iJRASET
International Journal For Research in
Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH
IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 14 Issue: I Month of publication: January 2026

DOI: https://doi.org/10.22214/ijraset.2026.76800

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 14 Issue I Jan 2026- Available at www.ijraset.com

# Framework for Data Analysis of Retail Market using Machine Learning Algorithms

Anshul Tipnis[1], Prof. Tejal Rachh[2]
[1]Student, [2]Asst. Professor, Atharva College of Engineering, Malad (W), Mumbai-400095, India

***Abstract: Presenting a comprehensive analytical framework for analyzing and forecasting supermarket sales data with the goal of understanding the seasonal sales trends and customer approach towards the product to improve overall business performance. The methodology begins with data preprocessing and cleaning to ensure quality, integrity and reliability towards the data which is followed by Exploratory Data Analysis (EDA) to uncover sales trends, seasonal patterns, and product and customer anomalies. Product price analysis with respect to different countries was conducted to detect customer approach. Customers are sectioned together using customer repetition analysis, K-means clustering with Elbow method, and customer lifetime value models to categorize customers into low, medium, and high-value customers. Market Basket Analysis with association mining rule to identify products which are collectively co-purchased. Ultimately in the end, sales were forecasted using Seasonal ARIMA (SARIMA) to predict future revenue to inform critical business decisions. The proposed system provides actionable insights analyzing the potential growth of a business on the surface of the global market.***
***Keywords: EDA, Market Basket Analysis, Associative rule mining, Apriori Algorithm, SARIMA, clustering, K-means, product analytics, customer lifetime value (CLV).***

## I. INTRODUCTION

Data-driven analytics framework endeavours to improve retail business outcomes through data science based and machine learning algorithms . In this research, the global sales data is examined, with a comparative focus on France and the Netherlands. It was observed that while France has a larger customer number ground, the Netherlands records a higher overall sales volume, telling differences in customer purchasing behavior and spending patterns for different products over the two nations.

Data is preprocessed and cleansed to ensure correctness of content that is to be used for analysis. Exploratory data analysis is performed for pattern and anomaly recognition, for finding seasonal trends and identifying correlations between data variables. Product price analysis is then performed to understand the sales drivers. Customer behavior towards a product or market is understood using RFM analysis, K-means clustering using the Elbow method to distribute customers based on their spending, and customer lifetime value (CLV) to partition customers into low, medium, and high-value customers. Market Basket Analysis is used in conjunction with the Apriori algorithm to record product linkages, producing rules based on lift, confidence, and support. Lastly, the SARIMA model is implemented for sales forecasting, allowing for accurate future sales estimates over a seasonal recorded data in chronological sequence. By integrating these algorithms, astute information about customer approach, performance statistics at the global level, and practical plans to improve marketing, inventory, and gross revenue planning is offered.

## II. PROBLEM STATEMENT

Retail businesses encounter growing setbacks in elucidating large volumes of sales data to understand customer behavior, contrast regional sales, and plan for upcoming ultimatums. In the analyzed dataset, a striking variance of the Netherlands having higher total revenue while having a smaller customer base than France was noted. The imbalance spotlights the importance of identifying high-value customers, evaluating product ultimatums, and examining regional dependencies in sales patterns. Unless data is analyzed properly, such intuitions remain discrete, leading to a restraint in a company's ability to design effective strategies.

Conventional reporting and elementary analytics have been often seen to fail to meet the requirement to capture these patterns, simultaneously leading to inefficient business strategies, poor inventory management, thus leading to loss of growth potential. To comprehend this, an eclectic framework is required that combines data preprocessing and cleaning, exploratory data analysis to identify trends, product and price analysis on global scale, customer sectionalization using RFM, K-Means with Elbow method, and customer lifetime value, market basket analysis using association rule mining with apriori algorithm, and sales prediction using SARIMA models. This algorithm enables businesses entities to discover hidden relationships and patterns in data, forecast future ultimatums precisely, and develop working strategies to improve business performance and overall sales architecture.

## III. PROPOSED METHODOLOGY

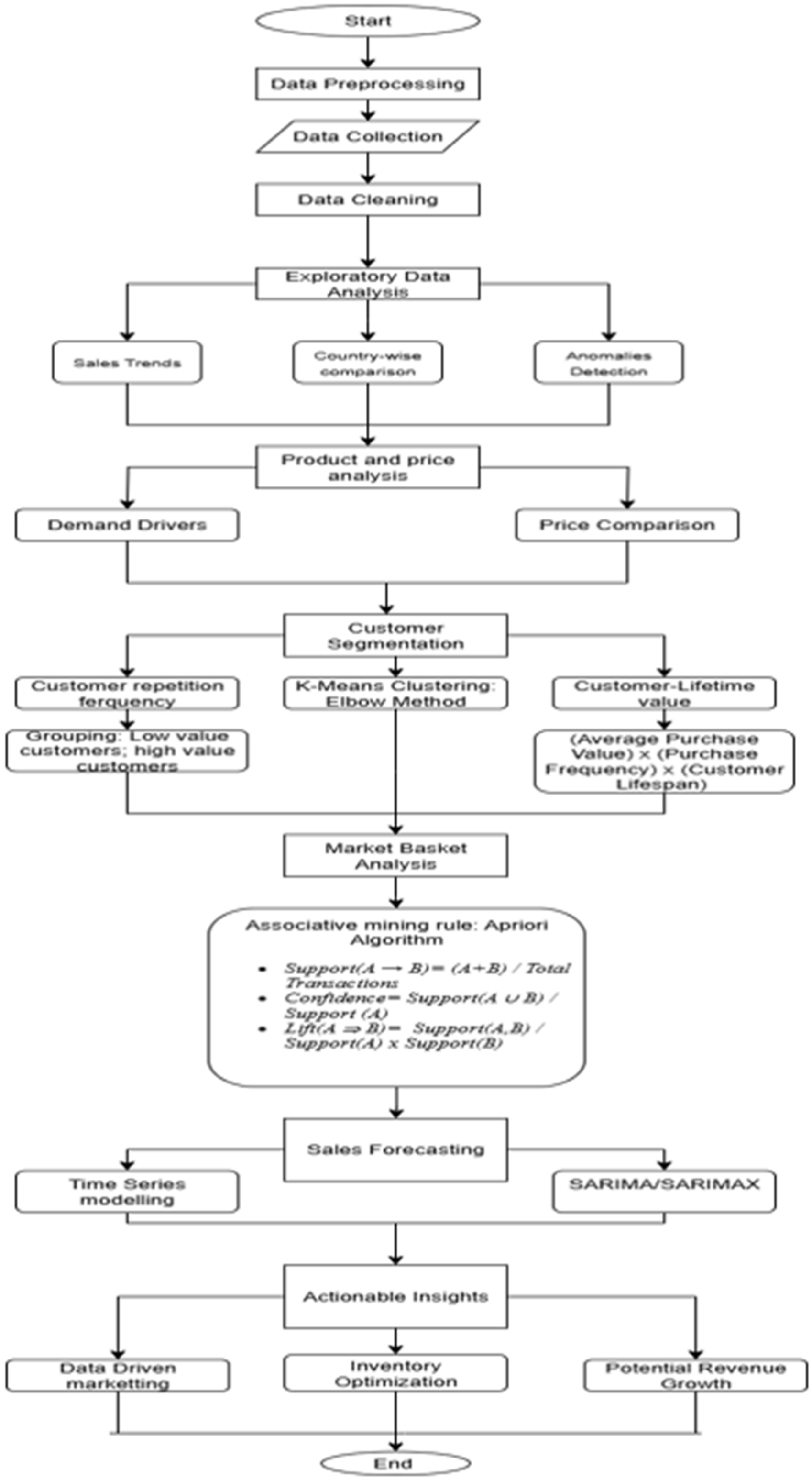



Fig.1. Data analytics architecture

## IV. PROPOSED ALGORITHM

Proposed System work for Data-Driven Retail Analytics Framework of the sales in the business following way:

1) Step 1: Import the raw retail sales dataset
2) Step 2: Inspect the dataset for inconsistencies, duplicate records, incorrect data types, crosses and wrong data.
3) Step 3: Preprocess and clean the dataset by removing duplicates and correcting data format issues.
4) Step 4: Identify and handle missing or null values using appropriate techniques.
5) Step 5: Perform basic statistical analysis to understand data distributions and ranges.
6) Step 6: Conduct Exploratory Data Analysis to visualize and understand overall sales behavior.
7) Step 7: Detect sales outliers and anomalies that may impact business insights.
8) Step 8: Analyze product-wise sales performance and revenue contribution.
9) Step 9: Perform time-based analysis to study daily, monthly, and yearly sales trends.
10) Step 10: Conduct country-wise analysis (France vs Netherlands) to compare total revenue.
11) Step 11: Compare average monthly sales between the selected countries.
12) Step 12: Analyze unit price variations across different countries and products.
13) Step 13: Segment customers based on purchase frequency and transaction history.
14) Step 14: Classify customers into repetitive buyers and one-time buyers.
15) Step 15: Extract customer-level features such as total spend and number of purchases.
16) Step 16: Apply data normalization techniques for clustering readiness.
17) Step 17: Implement the K-Means clustering algorithm for customer segmentation.
18) Step 18: Group customers into low-value, mid-value, and high-value (big spender) segments.
19) Step 19: Evaluate clustering performance using appropriate validation metrics.
20) Step 20: Calculate Customer Lifetime Value (CLV) for each customer.
21) Step 21: Rank customers based on their CLV scores.
22) Step 22: Identify the highest CLV customers in each country.
23) Step 23: Analyze purchasing behavior of high-value customers.
24) Step 24: Prepare transactional data for Market Basket Analysis (MBA).
25) Step 25: Identify frequently purchased item combinations.
26) Step 26: Apply the Apriori algorithm for association rule mining.
27) Step 27: Measure support values for frequent itemsets.
28) Step 28: Calculate confidence and lift to evaluate association strength.
29) Step 29: Interpret co-purchase patterns for business recommendations.
30) Step 30: Visualize sales trends using time series plots.
31) Step 31: Decompose the time series into trend, seasonality, and residual components.
32) Step 32: Apply transformations to make the sales data stationary.
33) Step 33: Perform stationarity tests to validate transformations.
34) Step 34: Build the SARIMA forecasting model incorporating seasonal patterns.
35) Step 35: Train the SARIMA model using historical sales data.
36) Step 36: Generate future sales forecasts for defined time horizons.
37) Step 37: Compare predicted values with actual sales to evaluate accuracy.
38) Step 38: Analyze forecasting errors using suitable performance metrics.
39) Step 39: Summarize insights derived from analytics and forecasting results.
40) Step 40: Present findings to support data-driven business decision-making.

### A. *Exploratory Data Analysis (EDA)*

The Exploratory Data Analysis is an analysis approach which is used to identify general, seasonal and trending patterns in a dataset. It is considered to be the first step after data preprocessing and cleaning [ 1]. The dataset was carefully preprocessed by removing missing value and replacing null values with mean or median from critical fields such as *CustomerID* and *Description* [2]. Only variables necessary for the framework are brought into consideration which may be *CustomerID, InvoiceNo, StockCode, Description, Quantity, UnitPrice, InvoiceDate,* and *Country* [3].

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 14 Issue I Jan 2026- Available at www.ijraset.com

Dataset understanding:

| | InvoiceNo | StockCode | Description | Quantity | InvoiceDate | UnitPrice | CustomerID | Country |
|---|---|---|---|---|---|---|---|---|
| 0 | 536365 | 85123A | WHITE HANGING HEART T-LIGHT HOLDER | 6 | 12/01/2010 08:26:00 | 2.55 | 17850.0 | United Kingdom |
| 1 | 536365 | 71053 | WHITE METAL LANTERN | 6 | 12/01/2010 08:26:00 | 3.39 | 17850.0 | United Kingdom |
| 2 | 536365 | 84406B | CREAM CUPID HEARTS COAT HANGER | 8 | 12/01/2010 08:26:00 | 2.75 | 17850.0 | United Kingdom |
| 3 | 536365 | 84029G | KNITTED UNION FLAG HOT WATER BOTTLE | 6 | 12/01/2010 08:26:00 | 3.39 | 17850.0 | United Kingdom |
| 4 | 536365 | 84029E | RED WOOLLY HOTTIE WHITE HEART. | 6 | 12/01/2010 08:26:00 | 3.39 | 17850.0 | United Kingdom |
| 5 | 536365 | 22752 | SET 7 BABUSHKA NESTING BOXES | 2 | 12/01/2010 08:26:00 | 7.65 | 17850.0 | United Kingdom |
| 6 | 536365 | 21730 | GLASS STAR FROSTED T-LIGHT HOLDER | 6 | 12/01/2010 08:26:00 | 4.25 | 17850.0 | United Kingdom |
| 7 | 536366 | 22633 | HAND WARMER UNION JACK | 6 | 12/01/2010 08:28:00 | 1.85 | 17850.0 | United Kingdom |
| 8 | 536366 | 22632 | HAND WARMER RED POLKA DOT | 6 | 12/01/2010 08:28:00 | 1.85 | 17850.0 | United Kingdom |
| 9 | 536367 | 84879 | ASSORTED COLOUR BIRD ORNAMENT | 32 | 12/01/2010 08:34:00 | 1.69 | 13047.0 | United Kingdom |

Fig.2. Sale transactional data

A new variable, TotalAmount, was created whose calculation formula is
TotalAmount = Quantity x UnitPrice
to accurately determine the contribution of each transaction towards total revenue attained .

| | CustomerID | InvoiceNo | StockCode | Description | Quantity | UnitPrice | TotalAmount | InvoiceDate | Country |
|---|---|---|---|---|---|---|---|---|---|
| 0 | 17850.0 | 536365 | 85123A | WHITE HANGING HEART T-LIGHT HOLDER | 6 | 2.55 | 15.30 | 12/01/2010 08:26:00 | United Kingdom |
| 1 | 17850.0 | 536365 | 71053 | WHITE METAL LANTERN | 6 | 3.39 | 20.34 | 12/01/2010 08:26:00 | United Kingdom |
| 2 | 17850.0 | 536365 | 84406B | CREAM CUPID HEARTS COAT HANGER | 8 | 2.75 | 22.00 | 12/01/2010 08:26:00 | United Kingdom |
| 3 | 17850.0 | 536365 | 84029G | KNITTED UNION FLAG HOT WATER BOTTLE | 6 | 3.39 | 20.34 | 12/01/2010 08:26:00 | United Kingdom |
| 4 | 17850.0 | 536365 | 84029E | RED WOOLLY HOTTIE WHITE HEART. | 6 | 3.39 | 20.34 | 12/01/2010 08:26:00 | United Kingdom |
| 5 | 17850.0 | 536365 | 22752 | SET 7 BABUSHKA NESTING BOXES | 2 | 7.65 | 15.30 | 12/01/2010 08:26:00 | United Kingdom |
| 6 | 17850.0 | 536365 | 21730 | GLASS STAR FROSTED T-LIGHT HOLDER | 6 | 4.25 | 25.50 | 12/01/2010 08:26:00 | United Kingdom |
| 7 | 17850.0 | 536366 | 22633 | HAND WARMER UNION JACK | 6 | 1.85 | 11.10 | 12/01/2010 08:28:00 | United Kingdom |
| 8 | 17850.0 | 536366 | 22632 | HAND WARMER RED POLKA DOT | 6 | 1.85 | 11.10 | 12/01/2010 08:28:00 | United Kingdom |
| 9 | 13047.0 | 536367 | 84879 | ASSORTED COLOUR BIRD ORNAMENT | 32 | 1.69 | 54.08 | 12/01/2010 08:34:00 | United Kingdom |

Fig.3. Sale transactional data with added total amount

A country-based analysis was carried out which compared the overall sales performance of all countries present in the dataset. The country of the United Kingdom is observed to have a global domination with the highest number of recorded Transactions, thus contributing the most to the highest total global revenue. Other than the UK, Germany, Netherlands and France appear to be powerful contributors to total sales revenue [4-6].

```
Total sales by country:
Country
United Kingdom        8187806.364
Netherlands            284661.540
EIRE                   263276.820
Germany                221698.210
France                 197403.900
Australia              137077.270
Switzerland             56385.350
Spain                   54774.580
Belgium                 40910.960
Sweden                  36595.910
Japan                   35340.620
Norway                  35163.460
Portugal                29367.020
Finland                 22326.740
Channel Islands         20086.290
Denmark                 18768.140
Italy                   16890.510
Cyprus                  12946.290
Austria                 10154.320
Hong Kong               10117.040
Singapore                9120.390
Israel                   7907.820
Poland                   7213.140
...
Czech Republic            707.720
Bahrain                   548.400
Saudi Arabia              131.170
```

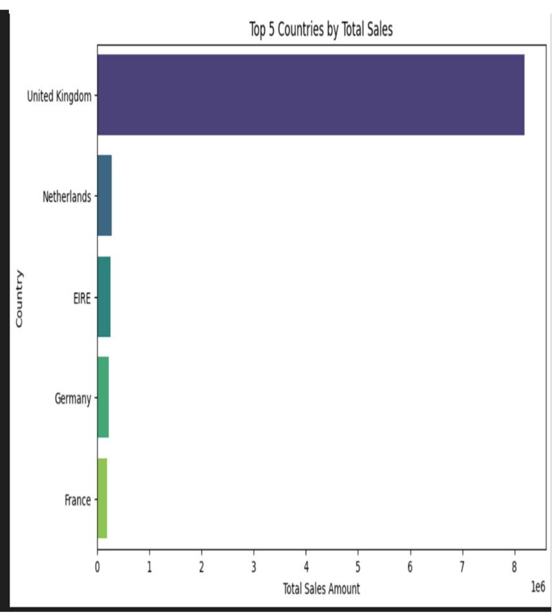



Fig.4. Country-Wise Sales Comparison

The average monthly sales of France and Netherlands were compared to properly analyze their revenue patterns. The analysis revealed the Netherlands consistently having higher average sales compared to the country of France. Dutch sales exhibited higher peaks amidst seasonal periods, emphasizing on having higher order volumes or larger basket sizes during demand surges. By contrast, France demonstrated relatively lower but stable revenue trends, with smaller fluctuations across the sales period months [7].

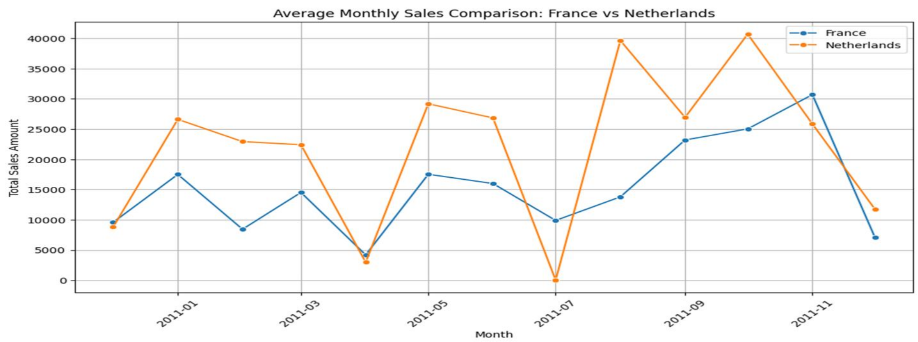



Fig.5. Average sales comparison of France and Netherlands

To Examine the differences in UnitPrices of items sold in France and Netherlands, comparison of product prices took place. The results demonstrate that the majority of top selling products in the Netherlands have a higher Unitprice than products of France. We can speculate about Dutch customers purchasing relatively more expensive items, which can be a supposed answer to their higher total sales revenue despite having a lower customer ground [8, 9].

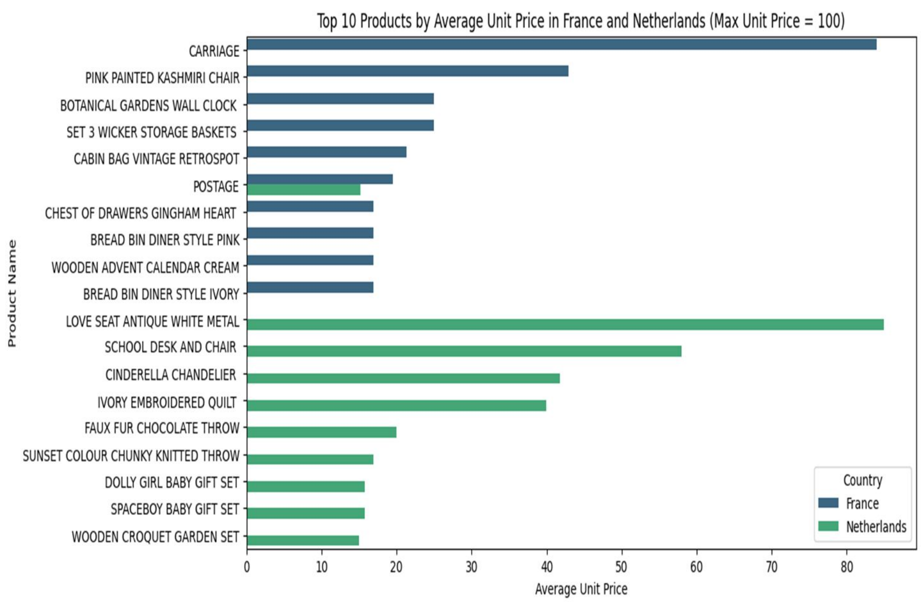



Fig.6. UnitPrice per product of both countries

### B. *Customer Segmentation*

Customer segmentation is the process of partitioning the customers of a supposed retail entity into individual groups based on shared characteristics such as buying behavior, demographics, or value. In analytics, it helps businesses identify high-value vs. low-value customers, tailored marketing strategies, and improve customer retention [10, 11].

When the dataset was segmented to analyze unique customers in each country, a massive difference was observed between France and the Netherlands. France had a much larger pool of unique customers, whereas the Netherlands had significantly fewer unique customers. However, despite this difference, the Netherlands consistently generated higher total sales revenue.This unique customer disparity emphasizes the singular appearance of that CustomerId and its presence in data.

```
country_unique_customers = df.groupby('Country')['CustomerID'].nunique().sort_values(ascending=False)
print("Total unique customers by country:")
country_unique_customers.head(20)
```
```
Total unique customers by country:

Country
United Kingdom     3950
Germany              95
France               87
Spain                31
Belgium              25
Switzerland          21
Portugal             19
Italy                15
Finland              12
Austria              11
Norway               10
Channel Islands       9
Netherlands           9
Australia             9
Denmark               9
Cyprus                8
Japan                 8
Sweden                8
Poland                6
Unspecified           4
Name: CustomerID, dtype: int64
```

Fig.6. Unique customers in countries

The customer distribution analysis revealed a striking imbalance between France and the Netherlands. France had 87 unique customers, whereas the Netherlands had only 9 unique customers. Another important finding from the customer analysis was that both France and the Netherlands exhibited a 100% repetition rate [12, 13]. This means that every customer in both countries has made more than one purchase during the observed period. Such a pattern highlights the presence of strong customer loyalty across both markets.

```
#finding how many customer ids are repeated and how many are unique

repeated_customers = df_france['CustomerID'].value_counts()[df_france['CustomerID'].value_counts() > 1]
unique_customers = df_france['CustomerID'].value_counts()[df_france['CustomerID'].value_counts() == 1]
print(f"Number of repeated customers: {len(repeated_customers)}")
print(f"Number of unique customers: {len(unique_customers)}")
```
```
Number of repeated customers: 87
Number of unique customers: 0
```
```
#finding how many customer ids are repeated and how many are unique

repeated_customers_netherlands = df_netherlands['CustomerID'].value_counts()[df_netherlands['CustomerID'].value_counts() > 1]
unique_customers_netherlands = df_netherlands['CustomerID'].value_counts()[df_netherlands['CustomerID'].value_counts() == 1]
print(f"Number of repeated customers in Netherlands: {len(repeated_customers_netherlands)}")
print(f"Number of unique customers in Netherlands: {len(unique_customers_netherlands)}")
```
```
Number of repeated customers in Netherlands: 9
Number of unique customers in Netherlands: 0
```

Fig.7. Repeated customers distribution

The Netherlands showed an even more remarkable trend in terms of customer return frequency. Specifically, one customer in the Netherlands demonstrated a very high rate of repeated purchases, far surpassing the activity levels seen in France. This single customer contributed significantly to the overall sales volume of the Netherlands, emphasizing the importance of high-value individuals in driving revenue.

```
Customer ID 14646.0 is repeated 2085 times in Netherlands.
Customer ID 12759.0 is repeated 95 times in Netherlands.
Customer ID 12775.0 is repeated 67 times in Netherlands.
Customer ID 12778.0 is repeated 51 times in Netherlands.
Customer ID 12802.0 is repeated 26 times in Netherlands.
Customer ID 12790.0 is repeated 21 times in Netherlands.
Customer ID 12787.0 is repeated 20 times in Netherlands.
Customer ID 12789.0 is repeated 4 times in Netherlands.
Customer ID 12791.0 is repeated 2 times in Netherlands.
```
```
Top 5 repeated customers:
CustomerID
12681.0     646
12682.0     525
12567.0     463
12637.0     394
12683.0     362
```

Fig.8. Maximum repeated customers

A significant difference between France and the Netherlands is that the most repeated customer in the Netherlands appeared an astonishing 2,085 times, whereas the highest repeated customer in France appeared 646 times. This clearly shows that Dutch customers, particularly a few key individuals, have much higher purchase frequencies compared to their French counterparts.

Clustering is an unsupervised machine learning technique used to combine data points of similar algorithms and group them together simultaneously into clusters. The idea is to categorize data points into clusters that share more similarities compared to those in other clusters [14,15].

```
Number of customers in each cluster:
   Cluster  Number of Customers
0        0                   87
1        1                    1
2        2                    1
```

```
Number of customers in each cluster in Netherlands:
   Cluster  Number of Customers
0        0                    9
1        1                    1
2        2                    8
```

Fig.9. Clustered customer data

K-means is a repetitious clustering algorithm that divides a dataset into *k* clusters by reducing the within-cluster sum of squared distances, also called inertia [16]. To identify big spenders, the K-means clustering algorithm was applied using the RFM model, where customer value is calculated as Monetary Value = Unit Price × Quantity [17].

The Elbow Method is a clustering technique used in k- means clustering to find the optimal number of clusters in a dataset. The Elbow Method was exercised to ascertain the optimum number of clusters, and customers were grouped into low, medium, and high-value spenders [18-20]. In France, the big spender cluster (Cluster 1) had an average spending of 4,161.06, with Customer ID 12536.0 emerging as the top spender. In contrast, the Netherlands showed a far stronger cluster, where Customer ID 14646.0 recorded a total spending of 28,276.21, making Dutch customers significantly more valuable despite being fewer in number. This comparison highlights that while France relies on broader participation, the Netherlands' revenue is dominated by a small cluster of extremely high-value customers.

```python
#finding who are the big spenders in each cluster
big_spenders = df_france.groupby('Cluster')['UnitPrice'].mean().reset_index()
big_spenders.columns = ['Cluster', 'Average Spending']
print("Average spending in each cluster:")
print(big_spenders)
```

```
Average spending in each cluster:
   Cluster  Average Spending
0        0          3.306386
1        1       4161.060000
2        2       1136.300000
```

```python
#biggest spender in france
biggest_spender = big_spenders.loc[big_spenders['Average Spending'].idxmax()]
print(f"The biggest spender in France is in Cluster {biggest_spender['Cluster']} with an average spending of {biggest_spender['Average Spending']:.2
#lets find out which customer id is the biggest spender
biggest_spender_id = df_france[df_france['Cluster'] == biggest_spender['Cluster']]['CustomerID'].value_counts().idxmax()
print(f"The biggest spender in France is Customer ID {biggest_spender_id}.")
#and what is his average spending
average_spending_biggest_spender = df_france[df_france['CustomerID'] == biggest_spender_id]['UnitPrice'].mean()
print(f"The average spending of the biggest spender in France is {average_spending_biggest_spender:.2f}.")
```

```
The biggest spender in France is in Cluster 1.0 with an average spending of 4161.06.
The biggest spender in France is Customer ID 12536.0.
The average spending of the biggest spender in France is 48.55.
```

```
   Cluster  Total Spending
0        0        28276.21
1        1        12483.18
2        2         2272.60
```

```python
#biggest spender in netherlands
biggest_spender_netherlands = top_spenders.loc[top_spenders['Total Spending'].idxmax()]
print(f"The biggest spender in Netherlands is in Cluster {biggest_spender_netherlands['Cluster']} with a total spending of {biggest_spender_netherla
(variable) biggest_spender_netherlands_id: int | str   in Netherlands
biggest_spender_netherlands_id = df_netherlands[df_netherlands['Cluster'] == biggest_spender_netherlands['Cluster']]['CustomerID'].value_counts().id
print(f"The biggest spender in Netherlands is Customer ID {biggest_spender_netherlands_id}.")
#and what is his average spending
average_spending_biggest_spender_netherlands = df_netherlands[df_netherlands['CustomerID'] == biggest_spender_netherlands_id]['UnitPrice'].mean()
print(f"The average spending of the biggest spender in Netherlands is {average_spending_biggest_spender_netherlands:.2f}.")
```

```
The biggest spender in Netherlands is in Cluster 0.0 with a total spending of 28276.21.
The biggest spender in Netherlands is Customer ID 14646.0.
The average spending of the biggest spender in Netherlands is 2.59.
```

Fig.10. K-Means Clustering for Big-Spenders

Customer lifetime value (CLV) is an important criterion used to forecast and project the gross revenue a business can anticipate from a customer over the course of their relationship [21, 22]. The formula for CLV calculation is standardized as:

CLV = (Average Purchase Value) × (Purchase Frequency) × (Customer Lifespan)

Average Purchase Value (APV): Calculated as the total revenue generated by a customer divided by the number of unique orders placed.

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
*ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538*
*Volume 14 Issue I Jan 2026- Available at www.ijraset.com*

```
Customer Lifetime Value (CLV) for all French customers:
    CustomerID          CLV
80     12731.0  3569.927227
52     12678.0  3341.001354
55     12681.0  2598.144825
56     12682.0  2334.225197
20     12567.0  1731.440568
..         ...          ...
42     12651.0    21.655022
82     12734.0    20.971179
46     12659.0    17.415197
60     12686.0    16.932707
9      12506.0    13.961790
```

```
Customer Lifetime Value (CLV) for all Netherlands customers:
   CustomerID           CLV
8     14646.0  24904.962178
0     12759.0    136.464950
1     12775.0    120.407525
2     12778.0     71.664059
3     12787.0     41.316238
7     12802.0     36.764554
5     12790.0     28.953267
6     12791.0     17.162376
4     12789.0      8.184653
```

Fig.11. CLV comparison between Netherlands and France

Purchase Frequency (PF): Determined by the number of orders made by a customer relative to the average order activity. Customer Lifespan (CL): Approximated as the observed time period of the dataset (1 year in this case). By applying this formula, CLV was computed for every customer in both France and the Netherlands.

The results emphasized a major distinction between the CLV's of two nations. France showed a relatively lower average CLV despite having a bigger customer base than Netherlands. Contrastingly, Netherland *CustomerID* 14646.0 showed a significantly bigger CLV due frequent purchases and larger order value. The top *CustomerID* from France 12731.0 showed a smaller CLV despite having higher per-unit spending.

```
Orders and prices for highest CLV customer in Netherlands (Customer ID 14646.0):
        InvoiceNo          InvoiceDate                     Description  \
37952      539491  12/20/2010 10:09:00      PACK OF 12 WOODLAND TISSUES
37953      539491  12/20/2010 10:09:00   PACK OF 12 PINK POLKADOT TISSUES
37954      539491  12/20/2010 10:09:00  SET OF 3 CAKE TINS PANTRY DESIGN
37955      539491  12/20/2010 10:09:00            JUMBO STORAGE BAG SUKI
37956      539491  12/20/2010 10:09:00       PACK OF 20 SPACEBOY NAPKINS
...           ...                  ...                               ...
534958     581176  12/07/2011 15:19:00      PACK OF 20 NAPKINS RED APPLES
534959     581176  12/07/2011 15:19:00   PACK OF 20 NAPKINS PANTRY DESIGN
534960     581176  12/07/2011 15:19:00           SPACEBOY BIRTHDAY CARD
534961     581176  12/07/2011 15:19:00                CARD DOLLY GIRL
537783     581338  12/08/2011 12:12:00        JUMBO BAG 50'S CHRISTMAS

        Quantity  UnitPrice  TotalPrice
37952         12       0.29        3.48
37953         12       0.29        3.48
37954          2       4.95        9.90
37955          1       1.95        1.95
37956          2       0.85        1.70
...          ...        ...         ...
534958        96       0.72       69.12
534959        96       0.72       69.12
534960        72       0.36       25.92
534961        72       0.36       25.92
...
```

```
Products bought by highest CLV customer in France (ID 12731.0):
        InvoiceNo          InvoiceDate                       Description  \
23273      538196  12/10/2010 10:56:00                           POSTAGE
23274      538196  12/10/2010 10:56:00       ASSORTED COLOUR BIRD ORNAMENT
23275      538196  12/10/2010 10:56:00        INFLATABLE POLITICAL GLOBE
23276      538196  12/10/2010 10:56:00          LUNCH BAG RED RETROSPOT
23277      538196  12/10/2010 10:56:00                 PICTURE DOMINOES
...           ...                  ...                               ...
472216     576672  11/16/2011 11:59:00        LUNCH BAG DOLLY GIRL DESIGN
472217     576672  11/16/2011 11:59:00    PACK OF 72 RETROSPOT CAKE CASES
472218     576672  11/16/2011 11:59:00                  SOLDIERS EGG CUP
472219     576672  11/16/2011 11:59:00  RED RETROSPOT CHILDRENS UMBRELLA
472220     576672  11/16/2011 11:59:00           MINI FUNKY DESIGN TAPES

        Quantity  UnitPrice  TotalPrice
23273          8      18.00       144.0
23274        120       1.69       202.8
23275         36       0.85        30.6
23276         10       1.65        16.5
23277         24       1.45        34.8
...          ...        ...         ...
472216        20       1.65        33.0
472217        24       0.55        13.2
472218        24       1.25        30.0
472219        24       3.25        78.0
```

Fig.12. Products by Highest CLV customers

The analysis of the highest CLV customer14646.0 of the Netherlands provides us with a valuable perspective on customer's purchase behaviour patterns. This *CustomerID* consistently purchased lower priced items but ordered in large volumes which eventually resulted in the observed exceptionally high CLV. In contrast, the ID 12731.0 from France showed a contrary behaviour. This customer approached to buy products with relatively higher UnitPricebut had comparatively had a very handful number of transactions.

### *C. Market Basket Analysis(MBA)*

Market Basket Analysis is a favoured data mining methodology used to recognize and identify relationships and patterns between products which are habitually bought in conjunction. The primary objective of market basket analysis is to detect hidden patterns and trends in customer approaches and transactions, which can be implemented to cross-selling, to bundle and products and design recommendation systems [23, 25]. In this aspect, MBA was performed using the Apriori Algorithm.

The Apriori algorithm is a classic method in Market Basket Analysis that helps discover frequent itemsets and generate association rules. The Apriori algorithm efficiently finds strong associations between products by eliminating impossible candidates early using the Apriori property. This makes it especially useful in retail and e-commerce to recommend product bundles, optimize store layouts, or design promotions [24, 25].

**Support:** Measures how frequently an item or itemset appears in the entire data inventory. It is calculated as the ratio of transactions that contain the item(s). A higher support value means the product or combination is commonly purchased, making it more significant for sales strategies. Support helps identify how common certain product combinations are specifically. A higher support value means the rule is based on a larger portion of the dataset, making it more statistically significant [23, 24, 25].

*Support (A → B) = Number of transactions containing {A, B} / Total Transactions*
(1)

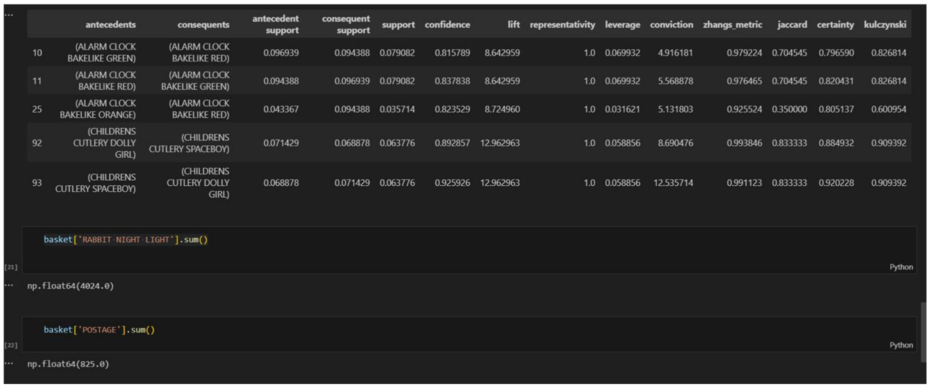
International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 14 Issue I Jan 2026- Available at www.ijraset.com

1) Confidence: It is a measure of the reliability of an association rule. It tells us the probability that a customer who bought item A will also buy item B. Confidence was one of the main metrics used to evaluate whether the association rules are actually meaningful [23, 24, 25].

$$Confidence = Support\ (A \cup B) / Support\ (A) \quad (2)$$

2) Lift: It is a measure that tells us how much more likely two items are to be bought together compared to them being bought independently. An example interpretation can be that the customers who buy bread are 1.875 times more likely to also buy butter compared to random chance [23, 24, 25].

$$Lift\ (A \Rightarrow B) = Support\ (A, B) / Support(A)\ x\ Support(B) \quad (3)$$



| | antecedents | consequents | antecedent support | consequent support | support | confidence | lift | representativity | leverage | conviction | zhangs_metric | jaccard | certainty | kulczynski |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 10 | (ALARM CLOCK BAKELIKE GREEN) | (ALARM CLOCK BAKELIKE RED) | 0.096939 | 0.094388 | 0.079082 | 0.815789 | 8.642959 | 1.0 | 0.069932 | 4.916181 | 0.979224 | 0.704545 | 0.796590 | 0.826814 |
| 11 | (ALARM CLOCK BAKELIKE RED) | (ALARM CLOCK BAKELIKE GREEN) | 0.094388 | 0.096939 | 0.079082 | 0.837838 | 8.642959 | 1.0 | 0.069932 | 5.568878 | 0.976465 | 0.704545 | 0.820431 | 0.826814 |
| 25 | (ALARM CLOCK BAKELIKE ORANGE) | (ALARM CLOCK BAKELIKE RED) | 0.043367 | 0.094388 | 0.035714 | 0.823529 | 8.724960 | 1.0 | 0.031621 | 5.131803 | 0.925524 | 0.350000 | 0.805137 | 0.600954 |
| 92 | (CHILDRENS CUTLERY DOLLY GIRL) | (CHILDRENS CUTLERY SPACEBOY) | 0.071429 | 0.068878 | 0.063776 | 0.892857 | 12.962963 | 1.0 | 0.058856 | 8.690476 | 0.993846 | 0.833333 | 0.884932 | 0.909392 |
| 93 | (CHILDRENS CUTLERY SPACEBOY) | (CHILDRENS CUTLERY DOLLY GIRL) | 0.068878 | 0.071429 | 0.063776 | 0.925926 | 12.962963 | 1.0 | 0.058856 | 12.535714 | 0.991123 | 0.833333 | 0.920228 | 0.909392 |

```
[21] basket['RABBIT NIGHT LIGHT'].sum()
```
Python

```
np.float64(4024.0)
```

```
[22] basket['POSTAGE'].sum()
```
Python

```
np.float64(825.0)
```

Fig.13. Market basket analysis on France from the Dataset

*D. Sales Forecasting using SARIMA models*

Sales forecasting was carried out using time series analysis. A time series is a chronological succession of numerical observations accumulated over time and arranged in chronological order. Mathematically, it can be defined as a stochastic process {Yt}, where each observation Yt corresponds to a value recorded at a specific time point $t \in T$. Time series data exhibit temporal dependence; that is, the value at time t is often correlated with values at preceding time points ($t-1$, $t-2$, … )[26-28, 31].

To evaluate country-level sales performance, a comparative analysis of Total Price was conducted between France and the Netherlands. The Total Price was computed as:

TotalPrice= UnitPrice x Quantity

For both countries, transactional records were filtered and aggregated on a daily basis. The resulting time series for France and the Netherlands were visualized using line graphs, allowing for a direct comparison of revenue patterns.

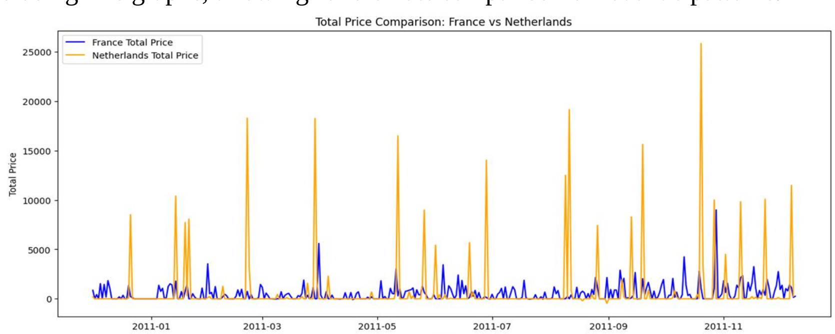



Fig.14. Total Price per item over sales period

The line plot shows that the Netherlands has a much higher total price compared to France, which indicates higher sales in the Netherlands despite having a lower number of customers than France.

For time series forecasting, Seasonal Autoregressive Integrated Moving Average (SARIMA) models were employed, as they exhibit favourable approaches for data mining on trends and seasonality.

The SARIMA model is an extension of the ARIMA model designed to handle time series data along with seasonal patterns. It combines non-seasonal components(p,d,q) with seasonal components (P,D,Q,s), where ‘s’ denotes the seasonal period [27, 29, 30].

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 14 Issue I Jan 2026- Available at www.ijraset.com

By incorporating both trend and seasonality, SARIMA effectively models complex temporal dependencies and cyclic behaviors. This makes it particularly useful for forecasting real-world data such as sales, climate, and financial time series that exhibit recurring seasonal fluctuations.

```
model = sm.tsa.statespace.SARIMAX(y,
                                  order = param,
                                  seasonal_order=param_seasonal,
                                  enforce_stationarity=False,
                                  enforce_invertibility=False)

#results
results = model.fit()
print(results.summary())

                                      SARIMAX Results
==========================================================================================
Dep. Variable:                     TotalPrice   No. Observations:                      374
Model:             SARIMAX(0, 1, 1)x(0, 1, 1, 30)   Log Likelihood               -2572.653
Date:                        Wed, 13 Aug 2025   AIC                           5151.307
Time:                                18:00:18   BIC                           5162.526
Sample:                            12-01-2010   HQIC                          5155.791
                                 - 12-09-2011
Covariance Type:                          opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
ma.L1         -1.0211      0.020    -50.852      0.000      -1.060      -0.982
ma.S.L30      -1.0172      0.047    -21.794      0.000      -1.109      -0.926
sigma2      7.396e+05   6.03e-08   1.23e+13      0.000     7.4e+05     7.4e+05
===================================================================================
Ljung-Box (L1) (Q):                   0.53   Jarque-Bera (JB):              8026.05
Prob(Q):                              0.47   Prob(JB):                         0.00
Heteroskedasticity (H):               2.74   Skew:                             3.56
Prob(H) (two-sided):                  0.00   Kurtosis:                        26.85
===================================================================================
```

Fig.15. SARIMAX results for France

The SARIMA model (0,1,1) × (0,1,1,30) was fitted to the French sales data, with 374 observations in the sample. The model achieved an AIC of 5151.30, BIC of 5162.53, and HQIC of 5155.79, indicating a reasonably good fit. The coefficients for the non-seasonal MA (1) and seasonal MA (30) terms were both significant, confirming the presence of short-term and seasonal dependencies in the data. Diagnostic tests such as the Ljung-Box and Jarque-Bera suggested residual non-normality, but the model remained effective in capturing the series' temporal dynamics for forecasting purposes.

```
                                      SARIMAX Results
==========================================================================================
Dep. Variable:                     TotalPrice   No. Observations:                      373
Model:             SARIMAX(0, 0, 1)x(0, 1, 1, 30)   Log Likelihood               -2957.543
Date:                        Wed, 13 Aug 2025   AIC                           5921.085
Time:                                18:01:55   BIC                           5932.305
Sample:                            12-01-2010   HQIC                          5925.570
                                 - 12-08-2011
Covariance Type:                          opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
ma.L1         -0.0131      0.086     -0.152      0.880      -0.182       0.156
ma.S.L30      -1.0694      0.023    -45.928      0.000      -1.115      -1.024
sigma2      9.325e+06   1.86e-09   5.02e+15      0.000    9.33e+06    9.33e+06
===================================================================================
Ljung-Box (L1) (Q):                   0.00   Jarque-Bera (JB):              8812.02
Prob(Q):                              0.97   Prob(JB):                         0.00
Heteroskedasticity (H):               1.47   Skew:                             4.71
Prob(H) (two-sided):                  0.05   Kurtosis:                        27.32
===================================================================================
```

Fig.16. SARIMAX results for Netherlands

SARIMAX stands for Seasonal Autoregressive Integrated Moving Average with eXogenous variables. It is an extension of the SARIMA model that not only accounts for trend, seasonality, and autocorrelation in a time series but also allows the inclusion of exogenous variables (X) — external factors that may influence the dependent variable [26,32].

The Diagnostic Plots for forecast data include

1) Standardized Residuals: Standardized residuals are the dissimilarities between the actual values and the model's predicted values, scaled by their standard deviation. They help in identifying whether errors fluctuate randomly around zero. If residuals show patterns or trends, it indicates the model has not fully captured the structure of the data.
2) Histogram plus Estimated Density: This is a histogram of the residuals overlaid with a smooth density curve (KDE) and a reference normal distribution. It is used to check whether residuals are normally distributed. If the histogram roughly matches the normal curve, the normality assumption holds; otherwise, deviations suggest non-normal residuals.
3) Normal Q-Q Plot: A Q-Q (Quantile-Quantile) plot compares the quantiles of the residuals with the quantiles of a theoretical normal distribution. If residuals are normally distributed, the points will lie close to the diagonal line. Systematic deviations (especially at the tails) indicate heavy tails, skewness, or non-normality.

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
*ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538*
*Volume 14 Issue I Jan 2026- Available at www.ijraset.com*

4) Correlogram (ACF of Residuals): A correlogram shows the autocorrelation function (ACF) of residuals at different lags, along with confidence intervals. Ideally, residual autocorrelations should be close to zero, lying inside the confidence bands. If significant spikes appear, it means the model has not fully captured time dependencies in the data.

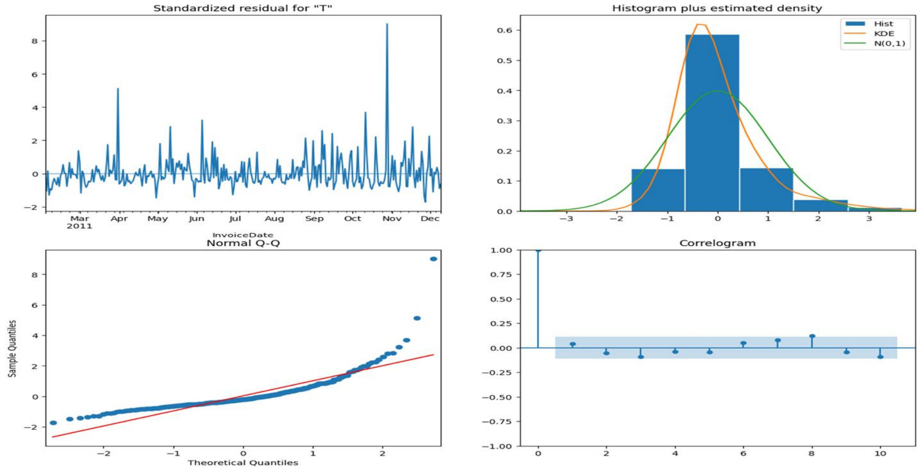



Fig.17. Diagnostic Plots for France

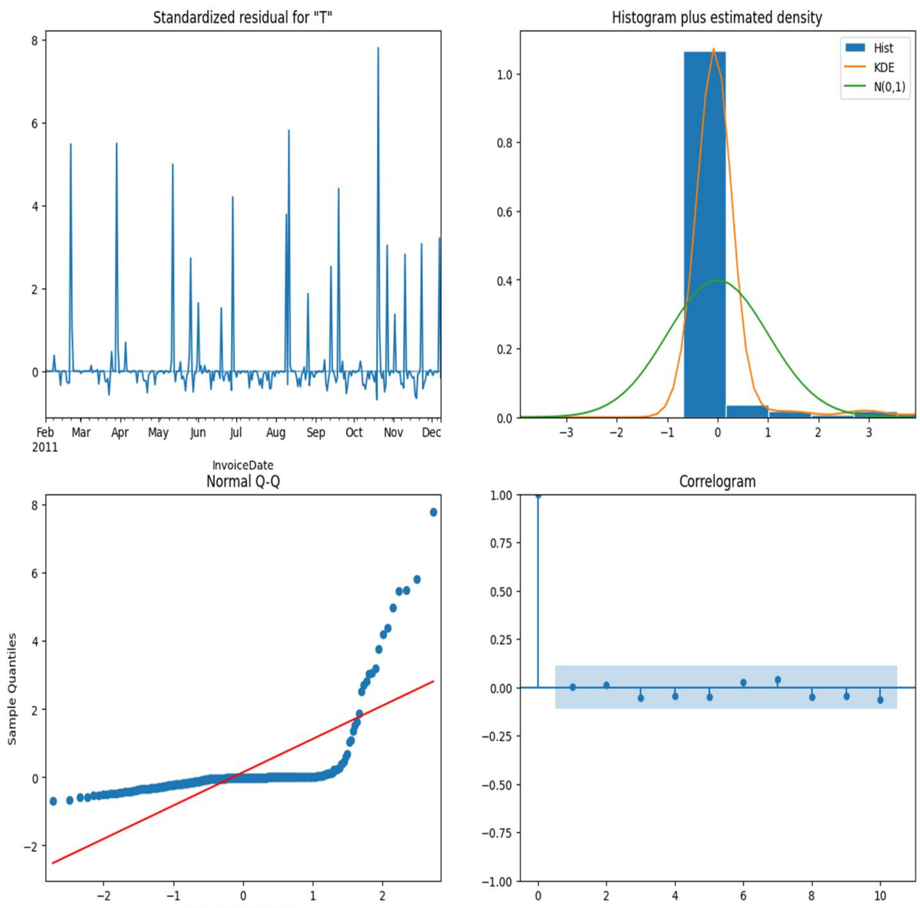



Fig.18. Diagnostic Plots for Netherlands

We performed a one-step ahead forecast using the SARIMAX model.
One-Step Ahead Forecast: The model predicts the next value in the series using all the information available up to the previous time point. After each new observation, the model updates and generates the next forecast [33].

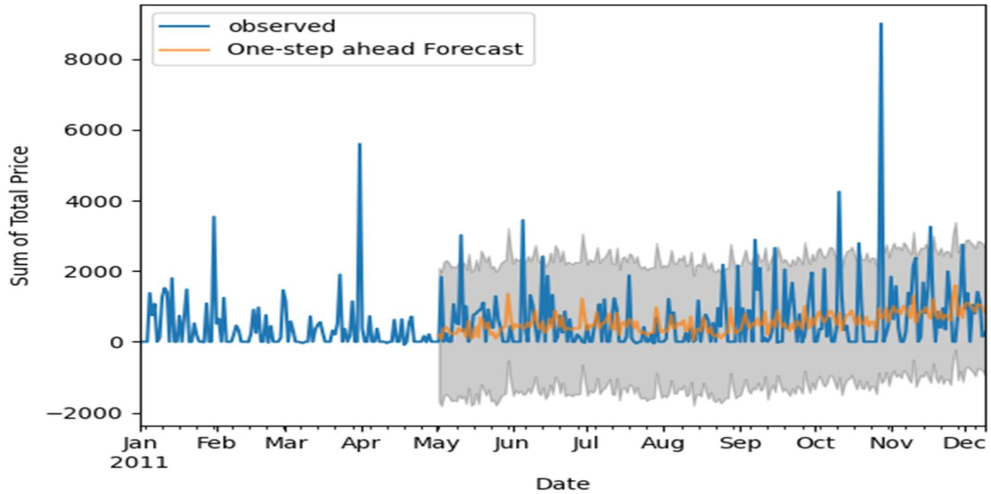



Fig.19. One-Step ahead forecast for France

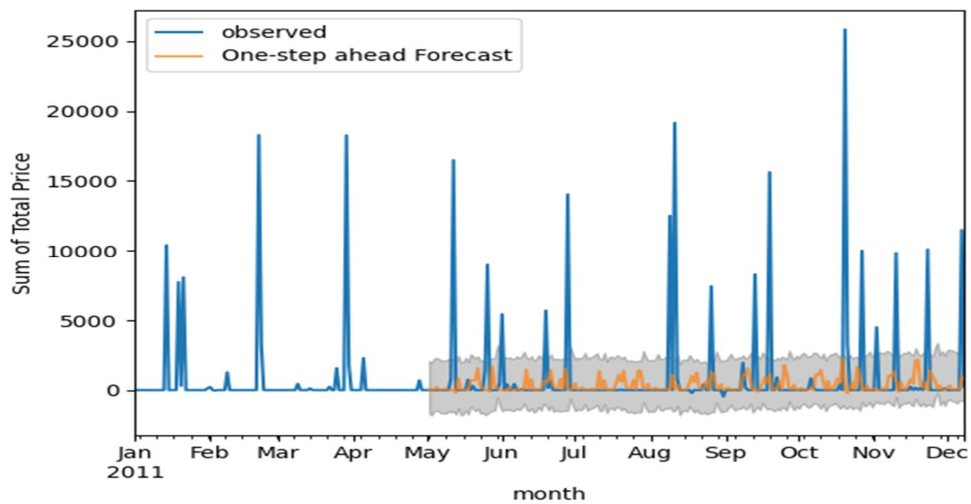



Fig.20. One-Step ahead forecast for Netherlands

In addition to one-step ahead forecasting, we also employed a dynamic forecast using the SARIMAX model.

Dynamic forecast: A long-term prediction method where the model generates future values by using its own previously predicted values, instead of actual observed data, as inputs for subsequent forecasts.

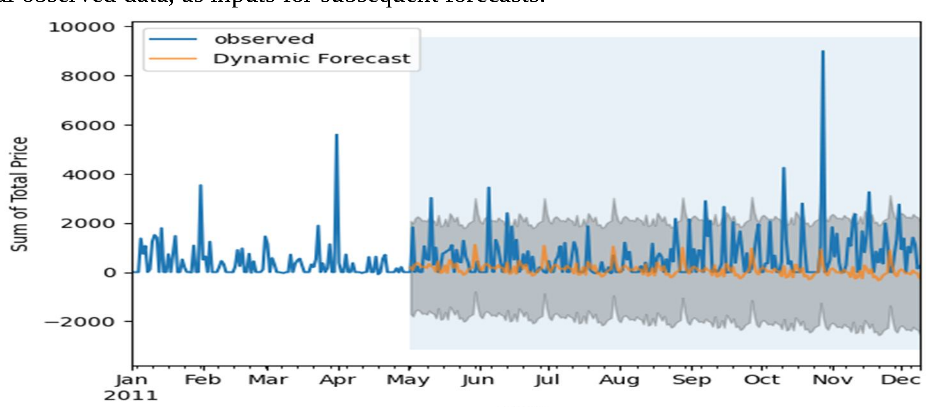



Fig.21. Dynamic forecast for France

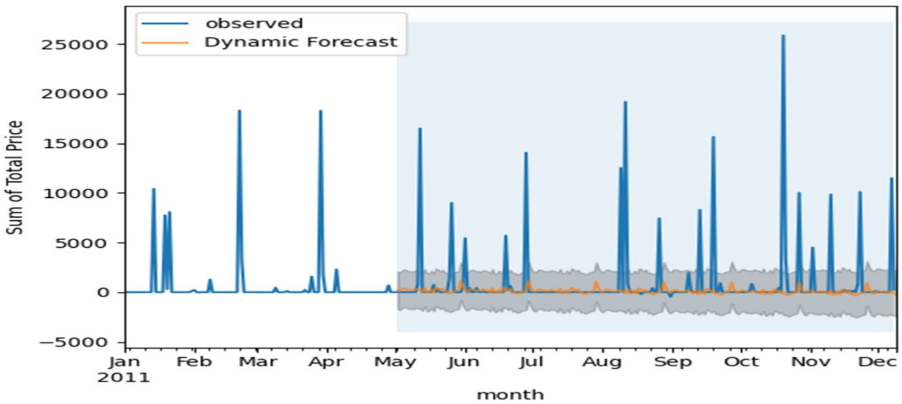



Fig.22. Dynamic forecast for Netherlands

The final forecast for both the countries represents the model's predicted future values of the time series based on historical acumulated data and fitted parameters. Along with point forecasts, the model also generates a confidence interval, which provides an estimated range within which the true future values are expected to lie with a specified probability.

The final forecast predicts an expected sales total for January 2012 which is a month ahead from the used data.

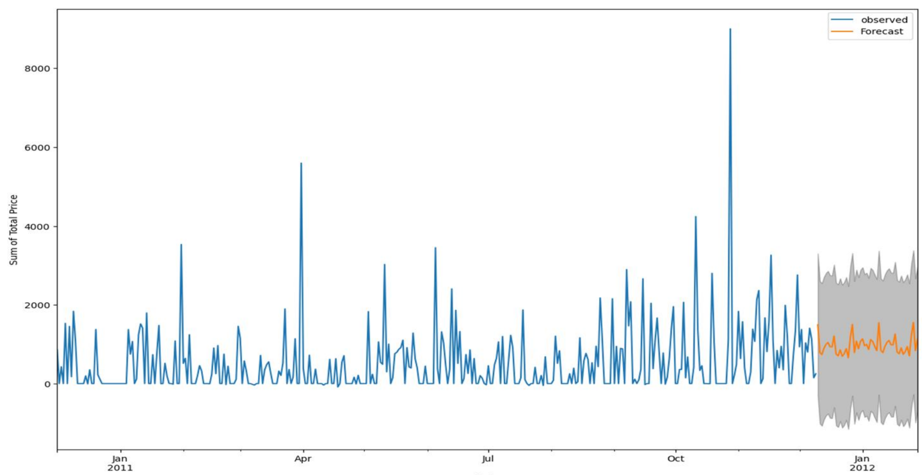



Fig.23. Final forecast for France

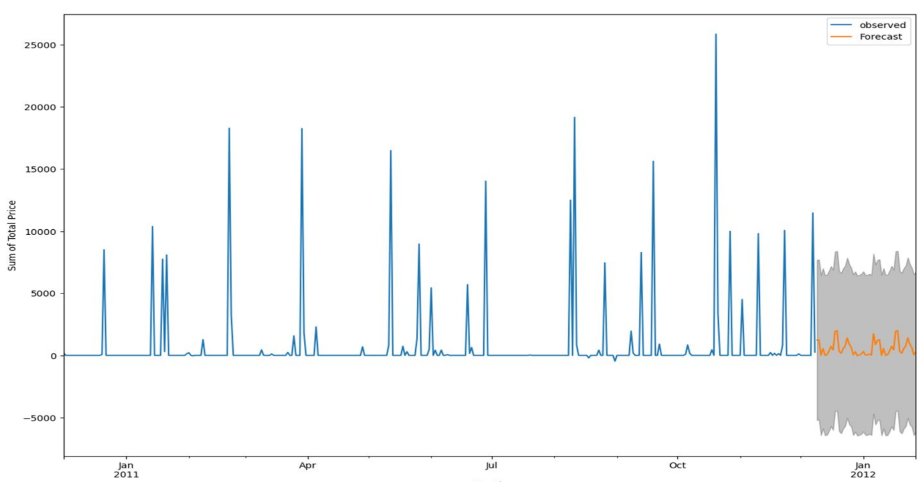



Fig.24. Final forecast for Netherlands

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 14 Issue I Jan 2026- Available at www.ijraset.com

The shaded region observed in the graph is the confidence interval.

A confidence interval is a statistical range, derived from sample data, that is likely to contain the true value of an unknown population parameter with a specified level of confidence. It is expressed as an interval estimate around a point estimate, accounting for sampling variability and uncertainty.

*1) Root Mean Squared Error (RMSE)*

RMSE is a commonly used metric to evaluate the accuracy of forecasting models. It measures the square root of the average squared differences between the predicted values and the actual observed values. A lower RMSE indicates that the model's predictions are closer to the true values, meaning the model performs better.

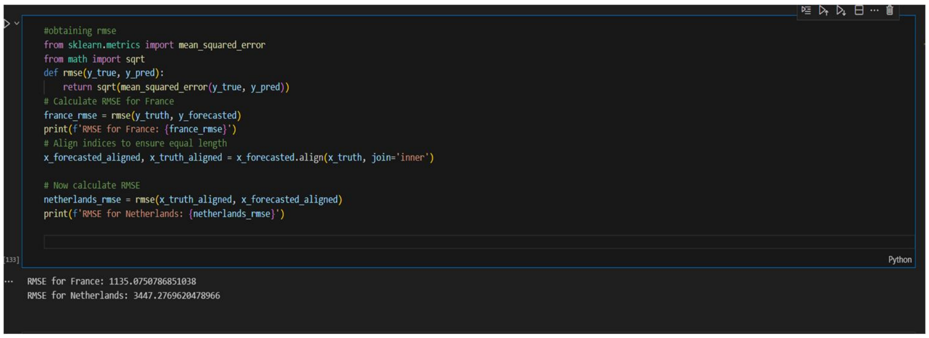



Fig.25. RMSE values

In this case, France has a much lower RMSE (1135.07) compared to the Netherlands (3447.28). This suggests that the forecasting model fits the French data more accurately, while the predictions for the Netherlands show higher deviations from the actual sales, implying more variability or difficulty in capturing the sales patterns due variability of reasons supposed as lower customer ratio yet a significant number of sales, singular customer repetition in excess, Greater Variability in Purchase Frequency, Overfitting to Outliers, etc.

## V. CONCLUSION

This paper presents a comprehensive study of Data Analysis and Forecasting on Supermarket Sales Transactions, where multiple analytical techniques were applied to extract meaningful insights and predict future sales behavior. Through Exploratory Data Analysis (EDA), the dataset was examined to uncover hidden patterns, identify anomalies, and summarize key attributes using statistical methods and visualization techniques. A country-wise comparison revealed that the Netherlands, despite having only 9 unique customers compared to France's 87, generated significantly higher total revenue, highlighting the impact of high-value, loyal customers. Further investigation into product price comparisons showed that the majority of top products in the Netherlands carried higher unit prices than those in France, reinforcing the revenue disparity. The computation of Customer Lifetime Value (CLV) provided a measure of long-term profitability, where the Netherlands exhibited customers with far superior CLV scores due to high purchase frequency and bulk ordering, while France displayed broader but lower-value spending behavior. Market Basket Analysis (MBA) and the Apriori algorithm were employed to detect strong product associations, enabling recommendations for product bundling and shelf placement to further drive sales. Finally, time-series forecasting using SARIMA was performed to model seasonal trends and predict future sales, producing accurate and reliable results that can guide inventory management and business planning. Altogether, the study concludes that advanced data-driven approaches such as EDA, CLV, clustering, MBA, and SARIMA forecasting can significantly improve business understanding, enhance customer targeting strategies, optimize product placements, and support robust forecasting for better decision-making in supermarket sales.

## REFERENCES

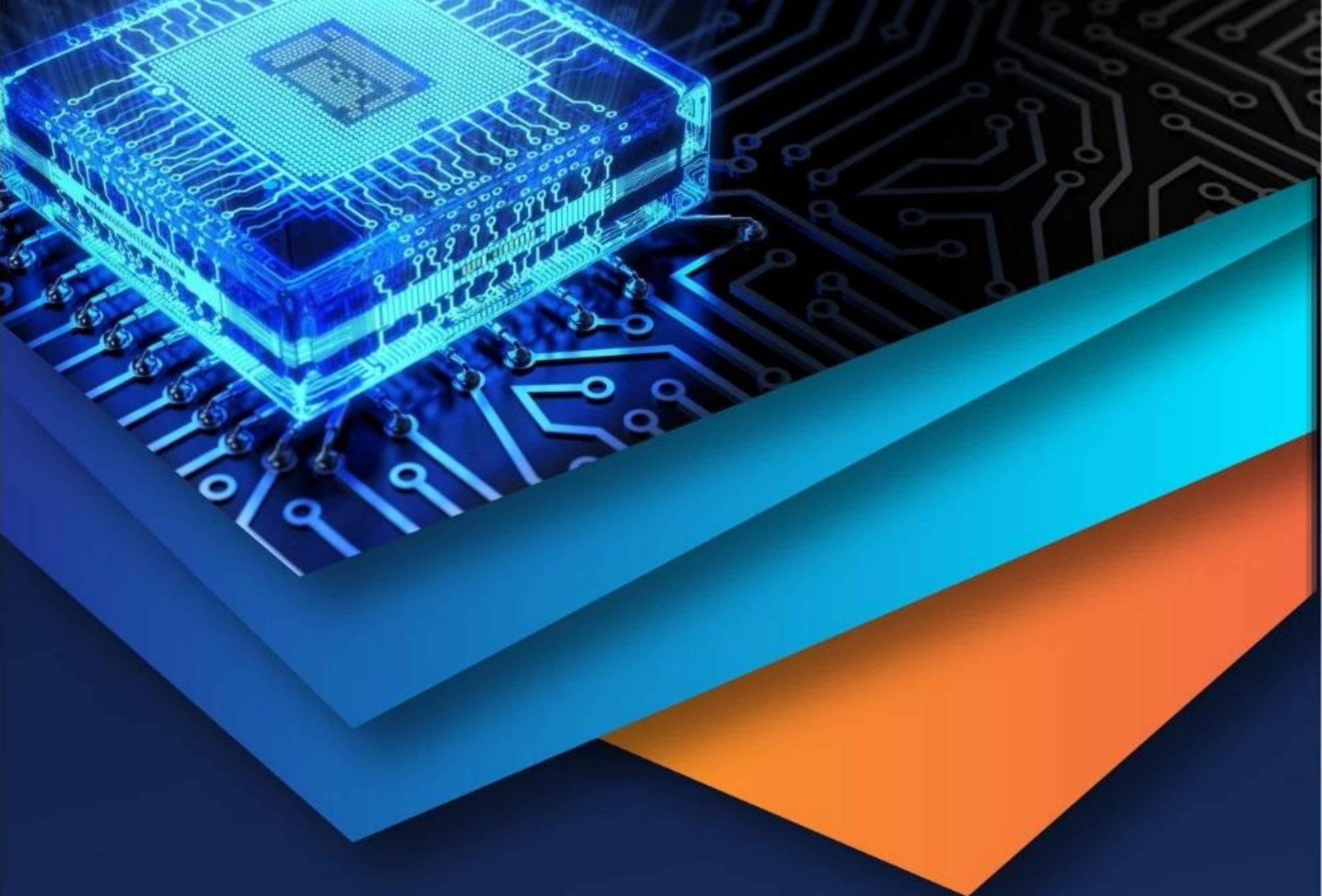
[1] Erni Widiastuti, Jani Kusanti, Herwin Sulistyowati, Destian Firnanda, Septian Muhamad Riski, Rizki Adhi Pratama: "Exploratory Data Analysis to Improve the Performance of Convolutional Neural Networks in Online Sales Product Image Prediction", IEEE, 2024.

[2] KangHui Ying, WenYu Hu, Jin Bo Chen, Guo Nong Li: "Research on Instance-Level Data Cleaning Technology", IEEE, 2021.

[3] M. V. Jerabandi, Keerti Rayaraddi, Achala Shirol, Sneha Mugalkod, Sushmita Tirlapur: "Analysis and Visualization of Supermarket Data using Data Science Techniques", International Research Journal in Global Engineering and Sciences (IRJGES), May 2021.

[4] G. S. Ramesh, T. V. Rajini Kanth, D. Vasumathi: "Analysis of Location Based Sales Data using Machine Learning Algorithms", International Journal on Emerging Technologies, February 2020.

[5] Yutao Li: "Application of Tableau in Visual Analysis Data of a US Supermarket Sales", IEEE, 2022.
[6] Zhao Mei, Li Mingjie: "Research on Supermarket Marketing Data Analysis Based on Business Intelligence", IEEE, 2023.
[7] Seyed Mojtaba Miri, Zohreh Dehdashti Shahrokh: "A Short Introduction to Comparative Research", Conference Paper, May 2019.
[8] Pranavi Satheesan, Prasanna S. Haddela, Jesuthasan Alosius: "Product Recommendation System for Supermarket",December 2020.
[9] Shuming Wang, Phisanu Chiawkhun: "Using data visualization for supermarket retail analysis".
[10] Aman Banduni, Ilavendhan A.: "Customer Segmentation Using Machine Learning", IJCRT, 2020.
[11] Marcin Majka: "Implementing Customer Segmentation in Marketing", 2024.
[12] Bharghav Madhiraju, Sukesh Reddy, Dr. G Sasikala:" CUSTOMER SEGMENTATION USING RFM ANALYSIS", EPRA International Journal of Economic and Business Review-Peer Reviewed Journal , July 2024.
[13] Mussadiq Abdul Rahim, Muhammad Mushafiq, Salabat Khan, Zulfiqar Ali Arain:"RFM-based repurchase behavior for customer classification and segmentation", July 2021.
[14] Pritika Talwar, Shubham, Komalpreet Kaur: "EXPLORING CLUSTERING TECHNIQUES IN MACHINE LEARNING", IJCRT, March 2024.
[15] Kamalpreet Bindra, Anuranjan Mishra: "A Detailed Study of Clustering Algorithms", IEEE, 2018.
[16] Mohiuuddin Ahmed, Raihan Seraj, Syed Mohammed Shamsul Islam: "The k-means Algorithm: A Comprehensive Survey and Performance Evaluation", MDPI, May 2020.
[17] Ms. Sarika Rathi, Prof. Vijay Karwande: "Review Paper on Customer Segmentation Approach Using RFM and K-Means Clustering Technique", International Journal of Creative Research Thoughts (IJCRT), December 2022.
[18] Km Vandna, Mr.Pawan Yadav, Mr.Vinod Kumar: "Elbow Method for Optimal Customer Segmentation Using K-Means Clustering", International Journal of Scientific Research and Engineering Development, May-June 2024.
[19] Darshan Anil Jethwa, Siya Milind Khamkar, Anish Anand Pachchhapur, Snehal Kulkarni. "Customer Segmentation Analysis using K-means Algorithm with Elbow Method and Dendrogram." IEEE, 2024.
[20] Fitri Marisa, Arie Restu Wardhani, Wiwin Purnomowati, Anik Vega Vitianingsih, Anastasia L Maukar, Erri Wahyu Puspitarini: " POTENTIAL CUSTOMER ANALYSIS USING K-MEANS WITH ELBOW METHOD",September 2023.
[21] Ms. Ramamani Venkatakrishna, Mr. Pradeempta Mishra, Ms. Sneha P Tiwari. "Customer Lifetime Value Prediction and Segmentation using Machine Learning." International Journal of Research in Engineering and Science (IJRES), August 2021.
[22] Mitra Bokaei Hosseni, Mohammad Jafar Tarokh. "Customer Segmentation Using CLV Elements." Journal of Service Science and Management, 2011.
[23] Venkata Harini, G. Venu, G. Vijay Kiran Reddy, et al. "Market Basket Analysis." International Journal for Research in Applied Science & Engineering Technology (IJRASET), May 2024.
[24] Edwin Omol, Dorcas Onyango, Lucy Mburu, Paul Abuonji: " Apriori Algorithm and Market Basket Analysis to Uncover Consumer Buying Patterns: Case of a Kenyan Supermarket",
[25] Himanshu Singh, Nikhil Shelke, Aniket Bavaskar, Shradha Nikam, Prof. Pradip Shewale, Prof. Deepa Mahajan: "Study on Market Basket Analysis with Apriori Algorithm Approach", International Research Journal of Engineering and Technology (IRJET), May 2021.
[26] Pooja Ghude, Mansi Padekar, Pradip Alam and Dr. Savita Sangam: "SALES FORECASTING PREDICTION USING MACHINE LEARNING", JETIR, June 2024.
[27] Yasaman Ensafi, Saman Hassanzadeh Amin, Guoqing Zhang, Bharat Shah: "Time-series forecasting of seasonal items sales using machine learning – A comparative analysis", International Journal of Information Management Data Insights, April 2022.
[28] Zhenyu Liu, Zhengtong Zhu, Jing Gao, Cheng Xu: "Forecast Methods for Time Series Data: A Survey." IEEE , June 2021.
[29] Peng Chen, Aichen Niu, Duanyang Liu, Wei Jiang, Bin Ma: "Time Series Forecasting of Temperatures using SARIMA: An Example from Nanjing." IOP Conference Series: Materials Science and Engineering, August 2018.
[30] P. Kabbilawsh, D. Sathish Kumar, N. R. Chithra: "Forecasting long-term monthly precipitation using SARIMA models." Journal of Earth System Science, Indian Academy of Sciences, March 2022.
[31] Sardar Usman, M. Usman Ashraf, Asad Hayat: . "Predictive Analysis of Retail Sales Forecasting using Machine Learning Techniques." Lahore Garrison University Research Journal of Computer Science and Information Technology, February 2023.
[32] Malde Ritik Vimal, Shaikh Mohammad Bilal Naseem: "Time Series Analysis: Forecasting with SARIMAX Model and Stationarity Concept." Journal of Emerging Technologies and Innovative Research (JETIR), December 2020.
[33] Georgia A Papacharalampous, Hristos Tyralis: "One-step ahead forecasting of annual precipitation and temperature using univariate time series methods (solicited)." European Geosciences Union General Assembly, April 2018.
[34] Gajendra Thakur, Anup Masurkar, Deepa Padwal: "A Review Of Superstore Sales And Customer Feedback Analysis Using Data And Information Visualization.", International Journal of Creative Research Thoughts (IJCRT), October 2024.




10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429


# INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY