There are five different steps in the data analysis flow. You'll learn how to apply them using Excel.

Normally, it is good to start your analysis by doing a data check to make sure the data you received makes sense and is ready to work with. For example, you can check for duplicate values or missing values and do a sense check with other internal data sources.

Then, you'll move onto some exploratory data analysis. This is where you'll be doing some initial investigations of the data to summarize it's main characteristics. You may already have some questions in mind that you'd like to answer such as: Does an increase in revenue also lead to an increase in profit? You'll also build your first visualizations in this step.

The next step is to analyze & visualize your data. It's key to choose the right visualization to convey a message. This step also enables you to dig deeper into certain topics to make sure you don't miss any insights.

Now that you have built an analysis, the next step is to portray your analysis clearly in a dashboard format.

The final step is to communicate your insights with stakeholders.

The problem you will be working on in this course is customer churn. You'll be using a fictitious churn dataset from a Telecom provider called Databel. You are hired as a consultant, and your task is to analyze why customers are churning, or in other words, leaving Databel.

But what is churn exactly? A good definition is the one from Investopedia: "The churn rate, also known as the rate of attrition or customer churn, is the rate at which customers stop doing business with an entity." You can compare churn with the leaky bucket problem. You can fill the bucket with more water (or new customers in this case), but your overall revenue won't increase if existing customers are leaving your company. It's easier to retain customers than to attract new customers, so for many companies it's a priority to reduce churn.

There are multiple methods to calculate churn, and depending on the industry, it might make sense for a company to slightly alter the formula. A traditional e-commerce platform might consider a certain customer a churner if he or she hasn't made a purchase in the last 12 months. The simplified formula for churn is to divide customers' lost by the total number of customers. If we have a total of 100 customers in a certain period, and 10 end up leaving, we have a churn rate of 10%.

The Databel dataset consists of 29 different columns and has one row per customer. You'll be analyzing a snapshot of the database at a specific moment in time, meaning there is no time dimension.

The dataset contains numerous dimensions, the first one being Customer\_id. The Customer\_id is a unique ID that identifies an individual customer. The second column is called Churn Label, and it indicates if a customer churned with "Yes" or "No". The dataset contains various other dimensions, such as demographic fields and information about premium plans.

The dataset contains more than just dimensions, so let's look at some measures. The Total Charges column, for example, takes the sum of all monthly charges billed to a customer.