</talentlabs>

# Chapter 5 Assignment - SQL Insight Generation

## Instructions

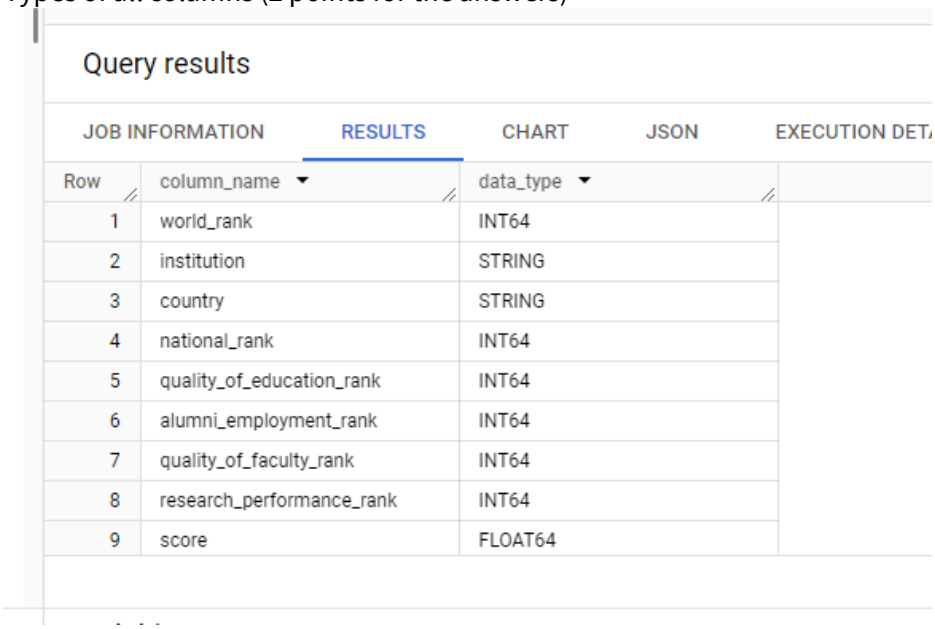
1. You can take help from the lecture notes to revise the concepts that we have covered
2. Choose the best suitable answer and submit the word document
3. You have been provided a csv file named “Top 2000 Universities of the World.csv”, this is your dataset for this assignment.
4. For these questions, you need to work on Google BigQuery and answer the questions in this document.
5. To get started with the assignment, you need to create a database and dataset in Google BigQuery using the csv file provided to you as a dataset. (You can take help from the Hands On exercise video from lectures)
6. For each question, apart from answering the questions, please also paste a screenshot of the SQL with the SQL output, as a proof of your work.
7. Please submit the assignment through TalentLabs Learning System. You will need to submit this word document.

</talentlabs>

**Question 1 (3 point):**
If you recall the SQL hands-on analysis video, you should remember that we need to create a table schema when we set up the data table before working on BigQuery.

What are the data types in your schema?

Types of all columns (2 points for the answers)



Query results

JOB INFORMATION | RESULTS | CHART | JSON | EXECUTION DET

| Row | column_name | data_type |
|---|---|---|
| 1 | world_rank | INT64 |
| 2 | institution | STRING |
| 3 | country | STRING |
| 4 | national_rank | INT64 |
| 5 | quality_of_education_rank | INT64 |
| 6 | alumni_employment_rank | INT64 |
| 7 | quality_of_faculty_rank | INT64 |
| 8 | research_performance_rank | INT64 |
| 9 | score | FLOAT64 |

SQL Query (1 point for the SQL query):

```
SELECT column_name, data_type
FROM `module3-chapter-5-assignment.topuni.INFORMATION_SCHEMA.COLUMNS`
WHERE table_name = 'ori';
```

</talentlabs>

**Question 2 (6 points):**

Data aggregation can help us in understanding how different groups of data compare to each other. In this exercise, we would like to understand which country is having the best ranking in terms of quality education.

To achieve this, we can calculate the average “quality education” ranking of each country, and see which country is having the best or worst average ranking.

(Hint: you can do a group by, and you should ignore countries with no data on quality education ranking)

Top 3 countries in terms of quality education (2 points)

Query results

JOB INFORMATION | RESULTS | CHART | JSON | EXECUTI

| Row | country | avg_quality_ranking |
|---|---|---|
| 1 | Switzerland | 113.1666666666... |
| 2 | Israel | 127.25 |
| 3 | Bulgaria | 154.0 |

Last 3 countries in terms of quality education (2 points)

JOB INFORMATION | RESULTS | CHART | JSON | E

| Row | country | avg_quality_ranking |
|---|---|---|
| 1 | Ghana | 508.0 |
| 2 | Chile | 500.0 |
| 3 | Greece | 474.0 |

SQL Query (2 points)

```
--- Top 3
SELECT country, AVG(quality_of_education_rank) AS avg_quality_ranking
FROM `module3-chapter-5-assignment.topuni.ori`
WHERE quality_of_education_rank != 0
GROUP BY country
ORDER BY avg_quality_ranking ASC
LIMIT 3;
```

```
--- Bottom 3
SELECT country, AVG(quality_of_education_rank) AS avg_quality_ranking
FROM `module3-chapter-5-assignment.topuni.ori`
WHERE quality_of_education_rank !=0
GROUP BY country
ORDER BY avg_quality_ranking DESC
LIMIT 3;
```

**Question 3 (7 points):**
In this dataset, there is a column named “National Ranking”, which shows the ranking of the universities within their own country. This can help us in identifying the best university in each of the countries.

Let’s try to find out the top universities of the countries listed below:

| **Country** | **Best University (2.5 points)** |
|---|---|
| India | Indian Institute of Management Ahmedabad |
| Denmark | University of Copenhagen |
| Malaysia | University of Malaya |
| Indonesia | University of Indonesia |
| Vietnam | Vietnam National University, Hanoi |

SQL Query (2 points):

```
SELECT country, institution, MIN(national_rank) AS best_ranking
FROM `module3-chapter-5-assignment.topuni.ori`
WHERE country IN ('India', 'Denmark', 'Malaysia', 'Indonesia', 'Vietnam')
GROUP BY country, institution
ORDER BY country, best_ranking;
```

</talentlabs>

Query results

JOB INFORMATION | RESULTS | CHART | JSON | EXECUTION DETAILS | EXECUTION GRAPH

| Row | country | institution | best_ranking |
|---|---|---|---|
| 1 | Denmark | University of Copenhagen | 1 |
| 2 | Denmark | Aarhus University | 2 |
| 3 | Denmark | Technical University of Denmark | 3 |
| 4 | Denmark | Aalborg University | 4 |
| 5 | Denmark | University of Southern Denmark | 5 |
| 6 | Denmark | Copenhagen Business School | 6 |
| 7 | Denmark | Roskilde University | 7 |
| 8 | India | Indian Institute of Managemen... | 1 |
| 9 | India | Indian Institute of Science | 2 |

**Question 4 (5 points):**
Data Summaries like mean, mode and media are great ways of summarising large datasets and generating insights.

In this question, we would like to do some analysis for universities in the UK. In order to do that,

1. you need to make a sub table for United Kingdom Institutions
2. summarise the column of interest using measures of location which should include Mean, Mode and Median.

Let's try to answer the following:

In terms of UK universities research performance ranking:

Mean (1 point)
775.884210526

Median (1 point)
707.0

Although the UK got really good university (e.g. Oxford University and Cambridge University), why is the mean ranking in research performance still bad? (1 point)

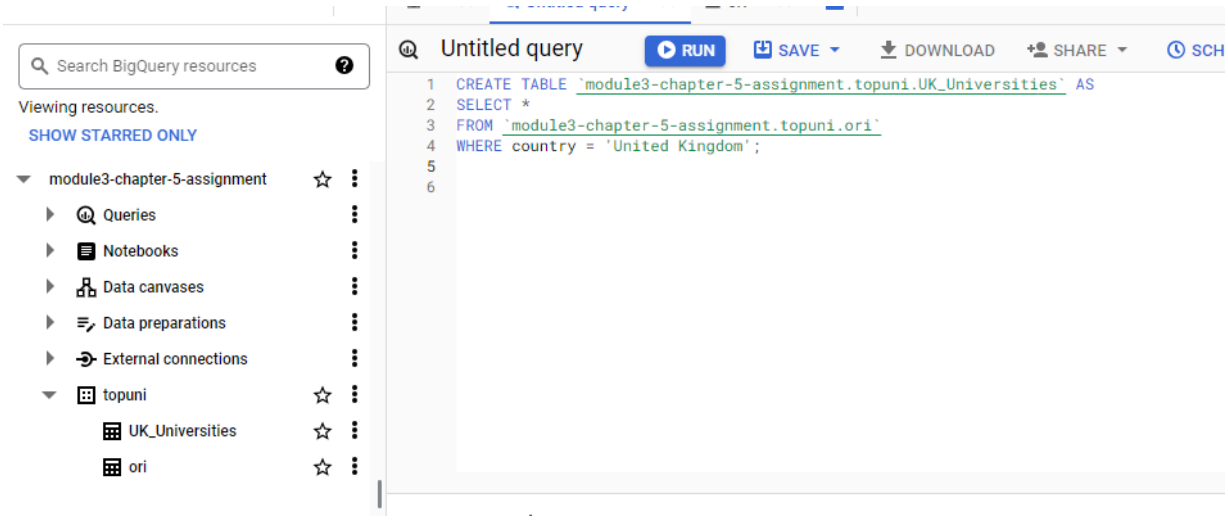
- **Explanation:** Even though the UK has top universities like Oxford and Cambridge, the mean ranking might not be very high because the dataset likely includes many universities with lower research performance rankings, which brings down the overall average.

</talentlabs>

SQL Query: (2 points)

1.

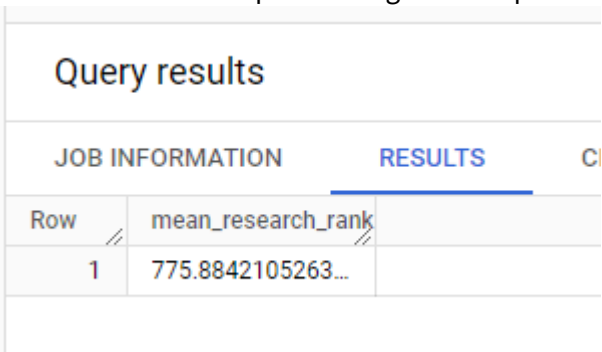
```
CREATE TABLE `module3-chapter-5-assignment.topuni.UK_Universities` AS
SELECT *
FROM `module3-chapter-5-assignment.topuni.ori`
WHERE country = 'United Kingdom';
```




2.

Mean

```
SELECT AVG( research_performance_rank) AS mean_research_ranking
FROM `module3-chapter-5-assignment.topuni.UK_Universities`;
```




Median

```
SELECT PERCENTILE_CONT(research_performance_rank, 0.5) OVER() AS median_research_ranking
FROM `module3-chapter-5-assignment.topuni.UK_Universities`;
```

</talentlabs>


Mode (no mode)

```
SELECT research_performance_rank, COUNT(*) AS frequency
FROM `module3-chapter-5-assignment.topuni.UK_Universities`
WHERE research_performance_rank !=0
GROUP BY research_performance_rank
ORDER BY frequency DESC
LIMIT 1;
```