Module 1: Course Orientation and Module 1: Analytics Mindset

Table of Contents

Lesson 1-0: Course Orientation

Lesson 1-0.1 Course Introduction

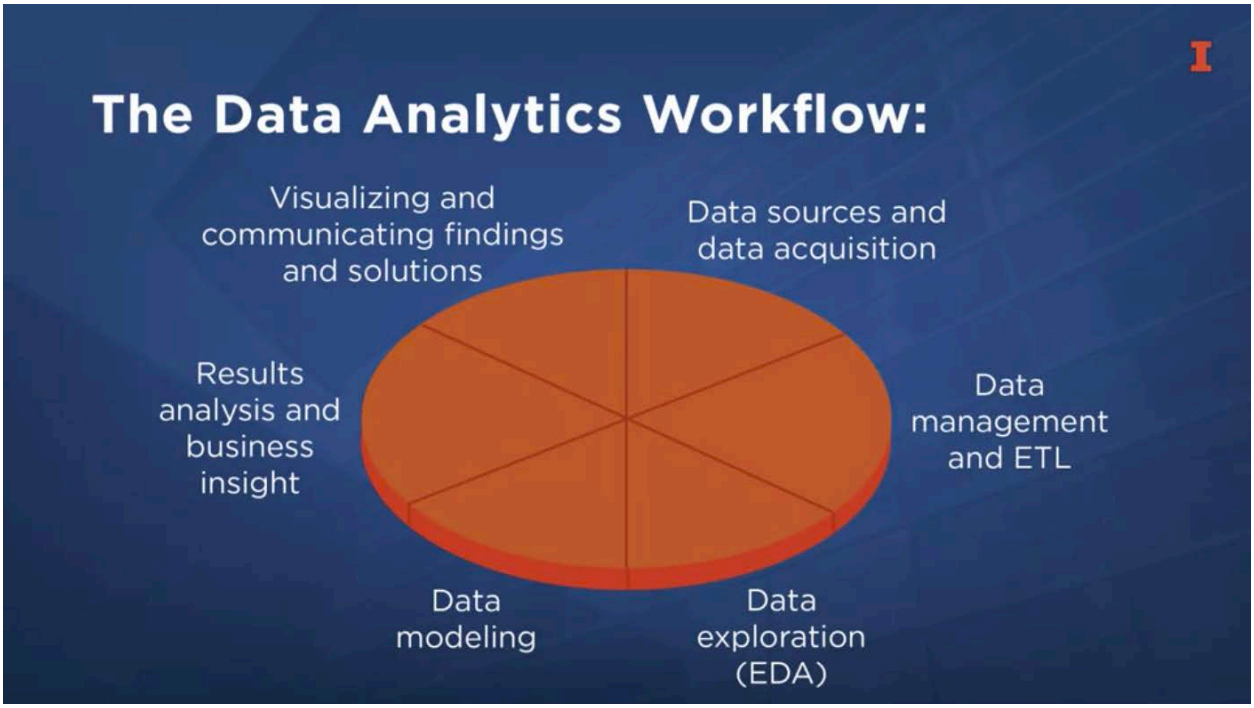
Prof. Jessen Hobson: This course is about giving you the tools you need to use data to gain actionable business insight.

Nearly everywhere we go and everything we do, from shopping online to typing a text, is infused with and enhanced by big data, machine learning and data analytics.

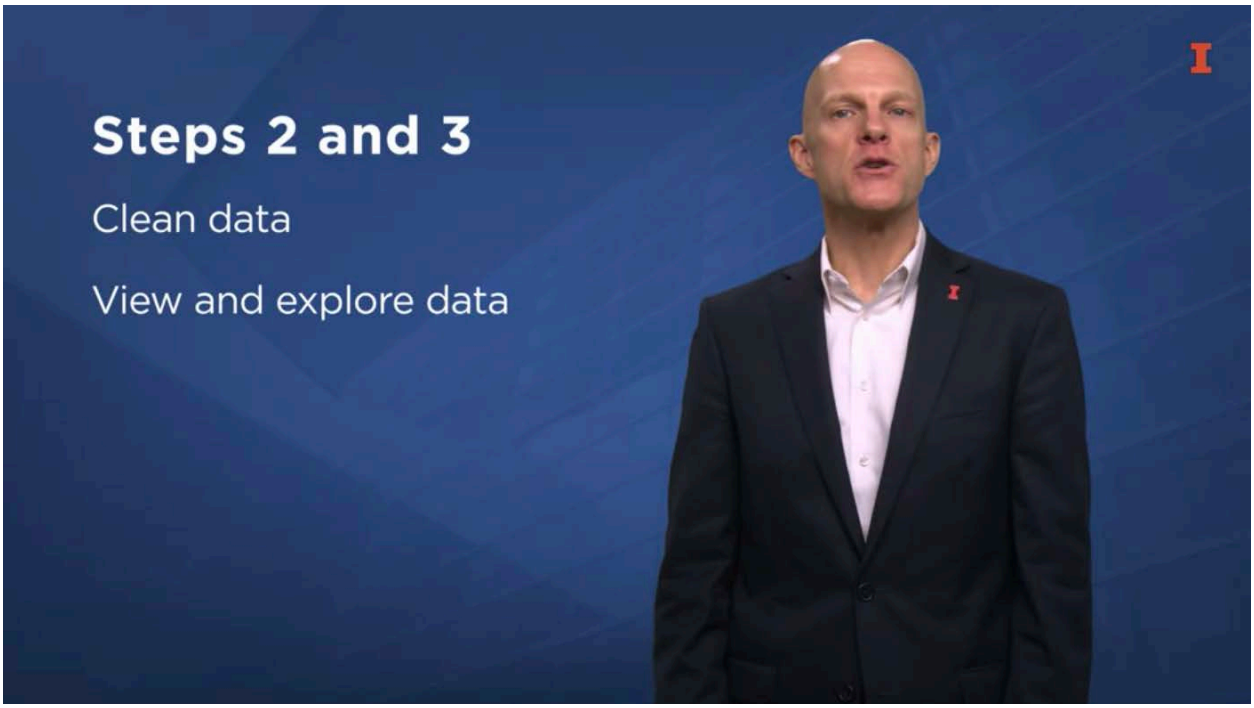
To win a business and to even be a successful participant, you and I need to learn how to master tools that help us take data and turn it into usable business insights.

Prof. Ronald Guymon: This course will give you four necessary and cutting edge tools to put you into the game. These tools are, one, your mind and how you process and think about data and data analytics. Two, power BI, a popular low coating business intelligence tool that is part of the core of Microsoft's business analytics tools. Three, R

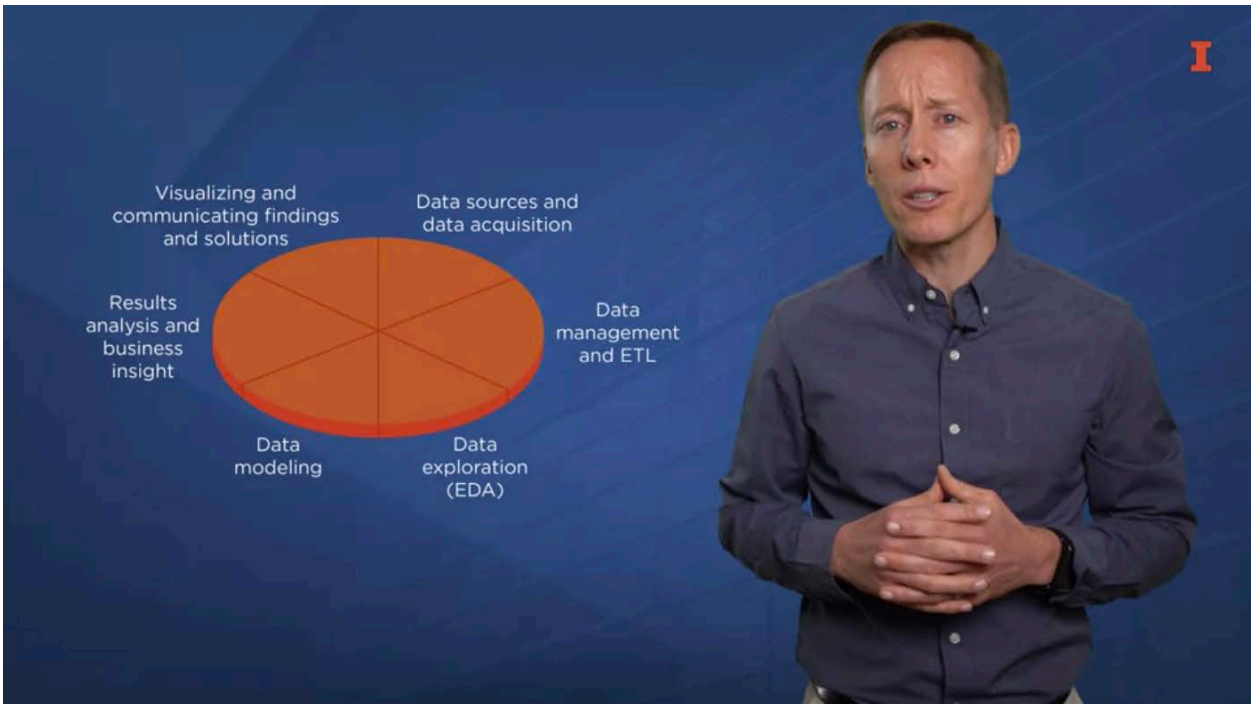
and RStudio, a popular coding language and code development environment that is relatively easy to use and was created to do analytics and statistical analyses. And four, Alteryx, another popular low coating business analytics and intelligence tool that focuses on automation and replicability.



In each module, we will use one of these tools to focus on the first half of the data analytics workflow. The data analytics workflow consists of the following parts, one, acquiring and maintaining data. Two, getting data ready for analysis. We often call this ETL, for extracting, transforming and loading data. Three, data exploration including the steps we take to understand and view our data. We often call this exploratory data analysis or EDA. Four, data modeling, including predicting future outcomes and inferring relationships from our data to other data and situations. Five, creating analysis, results and business insights. And six, visualizing and communicating findings and solutions.



Prof. Jessen Hobson: In particular, in this course will focus on steps two and three, and will use each tool in succession to learn and practice those steps. For example, imagine you've acquired some data and you want to view and explore it. The data is in some format, so you want to get it into your tool and clean it up, so that it's viewable and has as few mistakes and errors as possible. Then you want to explore it, so that you can start to see what stories it can tell, and how it can be useful to you in solving your business problems and creating business opportunities.



Prof. Guymon: Pictures are truly worth 1000 words. So, you will inevitably want to create some visualizations to help you and others interact with and understand the data. Our goal in this course is to provide you with a solid framework and foundation for how to understand and pursue business analytics. Thus in the future when you encounter a data analysis task that you have not encountered before, you will be able to slot it into that framework, understand it quickly, and move more rapidly through the process of assimilating the new information that you need to learn. More importantly, you will be able to put your new knowledge to work for you in gaining business insight and solving the business problem.

In each module we will use realistic business data to practice using these tools. Thus in each module, we will focus on solving business problems.

We are excited for you to expand your toolkit by learning more about these tools. Data cleaning, loading and exploration, are the foundation for all Data Analysis.

Opening a new data set and understanding it, is like unlocking a treasure chest and solving a puzzle.

It is like peeling an onion and learning layer by layer,

it is like opening a Matryoshka doll or nested doll and finding the truth at the center. As you go through these modules, dig in and play with the data. The more you practice, the more you expand your data tool kit and your ability to solve business problems.

Lesson 1-0.2 About Professor Jessen Hobson

Hi, my name is Jessen Hobson. I'm excited for you to take this class. I'd like to give you just a short biography, a little bit about my life, where I've come from, and what I've done. I grew up in Boise, Idaho, which is out in the West, now the Intermountain West in the Mountains. Went from there, and got my undergraduate education at Brigham Young University. I started there, and then took two-year hiatus to serve a mission for my church in Antofagasta, Chile. I came back to BYU and graduated with a bachelor's and a master's in accounting. Also, met my wonderful wife along the way. I went from there to Washington, D.C., and worked as an auditor for PricewaterhouseCoopers. But even then I knew I wanted to go and get a Ph.D., so I could teach. Shortly thereafter, I went to the University of Texas at Austin and got a Ph.D. in accounting. After graduating, I went and worked at Florida State University, and then came here to the University of Illinois. I've taught accounting audit and most recently, business analytics and data analytics. I'm really involved in data analytics and really enjoy it. Most recently, and really throughout my whole career, I've focused on data, and how I can teach and use data to solve business problems. I'm excited to be here, I'm excited that you're going to take this class. Be great.

Lesson 1-0.3 About Professor Ronald Guymon

Hi, my name is Ron Guymon and I'm on the faculty here in the School of Accountancy at the Gies College of Business. I have accounting degrees from Brigham Young University and the University of Iowa. My professional experience has been a mixture of academics and practice. In academics, my teaching and research has focused on management accounting, and data analytics. In practice, I've worked as a data scientist. There's a little bit about my professional experience. Let me tell you a little bit now about my personal life.

All right. Well, here I am. I'm at the top of a goblin here in Goblin Valley. I love it up here. The hike is really fun. It's also a majestic view. You see panoramic landscapes. I love being able to be out here. As a young man, I was able to come and hike and camp in this area. Now, at this stage in life, I spend a lot of time doing the job I love but not being outside as much. So I get to bring my children down here sometimes and watch them run and hike around. I love it. My wife and I a few years ago had a chance to run a race down here. We ran clear out and around some of the area here. My favorite part of the

race was ending here in what's called the Valley of the Hutus. At the very end, we ran up some stairs and finished where we could see all the goblins. It's one of my favorite races I've ever run. I'm here at Arches National Park, admiring these tremendously large arches. Behind me, we have a couple of arches that look like a bridge. When I think of bridges, I think of teachers. Why do I think of teachers when I think of bridges? Well, one of the greatest teachers I've ever known of, Thomas Monson, used to talk about the importance of building bridges for others to cross. He once shared a poem called The Bridge Builder, in which an old man is traveling and he gets to a giant chasm. At the bottom of this chasm is a river. Now this old man has a lot of experience and was able to find his way across the river and to the other side of the chasm. Once he gets across, he stops to build a bridge. Another traveler passing by asks him, "Why are you building a bridge if you've already crossed?" Field man replies that he's not building the bridge for himself, he's building the bridge for others so that when they have to cross, they'll have an easier go at it. I'm grateful for the opportunity to be a teacher and I aspire to be a good bridge builder. I hope you too as a result of your education, take time to build bridges for others.

So I hope that gives you a better idea of who I am. I'm grateful to be working at the Gies College of Business and I hope you enjoy the course.

Lesson 1-1: Module 1 Overview

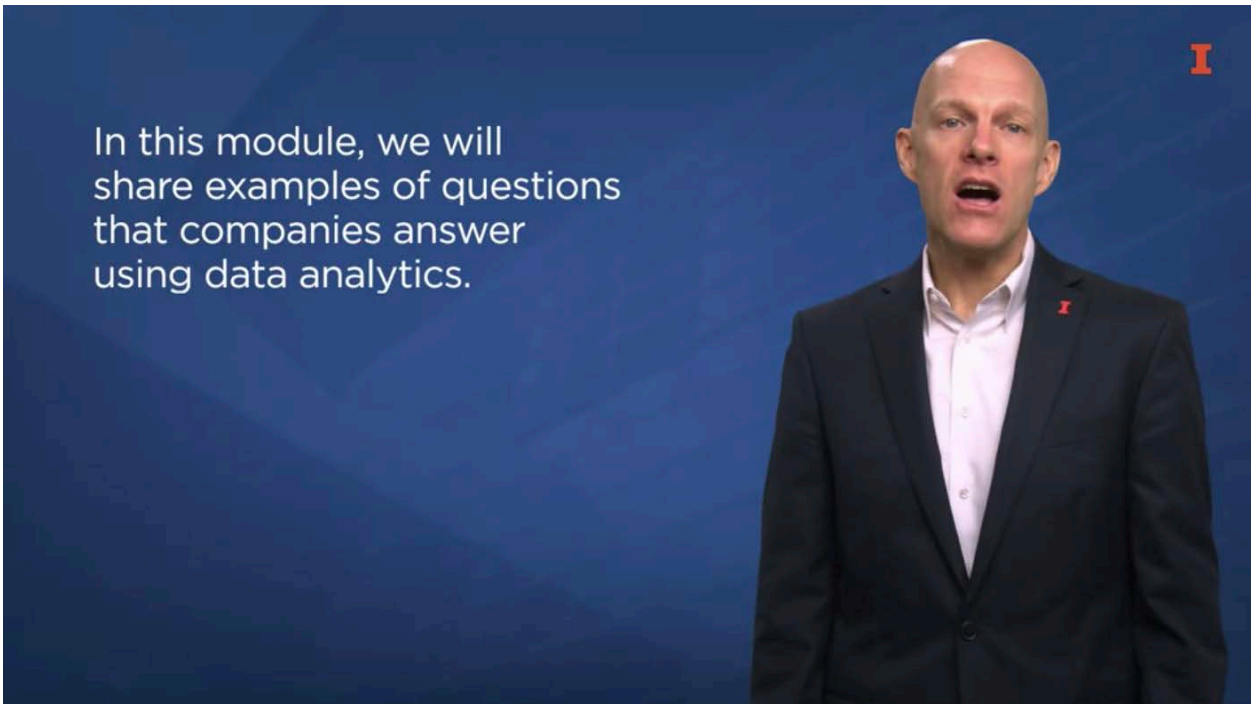
Lesson 1-1.1 Module 1 Introduction

Prof. Jessen Hobson: There are many powerful software tools for performing business analytic processes. In this set of lessons, we'll focus on the most important business analytic tool, which is your mind.

Prof. Ronald Guymon: There's a memorable nine word quip that highlights the importance of your mind. A fool with a tool is still a fool.

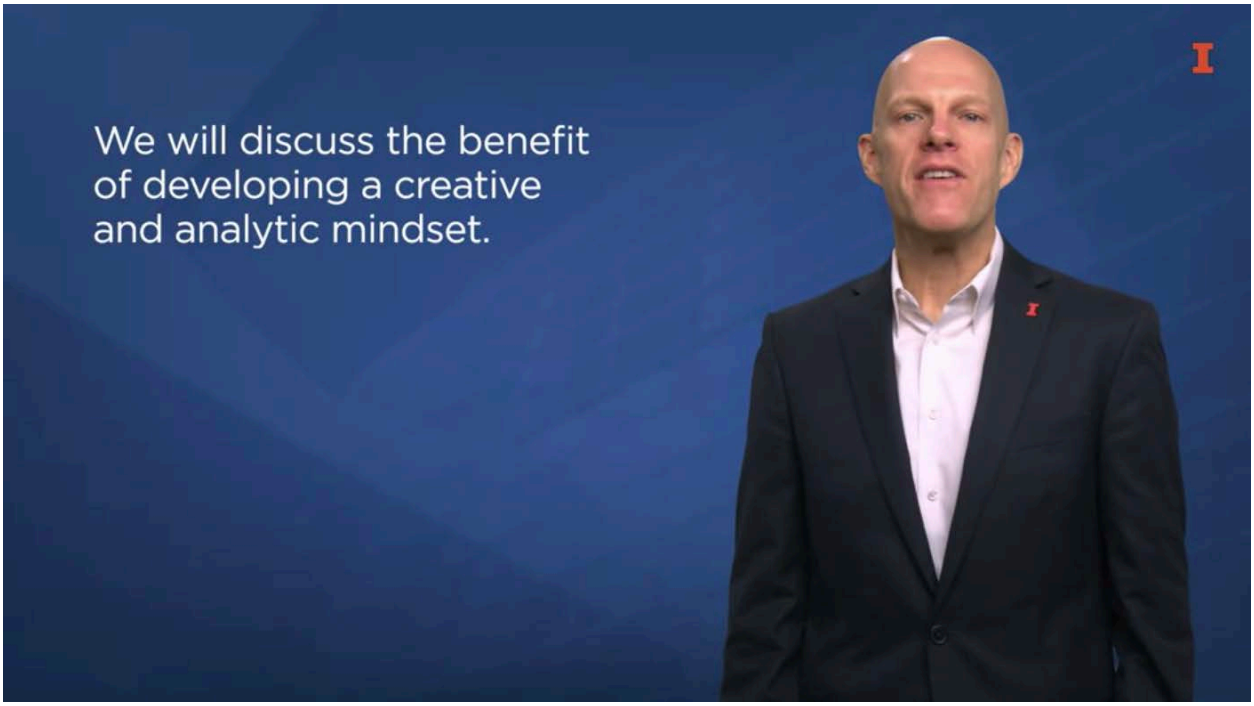
To illustrate that idea further. Imagine what would happen if a fool had a nice set of tools, but they used a hammer to pound in a screw or to crack open a soft boiled egg. They may actually do more harm than good.

On the other hand, someone who has a creative mind and some training with the hammer may find some really creative and productive ways to use that hammer. By the time you finish these lessons, your mind should be prepared in several ways.

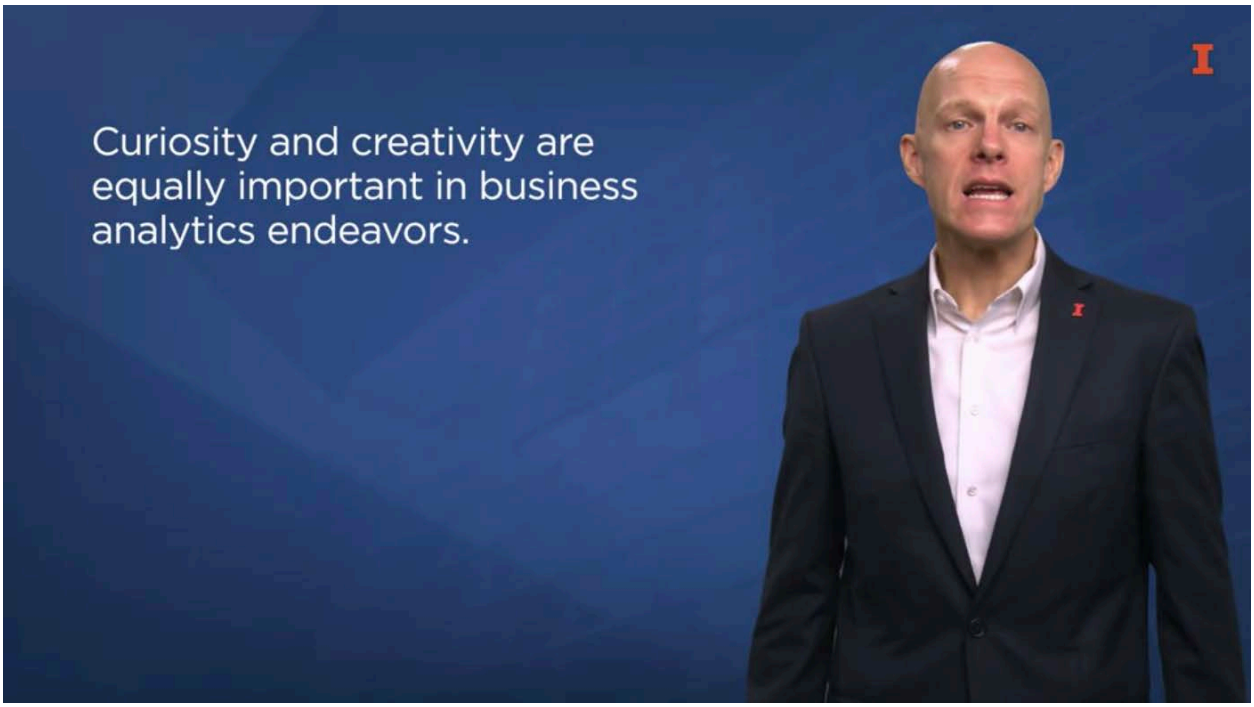


First, we'll prepare your mind by sharing several short examples of questions that companies answer using data analytics and highlighting the reality that the data itself

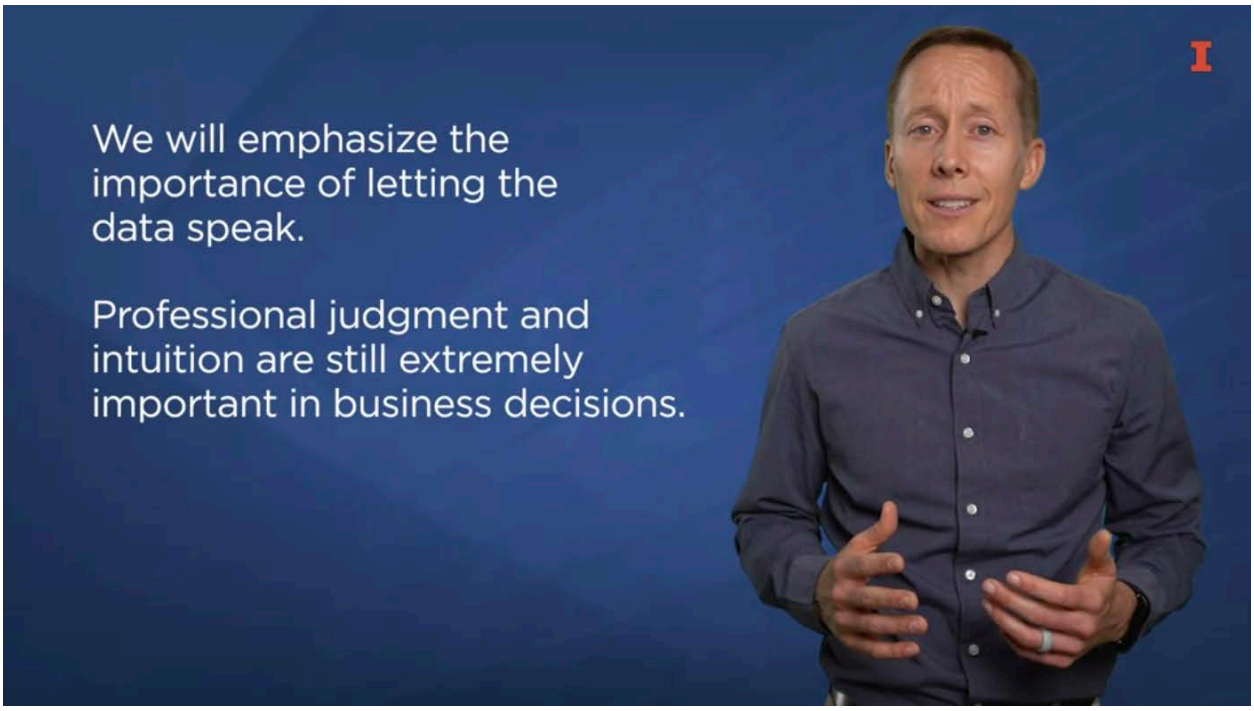
does not add value, but is more like a raw material that has value only after being processed correctly.



Second, we'll prepare your mind by discussing the importance of developing a creative and analytic mindset. Sometimes we hear those mindsets discussed as right and left brain activities and think that right brained activities are reserved more for the arts rather than the sciences.

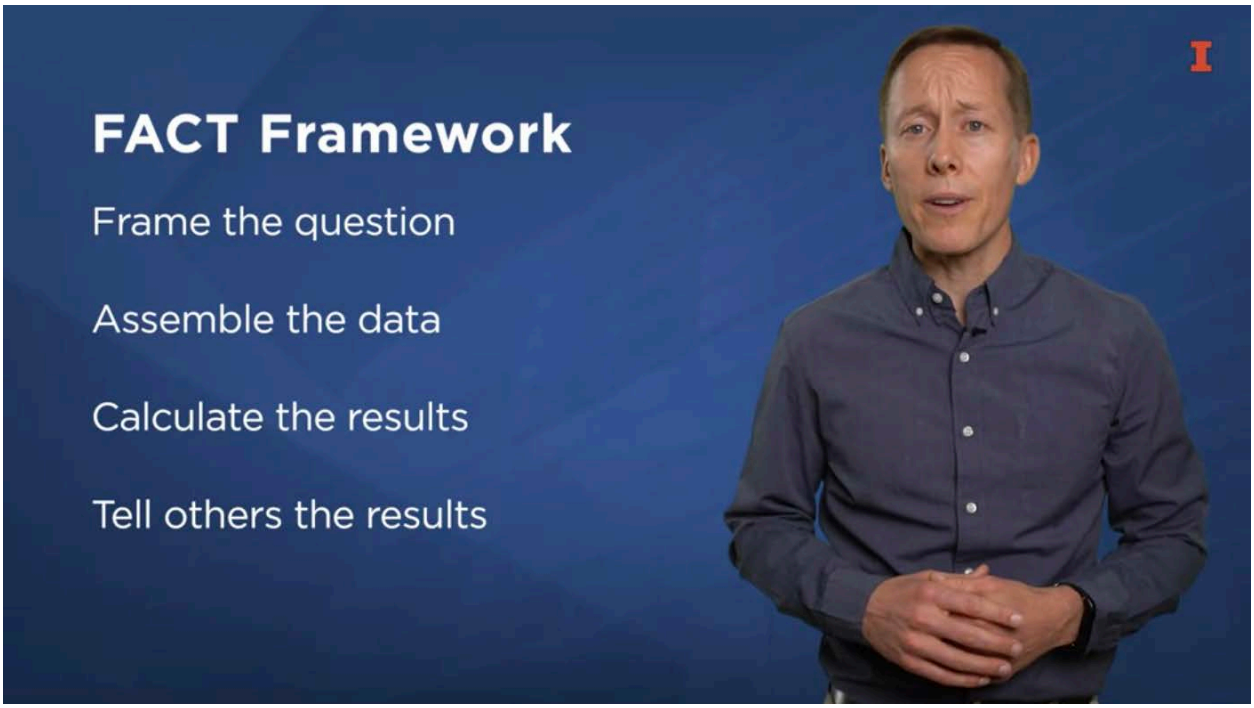


But the reality is that curiosity and creativity are just as important in business analytic endeavors, has the ability to break a problem down into smaller pieces and follow the structured process. This is partially because sometimes we pursue a business insight after recognizing patterns in the data, rather than starting with premises that are well laid out. The ability to connect those observe patterns with business processes, other questions and analytic processes requires a curious and a creative mind.



Third, we will prepare your mind by emphasizing the importance of letting the data speak. We don't want you to think that all decisions have to be made with data. Professional judgment and intuition are still extremely important in business decisions. However, it is important to be open to the possibility that business analytic results will oppose your beliefs. That is, it is important to let the data speak.

Fourth, we will prepare your mind by introducing a data set and elaborating on the fact framework for modeling that data. The data set is a sample of real business data and we want you to have some level of familiarity with that data and questions that it can answer. We hope that you'll recognize some business analytic applications for data sets with which you may be currently working.



We will then review the fact framework for analyzing data and elaborate on those steps to fill out a more nuanced data modeling pipeline.

Finally, we'll prepare your mind by reviewing some relevant business analytics terms and topics.

By the end of this module, we hope that you can say "A fool with a data tool is still a fool, but send that fool to data school and they'll be pretty cool."

References:

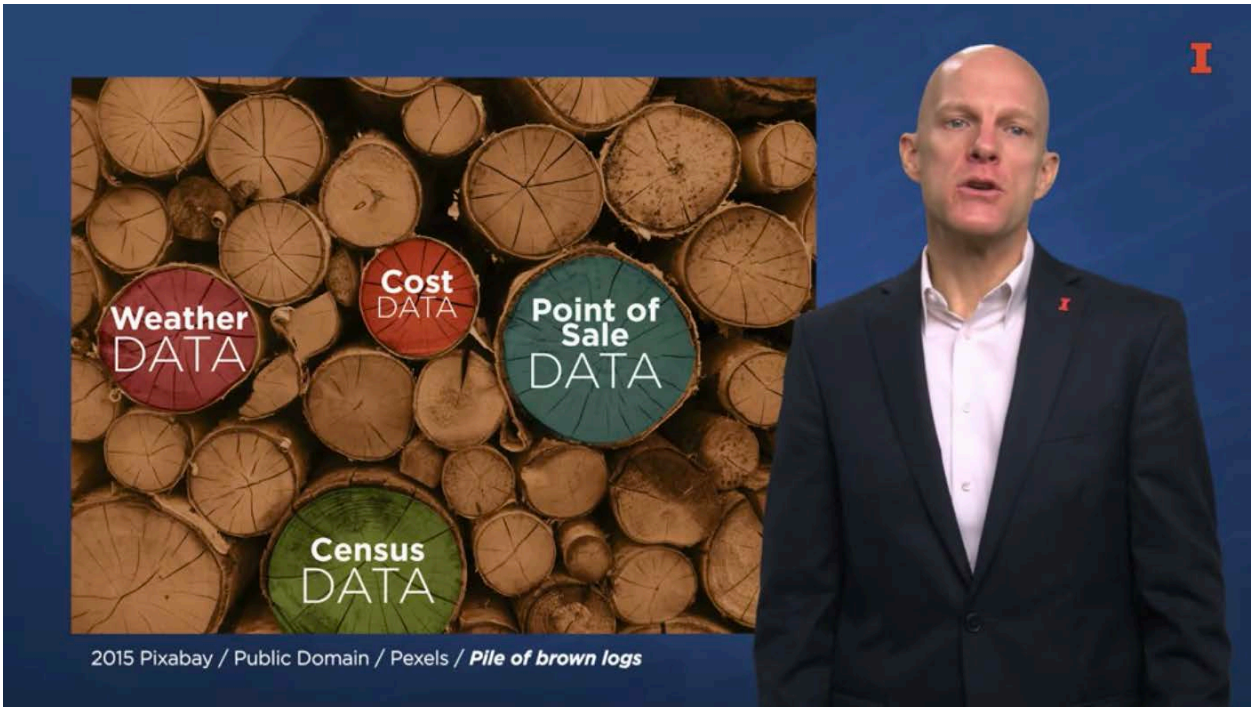
Myriams-Fotos (2016). ***Wooden board larch screw long screw hammer*** [Photograph]. Pixabay. https://pixabay.com/photos/wooden-board-larch-screw-long-screw-1337265/

stevepb (2014). ***Egg hammer hit beat fragile vulnerable threaten*** [Photograph]. Pixabay https://pixabay.com/photos/egg-hammer-hit-beat-fragile-583163/

Lesson 1-2: Business Context

Lesson 1-2.1 Turning Business Data Into Business Insights

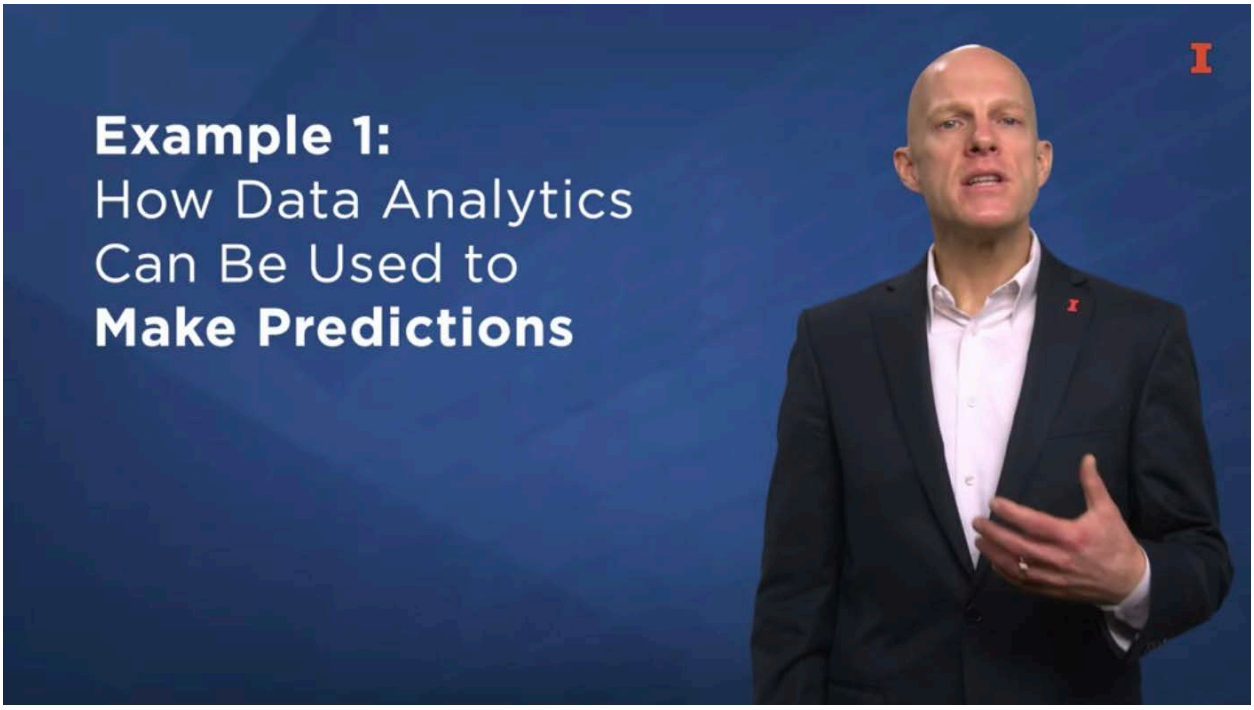
Data is valuable for many business decisions, but it often has to be processed before it can provide insight. Drawing on an analogy from the managerial accounting domain.



We want to emphasize that data is more like a raw material rather than a finished good that's ready to be consumed. In this lesson, we want to highlight some of the labor and overhead activities that are often applied to the data before it turns into insight.

In my experience, some people believe that data is a magical crystal ball that can provide insight into any problem. They may think that if you want to get more insight, then simply get more data regardless of the stores or quality of it. And without any thought about the tools that are used to convert the raw material data into a business insight. In this lesson, we hope to provide some high level details about the process of converting data into business insights. By sharing real life stories of companies that have transformed data into insight.

As we share these stories, please pay particular attention to these three things. First, the labor and other activities that were used to convert data to insight. Second, the variety of industries and tasks in which data provides insight. And third, the three categories of insight that data provides.



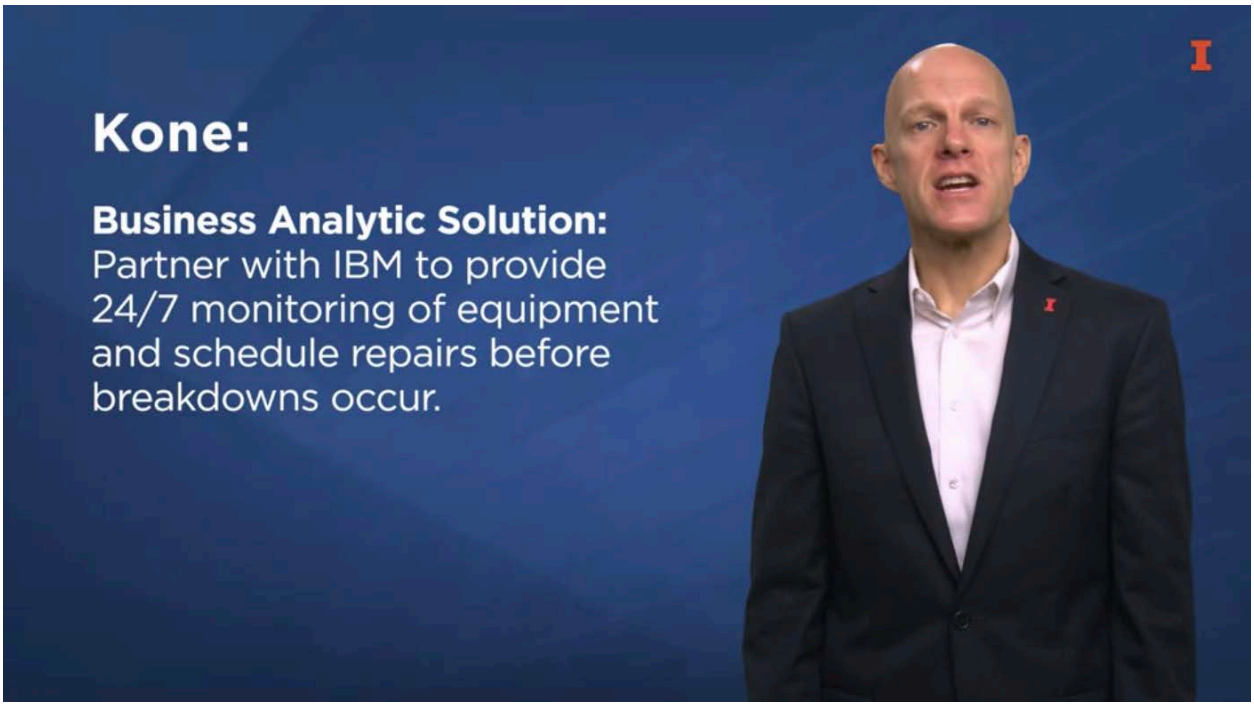
Let's first turn our attention to an example of how engineering and manufacturing

companies use data analytics to make predictions about when machines need maintenance.

In the manufacturing industry, machines are used to perform many tasks. But various parts of those machines wear out and need to be maintained or replaced or else they'll break. Even when machines have regular maintenance schedules, the unexpected breakdowns occur. Unexpected breakdowns carry high out of pocket costs, especially when wind broken part on the machine causes another part of the machine to break. Moreover, opportunity costs are high because of the time that must be spent identifying the source of the breakdown and then repairing the broken parts rather than making the product. One data driven solution is to use predictive maintenance.

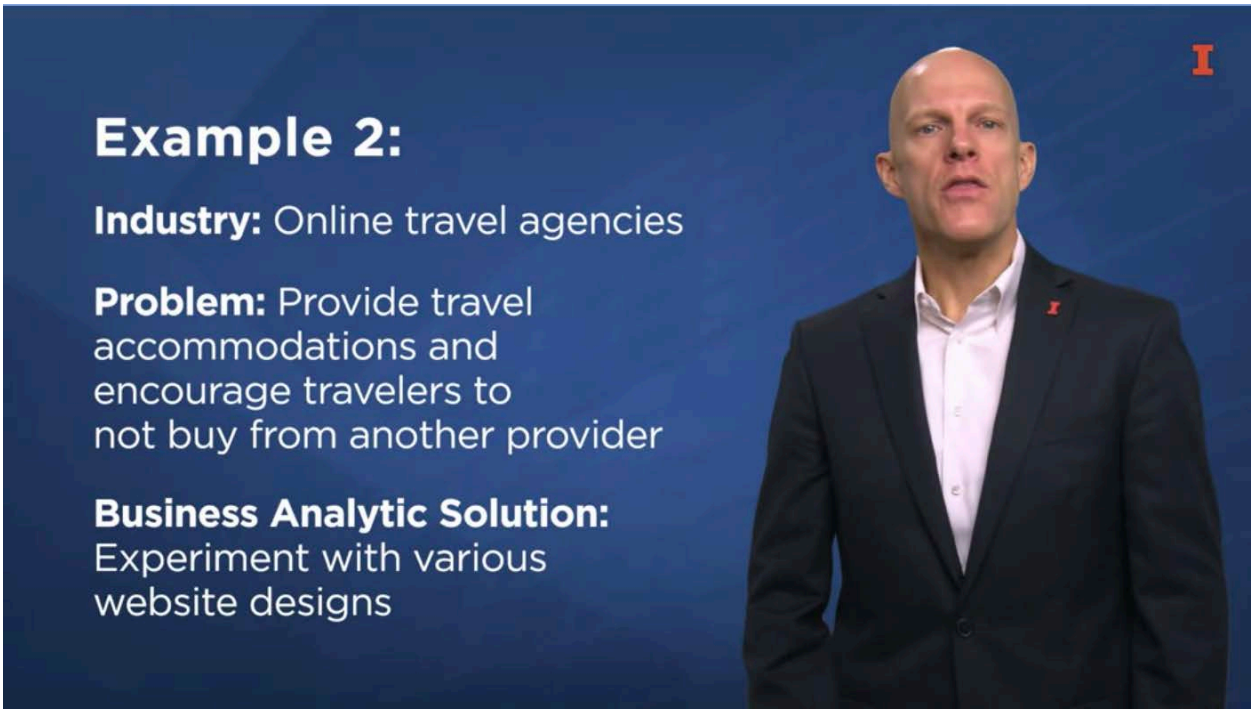
Kone is an engineering company that focuses on moving people, especially using elevators and escalators.

So if you've ever worked in a tall building, you've had to depend on an elevator to take you to your floor. Then you may know from personal experience that there can sometimes be long delays when an elevator goes out of service.



Kone has partnered with IBM to provide 24/7 monitoring of elevator equipment using sensors that are connected to the Internet. Now, rather than waiting for customers to call in and report that an elevator has been malfunctioning. They can use the sensor data to create models that will predict when a part will break and then schedule repaired during a time of low usage.

Alright example, to let's look at how data analytics can be used to effectively design websites.



Companies in the online travel agency industry [COUGH] tend to connect travelers with transportation places to stay, places to dine activities. As you may know from personal experience as a consumer, there's a lot to consider when planning a trip. So companies in this industry have to design a website that will provide the right amount of information.

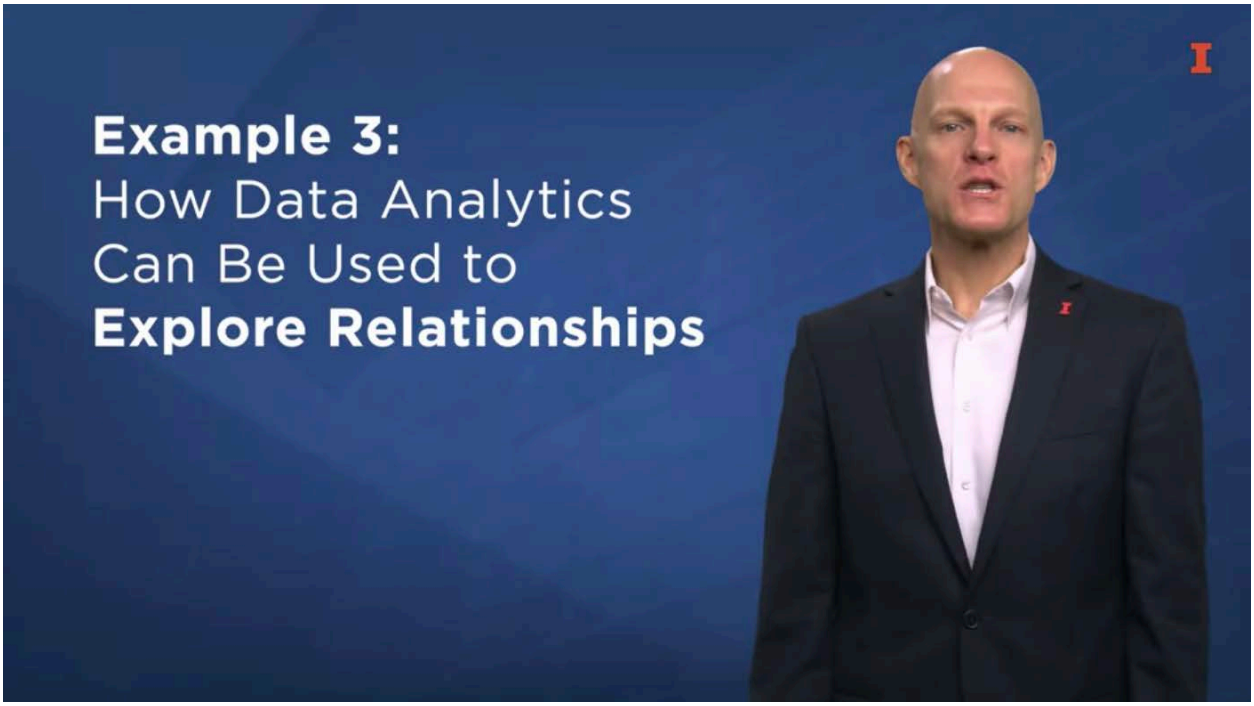
While also creating a sense of urgency so that the website visitor doesn't go off to some other website to make the purchase. So that's a that's a pretty tough balance to achieve, and there are bound to be many opinions.

So how do you decide on the right design? Well, companies can run A/B tests with various website designs, and then analyze the results. These A/B tests are essentially experiments in which consumers are randomly assigned to either the status quo version of the website version A, or to an experimental version of the website, version B. Customers behavior then can be quantified and analyzed to find out if the experimental version results in more desirable behavior than the status quo version. If so, then the experimental version becomes the new status quo.

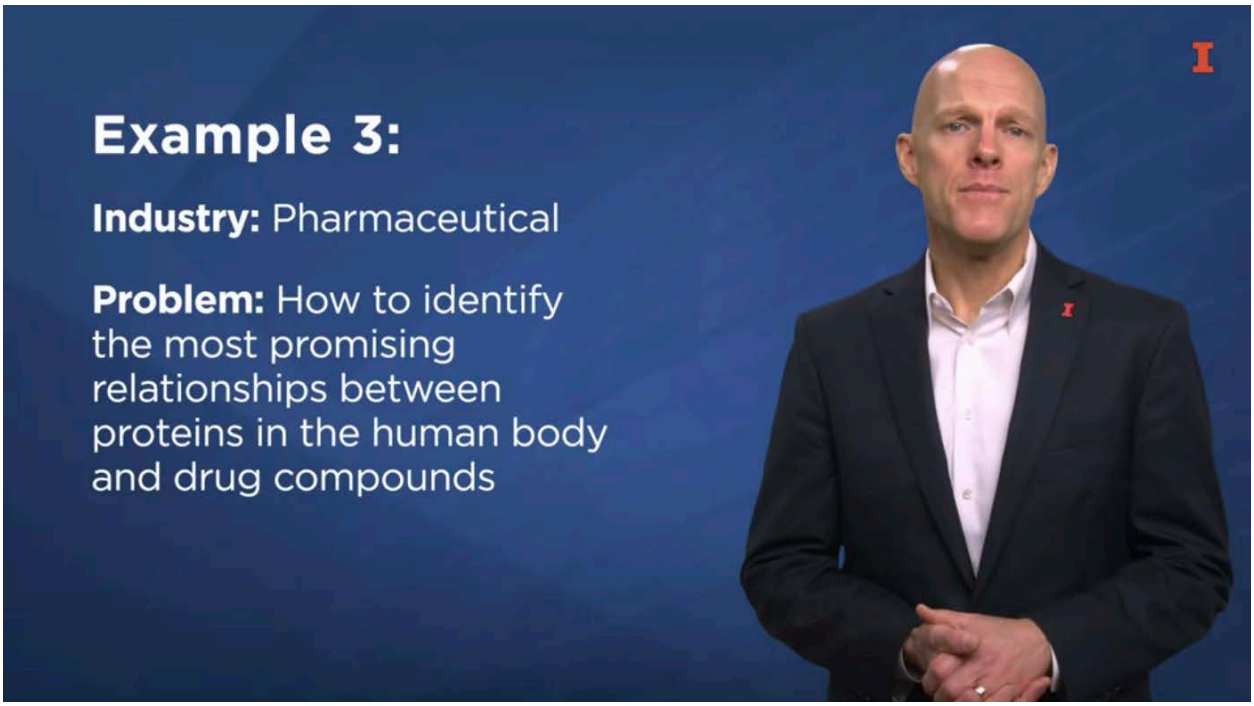
Booking.com is one company that takes this idea to the extreme. They enable every employee in their organizations to run experiments.

This freedom for all employees to run experiments results in thousands of experiments being run each year. Booking.com has a well-planned infrastructure in place to keep track of how customer behavior is affected by each version of the website. Even though

only about 10% or so of their tests results in improved conversion, they learn a lot about how to fine tune their website design and increase profits.

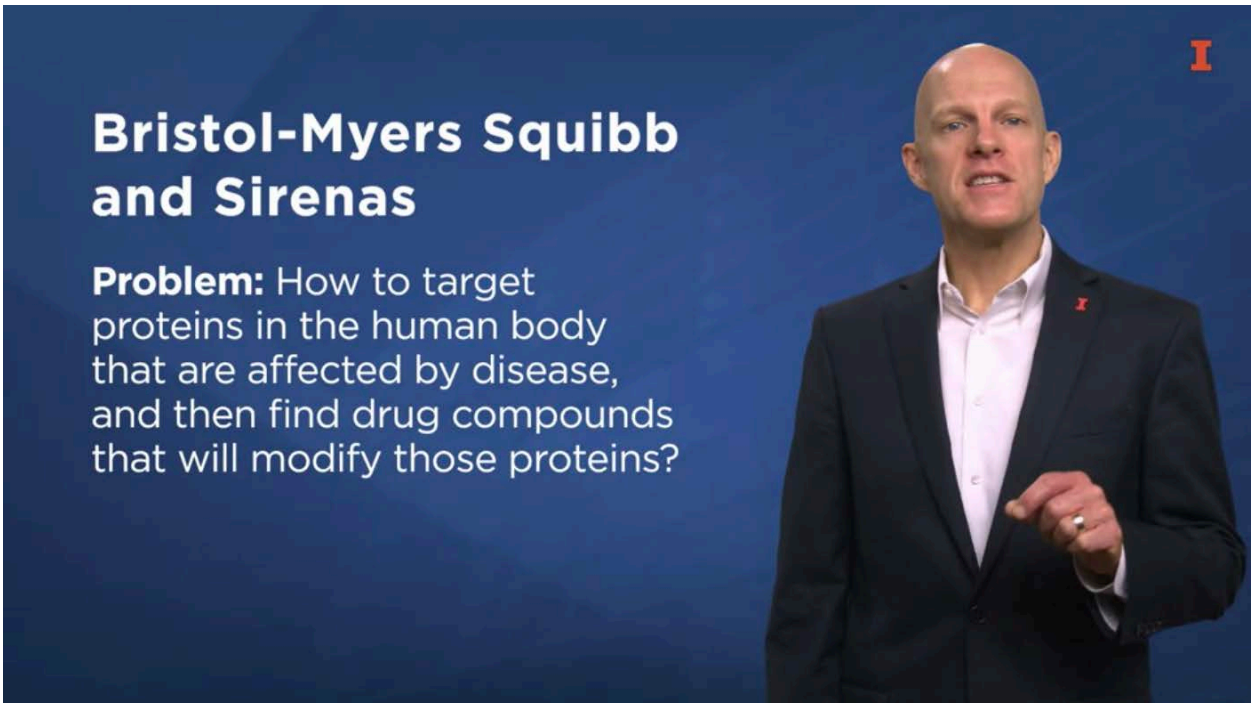


For the third example, we'll look at how data analytics can be used to explore relationships.

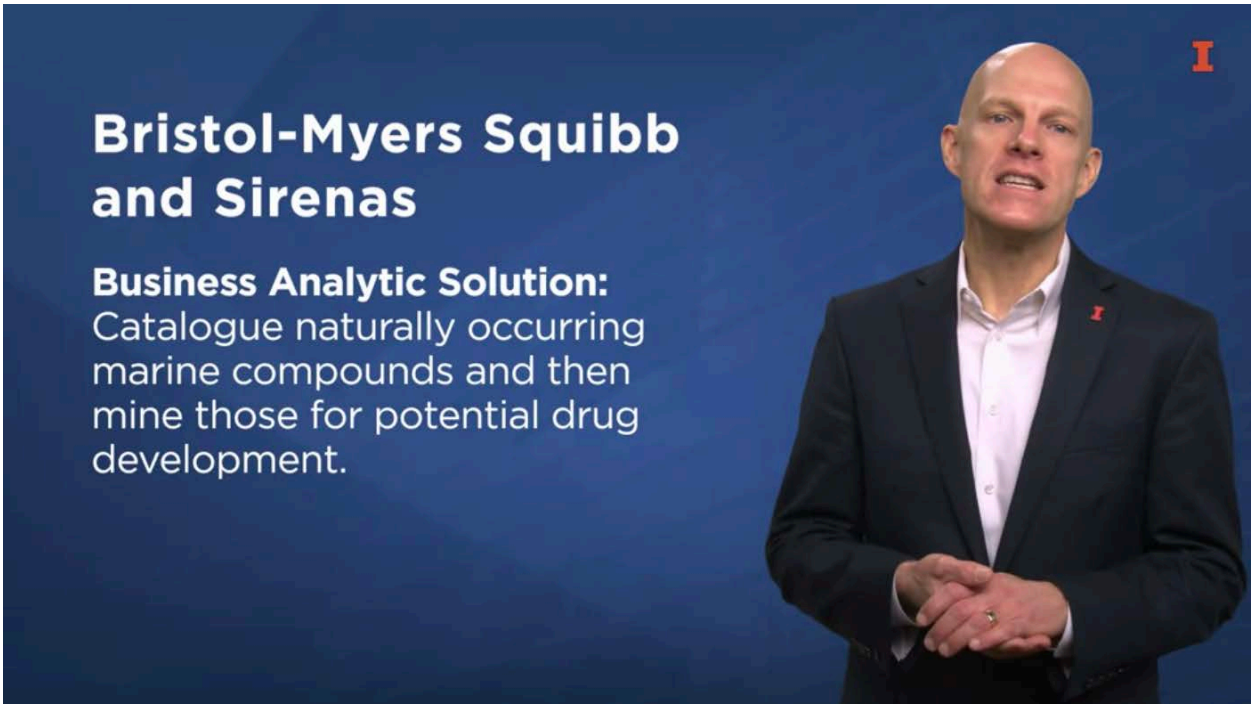


Drug discovery in the pharmaceutical industry depends heavily on exploring complex relationships between proteins in the body and chemical compounds. When you consider that there are tens of thousands of proteins in the human body that could be targeted for curing a disease, and zillions of molecules in the theoretical drug space. Identifying a potential drug to combat the disease is a mind blowing task.

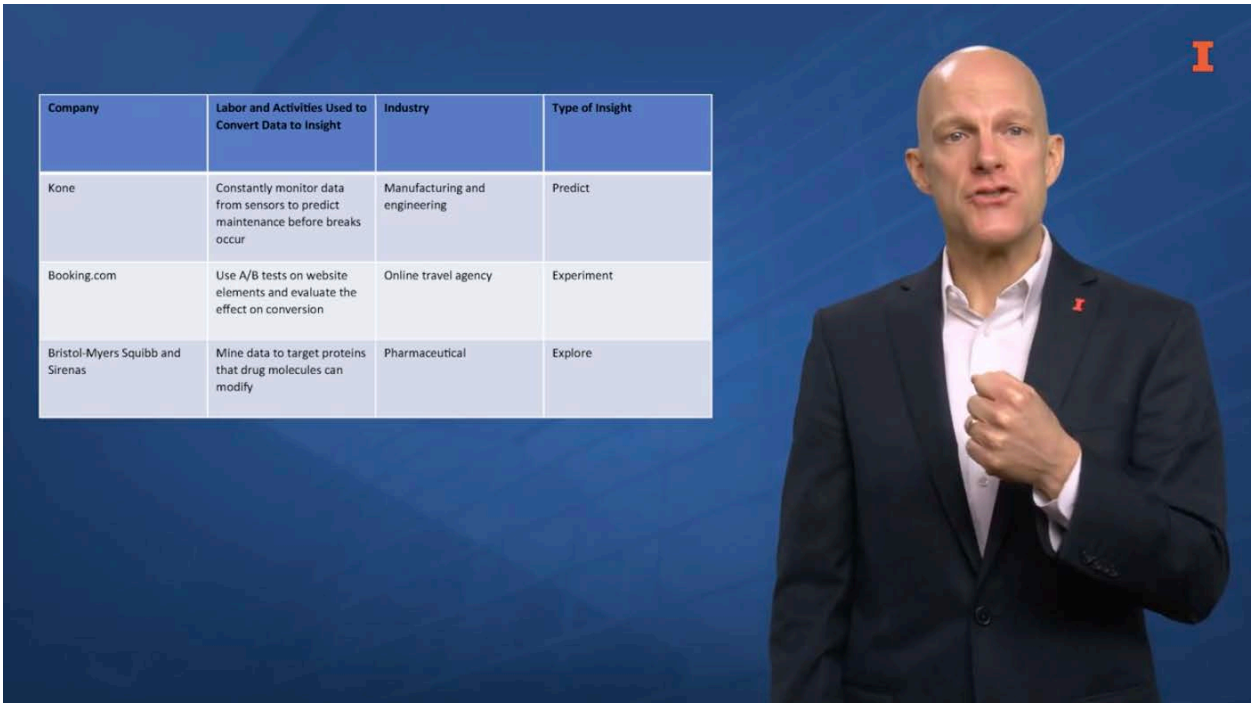
If we try to systematically go through each molecule and protein combination, then it would take forever. However, if companies can categorize proteins as well as compounds in the drug space. And then computational we explore a subset of the most promising relationships, then that process can move much faster.



Bristol Myers Squibb, a big pharma company, collaborated with Sirenas, a biotechnology company to discover effective therapeutics.



Sirenas has gathered a selection of marine compounds, and then apply their proprietary Atlantis data mining technology to quickly identify therapeutic candidates for treating diseases.



Now let's go back and review these three examples and appreciate the effort that went into converting the data to insight. In the first example, from an engineering and

manufacturing domain, data was created from sensors that were strategically placed on elevator equipment to provide continuous monitoring. The sensor data was then used to create models to predict future events. In the second example, from an online travel agency data was created from running experiments. That data has to be carefully created, stored and analyzed, so that Booking.com can quickly decide whether or not the new version of the website should be implemented. The data was used to provide insight about those experiments. And the third example, from a pharmaceutical industry, data was processed to quickly find relationships that could lead to therapeutic and effective treatments. The data was explored to find those relationships. So I think it's worth pointing out how important domain knowledge is and all of these examples. Without that domain knowledge and knowledge of the underlying business and its industry, the data would not be analysed in a way that leads to actionable insight. In conclusion, please remember that data is just a raw material. Sometimes data is not the right raw material to use to solve a business problem. However, we think that it often can be the right raw material, especially when it's combined with domain knowledge and effective data processing tools.

References:

Pixabay (2015). ***Pile of brown logs*** [Photograph]. Pexels. https://www.pexels.com/photo/bark-chopped-circle-dry-207296/

PublicDomainPictures (2012). ***Psychic psychics reading*** [Photograph]. Pixabay. https://pixabay.com/photos/psychic-psychics-psychic-reading-72085/

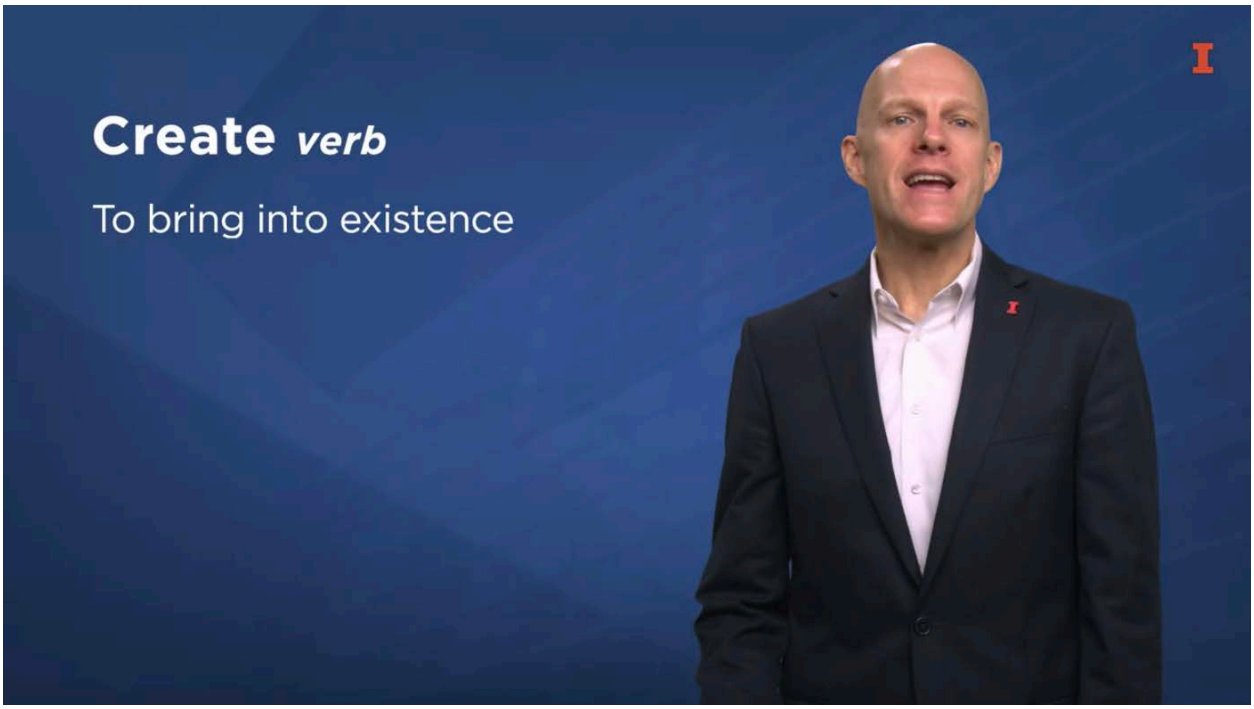
Kone Annual Report 2007 (2008). ***Kone Logo*** [Digital image]. Wikimedia. https://commons.wikimedia.org/wiki/File:KONE.svg

Booking.com (2017). ***Logo of Booking.com*** [Digital image]. Wikimedia. https://en.wikipedia.org/wiki/Booking.com#/media/File:Booking.com_logo.svg

Lesson 1-2.2 Your Mind is The Most Important Tool

In this lesson, we want to focus on creativity. That's right, creativity. That may seem unusual to some of you because creativity typically has connotations with artistic disciplines like painting. However, we think creativity is at the heart of business analytics. Let's investigate why by first defining creativity.

The Oxford Learners Dictionary defines the word creativity as the use of skill and imagination to produce something new or to produce art. As I mentioned, this definition supports the notion that creativity is something from the classic art domain. However, it also has a general application to something, perhaps anything that is new. Other definitions of creativity also include the idea of something new or original without specifying the domain. This definition also suggests that creativity requires skill and imagination.



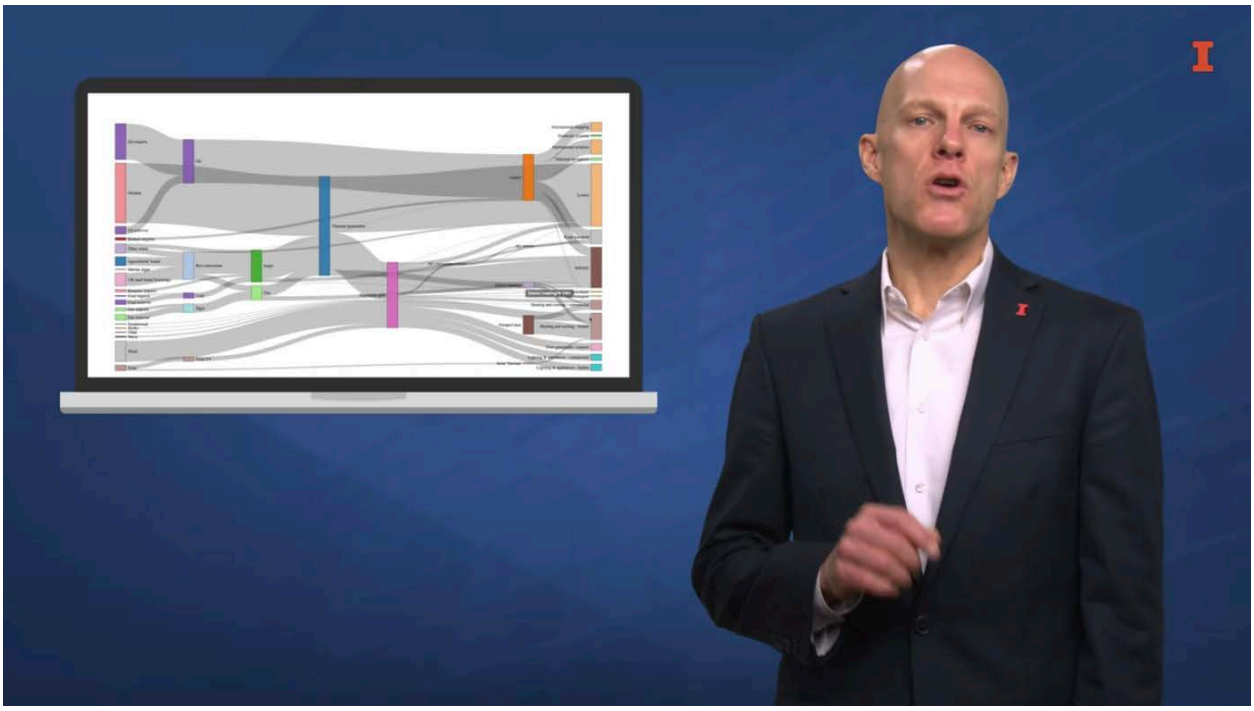
Another definition of creativity is effort to bring an idea into existence.

When you combine skill, imagination and effort, you're on the path to being creative.

Thus, creative artists such as Leonardo da Vinci and Vincent Van Gogh are skilled at using tools and raw materials such as paint brushes to paint. They also use their imagination, which I suspect is heavily influenced by their domain knowledge, such as

knowledge of the human body. Finally, they put forth an effort to apply their skills to bring forth what's in their imagination onto a Canvas resulting in an original painting.

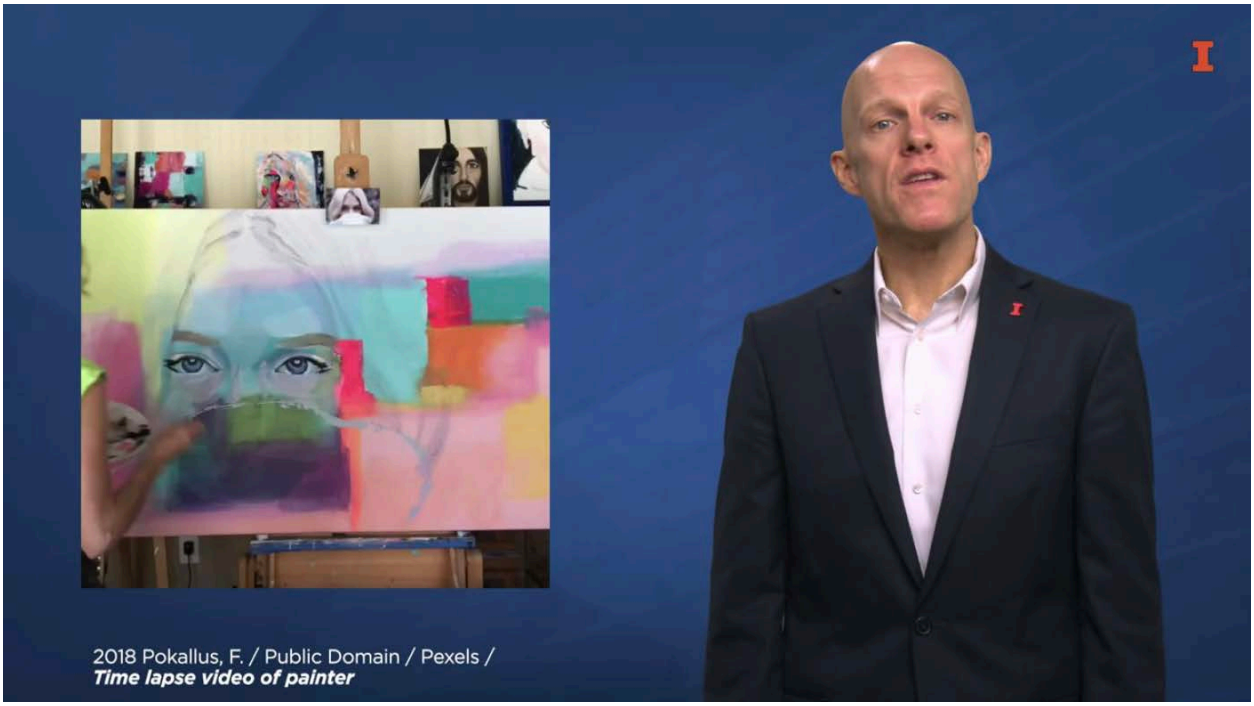
Now, with respect to business analytics, the tools are the software and the hardware. The raw material is the data.



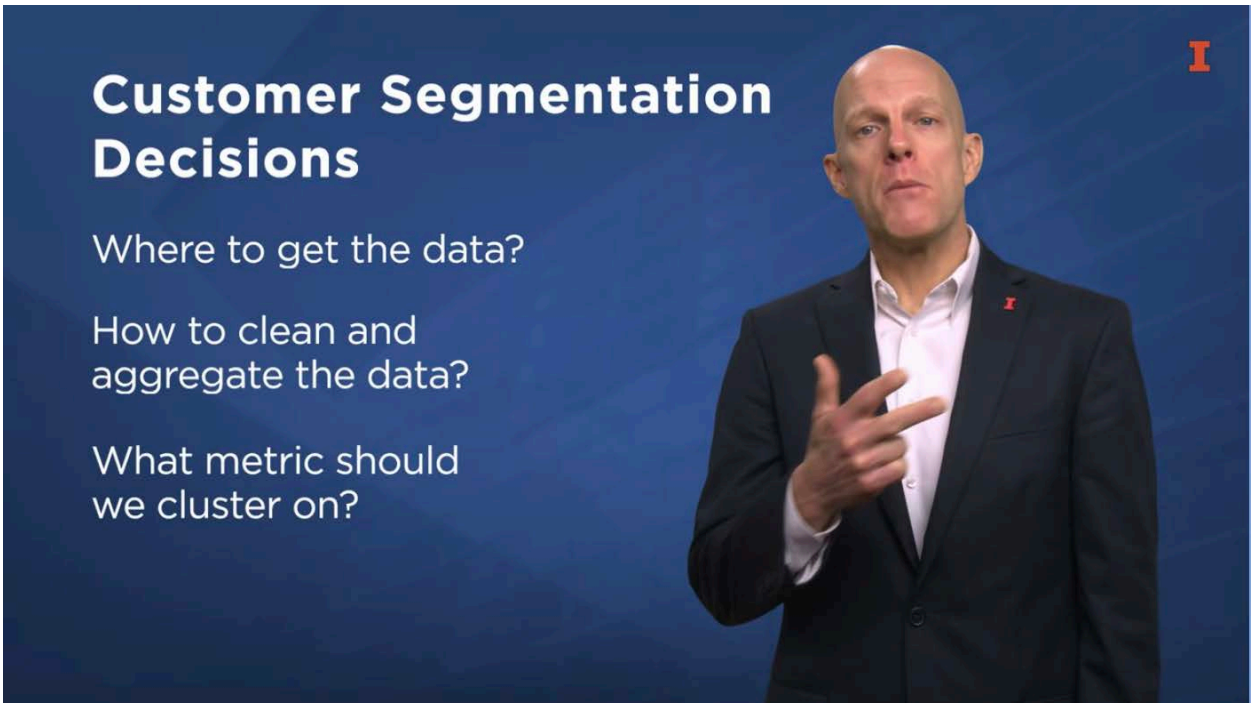
Imagination, I suspect, is heavily influenced by business domain knowledge, such as how a product has been marketed or how assets in the portfolio have been selected. By putting forth effort, a business analyst can apply their skills to identifying and fixing a business problem, resulting in an original solution.

Therefore, we want to emphasize that business analytics requires a lot of creativity. We hope that's good news to those of you who think that business analytics is simply

memorizing rules and equations and applying the same repetitive processes over and over. Now, we also want to emphasize that a creative mindset is not at odds with an analytic mindset in which you separate something down into smaller parts and work systematically on those parts such that they combine into a cohesive, singular output.



I suspect that artists like Leonardo da Vinci take an idea that is in their head and break it down into smaller parts and then work on implementing those parts in a structured, systematic way. When painting a face, the artist may first sketch the outline of the face, then sketch the eyes, nose, mouth, and ears, all before even adding paint. Then when adding paint, they may paint those parts before filling in the face with colors. Similarly, a business analyst needs to take a business problem and break it down into smaller parts and then work on implementing those parts in a structured, systematic way.



For example, when coming up with a customer segmentation strategy, a business analyst has to decide where to get the data, how to clean up and aggregate the data, what metrics are most important for segregating customers,

how many different customer segments to create, and how to communicate the results to the stakeholders in a meaningful way. With respect to business, analytics and art, achieving a creative and analytic mindset is just as important, if not more important,

than the tools and the raw materials. A creative mindset is heavily influenced by skill, imagination and a willingness to invest effort.

In fact, a creative analytic mindset is your most important tool. That's important to emphasize because in this course our focus is really on gaining skills with tools of secondary importance. That is, software tools for processing data. There's a wide array of tools that you can use to process data, and we encourage you to become familiar with and eventually even master some of those tools. These tools most often take the form of software applications like R, Python, PowerBI, Alteryx, and Tableau, all of which can be run on or at least from a personal computer. At some point, you may also start learning about the hardware like CPUs and GPUs, but that's for another course. The advancement in software and hardware makes data analytics possible, and without them we wouldn't be able to get much insight from all of the data that we amass. Just remember that while these tools are amazing, continuously evolving and very important, they don't create value on their own any more than laying a paintbrush next to a paint and Canvas will create an original work of art. The paintbrush and paint only result in an original work of art when a skilled and imaginative artist exerts effort to systematically apply paint to a Canvas. Our goal is that by the end of this course, you will have developed skills with that analytic tools so that if you intentionally apply these tools to data, you'll be able to come up with creative solutions to many business problems.

References:

Free-Photos (2016). ***Paints colorful painting art*** [Photograph]. Pixabay. https://pixabay.com/photos/paints-colorful-painting-arts-1149122/

WikImages (2013). ***Monalisa painting art oil painting*** [Digital image]. Pixabay. https://pixabay.com/photos/mona-lisa-painting-art-oil-painting-67506/

Prawny (2021). ***Vintage sky night starry night Vincent Van Gogh*** [Photograph]. Pixabay. https://pixabay.com/illustrations/vintage-sky-night-starry-night-5971661/

Microsoft (2021). ***Power BI sign in*** [Website]. Microsoft.com. https://powerbi.microsoft.com/en-us/landing/signin/

RStudio, PBC (2021). ***RSTudio*** [Website]. https://www.rstudio.com/

Pingwing (n.d.). ***Tableau Software Computer Icons Alteryx, Over And Over Again, blue, angle, text*** [Digital image]. Pingwing.com. https://www.pngwing.com/en/free-png-tdexp

References:

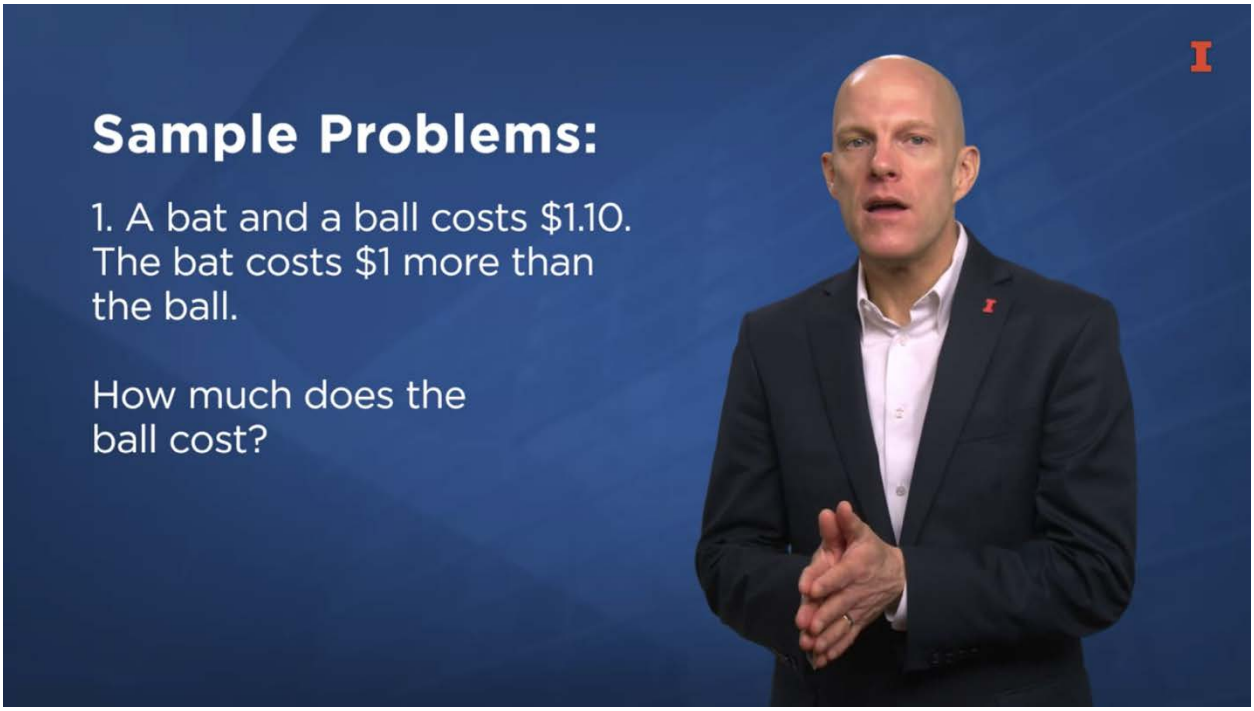
Pokallus, F. (2018). ***Time lapse video of painter*** [Video]. Pexels. https://www.pexels.com/video/time-lapse-video-of-painter-1051066/

Chiplanay (2017). ***Science brain bulb 3d poly triangle lights*** [Digital image]. https://pixabay.com/illustrations/science-brain-bulb-3d-poly-2953886/

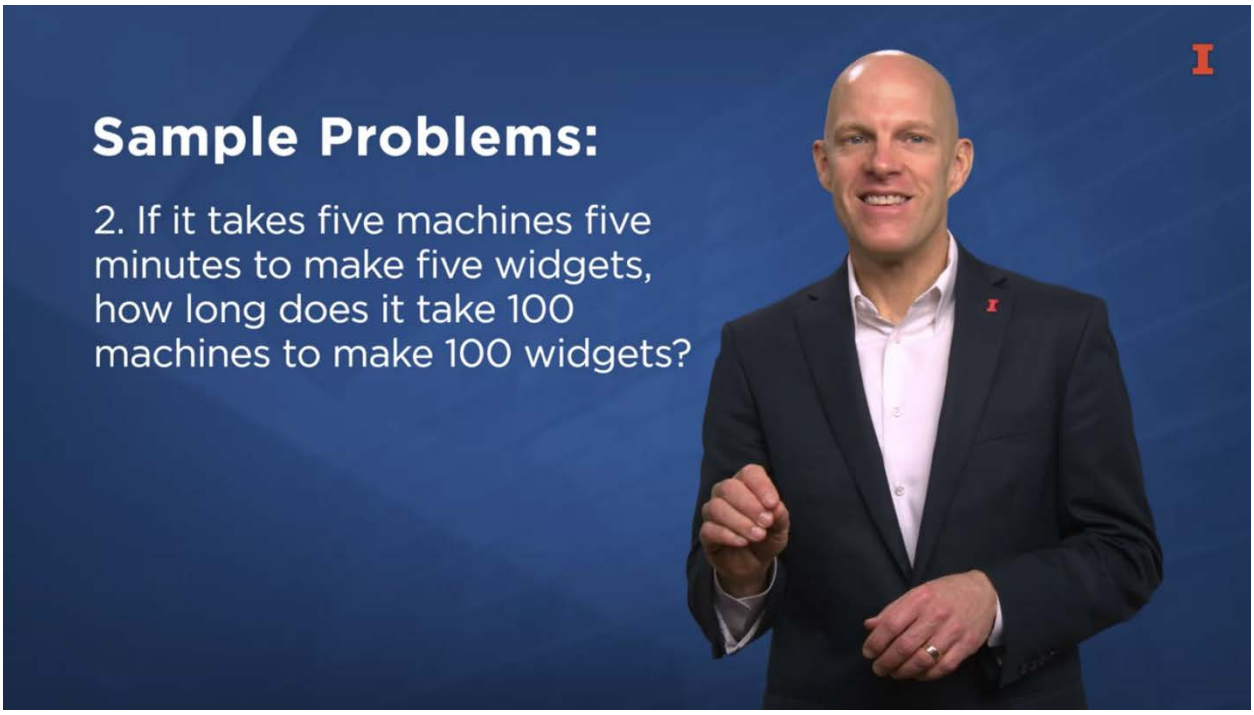
Wallstrand (2016). ***Screen laptop png computer illustration*** [Digital image]. https://pixabay.com/illustrations/screen-laptop-png-computer-1515324/

Lesson 1-3: TECA Data Set

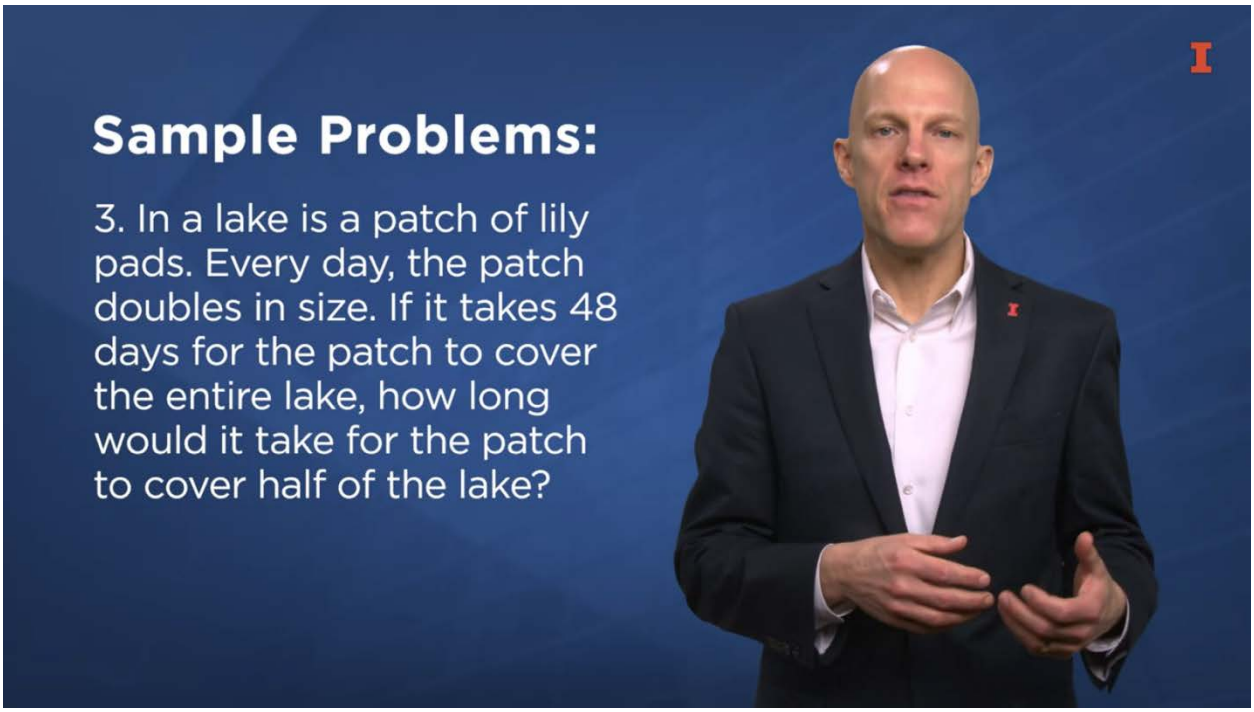
Lesson 1-3.1 Analytics Mindset: System 1 vs System 2



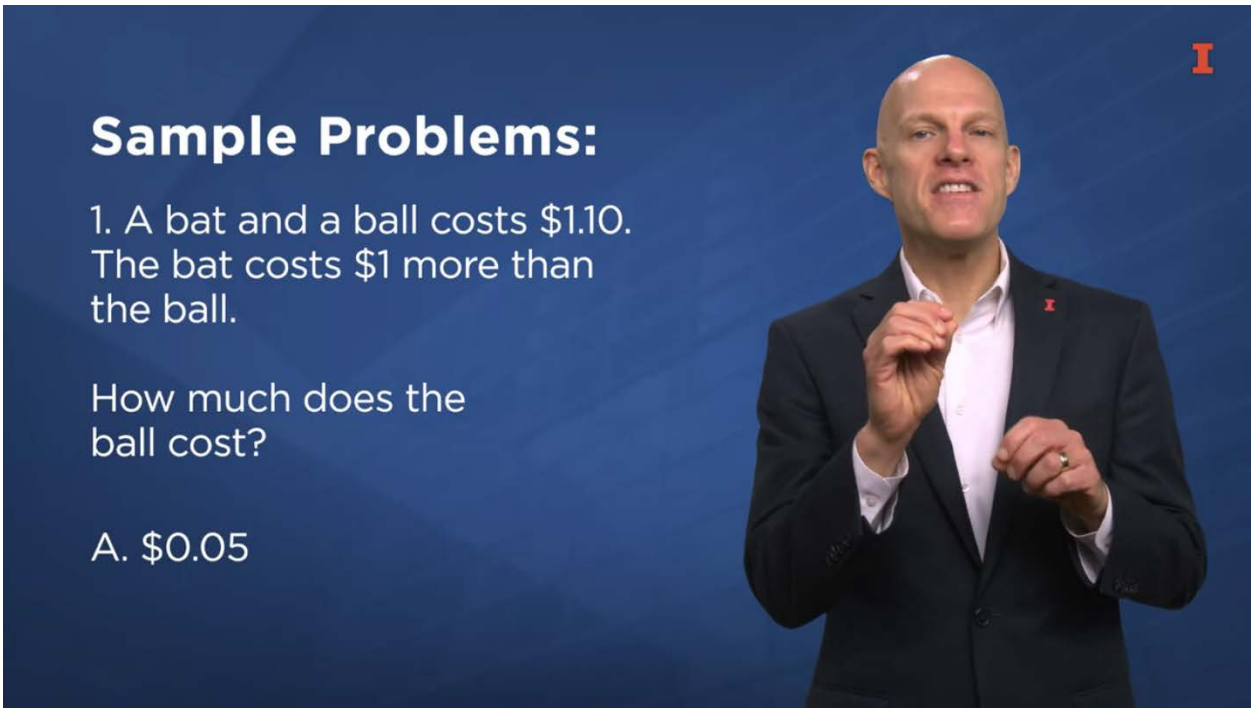
In this lesson we want to dive deeper into what it means to have an analytic mindset. So let's take a look at a few simple problems. I'm going to ask you these questions and then give you only a few seconds to write down a response, okay? Are you ready? Here it goes. Number 1, a bat and a ball cost a dollar and 10 cents. The bat costs $1 more than the ball. How much does the ball cost? You got that?



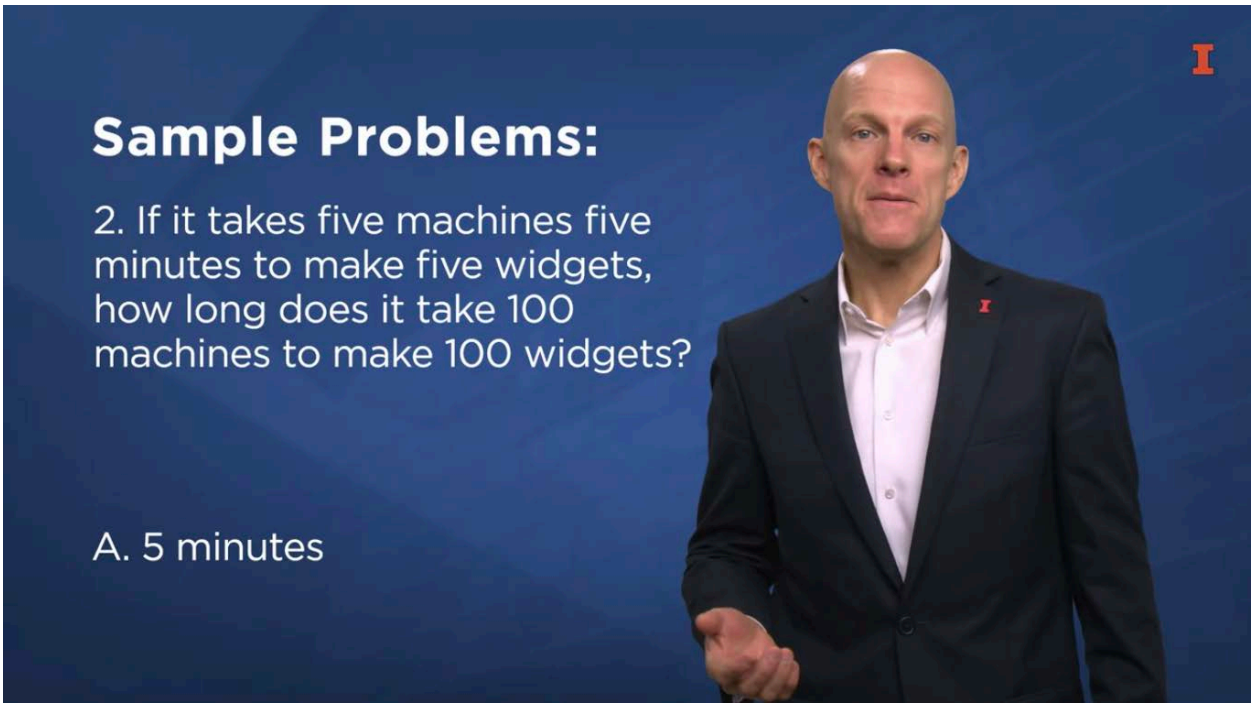
All right, number two. If it takes five machines five minutes to make five widgets, how long does it take 100 machines to make 100 widgets? Got it?



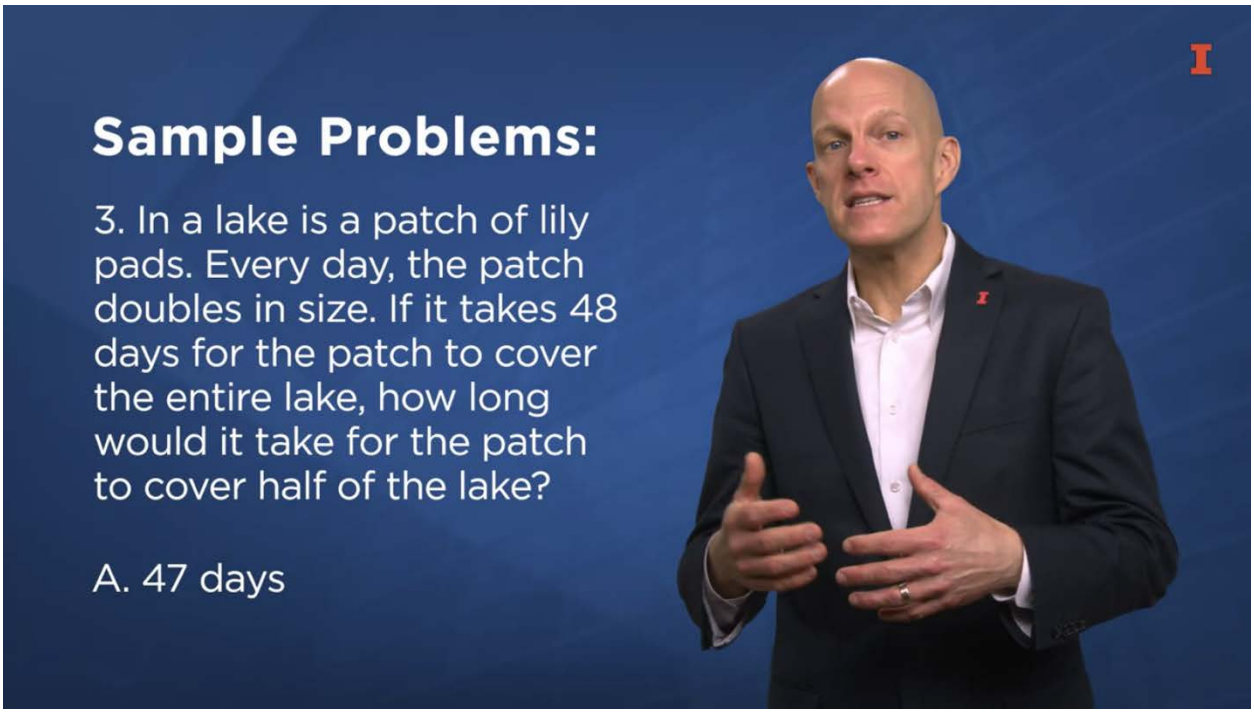
All right, number three, in a lake is a patch of lily pads. Every day, the patch doubles in size. If it takes 48 days for the patch to cover the entire lake, how long would it take for the patch to cover half of the lake?



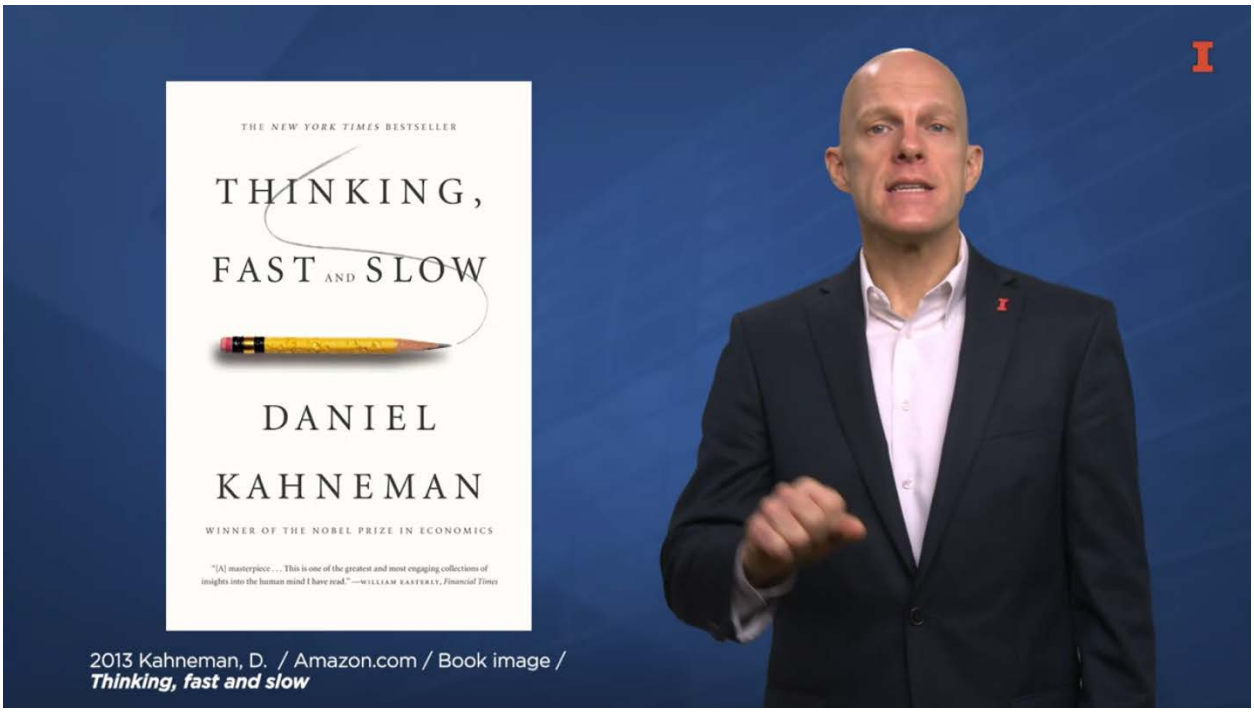
All right, let's see how you did. The answer to the first question is five cents.



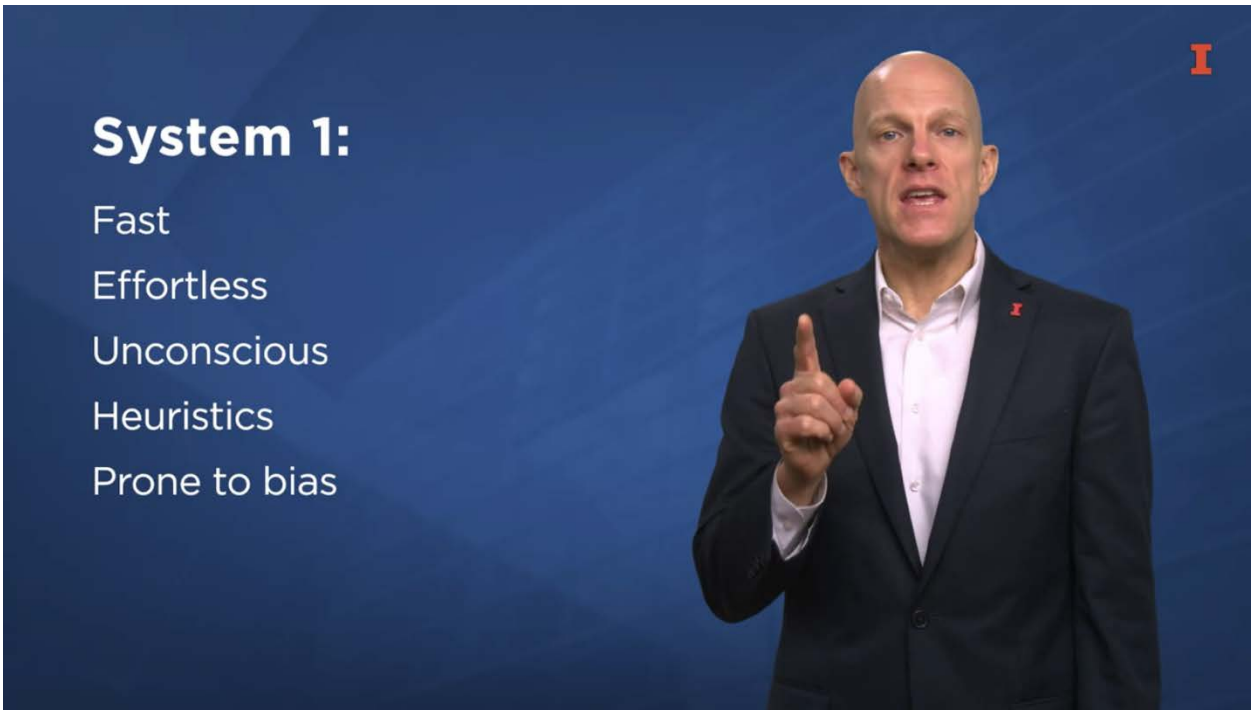
The answer to the second question is five minutes.



And the answer to the third question is 47 days. Now, [LAUGH] if you answered 10 cents, 100 widgets, or 24 days, then you're wrong. But you're in good company. Each of these questions has an obvious but incorrect answer that most people respond with. So when you go with the obvious answer, it's probably because you didn't have a reason to take the time to think deeply and carefully about them. And I didn't give you much time to think about it. Daniel Kahneman is a Nobel winning economist. His work has focused on judgment and decision making and has had a really big impact in many fields, including business. The findings from his research have challenged the traditional economic assumption that humans are rational decision makers

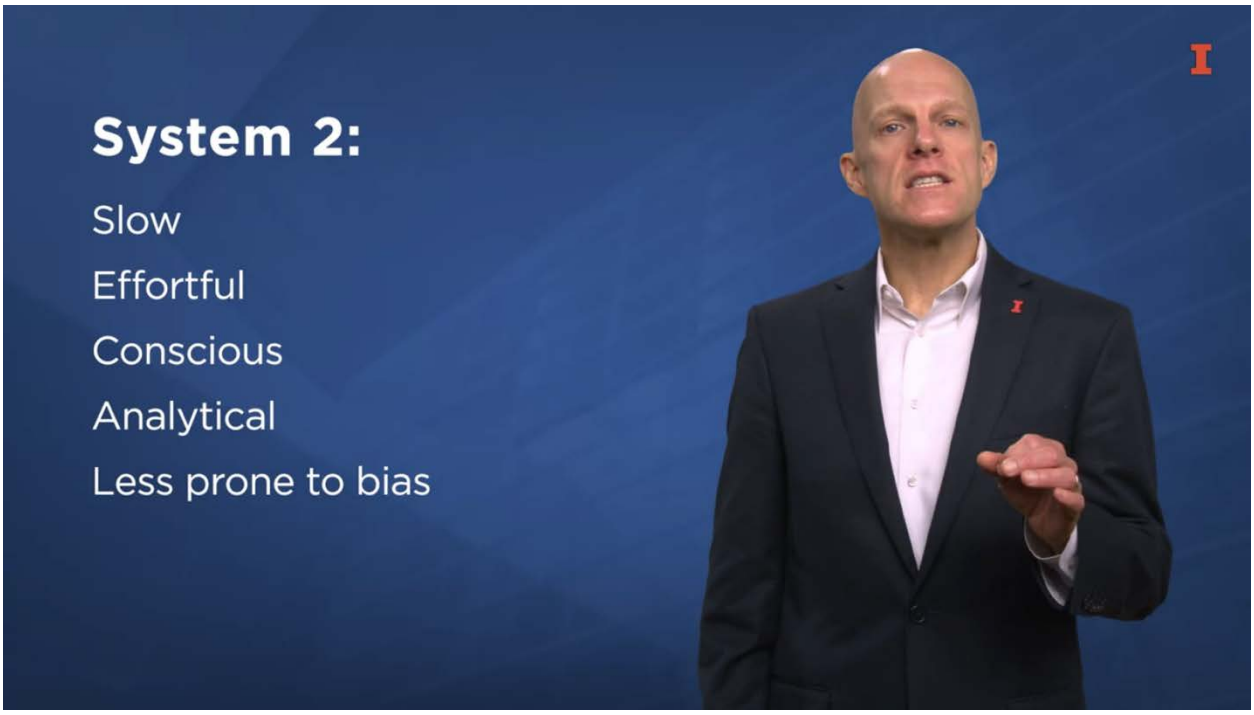


In his book, Thinking Fast and Slow, he talks about how humans have two systems for making decisions. System one is fast, effortless, and intuitive. In contrast system too is slow, effortful, and more analytical.

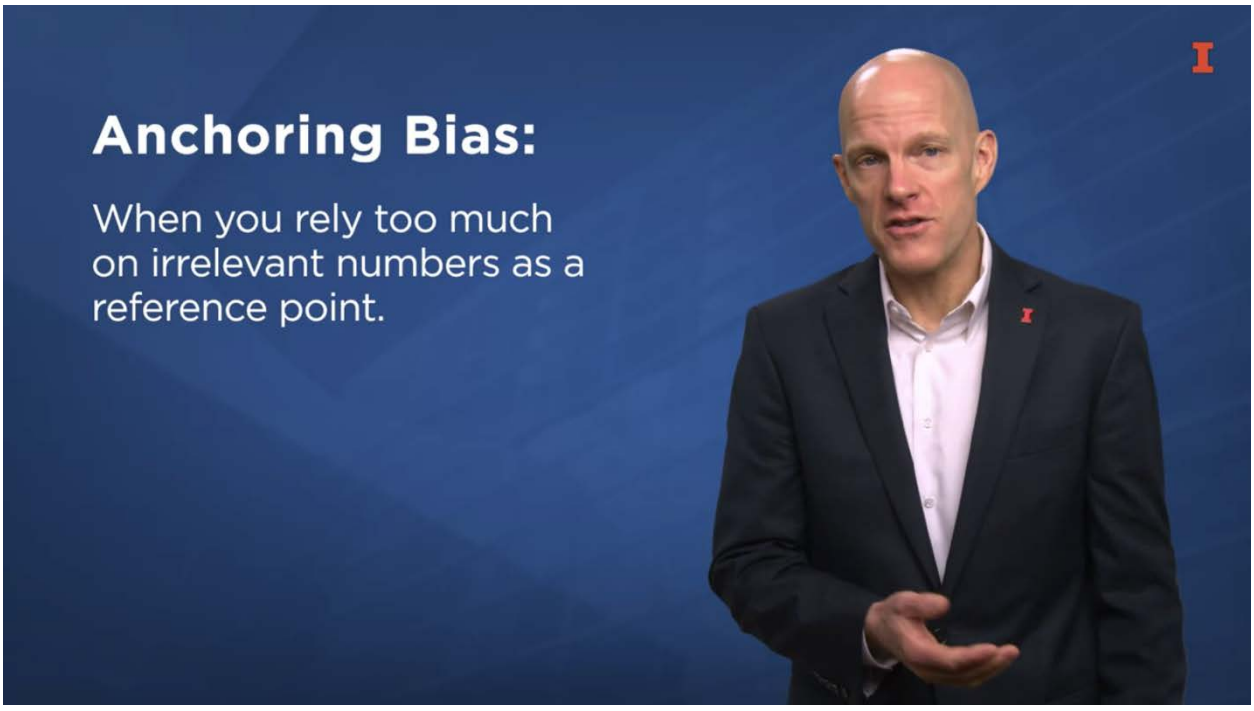


The system one pattern of thinking is fast and effortless and relies on cognitive shortcuts called heuristics or rules of thumb for making decisions. When we use heuristics, we try to fit observations into an existing pattern of thought rather than taking

time to create a new one. Heuristics are often very useful because we don't have time and energy to carefully consider all of the information that is presented to us. For example, they allow us to redesign on a billboard and quickly recognize how far away something is located from us without having to process every single detail of the sign. However, there are some common instances when these heuristics will lead to biased or incorrect answers.

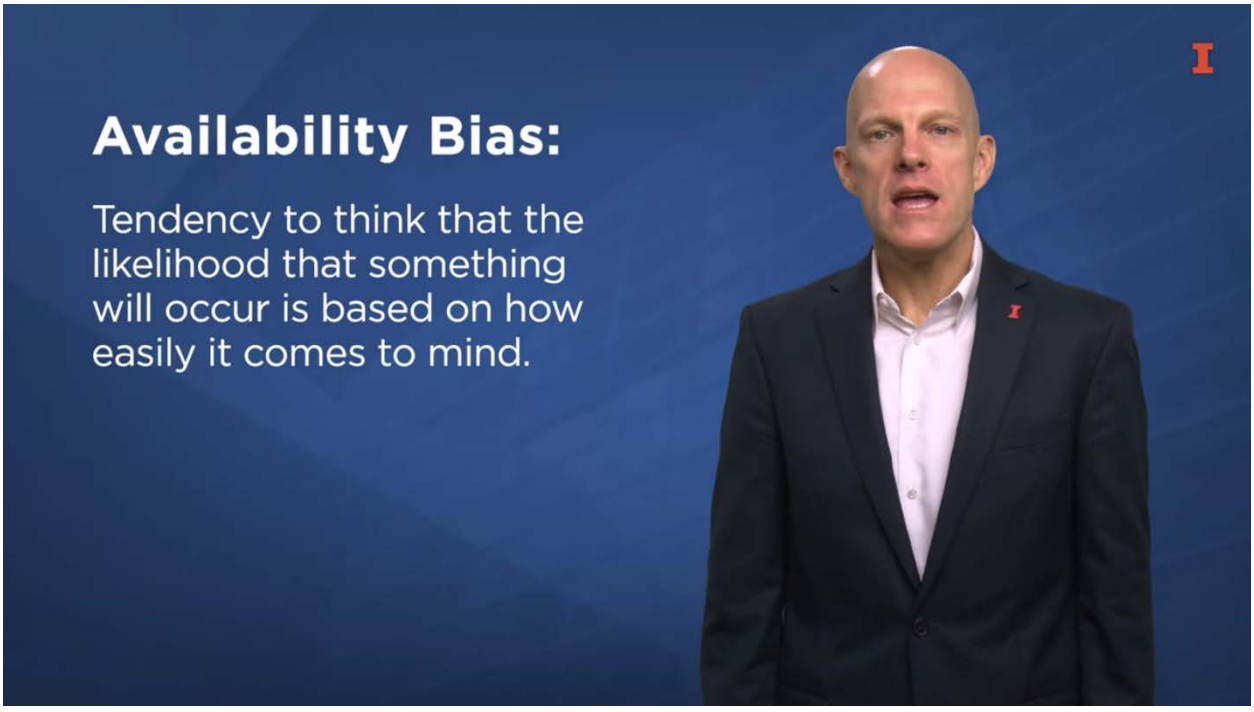


Now if we're aware of these situations, then hopefully we'll be able to switch over to system two, which is the slower, more conscious, deliberate, analytic mindset. Then we'll think more carefully before we arrive at a final unbiased decision. So let's take a look at some biases that you may face when dealing with big data so that you're better prepared to recognize them.



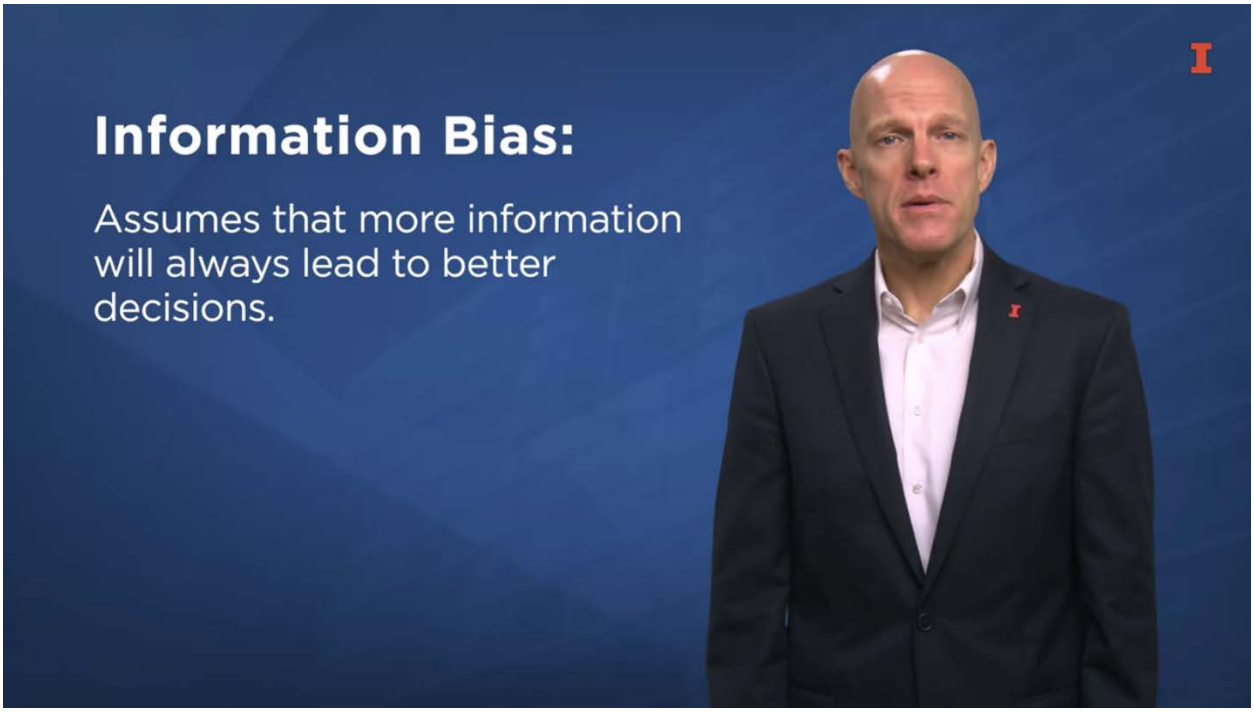
First, let's consider anchoring. The anchoring bias occurs when you rely too much on irrelevant numbers as a reference starting point and don't adjust quickly enough. For instance when you're buying a suit, you may find that the salesperson will first mention the regular price of the suit at $500. Before then letting you know that it's half off and only $250. The original anchor of $500 makes $250 seem like a great deal like a steel even if the same suit was being sold in another store for $200.

The clustering illusion is a tendency to see patterns in random data. This is important in the financial context when looking at the pattern of saying that income or the stock price over time. Your estimate of the amount of variability may be lower than what it really is.



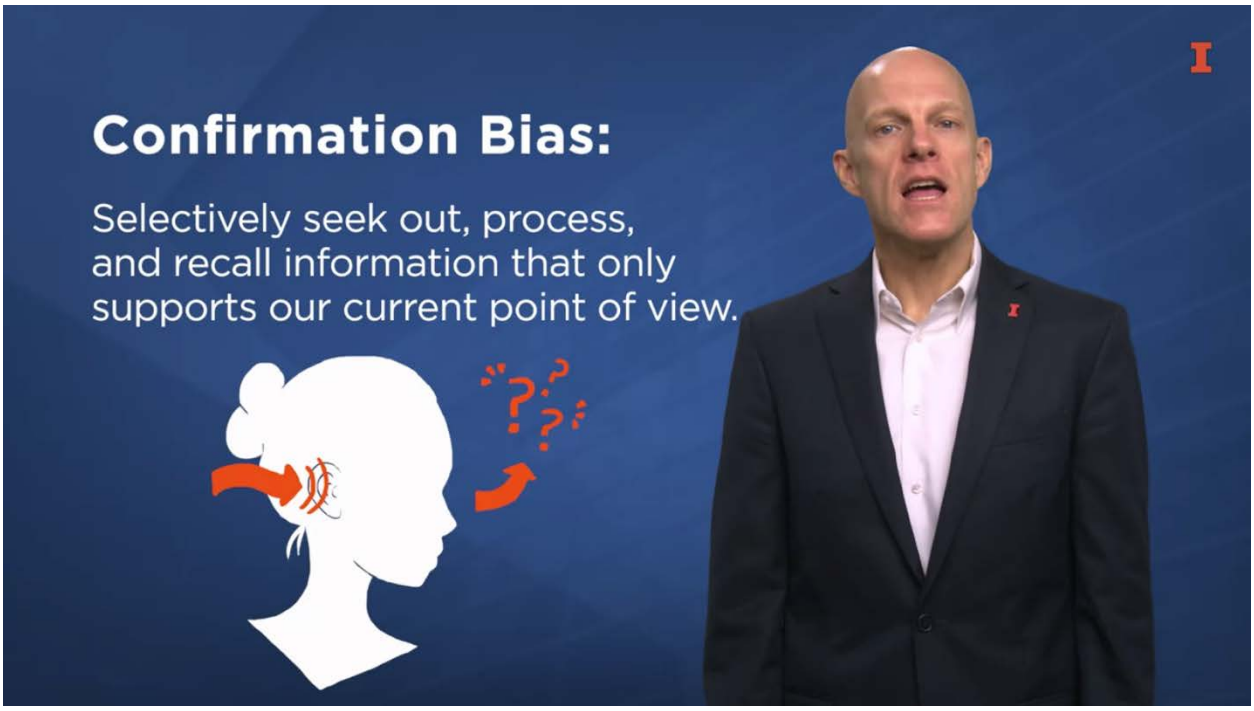
Next, the availability bias refers to the tendency to think that the likelihood that something will occur is based on how easily it comes to mind.

For instance, we may estimate the likelihood that there will be a significant downturn in the economy because it's easier to think of depressionary times rather than booming times next.

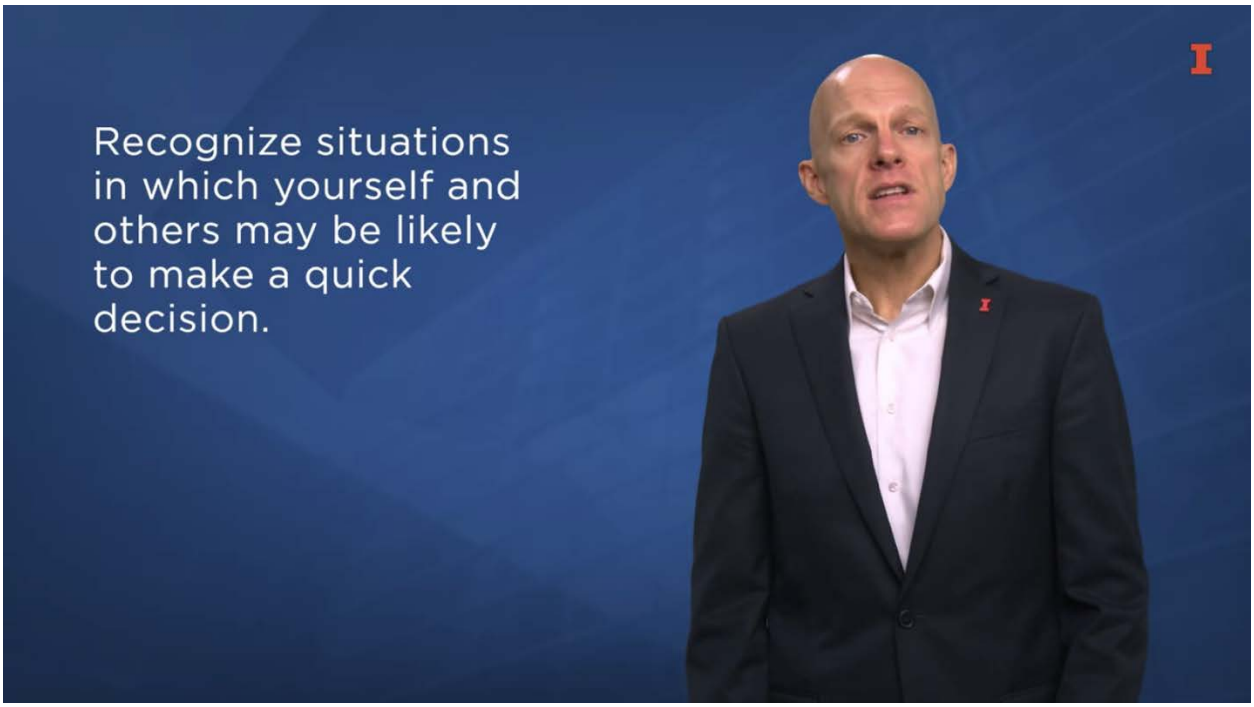


Next, the information bias assumes that more information will always lead to a better decision, even if the information is irrelevant. This type of bias is especially important in our day when we have the ability to access so much data. It's even easier to fall prey to

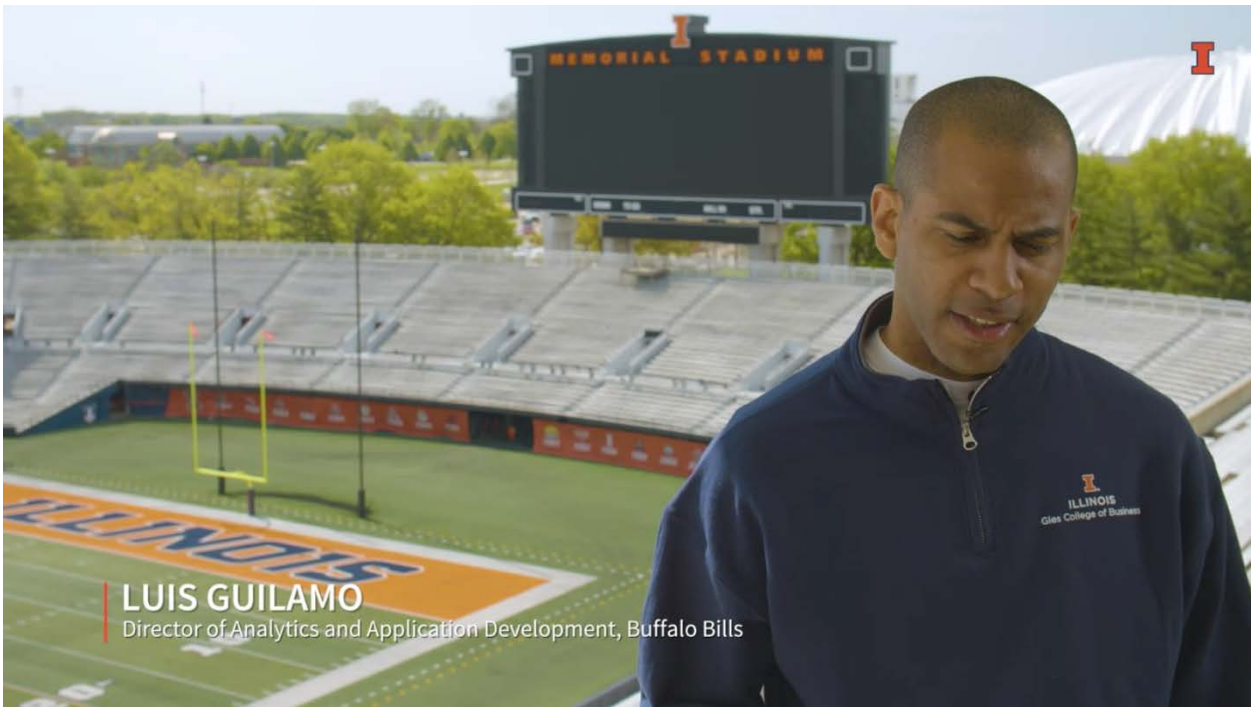
this bias when we hear the term big data repeated so often. Sometimes more data can be detrimental because it distracts us from the main issues or it causes us to get sidetracked.



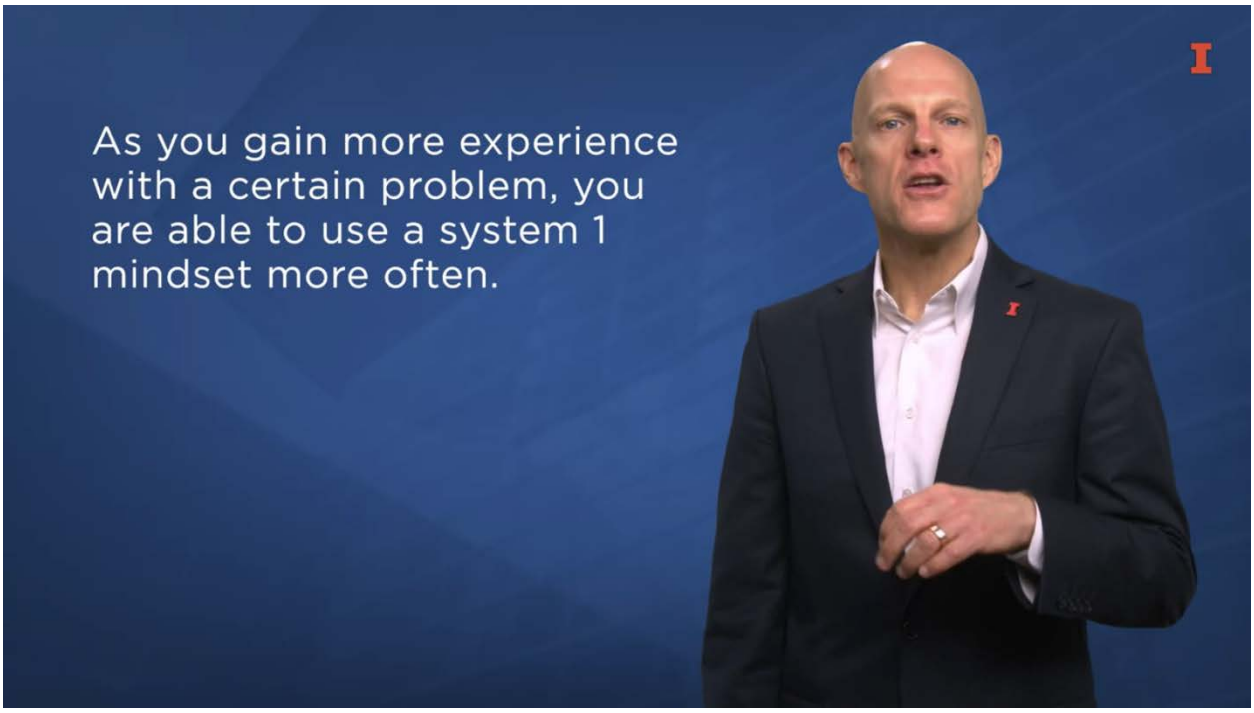
Next, the confirmation bias is to selectively seek out process and recall information that only supports our current point of view. This can lead to greater polarization. It's important to keep this in mind when communicating analytical results with others. Now while there are various biases that I didn't mention, I don't want you to think that biases are always bad. Heuristics can help us and are a wonderful part of how our minds work.



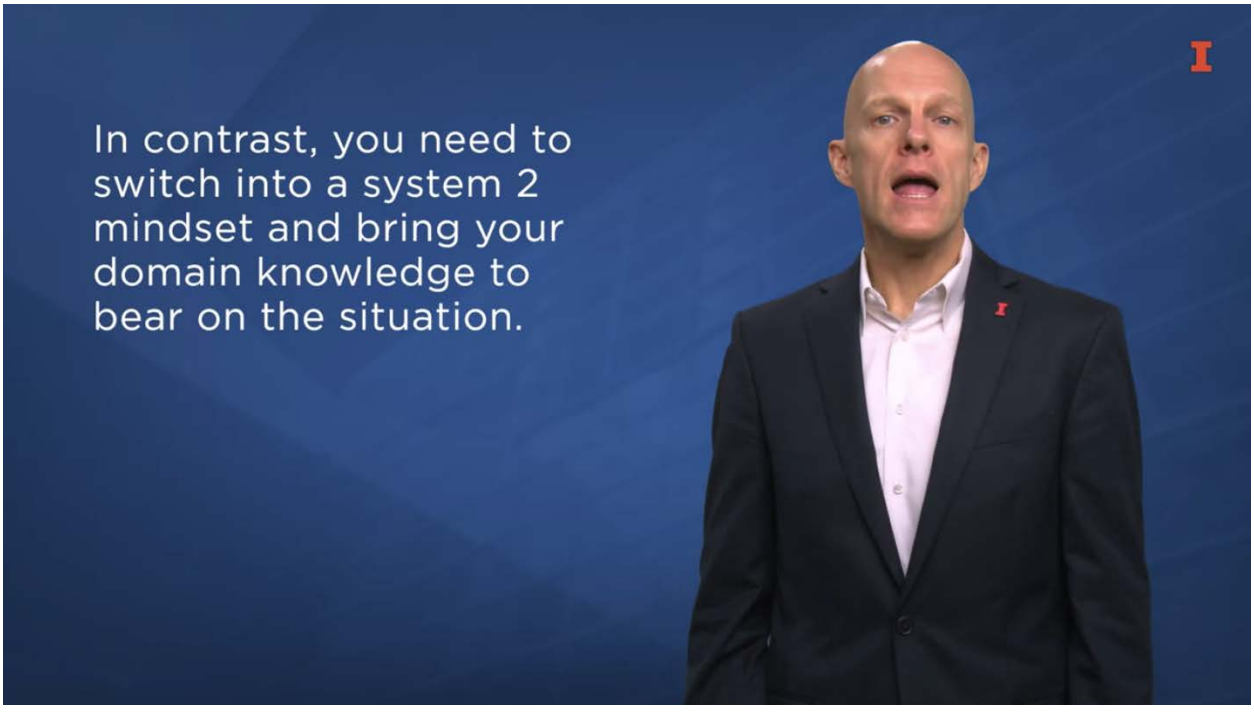
The main idea is to recognize situations in which yourself and others may be likely to make a quick decision. If you have time to consider the decision more carefully and if the decision matters, then you should pause and take a more analytic system two approach. We asked Luis, a university of Illinois alumnus who is also the director of analytics and applications development for the National Football League's Buffalo Bills team, if he ever makes quick and effortless decisions. And how he knows when to switch from a system one mindset to a system two mindset. Let's listen to his response.



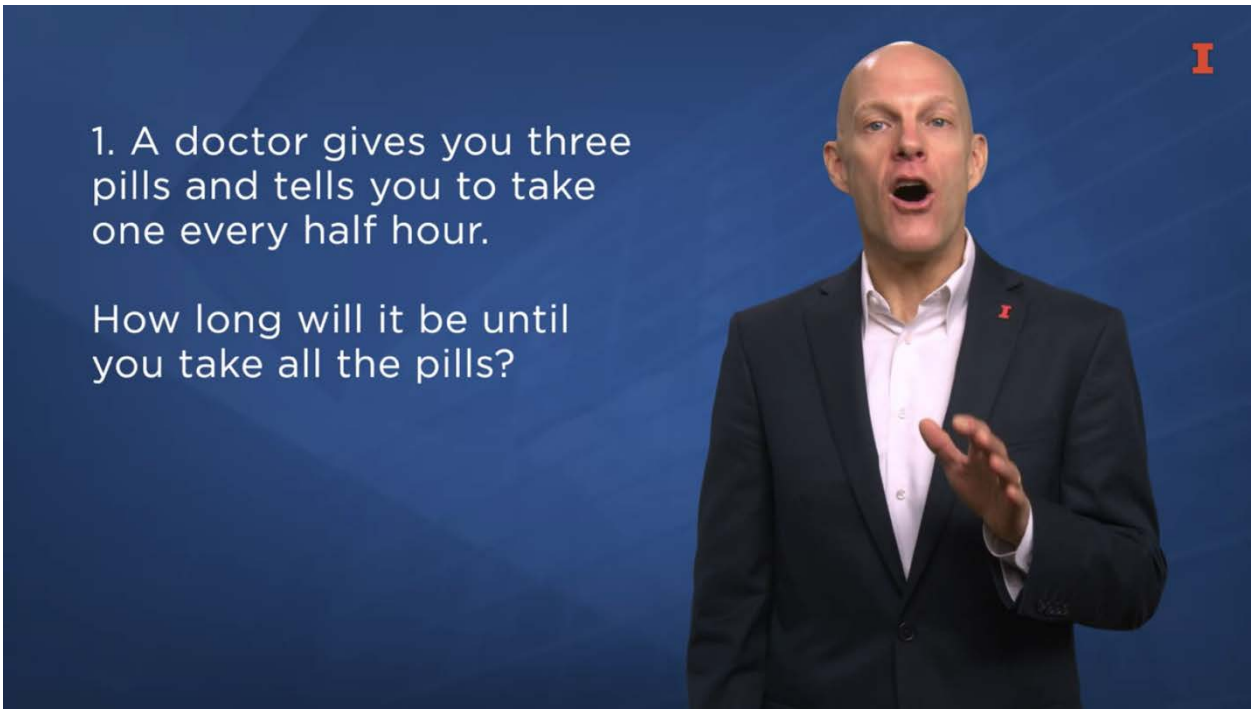
System one quick and effortless and system two kind of slow and methodical. I think they both have have their place. I think the further I go in my career, I utilize more and more system one just based off of experience. I've seen it before, I know what needs to be done. This is the way we need to do it. You execute it. So I think about it as from a process perspective, right? Not necessarily that it takes a short amount of time, which hopefully it does. But the process just knowing what the process that you need to take to answer the question quickly is extremely valuable because time is of the essence. So I utilize system on quite often actually, but that doesn't mean we don't validate the results, right? So from a validation perspective, I typically tell my employees when you think you're 100 correct, okay, check it again. But whenever you see anything new, a new question, new data, if it's a new domain, system two is pretty much the only way to go. You don't want to jump in head first and think you know what you're doing when you don't. That's when you ask questions, you try to understand your methodical about your approach. You ensure that every step of the way that you're doing it the right way and that the results are what you expect to see.



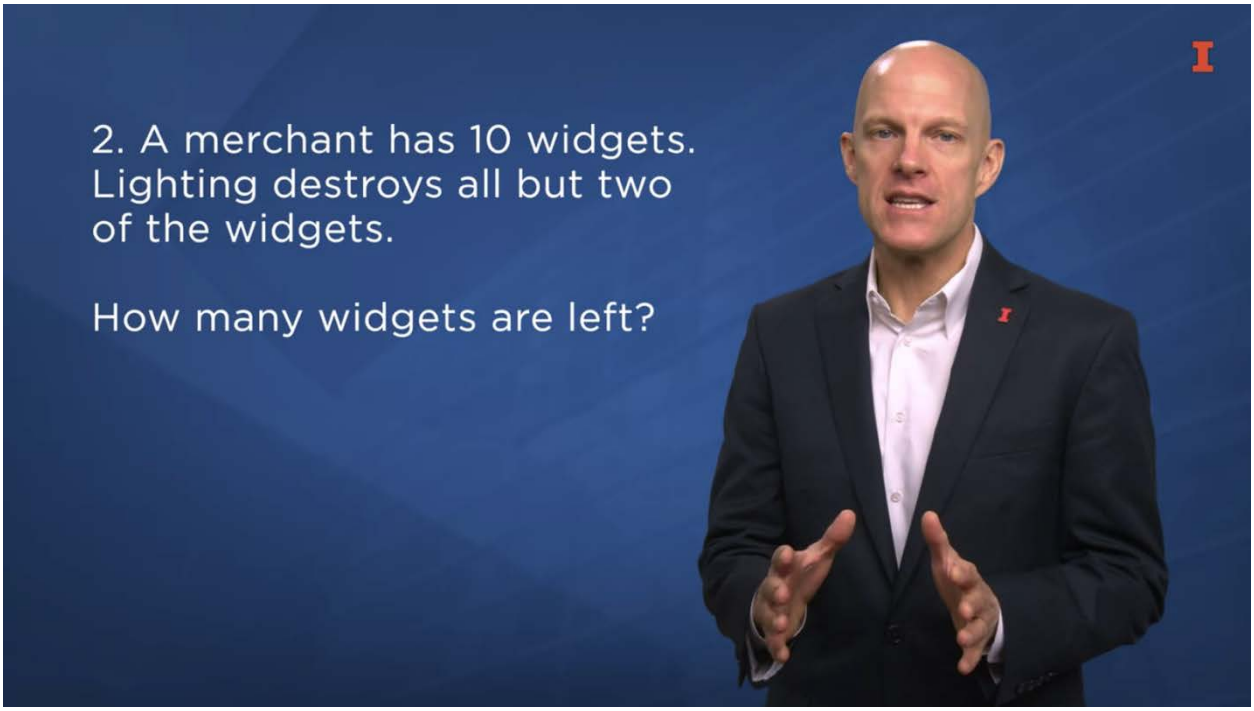
So I really like Luis's insight that as you gain more experience with a certain problem, you're able to use a system one mindset more often.



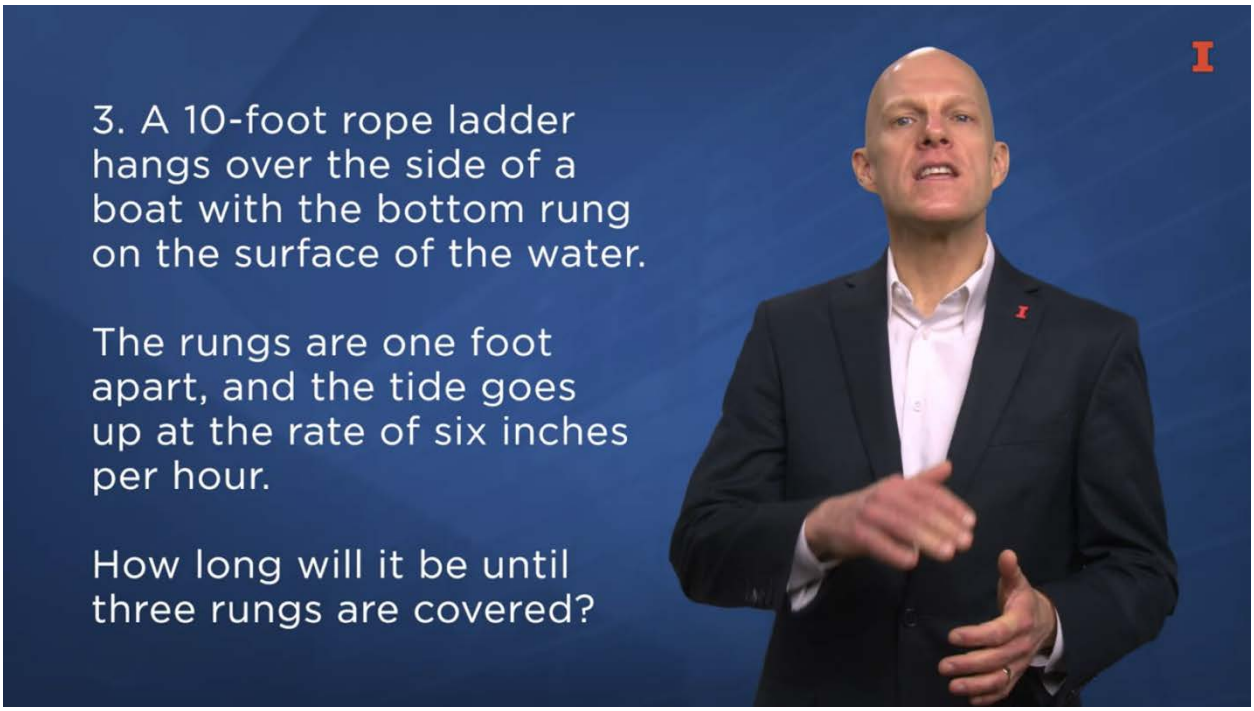
In contrast, when you face a new problem, you should switch into a system two mindset and bring your domain knowledge to bear on the situation. Now that you've learned about system one and system two let's see if you can correctly answer these new questions. Are you ready?



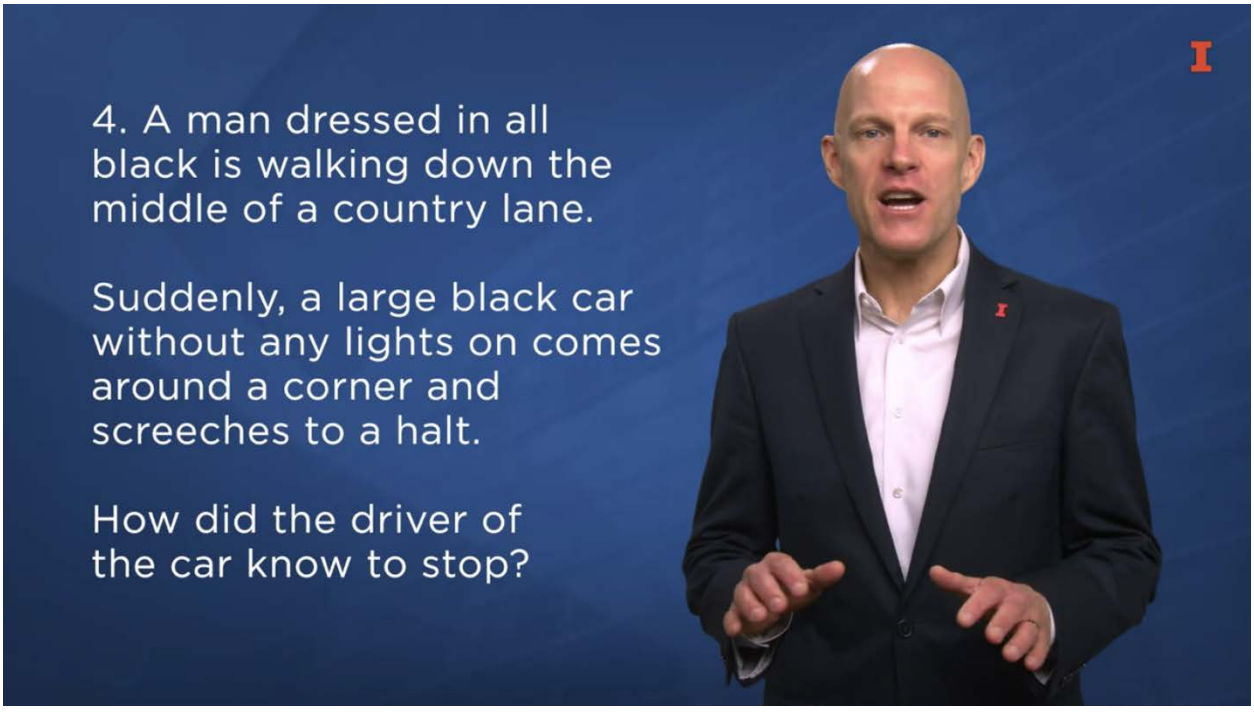
Number one, a doctor gives you three pills and tells you to take one every half hour. How long will it be until you take all of the pills?



Number two, a merchant has 10 widgets. Lightning destroys all but two of the widgets. How many widgets are left?



And number three, a 10-foot rope ladder hangs over the side of the boat with the bottom rung on the surface of the water. The rungs are one foot apart, and the tide goes up at the rate of six inches per hour. How long will it be until three rungs are covered?



And four, a man dressed in all black is walking down the middle of a country lane. Suddenly a large black car, without any lights on, comes around the corner and screeches to a halt. How did the driver of the car know to stop?

References:

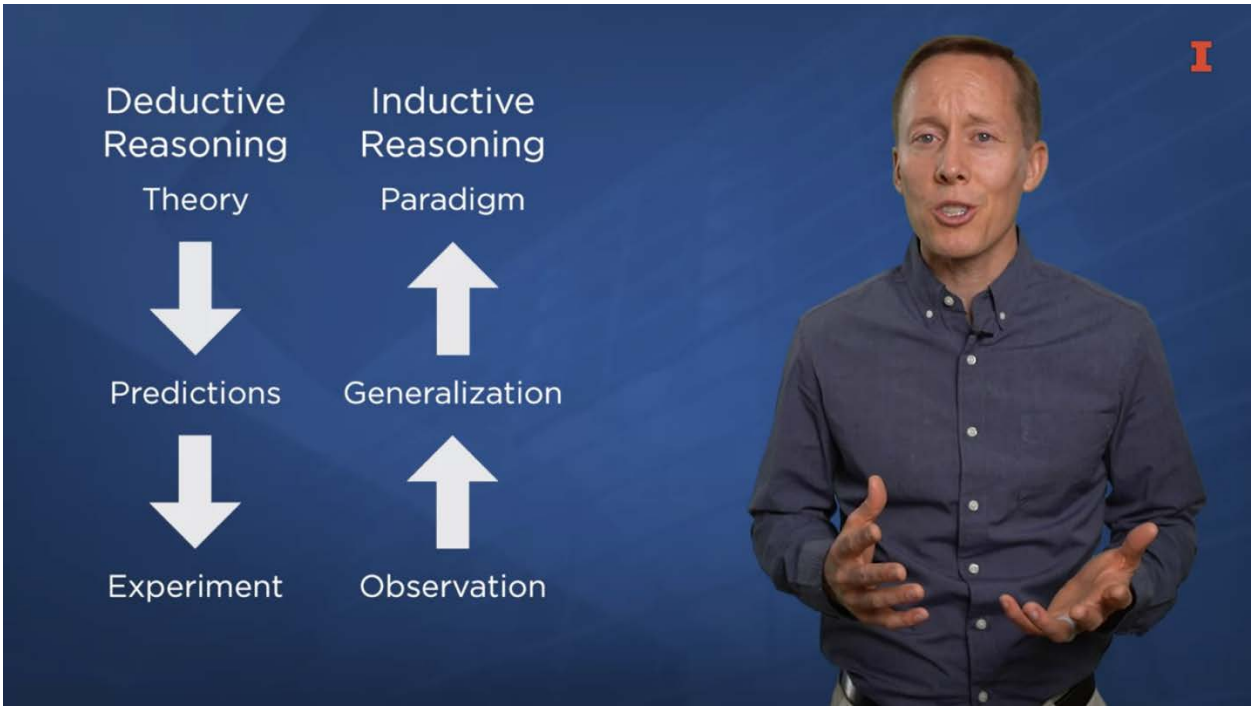
Kahneman, D. (2013). ***Thinking, fast and slow*** [Book image]. Amazon.com. https://images-na.ssl-images-amazon.com/images/I/61fdrEuPJwL.jpg

PublicDomainPictures (2012). ***Arrow business crisis decline depression down*** [Digital image]. Pixabay. https://pixabay.com/illustrations/arrow-business-crisis-decline-15589/

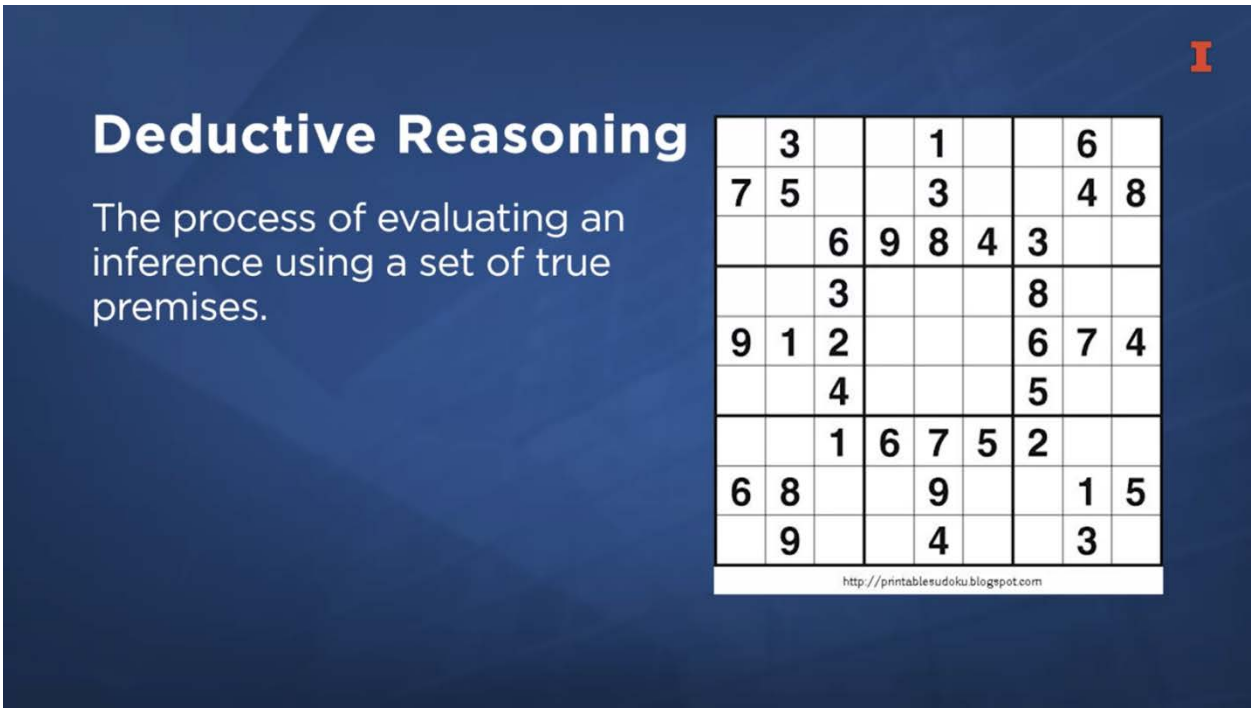
Lesson 1-4: Business Analytics

Lesson 1-4.1 Inductive Versus Deductive Reasoning

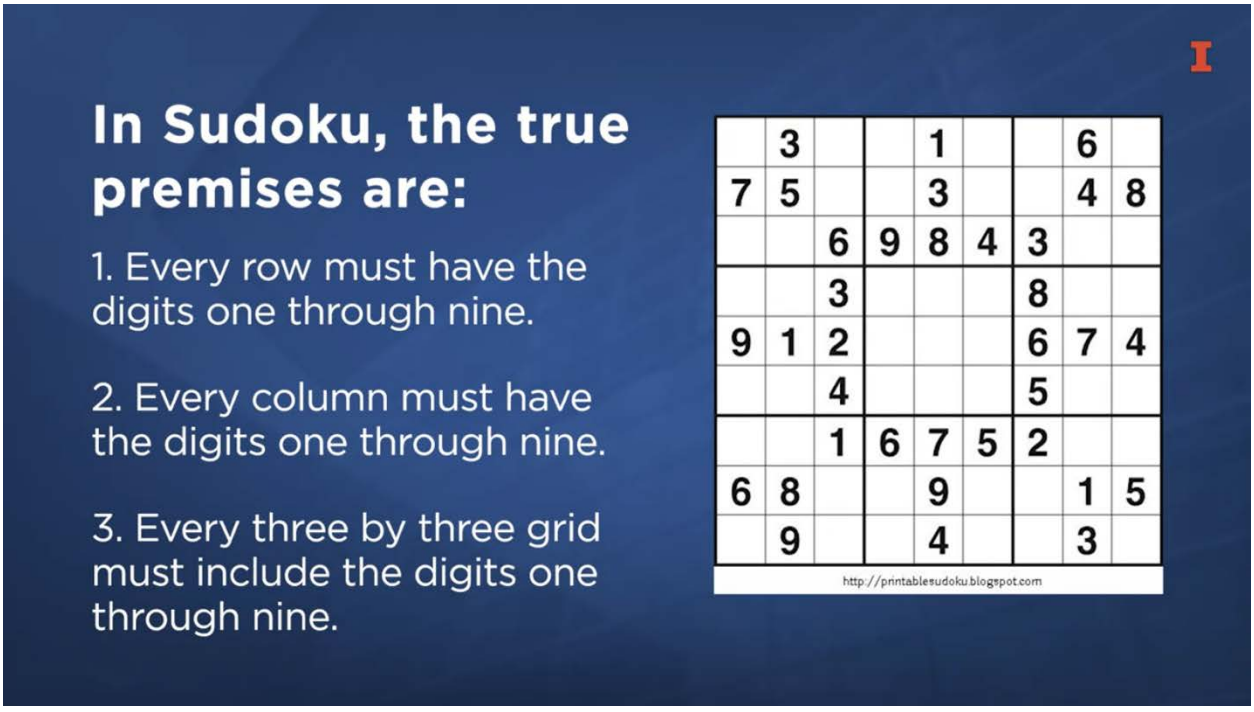
Once you've determined that a decision is worth spending some time on and you shift over into a system to mindset, the mindset in which you're going to think carefully and methodically about the problem. Let's consider some reasoning techniques that will be helpful when dealing with big data to logically arrive at an answer. Two systems of reasoning that are often compared and contrasted are deductive and inductive reasoning.



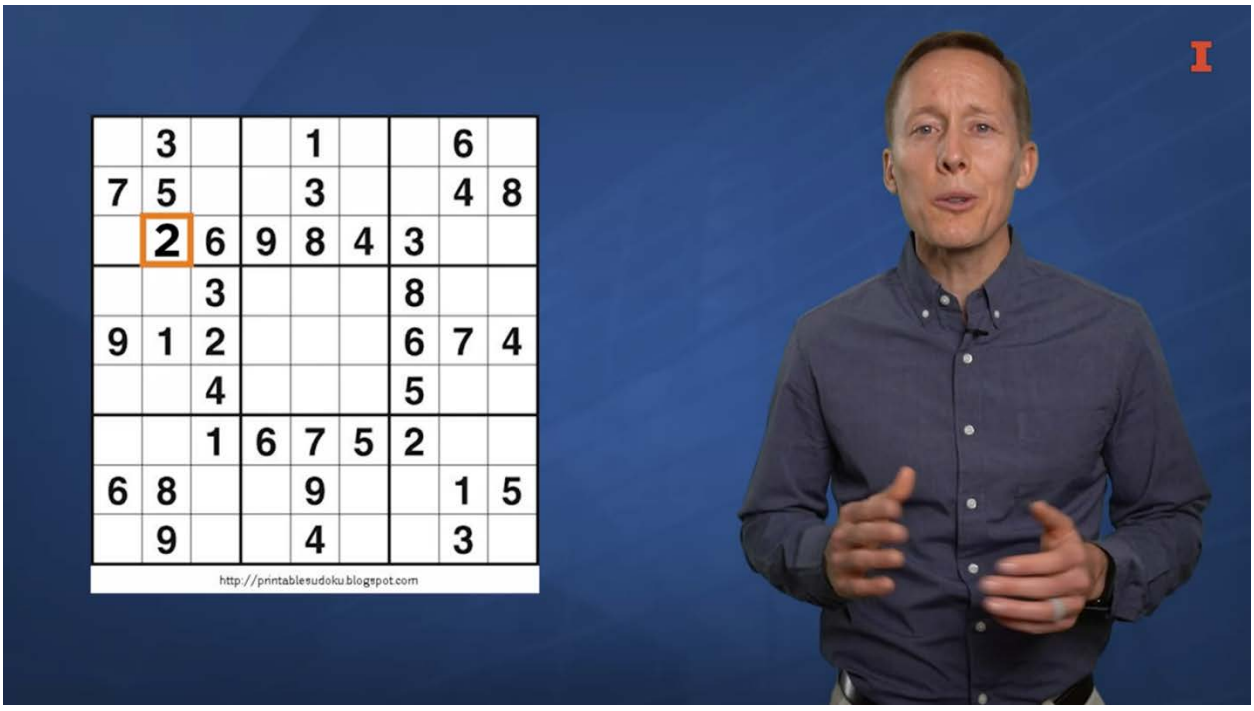
It's often said that deductive reasoning is a top-down approach starting with general rules from which you can make specific inferences, while inductive reasoning is a bottom-up approach that takes an observation and creates a more general rule. This is very helpful, but I think there's a little more to it. I think these terms refer to the way by which you evaluate an inference.



Deductive reasoning is the process of evaluating an inference using a set of true premises. If you've ever played Sudoku, you've been practicing deductive reasoning.

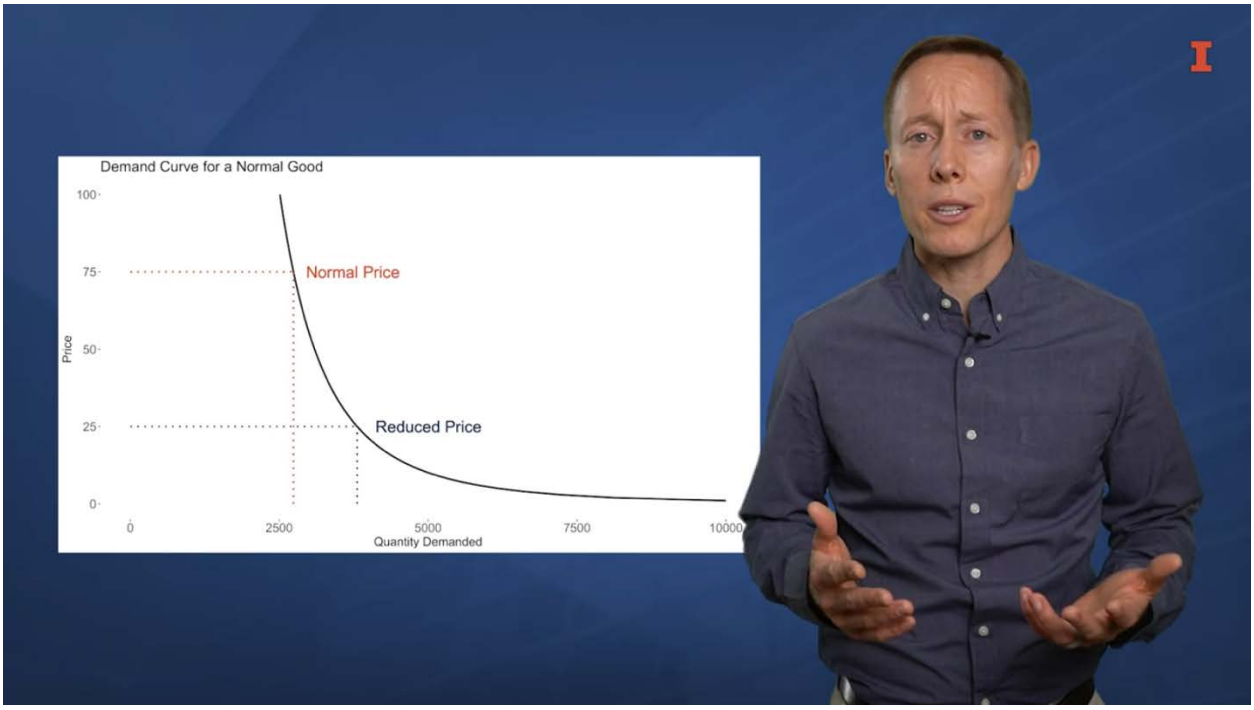


The true premises are: one, every row must have the digits one through nine, two every column must have the digits one through nine, and three every three by three grid must include the digits, one through nine. Using those premises, you can make inferences about the missing numbers.



For instance, let's look at the three by three grid in the top left-hand corner of this Sudoku puzzle and focus on the blank square under the Number 5. By looking at the other numbers in that row, we can rule out the possibility that the missing number is 6, 9, 8, 4, or 3. By looking at the numbers in the column, we can also rule out that the missing number is one or five. Finally, by looking at the other numbers in the three by three grid, we can rule out that the number is seven. That means the number has to be two. We use deductive reasoning in business all the time.

For example, from an accounting perspective, we know that a company's assets are equal to liabilities plus owners' equity. Therefore, if we know the value of assets and liabilities, then we can infer the value of owners' equity. Here's another example from the economics domain.



For most goods, it's generally accepted that as the price increases for a product, the demand decreases. Thus, as you set a production schedule for a product, you'll want to

consider the anticipated sales price. If the sales price increases, then you'll likely want to decrease the scheduled production and vice versa.

In contrast, when evaluating an inference using inductive reasoning, the premises are not intended to be valid. You may observe a pattern and then from that pattern, you infer a general rule recognizing that the pattern from which the rule was derived may be incomplete. One example of inductive reasoning is seen in Bongard problems. In these problems, the six squares on the left show an unknown rule that is not followed by the six squares on the right. Here's an example of one. Can you infer the rule? Feel free to pause the video if you want to spend time to figure it out. The road that is being followed by the six squares on the left but not by the ones on the right, is that four of the five shapes must be circles.

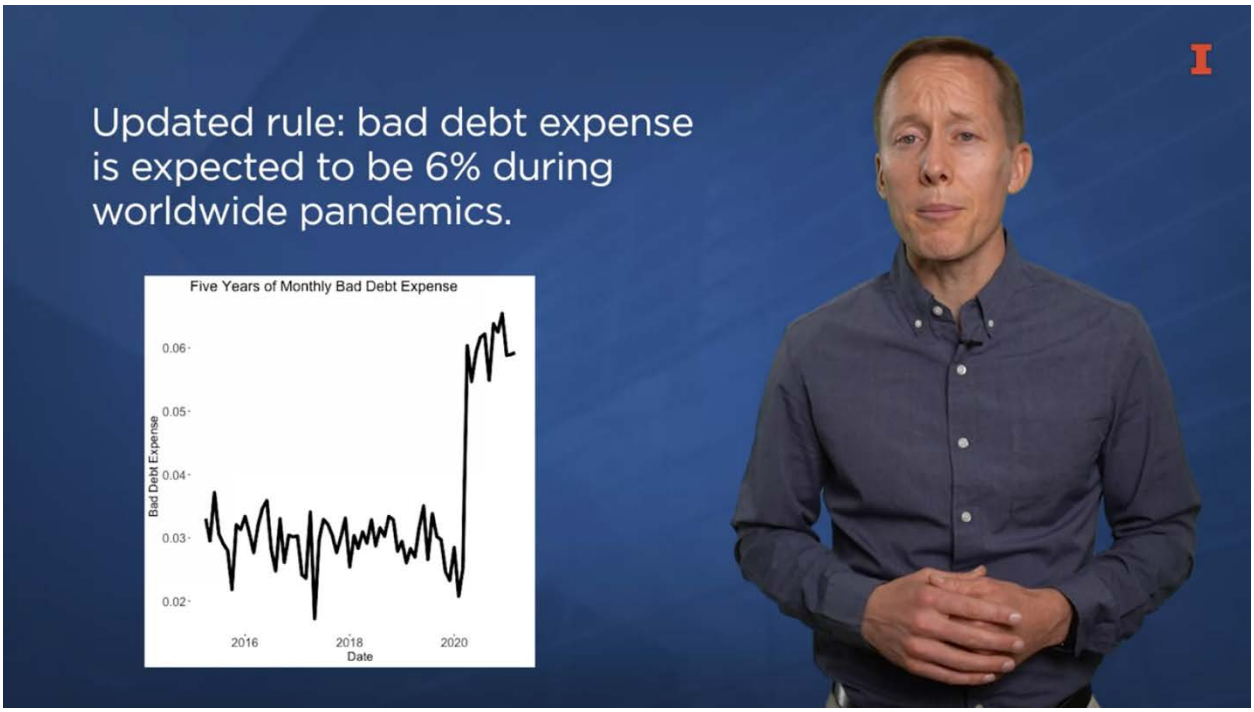
I'm standing in this barren, dry desert climate next to this goblin's shaped rock formation. What is now a desert, geologists believe, was once oceanfront property. The way geologists come to that conclusion is probably a good example of how inductive reasoning works. They could have seen the lines on this rock here and come up with the notion that what caused these lines is the result of water ebbing and flowing through the area. Now, this one piece of evidence isn't enough to provide conclusive support that this was oceanfront property but when they see this pattern repeated over and over dozens of times over here, it supports that hypothesis. Until they see evidence to suggest that this wasn't oceanfront property or until a better hypothesis presents itself to explain this pattern on the rocks, scientists are going to continue to believe that this was oceanfront property and that belief eventually becomes a generally accepted premise upon which additional hypotheses can be built.

Inductive reasoning is often used in business. One example of this is in calculating the bad debt expense. Since bad debt expense is an estimate of the amount of accounts receivable generated during a period that will never be collected, it's often inferred from observing historical collection patterns. Specifically, if we observe from historical data that for each month during the last five years, an average of three percent of accounts receivable were never collected, then we may establish a general rule that each month the bad debt expense will amount to three percent of monthly credit sales.

Both deductive and inductive reasoning are useful in analytically searching for answers to a problem. You may start with a premise that is believed to be true and then gather data to support that it's true. As you gather data to verify that it's true, you may find some evidence to suggest that it's not always true. In that case, you may make inferences from that pattern to generate a modified or complementary premise.

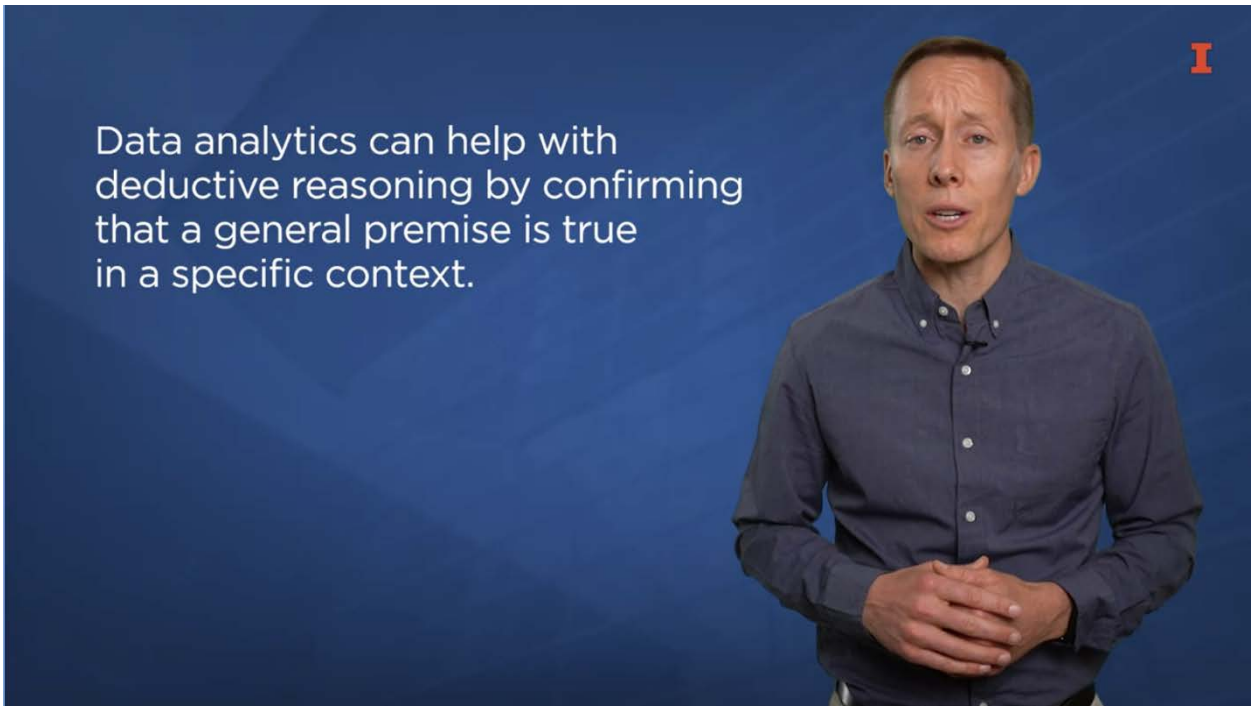
Here's an example. Let's suppose that you know the sales revenue is better than expected. You hypothesize that it was caused by paying for advertising from social media micro-influencers. As you investigate, you find that likes and comments on social media have indeed increased. However, you also discover that your marketing manager has been paying social media channels to promote the posts. Now, you have to try and figure out the impact of each individual action. Passing out the impact of each individual action is a type of problem that can be addressed by data analytic techniques.

It's important to recognize that data can confirm a premise, but it can never fully prove it. A popular example of this is the Black Swan problem. This problem is illustrated by a situation in which a general rule is made that all swans are white. However, it may be the case that black swans exist, but they just haven't been observed. Thus, the real way inductive reasoning is proven true is if it continually fails to be proven false.

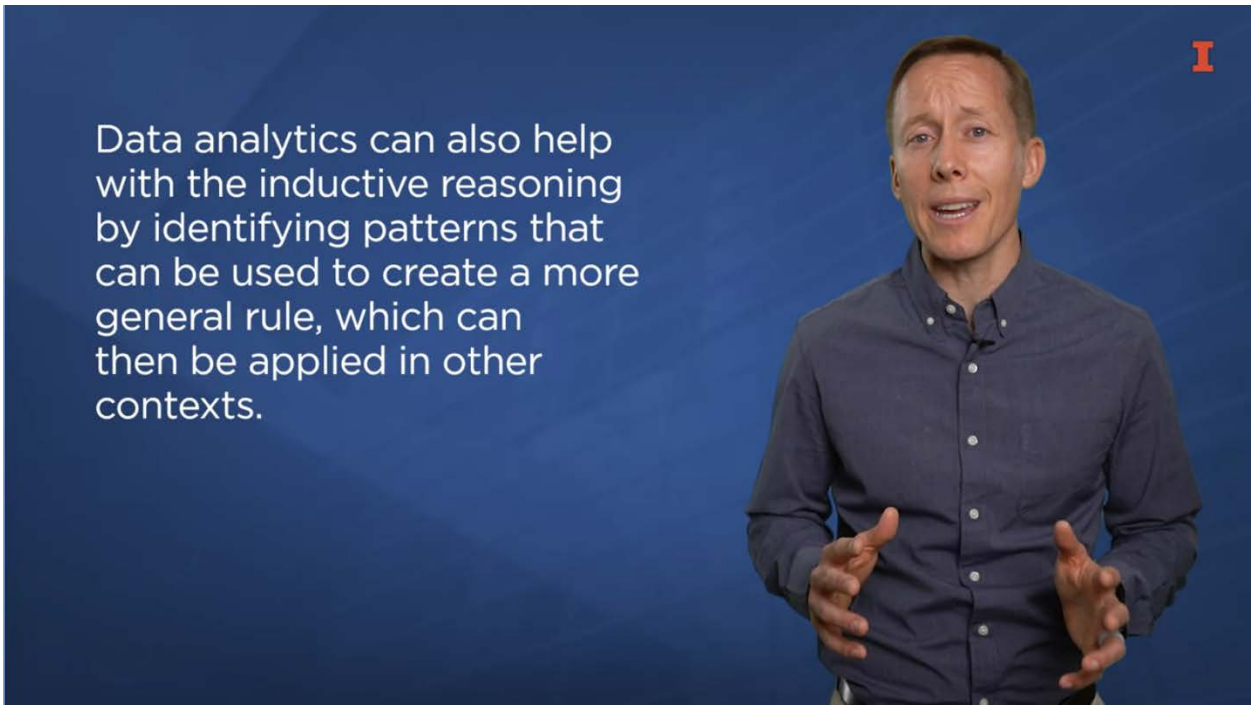


For instance, the premise that bad debt expense is three percent of monthly credit sales

may eventually be proven to be untrue if there's a month when bad debt expense is six percent of credit sales. When that's the case, a careful analysis of what led to the doubling of bad debt expense may result in a modified version of the rule to indicate its limitations when it is not true. In a business sense, it would be important to identify those limitations so that they can be prevented in the future.



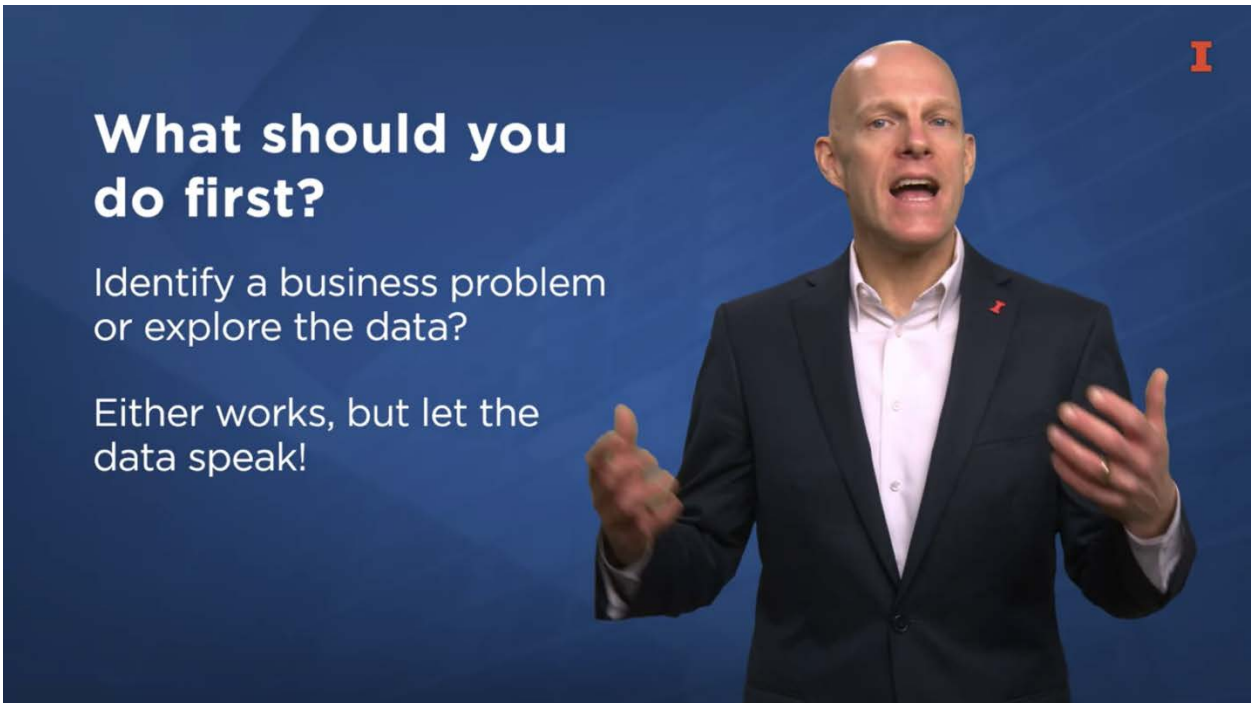
The takeaway is that data analytics can help with deductive reasoning by confirming that a general premise is true in a specific context.



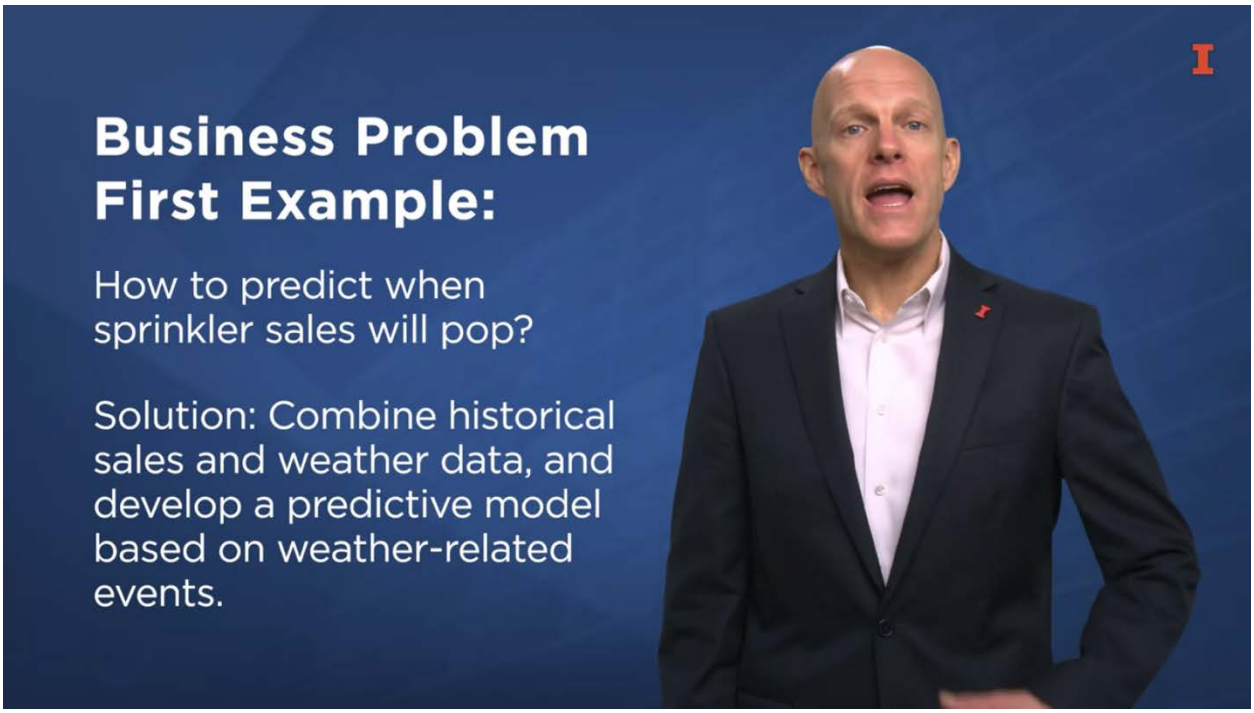
Data analytics can also help with inductive reasoning by identifying patterns that can be used to create a more general rule which can then be applied in other contexts. Finally, it's important to remember that intuition is still important because it's not likely for logical reasoning and data analytics to completely close the gap between what's known and what's unknown.

Lesson 1-4.2 Let the Data Speak

In a general sense deductive reasoning, and inductive reasoning are both valid approaches to applying data analytic tools to business problems. In this lesson, we want to look at a related but distinct dichotomy for creating data analytics solutions to business problems.

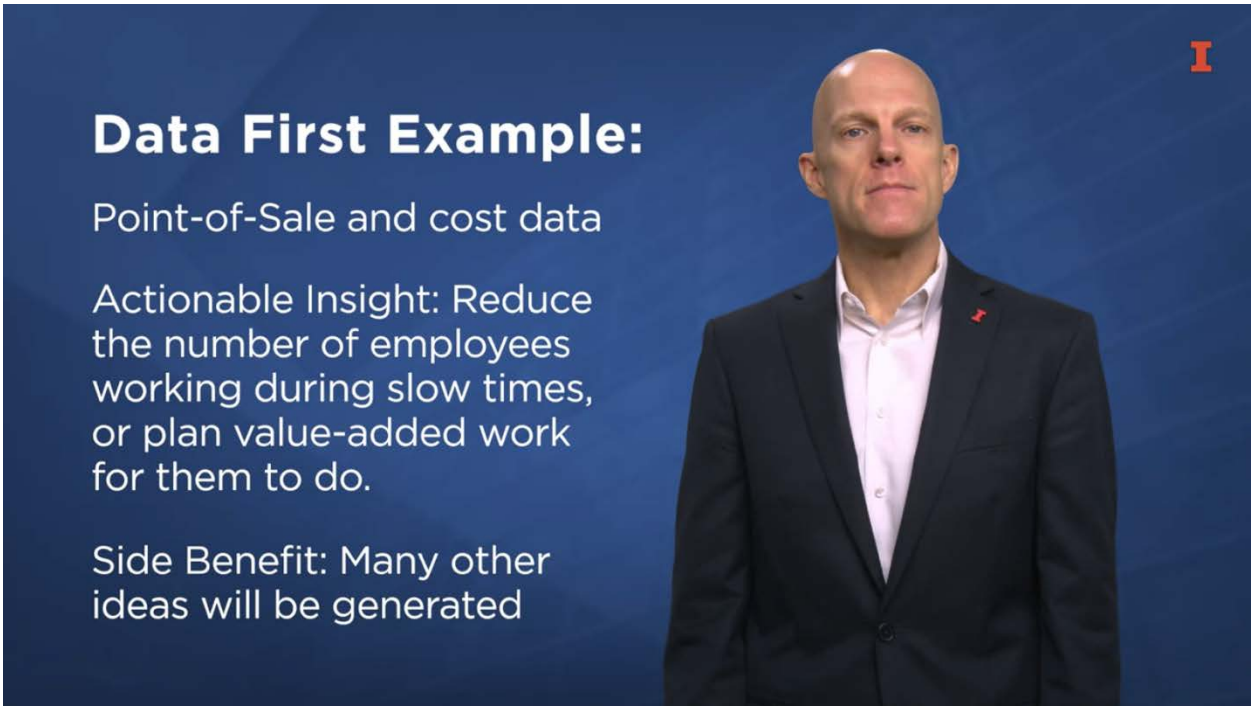


That dichotomy is whether you should identify a business problem before pursuing a data analytics solution or whether you should apply data analytic approaches to a data set and then find a business problem that it can solve. Our hope is that by the end of this lesson, you recognize that either approach can be appropriate. But the most important takeaway is that you should let the data speak.



Let's first consider a situation in which we start with a business problem and then pursue a data analytics solution. There's a large sprinkler manufacturing company that sells much of their product through stores like Lowe's and Home Depot. Sprinkler sales are very seasonal and they tend to spike up a lot during the spring months, right when people are turning their sprinkler systems back on and recognize that sprinklers have broken during the winter months. Those are also times when commercial construction and landscaping businesses start doing more yard work. The problem for this type of company is trying to make sure that sufficient sprinkler inventory is stocked at the Lowe's and Home Depot stores before sales spike in the spring or else they can miss out on a large number of sales. It's kind of like a black Friday situation, except that they don't know which weekend black Friday will be. What makes it even more complex is that the black Friday weekend varies from region to region because of the different weather conditions. Now, this company could obtain historical sales and weather data to identify the weather conditions that lead to a spike in sales. Applying a data analytic algorithm. They could then come up with a model for identifying the weekend when sales will spike for each region. For instance, there model may indicate that sales spike the weekend after the first day of the calendar year, that temperatures reach 70° foreign heights, 4 regions in which the annual precipitation is less than 12". Once they have the model, they could then get weather forecast data to plug into a model to create sales predictions. Using these predictions would allow them to improve their ability to supply the right amount of sprinkler product for each region so that they're prepared for the sale spikes. The way I described this scenario is an example of when a business problem comes first and then a predictive data analytic approach leads to the potential solution.

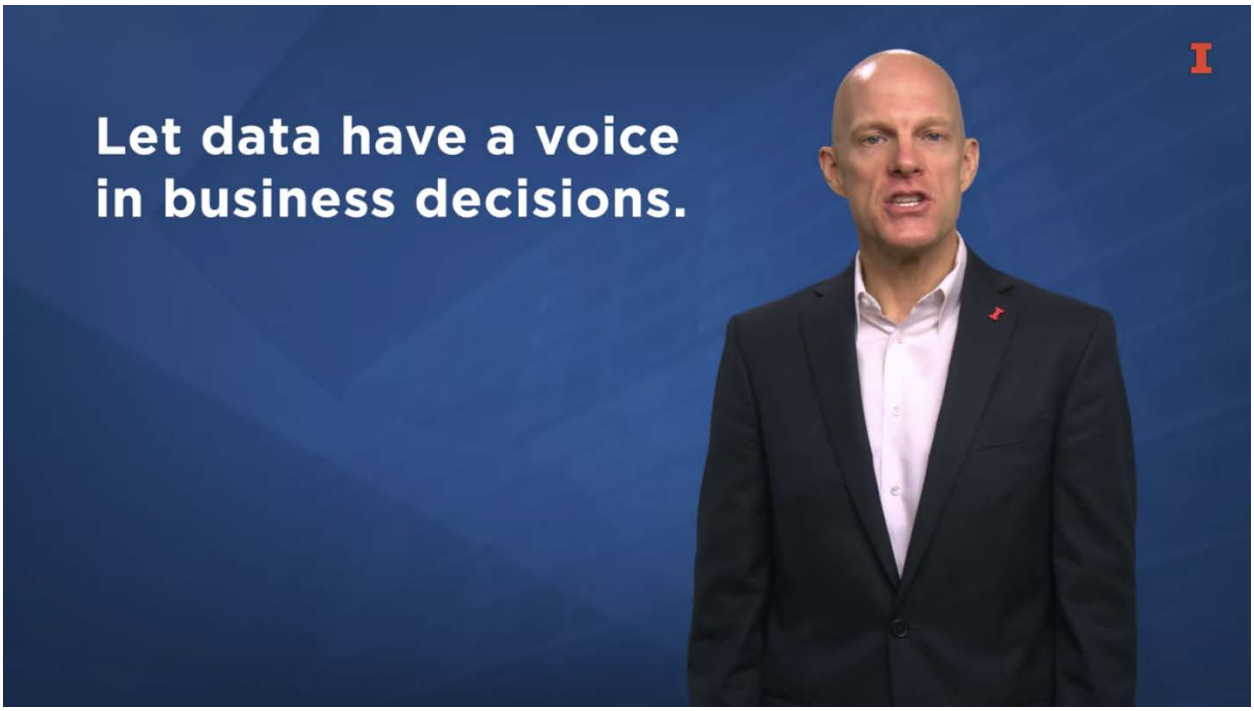
Now, let's consider the opposite approach when the data comes first and then is applied to a business problem.



Specifically, let's consider a small restaurant that has a mask point of sale data. Over several years. Let's assume that the owners of this company are doing fine. But they want to take their profitability to that next level by finding ways to reduce costs. Recognizing data as a valuable raw material. They could combine the point of sale data with cost data and apply visualizations and algorithms to identify their largest cost and other unnecessary costs. They may identify patterns in periods of the day and days of the week when they're overstaffed. Once they uncover those patterns, the owners could then plan more effectively by not scheduling workers to come in during those times or identify new value added activities that the employees could engage in during those times. Based on my experience, I am confident that as a result of mining the data, they would come up with many other ideas for how to improve performance. This is an example of when the data comes first, but provides insight about a business problem that can be solved. So we hope these two examples help illustrate the value of the business problem first versus data first dichotomy.

One thing that we haven't yet highlighted those the importance of considering the results of the data. This can be difficult if the data goes against strongly held beliefs. Booking.com is an online travel agency that gives all of their employees the ability to experiment with different versions of their website design. They run thousands of experiments each year and use data analytics to evaluate the effectiveness of the various website designs. Because anyone can run experiments, all employees, including executives, have to be prepared to back up their opinions with data analytic results and also be prepared to recognize that their opinions may be wrong.

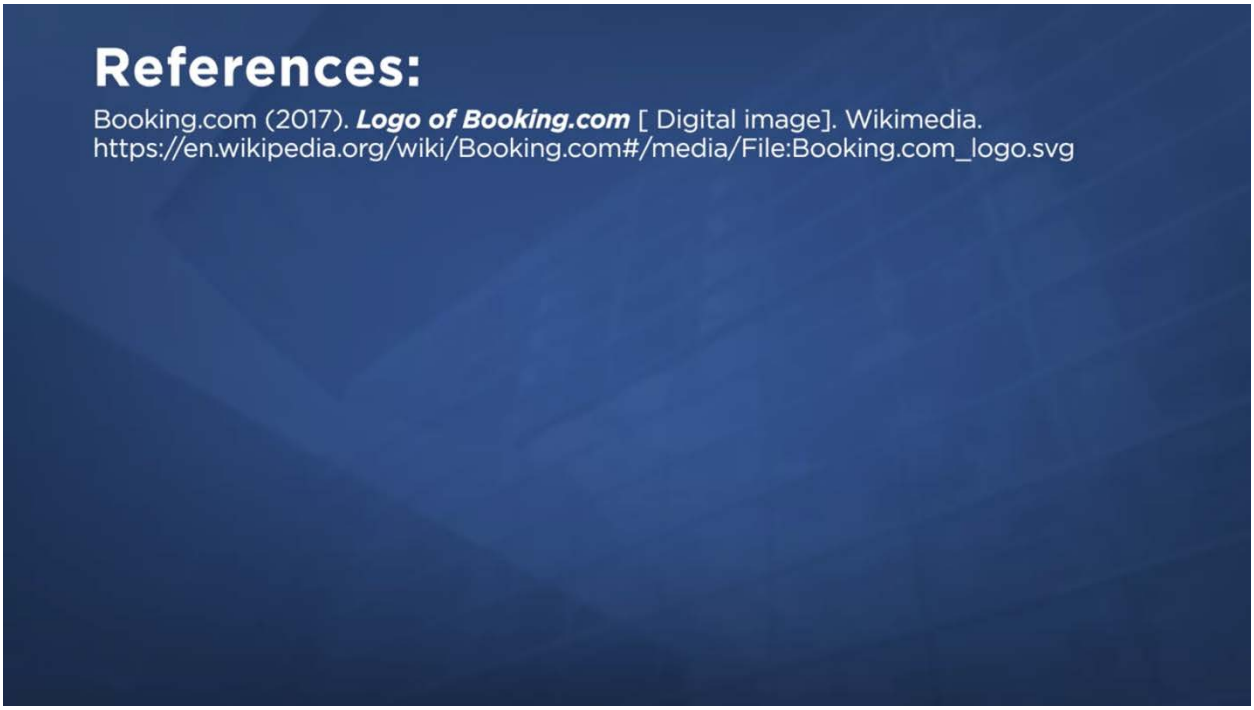
Emitting and believing that your opinion is wrong can be difficult for all of us. Fortunately for booking.com because they have established an experimental culture of letting the data speak because they run so many experiments, it can be easier for employees to admit that their opinions are wrong. All right to be clear. We're not trying to say that there's no room for opinion, professional judgment, heuristics and intuition.



What we're trying to emphasize is that because we're in an age in which we have so much data and so many data analytic tools, we should make sure to let data have a voice in business decisions.

In conclusion, whether you're starting with a business analytic problem or business analytics data, the phrase in my opinion should be heard a lot less often than the phrase the data says.

The phrase in my opinion it may be perfectly fine if data cannot be applied to a decision. But for most problems, we suggest that you consider how the phrase can be followed up with, and here's what the data says.



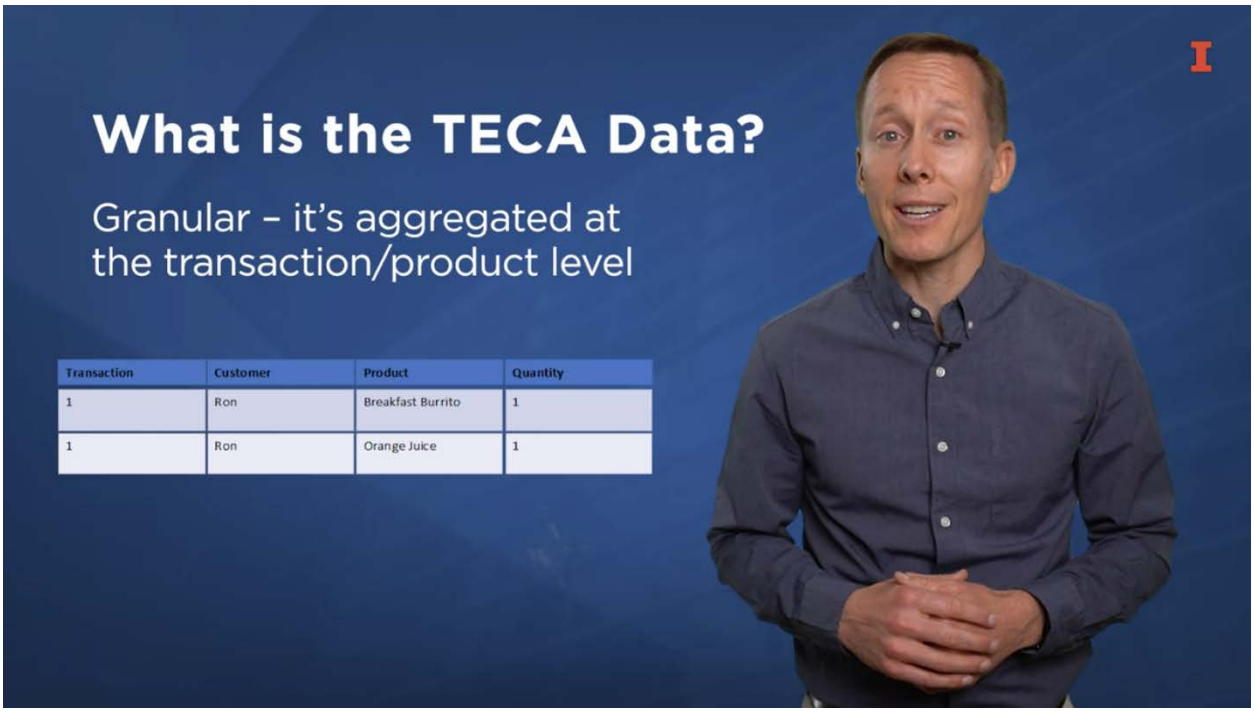
Lesson 1-5: Your Mind as a Tool

Lesson 1-5.1 Introduction to the Course Data Set: TECA

In this course, we're going to practice applying data analytic tools on some really cool point of sale data. We will use this data set with several different business analytic software platforms. In this lesson, we want to walk through that data set so that you can mentally prepare for how to approach it.

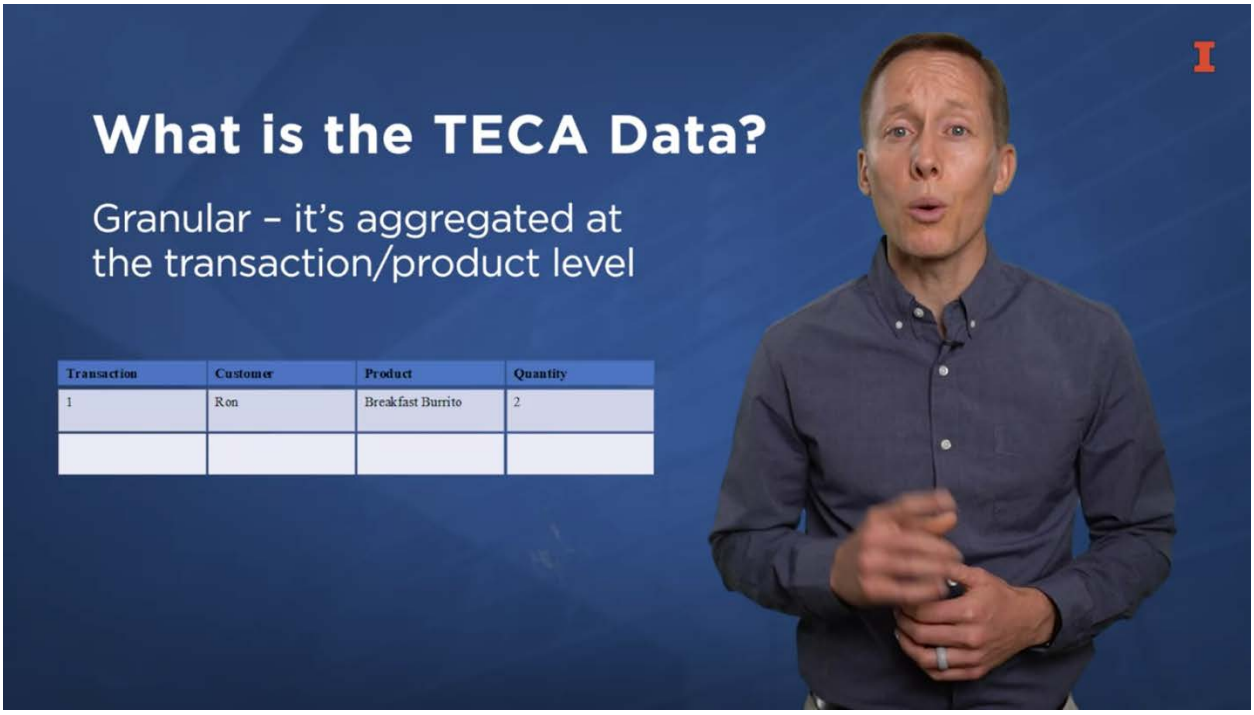
We will refer to the data set that we will be using as the TECA data set because it's from a company named TECA that owns over 150 convenience stores and gas stations throughout the middle of the United States. These stores sell typical convenience store items; gas, candy, soda, chips, lottery tickets, and so on.

While this TECA data set is pretty large, especially for Excel standards, it's a small portion of all the data that is generated. We have narrowed it down to a sample of the data from 2017-2019. If we used all of the data, it would consist of hundreds of millions of rows and would be too much for our personal computers to handle.

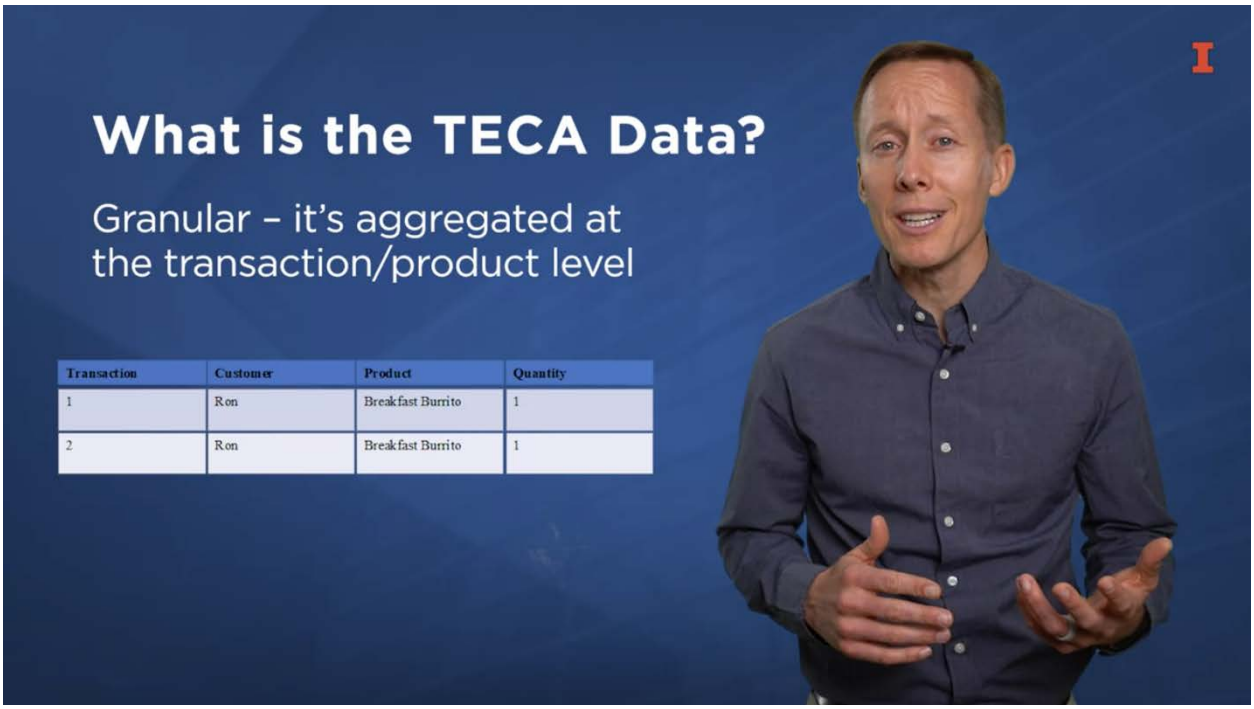


Let's talk about how this TECA data is aggregated. Each row of this data set represents a unique product of a single transaction. If a customer purchased two unique products

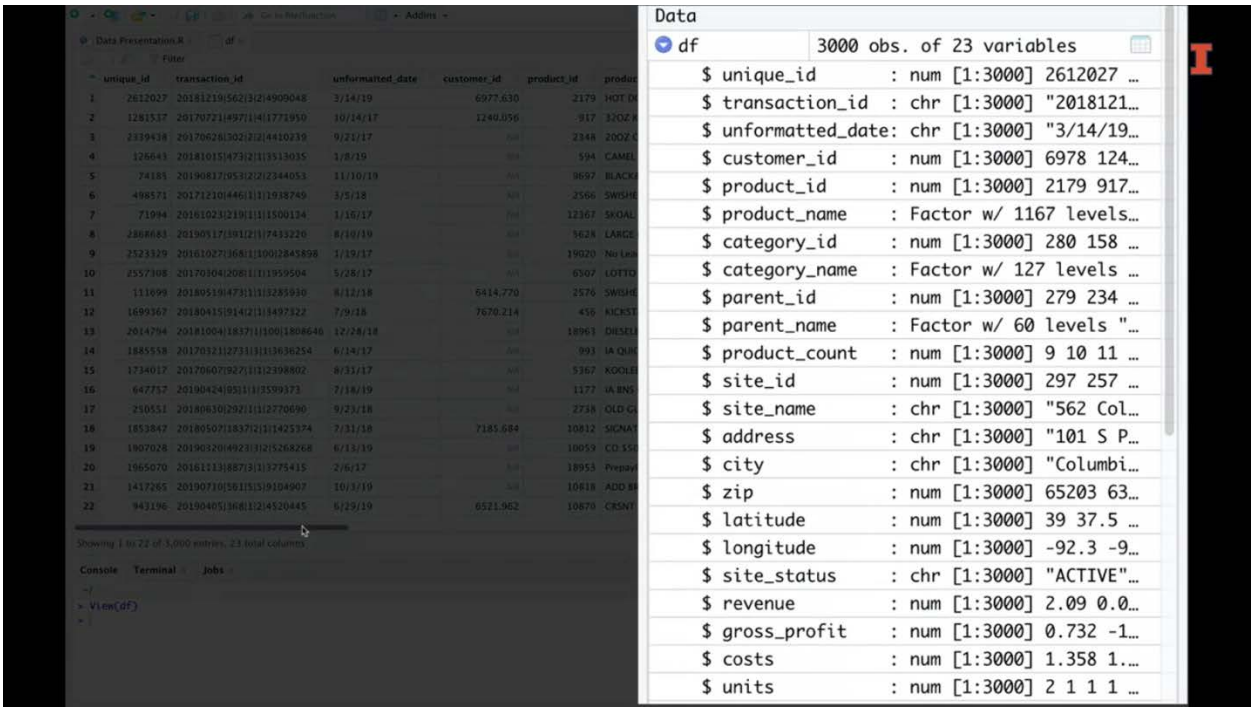
in a single transaction, then there would be two different rows of data for that customer's transaction.



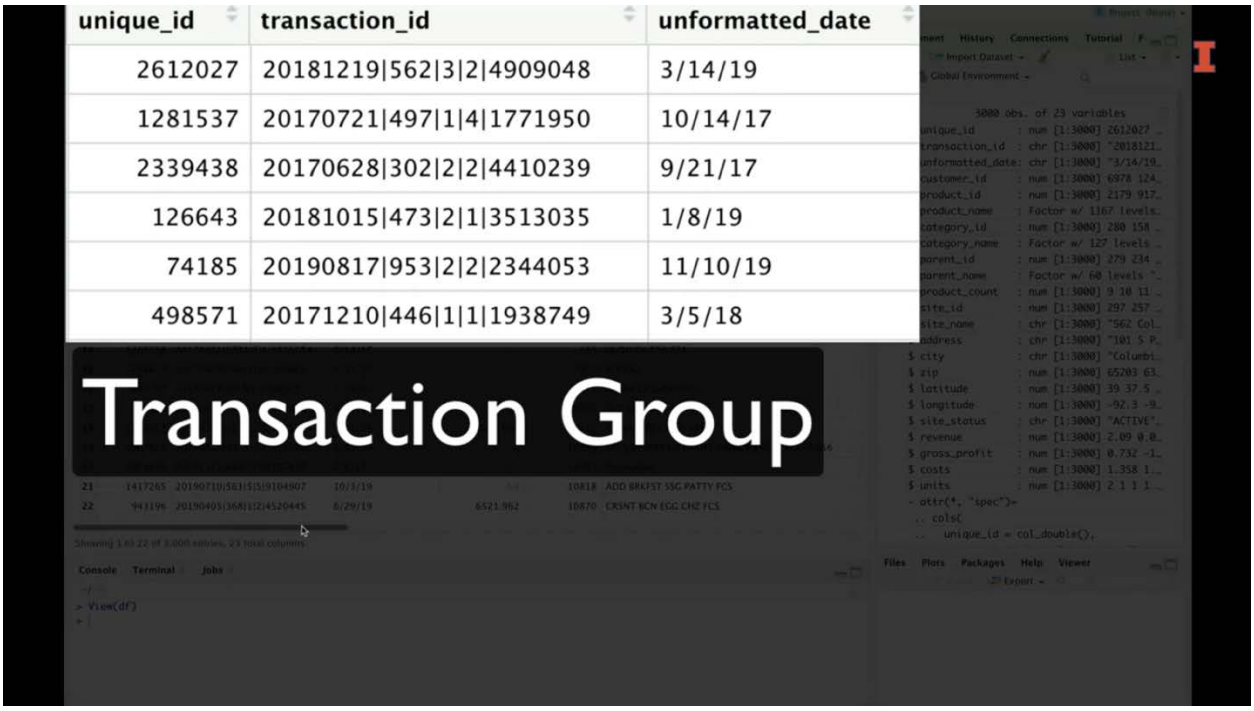
Importantly, if a customer purchased two of the same product, then there would only be one row for that transaction.



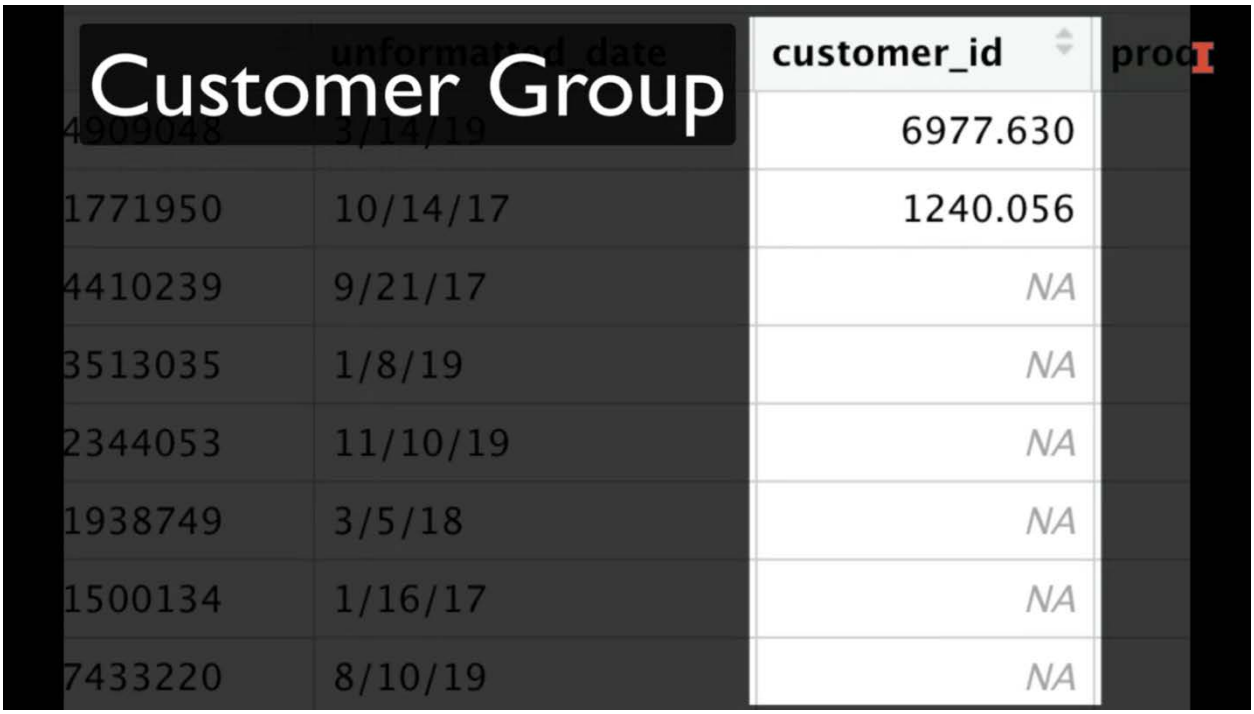
It's also worth noting that a customer could have more than one transaction during a visit. For example, if a customer purchased a single breakfast burrito and then before leaving, decided to go back and get another breakfast burrito for a friend or family member just a few minutes after the first transaction, then there would be a separate row for each transaction. That's what the rows represent. Let's talk about the columns.



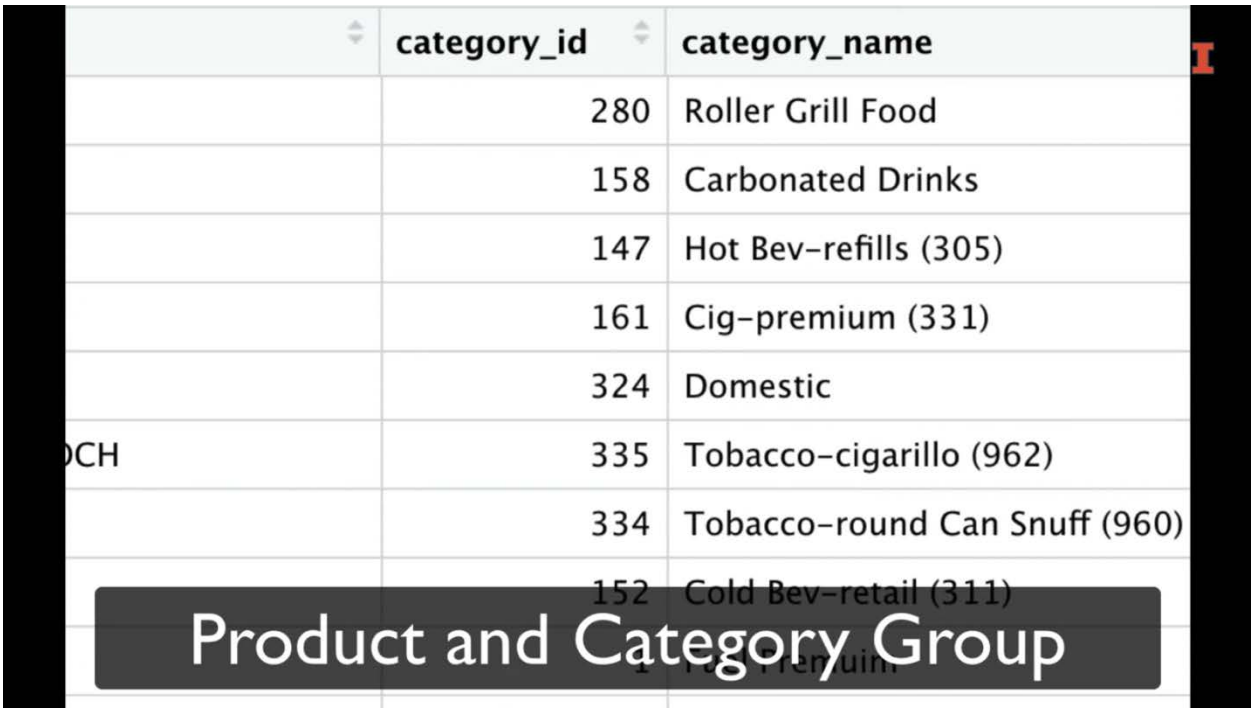
There are dozens of columns in this data set, and they can be grouped into five groups.



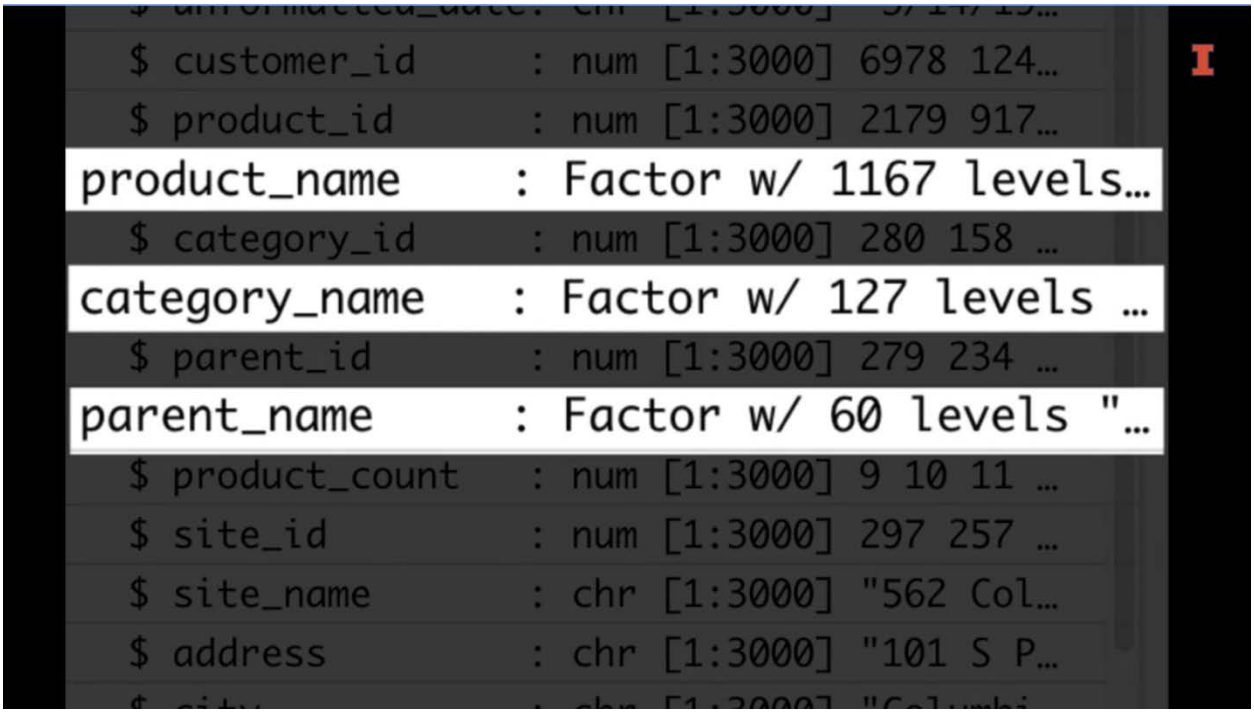
The first group is the transaction group. This group of columns includes a unique identifier for each row, the transaction number, and the date of the transaction.



The second group, the customer group, is really just a single column; customer_id. This unique identifier is only known for customers that are loyalty rewards members, otherwise it is blank.



The third group of columns is the product and category group. This group of columns includes information about the product's name and how it fits into the hierarchy of products.

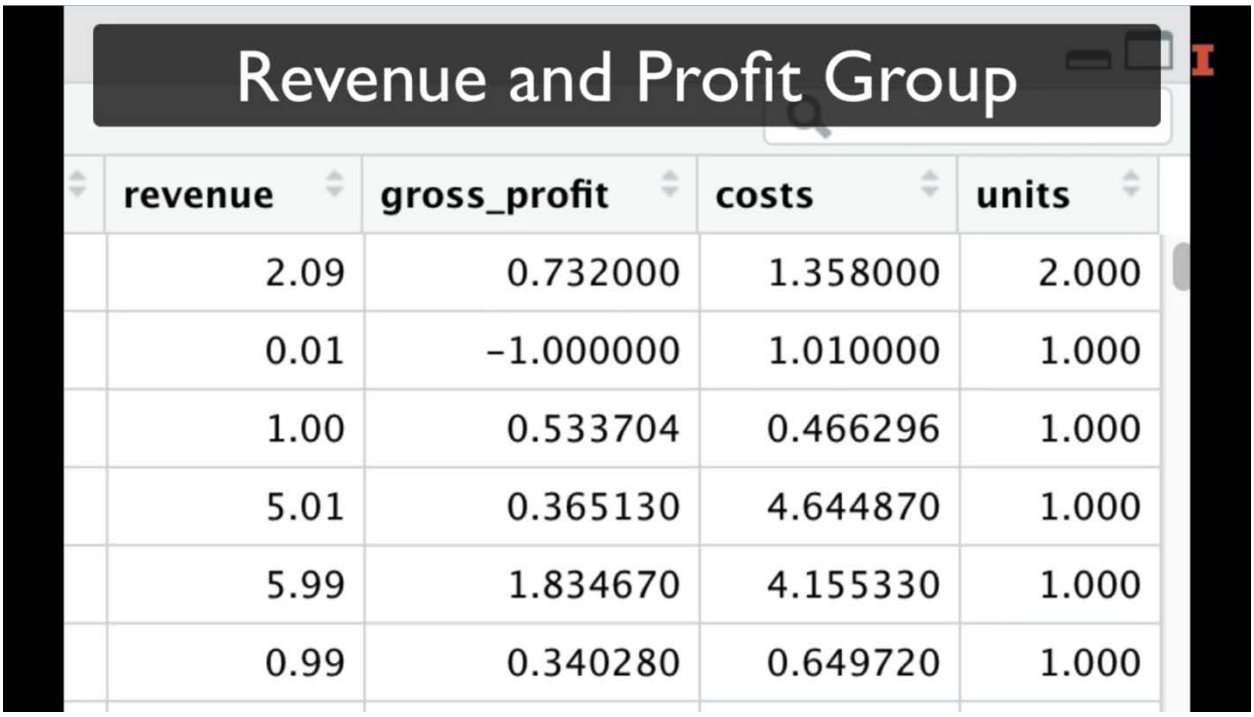


Specifically, all products belong to a category and all categories belong to a parent. Thus, the parent column has the fewest number of distinct values. The category column

has more distinct values than the parent. The product's column has the largest number of distinct values.

site_id	site_name	address
297	562 Columbia	101 S Providence Rd
257	497 Patton	1000 State Route 51
116	302 Parkersburg	602 Coates St
230	473 Hume	101 Main St
459	953 Laramie	152 N 5Th St
207	446 Bessemer	750 Academy Drive

The fourth group of columns is the site group. The columns in this group include information about where the store is located, such as city, state, zip code, and latitude and, longitude coordinates. To protect the actual convenience stores, the coordinates have been changed to match those of a nearby post office. There's also a store ID column.

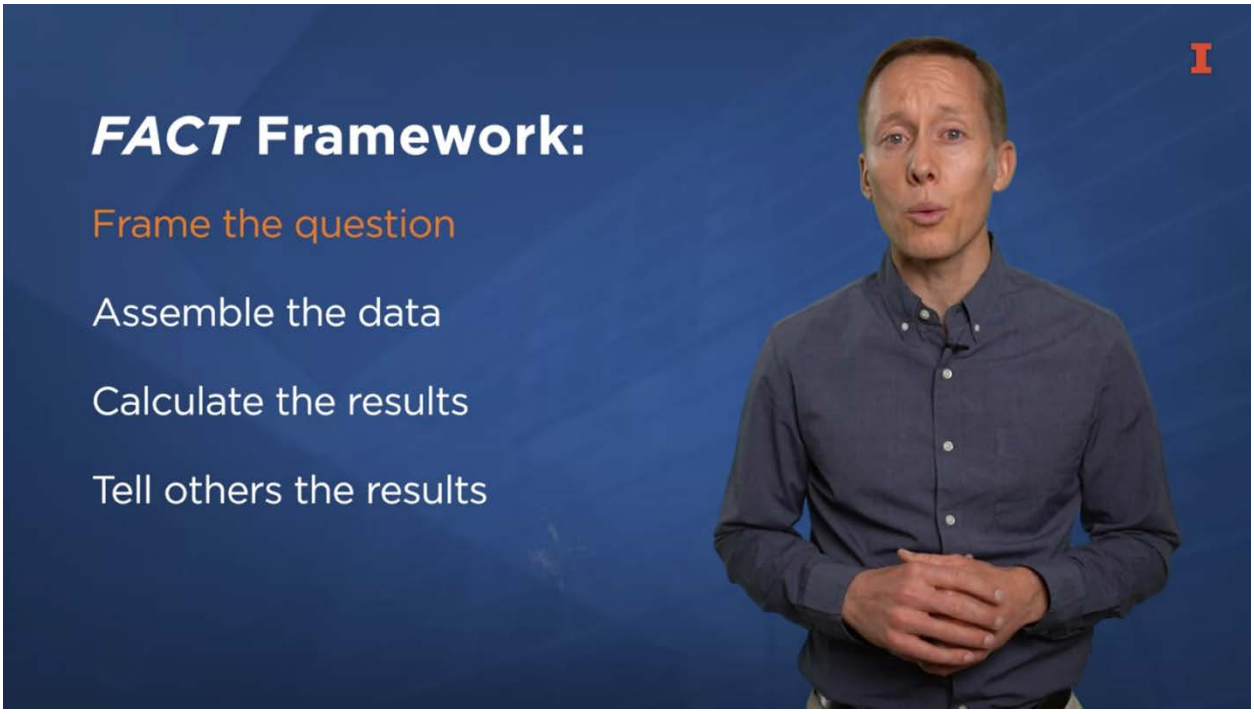


revenue	gross_profit	costs	units
2.09	0.732000	1.358000	2.000
0.01	-1.000000	1.010000	1.000
1.00	0.533704	0.466296	1.000
5.01	0.365130	4.644870	1.000
5.99	1.834670	4.155330	1.000
0.99	0.340280	0.649720	1.000

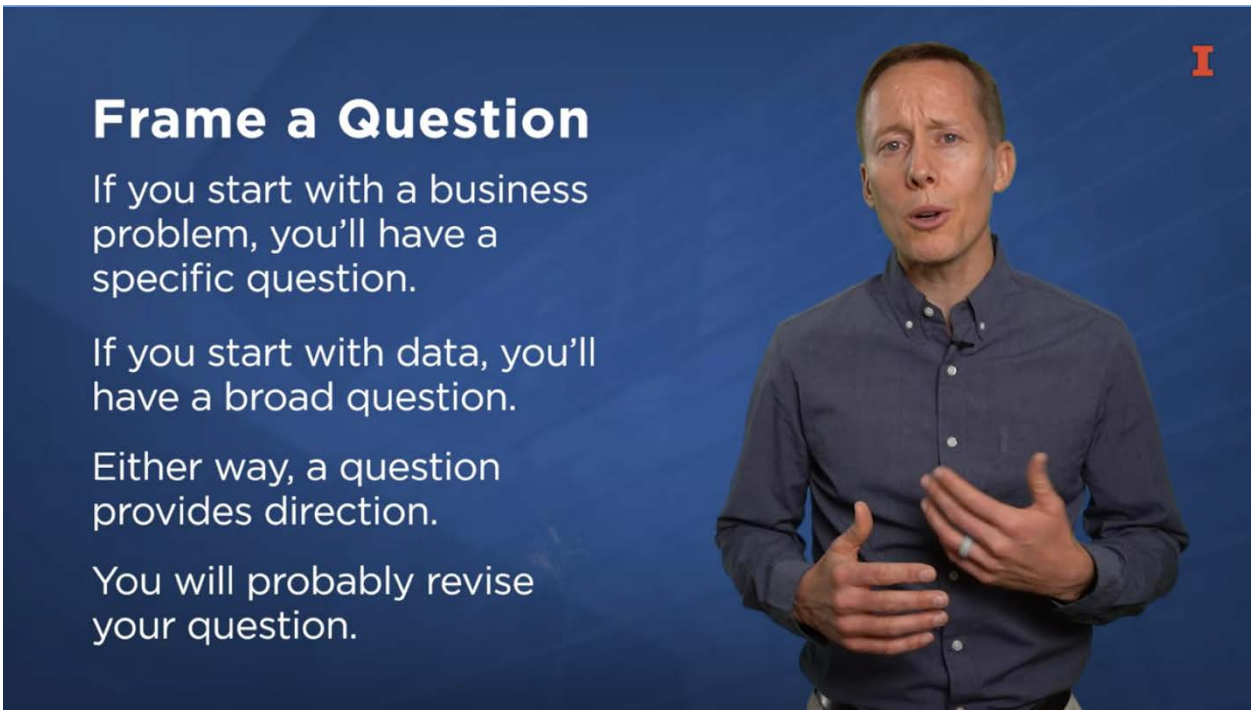
Finally, the fifth group of columns is the revenue and profit group. These columns indicate the amount of revenue and profit that are associated with each row of data. These columns will be critical in our analysis.

Because this data is so granular it can be aggregated at many different levels, such as by hour, day, month, year, by customer, by product, category, parent, profitability, and combinations of those. This means that there are many potential business insights that can be gained from it. While the data has already been cleaned up quite a bit, we have saved some of the data preparation tasks for you to complete. There's an introduction to the data set. You'll get to know it even further as you work on some data preparation and exploration tasks.

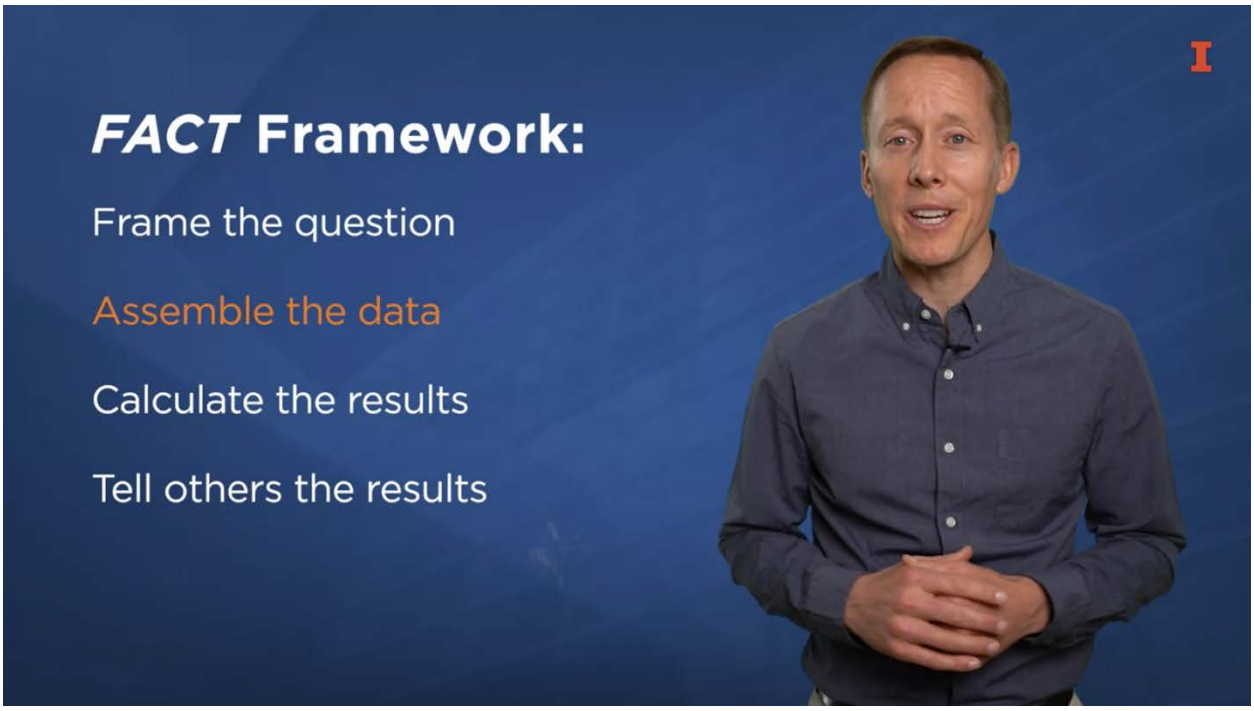
Lesson 1-5.2 Data Analytics Modeling Pipeline



In this lesson, we want to focus on a topic that is related to an analytic mind set by reviewing a procedural framework for applying data analytics to business problems so that you can arrive at actionable insight. Specifically, we will review the FACT framework to guide our discussion. Now, FACT is an acronym in which the F stands for Frame a Question. A stands for assemble data. C stands for calculate the results, and T stands for tell others. Each letter of this acronym actually represents a broad set of related steps. Let's start at the beginning with the F and discuss why it's important to frame a question. I know from personal experience that unless you have some idea for what to look for in a data set, it's easy to spend a lot of time analyzing data without finding actionable insights. I think we all recognize that we live in a world with an ever increasing amount of data. Whether you're using a deductive approach or an inductive approach, whether you're starting with the business problem or with data, and whether you're applying business analytics to predict, explore or experiment with data, framing a question is critical so that you don't drown in a sea of data.

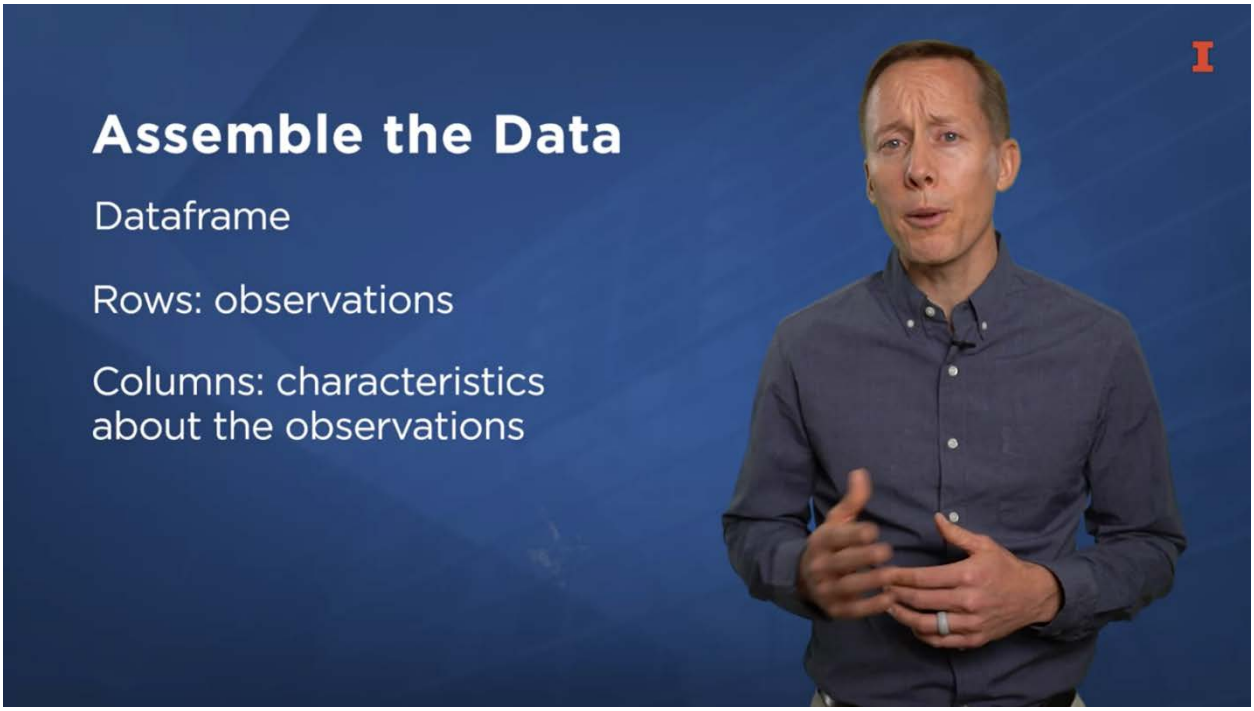


If you're starting with the business problem, then you will probably be able to start with a pretty specific question. In contrast, if you're starting with data, then you will probably have a fairly broad question. That's okay, because as you start analyzing the data, you will most likely either change or refine the way you frame your question.



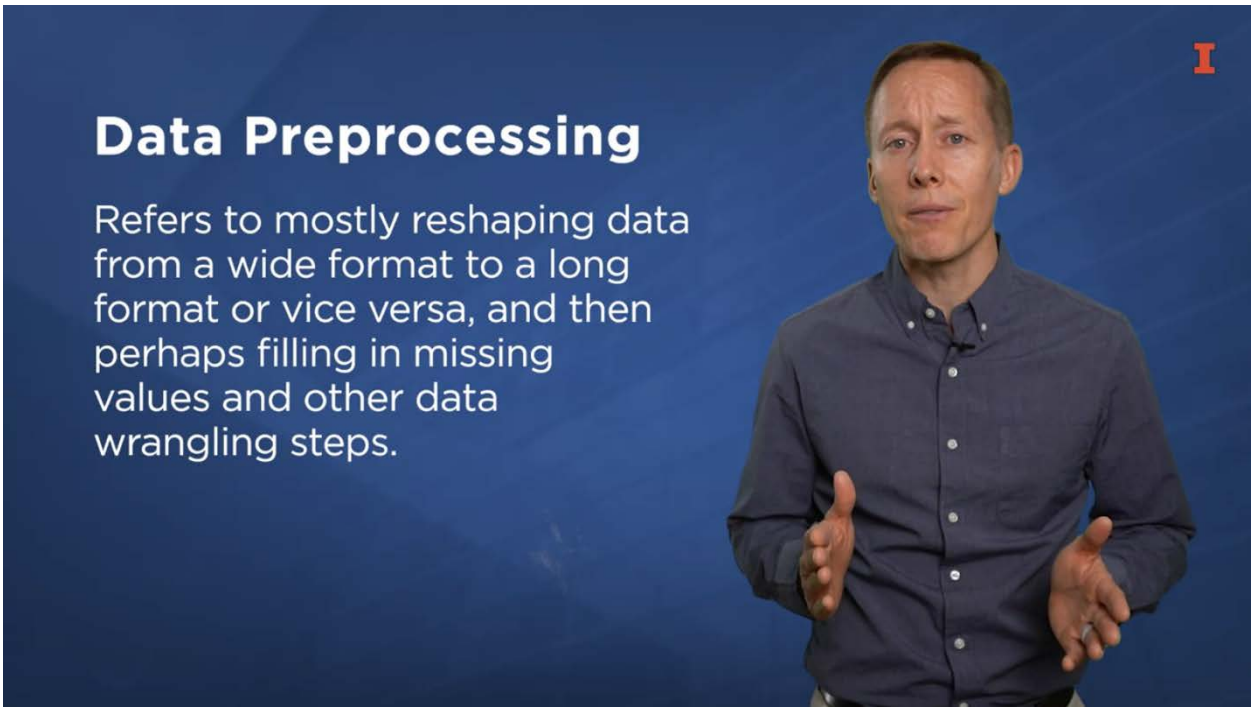
Let's now discuss four aspects of the A in the fact framework, Assemble Data. The first step is to identify data sources and how to acquire the data. Data is warehoused in a variety of locations and in a variety of formats. For instance, you may have some data that is stored in CSV files on your local machine, while other data could be stored in a relational database on a remote server. In that case, you'll need to get the host name and login credentials. The second step is data management and ETL. ETL stands for Extract, Transform, and Load, and this process refers to the steps required to extract

the necessary parts of the data from where it is stored. Transform the data into a format that can be read by data analytic software, and then load it into the software.

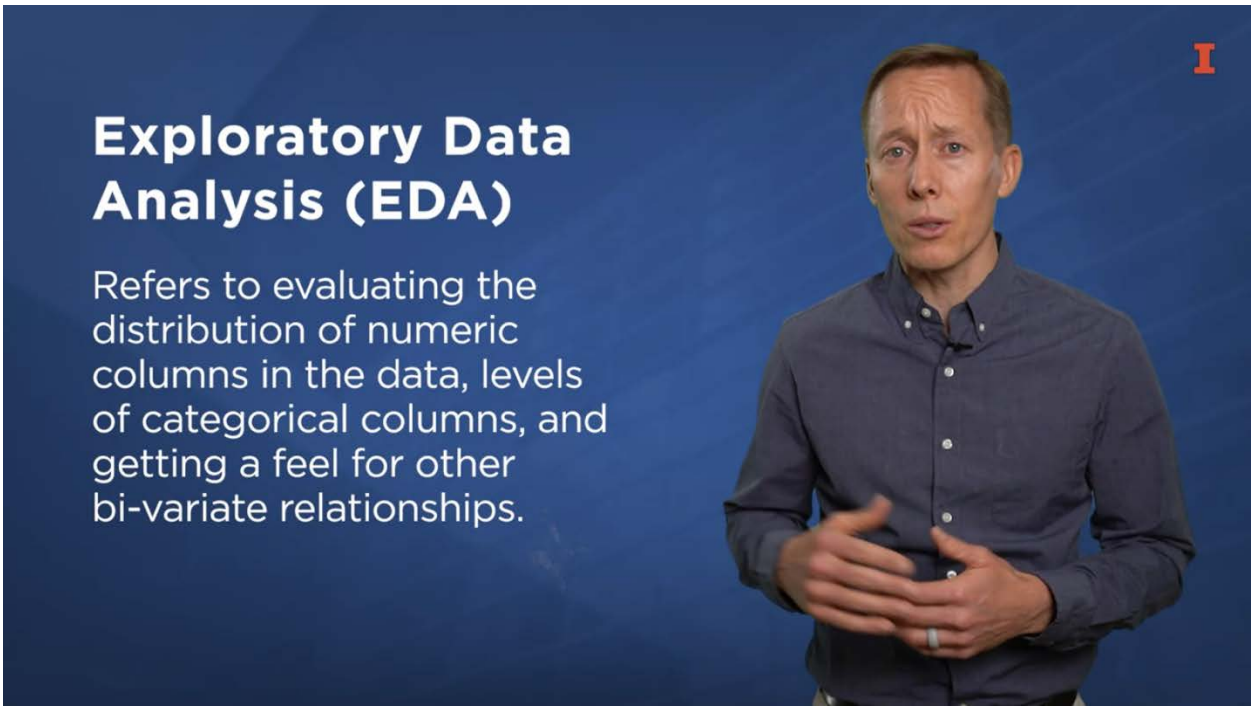


For our purposes, the ETL process will always result in a data frame in which rows represent observations and columns represent characteristics about the observations. There's typically some management that takes place both before and after you read in the data.

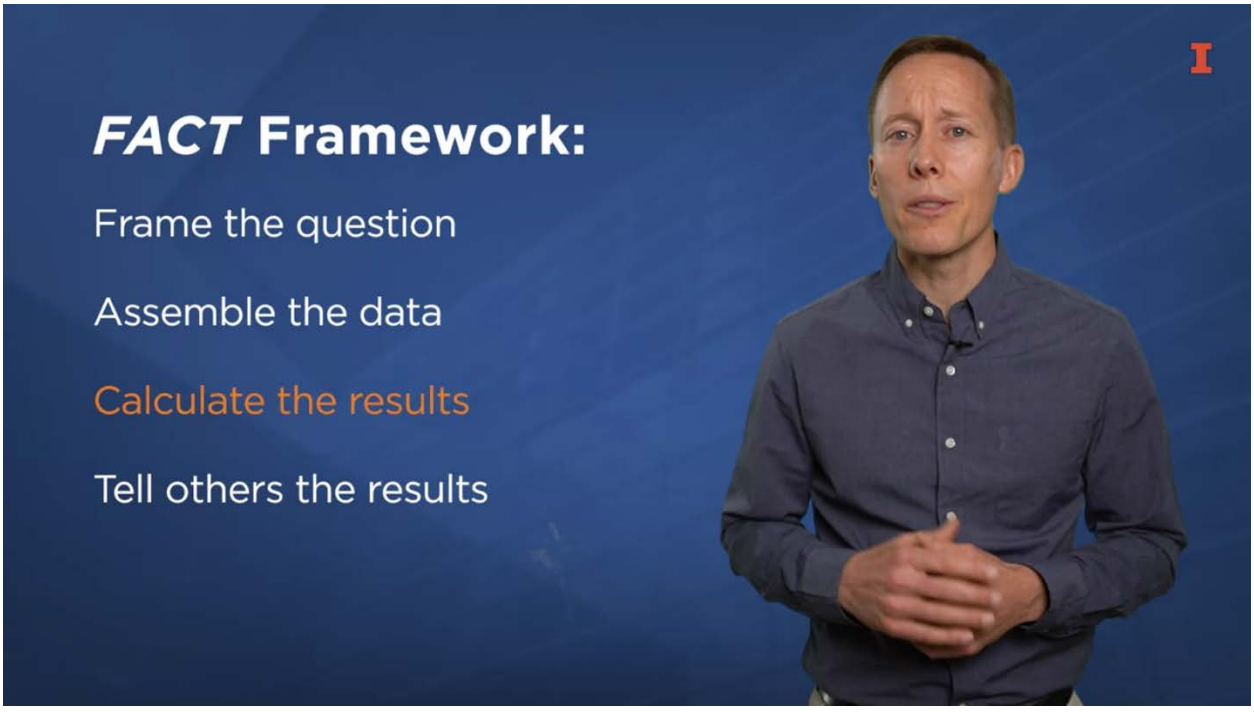
One of those management tasks is to wrangle the data. This refers to joining data, filling in or removing missing values, cleaning up data that has been stored inconsistently such as date values, converting columns of data to the right data type, preparing columns of data in two different data frames so that they can be joined together, and perhaps dealing with outliers.



Once the data is tidied, it typically needs to be pre-processed so that it's ready to be used in visualizations and algorithms. This refers mostly to reshaping the data from a wide format to a long format or vice versa, and then perhaps filling in missing values and other data wrangling steps again.

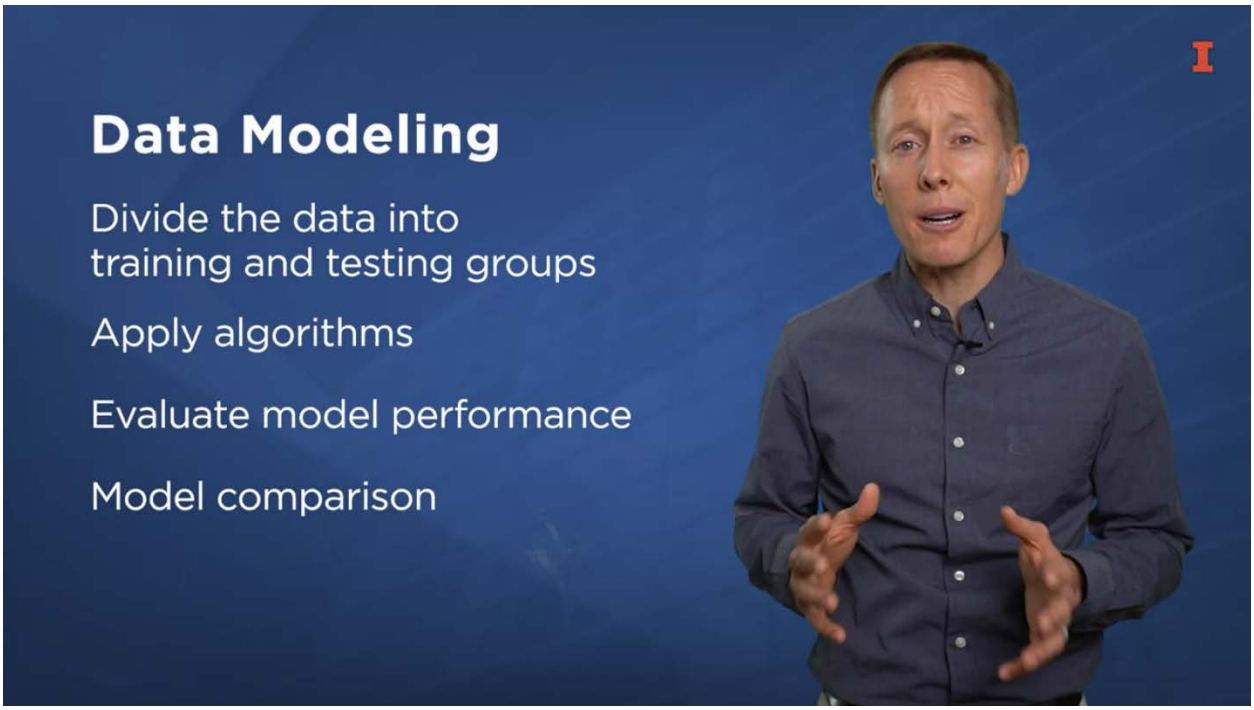


Another aspect of data assembly is known as EDA, which stands for Exploratory Data Analysis. This refers to evaluating the distribution of numeric columns in the data, the levels of categorical columns and getting a feel for other by variant relationships.



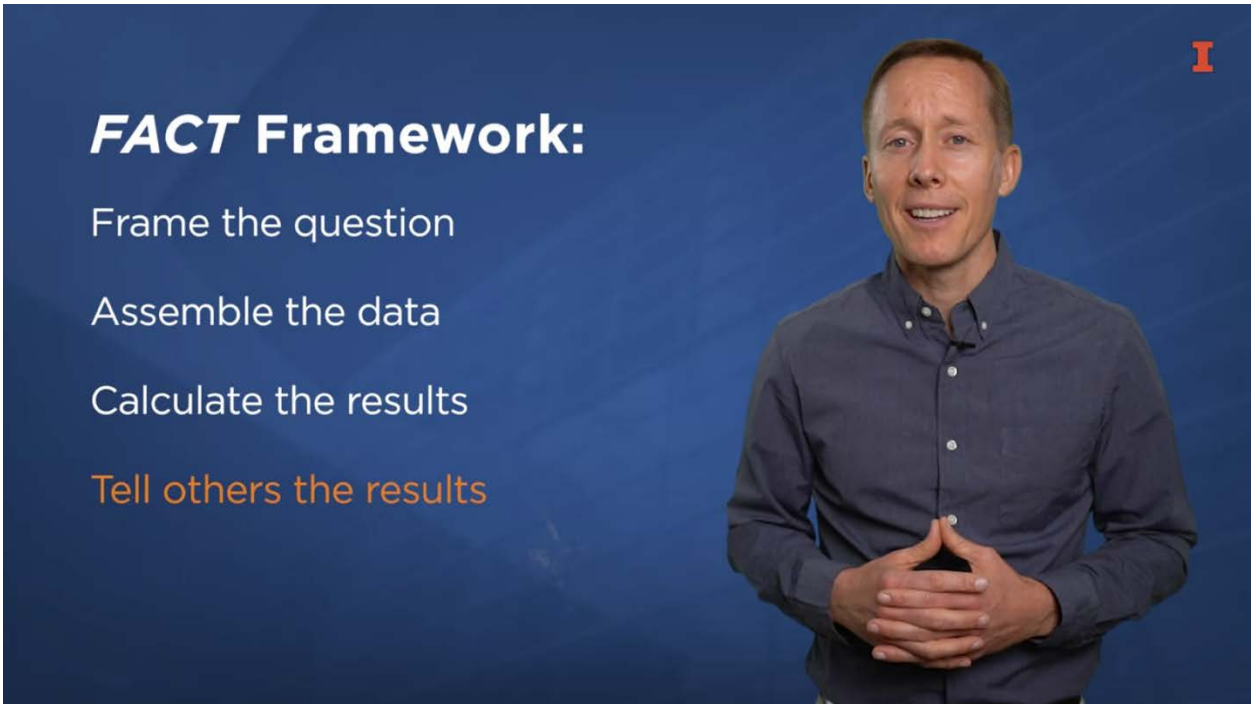
This step is often done with visualizations and summary statistics. This step will most likely highlight outliers and reveal other data wrangling steps. In my experience, EDA often leads to additional data wrangling and perhaps even a revision of the question.

This step can actually be part of calculating results because it may also reveal some insightful relationships and this brings us to the sea in the fact framework. Calculate results. Calculating results often is improved from advanced analytic algorithms that can evaluate complex relationships, calculating the results can be broken down into four steps.



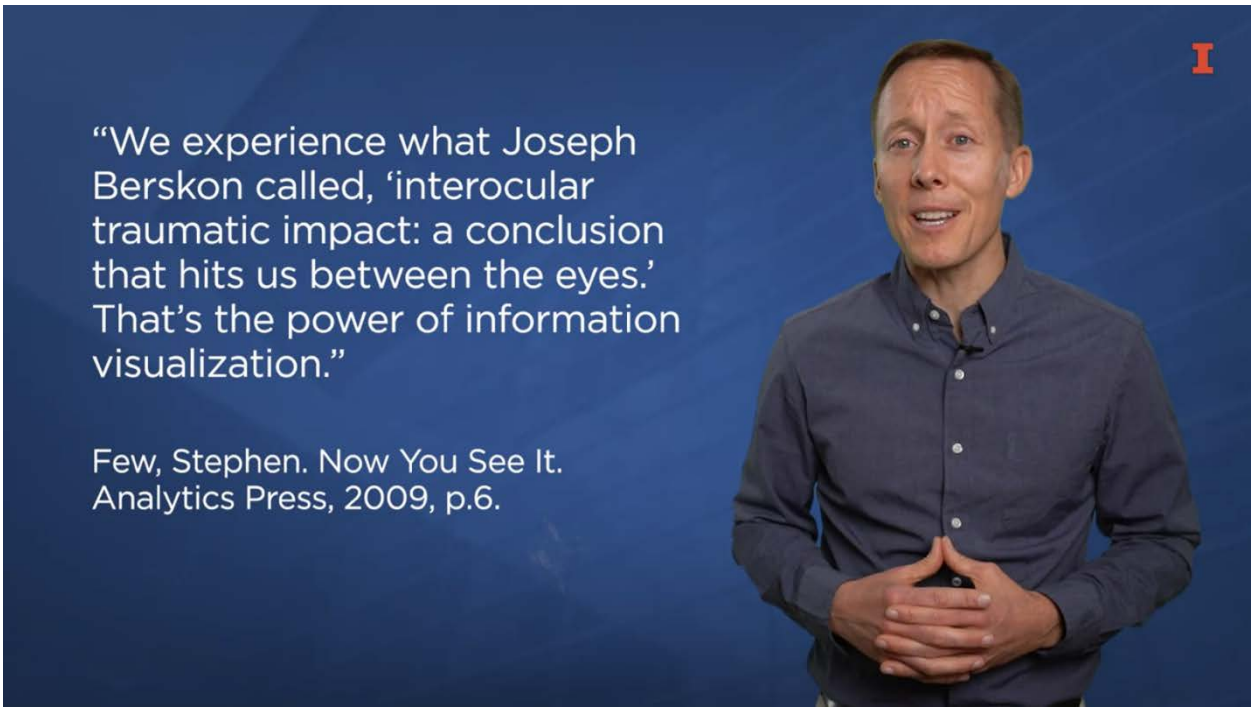
First, the data is often divided into at least two subgroups, typically around 60-75 percent of the data is selected to be processed by some kind of algorithm for developing a model. This is called the training data set. The remaining data set is known as the test data set, and it is used to evaluate model performance, which we'll talk about soon. Once the data has been divided into two groups. The next step is to apply algorithms to the training data set to come up with a model that can be used for explaining, predicting or exploring the data. Every algorithm has various hyper parameters that can be adjusted or tuned. Various combinations of hyper parameter values are typically compared at this point. Next is evaluating the performance of the model. While you can and should consider various metrics that are created based on the training data set, the most powerful way of evaluating model performance is to test how well the model works explaining, predicting or exploring the test dataset. Evaluating model performance on the test dataset is important to prevent over fitting or creating a model that fits one specific dataset, but that cannot be generalized to a broader population of data. The last step of data modeling is to create various models by tuning the hyper parameters and to run various algorithms and then compare the performance of the resulting models. Sometimes an ensemble method is used in which the results from multiple models are combined.

Comparing model performance leads to the next step, which is to analyze the results and identify business insight. This step is where you apply the model to a specific business domain. Hopefully this leads to recommendations for actionable insight.

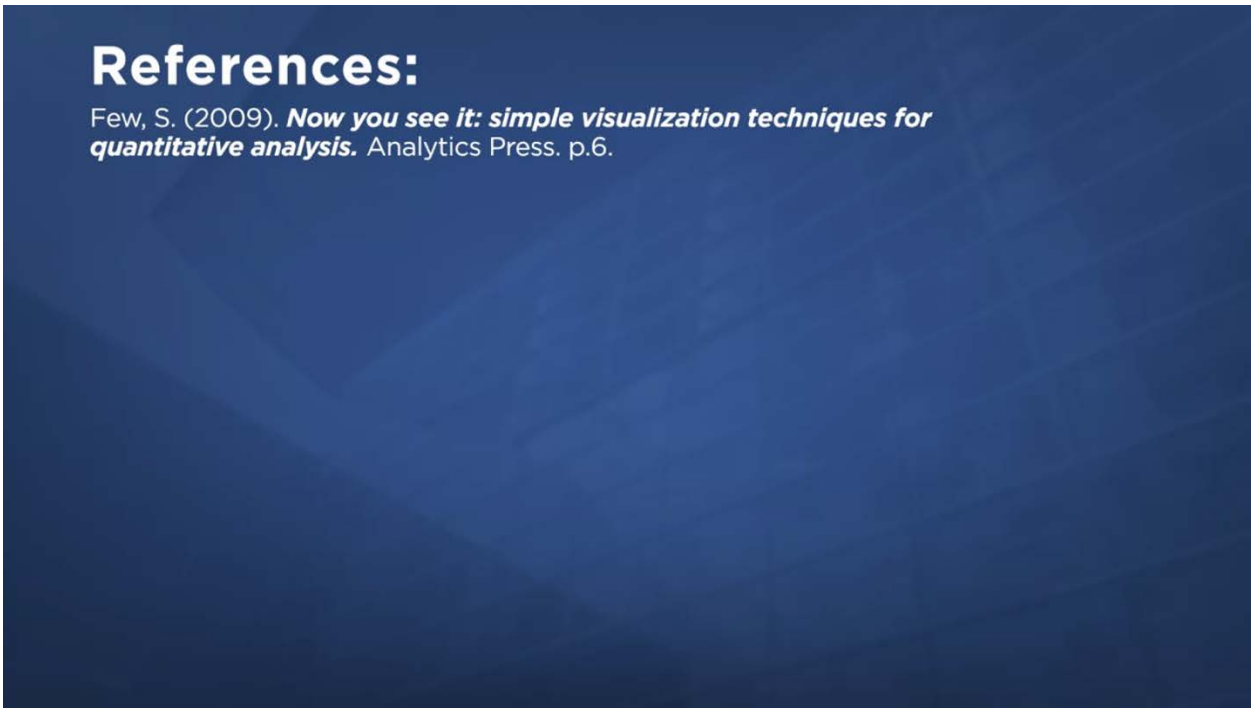


Once the results have been calculated, we need to tell others about the results. This can happen in many ways, but notebook's, reports, and dashboards are powerful ways to combine a narrative with visualizations and code.

It's really a creative endeavor at this point to identify what elements of the analysis and what elements of the results to highlight and to do it in an understandable way. If you share too many details, then you may lose some of the audience. If you share too few details, then your audience may overgeneralize the results or misinterpret them in some way. Visualizations are really helpful for communicating relationships in a memorable way.

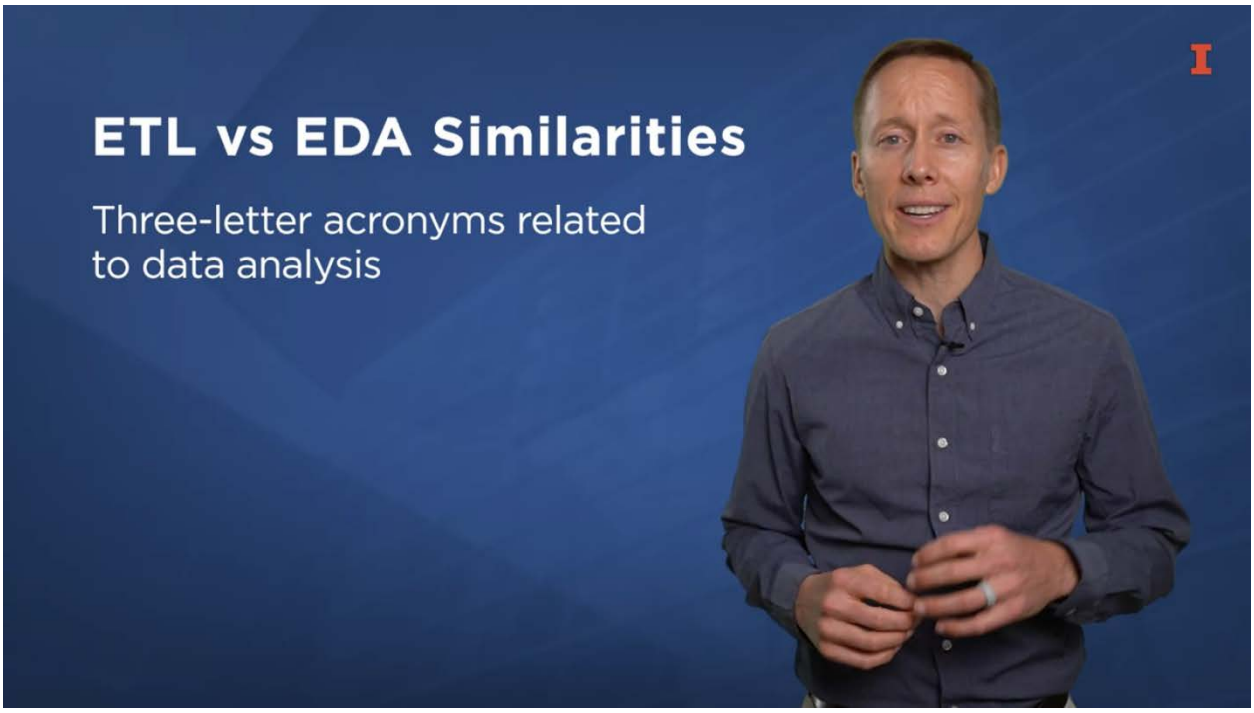


As Stephen Few a data visualization guru says about a good data visualization, he says, We experience what Joseph Bergson called intraocular traumatic impact: a conclusion that hits us between the eyes. That's the power of information visualization. If you're successful at communicating the results to others, then you may end up doing more analysis to evaluate additional questions that arise. Eventually, though, the model goes into production, meaning that you put your money where your mouth is and start acting on the inside. In conclusion, the fact framework is a helpful way to remember the general steps of the data modeling pipeline in this lesson, we have elaborated on these steps by breaking them down into more detailed parts.

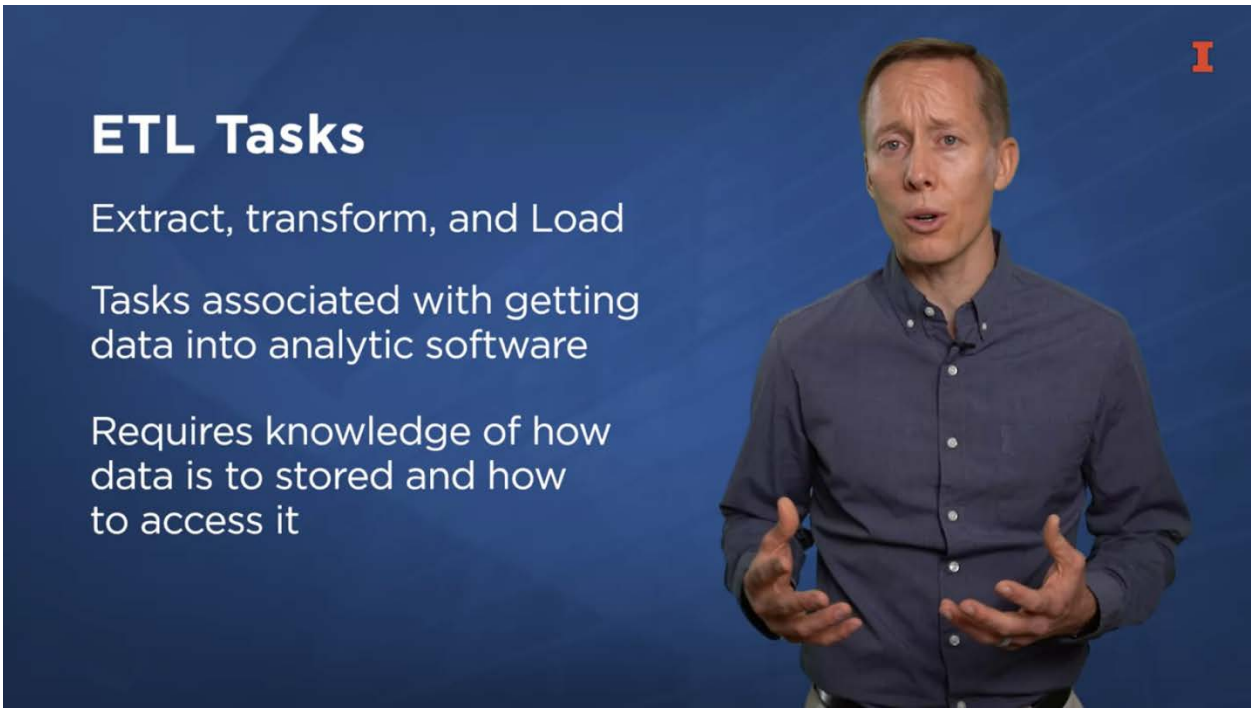

References:
Few, S. (2009). Now you see it: simple visualization techniques for quantitative analysis. Analytics Press. p.6.

Lesson 1-5.3 Business Analytics Key Terms

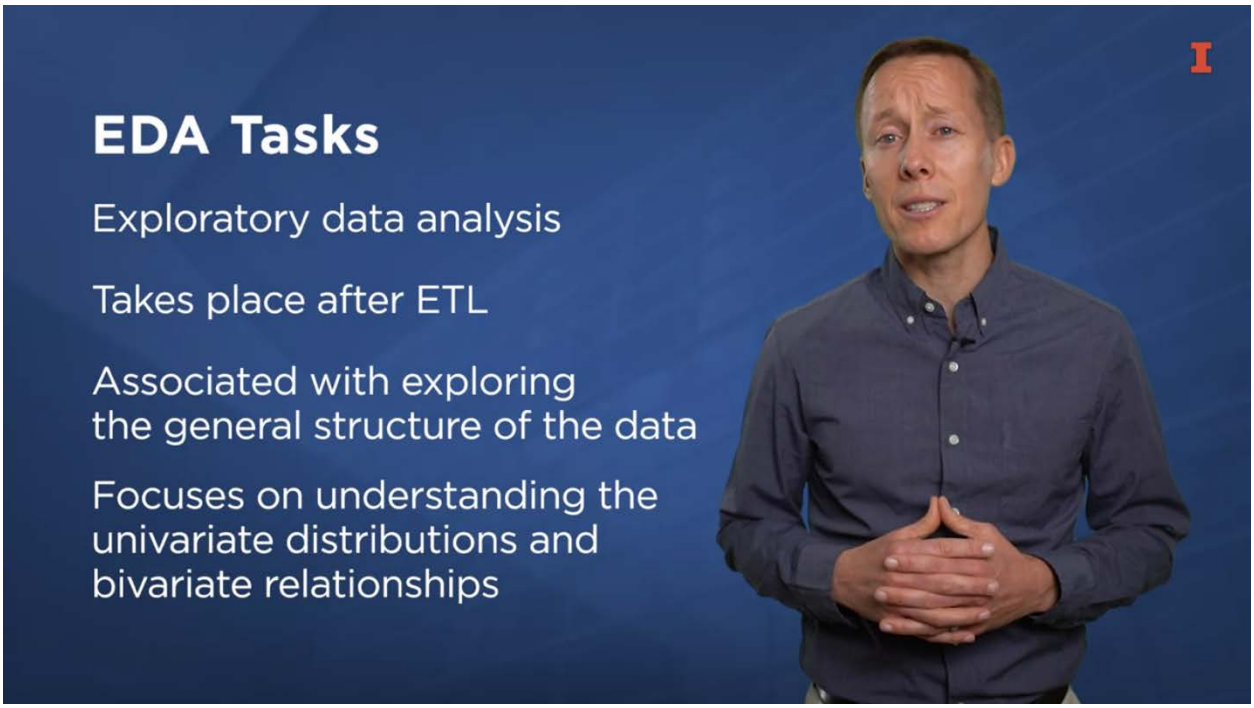
In this lesson, we will introduce some key terms, tools and concepts that are related to business analytics. We will present these terms in pairs to provide a baseline for comparing terms and to help differentiate similar ideas.



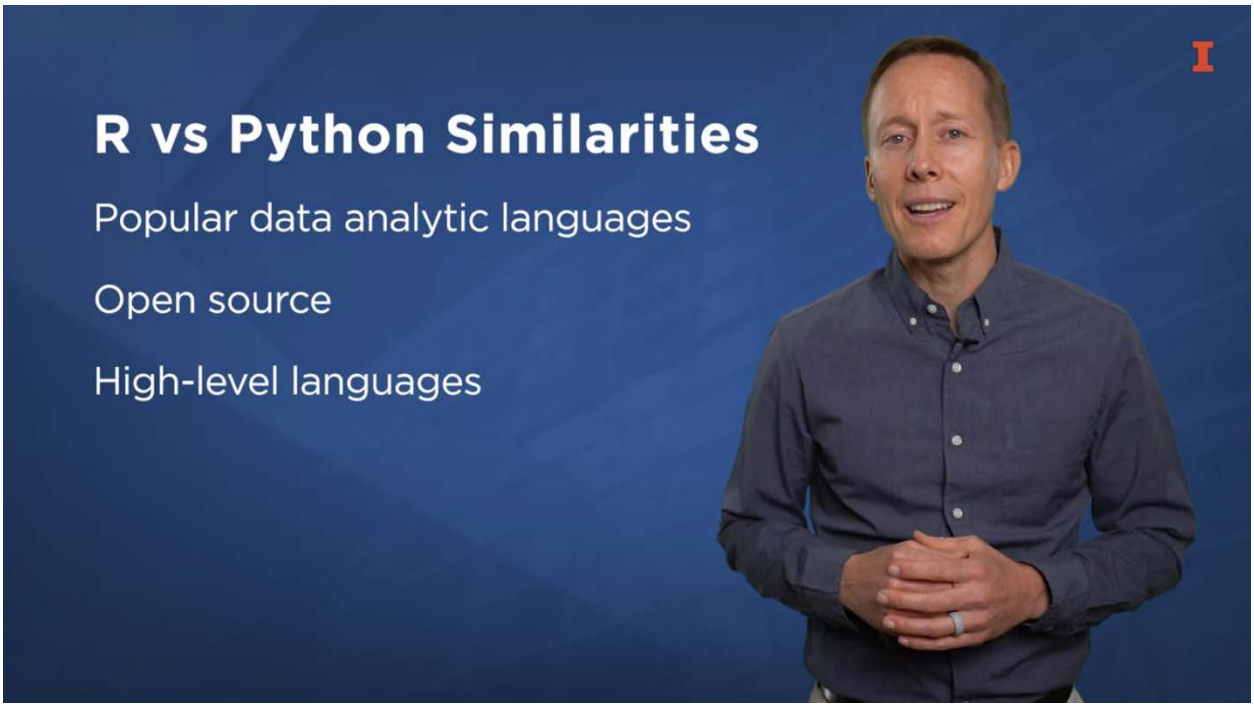
Let's start by comparing ETL and EDA. Both of these acronyms are three letters, and both of them have something to do with data, but that's where the similarities end. ETL stands for Extract, Transform and Load and broadly refers to tasks that are required to get data from where it's stored and into data analytics software.



ETL tasks typically require a knowledge of how data is stored and arranged. For instance, if data is stored on a computer and a series of CSV files, then it's helpful to know where those files are located and how they are named. There are many other ways in which data can be stored, oftentimes data is stored in a remote location. The ETL process begins by gathering credentials to access the data such as the host name, the username and password, and then after gaining access to the remote data storage, it's important to know how to extract the relevant portion of the data and convert it into a format, such as a data frame, that is useful for data analysis.

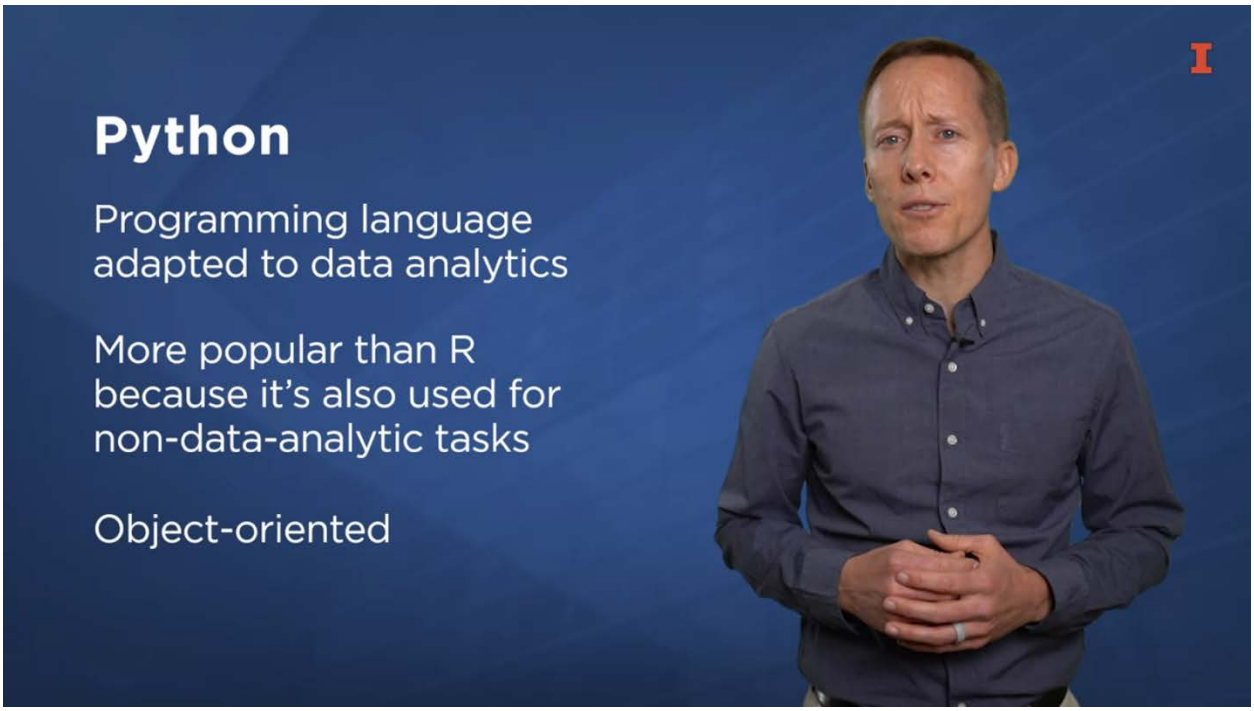


Now let's talk about EDA, which stands for Exploratory Data Analysis. What makes it exploratory? Well, once the ETL processes have taken place, it's really helpful to make sure the data looks like you expect it to look. It's also important to understand the distribution in the data and even some bivariate relationships. For instance, let's say that you're going to analyze a point of sale data that you just loaded into data analytics software. If you're interested in analyzing costs, then it's a good idea to identify how many observations there are for which cost values are missing. You may also want to make sure that there's variation in the cost values and find out how those values are related to values such as quantity. Finding answers to these types of questions are typically pursued using a combination of summary statistics and visualizations. EDA is typically a starting point for more in-depth analysis.

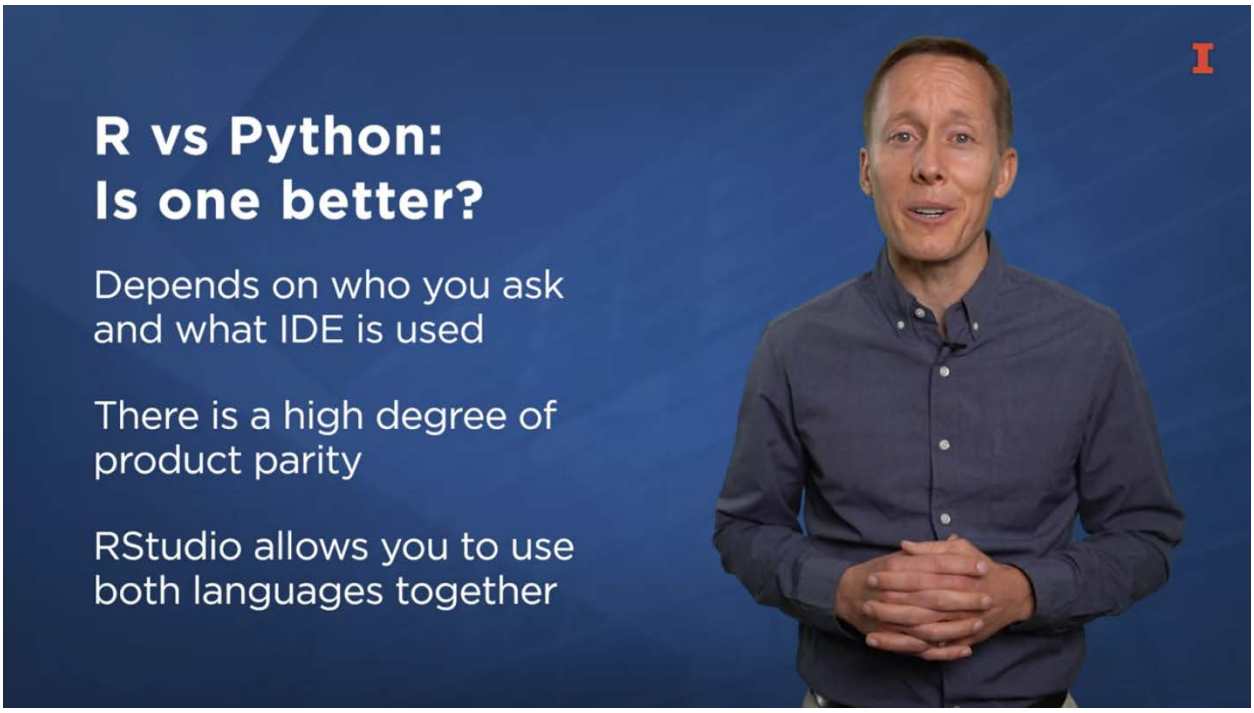


Let's now turn our attention to comparing two popular data analytic languages: R and Python. Both of these data analytic languages have become very popular, largely because both of them are open source, meaning that they're free and lots of smart people have opened their source code for others to use. They're both high level languages, meaning that you don't have to understand much about how computers translate binary signals in order to use these languages. The main benefit of high level languages is that they're easier for humans to read. You can search online for which one is better and find that there are many opinions for one over the other. Here's a little background about each one that might help you better understand some of those opinions.

R is a data analytic language that was built for statisticians who do not have a strong programming background. Therefore, many of the data analytic functions and error messages are often easier to use for those getting into data analytics without previous coding experience relative to Python. R is a very functional language, meaning that many processes are completed by using functions in a similar way that you use functions in Excel. Data analytics using R typically revolves around how to use functions to process data frames.

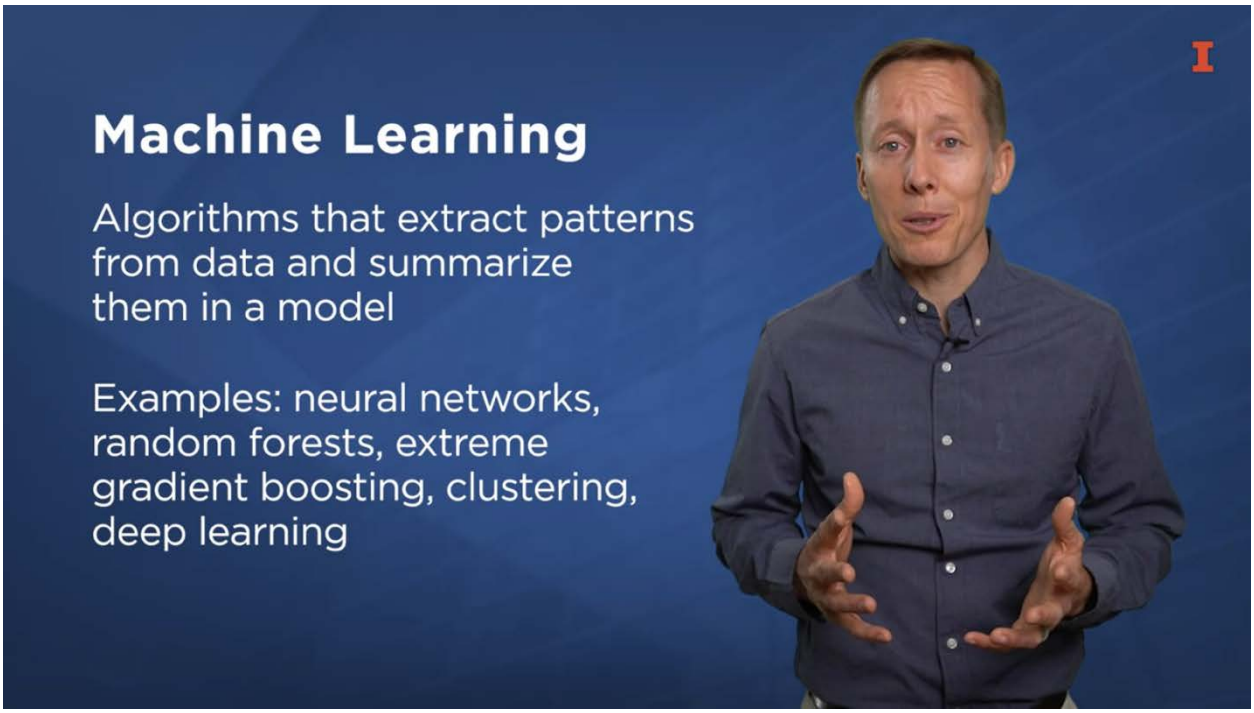


In contrast, Python is a general purpose programming language that was adapted to data analytics. As a general purpose programming language, it is more popular than R because it's used for many non-data analytic tasks. Many of Python's data analytic capabilities have been patterned after those of R. The error messages in Python probably make more sense for those who come from a computer science background than a statistical background. Python is an object oriented language. Different objects, inherent different capabilities and characteristics by virtue of being say a data frame object rather than an array object. The conversation in Python often revolves around different types of objects and the methods and functions to process those objects. Now, I find that R is often preferred by those who are stronger in statistics than in programming, while Python is often preferred by those who have a stronger programming background than statistics. Thus, many software engineers strongly prefer Python and find it easier to develop data analytics skills by using Python, since they are already familiar with Python from prior programming experience.

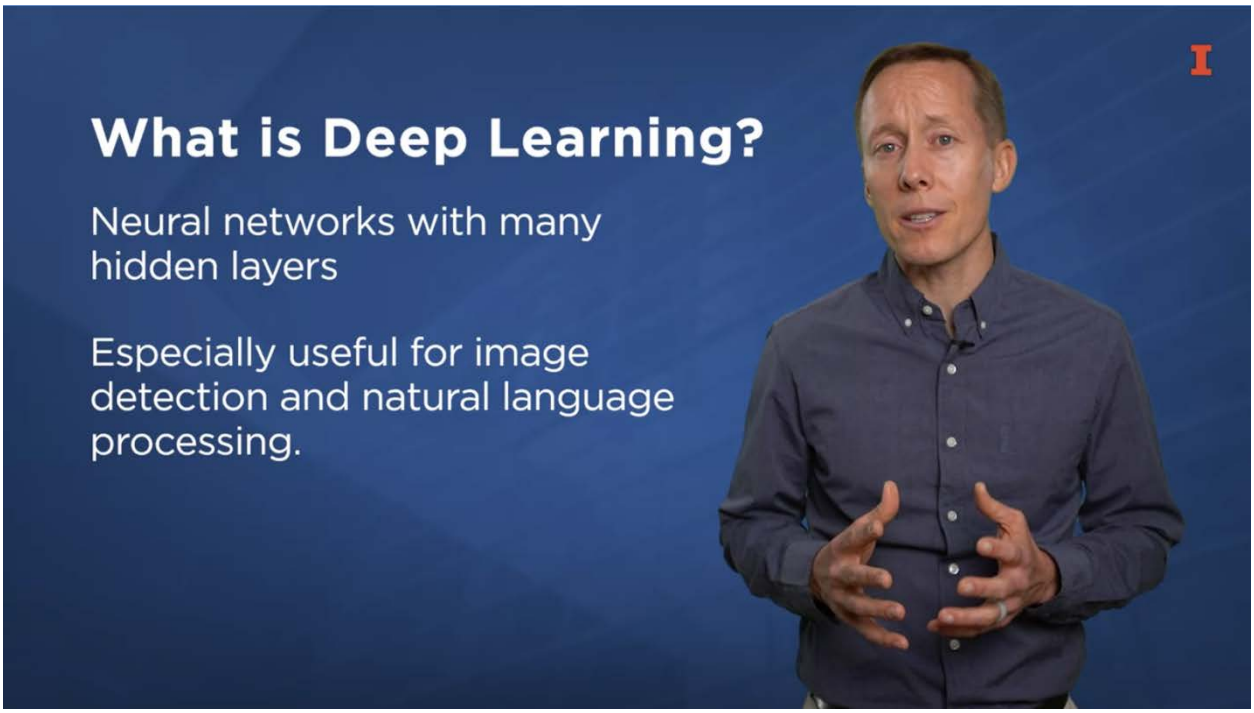


Surely there are some tasks that each data analytic language does better than the other, but it really depends on how you define "better". If you're referring to speed, then Python may be quicker with some tasks because it may require less interpretation. If you're referring to readability, then R may be better with some tasks because it wasn't intended for those with a deep programming background. In my opinion, the most important thing for those who are just getting started is to choose a language and stick with it. Both of these languages have been around long enough now that there is great parity in data analytic tasks. Once you gain a strong foundation and data analytics in one language, it's easier to switch over to the other.

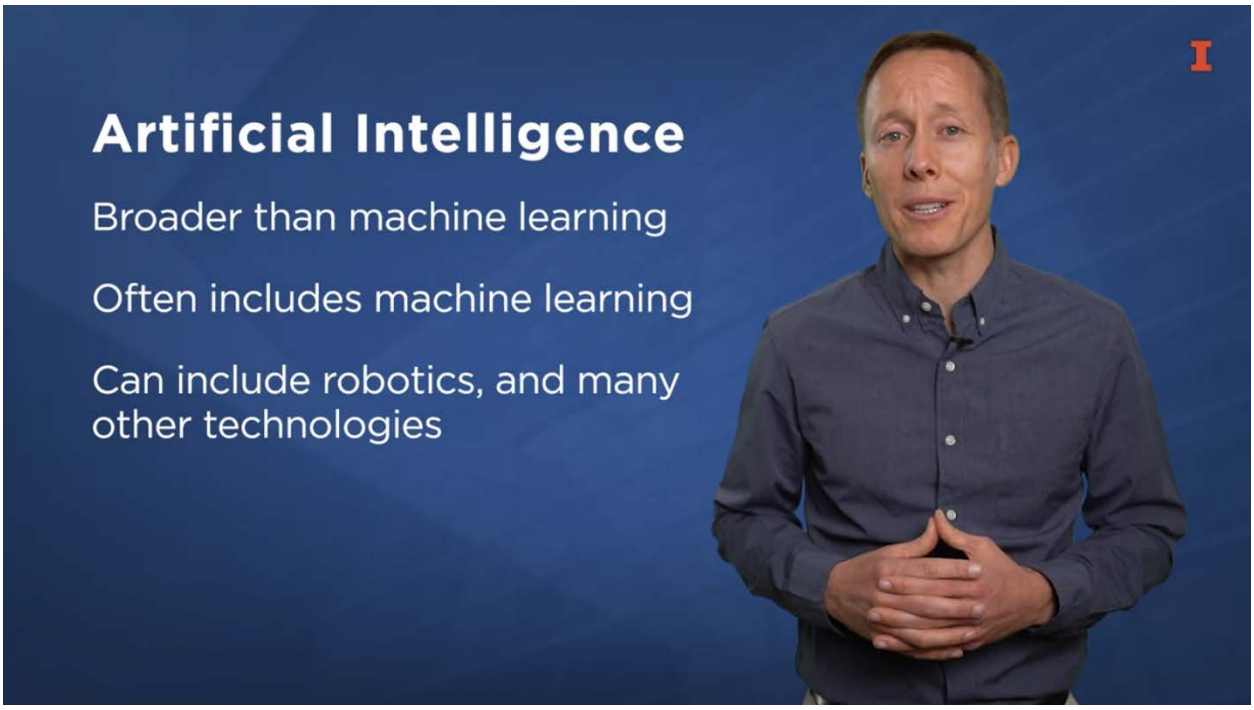
The debate about whether R or Python is better for data analytics is also confounded by the excellent integrated development environments, or IDEs, for creating data analytic code. And IDE is basically software for creating software. Our studio, for instance, allows you to create code in R while also keeping track of your files and projects, the variables and plots that you create, and accessing other tools like the terminal or command line. Our studio also has a bunch of built-in tools for highlighting syntax errors and prompts for helping you to complete your code. It is important to this R Vs Python debate that our studio also allows you to create code in Python so that you really don't have to choose one language over the other, rather you can use both languages together. There are some IDEs that are primarily used for Python, like Jupiter Lab and Spider, but I think our studio is the best one. Part of what makes it so great is that some of the common programming tasks, like publishing a dashboard file to the internet or knitting a markdown code into a PDF or HTML file can be done with the click of a button.



Let's shift gears now and turn our attention to comparing machine learning with artificial intelligence. Machine learning refers to the algorithms that extract patterns from data and then summarizes those patterns in a model that can be used to predict future outcomes, identify differences, or highlight relationships that might otherwise be hidden. So the machines learn these patterns by observing data, and the insight of the resulting model improves as a result of learning from more data. There are many machine learning algorithms, such as neural networks, random forests, extreme gradient boosting, and clustering. You've probably heard of deep learning.



How does deep learning fit into this conversation? Deep learning is a type of neural network that has many layers of data processing and it has been very useful for image detection and natural language processing. Thus, deep learning is a subset of machine learning.

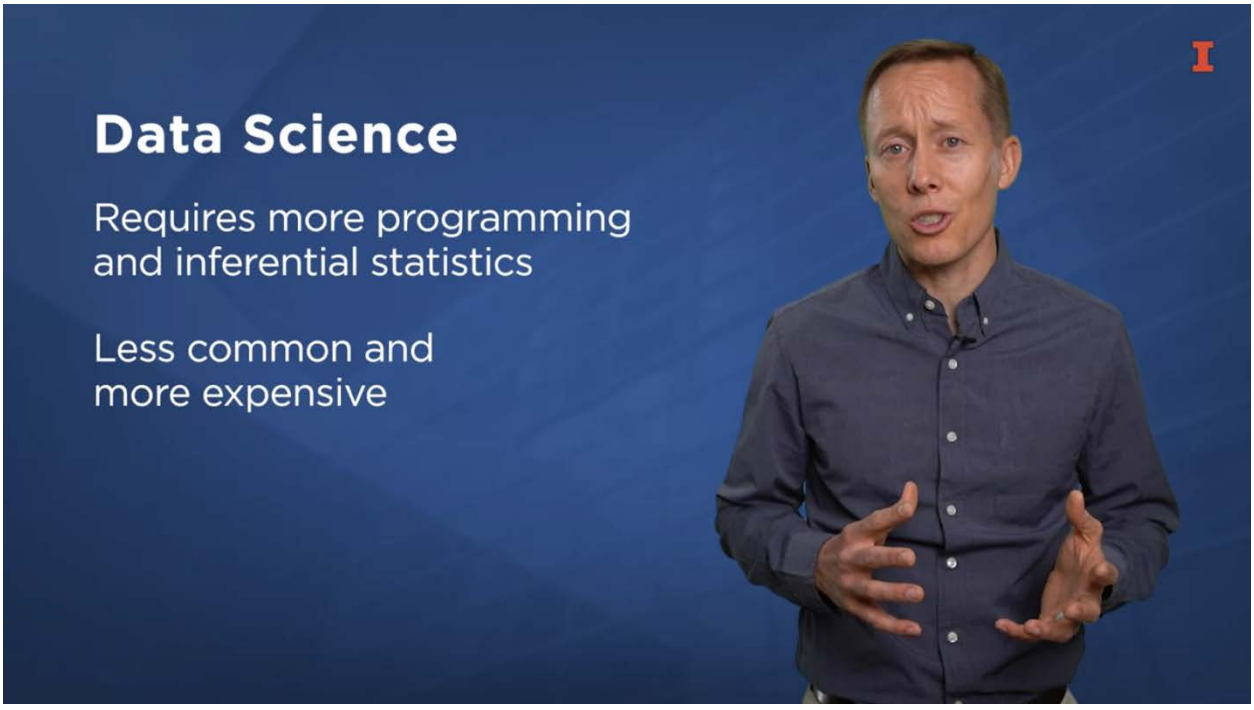


Artificial intelligence, or A.I., is broader than machine learning and is often built on machine learning algorithms as well as many other technologies, for example, A.I. may

be used to mimic human-like behavior in a robot by taking input, analyzing it with machine learning algorithms and then returning output in a humanistic way, for that to happen. Machine learning could certainly be used as one technology in conjunction with mechanical and electrical engineering technology so that a robot can translate audio and visual cues from its environment into gestures that mimic walking or audio signals that mimic talking. Finally, let's bring up data analytics versus data science. This is kind of a difficult comparison because to my knowledge, there's not a definitive source to which everyone subscribes. However, there are some commonalities in the way that people use these terms.

These titles definitely overlap in their responsibilities. Specifically, both data analysts and data scientists should be expected to be able to perform the ETL, EDA, data visualization and machine learning processes.

However, a data scientist is also expected to be more skilled at modeling, deploying models into production, and dealing with larger amounts of data and more complex problems.



Thus, a data scientist is likely to have more training and programming and applying the scientific method using inferential statistics. Based on a recent search on LinkedIn for the area in which I live, there were 70 percent more data analyst jobs compared to data

scientist jobs. A search for salaries indicated that data scientists are likely to receive higher salaries than data analysts, anywhere from 35 percent to 100 percent more than a data analyst.

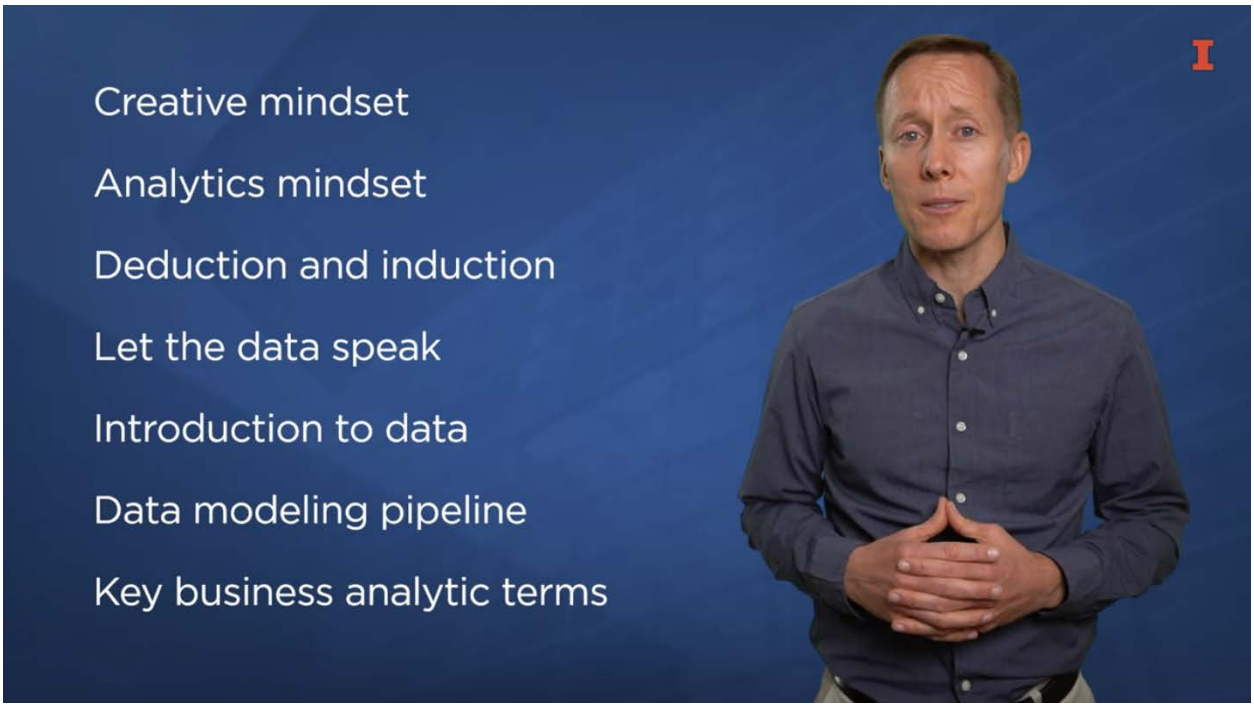
In conclusion, let's end with this question. What is business analytics? That's really the focus of this course. A business analyst is a data analyst that can perform the ETL, EDA, data visualization, and machine learning processes using business data rather than genomic data for instance. Combining your business knowledge of marketing, finance, accounting, logistics and organizational behavior domains with data analytics skills should make it possible for you to become a business analyst.

Lesson 1-6: Module 1 Review

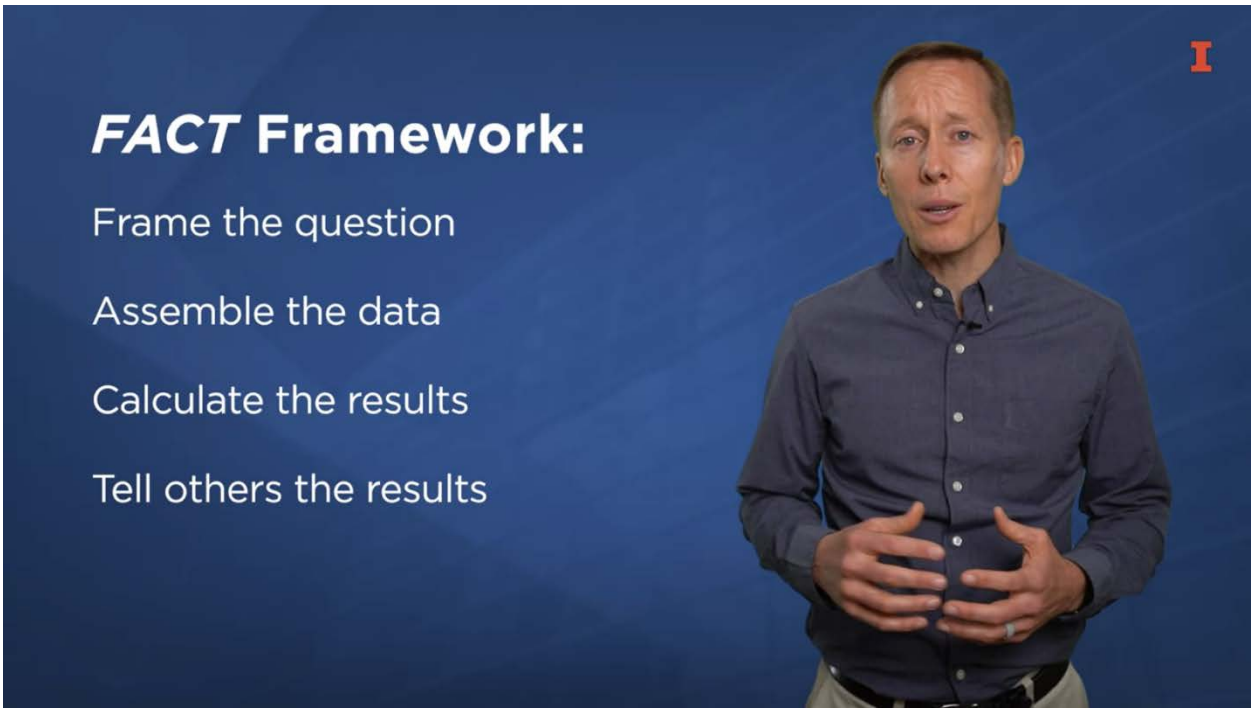
Lesson 1-6.1 Module 1 Conclusion

Prof. Jessen Hobson: Business analytic tools are pretty awesome, and we're excited for you to learn more about them so that you can drive actionable insight, the software tools are so powerful that you'll quickly find insight. However, if you're not careful, those insights will not necessarily be for the business questions that you're trying to answer. Even worse, you may find actionable insight, but that insight may worsen business performance.

Prof. Ronald Guymon: The purpose of these lessons has been to emphasize that your mind is the most important business analytic tool that you have.



We have discussed the importance of creativity, which is the ability to make connections between concepts, specifically data analytics and business domain knowledge. We have also affirmed the importance of having an analytic mindset so that you can take a big problem and break it down into manageable pieces. We have brought your attention to two different methods for making: inference, deduction, and induction. We have reminded you to be open to data analytic results that go against your business intuition but we've also reminded you that you should not totally ignore your intuition.



We introduced you to a data set and talked about the steps required to frame a question, assemble data, calculate results and tell others the results. Finally, we introduced you to some key business analytics terms.

Prof. Jessen Hobson: In short, your mind is now better prepared to use the powerful business analytic software tools that you'll soon learn more about. When you combine the right mindset with the right business analytic tools, you can do some really impressive things.

While a fool with a tool is still a fool. When tools and a creative mind are combined, benefits flow to humankind.