Telling a story with data

By the end of this lesson, you'll be able to explain what data visualization is and why it's valuable in business.

Looking at a large table of data and extracting information from it can be overwhelming. Even when you know that the data is rich with information, it's hard to pick out useful information just by looking at the data. You may get an idea of what fields and data types there are, but **you can't easily summarize the data or identify relationships within the data. This is where data visualization comes in.**

Data visualization is the visual representation of data, using infographics such as graphs, charts, and diagrams. Visualizations help you get a better understanding of your data because it lets you visually see it.

Data visualization: A chart, graph, diagram, or other visual representation of data used for exploring data or presenting data insights

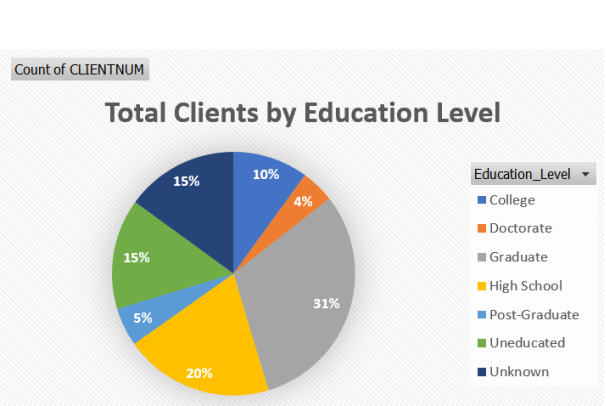
Visualizations can communicate a lot of information in an instant. They display relationships within the data. They can also show summary information, such as parts of a whole or comparisons between different groups. And they can show trends over time.

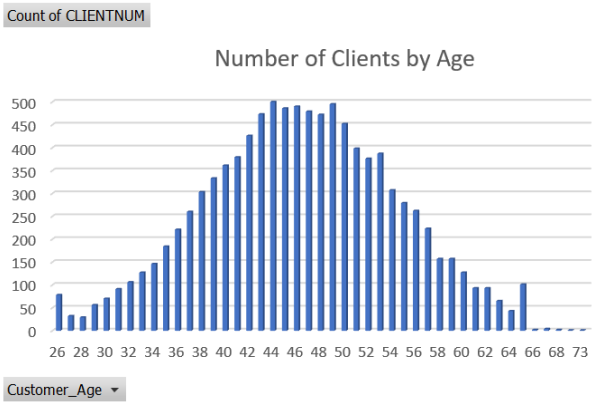
Data visualization for exploring a dataset

Take a look at the data below. Can you draw any meaningful conclusions from it?

	A	B	C	D	E	F	G	H	I	J
1	CLIENTNUM	Attrition_Flag	Customer_Age	Gender	Dependent_count	Education_Level	Marital_Status	Income_Category	Card_Category	Months_on_book
2	768805383	Existing Customer	45	M	3	High School	Married	$60K - $80K	Blue	39
3	818770008	Existing Customer	49	F	5	Graduate	Single	Less than $40K	Blue	44
4	713982108	Existing Customer	51	M	3	Graduate	Married	$80K - $120K	Blue	36
5	769911858	Existing Customer	40	F	4	High School	Unknown	Less than $40K	Blue	34
6	709106358	Existing Customer	40	M	3	Uneducated	Married	$60K - $80K	Blue	21
7	713061558	Existing Customer	44	M	2	Graduate	Married	$40K - $60K	Blue	36
8	810347208	Existing Customer	51	M	4	Unknown	Married	$120K +	Gold	46
9	818906208	Existing Customer	32	M	0	High School	Unknown	$60K - $80K	Silver	27
10	710930508	Existing Customer	37	M	3	Uneducated	Single	$60K - $80K	Blue	36
11	719661558	Existing Customer	48	M	2	Graduate	Single	$80K - $120K	Blue	36
12	708790833	Existing Customer	42	M	5	Uneducated	Unknown	$120K +	Blue	31
13	710821833	Existing Customer	65	M	1	Unknown	Married	$40K - $60K	Blue	54
14	710599683	Existing Customer	56	M	1	College	Single	$80K - $120K	Blue	36
15	816082233	Existing Customer	35	M	3	Graduate	Unknown	$60K - $80K	Blue	30
16	712396908	Existing Customer	57	F	2	Graduate	Married	Less than $40K	Blue	48
17	714885258	Existing Customer	44	M	4	Unknown	Unknown	$80K - $120K	Blue	37
18	709967358	Existing Customer	48	M	4	Post-Graduate	Single	$80K - $120K	Blue	36
19	753327333	Existing Customer	41	M	3	Unknown	Married	$80K - $120K	Blue	34
20	806160108	Existing Customer	61	M	1	High School	Married	$40K - $60K	Blue	56
21	709327383	Existing Customer	45	F	2	Graduate	Married	Unknown	Blue	37

You're probably struggling to come up with any insights, which makes sense—you're looking at a jumble of data. Now, take a look at the two charts below, which are based on the same data table.





It's probably easier for you to wrap your head around the data after looking at these charts. You can quickly recognize that this dataset has clients of many age groups and education levels. You can also see that the majority of clients have a graduate education level and are 40-50 years old. This is precisely why data visualization is so useful. In general, data visualization helps you visualize complex data, making it easier and faster to interpret.

Data visualization for storytelling

Data visualization is often used to tell a story about the data. A dataset isn't just an isolated or random set of values. Rather, every dataset has a background associated with it—a story or reason why the data was collected in the first place.

Similarly, every time that you analyze data, you try to solve a problem or answer a question.

After you analyze data, you can make conclusions about the data itself or about future actions that you or others should take.

All of this is a story that needs to be communicated effectively for others to understand it.

Not only can data visualization help you make sense of data, but it’s also especially helpful when you want to share your data and insights with others who are unfamiliar with the data.

For example, if you're trying to convince your boss to offer a new deal to customers, you can use visualizations to illustrate the positive impact that it would have on the business. Or, if your company has already started offering a new deal to customers, you can create visualizations that summarize the results for business leaders in your company.

The data analysis process

This lesson introduces the data analysis process and shows how data visualization fits into it.

Data analysis process

The process of cleaning, manipulating, and modeling data with the goal of getting meaningful insights and conclusions.

Raw data

Data that has been collected but hasn't had any additional processing or cleaning.

Data cleansing

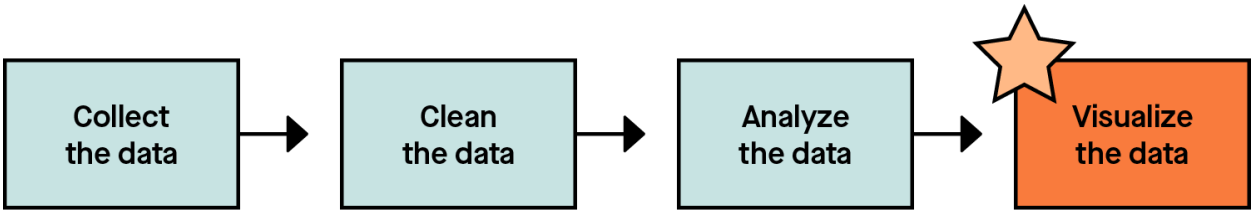
Also known as data cleaning, the process of identifying and correcting missing, inconsistent, or incorrect data.

Data visualization

The process of creating visual representations of the data or analysis, using graphs and charts.

Introduction

The data analysis process is a four-step process. It starts with data collection, followed by data cleansing and data analysis. The final step—and the focus of this course—is data visualization.



Step 1: Collect

To visualize data, you first need to have data. The first step in the data analysis process is data collection. Data can be collected in several ways, including surveys, online website tracking, or transaction tracking. Data that is collected from a source and hasn't undergone any additional processing is called raw data.

Step 2: Clean

Once you collect the data, you can't visualize it just yet. Most raw data isn't clean. It may have errors, missing values, or formatting issues. This can become a problem if it isn't fixed. If the data type of a field is wrong, then any formulas that you use on the data may produce wrong results. If there are blanks or strange symbols, then you may be missing important data from your analysis. Duplicate rows can give you incorrect totals and averages, which interfere with the accuracy of your visualizations.

Cleaning the data, or data cleansing, is the process of fixing raw data. You should make sure that data types are correct, and that they use consistent formatting. You should also check that there are no duplicate records, and you should fix any blank or missing values. Your job is to go through the data and look for any inconsistencies. This also gives you an opportunity to get more familiar with your data.

The table below shows a few rows of Walmart transaction data. As you can see, this data hasn't been cleaned. Customer 3 is missing a Total amount value and has a strange symbol in the Return column. So if you wanted to get a total or average of these columns, the result would be lower than it should be. The dates in the Date processed column are all in different formats. This creates more work to understand the processed dates.

Later in this course, you'll learn more about how to clean raw data so that it's ready to be used in charts and graphs.

Customer ID	Total amount	Date processed	Return
1	$35.87	10/1/2020	0
2	$87.90	2021-11-13	1
3		April 20, 2021	*
4	$42.50	1/15/21	0

Step 3: Analyze

Once your data is clean, it's ready for analysis. Analyzing data is the process of exploring and interpreting the data to draw conclusions. Statistical tests can highlight important relationships between fields and desired outputs. Similarly, dynamic models can also show how the data fields relate to one another, because input scenarios can be changed to see resulting outputs.

The goal of the analysis step is to find answers or explanations to the problem that you're trying to solve or the topic that you want to understand.

Step 4: Visualize

The last step in the data analysis process is visualizing the data. Or, to put it another way, data visualization. Data visualizations can assist with analyzing data. By turning the data into tables, graphs, and charts, you can view data from different angles and see patterns and relationships in the data.

Once you finish analyzing data, you probably need to share your results and findings with others. Because most people find pictures easier to interpret than statistical models and calculations, data visualizations can help you communicate your findings in a clear and engaging way.

For example, line graphs illustrate trends over time. Bar charts show comparisons between multiple groups. Pie charts show how each category makes up the whole.

The remainder of this course focuses on data visualizations. The next few modules will introduce commonly used chart types and show you how to create them using popular data visualization software. Then, you'll learn how to interpret data visualizations and come up with takeaways. Being able to use data visualizations effectively is a powerful skill because it allows you to better understand data and communicate complex information to people of all levels.

Conclusion

This lesson outlined the data analysis process: collecting data, cleaning data, analyzing data, and visualizing data.

Data visualization best practices

By the end of this lesson, you'll be able to identify data visualization best practices that help to show complex information in a digestible way.

Introduction

There is some creativity involved in making data visualizations, so there are no strict rules about how to create them. However, there are some best practices that you should follow. These best practices will help you create informative and beautiful visualizations that are also easy to understand.

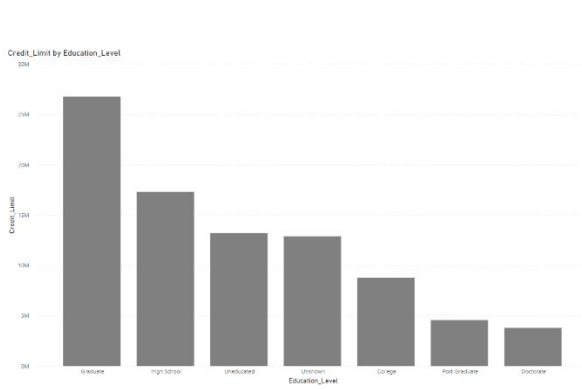
Choose an appropriate visualization

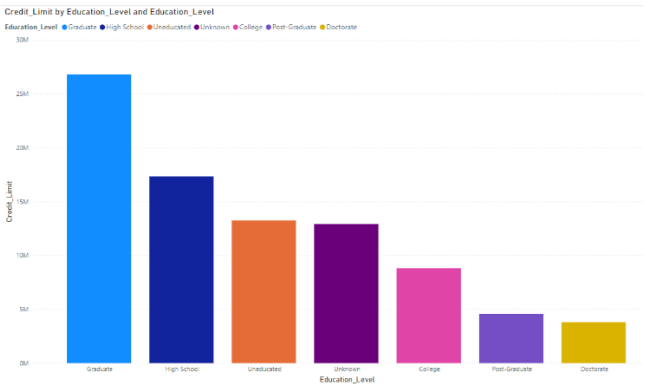
As you progress through this course, you'll be introduced to many different types of data visualizations, from simple bar charts to more complicated ones like waterfall charts. There's no shortage of visualizations that you can choose from. But because of this, it's important to select your visualizations wisely.

Each chart type has its own strengths and weaknesses; some are more effective than others in communicating a certain story. Try to select the visualization that is the most appropriate and most accurate for what you want to show. You'll learn more about when to use different visualizations in future lessons.

Use color

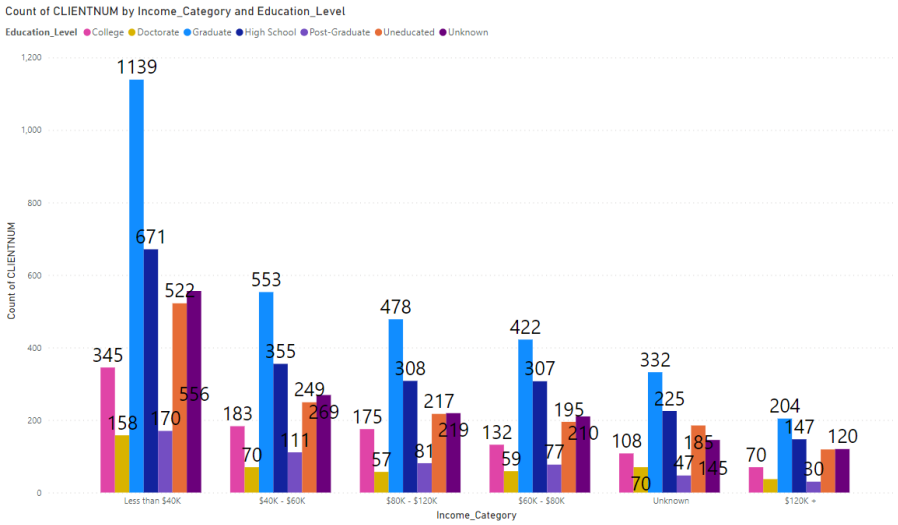
One of the best features of data visualizations is that they look like pictures. Pictures and diagrams are more engaging than data tables or models because people who don't have prior knowledge about the topic can understand them. **Using color in a visualization helps draw your audience's attention to specific information and helps keep your audience engaged.** Colors also visually separate categories and groups. This allows you to include multiple categories in one visualization and still maintain clarity. Below, there are two versions of the same chart: one without color and one with color. The chart with color is more engaging and is easier to visually separate each education level.



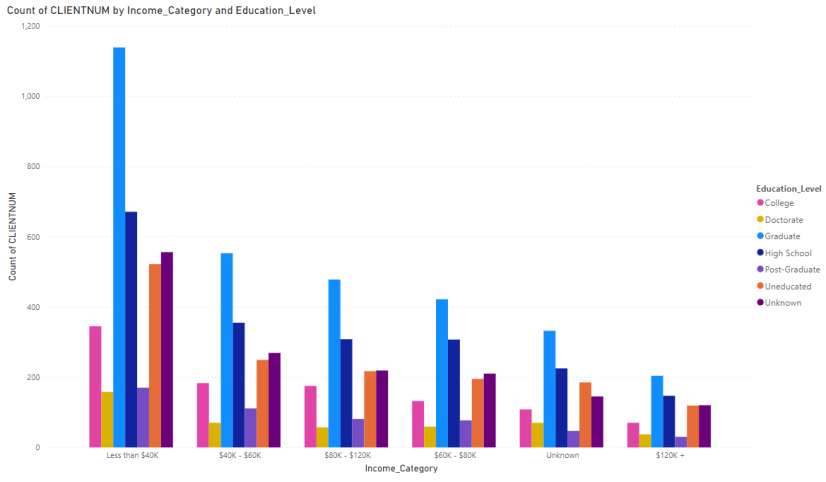


Avoid overcrowding

You've probably heard the expression "a picture is worth a thousand words," which means that a picture can communicate complex ideas better than a long description. This can be applied to visualizations of any kind. When you create a data visualization, you don't need lots of additional words and numbers; the visualization already includes a lot of that information. When creating a visualization, avoid overcrowding it. **A clean visualization draws people's attention more effectively than a crowded one**. To see this in action, take a look at the first clustered bar chart below.



There are lots of numbers crammed into the chart, making it hard to see each bar and process the overall information. The second chart, below, shows the same information, but it's much tidier. The data labels for each bar have been removed because those values can be identified using the y-axis labels.



Know your audience

If you're creating data visualizations, you're most likely planning on sharing them with someone. If that person is your colleague, and they've been working with the same dataset themselves, you can get away with more complicated and technical visualizations.

However, suppose you're planning on sharing your visualizations with a room full of people who have never worked with the dataset before. In that case, it's important to make your visualizations simple and clear.

In short, know your audience and tailor your visualizations based on them. You don't want to throw lots of technical details at people who are new to the topic or who aren't used to analyzing data. Similarly, you don't want to bore people already familiar with the data by showing them things that they already know. **Targeting your visualizations to your audience ensures that your audience receives the message that you're trying to convey.**

Data visualization tools

This lesson introduces some common data visualization tools. You'll use these tools throughout this course.

Introduction

There are many tools out there that allow you to create data visualizations with ease. In this lesson, you'll get an introduction to Excel, Tableau, and PowerPoint—three of the most popular tools used by industry professionals and individuals. Throughout this course, you'll learn how to use each of these tools to visualize data.

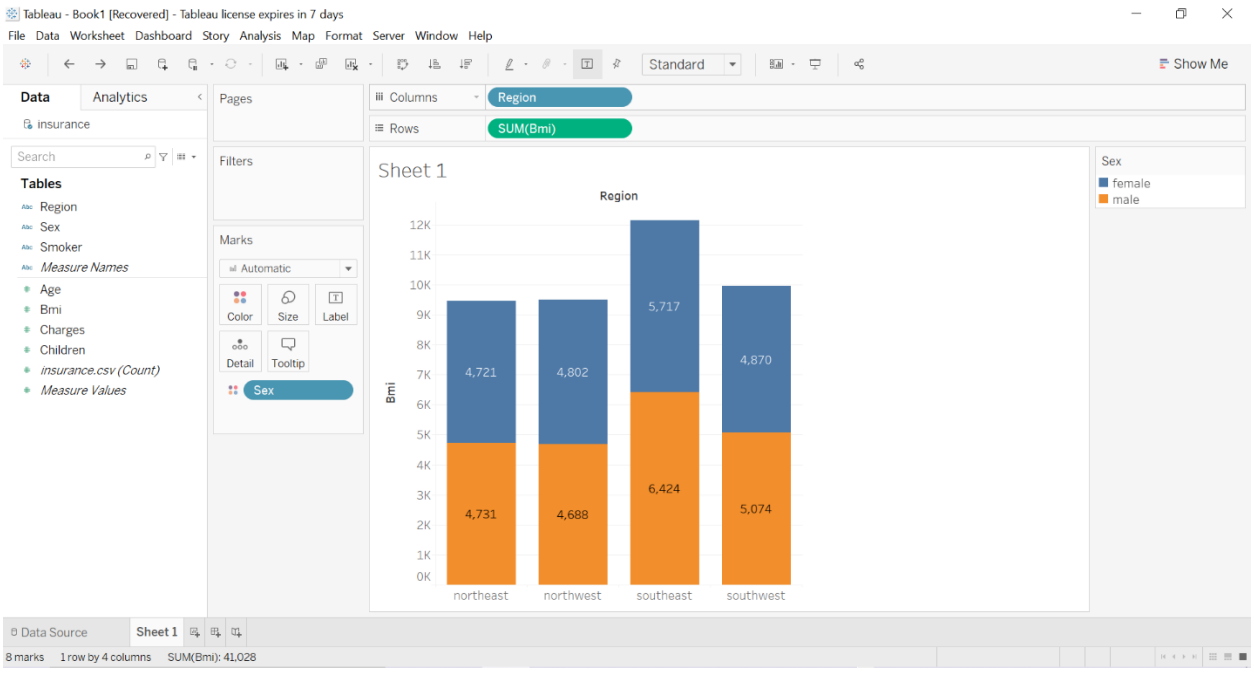
Excel

Microsoft Excel is one of the most widely used software programs for analyzing data and creating visualizations.

Tableau

Tableau is a popular data visualization tool because it creates very clean and professional-looking reports. Tableau is ideal for creating interactive visualizations and reports that update automatically as the user plays with the data.

However, Tableau is expensive, so it's mostly used by industry professionals whose companies pay for its licenses.



PowerPoint

While you can use PowerPoint to create simple visualizations, PowerPoint isn't a data visualization tool. Instead, **it's the preferred tool for creating effective presentations.**

Professionals often use PowerPoint to display their data visualizations. PowerPoint has many design options for slides, slide transitions, and styles. It's easy to insert images, charts, and animations to create engaging and customizable presentations.