Apuntes Modulo1

No es lo mismo información que Datos

Librerias de Python para analizar datos

3 grupos

Computo científico

Pandas (estructuras de datos)

NumPy (arreglos y matrices)

SciPy (Operaciones complejas, integrales etcs)

Visualizacion

Matplotlib (mas popular)

Seaborn (heatmaps, violín, tiempo)

Algoritmicas

Scikit-learn (Machine Learning)

Statsmodels Explorar datos

Importacion

Formatos, csv, json, xlsx..

Path

Dataframes

Head (n) muestra un numero de listas n

Tail(n) muestra las ultimas líneas

Se puede poner nombre a los headers, usando columns=[array of headers]

.describe da un resulmen con counts, avg, std, max y min y los cuartiles

.describe(include=”all”) para mayor detalles

.info() da los top 30 y last 30

.dtypes sirve para traes los tipos de datos