

# Group 3 - MOD Project:

Avanti Chandratre  
Chhaya Tundwal  
Jayesh Chaudhari

## 1. Executive Summary

Our objective is to identify optimal locations for opening a new Drug Store outlet in the United States. By analyzing demographic data, foot traffic patterns, and existing hospital information, we aim to highlight areas with high potential for drug store success. This data will then inform our recommendations for a suitable client seeking promising store locations.

### Key Findings:

- **Brand Focus:** There are 59 brands that match the NAICS code 446110, which represents pharmacies and drug stores. After a thorough analysis, we identified one standout brand: **Safeway Pharmacy**.
- **Optimum Location:** After detailed analysis, we found that **King County in Washington** is the optimum location to start the business. It has a high concentration of the adult population (25% aged 55-84) and excellent footfall in the nearby hospitals, making it a strong candidate for a successful store launch.
- **Outlets Analysis:** On average, drug store brands in the U.S. have 466.06 outlets. Major outliers in the data are **CVS** (9,906 outlets), **Walgreens** (8,605 outlets), and **RiteAid** (3,292 outlets). These outliers are dominant, while smaller, local brands such as **Clark's Rx** operate with as few as 1 outlet. After removing such outliers, the average number of outlets for the remaining brands provides a more realistic picture of expansion opportunities.
- **Geographic Dominance:** Many brands tend to have a stronghold in a specific state. This suggests a strategic opportunity for expansion in states where those particular brands have little or no presence.
- **Wholesaler Concentration:** The highest number of drug wholesalers are in **Oklahoma**, which could be a factor when considering supply chain logistics for new outlets.
- **Data Gaps:** It is important to note that we do not have population data for certain states, including Alabama, Alaska, Arizona, Arkansas, California, Colorado, and Connecticut, as well as specific counties like Bristol Bay Borough, Denali Borough, and Kenai Peninsula Borough. This limits our analysis in these areas, and they may offer additional opportunities that we were unable to evaluate.

This analysis positions **Safeway Pharmacy** as a strong candidate for expansion in underrepresented areas, particularly in **King County, Washington**, where demographic trends and

hospital foot traffic suggest high success potential. Detailed analysis of these key findings are discussed in this report.

## 2. Introduction

### **Background:**

Healthcare has always been a vital aspect of any country's well-being, and providing quality healthcare facilities is essential. These facilities should include easily accessible pharmacies and drug stores, ensuring that patients can conveniently obtain necessary medications when visiting hospitals. Motivated by this gap in healthcare access, we used data-driven analysis to identify ideal locations for drug stores near hospitals. Our goal is to enhance healthcare accessibility while offering strategic business solutions. By focusing on well-performing drug stores with fewer than average locations, we identified potential areas for expansion that can benefit both the business and the community.

**Objectives:** The purpose of this report is to determine the best location for opening a new store. Our team has focused on analyzing key data to answer the following questions:

1. Which brand should we focus on?
2. Which county and state will provide the optimal location to maximize profit while minimizing risk?
3. Which street should the store be opened on?

## 3. Methodology

### **Assumptions:**

- **Brand Expansion Potential:** Brands with fewer stores are more likely to expand, so we focused on those with fewer existing outlets.
- **Footfall as a Success Indicator:** Higher foot traffic in existing stores signals a successful brand, so we focused on brands with footfall above the industry average.
- **Healthcare Demand as a Key Driver:** Locations near hospitals with high footfall are likely to have greater demand for drug stores. The assumption is that areas with high footfall in hospitals are reflective of overall healthcare needs.
- **Targeting Older Populations:** Older demographics (ages 55 to 84) are more likely to frequent hospitals and pharmacies. Therefore, counties with a higher proportion of older populations were considered prime locations.

### **Thought Process:**

Our methodology revolves around maximizing profit while minimizing risk. To do this, we first focused on identifying a brand that has potential for expansion but is not overly saturated in the market. We targeted mid-sized brands with strong performance in terms of foot traffic, avoiding

outliers like CVS, Walgreens, and RiteAid, which already has broad coverage and limited potential for expansion.

Once the brand was selected, we moved to location analysis. We assumed that areas with high healthcare demand—measured by footfall near hospitals—are prime for pharmacy expansion. This assumption is based on the idea that drug stores will thrive near hospitals with high traffic. In parallel, we also focused on counties where the population is more likely to require frequent healthcare services, especially older individuals aged 55 to 84.

## 4. Results

### 1. Following are the results for Query 1: Deciding on optimum brand

- a. The 5 Drug Store Brands with the number of outlets just above the median number of Drug Store brand outlets are shown in table 1.1
- b. The 5 Drug Store Brands with the number of visits just above the median number of Footfall are shown in table 1.2

The combinations present in their intersection are Rite Aid, Safeway Pharmacy and Benzer Pharmacy. But since Rite Aid is an outlier we didn't go for that brand. Secondly, Benzer Pharmacy didn't have many location options that were closer to hospitals.

Therefore we went with the brand Safeway Pharmacy, which had a good opportunity to expand their business.

Brand	Outlet Count
Benzer Pharmacy	120
Duane Reade by Walgreens	136
The Medicine Shoppe	300
Safeway Pharmacy	400
Rite Aid	3292

Table 1.1: Based on Drug stores outlets

Brand	Outlet count
Discount Drug Mart	76
Benzer Pharmacy	114
Medicap Pharmacy	75
Rite Aid	3253
Safeway Pharmacy	424

Table 1.2: Based on Footfall(visits)

2. Following are the results for Query 2: Deciding on optimum state+county combination
  - a. The 5 State+County combinations which are just below the median number of Drug Store brand outlets are shown in table 2.1
  - b. The 10 State+County combinations which are just above the median hospital footfall are shown in table 2.2

The only state+county combination present in their intersection is King County in Washington state

State	County
WA	King County
VA	Fairfax County
OR	Multnomah County
MD	Montgomery County
WA	Pierce County

Table 2.1: Based on Drug stores outlets

State	County
OK	Cleveland County
MO	Laclede County
NM	Santa Fe County
IN	Dearborn County
MA	Hampden County
MT	Missoula County
OH	Greene County
NY	Bronx County
WA	King County
TN	Franklin County

Table 2.2: Based on nearby Hospital footfall

3. Following are the results for query 3: Deciding on exact location to open the store(Postal code and street address)  
State: Washington, County: King County

City	Postal Code	Street Address
Kirkland	98034	12040 NE 128th St Ste 1600
Renton	98055	400 S 43rd St
Renton	98055	4033 Talbot Rd S Ste 450
Renton	98055	4033 Talbot Rd S Ste 530
Auburn	98002	1 E Main St
Issaquah	98027	1740 NW Maple St Ste 100
Renton	98055	Main Hospital 400 S. 43rd St

Given that Renton has multiple potential locations, it presents more business opportunities. Opening a store on one of these streets in Renton would be a strategically sound decision based on the high hospital foot traffic and population dynamics in the area.

## 4. Discussion

**Implications of the Findings:** The analysis provides key insights for optimal store placement:

**Targeting the Right Population:**

Areas like Renton (98055), Seattle (98104), and Auburn (98002) have high percentages of older adults, a key demographic for drug stores. Locating the store in these areas ensures steady demand from frequent healthcare visitors.

**Leveraging Hospital Foot Traffic:**

Locations near high-traffic hospitals, such as 4033 Talbot Rd S in Renton and Main Hospital West Clinic in Seattle, attract consistent footfall from patients and healthcare workers, boosting store visibility.

**Suburban Growth Potential:**

Emerging areas like Issaquah and Maple Valley show moderate foot traffic and a growing aging population, offering long-term growth potential while avoiding overcrowded markets.

These insights help align the store's placement with demographic needs and high footfall areas, maximizing profit and reducing risk.

**Limitations:** Only locations with available population data were considered for this analysis. There may be other counties or states that could have been a better fit, but due to the absence of population data, they were not included in the evaluation. This limits the comprehensiveness of the findings and may overlook potentially more suitable locations.

## 5. Conclusion and Recommendations

**Actionable Recommendations:**

- **Target Brand Expansion:** Focus on partnering with brands like Safeway Pharmacy, which are mid-sized and looking to expand. These brands show strong performance but are not yet oversaturated in the market.
- **Prioritize King County, Washington:** Based on the analysis, King County has a high proportion of the adult population and significant hospital footfall, making it an optimal location for a new store.
- **Expand Demographic Analysis:** As future steps, consider collecting or purchasing more comprehensive demographic data for other counties and states to ensure no viable location is overlooked due to data gaps.
- **Monitor Competition:** Although RiteAid and Walgreens dominate in many locations, focus on regions where these brands have a weaker presence to minimize competition while capitalizing on unmet demand.

## Query Log for the Project

### Main queries

Q1. Which brand should be the focus for expansion?

-- loc\_counts will have the list of Pharmacy brands which have more than 50 outlets

```
WITH loc_counts AS (  
  -- Step 1: Get a count of distinct locations for each brand  
  SELECT  
    brands.safegraph_brand_id AS brand_id,  
    brands.brand_name AS brand,  
    COUNT(DISTINCT(places.safegraph_place_id)) AS location_count,  
    -- Generate row number for each brand, ordered by the count of locations  
    ROW_NUMBER() OVER (ORDER BY COUNT(DISTINCT(places.safegraph_place_id)))  
    AS row_num  
  FROM  
    `querious-jac-grp3-final.grp3_safegraph.brands` AS brands  
  INNER JOIN  
    `querious-jac-grp3-final.grp3_safegraph.places` AS places  
  ON  
    brands.safegraph_brand_id = places.safegraph_brand_ids  
  WHERE  
    -- Filter for Pharmacy category NAICS code  
    brands.naics_code = 446110  
  GROUP BY  
    brands.safegraph_brand_id, brands.brand_name  
  HAVING  
    -- Only include brands with more than 50 outlets  
    COUNT(DISTINCT(places.safegraph_place_id)) > 50  
,  
  -- Median will give us the median value from the number of outlets  
  median AS (  
    -- Step 2: Calculate the median row number from loc_counts  
    SELECT  
      MAX(row_num) / 2 AS med  
    FROM
```

```

        loc_counts
    ),
    -- Top5_brand_outlet will filter brands with number of outlets higher than the median
    value
    Top5_brand_outlet AS (
    -- Step 3: Select brands with row numbers around the median (5 above the median)
        SELECT
            brand_id,
            brand,
            location_count
        FROM
            loc_counts
        WHERE
            -- Brands just above the median
            row_num > (SELECT med FROM median)
        AND
            -- Up to 5 brands above the median
            row_num <= (SELECT med FROM median) + 5
        ORDER BY
            row_num ASC
    ),
    -- Step 4: Get footfall(visits) information (number of visits divided by the number of
    locations)
    outlet_footfall AS (
        SELECT
            brands.safegraph_brand_id AS of_brand_id,
            brands.brand_name AS of_brand,
            COUNT(DISTINCT(places.safegraph_place_id)) AS location_count,
            COUNT(DISTINCT(visits.date_range_start)) AS no_of_months,
            -- Generate a row number based on the ratio of visits to location count
            ROW_NUMBER() OVER (ORDER BY SUM(visits.raw_visit_counts) /
            COUNT(DISTINCT(places.safegraph_place_id)) *
            COUNT(DISTINCT(visits.date_range_start)) AS row_num
        FROM
            `querious-jac-grp3-final.grp3_safegraph.brands` AS brands
        INNER JOIN
            `querious-jac-grp3-final.grp3_safegraph.places` AS places
        ON
            brands.safegraph_brand_id = places.safegraph_brand_ids
        INNER JOIN
            `querious-jac-grp3-final.grp3_safegraph.visits` AS visits
        ON
            visits.safegraph_place_id = places.safegraph_place_id
        WHERE
            -- Filter for Pharmacy category NAICS code
            brands.naics_code = 446110
        GROUP BY
            brands.safegraph_brand_id, brands.brand_name
    )

```

```

HAVING
    -- Only include brands with more than 50 outlets
    COUNT(DISTINCT(places.safegraph_place_id)) > 50
),
-- Median will give us the median value from the Footfall
median_of AS (
-- Step 5: Calculate the median row number based on footfall data
SELECT
    MAX(row_num) / 2 AS med
FROM
    outlet_footfall
),
-- Top5_brand_footfall will filter brands with number of footfalls higher than the
median value
Top5_brand_footfall AS (
-- Step 6: Select brands with footfall data around the median (5 above the median)
SELECT
    of_brand_id,
    of_brand,
    outlet_footfall.location_count
FROM
    outlet_footfall
WHERE
    -- Brands just above the footfall median
    row_num > (SELECT med FROM median_of)
AND
    -- Up to 5 brands above the median
    row_num <= (SELECT med FROM median_of) + 5
ORDER BY
    row_num ASC
)

-- Step 7: Join the results from the Top5_brand_outlet and the Top5_brand_footfall
SELECT *
FROM Top5_brand_outlet
INNER JOIN Top5_brand_footfall
ON Top5_brand_outlet.brand_id = Top5_brand_footfall.of_brand_id;

```

Answer :

Row	brand_id	brand	location_count
1	SG_BRAND_dcb657b5f186 fff7e0f5c09b222db75d	Benzer Pharmacy	120
2	SG_BRAND_c7706c9b14f2 ca266c2a309dba059031	Rite Aid	3292



3	SG_BRAND_2e6163ca2b39 e204	Safeway Pharmacy	428
---	-------------------------------	------------------	-----

Q2. What county and state will maximize profit while minimizing risk?

-- Query to calculate footfall of hospitals above the median

-- Step 1: Create a granular dataset of hospital visits and population

```
WITH hospitals_granular (
    SELECT
        counties.county AS county,
        counties.state AS state,
        places.city AS city,
        places.postal_code as postal_code,
        places.street_address as street_address,
        -- Total footfall of hospitals
        sum(visits.raw_visit_counts) as footfall_of_hospitals,
        -- Total population of the region
        sum(demographics.pop_total) as pop,
        -- Footfall per capita per month (to measure traffic relative to
        population)
        (sum(visits.raw_visit_counts)/count(visits.date_range_start)) /
        nullif(sum(demographics.pop_total), 0) as footfall_per_month_per_capita
    FROM
        `querious-jac-grp3-final.grp3_safegraph.brands` as brands
    INNER JOIN
        `querious-jac-grp3-final.grp3_safegraph.places` as places
    ON
        brands.safegraph_brand_id = places.safegraph_brand_ids
    INNER JOIN
        `querious-jac-grp3-final.grp3_safegraph.visits` as visits
    ON
        visits.safegraph_place_id = places.safegraph_place_id
    INNER JOIN
        `querious-jac-grp3-final.grp3_safegraph.cbg_demographics` as demographics
    ON
        visits.poi_cbg = demographics.cbg
    INNER JOIN
        `querious-jac-grp3-final.grp3_safegraph.cbg_fips` as counties
    ON
        concat(counties.state_fips, counties.county_fips) = left(visits.poi_cbg,
5)
    WHERE
        brands.naics_code = 622110 -- Filter for hospitals
    GROUP BY
        county,
        state,
        city,
```

```

        postal_code,
        street_address,
        brand_name,
        places.latitude,
        places.longitude -- Group by region and hospital information
    ),
    -- Step 2: Rank hospitals by footfall per capita
    hospitals as (
        SELECT
            hospitals_granular.county,
            hospitals_granular.state,
            -- Calculate average footfall per capita
            avg(hospitals_granular.footfall_per_month_per_capita),
            -- Rank counties based on footfall
            ROW_NUMBER() OVER (ORDER BY
                (avg(hospitals_granular.footfall_per_month_per_capita))) as row_num
        FROM
            hospitals_granular
        GROUP BY
            county,
            state
    ),

    -- Step 3: Calculate the median footfall
    median1 as (
        SELECT
            max(row_num) / 2 as med -- Calculate the median row number (for footfall
            ranking)
        FROM
            hospitals
    ),

    -- Step 4: Select counties above the median footfall
    tab1 as (
        SELECT *
        FROM
            hospitals
        WHERE
            -- Filter for rows above the median footfall
            row_num > (SELECT med FROM median1)
        AND
            -- Select the next 12 rows after the median
            row_num <= (SELECT med FROM median1) + 12
        ORDER BY
            row_num ASC),

    -- Query to calculate brand outlet counts below average

```

-- Step 5: Count the number of drug store outlets for Safeway Pharmacy

```
drugs_stores as (  
  SELECT  
    places.region as state,  
    counties.county as county,  
    -- Rank counties by outlets per capita  
    ROW_NUMBER() OVER (ORDER BY count(places.safegraph_place_id)/  
nullif(sum(demographics.pop_total), 0)) as row_num,  
    -- Count total number of outlets  
    count(places.safegraph_place_id) as count_brand_outlets,  
    sum(demographics.pop_total),  
    -- Calculate outlets per capita  
    count(places.safegraph_place_id)/ nullif(sum(demographics.pop_total),0)  
  as Outlets_per_Capita  
  FROM  
    `querious-jac-grp3-final.grp3_safegraph.brands` as brands  
  INNER JOIN  
    `querious-jac-grp3-final.grp3_safegraph.places` as places  
  ON  
    brands.safegraph_brand_id = places.safegraph_brand_ids  
  INNER JOIN  
    `querious-jac-grp3-final.grp3_safegraph.visits` as visits  
  ON  
    visits.safegraph_place_id= places.safegraph_place_id  
  INNER JOIN  
    `querious-jac-grp3-final.grp3_safegraph.cbg_demographics` as demographics  
  ON  
    visits.poi_cbg=demographics.cbg  
  INNER JOIN  
    `querious-jac-grp3-final.grp3_safegraph.cbg_fips` as counties  
  ON  
    concat(counties.state_fips, counties.county_fips) = left(visits.poi_cbg,  
5)  
  WHERE  
    brands.brand_name = "Safeway Pharmacy"  
  GROUP BY  
    state,  
    county  
  ORDER BY  
    count_brand_outlets desc  
) ,
```

-- Step 6: Calculate the median outlet count

```
median as (  
  SELECT  
    max(row_num) / 2 as med -- Calculate the median row number (for outlet  
count)  
  FROM
```

```

        drugs_stores
    ),

-- Step 7: Select counties with outlet counts below the median
tab2 as (
    SELECT *
    FROM
        drugs_stores
    WHERE
        row_num < (SELECT med FROM median) -- Filter for rows below the median
        outlet count
    AND
        row_num >= (SELECT med FROM median) - 15 -- Select the next 15 rows
        before the median
    ORDER BY
        row_num ASC)

-- Step 8: Join results from footfall and outlet count analysis
SELECT *
FROM
    tab1
INNER JOIN
    tab2
ON
    tab1.county = tab2.county
AND
    tab1.state = tab2.state;

```

Q3. What specific street locations near hospitals will attract the most foot traffic?

```

-- Step 1: Create a dataset for hospital footfall, population, and aged population in
King County, WA
WITH tab AS (
    SELECT
        places.city as city,
        places.postal_code as postal_code,
        places.street_address as street_address,
        brands.brand_name as brand_hospital,
        -- Calculate the average hospital footfall per month by dividing total
        visit counts by the number of time periods
        (sum(visits.raw_visit_counts)/count(visits.date_range_start)) as
        footfall_of_hospitals_per_month,
        -- Calculate the average population per month by dividing total population by the
        number of time periods
        (sum(demographics.pop_total)/count(visits.date_range_start)) as total_pop_per_month,
        -- Calculate the average aged population per month for male and female (55-84 age
        range) by dividing by the number of time periods
        (sum(

```

```

`pop_m_55-59` + `pop_m_60-61` + `pop_m_62-64` + `pop_m_65-66` + `pop_m_67-69` +
`pop_m_70-74` + `pop_m_75-79` + `pop_m_80-84` + -- Male population in the age ranges
`pop_f_55-59` + `pop_f_60-61` + `pop_f_62-64` + `pop_f_65-66` + `pop_f_67-69` +
`pop_f_70-74` + `pop_f_75-79` + `pop_f_80-84` -- Female population in the age ranges
) / count(visits.date_range_start)) as aged_pop_per_month -- Dividing the aged
population by the number of months
FROM `querious-jac-grp3-final.grp3_safegraph.brands` as brands -- SafeGraph brands
dataset
INNER JOIN `querious-jac-grp3-final.grp3_safegraph.places` as places -- Join places
dataset to get location information
ON brands.safegraph_brand_id = places.safegraph_brand_ids
INNER JOIN `querious-jac-grp3-final.grp3_safegraph.visits` as visits -- Join visit
data to get footfall counts
ON visits.safegraph_place_id = places.safegraph_place_id
INNER JOIN `querious-jac-grp3-final.grp3_safegraph.cbg_demographics` as demographics -
- Join demographic data
ON visits.poi_cbg = demographics.cbg
INNER JOIN `querious-jac-grp3-final.grp3_safegraph.cbg_fips` as counties -- Join
county data
ON concat(counties.state_fips, counties.county_fips) = left(visits.poi_cbg, 5)
WHERE brands.naics_code = 622110 -- Filter for hospitals (NAICS code 622110)
AND counties.county = 'King County' -- Filter for King County
AND counties.state = 'WA' -- Filter for Washington state
GROUP BY city, postal_code, street_address, brand_name -- Group by city, postal code,
street address, and brand
ORDER BY footfall_of_hospitals_per_month DESC -- Order results by descending footfall
of hospitals per month
)

```

```

-- Step 2: Calculate the ratio of aged population to total population for each
hospital and select all fields from tab
SELECT *,
tab.aged_pop_per_month / tab.total_pop_per_month as aged_pop_ratio -- Calculate the
ratio of aged population to total population
FROM tab;

```

## -----Research Queries-----

```

-- Query 1: Fetch the number of brands in the dataset that match the NAICS code 446110
(representing pharmacies or drug stores)
-- The result should give 59 brands.

```

```

SELECT
    brands.safegraph_brand_id,
    brands.brand_name
FROM
    `querious-jac-grp3-final.grp3_safegraph.brands` AS brands

```

WHERE

brands.naics\_code = 446110;

-- Query 2: Count distinct place IDs in the dataset to check if all the place IDs are unique.

SELECT

DISTINCT(safegraph\_place\_id)

FROM

`querious-jac-grp3-final.grp3\_safegraph.places`;

-- Query 3: Calculate the ratio of each brand's location count to the sum of all location counts.

SELECT

safegraph\_brand\_id,

brand\_name,

location\_count,

location\_count / (SELECT SUM(location\_count) FROM loc\_counts) AS

loc\_count\_ratio

FROM

loc\_counts;

-- Query 4: Calculate the average number of outlets per brand.

-- For each brand, count the distinct place IDs (outlets), then calculate the average across all brands.

SELECT

AVG(loc\_count)

FROM (

SELECT

COUNT(DISTINCT p.safegraph\_place\_id) AS loc\_count

FROM

`querious-jac-grp3-final.grp3\_safegraph.brands` AS b

INNER JOIN

`querious-jac-grp3-final.grp3\_safegraph.places` AS p

ON

b.safegraph\_brand\_id = p.safegraph\_brand\_ids

WHERE

b.naics\_code = 446110

GROUP BY

b.safegraph\_brand\_id

);

-- Query 5: Find brands with a number of outlets greater than or equal to the average number of outlets per brand.

-- Brands are filtered based on this condition and then sorted by the number of outlets in descending order.

SELECT

```

        brands.safegraph_brand_id,
        brands.brand_name,
        COUNT(DISTINCT places.safegraph_place_id) AS locations
FROM
    `querious-jac-grp3-final.grp3_safegraph.brands` AS brands
INNER JOIN
    `querious-jac-grp3-final.grp3_safegraph.places` AS places
ON
    brands.safegraph_brand_id = places.safegraph_brand_ids
WHERE
    brands.naics_code = 446110
GROUP BY
    brands.safegraph_brand_id, brands.brand_name
HAVING
    COUNT(DISTINCT places.safegraph_place_id) >= (
        SELECT AVG(loc_count)
        FROM (
            SELECT
                COUNT(DISTINCT p.safegraph_place_id) AS loc_count
            FROM
                `querious-jac-grp3-final.grp3_safegraph.brands` AS b
            INNER JOIN
                `querious-jac-grp3-final.grp3_safegraph.places` AS p
            ON
                b.safegraph_brand_id = p.safegraph_brand_ids
            WHERE
                b.naics_code = 446110
            GROUP BY
                b.safegraph_brand_id
        )
    )
)
ORDER BY
    locations DESC;

```

-- Query 6: Retrieve place details along with brand and geographic information for wholesalers (NAICS code 424210).

```

SELECT
    counties.county, -- The county name from the cbg_fips table.
    places.region,
    places.city AS city,
    places.postal_code AS postal_code,
    places.street_address AS street_address,
    brands.brand_name AS brand_wholesaler, -- The brand/wholesaler name from the
                                           brands table.
    places.latitude,
    places.longitude
FROM
    `querious-jac-grp3-final.grp3_safegraph.brands` AS brands

```

```

INNER JOIN
    `querious-jac-grp3-final.grp3_safegraph.places` AS places
ON
    brands.safegraph_brand_id = places.safegraph_brand_ids
INNER JOIN
    `querious-jac-grp3-final.grp3_safegraph.visits` AS visits
ON
    visits.safegraph_place_id = places.safegraph_place_id
-- Join the cbg_demographics table to link demographic data based on the census block
-- group (cbg) of the place of interest (POI).
INNER JOIN
    `querious-jac-grp3-final.grp3_safegraph.cbg_demographics` AS demographics
ON
    visits.poi_cbg = demographics.cbg
-- Join the cbg_fips table to obtain county and state FIPS codes.
-- The FIPS code is combined to form a complete location identifier.
INNER JOIN
    `querious-jac-grp3-final.grp3_safegraph.cbg_fips` AS counties
ON
    CONCAT(counties.state_fips, counties.county_fips) = LEFT(visits.poi_cbg, 5)
-- Filter the results to only include places where the brand's NAICS code is 424210
-- (wholesalers).
WHERE
    brands.naics_code = 424210
-- Optional filter (commented out): you could restrict results to a specific county
-- like 'Prince William County'.
-- AND counties.county = 'Prince William County'
-- Filter the results to include only counties from certain states (NC, KY, TN, WV,
-- PA, OH, IN).
AND
    counties.state IN ('NC', 'KY', 'TN', 'WV', 'PA', 'OH', 'IN');

-- Query 7 : Selects the region and count of places based on the SafeGraph place ID
SELECT places.region, COUNT(places.safegraph_place_id)

-- Specifies the source table `places` from the dataset `querious-jac-grp3-
-- final.grp3_safegraph.places`
FROM `querious-jac-grp3-final.grp3_safegraph.places` AS places

-- Filters the records to only include places with NAICS code 424210 (Pharmaceutical
-- and Medical Distributors)
WHERE places.naics_code = 424210

-- Groups the results by region to get the count of places per region
GROUP BY places.region;

```



