RICHEST ATHLETES FROM 1990-2020

NAME: SMRUTI LOHAR

IMPORTING THE DATA

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
import plotly.express as px
```

```
[ ] df = pd.read_csv('Forbes Richest Atheletes (Forbes Richest Athletes 1990-2020).csv')
```

```
df
```

	S.NO	Name	Nationality	Current Rank	Previous Year Rank	Sport	Year	earnings ($ million)
0	1	Mike Tyson	USA	1	NaN	boxing	1990	28.6
1	2	Buster Douglas	USA	2	NaN	boxing	1990	26.0
2	3	Sugar Ray Leonard	USA	3	NaN	boxing	1990	13.0
3	4	Ayrton Senna	Brazil	4	NaN	auto racing	1990	10.0
4	5	Alain Prost	France	5	NaN	auto racing	1990	9.0
...	...	...	...	...	...	...	...	...
296	297	Stephen Curry	USA	6	9	Basketball	2020	74.4
297	298	Kevin Durant	USA	7	10	Basketball	2020	63.9
298	299	Tiger Woods	USA	8	11	Golf	2020	62.3

CLEANING THE DATA

```
df.isnull().sum()
```

```
S.NO                    0
Name                    0
Nationality             0
Current Rank            0
Previous Year Rank     24
Sport                   0
Year                    0
earnings ($ million)    0
dtype: int64
```

```
df=df.fillna('none')
df.head(3)
```

	S.NO	Name	Nationality	Current Rank	Previous Year Rank	Sport	Year	earnings ($ million)
0	1	Mike Tyson	USA	1	none	boxing	1990	28.6
1	2	Buster Douglas	USA	2	none	boxing	1990	26.0
2	3	Sugar Ray Leonard	USA	3	none	boxing	1990	13.0

EXPLOTARY DATA ANALYSIS

```
df['Nationality'].value_counts()
```

USA	206
UK	13
Germany	13
Switzerland	12
Portugal	10
Brazil	9
Argentina	9
Canada	6
Italy	4
Finland	3
France	3
Philippines	3
Russia	1
Australia	1
Dominican	1
Austria	1
Filipino	1
Spain	1
Serbia	1
Northern Ireland	1
Ireland	1
Mexico	1

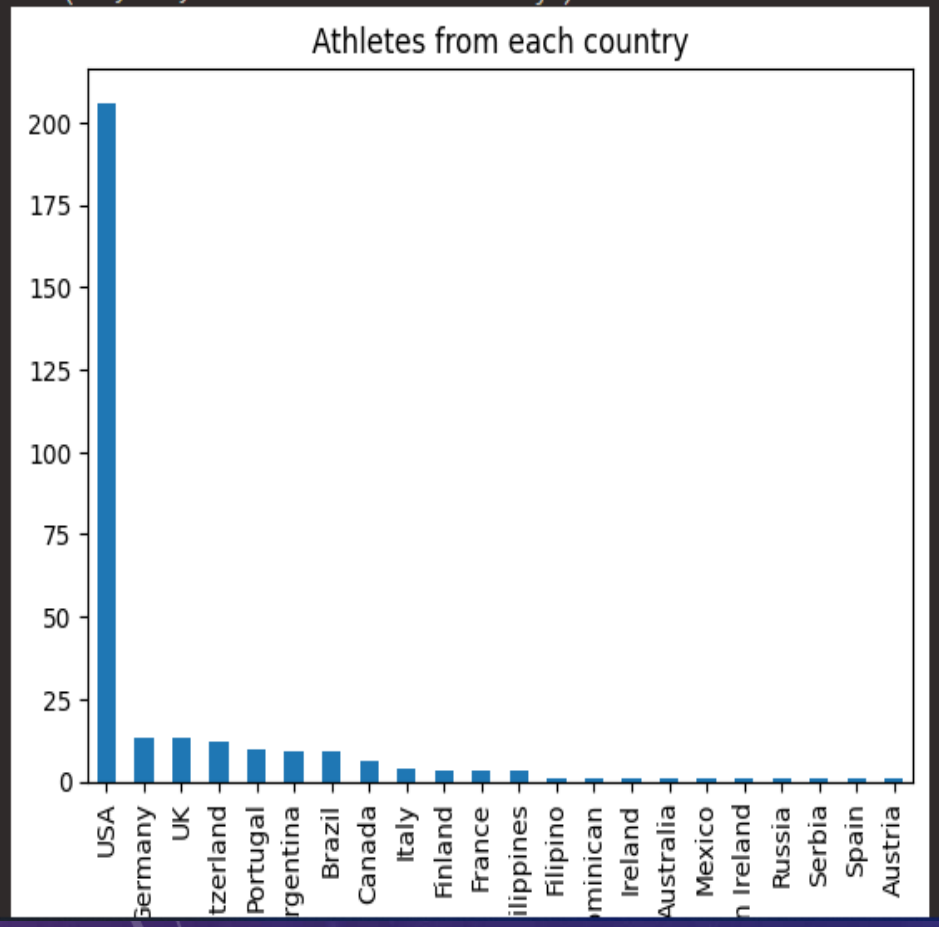
```
df.groupby("Nationality").count()
```

Nationality	S.NO	Name	Current Rank	Previous Year Rank	Sport	Year	earnings ($ million)
Argentina	9	9	9	9	9	9	9
Australia	1	1	1	1	1	1	1
Austria	1	1	1	1	1	1	1
Brazil	9	9	9	9	9	9	9
Canada	6	6	6	6	6	6	6
Dominican	1	1	1	1	1	1	1
Filipino	1	1	1	1	1	1	1
Finland	3	3	3	3	3	3	3
France	3	3	3	3	3	3	3
Germany	13	13	13	13	13	13	13
Ireland	1	1	1	1	1	1	1
Italy	4	4	4	4	4	4	4

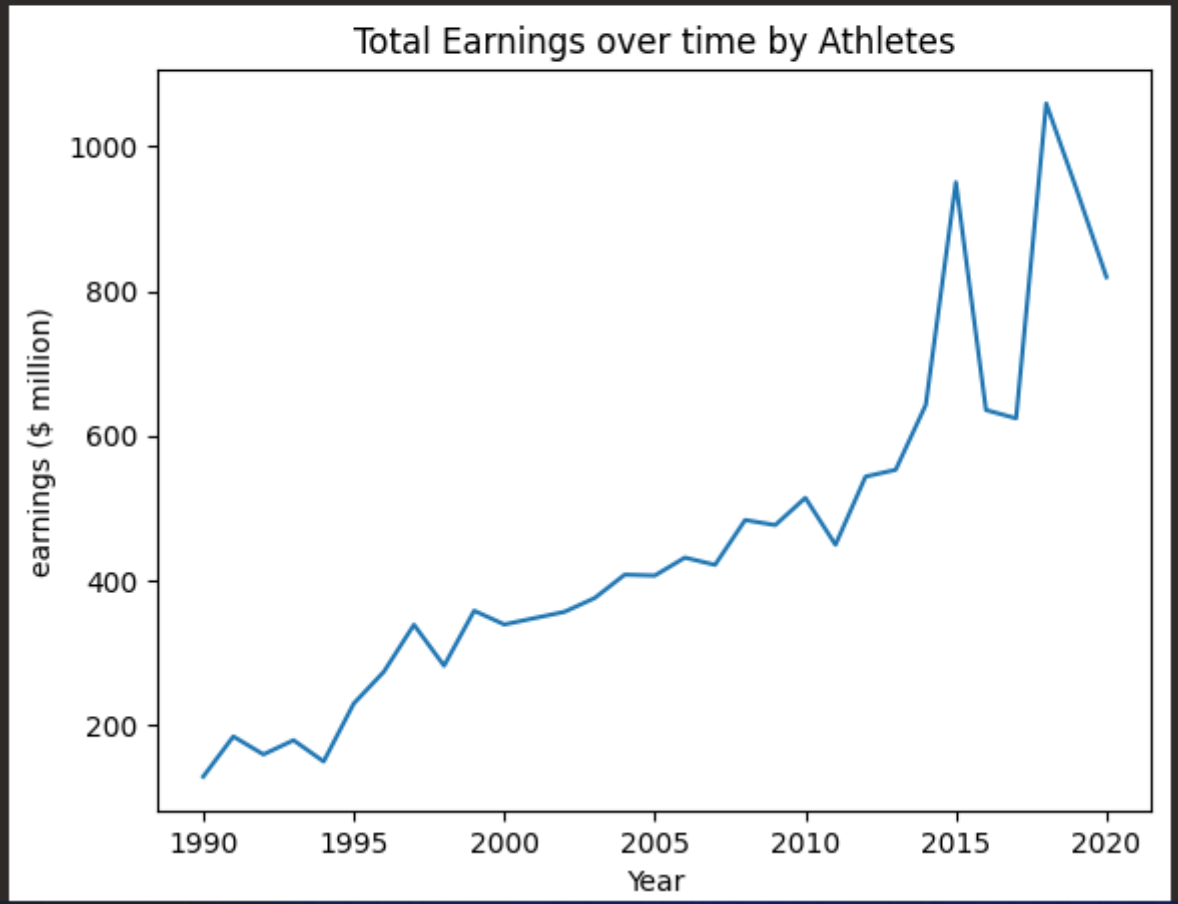
DATA ANALYSIS BY COUNTRY AND YEAR

```
df.groupby("Nationality")["S.NO"].count().sort_values(ascending=False).plot(kind='bar
plt.title("Athletes from each country")
```

```
Text(0.5, 1.0, 'Athletes from each country')
```



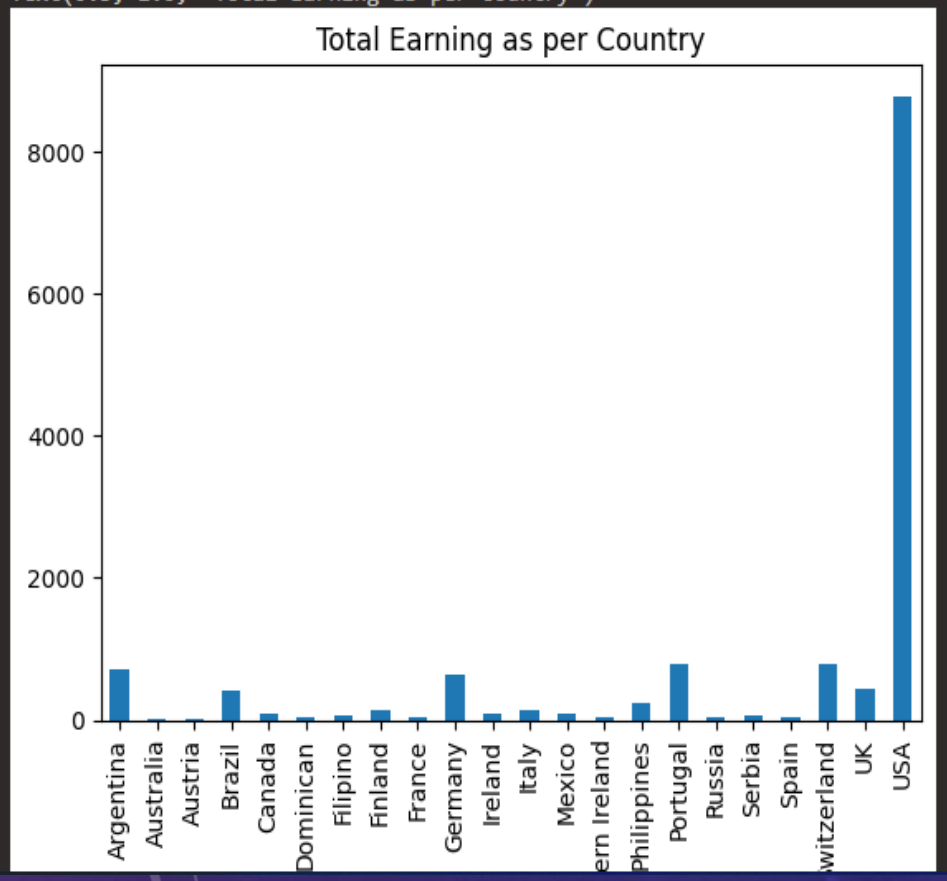

```
plt.plot(df.groupby("Year")["earnings ($ million)"].sum())
plt.xlabel('Year')
plt.ylabel('earnings ($ million)')
plt.title('Total Earnings over time by Athletes')
plt.show()
```



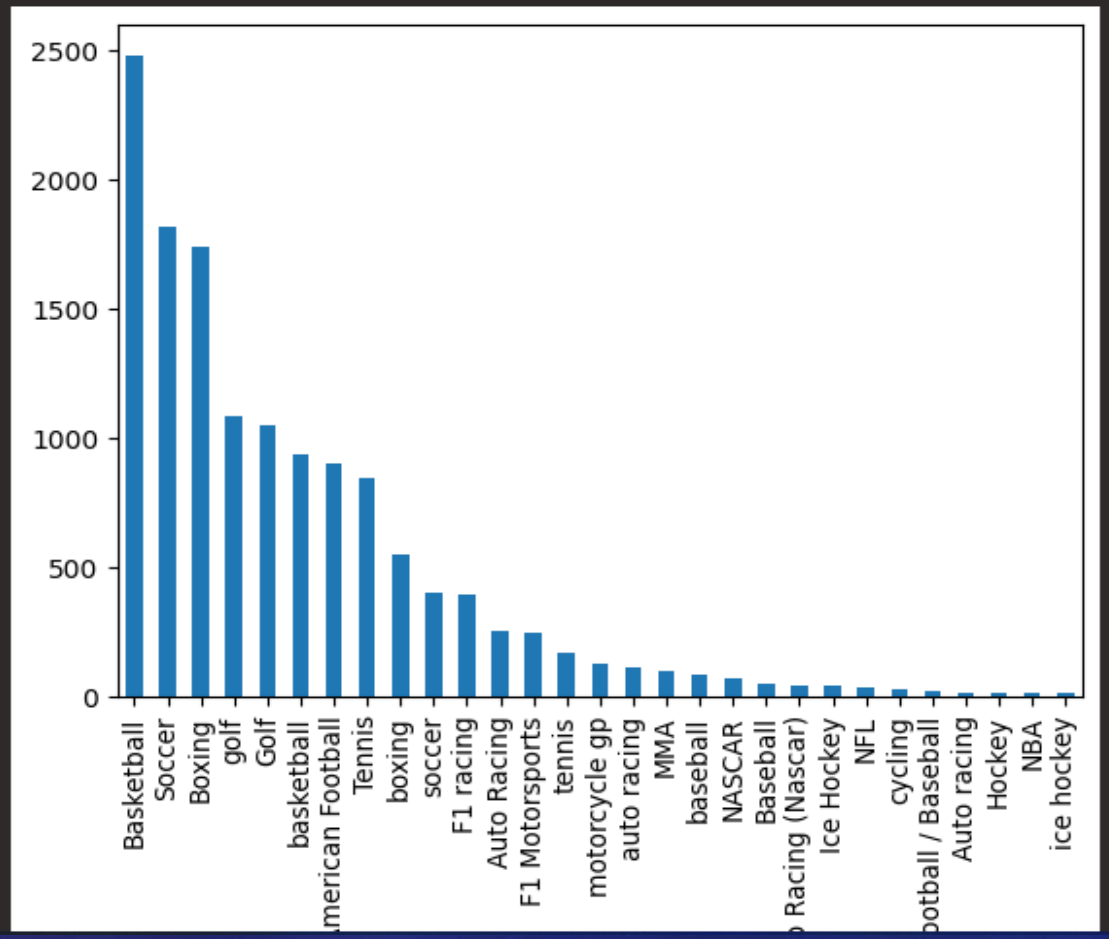
DATA ANALYSIS ON COUNTRY AND SPORTS

```
df.groupby("Nationality")["earnings ($ million)"].sum().plot(kind="bar")
plt.title("Total Earning as per Country")
```

```
Text(0.5, 1.0, 'Total Earning as per Country')
```




```
df.groupby("Sport")["earnings ($ million)"].sum().sort_values(ascending = False).plot(kind="bar")
plt.show()
```

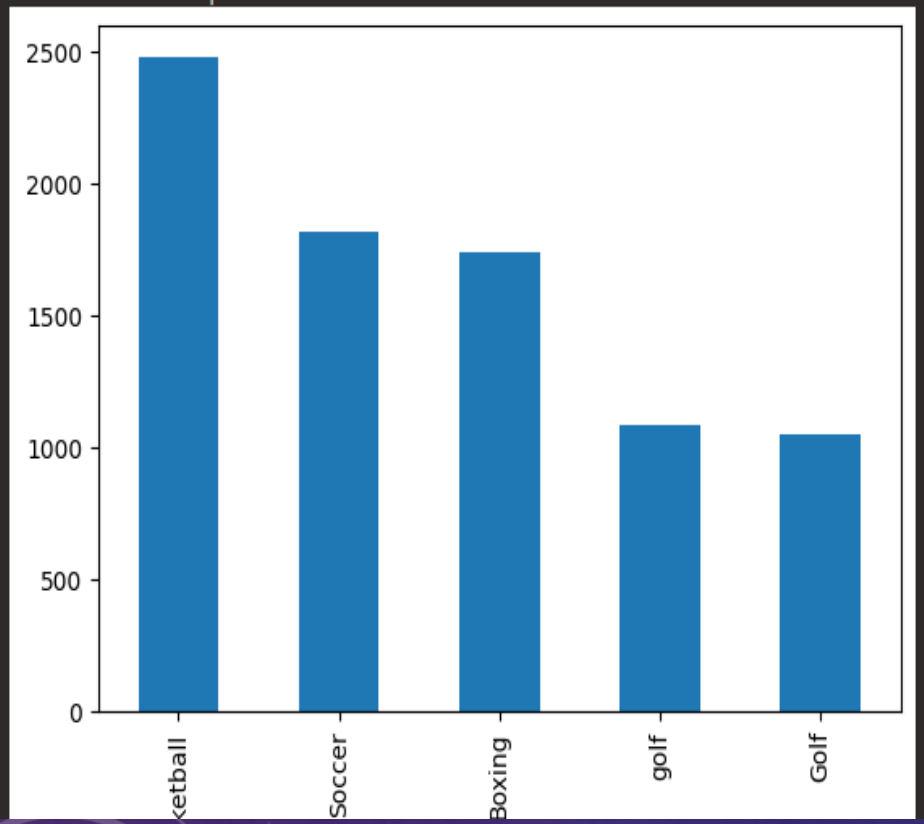


HIGHEST PAID SPORTS AND ATHLETES

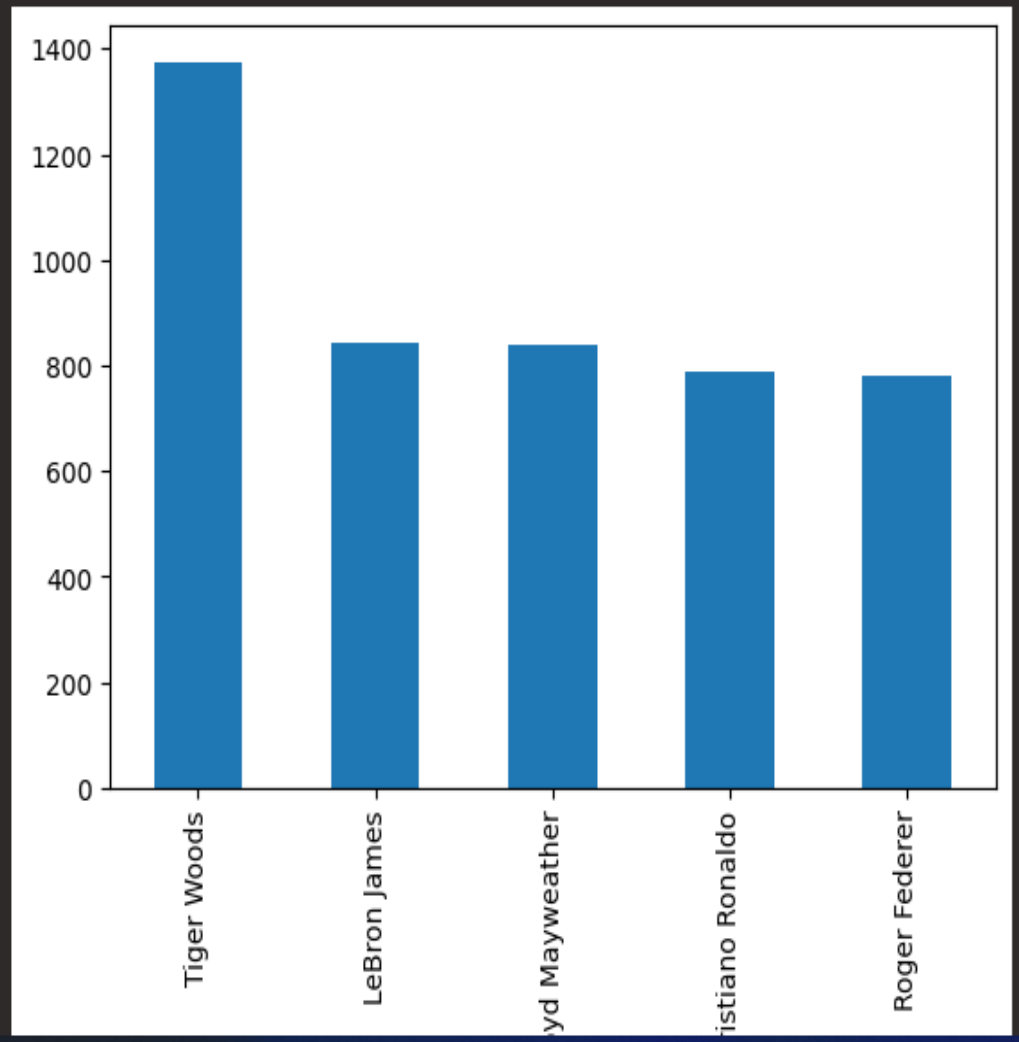
```
df1=df.groupby("Sport")["earnings ($ million)"].sum().sort_values(ascending = False)

# Top 5 Highest Paid Sports
df1.head().plot(kind="bar")
```

```
<Axes: xlabel='Sport'>
```




```
df.groupby("Name")["earnings ($ million)"].sum().sort_values(ascending=False).head().plot(kind="bar"
plt.show()
```



TOP 10 HIGHEST PAID ATHLETES IN A YEAR

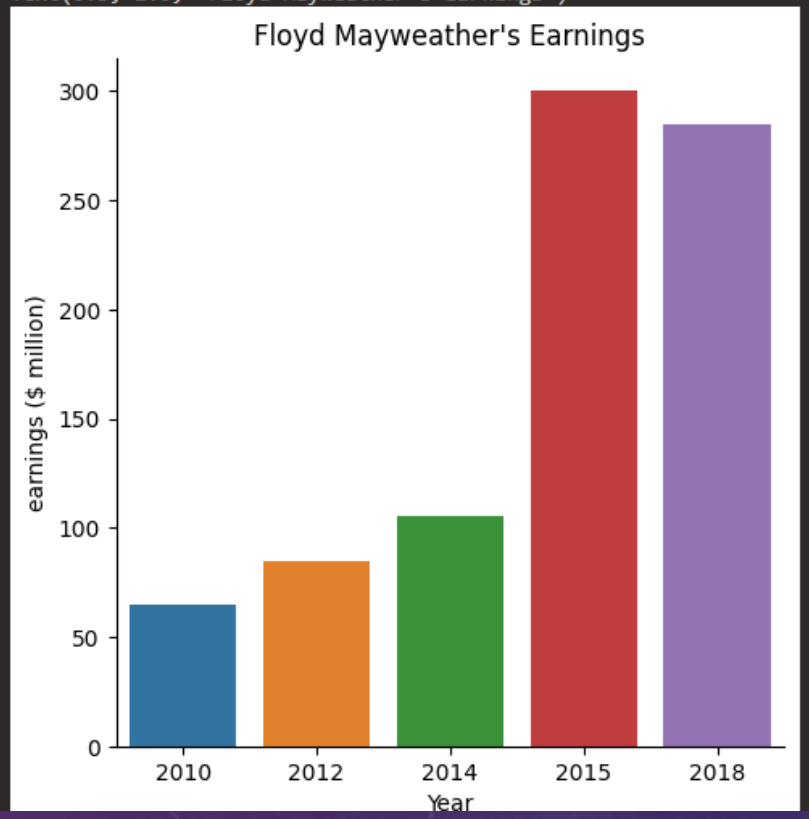
```
#Top 10 highest paid Athletes (1990-2020)
top_ten=df.sort_values('earnings ($ million)',ascending=False).head(10)
top_ten
```

	S.NO	Name	Nationality	Current Rank	Previous Year Rank	Sport	Year	earnings ($ million)
241	242	Floyd Mayweather	USA	1	1	Boxing	2015	300.0
271	272	Floyd Mayweather	USA	1	>100	Boxing	2018	285.0
242	243	Manny Pacquiao	Philippines	2	11	Boxing	2015	160.0
281	282	Lionel Messi	Argentina	1	2	Soccer	2019	127.0
171	172	Tiger Woods	USA	1	1	golf	2008	115.0
272	273	Lionel Messi	Argentina	2	3	Soccer	2018	111.0
181	182	Tiger Woods	USA	1	1	golf	2009	110.0
282	283	Cristiano Ronaldo	Portugal	2	3	Soccer	2019	109.0
273	274	Cristiano Ronaldo	Portugal	3	1	Soccer	2018	108.0
291	292	Roger Federer	Switzerland	1	5	Tennis	2020	106.3

DATA OF FLOYD EVERY YEAR

```
sns.catplot(x='Year',y='earnings ($ million)',kind='bar',data=floyd_mayweather)
plt.title("Floyd Mayweather's Earnings")
```

```
Text(0.5, 1.0, "Floyd Mayweather's Earnings")
```



HIGHEST PAID ATHLETE FROM 1990-2020

```
total_earnings_atheletes = pd.DataFrame(df.groupby("Name")["earnings ($ million)"].sum())
```

```
total_earnings_atheletes.sort_values(by='earnings ($ million)',ascending=False)
```

Name	earnings ($ million)
Tiger Woods	1373.8
LeBron James	844.8
Floyd Mayweather	840.0
Cristiano Ronaldo	787.1
Roger Federer	781.1
...	...
Cecil Fielder	12.7
Michael Moorer	12.1
Donovan "Razor" Ruddock	10.2
Greg Norman	8.5
Monica Seles	8.5

DATA OF TIGER WOODS EVERY YEAR

```
sns.catplot(x='Year',y='earnings ($ million)',kind='bar',data=tiger_woods)
plt.title("Tiger Woods's Earnings")
plt.xticks(rotation=90)
plt.show()
```