

```
In [ ]: from datasets import load_dataset

# dataset = load_dataset("azminetoushikwasi/cristiano-ronaldo-all-club-goals-stats", us
```

Using custom data configuration azminetoushikwasi--cristiano-ronaldo-all-club-goals-stat
s-a0623dcc1ee883c9
Found cached dataset csv (C:/Users/hakan/.cache/huggingface/datasets/azminetoushikwasi__
_csv/azminetoushikwasi--cristiano-ronaldo-all-club-goals-stats-a0623dcc1ee883c9/0.0.0/6b
34fb8fcf56f7c8ba51dc895bfa2bfbe43546f190a60fcf74bb5e8afdcc2317)
100%|██████████| 1/1 [00:00<00:00, 335.36it/s]

```
In [ ]: dataset
```

```
Out[ ]: DatasetDict({
  train: Dataset({
    features: ['Unnamed: 0', 'Goal_no', 'Season', 'Competition', 'Matchday', 'Venu
e', 'Team', 'Opponent', 'Result', 'Position', 'Minute', 'At_score', 'Type_of_goal'],
    num_rows: 698
  })
})
```

```
In [ ]: df = dataset['train'].to_pandas()
```

```
In [ ]: df
```

```
Out[ ]:
```

	Unnamed: 0	Goal_no	Season	Competition	Matchday	Venue	Team	Opponent	Result	Posit
0	0	1	2/3	Liga Portugal	6	H	Sporting CP	Moreirense FC	3:0	
1	1	2	2/3	Liga Portugal	6	H	Sporting CP	Moreirense FC	3:0	No
2	2	3	2/3	Liga Portugal	8	A	Sporting CP	Boavista FC	1:2	No
3	3	4	2/3	Taça de Portugal Placard	Fourth Round	H	Sporting CP	CD Estarreja	4:1	No
4	4	5	2/3	Taça de Portugal Placard	Fifth Round	H	Sporting CP	FC Oliveira do Hospital	8:1	No
...	
693	693	694	21/22	Premier League	33	H	Manchester United	Norwich City	3:2	
694	694	695	21/22	Premier League	33	H	Manchester United	Norwich City	3:2	
695	695	696	21/22	Premier League	34	A	Manchester United	Arsenal FC	3:1	

	Unnamed: 0	Goal_no	Season	Competition	Matchday	Venue	Team	Opponent	Result	Posit
696	696	697	21/22	Premier League	37	H	Manchester United	Chelsea FC	1:1	
697	697	698	21/22	Premier League	35	H	Manchester United	Brentford FC	3:0	

698 rows × 13 columns

In []:

```
# replace Ã§ with ç
df['Competition'] = df['Competition'].str.replace('Ã§', 'ç')

df['Competition'] = df['Competition'].str.replace('Italy Cup', 'Coppa Italia')

df['Competition'] = df['Competition'].str.replace('Liga Portugal', 'Primeira Liga')

df['Competition'] = df['Competition'].str.replace('Taça de Portugal Placard', 'Taça de
```

In []:

```
df
```

Out[]:

	Unnamed: 0	Goal_no	Season	Competition	Matchday	Venue	Team	Opponent	Result	Posit
0	0	1	2/3	Primeira Liga	6	H	Sporting CP	Moreirense FC	3:0	
1	1	2	2/3	Primeira Liga	6	H	Sporting CP	Moreirense FC	3:0	No
2	2	3	2/3	Primeira Liga	8	A	Sporting CP	Boavista FC	1:2	No
3	3	4	2/3	Taça de Portugal	Fourth Round	H	Sporting CP	CD Estarreja	4:1	No
4	4	5	2/3	Taça de Portugal	Fifth Round	H	Sporting CP	FC Oliveira do Hospital	8:1	No
...	
693	693	694	21/22	Premier League	33	H	Manchester United	Norwich City	3:2	
694	694	695	21/22	Premier League	33	H	Manchester United	Norwich City	3:2	
695	695	696	21/22	Premier League	34	A	Manchester United	Arsenal FC	3:1	
696	696	697	21/22	Premier League	37	H	Manchester United	Chelsea FC	1:1	
697	697	698	21/22	Premier League	35	H	Manchester United	Brentford FC	3:0	

698 rows × 13 columns

```
In [ ]: df.describe()
df = df.drop('Unnamed: 0', axis=1)
```

```
In [ ]: df
```

```
Out[ ]:
```

	Goal_no	Season	Competition	Matchday	Venue	Team	Opponent	Result	Position	Minute
0	1	2/3	Primeira Liga	6	H	Sporting CP	Moreirense FC	3:0	LW	34
1	2	2/3	Primeira Liga	6	H	Sporting CP	Moreirense FC	3:0	None	90+5
2	3	2/3	Primeira Liga	8	A	Sporting CP	Boavista FC	1:2	None	88
3	4	2/3	Taça de Portugal	Fourth Round	H	Sporting CP	CD Estarreja	4:1	None	67
4	5	2/3	Taça de Portugal	Fifth Round	H	Sporting CP	FC Oliveira do Hospital	8:1	None	13
...
693	694	21/22	Premier League	33	H	Manchester United	Norwich City	3:2	CF	32
694	695	21/22	Premier League	33	H	Manchester United	Norwich City	3:2	CF	76
695	696	21/22	Premier League	34	A	Manchester United	Arsenal FC	3:1	CF	34
696	697	21/22	Premier League	37	H	Manchester United	Chelsea FC	1:1	CF	62
697	698	21/22	Premier League	35	H	Manchester United	Brentford FC	3:0	CF	61

698 rows × 12 columns

```
In [ ]: # cluster the goals based on the competition
df['Competition'].value_counts()
```

```
Out[ ]:
```

LaLiga	311
UEFA Champions League	140
Premier League	102
Serie A	81
Copa del Rey	22
FA Cup	13
FIFA Club World Cup	7
EFL Cup	4
Supercopa	4

Coppa Italia	4
Primeira Liga	3
Taça de Portugal	2
UEFA Super Cup	2
Supercoppa Italiana	2
UEFA Champions League Qualifying	1
Name: Competition, dtype: int64	

In []:

```
# if there's a + sign in the minutes, replace it with 90 + the number after the + sign
for i in range(len(df['Minute'])):
    if '+' in df['Minute'][i]:
        if ('90+') in df['Minute'][i]:
            df['Minute'][i] = df['Minute'][i].replace('+', '')
            df['Minute'][i] = 90 + int(df['Minute'][i])
            df['Minute'][i] = df['Minute'][i] - 900
        elif ('45+') in df['Minute'][i]:
            df['Minute'][i] = df['Minute'][i].replace('+', '')
            df['Minute'][i] = 45 + int(df['Minute'][i])
            df['Minute'][i] = df['Minute'][i] - 450

df['Minute'] = df['Minute'].astype(int)
```

C:\Users\hakan\AppData\Local\Temp\ipykernel_4304\837626853.py:5: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy

```
df['Minute'][i] = df['Minute'][i].replace('+', '')
```

C:\Users\hakan\AppData\Local\Temp\ipykernel_4304\837626853.py:6: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy

```
df['Minute'][i] = 90 + int(df['Minute'][i])
```

C:\Users\hakan\AppData\Local\Temp\ipykernel_4304\837626853.py:7: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy

```
df['Minute'][i] = df['Minute'][i] - 900
```

C:\Users\hakan\AppData\Local\Temp\ipykernel_4304\837626853.py:9: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy

```
df['Minute'][i] = df['Minute'][i].replace('+', '')
```

C:\Users\hakan\AppData\Local\Temp\ipykernel_4304\837626853.py:10: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy

```
df['Minute'][i] = 45 + int(df['Minute'][i])
```

C:\Users\hakan\AppData\Local\Temp\ipykernel_4304\837626853.py:11: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy

```
guide/indexing.html#returning-a-view-versus-a-copy
df['Minute'][i] = df['Minute'][i] - 450
```

In []:

```
df
```

Out[]:

	Goal_no	Season	Competition	Matchday	Venue	Team	Opponent	Result	Position	Minute
0	1	2/3	Primeira Liga	6	H	Sporting CP	Moreirense FC	3:0	LW	34
1	2	2/3	Primeira Liga	6	H	Sporting CP	Moreirense FC	3:0	None	95
2	3	2/3	Primeira Liga	8	A	Sporting CP	Boavista FC	1:2	None	88
3	4	2/3	Taça de Portugal	Fourth Round	H	Sporting CP	CD Estarreja	4:1	None	67
4	5	2/3	Taça de Portugal	Fifth Round	H	Sporting CP	FC Oliveira do Hospital	8:1	None	13
...
693	694	21/22	Premier League	33	H	Manchester United	Norwich City	3:2	CF	32
694	695	21/22	Premier League	33	H	Manchester United	Norwich City	3:2	CF	76
695	696	21/22	Premier League	34	A	Manchester United	Arsenal FC	3:1	CF	34
696	697	21/22	Premier League	37	H	Manchester United	Chelsea FC	1:1	CF	62
697	698	21/22	Premier League	35	H	Manchester United	Brentford FC	3:0	CF	61

698 rows × 12 columns



In []:

```
# plot the goals based on the competition using plotly

import plotly.express as px
fig = px.scatter(df, x="Minute", y="Competition", color="Competition", size='Minute')

# cut competitions based on what country they're in

countries_ronaldo_played = ['Portugal', 'England', 'Spain', 'Italy']
competitions_pt = ['Primeira Liga', 'Taça de Portugal', 'Supertaca Cândido de Oliveira']
competitions_eng = ['Premier League', 'FA Cup', 'League Cup', 'Community Shield', 'EFL']
competitions_ita = ['Serie A', 'Coppa Italia', 'Supercoppa Italiana', 'Coppa Italia']
competition_esp = ['LaLiga', 'Copa del Rey', 'Supercopa de España', 'Supercopa de España']
competition_european = ['UEFA Super Cup', 'UEFA Champions League', 'UEFA Europa League']
df['Country'] = df['Competition'].apply(lambda x: 'Portugal' if x in competitions_pt else

# the other in this case is the FIFA Club World Cup
```

```
print(df['Country'].value_counts())

# export the dataframe to a csv file for checking
# df.to_csv('ronaldo_goals.csv')
```

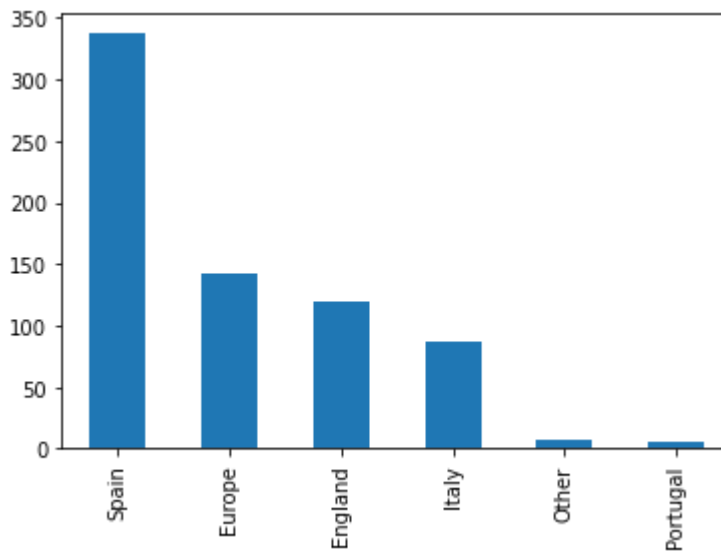
Spain	337
Europe	143
England	119
Italy	87
Other	7
Portugal	5

Name: Country, dtype: int64

```
In [ ]: # plot the goals based on country using matplotlib

df['Country'].value_counts().plot(kind='bar')
```

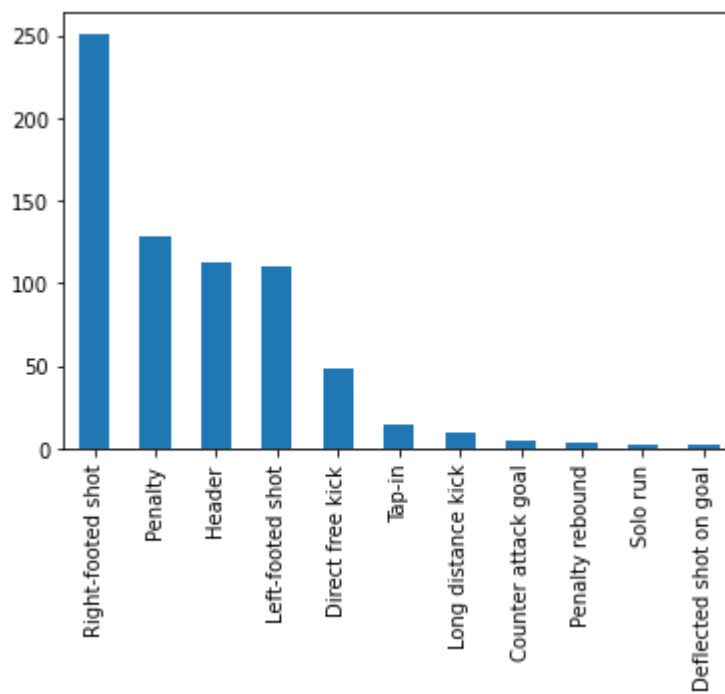
Out[]: <AxesSubplot:>



```
In [ ]: # fig.show()
```

```
In [ ]: # calculate the number of goals based on the type of goal
df['Type_of_goal'].value_counts().plot(kind='bar')
```

Out[]: <AxesSubplot:>



```
In [ ]: df['Minute'].value_counts()
```

```
Out[ ]: 90      17
45       15
89       15
23       14
76       14
..
72        2
95        2
93        2
120       1
1         1
Name: Minute, Length: 98, dtype: int64
```

```
In [ ]: #
df['Minute'].astype(int)

df['number_scored_at_that_minute'] = df.groupby('Minute')['Minute'].transform('count')
```

```
In [ ]: df
```

```
Out[ ]:
```

	Goal_no	Season	Competition	Matchday	Venue	Team	Opponent	Result	Position	Minute
0	1	2/3	Primeira Liga	6	H	Sporting CP	Moreirense FC	3:0	LW	34
1	2	2/3	Primeira Liga	6	H	Sporting CP	Moreirense FC	3:0	None	95
2	3	2/3	Primeira Liga	8	A	Sporting CP	Boavista FC	1:2	None	88
3	4	2/3	Taça de Portugal	Fourth Round	H	Sporting CP	CD Estarreja	4:1	None	67

	Goal_no	Season	Competition	Matchday	Venue	Team	Opponent	Result	Position	Minute	
	4	5	2/3	Taça de Portugal	Fifth Round	H	Sporting CP	FC Oliveira do Hospital	8:1	None	13
	
	693	694	21/22	Premier League	33	H	Manchester United	Norwich City	3:2	CF	32
	694	695	21/22	Premier League	33	H	Manchester United	Norwich City	3:2	CF	76
	695	696	21/22	Premier League	34	A	Manchester United	Arsenal FC	3:1	CF	34
	696	697	21/22	Premier League	37	H	Manchester United	Chelsea FC	1:1	CF	62
	697	698	21/22	Premier League	35	H	Manchester United	Brentford FC	3:0	CF	61

698 rows × 14 columns

```
In [ ]: fig = px.scatter(df, x="Minute", y="Competition", color="Competition", size='number_sco
fig.show()
```

In []: