Advice to Stakeholders

With the following hypothetical question in mind:

> *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?*

I would first suggest we narrow our data collection to just high tier players in Leagues 6-8. As an Esports Organization, we don't care so much about what differentiates a low tier player from a higher tier, but what the difference between a top tier and a professional player.

I would next offer three tools that could be made to help EG.

1. Using our model, we can see which of our players classified lower than a professional and then see which features caused them to be incorrectly classified. We can then take this information to tell our player's what skills they need to improve to perform better. **NOTE:** For this tool we would need a trustworthy model as close to 100% accuracy as possible.
2. Since our model also gives us importance of different player statistics, I would advise a way to build a tool that can take in our player's statistics to see where our player's fall against other pros. If they are below average, we they can use that knowledge to know where they can improve. This tool is slightly different than the above tool because it analyzes just the statistics without regards for rank. The above tool would be useful if a player is performing significantly below average to understand the overall picture in regards to rank.
3. For recruiting, we can use the model to intake high tier players or potential candidates. Then, if we encounter any players that are not already pros (anyone in League 6 or 7), that get classified as League 8, we might consider that they could be the next EG recruit. **NOTE:** For this tool we would need a trustworthy model as close to 100% accuracy as possible.

Finally, I would suggest the following for what we should try to collect in the future (in addition to the previous data) to support building these possible tools.

- **InGameRace:** After some research, I found that Starcraft II pros typically focus on one particular race and master it. With my limited knowledge of SC2, I know that the different races have different playstyles. Therefore, some might require more actions, more PACs, etc. This could help our model understand how to rank players that have a lower GapBetweenPACs and yet are still League 8 (pros). Likewise, when evaluating how our players are doing (tool 2), we can subset the data into the race our player mains to get a better idea of how they fall within their in game race rather than a generalization of all races.

- **AverageGameLength:** While I don't think this will be a good statistic for helping our players, I believe it will help our model and be relatively easy to collect. I imagine that the length of the game can help discern between higher tier players. For the Zerg race playstyle, professionals might be able to finish earlier than a typical high tier player.

On the contrary, professional games might also be longer because the skill levels are more evenly match and they have a better understanding of the game.

- **HoursPerWeek:** This statistic was missing for all professional players. So although it wasn't used in the previous models, I imagine it will be a tell for who is/isn't a professional player. Even if it doesn't help the model too much, it would be good to know how dedicated the player is and how consistent they are. It's a better statistic than Total Hours played because it shows how consistent they are rather than if they played a lot in the past but aren't as dedicated now.

Final Thoughts

Overall, I believe that the statistics collected now are good and I would need to do more research into how Starcraft II works and how the different races are played before I could make more suggestions for statistics to collect. I believe that the main drawbacks of our model is simply the lack of data. With at least 500-600 more datapoints on each of the undersampled leagues, I believe our model could achieve significant increases in performance. In the end, the only statistic I wouldn't advise to collect would be 'TotalHours', as I think that information is better captured in the 'HoursPerWeek' feature.