Final Project - Query Log

1. Query to find saturated cities

```
SELECT city, region, COUNT(*) as store_count
FROM `team-15-fa24-mgmt-582.safegraph.places`
WHERE top_category = 'Clothing Stores'
GROUP BY city, region
ORDER BY store_count DESC;
Identifies the number of clothing stores in each city and region.
```

2. Highest Avg visits for cities where population is more than 10000

```
SELECT
    DISTINCT p.city,
    AVG(v.raw_visit_counts) AS avg_visits,
    AVG(v.median_dwell) AS avg_dwell_time,
    SUM(pop.pop_total) AS pop_total_,
    SUM(clothing_store_counts.store_count) AS store_count
FROM
    `team-15-fa24-mgmt-582.safegraph.visits` AS v
JOIN
    `team-15-fa24-mgmt-582.safegraph.places` AS p ON v.safegraph_place_id =
p.safegraph_place_id
JOIN
    `team-15-fa24-mgmt-582.safegraph.cbg_demographics` AS pop ON v.poi_cbg = pop.cbg
JOIN
    (SELECT
         city,
         region,
         COUNT(*) AS store_count
     FROM
         `team-15-fa24-mgmt-582.safegraph.places`
     WHERE
         top_category = 'Clothing Stores'
     GROUP BY
         city, region) AS clothing_store_counts ON p.city = clothing_store_counts.city
WHERE
    TIMESTAMP_TRUNC(v.date_range_start, DAY) BETWEEN TIMESTAMP("2019-12-31") AND
TIMESTAMP("2020-04-01")
    AND p.top_category = 'Clothing Stores'
    AND p.region NOT IN ('GU', 'PR', 'AS', 'NI', 'MP')
```

```
GROUP BY
    p.city
HAVING
    SUM(pop.pop_total) >= 10000
    and sum(clothing_store_counts.store_count)<=1000
ORDER BY
    avg_visits DESC;
```

3. **Which cities have the highest average visits to clothing stores, and what are the demographic characteristics of these cities?**

```
    WITH age_population AS (
        SELECT
            p.cbg,pl.city,
            (`pop_m_25-29` + `pop_f_25-29` + `pop_m_30-34` + `pop_f_30-34` +
            `pop_m_35-39` + `pop_f_35-39` + `pop_m_40-44` + `pop_f_40-44`) AS
    pop_25_44,
            (`pop_m_15-17` + `pop_f_15-17` +`pop_f_18-19`+`pop_m_18-19`+ `pop_m_20` +
    `pop_f_20` +
            `pop_m_21` + `pop_f_21` + `pop_m_22-24` + `pop_f_22-24`) AS pop_15_24,
            (`pop_m_45-49` + `pop_f_45-49` + `pop_m_50-54` + `pop_f_50-54`) AS
    pop_45_54,
            (`pop_m_55-59` + `pop_f_55-59` + `pop_m_60-61` + `pop_f_60-61` +
            `pop_m_62-64` + `pop_f_62-64` + `pop_m_65-66` + `pop_f_65-66` +
            `pop_m_67-69` + `pop_f_67-69` + `pop_m_70-74` + `pop_f_70-74` +
            `pop_m_75-79` + `pop_f_75-79` + `pop_m_80-84` + `pop_f_80-84` +
            `pop_m_gte85` + `pop_f_gte85`) AS pop_55_plus
            ,p.pop_total
        FROM `team-15-fa24-mgmt-582.safegraph.cbg_demographics` p
        join  `team-15-fa24-mgmt-582.safegraph.visits` v on p.cbg= v.poi_cbg
        join `team-15-fa24-mgmt-582.safegraph.places` pl on v.safegraph_place_id=
    pl.safegraph_place_id
        WHERE pl.top_category = 'Clothing Stores'
    )
    SELECT
        city,
        Round(sum(a.pop_25_44)/NULLIF(sum(a.pop_total),0),2) as pop_25_44 ,
        ROUND(sum(a.pop_15_24) /NULLIF(sum(a.pop_total),0),2)as pop_15_24,
        ROUND(sum(a.pop_45_54)/NULLIF(sum(a.pop_total),0),2) as pop_45_54,
        ROUND(sum(a.pop_55_plus)/NULLIF(sum(a.pop_total),0),2) as pop_55_plus,
    FROM
    age_population a
    where city in ('Clifton Park',
    'Markham',
```

```
'Orange Park',
'Bethpage',
'Framingham',
'Catoosa',
'Doraville',
'Niceville',
'Trinity',
'Live Oak')
group by 1
```

4. **Which regions have the most clothing stores per capita?**

```
WITH store_count AS (
  SELECT
    p.city AS region,
    p.region AS state,  -- Selecting the state (region)
    COUNT(DISTINCT p.safegraph_place_id) AS store_count
  FROM `team-15-fa24-mgmt-582.safegraph.visits` v
  INNER JOIN `team-15-fa24-mgmt-582.safegraph.brands` b ON v.safegraph_brand_ids =
b.safegraph_brand_id
  INNER JOIN `team-15-fa24-mgmt-582.safegraph.places` p ON v.safegraph_place_id =
p.safegraph_place_id
  WHERE b.top_category = 'Clothing Stores'  -- Filtering for clothing stores
  GROUP BY p.city, p.region  -- Grouping by city and state
),
city_population AS (
  SELECT
    f.county_fips,  -- Keeping county FIPS for joining
    SUM(d.pop_total) AS total_population
  FROM `team-15-fa24-mgmt-582.safegraph.cbg_demographics` d
  INNER JOIN `team-15-fa24-mgmt-582.safegraph.cbg_fips` f ON RIGHT(d.cbg, 5) =
CONCAT(f.state_fips, f.county_fips)
  GROUP BY f.county_fips  -- Grouping by county FIPS
),
city_county_mapping AS (
  SELECT DISTINCT
    p.city AS region,
    p.region AS state,  -- Selecting state for mapping
    f.county_fips
  FROM `team-15-fa24-mgmt-582.safegraph.places` p
  INNER JOIN `team-15-fa24-mgmt-582.safegraph.cbg_fips` f ON p.city = f.county
  OR f.county LIKE CONCAT('%', p.city, '%')  -- Allowing for partial matches
)
```

```
SELECT
  sc.region AS city,
  sc.state,  -- Including state in the final selection
  SUM(sc.store_count) AS store_count,  -- Summing to avoid duplicates
  SUM(cp.total_population) AS total_population,  -- Summing population
  (SUM(sc.store_count) * 1.0 / NULLIF(SUM(cp.total_population), 0)) AS
stores_per_capita  -- Calculating stores per capita
FROM store_count AS sc
LEFT JOIN city_county_mapping AS cm ON sc.region = cm.region AND sc.state =
cm.state  -- Join on city and state
LEFT JOIN city_population AS cp ON cm.county_fips = cp.county_fips  -- Join on the
county FIPS
GROUP BY sc.region, sc.state  -- Grouping by city and state
ORDER BY stores_per_capita DESC;
```

5. Peak Shopping Hours

```
CREATE TABLE `team-15-fa24-mgmt-582.safegraph.clothing_store_popularity` (
 safegraph_place_id STRING,
 city STRING,
 region STRING,
 popularity_by_hour STRING,
 date_range_start TIMESTAMP,
 date_range_end TIMESTAMP,
 hour_0 INT64,
 hour_1 INT64,
 hour_2 INT64,
hour_3 INT64,
hour_4 INT64,
hour_5 INT64,
hour_6 INT64,
hour_7 INT64,
hour_8 INT64,
hour_9 INT64,
hour_10 INT64,
hour_11 INT64,
hour_12 INT64,
hour_13 INT64,
hour_14 INT64,
hour_15 INT64,
hour_16 INT64,
```

```
hour_17 INT64,
hour_18 INT64,
hour_19 INT64,
hour_20 INT64,
hour_21 INT64,
hour_22 INT64,
hour_23 INT64
);


INSERT INTO `team-15-fa24-mgmt-582.safegraph.clothing_store_popularity` (
 safegraph_place_id,
 city,
 region,
 date_range_start,
 date_range_end,
 popularity_by_hour,
 hour_0,
 hour_1,
 hour_2,
 hour_3,
 hour_4,
 hour_5,
 hour_6,
 hour_7,
 hour_8,
 hour_9,
 hour_10,
 hour_11,
 hour_12,
 hour_13,
 hour_14,
 hour_15,
 hour_16,
 hour_17,
 hour_18,
 hour_19,
 hour_20,
 hour_21,
```

```
 hour_22,
 hour_23
)
SELECT
 v.safegraph_place_id,
 p.city,
 p.region,
 v.date_range_start,
 v.date_range_end,
 v.popularity_by_hour,
 CAST(SPLIT(REPLACE(REPLACE(popularity_by_hour, '[', ''), ']', ''), ',')[OFFSET(0)]
AS INT64) AS hour_0,
 CAST(SPLIT(REPLACE(REPLACE(popularity_by_hour, '[', ''), ']', ''), ',')[OFFSET(1)]
AS INT64) AS hour_1,
 CAST(SPLIT(REPLACE(REPLACE(popularity_by_hour, '[', ''), ']', ''), ',')[OFFSET(2)]
AS INT64) AS hour_2,
 CAST(SPLIT(REPLACE(REPLACE(popularity_by_hour, '[', ''), ']', ''), ',')[OFFSET(3)]
AS INT64) AS hour_3,
 CAST(SPLIT(REPLACE(REPLACE(popularity_by_hour, '[', ''), ']', ''), ',')[OFFSET(4)]
AS INT64) AS hour_4,
 CAST(SPLIT(REPLACE(REPLACE(popularity_by_hour, '[', ''), ']', ''), ',')[OFFSET(5)]
AS INT64) AS hour_5,
 CAST(SPLIT(REPLACE(REPLACE(popularity_by_hour, '[', ''), ']', ''), ',')[OFFSET(6)]
AS INT64) AS hour_6,
 CAST(SPLIT(REPLACE(REPLACE(popularity_by_hour, '[', ''), ']', ''), ',')[OFFSET(7)]
AS INT64) AS hour_7,
 CAST(SPLIT(REPLACE(REPLACE(popularity_by_hour, '[', ''), ']', ''), ',')[OFFSET(8)]
AS INT64) AS hour_8,
 CAST(SPLIT(REPLACE(REPLACE(popularity_by_hour, '[', ''), ']', ''), ',')[OFFSET(9)]
AS INT64) AS hour_9,
 CAST(SPLIT(REPLACE(REPLACE(popularity_by_hour, '[', ''), ']', ''),
',')[OFFSET(10)] AS INT64) AS hour_10,
 CAST(SPLIT(REPLACE(REPLACE(popularity_by_hour, '[', ''), ']', ''),
',')[OFFSET(11)] AS INT64) AS hour_11,
 CAST(SPLIT(REPLACE(REPLACE(popularity_by_hour, '[', ''), ']', ''),
',')[OFFSET(12)] AS INT64) AS hour_12,
 CAST(SPLIT(REPLACE(REPLACE(popularity_by_hour, '[', ''), ']', ''),
',')[OFFSET(13)] AS INT64) AS hour_13,
```

```
 CAST(SPLIT(REPLACE(REPLACE(popularity_by_hour, '[', ''), ']', ''),
','')[OFFSET(14)] AS INT64) AS hour_14,
 CAST(SPLIT(REPLACE(REPLACE(popularity_by_hour, '[', ''), ']', ''),
','')[OFFSET(15)] AS INT64) AS hour_15,
 CAST(SPLIT(REPLACE(REPLACE(popularity_by_hour, '[', ''), ']', ''),
','')[OFFSET(16)] AS INT64) AS hour_16,
 CAST(SPLIT(REPLACE(REPLACE(popularity_by_hour, '[', ''), ']', ''),
','')[OFFSET(17)] AS INT64) AS hour_17,
 CAST(SPLIT(REPLACE(REPLACE(popularity_by_hour, '[', ''), ']', ''),
','')[OFFSET(18)] AS INT64) AS hour_18,
 CAST(SPLIT(REPLACE(REPLACE(popularity_by_hour, '[', ''), ']', ''),
','')[OFFSET(19)] AS INT64) AS hour_19,
 CAST(SPLIT(REPLACE(REPLACE(popularity_by_hour, '[', ''), ']', ''),
','')[OFFSET(20)] AS INT64) AS hour_20,
 CAST(SPLIT(REPLACE(REPLACE(popularity_by_hour, '[', ''), ']', ''),
','')[OFFSET(21)] AS INT64) AS hour_21,
 CAST(SPLIT(REPLACE(REPLACE(popularity_by_hour, '[', ''), ']', ''),
','')[OFFSET(22)] AS INT64) AS hour_22,
 CAST(SPLIT(REPLACE(REPLACE(popularity_by_hour, '[', ''), ']', ''),
','')[OFFSET(23)] AS INT64) AS hour_23

FROM `team-15-fa24-mgmt-582.safegraph.visits` v
JOIN `team-15-fa24-mgmt-582.safegraph.places` p
 ON v.safegraph_place_id = p.safegraph_place_id
WHERE p.top_category = 'Clothing Stores';

New Query:
SELECT
   city,
   SUM(hour_0) AS total_hour_0,
   SUM(hour_1) AS total_hour_1,
   SUM(hour_2) AS total_hour_2,
   SUM(hour_3) AS total_hour_3,
   SUM(hour_4) AS total_hour_4,
   SUM(hour_5) AS total_hour_5,
   SUM(hour_6) AS total_hour_6,
   SUM(hour_7) AS total_hour_7,
   SUM(hour_8) AS total_hour_8,
   SUM(hour_9) AS total_hour_9,
```

```
   SUM(hour_10) AS total_hour_10,
   SUM(hour_11) AS total_hour_11,
   SUM(hour_12) AS total_hour_12,
   SUM(hour_13) AS total_hour_13,
   SUM(hour_14) AS total_hour_14,
   SUM(hour_15) AS total_hour_15,
   SUM(hour_16) AS total_hour_16,
   SUM(hour_17) AS total_hour_17,
   SUM(hour_18) AS total_hour_18,
   SUM(hour_19) AS total_hour_19,
   SUM(hour_20) AS total_hour_20,
   SUM(hour_21) AS total_hour_21,
   SUM(hour_22) AS total_hour_22,
   SUM(hour_23) AS total_hour_23
FROM
   `team-15-fa24-mgmt-582.safegraph.clothing_store_popularity`

WHERE city in ('Clifton Park', 'Markham',
'Orange Park',
'Bethpage',
'Framingham',
'Catoosa',
'Doraville',
'Niceville',
'Trinity',
'Live Oak')

group by city
```