Task:
Once you’ve settled on your model, communicate your findings to non-technical stakeholders.

Hypothetical: after seeing your work, your stakeholders come to you and say that they can collect more data, but want your guidance before starting. How would you advise them based on your EDA and model results?

Answer:
EDA (Exploratory Data Analysis):

The analysis of the StarCraft player data reveals interesting insights about the distribution of players across different age categories and their corresponding skill levels. By examining the data, I observed that there is a variation in the age distribution among players at different skill levels, as represented by the LeagueIndex. To illustrate this, I created bar charts showing the count of players in each age category for each LeagueIndex value. From the visualizations, I can observe distinct patterns in the age distribution for each skill level. For example, players with LeagueIndex 1-3 are more popular with a younger crowd (16-20), while those with LeagueIndex 4-5 are more popular with slightly older crowds (20-23). I noticed that at LeagueIndex 7, there was no age group that was up to the age of 30-40s. The oldest in LeagueIndex 7 was 26 years olds. Lastly, no information was given for age in LeagueIndex 8 and I had to replace ‘?’ to not be available.

My analysis also highlights the potential implications of age in relation to player skill levels. By examining the count of players within each age category for different LeagueIndex values, I gain insights into the demographics of skilled players. These findings can be valuable for stakeholders interested in understanding the relationship between age and player performance in StarCraft, potentially informing decisions related to targeted marketing strategies.

Linear Regression Model:
The analysis of the StarCraft player data using linear regression has produced a model that demonstrates promising accuracy. The mean squared error (MSE) value of 1.0228 indicates that the model's predictions are relatively close to the actual LeagueIndex values. This low MSE suggests that the model captures the underlying patterns and relationships between the independent variables (such as age, hours per week, APM, etc.) and the dependent variable (LeagueIndex) quite well. It provides a reliable means of predicting a player's skill level based on the given features.

Mean Squared Error: 1.022769303371555
Optimal Number of Features: 12
MLR Equation: LeagueIndex = 4.08 + (0.06 * Age) + (0.12 * HoursPerWeek) + (0.17 * SelectByHotkeys) + (0.19 * AssignToHotkeys) + (0.06 * UniqueHotkeys) + (0.11 * MinimapAttacks) + (0.21 * NumberOfPACs) + (-0.20 * GapBetweenPACs) + (-0.40 * ActionLatency) + (0.07 * ActionsInPAC) + (0.12 * WorkersMade) + (0.04 * ComplexUnitsMade)

With the above, I created a multi linear equation using these optimal numbers of features. To visualize the data, I used a bar chart. The bar chart showed ActualLeagueIndex 1 vs the PredictedLeagueIndex to showcase a player’s success in obtaining the next rank. As the ActualLeagueIndex increased so did the PredictedLeagueIndex. I noticed that the line fell significantly when it came to LeagueIndex 7, I think this is because there are less players in LeagueIndex 7, the age that most players play is 23. At this point, these players have less time to play in the game, have commitments and are not able to play as often. All other LeagueIndex had more players, more time and the age range was larger.

The predicted model is significant for stakeholders, including players, coaches, and organizations involved in StarCraft. By accurately predicting a player's LeagueIndex based on their characteristics and gameplay statistics, stakeholders can gain valuable insights. For instance, knowing which LeagueIndex a player is likely to reach based on their current performance can help in identifying their strengths and weaknesses. Organizations can make informed decisions about player recruitment and team composition by considering the predicted LeagueIndex values. Lastly, the model allows stakeholders to identify players with the potential to progress to higher skill levels, such as LeagueIndex 8.