****

MARKETING & RETAIL ANALYTICS

PROJECT

****************** BUSINESS REPORT *******************

Prepared By : Sahid
COURSE : DSBA
DATE : 12-01-2025

Table of Contents

List of Tables

List of Figures

PART A: Exploratory Data Analysis & Customer Segmentation

1. Executive Summary & Data Overview:

The dataset consists of transactional data across various product lines, capturing details about orders, customer demographics, and sales trends. By performing Exploratory Data Analysis (EDA) and customer segmentation, the goal is to identify high-value customers and track product performance over time. The dataset can be used to derive insights on sales trends, customer behavior, and purchasing patterns.

Problem Statement:

The task at hand is to analyze sales data to identify key trends, understand customer segments, and provide insights for targeted marketing strategies. The dataset includes multiple attributes, such as product details, order values, and customer demographics, to help in customer segmentation and sales forecasting.

Data Information:

Shape of the Dataset:

The dataset contains the following number of rows and columns:

- **Number of rows:** 2747
- **Number of columns:** 20

Column Name	Description
ORDERNUMBER	Unique identifier for each order
QUANTITYORDERED	Number of items ordered in the transaction
PRICEEACH	Price per unit of the product
ORDERLINENUMBER	Line item number for the order
SALES	Total sales value for the order
ORDERDATE	Date the order was placed
DAYS_SINCE_LASTORDER	Number of days since the customer's last order
STATUS	Order status (e.g., Shipped, Pending)
PRODUCTLINE	Product category (e.g., Electronics, Apparel)
MSRP	Manufacturer's Suggested Retail Price (MSRP)
PRODUCTCODE	Product identifier
CUSTOMERNAME	Name of the customer
PHONE	Customer's phone number
ADDRESSLINE1	Customer's address (line 1)
CITY	City of the customer
POSTALCODE	Postal code of the customer
COUNTRY	Country of the customer
CONTACTLASTNAME	Last name of the customer's contact person
CONTACTFIRSTNAME	First name of the customer's contact person
DEALSIZE	Size of the deal (Small, Medium, Large)

(Table1 : Description of Auto Sales Data)

Statistical Summary:

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

(Table 2 : Statistical Summary of Auto Sales Data)

Order Quantity (QUANTITYORDERED): The average order consists of around 35 items, with a minimum of 6 and a maximum of 97. This indicates a generally moderate order size with some high-volume orders.

Price per Item (PRICEEACH): The average price per item is approximately 101.10, with values ranging from 26.88 to 252.87, suggesting a broad range of product prices being ordered.

Sales (SALES): Sales figures have a wide distribution, with an average of 3553.05 and a maximum of 14082.80. This shows that while most transactions are smaller, there are high-value sales as well.

Days Since Last Order (DAYS_SINCE_LASTORDER): The average time since the last order is around 1757 days, but it ranges from 42 to 3562 days, suggesting that some customers have been inactive for a while, while others have made recent purchases.

Manufacturer's Suggested Retail Price (MSRP): The MSRP varies between 33 and 214, with a mean of 100.69, showing that the dataset includes products of varying price points, which could be tied to customer segmentation strategies.

Overall, the data shows variability in customer orders, with both small and large purchases across different price points, and varying levels of engagement based on time since last order.

Assumptions and Limitations of the Data:

- **Time Period Limitation:** The data covers a period from 2018 to 2020, so it may not capture more recent trends or shifts in customer behavior.
- **External Factors:** The analysis assumes that external factors (e.g., market events, seasonality) influencing sales remain constant across the dataset, which may not be true.
- **Geographic Assumption:** The data focuses mainly on customers from the USA and Europe, with limited representation from other regions.

2. **Exploratory Data Analysis**

Key Trends and Insights (Weekly, Monthly, Quarterly Analysis)

1. **Product Line Analysis:**
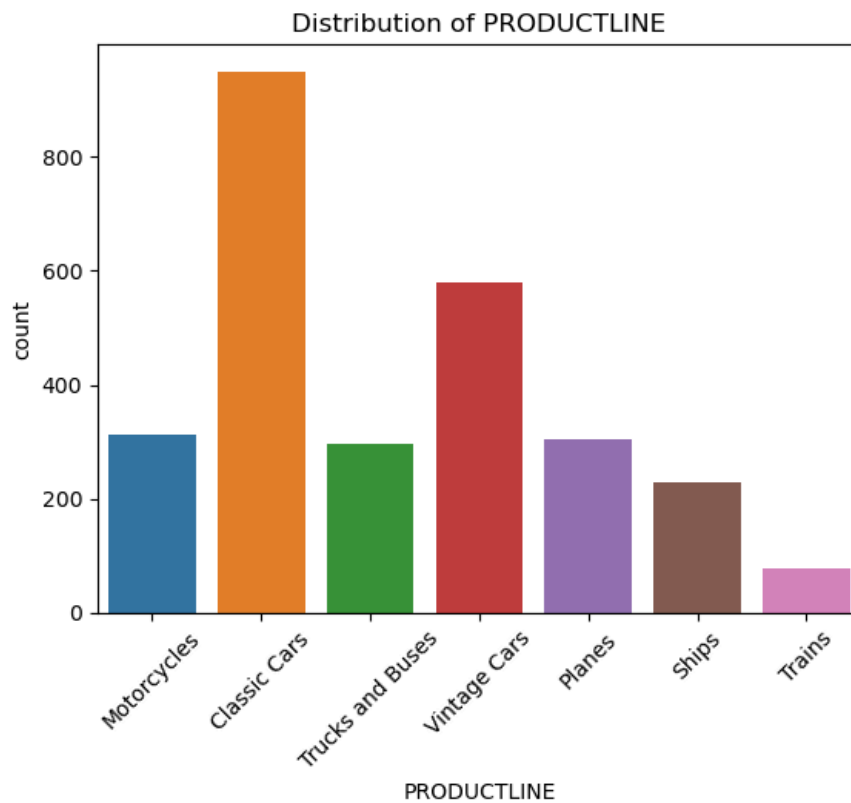
(Fig1 : showing Distribution of Product Line)
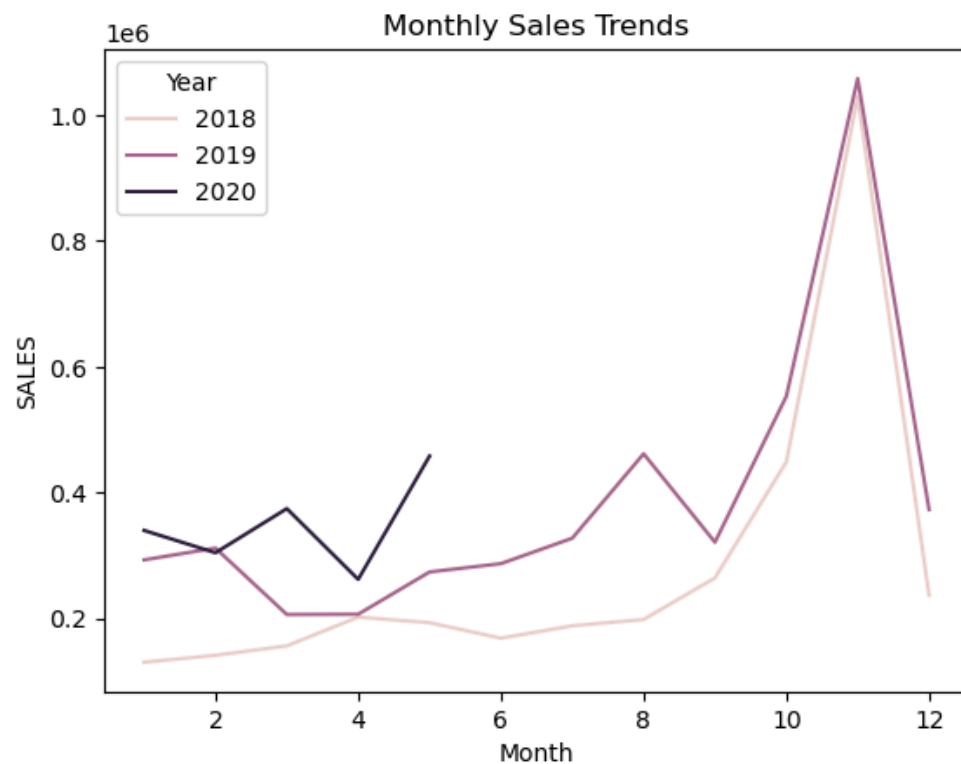
Top-Selling Product Lines:

Classic Cars: $3,842,868.54
Vintage Cars: $1,806,675.68
Trucks and Buses: $1,111,559.19
Motorcycles: $1,103,512.19
Planes: $969,323.42
Ships: $700,039.22
Trains: $226,243.47

Insight: "Classic Cars" is the top-selling product line, generating the highest sales, followed by "Vintage Cars" and "Trucks and Buses."

2. Sales Trend Over Time:

● **Monthly Analysis:**

(Fig 2 : showing monthly sales trends)

Monthly Sales Trend (Sample data for 2018, 2019, and 2020):

January 2018: $129,753.60
December 2018: $236,444.58
November 2019: $1,058,699.29
May 2020: $457,861.06

Insights:

● Sales fluctuated significantly, with peaks observed in November 2019 and December 2018, likely influenced by holiday promotions.
● Upward trend in 2019, but a dip in 2020, possibly due to external factors like the pandemic or market shifts.

- **Quarterly Analysis:**
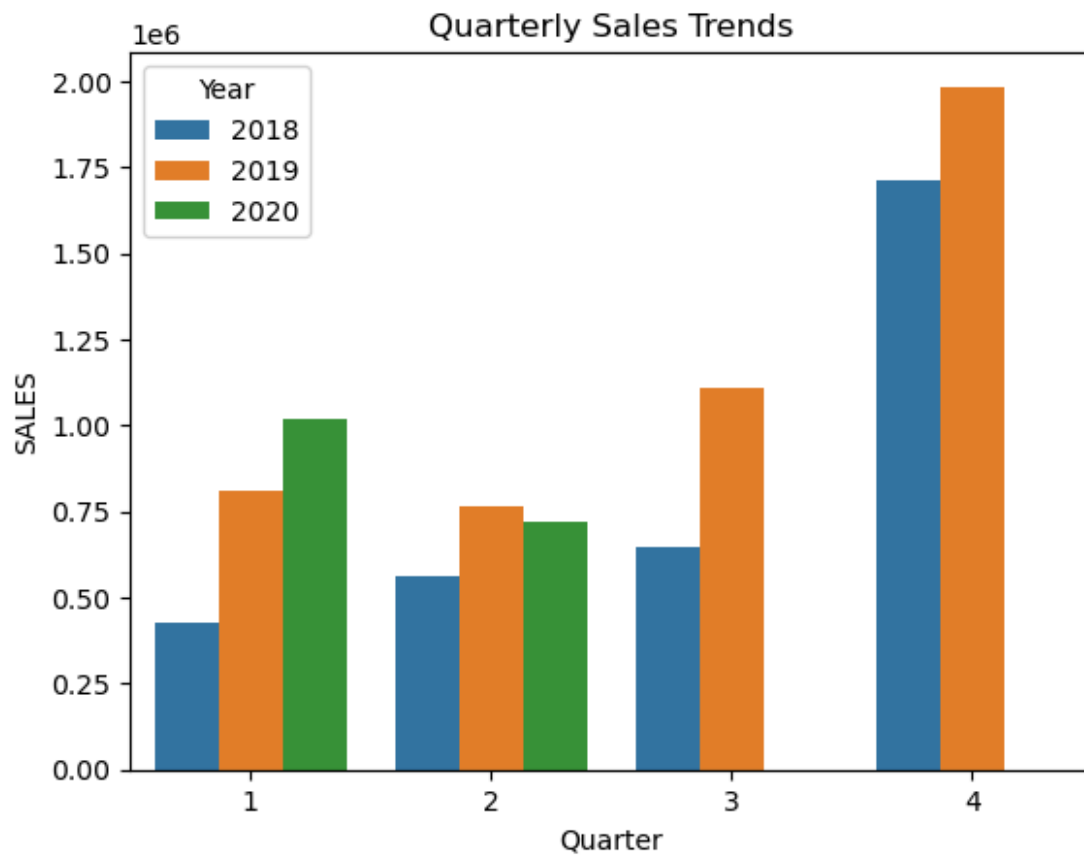
(Fig 3 : showing quarterly sales trends)

Quarterly Sales Trend:

Q1 2018: $426,399.11
Q2 2018: $562,365.22
Q4 2018: $1,714,735.19
Q1 2019: $809,841.36
Q4 2019: $1,984,426.20
Q1 2020: $1,017,788.74

Insight: Sales tend to peak in Q4, with Q1 showing consistent sales growth year-over-year. This suggests strong holiday-related sales and possibly promotions.

3. **Quantity Ordered vs. Sales:**
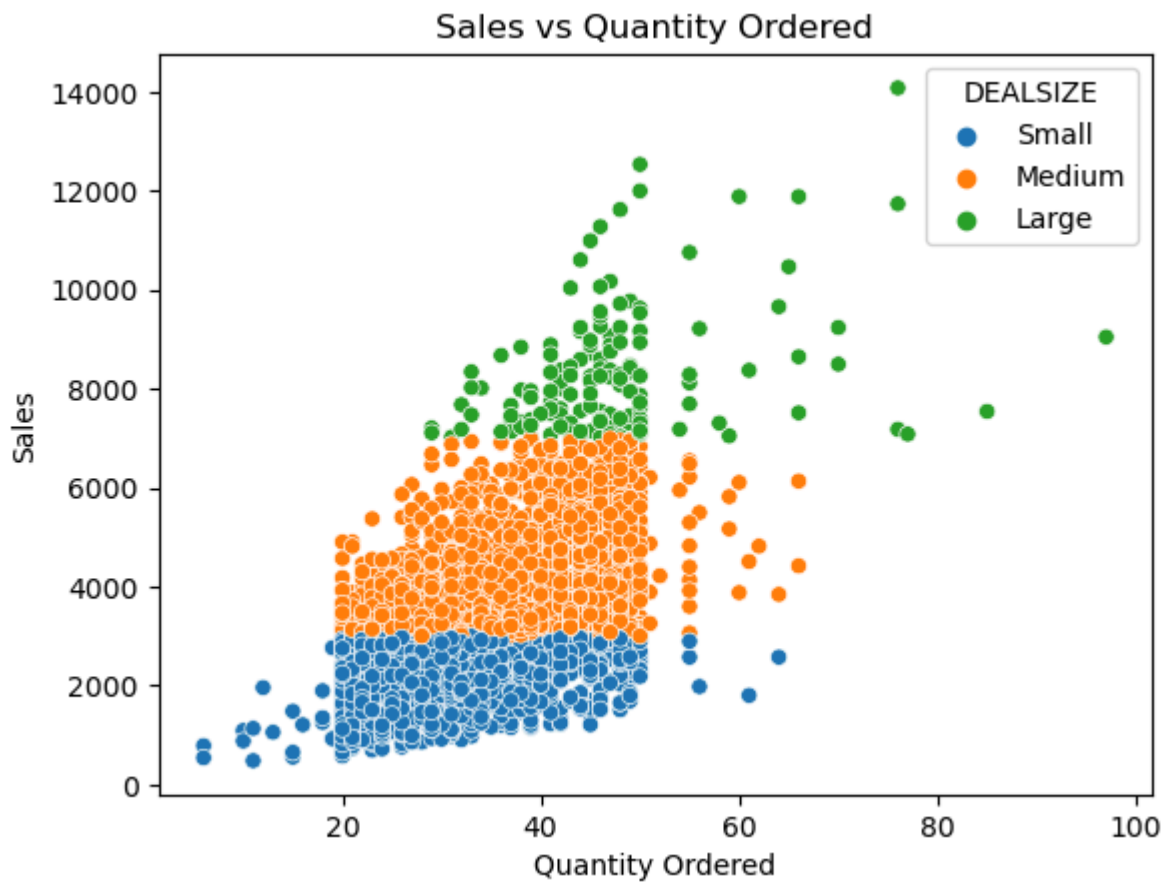
(Fig 4 showing correlation between quantity ordered & sales)

Insight: There is a moderate positive correlation between Quantity Ordered and Sales, meaning as the quantity ordered increases, sales tend to rise as well.

4. Discount Analysis:
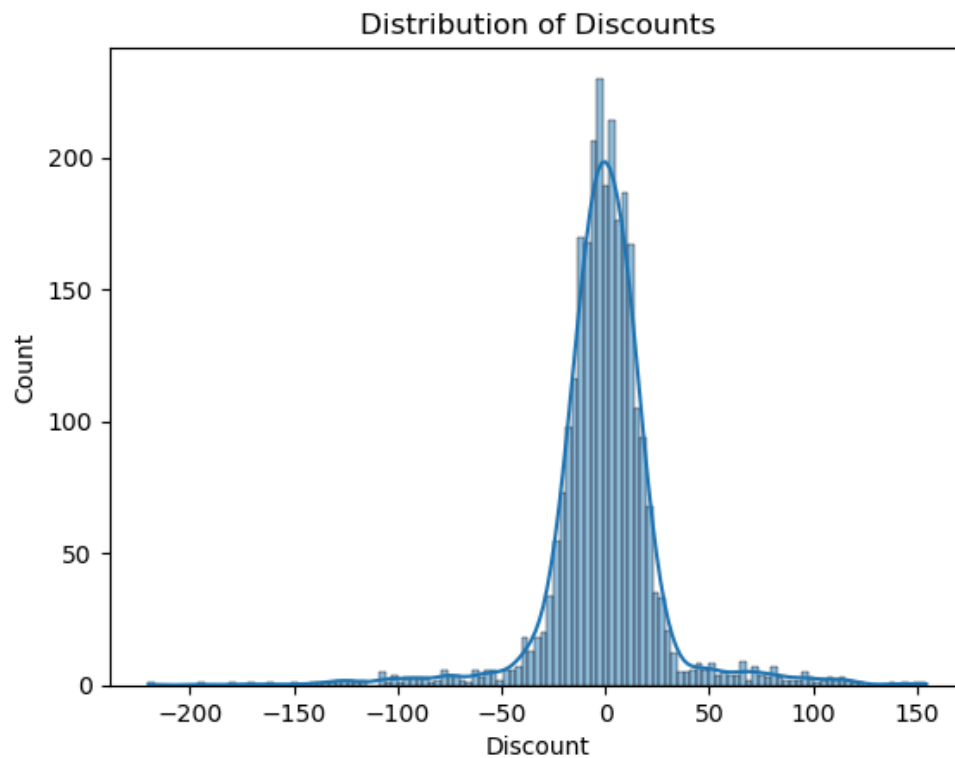
(Fig 5 showing distribution of discounts)

Sales by Discount Range:

(0, 10]   1860208.51

(10, 20]   1421611.26

(20, 30]   561845.65

(30, 40]   116450.26

(40, 50]   78797.50

Insight: The **0-10% discount range** generates the highest sales, with a clear decline in sales as the discount percentage increases.

5. **Customer Demographics:**

- ## **Regional Sales**
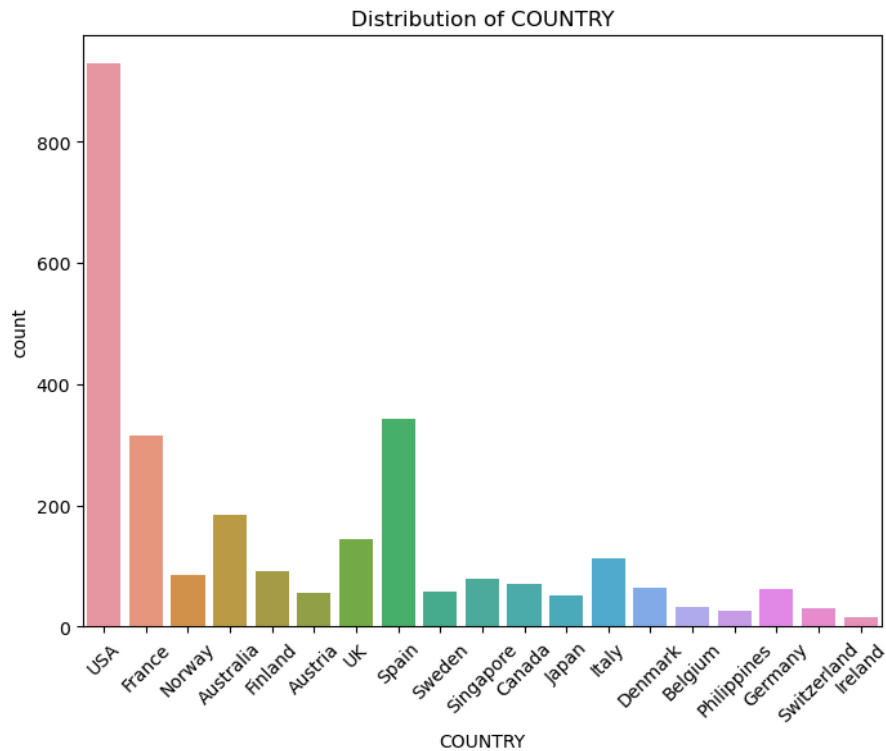
(Fig 6 showing distribution of sales by country)

Sales Distribution by countries

USA	3355575.69
Spain	1215686.92
France	1110916.52
Australia	630623.10
UK	478880.46
Italy	374674.31
Finland	329581.91
Norway	307463.70
Singapore	288488.41
Denmark	245637.15
Canada	224078.56
Germany	220472.09
Sweden	210014.21
Austria	202062.53
Japan	188167.81
Switzerland	117713.56
Belgium	108412.62

Philippines   94015.73
Ireland   57756.43

Insight from Sales Distribution by Countries:

1. **USA** leads with the highest sales, contributing **$3,355,575.69**, significantly outperforming other countries.
2. **Spain** and **France** follow, contributing around **$1,215,686.92** and **$1,110,916.52**, respectively.
3. Other countries such as **Australia**, **UK**, and **Italy** show moderate sales contributions.
4. Countries like **Philippines** and **Ireland** have relatively smaller sales contributions.

● **Deal Size Analysis:**
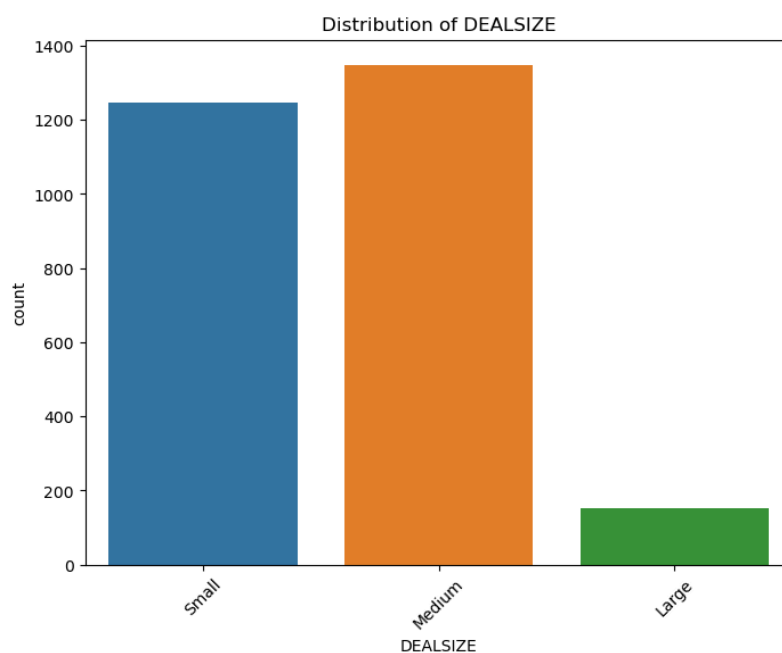
(Fig 7 showing the deal size by sales)

● **Medium-sized deals** contribute the most to overall sales, generating **$5,931,231.47**, which indicates this segment is a significant driver of revenue.
● **Small-sized deals** account for **$2,570,033.84**, showing a strong contribution despite their smaller individual transaction size.
● **Large-sized deals**, while having the highest value per transaction, generate only **$1,258,956.40**, indicating they are less frequent.

Insight:

Medium deals are the backbone of the business's revenue, suggesting efforts to maintain and grow this segment could be highly beneficial. Small deals also play a substantial role and could benefit from upselling strategies. Large deals, though fewer, could be targeted with premium services to increase their frequency and impact.

Summary of EDA Inferences Across Categories

1. Product Line Analysis

- **Top Performers:** Classic Cars and Vintage Cars are the highest contributors to sales.
- **Lower Performers:** Trains and Ships contribute the least to overall sales.

2. Sales Trends

- **Monthly:** Peak sales occur in November and December, likely driven by holiday promotions.
- **Quarterly:** Q4 consistently records the highest sales, underscoring holiday season impact.
- **Weekly:** End-of-quarter peaks indicate strong quarterly sales pushes.

3. Quantity Ordered vs. Sales

- **Correlation:** A moderate positive correlation (0.55) suggests higher quantities generally result in increased sales but not in a perfectly linear relationship.
- **Trends:** Larger quantities ordered (40–50) generate the highest sales.

4. Discount Impact

- **Sales by Discount Range:** Discounts between 0%-10% yield the highest sales. Sales drop significantly for discounts exceeding 20%.
- **Insights:** Small discounts are most effective in driving revenue, while deeper discounts are less impactful.

5. Customer Demographics

- **Regional Sales:**
  - **Dominant Markets:** USA leads with $3.35M in sales, followed by Spain ($1.21M) and France ($1.11M).
  - **Lower Contributors:** Countries like Ireland and the Philippines contribute less than $100K each.
- **Deal Size:**
  - **Medium-sized Deals:** The largest contributor, generating $5.93M in sales.
  - **Small-sized Deals:** Account for $2.57M, significant for smaller transactions.
  - **Large-sized Deals:** Generate $1.25M, representing high-value but infrequent transactions.

6. Insights on Average Sales by Quantity Ordered

- Sales and average transaction values increase as order quantities grow, peaking for orders in the range of 40–50 units.

3. <u>Customer Segmentation using RFM Analysis</u>

What is RFM Analysis?

RFM (Recency, Frequency, Monetary) analysis is a marketing technique used to segment customers based on their transactional behavior. It helps identify customer segments such as loyal customers, potential churners, and lost customers by analyzing three key metrics:

1. **Recency (R):** How recently a customer made a purchase.
2. **Frequency (F):** How often a customer makes purchases.
3. **Monetary (M):** How much revenue a customer contributes.

## Parameters Used and Assumptions Made

- **Recency:** Calculated as the number of days since the last transaction (date difference from the analysis date).
- **Frequency:** Number of purchases made by a customer within the observation period.
- **Monetary Value:** Total revenue generated by the customer during the observation period.
- **Assumptions:**
  - Transactions within the dataset accurately represent customer behavior.
  - Segmentation scores are divided into 4 groups (Best Customer , Churning Customer , Lost Customer , Loyal Customer for each R, F, and M).

<u>Workflow Implementation</u>

The **KNIME workflow** for RFM segmentation involves the following key steps:

1. **Excel Reader:** Load the transactional dataset.
2. **Date & Time Difference:** Calculate the **Recency (R)** for each customer by computing the difference between the last transaction date and the reference date.
3. **GroupBy:** Aggregate the data by customer ID to calculate:
   - **Recency (R):** Days since the last transaction.
   - **Frequency (F):** Total number of transactions for each customer.
   - **Monetary (M):** Sum of all transaction amounts for each customer.
4. **Column Reader:** Prepare the dataset by reading and validating R, F, and M columns for further processing.
5. **Math Formula:**
   - Assign **R, F, and M scores** based on quartiles or defined thresholds.
   - Combine R, F, and M scores into an **RFM score** (e.g., weighted or additive scoring).
6. **Normalizer:** Standardize the RFM scores to ensure consistency for clustering.
7. **KMeans Clustering:** Segment customers into 4 clusters based on their normalized RFM scores.
8. **Rule Engine:** Assign customer segments based on cluster labels, such as:
   - Best Customers
   - Loyal Customers
   - On the Verge of Churning
   - Lost Customers

KNIME Workflow Image
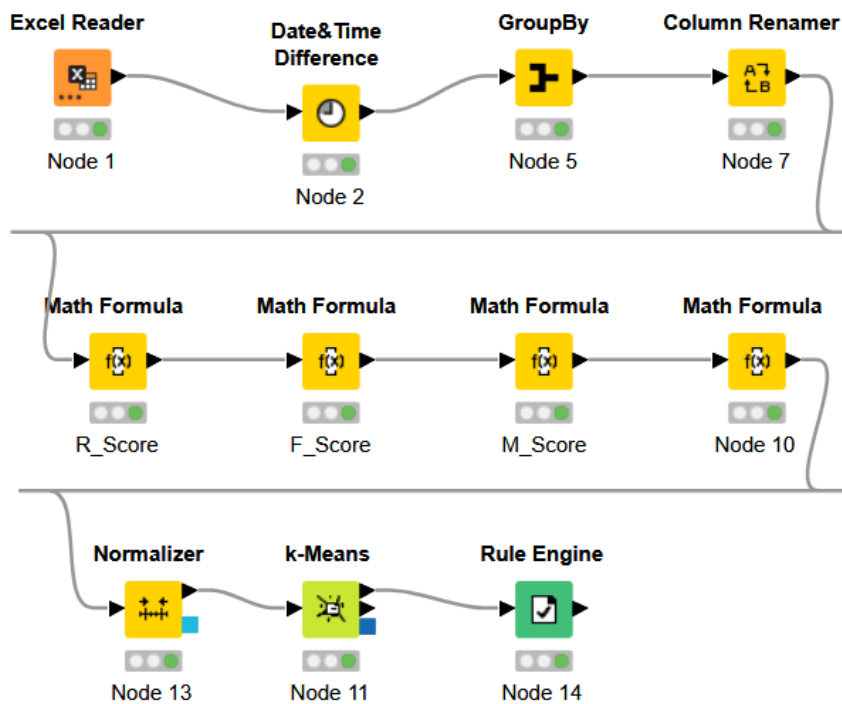
(Fig 8 showing the KNIME workflow used for RFM)

Results and Output

- Below is a snapshot of the output table showing the calculated RFM metrics and customer segments:

CUSTOMERNAME	Frequency	Monetary	Recency	R_Score	F_Score	M_Score	RFM_Score	Cluster	Customer_Segment
AV Stores, Co.	51	157807.81	1883	0.75	1	1	14	cluster_1	Loyal Customer
Alpha Cognac	20	70488.44	1751	1	0.25	0.5	10	cluster_0	Best Customer
Amica Models & Co.	26	94117.26	1952	0.5	0.5	0.5	9	cluster_2	Churning Customer
Anna's Decorations, Ltd	46	153996.13	1770	1	1	1	15	cluster_1	Loyal Customer
Atelier graphique	7	24179.96	1875	0.75	0	0	6	cluster_3	Lost Customer
Australian Collectables, Ltd	23	64591.46	1709	1	0.5	0.25	10	cluster_0	Best Customer
Australian Collectors, Co.	55	200995.41	1871	0.75	1	1	14	cluster_1	Loyal Customer
Australian Gift Network, C	15	59469.12	1806	0.75	0.25	0.25	8	cluster_3	Lost Customer
Auto Assoc. & Cie.	18	64834.32	1920	0.5	0.25	0.25	7	cluster_3	Lost Customer

(Table 3 showing output of the calculated RFM metrics and customer segments)

4. Inferences from RFM Analysis and Identified Segments

Segment Insights

1. **Best Customers:** These customers score high across R, F, and M, indicating recent and frequent purchases with significant revenue contributions.

CUSTOMERNAME	Customer_Segment
Alpha Cognac	Best Customer
Australian Collectables, Ltd	Best Customer
Auto Canal Petit	Best Customer
Baane Mini Imports	Best Customer
Blauer See Auto, Co.	Best Customer

(Table 4 showing 5 Best Customers)

2. **Customers on the Verge of Churning:** These customers have low recency scores but moderate frequency and monetary values, indicating they haven't engaged recently despite prior activity.

CUSTOMERNAME	Customer_Segment
Amica Models & Co.	Churning Customer
Canadian Gift Exchange Netw	Churning Customer
Collectable Mini Designs Co.	Churning Customer
Diecast Collectables	Churning Customer
Heintze Collectables	Churning Customer

(Table 5 showing customer on the verge of churning)

3. **Lost Customers:** These customers score low across all three metrics, indicating inactivity for a significant time and minimal engagement.

CUSTOMERNAME	Customer_Segment
Atelier graphique	Lost Customer
Australian Gift Network, Co	Lost Customer
Auto Assoc. & Cie.	Lost Customer
Auto-Moto Classics Inc.	Lost Customer
Bavarian Collectables Imports	Lost Customer

(Table 6 showing lost customers)

4. **Loyal Customers:** These customers have high frequency and monetary scores, indicating consistent engagement and significant contributions to revenue.

CUSTOMERNAME	Customer_Segment
AV Stores, Co.	Loyal Customer
Anna's Decorations, Ltd	Loyal Customer
Australian Collectors, Co.	Loyal Customer
Dragon Souveniers, Ltd.	Loyal Customer
Euro Shopping Channel	Loyal Customer

(Table 7 showing loyal customers)

PART B: Market Basket Analysis

1. Executive Summary

This report presents the results of the analysis of Point of Sale (POS) data for a grocery store. The primary objective was to identify frequently occurring sets of items in customer orders using Market Basket Analysis. Based on the insights derived from association rules, the report recommends targeted combo offers and discounts to increase the store's revenue by promoting popular item combinations.

2. Data Overview

The dataset, **"dataset_group.csv"**, includes transaction data for a grocery store. It contains records of individual customer purchases, which include product IDs, quantities, and transaction details. Key variables include:

- **Order ID**: Unique identifier for each transaction.
- **Products Purchased**: List of products bought in each transaction.

3. Exploratory Data Analysis (EDA)

3.1 Sales Trends and Key Findings

- **Total Orders**: 20,641 transactions
- **Unique Orders**: 1,139 unique products sold
- **Top 10 Most Sold Products**:
  - **Poultry**: 640
  - **Soda**: 597
  - **Cereals**: 591
  - **Ice Cream**: 579
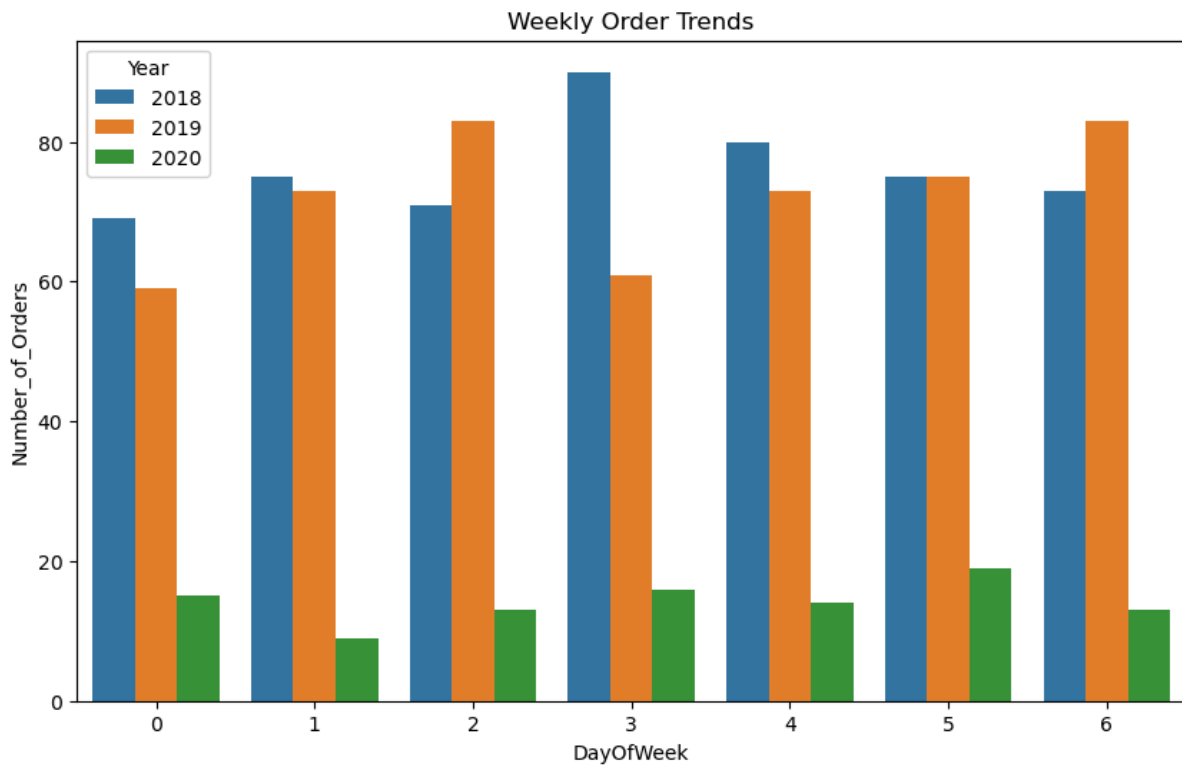 - **Cheeses**: 578

Monthly Trends:

(Fig 9 Showing monthly order trends by years)

- The monthly order distribution shows that January and May experienced higher-than-average sales, with peaks of 62 and 67 orders respectively.

Weekly Trends:

(Fig 10 showing weekly trends)

- **Thursday** recorded the highest sales volume, with 90 orders, indicating that it is a peak shopping day.

2. Association Rules

Relevance of Association Rules, KNIME Workflow, and Thresholds for Support and Confidence

Association rules reveal product pairs that customers frequently buy together, which can help in creating promotions, combo offers, and discounts. The following thresholds were used for the analysis:

- **Minimum Support**: 0.05 (indicating that the product pair appears in at least 5% of transactions)
- **Minimum Confidence**: 0.5 (indicating at least 50% likelihood that one item will be bought if the other is bought)
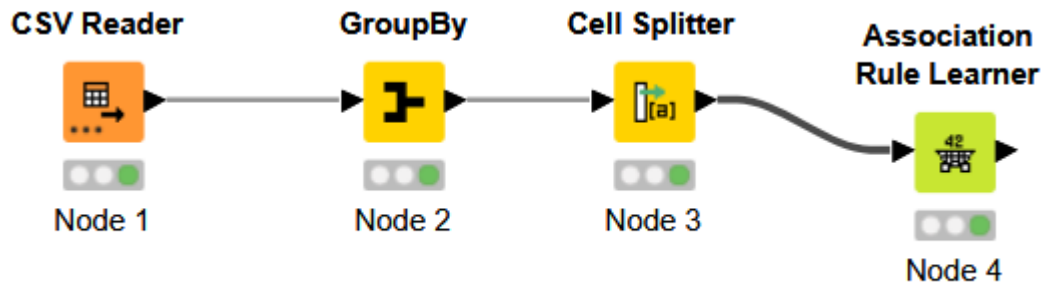
(Figure 11 showing KNIME Workflow Diagram used for MBA)

3. Identified Associations

The following table shows the association rules with Support, Confidence, and Lift values for various product pairs.

Support	Confidence	Lift	Consequent	implies	Items
0.05179982441	0.6860465116	1.627931202	poultry	<---	[dinner rolls, spaghetti sauce, ice cream]
0.05092186128	0.6744186047	1.726208518	cheeses	<---	[bagels, cereals, sandwich bags]
0.05355575066	0.6559139785	1.556429211	poultry	<---	[dinner rolls, spaghetti sauce, laundry detergent]
0.05531167691	0.6494845361	1.79119343	paper towels	<---	[eggs, ice cream, pasta]
0.05531167691	0.6428571429	1.73100304	pasta	<---	[paper towels, eggs, ice cream]
0.05355575066	0.6421052632	1.650920756	dinner rolls	<---	[spaghetti sauce, poultry, laundry detergent]
0.05179982441	0.6413043478	1.648861517	dinner rolls	<---	[spaghetti sauce, poultry, ice cream]
0.05004389816	0.6404494382	1.700400723	juice	<---	[yogurt, toilet paper, aluminum foil]
0.05092186128	0.6373626374	1.512408425	poultry	<---	[dinner rolls, spaghetti sauce, cereals]
0.05179982441	0.6344086022	1.627458103	eggs	<---	[paper towels, dinner rolls, pasta]

(Table 8 showing association rules for various product pairs)

Explanation of Key Metrics

1. Support: Represents the percentage of transactions in which the product pair appears. For example, the pair [Dinner Rolls, Spaghetti Sauce, Ice Cream] appears in 5.18% of all transactions.
2. Confidence: The probability that the consequent item is purchased given that the other items in the rule are purchased. For instance, the confidence of 0.6860 for "Poultry" suggests that if customers purchase [Dinner Rolls, Spaghetti Sauce, Ice Cream], there is a 68.6% chance they will also purchase Poultry.
3. Lift: Measures how much more likely the consequent item is to be bought when the antecedent items are bought together, compared to when they are bought independently. A lift value greater than 1 suggests a positive relationship. For example, the lift value of 1.628 for "Poultry" and [Dinner Rolls, Spaghetti Sauce, Ice Cream] indicates a strong association between these items.

4. Recommendations

Based on these associations, we suggest the following promotional strategies to increase sales:

1. **Combo Offer 1:** Poultry & Dinner Rolls & Spaghetti Sauce & Ice Cream
   - Offer: 10% Discount on Poultry, Dinner Rolls, Spaghetti Sauce, and Ice Cream combo.
   - Reason: Strong support and confidence indicate that these products are often purchased together, making it an ideal combo offer.
2. **Combo Offer 2:** Cheeses & Bagels & Cereals
   - Offer: Buy 1 Cheese, Get 1 Bagel Free
   - Reason: The high lift and confidence suggest that customers who buy cheese are likely to buy bagels and cereals as well.
3. **Combo Offer 3:** Paper Towels & Eggs & Ice Cream & Pasta
   - Offer: Buy Paper Towels and Eggs, Get 1 Free Pasta or Ice Cream.
   - Reason: Strong associations and positive lift values suggest that these items are frequently purchased together.
4. **Combo Offer 4:** Juice & Yogurt & Toilet Paper
   - Offer: 15% off on Juice, Yogurt, and Toilet Paper bundle.
   - Reason: Given the high support and confidence, these items show a strong association and can be sold together with a discount.
5. **Targeting High Sales Products:**
   - Focus promotions around Poultry, Ice Cream, and Paper Towels, as they have strong associations with other products and frequent purchases.

Conclusion

PART A:

Conclusion

The RFM (Recency, Frequency, Monetary) analysis has provided valuable insights into the customer base, allowing us to segment customers into four distinct groups: Best Customers, Loyal Customers, Customers on the Verge of Churning, and Lost Customers. This segmentation enables a more targeted approach to customer engagement, retention, and acquisition strategies. By focusing on the specific needs and behaviors of each segment, the business can optimize marketing efforts and enhance customer lifetime value.

Summary of Key Insights and Business Recommendations

1. **Best Customers**

- Insight: These customers have the highest RFM scores, indicating they are recent, frequent, and high-spending. They are the backbone of the business.
- Recommendation:
  - Retention Strategy: Offer personalized loyalty programs, exclusive offers, and early access to new products.
  - Referral Program: Encourage them to refer friends or colleagues by offering referral bonuses or discounts.

2. **Loyal Customers**
- Insight: Customers who are frequent buyers but with lower recency scores. They are consistently engaged but may be slipping into lower engagement.
- Recommendation:
  - Upsell & Cross-sell: Target them with relevant product recommendations based on past purchases.
  - Reward Engagement: Offer rewards for consistent purchases and engagement to maintain loyalty.

3. **Customers on the Verge of Churning**
- Insight: These customers exhibit reduced frequency and recency, yet their monetary value was once significant. They are at risk of disengagement.
- Recommendation:
  - Win-back Campaigns: Launch personalized re-engagement offers like discounts or product recommendations to bring them back.
  - Customer Feedback: Send surveys or direct communications to understand reasons for their decline in activity and address them accordingly.

4. **Lost Customers**
- Insight: Customers who have not interacted with the brand in a significant way for a while and have low scores in all RFM metrics.
- Recommendation:
  - Reactivation Campaigns: Use targeted email marketing with personalized discounts to re-engage them.
  - Surveys and Feedback: Gather feedback to understand why these customers left and develop strategies to address those concerns for potential future re-engagement.

PART B :

Conclusion

The analysis of the grocery store's transactional data has provided valuable insights into both customer purchasing behaviors and potential opportunities for increasing revenue. Below is a structured summary of the key findings and business recommendations based on the analysis:

Key Insights:

1. **Top-Selling Products:**
- The most frequently sold items include Poultry, Soda, and Cereals, indicating their strong demand in customer orders.
- Other notable top-sellers include Ice Cream, Cheeses, Waffles, and Soap.

2. **Sales Trends:**

- ○ Monthly Trends: Sales activity peaked in January (62 orders) and May (67 orders), suggesting these months may be ideal for targeted promotions.
- ○ Weekly Trends: Orders were highest on Thursday (90 orders), indicating a significant opportunity to align promotions with this peak shopping day.
3. **Common Product Associations**:
   - ○ Strong associations were found between commonly purchased items, such as:
     - ■ Poultry with Dinner Rolls and Spaghetti Sauce
     - ■ Eggs with Soda
   - ○ These associations present opportunities for bundling related products and cross-selling.

Business Recommendations:

1. **Promotions and Offers:**
   - ○ Leverage common product pairings, such as Poultry and Dinner Rolls, by offering bundle discounts like "Buy Poultry and Dinner Rolls Together for 10% Off".
   - ○ Target high-demand items like Poultry and Soda for combo offers, such as "Buy Poultry and Get Soda at 20% Off".
2. **Optimized Inventory and Sales Planning:**
   - ○ Focus inventory and promotions around January and May, based on observed high sales activity in these months.
   - ○ Since Thursday is the peak shopping day, plan special offers or discounts to drive higher sales during this time.
3. **Cross-Selling Opportunities:**
   - ○ Identify complementary products, such as Eggs and Ice Cream, and create promotional offers encouraging customers to purchase both together, driving higher average transaction value.

References :

- ● Previous Knime files
- ● Course Resource
- ● Google