

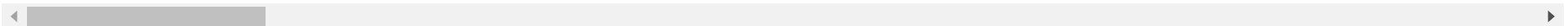
```
In [172...
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sb
```

```
In [173...
import pandas as pd
football = pd.read_csv(r'/reformed_dataset.csv', encoding='iso-8859-1')
football.head()
```

```
Out[173...
      Unnamed: 0  Unnamed: 0.1  ID  Name  Age  Photo  Nationality  Flag  Overall  Potent
```

0	0	0	158023	L. Messi	31	https://cdn.sofifa.org/players/4/19/158023.png	Argentina	https://cdn.sofifa.org/flags/52.png	94	
1	1	1	20801	Cristiano Ronaldo	33	https://cdn.sofifa.org/players/4/19/20801.png	Portugal	https://cdn.sofifa.org/flags/38.png	94	
2	2	2	190871	Neymar Jr	26	https://cdn.sofifa.org/players/4/19/190871.png	Brazil	https://cdn.sofifa.org/flags/54.png	92	
3	3	3	193080	De Gea	27	https://cdn.sofifa.org/players/4/19/193080.png	Spain	https://cdn.sofifa.org/flags/45.png	91	
4	4	4	192985	K. De Bruyne	27	https://cdn.sofifa.org/players/4/19/192985.png	Belgium	https://cdn.sofifa.org/flags/7.png	91	

5 rows × 90 columns



```
In [174...
df = football[['Name', 'Age', 'Overall', 'Potential', 'Value', 'Wage', 'Special', 'Internationality',
               'Weak Foot', 'Skill Moves', 'Jersey Number', 'Height', 'LS',
               'ST', 'RS', 'LW', 'LF', 'CF', 'RF', 'RW', 'LAM', 'CAM', 'RAM', 'LM', 'LCM', 'CM', 'RCM',
               'RM', 'LWB', 'LDM', 'CDM', 'RDM', 'RWB', 'LB', 'LCB', 'CB', 'RCB', 'RB', 'Crossing',
               'Finishing', 'HeadingAccuracy', 'ShortPassing', 'Volleys', 'Dribbling', 'Curve', 'FKAccuracy',
               'LongPassing', 'BallControl', 'Acceleration', 'SprintSpeed', 'Agility', 'Reactions', 'Balance',
               'ShotPower', 'Jumping', 'Stamina', 'Strength', 'LongShots', 'Aggression', 'Interceptions',
```

```

        'Positioning', 'Vision', 'Penalties', 'Composure', 'Marking', 'StandingTackle', 'SlidingTackle', 'GK
        'GKHandling', 'GK Kicking', 'GK Positioning', 'GK Reflexes', 'Release Clause'

    ]]

df = df[df.Overall > 86]
names = df.Name.tolist()

df = df.drop(['Name'], axis = 1)
df.head()

```

Out[174...

	Age	Overall	Potential	Value	Wage	Special	International Reputation	Weak Foot	Skill Moves	Jersey Number	Height	LS	ST	RS	LW	LF
0	31	94	94	110.5	0.565	2202	5.0	4.0	4.0	10.0	170.18	57.81547	57.81547	57.81547	59.03765	58.71939
1	33	94	94	77.0	0.405	2228	5.0	4.0	5.0	7.0	187.96	57.81547	57.81547	57.81547	59.03765	58.71939
2	26	92	93	118.5	0.290	2143	5.0	5.0	5.0	10.0	175.26	57.81547	57.81547	57.81547	59.03765	58.71939
3	27	91	93	72.0	0.260	1471	4.0	3.0	1.0	1.0	193.04	57.81547	57.81547	57.81547	59.03765	58.71939
4	27	91	92	102.0	0.355	2281	4.0	5.0	4.0	7.0	154.94	57.81547	57.81547	57.81547	59.03765	58.71939



In [175...

```

from sklearn import preprocessing

x = df.values # numpy array
mmScaler = preprocessing.MinMaxScaler()
xscaled = scaler.fit_transform(x)
Xnormalized = pd.DataFrame(xscaled)

```

In [178...

```

from sklearn.decomposition import PCA

pca = PCA(n_components = 2)
transform = pd.DataFrame(pca.fit_transform(Xnormalized))

```

In [179...

```

from sklearn.cluster import DBSCAN
db = DBSCAN(eps=1, min_samples=5)
db_clusters = db.fit_predict(transform)

```

In [180... `print(db_clusters)`

```
[0 0 0 1 0 0 0 0 0 1 0 0 0 0 0 0 0 1 1 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0
 1 0 0 1 1 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0]
```

In [181... `transform['cluster'] = db_clusters`
`transform['Name'] = names`
`transform.columns = ['x', 'y', 'cluster', 'Name']`
`transform.head()`

Out[181...

	x	y	cluster	Name
0	-1.237459	-1.390348	0	L. Messi
1	-1.082082	-0.972013	0	Cristiano Ronaldo
2	-1.099776	-1.277386	0	Neymar Jr
3	2.893399	-0.636772	1	De Gea
4	-1.184405	-0.340630	0	K. De Bruyne

In [187... `import matplotlib.pyplot as plt`
`import seaborn as sb`
`%matplotlib inline`

`ax = sb.lmplot(x="x", y="y", hue='cluster', data = transform, legend=False,`
`fit_reg=False, size = 12, scatter_kws={"s": 250})`

`texts = []`
`for x, y, s in zip(transform.x, transform.y, transform.Name):`
`texts.append(plt.text(x, y, s))`

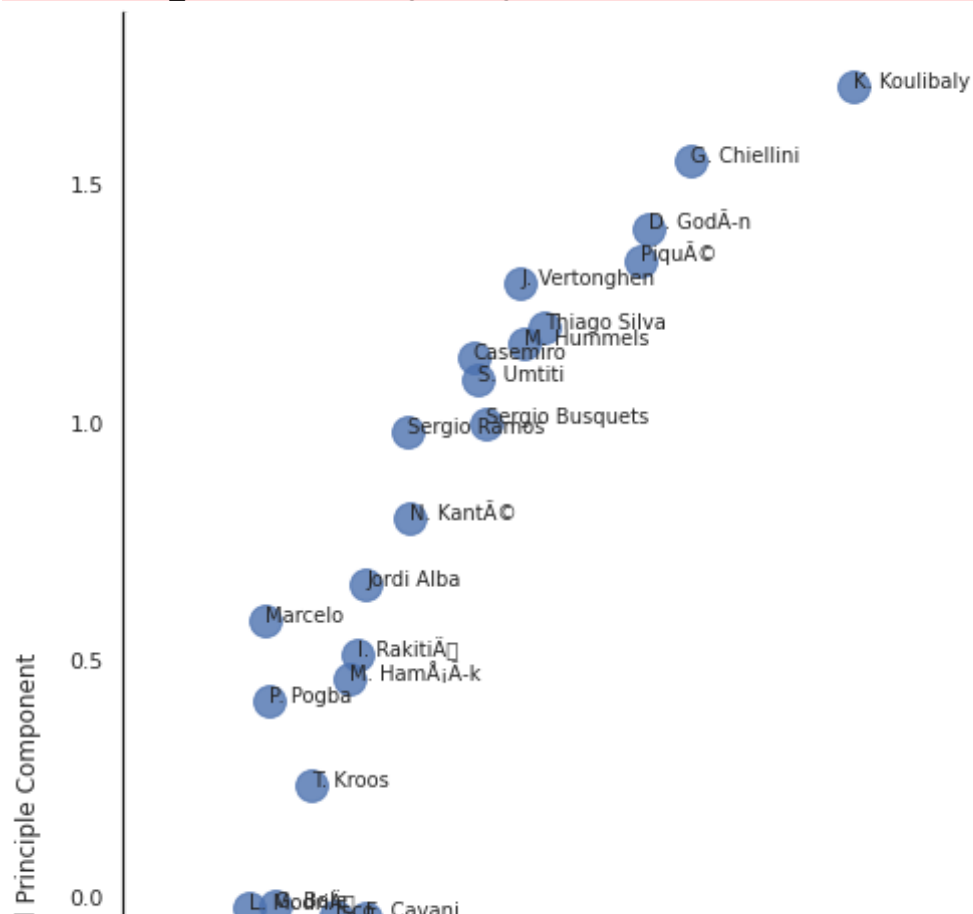
`plt.xlabel("first Principle Component")`
`plt.ylabel("second Principle Component")`

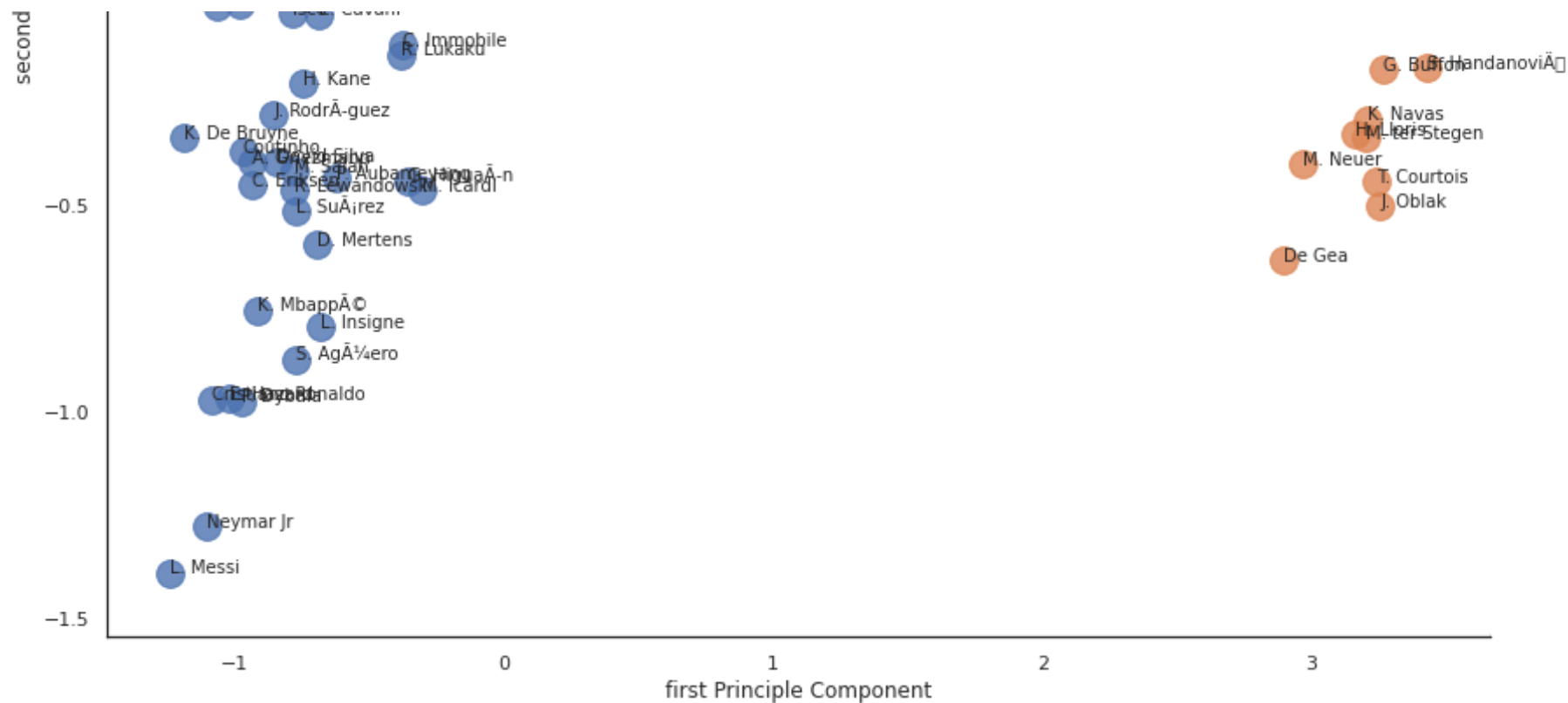
`plt.show()`

```

/usr/local/lib/python3.7/dist-packages/seaborn/regression.py:581: UserWarning: The `size` parameter has been renamed
to `height`; please update your code.
  warnings.warn(msg, UserWarning)
/usr/local/lib/python3.7/dist-packages/matplotlib/backends/backend_agg.py:214: RuntimeWarning: Glyph 135 missing from
current font.
  font.set_text(s, 0.0, flags=flags)
/usr/local/lib/python3.7/dist-packages/matplotlib/backends/backend_agg.py:214: RuntimeWarning: Glyph 141 missing from
current font.
  font.set_text(s, 0.0, flags=flags)
/usr/local/lib/python3.7/dist-packages/matplotlib/backends/backend_agg.py:183: RuntimeWarning: Glyph 135 missing from
current font.
  font.set_text(s, 0, flags=flags)
/usr/local/lib/python3.7/dist-packages/matplotlib/backends/backend_agg.py:183: RuntimeWarning: Glyph 141 missing from
current font.
  font.set_text(s, 0, flags=flags)

```





In []: