Can NAU Basketball Beat the Competition

Jacob Shigeta

2024-03-19

Abstract

This project delves into the performance of both the men's and women's basketball teams at Northern Arizona University over the past decade. Utilizing Python Selenium, data encompassing team statistics, player averages, head coaches, game statistics, and game outcomes were collected from the NAU Athletics website. The primary objective is to show how well the teams did over the past ten years as well as any possible changes to the teams making them perform better or worse than the previous seasons. Some of these changes include identifying the star player and the coach at the time as longitudinal data to highlight any significant shifts in team performance. This project aims to facilitate a deeper understanding of NAU Basketball performance dynamics.

Introduction

The data that was collected involves the NAU men's and women's basketball teams, more specifically, the stats for the season, the coach for that season, the player's stats, and the game stats. I have attached part of each data frame below.

Table 1: Transposed Men's and Women's Team Stats

	1	2	3	4	5	6
Team.Stats	14-18, 7-11	12-23, 5-13	9-23, 5-15	6-16, 4-10	16-14, 10-10	10-21, 8-12
Team	Mens	Mens	Mens	Mens	Mens	Mens
Year	2024	2023	2022	2021	2020	2019
Total.Points	2205	2581	2161	1390	2128	2262
Points.Per.Game	68.9	73.7	67.5	63.2	70.9	73.0
Scoring.Margin	-8.0	-1.0	-5.8	-11.3	1.5	-4.5
FG..Made.Attempted	800-1830	930-2123	770-1845	505-1197	770-1729	785-1819
FG..Percentage	0.437	0.438	0.417	0.422	0.445	0.432
FG..Per.Game	25.0	26.6	24.1	23.0	25.7	25.3
X3PT..Made.Attempted	201-646	295-815	277-773	139-397	216-600	259-749
X3PT..Percentage	0.311	0.362	0.358	0.350	0.360	0.346
X3PT..Per.Game	6.3	8.4	8.7	6.3	7.2	8.4
FT..Made.Attempted	404-579	426-576	344-518	241-364	372-531	433-614
FT..Percentage	0.698	0.740	0.664	0.662	0.701	0.705
FT..Per.Game	12.6	12.2	10.8	11.0	12.4	14.0
Total	7258	10368	9521	0	10215	12908
Per.Game	12-605	14-741	15-635	8-0	15-681	12-1,076
Assist.Turnover.Ratio	1.1	1.3	0.9	0.9	1.2	1.0
Points.Off.Turnovers	11.5	11.6	10.2	12.5	14.6	12.3
Head.Coach	Shane Burcar	Shane Burcar	Shane Burcar	Shane Burcar	Shane Burcar	Jack Murphy

Table 2: Men's and Women's Game Stats

X	Team	Opponent	Date	Score	W.L	FGM.A	PCT	X3FG.A	FTM.A	OFF	DEF	TOT	AVG
2024	Mens	at UConn	11/06/23	52-95	L	20-59	0.563	3-15	9-16	9	18	27	27.0
2024	Mens	at Grand Canyon	11/12/23	55-89	L	22-56	0.545	5-18	6-11	5	19	24	25.5
2024	Mens	at Seattle U	11/15/23	62-60	W	25-59	0.615	4-23	8-13	6	26	32	27.7
2024	Mens	vs Purdue Fort Wayne	11/17/23	67-77	L	27-59	0.500	8-28	5-10	7	29	36	29.8
2024	Mens	vs VMI	11/18/23	78-69	W	28-60	0.593	6-23	16-27	9	22	31	30.0
2024	Mens	at Hawaii	11/21/23	61-70	L	23-55	0.500	6-17	9-18	10	24	34	30.7

Table 3: Men's and Women's Player Stats

X	Team	X.	Player	GP	MIN	FG.	X3PT.	FT.	OREB	DREB	REB	AST	STL	BLK	PTS
2024	Mens	24	McLaughlin, Trent	32	34.8	0.402	0.329	0.769	0.4	2.8	3.2	2.0	1.4	0.2	16.7
2024	Mens	7	Basham, Carson	32	24.3	0.585	0.000	0.570	1.8	4.8	6.6	1.2	0.7	1.6	12.3
2024	Mens	4	Fort, Oakland	32	30.5	0.374	0.308	0.797	0.5	3.2	3.7	3.0	0.8	0.2	10.2
2024	Mens	3	Jackson, Jayden	32	29.2	0.502	0.212	0.752	1.6	2.7	4.3	2.0	0.8	0.2	10.0
2024	Mens	32	Lloyd, Liam	30	29.3	0.351	0.273	0.550	0.9	3.4	4.3	3.4	0.8	0.2	6.3
2024	Mens	12	Campisano, Diego	32	18.5	0.490	0.421	0.864	0.7	1.7	2.4	0.5	0.4	0.0	6.2

I am very interested in this data because, since I am a part of the NAU Marching and Pep Band and have been for my entire time here at NAU, we go to a lot of the sports events, mainly Football, Volleyball, and Basketball. Therefore, I get to see most, if not all, of the home games for each team for each season so I see how they do on the field/court. I have really noticed that, for basketball, the men's team usually does worse than the women's team and that has been consistent for the past three years where I have actually been to games. This got me wondering if this trend has been accurate throughout the years at NAU, or if it just started to happen for a reason that I can hopefully prove.

Methods

To scrape the data, I had to learn how to use Python Selenium to do the scraping and Pandas to save it to a data frame and export it. This was quite the feat, especially starting off. Thankfully I was not the only

one who was doing it so it made me feel better about struggling on it for as long as I did. Once I learned it and was able to successfully scrape one of the pages, it became very easy to scrape from the rest of them.

One of the main difficulties I had was figuring out how to install the programs and get them to run properly. This was because I had multiple versions of Python installed in different locations on my computer. It was saving it to the default one, but not the one in my editor. To fix this, I had to find the location where they were being downloaded, the location where they needed to be, and just move them from one location to another. This was tricky for me because it took me a while to figure out where it was being downloaded and even after figuring it out and deleting the older version of Python (had to reinstall it because the programs could not install properly through my editor), it was hard to remember where to go to find those files again.

Another difficulty I encountered while gathering the data was figuring out how to make Pandas concatenate data frames together and make it look neat. The way I went about this was to save the new lines to a new data frame, concatenate it to the main one, and delete the newer data frame. When it came to combining the larger data frames, I just saved them separately and combined them in R, due to it being more familiar to me and not taking as long between attempts.

The last difficulty I will talk about was issues with the scraping itself, mainly figuring out how to loop through the website easily to gather player data. I was running into an issue where the program will just break out of no where and I could never understand why until I looked deeper into it and it turns out that some of the players had a different value for their lines than others. This can be translated to other aspects of the other pages I scraped from as well, but that is the one that really caught me up. To solve this, I had to make a separate check to see if that issue happened and, if so, grab a different line in the website.

With the data I collected, I am planning to compare the amount of wins and losses between both teams, the total attendance per season for home games and the total points scored per season and who coached them, as well as something to do with the players that I have not decided yet but it will most likely be comparing the star players for each season (who scored the most points). I have attached below the wins and losses between both teams as a table. This will be the only table and the rest will be graphs.

Table 4: Men's and Women's Wins and Losses

	Mens	Womens	Sum
Loss	190	247	437
Win	86	177	263
Sum	276	424	700

As for the other three, I am planning to turn the attendance and amount of points scored per season into a time series graph. As for the player data, if I do go with the star player idea, I might compare them to each other based on gender, grab the best one for each, and then put those two against each other to see who was the best player NAU has ever had in the data I have and their stats.

Expected Outcomes

I am expected to find out which of the two basketball teams here at NAU really does better and the reason behind it. I am also expected to find some interesting facts between the teams or about each team separately. This can include if a certain player made a big impact on the performance of the team that year, if a certain coach made the team do better or worse, or if the number of people in attendance made a difference in the team's performance.