सिद्धिमूलं प्रबन्धनम्
भा. प्र. सं. इन्दौर
IIM INDORE

Integrated Program in Business Analytics

jigsaw
A UNEXT COMPANY

MYNTRA PRODUCTS RECOMMENDATION SYSTEM

IPBA_B17_GROUP L

PAVAN KUMAR T S
SAQLAIN SHEIKH
SHASHANK BOSE
SWARAJ DORA
CHAMAN MAMTANI
ANCHITA SINGH
ADITYA MODI
PRIYA NARUKA

सिद्धिमूलं प्रबन्धनम्
भा. प्र. सं. इन्दौर
IIM INDORE

Agenda

I. Business Problem and Objectives
II. Executive Summary
III. Data Overview and Key Business Assumptions (if any)
IV. Data Preparation and Pre-processing
   – Sanity checks, treatment and transformations for analytical dataset preparation
V. Exploratory Data Analysis
   – Key Business Findings and Insights
VI. Model Development and Validations
   – Model comparisons on key scoring metrics and model finalization
VII. Dashboarding (required only when it is in scope of analysis)
VIII. Business Recommendations and Potential Business Impact
IX. Next Steps
X. Appendix

jigsaw
A UNEXT COMPANY

सिद्धिमूलं प्रबन्धनम्
भा. प्र. सं. इन्दौर
IIM INDORE

Business Problem and Objectives

Overview of Myntra:

Myntra, founded in 2007, stands as a premier online fashion destination in India, offering a diverse range of apparel, accessories, and lifestyle products. Renowned for its extensive collection of domestic and international brands, Myntra provides customers with access to the latest fashion trends and styles. With innovative technology, personalized recommendations, and convenient shopping features, Myntra delivers a seamless online shopping experience, making it a preferred choice for millions of fashion enthusiasts across the country.

Business Problem:

- To effectively analyze user data and generate accurate recommendations.

- Consequently, the website experiences lower customer engagement and struggles to increase average order value.

Objectives:

Personalization

oved User Experience

Customer Retention

Increased Sales

jigsaw
A UNEXT COMPANY

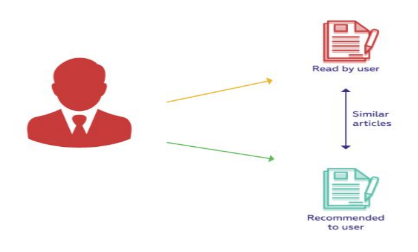
Executive Summary

Outlines a high-level overview of the presentation's content, allowing stakeholders to grasp the significance of Myntra's recommendation model and its impact on the company's objectives. The following components constitute the executive summary.

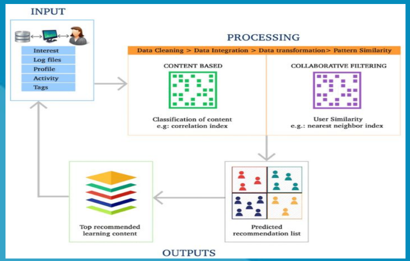
1) Recommendation Models used:

Popularity based

Content-based

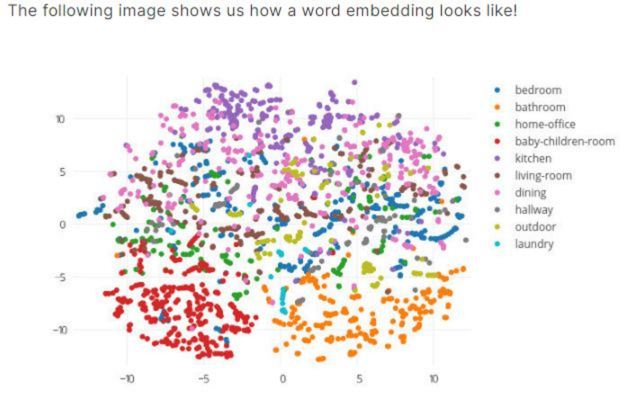


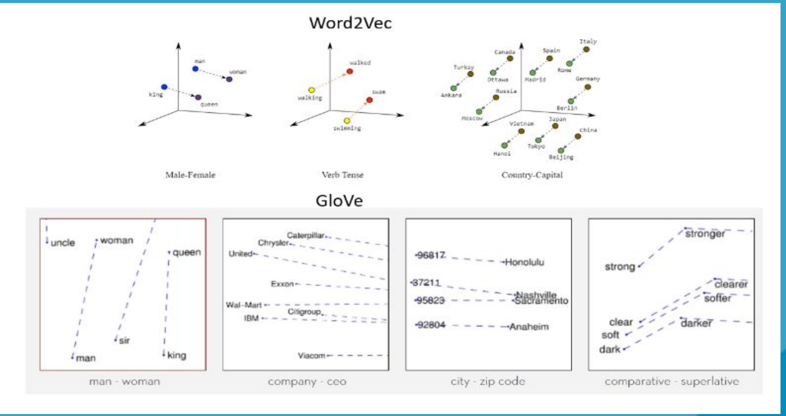
Hybrid approaches

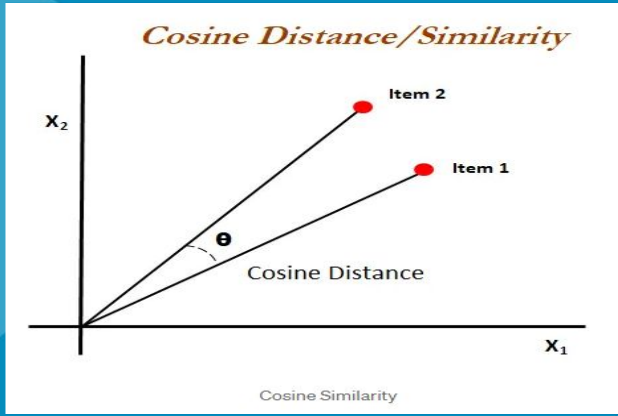


2) Technologies Used -> Python

3) Word Embeddings-> Count Vectorizer, TF-IDF, Word2Vec, Levenstien and GloVe models.







Data Overview and Key business assumptions

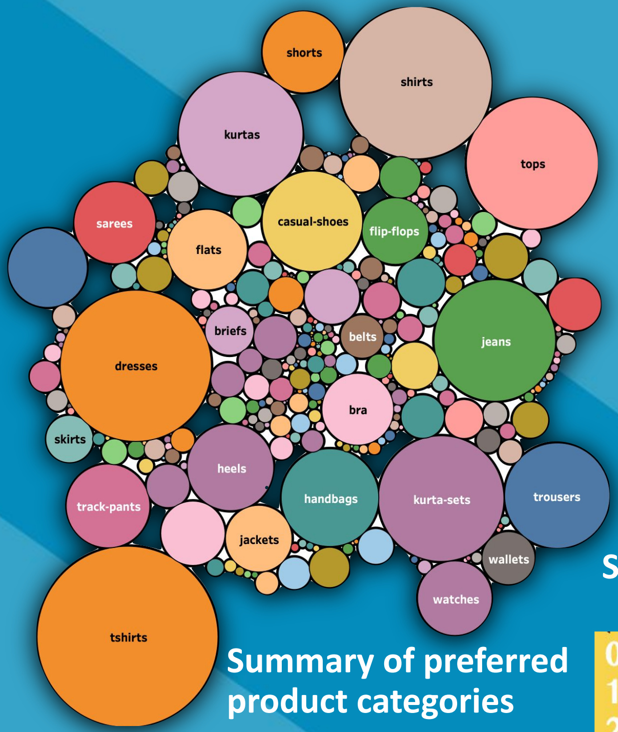
Data Overview:

Data Source: Myntra 168k Products (kaggle.com)
Product Data: Details about the products available on Myntra,including product name, brand name , rating of product, rating count, prices, sizes, links, and images.

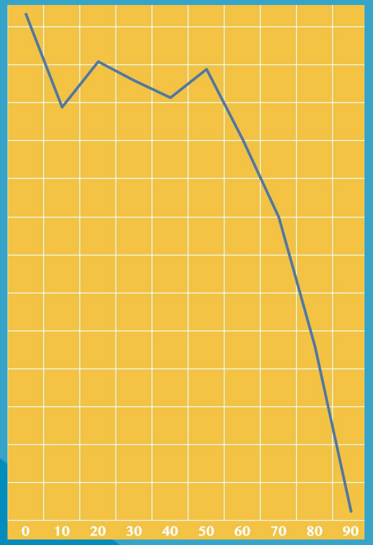
Ratings Analysis:
Mean Rating: 2.26
Mode Rating: 0

Price Distribution:
Average Discount: 37%, suggesting substantial savings for customers.

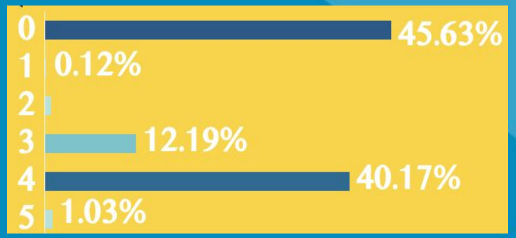
Spread of Prices:
Interquartile Range (IQR): Rs 109, indicating the variability in product prices.



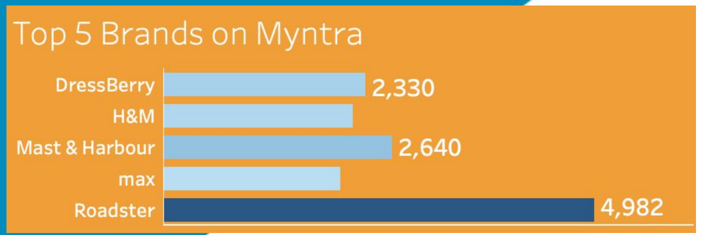
Summary of preferred product categories



Summary of discounts on products



Ratings summary of the data



Summary of most preferred brands

Key Business Assumptions:

1.Handling Data Sparsity: To manage data scarcity by excluding zero ratings and leveraging implicit feedback effectively.

2.Considering Price Sensitivity: Discounts (average 37%) have significant impact on purchase decisions, so the recommendation should consider the price fluctuations to prompt purchase decisions.

3.Accommodating Brand Preferences: By considering brand popularity and user inclinations.

4.Incorporating Size Preferences: taking into account size availability to suggest products in relevant sizes optimally.

5.Focusing on Popular Categories: Myntra caters heavily to clothing, so the model should prioritise recommendations within this category while considering user preferences for other categories as well.

Data Preparation and Pre-processing (Sanity checks, treatment and transformations for analytical dataset preparation)

Data preparation and pre-processing for a Myntra recommendation model involve several steps to ensure that the data is clean, structured, and suitable for analysis. Here's an overview of the process:

1) **Data Collection:**

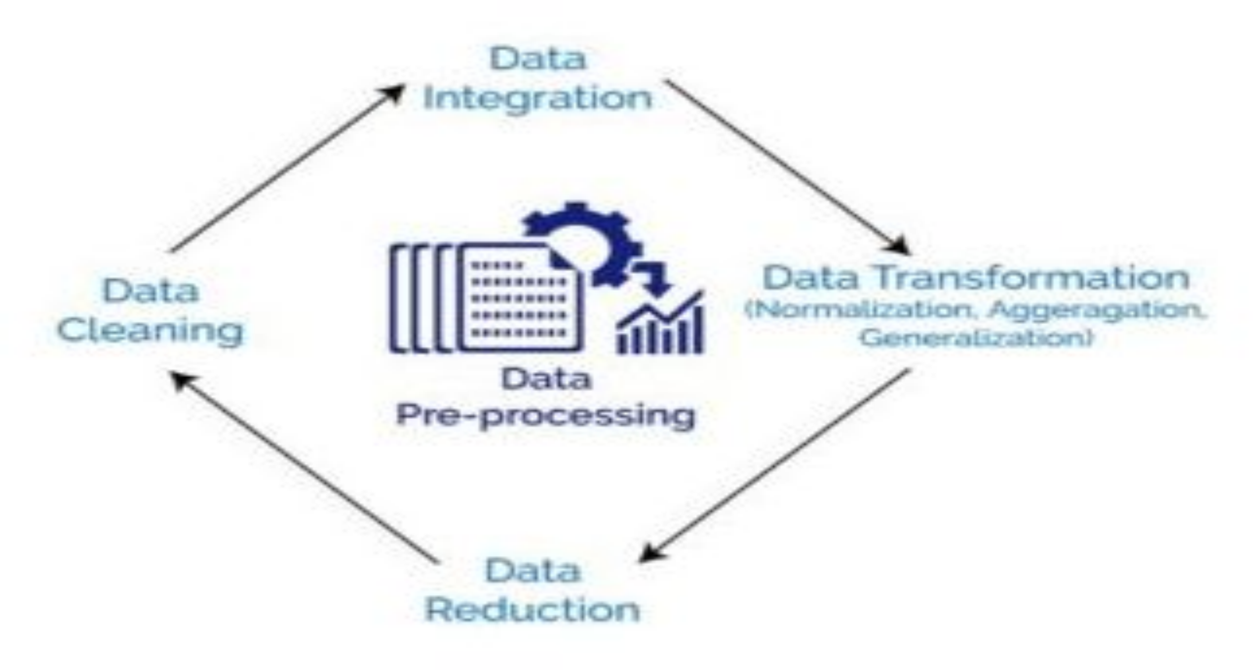
2) **Data Cleaning:**

3) **Data Integration:**

4) **Feature Engineering:**

5) **Text Pre-processing (if applicable):**

6) **Data Transformation:**



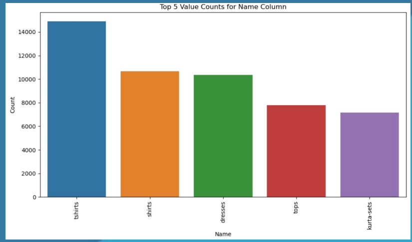
jigsaw
A UNEXT COMPANY

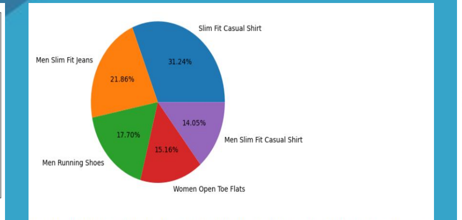
Exploratory Data Analysis (EDA)

In the exploratory data analysis (EDA) for the Myntra recommendation model, we aim to gain insights into the characteristics of the dataset, identify patterns, and understand the relationships between variables. Here's an outline of the EDA process for the Myntra recommendation model:

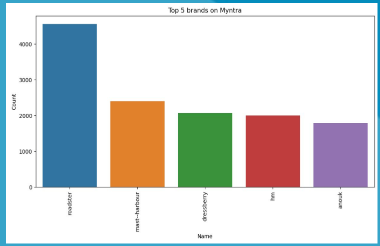
Dataset Overview: Myntra dataset with 168029 rows & 13 columns

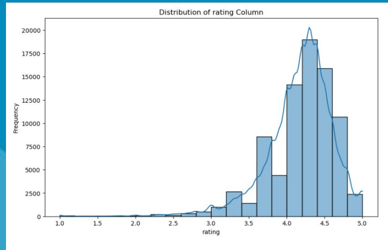
Product Analysis: Popularity based RM. Top products sold on Myntra. Logarithmic transformation .





Observation - Myntra's highest number of products sold pertain to 'Slim Fit casual shirt' followed by 'Men slim fit jeans' & 'Men running shoes'





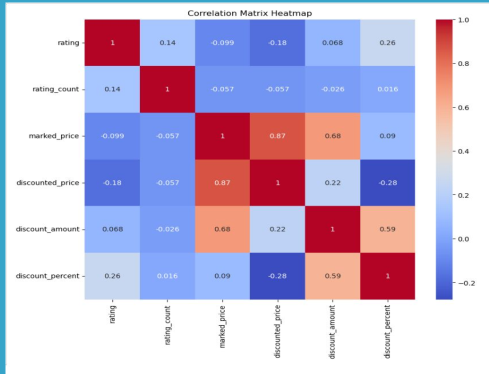
Maximum ratings fall between 4.0 and 4.8

```
myntra.columns

Index(['product_name', 'brand_name', 'rating', 'rating_count', 'marked_price',
       'discounted_price', 'sizes', 'product_link', 'img_link', 'product_tag',
       'brand_tag', 'discount_amount', 'discount_percent'],
      dtype='object')

myntra.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 168029 entries, 0 to 168028
Data columns (total 13 columns):
 #   Column            Non-Null Count   Dtype
---  ------            --------------   -----
 0   product_name      168029 non-null  object
 1   brand_name        168029 non-null  object
 2   rating            168029 non-null  float64
 3   rating_count      168029 non-null  int64
 4   marked_price      168029 non-null  int64
 5   discounted_price  168029 non-null  int64
 6   sizes             168029 non-null  object
 7   product_link      168029 non-null  object
 8   img_link          168029 non-null  object
 9   product_tag       168029 non-null  object
 10  brand_tag         168029 non-null  object
 11  discount_amount   168029 non-null  int64
 12  discount_percent  168029 non-null  int64
dtypes: float64(1), int64(5), object(7)
memory usage: 16.7+ MB
```

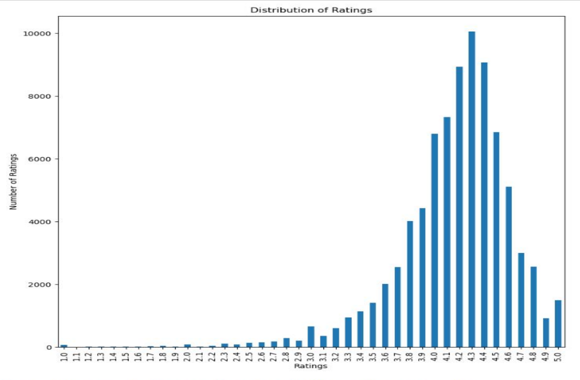
Price Analysis: The maximum discounted products on Myntra

Rating Analysis: Average rating given to most of the products is 4.17. Highest rated products on myntra is 5.0. Maximum ratings fall between 4.0 and 4.8. Highest Rated Product - products with the highest rating, considering the 'rating' column as the primary criterion and the 'rating count' column as the secondary criterion in case of a tie

Correlation Analysis: Rating and discounted price have a very high positive correlation.
We can see that histogram is skewed towards the right, which means that most of the products have a higher rating



As seen from above rating and discounted_price have a very high +ve correlation



We can see that histogram is skewed towards the right, which means that most of the products have a higher rating

Model Development and Validations

Algorithmic Approaches

Popularity based Recommendation Model by recommending the top N items to the user & category-wise (for each brand name)

Countvectorizer: Transforms product text data into numerical vectors by counting the occurrences of words, facilitating analysis of text data in popularity-based model.

TF-IDF Vectorizer: Utilized to convert product text data into numerical vectors based on term frequency-inverse document frequency, enabling analysis of textual features' importance for popularity-based models.

Word Embeddings: Embedding technique that transforms words into dense vectors, capturing semantic similarities and relationships between product names used in developing content-based model. E.g., Word2Vec, GloVe

GloVe Embeddings: Embedding method that generates word representations by leveraging global word co-occurrence statistics, enhancing semantic understanding and context awareness.

Fuzzy Wuzzy: Library used for fuzzy string matching, aiding in comparing and identifying similarities between strings, particularly helpful in data preprocessing and deduplication tasks.

Levenshtein: Algorithm used for measuring the dissimilarity between two strings by calculating the minimum number of operations required to transform one string into another, useful for tasks such as spell checking and string comparison.

Conclusion and Model Finalisation

With our final approach of Levenshtein Hybrid model below is working and understanding of the model.

In our recommendation system:

Model 1 - Cosine Similarity Recommendation:

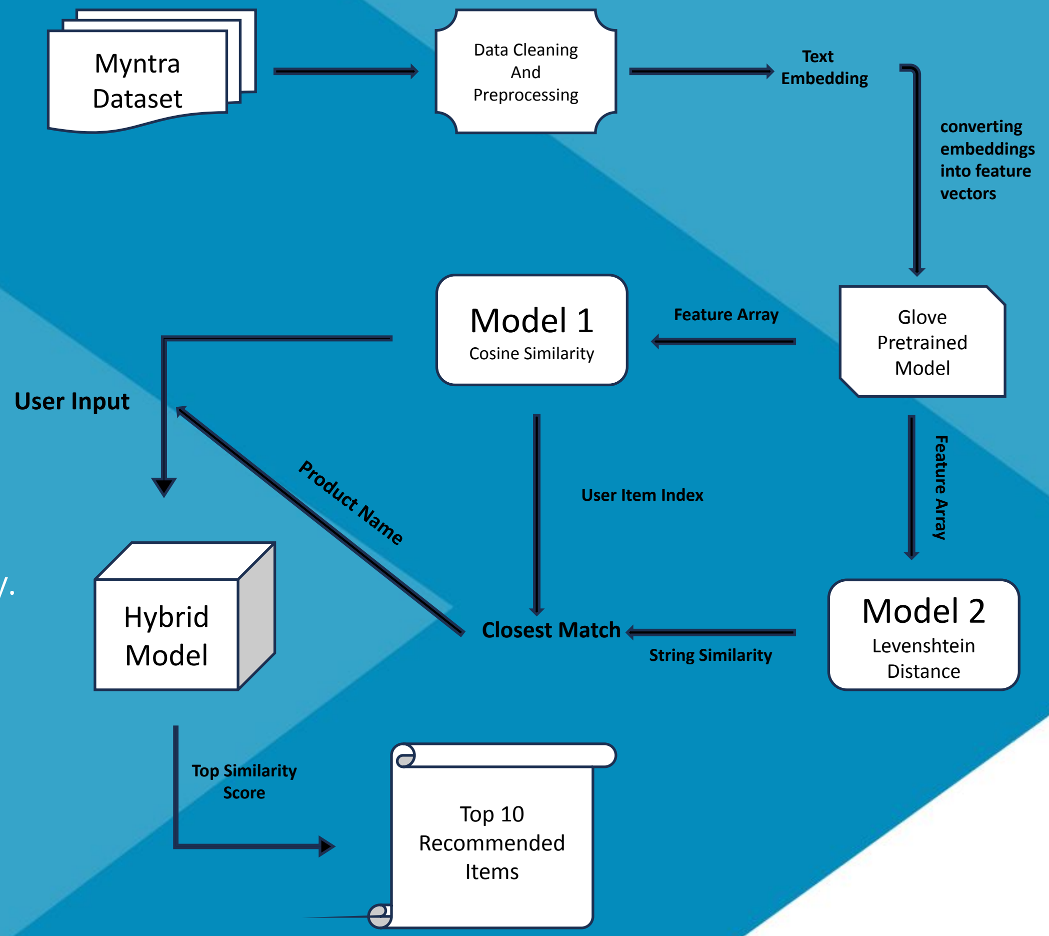
- Utilizes a GloVe embedding
- Computes cosine similarities
- Captures semantic relationships between products.

Model 2 - Levenshtein Distance Matching:

- Utilizes the Levenshtein Distance metric to measure string similarity.
- Directly matches user input to product names in the dataset.
- Facilitates user engagement by offering recommendations based on input similarity.

Hybrid Recommendation Approach:

- Combines both similarity-based and name-matching methods.
- leverages both GloVe embeddings and Levenshtein distance.
- Provides users with a comprehensive recommendation list tailored to their input.

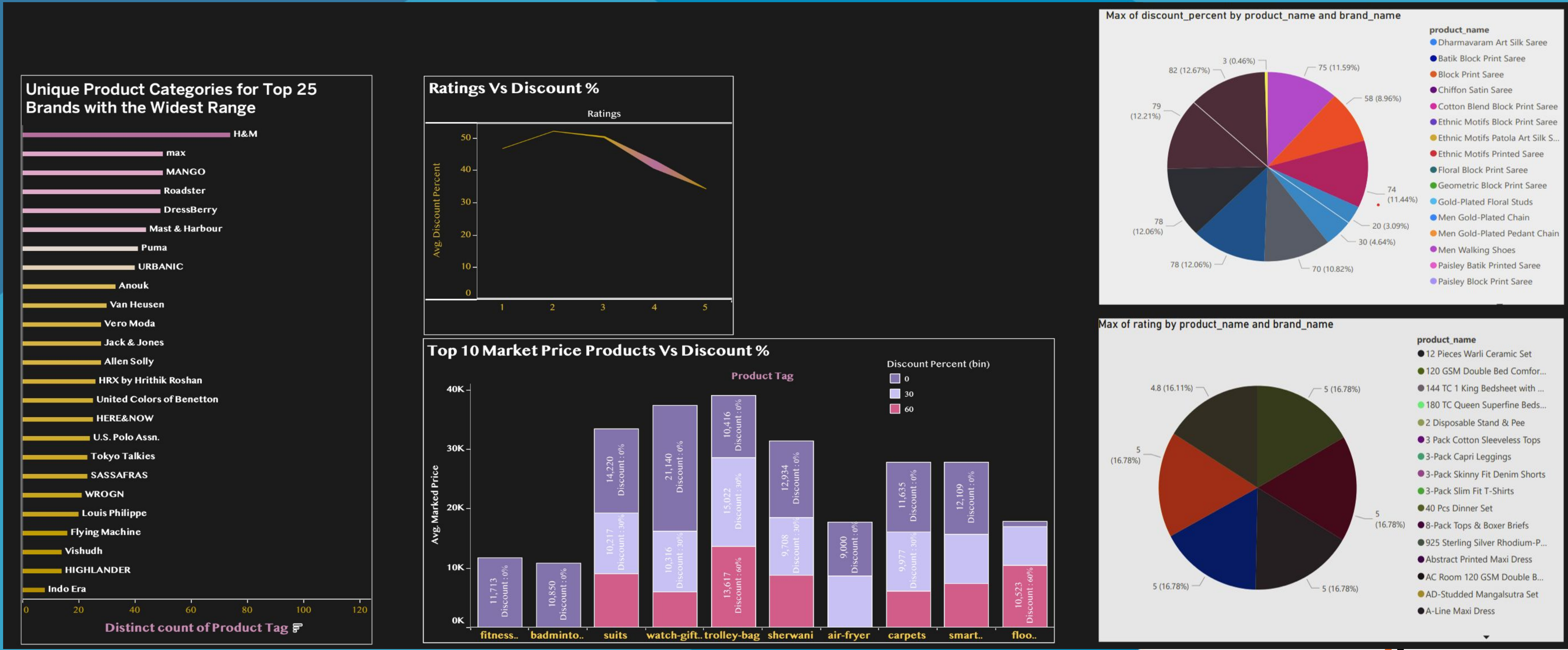


By integrating these models, our recommendation system offers personalized and accurate suggestions, enhancing user satisfaction and engagement.

Dashboarding

सिद्धिमूलं प्रबन्धनम्
भा. प्र. सं. इन्दौर
IIM INDORE

Here we are providing a detailed analysis of various brands and their products, showcasing average ratings, maximum discount percentages, marked prices, and rating counts. It covers brand performance and product-specific discount strategies, vital for understanding market trends and customer preferences. This information was possibly analysed using Power BI & Tableau Desktop, indicated by the mention of sharing a Power BI file for reference. The data is essential for insights into business strategies, pricing, and consumer satisfaction.



jigsaw
A UNEXT COMPANY

Business Recommendations and potential business impact

1. Personalized Recommendations:
   Utilize user data for tailored product suggestions.

2. Enhanced User Experience:
   Improve browsing experience with intuitive recommendations.

3. Reduced Cart Abandonment:
   Minimize abandonment rates by recommending complementary products.

4. Improved Inventory Management:
   Optimize inventory based on user preferences and purchase patterns.

5. Customer Loyalty:
   Enhance satisfaction and loyalty through personalized recommendations.

6. Competitive Edge Amplified:
   Propel Myntra ahead with a pioneering, personalized shopping journey.

Conclusion:
Implementing personalized recommendations on Myntra's platform can revolutionize user engagement, drive sales, and solidify its position as a leader in the e-commerce industry.

Next Steps

- Use other E-commerce datasets (E.g., AJIO, Amazon, Flipkart)
- Try new approaches viz Neural Network, Deep Learning & AI
- Enhance, evaluate and fine-tune recommendation accuracy
- Deploy the refined recommendation system into production
- Establish monitoring mechanisms to track its performance over time

Appendix

Title of Appendix

Thank You