Flavio Alvarez Penate

Ananth Jillepalli

CptS 451

July 6th, 2023

Milestone 2:

For Milestone 3, we will be required to be able to identify popular and successful businesses in the area in order for users to be able to narrow down their search category even more.

Popular businesses are businesses that seem to attract more customers compared to other businesses in the same category. In order to achieve this level of filtering, I plan to filter only businesses that have a True value for the "trendy" attribute. If a business is trending at a certain time, then it is more likely to be popular than not; therefore, this approach will provide a list of businesses that are likely to attract more customers than the others. Additionally, from this filtered list of businesses that are trendy, I plan to look at the number of checkins and compare it to the average number of checkins for other businesses in the same zip code. Since the trendy filtering method might give some false positives for popular businesses. Hence, to ensure that these false positives are weeded out, only trendy businesses that experience higher than average number of checkins for the area will be shown. If this combination of methods produces lists of businesses that are too narrow or too broad, some readjusting might have to be done.

Successful businesses are businesses which have been serving the community for a long time and which have loyal customers. Most of the time, when a business has

loyal customers that keep coming back time after time, these customers leave behind reviews with keywords and phrases such as "I have been coming here for [x amount of] years" or "Since [event happened], this business has provided [specific service]." Therefore, in order to determine whether a business has loyal customers or not, I plan to run various queries that try to find commonly used substrings in the Review.text attribute of the Review table. With this method, I plan to keep the substrings I will be comparing using the LIKE keyword in SQL as simple as possible in order to miss out on the least amount of reviews from loyal customers as possible.

In conclusion, for popular businesses I believe that it will be sufficient to make use of the statistics and attributes already provided by Yelp about a business' trendiness and number of checkins. On the other hand, in order to determine whether a business is successful and has loyal customers, I plan to run multiple queries comparing common substrings of phrases that are commonly said by customers who are loyal to a business. These queries will try to maintain the substrings as simple as possible in order to prevent missing out a large amount of loyal customers' reviews.