**Finding the best U.K. Postal Code in which to open a Restaurant Supply Store**

Max Allos

February 8th, 2021

1. **Introduction**
   1. **Background**

A friend of mine who runs a leading Restaurant Supply Store has found out that I am studying data science and has asked for help in trying to determine which Postal Code in the U.K. he should open his new store. As I am a man of the people, I sat down with him to discuss the problem and its requirements.

This is a multi-faceted problem whose solution very much varies from person to person. For this reason, I have decided to put the onus on my friend to outline various requirements that need to be met:

* It needs to be located in an area with a high density of restaurants.
* There should be heavy footfall of wealthy clients around the area.
* The number of COVID cases in the area should be taken into account.
* All assumptions should be confirmed by means of modelling and testing.

1. **Data Acquisition and Cleaning**

To address the main problem of finding a U.K. postcode in which to open the supply shop, I will need a list of U.K. postcodes which is very easy to get hold of: <https://en.wikipedia.org/wiki/List_of_postcode_areas_in_the_United_Kingdom_by_population>

In order to find the locations with a high density of restaurants we will use the Foursquare API: [https://api.foursquare.com](https://api.foursquare.com/)

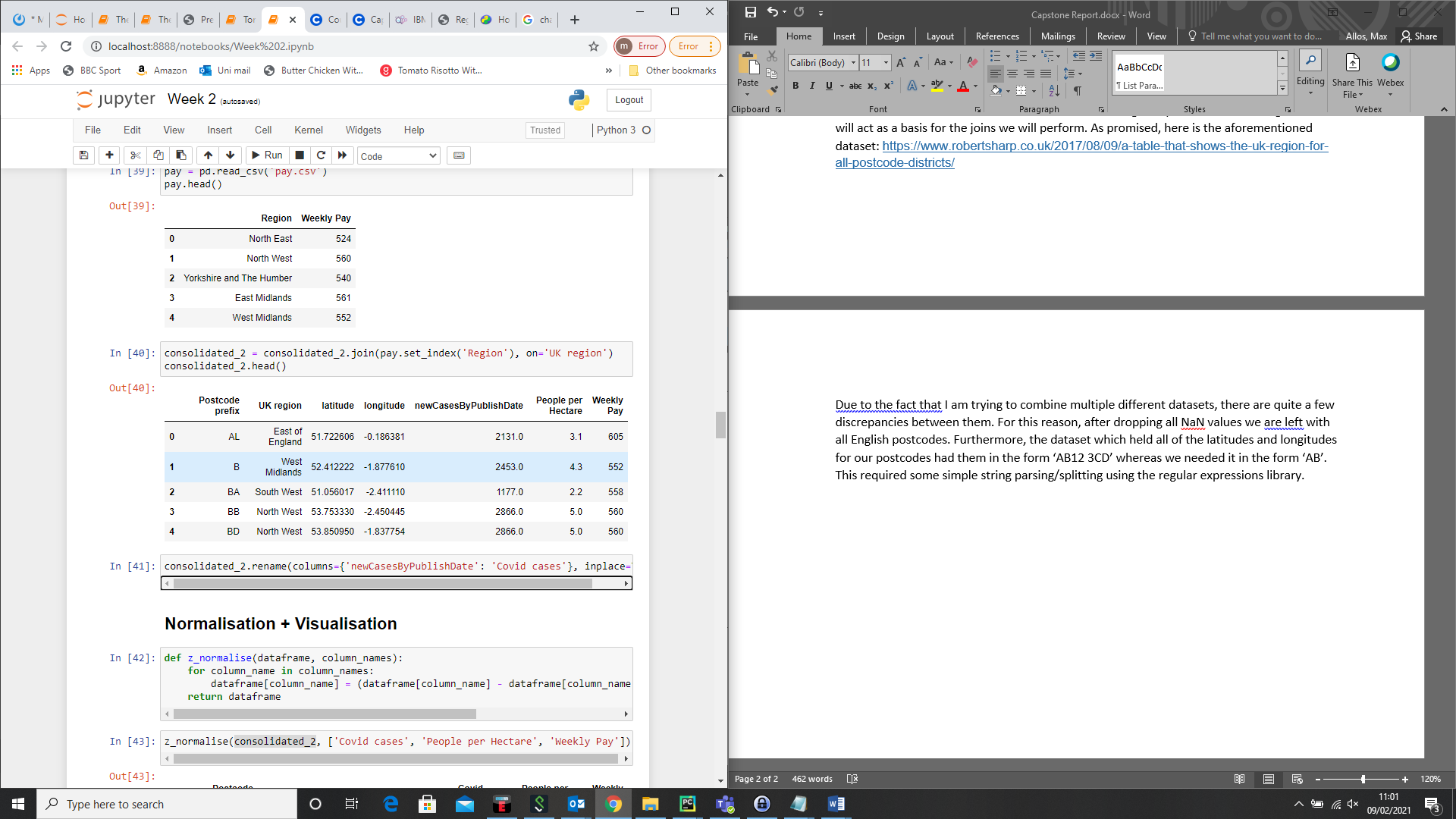
The second requirement is two-fold, it requires us to find an area of heavy footfall and wealthy clients. This will require two datasets, the first of which is the population density of each postcode and the second is the weekly income for each postcode. Unfortunately, neither of these are readily available but they are available per U.K. Region. Here are the two datasets we will use: <https://www.nomisweb.co.uk/census/2011/qs102ew> , <https://commonslibrary.parliament.uk/research-briefings/cbp-8456/>

Next, we will need to relate COVID cases to U.K. postcodes. As above, this information is not available at postcode level but is available at ‘region’ level so we have the following dataset: <https://coronavirus.data.gov.uk/details/download>

As you may have noticed, we now have a list of postcodes with a lot of ‘per region’ data. To address this issue we will need to find a dataset relating U.K. postcodes to U.K. regions. This will act as a basis for the joins we will perform. As promised, here is the aforementioned dataset: <https://www.robertsharp.co.uk/2017/08/09/a-table-that-shows-the-uk-region-for-all-postcode-districts/>

Due to the fact that I am trying to combine multiple different datasets, there are quite a few discrepancies between them. For this reason, after dropping all NaN values we are left with all English postcodes. Furthermore, the dataset which held all of the latitudes and longitudes for our postcodes had them in the form ‘AB12 3CD’ whereas we needed it in the form ‘AB’. This required some simple string parsing/splitting using the regular expressions library.

After performing a join on all of the relevant tables, we are left with the following (this is just showing the first four rows):



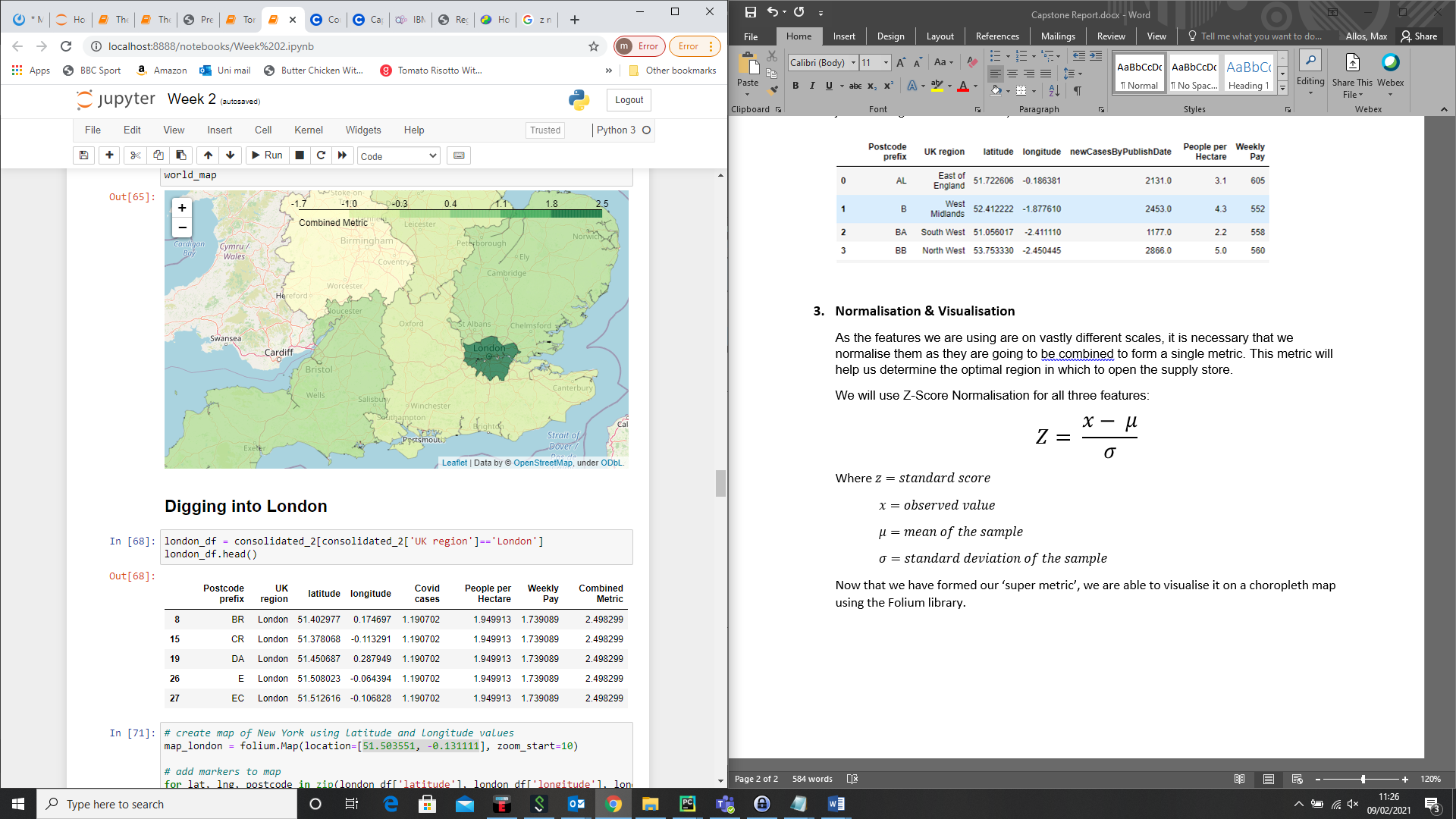
1. **Normalisation & Visualisation**

As the features we are using are on vastly different scales, it is necessary that we normalise them as they are going to be combined to form a single metric. This metric will help us determine the optimal region in which to open the supply store.

We will use Z-Score Normalisation for all three features:

Where

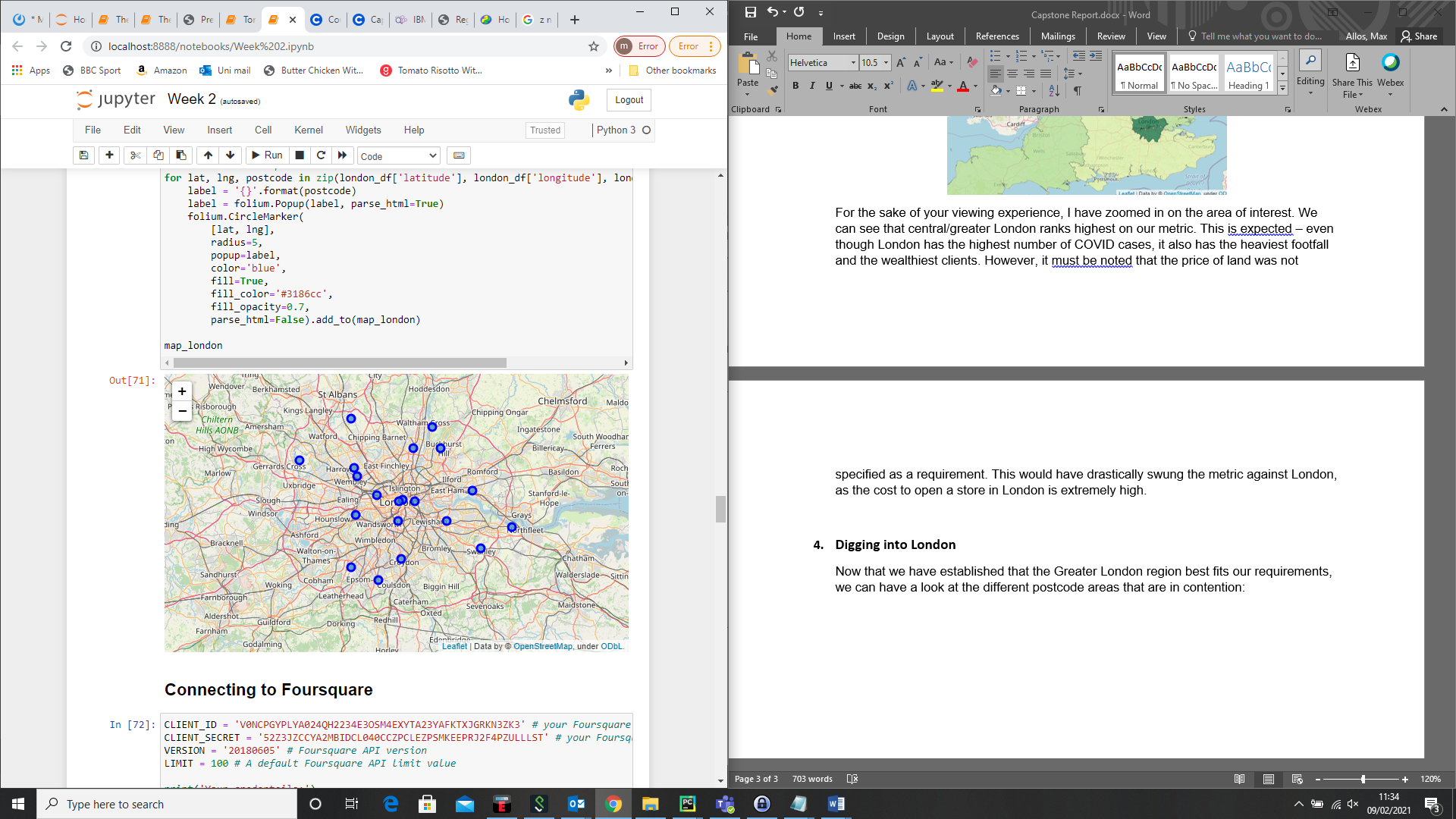
Now that we have formed our ‘super metric’, we are able to visualise it on a choropleth map using the Folium library.



For the sake of your viewing experience, I have zoomed in on the area of interest. We can see that central/greater London ranks highest on our metric. This is expected – even though London has the highest number of COVID cases, it also has the heaviest footfall and the wealthiest clients. However, it must be noted that the price of land was not specified as a requirement. This would have drastically swung the metric against London, as the cost to open a store in London is extremely high.

1. **Digging into London**

Now that we have established that the Greater London region best fits our requirements, we can have a look at the different postcode areas that are in contention:

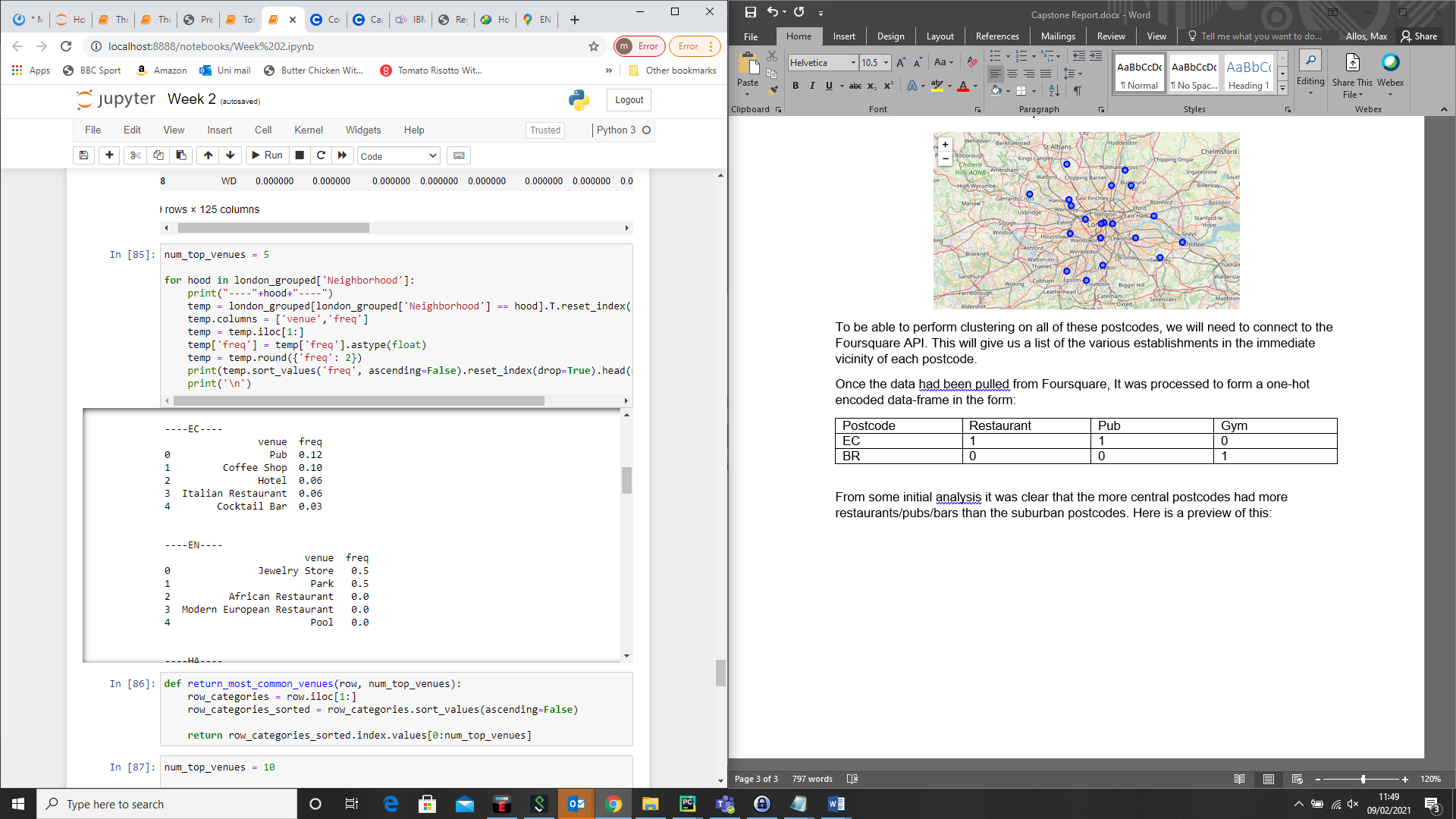


To be able to perform clustering on all of these postcodes, we will need to connect to the Foursquare API. This will give us a list of the various establishments in the immediate vicinity of each postcode.

Once the data had been pulled from Foursquare, It was processed to form a one-hot encoded data-frame in the form:

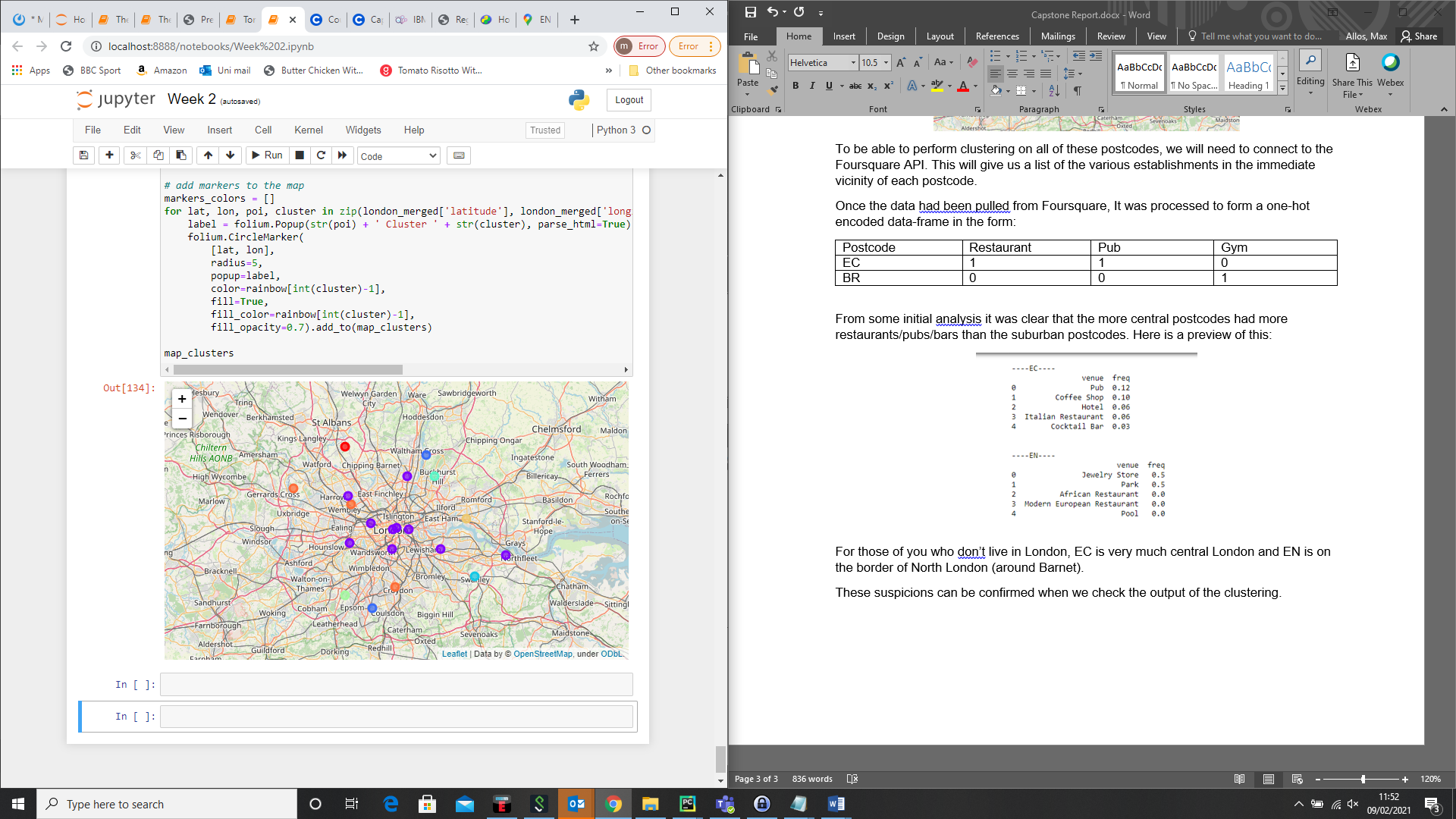
|  |  |  |  |
| --- | --- | --- | --- |
| Postcode | Restaurant | Pub | Gym |
| EC | 1 | 1 | 0 |
| BR | 0 | 0 | 1 |

From some initial analysis it was clear that the more central postcodes had more restaurants/pubs/bars than the suburban postcodes. Here is a preview of this:



For those of you who don’t live in London, EC is very much central London and EN is on the border of North London (around Barnet).

These suspicions can be confirmed when we check the output of the clustering.



From this, we can clearly see the trend of central postcodes grouped into the same cluster. Thus, can conclude that this would be the optimal area to open a restaurant supply store given the requirements stated at the start of this report.