

Analyst Intern Project Brief



Outline

Investigate schedule difficulty in the National Basketball Association.

All files required to complete the project can be found [here](#). You will not need any external sources of data or information beyond those provided to you. Please use R (preferred) or Python as the primary programming language to complete this project. Your work must be your own and original. You may use AI tools to help aid your work if you include a log with your prompts in your final submission.

The following is an outline of the provided datasets:

- 1) schedule_project_template.rmd / schedule_project_template.ipynb
 - .rmd (R) and .ipynb (Python) containing all questions required to complete the project
- 2) schedule.csv
 - Dataset containing team schedules from 2014-2023
- 3) schedule_24_partial.csv
 - Dataset containing 80 games for OKC and DEN for the 2024-25 season
- 4) locations.csv
 - Dataset containing team locations
- 5) team_game_data.csv
 - Dataset containing team stats from 2014-2023
- 6) data_dictionary.txt
 - Definitions for each column included in the above datasets

Assessment

Your submission will be assessed on the following criteria:

- 1) **Code:** Modularization, reproducibility, readability, conciseness & commentary
- 2) **Visualization:** Usefulness & effectiveness
- 3) **Model:** Sound theory & judgement
- 4) **Communication:** Ability to explain reasoning clearly & effectively

Deliverables

Please complete the project by the allotted deadline and share your submission via the Box link [here](#). If you submit multiple times, only the first submission will be assessed.

Your submission should consist of a compressed zip file with a naming convention of "Last Name, First Name (Email)". For example, Smith, Jane (jsmith@xyz.com). Your name and email must match what is stated on your application.

The compressed zip file must contain the following deliverables:

- 1) A single .rmd (R) or .ipynb (Python) containing the responses to the project questions
- 2) A single .html containing the knitted output of the .rmd (R) or .ipynb (Python)
- 3) [If you used AI to complete the project] A single text file containing an ordered list of any AI prompts, along with the specific model queried (e.g. ChatGPT 5 Thinking). Do not include the AI's output