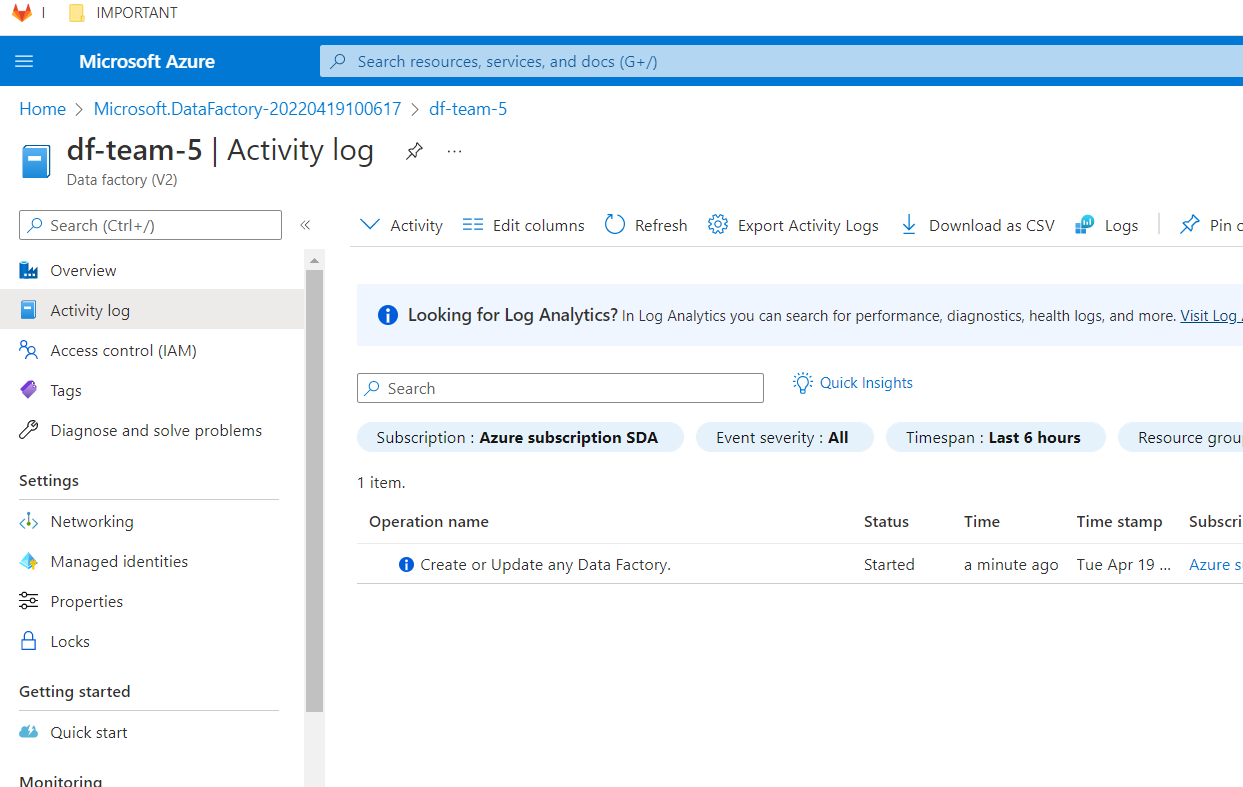
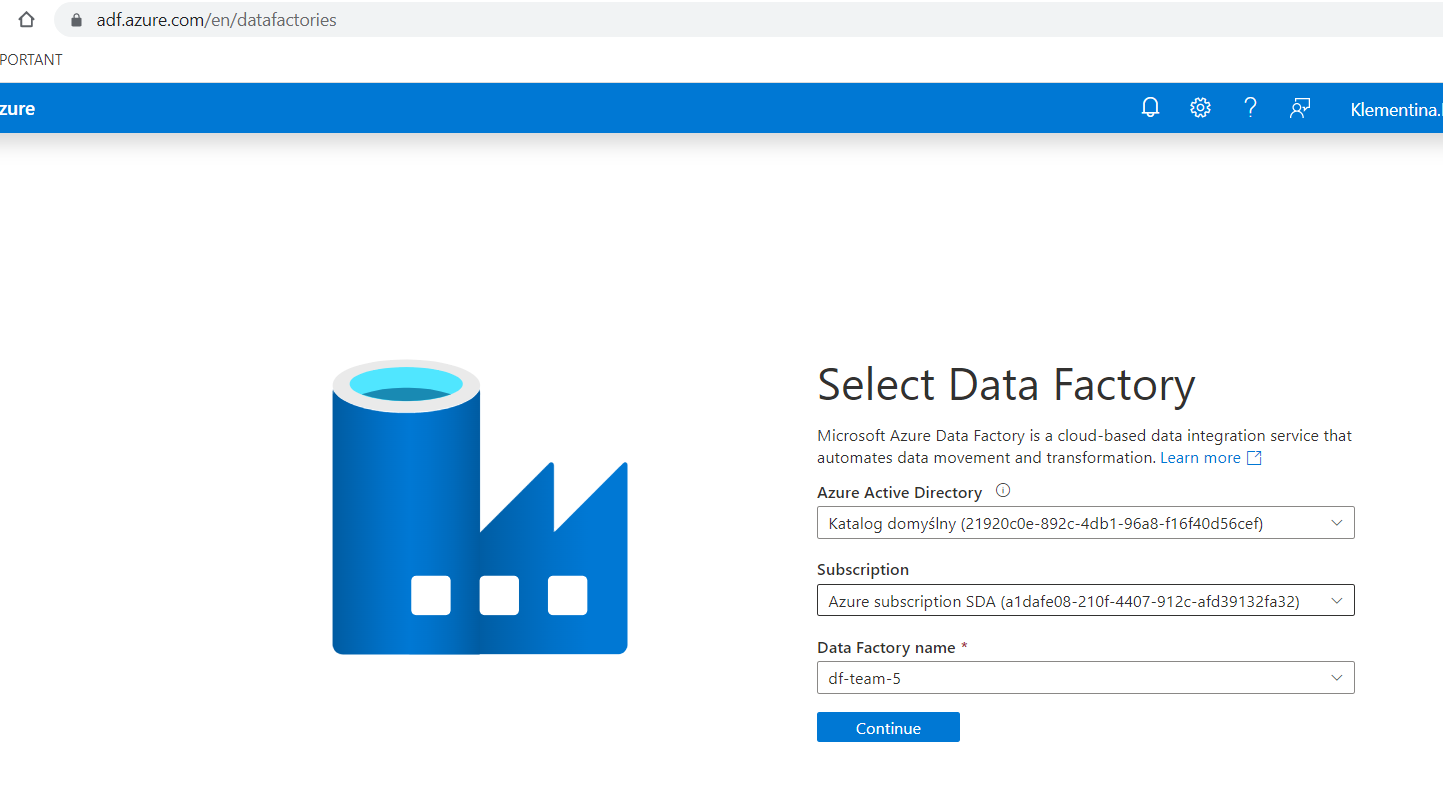
**Team 5**

**Solution:**

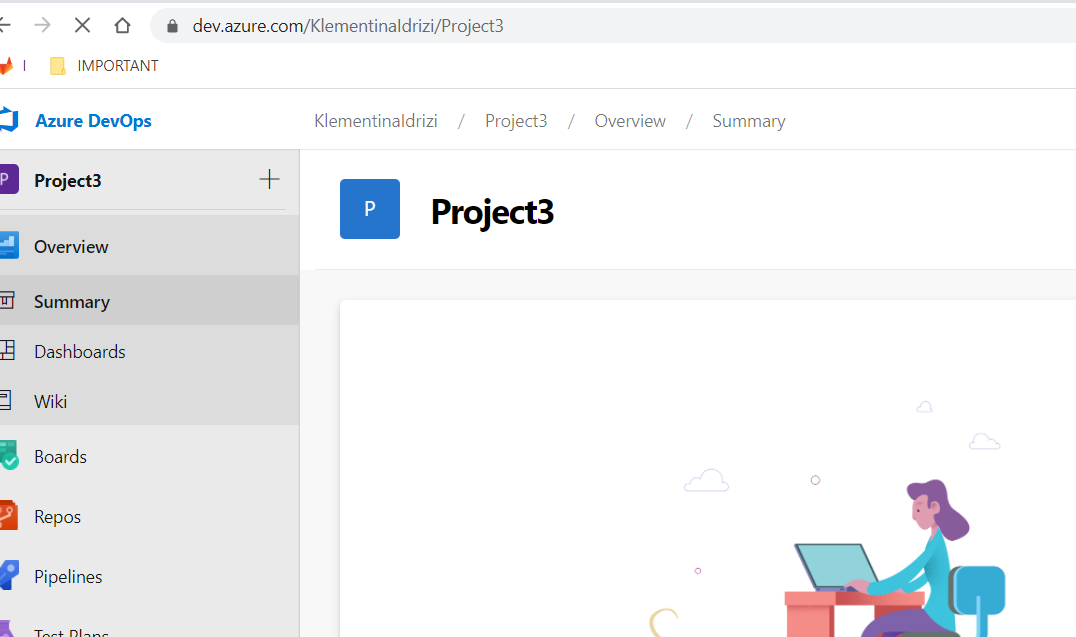
Create a Data Factory:



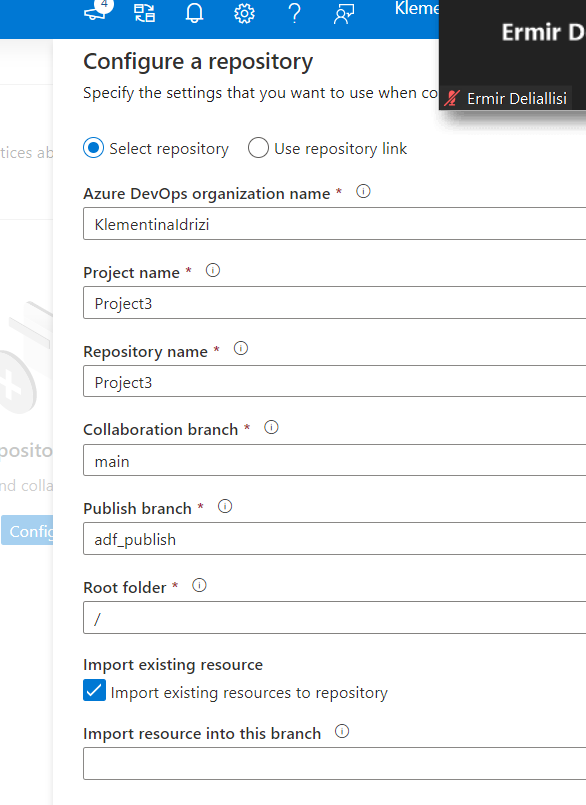
Connect the df into portal



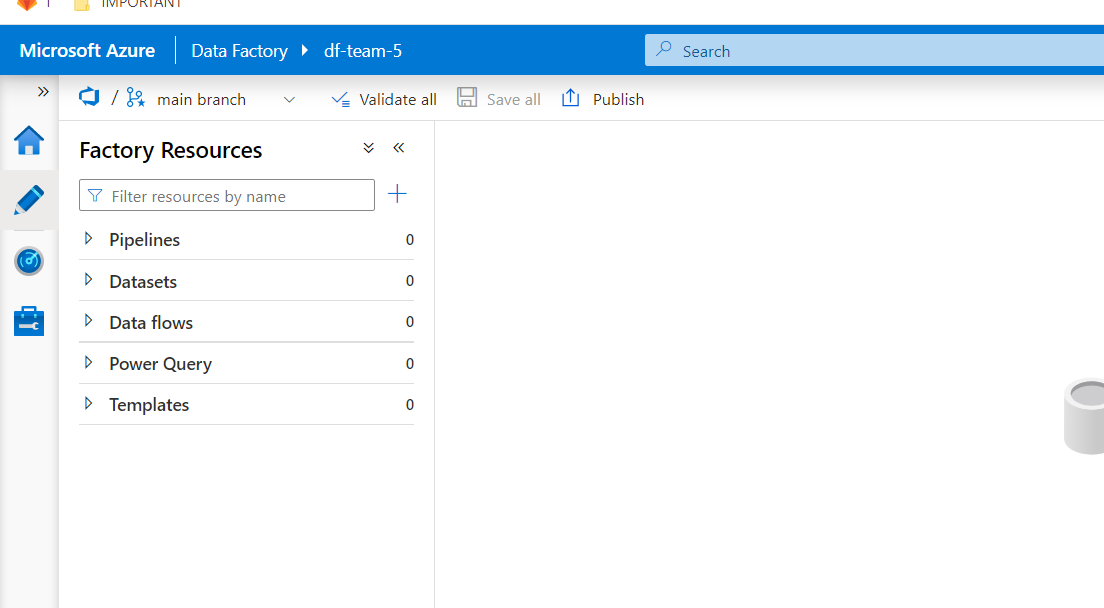
Create a new project in dev.azure.com



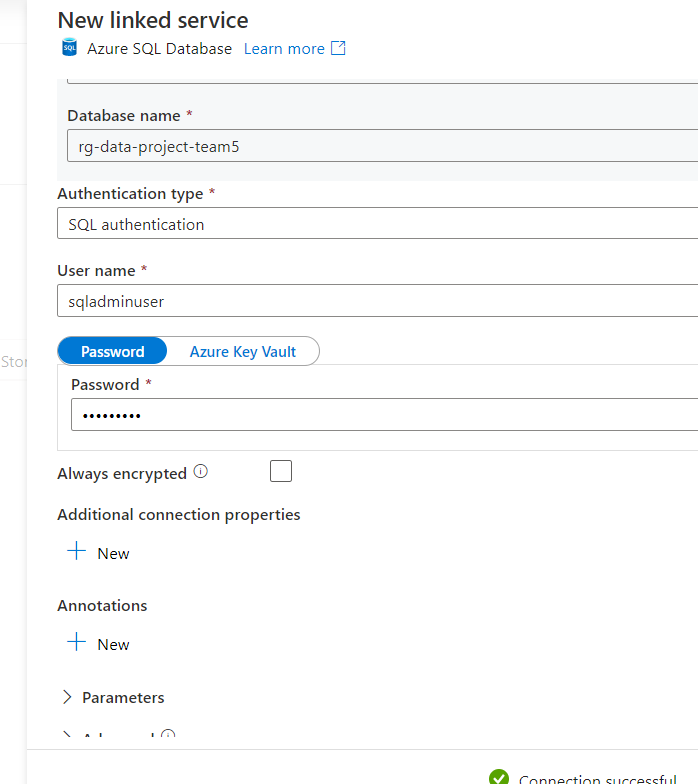
Configure git

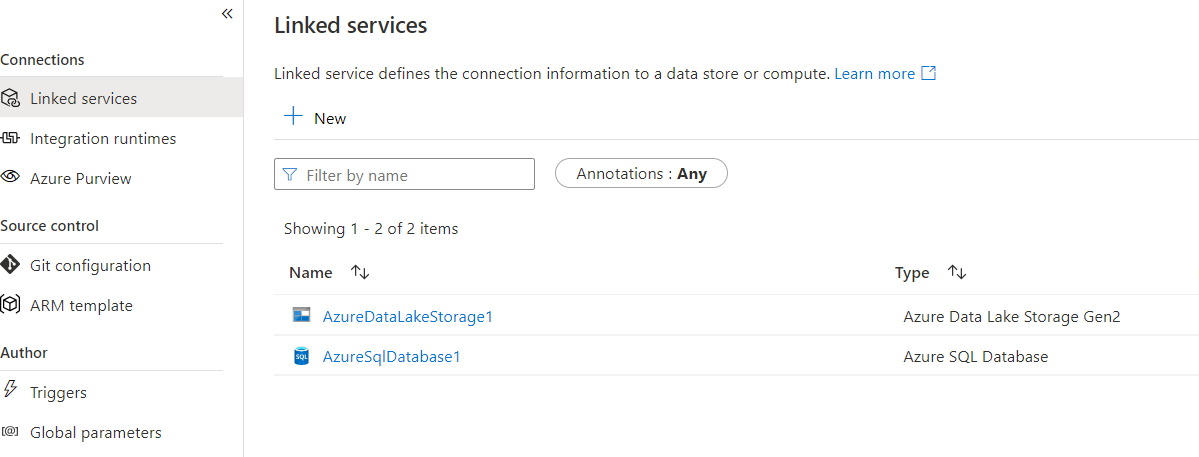


Create new linked services:

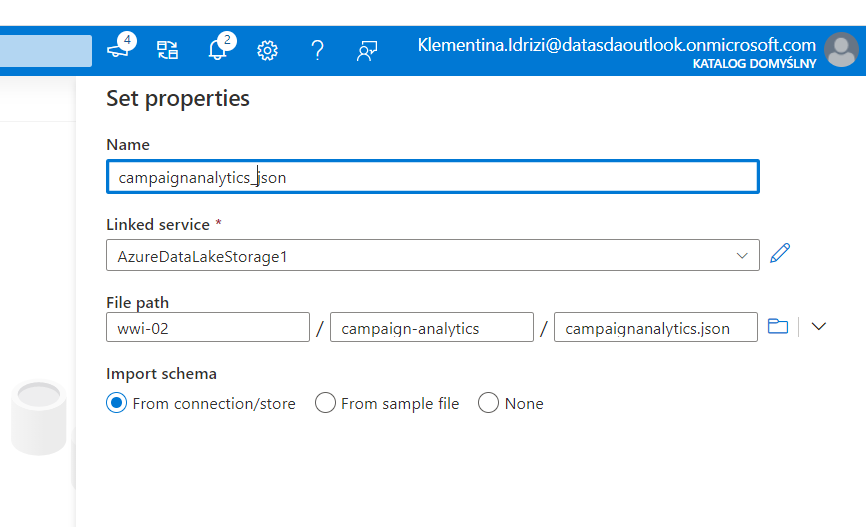


Create new linked service, sql server

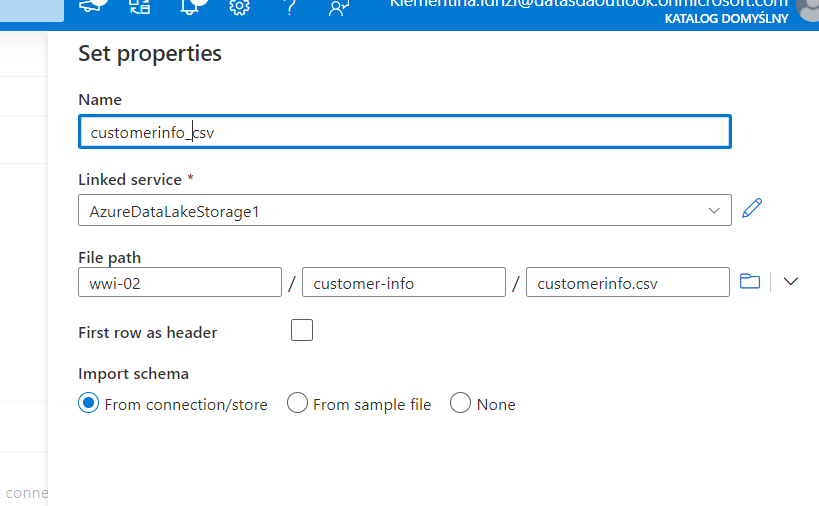




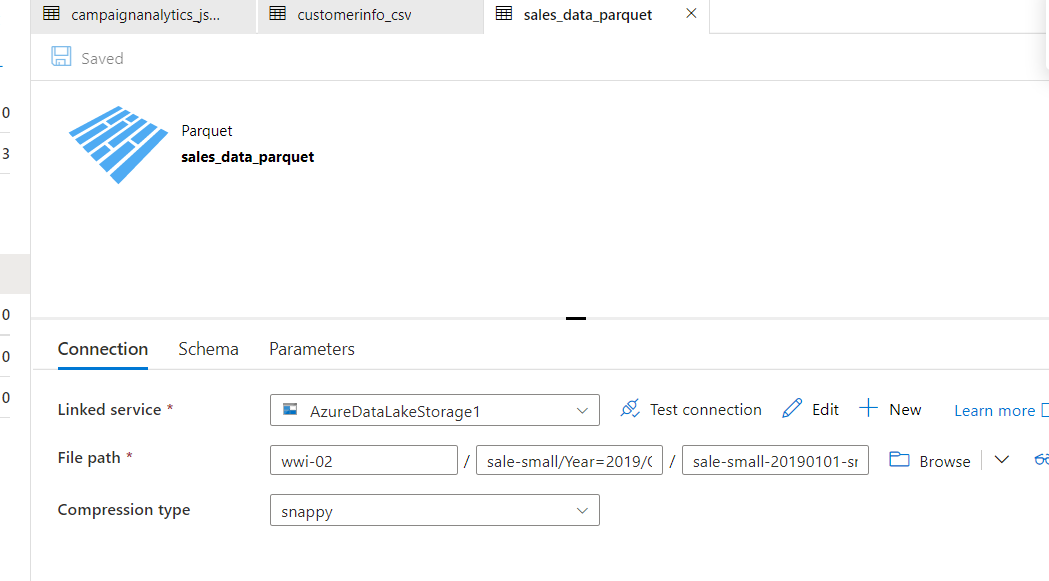
Create a dataset JSON



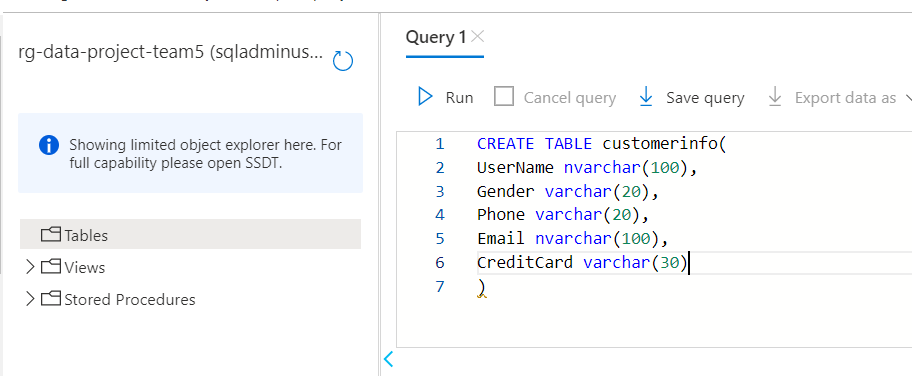
Create a new data set csv:

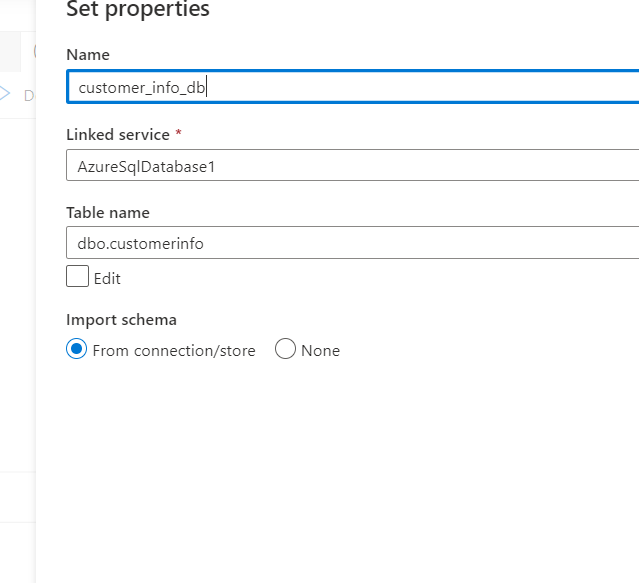


Create a new data set parquet:



Create a new sql azure data set



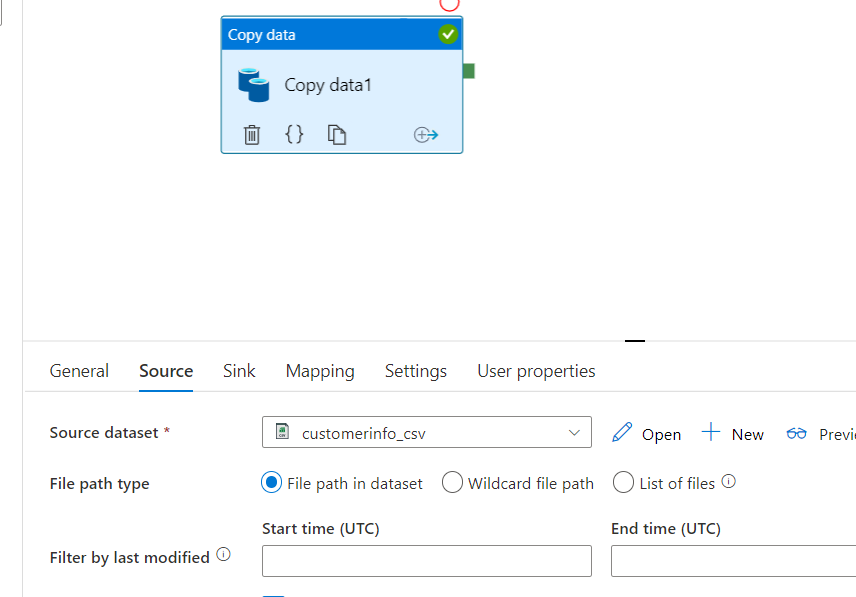


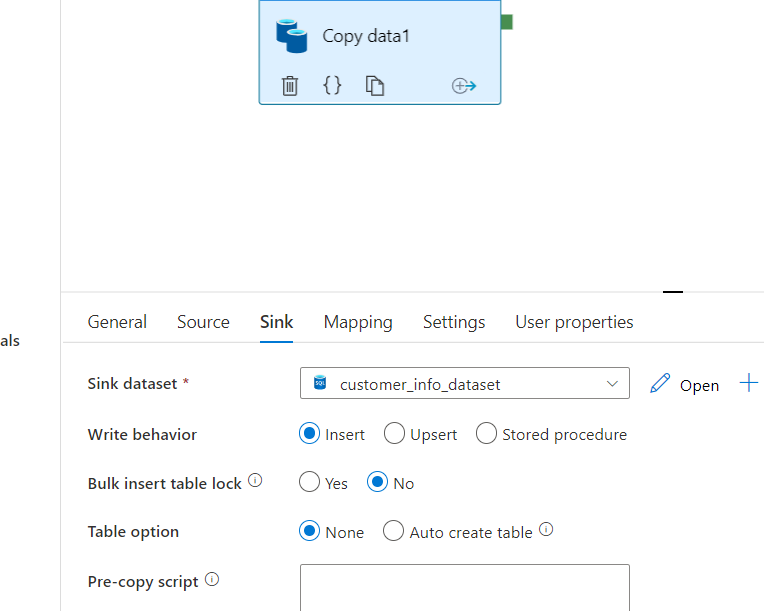
Create a new pipeline customerinfo

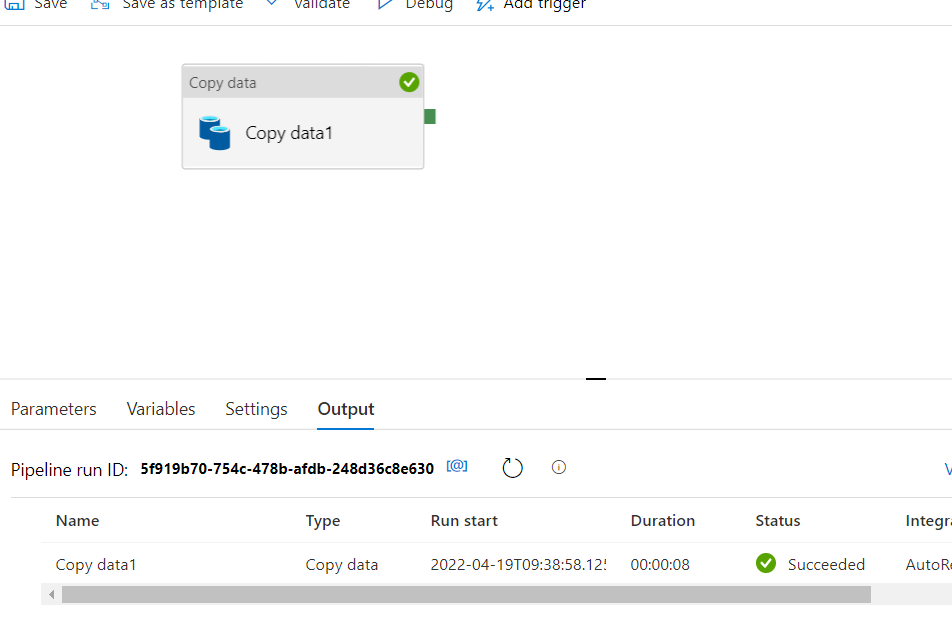
Create table

Create db

Connect with the pipeline





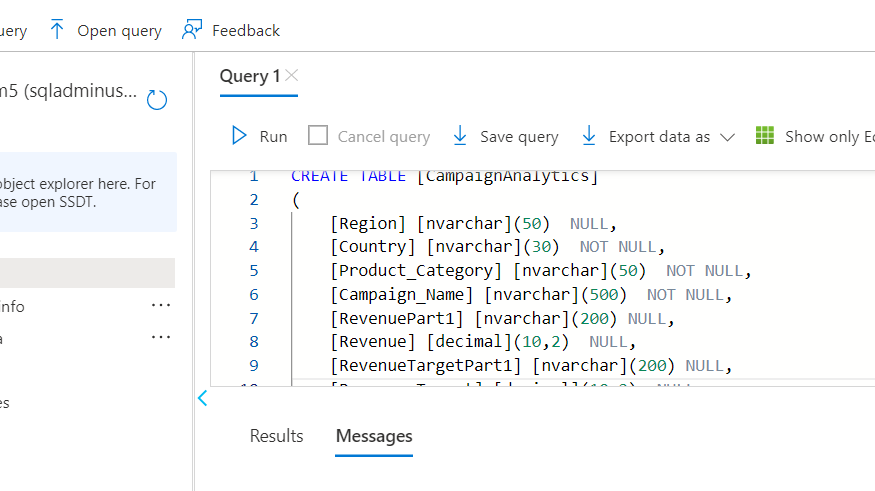


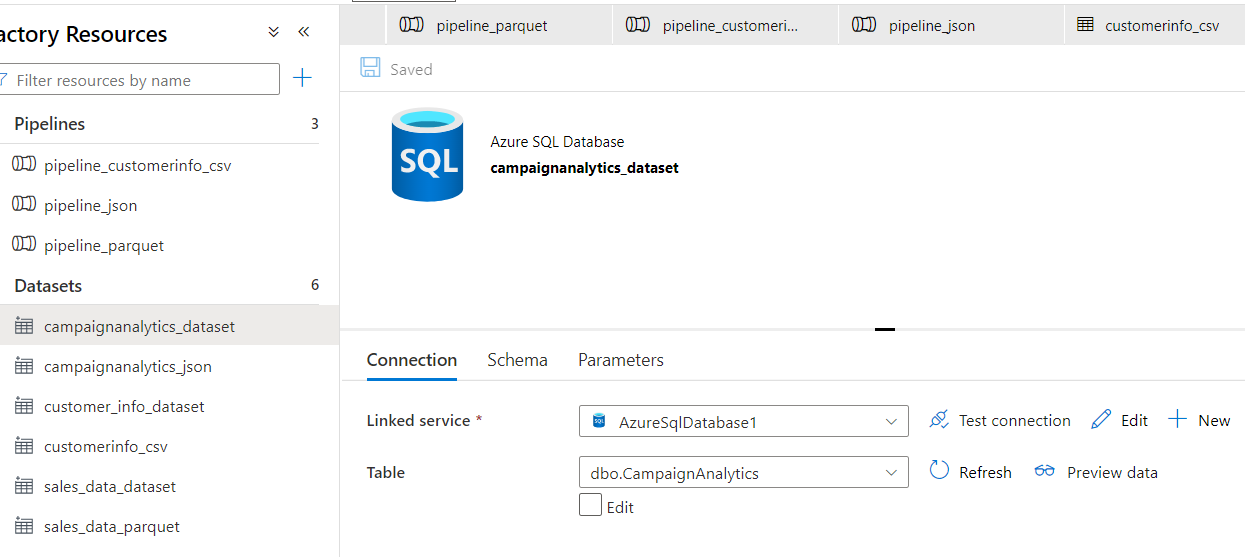
Create a new pipeline for json

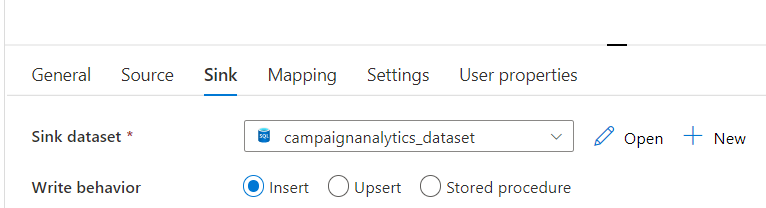
Create table

Create db

Connect with the pipeline

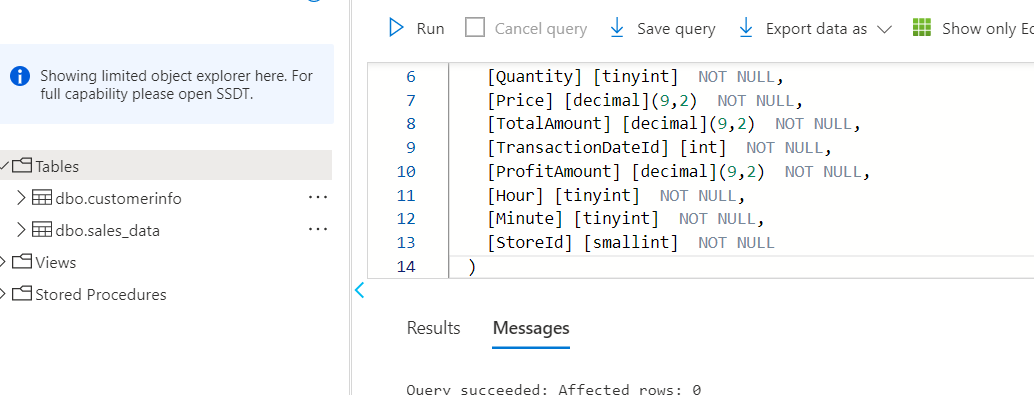




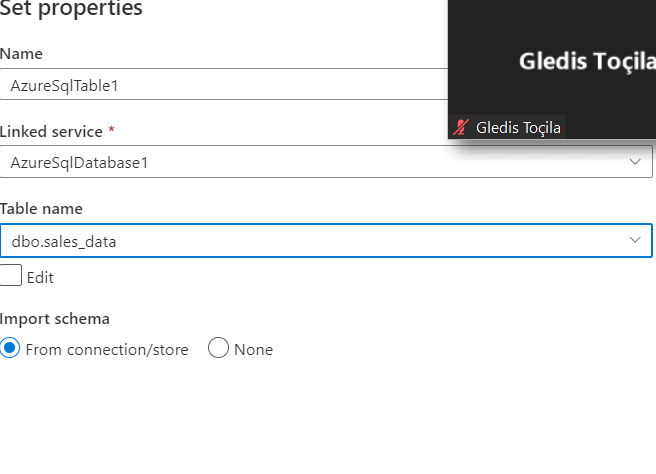


Create a new pipeline for parquet:

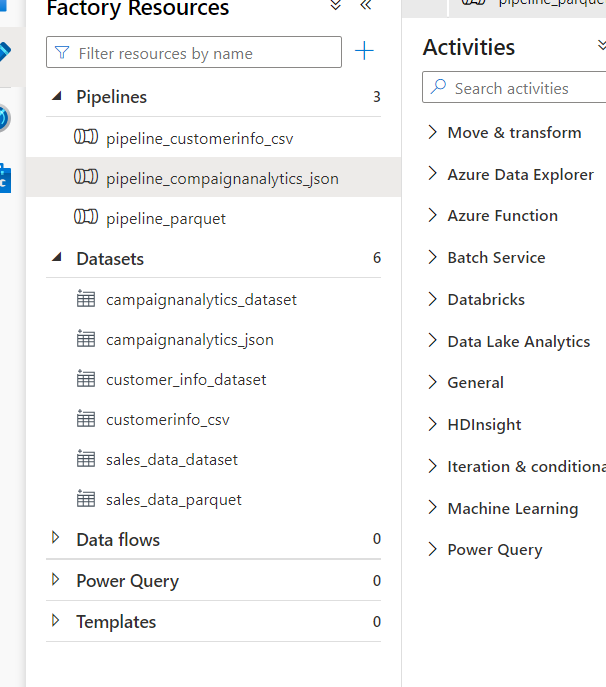
Create the table



Create db with the table we created in parquet:



Now we have the view:



Use the script to clean the tables

insert into [dbo].[CampaignAnalytics] (

[Region]

,[Country]

,[Product\_Category]

,[Campaign\_Name]

,[Revenue]

,[Revenue\_Target]

,[City]

,[State]

)

SELECT [Region]

,[Country]

,[Product\_Category]

,[Campaign\_Name]

,concat(SUBSTRING([RevenuePart1], 2, len([RevenuePart1])-2), Revenue) Revenue

,concat(SUBSTRING([RevenueTargetPart1], 2, len([RevenueTargetPart1])-2), Revenue\_Target) Revenue\_Target

,[City]

,[State]

FROM [dbo].[CampaignAnalytics\_ST]

Create a stored procedure

Move data in the final table

CREATE PROCEDURE [dbo].[move\_data\_to\_CampaignAnalytics]

AS

BEGIN

truncate table [dbo].[CampaignAnalytics];

insert into [dbo].[CampaignAnalytics] (

[Region]

,[Country]

,[Product\_Category]

,[Campaign\_Name]

,[Revenue]

,[Revenue\_Target]

,[City]

,[State]

)

SELECT [Region]

,[Country]

,[Product\_Category]

,[Campaign\_Name]

,concat(SUBSTRING([RevenuePart1], 2, len([RevenuePart1])-2), Revenue) Revenue

,concat(SUBSTRING([RevenueTargetPart1], 2, len([RevenueTargetPart1])-2), Revenue\_Target) Revenue\_Target

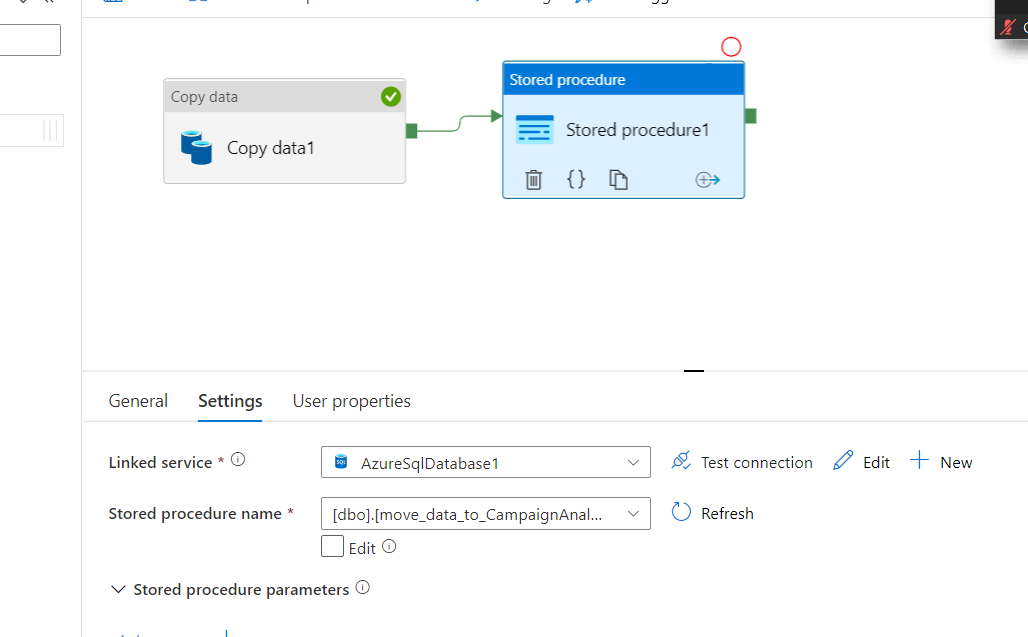
,[City]

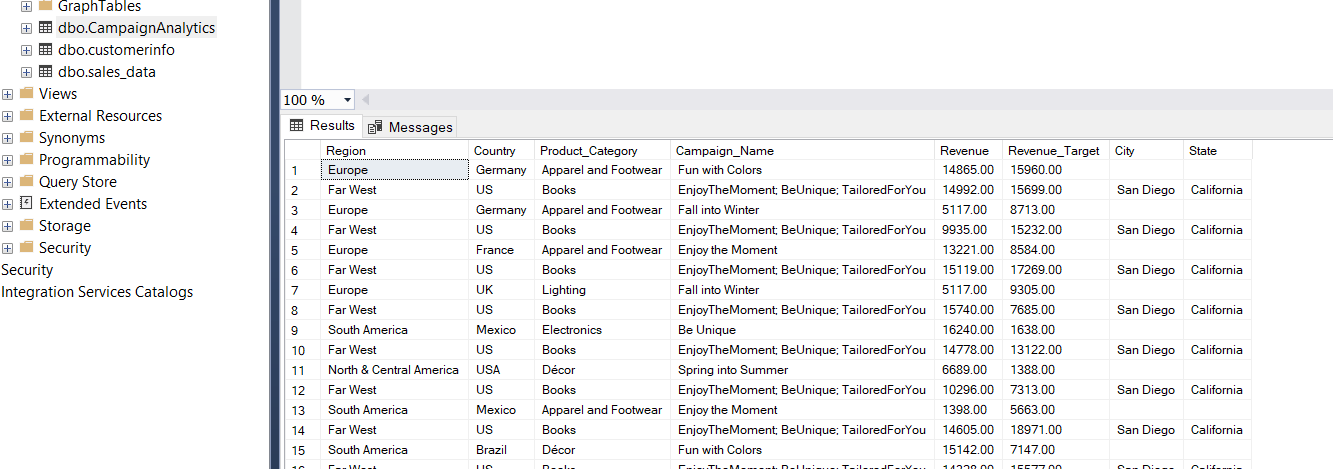
,[State]

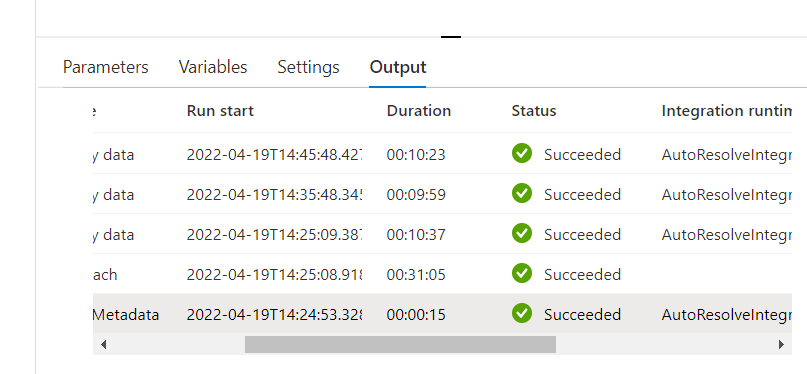
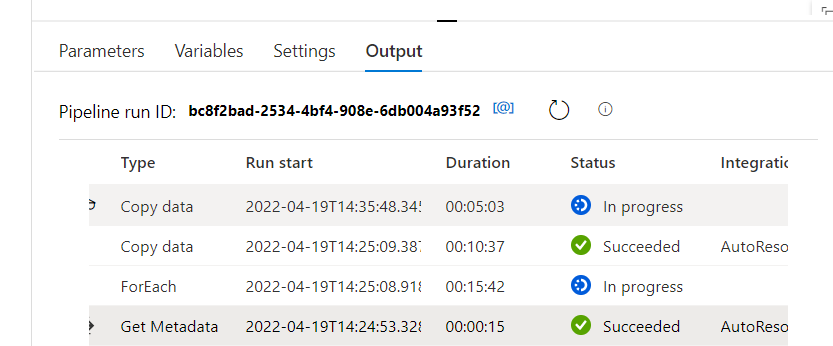
FROM [dbo].[CampaignAnalytics\_ST]

END

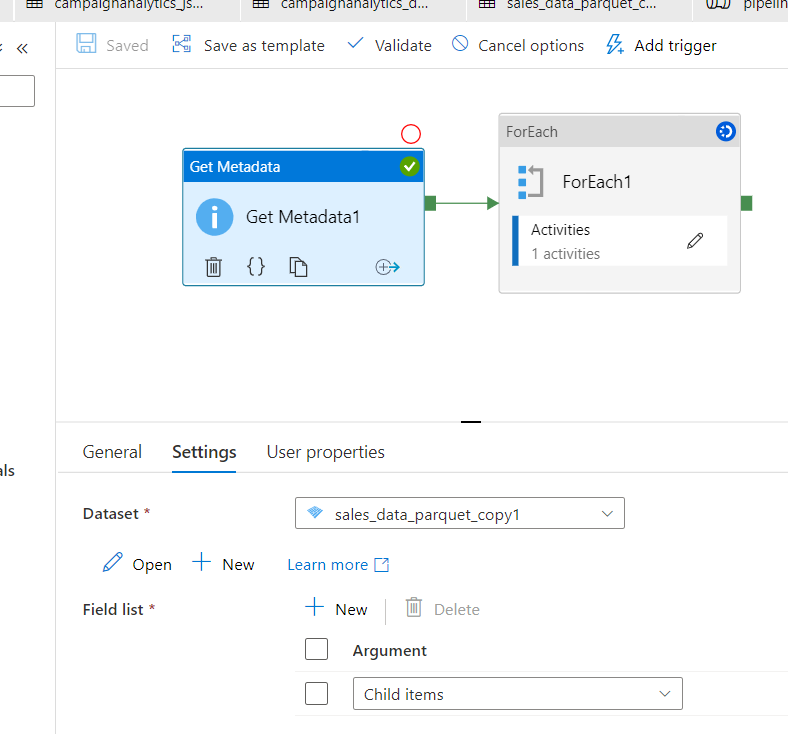
GO

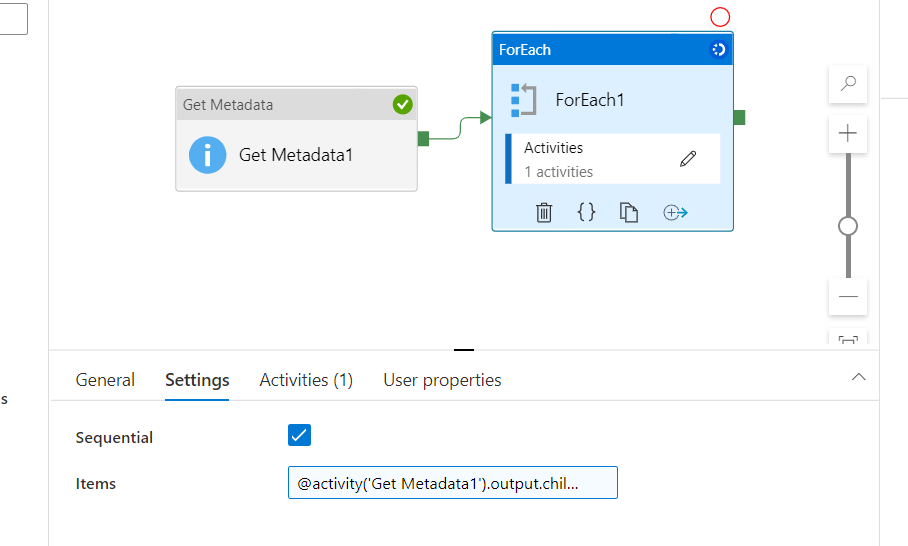




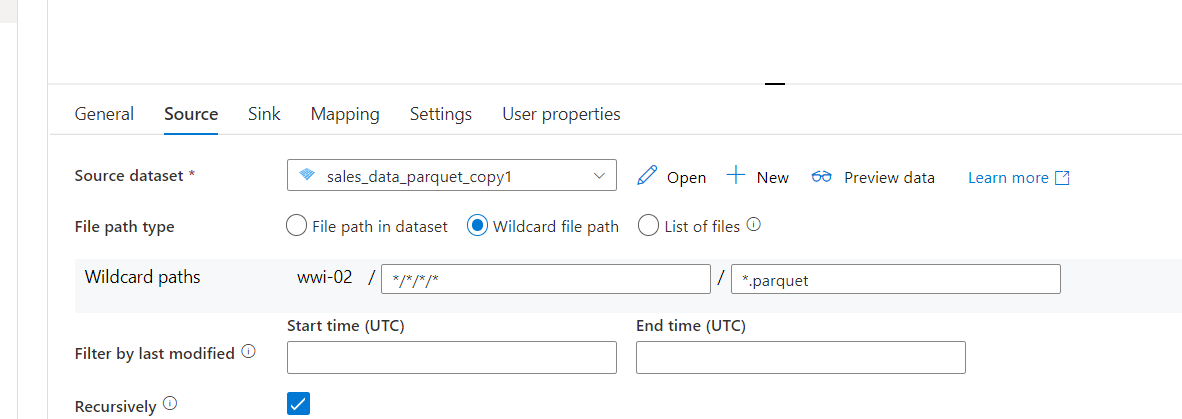
 

We need to create:

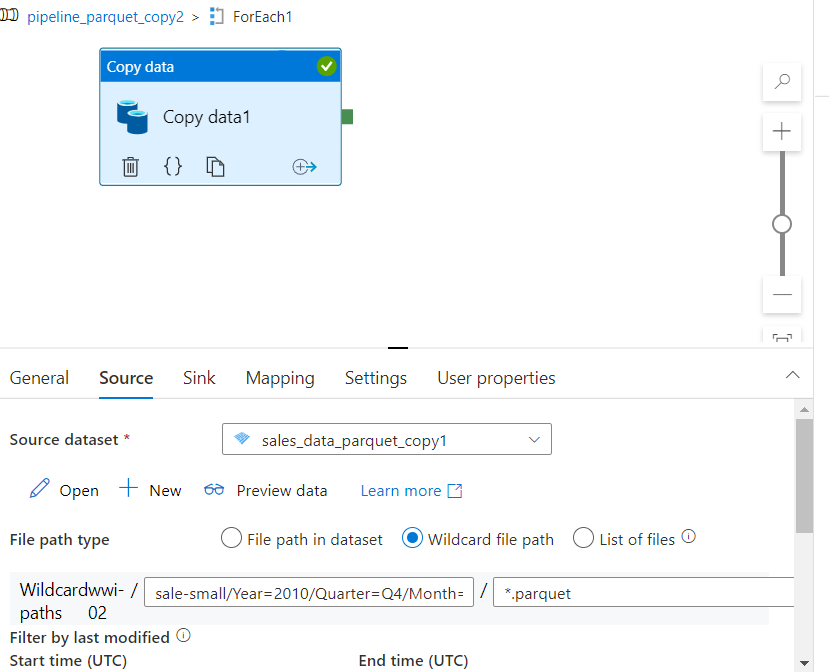


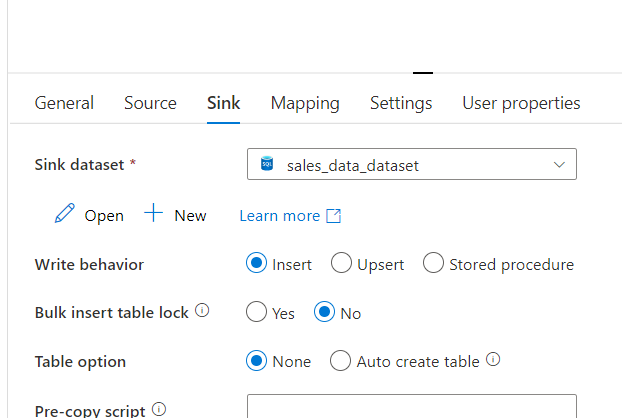


@**activity**('Get Metadata1').output.childItems



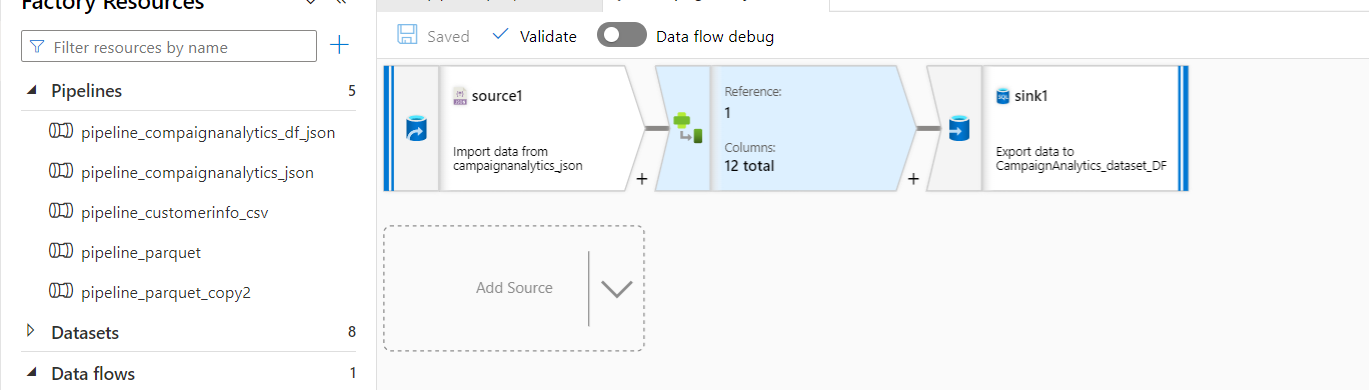
Another method to get only one folder by its path

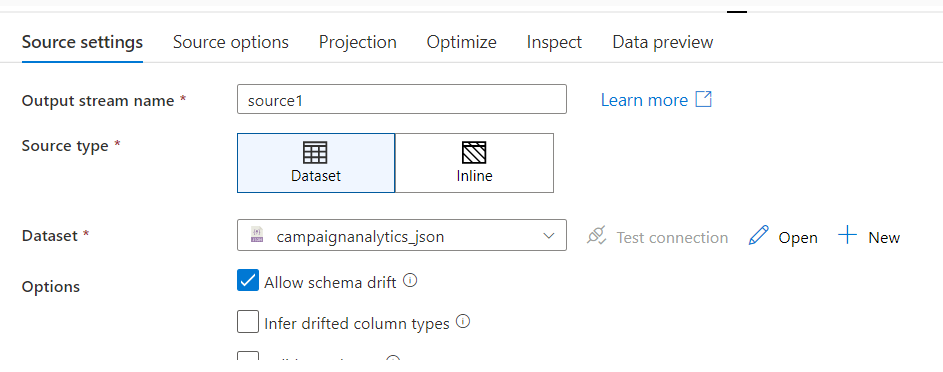




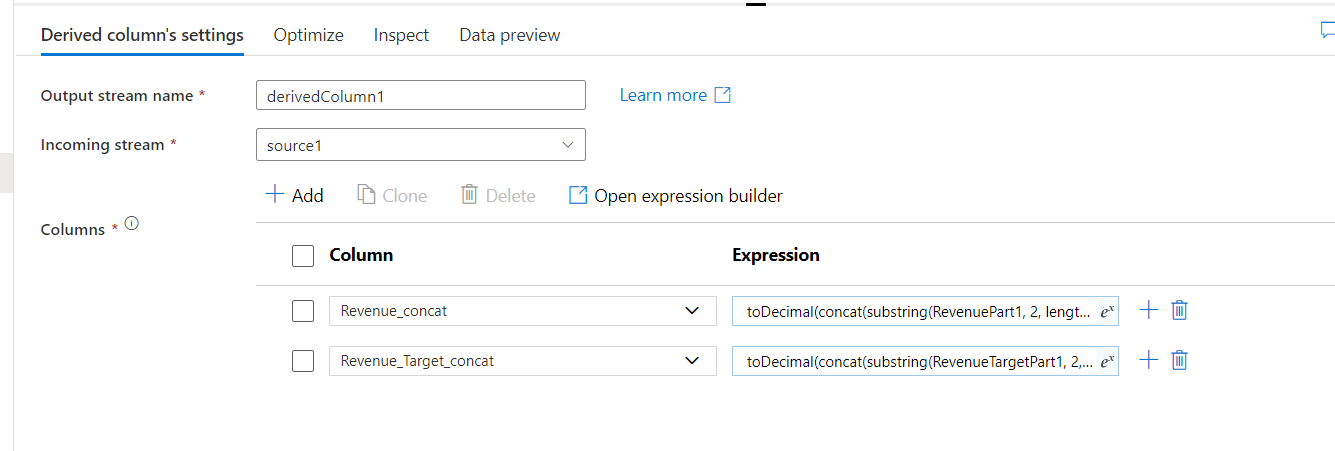
**Method 2:** Create data flows:

Source1





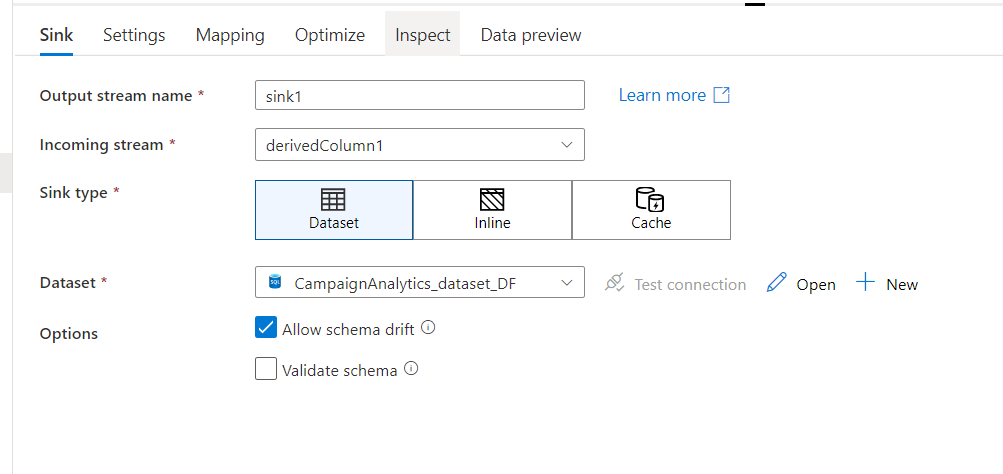
Drived column 1

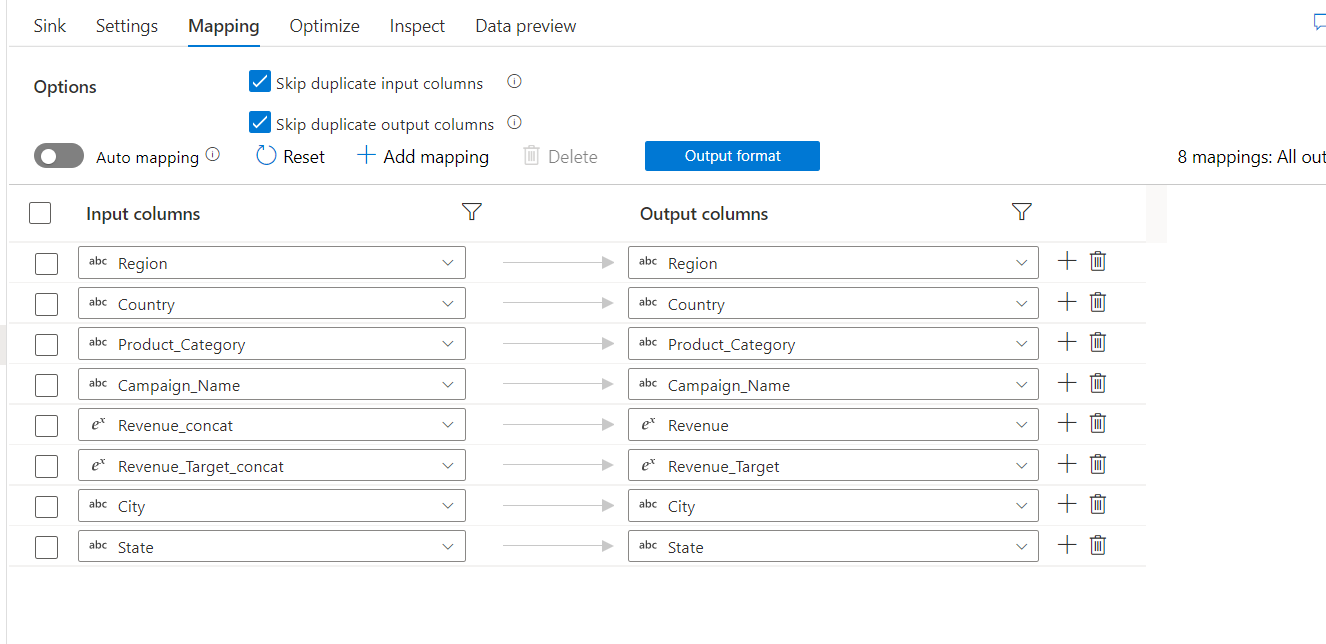


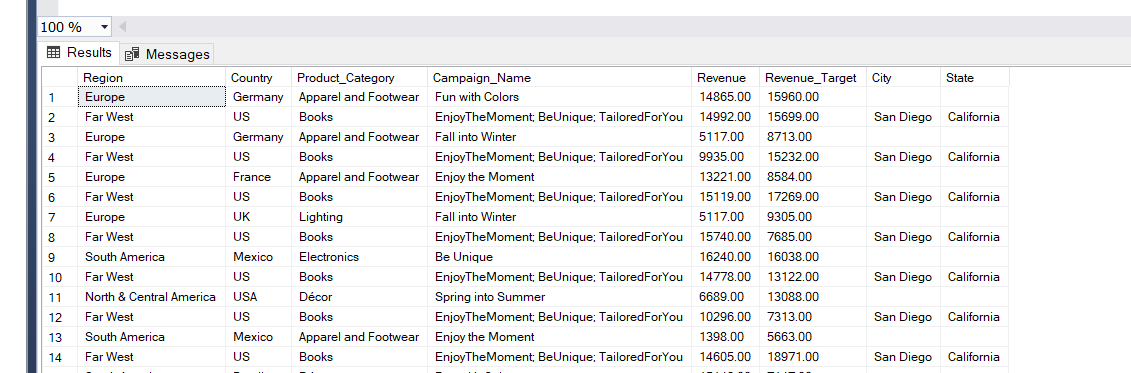
toDecimal(concat(substring(RevenuePart1, 2, length(RevenuePart1)-2), toString(Revenue)), 10, 2)

toDecimal(concat(substring(RevenueTargetPart1, 2, length(RevenueTargetPart1)-2), toString(Revenue\_Target)), 10, 2)

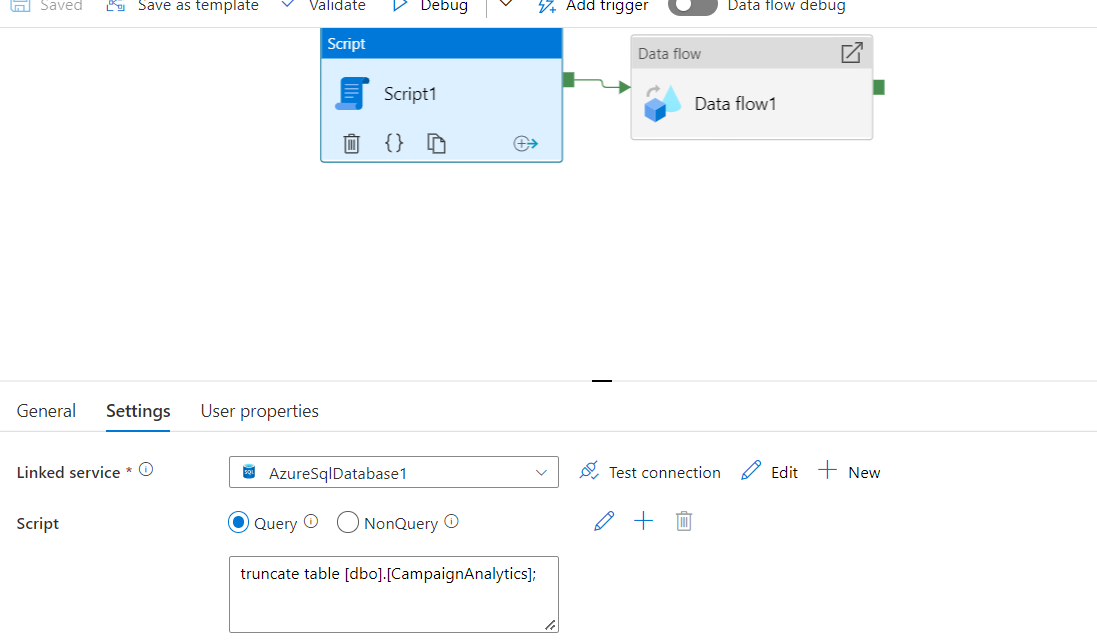
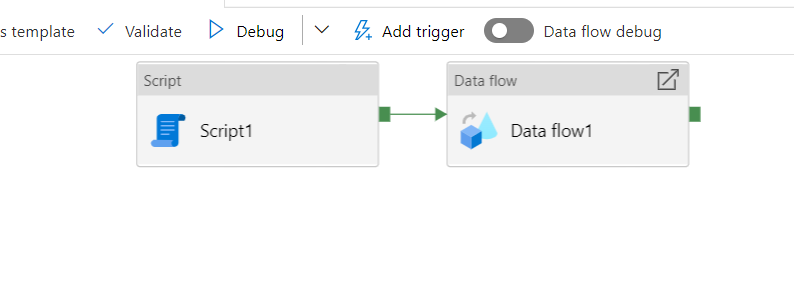
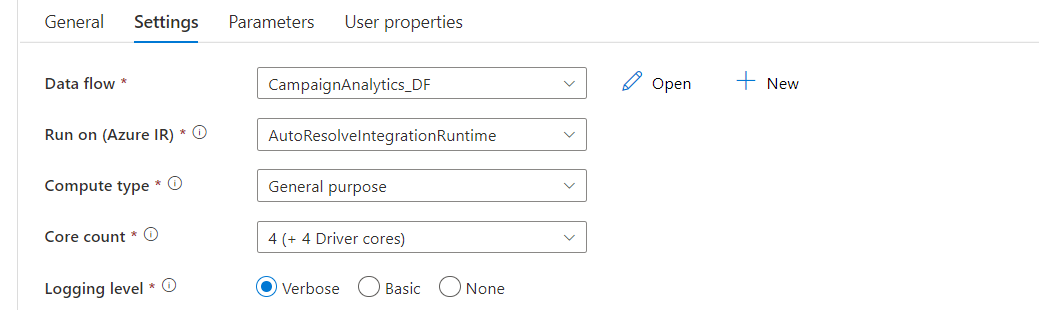
sink 1





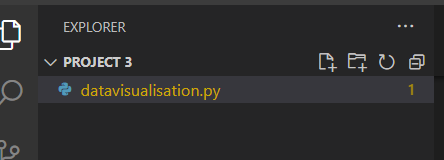


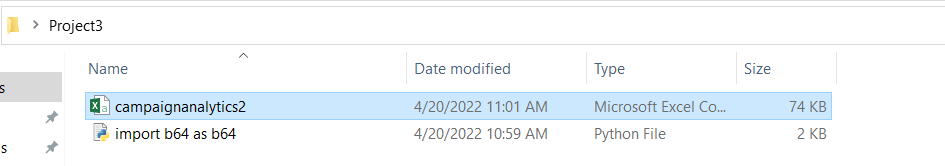
Create the pipeline for the data flow

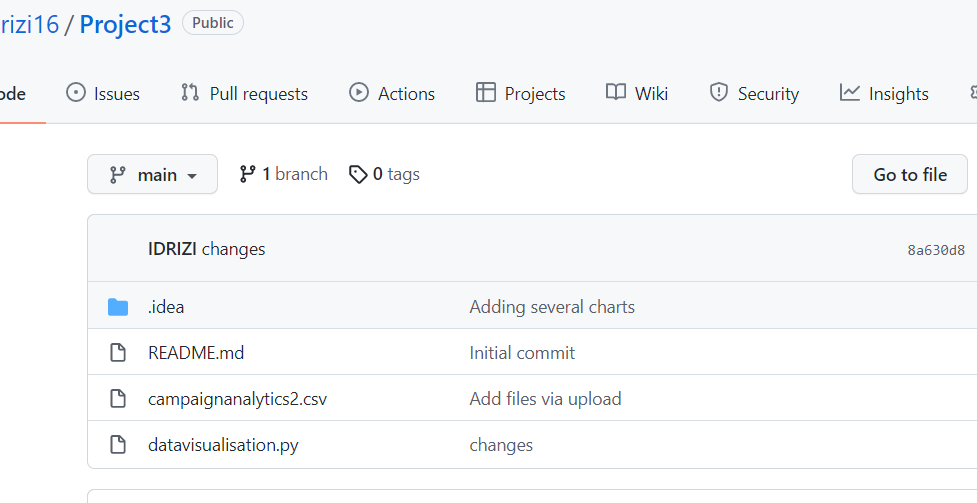
 

Bonus points :

Folders:







The code:

import streamlit as st

import pandas as pd

import pyodbc

server = 'rg-data-project.database.windows.net'

database = 'rg-data-project-team5'

username = 'sqladminuser'

password = 'Admin123+'

cnxn\_DB = pyodbc.connect('DRIVER={SQL Server};SERVER='+server+';DATABASE='+database+';UID='+username+';PWD='+ password)

header = st.container()

dataset = st.container()

features = st.container()

model\_training = st.container()

with header:

    st.title("Welcome to the best project of team 5")

    st.text("In this project we are going to analyze campaigns")

with dataset:

    st.header("Campaign analytics dataset")

    st.text("This dataset was given to us by our instructor")

    query = "SELECT Region, Country, Product\_Category, Campaign\_Name, Revenue, Revenue\_Target FROM dbo.CampaignAnalytics"

    campaign\_data = pd.read\_sql(query, cnxn\_DB)

    #campaign\_data = pd.read\_csv('campaignanalytics2.csv')

    st.write(campaign\_data.head())

    st.subheader('Several distributions')

    campaign\_data\_by\_Region = campaign\_data.groupby(['Region']).sum()[["Revenue", "Revenue\_Target"]]

    campaign\_data\_by\_Country = campaign\_data.groupby(['Country']).sum()[["Revenue", "Revenue\_Target"]]

    campaign\_data\_by\_Product\_Category = campaign\_data.groupby(['Product\_Category']).sum()[["Revenue", "Revenue\_Target"]]

    campaign\_data\_by\_Campaign\_Name = campaign\_data.groupby(['Campaign\_Name']).sum()[["Revenue", "Revenue\_Target"]]

    st.bar\_chart(campaign\_data\_by\_Region)

    st.area\_chart(campaign\_data\_by\_Country)

    st.line\_chart(campaign\_data\_by\_Product\_Category)

    st.bar\_chart(campaign\_data\_by\_Campaign\_Name)

with features:

    st.header("The features we created")

with model\_training:

    st.header("Time to train the model!")

    st.text("Here you get to choose the hyperparameters of the model and see how the performance changes.")

Run streamlit run datavisualisation.py

