



TEXTO: **CORRELACIONES** **DE DATOS**

NLTK (NATURAL LANGUAGE TOOLKIT)



ADHD



Librería de Python ampliamente utilizada en el procesamiento de lenguaje natural (NLP, por sus siglas en inglés). NLTK proporciona una amplia gama de herramientas y recursos para trabajar con texto y datos de lenguaje natural.



N-GRAMS

word_tokenize es la función proporcionada por la librería **NLTK** y se utiliza para llevar a cabo la tokenización de un texto, es decir, dividir un texto en palabras o términos individuales. Esta función se utiliza para la generación de bigrams para dividir un texto en secuencias de dos palabras consecutivas en lugar de palabras individuales.

	Cleaned	Body	Cluster	Texto_Completo	palabras_clave	bigram
94610	Same I have seriously fallen asleep some recen...	0	Same have seriously fallen asleep some recent...	seriously fallen asleep, right getting frustra...	[seriously fallen, fallen asleep, asleep , , ...]	
194624	Every woman who was diagnosed late in life aft...	0	Every woman who was diagnosed late in life aft...	seriously depressed wow shocker, every woman, ...	[seriously depressed, depressed wow, wow shock...	
82663	I ve found that having a buddy or something to...	0	ve found that having buddy or something to d...	times exersize stimulus visuals looking around...	[times exersize, exersize stimulus, stimulus v...	
174679	I leave glasses of water around the house and ...	0	leave glasses of water around the house and m...	water around, leave glasses, like, house, father	[water around, around , , leave, leave glasse...	
119775	Firstly happy birthday for Monday I don t have...	1	Firstly happy birthday for Monday don have a...	confidence shows self awareness sounds like, l...	[confidence shows, shows self, self awareness,...]	
...

NETWORKX

librería de Python diseñada para la creación, manipulación y análisis de estructuras de redes complejas. Se utiliza en una variedad de campos, incluyendo la ciencia de datos, la teoría de redes, la visualización de datos y la investigación en ciencias sociales, para modelar y estudiar sistemas complejos compuestos por nodos interconectados y arcos o conexiones entre esos nodos.

ADHD

GRÁFICO DE NODOS DE BIGRAMAS (UMbral DE FRECUENCIA = 1000)

