



# CHALLENGE 2: Euro Cup Soccer

---

SHAHZAD AZEEZ

## *Agenda*

1. Scenario
2. Challenging Questions
3. Summary



# *Scenario*

---

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

# Questions Solved

---

- 1.How many teams participated in the Euro2012?
- 2.What is the number of columns in the dataset?
- 3.View only the columns Team, Yellow Cards and Red Cards and assign them to a data frame called discipline.
- 4.Sort the teams by Red Cards, then to Yellow Cards.
- 5.Calculate the mean Yellow Cards given per Team.
- 6.Filter teams that scored more than 6 goals.
- 7.Select the teams that start with the letter G.
- 8.Select the first 7 columns.
- 9.Select all columns except the last 3.
- 10.Present only the Shooting Accuracy from England, Italy and Russia.

# Dataset

---

```
data = pd.read_csv('soccer_data.csv')
data.head()
```

	Team	Goals	Shots on target	Shots off target	Shooting Accuracy	% Goals- to- shots	Total shots (inc. Blocked)	Hit Woodwork	Penalty goals	Penalties not scored	...	Saves made	Saves- to- shots ratio	Fouls Won	Fouls Conceded	Offsides	Yellow Cards	Red Cards	S
0	Croatia	4	13	12	51.9%	16.0%	32	0	0	0	...	13	81.3%	41	62	2	9	0	
1	Czech Republic	4	13	18	41.9%	12.9%	39	0	0	0	...	9	60.1%	53	73	8	7	0	
2	Denmark	4	10	10	50.0%	20.0%	27	1	0	0	...	10	66.7%	25	38	8	4	0	
3	England	5	11	18	50.0%	17.2%	40	0	0	0	...	22	88.1%	43	45	6	5	0	
4	France	3	22	24	37.9%	6.5%	65	1	0	0	...	6	54.6%	36	51	5	6	0	

5 rows × 35 columns

# 1. How many teams participated in the Euro2012?

---

## 1. How many teams participated in the Euro 2012?

```
: # Count the number of unique teams in the 'Team' column
num_teams = data['Team'].nunique()

# Print the number of teams participated in Euro2012
print(f"The number of teams participated in Euro2012 is: {num_teams}")
```

The number of teams participated in Euro2012 is: 16

## 2. What is the number of columns in the dataset?

---

### 2. What is the number of columns in the dataset?

```
# Get the number of columns  
num_columns = data.shape[1]  
  
# Print the number of columns  
print(f"The number of columns in the dataset is: {num_columns}")
```

```
The number of columns in the dataset is: 35
```

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

---

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

```
: # Create a new DataFrame 'discipline' with selected columns
discipline = data[['Team', 'Yellow Cards', 'Red Cards']]

# Print the 'discipline' DataFrame
print(discipline)
```

	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



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

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

```
: # Sort the 'discipline' DataFrame by Red Cards and then Yellow Cards
discipline = discipline.sort_values(by=['Red Cards', 'Yellow Cards'], ascending=False)

# Print the sorted 'discipline' DataFrame
print(discipline)
```

	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

## 5. Calculate the mean Yellow Cards given per Team.

---

### 5. Calculate the mean Yellow Cards given per Team.

```
# Calculate the mean Yellow Cards per Team
mean_yellow_cards = data['Yellow Cards'].mean()

# Print the result
print(f"The mean Yellow Cards given per Team is: {mean_yellow_cards:.0f}")
```

The mean Yellow Cards given per Team is: 7

## 6. Filter teams that scored more than 6 goals.

---

### 6. Filter teams that scored more than 6 goals.

```
# Filter teams that scored more than 6 goals
teams_more_than_6_goals = data[data['Goals'] > 6]

# Display only the 'Team' and 'Goals' column
teams_column = teams_more_than_6_goals[['Team', 'Goals']]

# Print the filtered 'Team' and 'Goals' column
print(teams_column)
```

	Team	Goals
5	Germany	10
13	Spain	12

## 7. Select the teams that start with the letter G.

---

### 7. Select the teams that start with the letter G.

```
: # Display only the 'Team' column
teams_column = data[['Team']]

# Select the teams that start with the letter 'G'
teams_starting_with_G = teams_column[teams_column['Team'].str.startswith('G')]

# Print the selected teams
print(teams_starting_with_G)
```

```
      Team
5  Germany
6   Greece
```

## 8. Select the first 7 columns.

### 8. Select the first 7 columns.

```
# Select the first 7 columns
first_7_columns = data.iloc[:, :7]

# Print the selected columns
print(first_7_columns)
```

	Team	Goals	Shots on target	Shots off target	\
0	Croatia	4	13	12	
1	Czech Republic	4	13	18	
2	Denmark	4	10	10	
3	England	5	11	18	
4	France	3	22	24	
5	Germany	10	32	32	
6	Greece	5	8	18	
7	Italy	6	34	45	
8	Netherlands	2	12	36	
9	Poland	2	15	23	
10	Portugal	6	22	42	
11	Republic of Ireland	1	7	12	
12	Russia	5	9	31	
13	Spain	12	42	33	
14	Sweden	5	17	19	
15	Ukraine	2	7	26	

	Shooting Accuracy %	Goals-to-shots	Total shots (inc. Blocked)
0	51.9%	16.0%	32
1	41.9%	12.9%	39
2	50.0%	20.0%	27
3	50.0%	17.2%	40
4	37.9%	6.5%	65
5	47.8%	15.6%	80
6	30.7%	19.2%	32
7	43.0%	7.5%	110
8	25.0%	4.1%	60
9	39.4%	5.2%	48
10	34.3%	9.3%	82
11	36.8%	5.2%	28
12	22.5%	12.5%	59
13	55.9%	16.0%	100
14	47.2%	13.8%	39
15	21.2%	6.0%	38

## 9. Select all columns except the last 3.

---

### 9. Select all columns except the last 3.

```
# Select all columns except the last 3
all_except_last_3_columns = data.iloc[:, :-3]

# Print the selected columns
print(all_except_last_3_columns)
```

	Team	Goals	Shots on target	Shots off target	\
0	Croatia	4	13	12	
1	Czech Republic	4	13	18	
2	Denmark	4	10	10	
3	England	5	11	18	
4	France	3	22	24	
5	Germany	10	32	32	
6	Greece	5	8	18	
7	Italy	6	34	45	
8	Netherlands	2	12	36	
9	Poland	2	15	23	
10	Portugal	6	22	42	
11	Republic of Ireland	1	7	12	
12	Russia	5	9	31	
13	Spain	12	42	33	
14	Sweden	5	17	19	
15	Ukraine	2	7	26	

	Shooting Accuracy %	Goals-to-shots	Total shots (inc. Blocked)	\
--	---------------------	----------------	----------------------------	---

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

---

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

```
# Select the 'Team' and 'Shooting Accuracy' columns for England, Italy, and Russia
selected_teams = ['England', 'Italy', 'Russia']
shooting_accuracy_selected_teams = data[data['Team'].isin(selected_teams)][['Team', 'Shooting Accuracy']]

# Print the selected data
print(shooting_accuracy_selected_teams)
```

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

# Summary

---

- There are total 16 teams participated in euro 2012
- Italy leads in receiving the most yellow cards, with Portugal and Spain following in second and third positions, respectively.
- Greece, Poland, and the Republic of Ireland each hold the lead in red cards, with one red card per team.
- Mean Yellow Card shown per team is 7.
- The team scored more than 6 goals are Germany and Spain
- When shooting accuracy come into place between England, Italy and France we can say that England has higher shooting accuracy with 50% followed by Italy by 43% and Russia with 22.5%