

# Rapor

Elimizdeki verilerin bazı kısımlarında eksiklikler vardı. Öncelikle hangi sütunda kaç veri eksik olduğunu tespit ettim.

```
ID alanındaki boş veri sayısı, 0
Densite alanındaki boş veri sayısı, 0
Polarite alanındaki boş veri sayısı, 77
H2O_Skoru alanındaki boş veri sayısı, 9
O2_Skoru alanındaki boş veri sayısı, 0
Azot_Skoru alanındaki boş veri sayısı, 11
H2O_Oranı alanındaki boş veri sayısı, 7
O2_Oranı alanındaki boş veri sayısı, 2
Azot_Oranı alanındaki boş veri sayısı, 0
Kaynaklar alanındaki boş veri sayısı, 7
Isı_Oranı alanındaki boş veri sayısı, 2
Sınıf Bilgisi alanındaki boş veri sayısı, 3
```

Daha sonra, bu satırları yok saymak verinin büyük bir kısmının kaybına sebep olacağı için, eksik verileri kendi gruplarındaki (yaşanabilir / yaşanamaz) verilerin ortalamaları ile tamamlamaya karar verdim.

Ortalamaların hesaplarken yaşanabilir ortama ait olduğu bilinen fakat kendisi bilinmeyen veriler için yaşanabilir ve kendisi bilinen ortamların verilerinin ortalamasını ve diğer ihtimali kullandım. Ortalama alırken IQR hesabı ile ayırık verileri saptayıp bu verileri hesabın dışında tuttum.

```
ID için limit degerler: (-76.0, 232.0)
Densite için limit degerler: (-5.7250000000000005, 11.175)
Polarite için limit degerler: (-4.61175, 5.00625)
H2O_Skoru için limit degerler: (3.1499999999999986, 8.350000000000001)
O2_Skoru için limit degerler: (-17.16, 47.120000000000005)
Azot_Skoru için limit degerler: (2.05, 4.05)
H2O_Oranı için limit degerler: (-3.8500000000000005, 10.55)
O2_Oranı için limit degerler: (-1.95, 4.05)
Azot_Oranı için limit degerler: (-31.0625, 71.2775)
Kaynaklar için limit degerler: (5.200000000000001, 9.999999999999998)
Isı_Oranı için limit degerler: (-27.035714285714285, 57.82142857142857)
```

Yaşanabilir ortamlar için ort polarite değeri: -0.25542592592592633  
Yaşamamaz ortamlar için ort polarite değeri: 0.14750104166666694  
ort polarite = 0.011413125459897028

Yaşanabilir ortamlar için ort H2O oran değeri: 4.3533333333333335  
Yaşamamaz ortamlar için ort H2O oran değeri: 3.554747474747475  
ort azotkor = 3.8043055555555556

Bu işlemi yaptıktan sonra birbirinden bağımsız ortamlarda bu şekilde ortalama almanın doğruluğu konusunda tereddüte düştüm. Çünkü birbirinden uzak değerler farklı gruplarda bulunabildiği gibi yakın değerler de farklı gruplarda bulunabiliyor (ID = 19,23 ve 106'nın ısı değerleri gibi). İstatistikçi bir arkadaşım üzerinde hesap yapılması gereken verilerde bu tür varsayımsal çıkarımların yapılabileceğini söyledi.

Geriye eksik sınıf bilgilerini tespit etmek kaldı. Burada kullanılan bağıntıyı bilmediğimizden doğru bir sonuç elde etmek mümkün değil. Fakat eksik veri olmayan satırlardaki veriler ile sınıf bilgisi bilinmeyen veriler birbirleri ile oranlanıp belirli bir yüzdelikte benzerliğe sahip olanları yaşanabilir veya yaşanamaz olarak yorumlamak mümkün olabilir. Veya makine öğrenmesi algoritmalarından yardım alınarak bazı çıkarımlar yapılabilir. Ben bu konuda emin olamadığım için herhangi bir müdahalede bulunmadım.

Son olarak tüm bu işlemleri yaptıktan sonra verinin yorumlanabilmesi için bir özet tablosu oluşturdum.

								Kolonlar											
Veriler		Polarite		H2O Skoru		O2 Skoru		Azot Skoru		H2O Oranı		O2 Oranı		Azot Oranı		Kaynaklar		İsı Değeri	
		Yaşanabilir	Yaşanamaz	Yaşanabilir	Yaşanamaz	Yaşanabilir	Yaşanamaz	Yaşanabilir	Yaşanamaz	Yaşanabilir	Yaşanamaz	Yaşanabilir	Yaşanamaz	Yaşanabilir	Yaşanamaz	Yaşanabilir	Yaşanamaz	Yaşanabilir	Yaşanamaz
	Ortalama	-0,25542593	0,14750104	5,663406	5,35285	15,84745	14,61881	2,988889	3,1118	4	3,554747	1,55294	1,235844	22,73627	20,30368	7,020060332	7,06064	6,112129	21,30663
	Mod	(-0,25543, 24)	(0,147501, 52)	(5,6, 5)	(5, 8)	(0, 6)	(6,72, 4)	(3, 8)	(3, 18)	(4,3533333, 6)	(1,5, 13)	(1,3, 13)	(0,2 , 27)	(0, 8)	(0, 7)	(7,48717949, 6)	(7,7, 7)	384615384	(39,5 , 4)
	Medyan	-0,25542593	0,14750104	5,9	5,5	18,48	8,7365	2,806667	3,153333	4,353333333	4,65	1,3	1,45	24	8,955	7,4	7,7	6,5	11,625
	Varyans	0,868830458	1,07731605	2,750843	3,018702	102,9797	127,6491	1,735698	0,337539	0,350431373	4,712037	2,65779	1,777455	157,2328	260,1872	2,26497137	6,268661	2,262033	463,9746
	Standart Sapma	0,932110754	1,03793836	1,658567	1,737441	10,14789	11,29819	1,317459	0,580981	0,591972442	2,170723	1,63027	1,333212	12,53925	16,13032	1,504982183	2,503729	1,504006	21,54007
	IQR Sınırları	(-4.61175, 5.00625)		(3.1499999, 8.350000)		(-17.16, 47.12000)		(2.05, 4.05)		(-3.8500000, 10.55)		(-1.95, 4.05)		(-31.0625, 71.2775)		(5.200000, 9.999999)		(-27.03571, 57.82143)	