ECS 116
Databases for Non-Majors

Discussion 4
(4/25/24) |Spring ‘24

Today's Agenda

1. Quick Updates
2. Using Jupyter notebook to manipulate data using Py Pandas
3. Understanding what the script means

Basic Operators, Operations, Visualization, iloc

Quick Updates

Programming Assignment 1

- Last date is April 28th @ 11:59 pm
- Updated doc online

Creating Groups

- Assignment 2 onwards will be grouped (3 ppl)
- Go to the excel sheet and write names

Problem Set 2 Extension

- Folks who just joined
- Last date mostly sunday evening

Files

Today's Files:

1. Boilerplate: *Files->Juypter Notebooks -> DISC-4-BOILERPLATE-v01.ipynb*
2. Completed notebook: *Files->Juypter Notebooks -> DISC-4-Pandas-v01.ipynb*

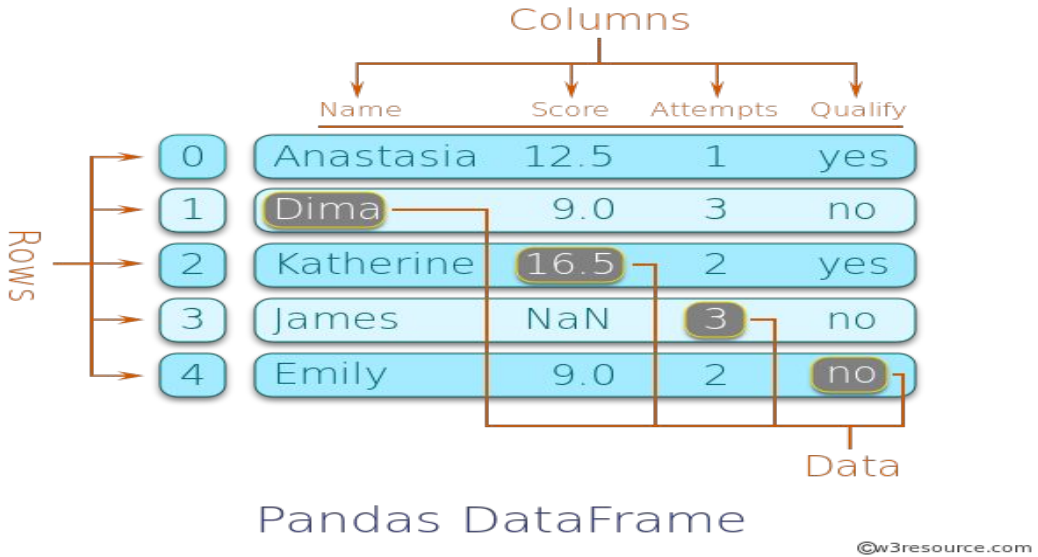
Tech

What is Python Pandas?

Pandas is a fast, powerful, flexible and easy to use open source data analysis and manipulation tool, built on top of the Python programming language.

Btw: If you haven't installed Anaconda use the installation doc to do so.

What are dataframes?



Two-dimensional, size-mutable, potentially heterogeneous tabular data.

Data structure also contains labeled axes (rows and columns). Arithmetic operations align on both row and column labels. Can be thought of as a dict-like container for Series objects. The primary pandas data structure

How does it all work?

1. Open Jupyter Notebook

2. Import required packages

3. Import the data

4. Manipulate data

5. Visualize data

6. Export file

Working on Jupyter Notebook

1. Download the BOILER PLATE code
2. Open Jupyter Notebook through the Anaconda navigator
3. In JN, open the downloaded boiler plate code
4. Go through the boiler plate code
5. Work on the data using Pandas

What did we learn today?

1. Quick Course Updates
2. Tech Used
3. Using pandas in jupyter notebook
4. Visualization using the data
5. Exporting the file

Thank You!

See you next Thursday!