2018 NBA Hackathon Application – Business Prompt

Task: We have provided a sample of 460 games (230 from the 2016-17 season and 230 from the 2017-18 season) for which your task is to predict total international viewers. To assist with your model, we have provided with viewership numbers from 1,000 games by country in each of the two seasons, plus stats on team and player performance and status. **Using these inputs, we would like you to predict the total number of international viewers for each of the 460 games in the test set.**

Team game stats and player stats are provided for all games in the training data set, but have been removed from the games in the test set. You are given four files:

- `game_data.csv`
    - This dataset includes incoming wins and losses (i.e., record going into a game), dates, and selected game stats for each game in the 2016-17 and 2017-18 seasons. Game stats are only included for games in the training set – not in the test set.
- `player_data.csv`
    - This dataset includes performance stats for each player in each game in the 2016-17 and 2017-18 seasons. It also includes indicators for whether the player was selected as an All-Star for the season in question and whether the player was active for the game. Game statistics are only included for games in the training set – not in the test set. All-Star status and active vs. inactive status is provided for all games.
- `training_set.csv`
    - This dataset includes total viewership for each international country for each game in the 2016-17 and 2017-18 seasons. Included are 1,000 games from each season.
- `test_set.csv`
    - This dataset includes a list of games in the 2016-17 and 2017-18 seasons for which you are expected to predict total international viewership (i.e., the sum across all countries). Included are 230 games from each season.

You will be graded on **Mean Absolute Percentage Error (MAPE)** on Total Viewers. We selected this metric due to natural scaling in the international viewership data. This metric is defined as:

$$\text{MAPE} = \frac{1}{n}\sum\nolimits_{i=1}^{n}\left|\frac{A_i - P_i}{A_i}\right|,$$

where n = *460* is the total observations in the test set, and A_i and P_i are the ith actual and predicted Total Viewers. Some tips to help you in your modeling:

- Consider all factors that may drive viewership. Team strength may be one, but there may be others such as market size and/or social following.
- Consider temporal/seasonal effects such as day of week, opening day/week and holidays.
- Consider using other public information like Google Trends or further historical team performance if you find it may helpful. Not required.

Please return to us a copy of `test_set_[Team_Name].csv` with the "Total Viewers" column filled in with your response variable. Please also return all code or relevant working files. Thank you!