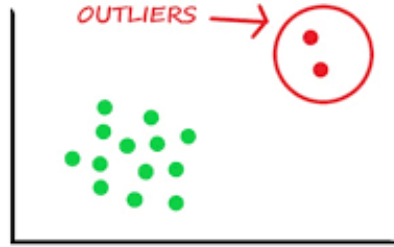




Aykırı değerler (Outliers)

Aykırı değer denilen değerler verilerin izlediği çizgiden bir sapma oluşturan ve verinin kimliğini doğru temsil etmeyen verilerdir. Aşağıda gösterilen veri setindeki kırmızı veriler çoğunluktan farklı bir davranış göstererek marjinal veri olmuşlardır. Yani Aykırı değer olmuşlardır.



Bunu örnek bir veri kümesi ile açıklayalım. Aşağıda bir şirkette çalışan insanların Ünvanları ve maaşları hakkındaki bilgileri görüyoruz. Ancak müdür ünvanındaki kişi doğrusal artışa zıtlık gösteriyor. Burda memur değeri marjinal davranarak outlier kategorisine giriyor.

ID	UNVAN	MAAŞ
1	Çaycı	4250
2	Sekreter	6000
3	Memur	8000
4	Müdür Yardımcısı	6000
5	Müdür	10000
6	Genel Müdür	14000
7	CEO	30000

Aykırı değerlerin oluşma sebepleri

1. Veri girişinde oluşabilecek hatalar.
2. Verideki bozulmalar
3. Herhangi bir hatadan kaynaklanmayan gerçekten aykırı davranan değerler (Örn. bir takımda oynayan üstün yetenekli bir oyuncunun üstün performans göstermesi.)

Aykırı değer tipleri nelerdir

1. **Noktasal aykırı değerler** - Verinin tümünden aykırı davranış gösteren tek bir ya da birkaç değer. *Örneğin sınıf ortalaması 24 olan bir sınıfta bir öğrencinin 100 alması noktasal aykırı değerdir*
2. **Bağlamsal aykırı değerler** - Bir gözlemin belirli bir bağlam olarak veri setinin geri kalanından farklı olduğu durumlara denir. *Örneğin ekonominin pandemi ile beraber kötüye gitmesi, Black Friday döneminde satışların fırlaması.*
3. **Toplu aykırı değerler** - Veri seti içerisindeki bir grubun kendi içinde karşılaştırıldığında normal tüm grupla karşılaştırıldığında marjinal olması toplu aykırı değerlere örnektir. *Örneğin bir ülkede satışlar durgunken ülkenin yalnızca bir şehrinde satışların yükselmesi.*

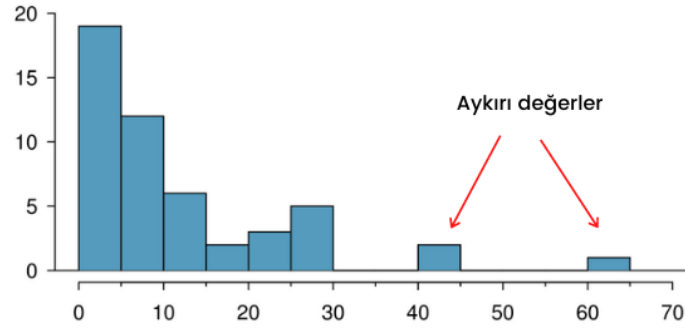
Aykırı değerler nasıl tespit edilir?

Yukarıda çalışmış olduğumuz veriler kümesinde bu tespiti yapmak bizim için zor olmadı çünkü az veri ile çalışıyorduk. Büyük verilerle çalıştığımız zamanlarda ise aykırı verileri bulmak için bazı yöntemler kullanacağız.

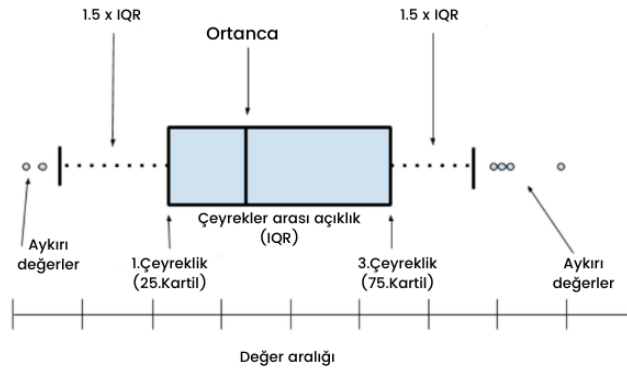
1. Grafikler

- Tek değişkenli grafikler

Histogram gibi grafikleri aykırı değerler için kullanmak mümkündür. Örneğin aşağıdaki grafik bir histogram grafiğidir. Bunları çizdirmek için ileride seaborn kütüphanesini kullanacağız.



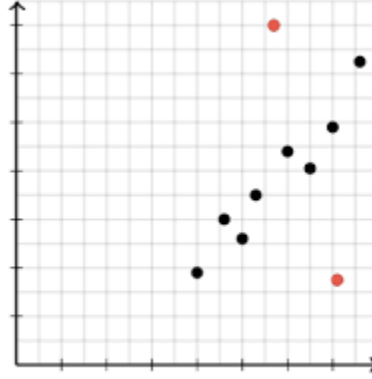
Kutu grafiklerindeki Aykırı değerleri bulmak için kullanabiliriz.



Aykırı değerler şekilde belirtilmiştir.

- Çok değişkenli grafikler

Scatter plot yani noktasal olarak saçılım grafiğindeki aykırı değerlerin tespiti için kullanmak mümkündür.



Şekilde kırmızı nokta ile belirtilen veriler aykırı değerlerdir.

Artıkları kullanarak eksik ve aykırı değerleri işleme

Öncelikle artık teriminin ne olduğuna değinelim. Artık, Modelin başarızılığını temsil eder. Sonuç ile tahmin arasındaki farklar olarakta tanımlanabilir.

Artık değerler için izlenebilecek yollar

- Aykırı değerleri kaldırabilirsiniz.
- Aykırı değerleri medyan ya da ortalama gibi değerlere atayabilirsiniz.
- Değerin ne olması gerektiğini tahmin edebilirsiniz.

Bunun için regresyon modellerini kullanabilirsiniz.