Angstrom Round 1 Data Science Assessment

Overview

We would like you to demonstrate your Data Science skills through completing this assessment. The assessment is designed to be open-ended, and there are no correct/wrong approaches in completing it. Please try to demonstrate your skills and mindset when it comes to exploring data you have not encountered in the past!

Assessment

There should be multiple csv files within the zip file (`AngstromDataScienceAssessment.zip`) that was attached. These files represent various fixture-level college basketball (NCAA Basketball) statistics from the past 2 years. The objective will be to **build a model that is able to predict the winners for the fixtures on the 25th of February, 2023**. How you go about this is entirely down to you, as long as you are demonstrating throughout the reasonings of your approach/thinking.

What to Send Back

Most likely, you will be working in Python, R or other languages suitable for the assessment:

- Python: **.ipynb** (if you worked on Google Colab, please download) or **.py** files.
- R: **.Rmd**/**.R** files.
- Other: Please send back files that we can easily navigate through and run your written code with.

Ideally, you should have proper commenting such that you do not need anything else - but if you do wish to add a supporting document, you may also send a **max 1 page** PDF to illustrate your thought process/findings.

Mini-Tour of the Datasets

Each of the fixtures are represented uniquely by a *FixtureKey*. This is in a format:

"<Team 1> v <Team 2> <Date>"

Please note, that if the game is played at the home of one of the teams (*IsNeutralSite*=0), then Team 1 will be listed as the home team.

1. `box_scores.csv`
   Box scores are tables that represent the match statistics at a player/team level. This dataset in particular, is displaying the main basketball statistics for each fixture, at the team level. There will be data on:
   - Number of shots made
   - Number of shots attempted
   - Number of rebounds
   - Number of shots assisted
   - Number of defensive steals forced
   - Number of opponent shots blocked
   - Number of turnovers committed
   - Number of fouls committed
2. `fixture_information.csv`
   Additional information of each of the fixtures that may be useful. The *TipOff* time is listed in Eastern Time. Please note that some of the data could be missing values for some of the fixtures.

3. `test_fixtures.csv`
   The set of *FixtureKey*s on the 25th of February, 2023, to predict the fixture outcomes for.
4. `test_fixtures_actuals.csv`
   Please use this where suitable, but this data shows what **actually** happened for the fixtures to predict, as well as some additional betting information provided by bookmakers. (Guide on *TeamHandicap*: -7.0 would mean that the team was predicted to win by 7 points, and +3.5 would mean that the team was predicted to lose by 3.5 points.)

Have fun, and we look forward to analysing your work!