E-Commerce Recommender System

Presented By : SAKSHI A

Tech Stack: Python, Pandas, Streamlit, Power BI, Scikit-learn

Contents

Objective

The goal of this project is to build a recommender system that **suggests the top 5 products to each user based on their past behavior.** I analyze what users have previously viewed, added to cart, or purchased. This information helps to understand their interests.

Recommender systems like this are used by Amazon, Netflix, and Flipkart **to improve user experience and sales.** I used data science techniques to replicate this in a simplified form using Python.

Dataset overview

Understanding the Data

The dataset used in this project contains 10,000 user interaction records. It includes user ID, product ID, event type (view, add-to-cart, purchase), price, quantity, and more. I transformed the event type into a numeric interaction score: view = 1, add-to-cart = 2, and purchase = 3. These scores help quantify user engagement with a product. This dataset gives a complete picture of user activity on the platform.

Data Cleaning & Preparation

Making the Data Model-Ready

The data was first cleaned by checking for null values and duplicates to ensure accuracy. I then mapped each event type to a numeric interaction score so it could be used in calculations. Using Pandas, we created a user-item matrix showing how users interacted with different products. This matrix is the core of collaborative filtering. It helps to compare users and identify patterns based on similar behaviors.

I used Collaborative Filtering based on user similarity to recommend products. Cosine similarity was calculated between users to find those with similar preferences. If one user liked a product, other similar users were likely to like it too. I avoided recommending products the user had already seen. This technique helps create personalized suggestions without needing explicit product features.

Model Building

Collaborative Filtering with Cosine Similarity

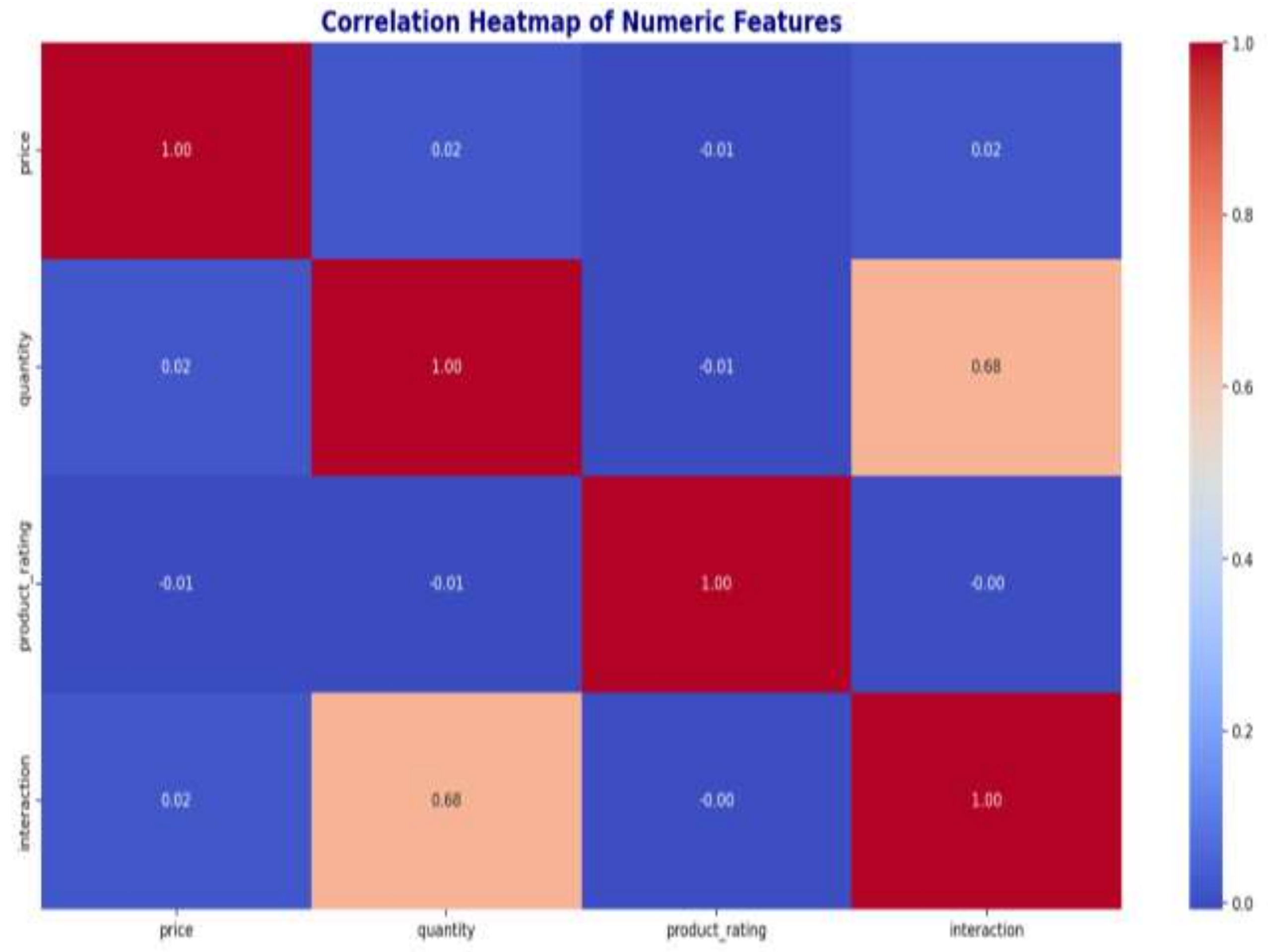
Model Evaluation

How I Evaluated the Model

To test the accuracy of our model, I used RMSE (Root Mean Squared Error), which tells how close the predicted interaction scores were to actual scores. A lower RMSE indicates better model performance. I also used Precision @K-style logic to check if recommended products matched real user purchases. These evaluations helped to confirm that my recommendations were meaningful and effective.

RMSE: 1.812

This heatmap shows the correlation between all numerical columns in the dataset. Correlation values range from -1 to +1: +1 indicates a strong positive correlation, -1 means a strong negative correlation, and 0 shows no correlation. Darker colors indicate stronger relationships. For example, interaction score and quantity may show moderate correlation. This analysis helps detect patterns and identify which features are more related, guiding model design and feature selection.



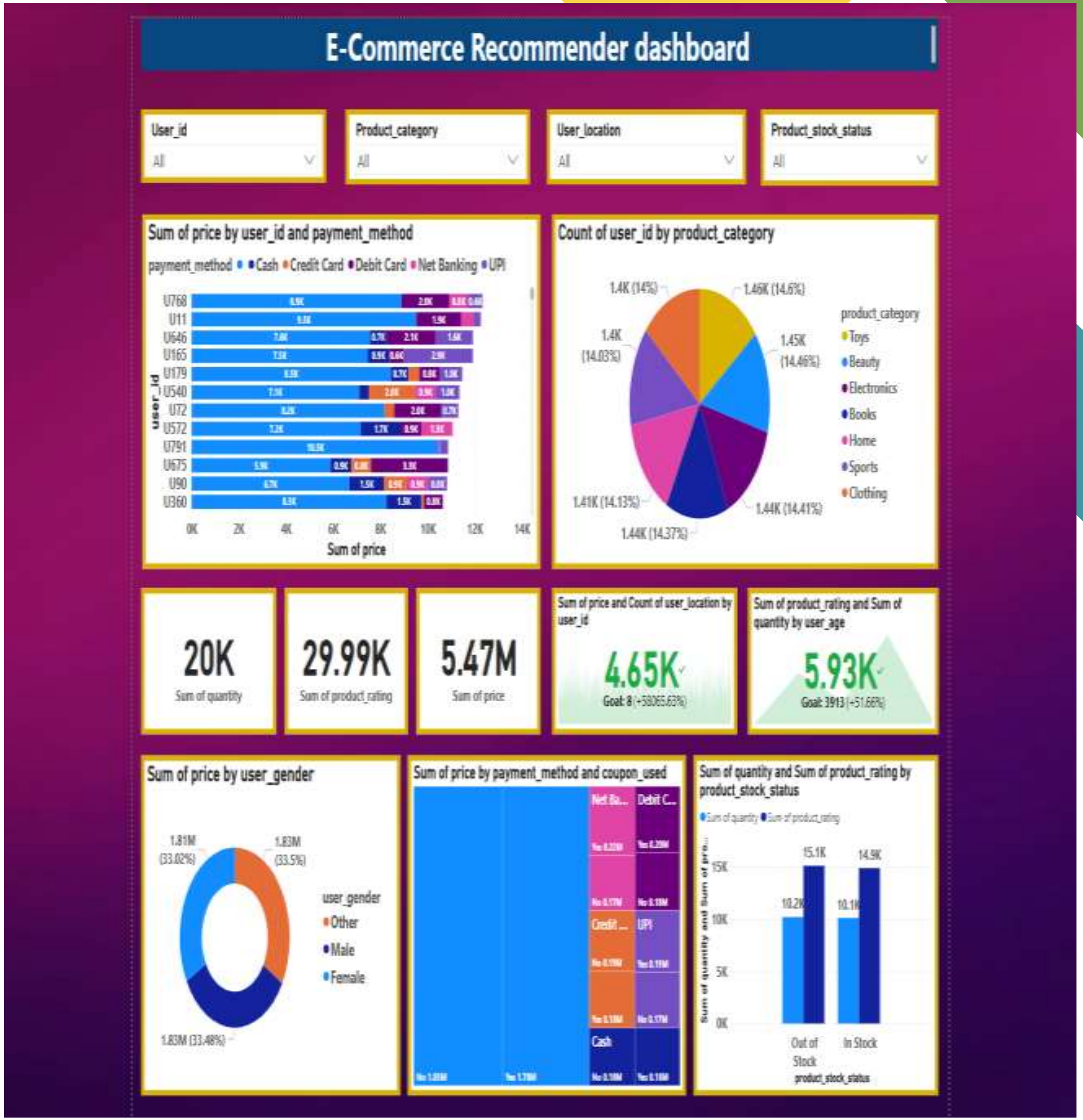
Correlation Heatmap Analysis

Understanding Relationships Between Numerical Columns

Power BI Dashboard

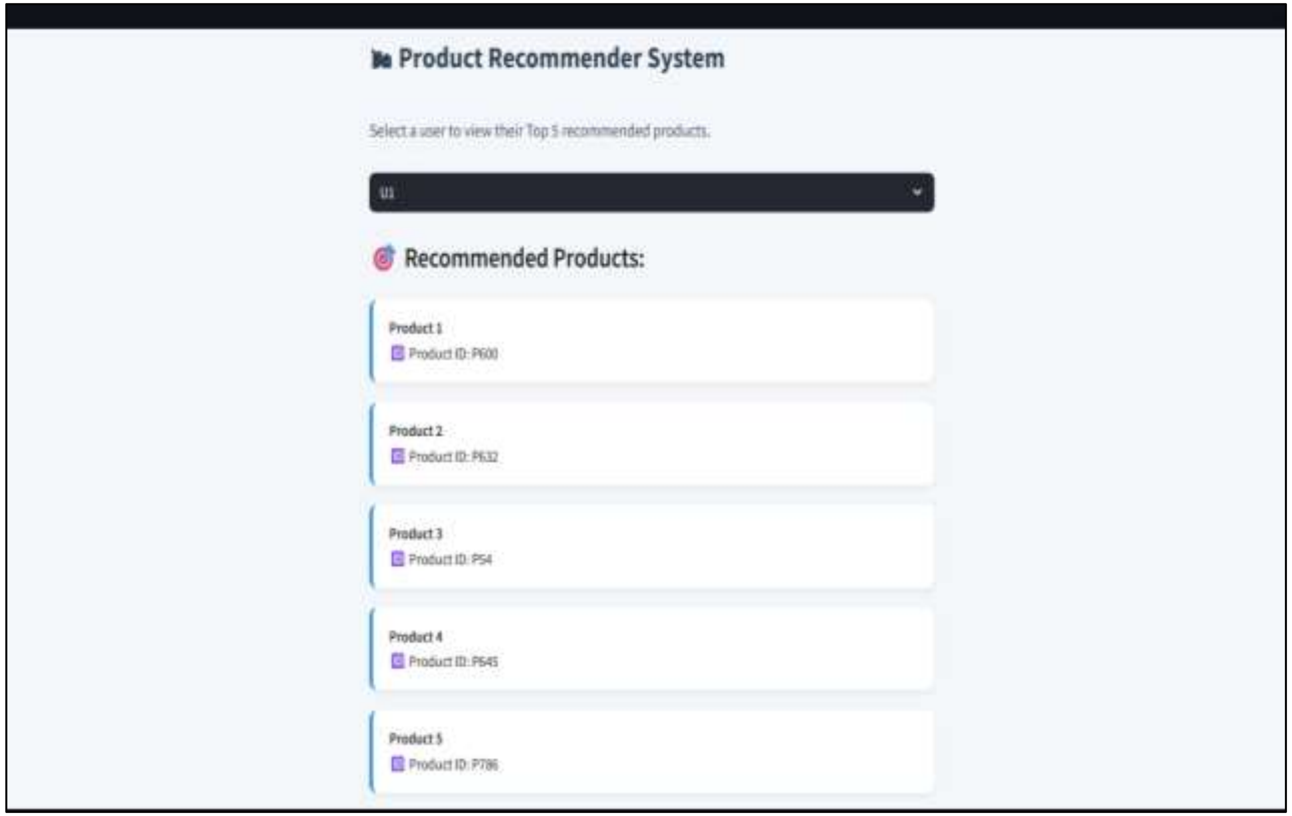
Visualizing Behaviour and Trends

The dashboard was created using Power BI to showcase product and user behavior. It includes filters for user ID, payment method, and product category. KPIs display total price, rating, and quantity. I used charts like pie, bar, donut, and stacked bar to visualize category-wise spending and preferences. This makes it easier for stakeholders to understand the data.



Streamlit Recommender App

Building a Simple Web App



I created a live app using Streamlit. The user selects a user ID and the app shows their top 5 recommended products. The backend uses collaborative filtering model to generate results. The interface is simple, clean, and interactive. This app demonstrates how the system would work in a real e-commerce environment.

Insights & Observations

What the Data Told Us

- I found that UPI and Credit Card were the most popular payment methods. Some users interacted with out-of-stock products, showing demand.
- Gender-wise spending was fairly equal. Most users viewed products before purchasing, and some frequently added items to the cart but didn't buy.
- Morning and evening were the most active times for user interactions.
- Users from metro cities like Delhi and Mumbai showed higher purchase activity.
- Product categories like Electronics and Clothing received the highest engagement. Certain users repeatedly engaged with the same products, suggesting potential loyalty or confusion.
- These insights can help in marketing, restocking, and user targeting strategies.

Thank you