

AMAZON ANALYSIS

PRASHANT UPADHYAY

26 AUGUST 2025



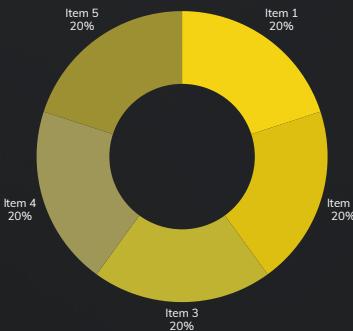
amazon

AMAZON, FOUNDED BY JEFF BEZOS IN 1994 AS AN ONLINE BOOKSTORE, HAS GROWN INTO ONE OF THE WORLD'S LARGEST MULTINATIONAL TECHNOLOGY COMPANIES. HEADQUARTERED IN SEATTLE, USA, AMAZON OPERATES ACROSS E-COMMERCE, CLOUD COMPUTING, DIGITAL STREAMING, AND ARTIFICIAL INTELLIGENCE. ITS FLAGSHIP PLATFORM, AMAZON.COM, REVOLUTIONIZED RETAIL BY OFFERING A WIDE RANGE OF PRODUCTS WITH FAST DELIVERY, SUPPORTED BY INNOVATIONS LIKE PRIME MEMBERSHIP AND AN ADVANCED GLOBAL LOGISTICS NETWORK. BEYOND E-COMMERCE, AMAZON WEB SERVICES (AWS) HAS BECOME A LEADER IN CLOUD COMPUTING, SERVING STARTUPS, ENTERPRISES, AND GOVERNMENTS WORLDWIDE. THE COMPANY ALSO OWNS ALEXA, KINDLE, AUDIBLE, AND STREAMING SERVICE PRIME VIDEO. AMAZON EMPHASIZES CUSTOMER OBSESSION, OPERATIONAL EXCELLENCE, AND INNOVATION AS ITS CORE PRINCIPLES. WITH A MARKET PRESENCE IN MULTIPLE COUNTRIES, AMAZON HAS TRANSFORMED SHOPPING HABITS AND CONTINUES TO REDEFINE INDUSTRIES THROUGH TECHNOLOGY, DATA-DRIVEN DECISIONS, AND EXPANSION INTO AREAS LIKE AUTOMATION, ROBOTICS, AND SUSTAINABILITY INITIATIVES.

Agenda

THE AGENDA OF THIS AMAZON E-COMMERCE ANALYTICS PROJECT IS TO DERIVE VALUABLE INSIGHTS FROM ORDERS, PRODUCTS, AND CUSTOMER DATA TO SUPPORT STRATEGIC DECISION-MAKING. THE PROJECT BEGINS WITH DATA PREPARATION, INCLUDING CLEANING, TRANSFORMATION, AND MODELING OF DATASETS FROM MULTIPLE SOURCES. IT THEN FOCUSES ON ORDER AND PRODUCT ANALYSIS TO IDENTIFY TOP-SELLING PRODUCTS, CATEGORY-WISE REVENUE, AND PAYMENT METHOD PREFERENCES. SALES PERFORMANCE IS STUDIED THROUGH MONTHLY SALES TRENDS, COD VS PREPAID PERFORMANCE, AND OVERALL REVENUE GROWTH. CUSTOMER INSIGHTS SUCH AS REPEAT PURCHASE BEHAVIOR, NEW CUSTOMER ACQUISITION, AND PRODUCT RETURN RATES ARE ALSO HIGHLIGHTED. FINALLY, INTERACTIVE DASHBOARDS IN POWER BI, ENHANCED WITH BUTTONS, SLICERS, AND URL ICONS, ALONGSIDE PYTHON VISUALIZATIONS, BRING THE ANALYSIS TO LIFE.

PROJECT OVERVIEW



DATA SOURCES

ORDERS DATA – ORDER DETAILS, STATUS, PAYMENT METHODS, REVENUE.

PRODUCT DATA – PRODUCT NAMES, CATEGORIES, PRICING.

CUSTOMER REVIEWS – RATINGS ON QUALITY, DELIVERY, AND PACKAGING.

KEY ANALYSIS AREAS

ORDER & PRODUCT INSIGHTS → TOP-SELLING PRODUCTS, SALES BY CATEGORY, AND PAYMENT TRENDS.

SALES PERFORMANCE → MONTHLY SALES, COD VS PREPAID ORDERS, CANCELLATION AND RETURN RATES.

CUSTOMER INSIGHTS → REPEAT PURCHASE RATE, NEW CUSTOMER ACQUISITION, AND SATISFACTION RATINGS.

TOOLS & TECHNIQUES

POWER BI → DATA MODELING, DAX MEASURES, INTERACTIVE DASHBOARDS WITH BUTTONS, SLICERS, AND URL ICONS.

PYTHON (MATPLOTLIB & SEABORN) → REPLICATED AND ENHANCED VISUALIZATIONS FOR DEEPER INSIGHTS.

EXCEL & SQL → DATA CLEANING, JOINS, AND PRELIMINARY ANALYSIS.

OUTCOME

DELIVERED AN END-TO-END ANALYTICS SOLUTION THAT TRANSFORMED RAW E-COMMERCE DATA INTO INTERACTIVE DASHBOARDS AND VISUAL REPORTS, HELPING IDENTIFY GROWTH OPPORTUNITIES AND CUSTOMER TRENDS.

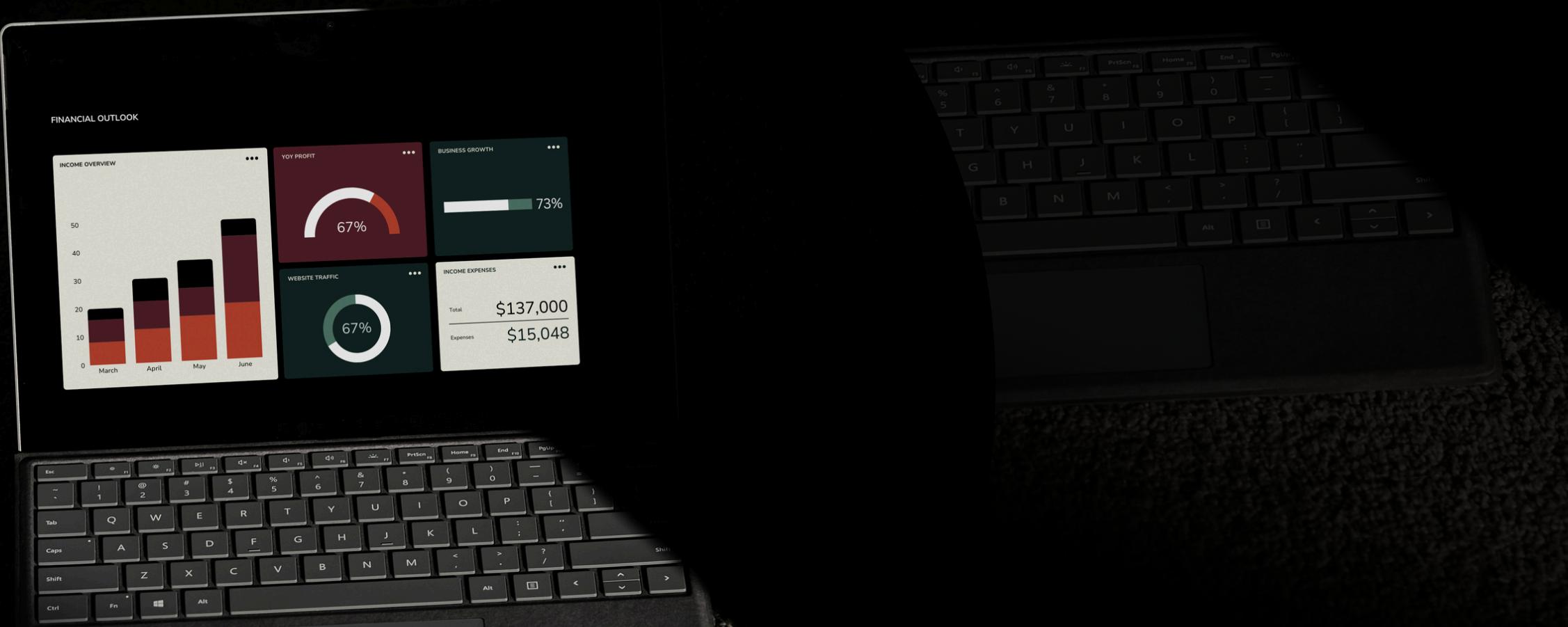
ACKNOWLEDGEMENT

I WOULD LIKE TO EXPRESS MY SINCERE GRATITUDE TO EVERYONE WHO SUPPORTED ME THROUGHOUT THE COMPLETION OF THIS AMAZON E-COMMERCE ANALYTICS PROJECT. I AM THANKFUL FOR THE GUIDANCE, ENCOURAGEMENT, AND VALUABLE FEEDBACK THAT HELPED ME SHAPE THIS PROJECT INTO A COMPREHENSIVE ANALYSIS.

SPECIAL THANKS TO MY MENTORS, PEERS, AND ONLINE LEARNING RESOURCES FOR PROVIDING KNOWLEDGE AND DIRECTION IN USING TOOLS LIKE POWER BI, PYTHON, SQL, AND EXCEL EFFECTIVELY. THEIR INSIGHTS GREATLY ENHANCED MY UNDERSTANDING OF DATA ANALYTICS AND VISUALIZATION.

I ALSO ACKNOWLEDGE THE DATASETS USED, WHICH ENABLED ME TO EXPLORE REAL-WORLD E-COMMERCE SCENARIOS AND BUILD MEANINGFUL DASHBOARDS. THIS PROJECT HAS BEEN A GREAT LEARNING EXPERIENCE IN CONNECTING BUSINESS PROBLEMS WITH DATA-DRIVEN SOLUTIONS.

ANALYSIS THROUGH DASHBOARDS



amazon

Sales Analysis



2022 2023 2024 2025

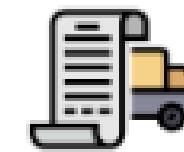
Total Customer

5K



Total Order

6K



Net Sales

576M



Sales Growth Rate

108.48%

Repeat Customer Rate

10.00

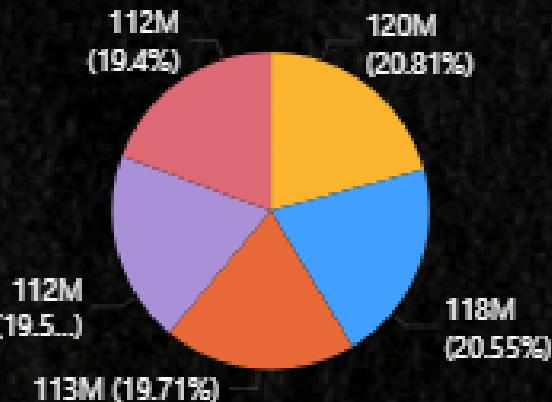


Sum of sales by month



Sum of TotalPrice by PaymentMethod

PaymentMethod
● UPI
● Net Banking
● Cash on Del..
● Debit Card
● Credit Card



Sum of TotalPrice by City



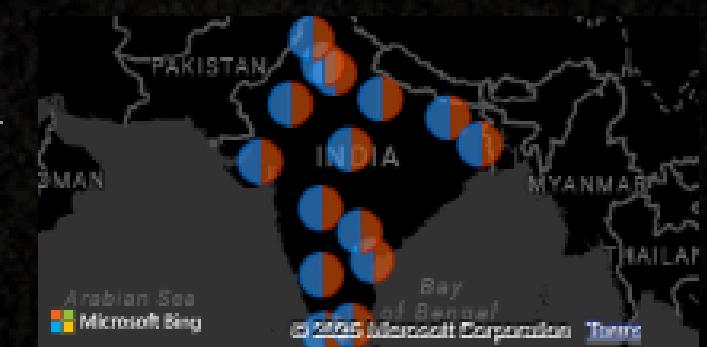
PaymentMethod Cancellation Rate countOrders Delivery Performance % Delivered Orders

| PaymentMethod | Cancellation Rate | countOrders | Delivery Performance % | Delivered Orders |
|------------------|-------------------|-------------|------------------------|------------------|
| Cash on Delivery | 0.27 | 1092 | 0.25 | 271 |
| Credit Card | 0.27 | 1071 | 0.25 | 270 |
| Debit Card | 0.27 | 1094 | 0.23 | 254 |
| Net Banking | 0.26 | 1093 | 0.23 | 255 |
| UPI | 0.24 | 1150 | 0.25 | 284 |
| Total | 0.26 | 5500 | 0.24 | 1334 |

State and PrimeMember

PrimeMember

● No
● Yes



Customer Analysis



amazon

CLV

575.68K

Avg Delviery Rating

2.97



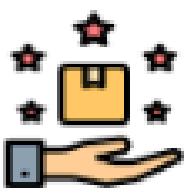
Avg Packaging Rating

3.04



Avg Prod Quality...

3.01

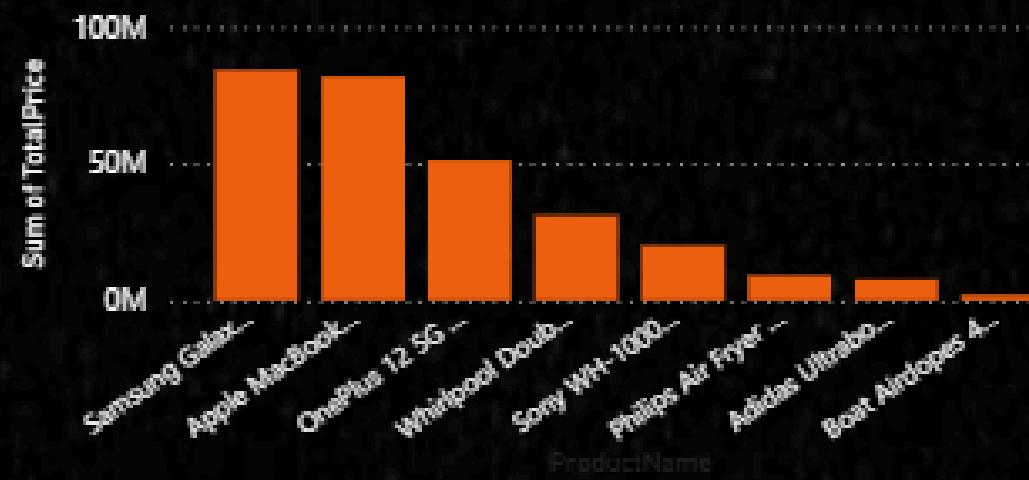


Avg Age

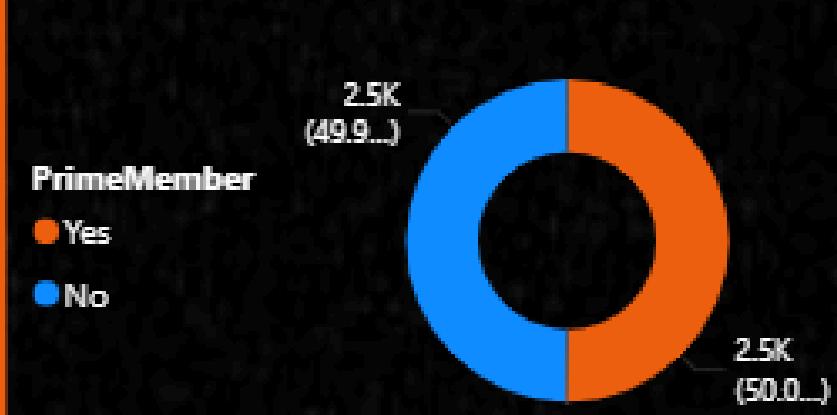
43.18



Loved Products



Count of PrimeMember by PrimeMember



CLV FirstName

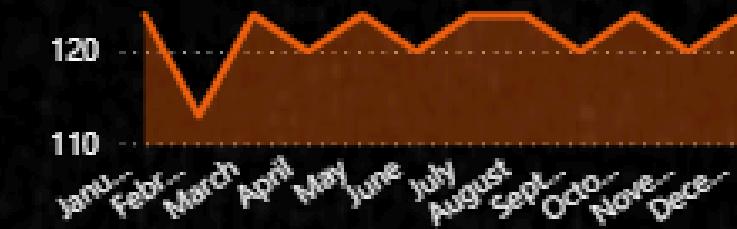
| CLV | FirstName |
|-------------|-----------|
| 7,09,447.95 | Sai |
| 6,83,948.75 | Mitali |
| 6,74,453.05 | Saanvi |
| 6,73,989.58 | Yash |
| 6,72,435.00 | Rajesh |
| 6,55,767.68 | Isha |
| 6,53,516.67 | Ananya |
| 6,50,341.12 | Kavya |
| 6,71,608.75 | |

Customer Retention Rate

0.10



New customer Acquired by Month



Product Return Rate

24.18



New customers

132



Order and Product Analysis

amazon

Total Products

30



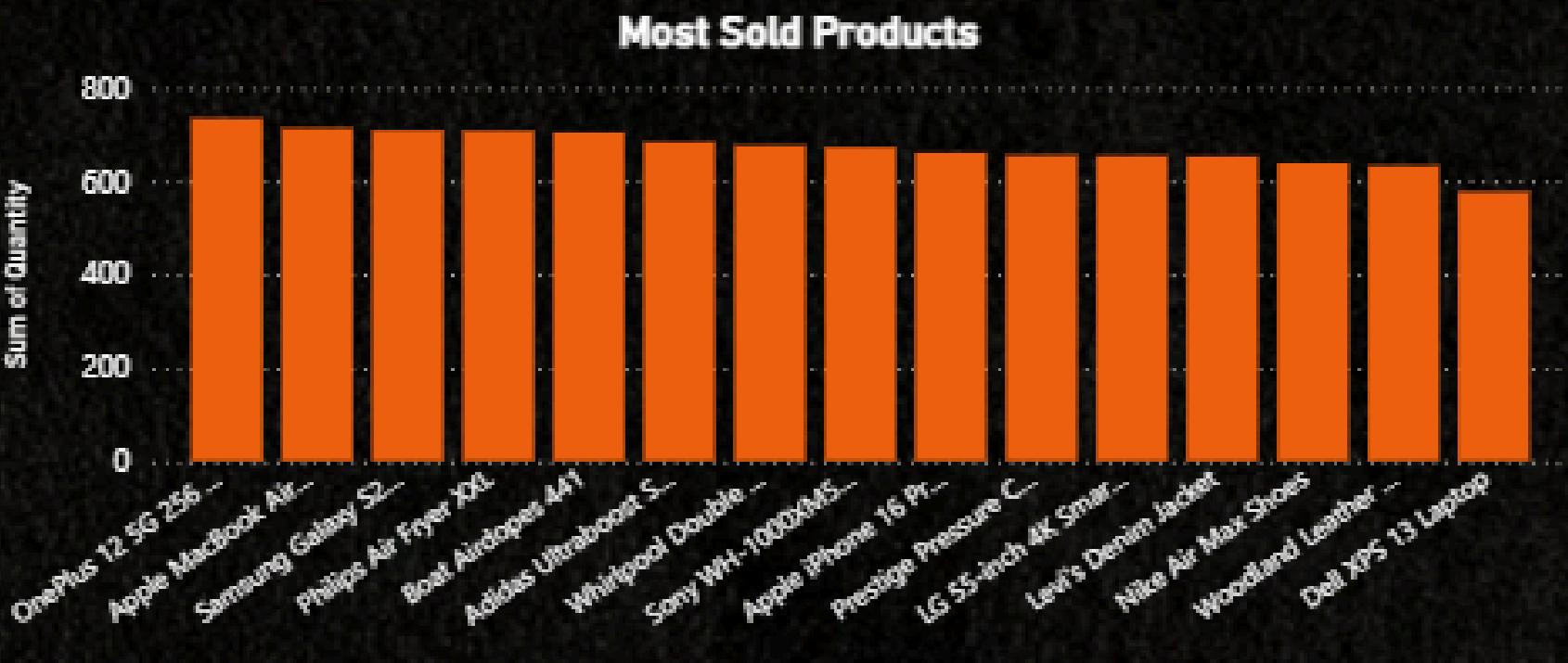
Average Shipping Time

3.18



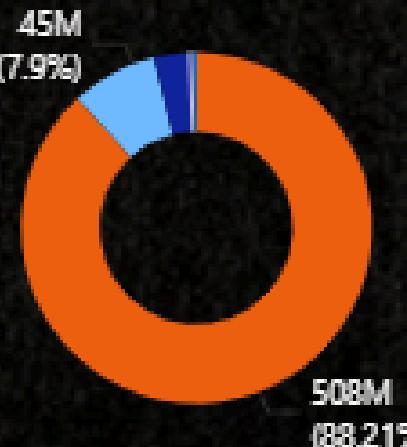
Avg Basket Size

2.10

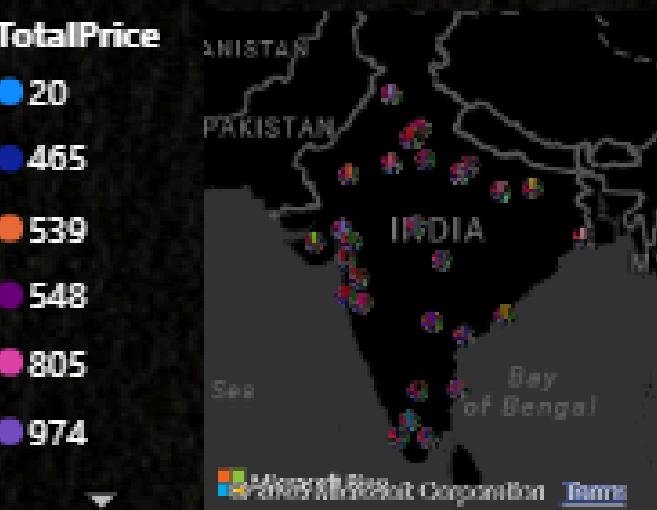


Categorized Total Price

- Category
- Electronics
- Home App...
- Fashion
- Home & K...
- Accessories



City and Total Sales



Avg Purchase Frequency

1.07



Total Quantity Sold

12K

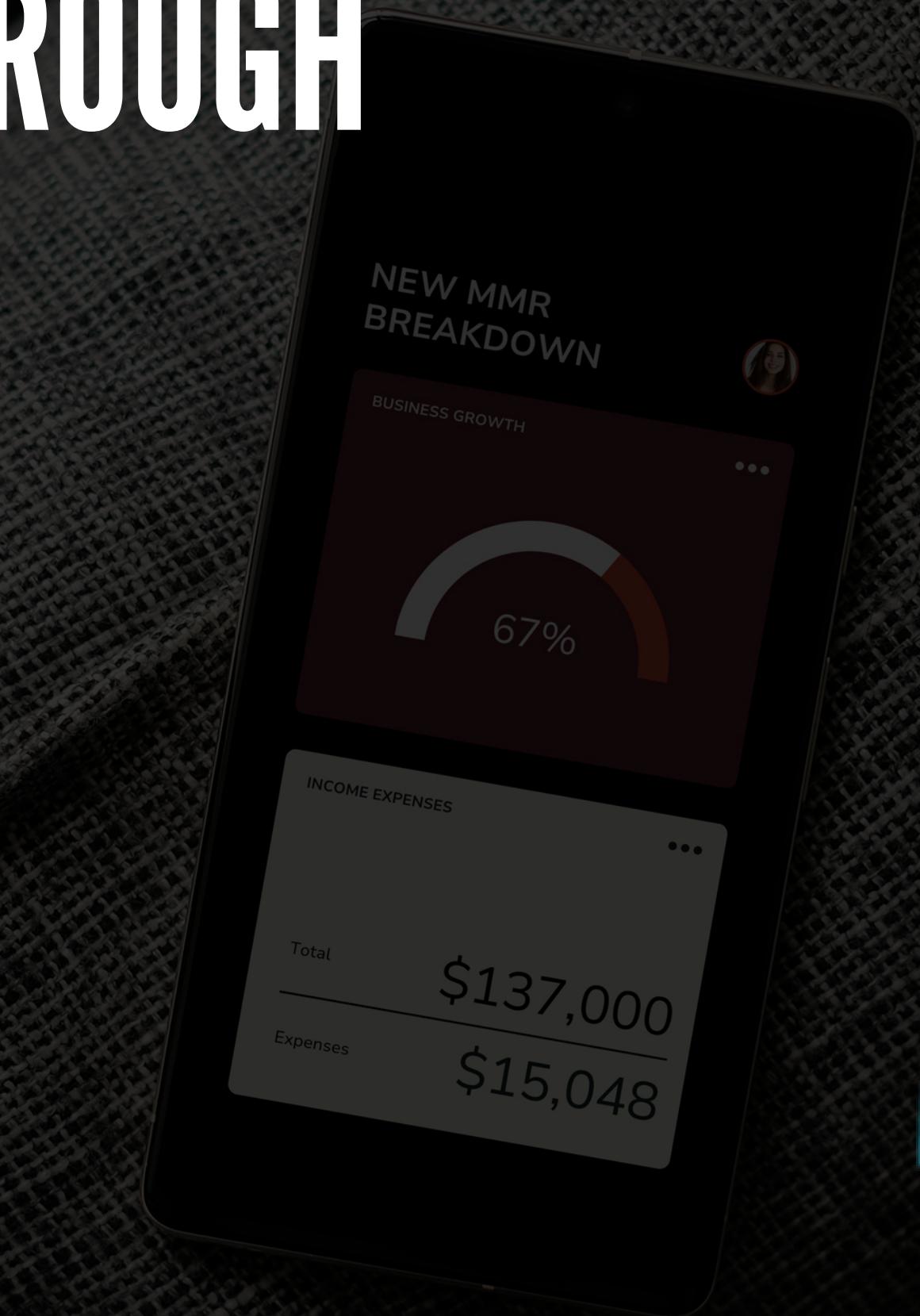


Cancellation Rate

0.26



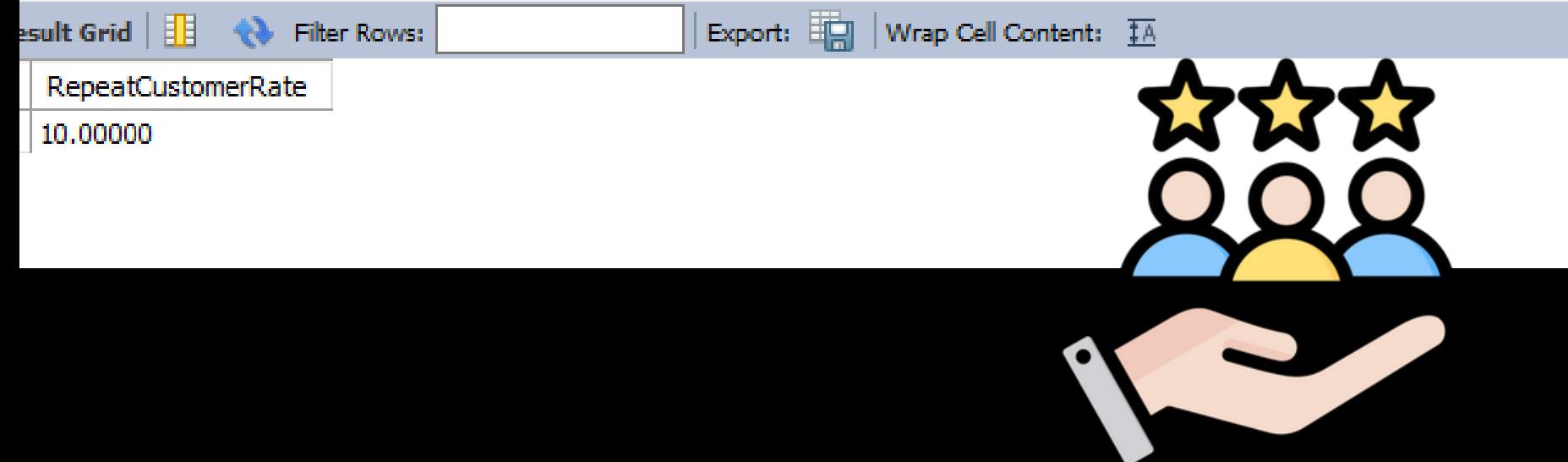
ANALYSIS THROUGH MySQL



MySQL™

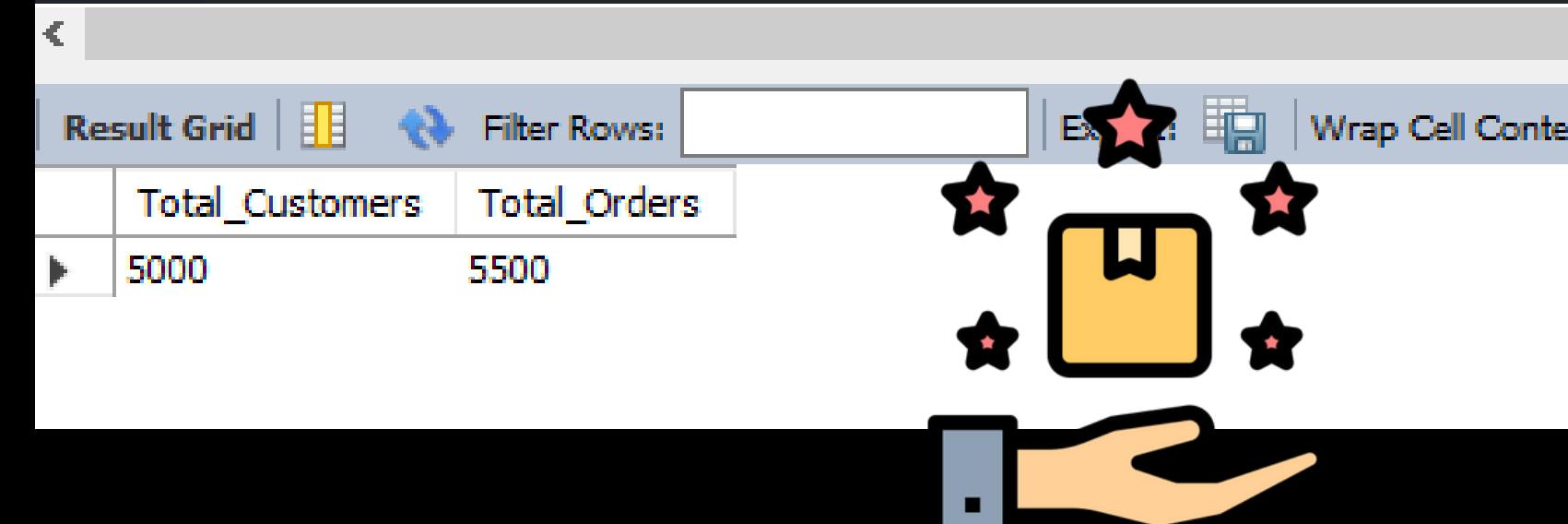
REPEAT CUSTOMER RATE

```
23  
24 •   SELECT COUNT(*) * 100.0 / (SELECT COUNT(*) FROM Customers) AS RepeatCustomerRate  
25   FROM (  
26     SELECT CustomerID  
27     FROM Orders  
28     GROUP BY CustomerID  
29     HAVING COUNT(OrderID) > 1  
30   ) t;  
31  
32
```



TOTAL CUSTOMER & ORDERS

```
11      ## Total Customers & Orders  
12 •   select count(distinct CustomerID) as Total_Customers,  
13      count(OrderID) as Total_Orders  
14      from orders;  
15  
16  
17  
18
```



MONTHLY SALES

```
32
33 *  SELECT FORMAT(OrderDate, 'yyyy-MM') AS Month,
34           SUM(od.Quantity * od.TotalPrice) AS MonthlySales
35   FROM Product o
36   JOIN orders od ON o.P_ID = od.P_ID
37 GROUP BY FORMAT(OrderDate, 'yyyy-MM')
38 ORDER BY Month;
39
```

Result Grid | Filter Rows: [] | Export: [] | Wrap Cell Content: []

| | Month | MonthlySales |
|----|----------|--------------|
| ▶ | 1 | 61449304 |
| 10 | 59194924 | |
| 11 | 49138120 | |
| 12 | 58244367 | |
| 13 | 49936953 | |
| 14 | 66655983 | |
| 15 | 52177805 | |
| 16 | 49128026 | |
| 17 | 55916014 | |
| 18 | 35033519 | |

Result 10 ×

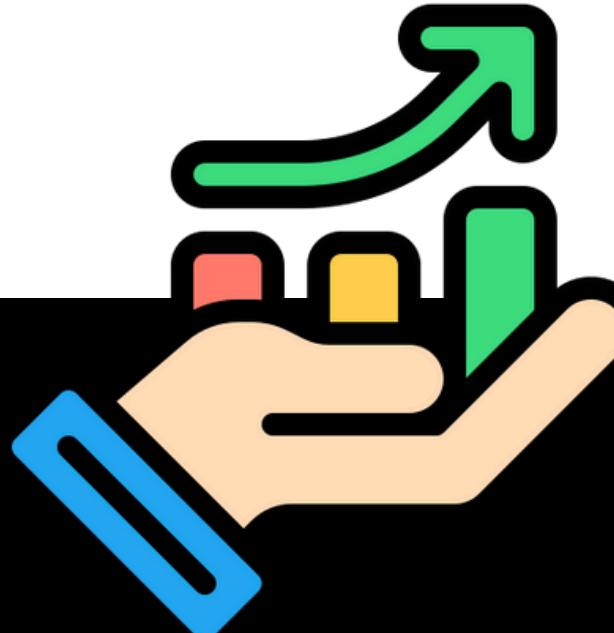


NET SALES

```
17
18      # Net Sales
19 *  select sum(TotalPrice) as Net_Sales
20      from orders;
21
22
```

Result Grid | Filter Rows: [] | Export: [] | Wrap Cell Content: []

| | Net_Sales |
|---|-----------|
| ▶ | 575678420 |



NEW CUSTOMER BY MONTH

```
94  
95 *  SELECT FORMAT(RegistrationDate, 'yyyy-MM') AS Month,  
96           COUNT(CustomerID) AS NewCustomers,  
97           SUM(COUNT(CustomerID)) OVER (ORDER BY FORMAT(RegistrationDate, 'yyyy-MM')) AS RunningTotalCustomers  
98     FROM Customers  
99   GROUP BY FORMAT(RegistrationDate, 'yyyy-MM')  
100  ORDER BY Month;  
101
```

Result Grid | Filter Rows: [] | Export: [] | Wrap Cell Content: []

| | Month | NewCustomers | RunningTotalCustomers |
|----|-------|--------------|-----------------------|
| 1 | 176 | 176 | |
| 10 | 146 | 322 | |
| 11 | 151 | 473 | |
| 12 | 174 | 647 | |
| 13 | 170 | 817 | |
| 14 | 157 | 974 | |
| 15 | 169 | 1143 | |
| 16 | 168 | 1311 | |
| 17 | 161 | 1472 | |
| 18 | 176 | 1648 | |
| 19 | 166 | 1814 | |
| 2 | 178 | 1992 | |
| 20 | 169 | 2161 | |
| 21 | 173 | 2334 | |
| 22 | 157 | 2491 | |

Result 18 x



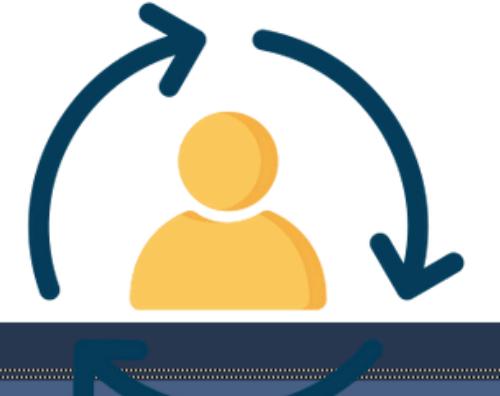
REPEAT CUSOMERS

```
78 *  WITH CustomerOrders AS (  
79       SELECT  
80             CustomerID,  
81             COUNT(OrderID) AS OrderCount,  
82             CASE WHEN COUNT(OrderID) > 1 THEN 1 ELSE 0 END AS IsRepeatCustomer  
83       FROM Orders  
84       GROUP BY CustomerID  
85     )  
86     SELECT  
87       CustomerID,  
88       OrderCount,  
89       IsRepeatCustomer,  
90       (SUM(IsRepeatCustomer) OVER () * 100.0 / COUNT(*) OVER ()) AS RepeatCustomerRate  
91     FROM CustomerOrders;
```

Result Grid | Filter Rows: [] | Export: [] | Wrap Cell Content: [] | Fetch rows: []

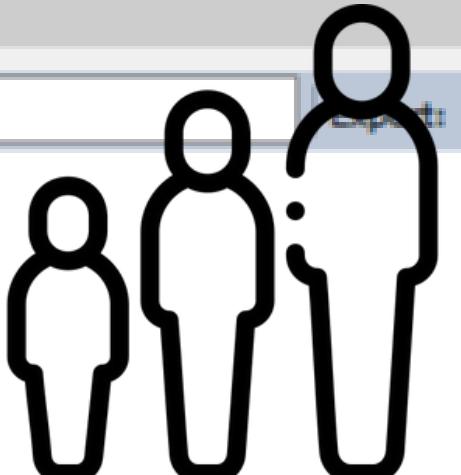
| | CustomerID | OrderCount | IsRepeatCustomer | RepeatCustomerRate |
|---|------------|------------|------------------|--------------------|
| 1 | CUST00001 | 1 | 0 | 10.00000 |
| 2 | CUST00002 | 1 | 0 | 10.00000 |
| 3 | CUST00003 | 1 | 0 | 10.00000 |
| 4 | CUST00004 | 1 | 0 | 10.00000 |
| 5 | CUST00005 | 1 | 0 | 10.00000 |
| 6 | CUST00006 | 1 | 0 | 10.00000 |
| 7 | CUST00007 | 1 | 0 | 10.00000 |

Result 17 x



AVERAGE AGE

```
71  
72      #Average Age  
73 •  SELECT AVG(Age) AS AvgAge  
74      FROM customers;  
75  
76
```

Result Grid | Filter Rows: 

| AvgAge |
|---------|
| 43.1772 |

AVERAGE RATING

```
64      # Average Delivery Rating / Packaging Rating / Product Quality Rating  
65 •  SELECT AVG(DeliveryRating) AS AvgDeliveryRating,  
66                  AVG(PackagingRating) AS AvgPackagingRating,  
67                  AVG(ProductQualityRating) AS AvgProductQualityRating  
68      FROM Reviews;
```

Result Grid | Filter Rows: Export: Wrap Cell Content: 

| AvgDeliveryRating | AvgPackagingRating | AvgProductQualityRating |
|-------------------|--------------------|-------------------------|
| 2.9698 | 3.0416 | 3.0108 |

CUSTOMER LIFE TIME VALUE

```
57  
58      # CUSTOMER LIFETIME VALUE  
59 •  select CustomerID,sum(Quantity * Price) as CLV  
60      from Product  
61      group by CustomerID;
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content: | Fetch

| | CustomerID | CLV |
|---|------------|-------|
| ▶ | CUST00001 | 32997 |
| | CUST00002 | 62999 |
| | CUST00003 | 91998 |
| | CUST00004 | 10998 |
| | CUST00005 | 59998 |
| | CUST00006 | 12999 |
| | CUST00007 | 5499 |
| | CUST00008 | 25998 |
| | CUST00009 | 3998 |
| | CUST00010 | 3998 |

Result 14 ×



PAYMENT WISE TOTAL SALES

```
41  
42      #Sales by Payment Method  
43 •  SELECT PaymentMethod,  
44                  SUM(Quantity * TotalPrice) AS TotalSales  
45      FROM Product  
46      GROUP BY PaymentMethod;
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

| | PaymentMethod | TotalSales |
|---|------------------|------------|
| ▶ | UPI | 301055055 |
| | Debit Card | 281119341 |
| | Net Banking | 301198866 |
| | Credit Card | 285448279 |
| | Cash on Delivery | 290906155 |



NET SALES \$

```
17  
18      # Net Sales  
19 •  select sum(TotalPrice) as Net_Sales  
20      from orders;  
21  
22
```

| | Net_Sales |
|---|-----------|
| ▶ | 575678420 |

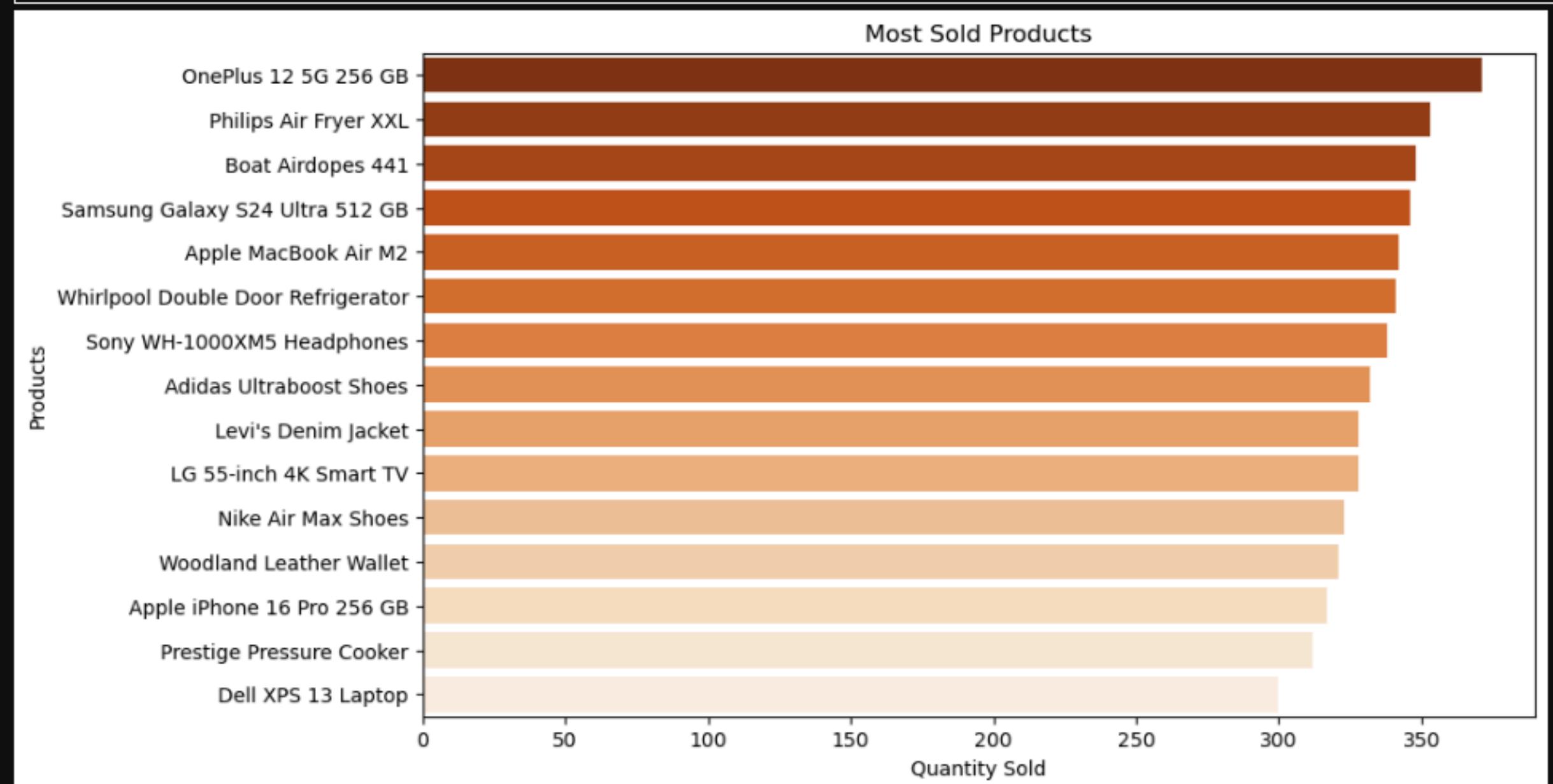


ANALYSIS THROUGH PYTHON



Most Soled Product

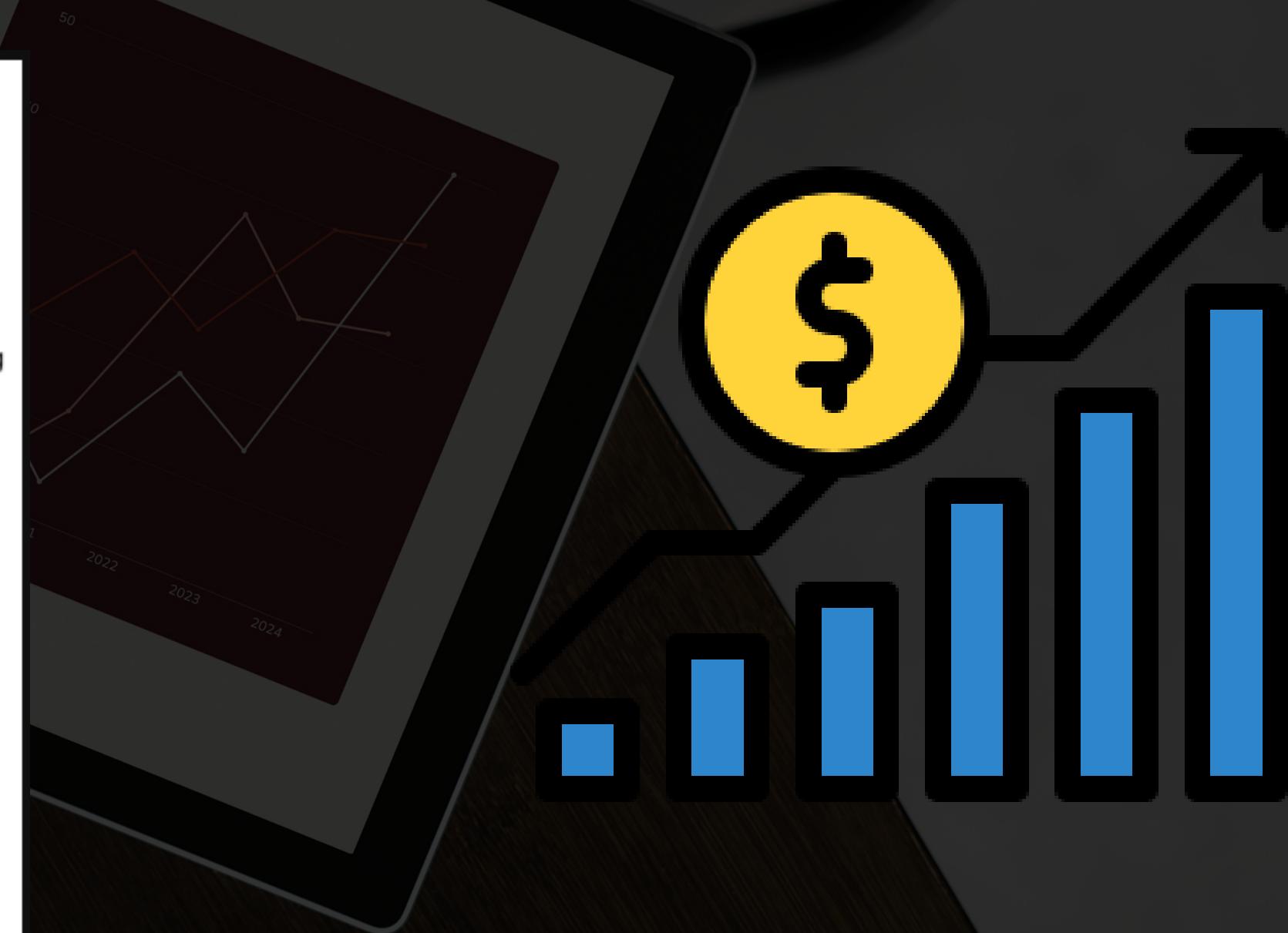
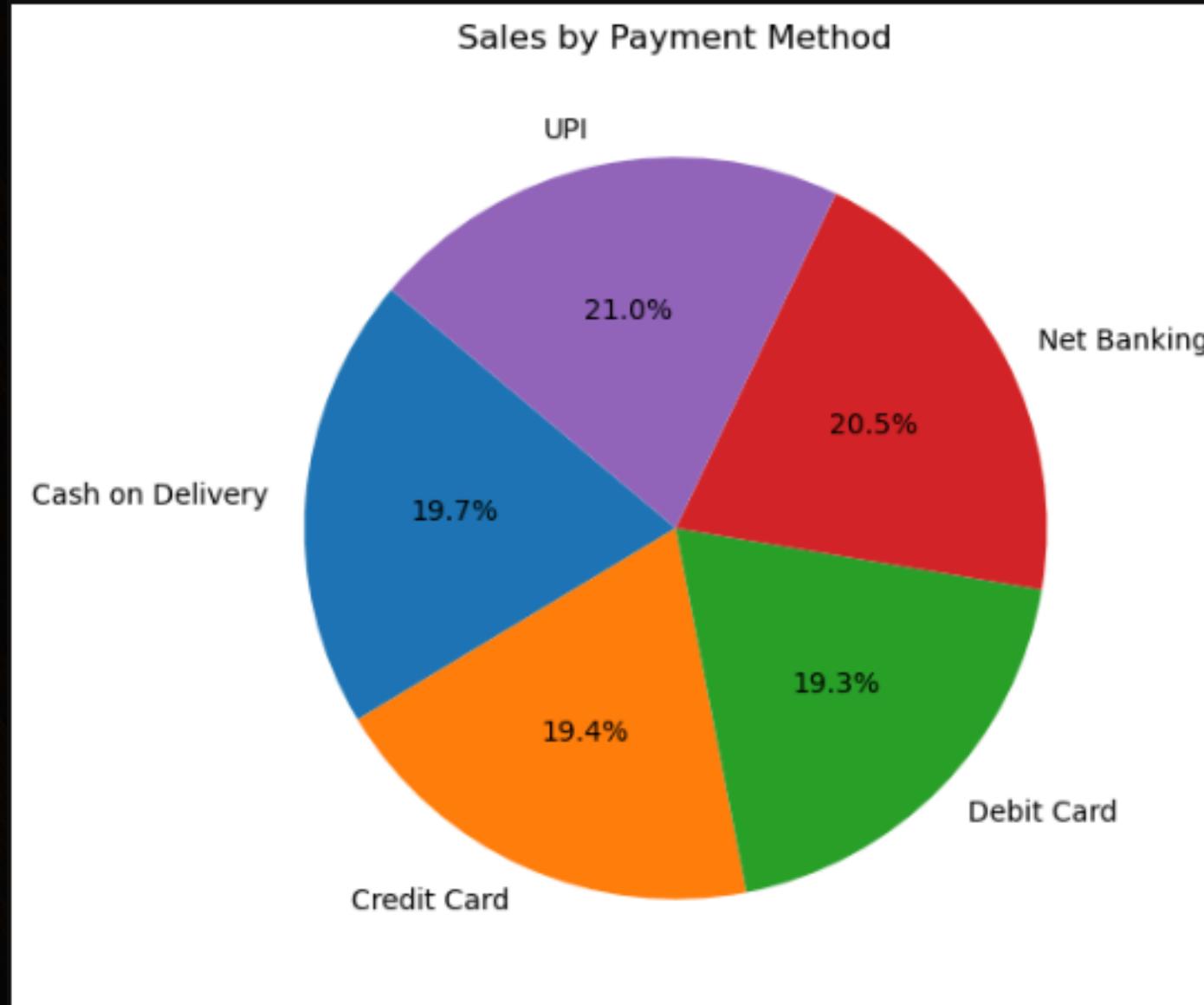
```
plt.figure(figsize=(10,6))
top_products = product['ProductName'].value_counts().head(15)
sns.barplot(x=top_products.values, y=top_products.index, palette="Oranges_r")
plt.title("Most Sold Products")
plt.xlabel("Quantity Sold")
plt.ylabel("Products")
plt.show()
```



Payment Method Share

```
payment_sales = df.groupby("PaymentMethod")["TotalPrice"].sum()

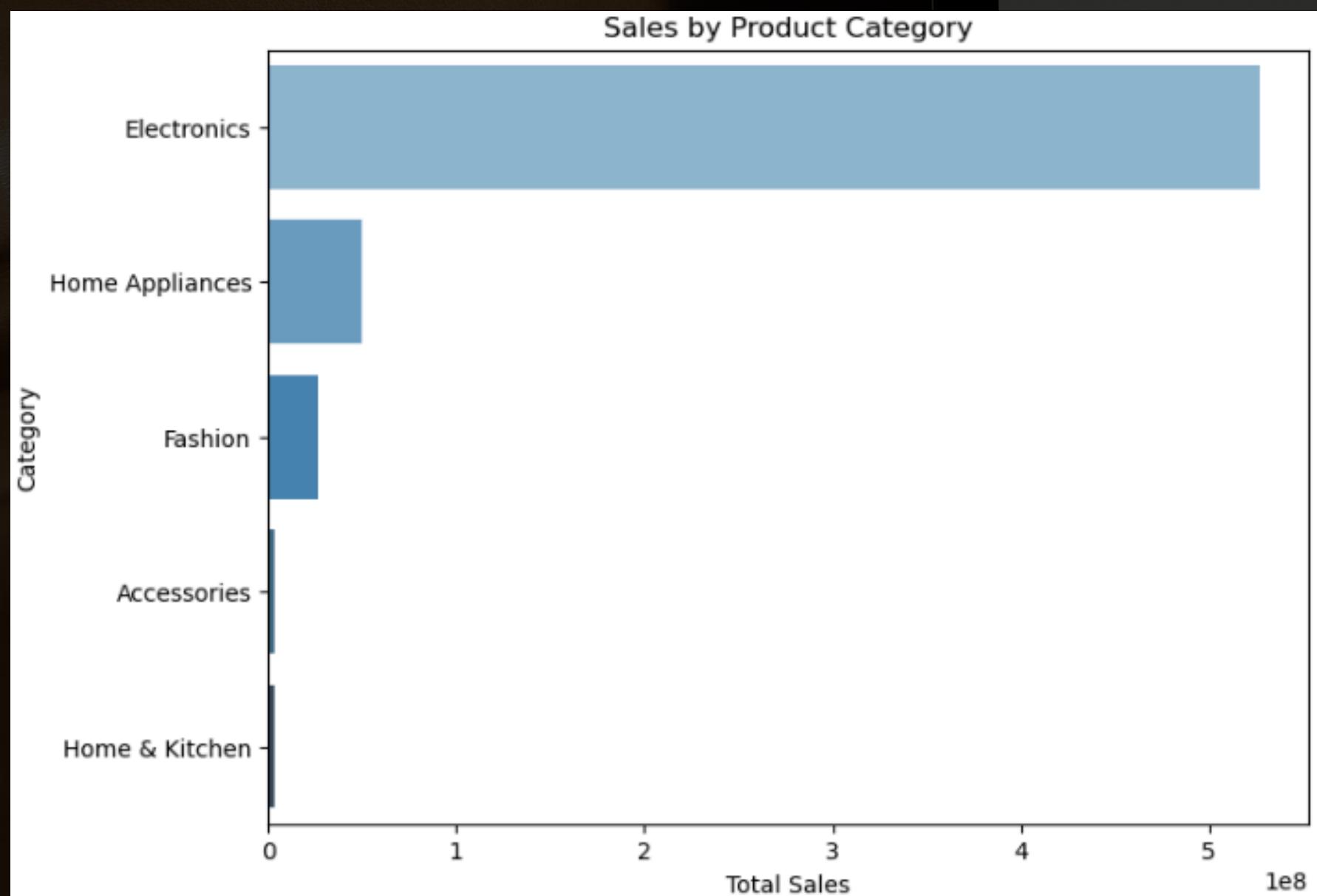
plt.figure(figsize=(6,6))
plt.pie(payment_sales, labels=payment_sales.index, autopct='%1.1f%%', startangle=140)
plt.title("Sales by Payment Method")
plt.show()
```



Sales by Category

```
category_sales = df.groupby("Category")["TotalPrice"].sum().sort_values(ascending=False)

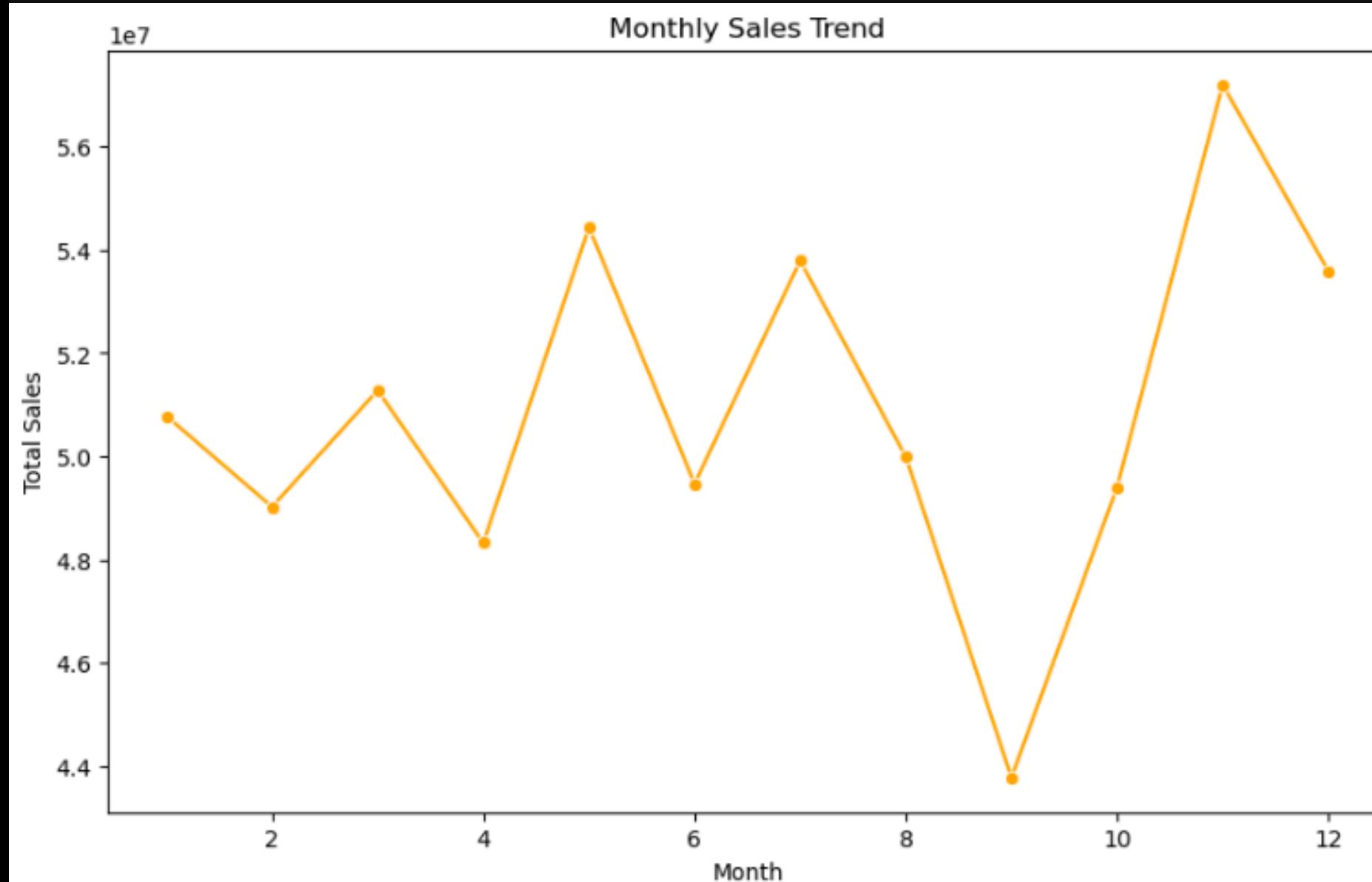
plt.figure(figsize=(8,6))
sns.barplot(x=category_sales.values, y=category_sales.index, palette="Blues_d")
plt.title("Sales by Product Category")
plt.xlabel("Total Sales")
plt.ylabel("Category")
plt.show()
```



Monthly Sales Trend

```
df[ "OrderDate" ] = pd.to_datetime(df[ "OrderDate" ], format="%d-%m-%Y")
monthly_sales = df.groupby(df[ "OrderDate" ].dt.month)[ "TotalPrice" ].sum()

plt.figure(figsize=(10,6))
sns.lineplot(x=monthly_sales.index, y=monthly_sales.values, marker="o", color="orange")
plt.title("Monthly Sales Trend")
plt.xlabel("Month")
plt.ylabel("Total Sales")
plt.show()
```



Customer Experience Rating

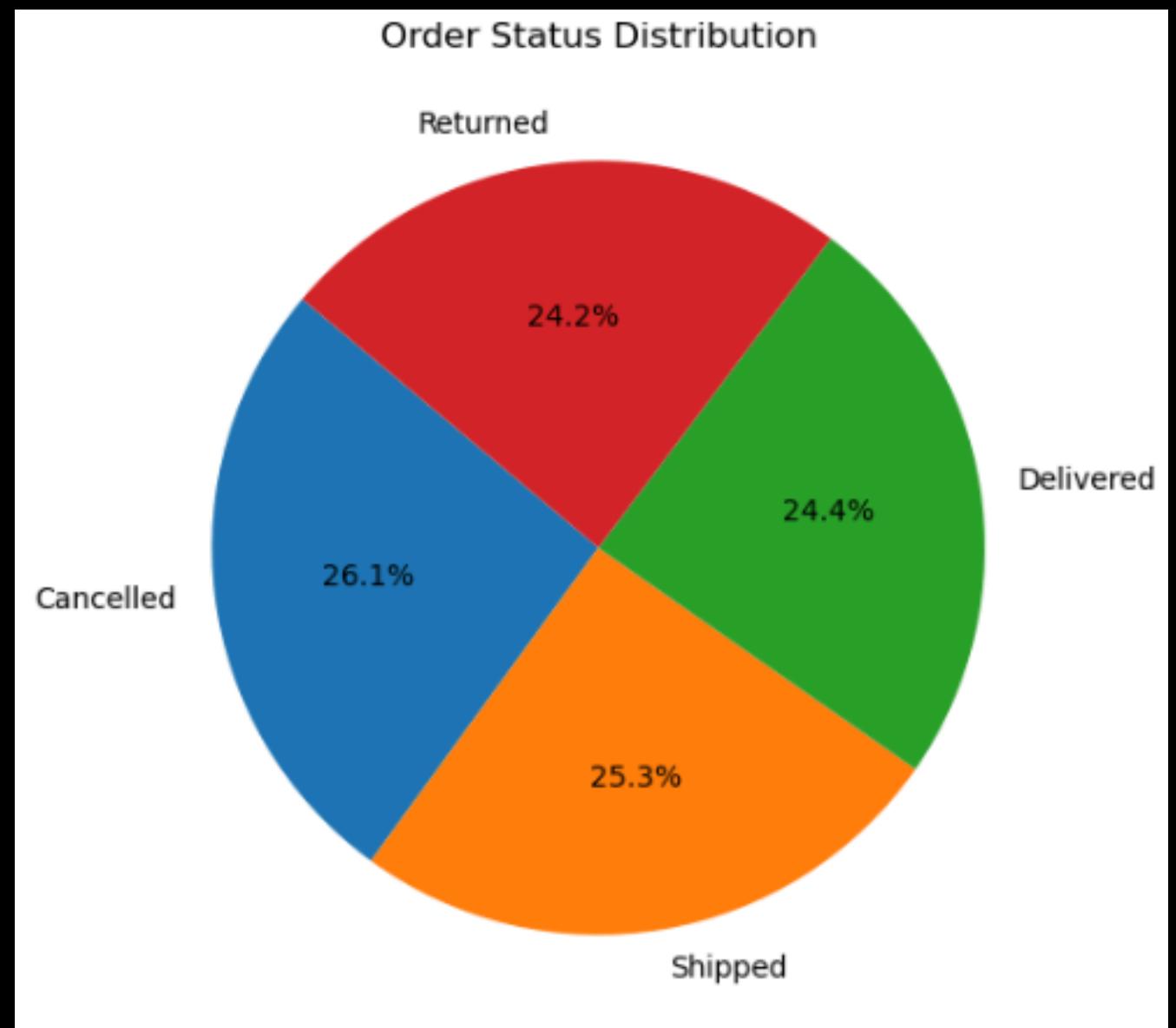
```
ratings = df[['ProductQualityRating', 'DeliveryRating', 'PackagingRating']]  
  
plt.figure(figsize=(8,5))  
sns.boxplot(data=ratings, palette="Set2")  
plt.title("Distribution of Customer Ratings")  
plt.show()
```



Return and Cancellation Rates

```
order_status_counts = df['OrderStatus'].value_counts(normalize=True) * 100

plt.figure(figsize=(6,6))
plt.pie(order_status_counts, labels=order_status_counts.index, autopct='%1.1f%%', startangle=140)
plt.title("Order Status Distribution")
plt.show()
```



GROUND

THE INDUSTRY'S HISTORY

THANK YOU

FOR YOUR ATTENTION

