William Zhu

Professor Vaghefi

10/29/2021

SELECT ROUND(sum(fare\_amount),30) as Total\_fare\_jan

FROM `cis-4400-homework-classwork.SQL\_2.Yellowcab\_2020`

where passenger\_count >= 2 and RatecodeID=1 and payment\_type= 1 or payment\_type=2

and fare\_amount >0

union distinct

SELECT ROUND(sum(fare\_amount),2) as Total\_fare\_feb

FROM `cis-4400-homework-classwork.SQL\_2.Yellowcab\_2020\_Feb`

where passenger\_count >= 2 and RatecodeID=1 and payment\_type= 1 or payment\_type=2

and fare\_amount >0

union distinct

SELECT ROUND(sum(fare\_amount),2) as Total\_fare\_mar

FROM `cis-4400-homework-classwork.SQL\_2.Yellowcab\_2020\_Mar`

where passenger\_count >= 2 and RatecodeID=1 and payment\_type= 1 or payment\_type=2

and fare\_amount >0

union distinct

SELECT ROUND(sum(fare\_amount),2) as Total\_fare\_apr

FROM `cis-4400-homework-classwork.SQL\_2.Yellowcab\_2020\_Apr`

where passenger\_count >= 2 and RatecodeID=1 and payment\_type= 1 or payment\_type=2

and fare\_amount >0

**Question 2**

SELECT sum(fare\_amount)/sum(trip\_distance) as Average\_fare #jan

FROM `cis-4400-homework-classwork.SQL\_2.Yellowcab\_2020`

where RatecodeID != 2 and RatecodeID != 3 and

trip\_distance between 8 and 20

union distinct

SELECT sum(fare\_amount)/sum(trip\_distance) as Average\_fare #feb

FROM `cis-4400-homework-classwork.SQL\_2.Yellowcab\_2020\_Feb`

where RatecodeID != 2 and RatecodeID != 3 and

trip\_distance between 8 and 20

union distinct

SELECT sum(fare\_amount)/sum(trip\_distance) as Average\_fare #mar

FROM `cis-4400-homework-classwork.SQL\_2.Yellowcab\_2020\_Mar`

where RatecodeID != 2 and RatecodeID != 3 and

trip\_distance between 8 and 20

union distinct

SELECT sum(fare\_amount)/sum(trip\_distance) as Average\_fare #Apr

FROM `cis-4400-homework-classwork.SQL\_2.Yellowcab\_2020\_Apr`

where RatecodeID != 2 and RatecodeID != 3 and

trip\_distance between 8 and 20

union distinct

SELECT sum(fare\_amount)/sum(trip\_distance) as Average\_fare # may and june return same value (3.001). Perhaps because of union distinct?

FROM `cis-4400-homework-classwork.SQL\_2.Yellowcab\_2020\_May`

where RatecodeID != 2 and RatecodeID != 3 and

trip\_distance between 8 and 20

union distinct

SELECT sum(fare\_amount)/sum(trip\_distance) as Average\_fare #June

FROM `cis-4400-homework-classwork.SQL\_2.Yellowcab\_2020\_Jun`

where RatecodeID != 2 and RatecodeID != 3 and

trip\_distance between 8 and 20

union distinct

SELECT sum(fare\_amount)/sum(trip\_distance) as Average\_fare #July

FROM `cis-4400-homework-classwork.SQL\_2.Yellowcab\_2020\_Jul`

where RatecodeID != 2 and RatecodeID != 3 and

trip\_distance between 8 and 20

union distinct

SELECT sum(fare\_amount)/sum(trip\_distance) as Average\_fare #August

FROM `cis-4400-homework-classwork.SQL\_2.Yellowcab\_2020\_Aug`

where RatecodeID != 2 and RatecodeID != 3 and

trip\_distance between 8 and 20

union distinct

SELECT sum(fare\_amount)/sum(trip\_distance) as Average\_fare #September

FROM `cis-4400-homework-classwork.SQL\_2.Yellowcab\_2020\_Sept`

where RatecodeID != 2 and RatecodeID != 3 and

trip\_distance between 8 and 20

union distinct

SELECT sum(fare\_amount)/sum(trip\_distance) as Average\_fare #October

FROM `cis-4400-homework-classwork.SQL\_2.Yellowcab\_2020\_Oct`

where RatecodeID != 2 and RatecodeID != 3 and

trip\_distance between 8 and 20

union distinct

SELECT sum(fare\_amount)/sum(trip\_distance) as Average\_fare #November

FROM `cis-4400-homework-classwork.SQL\_2.Yellowcab\_2020\_Nov`

where RatecodeID != 2 and RatecodeID != 3 and

trip\_distance between 8 and 20

union distinct

SELECT sum(fare\_amount)/sum(trip\_distance) as Average\_fare #December

FROM `cis-4400-homework-classwork.SQL\_2.Yellowcab\_2020\_Dec`

where RatecodeID != 2 and RatecodeID != 3 and

trip\_distance between 8 and 20

Returns one value for every month. Row != month

**Question 3**

**For this question I created two queries to achieve desired format**

SELECT extract(dayofweek from tpep\_pickup\_datetime) as weekday,

count(vendorID) as trips

FROM `cis-4400-homework-classwork.sql\_2\_2018.Jan`

where passenger\_count =1

group by weekday

union distinct

SELECT extract(dayofweek from tpep\_pickup\_datetime) as weekday,

count(vendorID) as trips

FROM `cis-4400-homework-classwork.sql\_2\_2018.Feb`

where passenger\_count =1

group by weekday

union distinct

SELECT extract(dayofweek from tpep\_pickup\_datetime) as weekday,

count(vendorID) as trips

FROM `cis-4400-homework-classwork.sql\_2\_2018.Mar`

where passenger\_count =1

group by weekday

union distinct

SELECT extract(dayofweek from tpep\_pickup\_datetime) as weekday,

count(vendorID) as trips

FROM `cis-4400-homework-classwork.sql\_2\_2018.Apr`

where passenger\_count =1

group by weekday

union distinct

SELECT extract(dayofweek from tpep\_pickup\_datetime) as weekday,

count(vendorID) as trips

FROM `cis-4400-homework-classwork.sql\_2\_2018.May`

where passenger\_count =1

group by weekday

union distinct

SELECT extract(dayofweek from tpep\_pickup\_datetime) as weekday,

count(vendorID) as trips

FROM `cis-4400-homework-classwork.sql\_2\_2018.Jun`

where passenger\_count =1

group by weekday

union distinct

SELECT extract(dayofweek from tpep\_pickup\_datetime) as weekday,

count(vendorID) as trips

FROM `cis-4400-homework-classwork.sql\_2\_2018.Jul`

where passenger\_count =1

group by weekday

union distinct

SELECT extract(dayofweek from tpep\_pickup\_datetime) as weekday,

count(vendorID) as trips

FROM `cis-4400-homework-classwork.sql\_2\_2018.Aug`

where passenger\_count =1

group by weekday

union distinct

SELECT extract(dayofweek from tpep\_pickup\_datetime) as weekday,

count(vendorID) as trips

FROM `cis-4400-homework-classwork.sql\_2\_2018.Sept`

where passenger\_count =1

group by weekday

union distinct

SELECT extract(dayofweek from tpep\_pickup\_datetime) as weekday,

count(vendorID) as trips

FROM `cis-4400-homework-classwork.sql\_2\_2018.Nov`

where passenger\_count =1

group by weekday

union distinct

SELECT extract(dayofweek from tpep\_pickup\_datetime) as weekday,

count(vendorID) as trips

FROM `cis-4400-homework-classwork.sql\_2\_2018.Dec`

where passenger\_count =1

group by weekday

union distinct

SELECT extract(dayofweek from tpep\_pickup\_datetime) as weekday,

count(vendorID) as trips

FROM `cis-4400-homework-classwork.sql\_2\_2018.Apr`

where passenger\_count =1

group by weekday

A picture containing application

Description automatically generated

Table

Description automatically generated

Results show sum of trips for 7 days of the week for 12 months

2nd code where I saved the results from the previous query to group.

SELECT weekday, sum(trips) as NumberOfTrips

FROM `cis-4400-homework-classwork.sql\_2\_2018.Query\_results\_Q3`

group by weekday

order by NumberOfTrips

limit 1;

Results show total trips for each day of the week in 2018

Table

Description automatically generated with low confidence

**Question 4**

SELECT

EXTRACT(DATE from tpep\_pickup\_datetime) as Date,

EXTRACT(TIME from tpep\_pickup\_datetime) as Pickup\_Time,

EXTRACT(TIME from tpep\_dropoff\_datetime) as Dropoff\_Time,

trip\_distance, RatecodeID, fare\_amount, round((total\_amount/trip\_distance),2) as Dollars\_per\_Mile

FROM `cis-4400-homework-classwork.SQL\_2.Yellowcab\_2020` as Yellowcab\_Jan

where

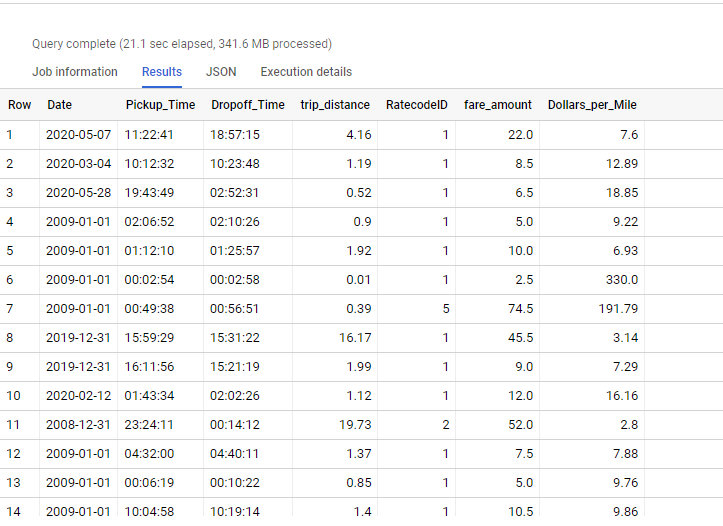
tolls\_amount = 0 and

trip\_distance < 30 and

trip\_distance >0 and #filtering negative trip distance

total\_amount > 0 and #filtering negative Total\_amount

fare\_amount > 0; # filtering out 1 instance where fare amount < 0 and total\_amount > 0



1. Write a paragraph to answer to explain:
   1. What you learned from the assignment.
   2. What did you find particularly challenging
   3. How long it took from you to complete it
   4. What would you change about the assignment if you could.

One thing I learned from this assignment is using Buckets to load multiple data sets into a cloud to be called upon for analysis later. This was one of the more repetitive tasks as well as calling the data later on using a ‘FROM’ clause. I felt like there should be an easier way to call upon multiple data sets but could not find a way. Getting used to extracting date/time took some time getting familiar with. This assignment took me about a day with a few times where I had to review sql code/formats. For question 3, I saved a query of 12 months x 7 weekdays, to group by weekdays. One thing I would change about this assignment is creating one dataset instead of 12