

Lab 3: Evaluate Your Data Agent

Pre-requisite

1. You have completed Lab 1.
2. You have downloaded all files from <https://aka.ms/fabconeu25-dataagent-tutorial>.

Learning Objectives

1. Apply your learning of Prompt engineering to build the best Data Agent possible for the Data Set
2. Use evaluation notebook to evaluate the quality of your Data Agent, and iteratively improve on your Data Agent using evaluation notebook results.

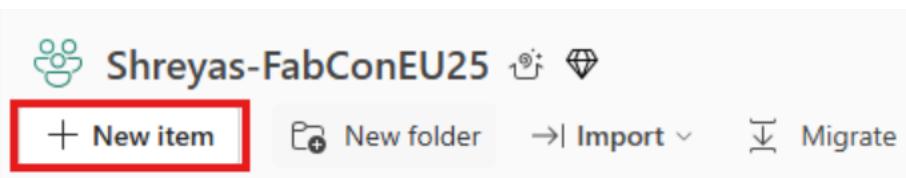
Lab Steps

Step 1: Create a NEW Data Agent

1. Return to your workspace through the Multi-tasking Nav:



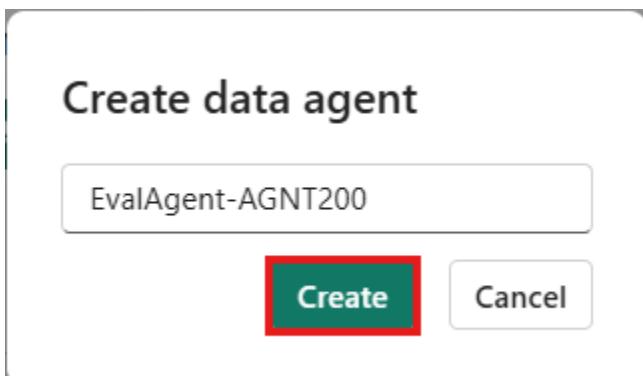
2. Select “+ New Item”



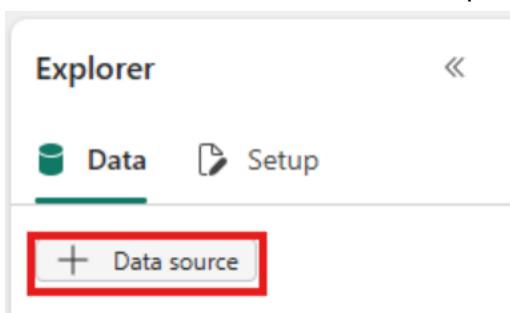
3. Search for “Data Agent” and Select “Data agent”

The image shows a search results page for "Data Agent". It includes a search bar with the text "Data Agent" (highlighted with a red box) and a "New item" button. Below the search bar, there are two tabs: "Favorites" and "All items" (selected). The results section shows a card for "Data agent (preview)" with a description: "Build a generative AI agent that understands your data and can answer complex questions in a variety of conversational interfaces." A star icon is at the bottom right of the card.

4. Name the Data Agent “EvalAgent-{WorkspaceName}” -> “Create”



5. Click on “+ Data source” in the Explorer



6. Search for your Data source using its name, select it, and hit “Add”

The screenshot shows the 'Add a data source' search results page. At the top, there is a search bar with the text 'Kaggle' and a 'Filter' button. Below the search bar is a table with columns: Name, Type, Owner, Refreshed, Endorsement, and Sensitivity. A single row is visible, showing 'Kaggle' as the Name, 'Lakehouse' as the Type, and 'Markus Cozowicz' as the Owner. This row is also highlighted with a red box.

Name	Type	Owner	Refreshed	Endorsement	Sensitivity
Kaggle	Lakehouse	Markus Cozowicz	—	—	—

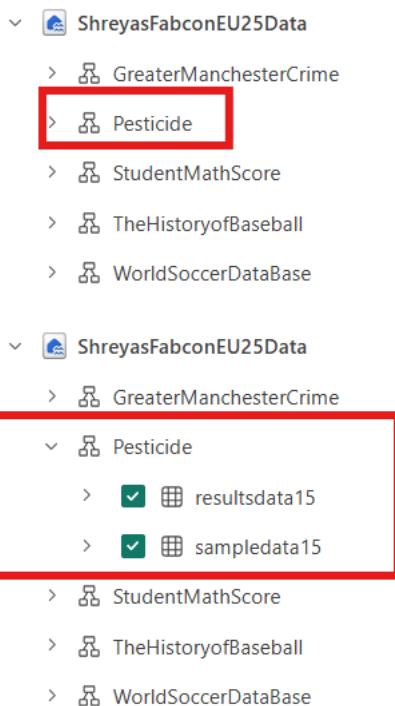
7. You should see the lakehouse with all the tables we've loaded added to the data agent:

The screenshot shows the Databricks Data Explorer interface. At the top, there are tabs for 'Data' (which is selected) and 'Setup'. Below the tabs, a message box states: 'No tables selected yet. The data agent can't answer questions until at least one table for a data source is selected.' A button labeled '+ Data source' is present. The main area displays a hierarchical list of data sources and their tables. One data source, 'ShreyasFabconEU25Data', is expanded and highlighted with a red box. It contains three tables: 'GreaterManchesterCrime' (with a single table 'greatermanchestercrime'), 'Pesticide' (with two tables 'resultsdata15' and 'sampledata15').

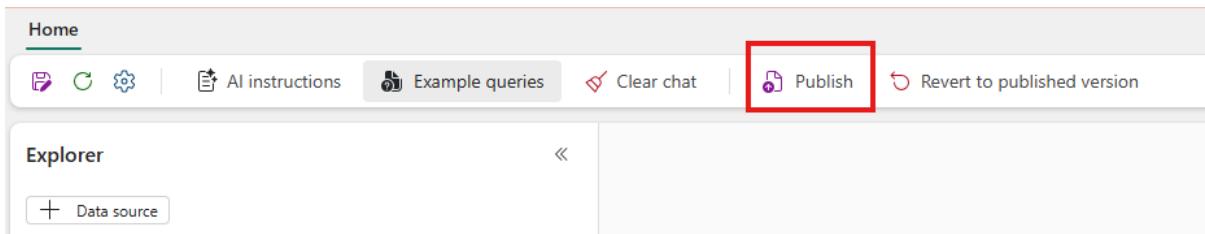
8. Pick 1 of the table Data Sets below, and select every table under it:

Dataset	Difficulty Rating
Pesticide	***
StudentMathScore	*
The History of Baseball	*
WorldSoccerDatabase	**

9.



10. Publish the Data Agent



Publishing the Data Agent allows Data Agent to be consumed by downstream channels. In this lab, we will use the published data agent and run an evaluation across benchmark questions, and evaluate how well the Data Agent performs.

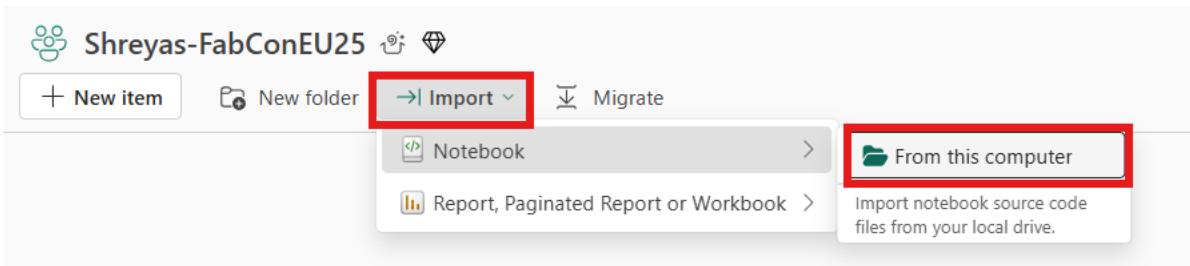
Step 2: Load the Evaluation notebook, and run it

TODO: Write the Evaluation notebook based on the Dataset

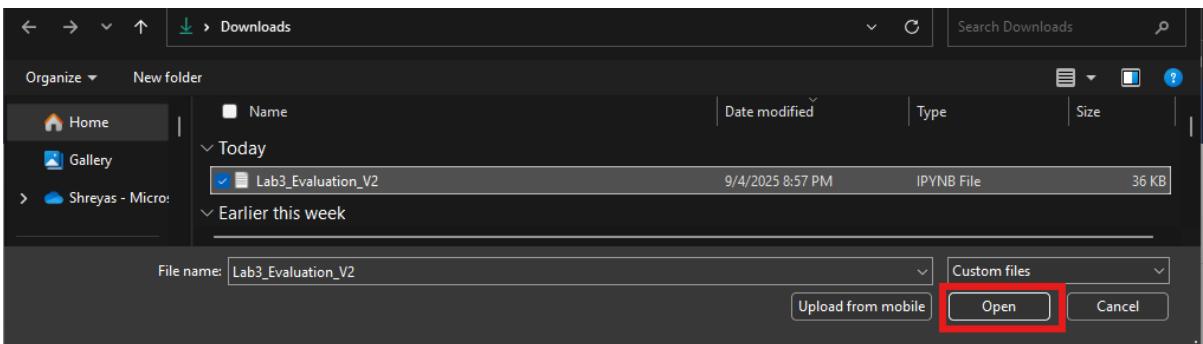
1. Return to your workspace using the Multi-tasking nav:



2. Select Import-> Notebook -> From this computer



3. Select the "Lab3_Evaluation_V2.ipynb" file



4. Open the Notebook:

Name
FabconEU25-141-Eval
FabconEU25-Shreyas
<input type="checkbox"/> Lab3_Evaluation_V2
<input type="checkbox"/> load_data_v2

5. Select “Add data Items” -> “New lakehoues”



No data sources added

Add data items ▾

New lakehouse

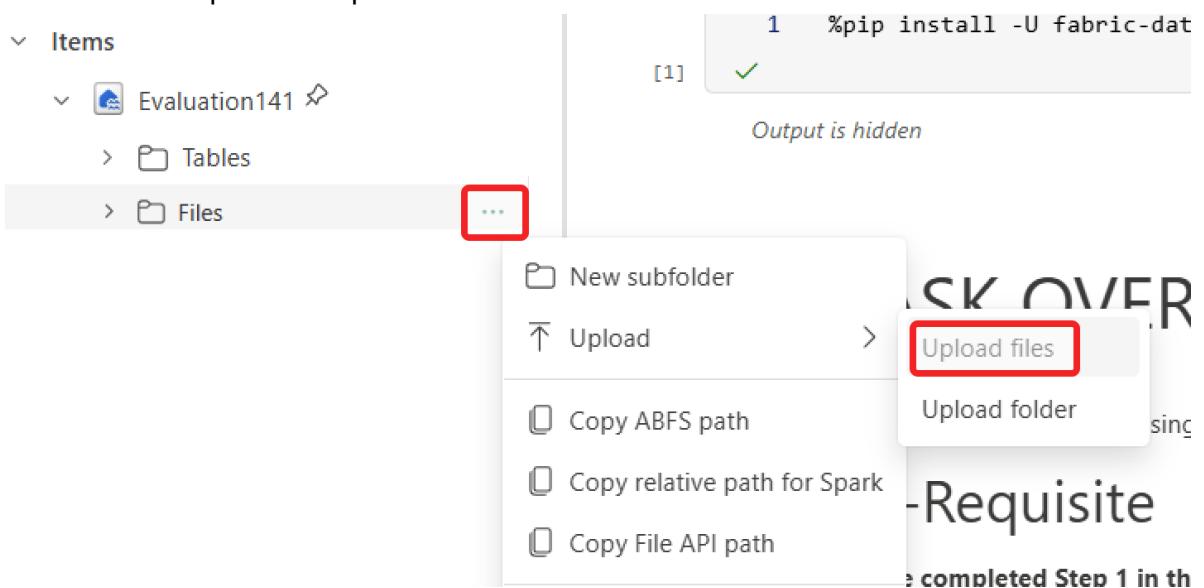
Existing data sources

6. Name the lake house “Evaluation###” where the 3 numbers is your account number

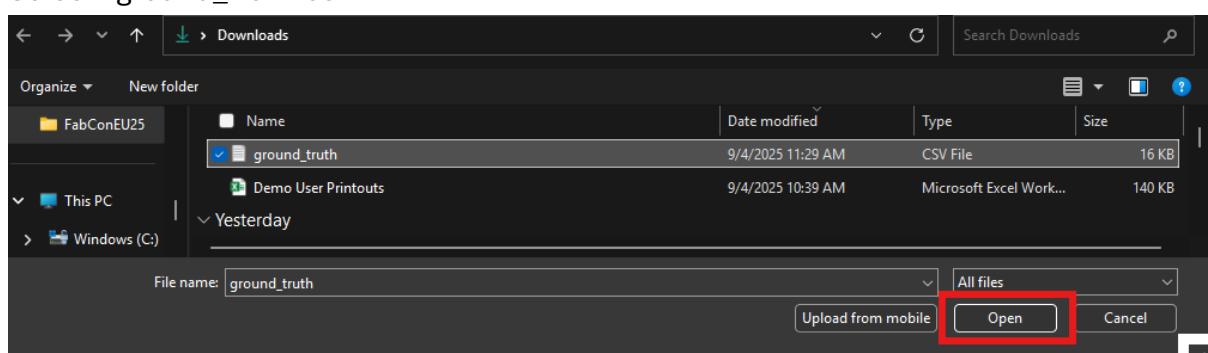
New lakehouse

Name *

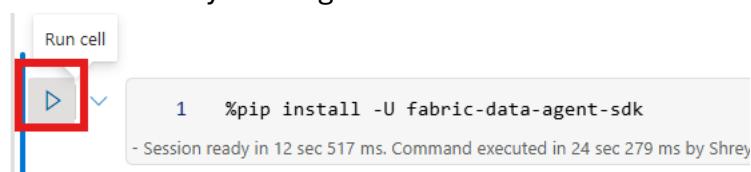
7. Go to: Files -> Upload -> Upload files



8. Select “ground_truth.csv”



9. Start this lab by clicking the “Run” Button in the Fisrt cell with “pip install” in it:



10. Update the “YOUR_DATA_AGENT_ARTIFACT_NAME” and “YOUR_ACCOUNT_NUMBER” variables in the Notebook:

 STEP 1: Use the Evaluation Notebook

1. **Update the YOUR_DATA_AGENT_ARTIFACT_NAME**
 - o Replace the placeholder with the name of your Data Agent Artifact from in Step 1 of the “Lab 3 Curate Your Data Agent” Document.
2. **Update the YOUR_ACCOUNT_NUMBER Variable**
 - o Replace the placeholder with your name

```
1  from fabric.dataagent.client import FabricDataAgentManagement
2  from fabric.dataagent.evaluation import evaluate_data_agent, get_evaluation_summary, get_evaluation_details, get_evaluation_summary_per_question
3  #from fabric.dataagent.evaluation.evaluator_api import add_ground_truth_batch
4  import openai
5  import os
6  import re
7
8  YOUR_DATA_AGENT_ARTIFACT_NAME = "EvalAgent-AGNT200"
9
10 YOUR_ACCOUNT_NUMBER = "200"
11
12 os.environ["OPENAI_API_VERSION"] = "2023-05-15"
13
14 data_agent = FabricDataAgentManagement(YOUR_DATA_AGENT_ARTIFACT_NAME)
```

✓ - Command executed in 18 sec 530 ms by Shreyas Canchi Radhakrishna on 10:55:47 PM, 9/08/25

- a. After this step, you should be able to run through each step in the notebook.
Please follow along and read the directions as you go!
- b. You should have been able to run all the cells in the Notebook, and get a result from the evaluation.

Step 3: Improve Configuration Set Up and Pass All Evaluation

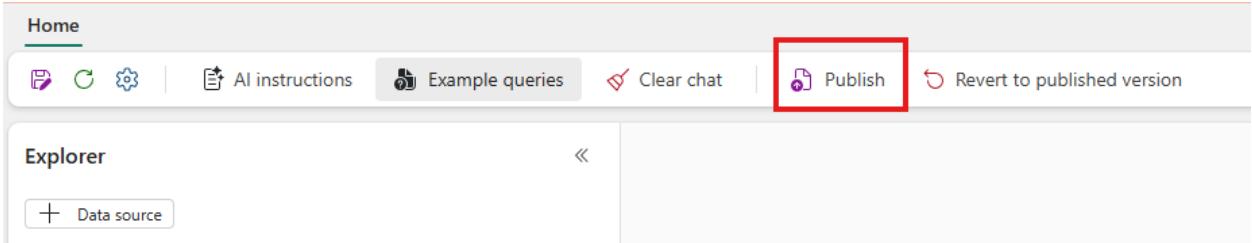
1. Identify all the Questions the Evaluation failed in:

index	T	F	?	%	failed_threads	question
0	1	0	0	100.0	No failed records	Average lifespan of hall of fame players
1	0	1	0	0.0	1	For award winners, what's average weight for each position
2	0	1	0	0.0	1	For award winners, which is the position that has the most hall of fame players?
3	0	1	0	0.0	1	For every award, who is the oldest winner?

Note: While the links above allow you to view a chat thread with the Data Agent, this link will only provide a **Consumer View**. This means that you will not be able to edit ANY configurations. In order to make configuration changes, open your Data Agent from your Workspace.

2. Open your Data Agent and update the configurations in the Data Agent to address the failed questions. Use the tools you discovered in Lab 2:

- a. Add Agent Instructions
 - b. Add Data Source Instructions
 - c. Add Example Queries
3. Publish the Data Agent:



4. Re-run the “ STEP 3: Evaluation” python cell in the Evaluation Notebook:

```
1 # The table used to store evaluation results
2 evaluation_table_name = f'evaluation_output_{YOUR_NAME}'
3
4 # run evaluation
5
6 eval_id = evaluate_data_agent(
7     df = df_pandas_groundtruth,
8     data_agent_name = YOUR_DATA_AGENT_ARTIFACT_NAME,
9     data_agent_stage = "production",
10    table_name = evaluation_table_name)
11
12 df = get_evaluation_summary_per_question(eval_id, verbose=True, table_name = evaluation_table_name)
13
14
```

Goal: The Data Agent should answer ALL questions in the Evaluation notebook.

Lab Feedback

Once you have completed the lab, please fill out the following form to provide feedback for the session so far: <https://forms.microsoft.com/r/2hnLST2yM0>