Berkeley Conkling

CPTS 451

Business Metrics Write Up

Measuring the success of a business can be done in many ways. In our case, we are only provided with the Yelp data, which can make this task more challenging since we are not provided any information on sales, profit, etc... These would normally be the best metrics to calculate the success of a business, however we can try to measure the relative success of a business through the amount of check-ins, total reviews, and star rating which the Yelp data provided us with. The first idea I had to measure the popularity of a business was to look at the average number of check-ins among all of the business, and then we could measure the popularity of any given business by comparing it to the average. I was able to find the average check-ins with the SQL query "SELECT AVG(num_checkins) AS average_checkins FROM Business;" This query gave an average check-in stat of 0.032 (rounded), which is extremely low, and obviously would not be a helpful metric to use. Since that would not be helpful, I looked to only average check-in values of businesses with a non-zero number of check-ins. Using the SQL query "SELECT AVG(num_checkins) AS non0_average_checkins FROM Business WHERE num_checkins > 0;" Using this query, I found that the non-zero average was 13.11 check-ins (This is among 28 non-zero check-in businesses. Obtained by the query "SELECT COUNT(*) AS non_zero_checkin_businesses FROM Business WHERE num_checkins > 0;"). Using this data, I conclude that we will be able to measure the popularity of the business by comparing its

number of check-ins to the non-zero average. The higher it is above the average, the more successful it is, and vice versa.

Now if we want to measure the success of a business, I think we should consider the popularity score which I described how we would determine in the last paragraph, but also we sohuld consider in conjunction with that metric the review_rating of the business could be used. However, this time I don't think the average should be used as a comparison to figure out what a good review rating is, since I think most would agree that a 3 star or below is not a very good review rating. Hence I think any review below 3.75 should subtract from the success score, while anything above 3.75 should improve it, with the amount of improvement increasing the higher above 3.75 it is. I will create a partial_success metric, and store this value there. Then to calculate the total success I will use the calculation (popularity * partial_success). Thus, if the restaurant has 0 check-ins, it will have a 0-success score, and if it's not 0, the partial success score will determine if it is positive (indicating a successful business), negative (indicating an unsuccessful business), or 0 (indicating neither succesful or failed)