☒Report Overview :

- This presents an analysis of e-commerce sales data, focusing on key areas such as revenue, customer behavior, and sales trends over time.
- The goal of this analysis is to derive insights that can help improve sales strategies, customer targeting, and business operations.

☒Key Insights:

#1# {from cleaning data} :

- Missing Data and Data Issues:
  - We found missing values in (Description, CustomerID)
  - We found negative values in Quantity , Zero & negative values in UnitPrice
  - 25% NaN in CustomerID, From our initial observations we got that we have CustomerID values with NaN entries, which could be valid as they may represent guest users who are not logged in
  - To handle this, we replace the NaN values with a negative number, such as -1, to indicate their presence, as -1 is not a logical value for CustomerID.
  - The InvoiceNo should be unique, but repeated values are present. This could be due to a single order containing multiple types of products.
  - Convert the InvoiceDate column to the DateTime type for more efficient use in our analysis
- Data Transformation (engineering):
  - We calculated TotalPrice from Quantity and UnitPrice, which will be helpful for calculating revenue and grouping in our analysis

#2# {Sales Overview} :

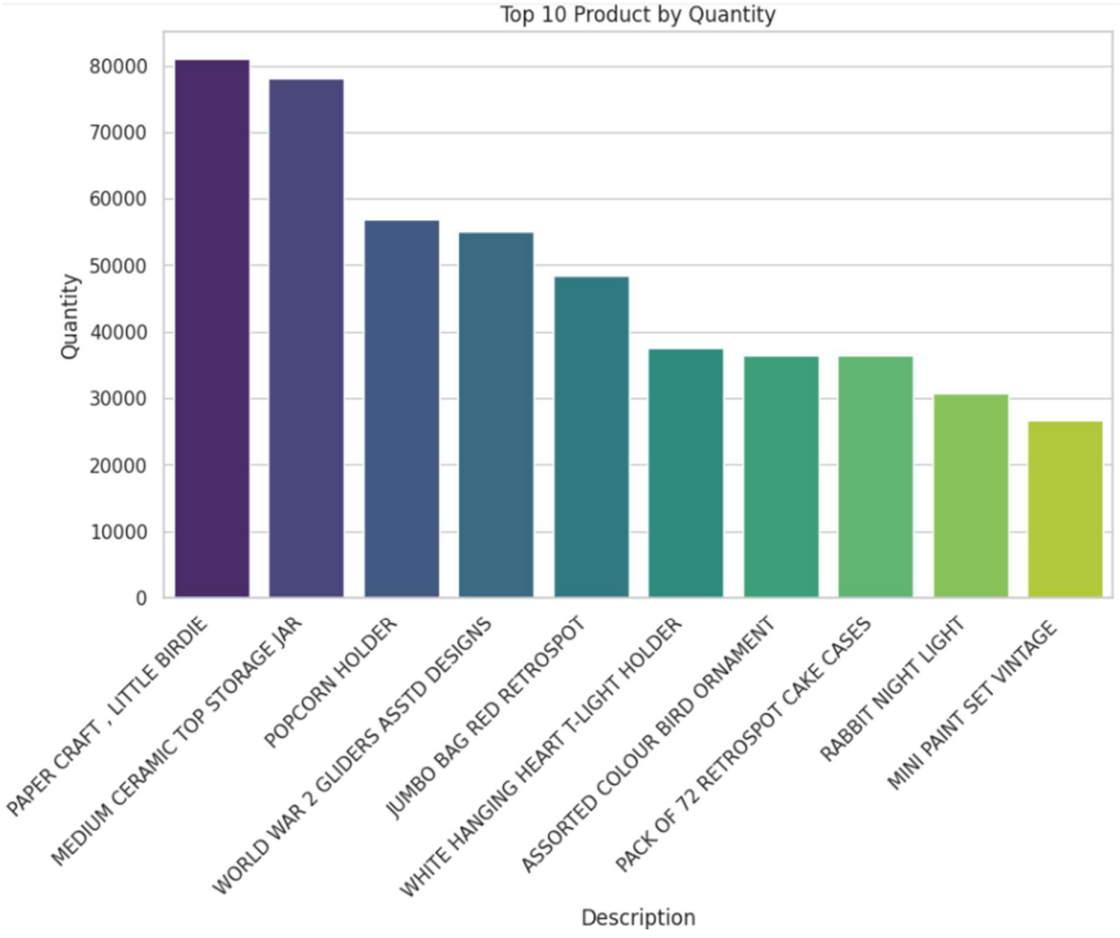
- Overall Revenue :
  - Total revenue :

    The total revenue generated from all transactions is ---> 10,666,684.544 $
  - Total number of transactions: ---> 19960
  - Number of Unique Customer ---> 4339

#3# Top selling products:

- List of top 10 products by quantity :
  1- PAPER CRAFT , LITTLE BIRDIE
  2- MEDIUM CERAMIC TOP STORAGE
  3- POPCORN HOLDER
  4- WORLD WAR 2 GLIDERS ASSTD DESIGNS
  5- JUMBO BAG RED RETROSPOT
  6- WHITE HANGING HEART T-LIGHT HOLDER
  7- ASSORTED COLOUR BIRD ORNAMENT
  8- PACK OF 72 RETROSPOT CAKE CASES
  9- RABBIT NIGHT LIGHT
  10-MINI PAINT SET VINTAGE

Note: for a more precise numbers we used ----> stock codes for grouping



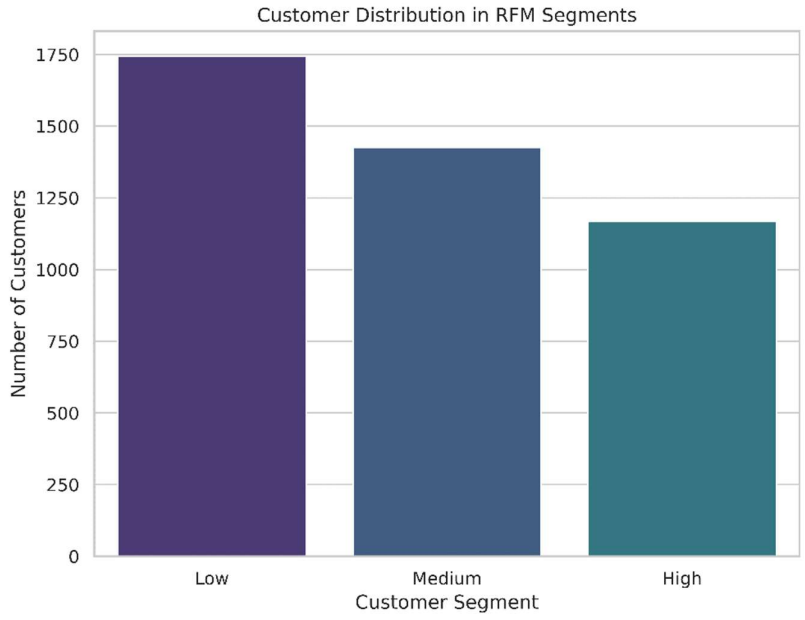
#4# Customer Segmentation (RFM Analysis):

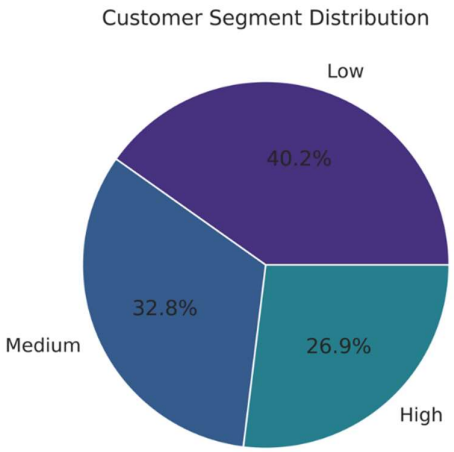
#Methodology

- Recency Calculation:
  {Lower recency receives a higher score}
  - We calculated it as the number of days since the last purchase for each customer. The current date was set to the latest InvoiceDate in the dataset plus one day.
- Frequency Calculation:
  {Higher frequency receives a higher score}
  - Frequency was calculated by counting the number of unique invoices (purchases) for each customer.
- Monetary Calculation:
  {Higher spending receives a higher score}
  - Monetary value was calculated by summing the total amount spent by each customer.
- RFM Scoring:
  - Calculated by summing the individual scores for Recency, Frequency, and Monetary.
- We segmented customers into three categories based on their RFM_Score
  - Low (Bottom 33% of RFM_Scores)
  - Medium (Middle 33% of RFM_Scores)
  - High (Top 33% of RFM_Scores)

#Key Findings:

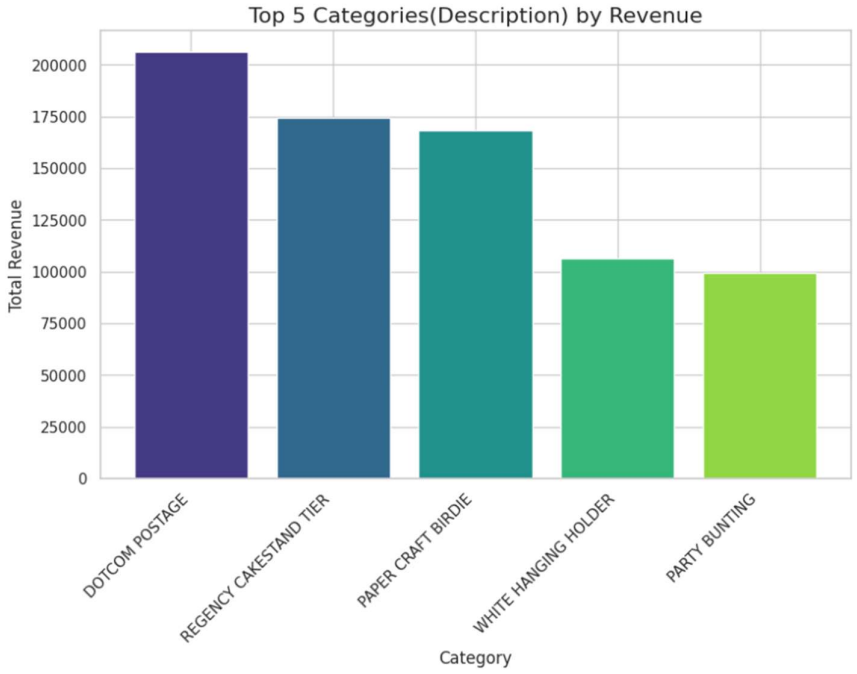
- High-Value (26.9%): Recent, frequent buyers who spend the most.
- Medium-Value (32.85%): Moderately engaged customers with average spending.
- Low-Value (40.2%): low spending and no recent activity.





#5# Product Category Analysis:

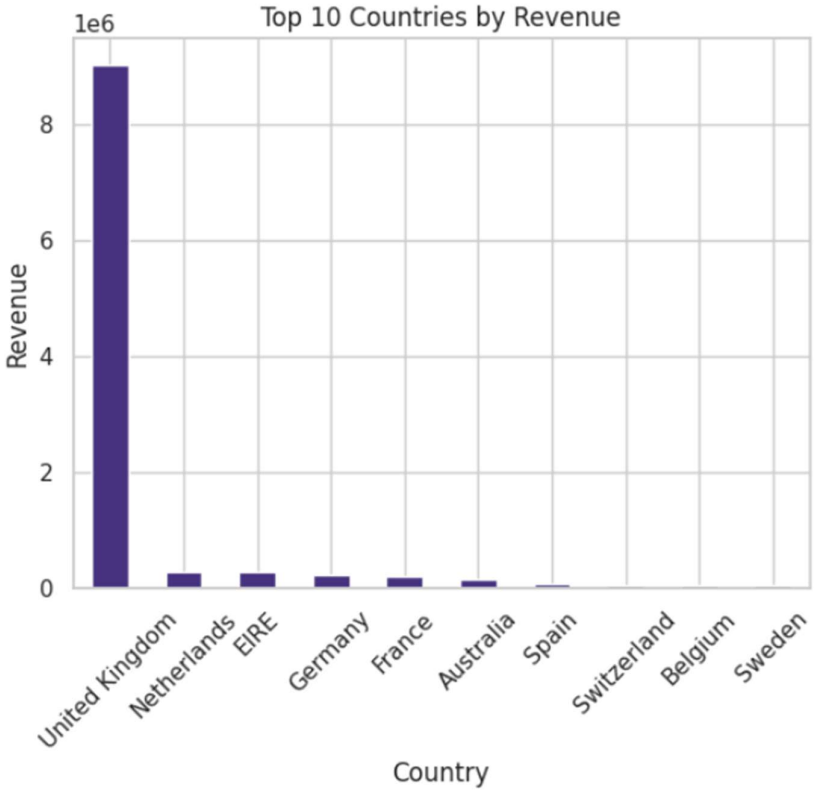
- List of 5 product categories by revenue:



--> Contribution of top 5 categories to overall revenue: 7.08%

#6# Geographical Insights:

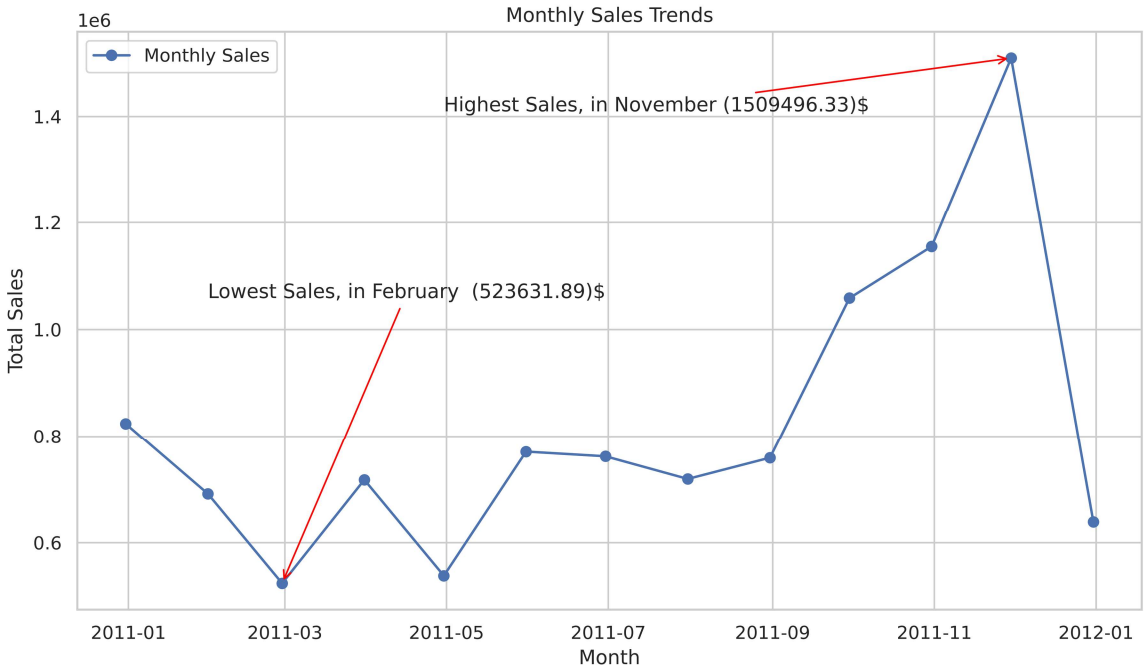
- Top 10 countries by revenue:



--> Contribution of Sales from Top 3 Countries to overall revenue: 89.94%

#7# Time Series Analysis Insights:

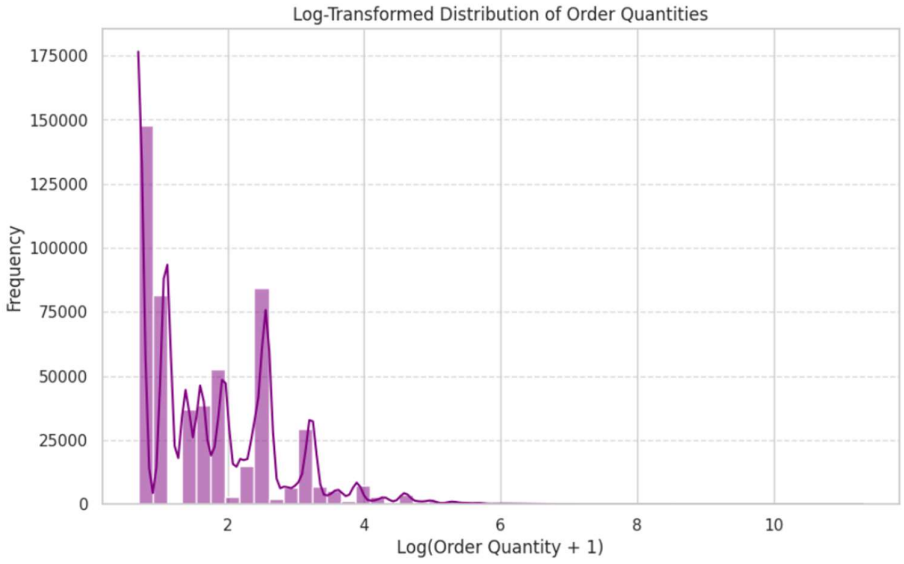
- Sales Trends Over Time :
  - Lowest Month with sales ----> February
  - Highest Month with sales ----> November
- This might happened due to large discounts at the end of the year.



#8# Customer Behavior Analysis:

- Order Quantity Distribution
  - The plot now shows a heavy skew toward smaller order sizes, with most quantities clustering at the lower end of the positive range.
  - A long tail indicates the presence of larger orders, though these are relatively infrequent compared to smaller ones.

#The log transformation normalizes the data, revealing a clearer peak in the most common order sizes. -The distribution becomes easier to interpret, with gradual declines in frequency for larger orders.
The plot now shows a heavy skew toward smaller order sizes, with most quantities clustering at the lower end of the positive range.



Concentration of Data

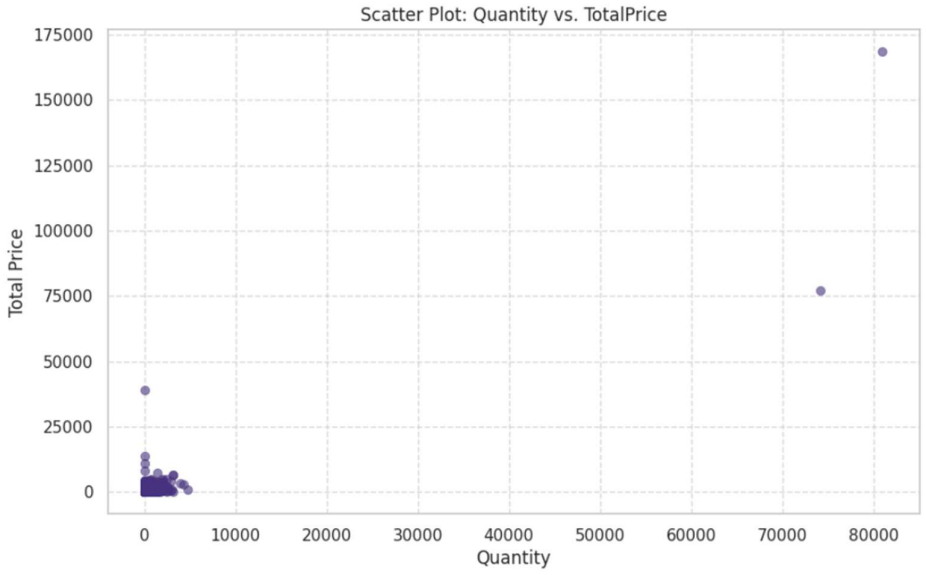
- The majority of the data points are concentrated in the lower ranges of Order Quantity (near 0–10k) and Total Price (below 150k).
- This indicates that most orders have relatively low quantities and generate smaller total sales.

Outliers:

- A few data points stand out as extreme values, with significantly high order quantities (60k–80k) or total prices (over 150k).
- These could represent bulk orders or high-value transactions

Pattern Observation:

- There is no clear linear relationship between Order Quantity and Total Price. This means that larger quantities do not always result in proportionately higher revenues.

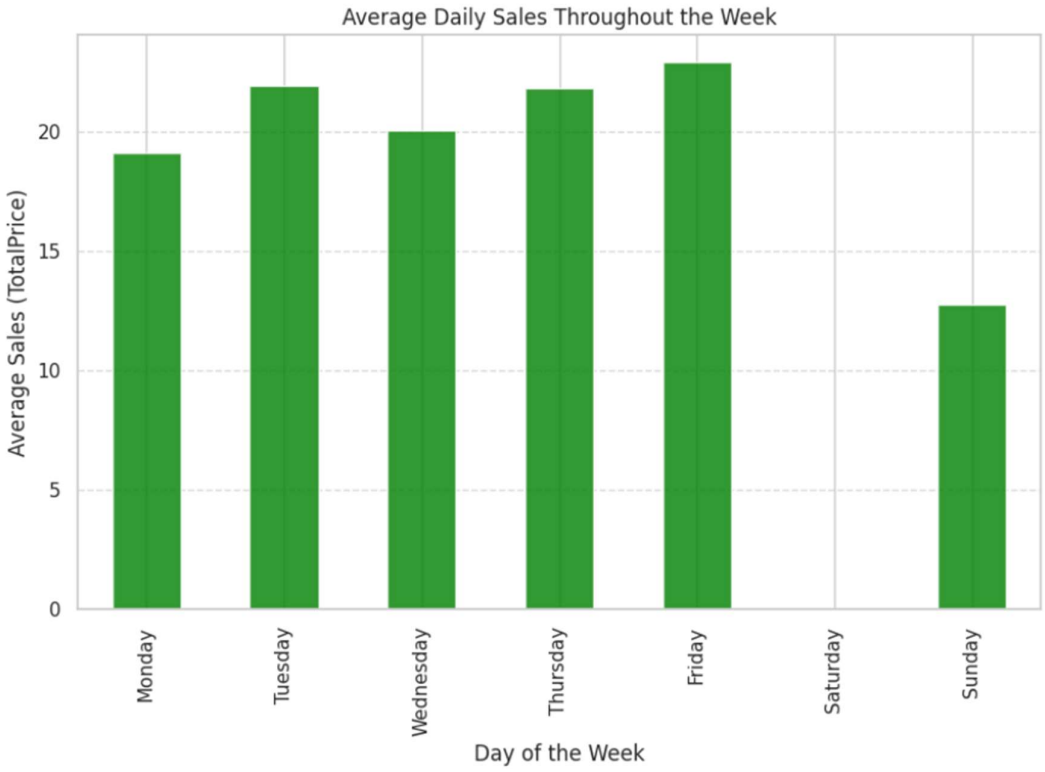


Sales Trends by Day:

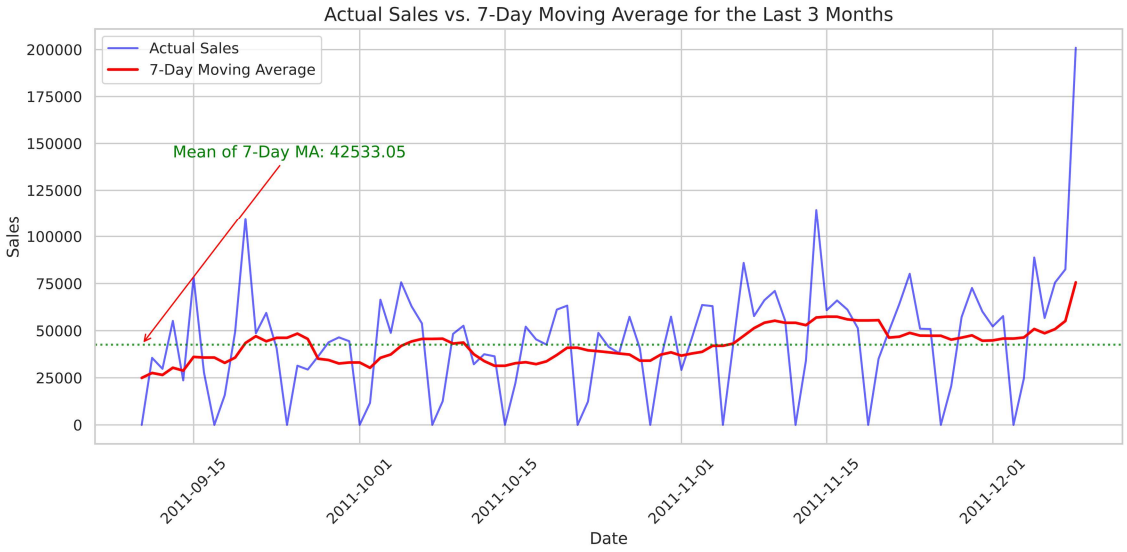
- Tuesday, Wednesday, Thursday, and Friday exhibit the highest average sales, all being relatively similar and slightly higher than other days.
- Monday has moderate sales, suggesting a slower start to the week.
- Sunday has the lowest average sales, indicating that customers are less active on this day.
- No sales on Saturday that's because it's a holiday day , so customers prefer to do offline activates more than active online.

Business Implications:

- The mid-week (Tuesday to Friday) is the most profitable period. Marketing campaigns or sales promotions during these days could further boost sales.



#9# Moving Average Forecasting:



The sharp peaks in the actual sales (blue line) suggest a Weekly seasonality.

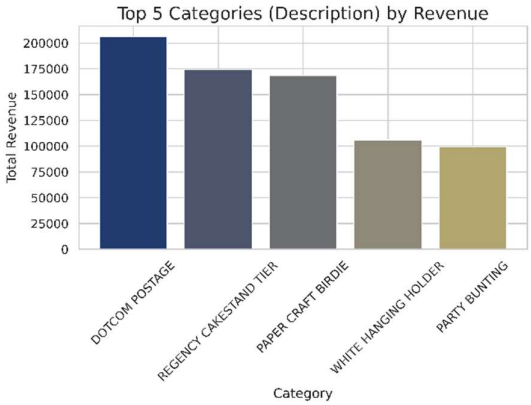
7-day moving average (red line) provides a smoother trend of sales over time.

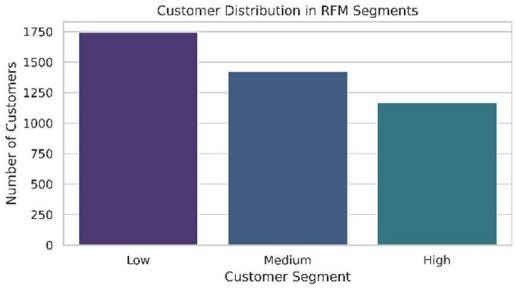
Mean value of 42533.05 indicates the average performance over the last 3 months.

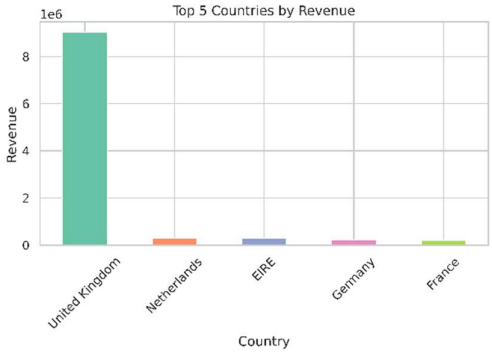
Fluctuations: there are significant peaks in the actual sales compared to Moving average, that may indicate variations or irregular spikes in sales.

The 7-day moving average can be a reliable metric for observing steady trends or patterns.

Summary Dashboard







☒Recommendations:

#For Customer segments:

- For High segment :
  - Focus on loyalty programs
  - Exclusive offers for them
- For Medium segment :
  - Promotions to increase spend/frequency
- For Low segment :
  - More campaigns/Discount to motivate purchases

#For Product strategy:

- Stock more of the Top selling products like (DOTCOM POSTAGE , PAPER CRAFT BIRDIE, …)
- Investigate the reasons behind the low sales of certain products like (PINK CRYSTAL CHARM, VINTAGE BLUE REEL, …) .

#For Geographical strategy:

- Prioritize marketing in top performing countries
  Like (UK, Netherlands, Eire,….)
- Look into the hidden potential of products with low sales in different areas.