#### Bienvenidos al Curso Programación I!!

#### Maestría en Ciencia de Datos

Universidad de Guadalajara

IT PhD Rocio del Carmen Chávez Álvarez

## Contenido del Programa

#### Presentación

El maestrante en Ciencia de Datos debe tener bases sólidas de programación que le permitan desarrollar sus propios sistemas de análisis de datos, a la par de ser capaz de extraer la información de diferentes repositorios (bases de datos y/o archivos), gestionar su código con sistemas de control de versiones (Git), y probar y depurar su propio código

Con este curso, el futuro maestro en Ciencia de Datos, obtendrá las habilidades necesarias para dominar el lenguaje de programación Python

#### Objetivo General

Adquirir las bases del lenguaje de programación Python que le permitan al alumno desarrollar y probar cualquier tipo de programas en Python, pudiendo hacerlo de manera colaborativa

#### **Temario**

- Unidad 1 Arquitectura de Python
- Unidad 2 Fundamentos de Programación en Python
- Unidad 3 Funciones de Cadena
- Unidad 4 Archivos
- Unidad 5 Estructuras de Datos en Python
- Unidad 6 Expresiones Regulares
- Unidad 7 Sockets y Servicios Web
- Unidad 8 Bases de Datos SQL
- Unidad 9 Herramientas para depuración y pruebas de Código
- Unidad 10 Gestión de Repositorios de Código para control de versiones y desarrollo colaborativo con Git y GitHub

#### Evaluación

- Ejercicios Prácticos en Python 60%
- Ejercicios intermedios de Python 25%
- Ejercicio de Git 15%

Arquitectura de Python

### Arquitectura de Python

Tipos de Lenguajes de Programación

Características de Python

Usos de Python

Historia

Algunas Implementaciones

Versiones de Python

**IDEs** 

# Tipos de lenguajes

Se clasifican con base en:

Su compilación

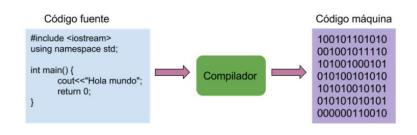
Necesidad de indicarle el tipo de variables a utilizar

Flexibilidad

Nivel de interpretabilidad

#### Tipos de lenguajes según su Compilación

*Compilado:* Primero se genera el script y después es traducido a un lenguaje que comprenda la máquina para crear un archivo ejecutable.



- Es necesario esperar a que el script sea compilado para saber si existen errores en el código
- Una vez compilados suelen ejecutarse más rápidamente

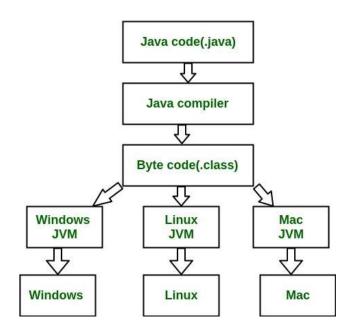
Ejemplos: C y C++

*Interpretado:* Requiere de un intérprete que va traduciendo el código al momento de irlo ejecutando. Ejemplos: Python y lenguaje R

*Intermedio:* Se compila el código fuente y luego se ejecuta dependiendo de la máquina en la que se esté ejecutando

Ejemplo: Java

#### Ejemplo de Lenguaje Intermedio



# Lenguajes Tipados y No Tipados

Lenguaje Tipado o Fuertemente Tipado: Tiene un sistema que define a qué tipo de dato pertenecen los valores y las expresiones utilizadas, así como la manera en la que los tipos se pueden manipular y como pueden interactuar.

Debido a que cualquier valor simplemente consiste en un conjunto de bits de un <u>ordenador</u>, el <u>hardware</u> no hace distinción entre <u>caracteres</u>, <u>enteros</u> y <u>números en coma flotante</u> o cualquier otro tipo de dato, por lo que es necesario informar a la computadora cómo deben ser tratados los bits utililzados para almacenar los datos.

```
Por ejemplo en C++ int f = 5
```

Lenguaje No Tipado o Débilmente Tipado: Si en una variable almaceno un número y después en esa misma variable almaceno un caracter, no hay problema.

Python pertenece a este tipo de lenguajes

# Lenguajes Estáticos y Dinámicos

*Lenguaje Estático:* El tipo de una variable no puede cambiar una vez habiendo sido declarado ya que la comprobación del tipo se lleva a cabo al momento de la compilación

*Lenguaje dinámico:* El intérprete asigna el tipo a las variables durante el tiempo de ejecución basado en el valor que tiene en ese momento

#### Lenguajes de Alto y Bajo Nivel

Lenguaje de Alto Nivel: Utilizan instrucciones que son fácilmente comprendidos por el ser humano, ya que son muy parecidas al idioma inglés

Lenguaje de Bajo Nivel: Contiene instrucciones básicas reconocibles solamente por la computadora



Imagen tomada de: https://moonantonio.github.io/post/2018/dev/004/

#### Características de Python

- Fácil de aprender
- Fácil de leer
- Desarrollo de programas en poco tiempo
- Interactivo
- Existen muchas librerías útiles para diferentes tipos de desarrollos
- Interpretado, no es necesario compilar el código. (Lenguaje de Scripting)
- Debido a que no es compilado, no es veloz
- De código abierto
- Utilizado en muchas ocasiones para construir prototipos

#### C++

```
#include <iostream>
using namespace std;
int main()
{
    cout << "Hello world!" << endl;
    return 0;
}</pre>
```

#### Java

```
public class HelloWorld {

/**
    * @param args
    */
public static void main(String[] args) {
    // TODO Auto-generated method stub
    System.out.println("Hello world!");
}
```

#### Python

print("Hello World")

#### Tiempo promedio requerido para crear un programa

# Mediana de horas necesarias para resolver un problema Python TCL Lisp Rexx C Java C++ 0 2 4 6 8 10 12

Fuente: Prechelt & Garret, tiempo total necesario para producir un programa de búsqueda/procesado de texto usando diferentes lenguajes.

#### Usos de Python

- Aplicaciones Web
- Ciberseguridad
- Desarrollo Back End
- Extracción de Información de sitios web (Web Scraping)
- Automatización
- Desarrollo de juegos
- Análisis de Datos
- Machine Learning
- Deep Learning
- IoT







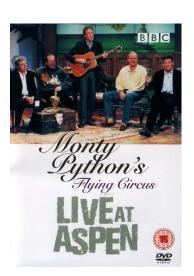


#### Algunas Empresas que Utilizan Python



#### Historia de Python

- Creado por Guido Van Rossum a finales de los 80's y principios de los 90's
- Está basado en otro lenguaje llamado ABC desarrollado por la compañía para que trabajaba Guido Van Rossum en los años 80's
- Lo llamó Python en honor a un grupo de cómicos famosos llamado Monty Python
- Versiones:
  - 0.9.0 febrero de 1991
  - 1.0 en 1994
  - 2.0 en octubre del año 2000
  - 3.0 en Diciembre del 2008
- Se encuentran vigentes las versiones 2 y 3, las cuales son incompatibles
- En la actualidad existen conferencias anuales llamadas PyCon



## ABC vs. Python

#### Recuperar las palabras de un documento en ABC

# HOW TO RETURN words document: PUT {} IN collection FOR line IN document: FOR word IN split line: IF word not.in collection: INSERT word IN collection RETURN collection

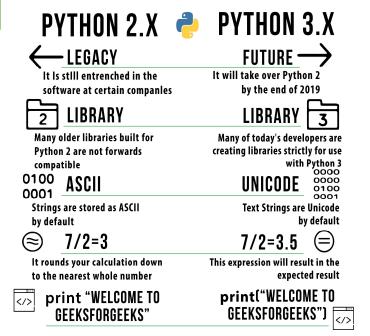
#### Recuperar las palabras de un documento en Python

```
def words(document):
    collection = set()
    for line in document:
        for word in line.split():
            if word not in collection:
                 collection.add(word)
    return collection
```

### Versiones de Python

- La última versión de Python 2, fue la 2.7
- Python 3
- MicroPython Corre en microcontroladores
  - Raspberry Pi
  - Micro:bit
  - Calliope

#### Python 2 vs Python 3



Existen muchas diferencias en el código de ambas vesiones, por lo que no pueden ser compatibles

Unicode tiene muchísimos caracteres en comparación con lo caracteres que contiene el código ASCII

En Python 2 la división arrojaba solamente el número entero del cociente, en Python 3, podemos obtener la cifra completa, incluyendo los decimales

En Python 2 print era una simple instrucción, en Pyhon 3, print es una función

# Algunas Implementaciones de Python CPython El intérprete está escrito en lenguaje C





IronPython Soportado por .Net y Mono(para móviles), las cuales son plataformas de Microsoft destinadas a desarrollar aplicaciones



PyPy Permite ejecutar código mucho más rápido que Cpython



*Cython* Optimiza a Python, ya que lo traduce a C++



#### Cython

# Python Python using namespace std; int main() { cout << "Hello World" << endl; return 0;</pre>

C++

#### Ejemplo de Python vs. Cython

#### Python

```
def f(x):
    return x**2-x

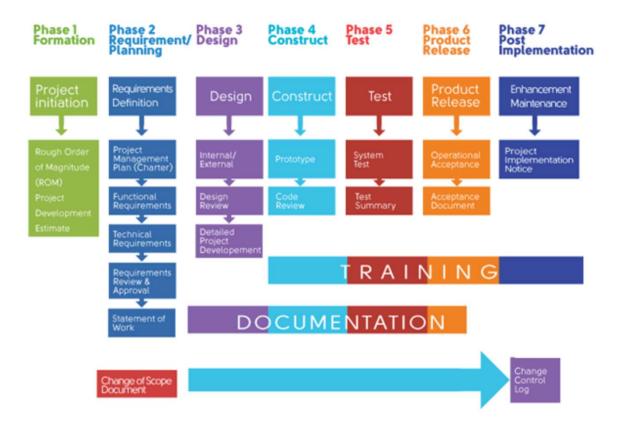
def integrate_f(a, b, N):
    s = 0
    dx = (b-a)/N
    for i in range(N):
        s += f(a+i*dx)
    return s * dx
```

#### Cython

```
cdef f(double x):
    return x**2-x

def integrate_f(double a, double b, int N):
    cdef int i
    cdef double s, x, dx
    s = 0
    dx = (b-a)/N
    for i in range(N):
        s += f(a+i*dx)
    return s * dx
```

#### Fases de Desarrollo de Software



#### Pseudocódigo

- No contiene códigoFácil de comprenderPermite concentrarse en la solución del problema
- Facilita la detección de errores

#### Ejemplo de Pseudocódigo

Obtención del promedio de los alumnos y del conteo de los aprobados

Inicializar contador de aprobados a cero

Inicializar contador de estudiantes a uno

Inicializar variable calificación a cero

Obtener la cantidad de estudiantes

Mientras el contador de estudiantes sea menor o igual al total de estudiantes

Agregar resultado del examen a la variable calificación

Si el resultado del examen es mayor a 59

Sumar 1 al contador de aprobados

Incrementar en uno el contador de estudiantes

Si no

Incrementar en uno el contador de estudiantes

Finaliza la sentencia if-else

Finaliza el ciclo while

Calcular el promedio

Imprimir en pantalla el promedio

Imprimir en pantalla la cantidad de aprobados

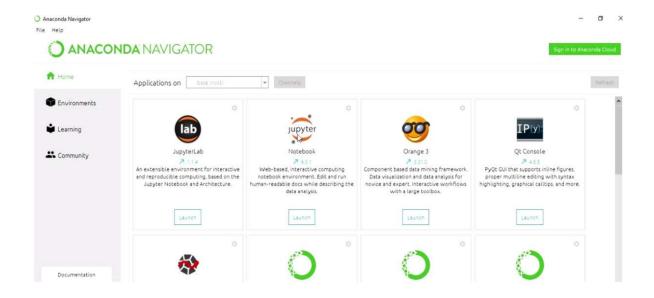
**Nota:** La indentación es muy importante en la claridad del pseudocódigo

#### IDEs para Python



IDE: Integrated Development Environment

#### Instalación de Anaconda



Ver instrucciones en el video Instalación de Jupyter Notebook (Anaconda) https://www.youtube.com/watch?v=mOCy5bq4AYw&ab\_channel=RocioChavezCienciadeDatos

#### Plan Propuesto de Trabajo

Anaconda Powershell Prompt (base) PS C:\Users\rocio> python Python 3.7.4 (default, Aug 9 2019, 18:34:13) [MSC v.1915 64 bit (AMD64)] :: Anaconda custom (64-bit) on win32 Type "help", "copyright", "credits" or "license" for more information. C:\Users\rocio\Python\_en\_terminal\prueba2.py - Sublime Text (UNREGISTERED) File Edit Selection Find View Goto Tools Project Preferences Help Anaconda Powershell Prompt print("Hola, estoy haciendo pruebas para correr python desde la terminal") (base) PS C:\Users\rocio> import numpy as np import pandas as pd Spyder Notebook 7 5.1.1 7 6.4.3 Scientific PYthon Development Web-based, interactive computing EnviRonment. Powerful Python IDE with notebook environment. Edit and run advanced editing, interactive testing, human-readable docs while describing the debugging and introspection features data analysis.

Utilizando Python desde la terminal de Anaconda

# Actualización de Paquetes

Desde al Anaconda Powershell Prompt

>conda update conda

>conda update --all

Nota: En windows es necesario abrir el prompt como administrador, dando clic en el botón derecho del mouse

#### Lista de paquetes instalados en Anconda

#### Anaconda Prompt

```
(base) C:\Users\rocio>conda list
  packages in environment at C:\Users\rocio\Anaconda3:
                          Version
                                                     Build Channel
 ipyw_jlab_nb_ext_conf
                          0.1.0
                                                    py37_0
absl-py
                          0.7.1
                                                    pypi_0
                                                              pypi
alabaster
                          0.7.12
                                                    py37_0
anaconda
                          2019.10
                                                    py37_0
anaconda-client
                          1.7.2
                                                    py37_0
anaconda-navigator
                          1.10.0
                                                    py37_0
anaconda-project
                          0.8.3
                                                              conda-forge
                                                      py_0
                          0.0.13
                                              pyh6c4a22f 0
                                                              conda-forge
anyqt
asn1crypto
                          1.0.1
                                                    py37_0
                                                              conda-forge
astor
                          0.8.0
                                                    pypi_0
                                                              pypi
astroid
                          2.3.1
                                                    py37_0
                                                              conda-forge
                                            py37he774522 0
astropy
                          3.2.1
atomicwrites
                          1.3.0
                                                    py37_1
                          19.2.0
                                                              conda-forge
attrs
                                                      py_0
babel
                                                              conda-forge
                          2.7.0
                                                      py_0
backcall
                          0.1.0
                                                    py37_0
backports
                          1.0
                                                      py_2
                                                              conda-forge
backports.os
                          0.1.1
                                                    py37 0
backports.shutil get terminal size 1.0.0
                                                             py37_2
beautifulsoup4
                          4.8.0
                                                              conda-forge
                                                    py37_0
bitarray
                                            py37he774522 0
                          1.0.1
bkcharts
                          0.2
                                                    py37_0
                                                       mk1
blas
                          1.0
bleach
                          3.1.0
                                                    py37 0
blosc
                          1.16.3
                                                h7bd577a 0
bokeh
                                                              conda-forge
                          1.3.4
                                                    py37_0
                          2.49.0
                                                    py37 0
boto
```

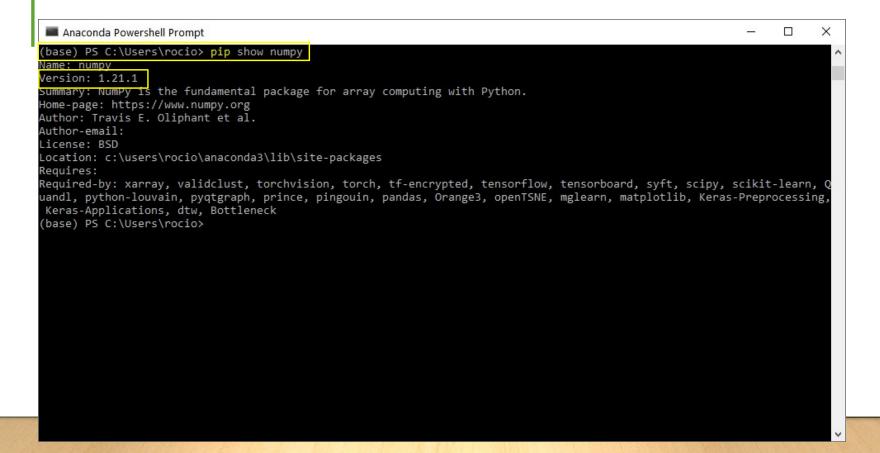
#### Instalando paquetes

```
(base) PS C:\Users\rocio> conda install matplotlib
Collecting package metadata (repodata.json): done
Solving environment: done
## Package Plan ##
 environment location: C:\Users\rocio\Anaconda3
 added / updated specs:
   - matplotlib
The following packages will be downloaded:
   package
                                    build
   ca-certificates-2021.5.30
                                h5b45459_0
                                                171 KB conda-forge
                                                142 KB conda-forge
  certifi-2021.5.30
                             py37h03978a9 0
  matplotlib-3.3.2
                             py37h03978a9_1
                                                 7 KB conda-forge
   matplotlib-base-3.3.2
                             py37h64d6e1c_1
                                                6.8 MB conda-forge
                                                5.7 MB conda-forge
                                h8ffe710 0
  openssl-1.1.1k
                                   Total:
                                               12.8 MB
The following NEW packages will be INSTALLED:
 matplotlib-base conda-forge/win-64::matplotlib-base-3.3.2-py37h64d6e1c_1
The following packages will be UPDATED:
 ca-certificates anaconda::ca-certificates-2020.10.14-0 --> conda-forge::ca-certificates-2021.5.30-h5b45459_0
                 anaconda::certifi-2020.6.20-py37_0 --> conda-forge::certifi-2021.5.30-py37h03978a9_0 pkgs/main::matplotlib-3.1.1-py37hc8f6\sim --> conda-forge::matplotlib-3.3.2-py37h03978a9_1
 certifi
 matplotlib
 openssl
                   anaconda::openssl-1.1.1h-he774522_0 --> conda-forge::openssl-1.1.1k-h8ffe710_0
Proceed ([y]/n)? y
Downloading and Extracting Packages openssl-1.1.1k | 5.7 MB |
matplotlib-base-3.3.
                 6.8 MB
                             matplotlib-3.3.2
                  7 KB
142 KB
                             certifi-2021.5.30
                             ca-certificates-2021 | 171 KB
```

Para instalar una versión en específico:

Ejemplo conda install matplotlib=3.1.1

#### Consultar versión de Paquetes



#### Utilizando el prompt de Anaconda

```
Anaconda Prompt - python
                                                                                                                X
 rror processing line 1 of C:\Users\rocio\Anaconda3\lib\site-packages\matplotlib-3.1.1-py3.7-nspkg.pth:
 Traceback (most recent call last):
   File "C:\Users\rocio\Anaconda3\lib\site.py", line 168, in addpackage
     exec(line)
   File "<string>", line 1, in <module>
   File "<frozen importlib._bootstrap>", line 580, in module_from_spec
 AttributeError: 'NoneType' object has no attribute 'loader'
Remainder of file ignored
(base) C:\Users\rocio>python
rror processing line 1 of C:\Users\rocio\Anaconda3\lib\site-packages\matplotlib-3.1.1-py3.7-nspkg.pth:
 Traceback (most recent call last):
   File "C:\Users\rocio\Anaconda3\lib\site.py", line 168, in addpackage
     exec(line)
   File "<string>", line 1, in <module>
   File "<frozen importlib._bootstrap>", line 580, in module_from_spec
 AttributeError: 'NoneType' object has no attribute 'loader'
Remainder of file ignored
Python 3.7.4 (default, Aug 9 2019, 18:34:13) [MSC v.1915 64 bit (AMD64)] :: Anaconda, Inc. on win32
Type "help". "copyright", "credits" or "license" for more information.
```

Utilizar ctrl+z para salir de python

#### Ayuda en el prompt de Python

```
>>> help()
Welcome to Python 3.7's help utility!
```

Utilizar ctrl+c para salir de la ayuda

```
help> matplotlib
Help on package matplotlib:
   matplotlib - An object-oriented plotting library.
DESCRIPTION
   A procedural interface is provided by the companion pyplot module,
   which may be imported directly, e.g.::
       import matplotlib.pyplot as plt
   or using ipython::
       ipython
   at your terminal, followed by::
       In [1]: %matplotlib
       In [2]: import matplotlib.pyplot as plt
   at the ipython shell prompt.
   For the most part, direct use of the object-oriented library is encouraged when
   programming; pyplot is primarily for working interactively. The exceptions are
   the pyplot functions `.pyplot.figure`, `.pyplot.subplot`, `.pyplot.subplots`, and `.pyplot.savefig`, which can greatly simplify scripting.
   Modules include:
       :mod:`matplotlib.axes`
            The `~.axes.Axes` class. Most pyplot functions are wrappers for
            `~.axes.Axes` methods. The axes module is the highest level of OO
            access to the library.
        :mod: `matplotlib.figure`
            The `.Figure` class.
        :mod:`matplotlib.artist`
 Más --
```

#### Limpiar el prompt de Python

import os

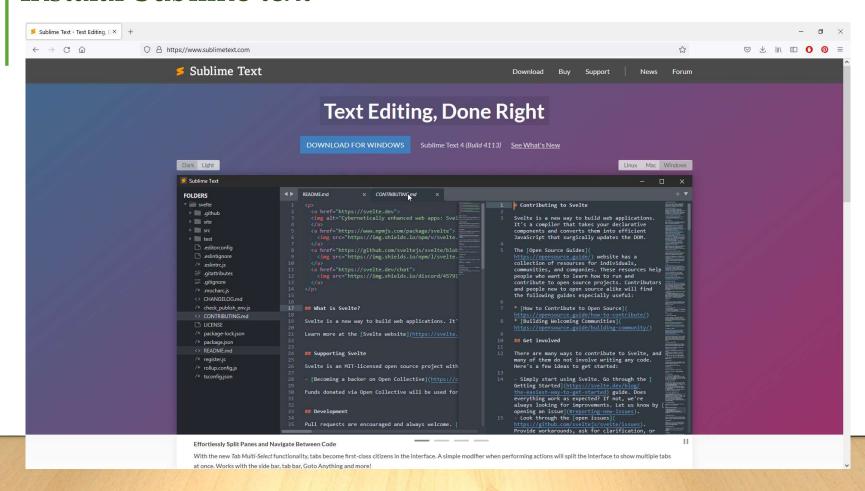
os.system("cls")

#### Ejecutando código en el prompt de Python

Anaconda Powershell Prompt

2
2
11.5
>>> 23/2
11.5
>>> 23//2
11
>>> \_

#### Instalar Sublime text



#### Creación de un script en Sublime Text

#### Ejecutando un script en el prompt de Anaconda Opción 1

- En caso de que se hayan generado archivos al correr el script con anterioridad, es mejor borrarlos antes de crearlos de nuevo
- Copiamos todas las líneas del script
- Vamos al prompt de Python
- Damos clic en el botón derecho del mouse o ctrl v

#### Anaconda Powershell Prompt

```
>> #Indicamos el directorio en el que vamos a trabajar
   import os
>>> os.chdir('C:/Users/rocio/Archivos en Jupyter/Bases de datos a analizar/')
>>> #Importamos los paquetes que vamos a necesitar
>>> import pandas as pd #Contiene funciones que nos ayudan en el análisis de datos
   import matplotlib.pyplot as plt
>>> #Leemos el archivo a analizar
   atletas = pd.read_csv('categorias de corredores.csv', index_col=0) #Con index_col=0 le indicamo
>>> atletas.info()
<class 'pandas.core.frame.DataFrame'>
Int64Index: 1000 entries, 3475 to 8527
Data columns (total 6 columns):
     Column
               Non-Null Count Dtype
    Lugar
               1000 non-null
                               int64
                1000 non-null
                               object
                1000 non-null
    Edad
                               int64
                999 non-null
                1000 non-null
                               float64
    Velocidad 1000 non-null
                               object
types: float64(1), int64(2), object(3)
memory usage: 54.7+ KB
>>> #Vemos las primeras filas del archivo
 .. atletas.head()
          Lugar Genero Edad Pais
                                        Tiempo Velocidad
Corredor
3475
          3592
                  Male
                              GBR 217.483333
                                                 Regular
13594
          13853
                Female
                          40
                               NY 272.550000
                                                 Regular
12012
         12256
                  Male
                          31 FRA 265.283333
                                                 Regular
10236
                Female
          10457
                               MI 256.150000
                                                 Regular
9476
          9686
                  Male
                                    252.250000
                                                 Regular
```

#### Ejecutando un script en el prompt de Anaconda

- Vamos al prompt de Anaconda
- Corremos el script con el comando "python" y el nombre del script entre comillas

#### Anaconda Powershell Prompt

# Creando un Archivo Ejecutable en el prompt de Anaconda

- Vamos al prompt de Anaconda
- Instalamos el paquete pyinstaller con el comando *pip install pyinstaller*

```
Anaconda Powershell Prompt

(base) PS C:\Users\rocio\Escritorio2\Pavillion-Rocio\Cursos impartidos Presenciales\Programacion I MCD\Scripts>

pip install pyinstaller

Collecting pyinstaller-4.5.1-py3-none-win_amd64.whl (1.9 MB)

| 1.9 MB 1.1 MB/s

Collecting altgraph
```

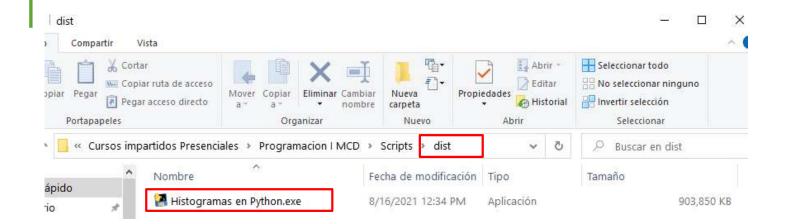
# Creando un Archivo Ejecutable en el prompt de Anaconda (Cont.)

- Dentro de la carpeta en donde se encuentra el archivo que queremos convertir a ejecutable, tecleamos

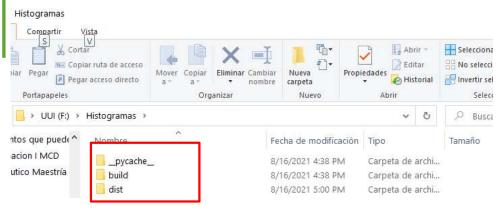
pyinstaller --onefile nombre\_del\_archivo.py



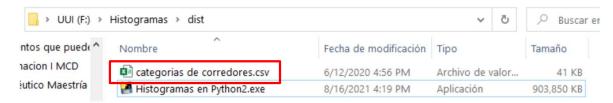
#### Ubicación del Archivo Ejecutable



#### Ubicación del Archivo Ejecutable en una USB



**Nota importante:** Es necesario colocar el archivo .csv en la misma carpeta en la que se encuentra el ejecutable



#### Tarea 1

- Crear un archivo ejecutable que realice lo siguiente:
  - Lea un archivo propio
  - Obtenga el histograma de una de las variables
  - Muestre el gráfico
  - Lo guarde en un archivo de tipo jpg
- El ejecutable deberá llevará tu nombre y apellido
- Subir a Google Drive:
  - El ejecutable
  - El archivo .csv que analizaste
  - El script .py del ejecutable
  - Un documento en Word con la captura de las pantallas del comando que utilizaste para crear el ejecutable
- Subir el link de Google Drive al classroom para que yo pueda revisar tu tarea