---

1. Write a program to initialize a dictionary (of your choice) and create a pandas DataFrame from it. Print the dataframe?

```
#Code here
import pandas as pd
data = {'Name': ['Harish', 'Mehul', 'Ayush'], 'Age': [20, 19, 22],
'City': ['Ambala', 'Delhi', 'Gorakhpur']}
df = pd.DataFrame(data)
print(df)
```

```
     Name  Age       City
0  Harish   20     Ambala
1   Mehul   19      Delhi
2   Ayush   22  Gorakhpur
```

---

2. Write a program to read the CSV file "summer.csv" into a pandas DataFrame "df", display the first 5 rows, and print the DataFrame's shape

```
#Code here
file_path = r"C:\Users\haris\Downloads\summer.csv"
df = pd.read_csv(file_path)
print(df.head())
print("DataFrame Shape:", df.shape)
```

```
   Year    City     Sport Discipline             Athlete Country
Gender  \
0  1896  Athens  Aquatics   Swimming       HAJOS, Alfred     HUN
Men
1  1896  Athens  Aquatics   Swimming    HERSCHMANN, Otto     AUT
Men
2  1896  Athens  Aquatics   Swimming   DRIVAS, Dimitrios     GRE
Men
3  1896  Athens  Aquatics   Swimming  MALOKINIS, Ioannis     GRE
Men
4  1896  Athens  Aquatics   Swimming  CHASAPIS, Spiridon     GRE
Men

                         Event   Medal
0              100M Freestyle    Gold
1              100M Freestyle  Silver
2  100M Freestyle For Sailors  Bronze
3  100M Freestyle For Sailors    Gold
4  100M Freestyle For Sailors  Silver
DataFrame Shape: (31165, 9)
```

---

3. Write a program to display the last 10 rows of the DataFrame?

```
#Code here
print(df.tail(10))
```

```
       Year    City      Sport           Discipline  \
31155  2012  London  Wrestling  Wrestling Freestyle
31156  2012  London  Wrestling  Wrestling Freestyle
31157  2012  London  Wrestling  Wrestling Freestyle
31158  2012  London  Wrestling  Wrestling Freestyle
31159  2012  London  Wrestling  Wrestling Freestyle
31160  2012  London  Wrestling  Wrestling Freestyle
31161  2012  London  Wrestling  Wrestling Freestyle
31162  2012  London  Wrestling  Wrestling Freestyle
31163  2012  London  Wrestling  Wrestling Freestyle
31164  2012  London  Wrestling  Wrestling Freestyle

                           Athlete Country Gender     Event   Medal
31155                AHMADOV, Emin     AZE    Men  Wg 74 KG  Bronze
31156          KAZAKEVIC, Aleksandr     LTU    Men  Wg 74 KG  Bronze
31157                KHUGAEV, Alan     RUS    Men  Wg 84 KG    Gold
31158  EBRAHIM, Karam Mohamed Gaber     EGY    Men  Wg 84 KG  Silver
31159              GAJIYEV, Danyal     KAZ    Men  Wg 84 KG  Bronze
31160           JANIKOWSKI, Damian     POL    Men  Wg 84 KG  Bronze
31161     REZAEI, Ghasem Gholamreza     IRI    Men  Wg 96 KG    Gold
31162                TOTROV, Rustam     RUS    Men  Wg 96 KG  Silver
31163            ALEKSANYAN, Artur     ARM    Men  Wg 96 KG  Bronze
31164               LIDBERG, Jimmy     SWE    Men  Wg 96 KG  Bronze
```

---

4. Write a program to get a summary of the DataFrame (such as the mean, count, etc.) and print the data types of all columns?

```
#Code here
print("Summary:")
print(df.describe())
print("\nData Types:")
print(df.dtypes)
```

```
Summary:
              Year
count  31165.000000
mean    1970.482785
std       33.158416
min     1896.000000
25%     1948.000000
50%     1980.000000
75%     2000.000000
max     2012.000000
```

```
Data Types:
Year           int64
City          object
Sport         object
Discipline    object
Athlete       object
Country       object
Gender        object
Event         object
Medal         object
dtype: object
```

---

5. Write a program to select a single column "Gender" from the DataFrame?

```
#Code here
gender_column = df['Gender']
print(gender_column)
```

```
0        Men
1        Men
2        Men
3        Men
4        Men
        ...
31160    Men
31161    Men
31162    Men
31163    Men
31164    Men
Name: Gender, Length: 31165, dtype: object
```

---

6. Write a program to select multiple columns "Year" and "City" from the DataFrame?

```
#Code here
selected_columns = df[['Year', 'City']]
print(selected_columns)
```

```
       Year    City
0      1896  Athens
1      1896  Athens
2      1896  Athens
3      1896  Athens
4      1896  Athens
...     ...     ...
```

```
31160  2012  London
31161  2012  London
31162  2012  London
31163  2012  London
31164  2012  London

[31165 rows x 2 columns]
```

---

7. Write a program to filter rows in a DataFrame based on the condition that Gender = Men ?

```
#Code here
filtered_df = df[df['Gender'] == 'Men']
print(filtered_df)
```

```
       Year    City      Sport           Discipline  \
0      1896  Athens   Aquatics             Swimming
1      1896  Athens   Aquatics             Swimming
2      1896  Athens   Aquatics             Swimming
3      1896  Athens   Aquatics             Swimming
4      1896  Athens   Aquatics             Swimming
...     ...     ...        ...                  ...
31160  2012  London  Wrestling  Wrestling Freestyle
31161  2012  London  Wrestling  Wrestling Freestyle
31162  2012  London  Wrestling  Wrestling Freestyle
31163  2012  London  Wrestling  Wrestling Freestyle
31164  2012  London  Wrestling  Wrestling Freestyle

                         Athlete Country Gender
Event  \
0                  HAJOS, Alfred     HUN    Men               100M
Freestyle
1                HERSCHMANN, Otto     AUT    Men               100M
Freestyle
2               DRIVAS, Dimitrios     GRE    Men  100M Freestyle For
Sailors
3              MALOKINIS, Ioannis     GRE    Men  100M Freestyle For
Sailors
4              CHASAPIS, Spiridon     GRE    Men  100M Freestyle For
Sailors
...                           ...     ...    ...
...
31160          JANIKOWSKI, Damian     POL    Men                    Wg
84 KG
31161  REZAEI, Ghasem Gholamreza     IRI    Men                    Wg
96 KG
31162             TOTROV, Rustam     RUS    Men                    Wg
96 KG
```

```
31163          ALEKSANYAN, Artur     ARM    Men                        Wg
96 KG
31164             LIDBERG, Jimmy     SWE    Men                        Wg
96 KG

        Medal
0        Gold
1      Silver
2      Bronze
3        Gold
4      Silver
...       ...
31160  Bronze
31161    Gold
31162  Silver
31163  Bronze
31164  Bronze

[22746 rows x 9 columns]
```

---

8. Write a program to select and print the first 5 rows and columns 2 to 6 (inclusive) of the DataFrame. Also, print the type of the resulting object?

```
#Code here
selected_data = df.iloc[:5, 1:6]
print(selected_data)
print("\nType of Resulting Object:", type(selected_data))

     City     Sport Discipline             Athlete Country
0  Athens  Aquatics   Swimming       HAJOS, Alfred     HUN
1  Athens  Aquatics   Swimming    HERSCHMANN, Otto     AUT
2  Athens  Aquatics   Swimming   DRIVAS, Dimitrios     GRE
3  Athens  Aquatics   Swimming  MALOKINIS, Ioannis     GRE
4  Athens  Aquatics   Swimming  CHASAPIS, Spiridon     GRE

Type of Resulting Object: <class 'pandas.core.frame.DataFrame'>
```

---

9. Write a program to delete a column "City" from the DataFrame and print the new dataframe??

```
#Code here
df = df.drop(columns=['City'])
print(df)

       Year      Sport           Discipline                    Athlete
\
```

```
0      1896   Aquatics              Swimming              HAJOS, Alfred

1      1896   Aquatics              Swimming            HERSCHMANN, Otto

2      1896   Aquatics              Swimming           DRIVAS, Dimitrios

3      1896   Aquatics              Swimming          MALOKINIS, Ioannis

4      1896   Aquatics              Swimming          CHASAPIS, Spiridon

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

31160  2012  Wrestling  Wrestling Freestyle          JANIKOWSKI, Damian

31161  2012  Wrestling  Wrestling Freestyle  REZAEI, Ghasem Gholamreza

31162  2012  Wrestling  Wrestling Freestyle             TOTROV, Rustam

31163  2012  Wrestling  Wrestling Freestyle           ALEKSANYAN, Artur

31164  2012  Wrestling  Wrestling Freestyle             LIDBERG, Jimmy


      Country Gender                         Event   Medal
0         HUN    Men              100M Freestyle    Gold
1         AUT    Men              100M Freestyle  Silver
2         GRE    Men  100M Freestyle For Sailors  Bronze
3         GRE    Men  100M Freestyle For Sailors    Gold
4         GRE    Men  100M Freestyle For Sailors  Silver
...       ...    ...                         ...     ...
31160     POL    Men                   Wg 84 KG  Bronze
31161     IRI    Men                   Wg 96 KG    Gold
31162     RUS    Men                   Wg 96 KG  Silver
31163     ARM    Men                   Wg 96 KG  Bronze
31164     SWE    Men                   Wg 96 KG  Bronze

[31165 rows x 8 columns]
```

---

10. Write a program to rename a column "Athlete" to "Participants" in the DataFrame?

```
#Code here
df = df.rename(columns={'Athlete': 'Participants'})
print(df)

       Year      Sport           Discipline               Participants
\
0      1896   Aquatics              Swimming              HAJOS, Alfred

1      1896   Aquatics              Swimming            HERSCHMANN, Otto
```

```
2      1896   Aquatics             Swimming          DRIVAS, Dimitrios
3      1896   Aquatics             Swimming        MALOKINIS, Ioannis
4      1896   Aquatics             Swimming        CHASAPIS, Spiridon
...     ...        ...                  ...                       ...
31160  2012  Wrestling  Wrestling Freestyle         JANIKOWSKI, Damian
31161  2012  Wrestling  Wrestling Freestyle  REZAEI, Ghasem Gholamreza
31162  2012  Wrestling  Wrestling Freestyle             TOTROV, Rustam
31163  2012  Wrestling  Wrestling Freestyle          ALEKSANYAN, Artur
31164  2012  Wrestling  Wrestling Freestyle             LIDBERG, Jimmy
```

	Country	Gender	Event	Medal
0	HUN	Men	100M Freestyle	Gold
1	AUT	Men	100M Freestyle	Silver
2	GRE	Men	100M Freestyle For Sailors	Bronze
3	GRE	Men	100M Freestyle For Sailors	Gold
4	GRE	Men	100M Freestyle For Sailors	Silver
...	...	...	...	...
31160	POL	Men	Wg 84 KG	Bronze
31161	IRI	Men	Wg 96 KG	Gold
31162	RUS	Men	Wg 96 KG	Silver
31163	ARM	Men	Wg 96 KG	Bronze
31164	SWE	Men	Wg 96 KG	Bronze

```
[31165 rows x 8 columns]
```

---

11. Write a program to select and print the "Participants", "Medal" and "Event" columns for index 10 and 201 from the DataFrame.?

```
#Code here
selected_data = df.loc[[10, 201], ['Participants', 'Medal', 'Event']]
print(selected_data)
```

```
          Participants   Medal           Event
10   PEPANOS, Antonios  Silver  400M Freestyle
201    LISTER, William    Gold      Water Polo
```

---

12. Write a program to display the first 10 rows of the df DataFrame and check for any missing values in each column.?

```
#Code here
print(df.head(10))
print("\nMissing Values:")
print(df.isnull().sum())
```

```
   Year     Sport Discipline           Participants Country Gender  \
0  1896  Aquatics   Swimming          HAJOS, Alfred     HUN    Men
1  1896  Aquatics   Swimming       HERSCHMANN, Otto     AUT    Men
2  1896  Aquatics   Swimming      DRIVAS, Dimitrios     GRE    Men
3  1896  Aquatics   Swimming     MALOKINIS, Ioannis     GRE    Men
4  1896  Aquatics   Swimming     CHASAPIS, Spiridon     GRE    Men
5  1896  Aquatics   Swimming  CHOROPHAS, Efstathios     GRE    Men
6  1896  Aquatics   Swimming          HAJOS, Alfred     HUN    Men
7  1896  Aquatics   Swimming       ANDREOU, Joannis     GRE    Men
8  1896  Aquatics   Swimming  CHOROPHAS, Efstathios     GRE    Men
9  1896  Aquatics   Swimming          NEUMANN, Paul     AUT    Men

                        Event   Medal
0              100M Freestyle    Gold
1              100M Freestyle  Silver
2  100M Freestyle For Sailors  Bronze
3  100M Freestyle For Sailors    Gold
4  100M Freestyle For Sailors  Silver
5             1200M Freestyle  Bronze
6             1200M Freestyle    Gold
7             1200M Freestyle  Silver
8              400M Freestyle  Bronze
9              400M Freestyle    Gold

Missing Values:
Year            0
Sport           0
Discipline      0
Participants    0
Country         4
Gender          0
Event           0
Medal           0
dtype: int64
```

---

13. Write a program to select and print the "Country" and "Medal" columns from the df DataFrame, and then print the number of unique values in each of these columns.?

```
#Code here
selected_columns = df[['Country', 'Medal']]
```

```
print(selected_columns)
print("\nNumber of Unique Values:")
print("Country:", selected_columns['Country'].nunique())
print("Medal:", selected_columns['Medal'].nunique())
```

```
      Country   Medal
0         HUN    Gold
1         AUT  Silver
2         GRE  Bronze
3         GRE    Gold
4         GRE  Silver
...       ...     ...
31160     POL  Bronze
31161     IRI    Gold
31162     RUS  Silver
31163     ARM  Bronze
31164     SWE  Bronze

[31165 rows x 2 columns]

Number of Unique Values:
Country: 147
Medal: 3
```

---

14. Write a program to group the df DataFrame by the "Country" column and calculate the total number of medals for each country. Print the result.?

```
#Code here
df = pd.read_csv(file_path)
medal_counts = df.groupby('Country')['Medal'].count()
print(medal_counts)
```

```
Country
AFG      2
AHO      1
ALG     15
ANZ     29
ARG    259
      ...
VIE      2
YUG    435
ZAM      2
ZIM     23
ZZX     48
Name: Medal, Length: 147, dtype: int64
```

---

15. Write a program to sort the df DataFrame by the "Year" column in descending order and print the first 5 rows of the sorted DataFrame.?

```
#Code here
sorted_df = df.sort_values(by='Year', ascending=False)
print(sorted_df.head())
```

```
       Year    City      Sport           Discipline
Athlete  \
31154  2012  London  Wrestling  Wrestling Freestyle      JULFALAKYAN,
Arsen
31095  2012  London  Wrestling  Wrestling Freestyle         GHASEMI,
Komeil
31126  2012  London  Wrestling  Wrestling Freestyle  GOUDARZI, Sadegh
Saeed
31164  2012  London  Wrestling  Wrestling Freestyle          LIDBERG,
Jimmy
31163  2012  London  Wrestling  Wrestling Freestyle       ALEKSANYAN,
Artur

      Country Gender     Event   Medal
31154     ARM    Men  Wg 74 KG  Silver
31095     IRI    Men  Wf 120KG  Bronze
31126     IRI    Men  Wf 74 KG  Silver
31164     SWE    Men  Wg 96 KG  Bronze
31163     ARM    Men  Wg 96 KG  Bronze
```

---

16. Write a program to add a new column "Century" to the df DataFrame. Classify each event as occurring in the "20th Century" if the "Year" is less than 2000, or the "21st Century" if the "Year" is 2000 or later. Print the first 5 rows to verify the change.

```
#Code here
df['Century'] = df['Year'].apply(lambda x: '20th Century' if x <
2000 else '21st Century')
print(df.head())
```

```
   Year    City     Sport Discipline             Athlete Country
Gender  \
0  1896  Athens  Aquatics   Swimming       HAJOS, Alfred     HUN
Men
1  1896  Athens  Aquatics   Swimming    HERSCHMANN, Otto     AUT
Men
2  1896  Athens  Aquatics   Swimming   DRIVAS, Dimitrios     GRE
Men
3  1896  Athens  Aquatics   Swimming  MALOKINIS, Ioannis     GRE
Men
4  1896  Athens  Aquatics   Swimming  CHASAPIS, Spiridon     GRE
Men
```

```
                         Event   Medal       Century
0              100M Freestyle    Gold  20th Century
1              100M Freestyle  Silver  20th Century
2  100M Freestyle For Sailors  Bronze  20th Century
3  100M Freestyle For Sailors    Gold  20th Century
4  100M Freestyle For Sailors  Silver  20th Century
```

---

17. Write a program to filter and print rows from the df DataFrame where the "Country" is "USA" and the "Medal" is "Gold". Also, count the number of rows that meet this condition.

```
#Code here
filtered_df = df[(df['Country'] == 'USA') & (df['Medal'] == 'Gold')]
print(filtered_df)
print("\nNumber of Rows:", len(filtered_df))
```

```
       Year    City       Sport           Discipline  \
13     1896  Athens   Athletics            Athletics
15     1896  Athens   Athletics            Athletics
21     1896  Athens   Athletics            Athletics
27     1896  Athens   Athletics            Athletics
29     1896  Athens   Athletics            Athletics
...     ...     ...         ...                  ...
30955  2012  London      Tennis               Tennis
30970  2012  London  Volleyball     Beach Volleyball
30971  2012  London  Volleyball     Beach Volleyball
31125  2012  London   Wrestling  Wrestling Freestyle
31133  2012  London   Wrestling  Wrestling Freestyle

                         Athlete Country Gender              Event Medal
\
13                 BURKE, Thomas     USA    Men               100M  Gold

15                CURTIS, Thomas     USA    Men      110M Hurdles  Gold

21                 BURKE, Thomas     USA    Men               400M  Gold

27               GARRETT, Robert     USA    Men      Discus Throw  Gold

29                 CLARK, Ellery     USA    Men         High Jump  Gold

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

30955          WILLIAMS, Serena     USA  Women           Singles  Gold

30970                MAY, Misty     USA  Women  Beach Volleyball  Gold

30971     WALSH JENNINGS, Kerri     USA  Women  Beach Volleyball  Gold
```

```
31125  BURROUGHS, Jordan Ernest     USA    Men          Wf 74 KG  Gold

31133     VARNER, Jacob Stephen     USA    Men          Wf 96 KG  Gold


             Century
13     20th Century
15     20th Century
21     20th Century
27     20th Century
29     20th Century
...             ...
30955  21st Century
30970  21st Century
30971  21st Century
31125  21st Century
31133  21st Century

[2235 rows x 10 columns]

Number of Rows: 2235
```

---

18. Write a program to create a new column "Medal Points" in the df DataFrame. Assign 3 points for "Gold", 2 points for "Silver", and 1 point for "Bronze". Print the first 5 rows to verify the changes.

```
#Code here
df['Medal Points'] = df['Medal'].map({'Gold': 3, 'Silver': 2,
'Bronze':
1})
print(df.head())
```

```
   Year    City     Sport Discipline             Athlete Country
Gender  \
0  1896  Athens  Aquatics   Swimming       HAJOS, Alfred     HUN
Men
1  1896  Athens  Aquatics   Swimming    HERSCHMANN, Otto     AUT
Men
2  1896  Athens  Aquatics   Swimming   DRIVAS, Dimitrios     GRE
Men
3  1896  Athens  Aquatics   Swimming  MALOKINIS, Ioannis     GRE
Men
4  1896  Athens  Aquatics   Swimming  CHASAPIS, Spiridon     GRE
Men

                       Event   Medal       Century  Medal Points
0             100M Freestyle    Gold  20th Century             3
```

```
1              100M Freestyle  Silver  20th Century             2
2  100M Freestyle For Sailors  Bronze  20th Century             1
3  100M Freestyle For Sailors    Gold  20th Century             3
4  100M Freestyle For Sailors  Silver  20th Century             2
```

---

19. Write a program to filter the df DataFrame to include only rows where the "Country" is "USA". Then, group the filtered DataFrame by "Year" and calculate the total number of medals won by the USA each year. Print the results.

```
#Code here
usa_df = df[df['Country'] == 'USA']
medals_by_year = usa_df.groupby('Year')['Medal'].count()
print(medals_by_year)
```

```
Year
1896     20
1900     55
1904    394
1908     63
1912    101
1920    193
1924    198
1928     84
1932    181
1936     92
1948    148
1952    130
1956    118
1960    112
1964    150
1968    149
1972    155
1976    155
1984    333
1988    193
1992    224
1996    260
2000    248
2004    264
2008    315
2012    250
Name: Medal, dtype: int64
```

---

20. Write a program to identify the athlete who has won the most medals. Display the athlete's name and the total number of medals they have won.

```
#Code here
```

---

21. Write a program to group the df DataFrame by the "Sport" column and calculate the number of each type of medal (Gold, Silver, Bronze) for each sport. Print the results.

```
#Code here
```

Good Luck!