

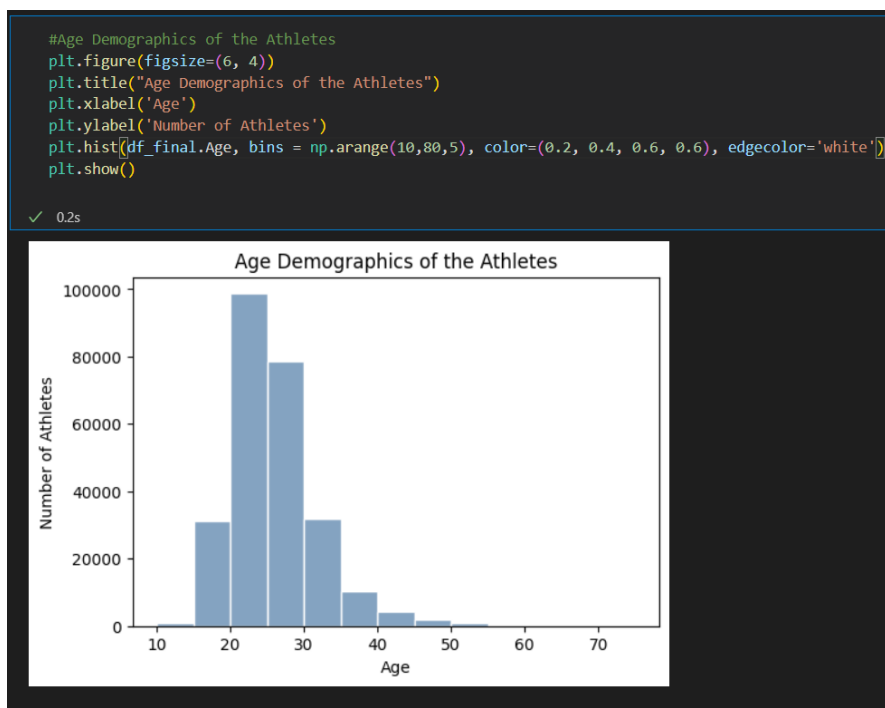
1. Provide a summary of the different descriptive statistics you looked at and WHY.

I analyzed age, weight, and height features because these are the critical factors that help in analyzing the performance of the athletes. Which nations have managed to have their athletes in the ideal Age, Weight, and Height range.

```
df_final[['Age', 'Height', 'Weight']].describe()
```

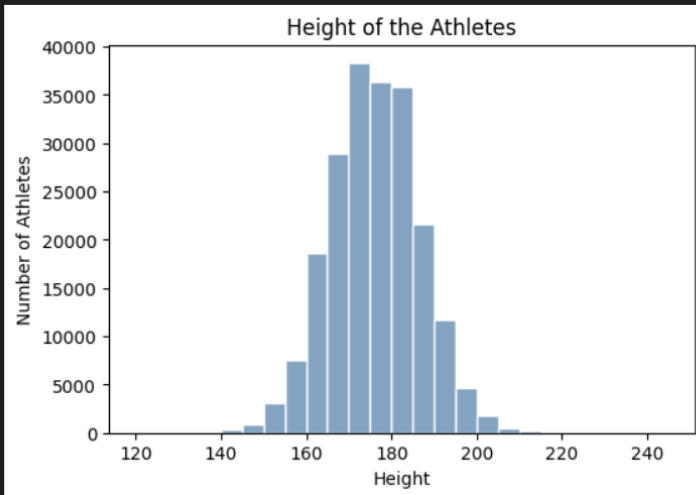
✓ 0.0s

	Age	Height	Weight
count	260079.000000	210656.000000	207945.000000
mean	25.457542	175.344234	70.708909
std	6.162508	10.519601	14.351103
min	10.000000	127.000000	25.000000
25%	21.000000	168.000000	60.000000
50%	24.000000	175.000000	70.000000
75%	28.000000	183.000000	79.000000
max	97.000000	226.000000	214.000000



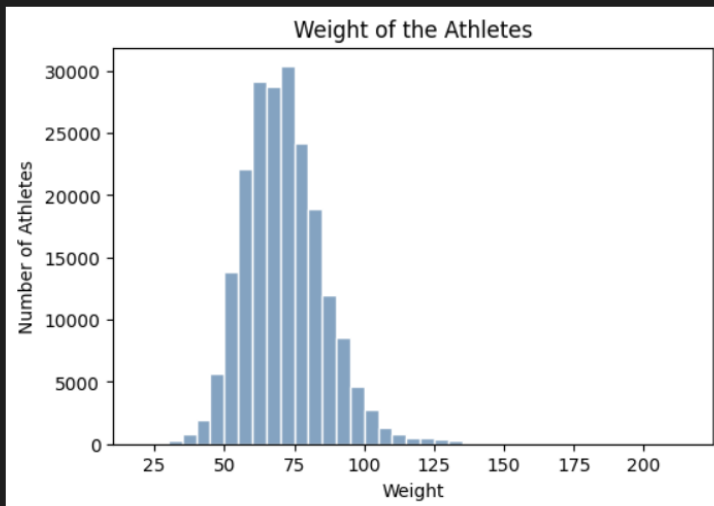
```
#Height Demographics of the Athletes
plt.figure(figsize=(6, 4))
plt.title("Height of the Athletes")
plt.xlabel('Height')
plt.ylabel('Number of Athletes')
plt.hist(df_final.Height, bins = np.arange(120,250,5), color=(0.2, 0.4, 0.6, 0.6), edgecolor='white')
plt.show()
```

✓ 0.2s



```
#Weight Demographics of the Athletes
plt.figure(figsize=(6, 4))
plt.title("Weight of the Athletes")
plt.xlabel('Weight')
plt.ylabel('Number of Athletes')
plt.hist(df_final.Weight, bins = np.arange(20,220,5), color=(0.2, 0.4, 0.6, 0.6), edgecolor='white')
plt.show()
```

✓ 0.2s



2. Submit 2-3 key points you may have discovered about the data, e.g. new relationships? Aha's! Did you come up with additional ideas for other things to review?

- Different sports in different seasons

Summer Sports :

```
df_final[df_final['Season'] == 'Summer']['Sport'].value_counts()
```

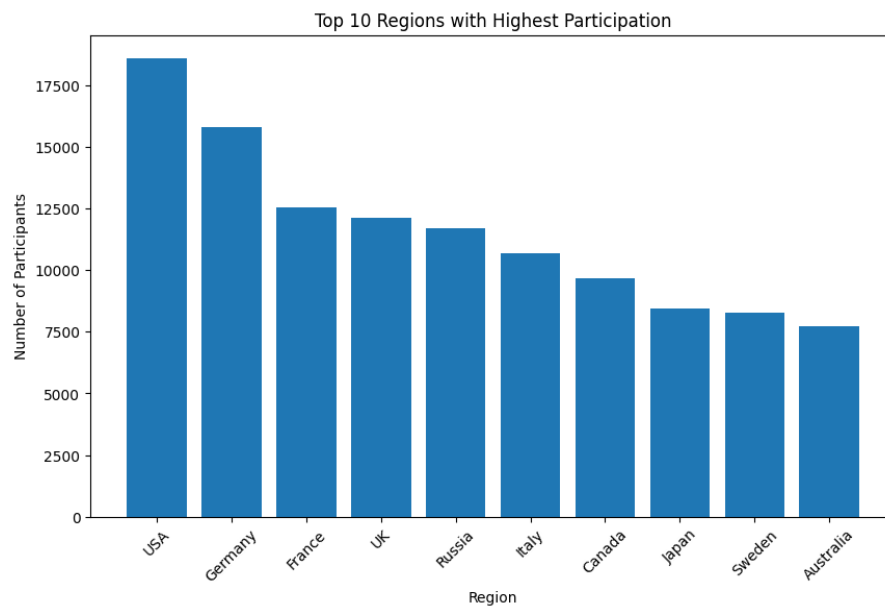
Sport	
Athletics	38596
Gymnastics	26703
Swimming	23052
Shooting	11432
Cycling	10827
Fencing	10731
Rowing	10594
Wrestling	7154
Football	6745
Sailing	6510
Equestrianism	6343
Canoeing	6169
Boxing	6046
Hockey	5400
Basketball	4526
Weightlifting	3926
Water Polo	3836
Judo	3799
Handball	3665
Volleyball	3404
Tennis	2862
Diving	2842

Winter Sports:

```
df_final[df_final['Season'] == 'Winter']['Sport'].value_counts()
```

Sport	
Cross Country Skiing	9133
Alpine Skiing	8829
Speed Skating	5613
Ice Hockey	5456
Biathlon	4893
Bobsleigh	3058
Ski Jumping	2401
Figure Skating	2244
Short Track Speed Skating	1534
Luge	1479
Nordic Combined	1344
Freestyle Skiing	937
Snowboarding	936
Curling	463
Skeleton	199
Military Ski Patrol	24
Alpinism	21

- Participation of different countries (highest participating nations)



- Number of participants in different sports

Athletics, Gymnastics, Swimming – very popular

Roque, Basque Pelota, Aeronautics – least popular

```
# Number of athletes in different sports
df_sports_participants = df_final.groupby(['Sport'])['ID'].count().sort_values(ascending=False)
```

✓ 0.0s

+ Code + Markdown

```
df_sports_participants.head()
```

✓ 0.0s

Sport	ID
Athletics	38596
Gymnastics	26703
Swimming	23052
Shooting	11432
Cycling	10827

Name: ID, dtype: int64

```
df_sports_participants.tail()
```

✓ 0.0s

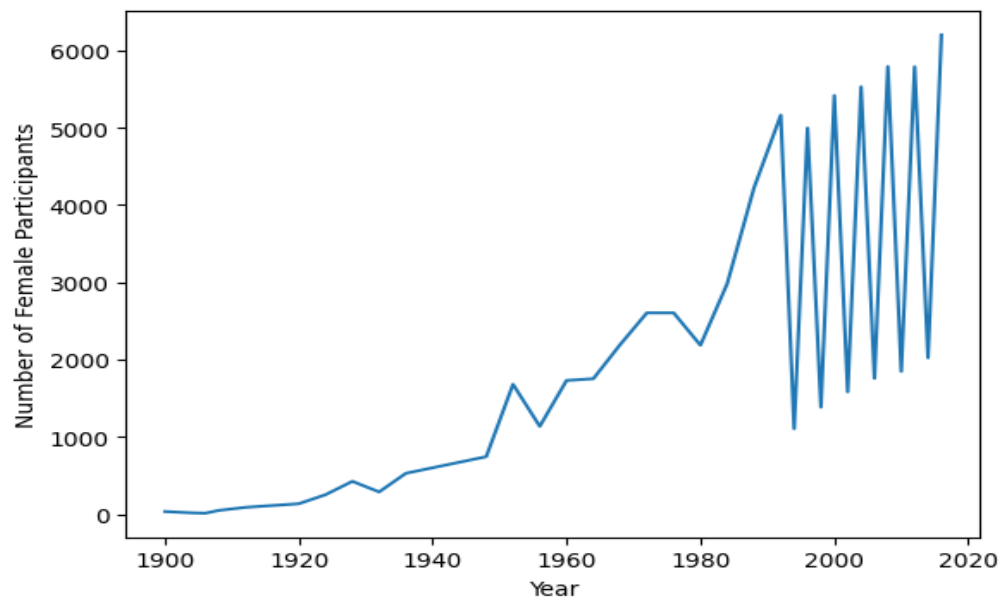
Sport	ID
Racquets	12
Jeu De Paume	11
Roque	4
Basque Pelota	2
Aeronautics	1

Name: ID, dtype: int64

3. Did you prove or disprove any of your initial hypotheses? If so, which one and what do you plan to do next?

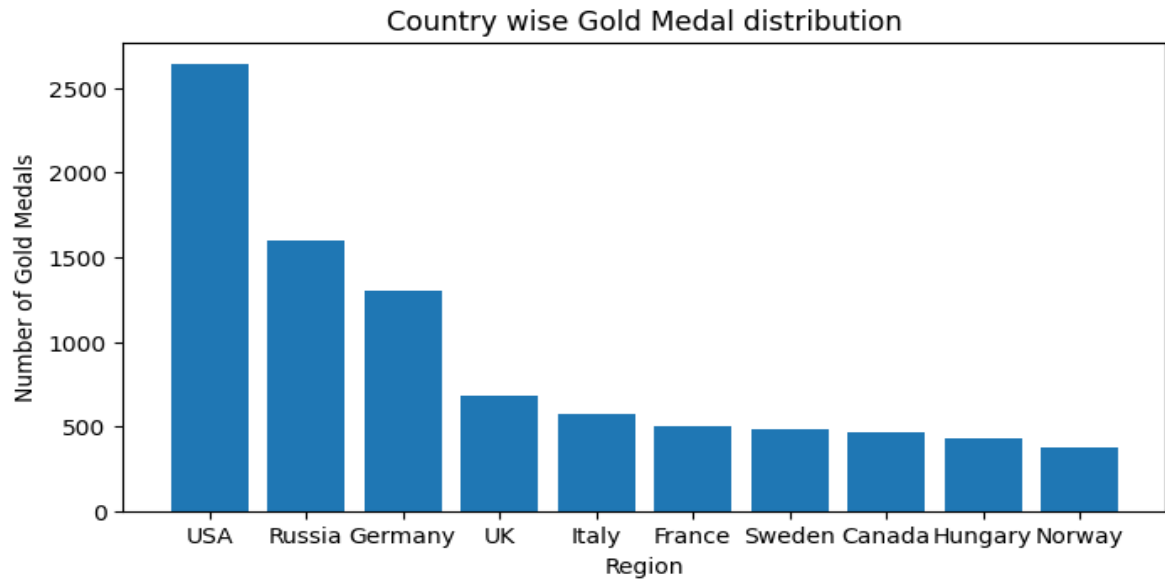
- Hypothesis: Participation of Females should have increased over the years

Outcome: My Hypothesis was correct, indeed the participation of females has increased over the years (1900 - 2020)



- Hypothesis : USA has the highest gold medal count

Outcome : My hypothesis was correct, USA has the highest gold medal count



- Next Steps

Female Participation: There is an abrupt pattern in female participation since 2000, In the winter season the number of female participants is lesser in comparison to the summer season of the respective year. I want to find out the reason behind this

Gold Medal distribution among nations : What is the reason why USA has the highest gold medal collection, I want to work on this.

4. What additional questions are you seeking to answer?

- Country-wise, sports performance (which country is performing best in what category)
- Role of Age,Weight, Height in Olympics Performance