Shopping Trends Analysis

Overview

This project analyzes customer shopping data to uncover insights about purchasing behavior, retention, and high-value customers. It demonstrates optimized data preprocessing, exploratory data analysis (EDA), and dashboard creation using Python, Pandas, Matplotlib, and Seaborn.

The analysis supports business decisions for targeting campaigns, improving retention, and tracking key performance metrics.

Key Features

Data Preprocessing & EDA

- Cleaned and explored data for 100+ customer transactions.
- Handled age segmentation, categories, subscription status, payment methods, and promo codes.
- Visualized distributions, averages, and trends.

Targeting High-Value Customers

- Identified the **top 20% of customers** by total spend.
- Calculated targeting precision: percentage of revenue contributed by high-value customers.
- Insights can guide focused marketing campaigns to improve revenue efficiency.

Retention Analysis

- Calculated repeat purchase rates for all customers.
- Compared subscribed vs non-subscribed customers.
- Provides actionable metrics to improve retention strategies.

Dashboards & Visualizations

Bar charts, line charts, and pie charts for:

- Age distribution
- Average purchase by category
- Purchases by gender
- Monthly spending trends
- Popular items, payment methods, and promo code usage
- High-value customer tracking

Can be extended to interactive dashboards using Plotly or Streamlit.

Data

The dataset includes columns:

CustomerID, Age, Gender, Category, Item, Purchase Amount, Purchase Date, Rating, Subscription Status, Payment Method, Promo Code Used

Sample dataset used for analysis: shopping_trends_large.csv

Note: CSV file should be excluded from GitHub or replaced with a sample file due to privacy.

How to Run in Google Colab

1. **Open the notebook in Colab**
    - Upload the notebook file or open from GitHub directly in Colab.
    - If your data is in Google Drive, mount Drive first.
2. **Mount Google Drive** (if needed)
3. from google.colab import drive
4. drive.mount('/content/drive')

This gives Colab access to your Drive files. It uses an authorization step when mounting.

5. **Set file path**
    - If your CSV is in Drive, adjust the path, for example:
    - file_path = '/content/drive/MyDrive/path_to/shopping_trends_large.csv'
    - Or upload the CSV directly to Colab and use the local path shown in the file browser.
6. **Install dependencies**
   Colab already includes many packages, but if you need to ensure versions:
7. !pip install pandas matplotlib seaborn
8. **Run the notebook cells**
   Execute the cells in order; plots will appear inline in Colab.

Insights

- Top 20% of customers contribute the majority of revenue (Targeting Precision).
- Subscribed customers show higher repeat purchase rates (Retention).
- Clothing and Electronics are high-performing categories.
- Age groups, promo codes, and payment methods can guide personalized marketing strategies.

Tech Stack

- Python for data analysis
- Pandas for preprocessing and aggregation
- Matplotlib & Seaborn for visualizations
- Extendable with Plotly / Streamlit for interactive dashboards