Project 1: Sports Analytics

Group 2 Analytics Avengers

Kayla Rivera
Payal Patel
Patrick Moon

Questions To Answer

- Does distance traveled for away games impact the winning percentage of NBA teams, more specifically, does the effect differ between teams in separate regions, such as the Chicago Bulls and the LA Lakers?
    - Motivation: Travel fatigue and jet lag can potentially affect the performance of players. Understanding if and how it impacts the team can lead to better scheduling and prep for away games/traveling.
- Do starters' individual stats compared to team scoring affect game impact?
    - Motivation: Understanding a starter's statistics (points scored, rebounds made, etc. vs the team's average can shed a light on the offensive dynamics and strategies of the team. This can help to evaluate if a balanced team offense or a star-centric approach is more effective.

Where and how we found the data we used to answer these questions

- **Basketball Reference**: https://www.basketball-reference.com/
  - Chicago Bulls & LA Lakers 2020 - 2022 seasons results
  - Player stats from 2022 season
- **NBA team stadium locations**: https://geojango.com/pages/list-of-nba-teams
- **GeoAPIfy** for stadium coordinates
- **Haversine** formula to calculate distance between coordinates

Regular Season

Share & Export ▼ Glossary

G	Date	Start (ET)			Opponent			Tm	Opp	W	L	Streak	Notes
1	Wed, Oct 19, 2022	7:30p	Box Score	@	Miami Heat	W		116	108	1	0	W 1	
2	Fri, Oct 21, 2022	7:00p	Box Score	@	Washington Wizards	L		100	102	1	1	L 1	
3	Sat, Oct 22, 2022	8:00p	Box Score		Cleveland Cavaliers	L		96	128	1	2	L 2	
4	Mon, Oct 24, 2022	8:00p	Box Score		Boston Celtics	W		120	102	2	2	W 1	
5	Wed, Oct 26, 2022	8:00p	Box Score		Indiana Pacers	W		124	109	3	2	W 2	
6	Fri, Oct 28, 2022	8:30p	Box Score	@	San Antonio Spurs	L		124	129	3	3	L 1	
7	Sat, Oct 29, 2022	8:00p	Box Score		Philadelphia 76ers	L		109	114	3	4	L 2	
8	Tue, Nov 1, 2022	7:30p	Box Score	@	Brooklyn Nets	W		108	99	4	4	W 1	
9	Wed, Nov 2, 2022	8:00p	Box Score		Charlotte Hornets	W		106	88	5	4	W 2	
10	Fri, Nov 4, 2022	7:30p	Box Score	@	Boston Celtics	L		119	123	5	5	L 1	
11	Sun, Nov 6, 2022	6:00p	Box Score	@	Toronto Raptors	L		104	113	5	6	L 2	
12	Mon, Nov 7, 2022	8:45p	Box Score		Toronto Raptors	W		111	97	6	6	W 1	
13	Wed, Nov 9, 2022	8:00p	Box Score		New Orleans Pelicans	L		111	115	6	7	L 1	
14	Sun, Nov 13, 2022	8:00p	Box Score		Denver Nuggets	L		103	126	6	8	L 2	
15	Wed, Nov 16, 2022	8:00p	Box Score	@	New Orleans Pelicans	L		110	124	6	9	L 3	
16	Fri, Nov 18, 2022	8:00p	Box Score		Orlando Magic	L		107	108	6	10	L 4	
17	Mon, Nov 21, 2022	8:00p	Box Score		Boston Celtics	W		121	107	7	10	W 1	
18	Wed, Nov 23, 2022	8:00p	Box Score	@	Milwaukee Bucks	W		118	113	8	10	W 2	
19	Fri, Nov 25, 2022	8:00p	Box Score	@	Oklahoma City Thunder	L	OT	119	123	8	11	L 1	
20	Mon, Nov 28, 2022	9:00p	Box Score	@	Utah Jazz	W		114	107	9	11	W 1	
G	Date	Start (ET)			Opponent			Tm	Opp	W	L	Streak	Notes
21	Wed, Nov 30, 2022	9:00p	Box Score	@	Phoenix Suns	L		113	132	9	12	L 1	
22	Fri, Dec 2, 2022	10:00p	Box Score	@	Golden State Warriors	L		111	119	9	13	L 2	
23	Sun, Dec 4, 2022	6:00p	Box Score	@	Sacramento Kings	L		101	110	9	14	L 3	
24	Wed, Dec 7, 2022	8:00p	Box Score		Washington Wizards	W		115	111	10	14	W 1	
25	Sat, Dec 10, 2022	8:00p	Box Score		Dallas Mavericks	W		144	115	11	14	W 2	
26	Sun, Dec 11, 2022	6:30p	Box Score	@	Atlanta Hawks	L	OT	122	123	11	15	L 1	
27	Wed, Dec 14, 2022	7:30p	Box Score		New York Knicks	L	OT	120	128	11	16	L 2	

2022-23 Regular Season

Share & Export ▼ Glossary

Rk	G	Date	Age	Tm		Opp		GS	MP	FG	FGA	FG%	3P	3PA	3P%	FT	FTA	FT%	ORB	DRB	TRB	AST	STL	BLK	TOV	PF	PTS	GmSc	+/-
1	1	2022-10-19	33-073	CHI	@	MIA	W (+8)	1	36:08	14	22	.636	2	3	.667	7	11	.636	1	5	6	9	2	1	1	3	37	34.6	+13
2	2	2022-10-21	33-075	CHI	@	WAS	L (-2)	1	36:47	11	23	.478	0	2	.000	10	11	.909	1	5	6	6	1	0	3	1	32	23.9	+1
3	3	2022-10-22	33-076	CHI		CLE	L (-32)	1	32:31	3	9	.333	0	0		7	7	1.000	0	2	2	1	2	0	0	2	13	10.4	-6
4	4	2022-10-24	33-078	CHI		BOS	W (+18)	1	32:45	10	17	.588	0	0		5	5	1.000	0	5	5	5	1	0	3	1	25	19.7	+1
5	5	2022-10-26	33-080	CHI		IND	W (+15)	1	32:17	6	14	.429	0	1	.000	5	5	1.000	0	1	1	6	0	0	3	5	17	9.1	-3
6	6	2022-10-28	33-082	CHI	@	SAS	L (-5)	1	35:33	11	20	.550	0	1	.000	11	12	.917	0	1	1	1	1	0	0	1	33	24.6	-15
7	7	2022-10-29	33-083	CHI		PHI	L (-5)	1	34:15	7	12	.583	1	2	.500	9	10	.900	0	3	3	4	0	0	4	2	24	16.9	+2
8	8	2022-11-01	33-086	CHI	@	BRK	W (+9)	1	32:45	8	21	.381	1	2	.500	3	3	1.000	0	4	4	1	3	0	1	2	20	11.6	-1
9	9	2022-11-02	33-087	CHI		CHO	W (+18)	1	28:17	2	11	.182	0	1	.000	5	5	1.000	2	6	8	5	1	1	5	2	9	4.7	-9
10	10	2022-11-04	33-089	CHI	@	BOS	L (-4)	1	35:52	13	23	.565	0	3	.000	20	22	.909	0	3	3	5	2	1	2	3	46	38.2	+10
11	11	2022-11-06	33-091	CHI	@	TOR	L (-9)	1	37:29	7	9	.778	0	1	.000	6	6	1.000	1	4	5	2	1	1	5	2	20	15.7	-8
12	12	2022-11-07	33-092	CHI		TOR	W (+14)	1	35:38	2	6	.333	0	1	.000	5	6	.833	0	6	6	7	1	0	2	4	9	9.3	+7
13	13	2022-11-09	33-094	CHI		NOP	L (-4)	1	36:38	14	26	.538	0	2	.000	5	6	.833	0	3	3	3	1	1	3	5	33	19.7	-3
14	14	2022-11-13	33-098	CHI		DEN	L (-23)	1	25:40	6	11	.545	0	1	.000	4	6	.667	0	2	2	4	0	1	0	3	16	12.8	-26
15	15	2022-11-16	33-101	CHI	@	NOP	L (-14)	1	30:53	11	15	.733	0	1	.000	6	7	.857	1	3	4	7	0	0	0	3	28	26.8	-14
16	16	2022-11-18	33-103	CHI		ORL	L (-1)	1	43:38	16	30	.533	1	2	.500	8	9	.889	1	3	4	2	0	0	2	2	41	26.2	+13
17	17	2022-11-21	33-106	CHI		BOS	W (+14)	1	35:03	11	24	.458	3	3	1.000	3	3	1.000	0	8	8	4	1	0	2	3	28	18.6	+20
18	18	2022-11-23	33-108	CHI	@	MIL	W (+5)	1	37:33	14	24	.583	2	3	.667	6	7	.857	0	4	4	8	1	2	2	0	36	31.6	+15
19	19	2022-11-25	33-110	CHI	@	OKC	L (-4)	1	42:16	12	27	.444	0	1	.000	6	8	.750	0	4	4	6	2	0	2	5	30	18.5	+6
20	20	2022-11-28	33-113	CHI	@	UTA	W (+7)	1	35:40	9	20	.450	0	1	.000	8	9	.889	2	2	4	6	1	0	0	4	26	20.8	+2
Rk	G	Date	Age	Tm		Opp		GS	MP	FG	FGA	FG%	3P	3PA	3P%	FT	FTA	FT%	ORB	DRB	TRB	AST	STL	BLK	TOV	PF	PTS	GmSc	+/-
21	21	2022-11-30	33-115	CHI	@	PHO	L (-19)	1	31:47	11	17	.647	0	1	.000	7	8	.875	0	7	7	4	0	1	2	3	29	23.5	-8
22	22	2022-12-02	33-117	CHI	@	GSW	L (-8)	1	35:45	4	15	.267	0	0		8	9	.889	1	5	6	7	0	0	1	4	16	11.2	-12
23	23	2022-12-04	33-119	CHI	@	SAC	L (-9)	1	35:24	6	18	.333	0	0		6	6	1.000	1	4	5	4	0	0	2	3	18	9.3	-8
24	24	2022-12-07	33-122	CHI		WAS	W (+4)	1	37:15	11	21	.524	0	1	.000	5	6	.833	0	7	7	4	0	1	5	3	27	15.7	+5
25	25	2022-12-10	33-125	CHI		DAL	W (+29)	1	31:34	9	13	.692	1	2	.500	9	9	1.000	0	9	9	5	0	0	2	1	28	26.3	+38
26	26	2022-12-11	33-126	CHI	@	ATL	L (-1)	1	45:19	10	21	.476	0	0		14	15	.933	1	12	13	8	0	0	1	2	34	31.0	+3
27	27	2022-12-14	33-129	CHI		NYK	L (-8)	1	39:48	8	18	.444	0	2	.000	16	17	.941	0	5	5	2	0	2	2	3	32	23.3	-9
28	28	2022-12-16	33-131	CHI		NYK	L (-23)	1	30:46	6	14	.429	0	1	.000	2	2	1.000	0	4	4	4	1	0	4	3	14	6.4	-23
29	29	2022-12-18	33-133	CHI	@	MIN	L (-24)	1	38:52	10	19	.526	2	4	.500	7	9	.778	1	3	4	6	2	0	1	2	29	24.9	-5
30	30	2022-12-20	33-135	CHI	@	MIA	W (+10)	1	39:31	9	14	.643	0	1	.000	6	6	1.000	1	4	5	5	0	0	2	2	24	20.4	+11
31	31	2022-12-21	33-136	CHI	@	ATL	W (+2)	1	37:50	12	23	.522	0	1	.000	4	6	.667	2	4	6	5	1	0	1	5	28	20.0	+5
32	32	2022-12-23	33-138	CHI	@	NYK	W (+1)	1	36:58	9	21	.429	1	3	.333	6	7	.857	0	7	7	10	2	1	0	5	25	23.3	+2
33	33	2022-12-26	33-141	CHI		HOU	L (-15)	1	40:57	11	18	.611	0	0		9	10	.900	1	4	5	9	1	0	2	3	31	28.4	-17

Data exploration and cleanup process

- Tried to create hypotheses on impacts that individual players have on a team's performance & game outcome
  - Explored what percentage of total points were scored by certain star players on each team in the 2022 season
  - Identify scoring leaders (DeMar DeRozan & LeBron James)
    - Explore if there are any correlations between scoring leaders and the number of games that the team won in the 2022 season
  - Identify rebounding players (Nikola Vucevic & Anthony Davis)
    - Explore if there are any correlations between rebounding and the number of games that the team won in the 2022 season
  - How many points are the scoring leaders scoring vs. average points per player in a game/season
  - Summarized the data using histograms, pie charts, and scatter plots
- Clean-up process
  - Excel functions to add additional columns to source file
  - Removed data where the players didn't play in that particular game
  - Renamed and dropped columns as needed
  - Added error message for when a city was not found in GeoAPIfy

Explore if there are any correlations between scoring leaders and the number of games that the team won in the 2022 season

```
demar_games_played = demar_stats_2022.loc[(demar_stats_2022['GS'] != 'Inactive') & (demar_stats_2022['GS'] != 'Did Not Dress')]
demar_games_played = demar_games_played[['Year', 'Game', 'Team', 'Opponent', 'Result', 'Tm', 'Opp', 'Player', 'GS', 'PTS']]
demar_games_played['PTS'] = demar_games_played['PTS'].astype(int)
demar_games_played['Score Difference'] = abs(demar_games_played['Tm'] - demar_games_played['Opp'])
demar_games_played_win = demar_games_played.loc[demar_games_played['Result'] == 'W']
demar_games_played_loss = demar_games_played.loc[demar_games_played['Result'] == 'L']
demar_games_played_win_avg_diff = demar_games_played_win['Score Difference'].mean()
demar_games_played_loss_avg_diff = demar_games_played_loss['Score Difference'].mean()
demar_games_played_avg_score = demar_games_played['Tm'].mean()

print(demar_games_played_win_avg_diff, demar_games_played_loss_avg_diff, demar_games_played_avg_score)
```

```
13.64864864864865 10.702702702702704 113.45945945945945
```

```
bins = [0,5,10,15,20,25,30]

plt.hist(demar_games_played_win['PTS'], bins, histtype='bar', rwidth=1, edgecolor = "black")
plt.yticks(np.arange(0, 18, 2))
plt.title('# of points scored by DeMar DeRozan vs # of games the Bulls won')
plt.xlabel('# of points scored by DeMar DeRozan')
plt.ylabel('# of games the Bulls won')
```

How many points are the scoring leaders scoring vs. average points per player in a game/season

```
number_of_bulls_players = 15
demar_pts_per_game = demar_games_played[['Game', 'Result', 'Tm', 'PTS']]
demar_pts_per_game['Avg pts per player'] = (demar_pts_per_game['Tm'] - demar_pts_per_game['PTS']) \
    / ((number_of_bulls_players) - 1)
demar_pts_per_game
plt.scatter(demar_pts_per_game['Game'], demar_pts_per_game['PTS'])
plt.scatter(demar_pts_per_game['Game'], demar_pts_per_game['Avg pts per player'])
plt.xticks(np.arange(0, 84, 2))
plt.yticks(np.arange(0, 55, 5))
plt.title('# of points scored by Demar DeRozan vs Average # of points scored by the rest of the team per game')
plt.xlabel('Game #')
plt.ylabel('# of points')
plt.legend(["DeMar DeRozan" , "Rest of Team"])
plt.show()
#plt.xlabel('Field Goal %')
#plt.ylabel('# of points the Bulls won by')
#plt.title('DeMar DeRozan Field Goal % vs. # of points the Bulls won by')
```

Clean-up process

	Year	Team	Game	Date	Start (ET)	Away gam	Opponent	Result	Tm	Opp	W	L	Streak	Notes	Home or Away
2	2022	Chicago Bu	1	Wed, Oct	7:30p	@	Miami He	W	116	108	1	0	W 1		=IF(ISBLANK(F2),"Home","Away")
3	2022	Chicago Bu	2	Fri, Oct 21	7:00p	@	Washingt	L	100	102	1	1	L 1		A IF(logical_test, [value_if_true], [value_if_false])
4	2022	Chicago Bu	3	Sat, Oct 22	8:00p		Cleveland	L	96	128	1	2	L 2		Home
5	2022	Chicago Bu	4	Mon, Oct :	8:00p		Boston Ce	W	120	102	2	2	W 1		Home

```
demar_games_played = demar_stats_2022.loc[(demar_stats_2022['GS'] != 'Inactive') & (demar_stats_2022['GS'] != 'Did Not Dress')]
```

```
#Drop duplicate columns
final_player_stats = final_player_stats.drop(['Date_y', 'Tm_y', 'Opp_y'], axis=1)
```

```
#Import geoapify
import requests
def get_coordinates(city_state):
    API_KEY = '141b8691f629447f822359b7dafbbb82'
    url = f'https://api.geoapify.com/v1/geocode/search?text={city_state}&apiKey={API_KEY}'
    response = requests.get(url)
    if response.status_code == 200:
        return response.json()['features'][0]['geometry']['coordinates']
    else:
        print(f"Error geocoding {city_state}: {response.text}")
        return None
#Get coordinates for cities each team travelled to
chicago_travelled_coordinates = [get_coordinates(city) for city in chicago_travelled_cities]
la_travelled_coordinates = [get_coordinates(city) for city in la_travelled_cities]
```

The analysis process

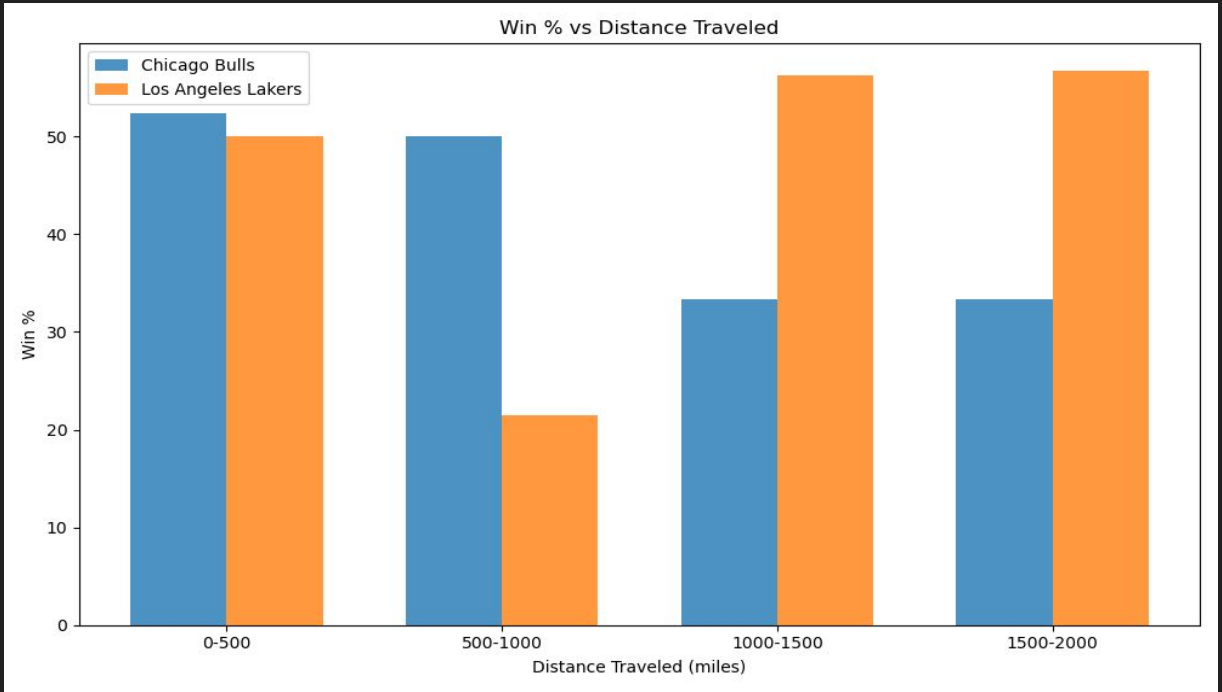
- - We merged 2 dataframes and made it into a combined data
  - We were able to see the team, year, date, start of game, home or away, who they were playing against arena, arena location, seat capacity, opening year
  - Defined our variables
- Defined our variables based on the data we needed and from there we used reset_index(drop=True) to clean out our data and the old index values were not added as a new column
- .append was used to combine the cities with its corresponding state
- created a [] (empty list) for each team to store arena location that they travelled to and iterates or loops through unique arena locations for each team
- Imported geoapify and used the geoapify key to get coordinates for the cities each team traveled to and looped through
- get_coordinates() takes the city and state as input and requests geoampify API to retrieve coordinates and return coordinates if successful, if not it prints as error
  - used import haversine to calculate distances between 2 points specifically latitude and longitude
- NumPy and matplotlib is used to create a bar chart to visualize between winning percentage and distances travelled by the two teams
  - bin represents a range of distances in miles
  - plt.show() displays the plot and its data by creating variables for the y and x axis and setting xticks for the graph
- plt.rcParam sets the matplotlib global parameter and when “figure.autolayout” is set to true then it automatically adjusts the layout of figures to fit the extra spaces

The analysis process - cont’d

- - scipy.stats was used to provide statistical data and create a scatter plot and calculate correlation coefficient and linregress is used to calculate the linear regression
- - plt.hist was used to create a histogram with bins, = [], plt.xlabel() and plt.ylabel(). These were given values to create the visualization of the histogram
    - - plt.scatter() for scatter plot
    - -plt.pie() for pie chart

Distance Traveled - Conclusions

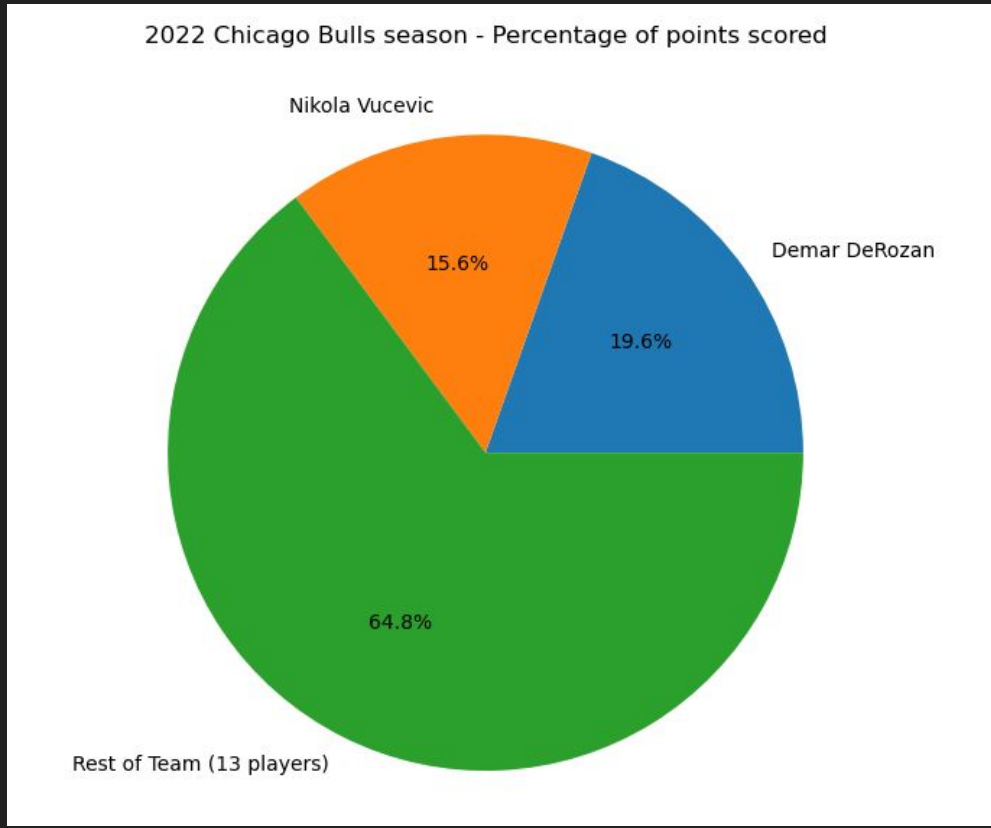
The Bulls exhibited a trend where the farther they traveled for away games, the lower their winning percentage became. Conversely, the Lakers winning percentage seemed to increase with longer travel distances.



Chicago Bulls - Player Stats analysis - Conclusions

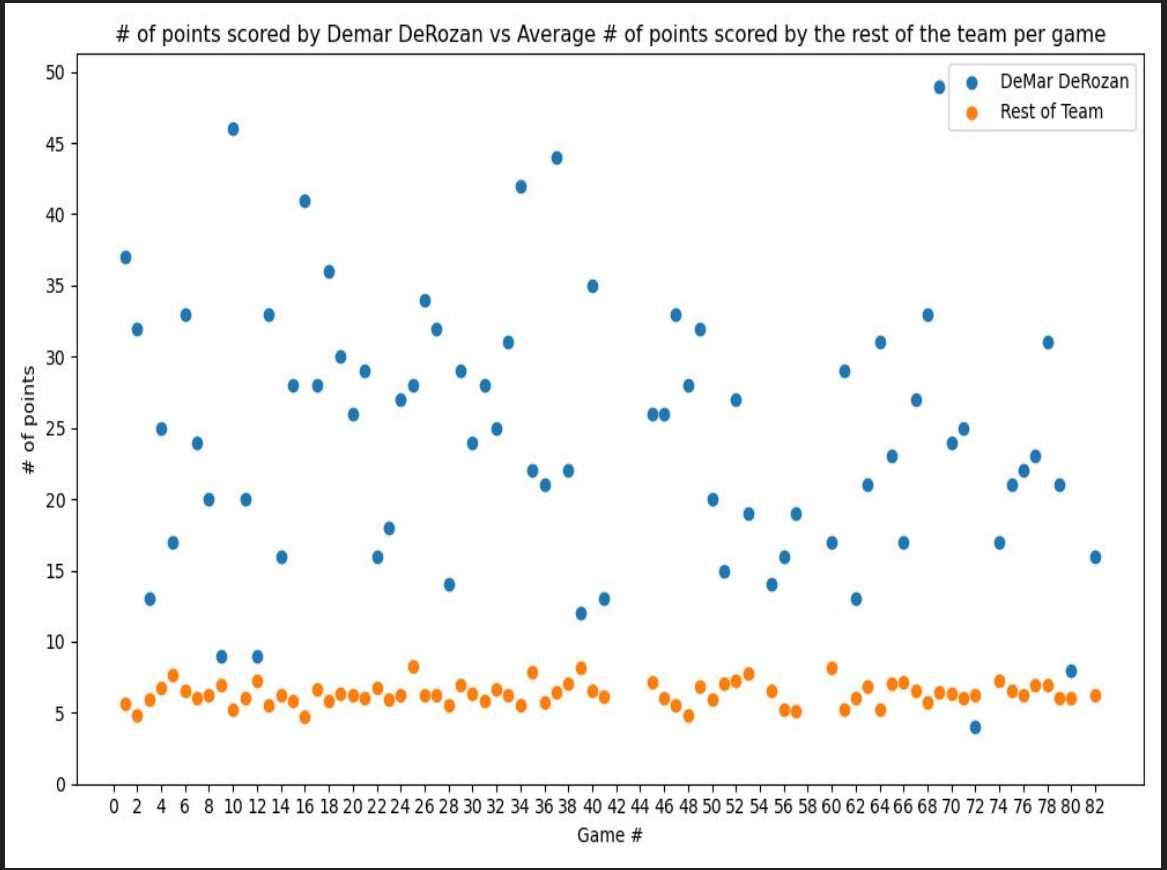
- In the 2022 season, the Chicago Bulls relied on DeMar DeRozan and Nikola Vucevic for their scoring abilities with the 2 players accounting for over 35% of all the points that the team scored.

	Team	Game	Tm	Nikola Vucevic Points	DeMar DeRozan Points	Rest of Team
0	Chicago Bulls	1	116	15	37	64
2	Chicago Bulls	2	100	24	32	44
3	Chicago Bulls	3	96	16	13	67
4	Chicago Bulls	4	120	18	25	77
5	Chicago Bulls	5	124	14	17	93
...	...	...	...	...	...	...
68	Chicago Bulls	77	121	21	23	77
38	Chicago Bulls	78	128	13	31	84
39	Chicago Bulls	79	105	19	21	65
50	Chicago Bulls	80	92	21	8	63
40	Chicago Bulls	82	103	8	16	79



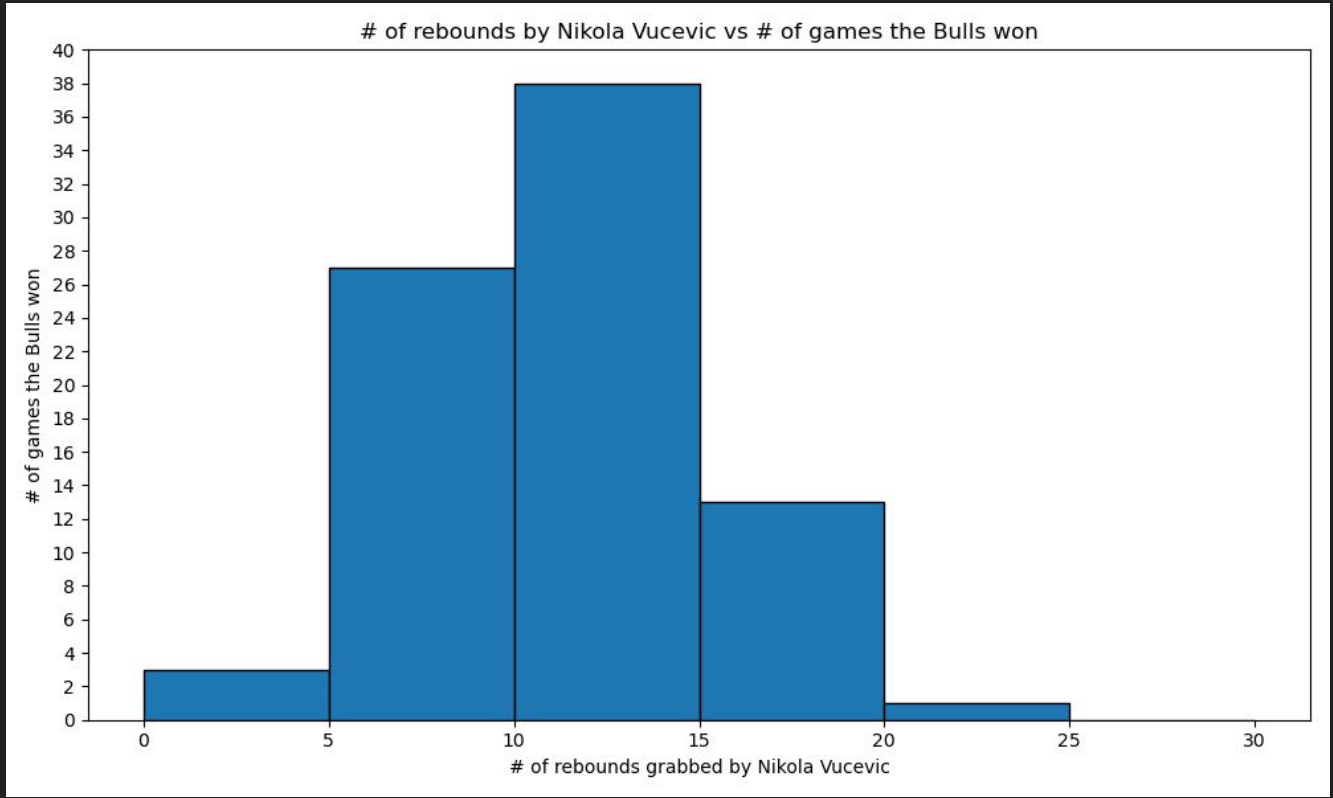
Chicago Bulls - Player Stats analysis - Conclusions - cont'd

- DeMar DeRozan led the team in scoring for majority of the games, scoring well above the average points per player in each game.



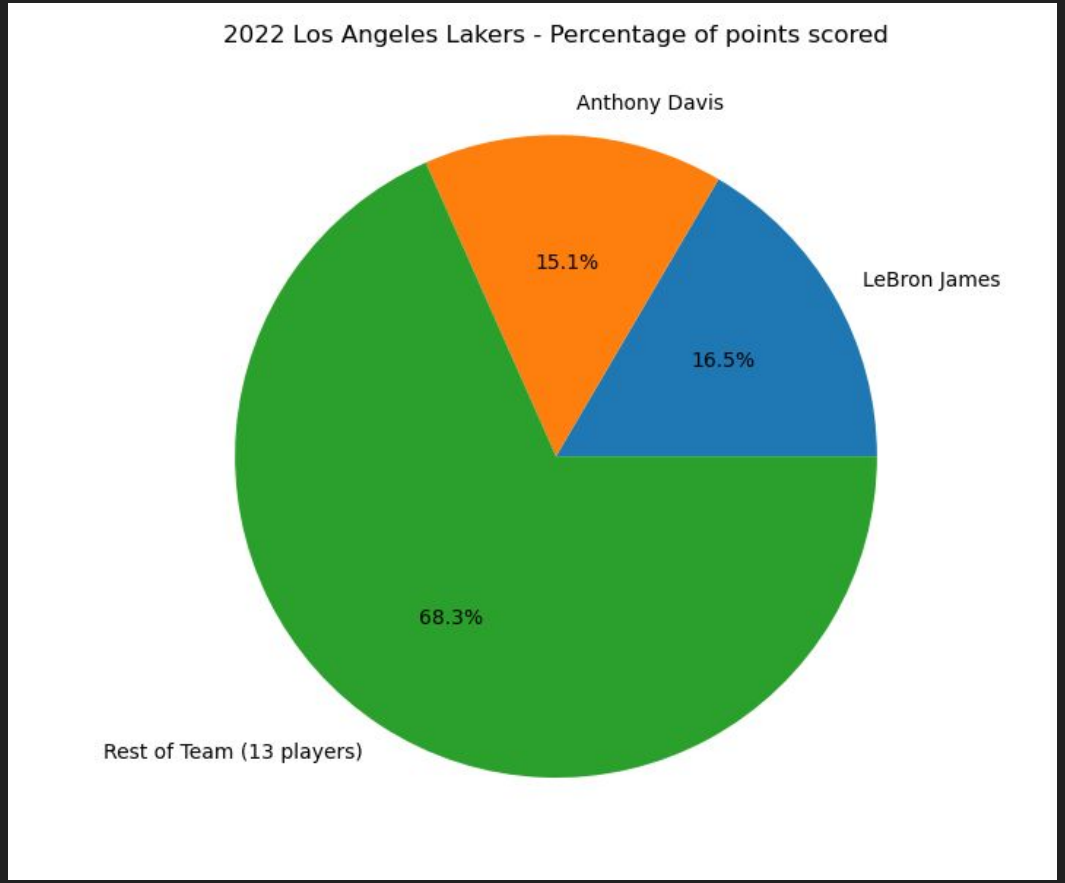
Chicago Bulls - Player Stats analysis - Conclusions - cont’d

- The Chicago Bulls won the most games in the 2022 season when Nikola Vucevic was grabbing between 10 - 15 rebounds per game. The more rebounds he grabbed up to 15 rebounds/game, the more games the Bulls won.



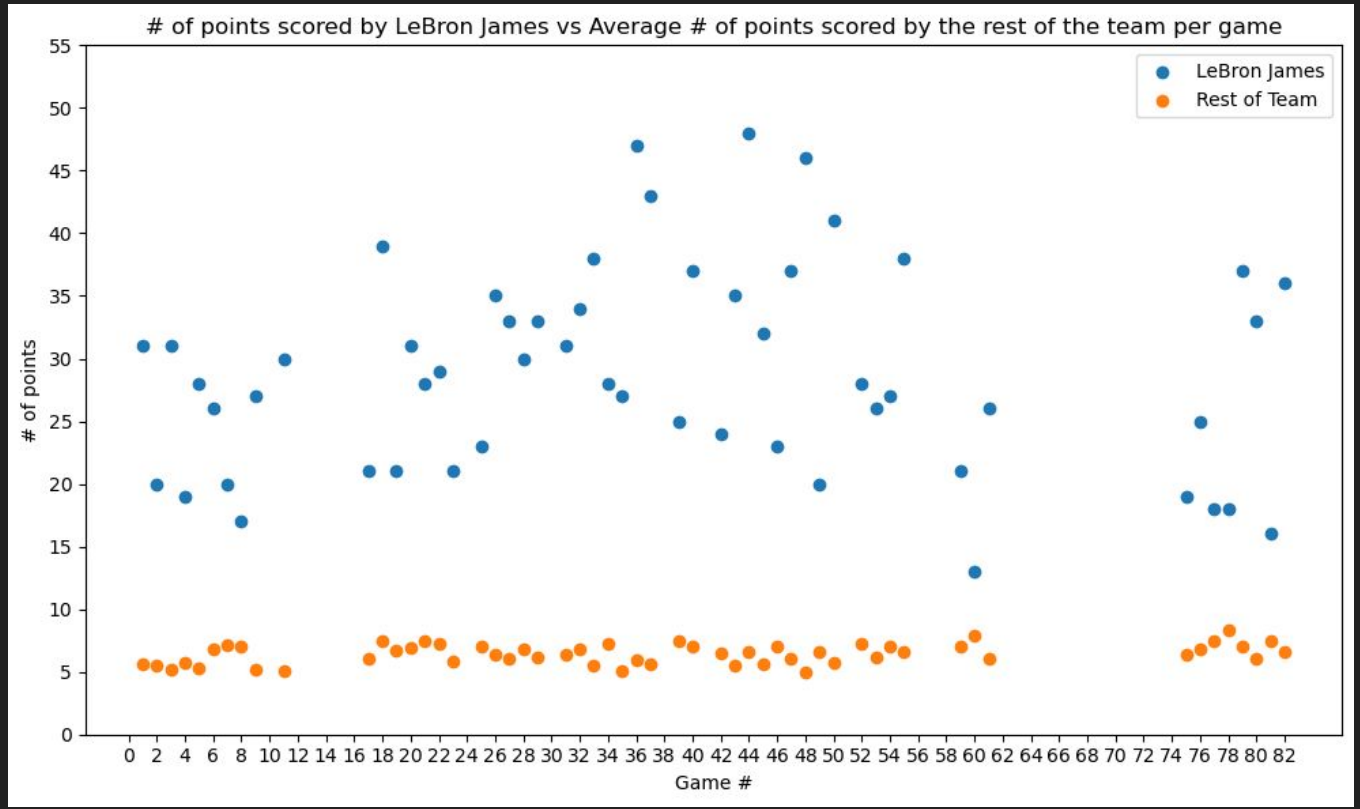
LA Lakers - Player Stats analysis - Conclusions

In the 2022-23 NBA season, Lebron James and Anthony Davis accounted for 31.6% of the total team points. Accounting for the missed games, they were the leading scorers for the Lakers throughout the season.



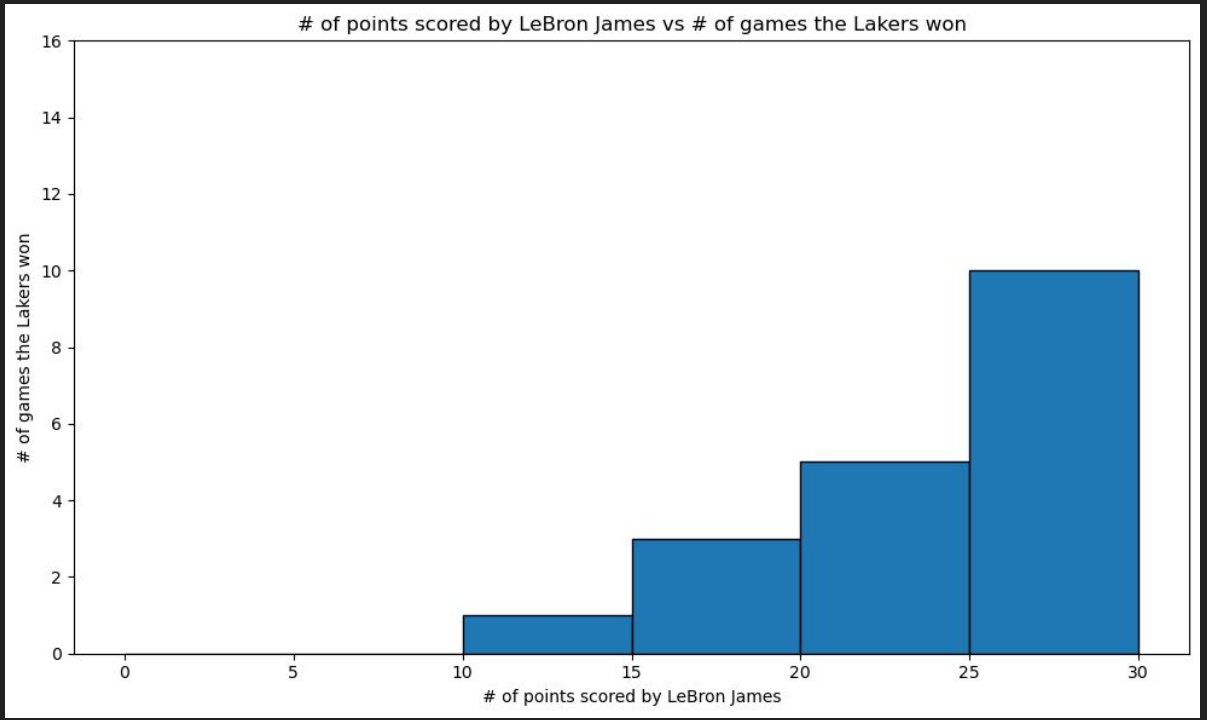
LA Lakers - Player Stats analysis - Conclusions Cont.

Lebron James led the team in scoring throughout the season, averaging 28.9 points per game in 55 games.



LA Lakers - Player Stats analysis - Conclusions Cont’d

Lebron James scored the most number of points in games that the Lakers won.



Implications of our findings

For the Bulls

- **Effect of Travel/Team Performance**
  - Correlation between longer travel distances and decreased win percentages suggests that fatigue, jet lag, or other related factors could be affecting the team. This could leave room to consider re-evaluating travel schedules, manage fatigue, more rest days
- **Influence of Key Players/Team Wins**
  - Wins tied to Vuc's rebound performance implies that team strategy might be heavily dependent on his presence in the paint. This could mean opposing teams target this to disrupt Bulls' strategy, or the Bulls may need other options for games where Vuc isn't available or marked up

For the Lakers

- **Effect of Travel/Team Performance**
  - Correlation between longer distances and increased win percentages suggests strong team resilience or effective adaptation strategies. This could mean current travel routines should be maintained, Lakers could be used as an example for other teams to understand travel strategies
- **Influence of Key Players/Team Wins**
  - Wins tied to LeBron's point scoring could indicate an over-reliance on him. It could imply that games where he is unavailable or not on his game are more likely to be lost. It could also imply the need for developing other options for scoring and offense in general.

The significant contribution of starting duos in both teams' total points emphasizes their offensive roles. This implies that both duos should be in good health as their performance heavily influences the team. And it implies how important contributions from bench players are, especially during crucial games where key players are fatigued or out.