CIS 9440 - Data Warehousing and Analytics

Get Started with Google BigQuery

Getting Started with Google BigQuery:

1. What is Google BigQuery?
2. Create an account
3. Create a new project
4. Load tables to new project (CSV files)

What is Google BigQuery?

Google BigQuery is a flexible, scalable, and cloud-based Enterprise Data Warehouse. With Google BigQuery, you can:

- Store data
- Query data
- Manage data
- Connect data to many tools (such as BI Platforms)

Who uses BigQuery: https://discovery.hgdata.com/product/google-bigquery

How is Google BigQuery different than Google Cloud Platform (or AWS)?

Google Cloud Platform (GCP) or Amazon Web Services (AWS) are suites of computing resources to deploy and operate applications on the web. Thus, these solutions require cloud-based infrastructure to be setup and maintained. If you want a storage server, you set it up.

Conversely, Google BigQuery automatically scales computing and storage infrastructure for the needs of the user.

Step 1: Redeem your Google Coupon

(Step 1) First, use the link below to request a Google Cloud Platform coupon (Google BigQuery is part of the Google Could Platform (GCP) suite). You will be asked to provide your school email address and name. An email will be sent to your school email address to confirm these details before a coupon is sent to you.

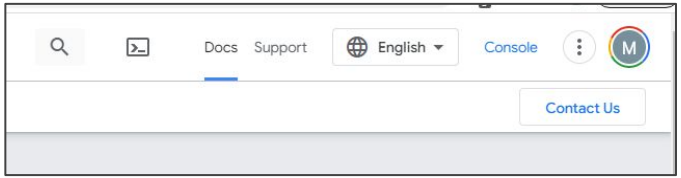
Student Coupon Retrieval Link

- You will be asked for a name and email address, which needs to match the domain (**@baruch.cuny.edu or @baruchmail.cuny.edu**). A confirmation email will be sent to you with a coupon code.
- You can request a coupon from the URL and redeem it until: 12/31/2022
- Coupon valid through: 8/31/2023
- You can only request ONE code per unique Baruch email address.

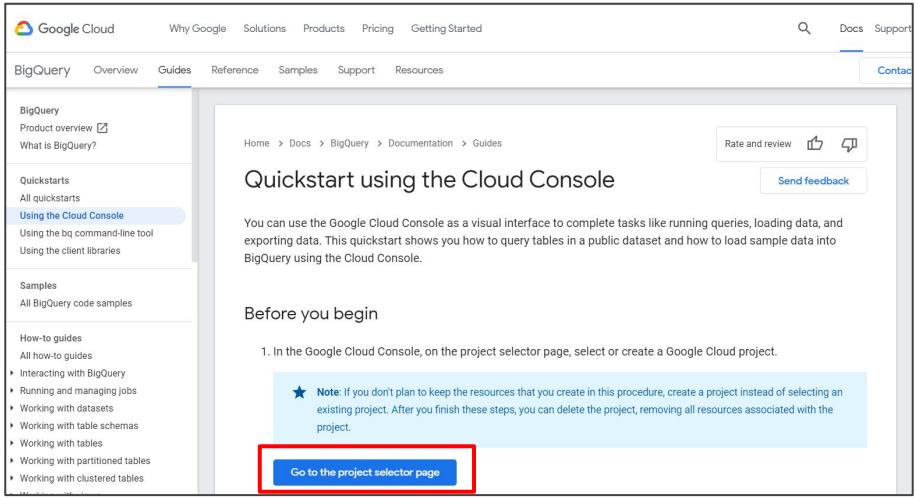
(Step 2) After you receive your coupon, please continue through this deck to get started with Google BigQuery. **NOTE**: when prompted for payment information, use the coupon code from step 1, no need to enter your credit card information.

Create Google Cloud Platform Account

1. Go to: https://cloud.google.com/bigquery/docs/quickstarts/quickstart-web-ui
2. Ensure you're logged into a Google Account, which will be shown on the top-right part of your browser. Example, mine shows an "M"

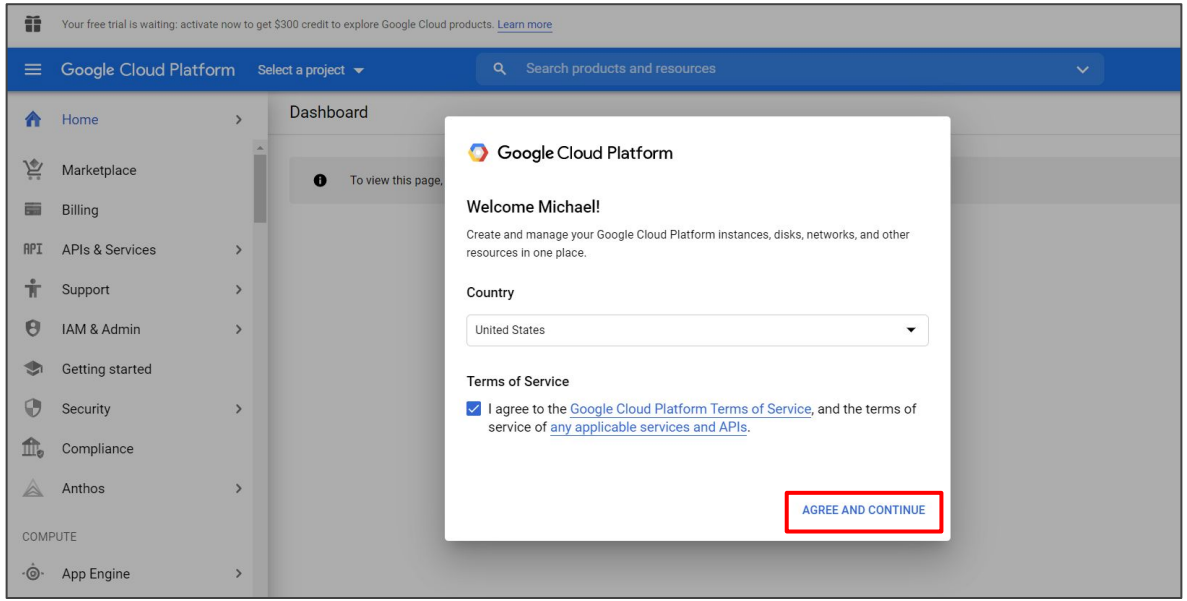


3. Click on "Go to the project selector page"



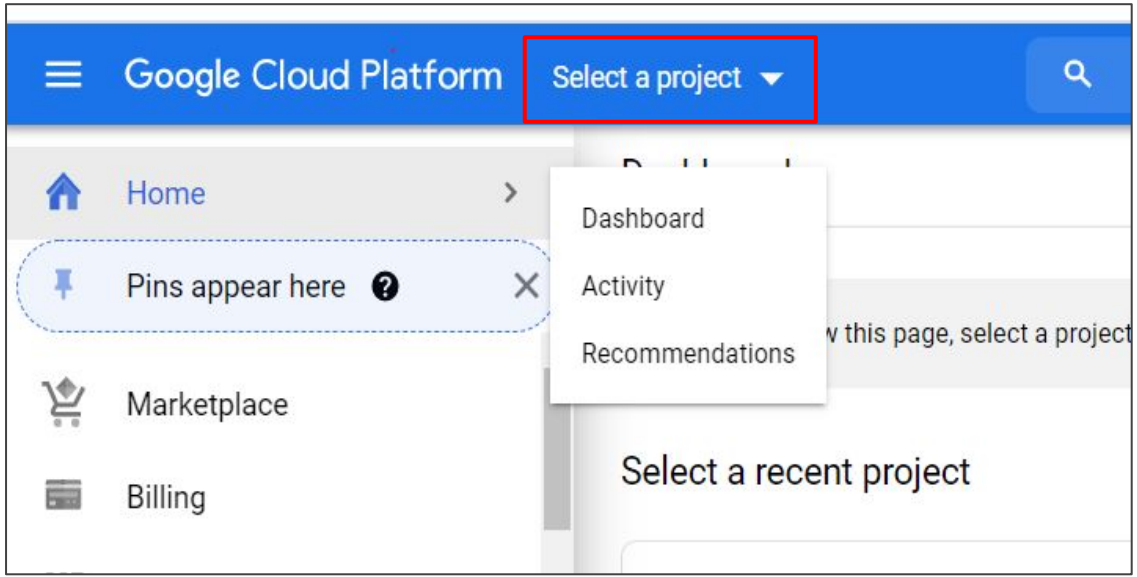
Create Google Cloud Platform Account

1. Check the Terms of Service box
2. Click “Agree and Continue”



Create a new project

1. Click Select a project

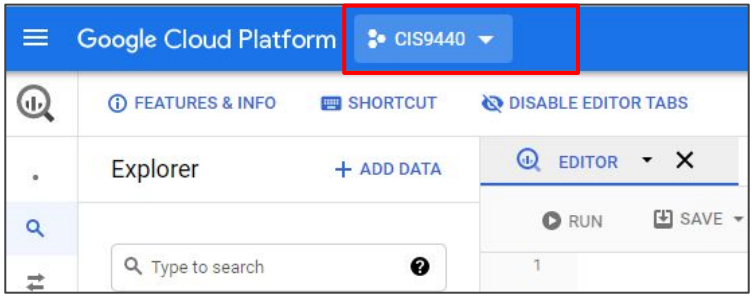


Create a new project

1. Click on “New Project”
2. Project name: CIS9440
3. Location: No organization

Select new project

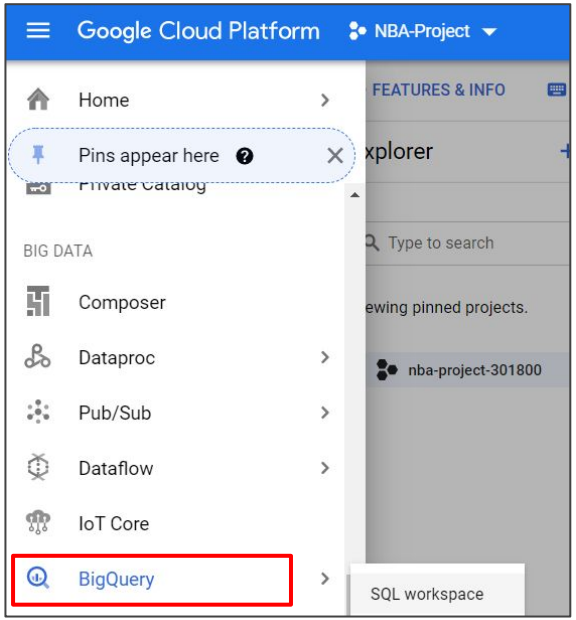
1. Click on the drop-down on the top-left to select a project



2. Choose the project you just created

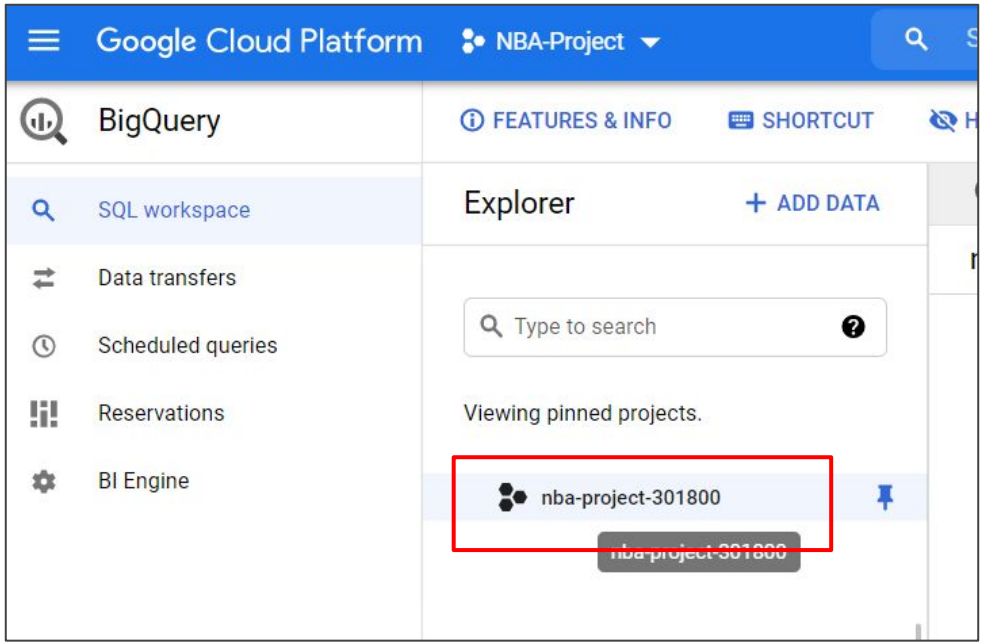
Navigate to BigQuery

1. On the left-hand navigation pane, scroll down and select “BigQuery” in the “Big Data” section



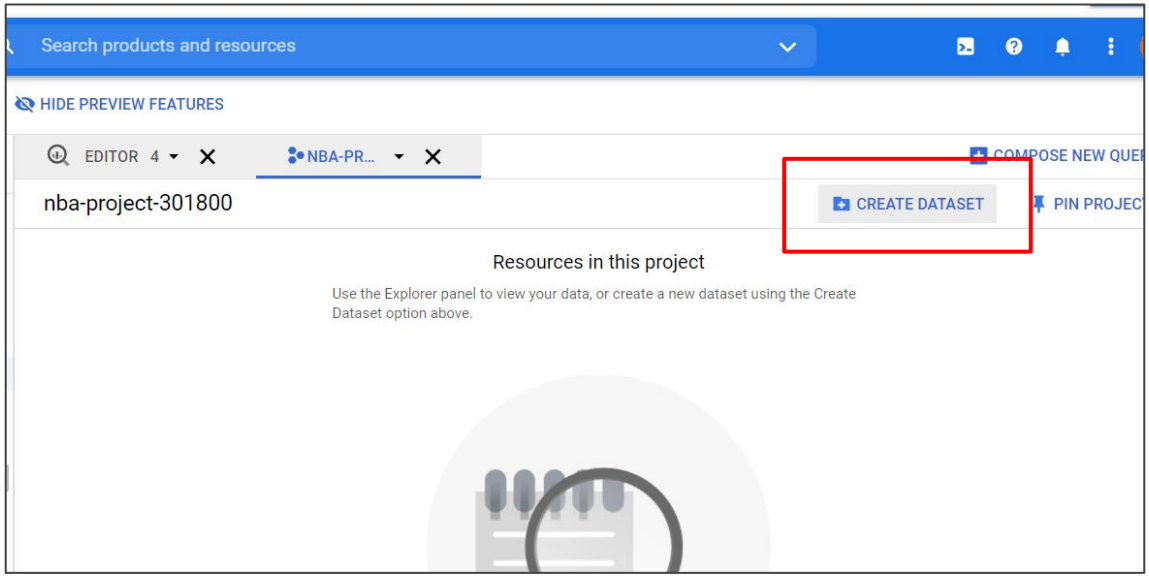
Select your project

1. Click on your project in the “Explorer” window



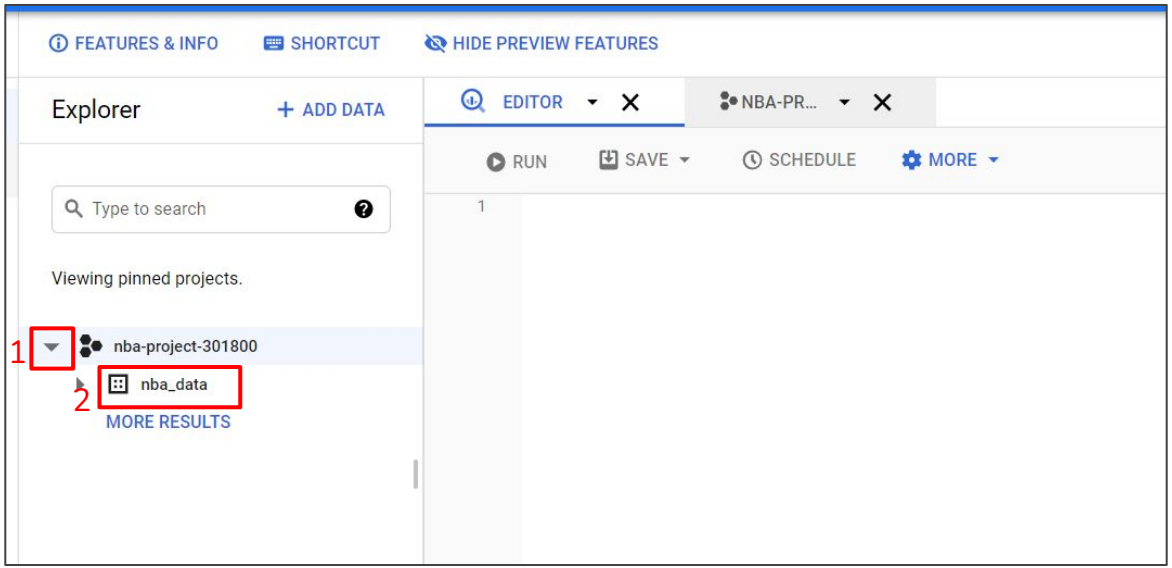
Create a dataset

1. Click on “Create Dataset”
2. Change the Dataset ID to “nba_data”
3. Click on “Create Dataset”



Create a new table

1. Expand your project
2. Click on your new dataset “nba_data”

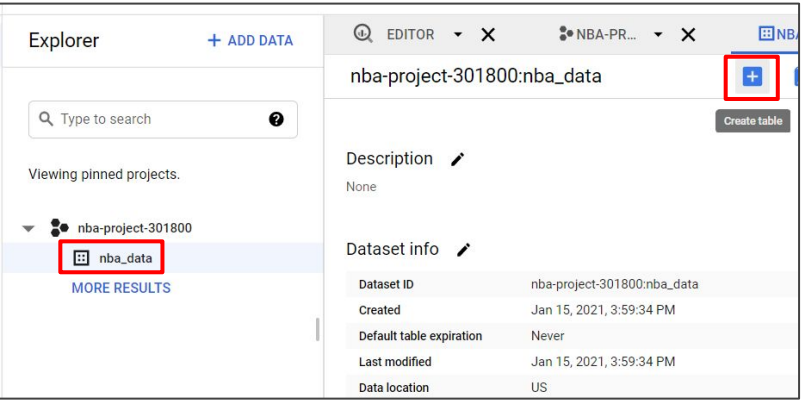


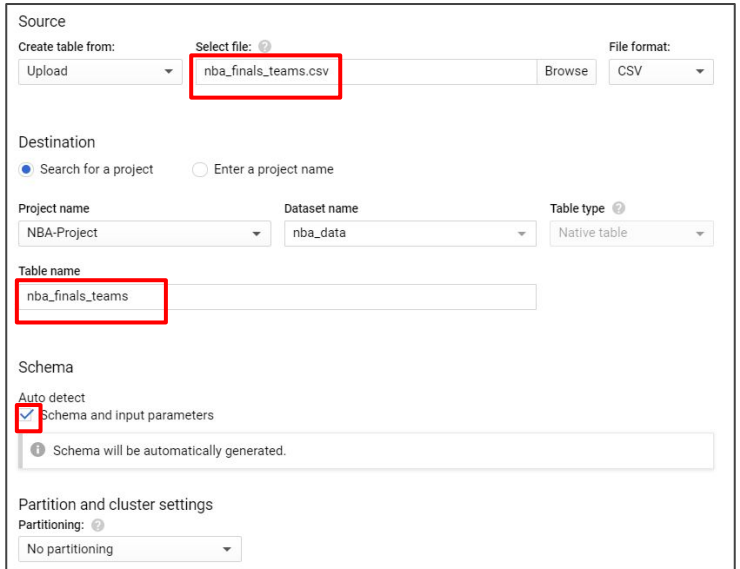
Download CSV files

1. Download the 3 CSV files onto your computer:
    a. https://drive.google.com/file/d/1uB59A7pwC0S0E8YQ5p9AAs2sAltv8D6s/view?usp=sharing
    b. https://drive.google.com/file/d/1EVA6J9g8WQGWESu3XJytVMVns1NJaTQR/view?usp=sharing
    c. https://drive.google.com/file/d/1MjcJDMAphBS2HuKrQN3dioWAejuhtGe2/view?usp=sharing

Create a new table

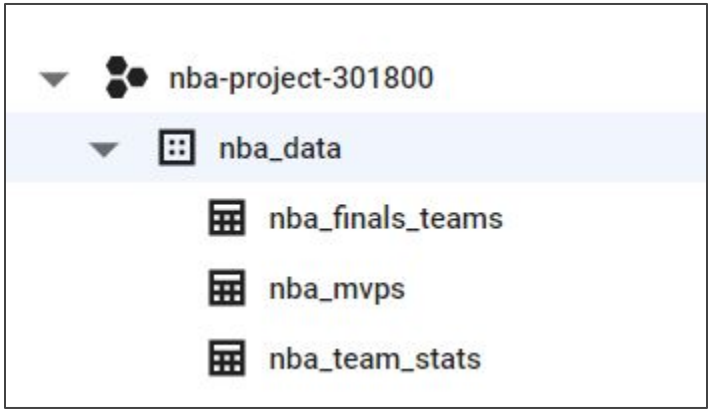
1. Back to BigQuery window
2. Click on your dataset, “nba data”
3. Click on the “+” icon to create a table
4. Create table from “upload”
5. Browse to 1 of the CSV files downloaded on the previous slide
6. Table name: same as CSV name without “.csv”
7. Check “Auto Detect Schema and input parameters”
8. Click on “Create Table”





Create 2 more new tables

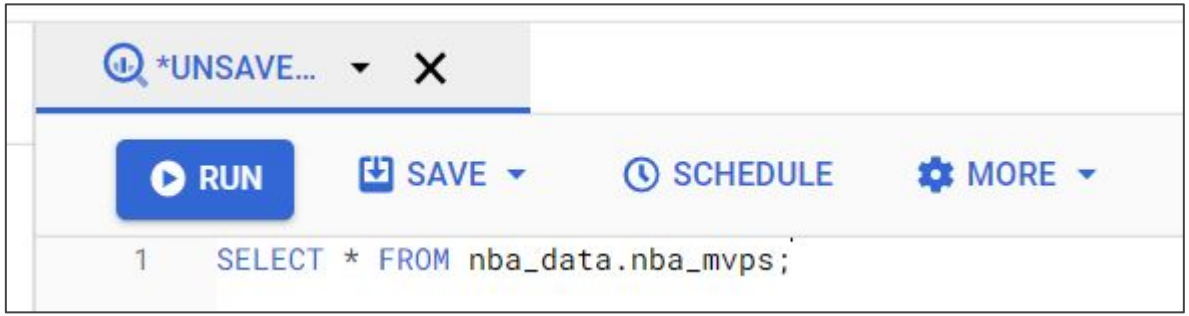
1. Repeat the process from the previous slide for the 2 CSV files you did not yet upload
2. You will then have 3 tables when you expand “nba_data” dataset:



Query your new dataset

1. Select “Compose new query” from the top-right

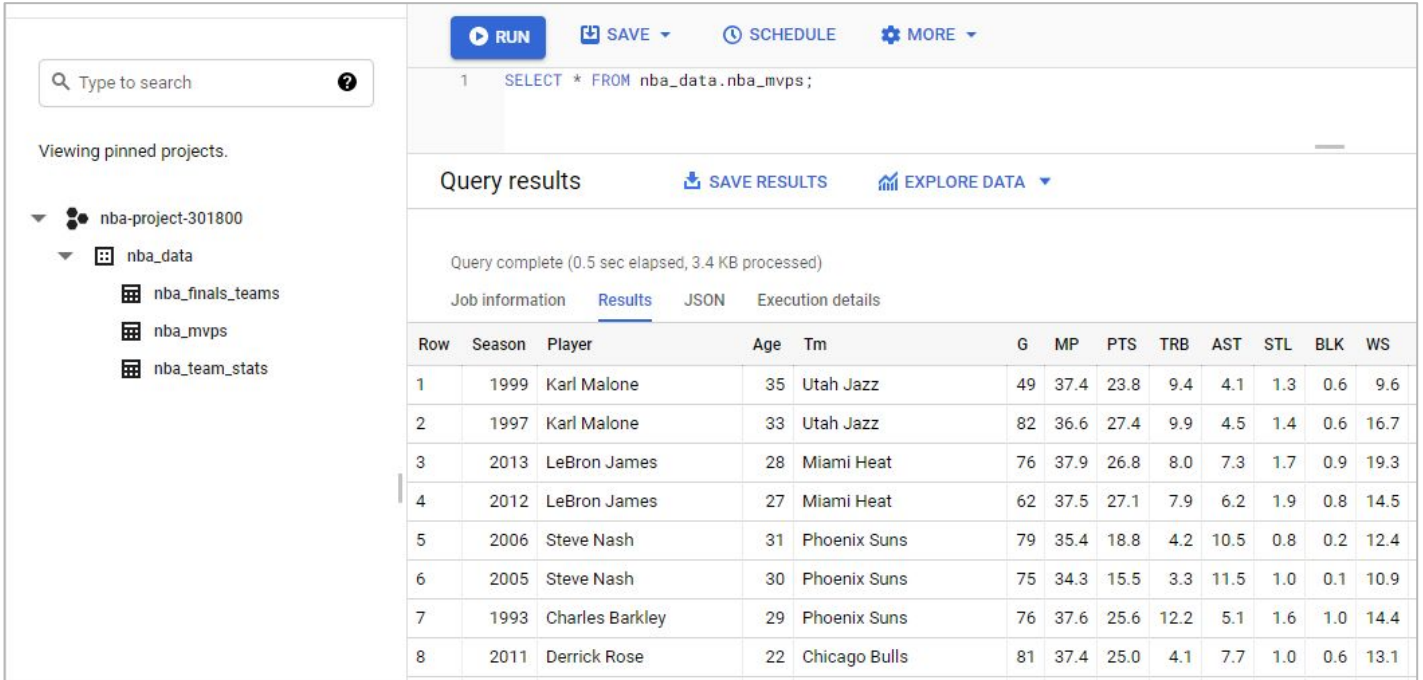
2. Type “SELECT * FROM nba_data.nba_mvps;”



3. Click on “Run”

Query Output

If everything was setup correctly, you should see the following output in your “Query results”



Row	Season	Player	Age	Tm	G	MP	PTS	TRB	AST	STL	BLK	WS
1	1999	Karl Malone	35	Utah Jazz	49	37.4	23.8	9.4	4.1	1.3	0.6	9.6
2	1997	Karl Malone	33	Utah Jazz	82	36.6	27.4	9.9	4.5	1.4	0.6	16.7
3	2013	LeBron James	28	Miami Heat	76	37.9	26.8	8.0	7.3	1.7	0.9	19.3
4	2012	LeBron James	27	Miami Heat	62	37.5	27.1	7.9	6.2	1.9	0.8	14.5
5	2006	Steve Nash	31	Phoenix Suns	79	35.4	18.8	4.2	10.5	0.8	0.2	12.4
6	2005	Steve Nash	30	Phoenix Suns	75	34.3	15.5	3.3	11.5	1.0	0.1	10.9
7	1993	Charles Barkley	29	Phoenix Suns	76	37.6	25.6	12.2	5.1	1.6	1.0	14.4
8	2011	Derrick Rose	22	Chicago Bulls	81	37.4	25.0	4.1	7.7	1.0	0.6	13.1