**February 7,2021**

**TakenMind Global Internship**

Assignment 1

**NumPy (Numerical python)**

NumPy is a Python library used for working with arrays, and has functions for working in domain of linear algebra, Fourier transform, and matrices.

In Python we have lists that serve the purpose of arrays, but they are slow to process. NumPy provides an array object **(ndarray)** faster than normal Python lists. it provides a lot of supporting functions that make working with very easy.

Arrays are very frequently used in data science, where speed and resources are very important.

NumPy installation using cmd: **pip install numpy**

Array can be of any dimensions (1D,2D and Higher dimensions)

**1D Array:** Example

### import numpy as np  
  
 arr1 = np. array ([1, 2, 3, 4, 5])  
  
 print(arr1)

**2D Array:**

### import numpy as np  
  
arr2 = np. array ([[1, 2, 3], [4, 5, 6]])  
  
print(arr2)

Shape: (arr2. shape) gives the dimensions of array

### (2,3)

* Zero array is array with all zeros (np. zeroes (5)) gives array of 5 zeroes.

[0,0,0,0,0]

* Ones array is array with all ones (np. ones (5)) gives array of 5 ones.

[1,1,1,1,1]

* Eye function creates an identity matrix (diagonal elements are one)
* arange function: np.arange(start:stop:step)

np. arange (2,10,2)

##[2,4,6,8]

**Saving and loading arrays:**

Arrays can be stored and can be loaded again

##np.save(“variable name”, array\_name)

Loading\_array:np.load(“variable name.npy”)

Print (Loading\_array)

**Delimiter:** Each value in the array is separated by delimiter

For Multiple Arrays

##np.savez(“saving\_array.npz”, x=array1,y=array2)

load\_npz=np. load (saving array.npz)

Print(load\_npz[‘x’]

Print(load\_npz[‘y’]

**Searching Arrays**

You can search an array for a certain value, and return the indexes that get a match.

To search an array, use the **where()** method.

##import numpy as np  
  
 arr = np.array([1, 2, 3, 4, 5, 4, 4])  
  
 x = np.where(arr == 4)  
  
 print(x)

**Sort():**Sorting means putting elements in an ordered sequence. The NumPy ndarray object has a function called sort(), that will sort a specified array.

**ufunc():**ufuncs stands for "Universal Functions" and they are NumPy functions that operates on the ndarray object.

**Stastical Processing and Sketching in Python:**

It is done by new library called matplotlib,

Syntax: import matplotlib.pyplot as plt.

Linear function heat map and Cos function heat map

Example : z = np.cos[x] + np.cos[y]

Imshow is used to print the map

Savefig can be used to save the figure to computer memory

**Pandas:**

Pandas is a Python library used for working with data sets.

It has functions for analyzing, cleaning, exploring, and manipulating data.

The name "Pandas" has a reference to both "Panel Data", and "Python Data

Analysis" and was created by Wes McKinney in 2008.

Installation : **pip install pandas**

* Pandas allows us to analyse big data and make conclusions based on statistical theories.
* Pandas can clean messy data sets, and make them readable and relevant.
* Relevant data is very important in data science.

**Pandas Series :**

A Pandas Series is like a column in a table.

##import pandas as pd  
  
 a = [1, 7, 2]  
  
 myvar = pd.Series(a)  
  
 print(myvar)

**Key/Value Objects as Series**

You can also use a key/value object, like a dictionary, when creating a Series.

## import pandas as pd  
  
calories = {"day1": 420, "day2": 380, "day3": 390}  
  
myvar = pd.Series(calories)  
  
print(myvar)

**Pandas DataFrames**

A Pandas DataFrame is a 2 dimensional data structure, like a 2 dimensional array, or a table with rows and columns.

## import pandas as pd  
  
 data = {  
    "calories": [560, 315, 640],  
   "duration": [60, 50, 55]  
 }  
  
 #load data into a DataFrame object:  
 df = pd.DataFrame(data)  
  
 print(df)

**Locate Row:** print(df.loc[0])

**Pandas Read CSV:**

A simple way to store big data sets is to use CSV files (comma separated files).

CSV files contains plain text and is a well know format that can be read by everyone including Pandas.

## import pandas as pd  
  
 df = pd.read\_csv('data.csv')  
  
 print(df.to\_string())

**Read JSON:**

Big data sets are often stored, or extracted as JSON.

JSON is plain text, but has the format of an object, and is well known in the world of programming, including Pandas.

## import pandas as pd  
  
 df = pd.read\_json('data.json')  
  
 print(df.to\_string())

**Viewing the Data:**

Mostly we use head(x) and x can be any number and the above values of data gets printed and tail() gets the last values of the data

The DataFrames object has a method called info (), that gives you more information about the data set.

Print(df.info())

**Pandas - Cleaning Data:**

Data cleaning means fixing bad data in your data set.

Bad Data could be:

* Empty cells
* Data in wrong format
* Wrong data
* Duplicates

**Pandas - Data Correlation**

**Finding Relationships**

A great aspect of the Pandas module is the corr() method.

The corr() method calculates the relationship between each column in your data set.

## df.corr()

**Pandas Plotting**

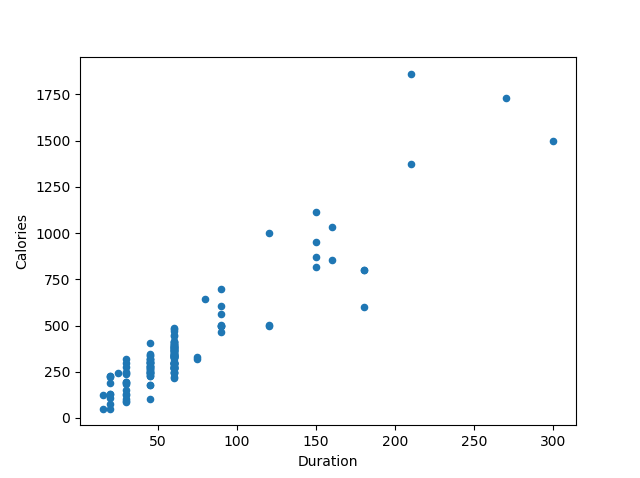
Pandas uses the plot() method to create diagrams.

Pythons uses Pyplot, a submodule of the Matplotlib library to visualize the diagram on the screen.

## import pandas as pd  
 import matplotlib.pyplot as plt  
  
 df = pd.read\_csv('data.csv')  
  
 df.plot()  
  
 plt.show()

**Scatter plot:**

**#** import pandas as pd  
 import matplotlib.pyplot as plt  
  
 df = pd.read\_csv('data.csv')  
  
 df.plot(kind = 'scatter', x = 'Duration', y = 'Calories')  
  
 plt.show()



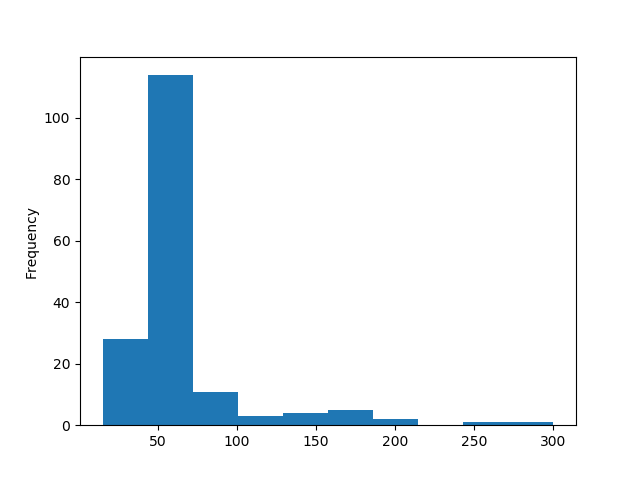
**Histogram:**

**#** Use the kind argument to specify that you want a histogram:

kind = 'hist'

A histogram needs only one column.

A histogram shows us the frequency of each interval, e.g. how many workouts lasted between 50 and 60 minutes?



Thank YOU !!!!