Marketing and Retail Analysis

Project 1
RFM Analysis

Sayyed Abdul Khaliq
Email : abdulkhaliq01112001@gmail.com

CONTENTS

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.

Overview

Objective:

To leverage data science techniques to analyze customer transaction data for an automobile parts manufacturing company, with the aim of identifying underlying buying patterns, understanding customer segments based on RFM (Recency, Frequency, Monetary) analysis, and recommending customized marketing strategies. The project seeks to enhance customer engagement, optimize marketing efforts, and improve overall business performance by delivering actionable insights from the data

Data Description

Column Name	Description
ORDERNUMBER	This column represents the unique identification number assigned to each order.
QUANTITYORDERED	It indicates the number of items ordered in each order.
PRICEEACH	This column specifies the price of each item in the order.
ORDERLINENUMBER	It represents the line number of each item within an order.
SALES	This column denotes the total sales amount for each order, which is calculated by multiplying the quantity ordered by the price of each item.
ORDERDATE	It denotes the date on which the order was placed.
DAYS_SINCE_LASTORDER	This column represents the number of days that have passed since the last order for each customer. It can be used to analyze customer purchasing patterns.
STATUS	It indicates the status of the order, such as "Shipped," "In Process," "Cancelled," "Disputed," "On Hold," or "Resolved"
PRODUCTLINE	This column specifies the product line categories to which each item belongs.
MSRP	It stands for Manufacturer's Suggested Retail Price and represents the suggested selling price for each item.
PRODUCTCODE	This column represents the unique code assigned to each product.
CUSTOMERNAME	It denotes the name of the customer who placed the order.
PHONE	This column contains the contact phone number for the customer.
ADDRESSLINE1	It represents the first line of the customer's address.
CITY	This column specifies the city where the customer is located.
POSTALCODE	It denotes the postal code or ZIP code associated with the customer's address.
COUNTRY	This column indicates the country where the customer is located.
CONTACTLASTNAME	It represents the last name of the contact person associated with the customer.
CONTACTFIRSTNAME	This column denotes the first name of the contact person associated with the customer.
DEALSIZE	It indicates the size of the deal or order, which are the categories "Small," "Medium," or "Large."

Table 1 – Data Dictionary

Summary of Data

- The dataset consists of 2,747 transactions made by customers of an automobile parts manufacturing company.
- The dataset consists of 20 columns, including 2 float, 6 integer, and 12 object data types, providing a diverse range of information about the transactions and customers.
- The dataset is free from duplicate entries.
- There are no null values in the dataset.

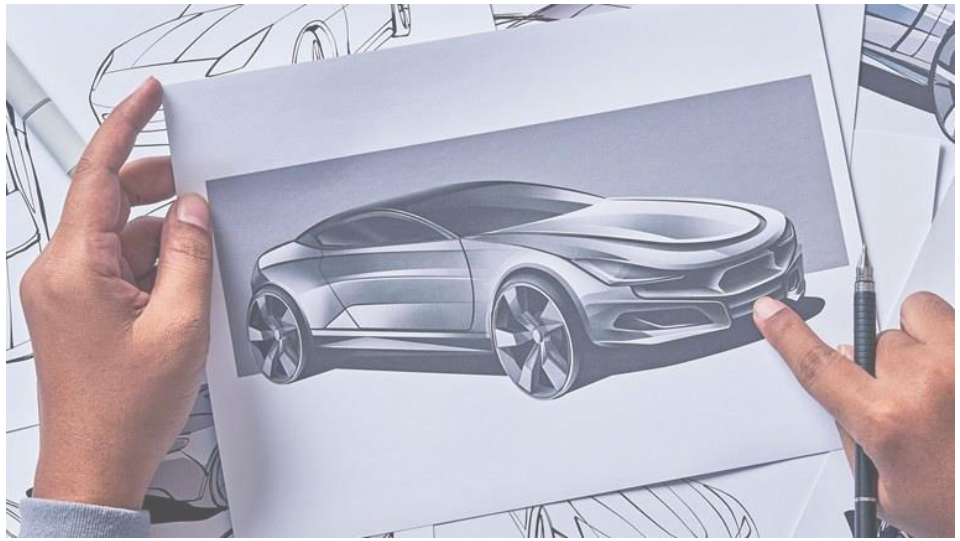
Descriptive Statistics:

	ORDERNUMBER	QUANTITYORDERED	PRICEEACH	ORDERLINENUMBER	SALES	DAYS_SINCE_LASTORDER	MSRI
:ount	2747.0	2747.0	2747.0	2747.0	2747.0	2747.0	2747.(
nean	10259.8	35.1	101.1	6.5	3553.0	1757.1	100.1
std	91.9	9.8	42.0	4.2	1839.0	819.3	40.1
min	10100.0	6.0	26.9	1.0	482.1	42.0	33.(
25%	10181.0	27.0	68.7	3.0	2204.3	1077.0	68.(
50%	10264.0	35.0	95.5	6.0	3184.8	1761.0	99.(
75%	10334.5	43.0	127.1	9.0	4503.1	2436.5	124.(
max	10425.0	97.0	252.9	18.0	14082.8	3562.0	214.(

Table 2 – Descriptive Stats

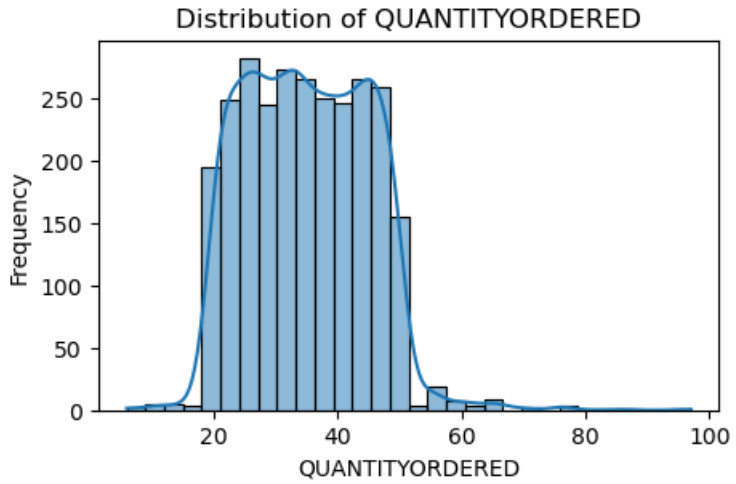
Data assumptions:

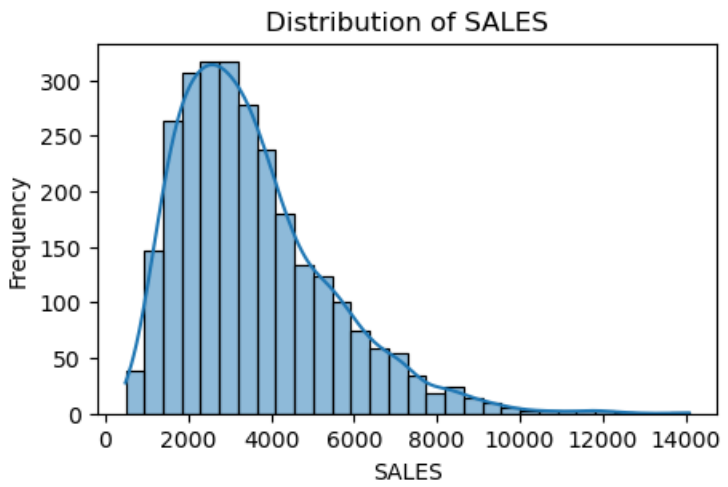
- The dataset contains information about orders placed by customers. Each order is identified by a unique ORDERNUMBER.
- The dataset contains information about customers' transactions and their associated details, such as the number of items purchased, days since the last order, and the MSRP of the products. This implies that the dataset includes transactional data related to customer purchases.
- The dataset does not contain duplicate entries. This means that each transaction entry is unique, and there are no repeated records for the same transaction or customer

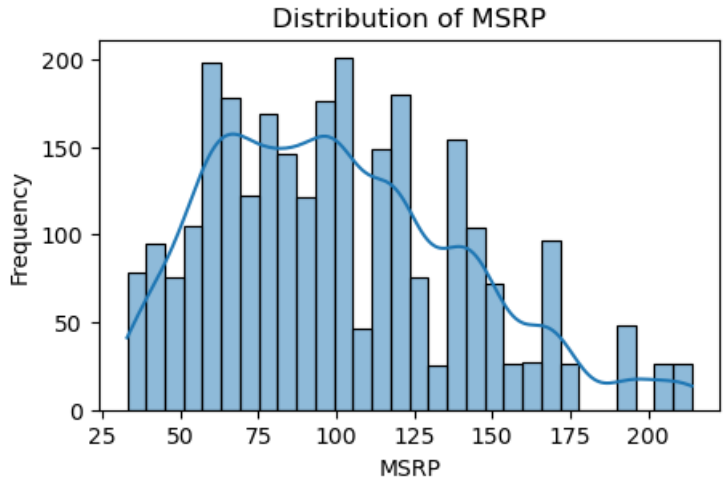


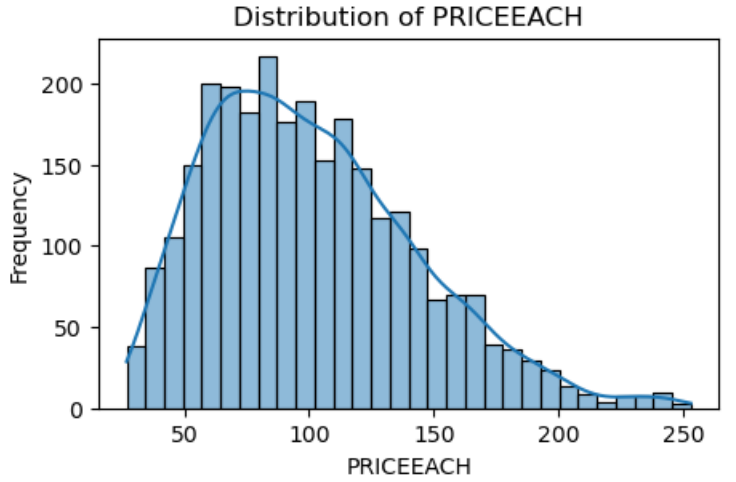
Exploratory Data Analysis

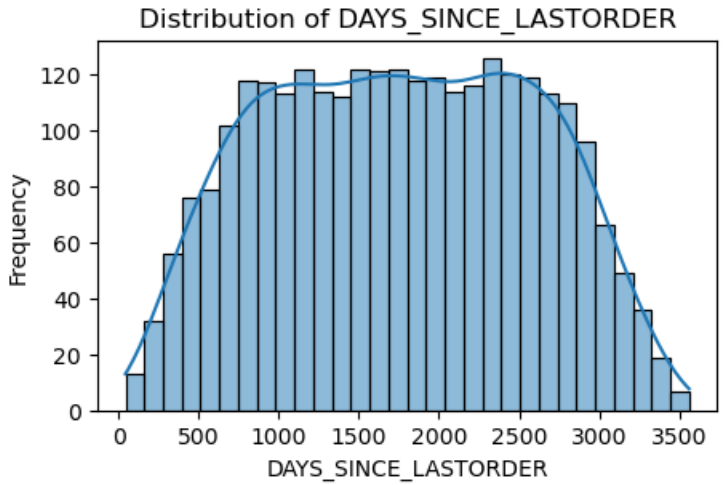
Univariate Analysis





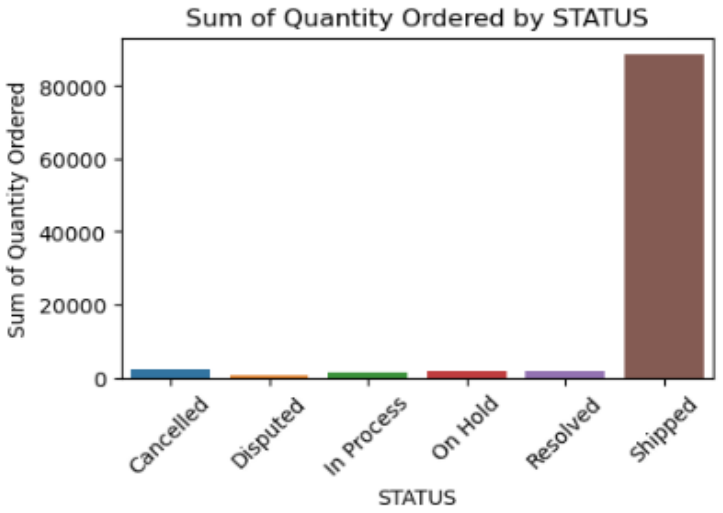


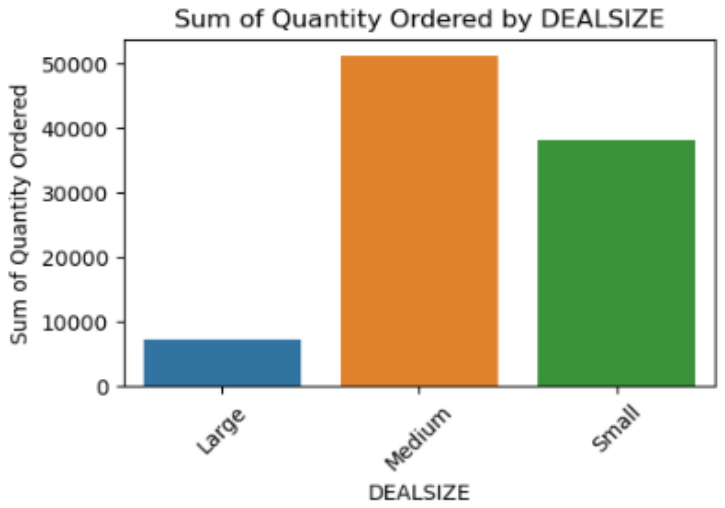


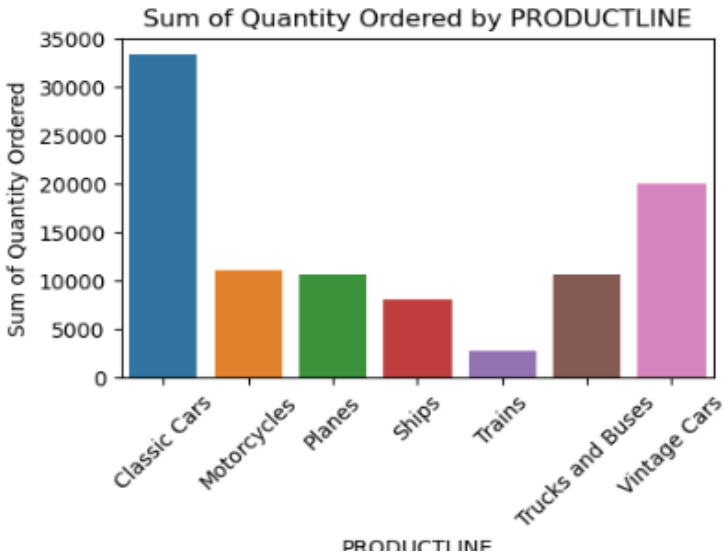


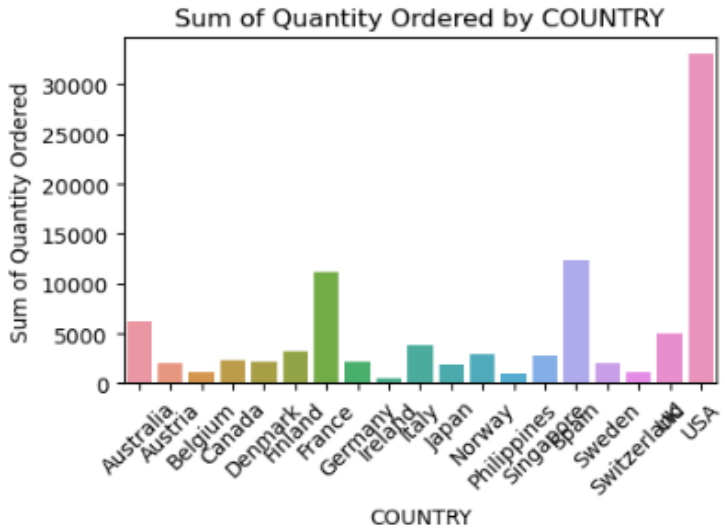
- Sales distribution seems positively skewed, with less higher values possibly due to larger quantities or higher-priced items.
- Price distribution is also Positively skewed with long right tail indicating higher priced items are sold less.

Univariate Analysis



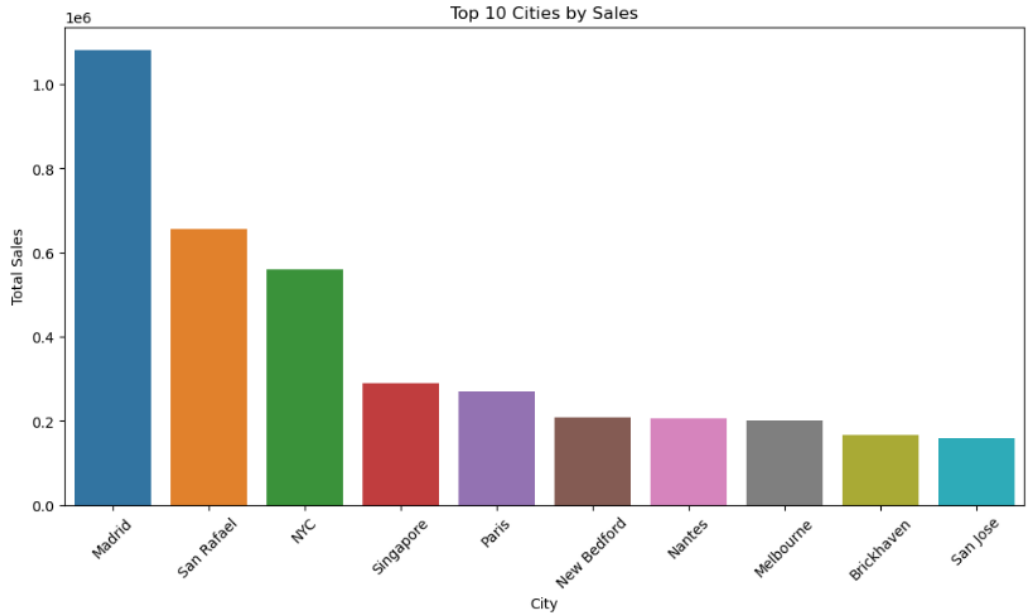


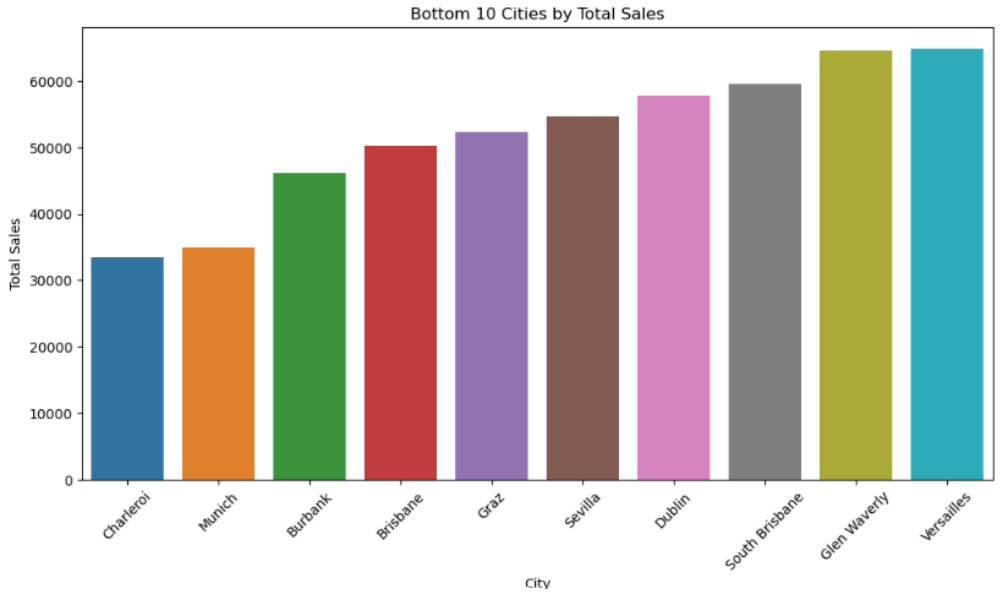




- The impressive count of successfully shipped orders highlights the company's robust order fulfillment capabilities.
- The majority of the deals fall into the small and medium categories indicating a focus on moderate to smaller deal sizes.
- Classic cars and vintage cars are the most popular product lines, reflecting strong demand for these vehicle models.
- Ships and trains have relatively lower counts, suggesting a niche market or specialized customer base for these product lines.
- The USA has the highest number of customers, followed by Spain, France, Australia, and the UK, indicating a strong market presence in these regions.

Univariate Analysis – Top 10 Countries

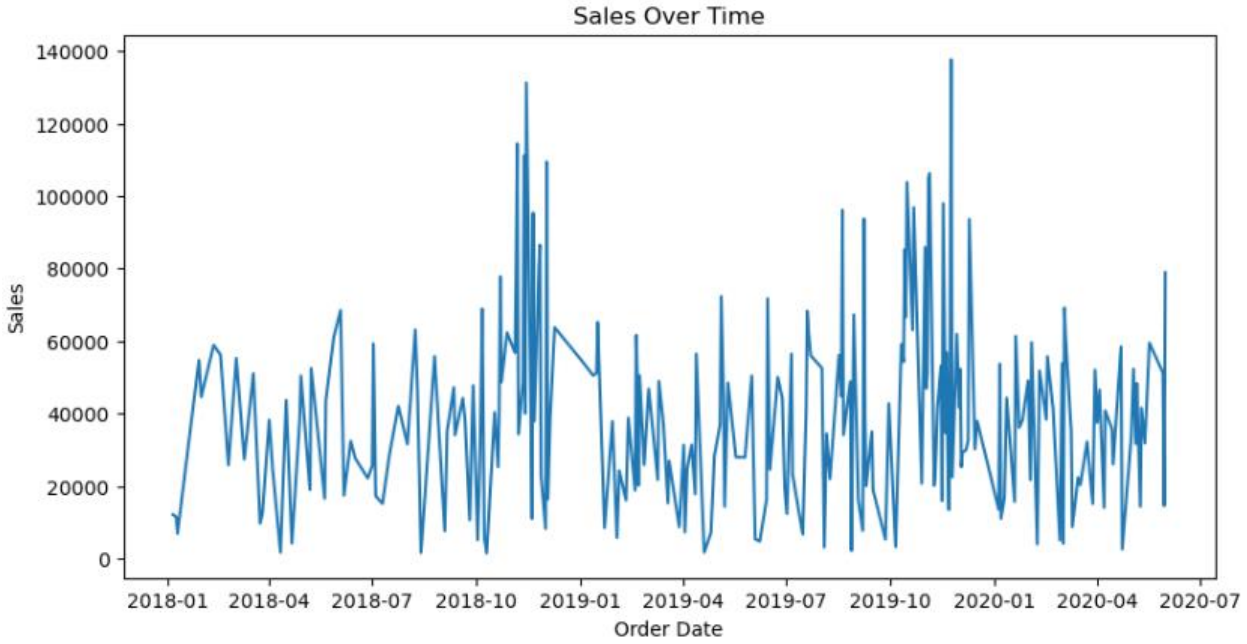


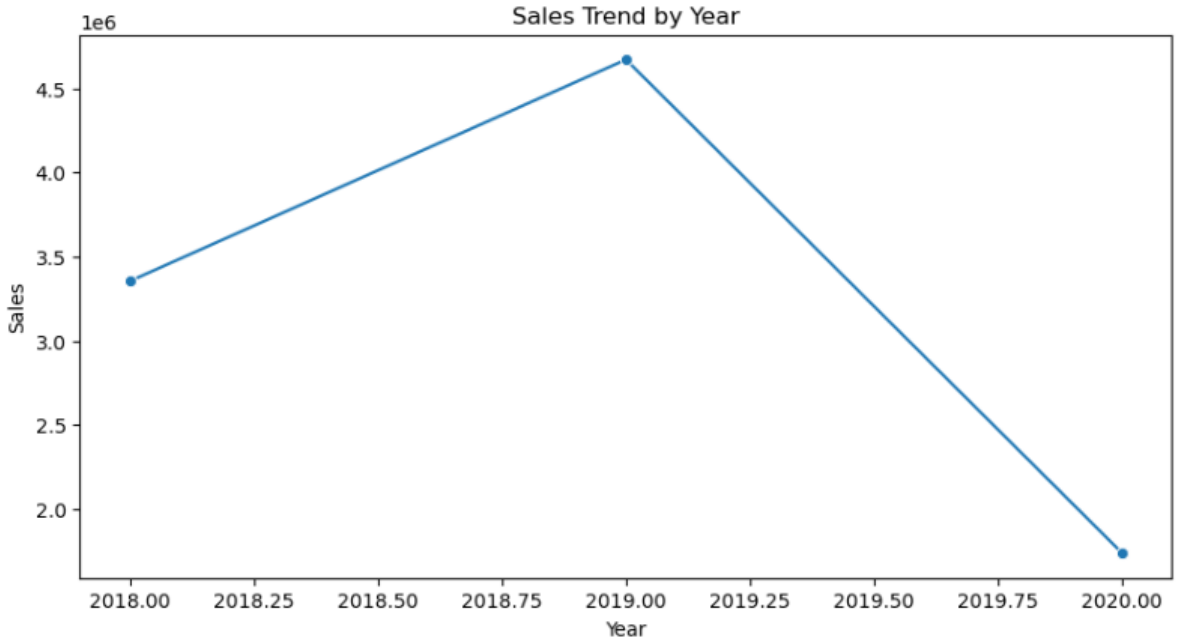


•Top 10 Cities: The top 10 cities are the major contributors to total sales, indicating where the bulk of the revenue is generated. These cities can be considered primary target markets for marketing campaigns and sales promotions due to their high sales volumes. More resources (such as inventory, customer service, and logistics) can be allocated to these cities to support the high demand. These cities might already have a high market penetration, so further sales growth may require innovative strategies or new product introductions.

•Bottom10 Cities: These cities represent potential growth areas where targeted efforts might increase sales. Investigating the reasons for low sales in these cities could uncover opportunities. epending on the strategic importance of these markets, the company might decide to optimize or reduce resource allocation to these cities to better manage costs. This could involve local market research, better distribution channels, or tailored marketing efforts.

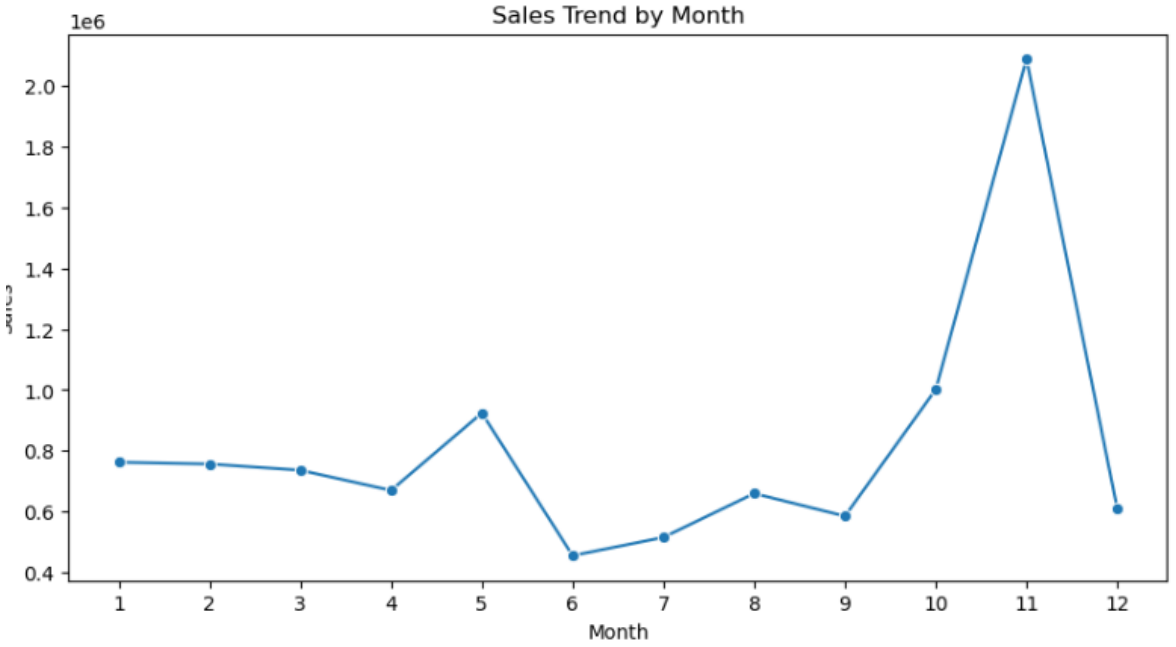
Trend Analysis

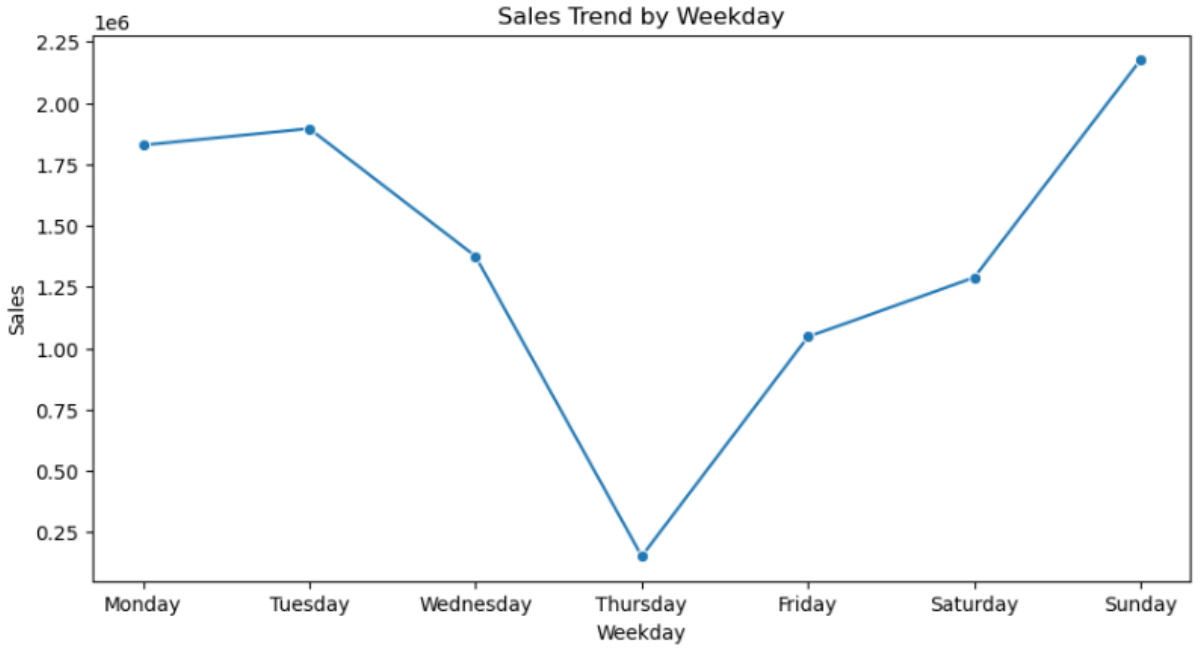




- Sales demonstrated a robust positive trend, growing substantially from 2018 to 2019. The total sales increased in 2019, indicating a strong business performance and heightened customer demand.
- However, the sales data for 2020 only covers up until May . Due to this incomplete data, it's challenging to draw definitive conclusions about the overall performance for 2020.

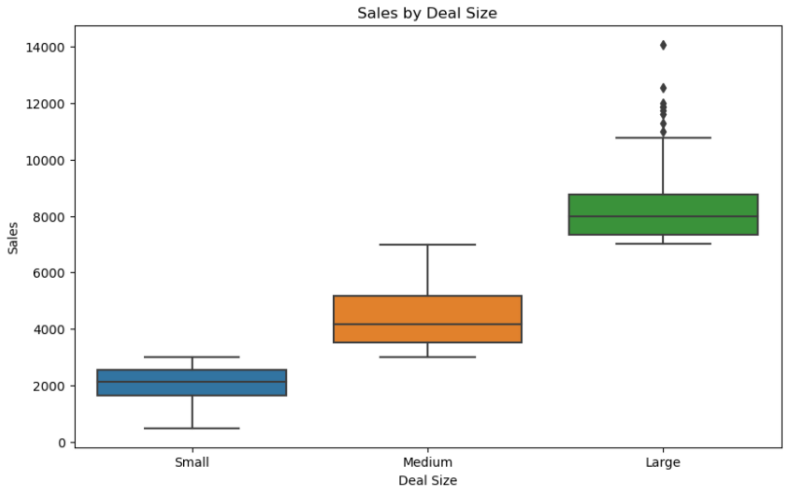
Trend Analysis

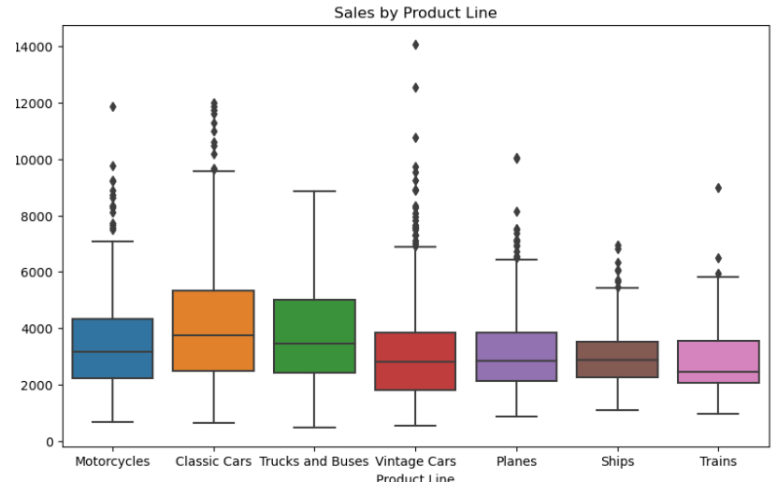


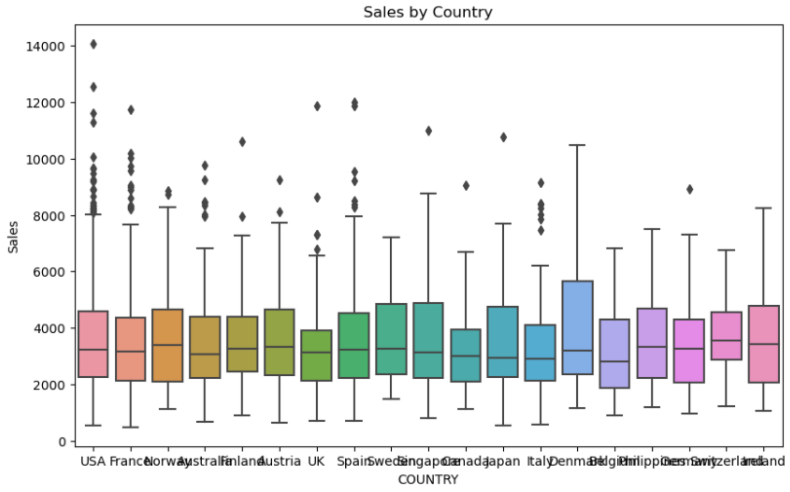


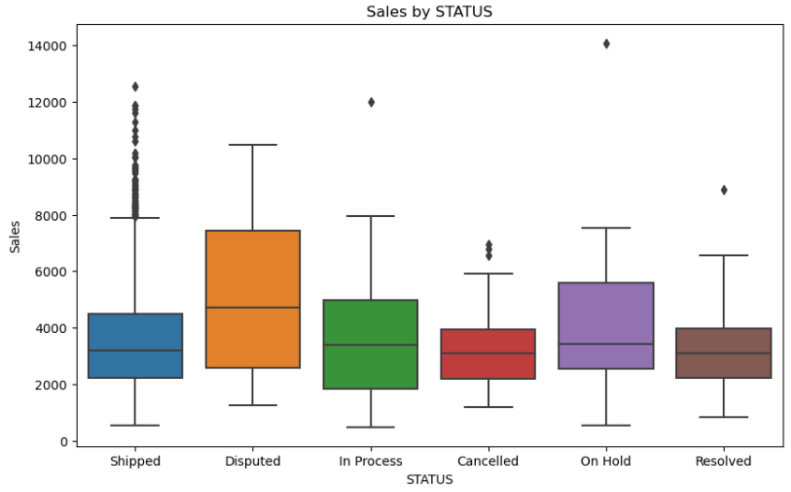
- The sales data exhibits notable fluctuations throughout the year, with peak sales recorded in November and October. This trend suggests heightened customer activity during these months, likely influenced by seasonal factors such as holiday shopping.
- In contrast, June records the lowest sales, indicating a potential decrease in customer demand during this period. This dip may reflect a seasonal lull or other factors affecting consumer spending.
- Analyzing weekly trends, Sundays consistently show the highest sales over the three-year period. This pattern indicates that customers are more active and likely prefer shopping on weekends.
- On the other hand, Wednesdays have the lowest sales among the weekdays, suggesting that midweek may see reduced customer engagement or be influenced by factors that affect their buying behavior on this specific day.

Bivariate Analysis



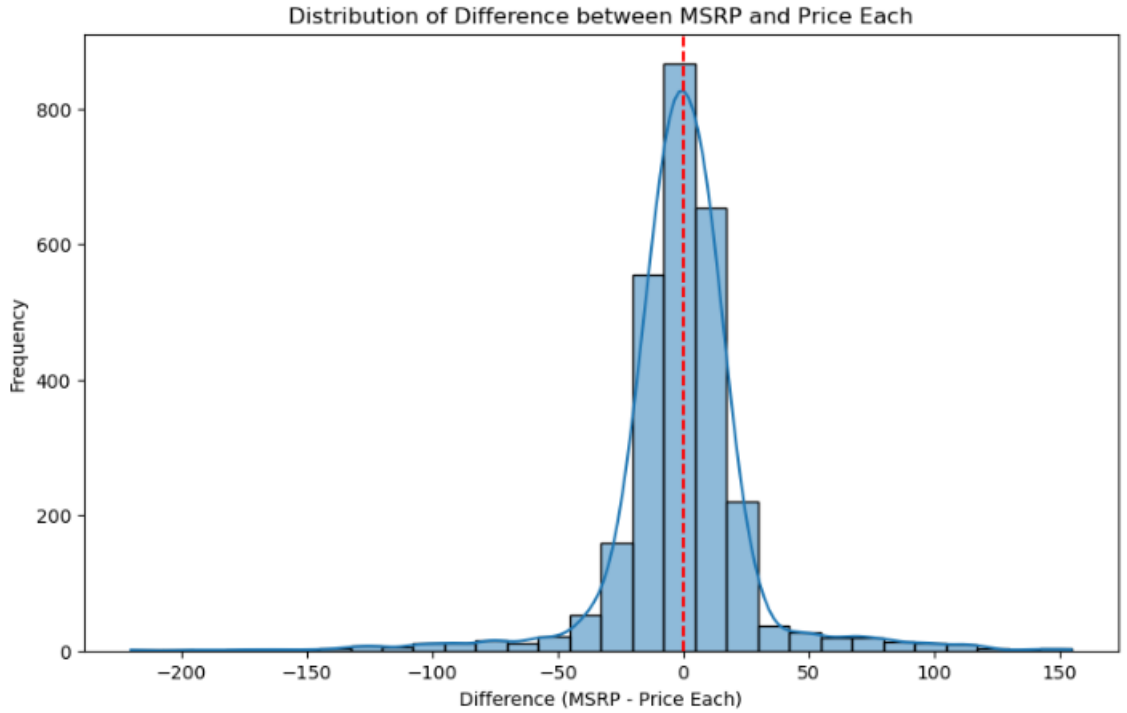






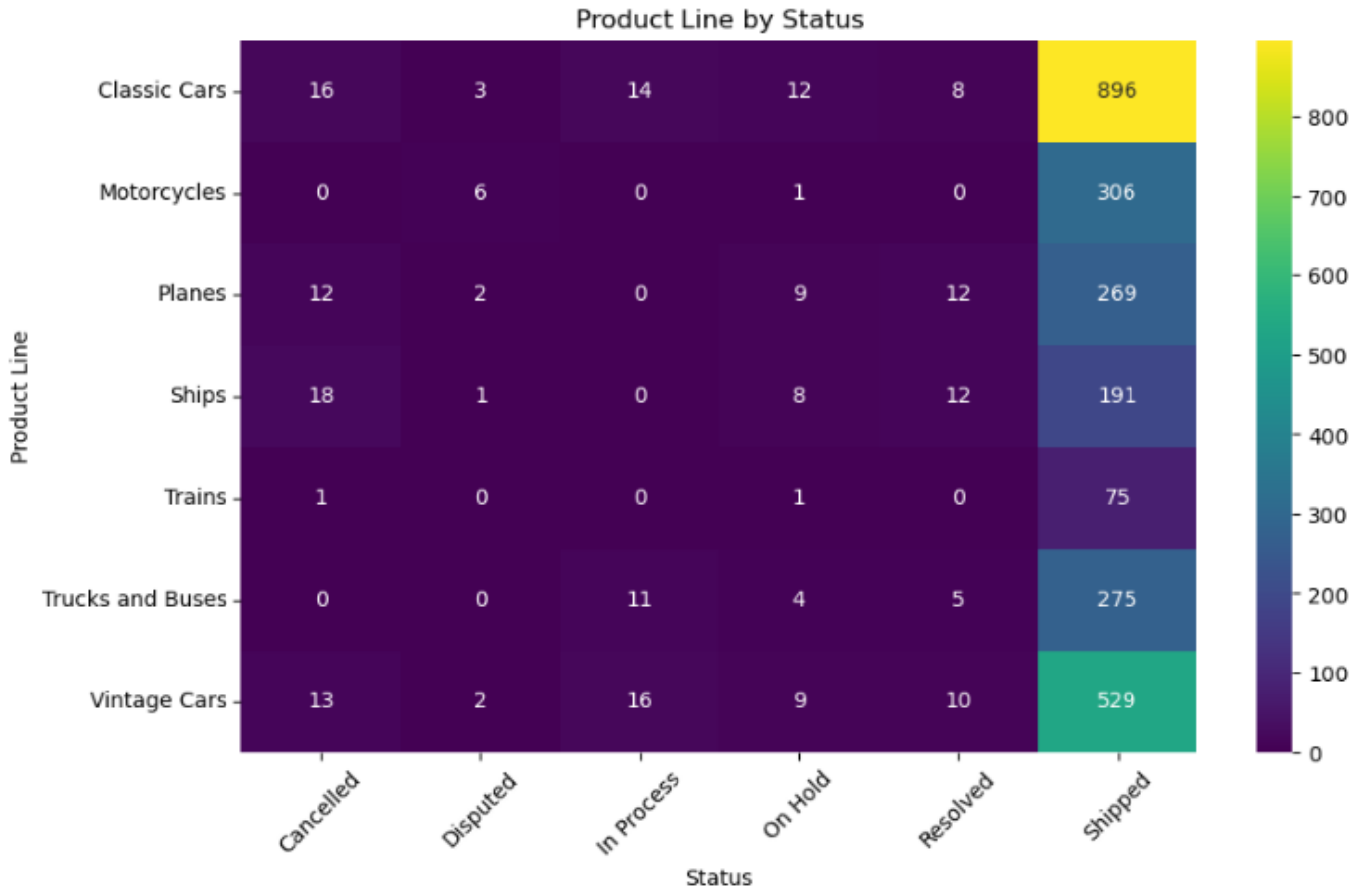
- The analysis confirms that deal size is categorized based on the sales amount, reflecting different transaction magnitudes.
- **Product Line Insights**:
  - Classic cars and Trucks and Buses stand out with higher sales compared to other product lines.
- **Geographical Sales Insights**:
  - While the USA has a higher quantity of products sold, the overall sales by country do not show significant variations. This suggests that despite the higher sales volume in the USA, other countries also maintain a relatively consistent sales performance.
- **Order Status Insights**:
  - Disputed orders exhibit higher sales amounts compared to other order statuses. This could indicate that disputes are more common in larger transactions or higher-value orders, possibly due to the complexities involved in such deals.

Bivariate Analysis – MSRP vs Sale Price



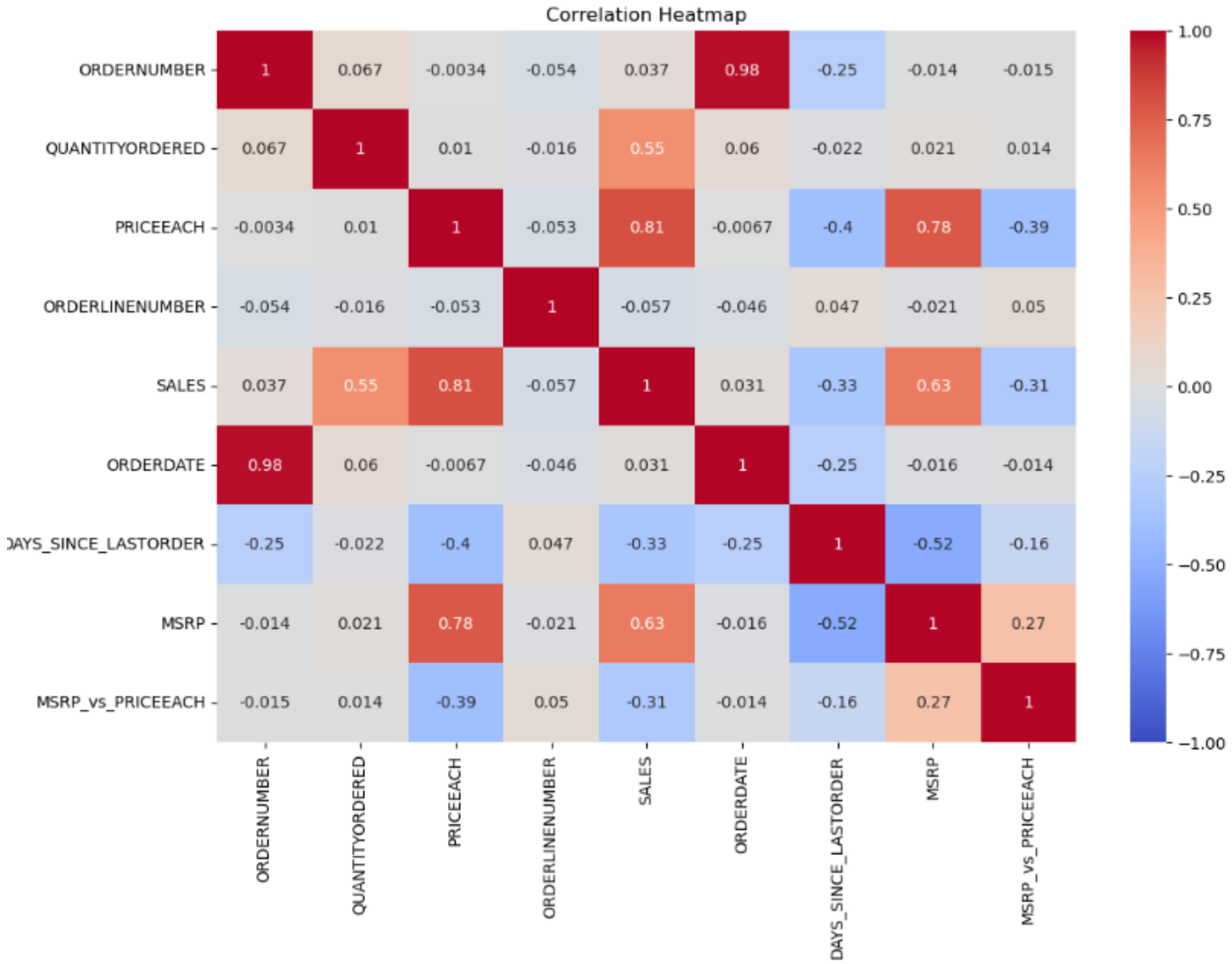
- The distribution of MSRP-PRICEEACH being symmetrical indicates that there is no significant skewness in the pricing deviations. This suggests that the prices are generally balanced around the MSRP.
- Distribution has long tails or outliers. This implies that there are occasional significant deviations from the MSRP, either much higher or lower.
- These outliers could be due to special promotions, discounts, or premium pricing strategies

Bivariate Analysis – MSRP vs Sale Price



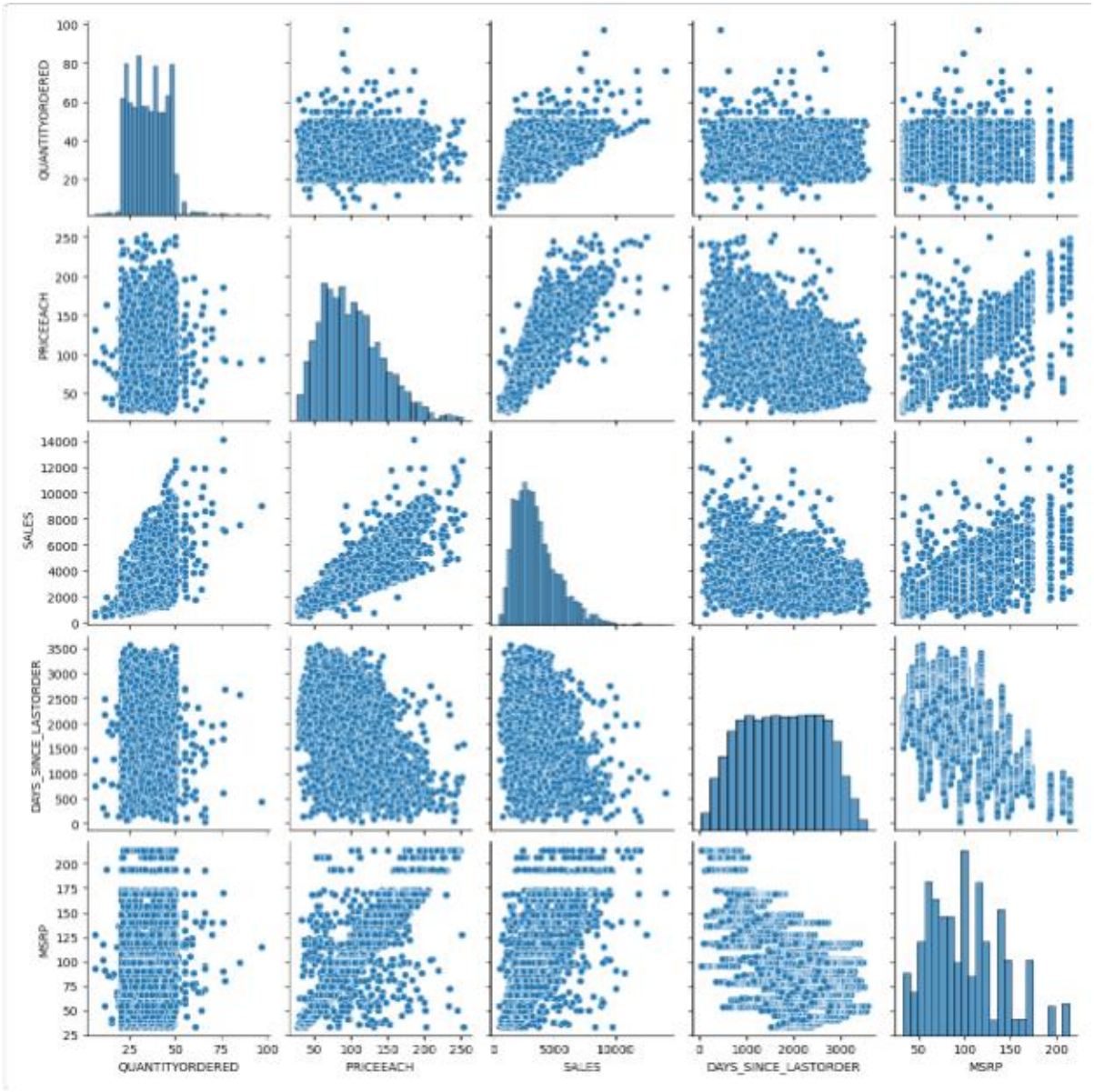
- **Classic Cars** have the highest number of shipped orders (896) but also show significant cancellations (16) and orders on hold (12), indicating potential issues in order fulfillment or customer satisfaction.
- **Motorcycles** and **Trucks and Buses** show a high proportion of orders shipped (306 and 275, respectively) with minimal issues, suggesting efficient order processing.
- **Planes** and **Ships** have notable numbers of cancellations (12 and 18, respectively) and disputed orders (2 and 1, respectively), which could indicate specific problems with these product lines.
- **Vintage Cars** also show a high number of shipped orders (529) but relatively high cancellations (13) and orders on hold (9), highlighting potential areas for improvement in managing these orders.

Multivariate Analysis – Correlation Map



- There is a moderate negative correlation between the number of days since the last order (DAYS_SINCE_LASTORDER) and SALES. This suggests that customers who order more frequently tend to contribute more to total sales.
- The strong positive correlation between PRICEEACH and SALES (0.8) indicates that increasing the price of items tends to lead to higher total sales. This suggests that customers might perceive higher-priced items as higher quality or more desirable, influencing their purchasing decisions.
- The positive correlation between MSRP and PRICEEACH (0.8) indicates that products with higher Manufacturer's Suggested Retail Prices (MSRP) tend to be sold at higher prices (PRICEEACH). This relationship underscores the importance of brand perception and pricing alignment with market expectations.

Multivariate Analysis – PairPlot



- As QUANTITYORDERED or PRICEEACH increases, SALES also tends to increase, showing that higher quantities or prices contribute to higher sales.
- Orders where longer days since the last order are associated with lower sales This suggests that customer purchasing patterns and sales performance are influenced by recency of orders.
- as PRICEEACH increases, MSRP_vs_PRICEEACH tends to decrease, and as MSRP increases, MSRP_vs_PRICEEACH tends to increase. This indicates deviations of the actual price from the manufacturer's suggested retail price and highlights potential pricing strategies or market behaviors.

Summary of Insights:

1. **Frequency and Sales**: Customers ordering more frequently tend to contribute more to total sales, showing a moderate negative correlation between DAYS_SINCE_LASTORDER and SALES.
2. **Price Influence**: A strong positive correlation (0.8) between PRICEEACH and SALES suggests that higher-priced items lead to higher total sales, likely due to perceived quality or desirability.
3. **Product Line Performance**:
   - Classic Cars and Trucks and Buses stand out with higher sales, indicating strong market demand.
   - Ships and Trains have lower sales, suggesting a niche or specialized market.
4. **Geographical Insights**: While the USA leads in product quantity sold, overall sales across countries show consistent performance, indicating a balanced market presence.
5. **Order Status Impact**:
   - Disputed orders have higher sales, possibly due to complexities in larger transactions, highlighting areas for improved dispute resolution.
   - High counts of successfully shipped orders reflect robust order processing and fulfillment capabilities.
6. **Deal Size Distribution**: Most deals fall into small and medium categories, indicating a focus on moderate transaction sizes.
7. **Price Distribution**: The distribution of MSRP vs. PRICEEACH shows symmetry, suggesting balanced pricing deviations around the MSRP with occasional outliers due to promotional or pricing strategies.
8. **Seasonal Sales Trends**:
   - Peak sales occur in November and October, likely driven by holiday shopping and seasonal promotions.
   - June records the lowest sales, suggesting a seasonal dip in customer demand during this period.
9. **Weekly Sales Patterns**:
   - Sundays consistently exhibit the highest sales, indicating increased customer activity and preference for weekend shopping.
   - Wednesdays show the lowest sales among weekdays, possibly influenced by reduced customer engagement midweek.

RFM Analysis

What is RFM?

- RFM (Recency, Frequency, Monetary) analysis is a behavior-based approach grouping customers into segments. It groups the customers on the basis of their previous purchase transactions. How recently, how often, and how much did a customer buy. RFM filters customers into various groups for the purpose of better service. There is a segment of customer who is the big spender but what if they purchased only once or how recently they purchased? Do they often purchase our product? Also, It helps managers to run an effective promotional campaign for personalized service.

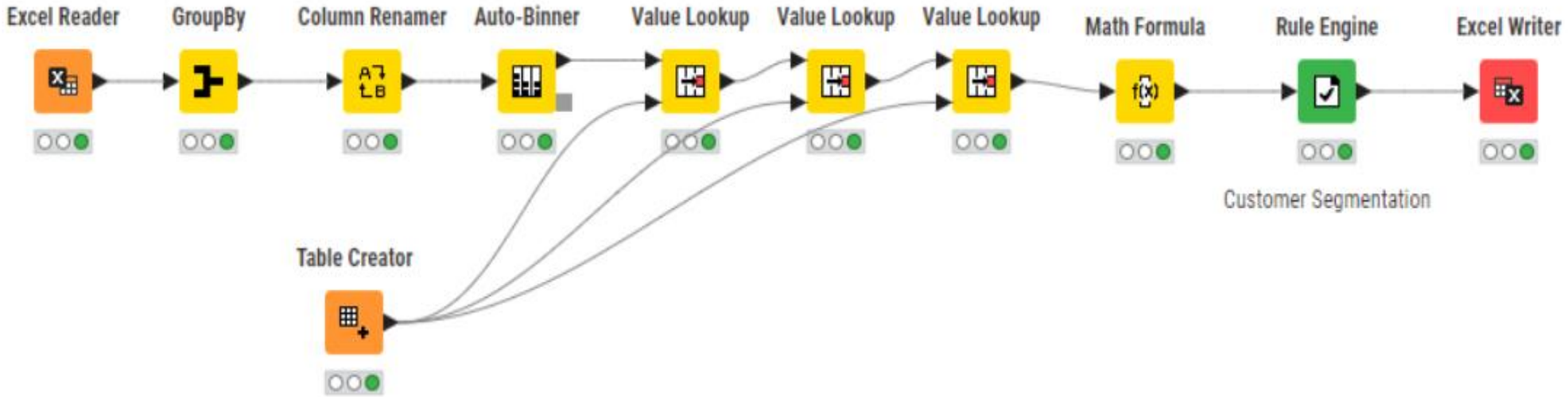
Parameters:

- **Recency (R):** Who have purchased recently? Number of days since last order(least recency)
- **Frequency (F):** Who has purchased frequently? It means the total number of purchases. (high frequency)
- **Monetary Value(M):** Who have high purchase amount? It means the total money customer spent i.e Sales (high monetary value)
- **Auto-binning:** Customers are segmented into four categories based on their above scores (1 to 4). This segmentation helps in categorizing customers based on their value and behaviour.

Assumptions:

- **Recent purchases are more indicative of customer engagement:** The assumption is that customers who have made purchases more recently are likely to be more engaged with the company and its offerings. They may have higher potential for repeat purchases or upselling/cross-selling opportunities.
- **Higher Frequency of purchases reflects customer loyalty:** The assumption is that customers who make purchases more frequently are more loyal to the company. They may have a stronger connection to the brand, higher customer satisfaction, and a higher likelihood of recommending the company to others.
- **Higher Monetary value indicates a higher spending customer:** The assumption is that customers who generate higher monetary value through their purchases are likely to be more valuable and potentially more profitable for the business.
- **Calculation of RFM Score:** Maximum weightage is given to recency then frequency and then monetary.

Knime Workflow :



Tool Used

Open for Innovation
KNIME

Output Table:

Rows: 89 | Columns: 12

#	RowID	CUSTOMER... String	Frequency Number (integer)	Monetary Number (double)	Recency Number (integer)	Frequency [... String	Monetary [... String	Recency [Bi... String	Recency_S... Number (integer)	Frequency_... Number (integer)	Monetary_S... Number (integer)	RFM_Score Number (double)	Customer_... String
1	Row0	AV Stores, Co.	3	157,807.81	421	Bin 2	Bin 4	Bin 3	2	2	4	224	Losing Customer
2	Row1	Alpha Cognac	3	70,488.44	675	Bin 2	Bin 1	Bin 4	1	2	1	121	Evasive Custom...
3	Row2	Amica Models ...	2	94,117.26	328	Bin 1	Bin 3	Bin 2	3	1	3	313	Late Bloomer
4	Row3	Anna's Decorati...	4	153,996.13	131	Bin 4	Bin 4	Bin 1	4	4	4	444	Platinum Custo...
5	Row4	Atelier graphique	3	24,179.96	312	Bin 2	Bin 1	Bin 2	3	2	1	321	Late Bloomer
6	Row5	Australian Colle...	3	64,591.46	1018	Bin 2	Bin 1	Bin 4	1	2	1	121	Evasive Custom...
7	Row6	Australian Colle...	5	200,995.41	229	Bin 4	Bin 4	Bin 1	4	4	4	444	Platinum Custo...
8	Row7	Australian Gift ...	3	59,469.12	190	Bin 2	Bin 1	Bin 1	4	2	1	421	Loyal
9	Row8	Auto Assoc. & ...	2	64,834.32	275	Bin 1	Bin 1	Bin 2	3	1	1	311	Late Bloomer
10	Row9	Auto Canal Petit	3	93,170.66	127	Bin 2	Bin 3	Bin 1	4	2	3	423	Loyal
11	Row...	Auto-Moto Clas...	3	26,479.26	1353	Bin 2	Bin 1	Bin 4	1	2	1	121	Evasive Custom...
12	Row...	Baane Mini Imp...	4	116,599.19	245	Bin 4	Bin 3	Bin 1	4	4	3	443	Very Loyal
13	Row...	Bavarian Collec...	1	34,993.92	801	Bin 1	Bin 1	Bin 4	1	1	1	111	Lost Customer
14	Row...	Blauer See Auto...	4	85,171.59	705	Bin 4	Bin 2	Bin 4	1	4	2	142	Almost Lost Cu...
15	Row...	Boards & Toys ...	2	9,129.35	410	Bin 1	Bin 1	Bin 2	3	1	1	311	Late Bloomer
16	Row...	CAF Imports	2	49,642.05	625	Bin 1	Bin 1	Bin 3	2	1	1	211	Almost Lost Cu...
17	Row...	Cambridge Coll...	2	36,163.62	484	Bin 1	Bin 1	Bin 3	2	1	1	211	Almost Lost Cu...
18	Row...	Canadian Gift E...	2	75,238.92	364	Bin 1	Bin 2	Bin 2	3	1	2	312	Late Bloomer
19	Row...	Classic Gift Ide...	2	67,506.97	344	Bin 1	Bin 1	Bin 2	3	1	1	311	Late Bloomer
20	Row...	Classic Legend...	3	77,795.2	309	Bin 2	Bin 2	Bin 2	3	2	2	322	Late Bloomer
21	Row...	Clover Collectio...	2	57,756.43	659	Bin 1	Bin 1	Bin 4	1	1	1	111	Lost Customer
22	Row...	Collectable Min...	2	87,489.23	575	Bin 1	Bin 2	Bin 3	2	1	2	212	High Risk Custo...
23	Row...	Collectables Fo...	3	81,577.98	179	Bin 2	Bin 2	Bin 1	4	2	2	422	Loyal
24	Row...	Corrida Auto Re...	3	120,615.28	407	Bin 2	Bin 4	Bin 2	3	2	4	324	Potential Custo...
25	Row...	Cruz & Sons Co.	3	94,015.73	971	Bin 2	Bin 3	Bin 4	1	2	3	123	Evasive Custom...
26	Row...	Daedalus Desig...	2	69,052.41	573	Bin 1	Bin 1	Bin 3	2	1	1	211	Almost Lost Cu...
27	Row...	Danish Wholes...	5	145,041.6	499	Bin 4	Bin 4	Bin 3	2	4	4	244	Losing Customer
28	Row...	Diecast Classic...	4	122,138.14	228	Bin 4	Bin 4	Bin 1	4	4	4	444	Platinum Custo...

Table 3– Output table

Customer Segmentation

Customers are segmented into 11 groups based on their RFM (Recency, Frequency, Monetary Value) scores.

1.Platinum Customer: Customers with the highest RFM scores (444), indicating they are highly engaged, frequent buyers, and contribute significantly to revenue.

2.Very Loyal: Customers with RFM scores between 433 and 443. They are very loyal and make frequent purchases, though slightly below the top tier in terms of overall engagement.

3.Loyal: Customers with RFM scores between 421 and 432. They are transitioning towards becoming very loyal customers, showing increasing frequency and monetary value.

4.Recent Customer: Customers with RFM scores between 344 and 420. These are customers who have made recent purchases but may not yet demonstrate high frequency or monetary value.

5.Potential Customer: Customers with RFM scores between 323 and 343. They show potential for becoming more engaged with the brand, possibly increasing their frequency and spending.

6.Late Bloomer: Customers with RFM scores between 311 and 322. These are customers who are showing improvement or increased engagement later than others, potentially indicating a recent uptick in activity.

7.Losing Customer: Customers with RFM scores between 224 and 310. They have shown a decline in engagement compared to previous periods, possibly indicating decreased frequency or spending.

8.High Risk Customer: Customers with RFM scores between 212 and 223. They are at higher risk of becoming inactive or churning due to declining engagement levels.

9.Almost Lost Customer: Customers with RFM scores between 124 and 211. These customers are at risk of being lost due to very low engagement levels and infrequent purchases.

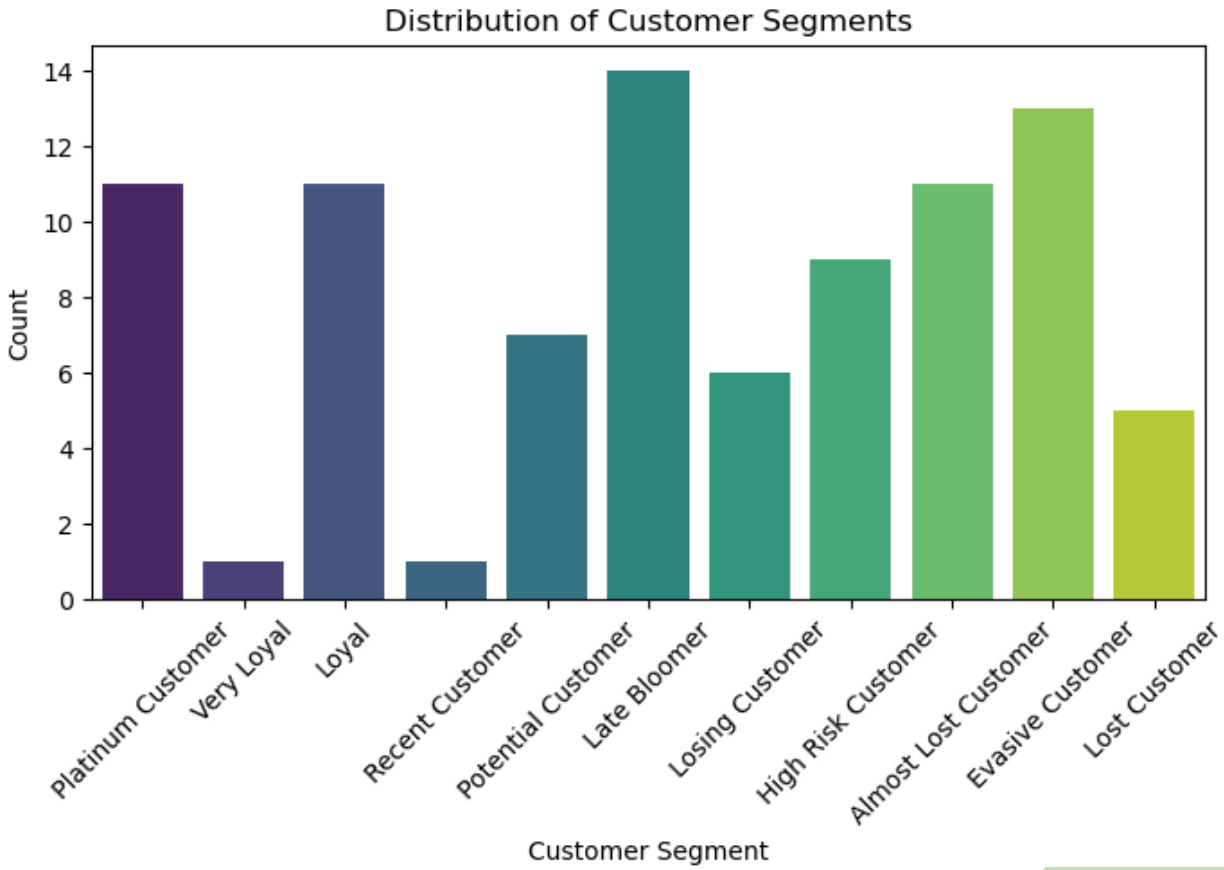
10.Evasive Customer: Customers with RFM scores between 112 and 123. They are actively disengaging or evading interactions with the brand, showing minimal activity.

11.Lost Customer: Customers with RFM scores below 112. These are customers who have not engaged recently and are considered lost or inactive.

Each category provides insights into customer behavior and helps in tailoring specific marketing strategies or retention efforts to maximize customer lifetime value and overall profitability. The categorization allows businesses to prioritize their efforts effectively based on the current engagement and loyalty levels of their customer base.

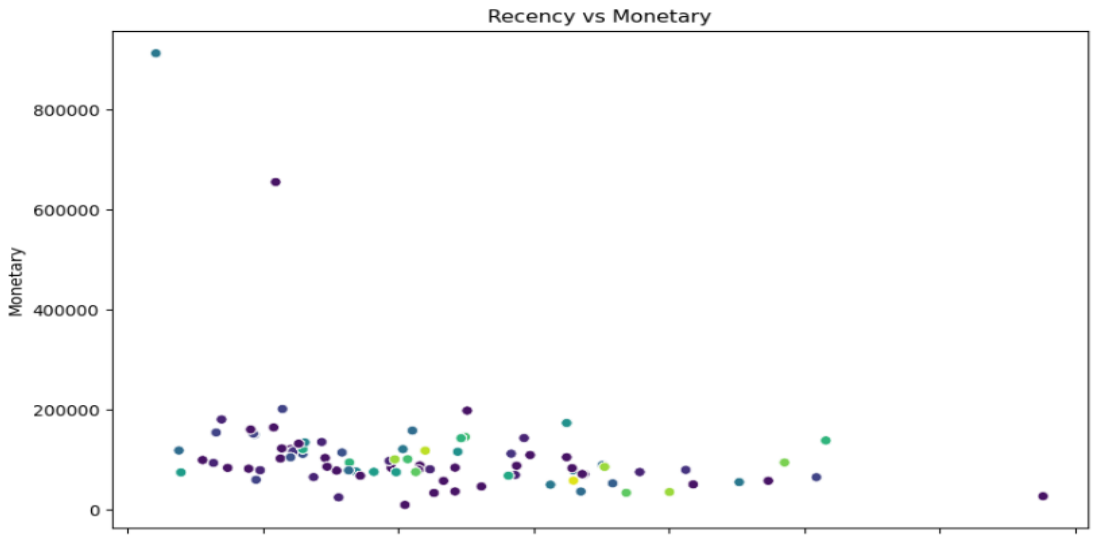
INFERENCES AND RECOMMENDATION

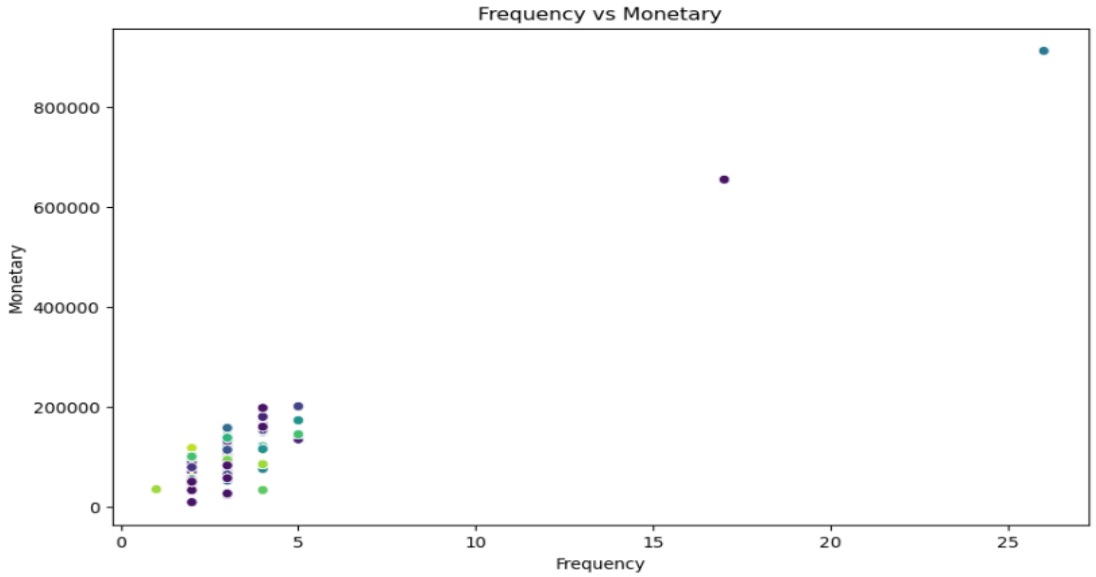
Customer Demographics based on RFM:

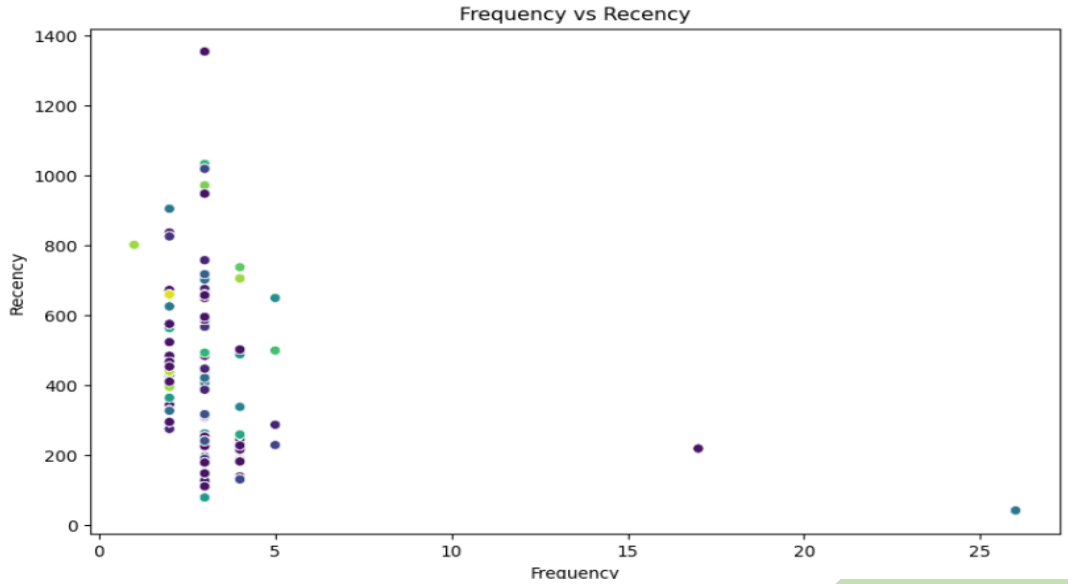


- **Late Bloomer (14 customers)**: Customers showing recent engagement improvements, ripe for targeted campaigns to boost loyalty.
- **Evasive Customer (13 customers)**: Minimal engagement; strategies needed to understand and address their disinterest.
- **Platinum Customer (11 customers)**: Top-tier, highly engaged customers warranting personalized attention to maintain loyalty and maximize revenue.
- **Loyal (11 customers)**: Potential for increased loyalty; strategies to nurture these customers can accelerate their transition to Platinum status.

Recency vs Frequency vs Monetary







There are few outliers in Frequency and monetary suggesting Customers with unusually high purchase frequencies and Customers with exceptionally high spending levels compared to others are present in the dataset

Best 5 Customers

Best Customers:

	CUSTOMERNAME	Recency_Score	Frequency_Score	Monetary_Score
3	Anna's Decorations, Ltd	4	4	4
6	Australian Collectors, Co.	4	4	4
27	Diecast Classics Inc.	4	4	4
32	Euro Shopping Channel	4	4	4
43	La Rochelle Gifts	4	4	4

The top 5 best customers are determined by calculating the RFM score, which is the sum of scores for Monetary, Recency, and Frequency. Higher RFM scores (444) indicate more valuable customers.

Customers on the Verge of Churning:

Customers on the Verge of Churning:

	CUSTOMERNAME	Recency_Score	Frequency_Score	Monetary_Score
13	Blauer See Auto, Co.	1	4	2
15	CAF Imports	2	1	1
16	Cambridge Collectables Co.	2	1	1
25	Daedalus Designs Imports	2	1	1
30	Dragon Souveniers, Ltd.	1	4	4

The customers with very low to low recency, high to moderate monetary value, and high to moderate purchase frequency are at an increased risk of churning

Top 5 Lost Customers

Lost Customers:

	CUSTOMERNAME	Recency_Score	Frequency_Score	Monetary_Score
12	Bavarian Collectables Imports, Co.	1	1	1
20	Clover Collections, Co.	1	1	1
29	Double Decker Gift Stores, Ltd	1	1	1
40	Iberia Gift Imports, Corp.	1	1	1
70	Signal Collectibles Ltd.	1	1	1

- Customers with low frequency of purchases, indicating reduced engagement.
- Customers with low recency, implying prolonged inactivity.
- Customers with low monetary value, suggesting decreased spending.

Top 5 Loyal Customers

Loyal Customers:

	CUSTOMERNAME	Recency_Score	Frequency_Score	Monetary_Score
7	Australian Gift Network, Co	4	2	1
9	Auto Canal Petit	4	2	3
11	Baane Mini Imports	4	4	3
22	Collectables For Less Inc.	4	2	2
33	FunGiftIdeas.com	4	2	3

Loyal customer refers to customers who are consistent and repeat purchases over time. They may not necessarily have the highest monetary value, but they demonstrate a strong commitment to the brand by repeatedly choosing to do business with the company.

.

Recommendations:

- **Platinum Customers and Very Loyal Customers**: Offer exclusive benefits, personalized experiences, and VIP treatment to strengthen loyalty and increase advocacy.
- **Loyal and Potential Customers**: Implement targeted nurturing campaigns with incentives for repeat purchases, personalized recommendations, and proactive customer support to solidify loyalty.
- **Almost Lost and High Risk Customers**: Launch targeted retention campaigns with personalized offers, proactive customer support, and incentives to prevent churn and re-engage these at-risk segments.
- **Evasive and Lost Customers**: Use targeted re-engagement tactics, such as personalized offers, reminders, and surveys to understand and address concerns that lead to avoidance behaviors.
- **Late Bloomers and Recent Customers**: Foster ongoing engagement with educational content, product trials, and personalized recommendations to capitalize on their recent increase in activity.
- **Continuous Monitoring:** Regularly review and update customer segmentation strategies based on evolving customer behavior and market dynamics

By implementing these recommendations, businesses can effectively nurture customer relationships, mitigate churn risks, and optimize lifetime customer value across diverse segments of their customer base.

THANK YOU

Sayyed Abdul Khaliq

abdulkhaliq01112001@gmail.com