Module 1 - Python P. 2

Scientifics Computeng Libraries in Python

1. Scientifics Computing Libraries
   - Pandas (Data structures & tools)
   - Numpy (Arrays & matrices)
   - SciPy (Integrals, solving differentials equations, optemization)

2. Visualization Libraries
   - Matplotlib (plots & grahps, most popular)
   - Seaborn

3. Algorithmic Libraries
   - Scikit-learn (Machene Learning : regressio classification)
   - Statsmodels (Explore data, estimates statistical models, and perform statistical tests).

Importing Data

2 important properties:

✓ Format
- Various formats: csv, Json, xlsx, .hdf..

✓ File Path of dataset
- Computer
- Internet

Printing the dataframe

df prints the entire dataset
df.head (n) to show the first n rows of data frame
df. tail (n) shows the bottom n rows of dataframe

adding headers

Replace default header (by df. columns = headers)

Exporting a Pandas dataframe to CSV

path = "C:\Windows\...\automobile.csv"
df.to_csv(path)

DATA FORMAT	Read	Save
CSV	pd.read_csv ()	df.to_csv
Json	pd.read_json ()	df.to_csv
excel	pd.read_excel()	df.to_csv
Sql	pd.read_sql ()	df.to_csv

Basic insights to data set - Data types

Panda Types	Native Python
Object	string
int64	int
float64	float
datetime (4, timedelta(ns)	NA

Why check?

- Potential info and type mismatch
- Compatibility with python methods

ai. numerical data - Math functions

In pandas, we use dataframe.dtypes to check data types

df.dtypes

- Returns a statistical summary
- full summary statistics

df. describe()

Count
Mean - Promedio
std - Standar devation
min
25% }
50% } limite de cada cuartil
75% }
max

df. describe(

→ Los mismos más

Unique
top
freq

df. info shows the top 30 rows and bottom 30 rows of a dataframe