Joshua Frey

3/30/2023

CPTS 451

Project Milestone 2 Metrics

In this project, we are creating a data search application based on Yelp.com’s business and review data. One of the major use cases or features of this project is to allow users to be able to see what businesses are classified as “popular” and “successful”. We want to allow users to select a business category and zip code which will then filter which businesses in that zip code and that business category is considered “successful” or “popular”. I will use some of the tables in my databases and it’s attributes to query an efficient set of metrics to determine what is a “popular” or “successful” business.

To determine whether a business is successful, I will use the example given in the milestone 2 description. I think writing a query over the check-in data and finding the average number of check-ins based on the business category and zip code is a very good metric. We can use the average number of check-ins to then compare the number of check-ins each business has. If a business has more check-ins, then the average than that business is deemed “successful”. To determine the total amount of check-ins a business has you would run this query over the check-in table “select business\_id, sum(count) from checkin group by business\_id”. This query will be used as a subquery to compare against the average number of check-ins. The table will show the business name and the number of check-ins.

To determine whether a business is popular, I will use the business table and use the stars rating and review count as the key metrics. I will do a similar query as the query to determine whether a business is successful, by finding the average stars rating and average review count based on the business category and zip code pair. I will then compare each business’s star rating and review count in that business category and zip code pair to determine whether it is popular. If the business star rating and review count is more than the average than it is deemed popular. I will implement this by writing a subquery that finds the star rating and review count of a business in that business category and zip code pair and then compare it with the outer query that gets the average star rating and review count in that business category and zip code pair. This query will return a table with the business name, stars, review rating, and review count to be displayed in the popular business section.