Drew Price, Caden Silberlicht

Milestone 2

4/4/24

Popular and Successful Businesses

When looking at the metrics for describing a popular and successful business, there are many factors to account for, including amount of reviews, average stars for the business, and the number of people visiting the business. We will classify popular and successful businesses as follows:

Popular - The amount of check-ins exceeds the average for that zipcode

Successful - The star rating is above 4 stars and the amount of reviews exceeds the average for the categories that describe the business.

We have determined that for a place to be popular, its total **numCheckins** should exceed the average for that zip code. Popularity is relative to the general area, so a zip code with a small population will not be made irrelevant because of a zip code like 90210 (an LA zip code). For a business to be considered successful, the rating must exceed 4 stars and the business must have more reviews in its designated categories than the average business in those categories. A niche business should get recognition for success in a category that has few entities.

For our next step, we created queries that would allow us to find what businesses meet these requirements.The following query returned us popular businesses by our metrics:

```
SELECT *
FROM Business
JOIN (
        SELECT AVG(numCheckins) AS avgCheckins, postal_code
        FROM Business
        GROUP BY postal_code) AS Average
ON Business.postal_code = Average.postal_code
WHERE numCheckins > avgCheckins
ORDER BY Business.postal_code
```

This query starts by finding the average number of **checkins** and the **postal_code** from the **Business** table, adding that result to the **Business** table in a new column called **Average**. The query then selects all values in the **Business** table where the **postal_code** from the business equals the **postal_code** of the average and where the **numCheckins** exceeds the **average number of checkins**. The results of this query are all of the businesses that are considered “popular” in their **postal_code** by our standards. From the results of this query, we can conclude that most “popular” businesses don’t necessarily meet the requirements for a “successful” business. Because of this, the need for a query to calculate successful businesses

is essential to decipher between the two. We used the following query to return successful businesses by our metrics:

```
SELECT business.business_id, name, city, reviewrating, review_count,
AVG(averageReviews) AS avgReviewCount
FROM Business
JOIN Categories AS Cat on Business.business_id = Cat.business_id
JOIN(
        SELECT AVG(review_count) AS averageReviews, category_name
        FROM Business
        JOIN Categories on Business.business_id = Categories.business_id
        GROUP BY category_name) AS Average
ON Average.category_name = Cat.category_name
GROUP BY Business.business_id, business.name
HAVING review_count > AVG(averageReviews) AND reviewrating >= 4.0
ORDER BY reviewrating DESC
```

This query works by creating a temporary table that joins the **Business** table and **Categories** table on the **business_id**, then finding the average **review_count** and grouping by the **category_name**. We join this table on **category_name** with another table which is made by joining the **Business** and **Categories** tables again. From here, we found the average of the average **review_count** for the categories making up each business while grouping by **business_id**. Finally, we filtered so only businesses with a **review_count** greater than the average of its categories, and a **reviewrating** of 4 stars and above, would appear. It is very evident, looking at the results of the query, that successful businesses are most typically found in major cities. This makes sense when we consider that successful businesses are partly looking for above average **review_count** without making it relative to location like we did in the query to determine popular businesses. Successful businesses are typically found in well populated areas, so this should not be concerning.

The difference between popular and successful businesses is subjective, however, our explanation provides a good foundation for how we will determine these attributes in the future. Likewise, the queries can provide a more technical explanation for our description of popular and successful.