

Exploring NCAA Data with BigQuery

GSP160



Google Cloud Self-Paced Labs

Overview

BigQuery is Google's fully managed, NoOps, low cost analytics database. With BigQuery you can query terabytes and terabytes of data without managing infrastructure or needing a database administrator. BigQuery uses SQL and takes advantage of the pay-as-you-go model. BigQuery allows you to focus on analyzing data to find meaningful insights.

We have a newly available dataset for NCAA Basketball games, teams, and players. The game data covers play-by-play and box scores back to 2009, as well as final scores back to 1996. Additional data about wins and losses goes back to the 1894-5 season in some teams' cases.

In this lab we will find and query the NCAA dataset using BigQuery.

What you'll learn

- Using BigQuery
- Query the NCAA Public Dataset
- Writing and executing queries

What you'll need

- A Google Cloud Project
- A Browser, such [Chrome](#) or [Firefox](#)

Setup and Requirements

Before you click the Start Lab button

Read these instructions. Labs are timed and you cannot pause them. The timer, which starts when you click **Start Lab**, shows how long Google Cloud resources will be made available to you.

This Qwiklabs hands-on lab lets you do the lab activities yourself in a real cloud environment, not in a simulation or demo environment. It does so by giving you new, temporary credentials that you use to sign in and access Google Cloud for the duration of the lab.

What you need

To complete this lab, you need:

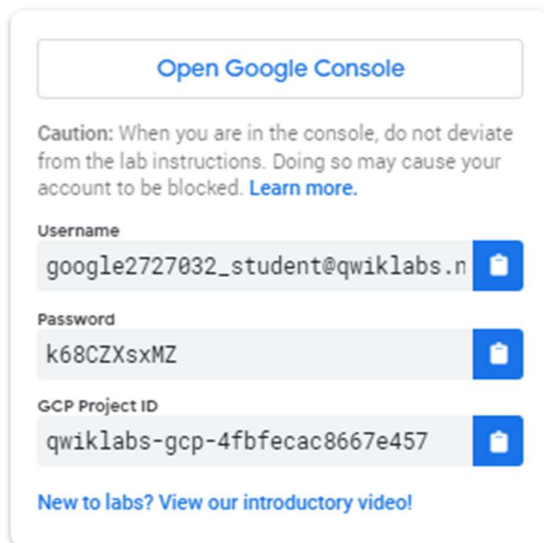
- Access to a standard internet browser (Chrome browser recommended).
- Time to complete the lab.

Note: If you already have your own personal Google Cloud account or project, do not use it for this lab.

Note: If you are using a Pixelbook, open an Incognito window to run this lab.

How to start your lab and sign in to the Google Cloud Console

1. Click the **Start Lab** button. If you need to pay for the lab, a pop-up opens for you to select your payment method. On the left is a panel populated with the temporary credentials that you must use for this lab.



A screenshot of a web panel titled "Open Google Console". It contains a caution message, a "Username" field with the value "google2727032_student@qwiklabs.n", a "Password" field with the value "k68CZxsxMZ", and a "GCP Project ID" field with the value "qwiklabs-gcp-4fbfecac8667e457". Each field has a copy icon to its right. At the bottom, there is a link "New to labs? View our introductory video!".

Open Google Console

Caution: When you are in the console, do not deviate from the lab instructions. Doing so may cause your account to be blocked. [Learn more.](#)

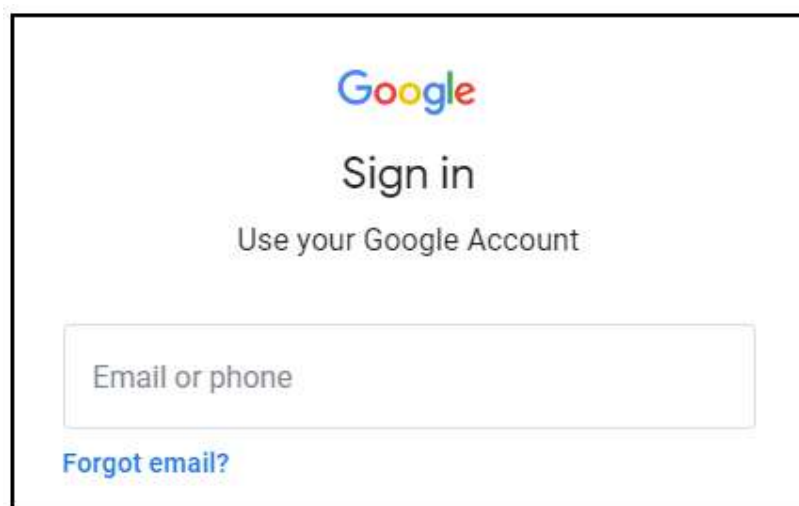
Username
google2727032_student@qwiklabs.n

Password
k68CZxsxMZ

GCP Project ID
qwiklabs-gcp-4fbfecac8667e457

[New to labs? View our introductory video!](#)

2. Copy the username, and then click **Open Google Console**. The lab spins up resources, and then opens another tab that shows the **Sign in** page.



A screenshot of the Google Sign in page. It features the Google logo at the top, followed by the text "Sign in" and "Use your Google Account". Below this is a text input field labeled "Email or phone". At the bottom left, there is a link "Forgot email?".

Google

Sign in

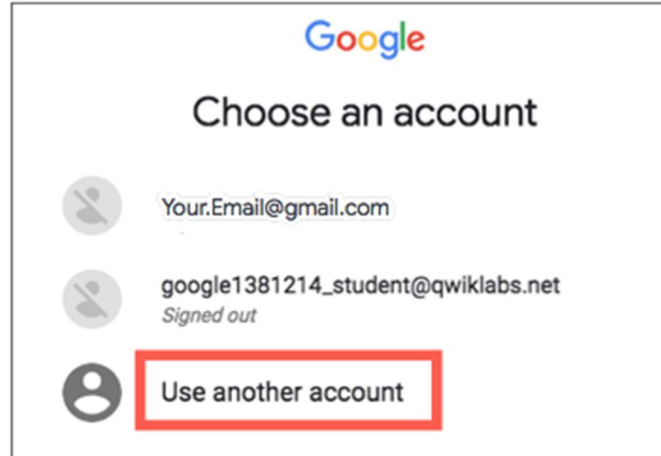
Use your Google Account

Email or phone

[Forgot email?](#)

Tip: Open the tabs in separate windows, side-by-side.

If you see the **Choose an account** page, click **Use Another**



Account.

3. In the **Sign in** page, paste the username that you copied from the Connection Details panel. Then copy and paste the password.

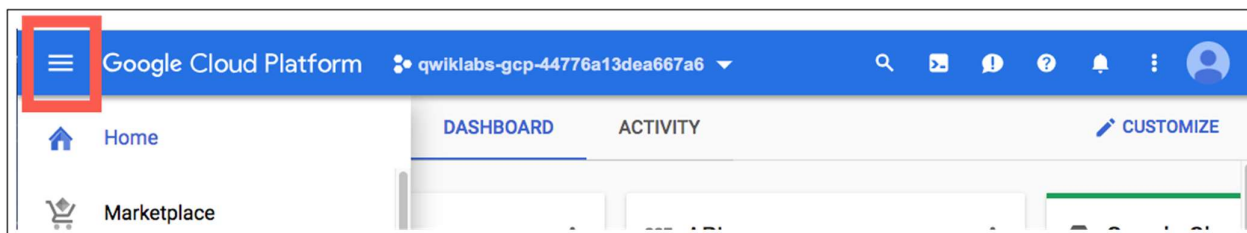
Important: You must use the credentials from the Connection Details panel. Do not use your Qwiklabs credentials. If you have your own Google Cloud account, do not use it for this lab (avoids incurring charges).

4. Click through the subsequent pages:

- Accept the terms and conditions.
- Do not add recovery options or two-factor authentication (because this is a temporary account).
- Do not sign up for free trials.

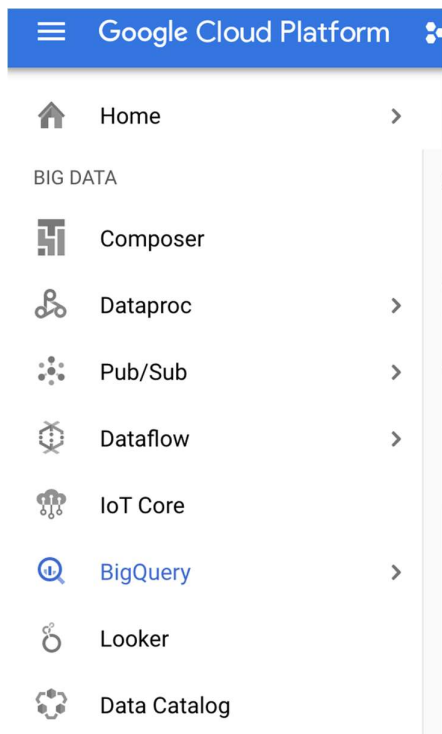
After a few moments, the Cloud Console opens in this tab.

Note: You can view the menu with a list of Google Cloud Products and Services by clicking the **Navigation menu** at the top-left.



Open BigQuery Console

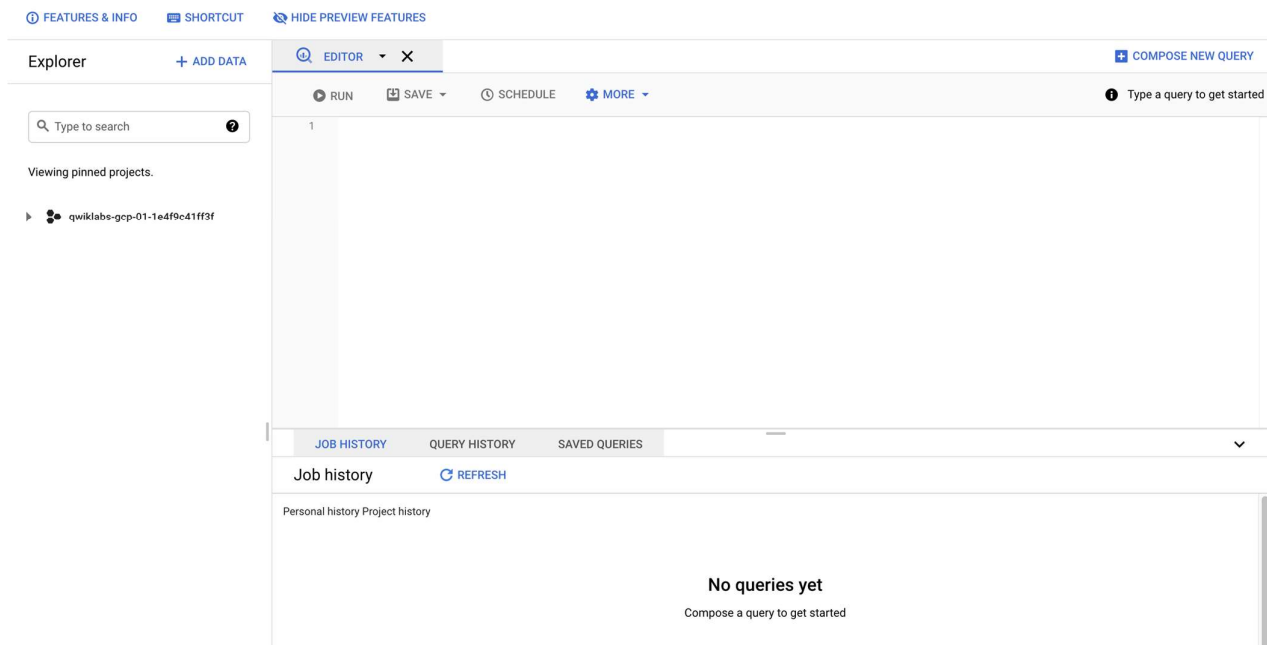
In the Google Cloud Console, select **Navigation menu** > **BigQuery**:



The **Welcome to BigQuery in the Cloud Console** message box opens. This message box provides a link to the quickstart guide and the release notes.

Click **Done**.

The BigQuery console opens.



BigQuery opens, but there's nothing in here! Luckily, there are tons of Open Datasets available in BigQuery for you to query, and of course you can upload your own data, which you'll do in the next section.

Find the NCAA public dataset in BigQuery

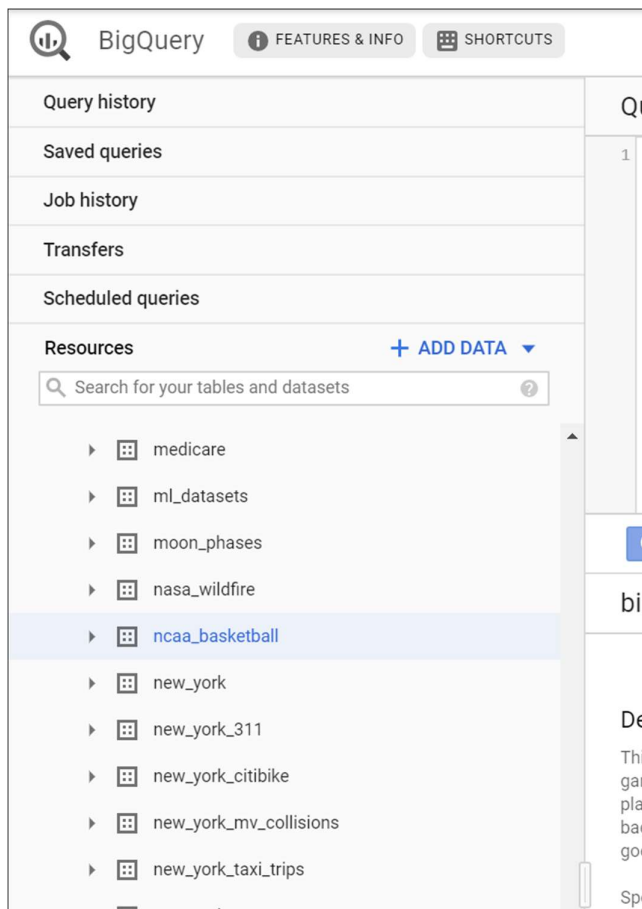
In this section, you pull in some public data so you can practice running SQL commands in BigQuery. Click on the **+ ADD DATA** link then select **Explore public datasets**:



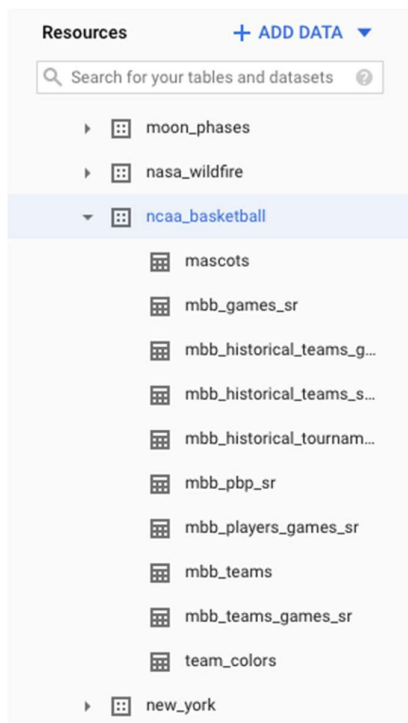
Type "ncaa basketball" in the searchbar and press **Enter**.

Click on the NCAA Basketball tile, then **View Dataset**.

A new browser tab opens, you now have a new project called `bigquery-public-data` added to the Resources panel, opened to `ncaa_basketball`:



Click on the dataset name to view the tables you can explore.



Click on `mbb_teams_games_sr` (men's NCAA game results table) and then click **Preview** tab to see sample rows of data. Click the **Details** tab to get metadata about the table.

▼

■

naaa_basketball

■

mscots

■

mbb_games_sr

■

mbb_historical_teams_g...

■

mbb_historical_teams_s...

■

mbb_historical_tournam...

■

mbb_pbp_sr

■

mbb_players_games_sr

■

mbb_teams

■

mbb_teams_games_sr

■

team_colors

▶

■

new_york

▶

■

new_york_311

▶

■

new_york_citibike

▶

■

new_york_mv_collisions

▶

■

new_york_taxi_trips

▶

■

new_york_trees

mbb_teams_games_sr

QUERY TABLE

COPY TABLE

DELETE TABLE

EXPORT ▼

Schema

Details

Preview

Description

Labels

Team-level box scores from every men's basketball game from the 2013-14 season to the 2017-18 season. Each row shows a single team's stats in one game. This data is identical to mbb_games_sr, but is organized differently to make it easier to calculate a single team's statistics.


None

Data provided by Sportradar LLC, see www.sportradar.com for more details.

Table info

Table ID	bigquery-public-data:naaa_basketball.mbb_teams_games_sr
Table size	95.36 MB
Number of rows	59,610
Created	Mar 17, 2018, 8:25:27 PM
Table expiration	Never
Last modified	Mar 15, 2019, 8:55:05 AM
Data location	US

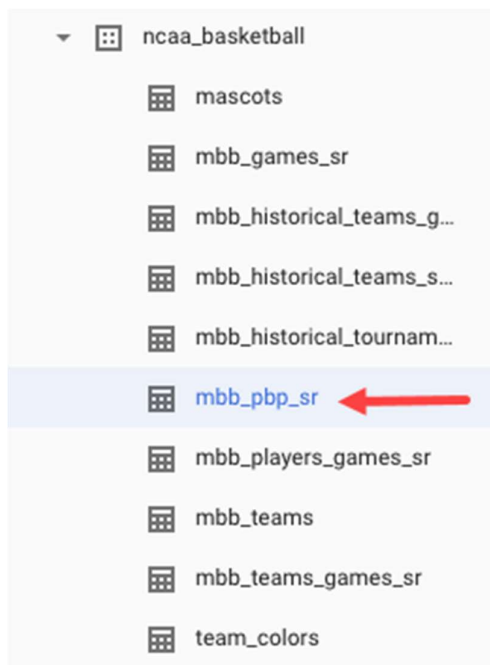
How many games does the dataset contain? How big is the table?

Table info 	
Table ID	bigquery-public-data:ncaa_basketball.mbb_games_sr
Table size	48.08 MB
Number of rows	29,805
Created	Mar 17, 2018, 8:24:34 PM
Table expiration	Never
Last modified	Feb 28, 2019, 7:35:04 AM
Data location	US

Answer: The table is about 48 MB and there are over 29k games for us to explore.

But how many individual plays can we analyze?

Hint: click on the `mbb` `pbp` `sr` (play-by-play) dataset:



Then click **Details**:

mbb_pbp_sr

Schema **Details** Preview

Description

Play-by-play information from men's basketball games, starting with the 2013-14 season. Each row shows a single event in a game.

Data provided by Sportradar LLC, see www.sportradar.com for more details.

Table info

Table ID	bigquery-public-data:ncaa_basketball.mbb_pbp_sr
Table size	3.14 GB
Long-term storage size	3.14 GB
Number of rows	4,160,393
Created	Mar 19, 2018, 10:10:51 AM
Table expiration	Never
Last modified	May 31, 2018, 10:47:26 AM
Data location	US

Answer: Over 4 million individual basketball plays.

Let's write some SQL to see what types of plays are there for us to explore.

Writing queries

What types of basketball play events are there?

Click **Compose Query**.

Copy and paste the below query into the editor:

```
#standardSQL
SELECT
  event_type,
  COUNT(*) AS event_count
FROM `bigquery-public-data.ncaa_basketball.mbb_pbp_sr`
GROUP BY 1
ORDER BY event_count DESC;
```

Now click **Run**.

Looking at your results, how many historical shots were TWOPOINTMADE or FREETHROWMISS?

Row	event_type	event_count
1	rebound	790490
2	twopointmiss	425422
3	turnover	410166
4	twopointmade	400648
5	freethrowmade	316637
6	assist	293403
7	threepointmiss	291935
8	personal foul	243613
9	shooting foul	183541
10	threepointmade	153967
11	freethrowmiss	134256
12	attempt blocked	81863
13	block	81863
14	tv timeout	79960
15	team timeout	63613
16	deadball	55249
17	offensive foul	48713
18	lineup change	35926
19	end period	23128

Click **Check my progress** to verify the objective.

Writing queries

Check my progress

Fun queries to run

**Which 5 games featured the most three point shots made?
How accurate were all the attempts?**

Click **+ Compose New Query** and add in the below query:

```
#standardSQL
#most three points made
SELECT
  scheduled_date,
  name,
  market,
  alias,
  three_points_att,
  three_points_made,
  three_points_pct,
  opp_name,
  opp_market,
  opp_alias,
  opp_three_points_att,
  opp_three_points_made,
  opp_three_points_pct,
  (three_points_made + opp_three_points_made) AS total_threes
FROM `bigquery-public-data.ncaa_basketball.mbb_teams_games_sr`
WHERE season > 2010
ORDER BY total_threes DESC
LIMIT 5;
```

Click **Run**.

scheduled_date	name	market	alias	three_points_att	three_points_made	three_points_pct	opp_name
2016-11-18	Knights	Middle Georgia	MGC	38	16	42.1	Tigers
2016-11-18	Tigers	Savannah State	SAV	52	24	46.2	Knights
2016-11-22	Tigers	Savannah State	SAV	45	23	51.1	Wildcats
2016-11-22	Wildcats	Fort Valley State	FVSU	39	14	35.9	Tigers
2017-12-10	Skyhawks	Tennessee-Martin	UTM	35	15	42.9	Cougars

Wow! The Tigers made over 50% of their three point shots on 11-22-2016.

Click **Check my progress** to verify the objective.

Query 1

Check my progress

Which 5 basketball venues have the highest seating capacity?

Click + **Compose New Query** and add the below query:

```
#standardSQL
SELECT
  venue_name, venue_capacity, venue_city, venue_state
FROM `bigquery-public-data.ncaa_basketball.mbb_teams_games_sr`
GROUP BY 1,2,3,4
ORDER BY venue_capacity DESC
LIMIT 5;
```

Click **Run**.

Row	venue_name	venue_capacity	venue_city	venue_state
1	AT&T Stadium	80000	Arlington	TX
2	University of Phoenix Stadium	72220	Glendale	AZ
3	NRG Stadium	71054	Houston	TX
4	Georgia Dome	71000	Atlanta	GA
5	Lucas Oil Stadium	70000	Indianapolis	IN

Imagine taking a shot with 80,000 people watching you!

Click **Check my progress** to verify the objective.

Query 2

Check my progress

Which teams played in the highest scoring game since 2010?

Click **Compose Query** and add the below query:

```
#standardSQL
#highest scoring game of all time
SELECT
  scheduled_date,
  name,
  market,
  alias,
  points_game AS team_points,
  opp_name,
  opp_market,
  opp_alias,
  opp_points_game AS opposing_team_points,
  points_game + opp_points_game AS point_total
FROM `bigquery-public-data.ncaa_basketball.mbb_teams_games_sr`
WHERE season > 2010
ORDER BY point_total DESC
LIMIT 5;
```

Click **Run**.

Row	scheduled_date	name	market	alias	team_points	opp_name	opp_market	opp_alias	opposing_team_points	point_total
1	2017-02-10	Bulldogs	Samford	SAM	127	Terriers	Wofford	WOF	131	258
2	2017-02-10	Terriers	Wofford	WOF	131	Bulldogs	Samford	SAM	127	258
3	2017-02-04	Vikings	Portland State	PRST	124	Eagles	Eastern Washington	EWU	130	254
4	2017-02-04	Eagles	Eastern Washington	EWU	130	Vikings	Portland State	PRST	124	254
5	2013-11-14	Pioneers	Sacred Heart	SHU	118	Crusaders	Holy Cross	HC	122	240

The Bulldogs and Terriers played in a game that scored 258 total points!

Click **Check my progress** to verify the objective.

Query 3

Check my progress

Since 2015, what was the biggest difference in final score for a National Championship?

Click **+ Compose New Query** and add the below query:

```
#standardSQL
#biggest point difference in a championship game
SELECT
  scheduled_date,
  name,
  market,
  alias,
  points_game AS team_points,
  opp_name,
  opp_market,
  opp_alias,
  opp_points_game AS opposing_team_points,
  ABS(points_game - opp_points_game) AS point_difference
FROM `bigquery-public-data.ncaa_basketball.mbb_teams_games_sr`
WHERE season > 2015 AND tournament_type = 'National Championship'
ORDER BY point_difference DESC
```

```
LIMIT 5;
```

Click **Run**.

Row	scheduled_date	name	market	alias	team_points	opp_name	opp_market	opp_alias	opposing_team_points	point_difference
1	2018-04-03	Wildcats	Villanova	VILL	79	Wolverines	Michigan	MICH	62	17
2	2018-04-03	Wolverines	Michigan	MICH	62	Wildcats	Villanova	VILL	79	17
3	2017-04-04	Tar Heels	North Carolina	UNC	71	Bulldogs	Gonzaga	GONZ	65	6
4	2014-04-08	Wildcats	Kentucky	UK	54	Huskies	Connecticut	CONN	60	6
5	2017-04-04	Bulldogs	Gonzaga	GONZ	65	Tar Heels	North Carolina	UNC	71	6

The finals games are surprisingly close! The biggest difference was recent in 2018 with a delta of 17 points.

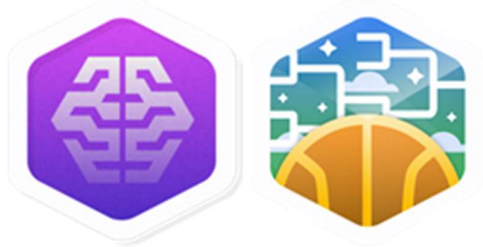
Click **Check my progress** to verify the objective.

Query 4

Check my progress

Congratulations!

You've learned how to query the NCAA basketball dataset inside of BigQuery. We encourage you to modify the above queries and write your own to further your understanding. Looking for more NCAA query practice? Checkout the GitHub repo [here](#).



Finish your Quest

Continue your Quest with [Google Cloud Solutions II: Data and Machine Learning](#) or [NCAA® March Madness®: Bracketology with Google Cloud](#). A Quest is a series of related labs that form a learning path. Completing this Quest earns you the badge above, to recognize your achievement. You can make your badge (or badges) public and link to them in your online resume or social media account. Enroll in a Quest and get immediate completion credit if you've taken this lab. [See other available Qwiklabs Quests](#).

Take your next lab

Continue your quest with, [Creating Custom Interactive Dashboards with Bokeh and BigQuery](#), or check out these suggestions:

- [Creating an Object Detection Application with TensorFlow](#)
- [Bracketology with Google Machine Learning](#)

Next steps/learn more

- Check out the [BigQuery subreddit](#) for how others are using BigQuery today.
- Find [open datasets available in BigQuery](#) and other interesting topics like [Visualizing BigQuery Data Using Google Data Studio](#).

Google Cloud Training & Certification

...helps you make the most of Google Cloud technologies. [Our classes](#) include technical skills and best practices to help you get up to speed quickly and continue your learning journey. We offer fundamental to advanced level training, with on-demand, live, and virtual options to suit your busy schedule. [Certifications](#) help you validate and prove your skill and expertise in Google Cloud technologies.

Manual Last Updated December 11, 2019

Lab Last Tested December 11, 2019

Copyright 2021 Google LLC All rights reserved. Google and the Google logo are trademarks of Google LLC. All other company and product names may be trademarks of the respective companies with which they are associated.