SmartStyleHub

Introduction

The primary objective of our project is to create a user-friendly and efficient platform that simplifies the process of outfit selection by offering personalized style suggestions. Specifically, our system utilizes NLP techniques to interpret user queries, extracting relevant keywords such as occasions ("interview," "casual outing") and clothing preferences ("men," "formal"). Additionally, our system integrates methods to search through a vast database of images featuring clothing items from renowned brands like H&M. By combining these technologies, our system effectively bridges the gap between user queries and visual representations of recommended outfits, offering a seamless and interactive experience.

Our project focuses on implementing a chatbot-style interface capable of understanding natural language queries related to fashion and providing visually appealing recommendations. The system will prioritize user input, analyzing keywords to identify specific clothing needs and occasions. Moreover, the integration of Computer Vision algorithms will enable the system to retrieve relevant images from the database, showcasing attire from recognized brands such as H&M. However, it's important to note that the system's recommendations will be based solely on the information provided by the user and the available database content. Future iterations of the project may explore additional features, such as real-time image recognition or personalized styling tips based on user preferences and feedback.

Motivation

In today's bustling lifestyle, the process of deciding what to wear can often be daunting and time-consuming, particularly when it comes to selecting attire for specific occasions. The endless array of choices in our wardrobe and the desire to present ourselves in a stylish and appropriate manner can lead to confusion and frustration. To address this common dilemma, our project aims to develop a sophisticated chatbot-style system that serves as a personalized fashion advisor. Leveraging the power of Natural Language Processing (NLP) and Computer Vision, our system will analyze user input in the form of text queries and provide tailored recommendations of outfits suitable for various situations.

Traditional methods of managing our wardrobe involve sorting clothes by type, color, or season, often leaving us overwhelmed by choices and settling for the ordinary. Modern fashion recommendation systems rely on purchase history or user-provided images, requiring the user to already have an idea in mind. We aim to revolutionize this process by combining the power of Natural Language Processing (NLP) with Computer Vision techniques. By leveraging Language Model capabilities and advanced image recognition, our system will allow users to directly ask fashion queries, enabling us to provide tailored recommendations from a vast fashion database without the need for predefined concepts.

Dataset

We used the H&M Personalized Fashion Recommendations dataset from Kaggle [1]. It consists of an images folder containing fashion article images and an articles.csv file with detailed metadata for each article id. The columns include - product name, section, department, garment group, color as well as a detailed textual description of each article. After performing exploratory data analysis [Fig.1, Fig.2], we selected adult men and women clothing along with footwear as categories. Data was preprocessed to only have article ids, their captions and image paths. Captions were created by concatenating product type, graphical appearance, gender and color columns, gender being ‘male’ and ‘women’. For improved accuracy, the longer detailed description was dropped. Final dataset has 55,000 unique fashion articles.

Methodology

The complete workflow of our project is shown in [Fig.3].

1. Low risk problem: The first step of our project involved selecting a large language model (LLM) to create a chatbot generating answers to user fashion queries. We decided to go with the Llama2 [2] - an open source LLM by Meta which can be easily accessed using Huggingface and LMStudio. To extract fashion related tokens from the Llama2 response text, we used prompt engineering and part-of-speech (POS) tagging to get relevant nouns and adjectives [Fig. 4]. The user interface was created using the Streamlit Python library [3].

2. Medium risk problem: To get the actual clothing suggestions in the form of images, we need to match the obtained tokens to the images in our fashion dataset. This text-to-image search was performed by OpenAI’s CLIP [4] - a pretrained multi-modal vision and language model. CLIP creates a shared embedding space for input texts and images, aiding in generating images most similar to the input text [Fig.5]. CLIP is pretrained on a large dataset consisting of diverse image-text pairs sourced from the internet. For improved model performance we finetuned it on the train dataset for 4 epochs, with a small learning rate of 0.000005 and cross entropy loss. CLIP processor takes significant time to create image and text encodings which again need to be passed to CLIP model for embeddings. To optimize this process, we stored image embeddings created by both base and finetuned models in Qdrant cloud [5] which is an open source vector database. To compare model performances we used R-precision metric [6] using cosine similarity between image and text embeddings as follows: Calculate how similar the model thinks images and their true captions (matching pairs) are and how different images and rest of the image captions (non-matching pairs) are. This R-precision was computed for both whole test data and test data segregated by product type and color name.

3. High risk problem: Currently we are using simple keyword extraction methods for tokenization. However, the majority of our work is going into prompt designing to ensure that the generated response is concise and in an appropriate format. This approach is not fully reliable as we might miss a few important tokens from time to time. As part of the future scope of this project, we could create a custom fashion keywords tokenizer – trained on a large dataset containing fashion reviews, suggestions, or comments having annotated keywords. As of now no such open-source dataset is available. Building it would require a lot of web scraping and manual annotation which could take weeks if not months. Therefore, the feasibility for this task is low at this stage.

Results

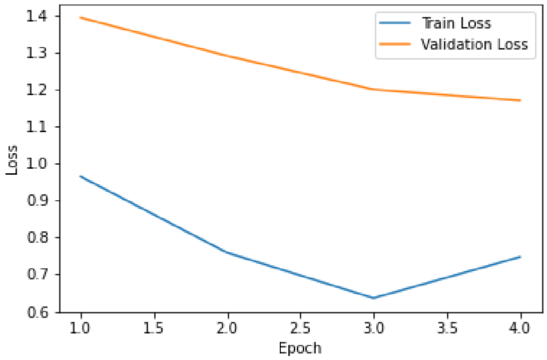


Fig 6. Loss Curve

Model-Test data pair	Similarity cosine score (higher the better)	Difference cosine score (lower the better)
Base-whole	0.3076	0.2043
Finetuned-whole	0.5930	0.4884
Base-segregated	0.3093	0.1868
Finetuned-segregated	0.3533	0.1862

Table 1. Performance Metrics

Both training and validation losses decrease with increase in number of epochs as evident from the loss curve. However after epoch 3 train loss starts to increase, signifying a slight overfit. This could be mitigated in future by increasing data size and adding more diverse samples. The R-precision table [Table. 1] shows that the on whole dataset, finetuned model has significantly higher similarity cosine but a higher difference cosine as well. High difference cosine could be explained by selected captions having similar values in them such as same color or product type name. The second dataset which was segregated on the basis of color and product type has dissimilar captions leading to improved metrics for finetuned model in terms of both similarity and difference cosines. Overall, the finetuned models still exhibit better performances than the base models. Additionally, the final Streamlit application output is shown in [Fig. 7]

Conclusion

In conclusion, the quest for curating a versatile and fitting wardrobe for various occasions persists as a shared challenge among many individuals. Fashion, transcending mere aesthetics, holds the power to imbue confidence and convey one's personal style. Our project envisages a transformative shift where the wardrobe evolves from a mere collection of garments to a personalized style advisor, facilitating individuals in effortlessly navigating their fashion journey. Through the integration of AI technology, our aim is to revolutionize the daily process of outfit selection, turning it into a pathway for self-expression and convenience. This initiative represents a significant stride toward enriching the fashion experience and fostering enhanced confidence and empowerment in individuals' daily lives.

We accomplished initial milestones, such as implementing the Llama2 model with LMStudio API, and integrating with Streamlit UI. Our medium-term objective involved the implementation of text-to-image search functionality using OpenAI's CLIP model, fine-tuned to bolster accuracy. However, we recognize the inherent limitations of our current reliance on basic keyword extraction methods. The development of a custom fashion tokenizer, trained on a dataset, poses feasibility challenges due to the scarcity of open-source data, necessitating arduous efforts in web scraping and manual annotation. Overcoming these hurdles is paramount to advancing the system's capabilities, ensuring the delivery of more refined and precise fashion recommendations to users.

References

[1] H&M Personalized Fashion Recommendations
[2] Llama2
[3] Streamlit
[4] CLIP
[5] Qdrant Vector Database
[6] R-Precision Metric

Appendix

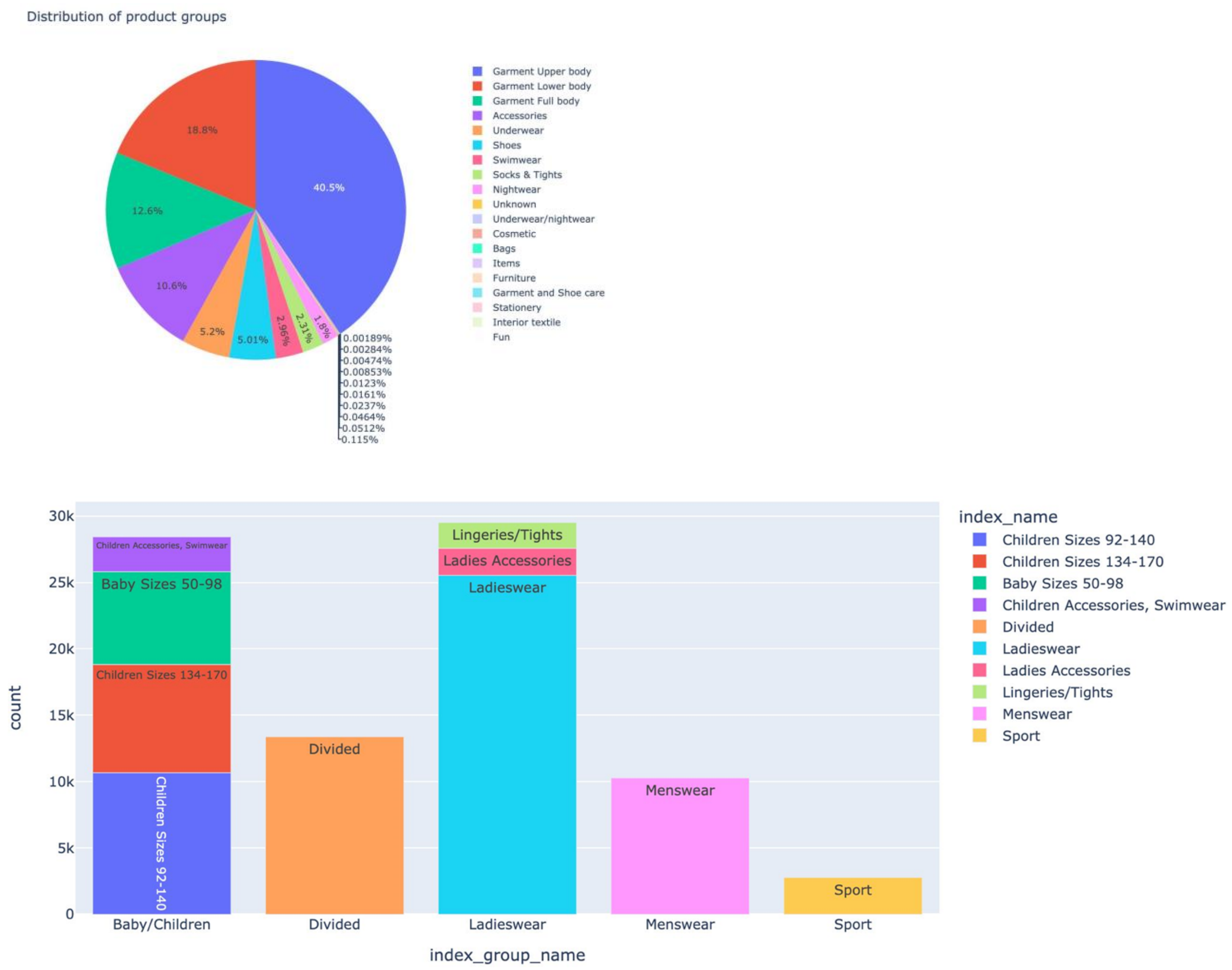


Fig 1. Exploratory data analysis (1)

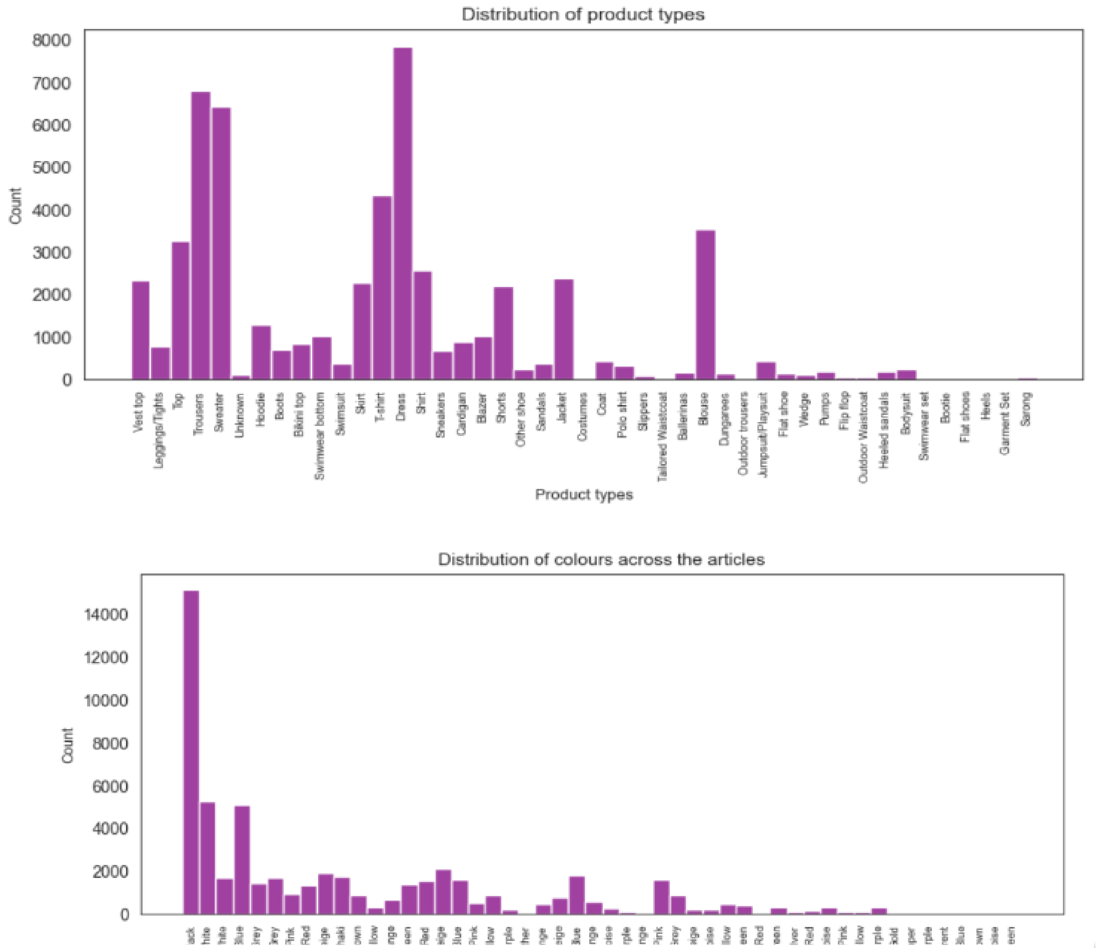


Fig 2. Exploratory data analysis (2)

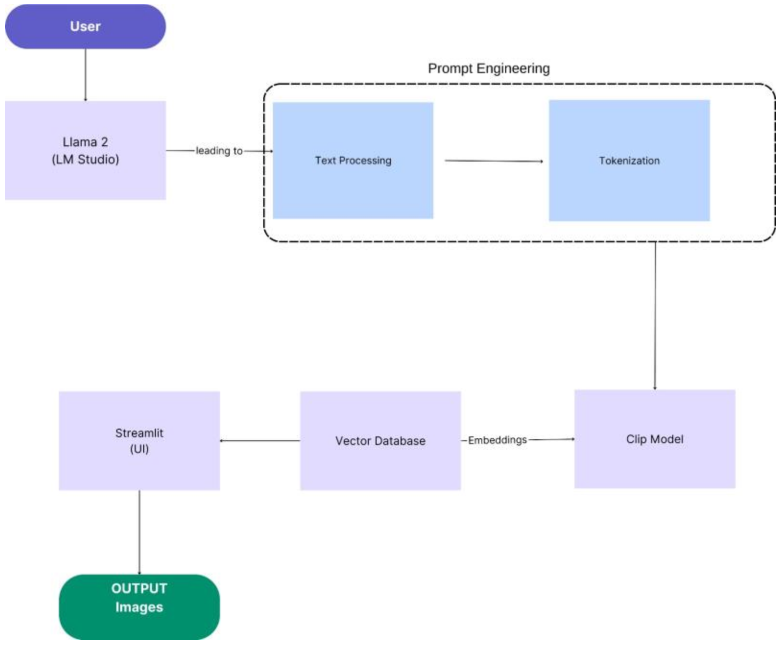
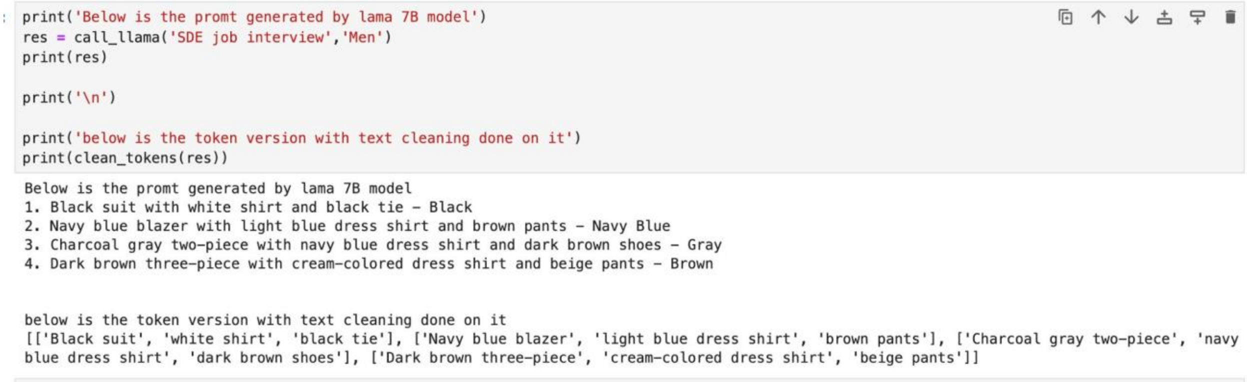


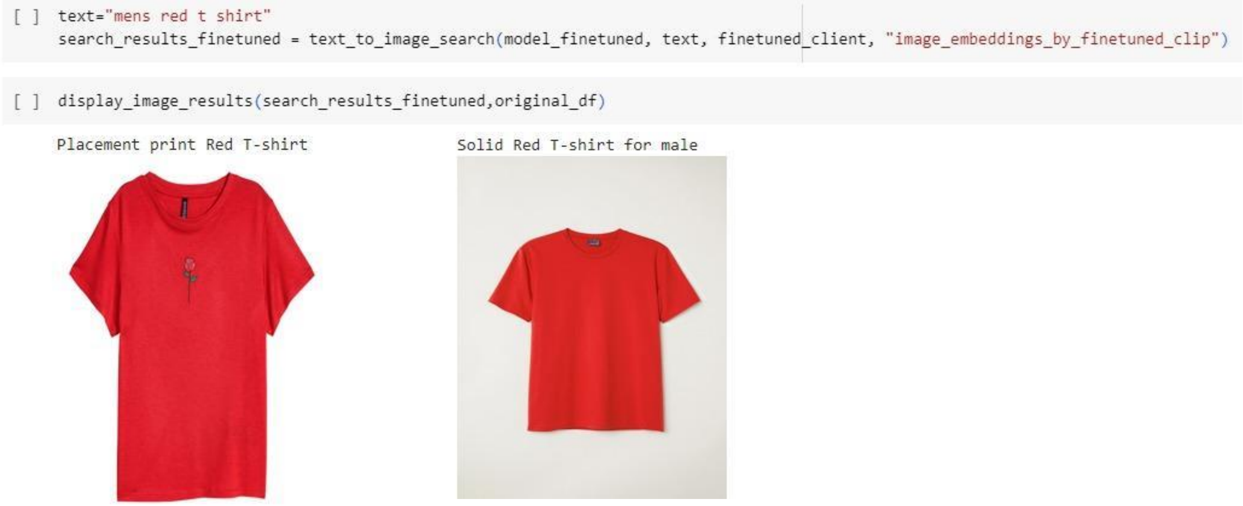
Fig 3. Workflow diagram



```
print('Below is the promt generated by lama 7B model')
res = call_llama('SDE job interview','Men')
print(res)

print('\n')

print('below is the token version with text cleaning done on it')
print(clean_tokens(res))
```

```
Below is the promt generated by lama 7B model
1. Black suit with white shirt and black tie - Black
2. Navy blue blazer with light blue dress shirt and brown pants - Navy Blue
3. Charcoal gray two-piece with navy blue dress shirt and dark brown shoes - Gray
4. Dark brown three-piece with cream-colored dress shirt and beige pants - Brown


below is the token version with text cleaning done on it
[['Black suit', 'white shirt', 'black tie'], ['Navy blue blazer', 'light blue dress shirt', 'brown pants'], ['Charcoal gray two-piece', 'navy blue dress shirt', 'dark brown shoes'], ['Dark brown three-piece', 'cream-colored dress shirt', 'beige pants']]
```

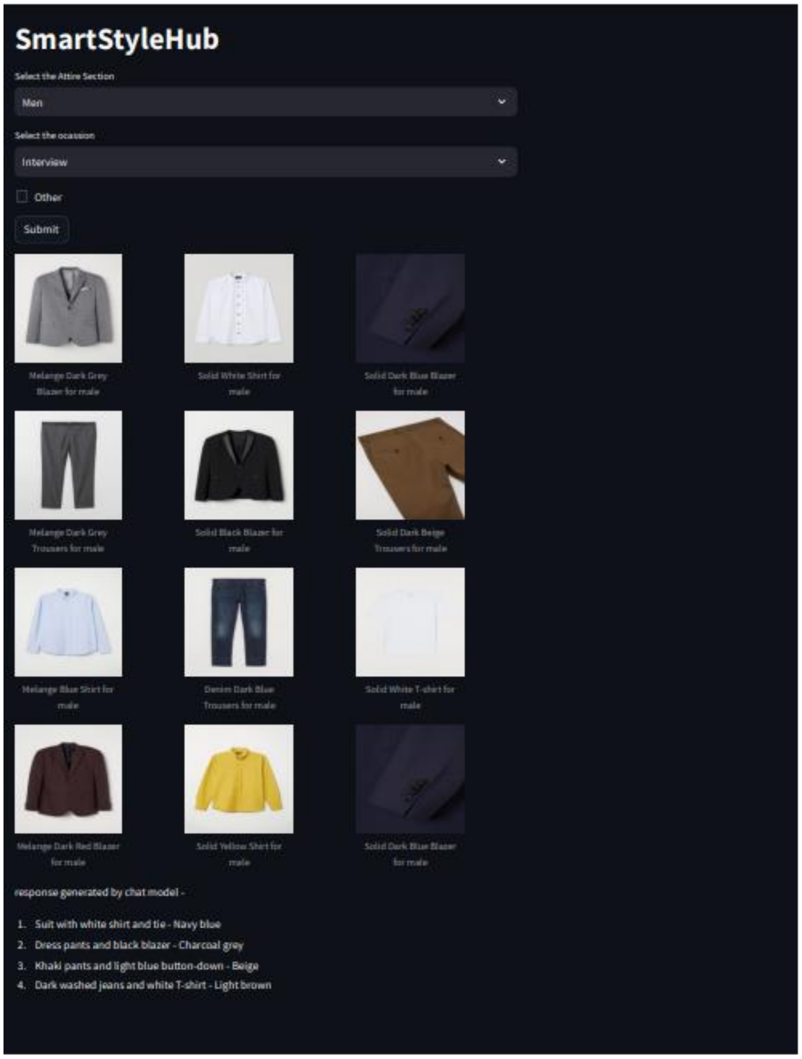
Fig 4. Llama2 model output



```
[ ]  text="mens red t shirt"
     search_results_finetuned = text_to_image_search(model_finetuned, text, finetuned_client, "image_embeddings_by_finetuned_clip")
```

```
[ ]  display_image_results(search_results_finetuned,original_df)
```

Fig 5. CLIP model output


SmartStyleHub
Submit
Melange Dark Grey Blazer for male
Solid White Shirt for male
Solid Dark Blue Blazer for male
Melange Dark Grey Trousers for male
Solid Black Blazer for male
Solid Dark Beige Trousers for male
Melange Blue Shirt for male
Denim Dark Blue Trousers for male
Solid White T-shirt for male
Melange Dark Red Blazer for male
Solid Yellow Shirt for male
Solid Dark Blue Blazer for male

Select the Attire Section

Men

Select the ocassion

Interview

□ Other

response generated by chat model -

1. Suit with white shirt and tie - Navy blue
2. Dress pants and black blazer - Charcoal grey
3. Khaki pants and light blue button-down - Beige
4. Dark washed jeans and white T-shirt - Light brown

Fig 7. Streamlit application output