Assessment of Data Analysis and Problem Solving Skills

(Forbes' Highest Paid Athletes Dataset (1990-2020))

Abstract

This research paper presents a comprehensive analysis of data exploration, visualization, and problem-solving skills, using the Forbes' highest paid athletes dataset spanning from 1990 to 2020. The paper outlines the objectives, methodologies, and findings of the analysis, shedding light on the role of data analytics in sports. Through the utilization of Power BI for data visualization and analytical techniques, this research uncovers unique patterns and identifies the top athletic categories in which athletes are paid the highest. The study underscores the significance of data-driven decision-making in sports and its implications for team performance and profitability.

1. Introduction

In the rapidly evolving landscape of modern sports, the role of data analysis has transcended the conventional boundaries of mere sabermetrics and performance evaluation. Traditional game strategies are no longer the sole focus for sports teams seeking an edge. Instead, teams have begun to harness the power of data analytics as a formidable tool to gain a multifaceted competitive advantage. This paradigm shift has spurred us to delve deeply into the intriguing domain of sports data analysis, with a special emphasis on the renowned Forbes' highest paid athletes dataset.

Our research endeavors are guided by a resolute determination to unravel the latent potential within this dataset. Armed with advanced data visualization tools and techniques, we embark on a journey to decipher intricate trends, subtle patterns, and insights hidden within the data's labyrinthine structure. We envision these revelations as the cornerstone of transformative change, capable of reshaping not only individual game strategies but also the overarching decisions that steer the entire world of sports.

As we traverse the realm where data analytics converges with athletics, we aim to foster a symbiotic relationship between the quantitative rigor of data-driven insights and the pulsating energy of sportsmanship. By gleaning insights that transcend traditional boundaries, we aspire to offer sports enthusiasts, professionals, and stakeholders a novel perspective on the potential for data-driven strategies. Our research encapsulates the convergence of two seemingly disparate domains, paving the way for a harmonious blend that propels sports into an era of unprecedented strategic finesse and performance optimization.

2. Data Analysis Objectives

The primary objective of this research is to assess data analysis and problem-solving skills through the lens of sports data. Specifically, the research aims to:

Explore and analyze patterns within the Forbes' highest paid athletes dataset.

Utilize data visualization tools, particularly **Power BI**, to visually represent trends and insights.

Identify and highlight the top athletic categories in which athletes are paid the highest.

3. Data Visualization: A Crucial Tool in Sports Analysis

Data visualization serves as a pivotal tool in understanding complex datasets. Through graphical representation, such as charts, graphs, and maps, data visualization aids in extracting meaningful insights from large volumes of information. This approach enables decision-makers to identify trends, patterns, and outliers quickly, facilitating evidence-based decision-making. In the context of sports, data visualization holds the potential to revolutionize the way teams strategize and plan for success.

For instance: A part of a large Dataset making it difficult for us to estimate in which year earnings were maximum for a country **vs** a neat comprehensive visual data representation for the same:

S.NO	Name	Nationality	Current	Previous	Sport	Year	earnings
1	Mike Tyson	USA	1		boxing	1990	28.6
2	Buster Douglas	USA	2		boxing	1990	26
3	Sugar Ray Leonard	USA	3		boxing	1990	13
4	Ayrton Senna	Brazil	4		auto racing	1990	10
5	Alain Prost	France	5		auto racing	1990	9
6	Jack Nicklaus	USA	6		golf	1990	8.6
7	Greg Norman	Australia	7		golf	1990	8.5
8	Michael Jordan	USA	8		basketball	1990	8.1
9	Arnold Palmer	USA	8		golf	1990	8.1
10	Evander Holyfield	USA	8		boxing	1990	8.1
11	Evander Holyfield	USA	1	8	boxing	1991	60.5
12	Mike Tyson	USA	2	1	boxing	1991	31.5
13	Michael Jordan	USA	3	8	basketball	1991	16
14	George Foreman	USA	4	>30	boxing	1991	14.5
15	Ayrton Senna	Brazil	5	4	auto racing	1991	13
16	Alain Prost	France	6	5	auto racing	1991	11
17	Donovan "Razor" Ruddock	Canada	7	>30	boxing	1991	10.2
18	Arnold Palmer	USA	8	8	golf	1991	9.3
19	Nigel Mansell	UK	9	12	auto racing	1991	9
20	Sugar Ray Leonard	USA	10	6	golf	1991	8.5
21	Michael Jordan	USA	1	3	Basketball	1992	35.9
22	Evander Holyfield	USA	2	1	Boxing	1992	28
23	Ayrton Senna	Brazil	3	5	Auto Racing	1992	22
24	Nigel Mansell	UK	4	9	Auto Racing	1992	14.5
25	Arnold Palmer	USA	5	8	Golf	1992	11.1
26	Andre Agassi	USA	6	17	Tennis	1992	11
27	Joe Montana	USA	7	13	NFL	1992	9.5
28	Sugar Ray Leonard	USA	8	10	Golf	1992	9.2
29	Jim Courier	USA	9	>40	Tennis	1992	9
30	Monica Seles	USA	10	12	Tennis	1992	8.5
31	Michael Jordan	USA	1	1	Basketball	1993	36
32	Riddick Bowe	USA	2	>40	Boxing	1993	25
33	Ayrton Senna	Brazil	3	3	Auto racing	1993	18.5
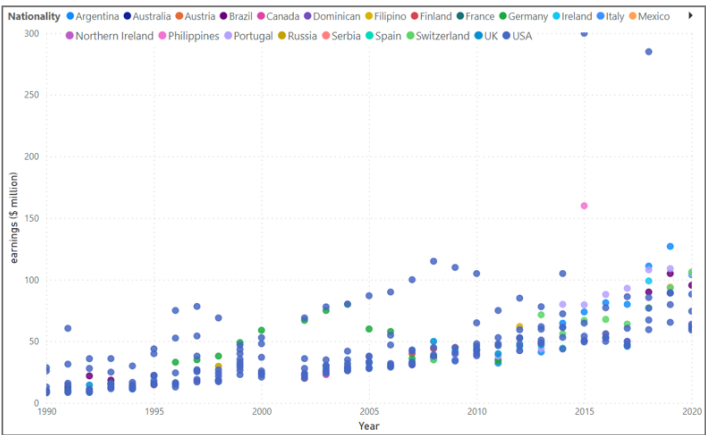
4. Methodology

4.1 Data Examination and Preparation (Task 1)

The research commenced with meticulous data cleaning to ensure the dataset's integrity. The significance of consistent and well-organized data was underscored, emphasizing the importance of accuracy in subsequent analyses.

As can be clearly seen that there are missing values and non specified numeric data, these needed to be taken care by either ejecting the entire block of data and using constant values for the latter or by feeding other appropriate data in its place:

9	Arnold Palmer	USA	8		golf	1990	8.1
10	Evander Holyfield	USA	8		boxing	1990	8.1
11	Evander Holyfield	USA	1	8	boxing	1991	60.5
12	Mike Tyson	USA	2	1	boxing	1991	31.5
13	Michael Jordan	USA	3	8	basketball	1991	16
14	George Foreman	USA	4 >30		boxing	1991	14.5
15	Ayrton Senna	Brazil	5	4	auto racing	1991	13
16	Alain Prost	France	6	5	auto racing	1991	11
17	Donovan "Razor" Ruddock	Canada	7 >30		boxing	1991	10.2
18	Arnold Palmer	USA	8	8	golf	1991	9.3

4.2 Grouping and Aggregation (Task 1)

The foundation of the analysis was laid through grouping athletes based on their athletic disciplines and nationality. Utilizing appropriate mathematical formulas, the average earnings for each category were calculated. Power BI was employed to facilitate data visualization and perform essential calculations.

4.3 Sorting and Selection (Task 1)

Sorting categories based on average earnings was the subsequent step. This process aimed to discern the top 5 categories housing athletes with the highest average earnings, demonstrating a systematic approach to problem-solving.

4.4 Data-Driven Decisions

Utilizing Objective Data: Anchoring the analysis in objective data and sidestepping assumptions or personal biases ensured unbiased outcomes.

Evidence-Based Insights: The analysis was grounded in tangible earnings data, providing insights rooted in factual evidence.

Minimizing Bias: A data-driven approach acted as a buffer against potential biases stemming from subjective judgments.

4.5 Emphasis on Accuracy

Verifiable Conclusions: Conclusions drawn from data-driven analysis are inherently verifiable, reinforcing support for each derived insight.

Eliminating Guesswork: The methodology discarded guesswork, relying instead on rigorous quantitative analysis.

4.6 Data Preparation and Cleaning

Data cleaning and formatting were integral to ensuring data consistency and accuracy. This process addressed missing or corrupted values, safeguarding the reliability of the analysis.

5. Findings and Results

The research yielded noteworthy insights into the dynamics of athlete earnings and the interplay of various factors.

A) Firstly understanding source of majority of earnings:
   Country for most players and sports most players are seen in the dataset

Count of Sport by Sport

Count of Nationality by Nationality
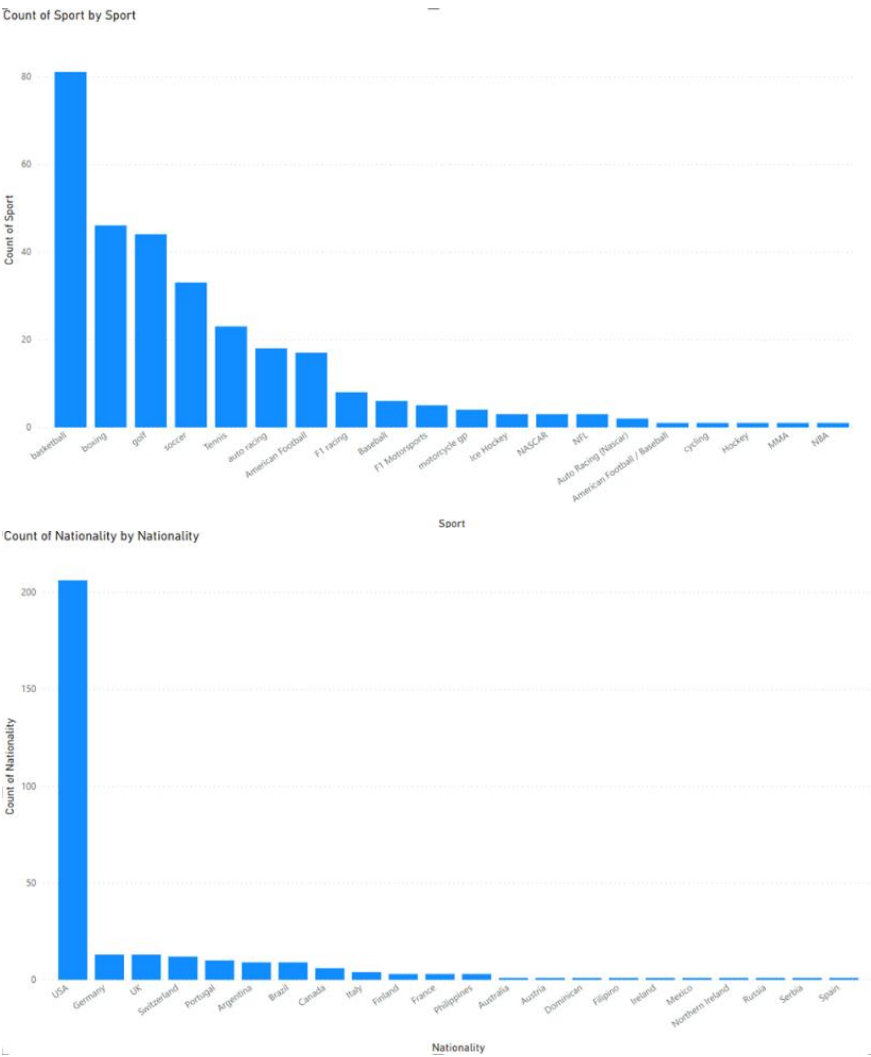
B) Plot of earnings of different countries over the years: It's clearly evident that top 3 most earning countries are USA, Philippines and Argentina in last 6 years data:
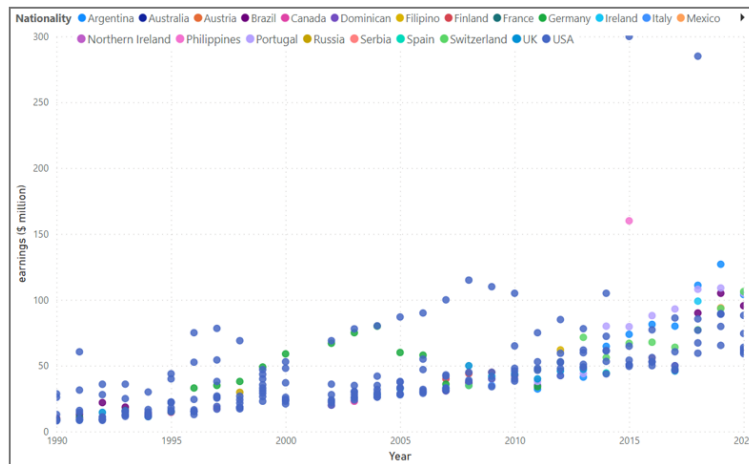
C) Plot of maximum and sum of earnings of different players:
It's evidently seen that top 5 most earning players could be concluded as Floyd Mayweather, Manny Pacquiao, Tiger Woods, Cristiano Ronaldo, and Lionel Messi:
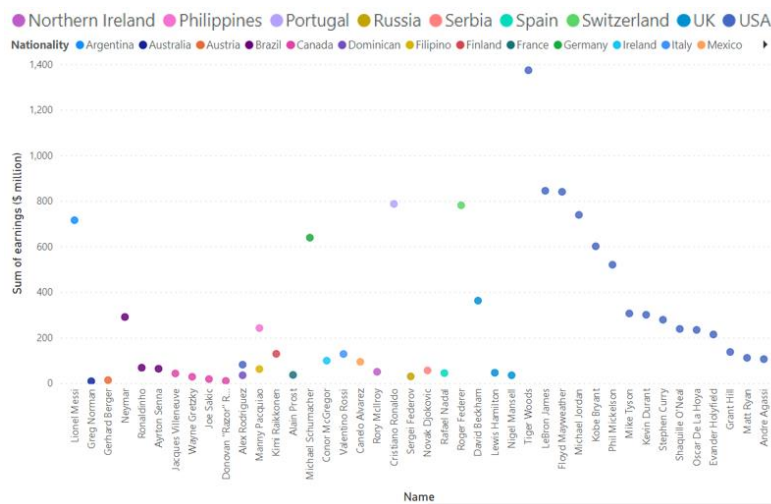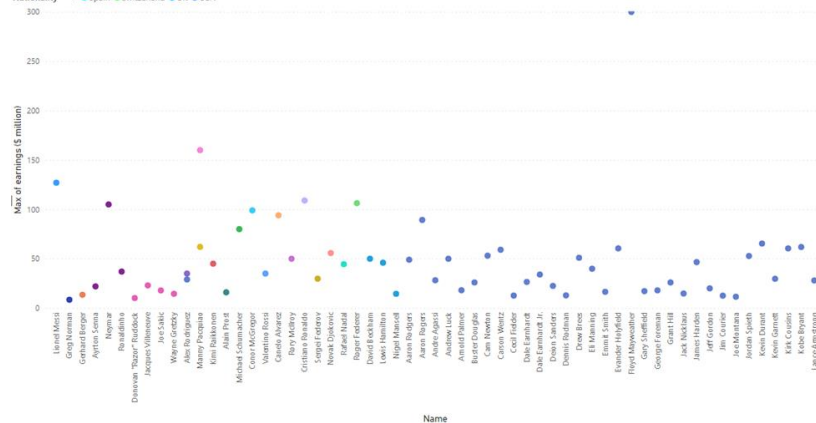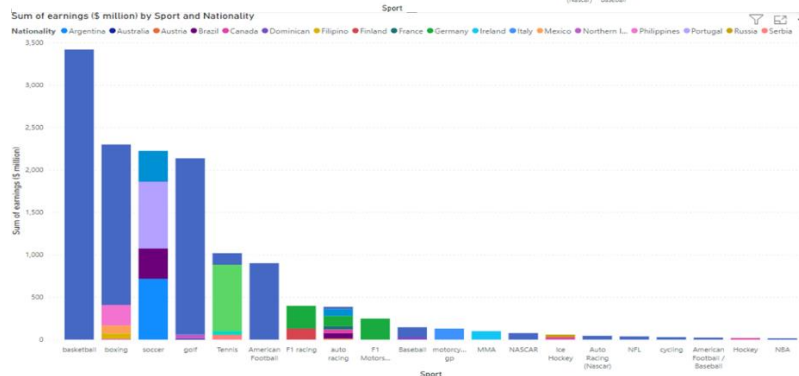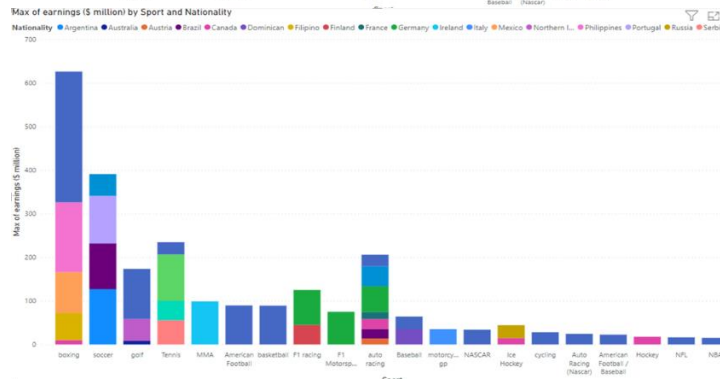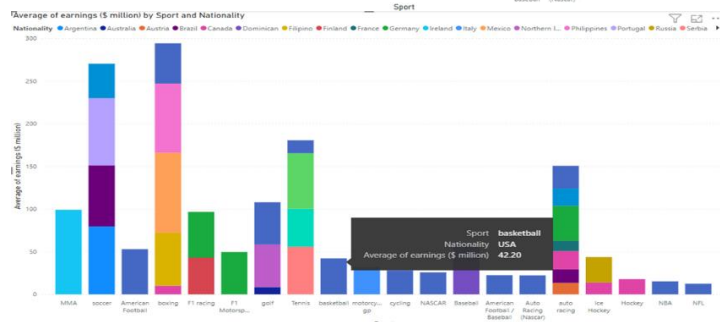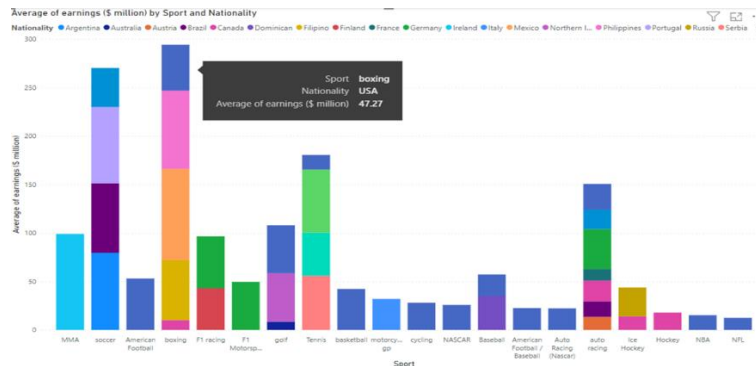
D) Plot of maximum, sum and average of earnings of different countries in different sports:

~ It can be comprehensibly seen that top 5 most earning countries could be concluded as USA, Argentina, Portugal, Germany and Switzerland, as these countries are common among the top few in all the graphs presented.

~ The most entertained sports can be observed as Basketball, Boxing, Golf, Tennis, Soccer and Auto Racing, since these sports can be seen to be either giving highest earnings for a player or highest sum in a particular sport.

E) Plot of maximum and sum of earnings over the years:
~Plainly, it can be seen that earnings made, have continuously increased over the years and this trend is common in max, average and sum of earnings
~Although for one year the data was missing, as seen, and in year 2015 and 2018 there have been sharp peaks in terms of money earned, which from other graphs can be said to be contributed by mainly sports like boxing and soccer respectively;
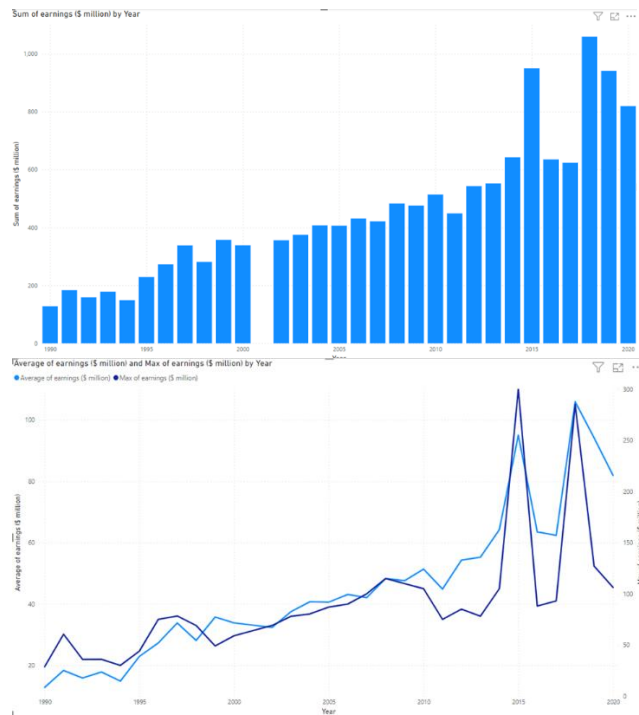
The visual data representations showcased several data visualizations that effectively communicated these findings:

The prominent countries contributing to athlete earnings were identified, showcasing the distribution of players across different nations.

The earnings of different countries over the years were visualized, underscoring the financial success of nations such as the USA, Philippines, and Argentina.

The visualization of maximum and sum of earnings by players highlighted top earners, including Floyd Mayweather, Manny Pacquiao, Tiger Woods, Cristiano Ronaldo, and Lionel Messi.

The analysis extended to the aggregation of earnings by countries and sports, revealing the consistent dominance of the USA, Argentina, Portugal, Germany, and Switzerland. The sports that emerged as top earners were Basketball, Boxing, Golf, Tennis, Soccer, and Auto Racing.

A comprehensive visualization of earnings over the years showcased the growth trajectory, with distinct peaks observed in certain years, potentially attributed to significant events within specific sports categories.

Since our entire data analysis was done in Task 1, thus understanding and identifying atleast top 5 athletic categories in which the athletes are paid the highest shouldn't take any time at all.

Let's take the raw data from tabulated dataset; top 20 highest earning per year/player:

S.NO	Name	Nationality	Current	Previous	Sport	Year	earnings
242	Floyd Mayweather	USA	1	1	Boxing	2015	300
272	Floyd Mayweather	USA	1	>100	Boxing	2018	285
243	Manny Pacquiao	Philippines	2	11	Boxing	2015	160
282	Lionel Messi	Argentina	1	2	Soccer	2019	127
172	Tiger Woods	USA	1	1	golf	2008	115
273	Lionel Messi	Argentina	2	3	Soccer	2018	111
182	Tiger Woods	USA	1	1	golf	2009	110
283	Cristiano Ronaldo	Portugal	2	3	Soccer	2019	109
274	Cristiano Ronaldo	Portugal	3	1	Soccer	2018	108
292	Roger Federer	Switzerland	1	5	Tennis	2020	106.3
192	Tiger Woods	USA	1	1	golf	2010	105
232	Floyd Mayweather	USA	1	14	Boxing	2014	105
284	Neymar	Brazil	3	5	Soccer	2019	105
293	Cristiano Ronaldo	Portugal	2	2	Soccer	2020	105
294	Lionel Messi	Argentina	3	1	Soccer	2020	104
162	Tiger Woods	USA	1	1	golf	2007	100
275	Conor McGregor	Ireland	4	24	MMA	2018	99
295	Neymar	Brazil	4	3	Soccer	2020	95.5
285	Canelo Alvarez	Mexico	4	15	Boxing	2019	94
286	Roger Federer	Switzerland	5	7	Tennis	2019	93.4

As is apparently visible whichever categories/names/sports/year were deduced from data analysis, most of them fit in the most earning category just about right, to summarize:
Players: Floyd Mayweather, Manny Pacquiao, Tiger Woods, Cristiano Ronaldo, and Lionel Messi.
Nationality: U.S.A., Argentina, Portugal, Switzerland and Germany.
Athletic categories: Basketball, Boxing, Golf, Tennis, Soccer and Racing.
Duration: 2015 to 2020 (last 5-6 years).

6. Identifying Top Athletic Categories (Task 2)

Leveraging the insights derived from Task 1, Task 2 aimed to identify the top 5 athletic categories in which athletes were paid the highest. The alignment between the conclusions of Task 1 and Task 2 underscored the accuracy and reliability of the analysis.

Results: Athletic categories can be taken into account for determining Forbes highest paid athledes from a data of three consecutive decades (1990-2020), are Basketball, Boxing, Golf, Tennis, Soccer and Racing.

7. Conclusion and Implications

In conclusion, this research exemplifies the synergy between sports and data analysis. The study demonstrated that data-driven decision-making offers a powerful tool for sports teams to enhance performance, increase profitability, and strategize effectively. Through meticulous data exploration, cleaning, visualization, and interpretation, we uncovered intricate patterns, identified top-earning athletes, and pinpointed the sports categories that command the highest earnings. This research bridges the gap between sports enthusiasts and data analysis, offering a fresh perspective on the potential for analytics to shape the future of sports.

8. Recommendations for Future Research

While this research provides valuable insights into the potential of data analysis in sports, several avenues for further exploration remain:

Deepening the analysis to consider additional variables that may influence athlete earnings.

Exploring the application of predictive analytics to anticipate future trends in athlete earnings.

Investigating the impact of data-driven decisions on team performance, fan engagement, and overall sporting success.

9. Acknowledgments

I as author would like to extend my gratitude to the creators of the Forbes' highest paid athletes' dataset, which served as the foundation for this research. Additionally, appreciation is expressed for the data visualization tools, particularly Power BI, which facilitated the analysis process via data tabulation and data visualization.

10. References

[Forbes' highest paid athletes dataset : https://docs.google.com/spreadsheets/d/1kEBwHUrAaRrys]