AIT
Asian Institute of Technology

NBA Player Career Dataset Analysis

Class Project Final Report

Group 9

Mingjun Lu st124339

Thitipong Saenpang st123038

24, Nov 2023

Introduction

This project constitutes a comprehensive exploration into the intricacies of shot selection and performance, focusing on the eminent NBA player, Jimmy Butler. Through a synthesis of Exploratory Data Analysis (EDA) and advanced Machine Learning (ML) techniques, the overarching goal is to construct predictive models capable of unraveling the multifaceted decision-making processes that players, exemplified by Jimmy Butler, employ during critical shots. By delving into the granular details of Butler's shot data, this analysis seeks to uncover patterns and insights that contribute to a deeper understanding of his gameplay dynamics.

Meanwhile, this endeavor extends beyond individual player analysis to encompass broader implications for team management. Incorporating a salary analysis of players within each position, the project aims to provide valuable insights for basketball clubs in optimizing their recruitment strategies. Understanding the correlation between player performance and salary distributions within specific positions can empower clubs to make informed decisions, enhancing team dynamics and overall competitiveness.

This dual-pronged approach, combining individual pl ayer analytics with broader team management insights, positions the project at the intersection of player-centric performance evaluation and strategic decision-making for sports organizations. As the analysis unfolds, it aspires to offer a nuanced perspective on the intricate interplay between player performance, shot selection, and the financial considerations that underscore effective team management in the NBA.

Problem Statement

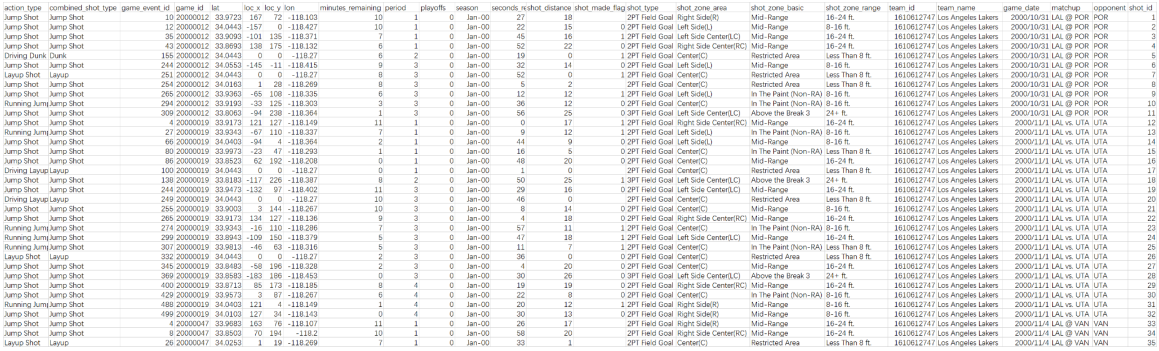
1. **Predictive Analytics:** The crux of our mission lies in foreseeing the upcoming season's performance of NBA teams, meticulously predicting win-loss records, points, and assists, all orchestrated through the lens of historical game data.
2. **Shooting Efficiency Analysis:** Our second frontier involves a deep dive into the intricate tapestry of factors influencing a player's shooting success. This entails a meticulous dissection of various parameters, including action types, shot types, shot distances, and an array of other variables.
3. **Salary Prediction:** Unfurling the future financial landscape, we endeavor to predict NBA player salaries for an imminent season, weaving a narrative around player statistics, positions, team affiliations, and historical salary data, with a pronounced focus on the unfolding drama of the 2021 season.

Related Works

1. **NBA Player Performance Analysis:** Our quest involves an exhaustive exploration of existing research and studies, peering into the multifaceted world of NBA player performance analysis. This encompasses a detailed examination of player statistics, efficiency metrics, and their far-reaching impact on team performance.
2. **Predictive Modeling in Sports Analytics:** With a discerning eye, we traverse through the annals of related works, unraveling the intricate tapestry of predictive modeling in sports analytics. Our focus spans player performance prediction, team win-loss prognosis, and the intricate factors that dance together to compose the symphony of success.
3. **Exploratory Data Analysis (EDA) in Sports Analytics:** Our journey into the heart of sports analytics literature includes a meticulous review of studies that have applied EDA techniques. We seek to glean insights into player performance nuances, team dynamics, and the capricious outcomes of riveting games.
4. **Machine Learning Models for NBA Data:** In our relentless pursuit of knowledge, we scrutinize research that has harnessed the power of machine learning models, notably the Random Forest algorithm, for the analysis of NBA data. Our gaze extends to the delicate art of feature engineering and the critical evaluation of model performance.

Datasets

1. Career Dataset: This dataset comprises 30,697 records from Kobe Bryant's entire career, providing detailed information about individual shots taken by NBA players during games. It includes features like action type, shot type, shot distance, and game-related information.



action_type	combined_shot_type	game_event_id	game_id	lat	loc_x	loc_y	lon	minutes_remaining	period	playoffs	season	seconds_remaining	shot_distance	shot_made_flag	shot_type	shot_zone_area	shot_zone_basic	shot_zone_range	team_id	team_name	game_date	matchup	opponent	shot_id
Jump Shot	Jump Shot	10	20000012	33.9723	167	72	-118.103	10	1	0	Jan-00	27	18		2PT Field Goal	Right Side(R)	Mid-Range	16-24 ft.	1610612747	Los Angeles Lakers	2000/10/31	LAL @ POR	POR	1
Jump Shot	Jump Shot	12	20000012	34.0443	-157	0	-118.427	10	1	0	Jan-00	22	15		2PT Field Goal	Left Side(L)	Mid-Range	8-16 ft.	1610612747	Los Angeles Lakers	2000/10/31	LAL @ POR	POR	2
Jump Shot	Jump Shot	35	20000012	33.9093	-101	135	-118.371	7	1	0	Jan-00	45	16		2PT Field Goal	Left Side Center(LC)	Mid-Range	16-24 ft.	1610612747	Los Angeles Lakers	2000/10/31	LAL @ POR	POR	3
Jump Shot	Jump Shot	43	20000012	33.8693	138	175	-118.132	6	1	0	Jan-00	52	22		2PT Field Goal	Right Side Center(RC)	Mid-Range	16-24 ft.	1610612747	Los Angeles Lakers	2000/10/31	LAL @ POR	POR	4
Driving Dunk	Dunk	155	20000012	34.0443	0	0	-118.27	6	2	0	Jan-00	19	0		2PT Field Goal	Center(C)	Restricted Area	Less Than 8 ft.	1610612747	Los Angeles Lakers	2000/10/31	LAL @ POR	POR	5
Jump Shot	Jump Shot	244	20000012	34.0553	-145	-11	-118.415	9	3	0	Jan-00	32	14		2PT Field Goal	Left Side(L)	Mid-Range	8-16 ft.	1610612747	Los Angeles Lakers	2000/10/31	LAL @ POR	POR	6
Layup Shot	Layup	251	20000012	34.0443	0	0	-118.27	8	3	0	Jan-00	52	0		2PT Field Goal	Center(C)	Restricted Area	Less Than 8 ft.	1610612747	Los Angeles Lakers	2000/10/31	LAL @ POR	POR	7
Jump Shot	Jump Shot	254	20000012	34.0163	1	28	-118.269	8	3	0	Jan-00	5	2		2PT Field Goal	Center(C)	Restricted Area	Less Than 8 ft.	1610612747	Los Angeles Lakers	2000/10/31	LAL @ POR	POR	8
Jump Shot	Jump Shot	265	20000012	33.9363	-65	108	-118.335	6	3	0	Jan-00	12	12		2PT Field Goal	Left Side(L)	In The Paint (Non-RA)	8-16 ft.	1610612747	Los Angeles Lakers	2000/10/31	LAL @ POR	POR	9
Running Jump	Jump Shot	294	20000012	33.9193	-33	125	-118.303	3	3	0	Jan-00	36	12		2PT Field Goal	Center(C)	In The Paint (Non-RA)	8-16 ft.	1610612747	Los Angeles Lakers	2000/10/31	LAL @ POR	POR	10
Jump Shot	Jump Shot	309	20000012	33.8063	-94	238	-118.364	1	3	0	Jan-00	56	25		3PT Field Goal	Left Side Center(LC)	Above the Break 3	24+ ft.	1610612747	Los Angeles Lakers	2000/10/31	LAL @ POR	POR	11
Jump Shot	Jump Shot	4	20000019	33.9173	121	127	-118.149	11	1	0	Jan-00	0	17		2PT Field Goal	Right Side Center(RC)	Mid-Range	16-24 ft.	1610612747	Los Angeles Lakers	2000/11/1	LAL vs. UTA	UTA	12
Running Jump	Jump Shot	27	20000019	33.9343	-67	110	-118.337	7	1	0	Jan-00	9	12		2PT Field Goal	Left Side(L)	In The Paint (Non-RA)	8-16 ft.	1610612747	Los Angeles Lakers	2000/11/1	LAL vs. UTA	UTA	13
Jump Shot	Jump Shot	66	20000019	34.0403	-94	4	-118.364	2	1	0	Jan-00	44	9		2PT Field Goal	Left Side(L)	Mid-Range	8-16 ft.	1610612747	Los Angeles Lakers	2000/11/1	LAL vs. UTA	UTA	14
Jump Shot	Jump Shot	80	20000019	33.9973	-23	47	-118.293	1	1	0	Jan-00	16	5		2PT Field Goal	Center(C)	In The Paint (Non-RA)	Less Than 8 ft.	1610612747	Los Angeles Lakers	2000/11/1	LAL vs. UTA	UTA	15
Jump Shot	Jump Shot	86	20000019	33.8523	62	192	-118.208	0	1	0	Jan-00	48	20		2PT Field Goal	Center(C)	Mid-Range	16-24 ft.	1610612747	Los Angeles Lakers	2000/11/1	LAL vs. UTA	UTA	16
Driving Layup	Layup	100	20000019	34.0443	0	0	-118.27	0	1	0	Jan-00	1	0		2PT Field Goal	Center(C)	Restricted Area	Less Than 8 ft.	1610612747	Los Angeles Lakers	2000/11/1	LAL vs. UTA	UTA	17
Jump Shot	Jump Shot	138	20000019	33.8183	-117	226	-118.387	8	2	0	Jan-00	50	25		3PT Field Goal	Left Side Center(LC)	Above the Break 3	24+ ft.	1610612747	Los Angeles Lakers	2000/11/1	LAL vs. UTA	UTA	18
Jump Shot	Jump Shot	244	20000019	33.9473	-132	97	-118.402	11	3	0	Jan-00	29	16		2PT Field Goal	Left Side Center(LC)	Mid-Range	16-24 ft.	1610612747	Los Angeles Lakers	2000/11/1	LAL vs. UTA	UTA	19
Driving Layup	Layup	249	20000019	34.0443	0	0	-118.27	10	3	0	Jan-00	46	0		2PT Field Goal	Center(C)	Restricted Area	Less Than 8 ft.	1610612747	Los Angeles Lakers	2000/11/1	LAL vs. UTA	UTA	20
Jump Shot	Jump Shot	258	20000019	33.9093	3	144	-118.267	10	3	0	Jan-00	8	14		2PT Field Goal	Center(C)	Mid-Range	8-16 ft.	1610612747	Los Angeles Lakers	2000/11/1	LAL vs. UTA	UTA	21
Jump Shot	Jump Shot	265	20000019	33.9173	134	127	-118.136	9	3	0	Jan-00	4	18		2PT Field Goal	Right Side Center(RC)	Mid-Range	16-24 ft.	1610612747	Los Angeles Lakers	2000/11/1	LAL vs. UTA	UTA	22
Running Jump	Jump Shot	274	20000019	33.9343	-16	110	-118.286	7	3	0	Jan-00	57	11		2PT Field Goal	Center(C)	In The Paint (Non-RA)	8-16 ft.	1610612747	Los Angeles Lakers	2000/11/1	LAL vs. UTA	UTA	23
Running Jump	Jump Shot	299	20000019	33.8943	-109	190	-118.379	5	3	0	Jan-00	47	18		2PT Field Goal	Left Side Center(LC)	Mid-Range	16-24 ft.	1610612747	Los Angeles Lakers	2000/11/1	LAL vs. UTA	UTA	24
Running Jump	Jump Shot	307	20000019	33.9813	-46	63	-118.316	5	3	0	Jan-00	11	7		2PT Field Goal	Center(C)	In The Paint (Non-RA)	Less Than 8 ft.	1610612747	Los Angeles Lakers	2000/11/1	LAL vs. UTA	UTA	25
Layup Shot	Layup	332	20000019	34.0443	0	0	-118.27	2	3	0	Jan-00	36	0		2PT Field Goal	Center(C)	Restricted Area	Less Than 8 ft.	1610612747	Los Angeles Lakers	2000/11/1	LAL vs. UTA	UTA	26
Jump Shot	Jump Shot	345	20000019	33.8493	-58	196	-118.328	2	3	0	Jan-00	4	20		2PT Field Goal	Center(C)	Mid-Range	16-24 ft.	1610612747	Los Angeles Lakers	2000/11/1	LAL vs. UTA	UTA	27
Jump Shot	Jump Shot	369	20000019	33.8983	-183	186	-118.453	0	3	0	Jan-00	30	26		3PT Field Goal	Left Side Center(LC)	Above the Break 3	24+ ft.	1610612747	Los Angeles Lakers	2000/11/1	LAL vs. UTA	UTA	28
Jump Shot	Jump Shot	400	20000019	33.8713	85	173	-118.185	9	4	0	Jan-00	19	19		2PT Field Goal	Right Side Center(RC)	Mid-Range	16-24 ft.	1610612747	Los Angeles Lakers	2000/11/1	LAL vs. UTA	UTA	29
Jump Shot	Jump Shot	429	20000019	33.9573	3	87	-118.267	6	4	0	Jan-00	22	8		2PT Field Goal	Center(C)	In The Paint (Non-RA)	8-16 ft.	1610612747	Los Angeles Lakers	2000/11/1	LAL vs. UTA	UTA	30
Running Jump	Jump Shot	488	20000019	34.0403	121	4	-118.149	1	4	0	Jan-00	20	12		2PT Field Goal	Right Side(R)	Mid-Range	8-16 ft.	1610612747	Los Angeles Lakers	2000/11/1	LAL vs. UTA	UTA	31
Jump Shot	Jump Shot	499	20000019	34.0103	127	34	-118.143	0	4	0	Jan-00	30	13		2PT Field Goal	Right Side(R)	Mid-Range	8-16 ft.	1610612747	Los Angeles Lakers	2000/11/1	LAL vs. UTA	UTA	32
Jump Shot	Jump Shot	4	20000047	33.9683	163	76	-118.107	11	1	0	Jan-00	26	17		2PT Field Goal	Right Side(R)	Mid-Range	16-24 ft.	1610612747	Los Angeles Lakers	2000/11/4	LAL @ VAN	VAN	33
Jump Shot	Jump Shot	8	20000047	33.8503	70	194	-118.2	10	1	0	Jan-00	58	20		2PT Field Goal	Right Side Center(RC)	Mid-Range	16-24 ft.	1610612747	Los Angeles Lakers	2000/11/4	LAL @ VAN	VAN	34
Layup Shot	Layup	26	20000047	34.0253	1	19	-118.269	7	1	0	Jan-00	33	1		2PT Field Goal	Center(C)	Restricted Area	Less Than 8 ft.	1610612747	Los Angeles Lakers	2000/11/4	LAL @ VAN	VAN	35

30,697 pieces of data from Kobe Bryant's entire career from the '96 season to the 2016 season

Column	Meaning
action_type	Action Type
combined_shot_type	Combined Shot Type
game_event_id	Game Event ID
lat	Latitude
loc_x	X Coordinate
loc_y	Y Coordinate
lon	Longitude
minutes_remaining	Minutes Remaining
period	Period
playoffs	Playoffs
season	Season
seconds_remaining	Seconds Remaining
shot_distance	Shot Distance
shot_made_flag	Shot Made Flag
shot_type	Shot Type
shot_zone_area	Shot Zone Area
shot_zone_basic	Shot Zone Basic
shot_zone_range	Shot Zone Range
team_id	Team ID
team_name	Team Name
game_date	Game Date
matchup	Matchup
opponent	Opponent
game_id	Game ID
shot_id	Shot ID

Column name and meanings

2. NBA Player Information Dataset: This dataset contains information about 4,550 professional players, including career start and end years, team affiliations, and player statistics per season from 1950 to 2017.

name	year_start	year_end	position	height	weight	birth_date	college
Alaa Abdelnaby	1991	1995	F-C	6月10日	240	24-Jun-68	Duke University
Zaid Abdul-Aziz	1969	1978	C-F	6月9日	235	7-Apr-46	Iowa State University
Kareem Abdul-Jabbar	1970	1989	C	7月2日	225	16-Apr-47	University of California, Los Ange
Mahmoud Abdul-Rauf	1991	2001	G	6月1日	162	9-Mar-69	Louisiana State University
Tariq Abdul-Wahad	1998	2003	F	6月6日	223	3-Nov-74	San Jose State University
Shareef Abdur-Rahim	1997	2008	F	6月9日	225	11-Dec-76	University of California
Tom Abernethy	1977	1981	F	6月7日	220	6-May-54	Indiana University
Forest Able	1957	1957	G	6月3日	180	27-Jul-32	Western Kentucky University
John Abramovic	1947	1948	F	6月3日	195	9-Feb-19	Salem International University
Alex Abrines	2017	2018	G-F	6月6日	190	1-Aug-93	
Alex Acker	2006	2009	G	6月5日	185	21-Jan-83	Pepperdine University

NBA Player Information Dataset

3. NBA Player Salaries Dataset: This Dataset provides a comprehensive overview of player compensation in the National Basketball Association (NBA) over a span of two decades, from 2000 to 2020. With 9,456 records, the dataset includes valuable information such as player rank based on salary, player names, positions (e.g., Center, Power Forward, Small Forward, Point Guard, Shooting Guard), and affiliated NBA teams. This dataset offers insights into the financial landscape of professional basketball, showcasing the evolution of player salaries over time. Analysts and enthusiasts alike can utilize this rich dataset to explore trends, compare player earnings, and gain a deeper understanding of the economic dynamics within the NBA during the specified period.

rank	name	position	team	salary	season
1	Shaquille O'N	C	Los Angeles L	17142000	2000
2	Kevin Garnet	PF	Minnesota Ti	16806000	2000
3	Alonzo Mour	C	Miami Heat	15004000	2000
4	Juwan Howa	PF	Washington	15000000	2000
5	Scottie Pippe	SF	Portland Trai	14795000	2000
6	Karl Malone	PF	Utah Jazz	14000000	2000
.					
.					
.					
523	Tyler Zeller	C	San Antonio	245686	2020
524	Jimmer Fred	G	Phoenix Suns	208509	2020
525	Jontay Porte	PF	Memphis Gri	197933	2020
526	Anthony Toll	PF	Memphis Gri	183115	2020
527	Tyler Johnso	SG	Phoenix Suns	183115	2020
528	Luguentz Do	SG	Oklahoma Ci	155647	2020

NBA Player Salaries Dataset

Methodology

Exploratory Data Analysis (EDA):

- **Objectives:** Our EDA serves as a compass, guiding us through the dense thicket of data intricacies. The mission is expansive, ranging from unraveling the underlying data structure to detecting outliers and anomalies, testing fundamental assumptions, and crystallizing optimal factor settings.
- **Graphical Techniques in EDA:** Armed with a formidable arsenal of graphical techniques, our EDA journey involves the orchestration of data traces, histograms, bivariate histograms, probability plots, lag plots, block plots, mean plots, standard deviation plots, box plots, main effects plots, probability plot correlation coefficient plots, univariate and multivariate control charts, 4-plots, block plots, scatter plots, and a myriad of visual tools tailored to the unique contours of our data.
- **Applications in Different EDA Problems:** Tailoring our approach to the specific nature of the EDA problem at hand, we deftly wield tools like probability plots, probability plot correlation coefficient plots, univariate and multivariate control charts, 4-plots, block plots, scatter plots, and more. Our arsenal is diverse, ensuring that we unravel the data mysteries through a kaleidoscope of analytical lenses.

Pre-processing:

- In the prelude to the grand symphony of machine learning, we navigate the preliminary steps of data preprocessing. Our voyage involves the delicate art of handling missing values, encoding categorical variables, and orchestrating the harmonious scaling or normalization of features to create a symphony of meaningful data for machine learning.

Machine Learning Training/Testing:

- **Random Forest Prediction:** The marquee event of our methodological extravaganza unfolds with the deployment of the Random Forest algorithm, an ensemble learning method that begets a "forest" of decision trees. The training process, infused with a judicious dose of randomness, births a diverse array of decision trees, enhancing model robustness and orchestrating a ballet that mitigates the specter of overfitting.

EDA Highlights:

- **Shot Type Distribution:** Our visual symphony takes a dramatic turn, depicting the distribution of Butler's shot types. This visual narrative unveils a predilection for

certain shot types, such as jump shots and layups, offering a tantalizing glimpse into the intricate dance of his offensive strategy.

- **3-Point vs. Mid-Range Shots:** Another act in our visual spectacle showcases the distribution between 3-point and mid-range shots, unfurling the canvas of Butler's proficiency in different shooting categories.
- **Shooting Range Analysis:** The canvas expands further with histograms and boxplots, offering a panoramic view of Butler's shooting range. This visual panorama underscores the critical importance of understanding shot distances in the rich tapestry of player performance.

Machine Learning Predictive Model:

- **Random Forest Classifier:** The pièce de résistance in our methodological saga, the Random Forest Classifier, takes center stage. Its performance, akin to a virtuoso performance, yields promising results in predicting Butler's shot outcomes. Through meticulous feature importance analysis, we identify the instrumental elements that orchestrate successful shots.

Conclusions and future works

In the grand finale of our project, the curtain falls, and we draw the threads of our analysis into a harmonious conclusion. The developed predictive models, crafted with precision and insight, offer a robust framework for the evaluation and forecasting of player performance. These insights, akin to a compass, guide teams, enabling them to tailor strategies steeped in the rich tapestry of historical performance data.

As we cast our gaze towards the future, the melody of future works beckons. The symphony of refinement calls for the incorporation of additional features, the exploration of advanced machine learning techniques, and an expansion of our analysis to include datasets from other players. This expansive vision seeks to create a more comprehensive understanding of the captivating world of basketball analytics, where each player's narrative enriches the collective saga.