1. Data engineers vs. data scientists

Great job on these exercises!

## 2. Data workflow

In the previous lesson we got acquainted with how the data flows through an organization, focused on the data engineer's responsibilities, and quickly mentioned data scientists. To prevent the confusion and assumptions that come with buzzwords, let's clarify how data engineers and data scientists contrast and compare.

## 3. Data engineers

You already know that data engineers focus on the first part of the workflow. Their role is to ingest and store the data so it's easily accessible and ready to be analyzed.

## 4. Data scientists

Data scientist intervene on the rest of the workflow: they prepare the data according to their analysis needs, explore it, build insightful visualizations, and then run experiments or build predictive models. Data engineers lay the groundwork that makes data science activity possible. Let's see how data engineers enable data scientists.

## 5. Data engineers enable data scientists

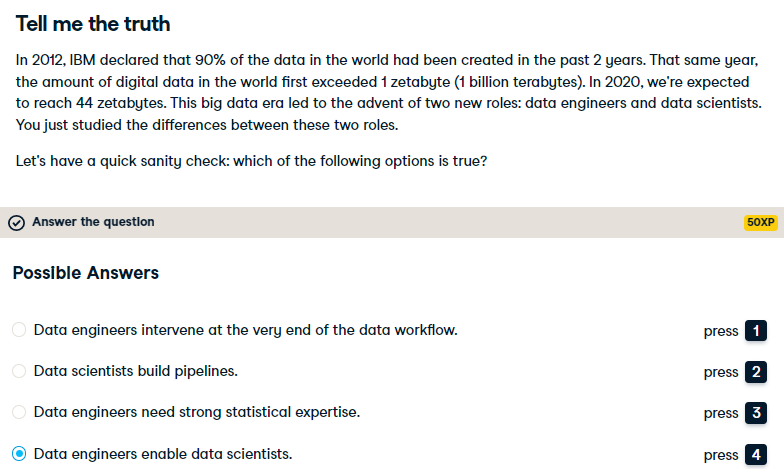
Vivian is a data engineer at Spotflix, our music streaming company, and Julian is a data scientist. Data engineers ingest and store collected data, so that data scientists can exploit it. At Spotflix, Vivian collects and store customer, artist, song data in their respective databases. Julian then uses these tables to understand listening patterns or build recommendation engines. Data engineers ensure that databases are optimized for analysis (correct table structure, information easy to retrieve) while data scientists access the databases to exploit the data it contains. At Spotflix, Vivian makes sure that Julian can easily access tracks, artist, listening sessions data, and can analyze it without too much preparation work. Data engineers build data pipelines. The next lesson is focused on this topic. Data scientists use the pipelines' outputs. At Spotflix, Vivian builds the pipeline that pulls listening sessions data, so that Julian's analyses remain up to date. Based on the above, it's no surprise that data engineers are software experts, while data scientists are analytics experts. In general, Vivian uses languages like software-oriented Python or Java, and SQL to create, update and transform databases, while Julian uses analytics-oriented Python or R, and SQL to query - or, in other words, request information from - databases.

## 6. Summary

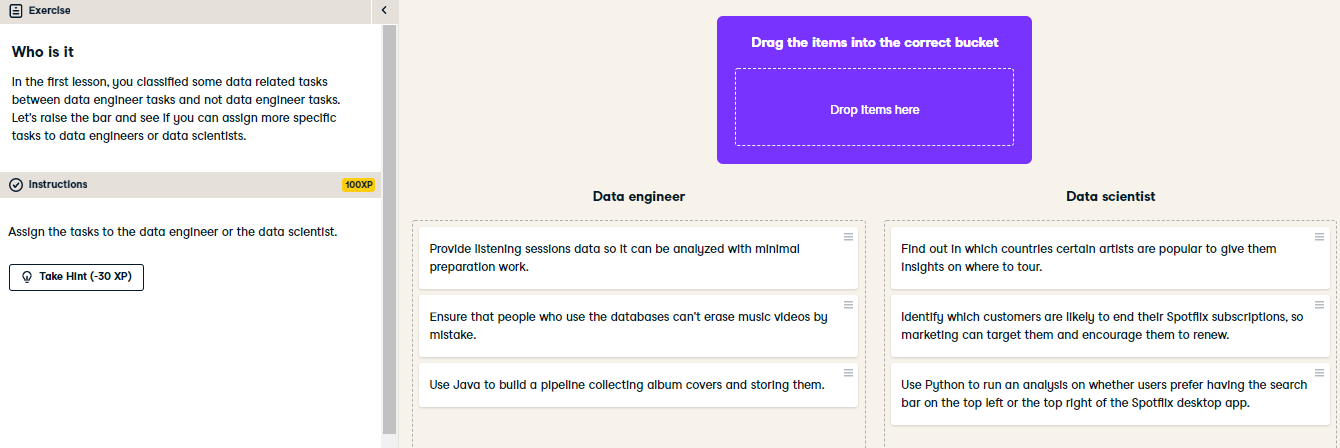
OK! Now you understand at which stages data engineers and data scientists intervene, and how data engineers enable data scientists.

## 7. Let's practice!

Time for a sanity check!



Yes! Data engineers lay the groundwork that makes data science possible.



All right! Your understanding of the world of data engineering is increasing by the minute.