

For this exercise, use the football dataset found in the file "corners_data.csv" and "corners_prices.csv".

Definitions:

- "match_id" is the unique identifier for the match
 - "home_team_id" and "away_team_id" are the unique identifiers for the home and away team
 - "competition_id" is the unique identifier for the league
 - "season_id" is the unique identifier for the season
 - "date_time" is the date and time of the match
 - "gameweek" is the gameweek of the match
 - "home_ft_score" and "away_ft_score" are number of home and away goals of the match
 - "home_corners" and "away_corners" are number of home and away corners of the match
-
- "odds_type" is the type of the prices, matchodds (1X2), over/under (OU) or handicap (HC)
 - "oa"
 - for 1X2 this is the price of away team having more corners
 - for OU this is the price of total corners being under the line ("od")
 - for HC this is the price of away team winning corners on the handicap line ("od")
 - "oh"
 - for 1X2 this is the price of away team having more corners
 - for OU this is the price of total corners being above the line ("od")
 - for HC this is the price of home team winning corners on the handicap line ("od")
 - "od"
 - for 1X2 this is the price of draw in corners
 - for OU this is the line
 - for HC this is the handicap line with respect to the home team (i.e. a -0.5 line means it is still considered away team having more corners even when the score is draw, because the handicap applies -0.5 to home corners)

Tasks:

Q1: Build a model to predict the number of home and away corners. Explain the process - including data exploration, features and model selection/comparison, model validation etc.

Q2: Using "corners_prices.csv", calculate the probability of the events happening (i.e. home most corners/away most corners/draw for 1X2, over/under the line for OU, home/away having most corners under the handicap line for HC).

- a) How many selections are you betting? Why?
- b) Calculate "EV" as the Expected Value of a bet

Now look at "corners_prices_results.csv". This file contains the match results of the games in "corners_prices.csv".

- c) Define "Stake" as the amount that you bet, calculate "PnL" as the Profit and Loss and "RoI" as PnL/Stake.
 - i) Calculate the overall PnL and RoI
 - ii) Plot the cumulative PnL and EV as a function of time in days
 - iii) Plot the 2-week rolling PnL and EV average
- d) Do you think that your assumed EV is an accurate representation of your real edge? (Hint: Run a Monte Carlo simulation)