GOOGLE CLOUD PLATFORM

ASSIGNMENT

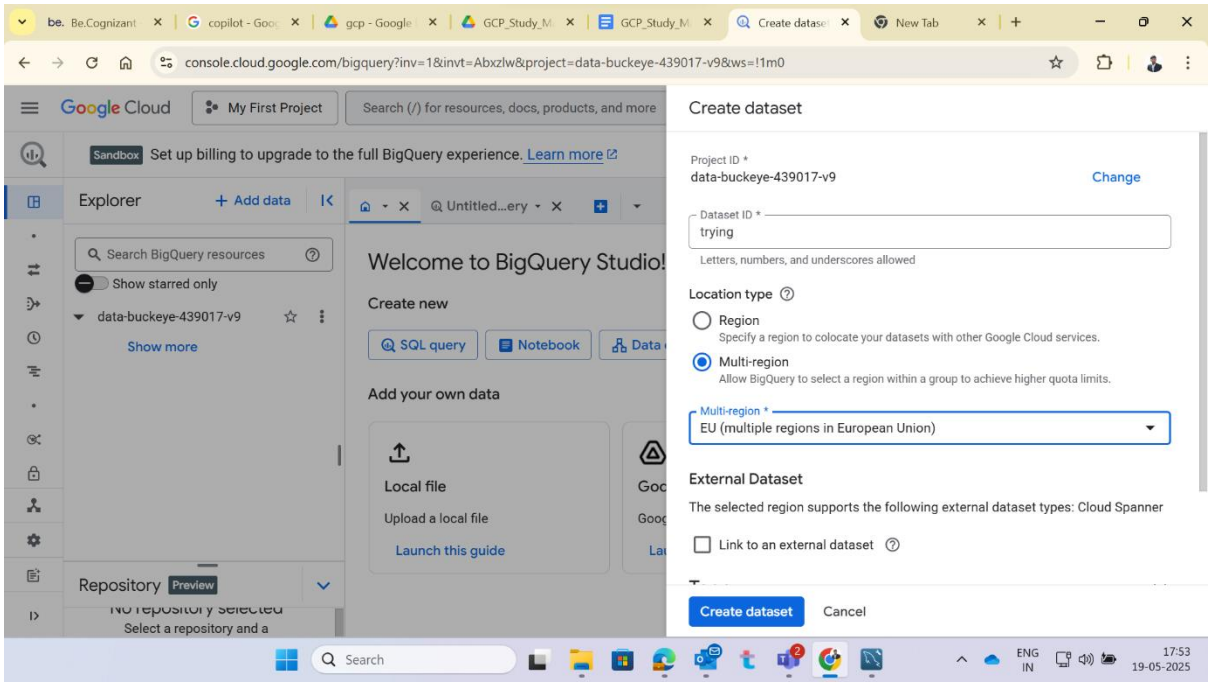
Name: Jebastin P
Employee ID: 2401171
Cohort Code: INTAIA25GCP003

Questions:

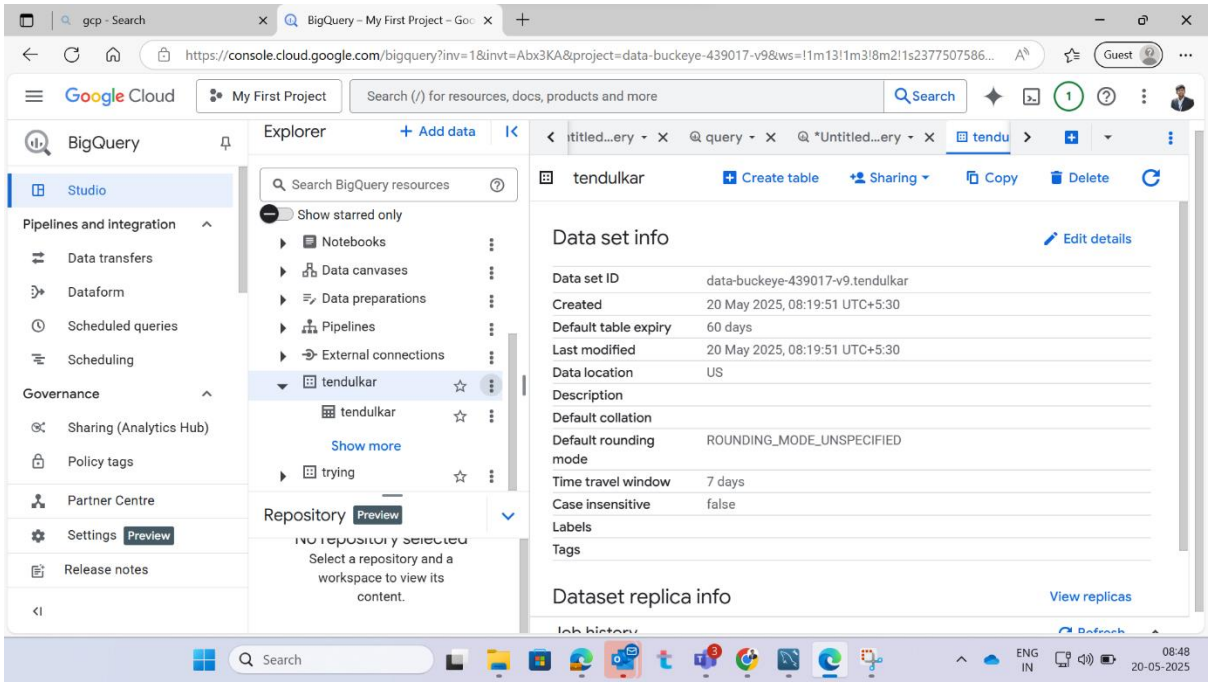
1. Sum of runs as Opener vs Other batting positions
2. Runs scored against Test playing Nation vs Non Test Playing Nation
3. Average Runs before opening for first time and after

Procedure:

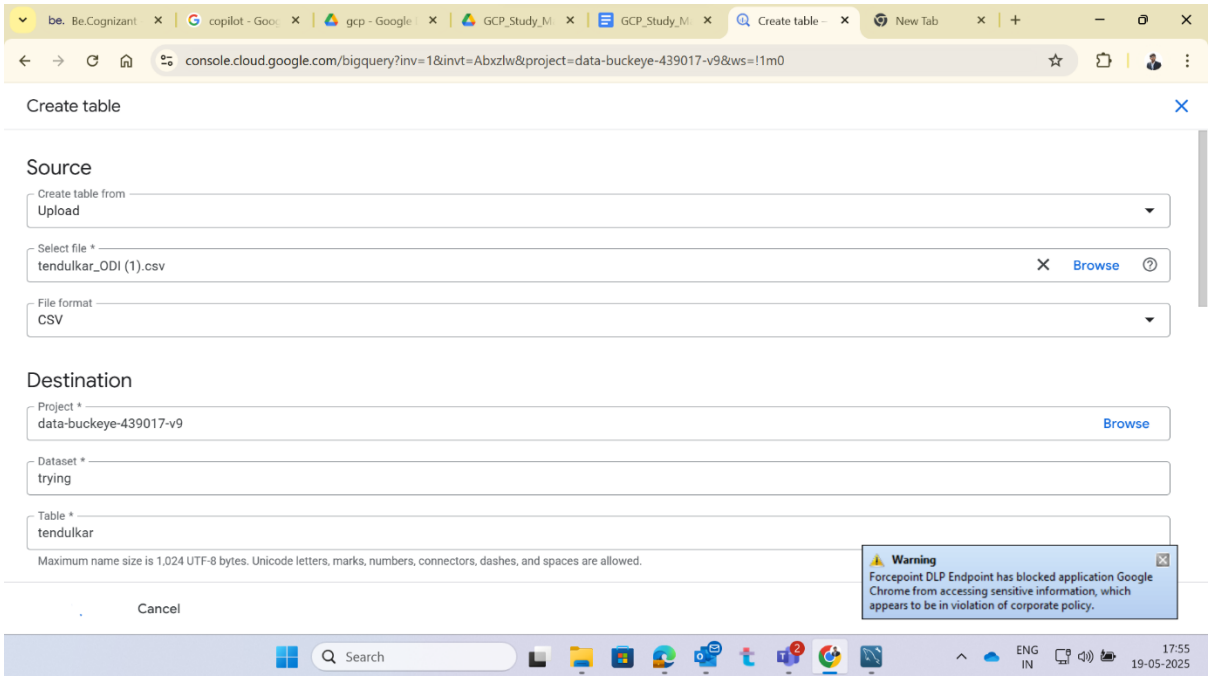
Step 1 : Opening Big Query and creating a dataset.



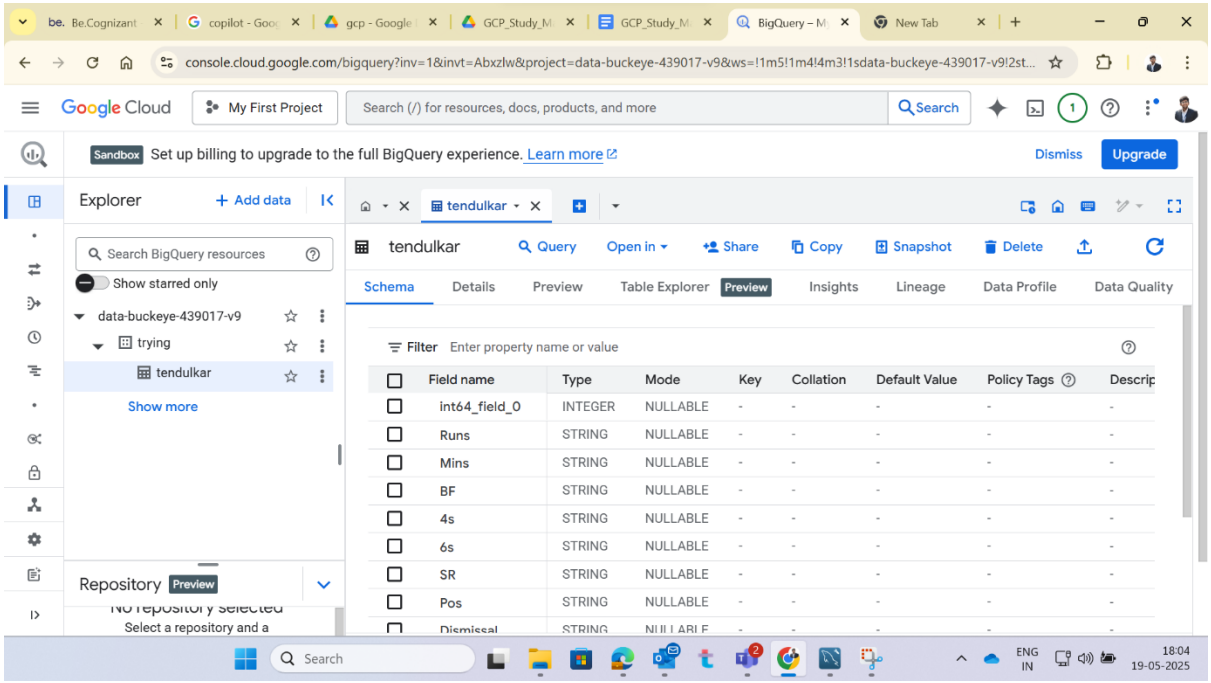
Step 2: Dataset Info



Step3: Creating a table by uploading the given dataset in the Big query:

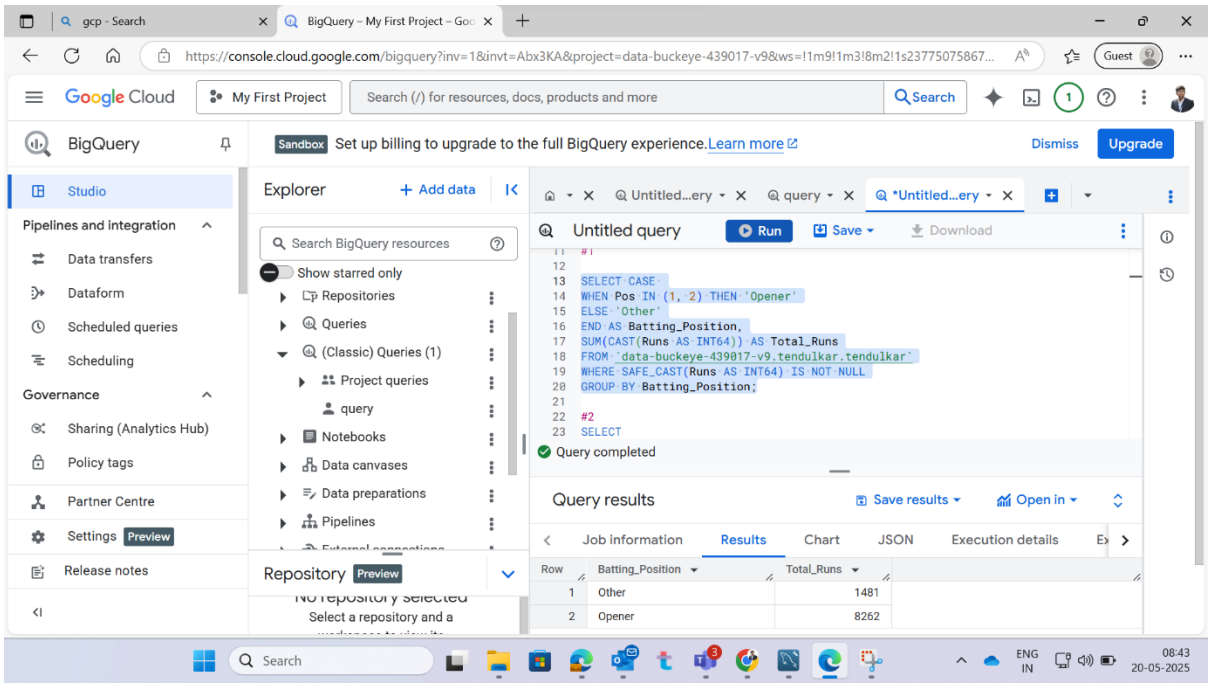


Step 4: These are the schema and details of the uploaded dataset

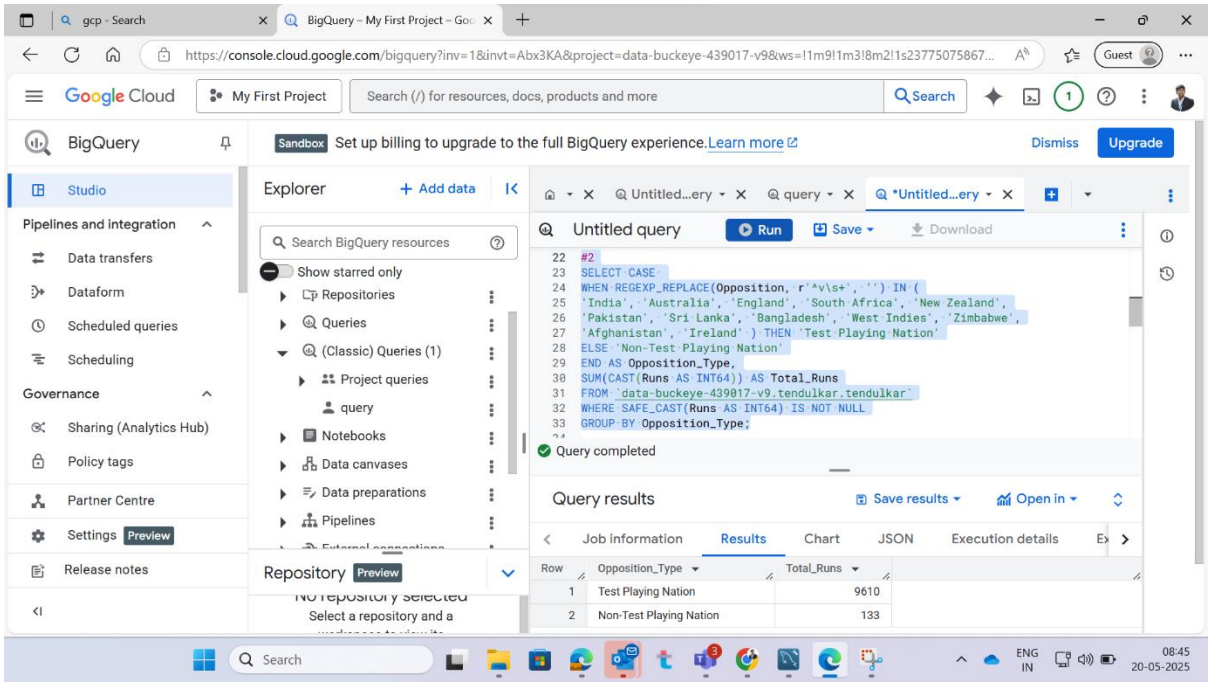


Step 5: Query space for data retrieval

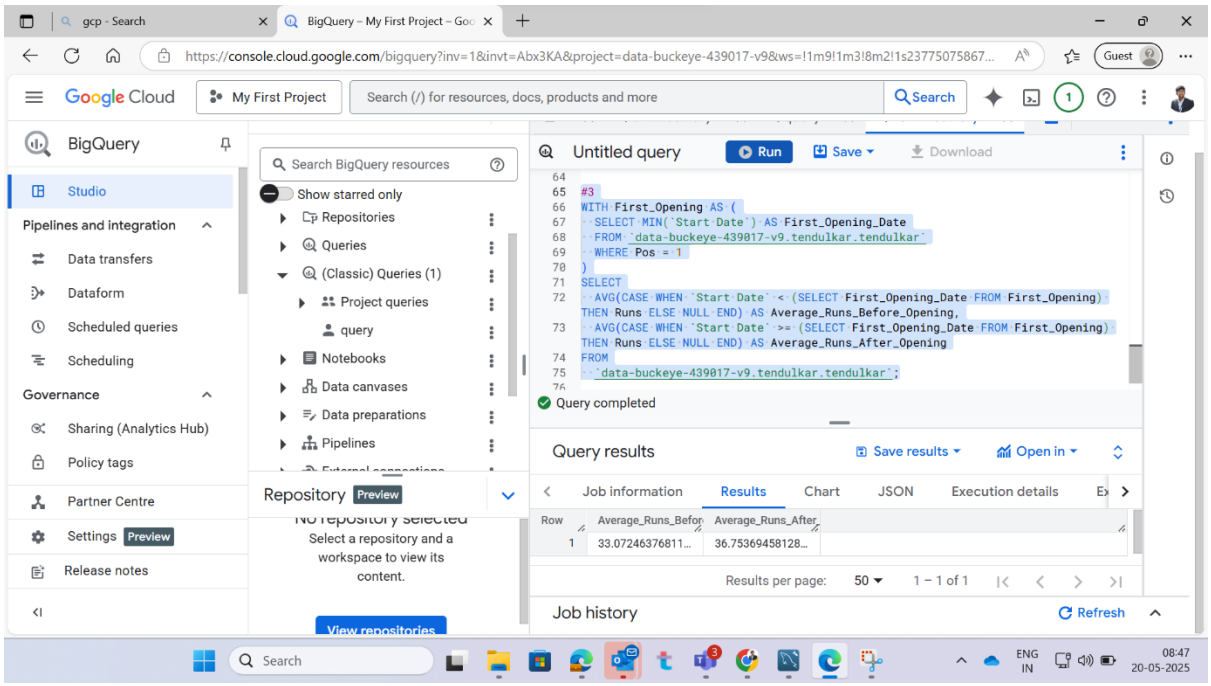
5.1 To display the sum of runs as Opener vs other batting positions



5.2 To display the runs scored against Test playing Nation vs Non-Test Playing Nation



5.3 To display the average runs before opening for first time and after

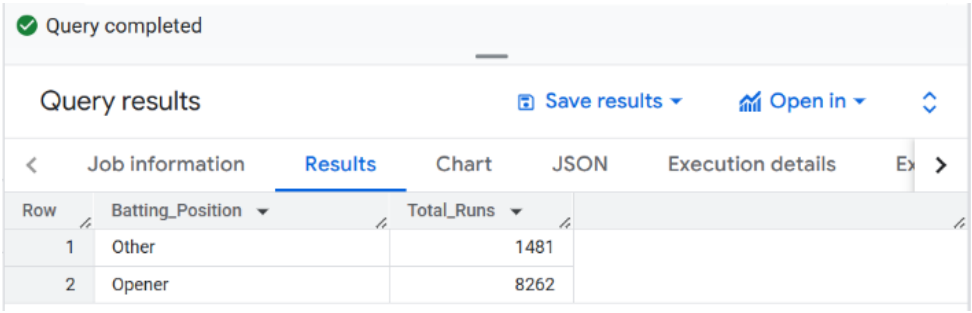


Query:

5.1 To display the sum of runs as Opener vs other batting positions

```
SELECT CASE
WHEN Pos IN (1, 2) THEN 'Opener'
ELSE 'Other'
END AS Batting_Position,
SUM(CAST(Runs AS INT64)) AS Total_Runs
FROM `data-buckeye-439017-v9.tendulkar.tendulkar`
WHERE SAFE_CAST(Runs AS INT64) IS NOT NULL
GROUP BY Batting_Position;
```

Solution:

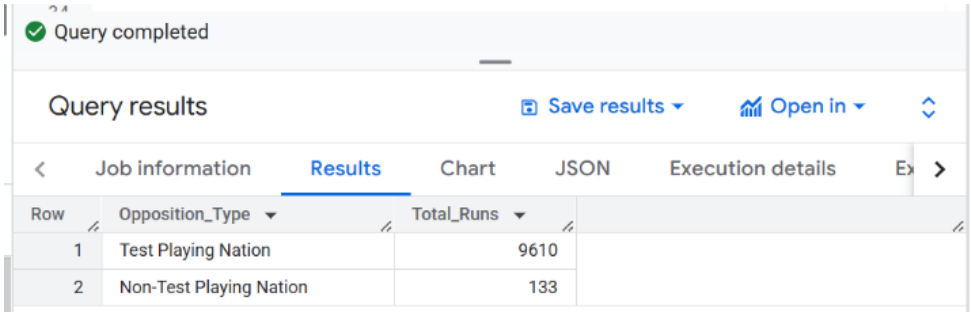


5.2 To display the runs scored against Test playing Nation vs Non Test Playing Nation

```
SELECT CASE
WHEN REGEXP_REPLACE(Opposition, r'^v\s+', '') IN (
'India', 'Australia', 'England', 'South Africa', 'New Zealand',
'Pakistan', 'Sri Lanka', 'Bangladesh', 'West Indies', 'Zimbabwe',
'Afghanistan', 'Ireland' ) THEN 'Test Playing Nation'
ELSE 'Non-Test Playing Nation'
END AS Opposition_Type,
SUM(CAST(Runs AS INT64)) AS Total_Runs
FROM `data-buckeye-439017-v9.tendulkar.tendulkar`
WHERE SAFE_CAST(Runs AS INT64) IS NOT NULL
```

GROUP BY Opposition_Type;

Solution:



5.3 To display the average runs before opening for first time and after

```
WITH First_Opening AS (
  SELECT MIN(`Start Date`) AS First_Opening_Date
  FROM `data-buckeye-439017-v9.tendulkar.tendulkar`
  WHERE Pos = 1
)
SELECT
  AVG(CASE WHEN `Start Date` < (SELECT First_Opening_Date FROM First_Opening) THEN Runs ELSE NULL END) AS Average_Runs_Before_Opening,
  AVG(CASE WHEN `Start Date` >= (SELECT First_Opening_Date FROM First_Opening) THEN Runs ELSE NULL END) AS Average_Runs_After_Opening
FROM
  `data-buckeye-439017-v9.tendulkar.tendulkar`;
```

Solution: