Code explain:

This notebook presents a weather-aware fashion outfit recommendation system built using the CLIP model and FAISS for fast similarity search. The process begins with data cleaning, where outfits are filtered to retain only those containing at least a specified number of "core" fashion items such as tops, coats, skirts, or pants. This is done by analyzing item metadata using keyword matching on item titles and descriptions. After filtering, the corresponding images are extracted and saved to a new directory to ensure consistency and efficiency in later processing. The CLIP model (openai/clip-vit-base-patch32) is then initialized from Hugging Face Transformers to generate embeddings for both text and images. For each outfit, embeddings of individual item images are computed and averaged to produce a unified "outfit vector" representing the semantic meaning of the entire outfit. These vectors are normalized and indexed using FAISS (IndexFlatIP) to enable cosine similarity-based retrieval. Two types of user inputs are supported: textual descriptions and uploaded images. For the **text + city** mode, the system retrieves real-time weather information using the OpenWeatherMap API, encodes both the user’s text and the weather description into embeddings, and uses these to filter and rank matching outfits via FAISS. For the **image + city** mode, the uploaded image is embedded using CLIP, and then matched against the indexed outfit vectors using the same weather-filtered similarity mechanism. The system is deployed using Gradio, with a clean user interface that provides two input modes—one for text + location and another for image + location—enabling users to explore personalized and seasonally appropriate fashion recommendations interactively. The logic ensures that recommended outfits are not only stylistically similar to the user’s input but also suitable for the weather conditions in a given city.

**Test：**

**text+ weather recommendation**

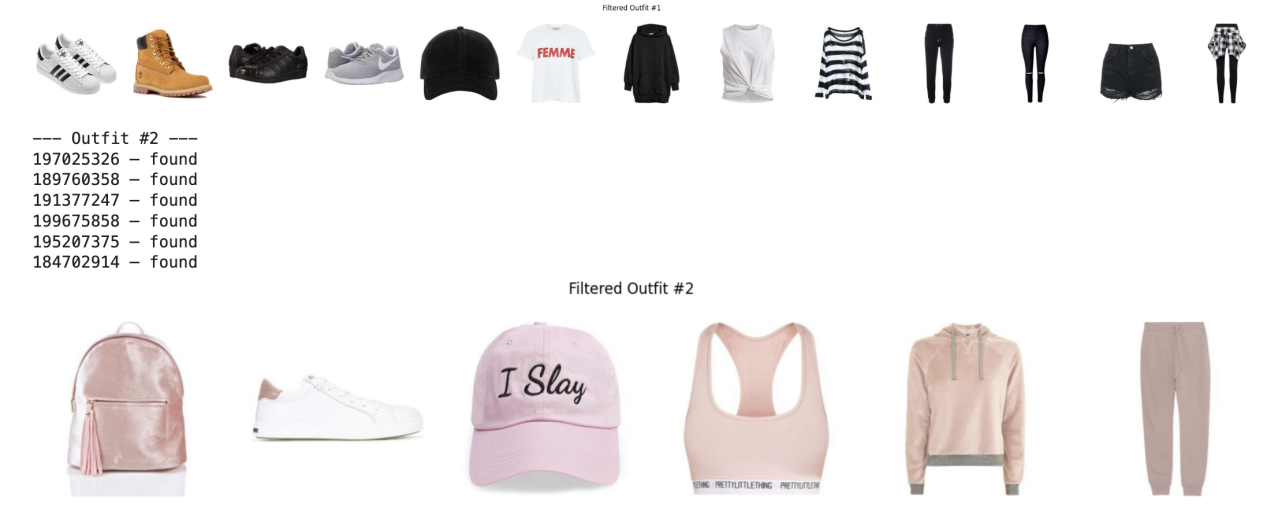
casual streetwear outfit suitable

Norilsk

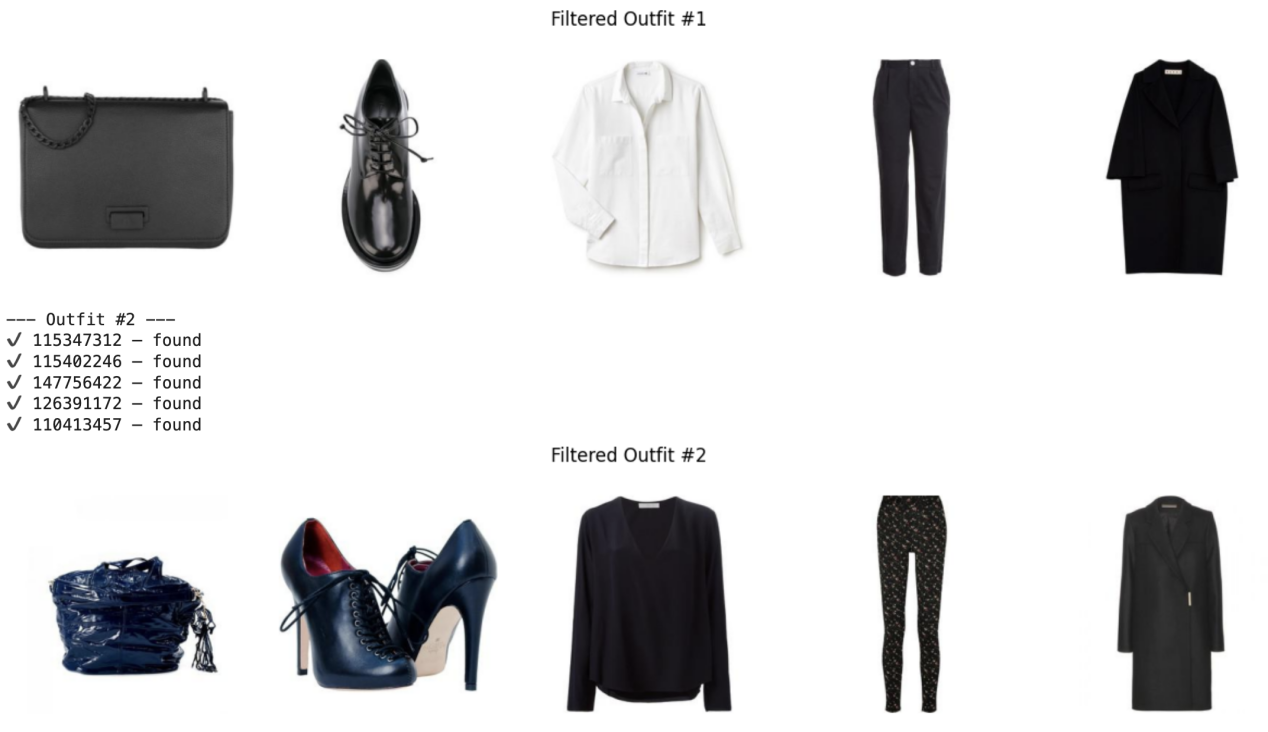


casual streetwear outfit suitable

London



image+ weather recommendation: test.png

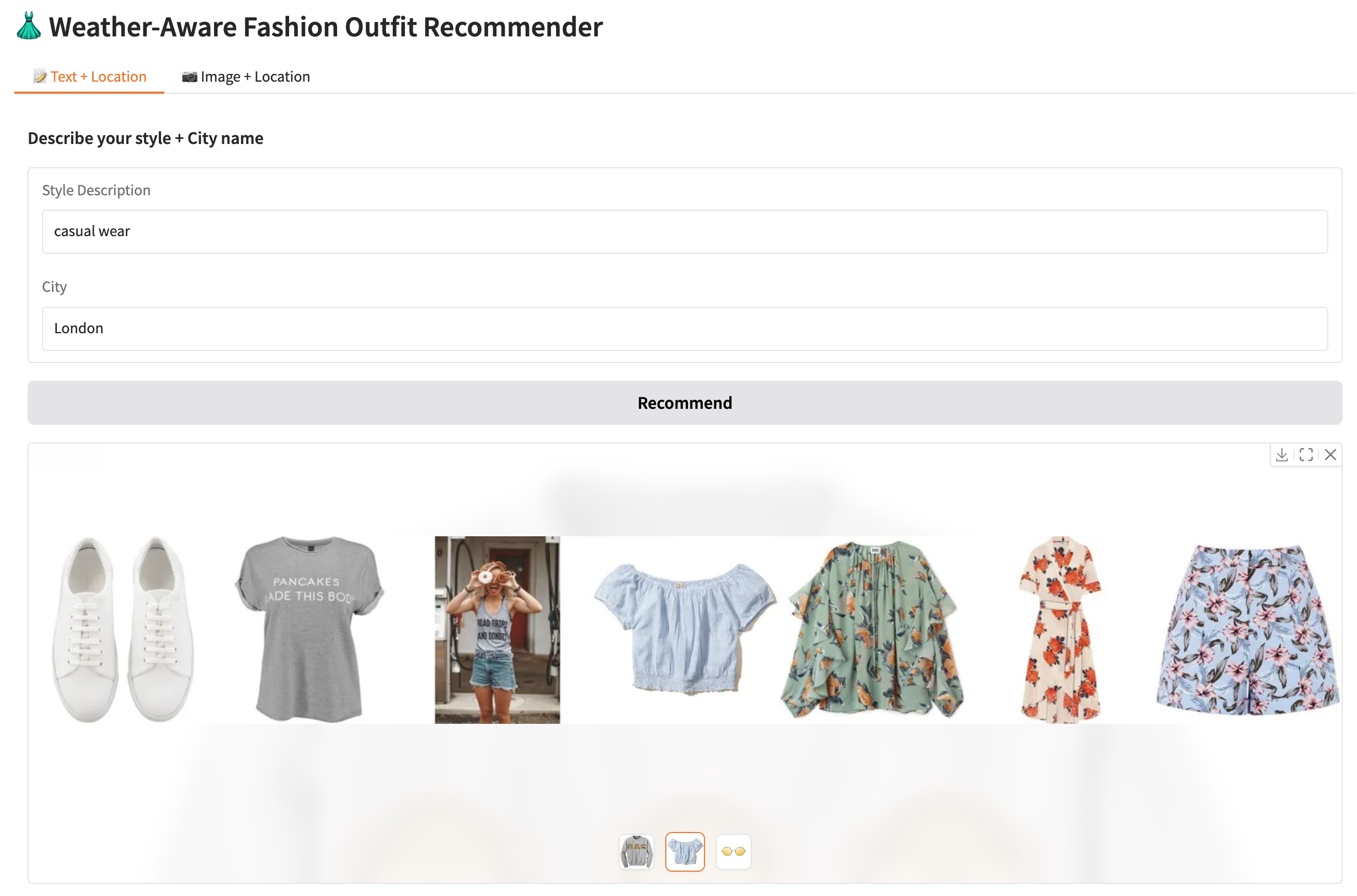


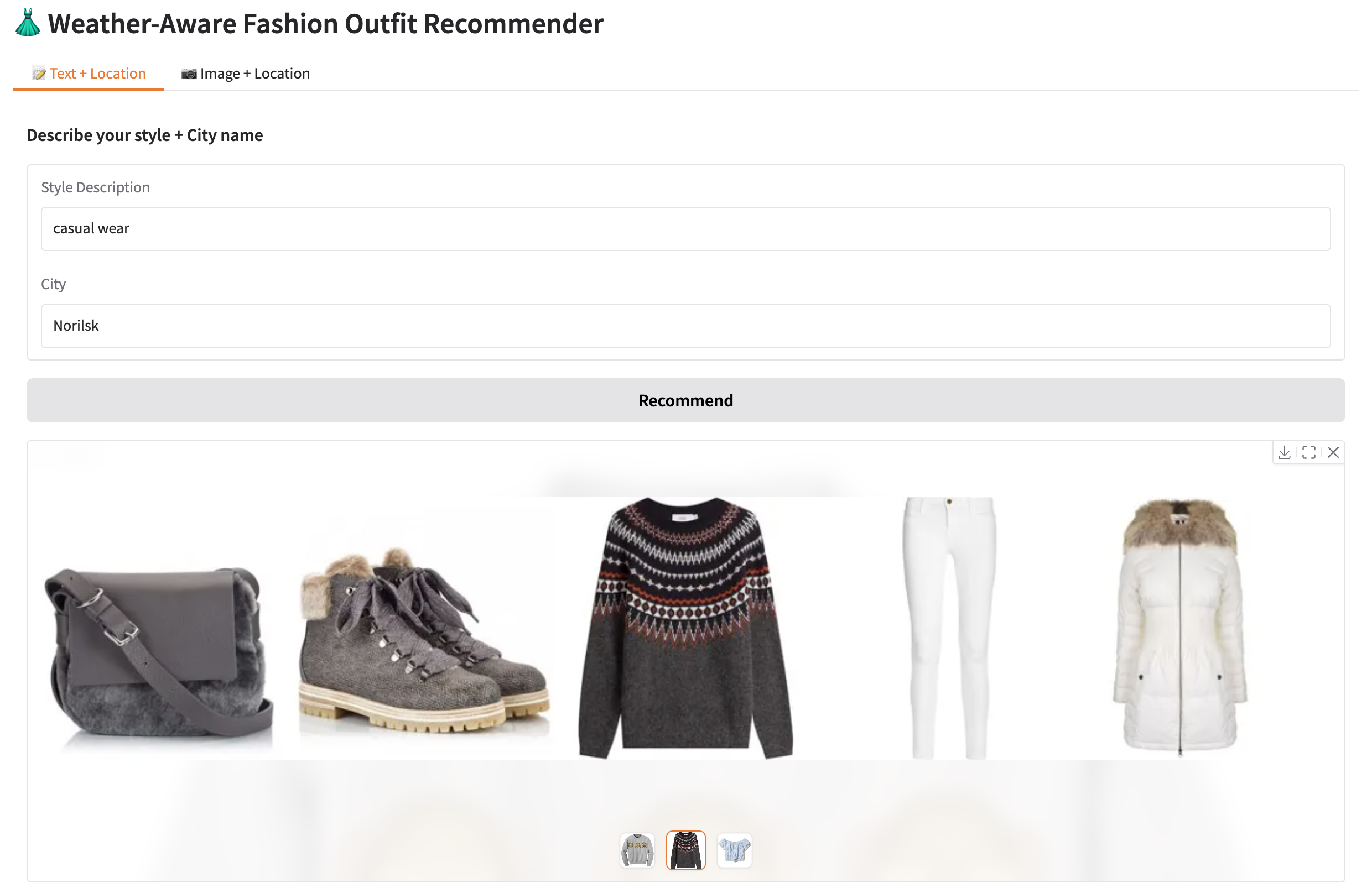
Disadvantages of this method：

This weather-aware outfit recommendation system effectively combines CLIP-based embeddings with user-provided input (text or image) and real-time weather conditions to generate personalized fashion suggestions. However, it has several limitations. First, the weather encoding is overly simplistic—using brief text descriptions like "cold" or "rainy" fails to capture nuanced weather conditions such as wind chill, humidity, or sudden temperature drops. Moreover, CLIP was not explicitly trained to model relationships between weather and clothing, which limits its ability to recommend contextually appropriate garments. Second, the outfit representation is constructed by averaging item embeddings, which ignores the structural composition of an outfit (e.g., top, bottom, shoes) and treats accessories with equal weight as main clothing pieces, leading to diluted semantic meaning. Third, the system does not incorporate user preferences or historical data, making the recommendations generic rather than truly personalized. Additionally, the reliance on image quality and dataset coverage introduces cold-start issues—outfits may be mismatched if the input image deviates from the training distribution. Lastly, the model lacks interpretability and interactivity; users are not provided with explanations for the recommendations and cannot give feedback to iteratively improve results. Future work could involve structured outfit modeling, user preference learning, fine-grained weather profiling, and feedback-driven optimization.

Future improvements：

Future improvements include using the average embeddings of only the main clothing pieces or even just the outerwear to compare with weather embeddings to reduce the effect of diluted semantic meaning when picking contextually appropriate clothing. This aids in ensuring the outfit is better suited for the weather. To better improve embeddings of both the training dataset and input images, fine-tuning can be done with CLIP to ensure that the model is better suited for modelling relationships between weather and clothing and captures stylistic elements more accurately. With this fine-tuning, extra nuance in the weather can be introduced by including more terms in the embedding space which capture different weather conditions. In terms of increasing personalization, future improvements could include example outfit inputs from the user to create an embedding for each unique user, allowing the model to capture the stylistic preferences of the user, which it can then use for future outfit generations to ensure more personalised suggestions. Introducing a feedback loop, where users are able to feedback on the outfit suggested, allows the model to adjust the embedding of an individual user to suit future recommendations to the person’s unique tastes. To prevent cold-start issues in this case, an average embedding of all users can be used for initial users.

Warm e

Cold、

Rain：

