

HENRY

Características de los datos

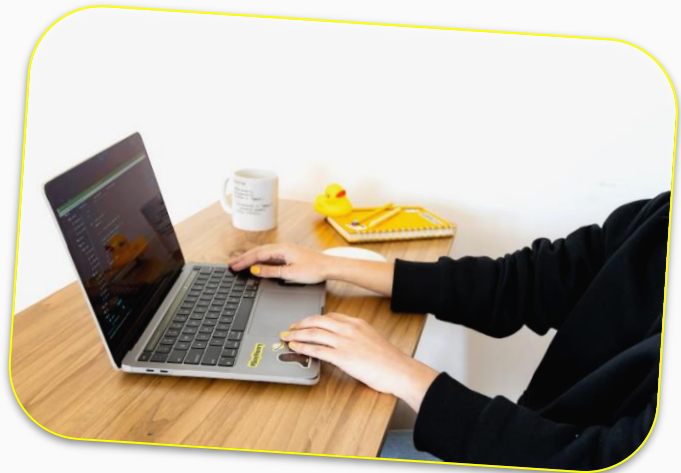


Data Science





Agenda



- Características de los datos
- Tipos de datos
- Falencias en los datos
- EDA (Análisis Exploratorio de Datos)
- Pipenv



OBJETIVOS DE LA CLASE

Al finalizar esta lecture estarás en la capacidad de...

- Conocer los distintos tipos de datos
- Identificar las falencias que se pueden encontrar en los datos
- Comprender de qué se trata el Análisis Exploratorio



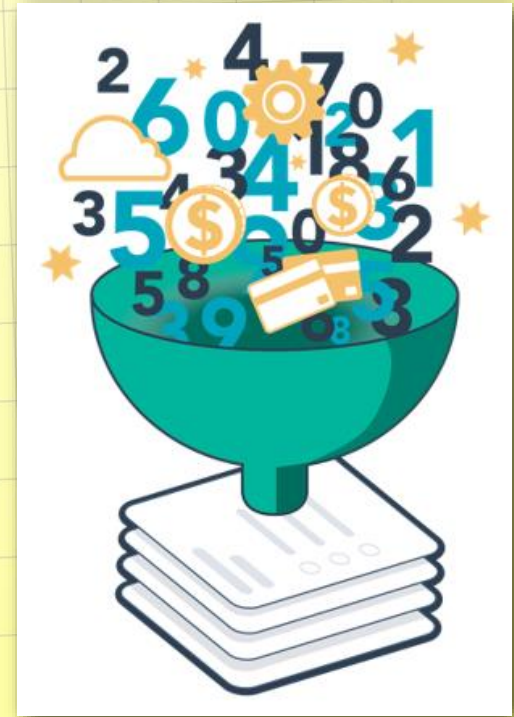
Características de los datos





¿Qué son?

Los **datos** con los que vamos a estar trabajando, son en definitiva, la **fuentes del conocimiento** necesario que debemos adquirir para poder **resolver** las preguntas que nos hacemos, entonces, es preciso conocer todas sus características.





Tipos de datos





Los datos no son más que **observaciones** del mundo en que vivimos, por tanto, los mismos pueden venir en diferentes formas.

Podemos incluir a los datos dentro de 2 categorías fundamentales:

**Datos
Cuantitativos**

Datos Cualitativos

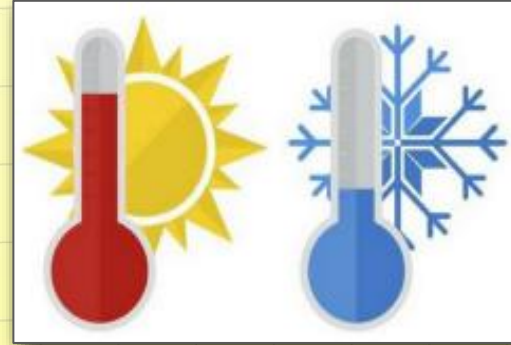


Datos Cuantitativos

Discretos



Continuos





Datos Cualitativos

Categoricos



Ordinales





Falencias en los **datos**





IMPORTANTE





Falencias en datos

Valores faltantes



Errores





Análisis Exploratorio



Preguntas orientadoras para analizar el dataset



- ¿Cuántos registros hay?
- ¿Son demasiado pocos?
- ¿Son muchos y no tenemos Capacidad (CPU+RAM) suficiente para procesarlo?
- ¿Están todas las filas completas o tenemos campos con valores nulos y por eso queda el resto de información inútil?
- ¿Qué datos son discretos y cuáles continuos?

Preguntas orientadoras para analizar el dataset



- ¿Cuáles parecen ser features importantes?
- ¿Cuáles podemos descartar?
- ¿Siguen alguna distribución?
- ¿Hay errores? ¿De qué tipo?
- ¿Podemos eliminarlos?
- ¿Es importante conservarlos? ¿son errores de carga o son reales?



Pipenv

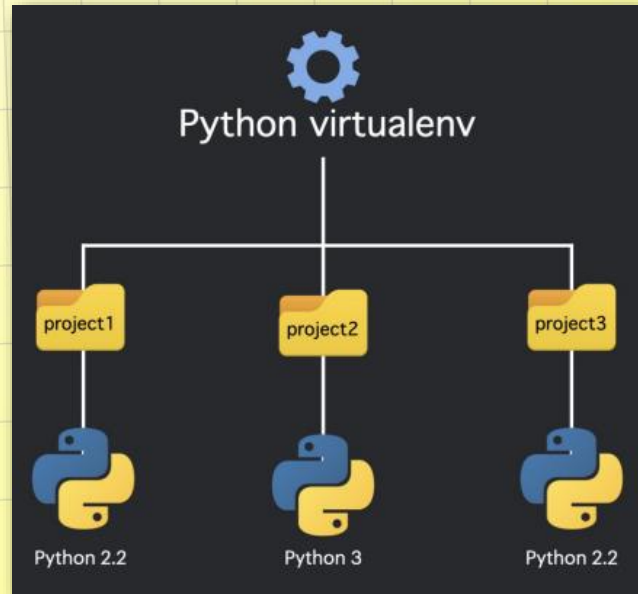




¿Qué es?

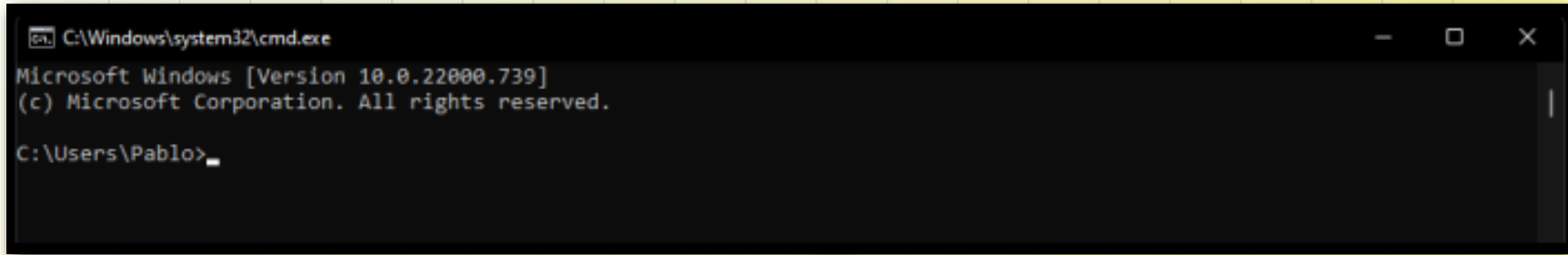
Pipenv es un **administrador de ambientes virtuales** para Python.

Un ambiente virtual es un directorio donde se almacenan librerías e información de Python, que es utilizado para tener diferentes versiones de módulos o Python mismo instalados.





Instalación



A screenshot of a Windows Command Prompt window. The title bar shows 'C:\Windows\system32\cmd.exe'. The window content displays the following text: 'Microsoft Windows [Version 10.0.22000.739]', '(c) Microsoft Corporation. All rights reserved.', and the command prompt 'C:\Users\Pablo>_'.

- Primero, abrimos la [consola/terminal/bash](#).
- En Windows podemos buscarla en el buscador de programas o utilizar la tecla de windows + R, donde escribiremos cmd y le daremos ejecutar.



Instalación

```
C:\Windows\system32\cmd.exe
C:\Users\Pablo>pip install pipenv
```

- Escribimos la instrucción `pip install pipenv` en la consola, lo que hace que se instale pipenv y las dependencias necesarias en el directorio donde tenemos Python.
- IMPORTANTE: Recuerden que deben tener Python instalado y agregado al PATH. Esto se hace descargando el instalador de python.org y siguiendo las instrucciones (recuerden chequear la opción Agregar Python al PATH).



Instalación

```
C:\Windows\system32\cmd.exe
C:\Users\Pablo>pipenv shell
```

- Una vez instalado pipenv, nos dirigimos al directorio donde queremos crear el ambiente e utilizamos el comando **pipenv shell**.
- Este comando inicializa el ambiente contenido en ese directorio y, si no existe ninguno, lo crea.



Comandos útiles

- `pipenv -h`
- `pipenv install <module>`
- `pipenv lock`
- `pipenv install --ignore-pipfile`
- `pipenv graph`
- `pipenv uninstall <module>`
- `pipenv uninstall --all`

HENRY

