Project Title - Analysis of UEFA Champions League Final (1956 - 2019)

By

Olalekan Rasaq

University of Ibadan, Ibadan Nigeria

2021

The dataset used for this project contains a table of Champions league winner and runner up from 1955/1956 season to 2018/2019 season. The table contain season, the club that won the champions league with the formation and their coach, the runner up, the country from which the winner and runner up located, and the most valuable player for the tournament.

We will be analysing the dataset to get some interesting information and answer some simple questions such as: -Who is the champions league winner for a particular season? -Which club has the highest record of winning the champions league? -Which coach has most champions league winning record? -Who is the most valuable player in champions league at a particular seasson? And many more...

Note: This project is part of requirement for the course "Data Analysis with Python: Zero to Pandas" offered by www.jovian.ai

Dowloading the dataset

The dataset we are going to use for this analysis will be downloaded form www.kaggle.com

In [1]:
```
!pip install jovian opendatasets --upgrade --quiet
```

In [2]:
```
import opendatasets as od
```

In [3]:
```
dataset_url = 'https://www.kaggle.com/egadharmawan/uefa-champion-league-final-all-season-19552019'
```

In [4]:
```
od.download(dataset_url)
```

```
Skipping, found downloaded files in "./uefa-champion-league-final-all-season-19552019" (use force=True to force download)
```

In [5]:
```
data_dir = './uefa-champion-league-final-all-season-19552019'
```

In [6]:
```
import os
os.listdir(data_dir)
```

Out[6]:
```
['UEFA Champion League All Season.csv']
```

Data Preparation and Cleaning

Let's load the dataset into panda dataframe and do some data cleaning.

In [7]:
```
import pandas as pd
```

In [8]:
```
uefa_df = pd.read_csv('./uefa-champion-league-final-all-season-19552019/UEFA Champion League All Season.csv')
uefa_df
```

Out[8]:

	club	nation	coach	formation	mvp	position	season
0	Real Madrid CF	Spain	José Villalonga	unknown	unknown	winner	'1955/1956'
1	Stade de Reims	France	Albert Batteux	unknown	unknown	runner up	'1955/1956'
2	Real Madrid CF	Spain	José Villalonga	unknown	unknown	winner	'1956/1957'
3	ACF Fiorentina	Italy	Fulvio Bernardini	unknown	unknown	runner up	'1956/1957'
4	Real Madrid CF	Spain	Luis Antonio Carniglia	unknown	unknown	winner	'1957/1958'
...	...	...	...	...	...	...	...
123	Juventus	Italy	Massimiliano Allegri	'3-4-3'	unknown	runner up	'2016/2017'
124	Real Madrid CF	Spain	Zinédine Zidane	'4-3-3'	Gareth Bale	winner	'2017/2018'
125	Liverpool FC	England	Jürgen Klopp	'4-3-3'	unknown	runner up	'2017/2018'
126	Liverpool FC	England	Jürgen Klopp	'4-3-3'	Virgil van Dijk	winner	'2018/2019'
127	Tottenham Hotspur	England	Jürgen Klopp	'4-4-2'	unknown	runner up	'2018/2019'

128 rows × 7 columns

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

Out[9]:
```
(128, 7)
```

In [10]:
```
uefa_df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 128 entries, 0 to 127
Data columns (total 7 columns):
 #   Column     Non-Null Count  Dtype
---  ------     --------------  -----
 0   club       128 non-null    object
 1   nation     128 non-null    object
 2   coach      128 non-null    object
 3   formation  128 non-null    object
 4   mvp        128 non-null    object
 5   position   128 non-null    object
 6   season     128 non-null    object
dtypes: object(7)
memory usage: 7.1+ KB
```

Let's sort our dataframe to show only the winners of the UEFA champions league from 1955/1956 to 2018/2019 season.

In [11]:
```
uefa_winner_df = uefa_df[uefa_df.position == 'winner']
uefa_winner_df
```

Out[11]:

	club	nation	coach	formation	mvp	position	season
0	Real Madrid CF	Spain	José Villalonga	unknown	unknown	winner	'1955/1956'
2	Real Madrid CF	Spain	José Villalonga	unknown	unknown	winner	'1956/1957'
4	Real Madrid CF	Spain	Luis Antonio Carniglia	unknown	unknown	winner	'1957/1958'
6	Real Madrid CF	Spain	Luis Antonio Carniglia	unknown	unknown	winner	'1958/1959'
8	Real Madrid CF	Spain	Miguel Muñoz Mozún	unknown	unknown	winner	'1959/1960'
...	...	...	...	...	...	...	...
118	Barcelona FC	Spain	Luis Enrique	'4-3-3'	Andrés Iniesta	winner	'2014/2015'
120	Real Madrid CF	Spain	Zinédine Zidane	'4-3-3'	Sergio Ramos	winner	'2015/2016'
122	Real Madrid CF	Spain	Zinédine Zidane	'4-3-3'	Cristiano Ronaldo	winner	'2016/2017'
124	Real Madrid CF	Spain	Zinédine Zidane	'4-3-3'	Gareth Bale	winner	'2017/2018'
126	Liverpool FC	England	Jürgen Klopp	'4-3-3'	Virgil van Dijk	winner	'2018/2019'

64 rows × 7 columns

From the output above, we can see that the index is not well sorted. Let us reset the index into default index

In [12]:
```
uefa_winner_df.reset_index(inplace=True, drop=True)
```

In [13]:
```
uefa_winner_df
```

Out[13]:

	club	nation	coach	formation	mvp	position	season
0	Real Madrid CF	Spain	José Villalonga	unknown	unknown	winner	'1955/1956'
1	Real Madrid CF	Spain	José Villalonga	unknown	unknown	winner	'1956/1957'
2	Real Madrid CF	Spain	Luis Antonio Carniglia	unknown	unknown	winner	'1957/1958'
3	Real Madrid CF	Spain	Luis Antonio Carniglia	unknown	unknown	winner	'1958/1959'
4	Real Madrid CF	Spain	Miguel Muñoz Mozún	unknown	unknown	winner	'1959/1960'
...	...	...	...	...	...	...	...
59	Barcelona FC	Spain	Luis Enrique	'4-3-3'	Andrés Iniesta	winner	'2014/2015'
60	Real Madrid CF	Spain	Zinédine Zidane	'4-3-3'	Sergio Ramos	winner	'2015/2016'
61	Real Madrid CF	Spain	Zinédine Zidane	'4-3-3'	Cristiano Ronaldo	winner	'2016/2017'
62	Real Madrid CF	Spain	Zinédine Zidane	'4-3-3'	Gareth Bale	winner	'2017/2018'
63	Liverpool FC	England	Jürgen Klopp	'4-3-3'	Virgil van Dijk	winner	'2018/2019'

64 rows × 7 columns

We want to rearrange the dataframe column into a more appropriate order

In [14]:
```
cols = list(uefa_winner_df.columns.values)
cols
```

Out[14]:
```
['club', 'nation', 'coach', 'formation', 'mvp', 'position', 'season']
```

In [15]:
```
uefa_winner_df = uefa_winner_df[['season', 'position', 'club', 'nation', 'coach', 'formation', 'mvp']]
uefa_winner_df
```

Out[15]:

	season	position	club	nation	coach	formation	mvp
0	'1955/1956'	winner	Real Madrid CF	Spain	José Villalonga	unknown	unknown
1	'1956/1957'	winner	Real Madrid CF	Spain	José Villalonga	unknown	unknown
2	'1957/1958'	winner	Real Madrid CF	Spain	Luis Antonio Carniglia	unknown	unknown
3	'1958/1959'	winner	Real Madrid CF	Spain	Luis Antonio Carniglia	unknown	unknown
4	'1959/1960'	winner	Real Madrid CF	Spain	Miguel Muñoz Mozún	unknown	unknown
...	...	...	...	...	...	...	...
59	'2014/2015'	winner	Barcelona FC	Spain	Luis Enrique	'4-3-3'	Andrés Iniesta
60	'2015/2016'	winner	Real Madrid CF	Spain	Zinédine Zidane	'4-3-3'	Sergio Ramos
61	'2016/2017'	winner	Real Madrid CF	Spain	Zinédine Zidane	'4-3-3'	Cristiano Ronaldo
62	'2017/2018'	winner	Real Madrid CF	Spain	Zinédine Zidane	'4-3-3'	Gareth Bale
63	'2018/2019'	winner	Liverpool FC	England	Jürgen Klopp	'4-3-3'	Virgil van Dijk

64 rows × 7 columns

Let us display the whole table of the champions league winner to see the full list

In [16]:
```
from IPython.display import display
with pd.option_context('display.max_rows', 100):
    display(uefa_winner_df)
```

	season	position	club	nation	coach	formation	mvp
0	'1955/1956'	winner	Real Madrid CF	Spain	José Villalonga	unknown	unknown
1	'1956/1957'	winner	Real Madrid CF	Spain	José Villalonga	unknown	unknown
2	'1957/1958'	winner	Real Madrid CF	Spain	Luis Antonio Carniglia	unknown	unknown
3	'1958/1959'	winner	Real Madrid CF	Spain	Luis Antonio Carniglia	unknown	unknown
4	'1959/1960'	winner	Real Madrid CF	Spain	Miguel Muñoz Mozún	unknown	unknown
5	'1960/1961'	winner	SL Benfica	Portugal	Béla Guttmann	unknown	unknown
6	'1961/1962'	winner	SL Benfica	Portugal	Béla Guttmann	unknown	unknown
7	'1962/1963'	winner	AC Milan	Italy	Nereo Rocco	unknown	unknown
8	'1963/1964'	winner	FC Internazionale	Italy	Helenio Herrera	unknown	unknown
9	'1964/1965'	winner	FC Internazionale	Italy	Helenio Herrera	unknown	unknown
10	'1965/1966'	winner	Real Madrid CF	Spain	Miguel Muñoz Mozún	unknown	unknown
11	'1966/1967'	winner	Celtic FC	Scotland	Jock Stein	unknown	unknown
12	'1967/1968'	winner	Manchester United FC	England	Sir Matt Busby	unknown	unknown
13	'1968/1969'	winner	AC Milan	Italy	Nereo Rocco	unknown	unknown
14	'1969/1970'	winner	Feyenoord	Netherlands	Ernst Happel	unknown	unknown
15	'1970/1971'	winner	AFC Ajax	Netherlands	Rinus Michels	unknown	unknown
16	'1971/1972'	winner	AFC Ajax	Netherlands	Stefan Kovacs	unknown	unknown
17	'1972/1973'	winner	AFC Ajax	Netherlands	Stefan Kovacs	unknown	unknown
18	'1973/1974'	winner	FC Bayern Munchen	Germany	Udo Lattek	unknown	unknown
19	'1974/1975'	winner	FC Bayern Munchen	Germany	Dettmar Cramer	unknown	unknown
20	'1975/1976'	winner	FC Bayern Munchen	Germany	Dettmar Cramer	unknown	unknown
21	'1976/1977'	winner	Liverpool FC	England	Robert Paisley	unknown	unknown
22	'1977/1978'	winner	Liverpool FC	England	Robert Paisley	unknown	unknown
23	'1978/1979'	winner	Nottingham Forest FC	England	Brian Clough	unknown	unknown
24	'1979/1980'	winner	Nottingham Forest FC	England	Brian Clough	unknown	unknown
25	'1980/1981'	winner	Liverpool FC	England	Robert Paisley	unknown	unknown
26	'1981/1982'	winner	Aston Villa FC	England	Anthony Barton	unknown	unknown
27	'1982/1983'	winner	Hamburg SV	Germany	Ernst Happel	unknown	unknown
28	'1983/1984'	winner	Liverpool FC	England	Joseph Fagan	unknown	unknown
29	'1984/1985'	winner	Juventus	Italy	Giovanni Trapattoni	unknown	unknown
30	'1985/1986'	winner	FCSB	Romania	Emeric Ienei	unknown	unknown
31	'1986/1987'	winner	FC Porto	Portugal	Artur Jorge Melo Teixeira	unknown	unknown
32	'1987/1988'	winner	PSV Eindhoven	Netherlands	Guus Hiddink	unknown	unknown
33	'1988/1989'	winner	AC Milan	Italy	Arrigo Sacchi	unknown	unknown
34	'1989/1990'	winner	AC Milan	Italy	Arrigo Sacchi	unknown	unknown
35	'1990/1991'	winner	FC crvena zvezda	Serbia	Ljubo Petrovic	unknown	unknown
36	'1991/1992'	winner	Barcelona FC	Spain	Johan Cruyff	unknown	unknown
37	'1992/1993'	winner	Olympique de marseille fc	France	Raymond Goethals	unknown	unknown
38	'1993/1994'	winner	AC Milan	Italy	Fabio Capello	'4-4-2'	unknown
39	'1994/1995'	winner	AFC Ajax	Netherlands	Louis van Gaal	unknown	unknown
40	'1995/1996'	winner	Juventus	Italy	Marcello Lippi	unknown	unknown
41	'1996/1997'	winner	Borussia Dortmund	Germany	Ottmar Hitzfeld	'4-4-2'	unknown
42	'1997/1998'	winner	Real Madrid CF	Spain	Jupp Heynckes	unknown	unknown
43	'1998/1999'	winner	Manchester United FC	England	Sir Alex Ferguson	'4-4-2'	David Beckham
44	'1999/2000'	winner	Real Madrid CF	Spain	Vicente del Bosque	unknown	Raúl
45	'2000/2001'	winner	FC Bayern Munchen	Germany	Ottmar Hitzfeld	unknown	Oliver Kahn
46	'2001/2002'	winner	Real Madrid CF	Spain	Vicente del Bosque	'4-4-2'	Zinedine Zidane
47	'2002/2003'	winner	AC Milan	Italy	Carlo Ancelotti	'4-4-2'	Paolo Maldini
48	'2003/2004'	winner	FC Porto	Portugal	José Mourinho	'4-3-1-2'	Deco
49	'2004/2005'	winner	Liverpool FC	England	Rafael Benítez	'5-3-2'	Steven Gerrard
50	'2005/2006'	winner	Barcelona FC	Spain	Frank Rijkaard	'4-3-3'	Samuel Eto'o
51	'2006/2007'	winner	AC Milan	Italy	Carlo Ancelotti	'4-4-2'	Filippo Inzaghi
52	'2007/2008'	winner	Manchester United FC	England	Sir Alex Ferguson	'4-4-2'	Edwin van der Sar
53	'2008/2009'	winner	Barcelona FC	Spain	Josep Guardiola	'4-3-2-1'	Xavi
54	'2009/2010'	winner	FC Internazionale	Italy	José Mourinho	'4-3-3'	Diego Milito
55	'2010/2011'	winner	Barcelona FC	Spain	Josep Guardiola	'4-3-3'	Lionel Messi
56	'2011/2012'	winner	Chelsea FC	England	Roberto Di Matteo	'4-4-2'	Didier Drogba
57	'2012/2013'	winner	FC Bayern Munchen	Germany	Jupp Heynckes	'4-4-2'	Arjen Robben
58	'2013/2014'	winner	Real Madrid CF	Spain	Carlo Ancelotti	'4-3-3'	Ángel di María
59	'2014/2015'	winner	Barcelona FC	Spain	Luis Enrique	'4-3-3'	Andrés Iniesta
60	'2015/2016'	winner	Real Madrid CF	Spain	Zinédine Zidane	'4-3-3'	Sergio Ramos
61	'2016/2017'	winner	Real Madrid CF	Spain	Zinédine Zidane	'4-3-3'	Cristiano Ronaldo
62	'2017/2018'	winner	Real Madrid CF	Spain	Zinédine Zidane	'4-3-3'	Gareth Bale
63	'2018/2019'	winner	Liverpool FC	England	Jürgen Klopp	'4-3-3'	Virgil van Dijk

We will create another dataframe with the list of champions league runner up.

In [17]:
```
uefa_runnerup_df = uefa_df[uefa_df.position == 'runner up']
uefa_runnerup_df.reset_index(inplace=True, drop=True)
uefa_runnerup_df = uefa_runnerup_df[['season', 'position', 'club', 'nation', 'coach', 'formation', 'mvp']]
```

In [18]:
```
uefa_runnerup_df
```

Out[18]:

	season	position	club	nation	coach	formation	mvp
0	'1955/1956'	runner up	Stade de Reims	France	Albert Batteux	unknown	unknown
1	'1956/1957'	runner up	ACF Fiorentina	Italy	Fulvio Bernardini	unknown	unknown
2	'1957/1958'	runner up	AC Milan	Italy	Giuseppe Viani	unknown	unknown
3	'1958/1959'	runner up	Stade de Reims	France	Albert Batteux	unknown	unknown
4	'1959/1960'	runner up	Eintracht Frankfurt	Germany	Paul Osswald	unknown	unknown
...	...	...	...	...	...	...	...
59	'2014/2015'	runner up	Juventus	Italy	Massimiliano Allegri	'4-4-2'	unknown
60	'2015/2016'	runner up	Club Atlético de madrid	Spain	Diego Simeone	'4-4-2'	unknown
61	'2016/2017'	runner up	Juventus	Italy	Massimiliano Allegri	'3-4-3'	unknown
62	'2017/2018'	runner up	Liverpool FC	England	Jürgen Klopp	'4-3-3'	unknown
63	'2018/2019'	runner up	Tottenham Hotspur	England	Jürgen Klopp	'4-4-2'	unknown

64 rows × 7 columns

Exploratory Analysis and Visualization

Now, we will do some exploratory analysis and visualization of our data to show some relationship

In [19]:
```
# Let's explore the number of league won by the clubs
club_win = uefa_winner_df['club'].value_counts()
club_win
```

Out[19]:
```
Real Madrid CF               13
AC Milan                      7
Liverpool FC                  6
FC Bayern Munchen             5
Barcelona FC                  5
AFC Ajax                      4
Manchester United FC          3
FC Internazionale             3
Nottingham Forest FC          2
Juventus                      2
FC Porto                      2
SL Benfica                    2
Olympique de marseille fc     1
Chelsea FC                    1
FCSB                          1
Celtic FC                     1
FC crvena zvezda              1
Aston Villa FC                1
Hamburg SV                    1
PSV Eindhoven                 1
Feyenoord                     1
Borussia Dortmund             1
Name: club, dtype: int64
```

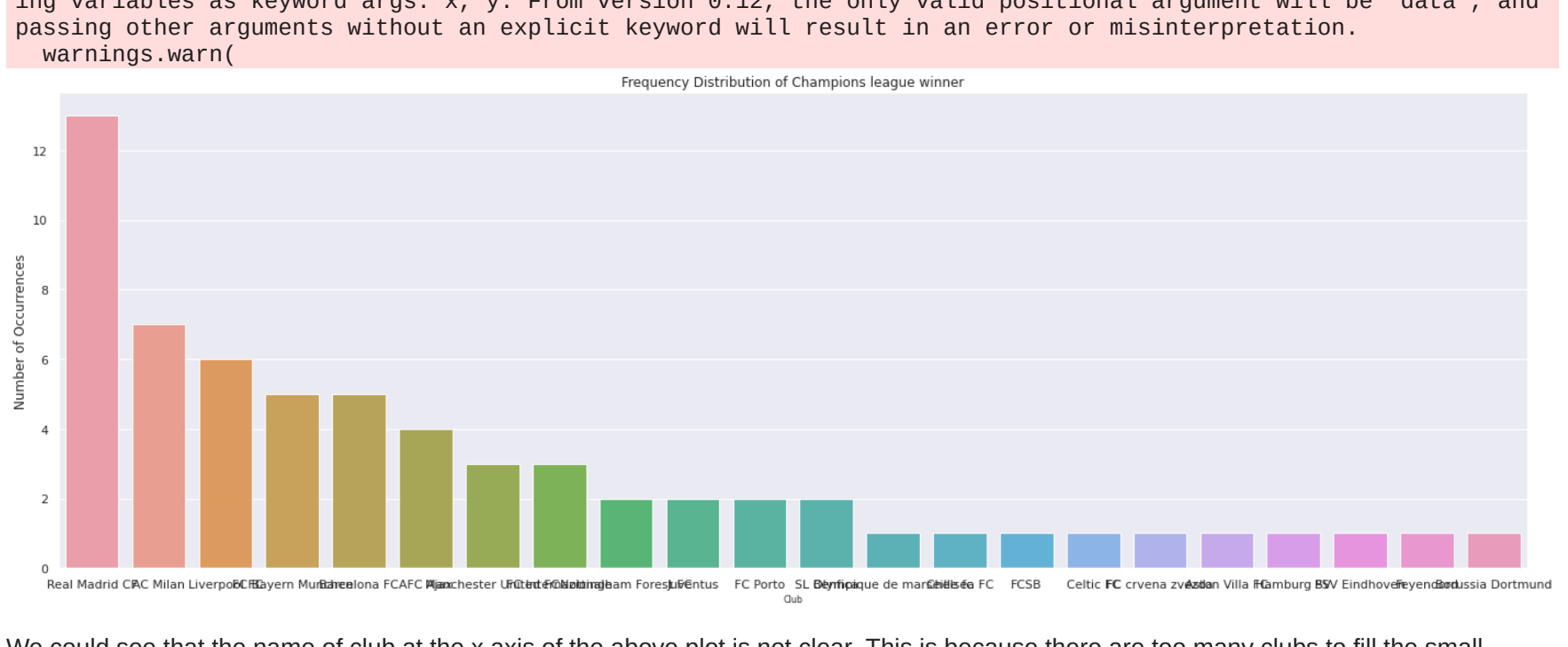
In [20]:
```
import seaborn as sns
import matplotlib
import matplotlib.pyplot as plt
%matplotlib inline

sns.set_style('darkgrid')
matplotlib.rcParams['font.size'] = 14
matplotlib.rcParams['figure.figsize'] = (24, 8)
matplotlib.rcParams['figure.facecolor'] = '#00000000'
```

Visual exploration is the most effective way to extract information between variables.

Below is a barplot of the frequency distribution of a champions league winner, which shows the frequency distribution of the club.

In [21]:
```
sns.set(style="darkgrid")
sns.barplot(club_win.index, club_win.values, alpha=0.9)
plt.title('Frequency Distribution of Champions league winner')
plt.ylabel('Number of Occurrences', fontsize=12)
plt.xlabel('Club', fontsize=8)
plt.show()
```

```
/home/akinkunmi/anaconda3/lib/python3.8/site-packages/seaborn/_decorators.py:36: FutureWarning: Pass the following variables as keyword args: x, y. From version 0.12, the only valid positional argument will be `data`, and passing other arguments without an explicit keyword will result in an error or misinterpretation.
  warnings.warn(
```



We could see that the name of club at the x axis of the above plot is not clear. This is because there are too many clubs to fill the small space at x axis.

To resolve this, we will only consider the frequency distribution plot of clubs with more than one champions league.

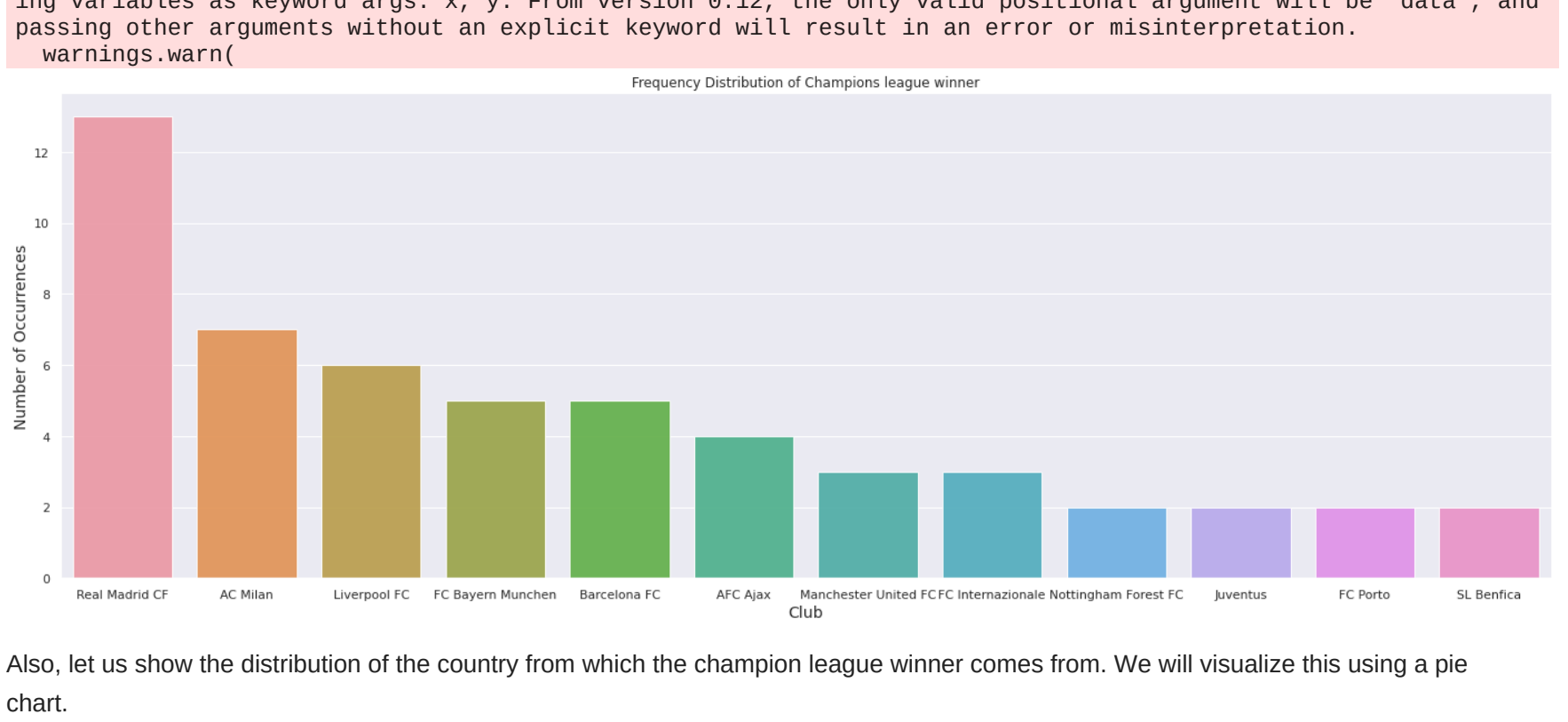
In [22]:
```
club_win = pd.DataFrame(club_win)
top_club = club_win[club_win.club > 1]
top_club
```

Out[22]:

	club
Real Madrid CF	13
AC Milan	7
Liverpool FC	6
FC Bayern Munchen	5
Barcelona FC	5
AFC Ajax	4
Manchester United FC	3
FC Internazionale	3
Nottingham Forest FC	2
Juventus	2
FC Porto	2
SL Benfica	2

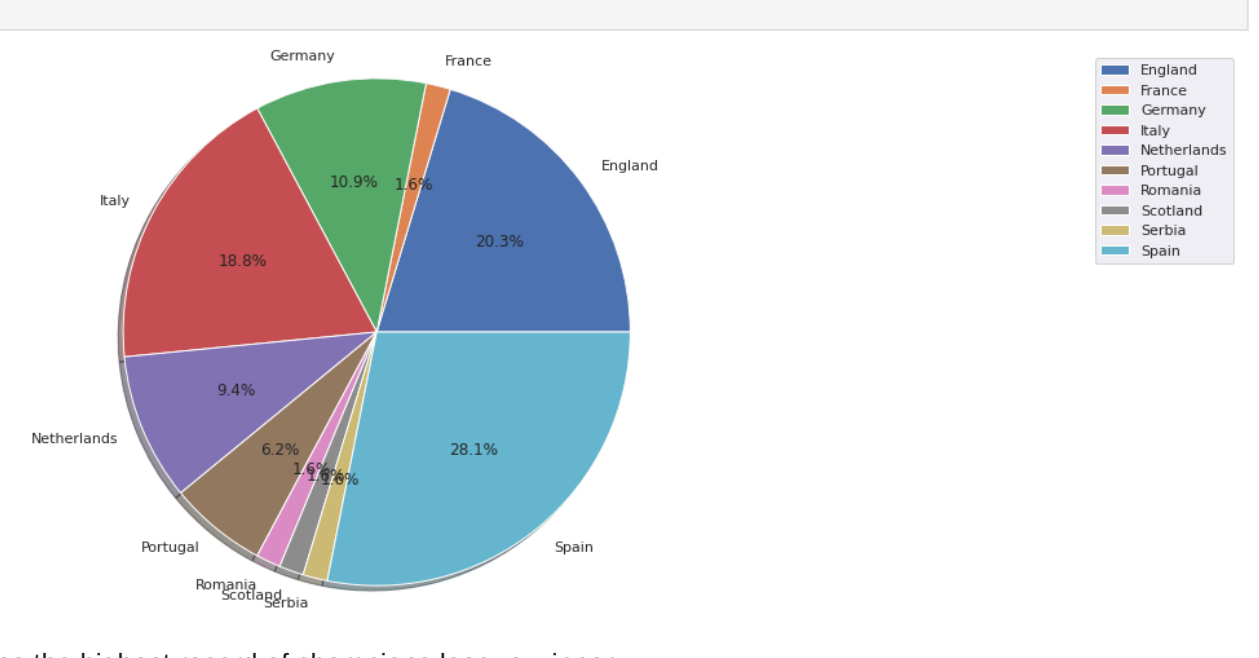
In [23]:
```
sns.set(style="darkgrid")
sns.barplot(top_club.index, top_club.club, alpha=0.9)
plt.title('Frequency Distribution of Champions league winner')
plt.ylabel('Number of Occurrences', fontsize=14)
plt.xlabel('Club', fontsize=14)
plt.show()
```

```
/home/akinkunmi/anaconda3/lib/python3.8/site-packages/seaborn/_decorators.py:36: FutureWarning: Pass the following variables as keyword args: x, y. From version 0.12, the only valid positional argument will be `data`, and passing other arguments without an explicit keyword will result in an error or misinterpretation.
  warnings.warn(
```



Also, let us show the distribution of the country from which the champion league winner comes from. We will visualize this using a pie chart.

In [24]:
```
labels = uefa_winner_df['nation'].astype('category').cat.categories.tolist()
counts = uefa_winner_df['nation'].value_counts()
sizes = [counts[var_cat] for var_cat in labels]
fig1, ax1 = plt.subplots()
ax1.pie(sizes, labels=labels, autopct='%1.1f%%', shadow=True) #autopct is show the % on plot
ax1.axis('equal')
plt.legend()
plt.show()
```



From the chart, we can see that Spain has the highest record of champions league winner.

Asking and Answering Questions

Now, let's ask some questions from our data and answer the question accordingly

Q1. Who is the champions league winner in the 1972/1973 season?

In [25]:
```
winner_1973 = uefa_winner_df[uefa_winner_df.season == "'1972/1973'"]
winner_1973
```

Out[25]:

	season	position	club	nation	coach	formation	mvp
17	'1972/1973'	winner	AFC Ajax	Netherlands	Stefan Kovacs	unknown	unknown

As we can see, the winner of 1972/1973 champions league is AFC Ajax.

Q2. Which coach has the record of highest no of champions league won?

In [26]:
```
winner_coach = uefa_winner_df['coach'].value_counts()
winner_coach
```

Out[26]:
```
Zinédine Zidane              3
Carlo Ancelotti              3
Robert Paisley               3
Sir Alex Ferguson            2
Miguel Muñoz Mozún           2
Ernst Happel                 2
Béla Guttmann                2
Jupp Heynckes                2
Nereo Rocco                  2
Brian Clough                 2
Luis Antonio Carniglia       2
Ottmar Hitzfeld              2
Arrigo Sacchi                2
Josep Guardiola              2
José Villalonga              2
Vicente del Bosque           2
José Mourinho                2
Stefan Kovacs                2
Dettmar Cramer               2
Helenio Herrera              2
Raymond Goethals             1
Rinus Michels                1
Louis van Gaal               1
Emeric Ienei                 1
Udo Lattek                   1
Jock Stein                   1
Artur Jorge Melo Teixeira    1
Marcello Lippi               1
Ljubo Petrovic               1
Anthony Barton               1
Fabio Capello                1
Sir Matt Busby               1
Guus Hiddink                 1
Roberto Di Matteo            1
Jürgen Klopp                 1
Giovanni Trapattoni          1
Luis Enrique                 1
Frank Rijkaard               1
Rafael Benítez               1
Johan Cruyff                 1
Joseph Fagan                 1
Name: coach, dtype: int64
```

Zinedine Zidane won the champions league most

Q3. Who is the champions league most valuable player in the 2011/2012 season?

In [27]:
```
mvp_2012 = uefa_winner_df[uefa_winner_df.season == "'2011/2012'"]
mvp = mvp_2012['mvp']
mvp
```

Out[27]:
```
56    Didier Drogba
Name: mvp, dtype: object
```

Q4. Which club has featured most in the final of champions leauge?

In [28]:
```
most_featured_club = uefa_df['club'].value_counts()
most_featured_club
```

Out[28]:
```
Real Madrid CF                  16
AC Milan                        11
FC Bayern Munchen               10
Juventus                         9
Liverpool FC                     9
Barcelona FC                     7
SL Benfica                       7
AFC Ajax                         6
FC Internazionale                5
Manchester United FC             5
Club Atlético de madrid          2
Chelsea FC                       2
Stade de Reims                   2
Nottingham Forest FC             2
Olympique de marseille fc        2
FC Porto                         2
Celtic FC                        2
FCSB                             2
Valencia CF                      2
Hamburg SV                       2
Borussia Dortmund                2
Bayer 04 Leverkusen              1
Leeds United                     1
Barcelona                        1
Malmo FF                         1
PSV Eindhoven                    1
Panathinaikos FC                 1
AS Monaco                        1
Club Brugge KV                   1
Vfl borussia mönchengladbach     1
Tottenham Hotspur                1
Club Atletico de Madrid          1
ACF Fiorentina                   1
As saint etienne                 1
UC sampdoria                     1
Arsenal FC                       1
FC crvena zvezda                 1
Feyenoord                        1
AS Roma                          1
Aston Villa FC                   1
Eintracht Frankfurt              1
FK Partizan                      1
Name: club, dtype: int64
```

Real Madrid CF has the highest record of champions league final.

Q5. How many times has the champion league final featured two clubs from England? And which season(s)?

In [33]:
```
x = uefa_winner_df[['season', 'position', 'club', 'nation']]
y = uefa_runnerup_df[['season', 'position', 'club', 'nation']]
y = y.rename(columns={'nation':'nation2'})
merged_df = x.merge(y, on='season') #we merge the winner and runner up dataframe together.
england = merged_df[merged_df.nation == 'England'][merged_df.nation2 == 'England']
england
```

```
/tmp/ipykernel_32670/1337420981.py:5: UserWarning: Boolean Series key will be reindexed to match DataFrame index.
  england = merged_df[merged_df.nation == 'England'][merged_df.nation2 == 'England']
```

Out[33]:

	season	position_x	club_x	nation	position_y	club_y	nation2
52	'2007/2008'	winner	Manchester United FC	England	runner up	Chelsea FC	England
63	'2018/2019'	winner	Liverpool FC	England	runner up	Tottenham Hotspur	England

Two England clubs had featured in champions league twice between 1956 and 2019; first time in 2007/2008 and second time in 2018/2019.

Q6. How many times has the champion league final featured two clubs from Spain? And which season(s)?

In [32]:
```
spain = merged_df[merged_df.nation == 'Spain'][merged_df.nation2 == 'Spain']
spain
```

```
/tmp/ipykernel_32670/1428149208.py:1: UserWarning: Boolean Series key will be reindexed to match DataFrame index.
  spain = merged_df[merged_df.nation == 'Spain'][merged_df.nation2 == 'Spain']
```

Out[32]:

	season	position_x	club_x	nation	position_y	club_y	nation2
44	'1999/2000'	winner	Real Madrid CF	Spain	runner up	Valencia CF	Spain
58	'2013/2014'	winner	Real Madrid CF	Spain	runner up	Club Atlético de madrid	Spain
60	'2015/2016'	winner	Real Madrid CF	Spain	runner up	Club Atlético de madrid	Spain

Three Spain clubs had featured in champions league thrice between 1956 and 2019; first time in 1999/200, second time in 2013/2014 and third time in 2015/2016.

Q7. With which club(s) has Carlo Ancelotti won champions league as a coach?

In [34]:
```
Ancelotti = uefa_winner_df[uefa_winner_df.coach == 'Carlo Ancelotti']
Ancelotti
```

Out[34]:

	season	position	club	nation	coach	formation	mvp
47	'2002/2003'	winner	AC Milan	Italy	Carlo Ancelotti	'4-4-2'	Paolo Maldini
51	'2006/2007'	winner	AC Milan	Italy	Carlo Ancelotti	'4-4-2'	Filippo Inzaghi
58	'2013/2014'	winner	Real Madrid CF	Spain	Carlo Ancelotti	'4-3-3'	Ángel di María

Inferences and Conclusion

From this analysis, we are able to arrive at the following conclusion:

1. Real Madrid CF has won the highest number of champions league title.
2. Real Madrid CF has featured in Champions league final more than any other club in the world.
3. Spain clubs has won more champions league than any other country.
4. Zinedine Zidane, Carlo Ancelotti and Robert Paisley have the highest record for winning champions league as coach. They won it three times each.
5. Two English clubs has featured in UEFA champions league final twice between 1956 and 2019 while Spanish clubs has done that three times.

Also, we were able to show the graphical distribution of champions league winner from 1956 to 2019. We also been able to use pie chart to analyze the country which has the highest number of champions league winning clubs.

References and Future Work

The dataset used for this analysis is downloaded from www.kaggle.com

The analysis is done based on the information provided in thie dataset. For future work, we hope to have a more comprehensive dataset that will include more information such as 'top scorer', 'host stadium for final' and so on.

Visit https://jovian.ai/learn/data-analysis-with-python-zero-to-pandas/lesson/lesson-4-analyzing-tabular-data-with-pandas to learn more about panda analysis.

Thanks for reading. Kindly contact Olalekan Rasaq on +2347063128038 for any question you may have.