Exploratory Data Analysis (EDA) with Pandas in E-commerce

The purpose of this project is to explore and analyze an e-commerce dataset using the Pandas framework to derive insights into customer behaviour, product trends, and sales performance.

Goals of the Project -

- Explore the Amazon dataset using Pandas.

- Compare how different payment methods impact total sales.
- Study customer behaviour, such as their ratings and purchasing habits.
- Create visual charts to show sales trends, customer ratings, and product performance.

Materials and Methods -

The data for this project is sourced from an e-commerce platform dataset, containing details about customer orders, products, payment methods, shipping, and ratings. It includes sales figures, product categories, order dates, customer satisfaction, and shipping costs. The project aims to analyze sales trends, customer purchasing behaviour, the impact of shipping delays, and the performance of different payment methods.

General Part –

● Libraries Import: Pandas, NumPy, Seaborn, Matplotlib

● **Dataset Exploration**: checking for missing values, duplicates, and generating summary statistics.

● **Feature Engineering**: Transformation of date columns and creation of new features like shipping delay and profit margin.

● **Visualization in Pandas**: Distribution analysis, relationships between variables, and time-based trends.

Project Outcome & Insights -

The project performs **Exploratory Data Analysis (EDA)** on an **e-commerce dataset** to gain meaningful insights into **sales performance, customer behaviour, and shipping efficiency**. Below are the key outcomes:

1. **Sales Performance-**

Sales by Region: The dataset reveals total sales by country, allowing businesses to identify key markets and regions contributing to their revenue.

Top Product Categories: The analysis highlights the most profitable product categories based on total sales and revenue, helping businesses focus on high-performing products.

Monthly Shipping Cost: A breakdown of shipping costs by month is provided, enabling better tracking of shipping expenses and budget planning.

2. **Customer Behaviour Analysis -**

Returning Customers: The analysis identifies patterns in returning customers, offering insights into customer retention and loyalty.

Top Spending Customers: Recognizing the highest spending customers helps businesses tailor marketing strategies and provide personalized offers.

3. **Shipping Performance -**

Shipping Delay Categorization: The project categorizes shipments based on delivery speed, such as Same Day, Fast, Moderate, and Delayed, which helps in improving logistics management.

Profitability by Shipping Method: Analysing which shipping methods are most profitable reveals insights into the most cost-effective ways to ship products.

4. **Profitability & Business Growth -**

Profit Margin Analysis: The project calculates profitability per order, providing actionable insights to optimize profit margins.

Year-over-Year Sales Growth: Tracks the annual sales growth rate, helping businesses in financial planning and setting future sales targets.

Feature Engineering:

Created new columns such as:-

- **Shipping_Date_year**: Extracts the year from Shipping_Date.
- **Shipping_order_month**: Extracts the month from Shipping_Date.
- **Order_Type**: Categorizes orders as "Expensive" or "Cheap" based on whether Total_Cost is above or below the mean.
- **Customer_Satisfaction**: Categorizes customers as "High" or "Low" satisfaction based on their rating.

Key Questions and Insights to be Addressed:

1.What is the total sales volume by category?

```
sales_by_category=df.groupby('Category')['Total_Cost'].sum().sort_values(ascending=False)
print("\nSales by Category:\n", sales_by_category)
```

Ans-

Category-

Electronics   77075.712

Books   75956.111

Beauty   71768.725

Clothing   71438.115

Sports   67245.073

Home & Kitchen   51315.992

2. What is the monthly profit trend?

ANS-monthly_sales=df.groupby(['order_year', 'order_month'])['sales_per_order'].sum().reset_index()
print("Monthly Sales Trend:\n", monthly_sales)

Monthly Profit Analysis:

Shipping_Date   Profit

0   2025-01 19093.184659

1   2025-02   1444.180875

3.Which country has the highest sales?

Ans-

df_top_countries_data=df.groupby('Country')['Total_Sales'].sum().sort_values(ascending=False).head(5).reset_index()

2 Total sales in Top 5 Countries

       Country   Total Sales

0   Canada 81494.258357

1   India 79401.207249

2   Germany 64464.411515

3 Australia 64226.364820

   4   UK 59645.617653

4.How do satisfaction levels vary depending on the product category?

Ans- df['Customer_Satisfaction'].value_counts()

5. How does final revenue vary by product category?
   ANS-
   product_wise_final_revenue=df.groupby('Product_Name')['Final_Rev
   enue'].sum().sort_values(ascending=False).head(10)
   print("\n:product wise final revenue\n",product_wise_final_revenue)

   1 Final Revenue of products
    Product_Name
   Laptop   38770.870333
   Jacket   30686.012477
   Smartphone   26115.586542
   Smartwatch   25442.264315
   Football   23237.125506
   Science Textbook   22614.409773
   Self-Help   21845.647118
   Face Cream   18548.092336
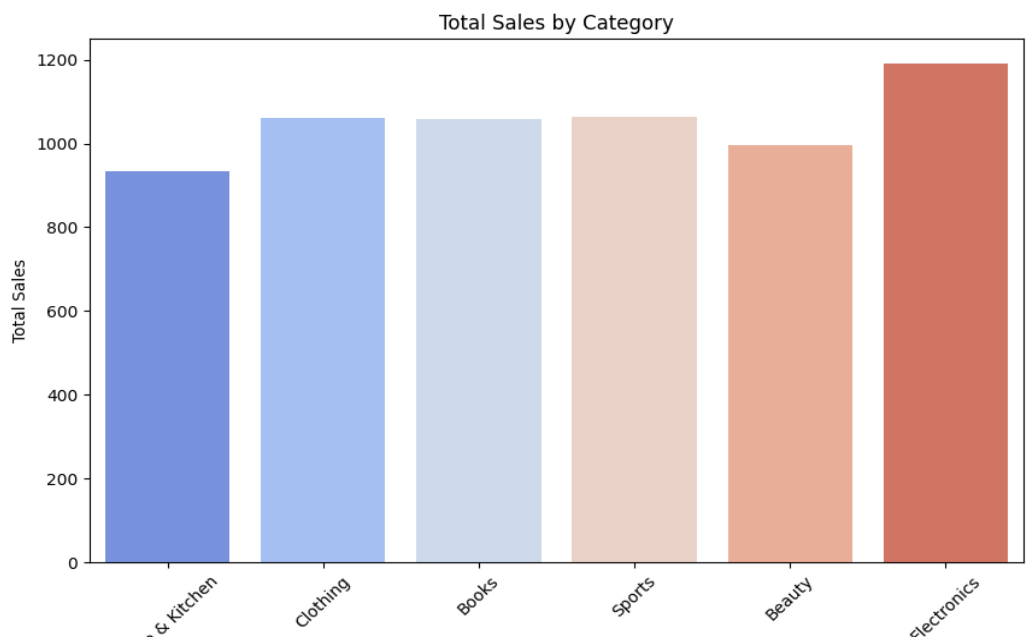   Lipstick   17490.996628
   Fiction Novel   16944.372120

   Visualization:
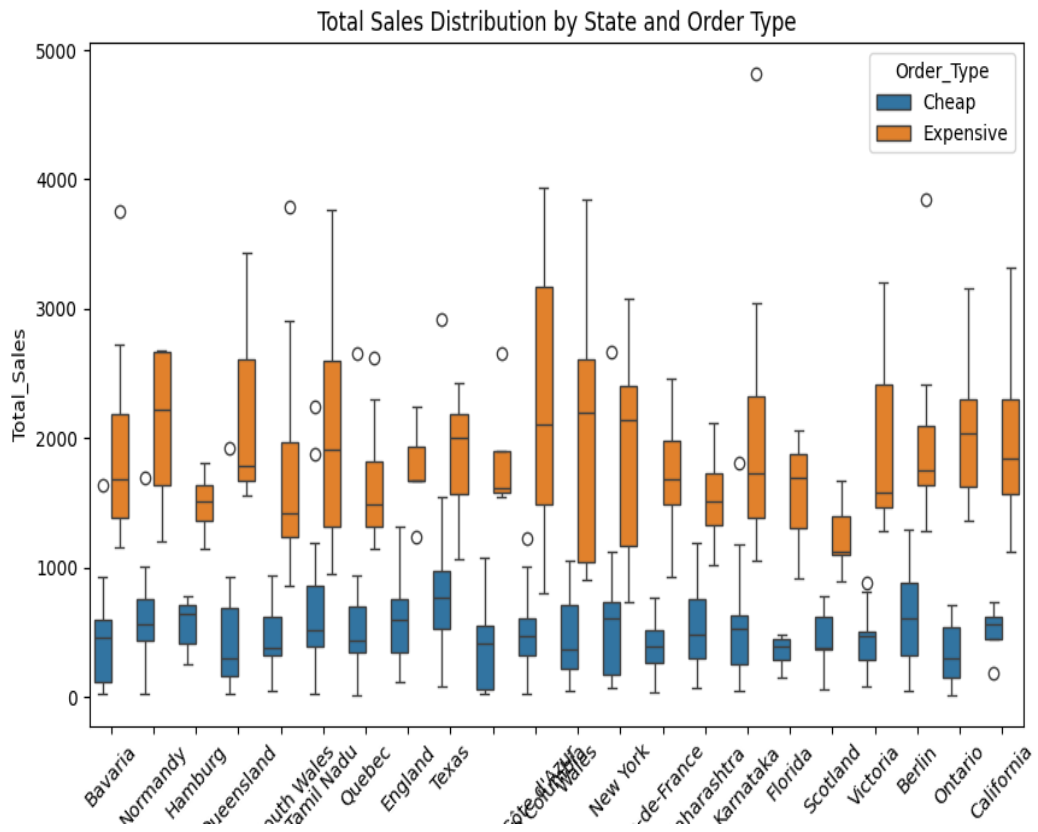   Several charts created to present inside including:-
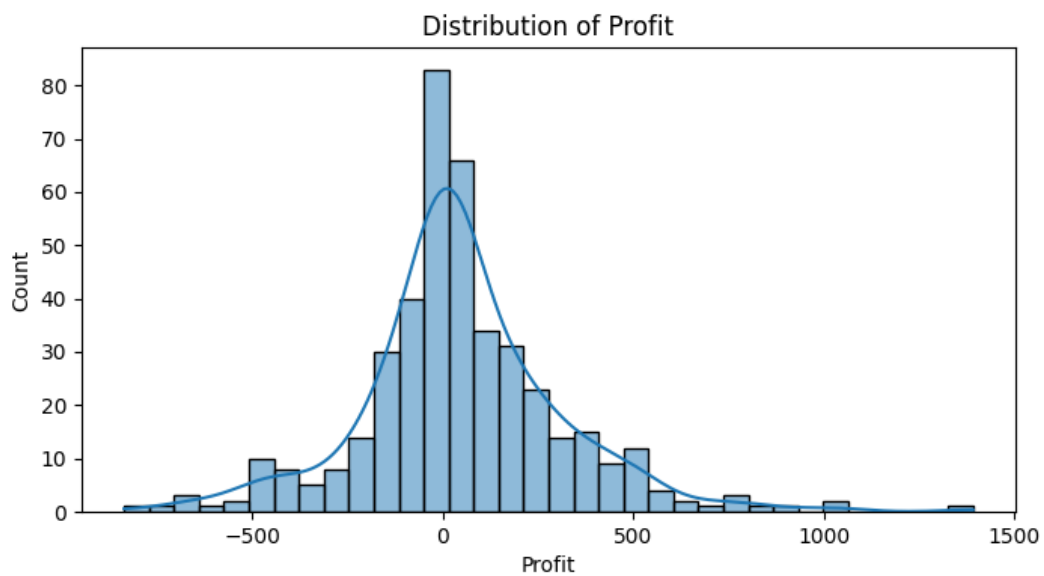
- Count of payment Method

Count of Payment Methods

- Total Sales By Category-
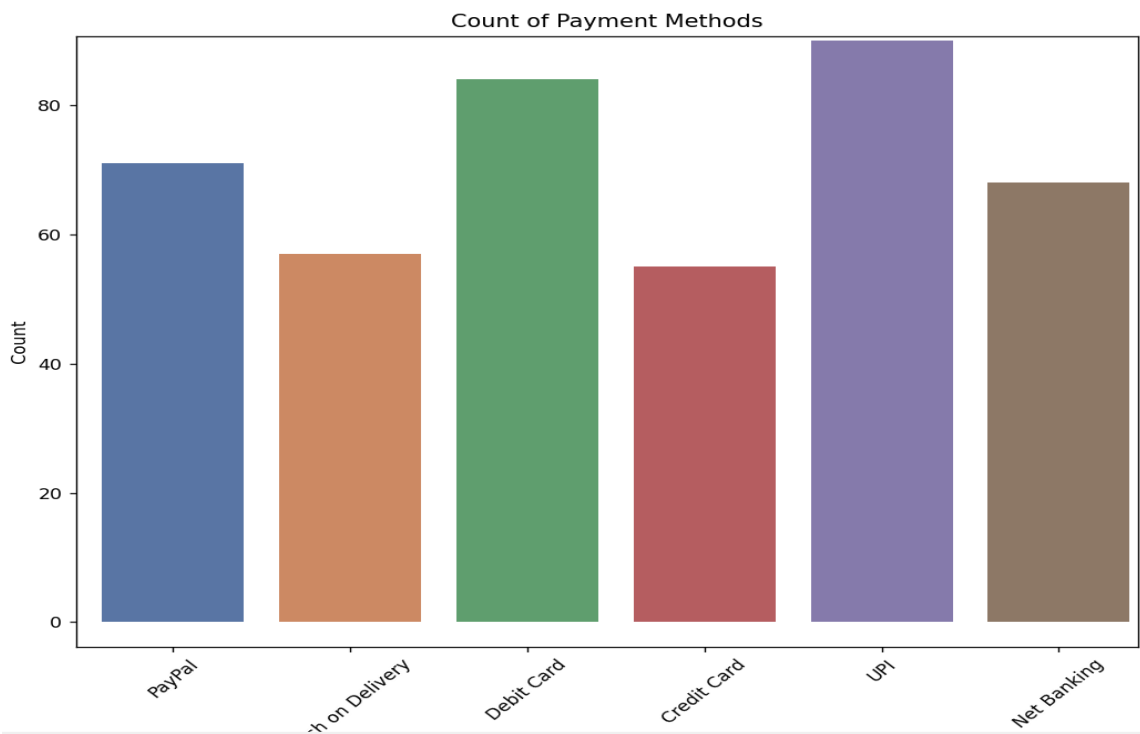
Total Sales by Category

- Showing box plot for Total sales Distribution by state & order type

Total Sales Distribution by State and Order Type

- Distribution Of profit-

Distribution of Profit

- Distribution of price-