IS THERE REALLY HOME COURT ADVANTAGE?

DIGHUM 100 - Theory and Methods in Digital Humanities by Dr. Adam G. Anderson || Kenneth Mao || June 19, 2021

Project Description:

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?

HYPOTHESIS:

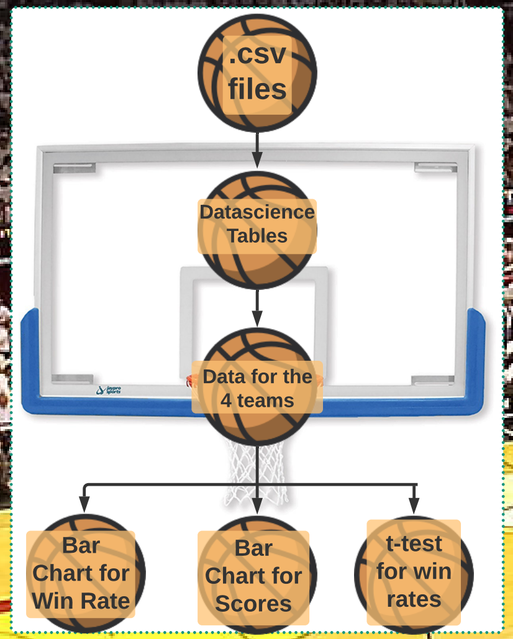
Home court advanrage is real; the teams are more likely to win if it is a home game; and the teams are more likely to score more for home games than away games.

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.

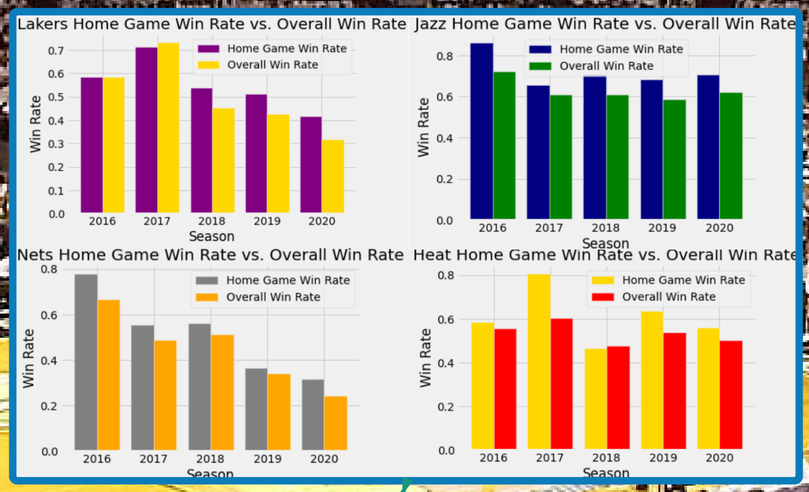
DATA USED EXAMPLE:

TEAM_ID	LEAGUE_ID	SEASON_ID	STANDINGSDATE	CONFERENCE	TEAM	G	W	L	W_PCT	HOME_RECORD	ROAD_RECORD	RETURNTOPLAY
1610612747	0	22020	2021-05-25	West	L.A. Lakers	72	42	30	0.583	21-15	21-15	nan
1610612762	0	22020	2021-05-25	West	Utah	72	52	20	0.722	31-5	21-15	nan
1610612751	0	22020	2021-05-25	East	Brooklyn	72	48	24	0.667	28-8	20-16	nan
1610612748	0	22020	2021-05-25	East	Miami	72	40	32	0.556	21-15	19-17	nan

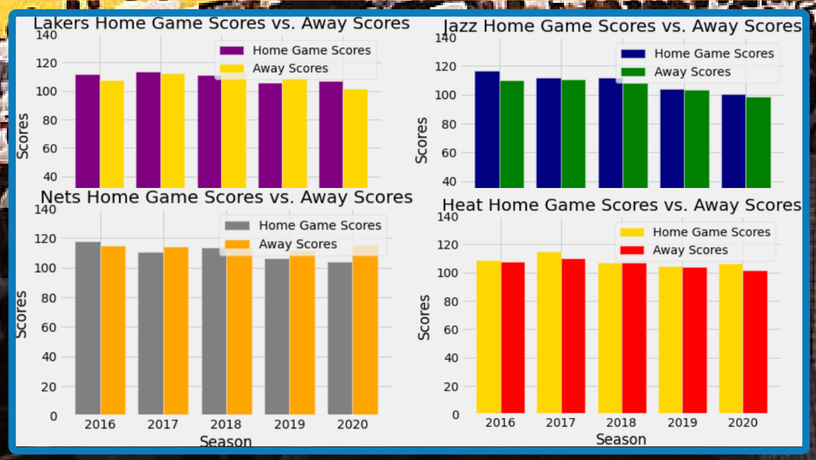


Average Overall Home Game Win Rate For All Teams For the Past 5 years

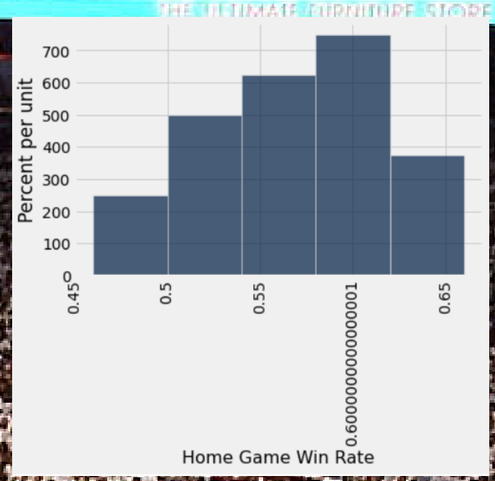
From the line plot, it appeared that the proportion of games won for home games out of all the games won ifor all teams is greater than 50%. Which indicates that if a team wins, it is more likely that it is a home game than an away game.



Home games win rate vs overall win rate for the past 5 seasons for the 4 teams
From this comparison, 18 out of 20 seasons have home game win rate greater than the overall win rate of the same season, which supports that it is more likely for a team to win if it is home game.



Home games scores vs away games scores for the past 5 seasons for the 4 teams
From this comparison, 13 out of 20 seasons have average home games scores greater than the away games scores of the same season. Which doesn't show that there is a significant difference between home game scores and away game scores.



T-test using home game win rates out of all games won: P (Home Game Win Rate < 0.5) = 0.1
Null: Home game win rate is the same as away game win rate. (Win rate = 0.5)
Alternative: Home game win rate is higher than away game win rate. (Win rate > 0.5)
P-cutoff = 0.2
P < P-cutoff: Accept alternative hypothesis; home game win rate is higher than away game win rate.

CONCLUSION:

From the line graph and the clustered bar charts, the home game win rate seems to be higher than away game win rates il was furthered proven by using t-test with a cut-off of 0.2, in which the althernative hypothesis was supported, where the home game win rate is higher than away game win rate. Although more than half of the seasons played have scores higher for home games than away games, there was a significant difference between the score. Therefore, it is to be concluded that home court advantage is real, but it only allows a team to have a higher chance of winning the game instead of scoring more than away games.

PROCESS:

1) Obtain accurate NBA statistics, in .csv format.
2) Upload files to google drive.
3) Use 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) Visualize the data to look for trend and pattern.
7) Use t-test to check if home game win rate is higher than away game win rate.
8) Check result with hypothesis.

TOOLS DESCIPTION:

Numpy: Mathmatical commands such as average, to find the mean, and arange, to get a list of number in increasing order.
Matplotlib: Visualize the comparison between home game win rates and overall win rate, also for home game scores and away game scores in clustered bar chart format.
Datascience: Used to make tables of the .csv files to see the informaiton contained. Also used to select the data needed for computation, using function such as .where.

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/