Scenario

You are a sports data analyst and you have been tasked with summarizing data from the matches from a previous EuroCup. Your manager would like the following questions answered.

In [5]:
```
## Import the library
import pandas as pd
import numpy  as np
import matplotlib.pyplot as   plt
import seaborn as sns
import plotly.express as px
from ydata_profiling import ProfileReport
%matplotlib inline
import warnings
warnings.filterwarnings("ignore")
plt.style.use('fivethirtyeight')
sns.set()

pd.options.display.float_format = '{:,.2f}'.format
pd.options.display.max_rows = None
pd.options.display.max_columns = None
```

In [6]:
```
data = pd.read_csv('https://raw.githubusercontent.com/guipsamora/pandas_exercises/master/02_Filtering_%26_Sorting/Euro12/Euro_201
```

In [7]:
```
data.sample(10)
```

Out[7]:

	Team	Goals	Shots on target	Shots off target	Shooting Accuracy	% Goals-to-shots	Total shots (inc. Blocked)	Hit Woodwork	Penalty goals	Penalties not scored	Headed goals	Passes	Passes completed	Passing Accuracy	Touches	Crosses	D
3	England	5	11	18	50.0%	17.2%	40	0	0	0	3	1488	1200	80.6%	2440	58	
0	Croatia	4	13	12	51.9%	16.0%	32	0	0	0	2	1076	828	76.9%	1706	60	
4	France	3	22	24	37.9%	6.5%	65	1	0	0	0	2066	1803	87.2%	2909	55	
1	Czech Republic	4	13	18	41.9%	12.9%	39	0	0	0	0	1565	1223	78.1%	2358	46	
2	Denmark	4	10	10	50.0%	20.0%	27	1	0	0	3	1298	1082	83.3%	1873	43	
10	Portugal	6	22	42	34.3%	9.3%	82	6	0	0	2	1891	1461	77.2%	2958	91	
13	Spain	12	42	33	55.9%	16.0%	100	0	1	0	2	4317	3820	88.4%	5585	69	
12	Russia	5	9	31	22.5%	12.5%	59	2	0	0	1	1602	1345	83.9%	2278	40	
8	Netherlands	2	12	36	25.0%	4.1%	60	2	0	0	0	1556	1381	88.7%	2163	50	
14	Sweden	5	17	19	47.2%	13.8%	39	3	0	0	1	1192	965	80.9%	1806	44	

In [8]:
```
data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 16 entries, 0 to 15
Data columns (total 35 columns):
 #   Column                     Non-Null Count  Dtype
---  ------                     --------------  -----
 0   Team                       16 non-null     object
 1   Goals                      16 non-null     int64
 2   Shots on target            16 non-null     int64
 3   Shots off target           16 non-null     int64
 4   Shooting Accuracy          16 non-null     object
 5   % Goals-to-shots           16 non-null     object
 6   Total shots (inc. Blocked) 16 non-null     int64
 7   Hit Woodwork               16 non-null     int64
 8   Penalty goals              16 non-null     int64
 9   Penalties not scored       16 non-null     int64
 10  Headed goals               16 non-null     int64
 11  Passes                     16 non-null     int64
 12  Passes completed           16 non-null     int64
 13  Passing Accuracy           16 non-null     object
 14  Touches                    16 non-null     int64
 15  Crosses                    16 non-null     int64
 16  Dribbles                   16 non-null     int64
 17  Corners Taken              16 non-null     int64
 18  Tackles                    16 non-null     int64
 19  Clearances                 16 non-null     int64
 20  Interceptions              16 non-null     int64
 21  Clearances off line        15 non-null     float64
 22  Clean Sheets               16 non-null     int64
 23  Blocks                     16 non-null     int64
 24  Goals conceded             16 non-null     int64
 25  Saves made                 16 non-null     int64
 26  Saves-to-shots ratio       16 non-null     object
 27  Fouls Won                  16 non-null     int64
 28  Fouls Conceded             16 non-null     int64
```

```
 29  Offsides                16 non-null     int64
 30  Yellow Cards            16 non-null     int64
 31  Red Cards               16 non-null     int64
 32  Subs on                 16 non-null     int64
 33  Subs off                16 non-null     int64
 34  Players Used            16 non-null     int64
dtypes: float64(1), int64(29), object(5)
memory usage: 4.5+ KB
```

In [9]:
```
data.shape
```

Out[9]:
```
(16, 35)
```

In [12]:
```
data.duplicated().sum()
```

Out[12]:
```
0
```

How many teams participated in the Euro2012?

In [30]:
```
data.Team.nunique()
```

Out[30]:
```
16
```

What is the number of columns in the dataset?

In [31]:
```
data.shape[1]
```

Out[31]:
```
35
```

View only the columns Team, Yellow Cards and Red Cards and assign them to a dataframe called discipline.

In [35]:
```
discipline=data[['Team','Yellow Cards','Red Cards']]
discipline
```

Out[35]:

	Team	Yellow Cards	Red Cards
0	Croatia	9	0
1	Czech Republic	7	0
2	Denmark	4	0
3	England	5	0
4	France	6	0
5	Germany	4	0
6	Greece	9	1
7	Italy	16	0
8	Netherlands	5	0
9	Poland	7	1
10	Portugal	12	0
11	Republic of Ireland	6	1
12	Russia	6	0
13	Spain	11	0
14	Sweden	7	0
15	Ukraine	5	0

Sort the teams by Red Cards, then to Yellow Cards.

In [41]:
```
discipline = discipline.sort_values(by=["Red Cards", "Yellow Cards"], ascending=False)
discipline
```

Out[41]:

	Team	Yellow Cards	Red Cards
6	Greece	9	1
9	Poland	7	1
11	Republic of Ireland	6	1
7	Italy	16	0
10	Portugal	12	0
13	Spain	11	0
0	Croatia	9	0
1	Czech Republic	7	0
14	Sweden	7	0
4	France	6	0
12	Russia	6	0
3	England	5	0
8	Netherlands	5	0
15	Ukraine	5	0

2	Denmark	4	0
5	Germany	4	0

Calculate the mean Yellow Cards given per Team.

In [43]:
```
discipline.groupby("Team")['Yellow Cards'].mean().reset_index()
```

Out[43]:

	Team	Yellow Cards
0	Croatia	9.00
1	Czech Republic	7.00
2	Denmark	4.00
3	England	5.00
4	France	6.00
5	Germany	4.00
6	Greece	9.00
7	Italy	16.00
8	Netherlands	5.00
9	Poland	7.00
10	Portugal	12.00
11	Republic of Ireland	6.00
12	Russia	6.00
13	Spain	11.00
14	Sweden	7.00
15	Ukraine	5.00

Filter teams that scored more than 6 goals.

In [112]:
```
data.query('Goals > 6')
```

Out[112]:

	Team	Goals	Shots on target	Shots off target	Shooting Accuracy	% Goals-to-shots	Total shots (inc. Blocked)	Hit Woodwork	Penalty goals	Penalties not scored	Headed goals	Passes	Passes completed	Passing Accuracy	Touches	Crosses	Drib
5	Germany	10	32	32	47.8%	15.6%	80	2	1	0	2	2774	2427	87.4%	3761	101	
13	Spain	12	42	33	55.9%	16.0%	100	0	1	0	2	4317	3820	88.4%	5585	69	

Select the teams that start with the letter G.

In [57]:
```
data[data.Team.str.startswith('G')]
```

Out[57]:

	Team	Goals	Shots on target	Shots off target	Shooting Accuracy	% Goals-to-shots	Total shots (inc. Blocked)	Hit Woodwork	Penalty goals	Penalties not scored	Headed goals	Passes	Passes completed	Passing Accuracy	Touches	Crosses	Dribb
5	Germany	10	32	32	47.8%	15.6%	80	2	1	0	2	2774	2427	87.4%	3761	101	
6	Greece	5	8	18	30.7%	19.2%	32	1	1	1	0	1187	911	76.7%	2016	52	

Select the first 7 columns.

In [61]:
```
data[data.columns[:7]]
```

Out[61]:

	Team	Goals	Shots on target	Shots off target	Shooting Accuracy	% Goals-to-shots	Total shots (inc. Blocked)
0	Croatia	4	13	12	51.9%	16.0%	32
1	Czech Republic	4	13	18	41.9%	12.9%	39
2	Denmark	4	10	10	50.0%	20.0%	27
3	England	5	11	18	50.0%	17.2%	40
4	France	3	22	24	37.9%	6.5%	65
5	Germany	10	32	32	47.8%	15.6%	80
6	Greece	5	8	18	30.7%	19.2%	32
7	Italy	6	34	45	43.0%	7.5%	110
8	Netherlands	2	12	36	25.0%	4.1%	60
9	Poland	2	15	23	39.4%	5.2%	48
10	Portugal	6	22	42	34.3%	9.3%	82
11	Republic of Ireland	1	7	12	36.8%	5.2%	28
12	Russia	5	9	31	22.5%	12.5%	59
13	Spain	12	42	33	55.9%	16.0%	100
14	Sweden	5	17	19	47.2%	13.8%	39

15	Ukraine	2	7	26	21.2%	6.0%	38

Select all columns except the last 3.

In [65]:
```
data[data.columns[:-3]]
```

Out[65]:

	Team	Goals	Shots on target	Shots off target	Shooting Accuracy	% Goals-to-shots	Total shots (inc. Blocked)	Hit Woodwork	Penalty goals	Penalties not scored	Headed goals	Passes	Passes completed	Passing Accuracy	Touches	Crosses	D
0	Croatia	4	13	12	51.9%	16.0%	32	0	0	0	2	1076	828	76.9%	1706	60	
1	Czech Republic	4	13	18	41.9%	12.9%	39	0	0	0	0	1565	1223	78.1%	2358	46	
2	Denmark	4	10	10	50.0%	20.0%	27	1	0	0	3	1298	1082	83.3%	1873	43	
3	England	5	11	18	50.0%	17.2%	40	0	0	0	3	1488	1200	80.6%	2440	58	
4	France	3	22	24	37.9%	6.5%	65	1	0	0	0	2066	1803	87.2%	2909	55	
5	Germany	10	32	32	47.8%	15.6%	80	2	1	0	2	2774	2427	87.4%	3761	101	
6	Greece	5	8	18	30.7%	19.2%	32	1	1	1	0	1187	911	76.7%	2016	52	
7	Italy	6	34	45	43.0%	7.5%	110	2	0	0	2	3016	2531	83.9%	4363	75	
8	Netherlands	2	12	36	25.0%	4.1%	60	2	0	0	0	1556	1381	88.7%	2163	50	
9	Poland	2	15	23	39.4%	5.2%	48	0	0	0	1	1059	852	80.4%	1724	55	
10	Portugal	6	22	42	34.3%	9.3%	82	6	0	0	2	1891	1461	77.2%	2958	91	
11	Republic of Ireland	1	7	12	36.8%	5.2%	28	0	0	0	1	851	606	71.2%	1433	43	
12	Russia	5	9	31	22.5%	12.5%	59	2	0	0	1	1602	1345	83.9%	2278	40	
13	Spain	12	42	33	55.9%	16.0%	100	0	1	0	2	4317	3820	88.4%	5585	69	
14	Sweden	5	17	19	47.2%	13.8%	39	3	0	0	1	1192	965	80.9%	1806	44	
15	Ukraine	2	7	26	21.2%	6.0%	38	0	0	0	2	1276	1043	81.7%	1894	33	

Present only the Shooting Accuracy from England, Italy and Russia.

In [106]:
```
Shooting=data[['Team',"Shooting Accuracy"]]
```

In [105]:
```
Shooting[Shooting['Team'].isin(['England', 'Italy', 'Russia'])]
```

Out[105]:

	Team	Shooting Accuracy
3	England	50.0%
7	Italy	43.0%
12	Russia	22.5%