-- Average trip distance

SELECT DATE(pickup\_datetime) date, AVG(trip\_distance) trip\_distance,

FROM [nyc-tlc:yellow.trips]

GROUP BY 1

ORDER BY 1

**Ans. Huge variation on the avg\_distance per day. Also, some negative values and not consistent**

-- clean data (only look at trips with fare\_amount/trip\_distance ratio between 1 and 10)

SELECT SUM( fare\_amount/trip\_distance BETWEEN 1 AND 10)/ COUNT(\*)

FROM [nyc-tlc:yellow.trips]

**Ans 0.9557608689222268 i.e. 95.58% of the data**

SELECT DATE(pickup\_datetime) date, AVG(trip\_distance) trip\_distance,

FROM [nyc-tlc:yellow.trips]

WHERE fare\_amount/trip\_distance BETWEEN 1 AND 10

GROUP BY 1

ORDER BY 1

**Ans. Now average trip distance is around 3 miles with less variations**

SELECT DAYOFWEEK(DATE(TIMESTAMP(pickup\_datetime))) dayofweek,

INTEGER(100\*AVG(trip\_distance/((dropoff\_datetime-pickup\_datetime)/3600000000)))/100 speed

FROM [nyc-tlc:yellow.trips]

WHERE fare\_amount/trip\_distance BETWEEN 1 AND 10

AND HOUR(pickup\_datetime) BETWEEN 8 AND 23

GROUP BY 1

ORDER BY 1

**Ans.**

**Row dayofweek speed**

**1 1 14.86**

**2 2 13.31**

**3 3 12.32**

**4 4 12.13**

**5 5 12.04**

**6 6 12.03**

**7 7 13.02**

SELECT DAYOFWEEK(DATE(TIMESTAMP(pickup\_datetime))) dayofweek,

INTEGER(100\*AVG(trip\_distance/((dropoff\_datetime-pickup\_datetime)/3600000000)))/100 speed

FROM [nyc-tlc:yellow.trips]

WHERE fare\_amount/trip\_distance BETWEEN 1 AND 10

AND HOUR(pickup\_datetime) BETWEEN 8 AND 23

AND (DAYOFWEEK(DATE(TIMESTAMP(pickup\_datetime))) = 2

OR DAYOFWEEK(DATE(TIMESTAMP(pickup\_datetime))) = 4

OR DAYOFWEEK(DATE(TIMESTAMP(pickup\_datetime))) = 6

OR DAYOFWEEK(DATE(TIMESTAMP(pickup\_datetime))) = 1)

GROUP BY 1

ORDER BY 1

Ans.

**Row dayofweek speed**

**1 1 14.86**

**2 2 13.31**

**4 4 12.13**

**6 6 12.03**