**What Did We Learn Through the Process?**

Throughout this project, we undertook a comprehensive analysis of soccer match data spanning multiple seasons from 2019 to 2024. The process involved several key steps, including data wrangling, exploratory data analysis (EDA), and predictive modelling. Here are some of the key learnings:

1. **Data Wrangling and Tidying**: We learned the importance of cleaning and preparing data to ensure accuracy and consistency. Handling missing values, correcting data types, and standardizing categorical values were crucial steps in this process.
2. **Exploratory Data Analysis**: EDA helped us uncover patterns and relationships within the data. Visualizations such as histograms and scatter plots provided valuable insights into the distribution of goals and their relationships.
3. **Predictive Modelling**: By performing regression analysis and A/B/n testing, we were able to build models that predict total goals based on first half goals and other variables. This step highlighted the importance of selecting relevant features and evaluating model performance.

**Are the Results What You Expected?**

The results of our analysis were both expected and surprising in some aspects:

* **Expected**: We anticipated finding a positive correlation between first half goals and total goals, which was confirmed by our analysis. The regression models showed that first half goals are a significant predictor of total goals.
* **Surprising**: The perfect fit in Scenario C, where both full-time home and away goals were included, was somewhat surprising. This result indicates that these variables together can explain all the variance in total goals, highlighting their strong predictive power.

**Key Findings/Takeaways**

1. **Correlation Between First Half and Total Goals**: There is a strong positive correlation (0.693) between first half goals and total goals, indicating that the number of goals scored in the first half is a good predictor of the total goals scored in a match.
2. **Predictive Models**: The regression models demonstrated varying levels of predictive power:
   * **Scenario A**: Using only first half goals, the model explained 48.1% of the variance in total goals.
   * **Scenario B**: Adding full-time home goals improved the model, explaining 69.5% of the variance.
   * **Scenario C**: Including both full-time home and away goals resulted in a perfect fit, explaining 100% of the variance.
3. **Data Quality and Preparation**: Proper data wrangling and tidying are essential for accurate analysis. Handling missing values, standardizing formats, and creating new features were critical steps in our process.
4. **Visual Insights**: Visualizations such as histograms and scatter plots were invaluable in understanding the data and communicating findings. They provided clear and intuitive representations of the relationships between variables.

**Final Thoughts**

This project provided valuable insights into the relationship between first half goals and total goals in soccer matches. The findings can be used to inform various applications, from marketing strategies to real-time match predictions. The process also underscored the importance of thorough data preparation and the power of visualizations in data analysis.

Moving forward, there are opportunities to explore additional features, use more advanced models, and develop real-time prediction systems to enhance the analysis further. This project serves as a solid foundation for future work in soccer analytics and predictive modelling.