📦E-Commerce Sales Analysis (Python)

📌 Project Overview

This project performs an end-to-end exploratory data analysis (EDA) on a realistic e-commerce sales dataset to uncover key business insights related to sales performance, customer behavior, product trends, and regional contribution.

The analysis follows an industry-style workflow starting from data loading and cleaning to advanced customer analytics and visualization.

🎯 Objectives

- ★ Analyze overall sales performance and key KPIs
- ★ Identify time-based sales trends (monthly & quarterly)
- ★ Evaluate product-wise and category-wise performance
- ★ Understand customer purchasing behavior
- ★ Identify high-value customers using revenue contribution
- ★ Apply Pareto (80/20) analysis for business insights

🛠️ Tools & Technologies

- ★ **Language:** Python
- ★ **Environment:** Google Colab
- ★ **Libraries:**
    - ○ pandas
    - ○ numpy
    - ○ matplotlib
    - ○ seaborn

📊 Dataset

- ★ **Type:** Synthetic but realistic e-commerce transaction data
- ★ **Records:** ~5,000 transactions
- ★ **Key Features:**
  - ○ Order date
  - ○ Customer ID & segment
  - ○ Product category & product name
  - ○ Quantity & unit price
  - ○ Sales amount
  - ○ Region & payment mode

The dataset was programmatically generated to closely resemble real-world e-commerce data and ensure full control over data quality and structure.

🔍 Analysis Performed

1. 🧹 Data Cleaning & Preprocessing

- ❖ Converted date columns to datetime format
- ❖ Verified missing values and duplicates
- ❖ Performed sanity checks on quantity and pricing
- ❖ Engineered time-based features (month, quarter, year)

2. 🧑‍💻 Exploratory Data Analysis (EDA)

- ❖ Overall KPIs (Total Sales, Orders, Customers, AOV)
- ❖ Monthly and quarterly sales trends
- ❖ Product-wise and category-wise sales analysis
- ❖ Region-wise revenue contribution

3. 👥 Customer Behavior Analysis

- ❖ Orders per customer distribution
- ❖ Average Order Value (AOV) distribution
- ❖ Customer segment performance

4. 🤖 Advanced Analysis

- ❖ Top 10 customers by revenue
- ❖ Pareto (80/20) revenue analysis
- ❖ Revenue concentration insights

📈 Key Insights

★ Sales show clear seasonality with noticeable quarterly fluctuations
★ A small group of customers contributes a large share of total revenue
★ Certain product categories consistently outperform others
★ Majority of customers are low-frequency buyers, indicating strong retention opportunities

---

📤 Deliverables

★ Quarterly sales summary (CSV)
★ Product performance report (CSV)
★ Region-wise sales report (CSV)
★ Top customers by revenue (CSV)

📌 Conclusion

This project demonstrates how data analysis can translate raw transactional data into actionable business insights.
The findings highlight the importance of customer retention, product optimization, and targeted marketing strategies in e-commerce businesses.

🚀 Future Enhancements

★ Customer segmentation using RFM analysis
★ Cohort analysis for retention tracking
★ Interactive dashboards (Power BI / Tableau)

☁️ Execution Environment

The entire project was developed and executed using **Google Colab**, leveraging its cloud-based environment for efficient experimentation, visualization, and reproducibility.

🧑🏻 Author

Anirudha Das
Aspiring Data Analyst / Business Intelligence Professional
📍 Bolpur, WB, India