

MRA Project

Business Report

Made By -
Soni Kumari / PGPDSBA.O.FEB23.A

Customer Insights: Data-Driven Strategies for Automobile Parts Manufacturing

Agenda :

In this project, we aim to leverage three years of transaction data to uncover customer buying patterns and provide tailored marketing strategies for an automotive parts manufacturing company. Our agenda covers the project scope, data overview, exploratory data analysis, customer segmentation, buying behavior analysis, and creation of customer profiles. The presentation will conclude with key insights, actionable recommendations, and an implementation plan, emphasizing the potential impact on the company's growth.

Executive Summary Of Data :

The dataset provides a comprehensive view of the sales transactions for an automotive parts company. Each order is uniquely identified by an ORDERNUMBER, with details such as the QUANTITYORDERED and PRICEEACH columns quantifying the number of items and their individual prices. The SALES column reflects the total revenue generated per order. ORDERDATE captures the date of order placement, while DAYS_SINCE_LASTORDER allows us to analyze customer purchasing patterns based on the time elapsed since their previous order. STATUS indicates the current status of each order, offering insights into operational efficiency. PRODUCTLINE categorizes items into specific product lines, and MSRP provides the Manufacturer's Suggested Retail Price. The dataset also includes information about customers, such as CUSTOMERNAME, PHONE, ADDRESSLINE1, CITY, POSTALCODE, COUNTRY, CONTACTLASTNAME, and CONTACTFIRSTNAME. Lastly, DEALSIZE categorizes orders into "Small," "Medium," or "Large," enabling a holistic understanding of deal sizes. Analyzing this data will unveil valuable insights into customer behavior, product performance, and overall business dynamics.

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.

Know Your Data

- Data has dimension of 2747 rows and 20 columns
- None of the column has a null value in it
- Below are the datatypes of corresponding columns in the dataset

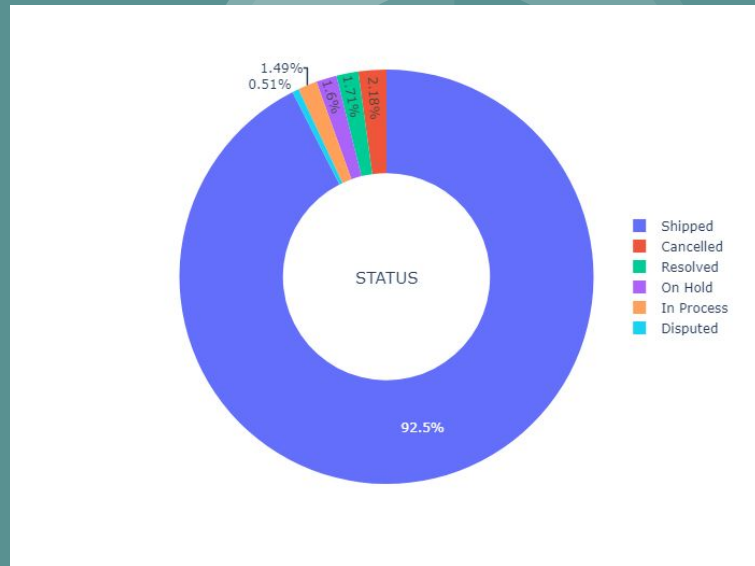
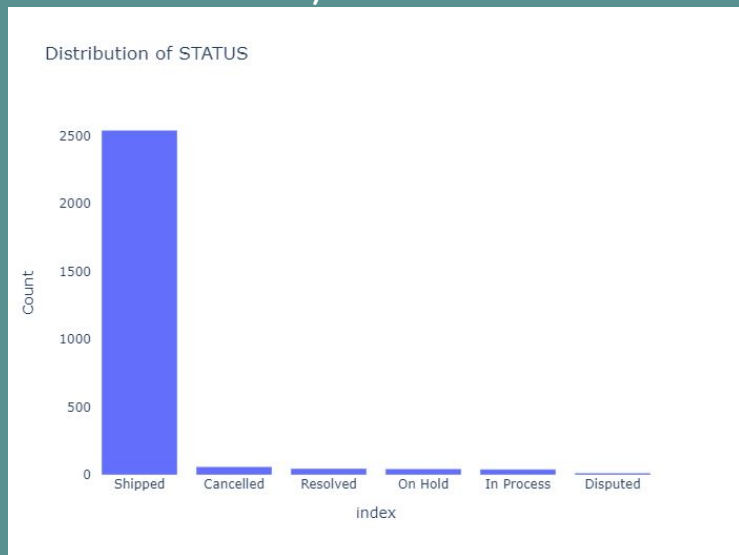
Data columns (total 20 columns):

#	Column	Non-Null Count	Dtype
0	ORDERNUMBER	2747 non-null	int64
1	QUANTITYORDERED	2747 non-null	int64
2	PRICEEACH	2747 non-null	float64
3	ORDERLINENUMBER	2747 non-null	int64
4	SALES	2747 non-null	float64
5	ORDERDATE	2747 non-null	int64
6	DAYS_SINCE_LASTORDER	2747 non-null	int64
7	STATUS	2747 non-null	object
8	PRODUCTLINE	2747 non-null	object
9	MSRP	2747 non-null	int64
10	PRODUCTCODE	2747 non-null	object
11	CUSTOMERNAME	2747 non-null	object
12	PHONE	2747 non-null	object
13	ADDRESSLINE1	2747 non-null	object
14	CITY	2747 non-null	object
15	POSTALCODE	2747 non-null	object
16	COUNTRY	2747 non-null	object
17	CONTACTLASTNAME	2747 non-null	object
18	CONTACTFIRSTNAME	2747 non-null	object
19	DEALSIZE	2747 non-null	object

dtypes: float64(2), int64(6), object(12)

Exploratory Data Analysis

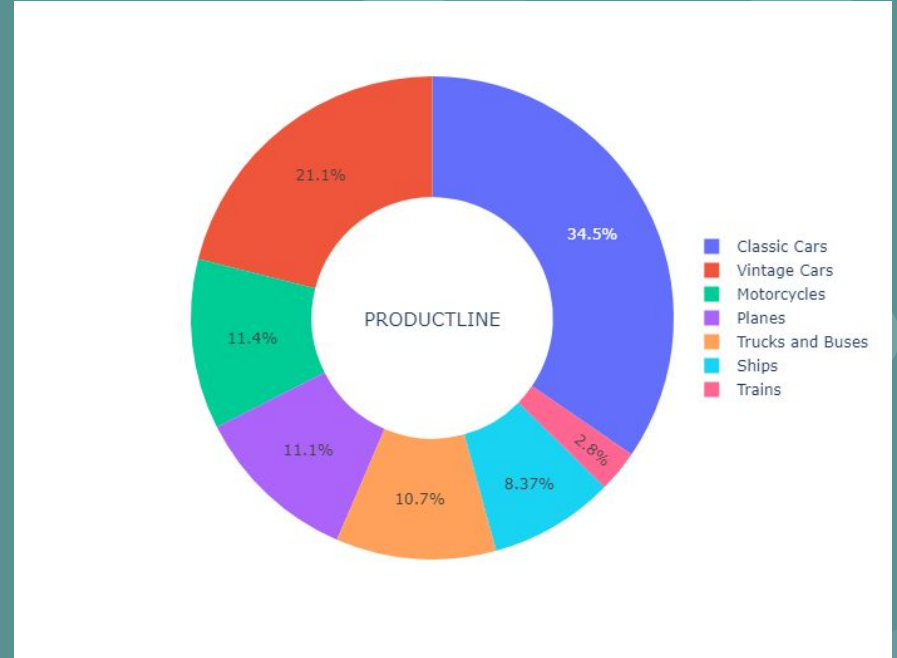
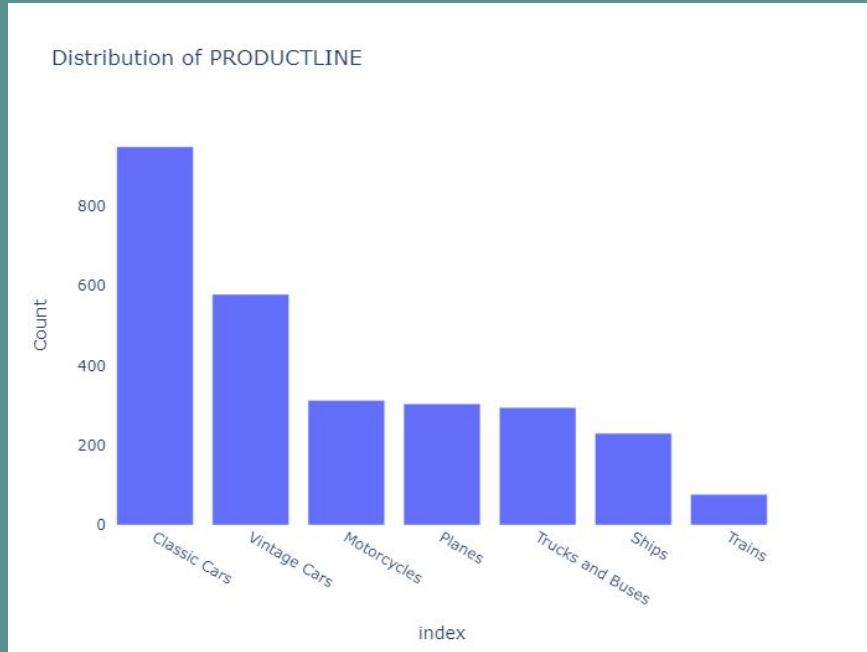
Univariate Analysis



Analysis :

- Graphs suggest that the majority of orders fall under the "Shipped" status, indicating successful completion and delivery of a significant portion of transactions.
- "Cancelled" orders are relatively lower, suggesting a generally stable and well-managed order fulfillment process.
- The "Resolved" status signifies successful resolution of certain issues. A higher count in this category suggests a proactive approach to addressing customer concerns, potentially contributing to overall customer satisfaction.

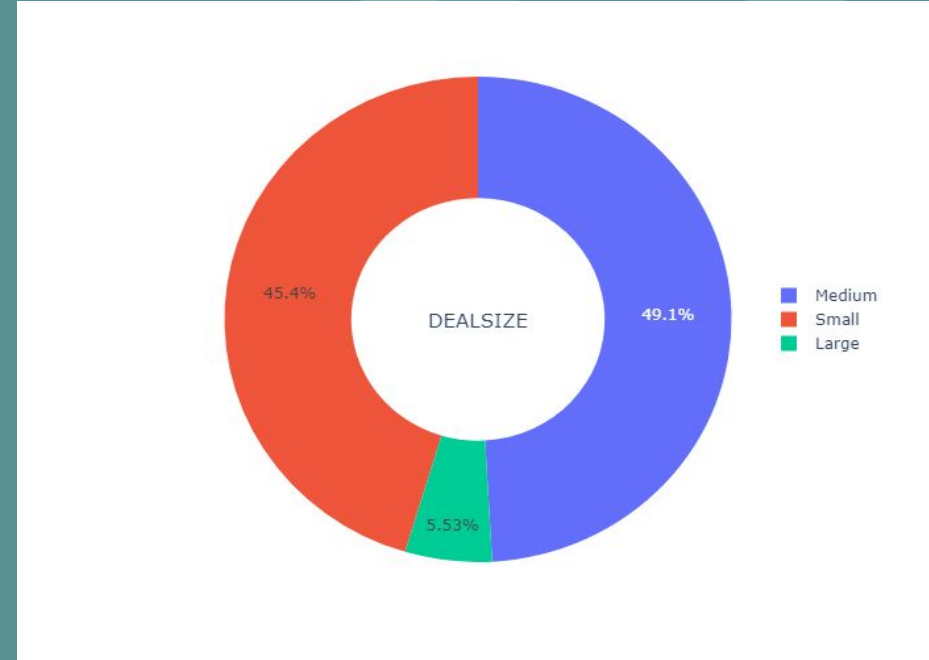
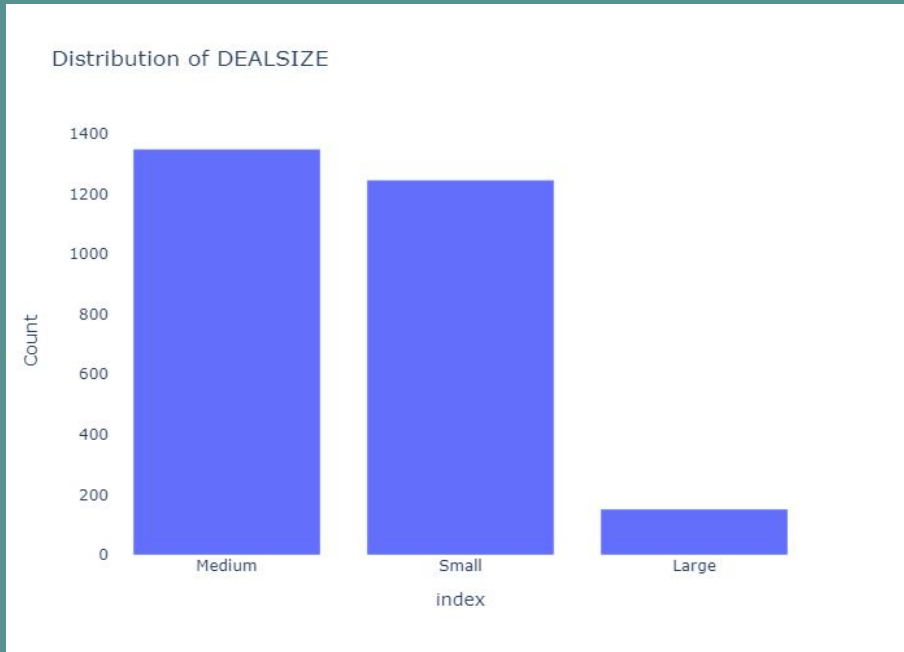
Univariate Analysis Ctnd..



Analysis :

- Graphs suggest that "Classic Cars" as the most popular product category, as it has the highest count. This indicates a strong demand for classic automotive products.
- The distribution across various categories ("Motorcycles," "Planes," "Trucks and Buses," "Ships," etc.) suggests a diverse customer base with interests in different types of vehicles.

Univariate Analysis Ctnd..



Analysis :

- Graphs suggest that the most common deal size by observing which category ("Medium," "Small," or "Large") has the highest count. This indicates the typical transaction size for the company and medium deal size is at the top in this .
- Although "Medium" deals may be the most common, assess whether "Small" deals significantly contribute to overall revenue.

Univariate Analysis Ctnd..

Inferences & Takeaway from Univariate Analysis:

Order Status Distribution:

- "Shipped" orders dominate, indicating a high success rate in order fulfillment, contributing to positive customer experiences.
- The relatively low count of "Disputed" orders suggests a generally smooth resolution process and a low incidence of customer dissatisfaction.
- "Cancelled" orders are present but in a lower count, pointing to effective measures in managing and minimizing cancellations.

Product Category Distribution:

- "Classic Cars" stand out as the most popular product category, showcasing a strong market demand for nostalgic automotive products.
- Niche categories like "Planes," "Ships," and "Trains" have lower counts, suggesting potential growth opportunities for targeted marketing and product development.
- Diverse product preferences across categories highlight the company's ability to cater to a broad range of customer interests.
-

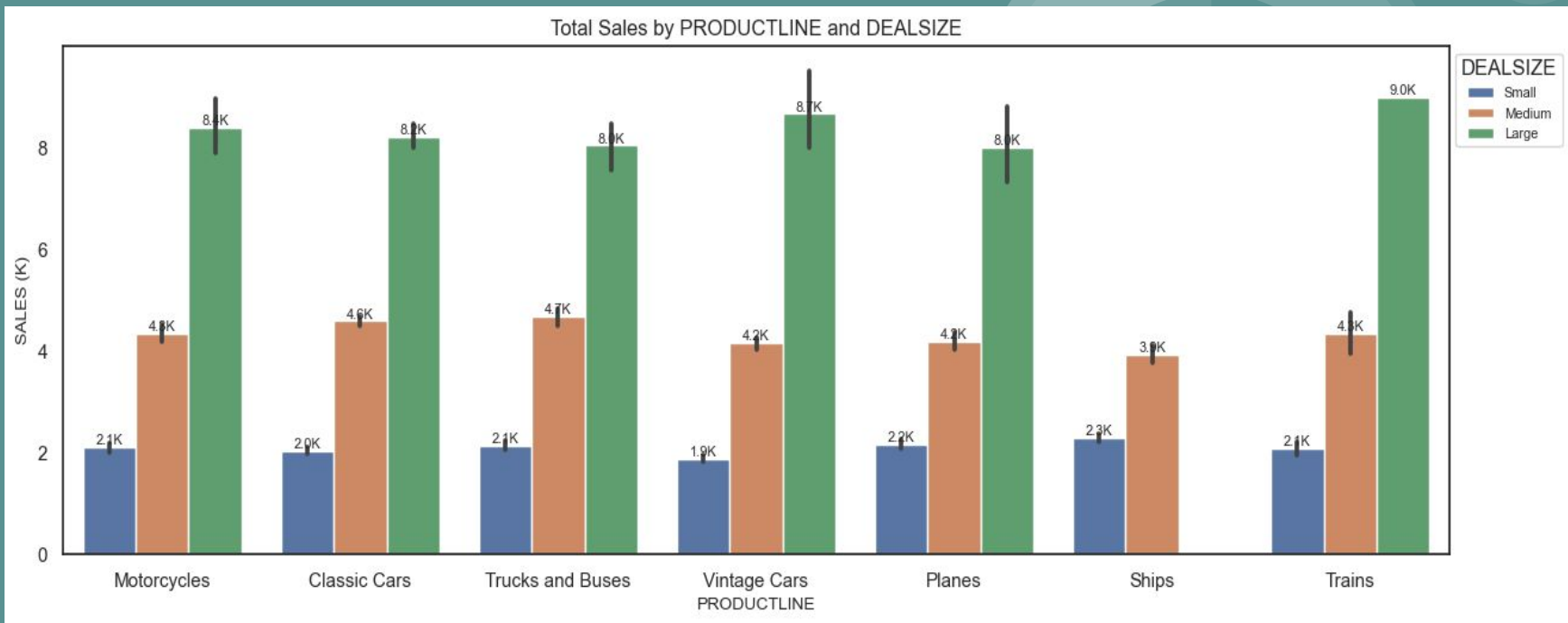
Deal Size Distribution:

- "Medium" deals are the most common, indicating a prevalent transaction size that aligns with the majority of customer preferences.
- The distribution suggests that while "Medium" deals are common, "Small" deals may play a crucial role in contributing to overall revenue.
- The presence of all three deal sizes signifies a well-balanced product offering that caters to customers with varying preferences and budgets.

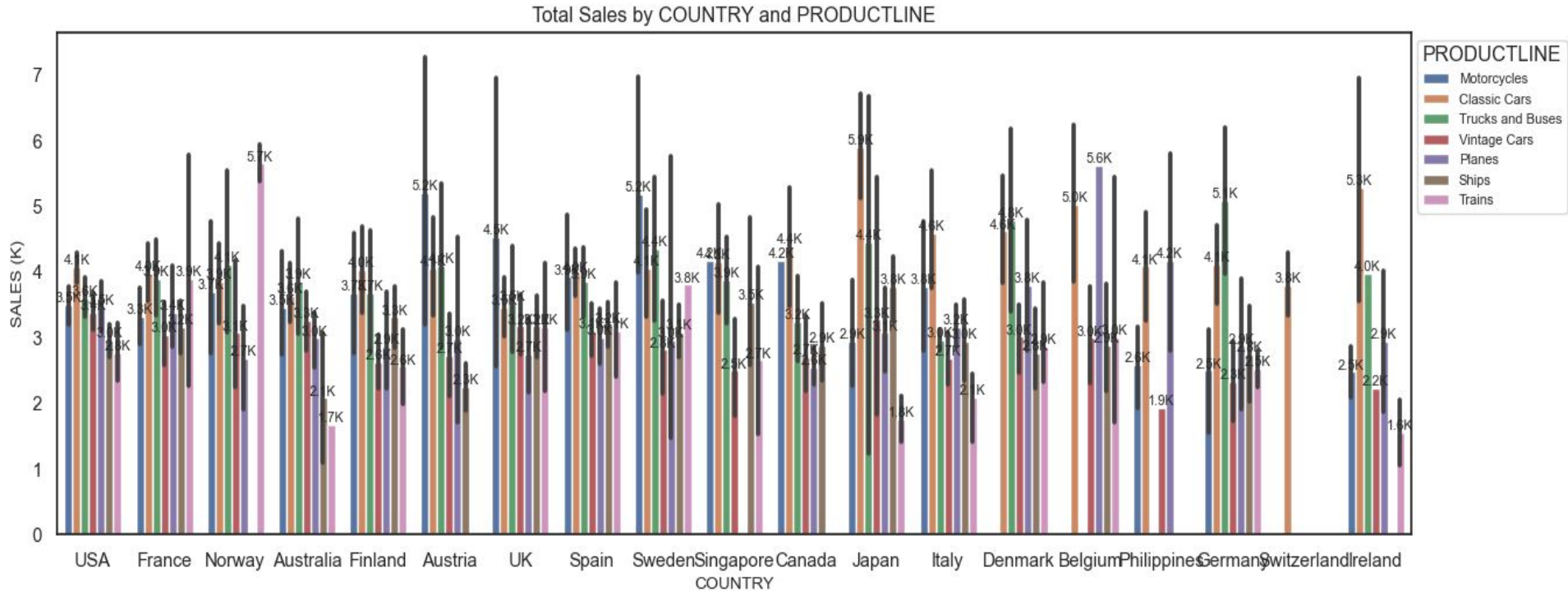
Bivariate Analysis



Multivariate Analysis



Multivariate Analysis Ctnd .

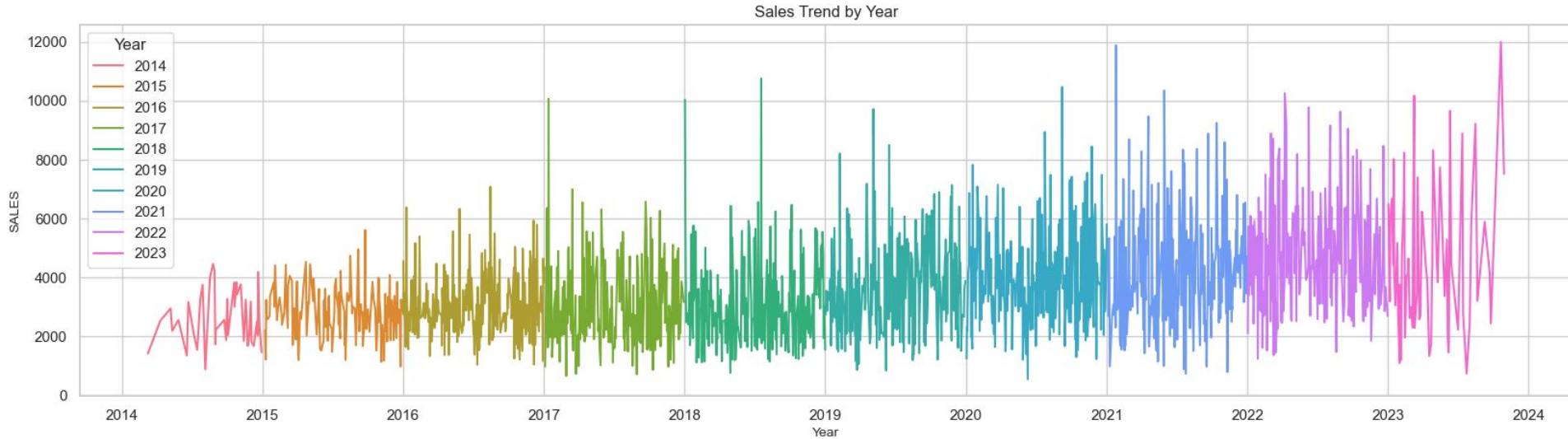


Multivariate Analysis Ctnd .

Inferences & Takeaway from Univariate Analysis:

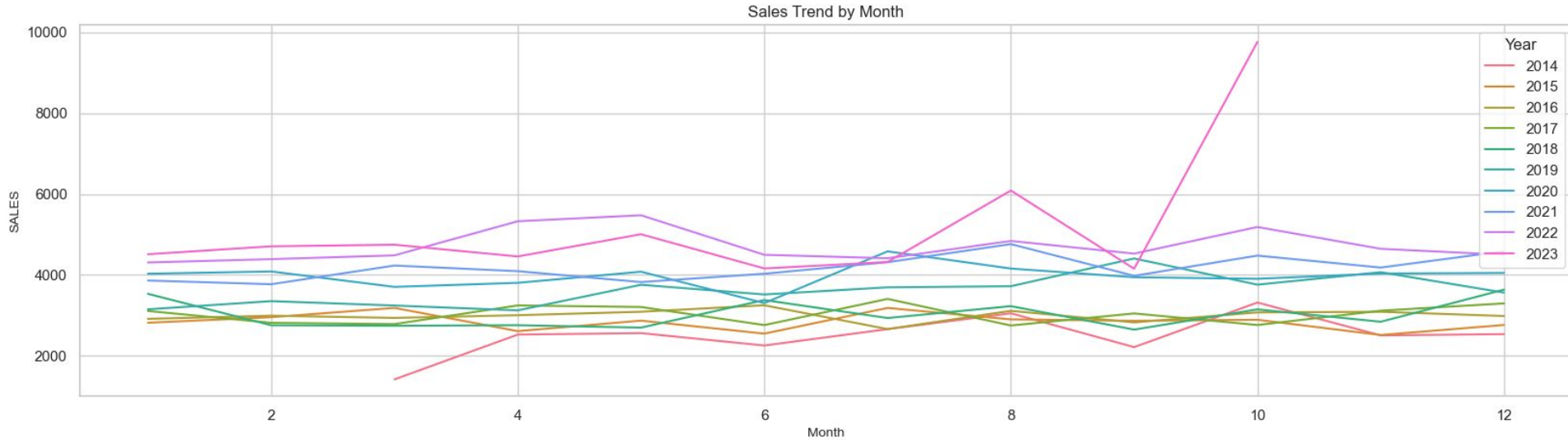
- Sales are highest in the United States, with Spain and France following closely. Noteworthy sales are also observed in Australia and Singapore, representing strong market presence in the Eastern Hemisphere.
- A notable surge in sales is evident in both the United States and France during the year 2019.

Weekly, Monthly, Quarterly, Yearly Trends in Sales and Sales Across different Categories of different features



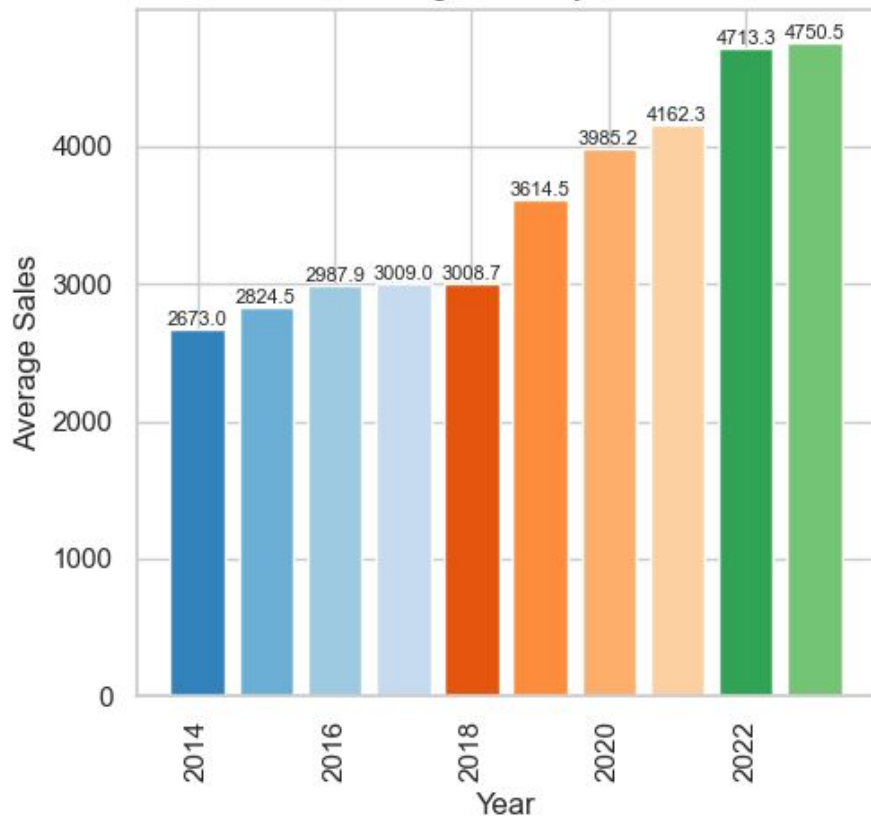
Tables Across different Categories of

Weekly, Monthly, Quarterly, Yearly Trends in Sales and Sales Across different Categories of different features Ctned .



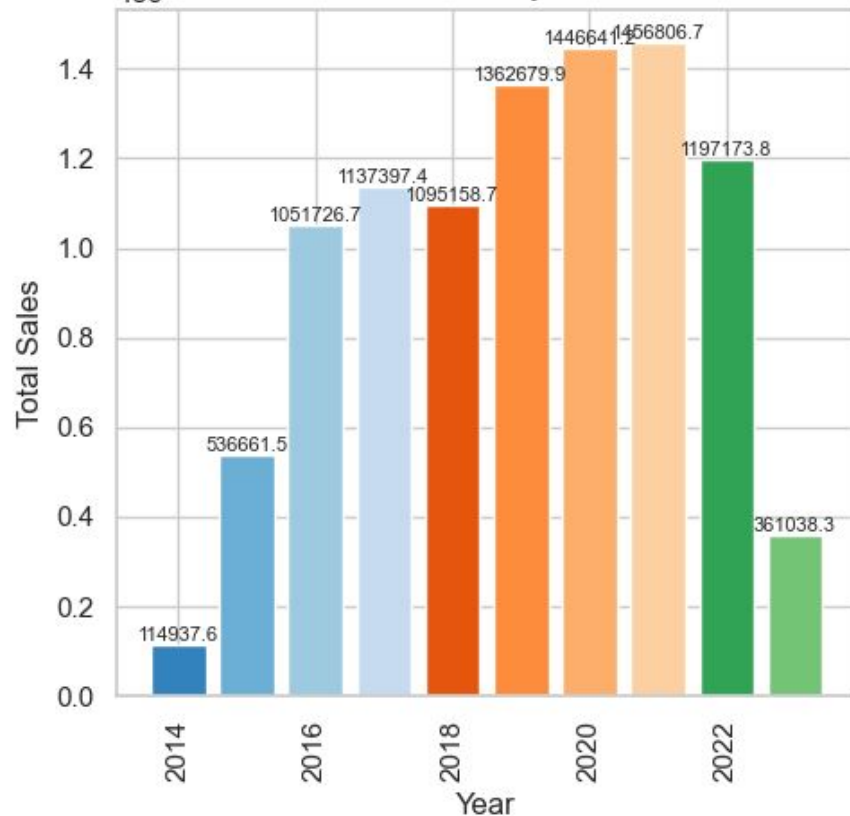
Sales Distribution

Average Sales by Year



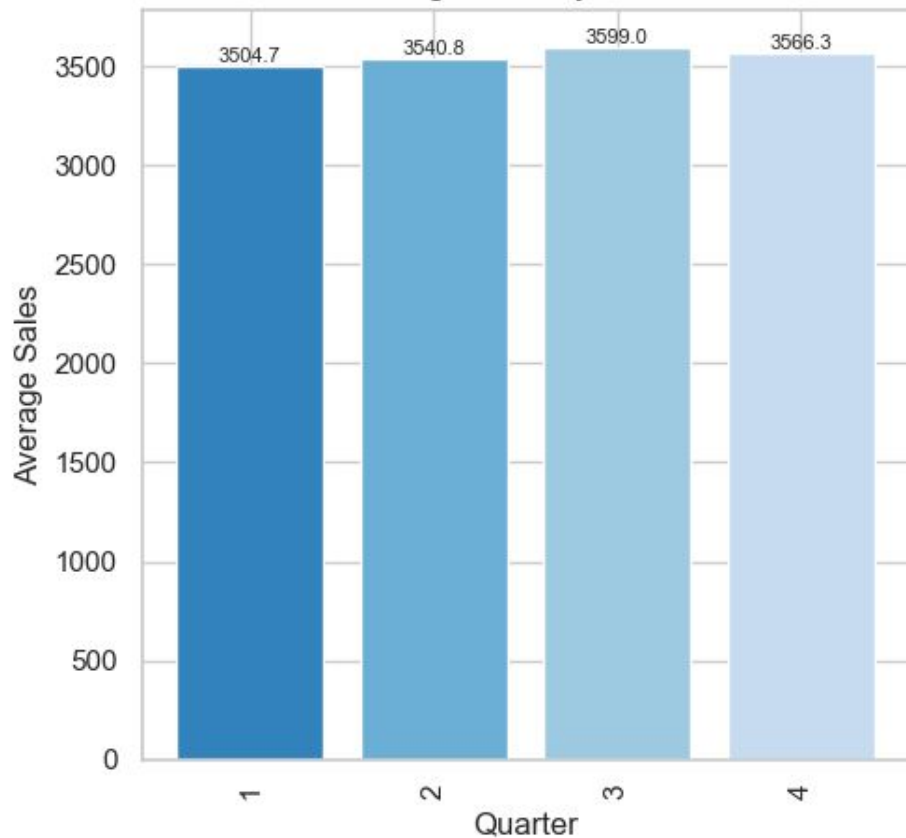
1e6

Total Sales by Year

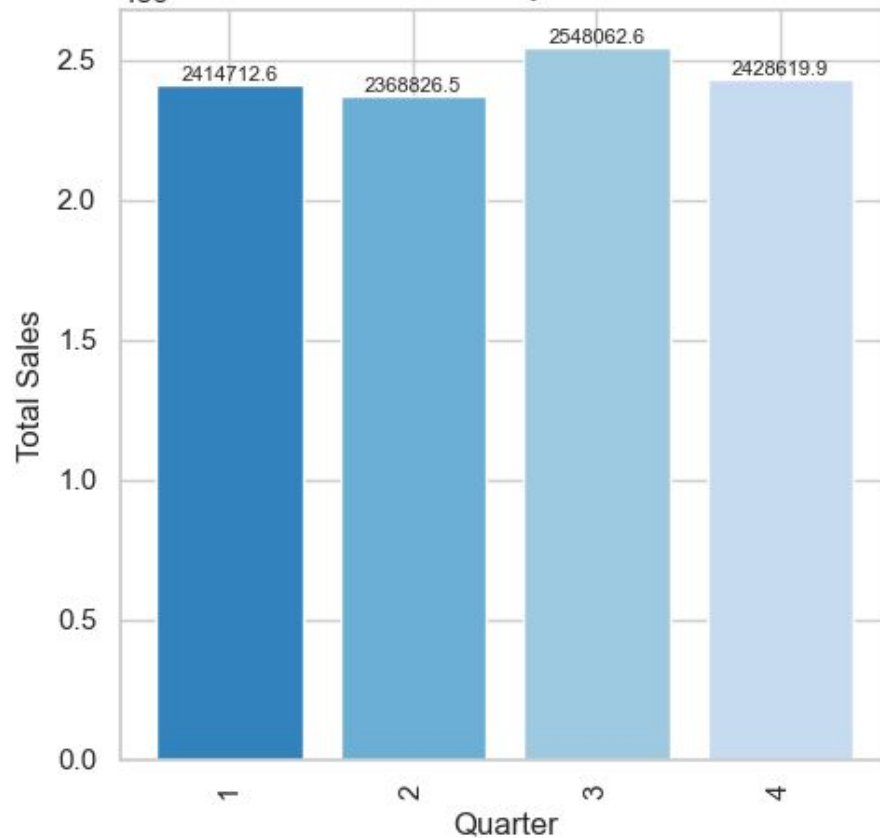


Sales Distribution Ctnd.

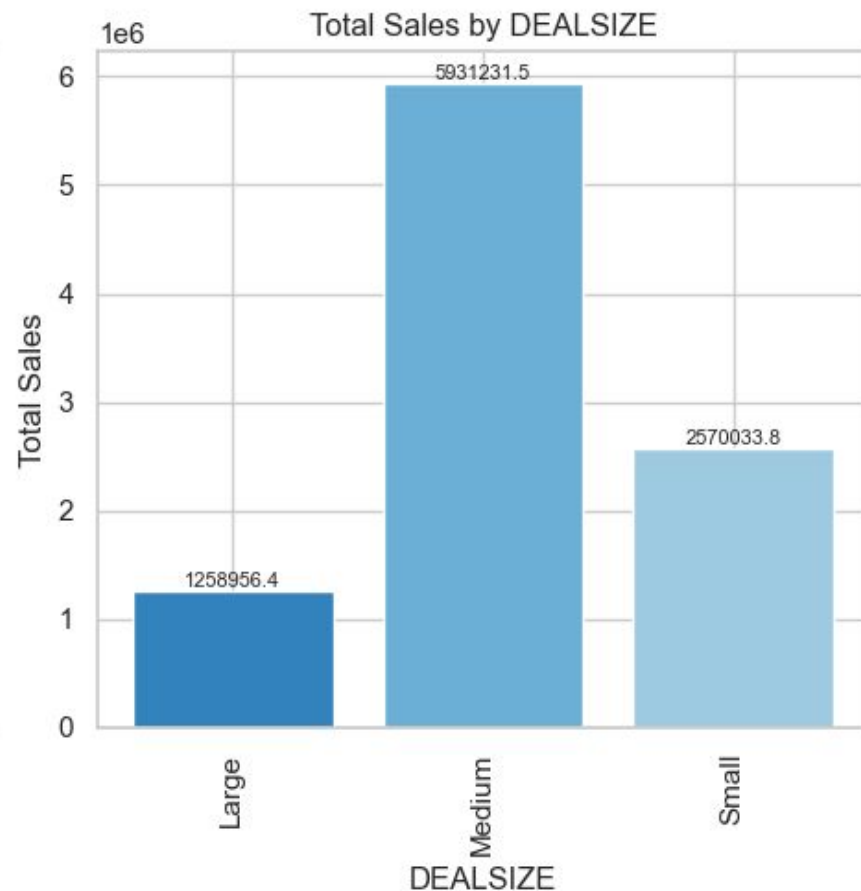
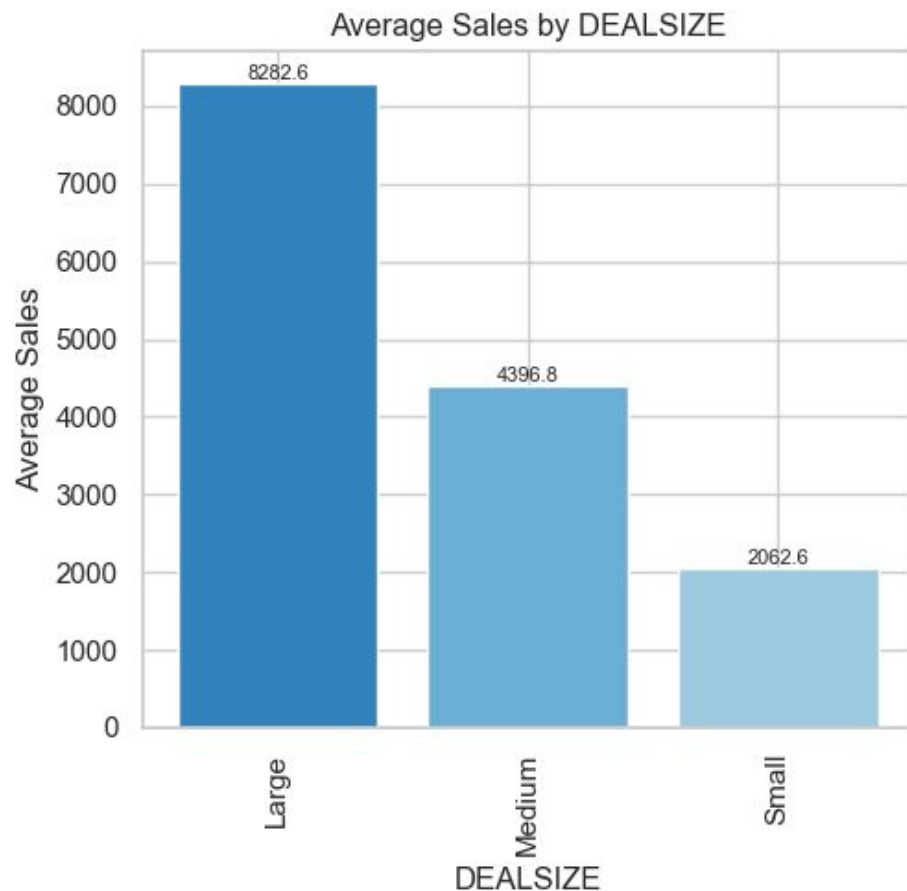
Average Sales by Quarter



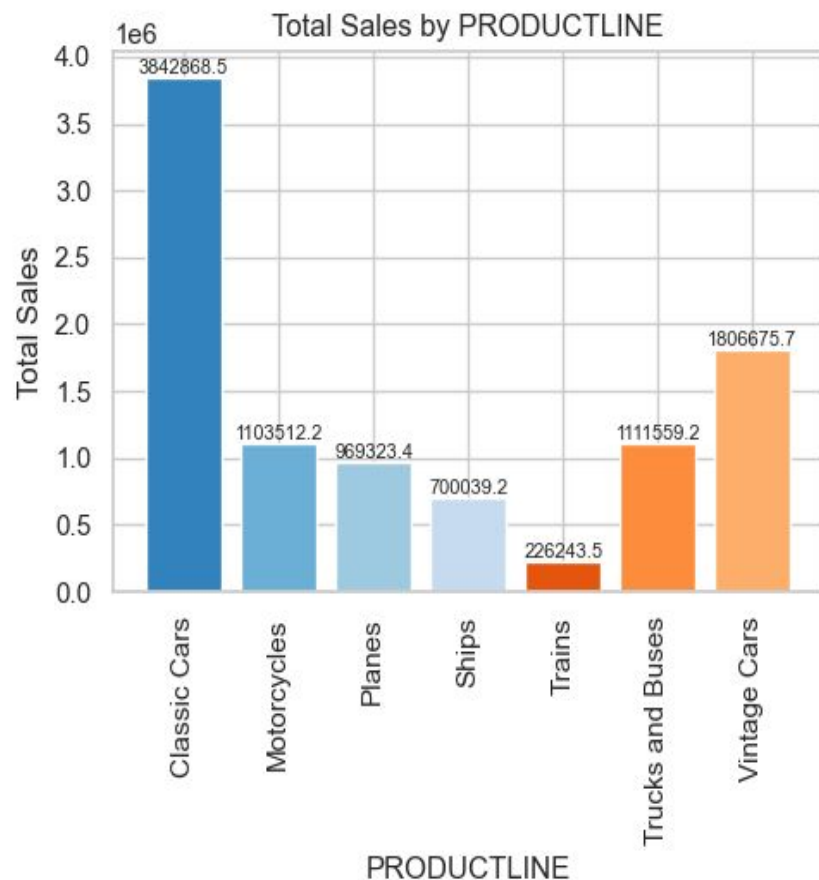
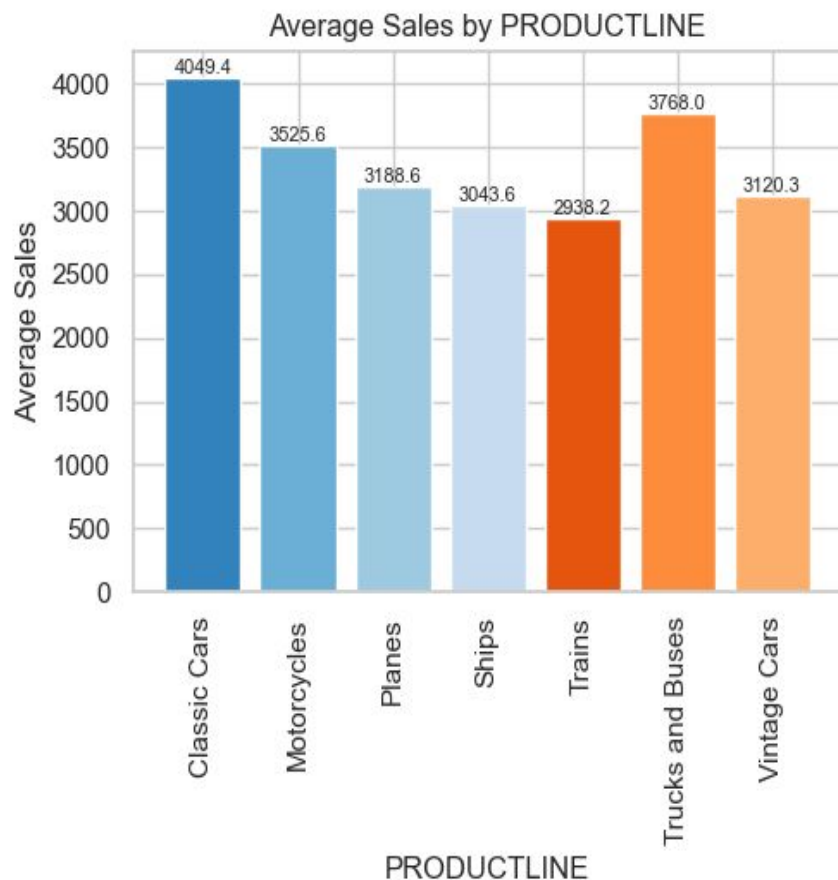
Total Sales by Quarter



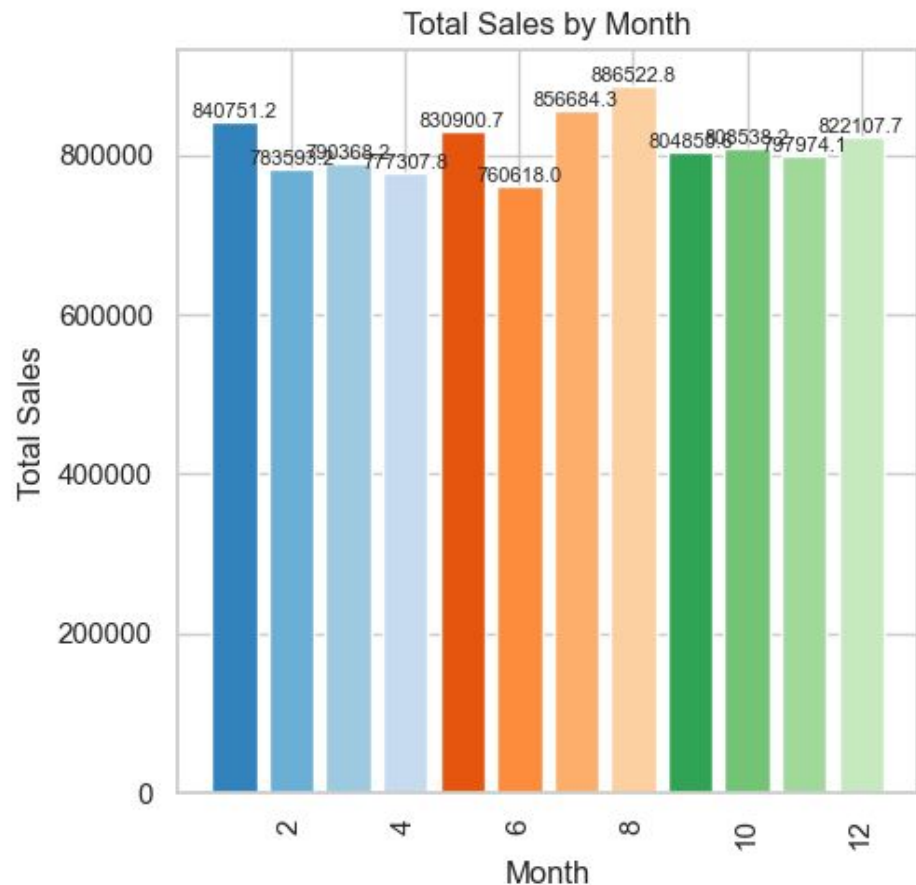
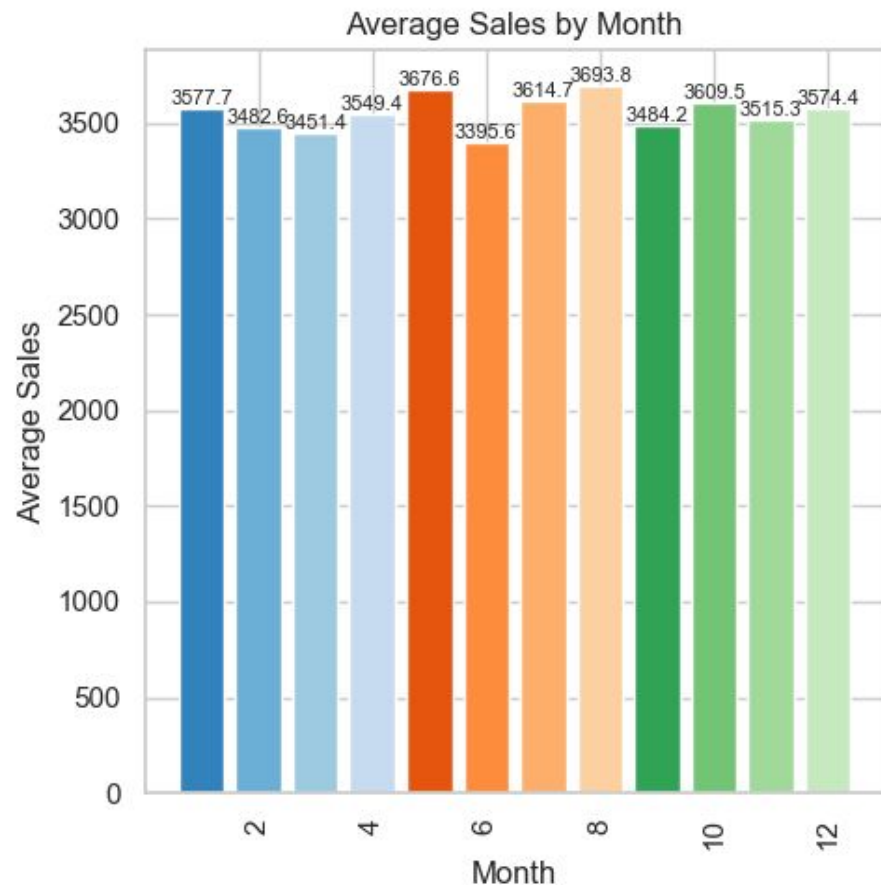
Sales Distribution Ctnl.



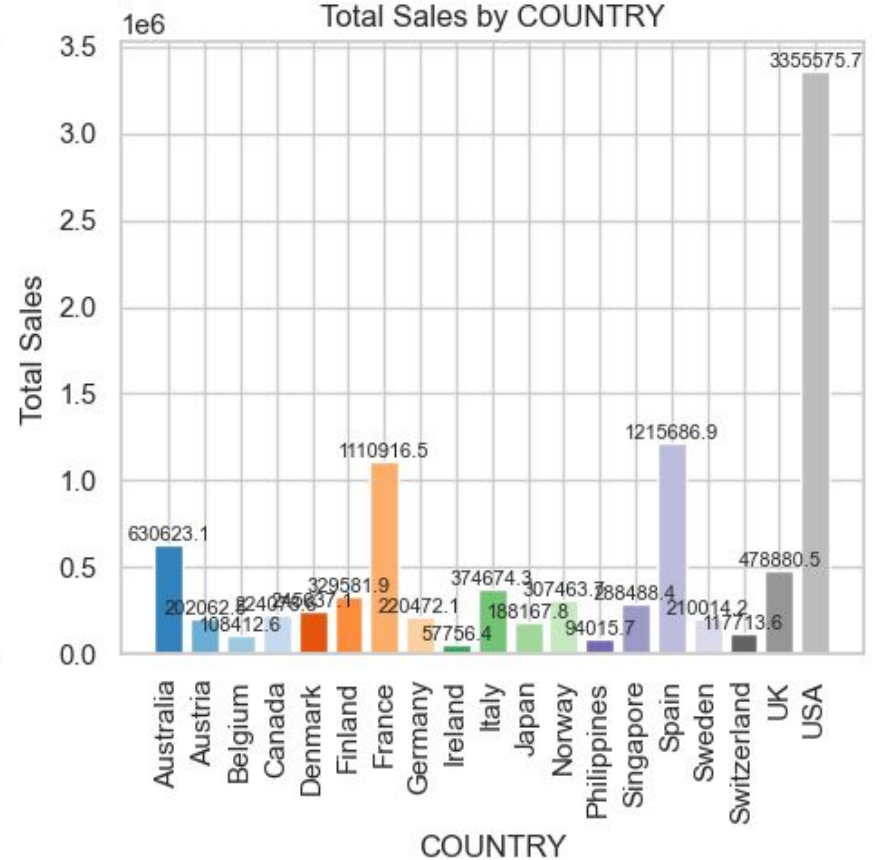
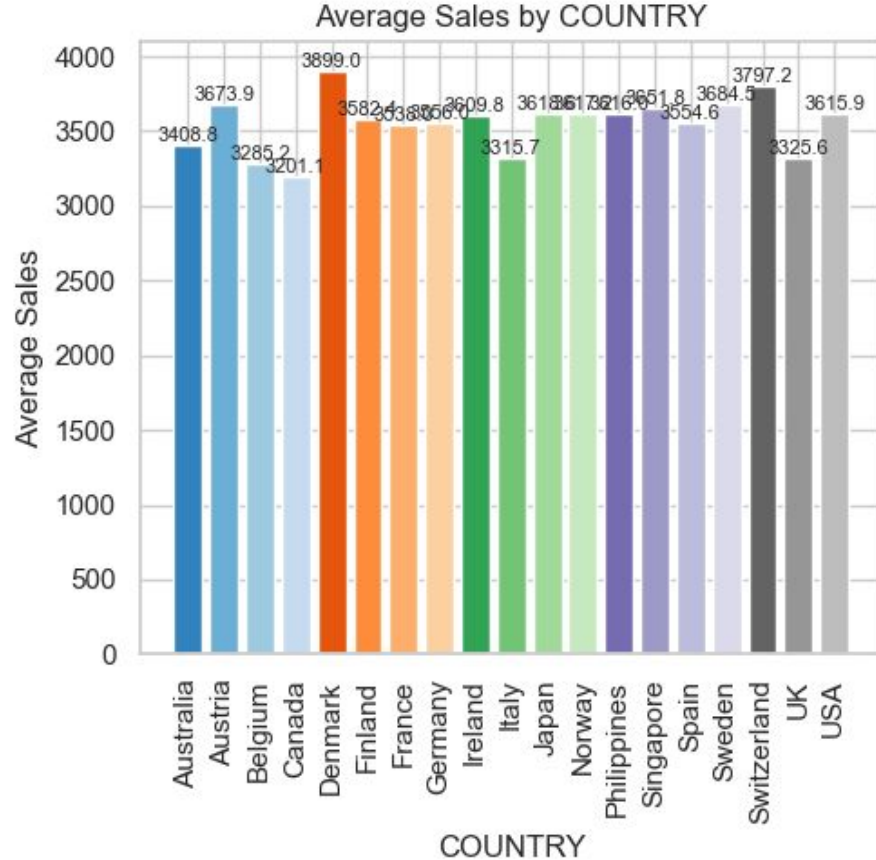
Sales Distribution Ctnl.



Sales Distribution Ctnl.



Sales Distribution Cntd.



Sales Distribution and EDA Analysis/Inferences :

Inferences & Takeaway

- The United States leads in sales, closely followed by Spain and France, while Australia and Singapore contribute significantly in the Eastern Hemisphere. In terms of city-wise performance, Madrid leads in sales, trailed by San Diego and New York City. Notably, there was a decline in sales from Australia and Singapore in the Eastern Hemisphere, but new customers from Japan were added in 2019. The year 2019 witnessed a substantial increase in sales in the United States and France.
- The company manufactures 109 automobile parts across 7 product lines, priced between 26 and 252. Notably, Car, Truck, Bus, and Motorcycle parts collectively contribute to 80% of the company's overall revenue.
- The sales trends exhibit an upward trajectory with monthly seasonality, reaching peak sales in October and November annually. The early months of 2020 saw increased sales and heightened customer activity compared to previous years.

Concerns:

- A notable concern arises from the absence of new customers added since September 2019, warranting attention to customer acquisition strategies.
- The majority of deals fall into the medium and small size categories, typically involving 20 to 50 items per transaction.

RFM Analysis

Lets Know RFM first

RFM stands for Recency, Frequency, and Monetary Value, and it's a method used in marketing and customer relationship management. Here's a simple breakdown:

Recency (R):

- This looks at how recently a customer has made a purchase or engaged with your business. The idea is that more recent interactions may indicate a higher level of interest or responsiveness.

Frequency (F):

- Frequency measures how often a customer interacts with your business. Customers who make frequent purchases or engage with your products/services more often are considered valuable.

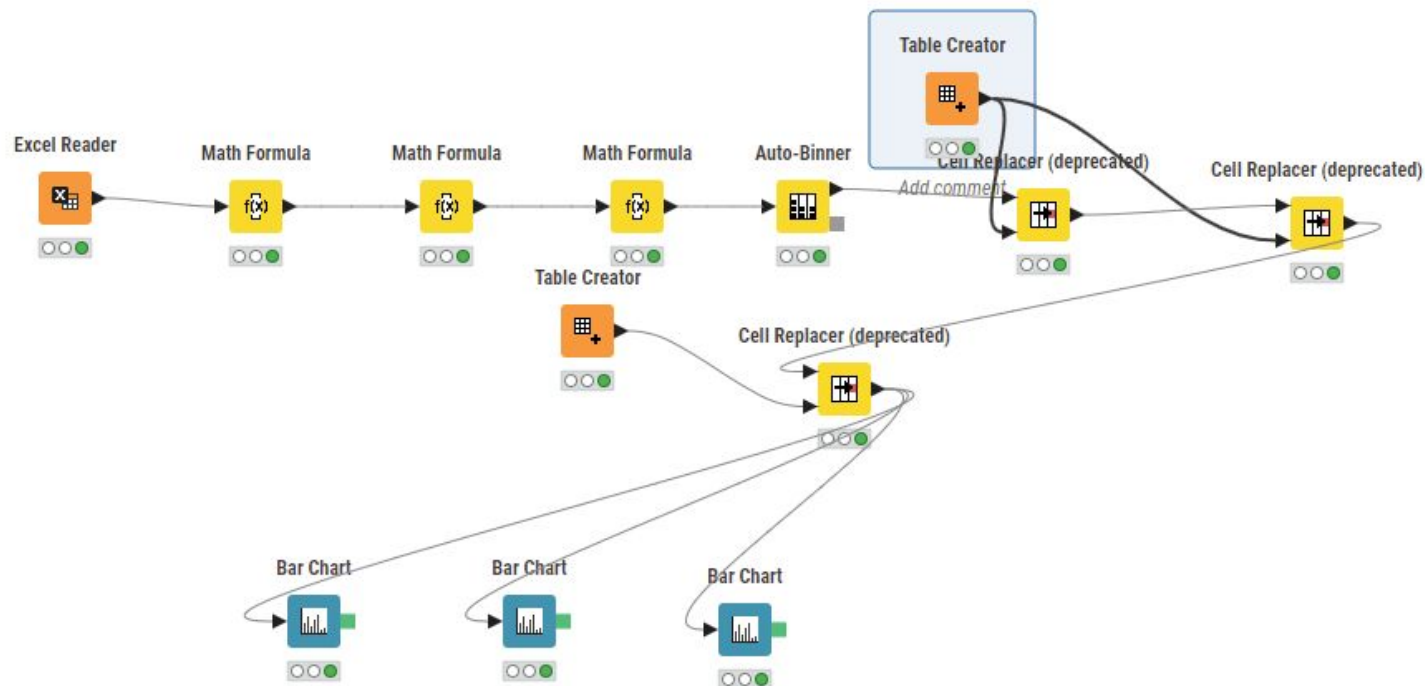
Monetary Value (M):

- Monetary Value focuses on how much money a customer has spent on your products or services. Customers who spend more are often considered more valuable to the business.

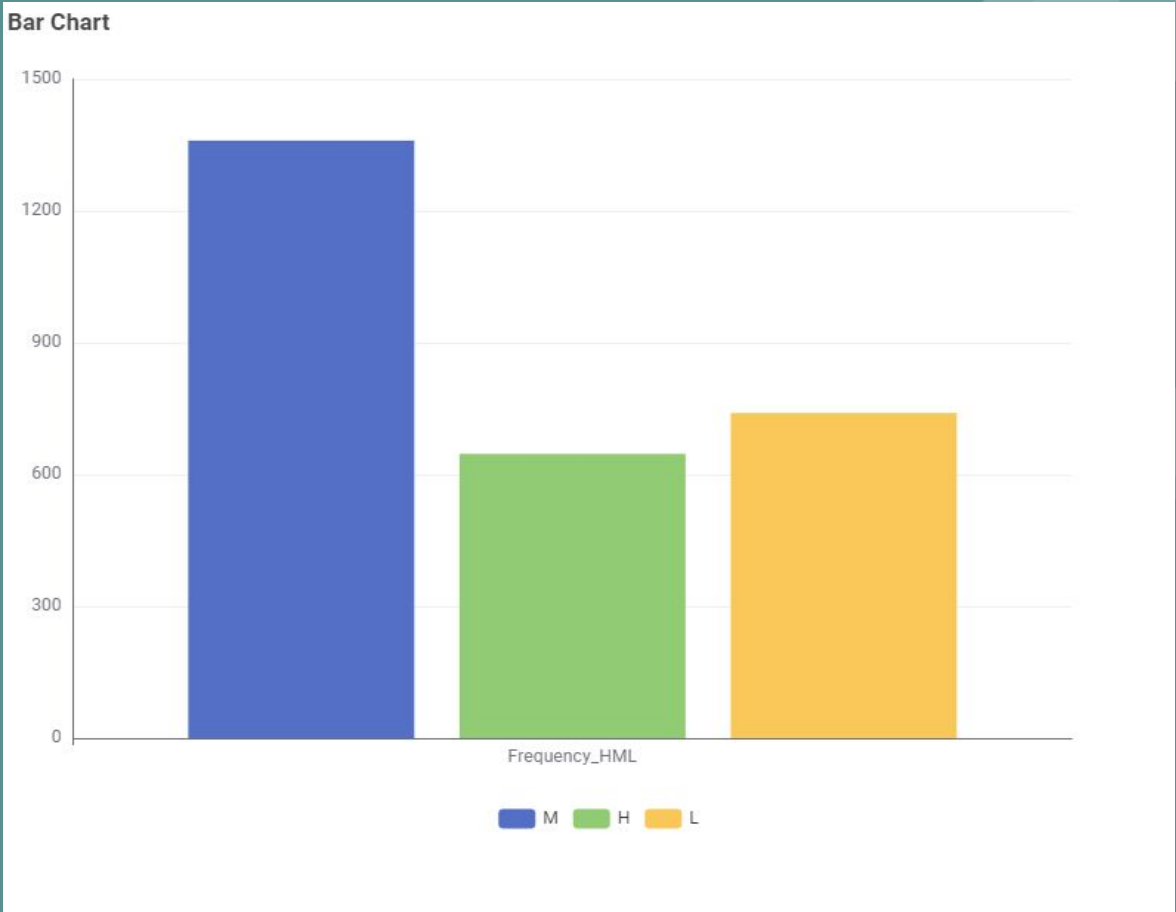
Parameters Used to do RFM

- DAYS_SINCE_LASTORDER is taken as Recency
- QUANTITYORDERED is taken as Frequency
- SALES is taken as Monetary
- ORDERNUMBER , RECENCY, FREQUENCY and MONETORY as taken into AutoBinner Node
- Two tables were created for Bin mapping to High low and medium .
- For FREQUENCY and MONETORY :
 - Bin 1 is mapped to L/Low , Bin 2 is mapped to M/Medium and Bin 3 is mapped to H/High
- Similarly it was opposite in case of RECENCY :
 - Bin 1 is mapped to H/High , Bin 2 is mapped to M/Medium and Bin 3 is mapped to L/Low

KNIME Workflow image for RFM



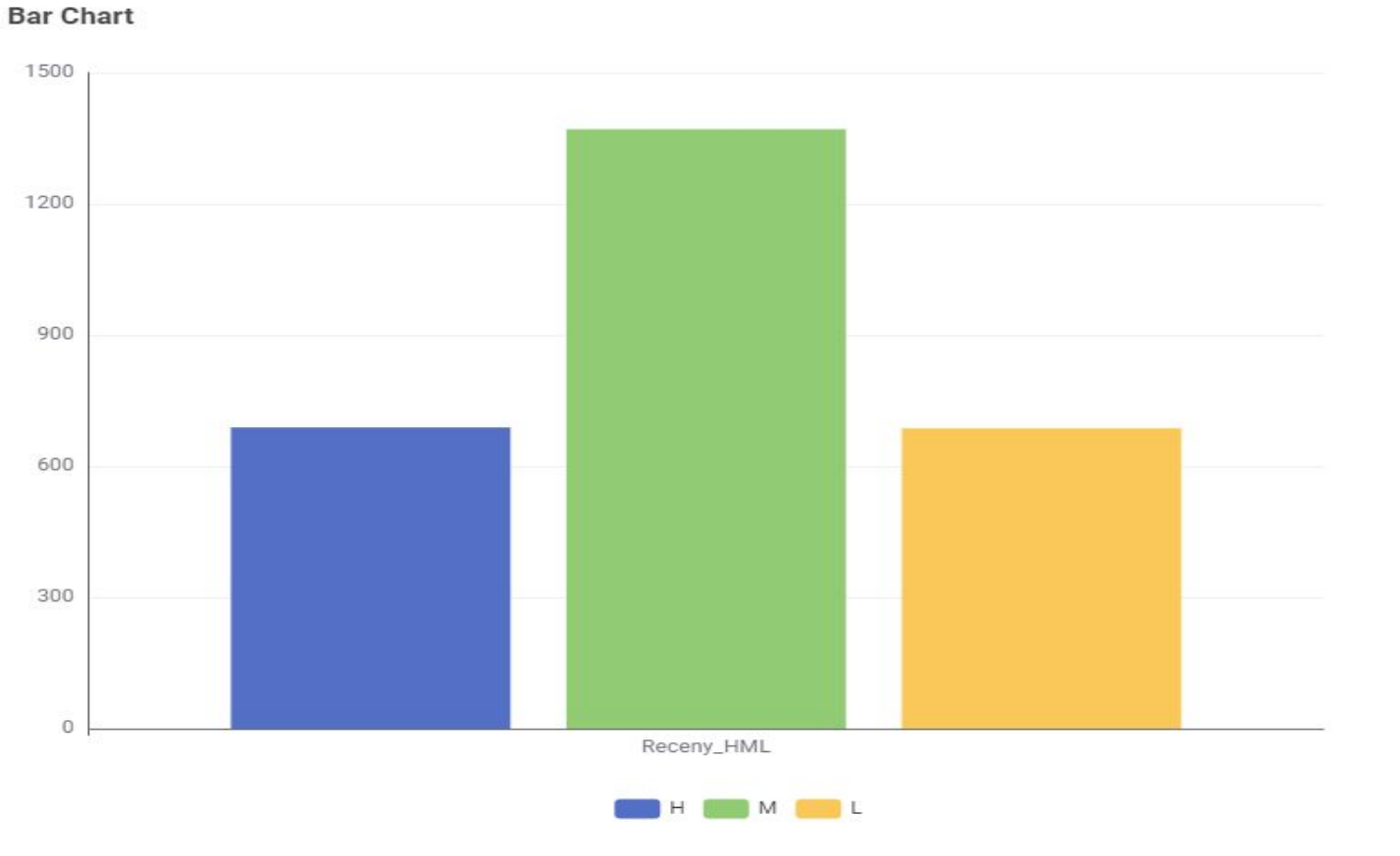
Output graphs from RFM KNIME



Output graphs from RFM KNIME CtnD .



Output graphs from RFM KNIME



Insights from RFM Analysis

Customer Demographics:

- The data suggests a diverse customer base from various countries, with the United States, Spain, and France leading in terms of sales.
- Notable sales are observed in Australia and Singapore, indicating a presence in the Eastern Hemisphere.

Order Status Analysis:

- The majority of orders are marked as "Shipped" (92.5%), highlighting a strong track record in successful order fulfillment.
- Instances of order cancellations ("Cancelled") are relatively low (2.18%), indicating effective management in preventing cancellations.

Product Category Preferences:

- "Classic Cars" emerge as the most preferred product category, contributing significantly to sales (34.55%).
- The diverse range of product categories, including "Vintage Cars," "Motorcycles," and "Planes," caters to varied customer interests.

Customer Transactions:

- Most customer transactions fall within the "Medium" and "Small" deal sizes (94.47%), with a smaller percentage in the "Large" category (5.53%).
- Analysis of transaction characteristics reveals that deals typically involve 20 to 50 items.

Insights from RFM Analysis

Key Customers:

- Euro Shopping Channel and Mini Gifts Distributors Ltd. stand out as the top customers, indicating their significant contribution to revenue.

Geographical Insights:

- Madrid leads in sales among cities, followed by San Rafael and NYC, providing insights into geographical sales distribution.
- The majority of customers are from the United States, Spain, and France, forming the core customer base.

Concerns and Opportunities:

- A notable concern is the absence of new customers added since September 2019, suggesting a need for focused customer acquisition strategies.
- Opportunities lie in exploring new markets or customer segments, especially in regions with lower sales representation.

Revenue and Product Line Analysis:

- The company manufactures 109 automobile parts across 7 product lines, with "Classic Car" parts generating the highest revenue (40%).
- Car, Truck, Bus, and Motorcycle parts collectively contribute to 80% of the company's overall revenue.

Sales Trends:

- The sales trends exhibit an upward trajectory with monthly seasonality, with peak sales observed in October and November annually.
- The early months of 2020 show increased sales and customer activity compared to previous years.

Customers on the Verge of Churning (Low Recency, Low Frequency, Low Monetary):

- Customers in "Recency_HML" bin labeled as "L" (Low Recency).
- Customers in "Frequency_HML" bin labeled as "L" (Low Frequency).
- Customers in "Monetary_HML" bin labeled as "L" (Low Monetary).
- Customers falling into the intersection of "L" bins across Recency, Frequency, and Monetary categories.
- Specific customer segments with lower RFM scores, indicating a potential decrease in engagement.

Some examples are below :

- La Corne D'abondance, Co.
- AV Stores, Co.
- Mini Gifts Distributors Ltd.
- Salzburg Collectables
- Norway Gifts By Mail, Co.
- Mini Creations Ltd.
- Alpha Cognac
- Enaco Distributors

Lost Customers (Low Recency, Low Frequency, Low Monetary):

- Customers in "Recency_HML" bin labeled as "L" (Low Recency).
- Customers in "Frequency_HML" bin labeled as "L" (Low Frequency).
- Customers in "Monetary_HML" bin labeled as "L" (Low Monetary).
- Customers falling into the intersection of "L" bins across Recency, Frequency, and Monetary categories.
- Customers with consistently low engagement across all three metrics, possibly indicating lost interest.

Below are few examples

- Auto Canal Petit
- Danish Wholesale Imports
- Land of Toys Inc.
- Cruz & Sons Co.
- Anna's Decorations, Ltd
- Vitachrome Inc.
- Gifts4AllAges.com

Loyal Customers (High Recency, High Frequency, High Monetary):

- Customers in "Recency_HML" bin labeled as "H" (High Recency).
- Customers in "Frequency_HML" bin labeled as "H" (High Frequency).
- Customers in "Monetary_HML" bin labeled as "H" (High Monetary).
- Customers falling into the intersection of "H" bins across Recency, Frequency, and Monetary categories.
- Customers with consistently high engagement across all three metrics, indicating loyalty and ongoing interest.

Below are few examples :

- Herkku Gifts
- Euro Shopping Channel
- Dragon Souvenirs, Ltd.
- Canadian Gift Exchange Network
- Auto Assoc. & Cie.
- Mini Gifts Distributors Ltd.
- Euro Shopping Channel

Revenue Growth through POS Data Analysis

Agenda

In this presentation, we will delve into the transformative potential of the grocery store's Point of Sale (POS) data to strategically enhance revenue. Commencing with an introduction to the project's objectives, we will navigate through the provided transactional data, emphasizing its significance in uncovering valuable insights. Following a meticulous exploration and cleaning of the dataset, our analysis will pivot towards identifying patterns in customer orders through item association and segmentation. The crux of our recommendations lies in the proposal of targeted combo offers and discount strategies, tailored to capitalize on prevalent customer behaviors. By the end of the presentation, we aim to provide a comprehensive implementation plan, empowering the grocery store to seamlessly integrate these insights and elevate its revenue through enticing promotions and strategic discounts.

Executive Summary Of Data :

The provided dataset represents a grocery store's transactional records on January 1, 2018. Each row corresponds to a specific order (Order_id), and the products purchased in that order are detailed. The dataset encompasses a diverse range of grocery items, including yogurt, pork, sandwich bags, lunch meat, all-purpose items, flour, soda, butter, beef, aluminum foil, dinner rolls, shampoo, and mixes. Notably, some products like "all-purpose" appear multiple times in a single order, suggesting potential variations or quantities. This dataset lays the foundation for exploring customer purchase patterns, identifying frequently co-occurring items, and formulating targeted recommendations for combo offers and discounts to optimize the grocery store's revenue strategy.

Problem Statement

A grocery store shared the transactional data with you. Your job is to conduct a thorough analysis of Point of Sale (POS) data, identify the most commonly occurring sets of items in the customer orders, and provide recommendations through which a grocery store can increase its revenue by popular combo offers & discounts for customers.

Know Your Data

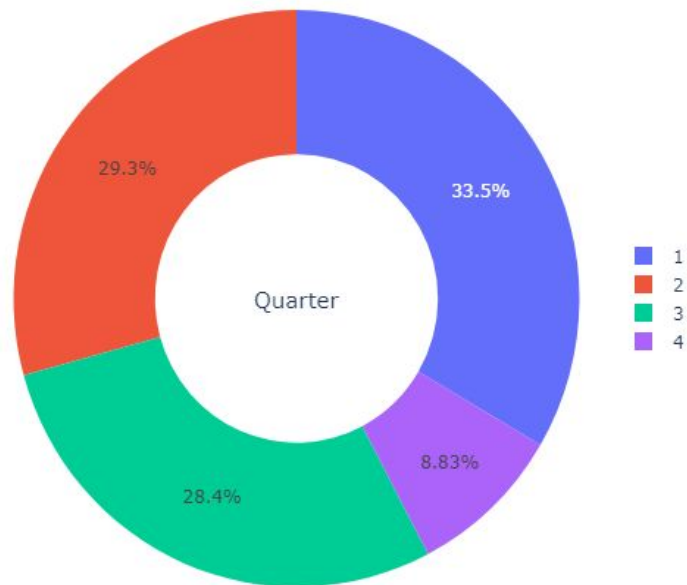
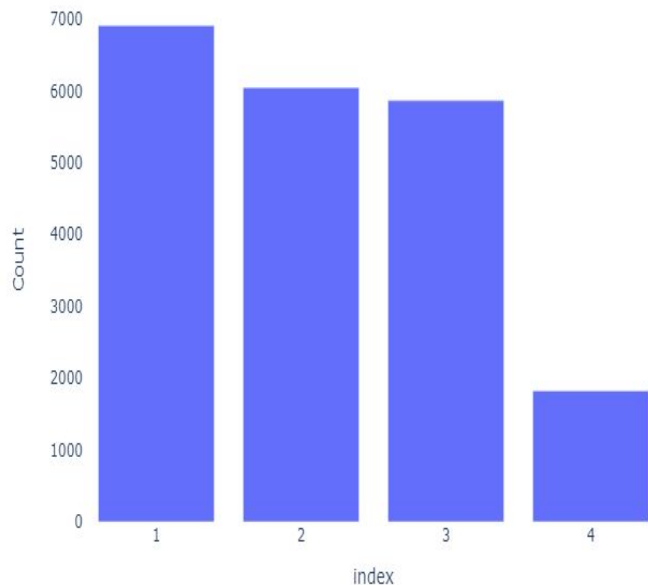
- Data has dimension of 20641 rows and 3 columns
- None of the column has a null value in it
- Below are the datatypes of corresponding columns in the dataset

```
Data columns (total 3 columns):  
#      Column      Non-Null Count  Dtype  
---  -  
0     Date          20641 non-null   object  
1     Order_id        20641 non-null   int64  
2     Product          20641 non-null   object  
dtypes: int64(1), object(2)
```

Exploratory Data Analysis

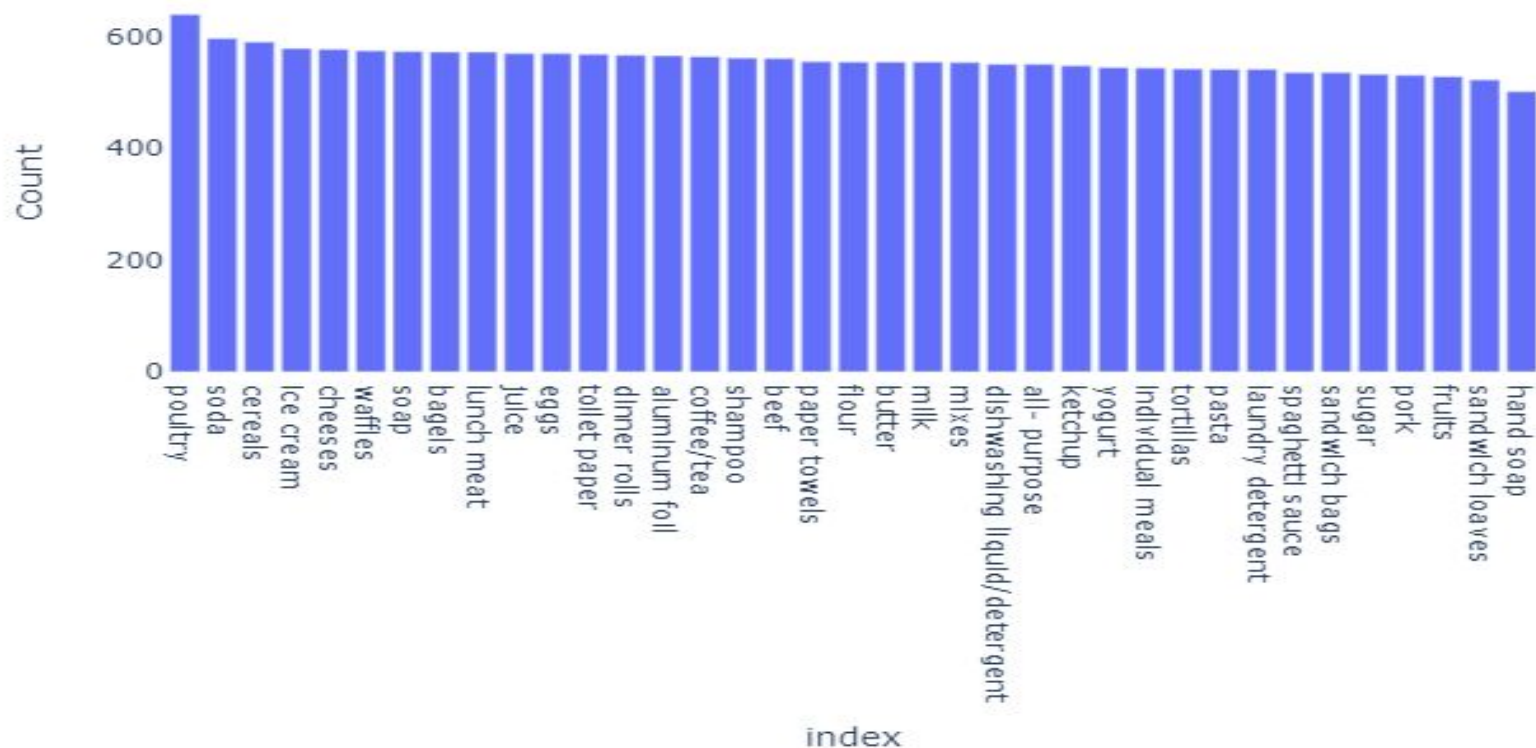
Univariate Analysis

Distribution of Quarter

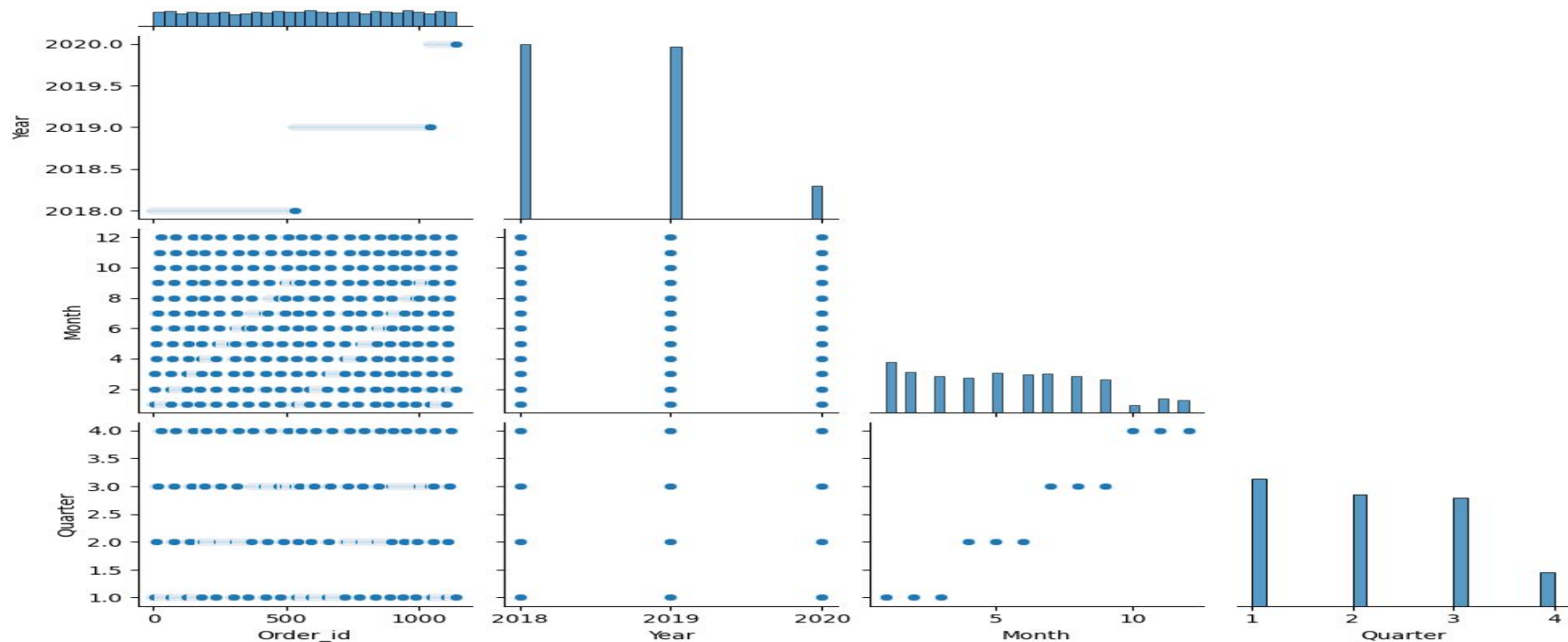


Univariate Analysis Ctnd..

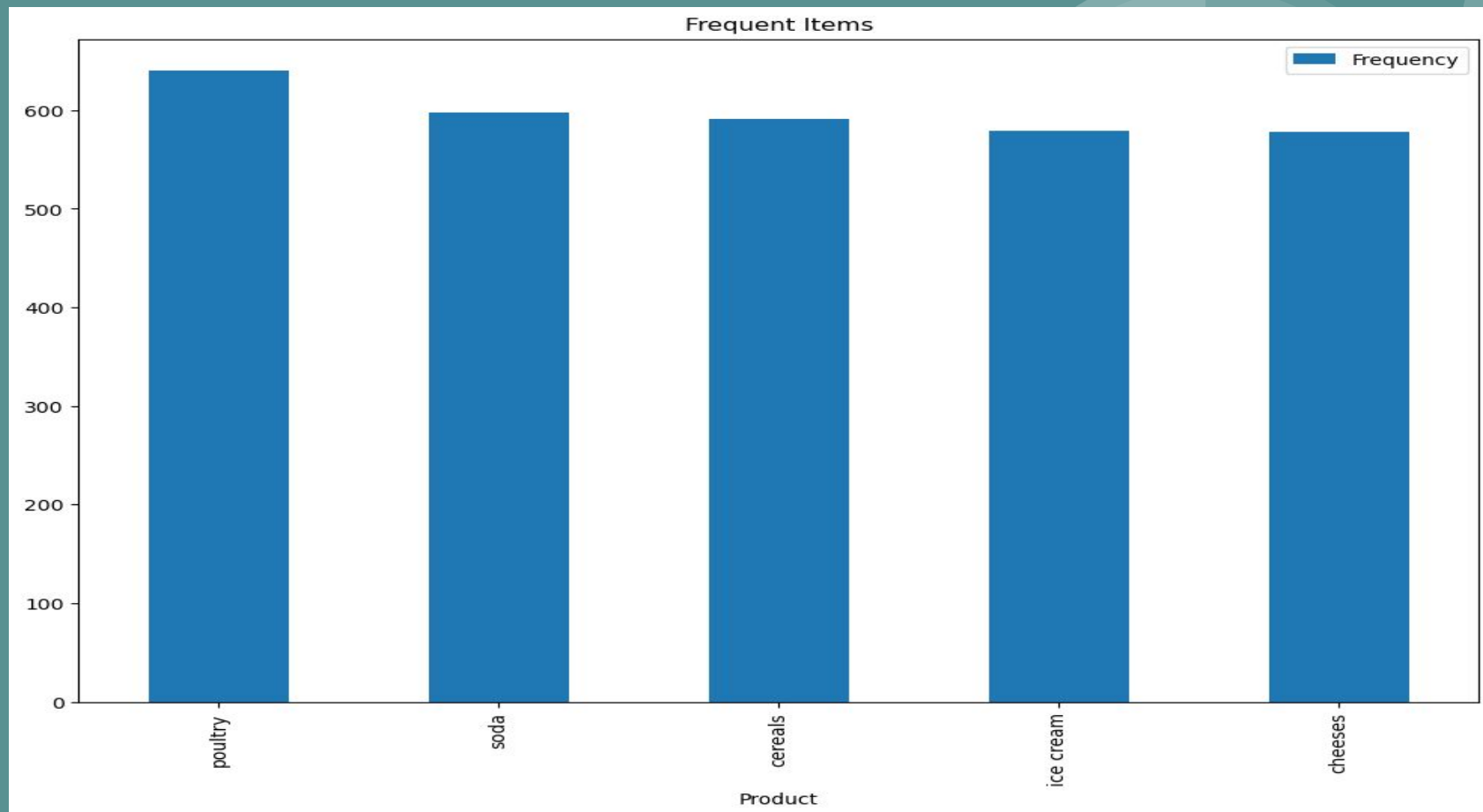
Distribution of Product



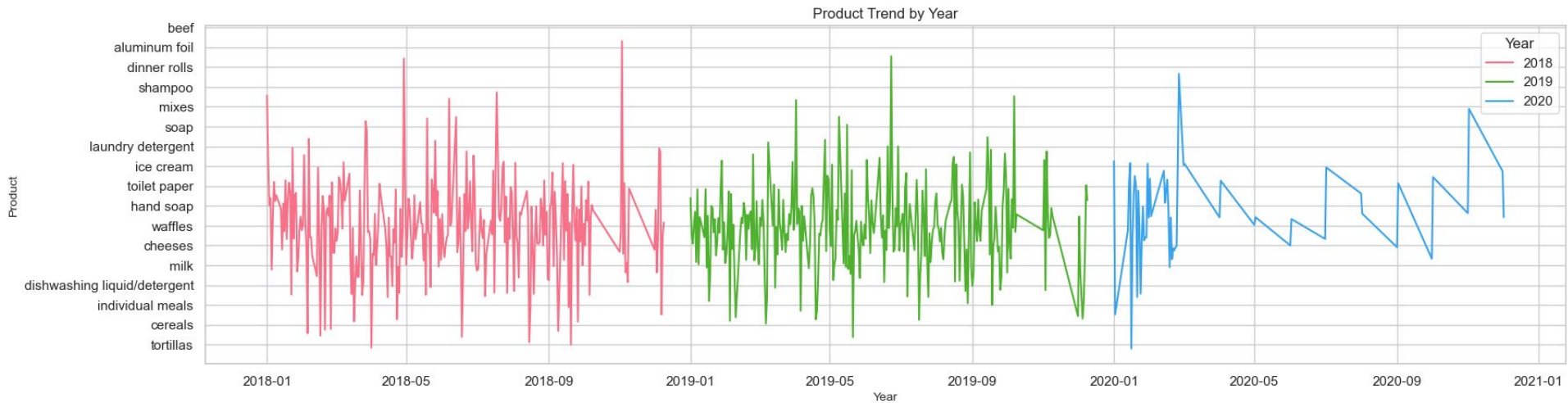
Bivariate Analysis Ctnd..



Frequency of Items Bought



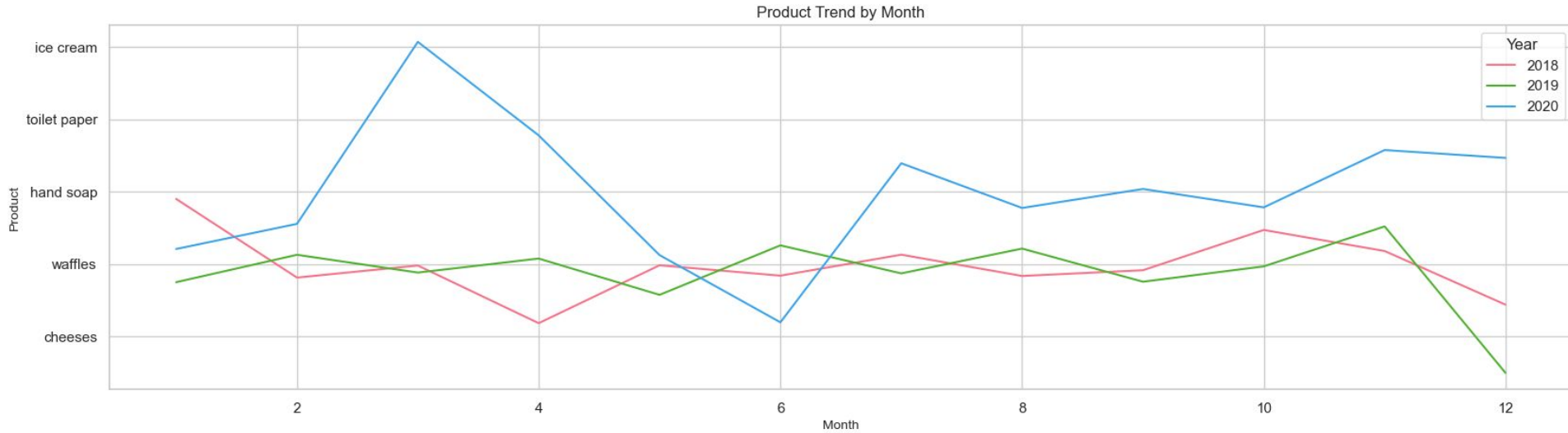
Weekly, Monthly, Quarterly, Yearly Trends in Sales and Sales Across different Categories of different features



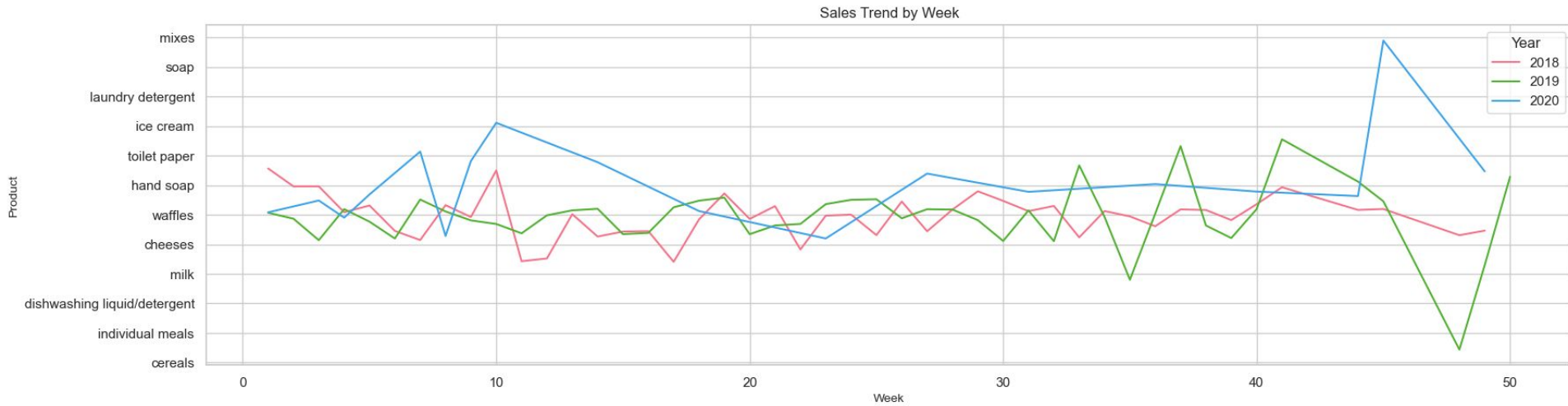
Sales Across different Categories of



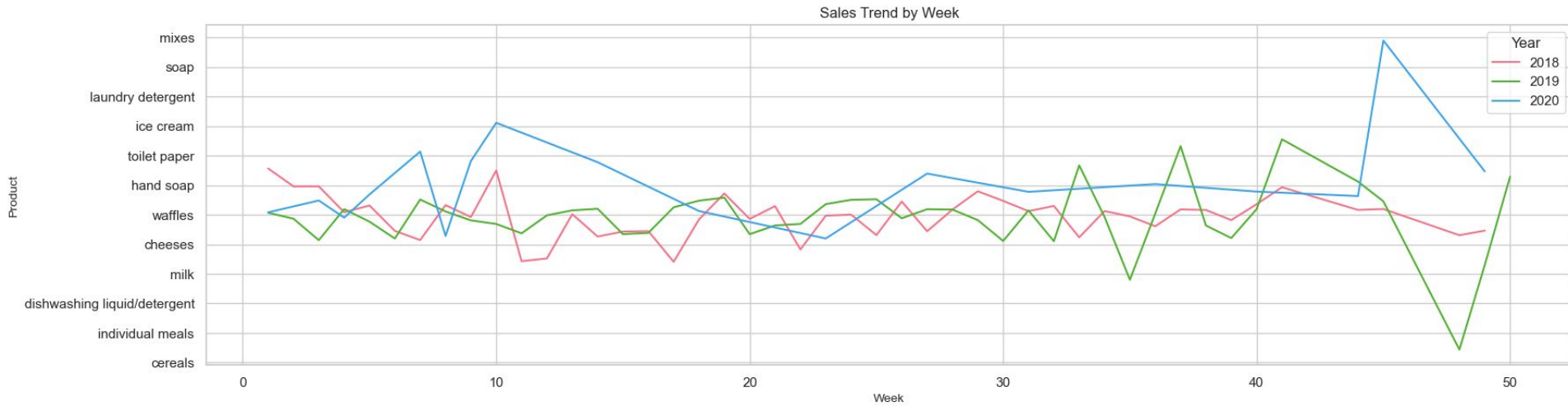
Weekly, Monthly, Quarterly, Yearly Trends in Sales and Sales Across different Categories of different features CtnD .



Weekly, Monthly, Quarterly, Yearly Trends in Sales and Sales Across different Categories of different features CtnD .



Weekly, Monthly, Quarterly, Yearly Trends in Sales and Sales Across different Categories of different features CtnD .



EDA Analysis/Inferences :

Inferences & Takeaway :

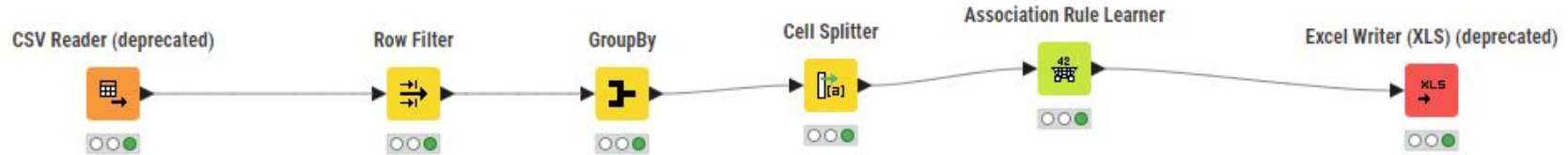
- Poultry is the most frequent bought item , followed by soda , cereals , icecream and cheeses .
- Association between "all-purpose" and "aluminum foil" is notable, with a confidence of 61.92% and a lift of 1.15.
- Conversely, "aluminum foil" also has a strong association with "all-purpose," reflecting a similar confidence and lift.
- "all-purpose" and "bagels" exhibit a significant association, with a confidence of 58.82% and a lift of 1.07.
- Reciprocally, "bagels" also show a notable connection with "all-purpose," emphasizing the bidirectional nature of the association.
- "all-purpose" and "beef" are associated, with a confidence of 60.06% and a lift of 1.09.
- The relationship between "beef" and "all-purpose" is mutual, with similar confidence and lift values.
- "all-purpose" appears to have versatile associations, including connections with "aluminum foil," "bagels," and "beef."
- The confidence levels indicate a relatively high likelihood of customers purchasing "aluminum foil" and "all-purpose" together.
- Similar patterns emerge for "all-purpose" and "bagels," suggesting potential cross-promotional opportunities.
- The association between "all-purpose" and "beef" indicates a potential for bundled promotions or suggestive selling strategies.

Market Basket Analysis (Association Rules)

Lets Know Market Basket Analysis first

- MBA (Market Basket Analysis) association rules play a pivotal role in extracting meaningful insights from transactional data, such as the provided grocery store dataset.
- In this case, the association rules reveal patterns in customer buying behavior, indicating which products tend to co-occur in shopping baskets.
- The relevance of MBA in this context lies in its ability to inform strategic decision-making for the grocery store.
- By understanding the associations between products, the store can design targeted combo offers and discounts to capitalize on these natural affinities.
- For instance, if "all-purpose" and "aluminum foil" are frequently purchased together, the store can create a bundled promotion or discount for these items, thereby incentivizing customers to buy both.
- MBA provides a data-driven foundation for the formulation of effective marketing and sales strategies, ultimately contributing to increased revenue and enhanced customer satisfaction in the competitive grocery retail landscape.

KNIME Workflow image for Market Basket Analysis



Threshold value of minimum support is taken at 3% and that of minimum confidence is at 0.4

Association Identified

Support	Confidence	Lift	Consequent	Implies
0.0307	0.7955	2.1938	poultry	<---
0.1949	0.4479	1.1814	cheeses	<---
0.0325	0.4762	1.2539	lunch meat	<---
0.0342	0.5072	1.3362	yogurt	<---
0.0386	0.4806	1.2641	soda	<---

Support, Confidence, & Lift values

Support:

- Definition: Support measures the frequency of a particular itemset in the dataset.
- Values based on output :
- Support values range from 0.0307 to 0.0398.
- For example, the support of 0.0307 for the association (all-purpose) → (aluminum foil) indicates that this combination appears in approximately 3.07% of the transactions.
- Interpretation: Higher support values signify that the corresponding itemsets are more frequently bought together.

Confidence:

- Definition: Confidence measures the likelihood that an item B is purchased when item A is purchased.
- Values based on output :
- Confidence values range from 0.4479 to 0.5072.
- For instance, the confidence of 0.5072 for the association (all-purpose) → (soda) suggests that when customers buy all-purpose, there is a 50.72% chance they will also buy soda.
- Interpretation: Higher confidence values indicate a stronger probability of items being purchased together.

Lift:

- Definition: Lift measures the strength of the association between two items, considering their individual occurrence probabilities
- Values based on output :
Lift values range from 1.1814 to 2.1938.
- For instance, a lift of 2.1938 for the association (all-purpose) → (poultry) indicates that customers are approximately 2.19 times more likely to buy poultry when they buy all-purpose than if these items were purchased independently.
- Interpretation: Lift values greater than 1 suggest a positive association, with higher values indicating stronger associations and potential opportunities for bundled promotions.

Suggestion of Possible Combos with Lucrative Offers

Poultry Combo:

- Offer a discount on poultry when purchased with any other item in the basket, as it demonstrates a high confidence of 79.55% and a significant lift of 2.1938.

Cheese Delight:

- Promote a cheese-themed combo or discount, leveraging the high support (19.49%) and a confidence of 44.79%, indicating a strong association with other items.

Suggestion of Possible Combos with Lucrative Offers Ctn'd.

Lunchtime Special:

- Introduce a lunchtime special combo with lunch meat, supported by a confidence of 47.62% and a lift of 1.1814.

Yogurt and More:

- Encourage customers to buy yogurt by offering a combo deal or discount, given its high confidence (50.72%) and lift (1.3362) when associated with other items.

Soda Mix and Match:

- Create promotions involving soda, capitalizing on its association with other products, supported by a confidence of 48.06% and a lift of 1.2641.

Thank You

Made By -
Soni Kumari /
PGPDSBA.O.FEB23.A