

Title

Objective

Data Description

Exploratory Data Analysis

RFM Analysis

Insights

Business recommendation

Part – A

RFM Analysis

Title

Objective

Data Description

Exploratory Data Analysis

RFM Analysis

Insights

Business recommendation

Business Context

An **automobile parts manufacturing company** has been actively selling products to a diverse range of customers for the past three years. Despite its growth, the company lacks the in-house expertise to derive actionable insights from its transaction data. As a result, they wish to uncover hidden patterns and trends in their customer transactions. By analyzing this data, the company aims to better understand customer behavior, improve customer segmentation, and implement targeted marketing strategies. These insights will help the company not only enhance customer satisfaction but also drive revenue growth by offering more personalized and efficient services.

Objective

The primary objective of this analysis is to leverage data science techniques to:

1. **Identify** underlying patterns in customer purchasing behavior.
2. **Segment** customers based on their transactional data.
3. **Provide** actionable insights to optimize the company's marketing efforts.
4. **Recommend** personalized marketing strategies for each customer segment to maximize sales and customer retention.

Your role as a Business Analyst is to use the provided dataset to achieve these goals and present findings in a manner that can guide the company's decision-making.

Title

Objective

Data Description

Exploratory Data Analysis

RFM Analysis

Insights

Business recommendation

Data Description

The dataset provided contains three years of transactional data from the company, with each row representing a unique order. There are 20 columns and 2747 rows. Below is an explanation of the key attributes:

- **ORDERNUMBER:** Unique identifier for each order.
- **QUANTITYORDERED:** Number of items ordered in a specific transaction.
- **PRICEEACH:** Price per unit of the product in the order.
- **ORDERLINENUMBER:** Sequence number of the product in the order.
- **SALES:** Total sales value for the order.
- **ORDERDATE:** Date when the order was placed.
- **DAYS_SINCE_LASTORDER:** Number of days since the customer's previous order.
- **STATUS:** Current status of the order (e.g., Shipped, Disputed).
- **PRODUCTLINE:** Product category to which the item belongs (e.g., Motorcycles, Classic Cars).
- **MSRP:** Manufacturer's Suggested Retail Price for the product.
- **PRODUCTCODE:** Unique identifier for the product.
- **CUSTOMERNAME:** Name of the customer placing the order.
- **PHONE:** Customer's contact phone number.
- **ADDRESSLINE1:** Customer's primary address.
- **CITY:** City of the customer's address.
- **POSTALCODE:** Postal code of the customer's address.
- **COUNTRY:** Country of the customer's address.
- **CONTACTLASTNAME:** Last name of the customer's contact person.
- **CONTACTFIRSTNAME:** First name of the customer's contact person.
- **DEALSIZE:** Size category of the transaction (e.g., Small, Medium, Large).

Title

Objective

Data Description

Exploratory Data Analysis

RFM Analysis

Insights

Business recommendation

What is Exploratory Data Analysis (EDA)?

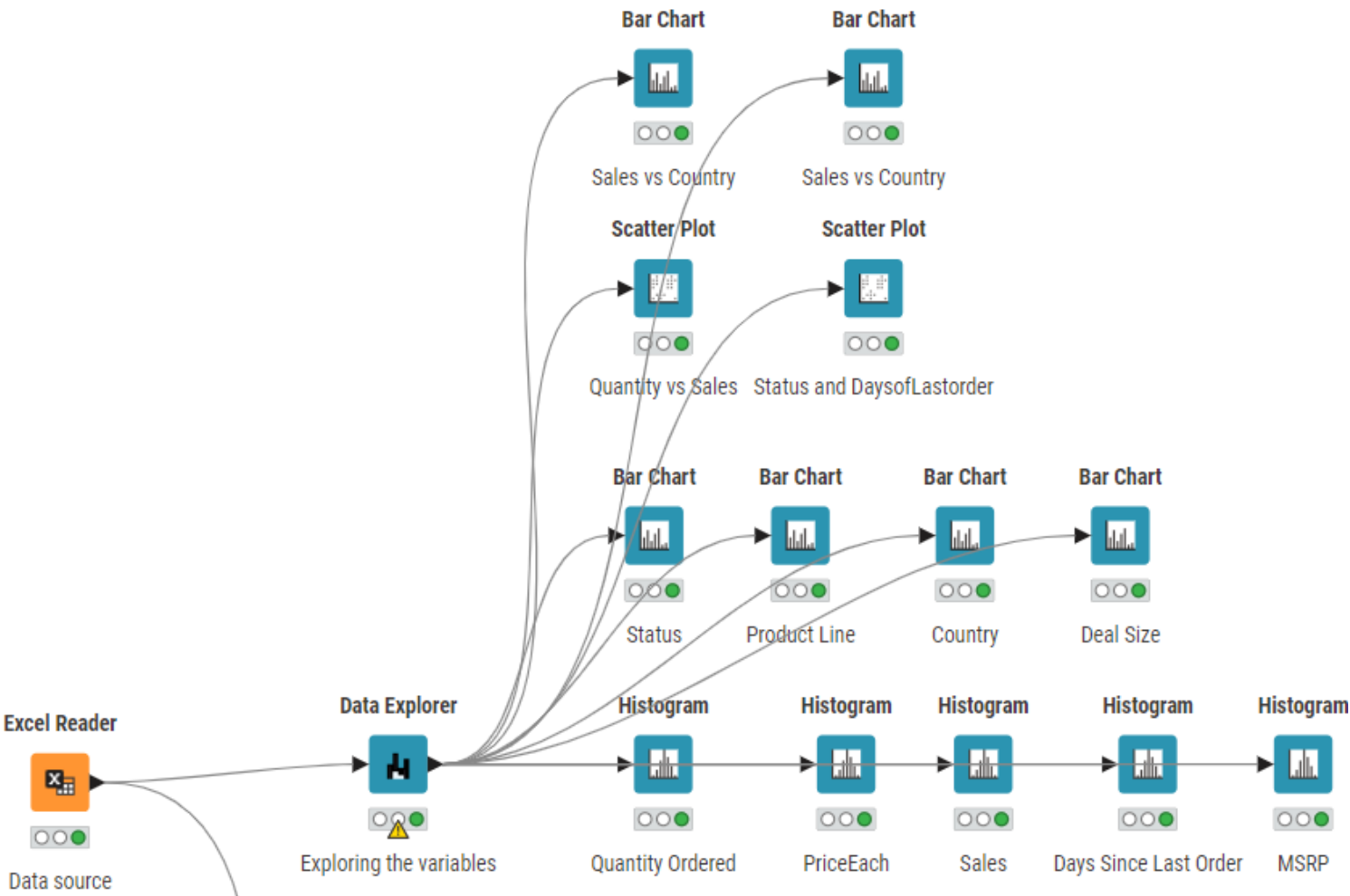
Exploratory Data Analysis (EDA) is a crucial step in the data analysis process that focuses on summarizing, visualizing, and understanding the structure of a dataset. The primary objective of EDA is to uncover patterns, detect anomalies, test hypotheses, and validate assumptions using statistical and graphical techniques. EDA lays the foundation for informed decision-making by revealing insights that guide data preprocessing, feature selection, and model development.

Why is EDA Important?

- **Data Quality Assessment:** Identifies missing values, outliers, and inconsistencies that may impact analysis.
- **Pattern Discovery:** Uncovers hidden relationships, trends, and correlations within the data.
- **Hypothesis Generation:** Helps develop data-driven hypotheses for deeper analysis and modeling.
- **Informed Preprocessing:** Guides data cleaning, normalization, and transformation strategies.
- **Model Readiness:** Ensures the dataset is suitable for modeling by understanding distributions and feature relevance.
- **Better Decision-Making:** Provides actionable insights that influence business strategies and objectives.

Exploratory Data Analysis (EDA) in KNIME

- Title
- Objective
- Data Description
- Exploratory Data Analysis
- RFM Analysis
- Insights
- Business recommendation

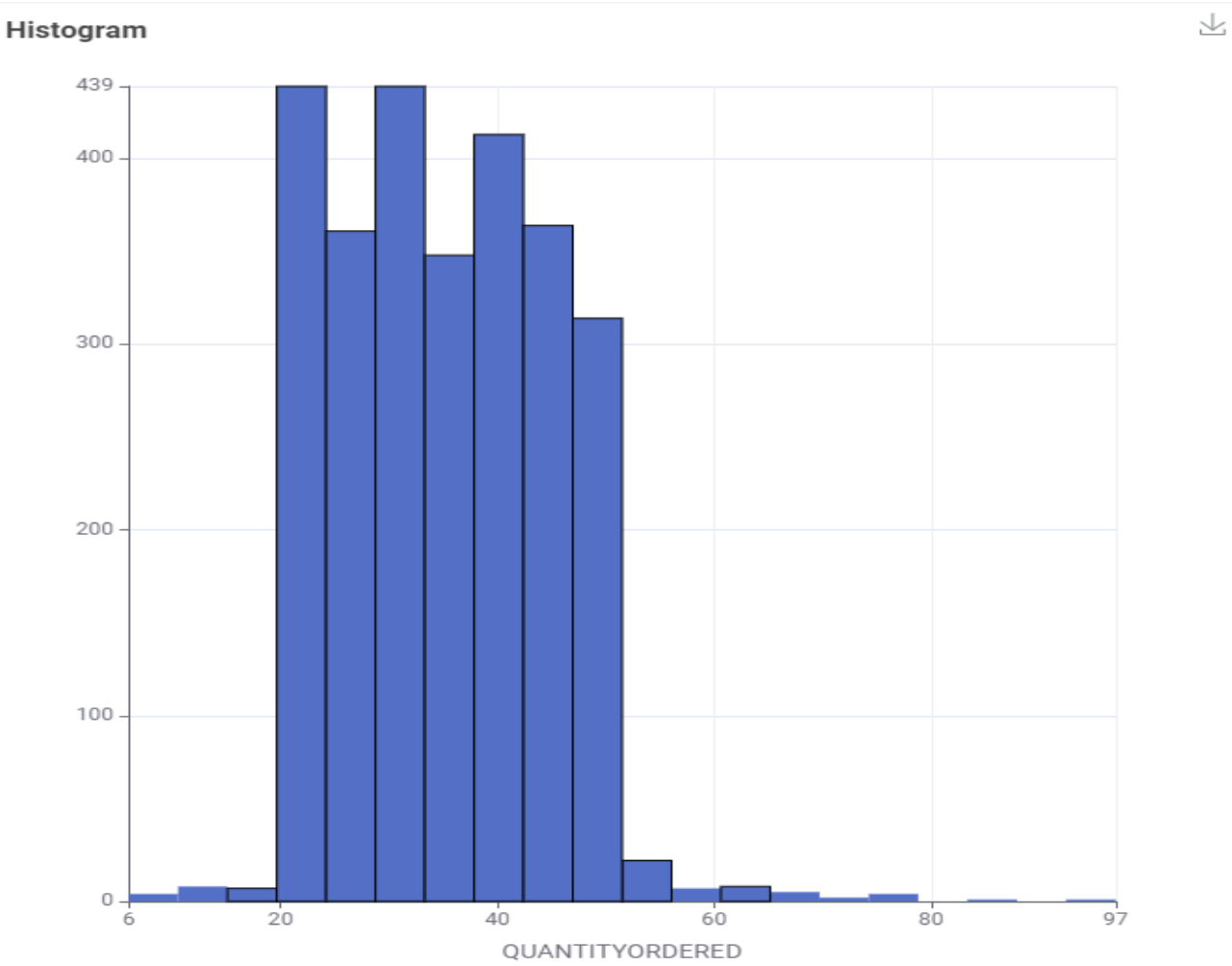


- Title
- Objective
- Data Description
- Exploratory Data Analysis
 - Univariate Analysis
 - Bivariate Analysis
- Conclusion on EDA
- RFM Analysis
- Insights
- Business recommendation

Univariate Analysis

Quantity Ordered

Majority of quantities can be observed between the range of 20 to 50. The minimum is 6 and the maximum is 97 and the average stands at 35.



Title

Objective

Data Description

Exploratory Data Analysis

Univariate Analysis

Bivariate Analysis

Conclusion on EDA

RFM Analysis

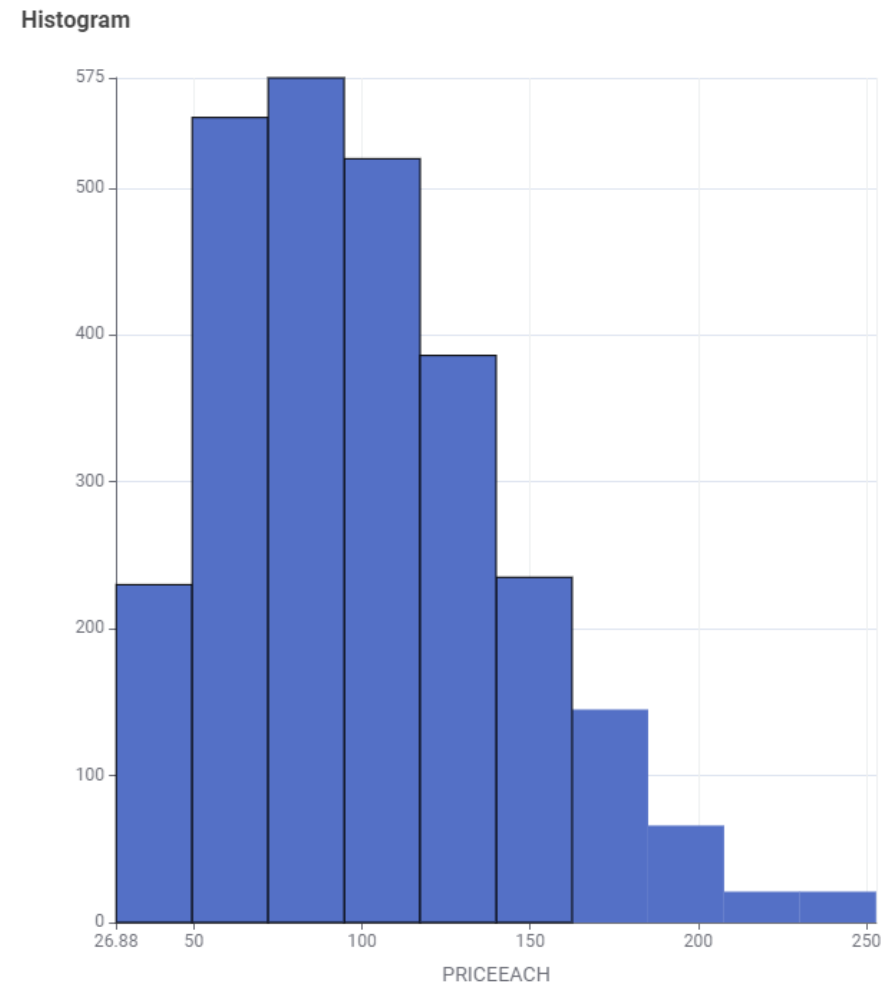
Insights

Business recommendation

Univariate Analysis

Price of Each

This shows the price of each item in an order. The majority of orders has the price range of 50 to 100. The minimum price is 26.88 and maximum is 252.87 where the average is 101.10.



Title

Objective

Data Description

Exploratory Data Analysis

Univariate Analysis

Bivariate Analysis

Conclusion on EDA

RFM Analysis

Insights

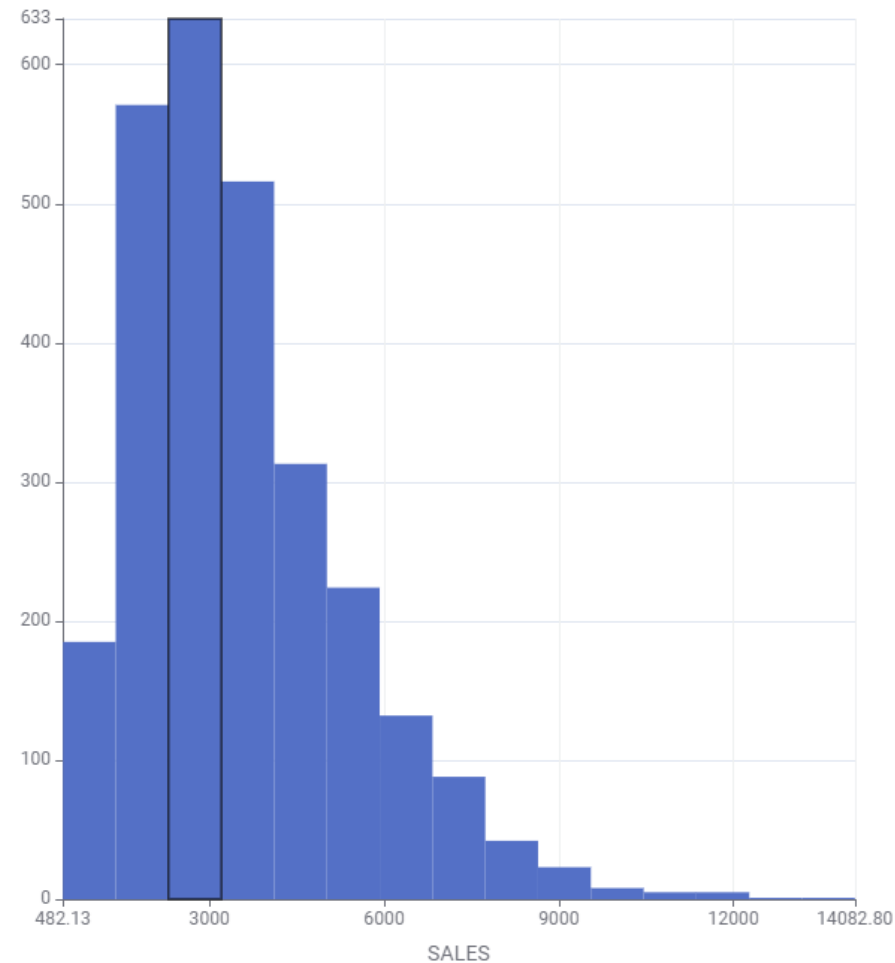
Business recommendation

Univariate Analysis

Sales

This sales variable is a left skewed distribution where the majority lies between 1388 to 4109. The minimum sale amount is 482.13 and the maximum is 14082.80 where the average sales stands at 3553.05.

Histogram



Title

Objective

Data Description

Exploratory Data Analysis

Univariate Analysis

Bivariate Analysis

Conclusion on EDA

RFM Analysis

Insights

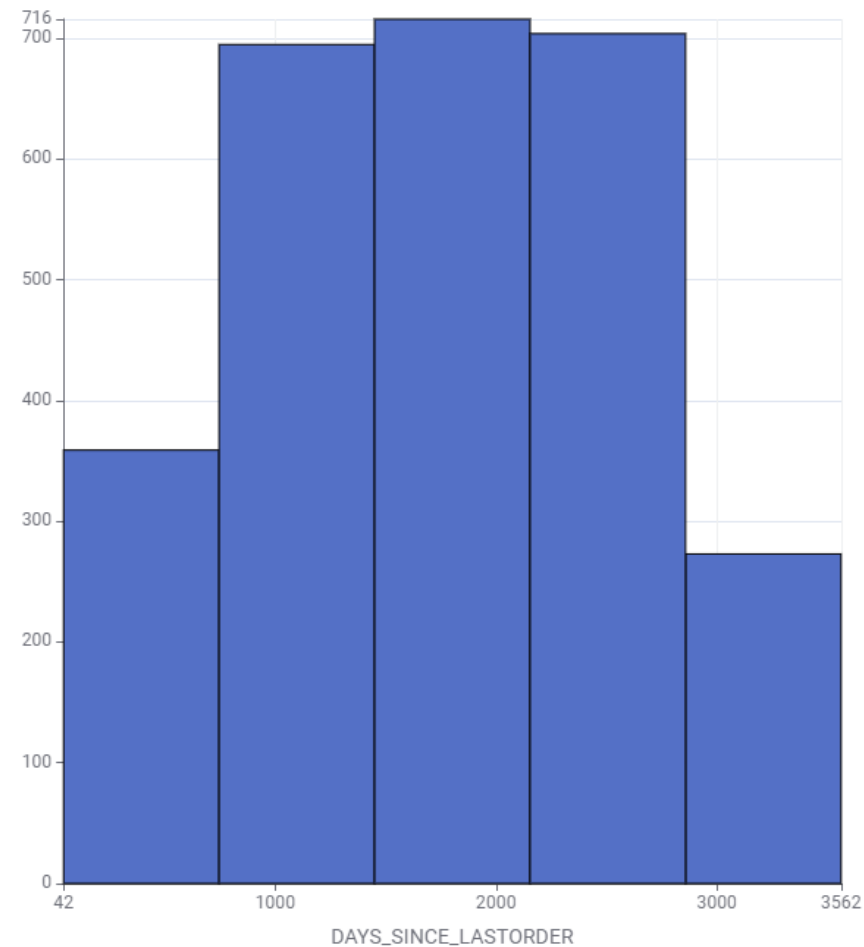
Business recommendation

Univariate Analysis

Days Since Last Order

Days since last order helps us get the recency of the customer. Majority of the customer recency lies between 746 days to 2858 days. The minimum days since last order is 42 and the maximum is 3562 days where the average days stands at 1757 days.

Histogram of Days Since Last Order



Title

Objective

Data Description

Exploratory Data Analysis

Univariate Analysis

Bivariate Analysis

Conclusion on EDA

RFM Analysis

Insights

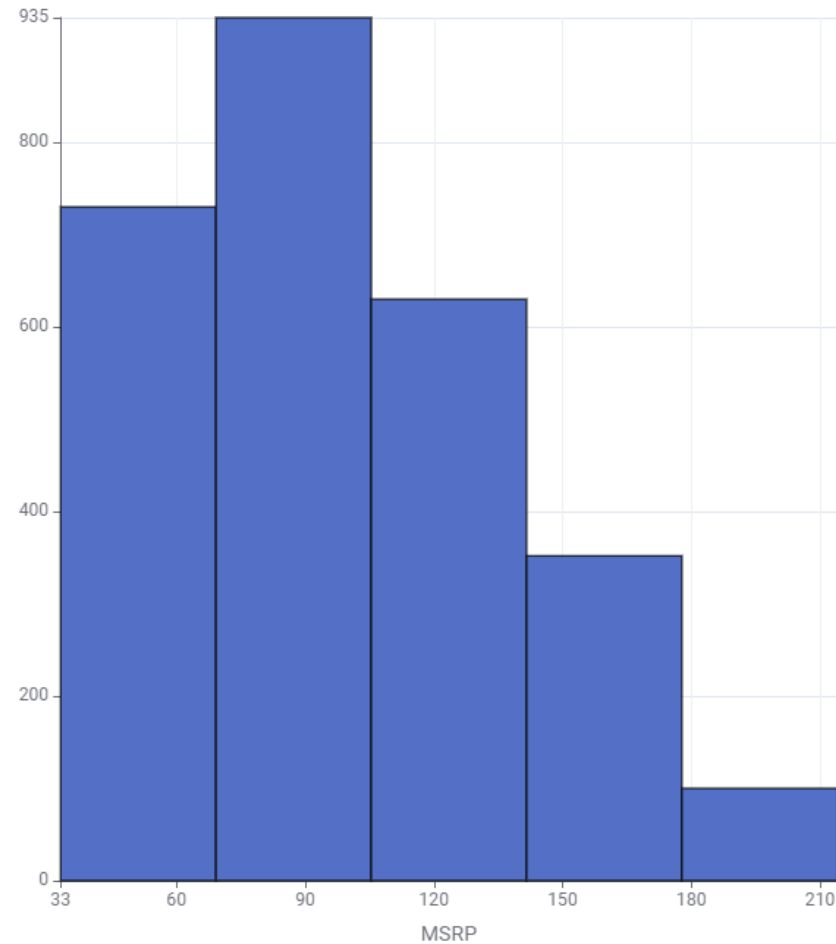
Business recommendation

Univariate Analysis

Manufacturer Suggested Retail Price (MSRP)

The majority of MSRP lies between 70 to 105. It follows a left skewed distribution. The minimum MSRP is 33 and the maximum goes till 214 where the average MSRP is 100.69.

Histogram of MSRP

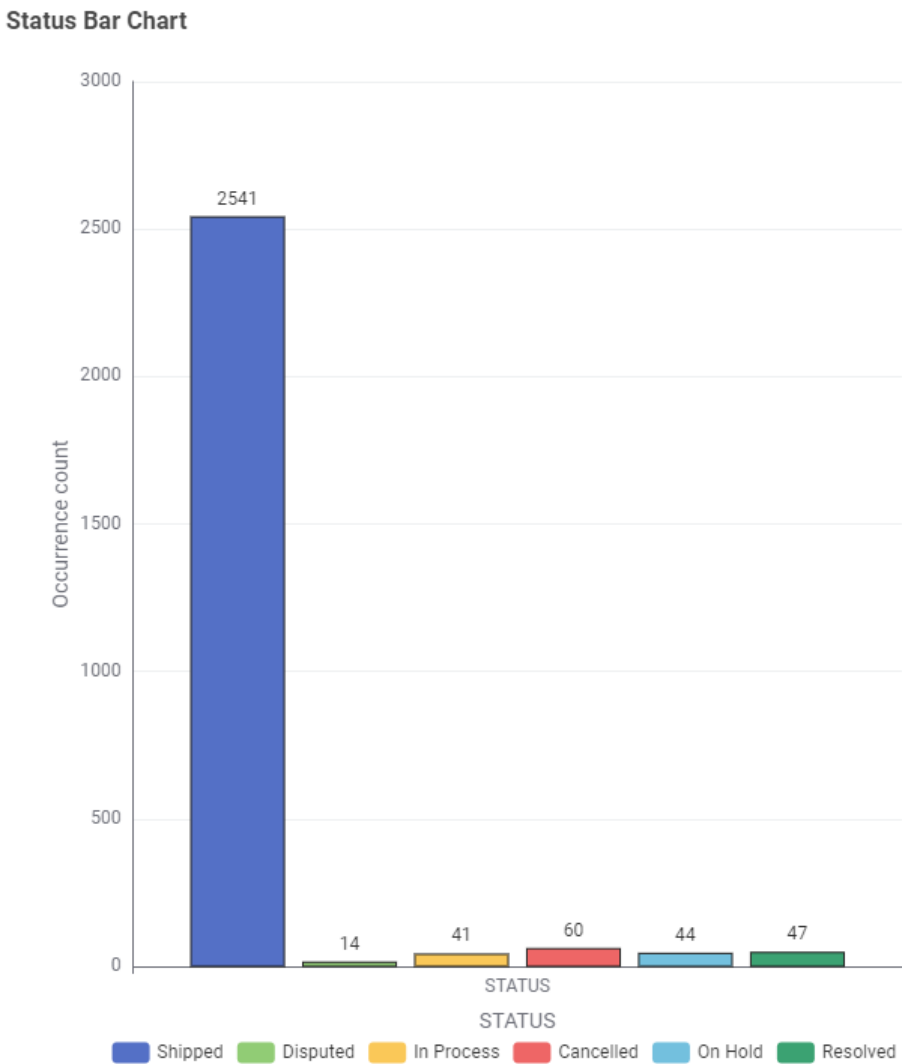


- Title
- Objective
- Data Description
- Exploratory Data Analysis
 - Univariate Analysis
 - Bivariate Analysis
- Conclusion on EDA
- RFM Analysis
- Insights
- Business recommendation

Univariate Analysis

Status

The status indicates the status of the order. Most of the orders are on shipped status followed by cancelled and resolved.



Title

Objective

Data Description

Exploratory Data Analysis

Univariate Analysis

Bivariate Analysis

Conclusion on EDA

RFM Analysis

Insights

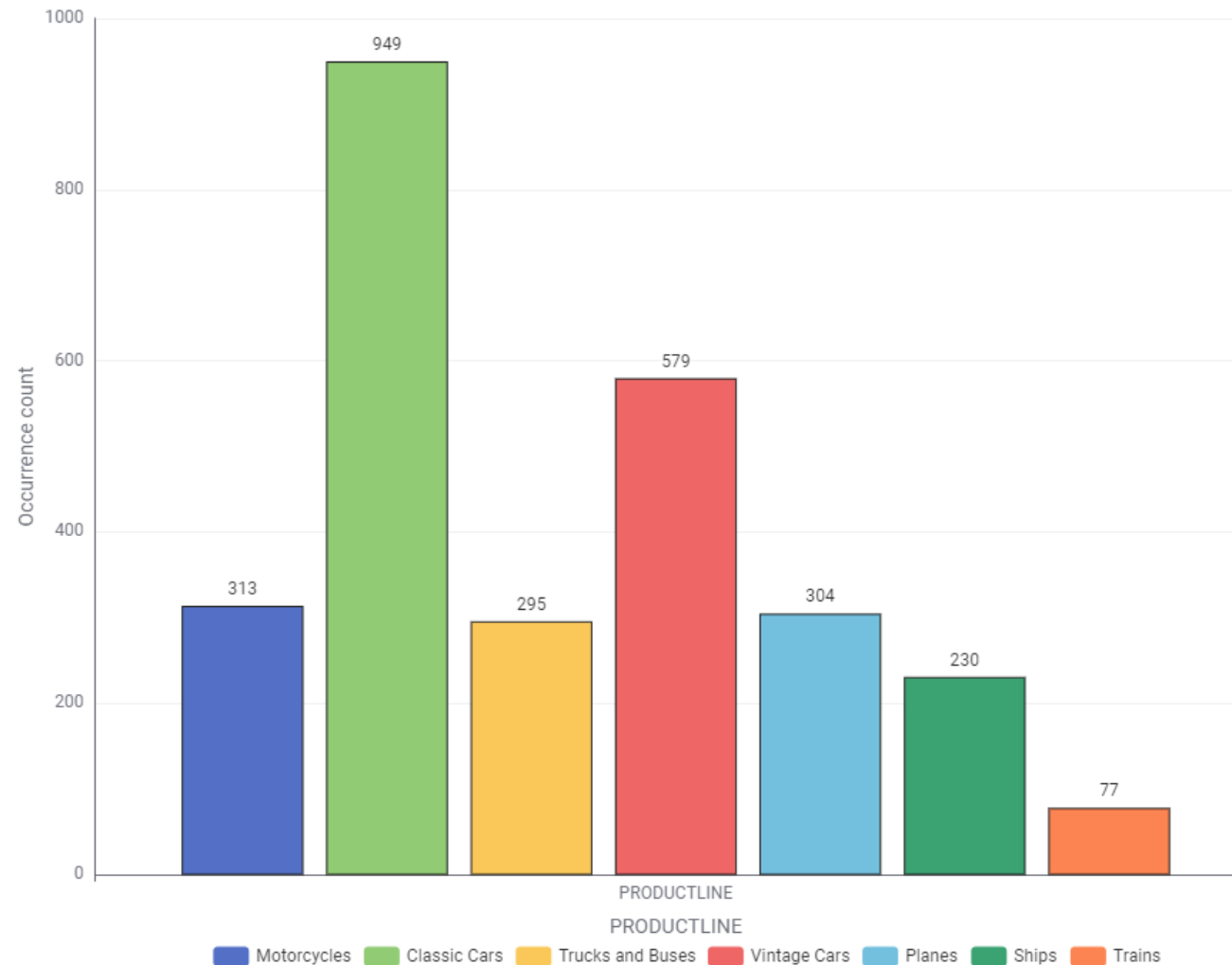
Business recommendation

Univariate Analysis

Product Line

We can see the parts of classic cars followed by vintage cars are sold most. The least is for trains and ships.

Product Line Bar Chart



Title

Objective

Data Description

Exploratory Data Analysis

Univariate Analysis

Bivariate Analysis

Conclusion on EDA

RFM Analysis

Insights

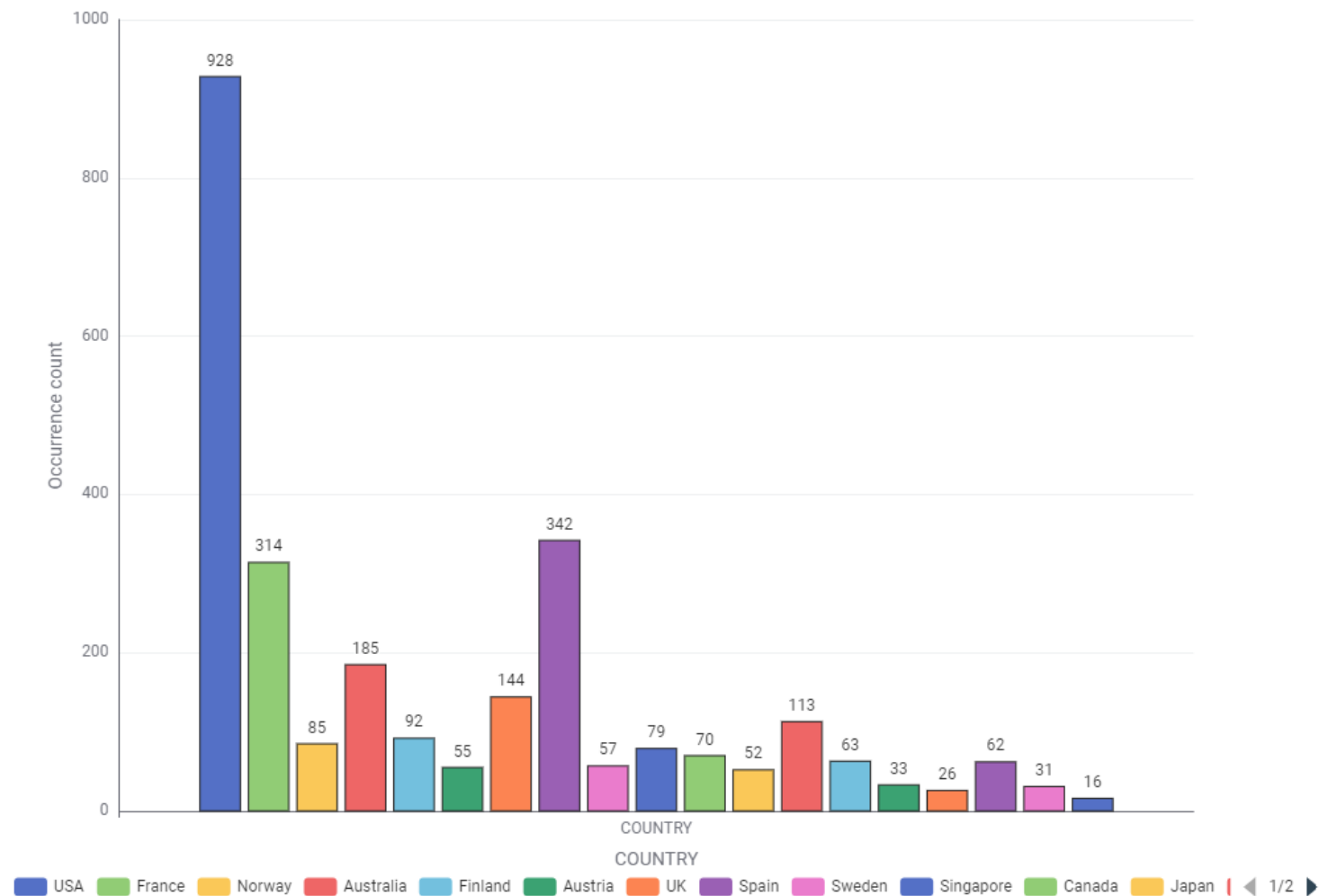
Business recommendation

Univariate Analysis

Country

Majority of the customer is located in USA followed by Spain. The least customers are in Ireland.

Country Bar Chart



Title

Objective

Data Description

Exploratory Data Analysis

Univariate Analysis

Bivariate Analysis

Conclusion on EDA

RFM Analysis

Insights

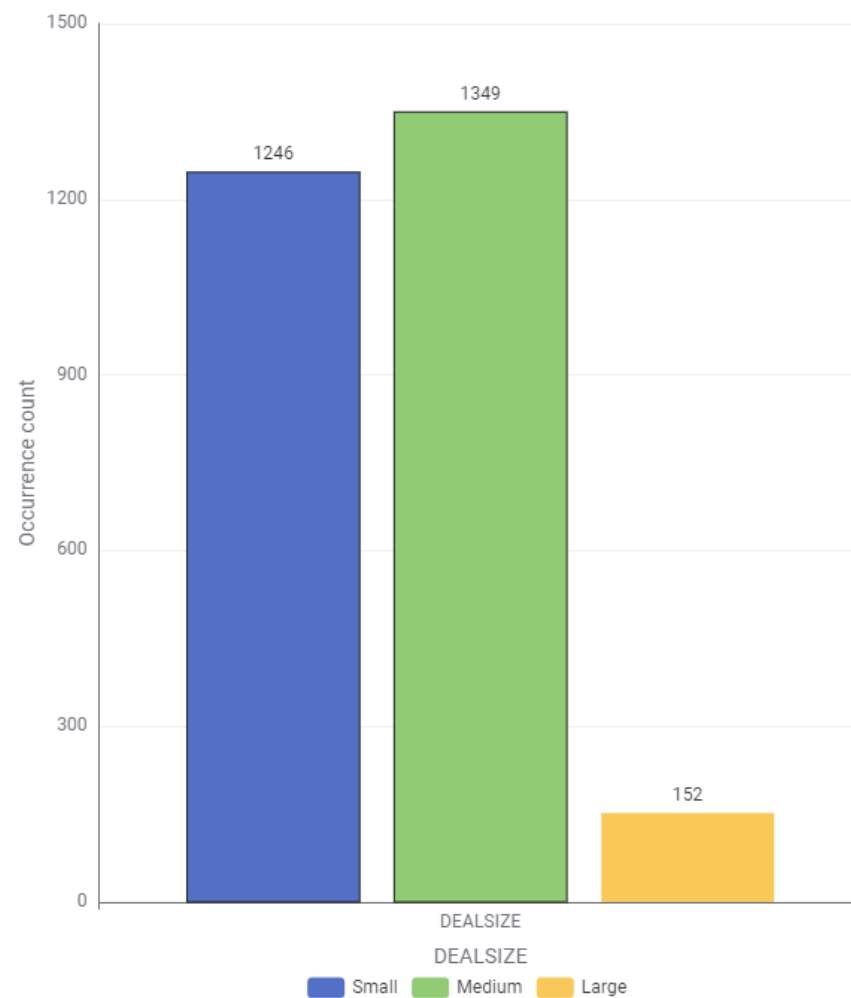
Business recommendation

Univariate Analysis

Deal Size

Majority of the orders fall under the medium deal size followed by small. Very few orders fall under Large deal size category.

Deal Size Bar Chart



Title

Objective

Data Description

Exploratory Data Analysis

Univariate Analysis

Bivariate Analysis

Conclusion on EDA

RFM Analysis

Insights

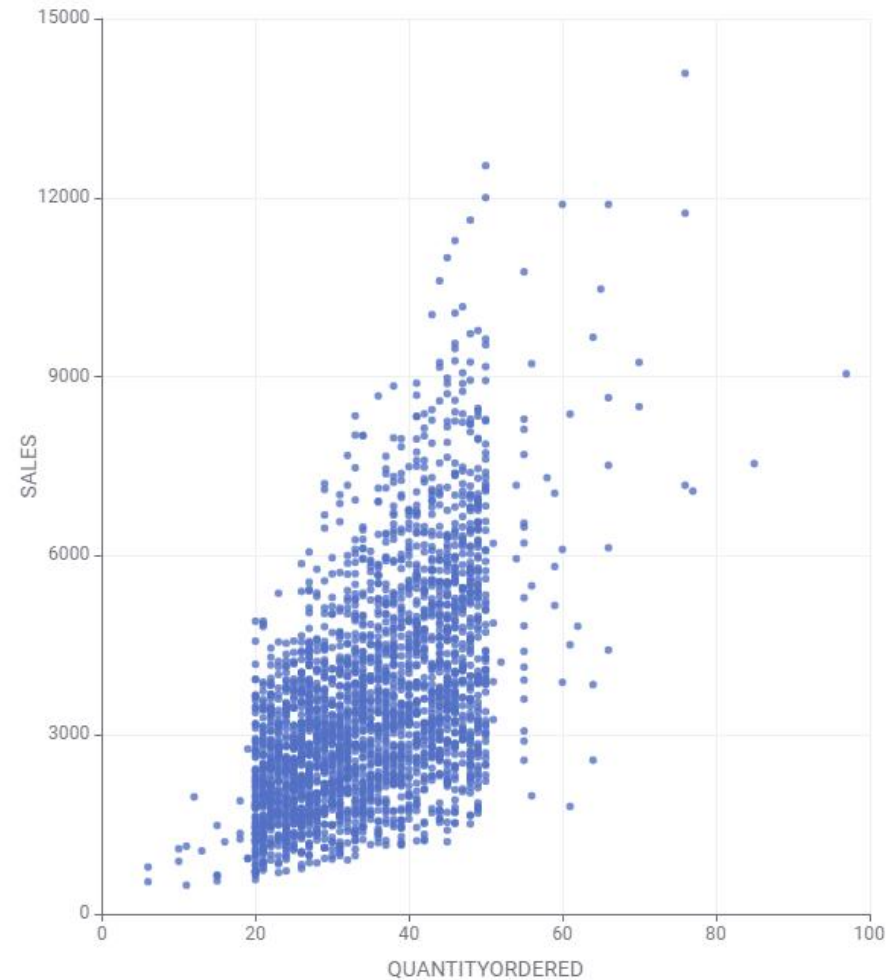
Business recommendation

Bivariate Analysis

Quantity Ordered vs Sales

We could see a linear positive correlation between these two variables where if the sales is more then the quantity ordered is also high.

Scatter Plot between Sales and Quantity Ordered

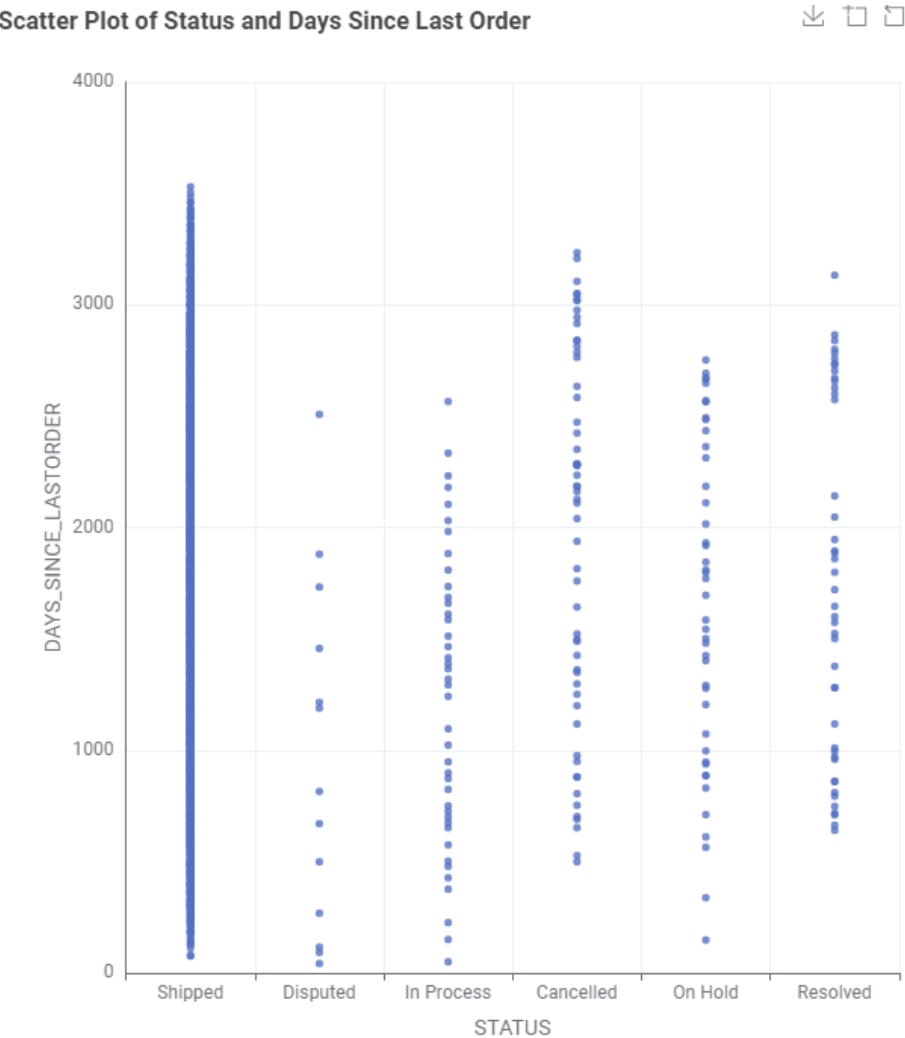


- Title
- Objective
- Data Description
- Exploratory Data Analysis
 - Univariate Analysis
 - Bivariate Analysis
- Conclusion on EDA
- RFM Analysis
- Insights
- Business recommendation

Bivariate Analysis

Status vs Days Since Last Order

We are checking to see if there are any correlation between status and days since last order in order to determine the churn. However, we could see that there is no pattern found.

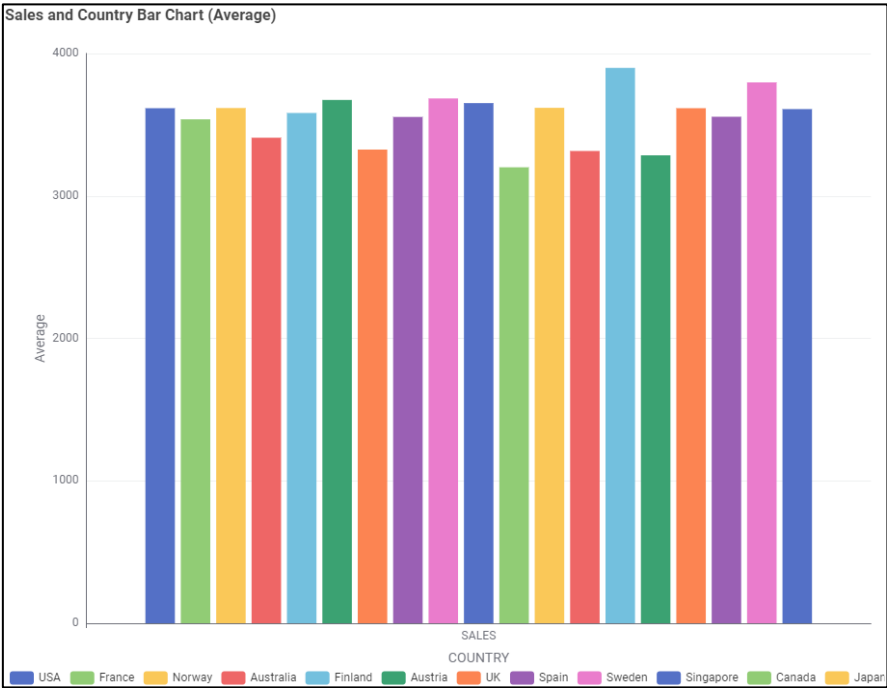
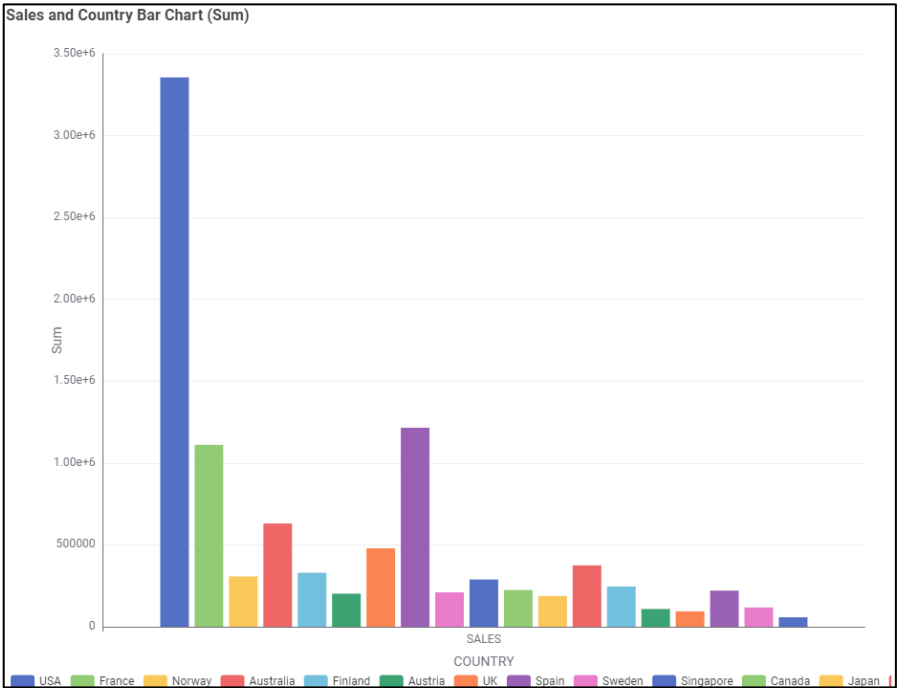


Bivariate Analysis

Sales vs Country

The highest sales in terms of sum is coming from USA followed by Spain. However, if we inspect the average sales from country, everything seems to be in a same range.

- Title
- Objective
- Data Description
- Exploratory Data Analysis
 - Univariate Analysis
 - Bivariate Analysis
- Conclusion on EDA
- RFM Analysis
- Insights
- Business recommendation



Title

Objective

Data Description

Exploratory Data Analysis

Univariate Analysis

Bivariate Analysis

Conclusion on EDA

RFM Analysis

Insights

Business recommendation

Conclusion

- **Quantity Ordered:**
 - Majority between 20 to 50 units.
 - Min: 6 | Max: 97 | Avg: 35
- **Price Each:**
 - Most orders priced between 50 to 100.
 - Min: 26.88 | Max: 252.87 | Avg: 101.10
- **Sales:**
 - Left-skewed distribution with most sales between 1,388 to 4,109.
 - Min: 482.13 | Max: 14,082.80 | Avg: 3,553.05
- **Recency (Days Since Last Order):**
 - Majority between 746 to 2,858 days.
 - Min: 42 | Max: 3,562 | Avg: 1,757 days.
- **MSRP (Manufacturer's Suggested Retail Price):**
 - Left-skewed, with most prices between 70 to 105.
 - Min: 33 | Max: 214 | Avg: 100.69
- **Order Status:**
 - Majority marked as Shipped, followed by Cancelled and Resolved
- **Product Line:**
 - Classic Cars and Vintage Cars lead sales. Trains and Ships are the least sold.
- **Geography:**
 - USA has the highest number of customers, followed by Spain.
 - Ireland has the least customer representation.
- **Deal Size:**
 - Medium deal size dominates, followed by Small.
 - Large deals are rare.
- **Correlations:**
 - Positive linear relationship between Sales and Quantity Ordered.
 - No significant pattern between Order Status and Recency (churn analysis)
- **Country-Wise Sales:**
 - USA leads in total sales, followed by Spain.
 - Average sales per country are within a similar range.

Title

Objective

Data Description

Exploratory Data Analysis

RFM Analysis

Insights

Business recommendation

What is RFM analysis?

RFM Analysis is a customer segmentation technique that evaluates customer behavior based on three key metrics:

- **Recency (R):** How recently a customer made a purchase.
- **Frequency (F):** How often a customer makes purchases.
- **Monetary (M):** How much a customer spends on purchases.

By analyzing these three dimensions, businesses can rank and segment customers to understand purchasing patterns and engagement levels.

Why is RFM analysis used?

- **Customer Segmentation:** Easily categorize customers into groups such as Loyal Customers, At-Risk, High-Value, and Churned.
- **Personalized Marketing:** Helps tailor marketing strategies to each segment, increasing campaign effectiveness.
- **Improved Retention:** Identifies at-risk customers, enabling proactive engagement to reduce churn.
- **Increased Sales:** Focuses efforts on high-value customers to maximize revenue potential.
- **Data-Driven Decisions:** Provides clear insights into customer behavior, helping optimize marketing strategies and resource allocation.

RFM Analysis is simple yet powerful, making it a go-to tool for businesses aiming to boost customer engagement, retention, and overall sales.

Title

Objective

Data Description

Exploratory Data Analysis

RFM Analysis

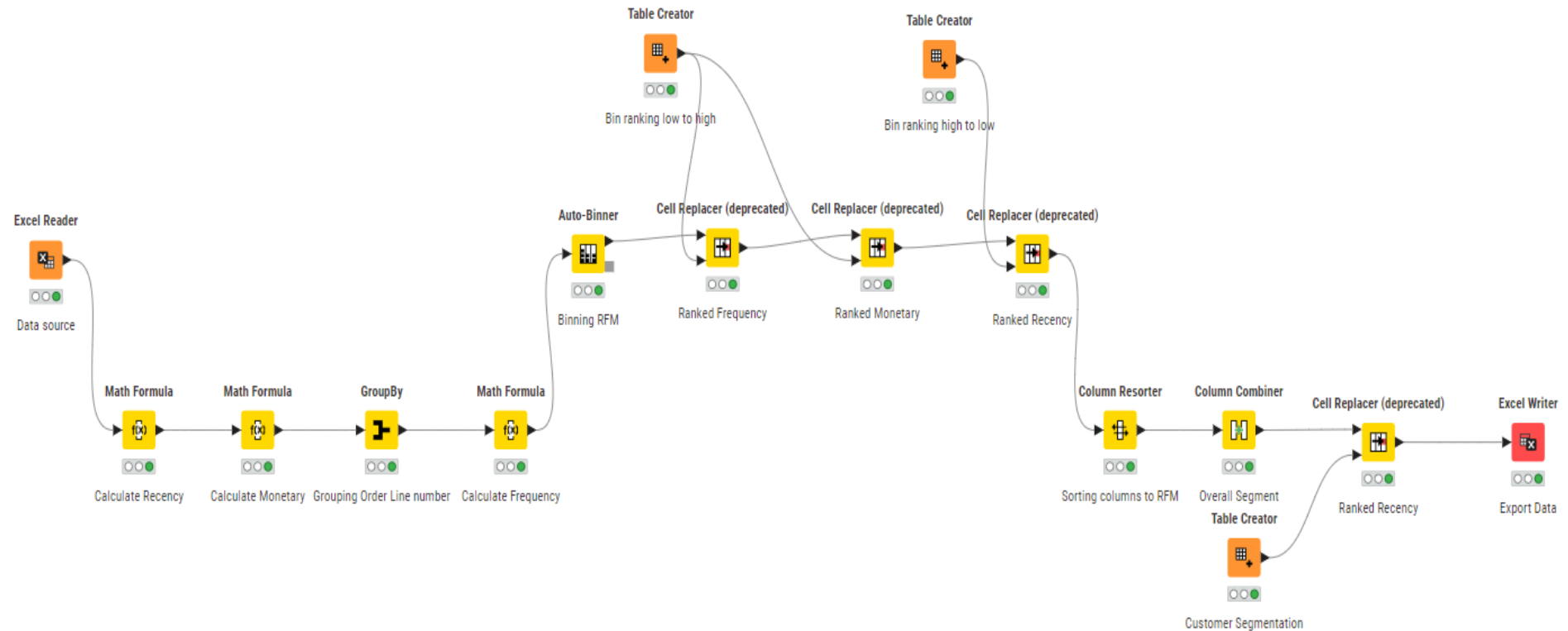
Insights

Business recommendation

RFM Analysis Workflow

KNIME is used to perform the RFM Analysis and here is the workflow diagram. We have used the following formula to compute RFM:

- **Recency:** We have used the days since last order column.
- **Frequency:** We have grouped the dataset by order line number and then used the order line number.
- **Monetary:** We have used sales amount as monetary.



Title
Objective
Data Description
Exploratory Data Analysis
RFM Analysis
Insights
Best customer
Loyal Customer
On Verge of Churn Customer
Lost Customer
Business recommendation

Best customer

- **Profile Overview:**
 - Customers with high recency, high frequency, and high monetary values.
 - They make frequent purchases, spend significantly, and have shopped recently.
- **Key Insights:**
 - Strong Brand Loyalty: Demonstrates consistent purchasing behavior and high engagement.
 - Top Revenue Contributors: Account for a major share of total sales, making them vital to business growth.
 - Low Churn Risk: Recent activity indicates strong brand affinity and continued engagement.
 - Preference for Premium Products: Likely to opt for higher-value automobile parts, contributing to higher sales per transaction.

ORDERNUMBER	CUSTOMERNAME	Recency_HML	Frequency_HML	Monetary_HML	RFM		Customer Cluster
10273		15 H	H	H	H,H,H		Best Customer
10275		18 H	H	H	H,H,H		Best Customer
10280		17 H	H	H	H,H,H		Best Customer
10287		17 H	H	H	H,H,H		Best Customer
10304		17 H	H	H	H,H,H		Best Customer
10306		17 H	H	H	H,H,H		Best Customer
10308		16 H	H	H	H,H,H		Best Customer
10310		17 H	H	H	H,H,H		Best Customer
10312		17 H	H	H	H,H,H		Best Customer
10314		15 H	H	H	H,H,H		Best Customer

Title
Objective
Data Description
Exploratory Data Analysis
RFM Analysis
Insights
Best customer
Loyal Customer
On Verge of Churn Customer
Lost Customer
Business recommendation

Loyal customer

- **Profile Overview:**
 - Customers with medium to high frequency, moderate recency, and moderate to high monetary values.
 - They purchase regularly but may not have made their most recent purchase as recently as the Best Customers.
- **Key Insights:**
 - Consistent Buyers: Demonstrate ongoing engagement and steady purchasing behavior.
 - Potential to Upsell: With strong brand affinity, they are ideal candidates for cross-selling and upselling strategies.
 - Moderate Churn Risk: Though loyal, longer gaps since their last purchase may indicate the need for re-engagement.
 - Valuable Revenue Stream: Contribute significantly to overall sales, ensuring a stable revenue flow.

ORDERNUMBER	CUSTOMERNAME	Recency_HML	Frequency_HML	Monetary_HML	RFM		Customer Cluster
10107		8 M	M	M	M,M,M		Loyal Customer
10114		10 M	M	M	M,M,M		Loyal Customer
10117		12 M	M	M	M,M,M		Loyal Customer
10119		14 M	M	M	M,M,M		Loyal Customer
10121		5 M	L	M	M,L,M		Loyal Customer
10129		9 M	M	M	M,M,M		Loyal Customer
10134		7 M	M	M	M,M,M		Loyal Customer
10140		11 M	M	M	M,M,M		Loyal Customer
10141		9 M	M	M	M,M,M		Loyal Customer
10143		16 M	H	M	M,H,M		Loyal Customer

Title
Objective
Data Description
Exploratory Data Analysis
RFM Analysis
Insights
Best customer
Loyal Customer
On Verge of Churn Customer
Lost Customer
Business recommendation

On Verge of Churn customer

- **Profile Overview:**
 - Customers with low recency (haven't purchased recently), combined with medium to high frequency and moderate to high monetary values in the past.
 - Historically valuable but showing signs of disengagement.
- **Key Insights:**
 - Declining Engagement: Once active buyers who are now at risk of churning due to extended inactivity.
 - High Recovery Potential: Given their strong purchase history, targeted reactivation campaigns could bring them back.
 - Warning Signs: A drop in recent interactions indicates a potential shift in loyalty or interest.
 - Opportunity for Retention: Personalized offers, loyalty programs, or exclusive deals can help reignite interest and prevent churn.

ORDERNUMBER	CUSTOMERNAME	Recency_HML	Frequency_HML	Monetary_HML	RFM	Customer Cluster
10104	13 L	M	M	L,M,M	Customer on the Verge of Churn	
10106	18 L	H	H	L,H,H	Customer on the Verge of Churn	
10109	6 L	M	M	L,M,M	Customer on the Verge of Churn	
10110	16 L	H	H	L,H,H	Customer on the Verge of Churn	
10124	13 L	M	M	L,M,M	Customer on the Verge of Churn	
10131	8 L	M	M	L,M,M	Customer on the Verge of Churn	
10133	8 L	M	M	L,M,M	Customer on the Verge of Churn	
10139	8 L	M	M	L,M,M	Customer on the Verge of Churn	
10148	14 L	M	H	L,M,H	Customer on the Verge of Churn	
10149	11 L	M	M	L,M,M	Customer on the Verge of Churn	
10172	8 L	M	M	L,M,M	Customer on the Verge of Churn	

Title
Objective
Data Description
Exploratory Data Analysis
RFM Analysis
Insights
Best customer
Loyal Customer
On Verge of Churn Customer
Lost Customer
Business recommendation

Lost customer

- **Profile Overview:**
 - Customers with low recency, low frequency, and low monetary values.
 - They have not made a purchase in a long time and showed minimal engagement even when active.
- **Key Insights:**
 - High Churn Risk: These customers are considered inactive and may have shifted to competitors or lost interest.
 - Low Retention Potential: Since their historical engagement was already low, re-engaging them might require significant effort.
 - Valuable Learnings: Analyzing their past interactions can help identify reasons for churn and improve future customer retention strategies.
 - Selective Targeting: Only high-value lost customers may warrant targeted win-back campaigns, while others can be deprioritized.

ORDERNUMBER	CUSTOMERNAME	Recency_HML	Frequency_HML	Monetary_HML	RFM	Customer Cluster
10100		4 L	L	L	L,L,L	Lost Customer
10101		4 L	L	L	L,L,L	Lost Customer
10102		2 L	L	L	L,L,L	Lost Customer
10112		2 M	L	L	M,L,L	Lost Customer
10113		4 M	L	L	M,L,L	Lost Customer
10115		5 L	L	M	L,L,M	Lost Customer
10116		1 L	L	L	L,L,L	Lost Customer
10118		1 L	L	L	L,L,L	Lost Customer
10123		4 L	L	L	L,L,L	Lost Customer
10125		2 L	L	L	L,L,L	Lost Customer

Title

Objective

Data Description

Exploratory Data Analysis

RFM Analysis

Insights

Business Recommendation

Business Recommendations

Best Customers

Your top customers who purchase frequently, spend significantly, and have ordered recently.

- **Exclusive Offers:** Provide early access to new product lines or limited-edition parts.
- **Loyalty Programs:** Offer reward points, discounts, or free shipping on future purchases.
- **Personalized Engagement:** Send personalized emails with product recommendations based on past purchases.
- **Referral Incentives:** Encourage them to refer peers by offering incentives like discounts or free accessories.

Loyal Customers

Repeat buyers who regularly purchase but may not be the biggest spenders.

- **Upselling/Cross-Selling:** Recommend complementary parts (e.g., filters with engine parts) or higher-end alternatives.
- **Special Promotions:** Offer bundle deals or volume discounts to increase their average order value.
- **Feedback Loop:** Engage them for product feedback or beta testing on new auto parts.
- **Loyalty Tiers:** Introduce tiered rewards to motivate more purchases and push them into the “Best Customer” category.

On Verge of Churn Customers

Previously active and valuable customers who haven’t purchased recently.

- **Win-Back Campaigns:** Send personalized re-engagement emails with special offers or discounts.
- **Maintenance Reminders:** For parts with typical wear/tear timelines, send proactive reminders to reorder (e.g., filters, brake pads).
- **Product Updates:** Inform them of new products or improvements in existing parts.
- **Reactivation Incentives:** Offer time-limited discounts or free shipping on their next order.

Lost Customer

Inactive customers who haven’t purchased in a long time and had minimal engagement.

- **Reactivation Offers:** Send deep discounts or bundle deals to reignite interest.
- **Exit Surveys:** Understand why they stopped buying—pricing, product issues, or competitors.
- **Highlight Innovations:** Share new product lines or company milestones that may attract them back.
- **“We Miss You” Campaigns:** Personalized messages with offers like “Get 20% off your next order—just for coming back!”