Communication to non technical audience:

- Our objective for this project is to predict a Starcraft player's rank based on various gaming performance metrics. We analyzed a dataset with information like age, hours played per week, total hours played, and specific in-game actions.
- Through our exploratory data analysis, we discovered some interesting patterns:
  - Age and Rank: Younger players tend to have a higher rank. This could suggest that younger players may have more time to dedicate to the game, or they might be quicker in their actions, leading to better performance.
  - Time Spent Playing and Rank: Players who spend more time playing the game each week and overall tend to have a higher rank. This indicates that practice and familiarity with the game play a crucial role in a player's performance.
- However, we found that the data for total hours played, age, and hours played per week were missing for players ranked in the highest league (LeagueIndex 8). To avoid discarding this valuable data, we replaced these missing values with median values from players ranked in the second-highest league (LeagueIndex 7).
- We also encountered some unrealistic data entries, like players who reported playing 168 hours in a week, which is technically the total number of hours in a week. To address this, we capped these values at a more realistic upper limit.
- After cleaning our data, we observed that some of the performance metrics were very closely related to each other, which could make it difficult for our model to determine the unique contribution of each metric. To simplify this, we removed some closely-related features.
- We experimented with various machine learning models like Random Forest, Neural Networks, Gradient Boost, and Logistic Regression models to choose our final model. All models performed relatively similarly, correctly predicting a player's rank in about half of the test cases. While this accuracy might not seem high, it's important to remember that we're trying to predict one out of eight possible ranks, so this level of performance is quite promising.
- In this specific case, we are trying to predict one out of eight possible ranks for each player. Since the ranks are closely related and can vary by only one position, we introduced an error range of plus or minus 1. This means that if our model predicts a rank that is one position higher or lower than the actual rank, we consider it an acceptable prediction.

- By incorporating this error range, our accuracy significantly improves to 88%. This means that in nearly 9 out of 10 cases, our model predicts the rank either correctly or within one position of the true rank. This level of performance is quite promising, considering the closely related nature of the ranks. Even if the model predicts a rank that is off by one, it is still considered a valuable prediction since it is very close to the actual rank.

Hypothetical :

- Data is missing for attributes TotalHours, Age and HoursPerWeek specifically for LeagueIndex 8 and therefore more data without missing features for LeagueIndex 8 should be gathered.
- Overall more data over all LeagueIndex's will help make the model more robust since the maximum testing accuracy that is being achieved is 50%. That points to the fact that if a new user's data is fed to the model to predict the League Index, there is only a 50% probability that the model would correctly predict the actual League Index (out of the 8 possible values that it can predict).
- Some more recommendations to get more features:
    - Collect Detailed Gaming Behavior: Expand the range of performance metrics, such as the number of units destroyed, number of games won vs lost, types of strategies employed (aggressive, defensive, etc.), and frequency of in-game communication. These could provide a more nuanced understanding of the player's gaming style and their performance.
    - Capture Hardware and Software Details: Information about the player's gaming setup, like PC specifications (CPU speed, RAM size, etc.), type of mouse/keyboard used, and network speed could also influence their gaming performance.
    - Longitudinal Data: Instead of a snapshot, collect data over a period of time. This will provide insights on how a player's performance changes over time, and could be useful for detecting patterns and trends.
    - Game Specific Events: Data related to participation and performance in in-game events or tournaments could provide valuable insights into a player's ability to perform under competitive pressure.
    - Data on Training or Practice Sessions: Besides actual game data, information on practice sessions could provide insights into how training influences performance.
    - More Demographic Data: More data about the player's background such as occupation, level of education, or even geographic location might provide additional insights.