

GOOGLE CLOUD PLATFORM

ASSIGNMENT

Name: Jebastin P

Employee ID: 2401171

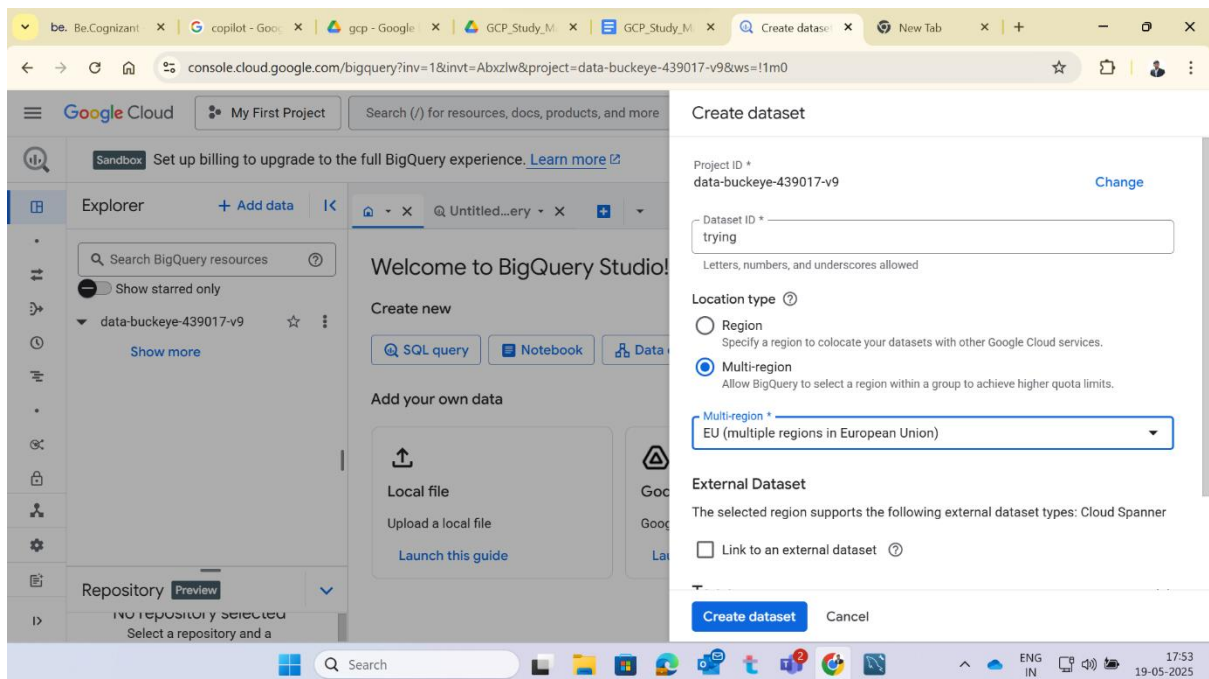
Cohort Code: INTAIA25GCP003

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

The screenshot shows the Google Cloud BigQuery console interface. On the left, the 'Studio' sidebar is visible with various navigation options. The main area displays the 'tendulkar' dataset information. The 'Data set info' section includes the following details:

Field	Value
Data set ID	data-buckeye-439017-v9.tendulkar
Created	20 May 2025, 08:19:51 UTC+5:30
Default table expiry	60 days
Last modified	20 May 2025, 08:19:51 UTC+5:30
Data location	US
Description	
Default collation	
Default rounding mode	ROUNDING_MODE_UNSPECIFIED
Time travel window	7 days
Case insensitive	false
Labels	
Tags	

Below the dataset info, there is a 'Dataset replica info' section with a 'View replicas' link. The Explorer on the left shows the 'tendulkar' dataset selected under 'External connections'.

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

The screenshot shows the 'Create table' dialog in the Google Cloud BigQuery console. The 'Source' section is set to 'Upload' with the file 'tendulkar_ODI (1).csv' selected. The 'File format' is set to 'CSV'. The 'Destination' section shows the 'Project' as 'data-buckeye-439017-v9', the 'Dataset' as 'trying', and the 'Table' as 'tendulkar'. A warning message is displayed at the bottom right:

Warning
Forcepoint DLP Endpoint has blocked application Google Chrome from accessing sensitive information, which appears to be in violation of corporate policy.

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

The screenshot shows the Google Cloud BigQuery console. The left sidebar displays the Explorer with the 'tendulkar' dataset selected under the 'data-buckeye-439017-v9' project. The main panel shows the 'Schema' tab for the 'tendulkar' dataset. The schema table lists the following fields:

Field name	Type	Mode	Key	Collation	Default Value	Policy Tags	Description
int64_field_0	INTEGER	NULLABLE	-	-	-	-	-
Runs	STRING	NULLABLE	-	-	-	-	-
Mins	STRING	NULLABLE	-	-	-	-	-
BF	STRING	NULLABLE	-	-	-	-	-
4s	STRING	NULLABLE	-	-	-	-	-
6s	STRING	NULLABLE	-	-	-	-	-
SR	STRING	NULLABLE	-	-	-	-	-
Pos	STRING	NULLABLE	-	-	-	-	-
Niemeical	STRING	NULLABLE	-	-	-	-	-

Step 5: Query space for data retrieval

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

The screenshot shows the Google Cloud BigQuery console in the 'Studio' view. A query is being executed, and the results are displayed in a table. The query is as follows:

```
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;
```

The query results table shows the following data:

Row	Batting_Position	Total_Runs
1	Other	1481
2	Opener	8262

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

The screenshot shows the Google Cloud BigQuery Studio interface. The Explorer pane on the left shows the project structure. The main editor displays a SQL query (Query #2) that filters runs based on the nation type. The query results are displayed in a table with two rows.

```
22 #2
23 SELECT CASE
24   WHEN REGEXP_REPLACE(Opposition, 'r'\s+', '') IN (
25     'India', 'Australia', 'England', 'South Africa', 'New Zealand',
26     'Pakistan', 'Sri Lanka', 'Bangladesh', 'West Indies', 'Zimbabwe',
27     'Afghanistan', 'Ireland') THEN 'Test Playing Nation'
28   ELSE 'Non-Test Playing Nation'
29 END AS Opposition_Type,
30 SUM(CAST(Runs AS INT64)) AS Total_Runs
31 FROM `data-buckeye-439017-v9.tendulkar.tendulkar`
32 WHERE SAFE_CAST(Runs AS INT64) IS NOT NULL
33 GROUP BY Opposition_Type;
```

Row	Opposition_Type	Total_Runs
1	Test Playing Nation	9610
2	Non-Test Playing Nation	133

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

The screenshot shows the Google Cloud BigQuery Studio interface. The Explorer pane on the left shows the project structure. The main editor displays a SQL query (Query #3) that calculates the average runs before and after the first opening. The query results are displayed in a table with one row.

```
64 #3
65 WITH First_Opening AS (
66   SELECT MIN('Start Date') AS First_Opening_Date
67   FROM `data-buckeye-439017-v9.tendulkar.tendulkar`
68   WHERE Pos = 1
69 )
70 SELECT
71   AVG(CASE WHEN 'Start Date' < (SELECT First_Opening_Date FROM First_Opening)
72     THEN Runs ELSE NULL END) AS Average_Runs_Before_Opening,
73   AVG(CASE WHEN 'Start Date' >= (SELECT First_Opening_Date FROM First_Opening)
74     THEN Runs ELSE NULL END) AS Average_Runs_After_Opening
75 FROM `data-buckeye-439017-v9.tendulkar.tendulkar`;
```

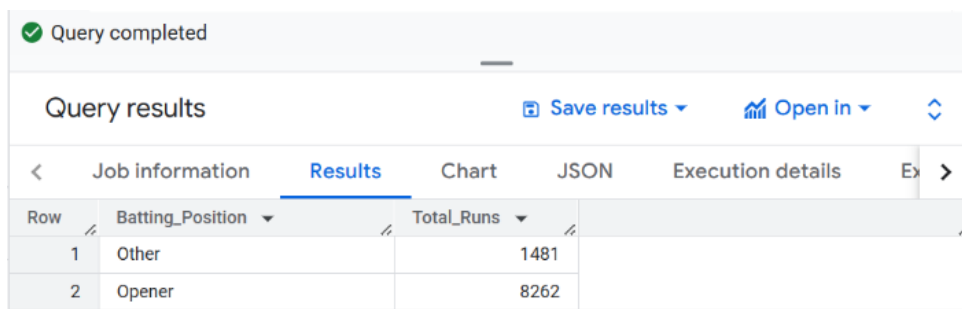
Row	Average_Runs_Before	Average_Runs_After
1	33.07246376811...	36.75369458128...

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:



Query completed

Query results [Save results](#) [Open in](#)

< Job information **Results** Chart JSON Execution details Ex >

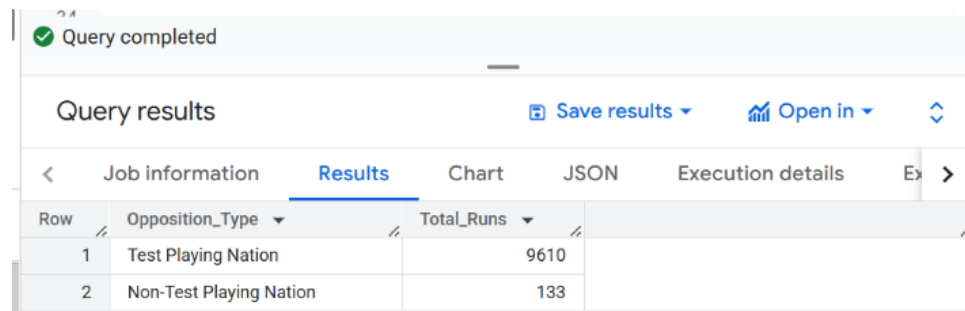
Row	Batting_Position	Total_Runs
1	Other	1481
2	Opener	8262

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:



Query completed

Query results

Save results Open in

Job information Results Chart JSON Execution details

Row	Opposition_Type	Total_Runs
1	Test Playing Nation	9610
2	Non-Test Playing Nation	133

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

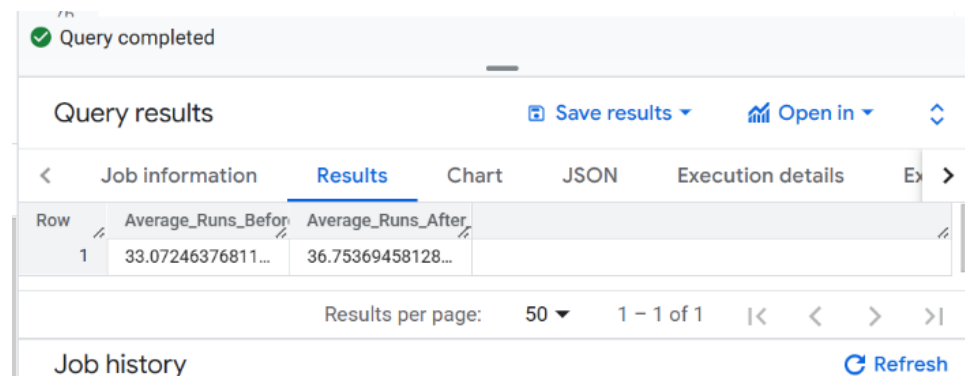
AVG(CASE WHEN 'Start Date' < (SELECT First_Opening_Date FROM First_Opening)
THEN Runs ELSE NULL END) AS Average_Runs_Before_Opening,

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:



Query completed

Query results

Save results Open in

Job information Results Chart JSON Execution details

Row	Average_Runs_Before	Average_Runs_After
1	33.07246376811...	36.75369458128...

Results per page: 50 1 - 1 of 1

Job history Refresh