

ge - Select or cr

Untitled1 - Jupyter Note

Python File Open

Convert Text Documents

python - AttributeError:

+

localhost:8888/notebooks/Untitled1.ipynb?kernel_name=conda-env-PythonAdv-py

Python [conda env:PythonAd

pyter

Untitled1

Last Checkpoint: 30 minutes ago (unsaved changes)

Edit

View

Insert

Cell

Kernel

Widgets

Help

⌂

📄

⬆

⬇

⬆

⬆

Run

⏮

⏪

⏩

⏭

Code

📄

In [1]:

```
import pandas as pd
import numpy as np
from sklearn.feature_extraction.text import TfidfVectorizer
```

In [2]:

```
#create the tf-idf function

def tf_idf_func(text_file):
    #create vectorizer
    vectorizer = TfidfVectorizer()
    #perform TG-IDF on file
    matrix = vectorizer.fit_transform(text_file)
    feature_names = vectorizer.get_feature_names()
    df = pd.DataFrame(
        data=matrix.toarray(),
        columns=feature_names)
    return df
```

In [3]:

```
#read in files
thor = open('thor_ragnarok.txt', 'r')
thor = thor.read()
thor = [thor]
```

In [4]:

```
matrix = tf_idf_func(thor)
matrix
```

Out[4]:

| | 10 | 12 | 142 | 18 | 30 | ability | about | absence | absolute | access | ... | yield | you | young | younger | your |
|---|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|-----|----------|----------|----------|----------|----------|
| 0 | 0.001267 | 0.001267 | 0.005067 | 0.001267 | 0.001267 | 0.001267 | 0.040536 | 0.001267 | 0.002534 | 0.011401 | ... | 0.001267 | 0.571305 | 0.001267 | 0.001267 | 0.076005 |

1 rows x 1566 columns

In []: