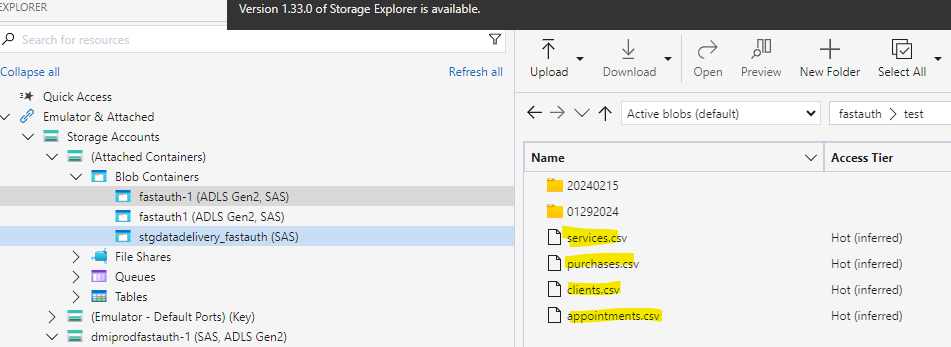
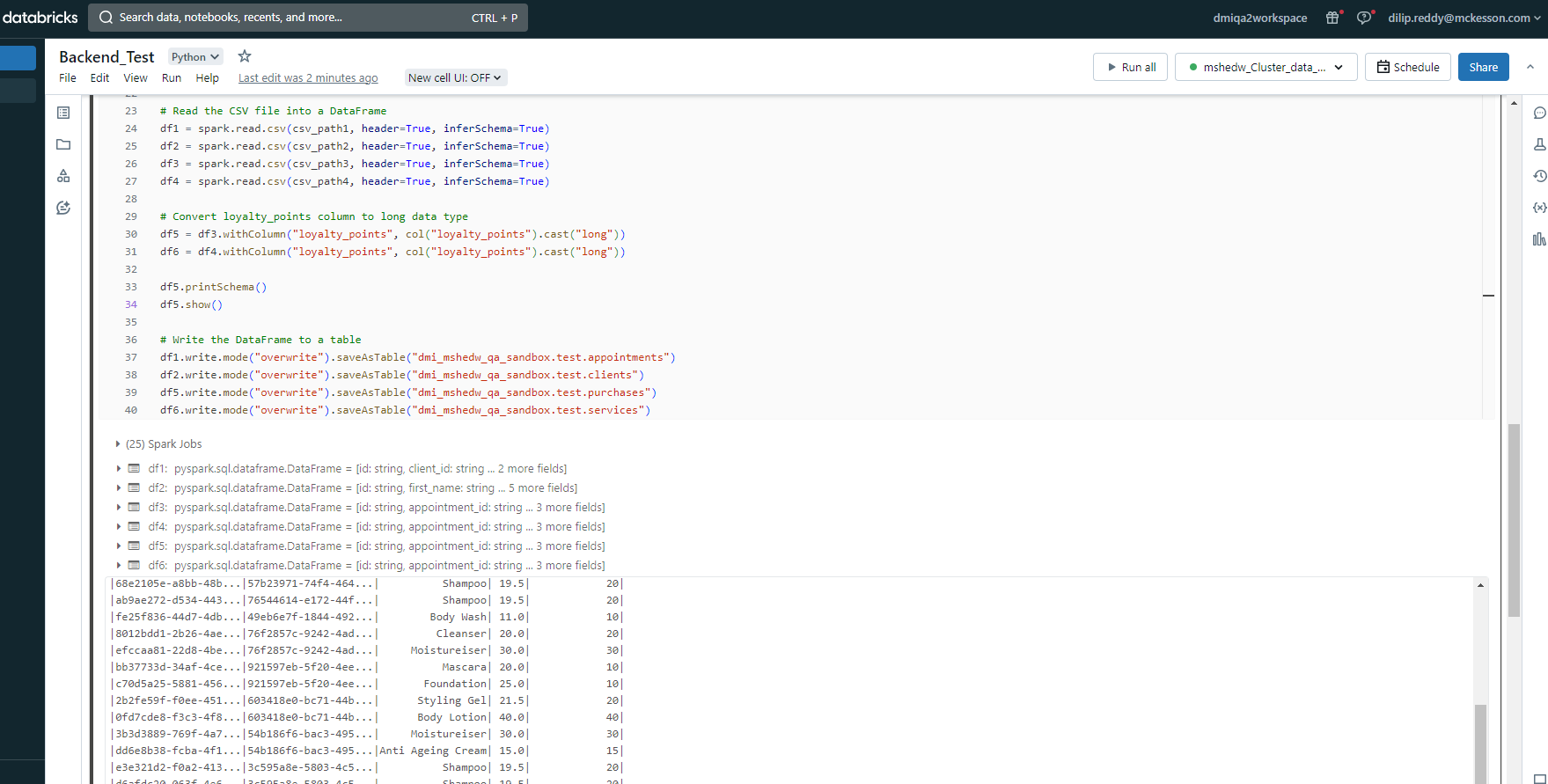
Step 1: Manually copied the files to azure blob storage.



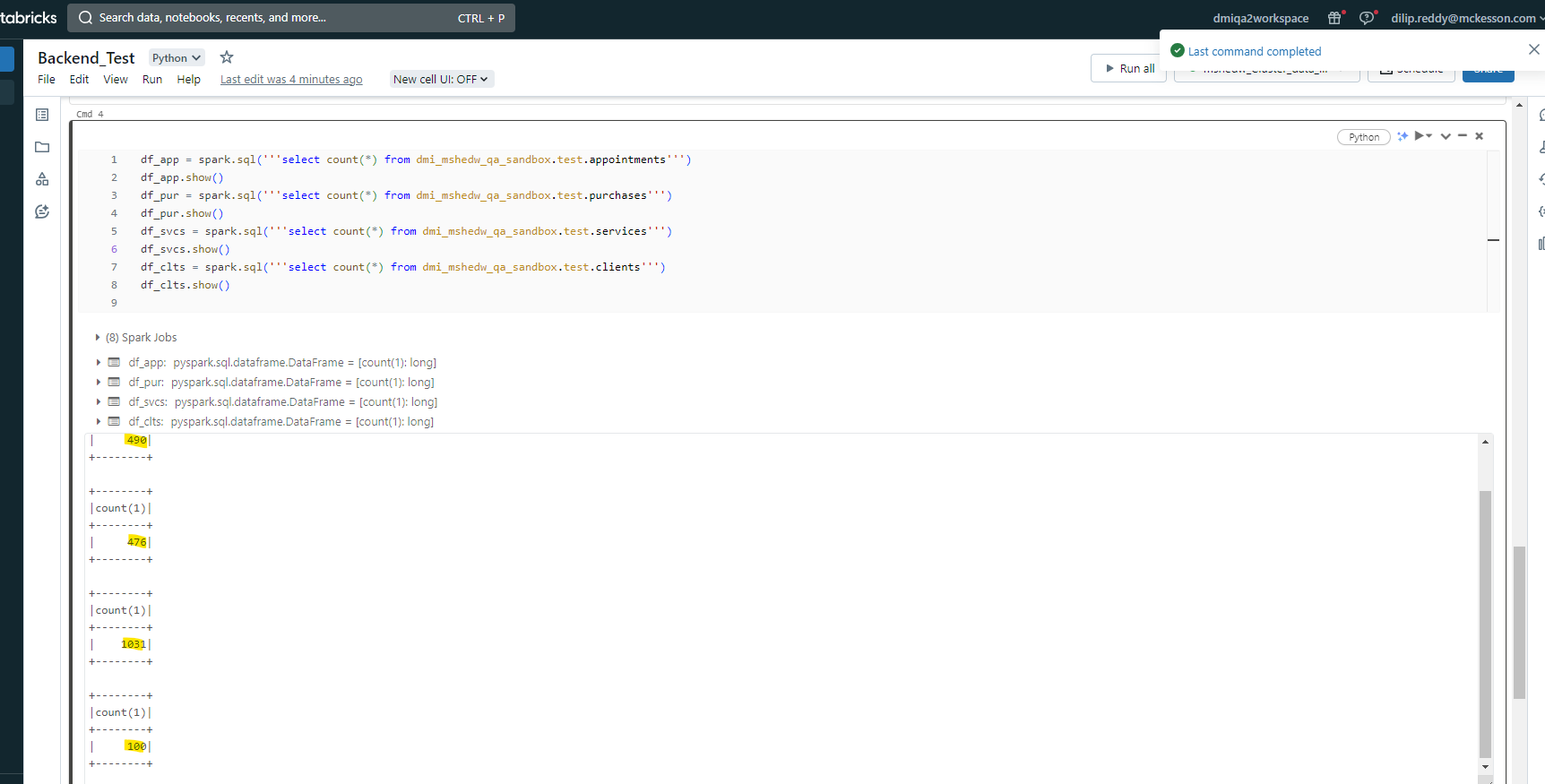
Step 2: Created 4 tables in databricks by executing the script.

Step 3: Created a notebook in databricks to read the above files. File read is successful, please see below output for one of the files.



Step 4: Loaded the data into the tables created in databricks.

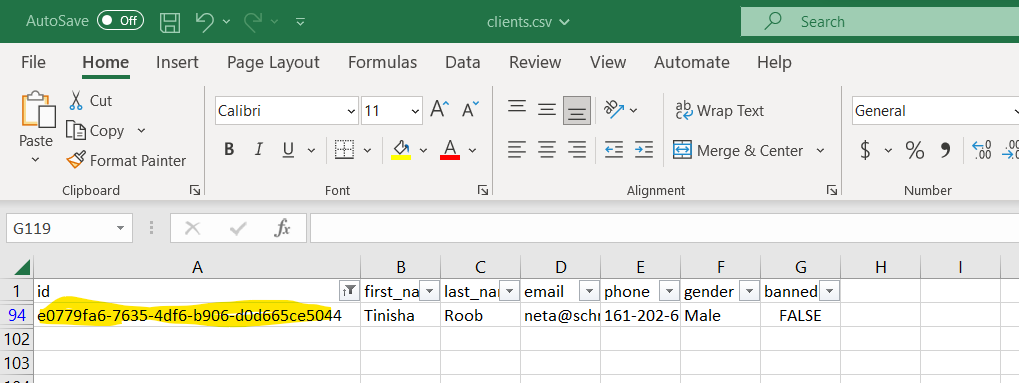
Please see below counts output for each table, the counts in table matched to the corresponding .csv file.



Step 5: Wrote code to identify the top 50 clients with most loyalty points.

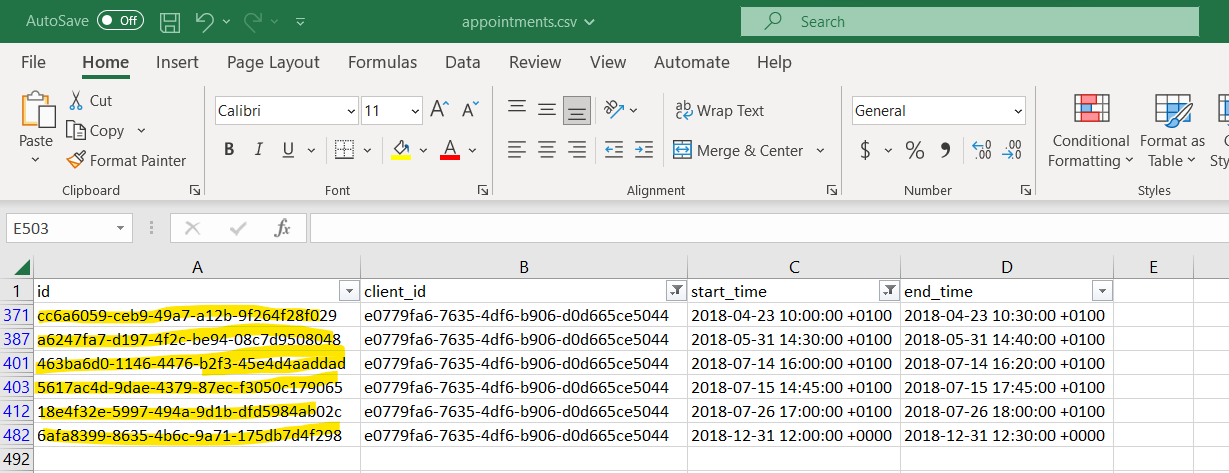
To test my code, I took one clientid and checked the loyalty points in csv file and compared it against to my query output.

Example: I picked below clientid from clients.csv file

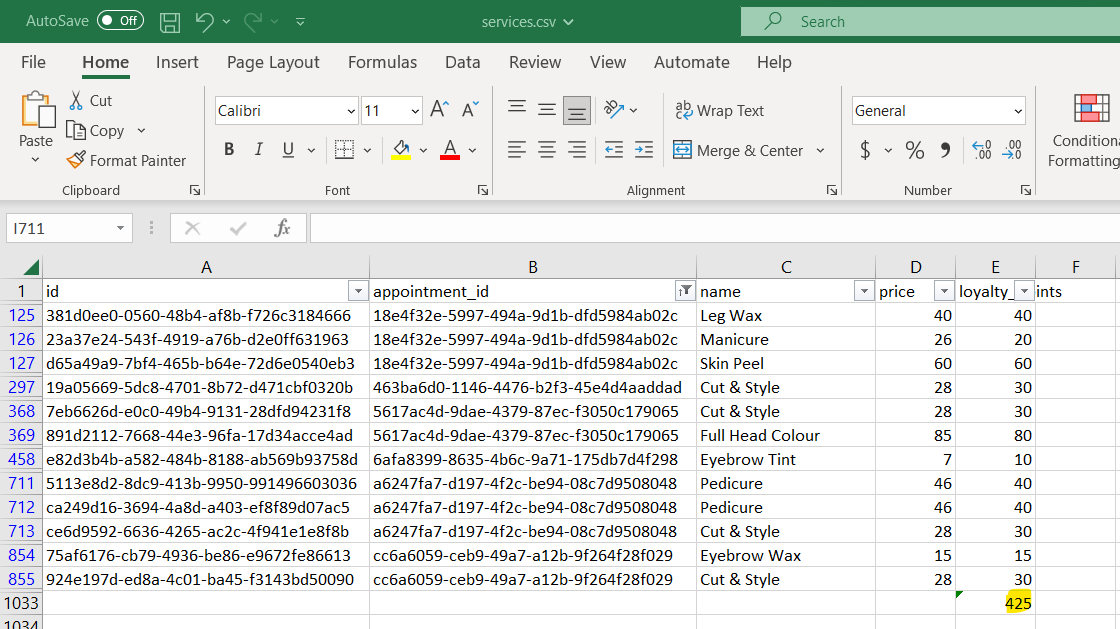


For the above clientid , I checked the appointmentid’s with starttime>= 2018-01-01 00:00:00,

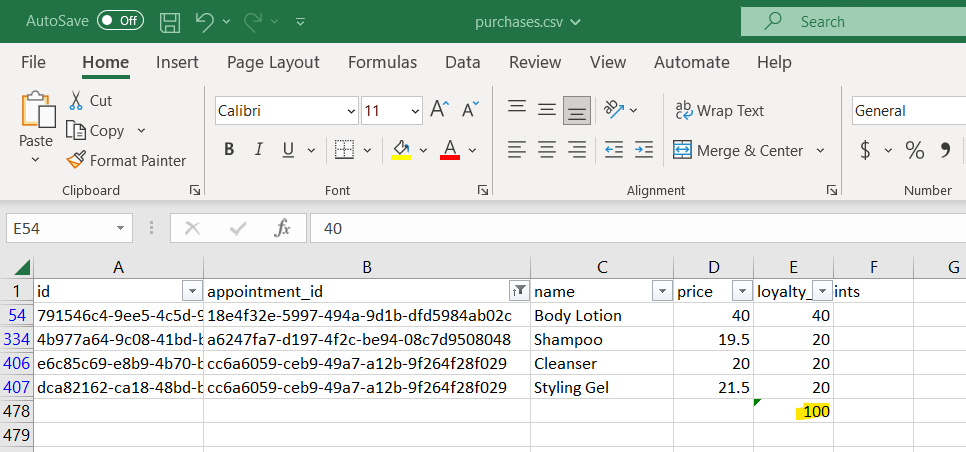
found below six appointmentid’s



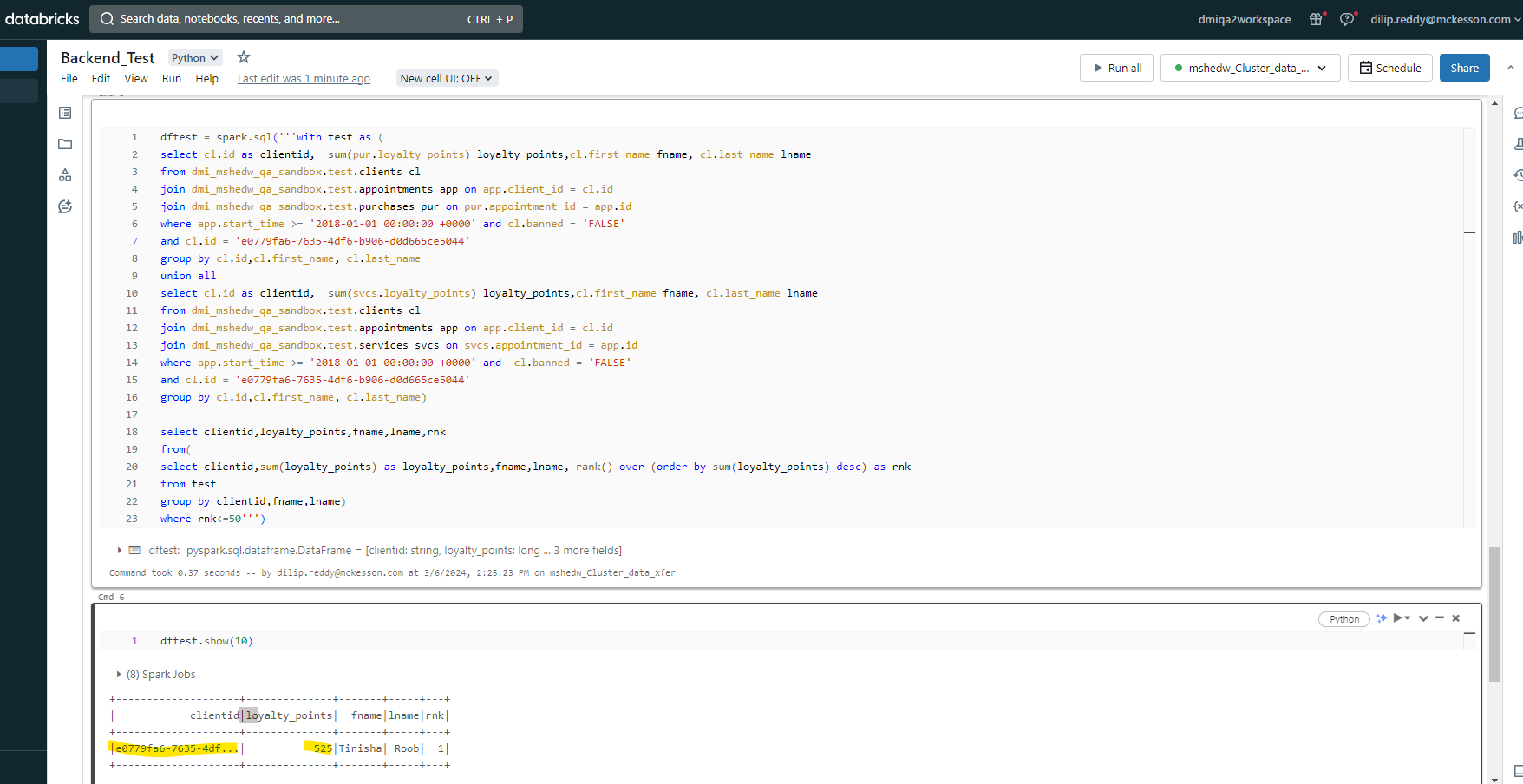
For the above appointmentid’s , I checked the loyalty points in services file. The total points are 425.



For the same appointmenid’s I checked the loyalty points in purchases file. The total points are 100.



The output of my query should show the loyalty points as 525 for the above clientid, because I am considering the loyalty points from both the files and adding them



The output from my notebook is matching to what is there in the file.

Test Passed Hurrrrayyyyyyyyyyyyyyyyyyyyy!!!! 😊

Then I ran my notebook for entire dataset by removing the clientid filter.

