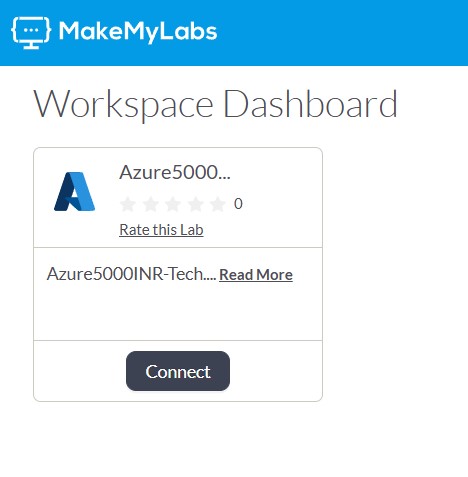
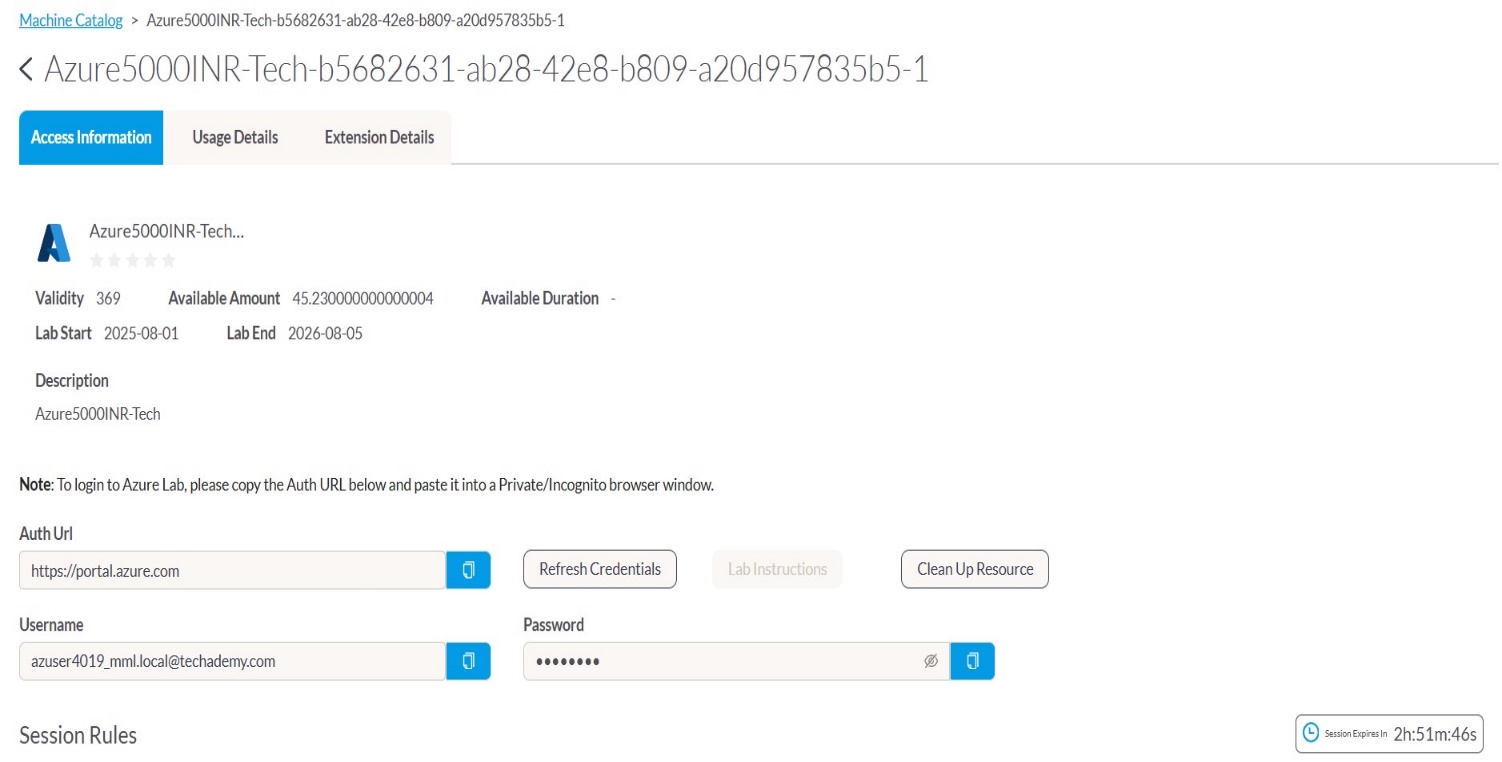
**AZURE ASSIGNMENT- 1**

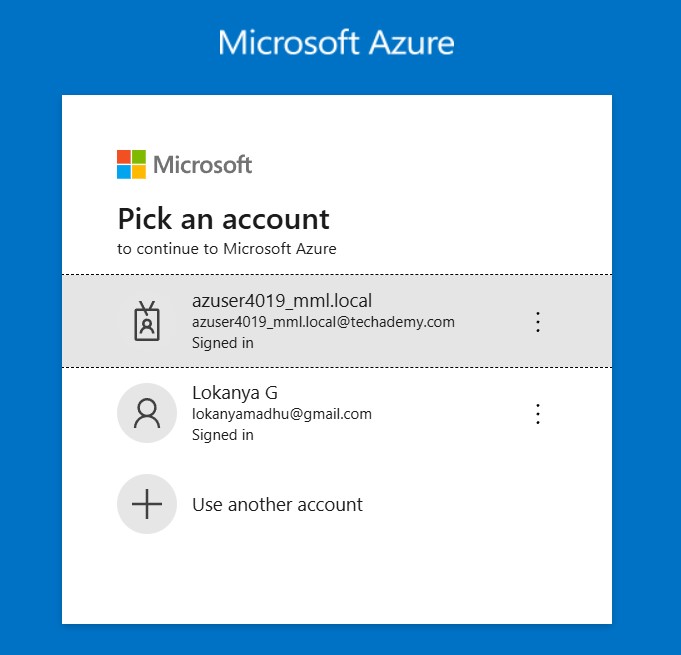
**STEPS TO USE AZURE WORKSPACE**

Copy the link provided in the mail and paste the url in chrome and open the page, give the user credentials provided in the mail and login.



Click on connect and the page will take you to another setup in makemylabs.

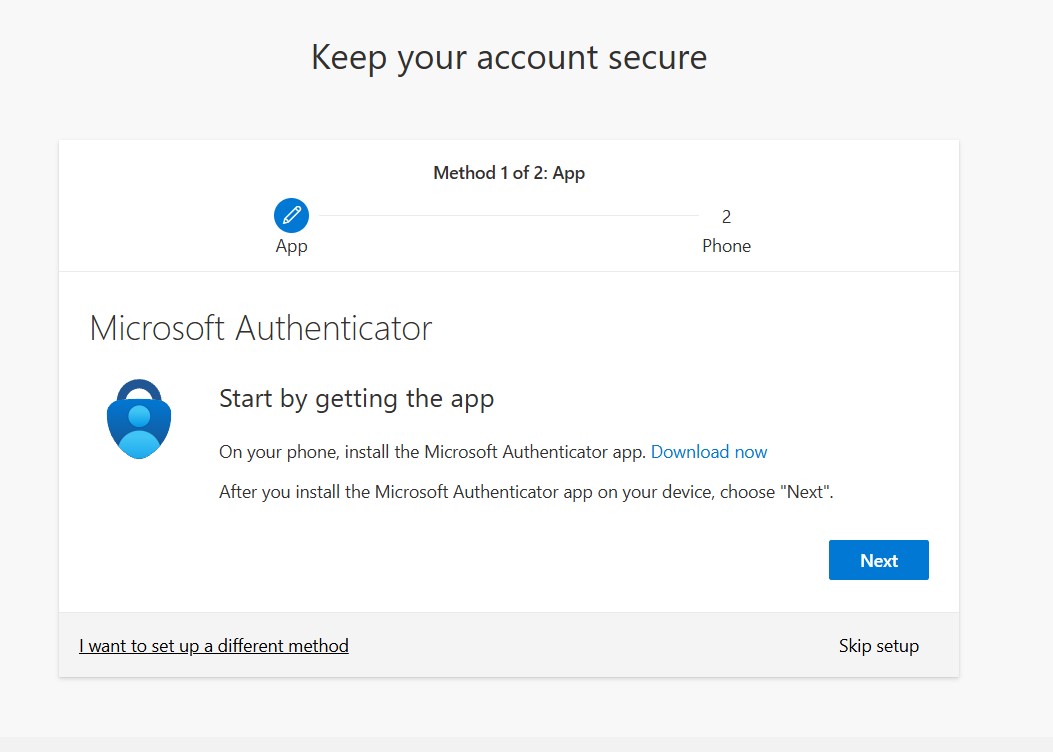
Copy and paste the Auth url in chrome to login in Azure.



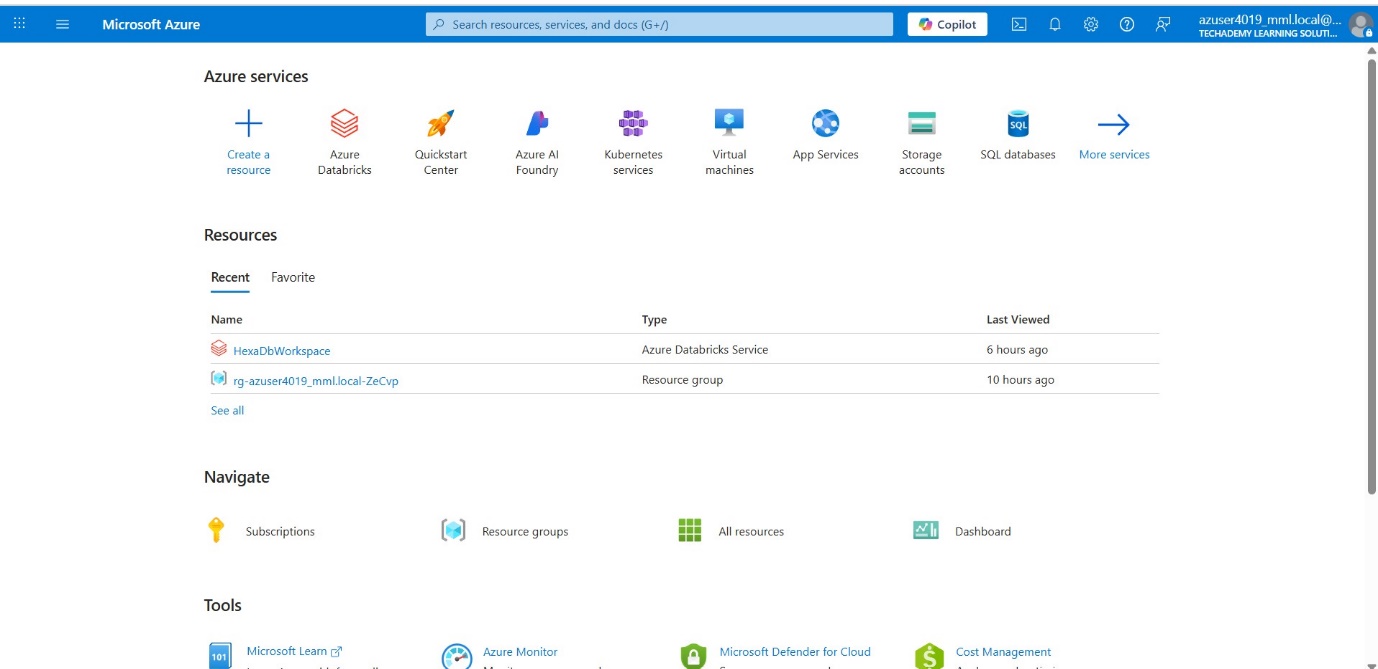
Click on the first azure account provided by Hexaware to continue.



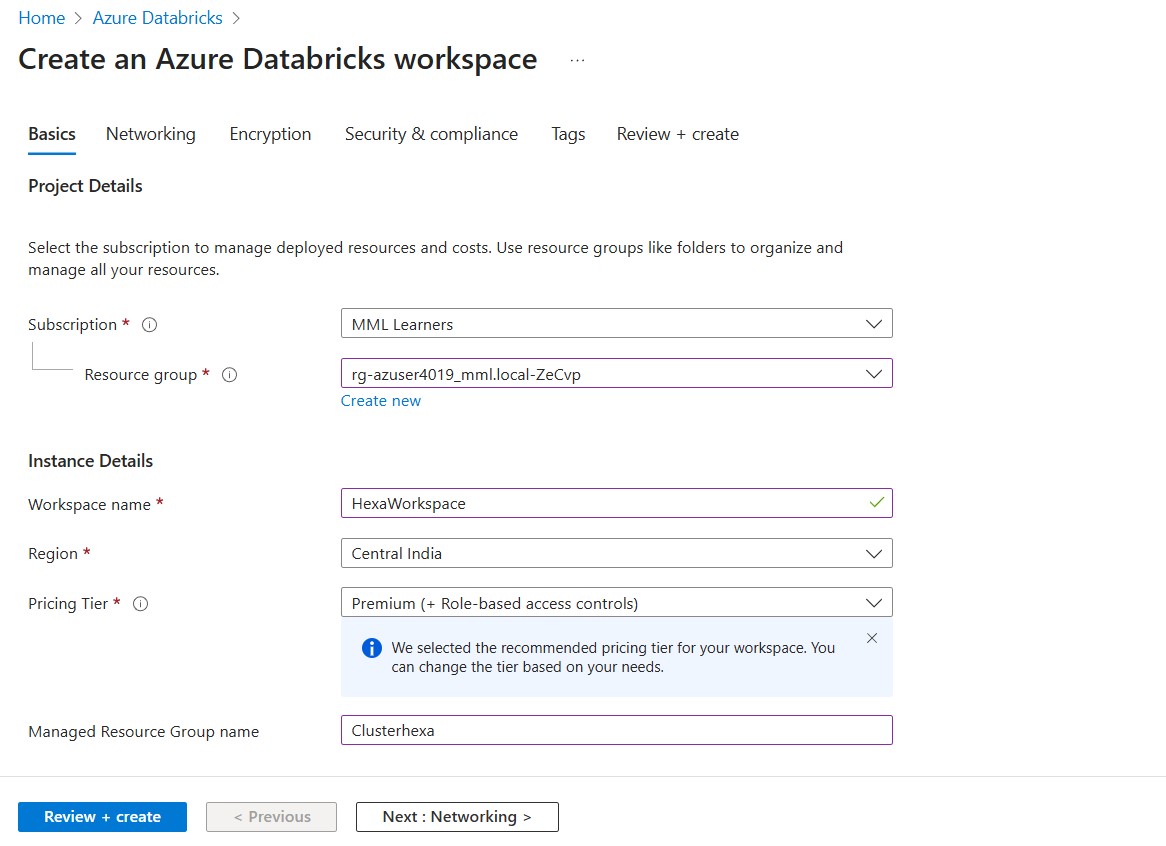
Click on Next to proceed.



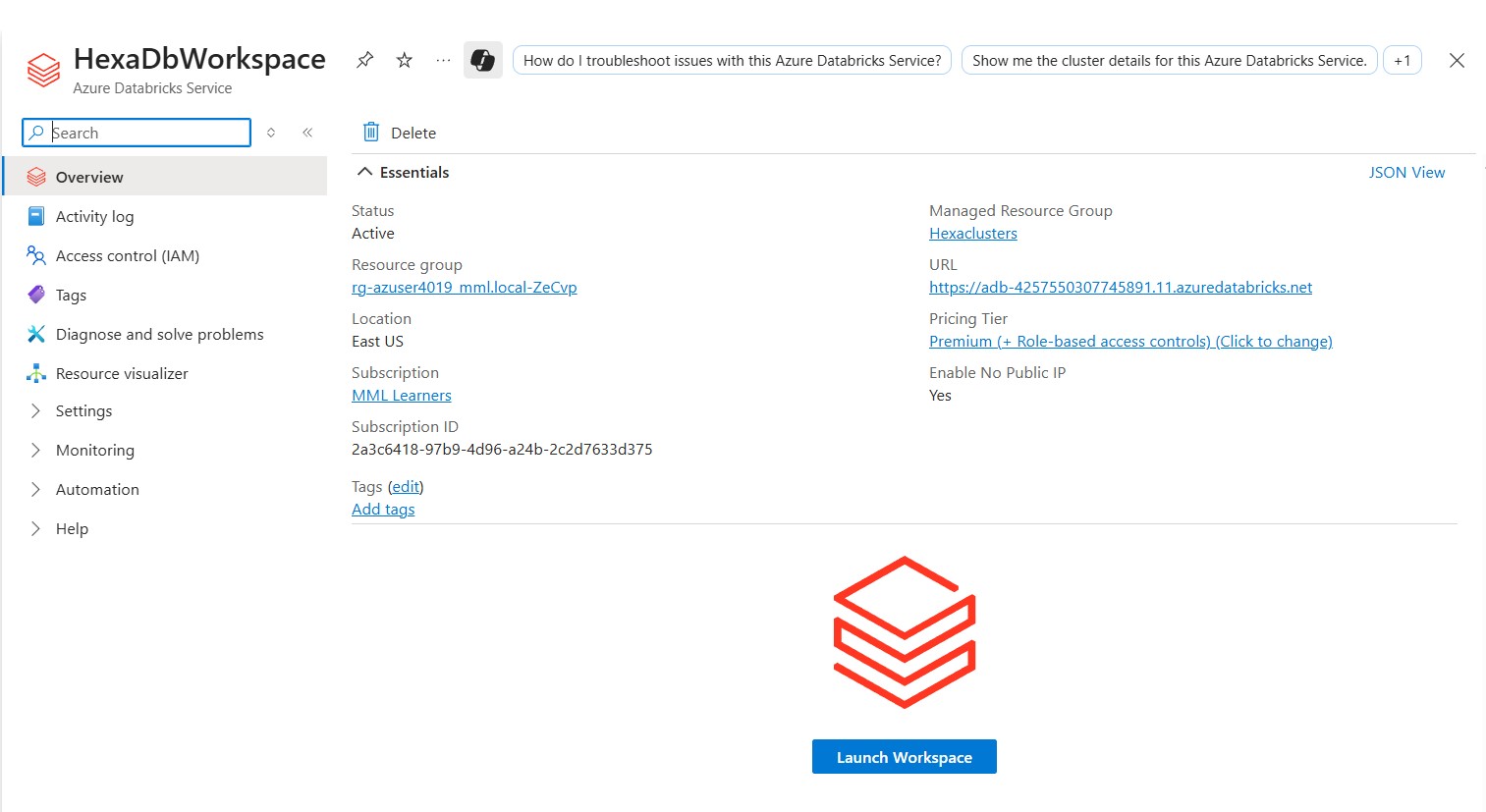
Click on skip setup ,inorder to avoid verification everytime you try to login inside the portal.



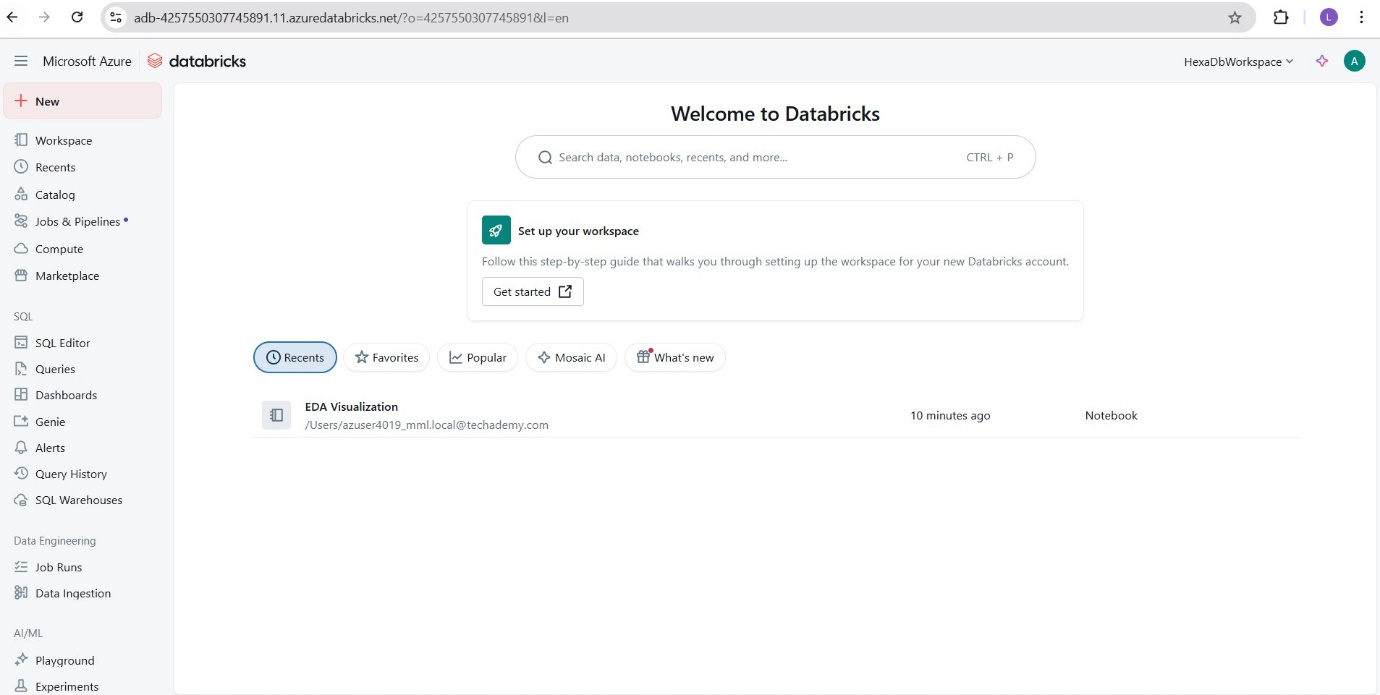
Now when you get inside the Azure page, Click on Azure databricks to do the initial operations.



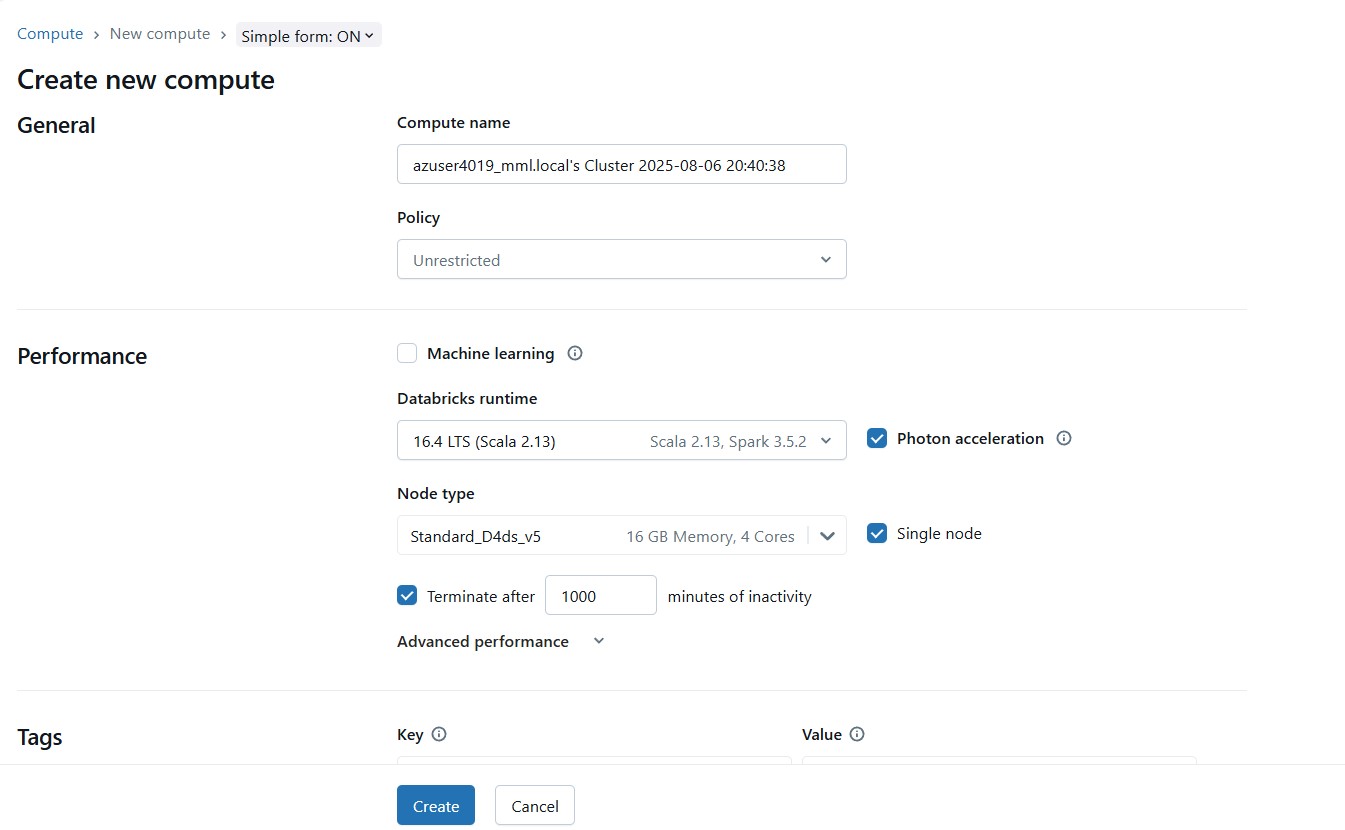
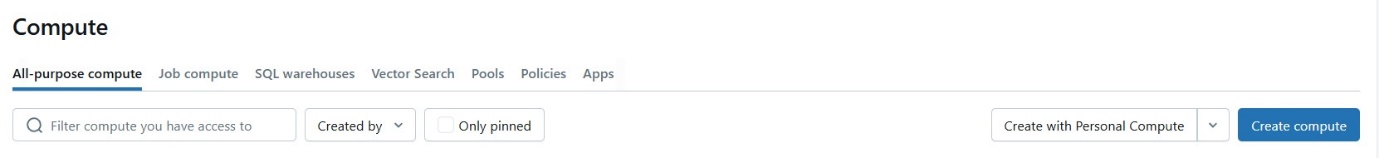
Click on create workspace in the Azure databricks page and give the required credentials like the workspace name and Managed resource group name ,finally click on Review+ reate to create a new databricks workspace in Azure.



Now click on launch Workspace and go to the databricks page.



Click on Compute present in the left corner and create a new Compute.



In the new Compute page just change the Databricks runtime to the latest LTS version and click on single node to avoid unnecessary costing, and change the time in Terminate after column as per your interest and click on create.

**Next create a new notebook and run the sql query in the first cell.**

%sql

USE CATALOG samples;

SELECT

   hour(tpep\_dropoff\_datetime) as dropoff\_hour,

   COUNT(\*) AS num

FROM samples.nyctaxi.trips

WHERE pickup\_zip IN ('10001', '10002')

GROUP BY 1;

**Next you can run the python code.**

from pyspark.sql.functions import hour, col

pickupzip = '10001'  # Example value for pickupzip

df = spark.table("samples.nyctaxi.trips")

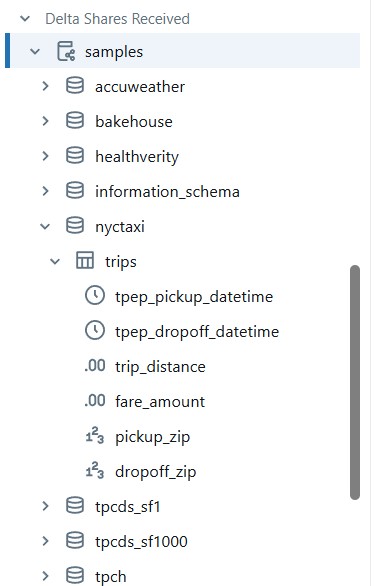
result\_df = df.filter(col("pickup\_zip") == pickupzip) \

              .groupBy(hour(col("tpep\_dropoff\_datetime")).alias("dropoff\_hour")) \

              .count() \

              .withColumnRenamed("count", "num")

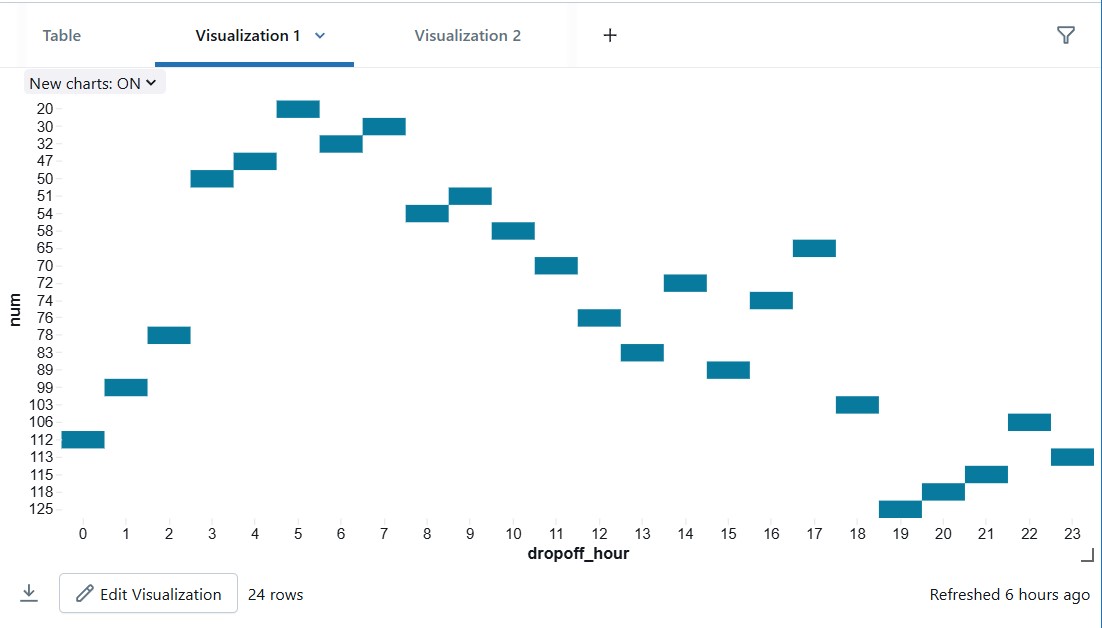
display(result\_df)



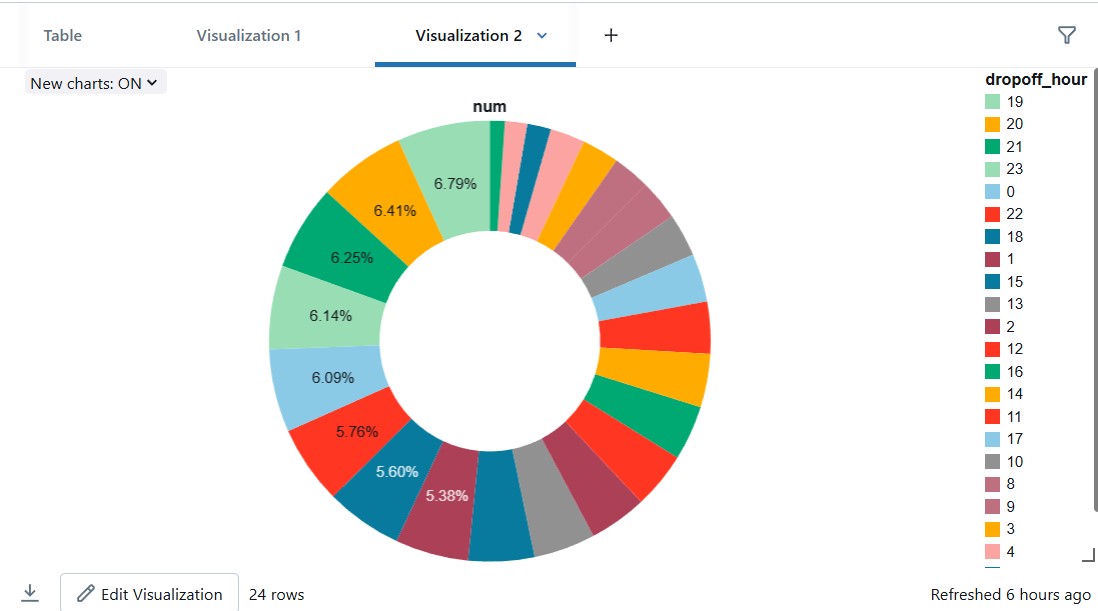
This code use the data present under samples named as nyctaxi.

**DIFFERENT TYPES OF VISUALIZATION**

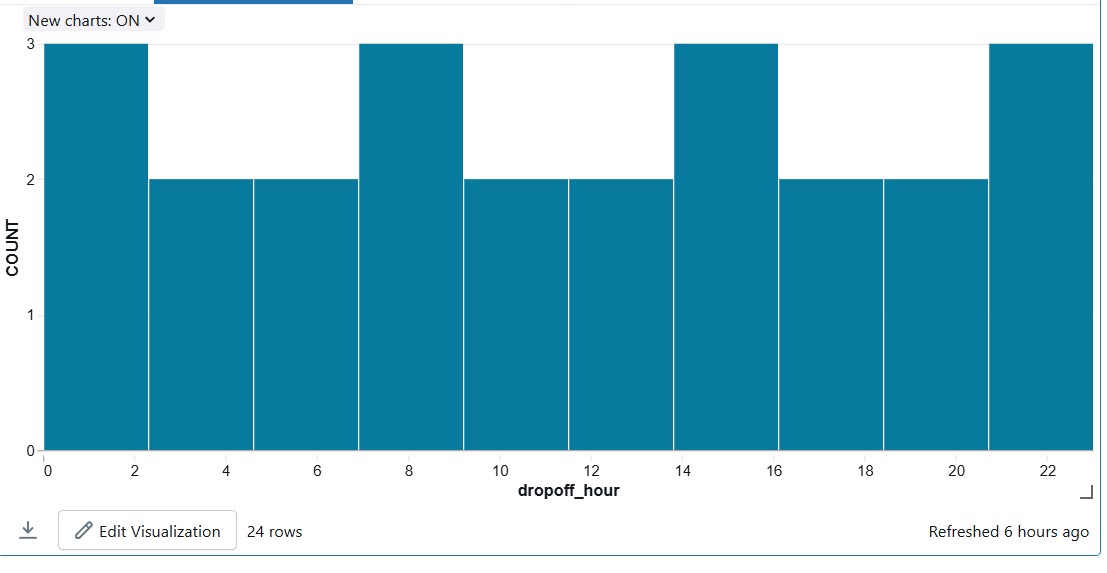
HEATMAPS



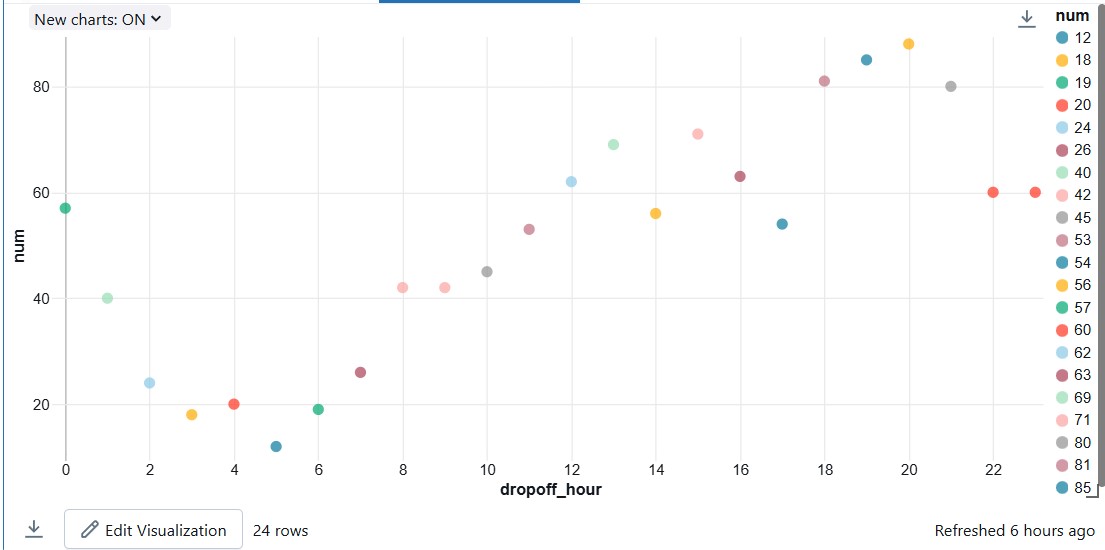
PIECHART



BAR GRAPHS



BUBBLE GRAPH



AREA GRAPH

