the first step is to **use** **this** **data model to** **identify which datasets will be required to answer your business question -**which is to to figure out the**top 5 categories with the largest popularity.**

So, let’s have a look at what data you have to work with. The client has sent through:

* **7 data sets** - each data set contains different columns and values
* **A data model** - this shows the relationships between all of the data sets, as well as any links that you can use to merge tables.

There is a lot of information here and it’s easy to get lost in the data. So, to make sure you are using the right data to answer the business questions you’ll follow these steps:

1. Requirements gathering
2. Data cleaning
3. Data modelling

This is exactly the task a data analyst would be focusing on and this is what you will be working on!

Analysis of sample data sets with visualizations to understand the popularity of different content categories.

Data sets - Quick Explanation

Great work! You’ve identified Reaction, Content, and Reaction Types as our relevant data sets.

Data sets - Quick Explanation

Great work! You’ve identified Reaction, Content, and Reaction Types as our relevant data sets.

To clarify why you made this selection:

The brief carefully it states that the client wanted to see “An analysis of their content categories showing the top 5 categories with the largest popularity”.

As explained in the data model, popularity is quantified by the “Score” given to each reaction type.

We therefore need data showing the content ID, category, content type, reaction type, and reaction score.

So, to figure out popularity, we’ll have to add up which content categories have the largest score.

But! Before we begin to work with the data sets, we’ll need to ensure that the data is clean and ready for analysis…

To clarify why you made this selection:

The brief carefully it states that the client wanted to see “An analysis of their content categories showing the top 5 categories with the largest popularity”.

As explained in the data model, popularity is quantified by the “Score” given to each reaction type.

We therefore need data showing the content ID, category, content type, reaction type, and reaction score.

So, to figure out popularity, we’ll have to add up which content categories have the largest score.

But! Before we begin to work with the data sets, we’ll need to ensure that the data is clean and ready for analysis…

You’ll learn to choose the right data visualizations that best support the story you want to tell.

Let’s bring the data to life using visualizations: