

2018 NBA Hackathon Application - Basketball Prompt

Task: Basketball Analytics: Use the attached data to calculate plus/minus for each player in each game. Plus/Minus is defined as the team's score differential while the player is on the court. Note: When a player is substituted before or during a set of free throws but was on the court at the time of the foul that caused the free throw, he is considered to be on the court for the free throws for the purposes of plus/minus. A player substituted in before a free throw but after a foul is not considered to be on the court until after the conclusion of the free throws.

Ex: In a Raptors vs. Knicks game, Kyle Lowry commits a shooting foul on Kristaps Porzingis. Porzingis shoots and scores the first free throw, and then Jonas Valanciunas enters the game for Serge Ibaka. Porzingis makes the second free throw. For the purposes of plus/minus, Ibaka (because he was on the court at the time of the foul) receives a value of -2 point differential for those free throws and Valanciunas is not considered on the court until after the free throws are completed.

This folder includes three data sets: Play by Play, Event Code Description and Game Lineups. Please submit your answer in a .csv file and save your code, spreadsheets and all other work in a zip file.

Please submit your answer in a spreadsheet titled "**Your_Team_Name_Q1_BBALL.csv**" as the title, substituting in the name of your team for "Your_Team_Name." The final product should have 3 columns. Column 1: Game_ID, Column 2: Player_ID, Column 3 Player_Plus/Minus.

Provided Data:

- *NBA Hackathon – Event Codes.txt*
 - This dataset provides look up values for the event message types and action types found in the play by play dataset. Each code is converted to an English language description of the event.
- *NBA Hackathon – Game Lineup Data Sample (50 Games).txt*
 - This dataset provides start of period player availability.
 - Game_id – A unique game code for each game
 - Period (Quarter) – The associated period of the line up (overtime period are indicated by values greater than 4)
 - Person_id – A unique identifier for each player
 - Team_id – A unique identifier for each team
 - Status – A variable indicating whether a player is active (A) or inactive (I)
- *NBA Hackathon – Play by Play Sample (50 Games).txt*
 - This dataset provides play by play information on the event level for each game.
 - To properly sort the events in a game use the following sequence of sorted columns: Period (ascending), PC_Time (descending), WC_Time (ascending), Event_Number (ascending)
 - Event_Number – An ordered counter for each event in a game. Note, this number may not be perfectly sequential so please use the sorting methodology outlined above
 - Event_Msg_Type, Action_Type – Coded descriptions of what happened during the event
 - WC_Time – The in-arena time of the event in Unix format. It is coded in tenths of a second.
 - PC_Time – The time on the game clock in tenths of a second (e.g. 7200 corresponds to 720 seconds/12 minutes remaining in the quarter)
 - Option 1 – On a shot attempt, this column will tell you the point value of the shot
 - On free throw attempts, if the value in this column is 1, it means it was a made free throw, otherwise, it was missed.
 - Person1, Person2 – The person_ids of the players who are directly associated with the event (e.g. If the event is an assisted made basket, Person1 is the shot maker and Person2 is the player who assisted)
 - In the case of a substitution, the Event_Msg_Type will be 8, Person1 will be the ID for the player leaving the game, and Person2 will be the ID for the player entering the game.
 - Team_id – In most scenarios, this is the Team ID associated with the Person1 column. However, there are instances when this is not the case. To accurately and consistently identify a player's team, we suggest merging in the Game Lineup dataset on the Person1 and Person2 columns.