Support-Kmeans-Spectral-Clustering

June 6, 2018

In [1]:
```
# Original Source: https://www.kaggle.com/teamaker/clustering-teams-based-on-style-of-pl
# Dataset https://www.kaggle.com/teamaker/clustering-teams-based-on-style-of-play/data
import pandas as pd
import numpy as np
import sqlite3
```

In [2]:
```
con = sqlite3.connect("./database.sqlite")
team=pd.read_sql_query('select * from Team', con)
team_attr=pd.read_sql_query('select * from Team_Attributes', con)
con.close()

df = pd.merge(team, team_attr, how='inner', left_on='team_api_id', right_on='team_api_id
print (df.shape)
```

```
(1458, 29)
```

In [3]:
```
df.sample(n=10)
```

Out[3]:

	id_x	team_api_id	team_fifa_api_id_x	team_long_name	\
111	3459	9825	1.0	Arsenal	
765	20531	8600	55.0	Udinese	
571	15630	8226	10029.0	TSG 1899 Hoffenheim	
432	11074	108893	111989.0	AC Arles-Avignon	
291	8779	8678	1943.0	Bournemouth	
784	21285	8537	1844.0	Livorno	
776	21280	9976	1848.0	Bari	
1247	43041	8388	477.0	CD Numancia	
347	9545	9748	66.0	Olympique Lyonnais	
1062	35288	10238	665.0	Vitória Setúbal	

	team_short_name	id_y	team_fifa_api_id_y	date	\
111	ARS	73	1	2012-02-22 00:00:00	
765	UDI	1302	55	2013-09-20 00:00:00	
571	HOF	596	10029	2015-09-10 00:00:00	
432	ARL	62	111989	2012-02-22 00:00:00	
291	BOU	227	1943	2012-02-22 00:00:00	
784	LIV	753	1844	2014-09-19 00:00:00	

```
776              BAR   127                1848  2012-02-22 00:00:00
1247             NUM   923                 477  2012-02-22 00:00:00
347              LYO   777                  66  2010-02-22 00:00:00
1062             SET  1357                 665  2013-09-20 00:00:00

      buildUpPlaySpeed buildUpPlaySpeedClass            ...             \
111                 25                  Slow            ...
765                 78                  Fast            ...
571                 69                  Fast            ...
432                 23                  Slow            ...
291                 46              Balanced            ...
784                 58              Balanced            ...
776                 50              Balanced            ...
1247                44              Balanced            ...
347                 60              Balanced            ...
1062                56              Balanced            ...

      chanceCreationShooting chanceCreationShootingClass  \
111                       30                      Little
765                       52                      Normal
571                       59                      Normal
432                       43                      Normal
291                       52                      Normal
784                       77                        Lots
776                       50                      Normal
1247                      57                      Normal
347                       70                        Lots
1062                      52                      Normal

      chanceCreationPositioningClass defencePressure defencePressureClass  \
111                       Free Form              57               Medium
765                       Free Form              40               Medium
571                       Organised              63               Medium
432                       Organised              44               Medium
291                       Organised              48               Medium
784                       Organised              29                 Deep
776                       Organised              45               Medium
1247                      Organised              50               Medium
347                       Organised              70                 High
1062                      Organised              37               Medium

      defenceAggression defenceAggressionClass  defenceTeamWidth  \
111                  57                  Press                52
765                  63                  Press                57
571                  61                  Press                38
432                  32                Contain                36
291                  50                  Press                62
784                  57                  Press                32
```

```
776                  45                  Press                  50
1247                 46                  Press                  62
347                  60                  Press                  70
1062                 37                  Press                  58

     defenceTeamWidthClass  defenceDefenderLineClass
111                 Normal                     Cover
765                 Normal                     Cover
571                 Normal                     Cover
432                 Normal                     Cover
291                 Normal                     Cover
784                 Narrow                     Cover
776                 Normal                     Cover
1247                Normal                     Cover
347                   Wide              Offside Trap
1062                Normal                     Cover

[10 rows x 29 columns]
```

```
In [4]: cols_to_keep = ['date', 'team_long_name', u'buildUpPlaySpeed', u'buildUpPlayDribbling',
                u'buildUpPlayPassing', u'chanceCreationPassing', u'chanceCreationCrossing',
              u'chanceCreationShooting', u'defencePressure', u'defenceAggression', u'defenceTea
        df = df[cols_to_keep]
```

```
In [5]: old_df = df.copy(deep=True)
```

```
In [6]: aggs = df.groupby('team_long_name')['date'].max().to_frame()
        df.drop('date', axis=1, inplace=True)
        df.drop_duplicates(subset='team_long_name', keep='last', inplace=True)
        df = df.merge(right=aggs, right_index=True, left_on='team_long_name', how='right')
        df = df.dropna()
        df.set_index('team_long_name', inplace=True)
        df.drop('date', axis=1, inplace=True)
        print (df.shape)
```

```
(260, 9)
```

```
In [37]: index = old_df.team_long_name[old_df.team_long_name.str.contains(pat = 'ES Troyes AC')]
         for i in index:
             print(old_df.iloc[[i]])
```

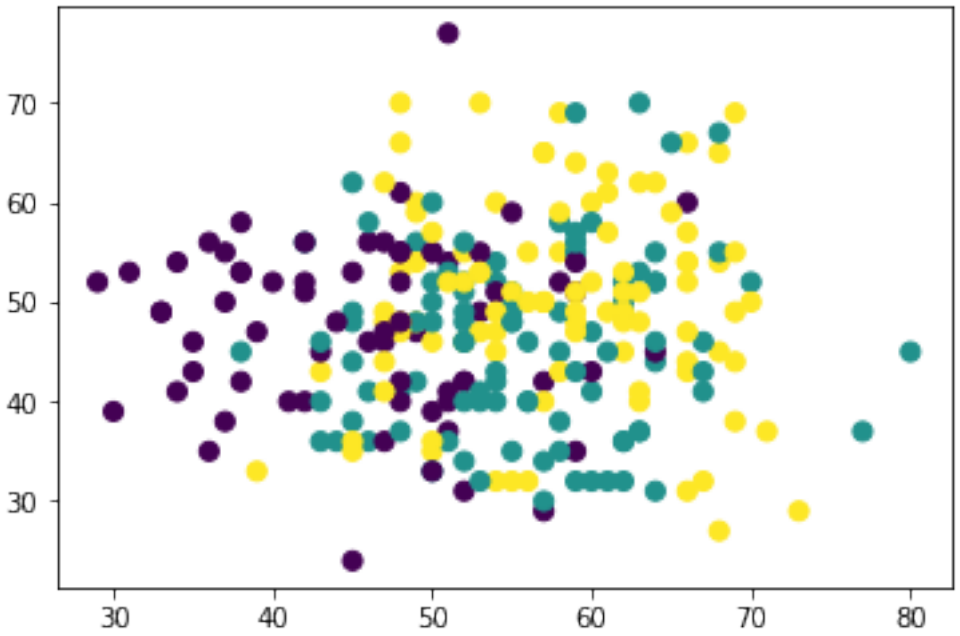
```
                  date team_long_name  buildUpPlaySpeed  \
469  2011-02-22 00:00:00   ES Troyes AC                35

     buildUpPlayDribbling  buildUpPlayPassing  chanceCreationPassing  \
469                   NaN                  50                     50

     chanceCreationCrossing  chanceCreationShooting  defencePressure  \
```

```
469                    50                     35               45

     defenceAggression  defenceTeamWidth
469                45                50
                   date team_long_name  buildUpPlaySpeed  \
470  2012-02-22 00:00:00   ES Troyes AC                49

     buildUpPlayDribbling  buildUpPlayPassing  chanceCreationPassing  \
470                   NaN                  51                     52

     chanceCreationCrossing  chanceCreationShooting  defencePressure  \
470                     49                      49               43

     defenceAggression  defenceTeamWidth
470                42                52
                   date team_long_name  buildUpPlaySpeed  \
471  2013-09-20 00:00:00   ES Troyes AC                49

     buildUpPlayDribbling  buildUpPlayPassing  chanceCreationPassing  \
471                   NaN                  51                     52

     chanceCreationCrossing  chanceCreationShooting  defencePressure  \
471                     49                      49               43

     defenceAggression  defenceTeamWidth
471                42                52
                   date team_long_name  buildUpPlaySpeed  \
472  2014-09-19 00:00:00   ES Troyes AC                49

     buildUpPlayDribbling  buildUpPlayPassing  chanceCreationPassing  \
472                  48.0                  51                     52

     chanceCreationCrossing  chanceCreationShooting  defencePressure  \
472                     49                      46               43

     defenceAggression  defenceTeamWidth
472                42                52
                   date team_long_name  buildUpPlaySpeed  \
473  2015-09-10 00:00:00   ES Troyes AC                49

     buildUpPlayDribbling  buildUpPlayPassing  chanceCreationPassing  \
473                  48.0                  51                     52

     chanceCreationCrossing  chanceCreationShooting  defencePressure  \
473                     49                      46               43

     defenceAggression  defenceTeamWidth
473                42                52
```

0.1 KMEANS

```
In [28]: from sklearn.cluster import KMeans
         from sklearn.preprocessing import MinMaxScaler
         from sklearn.decomposition import PCA
         import pylab as pl
         import matplotlib.pyplot as plt

         kmeans = KMeans(n_clusters=3)
         kmeans.fit(df)
         y_kmeans = kmeans.predict(df)
         plt.scatter(df.iloc[:, 0], df.iloc[:, 1], c=y_kmeans, s=50, cmap='viridis')
```

```
Out[28]: <matplotlib.collections.PathCollection at 0x7ff590a85080>
```

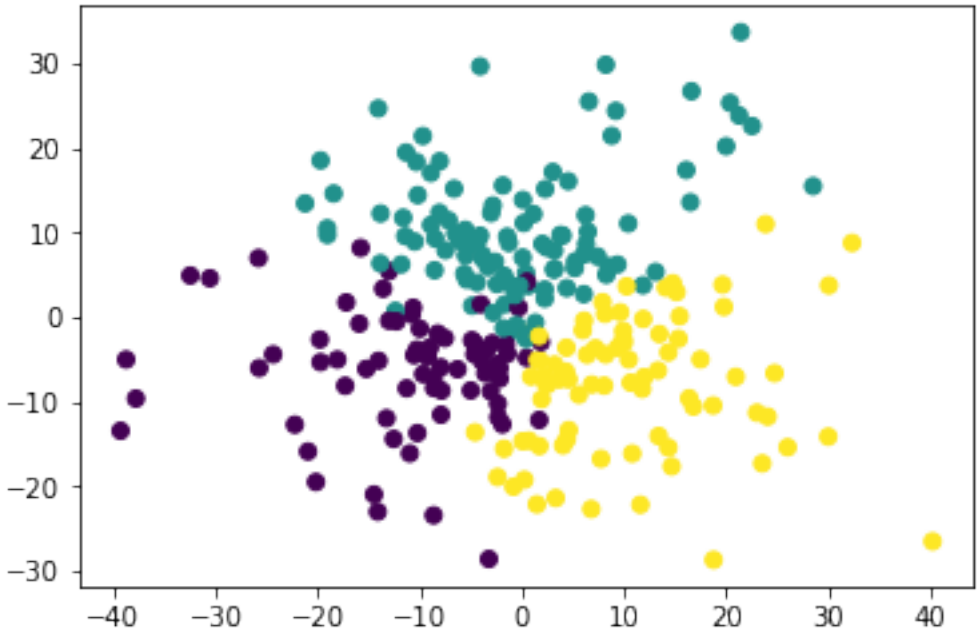


0.2 KMEANS & PCA

```
In [20]: pca = PCA(n_components=3).fit(df)
         print(pca.explained_variance_ratio_)
         print(np.sum([pca.explained_variance_ratio_]))
         pca_2d = pca.transform(df)
         kmeans = KMeans(n_clusters = 3, random_state=10)
         kmeans.fit(df)
         pl.figure('K-means with 3 clusters')
```

```
        pl.scatter(pca_2d[:, 0], pca_2d[:, 1], c=kmeans.labels_)
        pl.show()
```

```
[0.20589458 0.17268013 0.14630594]
0.5248806499908578
```

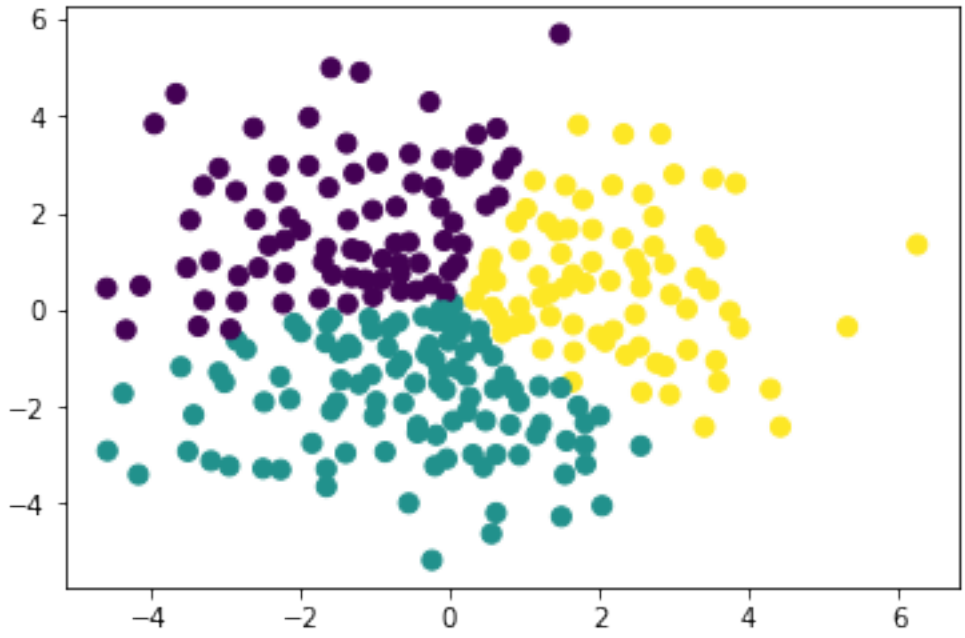


0.3 KMEANS & MDS

```
In [31]: from sklearn.manifold import MDS
         from sklearn.preprocessing import StandardScaler

         x = df.copy()
         x_normalized = StandardScaler().fit(x).transform(x)

         mds = MDS(n_components = 2, n_init = 10)
         mds_2 = MDS(n_components = 3, n_init = 10)
         x_mds = mds.fit_transform(x_normalized)
         x_mds_2 = mds_2.fit_transform(x_normalized)
```

```
In [35]: kmeans = KMeans(n_clusters=3)
         kmeans.fit(x_mds)
         y_kmeans = kmeans.predict(x_mds)
         plt.scatter(x_mds[:, 0], x_mds[:, 1], c=y_kmeans, s=50, cmap='viridis')
```

```
Out[35]: <matplotlib.collections.PathCollection at 0x7ff590d25a20>
```



In [39]: `old_df.iloc[[246, 114, 1244, 132, 473]]`

Out[39]:

	date	team_long_name	buildUpPlaySpeed	buildUpPlayDribbling	buildUpPlayPassing	chanceCreationPassing	chanceCreationCrossing	chanceCreationShooting	defencePressure	defenceAggression	defenceTeamWidth
246	2015-09-10 00:00:00	Swansea City	45	44.0	42	34	36	55	31	47	42
114	2015-09-10 00:00:00	Arsenal	59	51.0	30	28	44	46	51	44	52
1244	2015-09-10 00:00:00	Real Madrid CF	50	57.0	46	61	41	63	52		
132	2015-09-10 00:00:00	Liverpool	66	60.0	45	34	34	46	51		
473	2015-09-10 00:00:00	ES Troyes AC	49	48.0	51	52	49	46	43		

```
1244                60              63
132                 52              61
473                 42              52
```

1 Spectral Clustering

In [77]:
```
from sklearn import cluster

spectral = cluster.SpectralClustering(n_clusters=3, affinity='nearest_neighbors', eigen
sp_result= spectral.fit_predict(x_normalized)
plt.scatter(x_normalized[:, 0], x_normalized[:, 1], c=sp_result, s=50, cmap='viridis')
```

Out[77]:
```
<matplotlib.collections.PathCollection at 0x7ff590ebec88>
```