

Mo Rashid MT23047

```
[{"https://images-na.ssl-images-amazon.com/images/I/81q5vXEVUL_Sy88.jpg": array([0., 0.01246894, 0., ..., 0., 0., 0., ...], dtype=float32), "https://images-na.ssl-images-amazon.com/images/I/7IHSx4Y-5dL_Sy88.jpg": array([0.000000e+00, 0.000000e+00, 0.000000e+00, ..., 0.000000e+00, 8.240409e-05, 0.000000e+00], dtype=float32), "https://images-na.ssl-images-amazon.com/images/I/71dVsyEjzIL_Sy88.jpg": array([0., 0., 0., ..., 0., 0., 0., ...], dtype=float32), "https://images-na.ssl-images-amazon.com/images/I/7ZMDS1hUFL_Sy88.jpg": array([0., 0., 0., ..., 0., 0., 0., ...], dtype=float32), "https://images-na.ssl-images-amazon.com/images/I/7WbaodrTUL_Sy88.jpg": array([0., 0., 0., ..., 0., 0., 0., ...], dtype=float32), "https://images-na.ssl-images-amazon.com/images/I/81lviyxwDUL_Sy88.jpg": array([0., 0., 0., ..., 0., 0., 0., ...], dtype=float32), "https://images-na.ssl-images-amazon.com/images/I/51L6GFHJPL_Sy88.jpg": array([0., 0., 0., ..., 0., 0., 0., ...], dtype=float32), "https://images-na.ssl-images-amazon.com/images/I/71zh9tVhUL_Sy88.jpg": array([0., 0.00160999, 0., ..., 0., 0., 0., ...], dtype=float32), "https://images-na.ssl-images-amazon.com/images/I/81gk9wRLSL_Sy88.jpg": array([0., 0., 0., ..., 0., 0., 0., ...], dtype=float32), "https://images-na.ssl-images-amazon.com/images/I/81hhbeC8HL_Sy88.jpg": array([0., 0., 0., ..., 0., 0., 0., ...], dtype=float32), "https://images-na.ssl-images-amazon.com/images/I/51o2V4-wZAL_Sy88.jpg": array([0., 0.00050086, 0., ..., 0.00082602, 0., 0., ...], dtype=float32), "https://images-na.ssl-images-amazon.com/images/I/51zfFbd-xL_Sy88.jpg": array([0., 0., 0.00031377, ..., 0., 0., 0., ...], dtype=float32), "https://images-na.ssl-images-amazon.com/images/I/51Jg0SCFFSL_Sy88.jpg": array([0., 0., 0., ..., 0., 0., 0., ...], dtype=float32), "https://images-na.ssl-images-amazon.com/images/I/71raKRMMPL_Sy88.jpg": array([0., 0., 0., ..., 0., 0., 0., ...], dtype=float32), "https://images-na.ssl-images-amazon.com/images/I/71a3ozB8GL_Sy88.jpg": array([0., 0., 0., ..., 0., 0., 0., ...], dtype=float32), "https://images-na.ssl-images-amazon.com/images/I/71srB7SYLL_Sy88.jpg": array([0., 0.00014214, 0., ..., 0.00130017, 0., 0., ...], dtype=float32), "https://images-na.ssl-images-amazon.com/images/I/81lvefsmKL_Sy88.jpg": array([0., 0.00340899, 0., ..., 0., 0., 0., ...], dtype=float32), "https://images-na.ssl-images-amazon.com/images/I/614m5CUS7L_Sy88.jpg": array([0., 0., 0., ..., 0., 0., 0., ...], dtype=float32), "https://images-na.ssl-images-amazon.com/images/I/71ic7hkewL_Sy88.jpg": array([0., 0., 0., ..., 0., 0., 0., ...], dtype=float32), "https://images-na.ssl-images-amazon.com/images/I/61kdaA5uHL_Sy88.jpg": array([0.00194688, 0., 0., ..., 0., 0., 0., ...], dtype=float32), "https://images-na.ssl-images-amazon.com/images/I/71riuzspKtL_Sy88.jpg": array([0., 0.01758171, 0.00238352, ..., 0.00798625, 0., 0., ...], dtype=float32), "https://images-na.ssl-images-amazon.com/images/I/61llAAeAmzL_Sy88.jpg": array([0., 0., 0., ..., 0., 0., 0., ...], dtype=float32), "https://images-na.ssl-images-amazon.com/images/I/51fyLuNjynL_Sy88.jpg": array([0., 0., 0., ..., 0., 0., 0., ...], dtype=float32), "https://images-na.ssl-images-amazon.com/images/I/61se2PgT7L_Sy88.jpg": array([0., 0., 0., ..., 0.00132714, 0., 0., ...], dtype=float32), "https://images-na.ssl-images-amazon.com/images/I/61EkhxUuGL_Sy88.jpg": array([0., 0., 0., ..., 0., 0., 0., ...], dtype=float32), "https://images-na.ssl-images-amazon.com/images/I/21OKLIxGCL_Sy88.jpg": array([0., 0., 0., ..., 0., 0., 0., ...], dtype=float32), "https://images-na.ssl-images-amazon.com/images/I/51ouPPTtlWL_Sy88.jpg": array([0., 0., 0., ..., 0., 0., 0., ...], dtype=float32), "https://images-na.ssl-images-amazon.com/images/I/71xK9cSpnL_Sy88.png": array([0., 0., 0., ..., 0., 0., 0., ...], dtype=float32), "https://images-na.ssl-images-amazon.com/images/I/71d4cDYtDL_Sy88.jpg": array([0., 0.00027191, 0., ..., 0., 0., 0., ...], dtype=float32), "https://images-na.ssl-images-amazon.com/images/I/71TiVaR1dL_Sy88.jpg": array([0., 0., 0., ..., 0., 0., 0., ...], dtype=float32)
```

Performed text pre-processing steps, including HTML tag removal, tokenization, stop word removal, stemming, and lemmatization. Calculated term frequencies and TF-IDF scores from scratch. Saved the TF-IDF scores using pickle for later use.

3871
> <

We take the input URL and corresponding text URL and extract the features from the saved pickle file.

Image-Based Retrieval: In Image-Based Retrieval, we retrieve the top result based on the image's similarity and also calculate the retrieved text similarity using the input text review.



Question 4 :

Combined the individual similarity scores from image-based and text-based retrieval techniques to calculate a composite score. Ranked the pairs based on the composite score to identify top similar pairs. Methodologies:

Top Similar Images Based on Review Text Similarity:

1) Image URL: https://images-na.ssl-images-amazon.com/images/I/81q5+IxFVUL._SY88.jpg

Review: Loving these vintage springs on my vintage strat. They have a good tension and great stability. If you are floating your bridge and want the most out of your springs than t

Cosine similarity of images: 1.0000

Cosine similarity of text: 1.000

Composite similarity: 1.0000

1) Image URL: https://images-na.ssl-images-amazon.com/images/I/71I3nFp0MLL._SY88.jpg

Review: Loving these vintage springs on my vintage strat. They have a good tension and great stability. If you are floating your bridge and want the most out of your springs than t

Cosine similarity of images: 1.0000

Cosine similarity of text: 0.675

Composite similarity: 0.8377

1) Image URL: https://images-na.ssl-images-amazon.com/images/I/712LUK8v2+L._SY88.jpg

Review: Hot damn! I replaced the original neck position humbucker on my 1980 Les Paul (which I bought new) with a SPH90 Phat Cat. What a difference!! It is MUCH louder, and brig

Cosine similarity of images: 0.3451

Cosine similarity of text: 0.616

Composite similarity: 0.4807

Question 5:

Challenges:

1. Understanding what users want is hard because sometimes they don't explain well or use unclear words.
2. Words can mean different things, like "Python" can refer to both a programming language and a snake.
3. There's so much information online that it's easy to get overwhelmed with too many search results.
4. Finding images, videos, and audio can be tough because there's so much of it, and it's not just text.

Potential Improvements:

1. Use natural language processing to better understand what users are asking for.
2. Keep updating search results in real-time so users get the latest information.
3. Use context like location or past searches to give more relevant results.
4. Ask users for feedback to improve search results and ranking.