# Identifying Optimal Business Locations

## David Simon

# Introduction/Business Problem

## Background

Location of customer-facing businesses are crucial to success. This is evident by companies investing several thousands of dollars in footfall studies before deciding to open a store in a specific location. While marketing research highlights that the strength of a retail brand might exceed the advantages provided by location,[[1]](#footnote-1) it is seldom disputed that a new business has a substantially better chance of survival if its initial location is well-chosen.

## Business Problem

Where should one build a new coffee bar, a sandwich shop or a restaurant? The goal of this project is two-fold: (1) to identify locations of already successful food retail establishments in the City of San Francisco and (2) to attempt to identify locations in the City of San Francisco that possess similar characteristics and may be promising targets in which to build such establishments. The results of this project will be of interest to new entrepreneurs in the food business as well as existing restaurant chains exploring new locations.

# Data Sources

This analysis will crucially rely on data from FourSquare’s API as well as the geospatial data on the boundaries of San Francisco provided to us in an earlier lab. The two initial choices that the modeler has to confront are the following: (1) what constitutes a successful business?, and (2) how does one characterize the vicinity of businesses that makes them successful? I discuss these two questions next.

## Identifying Successful Businesses

Two potential measures of success for a business that can be assessed through the Foursquare API are (1) the rating associated with the business and (2) the number of tips associated with a business. It is beneficial to use these two measures in combination with one another. For example, a high rating with a low number of tips might indicate a place that has very few visitors and a high rating might be a statistical aberration. In a similar fashion, high number of tips but a low rating might mean a bad business that brings about a lot of intense negative reactions from visitors.

A candidate measure for the quality of the business would be to convert all ratings and number of tips into a z-score. That is, every from every rating and number of tips I subtract the mean for a comparable groups (for example, if a Chinese restaurant has a rating of 3.8 and the average Chinese restaurant has a 2.5 rating, this will lead to a score of 3.8-2.5 = 1.3) and then divide this difference with the standard deviation of the scores of all Chinese restaurants (in the above example, if the standard deviation of all Chinese restaurants’ score is 2, I arrive at a z-score of 1.3/2 = 0.65). This way, both the business ratings and the number of tips will have a zero-mean and standard deviation of one distribution. After this standardization, I can add up the z-scores associated with ratings and tips for a given restaurant to find the highest composite scores.

Alternatively, I can just use the number of tips as a measure for potential of the business. Some of the businesses are good, some are bad and that is determined by the quality of people who run the place, the quality of the business concept, etc. However, the number of tips, as far as they serve as a proxy for the number of visits might be a better indicator of exogenous factors to a restaurant’s success – that is, factors that are independent of the quality and the business concept of a place. I would expect that even a bland and boring sandwich place located among bustling office buildings would have a high number of tips.

This methodology allows me to identify the top 5-10 successful business in San Francisco within each category.

## Assessing the Neighborhood of Successful Businesses

## 

For purposes of this project, the assumption will be that the businesses found in the vicinity of one type of successful business will be very similar to each other. For example, as noted above, one would expect a popular lunch spot to be surrounded by office buildings. In the first step, I test the similarity of the neighborhoods of popular places in one category to each other. Unless they are drastically different, the “average” neighborhood of a successful business seems to be a good target for a search.

Once this target is identified, I can define a “coordinate grid” for the city of San Francisco and iterate through coordinates to find locations that are similar to the target. These locations will be suggested as potentially promising locations for a new business of a certain type.

1. See, for example, Swoboda, Bernhard, et al. "The importance of retail brand equity and store accessibility for store loyalty in local competition." Journal of Retailing and Consumer Services 20.3 (2013): 251-262. [↑](#footnote-ref-1)