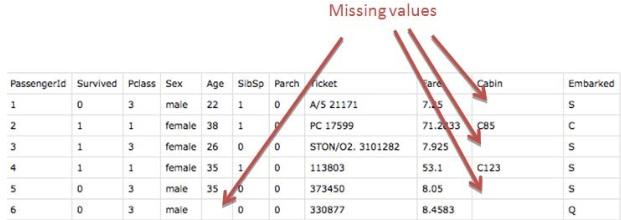
**Missing Data**

[jupyter notebook #1](file:///C:\Users\Arslan\Desktop\GitHub\daily_work_file\08.11.2022\08.11.2022.ipynb)



Gerçek dünya koşullarında, missing value her zaman bir sorundur. Makine öğrenimi ve veri madenciliği, kayıp değerlerin neden olduğu düşük veri kalitesi nedeniyle model tahmini doğruluğu açısından önemli zorluklarla karşı karşıyadır. Eksik değer tedavisi, modellerinin doğruluğunu ve geçerliliğini geliştirmek için bu alanlarda birincil odak noktasıdır.

Eksik veriler, manuel veri giriş teknikleri, ekipman arızaları ve yanlış ölçümler dahil olmak üzere çeşitli nedenlerle oluşur. Bir veri kümesinde NaN, Not a Number anlamına gelir. Pandas, eksik verileri algılamak için isnull() ve notnull() işlevlerini sağlar.

Bu derste şunları öğreneceksiniz:

DataFrame'de eksik değerler nasıl işlenir,

tablo içeren bir resim

Açıklama otomatik olarak oluşturuldu

dropna methodu axis argumentiyle missing value drop eder

metin içeren bir resim

Açıklama otomatik olarak oluşturuldu

metin içeren bir resim

Açıklama otomatik olarak oluşturuldu



fillna methodu, NA değerlerini boş olmayan verilerle "doldurabilir"

metin, tablo içeren bir resim

Açıklama otomatik olarak oluşturuldu

tablo içeren bir resim

Açıklama otomatik olarak oluşturuldu

replace methodu, fillna gibi NaN degerleri boş olmayan degerlerle değiştirir

metin içeren bir resim

Açıklama otomatik olarak oluşturuldu

interpolate yöntemi, eksik değerleri doldurmak için çeşitli interpolation teknikleri kullanır.

metin içeren bir resim

Açıklama otomatik olarak oluşturuldu

**Outliers**

EDA aşamasında, gelecekte Makine Öğrenimi modelleme sürecinin etkilenmemesi için verilere bazı önlemler uygulanacak. Eksik veri işlemenin yanı sıra, bir sonraki işlem verilerimizdeki aykırı değerleri ele almalıdır.

Aykırı değer, Wikipedia'ya göre diğer gözlemlerden uzak olan bir gözlem noktasıdır.

Veri kümemizdeki bazı değerlerin uç değere dönüşmesinin veri toplama hataları veya veri kümesindeki varyans gibi birçok farklı nedeni olabilir. Yukarıdaki resim, bir veri çerçeveleri tablosu sütunundaki aykırı değerleri göstermektedir. Kutu grafiğinin her iki tarafındaki siyah noktalar, aykırı değerleri gösterir.

Aykırı değerlerin bulunabilmesi için IQR yöntemi kullanılır. Wikipedia'dan;

Orta veya orta %50 veya teknik olarak H-spread olarak da adlandırılan çeyrekler arası aralık (IQR), istatistiksel dağılımın bir ölçüsüdür ve 75. ve 25. yüzdelikler arasındaki veya üst ve alt çeyrekler arasındaki farka eşittir, IQR = Q3 - S1.

[jupyter notebook #2](file:///C:\Users\Arslan\Desktop\GitHub\daily_work_file\08.11.2022\08.11.2022.ipynb)

tablo içeren bir resim

Açıklama otomatik olarak oluşturuldu

tablo içeren bir resim

Açıklama otomatik olarak oluşturuldu

Esas olarak bu aykırı değerlerden kurtulmanın iki yolu vardır.

1) IQR yöntemi ile aykırı değerlerin kaldırılması

metin içeren bir resim

Açıklama otomatik olarak oluşturuldu

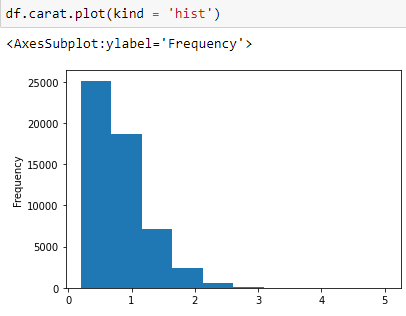
tablo içeren bir resim

Açıklama otomatik olarak oluşturuldu

2) Winsorize yöntemi ile bunları aykırı olmayan değerlere dönüştürmek

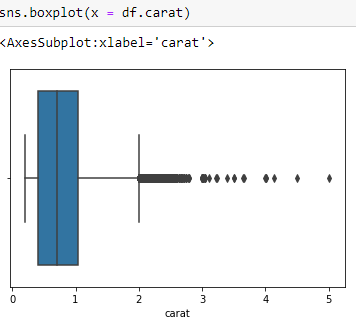


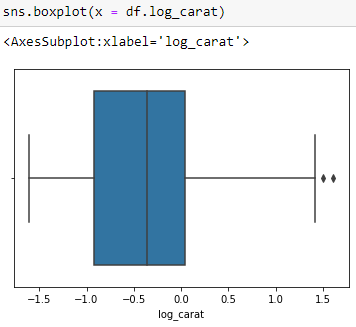
Log transformation method



tablo içeren bir resim

Açıklama otomatik olarak oluşturuldu





[Outlier detection and removal using IQR | Feature engineering tutorial python # 4](https://www.youtube.com/watch?v=A3gClkblXK8&t=1s)

[jupyter notebook #3](file:///C:\Users\Arslan\Desktop\GitHub\daily_work_file\08.11.2022\08.11.2022.ipynb)