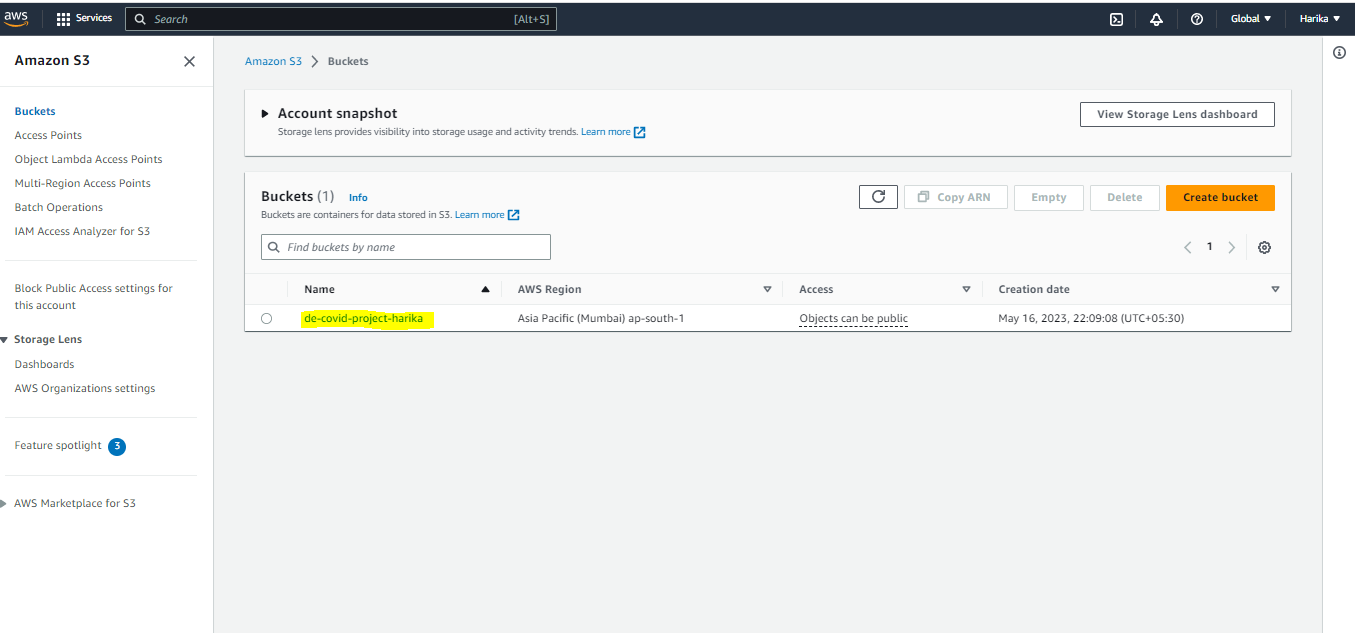
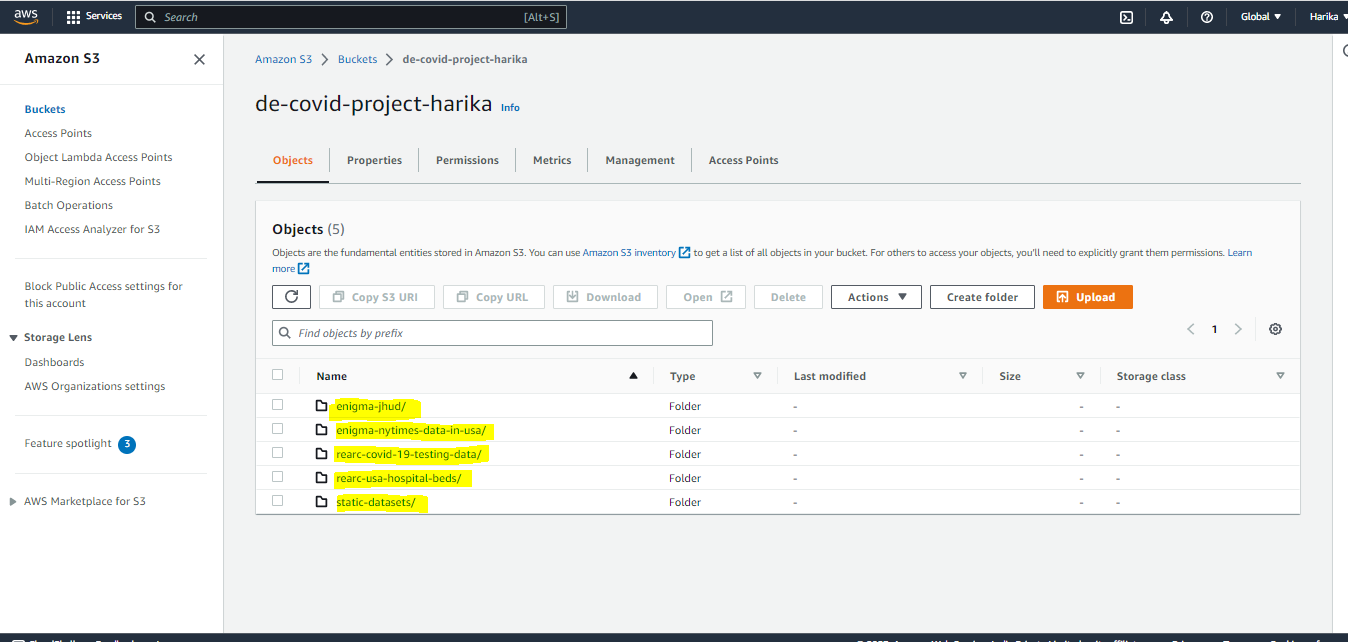
**Title of the project**

**Covid Data Analysis Project using Python, SQL and AWS**

**STEP-1**

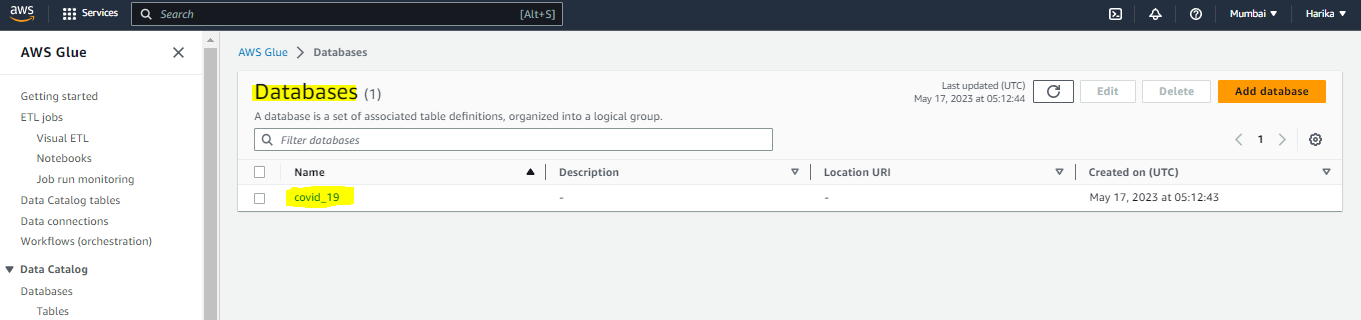
* **Extract the covid data and load the data into the AWS s3 bucket.**

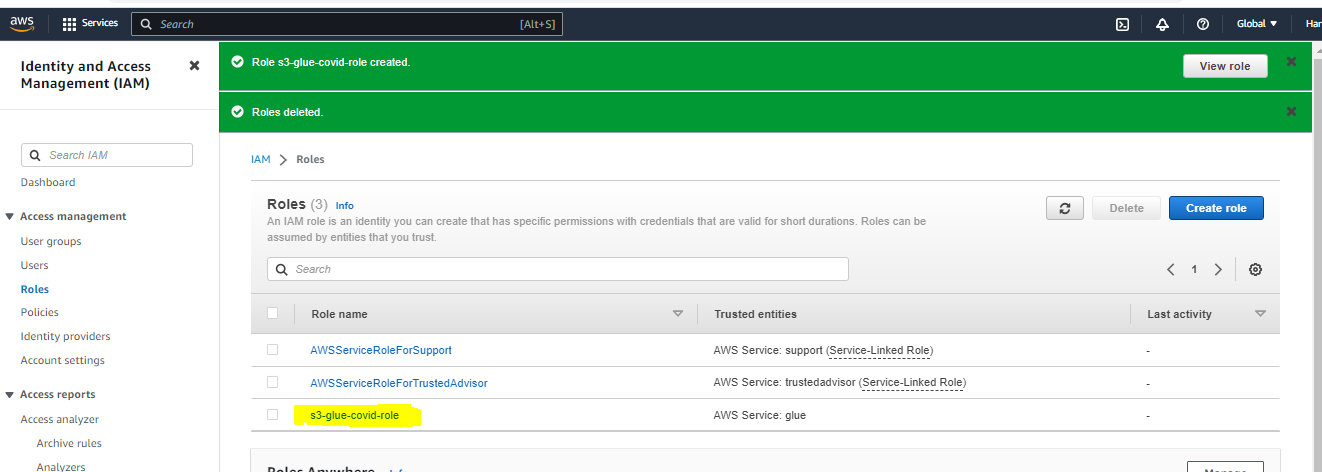




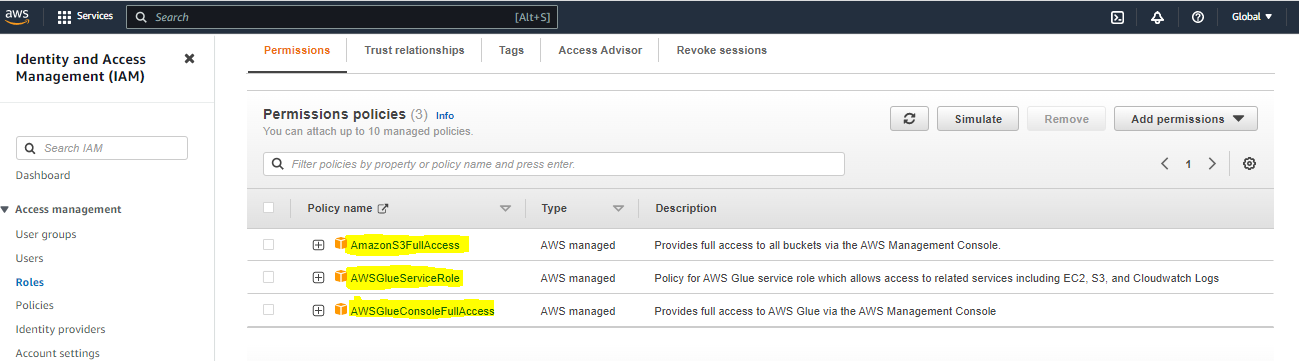
**STEP -2**

* Create crawler to query the data in Athena.
* To create a crawler firstly create a database and create an IAM role as shown below.

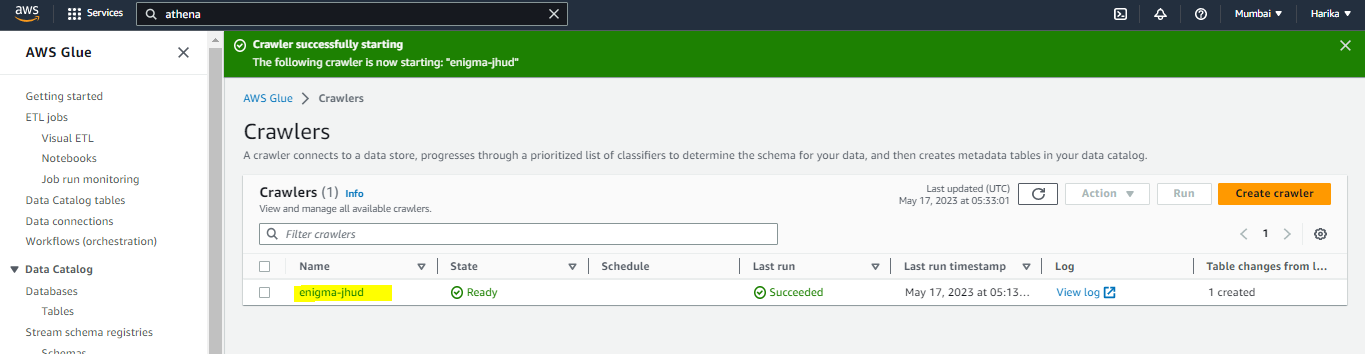




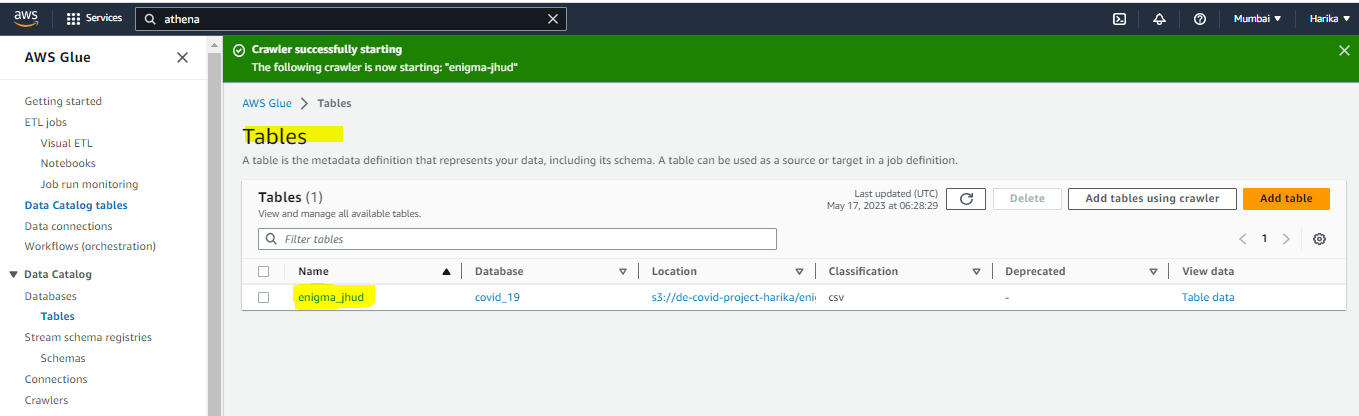
* In IAM role provide access to - [AmazonS3FullAccess](https://us-east-1.console.aws.amazon.com/iamv2/home?region=ap-south-1#/policies/details/arn%3Aaws%3Aiam%3A%3Aaws%3Apolicy%2FAmazonS3FullAccess), [AWSGlueServiceRole](https://us-east-1.console.aws.amazon.com/iamv2/home?region=ap-south-1#/policies/details/arn%3Aaws%3Aiam%3A%3Aaws%3Apolicy%2Fservice-role%2FAWSGlueServiceRole), AWSGlueConsoleFullAccess.

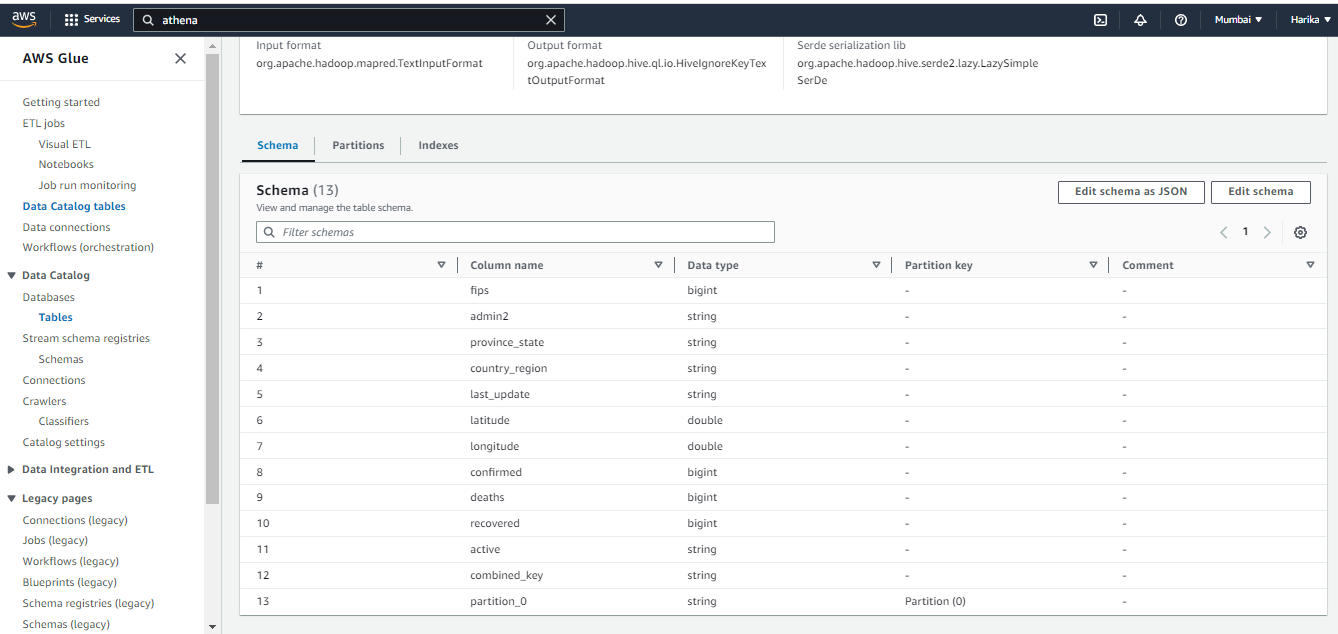


* Now create the crawler and run the crawler.

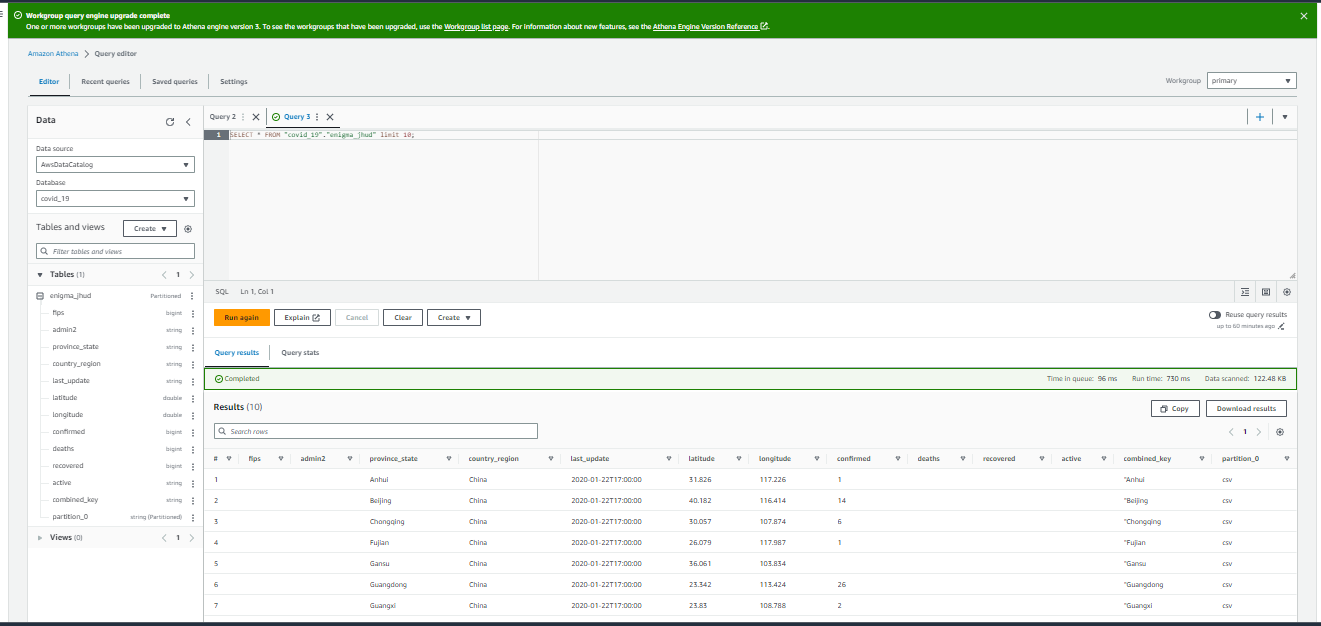


* Now we should be able to find the table once the crawler is ready.

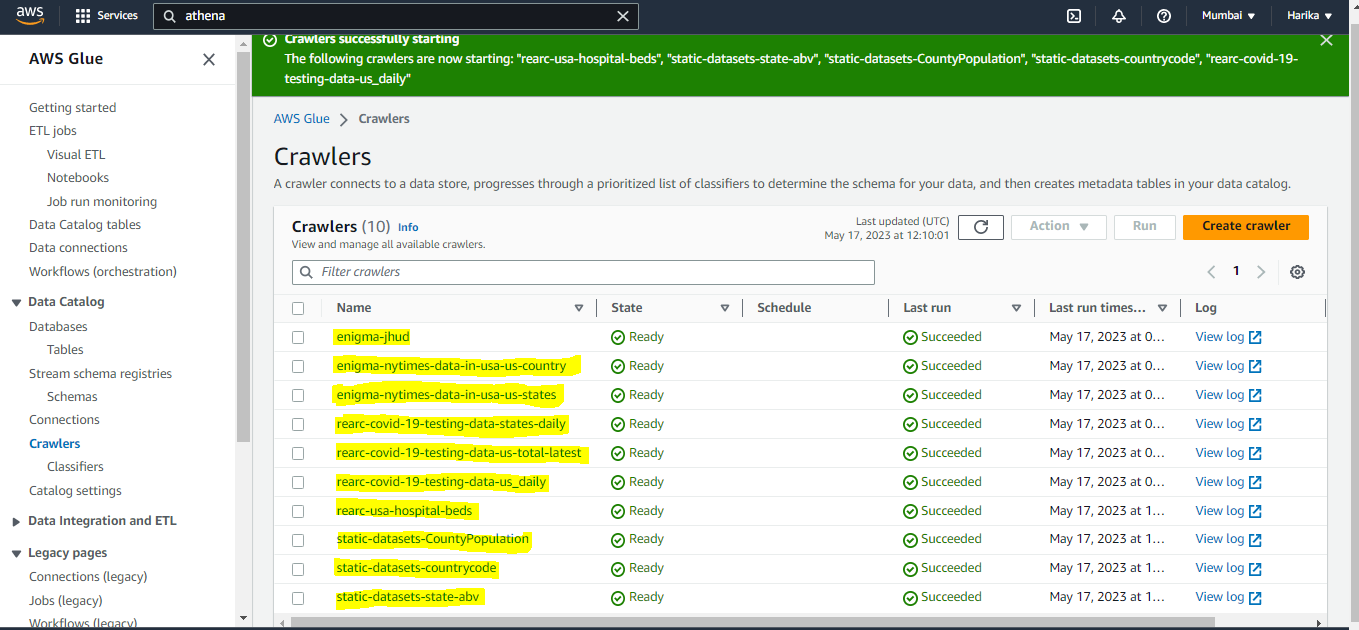




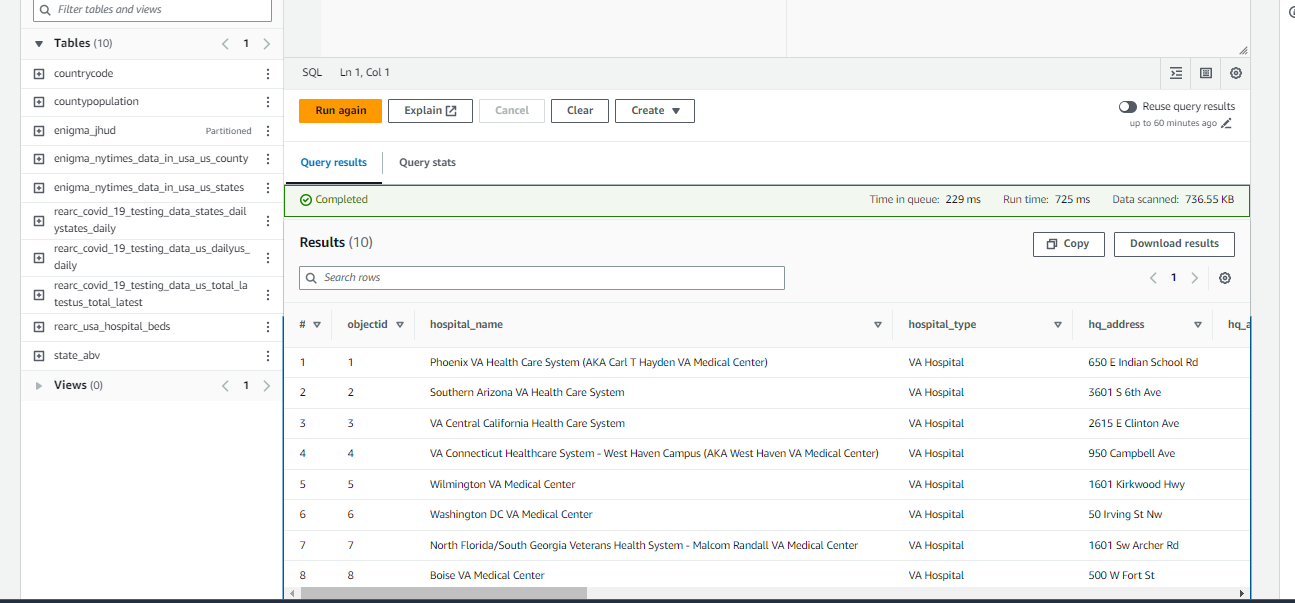
* Go to Athena and click on preview table.



* Similarly create the crawler for all the data present in the s3 bucket.

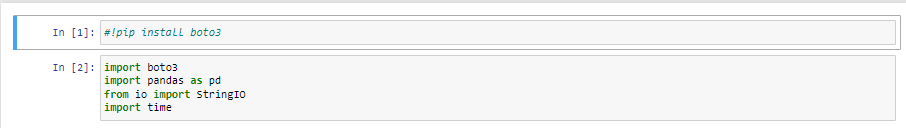


* Once all the crawlers are ready go to Athena and preview the data.



**STEP - 3**

* As now we can see that we have 10 tables with the data, however we can convert the data into star schema and reduce the redundancy of data across the different tables.
* Write the python code to connect to aws and connect to Athena and query the data
* Use the below code to import the libraries.



* To connect to the AWS use the below code:

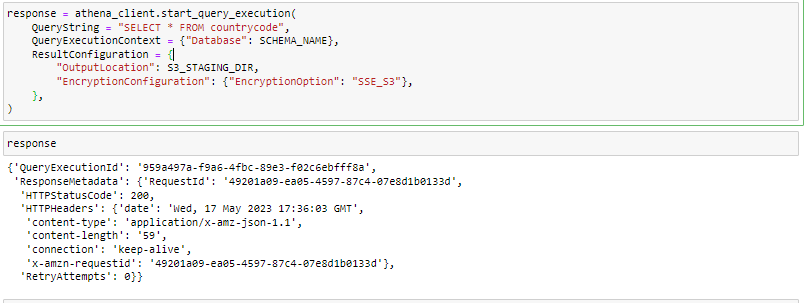


* Use the below code to connect to the Athena:

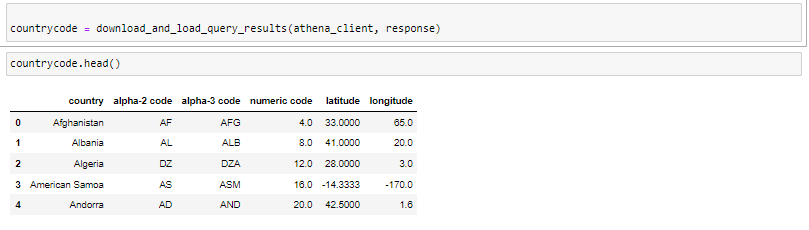


* Use the below code to query the data.

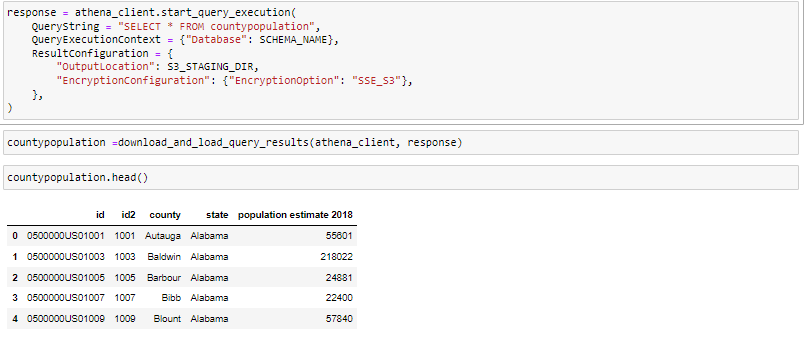


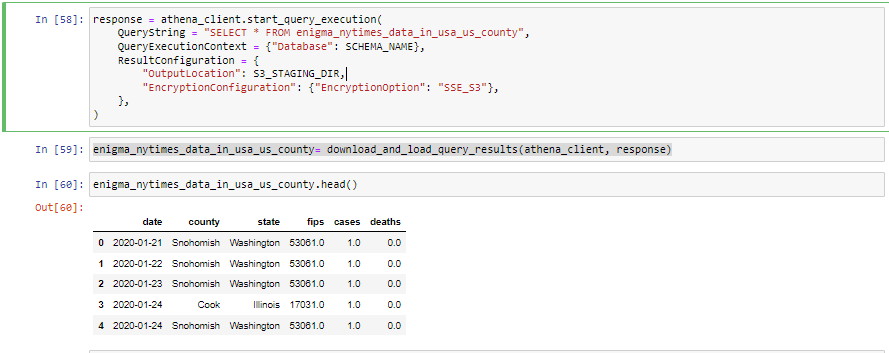


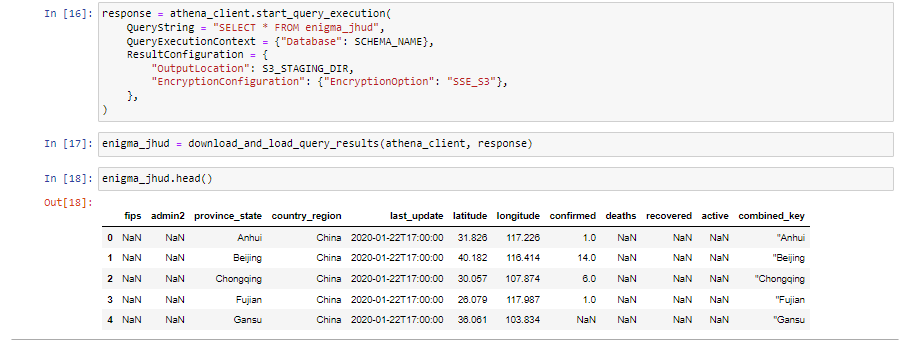
* Once we get the “HTTPSTATUSCODE” as 200 that means that we are able to establish the connection.
* Use the below code to get the data.

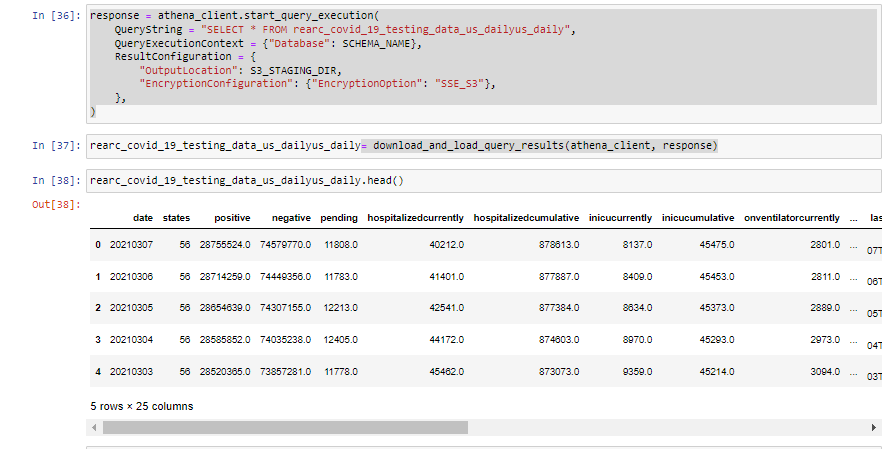


* Similarly, we can get the data of all the tables.





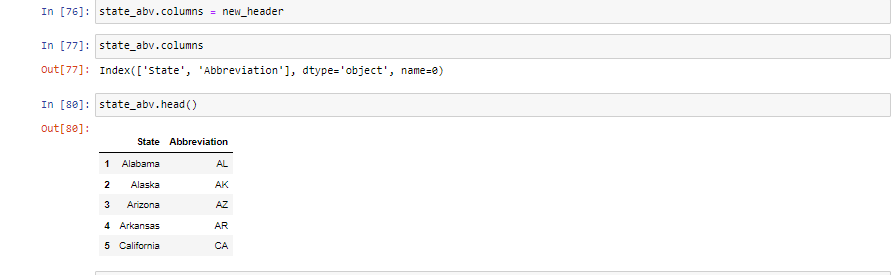






* As from the above data we can see that the columns names are not proper so we have to do some transformation to set the correct column names.
* Use the below code to set the column names.

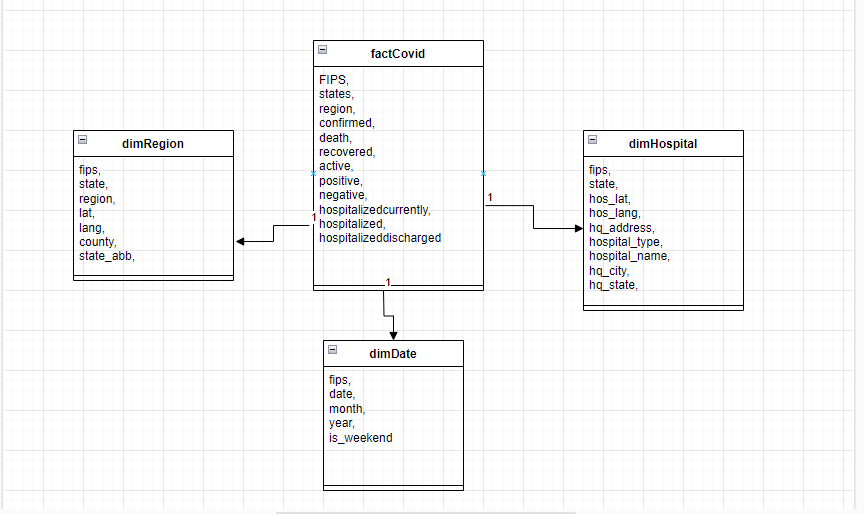




* So, we have successfully transformed the data into correct form.

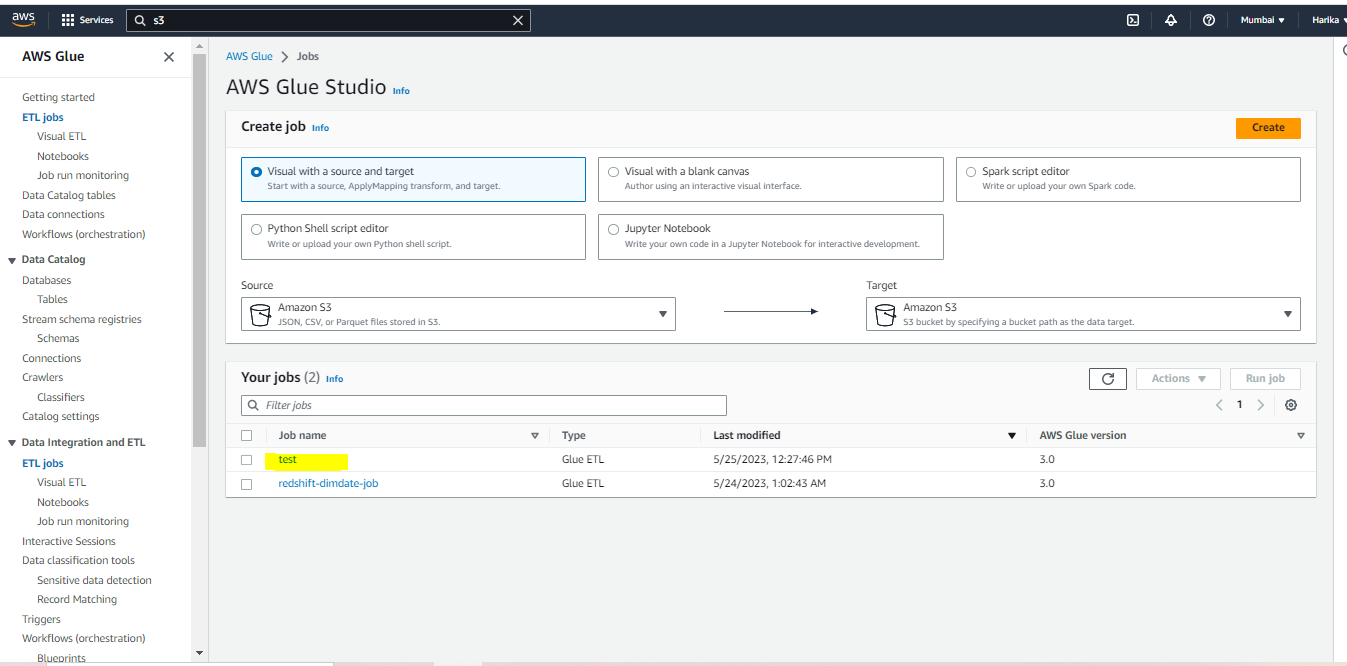
**STEP – 4**

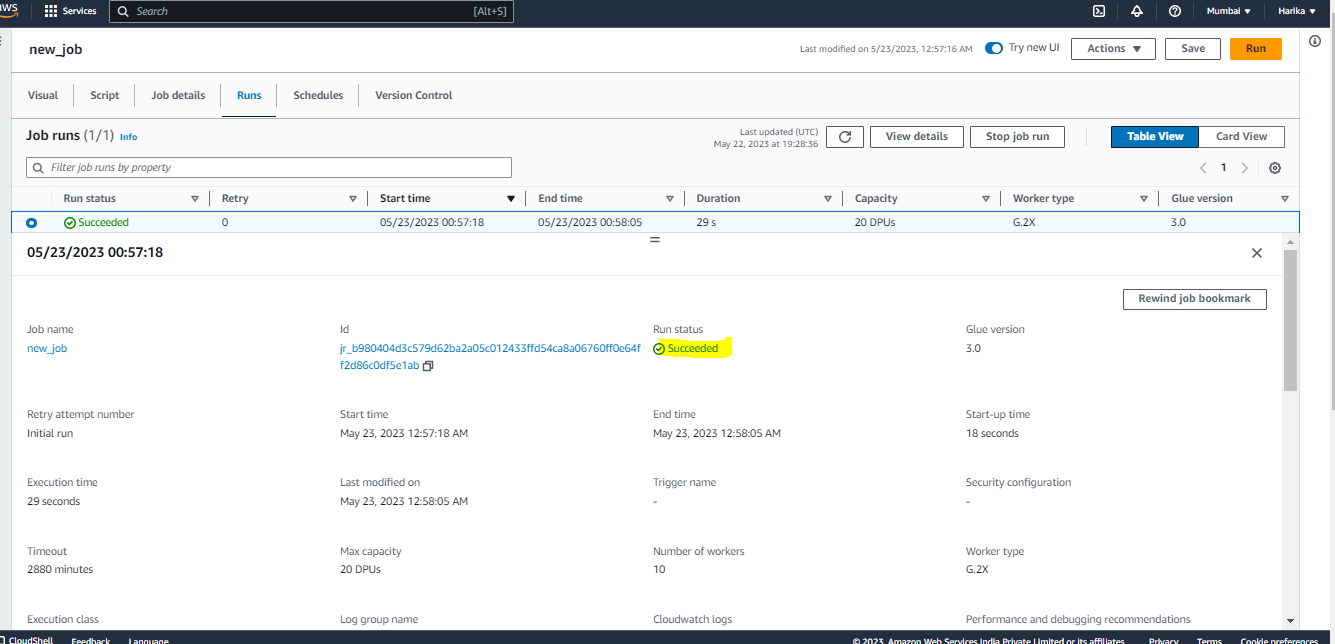
* Next, we have to work on dimension reduction and convert the data into star schema as shown below.

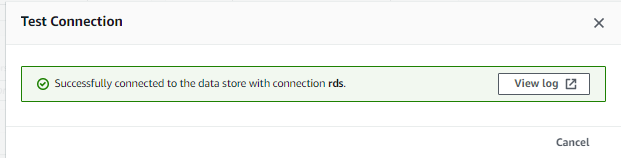


**Step-5**

* Deploy the transformed data into redshift.
* To do so, we can either use python code to deploy the data into redshift or we can use AWS GLUE.









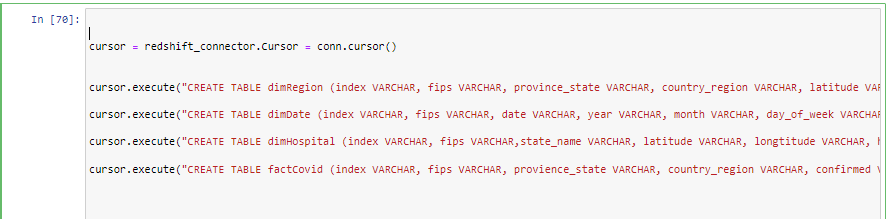
cursor = redshift\_connector.Cursor = conn.cursor()

cursor.execute("CREATE TABLE dimRegion (index VARCHAR, fips VARCHAR, province\_state VARCHAR, country\_region VARCHAR, latitude VARCHAR, longitude VARCHAR, country VARCHAR,state VARCHAR);")

cursor.execute("CREATE TABLE dimDate (index VARCHAR, fips VARCHAR, date VARCHAR, year VARCHAR, month VARCHAR, day\_of\_week VARCHAR);")

cursor.execute("CREATE TABLE dimHospital (index VARCHAR, fips VARCHAR,state\_name VARCHAR, latitude VARCHAR, longtitude VARCHAR, hq\_address VARCHAR,hospital\_name VARCHAR,hospital\_type VARCHAR,hq\_city VARCHAR,hq\_state VARCHAR);")

cursor.execute("CREATE TABLE factCovid (index VARCHAR, fips VARCHAR, provience\_state VARCHAR, country\_region VARCHAR, confirmed VARCHAR, deaths VARCHAR, recovered VARCHAR, active VARCHAR,date VARCHAR,postive VARCHAR,negative VARCHAR,hospitalizedcurrently VARCHAR,hospitalized VARCHAR,hospitalizeddischarged VARCHAR);")



cursor.execute("Copy dimRegion from 's3://harika-de-project-covid-output/dimRegion.csv' iam\_role 'arn:aws:iam::293486471273:role/service-role/AmazonRedshift-CommandsAccessRole-20230522T223424' delimiter ',' IGNOREHEADER 1 ")

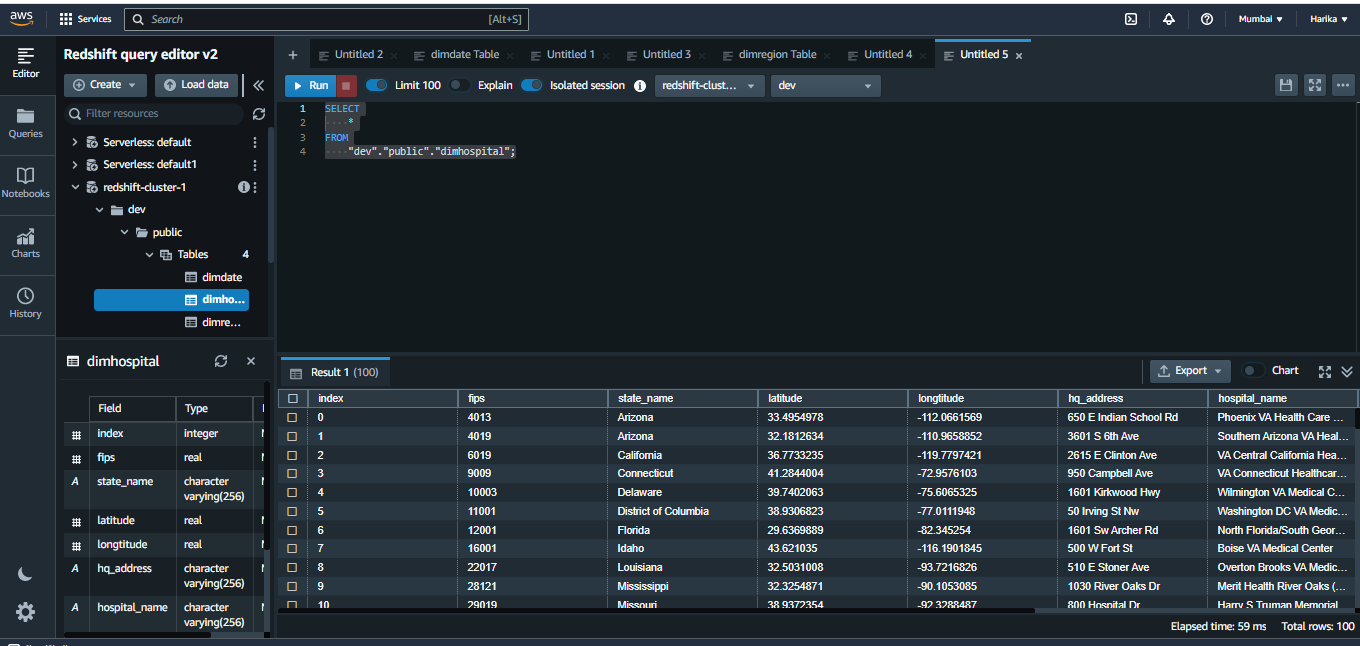
cursor.execute("Copy dimHospital from 's3://harika-de-project-covid-output/dimHospital.csv' iam\_role 'arn:aws:iam::293486471273:role/service-role/AmazonRedshift-CommandsAccessRole-20230522T223424' delimiter ',' IGNOREHEADER 1 ")

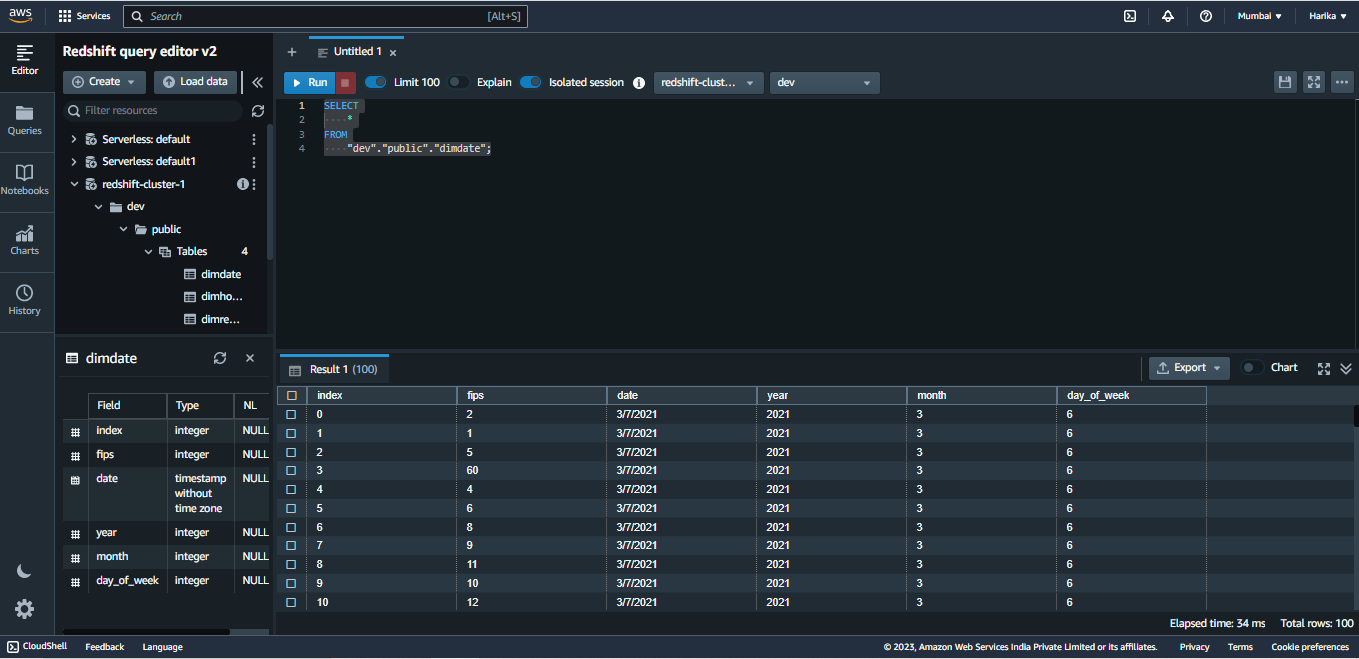
cursor.execute("Copy factCovid from 's3://harika-de-project-covid-output/factCovid.csv' iam\_role 'arn:aws:iam::293486471273:role/service-role/AmazonRedshift-CommandsAccessRole-20230522T223424' delimiter ',' IGNOREHEADER 1 ")

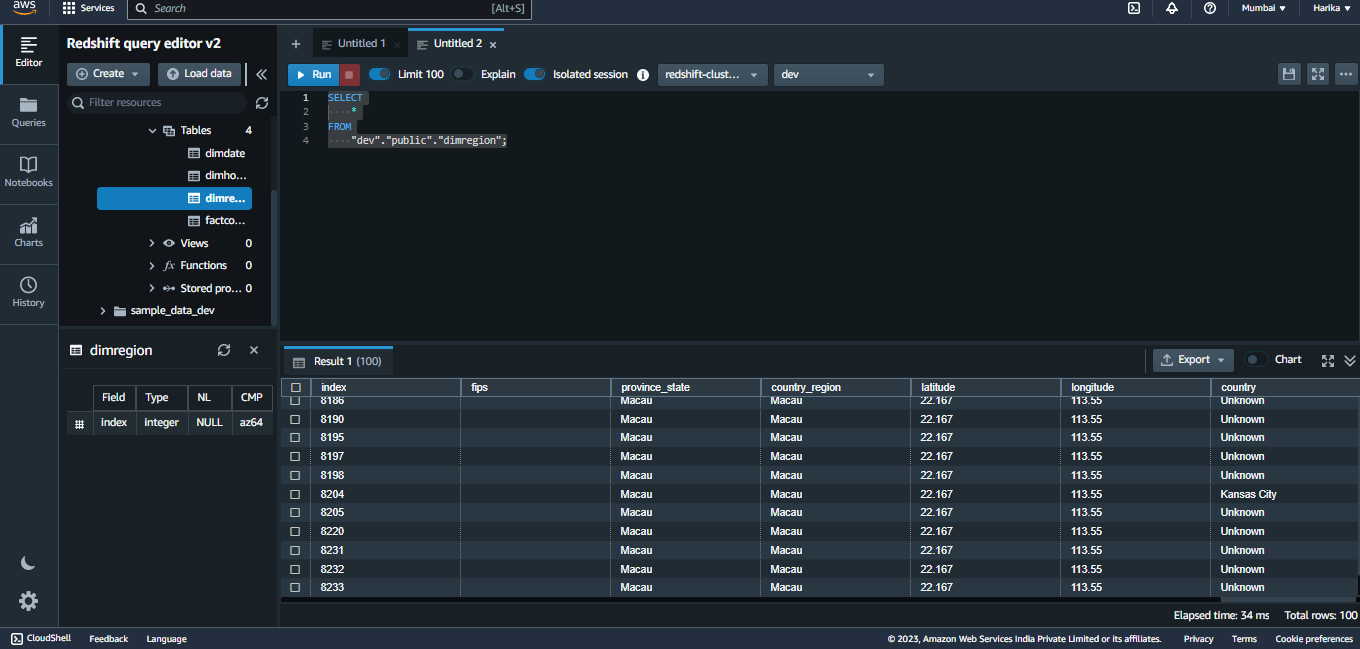
cursor.execute("Copy dimDate from 's3://harika-de-project-covid-output/dimDate.csv' iam\_role 'arn:aws:iam::293486471273:role/service-role/AmazonRedshift-CommandsAccessRole-20230522T223424' delimiter ',' IGNOREHEADER 1 ")

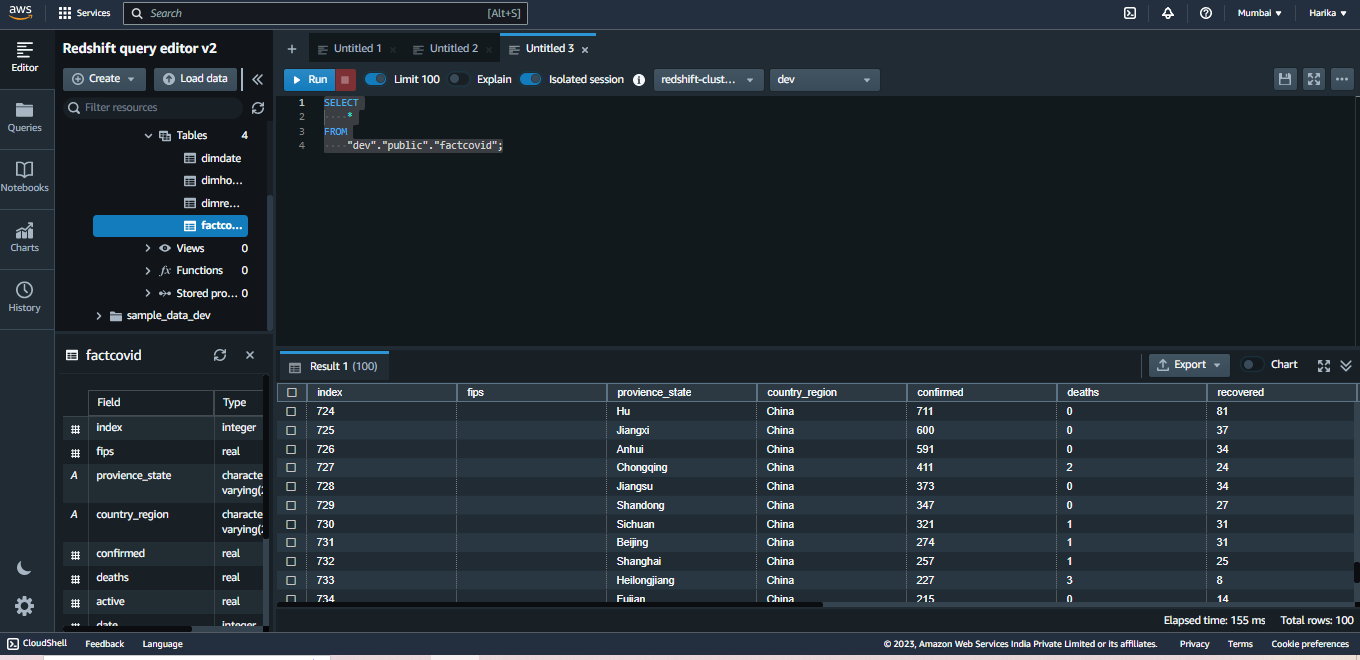


* After running the code go to redshift editor and check if the tables are created.









* We can query the data based on the requirement and prepare the dashboards from the data.

