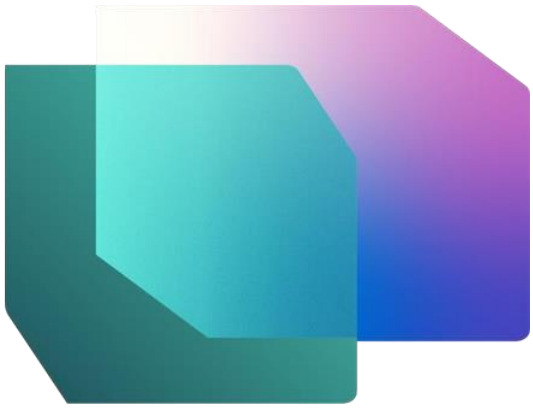
Microsoft Fabric

Chat with your Data in a Day

Lab #1

Version: September 2025



Document Structure 3
Scenario / Problem Statement 3
Introduction 3
Clean data with dataflows gen2 3
Task 1: Setting up a workspace 3
Task 2: Creating the dataflow gen2 7
Task 3: Uploading data 11
Task 4: transforming data with Copilot 13
Step 1: Format table 14
Step 2: Clean data 14
Step 3: remove redundant data 16
Step 4: Update to correct data types 17
Appendix 18
Answers for what need to be cleaned/transformed in the data 18
Troubleshooting 19
Failed to create the dataflow 19
Failed to get the existing dataflow names 22

Document Structure

The lab includes steps for the user to follow along with associated screenshots that provide visual aid. In each screenshot, sections are highlighted with red boxes to indicate the area(s) user should focus on.

Scenario / Problem Statement

You are getting data ready to create a Power BI report. The CSV you're working with does not have clean data. You'll use Fabric dataflow Gen2 to clean and transform the data with natural language

Current Challenges

- Data is not consistent within fields
- Null data
- Duplicate data

Introduction

Cleaning and transforming data with Microsoft Fabric is easier than ever. Instead of pointing and clicking to do the transformations, you can work with Copilot to get the desired results and clean and transform your data.

By the end of this lab, you will have learned:

- How to access Copilot in a dataflow Gen2
- Examples of natural language questions Copilot supports to transform
- How to use Copilot to clean and transform your data

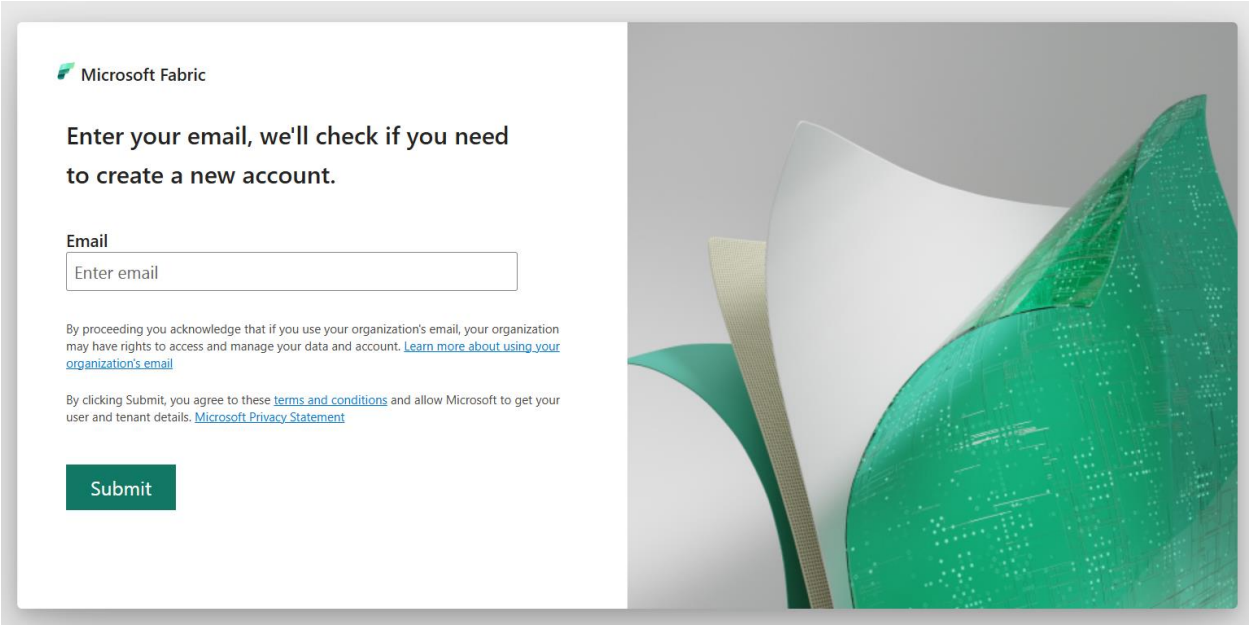
Clean data with dataflows gen2

Task 1: Setting up a workspace

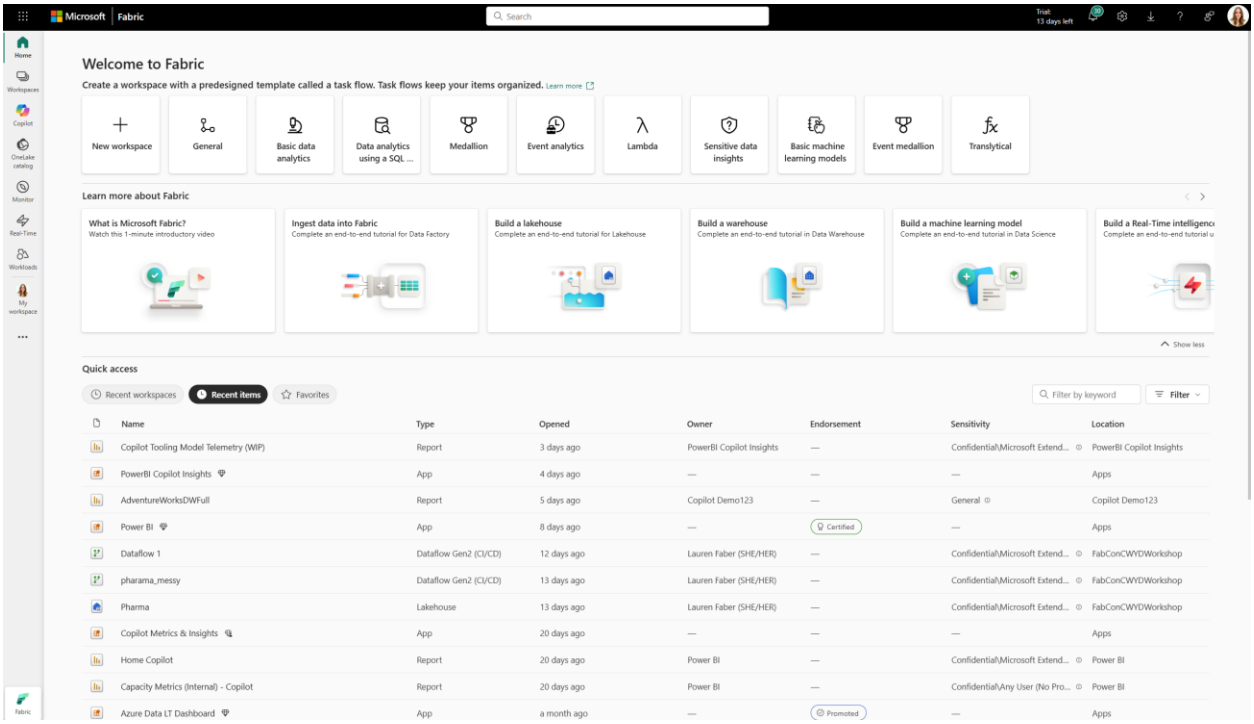
In this section we will get your Fabric workspace created so that you can upload your data to clean and transform it.

1. Navigate to https://app.fabric.microsoft.com

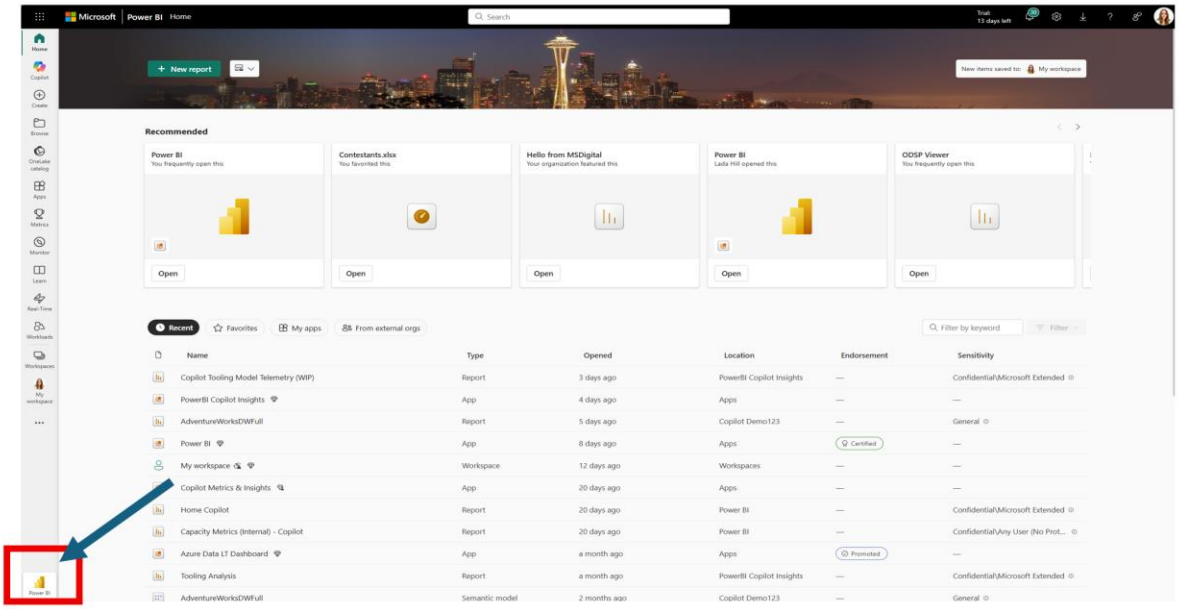
2. Enter the email provided on your printed credentials

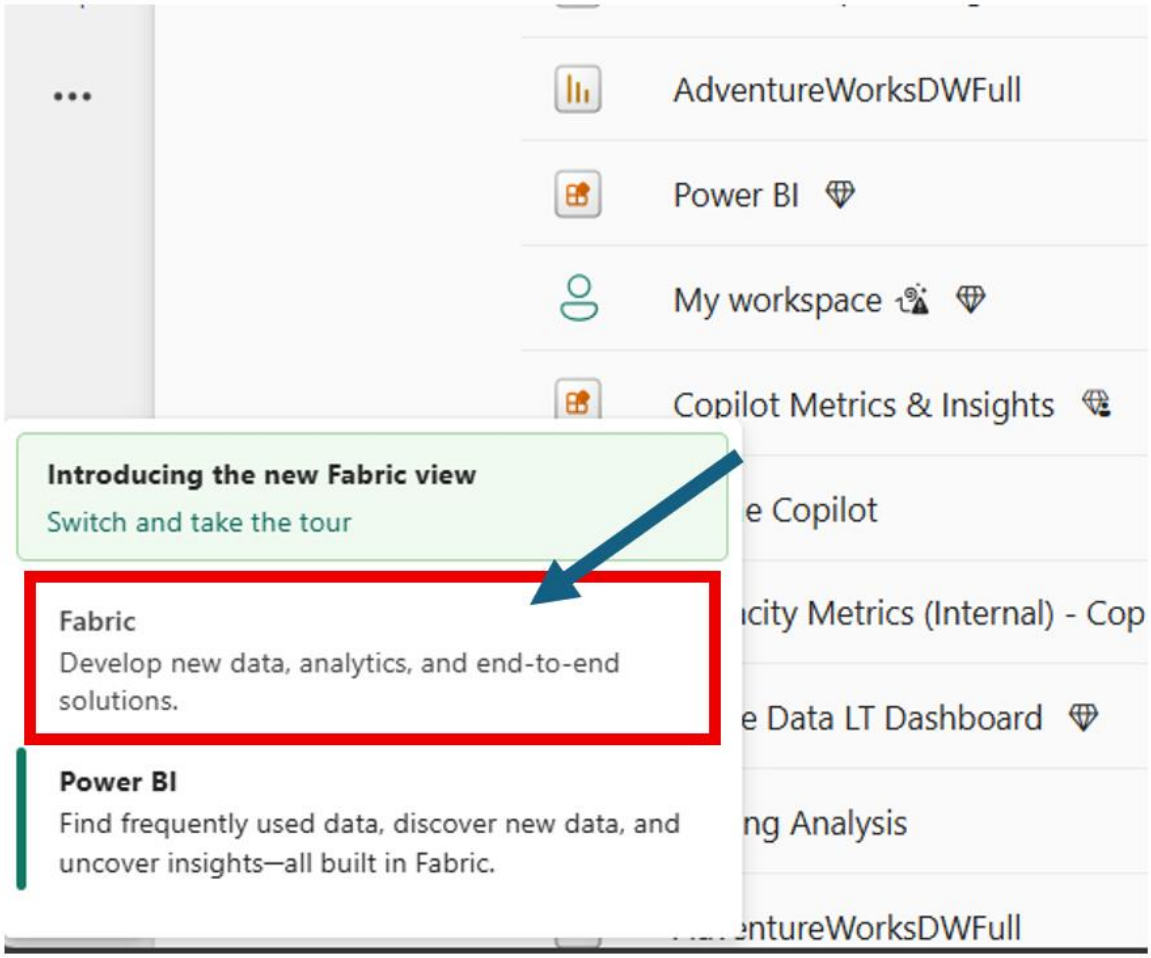


3. Enter the password provided on your printed credentials
4. Land on the Fabric homepage

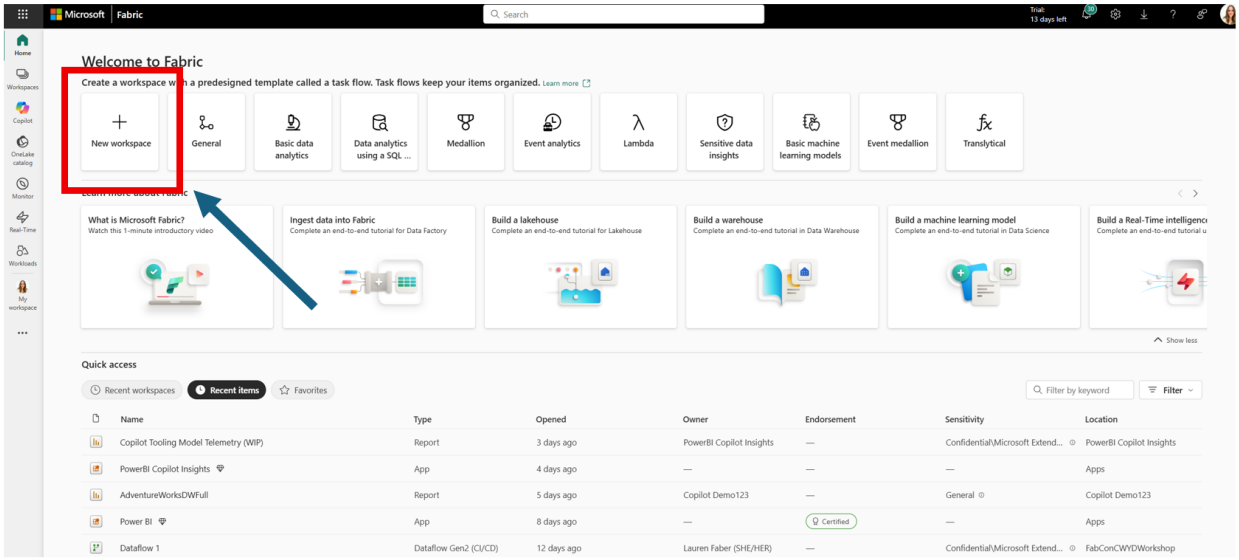


5. If you see the Power BI homepage instead of fabric, navigate to the Fabric home page by clicking on the Power BI button in the bottom left corner of the browser. Choose “Fabric”

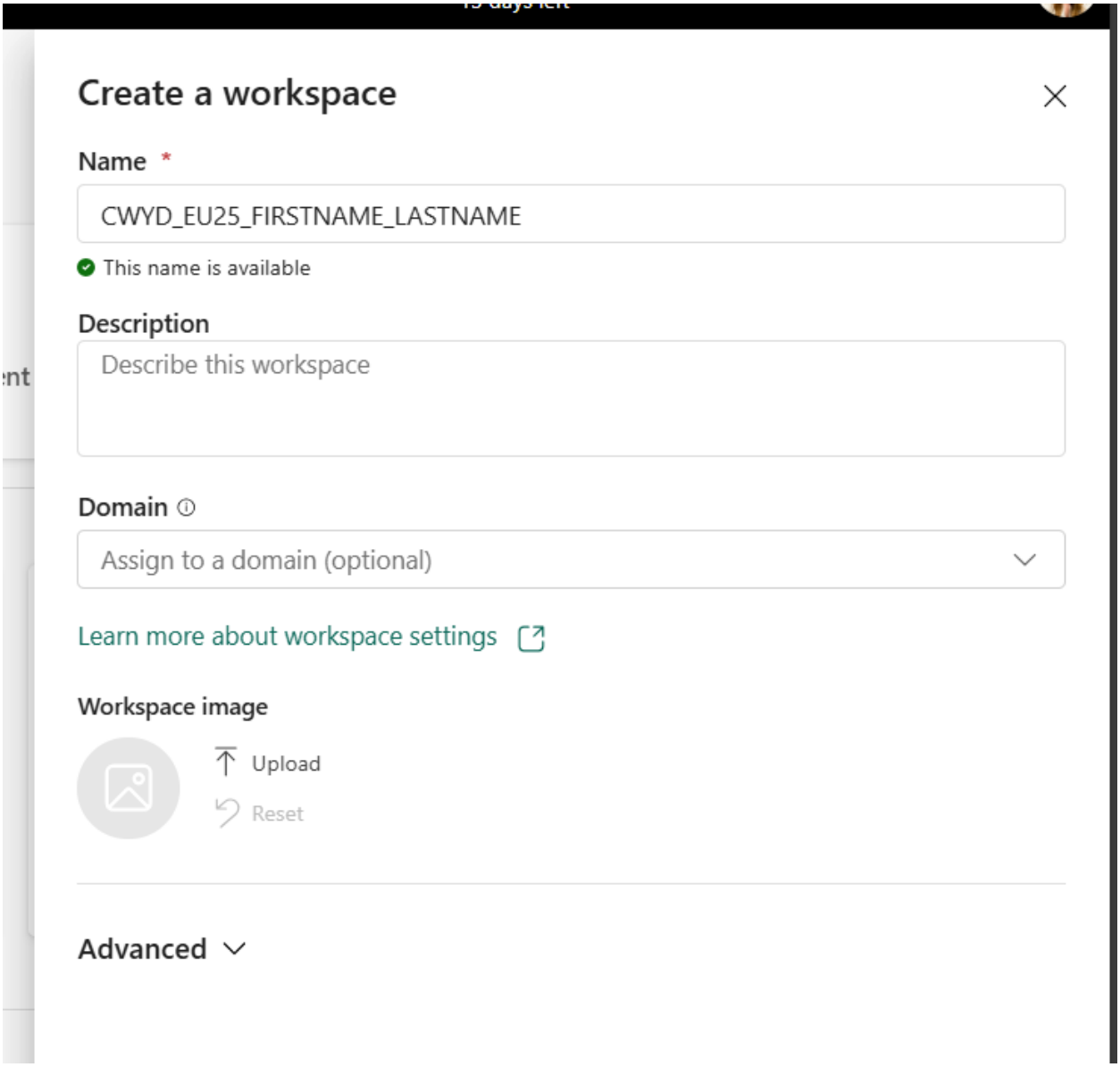

Microsoft | Power BI Home
New report
Recommended
Power BI
Contestants.xlsx
Hello from MSDigital
Power BI
ODSP Viewer
Open
Recent
Favorites
My apps
From external orgs
Name
Type
Opened
Location
Endorsement
Sensitivity
Copilot Tooling Model Telemetry (WIP)
PowerBI Copilot Insights
AdventureWorksDWFull
Power BI
My workspace
Copilot Metrics & Insights
Home Copilot
Capacity Metrics (Internal) - Copilot
Azure Data LT Dashboard
Tooling Analysis
AdventureWorksDWFull


AdventureWorksDWFull
Power BI
My workspace
Copilot Metrics & Insights
Introducing the new Fabric view
Switch and take the tour
Fabric
Develop new data, analytics, and end-to-end solutions.
Power BI
Find frequently used data, discover new data, and uncover insights—all built in Fabric.

6. Select “New workspace” in the top left.



7. Name the workspace your CWYD_EU25_FIRSTNAME_LASTNAME
    a. For example I would name my workspace CWYD_EU25_LAUREN_FABER



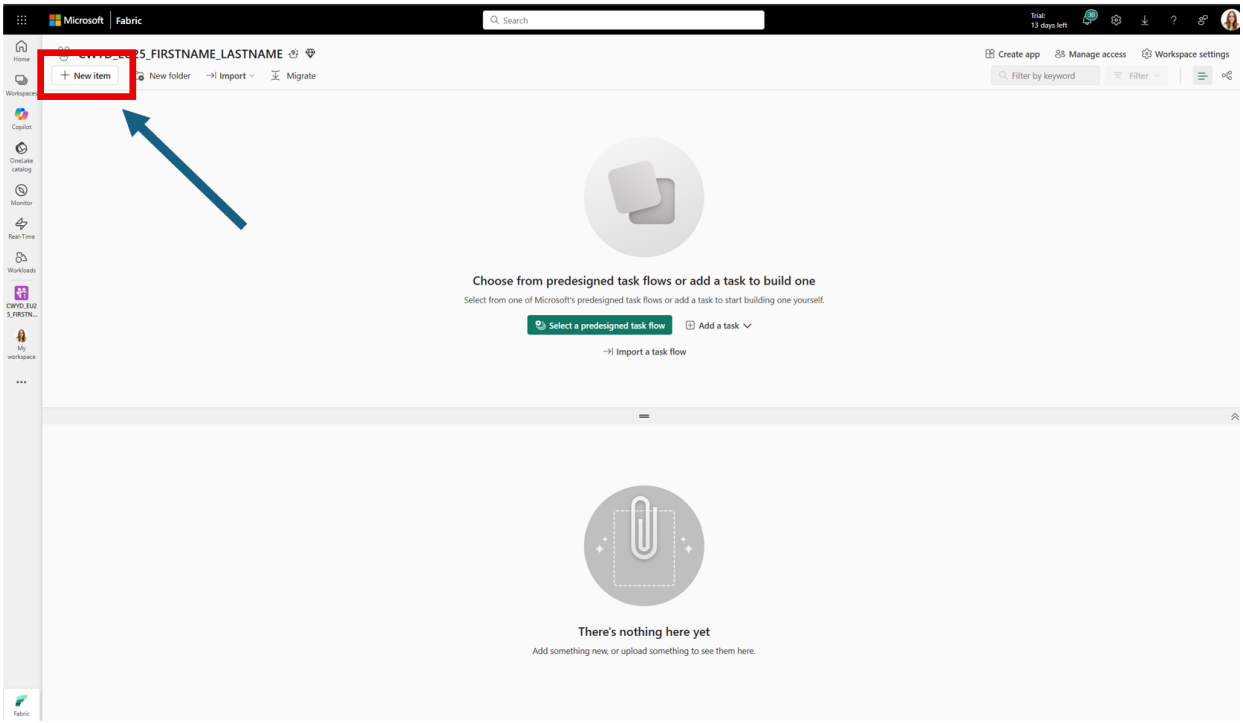
8. Click “Apply”

Congrats! You’ve completed the first step of the workshop. We recommend bookmarking this page in your browser for the remainder of the day.

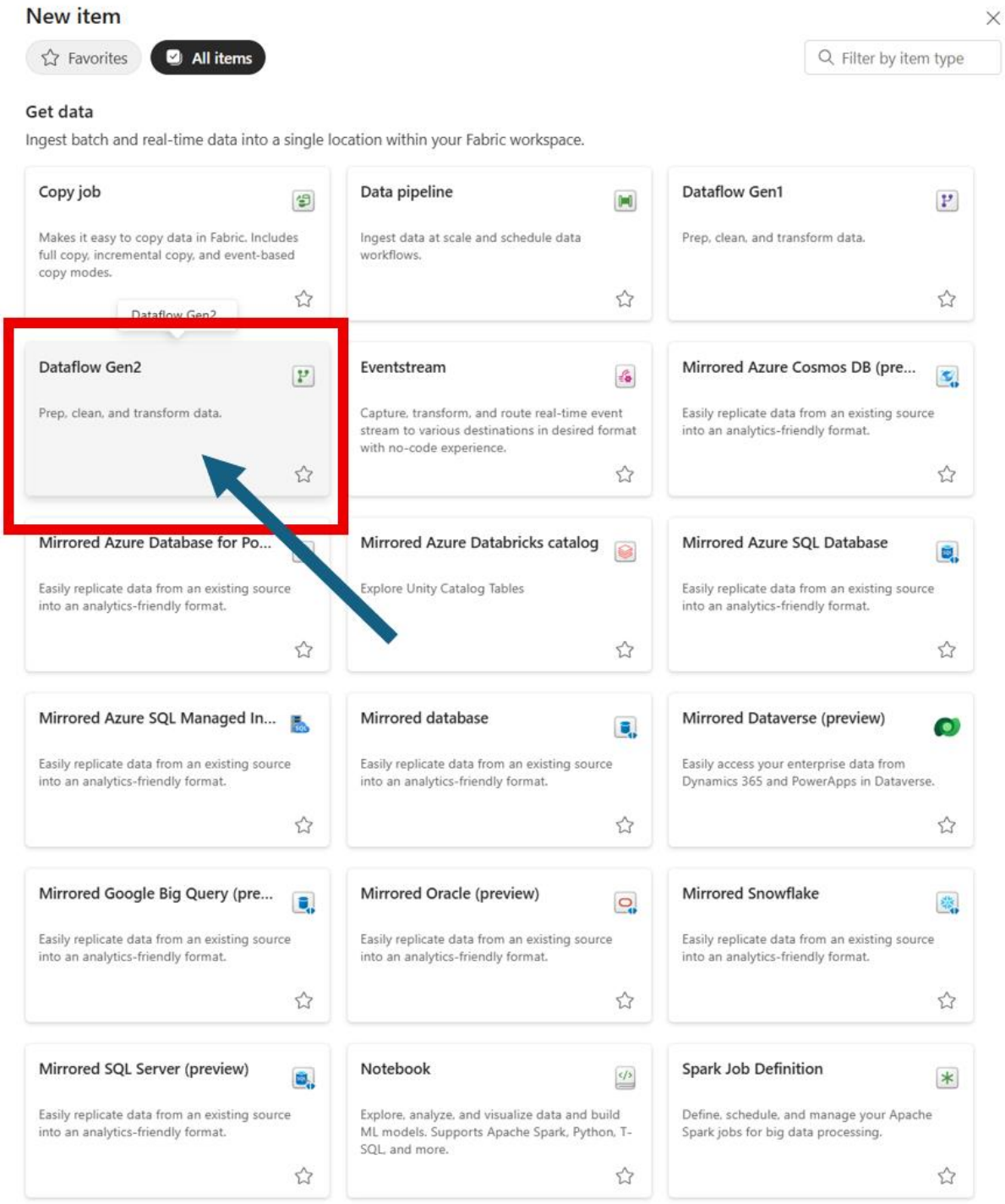
Task 2: Creating the dataflow gen2

This section will walk you through uploading a CSV file into a dataflow Gen2.

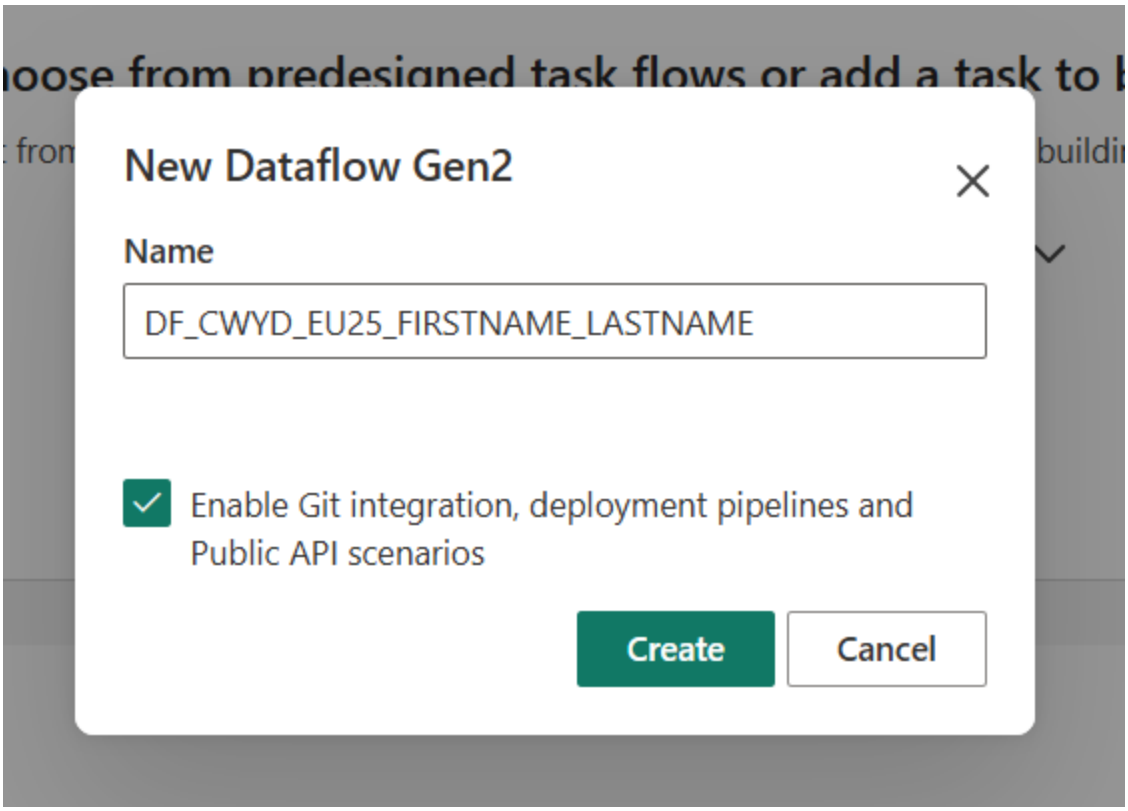
1. From within the workspace that you just created, select “New item” in the top left.



2. Select Dataflow Gen2
    a. Select the dataflow that has the description of Prep, clean, and transform data
    b. If you cannot scroll to find the item, search for dataflow in the search bar
    c. Select dataflow Gen2. You do not want to select Dataflow Gen1.
    d. It does not matter if you select the Dataflow Gen2 under the “Get data” or “Prepare data” headers.



3. Name the dataflow
    a. Name the dataflow as DF_CWYD_EU25_FIRSTNAME_LASTNAME
    b. For example I would name my dataflow DF_CWYD_EU25_LAUREN_FABER
    c. Name the dataflow

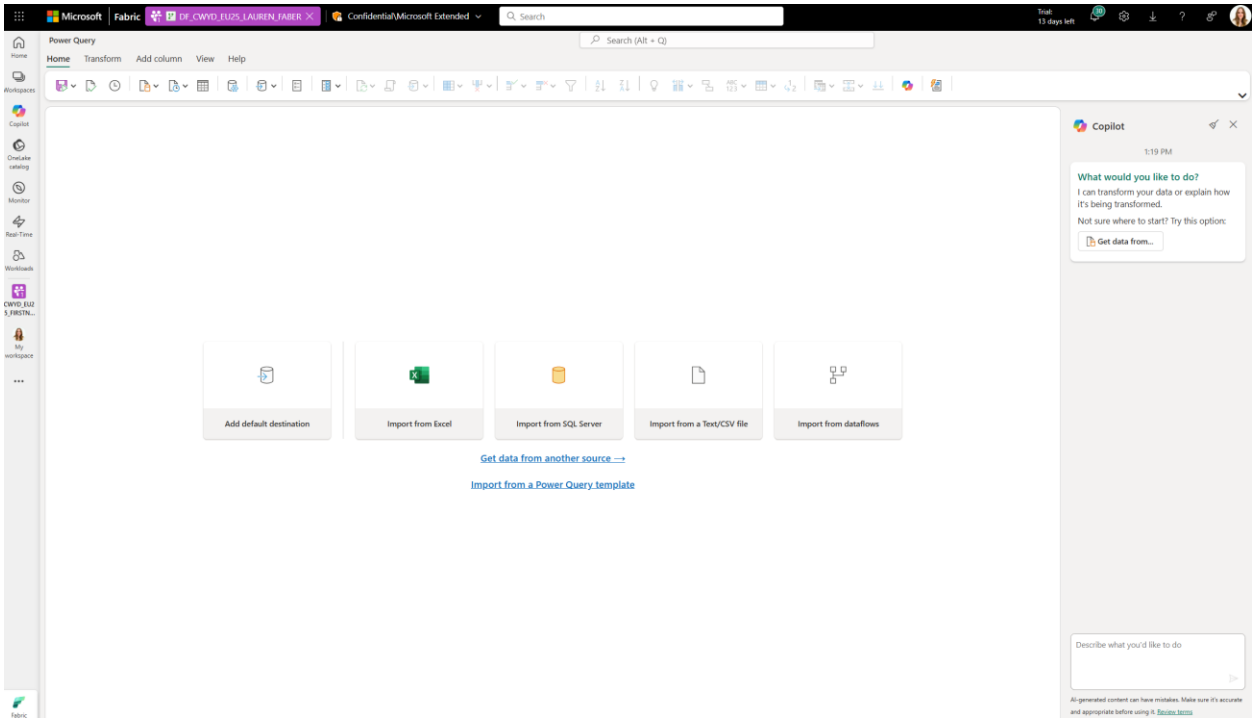


*You can leave the Enable Git integration box checked, or choose to uncheck it. For this workshop it does not matter.

4. Click Create

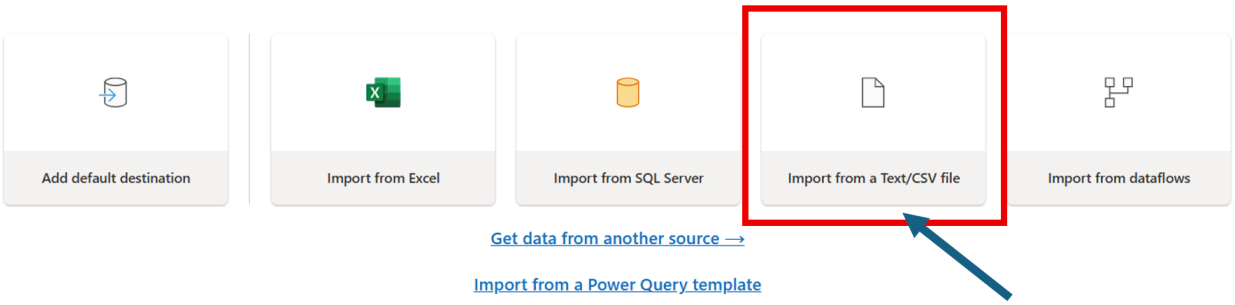
(If you run into the *Failed to create the dataflow* error or *failed to get the existing dataflow name* troubleshooting guidelines here.)

5. You should now see the dataflow created

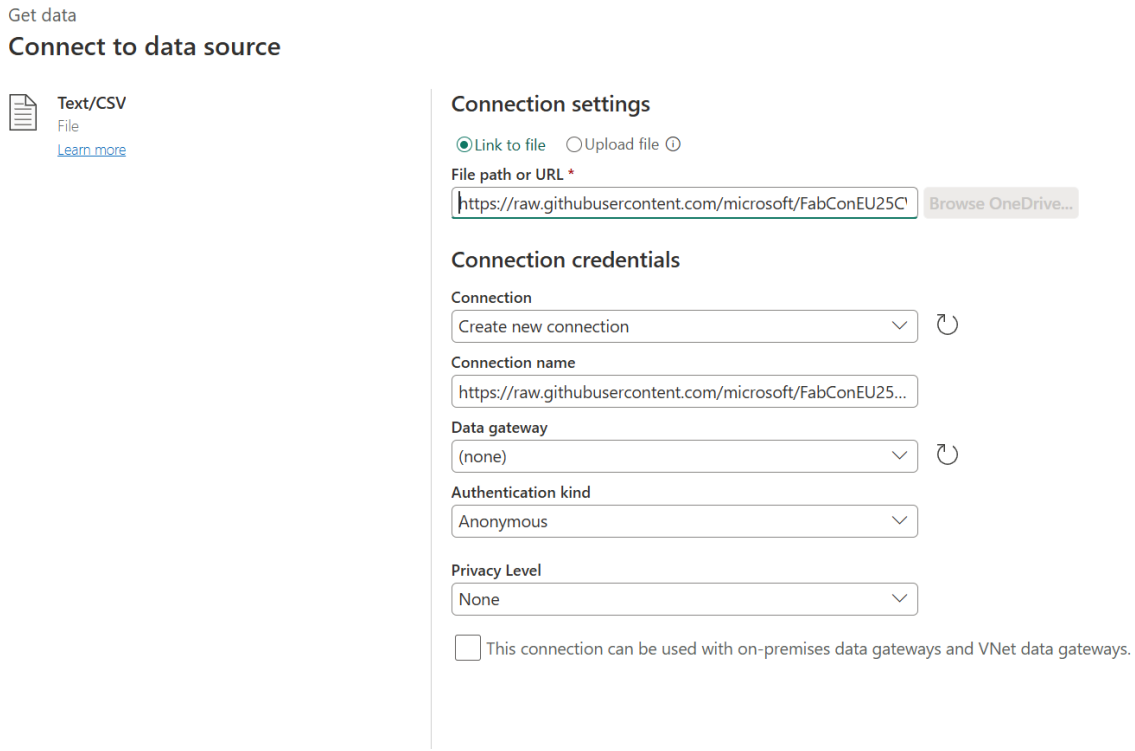


Task 3: Uploading data

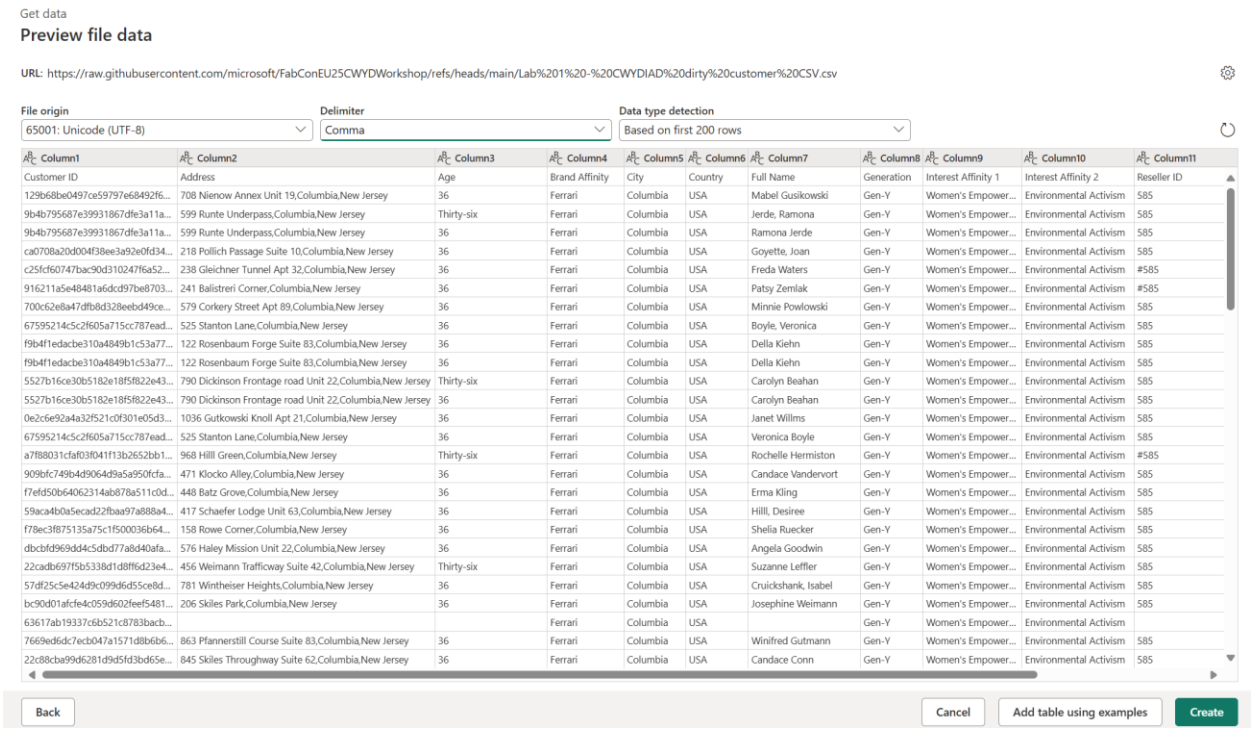
1. Select “Import from a Text/CSV file”



2. By default it will be set to “Link to file.” Paste the following link into the “File path or URL” input box: https://raw.githubusercontent.com/microsoft/FabConEU25CWYDWorkshop/refs/heads/main/Lab%201%20-%20CWYDIAD%20dirty%20customer%20CSV.csv



3. Click “Next”. You will see a screen like the following

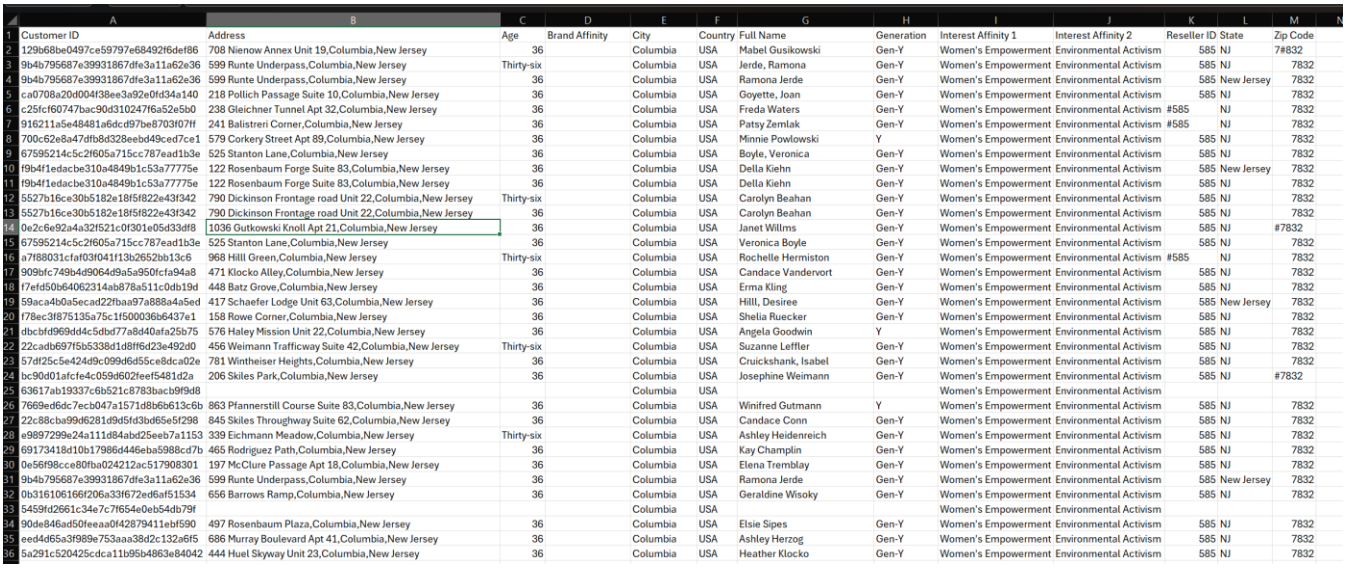


4. Click “Create”

You should now see all your data loaded into the dataflow!

Task 4: transforming data with Copilot

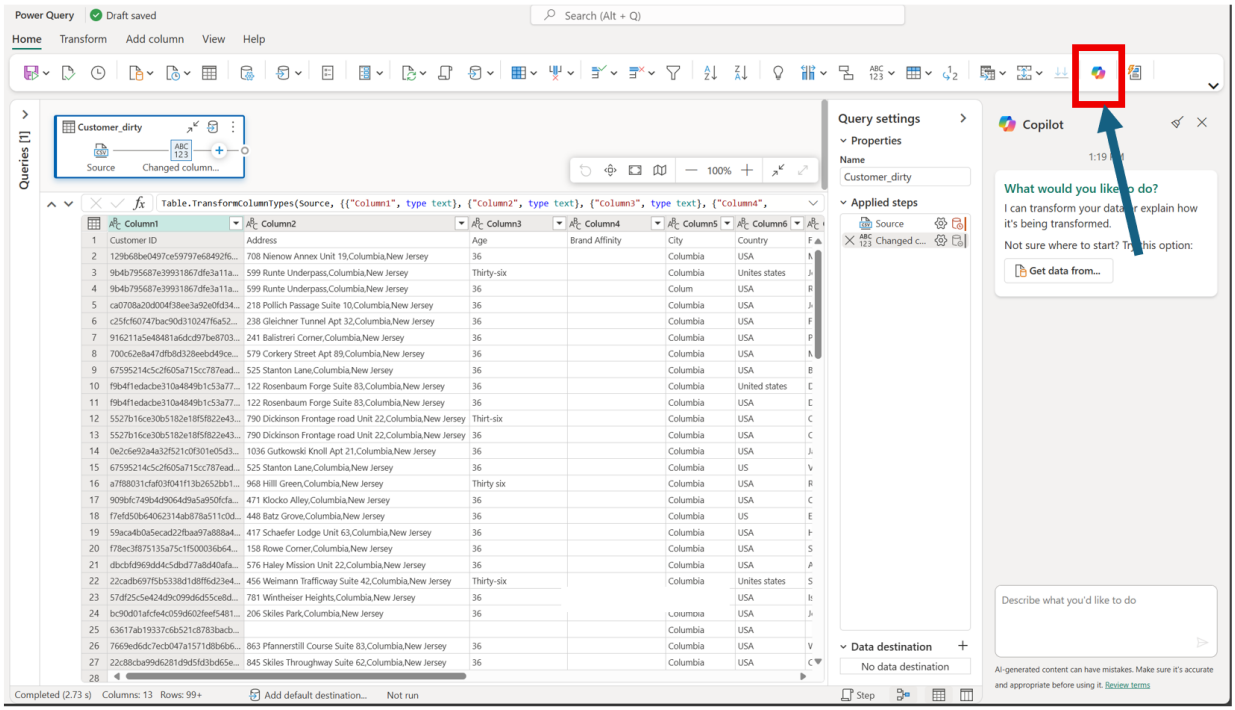
Now we will use Copilot to clean the messy data in our CSV. Before jumping into the rest of the lab see if you can identify problems with the data. There are at least 8 things that need to be fixed. How many can you find? (See answers at in index)



	A	B	C	D	E	F	G	H	I	J	K	L	M
1	Customer ID	Address	Age	Brand Affinity	City	Country	Full Name	Generation	Interest Affinity 1	Interest Affinity 2	Reseller ID	State	Zip Code
2	129b68be0497ce59797e68492f6def86	708 Nienow Annex Unit 19,Columbia,New Jersey	36		Columbia	USA	Mabel Gusikowski	Gen-Y	Women's Empowerment	Environmental Activism	585	NJ	7#832
3	9b4b795687e39931867dfe3a11a62e36	599 Runte Underpass,Columbia,New Jersey	Thirty-six		Columbia	USA	Jerde, Ramona	Gen-Y	Women's Empowerment	Environmental Activism	585	NJ	7832
4	9b4b795687e39931867dfe3a11a62e36	599 Runte Underpass,Columbia,New Jersey	36		Columbia	USA	Ramona Jerde	Gen-Y	Women's Empowerment	Environmental Activism	585	New Jersey	7832
5	ca0708a20d004f38ee3a92e0fd34a140	218 Pollich Passage Suite 10,Columbia,New Jersey	36		Columbia	USA	Goyette, Joan	Gen-Y	Women's Empowerment	Environmental Activism	585	NJ	7832
6	c25fcf60747bac90d310247f6a52e5b0	238 Gleichner Tunnel Apt 32,Columbia,New Jersey	36		Columbia	USA	Freda Waters	Gen-Y	Women's Empowerment	Environmental Activism	#585	NJ	7832
7	916211a5e48481a6dcd97be8703f07ff	241 Balistreri Corner,Columbia,New Jersey	36		Columbia	USA	Patsy Zemlak	Gen-Y	Women's Empowerment	Environmental Activism	#585	NJ	7832
8	700c62e8a47dfb8d328eebd49ced7ce1	579 Corkery Street Apt 89,Columbia,New Jersey	36		Columbia	USA	Minnie Powlowski	Y	Women's Empowerment	Environmental Activism	585	NJ	7832
9	67595214c5c2f605a715cc787ead1b3e	525 Stanton Lane,Columbia,New Jersey	36		Columbia	USA	Boyle, Veronica	Gen-Y	Women's Empowerment	Environmental Activism	585	NJ	7832
10	f9b4f1edacbe310a4849b1c53a77775e	122 Rosenbaum Forge Suite 83,Columbia,New Jersey	36		Columbia	USA	Della Kiehn	Gen-Y	Women's Empowerment	Environmental Activism	585	New Jersey	7832
11	f9b4f1edacbe310a4849b1c53a77775e	122 Rosenbaum Forge Suite 83,Columbia,New Jersey	36		Columbia	USA	Della Kiehn	Gen-Y	Women's Empowerment	Environmental Activism	585	NJ	7832
12	5527b16ce30b5182e18f5f822e43f342	790 Dickinson Frontage road Unit 22,Columbia,New Jersey	Thirty-six		Columbia	USA	Carolyn Beahan	Gen-Y	Women's Empowerment	Environmental Activism	585	NJ	7832
13	5527b16ce30b5182e18f5f822e43f342	790 Dickinson Frontage road Unit 22,Columbia,New Jersey	36		Columbia	USA	Carolyn Beahan	Gen-Y	Women's Empowerment	Environmental Activism	585	NJ	7832
14	0e2c6e92a4a32f521c0f301e05d33df8	1036 Gutkowski Knoll Apt 21,Columbia,New Jersey	36		Columbia	USA	Janet Willms	Gen-Y	Women's Empowerment	Environmental Activism	585	NJ	#7832
15	67595214c5c2f605a715cc787ead1b3e	525 Stanton Lane,Columbia,New Jersey	36		Columbia	USA	Veronica Boyle	Gen-Y	Women's Empowerment	Environmental Activism	585	NJ	7832
16	a7f88031cfaf03f041f13b2652bb13c6	968 Hilll Green,Columbia,New Jersey	Thirty-six		Columbia	USA	Rochelle Hermiston	Gen-Y	Women's Empowerment	Environmental Activism	#585	NJ	7832
17	909bfc749b4d9064d9a5a950fcfa94a8	471 Klocko Alley,Columbia,New Jersey	36		Columbia	USA	Candace Vandervort	Gen-Y	Women's Empowerment	Environmental Activism	585	NJ	7832
18	f7efd50b64062314ab878a511c0db19d	448 Batz Grove,Columbia,New Jersey	36		Columbia	USA	Erma Kling	Gen-Y	Women's Empowerment	Environmental Activism	585	NJ	7832
19	59aca4b0a5ecad22fbaa97a888a4a5ed	417 Schaefer Lodge Unit 63,Columbia,New Jersey	36		Columbia	USA	Hilll, Desiree	Gen-Y	Women's Empowerment	Environmental Activism	585	New Jersey	7832
20	f78ec3f875135a75c1f500036b6437e1	158 Rowe Corner,Columbia,New Jersey	36		Columbia	USA	Shelia Ruecker	Gen-Y	Women's Empowerment	Environmental Activism	585	NJ	7832
21	dbcbfd969dd4c5dbd77a8d40afa25b75	576 Haley Mission Unit 22,Columbia,New Jersey	36		Columbia	USA	Angela Goodwin	Y	Women's Empowerment	Environmental Activism	585	NJ	7832
22	22cadb697f5b5338d1d8ff6d23e492d0	456 Weimann Trafficway Suite 42,Columbia,New Jersey	Thirty-six		Columbia	USA	Suzanne Leffler	Gen-Y	Women's Empowerment	Environmental Activism	585	NJ	7832
23	57df25c5e424d9c099d6d55ce8dca02e	781 Wintheiser Heights,Columbia,New Jersey	36		Columbia	USA	Cruickshank, Isabel	Gen-Y	Women's Empowerment	Environmental Activism	585	NJ	7832
24	bc90d01afcfe4c059d602feef5481d2a	206 Skiles Park,Columbia,New Jersey	36		Columbia	USA	Josephine Weimann	Gen-Y	Women's Empowerment	Environmental Activism	585	NJ	#7832
25	63617ab19337c6b521c8783bacb9f9d8				Columbia	USA			Women's Empowerment	Environmental Activism			
26	7669ed6dc7ecb047a1571d8b6b613c6b	863 Pfannerstill Course Suite 83,Columbia,New Jersey	36		Columbia	USA	Winifred Gutmann	Y	Women's Empowerment	Environmental Activism	585	NJ	7832
27	22c88cba99d6281d9d5fd3bd65e5f298	845 Skiles Throughway Suite 62,Columbia,New Jersey	36		Columbia	USA	Candace Conn	Gen-Y	Women's Empowerment	Environmental Activism	585	NJ	7832
28	e9897299e24a111d84abd25eeb7a1153	339 Eichmann Meadow,Columbia,New Jersey	Thirty-six		Columbia	USA	Ashley Heidenreich	Gen-Y	Women's Empowerment	Environmental Activism	585	NJ	7832
29	69173418d10b17986d446eba5988cd7b	465 Rodriguez Path,Columbia,New Jersey	36		Columbia	USA	Kay Champlin	Gen-Y	Women's Empowerment	Environmental Activism	585	NJ	7832
30	0e56f98cce80fba024212ac517908301	197 McClure Passage Apt 18,Columbia,New Jersey	36		Columbia	USA	Elena Tremblay	Gen-Y	Women's Empowerment	Environmental Activism	585	NJ	7832
31	9b4b795687e39931867dfe3a11a62e36	599 Runte Underpass,Columbia,New Jersey	36		Columbia	USA	Ramona Jerde	Gen-Y	Women's Empowerment	Environmental Activism	585	New Jersey	7832
32	0b316106166f206a33f672ed6af51534	656 Barrows Ramp,Columbia,New Jersey	36		Columbia	USA	Geraldine Wisoky	Gen-Y	Women's Empowerment	Environmental Activism	585	NJ	7832
33	5459fd2661c34e7c7f654e0eb54db79f				Columbia	USA			Women's Empowerment	Environmental Activism			
34	90de846ad50feeaa0f42879411ebf590	497 Rosenbaum Plaza,Columbia,New Jersey	36		Columbia	USA	Elsie Sipes	Gen-Y	Women's Empowerment	Environmental Activism	585	NJ	7832
35	eed4d65a3f989e753aaa38d2c132a6f5	686 Murray Boulevard Apt 41,Columbia,New Jersey	36		Columbia	USA	Ashley Herzog	Gen-Y	Women's Empowerment	Environmental Activism	585	NJ	7832
36	5a291c520425cdca11b95b4863e84042	444 Huel Skyway Unit 23,Columbia,New Jersey	36		Columbia	USA	Heather Klocko	Gen-Y	Women's Empowerment	Environmental Activism	585	NJ	7832

**Troubleshooting tip. If Copilot starts to get overwhelmed, save your progress, exit the dataflow, re-enter, and try again.

1. Open the Copilot pane



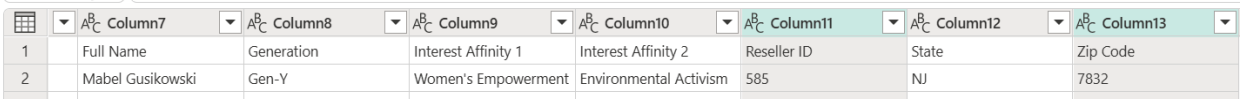
2. Now you'll work with Copilot to clean and transform the data. For each section use the phrases to ask Copilot to perform an action for you.

Step 1: Format table

First, your data needs some clean headers. Let's move the first row to be headers and use our company's naming format.

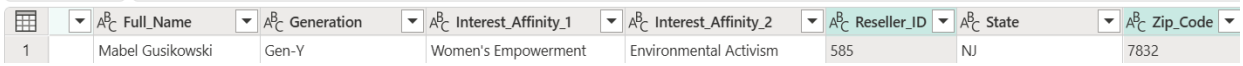
a. "Update first row as headers and add a "_" between spaces in the headers"

Before



	Column7	Column8	Column9	Column10	Column11	Column12	Column13
1	Full Name	Generation	Interest Affinity 1	Interest Affinity 2	Reseller ID	State	Zip Code
2	Mabel Gusikowski	Gen-Y	Women's Empowerment	Environmental Activism	585	NJ	7832

After



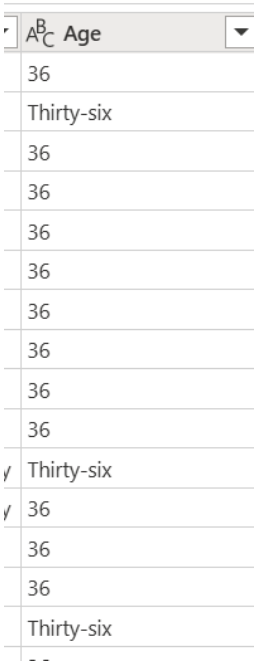
	Full_Name	Generation	Interest_Affinity_1	Interest_Affinity_2	Reseller_ID	State	Zip_Code
1	Mabel Gusikowski	Gen-Y	Women's Empowerment	Environmental Activism	585	NJ	7832

Step 2: Clean data

Next, let's focus on cleaning the data. We can see words where there should be numbers, columns without uniformity, and random symbols.

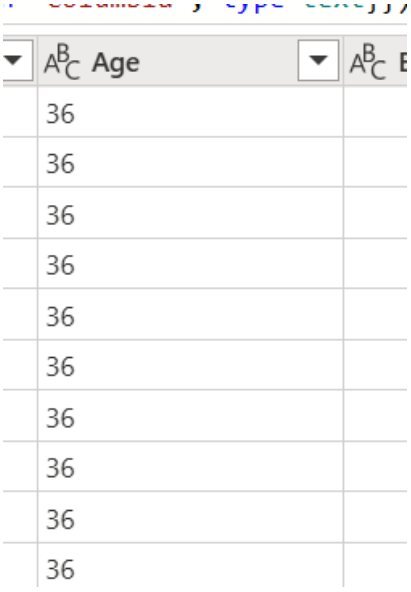
b. "In the Age column update anywhere an age is a word to be represented by a number"

Before



Age
36
Thirty-six
36
36
36
36
36
36
36
36
Thirty-six
36
36
36
Thirty-six

After



Age	
36	
36	
36	
36	
36	
36	
36	
36	
36	
36	

c. "In Full_Name switch order of names if there is a comma, then remove the comma"

Before

Full_Name
Mabel Gusikowski
Jerde, Ramona
Ramona Jerde
Joan Goyette
Freda Waters
Patsy Zemlak
Minnie Powlowski
Boyle, Veronica
Della Kiehn
Della Kiehn
Carolyn Beahan
Carolyn Beahan
Janet Willms
Veronica Boyle
Rochelle Hermiston
Candace Vandervort
Erma Kling
Hilll, Desiree
Shelia Ruecker
Angela Goodwin

After

Full_Name
Mabel Gusikowski
Ramona Jerde
Ramona Jerde
Joan Goyette
Freda Waters
Patsy Zemlak
Minnie Powlowski
Veronica Boyle
Della Kiehn
Della Kiehn
Carolyn Beahan
Carolyn Beahan
Janet Willms
Veronica Boyle
Rochelle Hermiston
Candace Vandervort
Erma Kling
Desiree Hilll
Shelia Ruecker

d. “Remove “#” anywhere in the data”

Before

After

Reseller_ID	
585	N
585	N
585	N
585	N
585	N
#585	N
585	N
585	N
585	N
585	N
585	N
585	N
585	N
585	N
#585	N
585	N

Reseller_ID	
585	N
585	N
585	N
585	N
585	N
585	N
585	N
585	N
585	N
585	N
585	N
585	N
585	N
585	N
585	N
585	N
585	N
585	N
585	N
585	N
585	N

Step 3: remove redundant data

Now the data looks uniform, but we still have some problems. We have empty of null values, and rows that are duplicates. Let's fix that!

e. "Remove rows with null or empty values"

Before

	Address	Age	Brar
a...	576 Haley Mission Unit 22,Columbia,New Jersey	36	
4...	456 Weimann Trafficway Suite 42,Columbia,New Jersey	36	
d...	781 Wintheiser Heights,Columbia,New Jersey	36	
1...	206 Skiles Park,Columbia,New Jersey	36	
b...			
6...	863 Pfannerstill Course Suite 83,Columbia,New Jersey	36	
e...	845 Skiles Throughway Suite 62,Columbia,New Jersey	36	
7...	339 Eichmann Meadow,Columbia,New Jersey	36	
8...	465 Rodriguez Path,Columbia,New Jersey	36	
0...	197 McClure Passage Apt 18,Columbia,New Jersey	36	
a...	599 Runte Underpass,Columbia,New Jersey	36	
5...	656 Barrows Ramp,Columbia,New Jersey	36	
d...			
b...	497 Rosenbaum Plaza,Columbia,New Jersey	36	

After

11	cecb7ef824f42e921a8c4a62bbac...	282 Krajcik Green Apt 27,Miami,Florida	18	Ferrari	Miami	
12	787b5a23a2b12f1838eaf76ad6c0...	109 Davis Heights Unit 93,Miami,Florida	18	Ferrari	Miami	
13	3a6c0c78f953a18bb9add670641...	900 Smitham Quay Unit 16,Miami,Florida	18	Ferrari	Miami	
14	7f9a53dd7db05197572131dc431...	517 Metz Bay Suite 94,Miami,Florida	23	Ferrari	Miami	
15	f420b7d3e35cb3cf639c69542610...	1079 Erdman Lock Unit 81,Miami,Florida	23	Ferrari	Miami	
16	511ad80d9b20e81e644a78a0daa...	667 Schowalter Skyway Suite 26,Miami,Florida	23	Ferrari	Miami	
17	66dacbbe43dae888b1f4fce505f0...	481 Nolan Rest Suite 50,Miami,Florida	23	Ferrari	Miami	
18	dd4a02296060f4786ebf8e39f0e0...	150 Durgan Station Apt 5,Miami,Florida	22	Ferrari	Miami	
19	341ef7ef735ce882da2da53973c1...	841 Kuvalis Throughway,Miami,Florida	22	Ferrari	Miami	
20	4b7ef26a042a3ef4b4488c22d03c...	1058 Lueilwitz Bay,Miami,Florida	22	Ferrari	Miami	
21	47628210fe3b9bc5216ef64d4009...	961 Schultz Gate Unit 57,Miami,Florida	27	Ferrari	Miami	
22	0bc121a4f35a3fb5c4b09b099027...	516 Heller Landing,Miami,Florida	27	Ferrari	Miami	
23	18ae8fa85845d12bb807178c452...	201 Zulauf Stravenue,Miami,Florida	27	Ferrari	Miami	
24	183e0e9a0cfa59175583cb9156b6...	897 Watsica Vista Unit 23,Miami,Florida	27	Ferrari	Miami	

f. “Remove duplicate rows”

Before

2	9b4b795687e39931867dfe3a11a...	599 Runte Underpass,Columbia,New Jersey	36		Columbia	USA
3	9b4b795687e39931867dfe3a11a...	599 Runte Underpass,Columbia,New Jersey	36		Colum	USA

After

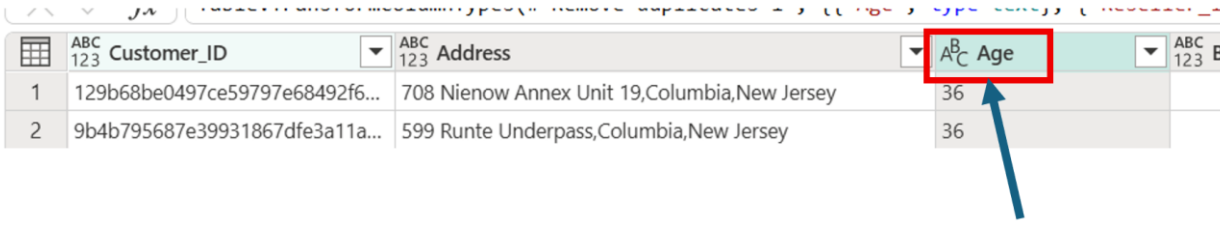
1	129b68be0497ce59797e68492f6...	708 Nienow Annex Unit 19,Columbia,New Jersey	36	
2	9b4b795687e39931867dfe3a11a...	599 Runte Underpass,Columbia,New Jersey	36	
3	ca0708a20d004f38ee3a92e0fd34...	218 Pollich Passage Suite 10,Columbia,New Jersey	36	
4	c25fcf60747bac90d310247f6a52...	238 Gleichner Tunnel Apt 32,Columbia,New Jersey	36	

Step 4: Update to correct data types

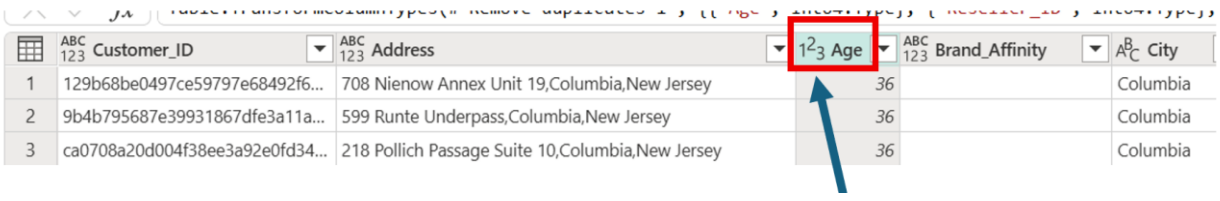
Lastly, we want to ensure our data is ready to be put into a semantic model, let’s update all the field types to be the right type.

g. “Update field type to whole number instead of text where applicable”

Before

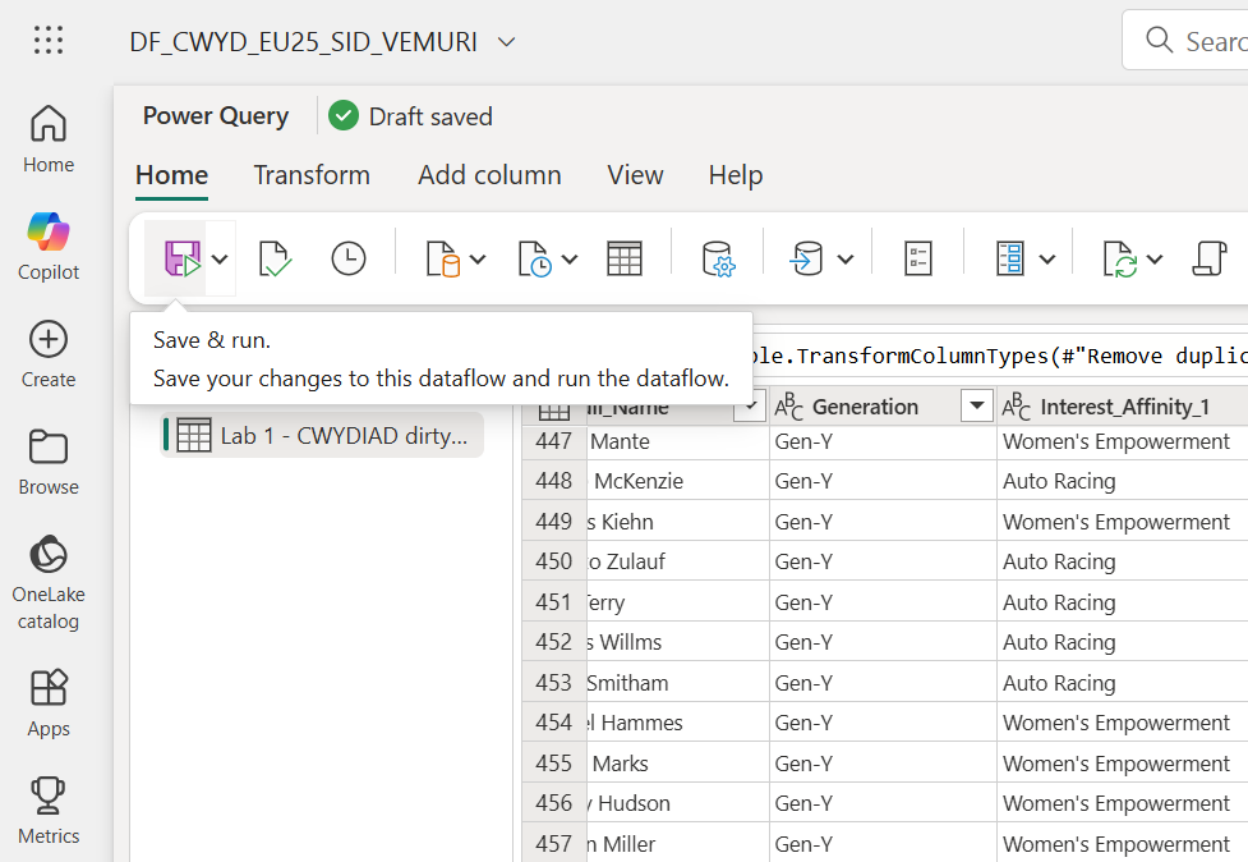


After



You did it! You successfully used Copilot to clean your data!

Now hit the save and run button in the top left corner



Appendix

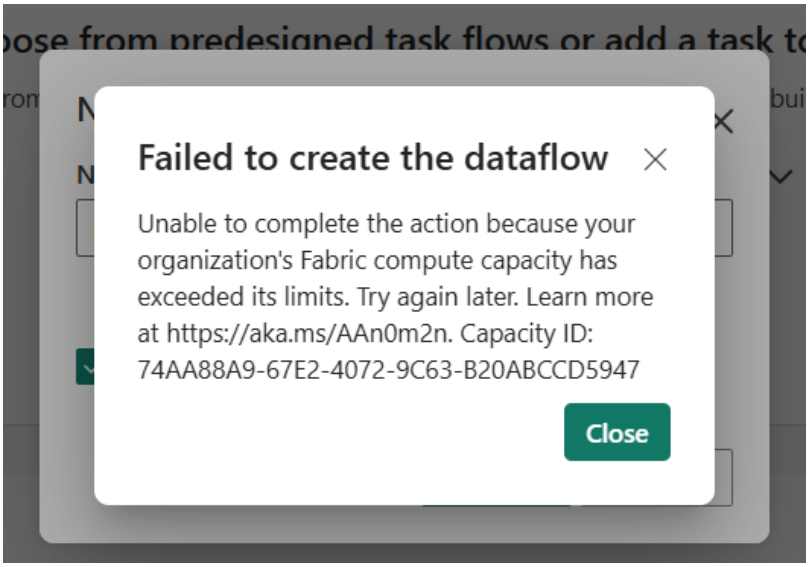
Answers for what need to be cleaned/transformed in the data

1. First row needs to be header
2. Age column has a mix of word and numbers
3. Generation is inconsistent
4. Reseller ID and Zip Code has random “#”s
5. State is inconsistent
6. Full Name has inconsistent format (some are First name Last name and others are Last Name, First name)
7. Null values
8. Duplicate items

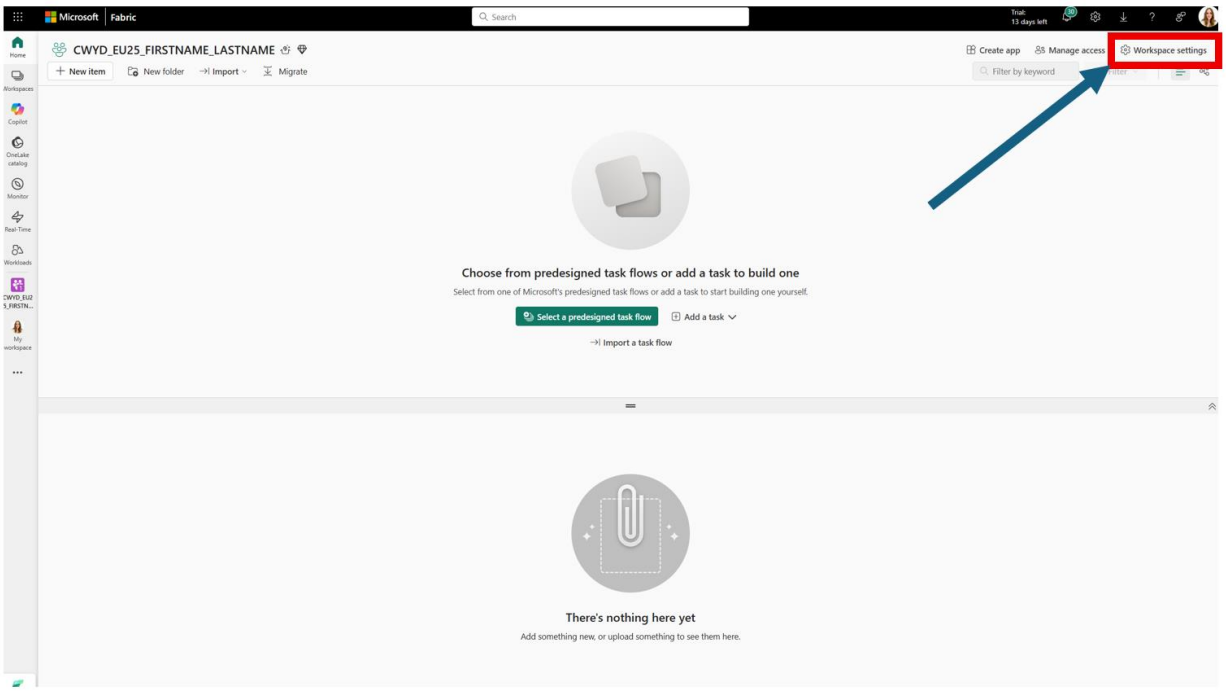
Troubleshooting

Failed to create the dataflow

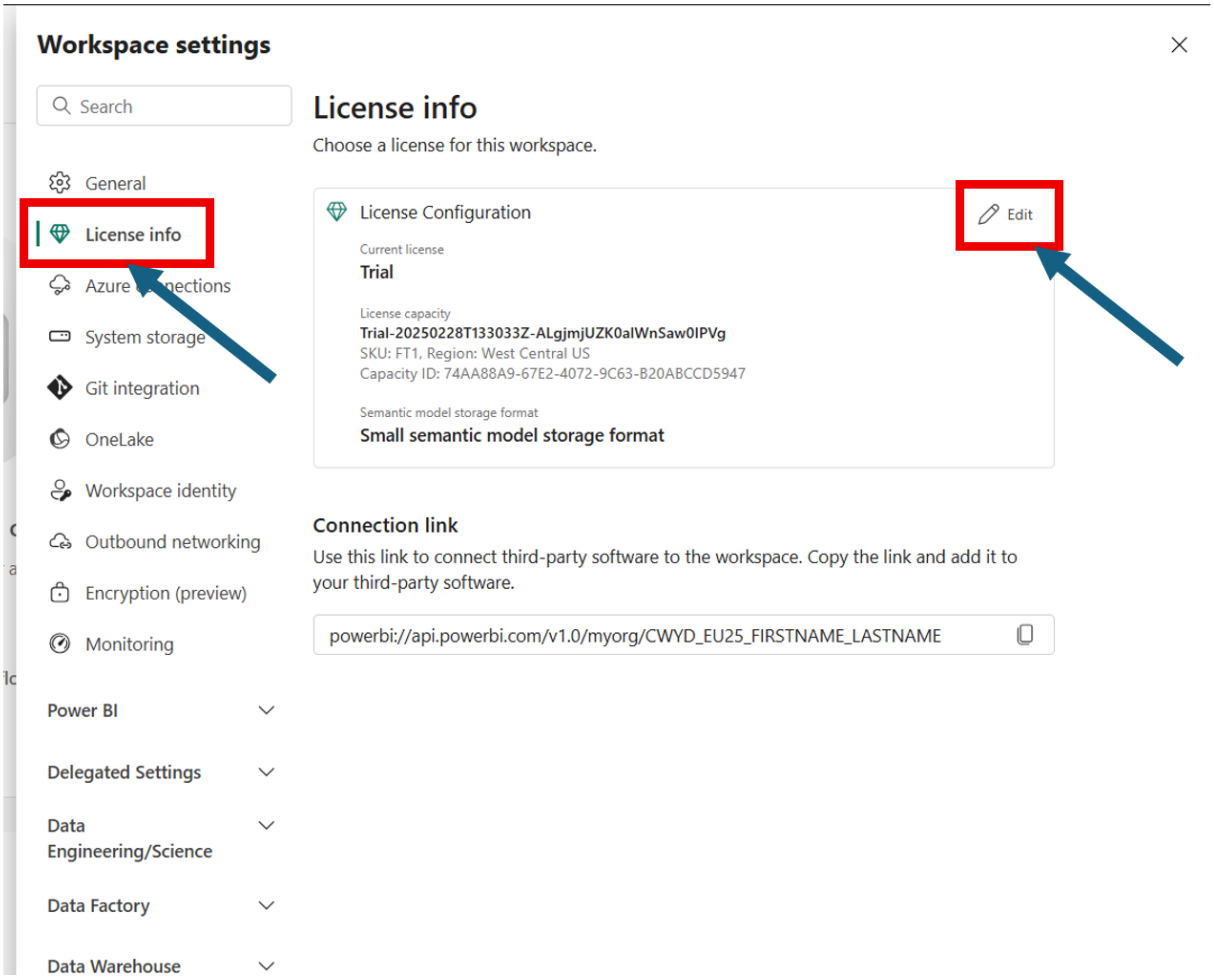
- If you run into the error failed to create the dataflow, follow these steps



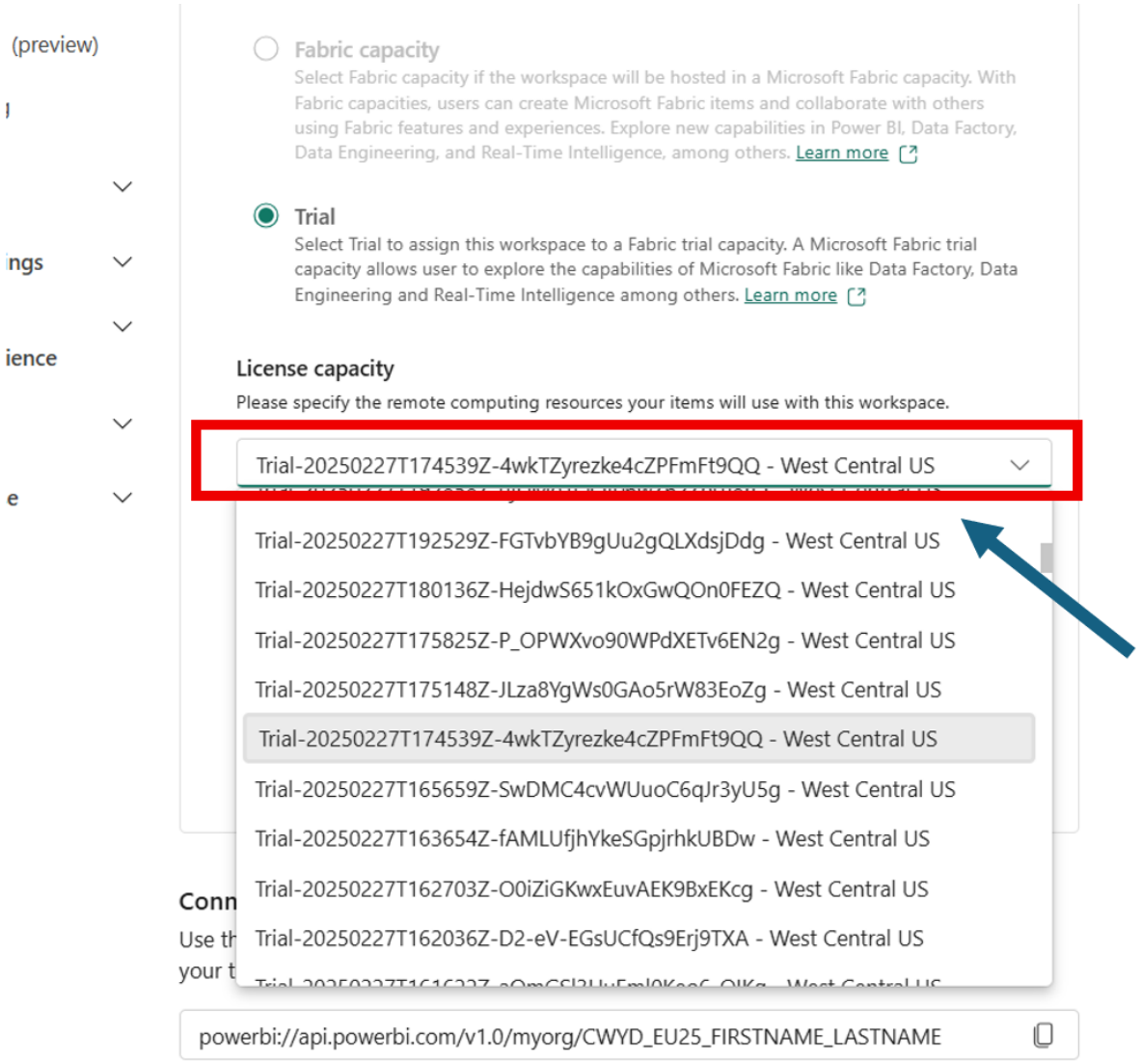
- Click on the workspace settings in the top right of the workspace



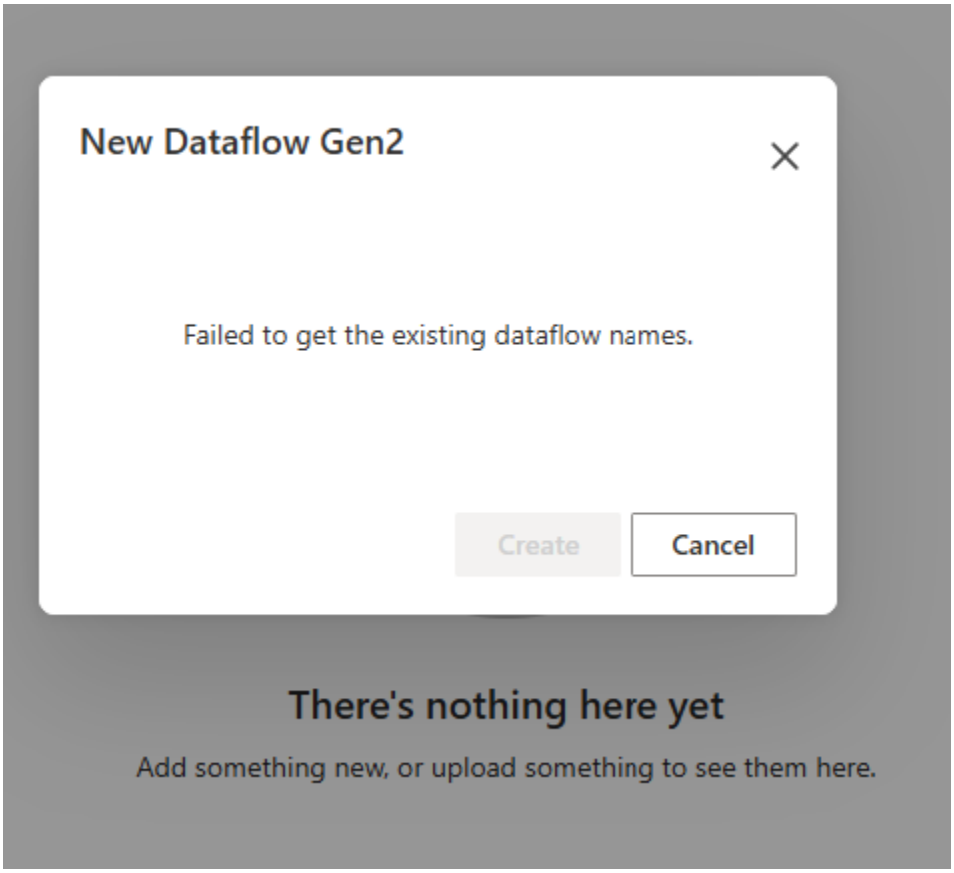
- Select “License info” and click the Edit icon



- Under license capacity, select the dropdown. Choose any new capacity and try again. If you still get the error try a few more capacities. If the error persists please raise your hand.



- Click “Select license”
- Try creating a dataflow Gen2 again

Failed to get the existing dataflow names

If you run into a “failed to get the existing dataflow names” error. Try creating the dataflow in a different workspace (like MyWorkspace)