Engineer Data in Google Cloud : Challenge Lab

Task 1: Clean your training data

```
CREATE OR REPLACE TABLE
  taxirides.taxi_training_data AS
SELECT
  (tolls_amount + fare_amount) AS fare_amount,
  pickup_datetime,
  pickup_longitude AS pickuplon,
  pickup_latitude AS pickuplat,
  dropoff_longitude AS dropofflon,
  dropoff_latitude AS dropofflat,
  passenger_count AS passengers,
FROM
  taxirides.historical_taxi_rides_raw
WHERE
  RAND() < 0.001
  AND trip_distance > 0
  AND fare_amount >= 2.5
  AND pickup_longitude > -78
  AND pickup_longitude < -70
  AND dropoff_longitude > -78
  AND dropoff_longitude < -70
  AND pickup_latitude > 37
  AND pickup_latitude < 45
  AND dropoff_latitude > 37
  AND dropoff_latitude < 45
  AND passenger_count > 0
```

Task 2: Create a BQML model called `taxirides.fare_model`

```
CREATE OR REPLACE MODEL taxirides.fare_model
TRANSFORM(
  * EXCEPT(pickup_datetime)

  , ST_Distance(ST_GeogPoint(pickuplon, pickuplat), ST_GeogPoint(dropofflon, dropofflat)) AS
euclidean
  , CAST(EXTRACT(DAYOFWEEK FROM pickup_datetime) AS STRING) AS dayofweek
  , CAST(EXTRACT(HOUR FROM pickup_datetime) AS STRING) AS hourofday
)
OPTIONS(input_label_cols=['fare_amount'], model_type='linear_reg')
AS

SELECT * FROM taxirides.taxi_training_data
```

Task 3: Perform a batch prediction on new data

```
CREATE OR REPLACE TABLE taxirides.2015_fare_amount_predictions
  AS
SELECT * FROM ML.PREDICT(MODEL taxirides.fare_model,(
  SELECT * FROM taxirides.report_prediction_data)
)
```

Congratulations! You completed this challenge lab.