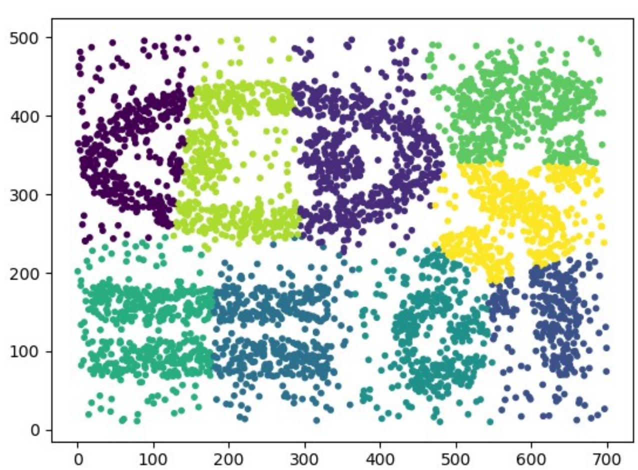
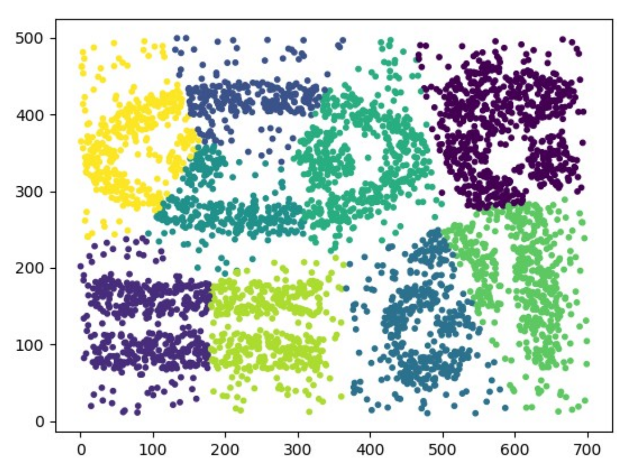
```python
#1
A="A"
B="B"
C="C"
D="D"
E="E"

#case 1
a=(0,1,1,1,1,2,2,3)
b=(A,A,A,E,E,D,D,C)
c=(8,8,8,4,4,5,5,6)
#case 2
a=(1,1,1,0,0,2,2,2)
b=(A,A,A,E,E,D,D,C)

def entropy(a, b):
    a=list(a)
    b=list(b)
    #remove outliers
    ind=[i for i in range(len(a)) if a[i]>0]
    a=[a[i] for i in ind]
    b=[b[i] for i in ind]

    uniquea, countsa=np.unique(a, return_counts=True)
    entsum=0

    for ele in countsa:
        templist=[]
        ratio=ele/sum(countsa)

        for i in range(0, ele):
            templist.append(b[i])

        b=b[ele:]
        uniquelist, listcount=np.unique(templist, return_counts=True)
        tempsum=0

        for i in listcount:
            r=i/sum(listcount)
            l=(-1*r)*math.log(r, 2)
            tempsum=tempsum+l

        entsum=entsum+(ratio*tempsum)
    return entsum

ent=entropy(a, b)
print("Entropy:", ent)

#2
def ordinal_variation(a,b):
    a=list(a)
    b=list(b)
    #remove outliers
    ind=[i for i in range(len(a)) if a[i]>0]
    a=[a[i] for i in ind]
    b=[b[i] for i in ind]
    c=[]
```

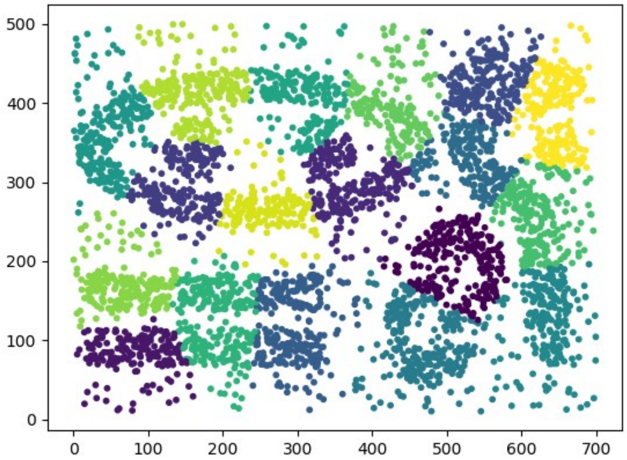
```python
    for i in b:
        if i == "A":
            c.append(4)
        elif i=="B":
            c.append(3)
        elif i=="C":
            c.append(2)
        elif i=="D":
            c.append(1)
        elif i=="E":
            c.append(0)

    uniquea, acount=np.unique(a, return_counts=True)
    ordvar=0

    for ele in acount:
        templist=[]
        ratio=ele/sum(acount)
        #print(ele)
        #ordinal variance
        d=c[0:ele]
        c=c[ele:len(c)]
        #print(d)
        summ=0

        for ele in d:
            for ele2 in d:
                if ele!=ele2:
                    try:
                        summ+=abs(ele-ele2)/(math.pow(abs(ele), 2)-abs(ele))
                    except Exception as e:
                        summ+=0
        #print(summ, ratio)
        ordvar+=(summ*ratio)
    return ordvar;

ov=ordinal_variation(a, b)
print("Ordinal Variation:",ov)

#3
def variance(a,b):
    a=list(a)
    b=list(b)
    #remove outliers
    ind=[i for i in range(len(a)) if a[i]>0]
    a=[a[i] for i in ind]
    b=[b[i] for i in ind]

    uniquea, acount=np.unique(a, return_counts=True)
    varsum=0

    for ele in acount:
        templist=[]
        ratio=ele/sum(acount)

        for i in range(0, ele):
            templist.append(b[i])
```

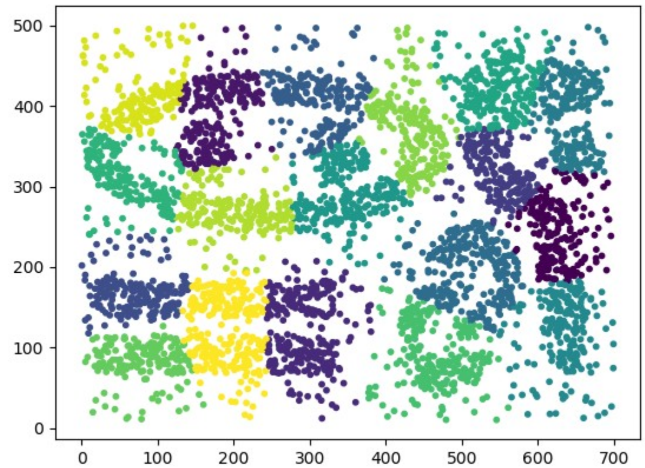
```
        b=b[ele:]
        #print(ratio, templist)
        v=ratio*np.var(templist)
        varsum=varsum+v
    return varsum

v=variance(a, c)
print("Variance:",v)


OUTPUT/RESULTS:
Case 1:
Entropy: 0.5714285714285714
Ordinal Variation: 0.7619047619047619
Variance: 2.2857142857142856

Case 2:
Entropy: 0.4591479170272448
Ordinal Variation: 0.5
Variance: 0.11111111111111112

#4
inp=[[1, 5], [2, 6], [3, 7], [4, 8]]
df=pd.DataFrame(inp)

def mdist(d):
    d=d.apply(zscore)
    dismat=pd.DataFrame(distance_matrix(d.values, d.values), index=d.index,
columns=d.index)
    return dismat

dismat=mdist(df)
print(dismat)

RESULT:

          0         1         2         3
0  0.000000  1.264911  2.529822  3.794733
1  1.264911  0.000000  1.264911  2.529822
2  2.529822  1.264911  0.000000  1.264911
3  3.794733  2.529822  1.264911  0.000000

#5
k=9
```





k=18





Interpretation: For k=9, the “=” shape as well as the backwards “s” has similar groupings. On the other hand, the “eye” shape and “cane” figures are grouped very differently. In the left, the “eye” is split into 3 even vertical cuts, while the right plot splits the “eye” into 4 groupings. Similarly, the “cane” on the left is split into more groups with 3, while the right “cane” is split into 2 even groups. For k=18, we see a similar trend to k=9. Again, the “=” and backwards “s” are grouped similarly in both plots. In this case, the “cane” is also split in about the same ways. The main difference lies in the “eye” shape, although both have 7 groupings, the groupings are shifted differently from one another.

Entropy: 0.0030007501875468873

#6
k=5

```
          fixed.acidity  volatile.acidity    ...       alcohol   quality
clusters                                     ...
0              6.411654          0.286688    ...     12.017070  6.717627
1              6.395822          0.277169    ...     10.111049  5.602667
2              7.007138          0.279714    ...      9.452198  5.605710
3              7.695896          0.263243    ...     10.831765  5.622623
4              6.688679          0.315142    ...      9.569811  5.500000
[5 rows x 12 columns]
```

k=10

```
          fixed.acidity  volatile.acidity    ...       alcohol   quality
clusters                                     ...
0              7.370990          0.314334    ...      9.322639  5.399317
1              6.946713          0.238711    ...      9.426384  5.949827
2              6.233659          0.350176    ...     12.494260  6.819961
3              6.828406          0.228869    ...     11.003428  6.406170
4              8.013465          0.257515    ...     10.486733  5.247525
5              6.249658          0.221772    ...     10.544551  5.946648
6              6.779841          0.286432    ...      9.593523  5.417772
7              6.691304          0.314239    ...      9.501087  5.434783
8              6.621345          0.444401    ...      9.828450  4.956140
9              7.042390          0.250007    ...     11.637530  6.422475
[10 rows x 12 columns]
```

mdist:

	0	1	2	...	4895	4896	4897
0	0.000000	5.386771	4.581024	...	5.309328	7.523470	6.958395
1	5.386771	0.000000	2.851156	...	2.783272	3.780493	3.438118
2	4.581024	2.851156	0.000000	...	3.548426	4.852179	4.110080
3	3.529110	3.254773	2.798696	...	3.447800	5.188067	4.632857
4	3.529110	3.254773	2.798696	...	3.447800	5.188067	4.632857
5	4.581024	2.851156	0.000000	...	3.548426	4.852179	4.110080
6	4.315951	2.255447	3.252105	...	2.178513	4.256442	4.117414
7	0.000000	5.386771	4.581024	...	5.309328	7.523470	6.958395
8	5.386771	0.000000	2.851156	...	2.783272	3.780493	3.438118
9	5.446206	2.891996	1.734287	...	3.461331	4.317455	3.493259
10	6.800478	4.247781	3.524325	...	3.901483	4.572140	3.992608
11	5.074709	3.237749	1.842164	...	3.488841	5.400241	4.549244
12	6.347620	3.280867	2.823433	...	3.442035	4.562729	3.936843
13	6.982846	3.930130	4.147811	...	5.005649	3.528553	3.545558
14	3.706473	6.275588	5.047728	...	6.747375	8.373299	7.926881
15	6.088038	2.632663	3.148065	...	3.086087	2.946831	2.414987
16	6.451387	3.532192	4.466405	...	3.307499	4.450890	4.632693
17	8.321620	5.301813	5.594298	...	6.211977	4.181077	4.750805
18	5.815819	2.826512	3.213381	...	3.468169	3.872381	3.594309
19	4.217589	2.475334	3.125003	...	2.442975	4.569987	4.426829
20	8.321620	5.301813	5.594298	...	6.211977	4.181077	4.750805
21	5.829290	2.376208	2.957044	...	2.792544	2.548836	1.817952
22	5.876728	2.360099	2.629538	...	3.920212	3.697574	3.327604
23	6.599452	4.882436	5.302206	...	4.975107	6.757520	6.937404
24	5.575400	1.266497	2.720749	...	3.736444	3.968608	3.589462
25	4.124698	4.145457	4.154074	...	4.576337	5.657181	5.453158
26	5.741360	1.997290	2.685511	...	3.545029	3.822824	3.400828
27	4.267854	2.364115	2.095607	...	3.926026	4.493601	4.224016
28	6.070590	2.715712	2.673640	...	3.611072	3.552420	2.991893
29	6.478552	3.962025	3.931279	...	3.864056	4.141176	4.206860
...	...	...	...	...	...	...	...
4868	5.488345	2.626849	3.644224	...	3.328377	3.756795	3.796098
4869	4.387556	2.580544	2.710353	...	3.220813	4.578833	4.017788
4870	5.290483	3.005482	3.548518	...	2.888374	2.510708	2.329066
4871	7.409169	3.694302	5.024422	...	4.549789	2.187046	2.489736
4872	3.554085	3.918386	4.235090	...	4.461252	5.649982	5.570561
4873	6.575335	2.953031	4.073275	...	3.530987	1.976984	2.180820
4874	6.264981	1.918066	3.806733	...	3.937740	3.425088	3.244241
4875	5.575164	2.474894	2.536447	...	1.785303	4.030380	3.115680
4876	6.335316	2.741449	3.558605	...	4.061361	2.985828	3.312245
4877	7.232193	4.206575	5.470264	...	4.192384	5.127401	5.365461
4878	7.249168	4.034586	5.178690	...	4.078085	4.733072	5.004566
4879	3.802700	3.764139	3.545548	...	3.838808	5.459512	5.069966
4880	3.802700	3.764139	3.545548	...	3.838808	5.459512	5.069966
4881	3.993089	3.454258	4.370043	...	3.095377	4.789592	4.446439
4882	6.082911	3.220167	4.540108	...	3.012337	3.189224	3.442974
4883	7.127709	4.664420	5.727096	...	4.584435	3.614166	4.289715
4884	3.765060	3.780715	3.647929	...	3.779737	5.443339	5.081868
4885	3.804846	3.764261	3.543814	...	3.839214	5.452938	5.065060
4886	7.868364	6.192888	6.659640	...	6.281935	6.642326	7.003658
4887	8.222449	5.248402	5.950681	...	5.078154	4.678148	5.274869
4888	4.879492	2.512140	2.935470	...	1.856939	4.808135	4.054067
4889	4.064558	3.508721	4.470139	...	3.164928	4.805694	4.479484
4890	6.382364	3.215129	3.934665	...	2.702450	2.405707	2.395319
4891	6.010470	2.424103	3.686499	...	2.548663	2.652463	2.213885
4892	5.505347	1.611090	2.703875	...	2.677663	3.629012	3.012278

```
4893  6.229407  2.300883  3.270559    ...     2.675435  2.420299  2.118802
4894  3.581634  3.171569  3.088140    ...     3.253202  4.994154  4.562829
4895  5.309328  2.783272  3.548426    ...     0.000000  4.328675  3.708491
4896  7.523470  3.780493  4.852179    ...     4.328675  0.000000  1.673130
4897  6.958395  3.438118  4.110080    ...     3.708491  1.673130  0.000000

[4898 rows x 4898 columns]

k=5
Entropy: 1.5938510683466758
Ordinal Agreement: 240364.6037106778
Variance: 1.2487170881627512

k=10
Entropy: 1.8194148566746986
Ordinal Agreement: 68433.19572929788
Variance: 7.739214642380813

#7
Complex9_RN32:
                   X         Y     CLASS
clusters
0         358.204497  258.8454  5.633408


White Wine:
          fixed.acidity  volatile.acidity    ...       alcohol   quality
clusters                                     ...
-1             8.071429          0.477857    ...     10.471429  5.000000
 0             6.853046          0.277955    ...     10.514328  5.879166

[2 rows x 12 columns]

#9
                          OLS Regression Results
==============================================================================
Dep. Variable:                quality   R-squared:                       0.282
Model:                            OLS   Adj. R-squared:                  0.280
Method:                 Least Squares   F-statistic:                     174.3
Date:                Sun, 21 Oct 2018   Prob (F-statistic):               0.00
Time:                        21:12:38   Log-Likelihood:                -5543.7
No. Observations:                4898   AIC:                         1.111e+04
Df Residuals:                    4886   BIC:                         1.119e+04
Df Model:                          11
Covariance Type:            nonrobust
==========================================================================================
===
                                coef    std err          t      P>|t|      [0.025
0.975]
------------------------------------------------------------------------------------------
---
Intercept                   150.1928     18.804      7.987      0.000     113.328
187.057
Q("fixed.acidity")            0.0655      0.021      3.139      0.002       0.025
0.106
Q("volatile.acidity")        -1.8632      0.114    -16.373      0.000      -2.086      -
1.640
```

```
Q("citric.acid")             0.0221      0.096      0.231      0.818      -0.166
0.210
Q("residual.sugar")          0.0815      0.008     10.825      0.000       0.067
0.096
chlorides                   -0.2473      0.547     -0.452      0.651      -1.319
0.824
Q("free.sulfur.dioxide")     0.0037      0.001      4.422      0.000       0.002
0.005
Q("total.sulfur.dioxide")   -0.0003      0.000     -0.756      0.450      -0.001
0.000
density                   -150.2842     19.075     -7.879      0.000    -187.679    -
112.890
pH                           0.6863      0.105      6.513      0.000       0.480
0.893
sulphates                    0.6315      0.100      6.291      0.000       0.435
0.828
alcohol                      0.1935      0.024      7.988      0.000       0.146
0.241
==============================================================================
Omnibus:                      114.161   Durbin-Watson:                   1.621
Prob(Omnibus):                  0.000   Jarque-Bera (JB):              251.637
Skew:                           0.073   Prob(JB):                     2.28e-55
Kurtosis:                       4.101   Cond. No.                     3.74e+05
==============================================================================

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly
specified.
[2] The condition number is large, 3.74e+05. This might indicate that there are
strong multicollinearity or other numerical problems.
```

The attributes with the highest p values are total sulfur dioxide, chlorides, and citric acid. Thus, they provide less evidence against the null hypothesis. We can observe that the other 8 attributes are important when deciding the quality attribute, because of their smaller p values. The R^2 value (0.282) is currently far from 1, thus this may not be a good model for the data.

```
optimization finished, #iter = 265331
obj = -0.435505, rho = -0.347698
nSV = 11, nBSV = 0
[LibSVM]SVR(C=1.0, cache_size=200, coef0=0.0, degree=3, epsilon=0.2, gamma='auto',
  kernel='linear', max_iter=-1, shrinking=True, tol=0.001, verbose=True)
```