Selecting a Borough in London

Client Problem/Project Overview

Moving cities or even within a city is often a daunting task, and selecting where to live within a large and cosmopolitan city like London is even more difficult as the rents can vary widely from borough to borough (what the Londoners refer to large neighbourhoods or groups of postcodes as – for example South Kensington neighbourhood is contained within the borough of Kensington and Chelsea) – so it becomes important for a potential resident to understand how the neighbourhoods differ (or are similar) to make an informed choice of where to live. Whether one is buying or renting will also impact this, for the purpose of this scenario we will take the point of view of a renter without children.

Towards this end we will consider how the boroughs of London cluster together – determining any differences that seem significant from one to the next, as well as consider the cost of living in each (as approximated by average weekly rent) borough, combined with the average salary to get a sense of how these variates can be predicted or correlate with the different classes of borough from clustering. Finally, we will suggest based on these factors and a deeper dive into a single borough in the selected class why it will make sense to live there.

We’ll conclude with a suggestion to the perspective renter:

Literature review:

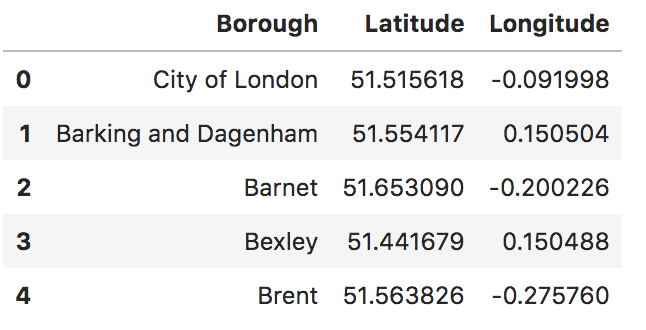
There has been a large body of literature on the boroughs of London mostly unscientific or fairly unanalytic which can be seen from articles in the popular press which focus on one or two metrics (combined with anecdotal and largely unverified opinion) which covers things like best neighbourhood for students, best neighbourhood for families, etc such as this one:

<https://www.movehub.com/us/international-moving/united-kingdom/london/where-to-live/>

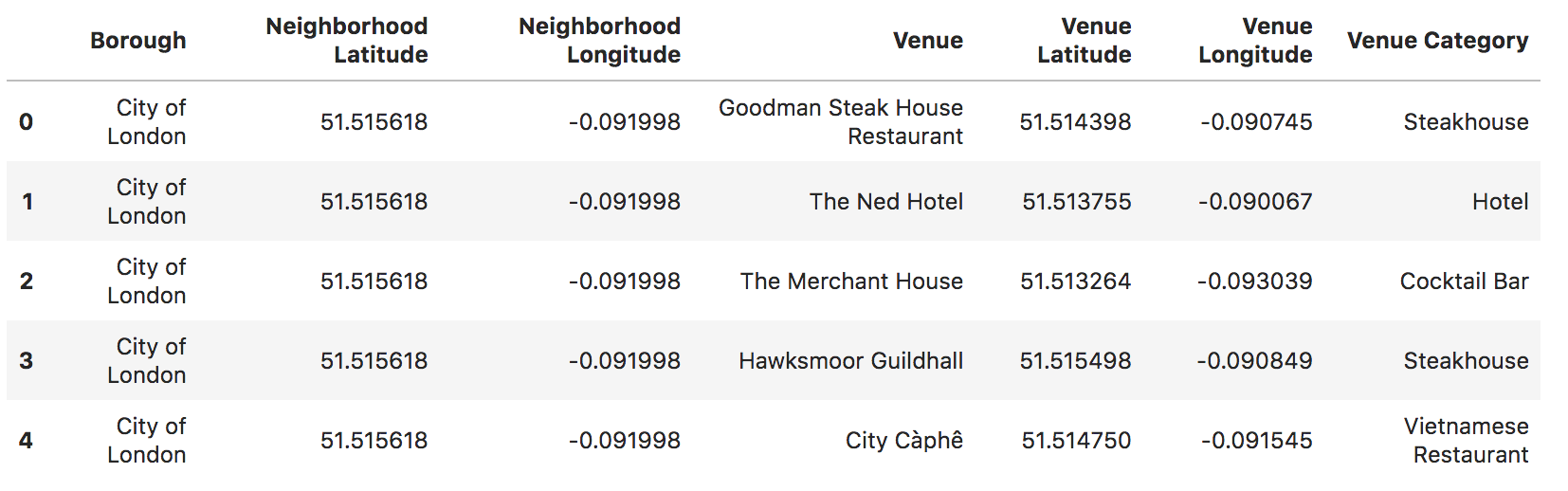
with very little by way of deep dives into neighbourhood structure as we are attempting in this study.

The data:

There are two primary sets of data that will be used – the first was venue information from foursquare – towards that end we collected geospatial data on the boroughs using a list of boroughs from the London Data Store, and then geocoder from the geopy package – next with our list (as below)



we collected venue data from to understand how the boroughs cluster – to discover which neighbourhoods are more similar to each other than others, we will also take a dive into the data to get a sense of lifestyle factors in each borough – ideally identifying and clustering the density of places such as gyms, schools, restaurants etc – and which neighbourhoods have the highest average restaurant reviews etc. – unfortunately school data was not available, and on the free license only basic venue information such as name, location, and category were available as seen below:



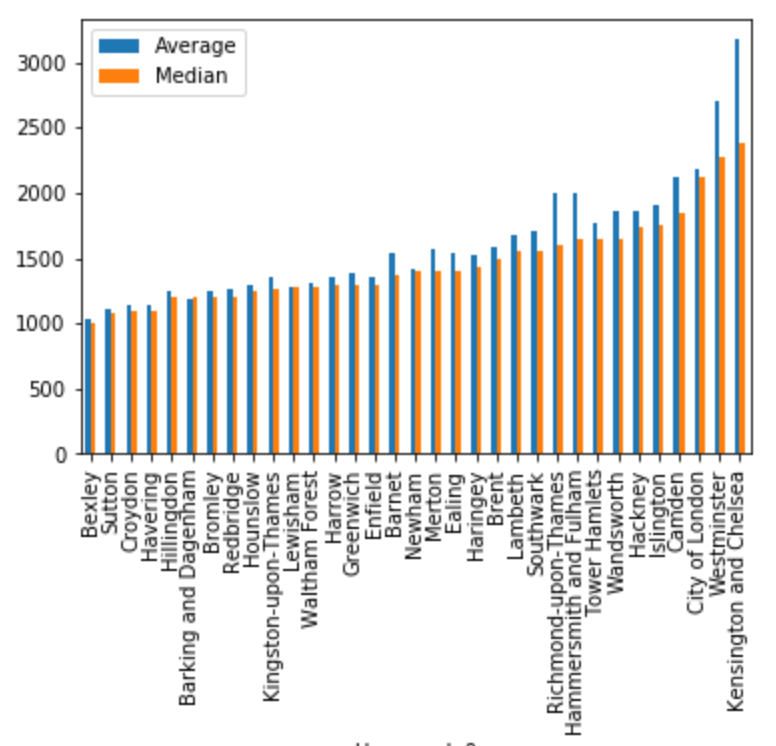
then, with what we had, we aggregated the data by borough and venues scoring (using the one hot vector approach) to get a borough analysis with relative density of the various venue types as seen in the following extract:



The next step after analyzing the data was to combine this with data from the Office of National Statistics (UK) acquired from the City of London in the form of the average and median monthly rent per borough in the year 2018 which we can find here:

<https://data.london.gov.uk/dataset/average-private-rents-borough>

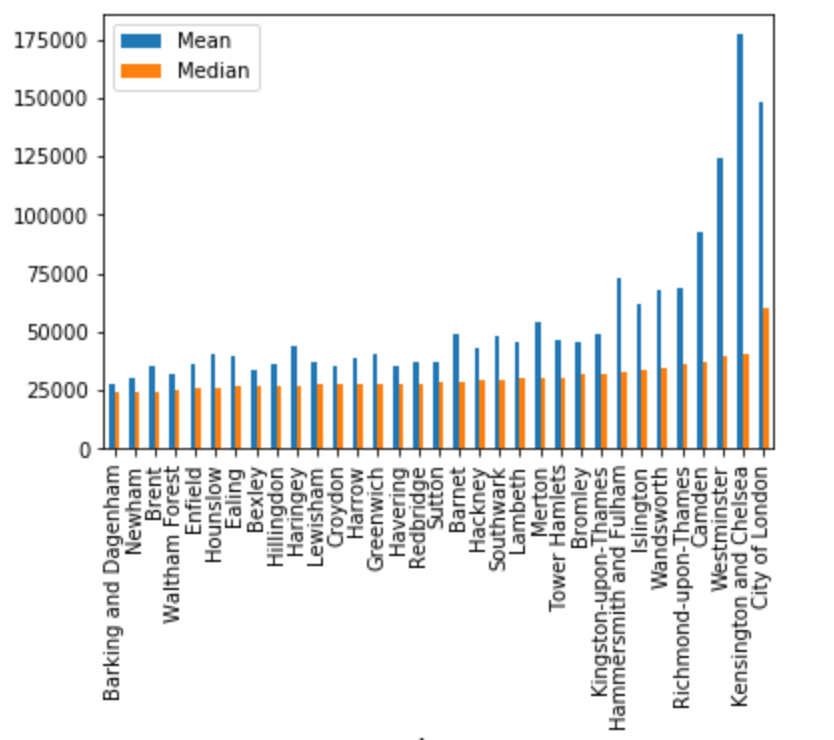
which if we look at a bar plot with the mean and median rent for each borough we can see rent as being quite sticky in many boroughs which breaks away at the more expensive boroughs with spikes in the averages as we would expect in neighbourhoods with a few properties commanding massive rents like in Chelsea for example.:



as well as the average and median annual salary (for also the year 2018) per borough which can be found here:

<https://data.london.gov.uk/dataset/earnings-place-residence-borough>

where we see that there are some massive outliers in the average salaries – as expected from the economic literature the distribution, like the distribution of wealth follows the so called power law.



It was for this reason that we used both the median monthly rent, and the median annual salary as our variates of choice.

After we processed and cleaned these two sets (and removing unwanted formatting) we attempted to answer some questions: do the different types of borough have different rents or incomes? Within a particular cluster which borough has the cheapest rent, etc?

\*\*Note – we only were able to query up to a one-mile radius in each borough, and were limited to the number of results by the free license – leading to further work to be done.

Methodology:

We will use clustering analysis in order to work out broad categories of similar neighbourhoods – with this in hand it means that using boroughs that are similar one can then optimize based on other factors such as rent or salary in this case – bearing in mind that it is a within category optimization – as between categories linear categorical regression will be used in order to reveal if there are any differences in salary or rents between the categories- before this however we will take a look through the boroughs (as displayed in the data section) and see how the rent and salary breaks down on a median and mean basis – in order to see if there is any interesting variation or obvious patterns, and finally a simple exploratory analysis of the suggested borough demonstrating the layout and distribution of venues in the borough.

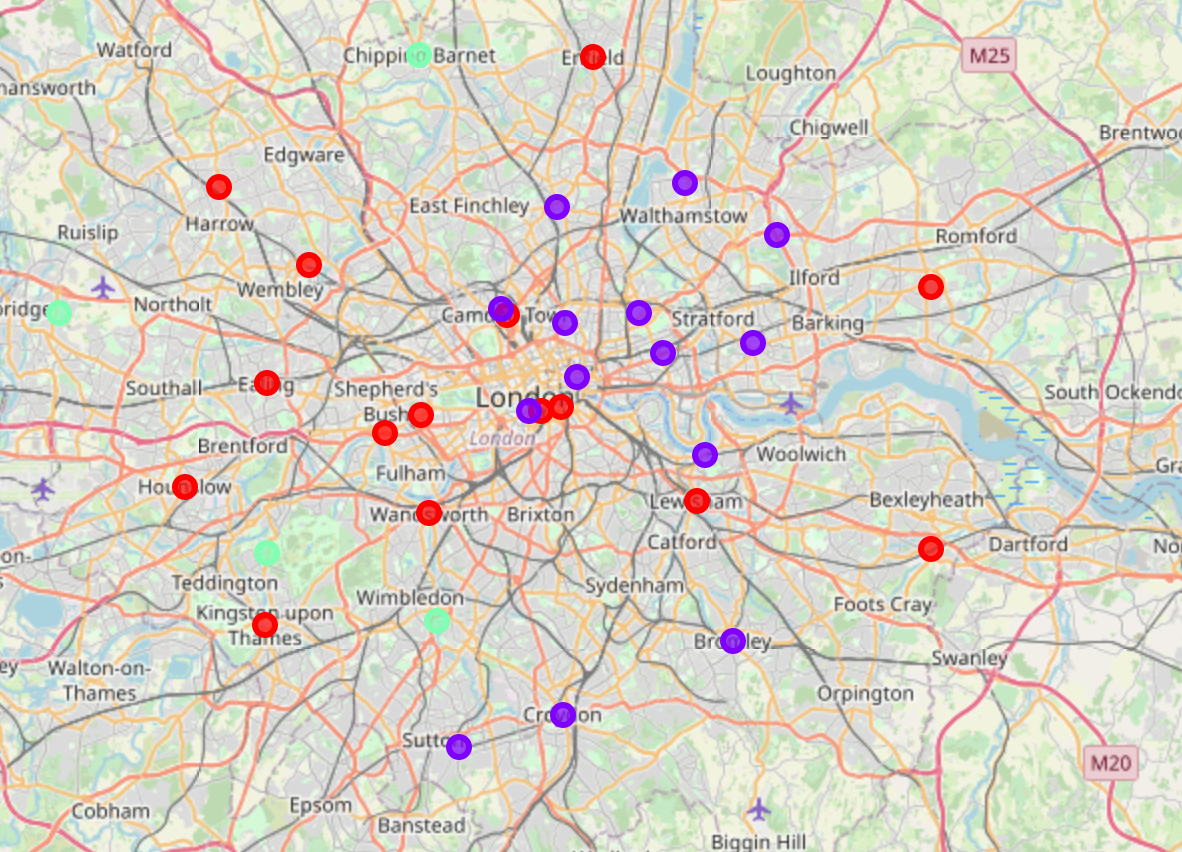
The results: our selection of data from foursquare:

Clustering:

Applying the clustering algorithm, we got somewhat robust results using a choice of three centers with the list being:

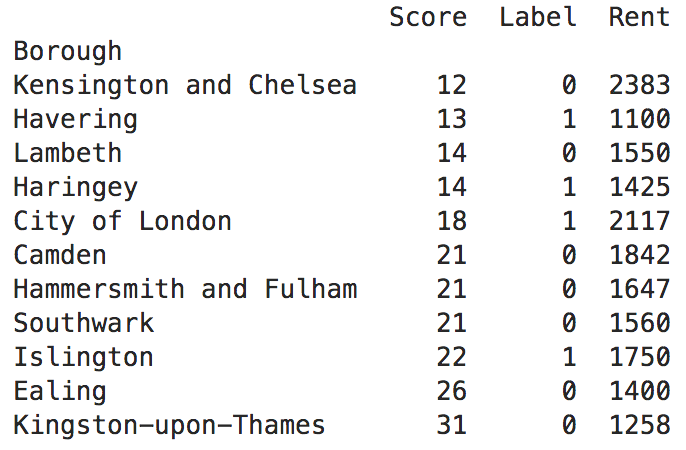


and visually:

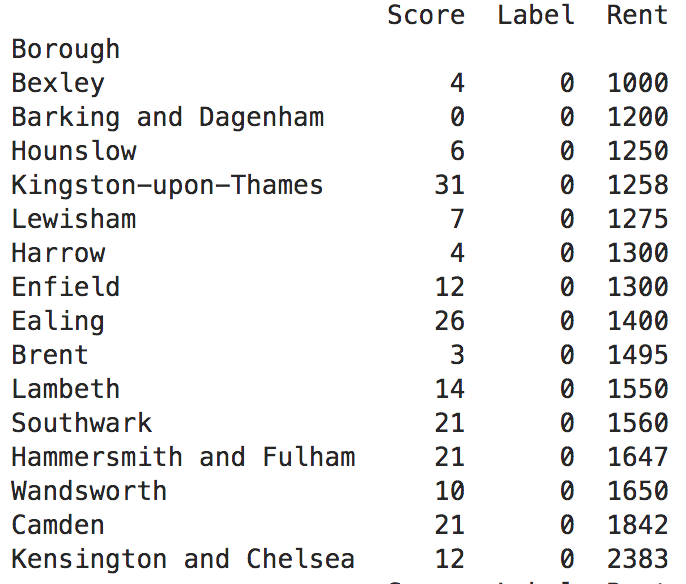


with an interesting east-west breakdown in the two dominant clusters – however to get a better sense of how the clusters were segregating we looked at a subcategory of venues (restaurants and cafes) and from this we rated each borough on the number of different restaurant (or eatery) category types (this is an important metric as for many people who live and move to cosmopolitan cities the choice and variety of cuisine available is very important) and what we find, when ranking based on this rating) is that cluster zero features heavily amoungst the top 3rd of boroughs relative to restaurant diversity:

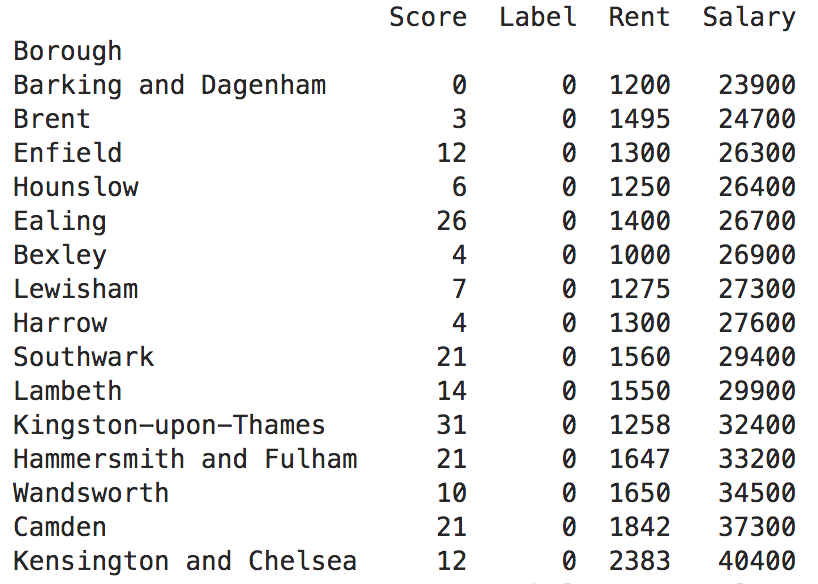
Continuing with restaurant diversity as one of the features that the prospective Londoner is considering, we counted up the number of different restaurant types found in each borough and then ranked them based on diversity vs rentals



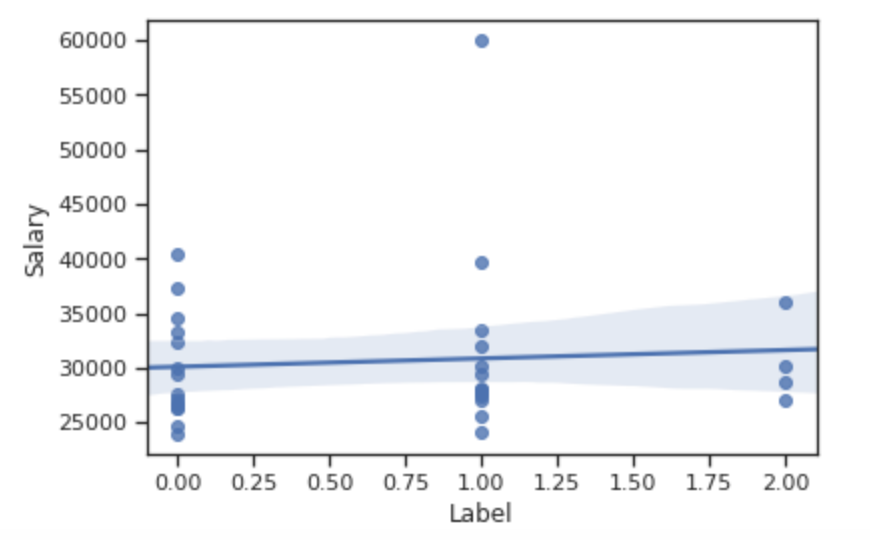
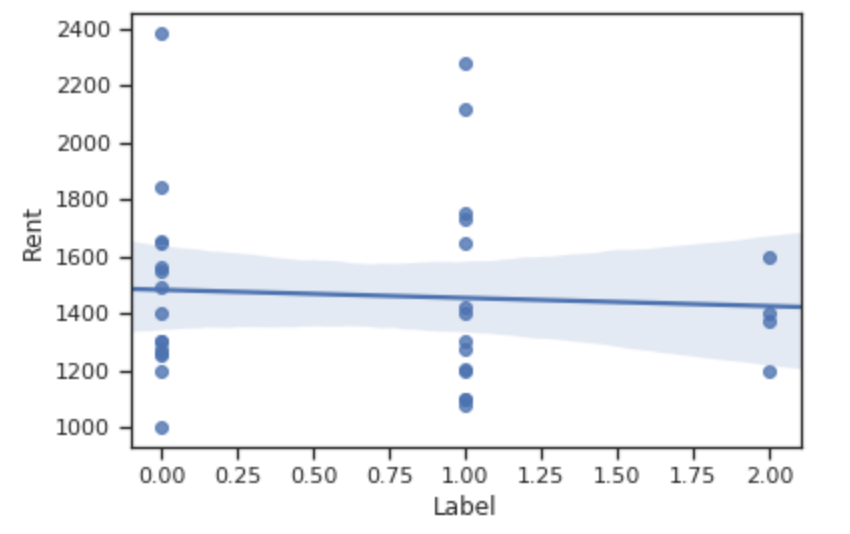
and so we will work as cluster or category 0 as being our set of boroughs under suggestion to the client, and here we will present category zero ordered both according to salary and rent we get:



From which we see that Kingston-upon-Thames is one of the most diverse boroughs with respect to dining, and on the other hand we see that the rent is in the bottom third of that cluster, and if we explore based on salary:



We find that despite having low rents the borough is close to the top third in terms of average salary – it’s possible that the diverse neighborhood adds so much to quality of life that it also translates into career success (though this does require a study of its own to be scientific) - that said, we can see that neither salaries nor rent overall seem predictable based off of the clustering categories (with a regression of rent or salary against the category) in the charts below:

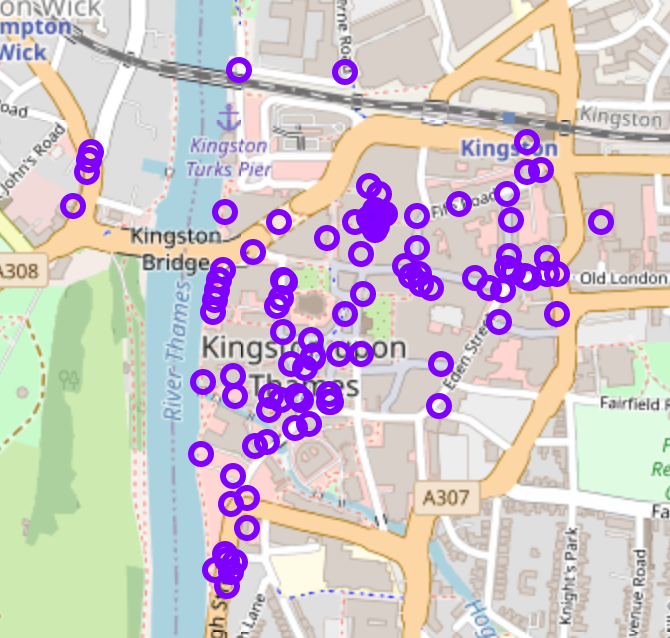


and for both categorical regressions the r squared was under 0.1 implying very little of rental nor salary is accounted for by clustering boroughs based on venues.

Discussion:

Based on our limited data from Foursquare it seems advisable to consider the more diverse (or desirable category 0 boroughs (as being more diverse) but noting that within that category there are widely varying rents and salaries – meaning that there is a place for all walks of economic life to find a borough that meets their financial constraints, and can satisfy their interests in dining or take out at the same time.

We suggested Kingston-upon-Thames for its diversity, and yet above average income, and below average rent with a good spread of venues in the borough making selection of a location within the borough also easier:



We should note though it is an outer borough, which means getting to other boroughs is more difficult from there, possibly accounting for the lower rent – a new study taking into account distance from other borough, commuter routes etc – may push the needle back towards the center of the city.

Conclusion:

The next stages would be to get more venue data, with richer content – reviews, cost rating to make a more informed decision, plus other local features not accounted for such as schools etc – this would greatly enhance, and further build the ability to make personalized suggestions based off of individual preferences.