Title: E-Commerce Data Analysis - Data Analyst Task

Your Name: Sanskriti Rai

Objectives & Approach

Objective : The task is to clean the data, analyze it, extract useful insights and then present your findings in a short report with visualizations. Dataset link: <u>eCommerce behavior data from multi category store</u>

Approach :

- The dataset given for the task was extremely huge with more than **42 Million rows**, hence the computation was resource-intensive.
- I used a parallel processing python library which has **similar syntax to Pandas**, named **Dask**. It is extremely efficient and fast. Dask works on the philosophy of lazy evaluation and does not do in-depth calculation unless called out to.
- Data Cleaning was handled using Dask then saved the file into a **parquet format**. Parquet stores files into binary format thus saving space and allows faster computation.
- Using this method of **Dask + Parquet**, I successfully **reduced 5.3 GB of dataset into 2.1 GB.** Furthermore, data analysis was done using **Matplotlib** for **Exploratory Data Analysis** and creating visualisations.
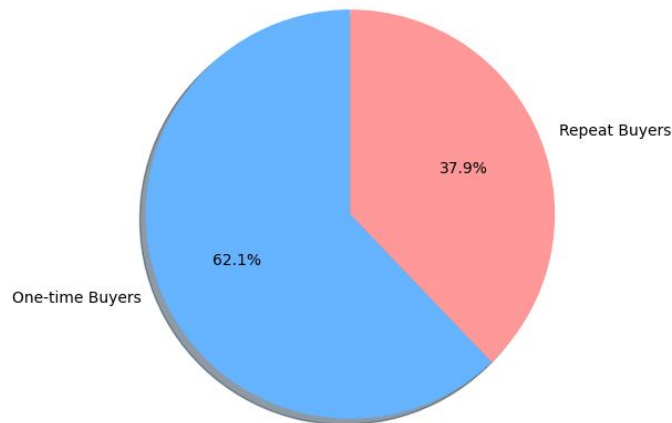
Key Insights

Revenue Volatility:

- Daily revenue ranged from ~6M to 9.7M
- Peak performance: Mid-October (14th-16th) with revenues exceeding $9.5M
- Later trend: Declining in the final week, dropping to ~$6.5M-$6M.
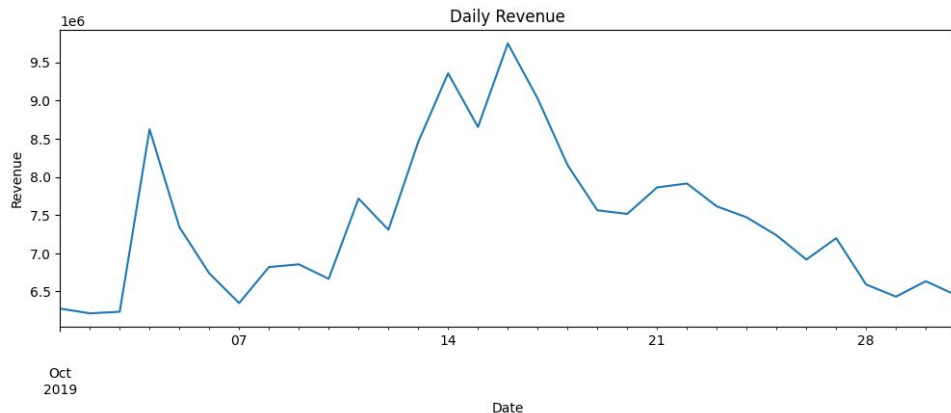
Proportion of Repeat vs. One-time Buyers

Repeat Buyers
37.9%

One-time Buyers
62.1%

Daily Revenue

Customer Engagement Metrics:

- Average events per user: 14.03 (shows moderate-low engagement)
- Repeat purchase rate: 37.86%
- Retention: Nearly 2 out of 3 customers are one-time purchasers

Key Insights

Category-Revenue:

- Electronics & Smartphones dominate: $150M+ revenue
- Second tier: Unknown category at $24M
- Market concentration: Top 2 categories account for approx 70-75% of total revenue.

Top 10 Categories by Revenue

Top 10 Brands by Revenue
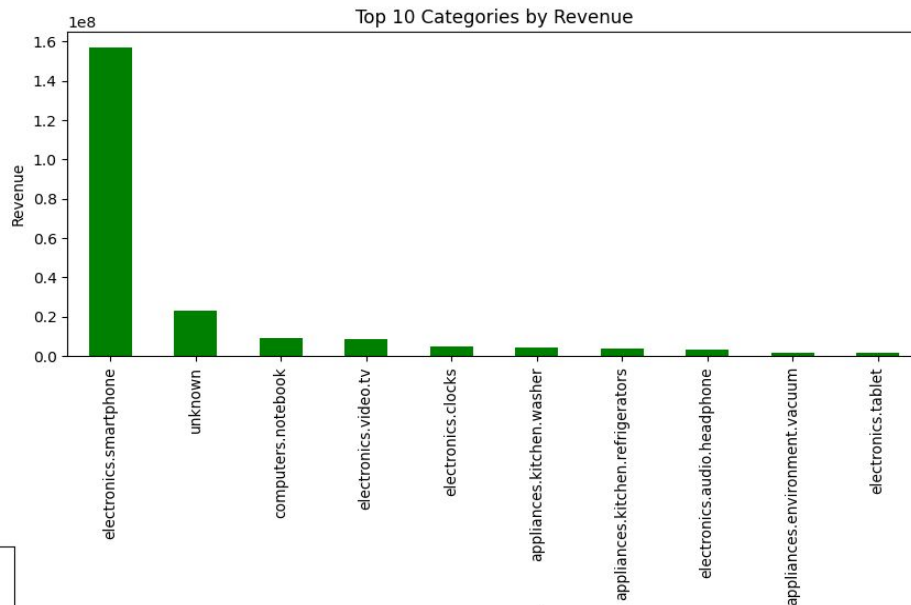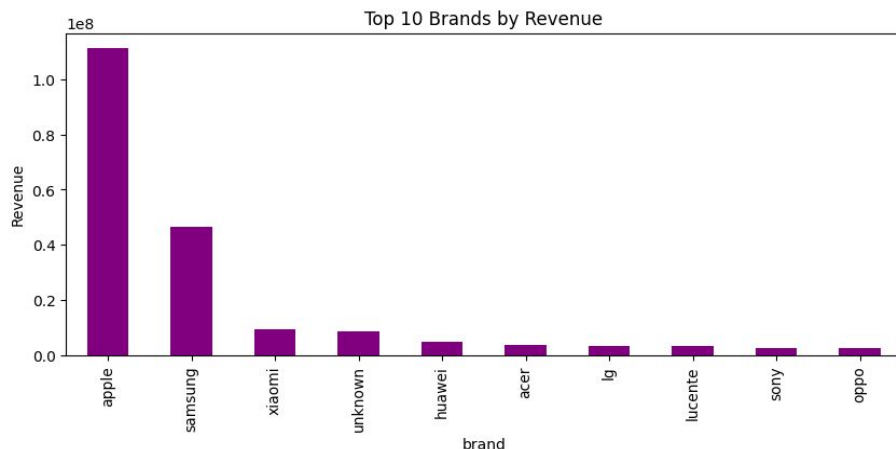
Brand-Revenue:

- Apple leads significantly: $100M+ revenue
- Samsung second: ~$50M revenue
- Market concentration: Top 2 brands account for approx 70-75% of total revenue.

Business Recommendations

Diversify Revenue Streams:

- Expand product categories: Invest in growing underperforming categories (computers, tv, clocks, etc).
- Reduce dependency: Reducing dependency on electronics/smartphone category to approx 50-55% of total revenue.
- Brand diversification: Developing relationships with emerging brands to reduce sole dependency on Apple/Samsung.
- It should be high priority work with a planned timeline of work.

Customer Retention Program:

- Target the 62% one-time purchasers: Developing personalized engagement campaigns.
- Loyalty program: Creating rewards system to increase repeat purchase rate from 38% to 50%+.
- Behavioral targeting: Use the 14+ average events per user to identify high-intent customers for retention efforts.
- It should also be high priority work with a planned timeline of work.