* Veri temizleme
  + Gürültülü veri(erkek hamile gözüküyorsa, biberon 1M dolar gözükütor)
  + Eksik veri analizi
  + Aykırı gözlem (outlier analysis)
* Veri standardizasyonu (data standardization, feature scaling)
  + 0-1 dönüşümü (normalizasyon)
  + Z-skoruna dönüştürme (standardizasyon)
  + Logaritmik dönüşüm
* Veri indirgeme (data reduction)
  + Gözlem(observation) sayısının azaltılması
  + Değişken sayısının azaltılması
* Değişken dönüşümleri (variable transformation)
  + Sürekli değişkenlerde dönüşümler
  + Kategorik değişkenlerde dönüşümler

Aykırı gözlem: genellenebilirlik kaygısıyla oluşturulan kural setlerini veya fonksiyonları yanıltır. Yanlılığa sebep olur.

Aykırı gözlem yakalama (genelde 1ve4)

1. Sektör bilgisi:1000 m2 evleri modele almamak tr de ort %97 evler 2+1
2. Standart sapma yaklaşımı: eşik değer belirlenir eşik d=ort+standart sapma, eşik d=ort+2\*standart sapma, eşik d=ort+3\*standart sapma
3. Z-skoru yaklaşımı: standart sapma yöntemine benzer şekilde çalışır. Değişken standart normal dağılıma uyarlanır, yani standartlaştırılır. Sonrasında -örneğin- dağılımın sağından ve solundan -+2,5 değerine göre bir eşik değer konulur ve bu değerin üzerinde ya da altında olan değerler aykırı değer olarak işaretlenir.
4. Boxplot(interquartile range - IQR) yöntemi: en sık kullanılan yöntemlerden birisi. D.lerin d.leri küç büyüğe sıralanır. Çeyreklerine yani Q1,Q3 d.lerine karşılık gelen d.ler üzerinden bir eşik değer hesaplanır ve bu eşik değere göre aykırı değer tanımı yapılır. IQR=1.5\*(Q3-Q1) , alt eşik değer=Q1-IQR , üst eşik değer=Q3+IQR