Theory Activity No. 1

NAME: Shruti Siddheshwar Thorat

PRN:202401040053

DIV:CS2-48

- **Data Set:** Kaggle Text Classification
  - **Dataset-FIFA dataset**
    - Google colab link- https://colab.research.google.com/drive/1eYPDF0ImZ47qaF1FyxsbBEEVKfuE2p2I?usp=sharing

20 operation on data set:

```
import pandas as pd
import numpy as np
import io
from google.colab import files
uploaded = files.upload()
filename = list(uploaded.keys())[0]

# Load the dataset into a pandas DataFrame
df = pd.read_csv(io.BytesIO(uploaded[filename]))

# Display the first few rows of the dataset to verify it's loaded correctly
df.head()
```

Choose files fifa.csv
- **fifa.csv**(text/csv) - 19247 bytes, last modified: 01/05/2025 - 100% done

Saving fifa.csv to fifa.csv

	team1	team2	possession team1	possession team2	possession in contest	number of goals team1	number of goals team2	date	hour	category	...	penalties scored team1	penalties scored team2	goal preventions team1	goal preventions team2	own goals team1	own goals team2	forced turnovers team1	forced turnovers team2	defensive pressures applied team1	defensive pressures applied team2
0	QATAR	ECUADOR	42%	50%	8%	0	2	20-Nov-22	17:00	Group A	...	0	1	6	5	0	0	52	72	256	279
1	ENGLAND	IRAN	72%	19%	9%	6	2	21-Nov-22	14:00	Group B	...	0	1	8	13	0	0	63	72	139	416
2	SENEGAL	NETHERLANDS	44%	45%	11%	0	2	21-Nov-22	17:00	Group A	...	0	0	9	15	0	0	63	73	263	251
3	UNITED STATES	WALES	51%	39%	10%	1	1	21-Nov-22	20:00	Group B	...	0	1	7	7	0	0	81	72	242	292
4	ARGENTINA	SAUDI ARABIA	64%	24%	12%	1	2	22-Nov-22	11:00	Group C	...	1	0	4	14	0	0	65	80	163	361

5 rows × 88 columns

Connected to Python 3 Google Compute Engine backend

Which team scored the most total goals?

```
#PROBLEM1: Which team scored the most total goals?
goals = pd.concat([df[['team1', 'number of goals team1']].rename(columns={'team1':'team', 'number of goals team1':'goals'}),
                   df[['team2', 'number of goals team2']].rename(columns={'team2':'team', 'number of goals team2':'goals'})])
goals.groupby('team')['goals'].sum().sort_values(ascending=False).head(1)
```

team	goals
FRANCE	16

dtype: int64

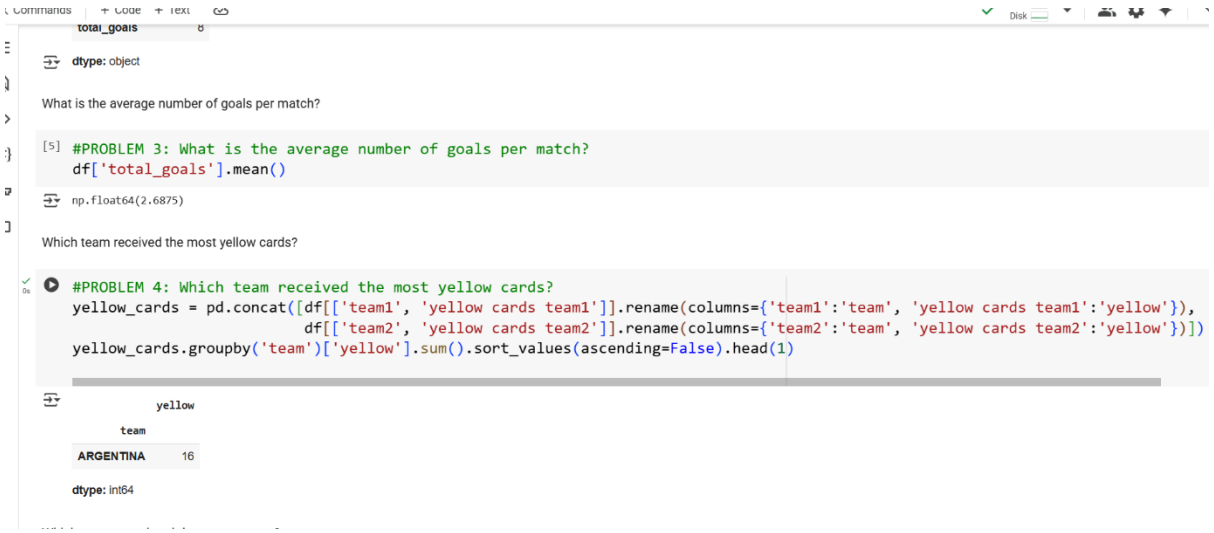
Which match had the highest total number of goals?

```
[4] #PROBLEM 2: Which match had the highest total number of goals?
df['total_goals'] = df['number of goals team1'] + df['number of goals team2']
df.loc[df['total_goals'].idxmax(), ['team1', 'team2', 'total_goals']]
```

	1
team1	ENGLAND
team2	IRAN
total_goals	8

dtype: object

Connected to Python 3 Google Compute Engine backend



total_goals 8

dtype: object

What is the average number of goals per match?

```
[5] #PROBLEM 3: What is the average number of goals per match?
    df['total_goals'].mean()
```

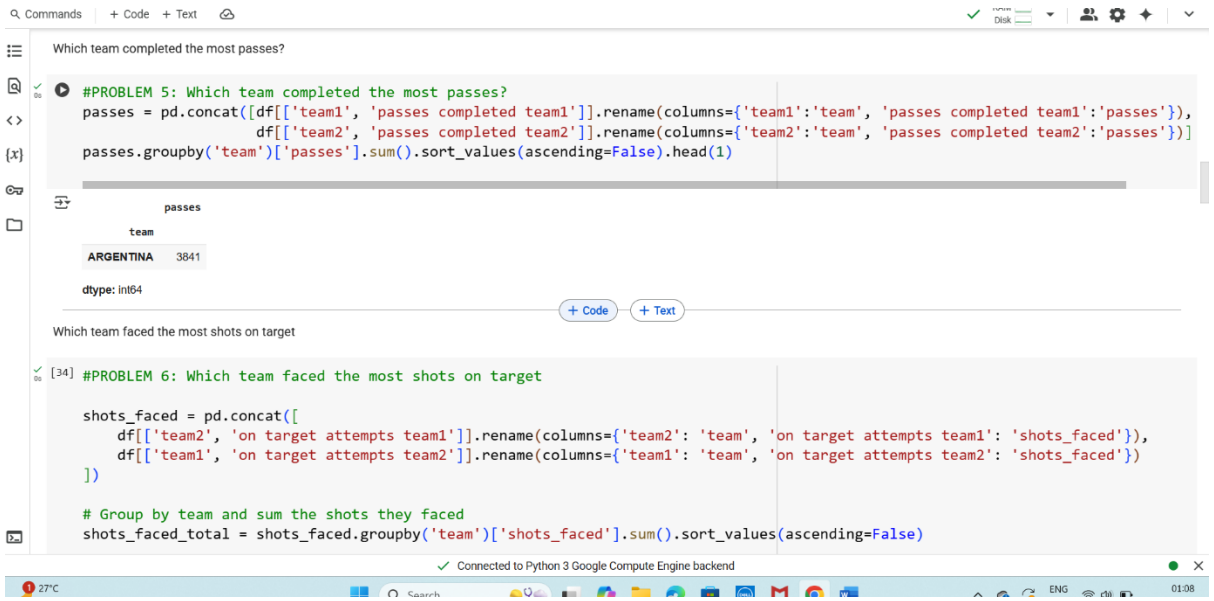
```
np.float64(2.6875)
```

Which team received the most yellow cards?

```
#PROBLEM 4: Which team received the most yellow cards?
yellow_cards = pd.concat([df[['team1', 'yellow cards team1']].rename(columns={'team1':'team', 'yellow cards team1':'yellow'}),
                          df[['team2', 'yellow cards team2']].rename(columns={'team2':'team', 'yellow cards team2':'yellow'})])
yellow_cards.groupby('team')['yellow'].sum().sort_values(ascending=False).head(1)
```

team	yellow
ARGENTINA	16

dtype: int64



Commands + Code + Text

Which team completed the most passes?

```
#PROBLEM 5: Which team completed the most passes?
passes = pd.concat([df[['team1', 'passes completed team1']].rename(columns={'team1':'team', 'passes completed team1':'passes'}),
                    df[['team2', 'passes completed team2']].rename(columns={'team2':'team', 'passes completed team2':'passes'})])
passes.groupby('team')['passes'].sum().sort_values(ascending=False).head(1)
```

team	passes
ARGENTINA	3841

dtype: int64

+ Code + Text

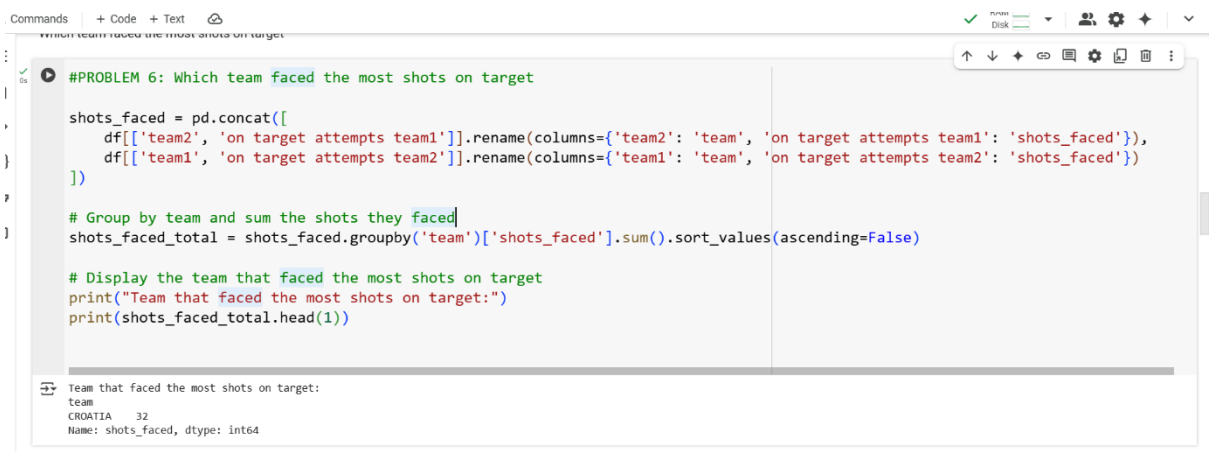
Which team faced the most shots on target

```
[34] #PROBLEM 6: Which team faced the most shots on target

shots_faced = pd.concat([
    df[['team2', 'on target attempts team1']].rename(columns={'team2': 'team', 'on target attempts team1': 'shots_faced'}),
    df[['team1', 'on target attempts team2']].rename(columns={'team1': 'team', 'on target attempts team2': 'shots_faced'})
])

# Group by team and sum the shots they faced
shots_faced_total = shots_faced.groupby('team')['shots_faced'].sum().sort_values(ascending=False)
```

Connected to Python 3 Google Compute Engine backend



Commands + Code + Text

```
#PROBLEM 6: Which team faced the most shots on target

shots_faced = pd.concat([
    df[['team2', 'on target attempts team1']].rename(columns={'team2': 'team', 'on target attempts team1': 'shots_faced'}),
    df[['team1', 'on target attempts team2']].rename(columns={'team1': 'team', 'on target attempts team2': 'shots_faced'})
])

# Group by team and sum the shots they faced
shots_faced_total = shots_faced.groupby('team')['shots_faced'].sum().sort_values(ascending=False)

# Display the team that faced the most shots on target
print("Team that faced the most shots on target:")
print(shots_faced_total.head(1))
```

```
Team that faced the most shots on target:
team
CROATIA    32
Name: shots_faced, dtype: int64
```

Which match had the highest number of total passes?

```
#PROBLEM 7: Which match had the highest number of total passes?
df['total_passes'] = df['passes team1'] + df['passes team2']
df.loc[df['total_passes'].idxmax(), ['team1', 'team2', 'total_passes']]
```

	56
team1	CROATIA
team2	BRAZIL
total_passes	1410

dtype: object

Find the average number of fouls per match.

```
[14] #PROBLEM 8: Find the average number of fouls per match.
(df['fouls against team1'] + df['fouls against team2']).mean()
```

```
np.float64(25.0)
```

Which teamcommited the most fouls in the tournaments?

```
#PROBLEM 9: Which team committed the most fouls in the tournament?
fouls = pd.concat([df[['team2', 'fouls against team1']].rename(columns={'team2':'team', 'fouls against team1':'fouls'}),
                   df[['team1', 'fouls against team2']].rename(columns={'team1':'team', 'fouls against team2':'fouls'})])
fouls.groupby('team')['fouls'].sum().sort_values(ascending=False).head(1)
```

team	fouls
ARGENTINA	118

dtype: int64

Which team had the highest number of draws in the tournament?

```
#PROBLEM 10: Which team had the highest number of draws in the tournament?
# Filter matches where the result is a draw
draws = df[df['number of goals team1'] == df['number of goals team2']]

# Count draws for each team
draw_teams = pd.concat([draws['team1'], draws['team2']])
most_draws = draw_teams.value_counts().head(1)

print("Team with most draws:")
print(most_draws)
```

```
Team with most draws:
CROATIA    4
Name: count, dtype: int64
```

Double-click (or enter) to edit

Average number of offsides per match.

```
#PROBLEM 11: Average number of offsides per match.
(df['offsides team1'] + df['offsides team2']).mean()
```

```
np.float64(3.9375)
```

Which match had the most defensive pressures applied (combined)?

```
[24] #PROBLEM 12: Which match had the most defensive pressures applied (combined)?
     df['total_def_pressure'] = df['defensive pressures applied team1'] + df['defensive pressures applied team2']
     df.loc[df['total_def_pressure'].idxmax(), ['team1', 'team2', 'total_def_pressure']]
```

	54
team1	MOROCCO
team2	SPAIN
total_def_pressure	790

dtype: object

Which team completed the most crosses?

```
#PROBLEM 13: Which team completed the most crosses?
crosses = pd.concat([df[['team1', 'crosses completed team1']].rename(columns={'team1':'team', 'crosses completed team1':'crosses'}),
                     df[['team2', 'crosses completed team2']].rename(columns={'team2':'team', 'crosses completed team2':'crosses'})])
crosses.groupby('team')['crosses'].sum().sort_values(ascending=False).head(1)
```

team	crosses
FRANCE	49

dtype: int64

Which match had the most red cards?

```
[26] #PROBLEM 14: Which match had the most red cards?
     df['total_reds'] = df['red cards team1'] + df['red cards team2']
     df.loc[df['total_reds'].idxmax(), ['team1', 'team2', 'total_reds']]
```

	16
team1	WALES
team2	IRAN
total_reds	1

dtype: object

What is the average number of shots (attempts) per match?

```
#PROBLEM 15: What is the average number of shots (attempts) per match?

(df['total attempts team1'] + df['total attempts team2']).mean()
```

```
np.float64(22.421875)
```

Which team had the highest average goals per game?

```
[28] #PROBLEM 16:Which team had the highest average goals per game?
     total_matches = pd.concat([df['team1'], df['team2']]).value_counts()
     total_goals = goals.groupby('team')['goals'].sum()
     (total_goals / total_matches).sort_values(ascending=False).head(1)
```

```
                0
ENGLAND       2.6

dtype: float64
```

Find matches where both teams scored at least 2 goals.

```
#PROBLEM 17: Find matches where both teams scored at least 2 goals.
df[(df['number of goals team1'] >= 2) & (df['number of goals team2'] >= 2)][['team1', 'team2', 'number of goals team1', 'number of goals team2'
```

	team1	team2	number of goals team1	number of goals team2
1	ENGLAND	IRAN	6	2
14	PORTUGAL	GHANA	3	2
28	CAMEROON	SERBIA	3	3
29	KOREA REPUBLIC	GHANA	2	3
43	COSTA RICA	GERMANY	2	4
46	SERBIA	SWITZERLAND	2	3
57	NETHERLANDS	ARGENTINA	2	2
63	ARGENTINA	FRANCE	3	3

Which team had the most successful defensive line breaks?

```
#PROBLEM 18: Which team had the most successful defensive line breaks?
def_breaks = pd.concat([df[['team1', 'completed defensive line breaksteam1']].rename(columns={'team1':'team', 'completed defensive line breaksteam1':'breaks'}),
                        df[['team2', 'completed defensive line breaks team2']].rename(columns={'team2':'team', 'completed defensive line breaks team2':'breaks'})])
def_breaks.groupby('team')['breaks'].sum().sort_values(ascending=False).head(1)
```

```
         breaks
team
ARGENTINA    95

dtype: int64
```

Which team made the most interceptions (forced turnovers)?

```
[31] #PROBLEM 19: Which team made the most interceptions (forced turnovers)?
     turnovers = pd.concat([df[['team1', 'forced turnovers team1']].rename(columns={'team1':'team', 'forced turnovers team1':'turnovers'}),
                            df[['team2', 'forced turnovers team2']].rename(columns={'team2':'team', 'forced turnovers team2':'turnovers'})])
     turnovers.groupby('team')['turnovers'].sum().sort_values(ascending=False).head(1)
```

```
          turnovers
team
CROATIA         548

dtype: int64
```

team	
CROATIA	548

dtype: int64

Which team had the most own goals?

```
[32] #PROBLEM 20: Which team had the most own goals?
     own_goals = pd.concat([df[['team1', 'own goals team1']].rename(columns={'team1':'team', 'own goals team1':'own'}),
                            df[['team2', 'own goals team2']].rename(columns={'team2':'team', 'own goals team2':'own'})])
     own_goals.groupby('team')['own'].sum().sort_values(ascending=False).head(1)
```

team	own
ARGENTINA	1

dtype: int64