

# Herramientas para cada etapa del análisis de datos

---

Entender cuál es la función de los programas más usados en cada etapa del estudio de la información.

## Etapa: EXTRACCIÓN DE INFORMACIÓN CON SQL



Extracción de la información.



Síntesis de la base de datos



Cuadros de control de la operación



# SQL

Esta etapa y estas herramientas son principalmente utilizadas por los **analistas** y los **ingenieros**, estas son las personas que crean bases de datos y las sintetizan.

El lenguaje que van a utilizar es SQL. SQL Nos ayuda a extraer información.

## Etapa: ANÁLISIS Y VISUALIZACIÓN CON R Y PYTHON

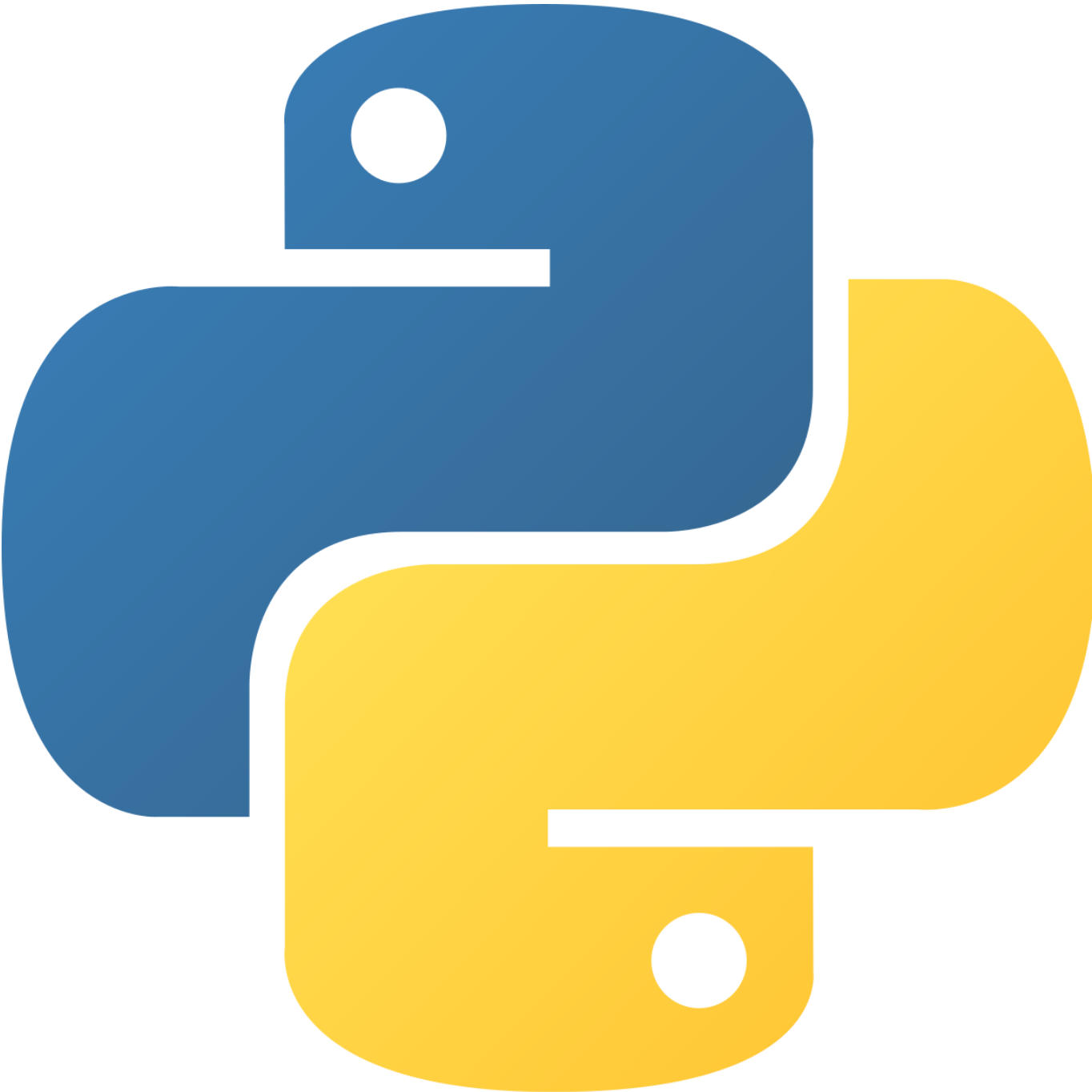
**0101****1001** Análisis de datos con enfoque estadístico

Más complejo al inicio



Análisis descriptivo y exploratorio

📁 Packages: ggplot2, dplyr





Análisis de datos con enfoque en ingeniería



Parecido a otros lenguajes



Análisis descriptivo y exploratorio



Librerías: Panda, Numpy

Estas herramientas son las que utiliza el **científico de datos**, estas herramientas son exploratorias son predictivas y se basan en modelos estadísticos, modelos matemáticos y también nos van a servir para hacer gráficos y visualización.

En realidad, puedes elegir el que más te guste, si tu rol deseado es el de científico de datos.

Ambos lenguajes son open-source por lo que podemos descargarlos de manera gratuita.

R tiene un enfoque más estadístico, nos ayuda a sacar ¿P Values?, Intervalos de confianza, es un enfoque más basado en la econometría. Y en cambio Python tiene un enfoque más ingenieril, como Java, C++ ese enfoque de ingeniería.

R Puede parecernos un tanto complejo al inicio, pero como todo, con práctica y al paso del tiempo nos familiarizamos y se nos va facilitar el trabajo. Python es más parecido a Java y a otros lenguajes, por lo que si ya sabemos programar es una mejor opción irnos por Python.

R vamos hacer un análisis descriptivo y exploratorio, en Python haremos lo mismo. ¿Qué es un Análisis Descriptivo? Como dice la palabra, nos ayuda a DESCRIBIR la información que ya tenemos, interpretarla, para ver cuánta información hay, y el EXPLORATORIO nos va servir para pasar a otra fase, para poder predecir. Ya sé que tengo esta información, pero quiero saber la información que voy a tener en un mes.

Ambos lenguajes tienen apoyos adicionales, tenemos packages R y Librerías Python.

R ggplot2 nos sirve para hacer gráficos increíbles, colors, funcionalidades.

R dplyr para reorganizar nuestra información.

Python Pandas y Numpy, los dos nos sirven para reorganizar la información, para ser más eficientes y no tener que escribir tanto código

## RETO EJERCICIO:

---

### Herramientas para cada etapa del análisis de datos

Acaban de informarte que las ventas de tu empresa cayeron mucho hoy. Debes extraer la información de los clientes que compraron ayer y los clientes que compraron hoy para entender por qué cayeron las ventas. ¿Qué herramienta debes usar para realizar la extracción de datos?

**Sería requerido que la empresa ya tuviera una base de datos estructurada, SQL Para realizar las consultas y extraer la información. Python para realizar un análisis descriptivo. Podríamos consultar perfil de clientes, comparar datos con otros años, etc.**

¿Qué clientes compraron ayer y hoy? Sin duda que SQL. Estamos con la mirada en el pasado y el presente, no nos están pidiendo que predigamos el futuro. Con SQL podremos saber quienes compraron ayer y hoy.

Reto de la clase: Extracción: SQL para obtener los datos(dependiendo de la DB, algún manejador que se adapte), extrayendo los datos de los clientes y las transacciones realizadas. Análisis: Usar una herramienta(Power BI/Tableau) para construir un reporte y tener de manera más visual los datos extraídos

