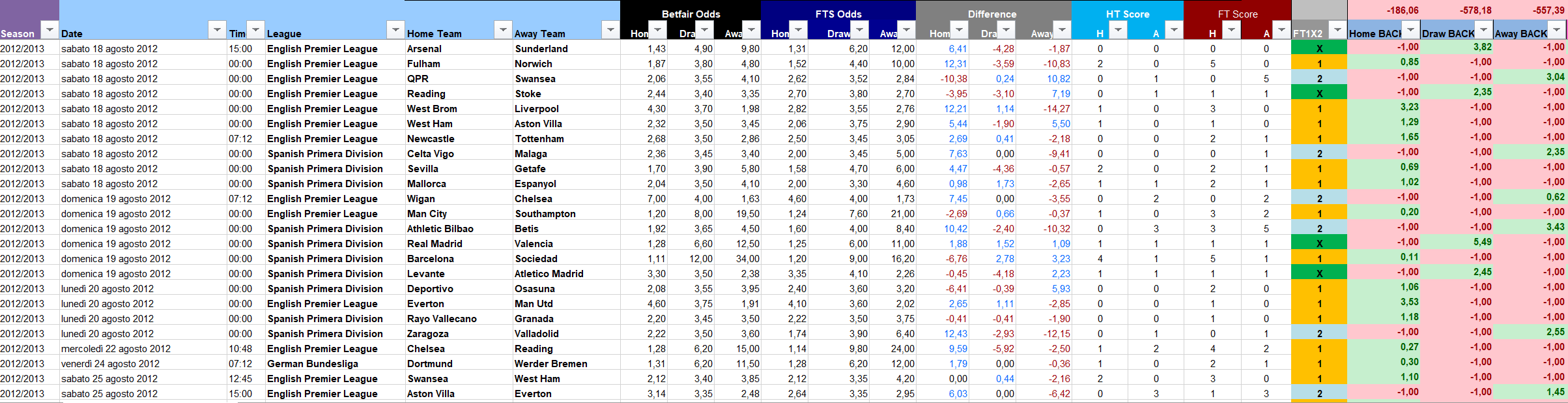
**Project Title**: Development of Advanced Predictive Models for Football Betting

**Detailed Project Description**: I am a sports trader and investor in the sports betting sector. I am looking for a data science expert for a long-term collaboration project. The project aims to develop sophisticated predictive models using advanced machine learning techniques, statistical analysis, and artificial intelligence, with the goal of predicting the results of football matches and other key performance indicators (KPIs) to optimize betting strategies and maximize investment returns.

**Available Data**:

* Complete historical data of football matches from various leagues, including details such as date, teams, final scores, and more. The data available are already cleaned and neatly formatted within dedicated databases, so there is no need for additional data cleaning.
* The databases available contain various types of data such as pre-match odds KPIs, goals scored, goals conceded, match results, home and away performance, etc.
* The databases include various types of data that can be worked on, including ELO models, linear regressions, and other statistical models.



**Specific Project Objectives**:

Starting from the available data, the main objective is to find patterns capable of predicting football match results, achieving measurable profit over the long term (complete football season). This process includes the creation of models that are trained through backtesting on about 70% of the available data and verifying the consistency of the backtest with the remaining 30% of the data. The available databases range from the 2012/2013 football season to the present and are continuously updated.

1. **Exploratory Data Analysis**: Conduct exploratory analysis to identify patterns, correlations, and key factors that influence match results.
2. **Feature Engineering**: Develop and select the most relevant variables (features) for the predictive models.
3. **Model Construction and Validation**: Develop various machine learning models (e.g., neural networks, decision trees, ensemble models) and evaluate their accuracy and reliability through cross-validation and backtesting techniques.
4. **Optimization and Tuning**: Optimize models to maximize accuracy and minimize overfitting.
5. **Results Interpretation**: Provide detailed analysis and interpretations of model results, including the probabilities of different outcomes and the most influential factors.
6. **Strategic Recommendations**: Based on the results of the models, provide recommendations on how to apply this information to improve betting and investment decisions.

At the end of the project, a complete list of KPIs and patterns should be provided to accurately apply investment strategies.

**Technical Requirements**:

* Deep knowledge in machine learning techniques and statistical modeling.
* Experience in working with complex datasets and conducting feature engineering.
* Ability to implement both simple (e.g., logistic regression) and complex models (e.g., deep learning).
* Programming skills, preferably in Python or R, and familiarity with libraries such as scikit-learn, TensorFlow, or PyTorch.

**Budget and Timeline**: Provide an appropriate estimate and project timeline.

**Collaboration and Communication**: I am looking for a long-term collaboration with regular updates on the project status and availability for periodic discussions on progress, strategies to be adopted, and possible maintenance.

# **Subject: Data Science Project Specifications for Sports Betting**

As discussed in our recent call, this document aims to outline clearly and concisely the specifications of the data science project you will be working on. The focus is on developing prediction algorithms for sports betting, with the goal of generating substantial profit through differentiated analysis of the main soccer leagues:



• English Premier League

• German Bundesliga

• Spanish Primera Division

• Italian Serie A

Main Objective: Develop predictive systems for the following betting outcomes:

column 20 Home BACK 15906 non-null float64

column 21 Draw BACK 15906 non-null float64

column 22 Away BACK 15906 non-null float64

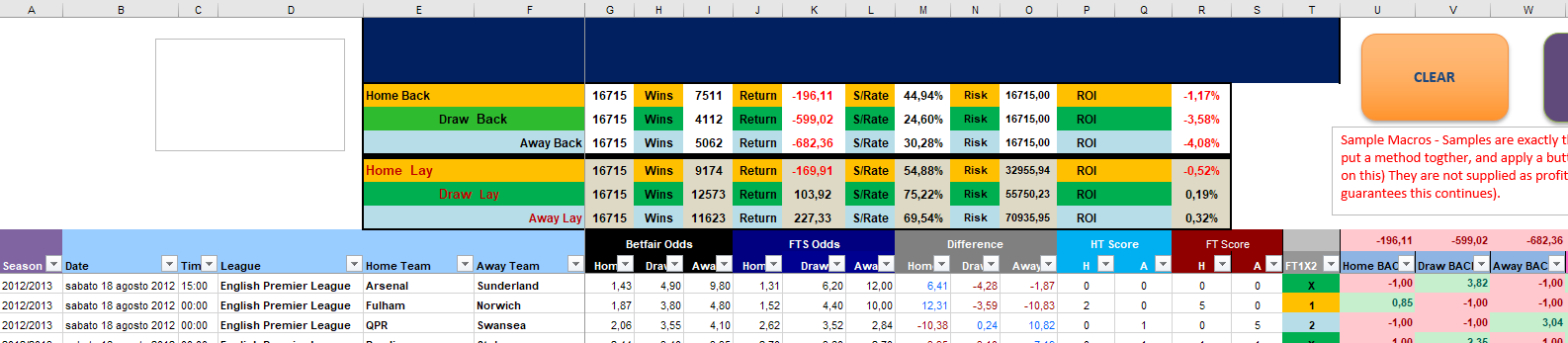


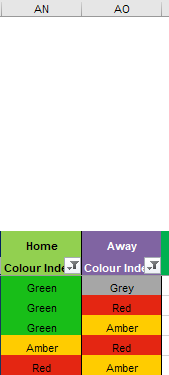
• BACK 1: home team win

• BACK X: draw

• BACK 2: away team win

Analysis Parameters: You will use a set of parameters extracted from our database, which can be employed individually or in combination with each other:





• LEAGUE: Identified in column D.

• BETFAIR ODDS: Available in columns G (for BACK 1), H (for BACK X), and I (for BACK 2).

• DIFFERENCE: Values located in columns M (for BACK 1), N (for BACK X), and O (for BACK 2).

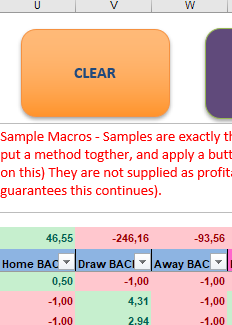
• COLOR INDEX HOME AND AWAY: Present in columns AN (for the home team) and AO (for the away team).

Success Metrics: The goal is to achieve a minimum of 70 annual profit points, measured as return on investment (ROI) in the following columns:

• Column U for BACK 1

• Column V for BACK X

• Column W for BACK 2



Data Volume: The data base for analysis should cover about 500-700 seasonal matches, combining all systems together.

Model Development: You will need to find combinations of data and KPIs that reveal effective predictive models. Each developed model will require detailed backtesting, broken down by seasons and months, to verify its reliability and profitability over time. It is essential to train the models on a significant sample of data and test them on the remaining database to ensure their robustness.

Work Phases: The work must be organized in stages, starting with the development of systems for BACK 1, followed by BACK X and BACK 2. This step-by-step approach will ensure methodical and structured progress in the project.

I trust in your competence and dedication to successfully complete this project. For any clarification or further discussion, do not hesitate to contact me.