**Regex**

Archivos de texto, un string en jupyter. Buscar información dentro de este texto.

Regex ayuda a buscar patrones que se repiten dentro de un texto. Por ejemplo un número de teléfono, un email.

Sirve para todos los lenguajes de programación.

La librería se llama regex. Se importa como re. Permite identificar patrones en un texto

Import re

Import panda as pd

Import numpy as np

El ‘.’ Sirve para cualquier cosa en regex. Como un comodín

^ : Inicia el string

$: para el final del string

\*Buscar entre 0 o muchas repeticiones

+ buscar que un texto aparezca 1 vez o muchas

? Interrogación: buscar que un texto aparezca una vez o ninguna

Cuando ^ está dentro de unos [] significa que lo que hay dentro no se incluya

Cuando quiero buscar los caracteres especiales indicados arriba hay que hacerlo precedidos de una barra invertida

\w: busca todos los caracteres alfanuméricos

\w+ busca secuencias continuas de caracteres alfanuméricos. Palabras

r’’ [0-9]{2}: que me busque números de 0 a 9 de dos dígitos seguidos

\d: para buscar dígitos del 0 al 9

\s: espacios en blanco

Estos símbolos en mayúscula es lo contrario. Que no sean dígitos

Lo primero que se escribe e$n la función es regex, el resto, no.

La función en sí misma no modifica el texto de origen. Habría que igualarla a una variable

Txt = re.sub(r’g’, ‘G’, txt) -> cambia en el texto las ‘g’ por ‘G’

Hay que decirle siempre el literal de lo que queremos cambiar o modificar

Txt = “The rain in Spain”

X = re.search(r’ sombreritoThe.\*Spain$’,txt)

Que busque un texto que empiece por Th, todo lo que haya en medio (.), Spain ha de estar al final y $ todo lo que haya detrás

Match hace lo mismo que el gorrito desde el principio del string

Pattern = r’Cookie’

Sequence = ‘I want a Cookie’

Sequence1 = ‘Cookie, I want you!’

re.match(pattern,sequence)