Submission I - Case Study

ON

Tokyo Olympics 2021

NAME: MANPREET SINGH

REG NO.: 202103680

SUBMITTED TO SCDL

SYMBIOSIS
Centre for Distance Learning (SCDL)

In partial fulfillment of the requirements for the award of degree of

Post Graduate Diploma in Data Science

DECLARATION

This is to declare that I ***"Manpreet Singh"*** have carried out this project work myself in part fulfillment of the ***Post Graduate Diploma in Data Science*** Program of SCDL.

The work is original, has not been copied from anywhere else and has not been submitted to any other University/Institute for an award of any degree/diploma.

Name: Manpreet Singh ***Date: 01.02.2023***

Signature: ***Reg No: 202103680***

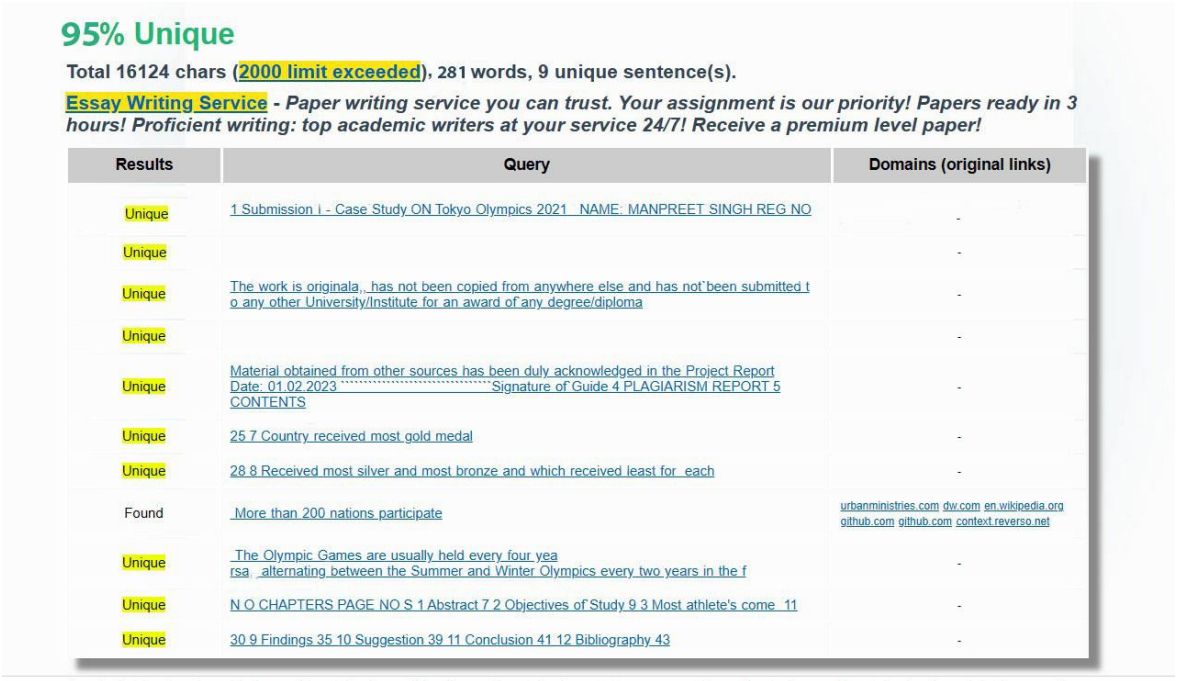
CERTIFICATE OF SUPERVISOR

Certified that the work incorporated in this Project Report on ***Tokyo Olympics 2021*** submitted by ***Manpreet Singh*** is his/her original work and completed under my supervision.

Material obtained from other sources has been duly acknowledged in the Project Report

Date: 01.02.2023 ***Signature of Guide***

PLAGIARISM REPORT



95% Unique

Total 16124 chars (2000 limit exceeded), 281 words, 9 unique sentence(s).

Essay Writing Service - Paper writing service you can trust. Your assignment is our priority! Papers ready in 3 hours! Proficient writing: top academic writers at your service 24/7! Receive a premium level paper!

Results	Query	Domains (original links)
Unique	1 Submission i - Case Study ON Tokyo Olympics 2021 NAME: MANPREET SINGH REG NO	-
Unique		-
Unique	The work is originala,, has not been copied from anywhere else and has not`been submitted t o any other University/Institute for an award of`any degree/diploma	-
Unique		-
Unique	Material obtained from other sources has been duly acknowledged in the Project Report Date: 01.02.2023 ``````````````````````````````Signature of`Guide 4 PLAGIARISM REPORT 5 CONTENTS	-
Unique	25 7 Country received most gold medal	-
Unique	28 8 Received most silver and most bronze and which received least for each	-
Found	More than 200 nations participate	urbanministries.com dw.com en.wikipedia.org github.com github.com context.reverso.net
Unique	The Olympic Games are usually held every four yea rsa alternating between the Summer and Winter Olympics every two years in the f	-
Unique	N O CHAPTERS PAGE NO S 1 Abstract 7 2 Objectives of`Study 9 3 Most athlete's come 11	-
Unique	30 9 Findings 35 10 Suggestion 39 11 Conclusion 41 12 Bibliography 43	-

CONTENTS		
S.NO	***CHAPTERS***	***PAGE NOS***
1	***Abstract***	***7***
2	***Objectives of Study***	***9***
3	***Most athlete's come***	***11***
4	***Country has the highest participants in it***	***18***
5	***Produces highest number of coaches***	***23***
6	***Gender across disciplines.***	***25***
7	***Country received most gold medal?***	***28***
8	***Received most silver and most bronze and which received least for each?***	***30***
9	***Findings***	***35***
10	***Suggestion***	***39***
11	***Conclusion***	***41***
12	***Bibliography***	***43***

Abstract

ABSTRACT

Olympics is one of the biggest sporting events featuring summer and winter sports competitions where thousands of athletes from around the world participate in a variety of competitions. More than 200 nations participate. The Olympic Games are usually held every four years, alternating between the Summer and Winter Olympics every two years in the four-year period. The International Olympic Committee (IOC) as formed in 1894 and the first modern Olympics was held in 1896.

Objectives of Study

OBJECTIVES OF THE STUDY

- **Which countries do most athlete's come from?**
- **Which Discipline is most popular and which country has the highest participants in it?**
- **Which country produces highest number of coaches?**
- **Gender across disciplines.**
- **Which country received most gold medal?**
- **Which received most silver and most bronze and which received least for each?**

Which countries do most athlete's come from?

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
```

In [2]:

```
athlete_data = pd.read_csv('Athletes.csv')
coach_data = pd.read_csv('Coaches.csv')
gender_data = pd.read_csv('EntriesGender.csv')
medal_data = pd.read_csv('Medals.csv')
team_data = pd.read_csv('Teams.csv')
```

In [3]:

```
athlete_data.head()
```

Out[3]:

	Name	NOC	Discipline
0	AALERUD Katrine	Norway	Cycling Road
1	ABAD Nestor	Spain	Artistic Gymnastics
2	ABAGNALE Giovanni	Italy	Rowing
3	ABALDE Alberto	Spain	Basketball
4	ABALDE Tamara	Spain	Basketball

In [4]:

```
gender_data.head()
```

Out[4]:

	Discipline	Female	Male	Total
0	3x3 Basketball	32	32	64
1	Archery	64	64	128
2	Artistic Gymnastics	98	98	196
3	Artistic Swimming	105	0	105
4	Athletics	969	1072	2041

In [5]:

```
medal_data.head()
```

Out[5]:

	Rank	Team/NOC	Gold	Silver	Bronze	Total	Rank by Total
0	1	United States of America	39	41	33	113	1
1	2	People's Republic of China	38	32	18	88	2
2	3	Japan	27	14	17	58	5
3	4	Great Britain	22	21	22	65	4
4	5	ROC	20	28	23	71	3

In [6]:

team_data.head()

Out[6]:

	Name	Discipline	NOC	Event
0	Belgium	3x3 Basketball	Belgium	Men
1	China	3x3 Basketball	People's Republic of China	Men
2	China	3x3 Basketball	People's Republic of China	Women
3	France	3x3 Basketball	France	Women
4	Italy	3x3 Basketball	Italy	Women

In [7]:

athlete_data.isnull().sum()

Out[7]:

Name 0
NOC 0
Discipline 0
dtype: int64

In [8]:

gender_data.isnull().sum()

Out[8]:

Discipline 0
Female 0
Male 0
Total 0
dtype: int64

In [9]:

medal_data.isnull().sum()

Out[9]:

```
Rank            0
Team/NOC        0
Gold            0
Silver          0
Bronze          0
Total           0
Rank by Total   0
dtype: int64
```

In [10]:

```
coach_data.isnull().sum()
```

Out[10]:

```
Name            0
NOC             0
Discipline      0
Event         145
dtype: int64
```

- **Which countries do most athlete's come from?**

In [11]:

```
athlete_data.head()
```

Out[11]:

	Name	NOC	Discipline
0	AALERUD Katrine	Norway	Cycling Road
1	ABAD Nestor	Spain	Artistic Gymnastics
2	ABAGNALE Giovanni	Italy	Rowing
3	ABALDE Alberto	Spain	Basketball
4	ABALDE Tamara	Spain	Basketball

In [12]:

```
athlete_data['NOC'].value_counts()
```

Out[12]:

```
United States of America       615
Japan                          586
Australia                      470
People's Republic of China     401
Germany                        400
                              ...
United Republic of Tanzania      2
Saint Kitts and Nevis            2
```

Marshall Islands 2

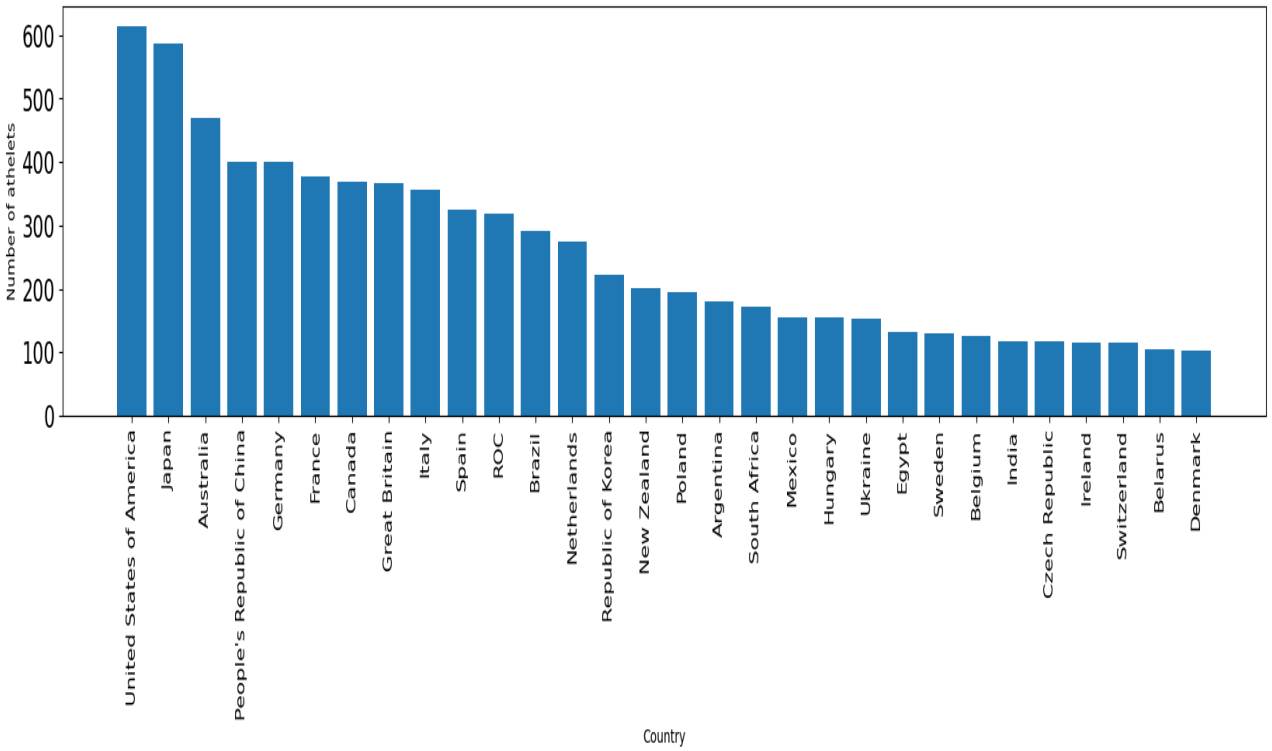
Vanuatu 2

South Sudan 2

Name: NOC, Length: 206, dtype: int64

In [13]:

```
y = athlete_data.NOC.value_counts().values[:30]
x = athlete_data.NOC.value_counts().index[:30]
plt.figure(figsize=(20,4))
plt.bar(x,y)
plt.xlabel("Country")
plt.ylabel("Number of athelets")
plt.xticks(rotation="vertical",size=12)
plt.yticks(size=16)
plt.show()
```



Which Discipline is most popular and which country has the highest participants in it?

- **Which Discipline is most popular and which country has the highest participants in it?**

In [14]:

```
athlete_data.head()
```

Out[14]:

	Name	NOC	Discipline
0	AALERUD Katrine	Norway	Cycling Road
1	ABAD Nestor	Spain	Artistic Gymnastics
2	ABAGNALE Giovanni	Italy	Rowing
3	ABALDE Alberto	Spain	Basketball
4	ABALDE Tamara	Spain	Basketball

In [15]:

```
athlete_data["Discipline"].value_counts()[:15]
```

Out[15]:

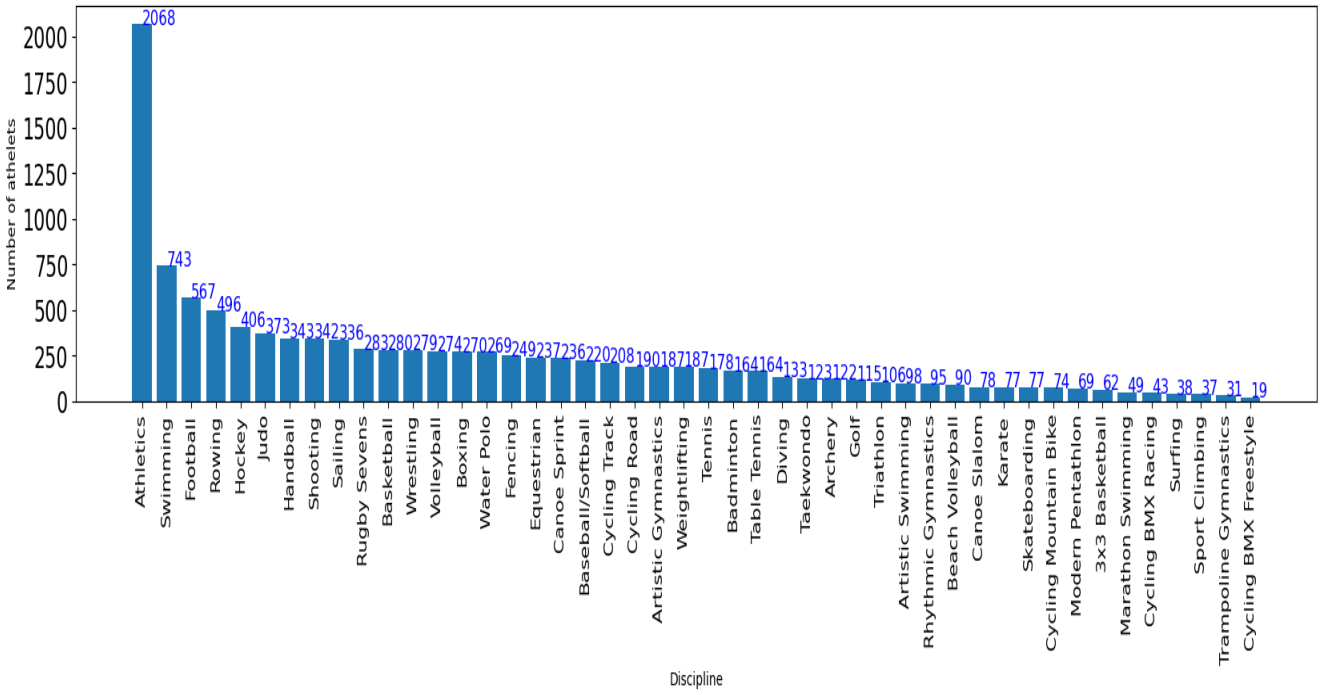
```
Athletics       2068
Swimming         743
Football         567
Rowing           496
Hockey           406
Judo             373
Handball         343
Shooting         342
Sailing          336
Rugby Sevens     283
```

```
Basketball       280
Wrestling        279
Volleyball       274
Boxing           270
Water Polo       269
Name: Discipline, dtype: int64
```

In [19]:

```
y = athlete_data.Discipline.value_counts().values
x = athlete_data.Discipline.value_counts().index
plt.figure(figsize=(20,4))
for index,value in enumerate(y):
    plt.text(index,value,str(value),color = "blue",size = 12)

plt.bar(x,y)
plt.xlabel("Discipline")
plt.ylabel("Number of athelets")
plt.xticks(rotation="vertical",size=12)
plt.yticks(size=16)
plt.show()
```



In [22]:

```
athlete_datas = team_data[(team_data["Discipline"]=="Athletics")]
athlete_datas["NOC"].value_counts()
```

Out[22]:

```
United States of America    5
Germany                     5
Italy                       5
Netherlands                 5
Great Britain               5
Jamaica                     5
Poland                      4
France                      4
Trinidad and Tobago         3
Japan                       3
Brazil                      3
Belgium                     3
```

Canada 2
People's Republic of China 2
Switzerland 2
South Africa 2
Nigeria 2
India 2
Ukraine 2
Denmark 2
Spain 1
Turkey 1
Australia 1
Ireland 1
Bahamas 1
Ecuador 1
Dominican Republic 1
Czech Republic 1
Cuba 1
Colombia 1
Botswana 1
Belarus 1
Ghana 1
Name: NOC, dtype: int64

Which country produces highest number of coaches?

- **Which country produces highest number of coaches?**

In [29]:

```
coach_data.isnull().sum()
```

Out[29]:

```
Name          0
NOC           0
Discipline    0
Event       145
dtype: int64
```

In [30]:

```
coach_data.NOC.value_counts()
```

Out[30]:

```
Japan                     35
United States of America  28
Spain                     28
Australia                 22
Canada                    16
                          ..
Portugal                   1
Belarus                    1
Slovakia                   1
Croatia                    1
Liechtenstein              1
Name: NOC, Length: 61, dtype: int64
```

Gender across disciplines.

- **Gender across disciplines.**

In [32]:

```
gender_data.head()
```

Out[32]:

	Discipline	Female	Male	Total
0	3x3 Basketball	32	32	64
1	Archery	64	64	128
2	Artistic Gymnastics	98	98	196
3	Artistic Swimming	105	0	105
4	Athletics	969	1072	2041

In [33]:

```
gender_data.isnull().sum()
```

Out[33]:

```
Discipline    0
Female        0
Male          0
Total         0
dtype: int64
```

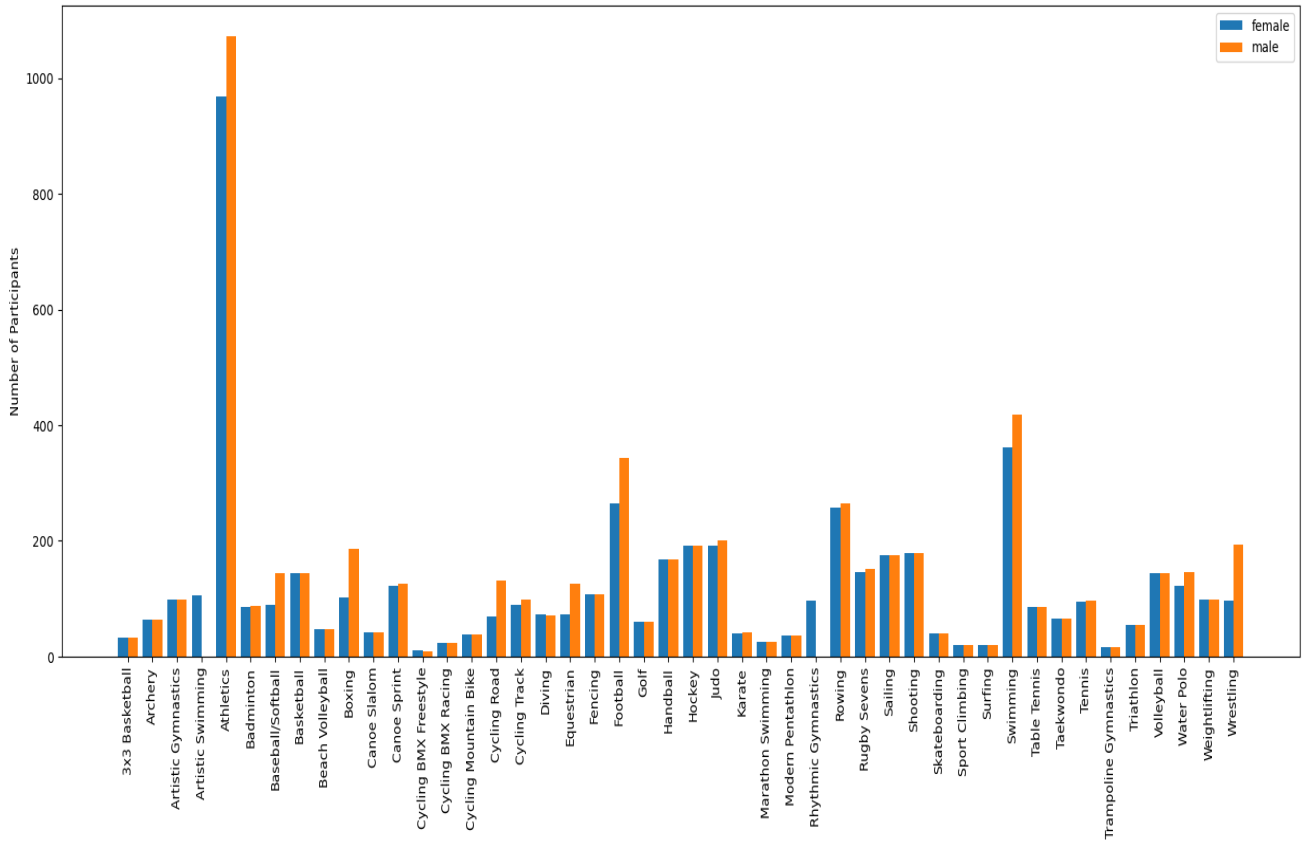
In [45]:

```
Disc = gender_data.groupby("Discipline")
x = [Discipline for Discipline,df in Disc]
female = gender_data.Female
male = gender_data.Male
plt.figure(figsize=(20,8))
```

```
x_axis = np.arange(len(x))
plt.bar(x_axis-0.2,female,0.4,label = "female")
plt.bar(x_axis+0.2,male,0.4,label = "male")
plt.ylabel("Number of Participants")
plt.xticks(x_axis,x,rotation="vertical")
plt.legend()
plt.show
```

Out[45]:

<function matplotlib.pyplot.show(close=None, block=None)>



Which country received most gold medal ? Which received most silver and most bronze and which received least for each?

- **Which country received most gold medal ? Which received most silver and most bronze and which received least for each?**

In [46]:

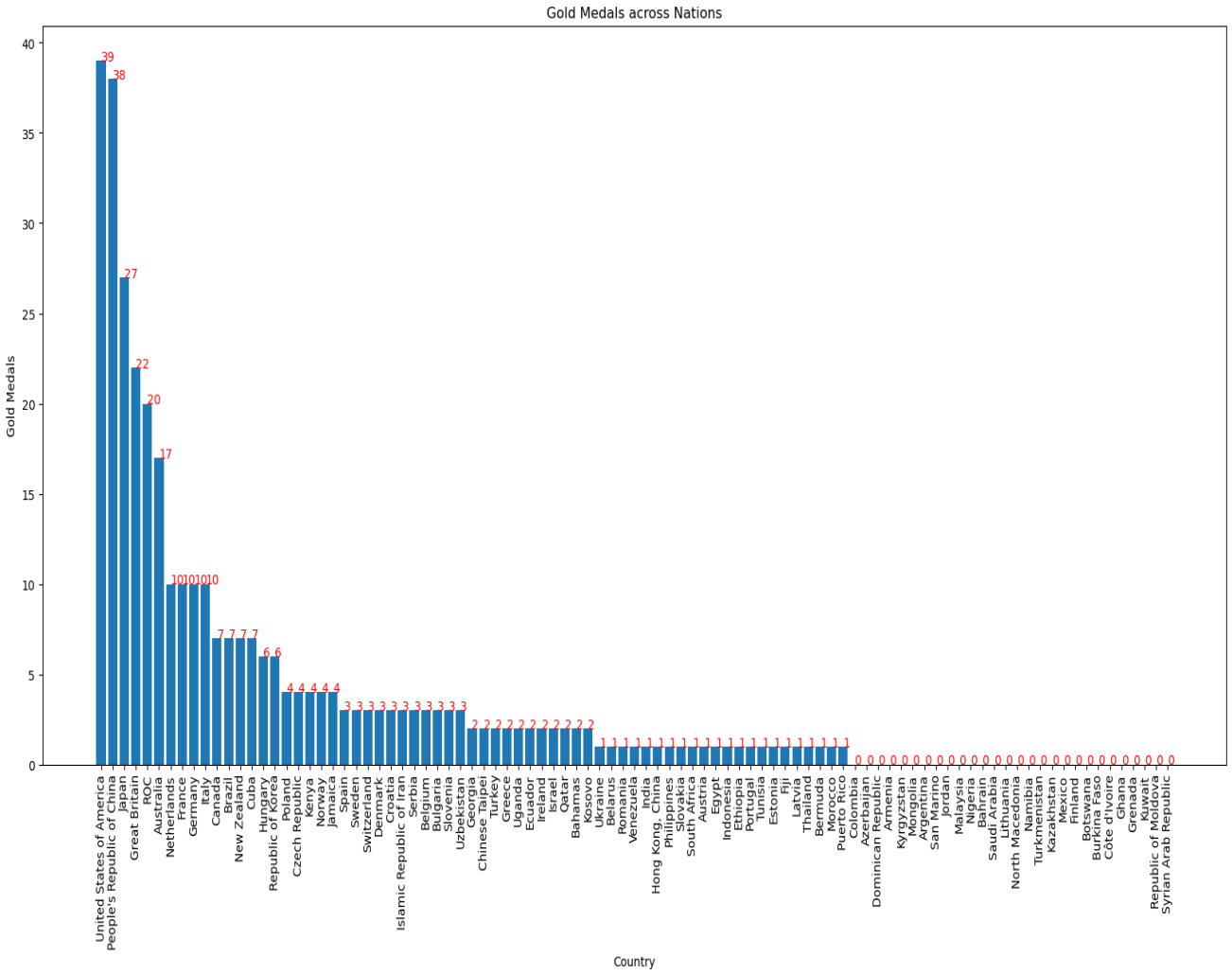
```
medal_data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 93 entries, 0 to 92
Data columns (total 7 columns):
 #   Column         Non-Null Count  Dtype
---  ------         --------------  -----
 0   Rank           93 non-null     int64
 1   Team/NOC       93 non-null     object
 2   Gold           93 non-null     int64
 3   Silver         93 non-null     int64
 4   Bronze         93 non-null     int64
 5   Total          93 non-null     int64
 6   Rank by Total  93 non-null     int64
dtypes: int64(6), object(1)
memory usage: 5.2+ KB
```

In [62]:

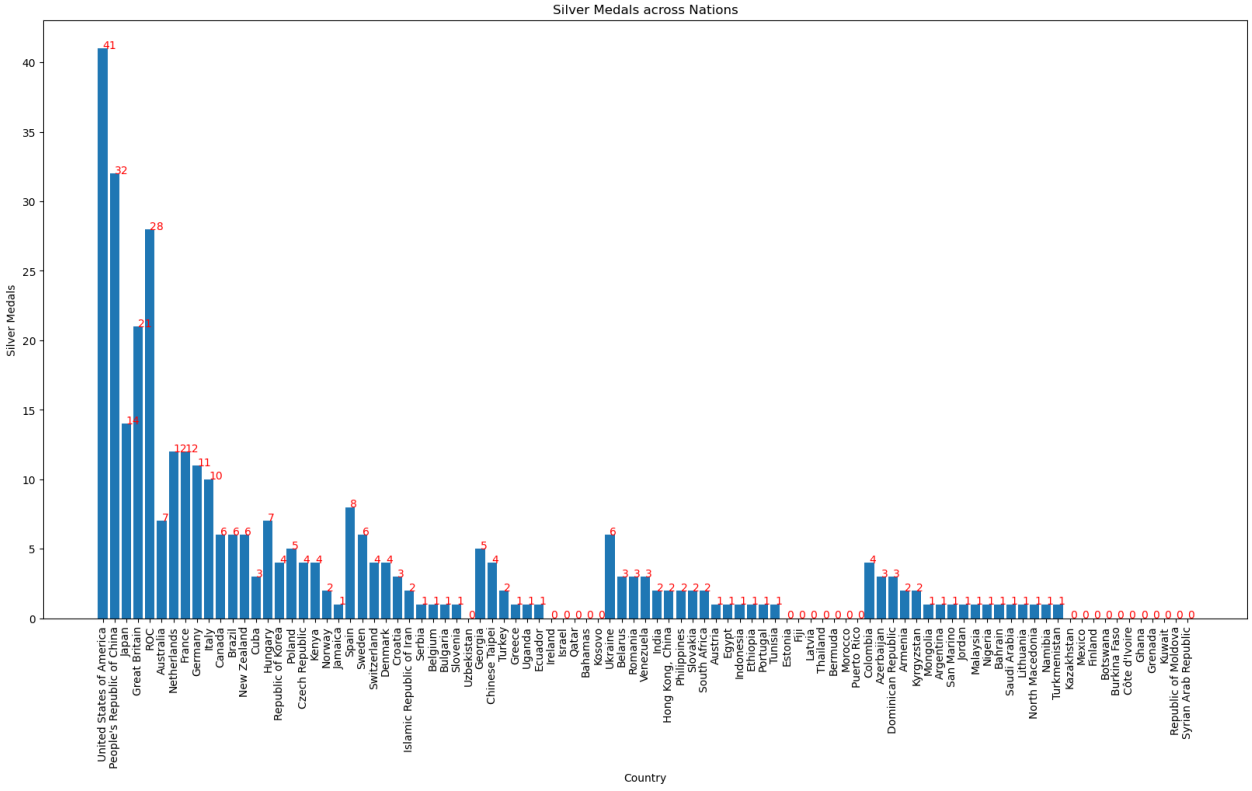
```
x = []
for team in medal_data["Team/NOC"]:
    x.append(team)
y = medal_data.Gold
plt.figure(figsize=(20,10))
plt.bar(x,y)
for index,value in enumerate(y):
   plt.text(index,value,str(value),color="Red",size=10)
plt.xlabel("Country")
plt.ylabel("Gold Medals")
```

```
plt.xticks(x,rotation = "vertical")
plt.title("Gold Medals across Nations")
plt.show()
```



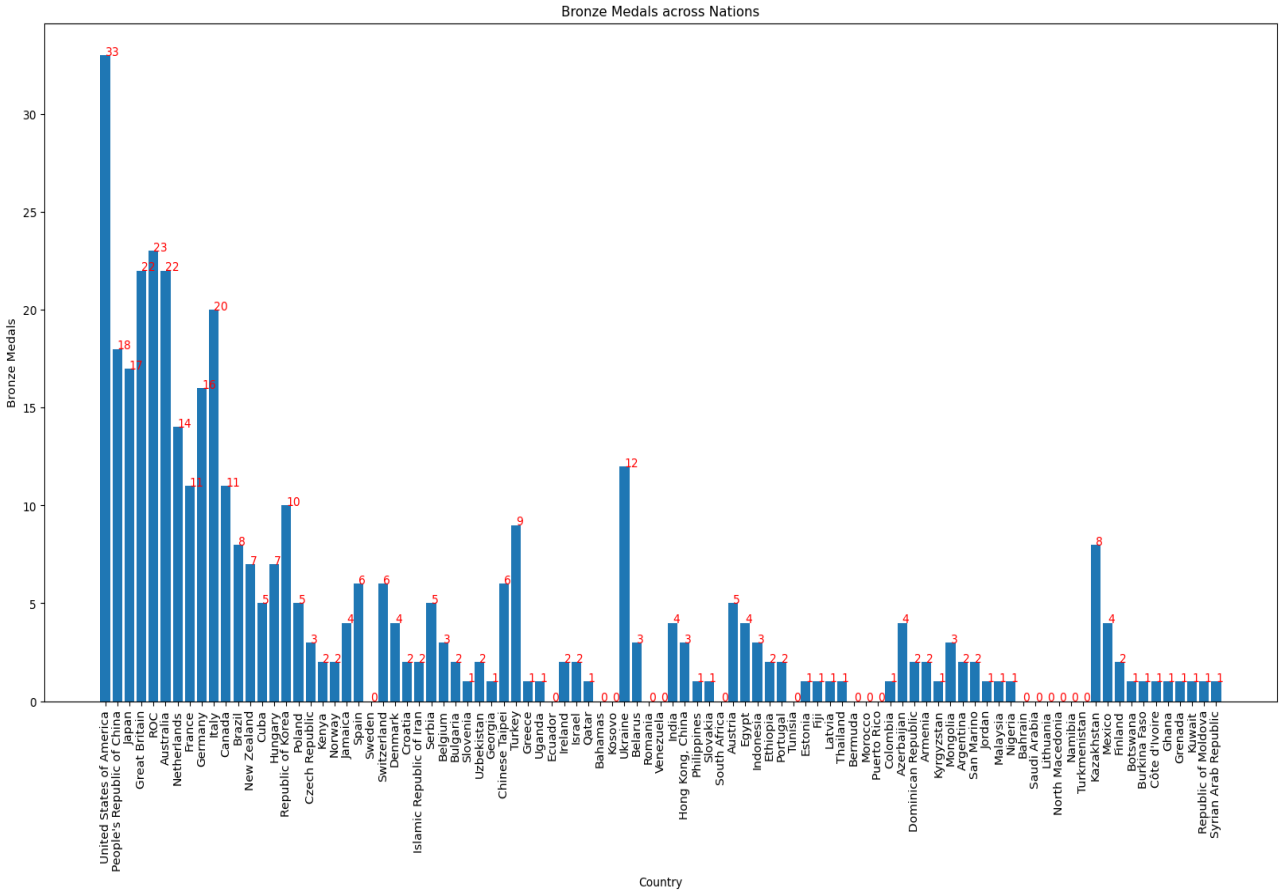
In [63]:

```
x = []
for team in medal_data["Team/NOC"]:
    x.append(team)
y = medal_data.Silver
plt.figure(figsize=(20,10))
plt.bar(x,y)
for index,value in enumerate(y):
   plt.text(index,value,str(value),color="Red",size=10)
plt.xlabel("Country")
plt.ylabel("Silver Medals")
plt.xticks(x,rotation = "vertical")
plt.title("Silver Medals across Nations")
plt.show()
```



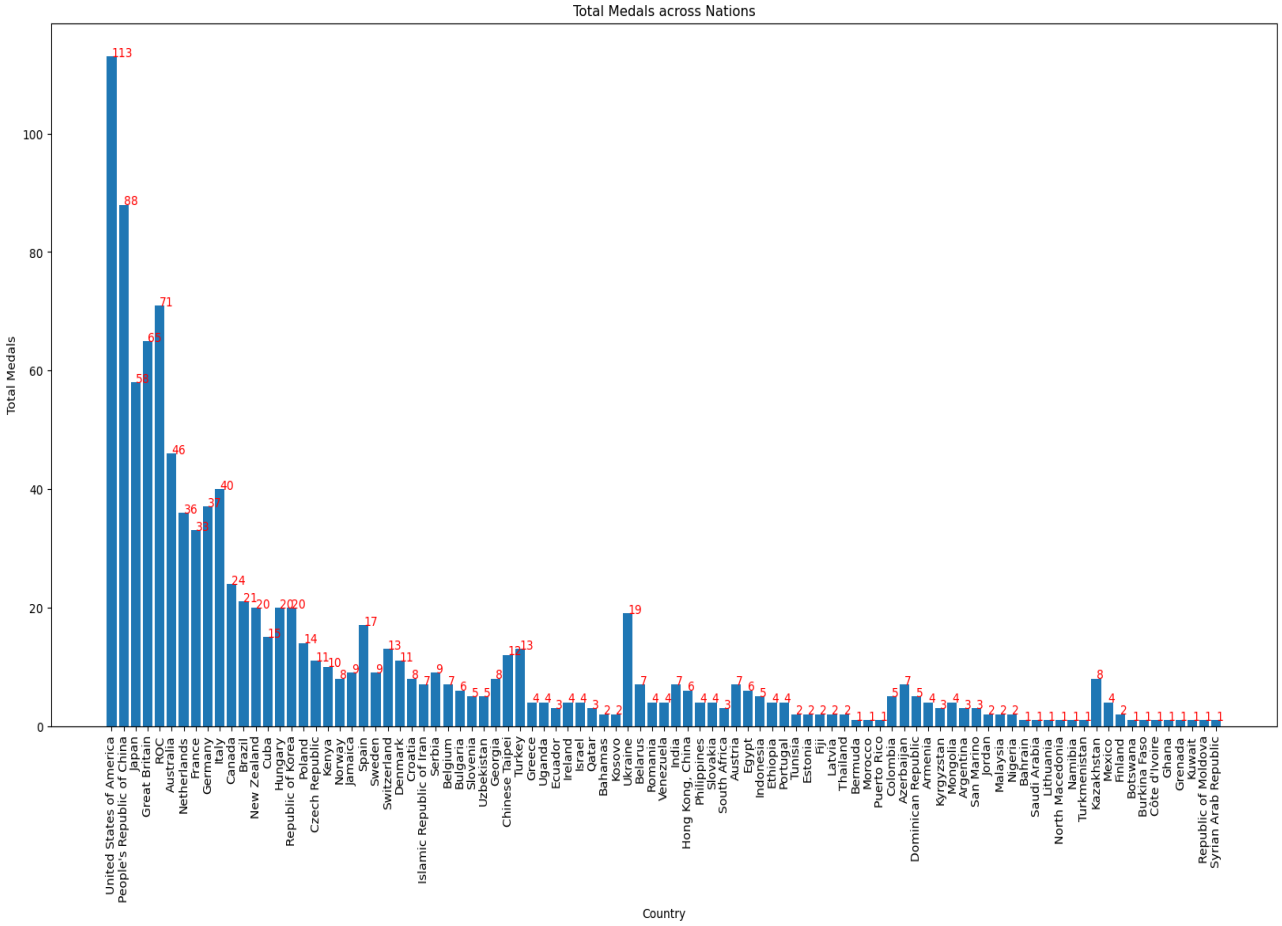
In [61]:

```
x = []
for team in medal_data["Team/NOC"]:
    x.append(team)
y = medal_data.Bronze
plt.figure(figsize=(20,10))
plt.bar(x,y)
for index,value in enumerate(y):
   plt.text(index,value,str(value),color="Red",size=10)
plt.xlabel("Country")
plt.ylabel("Bronze Medals")
plt.xticks(x,rotation = "vertical")
plt.title("Bronze Medals across Nations")
plt.show()
```



In [64]:

```
x = []
for team in medal_data["Team/NOC"]:
    x.append(team)
y = medal_data.Total
plt.figure(figsize=(20,10))
plt.bar(x,y)
for index,value in enumerate(y):
   plt.text(index,value,str(value),color="Red",size=10)
plt.xlabel("Country")
plt.ylabel("Total Medals")
plt.xticks(x,rotation = "vertical")
plt.title("Total Medals across Nations")
plt.show()
```


Total Medals across Nations
Total Medals
Country
United States of America
People's Republic of China
Japan
Great Britain
ROC
Australia
Netherlands
France
Germany
Italy
Canada
Brazil
New Zealand
Cuba
Hungary
Republic of Korea
Poland
Czech Republic
Kenya
Norway
Jamaica
Spain
Sweden
Switzerland
Denmark
Croatia
Islamic Republic of Iran
Serbia
Belgium
Bulgaria
Slovenia
Uzbekistan
Georgia
Chinese Taipei
Türkiye
Greece
Uganda
Ecuador
Ireland
Israel
Qatar
Bahamas
Kosovo
Ukraine
Belarus
Romania
Venezuela
India
Hong Kong, China
Philippines
Slovakia
South Africa
Austria
Egypt
Indonesia
Ethiopia
Portugal
Tunisia
Estonia
Fiji
Latvia
Thailand
Bermuda
Morocco
Puerto Rico
Colombia
Azerbaijan
Dominican Republic
Armenia
Kyrgyzstan
Mongolia
Argentina
San Marino
Jordan
Malaysia
Nigeria
Bahrain
Saudi Arabia
Lithuania
North Macedonia
Namibia
Turkmenistan
Kazakhstan
Mexico
Finland
Botswana
Burkina Faso
Côte d'Ivoire
Ghana
Grenada
Kuwait
Republic of Moldova
Syrian Arab Republic

In [ ]:

FINDINGS

FINDINGS

- ***Most athlete's come***

- ✓ United States of America 615
- ✓ Japan 586
- ✓ Australia 470
- ✓ People's Republic of China 401
- ✓ Germany 400

- ***Country has the highest participants in it***.

- ✓ Athletics 2068
- ✓ Swimming 743
- ✓ Football 567
- ✓ Rowing 496
- ✓ Hockey 406
- ✓ Judo 373
- ✓ Handball 343
- ✓ Shooting 342
- ✓ Sailing 336
- ✓ Rugby Sevens 283
- ✓ Basketball 280
- ✓ Wrestling 279
- ✓ Volleyball 274
- ✓ Boxing 270
- ✓ Water Polo 269

- ***Produces highest number of coaches***
  - ✓ Japan 35
  - ✓ United States of America 28
  - ✓ Spain 28
  - ✓ Australia 22
  - ✓ Canada 16

- ***Gender across disciplines***

Discipline	Female	Male	Total
3x3 Basketball	32	32	64
Archery	64	64	128
Artistic Gymnastics	98	98	196
Artistic Swimming	105	0	105
Athletics	969	1072	2041

- ***Which country received most gold medal ? Which received most silver and most bronze and which received least for each?***

  - ✓ ***most gold medal country received-*** **United States of America, People's Republic of China and Japan**
  - ✓ ***most Sllver medal country received-*** ***United States of America, People's Republic of China and Japan***
  - ✓ ***most bronze medal country received-*** ***United States of America, People's Republic of China and Japan***

Suggestions

SUGGESTIONS

As my suggestions for all country to focus on plyers Training and provide a best *coaches because which country won most medal that Produces highest number of coaches and Country should be highest participants in it*. Some of the solutions proposed recently are improvement of the sports infrastructure is the only solution, right attitude of the policy makers and implementers will do wonders and the upliftment of women in the society is the solution. All these solutions are valid and will definitely make an impact on the sports performance however these solutions will not work in isolation. There is a need to first identify the problems in the present system and understand how they are interconnected to each other to be finally able to propose non-isolated solutions and an implemention plan.

Conclusion

CONCLUSION

The study consisted with all the work flows of major Points, for all country to focus on plyers Training and provide a best *coaches because which country won most medal that Produces highest number of coaches and Country should be highest participants in it. and follow to highest won medal country strategies.*

Bibliography

BIBLIOGRAPHY

WEBSITES:

https://www.kaggle.com/datasets/arjunprasadsarkhel/2021-olympics-in-tokyo

Tool & Programming Language Use

- Python programming
- ✓ NumPy
- ✓ Pandas
- ✓ Matplotlib.pyplot
- Data cleaning – It is the process of identifying and removing the anomalies in the dataset (Pandas)
- Data Transformation – It involves changing the data type of the columns, creating derived columns or removing duplicate data to name a few.
- Exploratory data analysis – Perform univariate and multi-variate analysis on the datasets to find hidden insights and patterns in them.