Question 1:

Question 1: What is count of records for the 2022 Green Taxi Data??

- 65,623,481
- **840,402**
- 1,936,423
- 253,647

```
SELECT count(*) FROM `mindful-future-412612.ny_taxi.green_taxi_2022`
```

ery results

INFORMATION | RESULTS | CHART | PREVIEW | JSON | EXECUTION

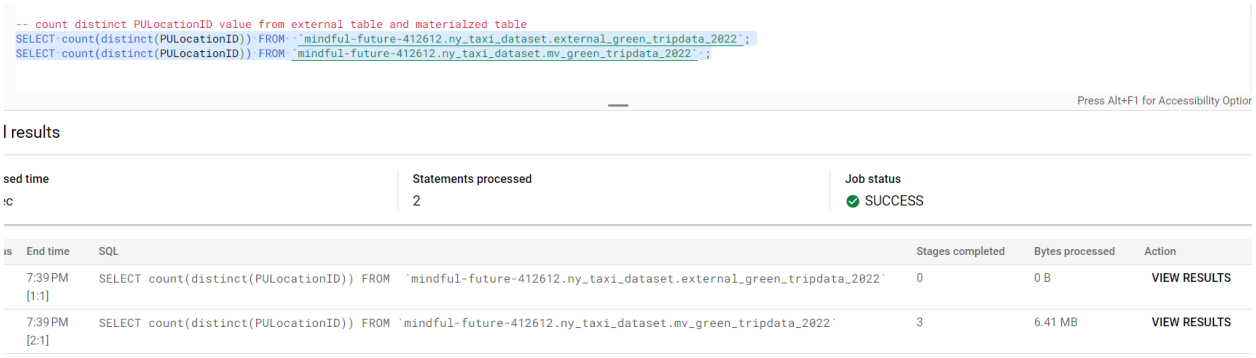
Metadata caching is disabled. You can accelerate queries over external tables by enabling m

f0_
840402

Question 2:

Write a query to count the distinct number of PULocationIDs for the entire dataset on both the tables.
What is the estimated amount of data that will be read when this query is executed on the External Table and the Table?

- **0 MB for the External Table and 6.41MB for the Materialized Table**
- 18.82 MB for the External Table and 47.60 MB for the Materialized Table
- 0 MB for the External Table and 0MB for the Materialized Table
- 2.14 MB for the External Table and 0MB for the Materialized Table



```
-- count distinct PULocationID value from external table and materialzed table
SELECT count(distinct(PULocationID)) FROM `mindful-future-412612.ny_taxi_dataset.external_green_tripdata_2022`;
SELECT count(distinct(PULocationID)) FROM `mindful-future-412612.ny_taxi_dataset.mv_green_tripdata_2022`;
```

Press Alt+F1 for Accessibility Optior

l results

sed time	Statements processed	Job status
:c	2	SUCCESS

ıs	End time	SQL	Stages completed	Bytes processed	Action
	7:39 PM [1:1]	SELECT count(distinct(PULocationID)) FROM `mindful-future-412612.ny_taxi_dataset.external_green_tripdata_2022`	0	0 B	VIEW RESULTS
	7:39 PM [2:1]	SELECT count(distinct(PULocationID)) FROM `mindful-future-412612.ny_taxi_dataset.mv_green_tripdata_2022`	3	6.41 MB	VIEW RESULTS

Question 3:

How many records have a fare_amount of 0?

- 12,488
- 128,219
- 112
- **1,622**

Question 4:

What is the best strategy to make an optimized table in Big Query if your query will always order the results by PUlocationID and filter based on lpep_pickup_datetime? (Create a new table with this strategy)

- Cluster on lpep_pickup_datetime Partition by PUlocationID
- **Partition by lpep_pickup_datetime Cluster on PUlocationID**
- Partition by lpep_pickup_datetime and Partition by PUlocationID
- Cluster on by lpep_pickup_datetime and Cluster on PUlocationID

Question 5:

Write a query to retrieve the distinct PULocationID between lpep_pickup_datetime 06/01/2022 and 06/30/2022 (inclusive)

Use the materialized table you created earlier in your from clause and note the estimated bytes. Now change the table in the from clause to the partitioned table you created for question 4 and note the estimated bytes processed. What are these values?

Choose the answer which most closely matches.

- 22.82 MB for non-partitioned table and 647.87 MB for the partitioned table
- **12.82 MB for non-partitioned table and 1.12 MB for the partitioned table**
- 5.63 MB for non-partitioned table and 0 MB for the partitioned table
- 10.31 MB for non-partitioned table and 10.31 MB for the partitioned table

```
-- Impact of partition
-- Scanning 1.6GB of data
SELECT DISTINCT(PULocationID)
FROM `mindful-future-412612.ny_taxi_dataset.green_trips_non_partitioned`
WHERE lpep_pickup_date BETWEEN '2022-06-01' AND '2022-06-30';

-- Scanning ~106 MB of DATA
SELECT DISTINCT(PULocationID)
FROM `mindful-future-412612.ny_taxi_dataset.green_trips_partitioned`
WHERE lpep_pickup_date BETWEEN '2022-06-01' AND '2022-06-30';
```

Press Alt+F1 for Accessibility Optior

All results

Elapsed time	Statements processed	Job status
1 sec	2	SUCCESS

Status	End time	SQL	Stages completed	Bytes processed	Action
✓	7:30 PM [1:1]	SELECT DISTINCT(PULocationID)	2	12.82 MB	VIEW RESULTS
✓	7:30 PM [6:1]	SELECT DISTINCT(PULocationID)	2	1.12 MB	VIEW RESULTS

Question 6:

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

- Big Query
- **GCP Bucket**
- Big Table
- Container Registry

Question 7:

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

- **True**
- False

(Bonus: Not worth points) Question 8:

No Points: Write a `SELECT count(*)` query FROM the materialized table you created. How many bytes does it estimate will be read? Why?

ANS: 0 bytes. Because `SELECT count(*)` query FROM the materialized only required retrieving the count metadata from the table's metadata rather than scanning the entire dataset