Take Home Test

Chenchen Liang
05.18.2016

Question 1: What's the average speed of yellow cabs on odd days of the week (Monday, Wednesday, Friday, Sunday)?

Query:

SELECT
  DAYOFWEEK(pickup_datetime) as day_of_week,
  ROUND (
  AVG(trip_distance/((dropoff_datetime-pickup_datetime)/3600000000)), 2)
as Speed
----convert service time to hour as denominator
FROM
  [nyc-tlc:yellow.trips]
WHERE
 trip_distance > 0 and DAYOFWEEK(pickup_datetime) in (1,3,5,7)
----exclude trips with bad record and even days (Tuesday, Thursday, and Saturday)
GROUP BY 1
ORDER BY 1

result:

Row	day_of_week	Speed
1	1	44.89
2	3	30.99
3	5	26.11
4	7	26.24

Question 2: Using data from the first half of 2015, compute the difference in average fare-per-mile for yellow cabs vs. green cabs.

Query:
SELECT a.Green_rate_2015, b.Yellow_rate_2015
FROM
(SELECT ROUND(avg(fare_amount/trip_distance),2) as Green_rate_2015
FROM [nyc-tlc.green.trips_2015]) as a
CROSS JOIN
----select unrelated two tables
(SELECT Round(avg(fare_amount/trip_distance),2) as Yellow_rate_2015
FROM [nyc-tlc.yellow.trips]
WHERE year(pickup_datetime) =2015) as b
----make sure both tables only contain data of first half of 2015

Result:

Row	Green_rate_2015	Yellow_rate_2015
1	8.28	6.45

In the month of March 2015, what was the average fare-per-mile for green cabs by day of week? Your results should include a date, yellow fare, and green fare.

I assume you are asking average fare-per-mile for both cabs in March 2015 by requiring results with both yellow fare, and green fare.

Query:

```
SELECT A.day_of_week, A.avg_green_fare_rate, B.avg_yellow_fare_rate
FROM
(SELECT
DAYOFWEEK(pickup_datetime) as day_of_week,
ROUND(AVG(fare_amount/trip_distance), 2) as avg_green_fare_rate
----calculate average fare-per-mile for green cabs
FROM [nyc-tlc.green.trips_2015]
WHERE
MONTH(pickup_datetime) =3
GROUP BY 1
ORDER BY 1
) AS A

LEFT JOIN

(SELECT
DAYOFWEEK(pickup_datetime) as day_of_week,
ROUND(AVG(fare_amount/trip_distance), 2) as avg_yellow_fare_rate
----calculate average fare-per-mile for yellow cabs
FROM [nyc-tlc.yellow.trips]
WHERE
YEAR(pickup_datetime) =2015 and MONTH(pickup_datetime) =3
----since this table contain data with multiple years, only choose 2015
GROUP BY 1
ORDER BY 1
) AS B
 ON A.day_of_week = B.day_of_week
```

Result:

Row	day_of_week	avg_green_fare_rate	avg_yellow_fare_rate
1	1	8.23	6.19
2	2	8.83	6.31
3	3	8.64	6.74
4	4	9.45	6.61
5	5	9.62	6.67
6	6	9.34	6.56

| 7 | 7 | 8.84 | 6.14 |