For this assessment, we'd like to see how you approach discovery, problem solving, and communication for a gaming data science project that involves *LeagueIndex* as the dependent variable.

Before You Start:

- Data Dictionary on page 2
- Please don't spend too much time on this assessment - it should take a reasonable amount of time that should not detract you from school+personal responsibilities given a week to complete.

Provided Material(s):

- csv file of player game data: *starcraft_player_data.csv*

Submission Material(s):

- Python script
- PDF of outputs and interpretation

OR

- If you choose to do this project in a notebook environment (ex: Jupyter notebook), you may do your interpretations/documentation using Markdown cells and submit the notebook as an exported PDF.

Assignment:
Attempt to solve the following problems with the provided dataset using Python (no restrictions on package usage):

1. Determine if this dataset needs any preprocessing. If so, clean the dataset and document your steps. If not, explain how you came to that conclusion.
2. Multicollinearity has a negative impact on many popular ML models. Check if this dataset experiences any multicollinearity. If so, reduce the impact until an acceptable point.
3. Determine what are the most important features that could help predict a player's rank? Interpret your results for a general audience (coaching staff, pro players, etc).
4. Your team's Starcraft2 coaching staff loved your project! They think this is perfect for scouting rising stars. Using your discoveries from (3), create a function to find players who should be given a chance to become professionals. Explain why your set of players make sense.
    a. **Hint:** Don't go overboard with complexity. More often than not, statistical reasoning is more efficient than taking an ML approach.
5. Hypothetically, if you were to move forward with creating a fully-fledged model to predict *LeagueIndex*, what model(s) would you consider and why? (Don't actually implement anything!)

Data Dictionary:

Feature	**Description**	**Datatype**
GameID	Unique ID number for each game	integer
LeagueIndex	Bronze, Silver, Gold, Platinum, Diamond, Master, GrandMaster, and Professional leagues coded 1-8	ordinal
Age	Age of each player	integer
HoursPerWeek	Reported hours spent playing per week	integer
TotalHours	Reported total hours spent playing	integer
APM	Action per minute	continuous
SelectByHotkeys	Number of unit or building selections made using hotkeys per timestamp	continuous
AssignToHotkeys	Number of units or buildings assigned to hotkeys per timestamp	continuous
UniqueHotkeys	Number of unique hotkeys used per timestamp	continuous
MinimapAttacks	Number of attack actions on minimap per timestamp	continuous
MinimapRightClicks	number of right-clicks on minimap per timestamp	continuous
NumberOfPACs	Number of PACs per timestamp	continuous
GapBetweenPACs	Mean duration in milliseconds between PACs	continuous
ActionLatency	Mean latency from the onset of a PACs to their first action in milliseconds	continuous
ActionsInPAC	Mean number of actions within each PAC	continuous
TotalMapExplored	The number of 24x24 game coordinate grids viewed by the player per timestamp	continuous
WorkersMade	Number of SCVs, drones, and probes trained per timestamp	continuous
UniqueUnitsMade	Unique unites made per timestamp	continuous
ComplexUnitsMade	Number of ghosts, infestors, and high templars trained per timestamp	continuous
ComplexAbilitiesUsed	Abilities requiring specific targeting instructions used per timestamp	continuous