WASHINGTON STATE UNIVERSITY

Amazon Product Recommendation System

Chinmay Chabbi 011858333

Yi Chou 011744816

CPTS 575 Data Science

PROJECT OVERVIEW

OBJECTIVE

Develop a recommendation system to provide personalized product suggestions to users based on their history and preferences.

METHODOLOGY

- Data Cleaning
- Feature Engineering
- Building and Evaluating the Recommendation System

DATASET OVERVIEW

The dataset comprises a rich collection of product data from Amazon across top 4 categories
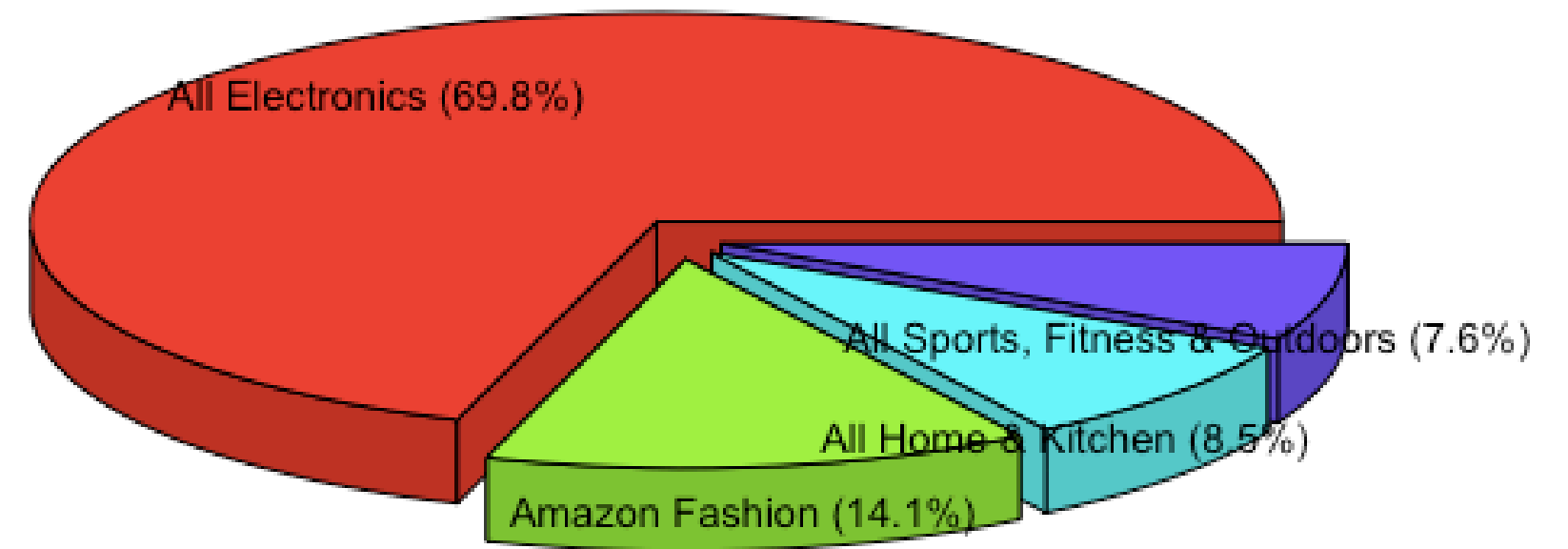
KEY DATASETS

01 **ELECTRONICS**

02 **SPORTS AND FITNESS**

03 **HOME & KITCHEN**

04 **FASHION**

05 **USER HISTORY**

All Electronics (69.8%)

All Sports, Fitness & Outdoors (7.6%)

All Home & Kitchen (8.5%)

Amazon Fashion (14.1%)

DATA CLEANING - OVERVIEW

Handling Missing Values

Removing Invalid Data

Standardizing Text

Removing Duplicates
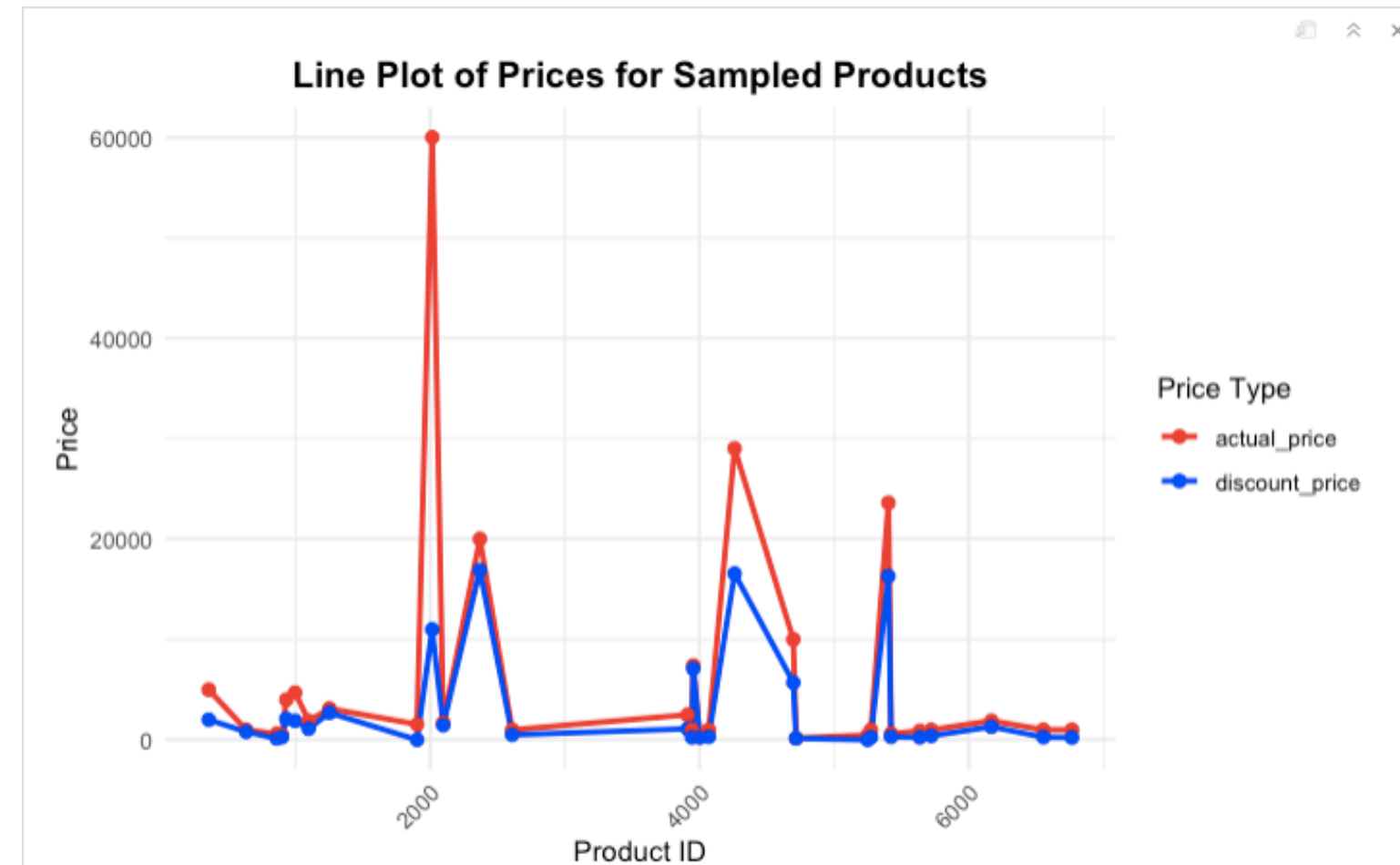
DATA CLEANING - VALIDATION

Comparison of original and cleaned data across datasets

Stage
- Original Rows
- Cleaned Rows
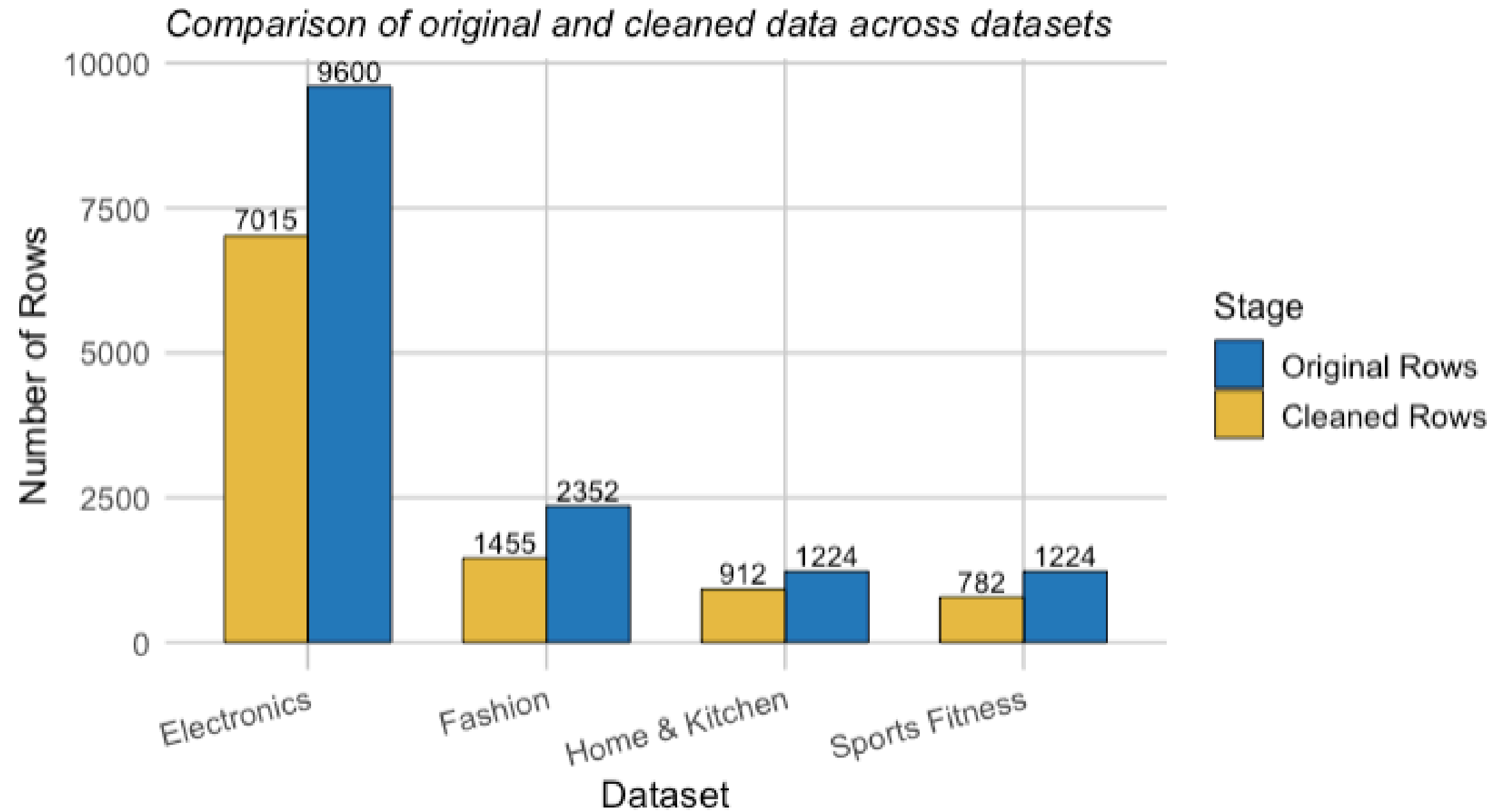
Source: Dataset Cleaning Process

FEATURE ENGINEERING

STEPS

01 DERIVED FEATURE

02 DISCOUNT PERCENTAGE

03 POPULARITY SCORE

04 TRANSFORM

05 COMBINING DATASETS

Line Plot of Prices for Sampled Products

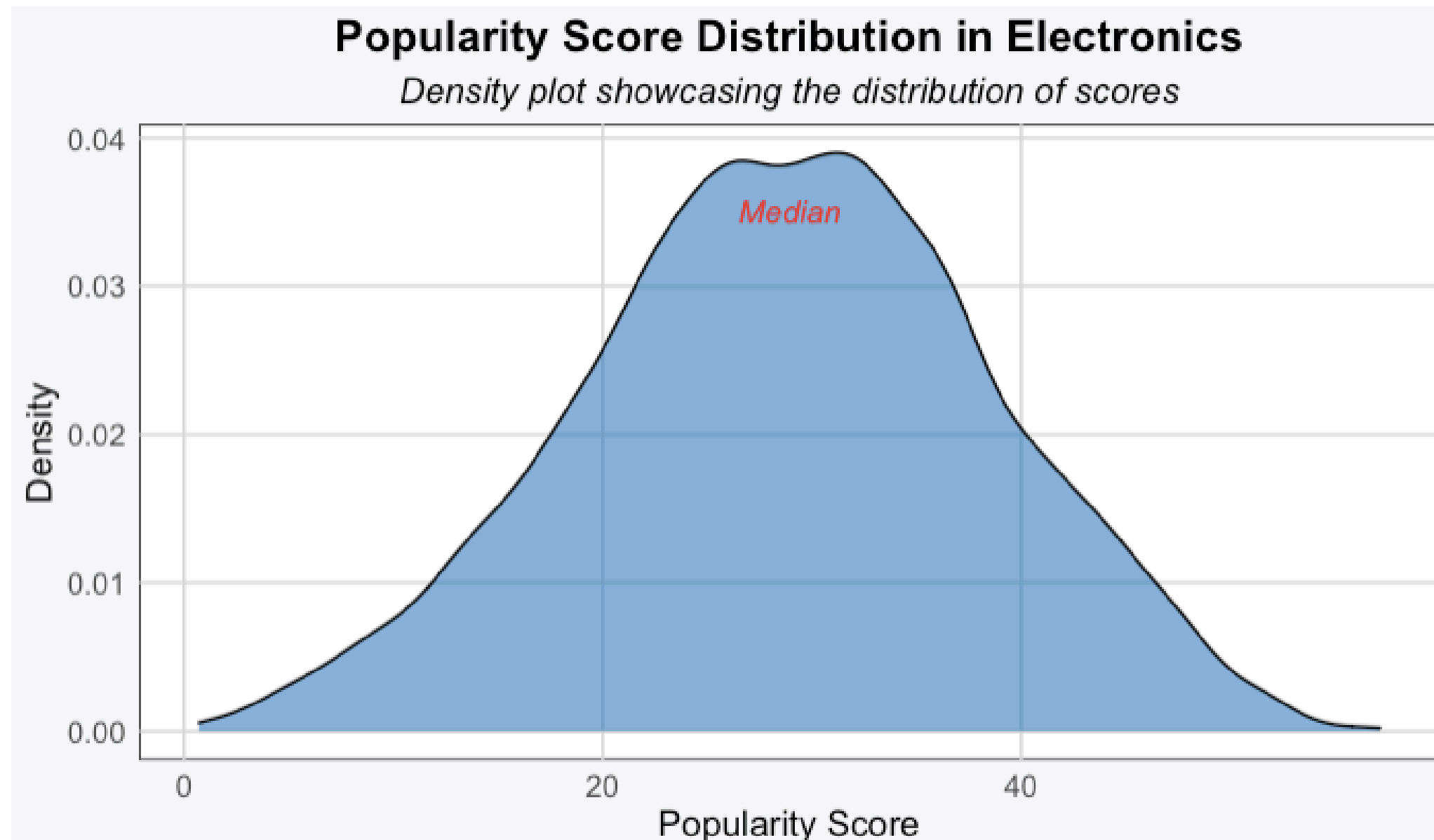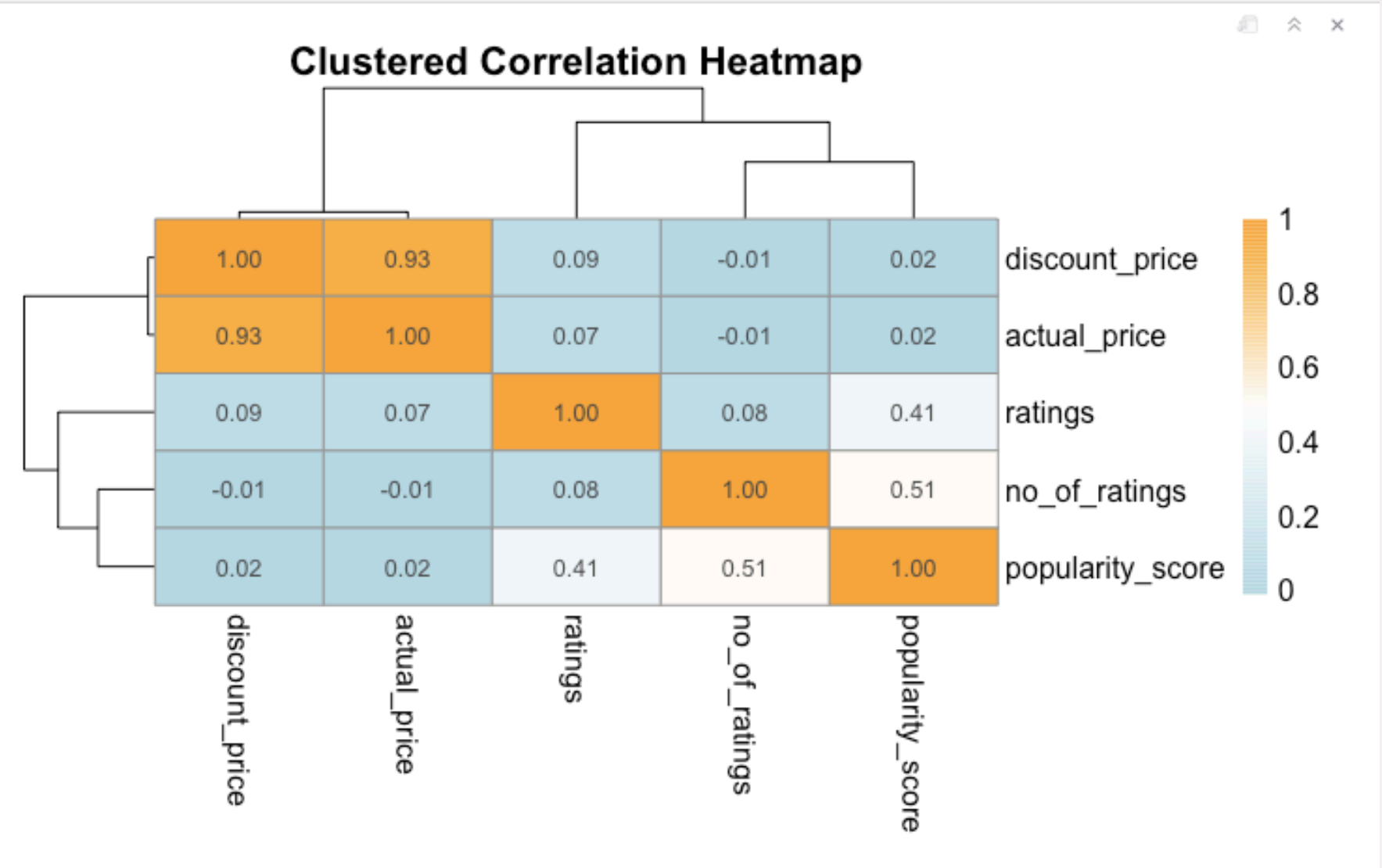$$Discount\ Percentage = \left(\frac{Actual\ Price - Discount\ Price}{Actual\ Price} \right) \times 100$$
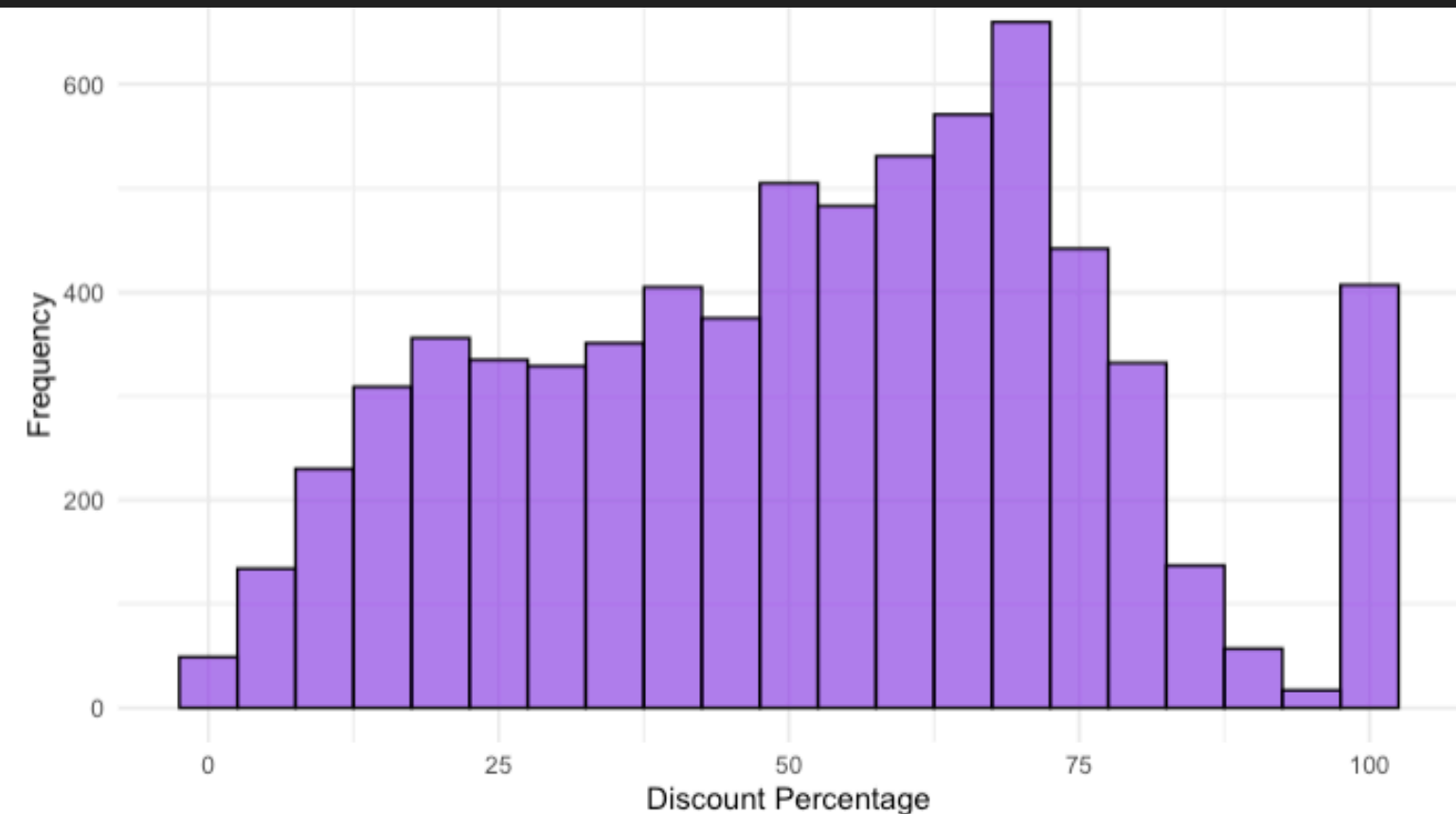
POPULARITY SCORE

FORMULA

$$\text{Popularity Score} = \text{Ratings} \times \log(1 + \text{Number of Ratings})$$

Popularity Score Distribution in Electronics
Density plot showcasing the distribution of scores

CLUSTER CO-RELATION HEATMAP

Clustered Correlation Heatmap

RECOMMENDATION SYSTEM APPROACH

WEIGHTED SCORE

50%
Popularity

10%
Discount

40%
Rating

Weighted Score=$(0.5×Popularity)+(0.4×Ratings)+(0.1×Discount)$

- **Analyze user history and match with available products.**
- **Normalize all the score**
- **Use weighted scoring to rank products for recommendations.**

RECOMMENDATION SYSTEM APPROACH

COMBINED SCORE

30%
category

40%
weighted

30%
Rating

Combined Score=(w1×Category Score)+(w2×Weighted Score)+(w3×Ratings)
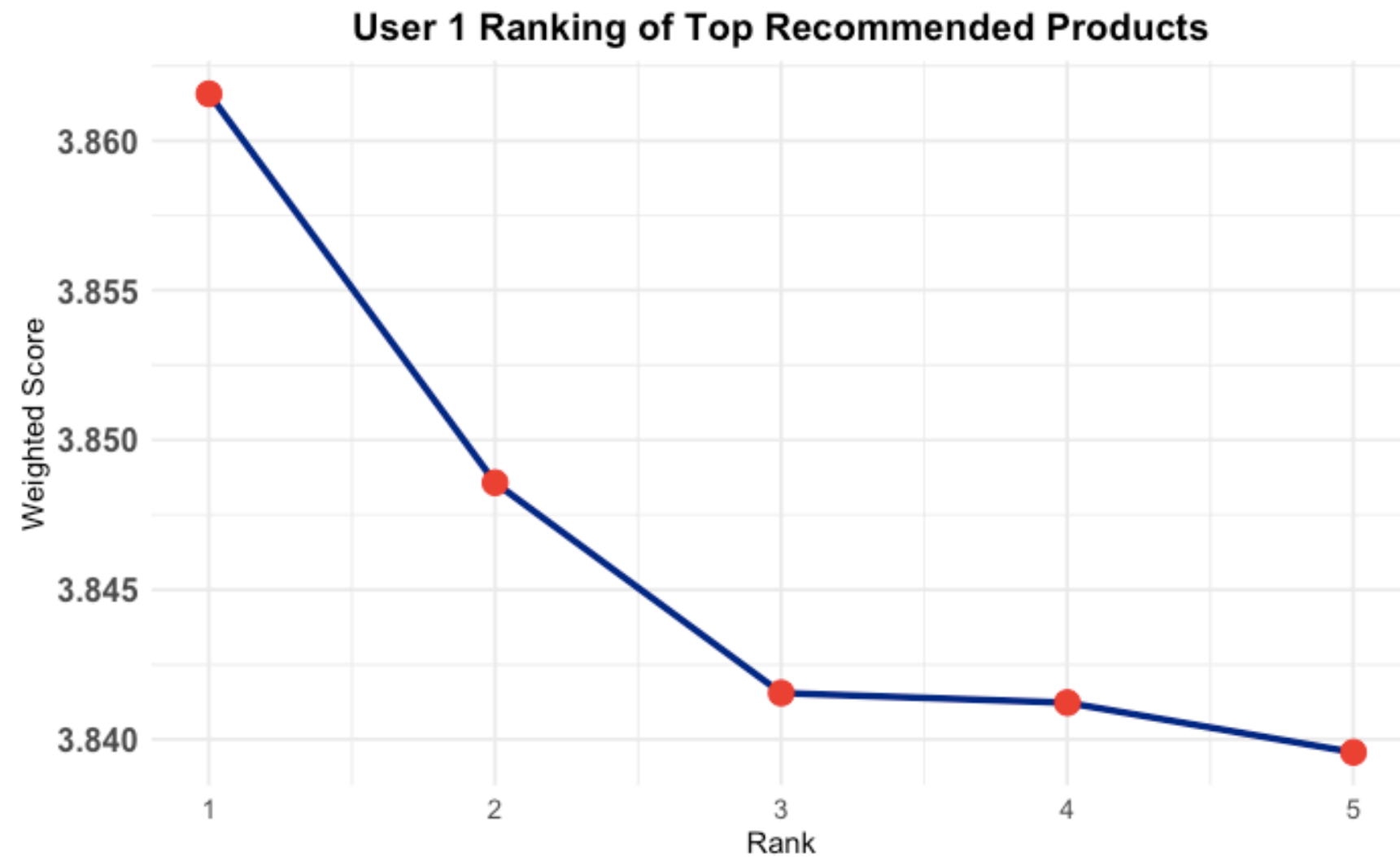
Matched user interactions (view, add-to-cart, purchase) with available products.
Combined scores based on category priority, weighted product features, and user preferences.
Displayed top-ranked products for each user.

RECOMMENDATION

User 1 Ranking of Top Recommended Products

01	**Air cooler**
02	**Fan Cleaner**
03	**logitech gaming mouse**
04	**Apple adapter**
05	**Apple pencil**

EVALUATION

Accuracy of Recommendations:
- To evaluate the system's effectiveness, we compared the predicted recommendations with actual user purchases or preferences.
- Calculated the following metrics:
  - True Positives (TP): Number of correctly recommended products that the user interacted with.
  - Overall Accuracy: Proportion of accurate recommendations out of total recommendations made.

ACCURACY: 35%