

Librerías de Python utilizadas

1. Librería re de Python

La librería 're' en Python proporciona soporte para expresiones regulares.

Algunas funciones importantes incluyen:

- a. `re.search(pattern, string, flags=0)`:
 - i. Busca el patrón en cualquier parte de la cadena.
 - ii. Devuelve un objeto Match si encuentra una coincidencia, o None si no la encuentra.
- b. `re.match(pattern, string, flags=0)`:
 - i. Intenta hacer coincidir el patrón solo al principio de la cadena.
 - ii. También devuelve un objeto Match o None.
- c. `re.findall(pattern, string, flags=0)`:
 - i. Encuentra todas las subcadenas no superpuestas donde el patrón coincide.
 - ii. Devuelve una lista de todas las coincidencias.
- d. `re.sub(pattern, repl, string, count=0, flags=0)`:
 - i. Reemplaza una o más coincidencias del patrón en la cadena.
 - ii. 'repl' puede ser una cadena o una función que devuelve una cadena.
- e. `re.compile(pattern, flags=0)`:
 - i. Compila una expresión regular en un objeto de patrón, que se puede reutilizar para búsquedas más eficientes.
- f. Grupos de captura:
 - i. Permiten extraer partes específicas de una coincidencia.
 - ii. Se definen con paréntesis en el patrón.
- g. Flags:
 - i. Modifican el comportamiento de la búsqueda.
 - ii. Algunos flags comunes: `re.IGNORECASE`, `re.MULTILINE`, `re.DOTALL`.

2. Pandas y PandasGUI:

a. Pandas:

- i. Es una librería de Python para análisis y manipulación de datos.
- ii. Proporciona estructuras de datos flexibles como DataFrame y Series.
- iii. Permite leer y escribir datos en diversos formatos (CSV, Excel, SQL, etc.).
- iv. Ofrece funciones para limpieza, transformación y análisis de datos.
- v. Funcionalidades clave de Pandas:
- vi. Lectura de datos: `pd.read_csv()`, `pd.read_excel()`, etc.

- vii. Selección y filtrado de datos: `df['columna']`, `df.loc[]`, `df.iloc[]`
- viii. Operaciones de agrupación: `df.groupby()`
- ix. Fusión y unión de datos: `pd.merge()`, `df.join()`
- x. Manejo de datos faltantes: `df.dropna()`, `df.fillna()`
- xi. Operaciones estadísticas: `df.describe()`, `df.mean()`, `df.median()`, etc.

b. PandasGUI:

- i. Es una interfaz gráfica para visualizar y explorar DataFrames de Pandas.
- ii. Permite ver, filtrar y editar datos sin necesidad de escribir código.
- iii. Ofrece funcionalidades como ordenamiento, filtrado y visualización de gráficos.

Referencias:

Getting started — pandas 2.2.2 documentation. (s/f). Pydata.org.

https://pandas.pydata.org/docs/getting_started/index.html

Jain, Y. (2022, febrero 17). *Text cleaning (using Regex) [Python]*. Medium.

<https://medium.com/@yashj302/text-cleaning-using-regex-python-f1dded1ac5bd>

re — Regular expression operations. (s/f). Python Documentation.

<https://docs.python.org/3/library/re.html>