

# **Retail Sales Analysis Using Python**

## **Problem Statement**

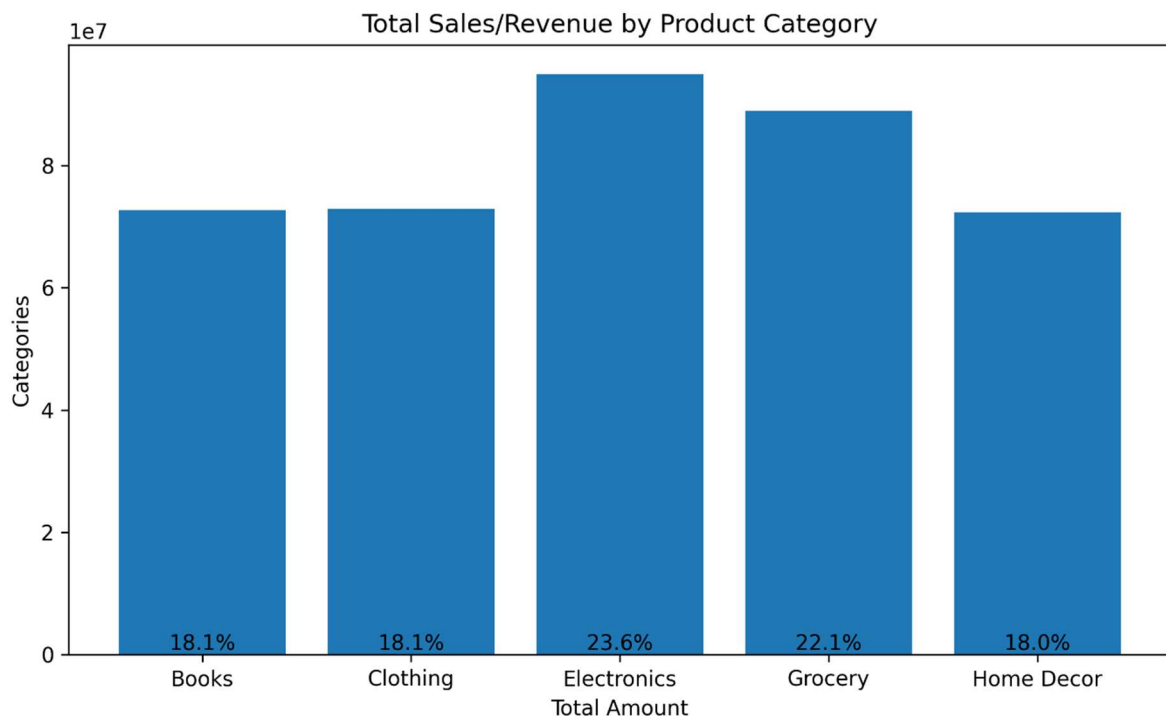
- 1) Our management team needs to focus marketing and supply chain efforts on the most profitable product categories and brands. We need clear insights into which categories and brands are driving revenue and which are underperforming.
- 2) We want to personalize marketing and improve ROI(Return on Investment). To do this, we need to segment our customers based on age, gender, income, and customer segment — and understand who brings the highest sales.
- 3) As we plan to expand operations, leadership wants insights into regional performance. We need to know which countries, states, and cities generate the most revenue and which areas need attention.
- 4) The marketing team wants to schedule campaigns more effectively. We need a clear understanding of how sales behave over time — by month and year — to identify peak seasons and low-performing periods.
- 5) Some products sell well but receive poor ratings. This may harm long-term customer trust. We want to cross-analyze sales volume with customer ratings to identify risky or high-potential products.
- 6) Customer feedback has highlighted concerns about delivery speed and reliability. We need to analyze how different shipping methods affect order status and customer satisfaction.
- 7) We want to reduce cart abandonment and improve checkout flow. To do this, we need to identify the most preferred payment methods across different customer groups and regions.

# Analysis And Findings

## 1) Problems Statement

- Our management team needs to focus marketing and supply chain efforts on the most profitable product categories and brands. We need clear insights into which categories and brands are driving revenue and which are underperforming.

### 1) Total Revenue By Product Category.



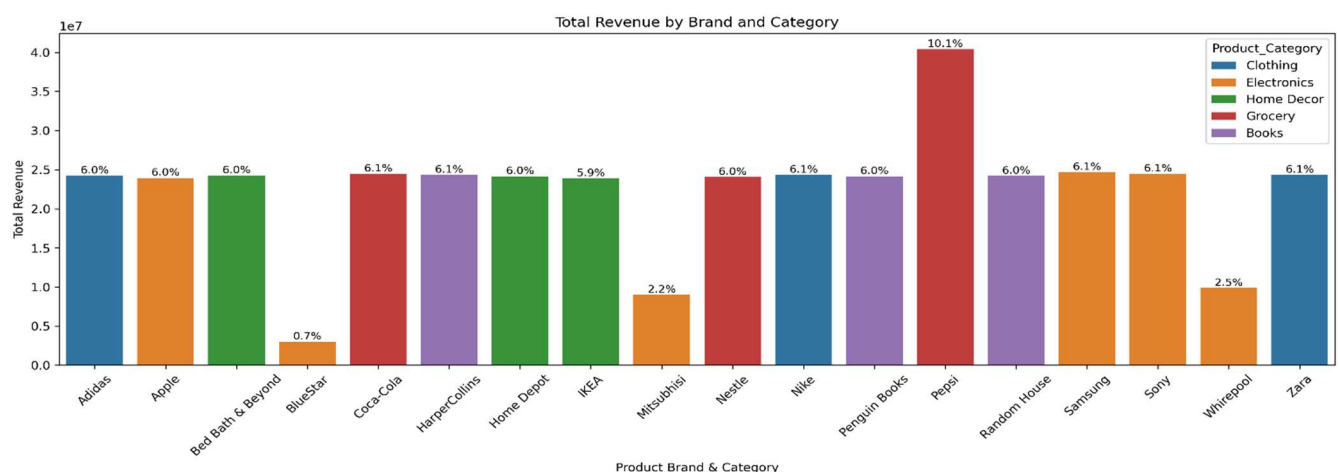
The total sales analysis reveals that Electronics is the top-performing category, contributing 23.6% of the total revenue, indicating strong consumer demand and value in this segment. Grocery closely follows with 22.1%, reflecting consistent high-volume sales. Books and Clothing have nearly equal shares, each at 18.1%, showing moderate but stable demand. Home Decor holds the smallest share at 18.0%,

which suggests room for growth in this category. Overall, Electronics and Grocery together account for nearly 46% of total revenue.

## Recommendation:

Focus on expanding Electronics and Grocery sales with targeted promotions. Boost marketing and product variety for Home Decor, Books, and Clothing to increase their revenue share.

## 2) Total Revenue by Brand and Category



Pepsi (Grocery) makes the most, with 10.1% of the total revenue. Most other big brands as Adidas (Clothing), Apple (Electronics), Bed Bath & Beyond (Home Decor), Coca-Cola (Grocery), HarperCollins (Books), and Samsung (Electronics) each make about 6.0% to 6.1%, which means their performance is quite steady. Some brands are making much less — Bluestar (Electronics) has only 0.7%, Mitsubishi has 2.2%, and Whirlpool has 2.5%, which shows they are not doing well. The Electronics category has the biggest difference between high and low performers. In contrast, categories like Books, Clothing, and Home Decor have more even earnings. Overall, most brands are

doing okay, but Pepsi is leading by a lot, and a few electronics brands need attention.

**Recommendation:**

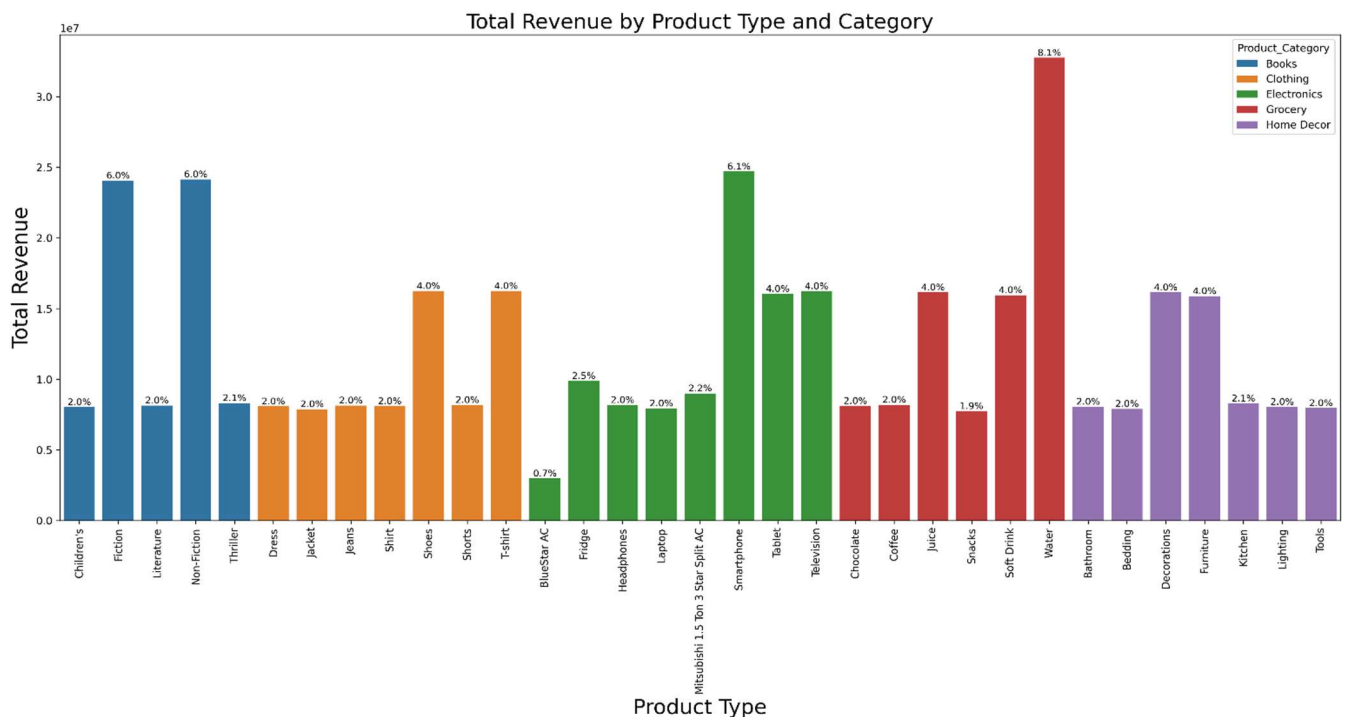
companies should focus on helping low-performing brands like Bluestar, Mitsubishi, and Whirlpool by improving marketing, product quality, or pricing.

Strong brands like Pepsi can be used for promotions or to launch new products.

the Electronics category shows mixed results, it's important to find out why some brands are doing well and others are not. At the same time, steady brands contributing around 6% should be supported to keep revenue stable.

It reducing too much dependence on Pepsi will help balance the business and lower risk.

**3) Total Revenue by Product Type and Category**



Water (Grocery) is the top revenue generator with 8.1%, followed by Smartphones (Electronics) and Fiction and Non-Fiction in (Books) at 6.1% . showing these are the most in-demand or high-value products.

Other strong performers include Shoes, Shirt, Soft Drink , Tablet, Television, Decorations, and Furniture, each contributing 4.0%, indicating solid sales within their categories.

Most other products contribute around 2.0%, including common items like Children's , literature , thriller(Books), Dress , Jacket, Jeans , Shirt , Shorts (Clothing) , Fridge , Headphones , Laptop , Ac (Electronics) , Chocolate , Coffee , Snacks (Grocery) , Bathroom , Kitchenware , Bedding , Lighting , Tools (Home Decor) , reflecting consistent but average demand.

A few items stand out for lower performance, especially BlueStar AC at 0.7% and Juice at 1.9%, which could point to weak sales.

### Recommendation:

The focus should be placed on boosting low-performing products like BlueStar AC and Juice through better marketing, discounts, or reviewing product quality.

High performers like Water and Smartphones or fiction and Non-Fiction should be further promoted or bundled to maximize profit. Since categories like Electronics and Grocery or Books show large performance gaps, it's important to analyze which product features or price points drive success.

It consistently performing categories like Clothing, and Home Decor should be maintained with regular product updates and seasonal promotions to keep customer interest steady.

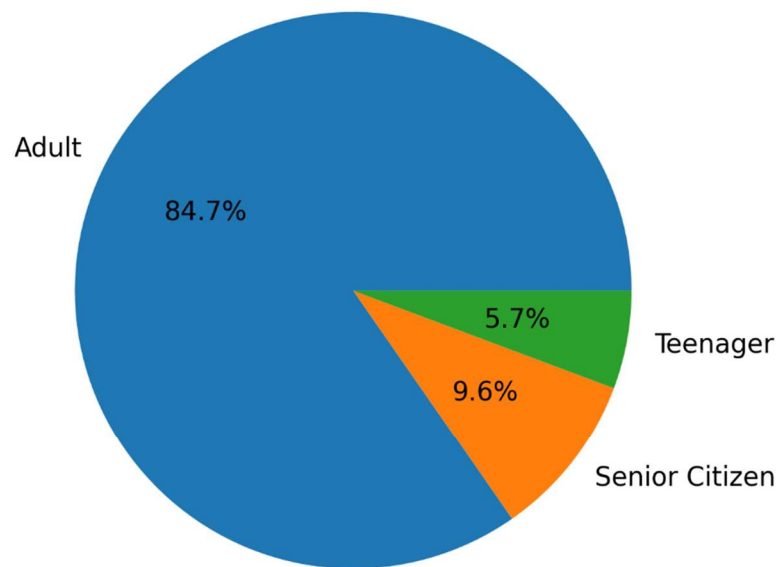
Use data to promote mid-range performers Shoes, Tablets, Soft Drinks with customized messaging, especially during peak seasons.

## **2) Problems Statement**

- We want to personalize marketing and improve ROI(Return on Investment). To do this, we need to segment our customers based on age, gender, income, and customer segment — and understand who brings the highest sales.

### **i) Age Group Based Total Sales**

Age Group Based Total Sales



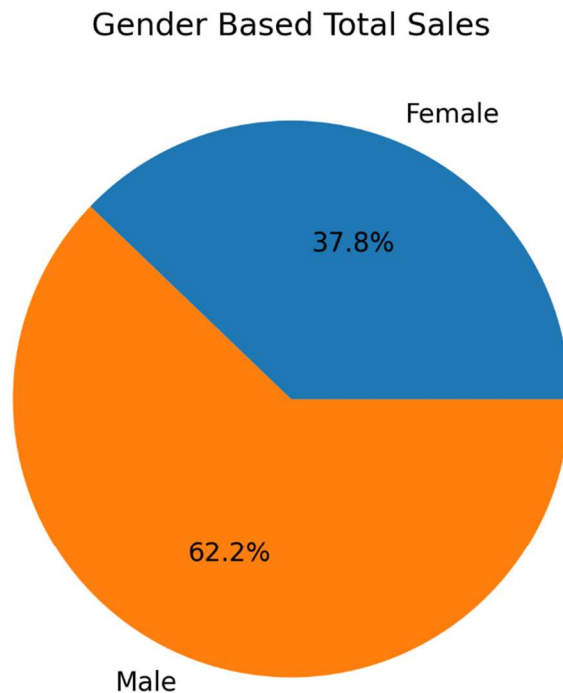
The chart shows that Adults dominate total sales, contributing 84.7%, making them the main target group. Senior Citizens account for 9.6%, while Teenagers make up just 5.7% of sales. This suggests that most purchasing activity comes from adults. Sales from teenagers and seniors are much lower, indicating potential areas to grow through targeted products or promotions designed specifically for these age groups.

**Recommendation:**

The Adult Age Group Target the most sales.

The Teenager and Senior Citizen to target Participate more sales using notification, marketing, discount.

## ii) Gender Based Total Sales



The chart indicates that males account for 62.2% of total sales, while females contribute 37.8%. This suggests that male customers are currently the dominant buyers. The noticeable gap implies a potential opportunity to increase sales from female customers by tailoring product offerings, marketing strategies, or promotions to better suit their preferences and shopping behaviours.

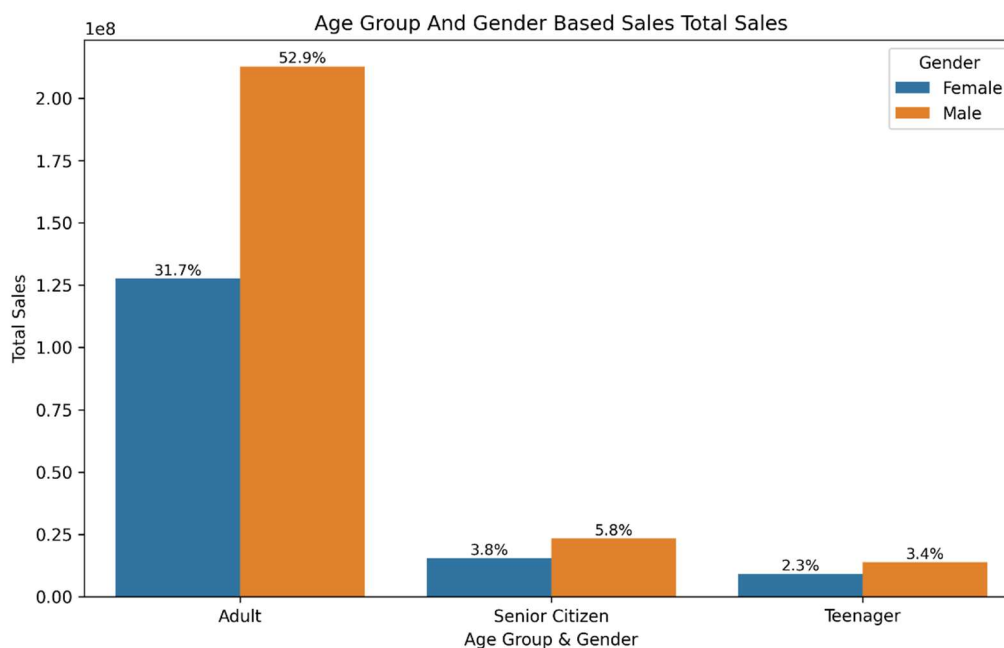
### **Recommendation:**

The Male to target more sales than marketing and offers.



The Female to target contribute more sales ads, marketing and offers or discount.

### iii) Age Group and Gender Based Total Sales



The chart shows that adult males dominate sales with a significant 52.9% share, followed by adult females at 31.7%. Senior citizens and teenagers contribute comparatively less, with male senior citizens at 5.8% and female senior citizens at 3.8%, while teenage males and females contribute 3.4% and 2.3% respectively. This clearly indicates that the adult age group, especially males, are the primary buyers.

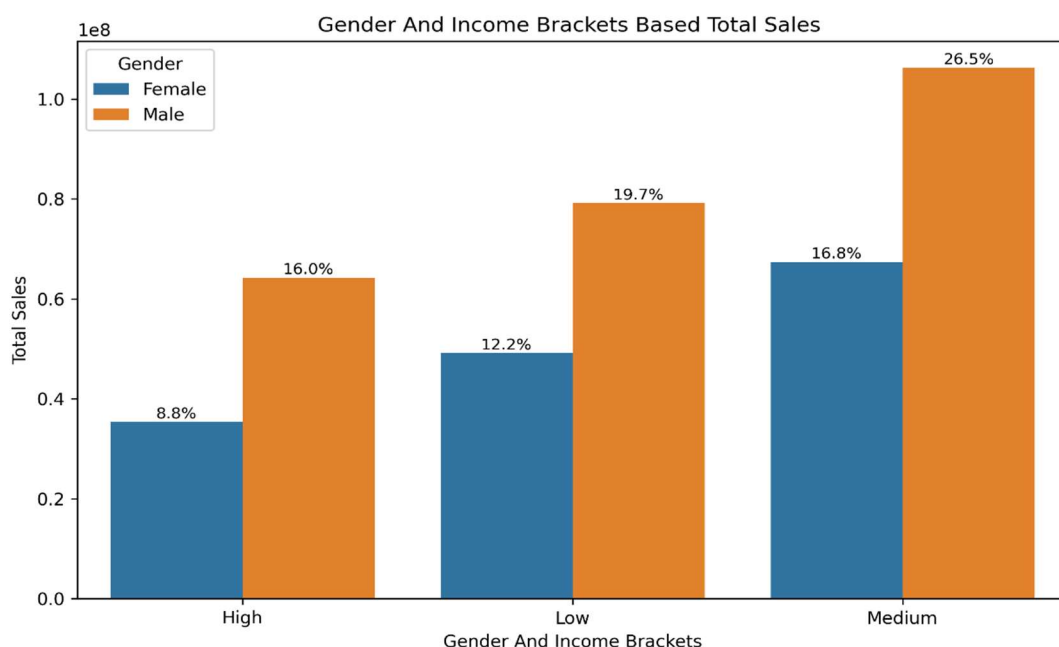
#### **Recommendation:**

Adult Males – They are the top buyers, focus campaigns and premium products on this group.

Adult Females More Second highest group, increase sales with female-focused marketing.

The Teenager and Senior Citizen is low sales they do the try this age group marketing and promotion to growth together.

#### iv) Gender And Income Brackets Based Total Sales.



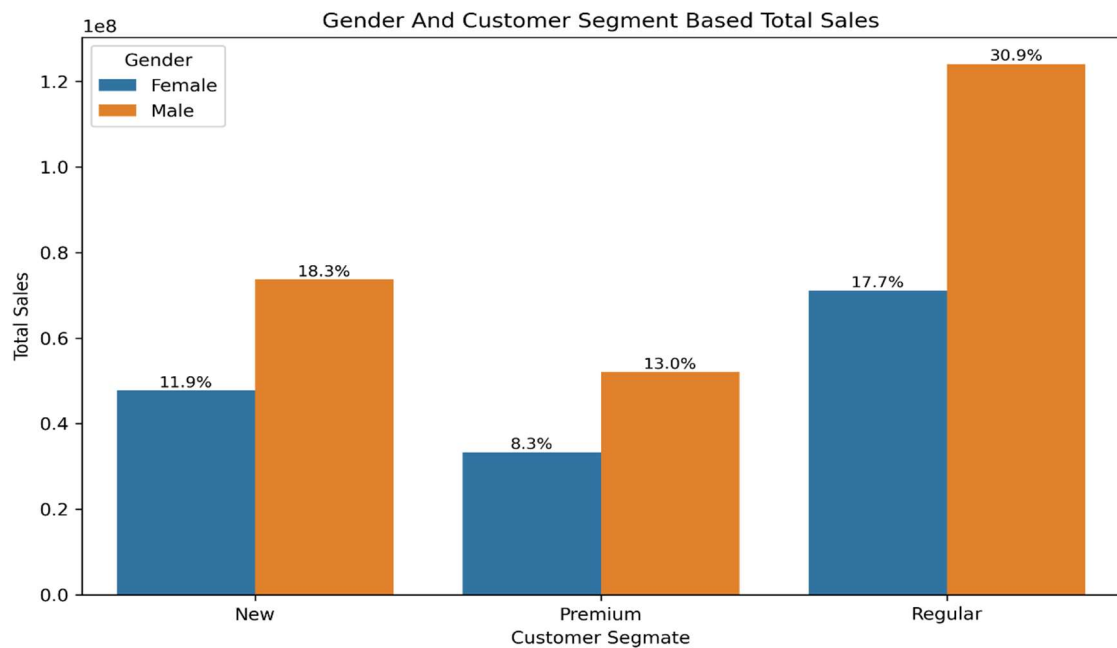
Men spent more than women in all income levels. The highest spending came from men in the medium-income group 26.5%, followed by women in the same group 16.8%. In the low-income group, men spent 19.7%, while women spent 12.2%. The least spending came from the high-income group, where men spent 16% and women only 8.8%. This means that medium-income men are the biggest buyers, and there's a chance to increase sales by encouraging more spending from women.

## Recommendation:

Focus on Medium Males Because They are the highest spenders and key revenue drivers.

Female spending is lower, targeted offers or promotions can help increase it.

### v) Gender And Customer Segment Based Total Sales.



The Male customers contribute more sales than Female customers across all segments. The Regular Male segment leads with the highest contribution at 30.9%, followed by Regular Female at 17.7%, making the Regular segment the most profitable overall. In the New

segment, Males account for 18.3% while Females contribute 11.9%. The Premium segment has the lowest share overall, with Males at 13.0% and Females at 8.3%.

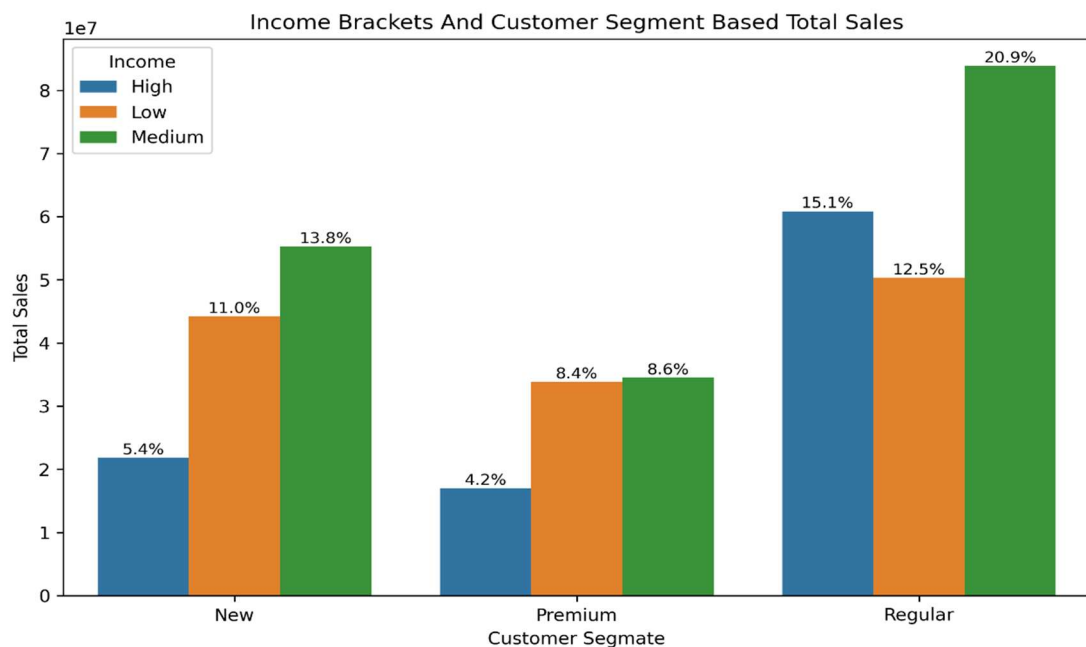
### **Recommendation:**

Focus marketing and loyalty efforts on Regular and New Male segments.

Use personalized campaigns to engage Premium and Female customers.

Explore ways to increase value and benefits in the Premium segment.

### **vi) Income Brackets and Customer Segment Based Total Sales**



The Medium income contribute the highest overall sales across all segments. In the Regular customer segment, Medium income users lead with 20.9%, followed by High income at 15.1%, and Low income at 12.5%. The New segment also shows strong sales from Medium income 13.8% and Low income 11.0% customers, while High income contributes just 5.4%. In the Premium segment, sales are almost and Medium income customers 8.6%, and Low income at 8.4% with High income contributing the least at 4.2%.

### **Recommendation:**

The Regular Segment Customer Medium Income Brackets more sales offers and Low Income Brackets participate offers and discount or High Income Brackets more sales do ads and marketing.

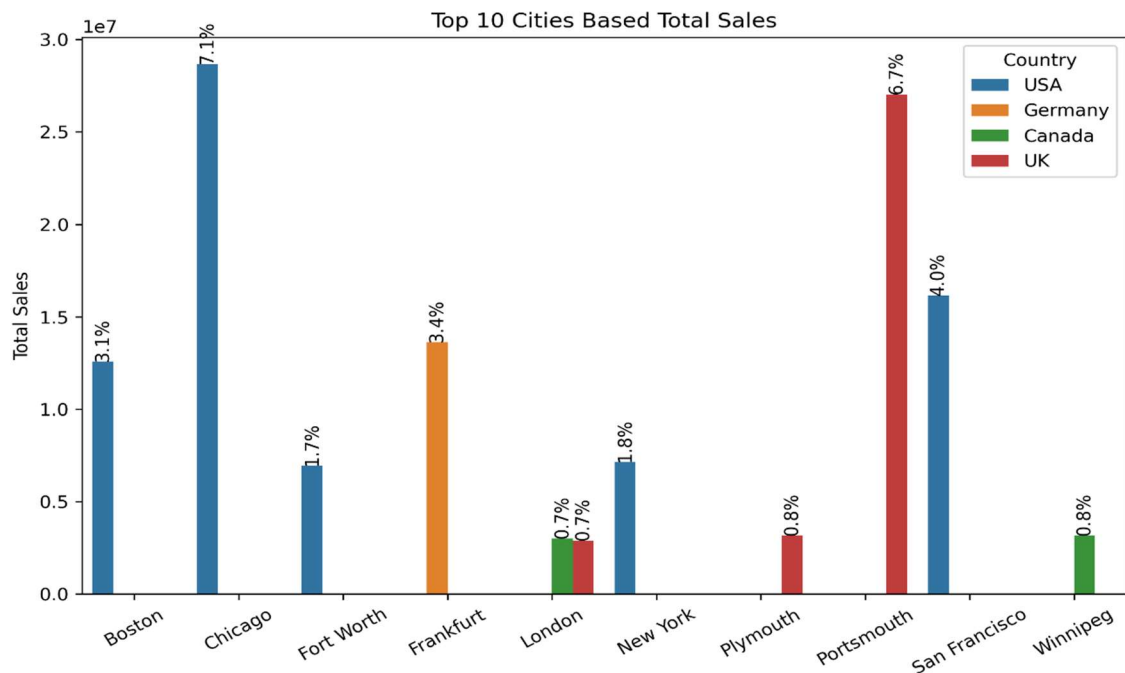
The Second high sales contribute is New Segment they Medium , Low and High Income engagement to more sales.

The Premium Segment all Income Brackets do the special offers and discount marketing.

### **3) Problems Statement.**

- As we plan to expand operations, leadership wants insights into regional performance. We need to know which countries, states, and cities generate the most revenue and which areas need attention.

#### **i) Top 10 Cities by Total Sales.**



Highest Sales City is First Chicago 7.1% Country is USA and second is Portsmouth 6.7 Country is UK.

The Middle level Sales contribute city is San Francisco USA with 4.0%. Frankfurt Germany adds 3.4%, and Boston USA adds 3.1%.

The Low level Sales Contribute city is New York USA with 1.8% and Fort Worth USA add 1.7% or London UK with 0.7% and Canada with 0.7 %, Plymouth UK add 0.8%, Winnipeg Canada with 0.8%.

Most of the top cities are in the USA, showing strongest market and Other Cities in the UK, Canada, and Germany have smaller sales.

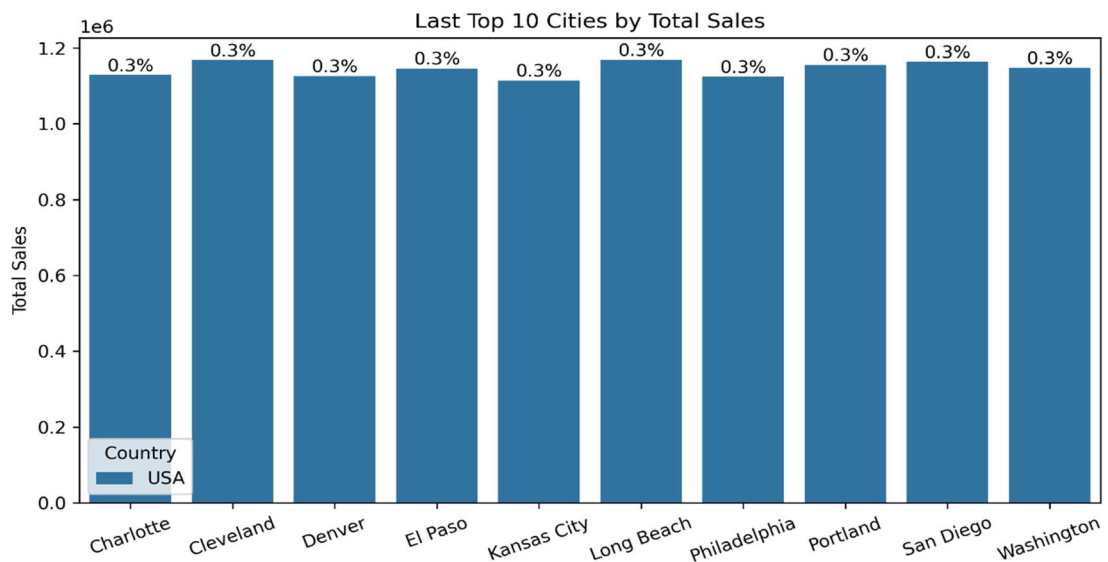
### **Recommendation:**

Invest more in the USA, especially in high-performing cities like Chicago and San Francisco.

Explore growth opportunities in underperforming cities like London Canada and Plymouth UK.

The lowest sales cities engage more sales using offers, ads.

## ii) Last Top 10 Cities by Total Sales.



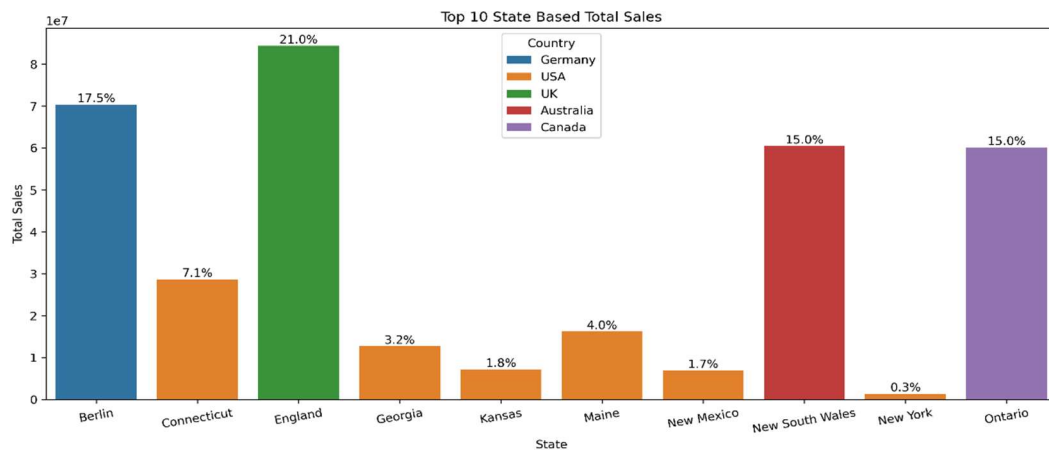
In all the USA Each city, including Charlotte, Cleveland, Denver, El Paso, Kansas City, Long Beach, Philadelphia, Portland, San Diego, and Washington, contributes just 0.3% of the total Sales. The revenue levels are very similar across these cities. Suggesting they are underperforming or less focused markets.

### Recommendation:

Run targeted campaigns Use promotions or local advertising to increase visibility in these cities.

Improve product and service access Check if distribution or service issues are limiting reach.

### iii) Top 10 State Based Total Sales.



The England, Berlin, these states contribute with 21.0%, 17.5% and New South Wales, and Ontario, as these states contribute 15.0%. These high-performing regions indicate strong market potential. On the other hand, New York is 0.3% and New Mexico is 1.7% or Kansas is 1.8% show very low revenue contribution. The Mid-level Sales contribute state is Georgia is 3.2% and Maine is 4.0%.

#### **Recommendation:**

Focus on the high Sales states England, Berlin, Ontario, and New South Wales to boost returns.

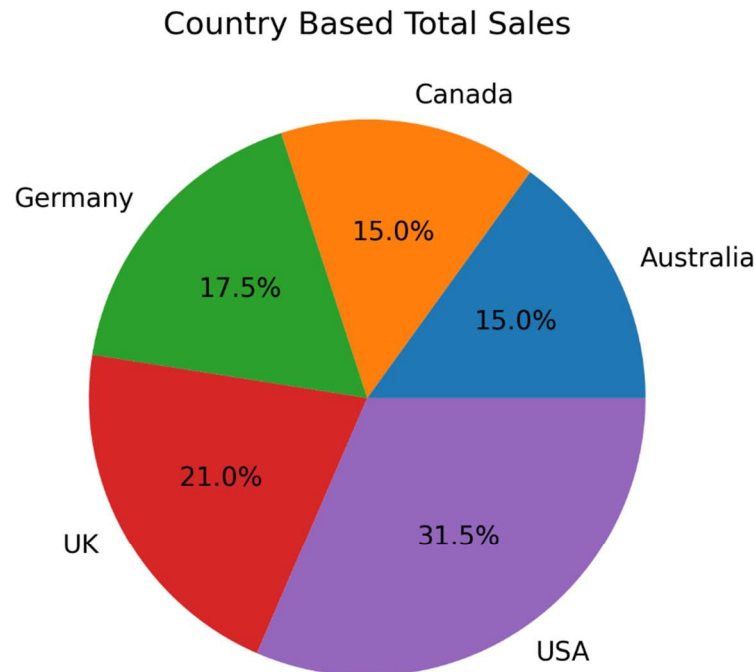
Improve marketing or service in low-performing areas New York and New Mexico, Georgia, Maine.

Explore why top cities Chicago and Portsmouth are performing well, and apply similar strategies elsewhere in this city.

Consider reducing investment in cities consistently showing low revenue.



#### iv) Country Based Total Sales



The USA country contributes the highest share of total sales with 31.5. UK is 21.0%, also showing strong performance. Germany holds 17.5%, a moderate but notable Sales. Canada and Australia both contribute 15.0%, showing balanced but comparatively lower performance.

#### **Recommendation:**

Focus the USA Country Because They Highest Sales.

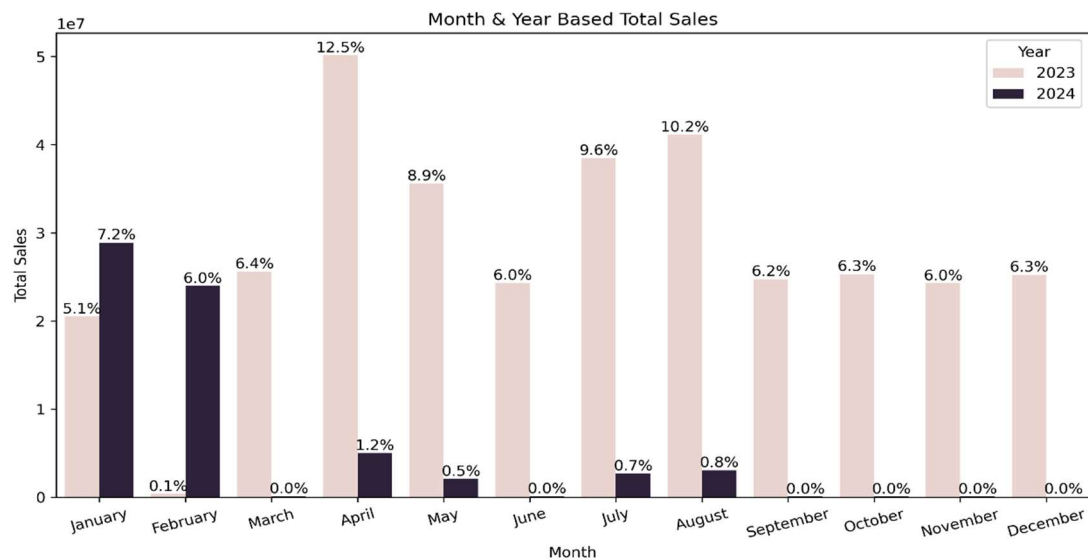
The UK and GERMANY strong markets that can be expanded with targeted strategies and more sales.

The Canada and Australia run marketing or promotional campaigns to increase market Sales.

#### 4) Problem Statement.

- 4) The marketing team wants to schedule campaigns more effectively. We need a clear understanding of how sales behave over time — by month and year — to identify peak seasons and low-performing periods.

##### i) Month & Year Based Sales.



2023 had strong, consistent sales from April to August, peaking in April (12.5%) and August (10.2%).

2023 had second high sales from May and July both May 8.9% and July 9.6%.

2023 had Low Sales is Month is January 5.1%.

2023 had without above the Month Average Sales is 6.0% to 6.4%.

2024 only shows partial data with January 7.2% and February 6.0% performing best

2024 Only shows partial data most highest sales month is January 7.2% and February 6.0%.

2024 Mid sales is April 1.2%, May 0.5%, July 0.7%, August is 0.8%.

2024 without above the month sales nothing sales 0%.

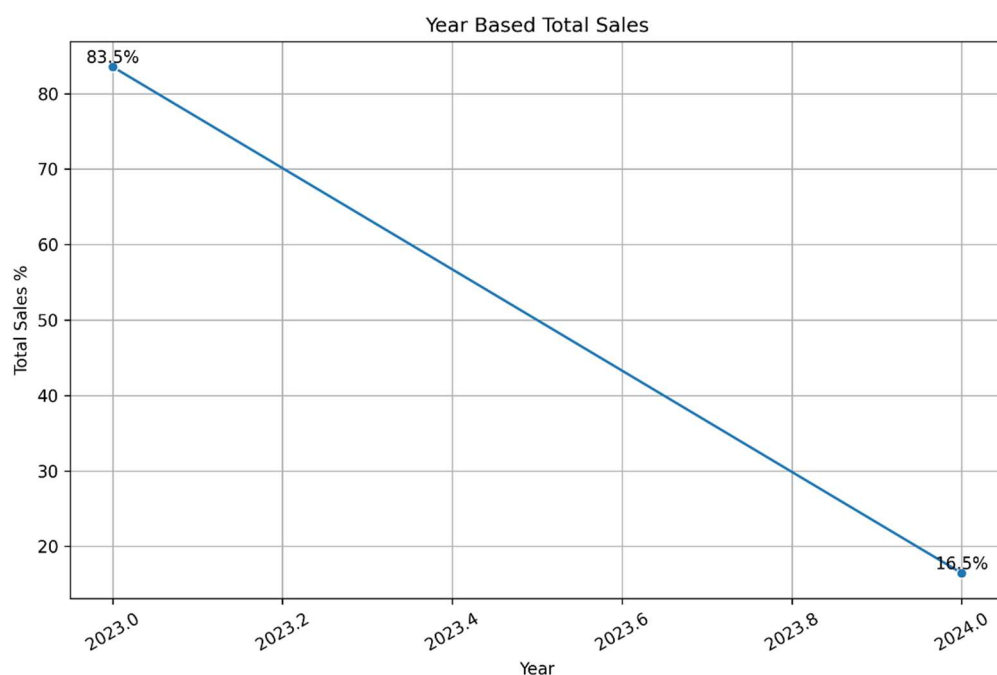
### **Recommendation:**

The April and August promotions and campaigns during these high performing months.

2024 this year mostly not data in database.

Explore why 2023 Month are performing well, and apply similar strategies elsewhere in this 2024 Month.

### **ii) Year Based Total Sales.**



There is a major different in total sales from 2023, 83.5% to 2024, 16.5%.

In 2024 incomplete data for the year.

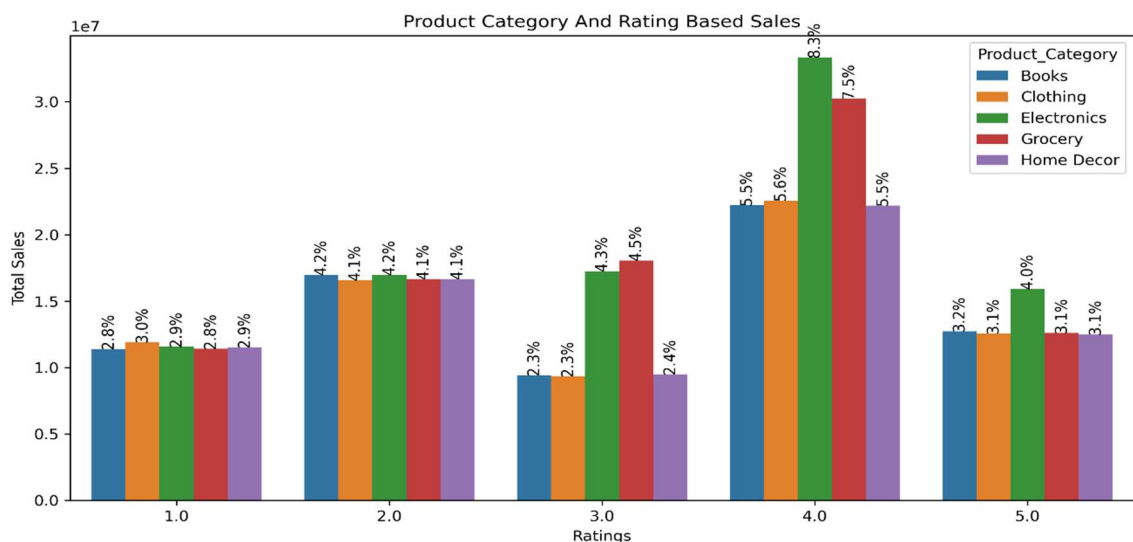
### Recommendation:

Implement corrective campaigns in upcoming months.

## 5) Problems Statement.

- Some products sell well but receive poor ratings. This may harm long-term customer trust. We want to cross-analyze sales volume with customer ratings to identify risky or high-potential products.

### i) Product Category and Rating Based Sales.



The highest total sales are observed for products rating 4.0, especially in Electronics 8.5% and Grocery 7.5% categories. These

two product types perform better at this rating level compared to others. Other products are average sales rating is 4.0.

Rating 5.0 show lower sales across all categories, with Electronics again leading 4.0%, but by a smaller sale. Other products are sales 3.1%.

Rating 1.0 is lower sales compare to rating 5.0, rating 1.0 all products are sales is 2.9%.

Rating 3.0 is better sales Electronics 4.3% and Grocery 4.5% perform. All products lower sales as compare to other 2.3%.

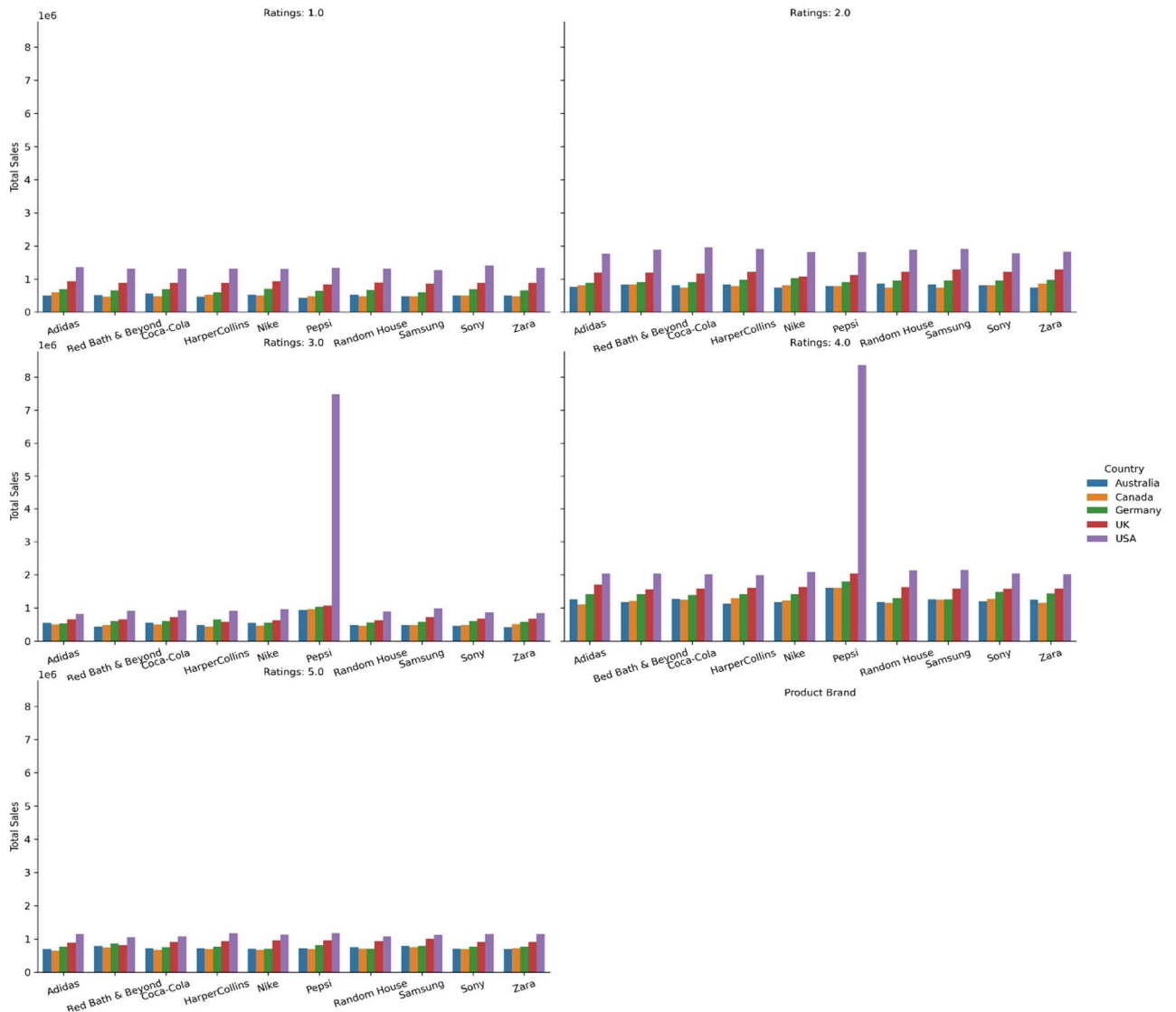
All categories have moderate and nearly equal sales at this rating 2.0, Books and Electrics have the highest sales at 4.2%. other products are sales is 4.1%.

### **Recommendation:**

Focus marketing on products with 4.0 ratings because they are give's more sales especially in Electronics & Grocery.

1 rating products don't have the highest sales, possibly due to low review count.

## ii) Product Brand and Rating or Country Based Sales



In Rating 3.0 and 4.0, Pepsi USA has extremely high sales far higher than any other brand-country combination.

In rating 4, highest sales all product brands as compare to other ratings.

Rating 2, sales begin to pick up slightly but remain modest all the product brands.

Rating 3, many brands, especially electronics and drink, show stronger sales across countries and other products brand as average sales contribute.

Rating 5, sales are good but not the highest. Drink, show the high sales in rating 5 but all products are slight dip in sales.

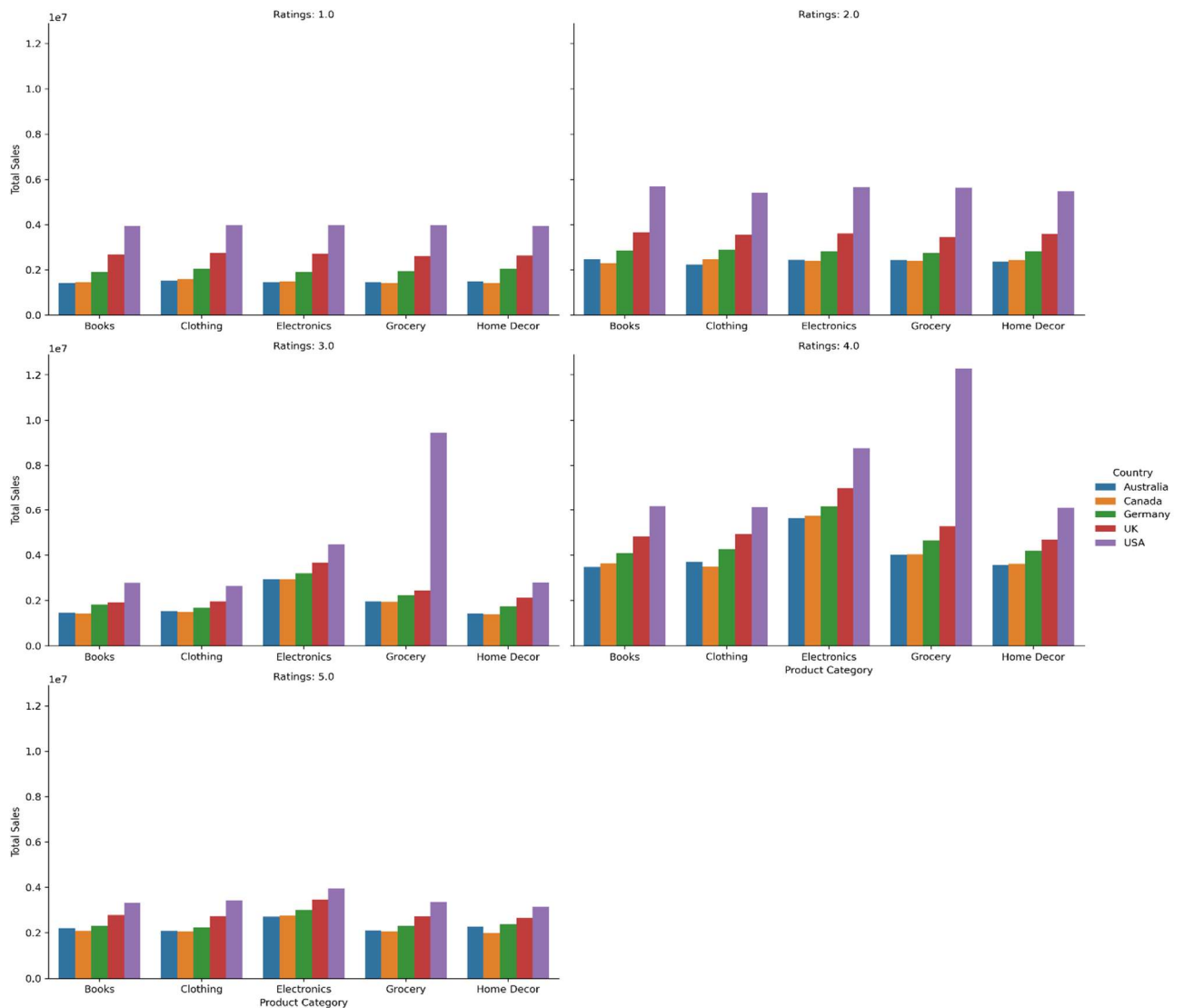
**Recommendation:**

Target and promote 4 ratings.

Avoid over optimizing for perfect ratings.

Boost engagement in high sales countries USA, UK.

### iii) Product Category and Rating or Country Based Sales



Rating 1, Sales are lowest across all product categories and countries. Even in strong markets the USA, all product category is sales lower



than other rating other rates, Customer dissatisfaction is clearly linked with poor sales.

Rating 2, Sales improve but still remain weak. Grocery and Electronics perform a bit better than others, especially in the USA second is UK, all products category is moderate level sales.

Rating 3, Sales begin to pick up. Electronics and Grocery see noticeable growth, especially in the USA and UK. This is the minimum acceptable rating level sales, and customer are mid-level trust this rating.

Rating 4, Sales peak sales across almost all product categories and countries. USA leads then UK, with Grocery and Electronics dominating, all products category are stills more sales than other ratings, and customer satisfaction is mostly high purchases.

Rating 5, Sales slightly dip compared to 4.0, though still high. USA remains the top performer, and Electronics and Clothing do well, then other products category is mid sales.

### **Recommendation:**

Focus on products with 4.0 ratings sales almost every category and country.

Using customer feedback to identify improvement areas that can push 3 to 4 ratings.

Run targeted marketing campaigns in high-performing regions USA, UK.

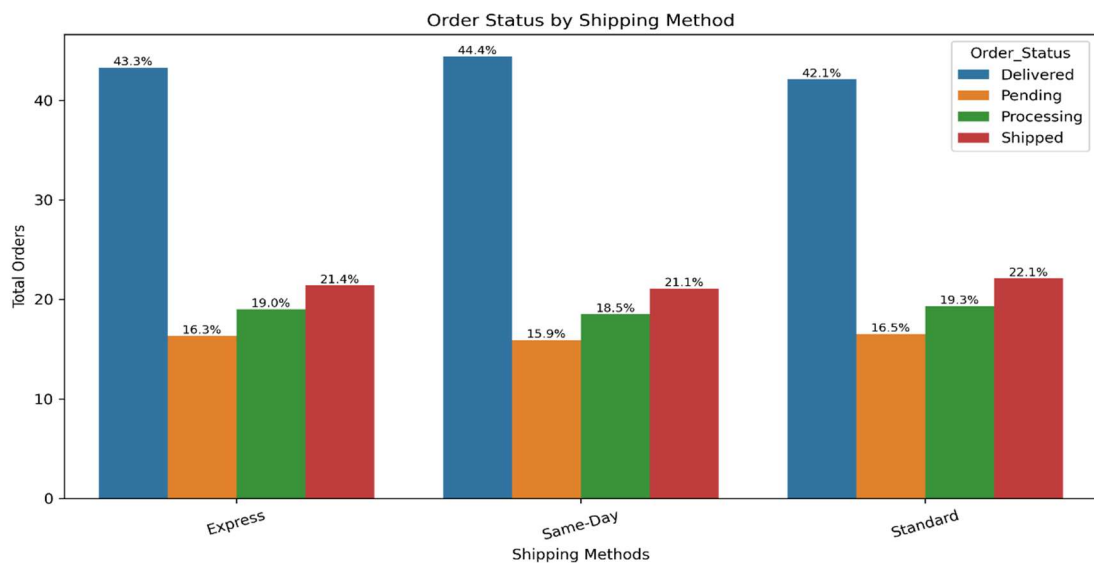
Rating 1 and 2 show the poor rating sales investigate customer service issues tied to these products.

Sales slightly decline at the 5 ratings, due to some reason like trust issues too perfect or fewer reviews.

## 6) Problem Statement.

- Customer feedback has highlighted concerns about delivery speed and reliability. We need to analyze how different shipping methods affect order status and customer satisfaction.

### i) Order Status by Shipping Method



The majority of orders are Delivered, ranging between 42.1% and 44.4%, showing strong logistics efficiency.

a notable share of orders is still in Pending 15.9%–16.5% all Shipping Methods. The Processing 18.5%–19.3% in an all-Shipping Methods, both order Pending or Processing do the speed this works.

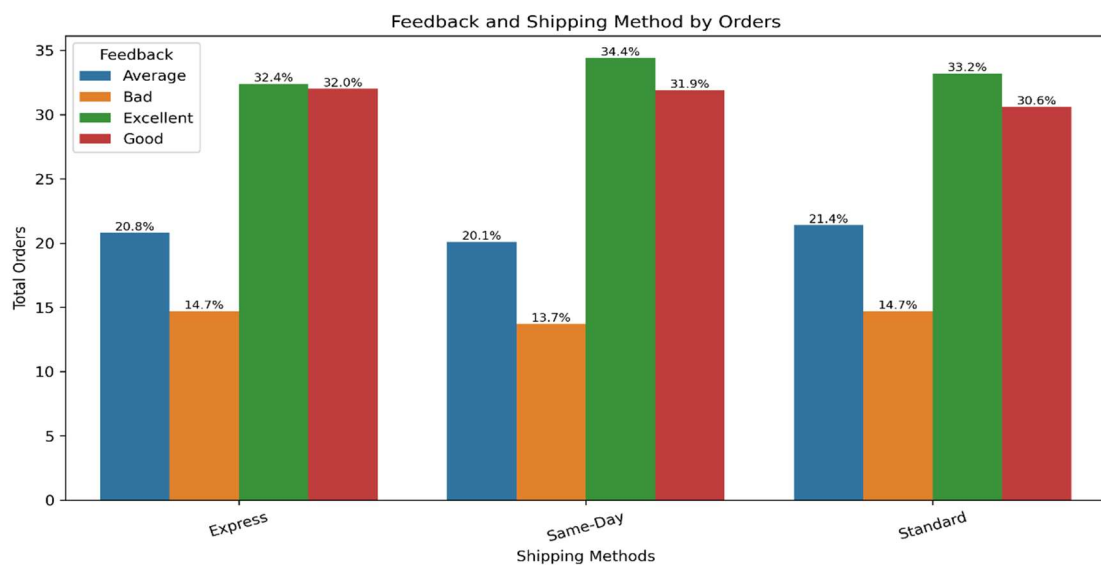
The Shipped status is also high 21.1%–22.1%, the many orders are still in transit and not delivered.

### Recommendation:

fulfilment operations to reduce the percentage of Pending and Processing orders, especially for Same-Day deliveries.

The shipping partners to ensure quicker transitions from Shipped to Delivered, especially in Standard shipping.

### ii) Feedback and Shipping Method by Orders



the highest of customer feedback is Excellent 32.4–34.4%, followed by Good is 30.6–32%. This indicates strong overall satisfaction with shipping.

The Average feedback is 20.1-21.4, and Bad feedback is 13.7-14.7 minimal dissatisfaction of the customers.

**Recommendation:**

Promote Same-Day and Express shipping in marketing they outperform in Excellent ratings.

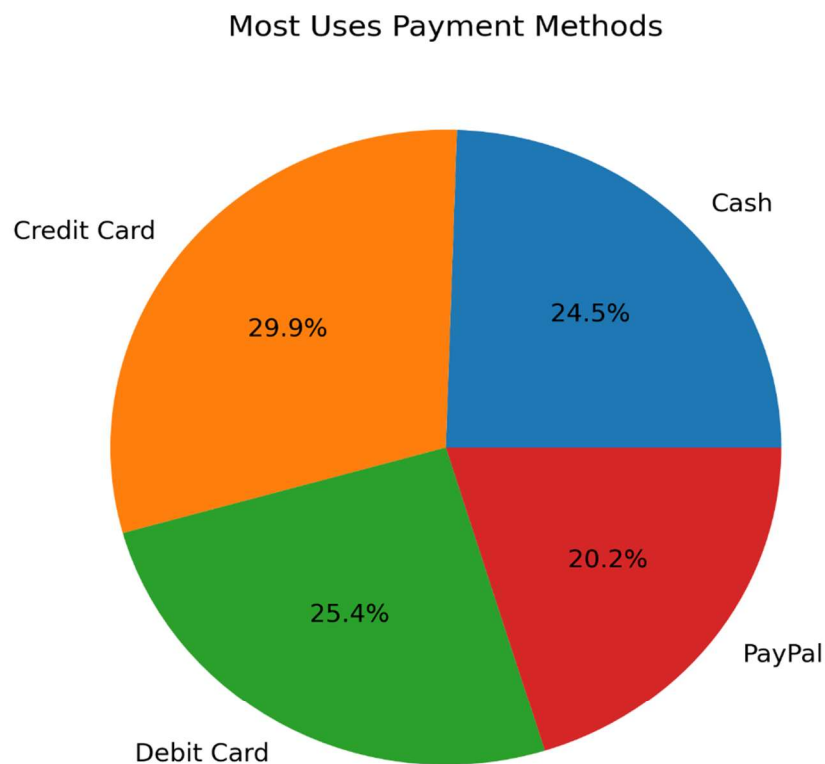
improve the communication or delivery accuracy.

Continue to maintain and enhance all the shipping methods is Good and Excellent ratings.

## 7) Problem Statement.

- We want to reduce cart abandonment and improve checkout flow. To do this, we need to identify the most preferred payment methods across different customer groups and regions.

### i) Most Uses Payment Methods



The most preferred payment method is Credit Card 29.9%, followed by Debit Card 25.4% and Cash 24.5%, while PayPal 20.2% is the least used.

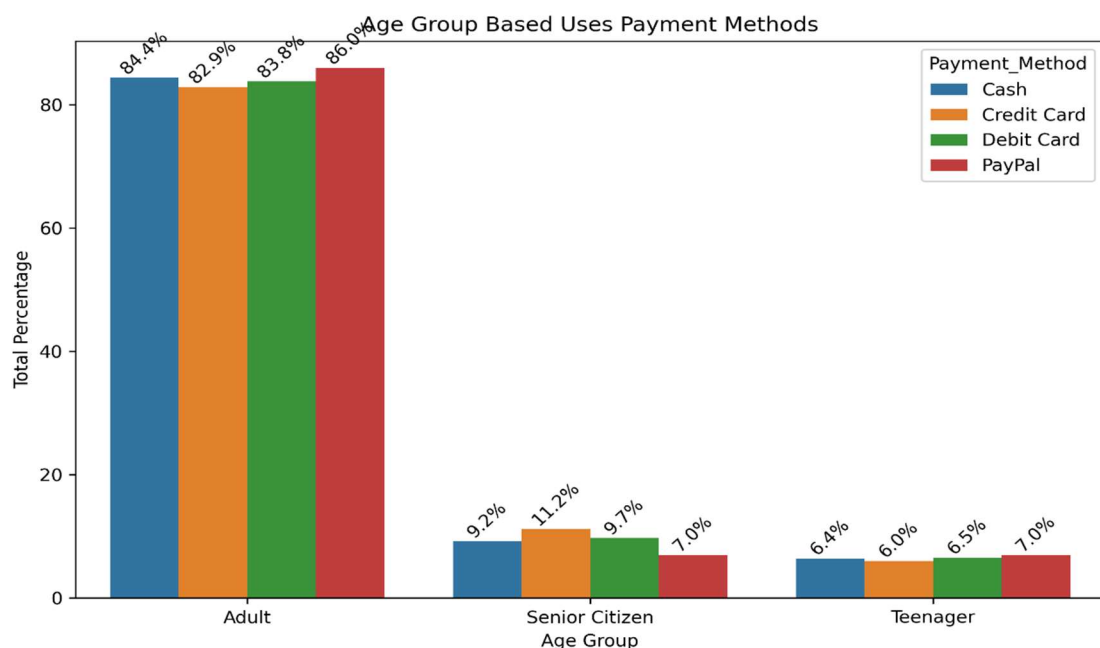
### **Recommendation:**

digital transactions are more dominant than cash or online wallets.

The give's discounts and offers for PayPal and Cash users to boost adoption customers.

The priority payment card options Credit and Debit on checkout for a smoother user experience.

## ii) Age Group Based Uses Payment Methods



Adults are the most uses users of all payment methods is PayPal 86.0%, Cash 84.4%, Debit Card 83.8%, Credit Card 82.9%.

Senior Citizens show most payment usage is Highest with Credit Card 11.2% and 9.7%, Cash 9.2%, and Lowest with PayPal 7.0%.

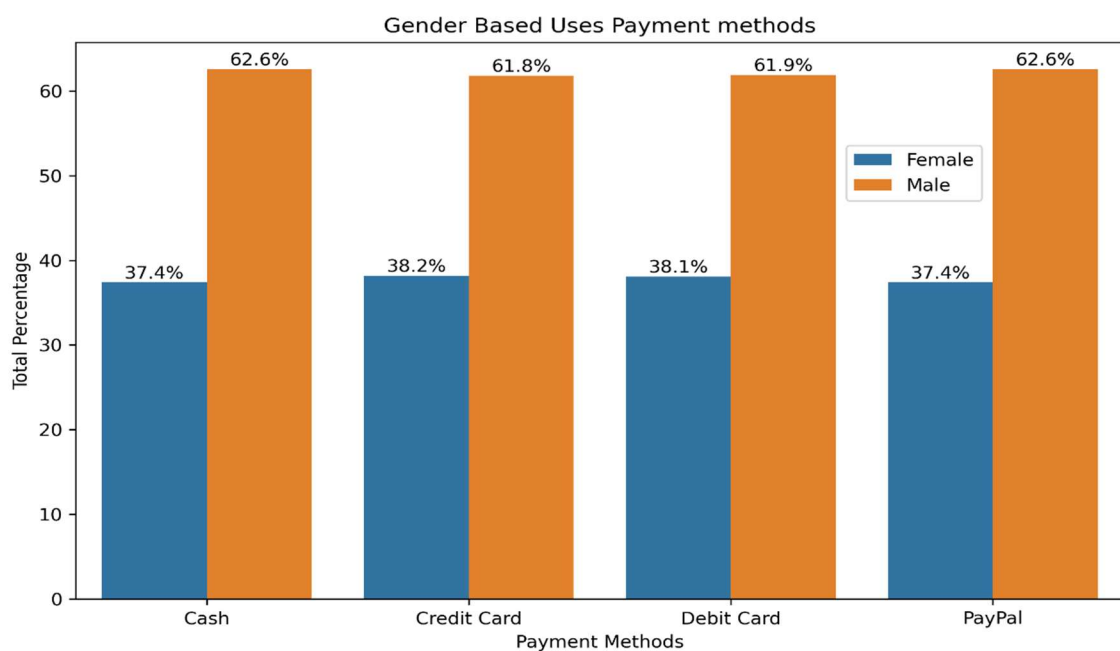
Teenagers have the least usage all payment methods around 6%–7% for all payment, with PayPal 7.0% slightly higher.

## Recommendation:

Focus marketing and loyalty programs on adults.

Consider teen-focused offers discounts via prepaid cards and senior-friendly payment tutorials to broaden user base.

### iii) Gender Based Uses Payment methods



Males consistently use all payment methods more than females.

Male usage payment methods are ranges from 61.8% to 62.6% across methods.

Females have lower usage across all payment types, the uses Between 37.4% and 38.2%.

**Recommendation:**

Run promotions or discounts for females using specific payment methods.

Collaborate with female-centric brands, influencers, or platforms.

Highlight security features of each method.

Create a feedback ask female users what would make them feel more confident and comfortable.



# **Tools & Technologies**

- - Python (Pandas, Matplotlib, Seaborn)
- - Jupyter Notebook
- - CSV Data Source (Retail Sales Data)

## **Data Overview**

- - 302010, rows of transaction data
- - Key columns (31): Transaction\_ID, Customer\_ID, Name, Email, Phone, Address, City, State, Zipcode, Country, Age, Gender, Income, Customer, Segment, Date, Year, Month, Time, Total\_Purchases, Amount, Total\_Amount, Product\_Category, Product\_Brand, Product\_Type, Feedback, Shipping\_Method, Payment\_Method, Order\_Status, Ratings, products, Age\_Group.

## **Data Cleaning**

- - Handled missing: whole data
- - Converted column: Transaction\_ID in Int, Customer\_ID in Int, Phone in Int, Zipcode in Int, Age in Int, Date in datetime, Year in Int, Time in timedelta, Total\_Purchases in Int, Amount in Int, Total\_Amount in Int.
- - Created new columns: Age\_Group.
- Remove the duplicates and null values.

**Name: Jignesh Hadiya.**