English Premier League

Dataset –

epl_matches_train • 2661 matches from season 2008 to 2015 • Team line up with squad formation (X, Y coordinates) • Detailed match events (red-card, yellow-card etc) ***epl_matches_test*** • Matches from season 2015-2016	***epl_players*** • Players' attributes (like attacking_work_rate, defensive_work_rate etc.) ***epl_teams*** • Teams' attributes (like defense_pressure, chance_creation etc.) ***epl_goals*** • Goals detail of all past matches (2008-2015) ***epl_potential_shots*** • Close shots from all past matches (2008-2015)

Please refer to ***column_description*** text file for understanding column names

Problem statement –

- Which are the 10 most and least valuable players for the session 2014-2015? –

  As a Data Scientist, your challenge is to investigate the available data and develop a metric(s) to identify the most and least "valuable" player.
  - Use number of goals, close shots and other relevant attributes

- Predict the match results for season 2015-2016 -

  Build a model to predict the match results (Win, Lose or Draw) for the 2015 season of the English Premier League (EPL). You would have to come up with metrics based on which you would predict the outcome.

  win – Home Team goals > Away Team goal
  draw – Home Team goals = Away Team goals
  lose – Home Team goals < Away Team goals

Submission –

- For first part – Submit 10 most and least valuable players in player_list_submission file.
- For second part – Provide prediction for the test data in prediction_submission file.
- Make short word document and explain –
  metrics you used for the first part
  procedure you followed for the second part
- Also provide your python or R scripts for first and second problem

Make a zipped folder with all required files listed above