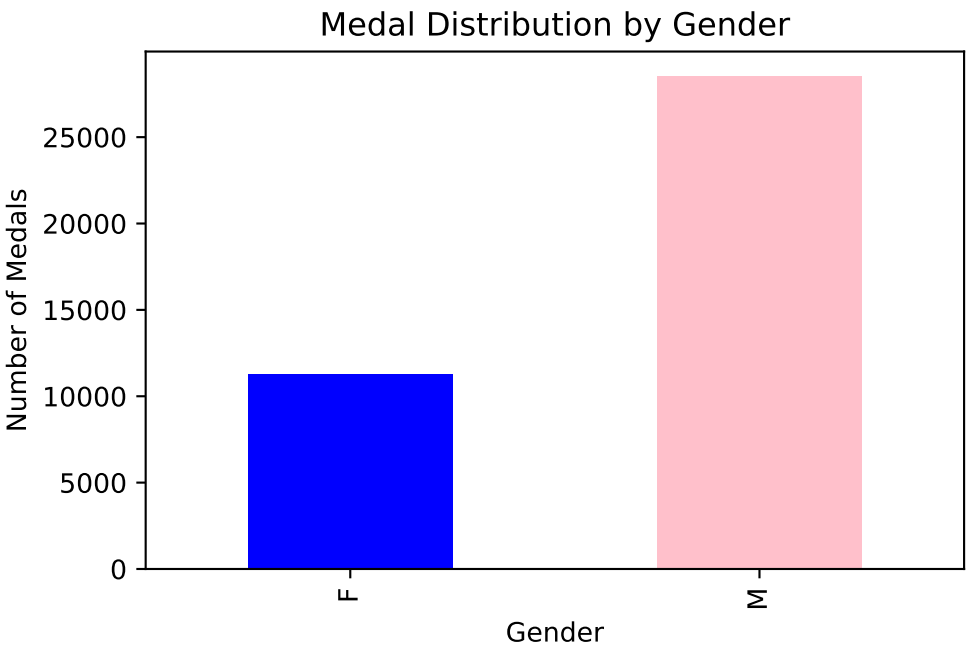
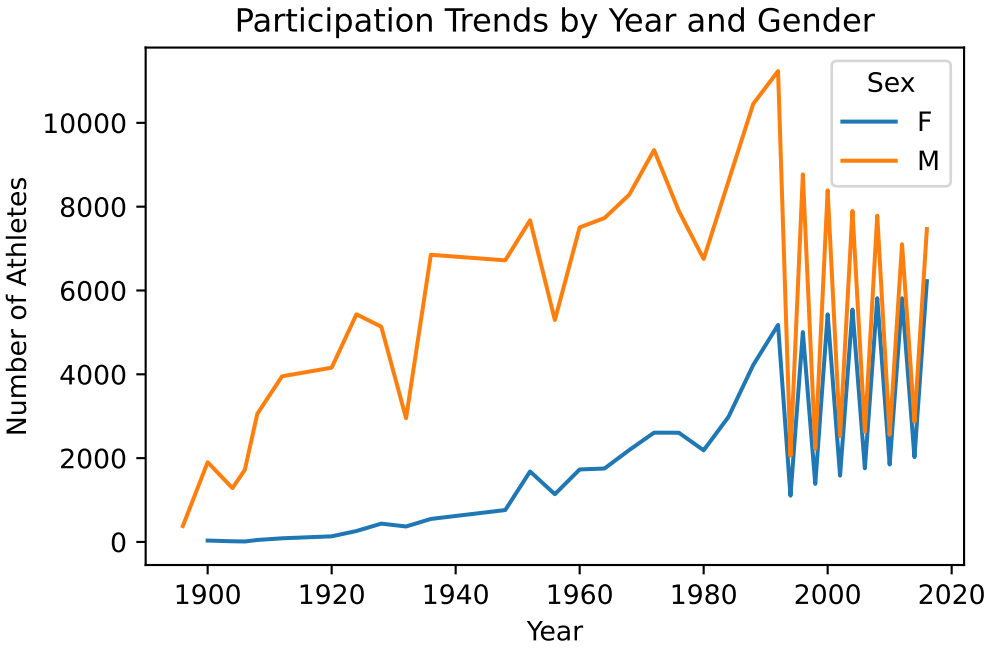
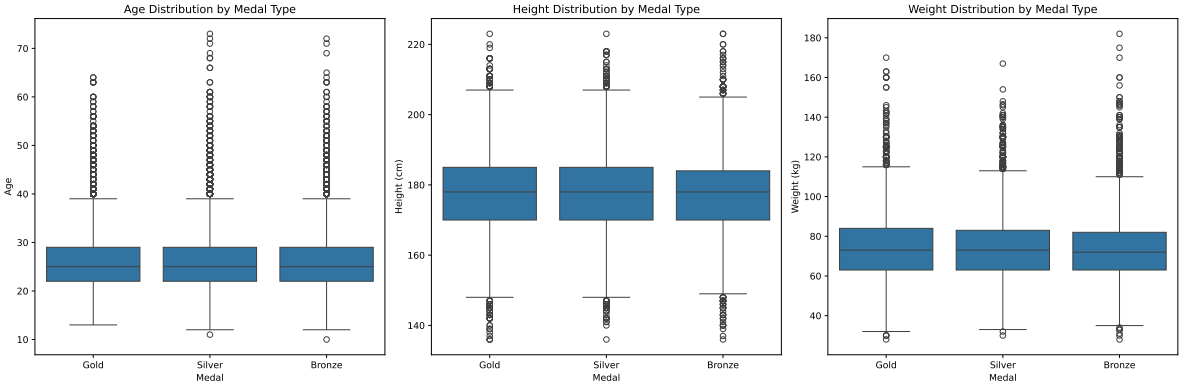
Some insights about Olympic Games from 1896 to 2020

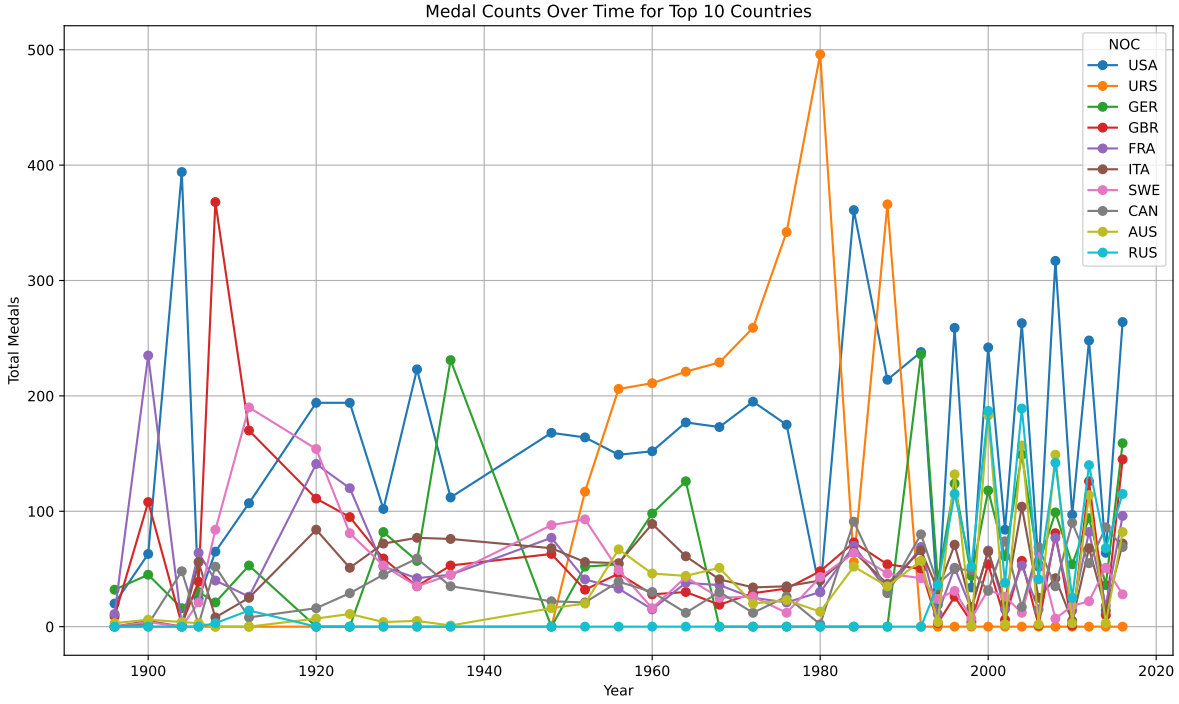
```
import pandas as pd
import matplotlib.pyplot as plt

# Load the dataset
file_path = '/home/edson/dataset/athlete_events.csv'
data = pd.read_csv(file_path)

# Filter data to include only rows where a medal was won
medal_data = data.dropna(subset=['Medal'])

# Count the number of medals won by gender
medal_by_gender = medal_data.groupby('Sex')['Medal'].count()

# Plotting the distribution
medal_by_gender.plot(kind='bar', color=['blue', 'pink'], title='Medal Distribution by Gender
plt.xlabel('Gender')
plt.ylabel('Number of Medals')
plt.show()

# Count the number of athletes participating by year and gender
participation_by_gender = data.groupby(['Year', 'Sex']).size().unstack()

# Plotting the participation trends
participation_by_gender.plot(kind='line', title='Participation Trends by Year and Gender')
plt.xlabel('Year')
plt.ylabel('Number of Athletes')
plt.show()

import seaborn as sns
```

```
# Plotting the distribution of Age, Height, and Weight by Medal type
plt.figure(figsize=(18, 6))

# Age Distribution
plt.subplot(1, 3, 1)
sns.boxplot(x="Medal", y="Age", data=medal_data, order=["Gold", "Silver", "Bronze"])
plt.title('Age Distribution by Medal Type')
plt.ylabel('Age')

# Height Distribution
plt.subplot(1, 3, 2)
sns.boxplot(x="Medal", y="Height", data=medal_data, order=["Gold", "Silver", "Bronze"])
plt.title('Height Distribution by Medal Type')
plt.ylabel('Height (cm)')

# Weight Distribution
plt.subplot(1, 3, 3)
sns.boxplot(x="Medal", y="Weight", data=medal_data, order=["Gold", "Silver", "Bronze"])
plt.title('Weight Distribution by Medal Type')
plt.ylabel('Weight (kg)')

plt.tight_layout()
plt.show()

# Aggregating total medal counts by country and year
total_medals_by_country_year = medal_data.groupby(['Year', 'NOC']).size().unstack(fill_value=

# Selecting top 10 countries by total medals won across all years
top_countries = total_medals_by_country_year.sum().sort_values(ascending=False).head(10).inde
medals_top_countries = total_medals_by_country_year[top_countries]

# Plotting the trends of medal counts for these top 10 countries over time
medals_top_countries.plot(kind='line', marker='o', figsize=(14, 8), title='Medal Counts Over
plt.ylabel('Total Medals')
plt.xlabel('Year')
plt.grid(True)
plt.show()
```


Medal Distribution by Gender
Number of Medals
25000
20000
15000
10000
5000
0
F
M
Gender


Participation Trends by Year and Gender
Number of Athletes
10000
8000
6000
4000
2000
0
1900
1920
1940
1960
1980
2000
2020
Year
Sex
F
M


Age Distribution by Medal Type
Age
Gold
Silver
Bronze
Medal
Height Distribution by Medal Type
Height (cm)
Weight Distribution by Medal Type
Weight (kg)


Medal Counts Over Time for Top 10 Countries
Total Medals
Year
NOC
USA
URS
GER
GBR
FRA
ITA
SWE
CAN
AUS
RUS