QUERY LOG

Data exploration:
1. **CBG Demographics, CBG FIPS**
   a. **What is CBG?**
      As per google, A Census Block Group is a statistical geographic unit used by the United States Census Bureau. It is a subdivision of a census tract and consists of a cluster of census blocks.

   b. **How are CBG connected to State and County FIPS**
      The CBG is supposed to be a 12 digit code, where the first two digits are the State FIPS, and the next three digits are the County FIPS. Hence the CBG_Demographics table gives us the demographic data while the CBG_Fips table gives us the state and county information.

      Query 1:
      ```sql
      SELECT
        cbg
      FROM
        `finalprojectmod.Finalmod.cbg_demographics`
        where cbg like '01%';
      ```
      Result: Null
      Insights: Some CBG are only 11 digits and the states which have FIPS starting from 0 (eg. 01, 02..) do not have any corresponding CBG codes. Therefore, we assume that the first digit has been truncated in these cases.

      Query 2:
      ```sql
      SELECT
        cbg
      FROM
        `finalprojectmod.Finalmod.cbg_demographics`
      WHERE
        cbg LIKE '1%' AND LENGTH(cbg) = 11;
      ```
      Result: 3925 records
      Insights: There are 3925 CBG present which should have started from 01 instead of 1, and therefore have only 11 digits instead of 12 as expected.

Query 3:

```sql
SELECT
  distinct state, state_fips
FROM
  `finalprojectmod.Finalmod.cbg_fips`
  ORDER BY state_fips;
```

Result: 56 records

Insights: Some state_Fips like '03' are not present.

Query 4:

```sql
SELECT
  cbg
FROM
  `finalprojectmod.Finalmod.cbg_demographics`
WHERE
  cbg LIKE '3%' AND LENGTH(cbg) = 11;
```

Result: Null

Insights: There are no records with CBG with state fips 03. We cross checked with google and found that codes like 03, 07 are officially not FIP State codes

c. **Our focus state is California, how many counties and CBGs are present in it**

Query 5:

```sql
SELECT count (distinct county_fips)
FROM
  `finalprojectmod.Finalmod.cbg_fips`
  where state = 'CA';
```

Result: 58

Query 6:

```sql
SELECT
  count (cbg)
FROM
  `finalprojectmod.Finalmod.cbg_demographics`
WHERE
  cbg LIKE '6%' AND LENGTH(cbg) = 11;
```

Result: 25607

d. **What are the demographic information that can be extracted from CBG Table**
   As per our observation from understanding the table, we have identified the following broad categories that are available to us:
   Male population , Male Age , Female population, Female age , Education, Income range, Rent range, Home value range

2. **Brands**
   a. **What is the significance of NAICS Code?**
      We found the distinct number of NAICS Code, top categories and subcategory codes. We found that the distinct number of NAICS codes is equal to the distinct number of sub categories and assume that each sub category has one particular NAICS code affiliated to it.

   Query 7:
   ```
   SELECT
     distinct top_category
   FROM
     `finalprojectmod.Finalmod.brands`;
   ```
   Result: 182 records

   Query 8:
   ```
   SELECT
     distinct naics_code
   FROM
     `finalprojectmod.Finalmod.brands`;
   ```
   Result: 348 records

   Query 9:
   ```
   SELECT
     distinct sub_category
   FROM
     `finalprojectmod.Finalmod.brands`;
   ```
   Result: 348 records

b. What are the most brands affiliated with distinct sub categories
   Query 10:

```sql
SELECT
   count (brand_name) as count_brand, sub_category, top_category
FROM
   `finalprojectmod.Finalmod.brands`
   group by sub_category, top_category
   order by count_brand desc ;
```

Result: The top Category 'Restaurants and Other Eating Places' with the sub category 'Full-Service Restaurants' has the highest count of brands, i.e 962

c. Our focus top category is clothing , how many sub categories and accordingly number of brands are available

   Query 11:

```sql
SELECT
   count (brand_name) as count_brand, sub_category
FROM
   `finalprojectmod.Finalmod.brands`
   where top_category= 'Clothing Stores'
   group by sub_category
   order by count_brand desc  ;
```

Result:

S.No.	Number of Brands	Sub categories
1	138	Family Clothing Stores
2	93	Women's Clothing Stores
3	37	Other Clothing Stores
4	18	Men's Clothing Stores
5	17	Clothing Accessories Stores
6	13	Children's and Infants' Clothing Stores

d. What is the significance of parent_safegraph_brand_id and parent_safegraph_place_id?
   We chose and arbitrary parent_safegraph_brand_id from preview table and searched it in the safegraph_brand_id

Query 12:
```
SELECT
  *
FROM
  `finalprojectmod.Finalmod.brands`
where safegraph_brand_id = 'SG_BRAND_e60bc2632185dce0'
;
```
Result: 1 Record

Query 13:
```
SELECT
  *
FROM
  `finalprojectmod.Finalmod.brands`
where parent_safegraph_brand_id = 'SG_BRAND_e60bc2632185dce0'
;
```
Result: 6 Records

Insights: There are a few parent record IDs, which have subsidiaries affiliated to them. We have interpreted this as a parent- child record, where if a particular brand_id is present as a parent_brand_id for other records, we assume other records to be a subsidiary of this brand.

3. **Places**

    a. **Can the NAICS codes in this table be joined with the brands table?**

    Query 14:
    ```
    SELECT
      distinct naics_code,  sub_category
    FROM
      `finalprojectmod.Finalmod.places`;
    ```
    Result: 350 records
    Insights: There are a few extra records in this table compared to brands table

b. **Which are extra NAICS codes here**

Query 15:
```sql
SELECT
  distinct naics_code
FROM
  `finalprojectmod.Finalmod.places`
  where naics_code not in (select distinct naics_code from
`finalprojectmod.Finalmod.brands`);
```
Result: 68 Records

Insight: There are 68 NAICS codes that are present in places table but are not present in brands table, while vice versa gives us null values, implying all NAICS codes in brands table are present in places table.

c. **Since our focus is on clothing brands, we will only be choosing brands that are present in both tables (brands, places) for comprehensive results**

Query 16:
```sql
SELECT
  distinct sub_category
FROM
  `finalprojectmod.Finalmod.places`
  where top_category = 'Clothing Stores' and sub_category not in (select
distinct sub_category from `finalprojectmod.Finalmod.brands`);
```

Result: Null (vice versa as well)

Insights: All sub_categories and naics codes from brands table are present in places table and vice versa is true as well (ran 4 different queries)

4. **Visits**
   a. **Since our focus is on clothing brands in California, how can we leverage this table's data for insight extraction - for preliminary purposes we have taken out raw visit counts by location**

   Query 17:
```sql
SELECT
  distinct location_name, sum(raw_visit_counts) as sum_visit
FROM
  `finalprojectmod.Finalmod.visits`
```

```sql
    where  region = 'CA' and safegraph_brand_ids in
    (select distinct safegraph_brand_id from
`finalprojectmod.Finalmod.brands` where top_category = 'Clothing Stores')
    group by location_name;
```

Result: The location 'Marshalls' has the highest sum of raw count visits, i.e. 196568