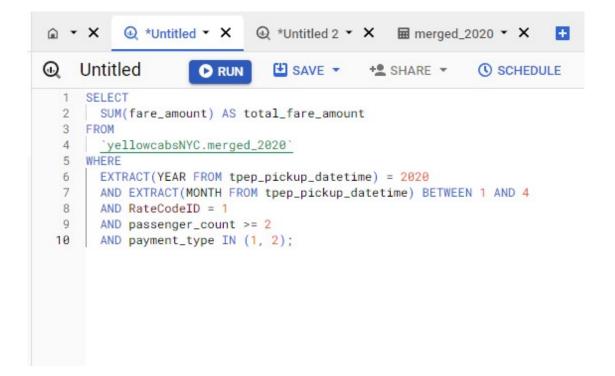
```
CREATE TABLE `yellowcabsNYC.merged_2020` AS
SELECT * FROM `yellowcabsNYC.jan2020`
UNION ALL
SELECT * FROM `yellowcabsNYC.feb2020`
UNION ALL
SELECT * FROM `yellowcabsNYC.march2020`
UNION ALL
SELECT * FROM `yellowcabsNYC.april2020`
UNION ALL
SELECT * FROM `yellowcabsNYC.may2020`
UNION ALL
SELECT * FROM `yellowcabsNYC.june2020`
UNION ALL
SELECT * FROM `yellowcabsNYC.july2020`
UNION ALL
SELECT * FROM `yellowcabsNYC.august2020`
UNION ALL
SELECT * FROM `yellowcabsNYC.sep2020`
UNION ALL
SELECT * FROM `yellowcabsNYC.oct2020`
UNION ALL
SELECT * FROM `yellowcabsNYC.nov2020`
UNION ALL
SELECT * FROM `yellowcabsNYC.dec2020`;
```

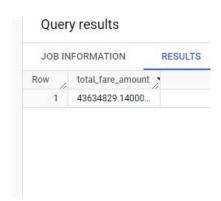
1) For the first four months of 2020, find the total fare amount (one record) for yellow taxi trips taken, charged at the standard rate, with 2 or more passengers who paid with a credit card or cash.

```
SELECT
   SUM(fare_amount) AS total_fare_amount
FROM
```

```
`yellowcabsNYC.merged_2020`
WHERE

EXTRACT(YEAR FROM tpep_pickup_datetime) = 2020
AND EXTRACT(MONTH FROM tpep_pickup_datetime) BETWEEN 1 AND 4
AND RateCodeID = 1
AND passenger_count >= 2
AND payment_type IN (1, 2);
```





2) What was the average fare amount per mile of a yellow taxi standard rate ride in 2020 travelled more than 8 miles but less than 20 miles (excluding trips made to airports i.e. RateCodeID 2, 3 and 4)? (Result should be 1 record).

```
SELECT

AVG(fare_amount / trip_distance) AS average_fare_per_mile
FROM
```

```
`yellowcabsNYC.merged 2020`
WHERE
 EXTRACT (YEAR FROM tpep pickup datetime) = 2020
  AND trip distance > 8
  AND trip distance < 20
  AND RateCodeID = 1;
Untitled

    SAVE ▼ + SHARE ▼
                                                        ( SCHE
                   RUN
 1 SELECT
    AVG(fare_amount / trip_distance) AS average_fare_per_mile
 4
      'yellowcabsNYC.merged_2020'
 5 WHERE
      EXTRACT(YEAR FROM tpep_pickup_datetime) = 2020
 6
 7
      AND trip_distance > 8
 8
      AND trip_distance < 20
     AND RateCodeID = 1;
 Query results
                                      CHART PREVIEW
 JOB INFORMATION
                        RESULTS
WOS
        average_fare_per_mj
```

3) Which day of the week (from Monday-Sunday) in 2020 has the lowest number of single rider trips on average? (Result should be one record)

```
WITH SingleRiderTrips AS (
SELECT
DATE(tpep pickup datetime) AS pickup date,
```

1

3.091674939942...

```
CASE WHEN passenger count = 1 THEN 1 ELSE 0 END AS
is single rider
    FROM
        `demo1-401820.yellowcabsNYC.merged 2020`
    WHERE
        EXTRACT(YEAR FROM tpep pickup datetime) = 2020
)
SELECT
    EXTRACT (DAYOFWEEK FROM pickup date) AS day of week,
   AVG(is_single_rider) AS average_single_rider_trips
FROM
    SingleRiderTrips
GROUP BY
   day of week
ORDER BY
    average single rider trips ASC
LIMIT 1;
```

Query results

JOB INFO	DRMATION	RESULTS	CHART PREVIEW	JSON
Row /	day_of_week ▼	average_sin	gle_rider	
1	7	0.67314427	2027	

```
O Untitled
                     O RUN
                               SAVE -
                                            + SHARE *
                                                           ( SCHEDULE
  1 WITH trip_counts AS (
  2
       SELECT
         EXTRACT(DAYOFWEEK FROM tpep_pickup_datetime) AS pickup_day,
  3
         COUNT(*) AS trip_count
  4
  5
       FROM
         'yellowcabsNYC.merged_2020'
  6
  7
       WHERE
  8
        EXTRACT(YEAR FROM tpep_pickup_datetime) = 2020
  9
        AND passenger_count = 1
 10
       GROUP BY
 11
        pickup_day
 12
     SELECT
 13
       CASE
 14
 15
         WHEN pickup_day = 1 THEN 'Sunday'
 16
         WHEN pickup_day = 2 THEN 'Monday'
 17
         WHEN pickup_day = 3 THEN 'Tuesday'
         WHEN pickup_day = 4 THEN 'Wednesday'
 18
         WHEN pickup_day = 5 THEN 'Thursday'
 19
         WHEN pickup_day = 6 THEN 'Friday'
 20
 21
         WHEN pickup_day = 7 THEN 'Saturday'
 22
      END AS day_with_lowest_single_rider_trips
 23 FROM
 24 trip_counts
 25 ORDER BY
 26 trip_count ASC
 27 LIMIT 1;
```

Query results JOB INFORMATION RESULTS Row day_with_lowest_single_rider_trips 1 Sunday

4) ETL: Create new dataset from Jan 2020 data which includes the following columns

```
DATE(tpep pickup datetime) AS pickup date,
  TIME (tpep pickup datetime) AS pickup time,
  TIME (tpep dropoff datetime) AS dropoff time,
  trip distance,
  RatecodeID,
  fare amount,
  IF(trip distance > 0, total amount / trip distance, NULL) AS
fare per mile
FROM
  `yellowcabsNYC.merged 2020`
WHERE
 tolls amount = 0
 AND trip distance < 30
  AND tpep pickup datetime IS NOT NULL
  AND tpep dropoff datetime IS NOT NULL
  AND trip distance IS NOT NULL
  AND RatecodeID IS NOT NULL
  AND fare amount IS NOT NULL
 AND fare amount > 0
  AND trip distance > 0;
(Another way):
WITH January2020Trips AS (
    SELECT
        DATE(tpep pickup datetime) AS pickup date,
        TIME (tpep pickup datetime) AS pickup time,
        TIME(tpep_dropoff_datetime) AS dropoff_time,
        trip distance,
       RatecodeID,
        fare amount,
        CASE
            WHEN trip distance > 0 THEN total amount /
trip distance
            ELSE NULL
        END AS fare per mile
        `demo1-401820.yellowcabsNYC.merged 2020`
    WHERE
```

```
EXTRACT (YEAR FROM tpep pickup datetime) = 2020
        AND EXTRACT (MONTH FROM tpep pickup datetime) = 1
        AND tolls amount = 0
        AND trip distance < 30
SELECT *
FROM January2020Trips;
Truncated version:
[ {
  "pickup date": "2020-10-24",
  "pickup time": "00:17:07",
  "dropoff time": "00:38:01",
  "trip distance": "6.7",
  "RatecodeID": "1.0",
  "fare amount": "21.0",
  "fare_per_mile": "3.7014925373134329"
}, {
  "pickup date": "2020-08-09",
  "pickup time": "00:47:26",
  "dropoff time": "01:10:50",
  "trip distance": "6.08",
  "RatecodeID": "1.0",
  "fare amount": "21.0",
  "fare per mile": "4.5723684210526319"
}, {
  "pickup date": "2020-02-27",
  "pickup time": "22:16:19",
  "dropoff time": "22:41:03",
  "trip distance": "7.0",
  "RatecodeID": "1.0",
  "fare amount": "24.0",
  "fare per mile": "4.6857142857142851"
}, {
  "pickup date": "2020-11-20",
  "pickup time": "21:57:18",
  "dropoff time": "22:28:06",
  "trip distance": "6.77",
```

```
"RatecodeID": "1.0",
  "fare amount": "25.0",
 "fare per mile": "5.104874446085673"
}, {
 "pickup date": "2020-11-11",
 "pickup time": "10:35:06",
  "dropoff time": "10:56:02",
  "trip distance": "8.03",
 "RatecodeID": "1.0",
  "fare amount": "25.0",
 "fare per mile": "4.5815691158156913"
}, {
 "pickup date": "2020-01-07",
 "pickup time": "13:53:14",
 "dropoff time": "14:13:08",
 "trip distance": "8.57",
 "RatecodeID": "1.0",
 "fare amount": "25.0",
 "fare per mile": "3.6126021003500584"
},
```

5) In this assignment, I learned how to use SQL queries to work with a dataset. I understood how to filter and transform data, which is crucial in dealing with real-world datasets. Debugging SQL queries was a challenge, requiring careful attention to detail to get the right results. The assignment took about three hours, including understanding the task, writing queries, and making them accurate. It would be helpful to practice complex queries in class before doing them as homework.