Summary of Findings

The objective was to determine which "kinds" of players are most likely to contribute a significant amount of data by predicting their experience level based on their age and the number of hours they play. Using a k-Nearest Neighbors (k-NN) classification model, we achieved a high accuracy on the test data after fine-tuning the number of neighbors (k) and standardizing the predictors. This result indicates that age and played hours are strong predictors of a player's experience level. The model successfully classified players into their respective experience categories, providing insights into their potential engagement and data contribution.

Was This Expected?

The findings were aligned with expectations, as both age and played hours are intuitively correlated with experience. Older players or those who play for more hours are likely to have higher experience levels, making these predictors natural choices. However, achieving a high level of accuracy required careful hyperparameter tuning and preprocessing steps, such as normalization, highlighting the importance of these techniques for k-NN performance. The results confirmed that a broader search for optimal k values was essential for improving the model's predictive accuracy.

Impact of Findings

The findings have significant implications for data-driven decision-making. Identifying player demographics that are most likely to contribute substantial data can help game developers and platforms tailor their strategies. For instance, targeted engagement campaigns can focus on age groups or activity levels identified as highly experienced, improving player retention and satisfaction. Moreover, understanding the predictors of experience can inform game design decisions, such as creating tutorials or difficulty levels suited to different player segments. These insights also allow for optimizing resources by focusing data collection efforts on players more likely to provide meaningful feedback or contributions.

Future Questions

1. **Additional Features:** Could including more features, such as game genre, time spent daily, or social interactions, enhance the model's performance?
2. **Population Generalization:** How well does the model generalize to new games or player populations with different behavior patterns?
3. **Dynamic Behavior:** Can we extend the model to dynamically predict experience over time as players age or change their playing habits?