Sierra Bettis
HAS Tools
October 29, 2021

Cheat Sheet 4 – Pandas

Pandas:

- Pandas is a dataframe and allows you to work with tabular data → matrix of values
    - Allows you to have as many rows and columns as you would like, you can store and manipulate tons of data
    - Can also be 1D series, which is a list of numbers
- Need to import the dataframe → *import pandas as pd*
- This is different from other object types previously learned because now the data is organized by rows and columns
    - Makes data easier to work with so that you can reference data by its respective row or column, or both at the same time
    - Makes is versatile for storing data
    - You are able to name the index row whatever you want to make it convenient for the user
- You can also manipulate the data with math functions → add, subtract, and multiply different rows and columns together

Importing data to the Pandas dataframe:

- Import data by creating a name and making it equal to a URL → then you can read that file as a pandas dataframe and it should read the csv as a dataframe with the indexing, rows, and columns set up

Applications of Pandas:

- You can use the data to make graphs
- How to index:
    - Location based by using iloc → *dataframe.iloc[row, column]*
        - Same as index location of [start:stop:step]
    - Label based by grabbing rows by name using loc → *dataframe.loc['row(s)']*
    - Grabbing out an individual column name → *dataframe['column(s)']*
    - You can reference rows and columns the same way as indexing with lists, but now you can utilize row and column names which can be easier, so you don't confuse indexing

Methods of Pandas:

- Use of head and tail
    - .head = returns the 1st value of the dataframe
    - .tail = returns the last value of the dataframe
    - Ex: *dataframe['row'].tail(3)* → will return the last 3 values from the dataframe in 'row'

    - This is where you can really manipulate the rows and columns to perform calculations with them if necessary
- Sort
    - You can use sort_values to sort the dataframe by ascending of descending
- Describe → summary of stats (useful when you want to get out the mean, average, max, min, etc.)
- Group by → group according to column values (useful for when you just want to pull out certain columns to analyze)