

WSU CPTS 451 - Milestone 2 Metrics Paper - March 31, 2020

Ben Schedin - Minju Lee

The primary goal of this analysis is to define a metric that indicates how popular each business is. Three specific measurements were selected to calculate an overall “popularity score” for each business: The total number of check-ins at each business, the total number of user generated tips for each business, and the average star rating for each business. These measurements are then used to calculate an overall popularity score for each business and a new table is created in the database that associates each business to its popularity rating. Listed below is the query that performs these actions as well as an explanation regarding why each measurement was selected and how the popularity formula was derived.

```
create table businesspopularity as (  
  select b.business_id, b.business_name, hc.category_type, b.business_avg_star, b.num_of_tips,  
    b.num_of_checkins, (b.num_of_checkins + b.num_of_tips) * b.business_avg_star as  
    popularity_rating  
  from business as b  
    join hascategory as hc on b.business_id = hc.business_id  
 group by b.business_id, b.business_name, hc.category_type, b.business_avg_star,  
    b.num_of_tips, b.num_of_checkins  
);
```

The first measure used to determine business popularity is the total number of recorded check-ins at each business. This measure allows us to determine how much service volume each business experiences. This measure was selected because businesses that are attended less often are, by definition, less popular than businesses that are attended more often.

The second measure that was taken into consideration was the total number of user generated tips for each business. This measure provides insight into how engaged each business’

customer base is and was selected because businesses that are more popular would expect to have received more tips than businesses that have received less tips.

The third measure used in determining business popularity is the average star rating for each business. This measure indicates how well each customer responded to the products and services provided by each business. This measure was chosen as a predictor of popularity because customers generally visit businesses with low customer satisfaction ratings less often, and are therefore less popular. Therefore, it would inversely be expected that businesses that are more popular would receive higher average customer satisfaction ratings.

Utilizing the three measurements discussed previously, the formula for popularity defined as $Popularity = (\# Check\ Ins + \# Tips) * Average\ Rating$ was chosen. The total number of check-ins and total number of user tips represent the total customer volume and engagement, and the average business rating represents how well each customer felt about their engagement with each business. This formula outputs a value that represents the overall business popularity, and can be used to rank order each business by its relative popularity score. In this formula, average star rating was chosen as a multiplicative value because it is highly unlikely that businesses that have low average customer satisfaction ratings will be as popular given that most people choose not to frequently visit businesses they dislike. Using this value multiplicatively also helps to scale the popularity score more significantly based on relatively small changes in value (e.g. 4.1 versus 4.7).