

Venue Location Recommender

Introduction/Business Problem

- When someone is looking to open a venue such as a shop, restaurant or bar, how do they decide where to open it?
- One option is to look for areas that have fewer of that type of venue nearby (lower density)
- On its own this strategy is likely to be insufficient because an area may have lower density of this type of venue because there is limited demand for it
- In order to make a better decision they also need to be able to identify areas where demand for the venue is likely to be high
- If we can quantify both the density of similar venues around a location and the likely demand, good potential locations could be characterised as having low density compared to demand
- Quantifying density of venues around a location is a fairly trivial task, but how can we estimate demand?
- One solution is to assume that demand is related to the density of other venues around the location that are not of the same type as the venue being opened. For example if you are seeking to open a café, demand may be indicated by the range and number of non-café venues such as work places, shops and other restaurants nearby.

Data: defining the search area

- Given an area where we want to open the venue we can split this area up into smaller areas
- Our starting point will be a venue type and location described by its longitude and latitude, a radius around the location we want to search in and a distance within which we want the returned location recommendations to be provided by
- For example we might want to know where to open a café within 1km of the centre of Toronto to the nearest 200m.
- Given these inputs we can covert them into a grid of points that are a distance of 200m apart in a circle of radius 1km around the centre of Toronto

Measuring density of target venues

- Using the search Endpoint of the Foursquare API we can pull a list of venues v of the type we are looking to open around each point p in the grid together with their distances d from p
- The radius r of this search will be the same distance as the distance between points in the grid
- This can be turned into a single measure of density D for the point in the grid as follows:

$$D(p) = \sum_v r - d$$

- So if $r = 200\text{m}$ a venue that is 0m from the point will add 200 to D , while a venue which is 200m from the point will add 0 to D

Estimating demand for venues

- We can also pull a list of other venues around p together with their categories and distances from p using the same Foursquare API Endpoint
- For each point and category we can calculate similar density metrics
- We will then use these densities as the independent variables in a regression model to predict the dependent variable, in this case the density of the target venues
- The theory is that a good location for a new venue will be one with a higher predicted density for that type of venue than the actual density we see
- So we can rank each point p by the predicted – actual and choose the best n points as location recommendations