Lab 6

How to import/access an existing data set into colab:

So far we have been using python's numpy to generate our data set (though lab 4). In Lab 5 we used a data set that was supplied.

In real life data science tasks, most data sets will be in the form of some existing data sets in some format, e.g., excel, or csv format.

For that reason, our next learning item is to be able to read/import an existing data set into our data analysis tool (in our current case, google colab)

 For this exercise we will import (upload) into colab a data set from a local drive (We can later learn how to accomplish this from other sources e.g., from google drive.

For that reason, I have added a csv dataset (testDataset.csv) for this assignment (under our canvas Data science Infusion folder: Labs/Lab 6 data set).

You will place this file somewhere in your local drive to use in the given instructions.

Reading csv data set into google colab

** **********************************

Reading data set from a local drive

```
from google.colab import files
uploaded = files.upload()
```

' You will have to 'Run' this to see next instruction to choose file

'Follow the instructions to locate your file location

'After choosing the file:

```
import pandas as pd
import io


df = pd.read_csv(io.BytesIO(uploaded['file.csv']))
```
'file.csv' will be **'replaced by the name of your file e.g., myFile.csv**

'Run this to see your read data set file

```
print(df)
```

'Explore your data set:

```
data.head()
```
      'prints 1st 5 rows…………………Run to see

```
data.shape()
```
      ' #Returns the size of the data set (i.e. #rows X #numcolumns)

Note: As said in class, the apostrophe (') in front of a sentence indicates a comment

****