# Draft - QuickSight Q Walkthrough (unedited)

# Lisa | Updated Draft for QuickSight Q Walkthrough

## Updated Draft 8/19/21 - MIGRATED & UPDATED IN XML

### <http://dev-dsk-lisaleun-2c-0f7808b8.us-west-2.amazon.com/QS/src/AWSSpaceNeedleDocs/build/server-root/quicksight/latest/user/quicksight-q-walkthrough.html>

Amazon QuickSight Q is a natural language query (NLQ) tool that uses metadata and machine learning algorithms to answer business questions instantly. QuickSight Q uses Topics to link datasets and metadata so you can ask questions naturally.

This walkthrough will cover how to conceptualize, develop, and test a QuickSight Q topic for a music streaming service. We’ll use a provided data model, dataset, and data dictionary to begin using QuickSight Q topics as quickly as possible.

To demonstrate how to use Amazon QuickSight Q and Topics, we’ll walk through how to develop Topics and make datasets natural language friendly. Then, we’ll analyze and test Topics to produce visualizations instantly.

By the end of the walkthrough, you’ll have an Amazon QuickSight Q Topic similar to {example}, which visualizes subscription data.

This walkthrough should take approximately 30 minutes and will cover the following steps:

* Step 1: Anticipate Questions and Identify Data
  + 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.
* Step 2: Create a Topic
  + We’ll determine potential topics from our list of questions and define the scope of our topic.
* Step 3: Update Your Topic to Be Natural Language Friendly
  + - We’ll add friendly names, synonyms, and metadata to ensure QuickSight Q can make connections between your fields.
* Step 4: Test Your Topic
  + We’ll go over potential questions to test with and improve QuickSight Q’s answers.

**Pre-reqs**

Before you begin this tutorial, you should have:

* completed Getting Started [link to main docs - getting started/prereqs for accessing Q]
* Downloaded the assets we will be using in this walkthrough [link to download]
  + Users 2018-2019.csv
  + Subscribers 2018-2019.csv
  + Artists 2018-2019.csv
  + Songs 2018-2019.csv
  + Streams 2018-2019.csv
  + Stakeholder Activity Topic - Data Dictionary 2018-2019.csv
  + Stakeholder Activity Topic - List of Questions

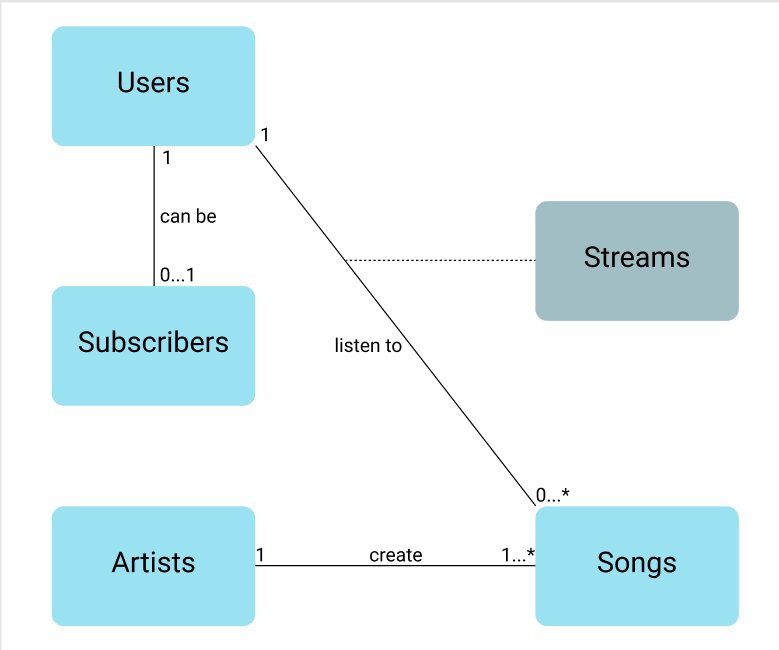
**Scenario**

Imagine you are an analyst at a music streaming company hoping to improve subscriptions and analyze traffic across platforms (mobile, desktop, app). With data on user traffic and campaigns, you want to allow your coworkers to efficiently get answers to their questions about the data.

**Description of sample data and model**

The data model is simplified for the purpose of this walkthrough [and is based off of a relational database]. The data model has 5 tables that provide crucial information to QuickSight Q:

* Users Table: records the information about stakeholders with accounts for the streaming service and provides the following information: user\_id, country, join\_date, age.
* Subscribers Table: records the information about stakeholder with paid membership for the streaming service and provides the following information: subscriber\_id, cost, start\_date, user\_id.
  + The user\_id in the Subscribers table is part of how the data model functions; it links the subscriber account to the user account.
  + For the purpose of this walkthrough, Subscribers are Users that pay a monthly fee. (The benefits of the membership can be left to the imagination.)
* Artists Table: lists all of the artists that create Songs and provides the following information: artist\_id, name, monthly\_listeners, monthly\_payment.
* Songs Table: lists all of the songs Users can listen to and provides the following information: song\_id, release\_date, genre, total\_streams, royalties, artist\_id.
* Streams Table: lists all of the streams when Users listen to Songs and provides the following information: date\_time, platform, song\_id, user\_id.



|  |  |  |
| --- | --- | --- |
| Column 1 | Column 2 | Column 3 |
|  |  |  |
| **Users Table** | **Artists Table** | **Streams Table** |
| user\_id | artist\_id | date\_time |
| country | name | platform |
| join\_date | monthly\_listeners | song\_id |
| age | monthly\_payment | user\_id |
|  |  |  |
| **Subscribers Table** | **Songs Table** |  |
| subscriber\_id | song\_id |  |
| cost | release\_date |  |
| start\_date | genre |  |
| user\_id | total\_streams |  |
|  | royalties |  |
|  | artist\_id |  |

**Description of the Topic that will be created**

Using the data model above, we will create a topic about Stakeholder Activity from 2018-2019. For the purpose of this walkthrough, we will define stakeholder activity as an interaction with the content being streamed or the stakeholder’s account.

**Glossary/Terminology**

* Data Model
* Dataset
* Field
* Value
* Topic
* Term/Key Term
* Friendly Name
* Synonym
* Metadata

### Outline

**Step 1: Anticipate Questions and Identify Data**

1. Anticipate questions
2. Identify terms
3. Identify data that will be needed to answer questions
4. Understand your data model

**Step 2: Create a Topic**

1. Connect to data
2. Create a topic

**Step 3: Update Your Topic to Be Natural Language Friendly**

1. Exclude fields you don’t want to use in the topic
2. Rename fields
3. Add synonyms
4. Configure metadata

**Step 4: Test Your Topic**

1. Using the Q bar
2. Troubleshooting wrong answers

## Steps

**Step 1: Anticipate Questions and Identify Data [currently being edited]**

1. Anticipate questions.
   1. 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.
   2. We'll go more into detail on using your questions to create topics later in step 2 [link to Step 2].
   3. Potential questions include (but are not limited to):
      1. information that needs to be communicated across multiple areas of the organization
         1. 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
         1. 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
         1. For the music streaming service, profit, revenue, and cost calculations are needed every quarter to determine the company’s performance.
   4. Consider reaching out to individuals involved in your organization to create a comprehensive list of questions.
   5. 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].
      1. Which songs were the most popular in 2019?
      2. Which platform (mobile, desktop, app) was the most popular in 2019?
      3. What was the subscription revenue in 2019?
      4. Show me the monthly traffic on each platform in 2019.
      5. What is the traffic on mobile by country?
      6. How many subscribers were recruited in 2019 by month?
   6. **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 Q can answer, see [link to main docs - questions Q can answer].
2. Identify key terms and calculations.
   1. 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.
   2. With 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 Q would still provide an answer, it would be incorrect.
   3. We recommend that you define these terms ahead of time and communicate their use with your organization. By using agreed upon terms, QuickSight Q will be able to more efficiently and accurately produce answers.
      1. Aside from the fields in the tables [link to data dictionary with only the tables], the walkthrough uses the following terms:
         1. subscription revenue (monthly): the amount made from subscribers for a given month (Jan - $10, Feb - $5, March - $20)
         2. moving subscription revenue (monthly): the amount made from all new subscribers each month (Jan $10, Feb - $15, March - $35)
         3. song popularity or top song: the highest number of **streams** for that **song**
         4. artist popularity or top artist: the highest number of **monthly song listeners** for that **artis**t
         5. platform popularity or top platform: the highest number of **streams** for that **platform value (mobile, desktop, app)**.
   4. Remember to use unique terms to improve Q’s accuracy.
   5. Generic terms such as “revenue” or “platform” will need additional information for 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".
   6. [INSERT GRAPHIC]
   7. For this walkthrough, we have defined fields, key terms, and calculations below for your reference.:
      1. [insert data dictionary snippet]
   8. **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.
   1. 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.
   2. For instance, for the question “Show me the monthly traffic on each platform in 2019?”, Q needs the following information:
      1. traffic - number of streams
      2. platform - how users stream music
      3. platform type - the field value for platform - mobile, app, desktop.
      4. Q’s algorithm will use the time periods “monthly” and “2019” to generate the visualization.
4. Understand your data model.
   1. A data model organizes your data and standardizes how it will be used.  By using the relationships defined in the data model, Q will be able to more accurately create visualizations.
      1. 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.
      2. Your main takeaways from your data model may include:
         1. How the field will be aggregated, if at all.
            1. Exp: **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.
            1. Exp: **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.
            1. Exp: **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.

**Step 2: Create a Topic [currently being edited]**

1. Determine potential topics.
   1. For the topic, the scope has been limited to one topic: Stakeholder Activity.
   2. Note: in practice, topics may need to be more narrow. However, due to the minimal calculations and data in the topic, we have included all of the data sets in the data model.
   3. In order to create a robust topic, we recommend gathering questions you and your organization have about your data. . Refer to step 1 of this walkthrough [link] to learn more.
   4. For this walkthrough example on music streaming service data, we have created the following list of questions:
      1. Which songs were the most popular in 2019?
      2. Which platform (mobile, desktop, app) was the most popular in 2019?
      3. What was the subscription revenue in 2019?
      4. Show me the monthly traffic on each platform.
      5. What is the traffic on mobile by country?
      6. How many subscribers were recruited in 2019 by month?
   5. From this list of questions, consider:
   6. Will the data be analyzed monthly? Quarterly? Yearly?
      1. For the questions “Which songs were the most popular in 2019?”, “Which platform (mobile, desktop, app) was the most popular in 2019?”, “What is the traffic on mobile by country?”, and “What was the subscription profit in 2019?” the data is analyzed within a single year.
      2. **Note:**Although data can be analyzed between multiple years, we’ll analyze the 2019 data only for the questions listed above.
      3. Is qualitative (categorical) or quantitative (numerical) data being analyzed?
         1. Which platform (mobile, desktop, app) was the most popular in 2019?
         2. What was the subscription revenue in 2019?
      4. Are some questions analyzing the same/similar data?
      5. The questions “Which platform was the most popular in 2019?” and “Show me the monthly traffic on each platform” both analyze traffic data for the platform. “Which platform was the most popular in 2019” determines which platform has the most traffic in a given year (by our definition of platform popularity), while “Show me the monthly traffic on each platform” shows trends in traffic data every month.
      6. Outside of this example, we suggest creating a topic with a specific period of time in mind (2018-2019, Q1 2019, etc.). Over time, the data might change to improve the organization or make the topic inconvenient for users. For platform traffic, it might be inconvenient to consistently list monthly platform traffic data over several years instead of the last 6 months.
   7. Is there enough data to answer the question? Is there anything missing?
      1. Make sure that you have all the data needed to answer your question. You might have to create calculations, or create new fields, or even entire tables. Without sufficient data to answer your question, Q may not be able to answer or give a correct answer. For this example, determining the organization’s profit would not be possible, as there is no data on employees or other sources of revenue and costs. However, determining subscription profit is possible, as we have defined it as the subscription revenue subtracted by the streaming costs and campaign costs.
      2. Topic organization depends on how often you and your organization want to view the data and how specific your list of questions is. We suggest using the considerations above to create topics that cover questions on a single subject and period of time.
      3. For this example, we could make topics on:
      4. Which songs were the most popular in 2019?
      5. Which platform (mobile, desktop, app) was the most popular in 2019?
      6. What was the subscription revenue in 2019?
      7. Show me the monthly traffic on each platform.
      8. What is the traffic on mobile by country?
      9. How many subscribers were recruited in 2019 by month?
         1. Song streams in 2019 to cover the question:
            1. Which songs were the most popular in 2019?
         2. Platform traffic in 2019 to cover the questions:
            1. Which platform (mobile, desktop, app) was the most popular in 2019?
            2. Show me the monthly traffic on each platform. (assuming this is for 2019 data, as the data provided for this example is limited to that year)
            3. What is the net traffic on mobile by country?
         3. Profit in 2019 to cover the questions:
            1. How much did it cost to stream songs in 2019?
            2. What was the subscription profit in 2019?
2. Create a topic.
   1. Upload the data into QuickSight. For this walkthrough, we’ll be using the datasets (CSV) that you can download here [link to download]. For more information about creating datasets in QuickSight, see [Creating Datasets](https://docs.aws.amazon.com/quicksight/latest/user/creating-data-sets.html).
      1. Open Amazon QuickSight.
      2. In the section on the left-hand side, choose Datasets [icon].
      3. Select Upload CSV.
      4. Select Streams.csv, all tables.
      5. Note: If adding from a static data source (CSV, JSON, etc.), add a date to the name of the file (Exp: sales.csv → sales\_07\_12\_21.csv). SPICE/QuickSight Q will not update a pre-existing static data source and will allow files of the same name to co-exist.
   2. Open Amazon QuickSight.
   3. In the navigation bar on the left-hand side, choose Topics ([icon]).
      1. Note: if you do not see the Topics icon, be sure to check that you have Q enabled here [link].
   4. On the **Q Topics** page that opens, choose **Create topic** at top right.
   5. In the **Create topic** page that opens, do the following:
      1. For **Topic name**, enter a descriptive name for the topic. For this example, let’s call this topic **2018-2019 Platform Traffic**.
      2. For **Description**, enter a specific description. For this example, let’s use the following description:   
           
         This topic includes data on users accessing the application with mobile, desktop, or the app. You will be able to determine platform popularity and look at trends from 2018-2019 for platform traffic.
      3. Choose **Continue**.
3. Connect data to a topic.
   1. In the **Add Data to Topic** page, do the following to add the CSVs for the walkthrough:
      1. Select **Datasets.**
      2. Choose [list of CSVs].
      3. Scroll to the bottom of the page. [Comment: may be unnecessary]
      4. Choose **Add data**. [This may take several minutes depending on the the data source. Reloading the page may help.].
      5. Note: datasets that have been configured in other topics will need to be configured for each new topic that they’re used in. To minimize inconsistency, keep track of the metadata used for each dataset and topic with a data dictionary or spreadsheet.
   2. [Future Work: consider adding a step for setting the default date and time basis as a baseline for how dates should be evaluated—potentially add a table with two date fields to show how the default date works]

**Step 3: Update Your Topic to Be Natural Language Friendly**

1. Exclude fields you don’t want to use in the topic
   1. Under the **Data** tab at the top of the page, choose the switch under **Include** on the right-hand side of the screen. Select the switch to include and exclude the field.
   2. As a reminder, we are tailoring the topic to answer the following questions from 2018-2019:
      1. Which songs were the most popular in 2019?
      2. Which platform (mobile, desktop, app) was the most popular in 2019?
      3. What was the subscription revenue in 2019?
      4. Show me the monthly traffic on each platform.
      5. What is the traffic on mobile by country?
      6. How many subscribers were recruited in 2019 by month?
      7. When Users stream songs
      8. Which platform Users prefer to access their music from
      9. Which Songs are the most popular
      10. Which Artists are the most popular
      11. How many Users created accounts
      12. How many Users became Subscribers
      13. How much revenue is generated from subscriptions
      14. How old Users are
      15. Where Users are located
   3. For this topic, select the switch [icon] on the right-hand side to exclude the following fields:
      1. Monthly Payment (Artists Table)
      2. Royalties (Songs Table)
      3. Total Streams (Songs Table) [may delete this]
      4. [Explanation] After you have excluded the monthly payment and royalties fields, the fields will be ignored by QuickSight Q. If you would like to include the fields for future use, select the switch on the right hand side.
      5. We are excluding the monthly payment and royalties fields because we want to limit our scope to answer the questions we selected for the topic in Step 2 [link to step 2].
         1. Which songs were the most popular in 2019?
         2. Which platform (mobile, desktop, app) was the most popular in 2019?
         3. What was the subscription profit in 2019?
         4. Show me the monthly traffic on each platform.
         5. What is the net traffic on mobile by country?
2. Rename fields to be reader-friendly.
   1. Readers are individuals that will have access to your topic but not the data itself.
   2. Under the **Data** tab at the top of the page, select the pencil icon [icon] next to the name of the field on the left hand side of the page.
   3. Highlight the box and change the field name to a friendly name.
   4. For this topic, select the pencil icon next to **user\_id** and change the name to **User ID**. Use the table below to change the field names to reader-friendly names. The reader-friendly names in this walkthrough will indicate which table the field is from (Subscriber Join Date, Subscriber Cost, Stream Platform, etc.) for clarity.
   5. [table at the bottom of the step so the formatting here isn’t messed up]
   6. Friendly names will be the only method of learning about the data for readers and provides information for QuickSight Q to make links. We recommend adding additional information about the field if needed.
   7. If QuickSight Q is unable to answer a question or needs clarification, the friendly names will help readers clear up any misunderstanding.  [SCREENSHOT: question - nothing; vis  - friendly names of fields]
   8. Note: For the purpose of this walkthrough, the original field names are similar to the reader-friendly field names. In practice, original field names may be far less descriptive. We recommend using a data dictionary to record both the original names and new names for reference.
3. Add synonyms to fields to be natural-language friendly.
   1. Under the **Data** tab at the top of the page, select the pencil icon [icon] beneath the **Synonyms** column. If you have not added any synonyms to the field yet, there will be the message “Add alternative names for field” next to the pencil icon [icon].
   2. Add a synonym. A synonym may be the friendly name reordered or how other individuals may refer to the field. If you have multiple synonyms, using the plus icon [icon] to add more. With more synonyms, QuickSight Q will be able to answer more questions. However, make sure that none of the synonyms are the same, or QuickSight Q will need additional clarification when answering your question.
   3. For this topic, select the pencil icon [icon] in the row named **User ID** and add the synonyms **unique user**, **user**, **user identification number**, **listeners**, and **customers**. Use the table below to add the remaining synonyms names to reader-friendly names. The natural-language friendly synonyms in this walkthrough will allow QuickSight Q to make connections between your data.
   4. [table at the bottom of the step so the formatting here isn’t messed up]
   5. Note: If the synonyms used for different fields are the same, QuickSight Q may need additional clarification when answering your questions. We recommend using a data dictionary to record both the original names and new names for reference.
   6. QuickSight Q uses field synonyms to identify key terms within questions.
      1. [Example] The question **What was the subscription revenue 2019?** could be asked in several ways:
         1. Show me the revenue from subscriptions by month in 2019.
         2. What is the annual subscription revenue?
         3. Membership revenue by quarter in 2019?
4. Add synonyms to values to be natural-language friendly
   1. Under the **Data** tab at the top of the page, select the [meatball??? icon] next to the include switch.
   2. Select **Field Value Synonyms.** Field value synonyms will link synonyms to specific values in your data. Unlike adding a synonym to the field, after you add field value synonyms, you will not be able to view the synonyms until you select the [meatball icon].
   3. Select **Add** to link values to synonyms.
   4. For this topic, select the [meatball icon] in the **Stream Platform** field.
   5. Select **Add** and enter the value **mobile**.
   6. For the value **mobile**, add the synonym **phone**.
   7. Create the value **app**, and add the synonyms **used the app** and **application**.
   8. Create the value **desktop**, and add the synonym **pc** and **computer.**
   9. This topic evaluates streaming traffic and uses the values **mobile, app, and desktop** to categorize how Users will access music. To ensure that QuickSight Q will be able to determine when the **Platform** field is being used, we added synonyms to values.
   10. QuickSight Q uses value synonyms to identify terms within questions.
       1. [Example] The question **Which streaming platform was the most popular in 2019?** could be asked in several ways:
          1. How many people used their app/mobile/desktop/phone in 2019?
          2. Show me platform access by month in 2019.
5. Configure metadata to be natural-language friendly.
   1. Under the **Data** tab, select the arrow icon [icon] under the **Include** column.
   2. Add the following:
      1. description of the field: what the field represents in the data, if there are limitations or exceptions.
      2. the field’s role: whether the field is a categorical (dimension) or quantitative (measure) value.
      3. the default aggregation of the field: how the field will be counted or operated upon with other fields.
      4. semantic type: what the field acts most similarly to
      5. semantic sub-type (if you are able to select anything):
      6. value format: how the field should be visualized
   3. For the topic, select the arrow icon [icon] under the **Include** column for the field **User ID.**
      1. In the **Description**, enter the following information:
         1. unique number assigned to each user
         2. A **description**should cover what the field represents in the data and any limitations or exceptions.
      2. Under **Role**, select the dropdown menu.
         1. choose **Dimension**.
         2. **Roles** determine if the field is something that can be analyzed or calculated**,** or a **Measure**, or if the field is something that will categorize the data or cannot be used for calculations, or a **Dimension.**
      3. Under **Default Aggregation**, select the dropdown menu and choose **Count Distinct.**
         1. A **default aggregation** will determine how the field will be counted for a **dimension.**
            1. **Count Distinct** will count each unique occurrence of **User ID** that apply.

Exp: from a list of the following **User IDs**, the aggregation would be 4, since there are 4 unique users: 1, 1,1,1,1,2,2,3,4,4,4,4,4,4,.

* + - * 1. **[NOTE] Count** will count each occurrence of any **User ID** that apply.

Exp: from a list of the following **User IDs,** the aggregation would be 12, since there are 12 occurrences of any **User ID:** 1,1,1,1,1,1,2,2,3,3,4,4.

* + - 1. A **default aggregation** will determine how the field will be used for calculations for a **measure.**
         1. (Subscription month membership will be the example)
         2. **Sum** will add all of the values that apply.

Exp: from a list of the following **Subscription Costs**, the aggregation would be 80, since it is taking the sum of the **Subscription Costs: 5, 5, 5, 5, 10, 10, 10, 15, 15.**

* + - * 1. **Average** will take the mean of all values

Exp: from a list of the following **Subscription Costs**, the aggregation would be 8.89, since it is taking the sum of the **Subscription Costs: 5, 5, 5, 5, 10, 10, 10, 15, 15.**

* + - * 1. **Max** will return the highest value of those that apply.

Exp: from a list of the following **Subscription Costs**, the aggregation would be 15, since it is taking the sum of the **Subscription Costs: 5, 5, 5, 5, 10, 10, 10, 15, 15.**

* + - * 1. **Min** will return the lowest value of those that apply.

Exp: from a list of the following **Subscription Costs**, the aggregation would be 5, since it is taking the sum of the **Subscription Costs: 5, 5, 5, 5, 10, 10, 10, 15, 15.**

* + 1. Under the **Not allowed aggregation,** select the dropdown menu and choose **Count** for **User ID.**
       1. The **Not allowed aggregation** is optional, and does not apply to all cases.
       2. A **not allowed aggregation** will determine how the field cannot  be counted for a **dimension.**
       3. A **not allowed aggregation** will determine how the field cannot be used for calculations for a **measure.**
    2. Under the **Semantic Type,** select the dropdown menu and choose **Identifier.**
       1. A **Semantic Type** identifies a general category of values that the field behaves most similarly to.
       2. For fields that have the role **Dimension**, they can have the following **Semantic Types:**
          1. **Duration:** fields that provide a length of time
          2. **Date/Date Part:** fields that specify a specific point in time (exp: 01/31/2019)

Be sure that the format of the date/date part is compatible with QuickSight Q [link to acceptable date/time formats]

* + - * 1. **Location:** fields that signify a geological region
        2. **Boolean:** fields that represent true, false, or null
        3. **Currency:** fields that represent money or used to signify quantities

exp: the **Currency** semantic type can be used for fields that represent revenue or cost, as well as **monthly artist listeners**

* + - * 1. **Distance:** fields that represent the amount of space
        2. **Person:**
        3. **Organization:**
        4. **Identifier:** fields that represent a unique entity.

Exp: **User ID 1** is assigned to Alice; since it is an identifier, no one else can have the same **User ID.** However, the **User ID** can be used to indicate how many times Alice streamed music in 2019, how many times Alice listened to a song, and whether she is Subscriber.

* + 1. Under **Semantic sub-type,** you should not be able to select any of the options for **User ID.**
       1. **A semantic sub-type** provides additional information about the **semantic type.**
    2. Under  **Value format**, select Number 12345.
       1. The **value format** will determine how the values are shown in the visualization and how they will be evaluated.
  1. Repeat **Part C (above**) for the contents of the table below.

1. Tip: Use Bulk Actions to configure metadata for multiple fields.
   1. Bulk Actions allow you to select multiple fields and apply the same metadata across them.
      1. Note: Bulk actions will override conflicting metadata.
         1. Exp: if no synonyms are added to the Bulk Action menu, they will not be altered in the individual fields. However, if a field is previously marked as a Measure and the Bulk Action is selected to make fields Dimensions, the field will be come a Dimension.
   2. For the topic, several fields have the name **[table name] ID.** For the purpose of this walkthrough, those fields have the same metadata (role, default aggregation, not allowed aggregation, semantic type, semantic sub-type, and value format).
   3. To apply the same metadata across all of the **[table name] ID** fields, select the box on the left-hand side of the screen for each field you will modify. For Bulk Actions, you must select at least two different fields.
   4. Select **Bulk Actions** at the top of the page.
   5. Apply the following:
      1. Role: **Dimension**
      2. Default Aggregation: Count Distinct
      3. Not allowed Aggregation: Count
      4. Semantic Type: Identifier
      5. Value Format: Number 1234

### Table for Step 2.2

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Column 1 | Column 2 | Column 3 | Column 4 | Column 5 | Column 6 |
| Original Field Name | **Reader-Friendly Field Name** | Original Field Name | **Reader-Friendly Field Name** | Original Field Name | **Reader-Friendly Field Name** |
| **Users Table** |  | **Artists Table** |  | **Streams Table** |  |
| user\_id | User ID | artist\_id | Artist ID | date\_time | Stream Date Time |
| country | n/a | name | Artist Name | platform | Stream Platform |
| join\_date | Join Date | monthly\_listeners | Artist Monthly Listeners | song\_id | Song ID |
| age | n/a | monthly\_payment | n/a | user\_id | User ID |
|  |  |  |  |  |  |
| **Subscribers Table** |  | **Songs Table** |  |  |  |
| subscriber\_id | Subscriber ID | song\_id | Song ID |  |  |
| cost | Subscription Cost | release\_date | Song Release Date |  |  |
| start\_date | Subscription Start Date | genre | Song Genre |  |  |
| user\_id | User ID | total\_streams | n/a |  |  |
|  |  | royalties | Song Royalties |  |  |
|  |  | artist\_id | Artist ID |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Column 1 | Column 2 | Column 3 | Column 4 | Column 5 | Column 6 | Column 7 |
| **Fields** | Dimension/Measure | Default Aggregation | Not allowed Aggregation | Semantic Type | Semantic Sub-Type | Value Format |
| **Users Table** |  |  |  |  |  |  |
| user\_id | Dimension | Count Distinct | Count | Identifier |  | Number |
| country | Dimension | Count Distinct |  | Location |  | Number |
| join\_date | Dimension | Count |  | Date Part |  | Number |
| age | Dimension | Count |  | Date Part |  | Number |
|  |  |  |  |  |  |  |
| **Subscribers Table** |  |  |  |  |  |  |
| subscriber\_id | Dimension | Count Distinct | Count | Identifier |  | Number |
| cost | Measure | Count |  | Currency | USD | Dollar |
| start\_date | Dimension | Count |  | Date Part |  | Number |
| user\_id | Dimension | Count Distinct | Count | Identifier |  | Number |
|  |  |  |  |  |  |  |
| Artists Table |  |  |  |  |  |  |
| artist\_id | Dimension | Count Distinct | Count | Identifier |  | Number |
| name | Dimension | Count | Count Distinct | Person |  | Number |
| monthly\_listeners | Measure | Sum |  | Currency |  | Number |
| monthly\_payment | Measure | Sum |  | Currency | USD | Dollar |
|  |  |  |  |  |  |  |
| **Songs Table** |  |  |  |  |  |  |
| song\_id | Dimension | Count Distinct | Count | Identifier |  | Number |
| release\_date | Dimension | Count |  | Date Part |  | Number |
| genre | Dimension | Count Distinct |  | Organization |  | Number |
| total\_streams | Measure | Count |  | Currency |  | Number |
| royalties | Measure | Count |  | Currency | USD | Dollar |
| artist\_id | Dimension | Count Distinct | Count | Identifier |  | Number |
|  |  |  |  |  |  |  |
| **Streams Table** |  |  |  |  |  |  |
| date\_time | Dimension | Count |  | Date Part |  | Number |
| platform | Dimension | Count | Count Distinct | Organization |  | Number |
| song\_id | Dimension | Count Distinct | Count | Identifier |  | Number |
| user\_id | Dimension | Count Distinct | Count | Identifier |  | Number |

**Step 4: Test Your Topic**

1. Ask QuickSight Q questions and iteratively test your Topic
   1. Asking QuickSight Q questions will ensure the friendly names, synonyms, and metadata behave as expected.
   2. Go to the QuickSight Q bar at the top of your screen. If you do not see the QuickSight Q icon [icon], refer to the [permissions/prereqs].
   3. Select the topic you would like to ask questions for. For this walkthrough, select **Stakeholder Activity 2018-2019.**
   4. Type in the question you would like to ask Q.
   5. While creating this topic, you may want to ask Q the following questions:
      1. How many users are there? (You may want to try this for every ID used in the topic: Artist, Song, Subscriber, etc.)
      2. How many streams are there?
      3. Show me streams in 2019.
      4. Show me users by quarter in 2019.
      5. When Users stream songs
      6. Which platform has the most streams
      7. Which Songs have the most number of streams
      8. Which Artists has the highest number of monthly listeners
      9. How many Users created accounts
      10. How many Users became Subscribers
      11. How much revenue is generated from subscriptions
      12. How old Users are
      13. Where Users are located
   6. By asking questions, you will be able to alter metatags as needed, and Q will learn from those corrections over time.
   7. After you have created your topic, try asking variations of the questions we selected for the topic in Step 2 [link to step 2].
      1. The questions we selected for the topic are listed below for your reference:
         1. Which songs were the most popular in 2019?
         2. Which platform (mobile, desktop, app) was the most popular in 2019?
         3. What was the subscription revenue in 2019?
         4. Show me the monthly traffic on each platform.
         5. What is the traffic on mobile by country?
         6. How many subscribers were recruited in 2019 by month?
      2. Potential variations include (but are not exclusive to):
         1. Top 10 songs
         2. Most popular platform by streams
         3. How many people used the app in 2019?
         4. Show me the subscription revenue by month in 2019.
         5. How many users are from \_\_country\_\_?
         6. how many streams were on desktop?
         7. How many subscribers joined in 2019?
         8. Show me the subscription revenue by quarter
2. Troubleshoot QuickSight Q Answers to Improve Q’s Performance
   1. If QuickSight Q provides an incorrect answer or uses an incorrect term, you can specify terms and alter visualizations to improve Q’s understanding.
   2. Note: if your question has to do with some period of time (month, quarter, 2019, year), try using the terms “in” and “by” before the period of time.
      1. \_\_\_\_ by month (groups the data by month)
      2. \_\_\_\_ in 2019 (filters out the data by year)
      3. \_\_\_\_ by month in 2019 (data grouped by month filtered in 2019)
   3. After asking the question **Which streaming platform was the most popular in 2019**, some terms may be underlined, such as streaming and popular, and some may not be underlined at all, such as **platform**. The underlined terms are terms that Q has designated to some field or terms that you or other users may have specified.
   4. Select the underlined term streaming. QuickSight Q will prompt you with the following actions:
      1. For this topic, select **skip this term**, which will ignore the streaming term. In this question, ignoring the term will allow QuickSight Q to focus on the **number of platform occurrences (mobile, desktop, app)** in **2019.** (With the term streaming, Q may incorrectly evaluate the question using the unique count of streams.
      2. Need to make a change? (search bar)
         1. This action will allow you to replace the term with your desired field. The current field (if any) that the term is referring to will be underneath the searchbar.
         2. [Q bar  need to make a change screenshot]
      3. Add a filter  - this will open a new browser tab
         1. This action will allow you to create a filter that will apply to the dataset. For more information, refer to the [main docs- fields] .
      4. Add a calculated field -  this will open a new browser tab
         1. This action will allow you to create a calculated field that will evaluate an expression you or other authors have created. For more information, refer to the [main docs- calculated fields] .
      5. Skip this term
         1. This action will ensure that Q will not factor the term into its visualization.
   5. If Q is unable to answer your question, it may prompt you with friendly names and ask that the questions be rephrased. In that case, attempt to structure your questions as closely to the list of Q-friendly question types as possible. For more information, refer to the [main docs - q-friendly question types ] .
      1. [SCREENSHOT - question: nothing; answer: options/rephrase the question]
      2. If a key term is not underlined, consider adding the key term to a field’s synonyms .
      3. Adjust the metadata and other details used to link the terms
         1. Add synonyms  (default and not allowed). For more information, refer to step 3 of this walkthrough.
         2. Adjust the aggregations (default and not allowed). For more information, refer to step 3 of this walkthrough.
      4. If you have a specific visualization you would like to link to the question, refer to the [main doc - linking visualizations]

**Conclusion**

* Reiterate everything the user just did
* Point to topics for next steps and to encourage them to continue learning.
* Link to additional resources, if applicable.