

Recommendation system for Sephora

CAPSTONE 2

Benazir Rowe
SPRINGBOARD | ONLINE DATA SCIENCE BOOTCAMP

Introduction

The power of personalization in retail is increasingly recognized, and the beauty industry is no exception. With a vast array of products to choose from, consumers can often feel overwhelmed and uncertain about their purchase decisions. This is where recommender systems come into play. In essence, a recommender system functions as a search ranking system, where the input comprises user and contextual data, and the output delivers a ranked list of items. The objective of a recommendation task, given a query, is to locate relevant items in a database and rank them according to specific goals, like clicks or purchases.

We want to build recommendation models capable of suggesting users the products based on various features, starting with just ratings, and later possibly incorporating user and item features into our model. The developed recommender system is expected to improve customer satisfaction, user engagement and conversion rates.

The data

The data was obtained from Kaggle.com¹ 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>

About Sephora

Sephora, a leading beauty retailer, offers an extensive selection of cosmetics, skincare, body, fragrance, nail color, and hair care products. Founded in France in 1969 and currently owned by luxury conglomerate LVMH, Sephora has grown into an internationally recognized brand with a significant online presence. With almost 340 brands, including its private label Sephora Collection, the retailer caters to a diverse clientele with varied beauty preferences and needs.

Overview of the recommender systems

Recommender systems have become integral to how businesses operate online, used extensively across various sectors, including retail, entertainment, and even news media. They aim to guide users in a personalized way to discover relevant content or products from a large pool of possibilities. Recommender systems utilize various methodologies to generate these recommendations:

Collaborative Filtering (CF): This method uses the past behavior of users, such as their ratings, preferences, or purchasing history, to make recommendations. Collaborative Filtering assumes that users who have agreed in the past tend to agree in the future. There are two types of CF - User-based (recommendations based on users who are similar to the target user) and Item-based (recommendations based on items that are similar to those that the target user rated).

Content-based Filtering: This method uses features of items and user profiles to make recommendations. For instance, if a user showed interest in a particular genre of movies, the system would recommend other movies of the same genre. Here, the focus is on the properties of items rather than on the behavior of users.

Hybrid Methods: These methods combine both collaborative and content-based filtering to leverage the strengths of both methods. Hybrid models can provide more accurate recommendations by overcoming the shortcomings of individual methods. For example, collaborative filtering might suffer from the cold start problem where it can't make accurate recommendations for new users due to lack of past behavior. In contrast, content-based filtering can address this problem as it doesn't require past user behavior data.

Matrix Factorization: This is a class of collaborative filtering algorithms used in recommender systems. Matrix factorization algorithms work by decomposing the user-item interaction matrix into the product of two lower-dimensionality rectangular matrices. Singular Value Decomposition (SVD) and Non-negative Matrix Factorization (NMF) are two commonly used matrix factorization methods in recommender systems.

Deep Learning-Based Methods: With the rise of deep learning, neural network-based models have also been applied to recommendation problems. These models can learn complex non-linear relationships and also deal with large-scale data.

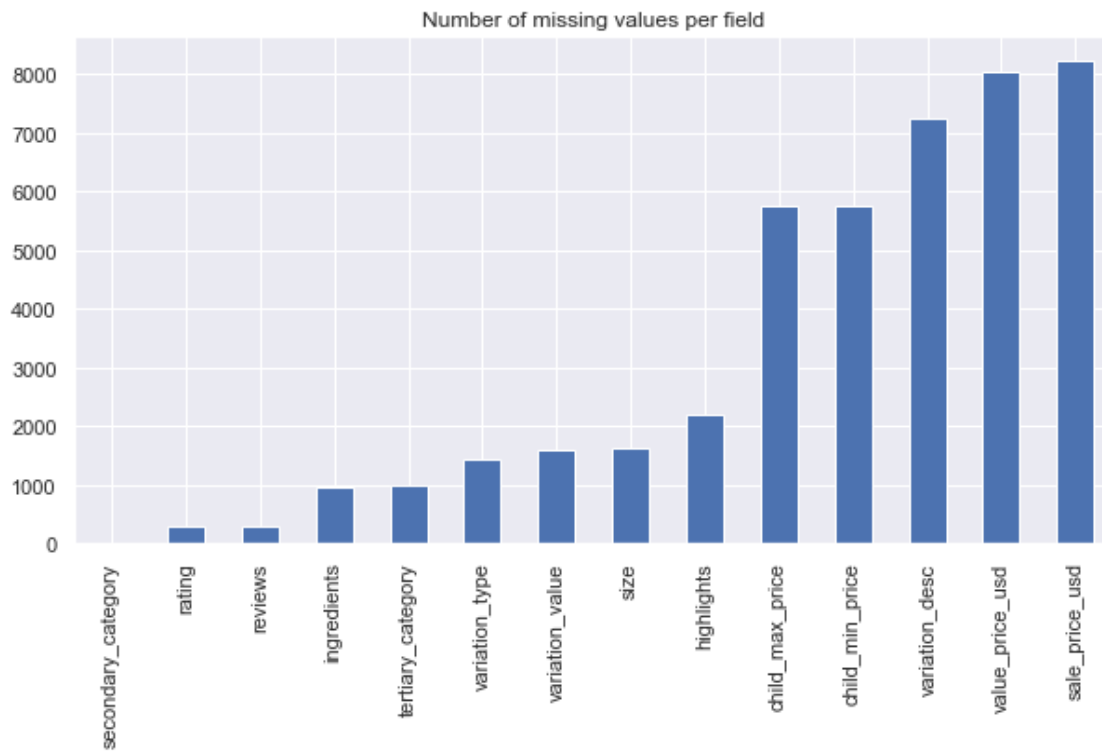
Evaluation of Recommender Systems: Accuracy of recommender systems is typically evaluated using measures like RMSE (Root Mean Square Error), MAE (Mean Absolute Error), Precision, Recall, F1-score, ROC-AUC, etc. Moreover, metrics like precision@k and recall@k are used to assess the quality of top-k recommendations.

For Sephora, a key business problem is how to personalize the user experience in a way that assists customers in discovering products they might love but aren't aware of. With over 8000+ products and more than a million user reviews in the Skincare category alone, the challenge lies in effectively leveraging this wealth of data to provide meaningful product recommendations.

As users interact with the Sephora platform by browsing products, writing reviews, and making purchases, they leave behind a trail of data that can be analyzed and used to understand their preferences and behaviors. This project aims to use this explicit feedback from user reviews to build a robust recommender system that can provide personalized product recommendations, enhancing the shopping experience for each user and potentially driving additional sales for Sephora. 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. The ultimate goal is to develop a system capable of offering top-notch product recommendations that resonate with individual user's needs and preferences, thereby enhancing their shopping experience at Sephora.

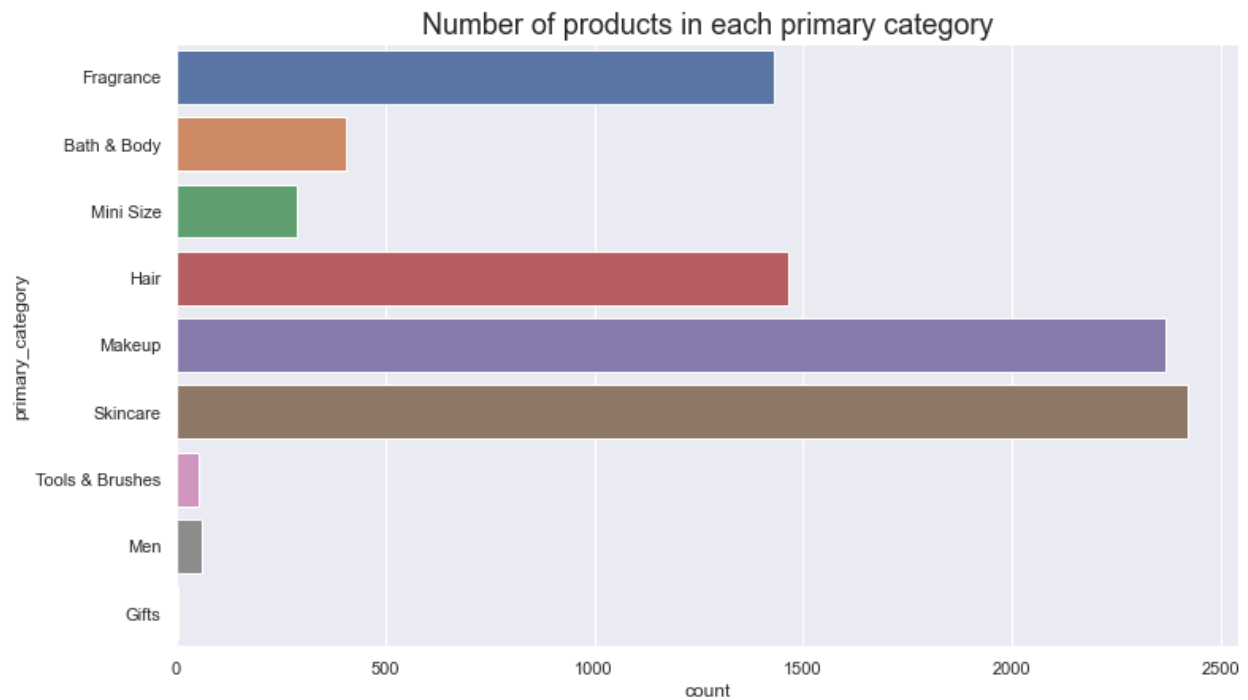
Exploratory data analysis

We conduct an initial exploratory data analysis to understand the nature of the data, identify any potential limitations, and address any data imbalance or quality issues. First, we explore the missing data.



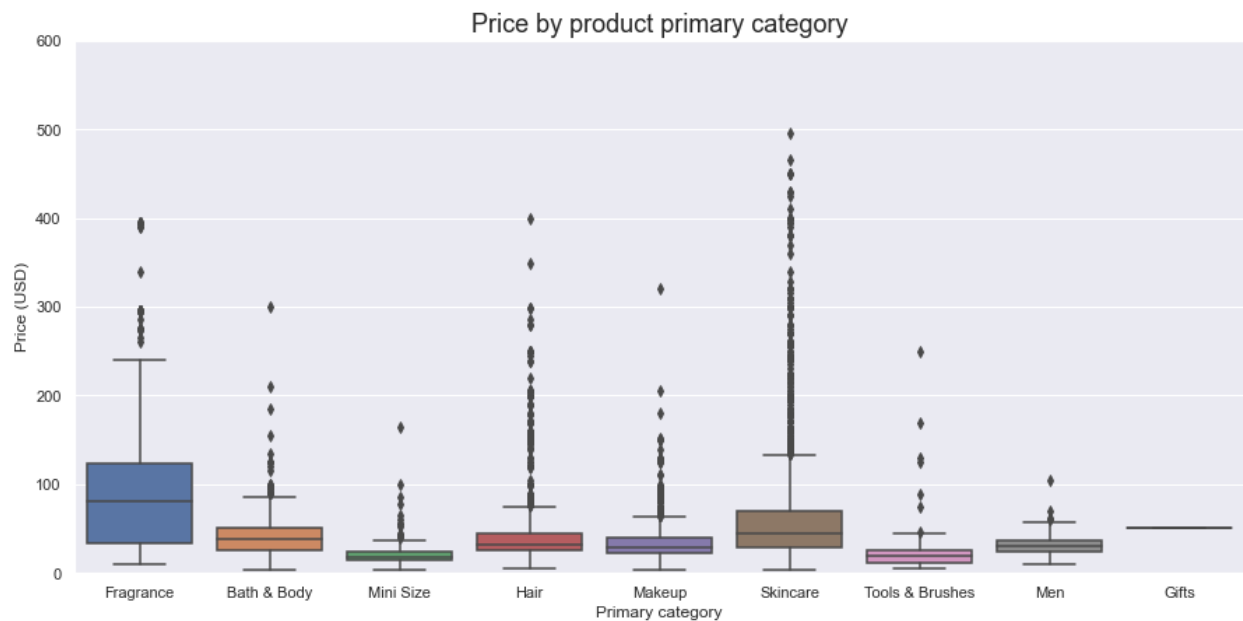
So many categories do not have values, but they are not exactly missing, more that the field is inapplicable to many products, for example sale_price_usd, value_price_usd, variation_desc, child_min_price, child_max_price, highlights, variation_type, variation_value, tertiary_category. We are not intending to analyze those fields, so we drop those columns.

Product features



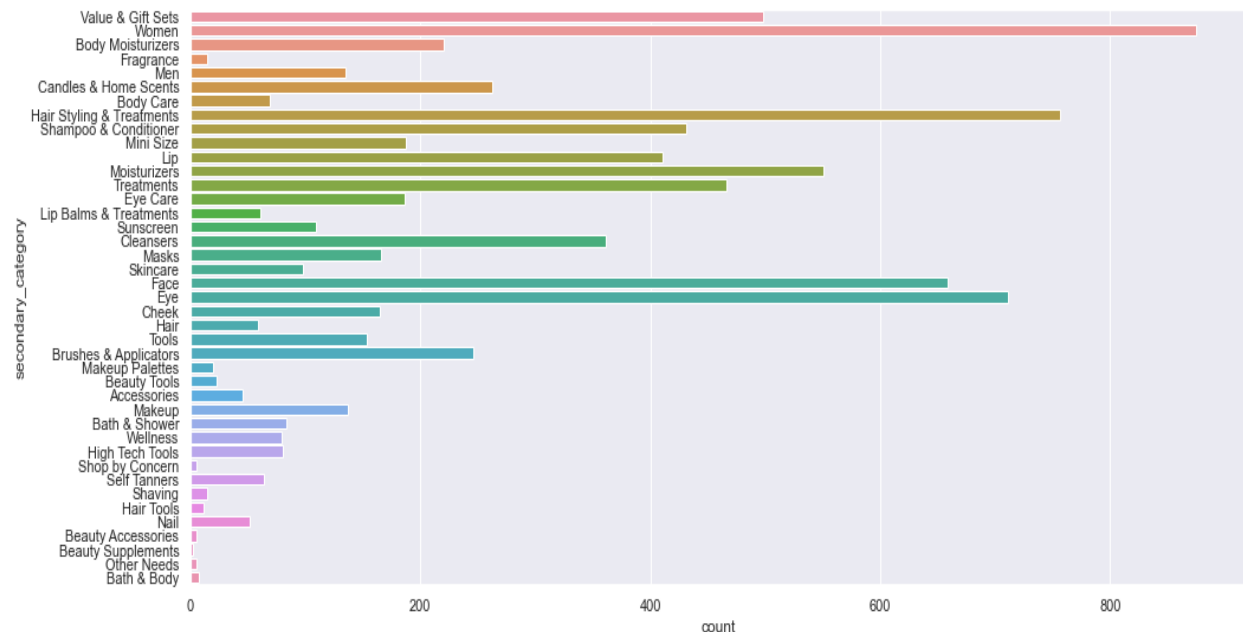
In this dataset, most products are from makeup and skincare category, followed by fragrance, hair, bath & body, and mini size. The smallest categories are tools and brushes, Men and Gifts.

Let's look at the price distribution for each category.

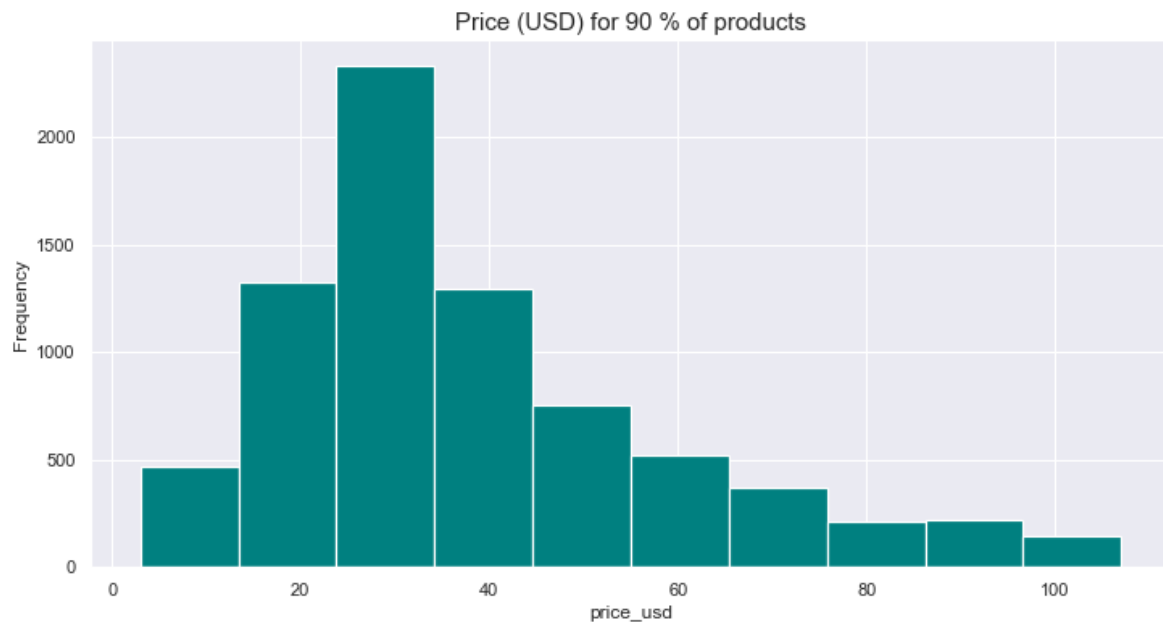


The most expensive category is fragrance, followed by skincare and bath&body, hair and makeup and other categories.

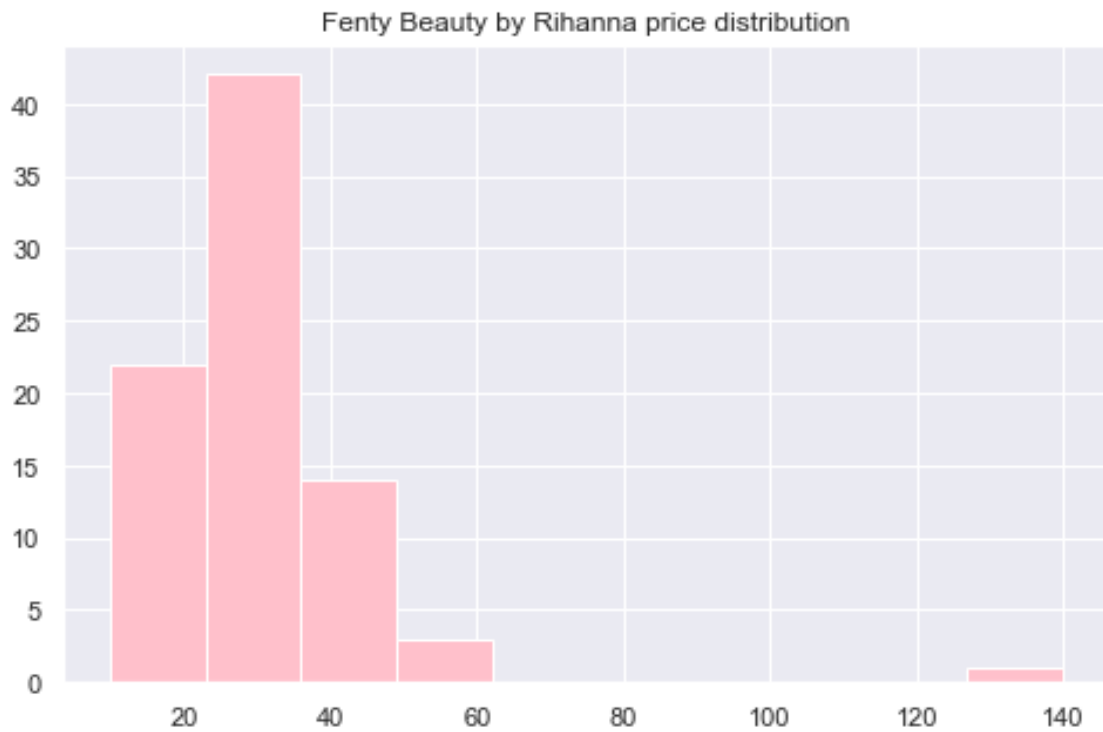
If we explore the secondary categories, it gives us more detailed breakdown, for example, for makeup we can see that eye is followed by the face and then lip and cheek etc.



If we look at 90% of the products, we can see that mode and median are around \$30 mark.

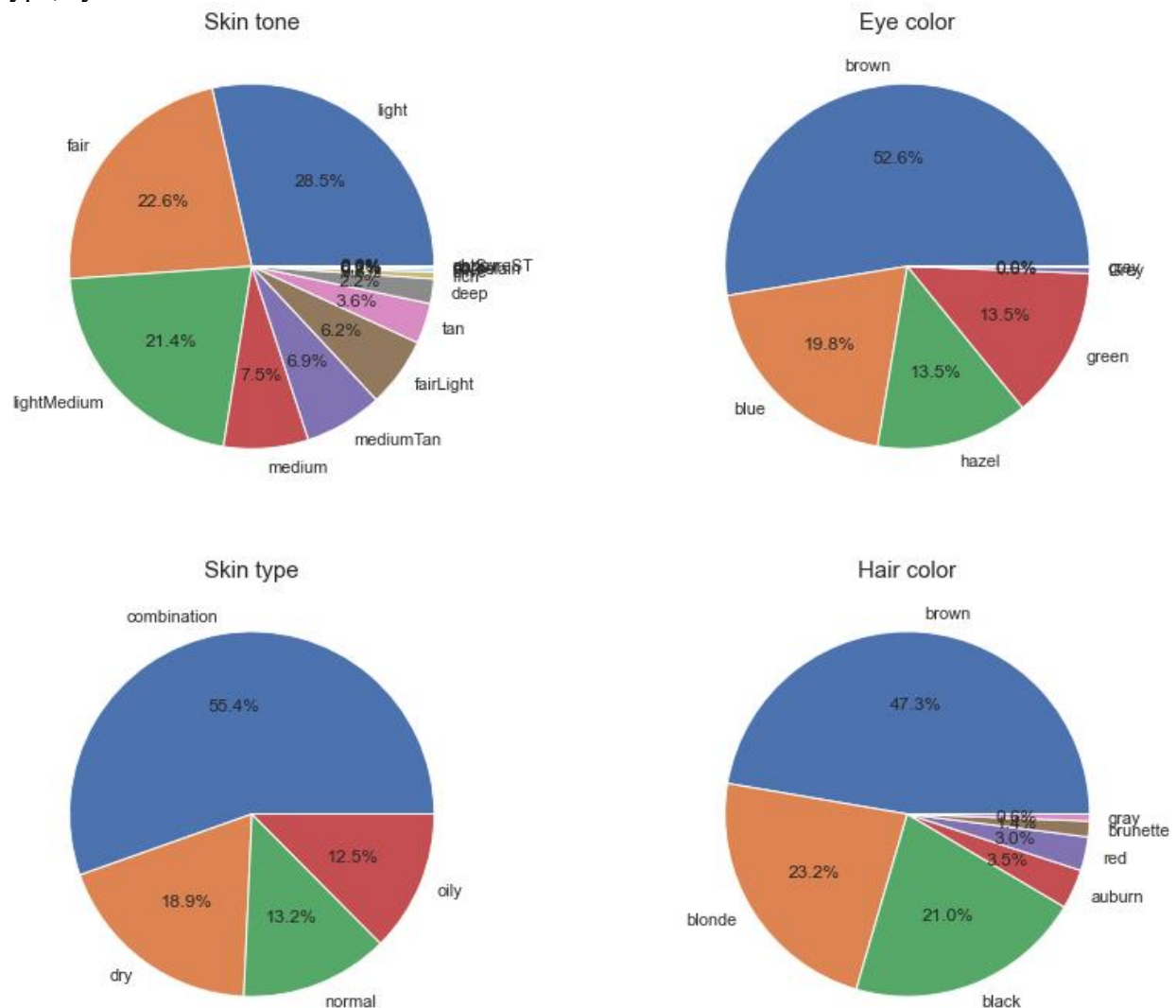


Let's look at an example of price distribution for a single brand.



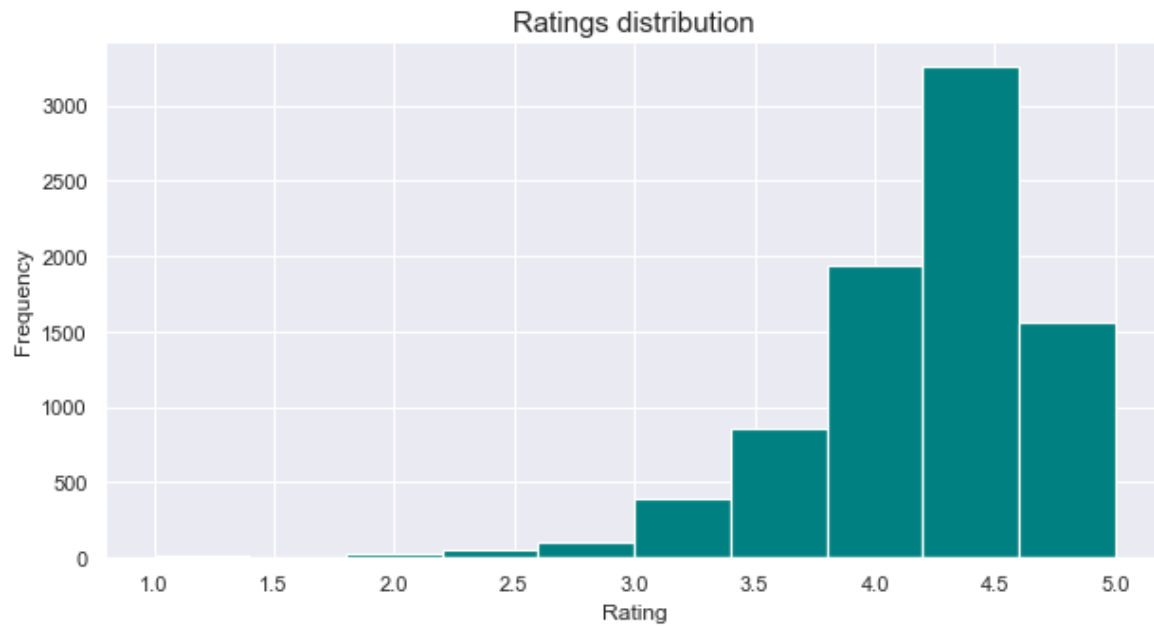
User features

There are many user features potentially relevant to the recommender engine, like skin tone, skin type, eye color and hair color.

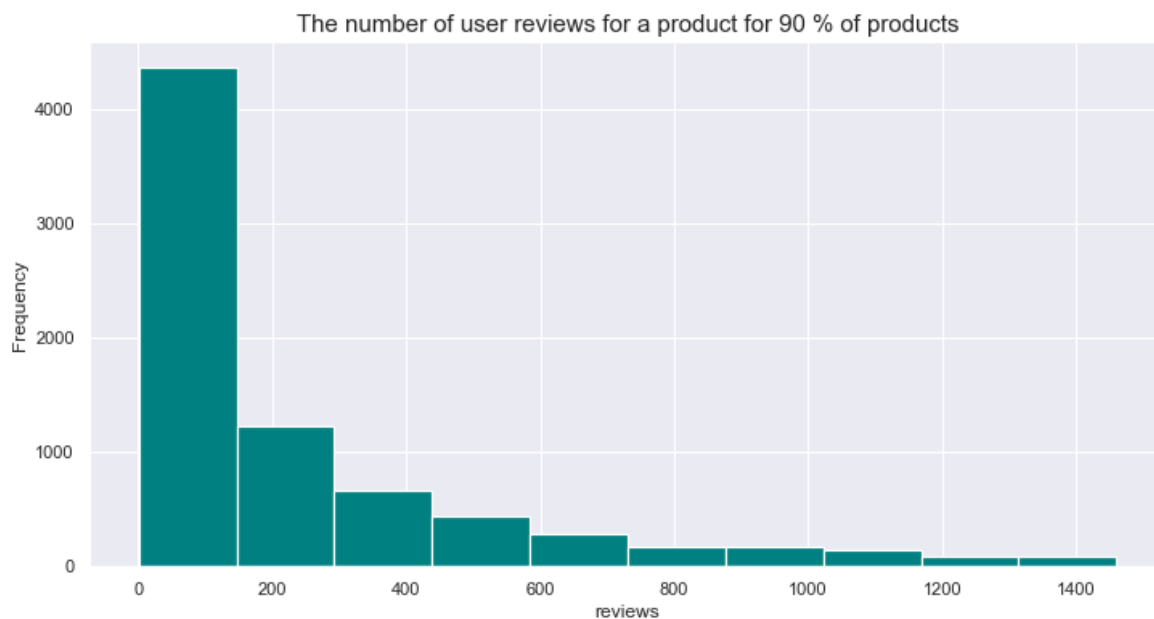


There are several dominant categories in each user feature. For example, we can see that around 75% of the users are in light/ fair/light medium skin tone category. Additionally, 52.6% of users whose info we have available have brown eyes, with other lighter eye color comprising the remainder. Next, most users have combination skin, followed by dry, normal, and oily. As for hair color, brown is at 47.3% with other large categories being blonde and black.

Ratings



If we look at average rating for all products, it looks like most products are around 4-4.5 star ratings, next there are many products with 5-star ratings, however, we do not see many low ratings, especially below 3 stars.



If we look at the number of reviews each product has, most products have under 200 ratings, with a few having over 500 ratings.

Modeling

In this section, we'll explore several different recommender system algorithms, each with its own strengths and applications. Our goal is to build models using these algorithms and evaluate them to determine which method provides the best recommendations for our Sephora dataset.

The four algorithms we will be using are Singular Value Decomposition (SVD), K-Nearest Neighbors (KNN), Non-negative Matrix Factorization (NMF), and a biased version of SVD. Each of these models is briefly introduced below:

Singular Value Decomposition (SVD): This is a popular matrix factorization method that decomposes a matrix into three other matrices. SVD can handle missing values in the user-item interaction matrix, allowing us to generate predictions for items that a user has not yet interacted with.

K-Nearest Neighbors (KNN): KNN is a memory-based method and one of the simplest methods of making recommendations. It finds users that are similar to the target user (or, alternatively, items that are similar to the target item) to generate recommendations.

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 can be useful for interpretability, as it allows for parts-based representations and can give a more intuitive understanding of the latent features.

Biased Singular Value Decomposition (biased SVD): Biased SVD is a variant of SVD that 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.

These models will be implemented using the Surprise Python library, which provides powerful tools for building and analyzing recommender systems. Each algorithm will be trained on our dataset, optimized through hyperparameter tuning, and then the model's predictions will be evaluated and compared to assess their performance. Through this process, we hope to identify the most effective recommender system for our Sephora product dataset.

The result of the 5- fold cross validation is presented in the table below:

Table 1. Model metrics.

	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

We can see that the lowest RMSE as well as MAE are for Biased SVD model. The fit time was longer; however, test time was shorter compared to regular SVD. NMF is losing on all metrics. Therefore, we choose to work with Biased SVD model since slightly longer fit time is not an issue for our problem.

We would like to demonstrate the top model's performance a little bit more in-depth.

RMSE (Root Mean Square Error): RMSE is a standard measure to compare the difference between predicted and actual ratings. Lower values of RMSE are better as they indicate smaller differences between the predicted and actual values. Our model has an RMSE of 0.6414. This indicates that, on average, our predicted ratings differ from the actual ratings by about 0.64 on a scale of 1-5. While this isn't a perfect score, it is quite decent for a recommender system, especially considering the difficulty of predicting exact ratings.

MAE (Mean Absolute Error): The MAE is another measure of prediction error. The MAE is the average over the test sample of the absolute differences between prediction and actual observation where all individual differences have equal weight. Our MAE of 0.4355 is lower than the RMSE, indicating that our model's predictions are relatively close to the actual ratings.

Binarized ratings classification metrics

Next, we would like to provide the classification metrics based on binarized ratings, where 4 or 5 stars is considered 'good' and 1-3 stars 'bad'. The results are below.

Metric	Value
Precision:	0.93
Recall:	0.99
F1:	0.96

Our model evaluation metrics show promising results. We'll examine each of these metrics to understand what they tell us about our recommender system's performance.

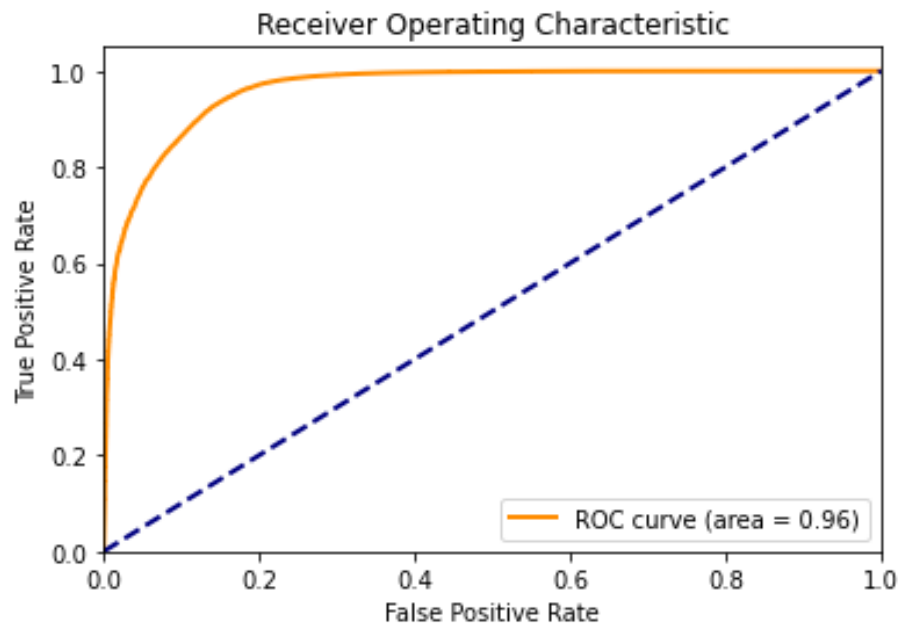
Precision: Precision is a measure of amongst all the items predicted positive, how many of them are actually positive. It is a critical metric in situations where False Positive is a high cost. In our case, a high precision score of 0.93 suggests that the products recommended by the model are indeed liked by the users most of the time.

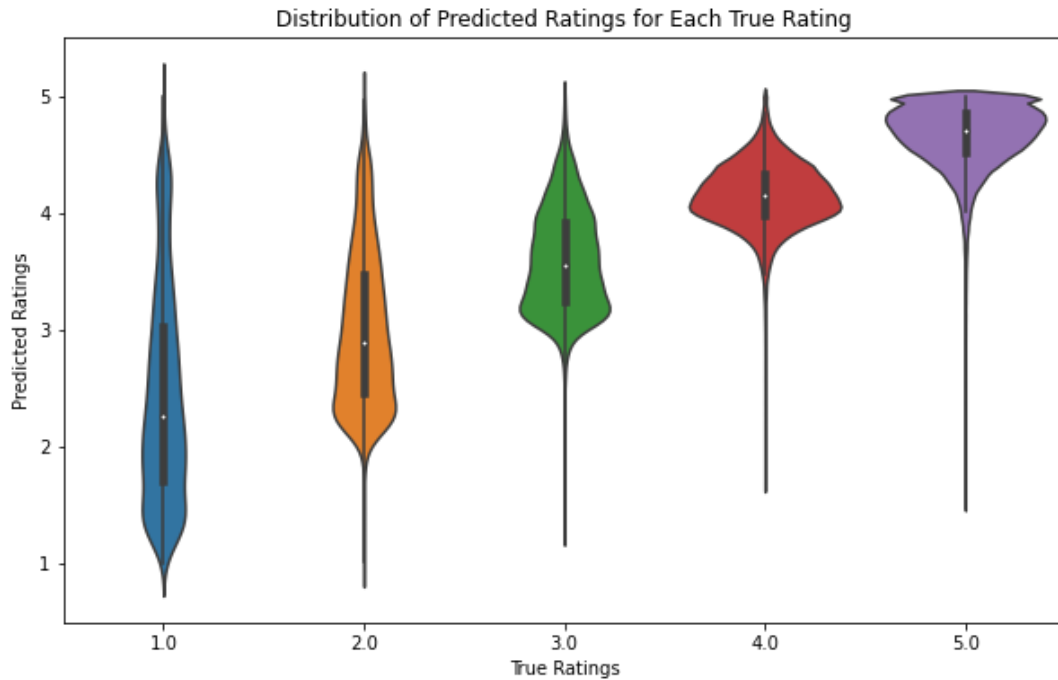
Recall: Recall is a measure of how many of the actual positive items are predicted as positive. It is an important metric when the cost of False Negative is high. The recall score of 0.99 indicates that our model is extremely good at capturing the products that users liked.

F1 Score: The F1 Score is the $2 * ((\text{precision} * \text{recall}) / (\text{precision} + \text{recall}))$. It is also called the F Score or the F Measure. Our high F1 score of 0.96 suggests that the model maintains an excellent balance between precision and recall, ensuring fewer false positives and fewer false negatives.

Given these evaluation metrics, our recommender system appears to be doing a solid job in recommending products that users are likely to enjoy and rate highly. The precision, recall, and F1 scores are particularly strong, suggesting the model is very effective at recommending relevant items and capturing user preferences. There is room to improve the model's RMSE and

MAE scores, which could be achieved through further hyperparameter tuning or trying different modeling techniques. However, considering the complexities and nuances inherent in human preferences, these results are quite encouraging.





Looking at the violin plots, it seems like the model captures the ratings accurately, perhaps with best performance for 4- and 5-star ratings. For 3-star ratings model does a good job, but often predicts 4-star ratings as well. For 2-star ratings the bias is again towards overestimation, sometimes giving 3 and occasionally 4-star ratings. For 1-star ratings the performance is the worst of all, model predicts 2s, 3s, even 4s. Overall, it seems like the model works best for predicting ‘good’ 4- and 5-star ratings.

Conclusion

Throughout this project, we have tried to build a robust recommender system for Sephora's skincare products. Using machine learning algorithms, we've navigated tasks of data preprocessing, model building, evaluation, and ultimately, predicting personalized product recommendations.

The key algorithms utilized in our recommendation system were Singular Value Decomposition (SVD), K-Nearest Neighbors (KNN), Non-negative Matrix Factorization (NMF), and Biased Singular Value Decomposition (biased SVD). Each of these models presented its own strengths, and their comprehensive evaluation provided us a deep understanding of their performance and potential applications.

The achieved evaluation metrics, particularly the high precision, recall, and F1 scores, indicate that the built recommender system is highly effective in suggesting relevant items and capturing user preferences. These results provide strong support for the utility of our recommender system, illustrating its potential to enhance user experience and drive business outcomes by personalizing product recommendations.

However, there are many ways to improve this work. The RMSE and MAE scores suggest that there is still room to improve the precision of our predicted ratings. Future work may explore additional algorithms, hyperparameter tuning, or feature engineering to reduce these errors.

Moreover, exploring hybrid models that combine the strengths of collaborative and content-based filtering could lead to further improvements. Additionally, incorporation of temporal dynamics, like evolving user preferences or changing product trends, could also refine the recommendations.

In conclusion, recommender systems are a powerful tool in the e-commerce industry, providing personalized user experiences and promoting product discoverability. This project on building a recommender system for Sephora's skincare products illuminates the potential of such systems in enhancing user satisfaction and in turn, driving business growth.

References

<https://surprise.readthedocs.io/en/stable/>

<https://www.kaggle.com/datasets/nadyinky/sephora-products-and-skincare-reviews>

<https://projects.scipy.org/scikits.html>

<https://builtin.com/data-science/recommender-systems>