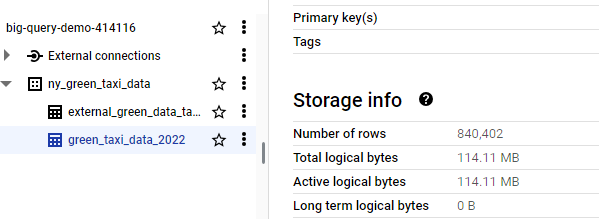
Big Query – Homework #4

SQL queries

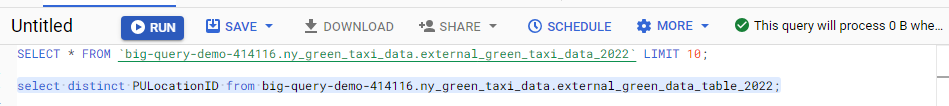
CREATE OR REPLACE TABLE ny\_green\_taxi\_data.green\_taxi\_data\_2022 AS

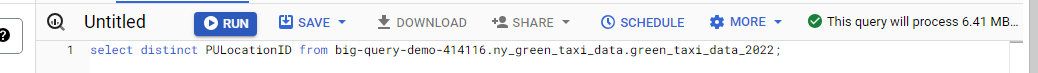
SELECT \* FROM big-query-demo-414116.ny\_green\_taxi\_data.external\_green\_taxi\_data\_2022

Question 1

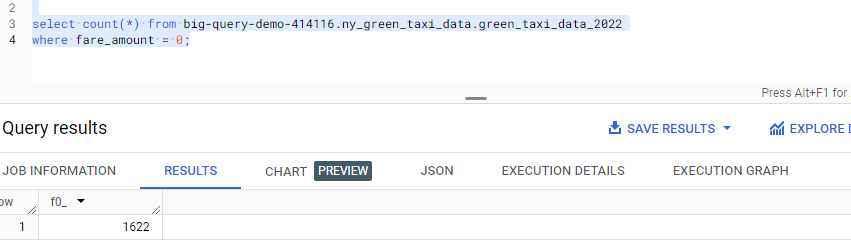


Question 2





Question 3



Question 4

Create or replace table ny\_green\_taxi\_data.green\_taxi\_data\_2022\_partitioned\_clustered

PARTITION BY DATE(LPEP\_PICKUP\_DATETIME)

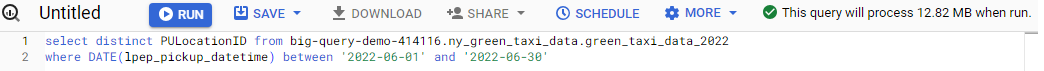
CLUSTER BY VendorID AS

SELECT \* FROM big-query-demo-414116.ny\_green\_taxi\_data.green\_taxi\_data\_2022;

Question 5

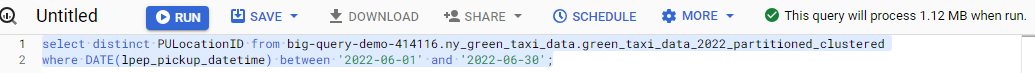
select distinct PULocationID from big-query-demo-414116.ny\_green\_taxi\_data.green\_taxi\_data\_2022

where DATE(lpep\_pickup\_datetime) between '2022-06-01' and '2022-06-30'



select distinct PULocationID from big-query-demo-414116.ny\_green\_taxi\_data.green\_taxi\_data\_2022\_partitioned\_clustered

where DATE(lpep\_pickup\_datetime) between '2022-06-01' and '2022-06-30';



Question 6

Where is the data stored in the External Table you created?

Ans : GCP bucket

Question 7

It is best practice in Big Query to always cluster your data:

Ans: False

Question 8:

Write a SELECT count(\*) query FROM the materialized table you created. How many bytes does it estimate will be read? Why?

select count(\*) from big-query-demo-414116.ny\_green\_taxi\_data.green\_taxi\_data\_2022;

It show 0 bytes as the query will return the results based on the metadata rather than do a scan of the entire table.