Understanding Coaching Options for Spark

> Alchemer Dashboard is currently waitlist only. Visit this page to learn more about Dashboard or join the waitlist!

Spark provides two different coaching options (in addition to the options already covered under data modeling):

- Reference questions

- Business terms

Reference questions

Reference questions are just a set of sample questions and their corresponding answers (in Alchemer Dashboard's Search keyword language).

How reference questions help Spark

The reference questions provided to Spark help it in the following ways :

Spark provides the verified answer when users ask the same question
The reference questions are designed in a way to provide the exact same answer when users ask the same question. We recommend adding the most commonly asked questions to the reference questions so that business users get a verified answer that is curated by an analyst.
For the questions specified under reference questions, Spark also remembers the visualization settings set by the analyst. Spark remembers the chart type selected and the axis settings of the visualizations (other visualization settings may not be retained).
Hence, one of the direct benefits of providing reference questions is to make sure Spark provides an accurate response for commonly asked questions.

Spark learns how different columns should be used to answer questions
An indirect benefit of providing reference questions is that it helps in warming up our usage-based ranking systems. The reference questions help our system learn which columns must be selected in answering questions.
We have observed that for newly created datasets where the usage-based ranking system isn't warmed up, Spark might struggle to pick the correct columns if there are similar-sounding column names. Adding reference questions coaches Spark to understand which columns should be used for answering questions.
We recommend that you add reference questions that use the majority of the columns present in your Model/Worksheet. Ensuring a good column coverage ensures that Spark understands when and how to use the columns available in your Model/Worksheet.

Guidelines for reference questions

Add commonly-asked questions in reference questions

Spark provides the answer you have curated when a user asks the same question. Hence, we recommend that you add the most commonly-asked questions by the business user as reference questions so that users get a verified answer every time they ask these questions.

Coached question templates

You should think of reference questions as coached question templates. For example, once you have provided "What is the sales for east this month ?" as a reference question, then Spark has the ability to generalize the learnings from this question to answer similar questions like "What is the sales for west this month ?" , "What are the sales for north this year ?" , "What is the quantity sold for east this week?", etc. (Please note that the visualization settings provided in query feedback may not extend to these questions)

Ensure good column coverage

As mentioned above, Spark learns how to use columns for analysis when they are used in reference questions. Hence, we recommend that you provide enough reference questions so that all the important columns in your Model/Worksheets are represented in the reference questions. For newly created datasets with no Dashbords, the reference questions play an extremely critical role in coaching Spark.

Business terms

Business terms give you an option to create a mapping between your data and the business term/vocabulary used in day-to-day operations. For example, the "active clusters" example we saw in Coaching Spark is an example of business terms.

How business terms help Spark

Business terms are a mapping of the data to the vocabulary used by the business users in day-to-day life. Every business has its own definition of various metrics and a very specific way for calculating these metrics. The business terms allow an analyst to add these mappings to the Spark knowledge to ensure that the LLM uses your definition of a metric instead of making assumptions to calculate these metrics.

We recommend using business terms for providing the following types of knowledge to Spark:

Define how to calculate certain metrics

In order to answer questions, you sometimes need to generate calculated fields or metrics. You can use the business term to teach the Spark on how to calculate specific metrics for your business. Spark learns how you calculate specific metrics and it has the ability to extrapolate these learnings to different scenarios.

Define synonyms for column values

Often, the same value might be referred to with different names. We recommend the use of business terms to define the synonyms for value. **While Spark has the ability to apply semantic**

matches based on publicly-known information, there are always some cases where you want to define synonyms or acronyms which are only applicable to your own business and cannot be guessed by a large language model.

Filters definition

Business terms can be used to define filters that should be applied for specific cases. For example, the active clusters example demonstrated how you can use the business term to define specific filters to be applied while calculating things.

Guidelines for business terms

Our system suggests business terms to you when you are coaching Spark for reference questions or correcting the answer during conversation. You must only add business terms which meet the following criteria :

Meaningful addition to Spark knowledge

The first thing you should evaluate is whether adding the business terms will lead to a meaningful addition to Spark knowledge about your data. Anything which is obvious or common knowledge can already be guessed by the LLM models as they are coached on large corpus of public datasets. Hence, try to avoid adding the definition for commonly available terms in business terms, as the LLM will already provide correct responses.

Business terms are most helpful for addition of definitions which are specific to your organization and cannot be guessed by business users. These are the business terms which will provide a meaningful addition to Spark knowledge.

Note there are terms like "sales contribution" which may seem fairly obvious at first, however, when you deep-dive into specifics of calculating it, you will realize that the way your business computes this is very specific. Hence, we recommend that you provide some sample to Spark to ensure these definitions are calculated according to your business requirements only.

Hold the same meaning in all contexts

Business terms, once defined, are considered to apply uniformly across all the question contexts. You must only use business terms for adding analytical definitions for the terms which hold the same meaning in all the different contexts it can be used for that Model/Worksheet.
You should avoid addition of business terms just to coach Spark on handling date intents better because in most datasets there may be more than one date column, and once you have used business terms to define date intent (e.g. this year → created_date.this year), then Spark might start using the same definition when `this year` is used in a different context.

Business terms and their analytical definition are correct

You should only accept the business terms whose analytical definition (represented by Alchemer Dashboard keyword-based search tokens) is correct.

Related Articles