**ASSIGNMENT 5.4**

**on**

**Data Architecture and Data Lakes in AWS**

**Submitted by:**

**Haseebullah Shaikh (2303.KHI.DEG.015)**

**and**

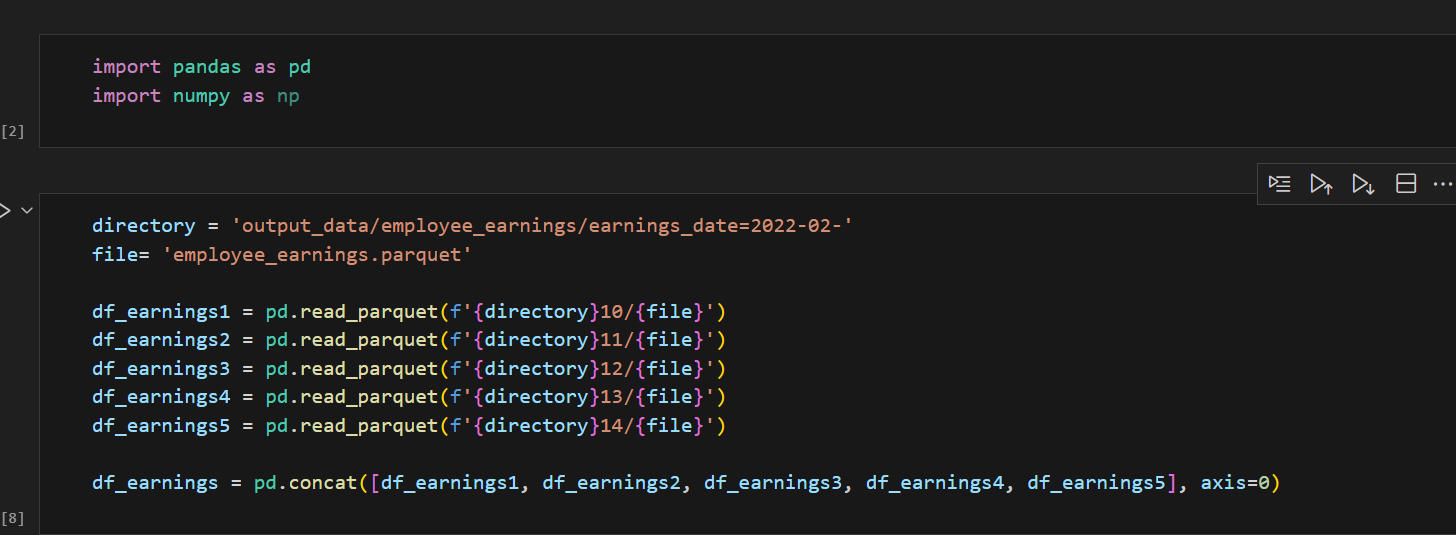
**Faiza Gulzar Ahmed (2303.khi.deg.001)**

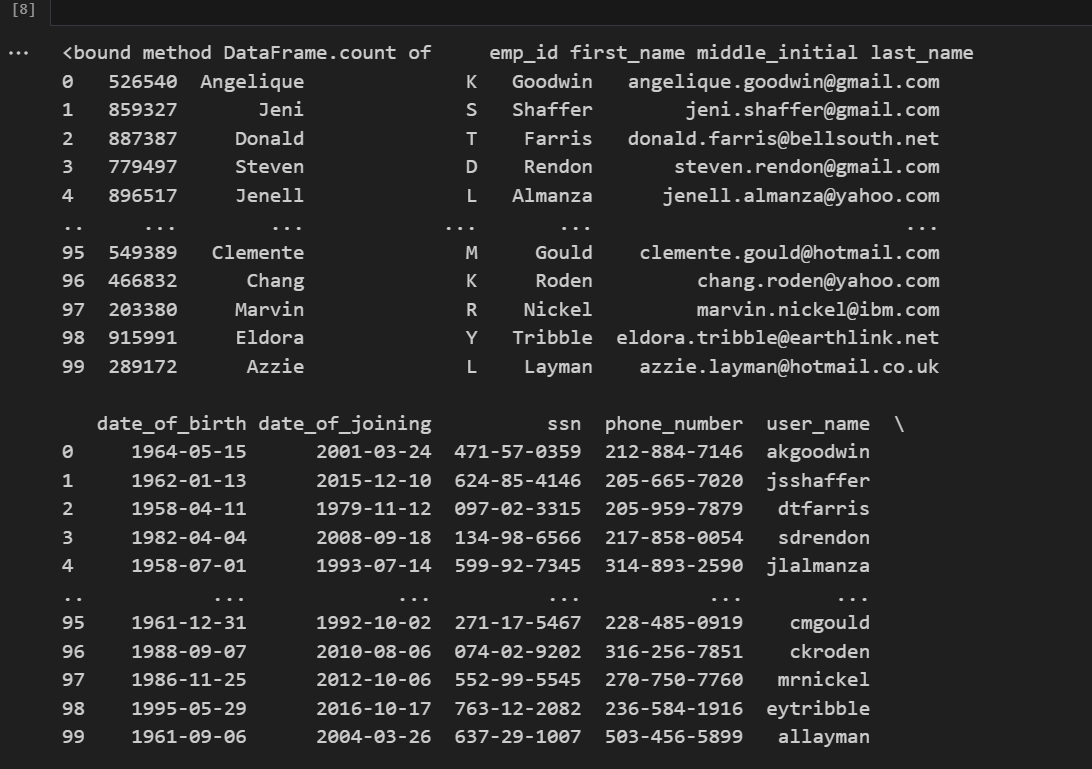
**Dated:** 19th May 2023

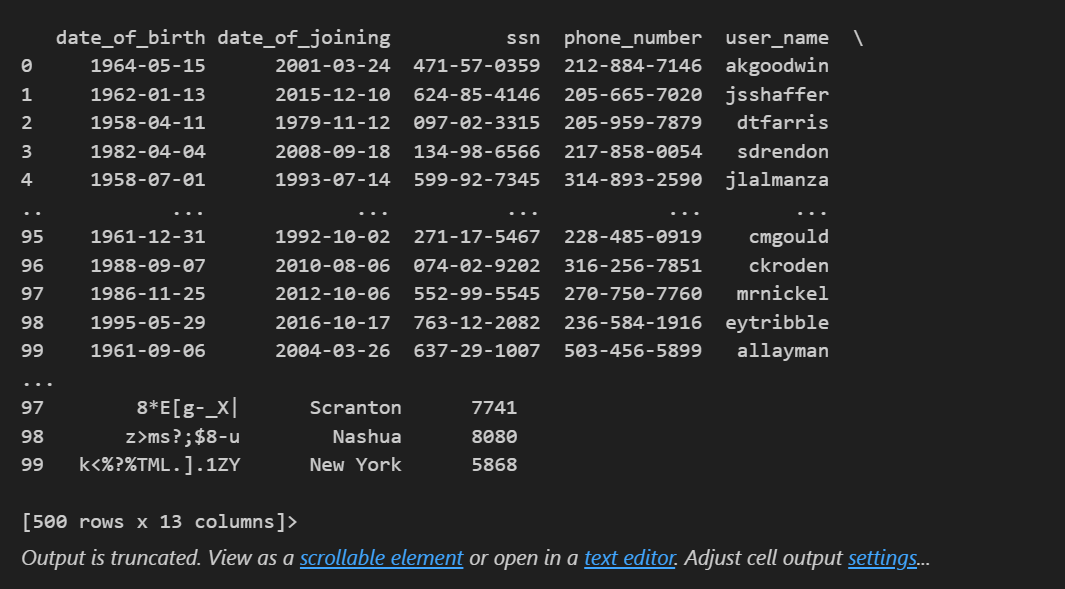
**Solution :**

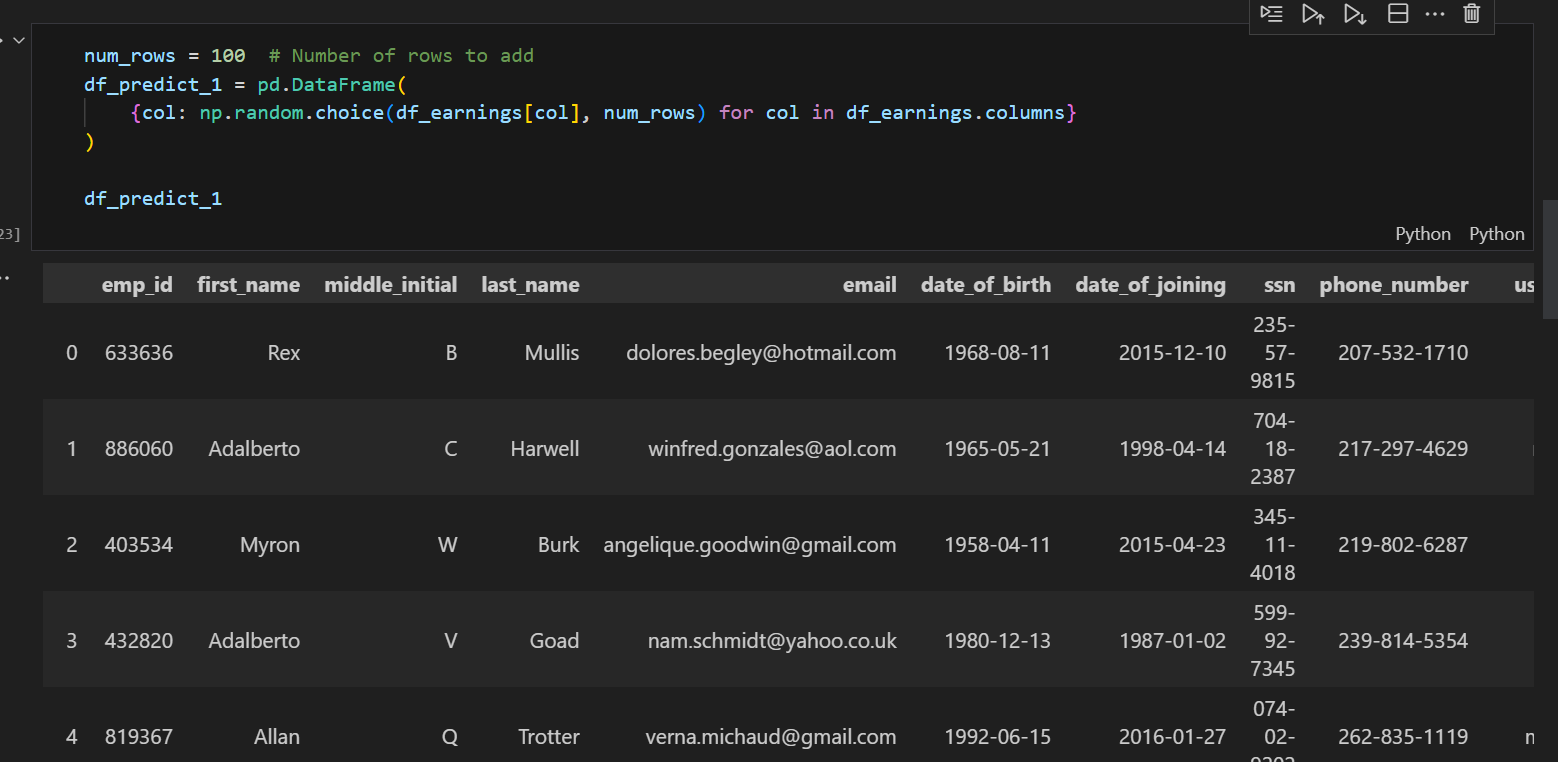
Simulated the Earning predictions for two more days, each day contains 100 entries or rows, created 200 new random entries for two days 100 for each day. We have used np.random function of NumPy library of python for generating random rows. Furthermore, random data is converted in to parquet format and stored in local directory.

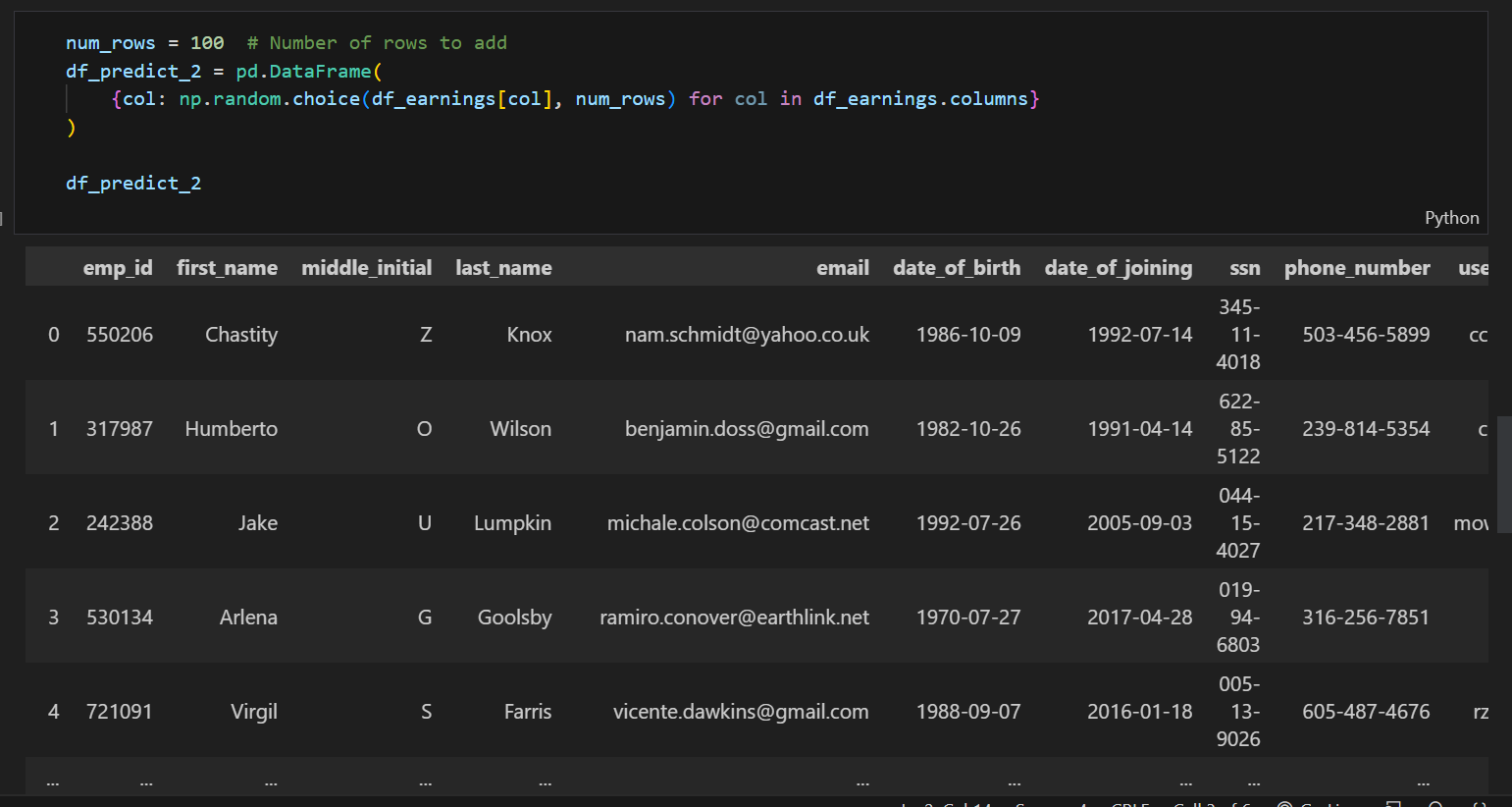
**Manipulation file**

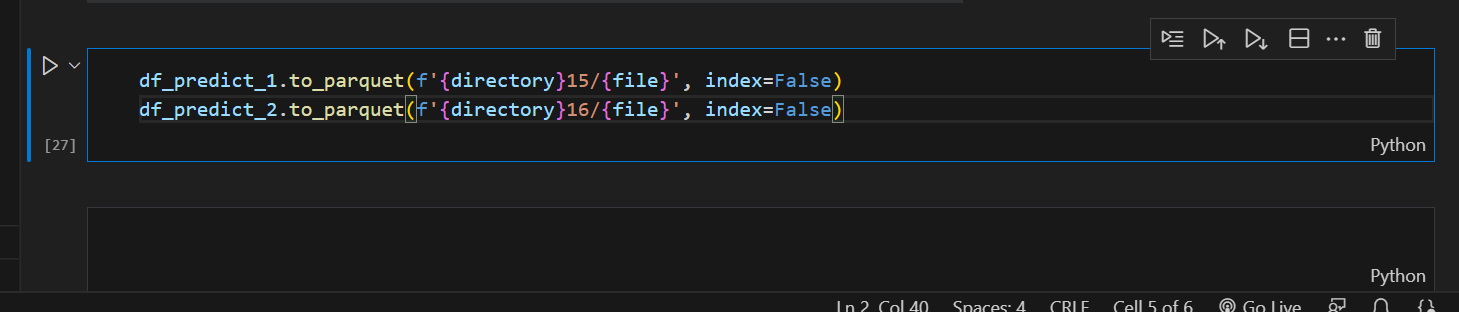




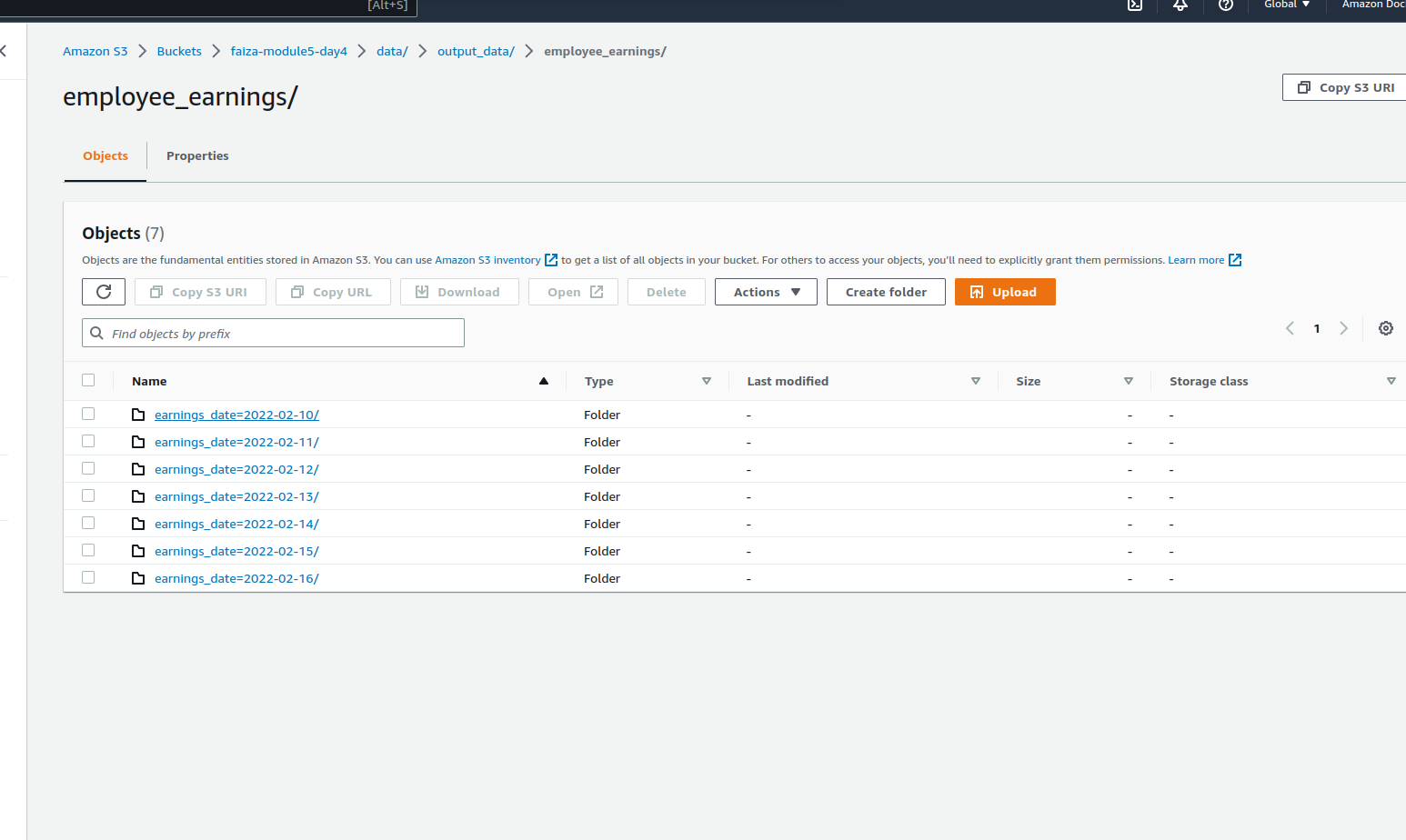




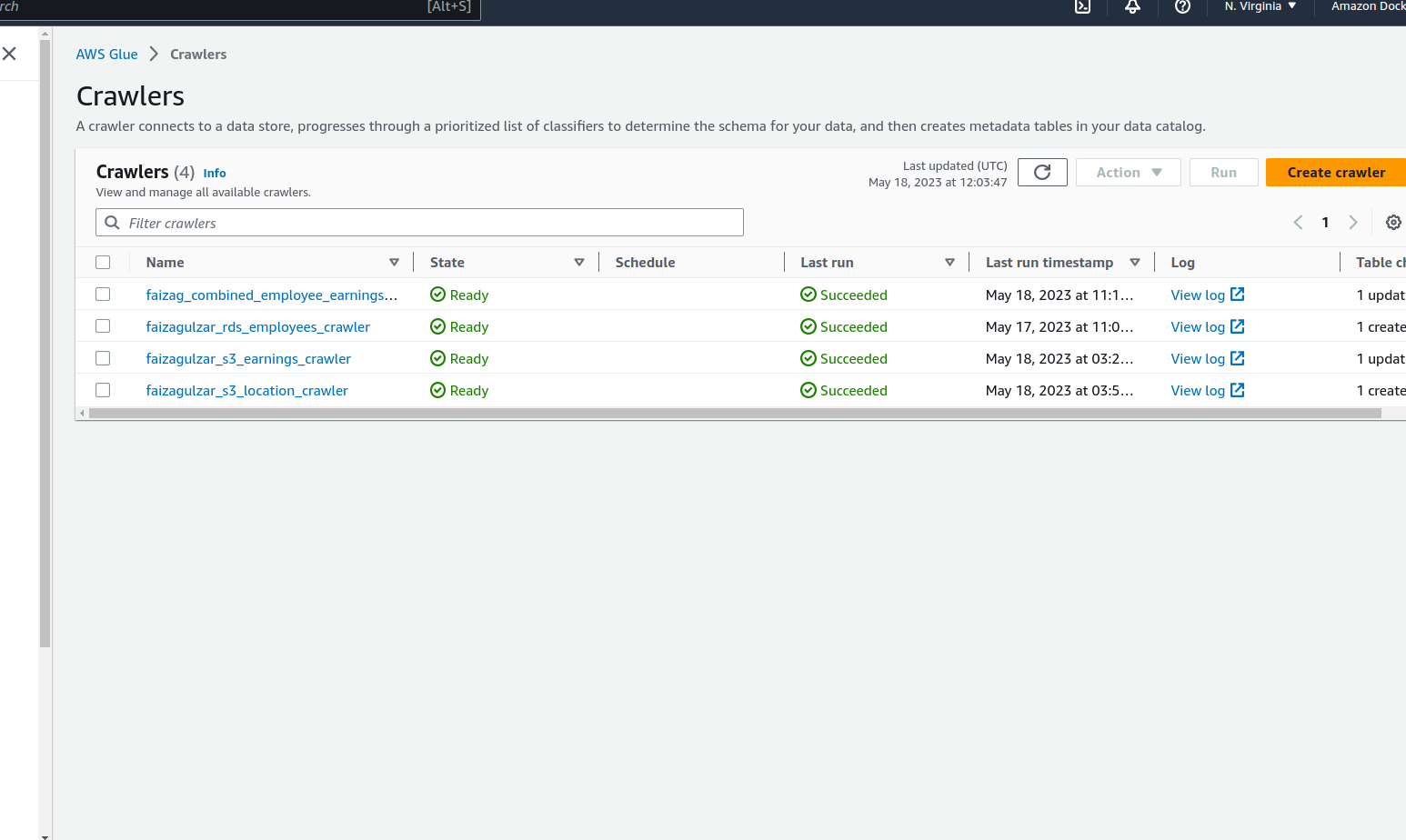




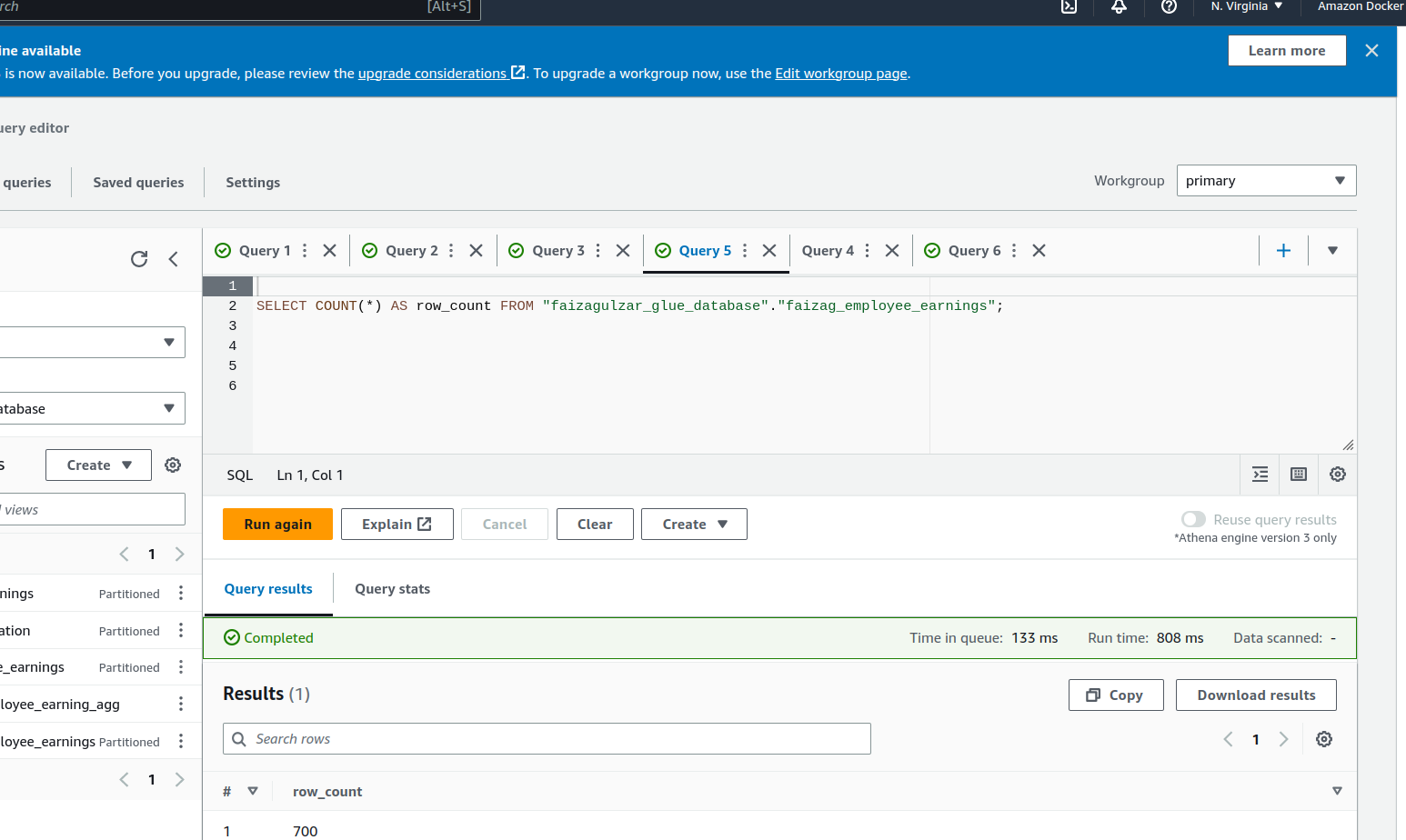
**Uploaded created random files in existing bucket which were created in task**

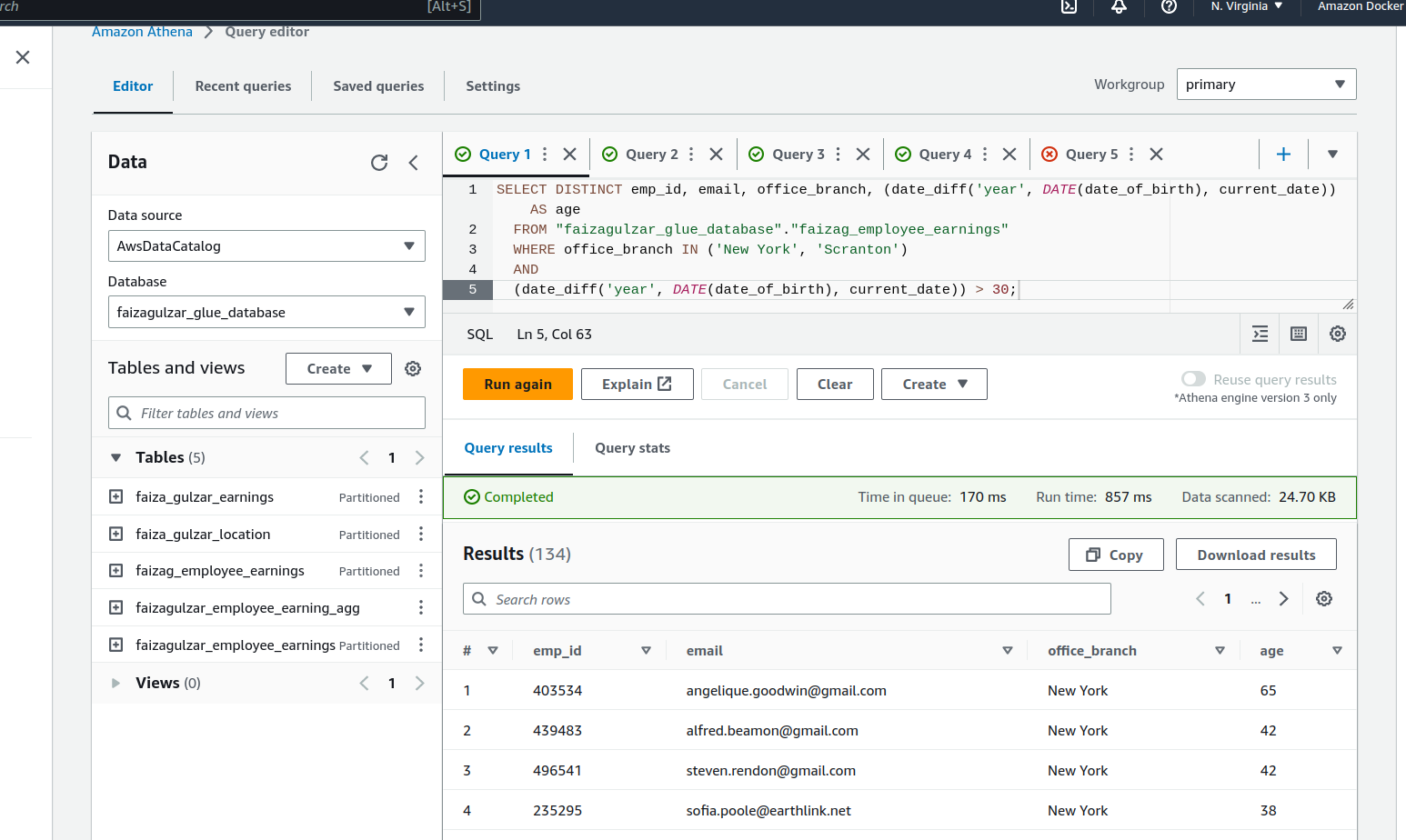


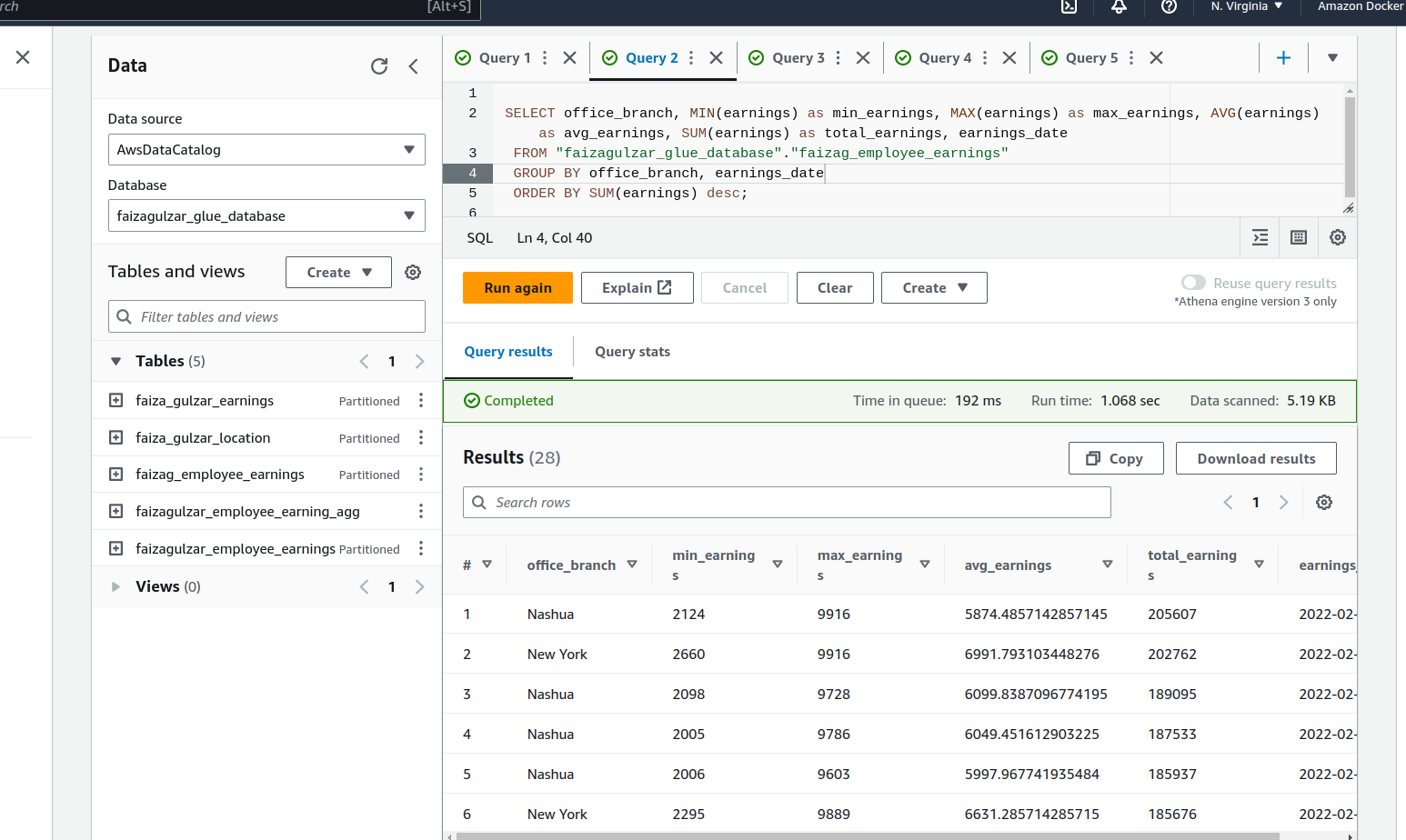
We ran again the crawler named faizag\_combined\_employee\_earnings which were created in the task for extracting the new uploaded data.

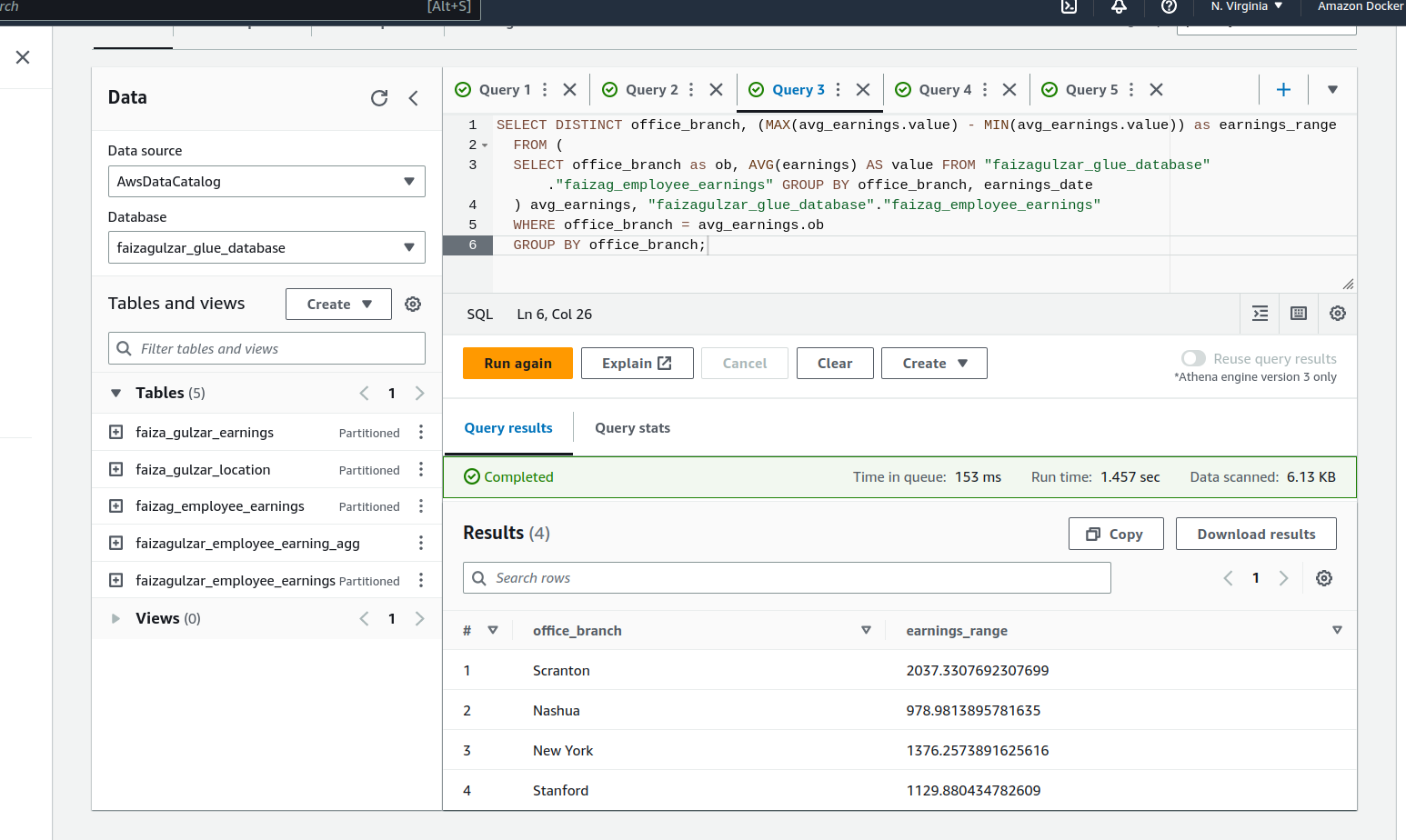


We ran the given SQL queries again in Anthena service to see the changes in data. As it can be seen there are 700 rows, previously there were 500 as we have now add 2 days random data which are 200 hundred rows.

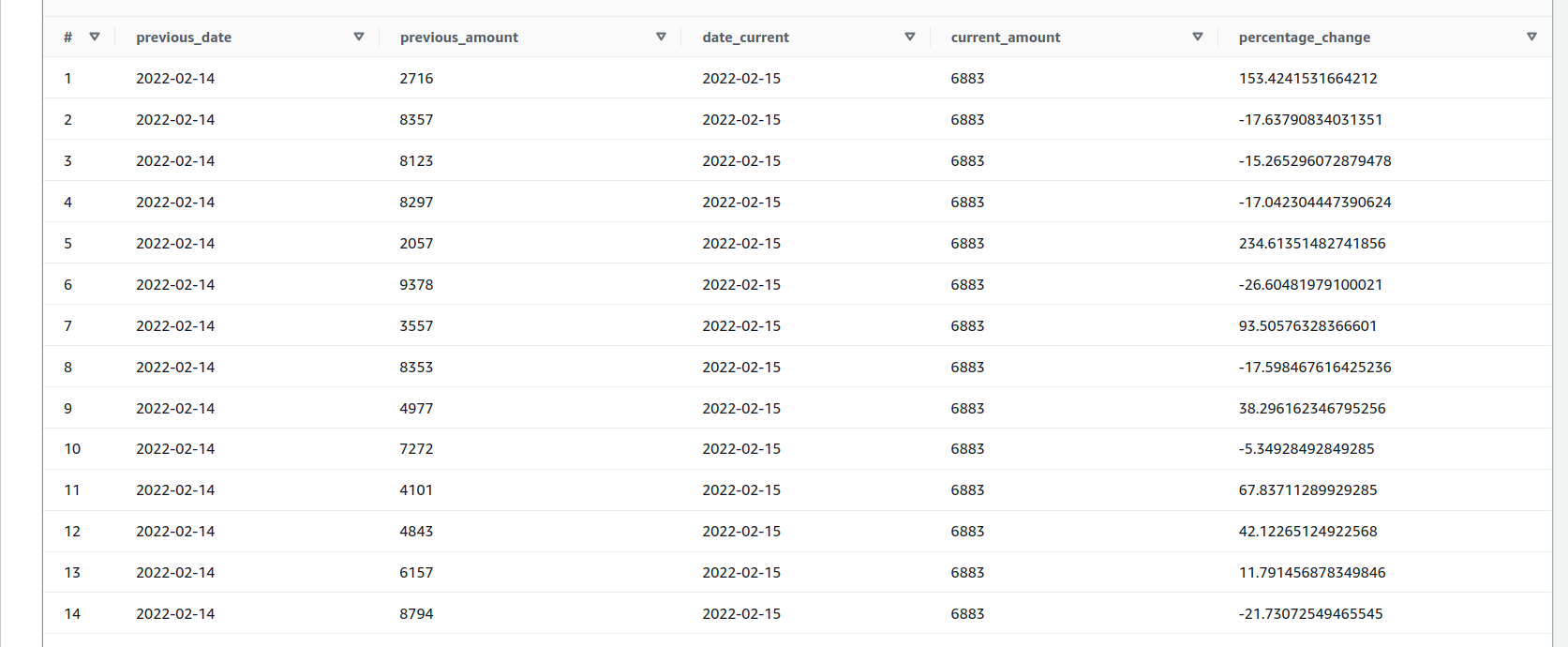
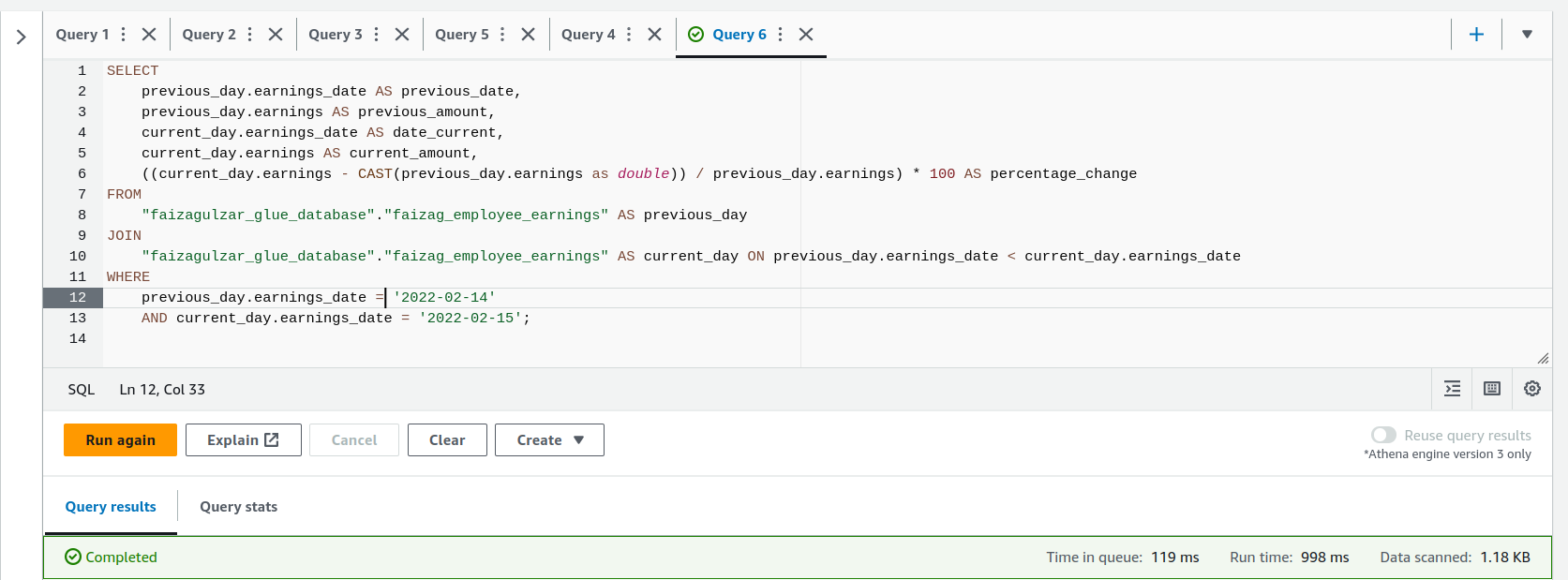








**Created new query which calculates the % change in earnings for every employee from a given day compared to the previous day**



**The End 😊**