# GSP341: Create ML Models with BigQuery ML

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

In Cloud Shell

1bq mk austin

### Task 2: Create a forecasting BigQuery machine learning model

In BigQuery Console Query Editor

1CREATE OR REPLACE MODEL austin.location\_model

2OPTIONS

3 (model\_type='linear\_reg', labels=['duration\_minutes']) AS

4SELECT

5 start\_station\_name,

6 EXTRACT(HOUR FROM start\_time) AS start\_hour,

7 EXTRACT(DAYOFWEEK FROM start\_time) AS day\_of\_week,

8 duration\_minutes,

9FROM

10 `bigquery-public-data.austin\_bikeshare.bikeshare\_trips` AS trips

11JOIN

12 `bigquery-public-data.austin\_bikeshare.bikeshare\_stations` AS stations

13ON

14 trips.start\_station\_name = stations.name

15WHERE

16 EXTRACT(YEAR FROM start\_time) = 2018

17 AND duration\_minutes > 0

### Task 3: Create the second machine learning model

In BigQuery Console Query Editor

1CREATE OR REPLACE MODEL austin.subscriber\_model

2OPTIONS

3 (model\_type='linear\_reg', labels=['duration\_minutes']) AS

4SELECT

5 start\_station\_name,

6 EXTRACT(HOUR FROM start\_time) AS start\_hour,

7 subscriber\_type,

8 duration\_minutes

9FROM `bigquery-public-data.austin\_bikeshare.bikeshare\_trips` AS trips

10WHERE EXTRACT(YEAR FROM start\_time) = 2018

### Task 4: Evaluate the two machine learning models

In BigQuery Console Query Editor

#### Query - 1

1*-- Evaluation metrics for location\_model*

2SELECT

3 SQRT(mean\_squared\_error) AS rmse,

4 mean\_absolute\_error

5FROM

6 ML.EVALUATE(MODEL austin.location\_model, (

7 SELECT

8 start\_station\_name,

9 EXTRACT(HOUR FROM start\_time) AS start\_hour,

10 EXTRACT(DAYOFWEEK FROM start\_time) AS day\_of\_week,

11 duration\_minutes

12 FROM

13 `bigquery-public-data.austin\_bikeshare.bikeshare\_trips` AS trips

14 JOIN

15 `bigquery-public-data.austin\_bikeshare.bikeshare\_stations` AS stations

16 ON

17 trips.start\_station\_name = stations.name

18 WHERE EXTRACT(YEAR FROM start\_time) = 2019)

19)

#### Query - 2

1*-- Evaluation metrics for subscriber\_model*

2SELECT

3 SQRT(mean\_squared\_error) AS rmse,

4 mean\_absolute\_error

5FROM

6 ML.EVALUATE(MODEL austin.subscriber\_model, (

7 SELECT

8 start\_station\_name,

9 EXTRACT(HOUR FROM start\_time) AS start\_hour,

10 subscriber\_type,

11 duration\_minutes

12 FROM

13 `bigquery-public-data.austin\_bikeshare.bikeshare\_trips` AS trips

14 WHERE

15 EXTRACT(YEAR FROM start\_time) = 2019)

16)

### Task 5: Use the subscriber type machine learning model to predict average trip durations

In BigQuery Console Query Editor

#### Query - 1

1SELECT

2 start\_station\_name,

3 COUNT(\*) AS trips

4FROM

5 `bigquery-public-data.austin\_bikeshare.bikeshare\_trips`

6WHERE

7 EXTRACT(YEAR FROM start\_time) = 2019

8GROUP BY

9 start\_station\_name

10ORDER BY

11 trips DESC

#### Query - 2

1SELECT AVG(predicted\_duration\_minutes) AS average\_predicted\_trip\_length

2FROM ML.predict(MODEL austin.subscriber\_model, (

3SELECT

4 start\_station\_name,

5 EXTRACT(HOUR FROM start\_time) AS start\_hour,

6 subscriber\_type,

7 duration\_minutes

8FROM

9 `bigquery-public-data.austin\_bikeshare.bikeshare\_trips`

10WHERE

11 EXTRACT(YEAR FROM start\_time) = 2019

12 AND subscriber\_type = 'Single Trip'

13 AND start\_station\_name = '21st & Speedway @PCL'))