In [4]:
```python
from sklearn.preprocessing import StandardScaler
from sklearn.cluster import KMeans
import pandas as pd
import numpy as np
from itertools import cycle, islice
import matplotlib.pyplot as plt
from pandas.plotting import parallel_coordinates

%matplotlib inline
```

In [5]:
```python
pwd
```

Out[5]:
```
'C:\\Users\\Hadi\\Desktop\\data science analystics\\week7 machin learning\\Week-7-MachineLear
ning\\weather'
```

In [6]:
```python
data = pd.read_csv('C:\\Users\\Hadi\\Desktop\\data science analystics\\week7 machin learning
\\Week-7-MachineLearning\\weather\\minute_weather.csv')
```

In [4]:
```python
data.shape
```

Out[4]:
```
(1587257, 13)
```

In [5]:
```python
data.head()
```

Out[5]:

	rowID	hpwren_timestamp	air_pressure	air_temp	avg_wind_direction	avg_wind_speed	max_wind_direction	max_wind_spe
0	0	2011-09-10 00:00:49	912.3	64.76	97.0	1.2	106.0	
1	1	2011-09-10 00:01:49	912.3	63.86	161.0	0.8	215.0	
2	2	2011-09-10 00:02:49	912.3	64.22	77.0	0.7	143.0	
3	3	2011-09-10 00:03:49	912.3	64.40	89.0	1.2	112.0	
4	4	2011-09-10 00:04:49	912.3	64.40	185.0	0.4	260.0	

In [6]:
```python
data[data.isnull().any(axis=1)]
```

Out[6]:

	rowID	hpwren_timestamp	air_pressure	air_temp	avg_wind_direction	avg_wind_speed	max_wind_direction	max_
0	0	2011-09-10 00:00:49	912.3	64.76	97.0	1.2	106.0	
34790	34790	2011-10-04 10:25:48	915.7	51.08	NaN	NaN	NaN	
35929	35929	2011-10-05 05:24:48	915.2	49.64	NaN	NaN	NaN	
36320	36320	2011-10-05 11:55:49	914.7	50.00	NaN	NaN	NaN	
36321	36321	2011-10-05 11:56:49	914.7	50.00	NaN	NaN	NaN	
36322	36322	2011-10-05 11:57:49	914.7	50.00	NaN	NaN	NaN	
36323	36323	2011-10-05 11:58:49	914.6	50.00	NaN	NaN	NaN	
36324	36324	2011-10-05 11:59:49	914.7	50.00	NaN	NaN	NaN	
36325	36325	2011-10-05 12:00:49	914.6	50.00	NaN	NaN	NaN	
36326	36326	2011-10-05 12:01:49	914.6	50.00	NaN	NaN	NaN	
36327	36327	2011-10-05 12:02:49	914.5	50.18	NaN	NaN	NaN	
36328	36328	2011-10-05 12:03:49	914.5	50.18	NaN	NaN	NaN	
36329	36329	2011-10-05 12:04:49	914.5	50.18	NaN	NaN	NaN	
36330	36330	2011-10-05 12:05:49	914.4	50.18	NaN	NaN	NaN	
36331	36331	2011-10-05 12:06:49	914.4	50.18	NaN	NaN	NaN	
64745	64745	2011-10-25 05:40:49	918.6	51.08	NaN	NaN	NaN	
79098	79098	2011-11-04 04:53:50	911.0	48.92	NaN	NaN	NaN	
79099	79099	2011-11-04 04:54:50	911.0	48.92	NaN	NaN	NaN	
79100	79100	2011-11-04 04:55:50	911.1	48.92	NaN	NaN	NaN	
79101	79101	2011-11-04 04:56:50	911.1	48.92	NaN	NaN	NaN	
79102	79102	2011-11-04 04:57:50	911.1	48.92	NaN	NaN	NaN	
79103	79103	2011-11-04 04:58:50	911.0	48.92	NaN	NaN	NaN	
79104	79104	2011-11-04 04:59:50	911.0	48.92	NaN	NaN	NaN	
79105	79105	2011-11-04 05:00:50	910.9	48.92	NaN	NaN	NaN	
79106	79106	2011-11-04 05:01:50	911.0	48.92	NaN	NaN	NaN	
79107	79107	2011-11-04 05:02:50	910.9	48.92	NaN	NaN	NaN	
79108	79108	2011-11-04 05:03:50	910.9	48.92	NaN	NaN	NaN	
79250	79250	2011-11-04 07:25:50	910.6	48.02	NaN	NaN	NaN	
79609	79609	2011-11-04 13:24:50	908.6	45.14	NaN	NaN	NaN	
79723	79723	2011-11-04 15:18:50	906.9	46.04	NaN	NaN	NaN	
...	...	...	...	...	...	...	...	
1346164	1346164	2014-03-27 08:46:32	917.1	44.78	NaN	NaN	NaN	
1346165	1346165	2014-03-27 08:47:32	917.1	44.78	NaN	NaN	NaN	
1346166	1346166	2014-03-27 08:48:32	917.1	44.78	NaN	NaN	NaN	
1346167	1346167	2014-03-27 08:49:32	917.1	44.96	NaN	NaN	NaN	
1346168	1346168	2014-03-27 08:50:32	917.1	44.78	NaN	NaN	NaN	
1346169	1346169	2014-03-27 08:51:32	917.2	44.78	NaN	NaN	NaN	
1346170	1346170	2014-03-27 08:52:32	917.1	44.60	NaN	NaN	NaN	
1346171	1346171	2014-03-27 08:53:32	917.1	44.60	NaN	NaN	NaN	
1346172	1346172	2014-03-27 08:54:32	917.1	44.78	NaN	NaN	NaN	
1346173	1346173	2014-03-27 08:55:32	917.1	44.96	NaN	NaN	NaN	
1346174	1346174	2014-03-27 08:56:32	917.1	44.78	NaN	NaN	NaN	
1346175	1346175	2014-03-27 08:57:32	917.0	44.78	NaN	NaN	NaN	
1346176	1346176	2014-03-27 08:58:32	917.0	44.96	NaN	NaN	NaN	
1346177	1346177	2014-03-27 08:59:32	917.1	45.14	NaN	NaN	NaN	
1346178	1346178	2014-03-27 09:00:32	917.1	44.96	NaN	NaN	NaN	
1346179	1346179	2014-03-27 09:01:32	917.2	44.96	NaN	NaN	NaN	
1346180	1346180	2014-03-27 09:02:32	917.1	44.96	NaN	NaN	NaN	
1346181	1346181	2014-03-27 09:03:32	917.2	44.78	NaN	NaN	NaN	
1346182	1346182	2014-03-27 09:04:32	917.2	44.78	NaN	NaN	NaN	
1346183	1346183	2014-03-27 09:05:32	917.2	44.60	NaN	NaN	NaN	
1346184	1346184	2014-03-27 09:06:32	917.3	44.60	NaN	NaN	NaN	
1346185	1346185	2014-03-27 09:07:32	917.3	44.60	NaN	NaN	NaN	
1346186	1346186	2014-03-27 09:08:32	917.3	44.78	NaN	NaN	NaN	
1346187	1346187	2014-03-27 09:09:32	917.4	44.96	NaN	NaN	NaN	
1346188	1346188	2014-03-27 09:10:32	917.3	45.14	NaN	NaN	NaN	
1346189	1346189	2014-03-27 09:11:32	917.4	45.14	NaN	NaN	NaN	
1346190	1346190	2014-03-27 09:12:32	917.4	45.14	NaN	NaN	NaN	
1346191	1346191	2014-03-27 09:13:32	917.4	44.96	NaN	NaN	NaN	
1346192	1346192	2014-03-27 09:14:32	917.5	44.96	NaN	NaN	NaN	
1394844	1394844	2014-04-30 06:21:49	916.7	62.06	NaN	NaN	NaN	

434 rows × 13 columns

In [25]:
```python
before_rows = data.shape[0]
print(before_rows)
```

```
1586823
```

In [26]:
```python
data = data.dropna()
```

In [27]:
```python
after_rows = data.shape[0]
print(after_rows)
```

```
1586823
```

In [28]:
```python
before_rows - after_rows
```

Out[28]:
```
0
```

In [7]:
```python
sampled_df = data[(data['rowID'] % 10) == 0]
sampled_df.shape
```

Out[7]:
```
(158726, 13)
```

In [8]:
```python
sampled_df.head()
```

Out[8]:

	rowID	hpwren_timestamp	air_pressure	air_temp	avg_wind_direction	avg_wind_speed	max_wind_direction	max_wind_sp
0	0	2011-09-10 00:00:49	912.3	64.76	97.0	1.2	106.0	
10	10	2011-09-10 00:10:49	912.3	62.24	144.0	1.2	167.0	
20	20	2011-09-10 00:20:49	912.2	63.32	100.0	2.0	122.0	
30	30	2011-09-10 00:30:49	912.2	62.60	91.0	2.0	103.0	
40	40	2011-09-10 00:40:49	912.2	64.04	81.0	2.6	88.0	

In [9]:
```python
sampled_df.describe()
```

Out[9]:

	rowID	air_pressure	air_temp	avg_wind_direction	avg_wind_speed	max_wind_direction	max_wind_spee
count	1.587260e+05	158726.000000	158726.000000	158680.000000	158680.000000	158680.000000	158680.00000
mean	7.936250e+05	916.830161	61.851589	162.156100	2.775215	163.462144	3.40055
std	4.582039e+05	3.051717	11.833569	95.278201	2.057624	92.452139	2.41880
min	0.000000e+00	905.000000	31.640000	0.000000	0.000000	0.000000	0.10000
25%	3.968125e+05	914.800000	52.700000	62.000000	1.300000	68.000000	1.60000
50%	7.936250e+05	916.700000	62.240000	182.000000	2.200000	187.000000	2.70000
75%	1.190438e+06	918.700000	70.880000	217.000000	3.800000	223.000000	4.60000
max	1.587250e+06	929.500000	99.500000	359.000000	31.900000	359.000000	36.00000

In [10]:
```python
sampled_df.describe().transpose()
```

Out[10]:

	count	mean	std	min	25%	50%	75%	max
rowID	158726.0	793625.000000	458203.937509	0.00	396812.5	793625.00	1190437.50	1587250.00
air_pressure	158726.0	916.830161	3.051717	905.00	914.8	916.70	918.70	929.50
air_temp	158726.0	61.851589	11.833569	31.64	52.7	62.24	70.88	99.50
avg_wind_direction	158680.0	162.156100	95.278201	0.00	62.0	182.00	217.00	359.00
avg_wind_speed	158680.0	2.775215	2.057624	0.00	1.3	2.20	3.80	31.90
max_wind_direction	158680.0	163.462144	92.452139	0.00	68.0	187.00	223.00	359.00
max_wind_speed	158680.0	3.400558	2.418802	0.10	1.6	2.70	4.60	36.00
min_wind_direction	158680.0	166.774017	97.441109	0.00	76.0	180.00	212.00	359.00
min_wind_speed	158680.0	2.134664	1.742113	0.00	0.8	1.60	3.00	31.60
rain_accumulation	158725.0	0.000318	0.011236	0.00	0.0	0.00	0.00	3.12
rain_duration	158725.0	0.409627	8.665523	0.00	0.0	0.00	0.00	2960.00
relative_humidity	158726.0	47.609470	26.214409	0.90	24.7	44.70	68.00	93.00

In [11]:
```python
sampled_df[sampled_df['rain_accumulation'] == 0].shape
```

Out[11]:
```
(157812, 13)
```

In [12]:
```python
sampled_df[sampled_df['rain_duration'] == 0].shape
```

Out[12]:
```
(157237, 13)
```

Drop all the Rows with Empty rain_duration and rain_accumulation

In [13]:
```python
del sampled_df['rain_accumulation']
del sampled_df['rain_duration']
```

In [14]:
```python
rows_before = sampled_df.shape[0]
sampled_df = sampled_df.dropna()
rows_after = sampled_df.shape[0]
```

In [15]:
```python
rows_before
```

Out[15]:
```
158726
```

In [16]:
```python
rows_after
```

Out[16]:
```
158680
```

In [17]:
```python
rows_before -rows_after
```

Out[17]:
```
46
```

In [18]:
```python
sampled_df.columns
```

Out[18]:
```
Index(['rowID', 'hpwren_timestamp', 'air_pressure', 'air_temp',
       'avg_wind_direction', 'avg_wind_speed', 'max_wind_direction',
       'max_wind_speed', 'min_wind_direction', 'min_wind_speed',
       'relative_humidity'],
      dtype='object')
```

selecy features of interest for clusterig

In [19]:
```python
features = ['air_pressure', 'air_temp', 'avg_wind_direction', 'avg_wind_speed', 'max_wind_di
rection',
        'max_wind_speed','relative_humidity']
```

In [20]:
```python
select_df = sampled_df[features]
```

In [21]:
```python
select_df.columns
```

Out[21]:
```
Index(['air_pressure', 'air_temp', 'avg_wind_direction', 'avg_wind_speed',
       'max_wind_direction', 'max_wind_speed', 'relative_humidity'],
      dtype='object')
```

In [22]:
```python
select_df
```

Out[22]:

	air_pressure	air_temp	avg_wind_direction	avg_wind_speed	max_wind_direction	max_wind_speed	relative_humidity
0	912.3	64.76	97.0	1.2	106.0	1.6	60.5
10	912.3	62.24	144.0	1.2	167.0	1.8	38.5
20	912.2	63.32	100.0	2.0	122.0	2.5	58.3
30	912.2	62.60	91.0	2.0	103.0	2.4	57.9
40	912.2	64.04	81.0	2.6	88.0	2.9	57.4
50	912.1	63.68	102.0	1.2	119.0	1.5	51.4
60	912.0	64.04	83.0	0.7	101.0	0.9	51.4
70	911.9	64.22	82.0	2.0	97.0	2.4	62.2
80	911.9	61.70	67.0	3.3	70.0	3.5	71.5
90	911.9	61.34	67.0	3.6	75.0	4.2	72.5
100	911.8	62.96	95.0	2.3	106.0	2.5	63.9
110	911.8	64.22	83.0	2.1	88.0	2.5	59.1
120	911.8	63.86	68.0	2.1	76.0	2.4	63.5
130	911.6	64.40	156.0	0.5	203.0	0.7	50.4
140	911.5	65.30	85.0	2.2	92.0	2.5	58.0
150	911.4	64.58	154.0	1.3	176.0	2.1	50.2
160	911.4	65.48	154.0	0.9	208.0	1.9	46.2
170	911.5	65.66	95.0	1.1	109.0	1.6	45.2
180	911.4	65.66	155.0	1.1	167.0	1.6	42.8
190	911.4	67.10	157.0	1.2	172.0	1.6	36.8
200	911.4	68.00	53.0	0.3	69.0	0.5	33.4
210	911.3	67.64	167.0	1.5	196.0	2.2	34.4
220	911.4	67.82	4.0	0.6	25.0	0.7	34.2
230	911.4	66.74	172.0	1.3	192.0	1.9	37.8
240	911.4	66.56	39.0	0.2	145.0	0.3	41.6
250	911.4	65.66	56.0	1.9	67.0	2.2	51.8
260	911.5	65.66	74.0	0.8	101.0	1.2	41.1
270	911.4	66.92	147.0	0.9	174.0	1.1	36.0
280	911.3	64.76	73.0	1.0	82.0	1.2	43.3
290	911.3	64.94	164.0	1.3	176.0	1.7	43.0
...	...	...	...	...	...	...	...
1586960	914.7	76.46	247.0	0.6	264.0	0.7	43.4
1586970	914.8	76.28	208.0	0.7	216.0	0.9	43.7
1586980	914.8	76.10	209.0	0.7	216.0	0.9	43.9
1586990	914.9	76.28	339.0	0.5	350.0	0.7	43.4
1587000	914.9	75.92	344.0	0.4	352.0	0.6	43.9
1587010	915.0	75.56	323.0	0.3	348.0	0.5	45.5
1587020	915.1	75.56	324.0	1.1	347.0	1.5	46.0
1587030	915.1	75.74	1.0	1.3	13.0	1.7	45.8
1587040	915.2	75.38	355.0	0.9	1.0	1.1	46.1
1587050	915.3	75.38	359.0	1.4	11.0	1.5	45.8
1587060	915.4	75.38	11.0	1.1	21.0	1.3	45.7
1587070	915.5	75.38	13.0	1.4	24.0	1.6	46.6
1587080	915.6	75.20	18.0	1.0	24.0	1.2	46.5
1587090	915.6	75.20	356.0	1.7	1.0	1.9	47.2
1587100	915.7	75.38	13.0	1.5	24.0	1.7	46.7
1587110	915.7	75.02	19.0	1.2	28.0	1.4	46.7
1587120	915.7	74.84	25.0	1.4	35.0	1.6	46.5
1587130	915.8	74.84	23.0	1.3	30.0	1.5	46.9
1587140	915.8	74.84	32.0	1.4	41.0	1.7	45.5
1587150	915.8	75.20	23.0	1.1	31.0	1.4	45.7
1587160	915.8	75.38	16.0	1.2	28.0	1.5	46.3
1587170	915.7	75.38	347.0	1.2	353.0	1.4	48.1
1587180	915.8	75.74	326.0	1.2	337.0	1.6	48.3
1587190	915.9	75.92	289.0	0.7	309.0	0.9	48.1
1587200	915.9	75.74	335.0	0.9	348.0	1.1	47.8
1587210	915.9	75.56	330.0	1.0	341.0	1.3	47.8
1587220	915.9	75.56	330.0	1.1	341.0	1.4	48.0
1587230	915.9	75.56	344.0	1.4	352.0	1.7	48.0
1587240	915.9	75.20	359.0	1.3	9.0	1.6	46.3
1587250	915.9	74.84	6.0	1.5	20.0	1.9	46.1

158680 rows × 7 columns

In [23]:
```python
X = StandardScaler().fit_transform(select_df)
X
```

Out[23]:
```
array([[-1.48456281,  0.24544455, -0.68385323, ..., -0.62153592,
        -0.74440309,  0.49233835],
       [-1.48456281,  0.03247142, -0.19055941, ...,  0.03826701,
        -0.66171726, -0.34710804],
       [-1.51733167,  0.12374562, -0.65236639, ..., -0.44847286,
        -0.37231683,  0.40839371],
       ...,
       [-0.30488381,  1.15818654,  1.90856325, ...,  2.0393087 ,
        -0.70306017,  0.01538018],
       [-0.30488381,  1.12776181,  2.06599745, ..., -1.67073075,
        -0.74440309, -0.04948614],
       [-0.30488381,  1.09733708, -1.63895404, ..., -1.55174989,
        -0.62037434, -0.05711747]])
```

In [24]:
```python
df = pd.DataFrame(X)
```

In [25]:
```python
df
```

Out[25]:

	0	1	2	3	4	5	6
0	-1.484563	0.245445	-0.683853	-0.765553	-0.621536	-0.744403	0.492338
1	-1.484563	0.032471	-0.190559	-0.765553	0.038267	-0.661717	-0.347108
2	-1.517332	0.123746	-0.652366	-0.376754	-0.448473	-0.372317	0.408394
3	-1.517332	0.062896	-0.746827	-0.376754	-0.653985	-0.413660	0.393131
4	-1.517332	0.184595	-0.851783	-0.085154	-0.816232	-0.206945	0.374053
5	-1.550101	0.154170	-0.631375	-0.765553	-0.480922	-0.785746	0.145113
6	-1.582869	0.184595	-0.830792	-1.008552	-0.675618	-1.033804	0.145113
7	-1.615638	0.199807	-0.841287	-0.376754	-0.718884	-0.413660	0.557205
8	-1.615638	-0.013166	-0.998722	0.255045	-1.010928	0.041112	0.912062
9	-1.615638	-0.043590	-0.998722	0.400845	-0.956846	0.330513	0.950218
10	-1.648407	0.093321	-0.704844	-0.230954	-0.621536	-0.372317	0.622071
11	-1.648407	0.199807	-0.830792	-0.328154	-0.816232	-0.372317	0.438919
12	-1.648407	0.169383	-0.988226	-0.328154	-0.946029	-0.413660	0.606808
13	-1.713945	0.215020	-0.064612	-1.105752	0.427659	-1.116489	0.106956
14	-1.746714	0.291082	-0.809801	-0.279554	-0.772966	-0.372317	0.396947
15	-1.779483	0.230232	-0.085603	-0.716953	0.135615	-0.537689	0.099325
16	-1.779483	0.306294	-0.085603	-0.911353	0.481741	-0.620374	-0.053302
17	-1.746714	0.321506	-0.704844	-0.814153	-0.589087	-0.744403	-0.091458
18	-1.779483	0.321506	-0.075108	-0.814153	0.038267	-0.744403	-0.183034
19	-1.779483	0.443205	-0.054116	-0.765553	0.092349	-0.744403	-0.411974
20	-1.779483	0.519267	-1.145660	-1.202952	-1.021744	-1.199175	-0.541707
21	-1.812251	0.488842	0.050840	-0.619753	0.351944	-0.496346	-0.503550
22	-1.779483	0.504055	-1.659945	-1.057152	-1.497668	-1.116489	-0.511182
23	-1.779483	0.412781	0.103318	-0.716953	0.308678	-0.620374	-0.373818
24	-1.779483	0.397568	-1.292599	-1.251552	-0.199695	-1.281861	-0.228822
25	-1.779483	0.321506	-1.114173	-0.425354	-1.043377	-0.496346	0.160375
26	-1.746714	0.321506	-0.925252	-0.959952	-0.675618	-0.909775	-0.247901
27	-1.779483	0.427993	-0.159073	-0.911353	0.113982	-0.951118	-0.442500
28	-1.812251	0.245445	-0.935748	-0.862753	-0.881131	-0.909775	-0.163956
29	-1.812251	0.260657	0.019353	-0.716953	0.135615	-0.703060	-0.175403
...	...	...	...	...	...	...	...
158650	-0.698110	1.234248	0.890489	-1.057152	1.087462	-1.116489	-0.160140
158651	-0.665341	1.219036	0.481160	-1.008552	0.568273	-1.033804	-0.148693
158652	-0.665341	1.203824	0.491655	-1.008552	0.568273	-1.033804	-0.141062
158653	-0.632572	1.219036	1.856085	-1.105752	2.017676	-1.116489	-0.160140
158654	-0.632572	1.188611	1.908563	-1.154352	2.039309	-1.157832	-0.141062
158655	-0.599804	1.158187	1.688155	-1.202952	1.996043	-1.199175	-0.080011
158656	-0.567035	1.158187	1.698651	-0.814153	1.985226	-0.785746	-0.060933
158657	-0.567035	1.173399	-1.691432	-0.716953	-1.627465	-0.703060	-0.068564
158658	-0.534266	1.142974	2.024015	-0.911353	-1.757262	-0.951118	-0.057117
158659	-0.501497	1.142974	2.065997	-0.668353	-1.649098	-0.785746	-0.068564
158660	-0.468728	1.142974	-1.586476	-0.814153	-1.540933	-0.868432	-0.072380
158661	-0.435959	1.142974	-1.565485	-0.668353	-1.508484	-0.744403	-0.038039
158662	-0.403190	1.127762	-1.513007	-0.862753	-1.508484	-0.909775	-0.041855
158663	-0.403190	1.127762	2.034511	-0.522553	-1.757262	-0.620374	-0.015145
158664	-0.370422	1.142974	-1.565485	-0.619753	-1.508484	-0.703060	-0.034223
158665	-0.370422	1.112549	-1.502511	-0.765553	-1.465218	-0.827089	-0.034223
158666	-0.370422	1.097337	-1.439537	-0.668353	-1.389503	-0.744403	-0.041855
158667	-0.337653	1.097337	-1.460529	-0.716953	-1.443585	-0.785746	-0.026592
158668	-0.337653	1.097337	-1.366068	-0.668353	-1.324605	-0.703060	-0.080011
158669	-0.337653	1.127762	-1.460529	-0.814153	-1.432769	-0.827089	-0.072380
158670	-0.337653	1.142974	-1.533998	-0.765553	-1.465218	-0.785746	-0.049486
158671	-0.370422	1.142974	1.940050	-0.765553	2.050125	-0.827089	0.019196
158672	-0.337653	1.173399	1.719642	-0.765553	1.877062	-0.744403	0.026827
158673	-0.304884	1.188611	1.331305	-1.008552	1.574202	-1.033804	0.019196
158674	-0.304884	1.173399	1.814103	-0.911353	1.996043	-0.951118	0.007749
158675	-0.304884	1.158187	1.761625	-0.862753	1.920328	-0.868432	0.007749
158676	-0.304884	1.158187	1.761625	-0.814153	1.920328	-0.827089	0.015380
158677	-0.304884	1.158187	1.908563	-0.668353	2.039309	-0.703060	0.015380
158678	-0.304884	1.127762	2.065997	-0.716953	-1.670731	-0.744403	-0.049486
158679	-0.304884	1.097337	-1.638954	-0.619753	-1.551750	-0.620374	-0.057117

158680 rows × 7 columns

In [26]:
```python
kmeans = KMeans(n_clusters=12)
model = kmeans.fit(X)
print("model\n", model)
```

```
model
 KMeans(algorithm='auto', copy_x=True, init='k-means++', max_iter=300,
    n_clusters=12, n_init=10, n_jobs=None, precompute_distances='auto',
    random_state=None, tol=0.0001, verbose=0)
```

In [27]:
```python
centers = model.cluster_centers_
centers
```

Out[27]:
```
array([[-0.70635479,  0.53403112,  0.17517676, -0.58447962,  0.34479891,
        -0.59791259, -0.10774813],
       [ 0.13262266,  0.84100449,  1.41242073, -0.6382135 ,  1.67658107,
        -0.58900661, -0.71382497],
       [-0.16370459,  0.8634041 , -1.31126777, -0.58955336, -1.16710698,
        -0.60484116, -0.64125855],
       [ 1.19007801, -0.25468312, -1.15493166,  2.12261318, -1.0534084 ,
         2.23936256, -1.13456932],
       [ 0.24471888, -0.99610161,  0.65809375, -0.547383  ,  0.8478918 ,
        -0.53015888,  1.16144381],
       [-0.83765454, -1.20120114,  0.37557175,  0.36904555,  0.47403811,
         0.35702563,  1.36243665],
       [-0.21057891,  0.63296915,  0.40858054,  0.73350062,  0.51675564,
         0.67164628, -0.1516111 ],
       [-1.18245544, -0.87048459,  0.44692904,  1.98306377,  0.53842715,
         1.94398173,  0.9088656 ],
       [ 0.71667149,  0.44894203,  0.28550213, -0.53631985,  0.47279917,
        -0.54255693, -0.76669252],
       [ 0.06113923, -0.78891905, -1.19730343, -0.5707786 , -1.0433185 ,
        -0.58536331,  0.87915471],
       [ 1.36768818, -0.08151489, -1.20649644, -0.0510817 , -1.07519624,
        -0.03084637, -0.97766689],
       [ 0.23405125,  0.32061871,  1.8879544 , -0.65181511, -1.55175807,
        -0.57669397, -0.28385048]])
```

In [28]:
```python
dfc = pd.DataFrame(centers)
dfc
```

Out[28]:

	0	1	2	3	4	5	6
0	-0.706355	0.534031	0.175177	-0.584480	0.344799	-0.597913	-0.107748
1	0.132623	0.841004	1.412421	-0.638214	1.676581	-0.589007	-0.713825
2	-0.163705	0.863404	-1.311268	-0.589553	-1.167107	-0.604841	-0.641259
3	1.190078	-0.254683	-1.154932	2.122613	-1.053408	2.239363	-1.134569
4	0.244719	-0.996102	0.658094	-0.547383	0.847892	-0.530159	1.161444
5	-0.837655	-1.201201	0.375572	0.369046	0.474038	0.357026	1.362437
6	-0.210579	0.632969	0.408581	0.733501	0.516756	0.671646	-0.151611
7	-1.182455	-0.870485	0.446929	1.983064	0.538427	1.943982	0.908866
8	0.716671	0.448942	0.285502	-0.536320	0.472799	-0.542557	-0.766693
9	0.061139	-0.788919	-1.197303	-0.570779	-1.043318	-0.585363	0.879155
10	1.367688	-0.081515	-1.206496	-0.051082	-1.075196	-0.030846	-0.977667
11	0.234051	0.320619	1.887954	-0.651815	-1.551758	-0.576694	-0.283850

plot

In [36]:
```python
# Function that creates a DataFrame with a column for Cluster Number

def pd_centers(featuresUsed, centers):
        colNames = list(featuresUsed)
        colNames.append('prediction')

        # Zip with a column called 'prediction' (index)
        Z = [np.append(A, index) for index, A in enumerate(centers)]

        # Convert to pandas data frame for plotting
        P = pd.DataFrame(Z, columns=colNames)
        P['prediction'] = P['prediction'].astype(int)
        return P
```

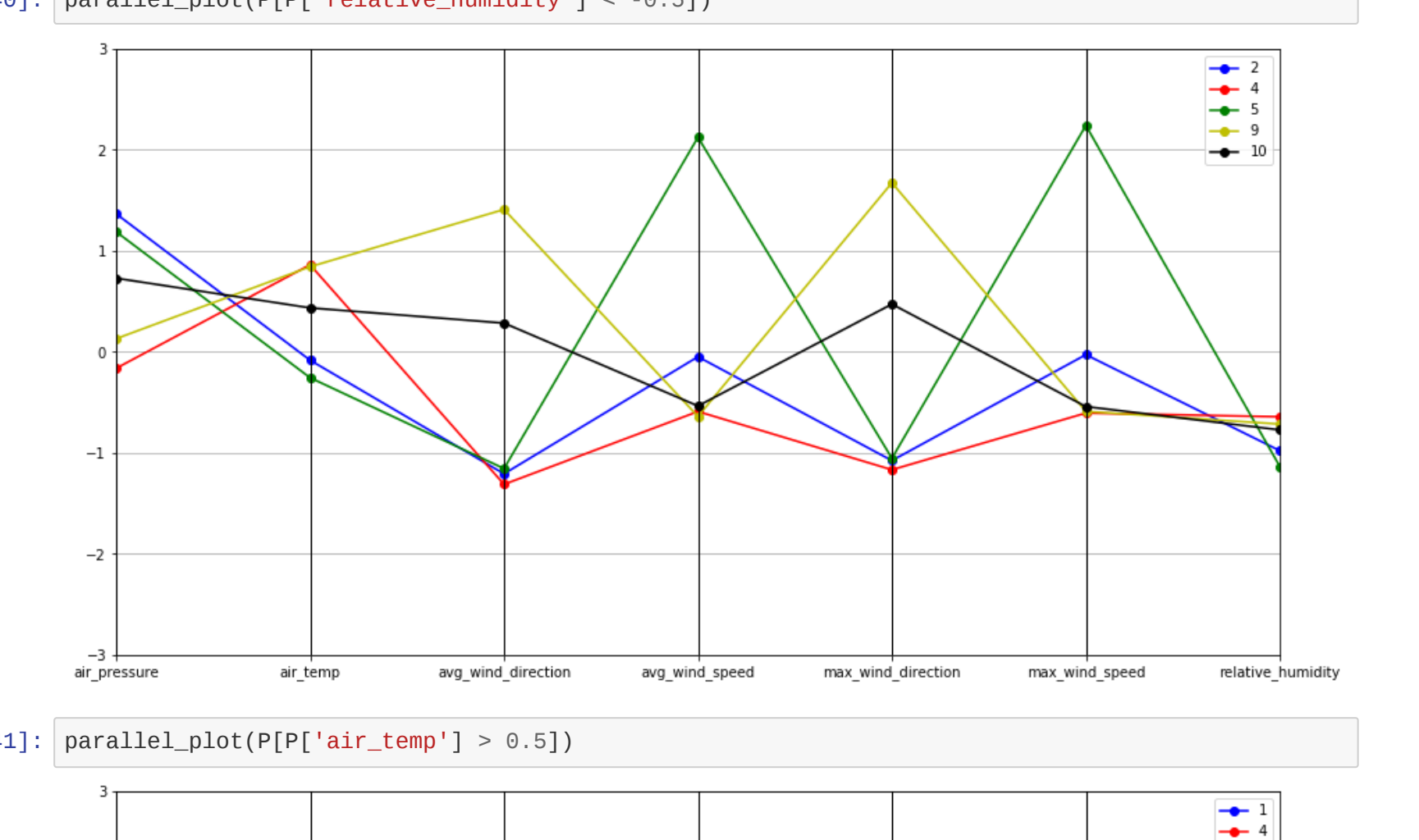
In [38]:
```python
# Function that creates Parallel Plots

def parallel_plot(data):
        my_colors = list(islice(cycle(['b', 'r', 'g', 'y', 'k']), None, len(data)))
        plt.figure(figsize=(15,8)).gca().axes.set_ylim([-3,+3])
        parallel_coordinates(data, 'prediction', color = my_colors, marker='o')
```

In [39]:
```python
P = pd_centers(features, centers)
P
```

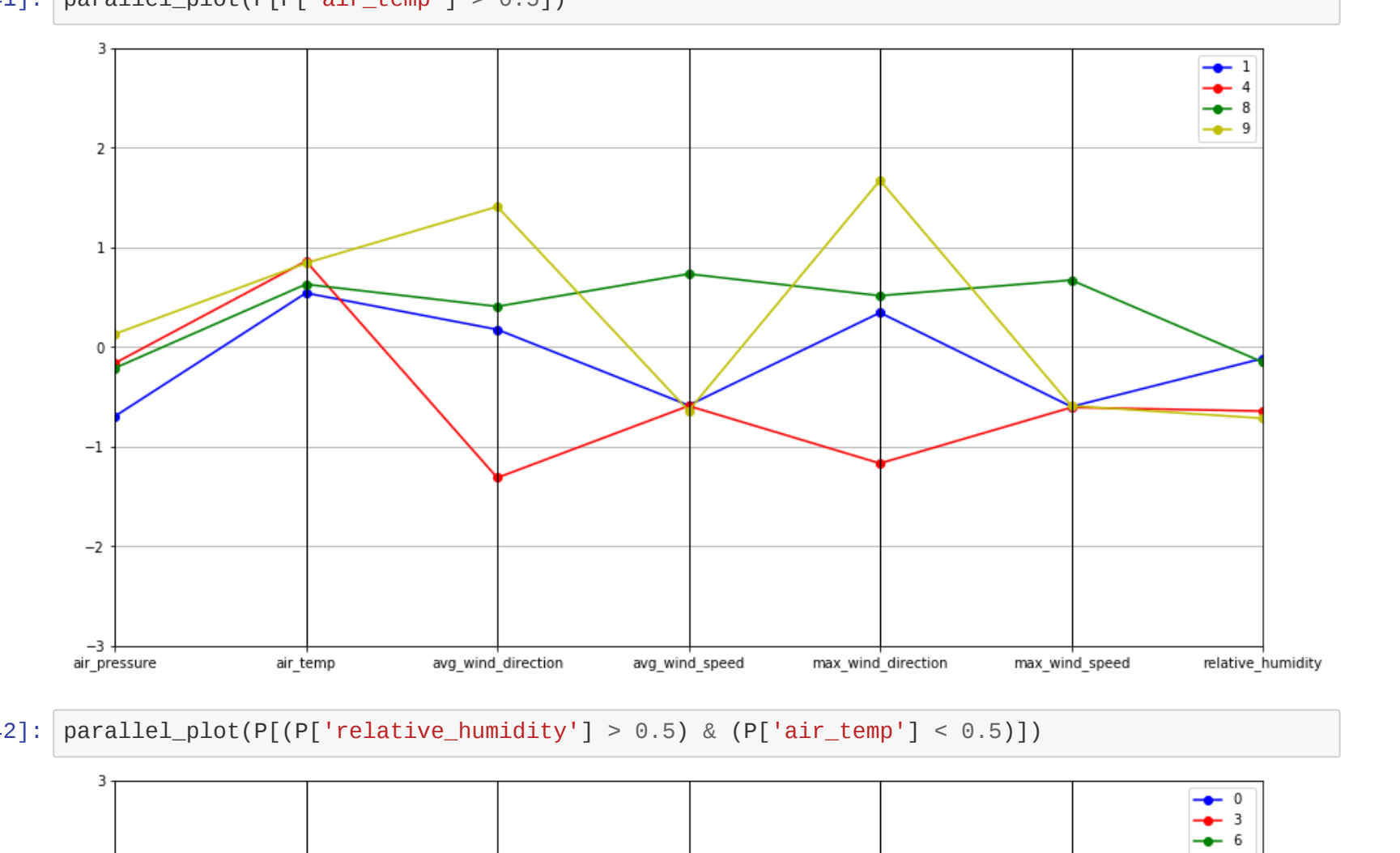
Out[39]:

	air_pressure	air_temp	avg_wind_direction	avg_wind_speed	max_wind_direction	max_wind_speed	relative_humidity	pre
0	-1.179342	-0.876563	0.446720	1.975899	0.538688	1.937175	0.914810	
1	-0.696010	0.543476	0.176912	-0.584115	0.346270	-0.597462	-0.114050	
2	1.368077	-0.081869	-1.206908	-0.047642	-1.075861	-0.027245	-0.977890	
3	0.060708	-0.787933	-1.197036	-0.570755	-1.043077	-0.585268	0.877645	
4	-0.162411	0.863347	-1.311167	-0.589835	-1.166840	-0.605155	-0.642042	
5	1.189960	-0.255285	-1.155018	2.125443	-1.053437	2.242580	-1.134262	
6	0.253092	-0.994446	0.660009	-0.547528	0.851442	-0.530134	1.157872	
7	0.233931	0.319096	1.887941	-0.651982	-1.551644	-0.576814	-0.282516	
8	-0.211274	0.631414	0.408542	0.734884	0.516678	0.672862	-0.149990	
9	0.130834	0.843663	1.410950	-0.638475	1.675046	-0.589204	-0.714227	
10	0.730135	0.436177	0.285486	-0.534751	0.473246	-0.541026	-0.772000	
11	-0.840131	-1.198434	0.375190	0.354031	0.473641	0.341912	1.362737	

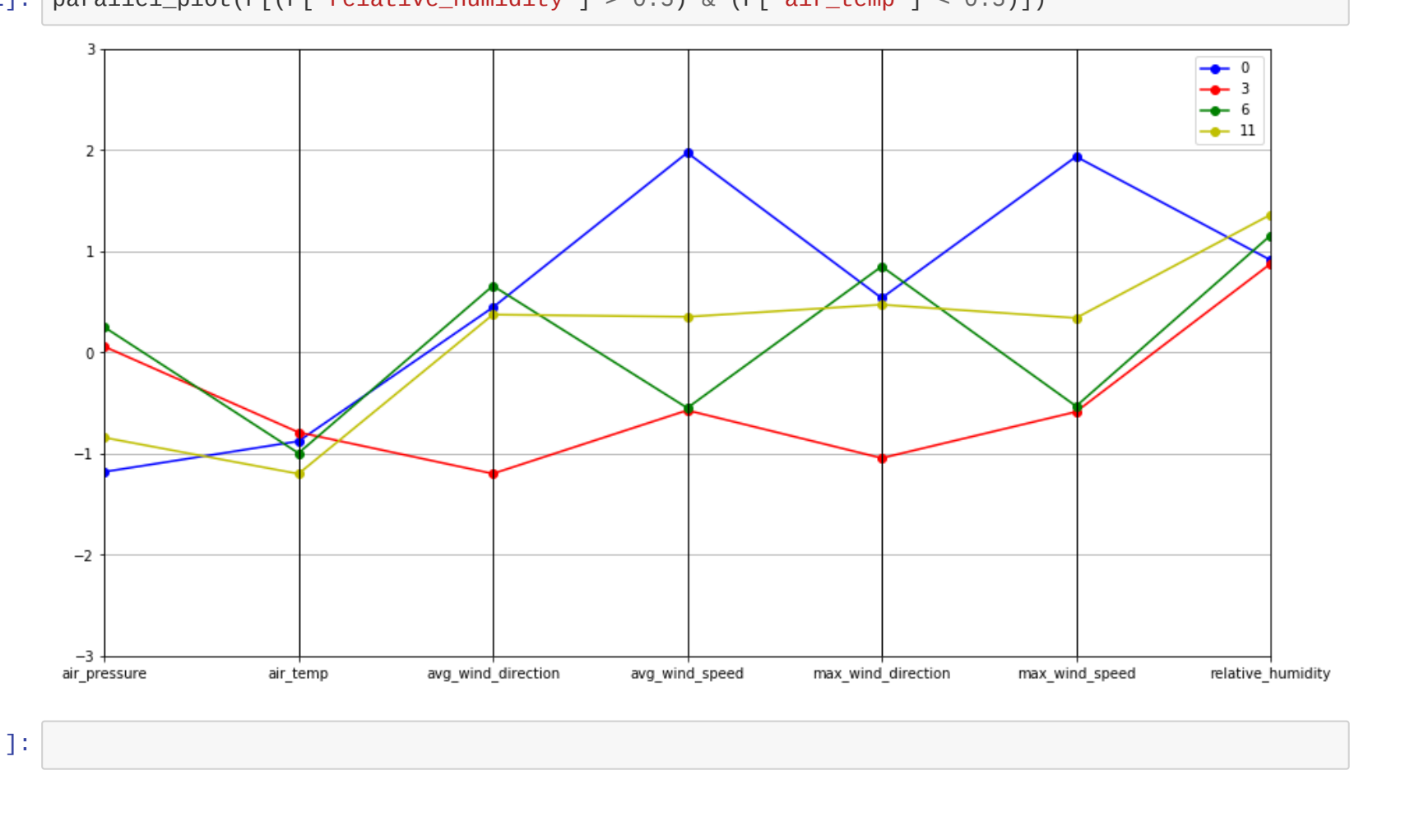
In [40]:
```python
parallel_plot(P[P['relative_humidity'] < -0.5])
```



In [41]:
```python
parallel_plot(P[P['air_temp'] > 0.5])
```



In [42]:
```python
parallel_plot(P[(P['relative_humidity'] > 0.5) & (P['air_temp'] < 0.5)])
```



In [ ]: