Manuel Santiago Gutierrez Plazas 506211011

Diseño de interfaces

Enlace del repositorio: https://github.com/MANUELSANT019/FitFusion

```
import numpy as np
import pandas as pd
from pandas_profiling import ProfileReport
#Esta seccion de codigo importa librerias como numpy que realiza calculo numerico
#la libreria pandas se usa para el manejo y estr
```

```
<ipython-input-2-f0702a0af64c>:3: DeprecationWarning: `import pandas_profiling` is going to be deprecated by April 1st. Please use `import ydata_profiling` instead.
  from pandas_profiling import ProfileReport
```

Esto solo imprime hola

```
data = pd.read_csv("scopus.csv")
#aca importamos o le decimos al programa cual es el archivo que contiene nuestra data
```

```
profile = ProfileReport(data, title="Profiling Report")
#aca usamos una biblioteca de la libreria pandas que se encarga de perfilar los datos
```

```
profile
#hacemos llamado para ejecutar
```

```
import numpy as np
import pandas as pd

import requests
from bs4 import BeautifulSoup

from PIL import Image
from wordcloud import WordCloud, ImageColorGenerator
import matplotlib.pyplot as plt

import string
import nltk
from nltk.corpus import stopwords
#aca se vuelve a importar las librerias anteriores junto a otras nuevas para extraer la informacion
#de una pagina web, pillow es para usar imagenes con python, wordcloud para generar nubes de palabras
# a partir de textos, matplotlib se utiliza para generar graficos en python
# nltk para procesar lenguaje natural y stopwords para remover palabras irrelevantes para el analisis
#se uso chat gpt para pillow y matplotlib
```

```
nltk.download('stopwords')
#se descarga el texto, se procesa el lenguaje natural y se remueven las palabras irrelevantes
```

```
[nltk_data] Downloading package stopwords to /root/nltk_data...
[nltk_data]   Unzipping corpora/stopwords.zip.
True
```

```
punctuation=[]
for s in string.punctuation:
    punctuation.append(str(s))
sp_punctuation = ["¿", "¡", """, """, "…", ":", "–", "»", "«"]

punctuation += sp_punctuation
#se inicia un ciclo para remover todos los signos de puntuacion ya que estos no son relevantes
```

```python
titulos = str
for titulo in data["Abstract"]:
    titulos = titulo + " " + str(titulos)
titulos
#se inicia una cadena str llamada titulo y esta lo que hace es usar los abstracts de la data
#para tener todo el abstract en una cadena para realizar un analisis posterior
```

```python
stop_words = stopwords.words('english')
#aca se remueven las palabras comunes en ingles que son irrelevantes para nuestro analisis
```

```python
for p in punctuation:
    clean_texto = titulos.lower().replace(p,"")

for p in punctuation:
    clean_texto = clean_texto.replace(p,"")
#el primer ciclo se encarga de volver a minuscula todo el texto
#el segundo ciclo es para reemplazar los signos de puntuacion por una cadena vacia
# basicamente limpia el texto
```

```python
for stop in stop_words:
    clean_texto_list = clean_texto.split()
    clean_texto_list = [i.strip() for i in clean_texto_list]
    try:
        while stop in clean_texto_list: clean_texto_list.remove(stop)
    except:
        print("Error")
        pass
    clean_texto= " ".join(clean_texto_list)
    #esta seccion de codigo analiza la data en busca de las palabras mas comunes en ingles
    # y las remueve del texto ya que son irrelevantes y solo ocasionaria
    # que las siguientes listas o diagramas estuvieran llenos de conectores
```

```
lista_texto = clean_texto.split(" ")

palabras = []

for palabra in lista_texto:
    if (len(palabra)>=3 and len(palabra)<18):
        palabras.append(palabra)
        #como se vio en secciones anteriores se limpia el texto para dividirlo en palabras individuales
        #luego se crea una lista que se llena con los datos "validos"
        #se inicia un bucle y el condicional adentro valida palabras con un numero de caracteres mayo a 3 y menos a 18
        #esto puede variar segun el numero de caracteres que consideremos valido
```
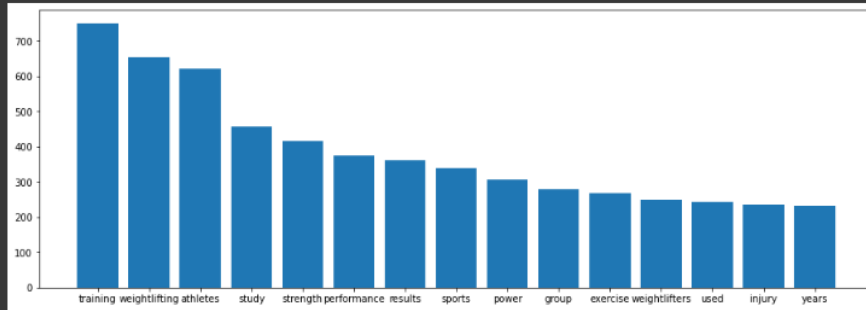
```
word_count={}
for palabra in palabras:
    if palabra in word_count.keys():
        word_count[palabra][0]+=1
    else:
        word_count[palabra]=[1]
#Esta seccion de codigo crea un diccionario, luego en el for por cada palabra en la lista anterior
#si palabra esta en word_count... increment el valor de su llave
```

```
df = pd.DataFrame.from_dict(word_count).transpose()
df.columns=["freq"]
df.sort_values(["freq"], ascending=False, inplace=True)
df.head(10)
#esta seccion me transpone "palabras clave", saca la frecuencia usando la libreria pandas y me muestra los 10 primeros
```

	freq
training	752
weightlifting	654
athletes	623
study	458
strength	416
performance	375
results	362
sports	340
power	306

```python
def plot_bar(data=df, top=5):
    fig = plt.figure()
    ax = fig.add_axes([0,0,2,1])
    ax.bar(x =df.iloc[:top,:].index, height = df.iloc[:top,0].values)
    plt.show()
    #en esta seccion se usa plotbar que es para realizar un grafico de barras
    #se definen en la funcion los 5 primeros, el tamaño de la grafica
    #y finalmente se muestra
```

```python
plot_bar(data=df, top=15)
#en esta seccion se llama a la funcion con sus parametros de entrada y me muestra los primeros 15 datos
```



```python
word_cloud = WordCloud(height=800, width=800, background_color='white',max_words=150, min_font_size=5, collocation_threshold=10).generate(clean_texto)
word_cloud.to_file("n1.png")
plt.figure(figsize=(10,8))
plt.imshow(word_cloud)
plt.axis('off')
plt.tight_layout(pad=0)
plt.show()
#esta seccion me crea la nube de palabras usando la biblioteca word_cloud, definimos los parametros, el tipo de archivo para mostrar
#finalmente se ejecuta el archivo y se muestra
```



```python
data.head()
#este muestra los el conjunto de nuestros primeros 5 datos de nuestra data
```

	Authors	Author(s) ID	Title	Year	Source title	Volume	Issue	Art. No.	Page start	Page end	...	Affiliations	Authors with affiliations	Abstract	Author Keywords	Index Keywords	Document Type	Publication Stage	Open Access	Source	EID
0	Brognara L., Mazzotti A., Rossi F., Lamia F.,...	57194237026;53164388700;58099428600;58099386670...	Using Wearable Inertial Sensors to Monitor Eff...	2023	Sensors	23	3	1636	NaN	NaN	...	Department of Biomedical and Neuromotor Scienc...	Brognara, L., Department of Biomedical and Neu...	Background: Dynamic balance plays a key role i...	foot orthosis; inertial sensor; instability; p...	Inertial navigation systems; Wearable sensors;...	Article	Final	NaN	Scopus	2-s2.0-85147856218
1	Ang C.L., Kong P.W.	58099372000;14056310700;	Field-Based Biomechanical Assessment of the Se...	2023	Sensors	23	3	1171	NaN	NaN	...	Physical Education and Sports Science Academic...	Ang, C.L., Physical Education and Sports Scien...	Traditionally, the biomechanical analysis of O...	barbell; ground reaction force; kinematic; kin...	Biomechanics; Kinematics; Laboratories; Video ...	Article	Final	NaN	Scopus	2-s2.0-85147843394
2	Huebner M., Riemann B., Hatchett A.	35749957700;7005706149;56976273800;	Grip Strength and Sports Performance in Compet...	2023	International Journal of Environmental Researc...	20	3	2033	NaN	NaN	...	Department of Statistics and Probability, Mich...	Huebner, M., Department of Statistics and Prob...	Grip strength (GS) is correlated with major mu...	aging; athletes; clean and jerk; hand symmetry...	aging; muscle; performance assessment; physica...	Article	Final	NaN	Scopus	2-s2.0-85147831973
3	Plachel F., Akgün D., Imiolczyk J.-P., Minkus ...	56410554100;57194329886;57201669702;5650739540...	Patient-specific risk profile associated with ...	2023	Archives of Orthopaedic and Trauma Surgery	143	2	NaN	699	706	...	Center for Musculoskeletal Surgery, Charité—Un...	Plachel, F., Center for Musculoskeletal Surger...	Introduction: Although age is considered to be...	Eccentric osteoarthritis; Primary glenohumeral...	case control study; complication; female; huma...	Article	Final	NaN	Scopus	2-s2.0-85112623370
4	Ames A., Shah S.S., Pettit R., Li L., Chilton ...	57960154000;57191668695;57203800947;5720923308...	Against surgeons' advice: the return to sport ...	2023	Journal of Shoulder and Elbow Surgery	NaN	NaN	NaN	NaN	NaN	...	New England Baptist Hospital, Boston, MA, Unit...	Ames, A., New England Baptist Hospital, Bostor...	Background: Return to sport in high-demand wei...	NaN	Case Series; glenohumeral arthritis; Level IV;...	Article	Article in Press	NaN	Scopus	2-s2.0-85148717379

5 rows × 24 columns