# Finding Replacement Player for Cristiano Ronaldo in Manchester United

## Import The Libraries

In [434]:

```
import pandas as pd
import seaborn as sns
import numpy as np
import matplotlib.pyplot as plt
import os
from sklearn.neighbors import NearestNeighbors
```

## Reading The Data

### DataFrame #1

In [435]:

```
fifa = pd.read_csv('Fifa 23 Players Data.csv')#loading the dataset using read.csv()
```

In [436]:

```
fifa.head(10) #show the first 10 rows of the dataframe
```

Out[436]:

| | Known As | Full Name | Overall | Potential | Value(in Euro) | Positions Played | Best Position | Nationality | |
|---|---|---|---|---|---|---|---|---|---|
| 0 | L. Messi | Lionel Messi | 91 | 91 | 54000000 | RW | CAM | Argentina | https://cdn.sofifa.net/players/1 |
| 1 | K. Benzema | Karim Benzema | 91 | 91 | 64000000 | CF,ST | CF | France | https://cdn.sofifa.net/players/1 |
| 2 | R. Lewandowski | Robert Lewandowski | 91 | 91 | 84000000 | ST | ST | Poland | https://cdn.sofifa.net/players/1 |
| 3 | K. De Bruyne | Kevin De Bruyne | 91 | 91 | 107500000 | CM,CAM | CM | Belgium | https://cdn.sofifa.net/players/1 |
| 4 | K. Mbappé | Kylian Mbappé | 91 | 95 | 190500000 | ST,LW | ST | France | https://cdn.sofifa.net/players/2 |
| 5 | M. Salah | Mohamed Salah | 90 | 90 | 115500000 | RW | RW | Egypt | https://cdn.sofifa.net/players/2 |
| 6 | T. Courtois | Thibaut Courtois | 90 | 91 | 90000000 | GK | GK | Belgium | https://cdn.sofifa.net/players/1 |
| 7 | M. Neuer | Manuel Neuer | 90 | 90 | 13500000 | GK | GK | Germany | https://cdn.sofifa.net/players/1 |
| 8 | Cristiano Ronaldo | C. Ronaldo dos Santos Aveiro | 90 | 90 | 41000000 | ST | ST | Portugal | https://cdn.sofifa.net/players/0 |
| 9 | V. van Dijk | Virgil van Dijk | 90 | 90 | 98000000 | CB | CB | Netherlands | https://cdn.sofifa.net/players/2 |

10 rows × 89 columns

In [437]:

```
#get the current shape of the dataframe
```

```
#get the current shape of the dataframe
fifa.shape
```

Out[437]:

```
(18539, 89)
```

In [438]:

```
#Checking the columns for na values
for column in fifa.columns:
    print(column)
    print(pd.isna(fifa[column]).sum())
```

```
Known As
0
Full Name
0
Overall
0
Potential
0
Value(in Euro)
0
Positions Played
0
Best Position
0
Nationality
0
Image Link
0
Age
0
Height(in cm)
0
Weight(in kg)
0
TotalStats
0
BaseStats
0
Club Name
0
Wage(in Euro)
0
Release Clause
0
Club Position
0
Contract Until
0
Club Jersey Number
0
Joined On
0
On Loan
0
Preferred Foot
0
Weak Foot Rating
0
Skill Moves
0
International Reputation
0
National Team Name
0
National Team Image Link
0
National Team Position
0
National Team Jersey Number
```

0
Attacking Work Rate
0
Defensive Work Rate
0
Pace Total
0
Shooting Total
0
Passing Total
0
Dribbling Total
0
Defending Total
0
Physicality Total
0
Crossing
0
Finishing
0
Heading Accuracy
0
Short Passing
0
Volleys
0
Dribbling
0
Curve
0
Freekick Accuracy
0
LongPassing
0
BallControl
0
Acceleration
0
Sprint Speed
0
Agility
0
Reactions
0
Balance
0
Shot Power
0
Jumping
0
Stamina
0
Strength
0
Long Shots
0
Aggression
0
Interceptions
0
Positioning
0
Vision
0
Penalties
0
Composure
0
Marking
0
Standing Tackle

```
0
Sliding Tackle
0
Goalkeeper Diving
0
Goalkeeper Handling
0
 GoalkeeperKicking
0
Goalkeeper Positioning
0
Goalkeeper Reflexes
0
ST Rating
0
LW Rating
0
LF Rating
0
CF Rating
0
RF Rating
0
RW Rating
0
CAM Rating
0
LM Rating
0
CM Rating
0
RM Rating
0
LWB Rating
0
CDM Rating
0
RWB Rating
0
LB Rating
0
CB Rating
0
RB Rating
0
GK Rating
0
```

In [439]:

```
#description of the data
fifa.describe()
```

Out[439]:

| | Overall | Potential | Value(in Euro) | Age | Height(in cm) | Weight(in kg) | TotalStats | BaseStats | Wa |
|---|---|---|---|---|---|---|---|---|---|
| count | 18539.000000 | 18539.000000 | 1.853900e+04 | 18539.000000 | 18539.000000 | 18539.000000 | 18539.000000 | 18539.000000 | 18 |
| mean | 65.852042 | 71.016668 | 2.875461e+06 | 25.240412 | 181.550839 | 75.173904 | 1602.114569 | 357.946221 | 8 |
| std | 6.788353 | 6.192866 | 7.635129e+06 | 4.718163 | 6.858097 | 7.013593 | 273.160237 | 39.628259 | 19 |
| min | 47.000000 | 48.000000 | 0.000000e+00 | 16.000000 | 155.000000 | 49.000000 | 759.000000 | 224.000000 | |
| 25% | 62.000000 | 67.000000 | 4.750000e+05 | 21.000000 | 177.000000 | 70.000000 | 1470.000000 | 331.000000 | 1 |
| 50% | 66.000000 | 71.000000 | 1.000000e+06 | 25.000000 | 182.000000 | 75.000000 | 1640.000000 | 358.000000 | 3 |
| 75% | 70.000000 | 75.000000 | 2.000000e+06 | 29.000000 | 186.000000 | 80.000000 | 1786.000000 | 385.000000 | 8 |
| max | 91.000000 | 95.000000 | 1.905000e+08 | 44.000000 | 206.000000 | 105.000000 | 2312.000000 | 502.000000 | 450 |

8 rows × 71 columns

8 rows × 71 columns

In [440]:

```
#print information about the data frame
fifa.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 18539 entries, 0 to 18538
Data columns (total 89 columns):
 #   Column                        Non-Null Count  Dtype
---  ------                        --------------  -----
 0   Known As                      18539 non-null  object
 1   Full Name                     18539 non-null  object
 2   Overall                       18539 non-null  int64
 3   Potential                     18539 non-null  int64
 4   Value(in Euro)                18539 non-null  int64
 5   Positions Played              18539 non-null  object
 6   Best Position                 18539 non-null  object
 7   Nationality                   18539 non-null  object
 8   Image Link                    18539 non-null  object
 9   Age                           18539 non-null  int64
 10  Height(in cm)                 18539 non-null  int64
 11  Weight(in kg)                 18539 non-null  int64
 12  TotalStats                    18539 non-null  int64
 13  BaseStats                     18539 non-null  int64
 14  Club Name                     18539 non-null  object
 15  Wage(in Euro)                 18539 non-null  int64
 16  Release Clause                18539 non-null  int64
 17  Club Position                 18539 non-null  object
 18  Contract Until                18539 non-null  object
 19  Club Jersey Number            18539 non-null  object
 20  Joined On                     18539 non-null  int64
 21  On Loan                       18539 non-null  object
 22  Preferred Foot                18539 non-null  object
 23  Weak Foot Rating              18539 non-null  int64
 24  Skill Moves                   18539 non-null  int64
 25  International Reputation      18539 non-null  int64
 26  National Team Name            18539 non-null  object
 27  National Team Image Link      18539 non-null  object
 28  National Team Position        18539 non-null  object
 29  National Team Jersey Number   18539 non-null  object
 30  Attacking Work Rate           18539 non-null  object
 31  Defensive Work Rate           18539 non-null  object
 32  Pace Total                    18539 non-null  int64
 33  Shooting Total                18539 non-null  int64
 34  Passing Total                 18539 non-null  int64
 35  Dribbling Total               18539 non-null  int64
 36  Defending Total               18539 non-null  int64
 37  Physicality Total             18539 non-null  int64
 38  Crossing                      18539 non-null  int64
 39  Finishing                     18539 non-null  int64
 40  Heading Accuracy              18539 non-null  int64
 41  Short Passing                 18539 non-null  int64
 42  Volleys                       18539 non-null  int64
 43  Dribbling                     18539 non-null  int64
 44  Curve                         18539 non-null  int64
 45  Freekick Accuracy             18539 non-null  int64
 46  LongPassing                   18539 non-null  int64
 47  BallControl                   18539 non-null  int64
 48  Acceleration                  18539 non-null  int64
 49  Sprint Speed                  18539 non-null  int64
 50  Agility                       18539 non-null  int64
 51  Reactions                     18539 non-null  int64
 52  Balance                       18539 non-null  int64
 53  Shot Power                    18539 non-null  int64
 54  Jumping                       18539 non-null  int64
 55  Stamina                       18539 non-null  int64
 56  Strength                      18539 non-null  int64
 57  Long Shots                    18539 non-null  int64
 58  Aggression                    18539 non-null  int64
```

```
 59  Interceptions              18539 non-null  int64
 60  Positioning                18539 non-null  int64
 61  Vision                     18539 non-null  int64
 62  Penalties                  18539 non-null  int64
 63  Composure                  18539 non-null  int64
 64  Marking                    18539 non-null  int64
 65  Standing Tackle            18539 non-null  int64
 66  Sliding Tackle             18539 non-null  int64
 67  Goalkeeper Diving          18539 non-null  int64
 68  Goalkeeper Handling        18539 non-null  int64
 69   GoalkeeperKicking         18539 non-null  int64
 70  Goalkeeper Positioning     18539 non-null  int64
 71  Goalkeeper Reflexes        18539 non-null  int64
 72  ST Rating                  18539 non-null  int64
 73  LW Rating                  18539 non-null  int64
 74  LF Rating                  18539 non-null  int64
 75  CF Rating                  18539 non-null  int64
 76  RF Rating                  18539 non-null  int64
 77  RW Rating                  18539 non-null  int64
 78  CAM Rating                 18539 non-null  int64
 79  LM Rating                  18539 non-null  int64
 80  CM Rating                  18539 non-null  int64
 81  RM Rating                  18539 non-null  int64
 82  LWB Rating                 18539 non-null  int64
 83  CDM Rating                 18539 non-null  int64
 84  RWB Rating                 18539 non-null  int64
 85  LB Rating                  18539 non-null  int64
 86  CB Rating                  18539 non-null  int64
 87  RB Rating                  18539 non-null  int64
 88  GK Rating                  18539 non-null  int64
dtypes: int64(71), object(18)
memory usage: 12.6+ MB
```

In [441]:

```
fifa['Contract Until'] = np.where(fifa['Contract Until'] =='-', '0', fifa['Contract Until'])
fifa['Contract Until'] = fifa['Contract Until'].astype(str)# convert data type of the 'contract until' column to string
```

In [442]:

```
#check for duplicate rows
fifa.duplicated().sum()
```

Out[442]:

```
119
```

In [443]:

```
#return the number of unique values in each column
fifa.nunique()
```

Out[443]:

```
Known As           17530
Full Name          18337
Overall               45
Potential             48
Value(in Euro)       257
                    ...
RWB Rating            72
LB Rating             71
CB Rating             73
RB Rating             71
GK Rating             64
Length: 89, dtype: int64
```

## Cleaning The Data

In [444]:

```
#1 - Remove irrelevant data
#2 - Remove duplicate data
#3 - fix structural errors (not found within this database)
#4 - deal with missing data (not applicable on this dataset as this dataset doesn't have missing data)
#5 - filter out data outliers
#6 - Validate data
```

In [445]:

```
#removing irrelevant data
fifa.drop(['Full Name', 'Image Link' , 'Club Position' , 'Club Jersey Number', 'International Reputation', 'National Team Name', 'National Team Image Link', 'National Team Position', 'National Team Jersey Number' , 'Attacking Work Rate' , 'Defensive Work Rate' , 'Goalkeeper Diving' , 'Goalkeeper Handling', ' GoalkeeperKicking' ,'Goalkeeper Positioning', 'Goalkeeper Reflexes', 'ST Rating', 'CF Rating', 'RW Rating', 'LW Rating', 'LF Rating', 'RF Rating', 'CAM Rating', 'LM Rating', 'CM Rating', 'RM Rating', 'LWB Rating', 'RWB Rating', 'CDM Rating', 'LB Rating', 'RB Rating', 'CB Rating', 'GK Rating' ],axis=1,inplace=True)
```

In [446]:

```
fifa.shape
```

Out[446]:

```
(18539, 56)
```

In [447]:

```
fifa.head(10)
```

Out[447]:

| | Known As | Overall | Potential | Value(in Euro) | Positions Played | Best Position | Nationality | Age | Height(in cm) | Weight(in kg) | ... | Long Shots | Aggress |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | L. Messi | 91 | 91 | 54000000 | RW | CAM | Argentina | 35 | 169 | 67 | ... | 91 | |
| 1 | K. Benzema | 91 | 91 | 64000000 | CF,ST | CF | France | 34 | 185 | 81 | ... | 80 | |
| 2 | R. Lewandowski | 91 | 91 | 84000000 | ST | ST | Poland | 33 | 185 | 81 | ... | 84 | |
| 3 | K. De Bruyne | 91 | 91 | 107500000 | CM,CAM | CM | Belgium | 31 | 181 | 70 | ... | 91 | |
| 4 | K. Mbappé | 91 | 95 | 190500000 | ST,LW | ST | France | 23 | 182 | 73 | ... | 82 | |
| 5 | M. Salah | 90 | 90 | 115500000 | RW | RW | Egypt | 30 | 175 | 71 | ... | 85 | |
| 6 | T. Courtois | 90 | 91 | 90000000 | GK | GK | Belgium | 30 | 199 | 96 | ... | 17 | |
| 7 | M. Neuer | 90 | 90 | 13500000 | GK | GK | Germany | 36 | 193 | 93 | ... | 16 | |
| 8 | Cristiano Ronaldo | 90 | 90 | 41000000 | ST | ST | Portugal | 37 | 187 | 83 | ... | 90 | |
| 9 | V. van Dijk | 90 | 90 | 98000000 | CB | CB | Netherlands | 30 | 193 | 92 | ... | 64 | |

**10 rows × 56 columns**

In [448]:

```
fifa.columns
```

Out[448]:

```
Index(['Known As', 'Overall', 'Potential', 'Value(in Euro)',
       'Positions Played', 'Best Position', 'Nationality', 'Age',
       'Height(in cm)', 'Weight(in kg)', 'TotalStats', 'BaseStats',
       'Club Name', 'Wage(in Euro)', 'Release Clause', 'Contract Until',
       'Joined On', 'On Loan', 'Preferred Foot', 'Weak Foot Rating',
       'Skill Moves', 'Pace Total', 'Shooting Total', 'Passing Total',
```

```
       'Dribbling Total', 'Defending Total', 'Physicality Total', 'Crossing',
       'Finishing', 'Heading Accuracy', 'Short Passing', 'Volleys',
       'Dribbling', 'Curve', 'Freekick Accuracy', 'LongPassing', 'BallControl',
       'Acceleration', 'Sprint Speed', 'Agility', 'Reactions', 'Balance',
       'Shot Power', 'Jumping', 'Stamina', 'Strength', 'Long Shots',
       'Aggression', 'Interceptions', 'Positioning', 'Vision', 'Penalties',
       'Composure', 'Marking', 'Standing Tackle', 'Sliding Tackle'],
      dtype='object')
```

In [449]:

```
#removing all 119 duplicates from the dataset
fifa.drop_duplicates()
```

Out[449]:

| | Known As | Overall | Potential | Value(in Euro) | Positions Played | Best Position | Nationality | Age | Height(in cm) | Weight(in kg) | ... | Long Shots | Aggr |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | L. Messi | 91 | 91 | 54000000 | RW | CAM | Argentina | 35 | 169 | 67 | ... | 91 | |
| 1 | K. Benzema | 91 | 91 | 64000000 | CF,ST | CF | France | 34 | 185 | 81 | ... | 80 | |
| 2 | R. Lewandowski | 91 | 91 | 84000000 | ST | ST | Poland | 33 | 185 | 81 | ... | 84 | |
| 3 | K. De Bruyne | 91 | 91 | 107500000 | CM,CAM | CM | Belgium | 31 | 181 | 70 | ... | 91 | |
| 4 | K. Mbappé | 91 | 95 | 190500000 | ST,LW | ST | France | 23 | 182 | 73 | ... | 82 | |
| ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | |
| 18534 | D. Collins | 47 | 56 | 110000 | ST,RM | CAM | Republic of Ireland | 21 | 174 | 68 | ... | 46 | |
| 18535 | Yang Dejiang | 47 | 57 | 90000 | CDM | CDM | China PR | 17 | 175 | 60 | ... | 35 | |
| 18536 | L. Mullan | 47 | 67 | 130000 | CM | RM | Northern Ireland | 18 | 170 | 65 | ... | 36 | |
| 18537 | D. McCallion | 47 | 61 | 100000 | CB | CB | Republic of Ireland | 17 | 178 | 65 | ... | 18 | |
| 18538 | N. Rabha | 47 | 50 | 60000 | LB | LB | India | 25 | 176 | 66 | ... | 28 | |

**18420 rows × 56 columns**

In [450]:

```
#drop the Goalkeeper positions because they're irrelevent to this analysis
fifa= fifa[fifa['Best Position'] != 'GK']
```

In [451]:

```
fifa.drop(['Positions Played' ],axis=1,inplace=True)
```

In [452]:

```
fifa.shape
```

Out[452]:

```
(16478, 55)
```

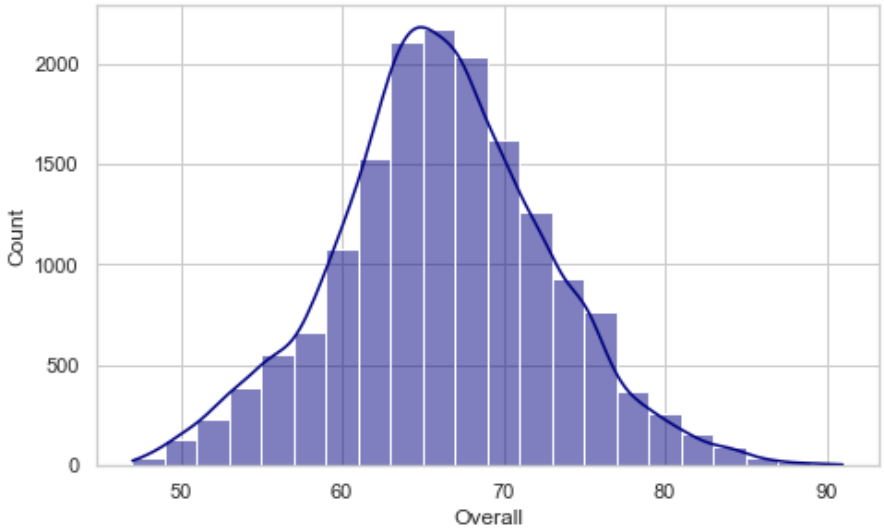
# Exploratory Data Analysis

In [453]:

```
#create a histogram to explore the general overall of all the players where most of the p
layers lay on the spectrum
sns.histplot(data = fifa['Overall'] , color = 'navy',binwidth = 2  , kde = True)
```

Out[453]:

```
<AxesSubplot:xlabel='Overall', ylabel='Count'>
```



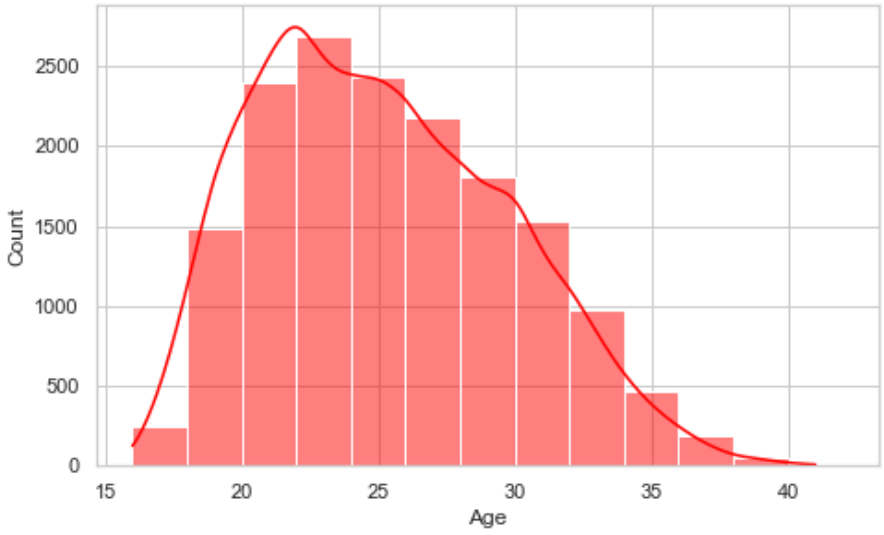

**The histogram shows that players that are 75 overall or over are a very small number of all the players, which would narrow up the candidates that can replace ronaldo a bit**

In [454]:

```
sns.histplot(data = fifa['Age'] , color = 'red' ,binwidth = 2 , kde = True)
```

Out[454]:

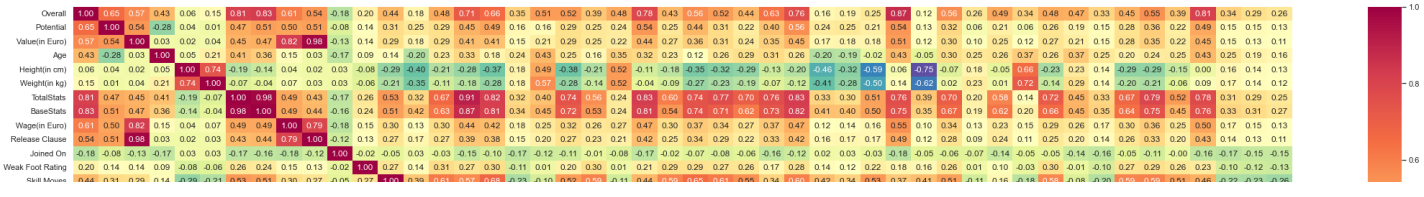
```
<AxesSubplot:xlabel='Age', ylabel='Count'>
```

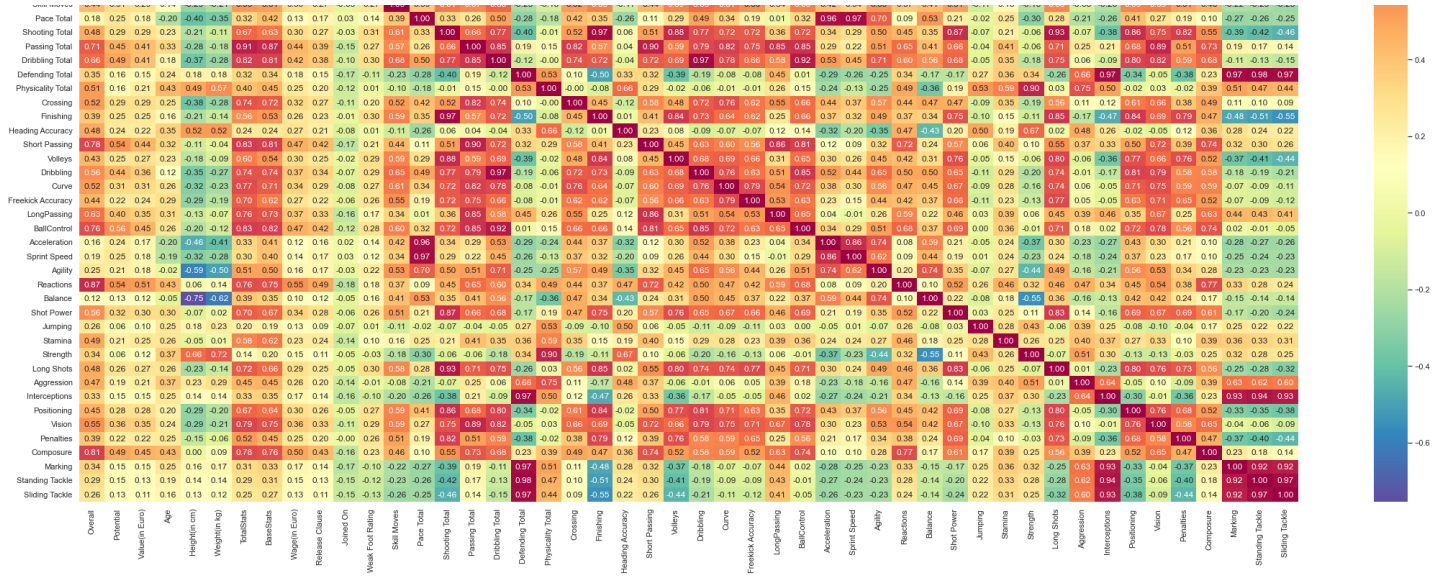



**This histogram shows that most players average in the mid-20's range which means they would have alot of potential to grow and alot of time to get better.**

In [455]:

```
plt.figure(figsize = (30,15))
fig = sns.heatmap(fifa.corr(),
                  cbar=True,
                  annot=True,
                  fmt='.2f',
                  cmap="Spectral_r"
                  )
plt.show()
```

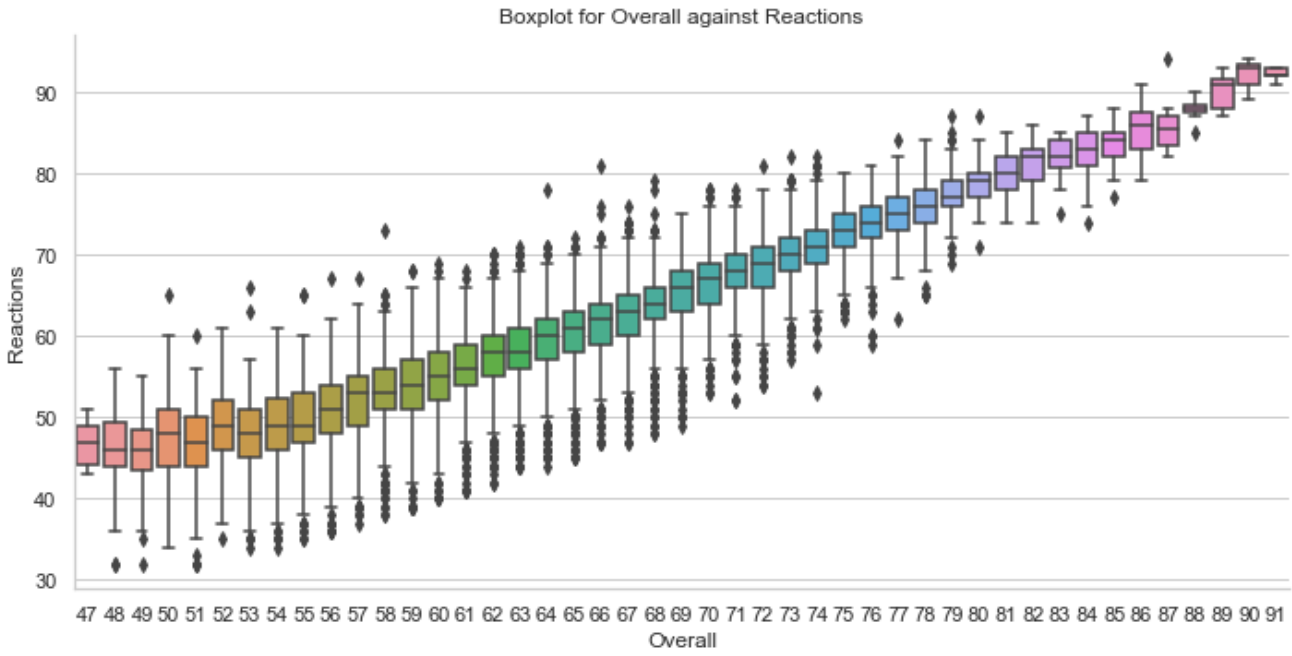




**this heatmap shows third interesting observation which are that there is a high correlation(0.84) between overall and Base stats which makes sense because when a player has high base stats he should have a high overall, the second observation is that there is a high correlation (0.87) between overall and reactions which means that usually players with high overalls have good reactions and know how to read other players which will help when looking for replacements for ronaldo which has 94 reactions. The final observation is that there is high correlation of 0.70 between composure and overall, which shows that good quality players know how to stay composed through out the 90 minutes of play time and not get affected by other factors. These observations will be kept in mind while looking for the player we are searching for.**

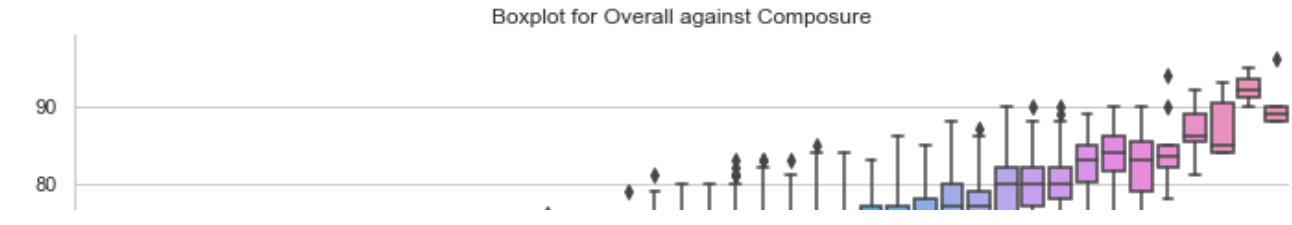
In [456]:

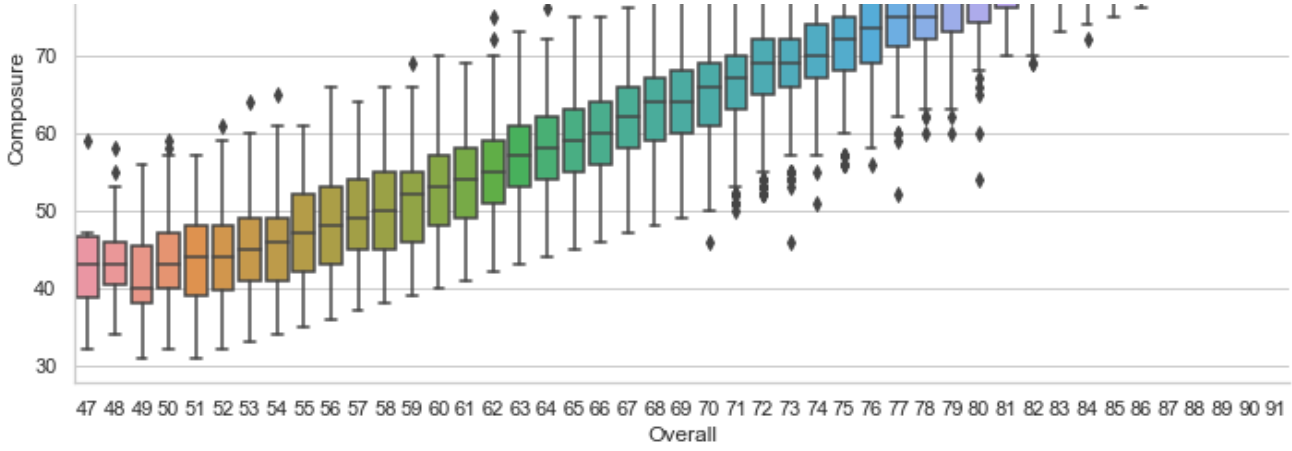
```
sns.catplot(x='Overall', y='Reactions', data = fifa , kind = 'box' , aspect = 2)
plt.title('Boxplot for Overall against Reactions')
plt.show()
```




In [457]:

```
sns.catplot(x='Overall', y='Composure', data = fifa , kind = 'box' , aspect = 2)
plt.title('Boxplot for Overall against Composure')
plt.show()
```






## Classifying positions into 3 different categories

**This step is crucial to avoid confusion and will help simplify the analysis where positions are classified to forward, midfielder, or defender instead of the multiple positions that are in the table Right now**

In [458]:

```
fifa.replace(['ST','RW','LW','CF','RM','LM'], 3,inplace= True) # Forward positions will be known as 3 in the dataset to make things easier for analysis
fifa.replace(['CM','CDM','CAM'], 2, inplace = True) # Midfielder positions will be known as 2 in the dataset
fifa.replace (['CB','RB','LB','RWB','LWB'], 1, inplace = True)# Defender positions will be known as 1 in the dataset to make things easier
```

In [459]:

```
fifa.head(7)
```

Out[459]:

| | Known As | Overall | Potential | Value(in Euro) | Best Position | Nationality | Age | Height(in cm) | Weight(in kg) | TotalStats | ... | Long Shots | Aggress |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | L. Messi | 91 | 91 | 54000000 | 2 | Argentina | 35 | 169 | 67 | 2190 | ... | 91 | |
| 1 | K. Benzema | 91 | 91 | 64000000 | 3 | France | 34 | 185 | 81 | 2147 | ... | 80 | |
| 2 | R. Lewandowski | 91 | 91 | 84000000 | 3 | Poland | 33 | 185 | 81 | 2205 | ... | 84 | |
| 3 | K. De Bruyne | 91 | 91 | 107500000 | 2 | Belgium | 31 | 181 | 70 | 2303 | ... | 91 | |
| 4 | K. Mbappé | 91 | 95 | 190500000 | 3 | France | 23 | 182 | 73 | 2177 | ... | 82 | |
| 5 | M. Salah | 90 | 90 | 115500000 | 3 | Egypt | 30 | 175 | 71 | 2226 | ... | 85 | |
| 8 | Cristiano Ronaldo | 90 | 90 | 41000000 | 3 | Portugal | 37 | 187 | 83 | 2159 | ... | 90 | |

**7 rows × 55 columns**

In [460]:

```
Cristiano = fifa.loc[(fifa['Known As']== 'Cristiano Ronaldo')] #saving ronaldo's row incase I lose him while further cleaning.
```

## Exploring Cristiano's best attributes

In [461]:

```
print(fifa.iloc[6])
```

```
Known As              Cristiano Ronaldo
```

```
Overall                          90
Potential                        90
Value(in Euro)                41000000
Best Position                        3
Nationality                   Portugal
Age                                 37
Height(in cm)                      187
Weight(in kg)                       83
TotalStats                        2159
BaseStats                          445
Club Name           Manchester United
Wage(in Euro)                  220000
Release Clause               77900000
Contract Until                   2023
Joined On                        2021
On Loan                             -
Preferred Foot                  Right
Weak Foot Rating                    4
Skill Moves                         5
Pace Total                         81
Shooting Total                     92
Passing Total                      78
Dribbling Total                    85
Defending Total                    34
Physicality Total                  75
Crossing                           80
Finishing                          93
Heading Accuracy                   90
Short Passing                      80
Volleys                            86
Dribbling                          85
Curve                              81
Freekick Accuracy                  79
LongPassing                        75
BallControl                        88
Acceleration                       79
Sprint Speed                       83
Agility                            77
Reactions                          94
Balance                            67
Shot Power                         93
Jumping                            95
Stamina                            76
Strength                           77
Long Shots                         90
Aggression                         63
Interceptions                      29
Positioning                        95
Vision                             76
Penalties                          90
Composure                          95
Marking                            24
Standing Tackle                    32
Sliding Tackle                     24
Name: 8, dtype: object
```

**This shows that the highest attributes (more than or equal 85) for Cristiano are : Finishing / Heading Accuracy / Volleys / Dribbling / Ball Control / Reactions / Shot Power / Jumping / Long Shots / Positioning / Penalties / Composure**

In [462]:

```
import plotly.express as px
df = fifa[fifa["Known As"] == "Cristiano Ronaldo"]
fig = px.line_polar(df, r=[df['Finishing'].values[0],df['Heading Accuracy'].values[0],
                           df['Volleys'].values[0],df['Dribbling'].values[0],df['BallCon
trol'].values[0],
                           df['Reactions'].values[0],
                          df['Shot Power'].values[0],df['Jumping'].values[0],df['Long Sh
ots'].values[0],df['Positioning'].values[0],df['Penalties'].values[0],df['Composure'].va
```

```
lues[0]],
                      theta=['Finishing','Heading','Volleys','Dribbling','Ball Control','R
eactions','Shot Power',
                                                    'Jumping','Long Shots','Positioning','Pena
lties','Composure'], line_close=True,
                      color_discrete_sequence=px.colors.sequential.Plasma_r,
                      template="plotly_dark",)
fig.show()
```

```
/Users/youssefayman/opt/anaconda3/lib/python3.9/site-packages/plotly/express/_core.py:271
: FutureWarning:

The frame.append method is deprecated and will be removed from pandas in a future version
. Use pandas.concat instead.
```

In [463]:

```
def stats(name):
    return fifa[fifa['Known As'] == name ][['Known As', 'Overall', 'Potential','Best Pos
ition','Age', 'Finishing','Heading Accuracy','Volleys','Dribbling','BallControl','Reactio
ns','Shot Power',
                                                    'Jumping','Long Shots','Positioning','Pena
lties','Composure']]
stats('Cristiano Ronaldo')
```

Out[463]:

| | Known As | Overall | Potential | Best Position | Age | Finishing | Heading Accuracy | Volleys | Dribbling | BallControl | Reactions | Shot Power | Jumping |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 8 | Cristiano Ronaldo | 90 | 90 | 3 | 37 | 93 | 90 | 86 | 85 | 88 | 94 | 93 | 9 |

# Forwards in Manchester United

**Before looking in the whole world, I must look first in the club to see if the option is already inside the club**

Before looking in the whole world, I must look first in the club to see if the option is already inside the club. Furthermore, I'm going to take a look at all the forwards available in Manchester United and compare them to Ronaldo.

In [464]:

```
manutdrep = fifa.loc[(fifa['Club Name']== 'Manchester United') & (fifa['Best Position'] == 3)]
manutdrep
```

Out[464]:

| | Known As | Overall | Potential | Value(in Euro) | Best Position | Nationality | Age | Height(in cm) | Weight(in kg) | TotalStats | ... | Long Shots | Aggressi |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 8 | Cristiano Ronaldo | 90 | 90 | 41000000 | 3 | Portugal | 37 | 187 | 83 | 2159 | ... | 90 | |
| 226 | Antony | 82 | 88 | 49000000 | 3 | Brazil | 22 | 174 | 63 | 2010 | ... | 76 | |
| 323 | M. Rashford | 81 | 85 | 37000000 | 3 | England | 24 | 186 | 70 | 2072 | ... | 78 | |
| 501 | A. Martial | 80 | 83 | 28000000 | 3 | France | 26 | 184 | 76 | 1973 | ... | 74 | |
| 2042 | A. Elanga | 74 | 85 | 9500000 | 3 | Sweden | 20 | 178 | 75 | 1829 | ... | 63 | |
| 4790 | F. Pellistri | 70 | 82 | 3800000 | 3 | Uruguay | 20 | 175 | 65 | 1754 | ... | 61 | |
| 10793 | Alejandro Garnacho | 64 | 85 | 1900000 | 3 | Argentina | 18 | 180 | 70 | 1586 | ... | 56 | |

7 rows × 55 columns

The table above shows that there are only 6 other forwards other than Cristiano. He plays alongside 2 of them which are usually Antony and Rashford. Pellistri, Garnacho and Elanga are still young to take on a huge role like this, of ourse they all have a high potential but they will not affect the team immediately, so this leaves us with Martial.

In [465]:

```
#the dataset wrote sancho as a midfielder but he plays for United as a forward
manutdrep2 = fifa.loc[(fifa['Known As']== 'J. Sancho')]
manutdrep2
```

Out[465]:

| | Known As | Overall | Potential | Value(in Euro) | Best Position | Nationality | Age | Height(in cm) | Weight(in kg) | TotalStats | ... | Long Shots | Aggression |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 144 | J. Sancho | 84 | 88 | 61500000 | 2 | England | 22 | 180 | 76 | 1977 | ... | 65 | 44 |

1 rows × 55 columns

Now we have two players that could potentially take ronaldo's place but lets compare ronaldo's best attributes to theirs.

In [466]:

```
manutd = [manutdrep, manutdrep2]
```

In [467]:

```
replacements = pd.concat(manutd)
```

In [468]:

```
replacements
```

Out[468]:

| | Known As | Overall | Potential | Value(in Euro) | Best Position | Nationality | Age | Height(in cm) | Weight(in kg) | TotalStats | ... | Long Shots | Aggressi |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 8 | Cristiano Ronaldo | 90 | 90 | 41000000 | 3 | Portugal | 37 | 187 | 83 | 2159 | ... | 90 | |
| 226 | Antony | 82 | 88 | 49000000 | 3 | Brazil | 22 | 174 | 63 | 2010 | ... | 76 | |
| 323 | M. Rashford | 81 | 85 | 37000000 | 3 | England | 24 | 186 | 70 | 2072 | ... | 78 | |
| 501 | A. Martial | 80 | 83 | 28000000 | 3 | France | 26 | 184 | 76 | 1973 | ... | 74 | |
| 2042 | A. Elanga | 74 | 85 | 9500000 | 3 | Sweden | 20 | 178 | 75 | 1829 | ... | 63 | |
| 4790 | F. Pellistri | 70 | 82 | 3800000 | 3 | Uruguay | 20 | 175 | 65 | 1754 | ... | 61 | |
| 10793 | Alejandro Garnacho | 64 | 85 | 1900000 | 3 | Argentina | 18 | 180 | 70 | 1586 | ... | 56 | |
| 144 | J. Sancho | 84 | 88 | 61500000 | 2 | England | 22 | 180 | 76 | 1977 | ... | 65 | |

8 rows × 55 columns

In [469]:

```
replacements = replacements[['Known As', 'Overall', 'Potential','Best Position','Age', '
Finishing','Heading Accuracy','Volleys','Dribbling','BallControl','Reactions','Shot Power
',
                                                    'Jumping','Long Shots','Positioning','Pena
lties','Composure']]
```

In [470]:

```
replacements.drop([2042,4790,10793,323,226],axis=0,inplace=True)
```

```
/var/folders/_5/fk2q70ks6cv1r4lh88zxfj480000gn/T/ipykernel_1815/2888388914.py:1: SettingW
ithCopyWarning:


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_g
uide/indexing.html#returning-a-view-versus-a-copy
```

In [471]:

```
#create a replacements table for manchester united replacements
replacements
```

Out[471]:

| | Known As | Overall | Potential | Best Position | Age | Finishing | Heading Accuracy | Volleys | Dribbling | BallControl | Reactions | Shot Power | Jump |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 8 | Cristiano Ronaldo | 90 | 90 | 3 | 37 | 93 | 90 | 86 | 85 | 88 | 94 | 93 | |
| 501 | A. Martial | 80 | 83 | 3 | 26 | 79 | 70 | 76 | 87 | 84 | 78 | 83 | |
| 144 | J. Sancho | 84 | 88 | 2 | 22 | 77 | 38 | 83 | 91 | 88 | 81 | 71 | |

**now that all the potential manchester united replacements are in one table, I'll display some data visualization to see if they really are competent enough to take his place**

**see if they really are competent enough to take his place**

In [472]:

```
import plotly.graph_objects as go
fig = go.Figure()
df= fifa[fifa["Known As"] == "Cristiano Ronaldo"]
fig.add_trace(go.Scatterpolar(
        r = [df['Finishing'].values[0],df['Heading Accuracy'].values[0],
        df['Volleys'].values[0],df['Dribbling'].values[0],df['BallControl'].values[0], d
f['Reactions'].values[0],
        df['Shot Power'].values[0],df['Jumping'].values[0],df['Long Shots'].values[0],
        df['Positioning'].values[0],df['Penalties'].values[0],df['Composure'].values[0]]
,
        theta = ['Finishing','Heading','Volleys','Dribbling','Ball Control','Reactions',
'Shot Power',
                                                'Jumping','Long Shots','Positioning','Pena
lties','Composure'],
        mode = 'lines',
        name = 'Cristiano Ronaldo',
        fill='toself',
        line_color = 'lightcoral'
    ))

s=fifa[fifa["Known As"] == "J. Sancho"]
fig.add_trace(go.Scatterpolar(
        r = [s['Finishing'].values[0],s['Heading Accuracy'].values[0],
        s['Volleys'].values[0],s['Dribbling'].values[0],s['BallControl'].values[0], s['R
eactions'].values[0],
        s['Shot Power'].values[0],s['Jumping'].values[0],s['Long Shots'].values[0],
        s['Positioning'].values[0],s['Penalties'].values[0],s['Composure'].values[0]],

        theta = ['Finishing','Heading','Volleys','Dribbling','Ball Control','Reactions',
'Shot Power',
                                                'Jumping','Long Shots','Positioning','Pena
lties','Composure'],
        mode = 'lines',
        name = 'Jadon Sancho',
        fill='toself',
        line_color = 'darkred'))

m =fifa[fifa["Known As"] == "A. Martial"]
fig.add_trace(go.Scatterpolar(
        r = [m['Finishing'].values[0],m['Heading Accuracy'].values[0],
        m['Volleys'].values[0],m['Dribbling'].values[0],m['BallControl'].values[0], m['R
eactions'].values[0],
        m['Shot Power'].values[0],m['Jumping'].values[0],m['Long Shots'].values[0],
        m['Positioning'].values[0],m['Penalties'].values[0],m['Composure'].values[0]],

        theta = ['Finishing','Heading','Volleys','Dribbling','Ball Control','Reactions',
'Shot Power',
                                                'Jumping','Long Shots','Positioning','Pena
lties','Composure'],
        mode = 'lines',
        name = 'Anthony Martial',
        fill='toself',
        line_color = 'red'
    ))
```

**The graph above shows that those two players do not come close to where Cristiano's attributes are, sancho does exceed him in dribbling, but in other aspects of the game he is far off. On the other hand, Martial also exceeds cristiano in dribbling but is more of a complete forward than Sancho. He's an all around better striker, maybe not the best but he's not missing an attribute like sancho is. All in all, this graph show that they are not competent enough to take cristiano's place so now we will have to start our analysis on the rest of the players**

## Removing irrelevent attributes and players (Data cleaning #2)

**In this part of the analysis, I will be ruling out some players that I know are unsignable by Manchester United for multiple reasons like: -**

- **age**
- **his contract / value**
- **the potential**
- **the position**
- **Current club**

  **As well as, dropping all the manchester united players because we already searched in the club.**

In [473]:

```
fifa=fifa[fifa['Club Name'] != 'Manchester United']
```

**Generally in football, Clubs are always looking to sign young talents because they have potential to grow and their longevity, so because of this we will narrow down the players to 28 years or younger. a player being 28 years and signing a five year contract would mean he'll play five year at his peak/highest level.**

In [474]:

```
fifa=fifa[fifa['Age']<=28]
```

**Another factor that would affect the transfer is if the player plays for either Manchester City or Liverpool who are Manchester United's direct rivals and it's highly unlikely that a direct transfer would happen with any of those clubs.**

In [475]:

```
fifa=fifa[fifa['Club Name'] != 'Manchester City']
fifa=fifa[fifa['Club Name'] != 'Liverpool']
```

**In this analysis, we're looking for a replacement for Cristiano which is a forward so it doesn't make sense to have midfielders and defenders in this dataset.**

In [476]:

```
fifa=fifa[fifa['Best Position'] != 2]
fifa=fifa[fifa['Best Position'] != 1]
```

In [477]:

```
fifa.shape
```

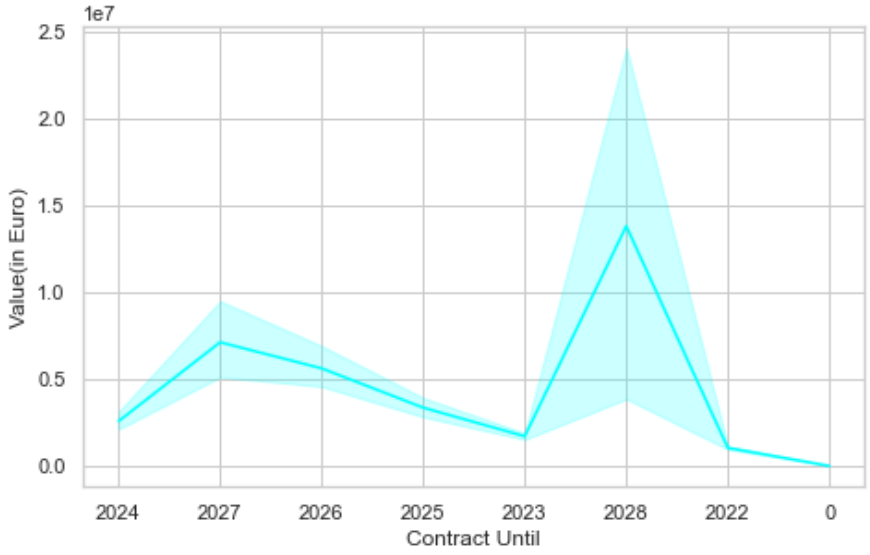
Out[477]:

```
(4129, 55)
```

**Players contract tend to have a relationship with the value. If a contract is still new or not terminating any time soon, the player tends to have a higher value. Especially in recent years, players are signing contracts with an unimaginably high release clause so the following graph will show the truth to this theory.**

In [478]:

```
g = sns.lineplot(x="Contract Until", y="Value(in Euro)", data=fifa,color='cyan')
```




**as this plot shows the prices spike at 2028 and 2027 which means the player that we would be lokking for would have a contract that ends within the next 4 years or by the end of 2026. In addition, recent news show that the transfer budget for manchester united next season would be around 100 million british pounds which should be around 116 million euros, so we will remove players with higher values.**

In [479]:

```
fifa=fifa[fifa['Contract Until'] != '2028']
fifa=fifa[fifa['Contract Until'] != '2027']
fifa=fifa[fifa['Value(in Euro)'] <= 116000000]
```

In [480]:

```
fifa.shape
```

Out[480]:

```
(3963, 55)
```

**Of course,cristiano fits into some of the exceptions that we dropped like being a manchester united player and being over 28, so I will add him again to the data frame with the exact same attributes.**

In [481]:

```
fifa = fifa.append(Cristiano)
```

```
/var/folders/_5/fk2q70ks6cv1r4lh88zxfj480000gn/T/ipykernel_1815/706814428.py:1: FutureWarning:

The frame.append method is deprecated and will be removed from pandas in a future version
```

```
. Use pandas.concat instead.
```

# Model Building

## K-means clustering

In [482]:

```
stats = ['Weak Foot Rating','Skill Moves','Crossing','Finishing','Heading Accuracy',

'Short Passing','Volleys','Dribbling','Curve','Freekick Accuracy','LongPassing','BallCont
rol','Acceleration',
'Sprint Speed' ,'Agility','Reactions','Balance' ,'Shot Power','Jumping','Stamina' ,'Stre
ngth' ,'Long Shots','Aggression' ,'Interceptions','Positioning','Vision','Penalties','Com
posure' ,'Marking','Standing Tackle' ,'Sliding Tackle'    ]
```

In [483]:

```
df = fifa[['Weak Foot Rating','Skill Moves','Crossing','Finishing','Heading Accuracy',

'Short Passing','Volleys','Dribbling','Curve','Freekick Accuracy','LongPassing','BallCont
rol','Acceleration',
'Sprint Speed' ,'Agility','Reactions','Balance' ,'Shot Power','Jumping','Stamina' ,'Stre
ngth' ,'Long Shots','Aggression' ,'Interceptions','Positioning','Vision','Penalties','Com
posure' ,'Marking','Standing Tackle' ,'Sliding Tackle'    ]]
```

In [484]:

```
df.head()
```

Out[484]:

| | Weak Foot Rating | Skill Moves | Crossing | Finishing | Heading Accuracy | Short Passing | Volleys | Dribbling | Curve | Freekick Accuracy | ... | Long Shots | Aggression | Interc |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **10** | 5 | 3 | 80 | 93 | 85 | 84 | 87 | 83 | 82 | 65 | ... | 86 | 80 | |
| **59** | 4 | 4 | 58 | 88 | 84 | 76 | 90 | 84 | 78 | 54 | ... | 76 | 90 | |
| **63** | 4 | 5 | 72 | 84 | 50 | 75 | 72 | 92 | 77 | 62 | ... | 76 | 58 | |
| **67** | 4 | 4 | 78 | 86 | 42 | 82 | 83 | 87 | 74 | 62 | ... | 82 | 63 | |
| **75** | 3 | 5 | 84 | 84 | 65 | 81 | 75 | 87 | 85 | 83 | ... | 83 | 72 | |

5 rows × 31 columns

In [485]:

```
from sklearn import preprocessing
x = df.values
scaler = preprocessing.MinMaxScaler()
x_scaled = scaler.fit_transform(x)
X_norm = pd.DataFrame(x_scaled)
```

In [486]:

```
from sklearn.decomposition import PCA
pca = PCA(n_components = 2)
r = pd.DataFrame(pca.fit_transform(X_norm))
```

In [487]:

```
from sklearn.cluster import KMeans

# specify the number of clusters
```

```
kmeans = KMeans(n_clusters=8, init='random', random_state=0)
# fit the data
kmeans = kmeans.fit(r)
# get the cluster labels
name= fifa['Known As'].tolist()
labels = kmeans.predict(r)
# centroid values
centroid = kmeans.cluster_centers_
# cluster values
clusters = kmeans.labels_.tolist()
```

In [488]:

```
r['cluster'] = clusters
r['names'] = name
fifa['cluster'] = clusters
r.columns = ['x','y','cluster','names']
r.head()
```

Out[488]:

| | x | y | cluster | names |
|---|---|---|---|---|
| **0** | **1.547070** | **1.112440** | **1** | **H. Kane** |
| **1** | **1.480716** | **0.873494** | **1** | **L. Martínez** |
| **2** | **1.254202** | **0.730178** | **1** | **Vinícius Jr.** |
| **3** | **1.420896** | **0.460077** | **4** | **S. Gnabry** |
| **4** | **1.420196** | **1.000743** | **1** | **M. Depay** |

In [489]:

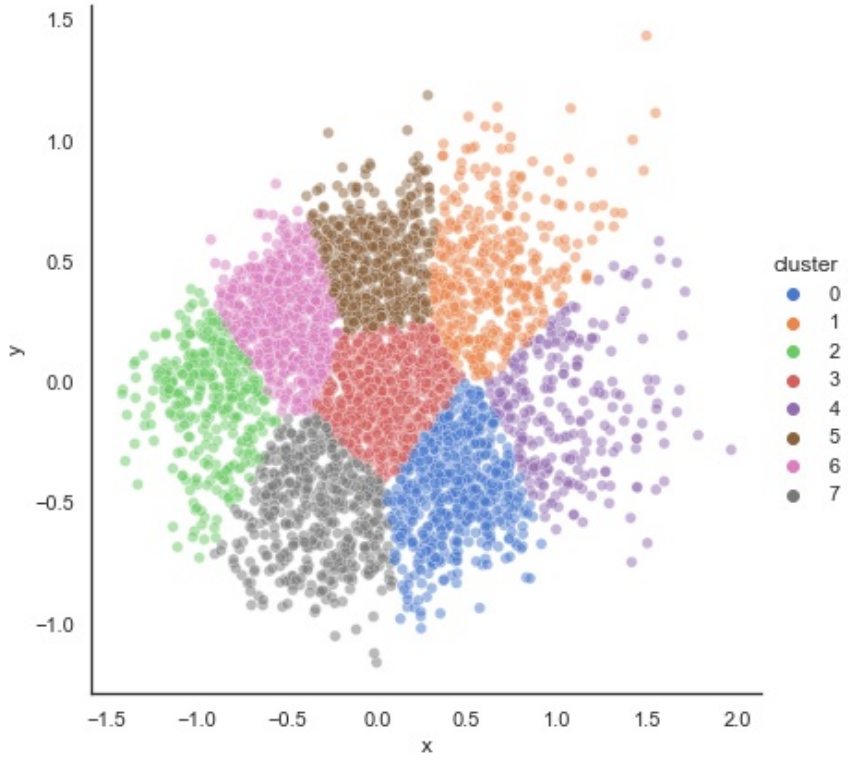
```
import seaborn as sns
sns.set_theme(style="white")



# Plot x against y with other semantics
sns.relplot(x="x", y="y", hue="cluster",
            sizes=(40, 400), alpha=.5, palette="muted",
            height=6, data=r)
```

Out[489]:

```
<seaborn.axisgrid.FacetGrid at 0x7fd1db5f1b80>
```



In [495]:

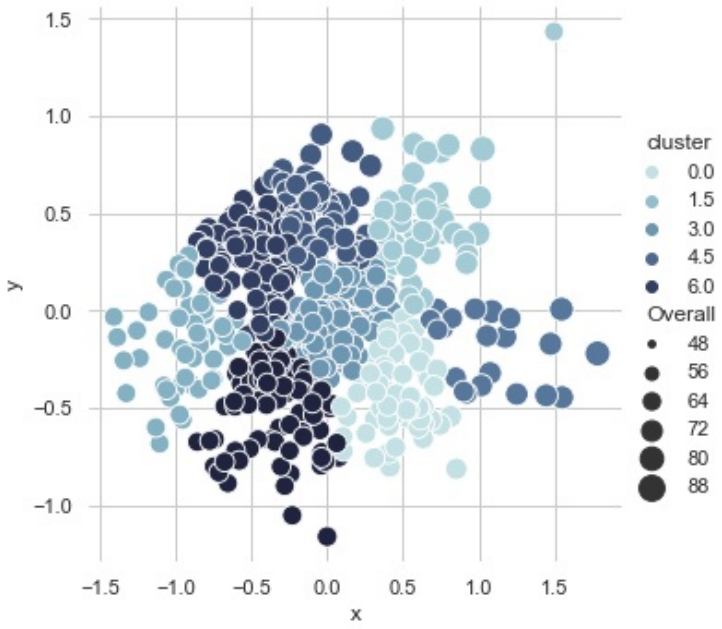
```
import seaborn as sns
sns.set_theme(style="whitegrid")


cmap = sns.cubehelix_palette(rot=-.2, as_cmap=True)
g = sns.relplot(
    data=r,
    x="x", y="y",
    hue="cluster", size=fifa['Overall'],
    palette=cmap, sizes=(10, 200),
)

g.ax.xaxis.grid(True, "minor", linewidth=.25)
g.ax.yaxis.grid(True, "minor", linewidth=.25)
g.despine(left=True, bottom=True)
```

Out[495]:

```
<seaborn.axisgrid.FacetGrid at 0x7fd1f2351040>
```




In [414]:

```
# Find Cristiano's Cluster
def Cluster(x):
    return fifa[fifa['Known As'] == x][['Known As', 'cluster']]
Cluster('Cristiano Ronaldo')
```

Out[414]:

| | Known As | cluster |
|---|---|---|
| 8 | Cristiano Ronaldo | 1 |

In [415]:

```
fifa[fifa['cluster'] == 1].head(12)
```

Out[415]:

| | Known As | Overall | Potential | Value(in Euro) | Best Position | Nationality | Age | Height(in cm) | Weight(in kg) | TotalStats | ... | Aggression | Int |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 10 | H. Kane | 89 | 89 | 105500000 | 3 | England | 28 | 188 | 89 | 2193 | ... | 80 | |
| 59 | L. Martínez | 86 | 90 | 98500000 | 3 | Argentina | 24 | 174 | 72 | 2163 | ... | 90 | |
| | Vinícius | | | | | | | | | | | | |

| | Known As | Overall | Potential | Value(in Euro) | Best Position | Nationality | Age | Height(in cm) | Weight(in kg) | TotalStats | ... | Aggression | Int |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 63 | Vinícius Jr. | 86 | 92 | 109000000 | 3 | Brazil | 21 | 176 | 73 | 1985 | ... | 58 | |
| 75 | M. Depay | 85 | 85 | 55000000 | 3 | Netherlands | 28 | 176 | 78 | 2089 | ... | 72 | |
| 128 | Rafael Leão | 84 | 90 | 66500000 | 3 | Portugal | 23 | 188 | 81 | 1971 | ... | 60 | |
| 135 | L. Sané | 84 | 85 | 49500000 | 3 | Germany | 26 | 183 | 80 | 2086 | ... | 63 | |
| 147 | Oyarzabal | 84 | 86 | 53000000 | 3 | Spain | 25 | 181 | 79 | 2044 | ... | 52 | |
| 152 | D. Vlahović | 84 | 91 | 86500000 | 3 | Serbia | 22 | 190 | 75 | 1912 | ... | 60 | |
| 166 | V. Osimhen | 83 | 89 | 58000000 | 3 | Nigeria | 23 | 185 | 78 | 1999 | ... | 69 | |
| 173 | C. Gakpo | 83 | 87 | 52000000 | 3 | Netherlands | 23 | 189 | 76 | 2057 | ... | 60 | |
| 206 | Gonçalo Guedes | 82 | 85 | 41000000 | 3 | Portugal | 25 | 179 | 68 | 2034 | ... | 64 | |
| 212 | T. Werner | 82 | 84 | 38000000 | 3 | Germany | 26 | 180 | 76 | 1994 | ... | 63 | |

12 rows × 56 columns

In [416]:

```
final_replacements = fifa[fifa['cluster'] == 1]
```

In [417]:

```
final_replacements.shape
```

Out[417]:

```
(396, 56)
```

In [418]:

```
final_replacements=final_replacements[final_replacements['Potential']>=85]
```

In [419]:

```
final_replacements.shape
```

Out[419]:

```
(27, 56)
```

In [420]:

```
final_replacements
```

Out[420]:

| | Known As | Overall | Potential | Value(in Euro) | Best Position | Nationality | Age | Height(in cm) | Weight(in kg) | TotalStats | ... | Aggression |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 10 | H. Kane | 89 | 89 | 105500000 | 3 | England | 28 | 188 | 89 | 2193 | ... | 80 |
| 59 | L. Martínez | 86 | 90 | 98500000 | 3 | Argentina | 24 | 174 | 72 | 2163 | ... | 90 |
| 63 | Vinícius Jr. | 86 | 92 | 109000000 | 3 | Brazil | 21 | 176 | 73 | 1985 | ... | 58 |
| 75 | M. Depay | 85 | 85 | 55000000 | 3 | Netherlands | 28 | 176 | 78 | 2089 | ... | 72 |
| 128 | Rafael Leão | 84 | 90 | 66500000 | 3 | Portugal | 23 | 188 | 81 | 1971 | ... | 60 |
| 135 | L. Sané | 84 | 85 | 49500000 | 3 | Germany | 26 | 183 | 80 | 2086 | ... | 63 |
| 147 | Oyarzabal | 84 | 86 | 53000000 | 3 | Spain | 25 | 181 | 79 | 2044 | ... | 52 |
| 152 | D. Vlahović | 84 | 91 | 86500000 | 3 | Serbia | 22 | 190 | 75 | 1912 | ... | 60 |
| 166 | V. Osimhen | 83 | 89 | 58000000 | 3 | Nigeria | 23 | 185 | 78 | 1999 | ... | 69 |

| | Known As | Overall | Potential | Value(in Euro) | Best Position | Nationality | Age | Height(in cm) | Weight(in kg) | TotalStats | ... | Aggression |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 173 | C. Gakpo | 83 | 87 | 52000000 | 3 | Netherlands | 23 | 189 | 76 | 2057 | ... | 60 |
| 206 | Gonçalo Guedes | 82 | 85 | 41000000 | 3 | Portugal | 25 | 179 | 68 | 2034 | ... | 64 |
| 221 | T. Abraham | 82 | 85 | 42000000 | 3 | England | 24 | 195 | 86 | 1927 | ... | 68 |
| 374 | Y. En-Nesyri | 80 | 85 | 31500000 | 3 | Morocco | 25 | 188 | 78 | 1881 | ... | 76 |
| 415 | N. González | 80 | 85 | 32000000 | 3 | Argentina | 24 | 180 | 72 | 1971 | ... | 67 |
| 512 | J. David | 79 | 85 | 28500000 | 3 | Canada | 22 | 180 | 77 | 1912 | ... | 37 |
| 537 | S. Chukwueze | 79 | 85 | 27500000 | 3 | Nigeria | 23 | 172 | 70 | 1898 | ... | 47 |
| 578 | D. Malen | 79 | 85 | 28000000 | 3 | Netherlands | 23 | 178 | 77 | 2018 | ... | 62 |
| 690 | Evanilson | 78 | 86 | 32000000 | 3 | Brazil | 22 | 181 | 77 | 1796 | ... | 45 |
| 763 | M. Kean | 78 | 86 | 32000000 | 3 | Italy | 22 | 182 | 72 | 1878 | ... | 57 |
| 1483 | B. Johnson | 75 | 85 | 12500000 | 3 | Wales | 21 | 181 | 71 | 1830 | ... | 48 |
| 1652 | K. Sulemana | 75 | 85 | 12500000 | 3 | Ghana | 20 | 175 | 70 | 1817 | ... | 52 |
| 1661 | Gonçalo Ramos | 75 | 85 | 12500000 | 3 | Portugal | 21 | 185 | 82 | 1900 | ... | 80 |
| 1759 | Gonçalo Ramos | 75 | 85 | 12500000 | 3 | Portugal | 21 | 185 | 82 | 1900 | ... | 80 |
| 2073 | Nico Williams | 74 | 86 | 10000000 | 3 | Spain | 19 | 181 | 67 | 1800 | ... | 36 |
| 3554 | B. Šeško | 72 | 85 | 5500000 | 3 | Slovenia | 19 | 193 | 77 | 1816 | ... | 62 |
| 5658 | Y. Moukoko | 69 | 88 | 3500000 | 3 | Germany | 17 | 175 | 73 | 1733 | ... | 43 |
| 8 | Cristiano Ronaldo | 90 | 90 | 41000000 | 3 | Portugal | 37 | 187 | 83 | 2159 | ... | 63 |

27 rows × 56 columns

**There are 27 possible replacements that were found using the K-means clustering algorithm with potential 85 or higher which is good, but we need players that have high potential for the future as well as immediate effect that will help the team so players like Y. Moukoko who is 17 years old and potential of 88 is perfect for the future but his rating now is still 69 so he won't be able to help the team right away. to solve this,we will drop all players with overall less that 83.**

In [421]:

```
final_replacements=final_replacements[final_replacements['Overall'] >= 83]
final_replacements=final_replacements[final_replacements['Known As']!= 'Cristiano Ronaldo
']
```

In [422]:

```
final_replacements
```

Out[422]:

| | Known As | Overall | Potential | Value(in Euro) | Best Position | Nationality | Age | Height(in cm) | Weight(in kg) | TotalStats | ... | Aggression | Int |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 10 | H. Kane | 89 | 89 | 105500000 | 3 | England | 28 | 188 | 89 | 2193 | ... | 80 | |
| 59 | L. Martínez | 86 | 90 | 98500000 | 3 | Argentina | 24 | 174 | 72 | 2163 | ... | 90 | |
| 63 | Vinícius Jr. | 86 | 92 | 109000000 | 3 | Brazil | 21 | 176 | 73 | 1985 | ... | 58 | |
| 75 | M. Depay | 85 | 85 | 55000000 | 3 | Netherlands | 28 | 176 | 78 | 2089 | ... | 72 | |

| | Known As | Overall | Potential | Value(in Euro) | Best Position | Nationality | Age | Height(in cm) | Weight(in kg) | TotalStats | ... | Aggression | Int |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 128 | Rafael Leão | 84 | 90 | 66500000 | 3 | Portugal | 23 | 188 | 81 | 1971 | ... | 60 | |
| 135 | L. Sané | 84 | 85 | 49500000 | 3 | Germany | 26 | 183 | 80 | 2086 | ... | 63 | |
| 147 | Oyarzabal | 84 | 86 | 53000000 | 3 | Spain | 25 | 181 | 79 | 2044 | ... | 52 | |
| 152 | D. Vlahović | 84 | 91 | 86500000 | 3 | Serbia | 22 | 190 | 75 | 1912 | ... | 60 | |
| 166 | V. Osimhen | 83 | 89 | 58000000 | 3 | Nigeria | 23 | 185 | 78 | 1999 | ... | 69 | |
| 173 | C. Gakpo | 83 | 87 | 52000000 | 3 | Netherlands | 23 | 189 | 76 | 2057 | ... | 60 | |

10 rows × 56 columns

**Now that we have the final 10 rows of players that can potentially replace ronaldo. let's eliminate a couple that might not be fit for the role like the following players: -**

**1 - Harry Kane / He is the star of tottenham and they will not let him go to a rival team easily.**

**2 - Vinicius junior / he is not the player that manchester united need as he is small and injury prone.**

**3 - Memphis Depay / he was in Manchester United a long time ago but he left on bad terms so seems unlikely to join again.**

**4 - Oyarzabal / Similar to Vinicius jr. he doesn't have a striker build as well as playing as a winger**

In [423]:

```
final_replacements = final_replacements[final_replacements['Known As']!='H. Kane']
final_replacements = final_replacements[final_replacements['Known As']!='Vinícius Jr.']
final_replacements = final_replacements[final_replacements['Known As']!='M. Depay']
final_replacements = final_replacements[final_replacements['Known As']!='Oyarzabal']
```

In [424]:

```
final_replacements
```

Out[424]:

| | Known As | Overall | Potential | Value(in Euro) | Best Position | Nationality | Age | Height(in cm) | Weight(in kg) | TotalStats | ... | Aggression | Inter |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 59 | L. Martínez | 86 | 90 | 98500000 | 3 | Argentina | 24 | 174 | 72 | 2163 | ... | 90 | |
| 128 | Rafael Leão | 84 | 90 | 66500000 | 3 | Portugal | 23 | 188 | 81 | 1971 | ... | 60 | |
| 135 | L. Sané | 84 | 85 | 49500000 | 3 | Germany | 26 | 183 | 80 | 2086 | ... | 63 | |
| 152 | D. Vlahović | 84 | 91 | 86500000 | 3 | Serbia | 22 | 190 | 75 | 1912 | ... | 60 | |
| 166 | V. Osimhen | 83 | 89 | 58000000 | 3 | Nigeria | 23 | 185 | 78 | 1999 | ... | 69 | |
| 173 | C. Gakpo | 83 | 87 | 52000000 | 3 | Netherlands | 23 | 189 | 76 | 2057 | ... | 60 | |

6 rows × 56 columns

# Who will replace Cristiano Ronaldo? (Final)

In [425]:

```
final_replacements
```

Out[425]:

| | Known As | Overall | Potential | Value(in Euro) | Best Position | Nationality | Age | Height(in cm) | Weight(in kg) | TotalStats | ... | Aggression | Inter |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 59 | L. Martínez | 86 | 90 | 98500000 | 3 | Argentina | 24 | 174 | 72 | 2163 | ... | 90 | |
| 128 | Rafael Leão | 84 | 90 | 66500000 | 3 | Portugal | 23 | 188 | 81 | 1971 | ... | 60 | |
| 135 | L. Sané | 84 | 85 | 49500000 | 3 | Germany | 26 | 183 | 80 | 2086 | ... | 63 | |
| 152 | D. Vlahović | 84 | 91 | 86500000 | 3 | Serbia | 22 | 190 | 75 | 1912 | ... | 60 | |
| 166 | V. Osimhen | 83 | 89 | 58000000 | 3 | Nigeria | 23 | 185 | 78 | 1999 | ... | 69 | |
| 173 | C. Gakpo | 83 | 87 | 52000000 | 3 | Netherlands | 23 | 189 | 76 | 2057 | ... | 60 | |

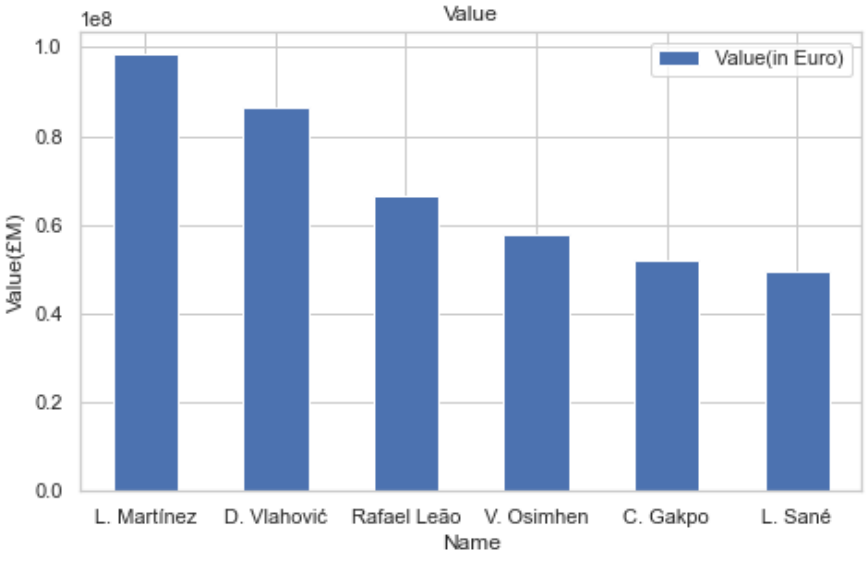
6 rows × 56 columns

**The table above leaves us with 6 candidates to replace this icon of football. lets compare some of their factors and attributes to see who would be the best and final choice.**

### Values

In [426]:

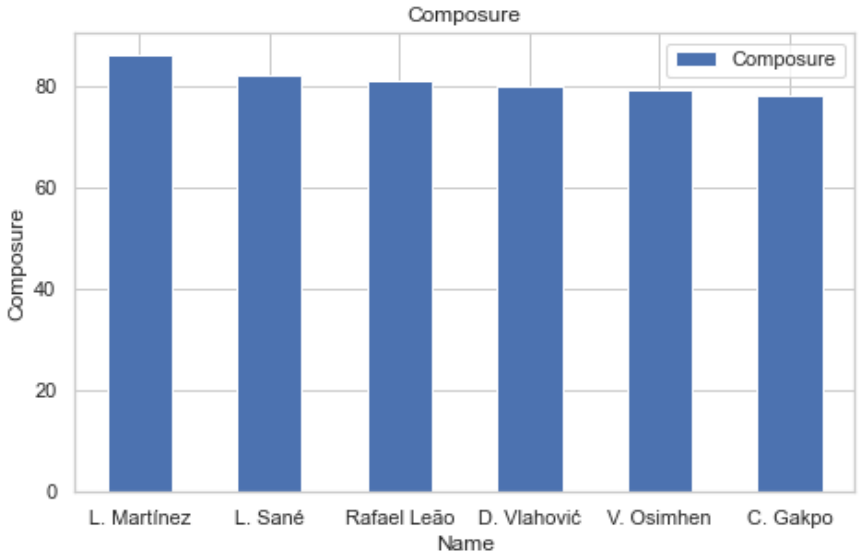
```
final_replacements_sorted= final_replacements.sort_values('Value(in Euro)',ascending=Fals
e)
ax = final_replacements_sorted.plot.bar(x='Known As', y='Value(in Euro)', rot=0)
plt.title('Value')
plt.xlabel('Name')
plt.ylabel('Value(£M)')
plt.show()
```




### Composure

In [496]:

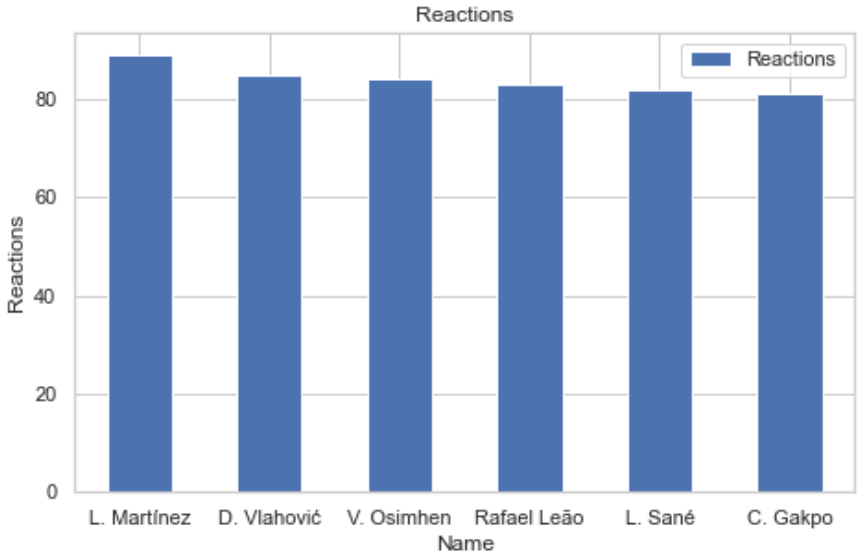
```
final_replacements_sorted= final_replacements.sort_values('Composure',ascending=False)
ax = final_replacements_sorted.plot.bar(x='Known As', y='Composure', rot=0)
plt.title('Composure')
plt.xlabel('Name')
plt.ylabel('Composure')
plt.show()
```




### Reactions

In [428]:

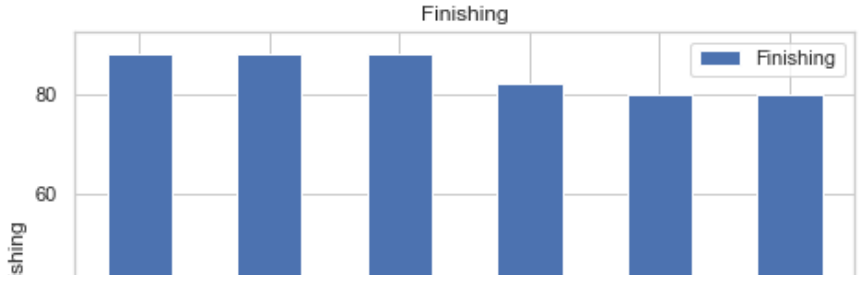
```
final_replacements_sorted= final_replacements.sort_values('Reactions',ascending=False)
ax = final_replacements_sorted.plot.bar(x='Known As', y='Reactions', rot=0)
plt.title('Reactions')
plt.xlabel('Name')
plt.ylabel('Reactions')
plt.show()
```

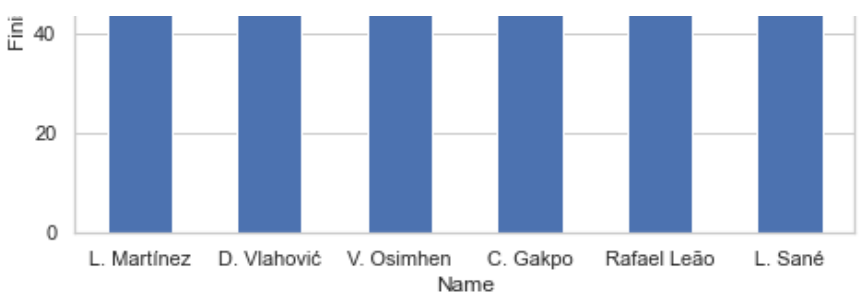



### Finishing

In [429]:

```
final_replacements_sorted= final_replacements.sort_values('Finishing',ascending=False)
ax = final_replacements_sorted.plot.bar(x='Known As', y='Finishing', rot=0)
plt.title('Finishing')
plt.xlabel('Name')
plt.ylabel('Finishing')
plt.show()
```

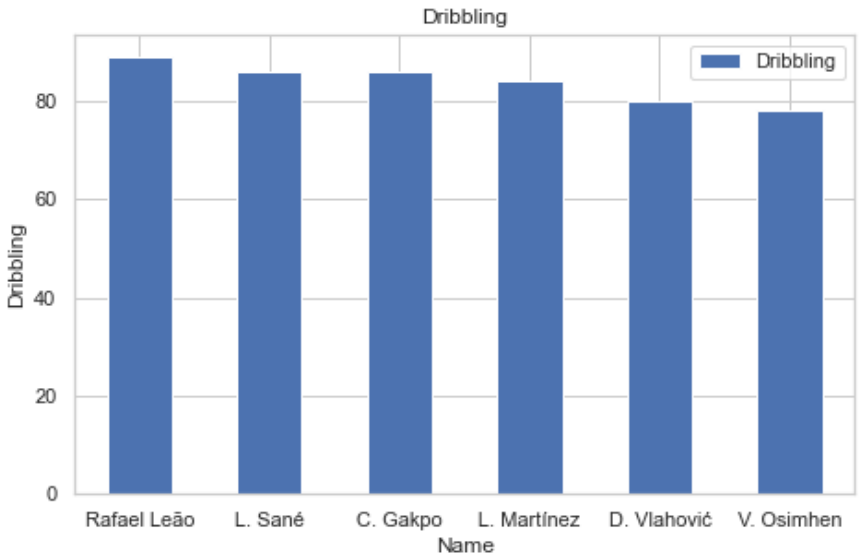





### Dribbling

In [430]:

```
final_replacements_sorted= final_replacements.sort_values('Dribbling',ascending=False)
ax = final_replacements_sorted.plot.bar(x='Known As', y='Dribbling', rot=0)
plt.title('Dribbling')
plt.xlabel('Name')
plt.ylabel('Dribbling')
plt.show()
```




In [431]:

```
final_replacements
```

Out[431]:

| | Known As | Overall | Potential | Value(in Euro) | Best Position | Nationality | Age | Height(in cm) | Weight(in kg) | TotalStats | ... | Aggression | Inter |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 59 | L. Martínez | 86 | 90 | 98500000 | 3 | Argentina | 24 | 174 | 72 | 2163 | ... | 90 | |
| 128 | Rafael Leão | 84 | 90 | 66500000 | 3 | Portugal | 23 | 188 | 81 | 1971 | ... | 60 | |
| 135 | L. Sané | 84 | 85 | 49500000 | 3 | Germany | 26 | 183 | 80 | 2086 | ... | 63 | |
| 152 | D. Vlahović | 84 | 91 | 86500000 | 3 | Serbia | 22 | 190 | 75 | 1912 | ... | 60 | |
| 166 | V. Osimhen | 83 | 89 | 58000000 | 3 | Nigeria | 23 | 185 | 78 | 1999 | ... | 69 | |
| 173 | C. Gakpo | 83 | 87 | 52000000 | 3 | Netherlands | 23 | 189 | 76 | 2057 | ... | 60 | |

6 rows × 56 columns

so in the past graphs I compared all of the players' values, composure, reactions, finishing and dribbling. They were all really similar as they came from the same clsuter but some players exceeded the others like lautaro

**martinez. Which would be my first choice if he was not very expensive and his body type was taller (he is 174 cm and Cristiano was 187cm)**

**The final choice of replacement for me was Dusan Vlahovic where he has the highest potential of 91 which exceeds Cristiano, has a tall body type, he is only 22 years old, and exceeds in the needed attributes like reactions and finishing. Although his price might be a bit high, but the club is getting what it is paying for.**

**On the other hand if the club is looking for a cheaper alternative, I would recommend Cody Gakpo. He is still young as well, has high growth potential, has a tall body type, and has TotalStats more than most of the other players. His attributes are on par with the other players and surely will get better in the future.**

In [432]:

```
no_name=final_replacements[['Weak Foot Rating','Skill Moves','Crossing','Finishing','Head
ing Accuracy',
'Short Passing','Volleys','Dribbling','Curve','Freekick Accuracy','LongPassing','BallCont
rol','Acceleration',
'Sprint Speed' ,'Agility','Reactions','Balance' ,'Shot Power','Jumping','Stamina' ,'Stre
ngth' ,'Long Shots','Aggression' ,'Interceptions','Positioning','Vision','Penalties','Com
posure' ,'Marking','Standing Tackle' ,'Sliding Tackle'    ]]
```

In [271]:

```
import plotly.graph_objects as go
def Convert(lst):
    return [ -i for i in lst ]
sr3 = Convert(no_name.iloc[6])
fig = go.Figure()
fig.add_trace(go.Bar(x=no_name.iloc[3], y=stats,orientation='h',
                base=0,
                marker_color='rgb(158,202,225)',
                name='Dusan Vlahovic',
                marker_line_color='rgb(8,48,107)',
                marker_line_width=1.5,
                opacity= 0.7,
                text = no_name.iloc[3],
                textposition='outside'
                ))
fig.add_trace(go.Bar(x=no_name.iloc[6], y=stats,orientation='h',
                base=sr3,
                marker_color='crimson',
                name='Cristiano Ronaldo',
                marker_line_color='red',
                marker_line_width=1.5,
                opacity= 0.7,
                text = sr3,
                textposition='auto'
                ))
fig.update_layout(
    height=500,
    title_text='Cristiano vs Vlahovic',
    barmode='overlay',xaxis_tickangle=-45,bargap=0.30
    )
fig.show()
```

In [270]:

```
import plotly.graph_objects as go
def Convert(lst):
    return [ -i for i in lst ]
sr3 = Convert(no_name.iloc[6])
fig = go.Figure()
fig.add_trace(go.Bar(x=no_name.iloc[5], y=stats,orientation='h',
                base=0,
                marker_color='rgb(158,202,225)',
                name='Cody Gakpo',
                marker_line_color='rgb(8,48,107)',
                marker_line_width=1.5,
                opacity= 0.7,
                text = no_name.iloc[5],
                textposition='outside'
                ))
fig.add_trace(go.Bar(x=no_name.iloc[6], y=stats,orientation='h',
                base=sr3,
                marker_color='crimson',
                name='Cristiano Ronaldo',
                marker_line_color='red',
                marker_line_width=1.5,
                opacity= 0.7,
                text = sr3,
                textposition='auto'
                ))
fig.update_layout(
    height=500,
    title_text='Cristiano vs Gakpo',
    barmode='overlay',xaxis_tickangle=-45,bargap=0.30
    )
fig.show()
```

In [ ]: