

In [2]:

```
#S08 T01: Tasca Feature Engineering
```

In [3]:

```
#Nivell 1
```

In [6]:

```
#Exercici 1
#Agafa un conjunt de dades de tema esportiu que t'agradi i normalitza els atributs categòrics en dummy.
#Estandaritza els atributs numèrics amb StandardScaler.

#Llibreries
import numpy as np
import pandas as pd
import random
import seaborn as sns
import warnings
import statistics

#Prenem el conjunt de dades de tema esportiu de la pàgina web
#https://www.kaggle.com/abecklas/fifa-world-cup/version/5

#Dataset FIFA World Cup
#Fitxer: Python/WorldCupPlayers.csv
#Fitxer: Python/WorldCups.csv
#Fitxer: Python/WorldCupMatches.csv

wcupplay_df = pd.read_csv('Python/WorldCupPlayers.csv', engine="python", error_bad_lines=False, warn_bad_lines=False, sep=',')
wcup_df = pd.read_csv('Python/WorldCups.csv', engine="python", error_bad_lines=False, warn_bad_lines=False, sep=',')
wcupmatch_df = pd.read_csv('Python/WorldCupMatches.csv', engine="python", error_bad_lines=False, warn_bad_lines=False, sep=',')

#wcupplay_df = pd.read_csv('Python/WorldCupPlayers.csv', engine="python", error_bad_lines=False, warn_bad_lines=False, sep=',')
#wcup_df = pd.read_csv('Python/WorldCups.csv', engine="python", error_bad_lines=False, warn_bad_lines=False, sep=',')
#wcupmatch_df = pd.read_csv('Python/WorldCupMatches.csv', engine="python", error_bad_lines=False, warn_bad_lines=False, sep=',')
```

In [7]:

```
#Una variable dummy es una variable binaria que ens indica que una variable categorica
#pren un valor especific. Podem crear dummy variables a python mitjançant el metode usi
ng get_dummies().
```

```
#pandas.get_dummies(data,
#                      prefix=None,
#                      prefix_sep='_',
#                      dummy_na=False,
#                      columns=None,
#                      sparse=False,
#                      drop_first=False,
#                      dtype=None)
```

In [8]:

```
#Normalització d'atributs categòrics en dummy.
# Prenem els dataframes:
# - wcup_df
# - wcupPlay_df
```

In [9]:

```
#Variables categoriques del dataframe wcup_df: cerca
wcup_df.dtypes
```

Out[9]:

```
Year          int64
Country      object
Winner       object
Runners-Up   object
Third         object
Fourth        object
GoalsScored  int64
QualifiedTeams int64
MatchesPlayed int64
Attendance    object
dtype: object
```

In [10]:

```
#Variables categoriques del dataframe wcup_df: filtre
wcup_df[['Country','Winner','Runners-Up','Third','Fourth','Attendance']].dtypes
```

Out[10]:

```
Country      object
Winner       object
Runners-Up   object
Third         object
Fourth        object
Attendance   object
dtype: object
```

In [11]:

```
# Atributs categòrics en dummy: 'Country'  
wcup_co_dummy = pd.get_dummies(wcup_df['Country'])  
wcup_co_dummy
```

Out[11]:

In [12]:

```
# Atributs categòrics en dummy: 'Winner'  
wcup_win_dummy = pd.get_dummies(wcup_df['Winner']) # World Cup Winners  
wcup_win_dummy
```

Out[12]:

	Argentina	Brazil	England	France	Germany	Germany FR	Italy	Spain	Uruguay
0	0	0	0	0	0	0	0	0	1
1	0	0	0	0	0	0	1	0	0
2	0	0	0	0	0	0	1	0	0
3	0	0	0	0	0	0	0	0	1
4	0	0	0	0	0	1	0	0	0
5	0	1	0	0	0	0	0	0	0
6	0	1	0	0	0	0	0	0	0
7	0	0	1	0	0	0	0	0	0
8	0	1	0	0	0	0	0	0	0
9	0	0	0	0	0	1	0	0	0
10	1	0	0	0	0	0	0	0	0
11	0	0	0	0	0	0	1	0	0
12	1	0	0	0	0	0	0	0	0
13	0	0	0	0	0	1	0	0	0
14	0	1	0	0	0	0	0	0	0
15	0	0	0	1	0	0	0	0	0
16	0	1	0	0	0	0	0	0	0
17	0	0	0	0	0	0	1	0	0
18	0	0	0	0	0	0	0	1	0
19	0	0	0	0	1	0	0	0	0

In [13]:

```
# Atributs categòrics en dummy: 'Runners-Up'  
wcup_run_dummy = pd.get_dummies(wcup_df['Runners-Up']) # Second position in World Cup  
wcup_run_dummy
```

Out[13]:

	Argentina	Brazil	Czechoslovakia	France	Germany	Germany FR	Hungary	Italy	Netherlar
0	1	0		0	0	0	0	0	0
1	0	0		1	0	0	0	0	0
2	0	0		0	0	0	0	1	0
3	0	1		0	0	0	0	0	0
4	0	0		0	0	0	0	1	0
5	0	0		0	0	0	0	0	0
6	0	0		1	0	0	0	0	0
7	0	0		0	0	0	1	0	0
8	0	0		0	0	0	0	0	1
9	0	0		0	0	0	0	0	0
10	0	0		0	0	0	0	0	0
11	0	0		0	0	0	1	0	0
12	0	0		0	0	0	1	0	0
13	1	0		0	0	0	0	0	0
14	0	0		0	0	0	0	0	1
15	0	1		0	0	0	0	0	0
16	0	0		0	0	1	0	0	0
17	0	0		0	1	0	0	0	0
18	0	0		0	0	0	0	0	0
19	1	0		0	0	0	0	0	0



In [14]:

```
# Atributs categòrics en dummy: 'Third'  
wcup_3_dummy = pd.get_dummies(wcup_df['Third']) # Third position in World Cup  
wcup_3_dummy
```

Out[14]:

	Austria	Brazil	Chile	Croatia	France	Germany	Germany FR	Italy	Netherlands	Poland
0	0	0	0	0	0	0	0	0	0	0
1	0	0	0	0	0	1	0	0	0	0
2	0	1	0	0	0	0	0	0	0	0
3	0	0	0	0	0	0	0	0	0	0
4	1	0	0	0	0	0	0	0	0	0
5	0	0	0	0	1	0	0	0	0	0
6	0	0	1	0	0	0	0	0	0	0
7	0	0	0	0	0	0	0	0	0	0
8	0	0	0	0	0	0	1	0	0	0
9	0	0	0	0	0	0	0	0	0	1
10	0	1	0	0	0	0	0	0	0	0
11	0	0	0	0	0	0	0	0	0	1
12	0	0	0	0	1	0	0	0	0	0
13	0	0	0	0	0	0	0	1	0	0
14	0	0	0	0	0	0	0	0	0	0
15	0	0	0	1	0	0	0	0	0	0
16	0	0	0	0	0	0	0	0	0	0
17	0	0	0	0	0	1	0	0	0	0
18	0	0	0	0	0	1	0	0	0	0
19	0	0	0	0	0	0	0	0	1	0



In [15]:

```
# Atributs categòrics en dummy: 'Fourth'  
wcup_4_dummy = pd.get_dummies(wcup_df['Fourth']) # Fourth position in World Cup  
wcup_4_dummy
```

Out[15]:

	Austria	Belgium	Brazil	Bulgaria	England	France	Germany FR	Italy	Korea Republic	Netherlar
0	0	0	0	0	0	0	0	0	0	0
1	1	0	0	0	0	0	0	0	0	0
2	0	0	0	0	0	0	0	0	0	0
3	0	0	0	0	0	0	0	0	0	0
4	0	0	0	0	0	0	0	0	0	0
5	0	0	0	0	0	0	1	0	0	0
6	0	0	0	0	0	0	0	0	0	0
7	0	0	0	0	0	0	0	0	0	0
8	0	0	0	0	0	0	0	0	0	0
9	0	0	1	0	0	0	0	0	0	0
10	0	0	0	0	0	0	0	1	0	0
11	0	0	0	0	0	1	0	0	0	0
12	0	1	0	0	0	0	0	0	0	0
13	0	0	0	0	1	0	0	0	0	0
14	0	0	0	1	0	0	0	0	0	0
15	0	0	0	0	0	0	0	0	0	0
16	0	0	0	0	0	0	0	0	0	1
17	0	0	0	0	0	0	0	0	0	0
18	0	0	0	0	0	0	0	0	0	0
19	0	0	1	0	0	0	0	0	0	0

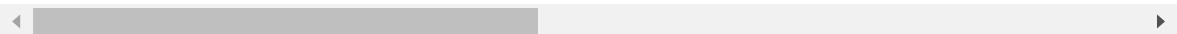


In [16]:

```
# Atributs categòrics en dummy: 'Attendance'  
wcup_attend_dummy = pd.get_dummies(wcup_df['Attendance']) # People in World Cup  
wcup_attend_dummy
```

Out[16]:

	1.045.246	1.545.791	1.563.135	1.603.975	1.865.753	2.109.723	2.394.031	2.516.215	2.70
0	0	0	0	0	0	0	0	0	0
1	0	0	0	0	0	0	0	0	0
2	0	0	0	0	0	0	0	0	0
3	1	0	0	0	0	0	0	0	0
4	0	0	0	0	0	0	0	0	0
5	0	0	0	0	0	0	0	0	0
6	0	0	0	0	0	0	0	0	0
7	0	0	1	0	0	0	0	0	0
8	0	0	0	1	0	0	0	0	0
9	0	0	0	0	1	0	0	0	0
10	0	1	0	0	0	0	0	0	0
11	0	0	0	0	0	1	0	0	0
12	0	0	0	0	0	0	1	0	0
13	0	0	0	0	0	0	0	0	1
14	0	0	0	0	0	0	0	0	0
15	0	0	0	0	0	0	0	0	0
16	0	0	0	0	0	0	0	0	0
17	0	0	0	0	0	0	0	0	0
18	0	0	0	0	0	0	0	0	0
19	0	0	0	0	0	0	0	0	0



In [17]:

```
#Variables categoriques del dataframe wcupplay_df: cerca  
wcupplay_df.dtypes
```

Out[17]:

```
RoundID          int64  
MatchID         int64  
Team Initials   object  
Coach Name      object  
Line-up          object  
Shirt Number    int64  
Player Name     object  
Position         object  
Event            object  
dtype: object
```

In [18]:

```
#Variables categoriques del dataframe wcupplay_df: filtre  
wcupplay_df[['Team Initials','Coach Name','Line-up','Player Name','Position','Event']].  
dtypes
```

Out[18]:

```
Team Initials   object  
Coach Name      object  
Line-up          object  
Player Name     object  
Position         object  
Event            object  
dtype: object
```

In [19]:

```
# Atributs categòrics en dummy: 'Team Initials'  
wcupplay_team = pd.get_dummies(wcupplay_df['Team Initials'])  
wcupplay_team.tail(n=10)
```

Out[19]:

	ALG	ANG	ARG	AUS	AUT	BEL	BIH	BOL	BRA	BUL	...	TUN	TUR	UAE	UKF
37774	0	0	0	0	0	0	0	0	0	0	...	0	0	0	(
37775	0	0	1	0	0	0	0	0	0	0	...	0	0	0	(
37776	0	0	0	0	0	0	0	0	0	0	...	0	0	0	(
37777	0	0	1	0	0	0	0	0	0	0	...	0	0	0	(
37778	0	0	0	0	0	0	0	0	0	0	...	0	0	0	(
37779	0	0	1	0	0	0	0	0	0	0	...	0	0	0	(
37780	0	0	0	0	0	0	0	0	0	0	...	0	0	0	(
37781	0	0	1	0	0	0	0	0	0	0	...	0	0	0	(
37782	0	0	0	0	0	0	0	0	0	0	...	0	0	0	(
37783	0	0	1	0	0	0	0	0	0	0	...	0	0	0	(

10 rows × 82 columns

In [20]:

```
# Atributs categòrics en dummy: 'Coach Name'  
wcupplay_coach = pd.get_dummies(wcupplay_df['Coach Name'])  
wcupplay_coach.head(n=10)
```

Out[20]:

	ACOSTA Nelson (URU)	ADSHEAD John (NZL)	AL JOHAR Nasser (KSA)	AL KARASHI Mohamed (KSA)	ALAMOS Luis (CHI)	ANDREOLI Franco (SUI)	ANTIC Radomir (SRB)	APPIAH James (GHA)	ARAGC Luis (
0	0	0	0	0	0	0	0	0	0
1	0	0	0	0	0	0	0	0	0
2	0	0	0	0	0	0	0	0	0
3	0	0	0	0	0	0	0	0	0
4	0	0	0	0	0	0	0	0	0
5	0	0	0	0	0	0	0	0	0
6	0	0	0	0	0	0	0	0	0
7	0	0	0	0	0	0	0	0	0
8	0	0	0	0	0	0	0	0	0
9	0	0	0	0	0	0	0	0	0

10 rows × 335 columns

In [21]:

```
# Atributs categòrics en dummy: 'Line-up'  
wcupplay_lup = pd.get_dummies(wcupplay_df['Line-up'])  
wcupplay_lup.head(n=10)
```

Out[21]:

	N	S
0	0	1
1	0	1
2	0	1
3	0	1
4	0	1
5	0	1
6	0	1
7	0	1
8	0	1
9	0	1

In [22]:

```
# Atributs categòrics en dummy: 'Player Name'  
wcupplay_player = pd.get_dummies(wcupplay_df['Player Name'])  
wcupplay_player.tail(n=10)
```

Out[22]:

	URI? I?	?	BAUTISTA	A COLE	A GUARDADO	A MEDINA	A A. AL- DOSSARY	A. AL- GANOUBI	A. ALMEIDA	A. AYE
37774	0	0	0	0	0	0	0	0	0	0
37775	0	0	0	0	0	0	0	0	0	0
37776	0	0	0	0	0	0	0	0	0	0
37777	0	0	0	0	0	0	0	0	0	0
37778	0	0	0	0	0	0	0	0	0	0
37779	0	0	0	0	0	0	0	0	0	0
37780	0	0	0	0	0	0	0	0	0	0
37781	0	0	0	0	0	0	0	0	0	0
37782	0	0	0	0	0	0	0	0	0	0
37783	0	0	0	0	0	0	0	0	0	0

10 rows × 7663 columns

In [23]:

```
# Atributs categòrics en dummy: 'Position'  
wcupplay_pos = pd.get_dummies(wcupplay_df['Position'])  
wcupplay_pos.head(n=10)
```

Out[23]:

C	GK	GKC
0	0	1
1	0	1
2	0	0
3	0	0
4	0	0
5	1	0
6	0	0
7	0	0
8	0	0
9	0	0

In [24]:

```
# Atributs categòrics en dummy: 'Event'  
wcupplay_event = pd.get_dummies(wcupplay_df['Event'])  
wcupplay_event.tail(n=10)
```

Out[24]:

	G1'	G1' G42'	G1' G6' G53'	G1' G67'	G1' O74'	G10'	G10' G13'	G10' G15' G50'	G10' G30'	G10' G90'	...	Y90'	Y90' G90'	Y90' O105'	Y90' O9'
37774	0	0	0	0	0	0	0	0	0	0	0	...	0	0	0
37775	0	0	0	0	0	0	0	0	0	0	0	...	0	0	0
37776	0	0	0	0	0	0	0	0	0	0	0	...	0	0	0
37777	0	0	0	0	0	0	0	0	0	0	0	...	0	0	0
37778	0	0	0	0	0	0	0	0	0	0	0	...	0	0	0
37779	0	0	0	0	0	0	0	0	0	0	0	...	0	0	0
37780	0	0	0	0	0	0	0	0	0	0	0	...	0	0	0
37781	0	0	0	0	0	0	0	0	0	0	0	...	0	0	0
37782	0	0	0	0	0	0	0	0	0	0	0	...	0	0	0
37783	0	0	0	0	0	0	0	0	0	0	0	...	0	0	0

10 rows × 1893 columns

In [25]:

```
#Estandarització d'atributs numèrics amb StandardScaler.  
# Prenem el dataframe:  
# - wcupmatch_df
```

In [26]:

```
#Variables numeriques del dataframe wcupmatch_df: cerca  
wcupmatch_df.dtypes
```

Out[26]:

```
Year                  int64  
Datetime             object  
Stage                object  
Stadium              object  
City                 object  
Home Team Name       object  
Home Team Goals      int64  
Away Team Goals      int64  
Away Team Name       object  
Win conditions       object  
Attendance          float64  
Half-time Home Goals int64  
Half-time Away Goals int64  
Referee              object  
Assistant 1          object  
Assistant 2          object  
RoundID              int64  
MatchID              int64  
Home Team Initials   object  
Away Team Initials   object  
dtype: object
```

In [27]:

```
#Variables numeriques del dataframe wcupmatch_df: filtre  
wcupmatch_df[['Year', 'Home Team Goals', 'Away Team Goals', 'Half-time Home Goals', 'Half-time Away Goals']].dtypes
```

Out[27]:

```
Year                  int64  
Home Team Goals      int64  
Away Team Goals      int64  
Half-time Home Goals int64  
Half-time Away Goals int64  
dtype: object
```

In [28]:

```
#Function StandardScaler  
#Transforma les dades per obtenir una mitjana de 0 i una desviació estàndard de 1.  
#La tècnica consisteix a calcular la mitjana estadística i la desviació estàndard  
#dels valors dels atributs, restar la mitjana de cada valor i dividir el resultat  
#per la desviació estàndard.  
  
#class sklearn.preprocessing.  
#StandardScaler(*,  
#                 copy=True,  
#                 with_mean=True,  
#                 with_std=True)
```

In [29]:

```
#StandardScaler per valors numèrics del dataframe wcupmatch_df  
from sklearn.preprocessing import StandardScaler  
#Implementation  
data = wcupmatch_df[['Year', 'Home Team Goals', 'Away Team Goals', 'Half-time Home Goals',  
'Half-time Away Goals']]  
data = data.dropna() #Esborrem valors NaN  
scaler = StandardScaler() #  
scaler.fit(data) #  
print('<-- Dataframe wcupmatch_df preprocessat --> \n', scaler.transform(data))  
  
<-- Dataframe wcupmatch_df preprocessat -->  
[[-2.45543187  1.36018996 -0.02051684  2.44547789 -0.6201146 ]  
 [-2.45543187  0.73880559 -0.9405351   1.37808693 -0.6201146 ]  
 [-2.45543187  0.11742122 -0.02051684  1.37808693 -0.6201146 ]  
 ...  
 [ 1.28860998 -1.12534751 -0.9405351   -0.756695    -0.6201146 ]  
 [ 1.28860998 -1.12534751  1.81951969 -0.756695     2.27488615]  
 [ 1.28860998 -0.50396314 -0.9405351   -0.756695    -0.6201146 ]]
```

In [32]:

```
scaled_data = scaler.transform(data)  
print(scaled_data.mean(axis = 0))  
  
[-2.60198748e-15  8.33970347e-17  5.00382208e-17 -3.33588139e-17  
 2.08492587e-17]
```

In [33]:

```
print('<-- Dataframe wcupmatch_df preprocessat --> \n', scaler.mean_)  
  
<-- Dataframe wcupmatch_df preprocessat -->  
[1.98508920e+03 1.81103286e+00 1.02230047e+00 7.08920188e-01  
 4.28403756e-01]
```

In [34]:

```
# Atributs categòrics en dummy: 'Year'  
wcupmatch_year = wcupmatch_df[['Year']]  
year_scaler = StandardScaler()  
scaler.fit(wcupmatch_year)
```

Out[34]:

```
StandardScaler()
```

In [35]:

```
scaler.transform(wcupmatch_year)
```


In [36]:

```
# Atributs categòrics en dummy: 'Home Team Goals'  
wcupmatch_htg = pd.get_dummies(wcupmatch_df['Home Team Goals'])
```

In [37]:

```
wcupmatch_htg
```

Out[37]:

	0	1	2	3	4	5	6	7	8	9	10
0	0	0	0	0	1	0	0	0	0	0	0
1	0	0	0	1	0	0	0	0	0	0	0
2	0	0	1	0	0	0	0	0	0	0	0
3	0	0	0	1	0	0	0	0	0	0	0
4	0	1	0	0	0	0	0	0	0	0	0
...
847	1	0	0	0	0	0	0	0	0	0	0
848	0	1	0	0	0	0	0	0	0	0	0
849	1	0	0	0	0	0	0	0	0	0	0
850	1	0	0	0	0	0	0	0	0	0	0
851	0	1	0	0	0	0	0	0	0	0	0

852 rows × 11 columns

In [38]:

```
# Atributs categòrics en dummy: 'Away Team Goals'  
wcupmatch_atg = pd.get_dummies(wcupmatch_df['Away Team Goals'])
```

In [39]:

```
wcupmatch_atg
```

Out[39]:

	0	1	2	3	4	5	7
0	0	1	0	0	0	0	0
1	1	0	0	0	0	0	0
2	0	1	0	0	0	0	0
3	0	1	0	0	0	0	0
4	1	0	0	0	0	0	0
...
847	1	0	0	0	0	0	0
848	0	0	0	0	0	0	1
849	1	0	0	0	0	0	0
850	0	0	0	1	0	0	0
851	1	0	0	0	0	0	0

852 rows × 7 columns

In [40]:

```
# Atributs categòrics en dummy: 'Half-time Home Goals'  
wcupmatch_hthg = pd.get_dummies(wcupmatch_df['Half-time Home Goals'])
```

In [41]:

```
wcupmatch_hthg
```

Out[41]:

	0	1	2	3	4	5	6
0	0	0	0	1	0	0	0
1	0	0	1	0	0	0	0
2	0	0	1	0	0	0	0
3	0	1	0	0	0	0	0
4	1	0	0	0	0	0	0
...
847	1	0	0	0	0	0	0
848	1	0	0	0	0	0	0
849	1	0	0	0	0	0	0
850	1	0	0	0	0	0	0
851	1	0	0	0	0	0	0

852 rows × 7 columns

In [42]:

```
# Atributs categòrics en dummy: 'Half-time Away Goals'  
wcupmatch_htag = pd.get_dummies(wcupmatch_df['Half-time Away Goals'])
```

In [43]:

```
wcupmatch_htag
```

Out[43]:

	0	1	2	3	4	5
0	1	0	0	0	0	0
1	1	0	0	0	0	0
2	1	0	0	0	0	0
3	1	0	0	0	0	0
4	1	0	0	0	0	0
...
847	1	0	0	0	0	0
848	0	0	0	0	0	1
849	1	0	0	0	0	0
850	0	0	1	0	0	0
851	1	0	0	0	0	0

852 rows × 6 columns

In [92]:

```
#StandardScaler per valors numèrics del dataframe wcup_df
from sklearn.preprocessing import StandardScaler
#Implementation
data = wcup_df[['GoalsScored','QualifiedTeams','MatchesPlayed','Attendance']]
data = data.dropna() #Esborrem valors NaN
scaler = StandardScaler() #
scaler.fit(data) #
print('<-- Dataframe wcupmatch_df preprocessat --> \n', scaler.transform(data))

<-- Dataframe wcupmatch_df preprocessat -->
[[-1.52312166 -1.16454471 -1.41812453 -1.2274627 ]
 [-1.52312166 -0.74107391 -1.4777096 -1.44527495]
 [-1.08749953 -0.88223084 -1.41812453 -1.43311838]
 [-0.96303607 -1.16454471 -1.17978427 -0.79222205]
 [ 0.65498899 -0.74107391 -0.94144402 -1.05702375]
 [ 0.21936686 -0.74107391 -0.40517844 -1.00801171]
 [-0.9319202 -0.74107391 -0.58393363 -0.93778885]
 [-0.9319202 -0.74107391 -0.58393363 -0.29649336]
 [-0.745225 -0.74107391 -0.58393363 -0.25740089]
 [-0.68299327 -0.74107391 -0.22642324 -0.00682429]
 [-0.52741394 -0.74107391 -0.22642324 -0.31309521]
 [ 0.84168419  0.38818157  0.60776766  0.22670631]
 [ 0.40606206  0.38818157  0.60776766  0.49884886]
 [-0.12290767  0.38818157  0.60776766  0.61580464]
 [ 0.68610485  0.38818157  0.60776766  1.64128606]
 [ 1.61958085  1.51743705  1.32278843  0.87318414]
 [ 1.30842218  1.51743705  1.32278843  0.79670016]
 [ 0.87280005  1.51743705  1.32278843  1.42294735]
 [ 0.81056832  1.51743705  1.32278843  1.25009144]
 [ 1.61958085  1.51743705  1.32278843  1.44914715]]
```

In [105]:

```
scaled_data = scaler.transform(data)
print(scaled_data.mean(axis = 0))

[-7.77156117e-17 -2.22044605e-17  1.77635684e-16 -6.66133815e-17]
```

In [106]:

```
print('<-- Dataframe wcupmatch_df preprocessat --> \n', scaler.mean_)

<-- Dataframe wcupmatch_df preprocessat -->
[1.18950000e+02 2.12500000e+01 4.18000000e+01 1.87288235e+06]
```

In []:

```
#Nivell 2
```

In [159]:

```
#Exercici 2  
#Continua amb el conjunt de dades de tema esportiu que t'agradi i aplica (PCA)  
#l'anàlisi de components principals.
```

```
import pandas as pd  
import numpy as np  
import matplotlib.pyplot as plt  
from sklearn.decomposition import PCA, KernelPCA  
from sklearn.datasets import make_circles  
from sklearn.preprocessing import StandardScaler  
%matplotlib inline
```

In [160]:

```
#Prenem el conjunt de dades wcup_df
```

In [161]:

```
wcup_df
```

Out[161]:

	Year	Country	Winner	Runners-Up	Third	Fourth	GoalsScored	Qual
0	1930	Uruguay	Uruguay	Argentina	USA	Yugoslavia	70	
1	1934	Italy	Italy	Czechoslovakia	Germany	Austria	70	
2	1938	France	Italy	Hungary	Brazil	Sweden	84	
3	1950	Brazil	Uruguay	Brazil	Sweden	Spain	88	
4	1954	Switzerland	Germany FR	Hungary	Austria	Uruguay	140	
5	1958	Sweden	Brazil	Sweden	France	Germany FR	126	
6	1962	Chile	Brazil	Czechoslovakia	Chile	Yugoslavia	89	
7	1966	England	England	Germany FR	Portugal	Soviet Union	89	
8	1970	Mexico	Brazil	Italy	Germany FR	Uruguay	95	
9	1974	Germany	Germany FR	Netherlands	Poland	Brazil	97	
10	1978	Argentina	Argentina	Netherlands	Brazil	Italy	102	
11	1982	Spain	Italy	Germany FR	Poland	France	146	
12	1986	Mexico	Argentina	Germany FR	France	Belgium	132	
13	1990	Italy	Germany FR	Argentina	Italy	England	115	
14	1994	USA	Brazil	Italy	Sweden	Bulgaria	141	
15	1998	France	France	Brazil	Croatia	Netherlands	171	
16	2002	Korea/Japan	Brazil	Germany	Turkey	Korea Republic	161	
17	2006	Germany	Italy	France	Germany	Portugal	147	
18	2010	South Africa	Spain	Netherlands	Germany	Uruguay	145	
19	2014	Brazil	Germany	Argentina	Netherlands	Brazil	171	



In [162]:

```
pca_wcup = wcup_df.iloc[:,6: ]  
pca_wcup
```

Out[162]:

	GoalsScored	QualifiedTeams	MatchesPlayed	Attendance
0	70	13	18	590549
1	70	16	17	363000
2	84	15	18	375700
3	88	13	22	1045246
4	140	16	26	768607
5	126	16	35	819810
6	89	16	32	893172
7	89	16	32	1563135
8	95	16	32	1603975
9	97	16	38	1865753
10	102	16	38	1545791
11	146	24	52	2109723
12	132	24	52	2394031
13	115	24	52	2516215
14	141	24	52	3587538
15	171	32	64	2785100
16	161	32	64	2705197
17	147	32	64	3359439
18	145	32	64	3178856
19	171	32	64	3386810

In [214]:

```
#Apliquem el PCA
pca = PCA(n_components=4) # Utilitzem 4 atributs del dataframe wcup_df
pca.fit(pca_wcup)
pca_transformed = pca.transform(pca_wcup)
pca_df = pd.DataFrame(data = pComponents, columns = ['GoalsScored', 'QualifiedTeams', 'MatchesPlayed', 'Attendance']).set_index(pca_wcup.index)
#Visualització de valors relacionats amb el PCA
print(pComponents)
print(pComponents[0])
print("Shape de dataframe PCA", pca_df.shape)
print("Components de PCA \n", pca.components_)
print("Explained variance de PCA \n", pca.explained_variance_)
print("Explained variance ratio de PCA \n", pca.explained_variance_ratio_)
print("Valors singulars de PCA \n", pca.singular_values_)
```

```
[[ -1.28233335e+06 -1.89002980e+01 6.23798208e-02 1.52683686e+00]
 [-1.50988235e+06 -1.26673403e+01 -2.55955496e+00 4.42950946e+00]
 [-1.49718235e+06 7.10505103e-01 3.17848484e-01 2.95201011e+00]
 [-8.27636351e+05 -1.27966748e+01 5.29508273e+00 1.45164218e-01]
 [-1.10427535e+06 4.63730907e+01 9.75965736e+00 3.09880480e-01]
 [-1.05307235e+06 3.32283095e+01 -9.42737568e-01 -3.18190816e+00]
 [-9.79710350e+05 -5.35726541e+00 -6.08937462e+00 -1.42293943e+00]
 [-3.09747351e+05 -2.33781405e+01 7.18700046e-01 -7.09226272e-01]
 [-2.68907351e+05 -1.86396617e+01 2.52086790e+00 -7.37532735e-01]
 [-7.12935061e+03 -2.25005976e+01 3.19398224e-01 -2.95853111e+00]
 [-3.27091350e+05 -9.02998770e+00 -1.77604860e+00 -3.35923189e+00]
 [ 2.36840651e+05 2.23352485e+01 -1.48750565e+00 -1.77581509e+00]
 [ 5.21148650e+05 1.06816225e+00 -1.83510585e+00 -1.30537648e+00]
 [ 6.43332650e+05 -1.87565647e+01 -4.52376712e+00 -9.71742247e-01]
 [ 1.71465565e+06 -2.22796548e+01 1.23738890e+01 -1.41650782e-01]
 [ 9.12217651e+05 3.18073560e+01 -2.68446848e+00 9.79044570e-01]
 [ 8.32314651e+05 2.42282688e+01 -5.80836081e+00 1.01361274e+00]
 [ 1.48655665e+06 -6.98941824e+00 -2.39674103e+00 1.87814290e+00]
 [ 1.30597365e+06 -4.07770866e+00 -4.69418713e+00 1.70970521e+00]
 [ 1.51392765e+06 1.56223715e+01 3.43002830e+00 1.62004765e+00]]
[-1.28233335e+06 -1.89002980e+01 6.23798208e-02 1.52683686e+00]
```

Shape de dataframe PCA (20, 4)

Components de PCA

```
[[ 2.38310021e-05 5.93597046e-06 1.49921741e-05 1.00000000e+00]
 [ 9.72834314e-01 1.06023443e-01 2.05797053e-01 -2.68983139e-05]
 [ 2.31192870e-01 -3.98982313e-01 -8.87334757e-01 1.01618666e-05]
 [-1.19689019e-02 9.10808511e-01 -4.12655549e-01 1.06530235e-06]]
```

Explained variance de PCA

```
[1.14884559e+12 4.58077367e+02 2.30362170e+01 4.25788032e+00]
```

Explained variance ratio de PCA

```
[1.00000000e+00 3.98728402e-10 2.00516216e-11 3.70622506e-12]
```

Valors singulars de PCA

```
[4.67205161e+06 9.32923897e+01 2.09209972e+01 8.99442750e+00]
```

In [215]:

```
pca_df
```

Out[215]:

	GoalsScored	QualifiedTeams	MatchesPlayed	Attendance
0	-1.282333e+06	-18.900298	0.062380	1.526837
1	-1.509882e+06	-12.667340	-2.559555	4.429509
2	-1.497182e+06	0.710505	0.317848	2.952010
3	-8.276364e+05	-12.796675	5.295083	0.145164
4	-1.104275e+06	46.373091	9.759657	0.309880
5	-1.053072e+06	33.228309	-0.942738	-3.181908
6	-9.797104e+05	-5.357265	-6.089375	-1.422939
7	-3.097474e+05	-23.378140	0.718700	-0.709226
8	-2.689074e+05	-18.639662	2.520868	-0.737533
9	-7.129351e+03	-22.500598	0.319398	-2.958531
10	-3.270914e+05	-9.029988	-1.776049	-3.359232
11	2.368407e+05	22.335248	-1.487506	-1.775815
12	5.211487e+05	1.068162	-1.835106	-1.305376
13	6.433326e+05	-18.756565	-4.523767	-0.971742
14	1.714656e+06	-22.279655	12.373889	-0.141651
15	9.122177e+05	31.807356	-2.684468	0.979045
16	8.323147e+05	24.228269	-5.808361	1.013613
17	1.486557e+06	-6.989418	-2.396741	1.878143
18	1.305974e+06	-4.077709	-4.694187	1.709705
19	1.513928e+06	15.622372	3.430028	1.620048

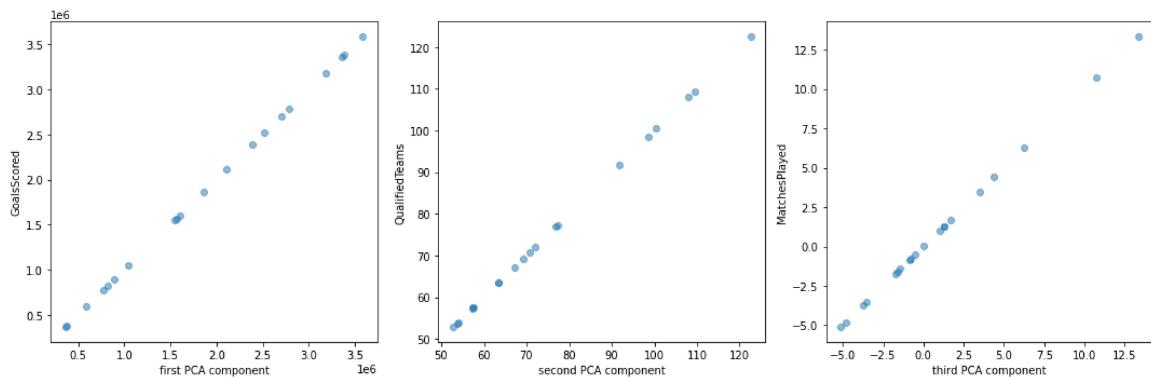
In [217]:

```
#Gràfic Plot de PCA
x = pca_wcup.dot(pca.components_[0])
y = pca_wcup.dot(pca.components_[1])
z = pca_wcup.dot(pca.components_[2])

fig, axes = plt.subplots(1,3,figsize=(15, 5))

axes[0].scatter(pca_wcup.dot(pca.components_[0]), x, alpha=0.5)
axes[0].set(xlabel="Primer PCA component", ylabel="GoalsScored")
axes[1].scatter(pca_wcup.dot(pca.components_[1]), y, alpha=0.5)
axes[1].set(xlabel="Segon PCA component", ylabel="QualifiedTeams")
axes[2].scatter(pca_wcup.dot(pca.components_[2]), z, alpha=0.5)
axes[2].set(xlabel="Tercer PCA component", ylabel="MatchesPlayed")

plt.tight_layout()
plt.show()
```



In [196]:

```
#Nivell 3
```

In [197]:

```
#Exercici 3
#Continua amb el conjunt de dades de tema esportiu que t'agradi
#i normalitza les dades tenint en compte els outliers.

#Un outlier és qualsevol dada o dades que destaca de la resta de dades del dataset.
#Hi ha 2 tipus d'outliers: Univariate i Multivariate.
#Un outlier Univariate és una dada que consisteix en valors extrems dins d'una única variable.
#Un outlier Multivariate és un resultat combinat d'almenys 2 variables.

#Per exemple, al dataframe wcup_df uns outliers podrien esser:
#Tipologia Univariate
#1994 USA Brazil Italy Sweden Bulgaria 141 24 52 3.587.5
#38
#2006 Germany Italy France Germany Portugal 147 32 64 3.359.4
#39
#2010 South Africa Spain Netherlands Germany Uruguay 145 32 64
#3.178.856
#2014 Brazil Germany Argentina Netherlands Brazil 171 32 64
#3.386.810

#Tipologia Univariate:
#El dataset no disposa d'outliers Multivariate.

#Atribut on hem trobat outliers
wcup_df['Attendance']
```

Out[197]:

```
0      590.549
1      363.000
2      375.700
3      1.045.246
4      768.607
5      819.810
6      893.172
7      1.563.135
8      1.603.975
9      1.865.753
10     1.545.791
11     2.109.723
12     2.394.031
13     2.516.215
14     3.587.538
15     2.785.100
16     2.705.197
17     3.359.439
18     3.178.856
19     3.386.810
Name: Attendance, dtype: object
```

In [199]:

```
#1. Dataframe: wcup_df : Conversió de l'atribut 'Attendance' de tipus objecte a tipus numèric
wcup_df['Attendance'] = wcup_df['Attendance'].str.replace('.', '').sort_values()
```

In [200]:

```
wcup_df['Attendance'] = wcup_df['Attendance'].astype(str).astype('int64')
```

In [202]:

```
#Esborrem les instàncies que hem indicat com a outliers  
wcup_df.drop(wcup_df[wcup_df.Attendance > 3000000].index, inplace=True)
```

In [204]:

```
#Disposem de les dades netes d'outliers (atribut Attendance).  
wcup_df
```

Out[204]:

	Year	Country	Winner	Runners-Up	Third	Fourth	GoalsScored	Qualifie
0	1930	Uruguay	Uruguay	Argentina	USA	Yugoslavia	70	
1	1934	Italy	Italy	Czechoslovakia	Germany	Austria	70	
2	1938	France	Italy	Hungary	Brazil	Sweden	84	
3	1950	Brazil	Uruguay	Brazil	Sweden	Spain	88	
4	1954	Switzerland	Germany FR	Hungary	Austria	Uruguay	140	
5	1958	Sweden	Brazil	Sweden	France	Germany FR	126	
6	1962	Chile	Brazil	Czechoslovakia	Chile	Yugoslavia	89	
7	1966	England	England	Germany FR	Portugal	Soviet Union	89	
8	1970	Mexico	Brazil	Italy	Germany FR	Uruguay	95	
9	1974	Germany	Germany FR	Netherlands	Poland	Brazil	97	
10	1978	Argentina	Argentina	Netherlands	Brazil	Italy	102	
11	1982	Spain	Italy	Germany FR	Poland	France	146	
12	1986	Mexico	Argentina	Germany FR	France	Belgium	132	
13	1990	Italy	Germany FR	Argentina	Italy	England	115	
15	1998	France	France	Brazil	Croatia	Netherlands	171	
16	2002	Korea/Japan	Brazil	Germany	Turkey	Korea Republic	161	

◀ ▶

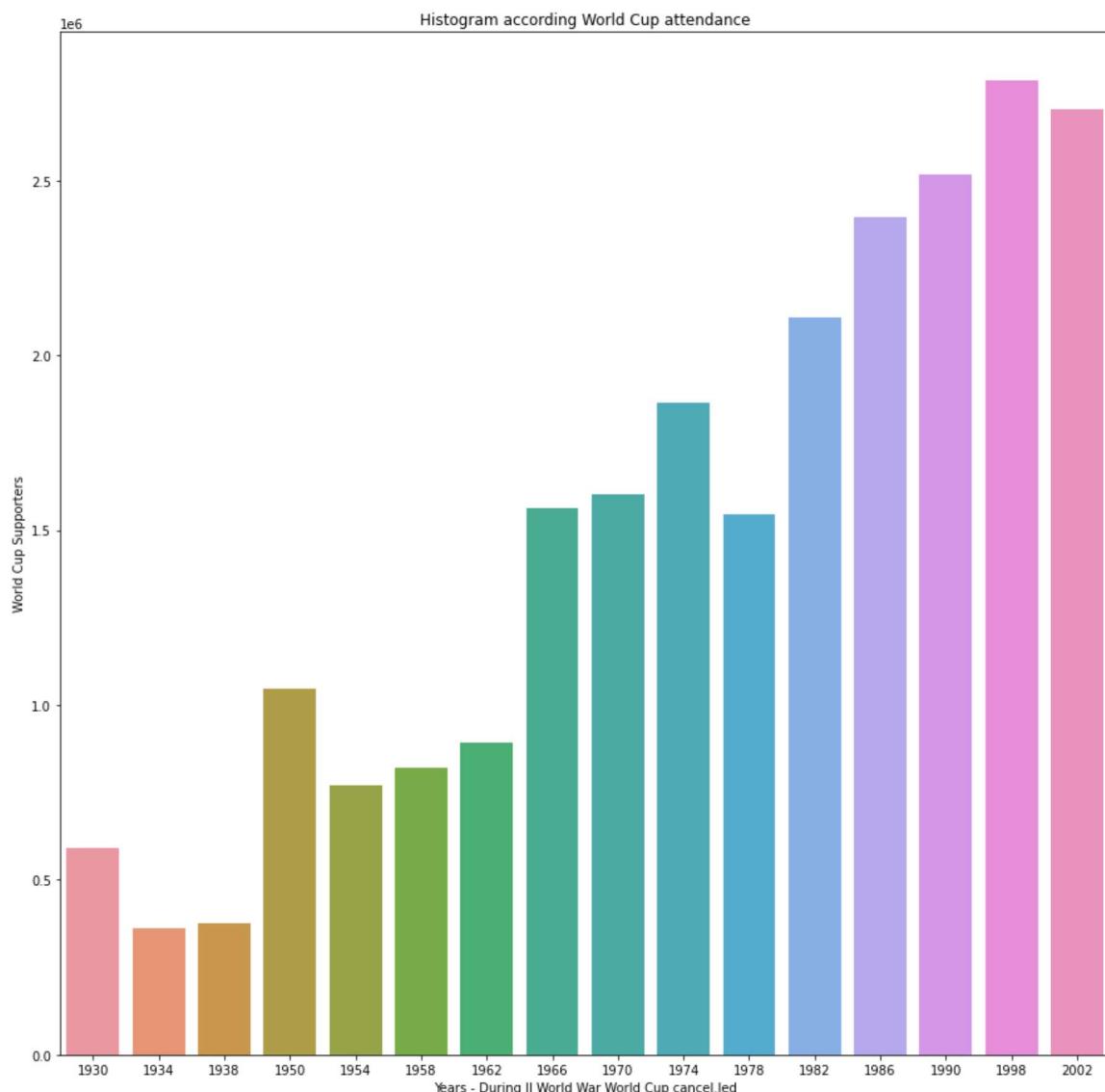
In [210]:

```
#World Cup - grafic sobre assistència (sense outliers).
import matplotlib.pyplot as plt

fig = plt.figure(figsize=(15,15))

wcup_df['Attendance'].dropna() #Esborrem els valors Nan

sns.barplot(data=wcup_df, x="Year", y="Attendance")
plt.xlabel("Years - During II World War World Cup cancelled")
plt.ylabel("World Cup Supporters")
plt.title("Histogram according World Cup attendance")
plt.show()
```



In [223]:

```
wcup_df['GoalsScored']
```

Out[223]:

```
0      70
1      70
2      84
3      88
4     140
5     126
6      89
7      89
8      95
9      97
10     102
11     146
12     132
13     115
14     171
15     161
Name: GoalsScored, dtype: int64
```

In []:

```
# Libraries - Normalization
from sklearn import preprocessing
import numpy as np
```

In [240]:

```
#Normalization of Atributes: Year
a_year = np.array([wcup_df['Year']])
a_year.reshape(1, -1)
normalized_arr = preprocessing.normalize(a_year)
print(normalized_arr)
```

```
[[0.24528286 0.24579122 0.24629957 0.24782465 0.24833301 0.24884137
 0.24934972 0.24985808 0.25036644 0.2508748 0.25138316 0.25189152
 0.25239987 0.25290823 0.25392495 0.25443331]]
```

In []:

```
#Atribut Normalitzat --> Tots els valors es troben entre 0 i 1.
```

In [236]:

```
#Normalization of Atributes: GoalsScored
a_goals = np.array([wcup_df['GoalsScored']])
a_goals.reshape(1, -1)
normalized_arr = preprocessing.normalize(a_goals)
print(normalized_arr)
```

```
[[0.15207225 0.15207225 0.1824867 0.19117655 0.30414451 0.27373006
 0.19334901 0.19334901 0.20638377 0.21072869 0.221591 0.31717927
 0.28676482 0.24983299 0.37149079 0.34976618]]
```

In []:

```
#Atribut Normalitzat --> Tots els valors es troben entre 0 i 1.
```

In [238]:

```
#Normalization of Atributes: QualifiedTeams
a_qteams = np.array([wcup_df['QualifiedTeams']])
a_qteams.reshape(1, -1)
normalized_arr = preprocessing.normalize(a_qteams)
print(normalized_arr)
```

```
[[0.16266529 0.20020343 0.18769072 0.16266529 0.20020343 0.20020343
 0.20020343 0.20020343 0.20020343 0.20020343 0.20020343 0.30030515
 0.30030515 0.30030515 0.40040687 0.40040687]]
```

In []:

```
#Atribut Normalitzat --> Tots els valors es troben entre 0 i 1.
```

In []:

```
#Normalization of Atributes: MatchesPlayed
a_matches = np.array([wcup_df['MatchesPlayed']])
a_matches.reshape(1, -1)
normalized_arr = preprocessing.normalize(a_matches)
print(normalized_arr)
```

In []:

```
#Atribut Normalitzat --> Tots els valors es troben entre 0 i 1.
```

In [239]:

```
#Normalization of Atributes: Attendance
a_attend = np.array([wcup_df['Attendance']])
a_attend.reshape(1, -1)
normalized_arr = preprocessing.normalize(a_attend)
print(normalized_arr)
```

```
[[0.08683608 0.0533766 0.05524405 0.15369609 0.11301826 0.1205473
 0.13133467 0.22984803 0.23585326 0.27434588 0.22729771 0.31021996
 0.35202545 0.36999175 0.4095294 0.39778023]]
```

In []:

```
#Atribut Normalitzat --> Tots els valors es troben entre 0 i 1.
```

In []:

```
#2. Dataframe: wcupmatch_df
```

In [315]:

```
wcupmatch_df.head()
```

Out[315]:

	Year	Datetime	Stage	Stadium	City	Home Team Name	Home Team Goals	Away Team Goals	Away Team Name	Win condition
0	1930	13 Jul 1930 - 15:00	Group 1	Pocitos	Montevideo	France	4	1	Mexico	
1	1930	13 Jul 1930 - 15:00	Group 4	Parque Central	Montevideo	USA	3	0	Belgium	
2	1930	14 Jul 1930 - 12:45	Group 2	Parque Central	Montevideo	Yugoslavia	2	1	Brazil	
3	1930	14 Jul 1930 - 14:50	Group 3	Pocitos	Montevideo	Romania	3	1	Peru	
4	1930	15 Jul 1930 - 16:00	Group 1	Parque Central	Montevideo	Argentina	1	0	France	

In [282]:

```
wcupmatch_df.tail()
```

Out[282]:

In [302]:

```
#Detectem una quantitat molt elevada de valors NaN que procedim a esborrar del datafram e.  
wcupmatch_df = wcupmatch_df.dropna()
```

Out[302]:

	Year	Datetime	Stage	Stadium	City	Home Team Name	Home Team Goals	Away Team Goals	Away Team Name
0	1930	13 Jul 1930 - 15:00	Group 1	Pocitos	Montevideo	France	4	1	Mexico
1	1930	13 Jul 1930 - 15:00	Group 4	Parque Central	Montevideo	USA	3	0	Belgium
2	1930	14 Jul 1930 - 12:45	Group 2	Parque Central	Montevideo	Yugoslavia	2	1	Brazil
3	1930	14 Jul 1930 - 14:50	Group 3	Pocitos	Montevideo	Romania	3	1	Peru
4	1930	15 Jul 1930 - 16:00	Group 1	Parque Central	Montevideo	Argentina	1	0	France
...
847	2014	05 Jul 2014 - 17:00	Quarter-finals	Arena Fonte Nova	Salvador	Netherlands	0	0	Costa Rica
848	2014	08 Jul 2014 - 17:00	Semi-finals	Estadio Mineirao	Belo Horizonte	Brazil	1	7	Germany
849	2014	09 Jul 2014 - 17:00	Semi-finals	Arena de Sao Paulo	Sao Paulo	Netherlands	0	0	Argentina
850	2014	12 Jul 2014 - 17:00	Play-off for third place	Estadio Nacional	Brasilia	Brazil	0	3	Netherlands
851	2014	13 Jul 2014 - 16:00	Final	Estadio do Maracana	Rio De Janeiro	Germany	1	0	Argentina

850 rows × 20 columns

In []:

```
#S'ha reduït la col. Lecció de dades de 4571 a 850 instàncies.
```

In [290]:

```
#Procedim a la conversió de valors numèrics en coma flotant a tipus numèric int64
#--Atribut Year
wcupmatch_df['Year'] = wcupmatch_df['Year'].astype("Int64")
```

In [294]:

```
#--Atribut Home Team Goals
wcupmatch_df['Home Team Goals'] = wcupmatch_df['Home Team Goals'].astype("Int64")
```

In [297]:

```
#--Atribut Away Team Goals
wcupmatch_df['Away Team Goals'] = wcupmatch_df['Away Team Goals'].astype("Int64")
```

In []:

```
#--Atribut Half-time Home Goals
wcupmatch_df['Half-time Home Goals'] = wcupmatch_df['Half-time Home Goals'].astype("Int64")
```

In [298]:

```
#--Atribut Half-time Away Goals
wcupmatch_df['Half-time Away Goals'] = wcupmatch_df['Half-time Away Goals'].astype("Int64")
```

In [308]:

```
#--Atribut Attendance
wcupmatch_df['Attendance'] = wcupmatch_df['Attendance'].astype("Int64")
```

In [310]:

```
wcupmatch_df
```

Out[310]:

	Year	Datetime	Stage	Stadium	City	Home Team Name	Home Team Goals	Away Team Goals	Away Team Name	condi
0	1930	13 Jul 1930 - 15:00	Group 1	Pocitos	Montevideo	France	4	1	Mexico	
1	1930	13 Jul 1930 - 15:00	Group 4	Parque Central	Montevideo	USA	3	0	Belgium	
2	1930	14 Jul 1930 - 12:45	Group 2	Parque Central	Montevideo	Yugoslavia	2	1	Brazil	
3	1930	14 Jul 1930 - 14:50	Group 3	Pocitos	Montevideo	Romania	3	1	Peru	
4	1930	15 Jul 1930 - 16:00	Group 1	Parque Central	Montevideo	Argentina	1	0	France	
...
4567	<NA>	NaN	NaN	NaN	NaN	NaN	<NA>	<NA>	NaN	
4568	<NA>	NaN	NaN	NaN	NaN	NaN	<NA>	<NA>	NaN	
4569	<NA>	NaN	NaN	NaN	NaN	NaN	<NA>	<NA>	NaN	
4570	<NA>	NaN	NaN	NaN	NaN	NaN	<NA>	<NA>	NaN	
4571	<NA>	NaN	NaN	NaN	NaN	NaN	<NA>	<NA>	NaN	

4572 rows × 20 columns

In [314]:

```
wcupmatch_df
```

Out[314]:

	Year	Datetime	Stage	Stadium	City	Home Team Name	Home Team Goals	Away Team Goals	Away Team Name
0	1930	13 Jul 1930 - 15:00	Group 1	Pocitos	Montevideo	France	4	1	Mexico
1	1930	13 Jul 1930 - 15:00	Group 4	Parque Central	Montevideo	USA	3	0	Belgium
2	1930	14 Jul 1930 - 12:45	Group 2	Parque Central	Montevideo	Yugoslavia	2	1	Brazil
3	1930	14 Jul 1930 - 14:50	Group 3	Pocitos	Montevideo	Romania	3	1	Peru
4	1930	15 Jul 1930 - 16:00	Group 1	Parque Central	Montevideo	Argentina	1	0	France
...
847	2014	05 Jul 2014 - 17:00	Quarter-finals	Arena Fonte Nova	Salvador	Netherlands	0	0	Costa Rica
848	2014	08 Jul 2014 - 17:00	Semi-finals	Estadio Mineirao	Belo Horizonte	Brazil	1	7	Germany
849	2014	09 Jul 2014 - 17:00	Semi-finals	Arena de Sao Paulo	Sao Paulo	Netherlands	0	0	Argentina
850	2014	12 Jul 2014 - 17:00	Play-off for third place	Estadio Nacional	Brasilia	Brazil	0	3	Netherlands
851	2014	13 Jul 2014 - 16:00	Final	Estadio do Maracana	Rio De Janeiro	Germany	1	0	Argentina

850 rows × 20 columns

In []:

In []:

```
# Libraries - Normalization
from sklearn import preprocessing
import numpy as np
```

In [318]:

```
#Normalization of Atributes: Year
a_year = np.array([wcupmatch_df['Year']])
a_year.reshape(1, -1)
normalized_arr = preprocessing.normalize(a_year)
print(normalized_arr)
```



```
0.03472912 0.03472912 0.03472912 0.03472912 0.03472912 0.03472912  
0.03472912 0.03472912 0.03472912 0.03472912 0.03472912 0.03472912  
0.03472912 0.03472912 0.03472912 0.03472912 0.03472912 0.03472912  
0.03472912 0.03472912 0.03472912 0.03472912 0.03472912 0.03472912  
0.03472912 0.03472912 0.03472912 0.03472912 0.03472912 0.03472912  
0.03472912 0.03472912 0.03472912 0.03472912 0.03472912 0.03472912  
0.03472912 0.03472912 0.03472912 0.03472912 0.03472912 0.03472912  
0.03472912 0.03472912 0.03472912 0.03472912 0.03472912 0.03472912  
0.03479823 0.03479823 0.03479823 0.03479823 0.03479823 0.03479823  
0.03479823 0.03479823 0.03479823 0.03479823 0.03479823 0.03479823  
0.03479823 0.03479823 0.03479823 0.03479823 0.03479823 0.03479823  
0.03479823 0.03479823 0.03479823 0.03479823 0.03479823 0.03479823  
0.03479823 0.03479823 0.03479823 0.03479823 0.03479823 0.03479823  
0.03479823 0.03479823 0.03479823 0.03479823 0.03479823 0.03479823  
0.03479823 0.03479823 0.03479823 0.03479823 0.03479823 0.03479823  
0.03479823 0.03479823 0.03479823 0.03479823 0.03479823 0.03479823  
0.03479823 0.03479823 0.03479823 0.03479823 0.03479823 0.03479823
```

In []:

```
#Atribut Normalitzat --> Tots els valors es troben entre 0 i 1.
```

In [320]:

```
#Normalization of Atributes: Home Team Goals
a_htgoals = np.array([wcupmatch_df['Home Team Goals']])
a_year.reshape(1, -1)
normalized_arr = preprocessing.normalize(a_htgoals)
print(normalized_arr)
```

[[0.05660818 0.04245614 0.02830409 0.04245614 0.01415205 0.04245614
0.05660818 0.04245614 0.01415205 0.01415205 0.08491227 0.05660818
0.01415205 0.05660818 0.04245614 0.08491227 0.08491227 0.05660818
0.04245614 0.05660818 0.04245614 0.04245614 0.07076023 0.04245614
0.09906432 0.02830409 0.04245614 0.02830409 0.01415205 0.02830409
0.01415205 0.01415205 0.04245614 0.04245614 0.02830409 0.01415205
0.08491227 0.04245614 0.04245614 0.02830409 0.08491227 0.04245614
0.02830409 0.05660818 0.01415205 0.02830409 0.11321636 0.04245614
0.02830409 0.07076023 0.02830409 0.05660818 0.05660818 0.05660818
0.02830409 0.04245614 0.04245614 0.04245614 0.02830409 0.05660818
0.02830409 0.02830409 0.01415205 0.02830409 0.01415205 0.02830409
0.11321636 0.02830409 0.07076023 0.02830409 0.09906432 0.08491227
0.04245614 0.04245614 0.02830409 0.02830409 0.01415205 0.07076023
0.01415205 0.05660818 0.12736841 0.05660818 0.02830409 0.09906432
0.07076023 0.04245614 0.01415205 0.11321636 0.09906432 0.02830409
0.05660818 0.09906432 0.05660818 0.05660818 0.09906432 0.02830409
0.05660818 0.08491227 0.05660818 0.04245614 0.04245614 0.04245614
0.02830409 0.01415205 0.01415205 0.09906432 0.01415205 0.04245614
0.01415205 0. 0.01415205 0.04245614 0.04245614 0.02830409
0.02830409 0.04245614 0.02830409 0. 0.02830409 0.02830409
0.05660818 0.04245614 0.02830409 0.02830409 0.08491227 0.01415205
0.02830409 0.02830409 0.01415205 0.01415205 0.02830409 0.05660818
0.04245614 0.07076023 0.08491227 0.07076023 0.02830409 0.02830409
0.01415205 0.04245614 0.02830409 0.01415205 0.02830409 0.
0.04245614 0. 0.04245614 0.02830409 0.05660818 0.01415205
0.08491227 0.02830409 0.02830409 0.02830409 0. 0.02830409
0.07076023 0.04245614 0. 0.04245614 0.02830409 0.04245614
0.01415205 0.01415205 0.04245614 0.05660818 0.01415205 0.04245614
0. 0.07076023 0.02830409 0.04245614 0.01415205 0.04245614
0.02830409 0.02830409 0.02830409 0.02830409 0.04245614 0.01415205
0.04245614 0. 0.01415205 0.02830409 0. 0.02830409
0.04245614 0.01415205 0.02830409 0.04245614 0.02830409 0.02830409
0.01415205 0.05660818 0.02830409 0.07076023 0.02830409 0.02830409
0.02830409 0.05660818 0. 0.02830409 0.04245614 0.01415205
0.01415205 0.02830409 0.05660818 0.04245614 0. 0.04245614
0.02830409 0.05660818 0.01415205 0.05660818 0. 0.01415205
0.02830409 0. 0. 0.04245614 0.04245614
0. 0. 0.01415205 0.12736841 0. 0.01415205
0. 0.01415205 0.01415205 0. 0. 0.01415205
0.01415205 0.05660818 0.04245614 0.02830409 0. 0.01415205
0.05660818 0. 0.01415205 0.02830409 0. 0.05660818
0. 0.02830409 0.01415205 0.02830409 0. 0.01415205
0. 0.02830409 0.04245614 0.02830409 0.01415205 0.02830409
0.04245614 0.04245614 0.04245614 0.01415205 0.08491227 0.02830409
0. 0.01415205 0.01415205 0. 0.04245614 0.04245614
0. 0.01415205 0.01415205 0.01415205 0.05660818 0.04245614
0. 0.07076023 0.04245614 0.02830409 0.01415205 0.01415205
0.02830409 0. 0.02830409 0.04245614 0.04245614 0.08491227
0.02830409 0.04245614 0. 0. 0.02830409 0.
0.14152046 0.07076023 0.01415205 0.04245614 0.01415205 0.
0.01415205 0. 0.01415205 0.05660818 0.05660818 0.
0.01415205 0.04245614 0.05660818 0.02830409 0.02830409 0.
0.05660818 0.01415205 0.07076023 0.01415205 0.02830409 0.01415205
0.02830409 0.05660818 0.04245614 0.01415205 0. 0.01415205
0.01415205 0.01415205 0. 0.04245614 0.02830409 0.
0.02830409 0. 0.01415205 0.02830409 0.05660818 0.
0.04245614 0. 0. 0.04245614 0.04245614 0.04245614
0.01415205 0. 0. 0.08491227 0.04245614 0.]

0.01415205 0.01415205 0.01415205 0.01415205 0. . 0.01415205
0.01415205 0.01415205 0.01415205 0.02830409 0.01415205 0.
0.01415205 0.01415205 0.01415205 0.01415205 0.08491227 0.02830409
0. . 0.02830409 0.02830409 0.02830409 0.02830409 0.
0.01415205 0.04245614 0. . 0. . 0.02830409
0.04245614 0.02830409 0.01415205 0.05660818 0. . 0.
0.04245614 0.01415205 0.01415205 0. . 0.01415205 0.02830409
0. . 0.02830409 0.05660818 0.04245614 0. . 0.
0. . 0.01415205 0.01415205 0.02830409 0.05660818 0.01415205
0.01415205 0.02830409 0.01415205 0. . 0.02830409 0.02830409
0.01415205 0.01415205 0. . 0.07076023 0.01415205 0.01415205
0. . 0. . 0.01415205 0.04245614 0.01415205 0.
0.01415205 0.05660818 0.02830409 0.02830409 0.01415205 0.01415205
0. . 0.01415205 0.01415205 0.01415205 0.02830409 0.05660818
0. . 0.02830409 0. . 0.02830409 0.01415205 0.01415205
0. . 0.01415205 0.01415205 0.04245614 0.01415205 0.01415205
0.02830409 0.01415205 0.02830409 0.01415205 0.01415205 0.
0.01415205 0.01415205 0.01415205 0.02830409 0.02830409 0.02830409
0.05660818 0.04245614 0.01415205 0.01415205 0.02830409 0.01415205
0. . 0.02830409 0.04245614 0.04245614 0.01415205 0.02830409
0.02830409 0.05660818 0. . 0. . 0.04245614 0.01415205
0. . 0.01415205 0.01415205 0.08491227 0.01415205 0.
0. . 0. . 0.04245614 0.04245614 0.01415205 0.04245614
0.02830409 0.01415205 0.01415205 0.01415205 0.02830409 0.02830409
0.02830409 0.02830409 0.01415205 0. . 0.05660818 0.
0.02830409 0.02830409 0.02830409 0.01415205 0. . 0.
0.04245614 0.02830409 0.01415205 0. . 0.01415205 0.01415205
0.01415205 0.02830409 0.01415205 0.02830409 0.01415205 0.04245614
0.01415205 0.04245614 0.01415205 0.05660818 0.01415205 0.
0. . 0.02830409 0.07076023 0.02830409 0.07076023 0.01415205
0.01415205 0.02830409 0.02830409 0.01415205 0. . 0.01415205
0.02830409 0.02830409 0.08491227 0.01415205 0.01415205 0.02830409
0.02830409 0. . 0.01415205 0.01415205 0.01415205 0.
0.01415205 0.05660818 0.01415205 0.01415205 0.02830409 0.02830409
0. . 0.02830409 0. . 0.04245614 0.02830409 0.
0.01415205 0.02830409 0.01415205 0. . 0. . 0.01415205
0.01415205 0.11321636 0.01415205 0.02830409 0.01415205 0.04245614
0.02830409 0.02830409 0. . 0. . 0.02830409 0.02830409
0.02830409 0.04245614 0.01415205 0.01415205 0.01415205 0.
0.02830409 0. . 0.04245614 0.01415205 0.05660818 0.01415205
0.01415205 0.01415205 0.02830409 0.01415205 0.05660818 0.01415205
0.02830409 0.04245614 0. . 0. . 0.01415205 0.
0.01415205 0.02830409 0.02830409 0.04245614 0.01415205 0.01415205
0. . 0. . 0.04245614 0.04245614 0. . 0.01415205
0.01415205 0.01415205 0.02830409 0. . 0. . 0.02830409
0.01415205 0.01415205 0. . 0. . 0.01415205 0.01415205
0.02830409 0. . 0.05660818 0. . 0.01415205 0.
0.02830409 0. . 0.04245614 0. . 0.04245614 0.
0.02830409 0.02830409 0. . 0.01415205 0.05660818 0.02830409
0.01415205 0.04245614 0.02830409 0.01415205 0.08491227 0.02830409
0. . 0.02830409 0. . 0.01415205 0. . 0.02830409
0.01415205 0. . 0. . 0.04245614 0. . 0.01415205
0.02830409 0.02830409 0.01415205 0.02830409 0. . 0.04245614
0. . 0.02830409 0.01415205 0.02830409 0.01415205 0.01415205
0.01415205 0. . 0.04245614 0.01415205 0.01415205 0.04245614
0. . 0. . 0. . 0.04245614 0.01415205
0.01415205 0. . 0.02830409 0.01415205 0.01415205 0.
0. . 0.05660818 0.02830409 0.01415205 0.01415205 0.01415205
0. . 0.02830409 0. . 0. . 0.05660818
0.02830409 0. . 0.02830409 0. . 0.02830409 0.01415205

```
0.01415205 0.01415205 0.          0.01415205 0.04245614 0.09906432
0.01415205 0.02830409 0.          0.01415205 0.02830409 0.
0.          0.01415205 0.          0.02830409 0.04245614 0.
0.01415205 0.01415205 0.          0.          0.01415205 0.
0.02830409 0.01415205 0.05660818 0.04245614 0.02830409 0.04245614
0.          0.01415205 0.02830409 0.01415205 0.          0.
0.02830409 0.          0.02830409 0.          0.04245614 0.01415205
0.01415205 0.04245614 0.04245614 0.01415205 0.01415205 0.02830409
0.02830409 0.04245614 0.02830409 0.05660818 0.          0.01415205
0.02830409 0.          0.01415205 0.02830409 0.          0.
0.02830409 0.02830409 0.          0.          0.02830409 0.01415205
0.01415205 0.02830409 0.01415205 0.01415205 0.02830409 0.02830409
0.          0.02830409 0.01415205 0.01415205 0.          0.
0.01415205 0.02830409 0.02830409 0.04245614 0.          0.
0.          0.02830409 0.          0.01415205 0.01415205 0.02830409
0.02830409 0.02830409 0.          0.01415205 0.          0.01415205
0.          0.          0.01415205 0.02830409 0.01415205 0.01415205
0.02830409 0.01415205 0.02830409 0.02830409 0.01415205 0.02830409
0.01415205 0.02830409 0.          0.02830409 0.01415205 0.
0.01415205 0.          0.          0.01415205]]
```

In []:

```
#Atribut Normalitzat --> Tots els valors es troben entre 0 i 1.
```

In [321]:

```
#Normalization of Atributes: Away Team Goals
a_atgoals = np.array([wcupmatch_df['Away Team Goals']])
a_atgoals.reshape(1, -1)
normalized_arr = preprocessing.normalize(a_atgoals)
print(normalized_arr)
```

```

[[0.02297182 0.          0.02297182 0.02297182 0.          0.
  0.          0.          0.          0.          0.06891546 0.
  0.          0.          0.02297182 0.02297182 0.02297182 0.04594364
  0.04594364 0.04594364 0.04594364 0.04594364 0.04594364 0.02297182
  0.02297182 0.02297182 0.04594364 0.02297182 0.02297182 0.02297182
  0.          0.          0.02297182 0.04594364 0.02297182 0.02297182
  0.          0.02297182 0.06891546 0.02297182 0.1148591 0.
  0.02297182 0.04594364 0.02297182 0.          0.          0.02297182
  0.02297182 0.02297182 0.02297182 0.04594364 0.04594364 0.
  0.          0.02297182 0.04594364 0.          0.04594364 0.02297182
  0.          0.04594364 0.          0.          0.          0.
  0.          0.02297182 0.04594364 0.04594364 0.02297182 0.02297182
  0.04594364 0.02297182 0.02297182 0.          0.          0.
  0.          0.02297182 0.          0.09188728 0.02297182 0.
  0.          0.04594364 0.02297182 0.06891546 0.          0.
  0.02297182 0.04594364 0.02297182 0.04594364 0.1148591 0.
  0.04594364 0.02297182 0.04594364 0.02297182 0.04594364 0.
  0.04594364 0.06891546 0.02297182 0.06891546 0.02297182 0.
  0.          0.          0.02297182 0.04594364 0.04594364 0.
  0.04594364 0.02297182 0.02297182 0.          0.          0.04594364
  0.          0.06891546 0.02297182 0.04594364 0.02297182 0.
  0.02297182 0.02297182 0.          0.          0.          0.
  0.02297182 0.04594364 0.06891546 0.04594364 0.02297182 0.
  0.          0.02297182 0.          0.02297182 0.          0.
  0.02297182 0.          0.          0.09188728 0.
  0.02297182 0.02297182 0.02297182 0.02297182 0.          0.
  0.          0.02297182 0.          0.          0.02297182 0.02297182
  0.          0.          0.02297182 0.04594364 0.          0.02297182
  0.          0.          0.          0.          0.02297182 0.02297182
  0.02297182 0.          0.02297182 0.02297182 0.02297182 0.02297182
  0.          0.          0.          0.          0.          0.
  0.02297182 0.          0.          0.02297182 0.02297182 0.02297182
  0.          0.          0.          0.          0.          0.
  0.02297182 0.04594364 0.          0.          0.04594364 0.
  0.          0.02297182 0.02297182 0.          0.          0.
  0.02297182 0.02297182 0.02297182 0.04594364 0.          0.
  0.          0.02297182 0.04594364 0.          0.02297182 0.02297182
  0.          0.06891546 0.          0.02297182 0.          0.
  0.          0.04594364 0.04594364 0.          0.02297182 0.04594364
  0.06891546 0.          0.02297182 0.          0.          0.02297182
  0.16080274 0.02297182 0.02297182 0.          0.06891546 0.
  0.09188728 0.02297182 0.          0.02297182 0.04594364 0.
  0.          0.02297182 0.04594364 0.02297182 0.04594364 0.04594364
  0.02297182 0.          0.02297182 0.02297182 0.02297182 0.04594364
  0.          0.02297182 0.02297182 0.02297182 0.02297182 0.02297182
  0.02297182 0.          0.02297182 0.          0.          0.02297182
  0.          0.          0.02297182 0.          0.02297182 0.02297182
  0.          0.          0.          0.          0.02297182 0.04594364
  0.          0.02297182 0.          0.          0.          0.
  0.04594364 0.          0.02297182 0.04594364 0.02297182 0.
  0.02297182 0.02297182 0.02297182 0.          0.02297182 0.
  0.02297182 0.04594364 0.04594364 0.02297182 0.02297182 0.02297182
  0.02297182 0.          0.02297182 0.02297182 0.02297182 0.
  0.          0.          0.02297182 0.          0.02297182 0.04594364
  0.02297182 0.02297182 0.02297182 0.02297182 0.04594364 0.02297182
  0.          0.          0.04594364 0.02297182 0.02297182 0.
  0.          0.          0.02297182 0.          0.02297182 0.
  0.04594364 0.02297182 0.06891546 0.02297182 0.02297182 0.
  0.04594364 0.          0.04594364 0.06891546 0.04594364 0.02297182
  0.02297182 0.02297182 0.02297182 0.          0.02297182 0.

```

0.02297182	0.04594364	0.	0.	0.02297182	0.02297182
0.02297182	0.02297182	0.02297182	0.	0.	0.
0.04594364	0.02297182	0.	0.04594364	0.02297182	0.02297182
0.06891546	0.	0.06891546	0.	0.04594364	0.02297182
0.06891546	0.	0.06891546	0.06891546	0.	0.
0.09188728	0.	0.	0.	0.04594364	0.02297182
0.	0.1148591	0.02297182	0.	0.02297182	0.02297182
0.04594364	0.	0.04594364	0.04594364	0.02297182	0.04594364
0.04594364	0.	0.1148591	0.02297182	0.02297182	0.
0.02297182	0.	0.02297182	0.	0.	0.02297182
0.	0.	0.02297182	0.02297182	0.	0.04594364
0.	0.	0.06891546	0.02297182	0.02297182	0.09188728
0.02297182	0.02297182	0.	0.02297182	0.	0.04594364
0.02297182	0.04594364	0.02297182	0.	0.02297182	0.02297182
0.02297182	0.02297182	0.	0.	0.04594364	0.
0.	0.	0.	0.04594364	0.02297182	0.02297182
0.02297182	0.	0.04594364	0.	0.02297182	0.02297182
0.06891546	0.	0.	0.04594364	0.02297182	0.
0.	0.	0.02297182	0.09188728	0.02297182	0.
0.	0.02297182	0.02297182	0.	0.	0.02297182
0.02297182	0.	0.02297182	0.04594364	0.04594364	0.06891546
0.	0.02297182	0.02297182	0.02297182	0.04594364	0.02297182
0.04594364	0.04594364	0.04594364	0.	0.06891546	0.04594364
0.	0.	0.04594364	0.02297182	0.02297182	0.06891546
0.02297182	0.04594364	0.04594364	0.02297182	0.	0.
0.02297182	0.04594364	0.04594364	0.02297182	0.	0.02297182
0.	0.06891546	0.06891546	0.	0.	0.
0.02297182	0.02297182	0.	0.	0.02297182	0.
0.02297182	0.04594364	0.	0.04594364	0.	0.04594364
0.	0.02297182	0.02297182	0.02297182	0.06891546	0.04594364
0.02297182	0.04594364	0.02297182	0.06891546	0.02297182	0.04594364
0.	0.02297182	0.04594364	0.	0.02297182	0.02297182
0.02297182	0.	0.02297182	0.	0.	0.
0.	0.02297182	0.02297182	0.02297182	0.06891546	0.04594364
0.02297182	0.04594364	0.	0.02297182	0.	0.02297182
0.	0.02297182	0.02297182	0.	0.	0.02297182
0.	0.	0.06891546	0.06891546	0.04594364	0.02297182
0.06891546	0.06891546	0.1148591	0.	0.	0.02297182
0.04594364	0.02297182	0.02297182	0.04594364	0.06891546	0.
0.02297182	0.04594364	0.	0.04594364	0.02297182	0.02297182
0.04594364	0.	0.02297182	0.	0.	0.
0.06891546	0.04594364	0.04594364	0.04594364	0.	0.
0.02297182	0.02297182	0.02297182	0.02297182	0.02297182	0.06891546
0.	0.02297182	0.	0.	0.	0.04594364
0.	0.	0.	0.	0.	0.02297182
0.	0.	0.04594364	0.02297182	0.	0.
0.02297182	0.04594364	0.09188728	0.02297182	0.06891546	0.04594364
0.04594364	0.	0.02297182	0.02297182	0.	0.04594364
0.04594364	0.02297182	0.09188728	0.04594364	0.	0.02297182
0.04594364	0.	0.	0.02297182	0.	0.
0.	0.	0.	0.06891546	0.02297182	0.
0.	0.02297182	0.04594364	0.02297182	0.02297182	0.02297182
0.02297182	0.	0.	0.	0.02297182	0.02297182
0.02297182	0.	0.	0.	0.02297182	0.02297182
0.	0.02297182	0.02297182	0.02297182	0.06891546	0.02297182
0.02297182	0.04594364	0.02297182	0.04594364	0.	0.

```
0.02297182 0.04594364 0.04594364 0.02297182 0.02297182 0.  
0.          0.          0.02297182 0.04594364 0.04594364 0.04594364  
0.02297182 0.          0.02297182 0.02297182 0.04594364 0.  
0.06891546 0.04594364 0.          0.06891546 0.04594364 0.  
0.02297182 0.04594364 0.02297182 0.02297182 0.02297182 0.  
0.          0.          0.02297182 0.02297182 0.09188728 0.02297182  
0.06891546 0.02297182 0.06891546 0.02297182 0.02297182 0.  
0.1148591  0.02297182 0.          0.06891546 0.04594364 0.02297182  
0.02297182 0.          0.02297182 0.          0.          0.04594364  
0.02297182 0.          0.02297182 0.06891546 0.04594364 0.09188728  
0.02297182 0.02297182 0.          0.02297182 0.1148591  0.04594364  
0.          0.04594364 0.          0.          0.09188728 0.04594364  
0.06891546 0.          0.09188728 0.06891546 0.02297182 0.  
0.09188728 0.02297182 0.06891546 0.02297182 0.06891546 0.  
0.02297182 0.02297182 0.02297182 0.02297182 0.02297182 0.  
0.          0.02297182 0.02297182 0.16080274 0.06891546 0.  
0.          0.          0.02297182 0.02297182 0.          0.  
0.02297182 0.02297182 0.          0.02297182 0.02297182 0.  
0.          0.02297182 0.02297182 0.02297182 0.          0.  
0.16080274 0.          0.06891546 0.          ]]
```

In []:

```
#Atribut Normalitzat --> Tots els valors es troben entre 0 i 1.
```

In [322]:

```
#Normalization of Atributes: Half-time Home Goals
a_hthgoals = np.array([wcupmatch_df['Half-time Home Goals']])
a_hthgoals.reshape(1, -1)
normalized_arr = preprocessing.normalize(a_hthgoals)
print(normalized_arr)
```

[0.08748178 0.05832118 0.05832118 0.02916059 0. 0.02916059
0. 0.05832118 0. 0. 0.08748178 0.02916059
0.02916059 0.11664237 0.05832118 0.02916059 0.08748178 0.02916059
0. 0.05832118 0.05832118 0.02916059 0.02916059 0.08748178
0.08748178 0. 0.02916059 0. 0. 0.02916059
0.02916059 0.02916059 0.02916059 0.08748178 0. 0.
0.11664237 0.05832118 0. 0. 0. 0.
0. 0.02916059 0. 0.02916059 0.11664237 0.02916059
0. 0.08748178 0. 0. 0.02916059 0.08748178 0.02916059
0.02916059 0. 0.05832118 0. 0.05832118 0.05832118
0.05832118 0.05832118 0.02916059 0.02916059 0. 0.02916059
0.11664237 0.05832118 0.05832118 0.02916059 0.08748178 0.08748178
0.02916059 0.05832118 0. 0. 0.02916059 0.11664237
0.02916059 0.02916059 0.11664237 0. 0. 0.02916059 0.05832118
0.11664237 0.02916059 0. 0.08748178 0.11664237 0.02916059
0.02916059 0.08748178 0.02916059 0.05832118 0.14580296 0.02916059
0.05832118 0.02916059 0. 0. 0.02916059 0.05832118 0.02916059
0.02916059 0.02916059 0.02916059 0.05832118 0.02916059 0.02916059
0.02916059 0. 0. 0.05832118 0.02916059 0.02916059
0. 0.02916059 0.02916059 0. 0. 0.02916059 0.02916059
0.02916059 0.02916059 0.05832118 0. 0.08748178 0.
0. 0. 0. 0.02916059 0. 0. 0.02916059
0.02916059 0.05832118 0.08748178 0.05832118 0. 0.
0.02916059 0.05832118 0.08748178 0.05832118 0. 0.
0.02916059 0.02916059 0. 0. 0.02916059 0. 0.
0.05832118 0. 0.05832118 0. 0.08748178 0.
0.11664237 0.02916059 0.02916059 0. 0. 0.02916059
0.05832118 0.05832118 0. 0. 0.02916059 0.05832118 0.02916059
0.02916059 0. 0. 0.05832118 0. 0. 0.02916059
0. 0.02916059 0.05832118 0. 0. 0.02916059 0.
0.05832118 0. 0. 0.02916059 0. 0. 0.
0.02916059 0.02916059 0.05832118 0.02916059 0.02916059
0. 0. 0. 0.05832118 0.02916059 0.02916059 0.
0.02916059 0.02916059 0. 0. 0.02916059 0.
0. 0.08748178 0.02916059 0.05832118 0. 0.
0. 0.02916059 0.05832118 0. 0. 0.02916059
0.05832118 0. 0. 0.02916059 0. 0. 0.
0.02916059 0. 0. 0.02916059 0.02916059 0. 0.02916059
0. 0. 0. 0. 0. 0. 0.05832118
0. 0. 0. 0. 0. 0. 0.
0. 0. 0. 0. 0.17496355 0. 0.
0. 0.02916059 0. 0. 0. 0. 0.
0. 0.05832118 0. 0. 0.05832118 0. 0.
0.05832118 0. 0. 0.02916059 0.02916059 0. 0.
0. 0. 0. 0.02916059 0.02916059 0. 0.02916059
0. 0. 0.02916059 0. 0. 0.02916059 0.02916059 0.02916059
0.02916059 0.02916059 0.05832118 0.02916059 0.11664237 0.02916059
0. 0.02916059 0.02916059 0. 0. 0.08748178 0.02916059
0. 0. 0.02916059 0. 0. 0.08748178 0.02916059
0. 0.08748178 0.05832118 0.02916059 0. 0.02916059
0.02916059 0. 0. 0. 0.02916059 0.05832118
0. 0. 0. 0. 0. 0. 0.
0.08748178 0.08748178 0. 0.02916059 0. 0.
0.02916059 0. 0.02916059 0.05832118 0.02916059 0.
0.02916059 0.02916059 0.02916059 0. 0.02916059 0.
0.05832118 0. 0. 0. 0. 0. 0.
0.02916059 0.05832118 0.08748178 0. 0. 0.02916059 0.02916059
0.02916059 0. 0. 0. 0.05832118 0. 0.
0. 0. 0. 0. 0. 0. 0.
0.05832118 0. 0. 0. 0. 0. 0.
0. 0. 0. 0. 0. 0. 0.]

0.	0.02916059	0.	0.02916059	0.	0.02916059
0.02916059	0.	0.	0.02916059	0.	0.
0.	0.02916059	0.	0.	0.05832118	0.02916059
0.	0.	0.	0.02916059	0.	0.
0.	0.08748178	0.	0.	0.	0.02916059
0.	0.02916059	0.02916059	0.02916059	0.	0.
0.02916059	0.02916059	0.	0.	0.	0.
0.	0.	0.	0.02916059	0.	0.
0.	0.	0.	0.02916059	0.05832118	0.
0.02916059	0.	0.	0.	0.02916059	0.
0.	0.02916059	0.	0.05832118	0.02916059	0.
0.	0.	0.02916059	0.05832118	0.	0.
0.	0.05832118	0.02916059	0.	0.	0.02916059
0.	0.02916059	0.	0.	0.	0.02916059
0.	0.	0.	0.	0.	0.
0.	0.02916059	0.02916059	0.	0.	0.
0.	0.	0.	0.	0.02916059	0.
0.	0.	0.	0.	0.	0.02916059
0.02916059	0.02916059	0.	0.02916059	0.	0.02916059
0.05832118	0.05832118	0.	0.02916059	0.02916059	0.
0.	0.02916059	0.02916059	0.02916059	0.	0.05832118
0.05832118	0.02916059	0.	0.	0.08748178	0.
0.	0.	0.	0.08748178	0.	0.
0.	0.	0.08748178	0.02916059	0.	0.05832118
0.05832118	0.	0.	0.	0.02916059	0.
0.	0.	0.	0.02916059	0.	0.11664237
0.02916059	0.02916059	0.02916059	0.	0.	0.
0.02916059	0.02916059	0.02916059	0.	0.	0.
0.02916059	0.02916059	0.02916059	0.	0.02916059	0.
0.02916059	0.02916059	0.02916059	0.	0.05832118	0.
0.	0.02916059	0.	0.02916059	0.02916059	0.
0.	0.02916059	0.05832118	0.	0.02916059	0.
0.	0.	0.	0.02916059	0.	0.
0.	0.02916059	0.	0.	0.02916059	0.
0.02916059	0.02916059	0.05832118	0.02916059	0.02916059	0.05832118
0.	0.	0.	0.02916059	0.	0.
0.02916059	0.08748178	0.	0.	0.	0.02916059
0.	0.	0.	0.05832118	0.02916059	0.
0.	0.	0.02916059	0.	0.	0.
0.	0.	0.	0.02916059	0.	0.
0.	0.	0.	0.11664237	0.	0.
0.	0.05832118	0.	0.	0.02916059	0.02916059
0.	0.08748178	0.02916059	0.02916059	0.	0.
0.02916059	0.	0.	0.02916059	0.08748178	0.
0.	0.	0.02916059	0.	0.02916059	0.02916059
0.02916059	0.08748178	0.	0.	0.	0.
0.02916059	0.02916059	0.02916059	0.05832118	0.	0.02916059
0.	0.	0.05832118	0.02916059	0.	0.
0.	0.	0.	0.	0.	0.
0.	0.	0.	0.	0.	0.
0.02916059	0.02916059	0.	0.	0.	0.
0.02916059	0.	0.05832118	0.	0.02916059	0.
0.	0.	0.02916059	0.	0.	0.
0.	0.	0.	0.02916059	0.	0.
0.	0.	0.	0.02916059	0.	0.
0.	0.	0.	0.05832118	0.	0.
0.02916059	0.	0.	0.02916059	0.05832118	0.02916059
0.	0.02916059	0.	0.	0.08748178	0.05832118
0.	0.	0.	0.02916059	0.	0.
0.02916059	0.	0.	0.	0.	0.02916059
0.	0.02916059	0.	0.05832118	0.	0.02916059
0.	0.	0.	0.02916059	0.	0.
0.	0.	0.	0.	0.	0.02916059
0.	0.	0.	0.	0.	0.
0.	0.	0.	0.	0.	0.
0.	0.	0.	0.	0.	0.02916059
0.	0.	0.	0.	0.	0.02916059
0.	0.	0.	0.	0.	0.
0.	0.	0.	0.	0.	0.05832118
0.02916059	0.	0.	0.	0.	0.

```
0.02916059 0.02916059 0.          0.02916059 0.02916059 0.02916059
0.          0.02916059 0.          0.          0.02916059 0.
0.          0.          0.          0.          0.02916059 0.
0.          0.          0.          0.          0.          0.
0.02916059 0.          0.05832118 0.05832118 0.02916059 0.05832118
0.          0.          0.          0.          0.          0.
0.02916059 0.          0.02916059 0.          0.02916059 0.
0.02916059 0.05832118 0.02916059 0.02916059 0.02916059 0.
0.          0.02916059 0.02916059 0.08748178 0.          0.
0.          0.          0.02916059 0.          0.
0.          0.02916059 0.          0.          0.          0.02916059
0.          0.          0.02916059 0.          0.          0.
0.          0.          0.02916059 0.          0.          0.
0.02916059 0.02916059 0.02916059 0.02916059 0.          0.
0.          0.02916059 0.          0.          0.          0.02916059
0.          0.02916059 0.          0.          0.          0.
0.          0.          0.02916059 0.          0.          0.
0.          0.          0.02916059 0.          0.          0.
0.          0.          0.          0.02916059 0.02916059 0.
0.          0.          0.          0.          0.          ]]
```

In []:

```
#Atribut Normalitzat --> Tots els valors es troben entre 0 i 1.
```

In [323]:

```
#Normalization of Atributes: Half-time Away Goals
a_htagoals = np.array([wcupmatch_df['Half-time Away Goals']])
a_htagoals.reshape(1, -1)
normalized_arr = preprocessing.normalize(a_htagoals)
print(normalized_arr)
```


0.04214498 0.08428995 0. 0. 0. 0.
0.04214498 0. 0.04214498 0. 0. 0.
0.08428995 0. 0. 0.08428995 0.04214498 0.04214498
0.04214498 0. 0.04214498 0. 0.04214498 0.
0.08428995 0. 0.08428995 0.04214498 0. 0.
0. 0. 0. 0. 0.04214498 0.
0. 0.04214498 0. 0. 0. 0.
0.04214498 0. 0. 0. 0. 0.04214498
0. 0. 0.08428995 0. 0. 0.
0. 0. 0. 0. 0. 0.
0. 0. 0.04214498 0. 0. 0.04214498
0. 0. 0.04214498 0. 0. 0.08428995
0. 0.04214498 0. 0. 0. 0.
0. 0.08428995 0.04214498 0. 0. 0.
0. 0. 0. 0. 0. 0.
0. 0. 0. 0. 0. 0.
0. 0. 0. 0. 0.04214498 0.04214498
0.08428995 0. 0. 0.04214498 0.04214498 0.
0. 0. 0.04214498 0.04214498 0. 0.
0. 0. 0.04214498 0. 0. 0.04214498
0.04214498 0. 0.04214498 0.04214498 0. 0.04214498
0. 0.04214498 0. 0. 0.04214498 0.04214498
0.04214498 0. 0.04214498 0. 0.04214498 0.04214498
0. 0. 0. 0. 0. 0.
0. 0. 0.08428995 0. 0. 0.
0.04214498 0.04214498 0.04214498 0. 0.
0. 0. 0.04214498 0.04214498 0. 0.04214498 0.08428995
0. 0. 0. 0.08428995 0. 0. 0.
0.04214498 0. 0. 0.04214498 0.04214498 0.04214498
0. 0. 0.08428995 0.08428995 0.04214498 0.04214498
0.04214498 0. 0. 0. 0. 0.
0.04214498 0. 0. 0. 0. 0.
0. 0. 0.04214498 0. 0. 0.
0. 0. 0.04214498 0. 0. 0.
0. 0. 0.04214498 0. 0. 0.
0. 0. 0.04214498 0. 0. 0.04214498
0. 0. 0.04214498 0. 0. 0.04214498
0. 0. 0. 0. 0. 0.
0. 0. 0.04214498 0.12643493 0. 0. 0.
0. 0. 0. 0. 0. 0.12643493 0.
0. 0. 0. 0. 0.04214498 0.04214498 0.
0.04214498 0. 0. 0. 0. 0.
0.12643493 0. 0.04214498 0.04214498 0. 0.
0. 0.04214498 0.04214498 0.04214498 0.04214498 0.08428995
0. 0.04214498 0. 0. 0. 0.
0. 0. 0. 0. 0. 0.04214498
0. 0. 0.04214498 0.04214498 0. 0.
0. 0. 0.04214498 0.08428995 0.04214498 0.08428995 0.04214498
0.04214498 0. 0. 0.04214498 0. 0.08428995
0.04214498 0.04214498 0.04214498 0.04214498 0. 0.04214498
0. 0. 0. 0. 0. 0.
0. 0. 0. 0. 0.04214498 0.
0. 0. 0. 0. 0.04214498 0.
0. 0. 0. 0. 0. 0.04214498 0.
0. 0. 0. 0. 0. 0.04214498 0.
0. 0. 0. 0. 0. 0.04214498 0.
0. 0. 0. 0. 0.04214498 0. 0.04214498 0.04214498
0.04214498 0. 0.04214498 0. 0.

```
0.04214498 0.04214498 0.04214498 0.04214498 0.          0.  
0.          0.          0.04214498 0.08428995 0.04214498 0.  
0.04214498 0.          0.          0.          0.          0.  
0.08428995 0.04214498 0.          0.08428995 0.08428995 0.  
0.          0.          0.04214498 0.          0.          0.  
0.          0.          0.04214498 0.          0.04214498 0.  
0.04214498 0.          0.04214498 0.          0.04214498 0.  
0.04214498 0.04214498 0.          0.          0.04214498 0.04214498  
0.04214498 0.          0.          0.          0.          0.04214498  
0.04214498 0.          0.          0.04214498 0.08428995 0.04214498  
0.          0.          0.04214498 0.12643493 0.04214498  
0.          0.          0.          0.12643493 0.04214498  
0.04214498 0.          0.08428995 0.          0.          0.  
0.04214498 0.          0.08428995 0.          0.08428995 0.  
0.          0.          0.          0.04214498 0.          0.  
0.          0.          0.04214498 0.21072488 0.08428995 0.  
0.          0.          0.          0.          0.          0.  
0.          0.          0.04214498 0.          0.          0.  
0.21072488 0.          0.08428995 0.          ]]
```

In []:

```
#Atribut Normalitzat --> Tots els valors es troben entre 0 i 1.
```

In [324]:

```
#Normalization of Atributes: Attendance
a_attend = np.array([wcupmatch_df['Attendance']])
a_attend.reshape(1, -1)
normalized_arr = preprocessing.normalize(a_attend)
print(normalized_arr)
```

[[0.00299468 0.01236283 0.01621265 0.0017177 0.01577463 0.00623263
0.01233587 0.01233587 0.0389059 0.00134774 0.02836994 0.01716078
0.00808644 0.04718575 0.02793799 0.04911571 0.05382 0.04605634
0.01078193 0.00606483 0.02223772 0.00943418 0.00539096 0.01415128
0.01684676 0.00606483 0.00808644 0.00202161 0.02358546 0.01549902
0.02897642 0.02358546 0.01010806 0.00471709 0.03706287 0.01829693
0.00606483 0.02052205 0.00471709 0.01280354 0.0090649 0.00741257
0.00539096 0.01349425 0.0148393 0.01010806 0.00471709 0.03939109
0.01222468 0.01347741 0.02223772 0.00808644 0.03032417 0.05502084
0.02001597 0.00640918 0.02459762 0.00494351 0.02832412 0.00746514
0.01333589 0.0053256 0.00684046 0.09597868 0.05017773 0.01739327
0.00356073 0.00241246 0.00572857 0.03019074 0.09359116 0.10294852
0.0053822 0.00756554 0.11715236 0.01381434 0.01684676 0.00907703
0.01078193 0.01886837 0.00876031 0.00943418 0.02897642 0.02291159
0.01752063 0.01280354 0.01660214 0.03773674 0.00269548 0.02931336
0.01617289 0.0114558 0.02021611 0.01886837 0.02358546 0.0114558
0.02695481 0.03908448 0.03032417 0.02156385 0.0421169 0.0229837
0.03325415 0.0209951 0.01033919 0.01113099 0.00646309 0.01198681
0.0071747 0.02755793 0.01020914 0.0078607 0.00823267 0.01431233
0.01684676 0.00955144 0.02617986 0.02040951 0.03431887 0.01481841
0.00896248 0.00882972 0.00913364 0.01069567 0.0110636 0.01562166
0.0041753 0.00190234 0.01746874 0.01351447 0.02149646 0.00795167
0.03333704 0.01826189 0.02188933 0.03351629 0.00532897 0.00706486
0.00480739 0.04380562 0.00648398 0.00855815 0.00534918 0.04409808
0.0059496 0.01004269 0.00659989 0.04451385 0.00541792 0.00800221
0.00501494 0.04374901 0.00672051 0.01261148 0.0053539 0.04530026
0.00482963 0.00717537 0.00384106 0.04031631 0.01163639 0.01195176
0.00787754 0.04267217 0.0039691 0.05161442 0.04494513 0.04628074
0.05872645 0.02434491 0.03187946 0.01550306 0.04665676 0.02013929
0.02879987 0.0183286 0.03077027 0.02158272 0.03462818 0.00929402
0.01714191 0.0313936 0.01872888 0.06238018 0.04118156 0.02164943
0.03940726 0.01201443 0.06622124 0.01625982 0.02842857 0.01080012
0.06104187 0.02695953 0.01808938 0.02712193 0.02579104 0.06367603
0.05909573 0.06531421 0.07221195 0.01391812 0.00927583 0.03407088
0.0090521 0.00872123 0.03564572 0.06213421 0.02019455 0.00912218
0.03828797 0.06419356 0.00648533 0.00856489 0.04504352 0.06944773
0.01223951 0.01204543 0.03423531 0.06063418 0.00666458 0.00828793
0.03321642 0.07290738 0.01809409 0.01573959 0.03654601 0.01757791
0.03454327 0.06903397 0.07035409 0.07238176 0.04177996 0.05465088
0.0114558 0.0181945 0.03713026 0.01603811 0.03571513 0.02203556
0.03591729 0.04177996 0.01907053 0.02136169 0.03618684 0.00902986
0.01704892 0.04723831 0.03773674 0.01172534 0.02439411 0.04056699
0.03591729 0.01745324 0.01907053 0.04723831 0.04540875 0.0403399
0.03810602 0.03029384 0.02655049 0.03908448 0.04605769 0.04568841
0.04177996 0.03618684 0.03656016 0.02783085 0.0519554 0.05269666
0.04553948 0.02855391 0.01172265 0.04825922 0.02194728 0.02752154
0.02555788 0.02252816 0.0178798 0.00648533 0.02375932 0.04829359
0.02343115 0.0279144 0.00534918 0.0189526 0.0155846 0.01526384
0.02066558 0.04832459 0.02373439 0.02839151 0.01432783 0.02367307
0.04551792 0.01688045 0.02107732 0.02499452 0.02377954 0.04494378
0.02746022 0.02515288 0.0454411 0.02582136 0.02667583 0.02514547
0.04694113 0.04817027 0.06401768 0.02223772 0.04582318 0.00741257
0.01549902 0.02425933 0.02830255 0.0297662 0.03339836 0.01516208
0.01684676 0.01684676 0.01684676 0.02162652 0.0319273 0.01280354
0.01010806 0.01280354 0.02830255 0.02771157 0.03234578 0.01482515
0.02024509 0.01010806 0.01684676 0.0249332 0.03032417 0.01347741
0.02190079 0.02897642 0.01078193 0.01886837 0.01684676 0.02762868
0.02675265 0.03339836 0.0249332 0.04380157 0.02897642 0.05054028
0.01347741 0.03032417 0.02965029 0.06070831 0.0249332 0.04380157
0.02965029 0.05054028 0.03369352 0.04717092 0.01886837 0.06064833
0.06469155 0.04413851 0.02408952 0.01111886 0.04043222 0.01341002

0.01482515 0.07412574 0.01549902 0.01617289 0.01212967 0.02055305
0.02156385 0.02462322 0.03032417 0.00929941 0.03234578 0.01361218
0.01886837 0.07722554 0.01342013 0.01347741 0.01785756 0.02021611
0.02117301 0.00956896 0.01347741 0.04380157 0.01078193 0.06992281
0.01886837 0.01529686 0.03436739 0.01615941 0.01347741 0.02425933
0.02175051 0.07721206 0.01752063 0.03032417 0.04717092 0.01334263
0.06652987 0.02594401 0.04380157 0.02810039 0.03032417 0.07721206
0.03032417 0.07715815 0.01415128 0.07722554 0.04971815 0.02891375
0.02074914 0.04939672 0.02241697 0.04220315 0.05038192 0.02080036
0.02374584 0.02209621 0.0224318 0.02406593 0.03757434 0.02607002
0.02173704 0.04947758 0.02625534 0.04795868 0.0390892 0.02144458
0.02376539 0.0224318 0.0220578 0.02274919 0.0355352 0.02514008
0.04886234 0.01875583 0.04939672 0.0234891 0.04211824 0.02036638
0.01956852 0.02422564 0.0224318 0.02355783 0.03371104 0.03212542
0.04136284 0.0502431 0.02144121 0.04939672 0.0239224 0.023262
0.0262614 0.04939672 0.04942637 0.03720101 0.0404174 0.04220315
0.03465446 0.04959888 0.03790319 0.04253267 0.04947893 0.05076804
0.06189903 0.04125367 0.03530744 0.06280067 0.03405404 0.0546246
0.03669628 0.02973925 0.04252998 0.04139451 0.06325554 0.0502869
0.03669426 0.04096458 0.0482006 0.05620146 0.04204075 0.05143113
0.03669426 0.04256165 0.06325554 0.05620146 0.04312635 0.04251381
0.04879091 0.05203425 0.03540178 0.05048232 0.04082172 0.0356875
0.0357158 0.04312635 0.04059799 0.03579667 0.04061888 0.06096438
0.04134531 0.05670417 0.03663631 0.04786501 0.03598468 0.04279077
0.04851866 0.05626817 0.04994053 0.06189903 0.06165914 0.06347454
0.05390963 0.02008134 0.02142908 0.02257466 0.02008134 0.02567446
0.03706287 0.0239224 0.02634833 0.05188802 0.02257466 0.02062043
0.02567446 0.03706287 0.02634833 0.0306611 0.02142908 0.0239224
0.02062043 0.02008134 0.02257466 0.05390963 0.0306611 0.02062043
0.0239224 0.02142908 0.03706287 0.02567446 0.0306611 0.02634833
0.02008134 0.02257466 0.05390963 0.0239224 0.02062043 0.03706287
0.02634833 0.02142908 0.02567446 0.02257466 0.0306611 0.02062043
0.02008134 0.0239224 0.02634833 0.02142908 0.05188802 0.02567446
0.03706287 0.0306611 0.02142908 0.05188802 0.02008134 0.02257466
0.02142908 0.02062043 0.05188802 0.0239224 0.03706287 0.02634833
0.036389 0.05121415 0.0306611 0.05390963 0.042158 0.02032191
0.02269528 0.02171075 0.02294529 0.0169721 0.03552712 0.01927134
0.02280512 0.02094456 0.02172491 0.01834073 0.03723538 0.03285792
0.020861 0.02513941 0.02416095 0.02931336 0.03526229 0.02580182
0.02439006 0.02421014 0.01617289 0.0318242 0.02476474 0.0245774
0.02850404 0.04454822 0.03073523 0.04095649 0.02088998 0.02675265
0.03241316 0.02269663 0.04401721 0.03172918 0.03084776 0.03023252
0.02033471 0.02090615 0.02596018 0.02938412 0.04438245 0.02647704
0.0304677 0.03385457 0.01784543 0.03142931 0.02734701 0.01696536
0.02623108 0.02678432 0.02725132 0.0245154 0.03077296 0.02600331
0.03196571 0.0251603 0.02980731 0.02837938 0.04397408 0.04114518
0.04277931 0.0465166 0.04447544 0.03504126 0.03234578 0.0424262
0.0333431 0.02897642 0.02762868 0.03032417 0.03099804 0.03504126
0.02897642 0.03234578 0.03504126 0.04851866 0.02897642 0.04447544
0.04380157 0.03369352 0.02762868 0.04851866 0.03504126 0.03504126
0.02897642 0.03234578 0.03032417 0.03099804 0.02762868 0.04447544
0.02897642 0.04380157 0.03369352 0.03504126 0.04851866 0.02897642
0.03032417 0.03099804 0.02560707 0.03504126 0.03234578 0.04447544
0.03369352 0.02762868 0.04380157 0.03504126 0.04851866 0.03099804
0.03032417 0.02897642 0.04447544 0.02897642 0.03504126 0.02762868
0.03099804 0.03032417 0.04380157 0.02897642 0.04851866 0.03369352
0.03504126 0.03234578 0.04380157 0.04447544 0.03504126 0.04649705
0.05693531 0.04319509 0.02123568 0.03752514 0.02604239 0.02043512
0.02616841 0.04222472 0.05624459 0.02063391 0.04236555 0.01608596
0.02495611 0.03661205 0.0220113 0.04208522 0.02874596 0.05537462
0.02128959 0.02383479 0.02580519 0.03071029 0.04319509 0.0417867

```
0.02345877 0.02565694 0.01795393 0.02576139 0.05691172 0.0428878  
0.02349921 0.03664911 0.02252412 0.0265606 0.04169505 0.02620749  
0.0248611 0.02414275 0.05619472 0.02549656 0.03599276 0.02348438  
0.01884613 0.0425165 0.04225976 0.02342575 0.02827425 0.01889667  
0.02061841 0.02356929 0.02729849 0.05685916 0.04175435 0.03645369  
0.02475934 0.04242351 0.02708015 0.05661656 0.04319509 0.03730479  
0.04210275 0.04107914 0.0244305 0.05693531 0.04184937 0.0264265  
0.03246236 0.02714013 0.03852786 0.03954204 0.02682004 0.02713474  
0.04605971 0.02898451 0.05036372 0.03442197 0.02633553 0.02679308  
0.03827584 0.04066268 0.02533955 0.02889354 0.04993447 0.02694268  
0.04632724 0.04216744 0.02660777 0.02714687 0.03436941 0.02643189  
0.03888097 0.04017682 0.02729107 0.04974443 0.02879583 0.0270377  
0.02653364 0.04245114 0.04657253 0.02777154 0.0267567 0.0389652  
0.02718393 0.03982237 0.02916848 0.03235319 0.0271718 0.04969726  
0.02821899 0.0455132 0.04137362 0.02649052 0.03889175 0.04973433  
0.04574367 0.04066268 0.05002813 0.0391795 0.04584609 0.05036372  
0.04263375 0.03448801 0.04619449 0.03963503 0.02779176 0.04262567  
0.03452036 0.03889175 0.04973433 0.03963503 0.02779176 0.04574367  
0.04262567 0.03452036 0.05002813 0.04066268 0.04619449 0.03448801  
0.0391795 0.04263375 0.04584609 0.05036372]]
```

In []:

```
#Atribut Normalitzat --> Tots els valors es troben entre 0 i 1.
```

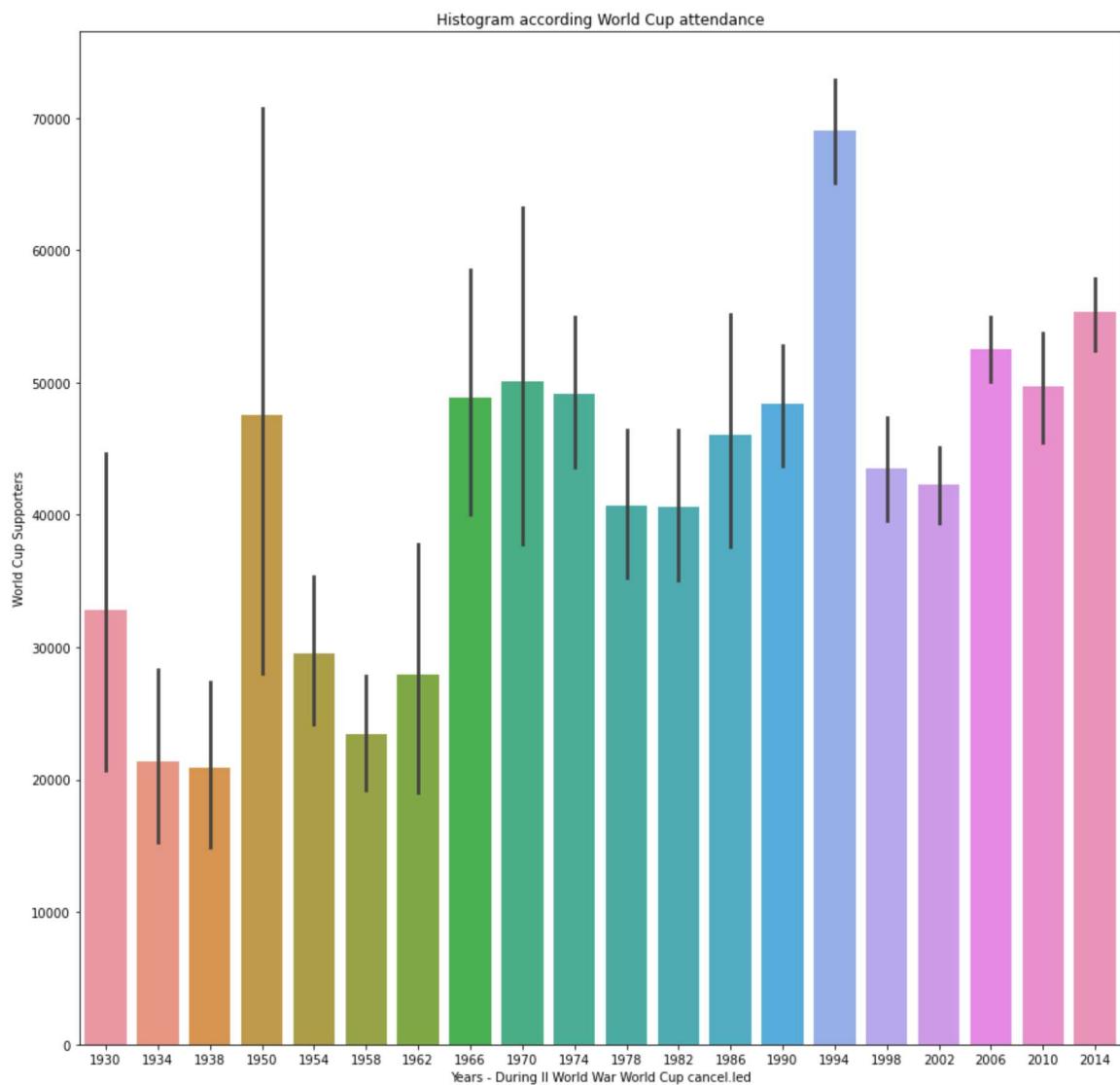
In [326]:

```
#World Cup Matches - grafic sobre assistència (sense outliers).
import matplotlib.pyplot as plt

fig = plt.figure(figsize=(15,15))

wcupmatch_df['Attendance'].dropna() #Esborrem els valors Nan

sns.barplot(data=wcupmatch_df, x="Year", y="Attendance")
plt.xlabel("Years - During II World War World Cup cancelled")
plt.ylabel("World Cup Supporters")
plt.title("Histogram according World Cup attendance")
plt.show()
```



In []: