Intake2 Pandas LAB#2

Question#1: (12.5 marks)

Kindly handle the premier league top scorer's information as the following:

- Use Pandas to read the Top_Scorers CSV dataset which contains the following information about top scorers: players names, rank, number of scored goals, and their nationalities
- Use Pandas to read the players_stats JSON dataset which contains the following information about top scorers: players names, height, appearances, wins, losses, goals per match ratio, assists, yellow cards, red cards
- Merge the two datasets
- Visualize a scatter plot about the relation between number of matched played (Appearances) and the number of scored goals

Question#2: (7.5 marks)

Use Pandas to read the epldata_final dataset and calculate per club the average age of players in addition to the average of their market value, then write the calculations to a new JSON file.

Author of Questions: Mahmoud Magdy

Question#3: (30 marks)

Kindly prepare the “Telco-customer-churn” dataset as the following:

- Use Pandas to read the WA_Fn-UseC_-Telco-Customer-Churn CSV dataset
- Kindly set index of the Pandas dataframe by value of CustomerID
- Kindly remove duplicates from the dataframe
- You may need to convert type of "TotalCharges" column from object to numeric, one of the possible suggestions is to do as the following:
  df['TotalCharges'] = pd.to_numeric(df['TotalCharges'], errors='coerce')
  - Take care that the above type conversion will leave some missing values for non-numeric values
  - Why we did this?!, you can try one-hot encoding to know 😊
- Kindly replace the missing values in "TotalCharges" by the average of this column
- Kindly do one-hot encoding for all the categorical columns, except "Contract" and "Churn" columns
- Kindly analyze "Contract" column, and do ordinal encoding for its values as the following:
  - Replace "Month-to-month" by 1
  - Replace "One year" by 2
  - Replace "Two year" by 3
- Kindly convert type of "Contract" column to numeric as we did before, and please don't forget to make sure that there's no missing values left in this column after conversion
- Kindly do label encoding for "Churn" column:
  - Replace "No" by 0
  - Replace "Yes" by 1
- Kindly convert type of "Churn" column to numeric as we did before, and please don't forget to make sure that there's no missing values left in this column after conversion
- Create a separate dataframe contains the “Churn” values, to act as Targets/Labels while training the machine learning model
  - Take care this new dataframe must has "CustomerID" as it's index also
  - kindly remove "Churn" from the original dataframe
- Kindly do normalization for all numerical columns
- Kindly save the two prepared dataframes as CSV files