E-Commerce Customer Analytics Report

Project: Marketing Analytics – E-Commerce Customers
Author: Adoksh Suryawanshi
Tools Used: Python (pandas, seaborn, scikit-learn), Jupyter Notebook
Dataset: Kaggle - E-commerce Behavior Data

Overview

This project analyzes user behavior on an e-commerce platform using 2019 event-level data to uncover key patterns in conversion drop-offs, product performance, temporal trends, and customer segmentation. By combining exploratory data analysis with RFM scoring and unsupervised clustering, it provides a 360° view of how users interact with the platform highlighting where and why they drop off, what they buy, and how often they return. These insights inform targeted marketing strategies aimed at improving retention, optimizing campaign timing, and boosting overall conversion efficiency.

Business Questions Answered

1. **Where are users dropping off in the purchase funnel?**
   Identified drop-off rates between `view → cart` (~4.7%) and `cart → purchase` (~20%), helping uncover friction points in the buyer journey.

2. **Which product categories and brands are most influential in driving purchases?**
   Smartphones under `electronics` dominate the conversion funnel; brands like Apple and Samsung are top performers.

3. **How does user behavior change over time (seasonality, campaigns, etc.)?**
   Time-series trends revealed sharp purchase spikes on Black Friday and weekly patterns that indicate timing for promotions.

4. **What customer segments exist based on purchasing behavior and how should they be targeted?**
   Using RFM segmentation and clustering, we uncovered 3 actionable customer types:

   - *Occasionals*,
   - *Big Spenders*
   - *Loyalists*.

5. **What marketing strategies can improve retention and conversion?**
   Strategy recommendations were derived per segment using funnel data and RFM clustering insights.

Column	Description
user_id	Anonymized unique user identifier
event_time	Timestamp of interaction
event_type	Interaction type: view, cart, purchase
product_id	Unique product SKU
category_code	Product category hierarchy
brand	Brand name
price	Product price in USD

- Total Events: ~42 million
- Time Period: Nov 2019 – March 2020
- Event Types: `view`, `cart`, `purchase`

Methodology

1. Data Cleaning & Wrangling

Goal: Prepare raw event-level data for analysis across funnel stages, category performance, and user segmentation.

Steps:

- Imported ~8M records from a Kaggle-provided CSV file (`2019-Nov.csv`).
- Parsed `event_time` to datetime format for temporal analysis.
- Removed null or malformed rows (especially missing product/category/brand).
- Extracted top-level product categories from `category_code` (e.g., `electronics.smartphone`).
- Created separate datasets for each `event_type`: `view`, `cart`, and `purchase`.
- Merged datasets using `user_id`, `product_id`, and `event_time` to model funnels and trace user journeys.
- Computed aggregate metrics (count, revenue, conversion) by event type and product/brand.

2. Funnel Analysis

Funnel Stage	Count	Conversion %
Product Views	5.5 million	100%
Added to Cart	~260,000	~4.7%
Completed Purchases	~52,000	~0.94%
Cart to Purchase	~52,000	~20%

Goal: Diagnose drop-off points in the conversion funnel and identify conversion bottlenecks.

Steps:

- Modeled the 3-stage funnel:
  `View → Cart → Purchase`
- Grouped and counted unique user-product interactions per stage.
- Merged funnel stages sequentially and tracked transition rates using set-based logic.
- Calculated:
  - View → Cart conversion rate
  - Cart → Purchase conversion rate
  - View → Purchase conversion rate
- Visualized funnel as a bar chart to represent user leakage across stages.

Insights Extracted:

- Only ~0.94% of viewed products were purchased.
- Cart abandonment rate: ~80%.
- Indicates strong opportunity for recovery tactics like cart reminders and promotions

Drop-off Insight:

> *The majority of users drop off between view and cart, indicating browsing-heavy sessions with low cart intent. Cart abandonment is typical but could be improved with targeted recovery strategies.*

3. Product & Brand-Level Analysis

Goal: Determine high-intent product categories and brands driving value.

Steps:

- Aggregated event counts (`view`, `cart`, `purchase`) by `category_code` and `brand`.
- Ranked categories and brands based on:
  - Purchase volume
  - View-to-purchase conversion efficiency

    - Revenue generated
- Used bar plots to visualize top 10 performers by brand and category.

Insights Extracted:

- `electronics.smartphone` dominated all stages of the funnel.
- Brands like *Apple* and *Samsung* showed consistent high performance.
- *High alignment between views, cart adds, and purchases in the smartphone category.*
- *Suggests strong product-market fit and brand loyalty in mobile tech.*

4. Time-Series Behavioral Trends

Goal: Identify temporal patterns in user activity to align campaigns and operations.

Steps:

- Grouped event data by day (using `event_time`) for each interaction type.
- Visualized daily trends for `view`, `cart`, and `purchase` volumes.
- Identified peak periods and weekly recurring patterns.
- Highlighted Black Friday (Nov 29, 2019) as a major spike.

Insights Extracted:

- Strong mid-week spikes (*Wednesdays/Thursdays*).
- Large purchase surge on *Black Friday*.
- Suggests timing marketing and inventory around peak events.

5. RFM Segmentation (Recency, Frequency, Monetary)

Goal: Segment users by value and engagement to enable personalized targeting.

Steps:

- Defined:

    - **Recency:** Days since last purchase
    - **Frequency:** Total purchases
    - **Monetary:** Total spend per user
- Grouped data at the `user_id` level to calculate RFM metrics.
- Binned each RFM metric into quartiles (1 to 4) and scored users accordingly.

Segment	Description	Strategy
Occasionals	Infrequent, low spenders	Win-back offers, retargeting
Big Spenders	High revenue but low repeat visits	Upsell campaigns, loyalty perks
Loyalists	Frequent and recent buyers	Early access, referrals, rewards

Tools Used: `pandas.qcut()`, histogram and boxplots for distribution inspection

Marketing Strategy to be considered:

> *Scheduling email campaigns and promotions to align with historical high-traffic periods and reduce friction by optimizing checkout during peak periods.*

6. K-Means Clustering

Goal: Derive unsupervised behavioral clusters for deeper marketing insights.

Steps:

- Selected scaled versions of `Frequency` and `Monetary` for clustering.
- Applied **Min-Max scaling** to normalize variables.
- Used:

    - **Elbow method** to find optimal number of clusters (k = 3)
    - **Silhouette score** to validate cluster compactness
- Trained **KMeans** model from `sklearn.cluster`.
- Labeled clusters based on dominant traits:
    - Cluster 0: Occasionals
    - Cluster 1: Big Spenders
    - Cluster 2: Loyalists
- Visualized cluster results using 2D scatter plots.

Tools Used: `scikit-learn`, `matplotlib`, `seaborn`