

# Data Science

w w w . d a t a s c i e n c e . p e

# Automatización de procesos en Python

Luiggi Silva – Data Scientist

Octubre - 2020

---

Mi nombre es Luiggi Silva

Mi experiencia :

**Sector Privado : Banca (BCRP-Pichincha)**

**Sector Publico: MEF**

**Academia: FMI – San Marcos**

**Voluntariado: Saturdays ai- DSRP**





# ¿Qué es Python?

# ¿Qué es Python?

Lenguaje de programación de alto nivel para programación general, creado por Guido van Rossum y lanzado por primera vez en 1991.



- Es fácil de leer y entender
- Productivo
- Opera en múltiples sistemas operativos
- Tiene muchas librerías
- Es un software abierto
- Amplia comunidad

¿Ventajas de  
python?



# ¿Por qué estudiar Python?

Python toma la codificación como lenguaje humano natural.

```
a = 10  
b = 12
```

```
sum = a + b  
print (sum)
```

No necesita  
definir el tipo de  
una variable en  
Python.

```
a = 10
b = "Ten"
c = 0.10
d = -12
e = False
```

Python le permite escribir programas que tengan una mayor funcionalidad con menos líneas de código.

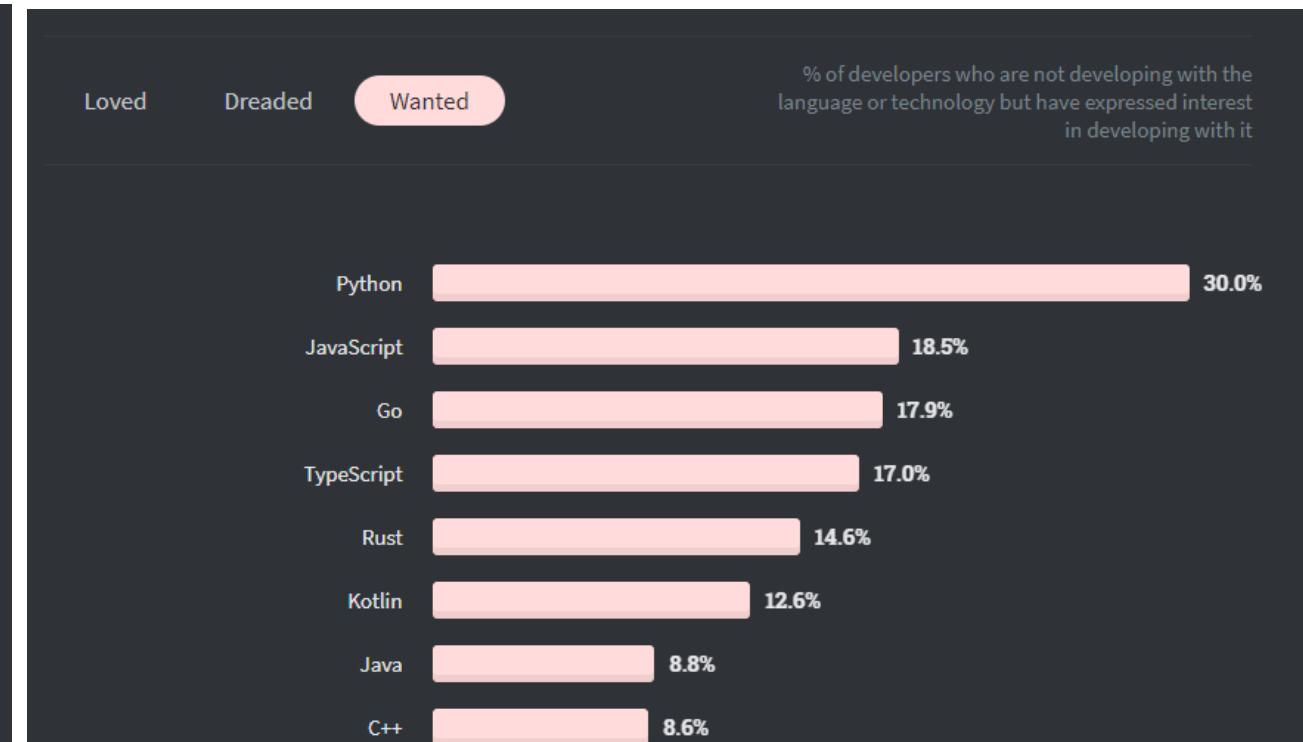
```
def getCount(inputStr):
    """
    return vowel count from the given string.
    """
    return sum(1 for letter in inputStr if letter in
               'aeiouAEIOU')
```

## Es flexible

Como lenguaje de tipo dinámico, Python es realmente flexible. Esto significa que no hay reglas estrictas sobre cómo crear funciones, y tendrá más flexibilidad para resolver problemas utilizando diferentes métodos.

Además, Python también es más indulgente con los errores, por lo que aún podrá compilar y ejecutar su programa hasta que llegue a la parte problemática.

# Su importancia en la industria



## ¿Qué empresas usas Python?

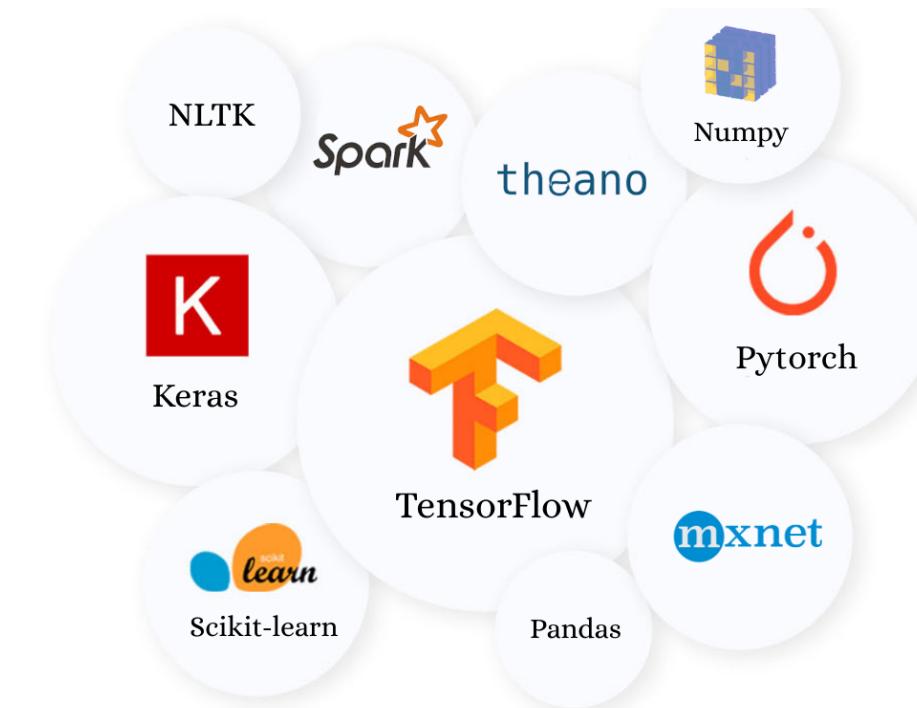
### Top Companies Using Python



- Automatización
- Data Science
- Web Development
- Machines (IOT-Robotica)
- Videojuegos

¿Qué se  
puede hacer  
con Python?

# Data Science framework y librerias



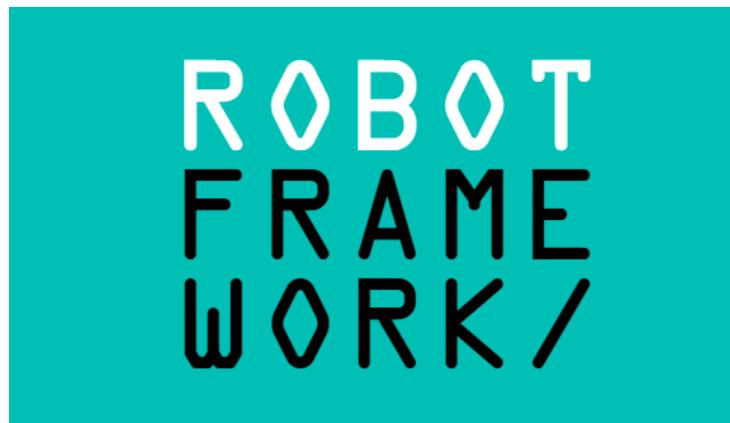
# Video Games framework



# Web Development framework



# Robotica/IOT



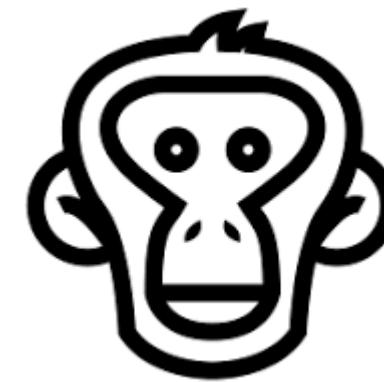
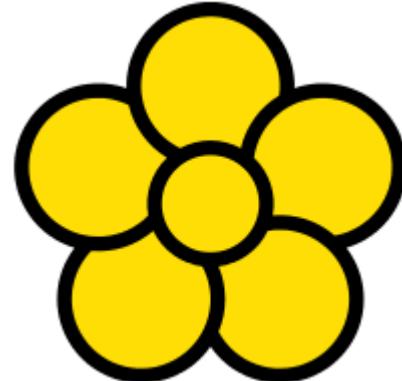
# Automatización de procesos



Apache  
Airflow



Pandas





# Automatización

# ¿Como se compila Python?

Python es un lenguaje interpretado es decir necesita un traductor intermedio para su ejecución.

Los lenguajes compilados son aquellos se ejecutan independientemente por ejemplo C , C++, Fortran.

Hoy en día, debido a la potencia de las maquinas, nos podemos dar el lujo de perder un poco de rendimiento por tener una mayor facilidad y rapidez para crear un código

# Ejemplo

CC++

```
#Declaramos los objetos int
```

```
Int a=7
```

```
int b=3
```

```
Int c
```

```
c=a+b
```

```
c=10
```

Python

```
a=7
```

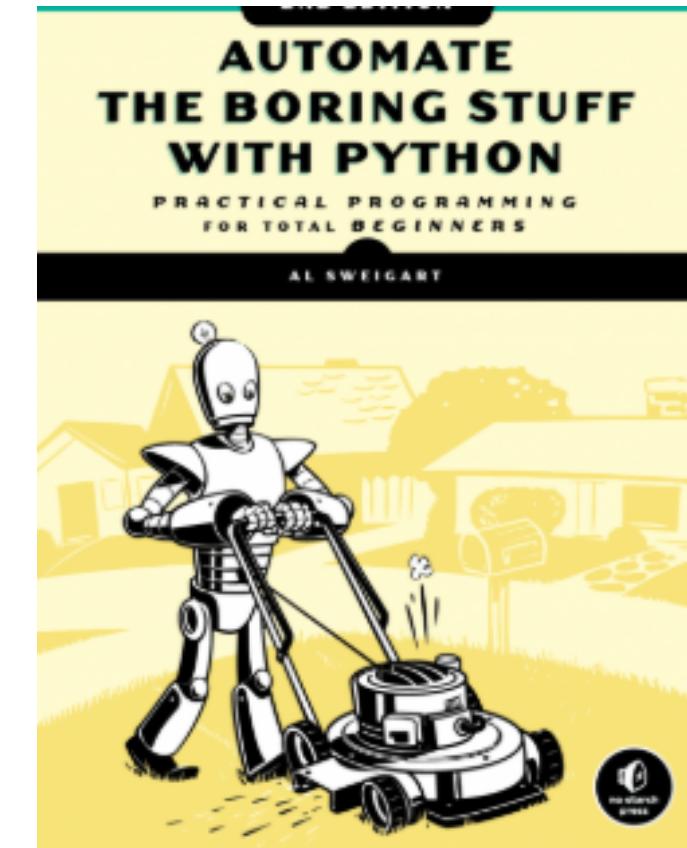
```
b=3
```

```
c=a+b
```

```
c=10
```

# ¿Qué podemos automatizar en Python?

Las cosas aburridas



- Descargar data de internet
- Ordenar archivos en carpetas
- Modelos de predicción
- Limpieza de base de datos
- Scrapear paginas web
- Reportes
- ETL de datos
- Y mas

# ¿Qué podemos empezar?

- 1) Identifiquemos una tarea que sea repetitiva
- 2) Validemos la oportunidad de automatizarla
- 3) Diseñemos el flujo de trabajo
- 4) Empecemos a programarla
- 5) Probemos su funcionalidad
- 6) Establezcamos alguna medida y/o acciones de mantenimiento del proceso de automatización

# Ejemplo

- 1) Identifiquemos una tarea que sea repetitiva:



*“Todos los días descargo información de muchas páginas web”*

- 2) Validemos la oportunidad de automatizarla :

*¿Consume mucho tiempo?*

*Dado los recursos actuales,  
¿se puede automatizar al  
100%?*

*¿Es muy costosa la  
automatización?*

*Si, me toma toda la mañana*

*Si, tenemos una pc y  
sabemos Python*

*No, en menos de una semana  
podemos terminar este  
proyecto*

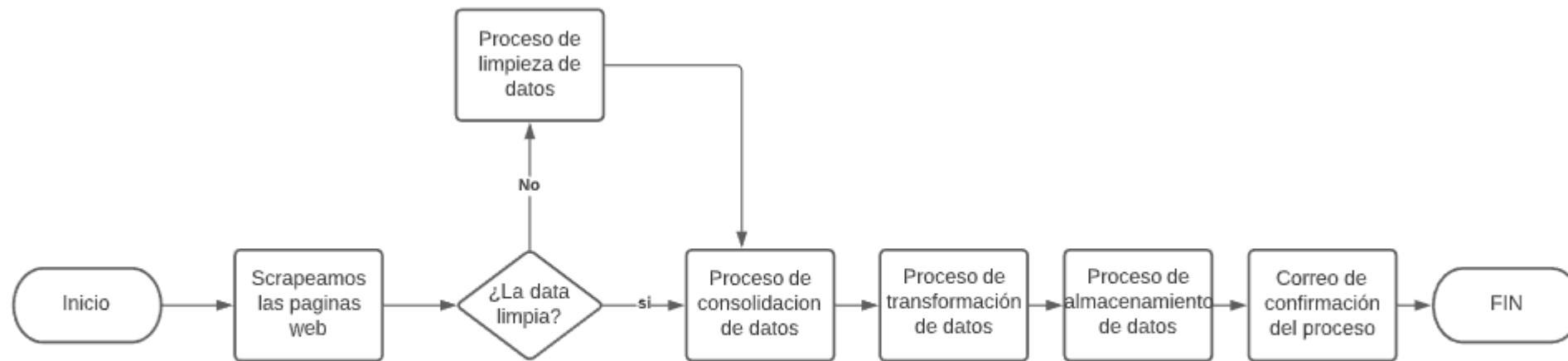
- 2) Validemos la oportunidad de automatizarla :

*Costo diario: 3 a 4 horas diarias aprox. 882 Horas al año*

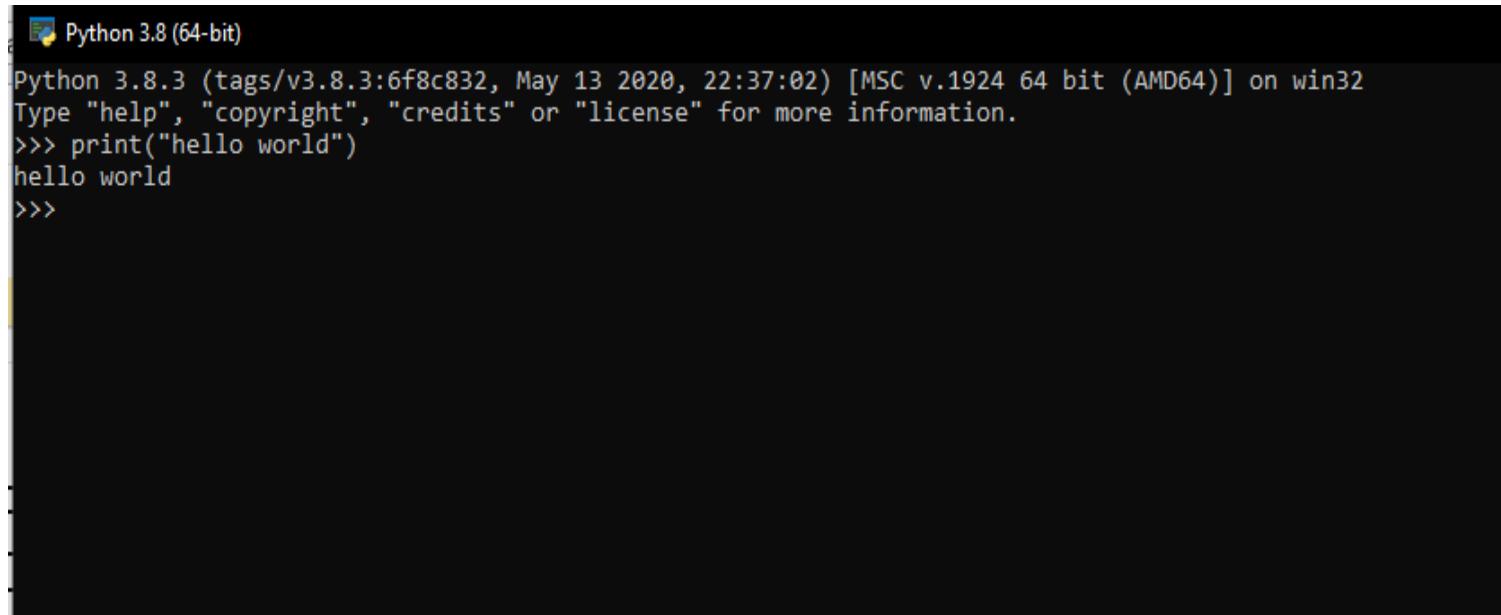
*Costo del proyecto: 3 a 4 horas diarias por una semana es 20 horas*

*La ganancia en este proyecto de automatización de muy alta*

- 3) Diseñemos el flujo de trabajo :



- 4) Empecemos a programarla:



```
Python 3.8 (64-bit)
Python 3.8.3 (tags/v3.8.3:6f8c832, May 13 2020, 22:37:02) [MSC v.1924 64 bit (AMD64)] on win32
Type "help", "copyright", "credits" or "license" for more information.
>>> print("hello world")
hello world
>>>
```



- 5) Probemos su funcionalidad :

¿Cumple con lo esperado?

¿Puede remplazar el trabajo manual sin alguna diferencia?

- 5) Establezcamos alguna medida y/o acciones de mantenimiento del proceso de automatización

\*Crear un campo con la ultima fecha de actualización correcta

\*Enviar un correo diario cuando el proceso termine y avisar si ocurrió un error.

\* Validar que la pagina scrapeada mantenga el formato en el tiempo.

# Proyecto

- Realizaremos un proceso de automatización end to end.

## Objetivo:

Lograr realizar un proceso de automatización de scraping que consolide los datos y almacene en la nube para la elaboración de un reporte web en tableau sobre el Covid 19.

# ¿Qué podemos empezar?

- 1) Identifiquemos una tarea que sea repetitiva

Entrar a la pagina

<https://github.com/CSSEGISandData/COVID-19>

Extraer los datos

Transformar los datos

Consolidar los datos

Pegarlos en un Gsheet (drive)

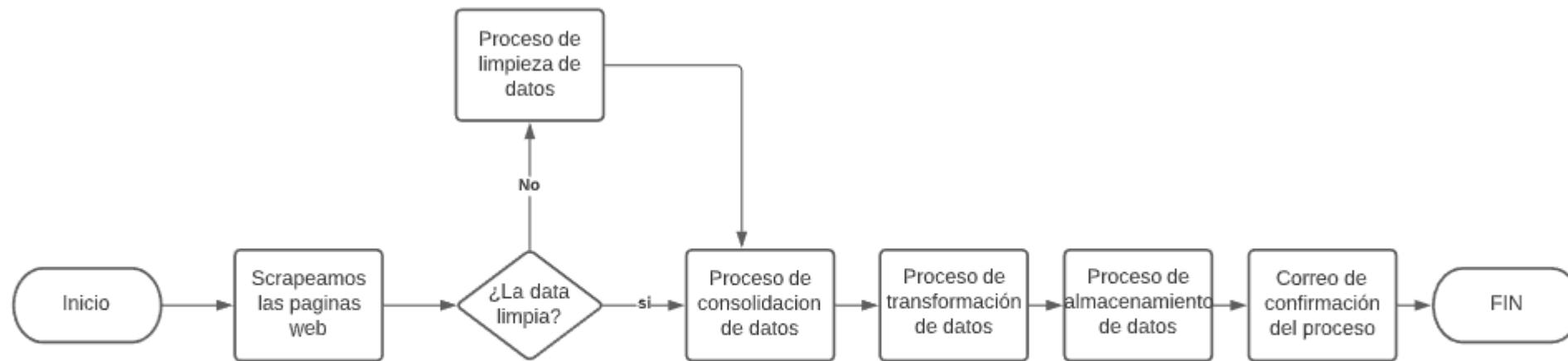
- 2) Validemos la oportunidad de automatizarla :

*Costo diario: 1 hora a mas diaria a partir de las 8 am*

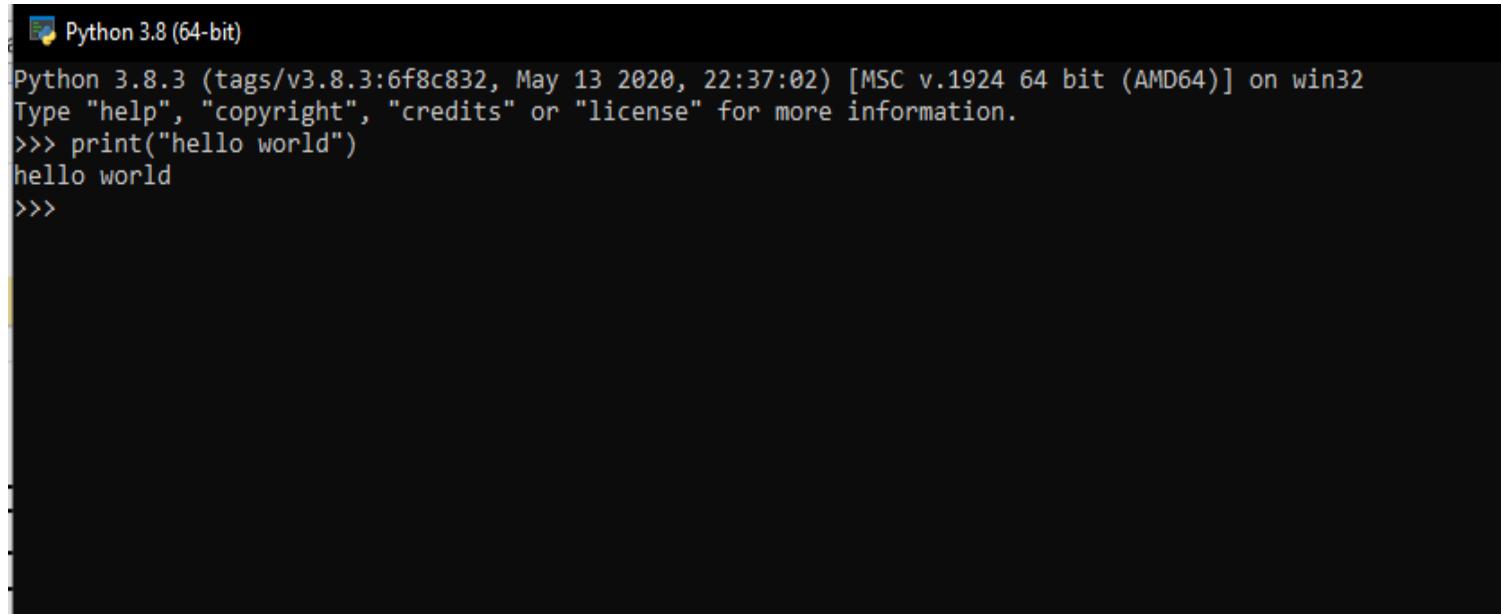
*Costo del proyecto: 1 dia*

*La ganancia en este proyecto de automatización de muy alta*

- 3) Diseñemos el flujo de trabajo :



- 4) Empecemos a programarla:



```
Python 3.8 (64-bit)
Python 3.8.3 (tags/v3.8.3:6f8c832, May 13 2020, 22:37:02) [MSC v.1924 64 bit (AMD64)] on win32
Type "help", "copyright", "credits" or "license" for more information.
>>> print("hello world")
hello world
>>>
```



- 5) Establezcamos alguna medida y/o acciones de mantenimiento del proceso de automatización

\*Crear un campo con la ultima fecha de actualización correcta

# Bibliografía

## 7 STEPS TO IMPLEMENTING AUTOMATION | TEAMHGS

(*7 Steps to Implementing Automation | TeamHGS, 2020*)

Teamhgs.com. 2020. *7 Steps To Implementing Automation | Teamhgs*. [online] Available at:  
<https://www.teamhgs.com/blog/7-steps-implementing-automation> [Accessed 10 October 2020].

## SMALLCOMBE, M.

*The Top 4 ETL Python Frameworks*

**En el texto:** (Smallcombe, 2020)

**Bibliografía:** Smallcombe, M., 2020. *The Top 4 ETL Python Frameworks*. [online] Xplenty. Available at:  
<https://www.xplenty.com/blog/top-etl-python-frameworks/> [Accessed 10 October 2020].

## SWEIGART, A.

*Automate the Boring Stuff with Python, 2nd Edition*

**En el texto:** (Sweigart, 2019)

**Bibliografía:** Sweigart, A., 2019. *Automate The Boring Stuff With Python, 2Nd Edition*. No Starch Press, Incorporated.



# Contacto

+51969559508

/luiggisilvaatto

Ejemplo@gmail.com

Luiggi Silva Atto

# Data Science

