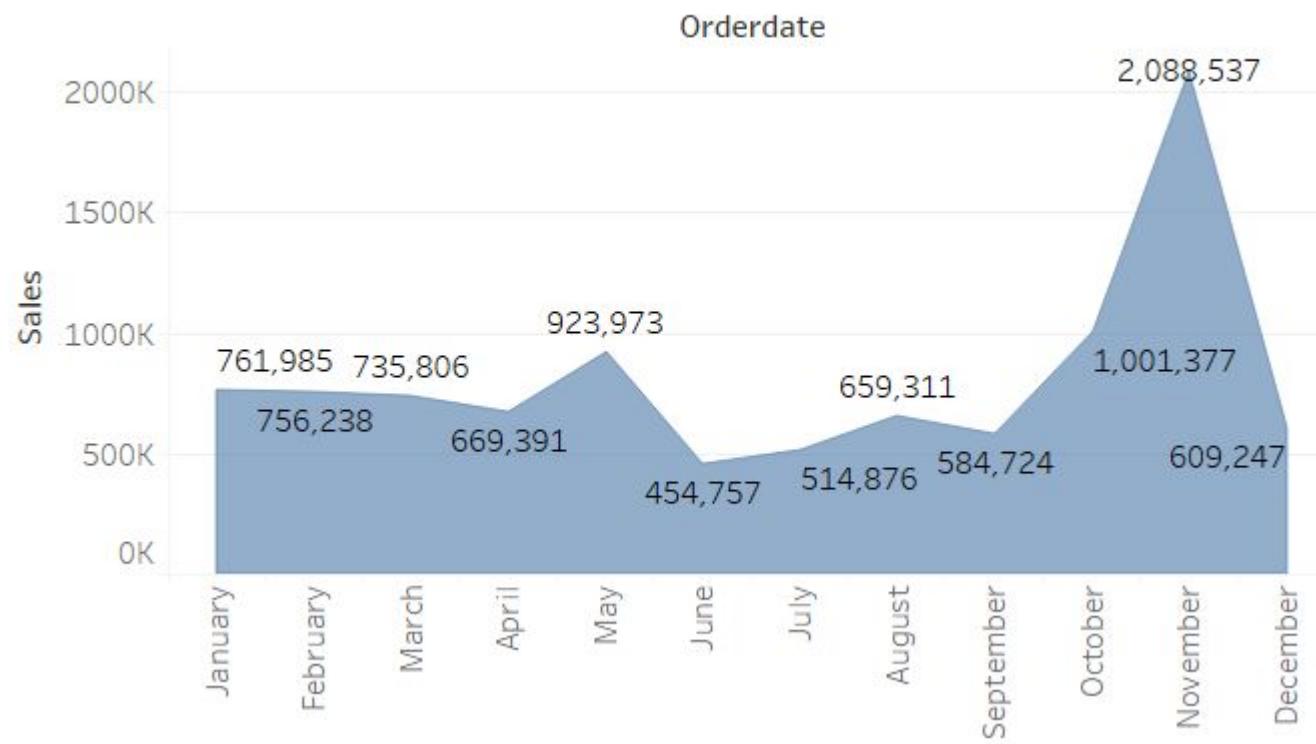
## Sales over Years

### Observation:

Sales exhibited a positive trend, showing significant growth from 2018, with total sales of 3,353,014, to 2019, reaching 4,669,925. This suggests that the business performed well and experienced an increase in customer demand.

However, since we only have sales data available up to May for 2020, with a total of 1,737,283, it is challenging to draw definitive conclusions about the overall performance for the entire year.

## Bivariate - Sales over Month



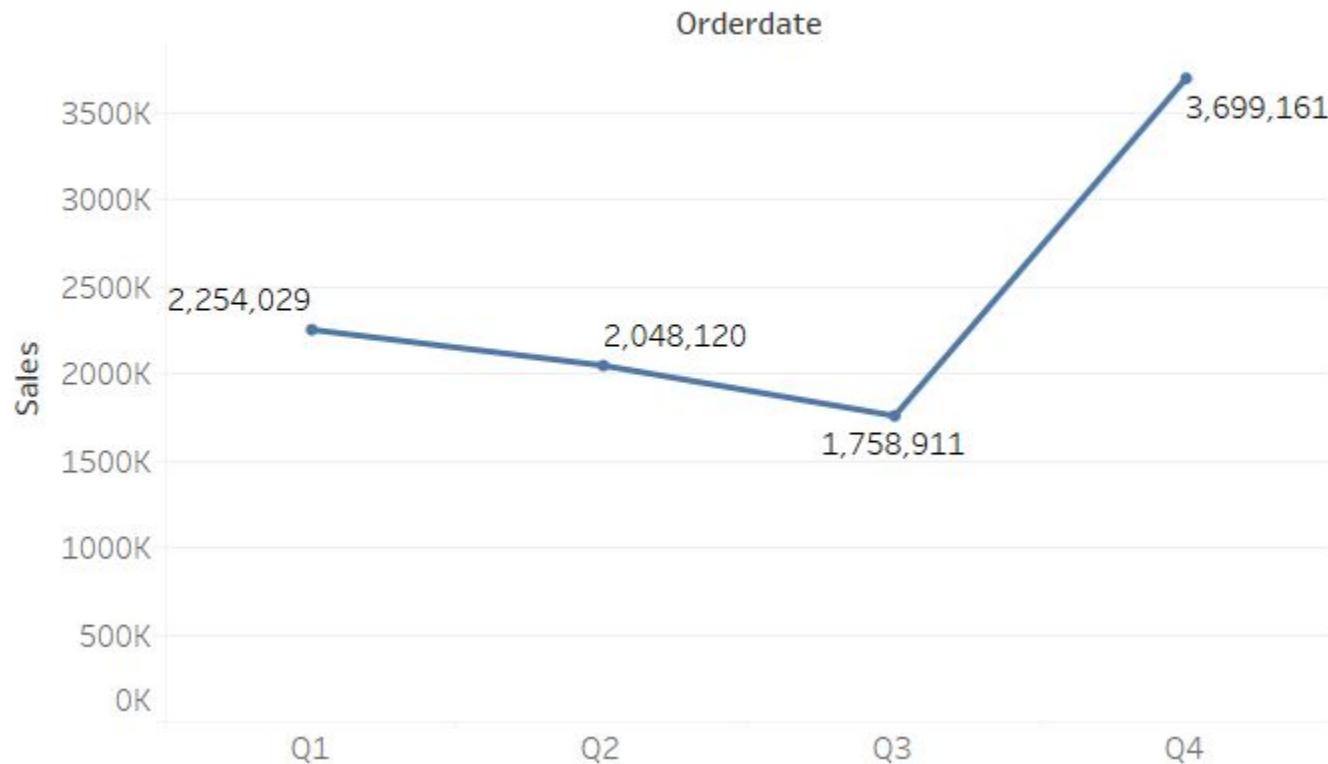
## Sales over Month

### Observation:

The sales data reveals fluctuations throughout the year, with peak sales occurring in November at 2,088,537 and in October at 1,001,377. This indicates increased customer activity, likely driven by seasonal factors.

In contrast, June shows the lowest sales figure at 454,757, suggesting a potential decline in customer demand during that period.

## Bivariate - Sales by Quarters



## Sales by Quarters

### Observation:

The quarterly sales data over the three-year period demonstrates a consistent pattern, with Q4 showing significantly higher sales at 3,699,161 compared to other quarters. This suggests a potential seasonal trend, indicating that the end of the year is marked by increased customer spending.

Q1 (2,254,029) and Q2 (2,048,120) exhibit relatively stable sales figures, while Q3 (1,758,911) records the lowest sales among the quarters.

## Bivariate - Sales by Week



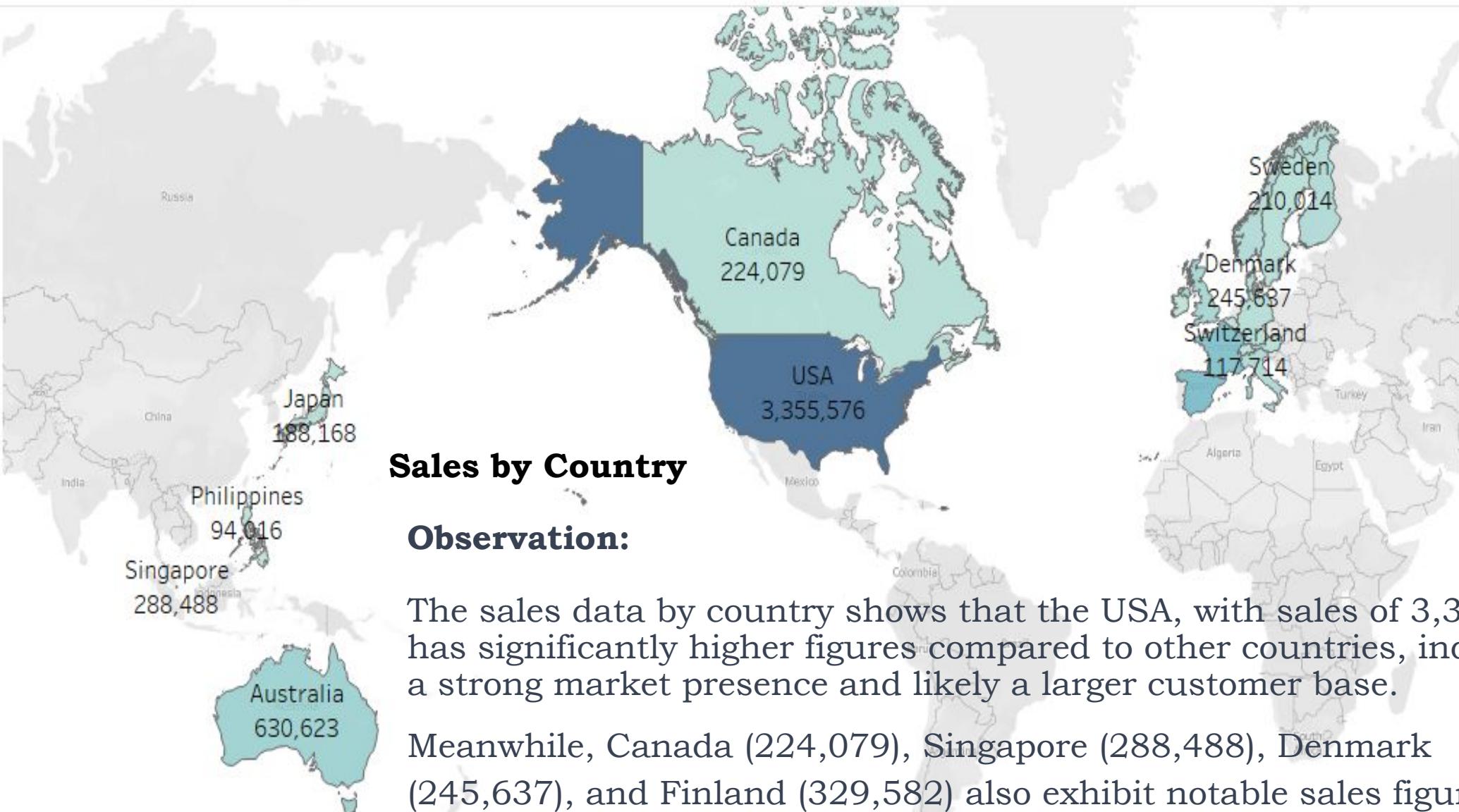
## Sales by Week

### Observation:

Sundays consistently record the highest sales at 2,175,121 over the three-year period, indicating that customers are more active and likely prefer shopping on weekends.

In contrast, Wednesdays have the lowest sales at 1,374,952 among the weekdays, suggesting that customers may be less engaged or influenced by factors that affect their buying behavior on this particular day.

## Bivariate - Sales by Country

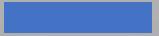


### Sales by Country

#### Observation:

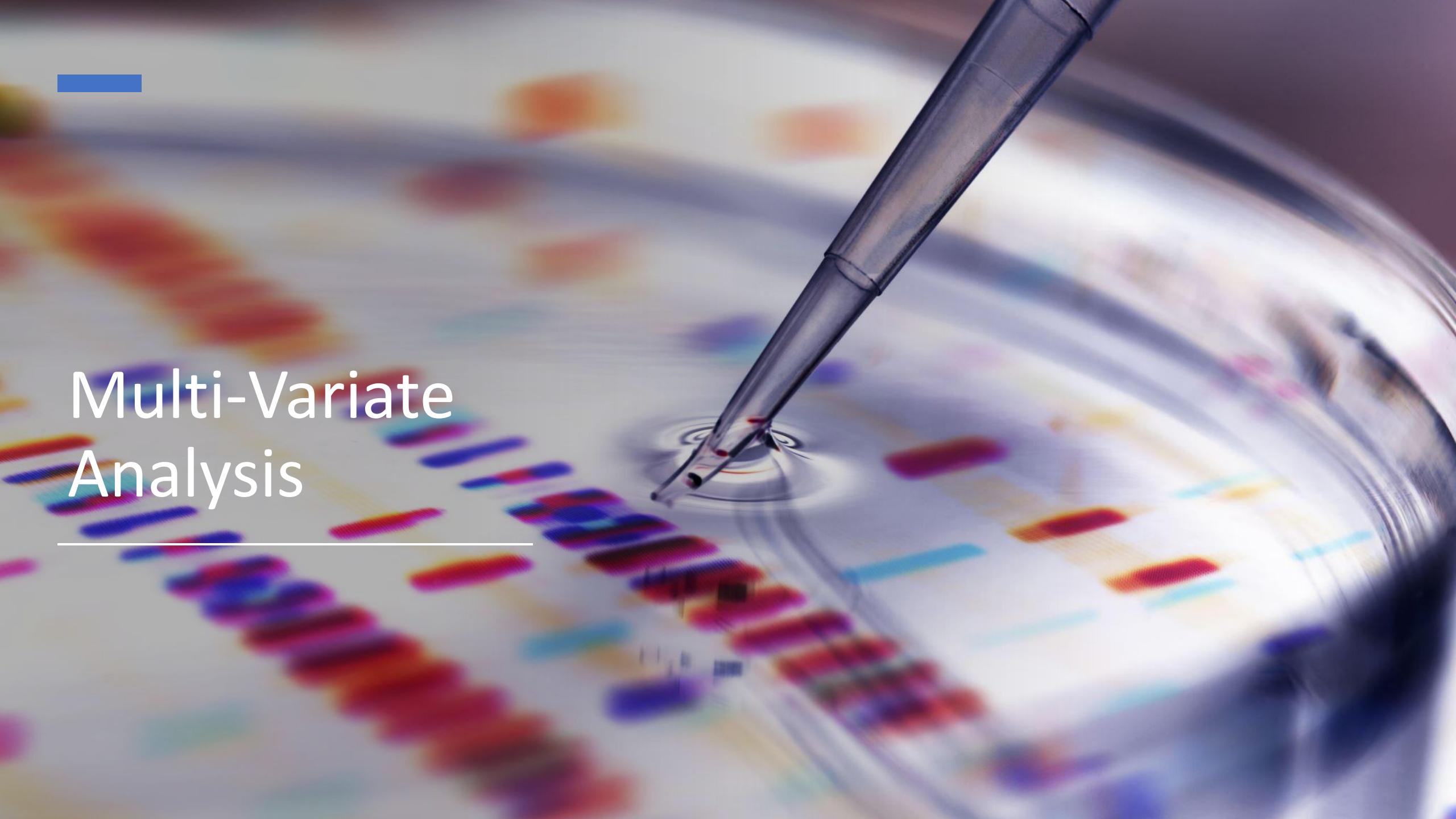
The sales data by country shows that the USA, with sales of 3,355,576, has significantly higher figures compared to other countries, indicating a strong market presence and likely a larger customer base.

Meanwhile, Canada (224,079), Singapore (288,488), Denmark (245,637), and Finland (329,582) also exhibit notable sales figures, highlighting significant market opportunities and customer demand in these respective regions.



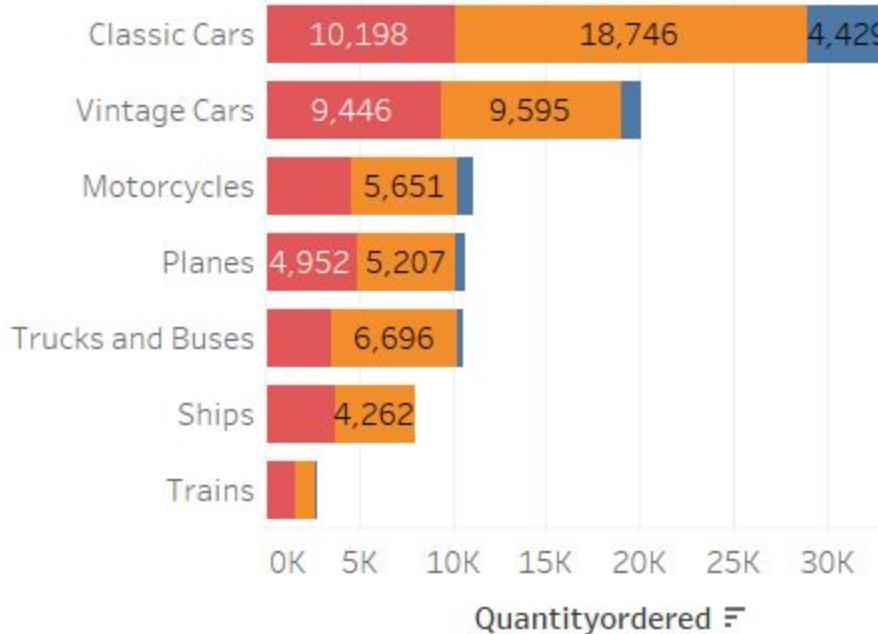
# Multi-Variate Analysis

---



## Multivariate - Sales and Deal Size by Product Line

Productline =

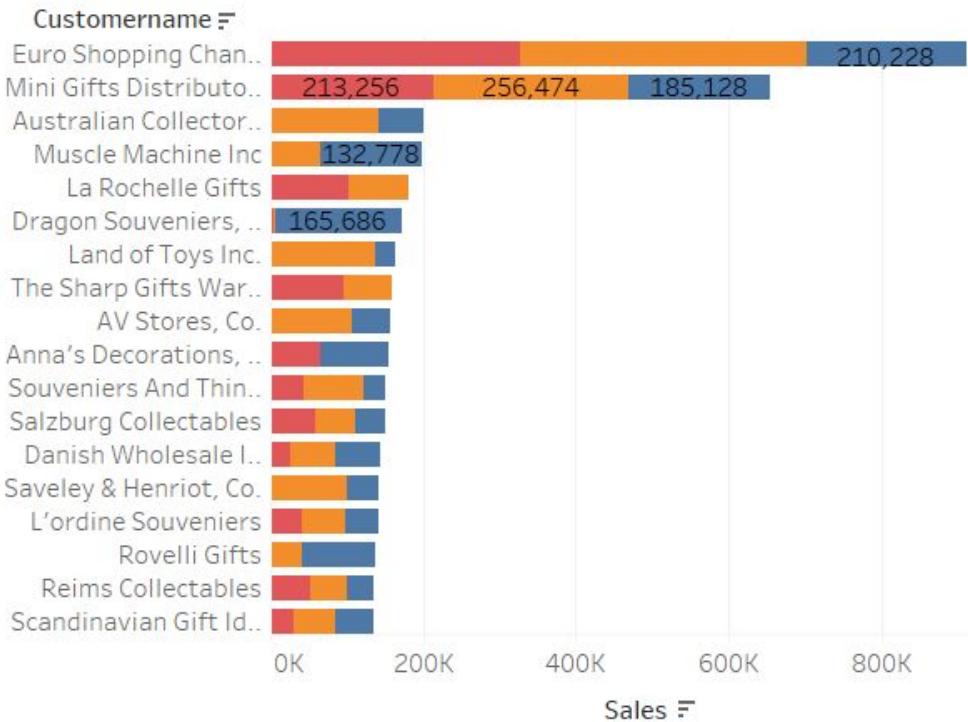


## Sales and Deal size by Product line

### Observation:

The graph indicates that Classic Cars and Vintage Cars are the most frequently ordered product lines, followed by Motorcycles, Planes, Trucks & Buses, and Ships. Conversely, Trains recorded the lowest number of orders. Notably, most orders across all product lines fall into the medium and small deal size categories, with higher deal sizes being relatively uncommon among these popular products.

## Multivariate - Customer Sales Analysis by Year



## Customer Sales Analysis by Year (2018-2020):

### Observation:

#### Euro Shopping Channel:

Achieved the highest sales across all three years and maintained consistent sales performance.

#### Mini Gift Distributor Ltd:

Ranked second in sales, demonstrating steady growth year over year.

#### Australian Collectors Co:

Ranked third in sales but had no recorded sales in 2020.

#### Muscle Machine Inc:

Fourth in sales, with no sales recorded in 2020.

#### La Rochelle Gifts:

Joined as a new company in 2019 and showed promising sales in its first year of operation.

## Multivariate - Sales Analysis by Country and Order Status

Country	Status					
	Cancelled	Disputed	In Process	On Hold	Resolved	Shipped
Australia		14,378	43,971			572,274
Austria					28,551	173,512
Belgium			8,412			100,001
Canada						224,079
Denmark		26,013			24,079	195,546
Finland						329,582
France			43,785			1,067,132
Germany						220,472
Ireland						57,756
Italy						374,674
Japan						188,168
Norway						307,464
Philippines						94,016
Singapore						288,488
Spain	50,011	31,822	35,133		53,816	1,044,905
Sweden	48,711		26,260			135,043
Switzerland						117,714
UK	50,408					428,472
USA	45,358		13,429	152,719	44,273	3,099,797

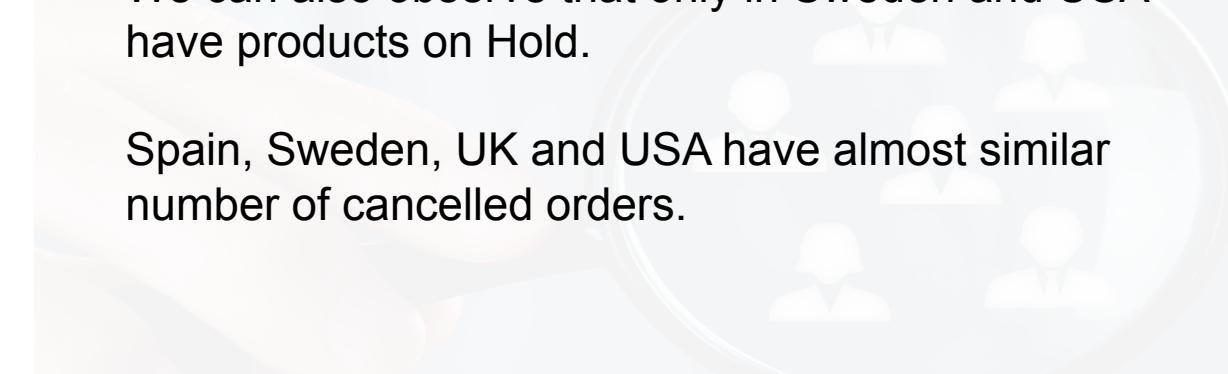
## Customer Sales Analysis by Country and Order Status:

### Observation:

As seen the maximum number orders shipped is in USA followed by France and Spain.

We can also observe that only in Sweden and USA have products on Hold.

Spain, Sweden, UK and USA have almost similar number of cancelled orders.



## **Summary:**

### **Product Line Analysis**

Classic cars and vintage cars are the most popular product lines, indicating strong demand for these vehicle models. In contrast, ships, trucks, buses, and trains have relatively lower sales counts, suggesting a niche market or specialized customer base for these categories.

### **Deal Size Analysis**

The majority of deals fall into the small and medium categories, with counts of 1,246 and 1,349, respectively. This indicates that the company primarily engages in transactions involving moderate to smaller deal sizes.

### **Order Status Analysis**

The significant count of 2,541 successfully shipped orders highlights the company's exceptional order fulfillment capabilities and its commitment to delivering orders promptly.

### **Quality Ordered Analysis**

Bins for quality orders of 21, 28, 35, and 42 have the highest counts, reflecting their popularity among customers. With counts ranging from 576 to 631, these quantities effectively meet the needs of a substantial portion of the customer base.

### **Sales Analysis**

Sales showed a positive trend, growing significantly from 2018 (3,353,014) to 2019 (4,669,925). This suggests that the business performed well and experienced increased customer demand. However, with sales data available only until May for 2020 (1,737,283), it is challenging to draw definitive conclusions about the overall performance for that year.

# INFERENCES AS PER EDA

---

Yearly sales have experienced a decline, which is concerning. However, Quarter 4 consistently outperforms other quarters. Sales remain stable during the first four months of the year, with Thursday showing the lowest sales and Sunday the highest.

Sales tend to increase from Friday to Sunday, while they dip from Monday to Thursday. Additionally, sales are generally higher in the early days of the month compared to the end days.

Most orders on hold are from the USA, with some from Sweden. Canceled orders are approximately equal among Spain, the USA, the UK, and Sweden. Spain also reports the highest number of disputes, although most of these disputes are resolved. Notably, classic car parts account for the highest percentage of sales.

# Recommendation as Per EDA

---

Further investigation is essential to understand the reasons behind the decline in yearly sales. To capitalize on the high sales observed in Quarter 4, businesses should focus on increasing their inventory during this peak period.

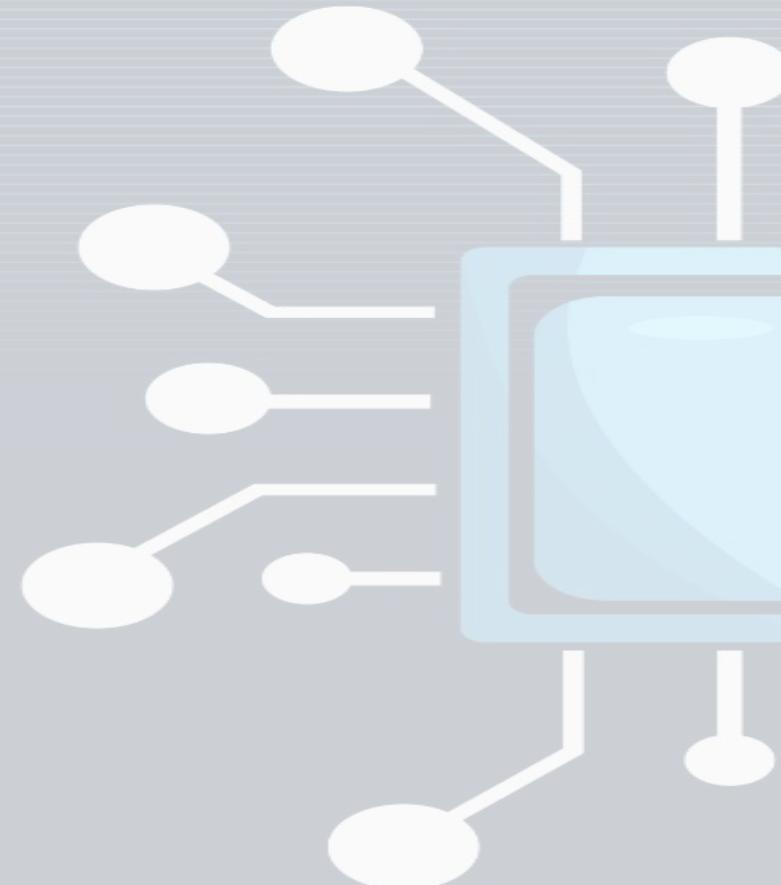
Additionally, it's important to identify the factors contributing to low sales in June and develop strategies to address them. Marketing efforts should be directed towards Thursdays to enhance sales on that day. Implementing weekend sales could also help boost sales from Friday to Sunday.

Offering discounts or promotions at the start of the month may increase early-month sales. Moreover, efforts should be made to resolve orders on hold to prevent potential revenue loss. Businesses should closely monitor canceled orders to identify any trends and take appropriate actions.

Finally, swift and efficient resolution of disputes is crucial for maintaining customer satisfaction. Given the high sales percentage of classic car parts, businesses should consider expanding their inventory in this category.

### **3. Customer Segmentation using RFM analysis (4 segments)**

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

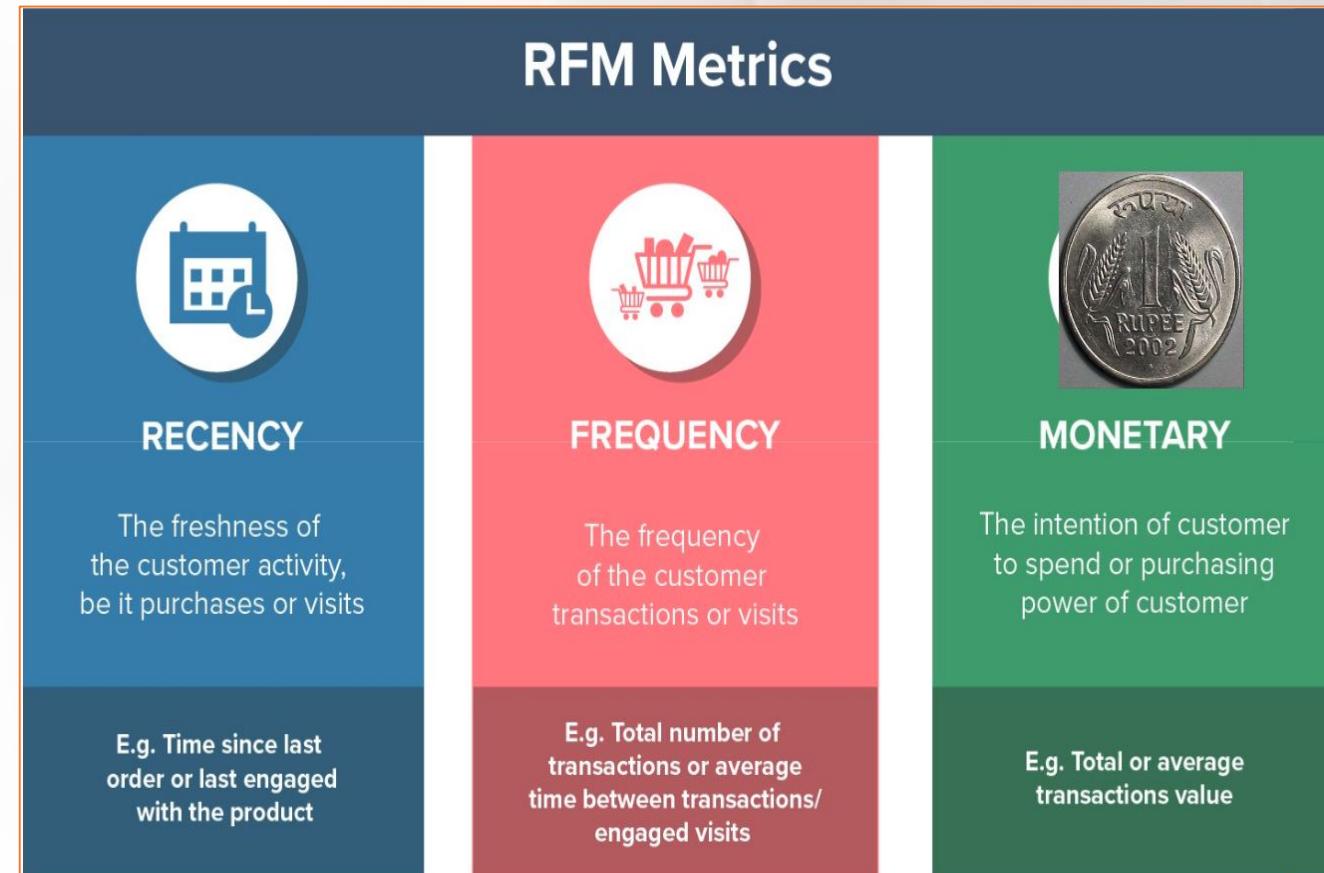


# What is RFM?

RFM analysis is a widely used customer segmentation technique in marketing and retail that evaluates and categorizes customers based on their purchasing behavior. RFM stands for Recency, Frequency, and Monetary Value—three key metrics that help assess customer engagement and profitability.

1. **Recency** refers to how recently a customer has made a purchase. Customers who have purchased more recently are typically more likely to respond to marketing efforts.
2. **Frequency** measures how often a customer makes purchases over a given period. Frequent buyers are often more loyal and valuable to the business.
3. **Monetary Value** assesses the total amount of money a customer has spent. Higher spending customers are usually considered more valuable.

By scoring customers in these three categories, businesses can identify their best clients based on their spending habits, allowing for more targeted marketing strategies and improved customer retention efforts. RFM analysis enables firms to focus on their most valuable customers and tailor their engagement strategies accordingly.



## RFM Analysis Parameters

- **Monetary Value:** Calculated by multiplying the price of each item by the quantity ordered. This metric represents the revenue generated by each customer.
- **Recency:** Determined by subtracting the order date from a fixed reference date, such as June 1, 2020. This metric indicates how recently a customer made a purchase.
- **Frequency:** Calculated as the count of purchases made by each customer. This reflects how often a customer engages in transactions.
- **Auto-binning:** Customers are segmented into four categories based on their RFM scores: High, Moderate, Low, and Very Low. This segmentation helps categorize customers based on their value and behavior.

These parameters are crucial in RFM analysis, allowing businesses to evaluate customer behavior, identify distinct customer segments, and make data-driven marketing and sales decisions.

## Assumptions

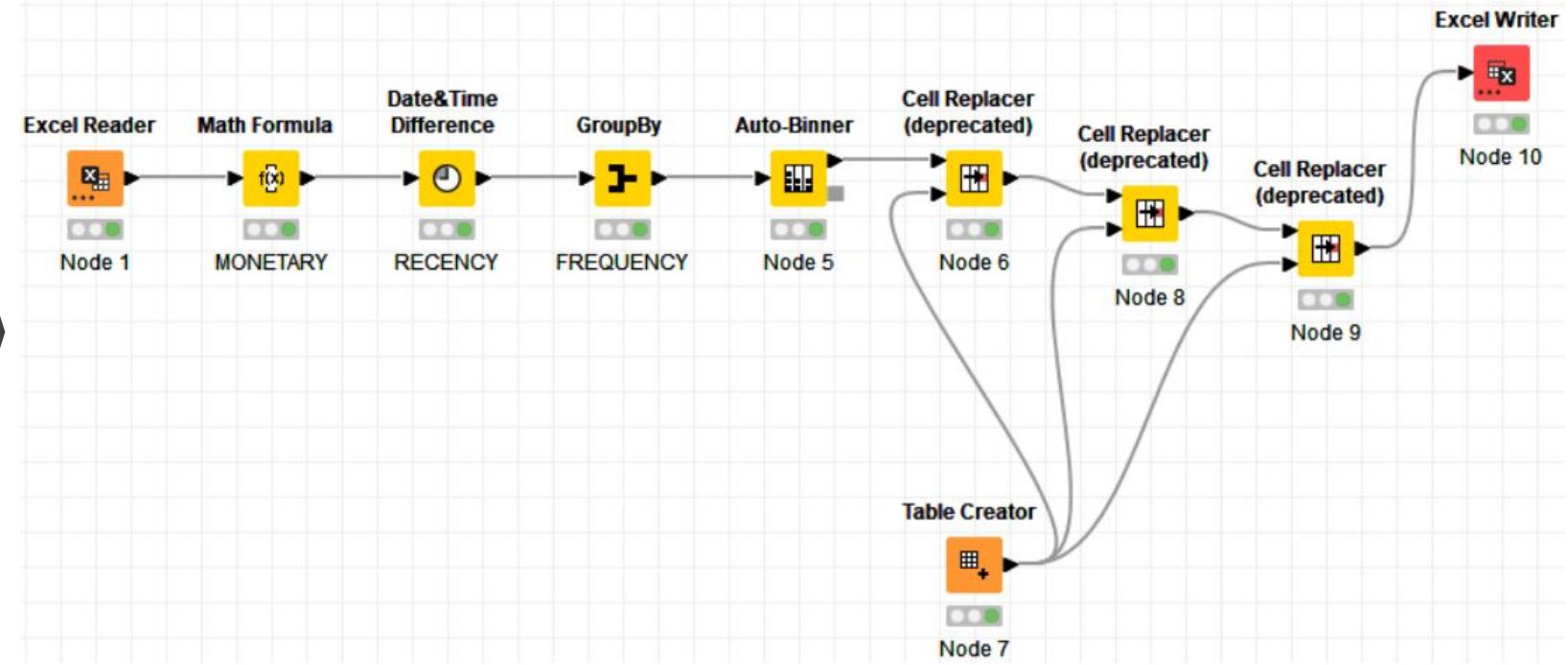
1. **Higher Monetary Value Indicates Higher Spending Customers:** It is assumed that customers generating higher monetary values through their purchases are likely to be more valuable and potentially more profitable for the business.
2. **Recent Purchases Reflect Customer Engagement:** The assumption is that customers who have made recent purchases are more engaged with the company and its offerings, suggesting a higher potential for repeat purchases or upselling/cross-selling opportunities.
3. **Higher Frequency of Purchases Reflects Customer Loyalty:** It is assumed that customers who purchase more frequently exhibit greater loyalty to the company. Such customers may have a stronger connection to the brand, higher satisfaction levels, and a greater likelihood of recommending the company to others.

# Tool used : KNIME

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



# KNIME WORKFLOW



# Output table head

...	S	CUSTOMERNAME	I	ORDERNUMBER	D	QUANTITYORDERED	D	PRICEEACH	D	ORDERLINENUMBER	D	SALES	31	ORDERDATE	D	DAYS_SINCE_LASTORDER	S	STATUS
...	AV Stores, Co.	51		34.863		91.085		9.02				3,094.271	2019-10-14		1,803.804		Shipped	
...	Alpha Cognac	20		34.35		101.16		4.95				3,524.422	2018-11-08		2,236.2		Shipped	
...	Amica Models & Co.	26		32.423		110.853		7.615				3,619.895	2019-08-17		1,318.615		Shipped	
...	Anna's Decorations, Ltd	46		31.935		106.424		6.435				3,347.742	2020-03-09		1,463.587		Shipped	
...	Atelier graphique	7		38.571		92.239		2				3,454.28	2019-09-27		1,424.429		Shipped	
...	Australian Collectables, Ltd	23		30.652		90.042		6.696				2,808.324	2018-11-21		1,689.696		Shipped	
...	Australian Collectors, Co.	55		35.018		104.59		7.036				3,654.462	2019-02-20		1,698.782		Shipped	
...	Australian Gift Network, Co	15		36.333		110.554		3.067				3,964.608	2018-11-06		1,349.133		Shipped	
...	Auto Assoc. & Cie.	18		35.389		99.488		8.556				3,601.907	2019-10-11		1,266.278		Shipped	
...	Auto Canal Petit	27		37.074		94.255		6.333				3,450.765	2019-01-15		1,763.444		Shipped	
...	Auto-Moto Classics Inc.	8		35.875		92.8		2				3,309.908	2018-06-16		2,281.75		Shipped	
...	Baane Mini Imports	32		33.812		108.574		6.344				3,643.725	2019-10-15		1,805.75		Shipped	
...	Bavarian Collectables Imports, Co.	14		28.643		84.289		7.5				2,499.566	2019-09-15		2,239.357		Shipped	
...	Blauer See Auto, Co.	22		36.864		108.031		3.864				3,871.436	2018-10-04		1,697.318		Shipped	
...	Boards & Toys Co.	3		34		89.807		1.333				3,043.117	2020-02-08		1,968.667		Shipped	
...	CAF Imports	13		36		104.963		5.308				3,818.619	2019-03-19		2,273		Shipped	
...	Cambridge Collectables Co.	11		32.455		101.329		3.273				3,287.602	2019-03-10		1,788.182		Shipped	
...	Canadian Gift Exchange Network	22		31.955		105.341		6				3,419.951	2018-12-05		1,620.045		Shipped	
...	Classic Gift Ideas, Inc	21		31.81		103.32		5.857				3,214.618	2018-11-13		1,668.095		Shipped	
...	Classic Legends Inc.	20		36		109.803		4.05				3,889.76	2018-10-20		1,931.6		Shipped	
...	Clover Collections, Co.	16		30.625		112.87		4.562				3,609.777	2019-02-12		1,939.062		Shipped	
...	Collectable Mini Designs Co.	25		38.16		91.535		7.96				3,499.569	2019-02-19		2,182.12		Shipped	
...	Collectables For Less Inc.	24		33.125		97.237		4.875				3,399.083	2020-01-20		1,546.125		Shipped	
...	Corrida Auto Replicas, Ltd	32		36.344		105.175		6.781				3,769.227	2018-05-28		1,761.156		Shipped	

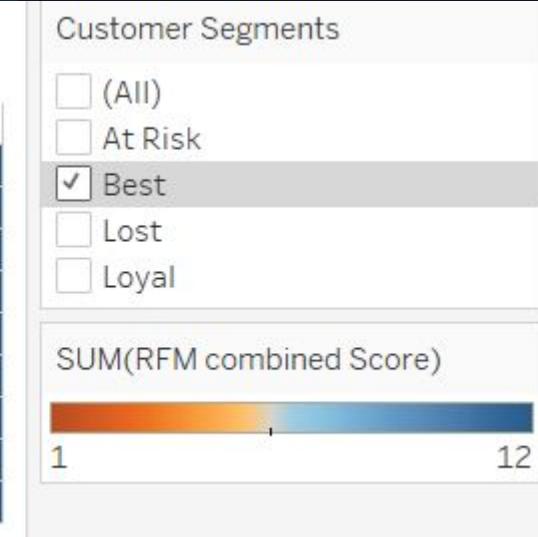
## **04. Inferences from RFM Analysis and identified segments:**

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

# TOP Best Customers

## RFM-Customers Division

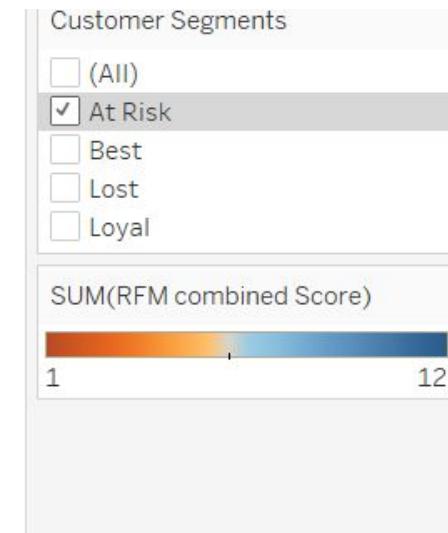
Customername	Recency Hml	Frequency Hml	Monetary Hml	
Danish Wholesale Imports	High	High	High	12
Euro Shopping Channel	High	High	High	12
L'ordine Souveniers	High	High	High	12
La Rochelle Gifts	High	High	High	12
Mini Gifts Distributors Ltd.	High	High	High	12
Reims Collectables	High	High	High	12
Salzburg Collectables	High	High	High	12
Souveniers And Things Co.	High	High	High	12
The Sharp Gifts Warehouse	High	High	High	12



- The top best customers are identified by calculating their RFM scores, which are the sum of their scores for Monetary Value, Recency, and Frequency. Higher RFM scores indicate more valuable customers.
  - Based on the RFM scores, we have grouped the top customers, placing greater emphasis on recency to prioritize those who have engaged with the company more recently. This approach allows us to identify and order customers according to their current value and engagement level.

RFM-Customers Division

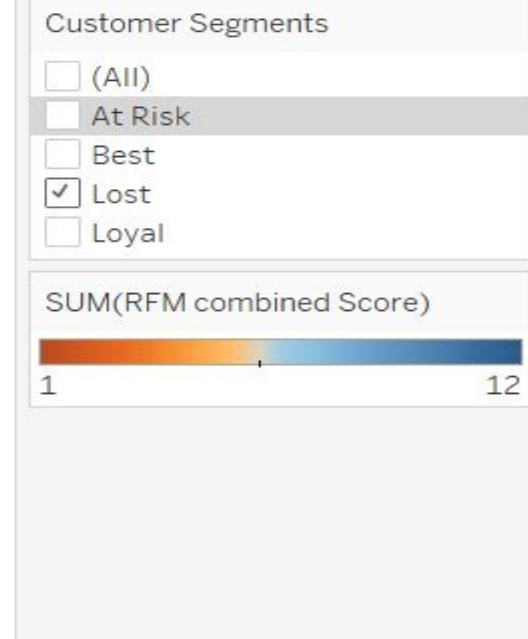
Customername	Recency Hml	Frequency Hml	Monetary Hml	
Alpha Cognac	High	Very Low	Very Low	6
Amica Models & Co.	Very Low	Low	Moderate	6
Atelier graphique	Low	Very Low	Very Low	4
Australian Collectables, Ltd	High	Low	Very Low	7
Australian Gift Network, Co	Moderate	Very Low	Very Low	5
Auto Canal Petit	High	Moderate	Moderate	10
Auto-Moto Classics Inc.	Moderate	Very Low	Very Low	5
AV Stores, Co.	Low	High	High	10
Baane Mini Imports	Low	Moderate	Moderate	8
Blauer See Auto, Co.	Low	Low	Low	6
Boards & Toys Co.	Moderate	Very Low	Very Low	5
Canadian Gift Exchange Network	Low	Low	Low	6



- Customers with very low to low recency, but high to moderate monetary value and purchase frequency, are at an increased risk of churning. The RFM analysis indicates that these are the top customers who are on the verge of disengagement.
  - To prevent losing these valuable customers, it is essential to focus on retaining them. Implementing a targeted action plan could help convert them into regular customers. This plan might include personalized outreach, special promotions, loyalty programs, or exclusive offers designed to re-engage these customers and encourage them to make additional purchases.

## RFM-Customers Division

Customername	Recency Hml	Frequency Hml	Monetary Hml	
Auto Assoc. & Cie.	Very Low	Very Low	Very Low	3
Bavarian Collectables Imports, Co.	Very Low	Very Low	Very Low	3
CAF Imports	Very Low	Very Low	Very Low	3
Cambridge Collectables Co.	Very Low	Very Low	Very Low	3
Clover Collections, Co.	Very Low	Very Low	Very Low	3
Daedalus Designs Imports	Very Low	Very Low	Very Low	3
Double Decker Gift Stores, Ltd	Very Low	Very Low	Very Low	3
Iberia Gift Imports, Corp.	Very Low	Very Low	Very Low	3
Online Mini Collectables	Very Low	Very Low	Very Low	3
Osaka Souveniers Co.	Very Low	Very Low	Very Low	3
Signal Collectibles Ltd.	Very Low	Very Low	Very Low	3
West Coast Collectables Co.	Very Low	Very Low	Very Low	3



Customers exhibiting low purchase frequency show reduced engagement, while those with low recency indicate prolonged inactivity. Additionally, customers with low monetary value suggest decreased spending. The RFM analysis indicates that these customers have effectively been lost, as they have not made purchases frequently and their recency is very low.

To better understand the reasons behind this loss, we should conduct a study or survey targeting these customers. Gathering insights on their experiences and reasons for disengagement will help us identify areas for improvement. Based on this feedback, we can implement strategies to re-engage these customers and reduce the likelihood of future losses.

# Top Loyal Customers

## RFM-Customers Division

Customername	Recency Hml	Frequency Hml	Monetary Hml	
Anna's Decorations, Ltd	Moderate	High	High	11
Australian Collectors, Co.	Moderate	High	High	11
Dragon Souveniers, Ltd.	Moderate	High	High	11
Mini Creations Ltd.	Moderate	High	Moderate	10
Muscle Machine Inc	Moderate	High	High	11
Scandinavian Gift Ideas	Moderate	High	High	11
Technics Stores Inc.	Moderate	High	High	11

## Customer Segments

(All)

At Risk

Best

Lost

Loyal

SUM(RFM combined Score)

1

12

- Loyal customers are those who consistently make repeat purchases over time. While they may not always represent the highest monetary value, their commitment to the brand is evident through their repeated business.
- According to RFM analysis, we have identified these loyal customers. By placing emphasis on their monetary value, we can develop targeted strategies to enhance their experience. Focusing on these customers presents an opportunity to nurture them further and potentially elevate them into our best customers.

# 05. RECOMMENDATIONS as Per RFM Analysis:

## Customer Retention Strategies

1. **Engagement and Retention:** Focus on customers at risk of churning by offering personalized incentives, loyalty programs, or exclusive promotions to encourage their continued loyalty.
2. **Continuous Monitoring and Analysis:** Regularly track customer behavior, sales trends, and key performance indicators (KPIs) to identify shifting patterns and proactively address potential issues.
3. **Upselling and Cross-Selling Opportunities:** Identify chances to upsell and cross-sell to existing customers, particularly those with high monetary value. Recommend relevant products or services based on their purchase history to boost average order value and enhance loyalty.

## Customer Segmentation According to RFM Analysis

- **Best Customers:** For these customers, provide personalized recognition, exclusive offers, and incentives to ensure they continue choosing our company. This approach helps maintain their loyalty and strengthens long-term relationships.
- **Loyal Customers:** Offer periodic discounts and promotions to keep them engaged and interested in our products or services. Keeping them engaged can transform them into our best customers and improve their satisfaction levels.
- **Customers on the Verge of Churn:** Develop an effective action plan to prevent these customers from leaving. Conduct surveys, offer incentives, and personalize communication to identify and address their concerns, thereby increasing their loyalty.
- **Lost Customers:** Analyze their behavior and preferences to understand the reasons for their departure. Identifying key drivers of customer churn will allow us to create targeted strategies to prevent future losses and improve overall retention rates.

# Part-B:Grocery Store Market Basket Analysis

## *Problem Statement*

---

A Grocery Store shared the transactional data with you. Your job is to identify the most popular combos that can be suggested to the Grocery Store chain after a thorough analysis of the most commonly occurring sets of items in the customer orders. The Store doesn't have any combo offers. Can you suggest the best combos & offers





# Contents:

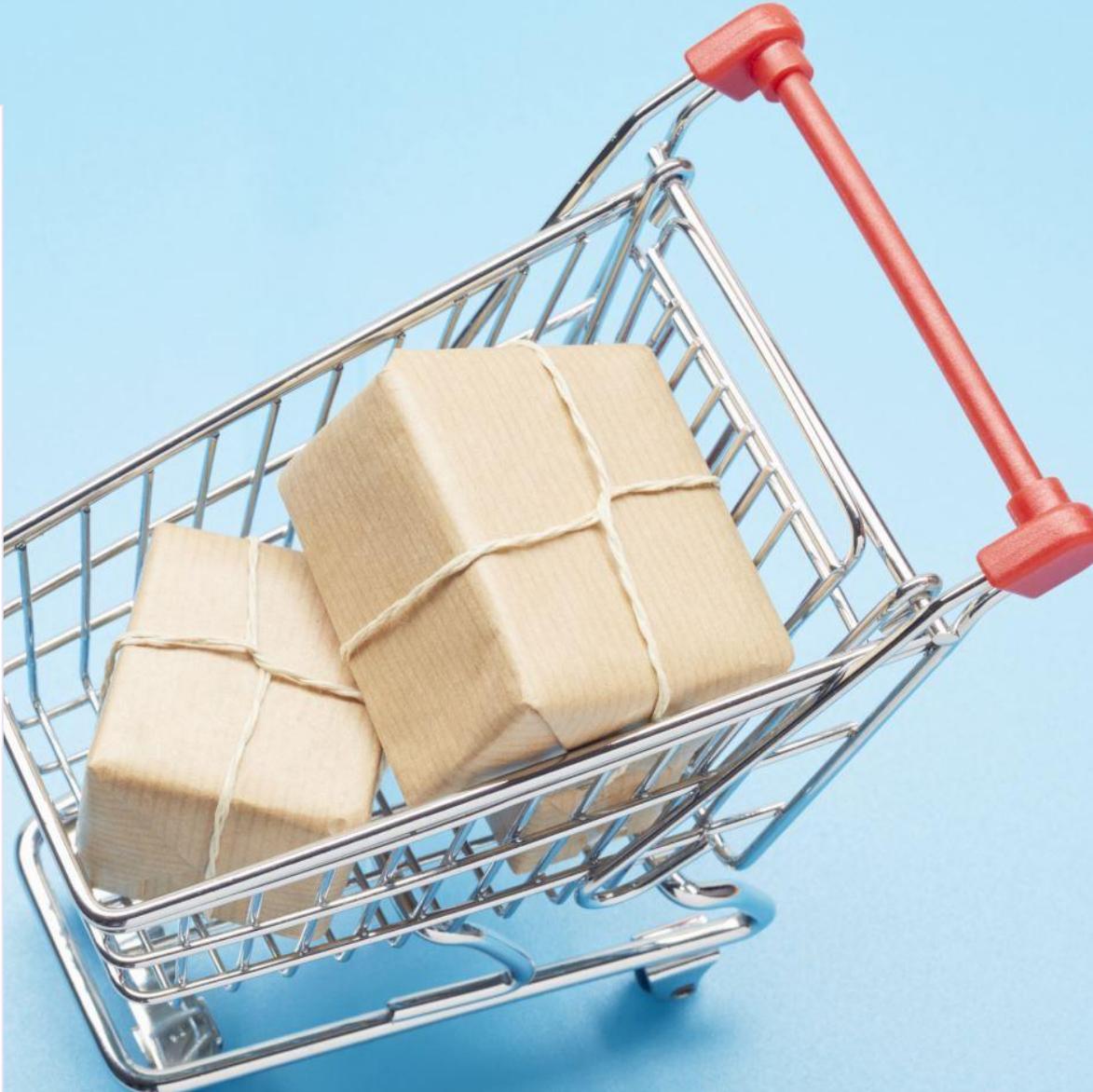
---

- Executive Summary of the data
  - Problem statement
  - Executive Summary & Data Dictionary
  - Assumptions about data
- Exploratory Analysis & Insights
  - Weekly, Monthly, Quarterly, Yearly Weekday Trends in Sales count
  - Products counts & Year Wise top products
  - Summary and Recommendations
- Market Basket Analysis
  - Market Basket Analysis Meaning
  - MRA KNIME WorkFlow & Output Table
- Associations Identified
  - Association Rule Parameters
  - MRA – values
  - Association Rules Table
- Recommendation
  - Recommendation
  - Summary

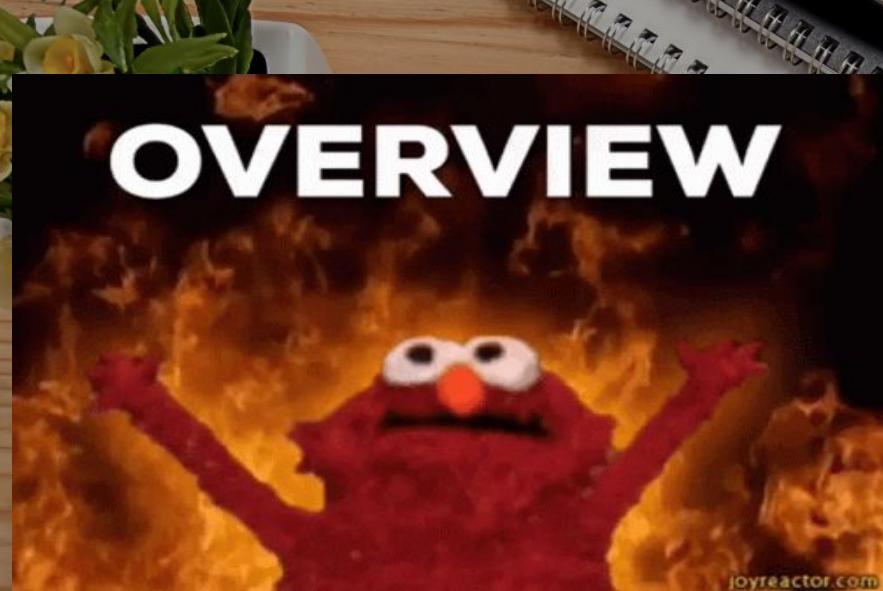
## Agenda:

---

- Executive Summary of the data
- Exploratory Data Analysis
- Market Basket Analysis
- Associations Identified
- Recommendation



# OVERVIEW



## 01 Executive Summary

- Problem statement
- Executive Summary & Data Dictionary
- Assumptions about data



# Executive Summary:

## *Dataset Description*

Total Rows: 20,641

Columns: 3 (Date, Order\_id, Product)

Missing Values: None

Duplicate Values: 4,730



## Project Overview

Data Range: January 1, 2018, to February 26, 2020

Objective: Conduct a comprehensive analysis of Point of Sale (POS) data to identify popular combo offers and discounts, ultimately aimed at increasing the grocery store's revenue.

While the dataset contains duplicate values, these reflect unique transactions where the same order ID corresponds to different products. Thus, the duplicates provide valuable insights into customer purchasing behavior and do not need to be removed.

## Analytical Approach

Exploratory Analysis: The exploratory analysis offers a comprehensive understanding of the data, revealing key trends and patterns in sales. This foundational analysis is crucial for identifying customer preferences and behaviors.

Market Basket Analysis: Utilizing association rules, this analysis uncovers relationships between products that customers frequently purchase together. By identifying these patterns, the grocery store can design attractive combo offers that encourage customers to buy complementary items.

## Actionable Recommendations

1. Create Combo Offers
2. Targeted Discounts
3. Promote High-Performing Combinations
4. Customer Engagement

# Data Dictionary

---



Date

date of product sold



Order\_id

Id of the order



Product

Name of the Product sold



# Duplicate Values:

In general, it is advisable to drop duplicate rows in a dataset to avoid skewing analysis results, as duplicates do not add additional information. However, in this specific case, removing duplicate rows may not be suitable.

The dataset includes a date, a customer ID, and a product purchased, and the same product can be bought by multiple customers on the same date. Therefore, if we were to drop duplicate rows, we might unintentionally eliminate valid information that reflects actual customer transactions.

As a result, the duplicate values will remain in the dataset, preserving the integrity of the data and ensuring that all relevant transactions are accounted for in the analysis.



# Assumptions:

The dataset captures items purchased at a grocery store over various dates. Each entry corresponds to a single item purchase.

- **First Column:** Represents the date of purchase.
- **Second Column:** Indicates the customer who made the purchase.
- **Third Column:** Specifies the item that was purchased.

It's important to note that the same item can be bought by multiple customers on different dates. The dataset does not include information regarding the quantity or price of each item. Additionally, we have chosen not to drop the duplicate values, as they reflect valid transactions and are essential for our analysis.



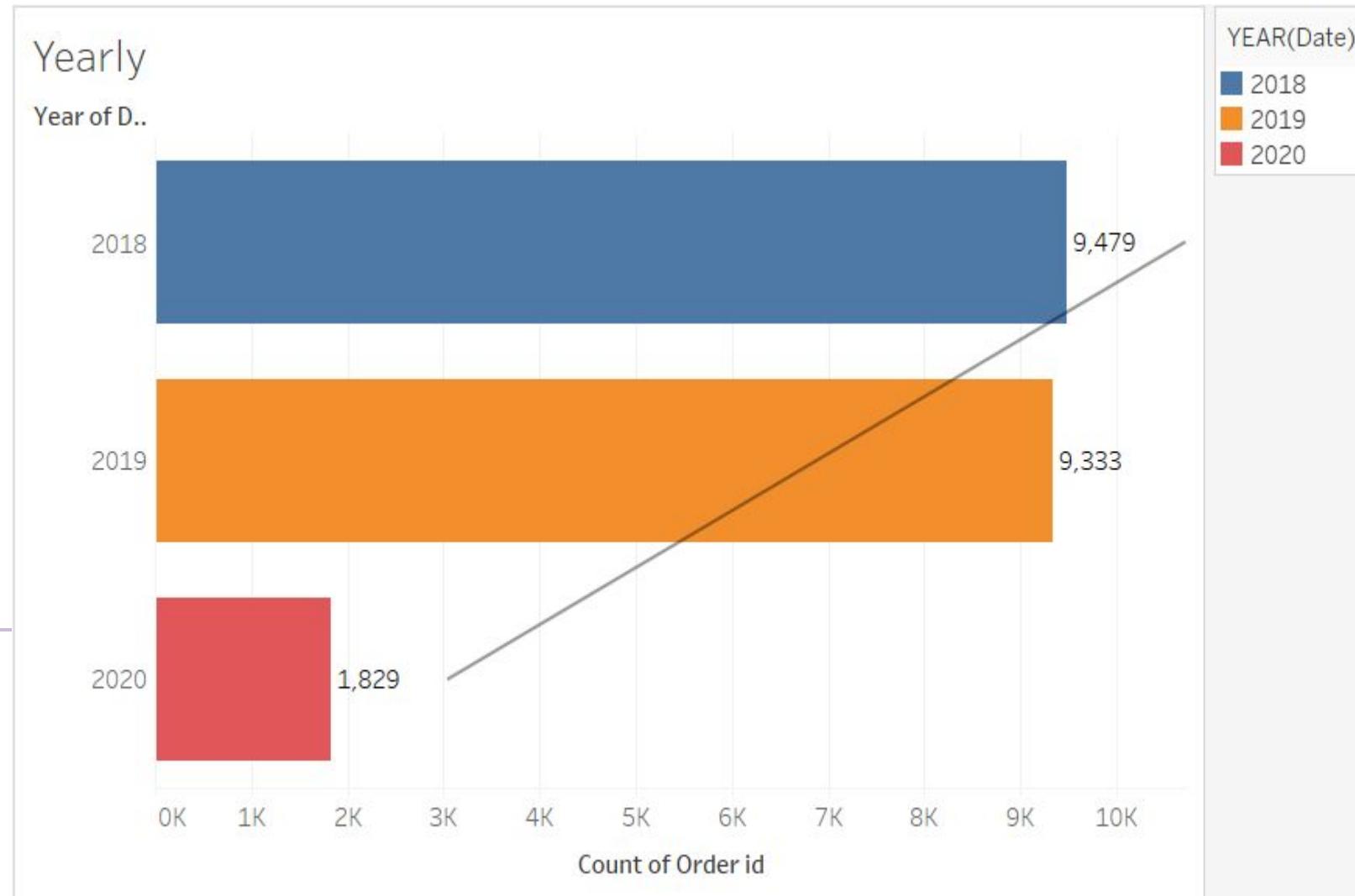
02

# Exploratory Data Analysis

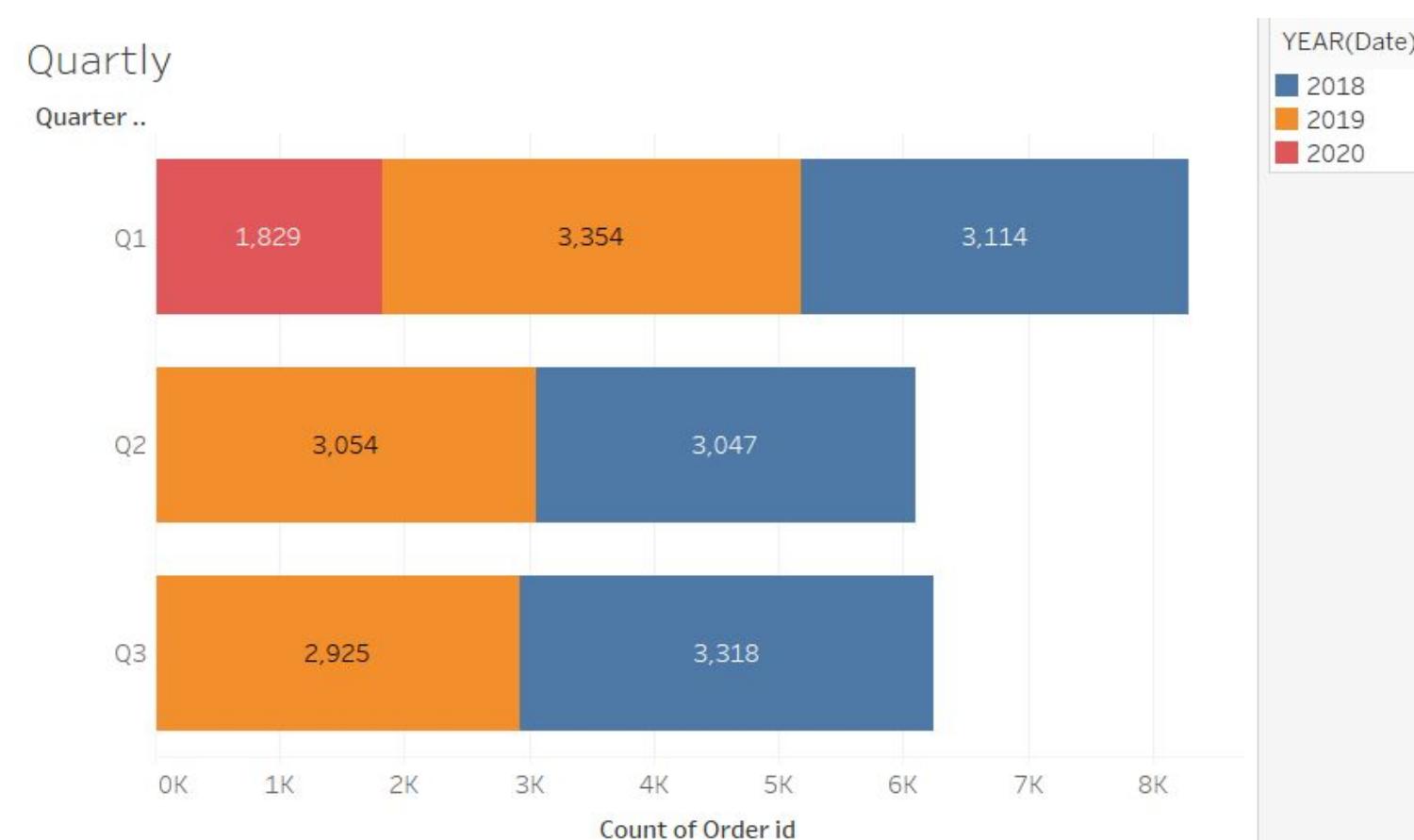
- Weekly, Monthly, Quarterly, Yearly Weekday Trends in Sales count
- Products counts & Year Wise top products
- Summary and Recommendations

# Yearly Count of Products Sold

As we have data till 26 feb 2020  
that's why the count of products  
sold in 2020 is low.



# Quarterly Count of Products Sold



Given that our dataset extends until February 26, 2020, the count of products sold in Q1 is notably high.

- **2019:** Q1 sales were the highest of the year.
- **2018:** Q3 recorded the highest sales.
- **Q2:** The count of products sold in Q2 remains approximately the same for both 2018 and 2019.

This trend suggests seasonal patterns in sales, with Q1 being particularly strong leading into the first quarter of 2020. Further analysis could help uncover the factors driving these variations in sales across different quarters.

# Monthly Count of Products Sold

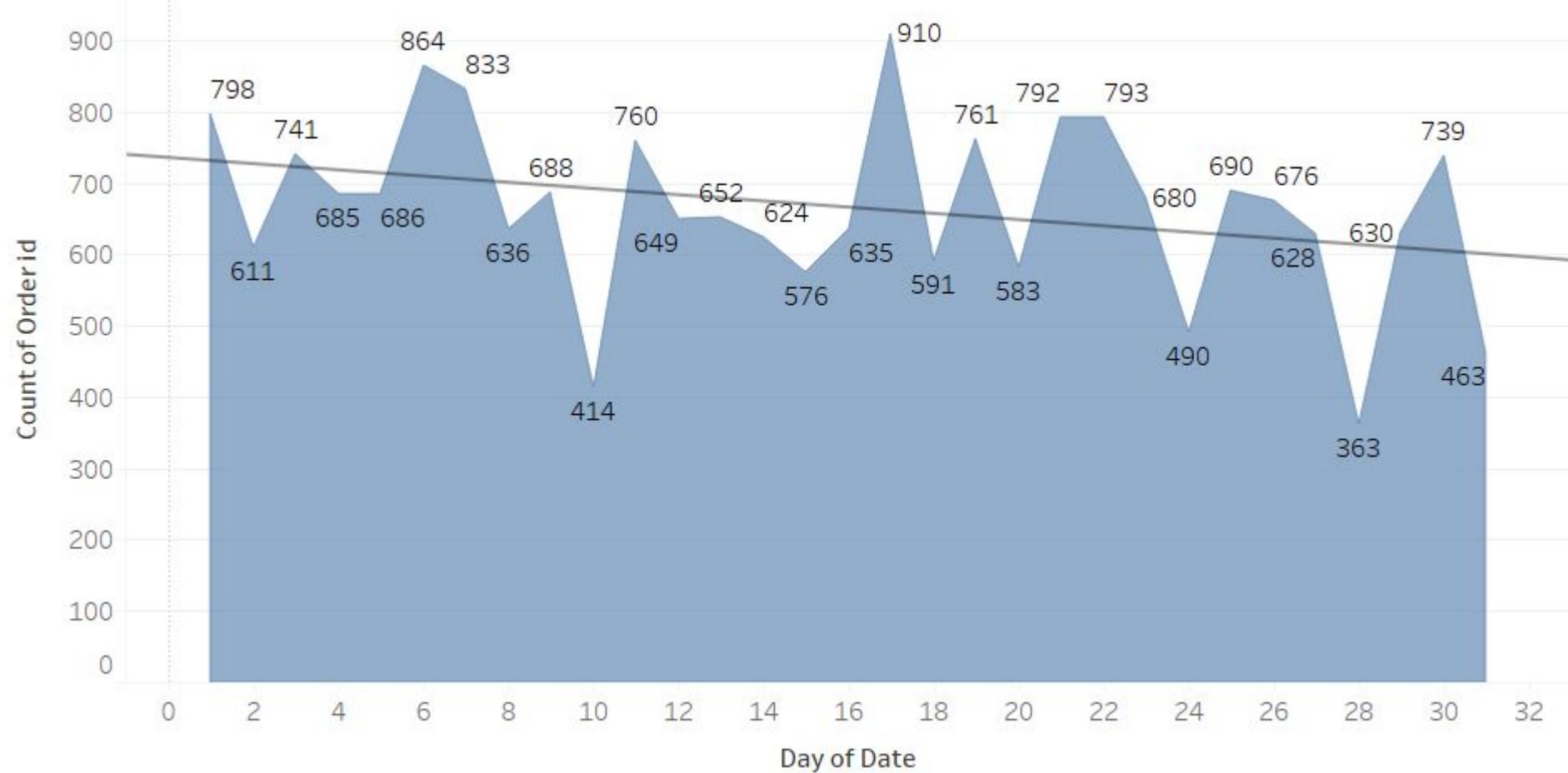
In 2018, the sales data indicates that the majority of products were sold in January, while February saw the least sales.

In contrast, 2019 experienced the highest sales in March, with January recording the lowest sales for that year.



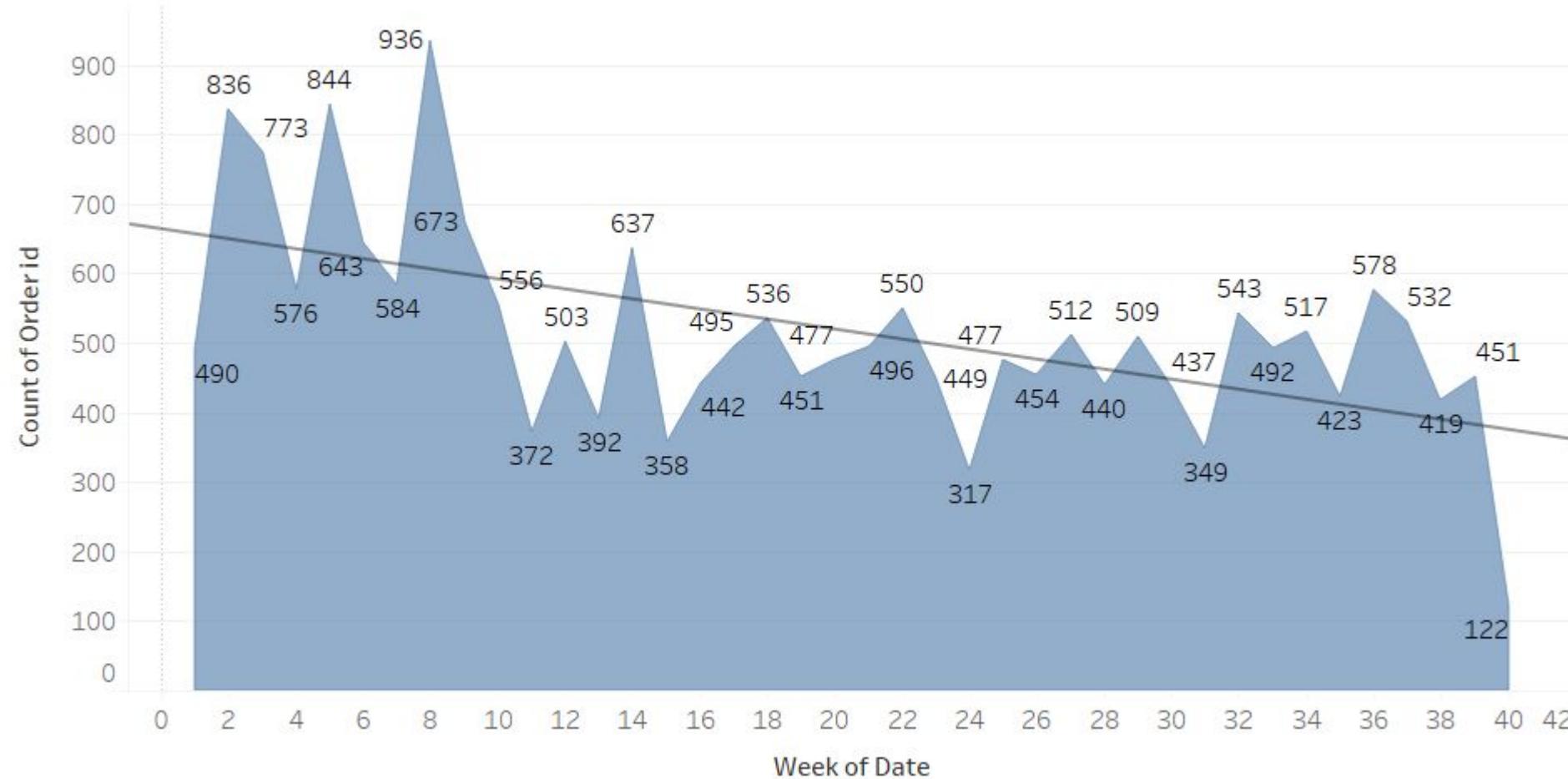
# Day of the Month Count of Products Sold

Daily



# Weekly Count of Products Sold

weekly

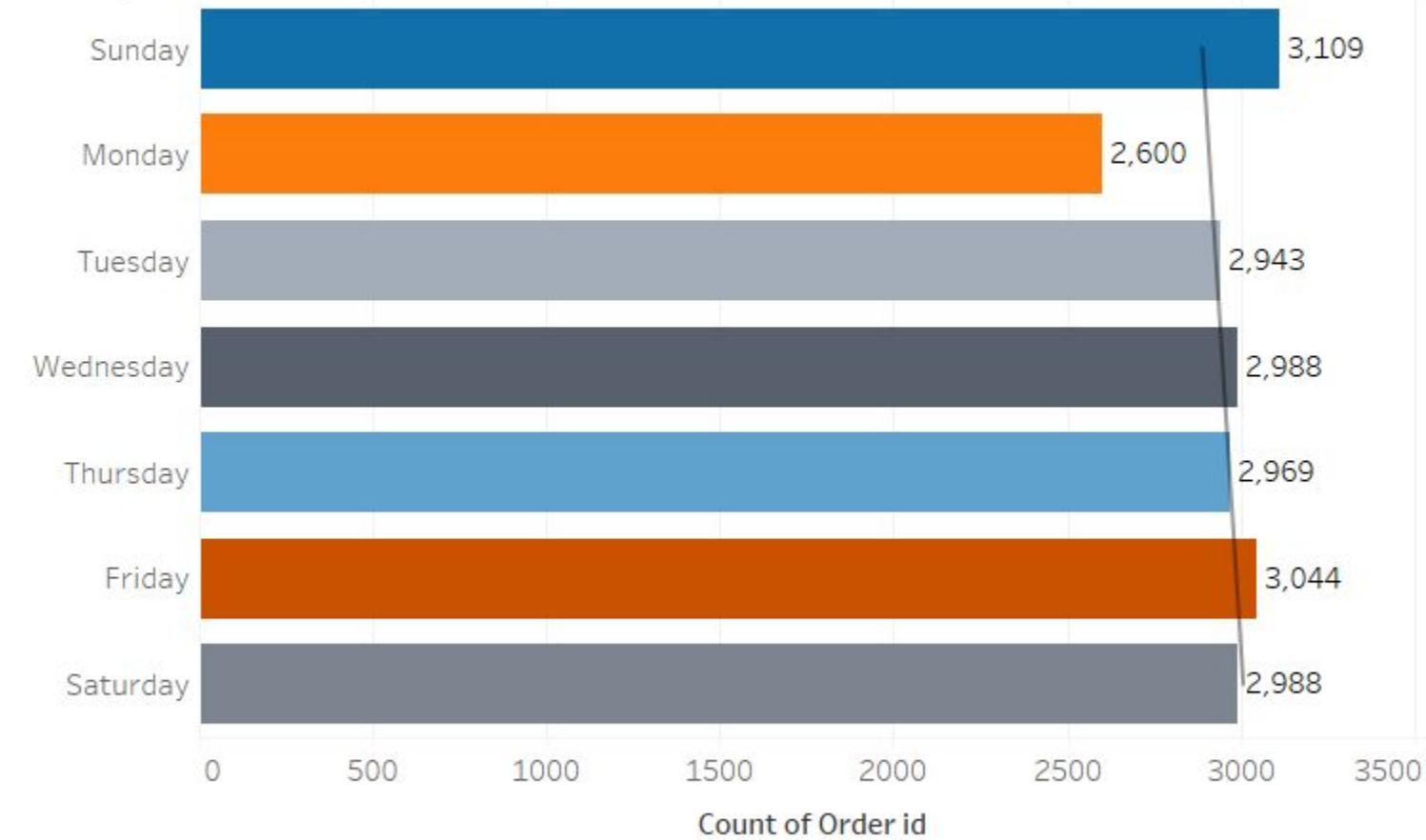


# Weekday Count of Products Sold

- The sales data reveals that most products were sold on Sundays, indicating a strong customer shopping tendency on that day. Conversely, Mondays recorded the least sales, suggesting that customer engagement may be lower at the beginning of the week.
- On other days, sales remained relatively consistent, showing a steady pattern of customer purchases throughout the week. This information can be valuable for planning promotions and staffing, particularly by focusing efforts on Sundays to capitalize on higher customer traffic.

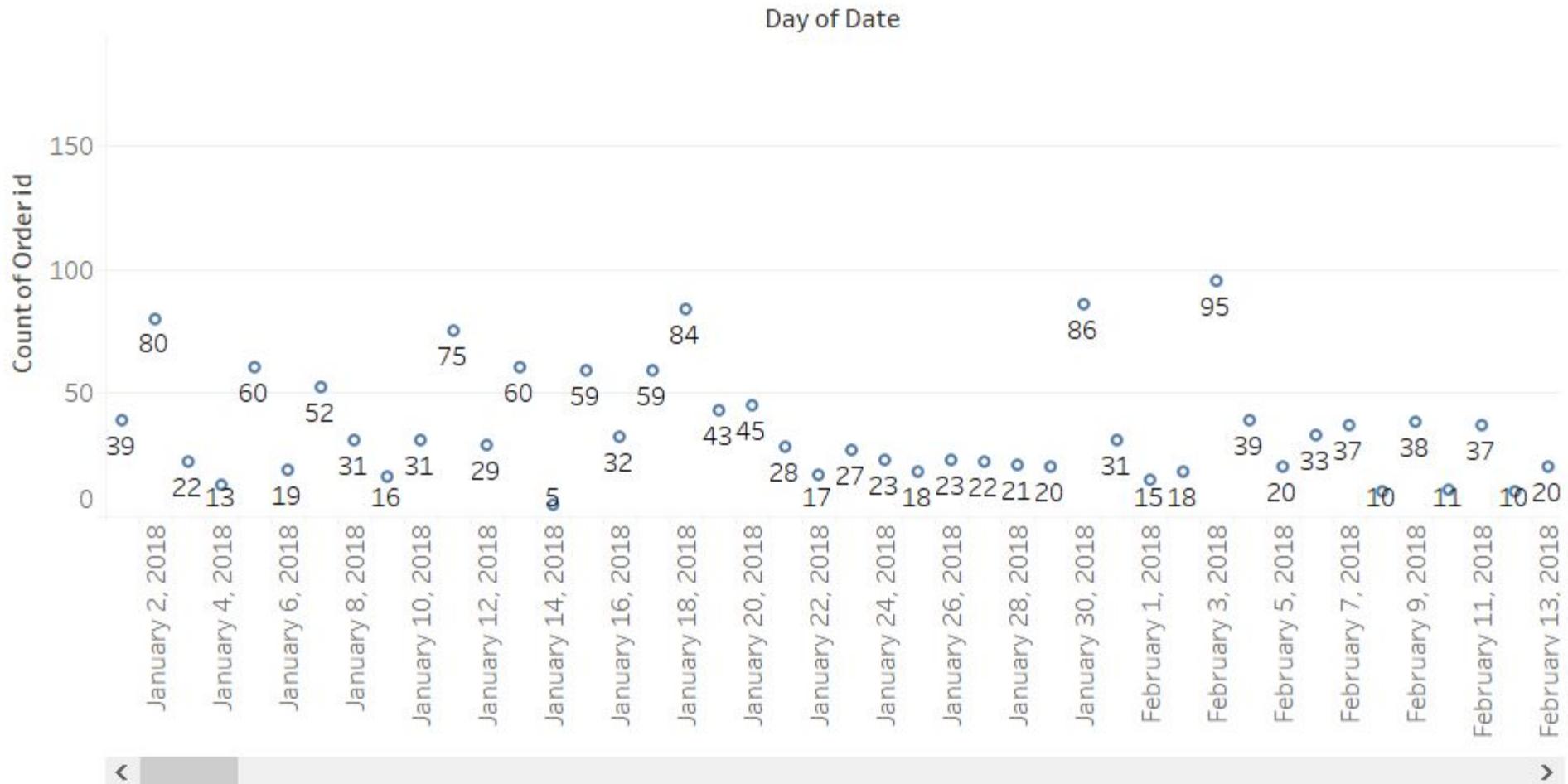
weekday part-1

Weekday of ..



# Daily Count of Sales

weekday part-2

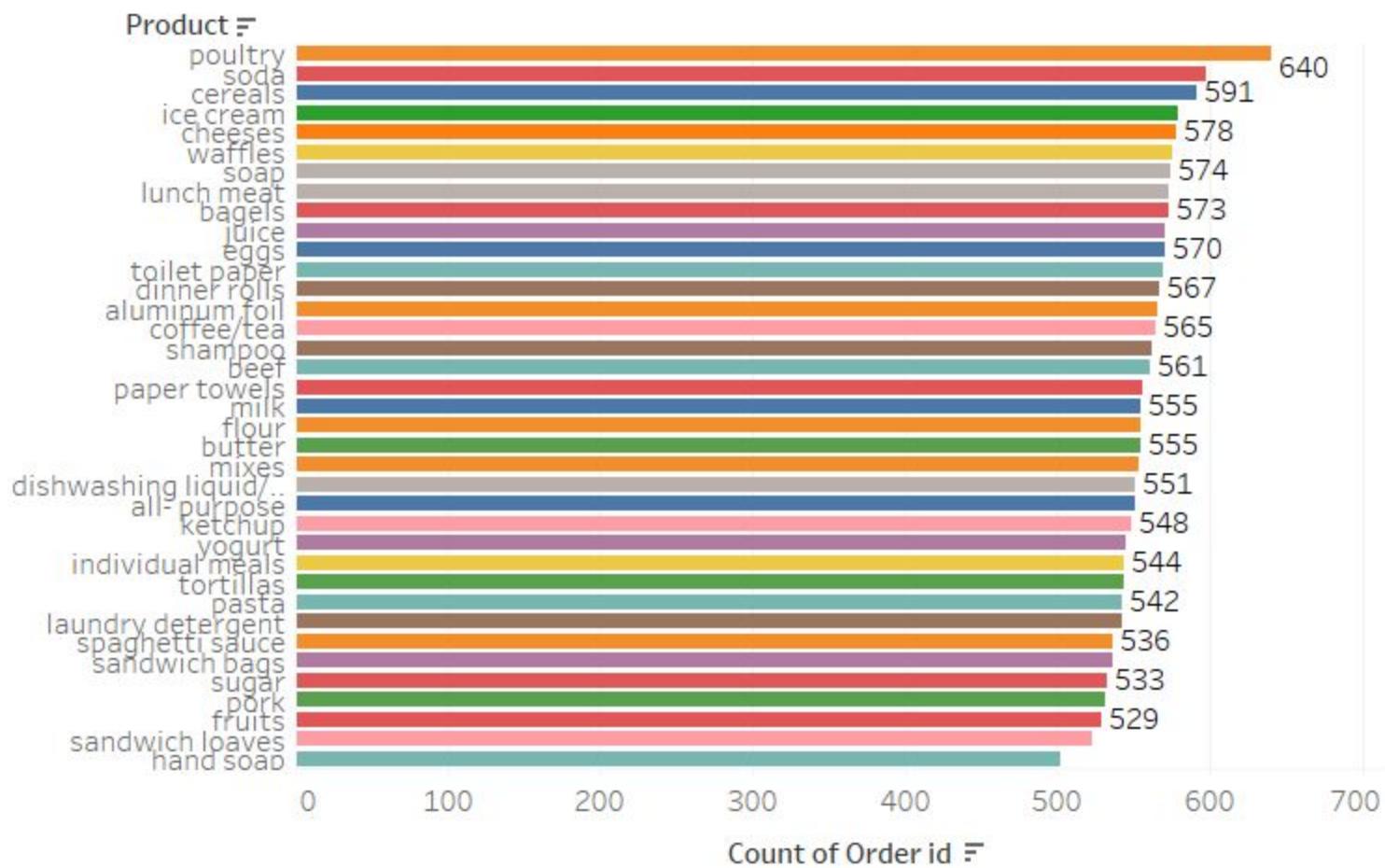


# Count of Products Sold

Treemap of products-1



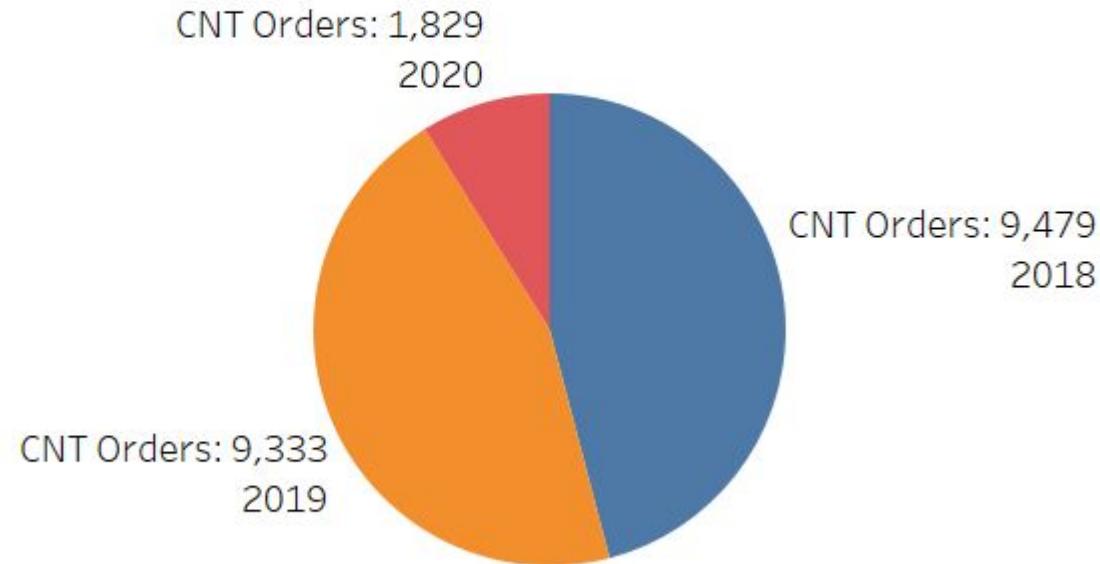
## Product ordered



# Count of Products Sold Yearly

Most of the Products were sold in 2018.

Pie chart



# Count of Eatable Products Sold:

- There are total 28 products in this category.
- Highest sold : poultry , Soda, cereals
- Least sold : pork, fruits, sandwich loaves

Eatables products

poultry 640	waffles 575	dinner rolls 567	milk 555	mixes 554	ketchup 548	yogurt 545
soda 597	bagels 573	coffee/tea 565	individual meals 544	spaghetti sauce 536	sugar 533	
cereals 591	lunch meat 573	beef 561	tortillas 543			
ice cream 579	eggs 570	butter 555	pasta 542	pork 531	fruits 529	
cheeses 578	juice 570	flour 555	sandwich bags 536			sandwich loaves 523

# Count of Non-Eatables Products Sold:

- There are 9 products in this category.
- Highest Sold Products : Soap, Toilet Paper.
- Lowest Sold Products: Hand soap.

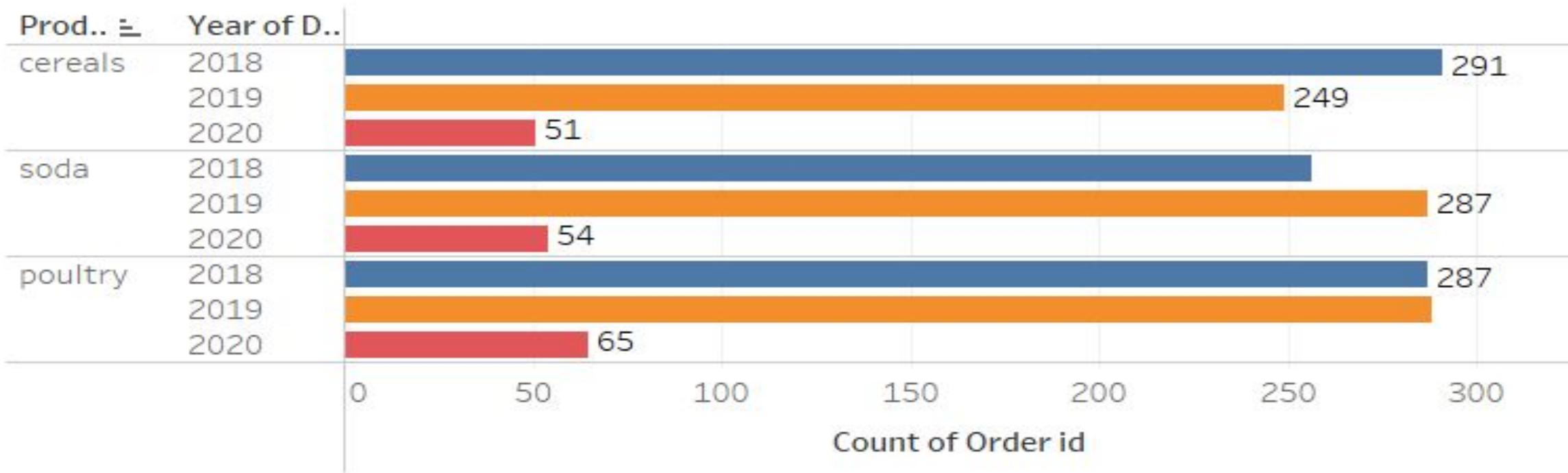
non eatables products

soap 574	shampoo 562	dishwashing	laundry detergent 542
toilet paper 569	paper towels 556		
aluminum foil 566	all-purpose 551	hand soap 502	

# Top three products over the years

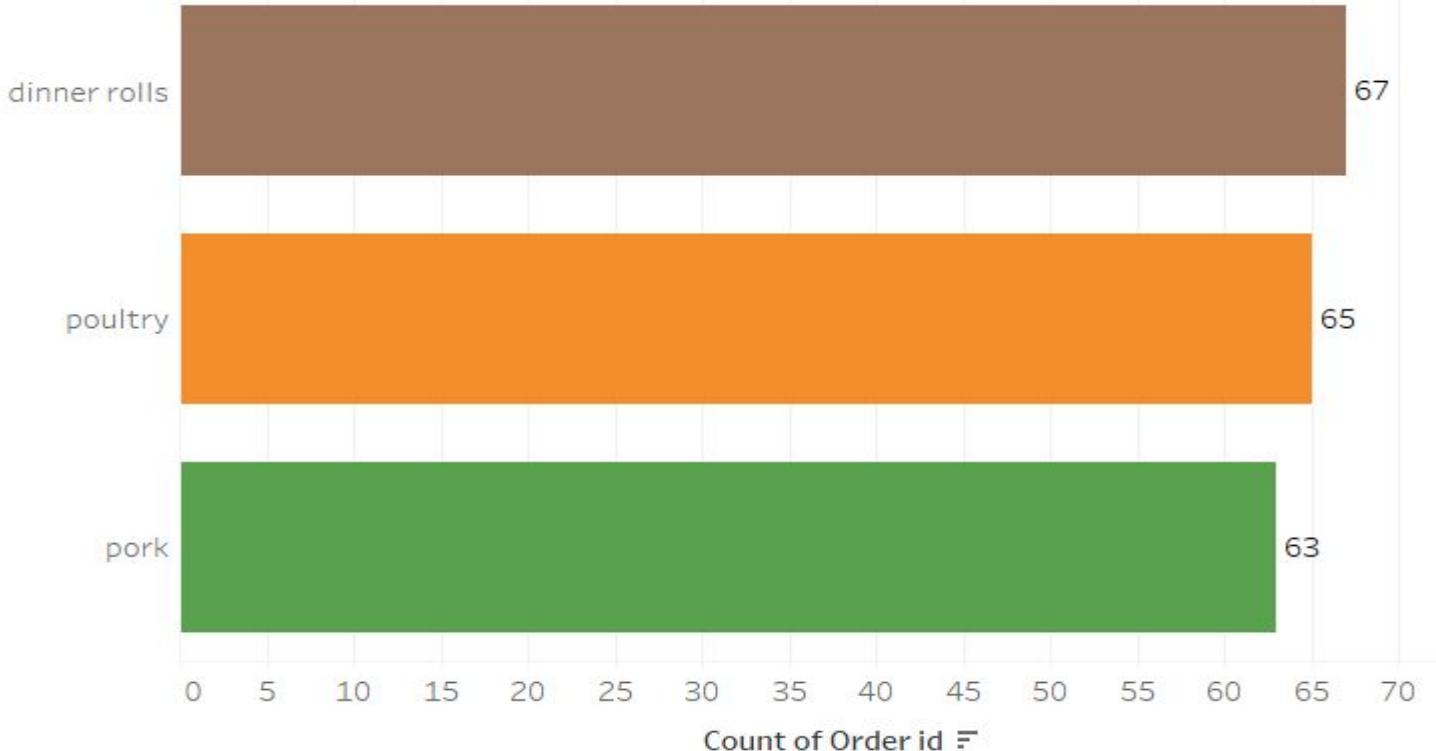
We can see poultry, cereals and soda are highly sold products over the years

yearly products



## top products in 2020

Product =



# Count of Products sold in 2020 January and February

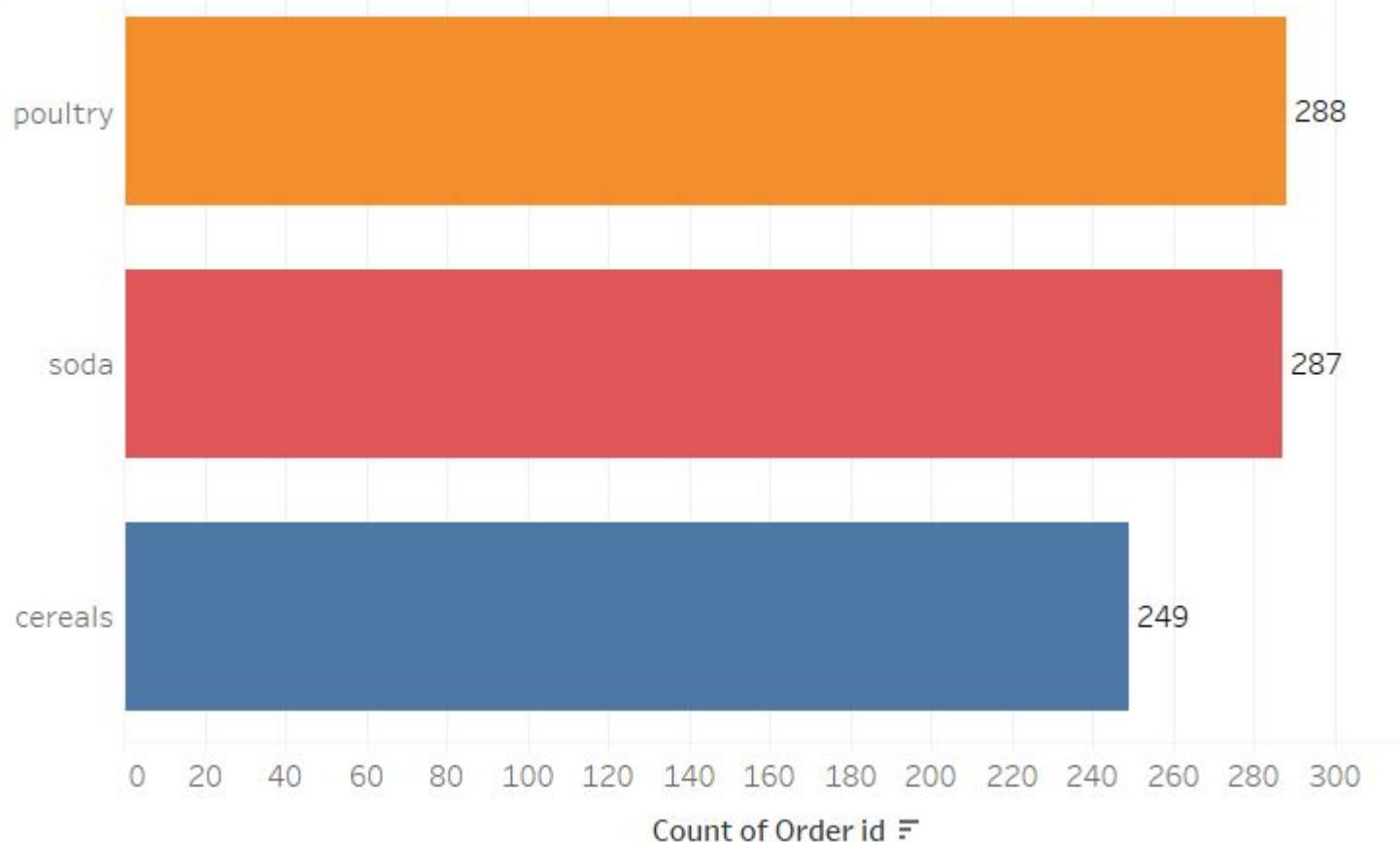
Top 3 products sold in 2020 January and February are dinner role, poultry and pork

# Count of Products sold in 2019

Poultry, Soda and Cereal are the top 3 Products sold in 2019

top product in 2019

Prod..



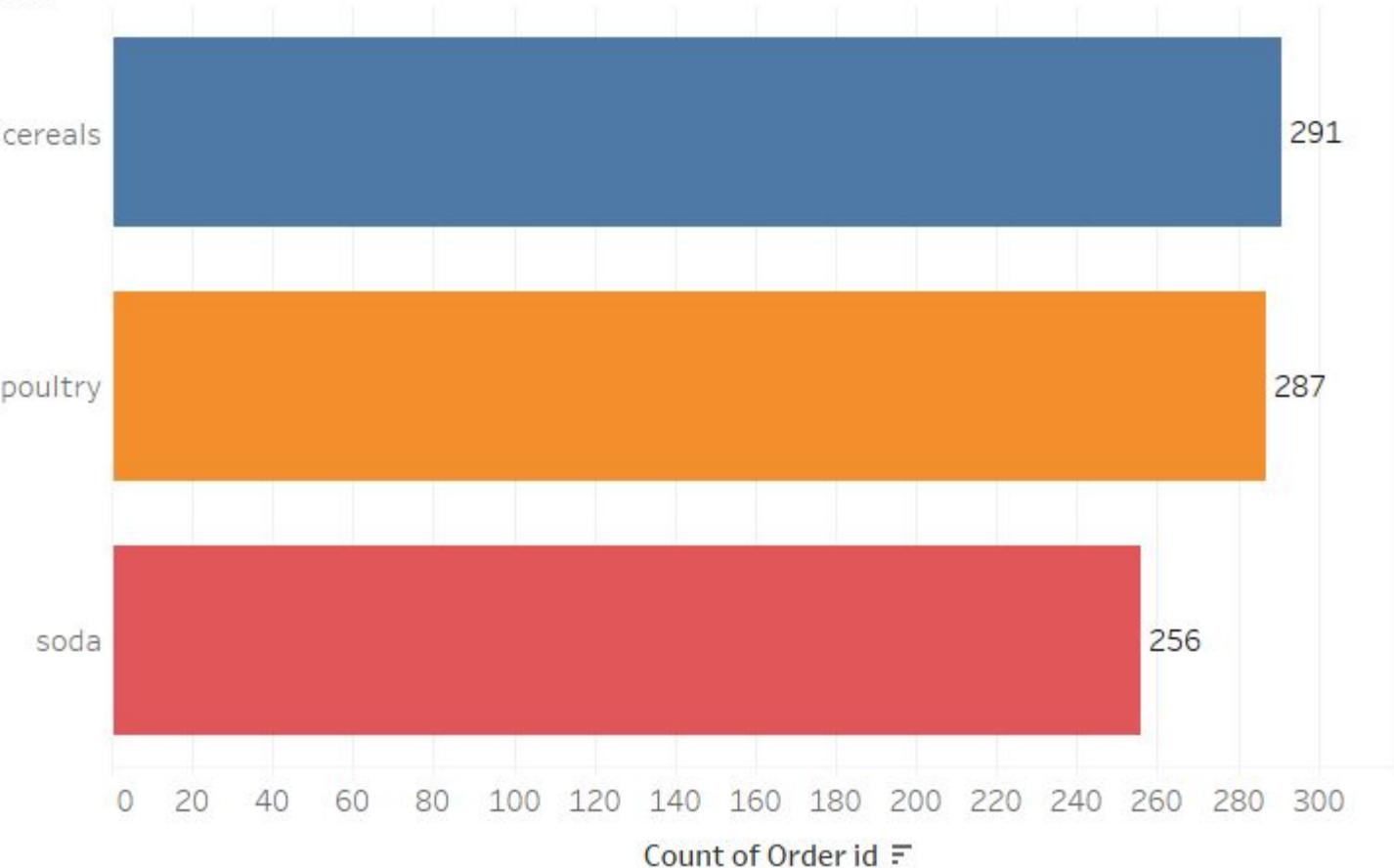
# Count of Products sold in 2018

---

Cereals, Poultry and Soda are the top 3 Products sold in 2018

top products in 2018

Prod..





# Summary as per EDA

1. **Top Products:**
  - **2018 & 2019:** Cereals, Poultry, Soda.
  - **Jan & Feb 2020:** Shift to Dinner Rolls, Poultry, Pork.
  - **Non-Eatables:** Highest sales in **Soap** and **Toilet Paper**; lowest in **Hand Soap**.
2. **Sales by Day:**
  - Highest sales on **Sundays**; lowest on **Mondays**.
  - Consistent weekday activity, peaking on **Wednesday to Friday**.
3. **Monthly Trends:**
  - **2018:** Highest in **January**, lowest in **February**.
  - **2019:** Peak in **March**, low in **January**.
  - Limited data for **2020** shows potential decline.
4. **Unique Orders:**
  - Downward trend in unique order IDs from **Q1 to Q3**.
  - Highest unique orders in **January, February, and May**; lowest in **June**.
5. **Evolving Preferences:**
  - **2018:** Everyday items like **Cereals** and **Poultry**.
  - **2019:** Mix of food and cleaning products, such as **Soda** and **Dishwashing Liquid**.
  - **2020:** Shift towards indulgent foods like **Ice Cream** and **Beef**.

# Recommendation as per EDA:

## Stock Key Products:

- Focus on promoting and stocking **Poultry**, **Soda**, and **Cereals**, as they are consistently top sellers.

## Non-Eatables:

- Increase inventory of **Soap** and **Toilet Paper**, the highest-selling non-eatable products.

## Address Hand Soap Sales:

- Evaluate reasons for low sales of **Hand Soap** and implement strategies to boost its sales.

## Sunday Promotions:

- Schedule promotions and offers on **Sundays** to capitalize on peak sales days.

## February Campaigns:

- Launch marketing campaigns and discounts in **February** to drive sales during historically low periods.

## January and March Promotions:

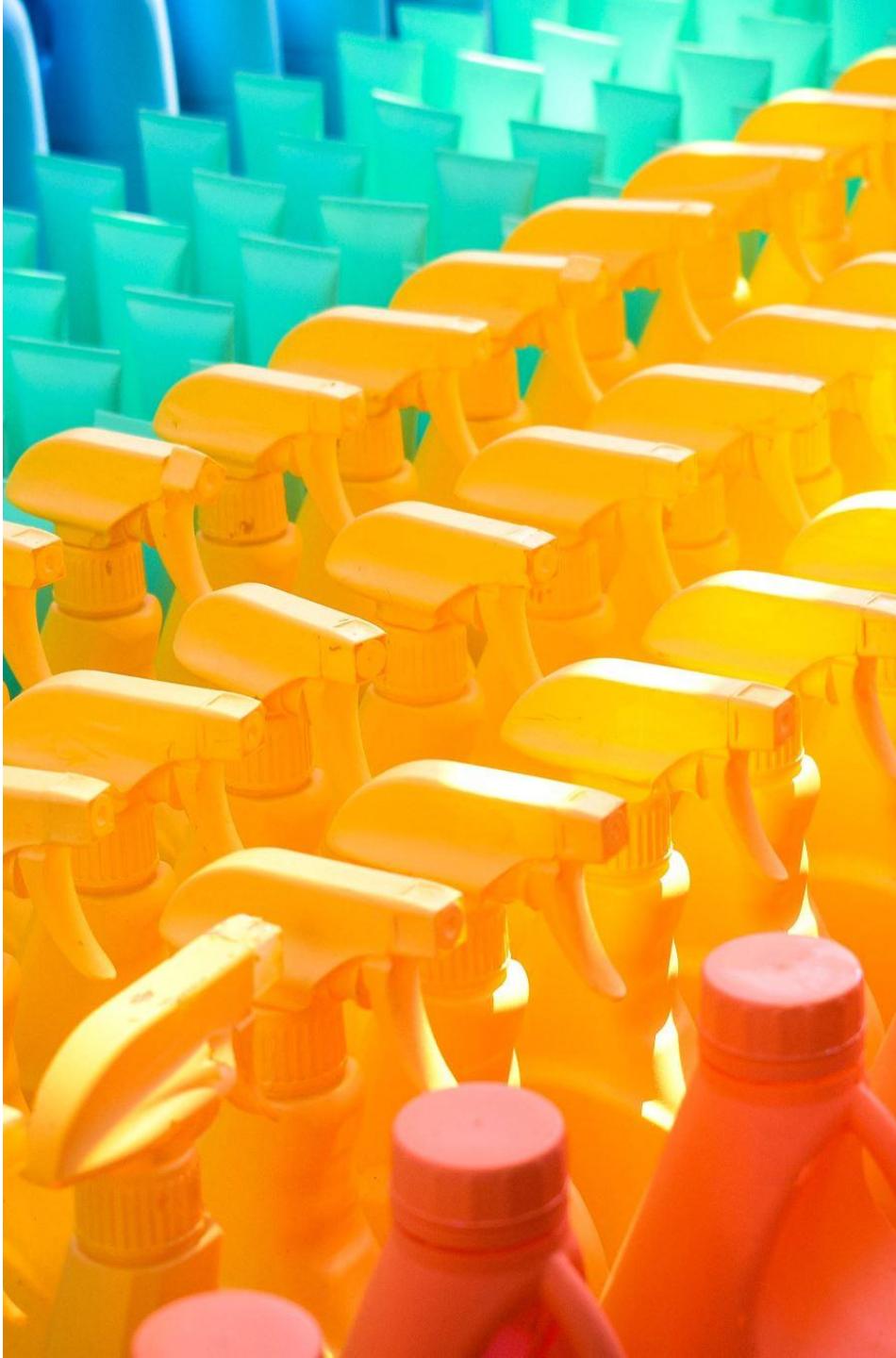
- Plan targeted promotions in **January** and **March** to enhance sales during high-demand months.

## Maintain Q2 Consistency:

- Keep Q2 stock levels consistent with previous years to sustain sales.

## Data Awareness:

- Consider the limited data for **2020** when making sales and marketing decisions.





# Market Basket Analysis

- Market Basket Analysis Meaning
- MRA KNIME WorkFlow & Output Table

# Market Basket Analysis

## Definition:

Market Basket Analysis is a statistical method used to analyze customer purchase patterns, revealing associations between different products. This technique helps businesses understand frequently co-purchased items and the impact of buying habits on sales.

## Data Requirements:

To perform Market Basket Analysis, businesses need transactional data, including:

- **Customer ID**
- **Product ID**
- **Transaction Date**

This data is organized into a matrix to illustrate relationships between products.

## Association Rules:

Association rules identify the strength of relationships between products, using metrics such as:

- **Support:** Frequency of items co-occurring in transactions.
- **Confidence:** Probability that a customer purchasing one item will buy another.
- **Lift:** Measures the correlation between two items, indicating how much more likely they are to be purchased together than by chance.

## Applications:

Market Basket Analysis is applied across various industries:

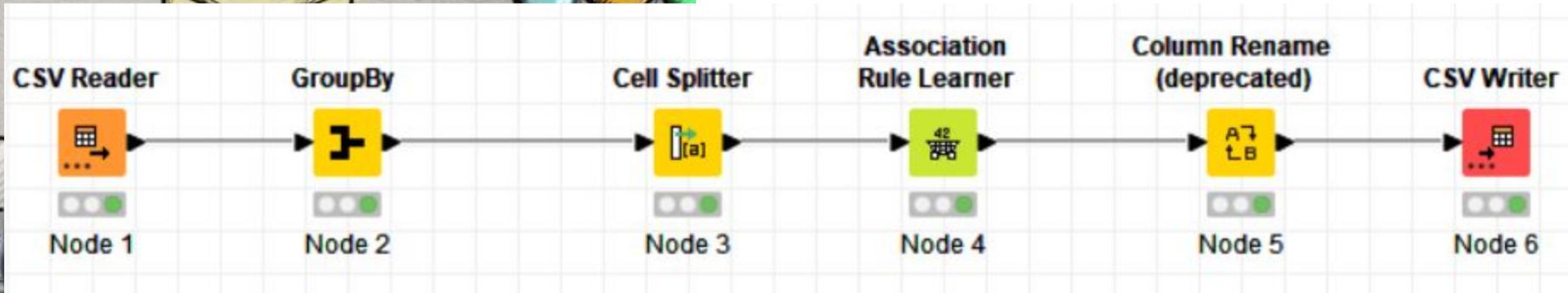
- **Retail:** Optimizes product placement and promotions.
- **E-commerce:** Personalizes product recommendations.
- **Marketing:** Develops targeted advertising campaigns.

## Benefits:

- Increases revenue by identifying cross-selling opportunities.
- Enhances customer satisfaction through personalized recommendations.
- Improves the overall shopping experience by tailoring promotions and product placements.



# MRA KNIME Workflow for Market Basket Analysis



# Associations Identified



*Let me see your identification.*



# Association Rules & its relevance.

## Insights into Customer Behavior:

Association rules highlight the relationships and co-occurrence patterns between items, offering valuable insights into customer purchasing behavior and preferences.

## Business Applications:

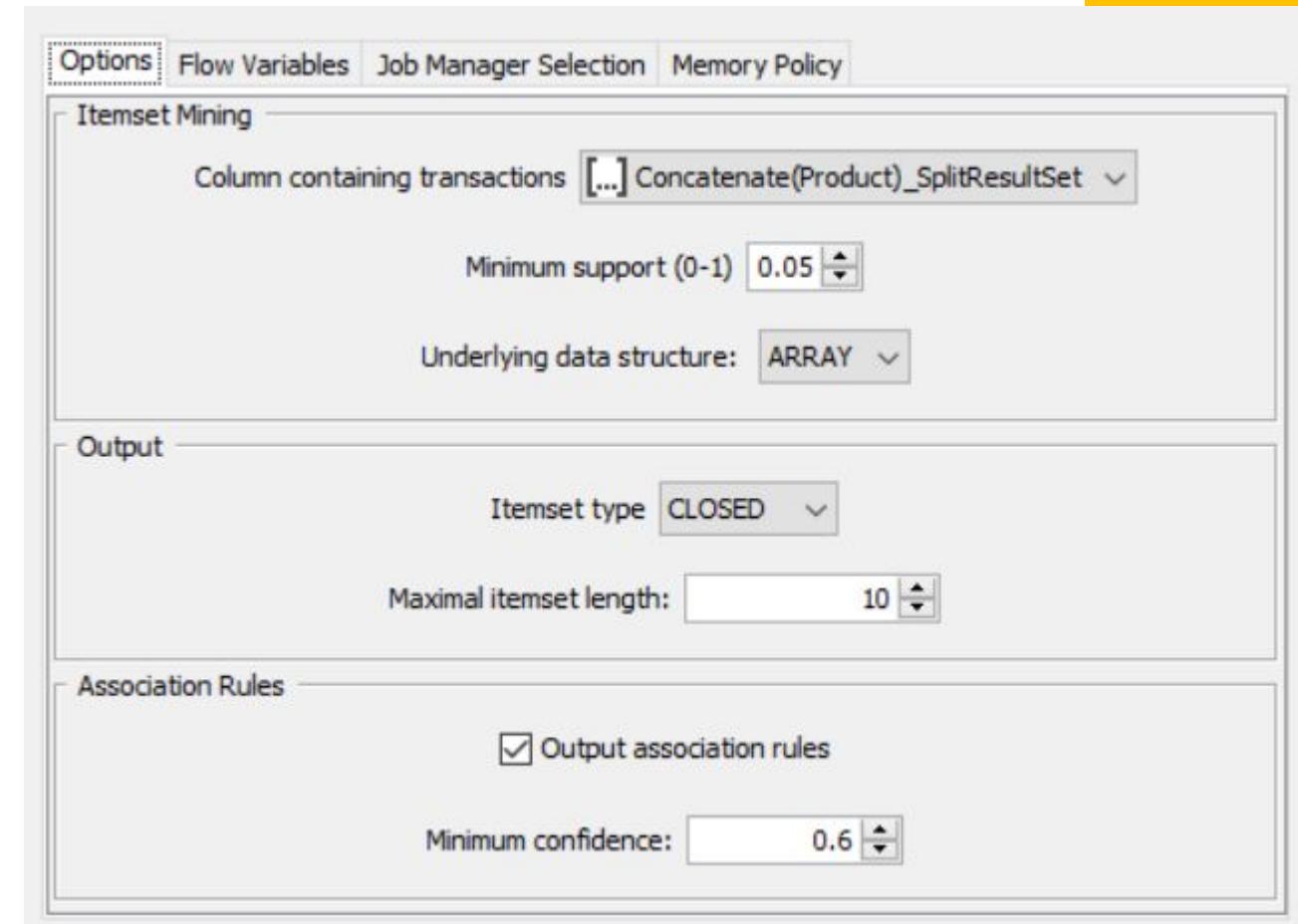
- **Product Placement:** Helps optimize the arrangement of products to encourage complementary purchases.
- **Targeted Marketing:** Guides the creation of focused marketing campaigns that resonate with customer interests.
- **Cross-Selling and Upselling:** Informs strategies to promote additional products to customers, enhancing their shopping experience.

## Benefits:

- Increases customer satisfaction by providing relevant product recommendations.
- Boosts revenue through improved sales strategies based on purchasing patterns.

# Threshold values of Support and Confidence

- The threshold for Minimum Support is set at 0.05, while the Minimum Confidence threshold is established at 0.6. These parameters help us identify which association rules are both significant and reliable. By applying these thresholds, we can exclude less relevant rules and concentrate on those with strong support and confidence levels, ensuring that our analysis yields valuable insights for informed decision-making.



# Output head Table

Row ID	D Support	D Confidence	D Lift	S Recommended_item	S Recommended_with	S Items_list
rule0	0.05	0.64	1.7	juice	<---	[yogurt, toilet paper, aluminum foil]
rule1	0.05	0.62	1.645	juice	<---	[yogurt, poultry, aluminum foil]
rule2	0.05	0.613	1.616	coffee/tea	<---	[yogurt, cheeses, cereals]
rule3	0.05	0.6	1.424	poultry	<---	[dishwashing liquid/detergent, laundry detergent, mixes]
rule4	0.051	0.63	1.678	mixes	<---	[yogurt, poultry, aluminum foil]
rule5	0.051	0.611	1.66	sandwich bags	<---	[cheeses, bagels, cereals]
rule6	0.051	0.674	1.726	cheeses	<---	[bagels, cereals, sandwich bags]
rule7	0.051	0.617	1.558	cereals	<---	[cheeses, bagels, sandwich bags]
rule8	0.051	0.63	1.621	dinner rolls	<---	[spaghetti sauce, poultry, cereals]
rule9	0.051	0.637	1.512	poultry	<---	[dinner rolls, spaghetti sauce, cereals]
rule10	0.051	0.604	1.589	milk	<---	[poultry, laundry detergent, cereals]
rule11	0.052	0.628	1.61	eggs	<---	[dinner rolls, poultry, soda]
rule12	0.052	0.641	1.649	dinner rolls	<---	[spaghetti sauce, poultry, ice cream]
rule13	0.052	0.686	1.628	poultry	<---	[dinner rolls, spaghetti sauce, ice cream]
rule14	0.052	0.628	1.614	dinner rolls	<---	[spaghetti sauce, poultry, juice]
rule15	0.052	0.602	1.429	poultry	<---	[dinner rolls, spaghetti sauce, juice]
rule16	0.052	0.634	1.627	eggs	<---	[paper towels, dinner rolls, pasta]
rule17	0.052	0.602	1.621	pasta	<---	[paper towels, eggs, dinner rolls]
rule18	0.054	0.642	1.651	dinner rolls	<---	[spaghetti sauce, poultry, laundry detergent]
rule19	0.054	0.656	1.556	poultry	<---	[dinner rolls, spaghetti sauce, laundry detergent]
rule20	0.055	0.624	1.565	ice cream	<---	[paper towels, eggs, pasta]
rule21	0.055	0.63	1.616	eggs	<---	[paper towels, ice cream, pasta]
rule22	0.055	0.643	1.731	pasta	<---	[paper towels, eggs, ice cream]
rule23	0.055	0.649	1.791	paper towels	<---	[eggs, ice cream, pasta]



## Market basket analysis, support, confidence, and lift values

In market basket analysis, support, confidence, and lift are key metrics used to evaluate the strength of associations between items in a transaction dataset.

**Support:** This metric represents the likelihood of seeing a particular combination of items in transactions. It's calculated by dividing the number of transactions containing both items by the total number of transactions. A high support value indicates that the itemset is frequently purchased together, suggesting it may be beneficial for promotion or joint placement.

**Confidence:** Confidence measures the probability that a transaction containing one item also includes another. It is determined by dividing the number of transactions with both items by the number of transactions containing the first item. A high confidence level suggests a strong likelihood that customers will buy these items together, making it useful for recommendations.

**Lift:** Lift assesses how much more frequently two items are purchased together than would be expected if they were independent. It is calculated by dividing the support of the itemset by the product of the individual supports of each item. A lift value greater than 1 indicates a positive association, suggesting that the items are likely to be bought together, while a value less than 1 implies a negative relationship. High lift values are particularly useful for cross-selling and bundling strategies.

## **4. Give Suggestion of Possible Combos with Lucrative Offers**

- Write recommendations

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



Recommendation in  
the next slide





1. **"Buy Two Get One Free" on Key Items:** Encourage bulk purchases by offering a "Buy Two Get One Free" promotion on yogurt, poultry, and aluminum foil.
2. **Discounted Combo Deal:** Create a special combo where customers can buy cereals, bagels, and sandwich bags at a discounted price, enhancing value.
3. **Mix and Match Discounts:** Provide discounts on mixes when purchased alongside yogurt, poultry, or aluminum foil to encourage complementary buying.
4. **Dinner Roll Promotions:** Offer a discount on dinner rolls when purchased with spaghetti sauce or poultry, creating meal pairings.
5. **Paper Products Bundle:** Introduce a "Paper Products Bundle" that includes paper towels, toilet paper, and tissues at a special price, encouraging bulk buying of essentials.
6. **Combo Deal with Juice:** Create a deal where customers can buy yogurt, poultry, and aluminum foil along with juice at a discounted rate or include an additional item.
7. **Limited-Time Offers:** Launch a limited-time promotion where purchasing poultry, laundry detergent, and mixes results in a percentage savings.
8. **Loyalty Program:** Implement a loyalty program where frequent buyers of recommended items or participants in the suggested combos earn rewards or exclusive discounts, fostering customer retention.

## Promotion Strategies

To maximize the effectiveness of these offers:

- Utilize in-store signage, advertisements, and social media campaigns to raise awareness.
- Highlight the value and savings customers receive by purchasing these combos.

These recommendations leverage market basket analysis insights to increase sales, enhance customer satisfaction, and encourage exploration of additional products.

# What we Learned?:

---

- The analysis revealed key product associations that can enhance the grocery store's strategy for product placement and promotions. Frequently purchased items include yogurt, poultry, aluminum foil, cheeses, cereals, and dinner rolls.
- Some surprising associations, like poultry with dishwashing liquid, laundry detergent, and mixes, highlight unique customer preferences.
- To leverage these insights, the store could implement promotions such as "buy two get one free" on selected items, encouraging larger purchases. Additionally, positioning complementary products near each other can increase the likelihood of additional sales.
- Overall, market basket analysis provides valuable insights into customer behavior, enabling informed decisions regarding product placement and promotional strategies.



# Tableau Links For Market and Retail analysis Projects EDA:

PART-A : RFM analysis link:

[https://public.tableau.com/views/MRA\\_Project\\_Part-A\\_Automobile\\_EDA/Bivariate-SalesbyQuarters?:language=en-GB&publish=yes&:sid=&:redirect=auth&:display\\_count=n&:origin=viz\\_share\\_link](https://public.tableau.com/views/MRA_Project_Part-A_Automobile_EDA/Bivariate-SalesbyQuarters?:language=en-GB&publish=yes&:sid=&:redirect=auth&:display_count=n&:origin=viz_share_link)

PART-B: Market Basket Analysis link:

[https://public.tableau.com/views/MRA\\_Project\\_Part-B\\_Grocery\\_EDA/Story1?:language=en-GB&publish=yes&:sid=&:redirect=auth&:display\\_count=n&:origin=viz\\_share\\_link](https://public.tableau.com/views/MRA_Project_Part-B_Grocery_EDA/Story1?:language=en-GB&publish=yes&:sid=&:redirect=auth&:display_count=n&:origin=viz_share_link)



**Let's Check It Out.**

Thank You

