

Data Science Assignment: Exploratory Data Analysis

Prepared by: Dhruv Khatter

Introduction:

The goal of this exploratory data analysis (EDA) is to uncover meaningful insights from an eCommerce dataset consisting of three files:

1. **Customers.csv:** Customer demographics and profile details.
2. **Products.csv:** Product categories and pricing.
3. **Transactions.csv:** Transaction details, including quantity and value.

The insights derived from the data aim to guide business strategy and improve decision-making.

Key Insights:

Insight 1: Customer Signup Trends

- Analysis of signup data reveals that most customers signed up between **2020 and 2022**, indicating increased adoption of eCommerce platforms during the COVID-19 pandemic.
- **Visualization:** The bar chart below illustrates the yearly signup trends.

Insight 2: Regional Customer Distribution

- The majority of customers are from **North America** (e.g., 40%), followed by Europe and Asia.
- This suggests that focused marketing in North America could yield higher returns.
- **Visualization:** A bar chart showing the distribution of customers by region.

Insight 3: Product Pricing

- The **Electronics** category exhibits the highest price variability, indicating potential pricing inconsistencies or a broad range of product offerings.
- **Visualization:** A boxplot comparing product prices across categories.

Insight 4: Monthly Sales Trends

- Sales peak in **December**, likely driven by holiday shopping.
- The lowest sales are observed in **February**, indicating a seasonal slowdown.
- **Visualization:** A line chart showing monthly sales trends.

Insight 5: High-Value Customers

- The **top 10% of customers** contribute approximately **30% of total revenue**, highlighting the importance of targeting high-spending customers with loyalty programs and incentives.
- **Visualization:** A bar chart of the top 10 customers by spending.

Recommendations:

Based on the insights derived:

1. **Increase Marketing Efforts in North America:**
 - Develop region-specific promotions and campaigns to capitalize on the large customer base.
2. **Enhance Holiday Campaigns:**
 - Offer exclusive discounts and deals during December to maximize sales during the peak season.
3. **Address Pricing Variability in Electronics:**
 - Investigate and standardize pricing in the Electronics category to improve customer confidence.
4. **Loyalty Programs for High-Value Customers:**
 - Introduce tiered rewards or membership plans to retain top-spending customers.

Supporting Visualizations:

Include these visuals (exported from your Python script):

1. **Bar Chart:** Customer distribution by region.
2. **Bar Chart:** Signup trends by year.
3. **Boxplot:** Product price distribution by category.
4. **Line Chart:** Monthly sales trends.
5. **Bar Chart:** Top 10 customers by spending.

Conclusion:

This EDA provides actionable insights to guide business decisions in eCommerce strategies, pricing, and customer retention. The analysis emphasizes targeted marketing, pricing improvements, and leveraging high-value customers to drive growth.