Step 1: Anticipate Questions and Identify Data

PDF (amazon-quicksight-user.pdf#anticipate-questions-quicksight-q-walkthrough)
Kindle (https://www.amazon.com/dp/B076453HGY)
RSS (amazon-quicksight-doc-release-notes.rss)

We'll use questions to identify necessary data fields and define key terms. We'll also walk through leveraging your data model for answering questions.

1. Anticipate questions.

   Before you start to prepare your datasets or create a topic, you and your organization need to determine the questions you want answered by the data. With detailed questions, you will be able to determine what data will need to be acquired and potential blockers.

   We'll go more into detail on using your questions to create topics later in step 2 [link to Step 2].

   Potential questions include (but are not limited to):

   1. Information that needs to be communicated across multiple areas of the organization.

      In this case, we will be communicating platform traffic (mobile, desktop, and app) for the purpose of optimizing the platform's performance.

   2. Information that specific roles need to complete their work.

      For our example, upper management may need access to profit calculations and visualize trends over time for the streaming service.

   3. Common visualizations or calculations that are used in the organization.

      For the music streaming service, profit, revenue, and cost calculations are needed every quarter to determine the company's performance.

   Consider reaching out to individuals involved in your organization to create a comprehensive list of questions. With detailed questions, you will be able to determine what data will need to be acquired and potential blockers.

   For this walkthrough, we'll be using the following questions to prepare our datasets for QuickSight Q and create a topic. The questions are based on the data model [link to data model].

   - Which songs were the most popular in 2019?
   - Which platform (mobile, desktop, app) was the most popular in 2019?
   - What was the subscription revenue in 2019?
   - What was the subscription profit in 2019?
   - Show me the monthly traffic on each platform in 2019.

- What is the traffic on mobile by country?
- How many subscribers were recruited in 2019 by month?

Note

The wording in these questions deliberately follows the format of QuickSight Q's usable question types. You'll want to reword your list of questions to match QuickSight Q's supported question types. For more information about the master list of question types QuickSight Q can answer, see [link to main docs question types].

2. Identify key terms and calculations.

   A term is a word or phrase that can be linked to a field in your dataset such as *User ID*, which represents unique users, or a calculated field such as *moving subscription revenue by month*, which calculates the total amount made from subscription costs per month.

   With Amazon QuickSight Q, terms are significant because they represent how data will be linked within a topic. In other words, if the term "artist" is used to represent "songs", then QuickSight Q would answer the question "Which songs were the most popular in 2019" using the artist information. While QuickSight Q would still provide an answer, it would be incorrect.

   Note

   We recommend that you define these terms ahead of time and communicate their use with your organization. By using agreed upon terms, Amazon QuickSight Q will be able to more efficiently and accurately produce answers. For an example data dictionary, we suggest referring to the data dictionary used in this walkthrough [link to data dictionary].

   Aside from the fields in the tables [link to data dictionary with only the tables], the walkthrough uses the following terms:

   - **Subscription Revenue:** the amount made from new subscriptions in a given month.
   - **Moving Subscription Revenue (monthly):** shows the total amount made from new subscriptions in a given year.
   - **Song Popularity or Top Song:** the highest number of *streams* for that *song*.
   - **Artist Popularity or Top Artist:** the highest number of *monthly song listeners* for that *artist.
   - **Platform Popularity or Top Platform:** the highest number of *streams* for that *platform value (mobile, desktop, app)*.

   Remember to use unique terms to improve QuickSight Q's accuracy. In this walkthrough, we specified a term by referring to the table name (monthly song listeners, user start data, etc.).

   Generic terms such as "revenue" or "platform" need additional information for QuickSight QuickSight Q's algorithms to work effectively

   For example, in the list of questions above, "popularity" was used in two separate questions. To clarify for QuickSight Q, "popularity" in the question "Which songs were the most popular in 2019" can be specified as "song popularity", "most listened song", "song with the most likes". On the other hand, "popularity" in the question "Which platform was the most popular in 2019" can be specified as "platform popularity", "most visited platform".

   For this walkthrough, we have defined fields, key terms, and calculations below for your reference:

   Note

Due to the use of common terms such as “popularity” and “revenue”, we suggest creating a data dictionary, a spreadsheet or other record of how terms will be used for consistency in your organization.

3. Identify data that will be needed to answer questions.

   Amazon QuickSight Q uses your data to create visualizations that answer someone's question there is not enough data, QuickSight Q will not be able to correctly answer your questions, if at all.

   For instance, for the question “Show me the monthly traffic on each platform in 2019?”, QuickSight Q needs the following information:

   - Traffic: the number of streams in a given period of time.
   - Platform: how Users stream music.
   - Platform Type: the field values for platform, of which there are only three (app, mobile, desktop).
   - Q’s algorithm will use the time periods “monthly” and “2019” to generate the visualization.

4. Understand your data model.

   A data model organizes your data and standardizes how it will be used. By using the relationships defined in the data model, Amazon QuickSight Q will be able to more accurately create visualizations.

   For instance, in the question “Which songs were the most popular in 2019?” the information provided in the Songs table identifies each unique song, while the Streams table determines how many times each song was streamed.

   Your main takeaways from your data model may include:

   1. How the field will be aggregated, if at all.

      The question *Which songs were the most popular in 2019?* is calculated with the Subscribers table. The subscription cost aggregated as a sum of the subscription costs per subscriber (for a subscription cost of $5 or $10, the total will be $5 + $10 + etc.) and the number of subscribers aggregated as distinct data points per month (for the subscriber_ids 1-5, QuickSight will count it as 5 users).

   2. How several fields may work together to answer a question.

      The question *Which songs were the most popular in 2019?* is calculated with cost_by_stream and licensing_cost from the Songs table and rate_royalties from the Artists table.

   3. Whether the datasets provided will be sufficient to answer potential questions.

      The question *Which songs were the most popular in 2019?* is calculated with the number of visits to the mobile platform, the location the mobile platform was accessed on, and the time it was accessed.