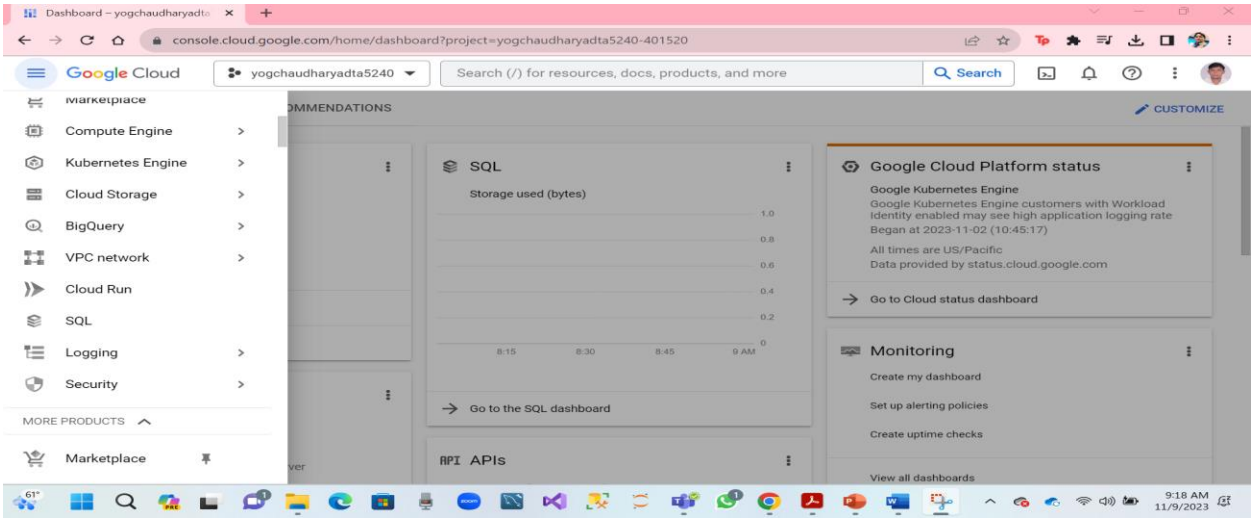
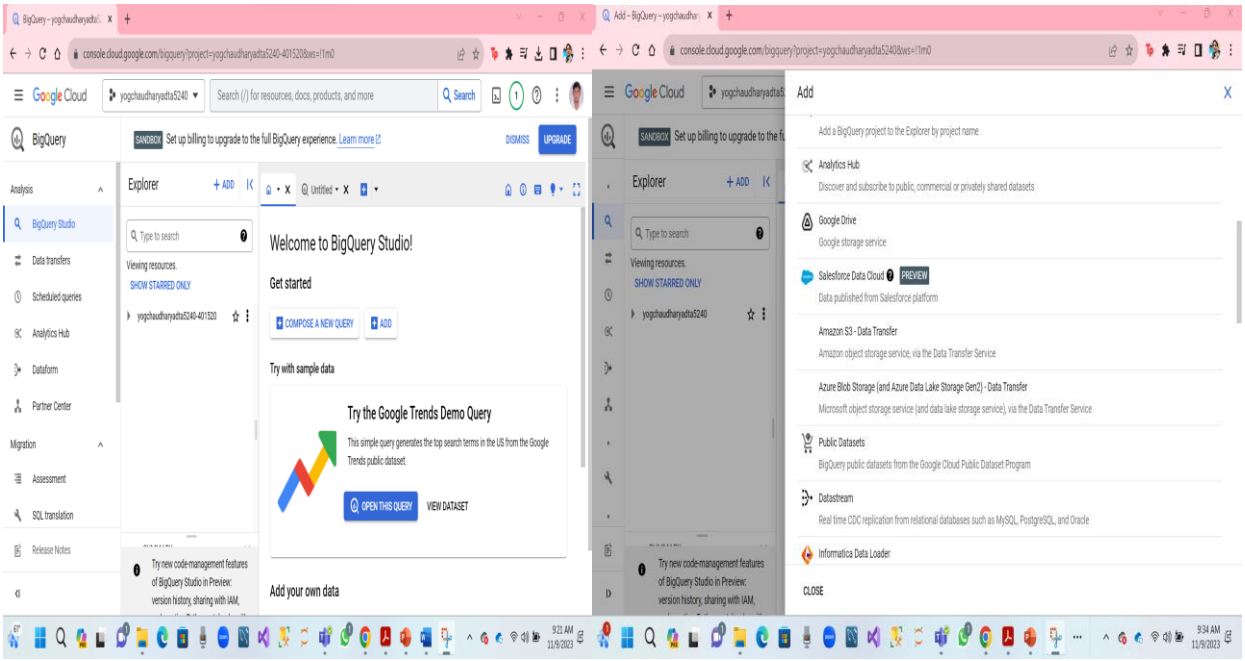
Yog Chaudhary
11727095
ADTA 5240 Week (Harvesting, Storing, And Retrieving Data)
Professor: Dr. Zeynep Orhan
University Of North Texas **Nov 09, 2023**

- **Big query Queries:**

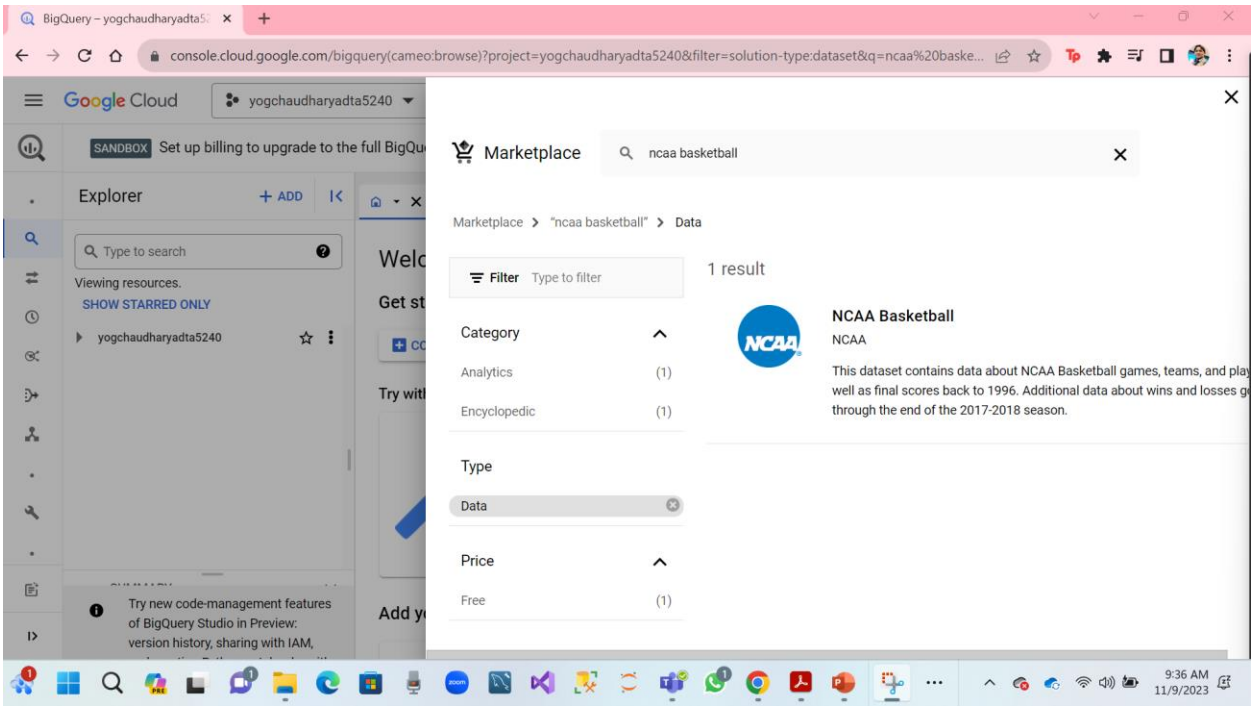
Using the public Big Query dataset. The NCCA basketball data. This dataset contains data about the NCCA basketball games, items, and players.

- First, I opened my GCP account and clicked on the navigation panel, then selected bid query.
- After clicking on that I was taking on this below page

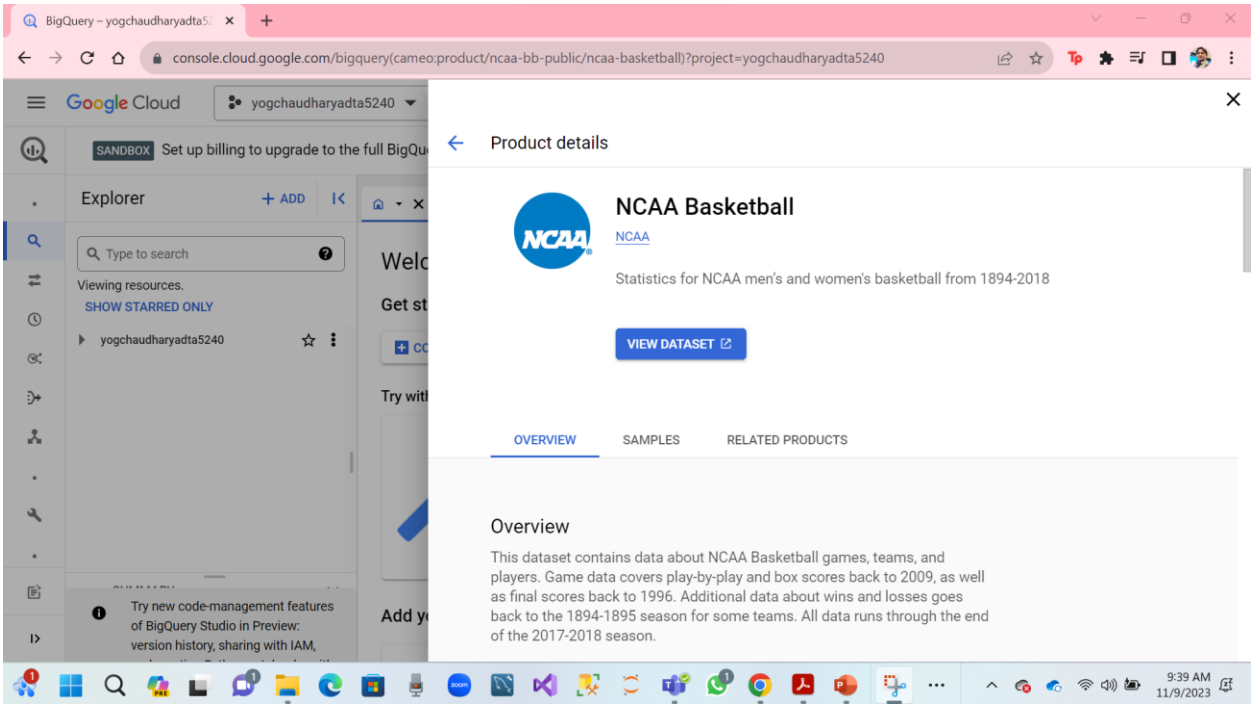




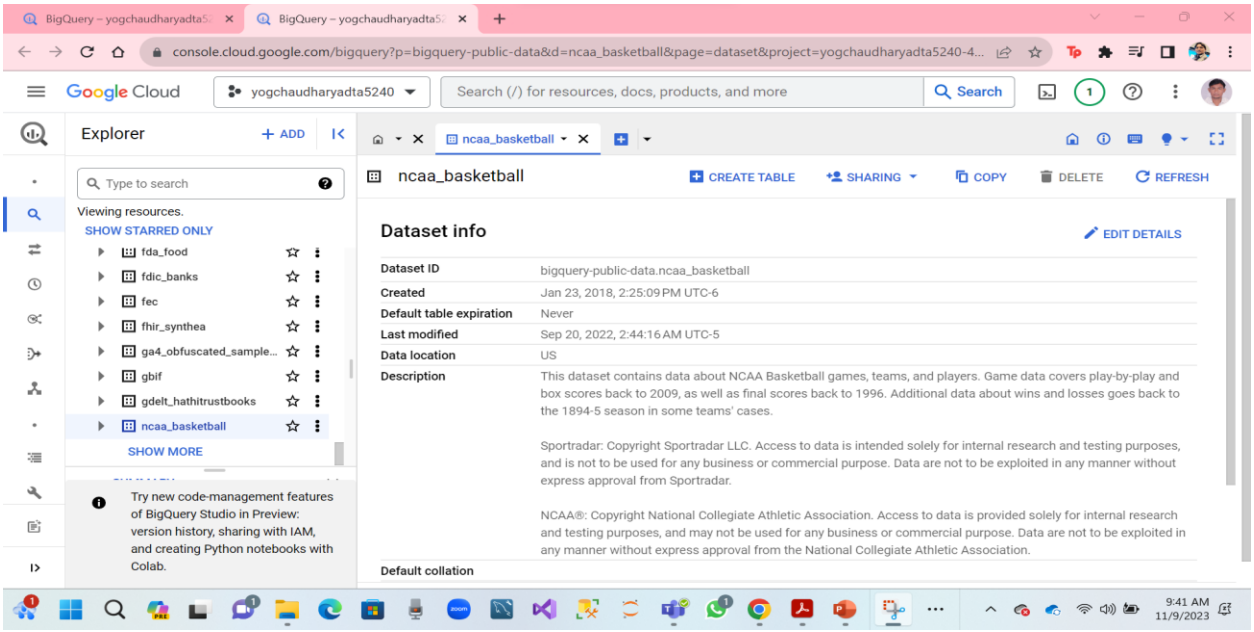
- I clicked on add data. Which has taken in this page. Then I selected a public dataset.
- Below is the screenshot of that.



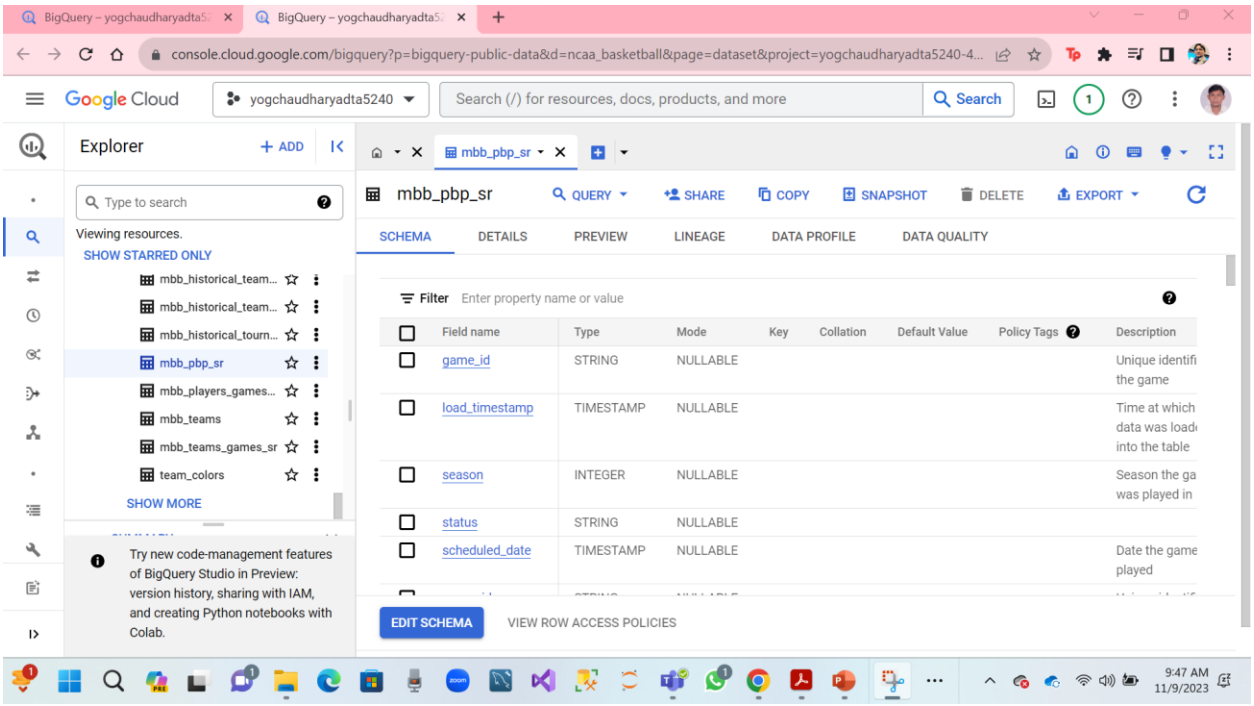
- Then I clicked on the NCAA basketball dataset.



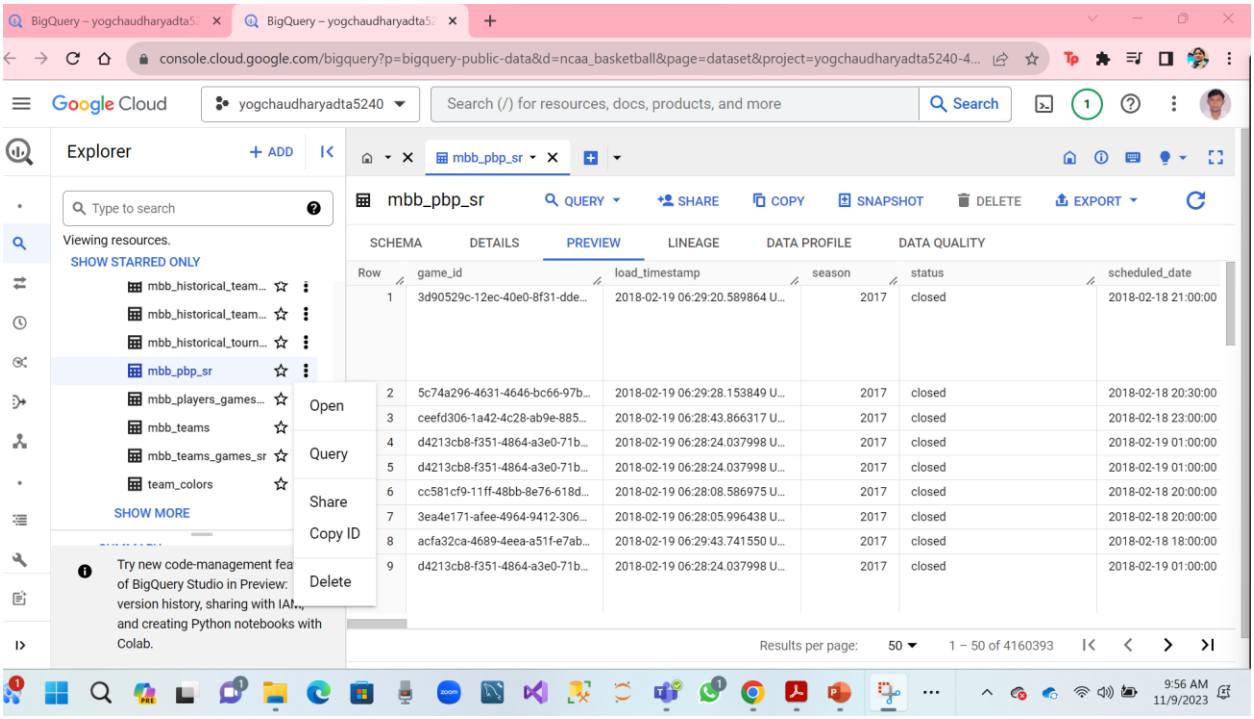
- Hare, I clicked on view dataset.



- Then I expanded the NCA basketball dataset.
- After, I click mbb_pbp_sr
- Click on open. It showed the schema of the predefined table.
- Here is a screenshot of that.



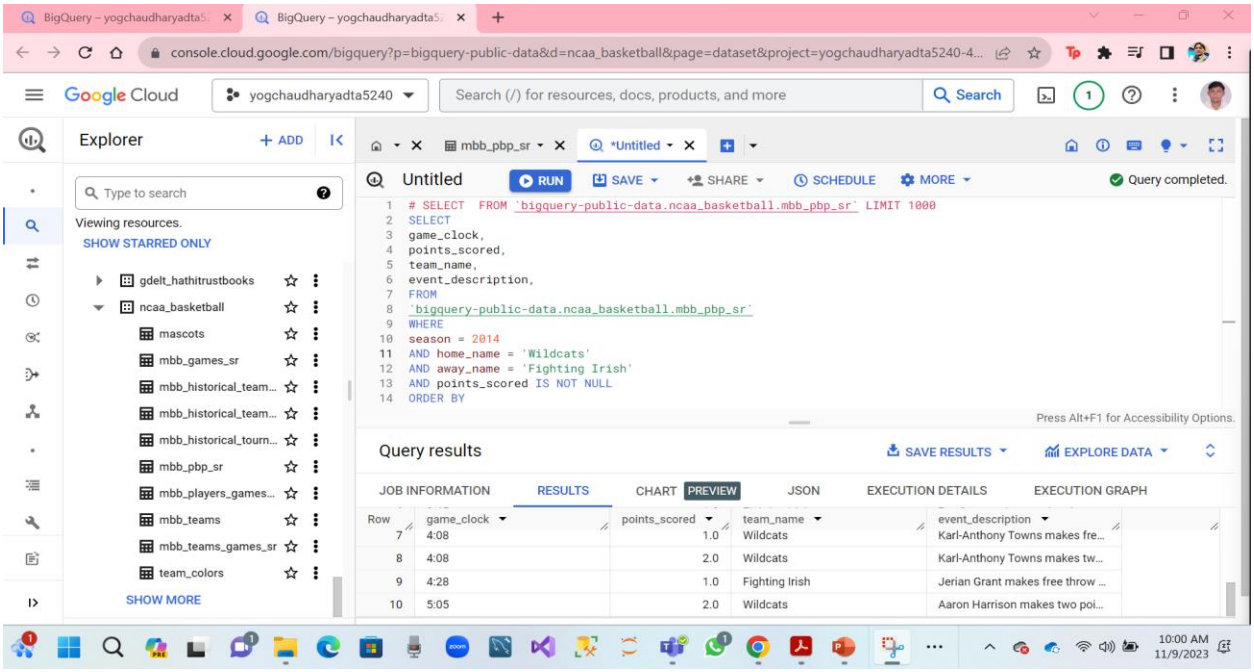
- After, I clicked preview to see more information about the dataset.



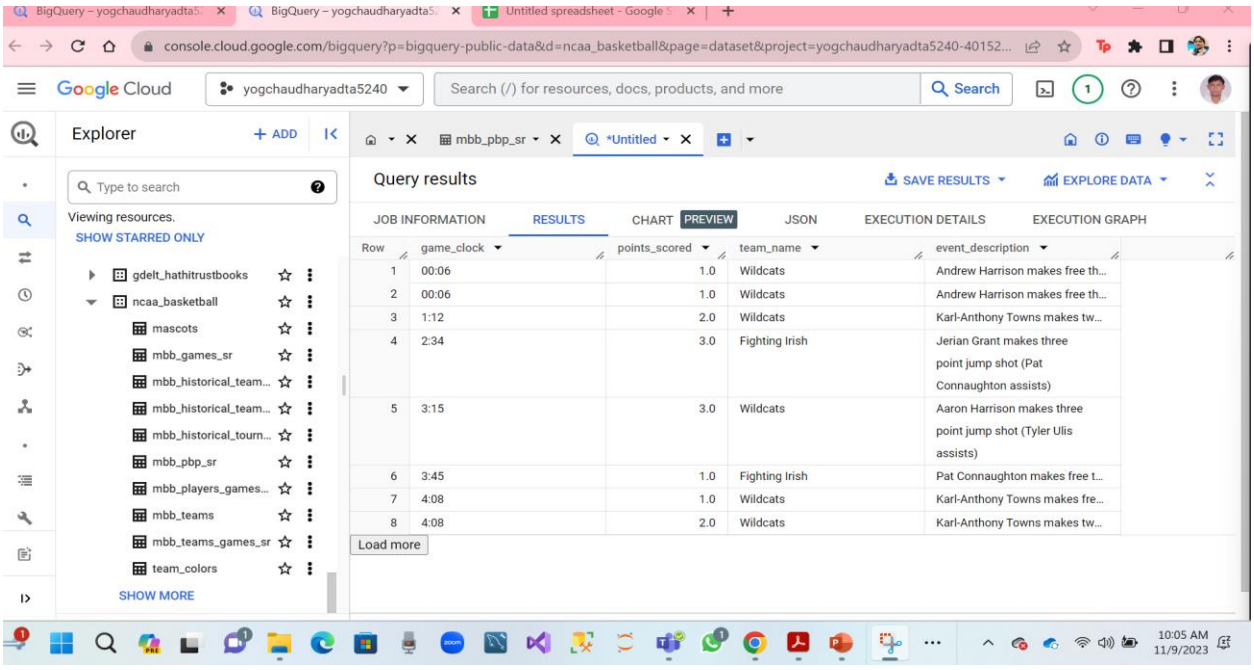
- I clicked on 3 dots in mbb_pbp_sr and clicked on Query.
- I have used SQL to query the dataset.
- For selecting the columns, I have used this code.
- In this have used SELECT, WHERE, ORDER BY, LIMIT statements.

```
SELECT
game_clock,
points_scored,
team_name,
event_description,
FROM
`bigquery-public-data.ncaa_basketball.mbb_pbp_sr`
WHERE
season = 2014
AND home_name = 'Wildcats'
AND away_name = 'Fighting Irish'
AND points_scored IS NOT NULL
ORDER BY
timestamp DESC
LIMIT
10;
```

- Here is a screenshot of that.

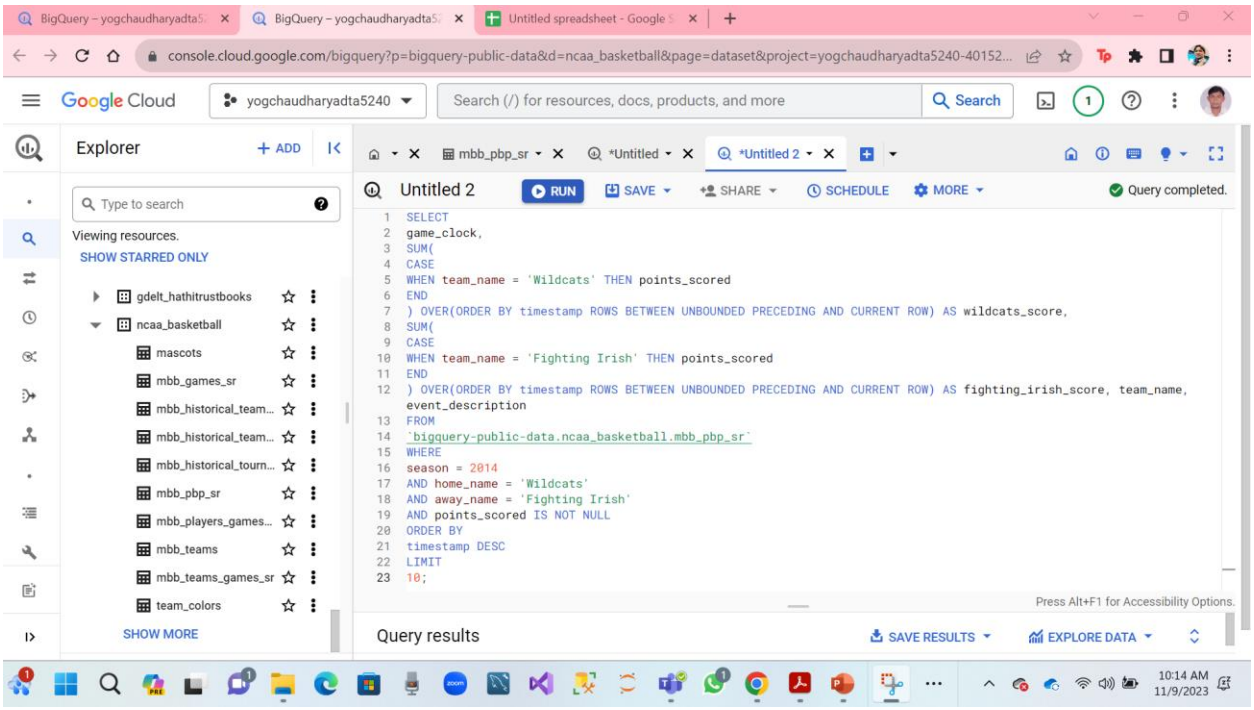


- Clicked on run to run the code. After clicking on that it shows the result.

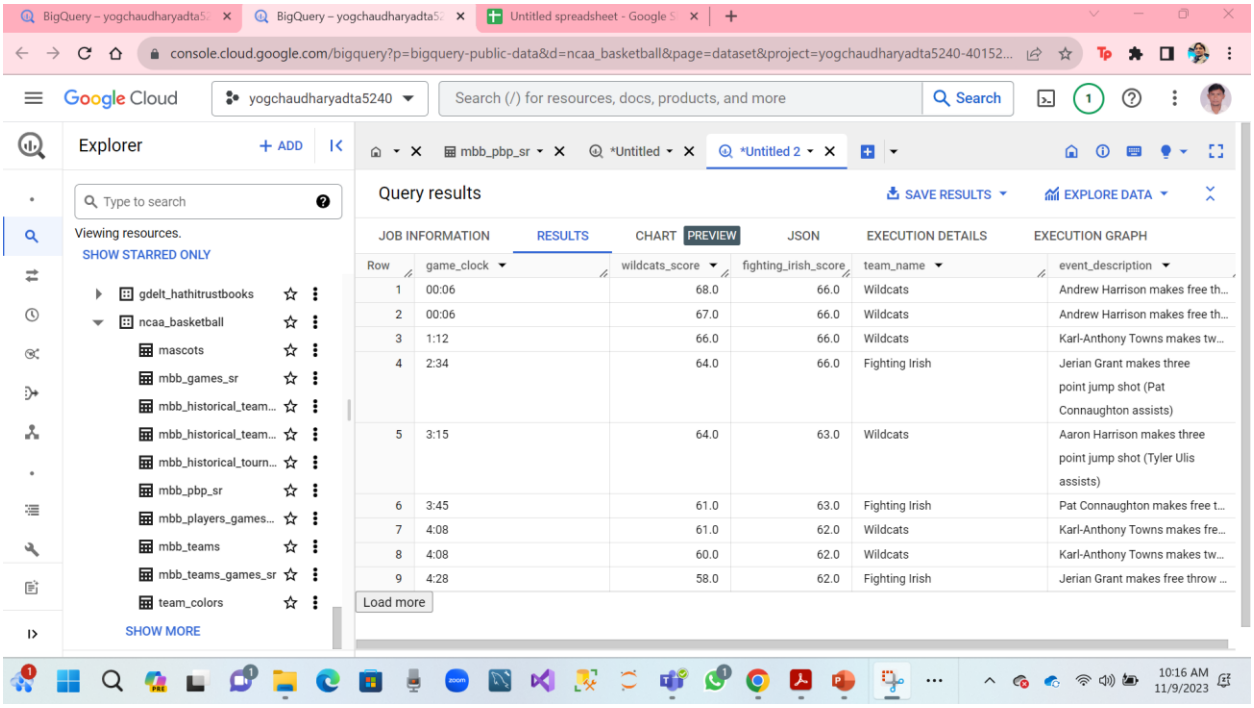


- Now I am changing the SQL code.
- By calculating the cumulative sum of scores for each team in the games.
- For clicked on composer new query sheet
- Run the query by adding two new columns: **wildcats_score and fighting_irish_score**

```
SELECT
game_clock,
SUM(
CASE
WHEN team_name = 'Wildcats' THEN points_scored
END
) OVER(ORDER BY timestamp ROWS BETWEEN UNBOUNDED PRECEDING AND CURRENT ROW) AS
wildcats_score,
SUM(
CASE
WHEN team_name = 'Fighting Irish' THEN points_scored
END
) OVER(ORDER BY timestamp ROWS BETWEEN UNBOUNDED PRECEDING AND CURRENT ROW) AS
fighting_irish_score, team_name, event_description
FROM
`bigquery-public-data.ncaa_basketball.mbb_pbp_sr`
WHERE
season = 2014
AND home_name = 'Wildcats'
AND away_name = 'Fighting Irish'
AND points_scored IS NOT NULL
ORDER BY
timestamp DESC
LIMIT
10;
```



- Click on run to run the code. After clicking on that it shows results.



- Hence, Finally, we can see the score throughout the game by using this code.