Group 7 Final Project

Team members: Chia-Wei Lin, Prashast Vaish, Soutik Banerjee, Sumant Chaurasiya

Executive Summary

This report explores the potential for setting up a chain of limited-service restaurants in Texas, leveraging Safegraph's traffic data to gain insights that could improve our decision-making. The primary goal is to identify optimal locations based on key factors such as income levels, age demographics, and the existing presence of limited-service restaurants. Our analysis also examines night-time eating patterns and dwell-time to uncover under-served areas that may offer a competitive edge.

By understanding consumer behavior through these metrics, we can ensure that success of the planned chain. Such success hinges on understanding supply and demand, and these specific insights enable us to pinpoint areas where demand is highest, ensuring that new restaurants are strategically located for maximum impact.

Premise

Our client wants to invest in opening a chain of limited-service restaurants, seeing the growing popularity of brands like Chipotle, Subway, Wendy's, Chick-fil-A in the US. After an initial discussion with consultants, the following decisions were made:

- The pilot project will be in Texas, as it has a fast-growing population and is the largest state in the contiguous United States

- The limited-service chain will be targeted at customers in the age bracket of 10 to 35 years and income bracket of <=$75,000 per year

- The unique selling point will be late-night service, when some restaurants may be closed but there will be customers willing to have a late-night snack. Advertisement for the new chain (not in the scope of this report) will be aimed at reinforcing this factor in the minds of prospective customers

- The outlets of this chain will expand, starting from one city to the rest of the county, and then to the rest of Texas before going national

With this premise in mind, we set out to gather insights from the Safegraph data available to us.

Research Questions and Methodology

Question 1:

Which counties in Texas are in the top 33rd percentile for both proportion of target age group and proportion of target income group? (Target age group: 10 – 35, target income group: <= 75k)

After running the query, we found 23 counties in Texas that are in the top 33% for both proportion of target age group and target income group.

Explanation of the query:

The query starts with cte_county_population, calculating the population of people aged 10-35 (population_10_35), total population, income below $75K (income_less_75), and total earners. It filters counties with over 10K total population and 2K earners.

- "cte_ranking" ranks counties by population proportion (population_10_35_perc) and income proportion (income_less_75_perc) for Texas counties using DENSE_RANK().
- In "ranked_data", percentiles for both proportions are calculated with PERCENT_RANK(), resulting in "population_10_35_percentile" and "income_less_75_percentile".
- The final selection filters counties in the top 33% for both population and income percentiles.
- In conclusion, this query identifies counties which meet the demographic criteria set by the client.

Question 2:

Out of the counties identified in Q1, which counties are under-served? The definition of 'under-served' being that they must be in the bottom 33rd percentile among all counties in Texas for restaurants per capita. Among the under-served counties, we will pick the one with the highest population in the target age group.

We identified that Hidalgo County in Texas is under-served, with the highest target population.

Explanation of the query:

- "cte_county_population" calculates the population aged 10-35 (population_10_35), total population, and total earning population
- "cte_lsr_restaurant" counts the number of limited-service restaurants (lsr_restaurant_count) per county
- "cte_pop_lsr" computes the number of limited-service restaurants per capita (lsr_restaurant_per_capita)
- "cte_ranking" ranks counties by LSR density using DENSE_RANK() and calculates their percentile

Finally, it intersects this data with previously identified counties (population_filter) and filters those with LSR per capita in the bottom 33%, selecting the county with the highest 10-35 population.

Question 3:

Among all the cities in the identified county (Q2), which city ranks at the top in terms of night-time popularity and has a minimum population of 10,000?

We identified the city of Pharr as the ideal city, with a population over 10,000 and a good culture of people visiting limited-service restaurants between 9 PM and 12 AM.

Explanation of the query:

- "cte_city" retrieves all distinct city and Census Block Group (CBG) combinations from visit data
- "cte_cbg_population" joins demographic data to get the total population for each CBG
- "cte_city_population" aggregates CBG data to calculate the total population for each city, filtering out cities with less than 10,000 people
- Main Query: Calculates the average late-night popularity (avg_late_night_popularity) for LSRs by summing visit popularity for 9 PM to midnight (hour_21_to_0_popularity) from SafeGraph data

Finally, it selects the city with the highest late-night restaurant popularity.

Question 4:

What is the average rent, average property price, average median dwell time (LSRs only) and number of visitors (LSRs only) for each CBG in that city? These are some of the key business metrics which will allow our client to choose a CBG where opening an outlet will be most profitable.

This query produces a comprehensive dataset that combines restaurants activity metrics to property value and rental information at granular level.

Explanation of the query:

- "cte_cbg_list" calculates the number of total restaurants (total_restaurant) and LSRs (total_lsr_restaurant) in each CBG, along with average visits (avg_visits_per_restaurant, avg_visits_per_lsr_restaurant) and dwell time (average_dwell_time_per_restaurant, average_dwell_time_per_LSR_restaurant)

- "cte_property_prices" computes the average property value (average_property_value) and rental prices (average_rental_rent) in each CBG based on demographic data
- Main Query: Joins the restaurant data with property prices, selecting key metrics such as total restaurants, visits, dwell time, property value, and rental prices for each CBG, rounding values for clarity

Conclusion and Recommendations

The analysis provides a data-driven approach for expanding limited-service restaurants in Texas, targeting the 10-35 age group with incomes up to $75,000. Key findings and recommendations include:

1. Target Market: Focus initial expansion efforts on Hidalgo County, which shows high potential with a large target demographic and underserved market

2. City Selection: Prioritize Pharr for the first restaurant location, given its strong night-time dining scene and suitable population size

3. Site Selection: Utilize the provided CBG-level metrics to identify specific neighborhoods in Pharr, balancing high visitor traffic with favorable rent and property prices

4. Competitive Advantage: Leverage the late-night service unique selling point, as the analysis confirms demand for night-time dining options

5. Scalable Approach: Apply this analytical framework to identify future expansion opportunities across Texas and eventually nationwide

By following these data-informed recommendations, the client can strategically position their limited-service restaurant chain for success, starting with a strong foothold in a high-potential Texas market.