BigQuery SQL Code For Data Exploration

Set up data

```
with tmp_tb as (
SELECT
    unique_key,
    DATETIME(trip_start_timestamp, 'America/Chicago') trip_start_timestamp,
    DATETIME(trip_end_timestamp, 'America/Chicago') trip_end_timestamp,
    trip_miles,
    pickup_census_tract,
    dropoff_census_tract,
    pickup_community_area,
    dropoff_community_area,
    payment_type,
    company,
    pickup_latitude,
    pickup_longitude,
    dropoff_latitude,
    dropoff_longitude,
    fare fare_dollars,
    (fare/trip_miles) as fare_per_mile,
    CAST(trip_start_timestamp AS DATE) trip_start_dt,
    CAST(trip_end_timestamp AS DATE) trip_end_dt,
    DATETIME_DIFF(trip_end_timestamp, trip_start_timestamp, MINUTE) trip_minutes,
    EXTRACT(YEAR FROM trip_start_timestamp) year,
    EXTRACT(MONTH FROM trip_start_timestamp) month,
    EXTRACT(DAY FROM trip_start_timestamp) day,
    EXTRACT(HOUR FROM trip_start_timestamp) hour,
    FORMAT_DATE('%a', DATE(trip_start_timestamp)) weekday,
    CASE WHEN (pickup_community_area not in (56, 64, 76)) OR (dropoff_community_area not in (5
6, 64, 76)) THEN 1 else 0 END is_outside_airport,
    CASE WHEN (pickup_community_area in (56,64,76)) OR (dropoff_community_area in (56,64,76))
THEN 1 else 0 END is_airport,
    CASE WHEN (pickup_community_area in (8,32,28)) OR (dropoff_community_area in (8,32,28)) TH
EN 1 else 0 END is_navy_pier
FROM
    `bigquery-public-data.chicago_taxi_trips.taxi_trips`
)
SELECT *
FROM tmp_tb
WHERE year > 2020 and trip_miles != 0 and fare_dollars != 0 and trip_miles is not null and pay
ment_type != "Unknown"


## Query number of rides per year
SELECT year, count(1) as number_of_rides
FROM `chicago-taxi-analysis-373620.374.Chicago_Taxi_Full`
group by year
order by year

/* Number of rides by month overall */
SELECT month, count(1) as number_of_rides
```

```
FROM `chicago-taxi-analysis-373620.374.Chicago_Taxi_Full`
group by month
order by month

/* Average fare per year */
SELECT year,avg(fare_dollars) as average_fare
FROM `chicago-taxi-analysis-373620.374.chicago-taxi-all-trips`
group by year
order by year

/* Total Fare sum overall per year */
SELECT year, sum(fare_dollars) as total_amount
FROM `chicago-taxi-analysis-373620.374.chicago-taxi-all-trips`
GROUP BY year
ORDER BY year

/* Number of rides by month overall */
SELECT month, avg(fare_dollars) as average_fare
FROM `chicago-taxi-analysis-373620.374.chicago-taxi-all-trips`
WHERE year = 2022
GROUP BY month
ORDER BY month

SELECT month, sum(fare_dollars) as total_amount
FROM `chicago-taxi-analysis-373620.374.chicago-taxi-all-trips`
WHERE year = 2022
GROUP BY month
ORDER BY month


/*Query number of rides per day overall*/
SELECT day, count(1) as number_of_rides
FROM `chicago-taxi-analysis-373620.374.chicago-taxi-all`
GROUP BY day
ORDER BY day

/* Query number of users by hour of day */
SELECT hour, count(1) as number_of_rides
FROM `chicago-taxi-analysis-373620.374.chicago-taxi-all`
GROUP BY hour
ORDER BY hour

SELECT hour, day, year, count(1) as number_of_rides
FROM `chicago-taxi-analysis-373620.374.chicago-taxi-all`
GROUP BY hour, day, year
ORDER BY hour, day, year

/* Load new table and create max variable */
SELECT year,hour, max(number_of_rides) as max_rides
FROM `chicago-taxi-analysis-373620.374.rides_info`
GROUP BY year, hour, number_of_rides
ORDER BY year, hour, number_of_rides

/* Find out top companies based on number of rides*/
```

```sql
SELECT company, count(1) AS number_of_rides
FROM `chicago-taxi-analysis-373620.374.chicago-taxi-all`
WHERE company IS NOT NULL
GROUP BY company
ORDER BY number_of_rides DESC
LIMIT 5


/* 2013 top companies*/
SELECT company, count(1) as number_of_rides
FROM `chicago-taxi-analysis-373620.374.chicago-taxi-all`
WHERE company is not null and year = 2013
GROUP BY company
ORDER BY number_of_rides desc
LIMIT 10

/*2014 top companies*/
SELECT company, count(1) as number_of_rides
FROM `chicago-taxi-analysis-373620.374.chicago-taxi-all`
WHERE company is not null and year = 2014
GROUP BY company
ORDER BY number_of_rides desc
LIMIT 10

/* Convert dropoff community area into new string to use for a new categorical variable
telling the area of the dropoff based on map */
SELECT *, CAST(dropoff_community_area as STRING) as area
FROM `chicago-taxi-analysis-373620.374.chicago-taxi-all`;

SELECT *,
    CASE

    WHEN area in ('76') THEN 'ORD Airport'
    WHEN area in ('1','2','3','4','5','6') THEN 'East Lincolnwood/Wrigley'
    WHEN area in ('7','8') then 'Navy Pier Area'
    WHEN area in ('9','10','11','12','13','14','15','16','17') then 'South Lincolnwood/Norridg
e'
    WHEN area in ('18','19','20','21','22','23','24','25') then 'Humboldt Park Area'
    WHEN area in ('26','27') then 'Garfield Park'
    WHEN area in ('28') then 'UIC'
    WHEN area in ('29','30','31') then 'Lawndale/Canalport'
    WHEN area in ('32','33') then 'Soldier Field/Grant Park'
    WHEN area in ('34','35','36','37','38','39') then 'Grand Boulevard/IIT'
    WHEN area in ('40','41','42') then 'UOC'
    WHEN area in ('43','44','45','46') then 'Chatham/Windsor Park Area'
    WHEN area in ('47','48') then 'Calumet Heights Area'
    WHEN area in ('49','50') then 'CSU Area'
    WHEN area in ('51','52','53') then 'Calumet Park/East Side'
    WHEN area in ('54','55') then 'Dolton Area'
    WHEN area in ('56','57','58','64') then 'MDW Airport'
    WHEN area in ('59','60','61','62','63') then 'Glendale/New City Area'
    WHEN area in ('66','67','68','69') then 'Englewood Area'
    WHEN area in ('70','71','72','73') then 'Evergreen Park/South Englewood'
    WHEN area in ('74','75') then 'Morgan Park Area'
```

```
    ELSE 'Loyola/Edgewater'
    END AS dropoff_area
FROM `chicago-taxi-analysis-373620.374.area_info`

SELECT *,

/*Create variables for cost per mile and cost per duration*/
SELECT *,
(fare_dollars/trip_miles) as cost_per_mile,
(fare_dollars/trip_minutes) as cost_per_minute
FROM `chicago-taxi-analysis-373620.374.chicago-taxi-all-area`
WHERE trip_miles !=0 and trip_minutes != 0

/* Query data from only 2021 and 2022 with no extreme fare prices */

SELECT *,
FROM `chicago-taxi-analysis-373620.374.chicago-taxi-all-trips`
WHERE year in (2021,2022) and company is not null and cost_per_mile < 20000
LIMIT 2500000

/* Create new categorical variable for pickup area based on map along with a mph variable*/
SELECT *, CAST(pickup_community_area as STRING) as pickup_area
FROM `chicago-taxi-analysis-373620.374.chicago-taxi-all-trips`
WHERE year in (2021,2022) and company is not null and cost_per_mile < 20000
LIMIT 2500000
SELECT *,
    CASE

    WHEN pickup_area in ('76') THEN 'ORD Airport'
    WHEN pickup_area in ('1','2','3','4','5','6') THEN 'East Lincolnwood/Wrigley'
    WHEN pickup_area in ('7','8') then 'Navy Pier Area'
    WHEN pickup_area in ('9','10','11','12','13','14','15','16','17') then 'South Lincolnwood/
Norridge'
    WHEN pickup_area in ('18','19','20','21','22','23','24','25') then 'Humboldt Park Area'
    WHEN pickup_area in ('26','27') then 'Garfield Park'
    WHEN pickup_area in ('28') then 'UIC'
    WHEN pickup_area in ('29','30','31') then 'Lawndale/Canalport'
    WHEN pickup_area in ('32','33') then 'Soldier Field/Grant Park'
    WHEN pickup_area in ('34','35','36','37','38','39') then 'Grand Boulevard/IIT'
    WHEN pickup_area in ('40','41','42') then 'UOC'
    WHEN pickup_area in ('43','44','45','46') then 'Chatham/Windsor Park Area'
    WHEN pickup_area in ('47','48') then 'Calumet Heights Area'
    WHEN pickup_area in ('49','50') then 'CSU Area'
    WHEN pickup_area in ('51','52','53') then 'Calumet Park/East Side'
    WHEN pickup_area in ('54','55') then 'Dolton Area'
    WHEN pickup_area in ('56','57','58','64') then 'MDW Airport'
    WHEN pickup_area in ('59','60','61','62','63') then 'Glendale/New City Area'
    WHEN pickup_area in ('66','67','68','69') then 'Englewood Area'
    WHEN pickup_area in ('70','71','72','73') then 'Evergreen Park/South Englewood'
    WHEN pickup_area in ('74','75') then 'Morgan Park Area'

    ELSE 'Loyola/Edgewater’
    END AS pick_up_area,
```

```
   (trip_miles/trip_minutes)*60 as mph
FROM `chicago-taxi-analysis-373620.374.chicago-taxi-all-trips-new`
```