## COURSE 5 - ANALYZE DATA TO ANSWER QUESTIONS

## 

## MODULE 1

Organizing data makes the data easier to use in your analysis. In this part of the course, you’ll learn the importance of organizing your data through sorting and filtering. You’ll explore these processes in both spreadsheets and SQL as you continue to prepare your data.

### **LEARNING OBJECTIVES**

* Describe what is involved in the data analysis process with reference to goals and key tasks
* Discuss the importance of organizing data before analysis with references to sorts and filters
* Describe sorting as it relates to data in a spreadsheet or database with reference to functionality and benefits
* Recall the steps involved in sorting and filtering data through the use of SQL queries

## LET’S GET ORGANIZED

### [INTRODUCTION TO GETTING ORGANIZED](https://www.coursera.org/learn/analyze-data/lecture/6chA9/introduction-to-getting-organized)

Hey there, future data analysts!

You've made a lot of progress so far. It's not an easy journey, but you're doing great. Before you started this program, something inside of you convinced you to get your Google Data Analytics Certificate.

You had an idea, did some research, and made the time to get started. Then you made the decision to commit to your goal. Now look where you are! That is something to be proud of.

Early on, we jumped right into the world of data analytics and saw how data played a part in your everyday life. You learned how to navigate spreadsheets and why structured thinking was key to solving problems. You also explored the best ways to collect and store your data. From there, you gained an understanding of clean data and data integrity. You've identified how to ask the right questions and learned to clean data.

Now we'll take your skills to the next level. Next up, you'll learn how to come up with clear and objective answers to any data question you encounter. Earlier, we learned about the data analysis process.

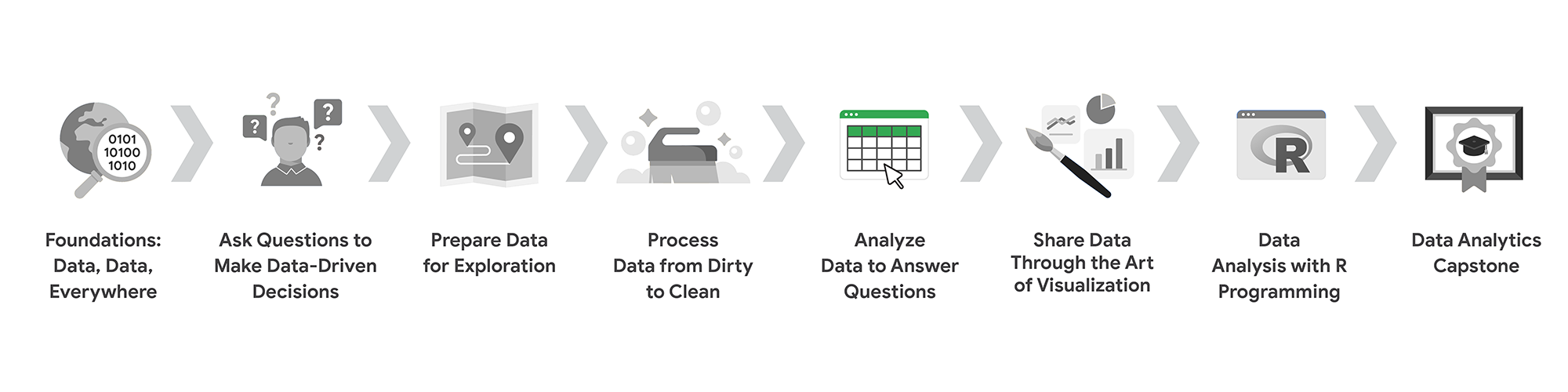
As a quick reminder, **the phases of that process are Ask, Prepare, Process, Analyze, Share, and Act**. We'll explore the Analyze phase more here, focusing on how to organize and format the data you have so that you can do all sorts of calculations.

Knowing how to analyze the data you've collected and cleaned is essential to your work as an analyst.

Before we get started, I'd like to introduce myself. My name is Ayanna, and I'm excited to be your instructor for this course. I'm a global insights manager at Google, and I've also taught at the Google Analytics Academy, which is a training resource for Google Analysts. In my job, I help advertisers determine the value of investing in Google products. When you search for something online, you'll often see an ad on the page. That's an investment an advertiser has made. I use data analysis to show advertisers the value they could gain from investing in those ads. That's what I love about being a data analyst: figuring out how to create value anytime I enter a situation. **The best way to know if you're creating value is if you have evidence.** **For me, that evidence is data**. Now that you know a little bit about my love for data, let's talk about what you'll learn here.

You'll start by covering best practices for organizing your data and the different ways you can sort through that data using spreadsheets and SQL. We'll also spend time learning three important ways to work with data that will boost your analytical skills. Then we'll talk about saving time. You'll discover tips and tricks that can help you analyze data more efficiently. Last but not least, we'll work together to identify techniques to help you be as fair and unbiased as possible. Coming up, we'll break down the basics of data analysis and bring **you one step closer to a future in data.**

### [COURSE 5 OVERVIEW: SET YOUR EXPECTATIONS](https://www.coursera.org/learn/analyze-data/supplement/Bbxbl/course-5-overview-set-your-expectations)



1. [Foundations: Data, Data, Everywhere](https://www.coursera.org/learn/foundations-data/home/welcome)
2. [Ask Questions to Make Data-Driven Decisions](https://www.coursera.org/learn/ask-questions-make-decisions/home/welcome)
3. [Prepare Data for Exploration](https://www.coursera.org/learn/data-preparation/home/welcome)
4. [Process Data from Dirty to Clean](https://www.coursera.org/learn/process-data/home/welcome)
5. **Analyze Data to Answer Questions** *(this course)*
6. [Share Data Through the Art of Visualization](https://www.coursera.org/learn/visualize-data/home/welcome)
7. [Data Analysis with R Programming](https://coursera.org/learn/data-analysis-r/home/welcome)
8. [Google Data Analytics Capstone: Complete a Case Study](https://coursera.org/learn/google-data-analytics-capstone/home/welcome)

Welcome to the fifth course in the series for the Google Data Analytics Certificate! The goal of data analysis is to make sense out of the data you collect and receive. Up until now, your focus has been on the preparations a data analyst goes through before entering the analysis phase. Specifically, in the last course, you learned about checking data for completeness and cleaning it for accuracy and reliability.

If you feel like a hiker who has climbed a great distance to get to higher ground, we are excited to tell you that you have arrived! You have reached a stage where you are ready to work directly with data. You will organize and format data. This will help you think about data in different ways. Similar to how the view from a hiker’s lookout is amazing, your view of data from this point on will be spectacular.

You will have hands-on practice organizing, sorting, filtering, formatting, converting, and combining data in spreadsheets. These are tasks you would complete in a real data analysis project. You will also learn how to sort and filter your data using SQL queries. You will be using functions and writing queries frequently as you continue your learning.

Course content

**Course 5 – Analyze Data to Answer Questions**

1. **Organize data to begin analysis.** Organizing data makes the data easier to use in an analysis. In this part of the course, you will learn the importance of organizing your data with sorting and filtering. You will explore organizing data in both spreadsheets and with SQL queries and temporary tables.
2. **Format and adjust data.** As you move closer to analyzing your data, you will want to have the data formatted and ready to go. In this part of the course, you will learn all about converting and formatting data, including how to use SQL queries to combine data. You will also discover the value of feedback and support from your colleagues and how it can lead to new insights that you can apply to your work.
3. **Aggregate data for analysis.** During an analysis, you might need to combine data to gain insights and complete business objectives. In this part of the course, you will explore the functions, procedures, and syntax to combine, or aggregate data. You will learn how to combine data within multiple cells in spreadsheets, and within multiple database tables using SQL queries.
4. **Perform data calculations.** Calculations are one of the more common tasks that data analysts perform during an analysis. In this part of the course, you will explore formulas, functions, and pivot tables in spreadsheets and SQL queries. All of these are used in data calculations. You will also learn about the benefits of using SQL to manage temporary database tables.

**What to expect**

You can expect to finish this course in about five weeks when you have completed all of the prescribed activities, which include:

* **Videos** of instructors teaching new concepts and demonstrating the use of tools
* **In-video questions** that pop up during or at the end of a video to check your learning
* **Readings** to introduce new ideas and build on the concepts from the videos
* [**Discussion forums**](https://www.coursera.org/learn/analyze-data/discussions) to discuss, explore, and reinforce new ideas for better learning
* **Discussion prompts** to promote thinking and engagement in the discussion forums
* **Hands-on activities** to introduce real-world, on-the-job situations, and the tools and tasks to complete assignments
* **Practice quizzes** to prepare you for graded quizzes
* **Hands-on activities** toreinforce learned skills for the graded quizzes
* **Graded quizzes** to measure your progress and give you valuable feedback

Hands-on activities promote additional opportunities to build your skills. Try to get as much out of them as possible. Assessments are based on the approach taken by the course to offer a wide variety of learning materials and activities that reinforce important skills. Graded and ungraded quizzes will help the content sink in. Ungraded practice quizzes are a chance for you to prepare for the graded quizzes. Both types of quizzes can be taken multiple times.

As a quick reminder, this course is designed for all types of learners, with no degree or prior experience required. Everyone learns differently, so the Google Data Analytics Certificate has been designed with that in mind. Personalized deadlines are just a guide, so feel free to work at your own pace. There is no penalty for late assignments. You'll see a **Reset deadlines** option on the Grades and Overview pages. Click it to switch to a new schedule for the course with updated deadlines.  [Contact Coursera](https://www.coursera.support/s/article/360036160591-How-to-contact-Coursera?) if you need additional assistance.

If you would like to review previous content or get a sneak peek of upcoming content, you can use the navigation links at the top of this page to go to another course in the program. When you pass all required assignments, you will be on track to earn your certificate.

Tips

* Try to complete all items in order. All new information builds on previous lessons.
* Treat every task as if it is real-world experience. Have a mindset that you are working at a company or in an organization as a data analyst. This will help you apply what you learn to the real world.
* Repeat demonstrated tasks on your own for extra practice and speed.
* Even though they aren’t graded, it is important to complete all practice items. They will help you build a strong foundation as a data analyst and prepare you for the graded assessments.
* Take advantage of all additional resources provided, including discussion forums and links to external articles for more information.
* When you encounter useful links in the course, remember to bookmark them so you can refer to the information later for study or review.
* Additional resources are free, but some sites place limits on how many articles you can access for free each month. Sometimes you can register on the site for full access, but you can always bookmark a resource and come back to view it later.
* Use a notebook or electronic journal to keep track of new formulas, functions, and syntax that you learn. That way, you will be able to refer back to any notes as needed.

Data analysts use spreadsheets and SQL queries a lot. If you perform all of the assigned activities in the application environments that are recommended, you will get a good idea of what you could be doing on the job as a data analyst.

**Updates to the course**

As you complete this course, you may notice updates to the content, like new practice materials and additional examples. These updates ensure the program provides up-to-date skills and guidance that will help you in your data analytics career. If you previously completed a graded activity, you *may* need to repeat the assessment in order to complete this course. For more information, check out [the course discussion forum.](https://www.coursera.org/learn/analyze-data/discussions)

## ORGANIZE DATA FOR ANALYSIS

### [THE ANALYSIS PROCESS](https://www.coursera.org/learn/analyze-data/lecture/olTet/the-analysis-process)

Welcome back. It's great to see you again. So let's talk about analysis. We've learned how to ask the right questions, prepare data for exploration, and then process that data to make sure it's squeaky clean.

Now it's time for the heart of the process: the actual analysis! Finally, right? **But what is analysis?** Basically, analysis is the process used to make sense of the data collected. It means taking the right steps to proceed and think about your data in different ways.

**The goal of analysis is to identify trends and relationships within the data so that you can accurately answer the question you're asking.**

To do this, you should **stick to the *4 phases of analysis***: **organize data, format and adjust data, get input from others, and transform data by observing relationships between data points and making calculations**.

Let's apply the ***4 phases of analysis*** to a real-world scenario.

**OFA - Organize, format and adjust data   
GIFO - Get Input from Others  
TD - Transform data**

Imagine you want to buy a gift for your friend Zara's wedding. The problem is you're not sure what to get her. Fortunately, you have a ton of data from her wedding website. But instead of reading all the data on her website and scrolling through a photo album of her and her partner, you go straight to the online registry, a wish list of gifts they'd enjoy.

The registry is like a dataset that you can analyze to make a decision. Now that you're checking out **organized data** in the registry, you want to make sure that the list of data, or gifts in this case, is formatted in a way that's easy to reference. **Formatting data streamlines things and saves you time**. Scrolling through hundreds of gifts can be time-consuming. Instead, you can **adjust the data** in a way that makes it easy to digest by filtering and sorting your data. You have a budget you want to stick to, so you sort the gift prices from low to high. You then filter prices to include gifts that are within your budget of $60.

You're working with a newly formatted list of data. At this point, it's good to remember that input from other people can also be really helpful when analyzing information and making decisions. You can check the list of gifts to figure out if anyone else has already bought any of the items. You realize a few of the items in the list have been purchased, and this informs your decision. When analyzing data, **gaining input from others** is important because it gives you a viewpoint you might not understand or have access to. On top of gaining input from other people, **it's also important to seek out others' perspectives early**. That way, if they predict any obstacles or challenges, you'll know beforehand. The people you'll look to for input don't have to be experts to be helpful. Sometimes all you need is for someone who's familiar with a topic or data you're considering. In our example, that would be Zara's wedding guests who are purchasing gifts from the same online registry. They probably aren't wedding gift experts, but their collaborative effort to mark off the item they purchase can help you figure out what not to buy, which will prevent Zara from getting the same gift twice. **In the end, getting input is valuable to your analysis**. This brings us to the last step of the analysis: **transforming data**.

**Transforming data** means identifying relationships and patterns between the data, and making calculations based on the data you have.

Going back to our example, you were able to find a gift that you knew Zara would like, and one that fits your budget. You were also able to choose a gift that wasn't already purchased by someone else. By finding the relationship between these data points, you chose, purchased, and sent a gift that would answer the problem you wanted to solve. The beauty of the analysis process is that you probably already analyze situations in your everyday life. Whether you're analyzing data in your personal life or in your career, these four tasks can help you make better decisions. The more you do it, the more comfortable you'll feel with the process. I hope this gives you a better understanding of the basics of analysis. As we move forward, we'll check out how to locate data for analysis, both in a spreadsheet and using SQL. When you're ready, you can go ahead. See you soon!

### 

### [AYANNA: STICKING WITH IT](https://www.coursera.org/learn/analyze-data/lecture/EzE9X/ayanna-sticking-with-it)

I think one of the coolest things about working with data at Google is that we have one of the world's most valuable datasets. People refer to **Google data as really a lens into human curiosity**.

We often look at **Google as really a proxy for what's happening in the world**. And so for many of our advertisers, they really, really value the data and the insights that we're able to give them from Google because they believe it's a proxy or a reflection of what's happening in their business or within their industry.

And so I think the value of the data that we're able to work with at Google really keeps me interested and excited about the work that I do. So I came to Google about three years ago after spending a few years in consulting. And so I was really interested in switching into a role that was really focused on sales and marketing. But at the same time, I still wanted to be able to leverage the analytical skill set that I had gained prior. This role was a great complement to the skill sets that I already had and the interest that I had in moving into the sales and marketing function. **I think one important thing for all students to realize is that no one learns this material overnight**. Many of your colleagues you may look at as experts, but most likely they've been able to gain that level of expertise through their years within the field.

**I think one of the biggest attributes that students should keep in mind is that the most important thing that they need to have throughout this learning journey is grit**.

**Grit to understand that it may be a struggle, it may be a challenge, but if you put in the work, you put in the time, these concepts will eventually click, and you'll be well on your way to becoming a data analyst**.

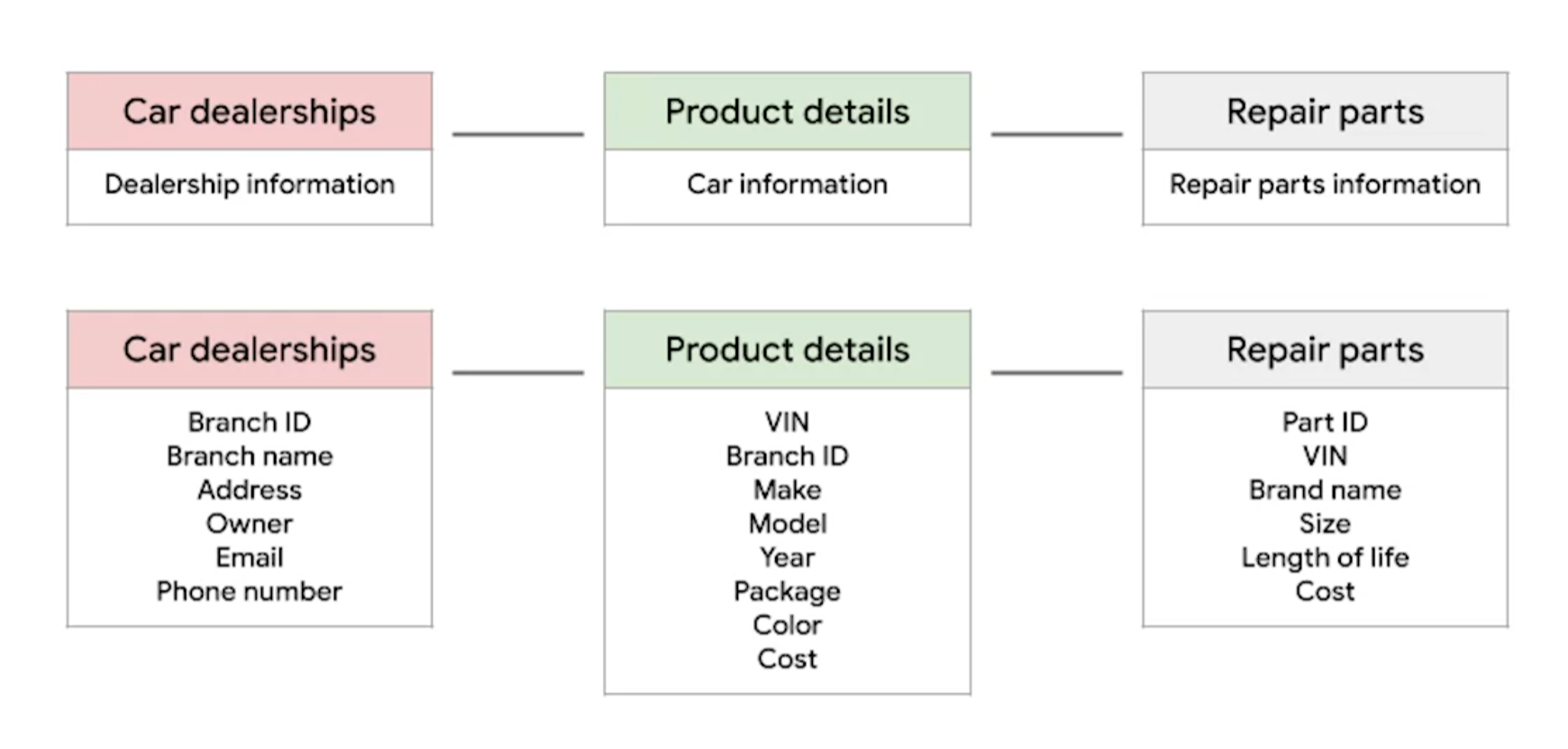
Hi, my name is Ayanna and I'm a global insights manager here at Google.

### [ALWAYS A NEED TO ORGANIZE](https://www.coursera.org/learn/analyze-data/lecture/4o37v/always-a-need-to-organize)

Hi again. Let's jump back in. Right now we're in the **Analyze phase of the data analysis process.** And even though each phase is unique, data analysts make decisions about organization throughout all of them.

That's what we're talking about here: **organization**. It's super important that you keep your data organized throughout your analysis.

How your data is classified and structured will impact your findings, whether you're working in a spreadsheet or a database. And once you know how your data is organized, you'll be able to capture or collect the information you need. Most of the data you'll use in your analysis will be organized in tables. Tables help you organize similar kinds of data into categories and subject areas that you can focus on as you analyze. For example, this basic database has tables for car dealerships, product details, and repair parts.

Each table then has several fields of data, like branch owner and the cost of repair parts. 

You can use these tables and fields to help you decide how to move forward with your analysis. The structure of this database can help you decide which data you need to pull to meet your objectives. For example, the total number of a particular brand of car sold, or a repair part for a specific make and model of a car at a certain branch.

Tables allow you to make decisions about data types. They help you to figure out what variables you need and the data type those variables should have. So if you have a database where you need to convert a data type during your analysis, you can do that by using the CAST command in SQL or any other method that you learn on the job or from your own research.

Like this example where we converted a purchase price column to be a FLOAT instead of a STRING so that it was in a numerical form we could use for calculations. **If you're performing your analysis in a spreadsheet, you want to make sure that the columns and rows are effectively organized**. You can even hide columns that you won't need for analysis or that show duplicate information. Once you have the data organized and formatted, you'll be ready to sort and filter it to find the data you need.

We'll cover sorting and filtering soon. **But for now, just know that both filters and sorts are affected by the type of data we're working with**. The bottom line is that **it's important to have your data in the right format**.

So always be prepared to adjust, no matter how far into your analysis you are. That's all for now. Coming up, we'll show you what filters are all about. Bye!

### [SORT AND FILTER DATA TO KEEP IT ORGANIZED](https://www.coursera.org/learn/analyze-data/supplement/RSNx9/sort-and-filter-data-to-keep-it-organized)

The first two phases of data analysis, **Organize data** and **Format and adjust data**, are important for data analysts because they can use these phases to manipulate their data in ways that make important patterns and trends more obvious. Most of the datasets you’ll use as a data analyst will be organized as tables. Tables are helpful because they let you manipulate and categorize your data. Having distinct categories and classifications lets you focus on, and differentiate between, the groups in your data quickly and easily.

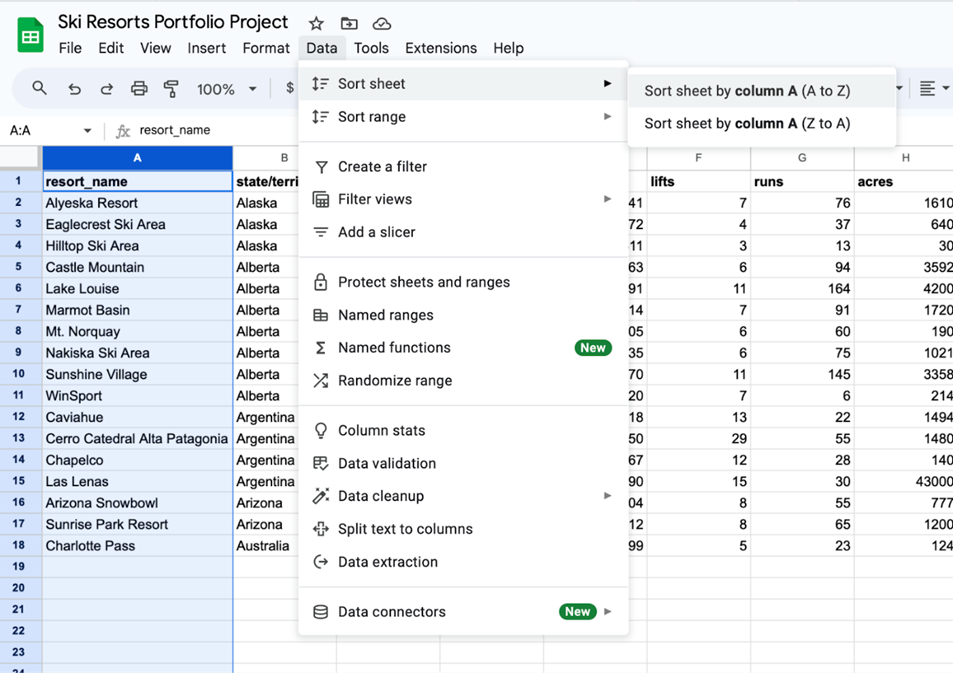
Sorting and filtering are two methods you can use to organize, format, and adjust data. For example, a filter can help you find errors or outliers so you can fix or flag them before your analysis. Outliers are data points that are very different from similarly collected data and might not be reliable values. The benefit of filtering the data is that after you fix errors or identify outliers, you can remove the filter and return the data to its original organization.

In this reading, you’ll review sorting and filtering and consider how they can be used together. You’ll also be introduced to how a particular form of sorting is done in a pivot table.

Sort data

Sorting is the process of arranging data into a meaningful order to make it easier to understand, analyze, and visualize. It ranks your data based on a specific metric you choose. You can sort data in spreadsheets, SQL databases (when your dataset is too large for spreadsheets), and tables in documents.

To rank items or create chronological lists, you can sort by ascending or descending order. Sorting arranges the data in a meaningful way and gives you immediate insights. Sorting also helps you to group similar data together by a classification. For example, if a ski resort design company wants to evaluate the resorts designed by a competitor, a data analyst can sort competitive resorts by locations, runs, acreage, and other factors. This way, the firm’s designers can visit the types of resorts they also design and gather information that could be used in its own future designs.

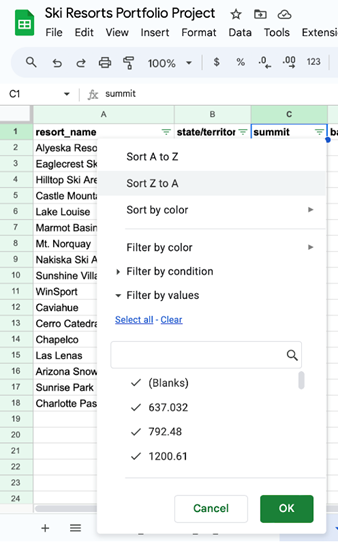
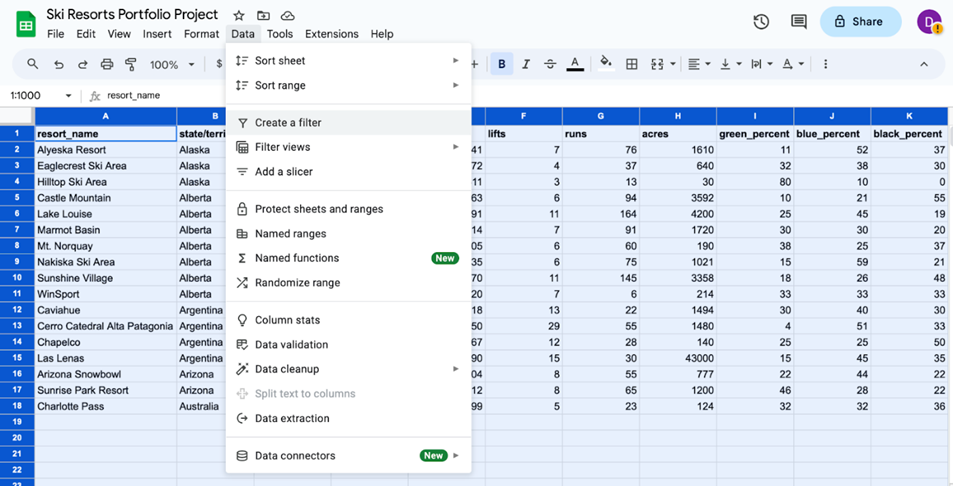


An example of sorting a spreadsheet of ski resorts, including information about resort name, state/territory/country, lifts, runs, and acres. The image taker has clicked into the Data menu option, selected Sort sheet, and is hovering over Sort sheet by column A (A to Z).

**Filter data**

Sometimes, an analysis may require only a subset of the data in your dataset. You can use a filter to show only the data that meets a specified criteria while hiding the rest. Filtering is useful when you have lots of data. You can save time by zeroing in on the data that’s important for your current analysis or the data that contains errors. Most spreadsheets and SQL databases allow you to filter your data in a variety of ways. Filtering gives you the ability to find what you are looking for without too much effort.

For example, if the ski resort design company wants to inspect specific criteria for the competitive ski resorts they intend to visit and evaluate, a data analyst can filter the competitive resort database to extract information about the number of runs compared to acreage to identify design trends or other insights.



An example of filtering data in a spreadsheet of ski resort information by specific evaluation criteria such as location, acreage, or number of runs

Sort a pivot table

A pivot table is a data summarization tool used to sort, reorganize, group, count, total, or average data. Items in the row and column areas of a pivot table are sorted in ascending order by any custom list first. If the items aren’t in a custom list, they will be sorted in ascending order by default. But, if you sort in descending order, you are setting up a rule that controls how the field is sorted even after new data points are added. For example, in the ski resort dataset, the pivot table allows locations to be sorted alphabetically by state, territory, or country.

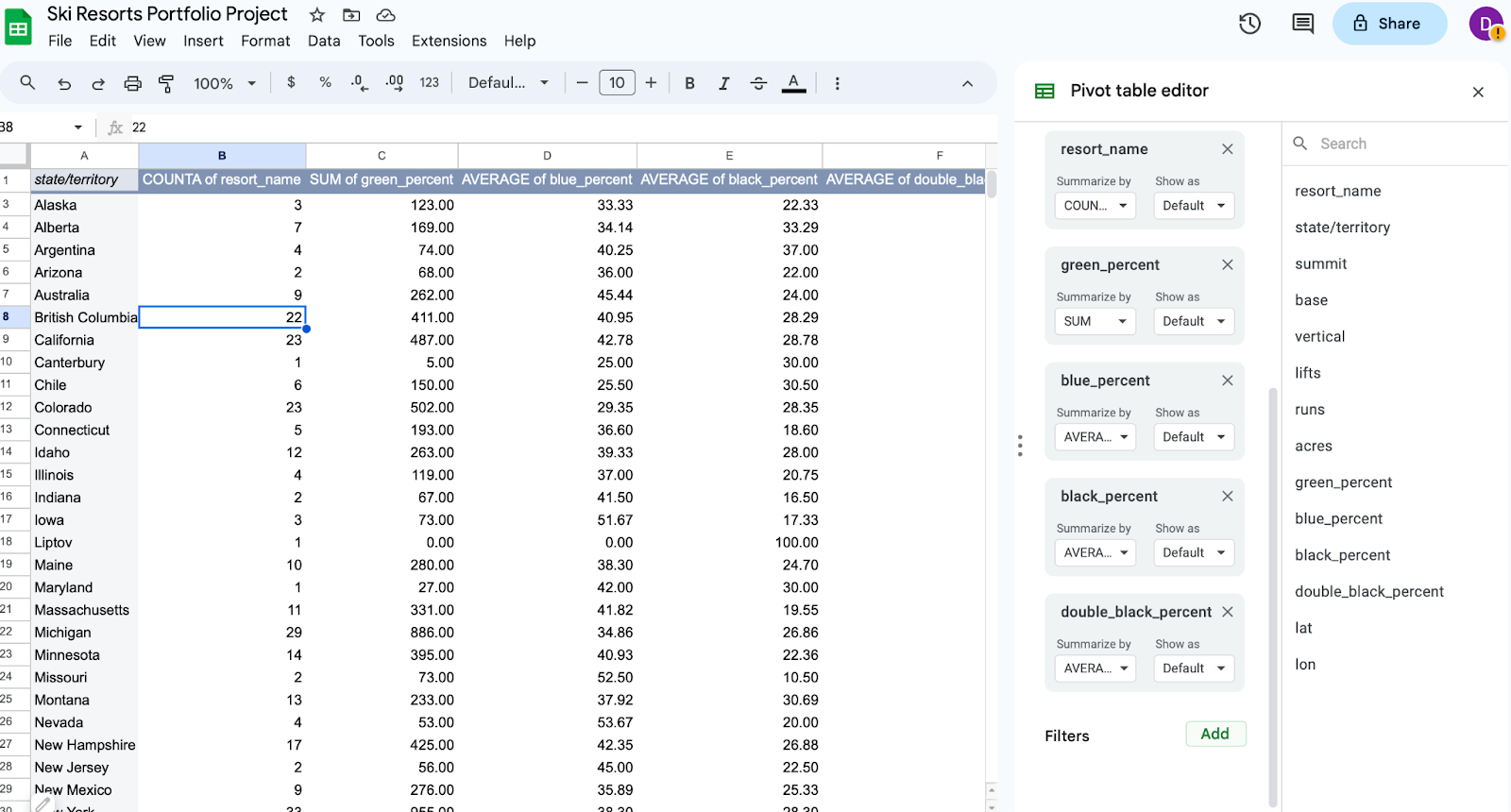


Image of a pivot table of the ski resort data, with the pivot table editor open and ready for parameters to be entered. The data is grouped by state, territory, or country.

**Key takeaways**

Data analysts filter and sort data to organize it for better understanding, analysis, and visualization. Sorting arranges data in a meaningful order, while filtering displays only data that meets specific criteria. Combining filtering and sorting allows for organizing only relevant data for analysis. Both spreadsheets and SQL databases allow for data filtering and sorting data.

### [UPLOAD THE MOVIE DATASET TO BIGQUERY](https://www.coursera.org/learn/analyze-data/supplement/sBFZn/upload-the-movie-dataset-to-bigquery)

### [STEP-BY-STEP: FILTER DATA WITH SQL](https://www.coursera.org/learn/analyze-data/supplement/CizF2/step-by-step-filter-data-with-sql)

This reading outlines the steps the instructor performs in the following video, [Filter data with SQL](https://www.coursera.org/learn/analyze-data/lecture/Y5Nmb/more-on-sorting-and-filtering). In the video, the instructor demonstrates filtering data with SQL using **WHERE** clauses.

Keep this step-by-step guide open as you watch the video. It can serve as a helpful reference tool if you need additional context or clarification while following the video steps. This is not a graded activity, but you can complete these steps to practice the skills demonstrated in the video.

**What you’ll need**

If you’d like to follow along with the instructor, you will need to log in to your BigQuery account and upload the Movies dataset. To do this, follow the instructions in the reading [**Upload the movie dataset to BigQuery**](https://www.coursera.org/learn/analyze-data/supplement/sBFZn/optional-upload-the-movie-dataset-to-bigquery).

empty alt text

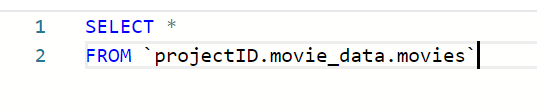
## **Example 1: Filter data in SQL**

Complete the following steps to use the **WHERE** clause to filter the database and narrow down the list to movies in the comedy genre.

1. In the BigQuery **Explorer pane**, select the **movie** dataset then the **movies** table.

2. Select the **Preview** tab from the **Details pane**.

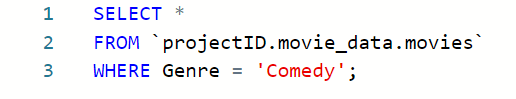
3. Select **Query** then **In new tab** and enter the following code into the query editor:



**Note:** If you’re completing this code in BigQuery, replace **projectID** in the code block to your own projectID.

4. Use the **WHERE** clause to filter the data. Enter **WHERE Genre = 'Comedy';** to filter for and select rows with 'Comedy' in the Genre column.

5. Your code should now match this code block:



6. Select **RUN** to run the query. The results display a shorter list of movies, all in the comedy genre.

Mark as completed

### [FILTER DATA WITH SQL](https://www.coursera.org/learn/analyze-data/lecture/Y5Nmb/filter-data-with-sql)

Hey, great to see you again. Earlier we talked about why you should organize your data, no matter what part of the lifecycle it's in. **Just like any collection, it's easier to manage and care for a group of things when there's structure around them.**

Now we should keep in mind that organization isn't just about making things look orderly. It's also about **making it easier to search and locate the data you need in a quick and easy way**.

As a data analyst, you'll find yourself rearranging and sifting through databases pretty often. **Two of the most common ways of doing this are with sorting and filtering**. We've briefly discussed sorting and filtering before, and it's important you know exactly what each one does.

**Sorting** is when you arrange data into a meaningful order to make it easier to understand, analyze, and visualize. Sorting ranks your data based on a specific metric that you can choose. You can sort data in spreadsheets and databases that use SQL. A common way to sort items when you're shopping on a website is from lowest to highest price, but you can also sort by alphabetical order, like books in a library. Or you can sort from newest to oldest, like the order of text messages in a phone. Or nearest to furthest away, like when you're searching for restaurants online. Another way to organize information is with a filter.

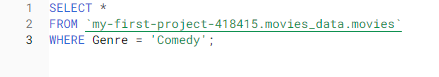
**Filtering** is showing only the data that meets a specific criteria while hiding the rest. Typically you can use filters when you want to narrow down the amount of data you want to sift through. Say you're searching for green sneakers online. To save time, you filter for green shoes only. Using a filter slims down larger data sets to smaller subsets that are relevant to what you need. Sorting and filtering are two actions you probably perform a lot online. Whether you're sorting movie showtimes from earliest to latest, or filtering your search results to just images, you're probably already familiar with how helpful they can be for making sense of data.

Now let's take that knowledge and apply it. When it comes to sifting through large, disorganized piles of data, filters are your friend. You might remember from a previous video that **you can use filters and spreadsheet programs, like Excel and Sheets, to only display data from rows that match the range or condition you've set. You can also filter data in SQL using the WHERE clause. The WHERE clause works similarly to filtering in a spreadsheet because it returns rows based on a condition you name**.

Let's learn how you can use a WHERE clause in a database. We'll use BigQuery to access the database and run our query. If you're joining us, open up your tool of choice for using SQL and reference the earlier resource on how to access the dataset. Otherwise, watch as the WHERE clause does its thing.

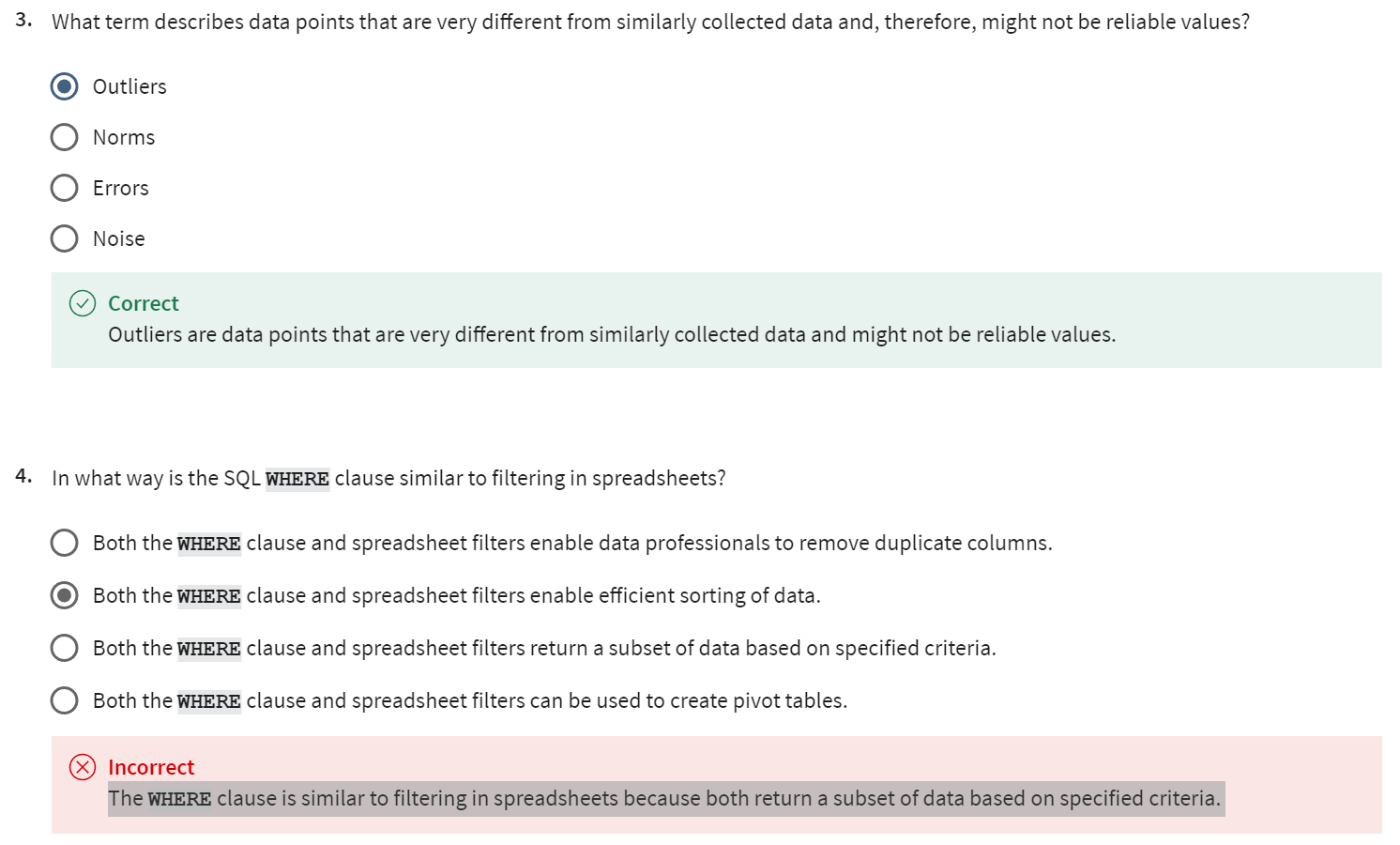
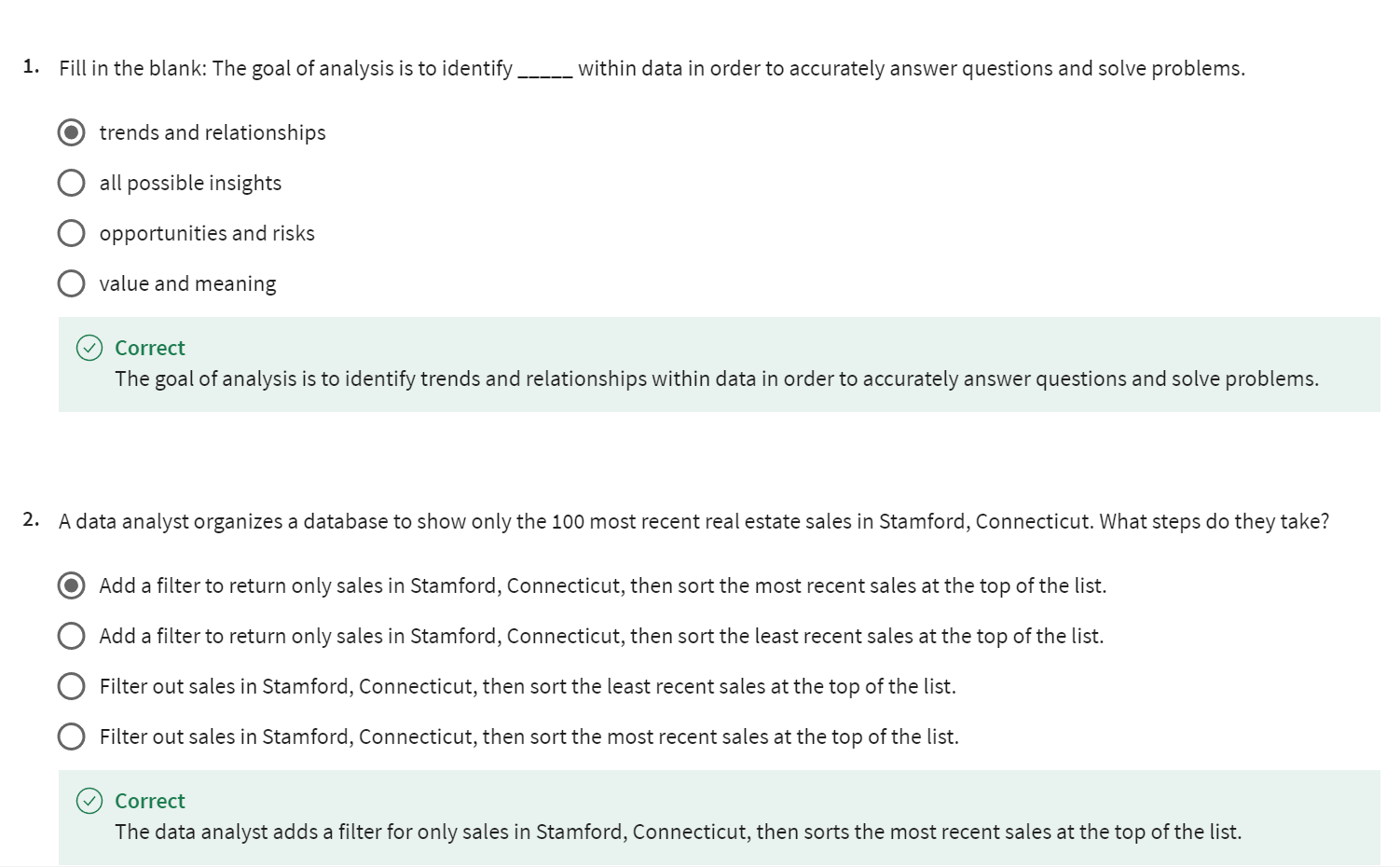
Here's the database.

You might recognize it from past videos. Basically, it's a long list of movies. Each row includes an entry for the columns named Movie\_Title, Release\_Date, Genre, Director, Cast\_Members, Budget, and Total\_Revenue. It also includes a link to the film's Wikipedia page. If you scroll down the list, the list goes on for a long time. Of course, we won't need to go through everything to find the data we want. That's the beauty of a filter! In this case, we'll use the WHERE clause to filter the database and narrow down the list to movies in the comedy genre. To start, we'll use the SELECT command followed by an asterisk. In SQL, an asterisk selects all of the data. On a new line, we'll type FROM and the name of the database: movie\_data.movies. To filter the movies by comedy, we're going to type WHERE, then list the condition, which is Genre.



Genre is a column in the dataset, and we only want to select rows where the cell in the Genre column exactly matches "Comedy." Next we'll type the equals sign and write the specific genre we're filtering for, which is comedy. Since the data in the Genre column is a string format, we have to use single or double quotations when writing it. And keep in mind that capitalization matters here, so we have to make sure that the letter casing matches the column name exactly. And now we can click Run to check out the results. What we're left with is a shorter list of comedy movies. Pretty cool, right? Here's something else you should know. You can apply multiple filters to a database. You can even sort and filter data at the same time for even more precise results. As a data analyst, knowing how to sort and filter data will make you a superstar. That's all for now. Coming up, we'll get down to the nitty-gritty of sorting functions in spreadsheets. See you there!

### [TEST YOUR KNOWLEDGE ON ORGANIZING DATA FOR ANALYSIS](https://www.coursera.org/learn/analyze-data/quiz/ED6hz/test-your-knowledge-on-organizing-data-for-analysis)

[](https://www.coursera.org/learn/analyze-data/quiz/ED6hz/test-your-knowledge-on-organizing-data-for-analysis)

## SORT DATA IN SPREADSHEETS

### [STEP-BY-STEP: SORT DATASETS IN SPREADSHEETS](https://www.coursera.org/learn/analyze-data/supplement/oafPC/step-by-step-sort-datasets-in-spreadsheets)

This reading outlines the steps the instructor performs in the following video, [Sort datasets in spreadsheets](https://www.coursera.org/learn/analyze-data/lecture/6f6R0/sorting-datasets). In this video, the instructor demonstrates how to sort data in spreadsheets with the data menu.

Keep this step-by-step guide open as you watch the video. It can serve as a helpful reference tool if you need additional context or clarification while following the video steps. This is not a graded activity, but you can complete these steps to practice the skills demonstrated in the video.

**What you’ll need**

If you’d like to access the spreadsheet the instructor uses in this video, click the link to the dataset to create a copy. If you don’t have a Google account, you may download the data directly from the attachments below.

Link to movie data starter project:  [Movie data starter project](https://docs.google.com/spreadsheets/d/1FLaUmMn62YlHYihV6pK1DJqWcFYCnuoqoxFWmm_o5b0/template/preview).

OR

[Movie Data Starter Project](https://d3c33hcgiwev3.cloudfront.net/NrzEyoWGRhix-2gQUzoZag_31090f90399f4347b98b63d02b95a9e1_Movie-Data-Starter-Project.xlsx?Expires=1712707200&Signature=EqTMQX9lvaDHdMvXxreLepIedLlgm182La4WETFJMNNmmvADYFQEvdxatLM-BX~WJCysITfY6F7funvw~FICWHBUcIDOle0KdtvMC-6pZ2g0fxkl4Iis9JBv1EGOkRB0OSTxRpOb9iUN4ySaExhE6oCSyuYNJ5XFzvlGVjtkh4M_&Key-Pair-Id=APKAJLTNE6QMUY6HBC5A)

[XLSX File](https://d3c33hcgiwev3.cloudfront.net/NrzEyoWGRhix-2gQUzoZag_31090f90399f4347b98b63d02b95a9e1_Movie-Data-Starter-Project.xlsx?Expires=1712707200&Signature=EqTMQX9lvaDHdMvXxreLepIedLlgm182La4WETFJMNNmmvADYFQEvdxatLM-BX~WJCysITfY6F7funvw~FICWHBUcIDOle0KdtvMC-6pZ2g0fxkl4Iis9JBv1EGOkRB0OSTxRpOb9iUN4ySaExhE6oCSyuYNJ5XFzvlGVjtkh4M_&Key-Pair-Id=APKAJLTNE6QMUY6HBC5A)

[empty alt text](https://d3c33hcgiwev3.cloudfront.net/NrzEyoWGRhix-2gQUzoZag_31090f90399f4347b98b63d02b95a9e1_Movie-Data-Starter-Project.xlsx?Expires=1712707200&Signature=EqTMQX9lvaDHdMvXxreLepIedLlgm182La4WETFJMNNmmvADYFQEvdxatLM-BX~WJCysITfY6F7funvw~FICWHBUcIDOle0KdtvMC-6pZ2g0fxkl4Iis9JBv1EGOkRB0OSTxRpOb9iUN4ySaExhE6oCSyuYNJ5XFzvlGVjtkh4M_&Key-Pair-Id=APKAJLTNE6QMUY6HBC5A)

## **Example 1: Sort a sheet with the menu**

Use the menu to quickly and easily sort an entire sheet, keeping rows together.

1. Open the [Movie data starter project](https://docs.google.com/spreadsheets/d/1FLaUmMn62YlHYihV6pK1DJqWcFYCnuoqoxFWmm_o5b0/template/preview) spreadsheet.
2. Select cell **B** to highlight all of **column B**.
3. Select **Data** from the menu.
4. Select **Sort sheet**.
5. Select **Sort sheet by column B (A to Z)**.
6. Notice that the movies in your spreadsheet are now arranged in chronological order based on the release date.

**Note:** In the video, the sorted results are not entirely correct as shown. While the release dates appear to be sorted in ascending order, many of the movie titles are also in alphabetical order. As a data analyst, you may suspect that this is too much of a coincidence and that something is not quite right with the data. In this example, the movie titles in column A may have been previously sorted alphabetically and the data wasn’t restored to the original unsorted format.

## **Example 2: Sort data in a specific column with the menu**

Use the menu to sort one column without affecting how the rest of the sheet is arranged. Each row in a table describes a single observation, so if you sort only one column you can introduce errors throughout your dataset. Use caution when using this option!

1. Select cell **A** to highlight all of **column A**.
2. Select **Data** from the menu.
3. Select **Sort range**.
4. Select **Sort range by column A (A to Z)**.
5. Notice that the movie data across the rows is now jumbled because sorting a single column in a sheet doesn’t keep data in a row together.

**Note:** As you’ve learned, each row in a table describes a single observation. Here, the values in Column A (movie titles) were sorted in A-to-Z order, but the rest of the sheet wasn’t sorted. For example, before sorting by column A, *The Devil Inside* was listed as having a release date of 2012-01-06. After sorting only the movie titles in Column A in A-to-Z order, however, *The Devil Inside* is listed as having a release date of 2015-06-16. This is incorrect.

### 

### [SORT DATA IN SPREADSHEETS](https://www.coursera.org/learn/analyze-data/lecture/6f6R0/sort-data-in-spreadsheets)

Hey there, data pro! Happy to see you back and ready to explore more of the organizational side of data analysis. In this video, we'll learn how to sort data in spreadsheets.

We've done some sorting in spreadsheets earlier in the program. Now it's time to build on what we've covered and introduce some more advanced sorting techniques.

**Sorting is amazing**. Not only does it **add order and meaning to your spreadsheets, it also gives you the power to reimagine data altogether**. When you sort data based on a specific metric, you can uncover new patterns and relationships within datasets you might not have otherwise noticed. This is especially true for spreadsheets, which you'll use a lot in your work as a data analyst.

**Knowing how to sort data in spreadsheets can make you a stronger and more confident analyst**. In many ways, sorting relies on your creativity to reimagine the information you have in front of you. In spreadsheets, you can sort data by ascending or descending order using numbers or letters. If cells are labeled with color, you can sort them by color, too. When sorting data in a spreadsheet, you can choose "**Sort sheet**" or "**Sort range**."

If "**Sort sheet**" is applied, all of the data in a spreadsheet is sorted by the conditions of a single column, but the related information across each row stays together.

On the other hand, "**Sort range**" doesn't keep the information across rows together. When you sort a range, you're selecting a specific collection of cells or the range that you want the sorting limited to. Nothing else on the spreadsheet gets rearranged but the specified cells. There's two methods for sorting spreadsheet data: one involves using the menu; the other involves writing out the sort function. For now, we'll focus on sorting with the menu. We'll get to writing out functions later on. Now, depending on the program you use, the process might seem slightly different, but the instructions and concepts we discuss will be basically the same.

Back to sorting with the data menu. To give you an idea of how to do it, we'll use the movie spreadsheet. Let's check it out. In this example, we'll sort movies by release date. We'll head to column B, which is listed as "Release Date." Click on the "B" button to highlight all the cells in the column. From there, we'll head to the Data tab in the menu.

Now you have two choices: **sort a sheet or a range of data**. You'll notice that we've selected just the release dates, but these release dates are specifically related to the movies in their row.

In this case, you want the release date and the movie title to stay in the same row as you sort because they're related. To do this, you'll want to "**Sort sheet**." **This will keep all the data together by row, no matter how you sort it**. Depending on the order you want the release dates to be in, you can sort from A to Z, which will also rank the dates numerically. Or you can sort from Z to A, which will sort data the opposite way. Since we want the release dates to be in order, we'll click "Sort sheet by column B" from A to Z. And there you go. You just sorted a sheet of data using the menu. Now the movies are arranged in chronological order based on release date.

Let's say you want to sort data in a specific column, but don't need the cells in that column tied to a specific row of information. Instead, you want to isolate the column's data and sort it on its own without affecting how the rest of the sheet's arranged. For fun, we'll use the Movie Title column in this example. First, we'll select the column we want to sort: column A. Clicking on column A highlights all the cells in the column which contain the movie titles. Then we'll go to the menu and click Data because we're isolating the column from the rest of the sheet. When we sort this time, we'll click "Sort range by column A." For this example, we'll sort the movie titles alphabetically from A to Z. And that's it! You'll notice that "Sort range" doesn't keep the rows together, so the data are a bit jumbled.

**You'll probably end up using "Sort sheet" more often**, but it's important to understand them both so you don't accidentally get them mixed up.

You've just sorted data in a spreadsheet using the menu, and you've learned how to sort data by an entire sheet or by a range of cells. That's something you'll be able to take with you wherever you go as a data analyst. Coming up, we'll learn about the second way to sort in a spreadsheet: by writing out a function. We'll also take sorting to the next level by custom sorting your data. See you there!

### 

### [STEP-BY-STEP: USE THE SORT FUNCTION IN SPREADSHEETS](https://www.coursera.org/learn/analyze-data/supplement/fyqMb/step-by-step-use-the-sort-function-in-spreadsheets)

This reading outlines the steps the instructor performs in the next video, [Use the SORT function in spreadsheets](https://www.coursera.org/learn/analyze-data/lecture/K6WB8/the-sort-function). In this video, the instructor demonstrates how you can use the **SORT** function to arrange data into a meaningful order to make it easier to understand, analyze, and visualize.

Keep this step-by-step guide open as you watch the video. It can serve as a helpful reference if you need additional context or clarification while following the video steps. This is not a graded activity, but you can complete these steps to practice the skills demonstrated in the video.

**What you’ll need**

If you’d like to follow along with the examples in this video, choose a spreadsheet tool. Google Sheets or Excel are recommended.

To access the spreadsheet the instructor uses in this video, click the link to the template to create a copy of the dataset. If you don’t have a Google account, download the data directly from the attachments below.

Link to dataset: [Party plan spreadsheet](https://docs.google.com/spreadsheets/d/1L1Z6b3X9WCpwzisHcxvjC5-Qett-BOMo9yg8Cz36NME/template/preview)

OR

[Party Plan Spreadsheet](https://d3c33hcgiwev3.cloudfront.net/SifBvA71QgKf-sX-0JyvSw_7a45f15631e54e65aa3541c76da0d3e1_Party-Plan-Spreadsheet.xlsx?Expires=1712707200&Signature=JDDDvnNklzlYBmcaL0nuc~nE3Lg8UZR3w1fswohwlA4SZqdnJj3sOCNwTZMMg-05L~dInRG9GeKllU2Wv~FiODfr~NVkk7dtwEaFmbzPNaKc5cSvGIxXw-rDrzO3Ut2aYc3TCz02-N3WdpyoQc2~br590p~11jKcS4U4q5EcauM_&Key-Pair-Id=APKAJLTNE6QMUY6HBC5A)

[XLSX File](https://d3c33hcgiwev3.cloudfront.net/SifBvA71QgKf-sX-0JyvSw_7a45f15631e54e65aa3541c76da0d3e1_Party-Plan-Spreadsheet.xlsx?Expires=1712707200&Signature=JDDDvnNklzlYBmcaL0nuc~nE3Lg8UZR3w1fswohwlA4SZqdnJj3sOCNwTZMMg-05L~dInRG9GeKllU2Wv~FiODfr~NVkk7dtwEaFmbzPNaKc5cSvGIxXw-rDrzO3Ut2aYc3TCz02-N3WdpyoQc2~br590p~11jKcS4U4q5EcauM_&Key-Pair-Id=APKAJLTNE6QMUY6HBC5A)

[empty alt text](https://d3c33hcgiwev3.cloudfront.net/SifBvA71QgKf-sX-0JyvSw_7a45f15631e54e65aa3541c76da0d3e1_Party-Plan-Spreadsheet.xlsx?Expires=1712707200&Signature=JDDDvnNklzlYBmcaL0nuc~nE3Lg8UZR3w1fswohwlA4SZqdnJj3sOCNwTZMMg-05L~dInRG9GeKllU2Wv~FiODfr~NVkk7dtwEaFmbzPNaKc5cSvGIxXw-rDrzO3Ut2aYc3TCz02-N3WdpyoQc2~br590p~11jKcS4U4q5EcauM_&Key-Pair-Id=APKAJLTNE6QMUY6HBC5A)

## **Example 1: Sort guests by table**

Use a spreadsheet function to sort guests by the table to which they’re assigned.

1. Open the [Party plan spreadsheet](https://docs.google.com/spreadsheets/d/1L1Z6b3X9WCpwzisHcxvjC5-Qett-BOMo9yg8Cz36NME/template/preview).
2. Select cell G1 and enter the equal sign **=**.
3. Enter **SORT** followed by an open parenthesis **(** to begin the **SORT** function.
4. Enter the first cell from the Guest column of the spreadsheet, **A2**, followed by a colon [**:**], then enter the last cell from the Sent Invitation column, **D6**, followed by a comma. **A2:D6** is the range of cells over which this function will sort.
5. Enter **2** followed by a comma to specify the column by which to sort the data. Note that the function doesn't recognize letters, so use the column’s number. Column A corresponds to 1, B corresponds to 2, and so on.
6. Enter **TRUE** followed by a close parenthesis **)**. ***TRUE means the function will return results in ascending order, so the tables will be listed starting with table one***. **Note:** To return results in descending order, enter **FALSE**.
7. Press **Enter** (Windows) or Return (Mac).

The party guests are now sorted by the table to which they’re assigned.

## **Example 2: Customized sort order**

When you **sort data in a spreadsheet using multiple conditions** with a customized sort order, **data is sorted based on the order in which the conditions are applied to the data**.

To sort party guests by whether or not they've been sent an invitation and to list the guests alphabetically:

1. Highlight all the data in the party plan spreadsheet: **cells A1 to D6**.
2. In the menu, select **Data > Sort range > Advanced range sorting options**. This opens the **Sort range** from A1 to D6 dialog box.
3. Check the **Data has a header row** checkbox to make sure column titles aren’t included in the sorted data.
4. From the **Sort by** drop-down list, select **Sent Invitation**.
5. Select the **A to Z** radio button to make sure the *No* responses are listed first and the *Yes* responses are listed second.
6. Select **Add another sort column** to add the additional sorting condition.
7. From the **Sort by** drop-down list, select **Guest Names** to list guests alphabetically.
8. Select the **A to Z** radio button.
9. Select **SORT**.

This returns your customized sort order, which lists the *No* invitations and those guests alphabetically, followed by the *Yes* invitations and those guests alphabetically.

### [USE THE SORT FUNCTION IN SPREADSHEETS](https://www.coursera.org/learn/analyze-data/lecture/K6WB8/use-the-sort-function-in-spreadsheets)

Happy to have you back. Earlier in the program, we covered some basics of sorting in spreadsheets. We learned the differences between sorting a range and an entire sheet, and how to sort a spreadsheet using the menu. Now that we've laid the groundwork, it's time to move on to more advanced ways to sort information.

We've talked about how there's two methods of sorting data in spreadsheets.

**The first method uses the Data tab in the menu of your spreadsheet program.**

**The second way to store information in a spreadsheet is by writing a SORT function.**

In spreadsheets, functions are preset commands that perform a specific process. So in this case, the SORT function, as you might be able to guess, sorts your data.

Let's check out this spreadsheet of party plans to witness the SORT function in action. The first arranged set of data is our original dataset of guests and some information about them. So let's say you want to sort the party guests by table to get an idea of who will be sitting where. To do that, start by typing a function in an empty cell. Just like any function, you do this by typing the equal sign, and then write SORT after it.

After your first open parenthesis, reference the first cell in which data is collected from. In this case, that's A2.

Then you'll include a colon and write the last cell you want included in the function, which is D6.

A2 colon D6 is the range for this function. Next, write a comma to separate the range from what we're sorting by, which is column B.

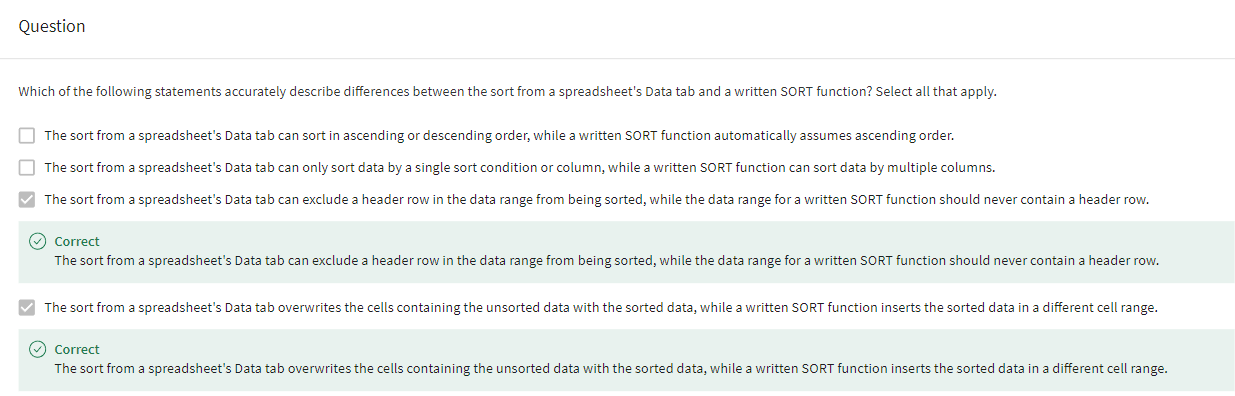
You should keep in mind that this part of the function doesn't recognize column letters. So in this case, we use the corresponding number instead, which is 2, since column B is the second column in our range.

Now add another comma.

In this next part you'll need to decide whether you want the data in this column to be in ascending or descending order. A TRUE statement is in ascending order, and FALSE is descending. Because we want the tables to be listed starting from table number one, we'll write TRUE for ascending, and then end the function with a closed parenthesis. Now, let's see our function play out.

Our party guests are now sorted by which table they're seated. Once you have an idea of the data you want to be sorted and how, applying functions to your data is simple. Now, you have two different tools in your tool belt for sorting data. After you've tackled writing SORT functions, you'll want to customize sort orders, too. A customized sort order is when you sort data in a spreadsheet using multiple conditions. This means that sorting will be based on the order of the conditions you select. Let's go back to our party spreadsheet. Imagine you want the guests to be sorted by whether or not they've been sent an invitation. And based on that, we want those guest names to be listed alphabetically. You can do that easily with the "Sort range" option under Data.

First, highlight all the data in the set from cells A1 to D6. Then under the Data tab in the menu, click "Sort range."In this case, check "Data has a header row," which makes sure that the title of the column isn't mixed into the sorting. Then, we'll make sure it's being sorted by "Sent invitation." Here, we want the "No" responses first and the "Yes" responses second, so we'll make sure A to Z is clicked to sort the responses in that order. Because we want to add an additional sorting condition, we'll now click on "Add another sort column." The guest names should be in alphabetical order. So let's select "Guest Names" and sort from A to Z.

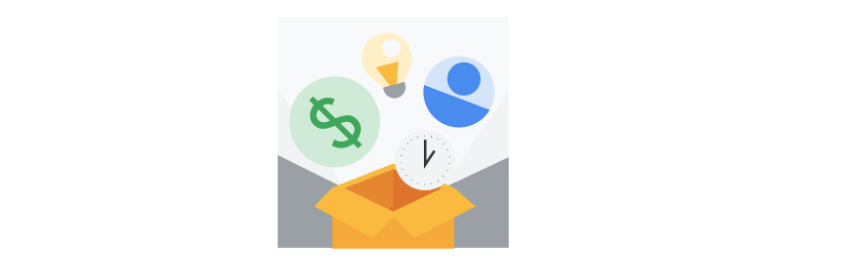


Then we'll click Sort. And voilà! You've officially applied a custom sort order like a champ.

Okay, so you've tackled sorting in spreadsheets by sheet, by range, through the menu, and by using a function. On top of all that, you've added to your organizational skills by learning how to create custom sort orders. Pretty soon you'll learn another powerful tool: how to sort data using SQL. Even though databases can sometimes be a lot to digest, learning these skills gives you the power to rearrange data in a way that makes sense to you. Once you've sorted data in a way that really clicks, you'll understand why it's so valuable to you as a data analyst. Bye for now!

### [SORT AND FILTER IN SHEETS AND EXCEL](https://www.coursera.org/learn/analyze-data/supplement/6xPhu/sort-and-filter-in-sheets-and-excel)

In this reading, you’ll examine the sorting and filtering options in Google Sheets and Microsoft Excel. Both offer basic sorting and filtering from set menu options. But if you need more advanced sorting and filtering capabilities, you can use their respective **SORT** and **FILTER** functions.



## **Sort and filter in Sheets**

Sorting in Google Sheets helps you quickly spot trends in numbers and text. For example, as the vice president of sales at a candy company, you want to improve chocolate sales in lower-performing regions—your company makes delicious chocolate and you know sales can improve. As a first step, you examine this by calculating gross (total) revenue of chocolate by sales region. In this case, you could sort the gross revenue column in **descending** (Z to A) order to find the top performing regions at the top, or sort the gross revenue column in **ascending** (A to Z) order to find the lowest performing regions at the top. Then, you can look at patterns in the best and worst regions to explore how to increase sales in the lower-performing regions.

If you want to learn more about the set menu options for sorting and filtering, start with these resources:

* [Sort and filter data](https://support.google.com/docs/answer/3540681) (Google Help Center): instructions to sort data in alphabetical or numerical order and create filter views
* [Sort data by selecting a range of data in a column](https://www.youtube.com/watch?v=VcRBHXBMKBU): video of steps to achieve the task
* [Sort a range of data using sort criteria for multiple columns](https://blog.sheetgo.com/google-sheets-formulas/sort-formula-google-sheets/): technical tip instructions by SheetGo company to sort data across multiple columns

In addition to the standard menu options, the **SORT** function allows you to do more advanced sorting. Use this function to create custom sorting rules. You can sort the rows of a given range of data by the values in one or more columns. And you can set the sort criteria for each column. Refer to the [SORT function](https://support.google.com/docs/answer/3093150?hl=en) page for the syntax.

Like the **SORT** function, use the [FILTER function](https://support.google.com/docs/answer/3093197?hl=en) to filter by any matching criteria you like. This creates a custom filter.

As you’ve learned, you can filter data and then sort the filtered results. Using the **FILTER** and **SORT** functions together in a range of cells can programmatically and automatically achieve these results for you.

## **Sort and filter in Excel**

You can also sort in ascending (A to Z) and descending (Z to A) order in Microsoft Excel. Excel offers **Smallest to Largest** and **Largest to Smallest** sorting options when you’re working with numbers.

Similar to the **SORT** function in Google Sheets, Excel includes custom sort capabilities that are available from the menu. After you select the data range, click the **Sort & Filter** button to select the criteria for sorting. You can even sort by the data in rows instead of by the data in columns if you select **Sort left to right** under **Options**. (**Sort top to bottom** is the default setting to sort the data in columns.)

If you want to learn more about sorting and filtering in Excel, start with these resources:

* [Sort data in a range or table](https://support.microsoft.com/en-us/office/sort-data-in-a-range-or-table-62d0b95d-2a90-4610-a6ae-2e545c4a4654) (Microsoft Support): instructions to perform sorting in a variety of use cases
* [Excel training: sort and filter data](https://support.microsoft.com/en-us/office/video-sort-data-in-a-range-or-table-ffb9fcb0-b9cb-48bf-a15c-8bec9fd3a472#ID0EAABAAA=Transcript) (Microsoft Support): sorting and filtering videos with transcripts
* [Excel: sorting data](https://www.youtube.com/watch?v=Ep5q1cUhQas): video of how to use the **Sort & Filter** and **Data** menu options for sorting

Excel also has [SORT](https://support.microsoft.com/en-us/office/sort-function-22f63bd0-ccc8-492f-953d-c20e8e44b86c), [SORTBY](https://support.microsoft.com/en-us/office/sortby-function-cd2d7a62-1b93-435c-b561-d6a35134f28f), and [FILTER](https://support.microsoft.com/en-us/office/filter-function-f4f7cb66-82eb-4767-8f7c-4877ad80c759) functions. Explore how you can use these functions to automatically sort and filter your data in spreadsheets without having to select any menu options at all.

## **Sort and filter manually with menus and buttons**

Both Sheets and Excel feature menu options designed to let you sort and filter without using functions. For example, after selecting the data you’d like to sort in Google Sheets, you can choose **Data > Sort sheet** or **Data > Sort range** to sort that data. To filter the data, select all the columns and rows and choose **Data > Create a filter**. In Excel, you can use the **Sort & filter** button to bring up a user-friendly interface that guides you through sorting or filtering.

Finally, when using menus or buttons, here are a couple of best practices:

* **Back up or make a copy of your data before making major changes.**
* **When filtering data, keep in mind that other users may also be accessing the spreadsheet. For example: Filters in Google Sheets can affect all viewers, so you should use filter views for personal filtering.**

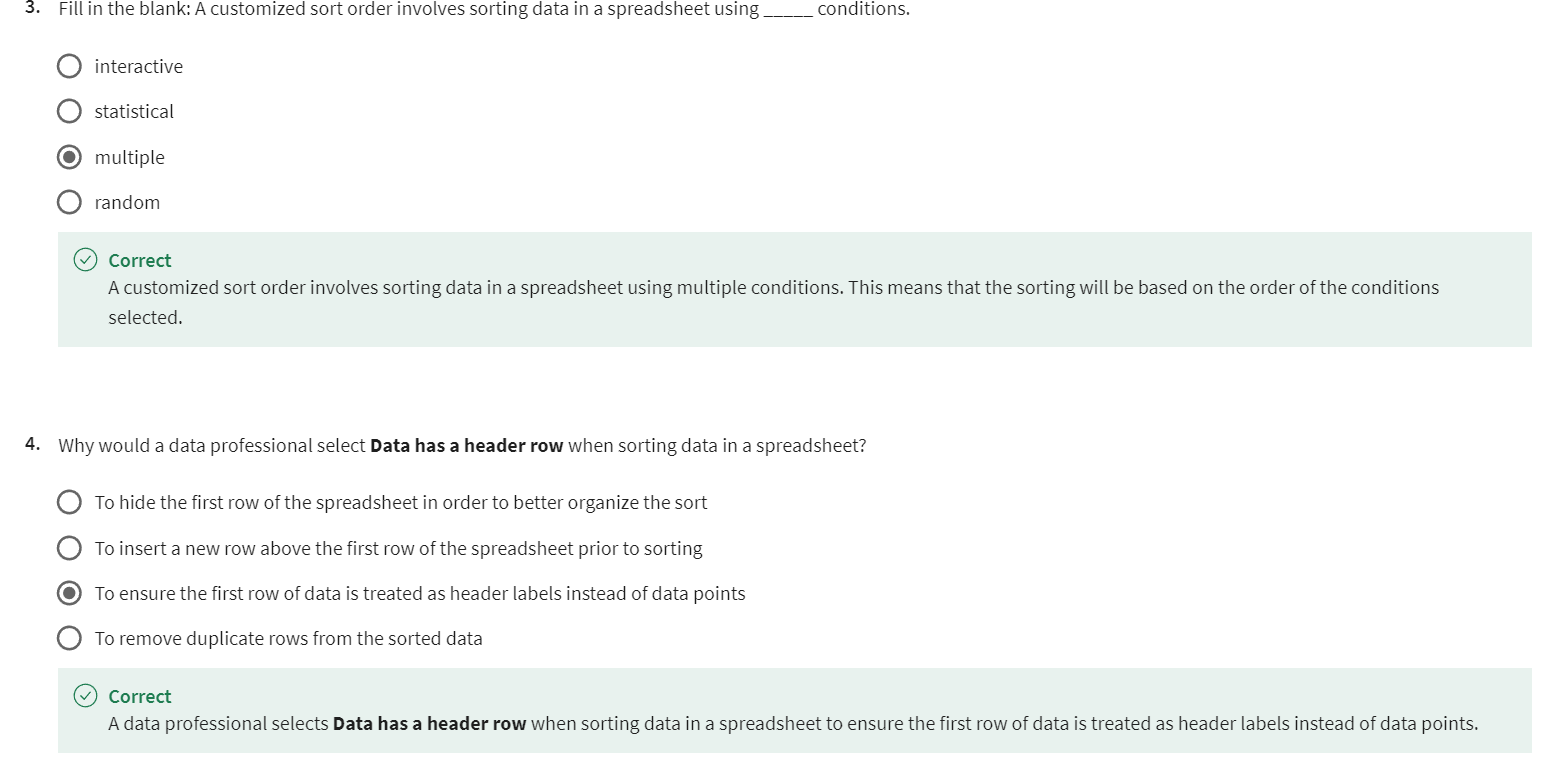
## **Key takeaways**

As you’ve learned, you can sort and filter data in Google Sheets and Excel with functions or by using menus and buttons. Sorting data is the process of arranging data into a meaningful order to make it easier to understand, analyze, and visualize. This can help you identify trends in the data. Filtering is the process of showing only the data that meets a specified criteria while hiding the rest. Once you’ve filtered data, you can sort it to find trends within those criteria. Menus and buttons offer the ability to do basic sort and filter functions, but you’ll need to use a function for custom sorting and filtering.

## 

### [TEST YOUR KNOWLEDGE ON SORTING DATA IN SPREADSHEETS](https://www.coursera.org/learn/analyze-data/quiz/XI0Vz/test-your-knowledge-on-sorting-data-in-spreadsheets)

### 



## 

## SORT DATA USING SQL

### [STEP-BY-STEP: SORT DATA WITH SQL](https://www.coursera.org/learn/analyze-data/supplement/wq3GZ/step-by-step-sort-data-with-sql)

This reading outlines the steps the instructor performs in the following video, [Sort data with SQL](https://www.coursera.org/learn/analyze-data/lecture/P6Yu3/sorting-queries-in-sql). In the video, the instructor sorts data with SQL using the **ORDER BY** command. Then, the instructor uses **WHERE** and **ORDER BY** together to filter and sort data.

Keep this step-by-step guide open as you watch the video. It can serve as a helpful reference tool if you need additional context or clarification while following the video steps. This is not a graded activity, but you can complete these steps to practice the skills demonstrated in the video.

## **What you’ll need**

If you’d like to follow along with the instructor, you will need to log in to your BigQuery account and upload the Movies dataset. To do this, follow the instructions in the reading [**Upload the movie dataset to BigQuery**](https://www.coursera.org/learn/analyze-data/supplement/sBFZn/optional-upload-the-movie-dataset-to-bigquery).

empty alt text

## **Example 1: Sort data by one column**

The **ORDER BY** command sorts data by column in a database. By default, the data is sorted in ascending order.

1. In the BigQuery **Explorer pane**, select the movie **dataset**, then the **movies** table.

2. Select the **Preview** tab from the **Details pane**.

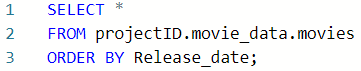
3. Select **Query** then **In new tab** and enter the following code into the query editor:



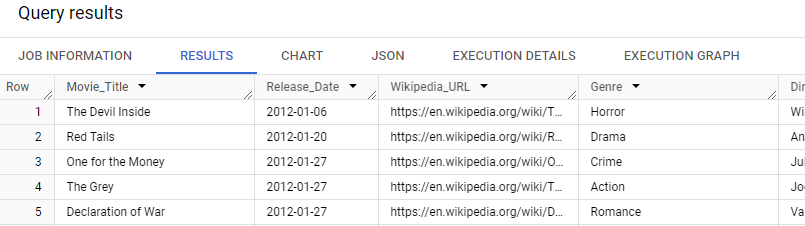
**Note:** If you’re completing this code in BigQuery, replace **projectID** in the code block to your own projectID.

4. Use the **ORDER BY** command to sort the data. Enter **ORDER BY Release\_Date;** to sort by the **Release\_Date** column.

5. Your code should now match this code block:



6. Select **RUN**. The results return all the movies in the database sorted from oldest to newest.

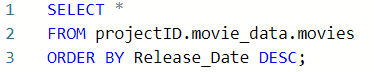


## **Example 2: Sort data in descending order**

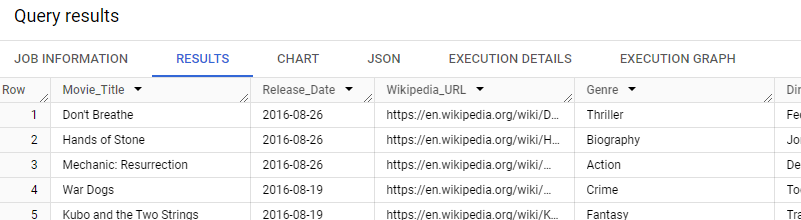
To use the **ORDER BY** command to sort data by descending order, specify **DESC** at the end of the **ORDER BY** command.

1. Enter **DESC** after **ORDER BY Release\_Date** in the query editor.

2. Your code should now match this code block:



3. Select **RUN**. The results include the same list of movies, this time sorted from newest to oldest.



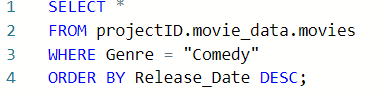
## **Example 3: Filter and sort data in descending order**

Use **WHERE** and **ORDER BY** together to filter, then sort, data.

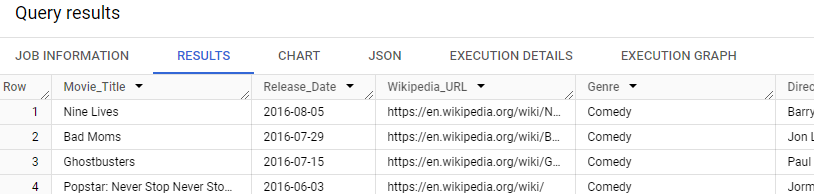
1. Select the beginning of the **ORDER BY** line and press Enter (Windows) or Return (Mac) to add a line between the **FROM** and **ORDER BY** clauses. The **ORDER BY** command is written on the last line to ensure all data is sorted.

2. Select the line you added. Enter **WHERE Genre = "Comedy"** to filter for rows in which the Genre column exactly matches **"Comedy"**.

3. Your code should now match this code block:



4. Select **RUN** to run the query. The results include only comedy movies sorted from newest to oldest.



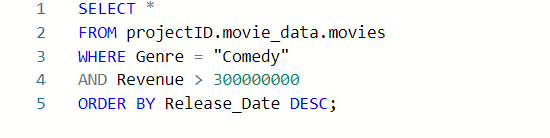
## **Example 4: Filter on two conditions, then sort data in descending order**

Use **WHERE**, **AND**, and **ORDER BY** to filter data on two conditions and then sort it.

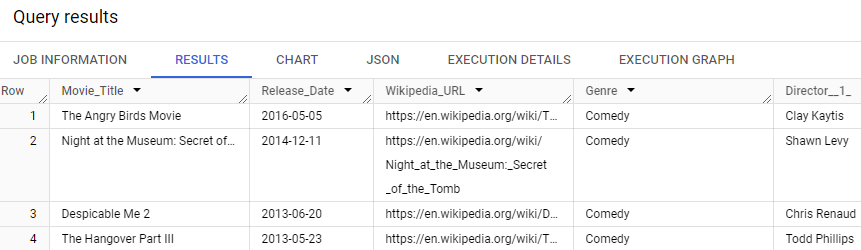
1. Select the beginning of the **ORDER BY** line and press Enter (Windows) or Return (Mac) to add a line between the **WHERE** and **ORDER BY** clauses.

2. Select the line you added. Enter **AND Revenue > 300000000** to add a condition to your query to return only rows where the Revenue column is greater than 300,000,000.

3. Your code should now match this code block:



4. Select **RUN**. The results are sorted from newest to oldest and include only comedy movies with revenues greater than $300,000,000.



### [SORT DATA WITH SQL](https://www.coursera.org/learn/analyze-data/lecture/P6Yu3/sort-data-with-sql)

Hello there! If you're hoping to learn about sorting—in SQL this time— you've definitely come to the right place. So far, we've sorted spreadsheets through the menu and with a written function. Which brings us to the next part of our learning: more sort functions, but this time in SQL. Data analysts love playing with the way data is presented. Sorting is a useful way to rearrange data because it can help you understand the data you have in a different light.

As you've probably already noticed, a lot of things you can do in spreadsheets can also be done in SQL. Sorting is one of those things. We've talked about using SQL with large datasets before. When a spreadsheet has too much data, you can get error messages, or it can cause your program to crash. That's definitely something we want to avoid.

SQL shortens processes that would otherwise take a very long time or be impossible to complete in a spreadsheet. Personally, I use SQL to pull and combine different data tables. It's much quicker than a spreadsheet, and that usually comes in handy. Here's something pretty helpful you can do with SQL.

You can use the ORDER BY clause to sort results returned in a query. Let's go back to our movie spreadsheet to get a better idea of how this works. Feel free to follow along in a SQL tool of your choice as we go. As a quick refresher, we have a database of movies listed with data like release date, director, and more. We can sort this table in lots of different ways using the ORDER BY function. For this example, let's sort by release date. First, we have the SELECT function and an asterisk.

Keep in mind that the asterisk means all columns are selected. Then we have FROM and the name of the database and table we're in right now. Now let's check out the next line. It's empty, but that's where we'll write our ORDER BY function. The ORDER BY command is usually the last clause in your query. Back to the actual sorting! We'll type ORDER BY with the space. With this clause, you can choose to order data by fields in a certain column. Because we want to sort by release date, we'll type Release\_Date. By default, the ORDER BY clause sorts data in ascending order. If you run the query as it is right now, the movies will be sorted from oldest to the most recent release dates. Let's run the query and see what we've got. You can also sort the release dates in the reverse order from the most recent dates to the oldest. To do this, just specify the descending order in the ORDER BY command written as DESC, D-E-S-C. Let's run this query.

As you'll notice, the most recently released films are now at the top of the database. In spreadsheets, you can combine sorts and filters to display information differently. You can do something similar in SQL too. You might remember that while sorting puts data in a specific order, filters narrow down data, so you only see data that fits the filter. For example, let's say we want to filter movies by genre so that we're only working with comedies. But we still want release dates to be sorted in descending order, from most recent to oldest films. We can do this with the WHERE clause. Let's try that now. First, we'll check that the ORDER BY clause is always the last line. That makes sure that all the results of the query you're running are sorted by that clause. Then, we'll add a new line for the WHERE clause after FROM and before ORDER BY.

Here's what we've got so far. From there, we want to type the column we're filtering for. In this case, we want to filter the database for comedies. After the WHERE clause, we'll type the column list's name as Genre. Now, we'll add an equal sign after Genre because we only want to include genres that match what we're filtering for. In this case, we're filtering for comedy, so we'll type Comedy between two apostrophes. Now, if you check out the entire query as a whole, you'll notice that we're selecting all columns, and we know it's all columns because that's what an asterisk means. The FROM clause specifies the name of the movie database we're using, and the WHERE clause filters the data to include entries whose genre is specified as comedy. Then in the last line, we have the ORDER BY clause, which will sort the data we've chosen to filter by release dates in descending order. This means when we run the query, we'll only have comedy movies listed from newest releases to oldest releases. Let's run it and figure out if that's the case.

Cool. Check out all those comedy movies and the way those dates are sorted.

Now, let's take this query a step further. We'll filter for two conditions at once using the AND filter. Working off the query we've been using, we'll add a second condition in the WHERE clause. We'll keep the sorting the same. Let's say you wanted to filter by comedy movies and movies that earned over 300 million in the box office. In this case, after the AND function, you'd add the revenue condition by typing Revenue. From there, you'll specify that you only want to return films with revenues over $300 million. To do that, type the greater than sign and then the complete number of 300 million without commas. Now let's run the query.

Here, the data only shows comedy movies with revenues of over $300 million, and it's sorted in descending order by release date. It looks really good. You just filtered and sorted a database like it's your job. And with practice, one day it can be. Just like that, you've finished another step in your data analyst journey. By now, you really dug and learned about the analysis process with a special emphasis on how organization can change how you go through your data. You also learned about both spreadsheets and SQL, and how to sort and filter data in both types of programs. To help you get more comfortable using spreadsheet and SQL features, you'll be getting some materials you can use as a resource. Coming up, we'll check out how an organizational mindset can take your analytical skills even further. We'll also cover converting, formatting, and adjusting data to combine information in a way that makes sense. Learning those skills early on can make your work as a data analyst much more efficient and effective in the long run.

[HANDS-ON ACTIVITY: SQL SORTING QUERIES](https://www.coursera.org/learn/analyze-data/quiz/iAg3O/hands-on-activity-sql-sorting-queries)



**Activity Overview**

****

So far, you’ve learned about SQL and how it’s used to retrieve data from databases. In this activity, you’ll practice sorting data with the ORDER BY clause in SQL. Sorting is a powerful tool for a data analyst. It enables you to:

* Organize and analyze your data in a meaningful way
* Find highest or lowest values in a dataset
* Compare data across different dimensions

By the time you complete this activity, you will be able to write SQL queries that sort data depending on your needs.

Scenario

Review the following scenario. Then complete the step-by-step instructions.

You’re a public health researcher with a state government agency. For your current project, you need to identify counties in the United States that have the most and least births in the 2016-2018 time frame. To do this, you’ll complete the following steps:

* Load the dataset.
* Query the data to explore its structure.
* Use ORDER BY to sort relevant data.
* Use sorted data to answer questions.

Step-By-Step Instructions

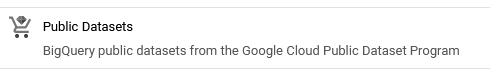
Follow the instructions to complete each step of the activity. Then answer the questions at the end of the activity before going to the next course item.

**Step 1: Load the CDC Births Data dataset**

1. Open the BigQuery console.

2. Select +ADD from the Explorer pane.

3. In the Add window, navigate to and then select Public Datasets.



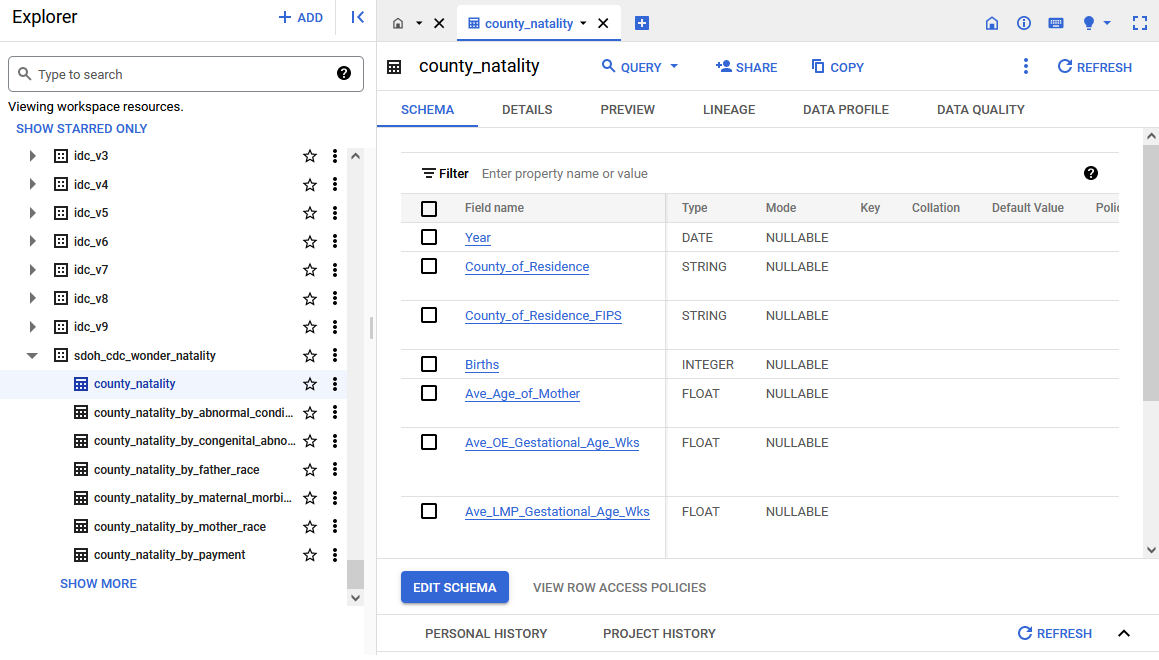
4. In the Marketplace search bar, enter  and press enter.

5. Select the result Births Data Summary from the CDC.

6. Select VIEW DATASET. This will bring you back to the console and open a Dataset info tab about the CDC dataset in the Details pane.

7. Select sdoh\_cdc\_wonder\_natality in the Explorer pane to examine the tables available within the dataset.

8. Select the table county\_natality and explore this table’s schema, details, and preview.



====================================================================================================

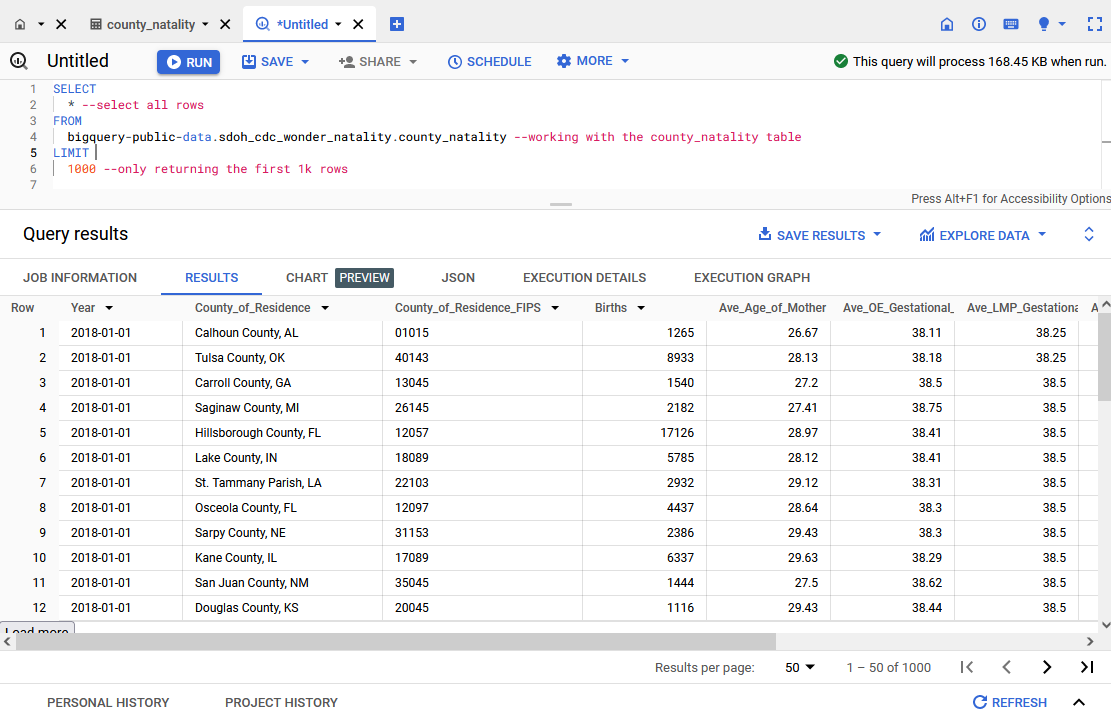
**Step 2: Query the data to explore its structure**

Now, it’s time to start working with the CDC births data. First, run a query to examine the dataset without sorting it.

1. Select Query, then in the New tab.

2. Enter the following query into the Query Editor to display the first 1,000 rows of the county\_natality table (see img)..

3. Select RUN.

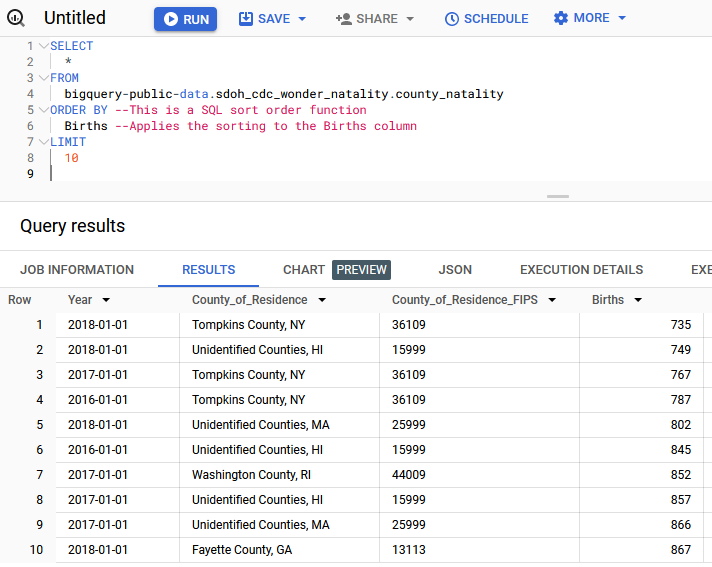


A screenshot of the Query results pane. The data returned include Year, County\_of\_Residence, County\_of\_Residence\_FIPS, Births, Avg\_Age\_of\_Mother, Ave\_OE\_Gestationa, and Ave\_LMP\_Gestational.

Examine the dataset you just loaded. Take a moment to familiarize yourself with the columns and fields available.

Note: Many of the public databases on BigQuery are living records and, as such, are periodically updated with new data. Throughout this course (and others in this certificate program), if your results differ from those you encounter in videos or screenshots, there's a good chance it is due to a data refresh.

**Step 3: Use ORDER BY to sort relevant data**

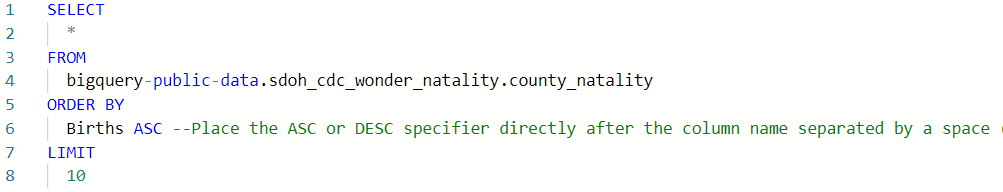
Now, sort the data with SQL’s ORDER BY function. Enter the following query into the Query Editor. The text preceded by two hyphens (--) are comments that explain the code. Run the Query.

Examine the Births column. Notice that it’s sorted from smallest to largest. When the ORDER BY function is applied to sort a given column, SQL will default to sorting in ascending order, which orders items from smallest to largest.

If you want the largest number to appear first, then you’d want to specify the sort order to be descending by adding a command to the ORDER BY clause. You can make your code easier to read by using a command to specify either sort order. Here are the corresponding commands:

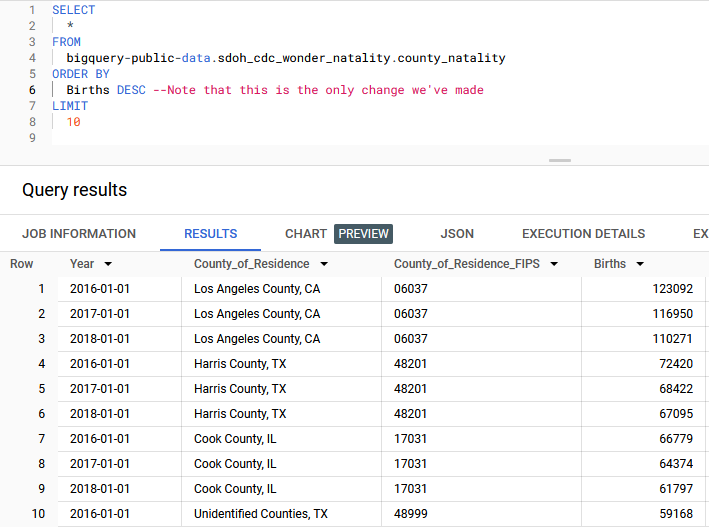
* ASC = Ascending
* DESC = Descending

Next, you’ll use the same query, but this time you’ll explicitly state the order of your ORDER BY function using ASC. Enter and run the following SQL query:



Notice that the results did not change. Tompkins County, NY, had just 735 births in 2018—the lowest birth count of any county in the US between 2016-2018.

Now, change the order from ascending (ASC) to descending (DESC) to find the most births. Enter and run this query:



The query returns the 10 rows with the largest values in the Births column. Los Angeles County takes up the top three spots.

**Step 4: Use sorted data to answer questions**

Now that you've become familiar with the basics of sorting functions, use them to answer questions about your data. This exercise will require you to apply both your previous learnings (especially filtering with the WHERE clause) and your new understanding of sorting.

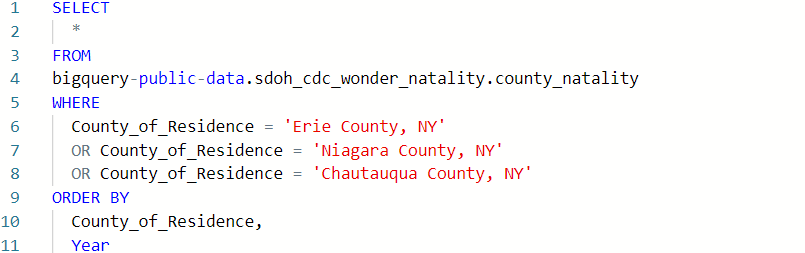
In your work as a public health researcher, you’re exploring whether the birth rate trends in several counties in upstate New York have been increasing or decreasing—and whether they follow the same pattern.

To answer this, you’ll need the following information:

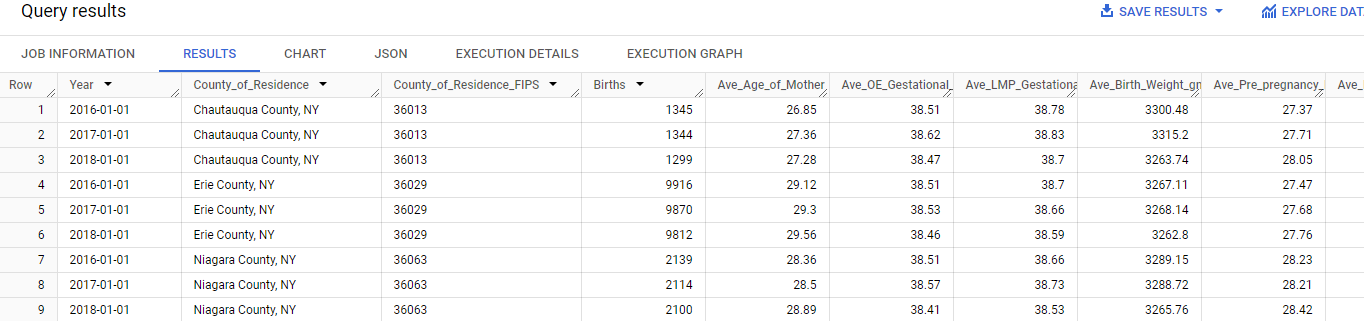
* Results from Erie, Niagara, and Chautauqua counties in New York state
* Results ordered by county of residence and year to find the trend

The following query will filter the results by county and sort the results byear and county. This will allow you to determine if the number of births is increasing or decreasing in each county.

Enter the following query into the Query Editor, then select RUN.



You’ve now successfully used both ORDER BY (sort) and WHERE (filter) clauses in the same query. Based on the results of this query, are births in these three counties following the same trend?



### [HANDS-ON ACTIVITY: ANALYZE WEATHER DATA IN BIGQUERY](https://www.coursera.org/learn/analyze-data/quiz/yRIIz/hands-on-activity-analyze-weather-data-in-bigquery)

### [TEST YOUR KNOWLEDGE ON SORTING DATA WITH SQL](https://www.coursera.org/learn/analyze-data/quiz/XAmut/test-your-knowledge-on-sorting-data-with-sql)

## MODULE 1 CHALLENGE

### [GLOSSARY TERMS FROM MODULE 1](https://www.coursera.org/learn/analyze-data/supplement/8eOci/glossary-terms-from-module-1)

### [MODULE 1 CHALLENGE](https://www.coursera.org/learn/analyze-data/exam/xkntO/module-1-challenge)