**Final Report: Analysis of Pharmacy Expansion Opportunities**

Shih Min Lin, Tzu-Yun Huang, Hung-Chen Hsu

Group 5

**Initial Research Questions and Industry of Interest**

Our team specializes in the healthcare sector within the United States, with a particular focus on pharmacies. As a real estate investment firm, our goal is to pinpoint areas with high demand for pharmacy services, characterized by a significant elderly population and a shortage of existing pharmacies. This analysis will guide strategic decisions on where to open new stores to maximize profitability. We are utilizing anonymized cell phone traffic data from SafeGraph to identify consumer movement patterns and uncover underserved regions.

**Specific Questions to Answer from the Data**

1. **Identify Popular Pharmacies: Which pharmacy brand in the healthcare sector has experienced the highest traffic over the past five years?**

* **Method**: Our query aims to find the top 5 pharmacies in the healthcare industry with the most traffic in the past 5 years. It retrieves the pharmacy names and the total number of visits, filtering companies in the "Health and Personal Care Stores" category. The query joins the 'places' and 'visits' tables based on SafeGraph place IDs, then groups and orders by total visit counts in the last 5 years.
* **Findings**: Our analysis shows that Walgreens leads in traffic among pharmacy brands, solidifying its position as a dominant market player. Studying its traffic patterns offers valuable insights and serves as a benchmark for future expansion strategies.

1. **Identify High-Demand Areas: Which U.S. regions with a substantial elderly population (aged 60 to 74) and a high median income have the highest elderly population per pharmacy ratio?**

* **Method**: Firstly, we created a temporary table (potential\_county\_table) to aggregate the elderly population (aged 60-74) and income distribution across counties. Then, by filtered counties by those having both a higher-than-average elderly population and higher-than-average income, we calculated the elderly population per pharmacy ratio for each county, ranking counties to determine which had the highest demand for pharmacies.
* **Findings**: Cook County, IL was identified as having the highest pharmacy capacity relative to its elderly population. On average, each pharmacy in Cook County serves a significantly larger elderly population than in other counties. This suggests a high demand for more pharmacies, making Cook County an ideal location for expansion.

1. **Determine Optimal Business Hours: Based on the findings from questions 1 & 2, what is the ideal business day for Walgreens in Cook County, IL?**

* **Method**: Firstly, we extracted daily popularity metrics from the popularity\_by\_day field for each day of the week using Json extract. Finally, we calculated the average visitor counts for each day by dividing the total daily popularity by the total raw visit counts, allowing us to determine which days experience the highest foot traffic.
* **Findings**: The results indicated that Friday has the highest traffic, while Sunday experiences the lowest. This suggests that we should consider closing on Sundays to reduce operational costs. Understanding this trend leads us to explore the age distribution in Cook County to further investigate why Sunday sees such low traffic.

1. **Determine the reason that Sunday has the fewest traffic: What is the age distribution in Cook County, and how does it compare to the United States as a whole?**

* **Method**: Firstly, we queried the Safegraph demographics data for Cook County to calculate the population distribution across various age groups, specifically focusing on children (age under 14), teenagers(15-24), middle-aged adults(25-44), older adults(45-64), and the elderly(65+). We aggregated the data to find the proportion of each age group relative to the total population in Cook County. Secondly, we executed a similar query for the entire United States to allow for a comparative analysis of age distributions.
* **Findings**: The results revealed that Cook County has a notably higher proportion of middle-aged residents (ages 25-64) compared to the national average. This demographic insight suggests that the population may be less inclined to engage in leisure activities on Sundays, as many are preparing for the work week ahead. With the anticipation of Monday responsibilities, individuals may prefer to spend their Sundays relaxing or completing necessary tasks rather than socializing or shopping.

**Conclusion**

Our analysis identifies Cook County, IL, as a prime location for pharmacy expansion, driven by its high elderly population relative to the number of pharmacies available. With Walgreens leading in foot traffic among pharmacy brands, the demand for additional pharmacy services in the region is evident.

Furthermore, our demographic insights reveal a higher proportion of middle-aged residents, which correlates with lower traffic on Sundays. As a result, we recommend closing or reducing operational hours on Sundays to minimize costs. This strategic approach enables us to optimize resources while positioning our investments effectively to meet community healthcare needs and maximize profitability.

**Reference**

1. Reuter, D. (n.d.). *Meet the typical CVS shopper: A white, gen X, college-educated city dweller earning a high income*. Business Insider. https://www.businessinsider.com/typical-cvs-shopper-demographic-urban-genx-earning-high-income-2021-9
2. Reuter, D. (n.d.-b). *Meet the typical Walgreens Shopper, a white college-educated suburban Baby Boomer earning $80,000*. Business Insider. https://www.businessinsider.com/typical-walgreens-shopper-demographic-suburban-boomer-earning-80k-income-2021-9#:~:text=Numerator%20found%20that%20Walgreens’%20typical,with%20similar%20education%20and%20income.
3. Institute of Medicine (US) Committee to Design a Strategy for Quality Review and Assurance in Medicare. (1990, January 1). *The elderly population*. Medicare: A Strategy for Quality Assurance: Volume 1. https://www.ncbi.nlm.nih.gov/books/NBK235450/#:~:text=Traditionally%2C%20the%20%E2%80%9Celderly%E2%80%9D%20are,persons%20age%2065%20and%20older
4. *Walgreens campaign*. Walgreens Campaign. (n.d.-a). https://walgreenscampaign.wordpress.com/#:~:text=The%20target%20audience%20for%20Walgreens,22%20through%20the%20elderly%2070%2B
5. *Demographic segmentation*. PickFu. (n.d.-a). https://www.pickfu.com/demographic-segmentation#:~:text=Age%20segmentation%20means%20focusing%20on,64%2C%20and%2065%20and%20older

**Query Log**

**A. Brand-Specific Questions (Identifying Potential Brands):**

(Calculate the total visit counts for each brand in a specific area, divided by the total population of that area, to determine the most popular brands in the market based on traffic.)

* **Which pharmacy brand in the healthcare sector has experienced the highest traffic over the past five years?**

-- Find a pharmacy brand in the healthcare industry that has the most traffic in the past 5 years.

SELECT p.location\_name AS pharmacy\_name,

SUM(v.raw\_visit\_counts) AS total\_visits

FROM `elemental-leaf-436616-q5.safegraph.places` AS p

JOIN `elemental-leaf-436616-q5.safegraph.visits` AS v

ON p.safegraph\_place\_id = v.safegraph\_place\_id

WHERE p.top\_category = 'Health and Personal Care Stores'

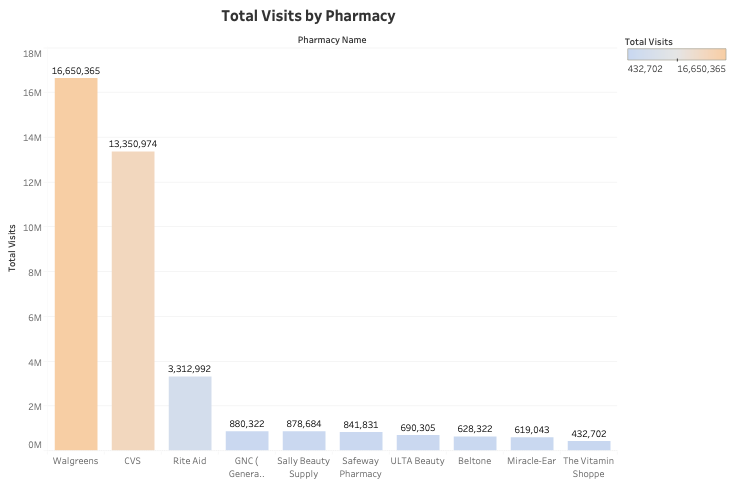
AND v.date\_range\_start >= TIMESTAMP(DATE\_SUB(CURRENT\_DATE(), INTERVAL 5 YEAR)) -- Convert to TIMESTAMP, and -- Filter for visits within the last 5 years

GROUP BY p.location\_name

ORDER BY total\_visits DESC

LIMIT 5;

**ANS: Walgreens**

****

**B. Geographic-Specific Questions(Identify High-Demand Areas)**

* **Which U.S. regions with a substantial elderly population (aged 60 to 74) and a high median income have the highest elderly population per pharmacy ratio?**

-- Find a county that has an elderly population higher than average and a high-income population higher than average, then calculate the pharmacy per capita ratio to identify the county with the highest pharmacy capacity.

CREATE TEMPORARY TABLE potential\_county\_table AS

SELECT f.county,f.state\_fips,f.county\_fips,f.state,

SUM(`pop\_m\_60-61` + `pop\_m\_62-64` + `pop\_m\_65-66` + `pop\_m\_67-69`+`pop\_m\_70-74`+

`pop\_f\_60-61` + `pop\_f\_62-64` + `pop\_f\_65-66`+`pop\_f\_67-69`+`pop\_f\_70-74`) AS older\_population,

SUM(`inc\_75-100` + `inc\_100-125` + `inc\_125-150`) AS higher\_income\_population

FROM `elemental-leaf-436616-q5.safegraph.cbg\_demographics` AS d

JOIN `elemental-leaf-436616-q5.safegraph.cbg\_fips` AS f

ON SUBSTRING(d.cbg, 1, 2) = f.state\_fips

AND SUBSTRING(d.cbg, 3, 3) = f.county\_fips

GROUP BY f.county,f.state\_fips,f.county\_fips,f.state

HAVING older\_population > (

SELECT AVG(older\_population)

FROM (

SELECT SUM(`pop\_m\_60-61` + `pop\_m\_62-64` + `pop\_m\_65-66` + `pop\_m\_67-69`+`pop\_m\_70-74`+

`pop\_f\_60-61` + `pop\_f\_62-64` + `pop\_f\_65-66`+`pop\_f\_67-69`+`pop\_f\_70-74`) AS older\_population,

FROM `elemental-leaf-436616-q5.safegraph.cbg\_demographics` AS d

JOIN `elemental-leaf-436616-q5.safegraph.cbg\_fips` AS f

ON SUBSTRING(d.cbg, 1, 2) = f.state\_fips

AND SUBSTRING(d.cbg, 3, 3) = f.county\_fips

GROUP BY f.county

) AS subquery)

AND higher\_income\_population> (

SELECT AVG(income)

FROM (

SELECT SUM(`inc\_75-100` + `inc\_100-125` + `inc\_125-150`) AS income,

FROM `elemental-leaf-436616-q5.safegraph.cbg\_demographics` AS d

JOIN `elemental-leaf-436616-q5.safegraph.cbg\_fips` AS f

ON SUBSTRING(d.cbg, 1, 2) = f.state\_fips

AND SUBSTRING(d.cbg, 3, 3) = f.county\_fips

GROUP BY f.county

) AS sub);

SELECT county, state,SUM (older\_population)/COUNT(v.safegraph\_place\_id) pharmacy\_percapita

FROM potential\_county\_table pc

JOIN `elemental-leaf-436616-q5.safegraph.visits` v

ON SUBSTRING(v.poi\_cbg, 1, 2) = pc.state\_fips

AND SUBSTRING(v.poi\_cbg, 3, 3) = pc.county\_fips

JOIN `elemental-leaf-436616-q5.safegraph.places` p

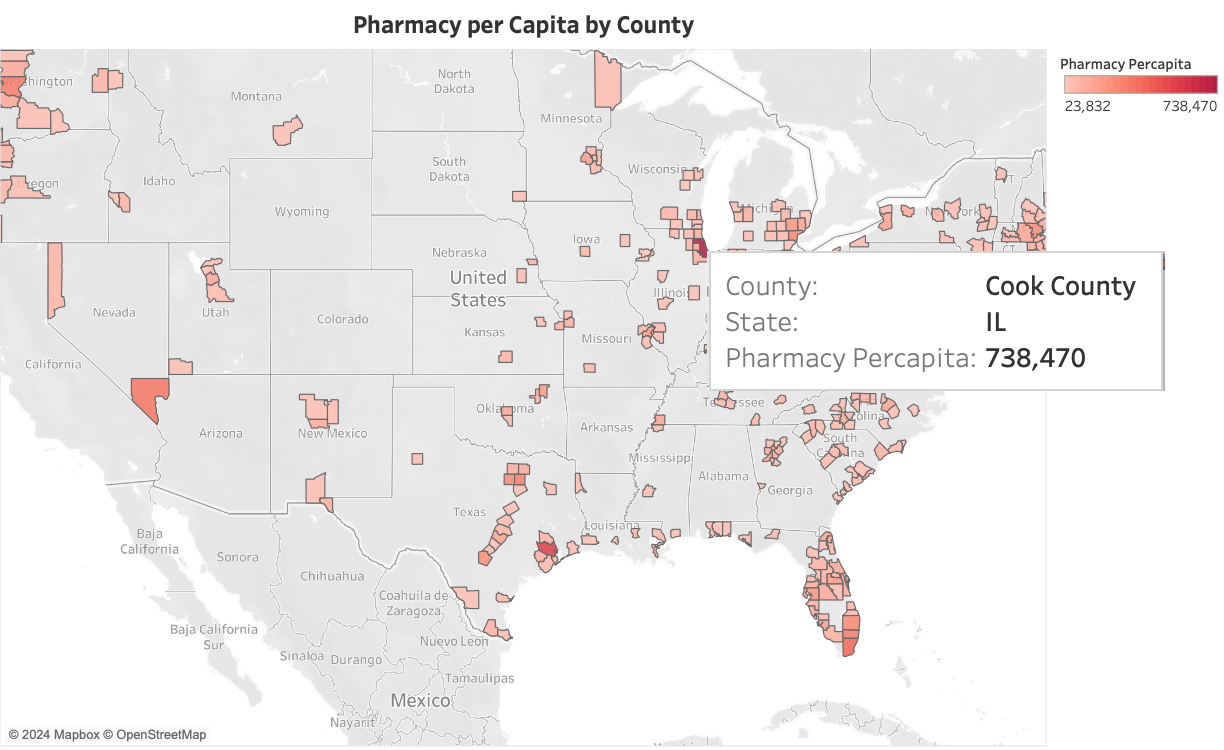
ON v.safegraph\_place\_id = p.safegraph\_place\_id

WHERE p.top\_category = 'Health and Personal Care Stores'

GROUP BY county,state

ORDER BY pharmacy\_percapita DESC;

**ANS: Cook County, IL**

****

**C. Business Hours (Determine Optimal Business Hours):**

* **Based on the answers to the first and second questions, what are the ideal business day and hours for Walgreens in Cook County?**

-- Finding the busiest day and hour for Walgreens as the ideal business day and hour.

SELECT

SUM(CAST(JSON\_EXTRACT\_SCALAR(popularity\_by\_day, '$.Monday') AS INT64))/SUM(raw\_visit\_counts) AS total\_monday,

SUM(CAST(JSON\_EXTRACT\_SCALAR(popularity\_by\_day, '$.Tuesday') AS INT64))/SUM(raw\_visit\_counts) AS total\_tuesday,

SUM(CAST(JSON\_EXTRACT\_SCALAR(popularity\_by\_day, '$.Wednesday') AS INT64))/SUM(raw\_visit\_counts) AS total\_wednesday,

SUM(CAST(JSON\_EXTRACT\_SCALAR(popularity\_by\_day, '$.Thursday') AS INT64))/SUM(raw\_visit\_counts) AS total\_thursday,

SUM(CAST(JSON\_EXTRACT\_SCALAR(popularity\_by\_day, '$.Friday') AS INT64))/SUM(raw\_visit\_counts) AS total\_friday,

SUM(CAST(JSON\_EXTRACT\_SCALAR(popularity\_by\_day, '$.Saturday') AS INT64))/SUM(raw\_visit\_counts) AS total\_saturday,

SUM(CAST(JSON\_EXTRACT\_SCALAR(popularity\_by\_day,'$.Sunday') AS INT64))/SUM(raw\_visit\_counts) AS total\_sunday

FROM `elemental-leaf-436616-q5.safegraph.visits` AS v

JOIN `elemental-leaf-436616-q5.safegraph.places` AS p

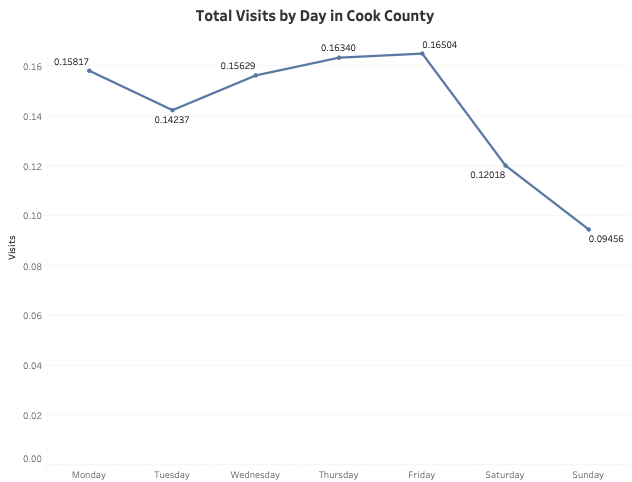
ON v.safegraph\_place\_id = p.safegraph\_place\_id

JOIN `elemental-leaf-436616-q5.safegraph.cbg\_fips` AS f

ON v.region = f.state

WHERE p.city LIKE '%Cook%'

AND state = 'IL';



* **What is the age distribution in Cook County?**

SELECT

f.county,

f.state,

SUM(

`pop\_m\_lt5` + `pop\_m\_10-14` + `pop\_f\_5-9` + `pop\_f\_lt5` +

`pop\_f\_10-14` + `pop\_m\_5-9`

) / SUM(pop\_total) AS kids\_pop,

SUM(

`pop\_m\_15-17` + `pop\_m\_18-19` + `pop\_m\_20` + `pop\_m\_21` + `pop\_m\_22-24` +

`pop\_f\_15-17` + `pop\_f\_18-19` + `pop\_f\_20` + `pop\_f\_21` + `pop\_f\_22-24`

) / SUM(pop\_total) AS teenager\_pop,

SUM(

`pop\_m\_25-29` + `pop\_m\_30-34` + `pop\_m\_35-39` + `pop\_m\_40-44` +

`pop\_f\_25-29` + `pop\_f\_30-34` + `pop\_f\_35-39` + `pop\_f\_40-44`

) / SUM(pop\_total) AS middleage\_pop,

SUM(

`pop\_f\_45-49` + `pop\_f\_50-54` + `pop\_f\_55-59` +

`pop\_m\_45-49` + `pop\_m\_50-54` + `pop\_m\_55-59` +

`pop\_f\_60-61` + `pop\_f\_62-64` + `pop\_m\_60-61` + `pop\_m\_62-64`

) / SUM(pop\_total) AS olderage\_pop,

SUM(

`pop\_m\_65-66` + `pop\_m\_67-69` + `pop\_m\_70-74` +

`pop\_f\_65-66` + `pop\_f\_67-69` + `pop\_f\_70-74`

) / SUM(pop\_total) AS elderly\_pop

FROM

`elemental-leaf-436616-q5.safegraph.cbg\_demographics` AS d

JOIN

`elemental-leaf-436616-q5.safegraph.cbg\_fips` AS f

ON

SUBSTRING(d.cbg, 1, 2) = f.state\_fips

AND SUBSTRING(d.cbg, 3, 3) = f.county\_fips

WHERE

state = 'IL'

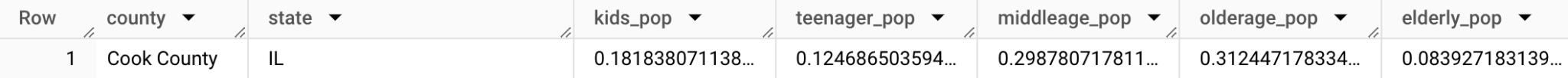
AND county LIKE '%Cook%'

GROUP BY

f.county,

F.state;

**ANS: The majority of the population falls within the 25-64 age group.**

****

* **In order to compare the age distribution in Cook County with that of the United States as a whole, what is the age distribution for the entire country?**

SELECT

SUM(

`pop\_m\_lt5` + `pop\_m\_10-14` + `pop\_f\_5-9` + `pop\_f\_lt5` +

`pop\_f\_10-14` + `pop\_m\_5-9`

) / SUM(pop\_total) AS kids\_pop,

SUM(

`pop\_m\_15-17` + `pop\_m\_18-19` + `pop\_m\_20` + `pop\_m\_21` +

`pop\_m\_22-24` + `pop\_f\_15-17` + `pop\_f\_18-19` + `pop\_f\_20` +

`pop\_f\_21` + `pop\_f\_22-24`

) / SUM(pop\_total) AS teenager\_pop,

SUM(

`pop\_m\_25-29` + `pop\_m\_30-34` + `pop\_m\_35-39` + `pop\_m\_40-44` +

`pop\_f\_25-29` + `pop\_f\_30-34` + `pop\_f\_35-39` + `pop\_f\_40-44`

) / SUM(pop\_total) AS middleage\_pop,

SUM(

`pop\_f\_45-49` + `pop\_f\_50-54` + `pop\_f\_55-59` +

`pop\_m\_45-49` + `pop\_m\_50-54` + `pop\_m\_55-59` +

`pop\_f\_60-61` + `pop\_f\_62-64` + `pop\_m\_60-61` + `pop\_m\_62-64`

) / SUM(pop\_total) AS olderage\_pop,

SUM(

`pop\_m\_65-66` + `pop\_m\_67-69` + `pop\_m\_70-74` +

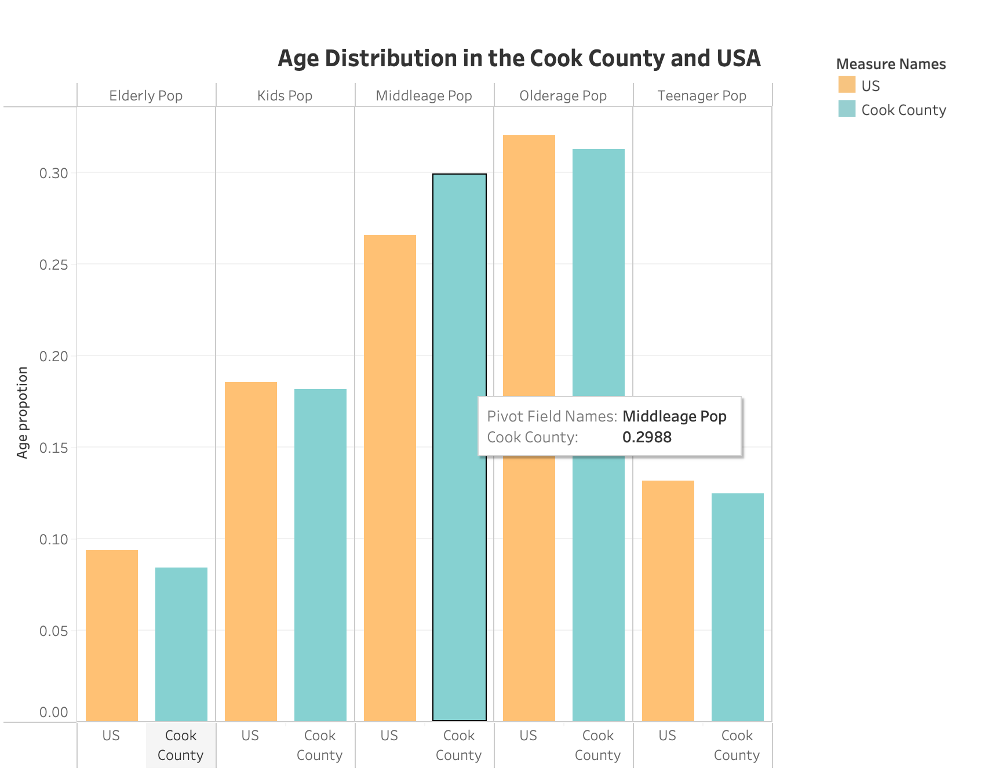
`pop\_f\_65-66` + `pop\_f\_67-69` + `pop\_f\_70-74`

) / SUM(pop\_total) AS elderly\_pop

FROM

`elemental-leaf-436616-q5.safegraph.cbg\_demographics`;

****

****

1. **Summing Population by Age Group:**
2. The query calculates the population proportions for different age groups across the entire U.S.:
   1. Kids: Ages less than 5 to 14 years old.
   2. Teenagers: Ages 15 to 24 years old.
   3. Middle Age: Ages 25 to 44 years old.
   4. Older Age: Ages 45 to 64 years old.
   5. Elderly: Ages 65 to 74 years old.
   6. Each group’s population is summed, and then divided by the total population (pop\_total) to determine its proportion of the overall U.S. population.
3. **No Location Filter:**
4. Unlike the Cook County query, this query does not filter by any specific location, allowing it to return results for the entire country.