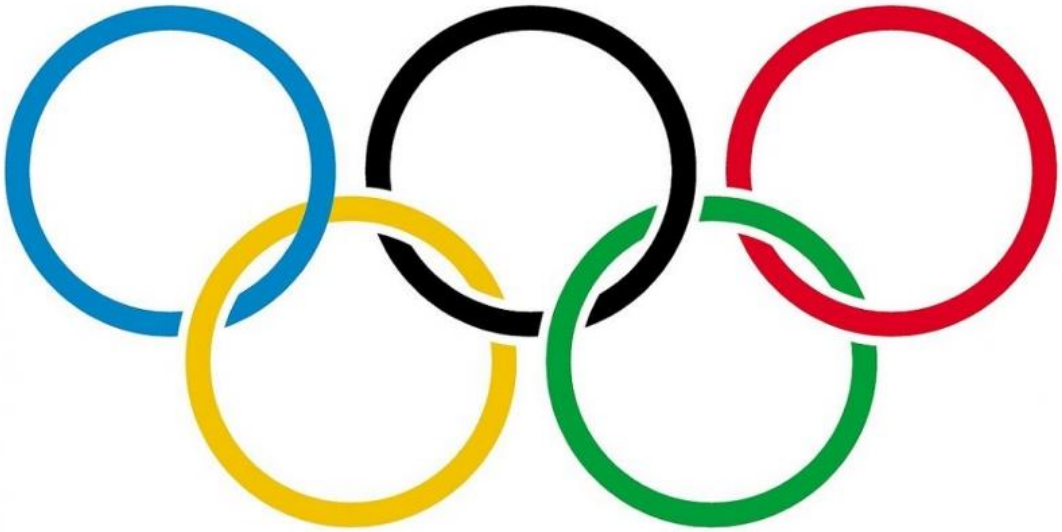
This Notebook Done by Sawsan Aljawni and Fatimah Al Yami

Summer Olympic Medalists 1896-2008



1. Dataset Introduction:

This dataset is about Summer Olympic Medalists from 1896 to 2008, it shows the host city, sport type, athlete name and gender, and the medal that the athlete won.

2. Python and Data Visualizations:

- **Importing the libraries and the dataset**

```
# In[14]:


#import the libraries
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns

```

In [22]:
```
#read the csv file
df=pd.read_csv("Summer_Olympic_medallists_1896-2008.csv")

df.head()
#df.shape
```

Out[22]:

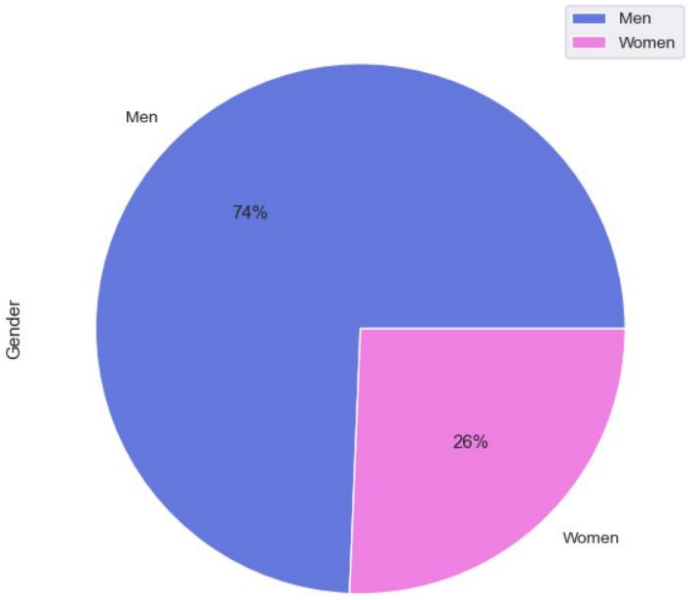
	City	Edition	Sport	Discipline	Athlete	NOC	Gender	Event	Event_gender	Medal
0	Athens	1896	Aquatics	Swimming	HAJOS, Alfred	HUN	Men	100m freestyle	M	Gold
1	Athens	1896	Aquatics	Swimming	HERSCHMANN, Otto	AUT	Men	100m freestyle	M	Silver
2	Athens	1896	Aquatics	Swimming	DRIVAS, Dimitrios	GRE	Men	100m freestyle for sailors	M	Bronze
3	Athens	1896	Aquatics	Swimming	MALOKINIS, Ioannis	GRE	Men	100m freestyle for sailors	M	Gold
4	Athens	1896	Aquatics	Swimming	CHASAPIS, Spiridon	GRE	Men	100m freestyle for sailors	M	Silver

- **Data Visualization**

1- Pie plot:

```
# In[108]:


plt.figure(figsize=(15,8), dpi=90)
df["Gender"].value_counts(normalize=True).plot(
                                        kind="pie",
                                        legend=True,
                                        autopct='%.f%%',
                                        colors = ['#6478DD', '#EF81E2']
                                        )
plt.title( "Total Number of Men and Female in Summer Olympic 1896-2008",c="black")

```

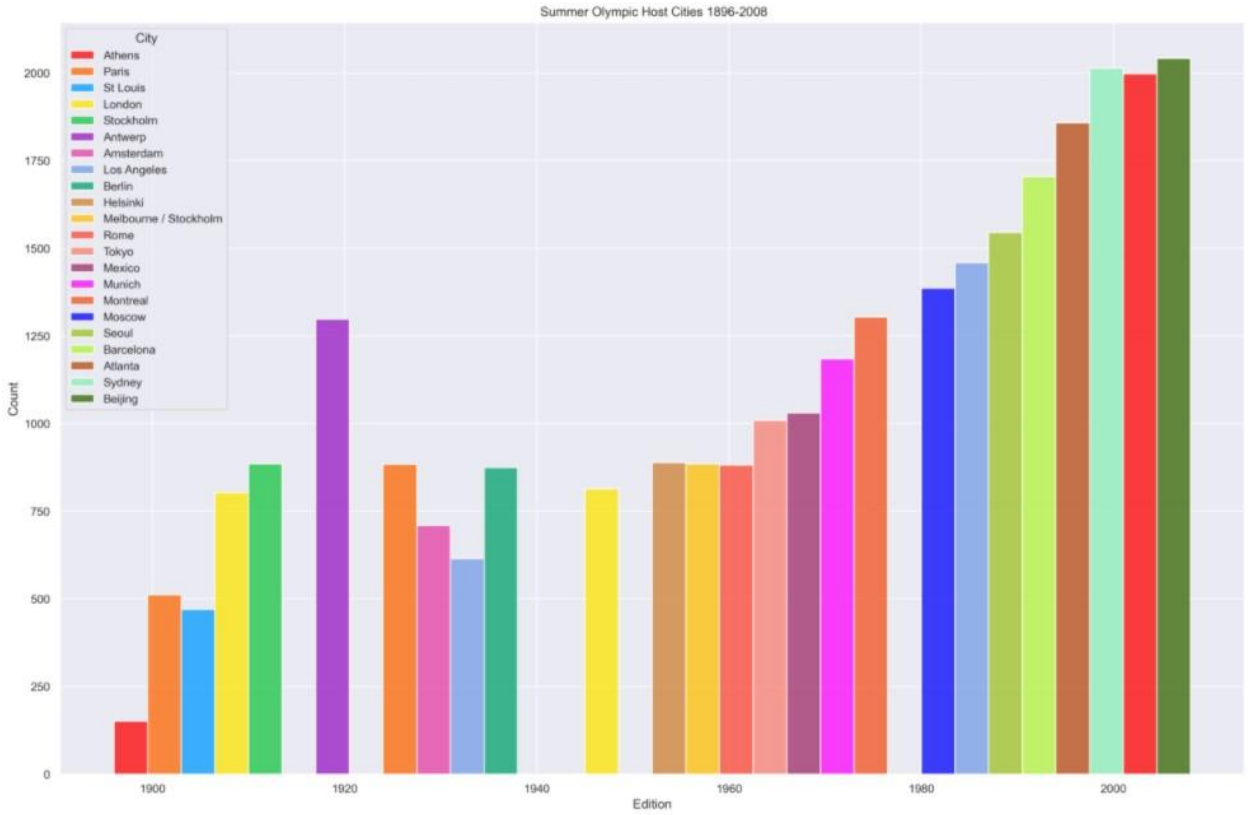
This pie plot illustrates the total percentage of men and women who is participated in summer Olympic from 1896 to 2008. The rate of men who is participated is 74 percent which is almost triple women rate. Whereas proportion of females was 26 percent.



2- Histogram plot

```
# In[115]:


plt.figure(figsize=(20,13), dpi=300)

city=sns.histplot(data=df,
             x="Edition",
              hue="City",
              binwidth=3.5,
             palette=sns.color_palette(['#FF0000', '#FF6500','#009AFF','#FFE500',
                                        '#16C540','#9816C5','#E53CA4','#749BE7',
                                        '#00A36C', '#CD7F32','#FFC000','#FF4433',
                                         '#FA8072','#9F2B68','#FF00FF','#F4511E',
                                         '#0000FF','#9DC223','#B4F43A','#B84811',
                                         '#87EFB8','#336600']),
              multiple="stack",
                 ).set_title(' Cities Total of Participating in the Summer Olympic 1896-2008')
```

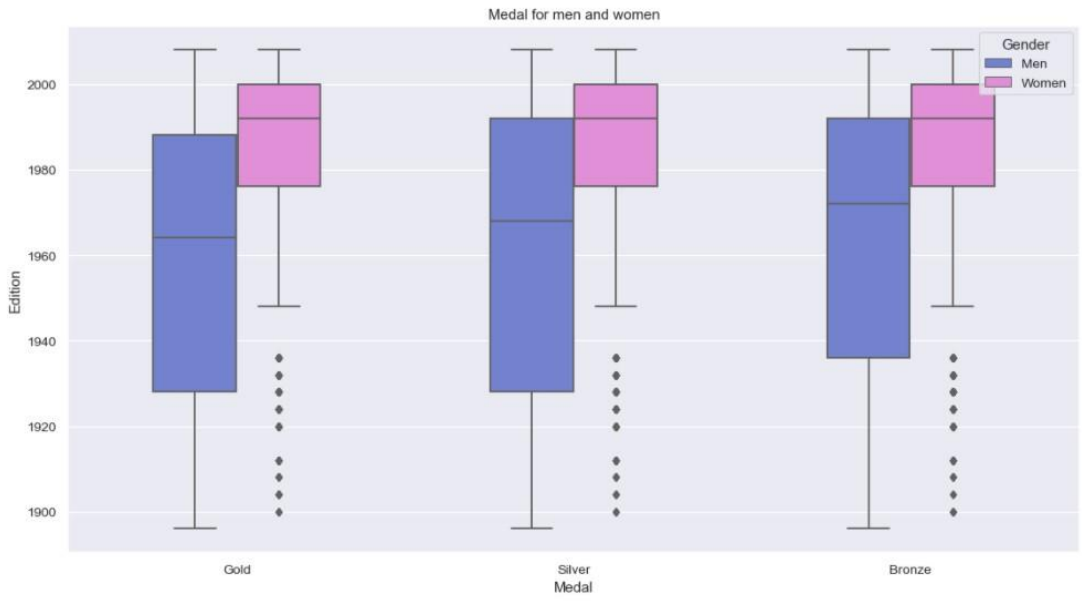
This chart shows the host cities in summer Olympic from 1896 to 2008. Athens has the lowest number of participants from other cities. Beijing has the highest number of participants from other cities. As for Rome, Melbourne and Helsinki has almost the same count.



3- Boxplot

```
# In[110]:


plt.figure(figsize=(15,8), dpi=90)
sns.set(style="darkgrid")

my_color={"Men":"#6478DD","Women":"#EF81E2"}

sns.boxplot(x="Medal",
            y="Edition",
            hue="Gender",
            data=df ,
            width=0.5,
           palette=my_color
           )

plt.title('Medal for men and women')
plt.show()
```

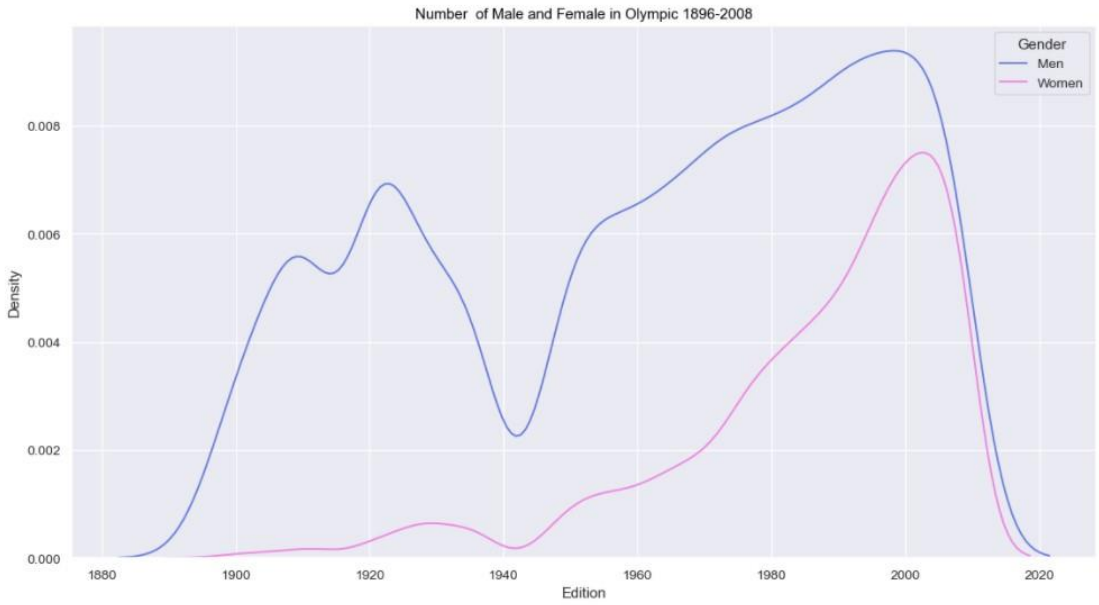
This plot shows the kind of medal for men and women from 1896 to 2008. From 1896 to 1900 the medal was for men but from 1900 to 2008 the medal women have earned it. Also, the bronze medal has less number than gold and silver.



4- line plot

```
# In[111]:


plt.figure(figsize=(15,8), dpi=90)
sns.kdeplot(
    data=df,
    x="Edition",
    hue="Gender",
    palette=my_color
)
plt.title( "Number  of Male and Female in Olympic 1896-2008",c="black")
plt.show()
```

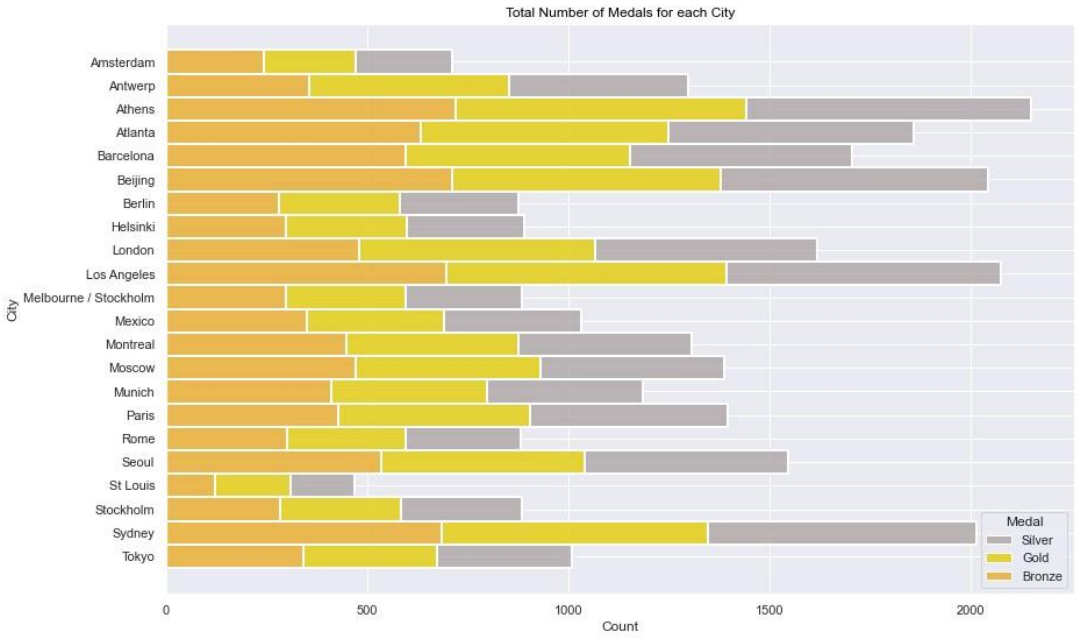
This line lot explains the comparison between the number of male and female in Olympic from 1896 to 2008. As we see in the chart, the men have the more participating than women.



5- Histplot

```
# In[112]:


plt.figure(figsize=(14,9))
med_color={"Gold":"#E3CC07","Silver":"#AEA6A6","Bronze":"#EBAC29"}
sns.histplot(
    df.sort_values('City').reset_index(),
    y="City",
    hue="Medal",
    multiple="stack",
    alpha=0.8,
    palette=med_color
)
plt.title("Total Number of Medals for each City",c='black');
plt.show()
```

This plot illustrates the number of medals for each city. Los Angeles, Sydney and Athens have the highest number of silver Medal, but St Louis has the lowest number of silver medal. Also, Los Angeles, Beijing and Athens have the highest number of gold Medal, but St Louis has the lowest number of gold medal. Athens and Beijing have the highest number of bronze Medal, but St Louis has the lowest number of bronze Medal.



7. Subplot

```
# In[113]:


f, sub = plt.subplots(2,2, figsize=(17,20),dpi=200)

# fig.tight_layout()
f.suptitle("Summer Olympic 1896-2008", fontsize="larger")

#1. pie
df["Gender"].value_counts(normalize=True).plot(
                                        kind="pie",
                                        legend=True,
                                        autopct='%.f%%',
                                        colors = ['#6478DD', '#EF81E2'],
                                        title="Total Number of Men and Female in Summer Olympic 1896-2008",
                                        ax=sub[0,0])

#-----------------------------------------------------------------------------------------------------
#2. Histogram
sns.kdeplot(
    data=df,
    x="Edition",
    hue="Gender",
    palette=my_color,
    ax=sub[1,1]
).set_title("Number of Male and Female in Olympic 1896-2008",c="black")


#-----------------------------------------------------------------------------------------------------
#3. BoxPlot
my_color={"Men":"#6478DD","Women":"#EF81E2"}

sns.boxplot(x="Medal",
            y="Edition",
            hue="Gender",
            data=df ,
            width=0.5,
           palette=my_color,
            ax=sub[0,1]
           ).set_title("Medal for men and women",c="black")

#-----------------------------------------------------------------------------------------------------
#4 Histogram
med_color={"Gold":"#E3CC07","Silver":"#AEA6A6","Bronze":"#EBAC29"}
sns.histplot(
    df.sort_values('City').reset_index(),
    y="City",
    hue="Medal",
    multiple="stack",
    alpha=0.8,
    palette=med_color,
    ax=sub[1,0]
).set_title("Total Number of Medals for each City",c="black")


```

Summer Olympic 1896-2008