**1. [1.] Categorical data [kategorik veriler]**

So far in this course, we've spent a lot of our time reviewing concepts and then testing our understanding of those concepts with coding exercises. [Bu kursta şimdiye kadar, kavramları gözden geçirmek ve ardından bu kavramları anlamamızı kodlama alıştırmalarıyla test etmek için çok zaman harcadık.] In this lesson, we'll take a more practical approach as we work through some exploratory data analysis with categorical data, similar to something you might see in a coding or take-home assessment. [Bu derste, bir kodlama veya eve götürme değerlendirmesinde görebileceğiniz bir şeye benzer şekilde, kategorik verilerle bazı keşifsel veri analizleri üzerinde çalışırken daha pratik bir yaklaşım izleyeceğiz.]

**2. [2.] Types of variables [Değişken türleri]**

Categorical features can only take on a limited, and usually fixed, number of possible values. [Kategorik özellikler yalnızca sınırlı ve genellikle sabit sayıda olası değer alabilir.] As we see here, there are different types of categorical data. [Burada gördüğümüz gibi, farklı kategorik veri türleri vardır.] The first type is ordinal. [İlk tip sıralı.] Ordinal data, like it sounds, takes some sort of order. [Sıralı veriler, göründüğü gibi, bir tür düzen alır.] An example is the number of stars given in a movie review, where 1 star represents a poor review and 5 stars represents an excellent review. [Bir film incelemesinde verilen yıldız sayısı buna bir örnektir; burada 1 yıldız kötü bir incelemeyi ve 5 yıldız mükemmel bir incelemeyi temsil eder.] For nominal variables, things are a bit different since order doesn't matter. [Nominal değişkenler için, sıra önemli olmadığından işler biraz farklıdır.] Examples of this include gender or eye color. [Bunun örnekleri arasında cinsiyet veya göz rengi bulunur.]

**3. [3.] Encoding categorical data [Kategorik verileri kodlama]**

It won't always come up during exploratory data analysis, but if you're performing any type of machine learning you may have to encode your categorical variables as something else. [Keşif amaçlı veri analizi sırasında her zaman ortaya çıkmaz, ancak herhangi bir tür makine öğrenimi yapıyorsanız, kategorik değişkenlerinizi başka bir şey olarak kodlamanız gerekebilir.] Let's start simple with label encoding. [Basit bir şekilde etiket kodlamasıyla başlayalım.] Label encoding involves mapping each value to a number as you can see here; note that these numbers have no relationship with each other. [Etiket kodlaması, burada görebileceğiniz gibi her bir değeri bir sayı ile eşleştirmeyi içerir; Bu sayıların birbirleriyle hiçbir ilişkisi olmadığını unutmayın.] Another popular technique is one-hot encoding. [Bir başka popüler teknik, tek sıcak kodlamadır.] One-hot encoding maps each category to its own column containing a 0 or 1 to indicate if the observation has that feature or not. [Tek-sıcak kodlama, gözlemin bu özelliğe sahip olup olmadığını belirtmek için her kategoriyi 0 veya 1 içeren kendi sütununa eşler.] The preprocessing package in scikit-learn and the fit-underscore-transform function are helpful here, along with the pandas get-underscore-dummies function as well. [scikit-learn'deki ön işleme paketi ve sığdır-alt çizgi-dönüştür işlevi burada, pandaların alt çizgi-aptallar işleviyle birlikte yararlıdır.]

1. 1 What is One Hot Encoding and How to Do It [1 Tek Sıcak Kodlama Nedir ve Nasıl Yapılır?]

**4. [4.] Example: laptop models [Örnek: dizüstü bilgisayar modelleri]**

Let's introduce a real dataset and work through some initial exploratory data analysis. [Gerçek bir veri kümesini tanıtalım ve bazı ilk keşifsel veri analizleri üzerinde çalışalım.] You can see from the head of this DataFrame, the dataset includes the company, model, and the accompanying price for popular laptops. [Bu DataFrame'in başlığından görebilirsiniz, veri kümesi popüler dizüstü bilgisayarlar için şirket, model ve beraberindeki fiyatı içerir.]

**5. [5.] Example: laptop models [Örnek: dizüstü bilgisayar modelleri]**

Once we see what we're working with, let's hone in on the company column and see how Apple, Lenovo, and Dell stack up against each other in terms of the number of observations for each value. [Neyle çalıştığımızı gördükten sonra, şirket sütununa odaklanalım ve Apple, Lenovo ve Dell'in her bir değer için gözlem sayısı açısından nasıl bir araya geldiğini görelim.] From the bar plot, it's clear that there are far more Lenovo and Dell models in our dataset than Apple models. [Çubuk grafiğinden, veri kümemizde Apple modellerinden çok daha fazla Lenovo ve Dell modeli olduğu açıktır.]

**6. [6.] Box plots [kutu araziler]**

Quick side note here: box plots are a great tool when you want to pack a bunch of information into one visualization. [Burada hızlı bir not: kutu çizimleri, bir grup bilgiyi tek bir görselleştirmede paketlemek istediğinizde harika bir araçtır.] The vertical marks on the lines represent the minimum, 25th percentile, median, 75th percentile, and the maximum, while any circles outside of this represent outliers. [Çizgilerdeki dikey işaretler minimum, 25. yüzdelik, medyan, 75. yüzdelik ve maksimum değerleri temsil ederken, bunun dışındaki tüm daireler aykırı değerleri temsil eder.]

1. 1 Wikimedia [1 Wikimedia]

**7. [7.] Example: laptop models [Örnek: dizüstü bilgisayar modelleri]**

Let's see how the laptops prices for each brand look using box plots. [Her marka için dizüstü bilgisayar fiyatlarının kutu çizimlerini kullanarak nasıl göründüğünü görelim.] You can see each company on the y-axis and the price in euros on the x-axis. [Her şirketi y ekseninde, fiyatı ise x ekseninde Euro cinsinden görebilirsiniz.] The plot tells us that, on average, Apple laptops are more expensive than Dell or Lenovo laptops. [Arsa, ortalama olarak Apple dizüstü bilgisayarların Dell veya Lenovo dizüstü bilgisayarlardan daha pahalı olduğunu söylüyor.] However, we can see from the outliers that Dell and Lenovo have some high-end models that are far more expensive. [Ancak, Dell ve Lenovo'nun çok daha pahalı bazı üst düzey modellere sahip olduğunu aykırı değerlerden görebiliyoruz.]

**8. [8.] Summary [Özet]**

Let's summarize what we learned here. [Burada öğrendiklerimizi özetleyelim.] We covered variable types, different encoding techniques, and walked through a surface-level exploration of a dataset. [Değişken türlerini, farklı kodlama tekniklerini ele aldık ve bir veri kümesinin yüzey düzeyinde keşfinden geçtik.]

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

Let's get to it and practice these concepts! [Hadi ona gidelim ve bu kavramları uygulayalım!]