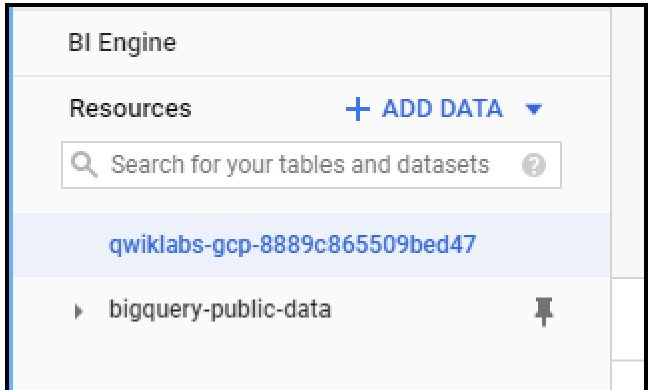
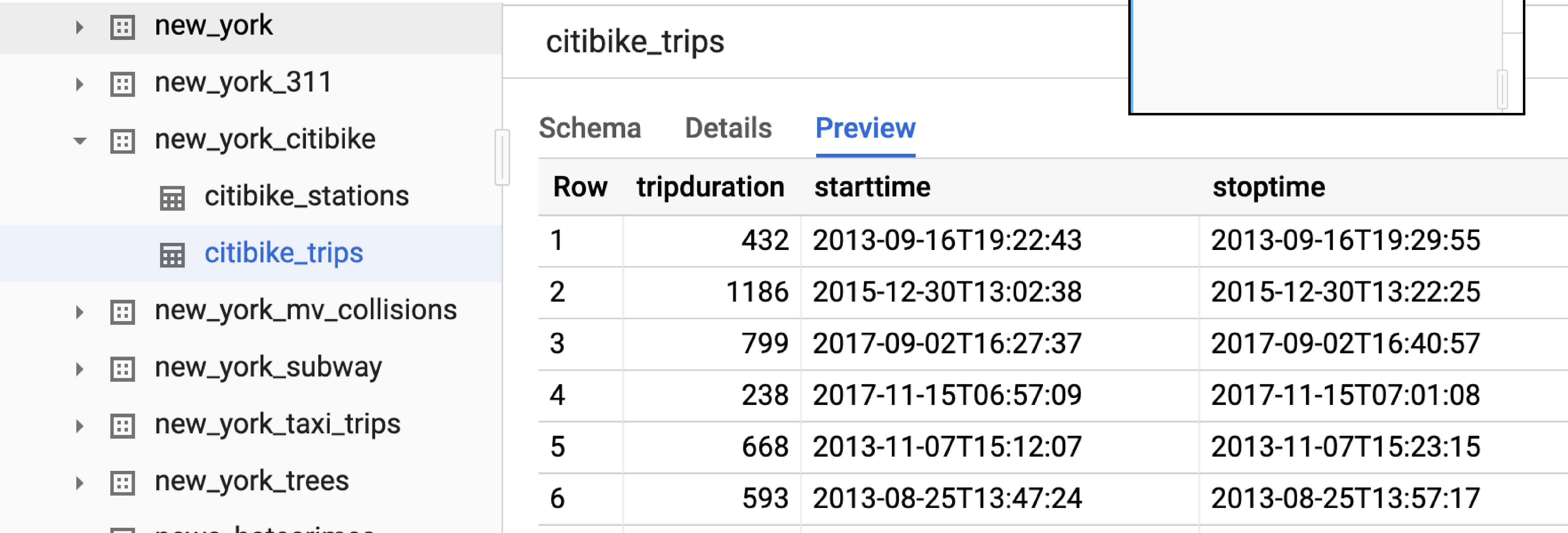
Using BigQuery to do Analysis

Access to public dataset and preview data

The Datasets window opens.
2.In the **Search** bar, type "NYC bike" then press **Enter**.
3.One result **NYC Citi Bike Trips** is returned. Click on the dataset name and then **View Dataset**.


BI Engine
Resources
+ ADD DATA
Search for your tables and datasets
qwiklabs-gcp-8889c865509bed47
bigquery-public-data



Row	tripduration	starttime	stoptime
1	432	2013-09-16T19:22:43	2013-09-16T19:29:55
2	1186	2015-12-30T13:02:38	2015-12-30T13:22:25
3	799	2017-09-02T16:27:37	2017-09-02T16:40:57
4	238	2017-11-15T06:57:09	2017-11-15T07:01:08
5	668	2013-11-07T15:12:07	2013-11-07T15:23:15
6	593	2013-08-25T13:47:24	2013-08-25T13:57:17

Explore data

```
SELECT
  MIN(start_station_name) AS start_station_name,
  MIN(end_station_name) AS end_station_name,
  APPROX_QUANTILES(tripduration, 10)[OFFSET (5)] AS
typical_duration,
  COUNT(tripduration) AS num_trips
FROM
  `bigquery-public-data.new_york_citibike.citibike_trips`
WHERE
  start_station_id != end_station_id
GROUP BY
  start_station_id,
  end_station_id
ORDER BY
  num_trips DESC
LIMIT
  10
```

(Hint: typical duration for the 10 most common one-way rentals)

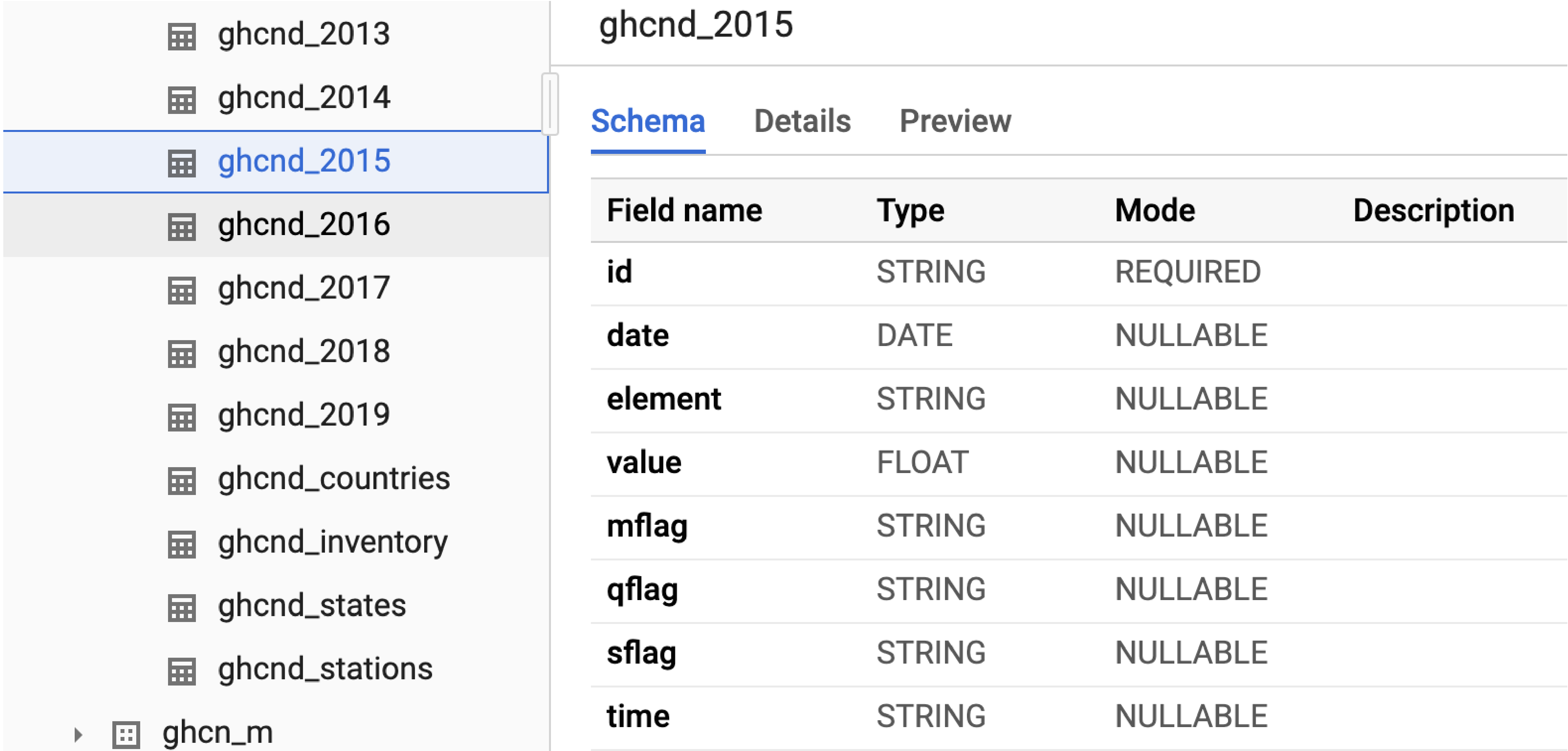
Explore data

```
WITH
 trip_distance AS (
SELECT
 bikeid,
 ST_Distance(ST_GeogPoint(s.longitude,
   s.latitude),
  ST_GeogPoint(e.longitude,
   e.latitude)) AS distance
FROM
 `bigquery-public-data.new_york_citibike.citibike_trips`,
 `bigquery-public-data.new_york_citibike.citibike_stations` as s,
 `bigquery-public-data.new_york_citibike.citibike_stations` as e
WHERE
 start_station_id = s.station_id
 AND end_station_id = e.station_id )
SELECT
 bikeid,
 SUM(distance)/1000 AS total_distance
FROM
 trip_distance
GROUP BY
 bikeid
ORDER BY total_distance DESC
LIMIT  5
```

total distance travelled by each bicycle in the dataset. Note that the query limits the results to only top 5

Access to the weather dataset

In the left pane of the BigQuery Console, select the newly added bigquery-public-data project and select **ghcn_d** > **ghcnd_2015**. Then click on the **Preview** tab. Your console should resemble the following:



Field name	Type	Mode	Description
id	STRING	REQUIRED	
date	DATE	NULLABLE	
element	STRING	NULLABLE	
value	FLOAT	NULLABLE	
mflag	STRING	NULLABLE	
qflag	STRING	NULLABLE	
sflag	STRING	NULLABLE	
time	STRING	NULLABLE	

Explore data

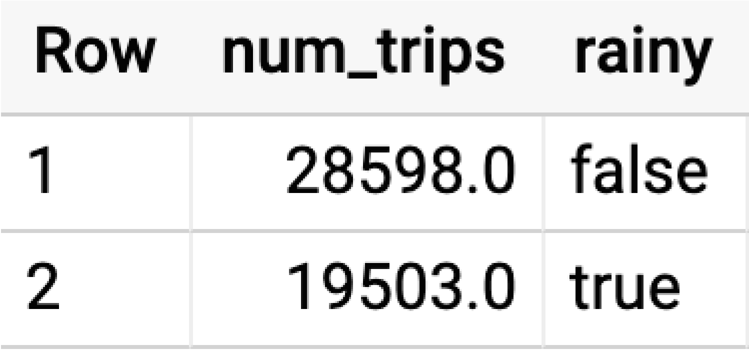
```
SELECT
  wx.date,
  wx.value/10.0 AS prcp
FROM
  `bigquery-public-data.ghcn_d.ghcnd_2015` AS wx
WHERE
  id = 'USW00094728'
  AND qflag IS NULL
  AND element = 'PRCP'
ORDER BY
  wx.date
```

rainfall (in mm) for all days in 2015 from a weather station in New York whose id is provided in the query

Explore data (**Find correlation between rain and bicycle rentals**)

```
WITH bicycle_rentals AS (
  SELECT
    COUNT(starttime) as num_trips,
    EXTRACT(DATE from starttime) as trip_date
  FROM `bigquery-public-data.new_york_citibike.citibike_trips`
  GROUP BY trip_date
),
rainy_days AS
(
SELECT
  date,
  (MAX(prcp) > 5) AS rainy
FROM (
  SELECT
    wx.date AS date,
    IF (wx.element = 'PRCP', wx.value/10, NULL) AS prcp
  FROM
    `bigquery-public-data.ghcn_d.ghcnd_2015` AS wx
  WHERE
    wx.id = 'USW00094728'
)
GROUP BY
  date
)
```

```
SELECT
  ROUND(AVG(bk.num_trips)) AS num_trips,
  wx.rainy
FROM bicycle_rentals AS bk
JOIN rainy_days AS wx
ON wx.date = bk.trip_date
GROUP BY wx.rainy
```



Row	num_trips	rainy
1	28598.0	false
2	19503.0	true