Popularity and Success Metric Proposal

Popularity:

We decided the best method for measuring the popularity of a business is using a metric that compares the number of "recent" check-ins with the average number of "recent" check-ins for businesses in the same category and zip code. The popularity of a business would be derived by its difference from the average. By using only recent check-ins (eg; check-ins that occurred over the past 6 months) we can account for businesses that were at one point popular but have since declined in popularity. This also allows us to avoid any bias that comes from one business being older than another (and therefore having a higher overall check-in count). We are allowed some flexibility with the specific date/time frame we use and may adjust it in the future if we feel it could be more accurate. In addition, specifying the zip code of the businesses ensures we aren't biased towards businesses in cities with high populations, which will naturally have a higher check-in count than businesses in communities with much lower populations. Finding the total number of check-ins for a business past a certain date will require us to aggregate the data from the "Checkins" table, counting each entry that contains the matching "business_id". We will be getting the number of check-ins past a certain date for a business by counting the number of entries in Checkins with a matching business_id and a greater or equal date.

Getting the recent check-in count for a business:

SELECT COUNT(*) AS recent_checkins

FROM Checkins

WHERE business_id = 'The ID of the Business we are analyzing'

AND date >= CURRENT_DATE - INTERVAL '6 months';

We will be finding the average check-in count for relevant businesses by first joining Checkins, Business, and Categories by the business_id key. Then we count all entries with a matching category name, matching zip code, and greater or equal date, and use this sum to find the average check-in count.

Getting the average recent check-in count for relevant businesses:

SELECT AVG(recent_checkins) AS avg_recent_checkins

FROM (

SELECT COUNT(*) AS recent_checkins

FROM Checkins

```
        INNER JOIN Businesses ON Checkins.business_id = Business.business_id

        INNER JOIN Categories ON Business.business_id = Categories.business_id

        WHERE Categories.category_name = (

            SELECT category_name

            FROM Categories

            WHERE business_id = 'The ID of the business we are analyzing'

        )

        AND Business.zipcode = (

            SELECT zipcode

            FROM Business

            WHERE business_id = 'The ID of the business we are analyzing'

        )

        AND Checkins.date >= CURRENT_DATE - INTERVAL '6 months'

        GROUP BY Checkins.business_id

    ) AS RecentCheckinsPerBusiness;
```

Success:

We decided to measure the success of a business by simply comparing the average number of stars in the business' reviews with a threshold of our choosing (eg, 2.5 stars). We considered multiple ways of weighing or comparing this metric but found that simply using the same rating system for all businesses was much more fair than, for example, comparing businesses in the same zip code or weighing a user's review based on their average review star count. However, we may find that all users are on average more or less critical than we initially expected. In this case we can adjust the star threshold (for all businesses) to be more in line with what the average user would consider a "successful" and well-liked business. To find the

average number of stars for a business' reviews we will simply take the average review_stars value in the Reviews table for all entries with a matching business id.

Getting the average number of stars for a business' reviews:

SELECT AVG(review_stars) AS average_review_stars

FROM Review

WHERE business_id = 'The ID of the business we are analyzing';