Overview

Using open sourced programming language(s) of your choice, answer the assessment questions to the best of your ability. Please spend **no more than five hours working on this assessment**. It is not necessary that you answer all the questions, rather, this is meant for you to showcase your programming proficiency and analytical acumen.

Make sure to double check your work and follow all the submission instructions.

Data Source

Your data source for this assessment will be https://eightthirtyfour.com/.

This website hosts 19 seasons of publicly available NBA play by play data. Data definitions can be found here: https://eightthirtyfour.com/data/extended_pbp_readme

---

Assessment

Your target audience for the below questions is the Head of Basketball Strategy & Analytics. Your response should be formatted in a report no longer than two pages.

Part 1: Data Comprehension *(estimated time: 1.5 hours)*

Data: Load the 2018-19 extended play-by-play dataset into your environment (2018-19 NBA Data).

(A) What values in the "`EVENTMSGTYPE`" column are associated with –
   a. Made shots
   b. Missed shots

(B) What was the most common five-man lineup played by each team? Use play clock time (columns: `["PCTIMESTRING", "TIME"]`) to rank each lineup. Display the output in a chart of your choosing.

(C) Which players led the league in three shot fouls drawn in the 4th quarter? Which players led the league in and-1s in the 4th quarter? Display the top 10 players for both questions in a table or chart of your choosing.

Part 2: Analytical Acumen *(estimated time: 3.5 hours)*

Pick **one** of the below prompts. Use your own judgement and interpretation when responding to any of the open-ended questions. Specify any assumptions you make, where applicable.

(A) Prompt 1: Select one team and analyze their performance in wins and losses.

   a. To select a team, filter the dataset using the "`HOME_TEAM`" and "`AWAY_TEAM`" columns.
   b. Conditional on the game result, what characteristics stand out? Calculate and examine both offensive and defensive 4 factors, team shot trends, opponent shot trends, and anything else you feel is relevant for this analysis.
   c. Explain and present your findings with tables and visuals. What additional research areas would you focus on if given the opportunity to expand this study?

(B) Prompt 2: Analyzing league level trends.

Load in additional extended play-by-play data from 2015-16 through 2017-18.

a. Calculate league level cumulative percentages for the following statistics: Free Throw Percentage, Two Point Percentage, Three Point Percentage, Assist to Turnover Ratio, and Free Throw Rate.
   - You must first sort the dataset by `GAME_ID`, `WCTIMESTRING`, and `EVENTNUM`. Then calculate cumulative values for each statistic, for each season.

b. Explore visualizing these cumulative statistics (you should be grouping by season). Include relevant visuals along with explanations in your final output.
   i. What trends do you observe in these statistics across seasons?
   ii. What trends do you observe within a given season?
   iii. Which of these statistics appears to stabilize, if at all, within a season?
   iv. How would you formalize a study of stabilization rate?

c. Time permitting, create and explore additional stabilization trends. Include takeaways.

Submission Instructions

Once you are prepared to submit your assessment, please create a ".zip" file with the following contents:

- A **final output** of your responses to the assessment in either a **PDF or HTML document**.
  - This output should be no longer than two pages (or the equivalent length in HTML format). Minimum font size is 11pt.
  - Written explanations, tables, and charts in response to the assessment questions should be included in this output. Do not include your code in this document.
- **All code** written to complete this assessment. Example file types to include are:

  `[".R", ".RMD", ".ipynb", ".py", ".jl", ".txt"]`

  - In your code script(s), please include a commented list of any resources you used to complete this assessment at the end of the script. For instance, Stackoverflow links should be included in your list of resources.
  - You are encouraged to write clean, easy to comprehend, functional code.
  - Your code will be reviewed. Your application will not be considered without including the scripts you wrote to complete this assessment.
- **Your updated resume.**
- **Do not** include any raw data files with your submission.

Final submission naming convention: [FirstInitial]_[LastName]_ATL2021.zip

- E.g. John Smith: J_Smith_ATL2021.zip

Submission Link: **https://hawks.com/basketballanalyticsinternship2021**