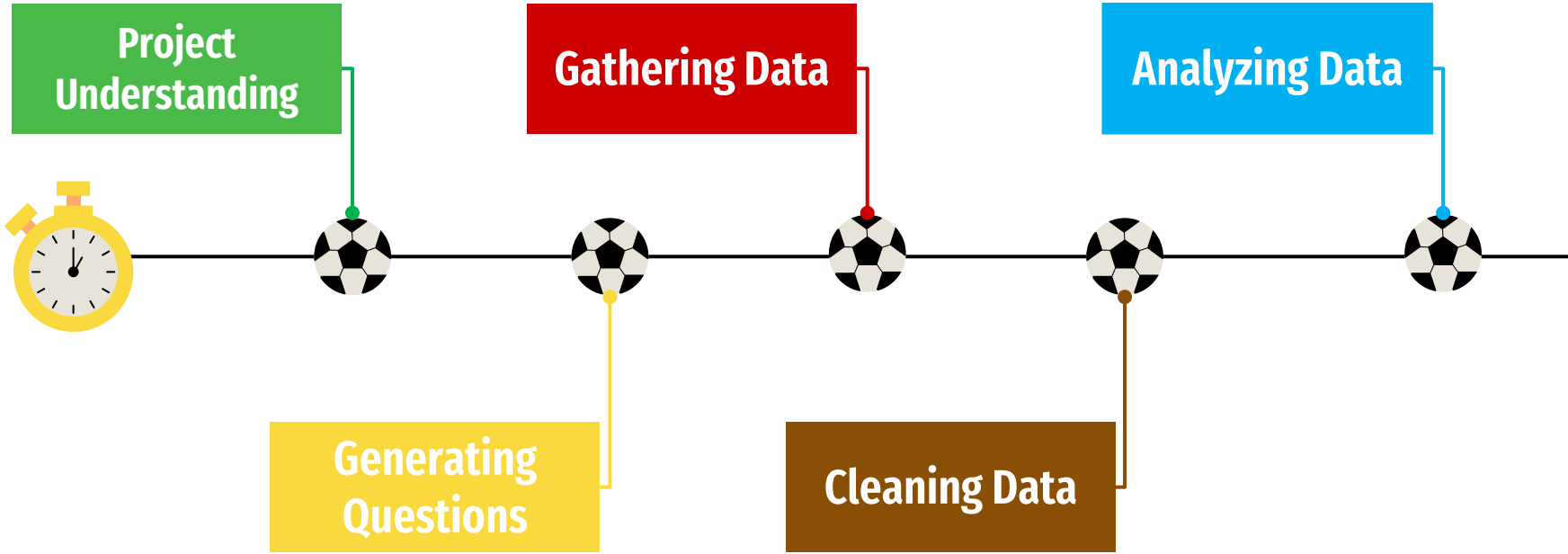
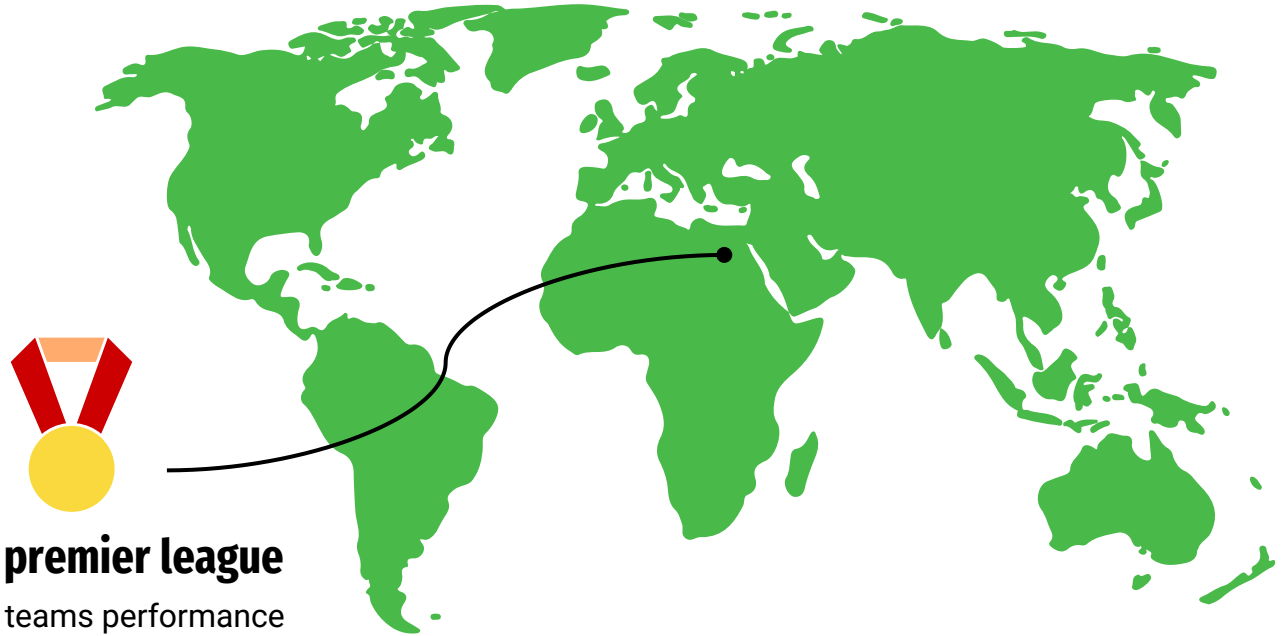
Egyptian Premier League Analysis

By: Mahmoud Ahmed Shimy

Table of content



Project Understanding



Egytian premier league

Analysing teams performance over seasons

Project Understanding

Goal

Match results	1	2	3	4
Al Ahly	✗	✓	✓	✓
Zamalek	✓	✓	✗	✓
El Ismaily	✓	✗	✓	✓
Al Ithad Al-Sakandary	✓	✓	✓	✗

Project Understanding

65 %

Stamina

75 %

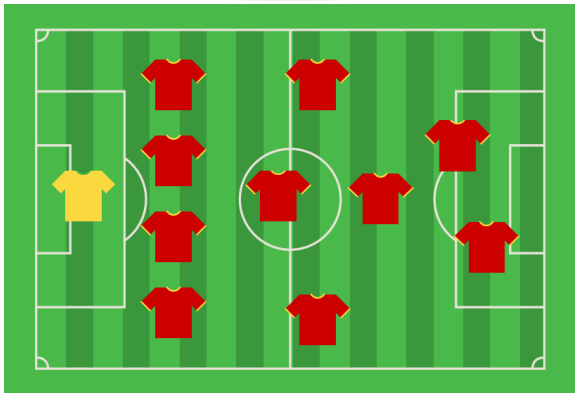
Sprint

30 %

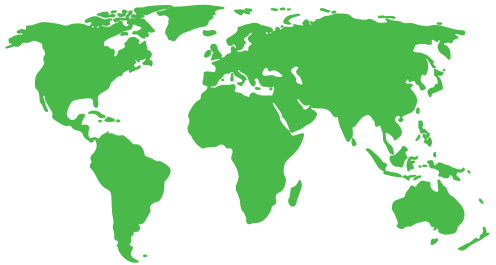
Defending

55 %

Attacking



Goal

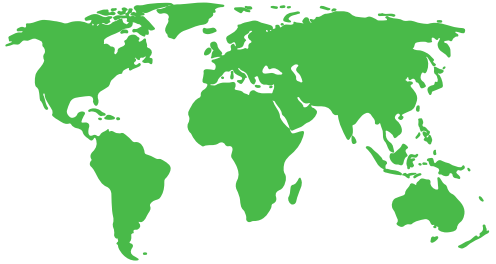


Match results	1	2	3	4
Al Ahly	✗	✓	✓	✓
Zamalek	✓	✓	✗	✓
El Ismaily	✓	✗	✓	✓
Al Ithad Al-Sakandary	✓	✓	✓	✗

Project Understanding

target audience

Goal



Match results	1	2	3	4
Al Ahly	✗	✓	✓	✓
Zamalek	✓	✓	✗	✓
El Ismaily	✓	✗	✓	✓
Al Ithad Al-Sakandary	✓	✓	✓	✗

Generating Questions

Season

For each season what is the different factors

Teams

For each team what is its performance over seasons

Win/Lose

What is the trends on winning and losing in the league

Qualification/ Relegation

Qualification and Relegation trends in the league

Gathering Data

AHMED IBRAHIM · UPDATED 8 MONTHS AGO

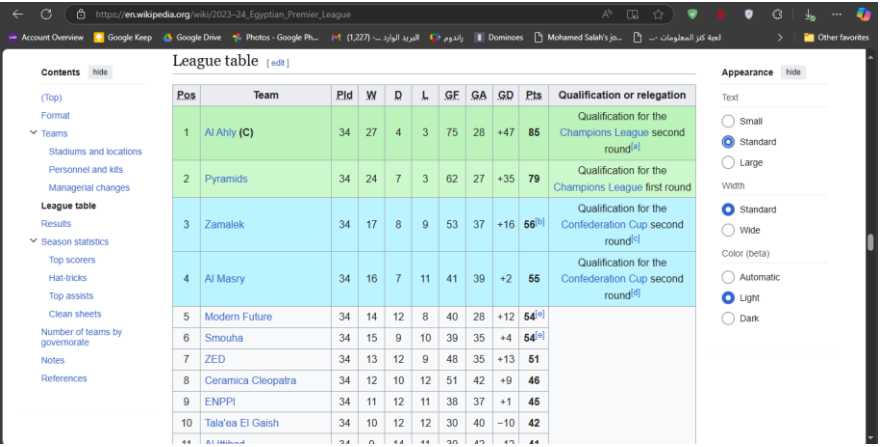
23 | New Notebook | Download

Egyptian Premier League from 2000 till 2022

Except seasons (2011-12,2012-13,2013-14)

EPL EGYPTIAN PREMIER LEAGUE

kaggle



Pos	Team	Pld	W	D	L	GF	GA	GD	Pts	Qualification or relegation
1	Al Ahly (C)	34	27	4	3	75	28	+47	85	Qualification for the Champions League second round[a]
2	Pyramids	34	24	7	3	62	27	+35	79	Qualification for the Champions League first round
3	Zamalek	34	17	8	9	53	37	+16	56[b]	Qualification for the Confederation Cup second round[c]
4	Al Masry	34	16	7	11	41	39	+2	55	Qualification for the Confederation Cup second round[d]
5	Modern Future	34	14	12	8	40	28	+12	54[e]	
6	Smouha	34	15	9	10	39	35	+4	54[e]	
7	ZED	34	13	12	9	48	35	+13	51	
8	Ceramica Cleopatra	34	12	10	12	51	42	+9	46	
9	ENPPI	34	11	12	11	38	37	+1	45	
10	Tala'ea El Gaish	34	10	12	12	30	40	−10	42	

Power BI

WIKIPEDIA
The Free Encyclopedia

Gathering Data

NTI Final Project.ipynb

File Edit View Insert Runtime Tools Help All changes saved

Share

+ Code + Text

RAM Disk

Gemini

Data Assessing

[69] df

	Season	Pos	Team	Pld	W	D	L	GF	GA	GD	Pts	Qualification or relegation
0	2000-01	1	Zamalek	26	20	5	1	54	18	36	65	Qualification for 2002 CAF Champions League
1	2000-01	2	Al Ahly	26	17	6	3	42	17	25	57	Qualification for 2002 CAF Champions League
2	2000-01	3	Al Masry	26	11	13	2	36	17	19	46	Qualification for 2002 CAF Cup
3	2000-01	4	Ismaily	26	11	10	5	32	19	13	43	Not Applicable
4	2000-01	5	El Qanah	26	11	9	6	26	18	8	42	Not Applicable
...	...	...	...	...	...	...	...	...	...	...	...	...
341	2023-24	14	El Ismaily	34	7	12	15	33	43	-10	33	NaN
342	2023-24	15	Pharco	34	6	15	13	32	43	-11	33	NaN
343	2023-24	16	Baladiyat El Mahalla	34	7	7	20	31	65	-34	28	Relegation to the Second Division
344	2023-24	17	Arab Contractors	34	5	11	18	32	57	-25	26	Relegation to the Second Division
345	2023-24	18	El Daklyeh	34	3	11	20	17	43	-26	20	Relegation to the Second Division

346 rows × 12 columns

Cleaning Data

Data Cleaning

In this stage we have 3 tasks:

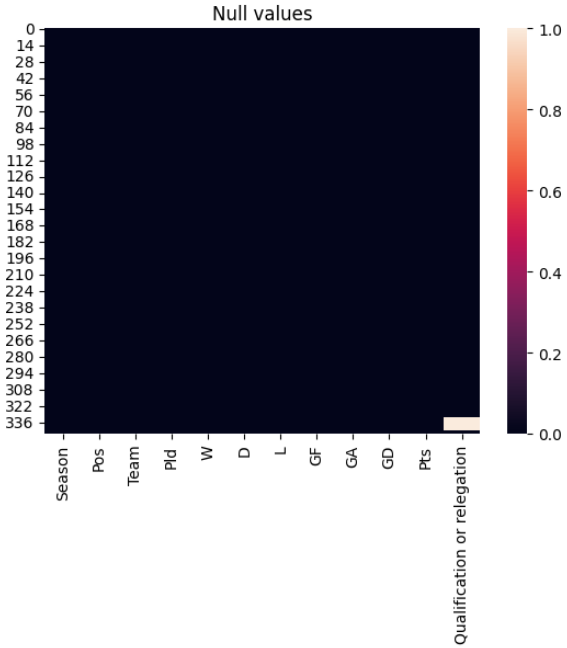
1. Rename columns
2. Handle Null values
3. Handle inconsistency in data values

```
# Renaming Column
df.rename(columns={'Qualification or relegation': 'Decision'}, inplace=True)
```

```
] df[df.Decision.isna()]
```

	Season	Pos	Team	Pld	W	D	L	GF	GA	GD	Pts	Decision
332	2023-24	5	Modern Sport	34	14	12	8	40	28	12	54	NaN
333	2023-24	6	Smouha	34	15	9	10	39	35	4	54	NaN
334	2023-24	7	ZED	34	13	12	9	48	35	13	51	NaN

```
sns.heatmap(df.isnull())
plt.title('Null values')
plt.show()
```



Cleaning Data

Data Cleaning

In this stage we have 3 tasks:

1. Rename columns
2. Handle Null values
3. Handle inconsistency in data values

```
df.Decision.value_counts()
```

Decision	count
Not Applicable	195
Relegation to the Second Division	58
Qualification for the Champions League	20
Qualification for the Confederation Cup	19
Remains in Premiere league	11
Qualification for CAF Champions League	4

```
qualcon = ['Qualification for CAF Confederation Cup','Qualification for the\xa0Confederation Cup',
           'Qualification for the\xa0Confederation Cup\xa0first or second round','Qualification to\xa0Confederation Cup',
           "Qualification for the\xa0CAF Cup", "Qualification to Confederation Cup",
           "Qualification for 2002 African Cup Winners' Cup", "Qualification for 2002 CAF Cup",
           "Qualification for CAF Confederation cup", "Qualification for the CAF Cup",
           "Qualification for the Confederation Cup first or second round"]
qualchamp = ['Qualification for the\xa0Champions League\xa0first or second round','Qualification to\xa0Champions League',
             'Qualification for the\xa0Champions League',"Qualification for the Champions League", "Qualification for CAF Champions League",
             "Qualification for the Champions League first or second round"]
```

```
df.loc[df['Decision'].isin(qualcon), 'Decision'] = 'Qualified for CAF Confederation cup'
df.loc[df['Decision'].isin(qualchamp), 'Decision'] = 'Qualified for CAF Champions League'
```

Making "Season" column has only year

```
[17] df['Season'] = df['Season'].str[:-3].astype(int)
```

Analyzing Data

Q1: The range of the Analysis?

```
[21] df.Season.min(), df.Season.max()
```

```
(2000, 2023)
```

Q2: Average Played matches?

```
[22] df.Pld.mean()
```

```
31.36416184971098
```

Q3: average Wins?

```
df.W.mean()
```

```
10.777456647398845
```

Q4: Average scored goals?

```
[24] df.GF.mean()
```

```
36.43063583815029
```

Q5: Average Received goals?

```
[25] df.GA.mean()
```

```
36.323699421965316
```

Q6: Average Goals Difference?

```
[26] df.GD.mean()
```

```
0.13583815028901733
```

Q7: Average Points?

```
[27] df.Pts.mean()
```

```
42.08092485549133
```

Q8: Minimum points?

```
[28] df.Pts.min()
```

```
9
```

Q9: Maximum points

```
df.Pts.max()
```

```
89
```

Q10: Maximum Received & Scored goals?

```
[30] df.GA.max(), df.GF.max()
```

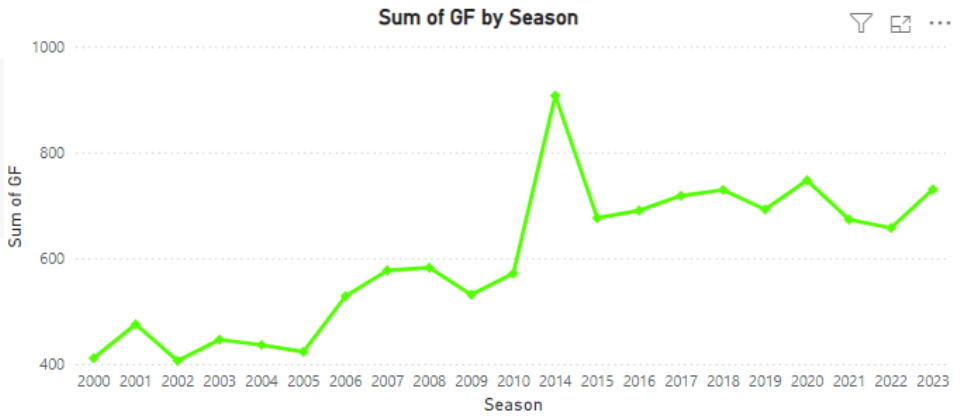
```
(88, 75)
```

Analyzing Data

Q11: Seasons with the most scored Goals?

```
print(df.groupby('Season')['GF'].sum().sort_values(ascending=False)[:5])
df.groupby('Season')['GF'].sum().plot(kind='line')
plt.title('Seasons with the most scored Goals')
plt.xlabel('Season')
plt.ylabel('Goals')
```

```
Season
2014    907
2020    747
2023    730
2018    729
2017    718
```

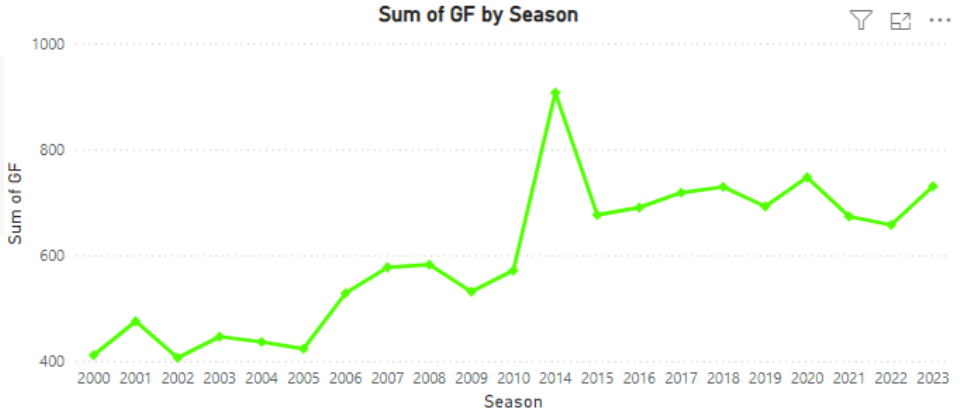


Analyzing Data

Q11: Seasons with the most scored Goals?

```
print(df.groupby('Season')['GF'].sum().sort_values(ascending=False)[:5])
df.groupby('Season')['GF'].sum().plot(kind='line')
plt.title('Seasons with the most scored Goals')
plt.xlabel('Season')
plt.ylabel('Goals')
```

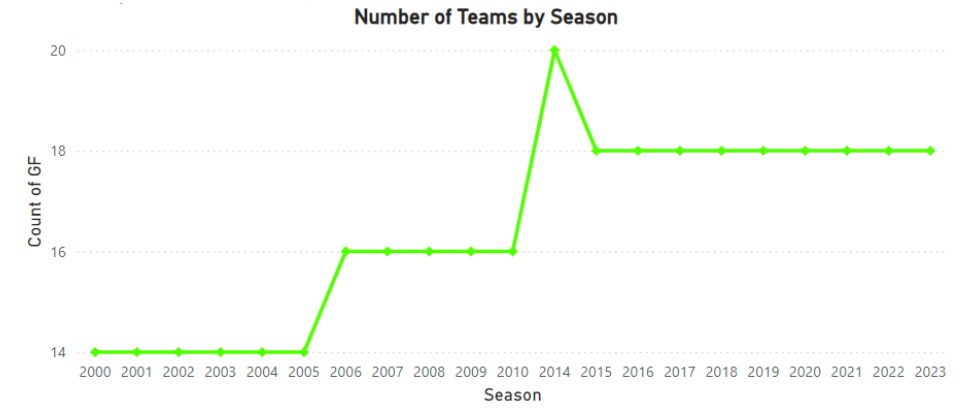
```
Season
2014    907
2020    747
2023    730
2018    729
2017    718
```



Q13: Number of matches in each Season?

```
print(df.groupby('Season').Season.count().sort_values(ascending=False)[:5])
df.groupby('Season').Season.count().plot(kind='line')
plt.title('Number of matches in each season')
plt.xlabel('Season')
plt.ylabel('Number of matches')
```

```
Season
2014    20
2023    18
2022    18
2021    18
2020    18
```

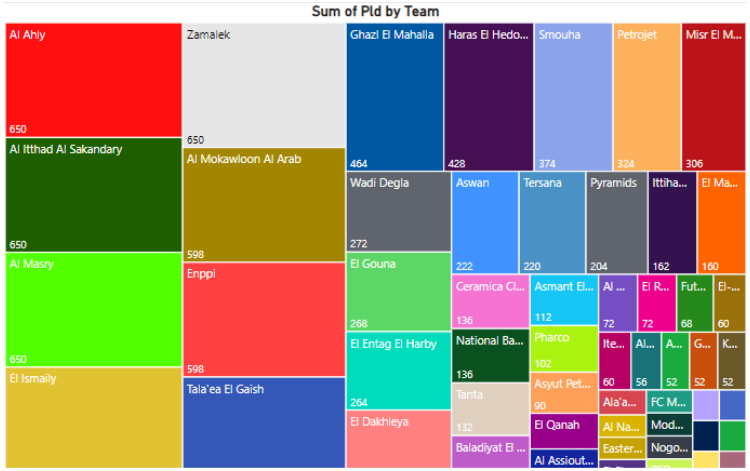


Analyzing Data

Q15: Top 5 teams with the most played matches?

```
print(df.groupby('Team')['Pld'].sum().sort_values(ascending=False)[:5])
sns.heatmap(data=df.groupby('Team')['Pld'].sum().sort_values(ascending=False)[:20].to_frame())
plt.title('Top teams with the most played matches')
plt.xlabel('Played Matches')
plt.ylabel('Team')
plt.show()
```

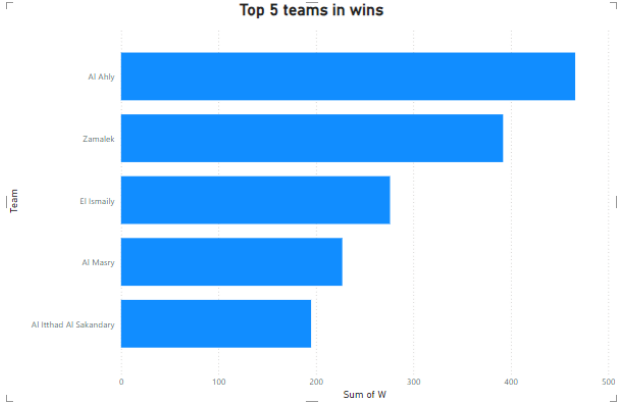
```
Team
Al Ahly                 650
Al Masry                650
El Ismaily              650
Zamalek                 650
Al Itthad Al Sakandary  650
```



Q16: Top 5 teams with the most wins?

```
print(df.groupby('Team')['W'].sum().sort_values(ascending=False)[:5])
df.groupby('Team')['W'].sum().sort_values(ascending=False)[:5].plot(kind='barh')
plt.title('Top 5 teams with the most wins')
plt.xlabel('Wins')
plt.ylabel('Team')
plt.show()
```

```
Team
Al Ahly       466
Zamalek       392
El Ismaily    276
Al Masry      227
Enppi         217
```

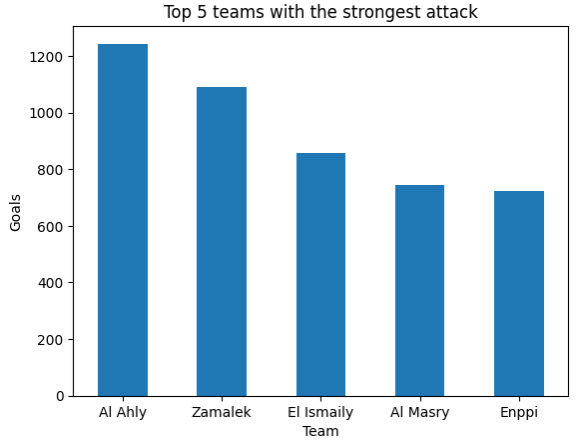


Analyzing Data

Q17: Top 5 teams with the strongest attack?

```
[ ] print(df.groupby('Team')['GF'].sum().sort_values(ascending=False)[:5])
    df.groupby('Team')['GF'].sum().sort_values(ascending=False)[:5].plot(kind='bar')
    plt.title('Top 5 teams with the strongest attack')
    plt.xlabel('Team')
    plt.ylabel('Goals')
    plt.xticks(rotation=0)
    plt.show()
```

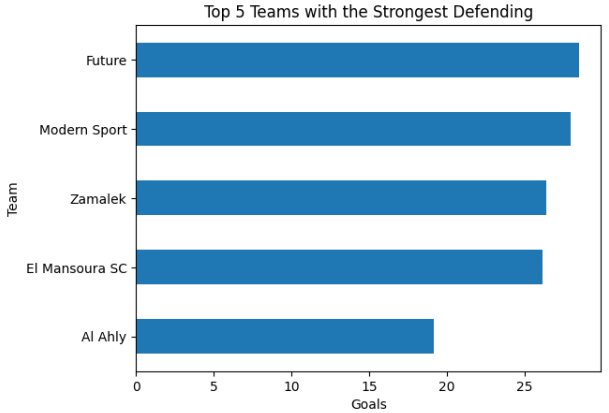
```
Team
Al Ahly       1243
Zamalek       1090
El Ismaily     858
Al Masry       743
Enppi          724
```



Q18: Top 5 Teams with the Strongest Defending?

```
[ ] print(df.groupby('Team')['GA'].mean().sort_values(ascending=True)[:5])
    df.groupby('Team')['GA'].mean().sort_values(ascending=True)[:5].plot(kind='barh')
    plt.title('Top 5 Teams with the Strongest Defending')
    plt.xlabel('Goals')
```

```
Team
Al Ahly           19.142857
El Mansoura SC    26.166667
Zamalek           26.380952
Modern Sport      28.000000
Future            28.500000
```

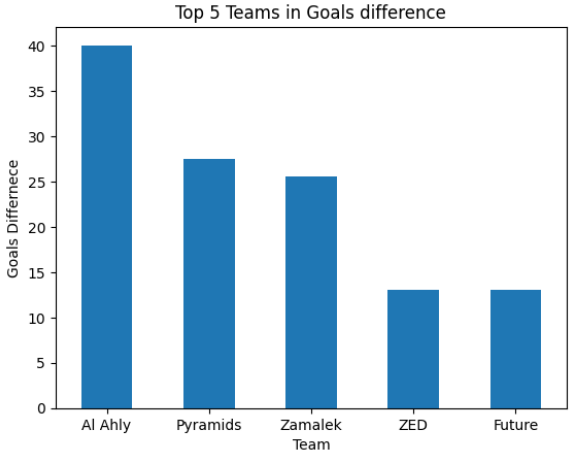


Analyzing Data

Q20: Top 5 Teams in Goals defference?

```
print(df.groupby('Team')['GD'].mean().sort_values(ascending=False)[:5])
df.groupby('Team')['GD'].mean().sort_values(ascending=False)[:5].plot(kind='bar')
plt.xticks(rotation=0)
plt.title('Top 5 Teams in Goals difference')
plt.xlabel('Team')
plt.ylabel('Goals Differnece')
```

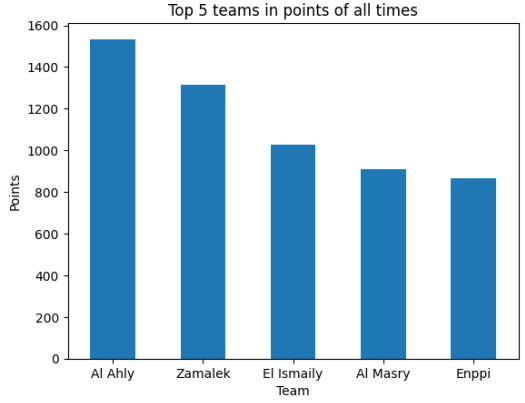
```
Team
Al Ahly     40.047619
Pyramids    27.500000
Zamalek     25.523810
ZED         13.000000
Future      13.000000
```



Q21: Top 5 teams in points of all times?

```
print(df.groupby('Team')['Pts'].sum().sort_values(ascending=False)[:5])
df.groupby('Team')['Pts'].sum().sort_values(ascending=False)[:5].plot(kind='bar')
plt.title('Top 5 teams in points of all times')
plt.xlabel('Team')
plt.ylabel('Points')
plt.xticks(rotation=0)
plt.show()
```

```
Team
Al Ahly       1534
Zamalek       1314
El Ismaily    1027
Al Masry       908
Enppi          865
```



Analyzing Data

Q22: Which Team get the maximum points in one season?

```
[ ] df.loc[df.Pts == df.Pts.max()]
```

	Season	Pos	Team	Pld	W	D	L	GF	GA	GD	Pts	Decision
256	2019	1	Al Ahly	34	28	5	1	74	8	66	89	Qualified for CAF Champions League

Q23: Which Team Scored the maximum Goals in one season?

```
[ ] df.loc[df.GF == df.GF.max()]
```

	Season	Pos	Team	Pld	W	D	L	GF	GA	GD	Pts	Decision
220	2017	1	Al Ahly	34	28	4	2	75	19	56	88	Qualified for CAF Champions League
328	2023	1	Al Ahly	34	27	4	3	75	28	47	85	Qualified for CAF Champions League

Analyzing Data

Q24: Top 5 Teams with win percentages?

```
[ ]  for team in df.Team.unique():
       df.loc[df.Team == team, 'Win%'] = df.loc[df.Team == team, 'W']/df.loc[df.Team == team, 'Pld']*100
```

```
[ ]  df.groupby('Team')['Win%'].mean().sort_values(ascending=False)[:5].plot(kind='bar')
     plt.title('Top 5 Teams with win percentages')
     plt.xlabel('Team')
     plt.ylabel('Win Percentage')
     plt.xticks(rotation=0)
     plt.show()
```

Future

46%

El Esmaily

43%

Al Ahly

72%

Zamalek

61%

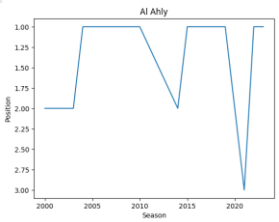
Pyramids

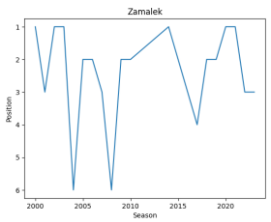
58%

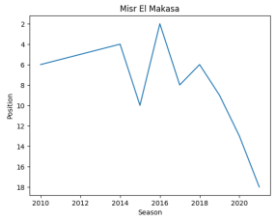
Analyzing Data

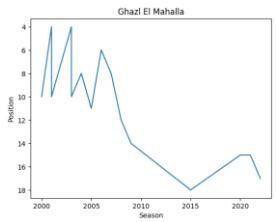
Q25: for each Team, how his position different by season?

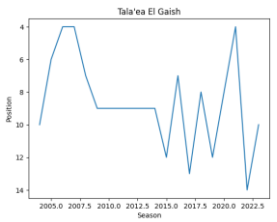
```
df.groupby(['Team', 'Season'])['Pos'].mean()
```

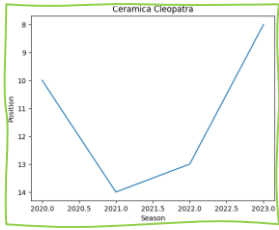


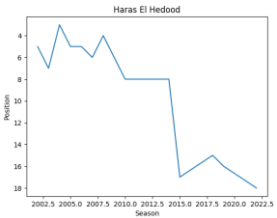


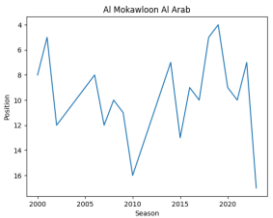


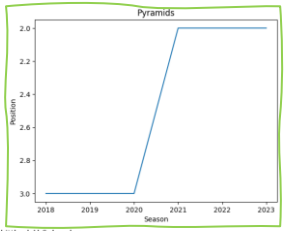


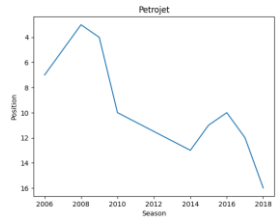


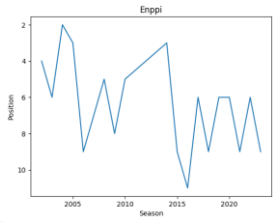


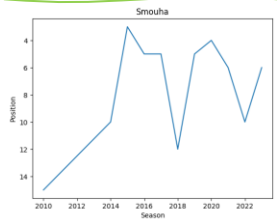


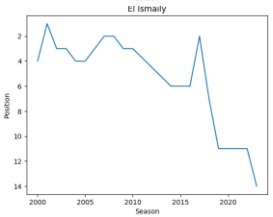


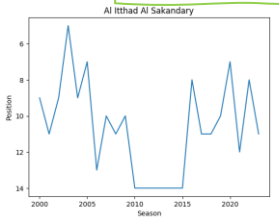


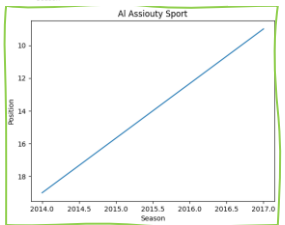










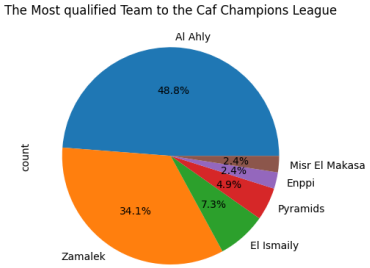


Analyzing Data

Q26: The most qualified Team to the Caf Champions League?

```
[ ] df.loc[df.Decision == 'Qualified for CAF Champions League'].Team.value_counts().plot(kind='pie', autopct = '%1.1f%%')
    plt.title('The Most qualified Team to the Caf Champions League')
    plt.show()
    df.loc[df.Decision == 'Qualified for CAF Champions League'].Team.value_counts()
```

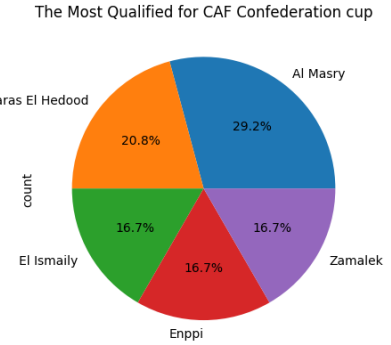
Team	
Al Ahly	20
Zamalek	14
El Ismaily	3
Pyramids	2
Enppi	1
Misr El Makasa	1



Q27: Top 5 most Qualified teams for CAF Confederation cup?

```
[ ] df.loc[df.Decision == 'Qualified for CAF Confederation cup'].Team.value_counts()[:5].plot(kind='pie', autopct = '%1.1f%%')
    plt.title('The Most Qualified for CAF Confederation cup')
    plt.show()
    df.loc[df.Decision == 'Qualified for CAF Confederation cup'].Team.value_counts()[:5]
```

Team	
Al Masry	7
Haras El Hedood	5
El Ismaily	4
Enppi	4
Zamalek	4



Analyzing Data

Q28: Top 5 teams remainig in the league?

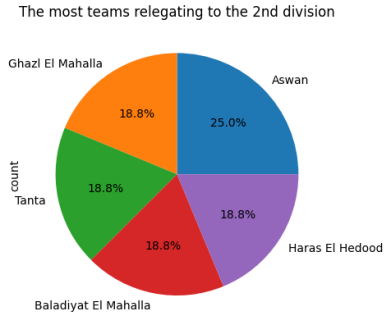
```
[ ]  df[df.Decision == 'Remains in Premiere league'].Team.value_counts().sort_values(ascending=False)[:5].plot(kind='pie', autopct = '%1.1f%%')
     plt.title('The most teams remainig in the league')
     plt.show()
     df[df.Decision == 'Remains in Premiere league'].Team.value_counts().sort_values(ascending=False)[:5]
```

Team	
Al Itthad Al Sakandary	20
Tala'ea El Gaish	17
Al Mokawloon Al Arab	16
El Ismaily	14
Enppi	14

Q29: Most teams relegating to the 2nd division?

```
[ ]  df[df.Decision == "Relegated to Second Division"].Team.value_counts().sort_values(ascending=False)[:5].plot(kind='pie', autopct = '%1.1f%%')
     plt.title('The most teams relegating to the 2nd division')
     plt.show()
     df[df.Decision == "Relegated to Second Division"].Team.value_counts().sort_values(ascending=False)[:5]
```

Team	
Aswan	4
Ghazl El Mahalla	3
Tanta	3
Baladiyat El Mahalla	3
Haras El Hedood	3



Conclusion

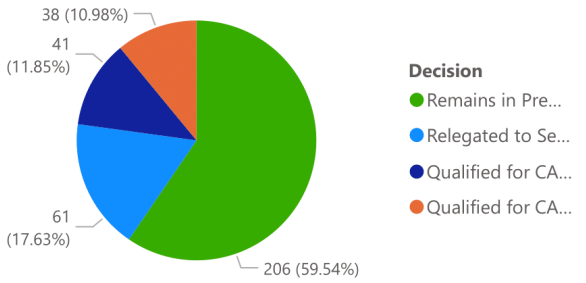
Conclusions

- Top Promising
  - Ceramica Cleopatra
  - Pyramids
- Top Safe
  - Al Itthad Al Sakandary
  - Tala'ea El Gaish
- Best Attack
  - Al Ahly
  - Zamalek
- Best Defending
  - Al Ahly
  - El Mansoura SC

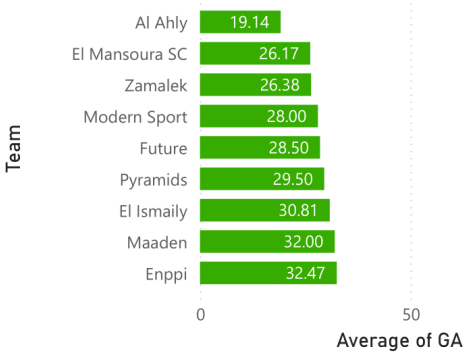
Egyptian Premier League Dashboard

By: Mahmoud Shimy

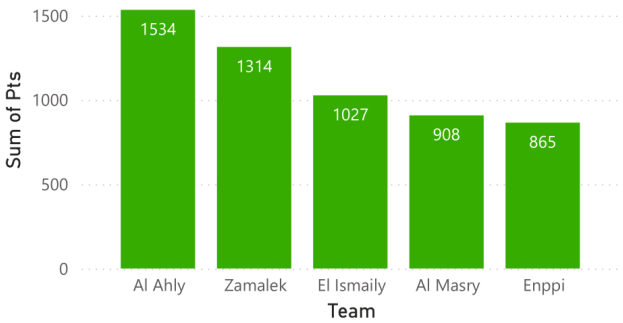
Count of Decision by Decision



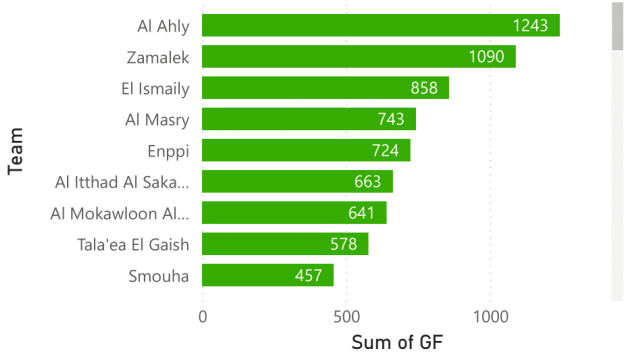
Average of GA by Team



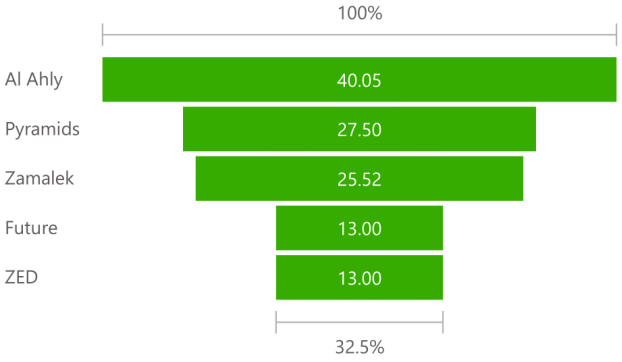
Sum of Pts by Team



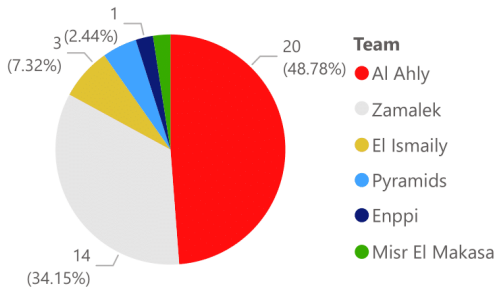
Sum of GF by Team



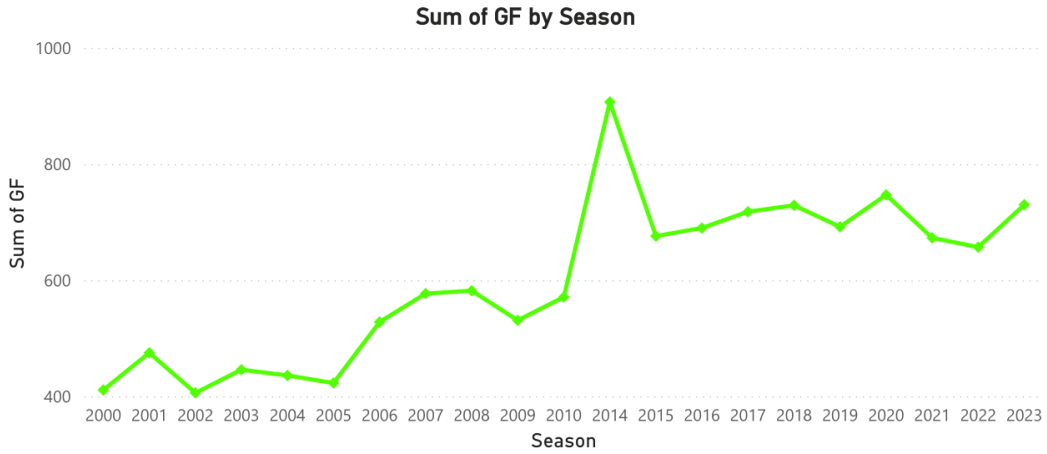
Average of GD by Team

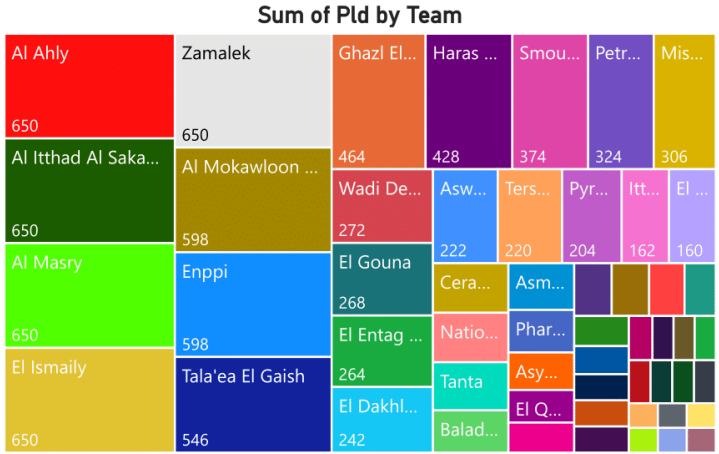


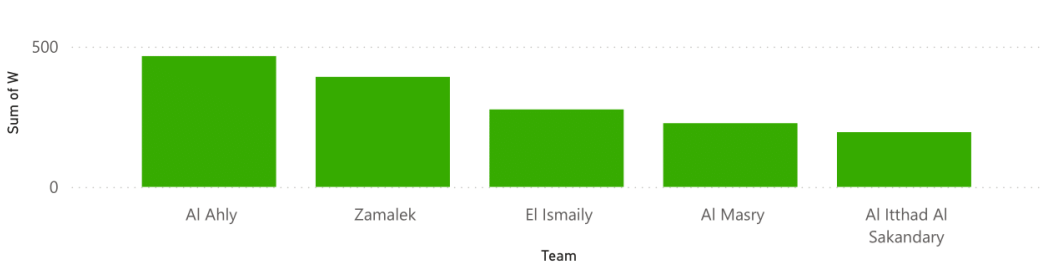
Teams in Champoins leage

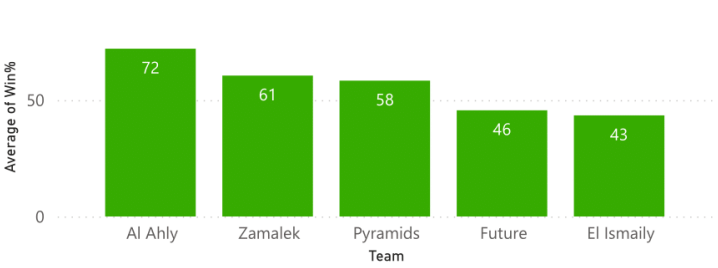


Average Scored Goals	Average Played Games	Average Received Goals	Average Points
36.43	31.36	36.32	42.08





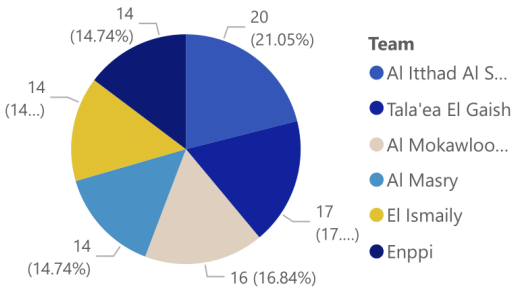




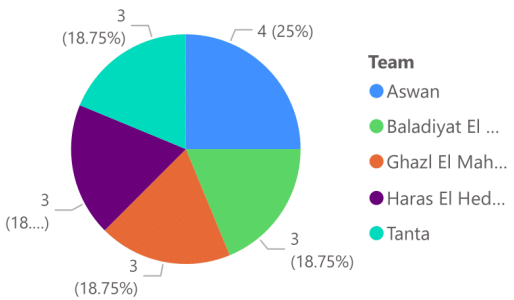
Egyptian Premier League Dashboard

By: Mahmoud Shimy

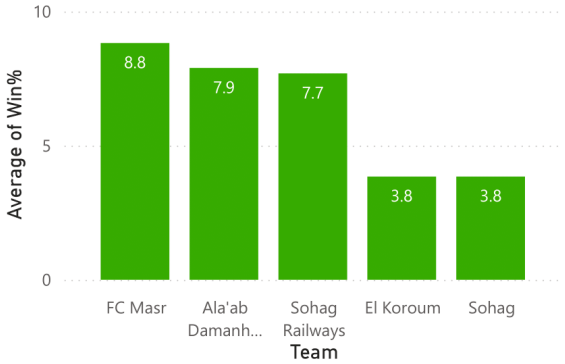
Teams remaining in the league



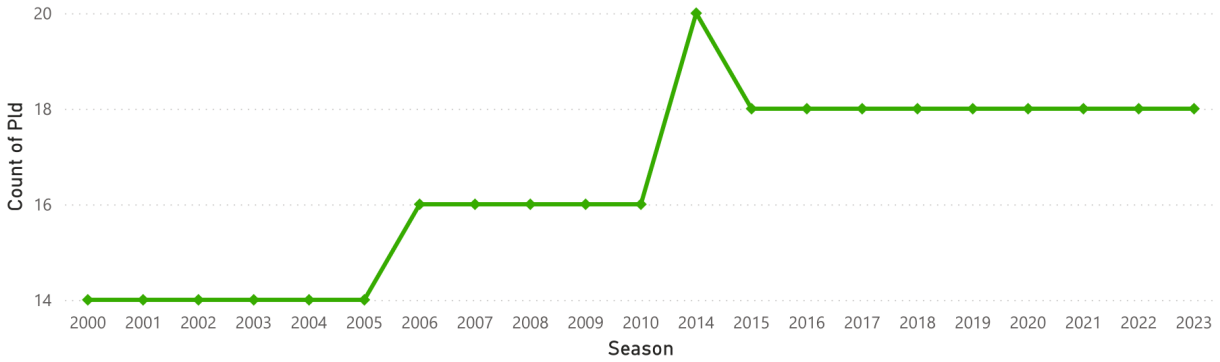
Top Teams Relegating down



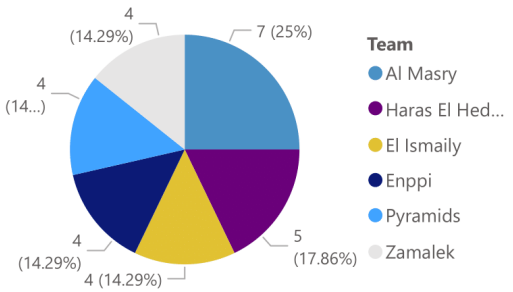
Minimun Teams in winning %



Count of Pld by Season



Teams in Confedration league



THANKS!

By: Mahmoud Shimy