Summer Olympics Data Analysis Assignment¶

In [49]:

import pandas as pd

import numpy as np

import matplotlib.pyplot as plt

In [50]:

df = pd.read\_csv("summer.csv")

In [53]:

df

Out[53]:

Year City Sport Discipline Athlete Country Gender Event Medal

0 1896 Athens Aquatics Swimming HAJOS, Alfred HUN Men 100M Freestyle Gold

1 1896 Athens Aquatics Swimming HERSCHMANN, Otto AUT Men 100M Freestyle Silver

2 1896 Athens Aquatics Swimming DRIVAS, Dimitrios GRE Men 100M Freestyle For Sailors Bronze

3 1896 Athens Aquatics Swimming MALOKINIS, Ioannis GRE Men 100M Freestyle For Sailors Gold

4 1896 Athens Aquatics Swimming CHASAPIS, Spiridon GRE Men 100M Freestyle For Sailors Silver

... ... ... ... ... ... ... ... ... ...

31160 2012 London Wrestling Wrestling Freestyle JANIKOWSKI, Damian POL Men Wg 84 KG Bronze

31161 2012 London Wrestling Wrestling Freestyle REZAEI, Ghasem Gholamreza IRI Men Wg 96 KG Gold

31162 2012 London Wrestling Wrestling Freestyle TOTROV, Rustam RUS Men Wg 96 KG Silver

31163 2012 London Wrestling Wrestling Freestyle ALEKSANYAN, Artur ARM Men Wg 96 KG Bronze

31164 2012 London Wrestling Wrestling Freestyle LIDBERG, Jimmy SWE Men Wg 96 KG Bronze

31165 rows × 9 columns

1. In how many cities Summer Olympics is held so far?

In [ ]:

len(df.City.unique())

Out[ ]:

22

In [ ]:

data = []

for City in df['City'].unique():

data.append([City , len(df[df['City'] == City])])

pd.DataFrame(data,columns = ['City','freq']).sort\_values(by='freq', ascending=False)

Out[ ]:

City freq

2 London 340

21 Beijing 307

20 Athens 298

19 Sydney 293

18 Atlanta 259

17 Barcelona 200

6 Los Angeles 183

16 Seoul 172

15 Moscow 143

14 Montreal 127

13 Munich 80

12 Mexico 71

11 Tokyo 63

10 Rome 49

9 Melbourne / Stockholm 46

8 Helsinki 46

5 Amsterdam 32

7 Berlin 28

0 Paris 20

4 Antwerp 15

3 Stockholm 10

1 St Louis 6

2. Which sport is having most number of Gold Medals so far? (Top 5)

In [ ]:

df= pd.read\_csv("summer.csv")

df=df[df["Medal"] =='Gold']

data = []

for Sport in df['Sport'].unique():

data.append([Sport, len(df[df['Sport'] == Sport])])

pd.DataFrame(data,columns = ['Sport',' Medal']).sort\_values(by=' Medal', ascending=False).head().plot(x = 'Sport', y = ' Medal', kind = 'bar', figsize = (5,5))

Out[ ]:

<matplotlib.axes.\_subplots.AxesSubplot at 0x7fef48089510>

3. Which sport is having most number of medals so far? (Top 5)

In [ ]:

df.groupby('Sport')['Medal'].count().nlargest(5).plot.bar(figsize = (5,5) , color = 'red')

Out[ ]:

<matplotlib.axes.\_subplots.AxesSubplot at 0x7efc9ae17d50>

4. Which player has won most number of medals? (Top 5)

In [ ]:

df.groupby('Athlete')['Medal'].count().nlargest(5).plot.bar(figsize = (5,5) , color = 'orange')

Out[ ]:

<matplotlib.axes.\_subplots.AxesSubplot at 0x7efc9ad04c90>

5. Which player has won most number Gold Medals of medals? (Top 5)

In [ ]:

df=df[df["Medal"] =='Gold']

df.groupby('Athlete')['Medal'].count().nlargest(5).plot.bar(figsize = (5,5) , color = 'green')

Out[ ]:

<matplotlib.axes.\_subplots.AxesSubplot at 0x7efc9a9f59d0>

6. In which Year India won first Gold Medal in Summer Olympics?

In [ ]:

df=df[df["Country"] =='IND']

df=df[df["Medal"] =='Gold']

data = []

for Year in df['Year'].unique():

data.append([Year , len(df[df['Year'] == Year])])

pd.DataFrame(data,columns = ['Year','Medal']).sort\_values(by='Year', ascending = True).head(1)

Out[ ]:

Year Medal

0 1928 15

7. Which event is most popular in terms on number of players? (Top 5)

In [52]:

df.groupby('Event')['Athlete'].count().head().sort\_values(ascending = False).plot.bar(figsize = (5,5), color = 'red')

Out[52]:

<matplotlib.axes.\_subplots.AxesSubplot at 0x7efc9a560110>

8. Which sport is having most female Gold Medalists? (Top 5)

In [54]:

df=df[df["Gender"] =='Women']

df=df[df["Medal"]== 'Gold']

data = []

for Sport in df['Sport'].unique():

data.append([Sport , len(df[df['Sport'] == Sport])])

pd.DataFrame(data,columns = ['Sport','Medal']).sort\_values(by='Medal', ascending=False).head().plot(x = 'Sport', y = 'Medal', kind = 'bar', figsize = (5,5) , color = 'green')

Out[54]:

<matplotlib.axes.\_subplots.AxesSubplot at 0x7efca293f350>

In [55]:

Women=df[df["Gender"] =='Women']

df=Women[Women["Medal"]== 'Gold']

data = []

for Sport in df['Sport'].unique():

data.append([Sport , len(df[df['Sport'] == Sport])])

pd.DataFrame(data,columns = ['Sport','Medal']).sort\_values(by='Medal', ascending=False).head()

Out[55]:

Sport Medal

4 Aquatics 589

7 Athletics 389

8 Gymnastics 268

14 Rowing 217

11 Volleyball 166

In [ ]: