File Structure

The project files were submitted in the following structure:

Three scripts, each designed for a specific task:

- Data scraping (01_scrape_results.py)
- Transforming data and CSV file creation for the dashboard (02_full_clean.py)
- Dashboard creation (03_dash_viz.py)
  Run 03_dash_viz.py to open dashboard in browser (http://127.0.0.1:XXXX)- choose port of your choice.

Each of these three scripts can run independently when directed to their respective files.

Two folders containing all our CSV files:

- One folder ('data') containing all CSV files generated by our scraping script.
- One folder ('data_clean') containing all CSV files created during data transformation, which afterwards will be used in the dashboard creation.

Script structure and content

For the scraping we balanced the need to minimize the impact on the scraped website and the overall data usage on our end. Given the potential creation of over 40,000 files, approximately amounting to 20GB, we opted for direct website scraping, incorporating sufficient sleep timers to mitigate website stress.

The script is designed with flexibility, allowing data scraping across different years. This approach not only reduces pressure on the website by enabling batch scraping but also simplifies the inclusion of future years.

We encountered challenges in extracting dynamic boxscore table data. This issue was resolved by accessing data hidden within HTML comments.

After scraping, we convert the data into annual CSV files, ensuring organized storage. The data is then transformed and integrated into Dash for further analysis.

Draw-related data was excluded, as draws have not occurred in the NBA.

Throughout the script we have added error handling to enhance script robustness and resilience, also providing insights into potential scraping challenges.

For additional details, please refer to the script comments (marked with '#')."

Additional notes

Frontpage image: Image is AI generated, no reference available.