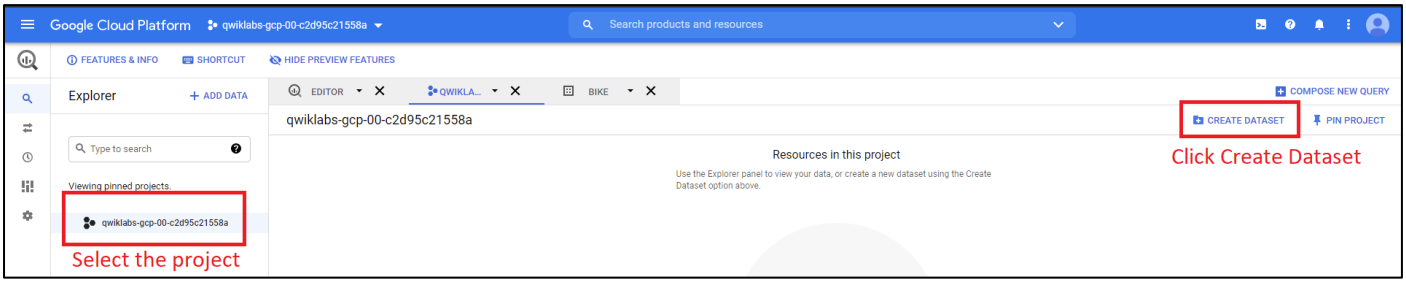
Create ML Models with BigQuery ML: Challenge Lab

Author: Vedant Kakde | **GitHub Profile:** github.com/vedant-kakde | **LinkedIn Profile:** linkedin.com/in/vedant-kakde/

Task 1: Create a dataset to store your machine learning models

This dataset will be used to store our BQML models for the challenge. The name of the dataset does not matter. Let's just call it bike. **Go to Bigquery -> Select Project -> Click Create Dataset**.



Task 2: Create a forecasting BigQuery machine learning model

Copy and paste these in Query editor then Run it one by one.

```
CREATE OR REPLACE
MODEL
bike.location_model
```

```
OPTIONS
  (model_type='linear_reg', labels=['duration_minutes']) AS
SELECT
    start_station_name,
    EXTRACT(HOUR FROM start_time) AS start_hour,
    EXTRACT(DAYOFWEEK FROM start_time) AS day_of_week,
    duration_minutes
FROM
    `bigquery-public-data.austin_bikeshare.bikeshare_trips` AS trips
JOIN
    `bigquery-public-data.austin_bikeshare.bikeshare_stations` AS stations
ON
    trips.start_station_name = stations.name
WHERE
    EXTRACT(YEAR FROM start_time) = 2018
    AND duration_minutes > 0
```

Author: Vedant Kakde | **GitHub Profile:** github.com/vedant-kakde | **LinkedIn Profile:** linkedin.com/in/vedant-kakde/

Task 3: Clean your training data

```
CREATE OR REPLACE
MODEL
bike.subscriber_model
                        OPTIONS
                          (model_type='linear_reg', labels=['duration_minutes']) AS
                        SELECT
                            start_station_name,
                            EXTRACT(HOUR FROM start_time) AS start_hour,
                            subscriber_type,
                            duration_minutes
                        FROM `bigquery-public-data.austin_bikeshare.bikeshare_trips` AS trips
                        WHERE EXTRACT(YEAR FROM start_time) = 2018
```

Task 4: Clean your training data

```
SELECT
          SQRT(mean_squared_error) AS rmse,
          mean_absolute_error
        FROM
          ML.EVALUATE(MODEL bike.location_model, (
          SELECT
            start_station_name,
            EXTRACT(HOUR FROM start_time) AS start_hour,
            EXTRACT(DAYOFWEEK FROM start_time) AS day_of_week,
            duration_minutes
          FROM
            `bigquery-public-data.austin_bikeshare.bikeshare_trips` AS trips
          JOIN
           `bigquery-public-data.austin_bikeshare.bikeshare_stations` AS stations
          ON
            trips.start_station_name = stations.name
          WHERE EXTRACT(YEAR FROM start_time) = 2019)
        )
```

```
SELECT
          SQRT(mean_squared_error) AS rmse,
          mean_absolute_error
        FROM
          ML.EVALUATE(MODEL bike.subscriber_model, (
          SELECT
            start_station_name,
            EXTRACT(HOUR FROM start_time) AS start_hour,
            subscriber_type,
            duration_minutes
          FROM
            `bigquery-public-data.austin_bikeshare.bikeshare_trips` AS trips
          WHERE
```

```
            EXTRACT(YEAR FROM start_time) = 2019)
        )
```

Task 5: Clean your training data

```
SELECT
          start_station_name,
          COUNT(*) AS trips
        FROM
          `bigquery-public-data.austin_bikeshare.bikeshare_trips`
        WHERE
          EXTRACT(YEAR FROM start_time) = 2019
        GROUP BY
          start_station_name
        ORDER BY
          trips DESC
```

```
  SELECT
AVG(predicted_duration_minutes)
AS
average_predicted_trip_length
FROM ML.predict(MODEL bike.subscriber_model, (
SELECT
    start_station_name,
    EXTRACT(HOUR FROM start_time) AS start_hour,
    subscriber_type,
    duration_minutes
FROM
  `bigquery-public-data.austin_bikeshare.bikeshare_trips`
WHERE
  EXTRACT(YEAR FROM start_time) = 2019
  AND subscriber_type = 'Single Trip'
  AND start_station_name = '21st & Speedway @PCL'))
```

Congratulations! You completed this challenge lab.

Author: Vedant Kakde | **GitHub Profile:** github.com/vedant-kakde | **LinkedIn Profile:** linkedin.com/in/vedant-kakde/