PROJECT: NBA DATASET CLEANING AND EXPORT TO CSV USING BIGQUERY SQL.

INTRODUCTION:

In this task, we performed data cleaning on the NBA dataset using BigQuery SQL. The cleaned dataset was then exported to a CSV file for further analysis or reporting.

DATASET OVERVIEW:

The original dataset contained player information such as Name, Team, College, Salary, Height, Weight, Age, and Position. Some rows had missing values and duplicate entries.

DATA CLEANING PROCESS:

STEP 1: Create a Dataset "nba_dataset"

STEP 2: Create a Table name as "Test_nba"

STEP 3: Give SQL Query to Cleaning the dataset

```sql
#head()
SELECT *
FROM `careful-synapse-461006-b5.test_dataset.test_nba`
LIMIT 5;
#tail()
SELECT *
FROM `careful-synapse-461006-b5.test_dataset.test_nba`
ORDER BY Name DESC
LIMIT 5;
#info()
  SELECT
  column_name,
  data_type,
  is_nullable
FROM `careful-synapse-461006-b5.test_dataset.INFORMATION_SCHEMA.COLUMNS`
WHERE table_name = 'test_nba';
#describe()
SELECT
  COUNT(*) AS total_rows,
  COUNT(Salary) AS salary_count,
  MIN(SAFE_CAST(Salary AS FLOAT64)) AS min_salary,
  MAX(SAFE_CAST(Salary AS FLOAT64)) AS max_salary,
  AVG(SAFE_CAST(Salary AS FLOAT64)) AS avg_salary,
  STDDEV(SAFE_CAST(Salary AS FLOAT64)) AS std_salary
FROM `careful-synapse-461006-b5.test_dataset.test_nba`;
#Null count per column
SELECT
```

```sql
  COUNTIF(Name IS NULL OR TRIM(Name) = '') AS null_names,
  COUNTIF(Team IS NULL OR TRIM(Team) = '') AS null_teams,
  COUNTIF(College IS NULL OR TRIM(College) = '') AS null_colleges,
  COUNTIF(SAFE_CAST(Salary AS FLOAT64) IS NULL) AS invalid_salaries
FROM `careful-synapse-461006-b5.test_dataset.test_nba`;
#Count of players by college()
SELECT College, COUNT(*) AS count
FROM `careful-synapse-461006-b5.test_dataset.test_nba`
GROUP BY College
ORDER BY count DESC;
#Duplicate()
SELECT Name, Team, College, Salary, COUNT(*) AS duplicate_count
FROM `careful-synapse-461006-b5.test_dataset.test_nba`
GROUP BY Name, Team, College, Salary
HAVING COUNT(*) > 1;
#Standardize Text Columns
SELECT
  INITCAP(TRIM(Name)) AS Name,
  INITCAP(TRIM(Team)) AS Team,
  INITCAP(TRIM(College)) AS College,
  SAFE_CAST(Salary AS FLOAT64) AS Salary
FROM `careful-synapse-461006-b5.test_dataset.test_nba`;
#Convert Salary to Numeric Format
SELECT
  SAFE_CAST(Salary AS FLOAT64) AS salary_numeric
FROM `careful-synapse-461006-b5.test_dataset.test_nba`;
```

Query completed

OUTPUT:

All results

Elapsed time	Statements processed	Job status
1 sec	4	SUCCESS

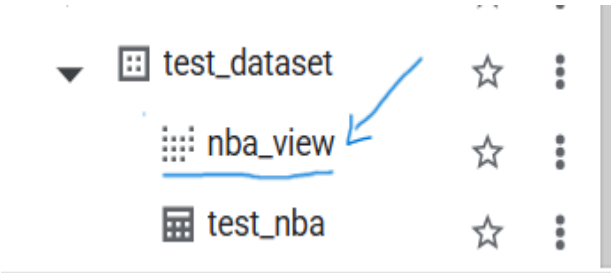
Status	End time	SQL	Stages completed	Bytes p	Action
	2:18 PM [2:1]	SELECT *	0	0 B	View results
	2:18 PM [6:1]	SELECT *	0	0 B	View results
	2:18 PM [11:3]	SELECT	1	10 MB	View results
	2:18 PM [18:1]	SELECT	2	3.48 KB	View results

STEP 4: Create View query to create view in table.

Untitled query Run Save Downloa

```
CREATE VIEW test_dataset.nba_view AS
SELECT *
 FROM `careful-synapse-461006-b5.test_dataset.test_nba`
 WHERE College IS NOT NULL
```

AFTER CREATED IT SHOULD BE:



STEP 5: VIEW THE "NBA_VIEW":

The view should show the cleaned data, not the raw data, so we can easily write queries to explore the entire process.

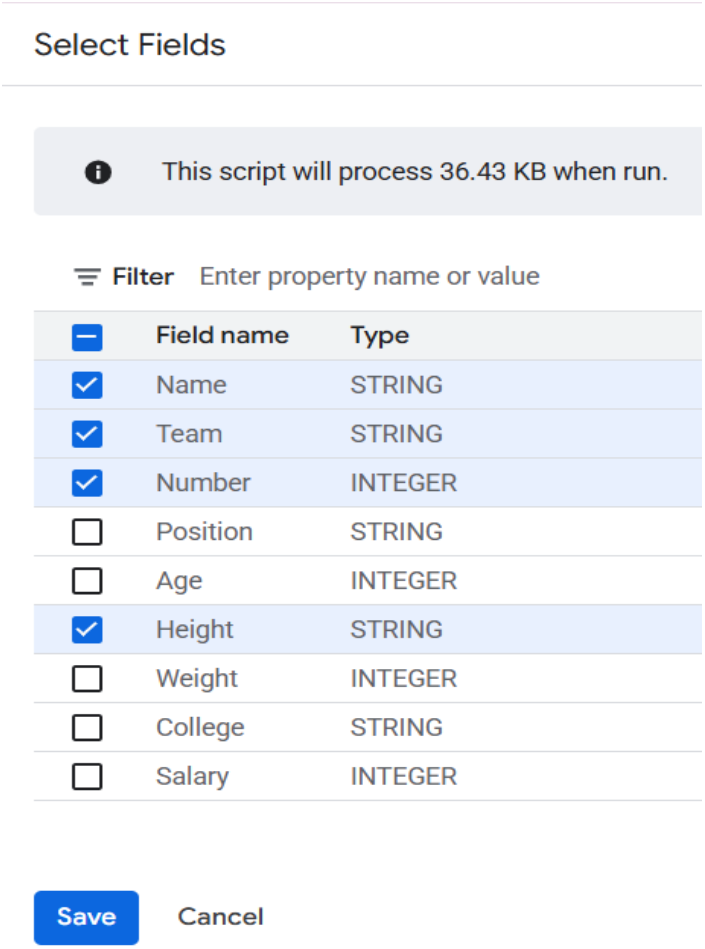
- **CLICK THE PREVIEW**

Schema | Details | Table Explorer | Preview | Insights | Lineage | Data Profile | Data Quality

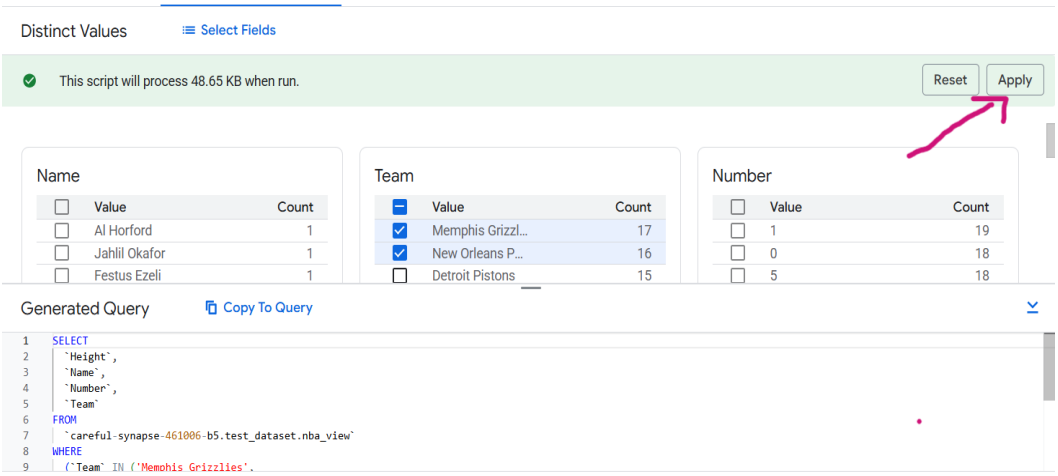
Filter Enter property name or value

Field name	Type	Mode	Key	Collation	Default Value	Policy Tags	Description
Name	STRING	NULLABLE	-	-	-	-	-
Team	STRING	NULLABLE	-	-	-	-	-
Number	INTEGER	NULLABLE	-	-	-	-	-
Position	STRING	NULLABLE	-	-	-	-	-
Age	INTEGER	NULLABLE	-	-	-	-	-
Height	STRING	NULLABLE	-	-	-	-	-
Weight	INTEGER	NULLABLE	-	-	-	-	-
College	STRING	NULLABLE	-	-	-	-	-
Salary	INTEGER	NULLABLE	-	-	-	-	-

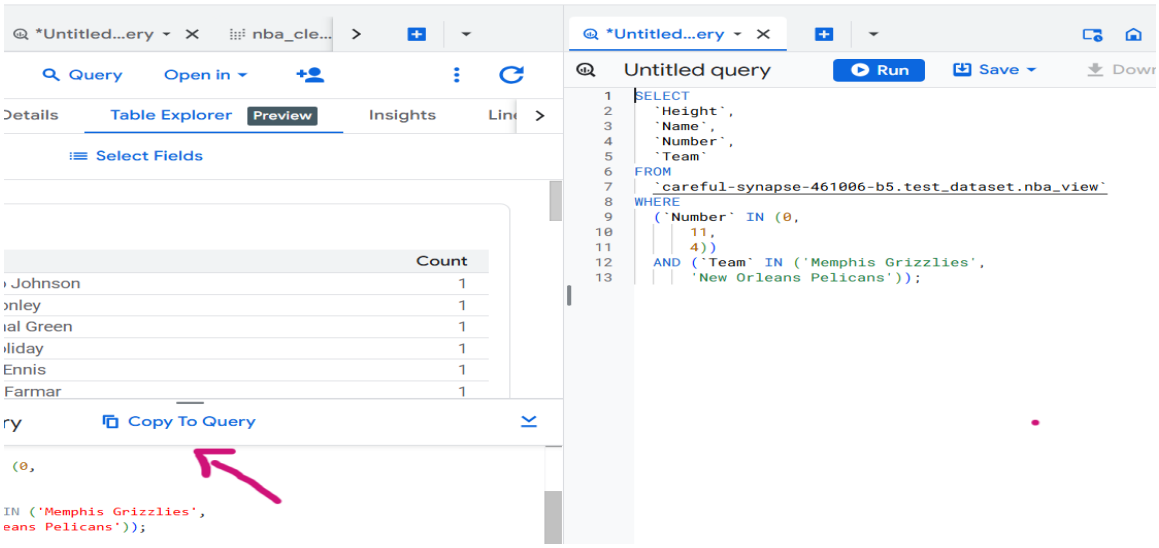
STEP 6: SELECT THE NECESSARY COLUMN TO EXPLORE:



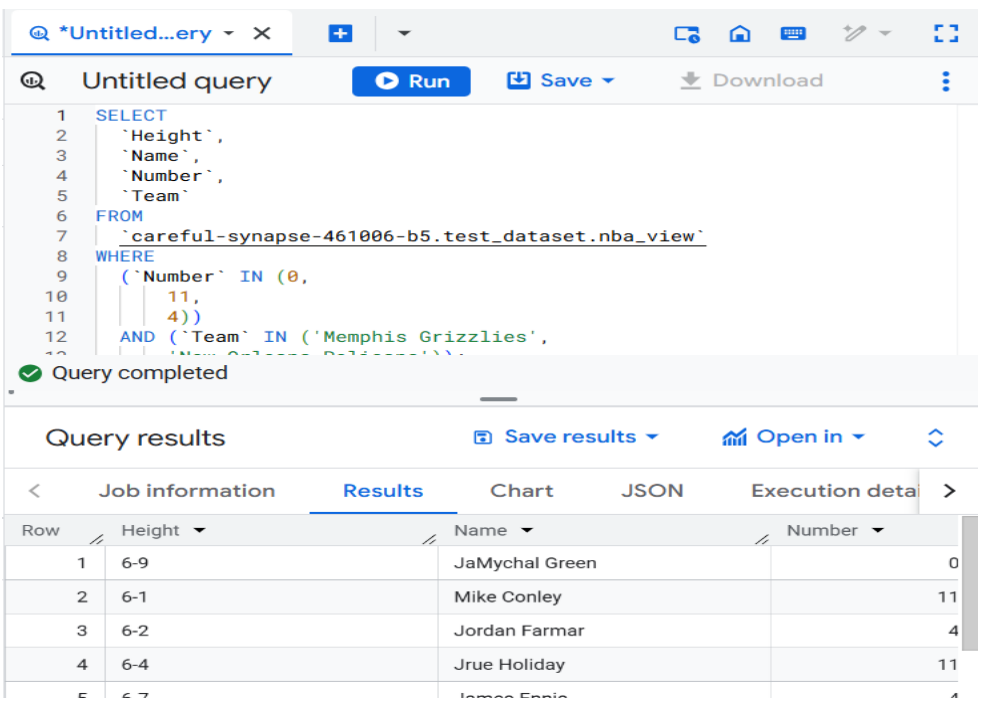
STEP7: Choose related column and click apply



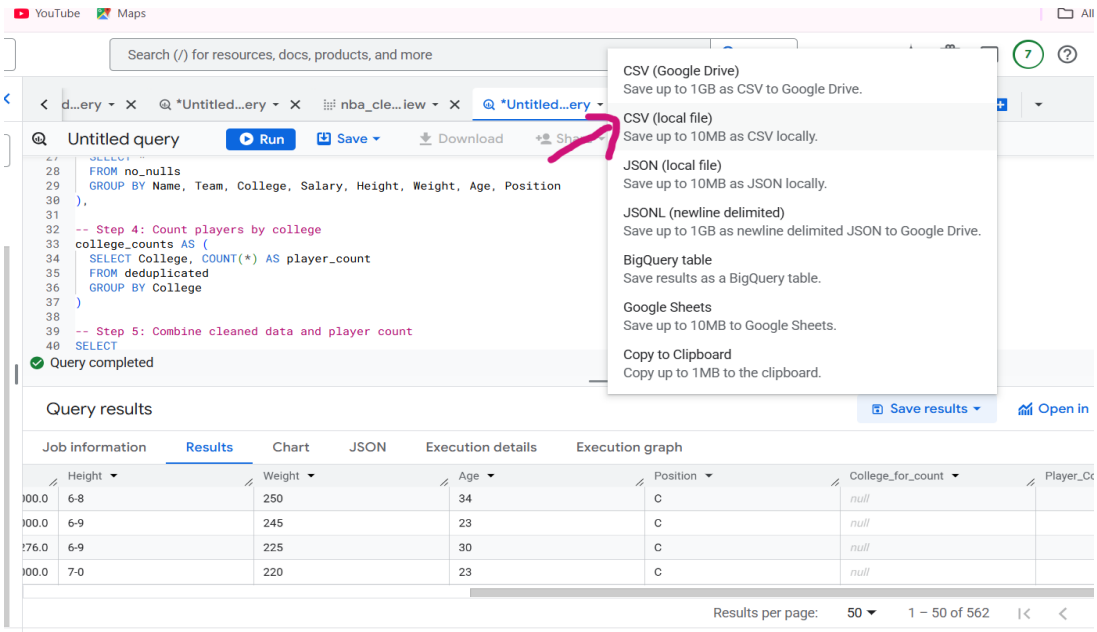
STEP 8: Click Copy the query to paste the query in execute page.



STEP 9: Run the Query



STEP 10: Download the Cleaned data into csv format for future analysis.



VISUALIZE THE CLEANED DATA IN LOOKER STUDIO:

I. POSITION BY SALARY (DONUT CHART)

QUESTION:
Which player position receives the highest overall salary percentage?

ANSWER:
The **Point Guard (PG)** position receives the highest overall salary percentage at **22.2%**.

II. SALARY BY NAME (BAR CHART)

QUESTION:
Who is the highest-paid NBA player among the listed names?

ANSWER:
Carmelo Anthony has the highest salary among the listed NBA players, earning over $20 million.

III. STAFF REPORT (TABLE)

QUESTION:
Name any two players listed in the Staff Report along with their college.

ANSWER:

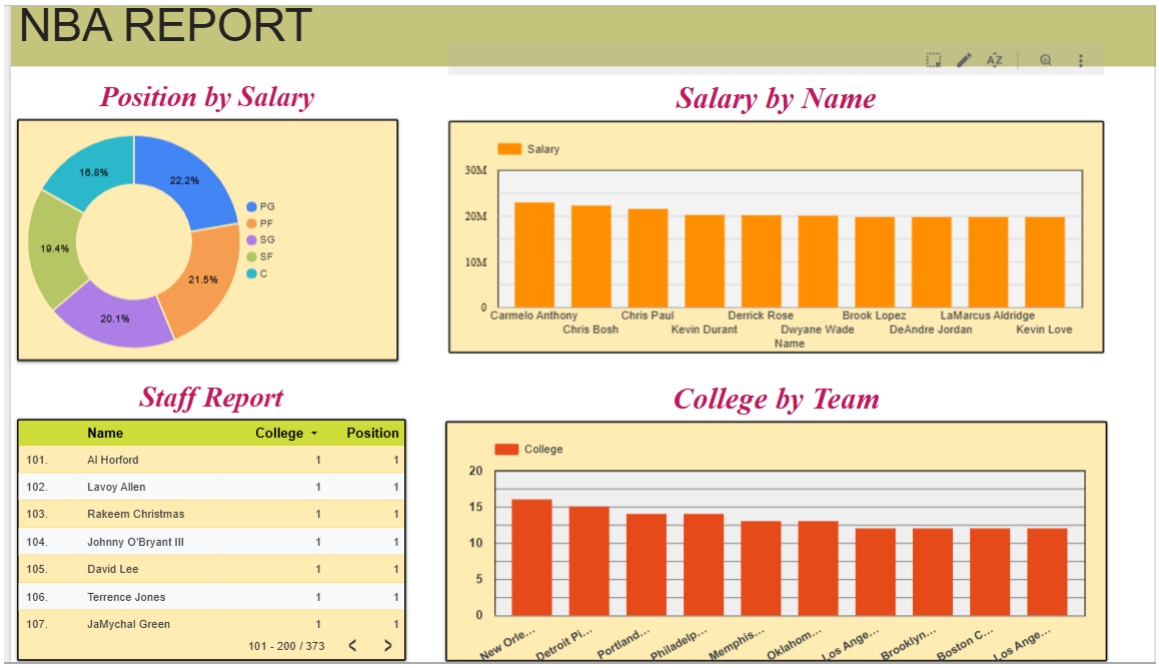
- Al Horford – College

- Lavoy Allen – College

IV. COLLEGE BY TEAM (BAR CHART)

QUESTION:
Which NBA team has the highest number of players from colleges?

ANSWER:
New Orleans has the highest number of college players among the listed teams, with about **16–17 players**.



CONCLUSION:

In this project, we carried out a structured data analysis process on an NBA dataset to uncover meaningful insights. The entire workflow included data cleaning, transformation, and visualization, which helped simplify raw information into clear, interpretable patterns.