Step 3: Update Your Topic to Be Natural Language Friendly

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

1. Exclude fields you don't want to use in the topic.

   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 [icon] to include and exclude the field.

   As a reminder, we are tailoring the topic to answer the following questions from 2018-2019:

   - 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?

   - Show me the monthly traffic on each platform.

   - What is the traffic on mobile by country?

   - How many subscribers were recruited in 2019 by month?

2. For this topic, select the switch [icon] on the right-hand side to exclude the following fields:

   - Monthly Payment (Artists Table)

   - Royalties (Songs Table)

   - Total Streams (Songs Table)

   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.

3. Rename fields to be reader-friendly.

   Readers are individuals that will have access to your topic but not the data itself.

   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.

   Highlight the box and change the field name to a friendly name.

   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.

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.

If Amazon 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]

ⓘ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.

4. Add synonyms to fields to be natural-language friendly.

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].

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.

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.

ⓘ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.

QuickSight Q uses field synonyms to identify key terms within questions.

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?

5. Add synonyms to values to be natural-language friendly

Under the *Data* tab at the top of the page, select the meatball [icon] next to the include switch.

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].

Select *Add* to link values to synonyms.

For this topic, select the meatball [icon] in the *Stream Platform* field.

1. Select *Add* and enter the value *mobile*.

2. For the value *mobile*, add the synonym *phone*.

3. Create the value *app*, and add the synonyms *used the app* and *application*.

4. Create the value *desktop*, and add the synonym *pc* and *computer.*

5. 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.

QuickSight Q uses value synonyms to identify terms within questions.

6. Configure metadata to be natural-language friendly.

Under the *Data* tab, select the arrow icon [icon] under the *Include* column.

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): a category below the semantic type. For more information, refer to the [step 3]

   6. value format: how the field should be visualized

7. For the topic, select the arrow icon [icon] under the *Include* column for the field *User ID.*

8. 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.

9. 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.*

10. 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.*

   Count Distinct will count each unique occurrence of *User ID* that apply.

   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,.

   2. *Count* will count each occurrence of any *User ID* that apply.

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.

11. A *default aggregation* will determine how the field will be used for calculations for a *measure.*

1. *Sum* will add all of the values that apply.

   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.*

2. *Average* will take the mean of all values

   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.*

3. *Max* will return the highest value of those that apply.

   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.*

4. *Min* will return the lowest value of those that apply.

   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.*

12. 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. For a *dimension*, a *not allowed aggregation* will determine how the field _cannot_ be counted.

3. For a *measure*, a *not allowed aggregation* will determine how the field _cannot_ be used for calculations.

13. Under the *Semantic Type,* select the dropdown menu and choose *Identifier.*

A *Semantic Type* identifies a general category of values that the field behaves most similarly to.

Fields can have the following *Semantic Types*. For a full list of semantic types, refer to [link to main docs - semantic types]

- *Duration:* fields that provide a length of time

- *Date/Date Part:* fields that specify a specific point in time (exp: 01/31/2019)
  Note
  Be sure that the format of the date/date part is compatible with QuickSight Q [link to acceptable date/time formats]

- *Location:* fields that signify a geological region

- *Boolean:* fields that represent true, false, or null

- *Currency:* fields that represent money or used to signify quantities
  Note

the *Currency* semantic type can be used for fields that represent revenue or cost, as well as *monthly artist listeners*. For this walkthrough, we will be using USD.

- *Distance:* fields that represent the amount of space

- Person: fields that represent an individual's personal information that are not used as identifiers.

- Organization: fields that represent how the data will be grouped.

- *Identifier:* fields that represent a unique entity.

  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.

14. Under *Semantic sub-type,* you should _not_ be able to select any of the options for *User ID.*

    A semantic sub-type provides additional information about the *semantic type.* For more information about semantic sub-types, refer to [link to main docs-semantic subtypes

15. Under *Value format*, select Number 12345.

    The *value format* will determine how the values are shown in the visualization and how they will be evaluated.

16. Repeat the step above for the contents of the table below.

17. Use Bulk Actions to configure metadata for multiple fields.

    Bulk Actions allow you to select multiple fields and apply the same metadata across them.
    ⓘNote
    Bulk actions will override conflicting metadata.

    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.

    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).

    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.

    1. Select *Bulk Actions* at the top of the page.

    2. Apply the following:

        - Role: *Dimension*

        - Default Aggregation: Count Distinct

        - Not allowed Aggregation: Count

        - Semantic Type: Identifier

- Value Format: Number 1234