**Title**- Big data Analytics of Healthcare-Stroke Data and Visualization using Microsoft Azure Databricks and PowerBI.

**Cloud Technologies Used**: Microsoft Azure Storage Gen2/ Azure BLOB Storage, Azure data bricks, Azure Data Factory (Thought of using it for designing a pipeline), Microsoft PowerBI.

**Team information**: members, how the collaboration has been done (if applicable):

* **Dharani Kurra**: Responsible for data collecting and gathering more data from other sources possible.
* **Krupa Sagar Reddy**: Responsible for platform and storage management on Azure.
* **Veera Poorna Prakash Reddy Chekkera**: Responsible for Data analytics and model building(optional).
* **Rajashekar Reddy Devireddy**: Responsible for visualization using PowerBI.

**Things that have been tried/accomplished**:

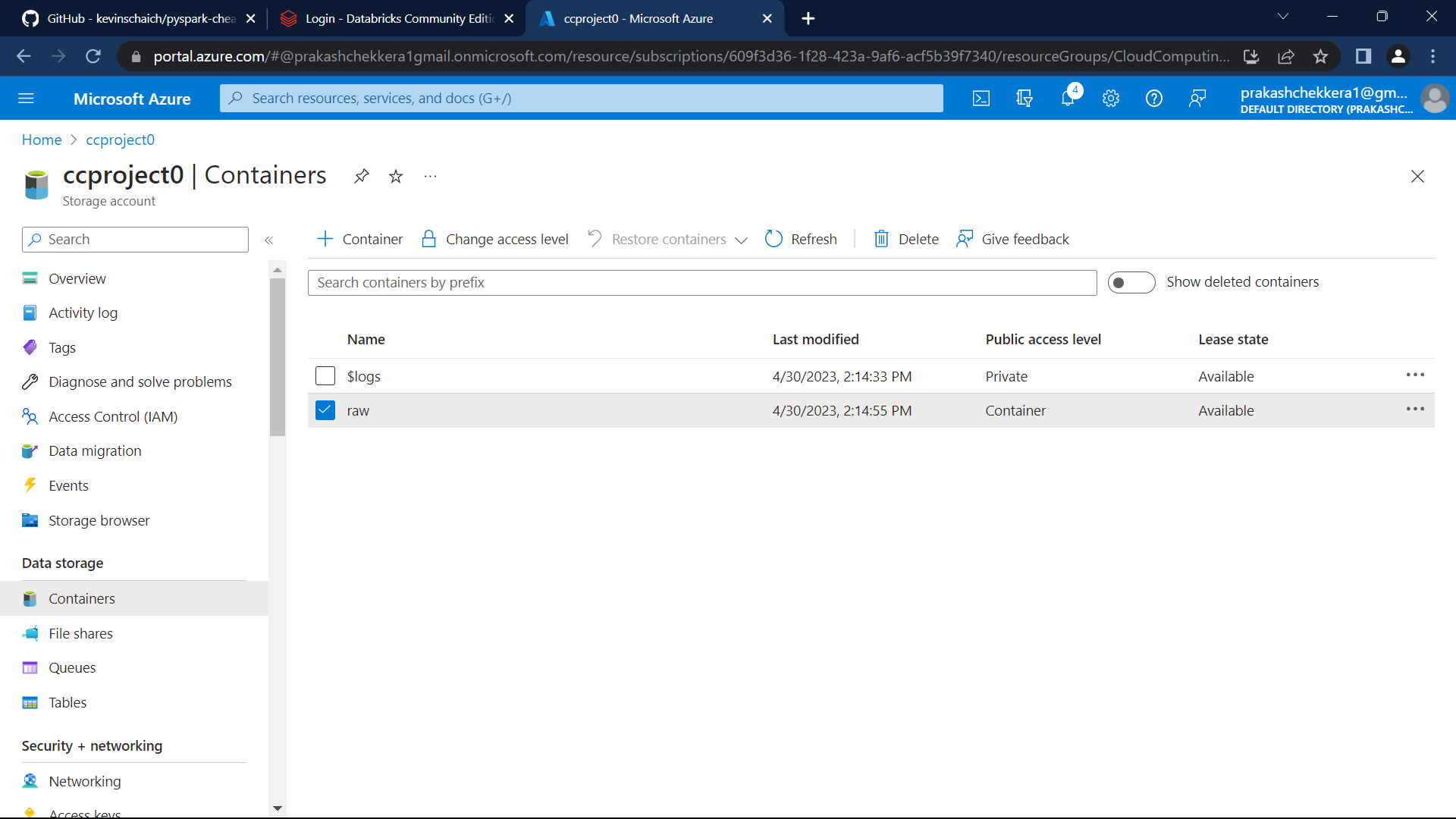
* We have successfully collected data from Kaggle and looking actively to collect more data from the internet as well.
* We have successfully created a storage account on Azure Data Lake Gen2 and Azure BLOB Storage.
* We have performed some initial analysis to identify relations among the features in the data by mounting the data storage onto Databricks platform on Azure.

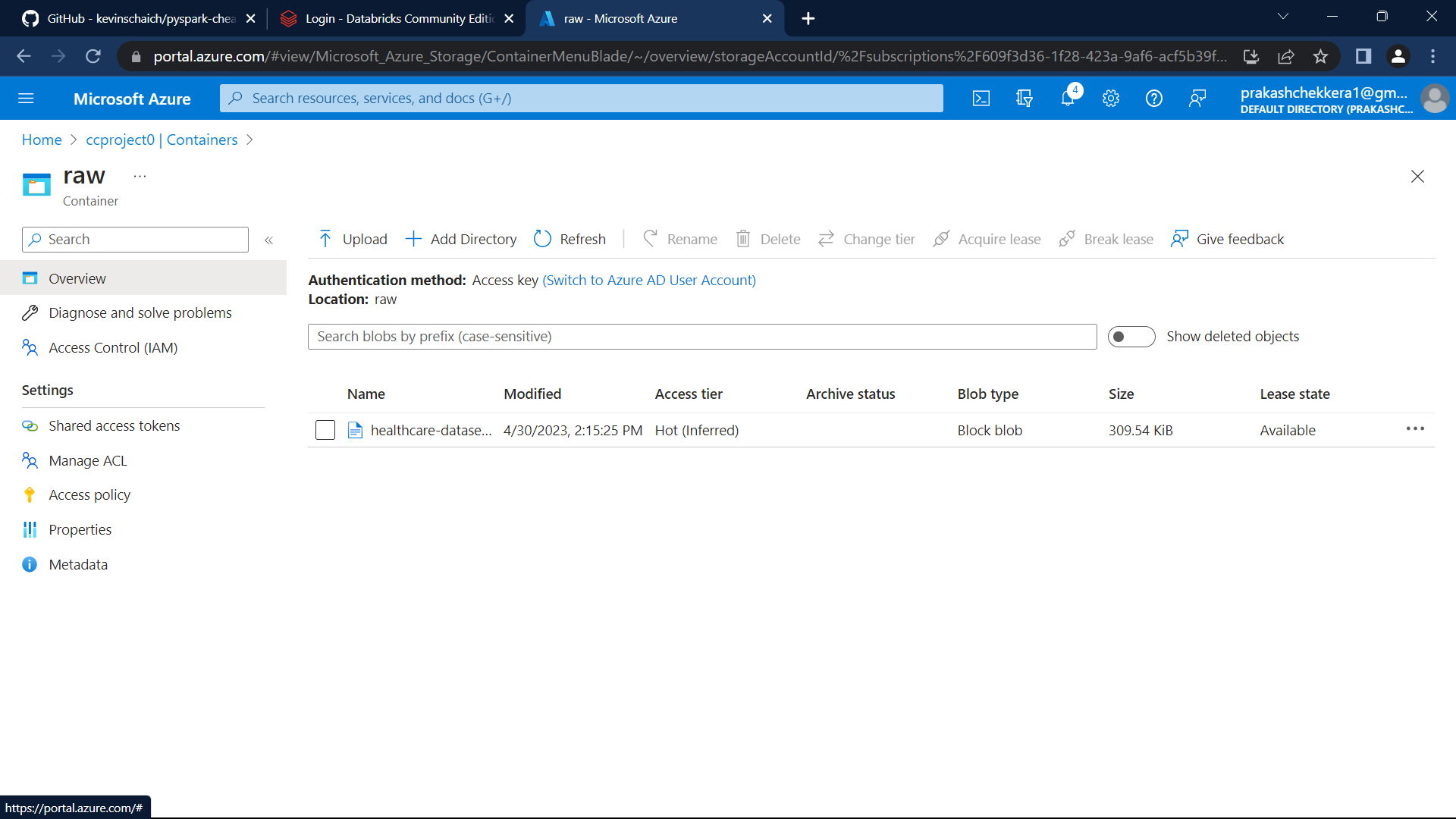
**Things to do**:

* We are yet to gather more data from the internet.
* We are yet to design a pipeline on Azure Data factory to automate the entire process.
* We are yet to perform further analysis on the gathered data to derive some more insights from it and optionally we have decided to build a ML model to predict the chances of a stroke.
* We are yet to perform beautiful visualizations on the cleaned final data on PowerBI.

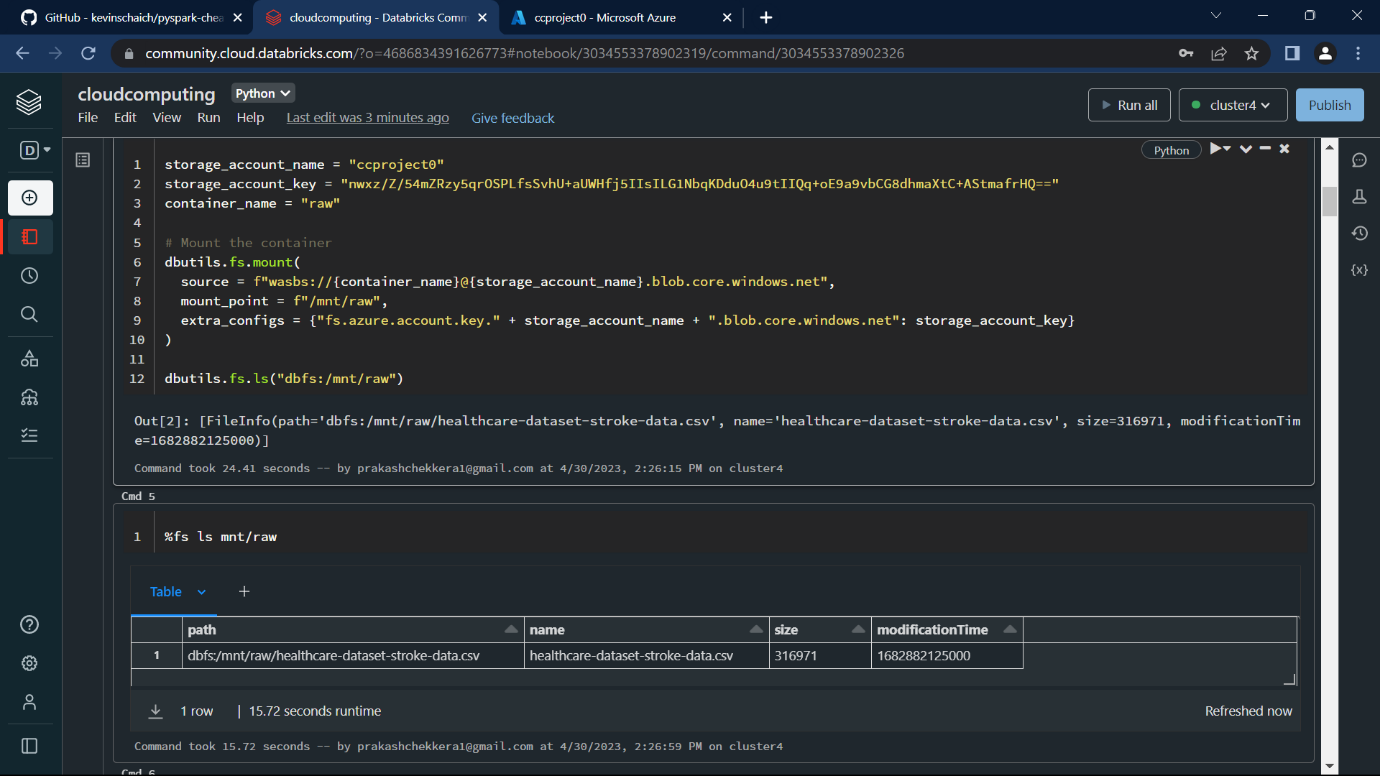
**Challenges/Comments:** So far, the entire process is smooth and we are trying our best to automate the entire process using Azure Data Factory. One challenge could be knowing how to design and implement a pipeline on ADF. Gathering more relevant data from the internet is the other challenge for us in this project. We are trying by all means to derive more insights from the huge dataset so as to derive at the general trends about reasons for the stroke in the people. The prior knowledge about the tools of ADF, Azure Data bricks, PowerBI is needed here to pace up the project.

Below are some of the snapshots for the project we accomplished till now:





Mounting is done as



We have successfully read data from mount point.

