Case Study

on

**Azure**

**Case study by,**

Mouli S,

Batch – 02,

moulisankar2002@outlook.com.

**Contents**

1. Understandings of Case Study ……………………………………………3
2. Case Study Questions ………………………………………………………….4-11
   1. Question 1 …………………………………………………………4
   2. Question 2 …………………………………………………………6
   3. Question 3 …………………………………………………………9

**1. Understandings of Case Study**

Coronavirus disease (COVID-19) is an infectious disease caused by the SARS-COV-2 virus.

Most people infected with the virus will experience mild to moderate respiratory illness and recover without requiring special treatment. However, some will become seriously ill and require medical attention. Older people and those with underlying medical conditions like cardiovascular disease, diabetes, chronic respiratory disease, or cancer are more likely to develop serious illness. Anyone can get sick with COVID-19 and become seriously ill or die at any age.

European Centre for Disease Prevention and Control (also known as ECDC) - an agency of the European Union tracks COVID-19 cases and vaccination status worldwide.

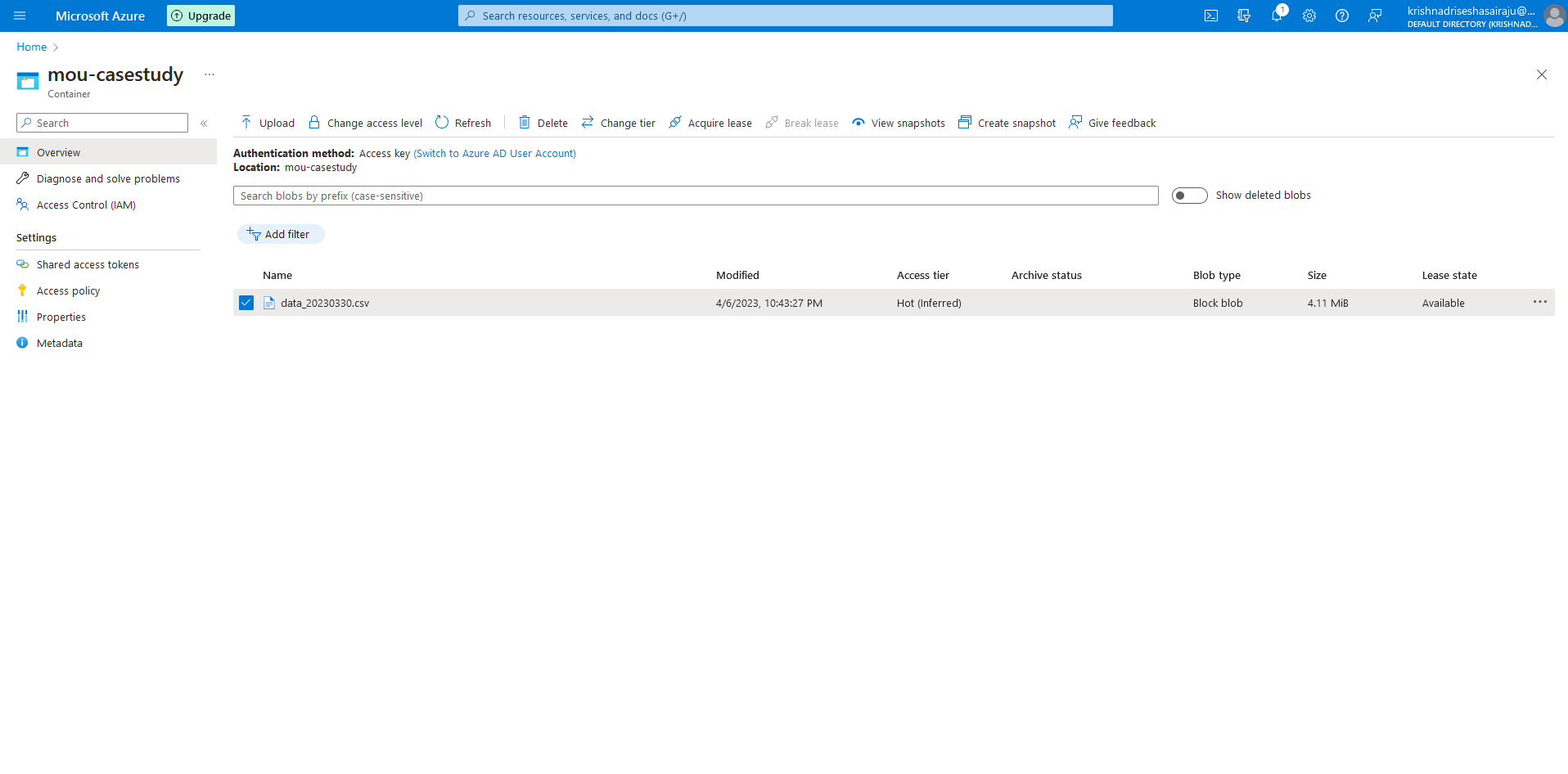
Reference Website - <https://www.ecdc.europa.eu/en/>

**2. Case Study Questions**

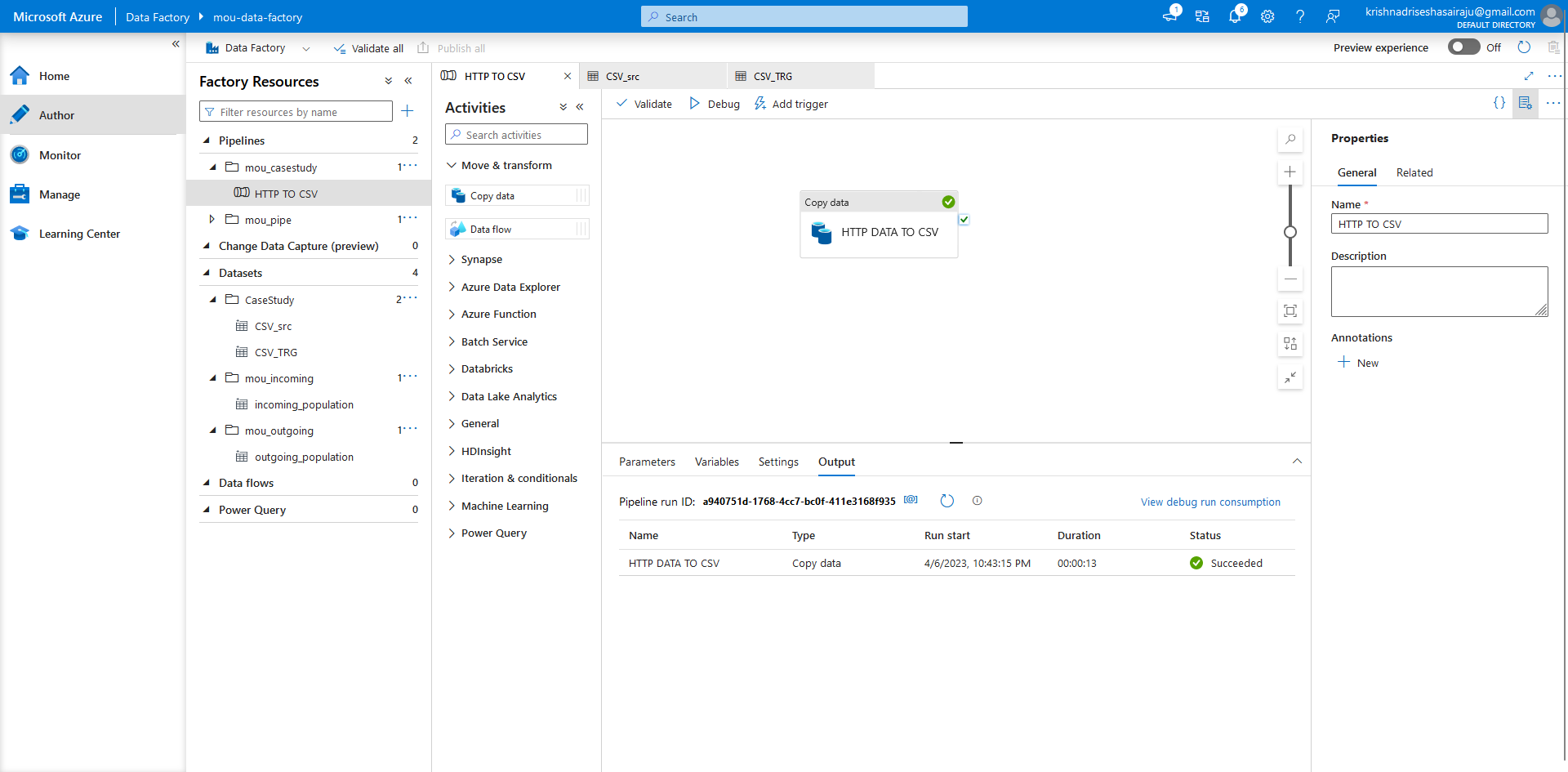
**2.1 Question 1**

**Design Job (using ADF) to read (data.csv) ECDC website (https://opendata.ecdc.europa.eu/covid19/casedistribution/) and consume data and store in Blob storage in Azure and rename with by appending timestamp like data\_20230330.csv.**

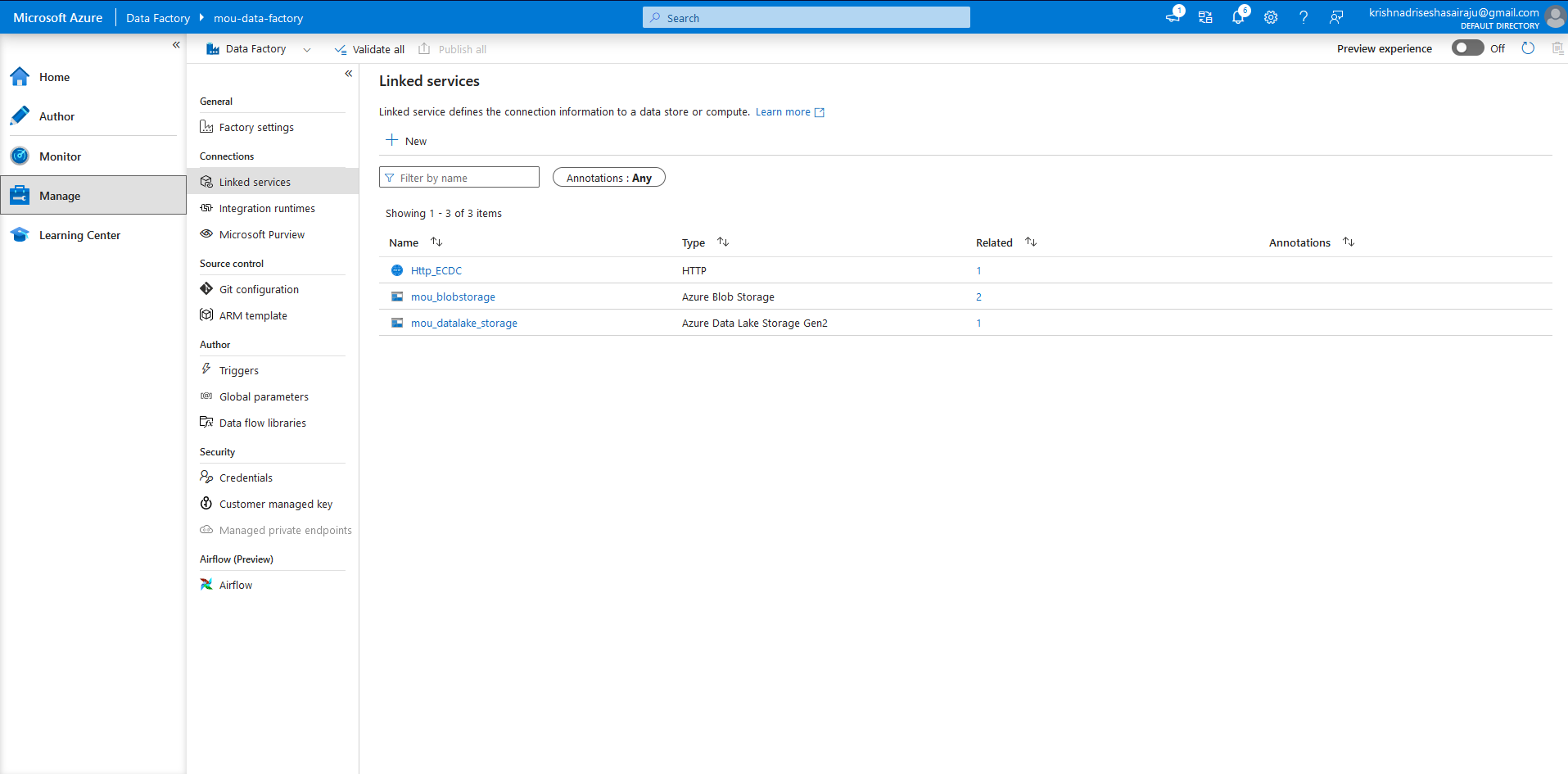
**File loaded in Blob storage:**

****

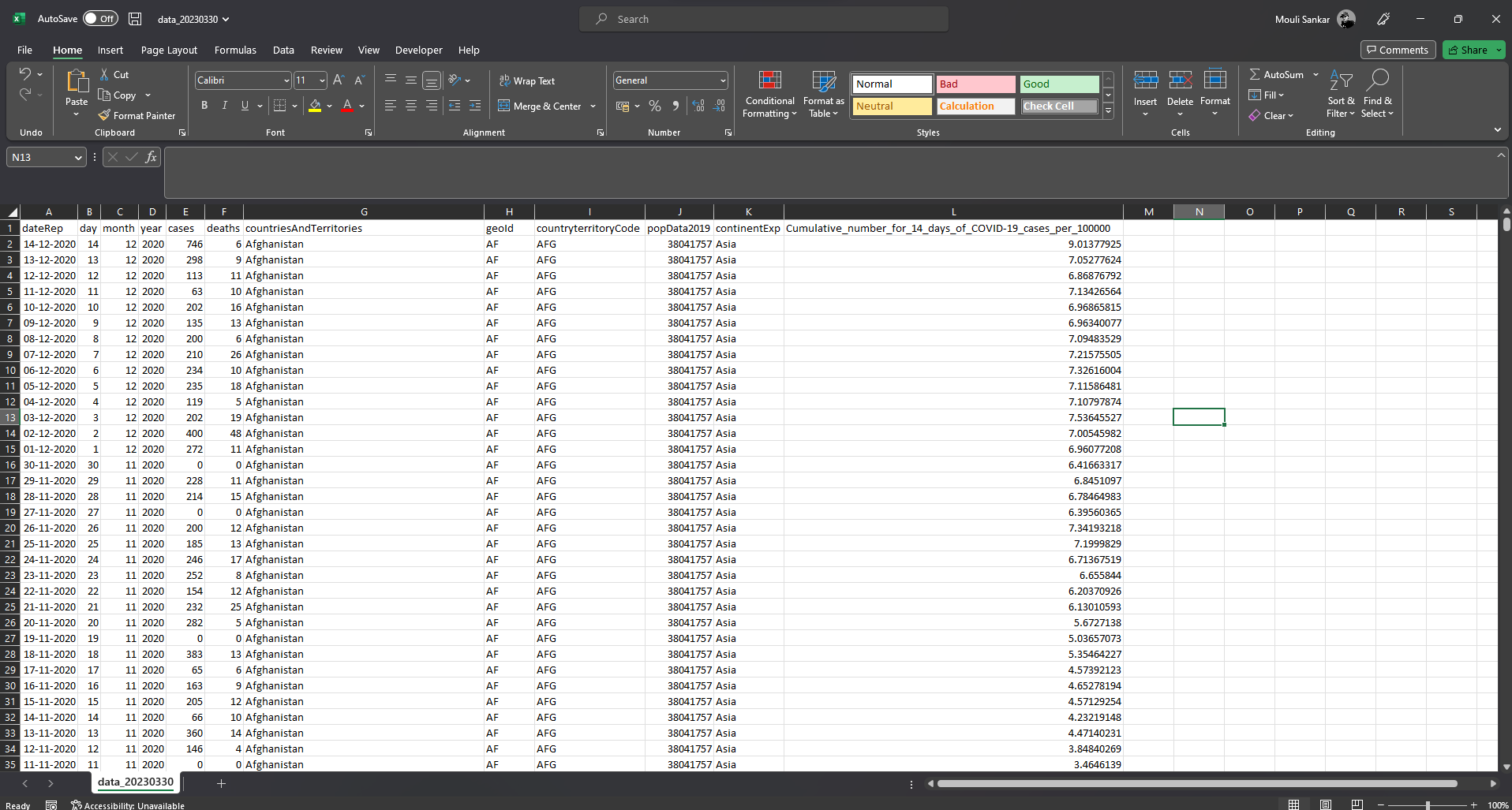
**Pipeline created for the ECDC Website to blob .csv file:**

****

**Linked services required for the pipeline and job run:**

****

**Output file in the format of CSV:**

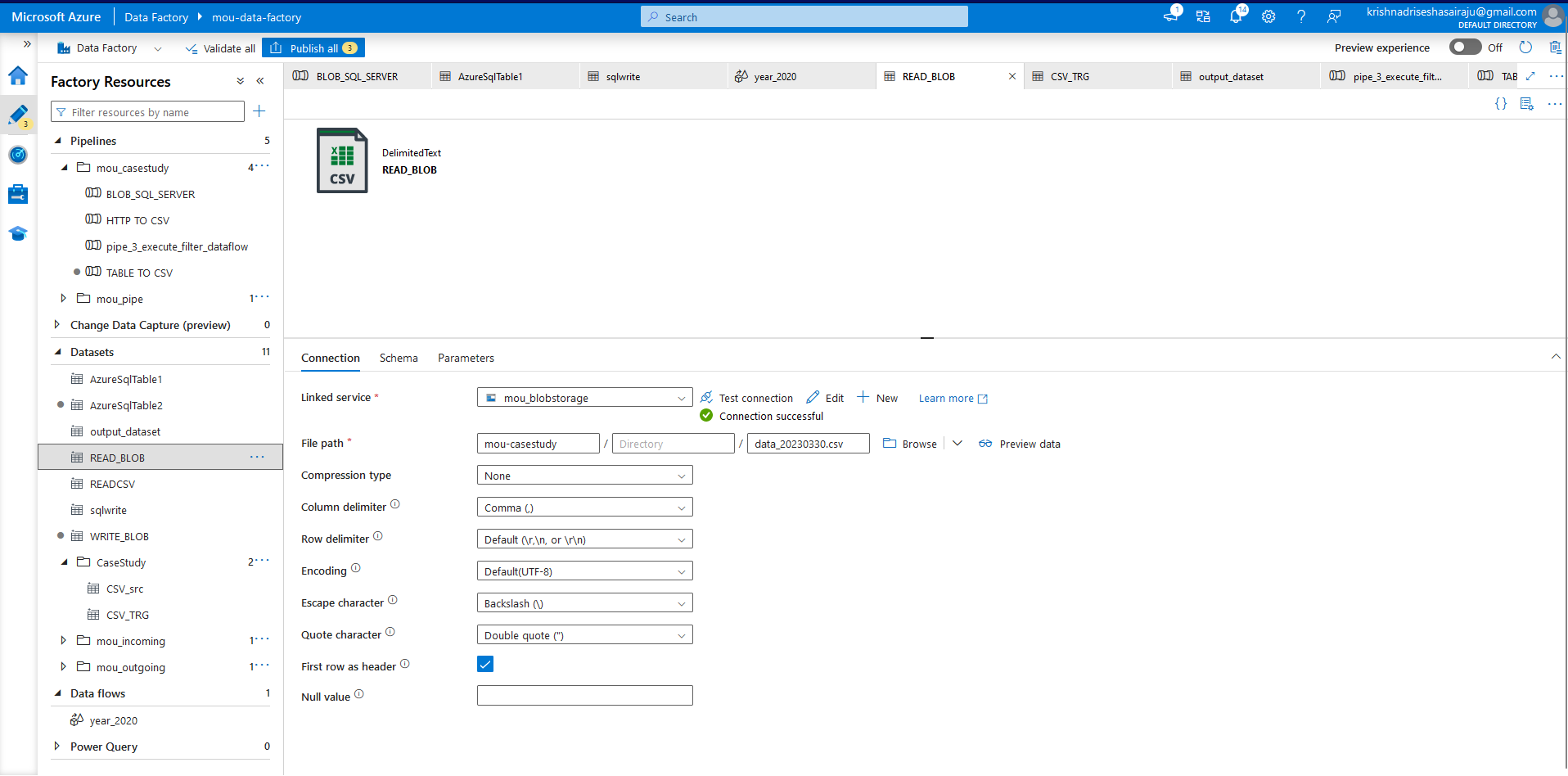
****

**2.2 Question 2**

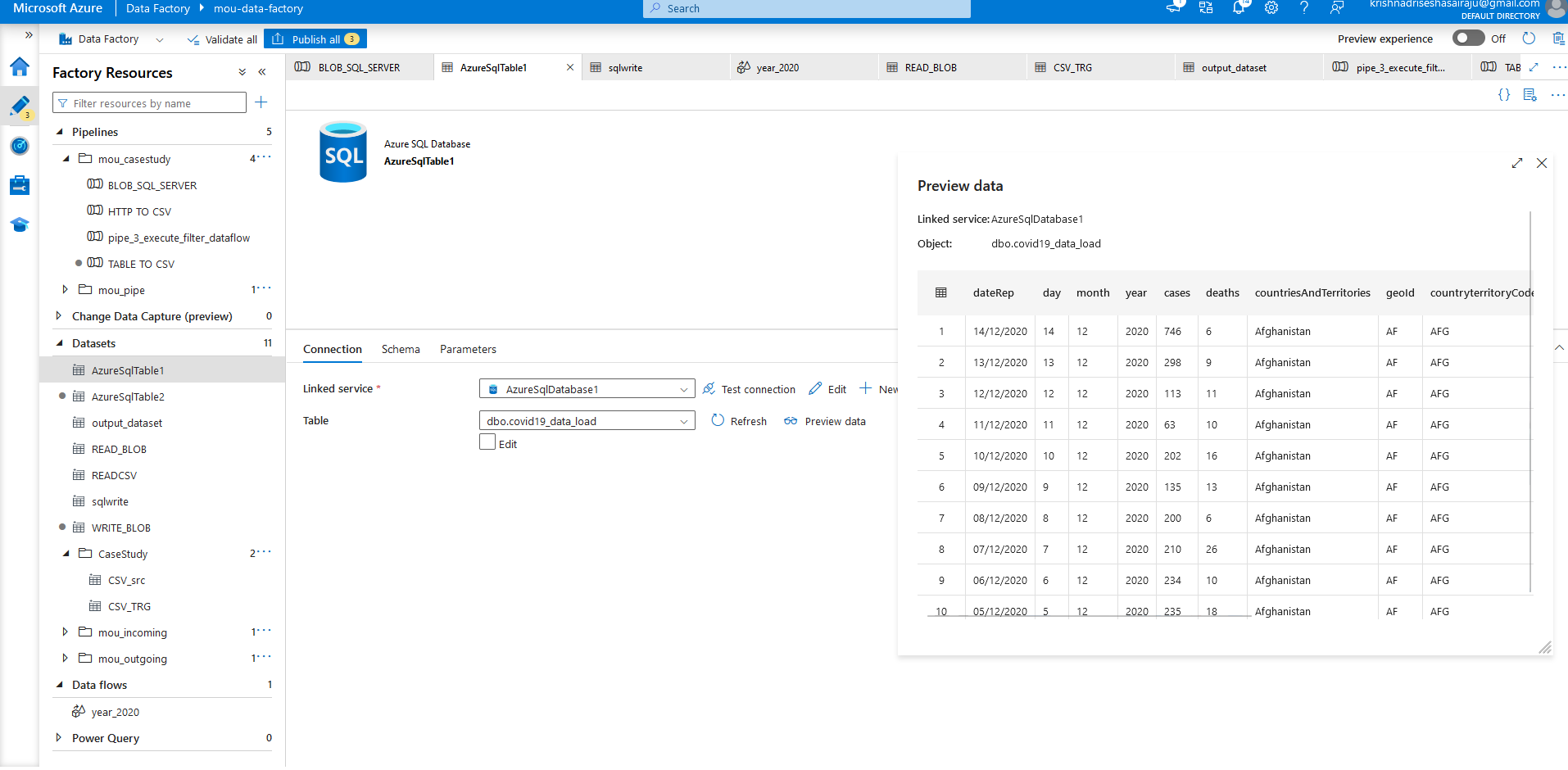
**Design Job (using ADF) to read the file (data\_20230330.csv) from blob storage and load into SQL Server.**

**List of Datasets:**

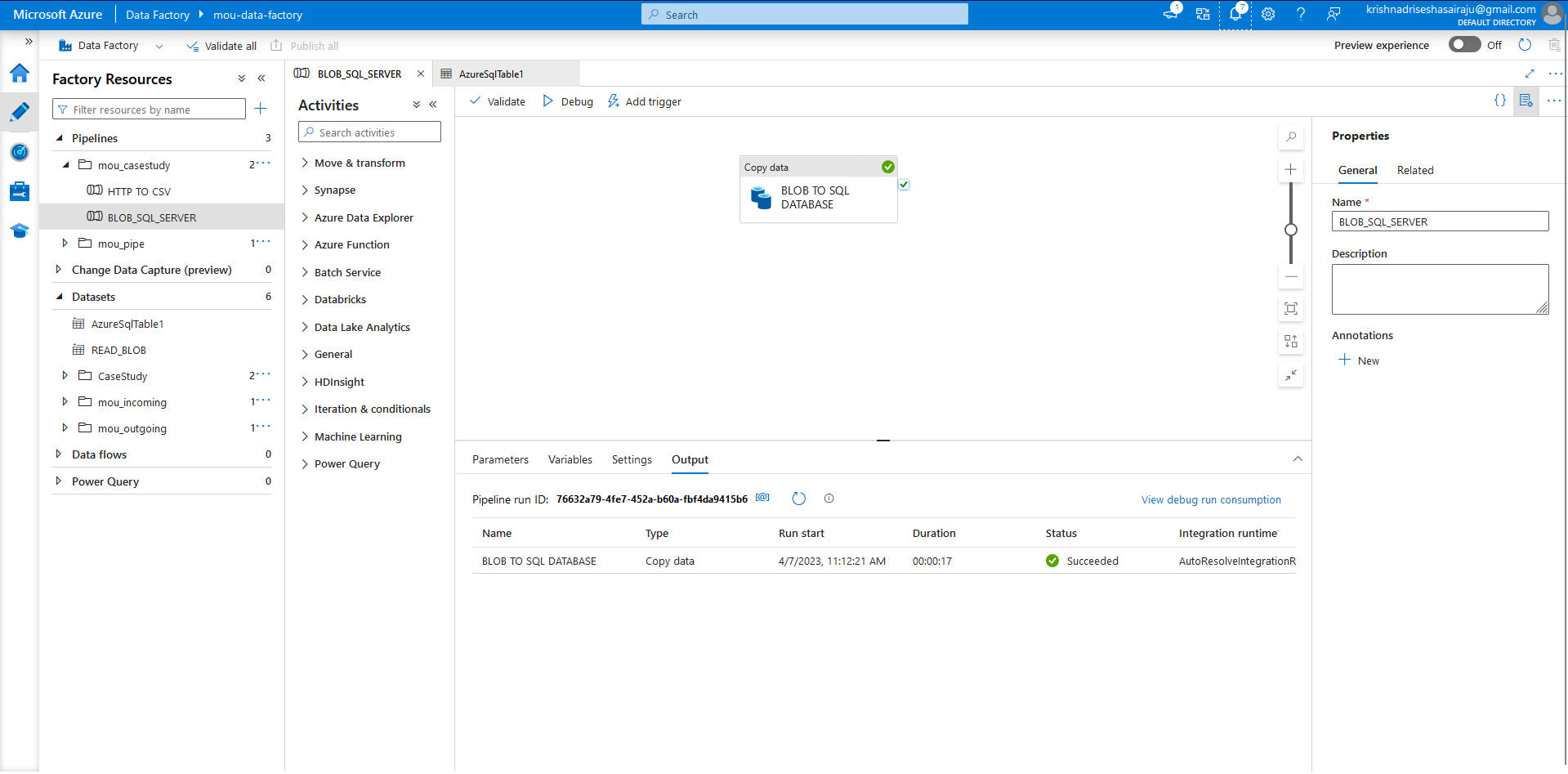
**1.**

****

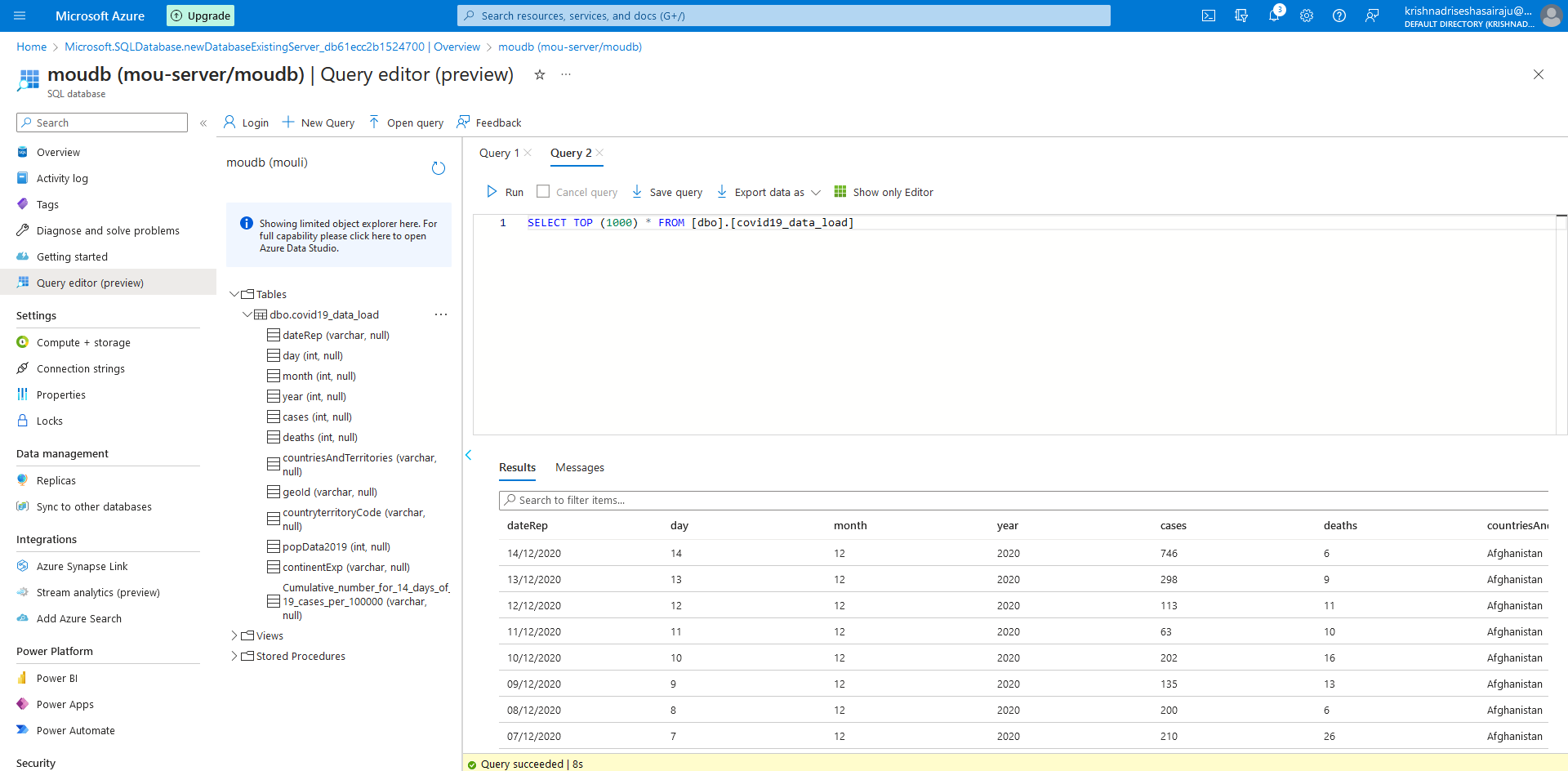
**2.**

****

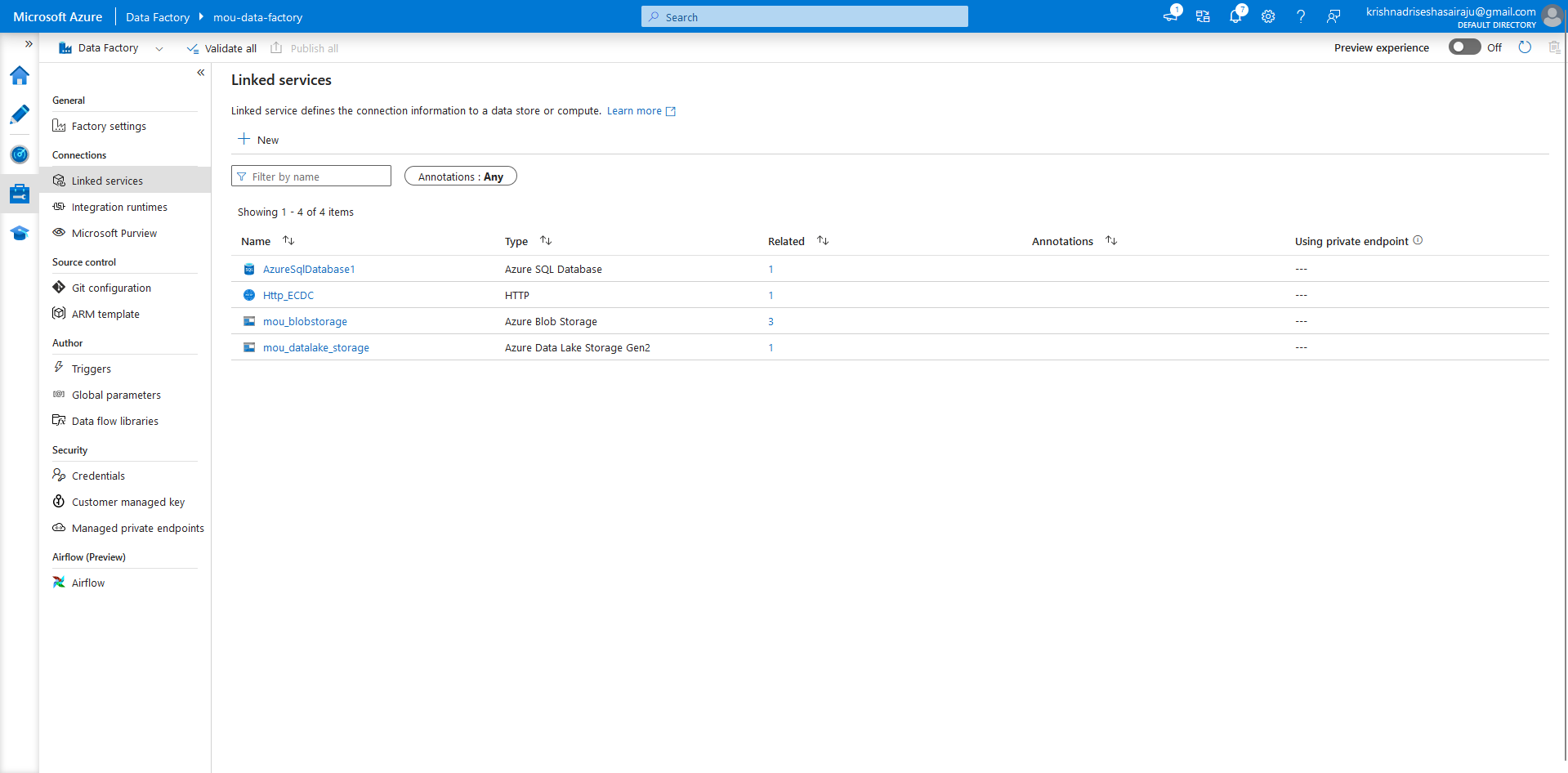
**Pipe line to read csv file from blob and load to SQL Database:**

****

**Loaded csv file queried in SQL Database:**

****

**List of linked services:**

****

**2.3 Question 3**

**Design a Job (using ADF) to summarize data for all the countries monthly for year 2020 in a table and store a copy in blob storage too having timestamp field like (out\_data\_20230330.csv). Output table and file will consist of following columns:**

**a. Continent (e.g... Asia)**

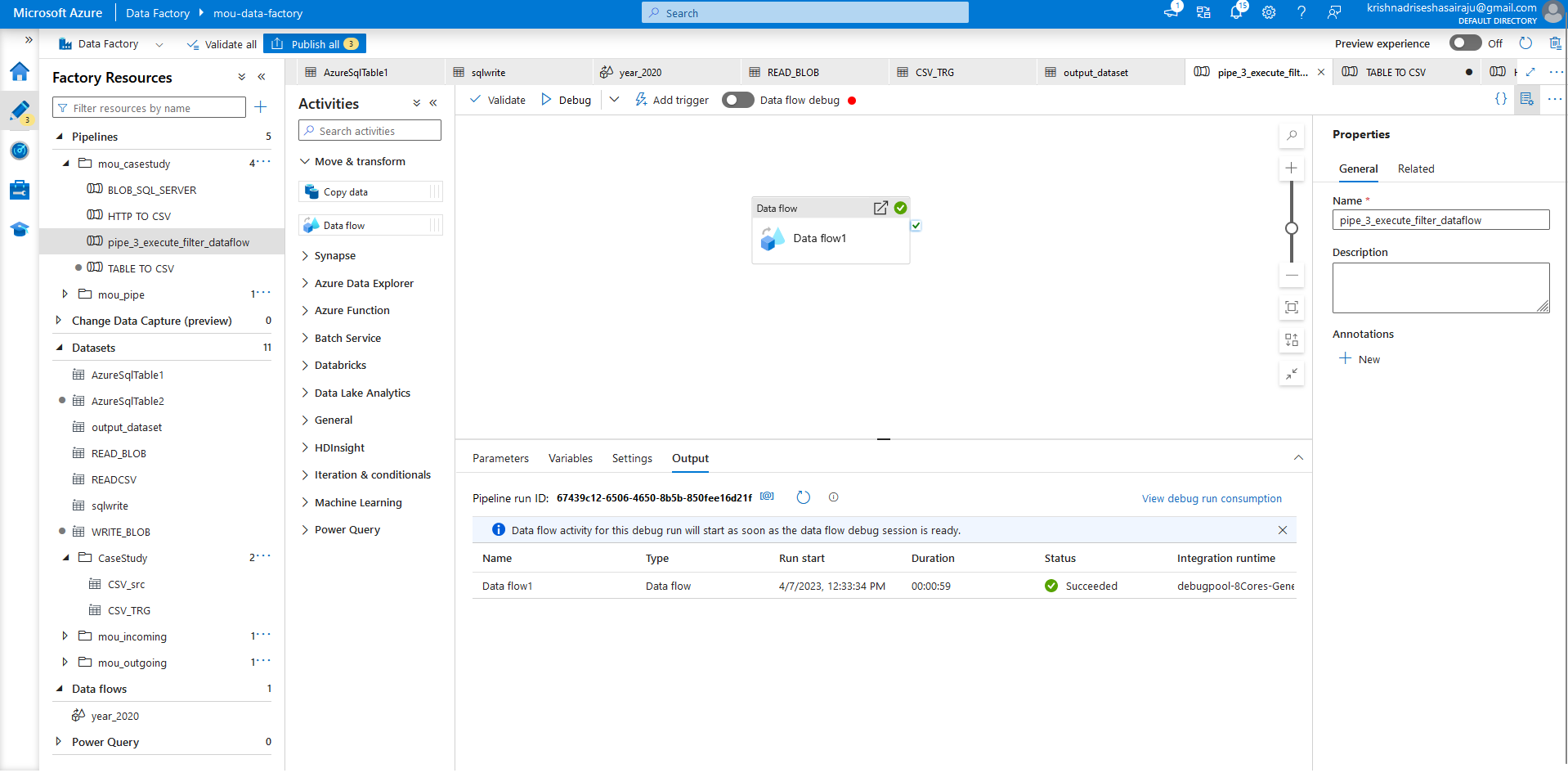
**b. Countries (e.g... Afghanistan)**

**c. Year Month (e.g., 2020-01)**

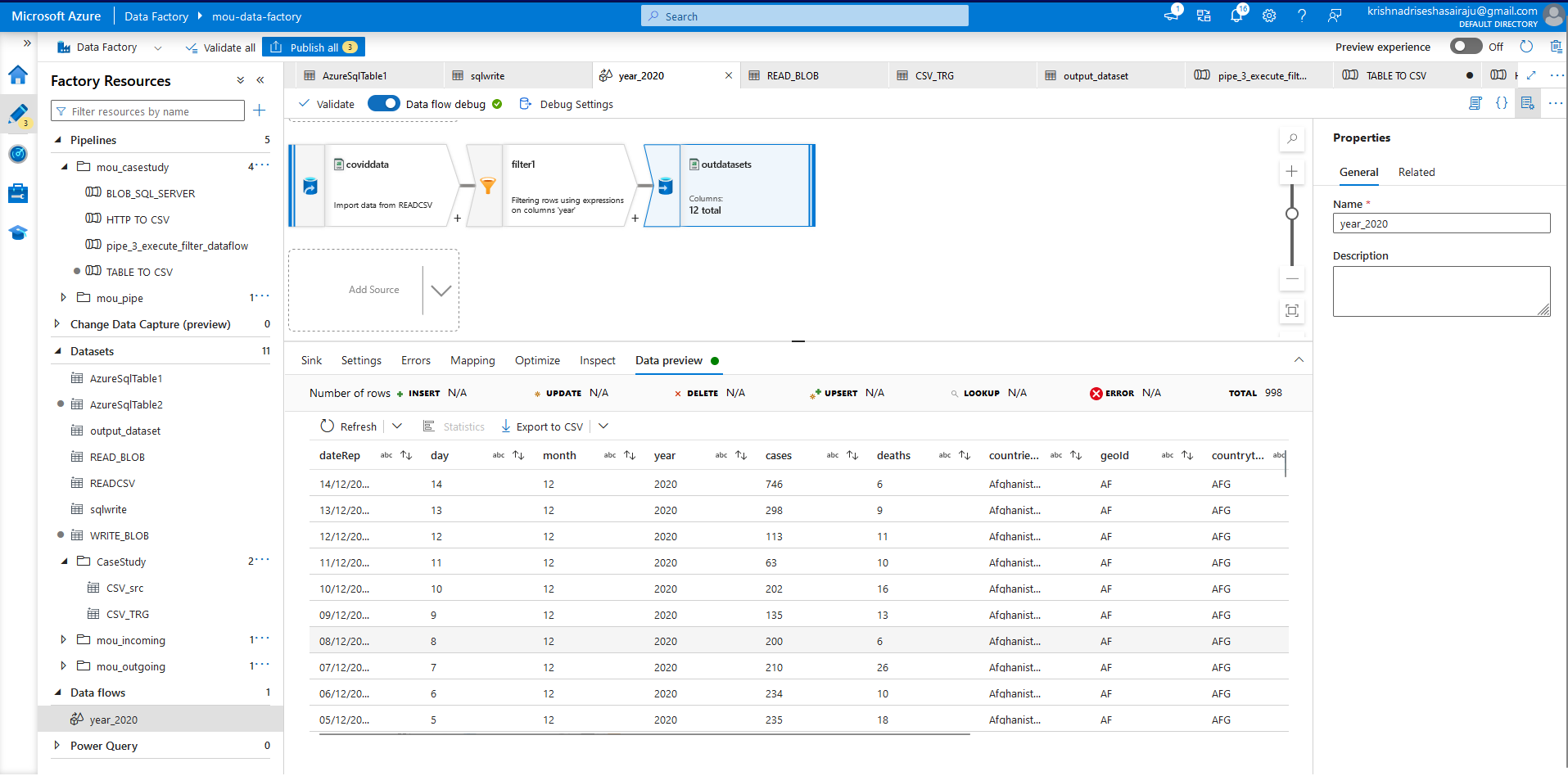
**d. Cases (e.g., 0)**

**e. Deaths (e.g., 0)**

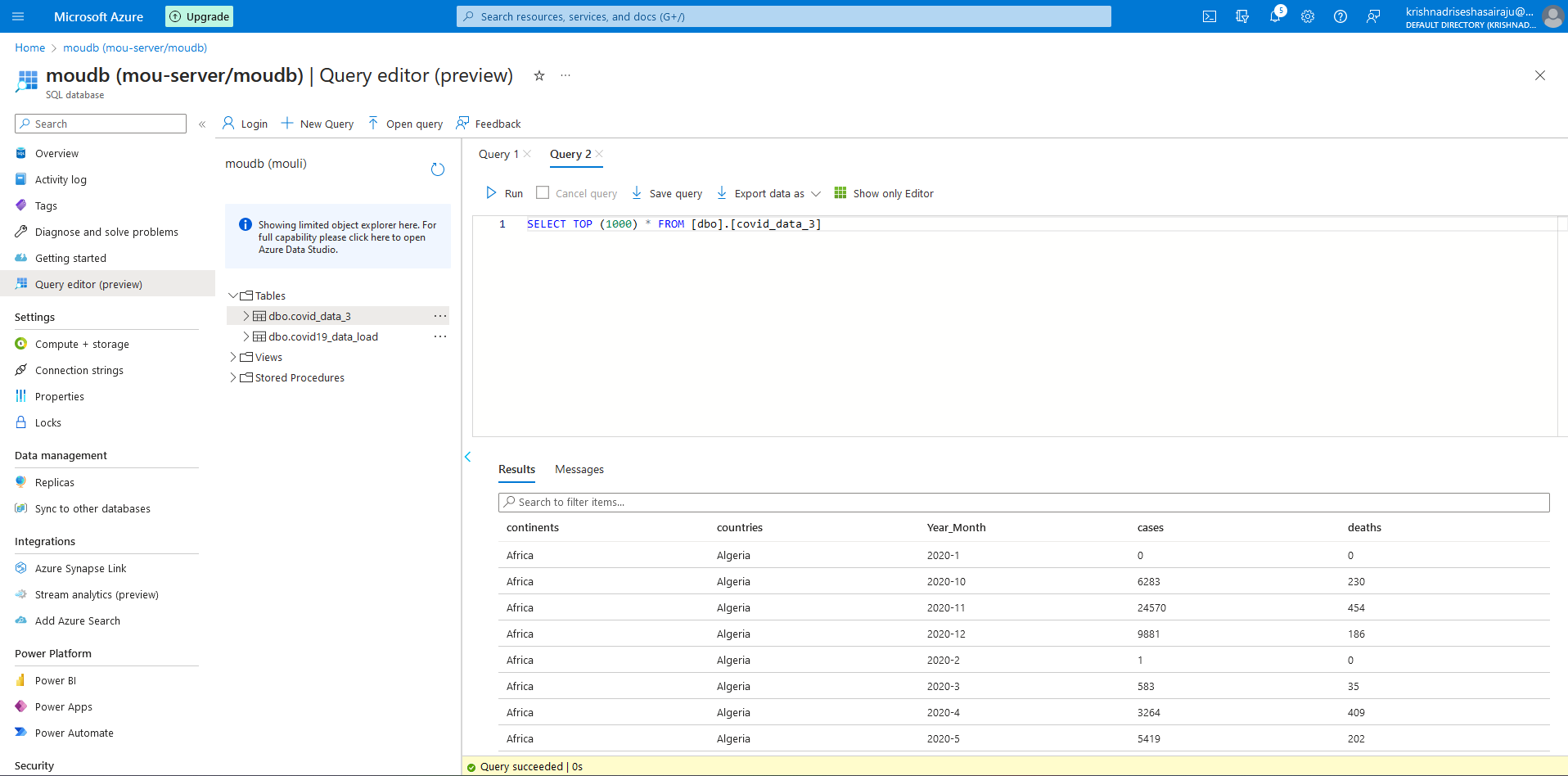
**Pipeline for the Dataflow:**

****

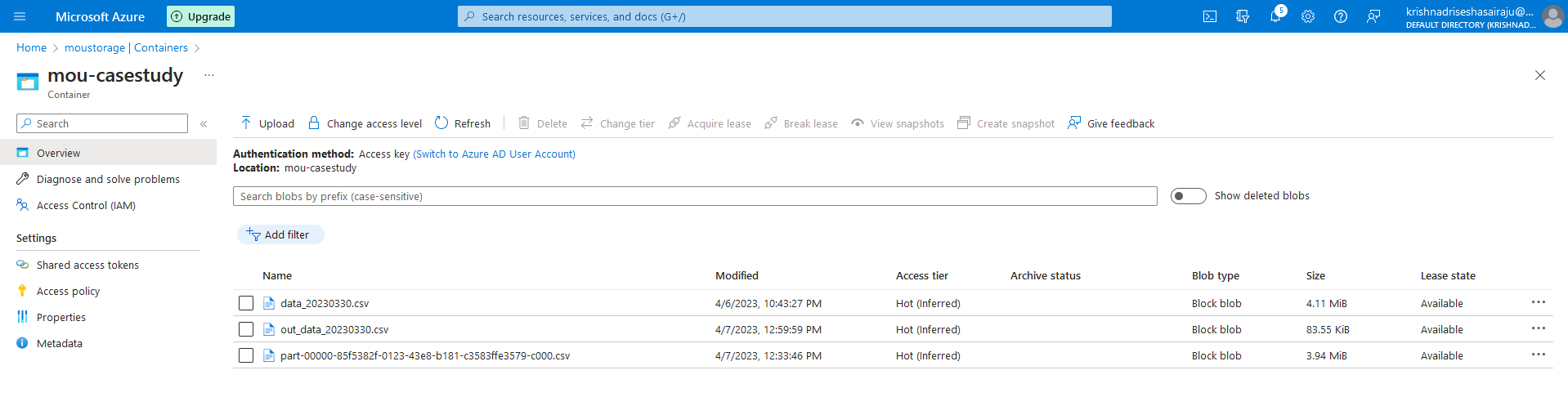
**Data flow creation for filter year 2020:**

****

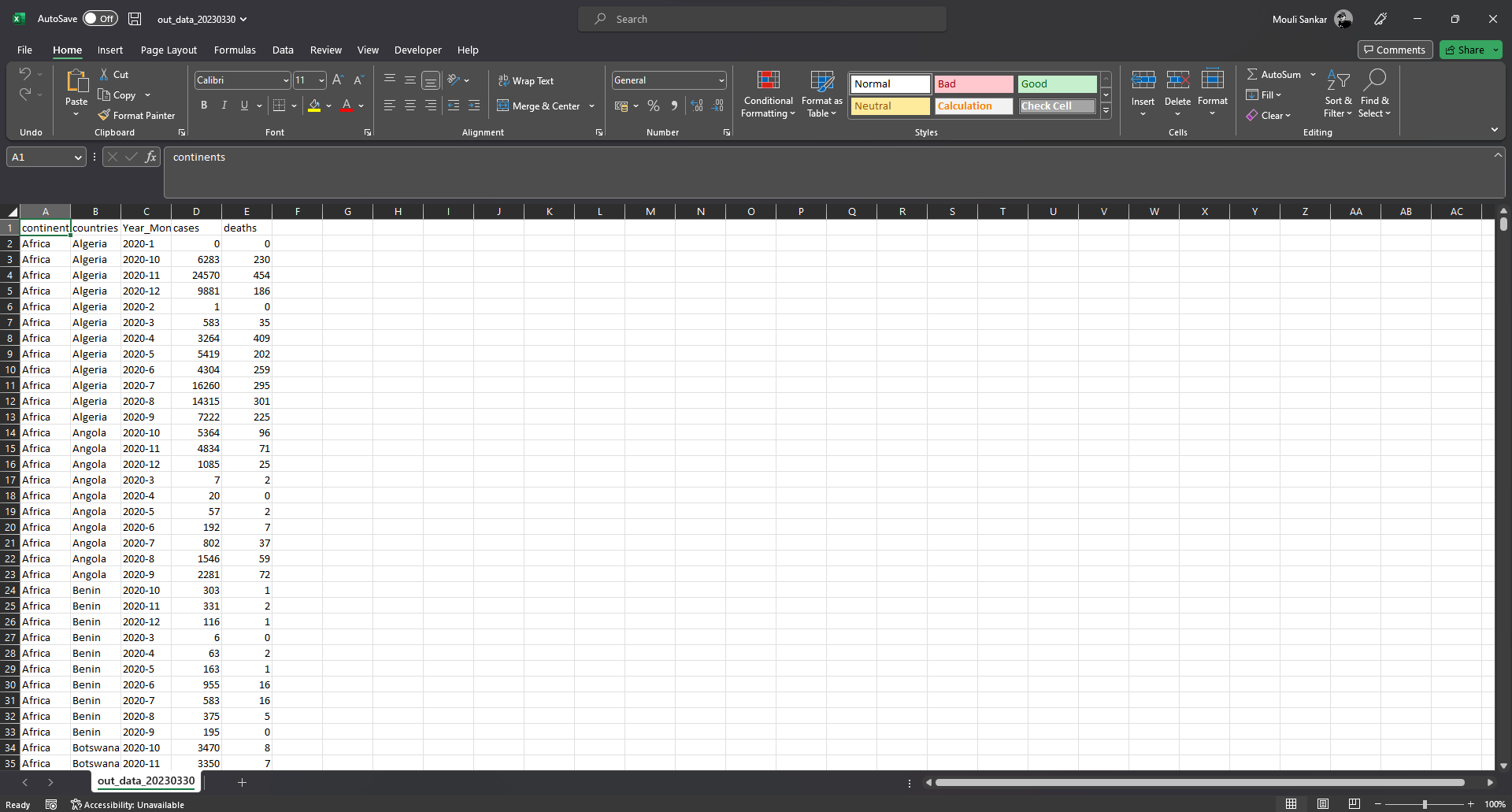
**Loaded data in SQL Database‘covid\_data\_3’**

****

**Generated file in Blob Storage:**

****

**Output file in the CSV Format:**

****