**Introduction:**

Soccer is a team sport played between two teams. It is a well known and loved sport, with approximately 3.5 billion fans around the world. But as much as people enjoy sitting in the crowd, at their favorite bars or at home and watching their favorite team plays, there is another aspect to this sport which makes the games even more interesting for their viewers – and that’s betting.

In general, sports betting refers to attempts at predicting sports results, while placing a wager on the outcome. In soccer, there are many aspects of betting – one can bet on simple outcomes in the game, such as the final result of the game (winning team), or the exact result (amount of points per team). However, betting can be much more complicated as one can also try and predict many other outcomes, such as number of “offsides”, number of “corners”, first team to achieve a “corner”, etc.

The possibilities for the gamblers are endless, but the one that gains the most, statistically, is the betting company. There are many soccer betting companies out there, aiming to increase their income. The main way in increasing the company’s outcome is by setting the betting odds in the optimal way.

The betting odds can be seen as a function of the likelihood of the outcome – each outcome receives a number, which would increase as the likelihood for that outcome decreases. For example, in a simple case, assume the value for team A’s win was set to and a gambler placed a bet of dollars on that outcome. If this outcome happened, the gambler would gain (after “losing” the they bet) dollars. Therefore, if the company set to be very low, they would lose less money to that gambler.

So, the question is, how does the company choose the optimal betting odds for a game? Many things can affect a game’s outcome, such as the teams’ levels, the players’ mood, the weather, etc.

In order to understand how to answer the question above, we try to solve the following question:   
**Does the outcome of a match affect the team’s outcome in their following match?**

By knowing the answer to this question, betting companies can get better at setting the odds, and thus increase their income. Also, gamblers can get better at trying to predict the outcome.

In this project we will use the different methods we learned in the course and calculate the ATE. By using matching and learners, and some more statistical hypotheses testing, we plan to answer this question.

**The Data:**

The data we use in this project is taken from Kaggle. It is a database, composed of multiple tables, but the tables we will be using are:

“Match” – a table of almost 26,000 records of matches, including the participating teams, dates, outcomes and betting odds given to each team by multiple betting companies.

* Add more tables if necessary.

The data can be found [here](https://www.kaggle.com/datasets/hugomathien/soccer?select=database.sqlite).

**Exploratory Data Analysis:**

First we preprocess “Match” table:

* We process the date column in order to be able to observe two consecutive games.
* We average the bets that each team received in all betting sites.
* We add a “goal\_diff” column: subtract away team’s goals from home team’s goals. This columns describes which team won (using the sign of the value) and by how many goals (using the distance from 0).
* We choose the relevant columns – columns that describe each team, columns that describe the results (including “goal\_diff”) and columns that describe the bets.

Then we create another table we call “Wins”:

* For each team, we summarize the wins and losses in the league.
* We calculate the total points for each team – 3 points per win, 1 point per draw and 0 points per loss.

We finally merge these two tables to a single dataset. This dataset is composed of a row for each match, which contains details about this match – participating teams, points and results, and bets.