**Goals**

1. **Primary Goal**: To analyze soccer match data to understand goal scoring patterns and predict total goals scored in a match based on first half goals.
2. **Secondary Goals**:
   * Identify key factors influencing match outcomes.
   * Develop predictive models to forecast total goals.
   * Provide actionable insights for marketing strategies and team performance analysis.

**Data**

1. **Source**: CSV files representing different soccer seasons from 2019 to 2024.
2. **Variables**:
   * **Div**: Division
   * **Date**: Date of the match
   * **Time**: Time of the match
   * **HomeTeam**: Home team
   * **AwayTeam**: Away team
   * **FTHG**: Full-time home goals
   * **FTAG**: Full-time away goals
   * **FTR**: Full-time result (H = Home win, D = Draw, A = Away win)
   * **HTHG**: Half-time home goals
   * **HTAG**: Half-time away goals
   * **HTR**: Half-time result (H = Home win, D = Draw, A = Away win)
   * **HS**: Home shots
   * **AS**: Away shots
   * **HST**: Home shots on target
   * **AST**: Away shots on target
   * **HF**: Home fouls
   * **AF**: Away fouls
   * **HC**: Home corners
   * **AC**: Away corners
   * **HY**: Home yellow cards
   * **AY**: Away yellow cards
   * **HR**: Home red cards
   * **AR**: Away red cards

**Analysis**

1. **Data Wrangling and Tidying**:
   * **Cleaning**: Handle missing values, correct data types, and standardize categorical values.
   * **Transformation**: Create new features such as total goals (Total Goals) and goal difference (Goal Difference).
   * **Aggregation**: Group and summarize data to the appropriate level of detail.
2. **Exploratory Data Analysis (EDA)**:
   * **Descriptive Statistics**: Calculate mean, median, range, and correlations for key variables.
   * **Visualizations**: Create histograms for univariate exploration and scatter plots for multivariate exploration.
3. **Predictive Modelling**:
   * **Correlation Analysis**: Assess the relationship between first half goals and total goals.
   * **Regression Analysis**: Build linear regression models to predict total goals based on first half goals and other variables.
   * **A/B/n Testing**: Compare different predictive models to determine the best approach.
4. **Communicating Findings**:
   * **Documentation**: Create detailed documentation of the process, findings, and insights.
   * **Visualizations**: Develop clear and informative visualizations to support the findings.
   * **Tailored Communication**: Present the results in a way that is accessible to different audiences, including the general public, marketing departments, managers, and other data analysts.