

PROBLEMAS PROPUESTOS

1. De una lista de deportes indicados en el archivo athlete_events.csv, liste todos los deportes que se han realizado en la historia de los juegos olímpicos en orden alfabético

Una solución posible:

```
import pandas as pd
import numpy

df = pd.read_csv('athlete_events.csv')
deportes = numpy.sort(df['Sport'].unique())
print(deportes)
```

2. Implemente una función que solicite al usuario un deporte, y busque en la base de datos todos los años en que se ha realizado dicho deporte.

Una solución posible:

```
import pandas as pd
import numpy

#funciones
def deporte_año(deporteA, df):

    df_sport= df[df['Sport']==deporteA]
    df_year = numpy.sort(df_sport['Year'].unique()). #ordena Los años

    print(deporteA, 'se ha jugado en los años:')
    for year in df_year:
        print(year)

#programa ppaL
df = pd.read_csv('athlete_events.csv')
deporteA= input('Ingrese un deporte: ')

#llamado a la función
deporte_año(deporteA, df)
```

3. Implemente una función que solicite al usuario un deporte, y busque en la base de datos el año con menor y mayor cantidad de atletas en dicho deporte.

Una solución posible:

```
import pandas as pd
import numpy

def deporte_anho(deporteA, df):

    df_sport= df[df['Sport']==deporteA]
    df_year = numpy.sort(df_sport['Year'].unique())

    players=[]
    for year in df_year:
        df_sel_A= df[(df['Sport']==deporteA) & (df['Year']==year)]
        stats_A= df_sel_A['Sport'].agg(['count'])
        players.append(stats_A['count'])
        print('Year:', year, ' # Jugadores:',stats_A['count'])

    players = numpy.array(players)
    return(df_year, players)

#programa ppaL
df = pd.read_csv('athlete_events.csv')
deporteA= input('Ingrese un deporte: ')

years, players= deporte_anho(deporteA, df)

index_min = numpy.argmin(players)
index_max = numpy.argmax(players)

print('Año con menos jugadores:',years[index_min], 'con:',players.min() )
print('Año con más jugadores:',years[index_max], 'con:',players.max() )
```

4. Implemente una función que solicite al usuario un deporte, y grafique la evolución de altura del deporte seleccionado por año.

Una solución posible:

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

def grafico_deporte(deporteA, df):

    df_sport= df[df['Sport']==deporteA]
    df_year = numpy.sort(df_sport['Year'].unique())

    players=[]
    anos = []
    for year in df_year:
        df_sel_A= df[(df['Sport']==deporteA) & (df['Year']==year)]
        stats_A= df_sel_A['Height'].agg(['mean', 'count'])

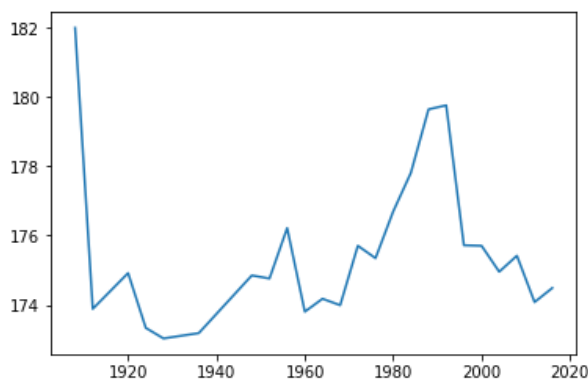
        #verificamos si hay datos en dicho año
        if stats_A['count']>0:
            anos.append(year)
            players.append(stats_A['mean'])
            print('Year:', year, ' # Promedio altura:', stats_A['mean'])

    #gráfica con mpl
    plt.plot(anos, players)

#programa ppal
df = pd.read_csv('athlete_events.csv')
deporteA= input('Ingrese un deporte: ')

#Llamado a la función
grafico_deporte(deporteA, df)
```

Ejemplo de gráfico:



5. Implemente una función que solicite al usuario un deporte, y grafique el histograma de peso del deporte seleccionado por año.

Una solución posible:

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

def histograma_deporte(deporteA, df):

    df_sport= df[df['Sport']==deporteA]
    df_year = numpy.sort(df_sport['Year'].unique())

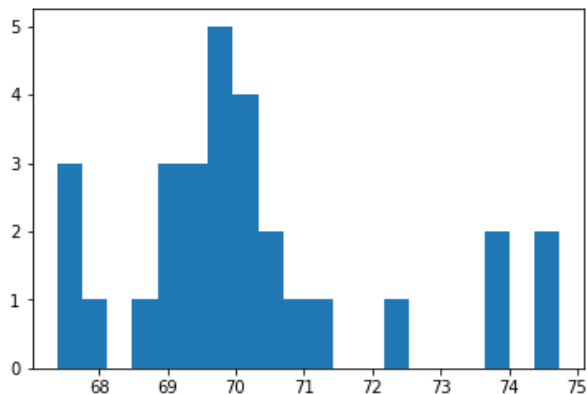
    players=[]
    años = []
    for year in df_year:
        df_sel_A= df[(df['Sport']==deporteA) & (df['Year']==year)]
        stats_A= df_sel_A['Weight'].agg(['mean', 'count'])
        if stats_A['count']>0:
            años.append(year)
            players.append(stats_A['mean'])
            print('Year:', year, ' # Promedio peso:', stats_A['mean'])

    plt.hist(players, bins=20)

#programa ppal
df = pd.read_csv('athlete_events.csv')
deporteA= input('Ingrese un deporte: ')

#llamado a la función
histograma_deporte(deporteA, df)
```

Ejemplo de gráfico:



6. Implemente una función que solicite dos deportes, y grafique la evolución de la altura de ambos deportes agrupados por año

Una solución posible:

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

def grafica_deporte(deporte, df, ax):
    datos= df[df['Sport']==deporte].groupby('Year')
    ax.plot(datos['Year'].unique(),datos['Age'].agg(['mean']), label=deporte)
    legend = ax.legend(loc='upper left', shadow=True, fontsize='medium')

#programa ppal
fig, ax = plt.subplots()

df = pd.read_csv('athlete_events.csv')
deporteA= input('Ingrese un deporte: ')
deporteB= input('Ingrese otro deporte: ')

#llamado a la función
grafica_deporte(deporteA, df, ax)
grafica_deporte(deporteB, df, ax)

plt.show()
```

Ejemplo de gráfico:

