



FUNDAMENTO DE NLTK



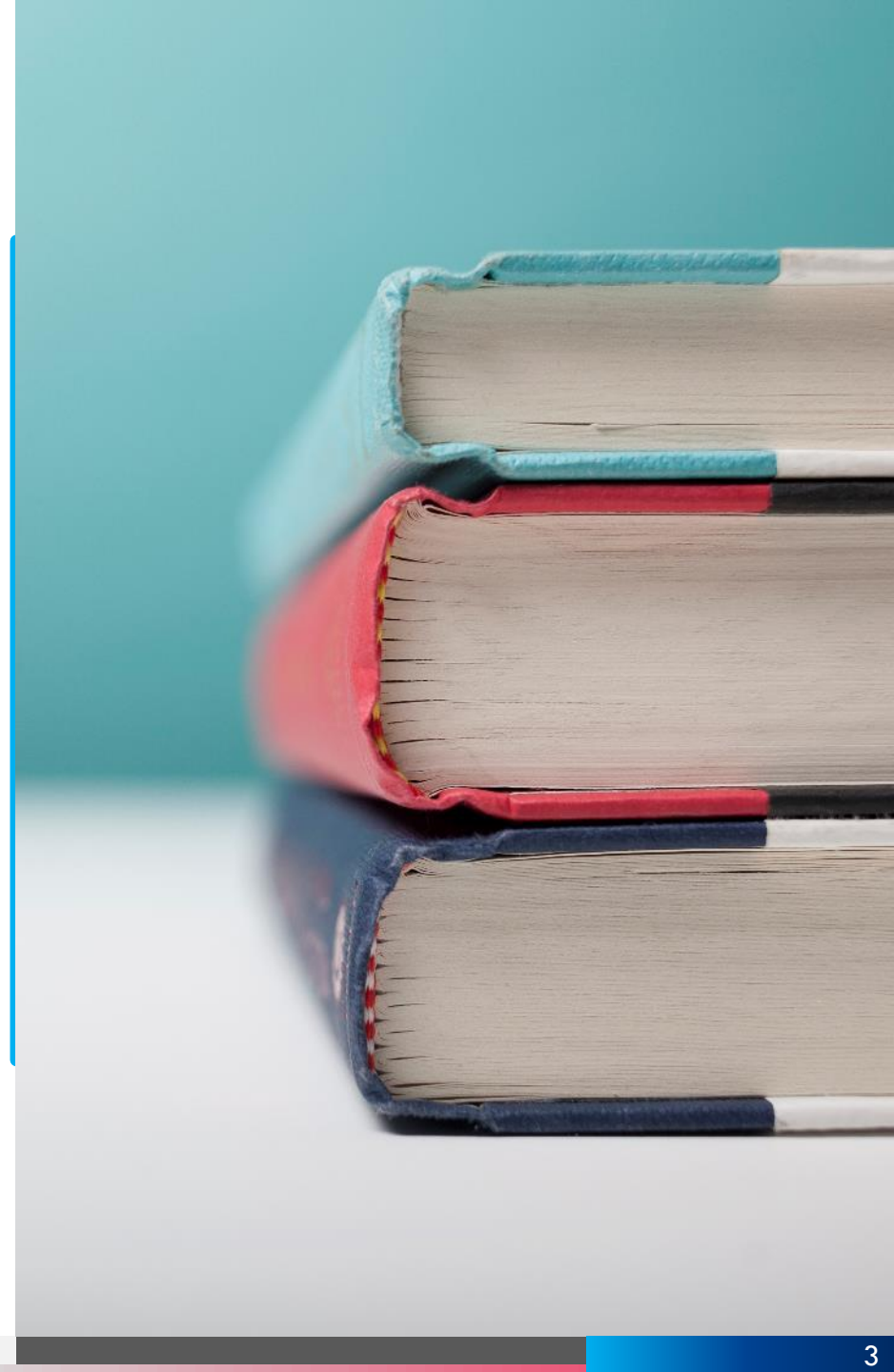
Introducción al Procesamiento del Lenguaje Natural



Contenido

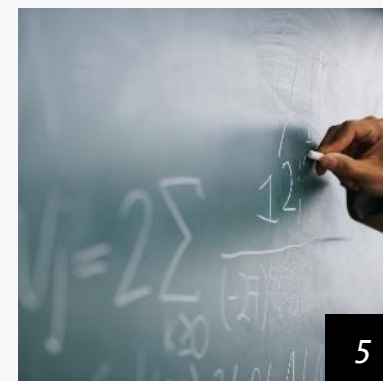
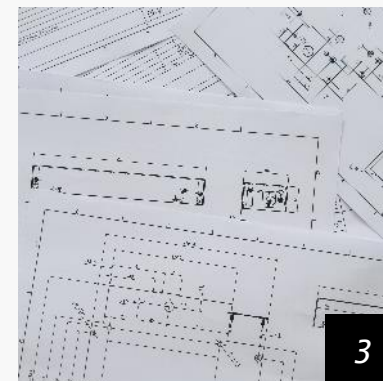
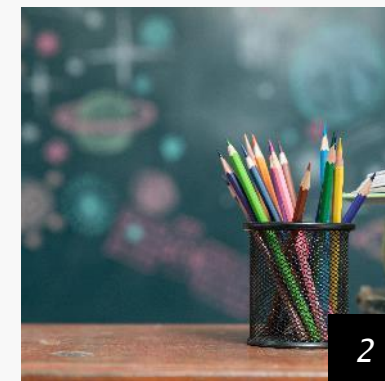
Introducción

- Introducción al procesamiento del lenguaje natural
- Evolución del NLP
- Conocimientos básicos de NLP



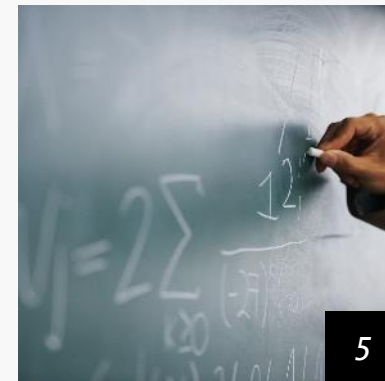
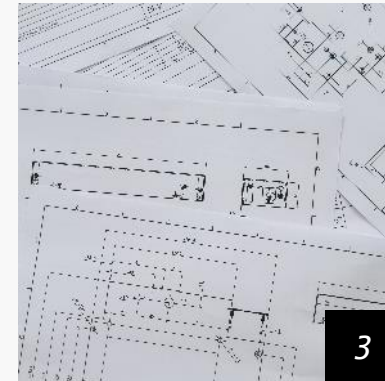
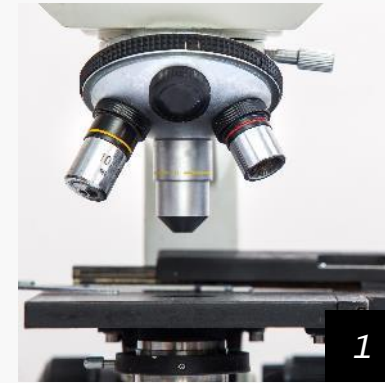
Introducción al proceso del lenguaje natural

Es una rama de la inteligencia artificial que ayuda a las computadoras a comprender, interpretar y manipular lenguajes humanos como el inglés para analizar y derivar su significado. El NLP permite a los desarrolladores a organizar y estructurar el conocimiento para la creación de traducción, resúmenes, reconocimiento de entidades nombradas, extracción de relaciones, reconocimiento de voz, segmentación de temas, Etc.



Evolucion del NLP

- 1950- El NLP comenzó cuando Alan Turing Publico un articulo titulado “MACHINERY AND INTELLIGENCE”
- 1950-Intentos de automatizar la traducción entre ruso e ingles
- 1960-El trabajo de Chomsky y otros sobre la teoría del lenguaje formal y la sintaxis generativa.
- 1990- Los modelos probabilísticos y basados en datos se había vuelto bastante estándar.
- 2000- Se dispone de una gran cantidad de datos y textuales
- El año 2020 ha sido el año de los generativos y a la cabeza de muchos ya habrá venido GPT-3 entrena un corpus con unos 175mil millones de parámetros en su versión completa, también GPT-3 que es de Microsoft ha sido capas de redactar artículos científicos enteros, noticias, programas scripts enteros en diferentes lenguajes y responder preguntas sobre si mismo.

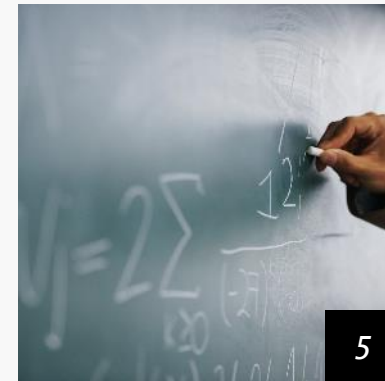
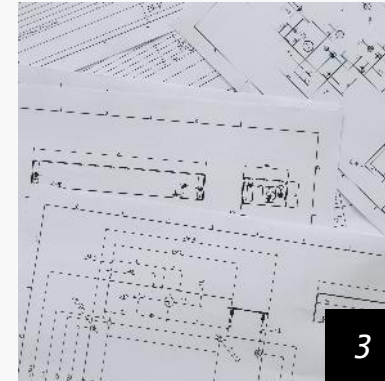
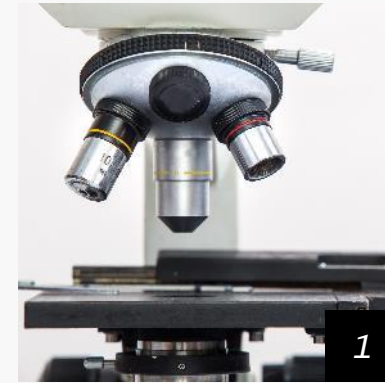


FUNDAMENTOS CON NLTK



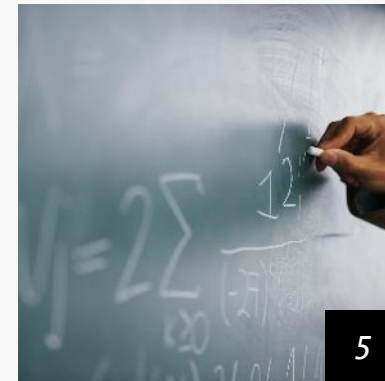
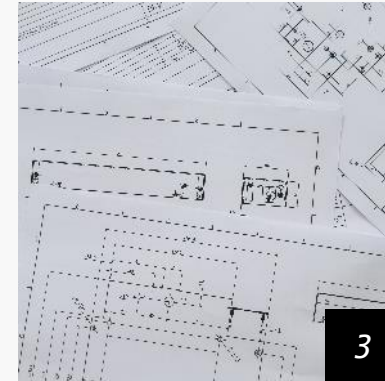
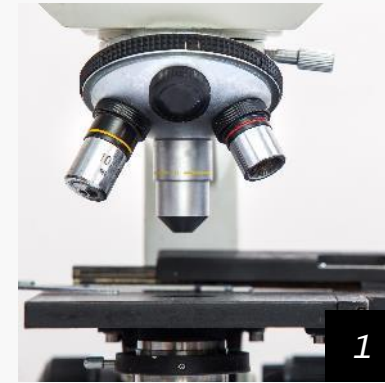
CONTENIDO

- Configuración de ambiente de trabajo
- Instalación de nltk y descargas de librerías de nltk
- Palabras, textos y vocabularios
- Tokenización con Expresiones Regulares
- Estadísticas básicas del lenguaje
- Distribuciones de frecuencia de palabras
- Refinamiento y visualización de cuerpos de texto
- N-gramas y Colocaciones del lenguaje
- ¿Cómo extraer n-gramas de un texto en Python?
- Colocaciones en Python
- Colocaciones en gráficos de dispersión
- Filtros y colocaciones en NLTK



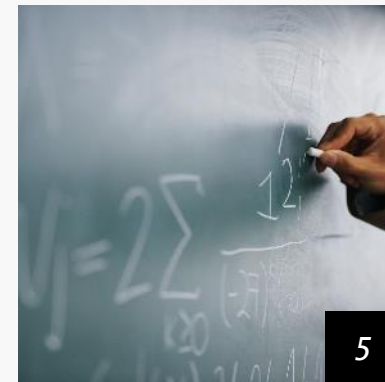
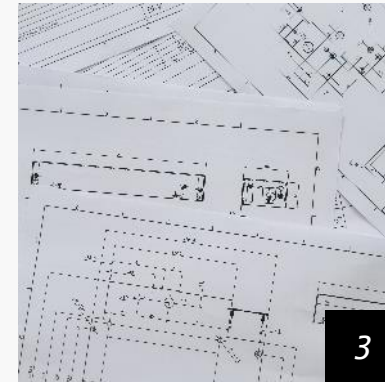
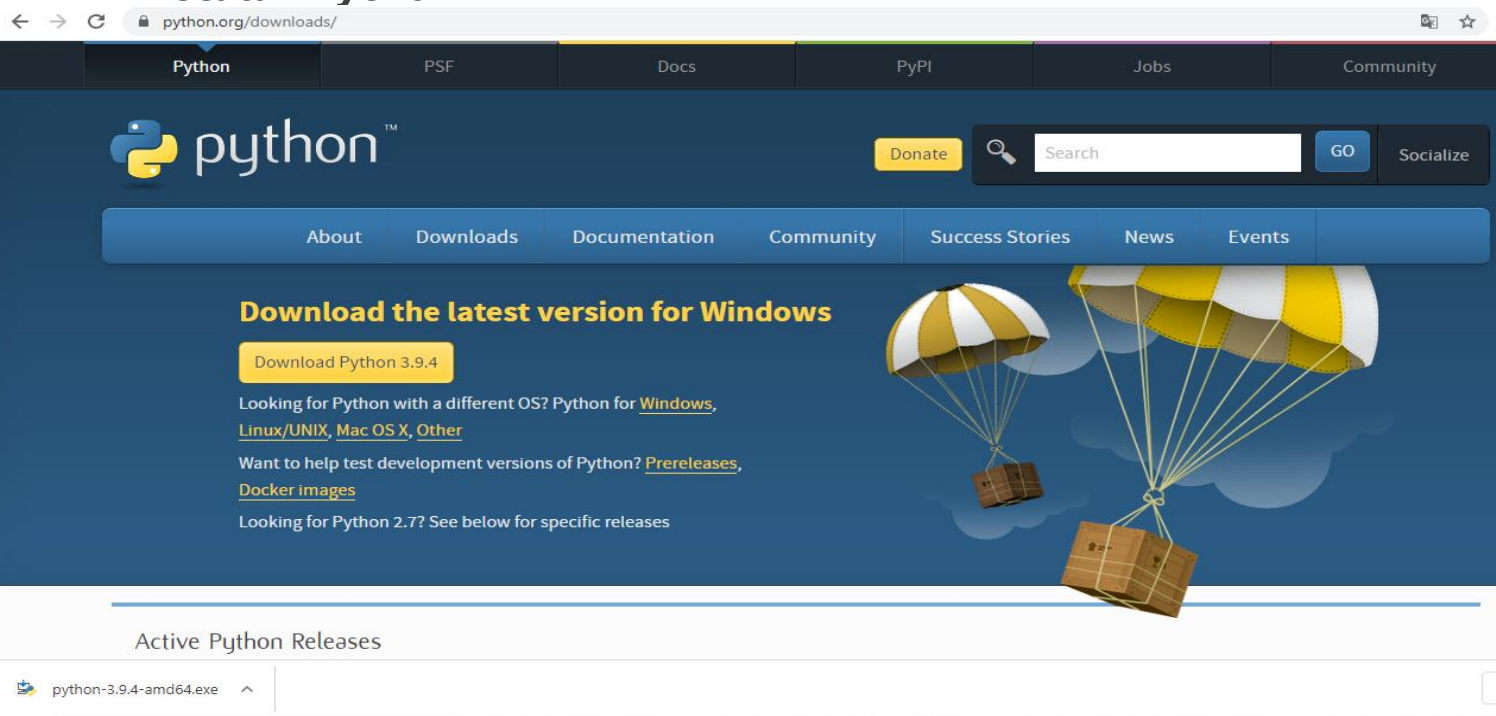
CONTENIDO

- Introducción a los recursos léxicos
- Recursos léxicos en NLTK
- NLTK para traducción de palabras
- Introducción a WordNet
- Explorando WordNet
- Similitud Semantica con WordNet
- Procesamiento de texto plano desde Web



Configuración de ambiente de trabajo

- **Windows**
- Ir a esta pagina para descargar Python <https://www.python.org/downloads/>
- **Instalar Python**

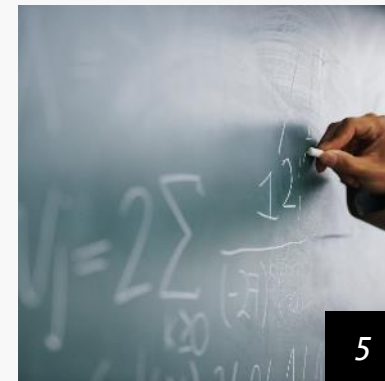
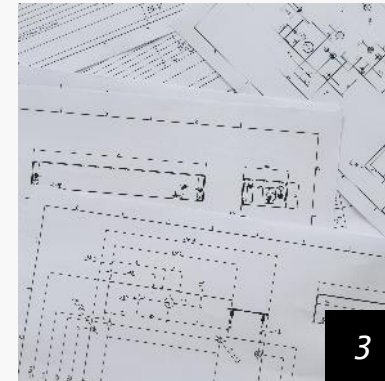
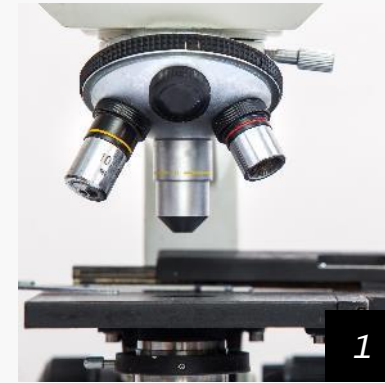


Configuración de ambiente de trabajo

- Ejecutamos lo que se descargo y sale esta pantalla

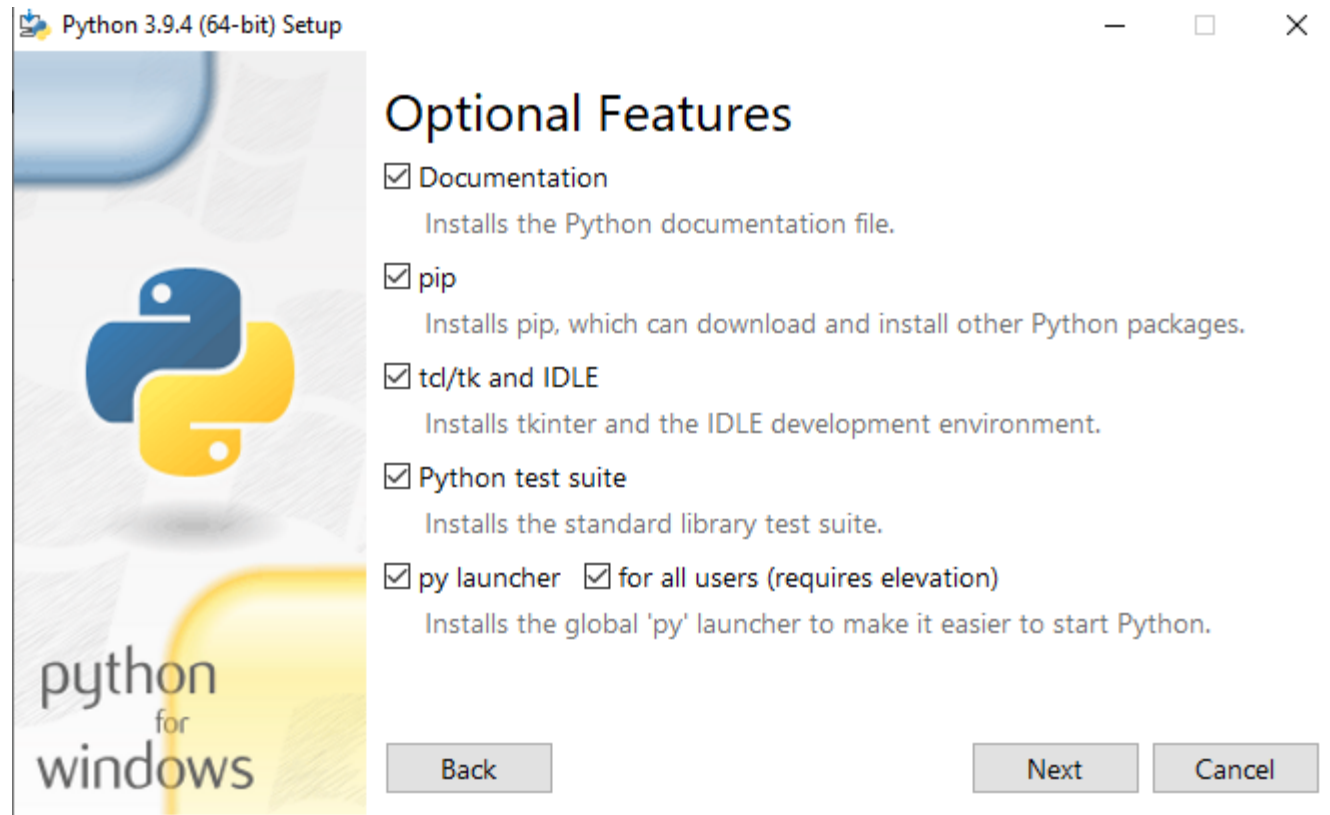


Seleccione customize installation



Configuración de ambiente de trabajo

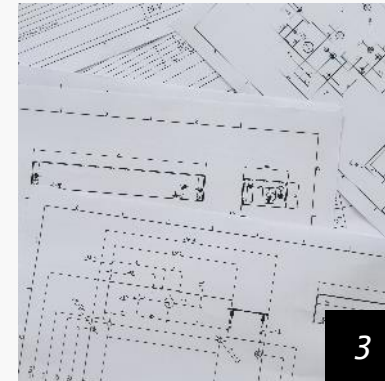
■ Haga clic en siguiente



1



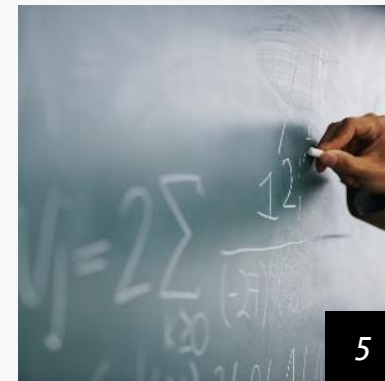
2



3



4



5



6

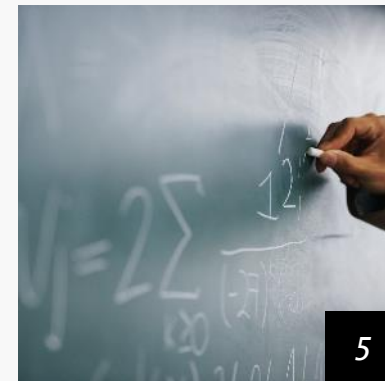
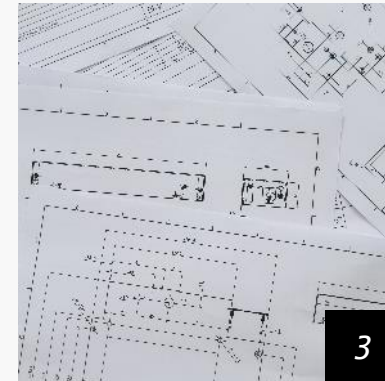
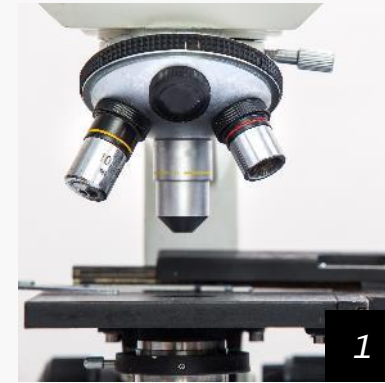
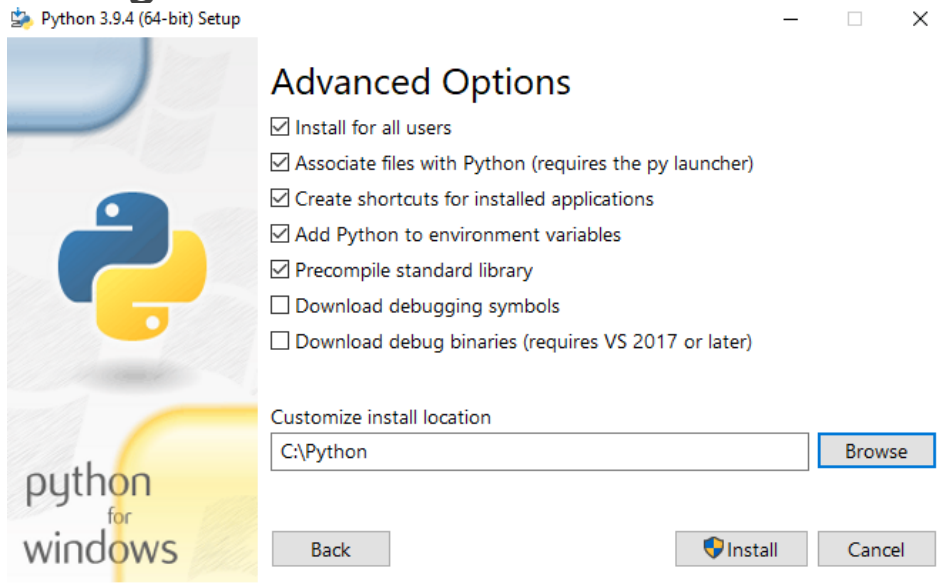
Configuración de ambiente de trabajo

▪ En la siguiente pantalla

1. Seleccione las opciones avanzadas

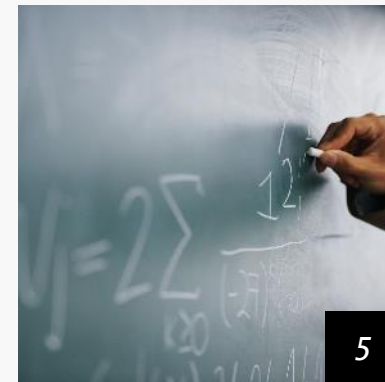
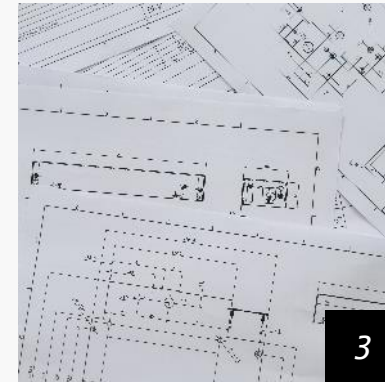
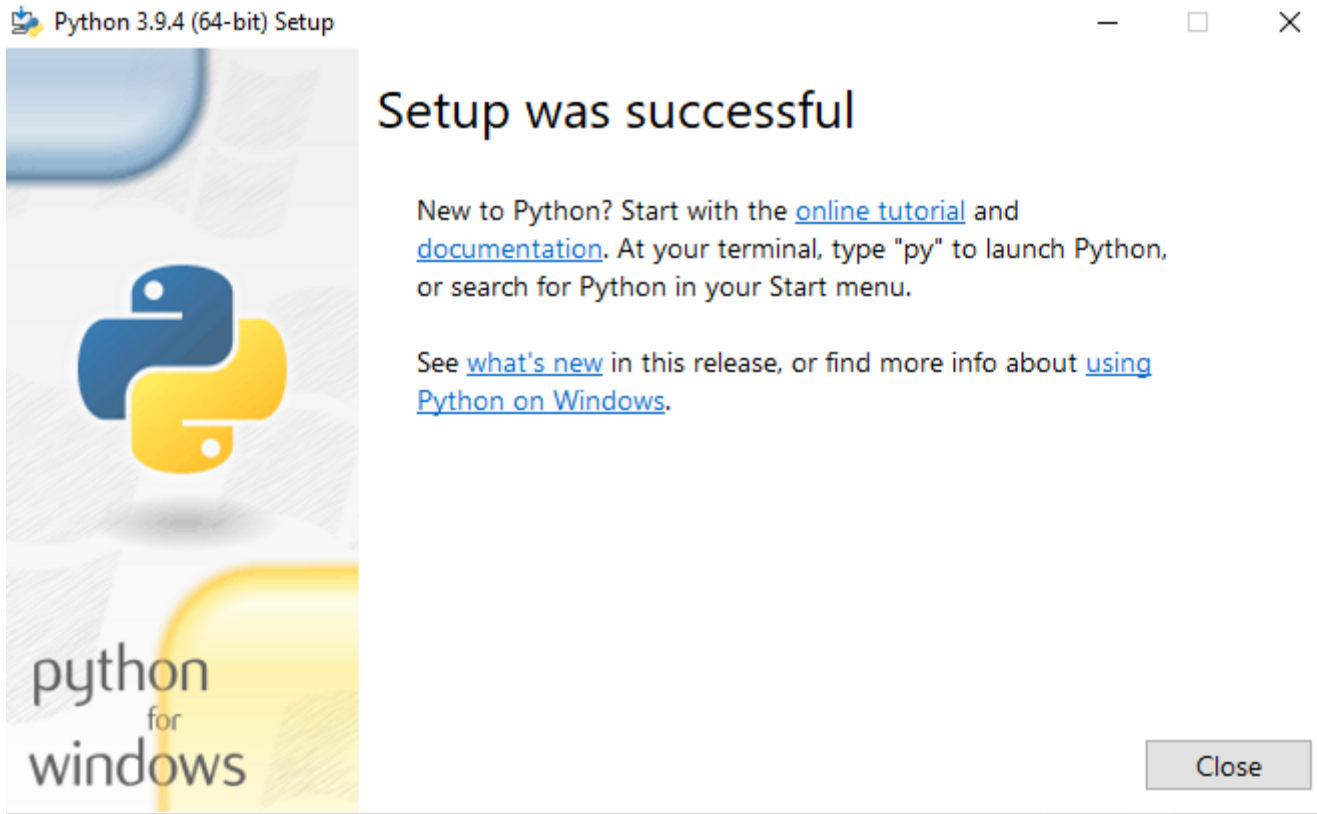
2. Dar una ubicación de instalación personalizada. En mi caso, se elige una carpeta en la unidad C para facilitar la operación

3. Haga clic en instalar



Configuración de ambiente de trabajo

- Haga clic en el botón close una vez que finalice la instalación.



Configuración de ambiente de trabajo

- copie la ruta de su carpeta de Scripts

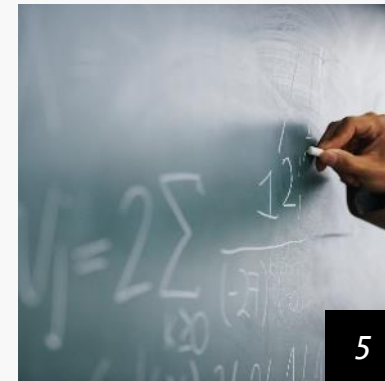
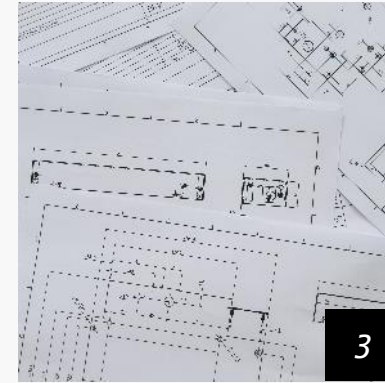
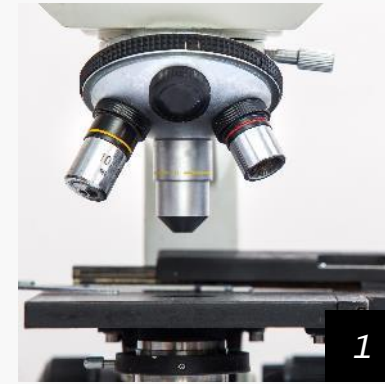
C:\Python\Scripts

- **si vamos a trabajar en Python entonces abrimos el símbolo de sistema**

- Navega hasta la ubicación de la carpeta pip
- Ingresa el comando para instalar NLTK

Pip3 install nltk

- La instalación debe realizarse con éxito



Configuración de ambiente de trabajo

```
Selecciónar Símbolo del sistema
Microsoft Windows [Versión 10.0.19042.867]
(c) 2020 Microsoft Corporation. Todos los derechos reservados.

C:\Users\andre>cd C:\Python\Scripts

C:\Python\Scripts>pip3 install nltk
Collecting nltk
  Using cached nltk-3.6.1-py3-none-any.whl (1.5 MB)
Collecting click
  Using cached click-7.1.2-py2.py3-none-any.whl (82 kB)
Collecting regex
  Using cached regex-2021.4.4-cp39-cp39-win_amd64.whl (270 kB)
Collecting tqdm
  Using cached tqdm-4.60.0-py2.py3-none-any.whl (75 kB)
Collecting joblib
  Using cached joblib-1.0.1-py3-none-any.whl (303 kB)
Installing collected packages: click, regex, tqdm, joblib, nltk
Successfully installed click-7.1.2 joblib-1.0.1 nltk-3.6.1 regex-2021.4.4 tqdm-4.60.0
WARNING: You are using pip version 20.2.3; however, version 21.0.1 is available.
You should consider upgrading via the 'c:\python\python.exe -m pip install --upgrade pip' command.

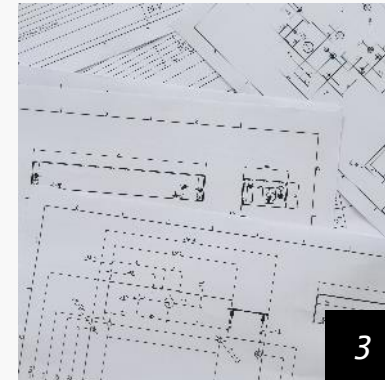
C:\Python\Scripts>
```



1



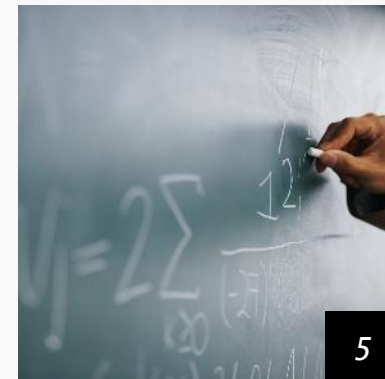
2



3



4



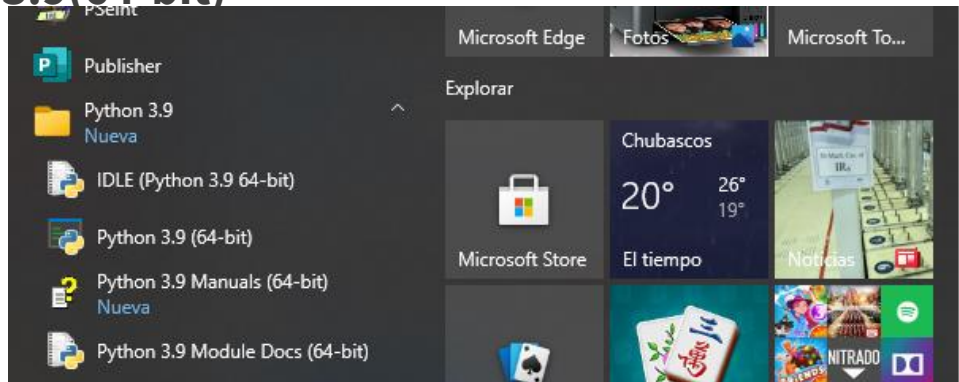
5



6

Configuración de ambiente de trabajo

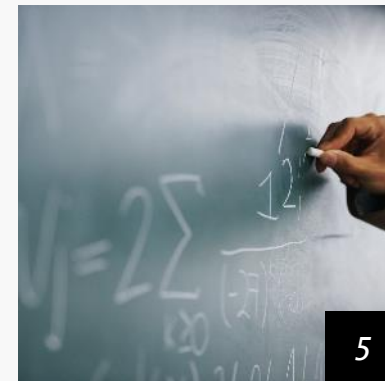
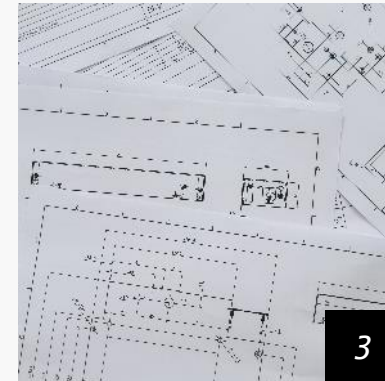
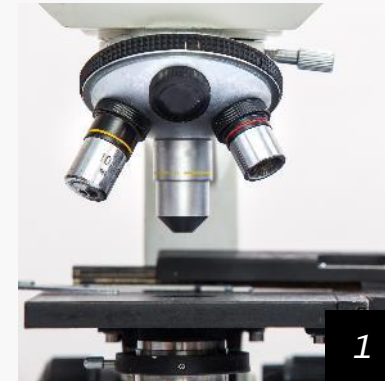
- En el menú inicio de Windows, busque y abra PythonShell(Python 3.9(64-bit))



- Puede verificar si la instalación es precisa proporcionando la siguiente comando

Import nltk

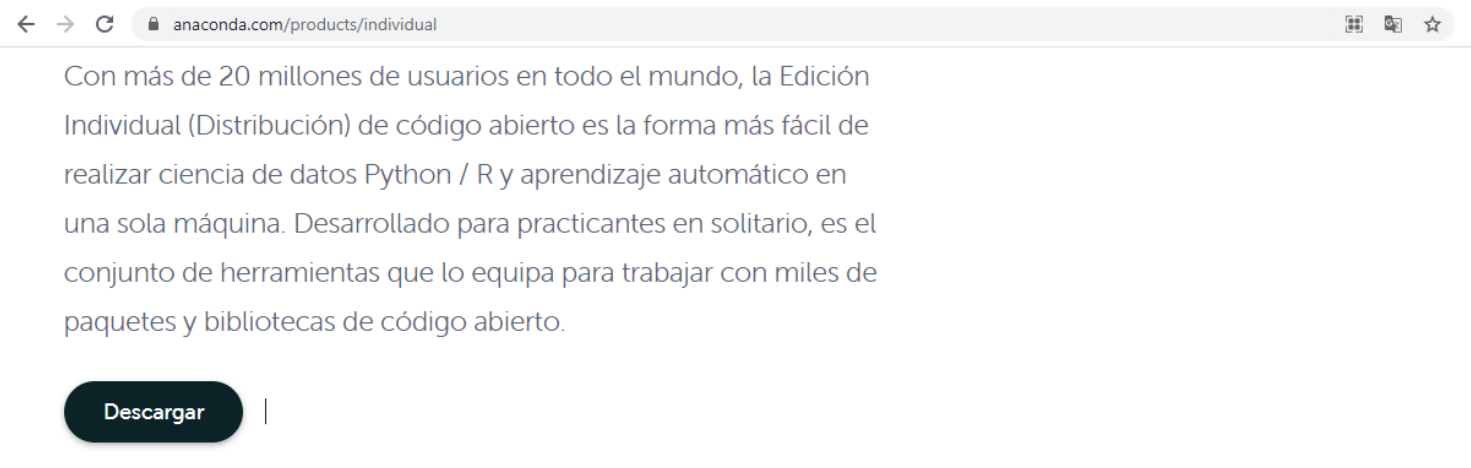
```
Python 3.9 (64-bit)
Python 3.9.4 (tags/v3.9.4:1f2e308, Apr 6 2021, 13:40:21) [MSC v.1928 64 bit (AMD64)] on win32
Type "help", "copyright", "credits" or "license" for more information.
>>> import nltk
>>>
```



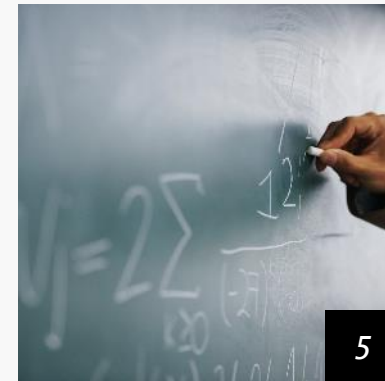
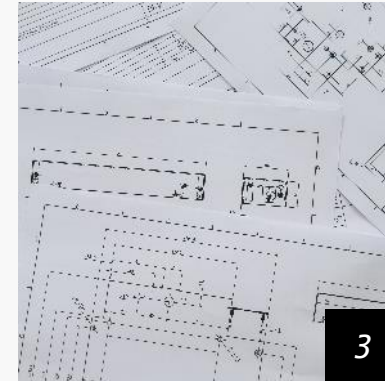
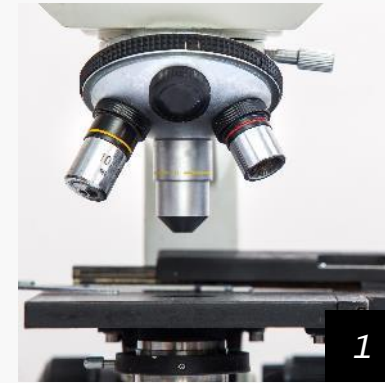
Configuración de ambiente de trabajo

Instalación de nltk a través de anaconda

- Ingresar a este enlace <https://www.anaconda.com/products/individual> para descargar anaconda ya sea para Windows , mac o linux



- En este link te explican como puedes instalarlo <https://www.guru99.com/download-install-r-rstudio.html>



Configuración de ambiente de trabajo

Instalación de nltk a través de anaconda

- Ya instalado anaconda
- Vaya a su símbolo del sistema y escriba el siguiente comando:

Conda install -c anaconda nltk



```
Anaconda Prompt (anaconda3) - conda install -c anaconda nltk

(base) C:\Users\andre>conda install -c anaconda nltk
Collecting package metadata (current_repodata.json): done
Solving environment: done

## Package Plan ##

  environment location: C:\Users\andre\anaconda3
  added / updated specs:
    - nltk

The following packages will be downloaded:

  package                        build      size  anaconda
  -----
  ca-certificates-2020.10.14    0         159 KB  anaconda
  nltk-3.5                      py_0      1.1 MB  anaconda
  openssl-1.1.1h                he774522_0 5.8 MB  anaconda
  -----
  Total:                          7.0 MB

The following packages will be SUPERSEDED by a higher-priority channel:

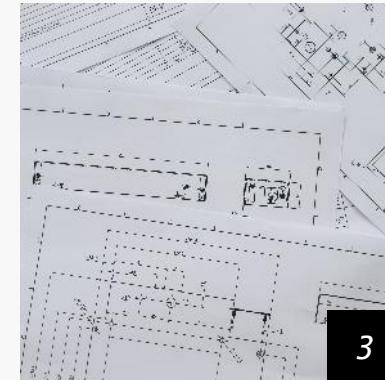
  ca-certificates  pkgs/main --> anaconda
  nltk             pkgs/main --> anaconda
  openssl         pkgs/main --> anaconda
```



1



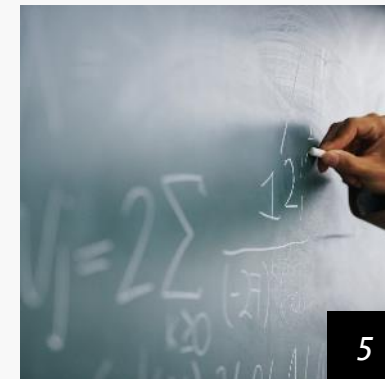
2



3



4



5



6

Configuración de ambiente de trabajo

Instalación de nltk a través de anaconda

- Le damos y

```
Anaconda Prompt (anaconda3)
The following packages will be downloaded:

package                                build                                size  channel
-----                                -
ca-certificates-2020.10.14             0                                  159 KB anaconda
nltk-3.5                               py_0                              1.1 MB anaconda
openssl-1.1.1h                         he774522_0                        5.8 MB anaconda
-----                                -
Total:                                7.0 MB

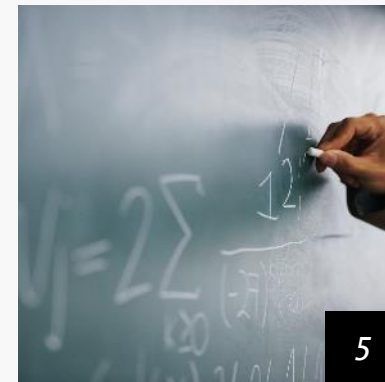
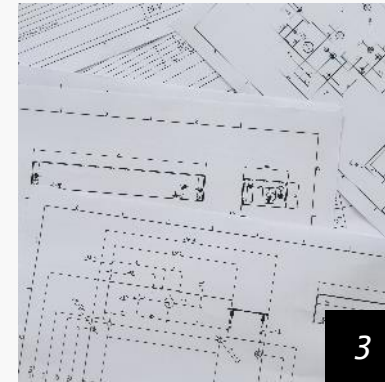
The following packages will be SUPERSEDED by a higher-priority channel:

ca-certificates                pkgs/main --> anaconda
nltk                          pkgs/main --> anaconda
openssl                        pkgs/main --> anaconda

Proceed ([y]/n)? y

Downloading and Extracting Packages
ca-certificates-2020 | 159 KB | ##### | 100%
openssl-1.1.1h       | 5.8 MB | ##### | 100%
nltk-3.5             | 1.1 MB | ##### | 100%
Preparing transaction: done
Verifying transaction: done
Executing transaction: done

(base) C:\Users\andre>
```



Comparación

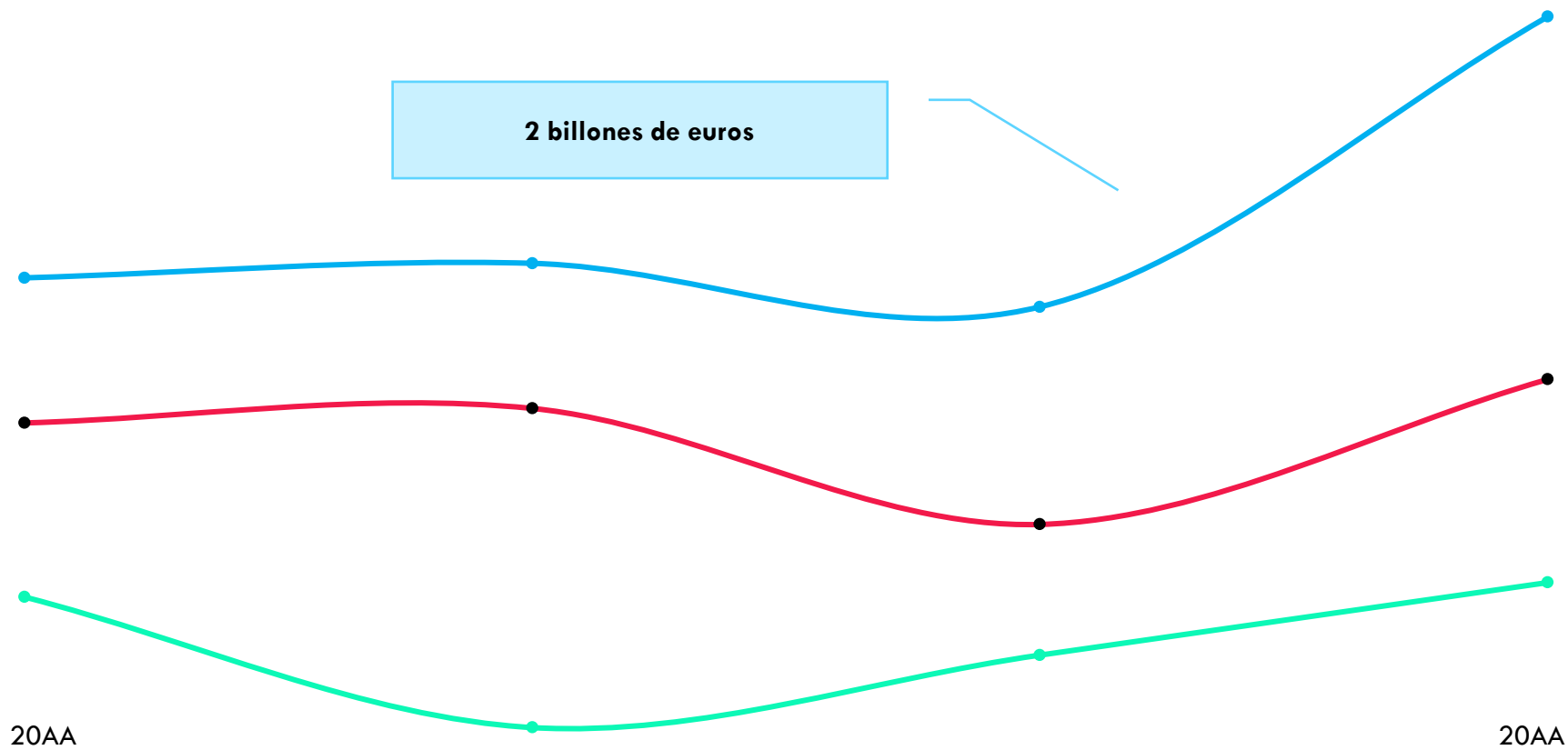
Comparar A

- Lorem ipsum Nulla a erat eget nunc hendrerit ultrices eu nec nulla. Donec viverra leo aliquet, auctor quam id, convallis orci.
 - Sed in molestie est. Cras ornare turpis at ligula posuere, sit amet accumsan neque lobortis.
 - Maecenas mattis risus ligula, sed ullamcorper nunc efficitur sed.

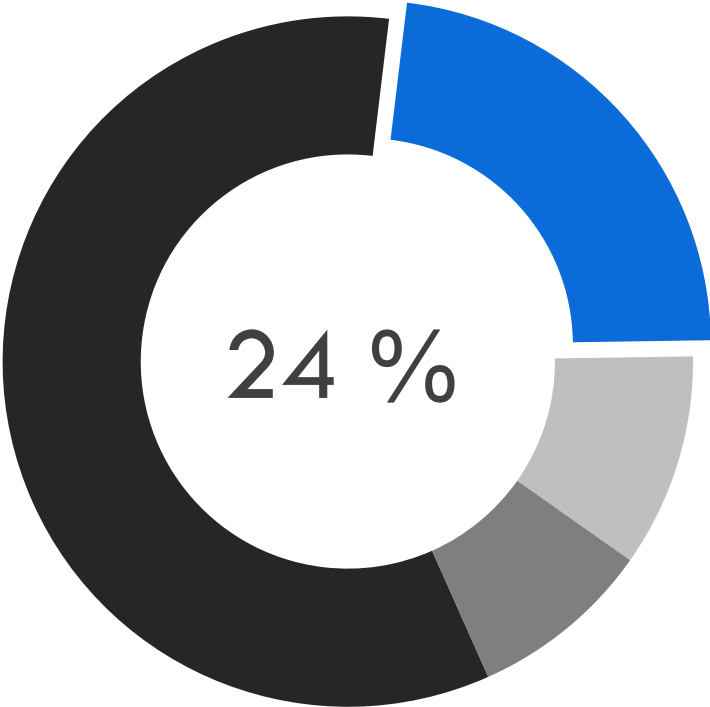
Comparar B

- Lorem ipsum Nulla a erat eget nunc hendrerit ultrices eu nec nulla. Donec viverra leo aliquet, auctor quam id, convallis orci.
 - Sed in molestie est. Cras ornare turpis at ligula posuere, sit amet accumsan neque lobortis.
 - Maecenas mattis risus ligula, sed ullamcorper nunc efficitur sed.

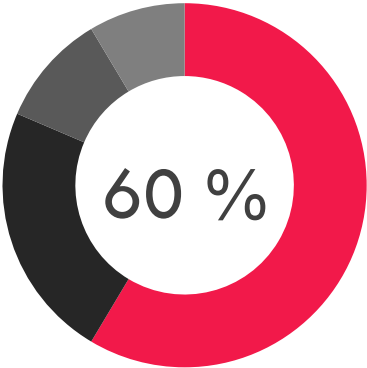
Diagramas y gráficos



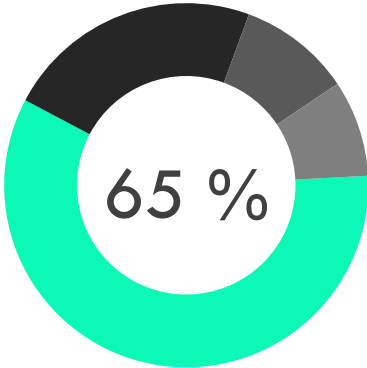
Diagramas y gráficos 1



Lorem ipsum dolor sit amet.

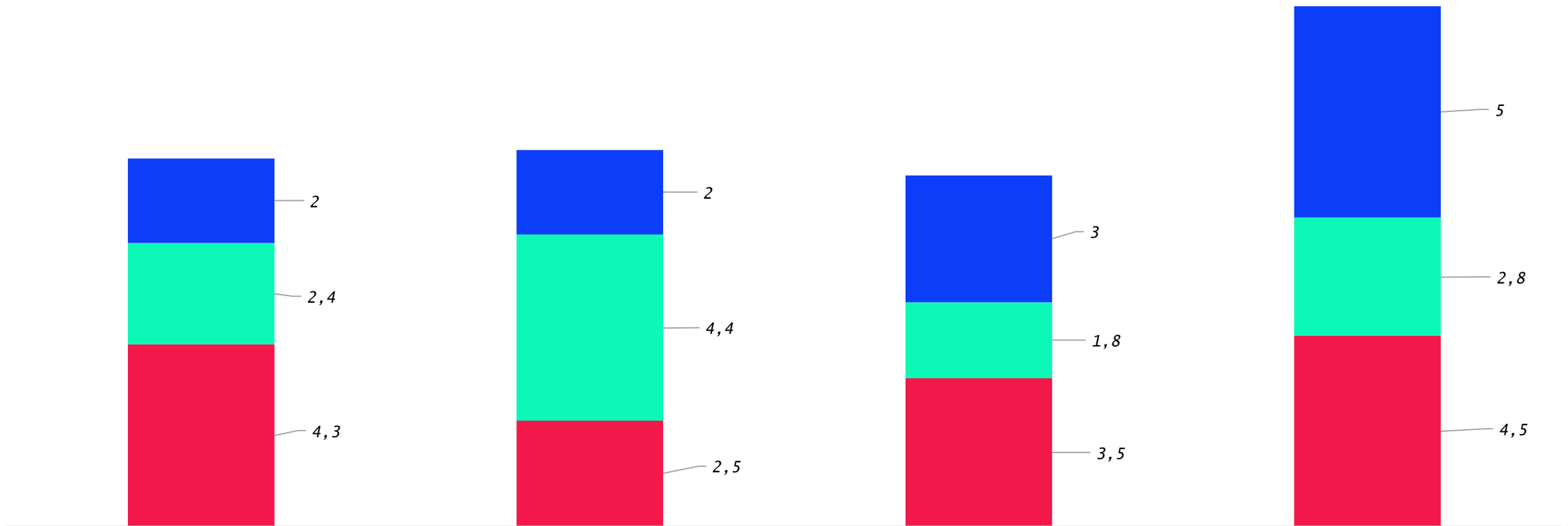


Lorem ipsum dolor
sit amet.



Lorem ipsum dolor
sit amet.

Diagramas y gráficos 2



Tabla

	TÍTULO DE LA COLUMNA UNO	TÍTULO DE LA COLUMNA DOS
Lorem Ipsum Dolor Sit Amet	5 000	40 000
Consectetur Adipiscing Elit	50 000	400 000
Etiam Aliquet Eu Mi Quis Lacinia	500 000	4 000 000
TOTAL	-1 968 750	-4 800 000



Caption lorem ipsum



Gracias

Pedro Armijo

✉ pedro@bellowscollege.com

