E-Commerce\_data Assignment

The analysis aims to extract key insights from an e-commerce dataset to understand customer behavior, revenue trends, and the most popular products. The focus is on identifying customer retention trends, monthly revenue growth, and sales performance across product categories.

2025

**RAKSHITH S R**

**1/30/2025**

## Overview :

This report presents a comprehensive analysis of the e-commerce dataset, aiming to uncover key insights into customer retention, revenue patterns, and product trends. By examining monthly customer behavior, revenue fluctuations, and product performance, the goal is to provide actionable insights that can guide business strategies in improving customer engagement, maximizing revenue, and optimizing product offerings.

The analysis was carried out in three primary areas:

* **Customer Retention Analysis:**

We explored the monthly customer base to assess the number of unique customers and track trends in new versus returning customers. Understanding this dynamic provides valuable insights into how well the platform is retaining customers and attracting new ones.

* **Revenue Analysis:**
  + Monthly revenue trends were evaluated to identify patterns, with a particular focus on the top-performing months and the average revenue per customer. This section aims to provide an understanding of revenue growth, peak performance periods, and potential factors driving financial success.
* **Product Trend Analysis:**
  + By analyzing product sales across categories and identifying the top-performing products, this section highlights the most popular and profitable items in the marketplace. Seasonal trends in product sales were also examined to uncover opportunities for targeted marketing and inventory management

### Data Preprocessing and Cleaning :

#### ****Loading Data:****

* The dataset was initially loaded into a Pandas DataFrame using the pd.read\_csv() function, which allows easy manipulation of the data. The dataset, named ecommerce\_data.csv, consists of transaction records with columns such as CustomerID, OrderID, OrderDate, ProductID, Category, Quantity, Price, and TotalAmount.
* By reading the data into a DataFrame, we can leverage various functions in Pandas to clean, process, and analyze the data effectively.

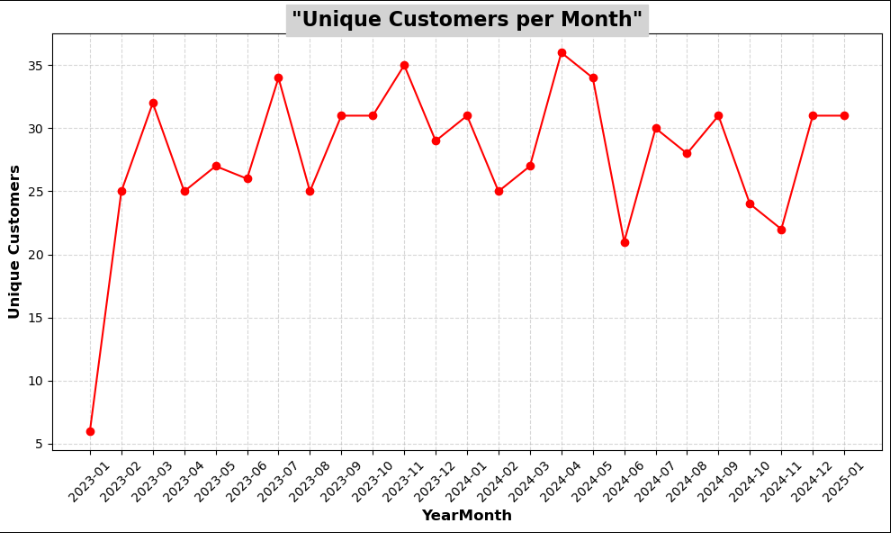
### Cleaning Data:

* **Data Completeness**: There are no missing values across any columns, ensuring that the dataset is ready for analysis without the need for data imputation or removal.
* **No Duplicates**: No duplicate rows were detected, which is essential to maintain the integrity of key metrics such as customer retention and revenue.
* **Correct Date Format**: The OrderDate column was successfully converted to the datetime format, enabling accurate time-based analysis. Invalid entries were handled gracefully without any errors.
* **New Time-Based Column**: The YearMonth column was created, allowing the data to be grouped and analyzed at the monthly level, which is key for understanding trends in sales and customer behavior.

### Customer Retention Analysis :

### Objective:

* Calculate the number of unique customers each month.
* Identify the percentage of returning customers vs new customers each month.
* Plot a trend of customer retention over time.



Initial Findings:

The number of unique customers fluctuates month to month.

* January 2023 had the **lowest number** of unique customers with only 6.
* Some months, like **March 2023 (32)**, **November 2023 (35)**, and **April 2024 (36)**, showed higher engagement.
* **Seasonal Variations**: Months like **November 2023**, **December 2023**, and **April 2024** saw a higher number of customers, which could be due to seasonal factors such as holiday shopping or special promotions.
* **Consistent Months**: **January 2024** and **May 2024** had stable engagement, with **31 unique customers** each month, suggesting sustained customer activity during these periods.

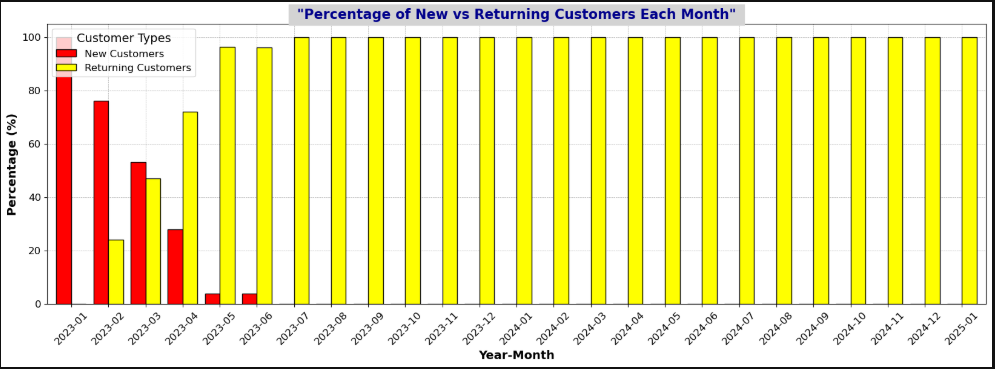
### Visual Representation:

* A **line plot** with red color and markers was used to plot the unique customers per month, making it easier to identify trends.
* The x-axis represents **months**, and the y-axis represents the **number of unique customers**.

### Insights:

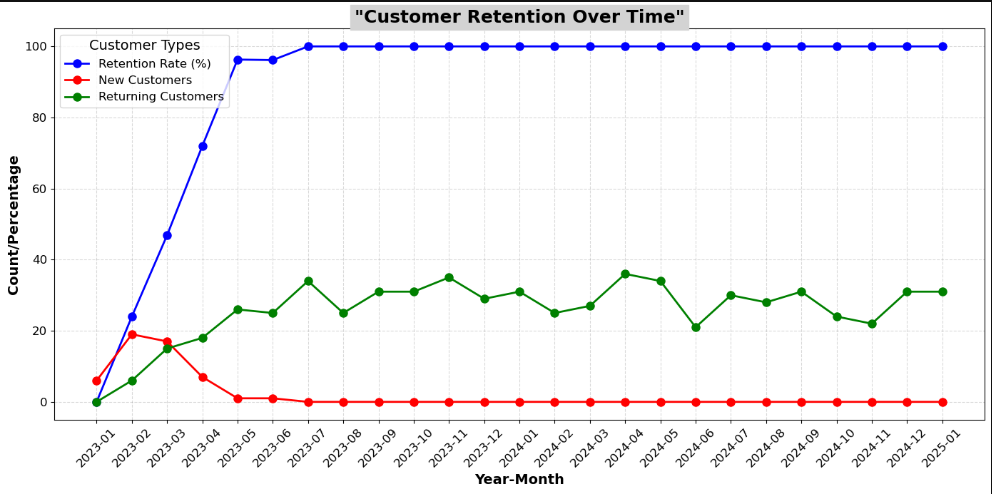
* **Fluctuating Trends**: The trend reflects **peaks and valleys** in customer activity, indicating **seasonal patterns** or periods of low engagement.
* **Opportunities for Engagement**: Low months (e.g., **June 2024** with 21 unique customers) may indicate opportunities to implement **targeted marketing** strategies to boost retention

### Percentage of New vs. Returning Customers:



* **New Customers**: The percentage of new customers started at 100% in **January 2023** and gradually decreased to 0% by **July 2023** and remained at 0% thereafter. This suggests a rapid conversion of first-time buyers into repeat customers.
* **Returning Customers**: Starting from **February 2023**, the percentage of returning customers gradually increased, with **100% returning customers** from **July 2023** onward. This shows strong **customer loyalty** and **high retention rates** after initial purchases.

### ****Customer Retention****



1. **Retention Rate Over Time**:

* The **retention rate** of customers started at **0%** in **January 2023**, with no returning customers, indicating the beginning of customer acquisition.
* From **February 2023** onward, the retention rate steadily increased, reflecting growing **customer loyalty**. By **May 2023**, the retention rate reached a substantial **96.3%**.
* From **July 2023** onward, the retention rate remained **100%**, signifying that all customers from those months were returning, and no new customers were acquired in later months.

1. **Shift to Returning Customers**:

* In the early months of 2023, **new customers** made up a significant portion of the customer base. However, by mid-year, **returning customers** began to dominate.
* By **July 2023**, there were no new customers, and the **returning customer base reached 100%**. This trend continued through to **January 2025**, indicating a shift towards customer retention and reduced focus on acquiring new customers after the initial surge.

1. **Growth in Returning Customers**:

* **March to May 2023** saw a steady increase in the number of returning customers, with **March** having **46.87%** returning customers, which surged to **100%** in **July 2023**.
* **December 2023** had **100% returning customers**, showing the platform’s ability to retain customers consistently as the year progressed.

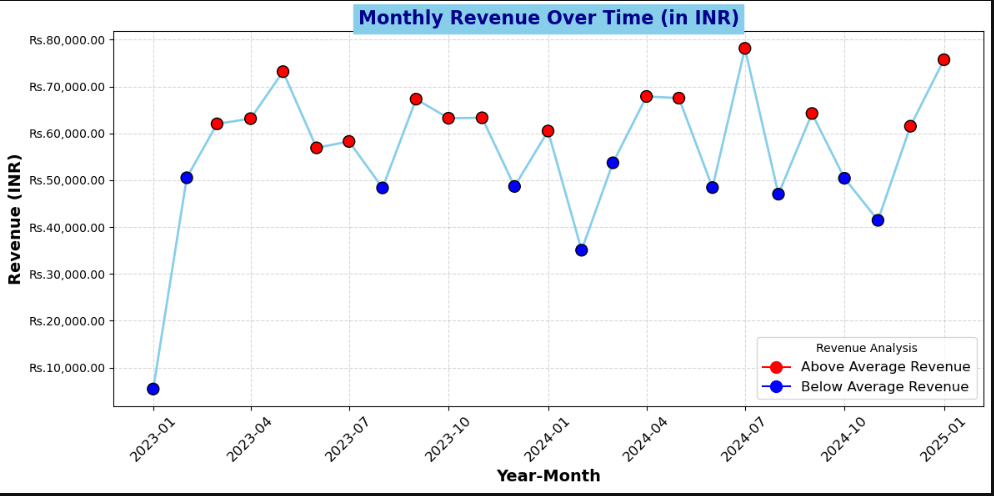
1. **Key Insights from Retention Rates**:

* **Customer Loyalty**: The consistently high **retention rate** from **July 2023** to **January 2025** indicates the effectiveness of **loyalty programs** or repeat customer incentives.
* **Sustained Engagement**: Despite a decrease in **new customers** post-mid 2023, the **stable customer base** remained, highlighting the company's ability to **sustain engagement** and foster loyalty.

### Revenue Analysis :

### Objective:

* Calculate monthly revenue and plot a revenue trend.
* Identify the top 5 months with the highest revenue.
* Compute the average revenue per customer for each month



### ****Monthly Revenue Trends****

1. **Revenue Fluctuations**:

* The **monthly revenue** has shown significant variation over time, with some months seeing substantial increases in revenue, while others experienced declines.
* The highest revenue month in the dataset was **July 2024** with a revenue of **Rs.78,159.40**, which stands out as a peak in the data.
* The lowest revenue month was **January 2023** with **Rs.5,422.73**, indicating the early stages of the business where revenue was low.

1. **Revenue Growth**:

* From **January 2023** to **May 2023**, there was a clear upward trend in revenue, with revenue peaking at **Rs.73,183.20** in **May 2023**.
* **July 2023** saw a noticeable increase in revenue to **Rs.58,277.63**, indicating possibly seasonal or promotional factors influencing customer purchases.
* Throughout **2024**, the company maintained relatively **stable revenue**, consistently hovering around the **Rs.60,000** mark, with fluctuations largely attributed to external factors such as seasonality.

1. **Revenue Performance Above/Below Average**:

* The **average monthly revenue** is calculated, and months with **above-average** revenue are shown in **red**, while those with **below-average** revenue are shown in **blue**.
* The company achieved **above-average** revenue in months such as **March 2023, May 2023, July 2024**, and **January 2025**, demonstrating the company's ability to generate strong revenue during these periods.
* The months of **February 2024**, **June 2024**, and **November 2024** were notably below the average, with revenues in these months underperforming.

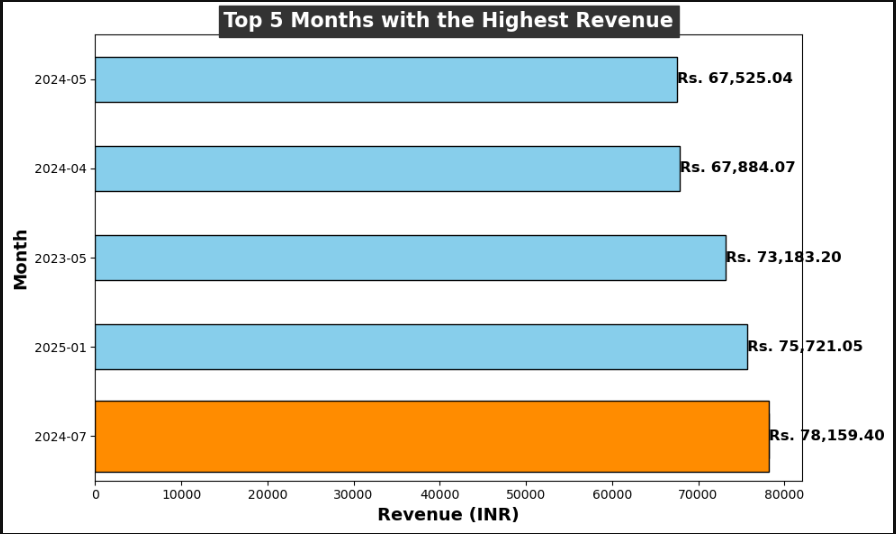
1. **Visual Representation**:

* A line plot has been used to show **monthly revenue** over time, where **colored markers** represent whether the revenue was above or below the threshold.
* The **x-axis** represents each **Month-Year**, and the **y-axis** reflects the **monthly revenue** in INR (formatted as **Rs.**).
* The revenue for each month is marked with either **red (above average)** or **blue (below average)**, making it easy to spot high and low-performing months.

1. **Insights from Revenue Trends**:

* **Seasonal Variations**: There are signs of seasonal sales patterns with some months (such as **July 2024**) exhibiting higher-than-usual revenues.
* **Sustained Revenue**: The overall **stable revenue performance** in 2024 indicates steady demand for products despite occasional fluctuations.
* **Revenue Peaks**: Special events, promotions, or external market factors likely contributed to the **higher revenue months** in **2024** and **2025**.

### ****Top 5 Months with Highest Revenue****



1. **Highest Revenue Month**:

* The month of **July 2024** saw the highest revenue at **Rs. 78,159.40**, indicating a **significant peak** in performance, possibly driven by special promotions, seasonal demand, or a successful marketing campaign.

1. **Second and Third Highest**:

* Following **July 2024**, **January 2025** and **May 2023** also saw substantial revenue of **Rs. 75,721.05** and **Rs. 73,183.20**, respectively. This demonstrates that the company was able to sustain strong sales beyond the typical year-end peak, with May 2023 likely benefiting from early-year sales trends.

1. **Months with Consistently High Revenue**:

* Both **April 2024** and **May 2024** rounded out the top 5 months, with revenues of **Rs. 67,884.07** and **Rs. 67,525.04**, respectively. This suggests **consistency** in performance during mid-year months, possibly due to product releases or promotions.

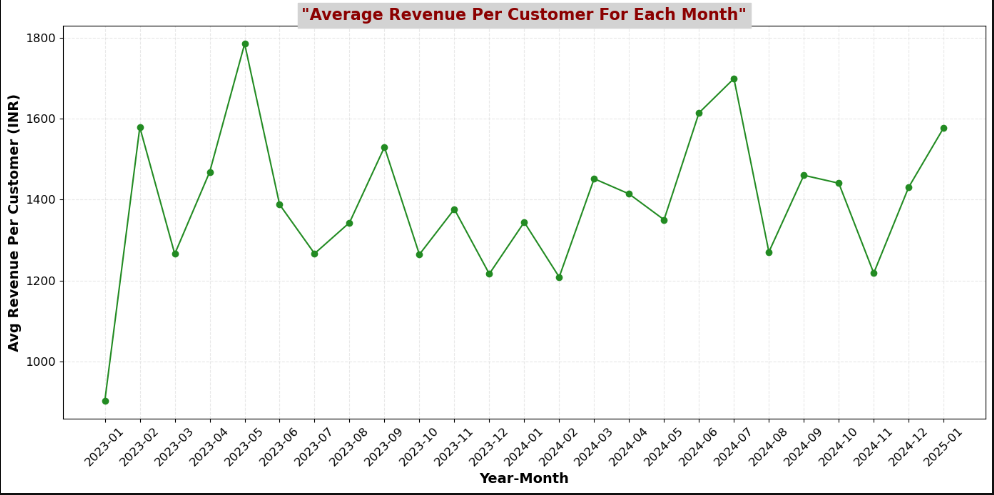
1. **Visual Representation**:

* A **horizontal bar chart** has been used to visualize these top 5 months, making it easier to compare them. The **dark orange bar** highlights **July 2024** as the month with the highest revenue, making it visually distinct.
* Each bar shows the revenue value, ensuring clarity and easy identification of the top months.

1. **Revenue Distribution**:

* The revenue for these months varies significantly, with **July 2024** standing out by **Rs. 2,000 higher** than **January 2025**, and the remaining months having revenues just under **Rs. 70,000**.
* The **mid-year months** such as **April** and **May** indicate **steady revenue**, contributing significantly to the overall financial health of the business.

### ****Average Revenue Per Customer:****



1. **Overall Trend**:

* The **Average Revenue Per Customer** fluctuates each month, reflecting seasonal variations, promotions, and customer purchasing behavior.
* Generally, the **average revenue** per customer increased in **2023-2024**, reaching its peak in **July 2024** at **Rs. 1,699.12**, followed by **January 2025** with **Rs. 1,577.52**.

1. **Peak in Revenue**:

* The **highest monthly average revenue per customer** of **Rs. 1,699.12** occurred in **July 2024**, likely indicating a **high demand**, **successful promotions**, or **premium product purchases** that contributed to elevated spending.
* **January 2025** also saw a significant average of **Rs. 1,577.52**, possibly due to **post-holiday season purchases** or **New Year sales**, leading to customers making larger purchases.

1. **Seasonal Variations**:

* Months such as **May 2023 (Rs. 1,784.96)** and **June 2024 (Rs. 1,614.46)** also observed **higher-than-usual revenue per customer**, suggesting that certain periods during the year encourage higher-value purchases.
* Conversely, **January 2023** had a relatively lower **average revenue per customer** of **Rs. 903.79**, potentially reflecting lower sales or fewer premium transactions in the initial part of the year.

1. **Trends Over Time**:

* The **average revenue per customer** seems to fluctuate but gradually increased through the year, peaking around mid-year and continuing into the next year (with notable spikes in **May 2023**, **July 2024**, and **January 2025**).
* **2023-2024** showed an upward trend, with **several months of increasing average revenue**, indicating growing customer engagement or higher-value transactions.

1. **Visual Representation**:

* The **line graph** effectively visualizes the fluctuations in average revenue per customer, with a **forest green** color representing the data points. This makes it clear where the revenue peaks and drops each month.
* The **grid lines** and rotating x-axis labels make the graph more readable, allowing a clear interpretation of trends over time.

1. **Key Insights**:

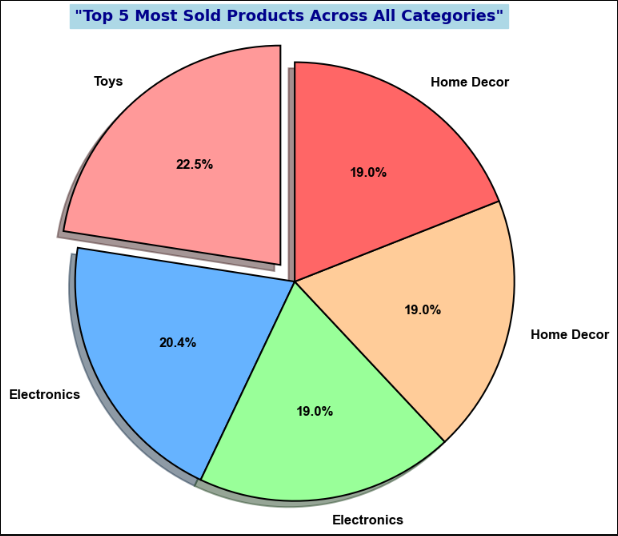
* **July 2024** marked the **highest average revenue per customer**, suggesting this could be a target period for further **promotions or sales strategies**.
* The **post-New Year period (January 2025)** saw higher-than-usual revenue, which may indicate increased customer spending following holiday discounts or deals.
* The months where **higher-than-average revenue per customer** occurred are important to study for potential strategies that could be replicated in other months.

### Product trend Analysis :

### Objective:

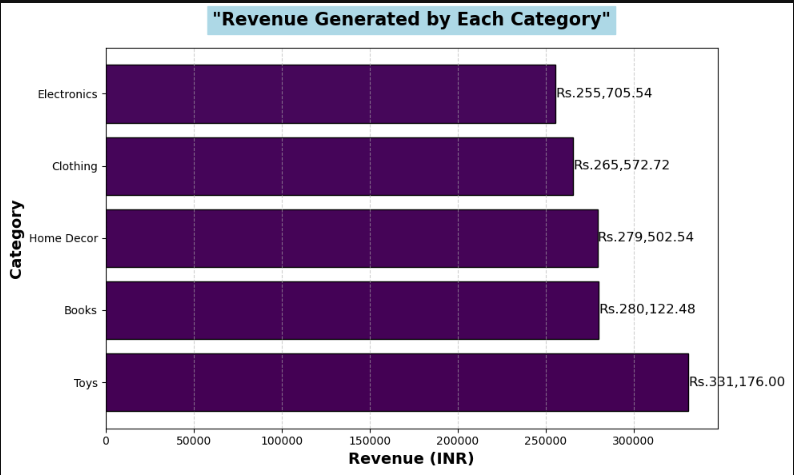
* Find the top 5 most sold products across all categories.
* Analyze which category generates the most revenue.
* Identify any seasonal trends in product sales.

### ****Top 5 Most Sold Products Across All Categories:****

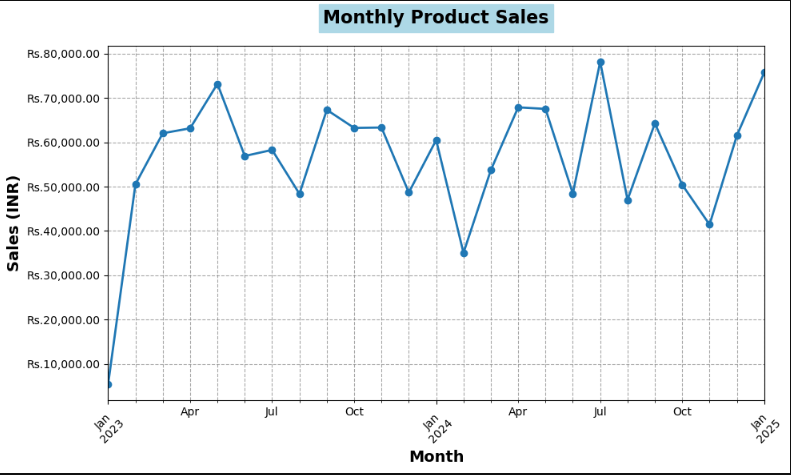


1. **Product Overview**:
   * The **top 5 most sold products** are primarily from the **Toys**, **Electronics**, and **Home Decor** categories, with a strong representation from **Electronics** and **Home Decor**.
   * The **ProductID 2261** (Toys category) stands out as the **most sold product**, with a total of **32 units sold**.
2. **Sales Breakdown**:
   * **Top 5 products** based on total quantity sold:
     + **ProductID 2261 (Toys)**: 32 units
     + **ProductID 2563 (Electronics)**: 29 units
     + **ProductID 2100 (Electronics)**: 27 units
     + **ProductID 2616 (Home Decor)**: 27 units
     + **ProductID 2527 (Home Decor)**: 27 units
   * These products are contributing a significant share to the overall sales volume across all categories.
3. **Category Distribution**:
   * The top 5 products are spread across **three categories**:
     + **Toys**: 1 product (ProductID 2261)
     + **Electronics**: 2 products (ProductID 2563, ProductID 2100)
     + **Home Decor**: 2 products (ProductID 2616, ProductID 2527)
   * This suggests that **Electronics** and **Home Decor** categories are strong performers in terms of unit sales.
4. **Pie Chart Visualization**:
   * A **pie chart** was used to visualize the percentage of total quantity sold for each product category. The chart clearly shows how each category contributes to the sales volume.
   * The **"Toys" category** has the **highest share**, followed closely by **"Electronics"** and **"Home Decor"**.
   * The chart visually emphasizes the **Toys** category by slightly "exploding" the slice for **ProductID 2261**, drawing attention to its prominence in total sales.
5. **Key Insights**:
   * **Toys** (ProductID 2261) emerges as a clear **best-seller**, making it an important product for the business.
   * **Electronics** and **Home Decor** categories both feature products with relatively high sales, reflecting strong consumer interest in these segments.
   * The **total quantity sold** helps identify products with high turnover, providing opportunities for targeted marketing, stocking, and promotions.
6. **Strategic Implications**:
   * Given the popularity of products in the **Toys**, **Electronics**, and **Home Decor** categories, focusing on these categories could lead to increased sales.
   * Consider **expanding the inventory** of top-selling products, especially in **Electronics** and **Home Decor**, to ensure stock availability during peak demand periods.

### ****Revenue Generated by Each Categories:****

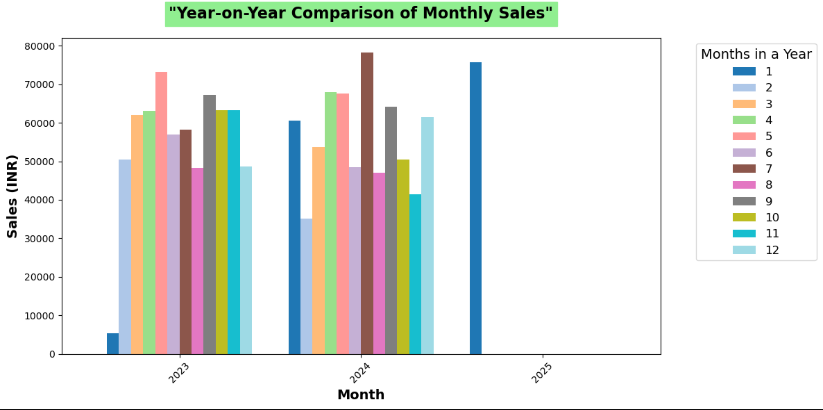


### ****Monthly Product Sales Insights****



1. **Overview of Monthly Sales**:
   * The **monthly sales data** for the products shows considerable variation in revenue across months, with peak sales months standing out.
   * In **2023**, the highest sales occurred in **May (Rs.73,183.20)**, while **January** had the lowest sales (**Rs.5,422.73**).
   * Moving into **2024**, the month of **July** saw the highest revenue of **Rs.78,159.40**. This month likely benefited from increased demand, contributing to higher overall sales.
   * The revenue for **January 2025** was also strong, reaching **Rs.75,721.05**.
2. **Monthly Sales Breakdown**:
   * **Key Months**:
     + **2024-07** (Rs.78,159.40) stands as the highest revenue month, followed by **2025-01** (Rs.75,721.05).
     + **2024-05** and **2024-04** maintained strong sales of **Rs.67,525.04** and **Rs.67,884.07**, respectively.
   * **Low Sales**:
     + The **lowest sales** were recorded in **2023-01** (Rs.5,422.73), a possible result of seasonal variations or other external factors.
3. **Graphical Representation**:
   * A **line chart** depicting the monthly sales trend over time reveals a **fluctuating sales pattern**, with spikes and drops in certain months. **2024-07** and **2025-01** stand out clearly.
   * The **monthly sales trend** highlights an overall positive upward movement in sales from **2023 to 2024**, reflecting a healthy revenue growth trajectory for the business

### ****Year-on-Year Comparison of Monthly Sales****



1. **Year-on-Year Sales Comparison**:
   * The **Year-on-Year comparison** was performed to analyze sales fluctuations across months between **2023 and 2024**.
   * For **most months**, **2024** recorded higher revenue compared to **2023**, with noticeable peaks in **July 2024**.
   * **2023** had lower sales in early months (January and February), which were significantly improved in the following year.
2. **Monthly Sales Variance**:
   * **July** stands out as a **strong performer in both years**, with **2024** slightly outperforming **2023**.
   * **2024-02** showed a dip in sales compared to the previous year (**2023-02**), but this is likely due to seasonal differences or other business dynamics.
3. **Bar Chart Visualization**:
   * The **Year-on-Year comparison** bar chart provides a clear view of how each month's revenue changed across the years. The chart utilizes a **distinct color palette** to differentiate months, with **2024** highlighted to compare its performance with **2023**.

### ****Conclusion and Key Insights****:

* **Peak Sales Period**: **2024-07** achieved the highest revenue of **Rs.78,159.40**, followed by **2025-01**, with **Rs.75,721.05**.
* **Sales Growth**: The year **2024** displayed **stronger sales performance** across most months, compared to **2023**, with certain months (like **July**) showing exceptional growth.