Methods to Classify Businesses as Successful and Popular.
Group CPDG

Successful and popular are important traits to businesses but are not immediately obvious. Utilizing the Yelp data, we will determine the metrics and algorithms that can identify businesses that fit in these classifications.

Definitions of popular and successful businesses: Popular businesses attract more customers compared to others in its category. Successful businesses are long standing and have many loyal customers.

The metrics used to determine a successful business are:

Average Rating, taken from an average of the "stars" field of all reviews given for a business. This field is already provided in business as reviewRating.

Business age, estimated by the difference in the "date" field between the oldest and newest reviews on a business. This metric is meant to measure how long a business has lasted.

The following query retrieves the business age for each businessId in business:

```
SELECT businessId, EXTRACT(days from MAX(rating.date)-MIN(rating.date)) as businessAge
FROM rating
GROUP BY businessId
ORDER BY businessAge;
```

Average Rating Difference, this is an average rating difference a business has achieved from the average of other businesses ratings containing any of its categories. This measures how much better a business is rated than its competitors on average. This is calculated by the formula: (average rating of business) / (average of the businesses average ratings sharing a category).

The following query retrieves the average rating difference for each businessId in business:

```
SELECT business.businessId as businessId, business.reviewrating - AVG(competetor.reviewrating) as ratingDifference
FROM business as competetor, business
WHERE business.businessId <> competetor.businessId and
    EXISTS (SELECT category
            FROM BusinessCategory
            WHERE BusinessCategory.businessId = business.businessId and
                category IN (SELECT category
                        FROM BusinessCategory
                        WHERE BusinessCategory.businessId = competetor.businessId))
GROUP BY business.businessId
ORDER BY ratingDifference;
```

A business that has maintained a high average rating difference over a long time period has demonstrated to be more capable than a business who still maintained a high average rating difference but over a short time period. This means that business age weighted by its average rating difference is a good quantitative measure of its success. This results in the formula:

Success Score = Average Rating Difference * Business age

The metrics used to determine a popular business are:

Review Count, the count of all reviews made on a particular business.

Review Frequency, the frequency at which reviews are made on a business. Determined by the number of reviews per day which is calculated from the review count and the estimated business age taken from the metric proposed for success. The formula for review frequency is: Review Count / Business Age.

The following query retrieves the review frequency for each businessId in business:

```
SELECT business.businessId as businessId, business.review_count/ages.businessAge as reviewFrequency
FROM business, (SELECT businessId, EXTRACT(days from MAX(rating.date)-MIN(rating.date)) as businessAge
                FROM rating
                GROUP BY businessId) ages
WHERE business.businessId = ages.businessId and ages.businessAge <> 0
ORDER BY reviewFrequency;
```

Check-ins per Person, the number of check-ins to a business per person in the population of the business's zip code. This metric estimates local popularity by how much of the community around the business visits it. The formula for check-ins per person is: (Total Check-ins) / (business zip code population). Total check-ins is the total of all check-ins to a business on all days.

The following query retrieves the check-ins per person for each businessId in business:

```
SELECT businessId, business.numCheckins/CAST(zipcodeData.population as FLOAT) as localPopularity
FROM business, zipcodeData
WHERE business.zipcode = zipcodeData.zipcode
ORDER BY localPopularity;
```

Together a business's review frequency and local popularity estimate business popularity with the formula:

Popularity Score = Review Frequency * Check-ins per Person

Together the formulas for popularity and success determine whether a business is popular or successful by its ranking among the popularity and success scores calculated for all other businesses. A business's highest ranking score determines the category it belongs to.