Final report

Team - 3

Structure of the code

Document_1 - EDA
           —---->Discount calculation
           —---->Category wise distribution of products(Gender)
           —---->Top 10 popular brands
           —---->Brands that offer highest discount
           —---->Top selling fabric
           —---->Top 7 expensive fabrics
           —---->Variation in discount with gender(including unisex)
           —---->Top selling products and least sold
           —---->Most expensive items

Document_2 —--> Installations
           —---> Data Preprocessing
  —-----> Popularity Based filtering
    —-----> Memory Based
—---->User-User
—--> Train-Test-Split
—--> Defining Evaluation Metric -RMSE
—--> User-User pivot creation
           —--> Using Cosine Similarity to generate a similarity weights
           —--> Collaborative filtering model
           —--> Generating final RMSE score and Predicted Ratings
           —--> Generating top 10 recommendations
           —--> Getting recommended items DataFrame
        —-----> Item-Item
        —---> Item based collaborative model
        —---> Evaluation -RMSE
        —---->Predicted ratings of the Items - Data Frame
        —---->Generating top 10 recommendations
    —-----> Model Based
—-----> KNN
—----->Parameter Tuning
—----->Model Fitting
—----->Top 10 Recommendations
—----->Model Evaluation - Cross Validation
—-----> Matrix Factorization using SVD
—------>Hyper Parameter Tuning
—------>Running the Model
—------>Top 10 Recommendations
—------>Evaluation - Cross Validation

Dataset : Myntra fashion data

Overview : Myntra is a one stop shop for all your fashion and lifestyle needs.Being India's largest e-commerce store for fashion and lifestyle products,Myntra aims at providing a hassle free and enjoyable shopping experience to shopper's across the country with the widest range of brand's and products on its portal.The brand is making a conscious effort to bring the power of fashion to shopper's with an array of the latest and trendiest products available in the country.
Headquarters - Bengaluru
Website - https://www.myntra.com/

Outline - In this dataset I analyze the trends of different products according to the ratings, their characteristics,brands and categories.

Initially this dataset has 15000 rows and 27 columns.

Column description: In the preprocessing phase, we performed data cleaning to ensure the quality and relevance of our dataset. This involved removing unnecessary columns that did not contribute to our analysis goals.The columns are - User_id , p_id , Ratings , brand , size , title , actual_color , variant_price , variant_compare_at_price , dominant_material , ideal_for , discount.

Analysis -

1. The pie chart reveals that the majority of the products, accounting for 58.6%, are designed for women, followed by 28.3% for men. A smaller proportion of the products are allocated for girls (4.69%) and boys (4%), with a similar share of 4.3% classified as unisex. These percentages provide valuable insights into the target demographics catered to by the Myntra fashion platform, highlighting a significant emphasis on offerings for women.
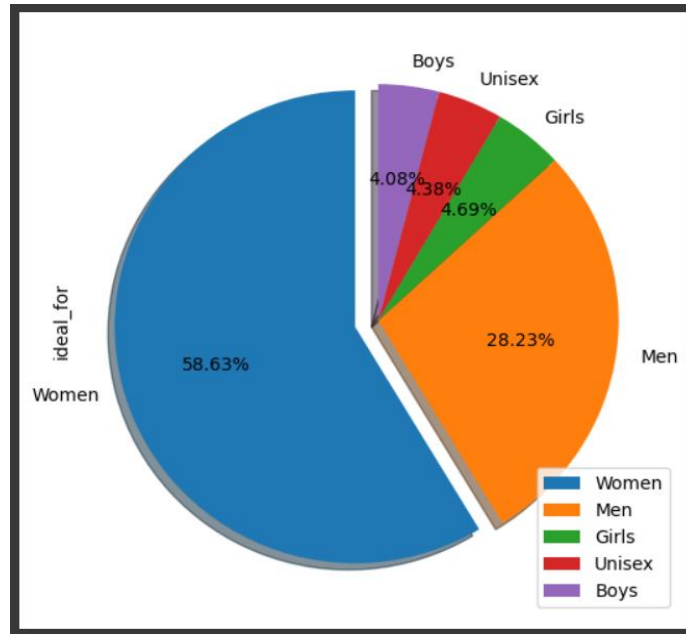
2. The analysis of top brands reveals that Anouk emerges as the most prominent brand, followed by Biba and Global Desi, respectively. These findings suggest a notable preference for these brands among consumers within the Myntra fashion platform.
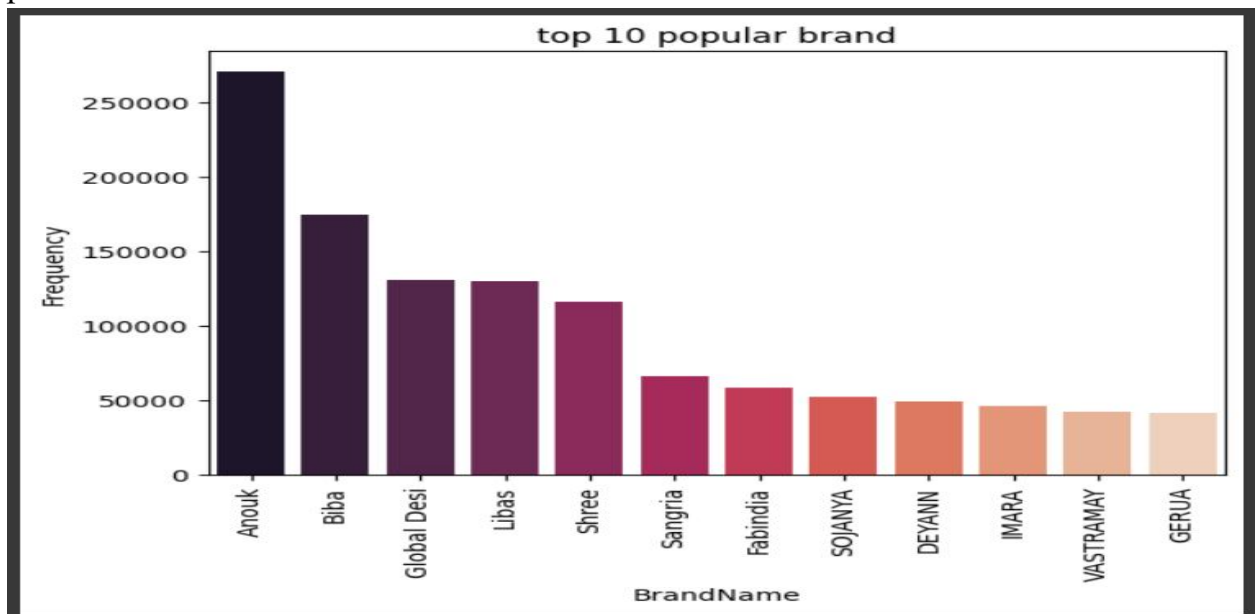

3. Upon analyzing the discounts offered by various brands, it is evident that Aaeeah takes the lead in providing the highest discounts, followed by Stylo Bug and I Am for You. This finding sheds light on the pricing strategies employed by different

brands within the Myntra fashion platform, with some brands prioritizing discounting as a means to attract and retain customers.
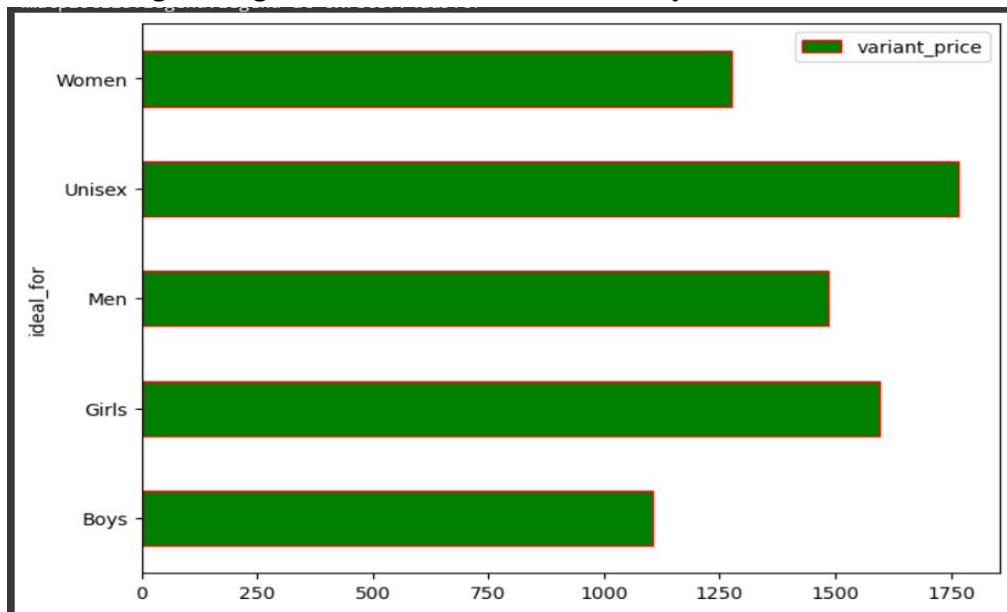

Top 15 Brands that offer highest Discount
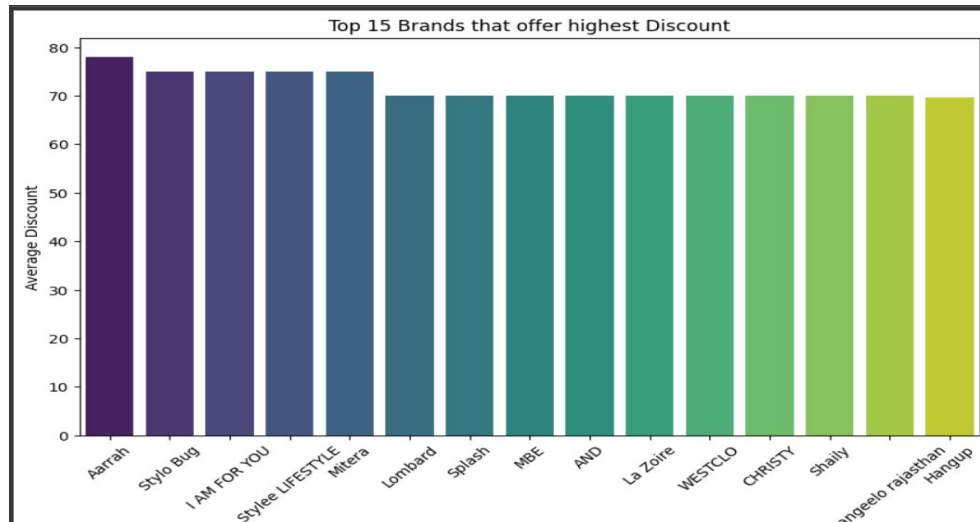
4. The analysis of cloth materials sold indicates that cotton emerges as the most popular choice among consumers, followed by viscose rayon and silk, respectively. This insight into material preferences provides valuable information for both retailers and manufacturers within the fashion industry, guiding decisions related to inventory management, product development, and marketing strategies.
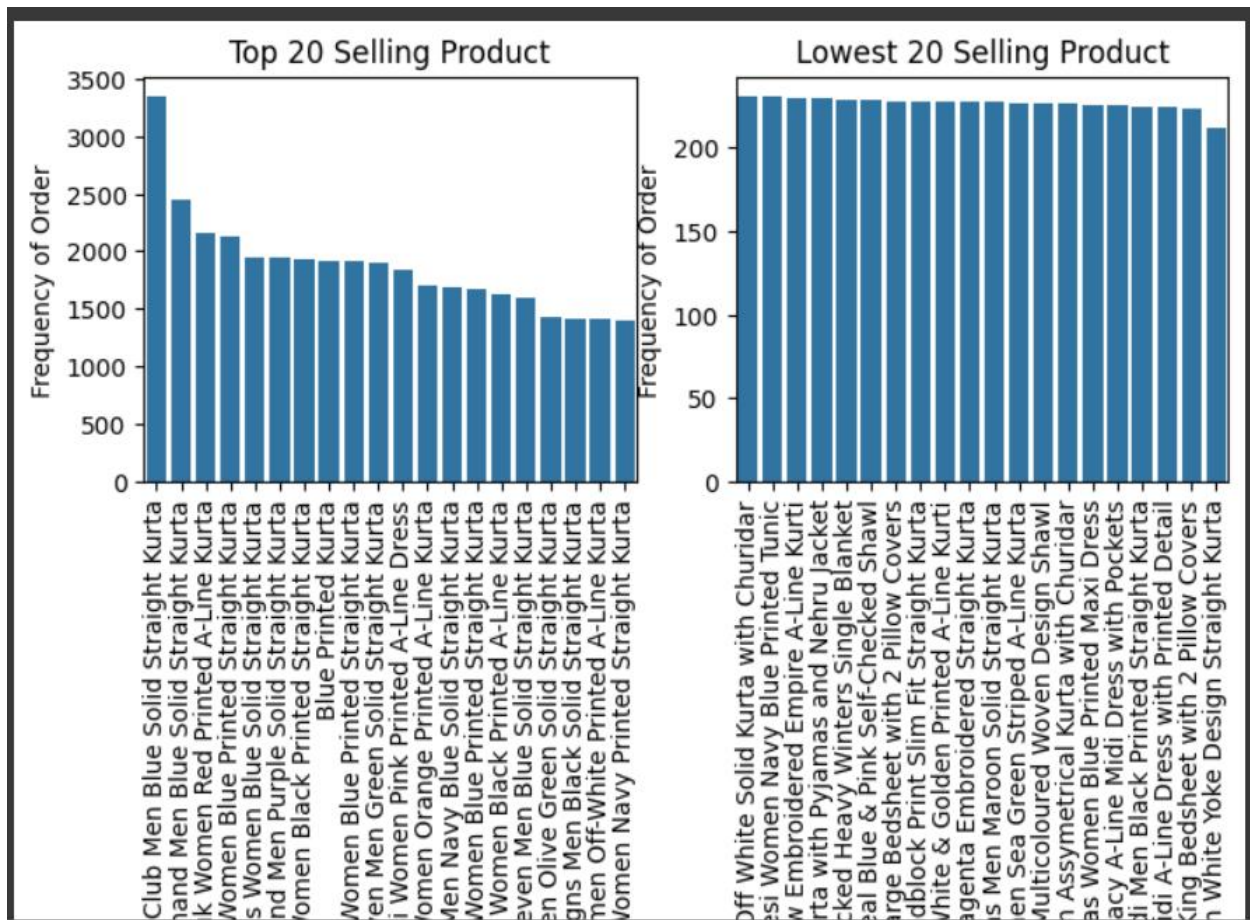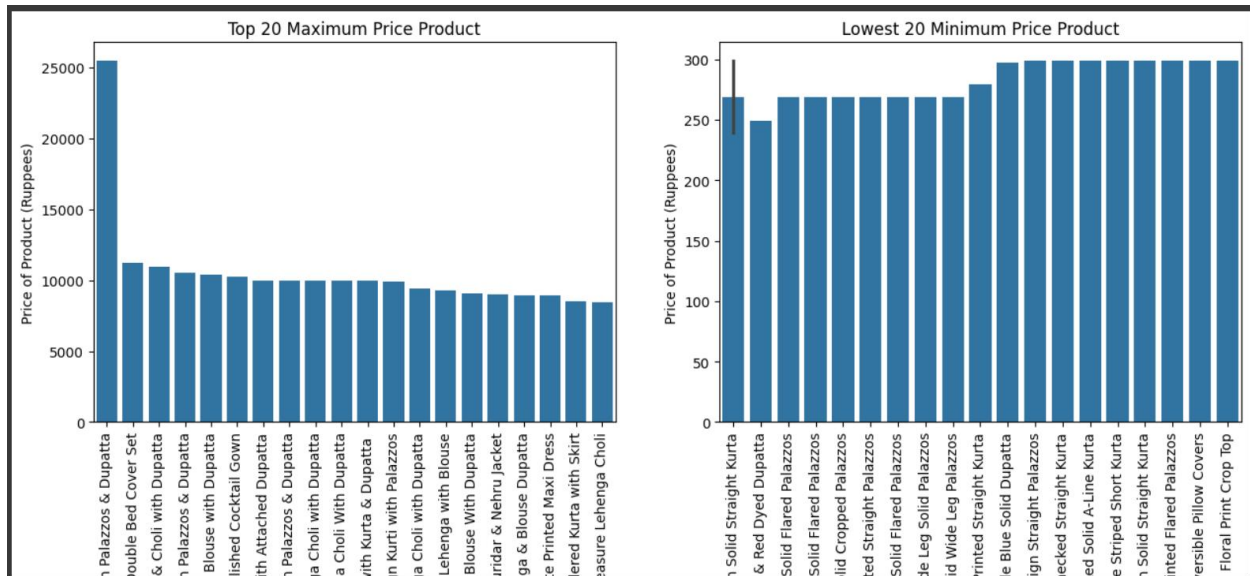

Top 5 Dominant Materials
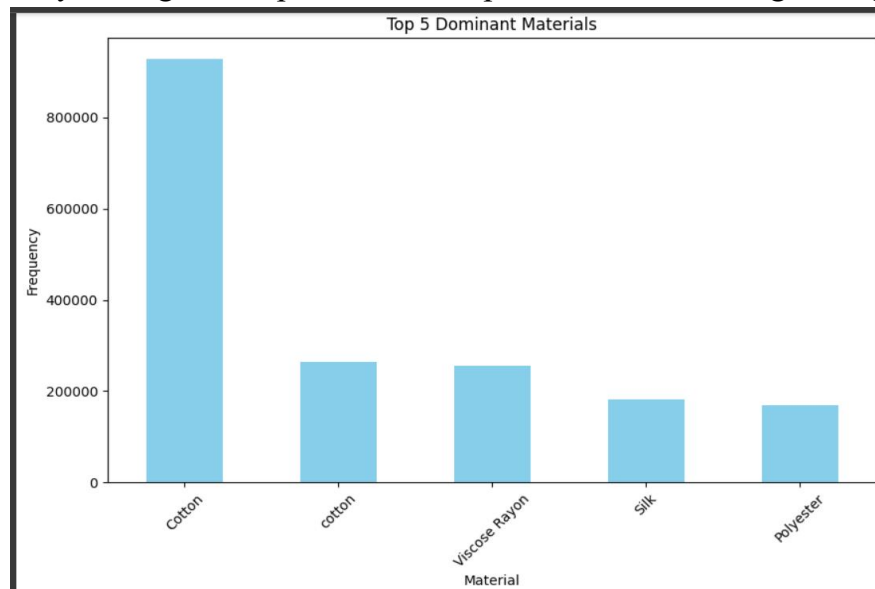
5. Velvet emerges as the most expensive material, followed by net and tensil, respectively. This insight into the pricing dynamics of different materials provides valuable information for both retailers and consumers, guiding decisions related to product selection, pricing strategies, and budget considerations.

Top 7 Most Expensive Dominant Materials
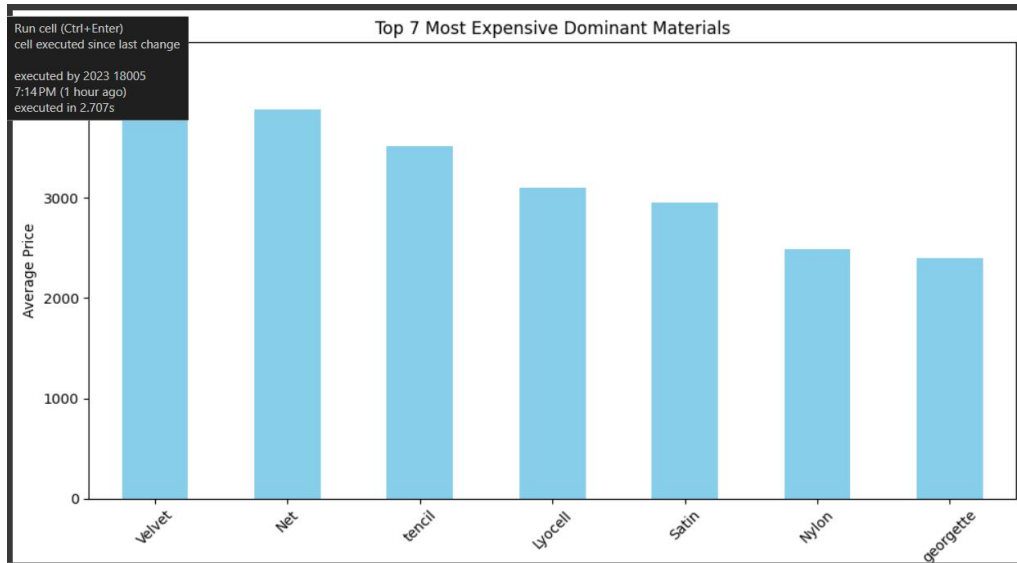
6. Unisex exhibit the highest average discount, followed by those targeted towards girls and men, respectively. This finding underscores the importance of considering demographic preferences and market dynamics when devising discounting strategies within the fashion industry.

Self explanatory plots

Top 20 Maximum Price Product

Lowest 20 Minimum Price Product

Top 20 Selling Product

Lowest 20 Selling Product

Insights and Inferences based on Model Evaluation

User-User Collaborative Filtering (CF):
  RMSE: 0.4107
Item-Item Collaborative Filtering (CF):
  RMSE: 0.2685

Inference:
The lower the RMSE, the better the model is at predicting ratings. Therefore, the Item-Item Collaborative Filtering model (0.2685) performs better than the User-User Collaborative Filtering model (0.4107) in terms of accuracy.

K-Nearest Neighbors (KNN)L
  k-value: 10
  min_k: 3
  similarity measure: Cosine similarity
  user-based: True
  min_support: 2
  Metrics:
  Test RMSE: 0.4563
  Test MAE: 0.3892
  Fit Time: 10.09 seconds
  Test Time: 13.02 seconds

Singular Value Decomposition (SVD)
  Metrics (mean of cross-validation):
  Test RMSE: ~0.4471 (average of [0.44790249, 0.44713923, 0.44691619, 0.44732718, 0.44717386])
  Test MAE: ~0.3835 (average of [0.38431032, 0.38364043, 0.38287316, 0.38387968, 0.38365732])
  Fit Time: ~4.14 seconds (average of [3.78, 4.00, 3.63, 4.58, 3.92])
  Test Time: ~0.37 seconds (average of [0.45, 0.28, 0.30, 0.50, 0.30])

Inference:
Performance: The SVD method tends to have a slightly lower RMSE and MAE compared to KNN based on the provided metrics. This suggests that, on average, SVD provides more accurate predictions.

Computational Efficiency: SVD appears to be computationally more efficient than KNN, as it has lower fit time and test time.

Additionally it can be observed that Memory Based is slightly easier to implement because it does not require extensive parameter tuning or use of any validation techniques. But model based is more powerful since it gives capability to work on larger datasets.