ANALYZING NBA GAME STATISTICS

DIGHUM 100 - Theory and Methods in Digital Humanities by Dr. Adam G. Anderson || Kenneth Mao || June 12, 2021

Project Desciption:

In team sports, the term home advantage, also called home-court advantage, describes the advantahe that the home team is said to gain over the visiting team. This benefit has been attributed to psychological effects supporting fans have on the competitors or referees; to psychological or physiological advantages of playing near home in familiar situations; to the disadvantages away teams suffer from changing time zones or climates, or from the rigors of trave.

Using data from previous seasons, including the 2020 season and excluding all the playoff (post-season) games, this project will reveal if home-court advantage really exist.

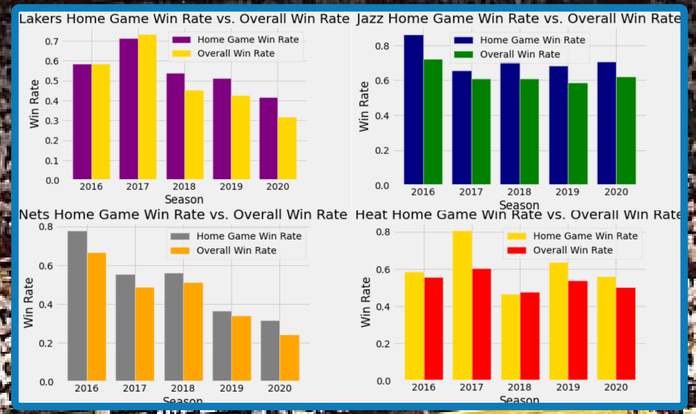
Research Questions:

- Is home court advantage real?
- Is it more likely for a team to win if it was a home game.?
- Do teams score more for home games regardless of winning or losing?

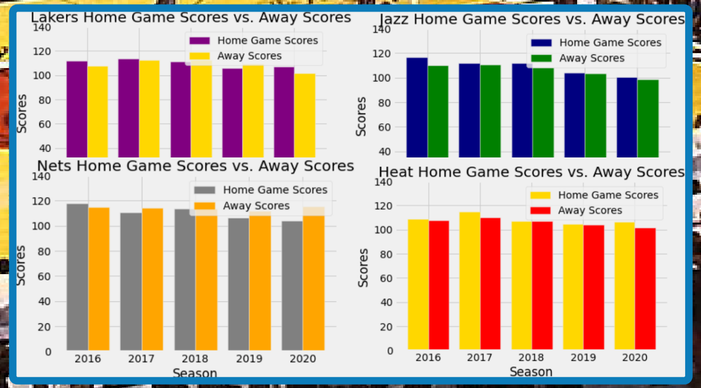
DATA DESCRIPTION:

games.csv: all games from 2004 season to 2020 season with the dates, teams and some details like number of points for each game played.
games_details.csv: details of each game, such as player scores, time playerd, etc.
players.csv: player's name and team of each year.
ranking.csv: includes the ranking of teams for the western and eastern conference for each season, contains games won, total, home, and away.
teams.csv: includes detail of each team, such as year founded, owner, arena name, nicknames, etc.

Average Overall Home Game Win Rate for the Past 5 years



Home games win rate vs overall win rate for the past 5 seasons for each of the 4 teams chosen



Home games scores vs away games scores for the past 5 seasons for each of the 4 teams chosen

PROCESS:

1) Obtain accurate NBA statistics, in .csv format.
2) Upload files to google drive.
3) Use pandas, numpy, matplotlib, and datascience library to analyze and visualize the data.
4) Data cleaning.
5) Sort data to look for trend and pattern. Use data from 4 teams; 2 from western conference and 2 from eastern conference: Lakers, Jazz, Nets, Heat
6) Use direct comparion to check if home court advantage exist.
7) Plot data to look for trend and pattern.
8) Use result to come up with an hypothesis.

References

1) .csv files with all the NBA Statistics from 2004 to 2020 seasons: https://www.kaggle.com/nathanlauga/nba-games
2) Insipiration of using NBA data for poster: Poster by Edress Gul from Summer 2020
3) Source of the average Field Goal percentage graph: https://weaksideawareness.wordpress.com/2012/01/12/how-aggregate-nba-stats-change-through-season/
4) Reference for setting up Github: https://github.com/AshQTan/DH100/blob/main/DH100_demo.ipynb
5) References of datascience functions: http://data8.org/sp21/python-reference.html
6) Coding reference for making the clustered bar charts: https://www.geeksforgeeks.org/create-a-grouped-bar-plot-in-matplotlib/