Homework #2

Task: In this homework you will access two files from my website. These files contain information on video game players and stats about these players sessions. You will do some analysis on these files and create various outputs. For submission, you only need to submit your .py file. **NOTE, there are questions that need to be answered. I ask that these be answered at the end of your python file using comments.**

Submission Requirements:

- Python file used to perform the requested task.
- File locations:
  - File 1: http://drd.ba.ttu.edu/isqs6339/hw/hw2/players.csv
  - File 2: http://drd.ba.ttu.edu/isqs6339/hw/hw2/player_sessions.csv
- *Note* your python file should load your pandas data frames directly from the website using techniques we have discussed in class.

Grading:

- You will be graded based upon:
  - Quality of your code
    - Note, I am not looking for the most efficient code. I am looking for code that is well documented (i.e. commented) and follows a logical progression. Your goal is to write code that another developer could pick up and know what you are doing.
  - Adhering to best practices listed in the lectures
    - Example: Variables that I can change to run your code in my environment. I should not have to look through your code for items to change.
  - Correctly generating outputs as requested below.

Task:

- You will need to read both files from the web locations into pandas data frames and merge based on playerId.
- Clean the data
  - Note, while this does not represent all issues, some "clan" data is missing. If player's name starts with the letter "V", then they should be in the guild "LoD".
- Encode a new column (called player_performance_metric) using the following function: ((damage_done * 3.125) + (healing_done * 4.815)) / 4
- Encode a new column (called dps_quality) based on the following criteria:
  - damage_done > 600000 -> "High"
  - damage_done between 400000 and 599999 -> "Medium"
  - damage below 400000 -> "Low"
- Encode a new column (called player_dkp_gen_rate) using the following criteria:
  - If dps_quality is "High": player_performance_metric * 1.25
  - if dps_quality is "Medium": player_performance_metric * 1.15
  - if dps_quality is "Low" and clan is not "LoD": player_performance_metric * .85
  - if dps_quality is "Low" and clan is "LoD": player_performance_metric * 2.35
- Produce the following csv outputs (Note, you should include group by fields when reporting):

- Average damage and healing per session by clan
- Average damage and healing by position
- Output of all merged data sorted by “player_dkp_gen_rate”

Questions:
These are the questions you should answer at the bottom of your .py file. (When possible Justify your answer with metrics from your data)

- What is the quality of your data? i.e. how clean is your data?
- What steps did you take to clean your data and why did you choose those options?
- Are there other potential ways you could have cleaned your data?
- Which player has the highest values for:
  - damage_done
  - healing_done
  - player_performance_metric
  - player_dkp_gen_rate