Outfit Recommendation Application

(ORA)

August 12, 2024 **Prabhraj Singh**

Step 1: Prototype Selection

1. Project Overview

The Outfit Recommendation System is designed to cater initially to a female audience by having a market analysis in the outfit market particularly for jeans, then helping users find apparel that matches their style preferences. The recommendation system will utilize natural language processing (NLP) techniques such as Bag of Words and TF-IDF to recommend similar outfits. Over time, the platform aims to grow into a comprehensive e-commerce site similar to Amazon, offering a wide range of products beyond clothing.

2. Problem Statement

In today's digital age, customers expect personalized shopping experiences. While many platforms offer vast product selections, they often lack the ability to recommend products based on individual preferences, particularly in the fashion domain. Our goal is to bridge this gap by providing a recommendation system that not only understands the user's style but also suggests outfits accordingly.

3. Market Need

The fashion industry is rapidly evolving, with e-commerce playing a significant role. However, many platforms fail to deliver personalized experiences, especially for female shoppers. By focusing on personalized outfit recommendations, we aim to capture a significant market share, providing users with a tailored shopping experience that encourages repeat visits and purchases.

4. Target Audience

- **Primary Audience:** Initially, the platform will focus on female shoppers aged 18-35, who are tech-savvy and frequently shop online.
- **Secondary Audience:** As the platform expands, it will target a broader audience, including males and different age groups.

5. External Search

These are some of the sources I visited for more information and need for shopping pattern analysis of customers.

1. https://jayambe36.medium.com/fashion-recommendation-system-using-image-features-460b5cc88c7b

2. https://towardsdatascience.com/building-a-personalized-real-time-fashion-collection-recommender-22dc90c150cb

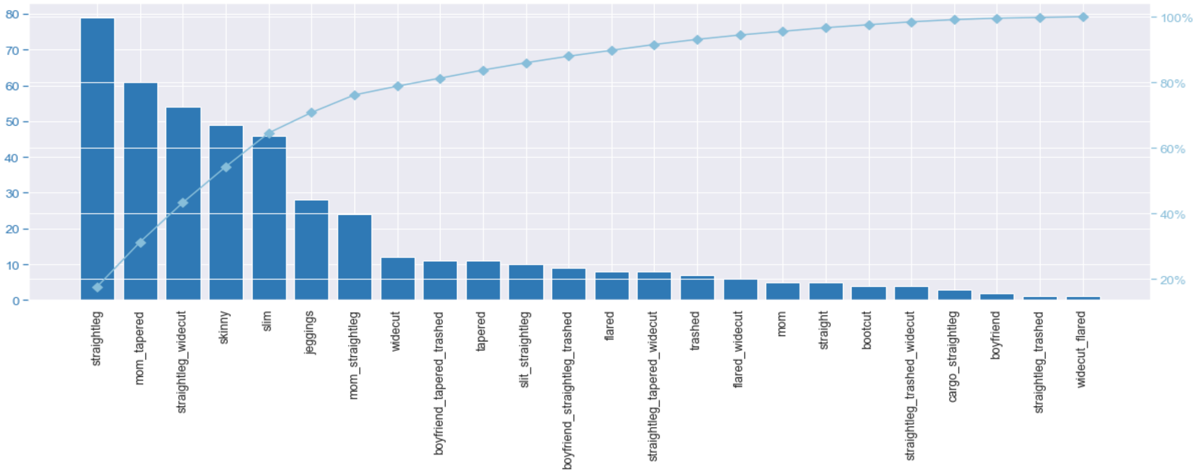
3. https://statso.io/fashion-recommendations-using-image-features-case-study/

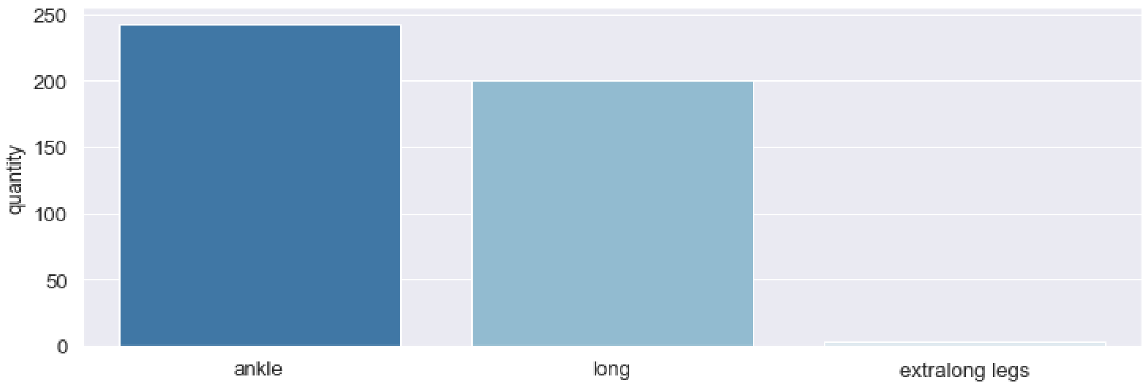
Dataset Description:

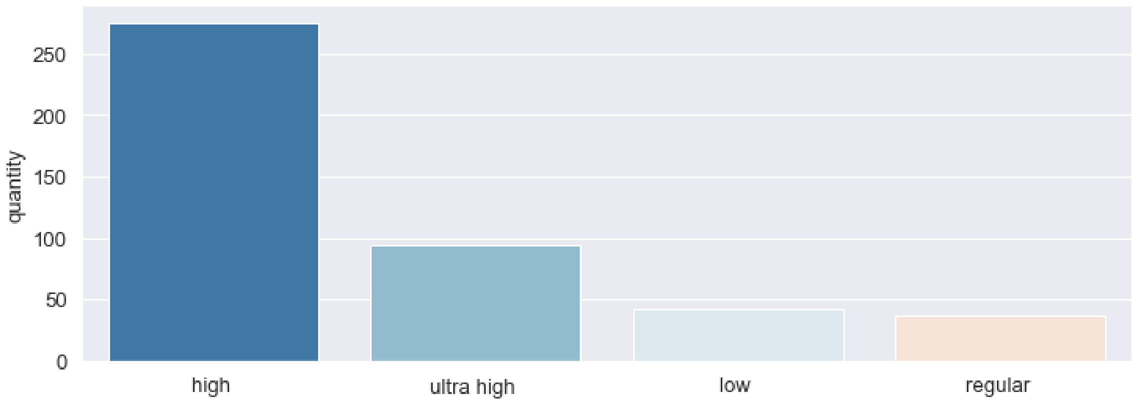
The dataset used in these models contains an item that a customer can look and buy if he/she likes it. Each row corresponds to the item that can be bought by one customer in one invoice.

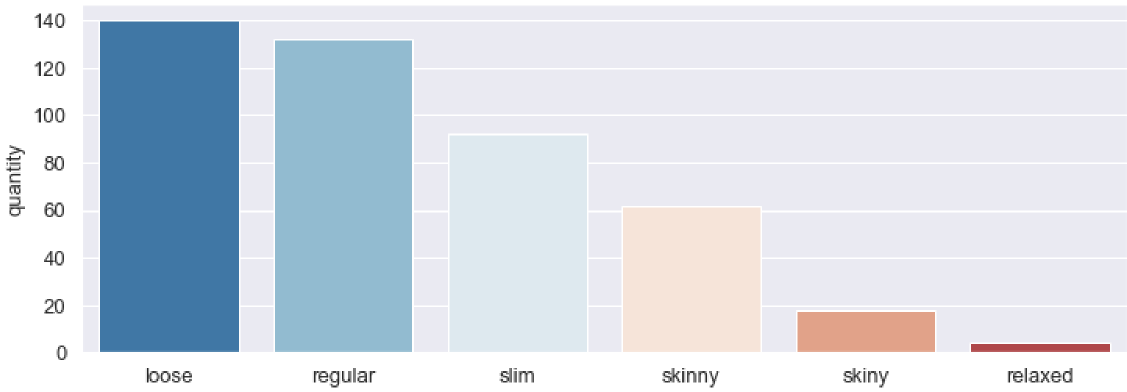
Data Overview

The medium price of the competitor's products is 29.99 USD. 75% of the products in the dataset are between 17.99 USD and 34.99 USD.
Pareto diagram was used to define which are the most significant styles and colors, i.e. those that represent 80% of the number of products in the dataset, and that according to Pareto reserve around 20% of the available styles and colors.
Of the 24 styles present in the dataset, 8 represent 80% of the products in the dataset: straight leg, tapered mom, straight leg wide-cut, skinny, slim, jeggings, mom straight leg, and wide-cut. In terms of length and waist, the ankle and long lengths represent practically 100% of the products, and 82.37% have either a high or ultra-high waist. About the fit, 80% of the products have fit loose, regular, or slim.









6. Technology and Methodology

1. Bag of Words

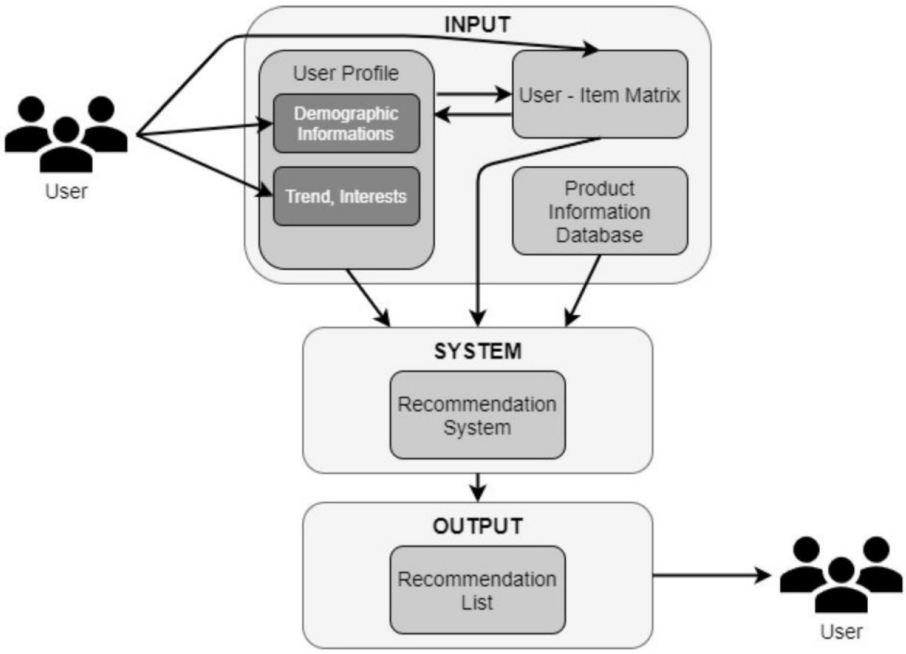
Bag of Words is a representation of text that describes the occurrence of words within a document. It is used to extract features from the text data, which can then be used in machine learning algorithms for outfit recommendations.

2. TF-IDF

Term Frequency-Inverse Document Frequency (TF-IDF) is another technique to convert text data into a numerical format. It helps in identifying the importance of a word in a document relative to a collection of documents. This method will be used to analyze product descriptions and user reviews to recommend similar apparel.

3. Recommendation Algorithms

The recommendation engine leverages NLP models such as Bag of Words and TF-IDF to identify and suggest products that closely match the user's preferences. By analyzing the vectorized text of product titles and descriptions, the system can identify products with similar features and recommend them to the user.



4. Metrics that can be used in Fashion Recommendation System Evaluation

The performance of a recommendation algorithm is evaluated by using some specific metrics that indicate the accuracy of the system.

- ***Root-mean square error (RMSE)***. RMSE is widely used in evaluating and comparing the performance of a recommendation system model compared to other models. A lower RMSE value indicates higher performance by the recommendation model. RMSE, as mentioned by can be as represented as follows:

$$RMSE = \sqrt{\frac{1}{N_p} \sum_{u,i} (p_{ui} - r_{ui})^2}$$

  where, N_p is the total number of predictions, p_{ui} is the predicted rating that a user u will select an item i and r_{ui} is the real rating.

- ***Precision***. Precision can be defined as the fraction of correct recommendations or predictions (known as True Positive) to the total number of recommendations provided, which can be as represented as follows:

$$Precision = \frac{True\ Positive\ \ (TP)}{True\ Positive\ (TP) + False\ Positive\ \ (FP)}$$

  It is also defined as the ratio of the number of relevant recommended items to the number of recommended items expressed as percentages.

- ***Recall***. Recall can be defined as the fraction of correct recommendations or predictions (known as True Positive) to the total number of correct relevant recommendations provided, which can be as represented as follows:

$$Recall = \frac{True\ Positive\ \ (TP)}{True\ Positive\ (TP) + False\ Negative\ \ (FN)}$$

  It is also defined as the ratio of the number of relevant recommended items to the total number of relevant items expressed as percentages.

- ***F1 Score***. F1 score is an indicator of the accuracy of the model and ranges from 0 to 1, where a value close to 1 represents higher recommendation or prediction accuracy. It represents precision and recall as a single metric and can be as represented as follows:

$$F1\ score = 2 \times \frac{Precision * Recall}{Precision + Recall}$$

- ***Coverage***. Coverage is used to measure the percentage of items which are recommended by the algorithm among all of the items.
- ***Accuracy***. Accuracy can be defined as the ratio of the number of total correct recommendations to the total recommendations provided, which can be as represented as follows:

$$Accuracy = \frac{TP + FN}{TP + FN + TN + FP}$$

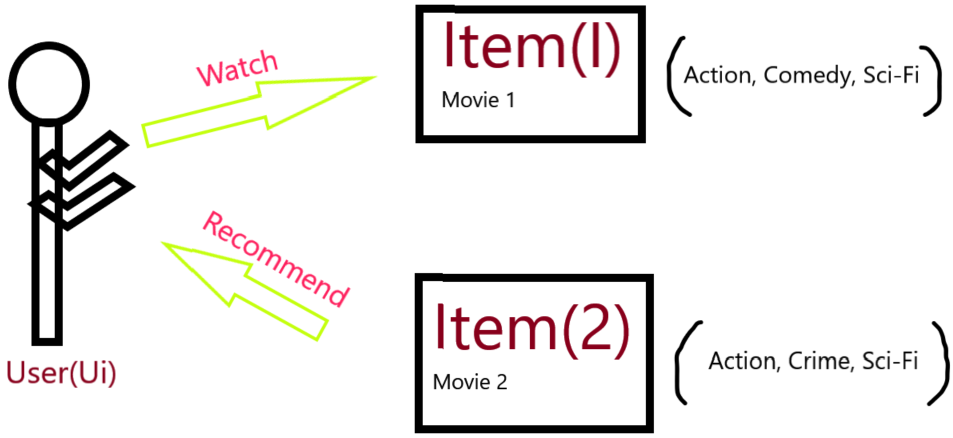
7. Business Model

There are several methods of how to implement recommender systems, and, in this case, we used a hybrid model of:

- Collaborative filtering model
- Content based model

For the current model we are using content-based model works under the assumption that what customer liked/bought in the past, would probably be liked/bought in the future.

For example, if a user watched a particular movie of a particular genre, then we can recommend him/her another movie of the same genre.



8. Monetization strategy

- **Freemium to Premium:** Offer a free basic version of the recommendation system to attract users. Monetize through premium subscriptions that provide advanced features and exclusive offers.
- **Affiliate Marketing:** Partner with fashion brands and retailers to earn commissions on sales generated through the platform.
- **In-App Purchases:** Allow users to purchase curated outfits directly through the platform.

9. Benchmarking

Lot of ecommerce stores like Amazon, Snapdeal, and Flipkart use these techniques to improvise their sales and to also create a smooth shopping experience for customers. Generally, Benchmarking involves comparing project processes and performance metrics to either industry best standards and practices or successful completed projects. For this there is a need to continuously search for implementation of better techniques which lead to better results or outputs.

We used two methods BOW and TF-IDF to check which method gave us more accurate results.

The BOW model is giving us quite a good variety of products and also displaying different types of brands.

In TF-IDF model, the products are quite restricted to being similar and of the same brand but BOW is giving us a better variety since it is changing the brands.

10. Applicable Regulations (Government and Environmental)

1. Data collection and Privacy of Regulations of Customers.
2. Government norms for Small Businesses and Street Vendors
3. Rules against False Marketing
4. Employment Schemes and laws created by government

11. Applicable Constraints

1. Lack of initial data to perform algorithms.
2. Convincing Shopkeepers and vendors to use this technique of selling over traditional means.
3. Lack of technical knowledge of vendors.

4. Rarely bought items will not be detected by algorithm, so it won't be generated as an output, so shopkeepers need to note which items are rarely bought and buy them in small quantities.

5. Need to continuously update and manage the data and model.

12. Business Opportunity

Though the process of getting started to breaking even and then towards profit may require some time it will be improving day by day.

When small shop owners and vendors start using these techniques, they will not only improve their sales but they will also have an in-depth analysis of what things customers are buying and what they are not buying.

That will also help them with maintaining their budget and which will eventually help them increase their reach and have growth in their business.

Gradually we can expand from just outfits to accessories that can go with those outfits, to giving more products that may not necessarily be in relation to outfits.

13. Concept Generation

It aims to provide personalized recommendations for clothing items to users by analyzing a comprehensive dataset of apparel products available. The system is built using natural language processing (NLP) techniques and machine learning models to analyze product descriptions and suggest similar items based on user preferences.

Data Acquisition

The system begins by sourcing data from a large dataset containing detailed information on Amazon's apparel products. This includes fields such as product SKUs, ASINs, titles, descriptions, prices, brands, colors, and customer reviews. The data is stored in a JSON format, allowing for structured analysis and manipulation.

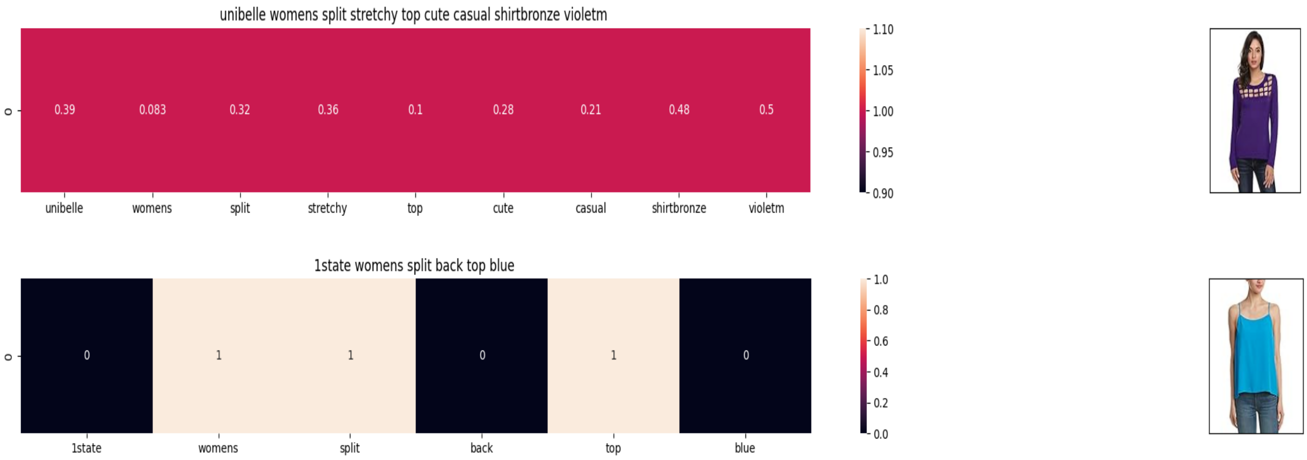
Data Preprocessing

Before any recommendation can be made, the raw data undergoes preprocessing, which includes:

- **Handling Missing Values:** Addressing incomplete data entries by filling in missing values or discarding irrelevant columns.
- **Text Normalization:** Cleaning and standardizing textual data like product titles and descriptions by removing stop words, punctuation, and applying tokenization.
- **Feature Extraction:** Converting text data into numerical representations using techniques like Bag of Words and TF-IDF (Term Frequency-Inverse Document Frequency).
- **Feature Engineering:** Feature engineering is a critical step in transforming raw data into meaningful inputs for the recommendation model. This includes:
- **Text Vectorization:** Converting text descriptions into vector representations that capture the frequency and importance of words.

- **Similarity Measures:** Utilizing cosine similarity to measure the distance between product vectors and recommend similar items.
- Our code-flow for a particular method (BOW or TF-IDF) will go as follows: ***get_result => text_to_vector => plot_heatmap_image => plot_heatmap => display_image***

By converting the text to vectors then using the Euclidean distance to measure the similarity we can generate similar styled outfits that the user may like.



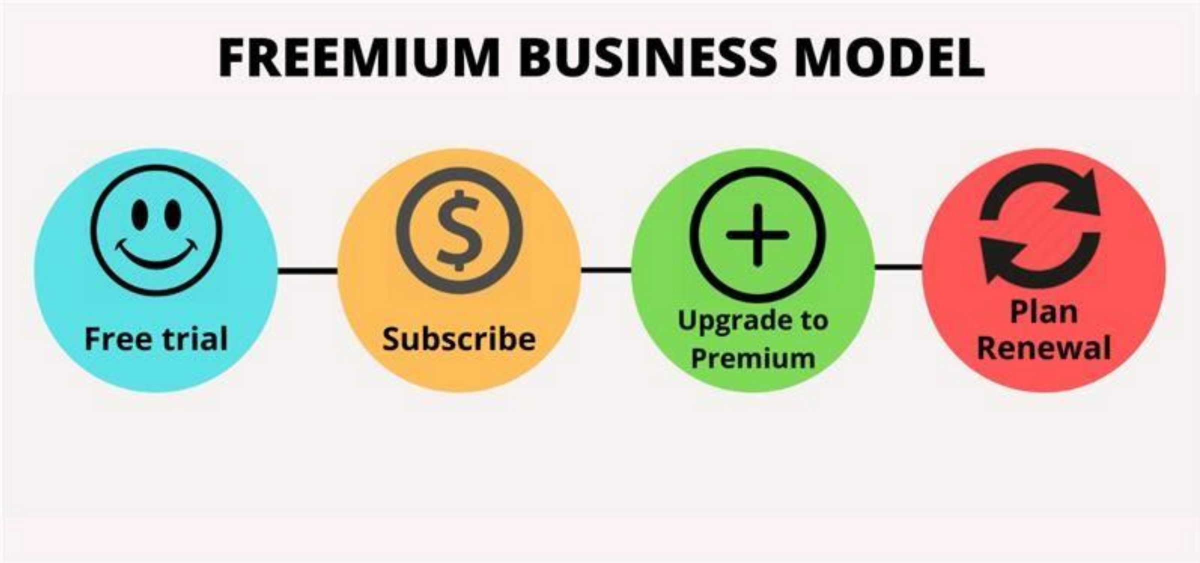
This is similar to query in retrieval systems. For example, when you type a query to search for some product on Amazon, the search engine will map your query against a set of keys (features and feature matrix) associated with candidate products in the database, then present you the best matched result.

Step 2: Prototype Development

Github link: https://github.com/prabhrajsingh/FeynnLabs/tree/main/ORA

Step 3: Business Modeling

- Freemium is a business model in which a company offers basic or limited features to users at no cost and then charges a premium for supplemental or advanced features.
- This type of business model has the advantage of acquiring a large set of initial users, especially when there’s no cost associated with trying out an app or a service.
- For the freemium model to work, companies must ensure that their premium users can access more upgraded features, such as increased storage or customizations, and additional customer service.



Step 4: Financial Modeling

Evaluating the Proposed Financial Equation for an Outfit Recommendation System

Understanding the Equation:

- **y = Total Profit**
- **m = Price of the product**
- **x(t) = Rate as a function of time**
- **c = Total cost**

Though the equation given is quite simple for a dynamic system like outfit recommendation system we would be taking assumptions on number of things for the equation to work.

1. **Product Variety:**
    - **Multiple 'm' values:** Introduce a price vector or distribution to account for different apparel prices.
    - **Product-specific costs:** Consider differentiating production and maintenance costs based on product type (e.g., dresses, shoes, accessories).

2. **Revenue Model:**
    - **Commission-based:** If you earn a commission on sales, the equation might involve a percentage of total sales rather than a fixed price.
    - **Subscription model:** For subscription-based services, revenue structure differs significantly.

3. **User Behavior and Recommendations:**
   - **Conversion rates:** Incorporate factors like click-through rates, purchase rates, and average order value to measure the effectiveness of recommendations.
   - **Recommendation algorithm impact:** Quantify the impact of recommendation accuracy on revenue.
4. **Cost Dynamics:**
   - **Variable costs:** Consider costs that fluctuate with sales volume (e.g., shipping, packaging).
   - **Economies of scale:** Factor in potential cost reductions as the business grows.
5. **Time-Dependent Factors:**
   - **Seasonality:** Account for fluctuations in demand and costs due to seasons.
   - **Fashion trends:** Incorporate the impact of changing fashion trends on sales and costs.

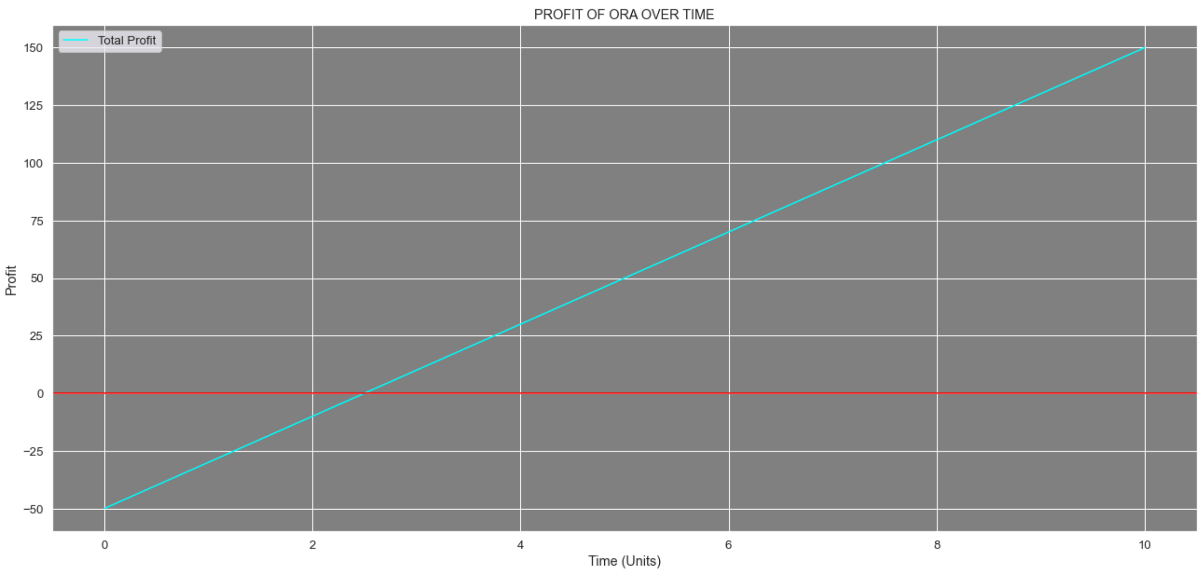
To plot the graph of y = mx(t) - c, we would need specific values for m, c, and a range for t.

For simplicity, let's assume:

- m = 10 (price per unit)
- c = 50 (total cost)
- x(t) = 2t (rate increases by 2 units per time period)

Equation becomes:

- y = 10 * 2t - 50
- y = 20t - 50



Conclusion

The Outfit Recommendation System represents a significant opportunity in the e-commerce space by addressing the growing demand for personalized shopping experiences. By starting with a focused approach and gradually expanding, the platform aims to become a leading player in the fashion e-commerce industry.

The Recommendation System provides a solid foundation for recommending products based on user preferences and product features. With further refinement and expansion, it has the potential to significantly enhance the shopping experience.