**1. [1.] Missing data and outliers [Eksik veriler ve aykırı değerler]**

In this lesson, we'll take a step back and explore a few scenarios prior to the model building stage. [Bu derste, bir adım geriye gideceğiz ve model oluşturma aşamasından önceki birkaç senaryoyu inceleyeceğiz.] We'll look at how to handle missing data and then deal with outliers. [Eksik verilerin nasıl ele alınacağına bakacağız ve ardından aykırı değerlerle ilgileneceğiz.]

**2. [2.] Handling missing data [Eksik verileri işleme]**

How do we identify and correct for null values in our dataset? [Veri kümemizdeki boş değerleri nasıl belirler ve düzeltiriz?] There are two common approaches. [İki ortak yaklaşım vardır.] First is dropping the whole row when you detect a missing value, and second is imputing the missing values with some other value. [Birincisi, eksik bir değer tespit ettiğinizde tüm satırı bırakmak, ikincisi ise eksik değerleri başka bir değerle yüklemektir.]

**3. [3.] Drop the whole row [Tüm satırı bırak]**

Dropping the whole row is likely the simplest approach of correcting for null values, as it can be done in one line of code. [Tüm satırı bırakmak, tek bir kod satırında yapılabileceğinden, boş değerleri düzeltmenin en basit yaklaşımıdır.] However, there are some trade offs to consider. [Ancak, dikkate alınması gereken bazı takaslar var.] By dropping any rows with a null value, you could potentially lose a significant portion of your dataset and exclude information that could strengthen your model or produce insights. [Boş değere sahip herhangi bir satırı bırakarak, potansiyel olarak veri kümenizin önemli bir bölümünü kaybedebilir ve modelinizi güçlendirebilecek veya içgörüler üretebilecek bilgileri hariç tutabilirsiniz.] In this example, the line of code would drop rows 4 and 6, since they contain at least one null value. [Bu örnekte, kod satırı en az bir boş değer içerdiğinden 4. ve 6. satırları bırakacaktır.]

**4. [4.] Impute missing values [Eksik değerleri yükle]**

Another option is to impute missing values for the nulls. [Başka bir seçenek de boş değerler için eksik değerler yüklemektir.] This approach takes a little more thought, but allows you to preserve the information contained in the rows with some null values. [Bu yaklaşım biraz daha düşünmeyi gerektirir, ancak satırlarda yer alan bilgileri bazı boş değerlerle korumanıza izin verir.] There are a few popular ways to impute values. [Değerleri empoze etmenin birkaç popüler yolu vardır.] You can insert a constant value, like 0, insert a randomly selected record from another observation, use the mean, median, or mode, or use another model to predict the value and impute that as well. [0 gibi sabit bir değer ekleyebilir, başka bir gözlemden rastgele seçilmiş bir kayıt ekleyebilir, ortalamayı, medyanı veya modu kullanabilir veya değeri tahmin etmek ve onu da ima etmek için başka bir model kullanabilirsiniz.]

**5. [5.] A few useful functions [Birkaç kullanışlı fonksiyon]**

Let's discuss a few useful functions for these techniques. [Bu teknikler için birkaç yararlı işlevi tartışalım.] Isnull is a pandas function that identifies any rows in that have a null value. [Isnull, boş değeri olan tüm satırları tanımlayan bir panda işlevidir.] You can also take it a step further and specify which fields must be null. [Ayrıca bir adım daha ileri gidebilir ve hangi alanların boş olması gerektiğini belirleyebilirsiniz.] Similarly, you can use the pandas dropna function if you want to drop the rows outright; either all of them or a specified subset. [Benzer şekilde, satırları doğrudan bırakmak istiyorsanız, pandas dropna işlevini kullanabilirsiniz; ya hepsi ya da belirli bir alt küme.] Fillna is useful for imputation; you specify a technique and the DataFrame fills in the nulls. [Fillna, isnat için kullanışlıdır; bir teknik belirlersiniz ve DataFrame boş değerleri doldurur.]

**6. [6.] Dealing with outliers [Aykırı değerlerle başa çıkmak]**

Moving on to outliers, there are a few different ways to use statistics to identify outliers, including standard deviation, or z-scores, as well as interquartile range, commonly referred to as IQR. [Aykırı değerlere geçersek, standart sapma veya z puanları ve ayrıca genellikle IQR olarak adlandırılan çeyrekler arası aralık dahil olmak üzere aykırı değerleri belirlemek için istatistikleri kullanmanın birkaç farklı yolu vardır.]

**7. [7.] Standard deviations [Standart sapma]**

Using standard deviations is a popular, straightforward method for identifying outliers and is the most likely to come up in your interview. [Standart sapmaları kullanmak, aykırı değerleri belirlemek için popüler ve basit bir yöntemdir ve görüşmenizde ortaya çıkması en olası yöntemdir.] Quite simply, any observation that falls outside of 3 standard deviations of the mean is deemed an outlier. [Oldukça basit bir şekilde, ortalamanın 3 standart sapmasının dışında kalan herhangi bir gözlem aykırı değer olarak kabul edilir.] On the normal curve shown, the tails make up around 0 point 1 percent of the population; anything past this threshold is considered an outlier. [Gösterilen normal eğride, kuyruklar popülasyonun yaklaşık yüzde 0 virgül 1'ini oluşturur; bu eşiği aşan her şey aykırı değer olarak kabul edilir.]

1. 1 Wikimedia [1 Wikimedia]

**8. [8.] Interquartile range (IQR) [Çeyrekler arası aralık (IQR)]**

Using IQR is another way to determine whether or not a value is an outlier. [IQR kullanmak, bir değerin aykırı değer olup olmadığını belirlemenin başka bir yoludur.] Recall box plots from earlier in the course. [Kursun önceki bölümlerinden kutu çizimlerini hatırlayın.] You can summarize your data pretty effectively in one plot using the median, quartiles, and range. [Verilerinizi medyan, çeyrekler ve aralığı kullanarak tek bir grafikte oldukça etkili bir şekilde özetleyebilirsiniz.] The IQR is computed by subtracting the first quartile from the third quartile. [IQR, üçüncü çeyrekten ilk çeyreği çıkararak hesaplanır.] Using this value, you can set outlier thresholds with the formula you see at the tail ends of the boxplot. [Bu değeri kullanarak, boxplot'un kuyruk uçlarında gördüğünüz formül ile aykırı değer eşikleri belirleyebilirsiniz.] You take 1 point 5 times the IQR and then add and subtract that from the first and third quartile. [IQR'nin 5 katı 1 puan alırsınız ve ardından bunu birinci ve üçüncü çeyrekten toplar ve çıkarırsınız.] When you generate a box plot, you'll see these outliers represented as dots outside of the end points in the picture. [Bir kutu grafiği oluşturduğunuzda, bu aykırı değerlerin resimdeki bitiş noktalarının dışında noktalar olarak temsil edildiğini göreceksiniz.]

1. 1 Wikimedia [1 Wikimedia]

**9. [9.] Summary [Özet]**

Let's summarize what we learned. [Öğrendiklerimizi özetleyelim.] We discussed ways to handle null values, including dropping a whole row or deciding on a value to impute in place of a null value. [Bir satırın tamamını bırakmak veya boş bir değer yerine atılacak bir değere karar vermek de dahil olmak üzere boş değerleri işlemenin yollarını tartıştık.] Then, we talked about identifying outliers using the three standard deviation threshold and interquartile range. [Ardından, üç standart sapma eşiğini ve çeyrekler arası aralığı kullanarak aykırı değerleri belirlemekten bahsettik.]

**10. [10.] Let's prepare for the interview! [Röportaj için hazırlanalım!]**

Let's get to work, and go try this out on some exercises! [Hadi işe koyulalım ve gidip bunu bazı egzersizlerde deneyelim!]