**SPI:** this variable is calculated by … and … and is an advanced metric of team strength of both the home and away teams. The metric incorporates recent results and the strength of the teams played previously. It therefore is also a measure of recent form, with a string of bad results against lesser rated teams lowering the rating of a team possibly.

**Rating\_difference;** this variable is calculated by subtracting the away team spi rating from the home team spi rating. A large positive(negative) value entails a large quality difference in favour of the home(away) team. In case of a large quality difference it is thus more likely for a certain team to win the match.

**Team average age:** This variable captures the average age of the entire squad at the beginning of the season.

**Old or young:**  this variable captures the differences in age between the home and away team, where a positive value implies the home team have an older squad on average than the away team, and a negative value implies the away team have older squad players.

**Importance of match:** This variable is calculated by … and …. and represents the importance of the match outcome on the league developments for both teams their perspective. There are 2 values, one value for the away team and one value for the home team

**Spread importance:** this variable is calculated by subtracting the away importance value from the home importance. A positive value thus can be viewed as a match where the match is relatively more important for the home team whereas a negative value implies the contrary. Values close to 0 signify a situation where the match is of approximately equal importance to both teams.

**Covid:** a dummy variable that indicates whether or not the game was played before or after the start of the covid-19 pandemic. A value of 1 represents matches that are played behind closed doors, with no spectators attending, whereas a 0 implies fans were present.

**Crowd occupancy:** The ratio of the number of attendants to the number of seats in the stadium. Due to data collection and computation limits, the crowd occupancy is calculated as the total number of attendants within a league divided by the average capacity for all teams in the league. For example a league where the clubs on average have a stadium capacity of 50.000 with 4.000.000 fans in total visiting 200 matches means an average attendance of ((4000000/200))/50.000 = 0.4.

**Expected goals:** Expected goals are measured as the expected goals that would have been made from a teams chances within a match, based on average player and average situation. It is a sum of the quality of chances created by a team. A shot for open goal with no opponent player in tackling distance has a very high xg chance since it will be converted into a goal almost all the time. A 40 yard shot that gets deflected by a defender on the way and flies in with luck has a low xg value, as such shots have a low probability of producing a goal. Expected goals metric thus takes out luck out of the performance of a team and can be seen as a good metric of match performance and match outcomes.

**Spread in fouls:** The number of fouls committed by the home team - the number of fouls committed by the away team. A positive number means the home team has committed more fouls than the away team in the specific match, and a negative number implies the away team committed more fouls. 0 implies an equal number of fouls for both teams. Based on a non biased referee and controlling for team dominance and team quality we expect the number of fouls for home and away teams to not be significantly different from each other.

**Spread in yellow cards :** The number of yellow cards for the home team - the number of yellow cards for the away team. A negative number can be interpreted as the away team receiving more yellow cards than the home team. A positive number means the number of yellow cards received by the home team exceeds those of the away team. 0 implies an equal number of yellow cards for home and away team. This could be 0 for both but also 1 or 2 or more for each side. Based on a non biased referee and controlling for team dominance and team quality we expect home and away teams to have no significant difference in the number of yellow cards.

**Spread in red cards:** The number of red cards for the home team - the number of red cards for the away team. A negative number implies the away team received more red cards than the home team. A positive number signifies the number of red cards received by the home team exceeds those of the away team. 0 implies an equal number of red cards for home and away team. This could be 0 for both but also 1 or 2 for each side. Based on a non biased referee and controlling for team dominance and team quality we expect home and away teams to have equal number of red cards on average

**Absolute crowd size:** The absolute number of supporters present at the game**.**

**Corners:** The amount of corner kicks for a team within the match. This can be seen as a measure of attacking dominance since corners are often results of defending clearances or goalkeeper saves.

**Shots on target:** The number of shots undertaken by a team that without interference of any player would be a goal. Examples include a shot that would have been on target but was blocked by an opponents defender, or a shot saved by the keeper that was going in the goal otherwise.

**Shots/shots on target ratio:** The ratio between the total number of shots undertaken by a team to the total number of shots on target from that same team. A higher ratio means better quality shooting of a team and more efficient use of shots.

**Shots:** The total number of shots undertaken by a team within the match. This includes all of shots wide, shots blocked, shots on target, shots on woodwork.

**Home Goals:** This variable denotes the number of goals scored by the home team in the specific games

**Away Goals:** This variable denotes the number of goals scored by the away team in teh specific game.

**Points home:** The number of points for the home team in this specific match.

**Points away:** The number of points for the away team in this specific match.

**Percentage Points home from total:** Home points as a fraction of total points obtained in the game. A value above 50 percent implies a home advantage as more points are collected on average at home

**Points away from total:** Away points as a fraction of total points obtained in the game. A value above 50 percent implies a home disadvantage as more points are collected away on average than at home

**Difference in goals home/away:** DIfferences in goals scored by the home and away team per match. A 4-0 home win implies a value of +4, whereas a 3-0 away win implies a value of -3. A draw signifies a value of 0. A negative value implies more goals scored by home teams than away teams, although not necessarily more home wins than away wins.

**Difference in points home/away:** the difference in points for the home and away team for a single match, calculated as points of the home team - points of the away team. A home win means a +3 difference, whereas an away win gives a -3 difference. A draw results in 0. Thus a overall positive number for this metric can be interpreted as more home than away wins, whereas a negative number implies more away than home wins

**Difference in expected goals home/away:** Difference in the number of expected goals for home and away teams. A match where the home team had chances created such that their expected goals totaled up to 2.93, and the away team had expected goals of 1.39 means a spread of 2.54, the higher this spread the more dominant the home team was and the more chance the home team had of winning the match. Vice versa, a large negative value signifies an away team that dominated the match and chances and had a high probability of winning the match.

**Home goals HT:** The number of goals scored by the home team in the first halfe

**Away goals HT:** number of goals scored by the away team at half-time

**HT result:** whether the first half result would have resulted in a home win, draw or away win.

**Referee bias:** since referee bias is a latent concept and not easily capturable in a single number, we use SEM to construct a latent variable Referee bias: which takes in multiple numbers on the referee assessment of home and away teams. Based on a non biased referee and controlling for team dominance and team quality we expect home and away teams to have equal number of fouls/red cards and yellow cards.

**Team performance home advantage:** Team performance is difficult to capture in a single variable, goals can be result of pure luck. By combining, goals, expected goals, corners, shots, shots on targets into a single measure, we obtain a richer definition of team performance, to better capture the different dynamics behind team performance.

**Outcome home advantage:** a variable constructed by combining the different measures of match outcomes to obtain a better measurement of outcome. It combines, points, expected goals, goals and expected points to get a very rich definition of home advantage.

Uitgebreide