

Assignment 3 (Beginner)

May 23, 2020

Feel free to refer to the class Jupyter Notebook, we do not expect you to know the syntax by memory.

We only expect you to understand how the libraries work and which function to use where. It is alright if you copy paste from the class Jupyter notebook. Almost everybody copy pastes in programming. The difference is some know what to copy paste and others don't.

```
[ ]: import pandas as pd
import matplotlib.pyplot as plt
import numpy as np
```

0.1 Matplotlib

```
[ ]: %matplotlib inline
```

```
[ ]: # Plot the following array and label both the axis. Also give the plot a title
# x = np.linspace(0, 100, 20)
```

```
[ ]: # Plot  $y = x^5$  and use:
# 1. A black line
# 2. Red circles
# 3. Green squares
# to represent the graph
```

```
[ ]: # Plot  $y = x^2$  and  $y = x^3$  in two different subplots (of shape (2,1)) in a
→figure
# x = np.arange(0., 10., 0.2)
```

```
[ ]: # Redo the above plot, but limit the axis to [5,6] in the x-axis and [20,100]
→in the y-axis
```

```
[ ]: # In three subplots (of shape (3,1)), plot a bar plot, a scatter plot and a
→line plot
```

```
[ ]: # Plot three different figures of varying sizes (take any size you want)
```

```
[ ]: # Plot the following images in a loop.
# If you are doing this on Kaggle, please enable internet in settings
# (see telegram group for instructions)
# https://i.ibb.co/TwsG391/Screenshot-from-2020-05-21-19-17-26.png
# https://i.ibb.co/bR4XSQ8/Screenshot-from-2020-05-23-00-03-32.png
# https://i.ibb.co/SQbPCDH/Screenshot-from-2020-05-22-17-30-38.png
# https://i.ibb.co/JntTQ9v/Screenshot-from-2020-05-22-18-35-22.png
# Hint, the below code allows you to open images from URL

from PIL import Image
import requests
from io import BytesIO

url = "https://i.ibb.co/TwsG391/Screenshot-from-2020-05-21-19-17-26.png"

response = requests.get(url)
img = Image.open(BytesIO(response.content))
```

0.2 Pandas

```
[ ]: # Create a new series from a numpy array
```

```
[ ]: # Perform the following computations on the series:
# Add, subtract, multiply (with a constant), divide, exponent

s = pd.Series(np.random.randint(5, size=10))
```

```
[ ]: # this is how you import a csv file

df = pd.read_csv("https://raw.githubusercontent.com/cognibot-official/data/
↳master/occupation.csv",
                 index_col = 'user_id')
```

```
[ ]: # Display the first 10 rows and last 10 rows of the above dataframe df
```

```
[ ]: # Extract only the occupation column from the dataframe

df.head()
```

```
[ ]: # Display the first 4 rows and 3 columns
```

```
[ ]: # Display till the row index is 6 and column name is gender
```

```
[ ]: # Display all columns of only those whose occupation is writer
```

```
[ ]: # Display the summary statistics of the dataframe
```

```
[ ]: # Calculate the mean age in the dataset
```

```
[ ]: # Add a new column to the dataframe based on the following rules:  
# If the person is less than 30 years and is an engineer/programmer/scientist  
# label them as rich, label the others as poor :)
```

```
[ ]: # Find the mean (average) age of a :  
# Student  
# Lawyer  
# Engineer
```

```
[ ]: # Bonus: Find out which field has the youngest people  
# and which field has more men compared to women and vice versa
```