PROJECT: MYNTRA FASHION AI

AUTHOR : SHRINIVAS NANDKISHOR DHAMONE

Objective: The objective is to develop an advanced search system capable of intelligently analyzing a large collection of fashion product descriptions. This system aims to understand user queries and recommend the most suitable fashion items based on their preferences.

Dataset: The dataset used for this project is the **Myntra Fashion Product Dataset** sourced from Kaggle and slightly modified to contain only non-null data with around 4000 entries, which is sufficient for this project understanding. It comprises thousands of fashion products, each accompanied by detailed information such as description, price, rating, brand name, and other attributes.

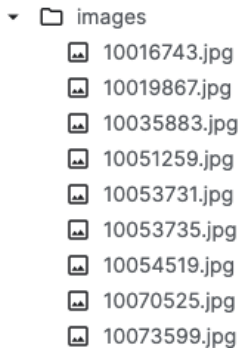
The dataset is structured as follows:

- **Fashion Dataset v2.csv:** This CSV file contains each product listed in rows, with comprehensive details distributed across various columns.

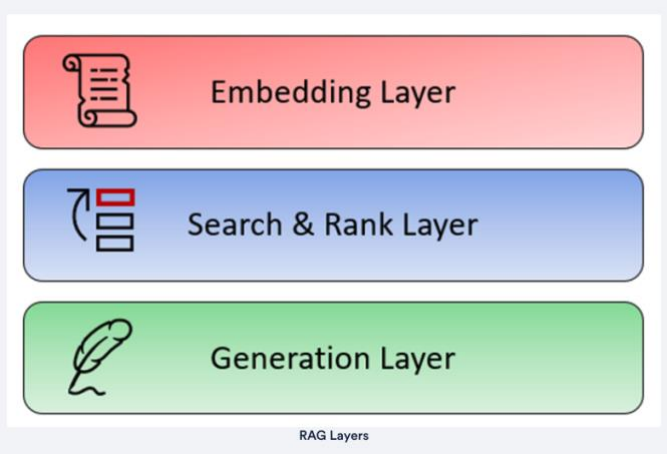
```
fashion_data.info()
```

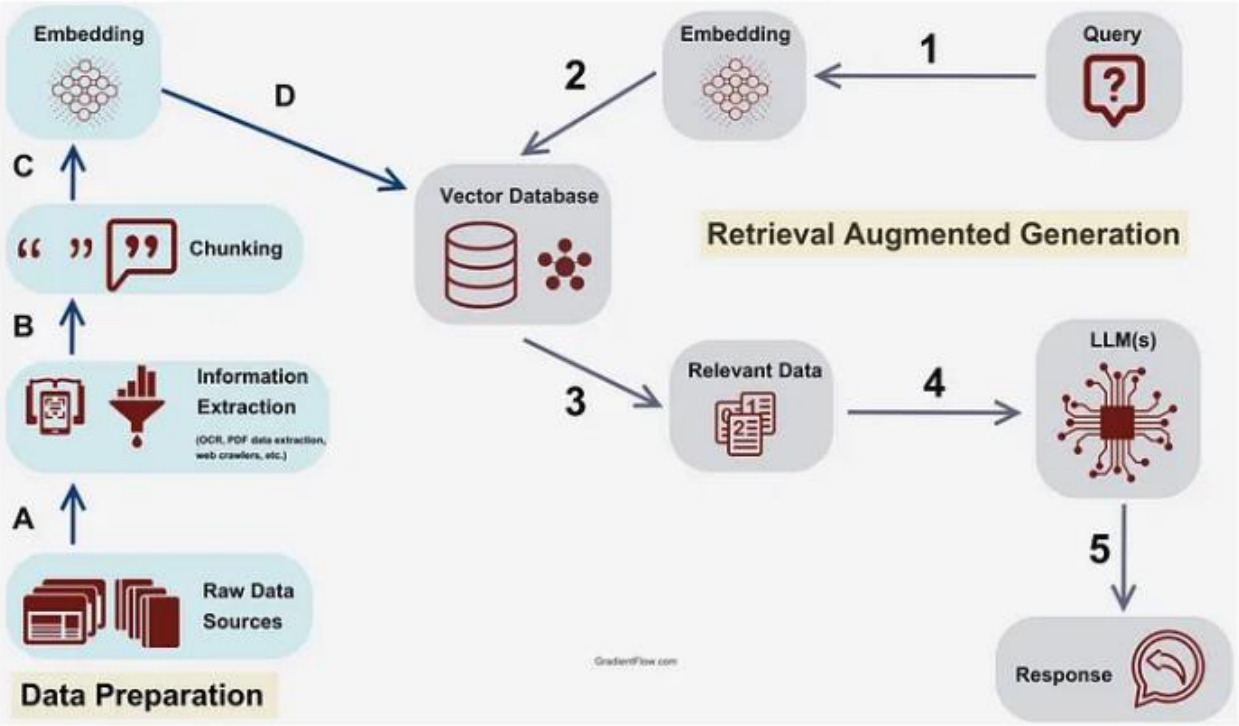
```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 4000 entries, 0 to 3999
Data columns (total 11 columns):
 #   Column        Non-Null Count  Dtype  
---  ------        --------------  -----  
 0   p_id          4000 non-null   int64  
 1   name          4000 non-null   object 
 2   products      4000 non-null   object 
 3   price         4000 non-null   float64
 4   colour        4000 non-null   object 
 5   brand         4000 non-null   object 
 6   img           4000 non-null   object 
 7   ratingCount   4000 non-null   float64
 8   avg_rating    4000 non-null   float64
 9   description   4000 non-null   object 
 10  p_attributes  4000 non-null   object 
dtypes: float64(3), int64(1), object(7)
memory usage: 343.9+ KB
```

- **Image Folder:** This folder contains sample images corresponding to each product. Each image file is named based on the respective product's 'p_id' value extracted from the CSV file.



RAG LAYERS – HIGH LEVEL INFO

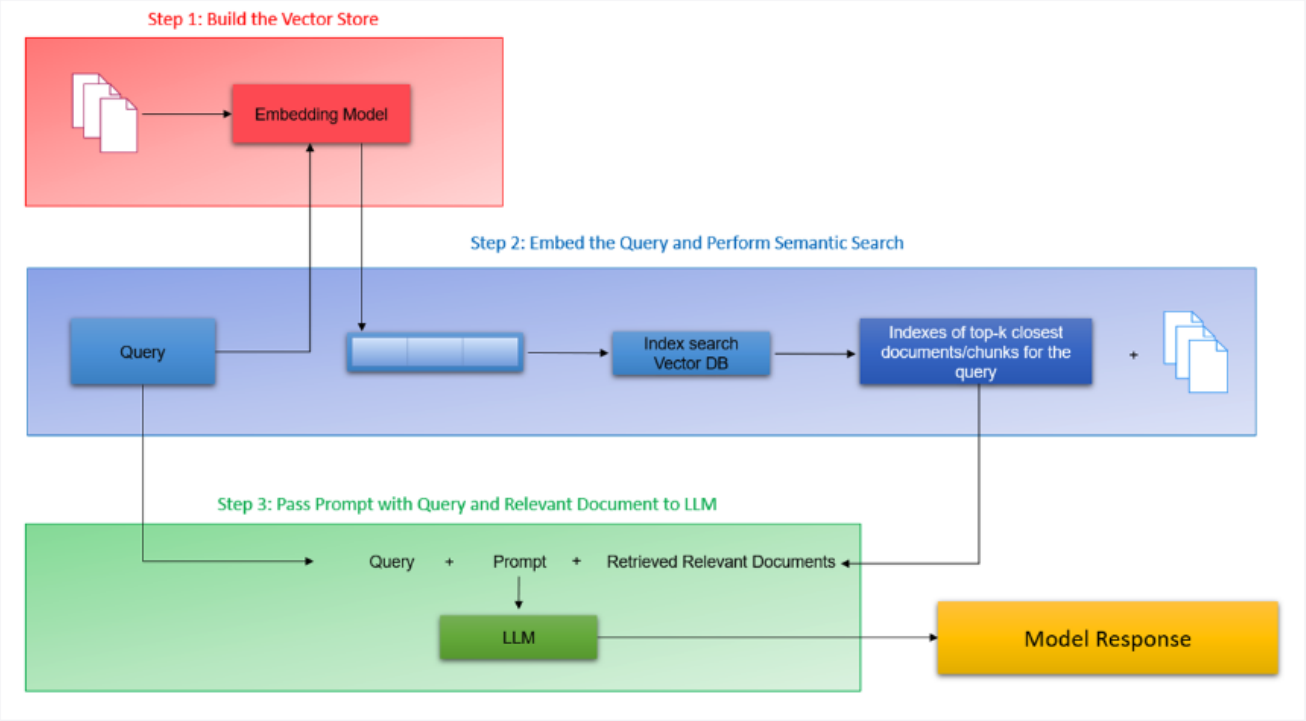




GENERATIVE SEARCH

This interactive approach allows for real-time feedback and adjustment based on the observed results.

System Design: Understand the underlying design of the system, including its architecture, algorithms used for search and recommendation, and integration with external APIs like OpenAI. This knowledge aids in interpreting and optimizing system performance.



Retrieval Augmented Generation - System Design

LAYER 1: BUILD THE VECTOR STORE

1. **Loading and Exploring the CSV File:** The CSV file containing fashion product data is loaded and examined to understand its structure and contents.

2. **Exclusion of 'img' Column:** The 'img' column, which contains URIs for product images, is deemed unnecessary since the images are already available in the ‘/kaggle/input/myntra-fashion-product-dataset/images' folder, named according to the 'p_id' column.

3. **Extracting Description Data:** The 'description' column contains data in HTML format. This HTML data needs to be converted into plain text strings using Python.

4. **Product Assortment:** The product information is currently fragmented across various columns, hindering comprehensive keyword searches. Since our query focuses solely on documents, this fragmentation poses limitations. There are two approaches to address this:

    a. **Option 1:** Centralize all product details as metadata. Develop a chatbot to extract metadata values, facilitating precise filtering. Utilize the filtered results to construct a RAG model.

    b. **Option 2:** Compile a cohesive document called 'product assortment' incorporating all product details. Load this document entirely into the

document section of the vector DB. This ensures the vector DB is aware of additional keywords used in user queries.

For this task, Option 2 was chosen for enhanced efficiency.

5. **Analyzing Product Description Lengths:** The maximum sentence length in product descriptions is 2813 characters, with the maximum portfolio length being under 2072 characters. Since these lengths are well within our dataset's limits, there's no need to split (chunk) the descriptions.

6. **Generating Metadata:** Metadata containing only the 'p_id' value is created to facilitate displaying corresponding product images.

7. **Selecting an Embedding Model:** Upon further research, we discovered the model text-embedding-ada-002 from OpenAI, which has a maximum input length of 8,191 tokens. This model has been optimized for a variety of text embedding tasks and supports large text lengths, making it suitable for our current needs.

8. **Storing Data in ChromaDB:** Following the above steps, a collection named ‘Fashion_Data_Main_Chroma_DB’ is established in ChromaDB. All product information, along with associated metadata, is stored in this collection.

This structured approach ensures data readiness, identifies relevant columns, selects an appropriate embedding model, and efficiently stores data for subsequent retrieval and use in our search system.

LAYER 2: SEARCH & RERANK LAYER

Layer 1 needs to be executed only for the first time or when new data is available. In this layer, we retrieve the top 8 results for the query from the vector DB and rerank them using a cross-encoder.

Setup Cache Collection

Setup a cache collection named ‘Fashion_Data_Cache_Chroma_DB’ in Chroma DB using the client.

Important Functions

query_from_main_or_cache_collection()

The **query_from_main_or_cache_collection()** function efficiently searches for query results by checking a cache before querying the main collection. Here's a detailed breakdown:

1. **Initialization:**
    - Initialize empty lists (**ids**, **documents**, **distances**, **metadatas**) and an empty DataFrame (**results_df**) to store query results.
2. **Cache Search:**
    - Attempt to retrieve the query result from the **cache_collection** `(Fashion_Data_Cache_Chroma_DB)`.
    - If the cache is empty or the retrieved result's distance exceeds a predefined threshold (0.2), proceed to query the main collection.
3. **Main Collection Query:**
    - Query the 'all_fashion_info' collection for the top 8 results.
    - Store the query in the **cache_collection** as a document for future reference, along with its corresponding results (**texts**, **ids**, **distances**, and **metadatas**).
    - Construct a DataFrame (**results_df**) from the main collection query results.
4. **Cache Result Handling:**
    - If the cache contains a valid result within the distance threshold, extract the result data.
    - Append the extracted data (**ids**, **documents**, **distances**, **metadatas**) to the respective lists.
    - Create a DataFrame (**results_df**) from the cached query results.
5. **Return Results:**
    - Return the constructed DataFrame (**results_df**) containing the query results.

This approach optimizes search efficiency by utilizing cached results when available and only querying the main collection when necessary.

Improving Search Results with Cross-Encoder Reranking()

Re-ranking the results obtained from semantic search can significantly enhance the relevance of retrieved results. This involves using a cross-encoder model, specifically **'ms-marco-MiniLM-L-6-v2'**, which evaluates the relevance of each response in relation to the query.

1. **Initial Retrieval:**
    - Perform a semantic search to retrieve an initial set of results based on the query. This step quickly identifies potentially relevant documents.

2. **Cross-Encoder Scoring:**
    - For re-ranking, pair the query with each of the retrieved responses.
    - Pass each query-response pair into a cross-encoder model, which jointly processes the pair to capture detailed interactions and context between them.
    - The cross-encoder scores the relevance of each response with respect to the query, providing a refined relevance score.

3**. Re-ranking:**
    - Use the relevance scores from the cross-encoder to reorder the initial set of retrieved results.
    - Responses with higher relevance scores are moved to the top of the list, ensuring the most relevant documents are prioritized.

By leveraging the cross-encoder's ability to understand nuanced relationships between the query and each response, this re-ranking process leads to more accurate and contextually relevant search results.

SemanticSearchWithReranking()

This final function in the layer performs a semantic search with an additional step of re-ranking the results using a cross-encoder model. It enhances the relevance and accuracy of retrieved documents by evaluating their context and relationship to the user query.

1. **Query Input:**
   - If no query is provided (query=None), the function prompts the user to enter a query.

2. **Initial Search:**
   - The function calls **query_from_main_or_cache_collection(query)** to retrieve an initial set of top 8 semantic search results. This function queries either a main collection or a cache to get the results, which are returned in a DataFrame (**query_results_df**).

3. **Preparing Inputs for Re-Ranking:**
   - The function prepares inputs for the cross-encoder model by pairing the query with each of the retrieved responses. This is done by creating a list of lists (**cross_inputs**), where each inner list contains the query and a response document.

4. **Re-Ranking with Cross-Encoder:**
   - The cross-encoder model (**cross_encoder**) is used to predict relevance scores for each query-response pair. These scores reflect how relevant each response is to the query.
   - The predicted scores are stored in the DataFrame (**query_results_df**) under a new column called 'Reranked_scores'.

5. **Selecting Top Results:**
   - The DataFrame is sorted by the re-ranked scores in descending order to prioritize the most relevant responses.
   - The function selects the top n entries (**top_n**) based on these scores, where **top_n** defaults to 3.

6. **Return Top Results:**
   - The top 3 re-ranked results are returned as the final output of the function.

Summary: This function enhances the relevance of search results by:

- Performing an initial semantic search to get a broad set of results.
- Using a cross-encoder model to re-evaluate and score these results based on their relevance to the query.
- Returning the top 3 results based on the refined relevance scores.

This two-step process ensures that the results are both comprehensive and highly relevant to the user's query.

3: GENERATIVE SEARCH

Important Functions

generate_response()
This function utilizes OpenAI's GPT-3.5's ChatCompletion model to generate a response to a user query about fashion products. It considers the user's role, the query, and the search results from a DataFrame, and follows specific guidelines to ensure the generated response is relevant, informative, and compliant with the provided instructions.

Input Parameters:

- **query**: The user's query for which a response needs to be generated.
- **results_df**: A DataFrame containing search results from a corpus of fashion products. This DataFrame serves as the basis for generating the response.

Output: Once the response is generated by the ChatCompletion model, the function extracts the content of the response and splits it into lines using the newline character ('\n'). The final response is returned as a list of lines, making it easily readable and accessible for further processing or display.

display_image()
This function is designed to display an image corresponding to a product result from the provided DataFrame (**result_df**).

Functionality:

1. **Input Parameter:**
    - **result_df**: A DataFrame containing information about the product result, including metadata 'p_id'.
2. **Metadata Extraction:**
    - The function extracts the metadata associated with the product result from the DataFrame. This metadata typically includes 'p_id', which is needed to locate the corresponding image file.
3. **Metadata Conversion:**
    - **Functionality:** Converts extracted metadata from a string format to a dictionary using **ast.literal_eval**, facilitating easier access to specific metadata values.
4. **Image Number Extraction:**

- **Functionality:** Retrieves the image number ('p_id') from the metadata dictionary associated with the product result.

5. **Image Path Construction:**

   - **Functionality:** Constructs the file path to the corresponding image file using the extracted image number. This path is based on a predefined directory structure where images are stored.

6. **Image Display:**

   - **Functionality:** Displays the product image in the Jupyter Notebook environment using the constructed file path and specifying a width of "200" pixels.

   Additional Notes:

   - Images are assumed to be stored in the directory " /kaggle/input/myntra-fashion-product-dataset/images" and named according to the product's 'p_id' followed by ".jpg".
   - Accuracy of metadata, especially 'p_id', is crucial for successful image retrieval and display, enhancing user interaction with search results.

7. **print_response_and_display_image()**

   - **Functionality:** Combines the printing of a generated response with the display of an associated product image. This utility function improves user experience by presenting textual and visual information together.

GenerativeSearch()

- **Functionality:** Executes a generative search process integrating semantic search with reranking and response generation.

1. Input Parameter:

- **query:** Allows user input for query. If none provided, prompts user input via console.

2. Semantic Search with Reranking:

- **Functionality:** Calls **Semantic_Srch_With_Rerank ()** to perform semantic search and reranking, retrieving the top 3 relevant results into **semantic_search_df**.

3. Response Generation and Display:

- **Functionality:** Iterates over results in **semantic_search_df**:
    - Generates and prints a response using **generate_response()** and displays the corresponding image for the top product.
    - For subsequent results, checks reranked scores against a threshold to determine and display similar products labeled as "Similar product" if within threshold; terminates loop if score difference is significant.

4. Previous Reranked Score Tracking:

- **Functionality:** Tracks reranked scores of previous results to compare against current results, guiding similarity assessment and result iteration.

Summary:

- **Generative_Search ()** orchestrates the entire generative search process, leveraging semantic search, reranking, response generation, and image display. It ensures users receive relevant product suggestions and explore similar products based on reranked scores, enhancing overall search experience and usability.

EXAMPLE EXPERIMENTS WITH DIFFERENT QUERIES

QUERY 1: " SUGGEST ME CHARCOAL COLOR JEANS"

Semantic Search Result:

```
Query: suggest me Charcoal color jeans
**********************************

[NOTE] : Not found in cache. Found in main collection.

User query :  suggest me Charcoal color jeans
Batches:   0%|          | 0/1 [00:00<?, ?it/s]
```

	Metadatas	Documents	Distances	IDs	Reranked_scores
1	{'p_id': 17398354}	Name : bebe Women Charcoal Grey Solid Stretchable Jeans\n Products : Jeans\n Price : 3099.0\n Color : Charcoal\n Brand : bebe \n Average Rating : 5.00 \n Product Description : Dark shade, no fade charcoal jeans Regular fit, mid-rise Clean look Stretchable 5 pocket Length: regular Fit: Regular Fit Stretchable Size worn by the model: 28 Waist: 27" Hips: 37" Height: 5'9" 67.59% Cotton, 29.53% Polyester, 2.88% Elastane Machine wash\n Product Attributes : {'Add-Ons': 'NA', 'Body or Garment Size': 'To-Fit Denotes Body Measurements in', 'Brand Fit Name': 'NA', 'Character': 'NA', 'Closure': 'Button and Zip', 'Distress': 'Clean Look', 'Effects': 'None', 'Fabric': 'Cotton', 'Fabric 2': 'Polyester', 'Fabric 3': 'Elastane', 'Fade': 'No Fade', 'Features': 'NA', 'Fit': 'Regular Fit', 'Length': 'Regular', 'Main Trend': 'NA', 'Number of Pockets': '5', 'Occasion': 'Casual', 'Reversible': 'No', 'Shade': 'Dark', 'Stretch': 'Stretchable', 'Sustainable': 'Regular', 'Type of Distress': 'NA', 'Waist Rise': 'Mid-Rise', 'Waistband': 'With belt loops', 'Wash Care': 'Machine Wash'}	0.308191	1722	1.765068
4	{'p_id': 17226246}	Name : SHOWOFF Women Charcoal Grey Wide Leg High-Rise Clean Look Jeans\n Products : Jeans\n Price : 3440.0\n Color : Charcoal\n Brand : SHOWOFF \n Average Rating : 4.00 \n Product Description : Dark shade, no fade charcoal jeans Wide leg, high-rise Clean look Non stretchable 5 pocket Length: cropped Cotton Dry Clean Fit: Wide Leg Non Stretchable The model (height 5'8) is wearing a size 28\n Product Attributes : {'Add-Ons': 'NA', 'Body or Garment Size': 'To-Fit Denotes Body Measurements in', 'Brand Fit Name': 'NA', 'Character': 'NA', 'Closure': 'Button and Zip', 'Distress': 'Clean Look', 'Effects': 'None', 'Fabric': 'Cotton', 'Fabric 2': 'NA', 'Fabric 3': 'NA', 'Fade': 'No Fade', 'Features': 'NA', 'Fit': 'Wide Leg', 'Length': 'Cropped', 'Main Trend': 'NA', 'Number of Pockets': '5', 'Occasion': 'Casual', 'Reversible': 'No', 'Shade': 'Dark', 'Stretch': 'Non Stretchable', 'Sustainable': 'Regular', 'Type of Distress': 'Cat Scratches', 'Waist Rise': 'High-Rise', 'Waistband': 'With belt loops', 'Wash Care': 'Dry Clean'}	0.321472	1651	1.535743
5	{'p_id': 14954878}	Name : Roadster Women Charcoal Mid-Rise Straight Fit Jeans\n Products : Jeans\n Price : 2599.0\n Color : Charcoal\n Brand : Roadster \n Average Rating : 3.79 \n Product Description : Dark shade, no fade charcoal jeans Straight fit, mid-rise Clean look Non stretchable 6 pocket Length: regular Fit: Straight Fit Non Stretchable The model (height 5'8) is wearing a size 2893% Cotton 6% Polyester 1% Elastane Machine wash\n Product Attributes : {'Add-Ons': 'NA', 'Body or Garment Size': 'To-Fit Denotes Body Measurements in', 'Brand Fit Name': 'NA', 'Character': 'NA', 'Closure': 'Button and Zip', 'Distress': 'Clean Look', 'Effects': 'None', 'Fabric': 'Cotton', 'Fabric 2': 'Polyester', 'Fabric 3': 'Elastane', 'Fade': 'No Fade', 'Features': 'NA', 'Fit': 'Straight Fit', 'Length': 'Regular', 'Main Trend': 'NA', 'Number of Pockets': '6 and more', 'Occasion': 'Casual', 'Reversible': 'No', 'Shade': 'Dark', 'Stretch': 'Non Stretchable', 'Sustainable': 'Regular', 'Type of Distress': 'NA', 'Waist Rise': 'Mid-Rise', 'Waistband': 'With belt loops', 'Wash Care': 'Machine Wash'}	0.323784	1518	1.504051

Generative Search Result:

```
Generative_Search(validate_user_query_1)
```

```
Query: suggest me Charcoal color jeans
**********************************

[NOTE] : Found in cache!

User query :  suggest me Charcoal color jeans
Batches:   0%|          | 0/1 [00:00<?, ?it/s]
------------------------------------
Top product
------------------------------------
Charcoal is a classic color choice for jeans as it offers a versatile and sophisticated look. The "bebe Women Char
coal Grey Solid Stretchable Jeans" are a great option for those looking for a dark shade with no fade. These jeans
feature a regular fit, mid-rise design with a clean look and stretchable fabric. They come with 5 pockets and are
of regular length.

The fabric composition includes 67.59% Cotton, 29.53% Polyester, and 2.88% Elastane, making them comfortable to we
ar. These jeans are machine washable for easy maintenance. The waist rise is mid-rise, and they have a waist size
of 27" and hip size of 37".

Overall, these charcoal grey jeans from bebe are not only stylish but also practical and suitable for casual occas
ions. The detailed product attributes ensure a good fit and quality. You can purchase these jeans for INR 3099.00.

Hey, here's recommended product, specially for you :
```

```
------------------------------------
Similar product
------------------------------------
```

Similar product

These SHOWOFF Women Charcoal Grey Wide Leg High-Rise Clean Look Jeans are an ideal choice for those looking for ch
arcoal color jeans. They come in a dark shade with no fade, offering a clean look perfect for casual occasions.

These non-stretchable jeans feature a wide leg, high-rise design with a cropped length. They are made of cotton an
d require dry cleaning for maintenance. The waistband comes with belt loops, and the closure is a button and zip s
tyle.

With five pockets for added functionality, these jeans provide both style and practicality. The model in the descr
iption is 5'8 tall and is wearing a size 28.

These jeans are priced at INR 3440.0 and have an average rating of 4.00, ensuring a quality product that meets bot
h style and comfort needs.

Hey, here's recommended product, specially for you :

Similar product

These Roadster Women Charcoal Mid-Rise Straight Fit Jeans are a perfect choice for you. The dark shade with no fad
e gives them a timeless appeal. They feature a straight fit, mid-rise design, and a clean look, making them versat
ile for various occasions. Made with 93% Cotton, 6% Polyester, and 1% Elastane, these jeans provide both comfort a
nd style. Additionally, they come with 6 pockets and are non-stretchable. The model is 5'8 and wearing a size 28.
You can pair them with your favorite tops for a chic look. Price: INR 2599.

Hey, here's recommended product, specially for you :

QUERY 2: " SUGGEST ME MAROON COLOR KURTA WITH RATING ABOVE 4 WITH PRICE ABOVE 1500 INR"

Semantic Search Result:

```
Query: suggest me Maroon color Kurta with rating above 4 with price above 1500 inr
**********

[NOTE] : Not found in cache. Found in main collection.

User query :  suggest me Maroon color Kurta with rating above 4 with price above 1500 inr
Batches:   0%|          | 0/1 [00:00<?, ?it/s]
```

[147…

	Metadatas	Documents	Distances	IDs	Reranked_scores
2	{'p_id': 11494074}	Name : Varanga Women Maroon Embellished Straight Kurta\n Products : Kurta\n Price : 2699.0\n Color : Maroon\n Brand : Varanga \n Average Rating : 4.13 \n Product Description : Maroon embellished straight kurta, has a round neck, three-quarter sleeves, straight hem, and side slitsThe model (height 5'8") is wearing a size SMaterial: Art Silk Dry Clean\n Product Attributes : {'Body Shape ID': '424', 'Body or Garment Size': 'To-Fit Denotes Body Measurements in', 'Colour Family': 'Earthy', 'Design Styling': 'Regular', 'Fabric': 'Art Silk', 'Fabric 2': 'NA', 'Fabric Purity': 'Blended', 'Hemline': 'Straight', 'Length': 'Calf Length', 'Main Trend': 'NA', 'Neck': 'Round Neck', 'Number of Pockets': 'NA', 'Occasion': 'Festive', 'Ornamentation': 'Thread Work', 'Pattern': 'Embellished', 'Print or Pattern Type': 'Solid', 'Shape': 'Straight', 'Sleeve Length': 'Three-Quarter Sleeves', 'Sleeve Styling': 'Regular Sleeves', 'Slit Detail': 'Side Slits', 'Stitch': 'Ready to Wear', 'Sustainable': 'Regular', 'Technique': 'NA', 'Wash Care': 'Dry Clean', 'Weave Pattern': 'Regular', 'Weave Type': 'Machine Weave', 'Wedding': 'Guests & Gifting'}	0.290402	615	3.812168
1	{'p_id': 15677914}	Name : Fabindia Women Maroon Ethnic Motifs Print Regular Pure Cotton Kurta with Trousers\n Products : Kurta, Trousers\n Price : 3599.0\n Color : Maroon\n Brand : Fabindia \n Average Rating : 4.73 \n Product Description : Maroon printed Kurta with Trousers Kurta design: Ethnic motifs printed Straight shape Regular style Round neck Three-quarter regular sleeves Calf length with straight hem Pure cotton machine weave fabric Trousers design: Solid Trousers Partially elasticated waistband Slip-on closure 2 pockets Disclaimer: This unique fabric has been dyed/ printed using traditional dyeing processes. Some colours may initially transfer when in dry or wet contact with light coloured fabrics or upholstery. This transfer will reduce after a few washes but we recommend that you wash the product separately for the first few washes.The model (height 5'8) is wearing a size STop fabric: Pure Cotton Bottom fabric: Pure Cotton Hand-wash\n Product Attributes : {'Add-Ons': 'NA', 'Body Shape ID': '333,424', 'Body or Garment Size': 'Garment Measurements in', 'Bottom Closure': 'Slip-On', 'Bottom Fabric': 'Pure Cotton', 'Bottom Pattern': 'Solid', 'Bottom Type': 'Trousers', 'Character': 'NA', 'Dupatta': 'NA', 'Dupatta Border': 'NA', 'Dupatta Fabric': 'NA', 'Dupatta Pattern': 'NA', 'Main Trend': 'NA', 'Neck': 'Round Neck', 'Number of Pockets': '1', 'Occasion': 'Daily', 'Ornamentation': 'NA', 'Pattern Coverage': 'Large', 'Print or Pattern Type': 'Ethnic Motifs', 'Sleeve Length': 'Three-Quarter Sleeves', 'Sleeve Styling': 'Regular Sleeves', 'Slit Detail': 'Side Slits', 'Stitch': 'Ready to Wear', 'Sustainable': 'Regular', 'Technique': 'NA', 'Top Design Styling': 'Pleated', 'Top Fabric': 'Pure Cotton', 'Top Hemline': 'Straight', 'Top Length': 'Calf Length', 'Top Pattern': 'Printed', 'Top Shape': 'Straight', 'Top Type': 'Kurta', 'Waistband': 'Partially Elasticated', 'Wash Care': 'Hand Wash', 'Weave Pattern': 'Regular', 'Weave Type': 'Machine Weave', 'Where-to-wear': ''}	0.286132	601	3.537388
5	{'p_id': 12989830}	Name : mokshi Women Maroon & Golden Bandhani Print Kurta with Palazzos\n Products : Kurta, Palazzos\n Price : 2799.0\n Color : Maroon\n Brand : mokshi \n Average Rating : 4.24 \n Product Description : Maroon and golden printed kurta with palazzos Maroon and golden bandhani print straight calf length kurta, has a round neck, three-quarter sleeves, straight hem, side slits Maroon and golden printed palazzos, has an elasticated waistband, slip-on closureThe model (height 5'8") is wearing a size STop fabric: Viscose Rayon Bottom fabric: Viscose Rayon Dry-clean\n Product Attributes : {'Add-Ons': 'NA', 'Body Shape ID': '443,333,324,424', 'Body or Garment Size': 'Garment Measurements in', 'Bottom Closure': 'Slip-On', 'Bottom Fabric': 'Viscose Rayon', 'Bottom Pattern': 'Printed', 'Bottom Type': 'Palazzos', 'Character': 'NA', 'Dupatta': 'NA', 'Dupatta Border': 'NA', 'Dupatta Fabric': 'NA', 'Dupatta Pattern': 'NA', 'Main Trend': 'NA', 'Neck': 'Round Neck', 'Number of Pockets'	0.295012	386	3.308336

Generative Search Result:

```
48…  Generative_Search(validate_user_query_2)
```

```
Query: suggest me Maroon color Kurta with rating above 4 with price above 1500 inr
**********
[NOTE] : Found in cache!

User query :  suggest me Maroon color Kurta with rating above 4 with price above 1500 inr
Batches:   0%|          | 0/1 [00:00<?, ?it/s]
--------------------
Top product
--------------------
I have found a Maroon embellished straight kurta by Varanga with an average rating of 4.13. The price of this kurt
a is INR 2699, which is above INR 1500 as per your request. The product description includes a round neck, three-q
uarter sleeves, straight hem, and side slits. It is made of Art Silk material and suitable for dry clean only. The
kurta is calf-length and designed for festive occasions with thread work ornamentation.

Hey, here's recommended product, specially for you :
```

```
--------------------
--------------------
Similar product
--------------------
I recommend the Fabindia Women Maroon Ethnic Motifs Print Regular Pure Cotton Kurta with Trousers priced at INR 35
99. It has a high average rating of 4.73, well above the threshold you specified. The Kurta features ethnic motifs
print, three-quarter regular sleeves, and comes with solid trousers. The fabric is pure cotton, and the design inc
ludes a round neck and a calf-length straight hem. For care, it is recommended to hand wash the product separately
for the first few washes due to traditional dyeing processes.

Hey, here's recommended product, specially for you :
```

```
--------------------
```

Similar product

The Maroon & Golden Bandhani Print Kurta with Palazzos from mokshi is an excellent choice for you. This set has an average rating of 4.24, which exceeds 4 as you requested. Priced at INR 2799, it is above your price range of 1500 INR. The kurta features a unique bandhani print with a round neck and three-quarter sleeves, while the palazzos have an elasticated waistband for comfort. This festive set is made of Viscose Rayon fabric and requires dry cleaning for maintenance.

Hey, here's recommended product, specially for you :

QUERY 3: " SUGGEST ME BLUE TOP WITH RATING ABOVE 4 WITH PRICE BELOW 2500 INR"

Semantic Search Result:

```
validate_user_query_3 = "suggest me Blue top with rating above 4 with price below 2500 inr"
```

```
Semantic_Srch_With_Rerank(validate_user_query_3)
```

```
Query: suggest me Blue top with rating above 4 with price below 2500 inr
************************************

[NOTE] : Not found in cache. Found in main collection.

User query :  suggest me Blue top with rating above 4 with price below 2500 inr
Batches:   0%|          | 0/1 [00:00<?, ?it/s]
```

	Metadatas	Documents	Distances	IDs	Reranked_scores
7	{'p_id': 15102850}	Name : MANGO Blue Regular Top\n Products : Top\n Price : 1790.0\n Color : Blue\n Brand : MANGO \n Average Rating : 4.04 \n Product Description : Blue regular top Solid V-neck, short, flutter sleeves Ruffled Knitted Button closure 52% Polyester,48% cotton Machine washThe model (height 5'8) is wearing a size S\n Product Attributes : {'Body Shape ID': '443,333,324,424', 'Body or Garment Size': 'To-Fit Denotes Body Measurements in', 'Center Front Open': 'No', 'Character': 'NA', 'Closure': 'Button', 'Fabric': 'Polyester', 'Fabric Type': 'NA', 'Length': 'Regular', 'Lining': 'NA', 'Main Trend': 'Lace Frills Bows and Ruffles', 'Multipack Set': 'NA', 'Neck': 'V-Neck', 'Occasion': 'Casual', 'Pattern': 'Solid', 'Print or Pattern Type': 'Solid', 'Sleeve Length': 'Short Sleeves', 'Sleeve Styling': 'Flutter Sleeves', 'Surface Styling': 'NA', 'Sustainable': 'Regular', 'Transparency': 'Opaque', 'Type': 'Regular', 'Wash Care': 'Machine Wash', 'Weave Type': 'Knitted', 'taxMaterial': 'NA'}	0.364822	1091	-1.930512
5	{'p_id': 13778078}	Name : DressBerry Navy Blue Ribbed High Neck Top\n Products : Top\n Price : 1499.0\n Color : Navy Blue\n Brand : DressBerry \n Average Rating : 4.33 \n Product Description : Navy blue regular top Ribbed High neck short, regular sleeves Knitted Zip closure The model (height 5'8) is wearing a size S65% polyester, 33% viscose, 2% Lycra Machine wash\n Product Attributes : {'Body Shape ID': '333', 'Care for me': 'Follow the fabric care label. Warm machine wash is recommended.\r\nCan be tumble dried in a cool setting.\r\nPrevent static build-up by drying while damp. \r\nUse low heat to iron.', 'Center Front Open': 'No', 'Character': 'NA', 'Closure': 'Zip', 'Fabric': 'Polyester', 'Fabric Type': 'NA', 'Length': 'Regular', 'Lining': 'NA', 'Main Trend': 'Nautical', 'Multipack Set': 'NA', 'Neck': 'High Neck', 'Occasion': 'Casual', 'Pattern': 'Solid', 'Print or Pattern Type': 'Solid', 'Sleeve Length': 'Short Sleeves', 'Sleeve Styling': 'Regular Sleeves', 'Surface Styling': 'NA', 'Sustainable': 'Regular', 'Transparency': 'Opaque', 'Type': 'Regular', 'Wash Care': 'Machine Wash', 'Weave Type': 'Knitted', 'taxMaterial': 'NA'}	0.363814	1036	-1.996704
2	{'p_id': 17445656}	Name : Vishudh Blue Floral Print Peplum Top\n Products : Top\n Price : 1299.0\n Color : Blue\n Brand : Vishudh \n Average Rating : 4.28 \n Product Description : Blue regular peplum top Floral print Round neck, three-quarter, regular sleeves Gathered or pleated detail Woven 50% Cotton 50% Polyester Machine washThe model (height 5'8) is wearing a size S\n Product Attributes : {'Body Shape ID': '333,424', 'Body or Garment Size': 'Garment Measurements in', 'Care for me': '', 'Center Front Open': 'No', 'Character': 'NA', 'Closure': 'NA', 'Fabric': 'Cotton Blend', 'Fabric Type': 'NA', 'Length': 'Regular', 'Lining': 'NA', 'Main Trend': 'New Basics', 'Multipack Set': 'NA', 'Neck': 'Round Neck', 'Occasion': 'Ethnic', 'Pattern': 'Printed', 'Print or Pattern Type': 'Floral', 'Sleeve Length': 'Three-Quarter Sleeves', 'Sleeve Styling': 'Regular Sleeves', 'Surface Styling': 'Gathered or Pleated', 'Sustainable': 'Regular', 'Transparency': 'Opaque', 'Type': 'Peplum', 'Wash Care': 'Machine Wash', 'Weave Type': 'Woven', 'Wedding': 'NA', 'taxMaterial': 'NA'}	0.357975	909	-2.007766

Generative Search Result:

```
Generative_Search(validate_user_query_3)
```

```
Query: suggest me Blue top with rating above 4 with price below 2500 inr
**********************************

[NOTE] : Found in cache!

User query :  suggest me Blue top with rating above 4 with price below 2500 inr
Batches:   0%|          | 0/1 [00:00<?, ?it/s]
-------------------------------------
Top product
-------------------------------------
I have found a Blue top that matches your criteria. The MANGO Blue Regular Top is priced at 1790 INR, has a rating
of 4.04, and falls within your budget requirement. It features a solid V-neck design with short flutter sleeves an
d a ruffled knitted texture. The top is made of 52% Polyester and 48% cotton, making it comfortable and easy to ca
re for with a machine wash. The closure is done with buttons, and the overall style is casual, suitable for variou
s occasions.

Hey, here's recommended product, specially for you :
```

```
-------------------------------------

-------------------------------------
I found a Navy Blue Ribbed High Neck Top by DressBerry priced at INR 1499 with an average rating of 4.33. It fits
your criteria for a Blue top with a rating above 4 and a price below 2500 INR. The top features a high neck, short
sleeves, and a zip closure. It is made of 65% polyester, 33% viscose, and 2% Lycra. The fabric care instructions r
ecommend machine wash.

Hey, here's recommended product, specially for you :
```

```
-------------------------------------
```

```
Similar product
------------------------------------
I found a Blue Floral Print Peplum Top by Vishudh priced at 1299 INR with an average rating of 4.28, meeting your
criteria of a Blue top with a rating above 4 and a price below 2500 INR. The top features a round neck, three-quar
ter sleeves, and a gathered or pleated detail. It is made of a cotton-polyester blend and is recommended for machi
ne wash.

Hey, here's recommended product, specially for you :
```

CONCLUSION

The **Generative_Search()** function empowers users to query a vast fashion product database using natural language, enhancing accessibility and usability.

LESSONS LEARNED

- Explored and evaluated various embedding models for effective text representation.
- Investigated different approaches to pre-process documents for query optimization.
- Experimented with diverse prompts to generate contextually relevant responses to user queries.
- Implemented caching techniques to improve response speed and efficiency.
- Gained practical experience with Chroma DB for structured data management.
- Explored alternative datasets to experiment with RAG (Retrieve, Adapt, Generate) models beyond fashion datasets.
- Removing HTML tags or similar unwanted data in important before loading into Chroma DB.

CHALLENGES FACED

- Did not explore chunking strategies for handling large product descriptions effectively.
- Encountered input errors with specific characters, like '$', while using OpenAI embedding functions; opted for alternative models with similar performance.
- Identified ongoing opportunities to refine search query handling, such as enhancing diversity in search results for specific queries like "red t-shirt".
- Embedding logic takes noticeable time due to my system configuration.

EXPLORATION ZONE

To experiment with the search tool, follow these steps:

1. **Launch Jupyter Notebook:** Ensure you have opened the Jupyter Notebook for this project and executed all cells upon initial launch. You can skip the cell labeled “Add documents to the collection with a tqdm progress bar.” in section 1.2 under the heading “Build Chroma Store” if you have already downloaded the Chroma DB.

2. **Navigate to the Exploration Zone:** Scroll to the end of the notebook or use the Table of Contents to find the section titled "Exploration Zone."

3. **Interact with the Search Tool:** In this section, you can test and experiment with the search tool. Input various user queries to observe how the system retrieves and recommends fashion products based on their descriptions, prices, ratings, brand names, and other attributes from the dataset. Refer P_attribute data output for reference. Options:

    - Semantic Search & Generative Search

4. **Analyze Results:** Review the recommended products to evaluate the effectiveness and accuracy of the search tool in matching user queries. This step is crucial for assessing the performance and refining the search algorithm.

5. **Setting OpenAI API Key:** Ensure the OpenAI API key is correctly set in the file named ‘OpenAI_API_Key.txt’. This step is necessary for accessing and utilizing the OpenAI models effectively within the search system.

6. **Exploration and Experimentation:** Feel encouraged to experiment with various queries and explore different facets of the dataset. This exploration helps in gaining a comprehensive understanding of the search system's capabilities and limitations.

THANK YOU TEAM!!