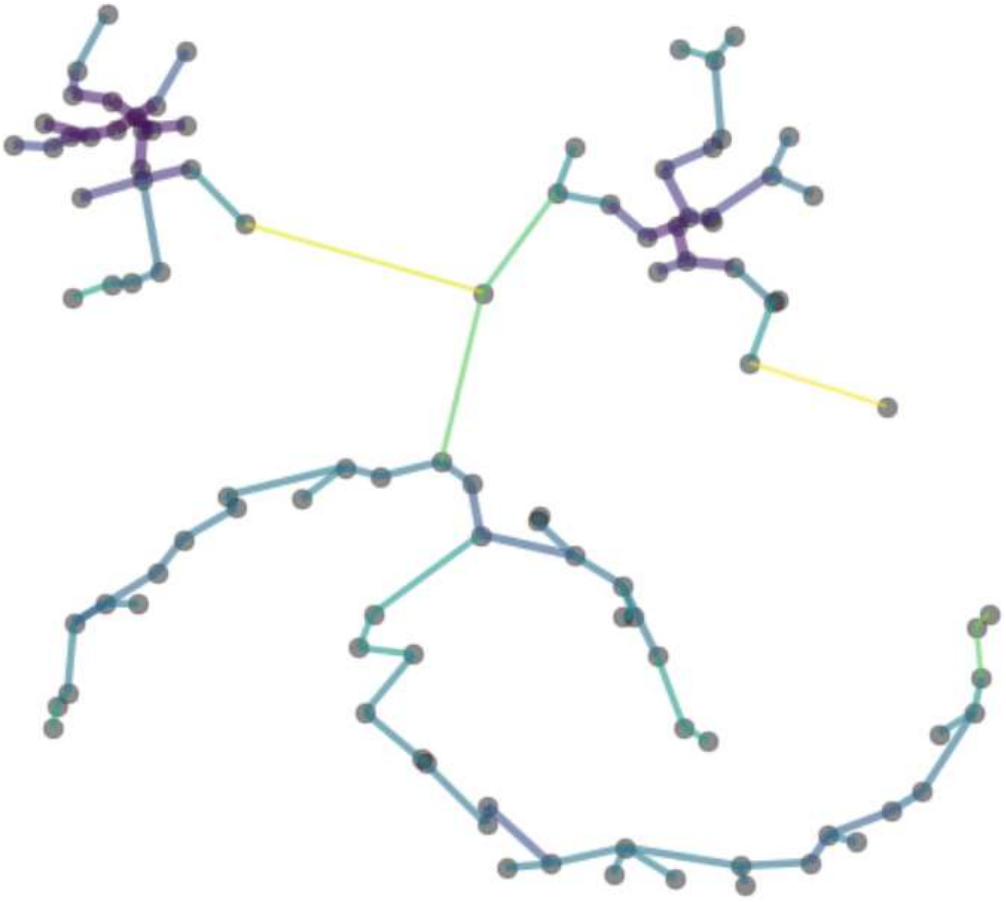
MARCH 2020

Report

London Urban Farming Placement Plan



Picture source: https://hdbscan.readthedocs.io/en/latest/how_hdbscan_works.html

Authored by Claude Nortje

Introduction

Food security is an important topic in general especially with the ever-expanding population. Feeding our growing population can be lucrative to those market participants with capital and knowledge. Thorough planning is necessary prior to setup. Constructing an urban farm is extremely capital intensive at commencement and as such making informed decisions is essential for successful implementation. The value that data science provides to entrepreneurs and other market participants during planning far outweighs the costs of making poor decisions.

For my topic I have chosen to identify areas in the greater London area best suited for urban farming. I will do so by applying Data Science methodology, the python programming language and machine learning techniques to current publicly available data to identify ideal locations.

Target Audience

This project is aimed at those entrepreneurs building a proposal for their finance application and other market participants with capital wanting to invest in an urban farm. The scope of this project is limited to finding optimal locations and would form part of an overall business plan.

Background

What is urban farming? Let me begin by expanding on this business use case

Urban farming is the practice of sustainable food production in towns, city centers and other densely populated areas[1]. Advanced modern farming techniques and nontraditional materials are used in the construction of these farms which enable farming operations to produce food continuously and with more flexibility. By providing optimal growing conditions tailored to selected crops farmers can achieve highly nutritious produce much faster than traditional farming. Environmental control and management form the basis for these achievements.

Vertical farming is one such design as illustrated in the below pictures.

These modern farming designs are well suited for urban areas. Existing buildings and surrounding infrastructure are leveraged in setting up these controlled environments.

Densely populated urban areas offered other advantages such as proximity to distribution channels and potentially low-cost existing infrastructure[2].

[1] https://en.wikipedia.org/wiki/Urban_agriculture

[2] https://www.researchgate.net/publication/289524313_Urban_Hydroponics_for_Green_and_Clean_Cities_and_for_Food_Security

There are many positive social impacts for nearby communities. ie More affordable and consistent supply of highly nutritious food. Communities with higher levels of unemployment benefit from having businesses in their area providing job opportunities. Professional development through training and other education programs compound these benefits.

Surrounding property values can also benefit from having an urban farm developing and improving the farming premises especially where the building was in a bad state. Building selection is one of the main focal points of this project and will be discussed in more detail below.

Problem statement

Phrasing this project's problem into questions for better understanding of the solution/s needed

1. Where are potential customers situated and can we group their locations?
2. Which communities would provide employees and would benefit most from having a business like this nearby?
3. Based on the results from the first two questions, which buildings offer the best opportunities an urban farm?

Factors in solving for location

1. Customers and distribution
2. Building and infrastructure
3. Employees

Customers

Customers of an urban farm include wholesalers, food manufacturers, supermarkets, grocery stores, fruit and Veg markets, stall owners at farmers markets and restaurants. Delivering produce to these outlets can be costly

As mentioned in the introduction establishing an urban farm is capital intensive at inception. Not mentioned thus far but equally relevant is the delay of revenue from sale. Usually produce is not ready at commencement. As such there is a huge focus on keeping operational costs low. Distribution cost is the cost for transporting goods to customers and can be a substantial cost for any business. As such an urban farm's location can benefit from low distribution cost if located near many customers.

Building and Infrastructure

To be able to control the environment in the context of an urban farm, a building fit for purpose is needed with features such as size, heating/cooling systems, water supply, drainage, office space, collection and storage areas. Renting or purchasing a suitable building Is another important decision which will impact on capital allocation. It is advantageous to select a building that is structurally sound with most if not all of the feature but does not have aesthetic appeal nor situated in a premium location.

Employees

Another factor for consideration is the need for human resources. Most of the work within an urban farm that cannot be automated would require employees. Finding employees is an important factor for any business. Providing job opportunities as well as developing social cohesion between the surrounding community and an urban farm has great benefit to both business and community. To maximize this benefit, levels of unemployment per area need to be analyzed.

Data

Acquiring and use of data

Step 1 in this process is to determine what data is needed to solve for the problem questions

For C*ustomers* (and in line with guidelines) the Foursquare platform will be used for the collection of a large sample of potential customers.

For *Building and Infrastructure* web-scraping will be collected from prominent realtor site as well as property reports from the UK government website.

For *Employees*, such demographic data will be sources from the UK government website.

Step 2 is extracting relevant part of the data for its usefulness and determining if further information is needed

For Customers

At a glance, this data shows potential customers per category as well as the coordinates per customer. This data comes from Foursquare.

	name	categories	lat	lng
0	Sainsbury's	Supermarket	51.492824	0.120724
1	Lesnes Abbey	Historic Site	51.489526	0.125839
2	Lidl	Supermarket	51.496152	0.118417
3	Abbey Wood Railway Station (ABW)	Train Station	51.490825	0.123432
4	Co-op Food	Grocery Store	51.487650	0.113490

For Building and Infrastructure

Here is a visual of the geographic distribution of industrial land in the London area.

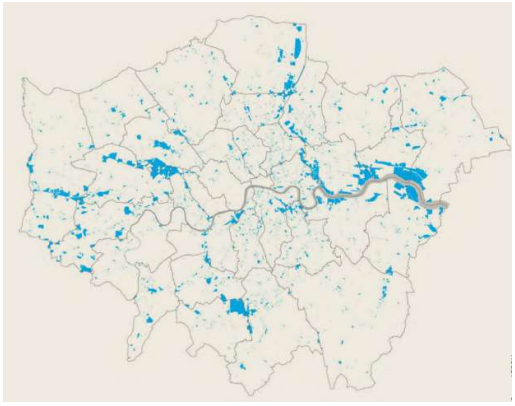


Figure 1 Source: www.london.gov.uk

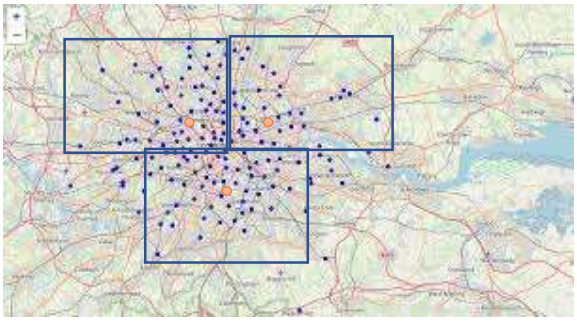
For Employees,

This data contains unemployment levels per area as well as the amount pay.

Code	Area	2017		2018		2019	
		Pay (£)	conf %	Pay (£)	conf %	Pay (£)	conf %
00AA	City of London	#	#	902	19	#	#
00AB	Barking and Dagenham	461.0	5.1	479.1	4.9	472.9	6.2
00AC	Barnet	522.6	4.5	536.6	4.8	536.6	4.4
00AD	Bexley	513.0	4.6	513.8	5.3	550.2	5.2
00AE	Brent	471.0	4.7	480.0	3.9	524.6	4.0

Step 3 How this data will be used for solving this projects problem.

In this final step machine learning will be used to calculate the optimal locations. This method involves clustering of similar locations. These clusters are determined based on the feature set built in the above steps. Distances between locations are measured (Ie Euclidean distance) and resampled to optimize for nearest “best results” distances.



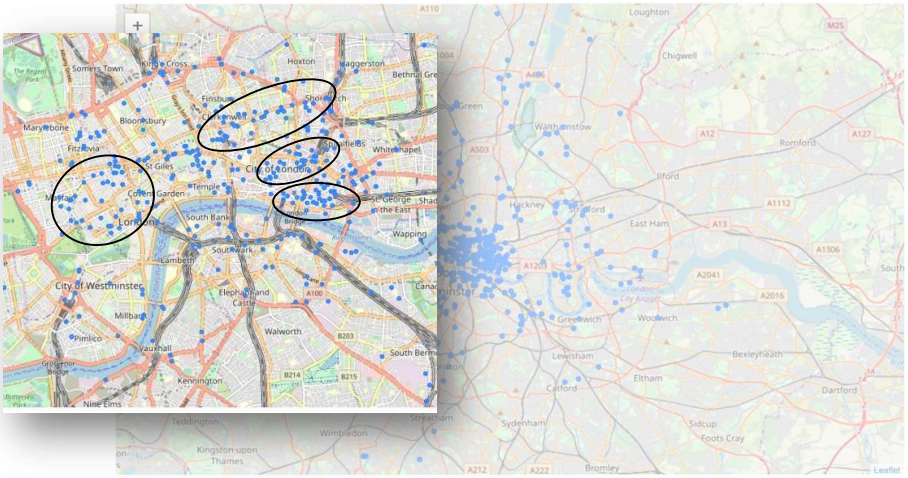
Methodology

Crisp DM methodology was use throughout this project, though the exact order of each step was not necessarily followed. The general flow of the information reflects that of the Crisp DM methodology. Crisp DM was selected for its ease of use and understanding as well as supporting documentation on IBM’s website.

Collecting the initial data consisted of downloading a CSV file, web scraping and using the Foursquare API.

This data was mapped onto a folium map at each point of data collection. Visualizing data was an important process to enable the spotting of patterns, easily recognize aspect to consider later in the analysis as well as identifying data that was not be relevant.

The data was of particularly high-quality with very little cleaning required.



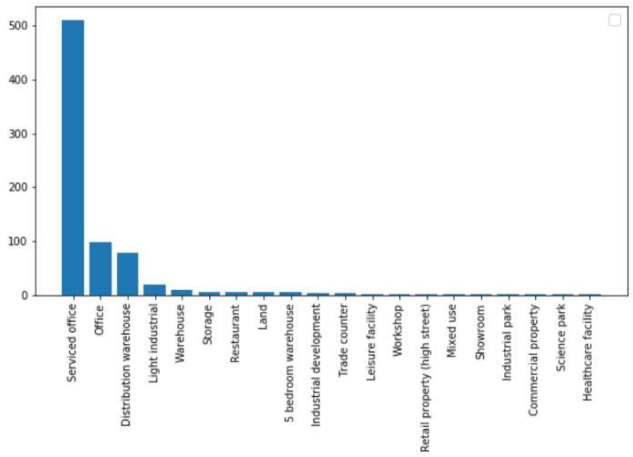
Some notable stats and patterns:

Commercial properties

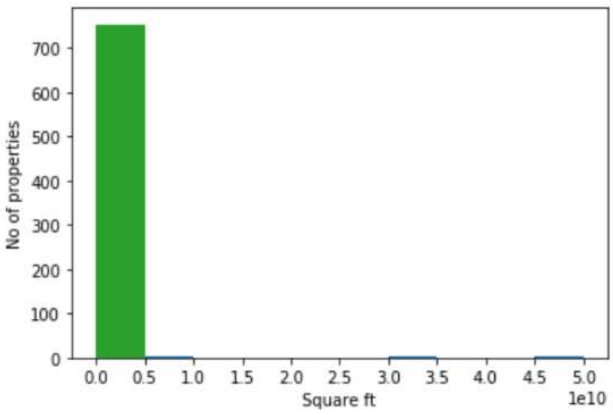
Bar chart view

From the above we can note that small offices dominate the commercial properties available followed by warehousing. This could allow the business to split recourses into a node-like structure. Starting small and growing in scale may be more viable. Taking advantage of greater reach with a node-like structure would avoiding saturating opportunities close to multiple sites. The caveat to these stated advantages is that a large portion of office space aren't suitable. For instance, above ground level would need an appropriate lift or ramp for the movement of product. Depending on the scope of the business, receiving and dispatch access would be appropriate. Weight of equipment may impact choice as well. Lastly, regulations would also influence the building selection. All these nuances are outside of this analysis and would need further consideration in the next phase of planning once a feasibility study of highlighted buildings have been done.

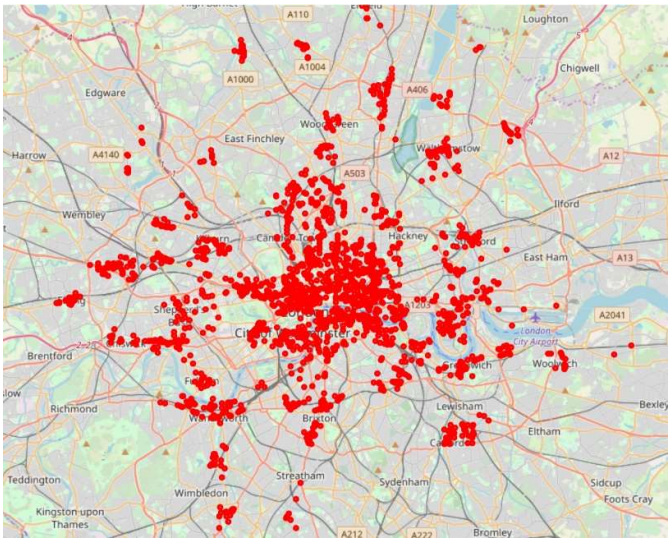
Types of properties



Property size distribution



Potential Customers



Mapped data points view

The data points appear in highly dense clusters.

```
Top 15 customer categories of 2779 entries
'Grocery Store'            465
'Coffee Shop'              270
'Café'                     259
'Restaurant'                93
'Italian Restaurant'        86
'Pizza Place'               82
'Indian Restaurant'         75
'Bakery'                    74
'Fast Food Restaurant'      73
'Chinese Restaurant'        68
'Market'                    67
'Burger Joint'              56
'Sandwich Place'            53
'Supermarket'               42
'Pub'                       35
```

Employees

Choropleth map view

This choropleth map clearly displays areas of all low earning individuals. The income range is between £12,000 and £80,000 per annum. This map consequently assists in offering appropriate salaries as well knowing which areas need uplifting.

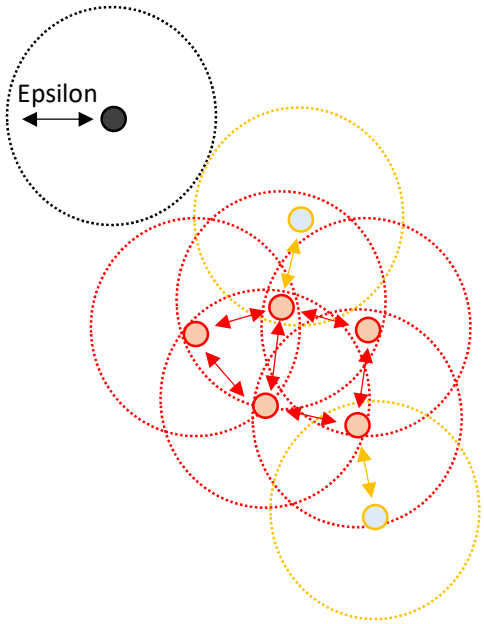
Modelling

Density based spatial clustering of applications with noise or DBscan was selected for this project. The added benefit of using the DB scan algorithm Is its robustness against outline data points. The algorithm traverses each data point and allocates one of three classes for each data point based on the neighboring data points distance within a set radius. the three classes are a core point a border point poor noise (an irrelevant point). The algorithm distinguishes between a core point and a border point by counting the number of

neighboring points within a radius to determine whether the total number of neighboring points are equal to or more than a predefined minimum value.

Should the total number of neighbouring points within the set radius be equal to or more than the predefined minimum value then the labels the selected datapoint as a core point. Should the total number of neighbouring points within the set radius be less than the predefined minimum value then the selected data point is labeled as a border point.

See sketch below.



Radius = Epsilon

MinPts = 3 (Selected point plus neighbours within its radius)

Core Point

Red point has at least 2 neighbouring points within its radius

Border Point

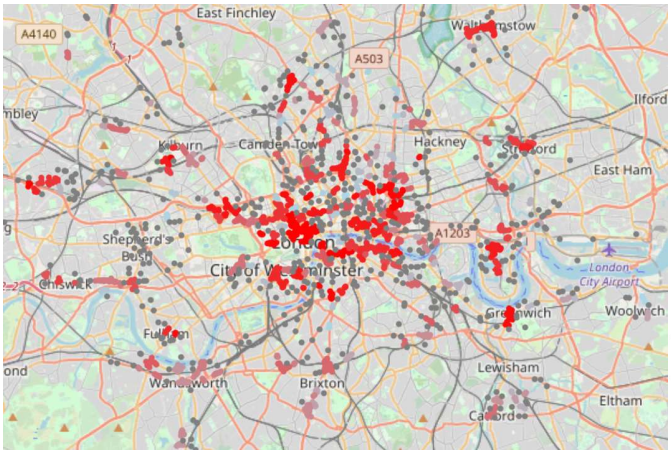
Yellow point has less than 2 neighbouring points within its radius

Noise Point

Black point does not have any neighbouring points within its radius

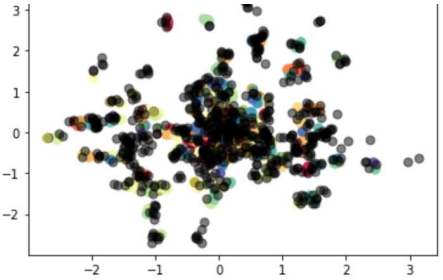
Results

DBscan revealed 225 (the additional is the group of outliers) clusters of potential customers in close proximity to one another with many outliers removed.

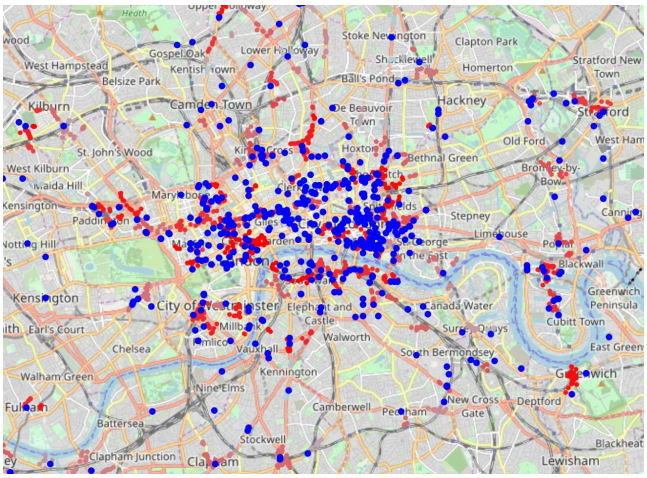


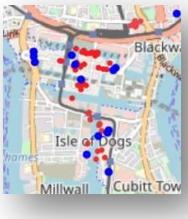

```
# Number of clusters in labels,
n_clusters = len(set(labels)) -
n_clusters
```

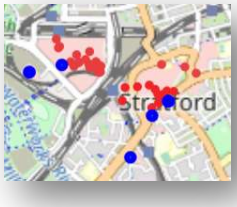
226



The potential commercial properties were added to the clusters to indicate which were near and should be considered.







Using the interactive map one can locate each address.

Discussion

The model I selected took into consideration the problem statement and business background. The important factors with that of customer location business premises and employees. during the visual representation of each data set it was clear that clustering would be useful but also regarding the density between data points. To maximize the benefit to this problem the differentiation between less dense and most dense would assist in solving this problem. Therefore, the model selected took into consideration not only clustering but also density.

Conclusion

From the results clustered customers and potential commercial properties in all areas of London are easily distinguishable. Priority would be given to those combinations in or nearest low-income groups/boroughs as per the choropleth map.

The next process would be to apply these results to a feasibility study as well as business plan to further narrow results based on the monthly budget as well as square meters of the potential commercial building required.

Thank you for taking the time to read through my work!