VALENTINA SANCHEZ

Skincare Recommendation System

MENTOR: BEN BELL

Introduction to the Problem

ISSUE

The skincare industry becoming a $231.3 billion market, driven by the impacts of COVID-19 and social media influence. Consumers are prioritising skin health, and buying products endorsed online.

CONCERN

A great concern regarding consumers' limited knowledge about their skin and product ingredients. Often, purchases are influenced by paid endorsements from influencers, raising concerns about transparency and informed decision-making.

Project Overview

Our goal is to align users with the skincare items that most closely align with their individual needs and preferences by building a supervised learning recommendation system.

The dataset was sourced from a Kaggle user who performed web scraping on the Sephora website, resulting in the creation of multiple tables.

The dataset is primarily organised into five review tables and product information table that we will merge and use.

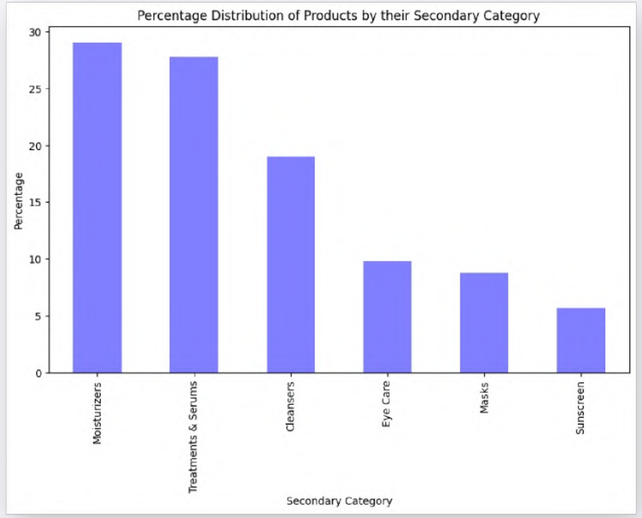
COLLABORATIVE

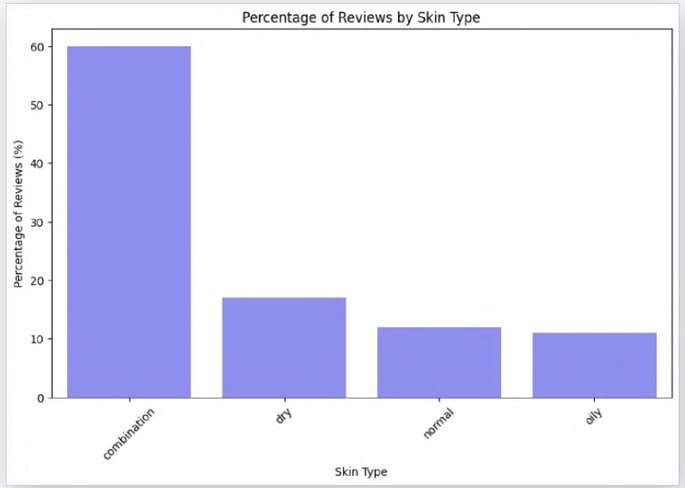
CONTENT

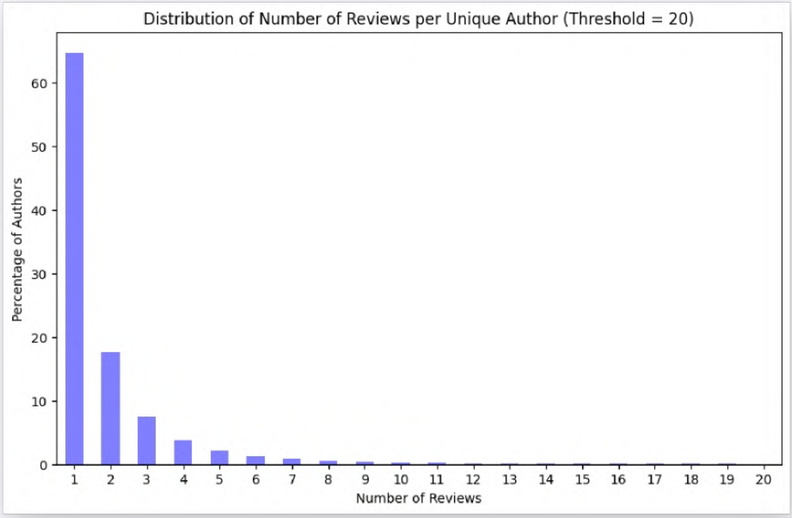
COLLAB/CONTENT

TAKING THESE IMPORTANT PRINCIPLES INTO ACCOUNT

Exploratory Data Analysis







CATEGORY DISTRIBUTION

Refinement of secondary categories. Most populated categories on the dataset.

SKIN-TYPE DISTRIBUTION

Observing which skin type has made the most reviews.

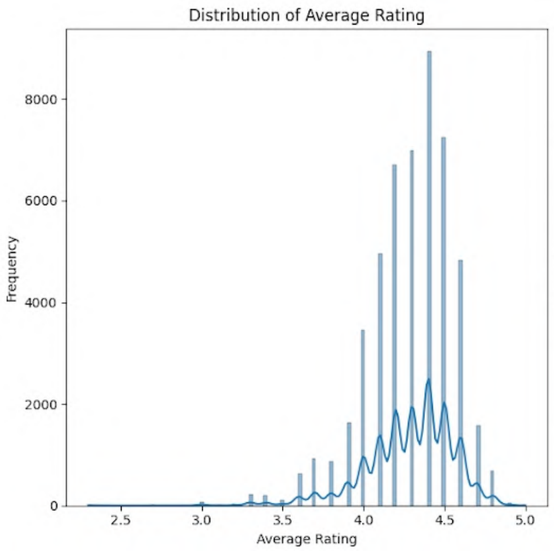
NO. OF REVIEWS / UNIQUE AUTHOR

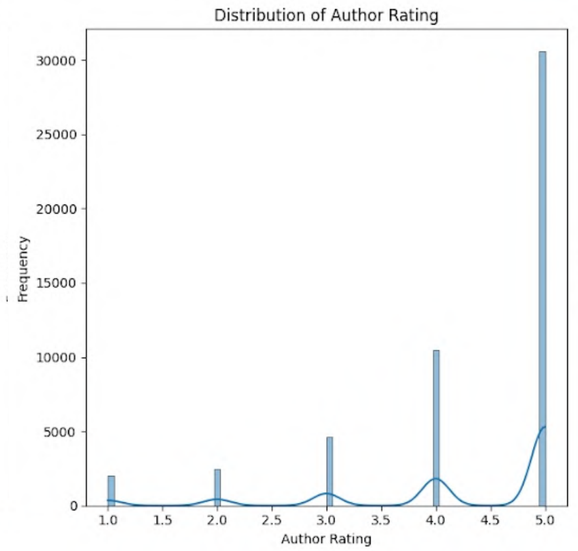
Most user make one review and less likely more than five.

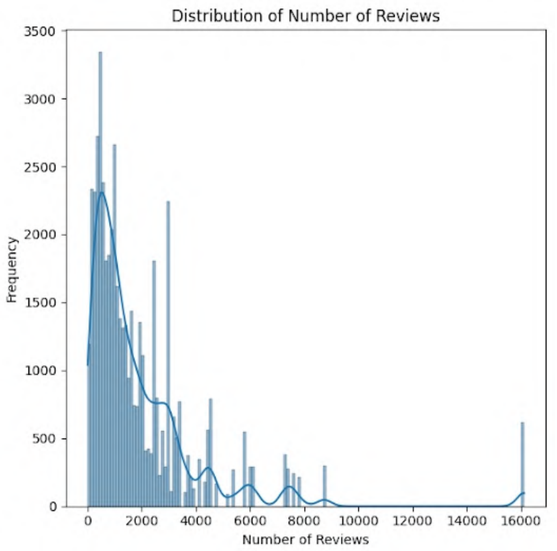
ANALYSIS

We consider the skewness and normalise these values later on.

Distribution of:
Average Rating | Author Rating | No. Reviews of Product

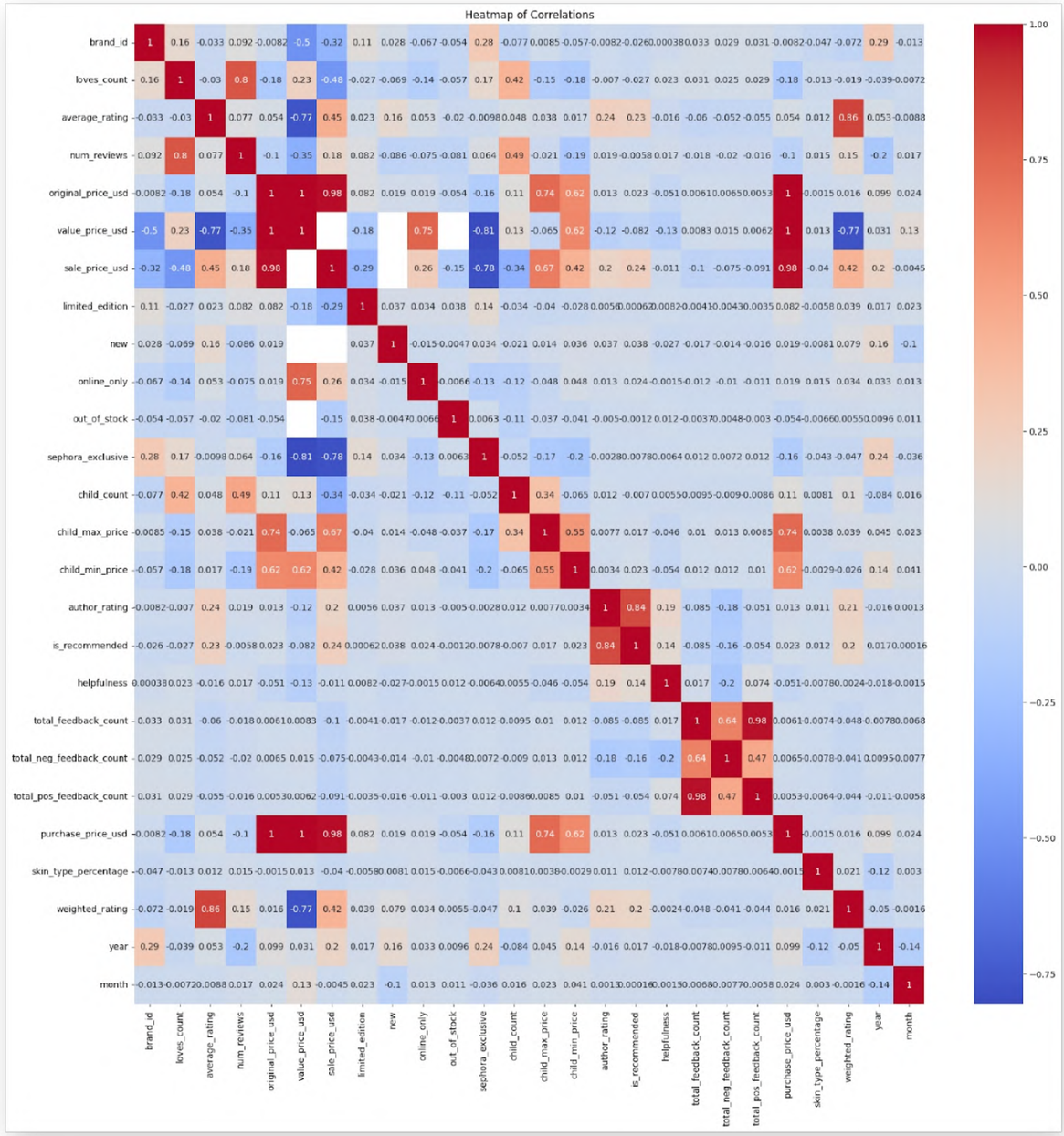






HEATMAP

Many features operate independently of each other. However, due to complexity concerns and the presence of missing values in certain columns, most of these features will not be utilized.



Try Pitch

Data Preprocessing

01

ENCODING SKIN TYPE & PRODUCT CATEGORY WITH LABEL ENCODING

02

CLEANING INGREDIENTS AND APPLY TF-IDF (TERM FREQUENCY-INVERSE DOCUMENT FREQUENCY) VECTORIZATION

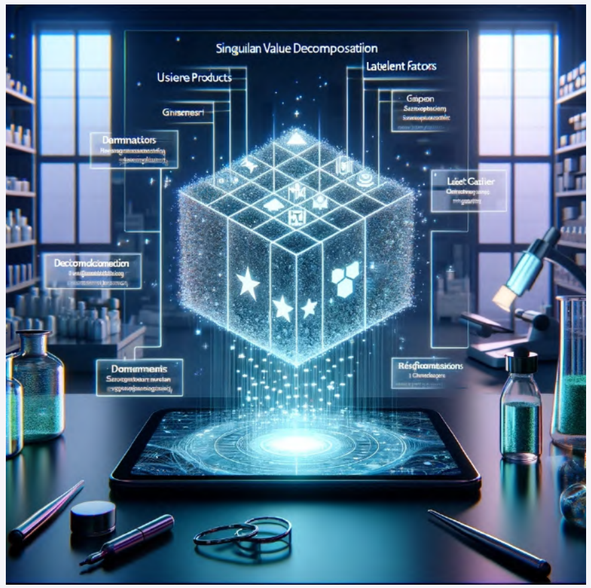
03

NORMALISING BOTH RATINGS AND NUMBER OF REVIEWS

Modeling

LIGHTFM

ability to combine collaborative filtering and content-based methods, providing personalized skincare product recommendations by considering both user preferences and item features.



SVD

uncover underlying patterns in skincare product ratings, aiding in recommendation accuracy by capturing latent factors in user-item interactions.

K-NN

simplicity and effectiveness in identifying similar users or items based on their ratings, contributing to personalized skincare product recommendations.

Evaluation

01

LIGHTFM

no_components	20
epochs	20
num_threads	4
auc	0.82

performs the best in distinguishing between relevant and irrelevant products for users.

02

SVD

lr_all	0.01
epochs	30
n_factors	70
auc	0.705

even though it performs well it only has the item, the product id and the author rating

03

KNN

k	20
sim_options	name_based
user_based	false
rmse	0.28

predictive accuracy in terms of minimizing the difference between predicted and actual ratings

Conclusion

The **objective** of using LightFM for the skincare hybrid recommendation system is to enhance the personalization of product suggestions for users.

SVD and KNN suffer from the limitation potentially biases recommendations towards popular items, neglecting the comprehensive benefits of considering product attributes.

We recommend users to prioritise product effectiveness and suitability for their skin type over popularity. Based on factors such as ingredients, user reviews, and suitability for specific skin concerns rather than solely relying on popularity metrics.

Further Work

- Incorporation of additional dataset attributes for deeper insights.
- Exploration of ensemble approach combining SVD and k-NN and comparing it to the LightFM model we currently have.
- Integration of textual review analysis and diverse product attributes.
- Development of user-friendly interface for users who may not know what to use for their skin.
- Emphasis on diversity and novelty in recommendations.