Lab 3: Evaluate Your Data Agent

Pre-requisite

1. You have completed Lab 1.

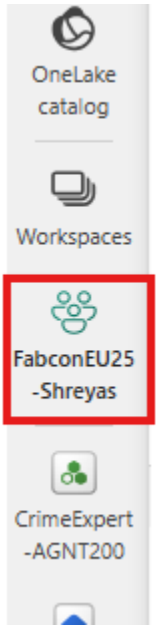
Learning Objectives

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.

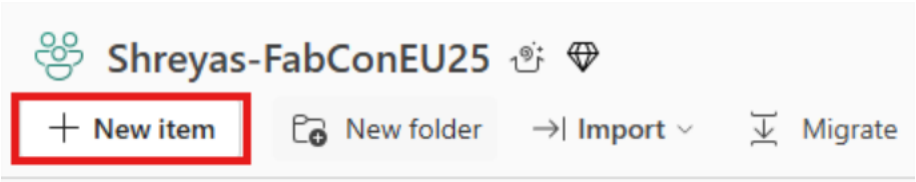
Lab Steps

Step 1: Create a NEW Data Agent

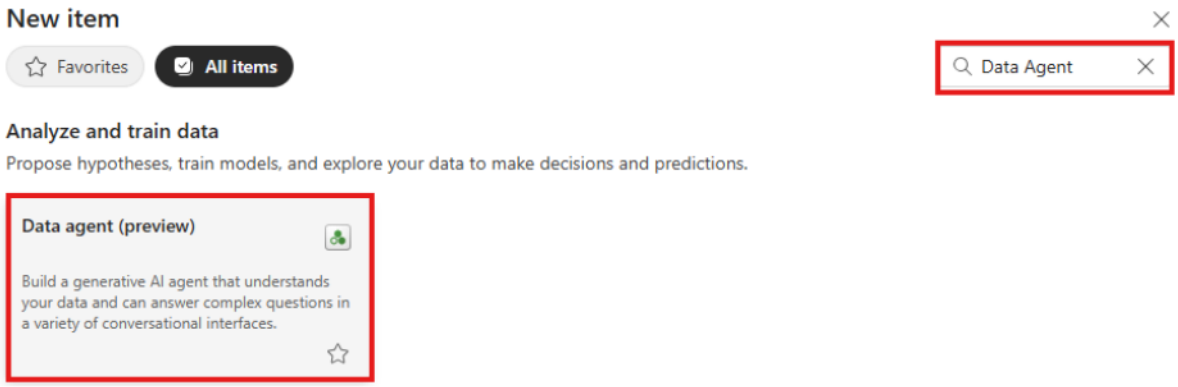
1. Return to your workspace through the Multi-tasking Nav:



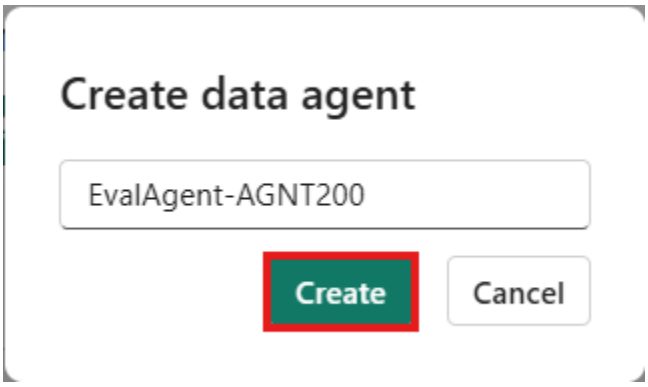
2. Select "+ New Item"



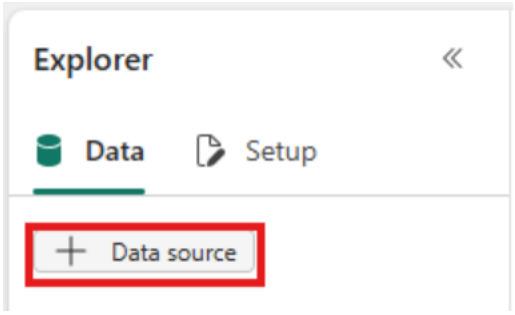
3. Search for "Data Agent" and Select "Data agent"



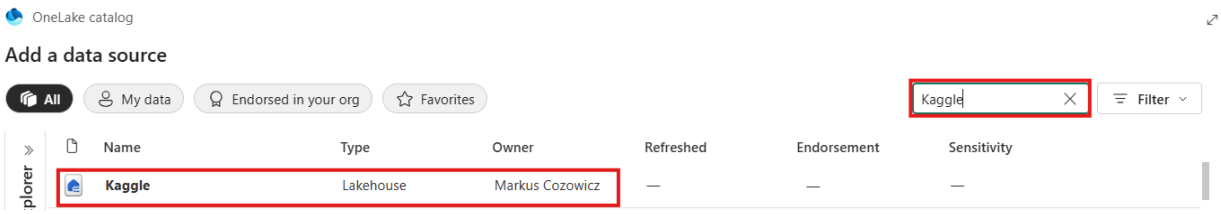
4. Name the Data Agent “EvalAgent-{WorkspaceName}” -> “Create”



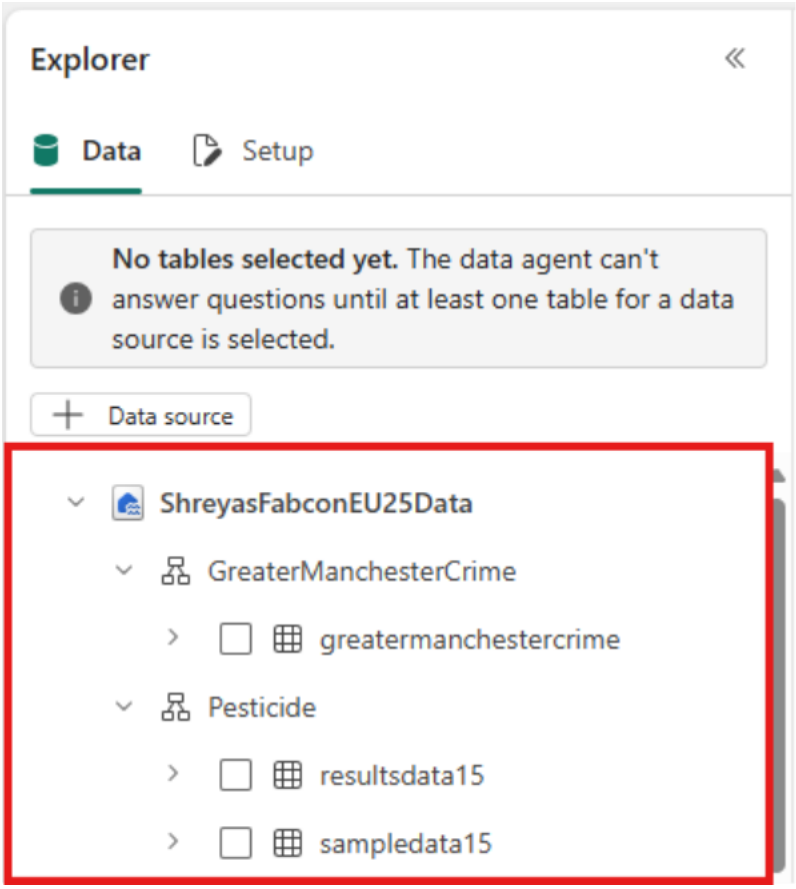
5. Click on “+ Data source” in the Explorer



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



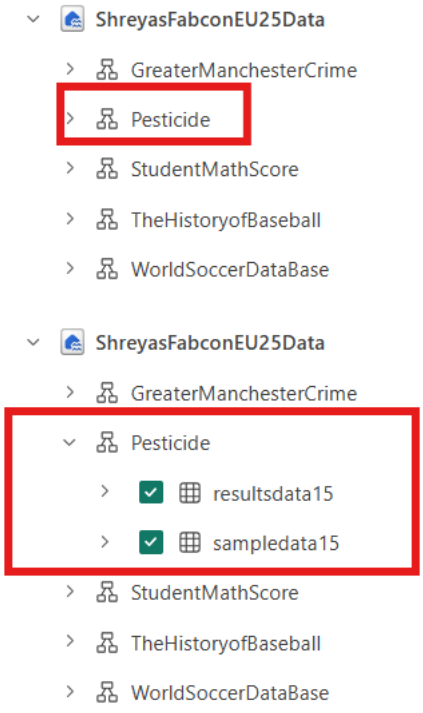
7. You should see the lakehouse with all the tables we've loaded added to the data agent:



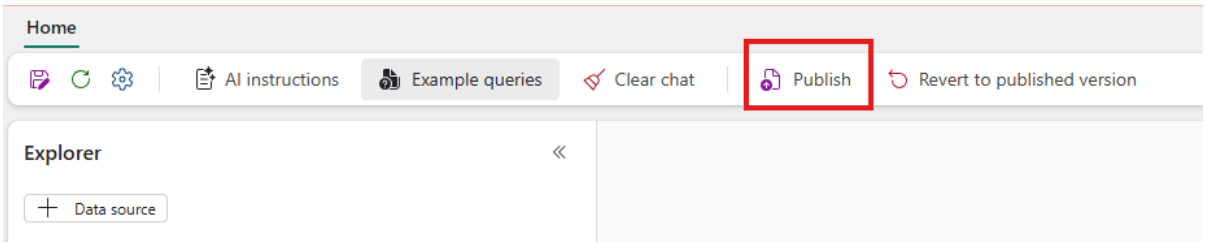
8. Pick 1 of the table Data Sets below.:

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

9. Select ALL the Tables under the Dataset you chose.



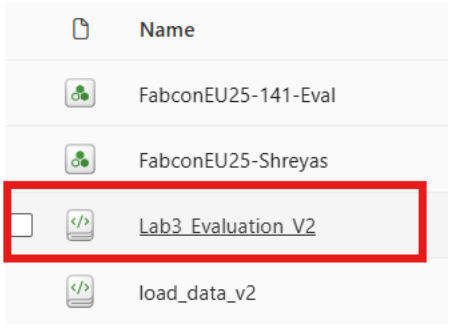
10. Publish the Data Agent



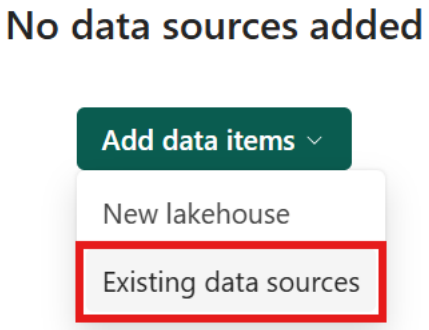
Publishing the Data Agent allows Data Agent to be consumed by downstream channels. In this lab, we will use the published data agent and run an evaluation across benchmark questions, and evaluate how well the Data Agent performs.

Step 2: Load the Evaluation notebook, and run it

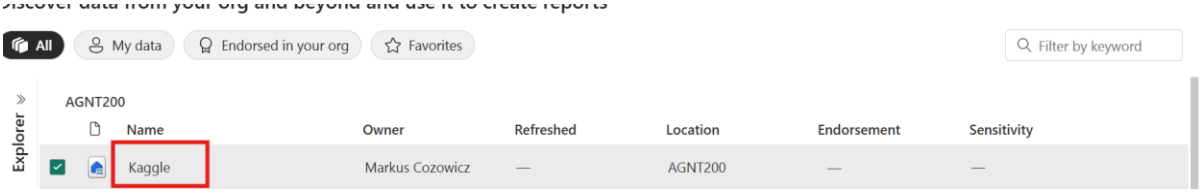
1. Open the Evaluation Notebook “Lab3 Evaluation V2”:



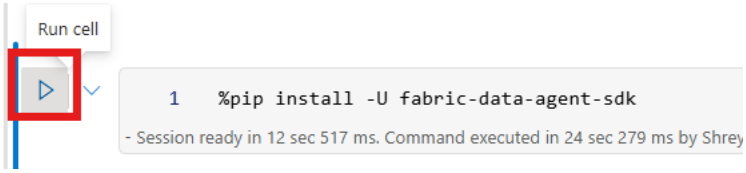
2. Select “Add data Items” -> “Existing data sources”



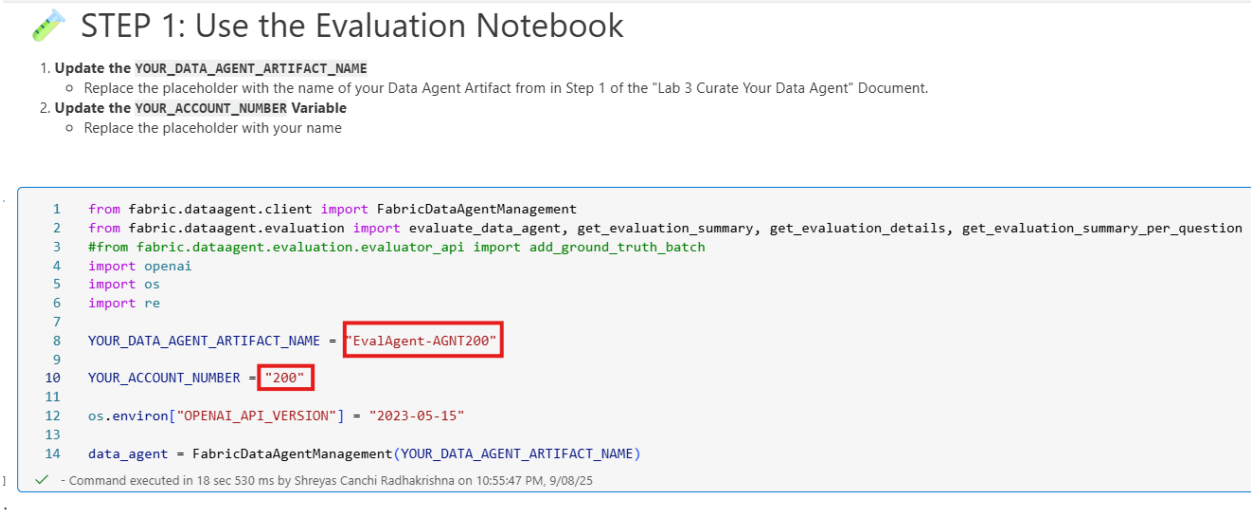
3. Select “Kaggle” -> “Connect”



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



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



   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:

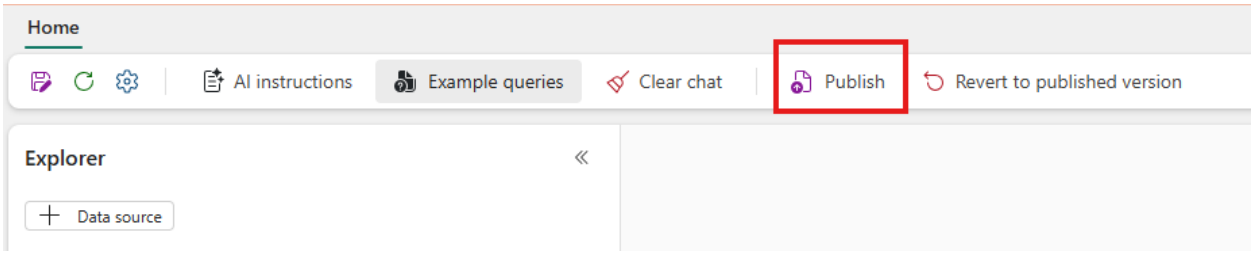
index	T	F	?	%	failed_threads	question
0	1	0	0	100.0	No failed records	Average lifespan of hall of fame players
1	0	1	0	0.0	1	For award winners, what's average weight for each position
2	0	1	0	0.0	1	For award winners, which is the position that has the most hall of fame players?
3	0	1	0	0.0	1	For every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

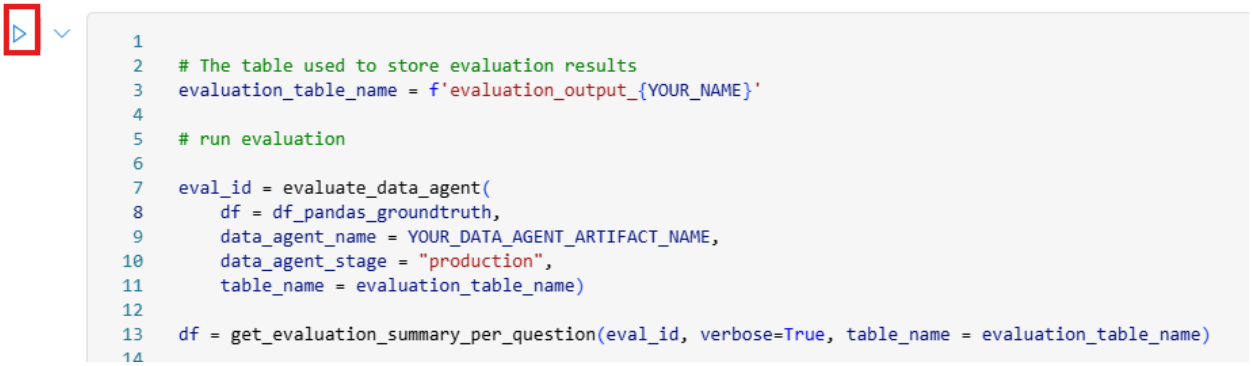
2. Open your Data Agent and update the configurations in the Data Agent to address the failed questions. Use the tools you discovered in Lab 2:

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:



```
# The table used to store evaluation results
evaluation_table_name = f'evaluation_output_{YOUR_NAME}'

# run evaluation

eval_id = evaluate_data_agent(
    df = df_pandas_groundtruth,
    data_agent_name = YOUR_DATA_AGENT_ARTIFACT_NAME,
    data_agent_stage = "production",
    table_name = evaluation_table_name)

df = get_evaluation_summary_per_question(eval_id, verbose=True, table_name = evaluation_table_name)
```

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 this lab: https://forms.microsoft.com/r/2hnLST2yM0