Data Mining Assignment 1
Vrishti Jain: 661983527 CSCI 6390

Part 1

```
enter name of fileairfoil_self_noise.dat
enter epsilon0.001
<class 'pandas.core.frame.DataFrame'>
mean
              0
0  2886.380572
1     6.782302
2     0.136548
3    50.860745
4     0.011140
5   124.835943
---------------
vairance
              0
0  9.938717e+06
1  3.502424e+01
2  8.749868e-03
3  2.425116e+02
4  1.729287e-04
5  4.759146e+01
total variance  9939042.519937238

    inner covarinace
              0            1         2            3         4            5
 0  9.932105e+06 -5085.672506 -1.078782  6557.771759 -9.533231 -8491.741152
 1 -5.085673e+03    35.000938 -0.279302     5.411780  0.058594    -6.369183
 2 -1.078782e+00    -0.279302  0.008744     0.005512 -0.000271    -0.152295
 3  6.557772e+03     5.411780  0.005512   242.350262 -0.000813    13.431013
 4 -9.533231e+00     0.058594 -0.000271    -0.000813  0.000173    -0.028346
 5 -8.491741e+03    -6.369183 -0.152295    13.431013 -0.028346    47.559799
 _____
    outer covraince
              0            1         2            3         4            5
 0  9.932105e+06 -5085.672506 -1.078782  6557.771759 -9.533231 -8491.741152
 1 -5.085673e+03    35.000938 -0.279302     5.411780  0.058594    -6.369183
 2 -1.078782e+00    -0.279302  0.008744     0.005512 -0.000271    -0.152295
 3  6.557772e+03     5.411780  0.005512   242.350262 -0.000813    13.431013
 4 -9.533231e+00     0.058594 -0.000271    -0.000813  0.000173    -0.028346
 5 -8.491741e+03    -6.369183 -0.152295    13.431013 -0.028346    47.559799
 _____
    correlation matrix
           0         1         2         3         4         5
 0  1.000000 -0.272765 -0.003661  0.133664 -0.230107 -0.390711
 1 -0.272765  1.000000 -0.504868  0.058760  0.753394 -0.156108
 2 -0.003661 -0.504868  1.000000  0.003787 -0.220842 -0.236162
 3  0.133664  0.058760  0.003787  1.000000 -0.003974  0.125103
 4 -0.230107  0.753394 -0.220842 -0.003974  1.000000 -0.312670
 5 -0.390711 -0.156108 -0.236162  0.125103 -0.312670  1.000000
```

Part 2

```
Eigen Values
                0              1
0   9.925075e+06   9.925075e+06

Eigen Vector
                0              1
0   1.000000e+00  -1.000000e+00
1  -5.089987e-04   5.089987e-04
2  -1.507691e-07   1.507691e-07
3   6.870447e-04  -6.870447e-04
4  -9.575641e-07   9.575641e-07
5  -8.500994e-04   8.500994e-04
```
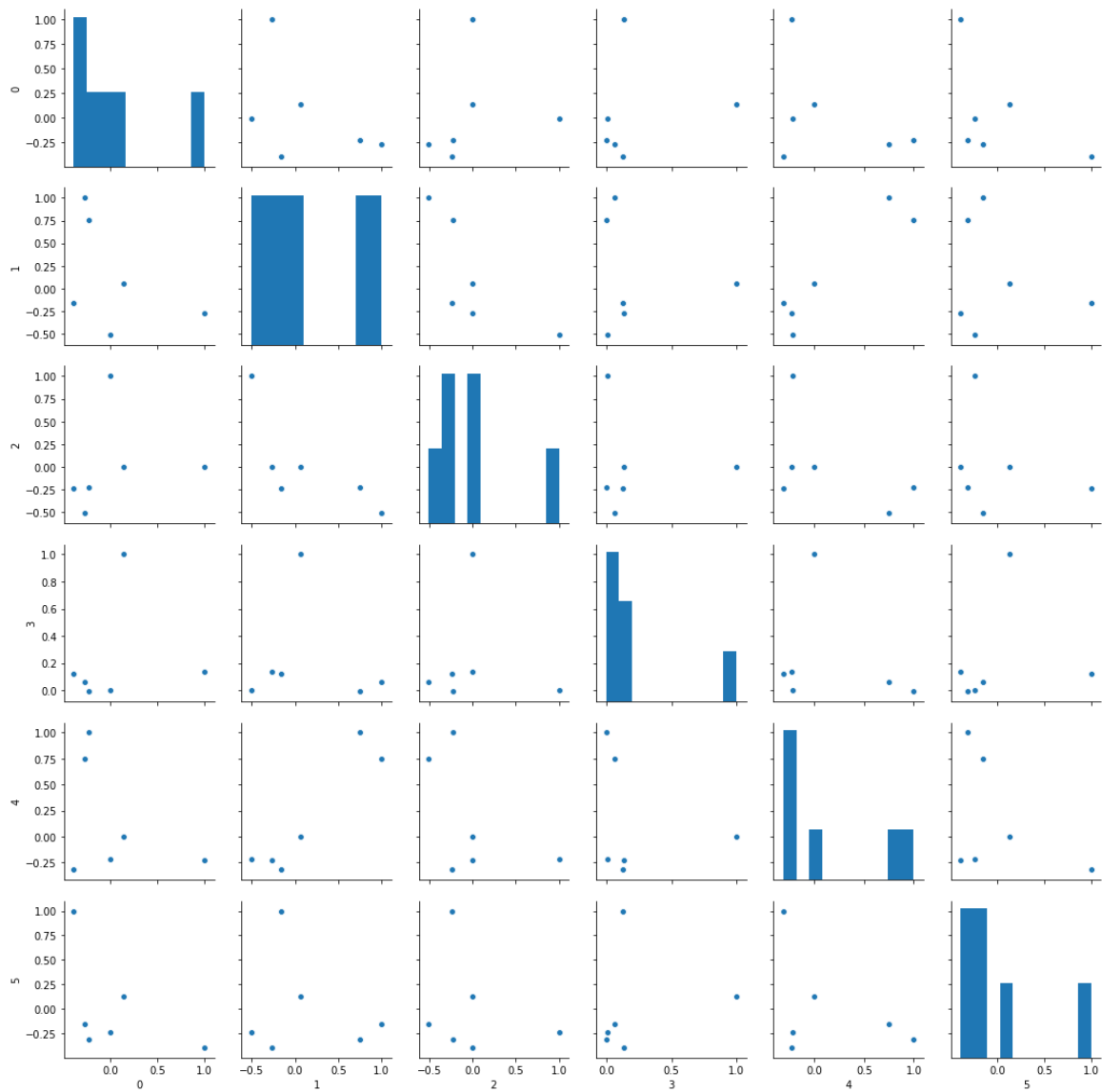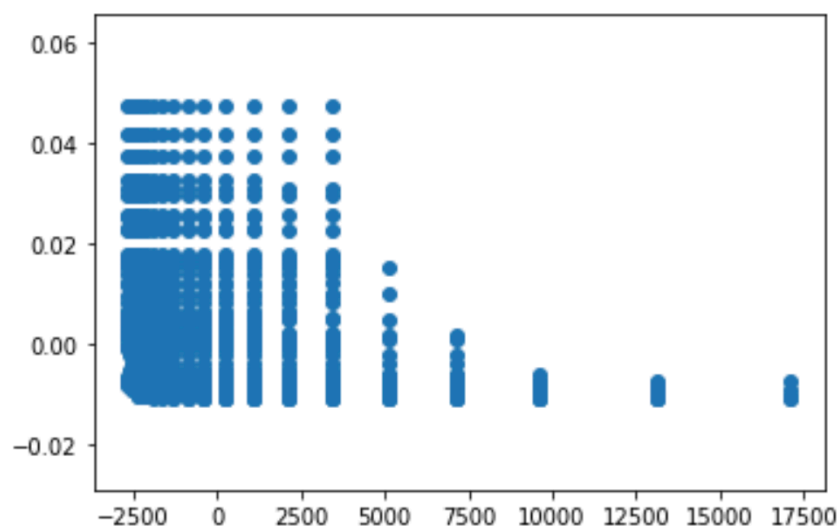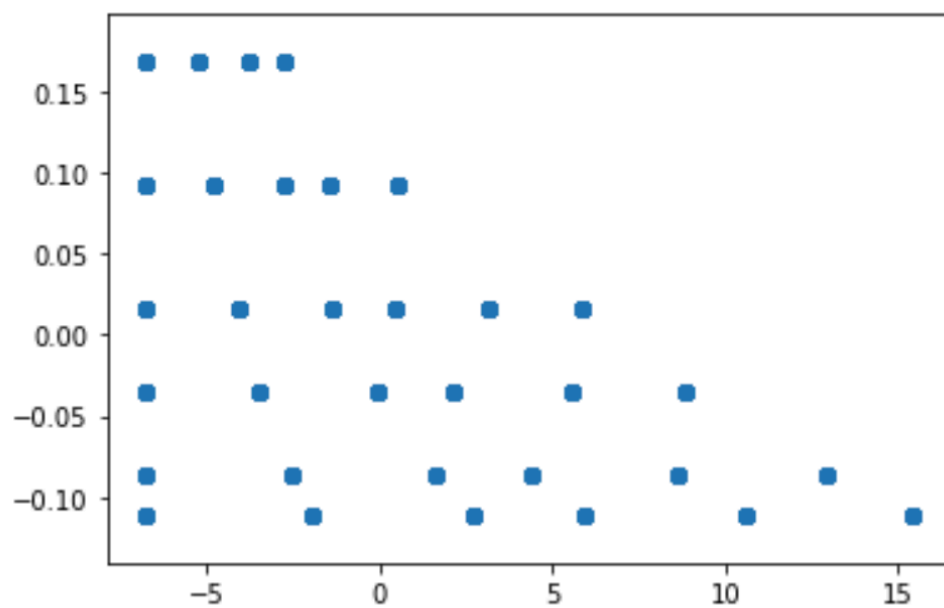
```
#### scatter plots
#between  most correlated attributes 0 and 4
plt.scatter(center_df[:,0],center_df[:,4])
plt.show()
```



```
# most anti correlated 1 and 2
plt.scatter(center_df[:,1],center_df[:,2])
plt.show()
```

Projection of original data on new dimensions