Assessment of Data Analysis and Problem Solving Skills

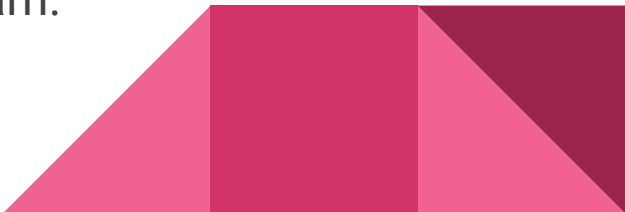Forbes' Highest Paid Athletes Dataset (1990-2020)

-Alisha Verma

Objective

- The presentation showcases the results of data analysis and problem-solving using the Forbes' highest paid athletes dataset.
- Plotting various graphs to analyze and catch patterns using Power BI.
- Laying forth the findings from Task 1 and Task 2, highlighting unique patterns and top-paying athletic categories.

Power BI

Overview

- Any sports enthusiast knows that analyzing data is a big part of the experience. Be it cricket or car racing. But data analysis in sports is now taking teams far beyond old-school sabermetrics and game performance.
- So how exactly does data analysis help?
- Sports is just another field into which we can incorporate our analytical and logical abilites to analyse data and predict interesting results.
- Data analytics seeks to assist decision-makers in sporting teams make better decisions directed toward increased growth and profitability.
- Analysts may also help teams to get valuable information from the data and practise them to enhance the performance of the team.

Data Visualization

- What do we mean by Data Visualization and how does it help us?
- Data visualization is the graphical representation of information and data in a graphical format(Example: charts, graphs, and maps). Data visualization tools provide an accessible way to see, understand and analyse trends, patterns in data, and outliers. Data visualization tools and technologies are essential to analyzing massive amounts of information and making data-driven decisions. General types of data visualization are Charts, Tables, Graphs, Maps.
- The most important thing that data visualization does is discover the trends in data. It is definitely faster to gather some insights from the data using data visualization rather than just studying a chart.
- Let's see a simple example with the data provided for the tasks:

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
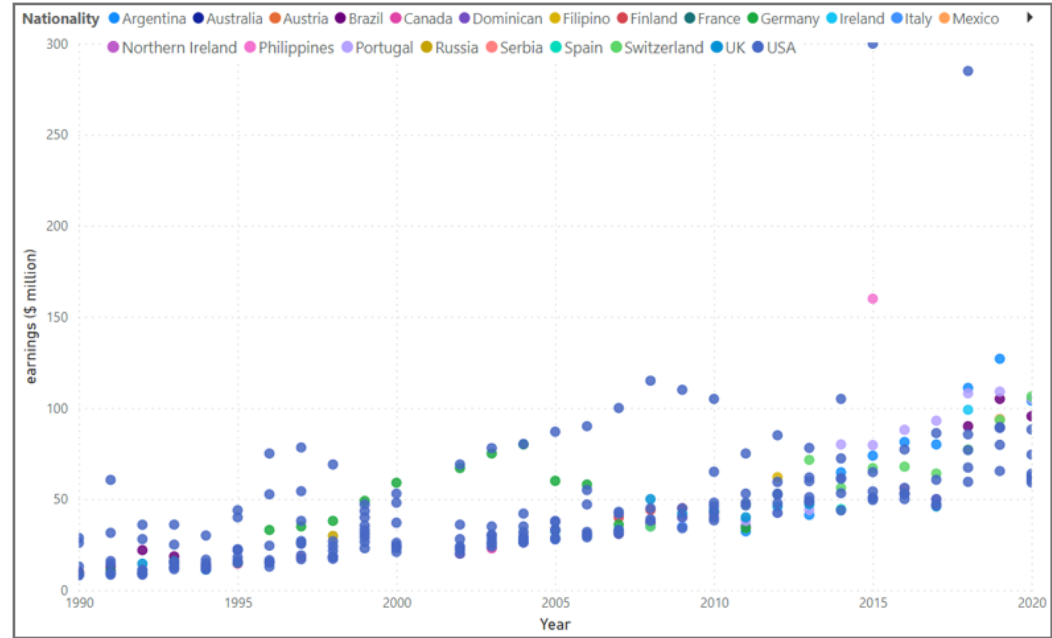
Problem solving Approach:

- **Step 1:** Data Examination and Preparation
  We initiated by meticulously cleaning our dataset. Highlighting the significance of consistent and well-organized data for ensuring precise outcomes.
- **Step 2:** Grouping and Aggregation
  The various columns emerged as the linchpin for categorizing athletes based on their athletic disciplines and nationality etc. Then calculated and plotted the average earnings for each category using appropriate mathematical formulas. Using Power BI to plot all types of required data visualizations do required calculations.
- **Step 3:** Sorting and Selection
  The subsequent stage involved sorting the categories in order based on average earnings. Our final goal is to pinpoint the top 5 categories harboring athletes with the highest average earnings.

Data-Driven Decisions:

- **Utilizing Objective Data:** Anchoring the analysis in objective data, consciously sidestepping assumptions or personal biases for an unbiased outcome.
- **Evidence-Based Insights:** The analysis hinged on interpreting tangible earnings data, furnishing us with insights firmly rooted in factual evidence.
- **Minimizing Bias:** This recourse to a data-driven approach acted as a buffer against potential biases stemming from subjective judgments.

Emphasis on Accuracy:

- **Verifiable Conclusions:** Conclusions drawn from data-driven analysis are inherently verifiable, offering support to every conclusion resulted from the dataset.
- **Eliminating Guesswork:** The Methodology used excised guesswork, relying instead on rigorous quantitative analysis.

Data Preparation and Cleaning

- Why do we need to clean the data?

  ~ Cleaning and formatting the data is required to ensure consistency.

  ~ Addressing missing or corrupted values for ensuring data accuracy.

- Importing the entire dataset of Forbe's highest paid athletes from 1990-2020:
  (https://docs.google.com/spreadsheets/d/1kEBwHUrAaRrysL0o5oHDXyiHZLmnj3Wx/edit#gid=57182323)

- 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

Findings and result, alongwith supported Data visuals
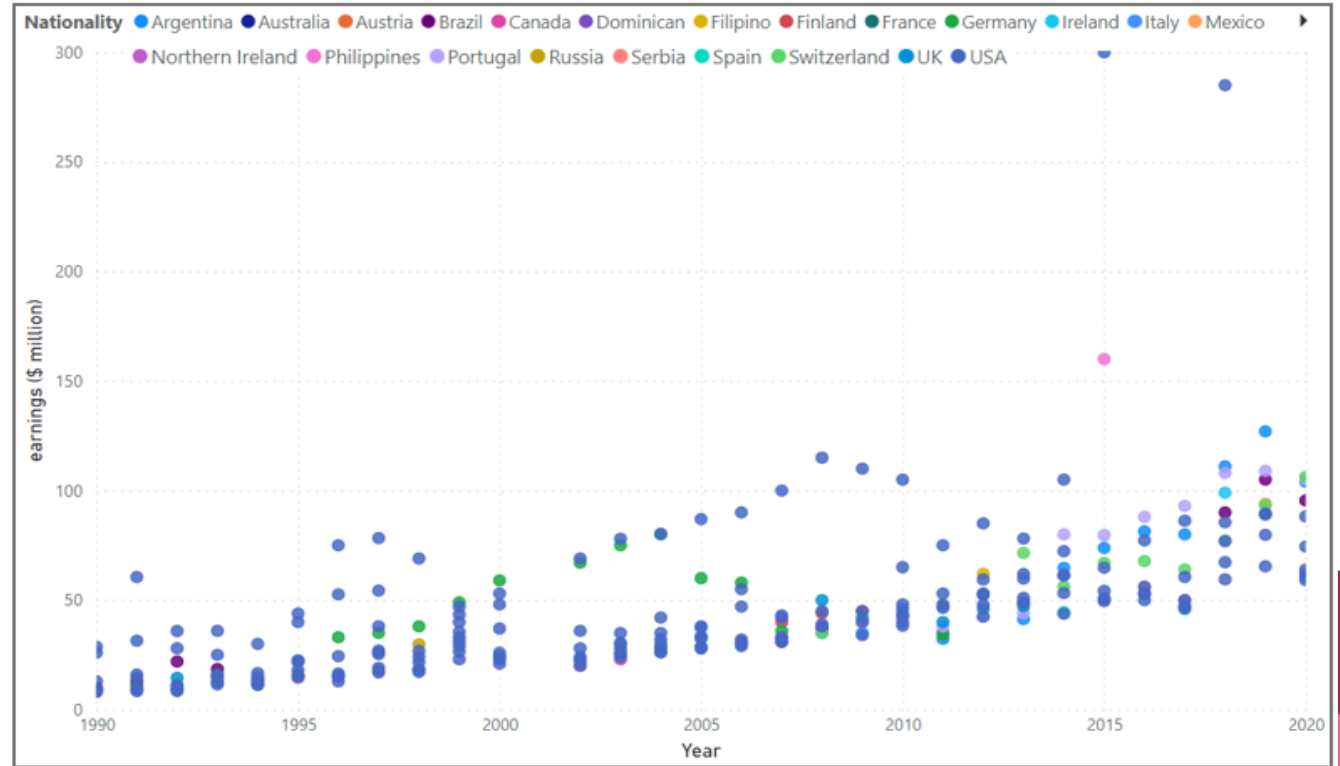
Task 1: Identifying and Analyzing Patterns

Without further ado let's present the analysis report:

● Firstly understanding source of majority of earnings:
● Country for most players and sports most players are seen in the dataset:
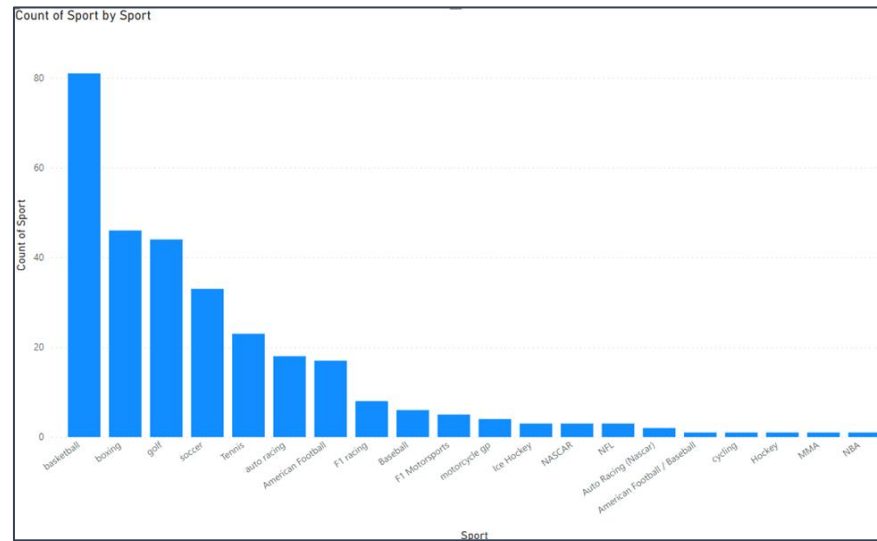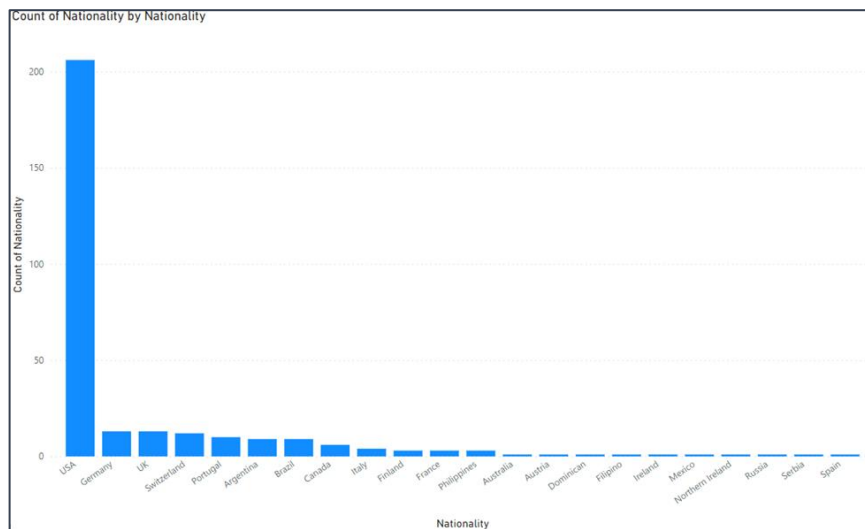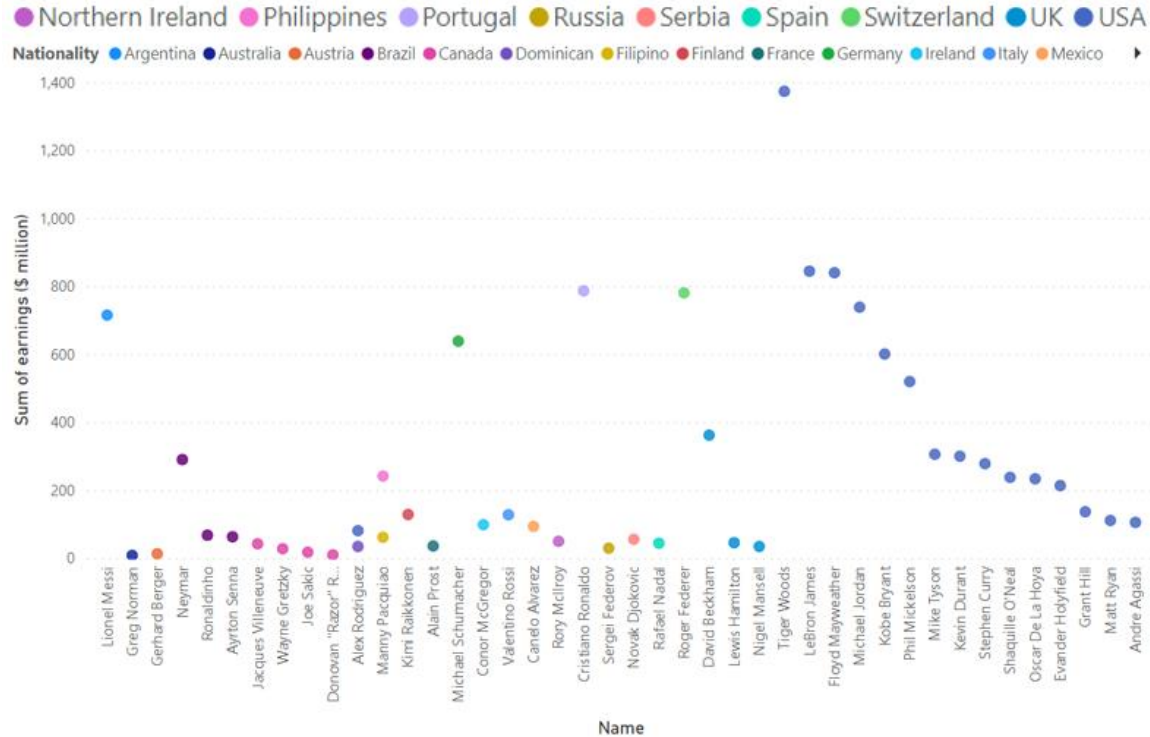
- Plot of earnings of different countries over the years:

  It's clearly evident that top 3 most earning countries are USA, Philippines and Argentina in last 6 years data:

- 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:
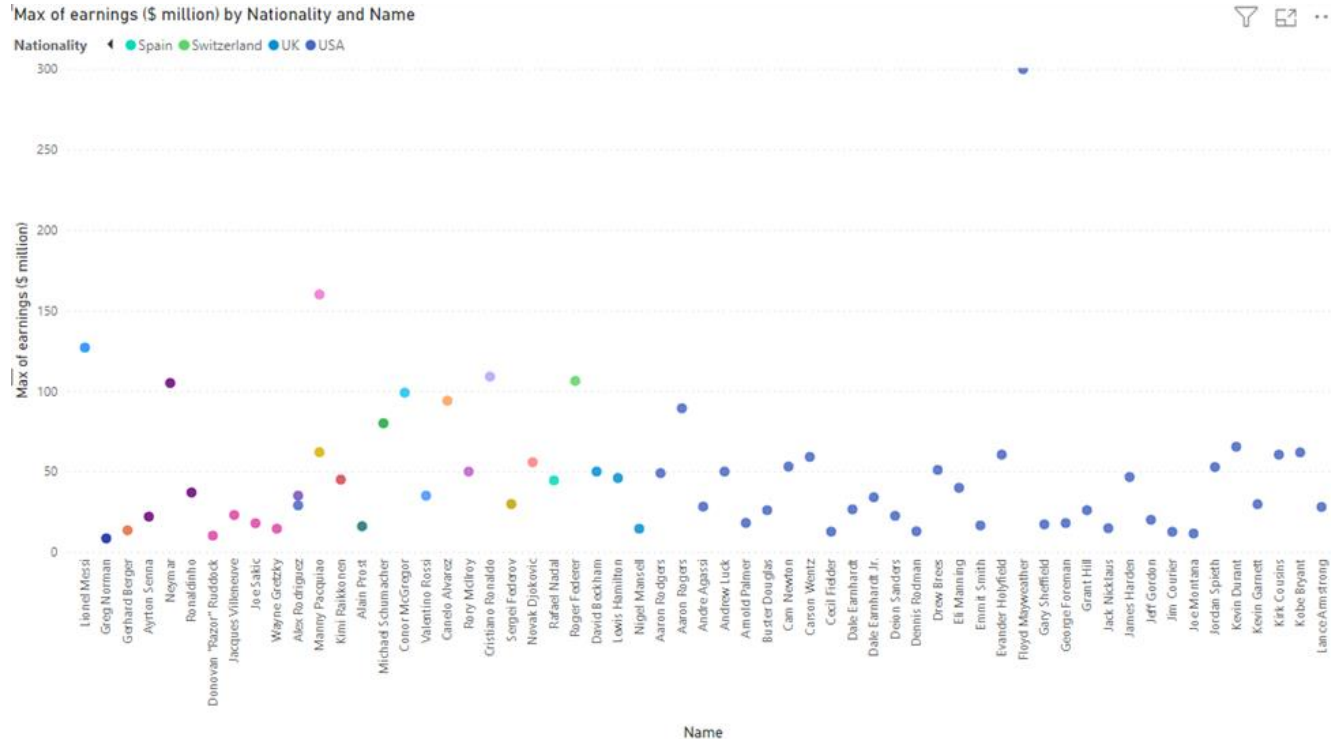
- Plot of maximum and sum of earnings of different players:
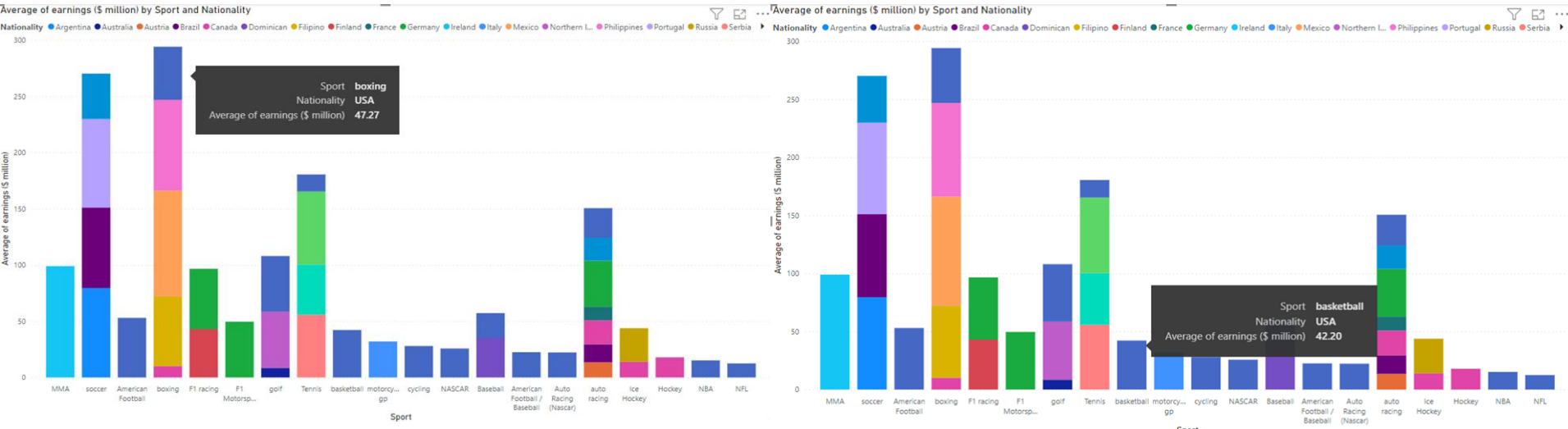
It's evidently seen that top 5 most earning players could be concluded as Floyd Mayweather, Manny Pacquiao, Tiger Woods, Cristiano Ronaldo, and Lionel Messi:

- 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.

Average of earnings ($ million) by Sport and Nationality

Sport **boxing**
Nationality **USA**
Average of earnings ($ million) **47.27**

Average of earnings ($ million) by Sport and Nationality

Sport **basketball**
Nationality **USA**
Average of earnings ($ million) **42.20**

- 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.
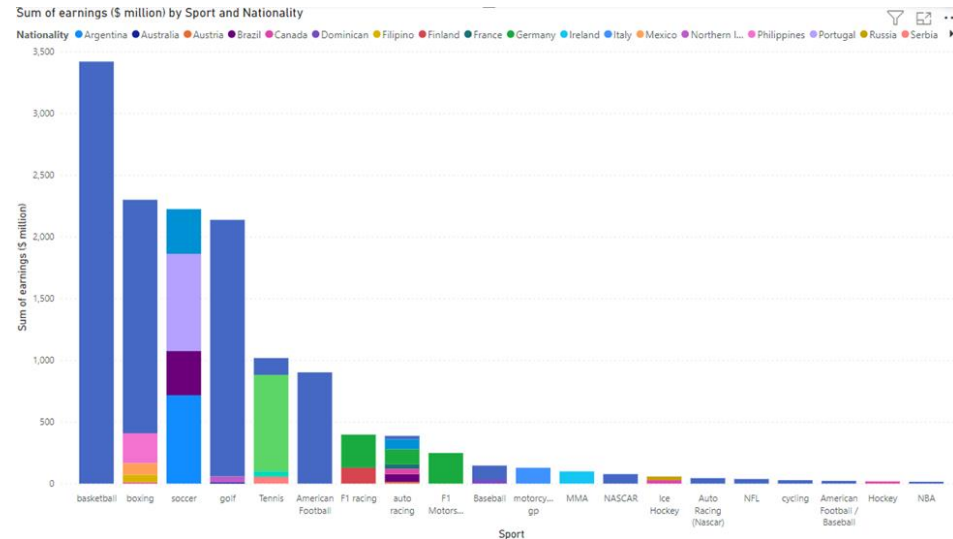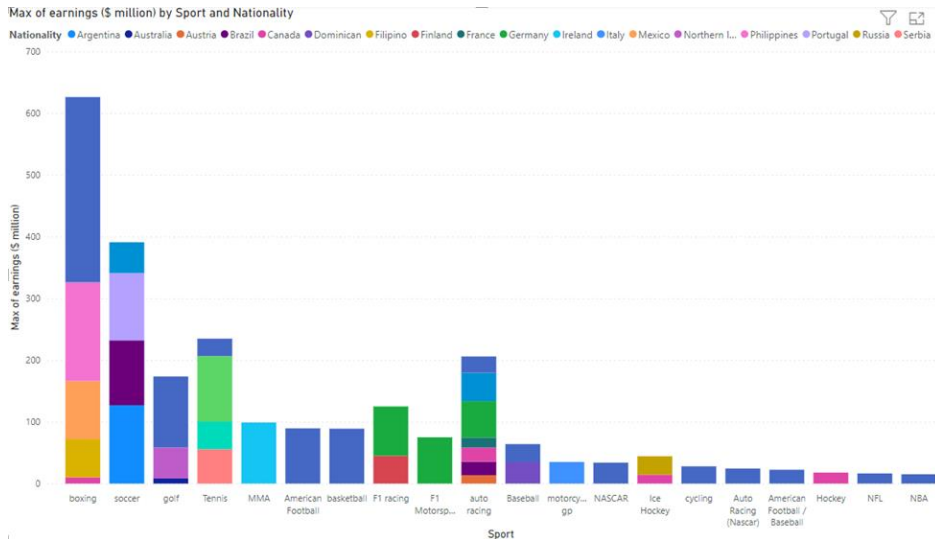
- Plot of maximum and sum of earnings over the years:

  ~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;
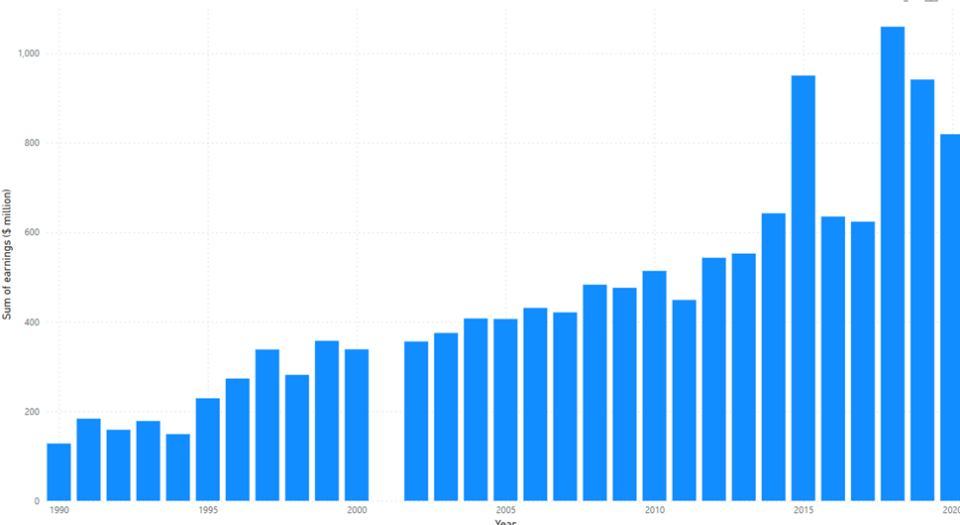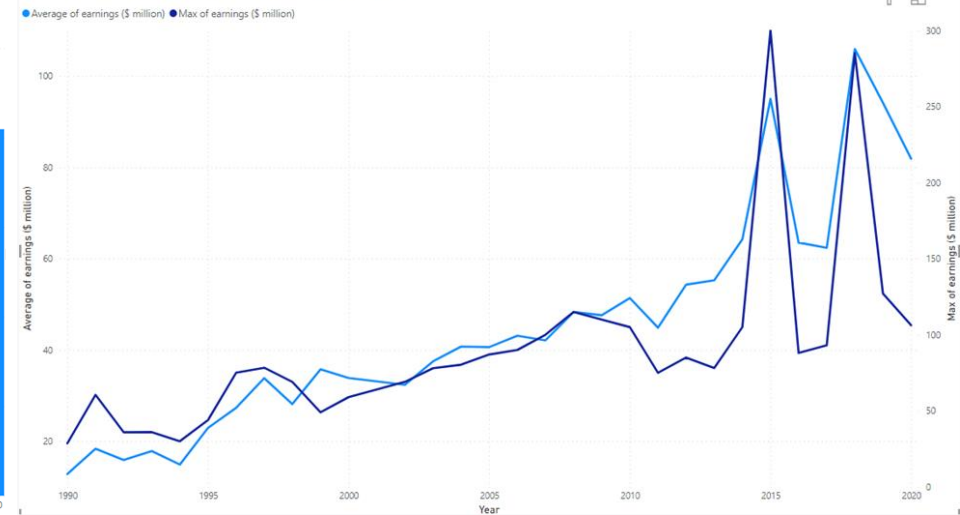
Task 2: Identifying top athletic categories

- 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:

- If we observe, these all players are at the top, mostly ranked 1/2/3, and having a very high earning.

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	Switzerlan	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	Switzerlan	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).

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	Switzerlan	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	Switzerlan	5	7	Tennis	2019	93.4

Conclusion and Submission

- In conclusion, the analysis of the Forbes' highest paid athletes dataset spanning from 1990 to 2020 reveals the trends, patterns and the top 5 athletic categories that have consistently yielded the highest average earnings for athletes. These categories represent a diverse range of sports, showcasing the financial success attained by athletes in these disciplines over the past three decades.
- Regarding Task 1, the data visualizations presented showcase various observable patterns and trends. In Task 2, the analysis determined the top 5 athletic categories to be basketball, boxing, golf, tennis, and soccer.

THANK YOU!