QUERYS DE BIG DATA

1

```sql
SELECT*
FROM
 ml.PREDICT(MODEL `taxi.taxifare_model`,
  (
 WITH params AS (
   SELECT
   1 AS TRAIN,
   2 AS EVAL
   ),

   daynames AS
    (SELECT ['Sun', 'Mon', 'Tues', 'Wed', 'Thurs', 'Fri', 'Sat'] AS daysofweek),

   taxitrips AS (
   SELECT
    (tolls_amount + fare_amount) AS total_fare,
    daysofweek[ORDINAL(EXTRACT(DAYOFWEEK FROM pickup_datetime))] AS dayofweek,
    EXTRACT(HOUR FROM pickup_datetime) AS hourofday,
    pickup_longitude AS pickuplon,
    pickup_latitude AS pickuplat,
    dropoff_longitude AS dropofflon,
    dropoff_latitude AS dropofflat,
    passenger_count AS passengers
   FROM
    `nyc-tlc.yellow.trips`, daynames, params
```

```
        WHERE

          trip_distance > 0 AND fare_amount > 0

          AND MOD(ABS(FARM_FINGERPRINT(CAST(pickup_datetime AS
STRING))),1000) = params.EVAL

        )


        SELECT *

        FROM taxitrips


        ));
```

2

```
SELECT

  SQRT(mean_squared_error) AS rmse

FROM

  ML.EVALUATE(MODEL taxi.taxifare_model,

  (

  WITH params AS (

    SELECT 1 AS TRAIN,2 AS EVAL),

  daynames AS

    (SELECT ['Sun', 'Mon', 'Tues', 'Wed', 'Thurs', 'Fri', 'Sat'] AS daysofweek),

  taxitrips AS (

  SELECT

    (tolls_amount + fare_amount) AS total_fare,

    daysofweek[ORDINAL(EXTRACT(DAYOFWEEK FROM pickup_datetime))] AS
dayofweek,

    EXTRACT(HOUR FROM pickup_datetime) AS hourofday,

    pickup_longitude AS pickuplon,

    pickup_latitude AS pickuplat,
```

```
    dropoff_longitude AS dropofflon,

    dropoff_latitude AS dropofflat,

    passenger_count AS passengers
  FROM
   `nyc-tlc.yellow.trips`, daynames, params
  WHERE
   trip_distance > 0 AND fare_amount > 0
   AND MOD(ABS(FARM_FINGERPRINT(CAST(pickup_datetime AS STRING))),1000) =
params.EVAL)


  SELECT *
  FROM taxitrips))
```

3

```
CREATE or REPLACE MODEL taxi.taxifare_model
OPTIONS
  (model_type='linear_reg', labels=['total_fare']) AS
WITH params AS (SELECT 1 AS TRAIN,2 AS EVAL),
    daynames AS
    (SELECT ['Sun', 'Mon', 'Tues', 'Wed', 'Thurs', 'Fri', 'Sat'] AS daysofweek),
    taxitrips AS (SELECT
    (tolls_amount + fare_amount) AS total_fare,
    daysofweek[ORDINAL(EXTRACT(DAYOFWEEK FROM pickup_datetime))] AS
dayofweek,
    EXTRACT(HOUR FROM pickup_datetime) AS hourofday,
    pickup_longitude AS pickuplon,
    pickup_latitude AS pickuplat,
    dropoff_longitude AS dropofflon,
    dropoff_latitude AS dropofflat,
```

```
    passenger_count AS passengers

    FROM `nyc-tlc.yellow.trips`, daynames, params

    WHERE

    trip_distance > 0 AND fare_amount > 0

    AND MOD(ABS(FARM_FINGERPRINT(CAST(pickup_datetime AS STRING))),1000) =
params.TRAIN)

  SELECT *

  FROM taxitrips
```

4

```
CREATE or REPLACE MODEL taxi.taxifare_model

OPTIONS

 (model_type='linear_reg', labels=['total_fare']) AS

WITH params AS (SELECT 1 AS TRAIN,2 AS EVAL),

    daynames AS

    (SELECT ['Sun', 'Mon', 'Tues', 'Wed', 'Thurs', 'Fri', 'Sat'] AS daysofweek),

    taxitrips AS (SELECT

    (tolls_amount + fare_amount) AS total_fare,

    daysofweek[ORDINAL(EXTRACT(DAYOFWEEK FROM pickup_datetime))] AS
dayofweek,

    EXTRACT(HOUR FROM pickup_datetime) AS hourofday,

    pickup_longitude AS pickuplon,

    pickup_latitude AS pickuplat,

    dropoff_longitude AS dropofflon,

    dropoff_latitude AS dropofflat,

    passenger_count AS passengers

    FROM `nyc-tlc.yellow.trips`, daynames, params

    WHERE

    trip_distance > 0 AND fare_amount > 0
```

```
    AND MOD(ABS(FARM_FINGERPRINT(CAST(pickup_datetime AS STRING))),1000) =
params.TRAIN)

 SELECT *

 FROM taxitrips
```

5

```
WITH params AS (SELECT 1 AS TRAIN,2 AS EVAL),

   daynames AS

   (SELECT ['Sun', 'Mon', 'Tues', 'Wed', 'Thurs', 'Fri', 'Sat'] AS daysofweek),

    taxitrips AS (SELECT

    (tolls_amount + fare_amount) AS total_fare,

    daysofweek[ORDINAL(EXTRACT(DAYOFWEEK FROM pickup_datetime))] AS
dayofweek,

    EXTRACT(HOUR FROM pickup_datetime) AS hourofday,

    pickup_longitude AS pickuplon,

    pickup_latitude AS pickuplat,

    dropoff_longitude AS dropofflon,

    dropoff_latitude AS dropofflat,

    passenger_count AS passengers

    FROM `nyc-tlc.yellow.trips`, daynames, params

    WHERE

    trip_distance > 0 AND fare_amount > 0

    AND MOD(ABS(FARM_FINGERPRINT(CAST(pickup_datetime AS STRING))),1000) =
params.TRAIN)
```