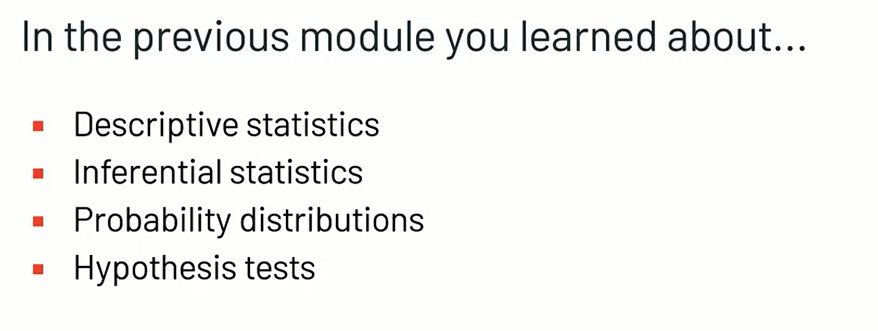
**Learning Objectives**

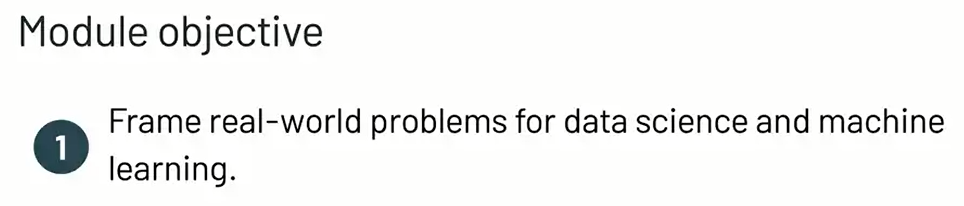
* Develop multiple research questions from your own industry or interest area that can be solved with hypothesis tests, regression, classification, and clustering
* Give examples of clustering problems
* Differentiate between regression and classification problems
* Differentiate between supervised learning and unsupervised learning
* Describe, at a high-level, how machine learning can be used to solve a given scenario
* Differentiate between machine learning and other ways to solve data science questions
* Describe the challenges associated with answering research questions using only hypothesis testing
* Explain why it is important to frame questions in a way that enable them to be answered with data

# **Module and Lesson Intro**

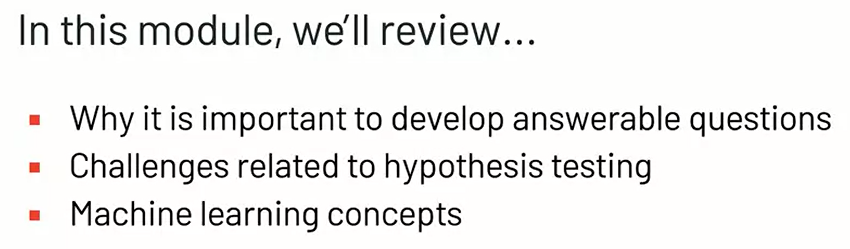
Hi everyone, welcome to another module of Data Science Fundamentals for data analyst. [Herkese merhaba, veri analisti için başka bir Veri Bilimi Temelleri modülüne hoş geldiniz.] We hope that you're enjoying the course so far and we want to take a few moments to review what you did in the previous module. [Şimdiye kadar kurstan keyif aldığınızı umuyoruz ve bir önceki modülde yaptıklarınızı gözden geçirmek için birkaç dakikanızı ayırmak istiyoruz.]



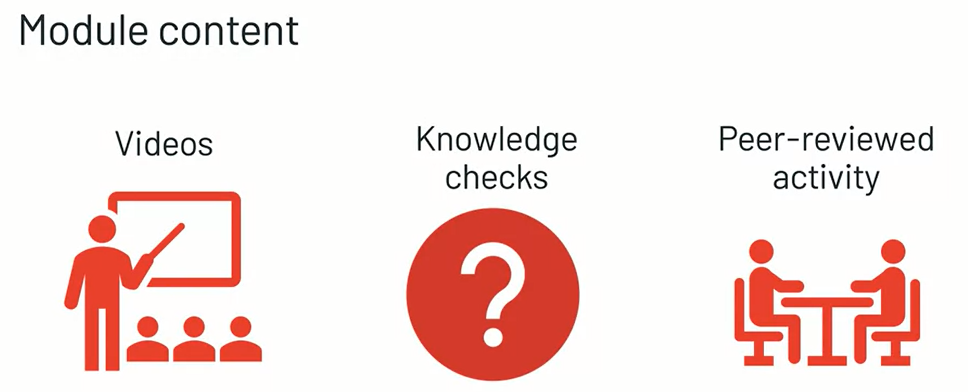
In the previous module, you learned about descriptive and inferential statistics, probability distributions and how probability can be used to design and conduct hypothesis tests to learn from data. [Önceki modülde, tanımlayıcı ve çıkarımsal istatistikler, olasılık dağılımları ve veriden öğrenmek için hipotez testleri tasarlamak ve yürütmek için olasılığın nasıl kullanılabileceğini öğrendiniz.] Building these hands-on technical skills has you on the right path and developing your data science skills. [Bu uygulamalı teknik becerileri oluşturmak, doğru yolda olmanızı ve veri bilimi becerilerinizi geliştirmenizi sağlar.] While there more hands on skills to be learned, we need to spend some time considering one of the most important learning objectives in this entire course. [Öğrenilecek daha çok uygulamalı beceri olsa da, bu kursun tamamındaki en önemli öğrenme hedeflerinden birini düşünmek için biraz zaman harcamamız gerekiyor.]



That objective is framing real-world problems for data science and machine learning. [Bu hedef, veri bilimi ve makine öğrenimi için gerçek dünyadaki sorunları çerçevelemek.] That's what you'll be able to do by the end of this module. [Bu modülün sonunda yapabilecekleriniz bu.]



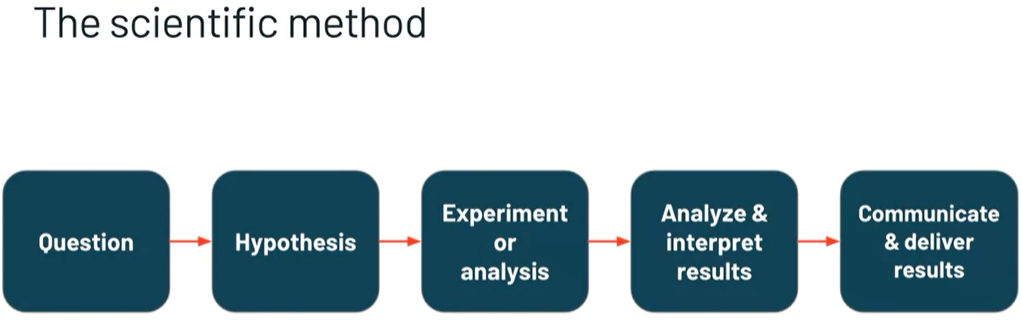
Along the way, we'll describe the importance of developing answerable questions. [Bu arada, cevaplanabilir sorular geliştirmenin önemini açıklayacağız.] Describe the challenges associated with hypothesis testing and also introduce some machine learning concepts. [Hipotez testiyle ilgili zorlukları tanımlayın ve ayrıca bazı makine öğrenimi kavramlarını tanıtın.]



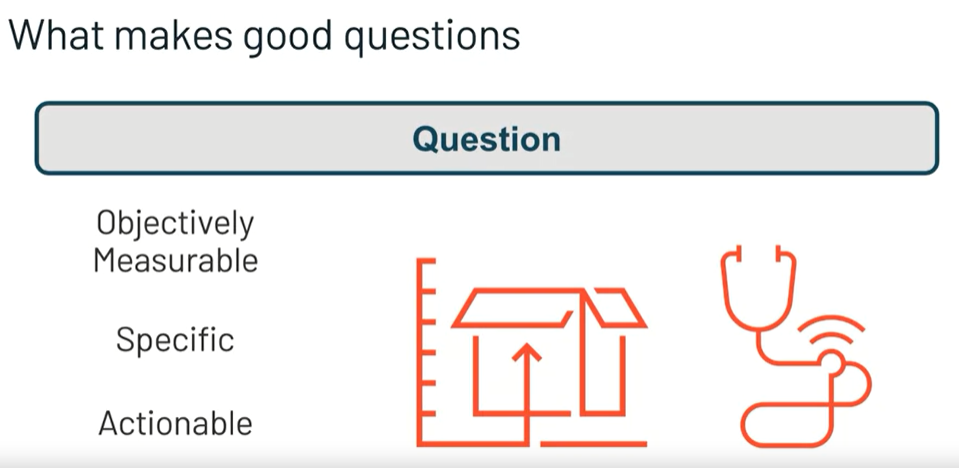
This module will be taught in a series of videos, knowledge checks and a peer reviewed activity. [Bu modül, bir dizi video, bilgi kontrolleri ve akran tarafından gözden geçirilen bir etkinlikle öğretilecektir.] When you're ready to learn how to frame real-world scenarios as question to begin to data science and machine learning process, click onto the next video. [Veri bilimi ve makine öğrenimi sürecine başlamak için gerçek dünya senaryolarını soru olarak nasıl çerçeveleyeceğinizi öğrenmeye hazır olduğunuzda, sonraki videoya tıklayın.] Good luck. [İyi şanlar.]

# **Why Good Questions Matter**

Hello and welcome back, in this video we're going to discuss why it's important to frame real-world scenarios so that they can be easily answered with data. [Merhaba ve tekrar hoş geldiniz, bu videoda gerçek dünya senaryolarını, verilerle kolayca yanıtlanabilmeleri için çerçevelemenin neden önemli olduğunu tartışacağız.]



Remember the scientific method that we've been using to design data science processes. [Veri bilimi süreçlerini tasarlamak için kullandığımız bilimsel yöntemi hatırlayın.] We talked about how the entire data science process starts with a question, and for that reason it's crucial for that question to be good. [Tüm veri bilimi sürecinin bir soruyla nasıl başladığını konuştuk ve bu nedenle bu sorunun iyi olması çok önemli.] But what actually makes a question good? [Ama aslında bir soruyu iyi yapan nedir?]



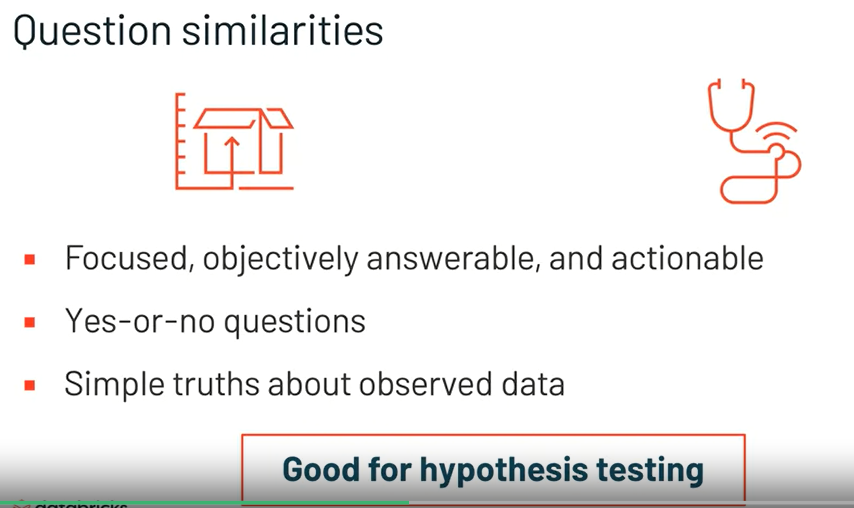
In the first module, we said that good questions provide answers that are objectively measurable. [İlk modülde iyi soruların objektif olarak ölçülebilir cevaplar verdiğini söylemiştik.] In data science, we want to apply our own knowledge in the areas of applied statistics, computer science, and within the domain. [Veri biliminde, uygulamalı istatistik, bilgisayar bilimi ve alan içinde kendi bilgimizi uygulamak istiyoruz.] But we want our answers to come from the data we have available. [Ancak cevaplarımızın elimizdeki verilerden gelmesini istiyoruz.] This is where tools like hypothesis tests come in handy. [Hipotez testleri gibi araçların kullanışlı olduğu yer burasıdır.] When we ask questions like, are out of stock items a problem? [Stokta olmayan ürünler sorun olur mu gibi sorular sorduğumuzda.] Or does medicine work? [Yoksa tıp işe yarıyor mu?] We're asking questions that are really large in scope and they're tough to answer objectively. [Kapsamı gerçekten geniş olan ve objektif olarak yanıtlanması zor sorular soruyoruz.] What is a problem when it comes to out of stock items? [Stokta olmayan ürünler söz konusu olduğunda sorun nedir?] Is it a problem if a single item is out of stock? [Tek bir ürün stokta yoksa sorun olur mu?] What about if 100 items are out of stock? [Peki ya 100 ürün stokta yoksa?] What about if items are more likely to be out of stock in certain communities, is that a problem? [Peki ya belirli topluluklarda ürünlerin stokta kalma olasılığı daha yüksekse, bu bir sorun mu?] We can ask similar questions about medicine. [Tıp hakkında da benzer sorular sorabiliriz.] What does it mean for medicine to work? [Tıbbın işe yaraması ne anlama geliyor?] Is medicine supposed to eliminate all health issues? [Tıbbın tüm sağlık sorunlarını ortadan kaldırması mı gerekiyor?] Is it supposed to extend life expectancy? [Ömrü uzatmak mı gerekiyor?] And is medicine the entire medical field or is it just something that's administered to people who need treatment for some illness? [Ve tıp tüm tıp alanı mı yoksa sadece bir hastalık için tedaviye ihtiyacı olan insanlara verilen bir şey mi?] The major issue with these questions is their lack of specificity. [Bu sorularla ilgili en büyük sorun, özgüllüklerinin olmamasıdır.] When our questions are too broad in scope, it's difficult to arrive at answers that are objective, easily understood and actionable. [Sorularımızın kapsamı çok geniş olduğunda, nesnel, kolay anlaşılır ve eyleme geçirilebilir yanıtlara ulaşmak zordur.] Even if we deemed that medicine does not work, what would we actually recommend be done about that? [İlacın işe yaramadığını düşünsek bile bu konuda aslında ne yapılmasını tavsiye ederiz?] We wouldn't know where to start, and one of the key goals of data science is to further our understanding in application within a specific domain. [Nereden başlayacağımızı bilemezdik ve veri biliminin temel hedeflerinden biri, belirli bir alandaki uygulama anlayışımızı ilerletmektir.] Remember, that's right there in our data science definition, so what do we do? [Unutmayın, bu bizim veri bilimi tanımımızda var, o halde ne yapıyoruz?] We ask more pointed questions, when we ask questions like do we

have more out of stock items in rural communities than in suburban communities? [Kırsal topluluklarda banliyö topluluklarından daha fazla stokumuz var mı? gibi sorular sorduğumuzda daha anlamlı sorular soruyoruz.] Or does this new medical treatment reduced the average recovery time for a patient? [Yoksa bu yeni tıbbi tedavi, bir hasta için ortalama iyileşme süresini kısaltıyor mu?] We're asking questions that can be answered objectively using hypothesis tests. [Hipotez testleri kullanılarak objektif olarak cevaplanabilecek sorular soruyoruz.] We can design and carry out experiments, collect data, and perform analysis on that data to objectively answer those questions. [Bu soruları nesnel olarak yanıtlamak için deneyler tasarlayabilir ve yürütebilir, veri toplayabilir ve bu veriler üzerinde analiz yapabiliriz.] This is because they're specific in their descriptions and

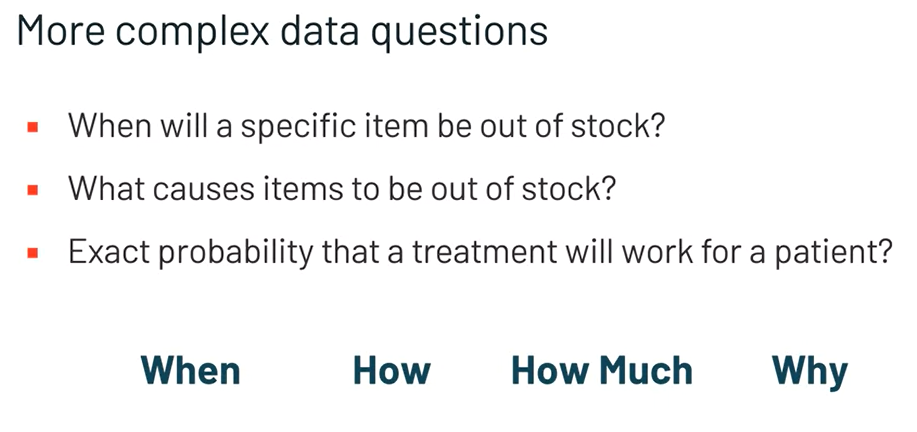
their potential answers don't allow any room for interpretation. [Bunun nedeni, açıklamalarında belirli olmaları ve olası cevaplarının yoruma yer bırakmamasıdır.] Another key point to make about these questions is that they're actionable. [Bu sorularla ilgili olarak belirtilmesi gereken bir diğer önemli nokta, eyleme geçirilebilir olmalarıdır.] If we find that there are more out of stock items in rural communities than suburban communities, we can design more data science projects to determine why that might be. [Kırsal topluluklarda banliyö topluluklarından daha fazla stokta olmayan ürün olduğunu tespit edersek, bunun neden olabileceğini belirlemek için daha fazla veri bilimi projesi tasarlayabiliriz.] And then, we can take action to make product distribution more equitable. [Ardından, ürün dağıtımını daha adil hale getirmek için harekete geçebiliriz.] This might look like investing more in delivery trucks and delivery drivers in rural or urban communities. [Bu, kırsal veya kentsel topluluklarda teslimat kamyonlarına ve teslimat sürücülerine daha fazla yatırım yapmak gibi görünebilir.] Or it might even mean investing in a distribution center more central to the affected communities. [Veya etkilenen topluluklar için daha merkezi bir dağıtım merkezine yatırım yapmak anlamına da gelebilir.] Similarly, if we find that the new medical treatment does reduce the average recovery time for a patient, we could begin educating medical professionals on that treatment and distributing the treatment to medical centers around the world. [Benzer şekilde, yeni tıbbi tedavinin bir hasta için ortalama iyileşme süresini azalttığını tespit edersek, tıp uzmanlarını bu tedavi konusunda eğitmeye ve tedaviyi dünya çapındaki tıp merkezlerine dağıtmaya başlayabiliriz.] This is why good questions matter, they enable the use of data, the use of analysis techniques and objective in actionable results. [Bu nedenle iyi sorular önemlidir, verilerin kullanılmasını, analiz tekniklerinin kullanılmasını ve eyleme geçirilebilir sonuçlarda objektif olmasını sağlar.] Many data science projects fail because they don't start with the right question. [Birçok veri bilimi projesi, doğru soruyla başlamadıkları için başarısız olur.] By following these questions guidelines, you can ensure your data science projects don't share that same thing. [Bu soru yönergelerini izleyerek veri bilimi projelerinizin aynı şeyi paylaşmamasını sağlayabilirsiniz.] In the next video, we'll talk about how even the most powerful hypothesis can answer all data science questions. [Bir sonraki videoda, en güçlü hipotezin bile tüm veri bilimi sorularına nasıl cevap verebileceğinden bahsedeceğiz.]

# **Challenges with Solving Real-World Problems with Hypothesis Testing**

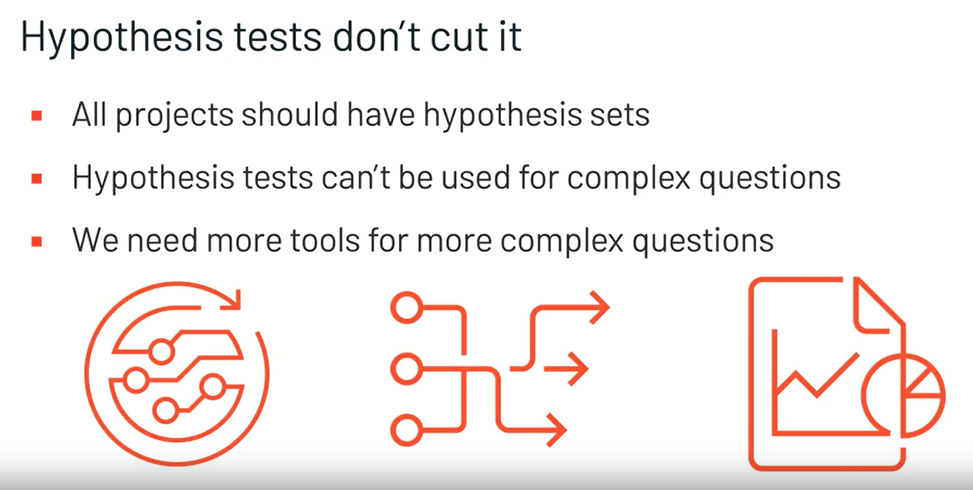
Hello again, in our last video we discussed why it's important to frame real-world scenarios so that they can be easily answered with data. [Tekrar merhaba, son videomuzda, verilerle kolayca yanıtlanabilmeleri için gerçek dünya senaryolarını çerçevelemenin neden önemli olduğunu tartıştık.] In this video, we're going to talk about some of the challenges associated with answering these types of research questions using hypothesis tests. [Bu videoda, bu tür araştırma sorularını hipotez testleri kullanarak yanıtlamayla ilgili bazı zorluklardan bahsedeceğiz.]



Think back to the questions that we asked in our previous video. [>> Bir önceki videomuzda sorduğumuz soruları tekrar düşünün.] Do we have more out of stock items in rural or urban communities than we do in suburban communities? [Kırsal veya kentsel topluluklarda, banliyö topluluklarında olduğundan daha fazla stokta olmayan ürünümüz var mı?] And does this new medical treatment reduce the average recovery time for a patient? [Ve bu yeni tıbbi tedavi, bir hasta için ortalama iyileşme süresini kısaltıyor mu?] What do these questions have in common? [Bu soruların ortak noktası nedir?] Well, we know they're both good because they're focused, objectively answerable, and actionable, but they have a few other similarities too. [Her ikisinin de iyi olduğunu biliyoruz çünkü odaklanmışlar, nesnel olarak cevap verebilirler ve eyleme geçirilebilirler, ancak başka benzerlikleri de var.] Both of these questions are yes or no questions. [Bu soruların her ikisi de evet veya hayır sorularıdır.] The answer to each of these questions is either going to be yes, or it's going to be no. [Bu soruların her birinin cevabı ya evet olacak ya da hayır olacak.] While we can provide various explanations, the answers can vary a whole lot. [Çeşitli açıklamalar sunabilsek de cevaplar çok farklı olabilir.] And this is partially what makes them objectively answerable using data and tools like hypothesis tests. [Ve bu kısmen, hipotez testleri gibi verileri ve araçları kullanarak onları nesnel olarak yanıtlanabilir kılan şeydir.] These questions are also focused on understanding simple truths about observed data. [Bu sorular ayrıca gözlemlenen verilerle ilgili basit gerçekleri anlamaya odaklanır.] They are asking about simple aggregations like counts and averages for various samples of data. [Çeşitli veri örnekleri için sayımlar ve ortalamalar gibi basit toplamalar hakkında sorular soruyorlar.] And in the case of hypothesis testing, they're using those aggregations and statistical inference to make claims about a population. [Ve hipotez testi durumunda, bir popülasyon hakkında iddialarda bulunmak için bu toplamaları ve istatistiksel çıkarımları kullanıyorlar.] These similarities are what make these questions suitable for hypothesis testing. [Bu benzerlikler, bu soruları hipotez testi için uygun kılan şeydir.] But this is also pretty limiting. [Ama bu aynı zamanda oldukça sınırlayıcı.]



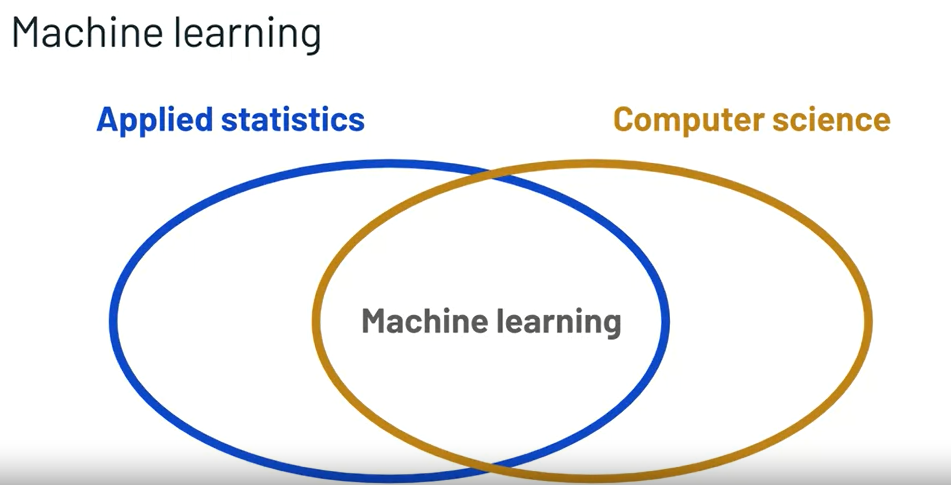
What if we have a different data science question? [Ya farklı bir veri bilimi sorumuz varsa?] What if we want to know something more, or something else? [Ya daha fazlasını veya başka bir şeyi bilmek istiyorsak?] For example, what if we want to estimate when a specific item is going to be out of stock in a specific store? [Örneğin, belirli bir ürünün belirli bir mağazada ne zaman stokta olmayacağını tahmin etmek istersek ne olur?] Or what if we want to understand what causes items to be out of stock in the first place? [Ya da ilk etapta ürünlerin stokta kalmamasına neyin sebep olduğunu anlamak istersek?] In the medical example, we might want to estimate the exact probability that a new medical treatment will affect the recovery time for a specific patient. [Tıbbi örnekte, yeni bir tıbbi tedavinin belirli bir hasta için iyileşme süresini etkileme olasılığını tam olarak tahmin etmek isteyebiliriz.] These are complex questions, but they're still good questions. [Bunlar karmaşık sorular, ancak yine de iyi sorular.] They're still specific, they're still objectively answerable, and they're still actionable. [Hâlâ spesifikler, hâlâ nesnel olarak yanıtlanabilirler ve hâlâ eyleme geçirilebilirler.] They're also incredibly similar to countless data science projects that you'll come across in the real world. [Ayrıca gerçek dünyada karşılaşacağınız sayısız veri bilimi projesine inanılmaz derecede benziyorlar.] These questions are focused less on the what of the data, and more of the when, the how, how much, and why. [Bu sorular, verilerin ne olduğuna daha az odaklanır ve daha çok ne zaman, nasıl, ne kadar ve neden sorularına odaklanır.] Questions that data scientists and key stakeholders ask all the time to learn more about their specific domains and industries. [Veri bilimcilerin ve kilit paydaşların kendi özel alanları ve endüstrileri hakkında daha fazla bilgi edinmek için her zaman sordukları sorular.]



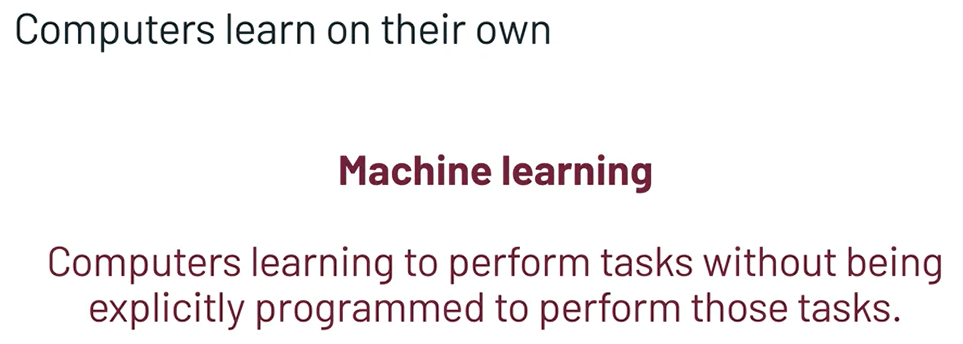
We can and should set hypotheses for these types of projects. [Bu tür projeler için hipotezler kurabiliriz ve koymalıyız.] We can hypothesize that items go out of stock at regular intervals. [Öğelerin düzenli aralıklarla stoktan çıktığını varsayabiliriz.] We can hypothesize that items go out of stock on specific days of the week or months of the year. [Ürünlerin haftanın belirli günlerinde veya yılın aylarında stokta kalmadığını varsayabiliriz.] But hypotheses can't help us with these more complex data science projects. [Ancak hipotezler, bu daha karmaşık veri bilimi projelerinde bize yardımcı olamaz.] How would we use a hypothesis test to estimate when a certain item is going to be out of stock? [Belirli bir ürünün ne zaman stokta kalmayacağını tahmin etmek için bir hipotez testini nasıl kullanırız?] There's no realistic mechanism for that, especially when the scenario doesn't appropriately match with one of the probability distributions we learned about. [Bunun için gerçekçi bir mekanizma yoktur, özellikle de senaryo, öğrendiğimiz olasılık dağılımlarından biriyle uygun şekilde eşleşmediğinde.] The same can be said for understanding why items go out of stock, or estimating the exact probability that a new medical treatment will affect the recovery time of a specific patient. [Aynı şey, ürünlerin neden stokta kalmadığını anlamak veya yeni bir tıbbi tedavinin belirli bir hastanın iyileşme süresini etkileme olasılığını tam olarak tahmin etmek için de söylenebilir.] For these questions, hypothesis tests just aren't enough. [Bu sorular için hipotez testleri yeterli değildir.] We need more technical tools in our data science tool belt. [Veri bilimi alet kemerimizde daha fazla teknik araca ihtiyacımız var.] Throughout the rest of this lesson, we'll start to look into the basic concepts of machine learning. [Bu dersin geri kalanında, makine öğreniminin temel kavramlarını incelemeye başlayacağız.] And learn how to frame these more complex data science questions. [Ve bu daha karmaşık veri bilimi sorularını nasıl çerçeveleyeceğinizi öğrenin.]

# **Introduction to Machine Learning, Part 1**

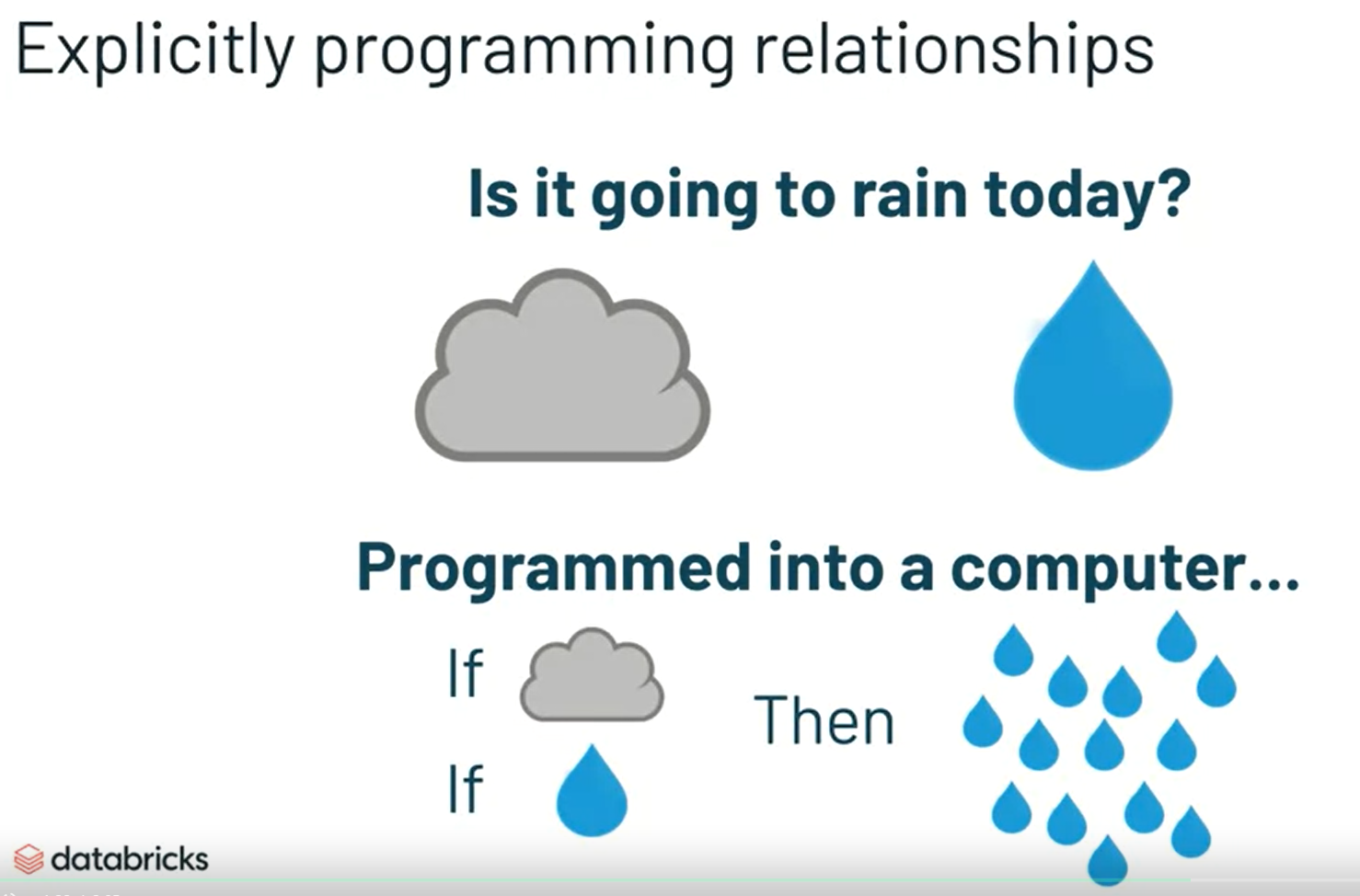
In the previous video, we described how the technical tools we've learned so far, like hypothesis tests, can be limiting for more complex data science questions. [Önceki videoda, hipotez testleri gibi şimdiye kadar öğrendiğimiz teknik araçların daha karmaşık veri bilimi soruları için nasıl sınırlayıcı olabileceğini açıkladık.] In this video, we'll describe the field of machine learning. [Bu videoda makine öğrenimi alanını anlatacağız.]



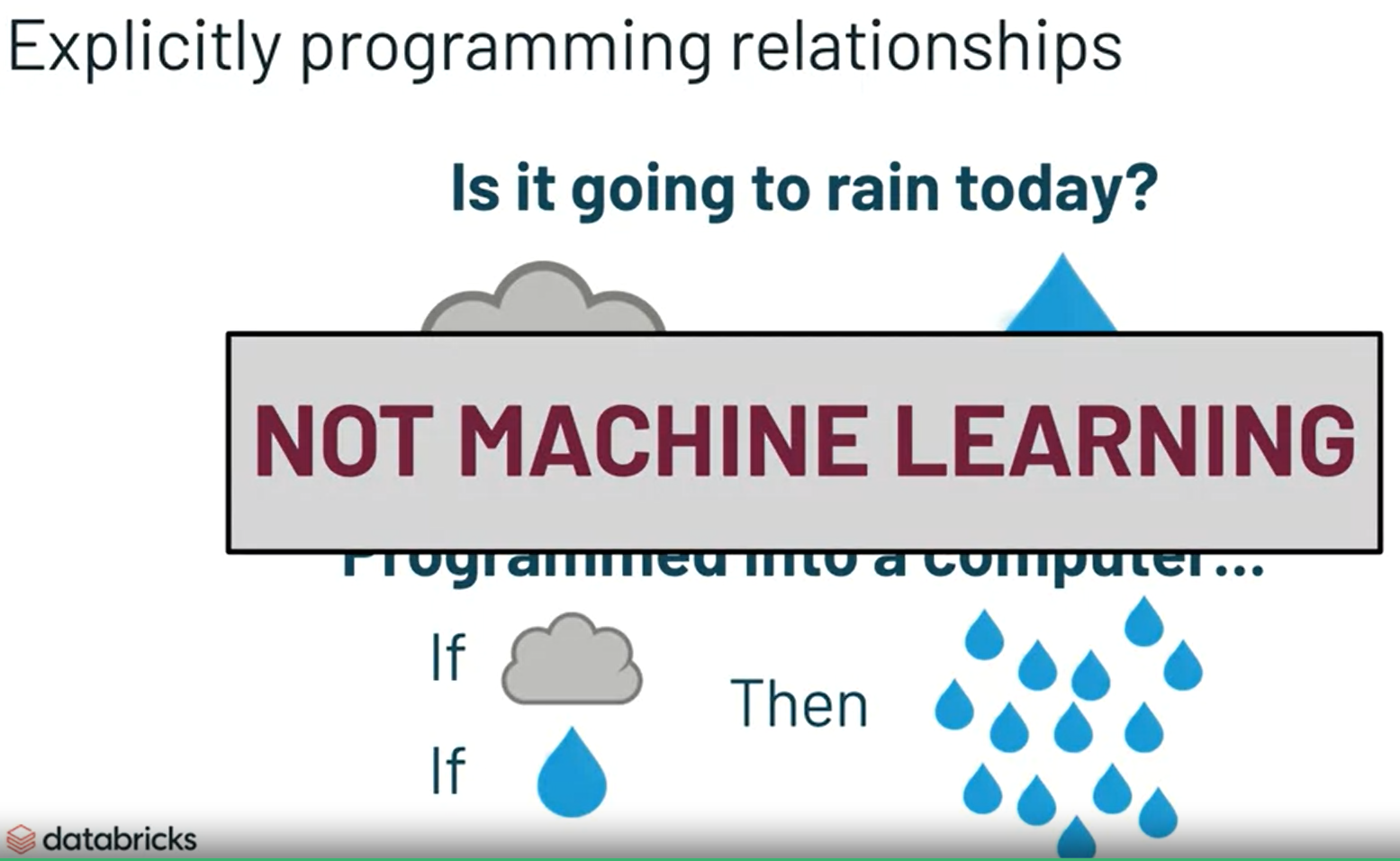
If you remember back to the first module we talked about how machine learning is at the intersection of applied statistics and computer science. [İlk modülü hatırlarsanız, makine öğreniminin uygulamalı istatistik ve bilgisayar biliminin kesişim noktasında nasıl olduğundan bahsetmiştik.] And that's exactly right, but there's a little bit more to it than that. [Ve bu kesinlikle doğru, ama bundan biraz daha fazlası var.]



Machine learning is centered around the idea of computers learning to perform tasks without being explicitly programmed to perform those tasks. [Makine öğrenimi, bilgisayarların bu görevleri gerçekleştirmek için açıkça programlanmadan görevleri gerçekleştirmeyi öğrenmesi fikri etrafında toplanmıştır.]



Before we go through an example of machine learning in the next video, let's take a look here at what isn't machine learning. [Bir sonraki videoda bir makine öğrenimi örneğine geçmeden önce, burada makine öğreniminin ne olmadığına bir göz atalım.] Let's say that we wake up in the morning and we need to determine whether or not it's going to rain throughout the day. [Diyelim ki sabah uyandık ve gün boyunca yağmur yağıp yağmayacağını belirlememiz gerekiyor.] What might we look at to make that prediction? [Bu öngörüyü yapmak için neye bakabiliriz?] I would probably look at the sky. [Muhtemelen gökyüzüne bakardım.] If the sky is gray, I assume it's more likely to rain. [Gökyüzü griyse, yağmur yağma olasılığının daha yüksek olduğunu varsayıyorum.] I also might check out the humidity levels, I know that humidity can bring rain as well. [Ayrıca nem seviyelerine de bakabilirim, nemin yağmur da getirebileceğini biliyorum.] On the other hand, if it's clear and sunny with dry air, I know that the likelihood of rain throughout the day is probably going to be relatively low. [Öte yandan, hava açık ve güneşliyse, kuru hava varsa, gün boyunca yağmur yağma olasılığının muhtemelen nispeten düşük olacağını biliyorum.] Now we could take these learnings and this knowledge that we have and we could explicitly program all of these decision points into a computer. [Şimdi bu öğrenmeleri ve sahip olduğumuz bu bilgiyi alabilir ve tüm bu karar noktalarını bir bilgisayara açıkça programlayabiliriz.] We could have the computer take a picture of the sky, we could have the computer measure the humidity. [Bilgisayarın gökyüzünün resmini çekmesini sağlayabiliriz, bilgisayarın nemi ölçmesini sağlayabiliriz.] And we could program it to use that information to tell us if it's going to rain if it finds the sky to be gray and the humidity to be high. [Ve onu, gökyüzünü gri ve nemi yüksek bulursa, yağmur yağacağını bize söylemek için bu bilgiyi kullanmak üzere programlayabiliriz.] While this might work, and it's an automated process, it's not machine learning. [Bu işe yarayabilir ve otomatikleştirilmiş bir süreç olsa da, makine öğrenimi değildir.]

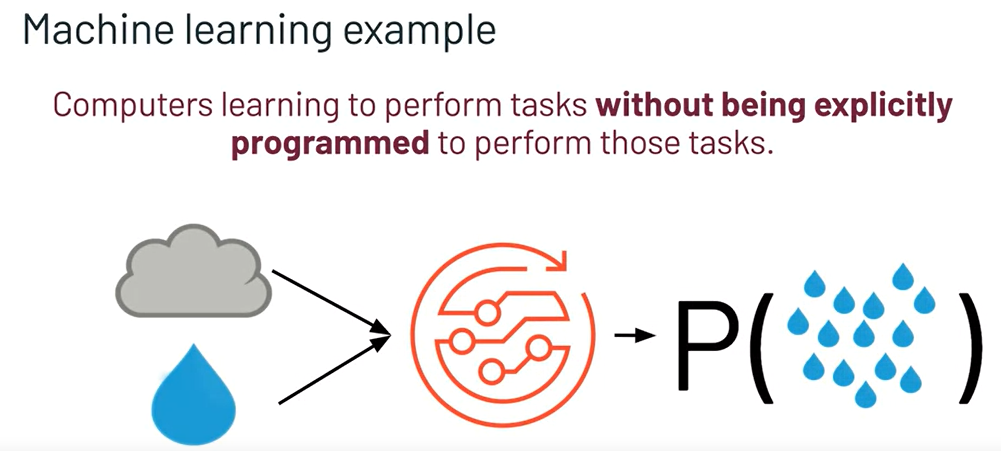


It's not machine learning because we're directly telling the computer how to use information, like the cloud coverage and the humidity level, to predict whether or not it's going to rain. [Bu makine öğrenimi değil çünkü bilgisayara bulut kapsamı ve nem seviyesi gibi bilgileri nasıl kullanacağını, yağmurun yağıp yağmayacağını tahmin etmesi için doğrudan söylüyoruz.] In the next video, we'll look at an example of how this task could be

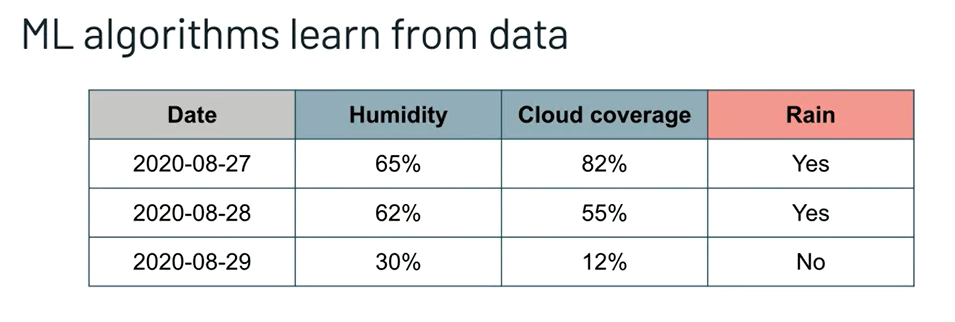
accomplished with machine learning. [Bir sonraki videoda, bu görevin makine öğrenimi ile nasıl gerçekleştirilebileceğine dair bir örneğe bakacağız.]

# **Introduction to Machine Learning, Part 2**

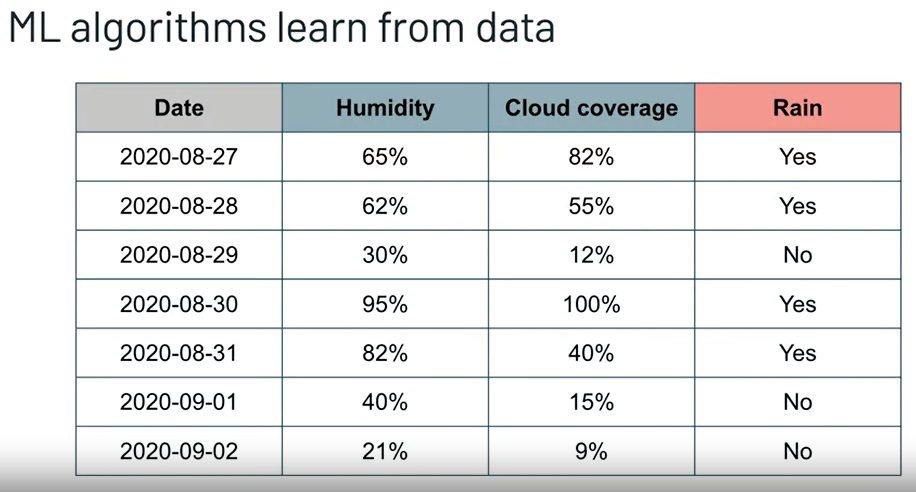
We just began our introductory look at machine learning. [Makine öğrenimine giriş niteliğindeki incelememize yeni başladık.] In this video, we'll continue with our weather example to describe how machine learning can be used to complete this task. [Bu videoda, bu görevi tamamlamak için makine öğreniminin nasıl kullanılabileceğini açıklamak için hava durumu örneğimizle devam edeceğiz.]



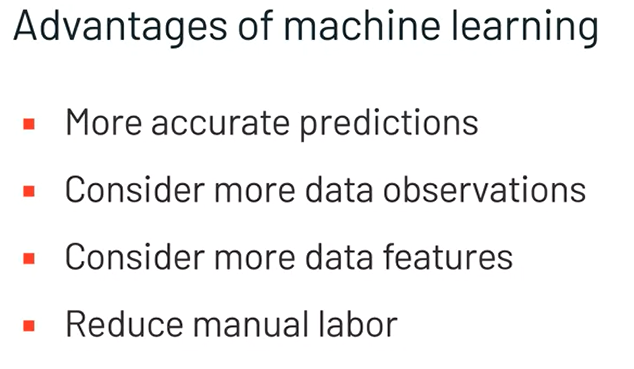
Remember from our last video that machine learning is computers learning to perform tasks without being explicitly programmed to do so. [Son videomuzdan hatırlayalım ki, makine öğrenimi, açıkça programlanmadan görevleri gerçekleştirmeyi öğrenen bilgisayarlardır.] You might be thinking, but how can that happen? [Düşünüyor olabilirsiniz, ama bu nasıl olabilir?] Don't we need to tell the computer something? [Bilgisayara bir şey söylememiz gerekmiyor mu?] I think you're right, we do need to direct computers a bit. [Sanırım haklısın, bilgisayarları biraz yönlendirmemiz gerekiyor.] Let's stick with our example where we're trying to predict whether or not it's going to rain. [Yağmur yağıp yağmayacağını tahmin etmeye çalıştığımız örneğimize bağlı kalalım.] We've been considering cloud coverage and the humidity level as predictors or what we call features of whether or not it's going to rain. [Bulut kapsamını ve nem seviyesini tahmin ediciler olarak veya yağmur yağıp yağmayacağına dair özellikler dediğimiz özellikleri dikkate aldık.] If we wanted to use machine learning to solve this, we can still use those features, but we want to let the computer itself determine the relationship between cloud coverage and humidity level and whether or not it's going to rain. [Bunu çözmek için makine öğrenimini kullanmak isteseydik, yine de bu özellikleri kullanabiliriz, ancak bulut kapsamı ve nem seviyesi arasındaki ilişkiyi ve yağmur yağıp yağmayacağını bilgisayarın kendisinin belirlemesine izin vermek istiyoruz.] Instead of us explicitly programming it's going to rain if the sky is gray and the humidity is high, the computer will use a machine learning algorithm to learn those relationships itself and then predict the chance of rain. [Gökyüzü griyse ve nem yüksekse yağmur yağacağını açıkça programlamak yerine, bilgisayar bu ilişkileri öğrenmek için bir makine öğrenme algoritması kullanacak ve ardından yağmur olasılığını tahmin edecek.] This is machine learning. [Bu makine öğrenmesidir.] But how does it work? [Ama nasıl çalışıyor?] How can computers and algorithms learn these relationships? [Bilgisayarlar ve algoritmalar bu ilişkileri nasıl öğrenebilir?] Well, it's all centered around data and usually a lot of data. [Şey, her şey veri etrafında toplanmış ve genellikle çok fazla veri var.]



In order for our machine learning algorithm to learn how cloud coverage and humidity are related to whether or not it rains, we need to provide the algorithm with this information from past days. [Makine öğrenimi algoritmamızın bulut kapsamı ve nemin yağmur yağıp yağmamasıyla nasıl ilişkili olduğunu öğrenebilmesi için algoritmaya geçmiş günlerden bu bilgileri sağlamamız gerekiyor.] This is something that looks like a SQL table or a data frame. [Bu, SQL tablosuna veya veri çerçevesine benzeyen bir şeydir.] There are many different machine learning algorithms that will run those relationships in different ways, but they're all going to use this data. [Bu ilişkileri farklı şekillerde yürütecek birçok farklı makine öğrenimi algoritması var, ancak hepsi bu verileri kullanacak.] This is what we call training data. [Buna eğitim verileri diyoruz.] Machine learning algorithms use training data to learn those relationships and in result, learn to predict whether or not it's going to rain based on the humidity level and the cloud coverage. [Makine öğrenimi algoritmaları, bu ilişkileri öğrenmek için eğitim verilerini kullanır ve sonuç olarak, nem düzeyine ve bulut kapsamına bağlı olarak yağmur yağıp yağmayacağını tahmin etmeyi öğrenir.] This is where the applied statistics comes into machine learning. [Uygulamalı istatistiklerin makine öğrenimine girdiği yer burasıdır.] These algorithms are usually rooted in the mathematical concepts that are prevalent in applied statistics. [Bu algoritmalar genellikle uygulamalı istatistiklerde yaygın olan matematiksel kavramlara dayanır.]



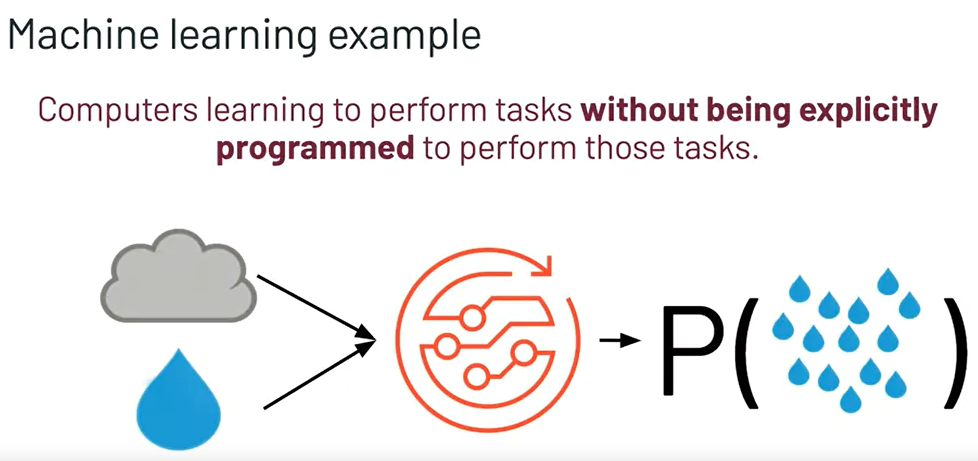
Just as with hypothesis tests and applied statistics, it's usually best to have as much data as possible when doing machine learning. [Hipotez testleri ve uygulamalı istatistiklerde olduğu gibi, makine öğrenimi yaparken genellikle mümkün olduğunca fazla veriye sahip olmak en iyisidir.] This gives the machine learning algorithm more information from which to learn. [Bu, makine öğrenimi algoritmasına öğrenebileceği daha fazla bilgi verir.] More days of whether to observe, to learn to predict whether or not it's going to rain. [Yağmur yağıp yağmayacağını tahmin etmeyi öğrenmek, gözlemlemek için daha fazla gün.] This is similar to what humans call wisdom. [Bu, insanların bilgelik dediği şeye benzer.] The more experience somebody has with something, the more wisdom they have, the more they're likely going to know about it. [Birisi bir şeyle ilgili ne kadar çok deneyime sahipse, o kadar fazla bilgeliğe sahipse, muhtemelen o şey hakkında daha fazla şey bilecektir.] As data gets larger and larger, think billions or even trillions of rows and hundreds of thousands of columns, which again, we call features. [Veri büyüdükçe ve büyüdükçe, milyarlarca hatta trilyonlarca satır ve yüz binlerce sütun düşünün ki buna yine özellik diyoruz.] The algorithms need to be implemented in more computationally efficient ways. [Algoritmaların hesaplama açısından daha verimli yollarla uygulanması gerekir.] This, along with the programming to direct this process, is the computer science of machine learning. [Bu, bu süreci yönlendirmek için programlama ile birlikte, makine öğreniminin bilgisayar bilimidir.]



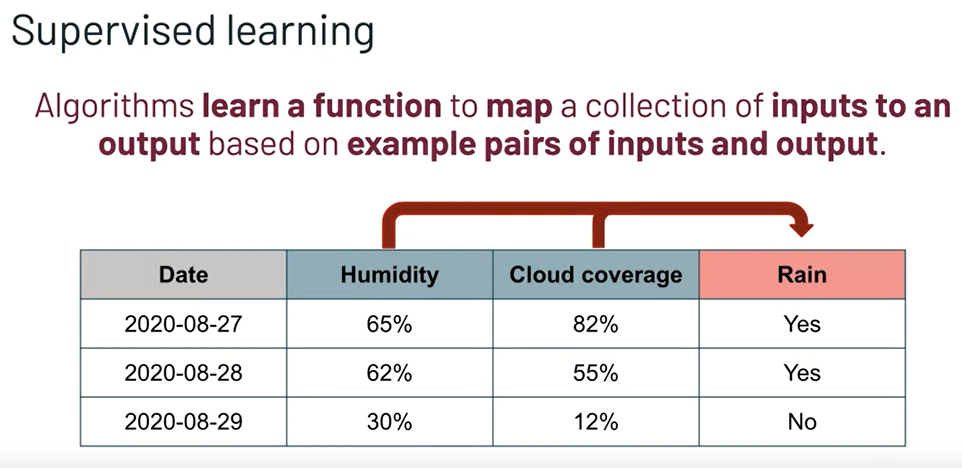
What are the advantages of using machine learning to determine whether or not it's going to rain? [Yağmur yağıp yağmayacağını belirlemek için makine öğrenimini kullanmanın avantajları nelerdir?] You might be thinking, I know whether or not it's going to rain in the morning, why do I need something else to tell me? [Sabah yağmur yağıp yağmayacağını biliyorum, neden başka bir şey söylemem gerekiyor diye düşünebilirsiniz.] That's a good question. [Bu iyi bir soru.] As humans, we can reasonably tell whether or not it's going to rain most of the time, but not all of the time. [İnsanlar olarak, çoğu zaman yağmur yağıp yağmayacağını makul bir şekilde söyleyebiliriz, ancak her zaman değil.] It's reasonable that the machine learning algorithm will be better at predicting this than us when we give it enough information to learn from. [Öğrenecek kadar bilgi verdiğimizde, makine öğrenme algoritmasının bunu tahmin etmede bizden daha iyi olması mantıklı.] So we'll likely get more accurate predictions. [Böylece muhtemelen daha doğru tahminler alacağız.] This is really driven by the algorithm's ability to learn from all of the different days of data. [Bu, gerçekten, algoritmanın tüm farklı veri günlerinden öğrenme yeteneği tarafından yönlendirilir.] It can record each individual day of weather, learn from each of them and combine the learnings from each of them to create a robust prediction model. [Her bir hava gününü kaydedebilir, her birinden öğrenebilir ve sağlam bir tahmin modeli oluşturmak için her birinden öğrendiklerini birleştirebilir.] We, as humans, we're going to have trouble comprehending all of those learnings ourselves. [Biz insanlar olarak, tüm bu öğrendikleri kendimiz anlamakta güçlük çekeceğiz.] We're going to have trouble remembering each day and even if we remember them subconsciously, we're not going to remember them as well as a computer will. [Her gün hatırlamakta güçlük çekeceğiz ve onları bilinçaltında hatırlasak bile bir bilgisayarın hatırlayacağı kadar iyi hatırlamayacağız.] We might recall certain notable days, but we won't be able to value each individual day and that's a key point about the value of using machine learning algorithms. [Bazı önemli günleri hatırlayabiliriz, ancak her bir günü ayrı ayrı değerlendiremeyeceğiz ve bu, makine öğrenimi algoritmalarını kullanmanın değeri hakkında önemli bir nokta.] In addition, we can provide the algorithm additional features or inputs the model we'll use to make predictions like whether or not it rained the day before, the dew-point, the types of clouds that are in the sky, the temperature, the season, and so on. [Ayrıca, algoritmaya, önceki gün yağmur yağıp yağmadığı, çiy noktası, gökyüzündeki bulut türleri, sıcaklık, hava sıcaklığı gibi tahminler yapmak için kullanacağımız modele ek özellikler veya girdiler sağlayabiliriz. sezon falan.] This additional information is really helpful to machine learning algorithms because it provides more information to make predictions. [Bu ek bilgi, tahmin yapmak için daha fazla bilgi sağladığı için makine öğrenimi algoritmaları için gerçekten yararlıdır.] But as this number features grows, it can be hard for human minds to comprehend all of this in real time and use it to make a prediction themselves. [Ancak bu sayı özellikleri büyüdükçe, insan zihninin tüm bunları gerçek zamanlı olarak kavraması ve bunu kendi başına bir tahminde bulunmak için kullanması zor olabilir.] Finally, using machine learning can increase the accuracy of our predictions but it can also reduce the labor required of humans. [Son olarak, makine öğrenimini kullanmak tahminlerimizin doğruluğunu artırabilir, ancak aynı zamanda insanlar için gereken emeği de azaltabilir.] Think about it this way. [Bu şekilde düşün.] Do you look outside in the morning every day to determine whether or not it's going to rain? [Her sabah yağmur yağıp yağmayacağını anlamak için dışarı bakar mısınız?] I know that I don't. [Yapmadığımı biliyorum.] Even if you do, you likely don't spend much time making your prediction. [Yapsanız bile, tahmininizi yapmak için fazla zaman harcamazsınız.] You might look at a weather app, receive a notification, see something on the news, or hear about it on the radio. [Bir hava durumu uygulamasına bakabilir, bir bildirim alabilir, haberlerde bir şey görebilir veya radyoda duyabilirsiniz.] Instead, we rely on machine learning algorithms developed by experts to predict the probability of rain. [Bunun yerine, yağmur olasılığını tahmin etmek için uzmanlar tarafından geliştirilen makine öğrenimi algoritmalarına güveniyoruz.] This is a major advantage that applies in nearly every scenario in which machine learning algorithms are applied. [Bu, makine öğrenimi algoritmalarının uygulandığı hemen hemen her senaryoda geçerli olan büyük bir avantajdır.] Remember, machine learning is the process of algorithms rooted in applied statistics and computer science, learning from the data provided to them. [Unutmayın, makine öğrenimi, uygulamalı istatistik ve bilgisayar bilimine dayanan ve kendilerine sağlanan verilerden öğrenen algoritmaların sürecidir.] This can open up a whole new set of real-world problems that can be solved using data science. [Bu, veri bilimi kullanılarak çözülebilecek yepyeni bir dizi gerçek dünya sorununun kapısını aralayabilir.] In the next video, we'll start to categorize these problems into different classes of machine learning. [Bir sonraki videoda, bu sorunları farklı makine öğrenimi sınıflarına ayırmaya başlayacağız.] Stay with us. [Bizimle kal.]

# **Supervised and Unsupervised Learning**

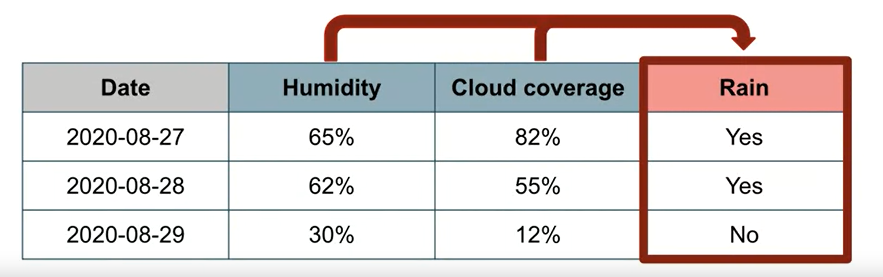
Hi. [Merhaba.] In the previous video, we defined and described machine learning with an example related to predicting whether or not it's going to rain on a given day. [Bir önceki videoda, belirli bir günde yağmur yağıp yağmayacağını tahmin etmekle ilgili bir örnekle makine öğrenmesini tanımladık ve anlattık.] In this video, we'll stick with that example to define supervised learning and unsupervised learning. [Bu videoda, denetimli öğrenmeyi ve denetimsiz öğrenmeyi tanımlamak için bu örneğe bağlı kalacağız.]



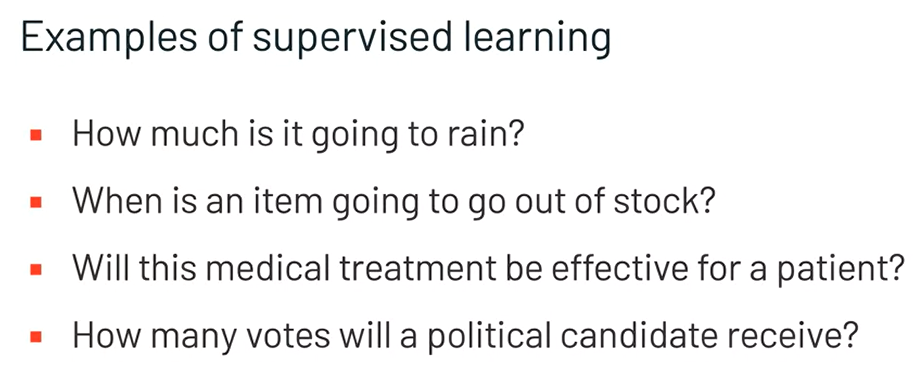
Recall from the previous video that machine learning is centered around the idea of computers learning to perform tasks without being explicitly programmed to do so. [Önceki videodan, makine öğreniminin, bilgisayarların açıkça programlanmadan görevleri gerçekleştirmeyi öğrenmesi fikri etrafında toplandığını hatırlayın.] We were talking about using machine learning to predict whether or not it's going to rain. [Yağmurun yağıp yağmayacağını tahmin etmek için makine öğrenimini kullanmaktan bahsediyorduk.] This is an example of what we call supervised machine learning. [Bu, denetimli makine öğrenimi dediğimiz şeyin bir örneğidir.] A very common subset of machine learning that's used in practice pretty frequently. [Pratikte oldukça sık kullanılan, çok yaygın bir makine öğrenimi alt kümesi.]



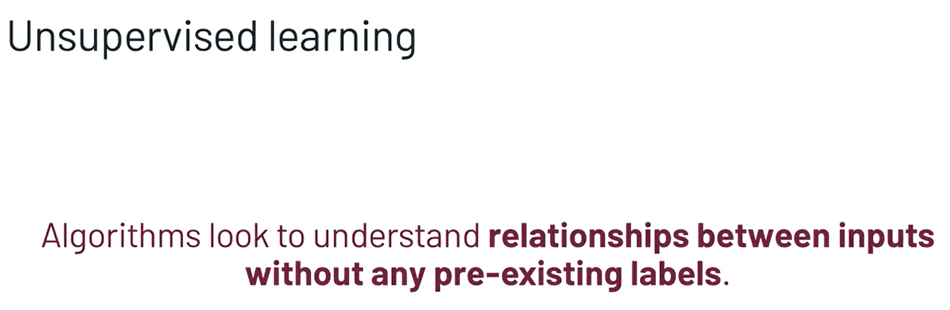
Supervised machine learning, which is commonly referred to as just simply supervised learning, is when an algorithm is learning a function then maps a collection of inputs to an output based on example pairs of inputs and output, which we refer to as our training data. [Genellikle yalnızca denetimli öğrenme olarak adlandırılan denetimli makine öğrenimi, bir algoritmanın bir işlevi öğrenmesi ve ardından eğitim verilerimiz olarak adlandırdığımız örnek girdi ve çıktı çiftlerine dayalı olarak bir girdi koleksiyonunu bir çıktıya eşlemesidir.] In a way, our algorithm is being supervised by certain truth values for the function output. [Bir bakıma algoritmamız, fonksiyon çıktısı için belirli doğruluk değerleri tarafından denetleniyor.] In this case, our rain problem is a supervised learning problem because we're trying to use an algorithm to map inputs of cloud coverage and humidity levels to the probability that it's going to rain. [Bu durumda, yağmur problemimiz denetimli bir öğrenme problemidir çünkü bulut kapsamı ve nem seviyelerinin girişlerini yağmur yağma olasılığına eşlemek için bir algoritma kullanmaya çalışıyoruz.] Our supervising truth is whether or not it rained on each day within our training data. [Denetleyici gerçeğimiz, eğitim verilerimiz dahilinde her gün yağmur yağıp yağmadığıdır.]



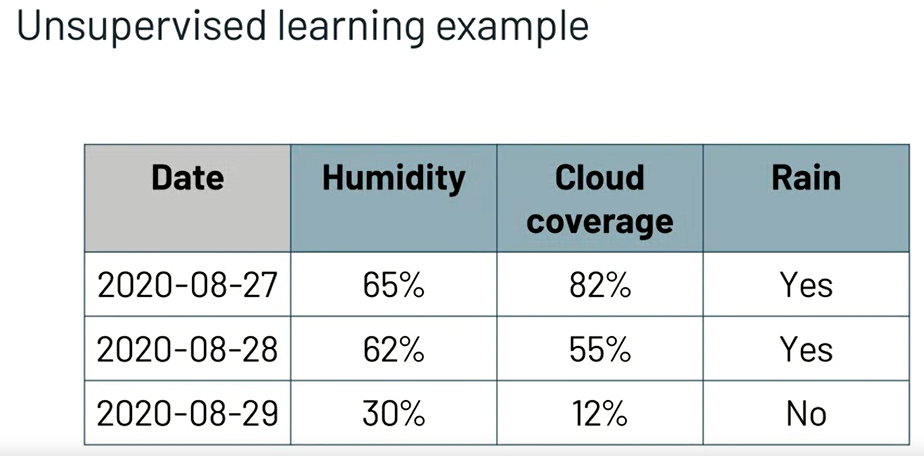
These truth values, also known as labels, are the key behind supervised learning. [Etiketler olarak da bilinen bu doğruluk değerleri, denetimli öğrenmenin ardındaki anahtardır.] We cannot do supervised learning without these supervising labels, because they're supervising our machine learning algorithm. [Bu denetleyici etiketler olmadan denetimli öğrenme yapamayız, çünkü bunlar makine öğrenimi algoritmamızı denetlerler.]



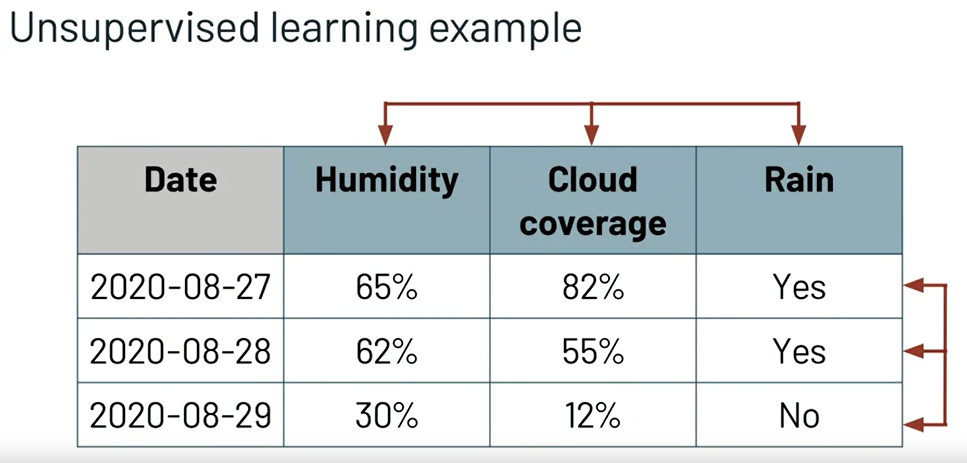
There are plenty of other examples of supervised learning. [Denetimli öğrenmenin başka birçok örneği vardır.] We could try to predict how much it's going to rain. [Ne kadar yağmur yağacağını tahmin etmeye çalışabiliriz.] This is different than the probability of it raining at all, but it still contains labels to supervise the algorithms learning process. [Bu, yağmur yağma olasılığından farklıdır, ancak yine de algoritmaların öğrenme sürecini denetlemek için etiketler içerir.] Those labels are the amount that it rained on previous days instead of just whether or not it did rain. [Bu etiketler, sadece yağmur yağıp yağmadığı yerine, önceki günlerde yağan yağmur miktarıdır.] Thinking back to our supply chain example, predicting when a specific item might go out of stock is also a supervised learning problem. [Tedarik zinciri örneğimize geri dönersek, belirli bir ürünün ne zaman stokta kalmayacağını tahmin etmek de denetimli bir öğrenme sorunudur.] When items have gone out of stock in the past, could serve as supervising labels. [Öğeler geçmişte stokta bittiğinde, denetleyici etiketler olarak kullanılabilir.] Predicting whether or not a medical treatment will be effective for a particular patient is another example, and estimating how many votes a political candidate might get in an election year is yet another. [Bir tıbbi tedavinin belirli bir hasta için etkili olup olmayacağını tahmin etmek başka bir örnektir ve bir siyasi adayın bir seçim yılında kaç oy alabileceğini tahmin etmek başka bir örnektir.] Overall, it's a good rule of thumb that if the word's estimate or predict are involved in the data science questionnaire design. [Genel olarak, kelimenin tahmini veya tahmininin veri bilimi anket tasarımına dahil olması iyi bir kuraldır.] It's a supervised learning problem and task. [Bu denetimli bir öğrenme problemi ve görevidir.] You might be wondering, what if we don't have labels to supervise our learning? [Merak ediyor olabilirsiniz, ya öğrenmemizi denetleyecek etiketlerimiz yoksa?] What if we aren't trying to estimate or predict something as part of our question or design? [Sorumuzun veya tasarımımızın bir parçası olarak bir şeyi tahmin etmeye veya tahmin etmeye çalışmıyorsak ne olur?]



The opposite of supervised machine learning is unsupervised machine learning. [Denetimli makine öğreniminin tersi, denetimsiz makine öğrenimidir.] Many data science students wonder, how do we do machine learning if we don't have any labels on which to train our algorithm? [Birçok veri bilimi öğrencisi, algoritmamızı eğitecek herhangi bir etiketimiz yoksa makine öğrenimini nasıl yapacağımızı merak ediyor.] That's a good question. [Bu iyi bir soru.] Unsupervised learning is a completely different problem set. [Denetimsiz öğrenme tamamen farklı bir problem setidir.] It's a type of machine learning that looks to understand relationships between inputs without any preexisting labels. [Bu, önceden var olan herhangi bir etiket olmaksızın girdiler arasındaki ilişkileri anlamaya çalışan bir tür makine öğrenimidir.] So unsupervised algorithms are not being supervised by some label, they're not trying to predict anything and they're not trying to estimate anything. [Yani denetimsiz algoritmalar bir etiket tarafından denetlenmiyor, hiçbir şeyi tahmin etmeye çalışmıyorlar ve hiçbir şeyi tahmin etmeye çalışmıyorlar.] Instead, they're trying to understand how features, or columns, and records, or rows are related to one another. [Bunun yerine, özelliklerin, sütunların, kayıtların veya satırların birbirleriyle nasıl ilişkili olduğunu anlamaya çalışıyorlar.]



What would an example of unsupervised learning look like? [Denetimsiz öğrenme örneği neye benzer?] Let's stick with our weather domain. [Hava durumu alanımıza bağlı kalalım.] Instead of trying to predict whether or not it's going to rain, we might try to group days together if they have similar weather. [Yağmur yağıp yağmayacağını tahmin etmeye çalışmak yerine, benzer hava koşullarına sahip günleri birlikte gruplandırmaya çalışabiliriz.] We can use an unsupervised learning algorithm to group these days together based on their features. [Bu günleri özelliklerine göre gruplandırmak için denetimsiz bir öğrenme algoritması kullanabiliriz.] Cloud coverage, humidity level, and whether or not it rained, which we can now use as a feature rather than labels in our unsupervised learning problem. [Denetimsiz öğrenme problemimizde artık etiketler yerine bir özellik olarak kullanabileceğimiz bulut kapsamı, nem seviyesi ve yağmur yağıp yağmadığı.] Days that are clear and dry might be grouped together. [Açık ve kuru günler birlikte gruplandırılabilir.] Days that are cloudy and dry might be grouped together, and days that are cloudy and rainy might be grouped together. [Bulutlu ve kuru günler birlikte gruplandırılabilir ve bulutlu ve yağmurlu günler birlikte gruplandırılabilir.] This can be useful because we can look at the days in each group and then label those groups ourselves based on their average of weather. [Bu yararlı olabilir, çünkü her gruptaki günlere bakabilir ve ardından bu grupları ortalama hava durumuna göre kendimiz etiketleyebiliriz.]



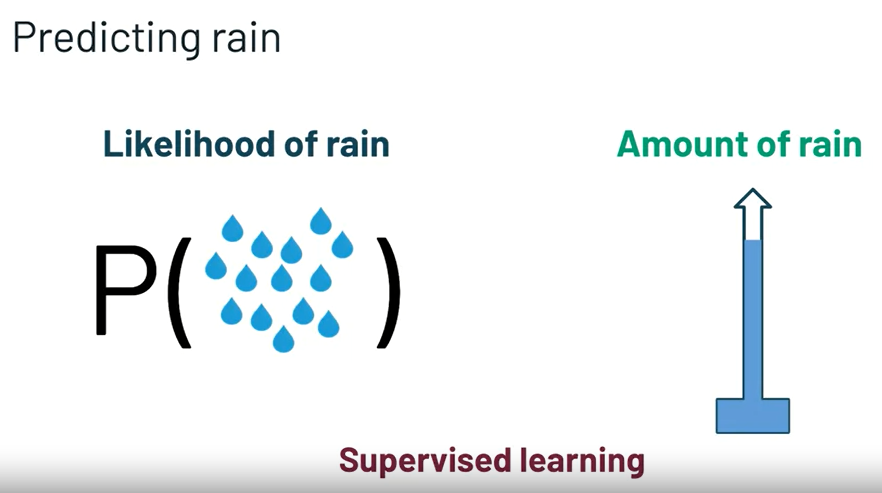
Some days might be sunny, some days might be overcast, and some days might be raining, or there might be some other type of weather pattern based on the patterns identified in our unsupervised learning process. [Bazı günler güneşli olabilir, bazı günler bulutlu ve bazı günler yağmurlu olabilir veya denetimsiz öğrenme sürecimizde tanımlanan kalıplara dayalı olarak başka türde bir hava durumu kalıbı olabilir.]



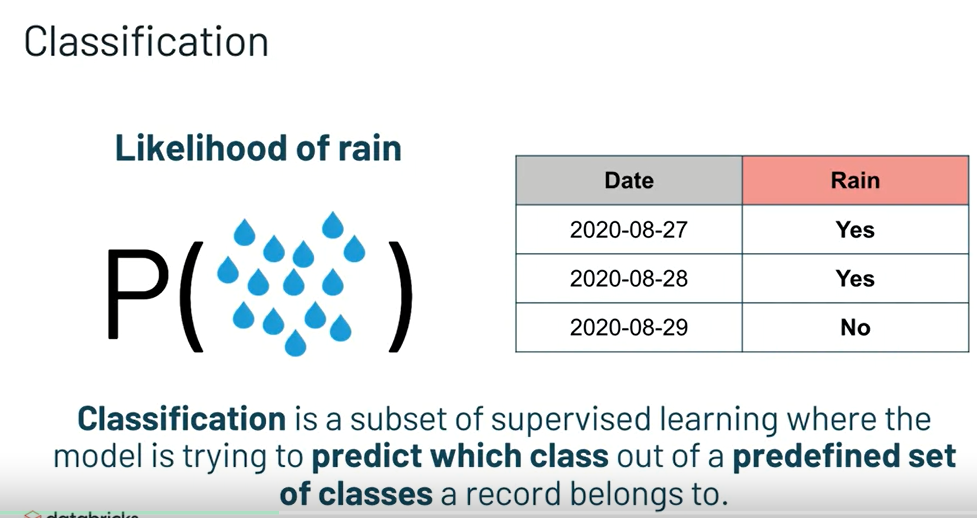
Other examples of unsupervised learning in the real world, group customers together based on their purchasing habits, group patients together based on their health conditions, and identify transactions that might be fraudulent because they're so dissimilar from other transactions. [Gerçek dünyada denetimsiz öğrenmenin diğer örnekleri, müşterileri satın alma alışkanlıklarına göre gruplandırmak, hastaları sağlık koşullarına göre gruplandırmak ve diğer işlemlerden çok farklı oldukları için sahte olabilecek işlemleri belirlemek.] Unsupervised learning can be especially helpful when data labels aren't available or we're trying to learn more about our data and its relationships. [Denetimsiz öğrenme, özellikle veri etiketleri bulunmadığında veya verilerimiz ve ilişkileri hakkında daha fazla bilgi edinmeye çalıştığımızda yardımcı olabilir.] You will probably see this a lot during the exploratory data analysis phase of a data science project. [Bir veri bilimi projesinin keşifsel veri analizi aşamasında muhtemelen bunu çok göreceksiniz.] Being able to frame real-world problems into supervised learning and unsupervised learning problems opens up a world of possibilities for applying machine learning to data science projects. [Gerçek dünyadaki sorunları denetimli öğrenme ve denetimsiz öğrenme sorunları olarak çerçeveleyebilmek, makine öğrenimini veri bilimi projelerine uygulamak için bir fırsatlar dünyası açar.] Instead of just testing aggregated data with hypothesis tests, we can now predict, estimate, group, and find relationships using our data. [Birleştirilmiş verileri yalnızca hipotez testleri ile test etmek yerine, artık verilerimizi kullanarak ilişkileri tahmin edebilir, tahmin edebilir, gruplayabilir ve bulabiliriz.] Throughout the rest of this lesson, we'll dig down deeper into supervised learning and unsupervised learning before we move on to learning how to apply them in the next module. [Bu dersin geri kalanında, bir sonraki modülde bunların nasıl uygulanacağını öğrenmeye geçmeden önce, denetimli öğrenme ve denetimsiz öğrenme konularına derinlemesine ineceğiz.] But first, we have a brief knowledge check for you to make sure you are grasping the key points so far. [Ama önce, şu ana kadar kilit noktaları kavradığınızdan emin olmak için kısa bir bilgi kontrolümüz var.]

# **Regression and Classification**

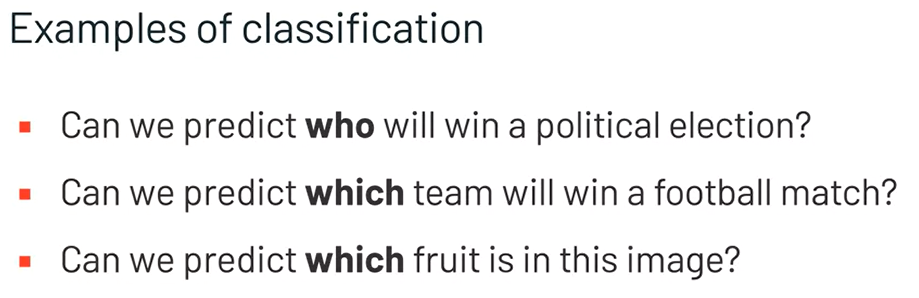
To this point, we've talked about how machine learning problems can be divided into two primary categories, supervised learning and unsupervised learning. [Bu noktaya kadar, makine öğrenimi problemlerinin denetimli öğrenme ve denetimsiz öğrenme olmak üzere iki ana kategoriye nasıl ayrılabileceğinden bahsettik.] We described how framing real world problems in this way can open up a world of possibilities in terms of what can be solved using a data science project. [Gerçek dünya problemlerini bu şekilde çerçevelemenin, bir veri bilimi projesi kullanılarak çözülebilecek şeyler açısından nasıl bir olasılıklar dünyası açabileceğini anlattık.] In this video, we're going to breakdown supervised learning a bit further into regression and classification. [Bu videoda, denetimli öğrenmeyi biraz daha ayrıntılı olarak regresyon ve sınıflandırmaya ayıracağız.]



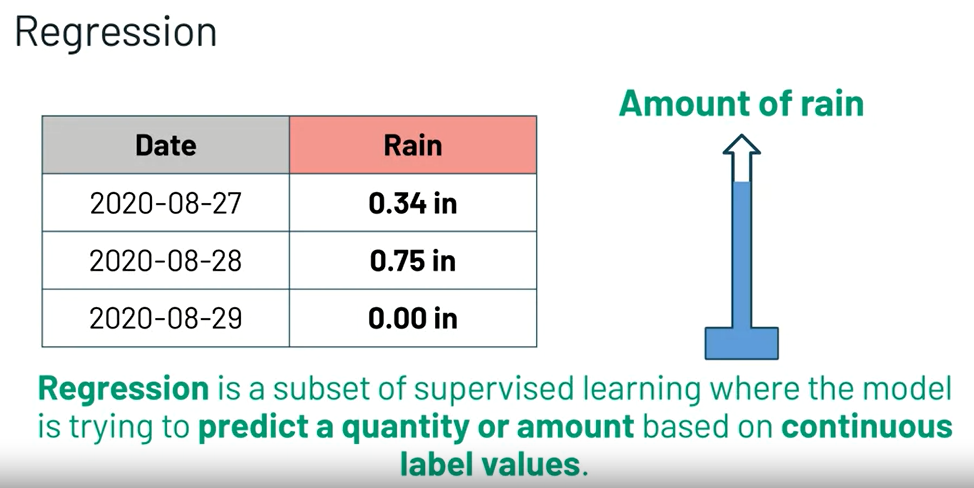
If you remember back a few videos, we talked about 2 examples of predicting things related to weather. [Birkaç videoyu hatırlarsanız, hava durumuyla ilgili şeyleri tahmin etmenin 2 örneğinden bahsettik.] We talked about predicting the likelihood that it was going to rain and predicting the amount that it's going to rain. [Yağmur yağma olasılığını tahmin etmekten ve yağacak miktarı tahmin etmekten bahsettik.] Now these are both examples of supervised learning as we learned in the last video. [Şimdi bunların ikisi de son videoda öğrendiğimiz gibi denetimli öğrenme örnekleridir.] They're both using truthful labels to learn to map input data to some type of output prediction. [Her ikisi de girdi verilerini bir tür çıktı tahminiyle eşleştirmeyi öğrenmek için doğru etiketleri kullanıyor.]



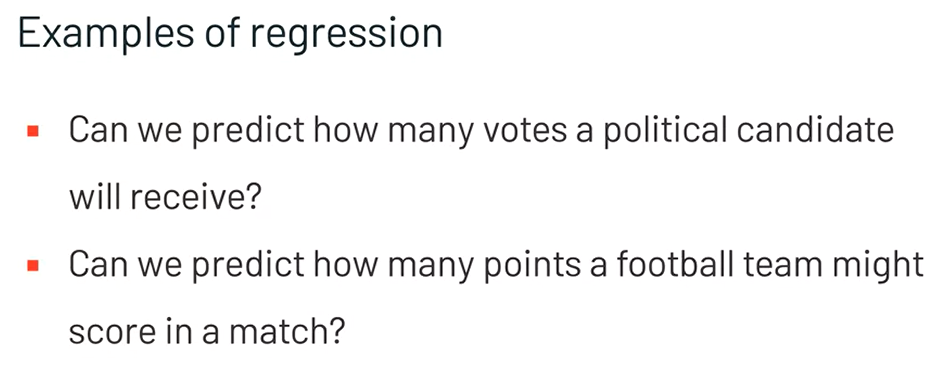
But there are labels differ from one another. [Ama birbirinden farklı etiketler var.] On the one hand, predicting whether or not it rains is focused on predicting a discrete class. [Bir yandan, yağmur yağıp yağmayacağını tahmin etmek, ayrı bir sınıfı tahmin etmeye odaklanır.] It either rains or it doesn't rain. [Ya yağmur yağar ya da yağmaz.] These two classes can be thought of like mutually exclusive categories. [Bu iki sınıf birbirini dışlayan kategoriler gibi düşünülebilir.] The process is known as classification. [İşlem sınıflandırma olarak bilinir.] Classification is a subset of supervised learning, where the model is trying to predict which class out of a predefined set of classes, a record belongs to. [Sınıflandırma, modelin önceden tanımlanmış bir sınıf kümesinden hangi sınıfa ait olduğunu tahmin etmeye çalıştığı denetimli öğrenmenin bir alt kümesidir.] Because classification models output predictions, they frequently output a probability their record belongs in one of or each of the predefined classes. [Sınıflandırma modelleri tahminler ürettiği için, sıklıkla kayıtlarının önceden tanımlanmış sınıflardan birine veya her birine ait olma olasılığını verirler.] It's up to the data scientists then to translate these predictions into actual classes. [O zaman bu tahminleri gerçek sınıflara çevirmek veri bilimcilerine kalmış.] So taking the probability that it's going to rain into either the class, yes, it's going to rain or the class, no, it's not going to rain. [Yani ya sınıfa yağmur yağma olasılığını alırsak, evet, yağmur yağacak ya da sınıf, hayır, yağmur yağmayacak.]



Other examples of classification include predicting who's going to win a political election, predicting whether or not a football team will win a match, an predicting which fruit might be an image. [Diğer sınıflandırma örnekleri arasında siyasi bir seçimi kimin kazanacağını tahmin etme, bir futbol takımının bir maçı kazanıp kazanmayacağını tahmin etme, hangi meyvenin bir görüntü olabileceğini tahmin etme sayılabilir.] Classification problems are really useful in the data science process because they have concrete outcomes that are very actionable. [Sınıflandırma problemleri, veri bilimi sürecinde gerçekten faydalıdır çünkü çok uygulanabilir somut sonuçları vardır.] The probability that it's going to rain is something that's easily interpretable, and we can easily shift our behavior based on that probability. [Yağmur yağma olasılığı kolayca yorumlanabilen bir şeydir ve bu olasılığa dayalı olarak davranışlarımızı kolayca değiştirebiliriz.] We can bring a rain jacket we can bring in umbrella, or we might not bring a rain jacket or an umbrella. [Şemsiye getirebileceğimiz bir yağmurluk getirebiliriz ya da bir yağmurluk ya da şemsiye getirmeyebiliriz.]



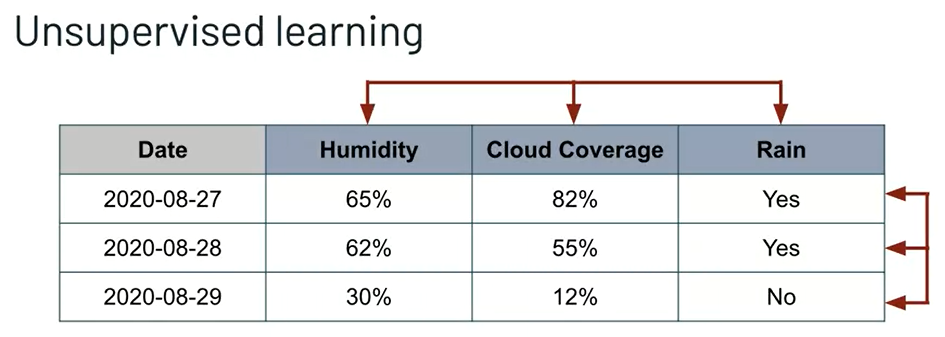
On the other hand, predicting how much it's going to rain this focused on predicting a continuous value. [Öte yandan, bunun ne kadar yağacağını tahmin etmek, sürekli bir değer tahmin etmeye odaklandı.] It could rain any amount greater than 0 inches, even if some of those amounts are incredibly unlikely. [Bu miktarların bazıları inanılmaz derecede düşük olsa bile, 0 inçten fazla herhangi bir miktarda yağmur yağabilir.] This continuous output is commonly associated with estimating a quantity or an amount of something. [Bu sürekli çıktı, genellikle bir miktarın veya bir şeyin miktarının tahmin edilmesiyle ilişkilendirilir.] This process is known as regression. [Bu süreç regresyon olarak bilinir.] Regression is a subset of supervised learning where the model is trying to predict the quantity or amount based on the continuous label values. [Regresyon, modelin sürekli etiket değerlerine dayalı olarak miktarı veya miktarı tahmin etmeye çalıştığı denetimli öğrenmenin bir alt kümesidir.]



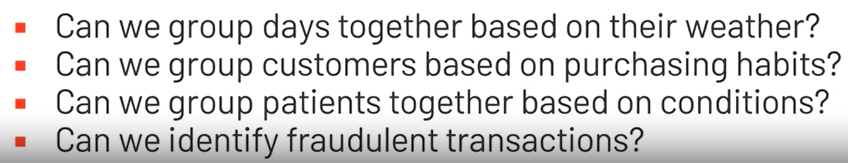
Other examples of regression include predicting how many votes a particular candidate might receive in a political election, and predicting how many points of football team might score in a match instead of whether or not they're going to win. [Diğer regresyon örnekleri, belirli bir adayın siyasi bir seçimde kaç oy alabileceğini tahmin etmeyi ve bir futbol takımının kazanıp kazanamayacaklarını değil, bir maçta kaç puan alabileceğini tahmin etmeyi içerir.] Regression problems are also practically relevant in many organizations. [Regresyon sorunları da birçok organizasyonda pratik olarak geçerlidir.] Across different industries, predicting continuous values can open up actionable applications of machine learning and data science that wouldn't otherwise be possible. [Farklı endüstrilerde, sürekli değerleri tahmin etmek, aksi halde mümkün olmayacak şekilde makine öğrenimi ve veri biliminin eyleme geçirilebilir uygulamalarını açabilir.] As an example, understanding how much it's going to rain can lead to things like taking preventative flood measures, or even in extreme cases leading to evacuations. [Örnek olarak, ne kadar yağmur yağacağını anlamak, önleyici sel önlemleri almak gibi şeylere ve hatta aşırı durumlarda tahliyelere yol açabilir.] Classification and regression are both forms of supervised learning, but they have different applications. [Sınıflandırma ve regresyon, her ikisi de denetimli öğrenmenin biçimleridir, ancak farklı uygulamaları vardır.] Hopefully you found the examples in this video helpful in understanding how they can be used to insert different types of data science questions. [Umarız bu videodaki örnekleri, farklı türde veri bilimi sorularını eklemek için nasıl kullanılabileceğini anlamada yardımcı bulmuşsunuzdur.] In the next video we'll dive deeper into a common form of unsupervised learning, clustering. [Bir sonraki videoda, ortak bir denetimsiz öğrenme biçimi olan kümelemenin daha derinlerine ineceğiz.]

# **Clustering**

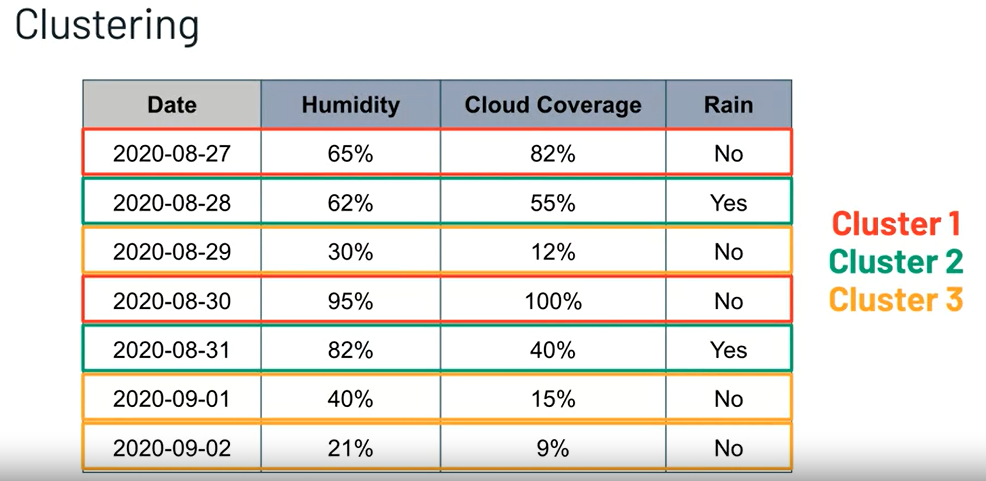
A few videos ago we talked about supervised and unsupervised machine learning. [Birkaç video önce denetimli ve denetimsiz makine öğrenmesinden bahsetmiştik.] And we described that supervised learning uses label values weather from a predefined set of classes or continuous values to map a function from a set of inputs to an output. [Ve denetimli öğrenmenin, bir işlevi bir girdi kümesinden bir çıktıya eşlemek için önceden tanımlanmış bir sınıf kümesinden hava durumu etiket değerlerini veya sürekli değerleri kullandığını tanımladık.] In this video, we'll look further into a common example of unsupervised learning known as clustering. [Bu videoda, kümeleme olarak bilinen denetimsiz öğrenmenin yaygın bir örneğine daha yakından bakacağız.]



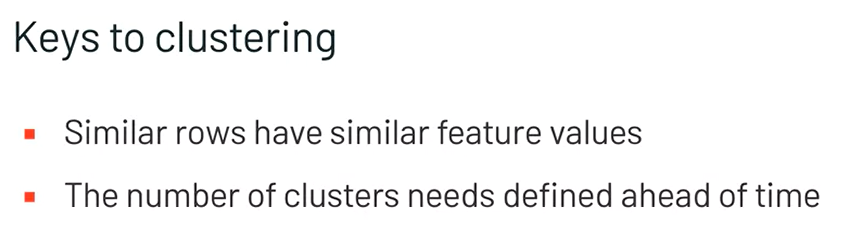
Recall that unsupervised learning looks to understand relationships between inputs without any pre existing labels. [Denetimsiz öğrenmenin, önceden var olan herhangi bir etiket olmaksızın girdiler arasındaki ilişkileri anlamaya çalıştığını hatırlayın.] This lack of labels is what makes it unsupervised. [Bu etiket eksikliği, onu denetimsiz yapan şeydir.]



We gave a few examples of unsupervised learning like grouping days together based on their weather. [Günlerin hava durumuna göre gruplandırılması gibi denetimsiz öğrenmeye birkaç örnek verdik.] Grouping customers together based on their purchasing habits, grouping patients together based on their health conditions. [Müşterileri satın alma alışkanlıklarına göre gruplandırmak, hastaları sağlık durumlarına göre gruplandırmak.] And identifying transactions than might be fraudulent because they're so dissimilar from other transactions. [Ve diğer işlemlerden çok farklı oldukları için sahte olabilecek işlemlerden daha fazlasını belirlemek.] The first three of these examples fall into a subset of unsupervised learning that we call clustering. [Bu örneklerin ilk üçü, kümeleme dediğimiz denetimsiz öğrenmenin bir alt kümesine girer.]



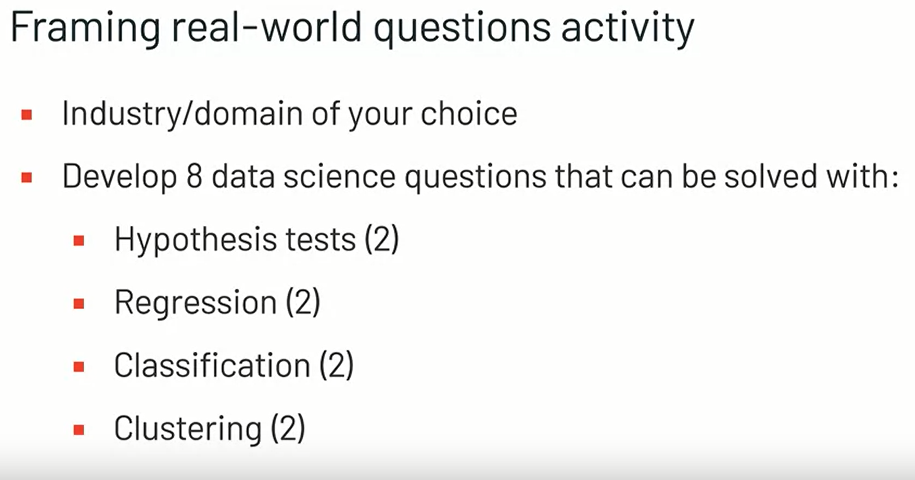
Clustering is an unsupervised learning approach that aims to group records together based on the similarity of their inputs or features. [Kümeleme, girdilerinin veya özelliklerinin benzerliğine dayalı olarak kayıtları birlikte gruplandırmayı amaçlayan denetimsiz bir öğrenme yaklaşımıdır.] Looking at this data here, we see that records with similar values are placed into the same cluster. [Buradaki verilere baktığımızda benzer değerlere sahip kayıtların aynı kümeye yerleştirildiğini görüyoruz.] This is what a clustering algorithm learns to do. [Bir kümeleme algoritmasının yapmayı öğrendiği şey budur.] In practice, clustering is a great way to learn more about the relationships within your data. [Pratikte, kümeleme, verileriniz içindeki ilişkiler hakkında daha fazla bilgi edinmenin harika bir yoludur.] It's also a common practice in making large data more actionable, like clustering millions of customers together into an actionable group. [Milyonlarca müşteriyi eyleme dönüştürülebilir bir grup halinde bir araya getirmek gibi, büyük verileri daha eyleme dönüştürülebilir hale getirmek için de yaygın bir uygulamadır.]



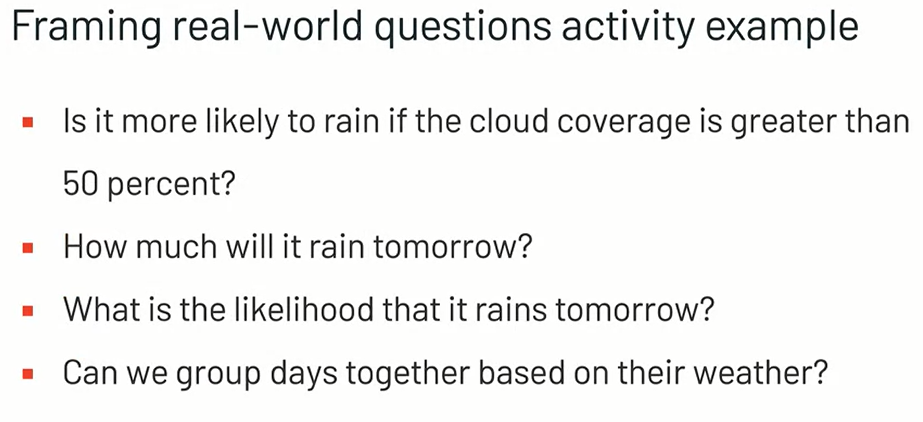
There are a variety of clustering algorithms, but most of them follow a couple of key principles that are very important. [Çeşitli kümeleme algoritmaları vardır, ancak çoğu çok önemli olan birkaç temel ilkeyi takip eder.] First, most clustering algorithms follow the assumption that similar records or observations will be close in proximity in the feature space. [İlk olarak, çoğu kümeleme algoritması, benzer kayıtların veya gözlemlerin özellik uzayında yakın olacağı varsayımını takip eder.] In other words, if we plotted all of the data records on a multi dimensional plot, points that are close in proximity will be more similar than points that are not close in proximity. [Diğer bir deyişle, tüm veri kayıtlarını çok boyutlu bir çizim üzerine çizersek, yakın olan noktalar yakın olmayan noktalara göre daha benzer olacaktır.] Then the points are placed in clusters and given a type of label to represent them being in that cluster. [Daha sonra noktalar kümelere yerleştirilir ve o kümede olduklarını göstermek için bir tür etiket verilir.] Another key principle of clustering is that the data scientist usually needs to define the number of clusters ahead of the learning process, and that can be challenging. [Kümelemenin diğer bir temel ilkesi, veri bilimcisinin genellikle öğrenme sürecinden önce küme sayısını tanımlaması gerektiğidir ve bu zor olabilir.] It's tough to know whether data naturally fits into two clusters, three clusters four clusters or even more. [Verilerin doğal olarak iki kümeye mi, üç kümeye dört kümeye mi yoksa daha fazla kümeye mi sığdığını bilmek zor.] As a result, it's common to perform multiple clustering analysis, each with a different number of clusters. [Sonuç olarak, her biri farklı sayıda kümeye sahip birden çok kümeleme analizi yapmak yaygındır.] Then the results of each analysis are evaluated to identify the optimal number of clusters. [Ardından, optimal küme sayısını belirlemek için her analizin sonuçları değerlendirilir.] It's also common to use domain knowledge at this point to help understand the possible numbers of clusters that the data might naturally fit into. [Bu noktada, verilerin doğal olarak sığabileceği olası küme sayılarını anlamaya yardımcı olması için alan bilgisinin kullanılması da yaygındır.] While there are other examples of unsupervised learning, clustering is the most broadly used in practice. [Denetimsiz öğrenmenin başka örnekleri olsa da, uygulamada en yaygın olarak kullanılan kümelemedir.] And is absolutely essential to understanding the basics of machine learning and data science. [Makine öğrenimi ve veri biliminin temellerini anlamak için kesinlikle gereklidir.] It can be incredibly useful and exploratory data analysis and when labels aren't available. [Etiketlerin mevcut olmadığı durumlarda inanılmaz derecede yararlı ve keşif amaçlı veri analizi olabilir.] This concludes our lecture videos on framing real world problems for data science and machine learning. [Bu, veri bilimi ve makine öğrenimi için gerçek dünya problemlerini çerçeveleme konusundaki ders videolarımızı sonlandırıyor.] Its vital to understand the limitations of hypothesis tests and how complex problems can be organized into different fields of machine learning. [Hipotez testlerinin sınırlarını ve karmaşık problemlerin makine öğreniminin farklı alanlarında nasıl organize edilebileceğini anlamak hayati önem taşır.] So in this lesson with the knowledge check and in activity where you develop real world questions from your own industry for different types of machine learning. [Bu derste, farklı makine öğrenimi türleri için kendi sektörünüzden gerçek dünya soruları geliştirdiğiniz bilgi kontrolü ve aktivitede.]

# **Framing Real-World Questions Activity**

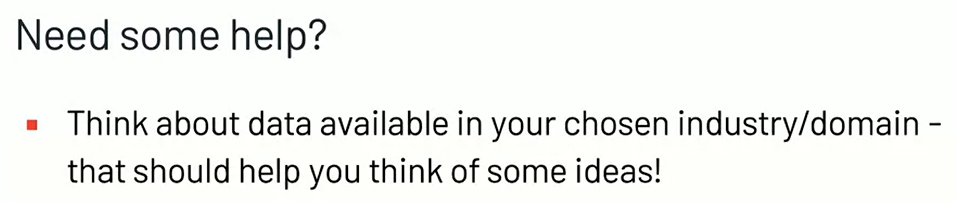
Hi. [Merhaba.] In this video, we'll describe the framing real-world questions activity that's coming up. [Bu videoda, yaklaşan gerçek dünya sorularını çerçeveleme etkinliğini anlatacağız.]



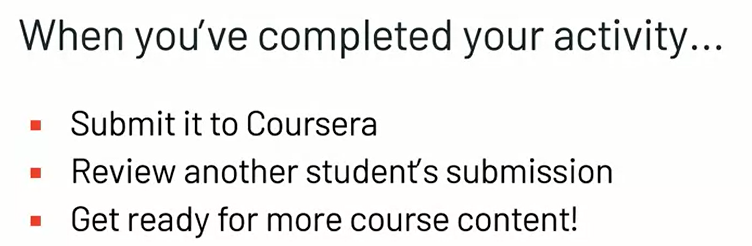
In this activity, you will be developing data science questions within an industry or domain of your choice. [Bu aktivitede, seçtiğiniz bir endüstri veya alan içinde veri bilimi soruları geliştireceksiniz.] You should develop two questions that can be solved with hypothesis tests, two questions that can be solved with regression, two problems that can be solved with classification, and two problems that can be solved with clustering. [Hipotez testleri ile çözülebilecek iki soru, regresyon ile çözülebilecek iki soru, sınıflandırma ile çözülebilecek iki soru ve kümeleme ile çözülebilecek iki soru geliştirmelisiniz.]



For example, if we stick with our weather examples, we might have the following questions for each category. [Örneğin, hava durumu örneklerimize bağlı kalırsak, her kategori için aşağıdaki sorularımız olabilir.] For hypothesis tests, is it more likely to rain if the cloud coverage is greater than 50 percent? [Hipotez testleri için, bulut kapsamı yüzde 50'den fazlaysa yağmur yağması daha olası mı?] For regression, how much will it rain tomorrow? [Regresyon için yarın ne kadar yağmur yağacak?] For classification, what is the likelihood that it rains at all tomorrow? [Sınıflandırma için, yarın yağmur yağma olasılığı nedir?] And for clustering, can we group days together based on their cloud coverage and humidity levels? [Ve kümeleme için günleri bulut kapsamına ve nem seviyelerine göre gruplandırabilir miyiz?]



If you're having trouble, think about what kind of data might be available in the industry or domain that you've chosen. [Sorun yaşıyorsanız, seçtiğiniz sektörde veya alanda ne tür verilerin bulunabileceğini düşünün.] That should help spur some ideas. [Bu, bazı fikirleri teşvik etmeye yardımcı olmalıdır.]



When you've completed your activity, submit it to Coursera, and then you'll have the opportunity to assess another student's submission. [Aktivitenizi tamamladığınızda onu Coursera'ya gönderin ve ardından başka bir öğrencinin gönderimini değerlendirme fırsatınız olacak.] A detailed [inaudible] will be available in the activity instructions. [Etkinlik talimatlarında ayrıntılı bir [duyulmuyor] bulunacaktır.] At that point, you will have finished the Connecting Data Science to the Real-World module. [Bu noktada, Veri Bilimini Gerçek Dünyaya Bağlama modülünü bitirmiş olacaksınız.] Be sure to join in the next module to learn more about the actual machine learning methodologies that can be used to answer some of these questions. [Bu sorulardan bazılarını yanıtlamak için kullanılabilecek gerçek makine öğrenimi metodolojileri hakkında daha fazla bilgi edinmek için bir sonraki modüle katıldığınızdan emin olun.]