Answers PDF

All code is preceded by:

```
import pandas as pd
import numpy as np

data = pd.read_csv("https://elasticbeanstalk-us-east-1-712866102994.s3.amazonaws.com/data/OLYMPICS_athlete_events.csv")
```

1. How many years of data and how many different countries are present in the data set.

Code:

```
numberOfYears = len(pd.unique(data["Year"]))
print("There are", numberOfYears, "different years represented in the data")

numberOfCountries = len(pd.unique(data["NOC"]))
print("There are", numberOfCountries, "different countries represented in the data")
print()
```

Result:

```
There are 35 different years represented in the data
There are 230 different countries represented in the data
```

2. Provide a medal count of USA, Canada and Mexico for all of the years in the data set. Make sure you distinguish between counts for Gold Silver and Bronze.

Code:

```
#2
usaGold = 0
usaSilver = 0
usaBronze = 0
mexGold = 0
mexSilver = 0
mexBronze = 0
canGold = 0
canSilver = 0
canBronze = 0

for index, row in data.iterrows():
  if (row["NOC"] == "USA"):
    if(row["Medal"] == "Gold"):
      usaGold += 1
    elif (row["Medal"] == "Silver"):
      usaSilver += 1
    elif (row["Medal"] == "Bronze"):
      usaBronze += 1
    else:
      continue
  elif (row["NOC"] == "MEX"):
    if(row["Medal"] == "Gold"):
      mexGold += 1
    elif (row["Medal"] == "Silver"):
      mexSilver += 1
    elif (row["Medal"] == "Bronze"):
      mexBronze += 1
    else:
      continue
  elif (row["NOC"] == "CAN"):
    if(row["Medal"] == "Gold"):
      canGold += 1
    elif (row["Medal"] == "Silver"):
      canSilver += 1
    elif (row["Medal"] == "Bronze"):
      canBronze += 1
    else:
      continue


print("The number of USA gold medals is", usaGold)
print("The number of USA silver medals is", usaSilver)
print("The number of USA bronze medals is", usaBronze)
print("The number of MEX gold medals is", mexGold)
print("The number of MEX silver medals is", mexSilver)
print("The number of MEX bronze medals is", mexBronze)
print("The number of CAN gold medals is", canGold)
print("The number of CAN silver medals is", canSilver)
print("The number of CAN bronze medals is", canBronze)
print()
```

Result:

```
The number of USA gold medals is 2638
The number of USA silver medals is 1641
The number of USA bronze medals is 1358
The number of MEX gold medals is 30
The number of MEX silver medals is 26
The number of MEX bronze medals is 54
The number of CAN gold medals is 463
The number of CAN silver medals is 438
The number of CAN bronze medals is 451
```

3. How many Olympic events took place in the summer and winter after the year 2000? Distinguish the counts between them.

Code:

```
eventData = data[["Season", "Year", "Event"]]

summerData = eventData[(eventData["Season"] == "Summer") & (eventData["Year"] > 2000)]
winterData = eventData[(eventData["Season"] == "Winter") & (eventData["Year"] > 2000)]

summerEvents = len(pd.unique(summerData["Event"]))
winterEvents = len(pd.unique(winterData["Event"]))

print("The number of different Summer Olympic events taking place after 2000 is", summerEvents)
print("The number of different Winter Olympic events taking place after 2000 is", winterEvents)
print()
```

Result:

```
The number of USA gold medals is 2638
The number of USA silver medals is 1641
The number of USA bronze medals is 1358
The number of MEX gold medals is 30
The number of MEX silver medals is 26
The number of MEX bronze medals is 54
The number of CAN gold medals is 463
The number of CAN silver medals is 438
The number of CAN bronze medals is 451
```

4. Compare medal counts across all 3 (Gold, Silver and Bronze) for summer olympics for China and the USA for the year 2016.

Code:

```
summer2016_data = olympic_data[(olympic_data["Year"] == 2016) & (olympic_data["Season"] == "Summer")]

china_gold = 0
china_silver = 0
china_bronze = 0

usa_gold = 0
usa_silver = 0
usa_bronze = 0

for index,row in summer2016_data.iterrows():
  if row["NOC"] == "CHN":
    if row["Medal"] == "Gold":
      china_gold += 1
    elif row["Medal"] == "Silver":
      china_silver += 1
    elif row["Medal"] == "Bronze":
      china_bronze += 1
    else:
      continue
  elif row["NOC"] == "USA":
    if row["Medal"] == "Gold":
      usa_gold += 1
    elif row["Medal"] == "Silver":
      usa_silver += 1
    elif row["Medal"] == "Bronze":
      usa_bronze += 1
    else:
      continue
  else:
      continue

print("China's medals in 2016 Summer Olympics:", china_gold, "Gold, ", china_silver, "Silver, ", china_bronze, "Bronze.")
print("USA's medals in 2016 Summer Olympics:", usa_gold, "Gold, ", usa_silver, "Silver, ", usa_bronze, "Bronze.")
```

Result:

```
China's medals in 2016 Summer Olympics: 46 Gold,  30 Silver,  37 Bronze.
USA's medals in 2016 Summer Olympics: 139 Gold,  54 Silver,  71 Bronze.
```

5. How many medals did the US win in Summer vs Winter games after the year 2000.

Code:

```
USA_medals_data = olympic_data[(olympic_data["Year"] > 2000) & (olympic_data["NOC"] == "USA")]
Summer_medals= 0
Winter_medals= 0

for index,row in USA_medals_data.iterrows():
  if row["Season"] == "Summer":
    if row["Medal"] == "Gold" or row["Medal"] =="Silver" or row["Medal"] =="Bronze":
      Summer_medals += 1
    else:
      continue
  elif row["Season"] == "Winter":
    if row["Medal"] == "Gold" or row["Medal"] =="Silver" or row["Medal"] =="Bronze":
      Winter_medals += 1
    else:
      continue
  else:
      continue
print("The US won",Summer_medals,"medals in the Summer games after the year 2000")
print("The US won",Winter_medals,"medals in the Winter games after the year 2000")
```

Result:

```
The US won 1092 medals in the Summer games after the year 2000
The US won 297 medals in the Winter games after the year 2000
```

6. What is the average age of male vs female contestants for the USA and China in games after the year 2000.

Code:

```
usa_contestants = olympic_data[(olympic_data["Year"] > 2000) & (olympic_data["NOC"] == "USA")]
usa_age_avg = usa_contestants.groupby(["Sex"]).mean()

print("The average age of USA contestants after 2000:", usa_age_avg["Age"])

china_contestants = olympic_data[(olympic_data["Year"] > 2000) & (olympic_data["NOC"] == "CHN")]
china_age_avg = china_contestants.groupby(["Sex"]).mean()

print("The average age of a China contestants after 2000:", china_age_avg["Age"])
```

Result:

```
The average age of USA contestants after 2000: Sex
F    25.745737
M    26.888395
Name: Age, dtype: float64
The average age of a China contestants after 2000: Sex
F    23.216797
M    24.257627
```

7. List the different sports in which the US has won a Gold medal in.

Code:

```
data_3 = data[["Sport","NOC","Medal"]]
#print(data_3)

data_USGoldSports = data_3[(data_3["NOC"] == "USA") & (data_3["Medal"] == "Gold")]
#print(Data_USGoldSports)

data_USGoldDiffSports = data_USGoldSports.drop_duplicates(subset = ["Sport"])
data_USGoldDiffSports
```

Result:

	Sport	NOC	Medal
283	Golf	USA	Gold
609	Basketball	USA	Gold
710	Rowing	USA	Gold
739	Baseball	USA	Gold
787	Athletics	USA	Gold
1239	Wrestling	USA	Gold
1472	Shooting	USA	Gold
1727	Boxing	USA	Gold
1843	Swimming	USA	Gold
2031	Tennis	USA	Gold
2911	Football	USA	Gold
4157	Figure Skating	USA	Gold
4965	Sailing	USA	Gold
5988	Softball	USA	Gold
6774	Equestrianism	USA	Gold
6812	Snowboarding	USA	Gold
6863	Weightlifting	USA	Gold
9382	Water Polo	USA	Gold
9390	Alpine Skiing	USA	Gold

8. Compare the countries of Denmark, Norway and Sweden in medal counts across the entire data set.

Code:

```
#Denmark
Den = data["Team"] == "Denmark"
data["Den"]= Den

den_Medal = data[["Den","Medal"]].groupby("Den").count()
print("Denmark's medal counts is:")
print(den_Medal)
print()

#Norway
Nor = data["Team"] == "Norway"
data["Nor"] = Nor

nor_Medal = data[["Nor","Medal"]].groupby("Nor").count()
print("Norway's medal counts is:")
print(nor_Medal)
print()

#Sweden
Swe = data["Team"] == "Sweden"
data["Swe"] = Swe

swe_Medal = data[["Swe","Medal"]].groupby("Swe").count()
print("Sweden's medal counts is:")
print(swe_Medal)
print()
```

Result:

```
Denmark's medal counts is:
       Medal
Den
False  39230
True     553

Norway's medal counts is:
       Medal
Nor
False  38873
True     910

Sweden's medal counts is:
       Medal
Swe
False  38349
True    1434
```

9. Identify which countries have won medals in Long Jump (Event column data use str.contains()) and list how many they have won.

Code:

```
data_Long = data[data.Event.str.contains("Long Jump")]
#print(data_Long)


groupedLongTeam = data_Long[["Team","Medal"]].groupby("Team").count()
#print(groupedLongTeam)
df = groupedLongTeam[groupedLongTeam['Medal'] != 0]
df
```

Result:

Team	Medal
Argentina	1
Australia	4
Belgium	1
Brazil	1
Bulgaria	1
Canada	2
Cuba	2
Czechoslovakia	1
East Germany	7
Finland	1
France	2
Germany	4
Great Britain	9
Greece	2
Haiti	1
Hungary	2
Italy	4
Jamaica	1
Japan	2
New Zealand	1
Nigeria	2
Norway	1
Panama	1
Poland	3
Romania	3
Russia	6
Serbia	1
South Africa	2

10. List the countries that have won a medal in Handball.

Code:

```
datahandball = data[["Sport","Team","Medal"]]

handballcountriesgold = datahandball[(datahandball["Sport"] == "Handball") & (datahandball["Medal"] == "Gold")]
handballcountriessilver = datahandball[(datahandball["Sport"] == "Handball") & (datahandball["Medal"] == "Silver")]
handballcountriesbronze = datahandball[(datahandball["Sport"] == "Handball") & (datahandball["Medal"] == "Bronze")]

Data_handballcountriesgold = handballcountriesgold.drop_duplicates(subset = ["Team"])
Data_handballcountriessilver = handballcountriessilver.drop_duplicates(subset = ["Team"])
Data_handballcountriesbronze = handballcountriesbronze.drop_duplicates(subset = ["Team"])

print("The following countries have won a gold medal in handball:")
print(Data_handballcountriesgold)
print("The following countries have won a silver medal in handball:")
print(Data_handballcountriessilver)
print("The following countries have won a bronze medal in handball:")
print(Data_handballcountriesbronze)
print()
```

Result:

```
The following countries have won a gold medal in handball:
         Sport          Team Medal
79     Handball        Norway  Gold
173    Handball        France  Gold
6518   Handball    Yugoslavia  Gold
6597   Handball       Denmark  Gold
8041   Handball  Soviet Union  Gold
13356  Handball       Croatia  Gold
13803  Handball       Germany  Gold
14321  Handball  Unified Team  Gold
21092  Handball  East Germany  Gold
23362  Handball        Russia  Gold
37932  Handball   South Korea  Gold
The following countries have won a silver medal in handball:
         Sport            Team   Medal
175    Handball          France  Silver
6517   Handball      Yugoslavia  Silver
6702   Handball          Norway  Silver
7020   Handball          Sweden  Silver
7608   Handball          Russia  Silver
8042   Handball    Soviet Union  Silver
10102  Handball         Iceland  Silver
12333  Handball    East Germany  Silver
13596  Handball         Hungary  Silver
14614  Handball      Montenegro  Silver
15350  Handball         Austria  Silver
16418  Handball         Germany  Silver
18961  Handball  Czechoslovakia  Silver
22298  Handball         Romania  Silver
40790  Handball     South Korea  Silver
71906  Handball    West Germany  Silver
The following countries have won a bronze medal in handball:
         Sport          Team   Medal
2263   Handball         Spain  Bronze
4793   Handball       Croatia  Bronze
5458   Handball        Norway  Bronze
6393   Handball   South Korea  Bronze
7868   Handball       Hungary  Bronze
7923   Handball  Soviet Union  Bronze
7924   Handball  Unified Team  Bronze
8142   Handball        Poland  Bronze
15678  Handball    Yugoslavia  Bronze
17460  Handball       Romania  Bronze
23400  Handball   Switzerland  Bronze
26025  Handball       Ukraine  Bronze
39514  Handball         China  Bronze
40444  Handball        Russia  Bronze
52363  Handball        France  Bronze
56128  Handball       Germany  Bronze
93365  Handball  East Germany  Bronze
```

11. What is the average height and weight of individuals that compete in Wrestling and Weightlifting. Distinguish between male and female competitors for each of the sports.

Code:

```
wrestling_contestants = data[(data["Sport"] == "Wrestling")]
wrestling_sex = wrestling_contestants.groupby(["Sex"]).mean()
print("The average height of wrestling contestants:", wrestling_sex["Height"])
print("The average weight of wrestling contestants:", wrestling_sex["Weight"])

weightlifting_contestants = data[(data["Sport"] == "Weightlifting")]
weightlifting_sex = weightlifting_contestants.groupby(["Sex"]).mean()
print("The average height of weightlifting contestants:", weightlifting_sex["Height"])
print("The average weight of weightlifting contestants:", weightlifting_sex["Weight"])
```

Result:

```
The average height of wrestling contestants: Sex
F    163.865132
M    172.870686
Name: Height, dtype: float64
The average weight of wrestling contestants: Sex
F    60.554455
M    76.400640
Name: Weight, dtype: float64
The average height of weightlifting contestants: Sex
F    160.467391
M    169.153061
Name: Height, dtype: float64
The average weight of weightlifting contestants: Sex
F    67.724622
M    80.251796
Name: Weight, dtype: float64
```

12. (Custom): Which cities have hosted the olympic games and when did they do so?

Code:

```
datacity = data[["Year","Season","City"]]
Data_city = datacity.drop_duplicates(subset = ["City"])
print("The following cities hosted the Summer/Winter Olympics in their respective years:")
print(Data_city)
```

Result:

```
The following cities hosted the Summer/Winter Olympics in their respective years:
      Year  Season                  City
0     1992  Summer              Barcelona
1     2012  Summer                 London
2     1920  Summer              Antwerpen
3     1900  Summer                  Paris
4     1988  Winter                Calgary
6     1992  Winter             Albertville
8     1994  Winter             Lillehammer
26    1932  Summer             Los Angeles
28    2002  Winter          Salt Lake City
29    1952  Summer                Helsinki
30    1980  Winter             Lake Placid
31    2000  Summer                  Sydney
32    1996  Summer                 Atlanta
35    1912  Summer               Stockholm
40    2014  Winter                   Sochi
68    1998  Winter                  Nagano
77    2006  Winter                  Torino
79    2008  Summer                 Beijing
80    2016  Summer          Rio de Janeiro
82    2004  Summer                  Athina
83    1960  Winter            Squaw Valley
85    1964  Winter               Innsbruck
87    1984  Winter                Sarajevo
89    1968  Summer              Mexico City
90    1972  Summer                  Munich
91    1988  Summer                   Seoul
94    1936  Summer                  Berlin
110   1952  Winter                    Oslo
111   1956  Winter       Cortina d'Ampezzo
128   1956  Summer               Melbourne
129   1960  Summer                    Roma
133   1928  Summer               Amsterdam
145   1976  Summer                 Montreal
188   1980  Summer                   Moskva
192   1964  Summer                    Tokyo
245   2010  Winter                Vancouver
266   1968  Winter                Grenoble
626   1972  Winter                 Sapporo
672   1924  Winter                Chamonix
711   1904  Summer                St. Louis
935   1928  Winter             Sankt Moritz
1327  1936  Winter  Garmisch-Partenkirchen
```

Custom Questions:

13. Compare the total number of medals won by the US in 2016 and 2020

Code:

```
US2016_medals = olympic_data[(olympic_data["Year"] == 2016) & (olympic_data["NOC"] == "USA")]
US2016_medals = US2016_medals[US2016_medals['Medal'].notna()]
print('In 2016, the US recieved',US2016_medals['Medal'].count(),'medals')

US2000_medals = olympic_data[(olympic_data["Year"] == 2000) & (olympic_data["NOC"] == "USA") & (olympic_data["Medal"].notna())]
print('In 2000, the US recieved',US2000_medals['Medal'].count(),'medals')
print('The US recieved',US2016_medals['Medal'].count()-US2000_medals['Medal'].count(), 'more medals in 2016 than in 2000')
```

Result:

```
In 2016, the US recieved 264 medals
In 2000, the US recieved 242 medals
The US recieved 22 more medals in 2016 than in 2000
```

14. What is the average BMI of all competitors in the dataset?

Code:

```
86 bmi = data["Weight"] / (.01 * data["Height"])**2
87 avgBMI = bmi.mean()
88
89 print("The average BMI of all competitors in the dataset is", avgBMI)
```

Result:

```
The average BMI of all competitors in the dataset is 22.784519726473317
```

15. Which countries won the different medals in Swimming?

Code:

```
dataswimming = data[["Sport","Team","Medal"]]

swimmingcountriesgold = dataswimming[(dataswimming["Sport"] == "Swimming") & (dataswimming["Medal"] == "Gold")]
swimmingcountriessilver = dataswimming[(dataswimming["Sport"] == "Swimming") & (dataswimming["Medal"] == "Silver")]
swimmingcountriesbronze = dataswimming[(dataswimming["Sport"] == "Swimming") & (dataswimming["Medal"] == "Bronze")]

Data_swimmingcountriesgold = swimmingcountriesgold.drop_duplicates(subset = ["Team"])
Data_swimmingcountriessilver = swimmingcountriessilver.drop_duplicates(subset = ["Team"])
Data_swimmingcountriesbronze = swimmingcountriesbronze.drop_duplicates(subset = ["Team"])

print("The following countries have won a gold medal in swimming:")
print(Data_swimmingcountriesgold)
print("The following countries have won a silver medal in swimming:")
print(Data_swimmingcountriessilver)
print("The following countries have won a bronze medal in swimming:")
print(Data_swimmingcountriesbronze)
print()
```

Result:

The following countries have won a gold medal in swimming:

	Sport	Team	Medal
1772	Swimming	Great Britain	Gold
1843	Swimming	United States	Gold
2110	Swimming	France	Gold
6652	Swimming	Denmark	Gold
7948	Swimming	East Germany	Gold
8449	Swimming	Japan	Gold
9394	Swimming	Australia	Gold
9915	Swimming	Sweden	Gold
13124	Swimming	Kazakhstan	Gold
15881	Swimming	Germany	Gold
16285	Swimming	Canada	Gold
18414	Swimming	Spain	Gold
22546	Swimming	Yugoslavia	Gold
23696	Swimming	Australasia	Gold
24389	Swimming	Bulgaria	Gold
41794	Swimming	Brazil	Gold
45642	Swimming	Zimbabwe	Gold
46768	Swimming	Hungary	Gold
48927	Swimming	New York Athletic Club #1-1	Gold
50497	Swimming	Netherlands	Gold
52367	Swimming	Belgium	Gold
68360	Swimming	South Africa	Gold
68740	Swimming	Soviet Union	Gold
69568	Swimming	Italy	Gold
85043	Swimming	West Germany	Gold
88394	Swimming	Deutscher Schwimm Verband Berlin	Gold
102447	Swimming	Russia	Gold
107146	Swimming	Poland	Gold
108415	Swimming	China	Gold
121306	Swimming	Ukraine	Gold
128404	Swimming	Unified Team	Gold
141263	Swimming	New Zealand	Gold
147806	Swimming	Greece	Gold
155890	Swimming	Lithuania	Gold
156288	Swimming	Tunisia	Gold
160910	Swimming	Romania	Gold
165901	Swimming	Mexico	Gold
169888	Swimming	Suriname	Gold
170080	Swimming	Austria	Gold
182269	Swimming	South Korea	Gold
190858	Swimming	Costa Rica	Gold
213955	Swimming	Singapore	Gold
223711	Swimming	Ireland	Gold
270403	Swimming	Argentina	Gold

The following countries have won a silver medal in swimming:

	Sport	Team	Medal
803	Swimming	Soviet Union	Silver
1462	Swimming	United States	Silver
2112	Swimming	France	Silver
5434	Swimming	Sweden	Silver
6654	Swimming	Denmark	Silver
7353	Swimming	Greece	Silver
7412	Swimming	Australia	Silver
8708	Swimming	Japan	Silver
12549	Swimming	Germany	Silver
13678	Swimming	East Germany	Silver
14244	Swimming	Hungary	Silver
14638	Swimming	Great Britain	Silver
14838	Swimming	Netherlands	Silver
15629	Swimming	Unified Team	Silver
16970	Swimming	Australasia	Silver
18407	Swimming	Spain	Silver
20703	Swimming	Tritons Lillois-2	Silver
21118	Swimming	West Germany	Silver
22547	Swimming	Yugoslavia	Silver
25790	Swimming	Brazil	Silver
30644	Swimming	Italy	Silver
34375	Swimming	Argentina	Silver
37266	Swimming	Serbia	Silver
38522	Swimming	China	Silver
42709	Swimming	Canada	Silver
42850	Swimming	Romania	Silver
45641	Swimming	Zimbabwe	Silver
50650	Swimming	Belgium	Silver
58353	Swimming	Croatia	Silver
66204	Swimming	Cuba	Silver
72966	Swimming	Bulgaria	Silver
78127	Swimming	Belarus	Silver
80788	Swimming	Chicago Athletic Association-2	Silver
95166	Swimming	Austria	Silver
103625	Swimming	Slovenia	Silver
104507	Swimming	Russia	Silver
107144	Swimming	Poland	Silver
121305	Swimming	Ukraine	Silver
134047	Swimming	South Africa	Silver
141262	Swimming	New Zealand	Silver
163011	Swimming	Slovakia	Silver
175245	Swimming	Norway	Silver
182268	Swimming	South Korea	Silver
190870	Swimming	Costa Rica	Silver
219616	Swimming	Finland	Silver

The following countries have won a bronze medal in swimming:

	Sport	Team	Medal
37	Swimming	Finland	Bronze
210	Swimming	Hungary	Bronze
476	Swimming	Canada	Bronze
740	Swimming	South Africa	Bronze
804	Swimming	Soviet Union	Bronze
814	Swimming	Australia	Bronze
892	Swimming	Great Britain	Bronze
1739	Swimming	Austria	Bronze
1847	Swimming	United States	Bronze
5436	Swimming	Sweden	Bronze
8713	Swimming	Japan	Bronze
13552	Swimming	France	Bronze
14500	Swimming	Argentina	Bronze
16042	Swimming	Italy	Bronze
16971	Swimming	Australasia	Bronze
17303	Swimming	West Germany	Bronze
18416	Swimming	Spain	Bronze
19318	Swimming	Cuba	Bronze
21012	Swimming	Netherlands	Bronze
23347	Swimming	Belgium	Bronze
23612	Swimming	Denmark	Bronze
25795	Swimming	Brazil	Bronze
27411	Swimming	Trinidad and Tobago	Bronze
30349	Swimming	East Germany	Bronze
30710	Swimming	Germany	Bronze
33513	Swimming	China	Bronze
41558	Swimming	Russia	Bronze
42851	Swimming	Romania	Bronze
45643	Swimming	Zimbabwe	Bronze
47664	Swimming	Poland	Bronze
48148	Swimming	Switzerland	Bronze
58733	Swimming	Greece	Bronze
65352	Swimming	Missouri Athletic Club-3	Bronze
72967	Swimming	Bulgaria	Bronze
78131	Swimming	Belarus	Bronze
119570	Swimming	New Zealand	Bronze
130278	Swimming	Unified Team	Bronze
137392	Swimming	Pupilles de Neptune de Lille-1	Bronze
156289	Swimming	Tunisia	Bronze
169890	Swimming	Suriname	Bronze
172876	Swimming	Norway	Bronze
190860	Swimming	Costa Rica	Bronze
196533	Swimming	Mexico	Bronze
216657	Swimming	Ukraine	Bronze
223712	Swimming	Ireland	Bronze
252357	Swimming	Venezuela	Bronze
265810	Swimming	Philippines	Bronze