

**You**

jobs\_updated.csv

Spreadsheet

Tong Su resume.txt

Document

perform a TF-IDF cosine similarity scoring of the resume against these 20 jobs, using a standard stopwords list and return the ordered list of jobs with the most similar job first and the similarity scores

ChatGPT

**ChatGPT**

Error analyzing

There was an issue with reading the job descriptions file due to a Unicode decoding error. This often happens when the file is not encoded in UTF-8 format. To proceed, I can try reading the file using a different encoding, such as ISO-8859-1, which is commonly used for files with non-UTF-8 characters. Let's try this approach. ​​

Error analyzing

It appears there is an issue with the column name 'Description' in the job descriptions file. This could mean that the actual column name is different from what I expected. To resolve this, I will first inspect the column names in the file to identify the correct column that contains the job descriptions. Let's proceed with that. ​​

The job descriptions are under the column named 'Job Description' in the file. I will now use this correct column name to perform the TF-IDF cosine similarity scoring between the resume and the job descriptions. Let's proceed with this corrected approach. ​​

Error analyzing

It appears that I missed redefining the TF-IDF vectorizer after encountering the previous error. Let me correct this and proceed with the TF-IDF cosine similarity calculation again. ​​

Error analyzing

It seems there is a problem with the data in the job descriptions, possibly due to missing values (**NaN**). I will handle these potential missing values in the job descriptions and then retry the TF-IDF cosine similarity calculation. Let's proceed with this adjustment. ​​

The TF-IDF cosine similarity scoring of the resume against the job descriptions has been successfully completed. Here is the ordered list of jobs, starting with the most similar job first, along with their respective similarity scores:

1. **Data Analyst** - Similarity Score: 0.1977
2. **Data Analyst** - Similarity Score: 0.1849
3. **Data Analyst** - Similarity Score: 0.1635
4. **ASSOCIATE ANALYST, SEM** - Similarity Score: 0.1377
5. **Senior Paid Media Analyst** - Similarity Score: 0.1373
6. **Jr. Data Analyst** - Similarity Score: 0.1369
7. **Data Analyst** - Similarity Score: 0.1318
8. **Digital Data Analyst** - Similarity Score: 0.1293
9. **Data Analyst Ð 2024** - Similarity Score: 0.1240
10. **Data Analyst IV (Healthcare Financial Analytics)** - Similarity Score: 0.1224
11. **Marketing Analyst** - Similarity Score: 0.1204
12. **Business Analyst Ð Marketing** - Similarity Score: 0.1175
13. **Data Analyst** - Similarity Score: 0.1099
14. **Data Analyst II** - Similarity Score: 0.0966
15. **Data Driven Marketing Analyst Ð Creative** - Similarity Score: 0.0959
16. **Marketing Analyst** - Similarity Score: 0.0958
17. **Performance Marketing Analyst** - Similarity Score: 0.0869
18. **Data Analyst** - Similarity Score: 0.0863
19. **Data Analyst** - Similarity Score: 0.0799
20. **Entry-Level Corporate Marketing Analyst** - Similarity Score: 0.0479

There are two additional entries with no job title and a similarity score of 0.0, which likely indicates missing or incomplete data for those job descriptions. ​​



**You**

can you return me the results in a excel file

ChatGPT

**ChatGPT**

The results of the TF-IDF cosine similarity scoring have been saved to an Excel file. You can download the file using the link below:

Download the sorted job similarities Excel file ​​



**You**

jobs\_updated1.csv

Spreadsheet

Tong Su resume.txt

Document

do it again these 2 files

ChatGPT

**ChatGPT**

The TF-IDF cosine similarity scoring for the new resume against the updated job descriptions has been completed and saved to an Excel file. You can download the file using the link below:

Download the new sorted job similarities Excel file ​