

Ödevi yaparken **Python** kullandım ve Python kullanırken **pandas** kütüphanelerinden yardım aldım. Verilen **xml** uzantılı dosyayı **csv** uzantılı dosyaya çevirip başlıklarını sırasıyla Sepal.Length,Sepal.Width,Petal.Length,Petal.Width şeklinde isimlendirdim. Aşağıda gösterilen kod yardımıyla .csv uzantılı dosyayı Python’a import ettim.

```
#irisdata.csv dosyasını aşağıdaki komut ile importluyorum
iris_data = pd.read_csv("irisdata.csv")
```

İmport edilen veriler aşağıdaki gibidir:

Index	sepal.Length	sepal.Width	petal.Length	petal.Width	Cicek Ismi
0	5.1	3.5	1.4	0.2	Iris-setosa
1	4.9	3	1.4	0.2	Iris-setosa
2	4.7	3.2	1.3	0.2	Iris-setosa
3	4.6	3.1	1.5	0.2	Iris-setosa
4	5	3.6	1.4	0.2	Iris-setosa
5	5.4	2.9	1.7	0.4	Iris-setosa
6	4.6	3.4	1.4	0.3	Iris-setosa
7	5	3.4	1.5	0.2	Iris-setosa
8	4.4	2.9	1.4	0.2	Iris-setosa
9	4.9	3.1	1.5	0.1	Iris-setosa
10	5.4	3.7	1.5	0.2	Iris-setosa
11	4.8	3.4	1.6	0.2	Iris-setosa
12	4.8	3	1.4	0.1	Iris-setosa
13	4.3	3	1.1	0.1	Iris-setosa
14	5.8	4	1.2	0.2	Iris-setosa
15	5.7	4.4	1.5	0.4	Iris-setosa
16	5.4	3.9	1.3	0.4	Iris-setosa
17	5.1	3.5	1.4	0.3	Iris-setosa
18	5.7	3.8	1.7	0.3	Iris-setosa
19	5.1	3.8	1.5	0.3	Iris-setosa
20	5.4	3.4	1.7	0.2	Iris-setosa
21	5.1	3.7	1.5	0.4	Iris-setosa
22	4.6	3.6	1	0.2	Iris-setosa
23	5.1	3.3	1.7	0.5	Iris-setosa
24	4.8	2.6	1.9	0.2	Iris-setosa
25	5	3	1.6	0.2	Iris-setosa
26	5	3.4	1.6	0.4	Iris-setosa
27	5.2	3.5	1.5	0.2	Iris-setosa
28	5.2	3.4	1.4	0.2	Iris-setosa
29	4.7	3.2	1.6	0.2	Iris-setosa
30	4.8	3.1	1.6	0.2	Iris-setosa
31	5.4	3.4	1.5	0.4	Iris-setosa
32	5.2	4.1	1.5	0.1	Iris-setosa
33	5.5	4.2	1.4	0.2	Iris-setosa
34	4.9	3.1	1.5	0.1	Iris-setosa
35	5	3.2	1.2	0.2	Iris-setosa
36	5.5	2.5	1.3	0.2	Iris-setosa
37	4.9	3.1	1.5	0.1	Iris-setosa
38	4.4	3	1.3	0.2	Iris-setosa
39	5.1	3.4	1.5	0.2	Iris-setosa
40	5	3.5	1.3	0.3	Iris-setosa
41	4.5	2.3	1.3	0.3	Iris-setosa
42	4.4	3.2	1.3	0.2	Iris-setosa
43	5	2.5	1.6	0.6	Iris-setosa
44	5.1	3.8	1.9	0.4	Iris-setosa
45	4.8	3	1.4	0.3	Iris-setosa
46	5.1	2.8	1.6	0.2	Iris-setosa
47	4.6	3.2	1.4	0.2	Iris-setosa
48	5.3	2.7	1.5	0.2	Iris-versicolae
49	5	3.3	1.4	0.2	Iris-versicolae
50	7	5.2	4.7	1.4	Iris-versicolae
51	6.4	3.2	4.5	1.5	Iris-versicolae
52	6.9	3.3	4.9	1.5	Iris-versicolae
53	5.5	2.3	4	1.3	Iris-versicolae
54	6.5	2.8	4.6	1.5	Iris-versicolae
55	5.7	2.8	4.5	1.3	Iris-versicolae
56	6.3	3.3	4.7	1.6	Iris-versicolae
57	4.9	2.4	3.4	1	Iris-versicolae
58	6.6	2.9	4.4	1.3	Iris-versicolae
59	5.2	3.7	3.9	1.4	Iris-versicolae
60	5	2	3.5	1	Iris-versicolae
61	5.9	3	4.2	1.5	Iris-versicolae
62	6	2.2	4	1	Iris-versicolae
63	6.1	2.9	4.7	1.4	Iris-versicolae
64	5.6	2.9	3.6	1.3	Iris-versicolae
65	6.7	3.1	4.4	1.4	Iris-versicolae
66	5.6	3	4.5	1.5	Iris-versicolae
67	5.8	2.7	4.1	1	Iris-versicolae
68	5.2	2.2	4.5	1.5	Iris-versicolae
69	5.6	2.5	3.9	1.1	Iris-versicolae
70	5.9	3.2	4.8	1.8	Iris-versicolae
71	6.1	2.8	4	1.3	Iris-versicolae
72	6.3	2.5	4.9	1.5	Iris-versicolae
73	6.1	2.8	4.7	1.2	Iris-versicolae
74	6.4	2.9	4.3	1.3	Iris-versicolae
75	6.6	3	4.4	1.4	Iris-versicolae
76	6.8	2.8	4.8	1.4	Iris-versicolae
77	6.7	3	5	1.7	Iris-versicolae
78	6	2.9	4.5	1.5	Iris-versicolae
79	5.7	2.6	3.5	1	Iris-versicolae
80	5.5	2.4	3.8	1.1	Iris-versicolae
81	5.5	2.4	3.7	1	Iris-versicolae
82	5.8	2.7	3.9	1.2	Iris-versicolae
83	6	2.7	5.1	1.6	Iris-versicolae
84	5.4	3	4.5	1.5	Iris-versicolae
85	6	3.4	4.5	1.6	Iris-versicolae
86	6.7	3.1	4.7	1.5	Iris-versicolae
87	6.3	2.3	4.4	1.3	Iris-versicolae
88	5.6	3	4.1	1.3	Iris-versicolae
89	5.5	2.5	4	1.3	Iris-versicolae
90	5.5	2.6	4.4	1.2	Iris-versicolae
91	6.1	3	4.6	1.4	Iris-versicolae
92	5.8	2.6	4	1.2	Iris-versicolae
93	5	2.3	3.3	1	Iris-versicolae
94	5.6	2.7	4.2	1.3	Iris-versicolae
95	5.7	3	4.2	1.2	Iris-versicolae
96	5.7	2.9	4.2	1.3	Iris-versicolae
97	6.2	2.9	4.3	1.5	Iris-versicolae
98	5.1	2.5	3	1.1	Iris-versicolae
99	5.7	2.8	4.3	1.3	Iris-versicolae
100	6.3	3.3	6	2.5	Iris-versicolae
101	5.8	2.7	5.1	1.9	Iris-versicolae
102	7.1	3	5.9	2.1	Iris-versicolae
103	6.3	2.9	5.6	1.8	Iris-versicolae
104	6.5	3	5.8	2.2	Iris-versicolae
105	7.6	3	6.6	2.1	Iris-versicolae
106	4.9	2.5	4.5	1.7	Iris-versicolae
107	7.3	2.9	6.3	1.8	Iris-versicolae
108	6.7	2.5	5.8	1.8	Iris-versicolae
109	7.2	3.6	6.1	2.5	Iris-versicolae
110	6.5	3.2	5.1	2	Iris-versicolae
111	6.4	2.7	5.3	1.9	Iris-versicolae
112	6.8	3	5.5	2.1	Iris-versicolae
113	5.7	2.5	5	2	Iris-versicolae
114	5.8	2.8	5.1	2.4	Iris-versicolae
115	6.4	3.2	5.3	2.3	Iris-versicolae
116	6.5	3	5.5	1.8	Iris-versicolae
117	7.7	3.8	6.7	2.2	Iris-versicolae
118	7.7	2.6	6.9	2.3	Iris-versicolae
119	6	2.2	5	1.5	Iris-versicolae
120	6.9	3.2	5.7	2.3	Iris-versicolae
121	5.6	2.8	4.9	2	Iris-versicolae
122	7.7	2.8	6.7	2	Iris-versicolae
123	6.3	2.7	4.9	1.8	Iris-versicolae
124	6.7	3.3	5.7	2.1	Iris-versicolae
125	7.2	3.2	6	1.8	Iris-versicolae
126	6.2	2.8	4.8	1.8	Iris-versicolae
127	6.1	3	4.9	1.8	Iris-versicolae
128	6.4	2.8	5.6	2.1	Iris-versicolae
129	7.2	3	5.8	1.6	Iris-versicolae
130	7.4	2.8	6.1	1.9	Iris-versicolae
131	7.9	3.8	6.4	2	Iris-versicolae
132	6.4	2.8	5.6	2.2	Iris-versicolae
133	6.3	2.8	5.1	1.5	Iris-versicolae
134	6.1	2.6	5.6	1.4	Iris-versicolae
135	7.7	3	6.1	2.3	Iris-versicolae
136	6.3	3.4	5.6	2.4	Iris-versicolae
137	6.4	3.1	5.5	1.8	Iris-versicolae
138	6	3	4.8	1.8	Iris-versicolae
139	6.9	3.1	5.4	2.1	Iris-versicolae
140	6.7	3.1	5.6	2.4	Iris-versicolae
141	4.9	3.1	5.1	2.3	Iris-versicolae
142	5.8	2.7	5.1	1.9	Iris-versicolae
143	6.8	3.2	5.9	2.3	Iris-versicolae
144	6.7	3.3	5.7	2.5	Iris-versicolae
145	6.7	3	5.2	2.3	Iris-versicolae
146	6.3	2.5	5	1.9	Iris-versicolae
147	6.5	3	5.2	2	Iris-versicolae
148	6.2	3.4	5.4	2.3	Iris-versicolae
149	5.9	3	5.1	1.8	Iris-versicolae

150 satırdan oluşan veriyi aşağıdaki gibi yazılan kod yardımıyla 3 parçaya böldüm.

```
#irisdata.csv dosyasında verilen çiçek türlerini 3 ayrı sınıfa ayırıyoruz.
setosa = iris_data.iloc[0:50]
versicolae = iris_data.iloc[50:100]
virginica = iris_data.iloc[100:150]
```

a-) `pd.DataFrame(df.mean())` komutuyla her serinin 4 sütünü ayrı ayrı ortalamasını hesaplatarak ortalama vektörünü elde ettim.

```
#pandas kütüphanesinde bulunan .mean() komutuyla b
avg_setosa = pd.DataFrame(setosa.mean())
avg_versicolor = pd.DataFrame(versicolor.mean())
avg_virginica = pd.DataFrame(virginica.mean())
```

Sonuçlar aşağıdaki şekilde gösterilmiştir.

avg_setosa - DataFrame

Index	0
Sepal.Length	5.006
Sepal.Width	3.418
Petal.Length	1.464
Petal.Width	0.244

avg_versicolor - DataFrame

Index	0
Sepal.Length	5.936
Sepal.Width	2.77
Petal.Length	4.26
Petal.Width	1.326

avg_virginica - DataFrame

Index	0
Sepal.Length	6.588
Sepal.Width	2.974
Petal.Length	5.552
Petal.Width	2.026

b-) Ardından `pd.cov()` komutunu kullanarak kovaryans matris hesabı yaptım. Aşağıda ilgili kodlar ve kovaryans matrisler belirtilmiştir.

```
#numpy kütüphanesinde bulunan np.cov ve np.array komu
setosa_cov_matrix= setosa.iloc[:, :4].cov()
versicolor_cov_matrix= versicolor.iloc[:, :4].cov()
virginica_cov_matrix= virginica.iloc[:, :4].cov()
```

setosa_cov_matrix - DataFrame

Index	Sepal.Length	Sepal.Width	Petal.Length	Petal.Width
Sepal.Length	0.124249	0.100298	0.0161388	0.0105469
Sepal.Width	0.100298	0.14518	0.0116816	0.0114367
Petal.Length	0.0161388	0.0116816	0.0301061	0.00569796
Petal.Width	0.0105469	0.0114367	0.00569796	0.0114939

versicolor_cov_matrix - DataFrame

Index	Sepal.Length	Sepal.Width	Petal.Length	Petal.Width
Sepal.Length	0.266433	0.0851837	0.182898	0.0557796
Sepal.Width	0.0851837	0.0984694	0.0826531	0.0412041
Petal.Length	0.182898	0.0826531	0.220816	0.073102
Petal.Width	0.0557796	0.0412041	0.073102	0.0391061

virginica_cov_matrix - DataFrame

Index	Sepal.Length	Sepal.Width	Petal.Length	Petal.Width
Sepal.Length	0.404343	0.0937633	0.30329	0.0490939
Sepal.Width	0.0937633	0.104004	0.0713796	0.0476286
Petal.Length	0.30329	0.0713796	0.304588	0.0488245
Petal.Width	0.0490939	0.0476286	0.0488245	0.0754327

Son aşama olarak kovaryans matrislerin doğru hesaplanıp hesaplanmadığının sağlamasını yapmak için varyans hesabı yaptım. Varyans hesabı sonucunda yukarıda belirtilen matrislerdeki sayılarla eşleşmesini bekliyorum. Aşağıda açıklamaya göre ilgili kod ve çıktıları belirtilmiştir.

```
#Sağlama yapmak için varyans hesabı yapalım. Cov matrisind
#Setosa'nın varyansını hesaplama
setosa_matrix = pd.DataFrame(setosa.iloc[:, :4])
setosa_var= setosa_matrix.var() #axis=0 değeriyle her sütü

#Versicolor'ın varyansını hesaplama
versicolor_matrix = pd.DataFrame(versicolor.iloc[:, :4])
versicolor_var= versicolor_matrix.var()

#Virginica'nın varyansını hesaplama
virginica_matrix = pd.DataFrame(virginica.iloc[:, :4])
virginica_var= virginica_matrix.var()
```

setosa_var - Series		versicolor_var - Series		virginica_var - Series	
Index	0	Index	0	Index	0
Sepal.Length	0.124249	Sepal.Length	0.266433	Sepal.Length	0.404343
Sepal.Width	0.14518	Sepal.Width	0.0984694	Sepal.Width	0.104004
Petal.Length	0.0301061	Petal.Length	0.220816	Petal.Length	0.304588
Petal.Width	0.0114939	Petal.Width	0.0391061	Petal.Width	0.0754327

Yukarıdaki varyans hesapları ile kovaryans matrislerinin köşegenlerinin karşılaştırılması yapıldığında birebir eşit çıktığını görüyoruz. Bu da işlemi doğru yaptığımızı gösteriyor.

Kod:

```

1 import pandas as pd
2
3 #Irisdata.csv dosyasını aşağıdaki komut ile importluyoruz
4 iris_data = pd.read_csv("irisdata.csv")
5
6 #Irisdata.csv dosyasında verilen çiçek türlerini 3 ayrı sınıfa ayırıyoruz.
7 setosa = iris_data.iloc[0:50]
8 versicolor = iris_data.iloc[50:100]
9 virginica = iris_data.iloc[100:150]
10
11 #pandas kütüphanesinde bulunan .mean() komutuyla birlikte 3 çiçeğin ayrı ayrı hesaplayarak ortalama değer
12 avg_setosa = pd.DataFrame(setosa.mean())
13 avg_versicolor = pd.DataFrame(versicolor.mean())
14 avg_virginica = pd.DataFrame(virginica.mean())
15
16 #numpy kütüphanesinde bulunan np.cov ve np.array komutlarını kullanarak kovaryans matrislerini hesapladım.
17 setosa_cov_matrix = setosa.iloc[:, :4].cov()
18 versicolor_cov_matrix = versicolor.iloc[:, :4].cov()
19 virginica_cov_matrix = virginica.iloc[:, :4].cov()
20
21 #Saglama yapmak için varyans hesabı yapalım. Cov matrisinde bulunan köşegen değerler varyansı ile eşleşip
22 #setosa'nın varyansını hesaplama
23 setosa_matrix = pd.DataFrame(setosa.iloc[:, :4])
24 setosa_var = setosa_matrix.var() #axis=0 değeriyle her sütünün varyans değerini hesaplıyorum.
25
26 #Versicolor'in varyansını hesaplama
27 versicolor_matrix = pd.DataFrame(versicolor.iloc[:, :4])
28 versicolor_var = versicolor_matrix.var()
29
30 #Virginica'nın varyansını hesaplama
31 virginica_matrix = pd.DataFrame(virginica.iloc[:, :4])
32 virginica_var = virginica_matrix.var()

```

c-) 4 öz niteliğimiz olduğundan dolayı kovaryans matrisi 4x4'lük matris olarak ortaya çıkmıştır. Kovaryans matrisinin köşegenindeki her eleman öz niteliğe bağlı ilgili varyansları vermektedir.

d-) Köşegen dışındaki elemanlar 0'dan farklı değerlerdir çünkü öz nitelikler birbiriyle ilintilidir.