

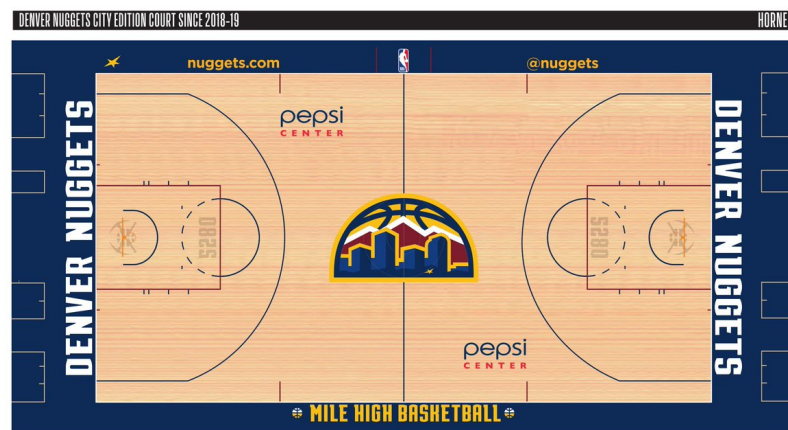
Denver Nuggets Home Court Advantage

By Jonah Chambers

Introduction:

The summer is a time for rest and recovery in the National Basketball Association. The champions are crowned in June and then players and teams have 4 months to rest, organize rosters and prepare for another 8 months of games. The offseason can also be a time for change. Changes in players, coaches, and front office roles dominate the leagues news cycle, but changes occur at every level. New jerseys, new arenas, new color schemes, new mascots, anything to increase effectiveness on the court, and increase revenue.

With all the change and commotion it wouldn't be surprising if one didn't notice the subtle update the Denver Nuggets made to the design of their home court in 2018. Four numbers just under the free throw line reading '5,280'. This number represents the altitude in feet of Ball Arena, home to the Nuggets. The court decal was added to remind opponents that they are playing in the highest altitude arena in the NBA. The higher altitude leads to lower air pressure and in turn lower oxygen in the air, causing increased fatigue for some and can even lead to altitude sickness. A collection of symptoms caused from the body adjusting to altitude that affect the body - especially through the cardiovascular system.



Fans and team media tout this as the greatest home court advantage in sports. While other teams may have rowdy crowds or cavernous arena designs that make it impossible to hear, Denver's advantage is a fact of science that starts imposing itself on the visiting team from the time their planes land.

Still it's fair to question how much the high elevation affects the game's result. Risk of altitude sickness is low below 6,000 feet, and in any case we are talking about some of the best conditioned athletes in the world. It is fair to wonder if this is a true advantage or just a talking point for TV announcers to fill broadcast space.

Thus brings us to the question of this analysis; Does Denver's altitude give it the greatest home court advantage in the NBA? To conduct this analysis, data was analyzed from 20 seasons

worth of regular season games spanning from the 2002-03 season to the 2021-22 season. All data was acquired from the NBA's official website, and in part thanks to the NBA API developed by Swar Patel (github username swar).

Dataset:

Data was acquired from two different modules of the NBA data pages, traditional and advanced. Traditional data is the most basic level of NBA data, where the statistics are largely counting statistics or percentages of counting statistics. This includes things like points, assists, rebounds, as well as shots made and missed. Advanced statistics are usually calculated from traditional statistics, and allow more rich data points for teams and fans to use to understand the game. The two advanced statistics included in this analysis were *net rating* and *pace*. Net rating is the number of points scored minus the number of points allowed per 100 possessions. At this time, it is generally considered the best indicator over a season of how well a team has performed - analytics often point to it being as rich of a statistic as wins. This is because it can be used to see if a team is winning by a lot or just barely winning (the per 100 possessions is to normalize across teams, so fast or slow playing teams aren't misrepresented in effectiveness). Pace is simply the number of times the team possess the ball over the course of the game.

All data featured in this analysis is performed at the team level, meaning each row contains the statistics for one team in one game.

Acquisition

All data acquisition was performed in python and used the following packages : requests, nbaapi, pandas, time, numpy. Traditional data acquisition was considerably easier than the advanced, and could be done with the code below.

```
url = 'https://stats.nba.com/stats/leaguegameolog'
headers= { 'user-agent': 'Mozilla/5.0 (Windows NT 10.0; Win64; x64)\n\nAppleWebKit/537.36 (KHTML, like Gecko) Chrome/96.0.4664.110 Safari/537.36',\n          'Referer': 'https://www.nba.com/' }\n\npayload = {\n    'Counter': '1000',\n    'DateFrom': '',\n    'DateTo': '',\n    'Direction': 'DESC',\n    'LeagueID': '00',\n    'PlayerOrTeam': 'T',\n    'Season': '2012-13',\n    'SeasonType': 'Regular Season',\n    'Sorter': 'DATE'}\n\njsonData = requests.get(url, headers=headers, params=payload).json()\n\nrows = jsonData['resultSets'][0]['rowSet']\ncolumns = jsonData['resultSets'][0]['headers']\n\ndf = pd.DataFrame(rows, columns=columns)
```

The final variable `df`, was a pandas dataframe containing 2 rows for each game, one for each team. The table below includes all the relevant fields included in this dataframe.

| Name | Description |
|-------------------|--|
| Season_ID | 6 digit key unique to each season |
| Team_ID | 9 digit key unique to each team |
| Team_Name | String, containing team city and mascot name |
| Team_Abbreviation | String, 3 letter team name abbreviation |
| Game_ID | 8 digit key unique to each team, generated in chronological order |
| Matchup | String with team abbreviation for both teams as well as a character (@ or v.) designated where the game took place |
| WL | Single character indicating a Win or loss |

Acquisition for the advanced data was done in a different manner due to the way the NBA.com organizes data. While advanced data is more fruitful for understanding the game and growing in popularity, it's still of less interest to the general fan since the numbers are not quite as easy to interpret. Luckily, the open source NBA data community has created the NBA API which can be used to acquire the data from the NBA website that isn't as easily available. Included in this package is a `boxscoreadvancedV2` endpoint that was used to acquire the data needed. The full code can be seen in the coding files linked in the portfolio but it can be described in 3 major steps. First a list of games was generated for the season of interest, then a loop was used to scrape the data for that game (once again 2 rows per game, one for each team), and finally each row was appended to a corresponding row in the master dataframe. To handle failed requests and timeouts, a second loop was required. Prior to running this loop, a list of game ids were generated that were included in the original list but weren't in the master dataframe. Once the entire season was looped over the master dataframe for the season was exported to a .csv file. The following fields were included in this analysis.

| | |
|------------|--|
| Season_ID | 6 digit key unique to each season. |
| Team_ID | 9 digit key unique to each team. |
| Game_ID | 8 digit key unique to each team, generated in chronological order. |
| Net rating | Numerical, difference between teams points scored per 100 possessions, and points allowed per 100 possessions. |
| Pace | Integer, possessions per game. |

In total, advanced and traditional data was acquired for 23,955 games, spanning over 20 seasons.

Data verification, cleaning, and variable creation

The primary mode of data verification was ensuring that every game was included in our dataset and that there were no duplicate values. Duplicate values were dropped if found for a game and team id. This was done in a single line of code. The .shape attribute was used to verify that the appropriate amount of games were included for each season. This was quite easy considering that each team plays 82 games each season. There have been some exceptions including the pandemic, labor negotiation lockouts, and a single game cancellation in 2013. It should be noted that one game was not included in this analysis due to inability to acquire the advanced data. Another form of data verification was performed using the isna() method of pandas dataframes. Finally, all datatypes were declared in the SQL table prior to copying the data from the csv into the SQL table. This would validate that the data conforms to the data requirements for that field. All of these methods revealed no problems in our dataset, and this is not surprising due to the robustness of the nba data modules and the nba.com website.

Only two variables were created; the opponent field, a team_abbreviation for the opponent of the team in the row of interest, and a home_away variable indicating whether this was a home or away game for the row of interest. In both cases this was done using a loop and splicing of the 'matchup' field which was acquired in the traditional dataset.

Analysis and Findings:

With 20 years worth of data saved to csv files all that was left was vertical concatenation by year, and then joining our traditional and advanced tables by game_id and team_id, done in PgAdmin4 (a postgresql management tool). The first question we sought to answer was which team has the highest win percentage at home (note: win percentage is important over raw wins because there has been relocation over the past 20 years, leading to an uneven amount of home games for some franchises). The best way to show this was with a simple bar plot as in Figure 1.

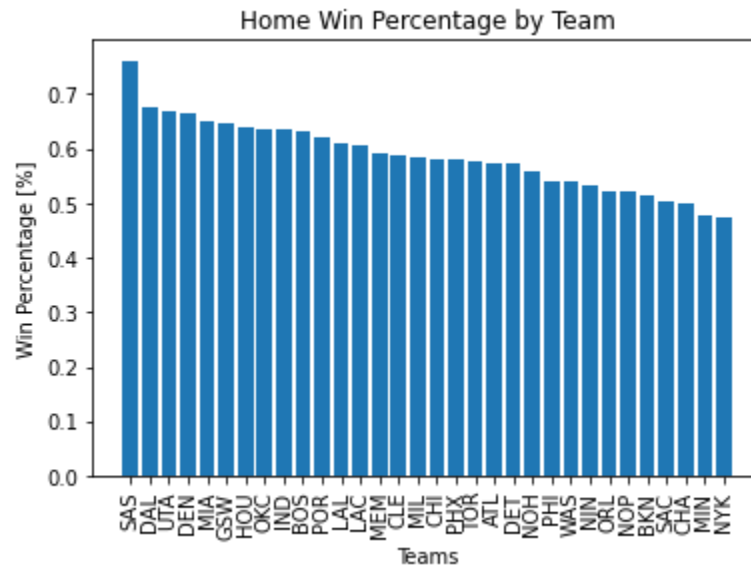


Figure 1: Home Win Percentage by Team

The San Antonio Spurs are the NBA's most successful team at home, winning over 75% of their games. No other team has won more than 67.6%. Denver is the 4th most winning team at home with a HW% of 66%. Interestingly, one of the only teams better than Denver by HW% is the NBA's second highest altitude team Utah (UTA). The Utah Jazz play in Salt Lake City, Utah at an altitude of roughly 4,200 feet. It's also worth noting that only 2 teams, the Minnesota Timberwolves, and the New York Knicks, have won less than half of their home games (48% and 47% respectively). This initial result suggests that while Denver is in the top 87th percentile of HW% they aren't the best team at home, as they like to boast.

To get a better idea of how these home win percentages compared to the teams records overall the total win percentage was plotted against the home win percentage resulting in Figure 2.

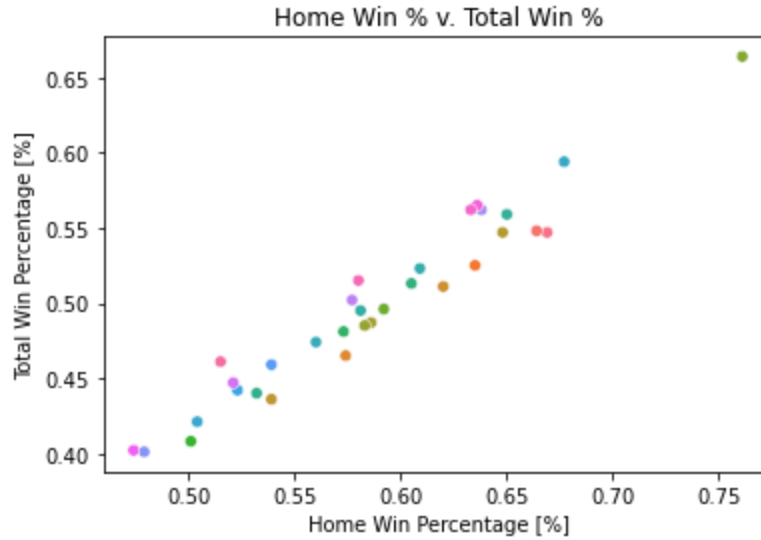


Figure 2 : Home Win % v. Total Win %

Figure 2 shows home win percentage is highly correlated with total win percentage, with a Pearson coefficient of $r = 0.97$. What this tells us is fairly intuitive, the teams that are best at home are the teams that are best overall, and most teams are better at home than they are overall. This means we need a different metric to assess which team has the best home court advantage. This is where the home win percentage differential (HWD) comes into play. HWD subtracts the teams total win percentage (W%) from their home win percentage (HW%) to give us an idea of which teams are the best at home, adjusting for how good those teams are as a whole.

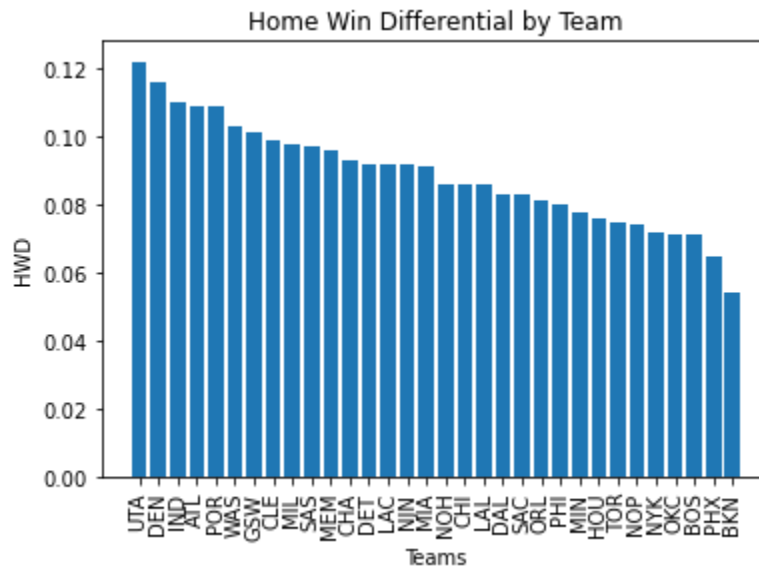


Figure 3: Home Win Differential by Team

Figure 3 shows Utah and Denver at 1 and 2 by HWD, followed by Indiana, Atlanta, Portland and Washington. When we plot these results versus teams total win percentage we see that HWD doesn't have the same correlation with W% as we saw with the raw HW%; Pearson correlation $r=0.18$.

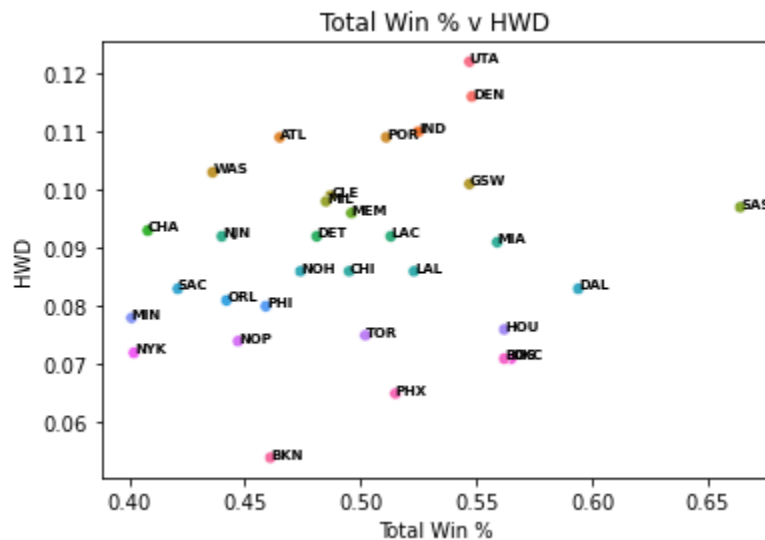


Figure 4 : Total Win % v HWD

This is a pretty strong indication that Denver has one of the best home court advantages in the NBA, with only their high altitude neighbors, the Utah Jazz, leading them in HWD.

Next we wanted to look at how the advanced data supported the findings we have seen in HWD. Wins are a strong way to tell how teams perform, but they are limiting in the amount of information it provides. As mentioned earlier Net Rating is often used as a more nuanced way that a team has performed over the course of a season. In this case we were able to analyze Net Rating over the course of the 20 seasons of data we have for each team.

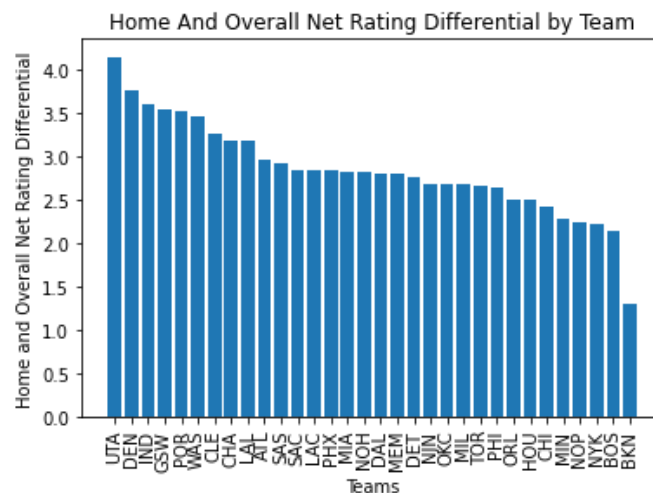


Figure 5: Home and Overall Net Rating Differential by Team

Net rating analysis supports our findings by win percentage differential showing that UTA and DEN are once again the top 2 teams.

While it's not possible to prove through this form of statistical analysis that the altitude is what is creating these teams' advantage, it is certainly interesting that the two highest altitude teams by far are the leaders. Out of curiosity, we decided to plot the differential versus altitude, as well as calculating the pearson coefficient.

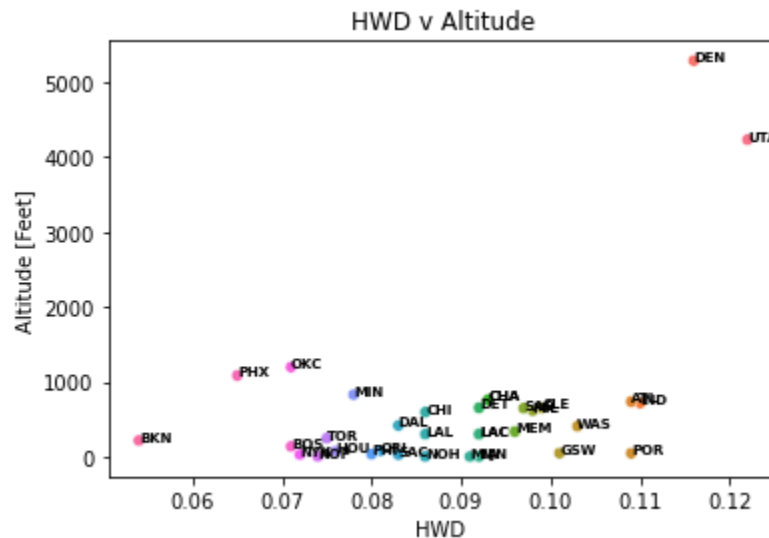


Figure 6: HWD v Altitude

We see a moderate positive correlation with a pearson coefficient of $r = 0.49$. The main limitation of this comparison is the lack of NBA teams in moderate to high altitude cities. This is due to the fact that most NBA teams are located in the US (and Canada's) most populated cities, which due to geographical and historical reasons are often near the sea and or rivers and usually are much closer to sea level.

The final goal of this statistical analysis was to try and figure out a potential mechanism by which the altitude is creating an advantage. Since altitude primarily affects the cardiovascular system, it would make sense to try and determine if opponents were moving slower or traveling less when playing at higher altitudes. While we don't have access to the NBA's miles run per game data, which is behind a paywall, we do have access to the game pace, which as mentioned before, is the number of possessions per game. The goal of this analysis was to show that teams have less possessions per game indicating a slower pace of play, due to the fatigue caused by playing at altitude. In order to do this the average pace of a team per season was calculated. Then the difference between the opponent's pace in the current game, and their season average was calculated, and then this difference was averaged by the home team. The results are shown in Figure 7 below.

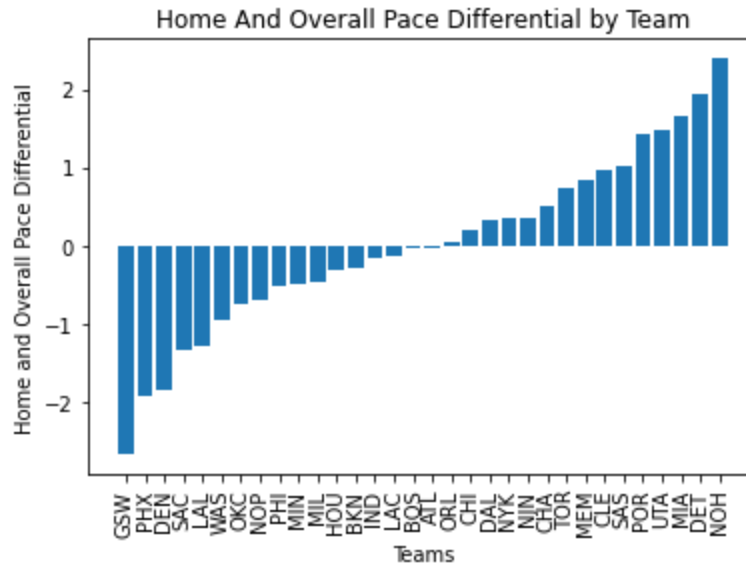


Figure 7: Home and Overall Pace Differential by Team

Teams have played slowest in Golden State, located in Oakland, Ca until 2018 and now in San Francisco, followed by Phoenix, and then Denver. Utah however has the 4th highest differential, on the complete other side of the graph from Denver, and thus we have a pearson correlation coefficient of $r = -0.12$, suggesting a weak negative relationship between altitude and pace. This doesn't come as a complete surprise, seeing as pace is a style of play variable. By this I mean that coaches and teams will try to play with pace to match the players of their team and their specific game plan. Since we can't assume teams will always play their fastest speed possible, and will instead try to control pace to where they are comfortable, we are not able to say that the altitude isn't affecting pace.

Conclusions:

In this analysis we sought to gain an answer to the question of whether Denver has the greatest home court advantage in the NBA, via its league leading altitude. What we found is that by Home Win Percent Differential (described in analysis and findings) and Net Rating, Denver ranks second in the NBA, trailing only the Utah Jazz in both statistics. What is so interesting about this is Utah plays at the second highest elevation in the NBA at 4,300 feet, about a 1000 feet below Denver and about 4 times higher than the 3rd highest team (The Oklahoma City Thunder). Correlation showed a moderate positive linear relationship between altitude and net rating differential.

The major limitation of this study is we have no way to prove through this data that it is the altitude causing Denver and Utah's home court advantage. Indiana performed quite well by the metrics described in this analysis, but they play at only 700 feet in elevation. There are no doubt other variables that contribute to the home court advantage, such as average fan attendance, the design of the arena, the travel time from the home court to wherever the away team had to travel from and plenty of other factors.

As far as the Denvers media claims, they might have to relinquish their title to their mountainous western counterparts the Utah Jazz. While this analysis might not have a huge impact on teams decision making, there are ways in which the higher altitude teams can use this to their advantage. In the NBA playoffs the higher seeded team (the team who won more games in the regular season) gets to play the first two games, and the last game of a 7 games series at home(4 of 3 of the games at home in the series total). In recent years teams have sacrificed some regular season winning in order to give players more rest over the course of the season. Based on this analysis we can advise Denver and Utah to weigh regular season wins higher than other teams might when it comes to planning on resting players during the regular season, since they have these large differentials between home and away win percentages, which can help swing games and series in the playoffs.