# Capstone 2

## Recommendation system for Sephora

Benazir Rowe, Springboard Data Science Fellow

# Purpose, Motivation, Description

- **Purpose:** use explicit feedback from user reviews to build a recommender system that can provide personalized product recommendations for Sephora online platform.
- **Goal:** to develop a system capable of offering product recommendations that resonate with individual user's needs and preferences, thereby enhancing their shopping experience at Sephora, boost online sales, improve customer satisfaction, increase online revenue.
- **Description:** By exploring various machine learning algorithms, such as Matrix Factorization techniques like SVD and NMF, neighborhood-based methods like KNN, we aim to create an effective model that can scale with the size of Sephora's extensive product range and user base.

# Data acquisition

The data was obtained from Kaggle.com [1] containing about 8000+ products and over 1 million user reviews from the skincare category. This dataset was collected via Python scraper in March 2023 and contains:
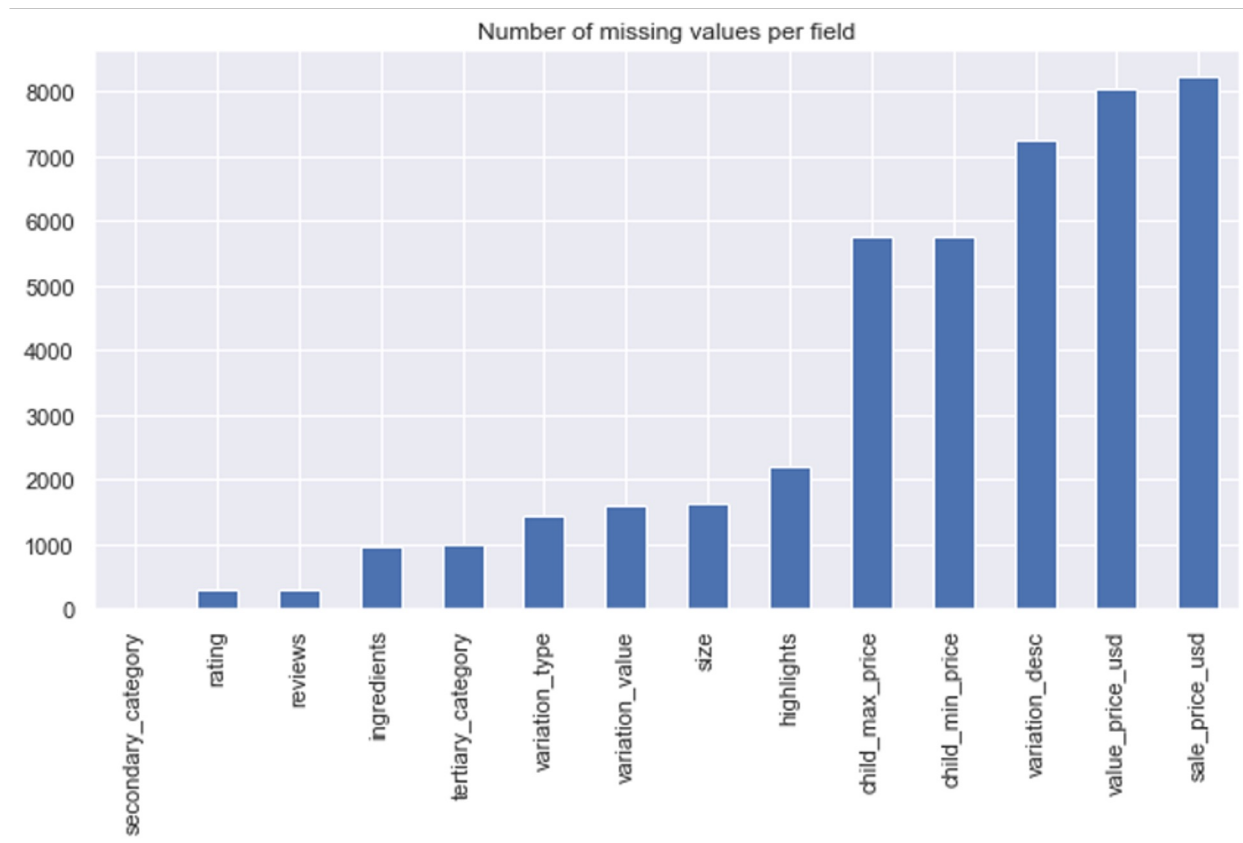
- information about all beauty products (over 8,000) from the Sephora online store, including product and brand names, prices, ingredients, ratings, and all features.
- user reviews (over 1 million on over 2,000 products) of all products from the skincare category, including user appearances, and review ratings by other users.

[1] https://www.kaggle.com/datasets/nadyinky/sephora-products-and-skincare-reviews
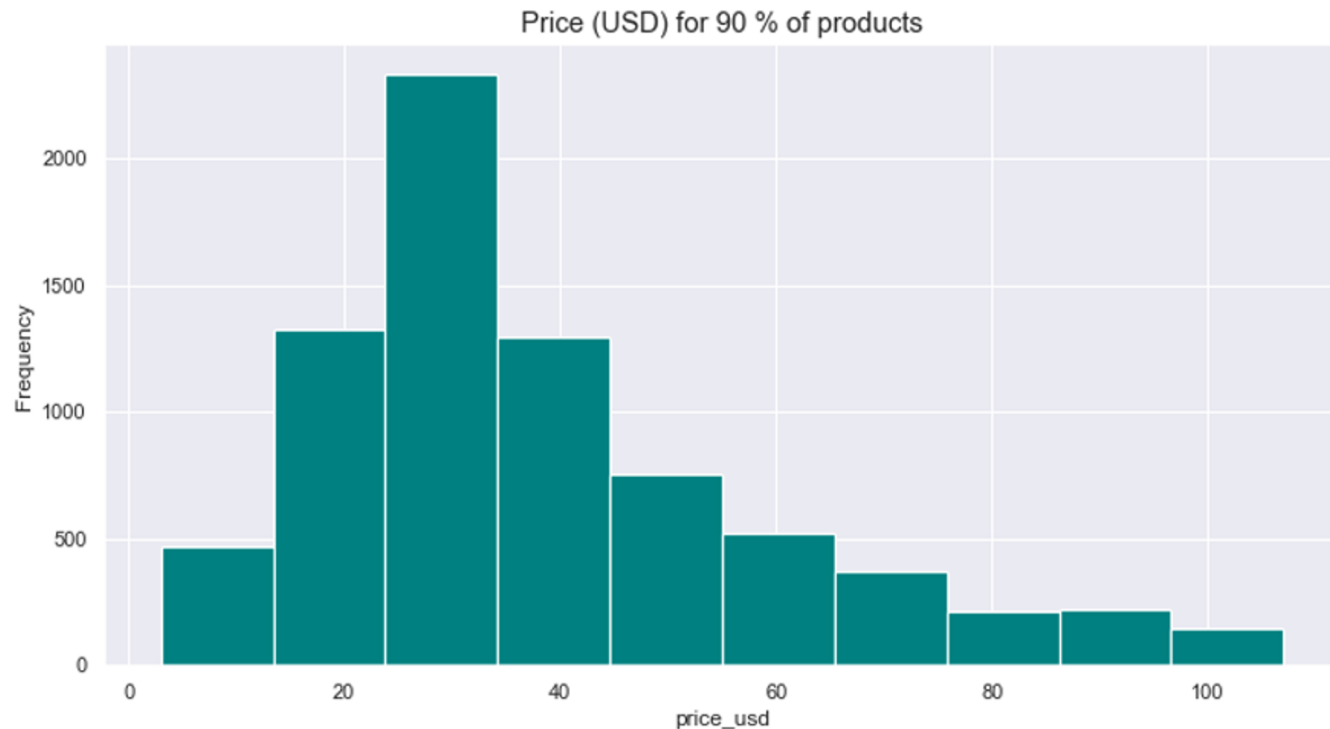
# Missing data

Many categories have missing values, but they are not exactly missing, more that the field is not applicable to many products.

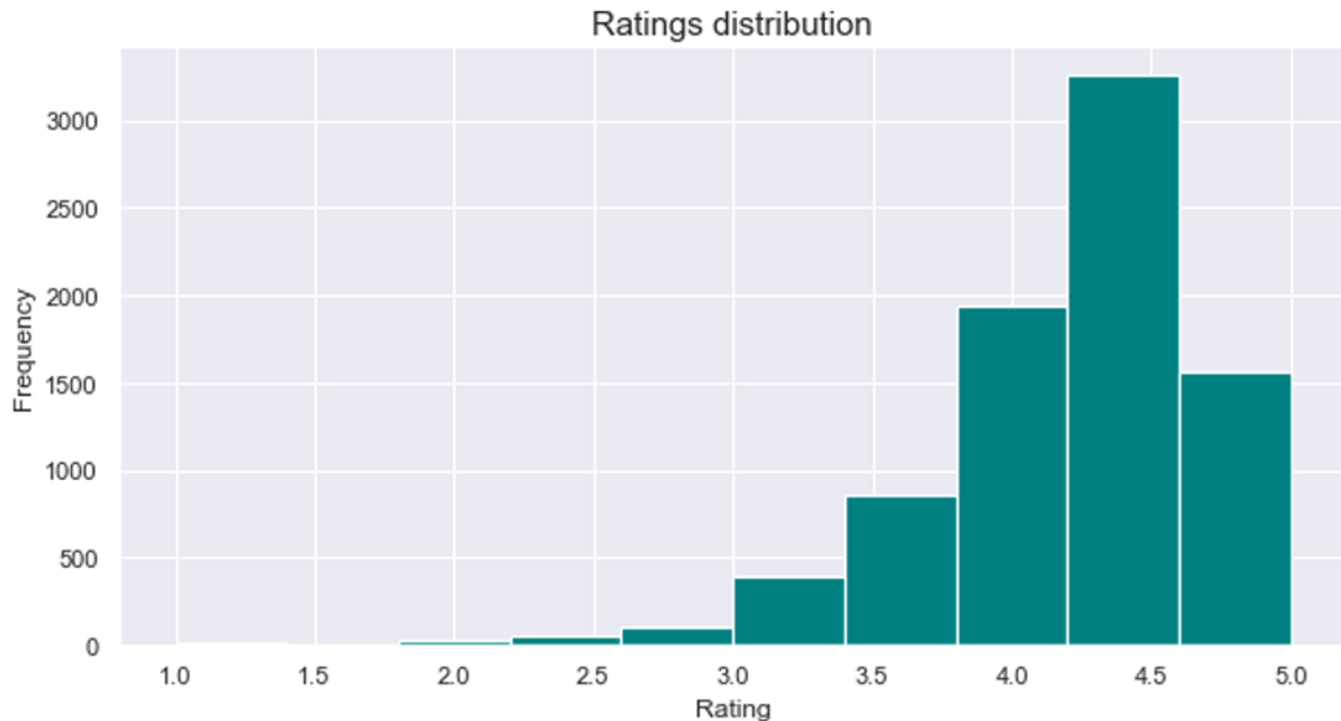We are not intending to analyze those fields, so we drop those columns.



Number of missing values per field

# Exploratory Data Analysis

Most products cost around $30 and the distribution of product price is skewed left.



Price (USD) for 90 % of products

# Exploratory data analysis

Most products have average rating of 4.5 stars, with some having 5-star and 3-star reviews.

Interestingly, there are not many products with 1-star or 2-star average ratings.



Ratings distribution

# Feature selection

For matrix factorization models, we focus on product, user and rating features.

We separate those features, remove any missing values and user, item duplicates.

# Modeling

**Singular Value Decomposition (SVD):** This method decomposes a matrix into 3 other matrices. SVD can handle missing values in the user-item interaction matrix, allowing prediction generation for items that a user has not yet interacted with.

**Non-negative Matrix Factorization (NMF):** NMF is another matrix factorization method similar to SVD, but with the constraint that the matrices are non-negative. NMF allows for parts-based representations and can give a more intuitive understanding of the latent features.

**Biased SVD:** Biased SVD includes user and item biases in its model to account for the systematic tendencies of some users to always give high or low ratings and for some items to receive higher or lower ratings than others.
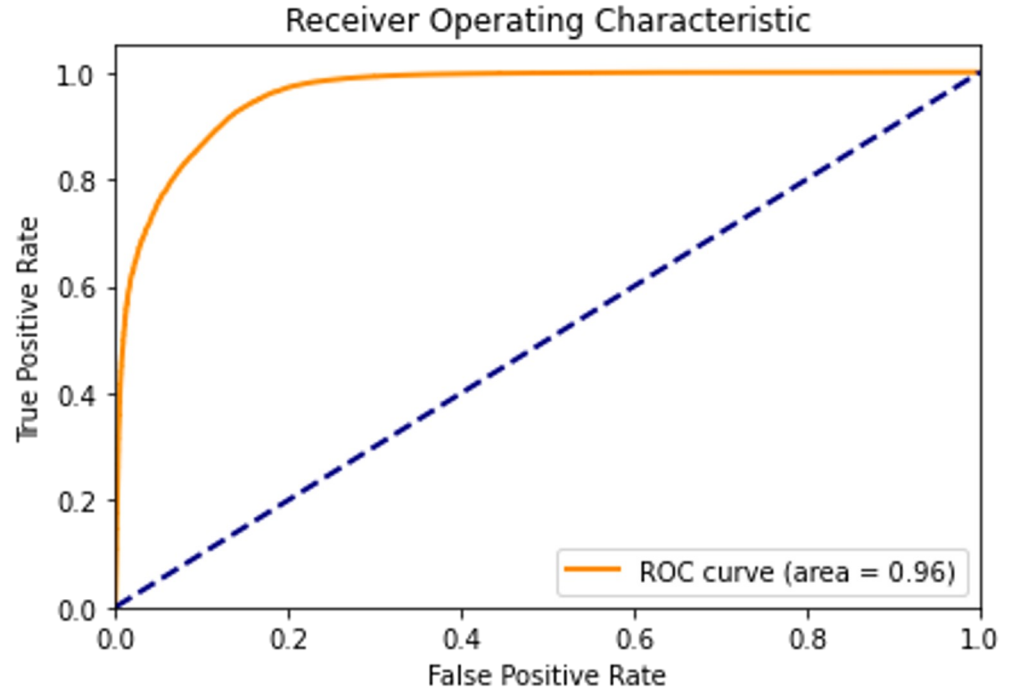
# Model evaluation

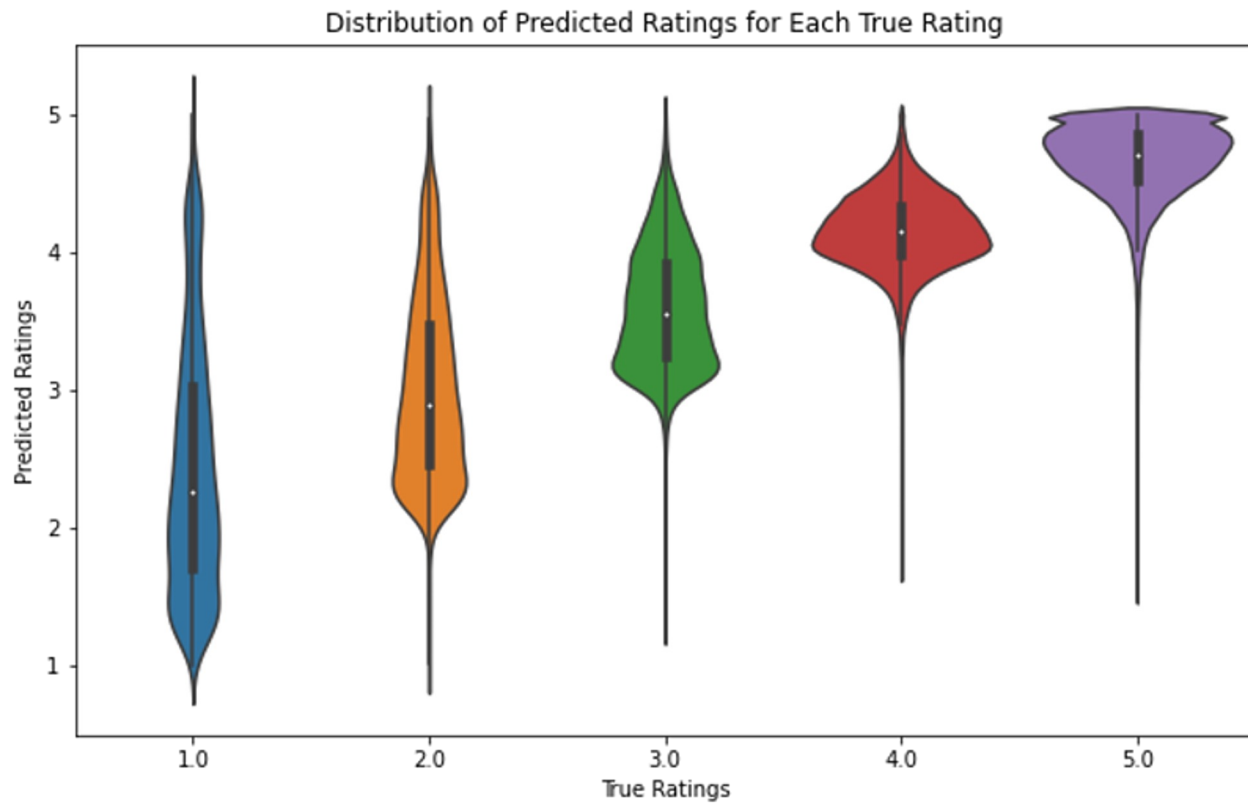| | RMSE | MAE | Fit time | Test time |
|---|---|---|---|---|
| **SVD** | 1.0313 | 0.7582 | 8.86 | 1.89 |
| **Biased SVD** | 0.916 | 0.6464 | 10.24 | 1.20 |
| **NMF** | 1.1537 | 0.8545 | 32.30 | 2.83 |

Based on the results of 5-fold cross validation, Biased SVD model performed best in all metrics except fit time, however 2 second fit time difference is not an issue.

# Classification metrics for binarized predictions

| Metric | Value |
|--------|-------|
| Precision: | 0.93 |
| Recall: | 0.99 |
| F1: | 0.96 |



Receiver Operating Characteristic

# True vs Predicted ratings

# Sample recommendation for user '6538326896'

| product_name | brand_name | loves_count | rating | reviews | price_usd | limited_edition | new | online_only | out_of_stock | sephora_exclusive | primary_category | secondary_category |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Dramatically Different Moisturizing Gel | CLINIQUE | 95864 | 4.4383 | 3977.0 | 32.5 | 0 | 0 | 0 | 0 | 0 | Skincare | Moisturizers |
| Mini Daily Microfoliant Exfoliator | Dermalogica | 110141 | 4.696 | 4598.0 | 18.0 | 0 | 0 | 0 | 0 | 0 | Skincare | Mini Size |
| Daily Microfoliant Exfoliator | Dermalogica | 95311 | 4.696 | 4598.0 | 65.0 | 0 | 0 | 0 | 0 | 0 | Skincare | Cleansers |
| Plantscription Retinol Night Moisturizer | Origins | 11217 | 4.5385 | 39.0 | 73.0 | 0 | 0 | 1 | 0 | 0 | Skincare | Moisturizers |
| Salicylic Acid Acne Healing Dots | Peace Out | 316297 | 4.1779 | 4560.0 | 19.0 | 0 | 0 | 0 | 0 | 0 | Skincare | Treatments |
| Oil-Absorbing Pore Treatment Strips | Peace Out | 102327 | 3.975 | 1562.0 | 19.0 | 0 | 0 | 0 | 0 | 0 | Skincare | Treatments |
| Brightening Saffron Serum - Radiant Rani | RANAVAT | 7269 | 4.7824 | 216.0 | 135.0 | 0 | 0 | 0 | 0 | 0 | Skincare | Treatments |
| Unseen Sunscreen SPF 40 PA+++ | Supergoop! | 302951 | 4.26 | 3262.0 | 48.0 | 1 | 0 | 0 | 0 | 1 | Skincare | Sunscreen |
| Mini Unseen Sunscreen SPF 40 PA+++ | Supergoop! | 66960 | 4.2604 | 3260.0 | 22.0 | 0 | 0 | 0 | 0 | 1 | Skincare | Mini Size |
| Violet-C Radiance Mask | Tatcha | 96300 | 4.0507 | 651.0 | 70.0 | 0 | 0 | 0 | 0 | 0 | Skincare | Masks |

The recommendation has a problem skincare theme this user may be interested in and looks relevant and comprehensive.