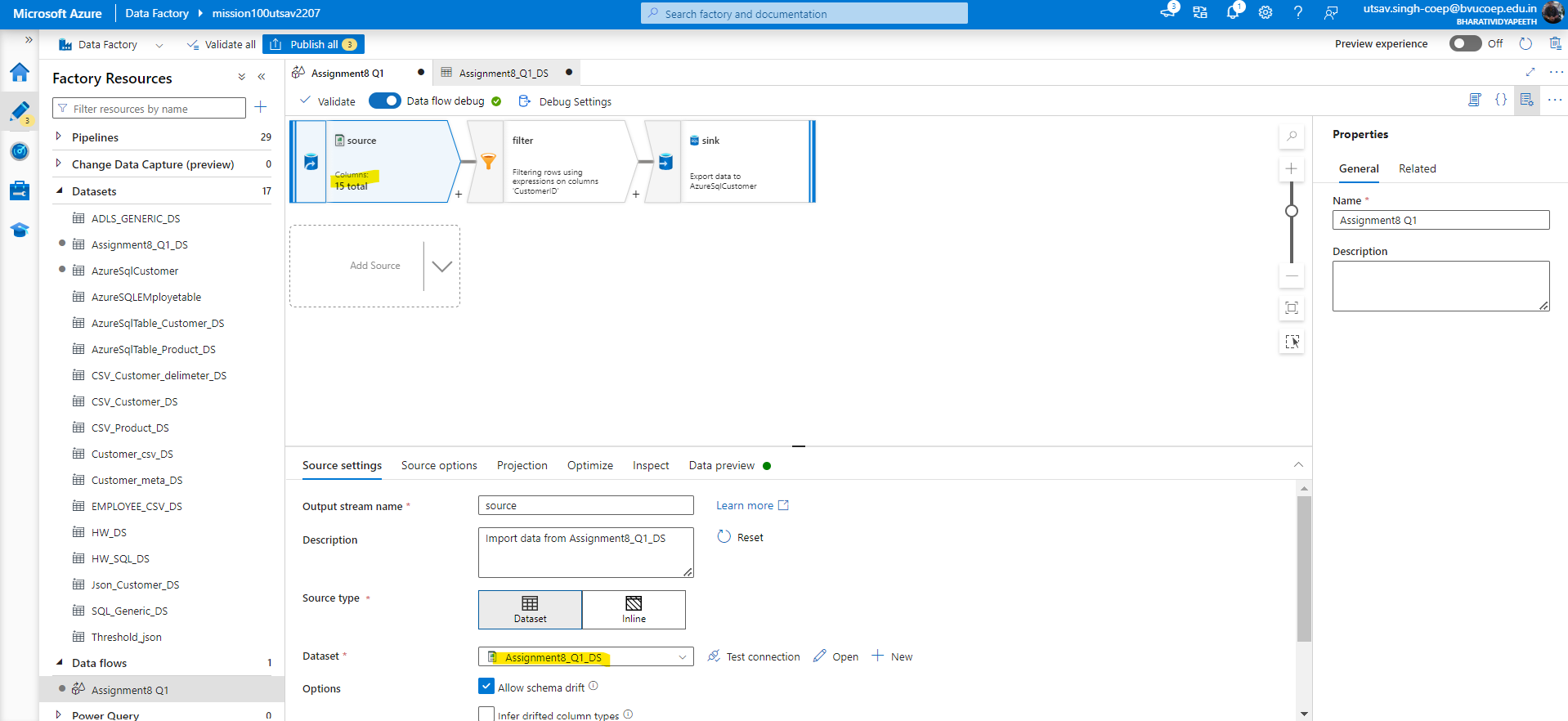
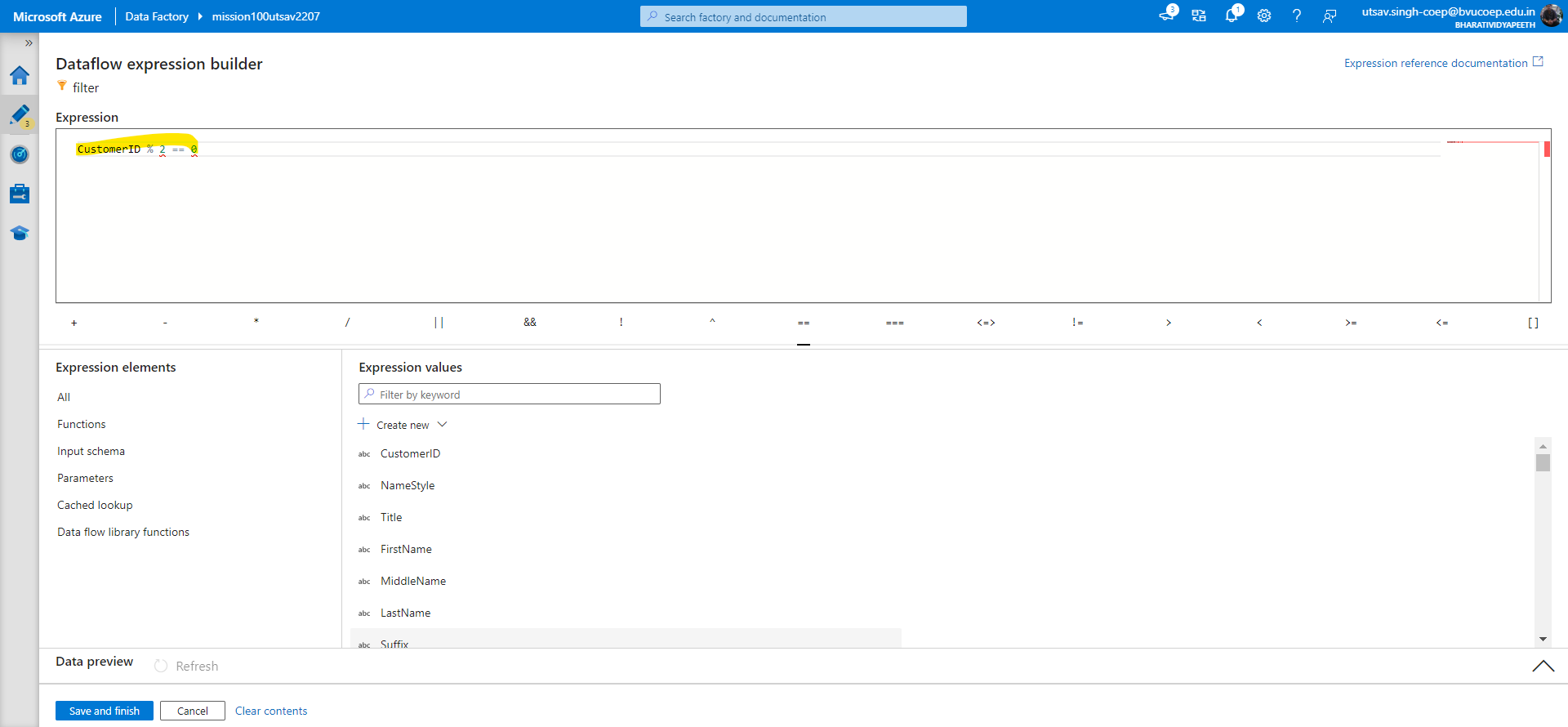
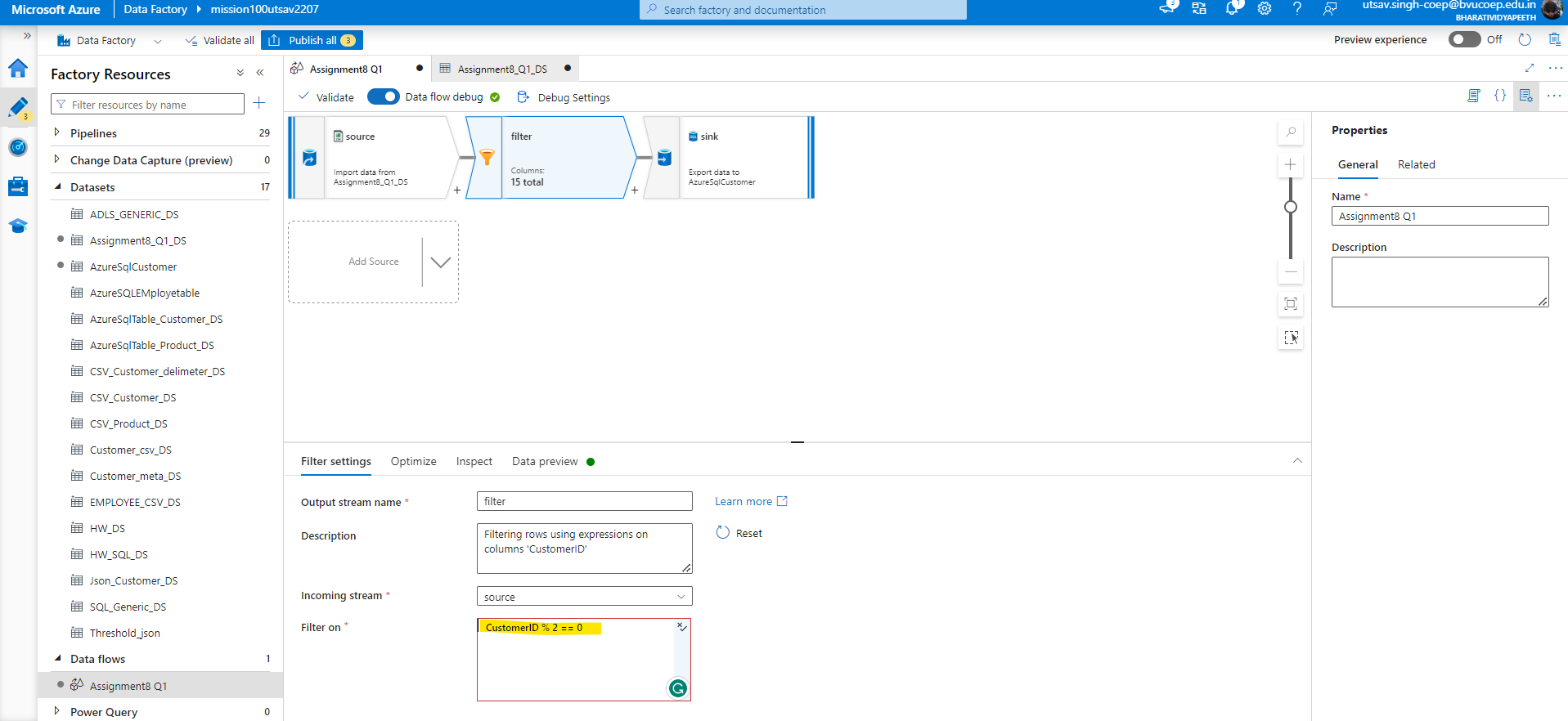
## Question 1: Create a pipeline to copy the customer data from csv file to SQL where the customer id is an even number.

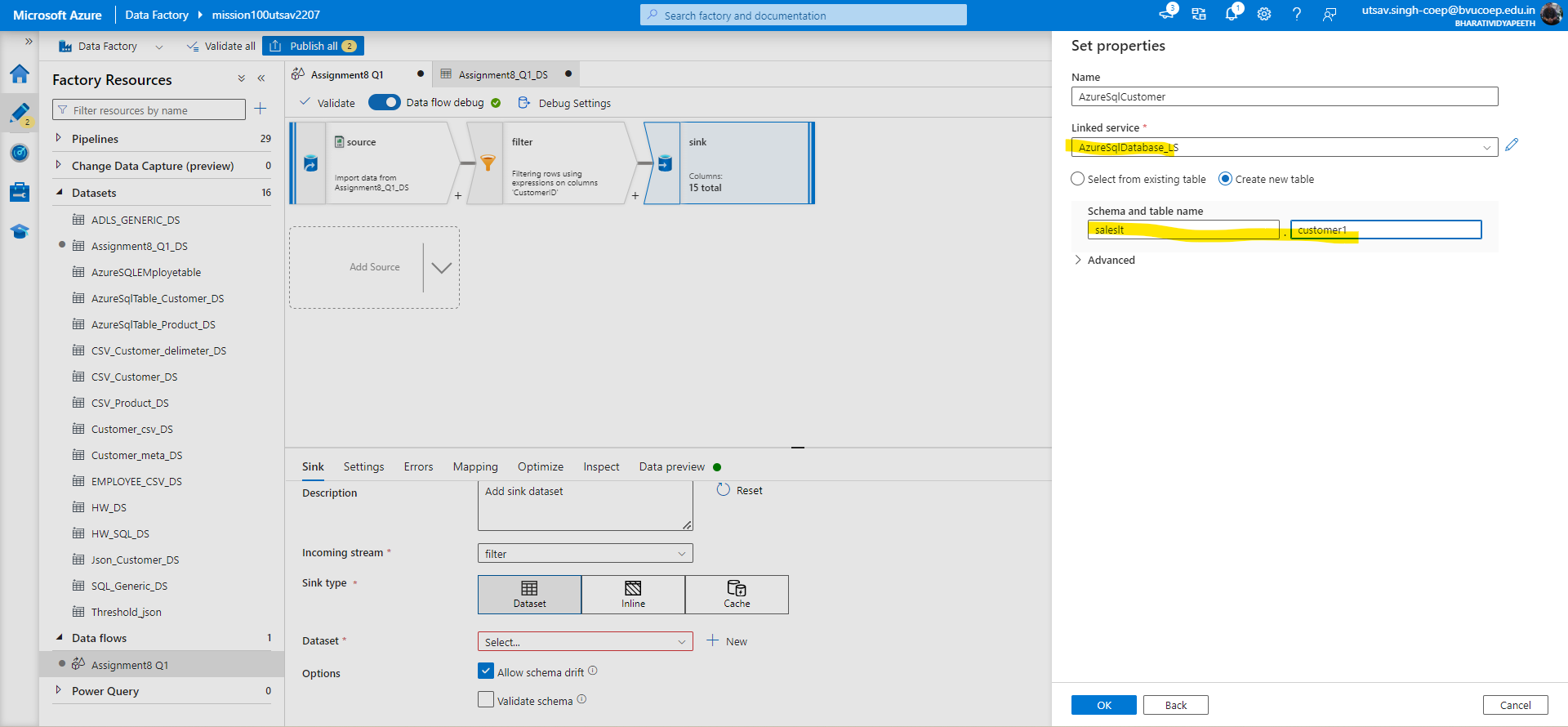
## (Attaching the Customer File)

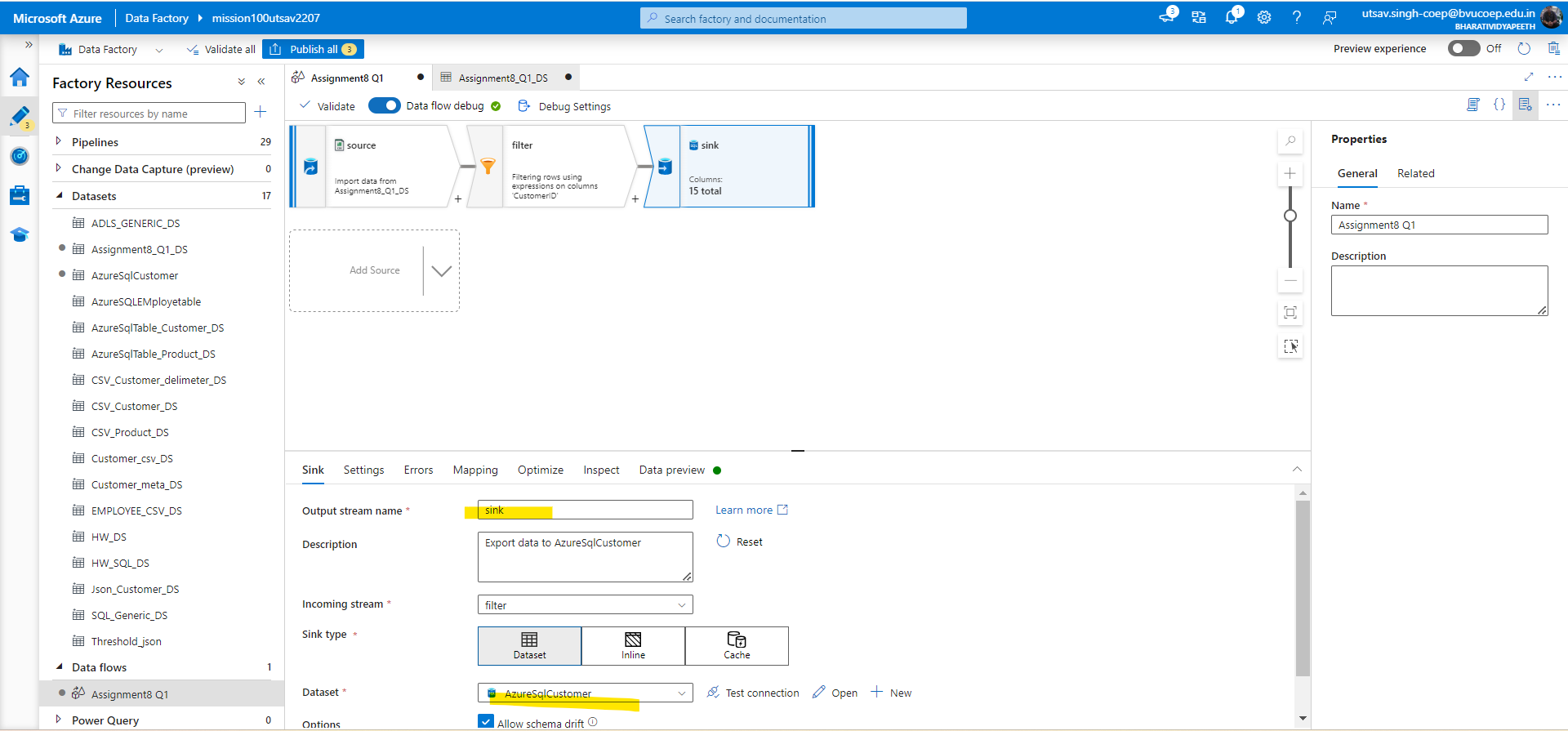
## Solution:- Firstly upload customer file to adls location and create dataset then create a new dataflow and add source then select the datasource created then add filter use expression builder and give customer\_id%2==0 then create sink and create new sql dataset to save as new table.

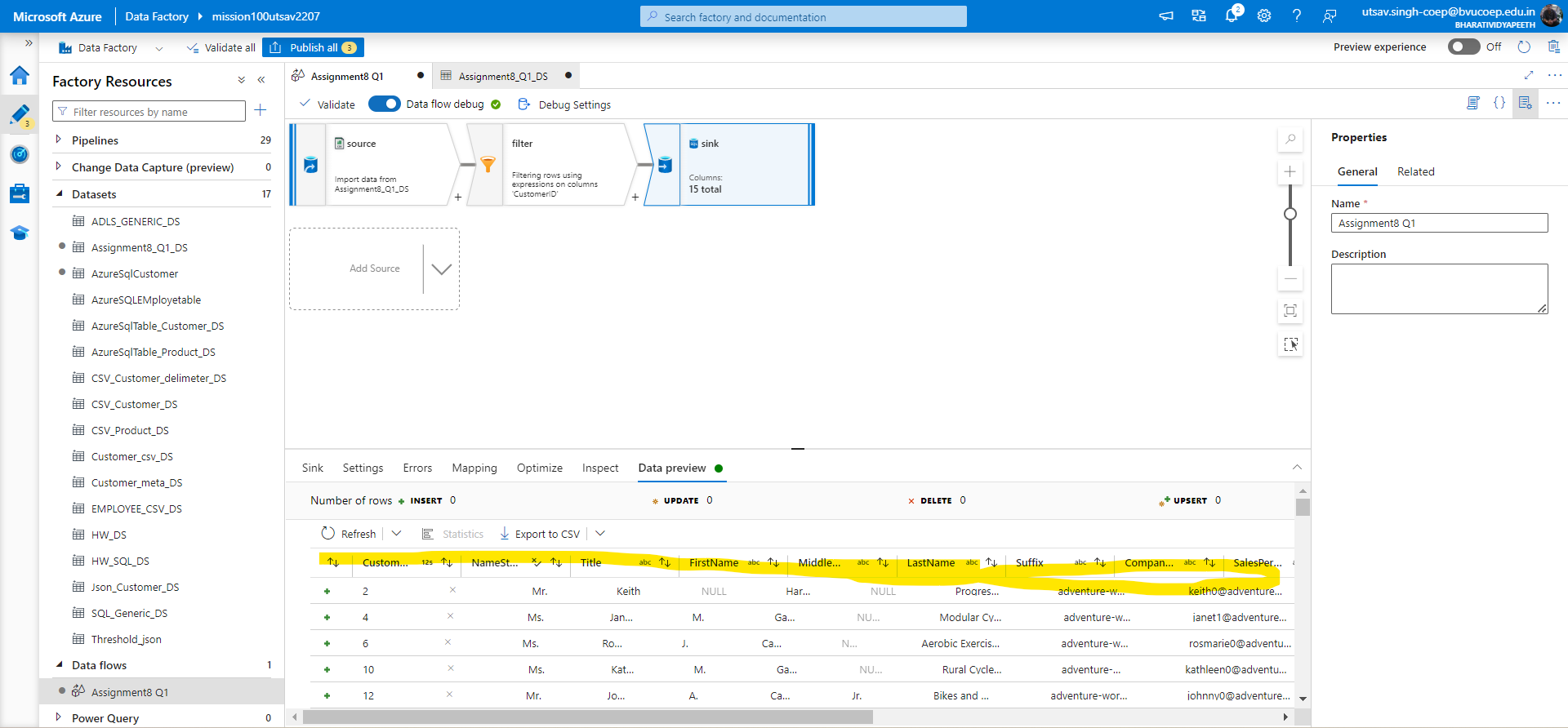




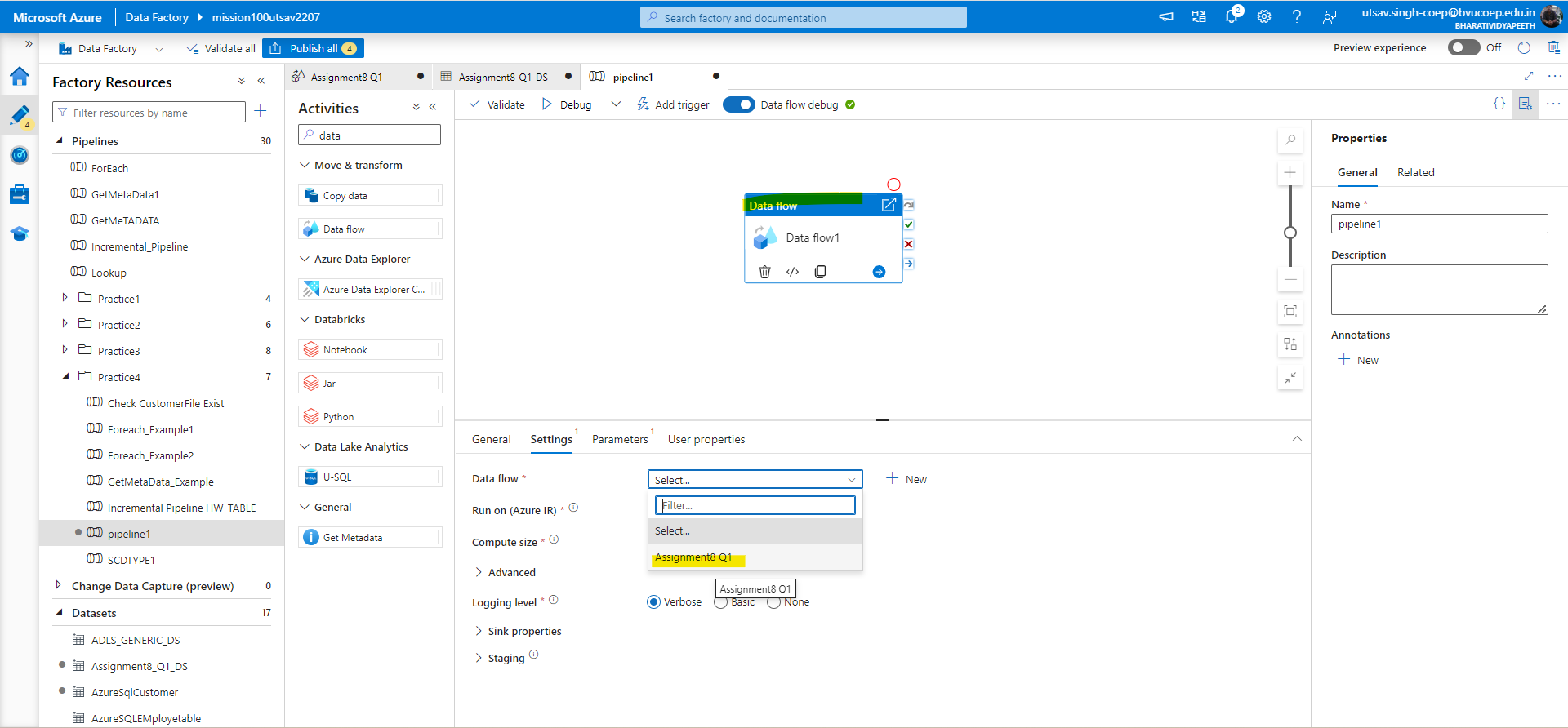






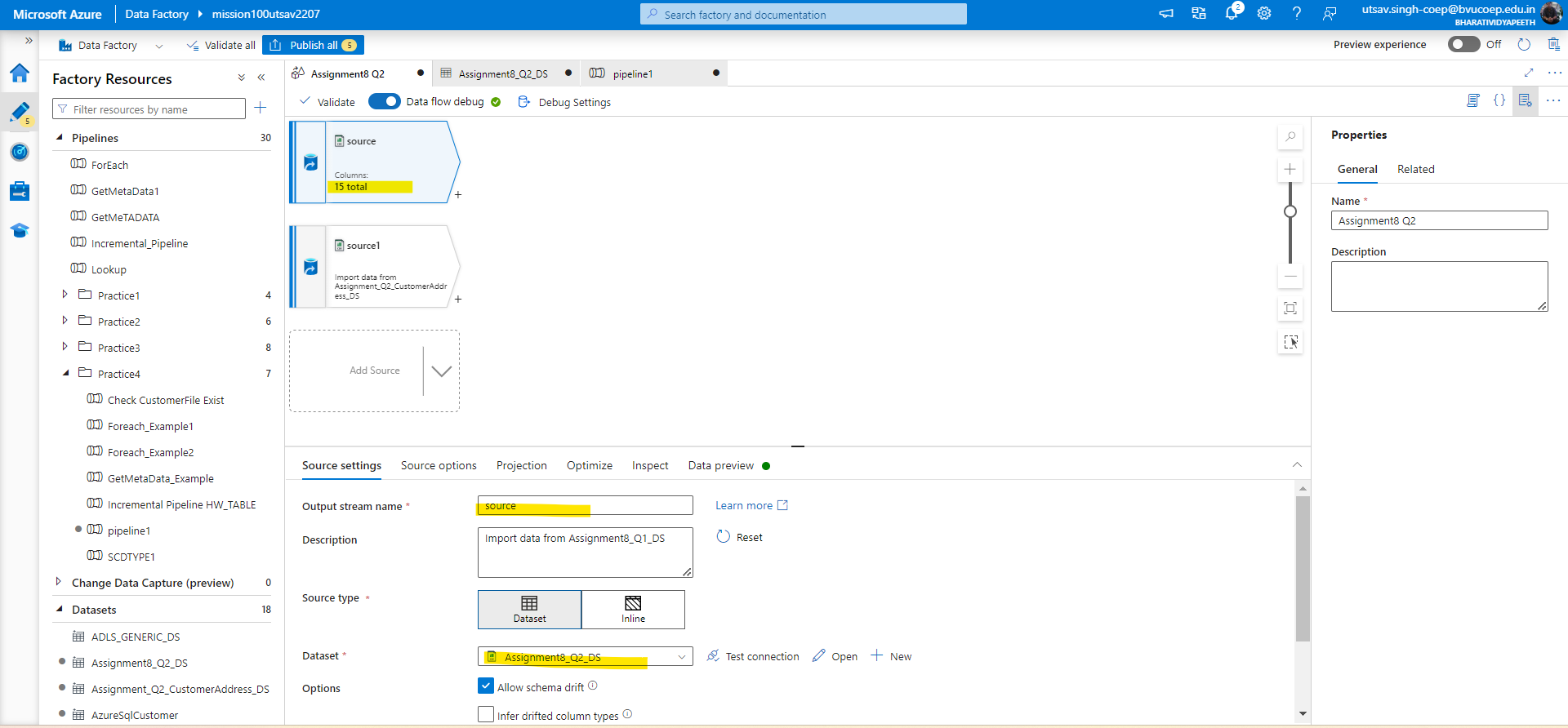


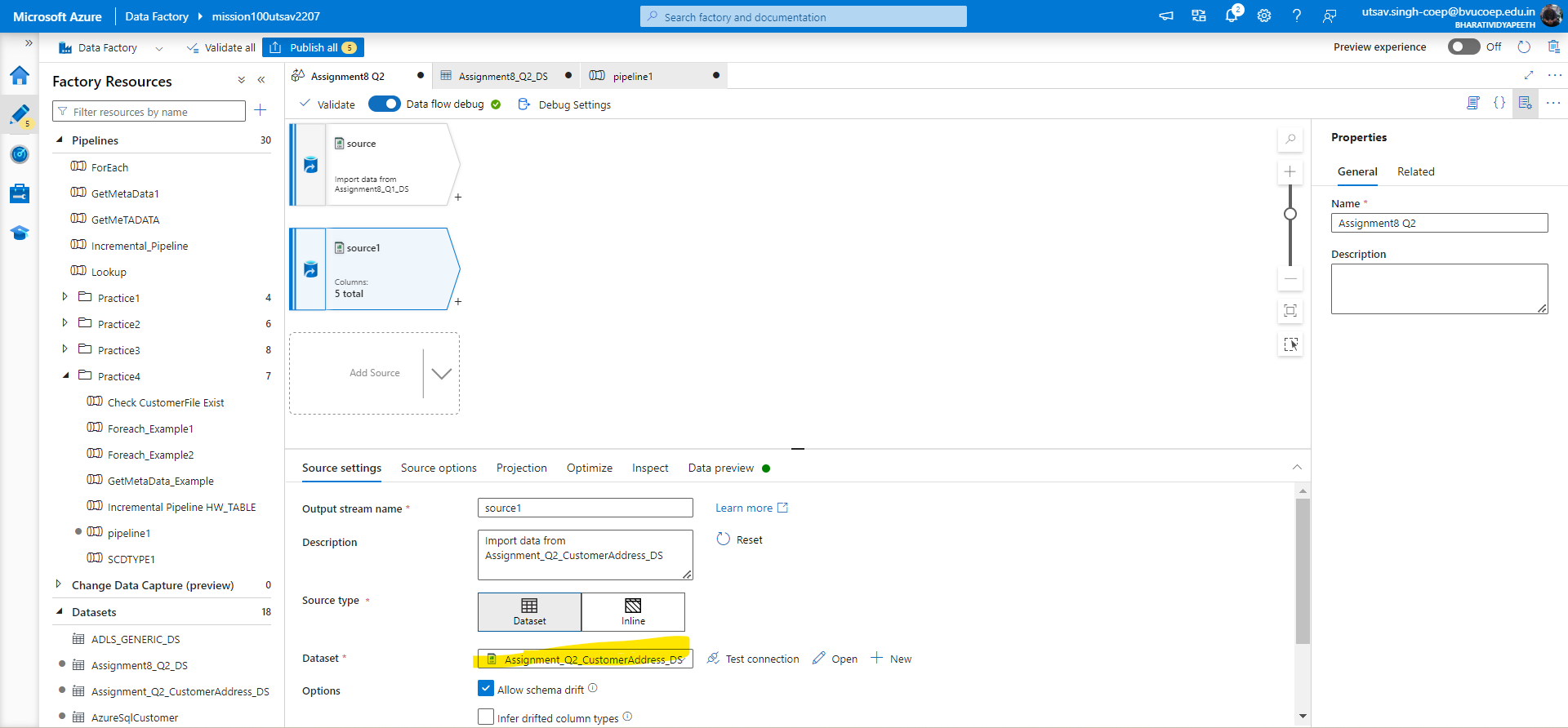
Finally create a dataflow activity and call the dataflow.

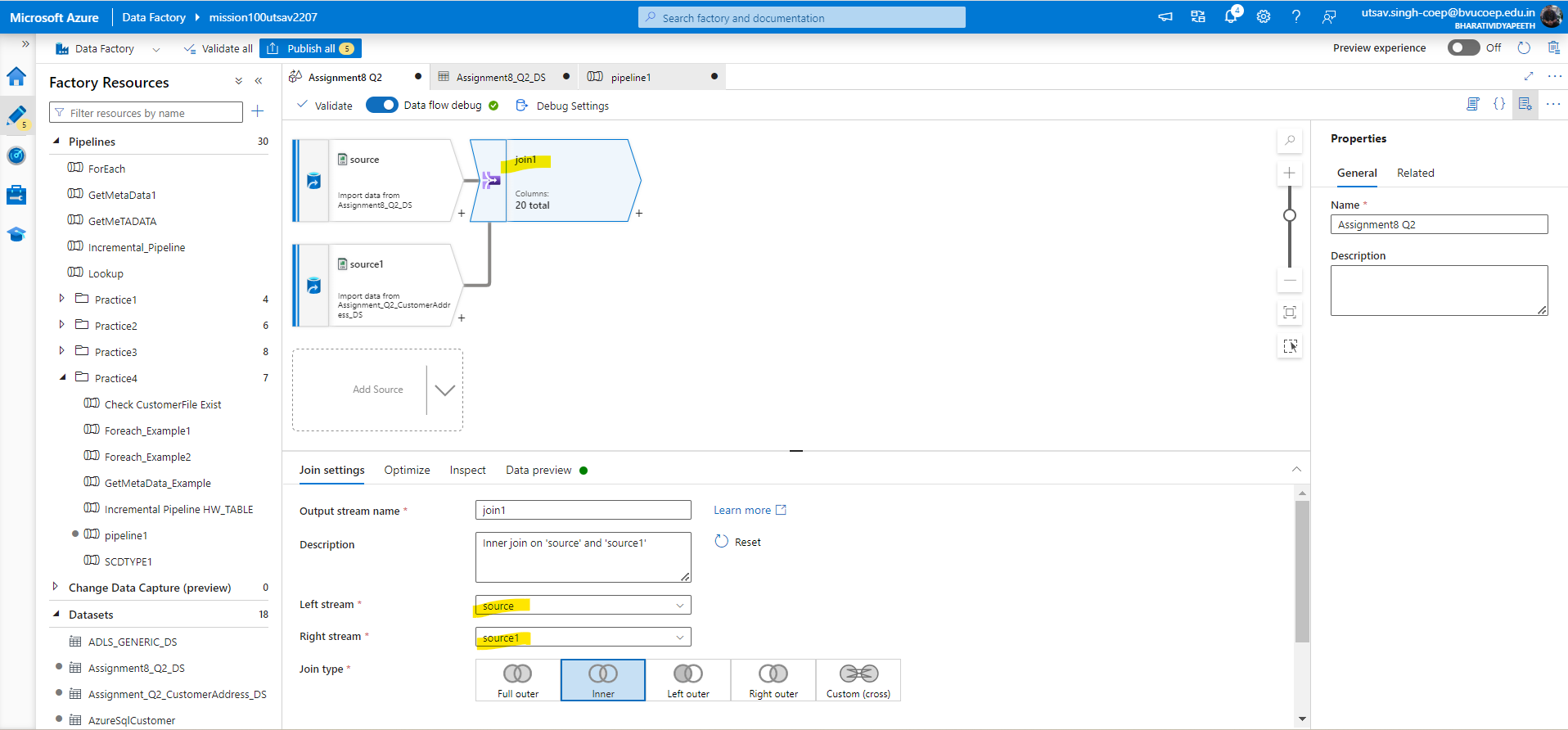


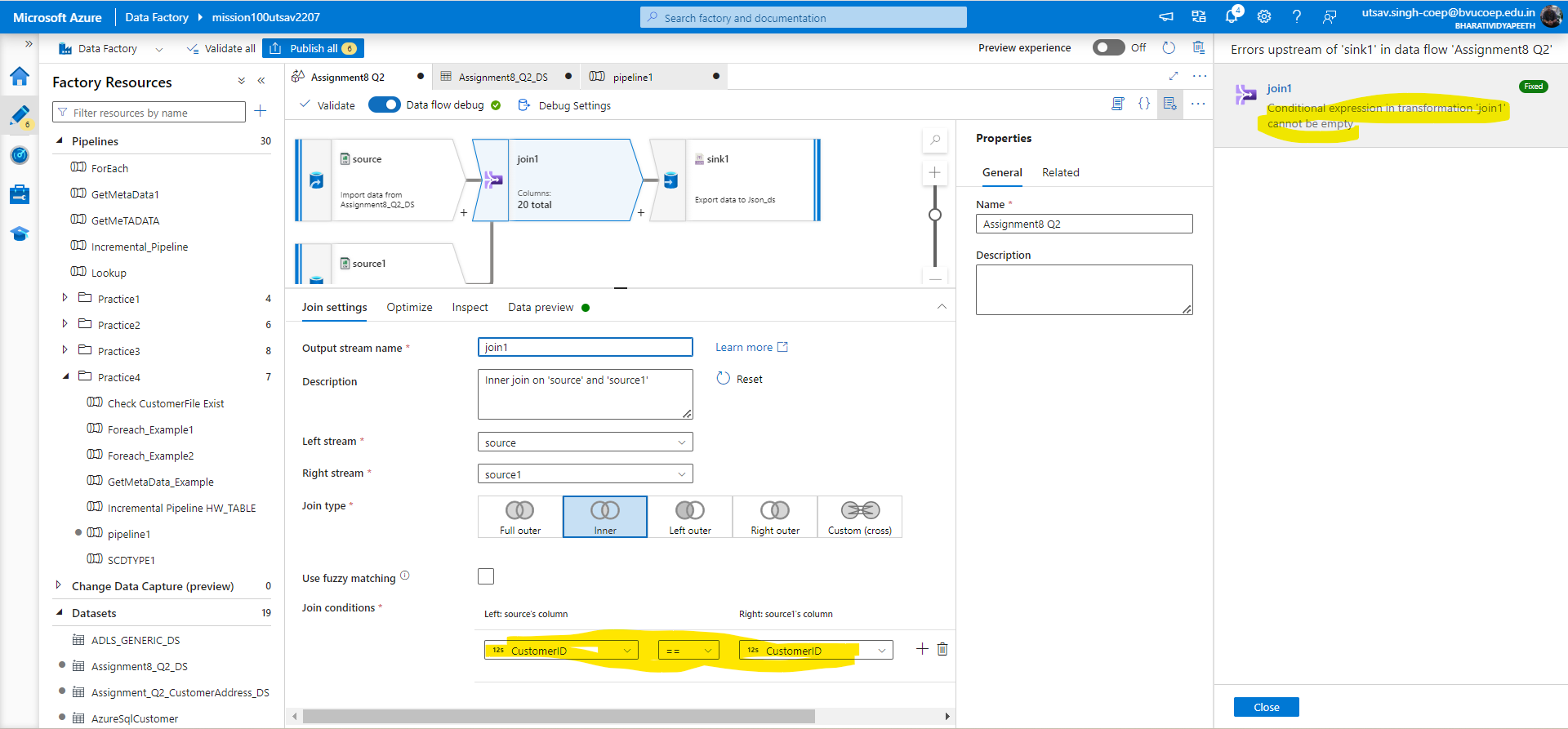
## Question 2: Create a pipeline to join the two files (Customer, Customer Address) based on customer id and save the result as a JSON file.

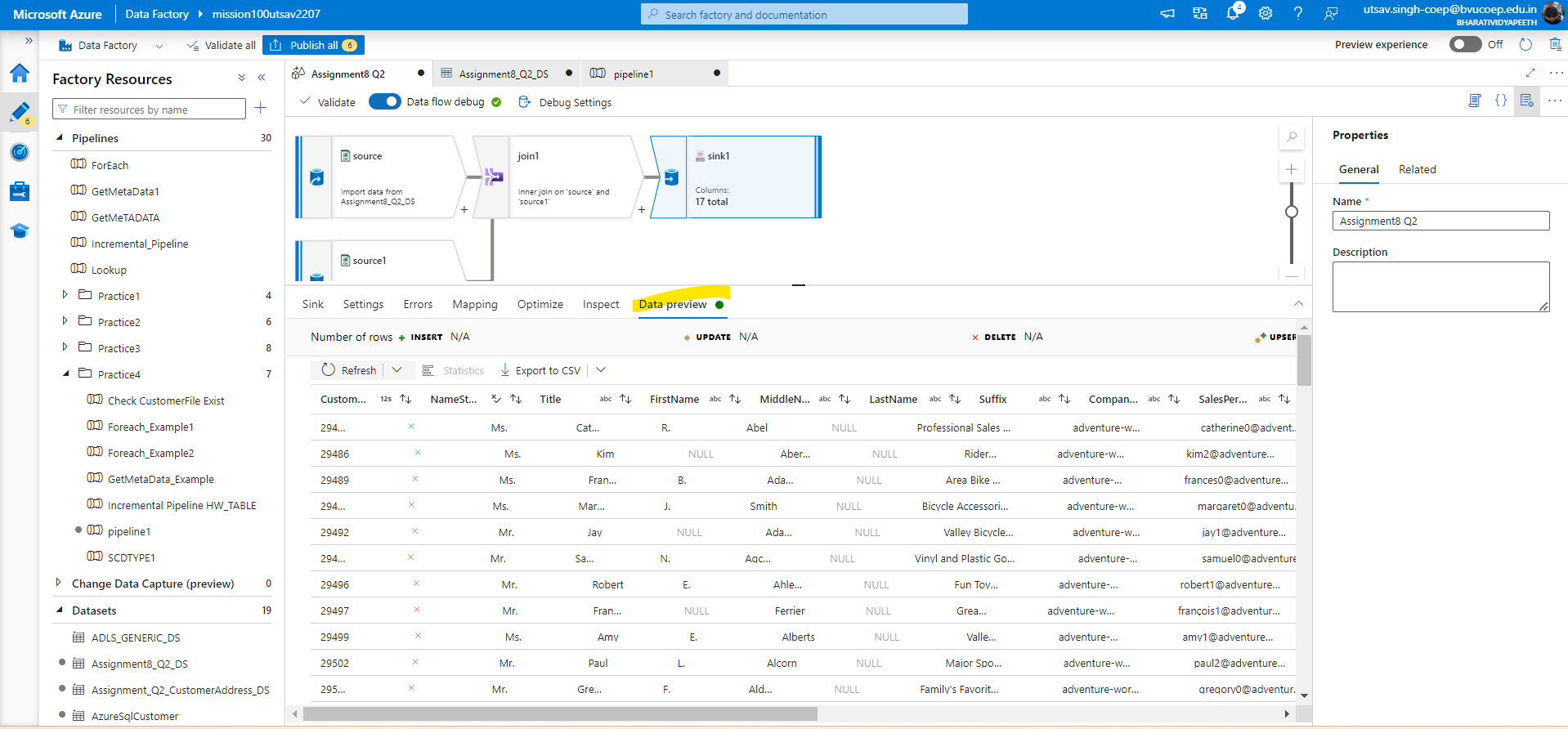
Solution:-Create dataflow and create two datasource one for customer and other for customerAddress then create join and join both table using inner and create sink then preview data and recall using dataflow activity.









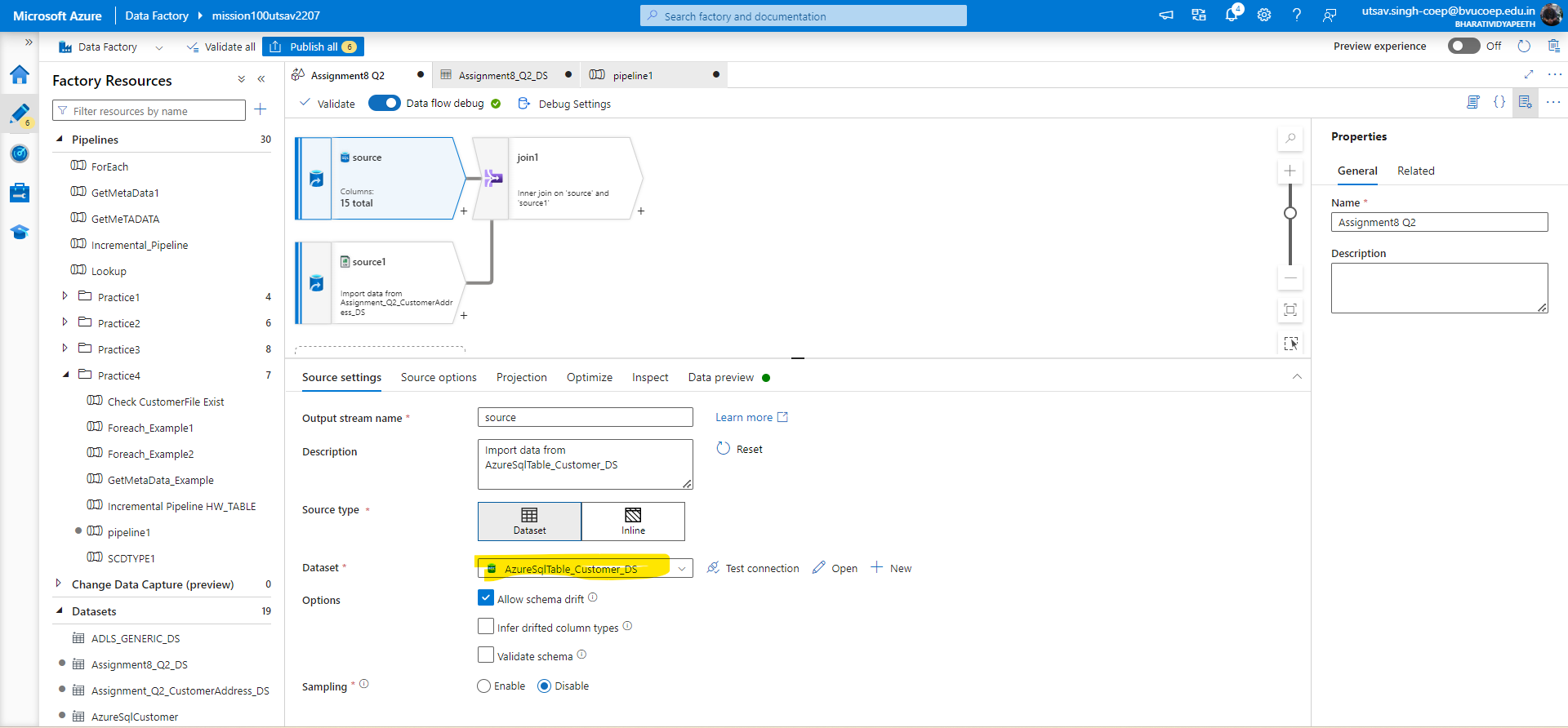


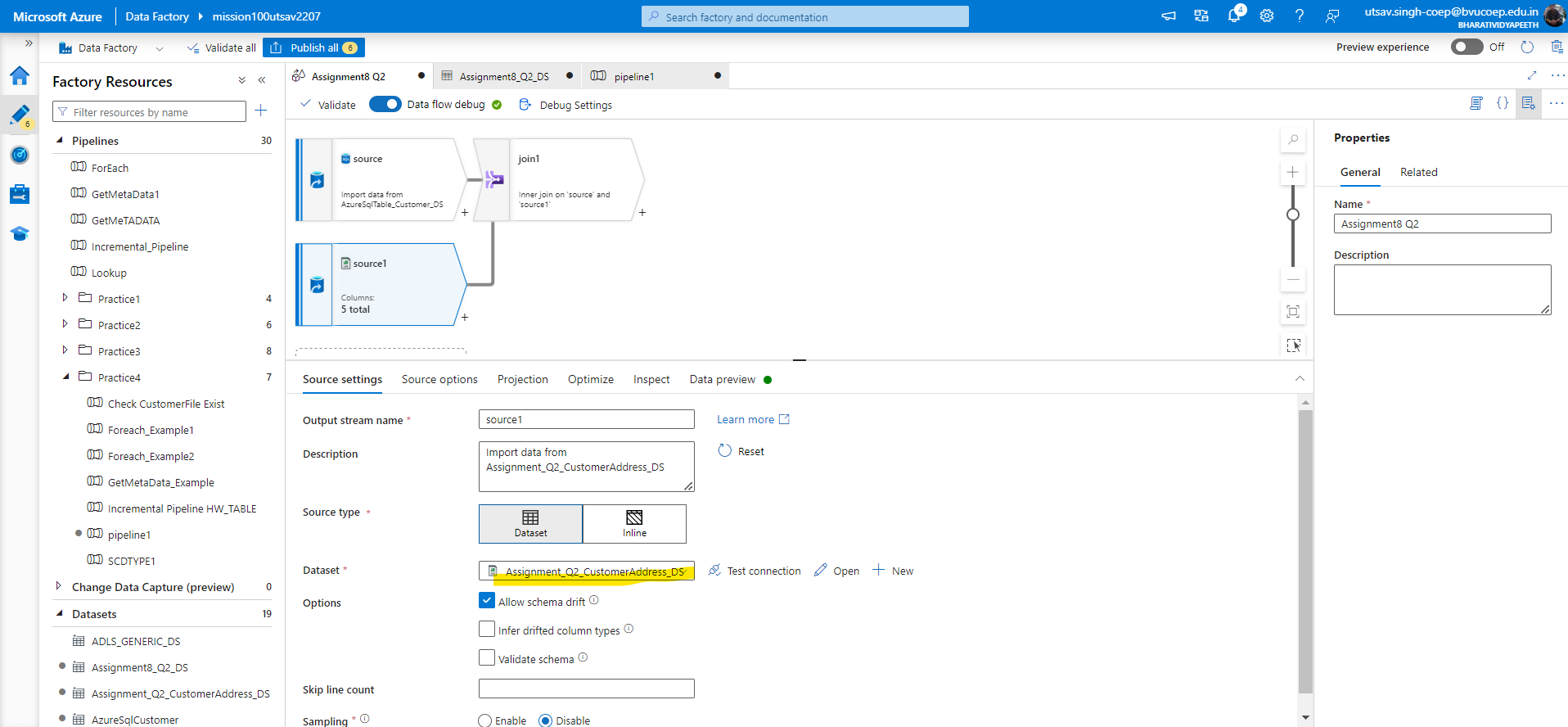
## 

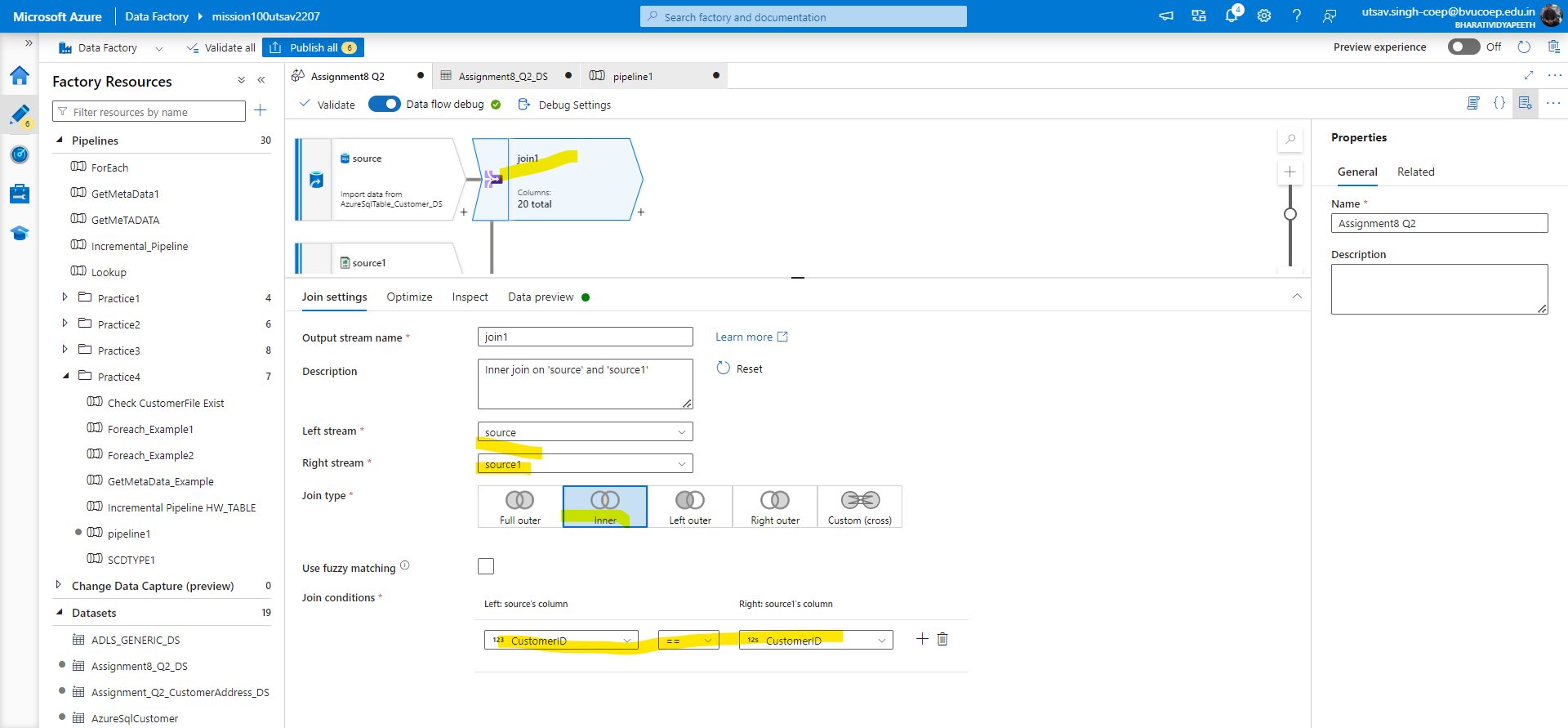
## Question 3: Create a pipeline to read the Customer table data from SQL and CustomerAddress data from CSV, join both of them,

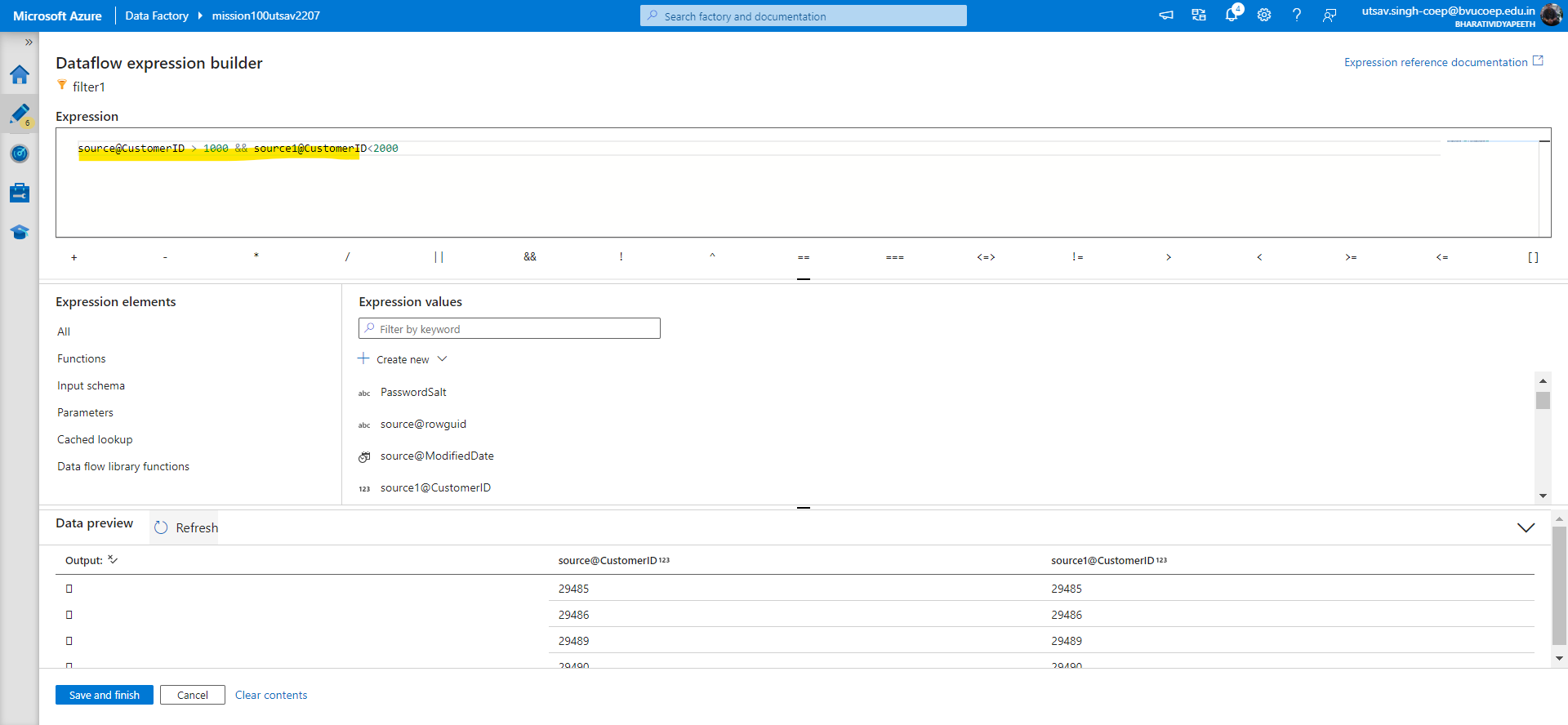
## and then save the result where customer id> 1000 & Customer id <2000 in ascending order as a Parquet file.

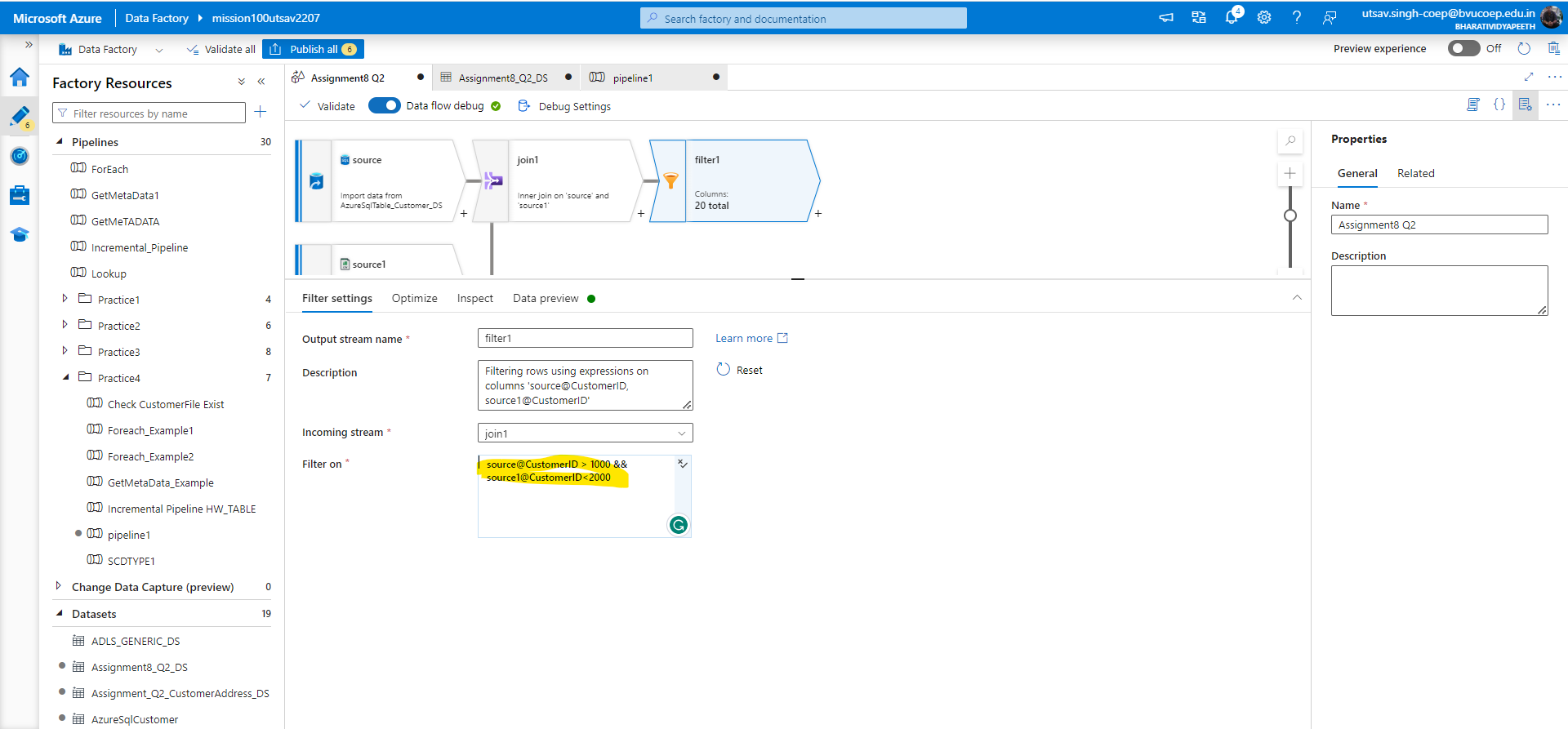
Solution:-Create a two datasource one will point to customer table DS and second will point to customerAddress CSV. I’m using previous dataflow for customerADRESS.Then use filter for checking customer id> 1000 & Customer id <2000 then will create sort for sorting and in sort give the col name and sort by asc then use sink there you create the dataset which will save in parquet format then create dataflow activity and call the DATAFLOW.

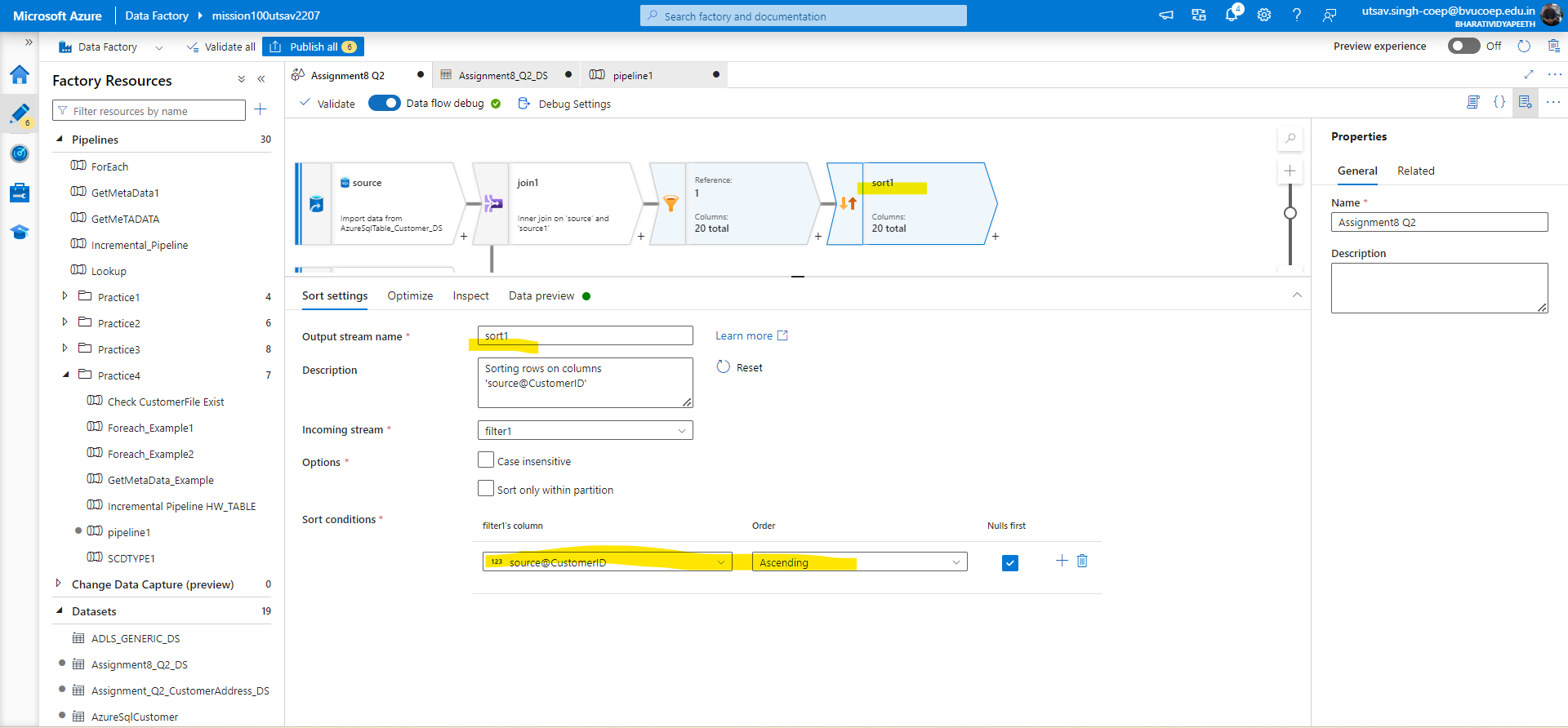


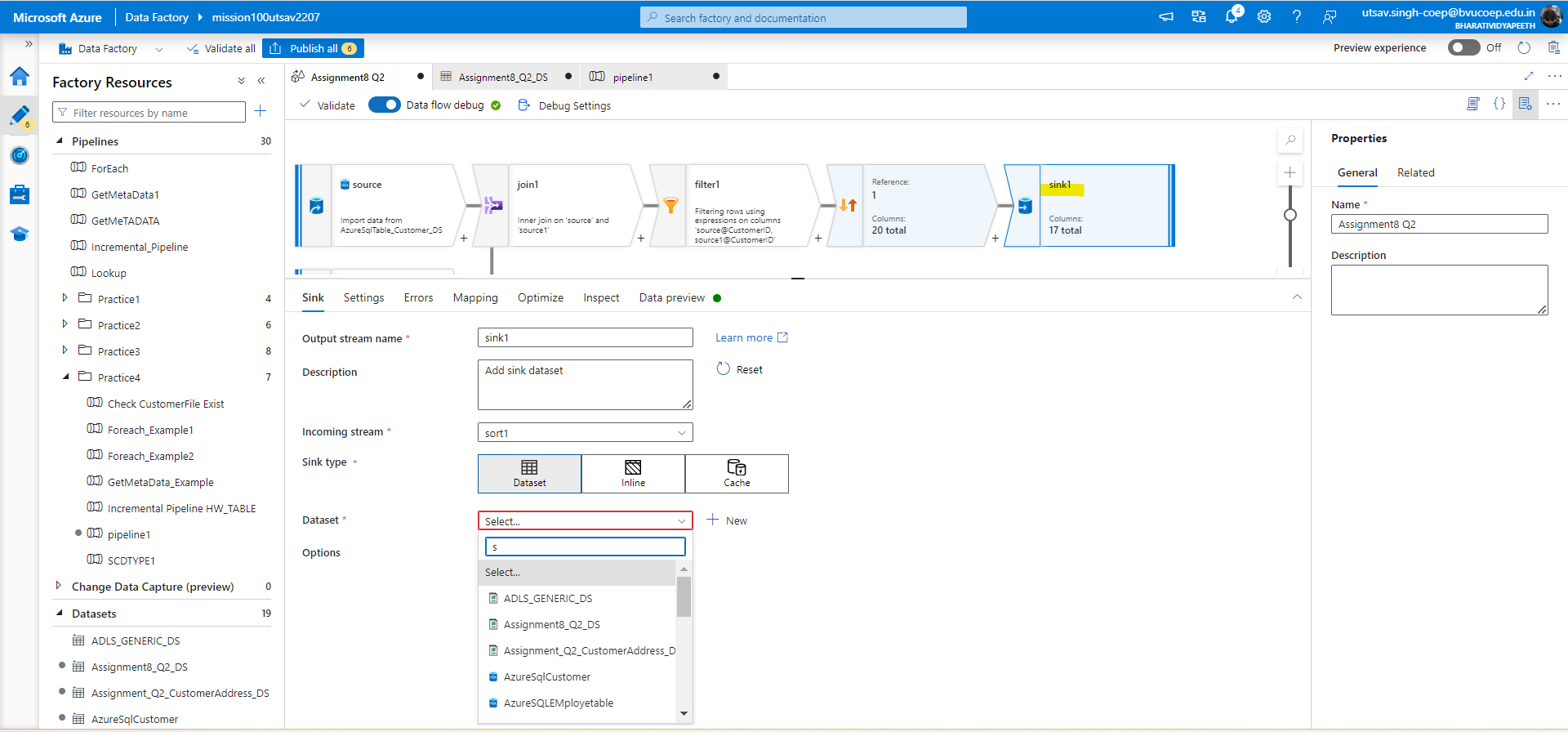










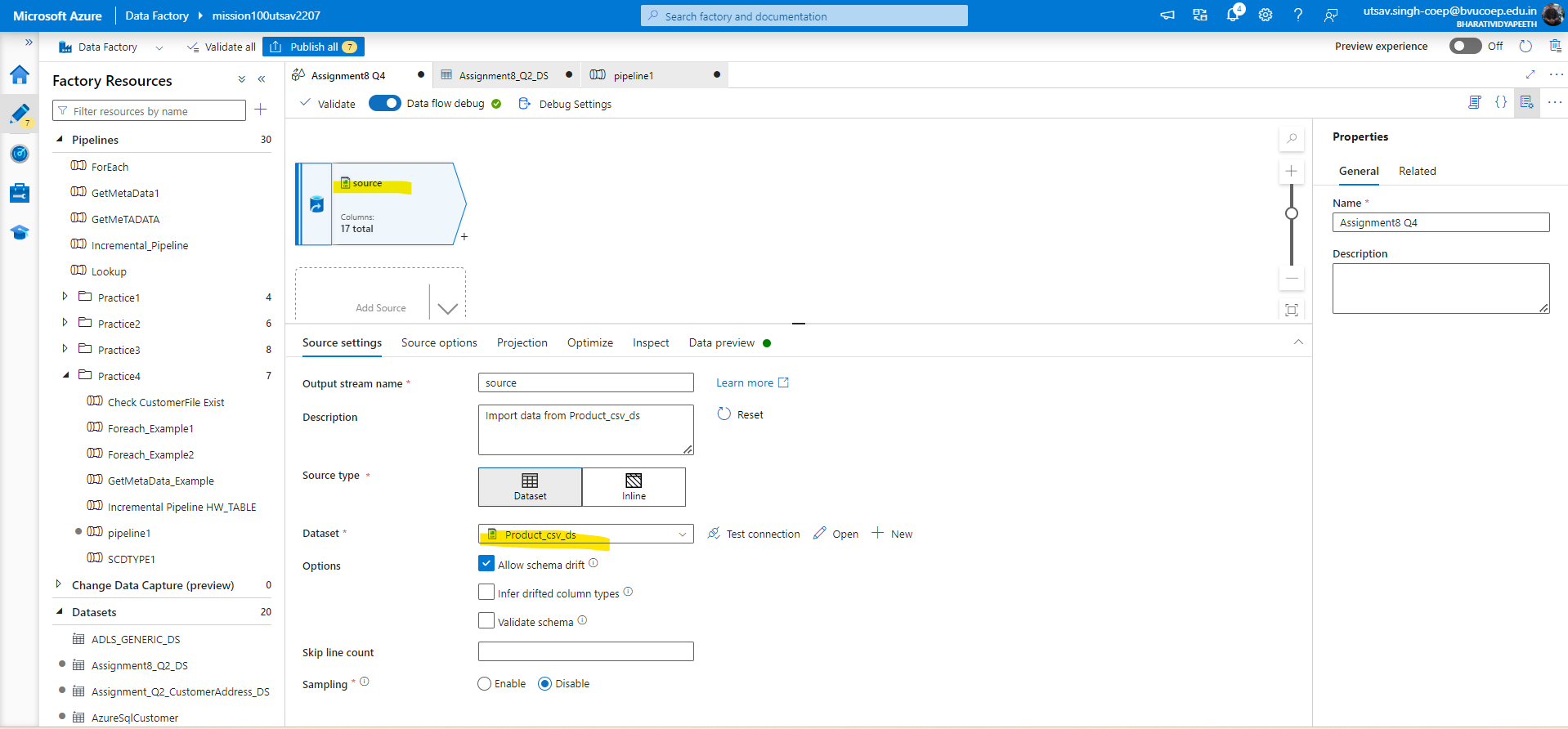


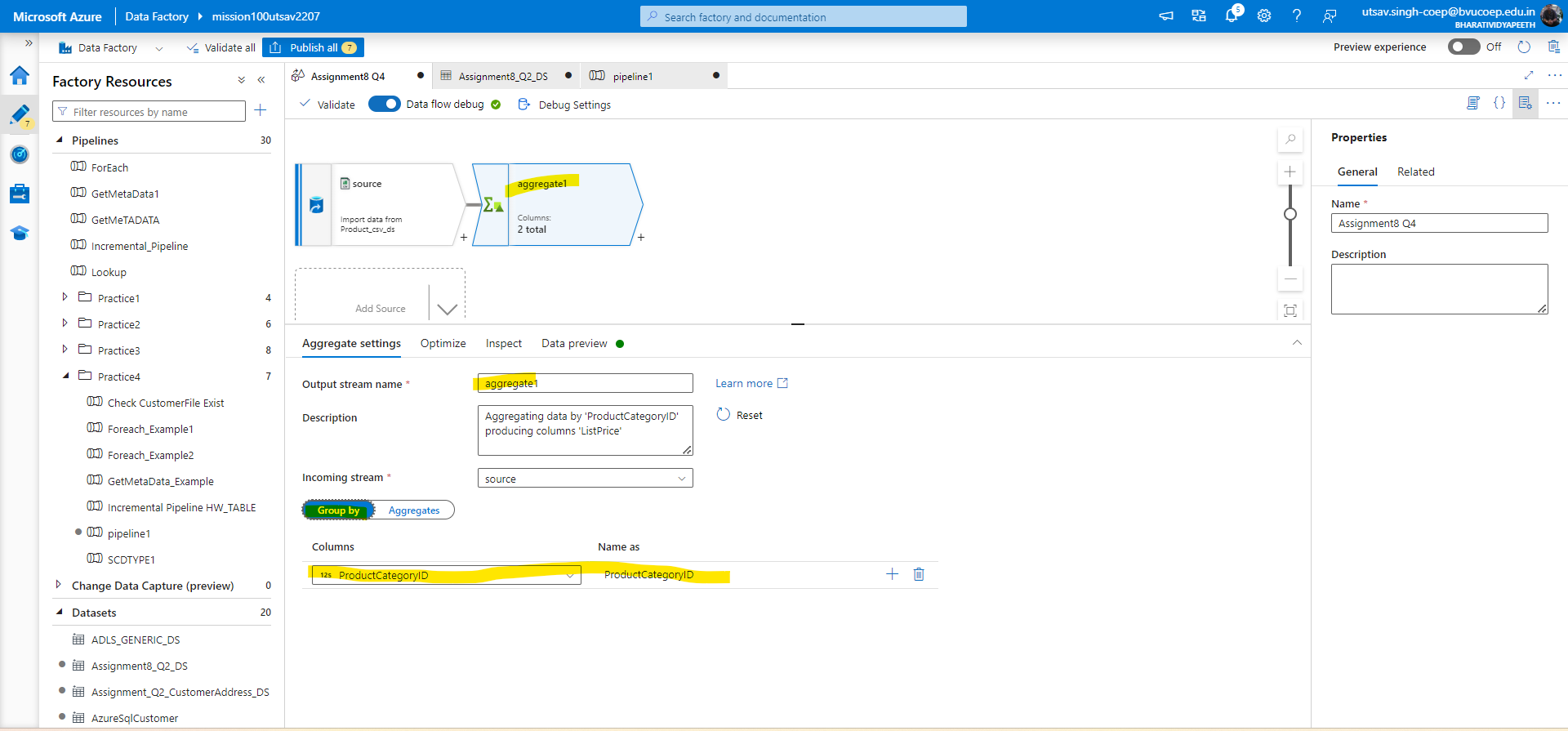
## 

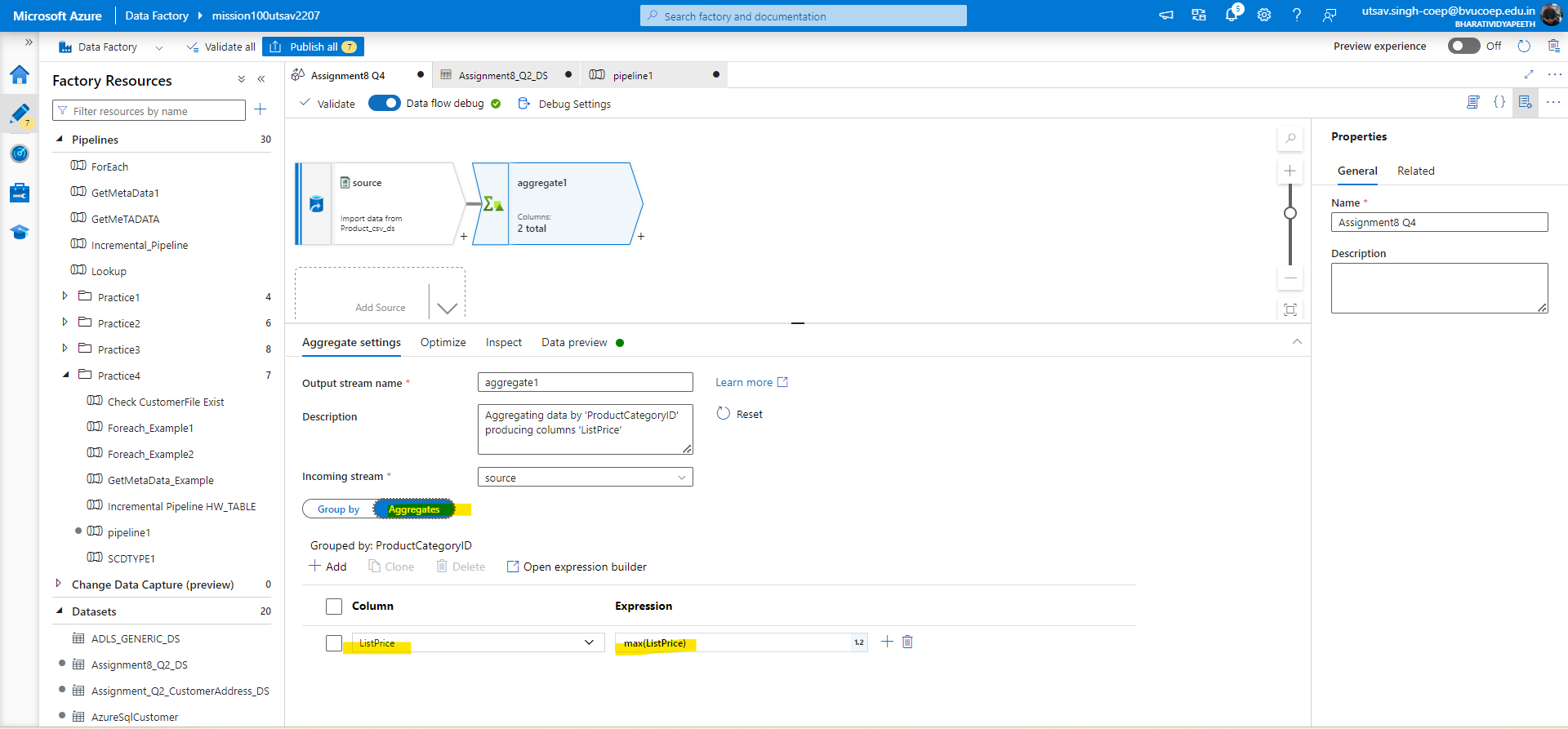
## Question 4: create a pipeline to read the Product CSV file, and calculate the highest listPrice of any product under each productcategory.

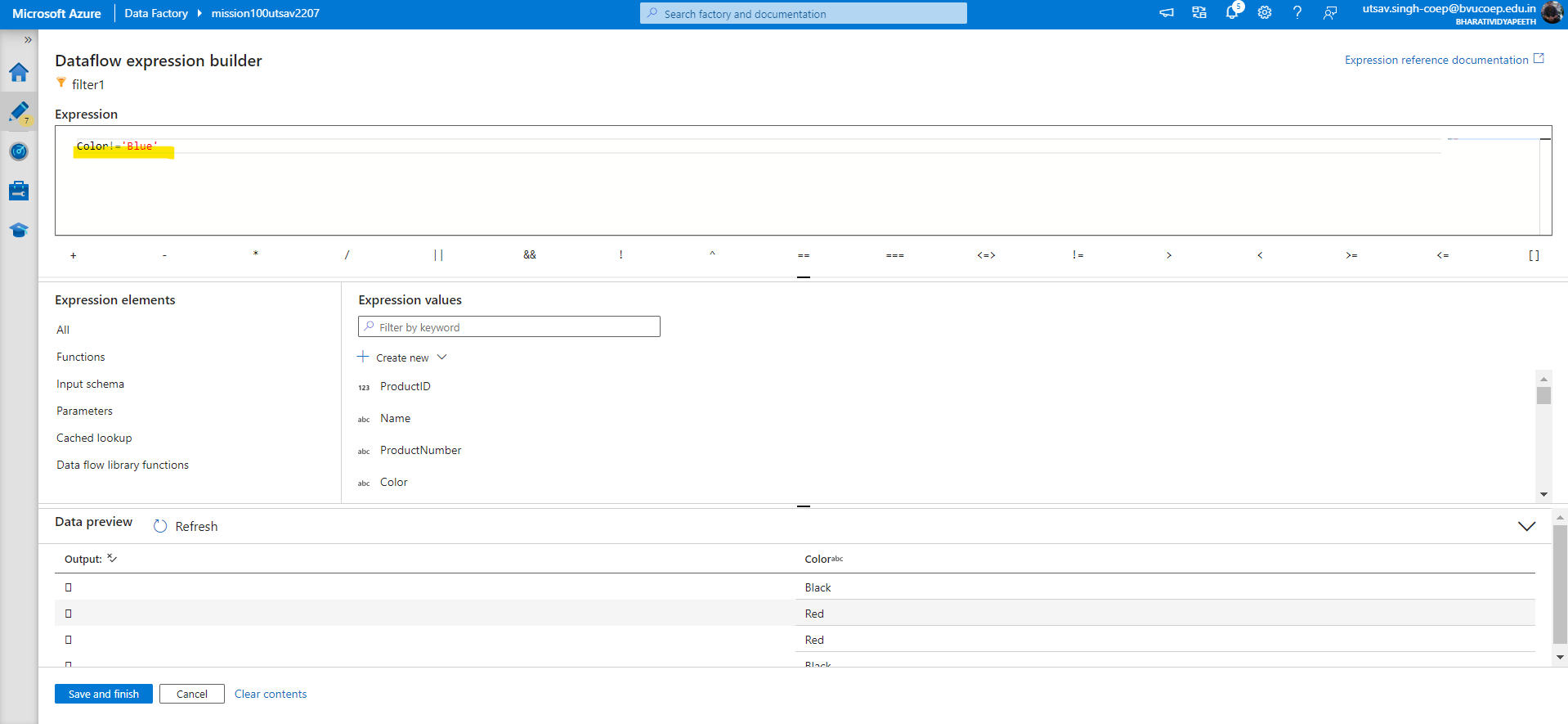
## Ensure that product shouldn't be of blue in color and save the result as CSV file inside ProductResult folder.

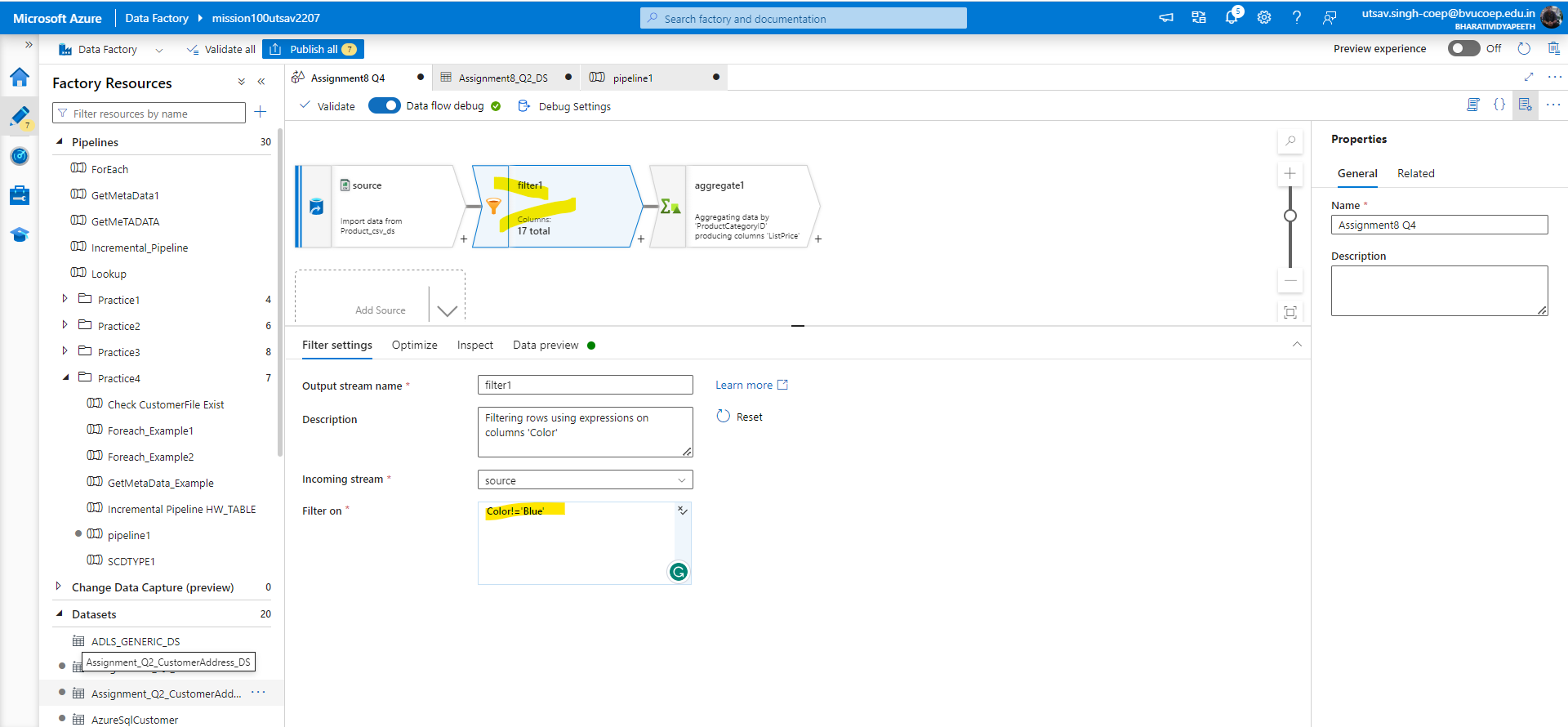
Solution:- Create dataflow and select the source as product dataset then create aggregate function for group by productcategory and fetch highest listPrice. AND create a filter condition before aggregate and in expression builder check color col- coloe!=’Blue’ then lastly use sink and save as per your desired location in any format.











## 

## Question 5: create a pipeline to read the Product CSV file, and calculate the highest listPrice of any product under each productcategory.

## Ensure that product shouldn't be of blue in color and save the result as a SINGLE CSV file inside ProductSingleResult folder.

Solution:-In this question we have to solve same as ques4 only save the result as single partition using optimize under the sink.

