Capstone Project - The Battle of Neighborhoods

By Mohammed Shah

Introduction to the problem

As you may already know, child obesity is a serious concern in most western counties and particularly in the U.K. The U.K government released a report stating that 1 in 5 children in reception year (4-5 year olds) were obese. As children grow older, the rates of obesity increases. By year 6 (10-11 year olds), over a third of children were classed as overweight or obese. The obesity figures are around 20%.

> **Childhood obesity is more prevalent in London than England overall. In 2018/19, some 23.2% of children in Year 6 were considered obese in London, compared to 20.2% in England.** - *Trust for London*

For this project, I will be looking at the boroughs of Greater London in the U.K. I will be aiming to answer the question:

Is there a correlation between average income in a Greater London borough, the level of child obesity and the types and frequency of venues in the borough?

I will be doing this for public health UK to see if the local authorities should limit number of business license's it gives or to promote to certain types of venues in order to improve child health in their boroughs.

Who will this information be useful to?

There are 32 boroughs in Greater London with a population of 8.92 million people.

If I can prove a correlation between average income, child obesity levels and the number of unhealthy venues in a borough:

1. The local councils can use this information to limit the number of licences it grants to unhealthy venues in the borough, like fast food outlets and increase the spending on healthy venues like gyms and outdoor spaces like parks.

2. The local health services could also use the information to better educate the children in the boroughs with the highest density of unhealthy venues, on healthy eating and life style choices.

This will allow the local councils and health services to improve the health of the children in its borough and improve the future health and wellbeing of the children, thus saving the councils and health services vast sums of money by tackling childhood obesity at an early stage before any of future long term health and employment issues start to have a serious impact on their wellbeing.

A map of Greater London

Is there any correlation between Average Income in a borough and the levels of childhood obesity?

Let’s get some data

First we will need some economic data about the average income for each Greater London borough. I got the data from https://data.london.gov.uk/dataset/earnings-place-residence-borough (https://data.london.gov.uk/dataset/earnings-place-residence-borough) . This file contains the weekly average income per borough from the years 2002 to 2019. We are only interested in the most recent data, the 2019 data. I downloaded a xlsx file on to my local storage for convenience.

	Code	Area	2002	Unnamed: 3	2003	Unnamed: 5	2004	Unnamed: 7	2005	Unnamed: 9	2006	Unnamed: 11	2007	Unnamed: 13	2008	Unname
			Pay		Pay		Pay		Pay		Pay		Pay		Pay	
0	NaN	NaN	conf %	conf %	conf %	conf %	conf %	conf %	conf							
			(£)		(£)		(£)		(£)		(£)		(£)		(£)	
1	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	Na
City of																
2	00AA	!	!	!	!	#	#	#	#	#	#	#	#	#		
London																
Barking																
3	00AB	and	383.3	5.5	354.9	6.5	360.1	5.2	375.1	5.1	417.6	6	446.7	6.4	449.9	6
Dagenham																

4	00AC	Barnet	427.4	5.1	450.1	5	453.3	5.6	442.3	5.3	466.1	5.7	460	5.6	502.2

As you can see the data contains a lot of junk, lets clean up the data. We only need the columns 'Area', which contains the name of the boroughs and '2019', which contains the average weekly income.

	Borough	Average Weekly Income
0	Barking and Dagenham	472.9
1	Barnet	536.6
2	Bexley	550.2
3	Brent	524.6
4	Bromley	641.3

```
'income' dataframe contains 32 rows/boroughs
```

As you can see we've got all 32 Greater London Boroughs and the data is in a usable form now.

Now let’s get the childhood obesity data.

I got the childhood obesity data from https://www.trustforlondon.org.uk/data/child-obesity/ (https://www.trustforlondon.org.uk/data/child-obesity/) I downloaded a csv of the data on to my local storage for convenience. Let’s have a look at the uncleaned data

	Area	ONS-code	Proportion of obese children in Year 6 (2008/09)	Proportion of obese children in Year 6 (2018/19)	Percentage points change between Area 2008/09 and 2018/19
0	England	E92000001	18.30%	20.20%	1.90%
1	London	E12000007	21.30%	23.20%	1.90%

2	Barking and Dagenham	E09000002	24.10%	29.60%	5.50%
3	Barnet	E09000003	18.30%	19.30%	1%
4	Bexley	E09000004	21.50%	22.70%	1.20%

Now I will clean the data. We are only interested in the **'Area'** and **'Proportion of obese children in Year 6 (2018/19)'** columns.

	Borough	Proportion obese
0	Barking and Dagenham	29.6
1	Barnet	19.3
2	Bexley	22.7
3	Brent	26.0
4	Bromley	17.1

```
'obesity' dataframe contains 32 rows/boroughs
```

As you can see we've got all 32 Greater London Boroughs and the data is in a usable form now.

Now I will merge the 'income' and 'obesity' dataframes into a new dataframe called **'lon_in_ob'**

	Borough	Proportion obese	Average Weekly Income
0	Barking and Dagenham	29.6	472.9
1	Barnet	19.3	536.6

2	Bexley	22.7	550.2
3	Brent	26.0	524.6
4	Bromley	17.1	641.3

```
The 'lon_in_ob' dataframe contains 32 rows/boroughs
```

Now let’s analyse this data.

First I will plot some Choropleth maps to show the income and childhood obesity levels for each of the 32 Greater London boroughs.

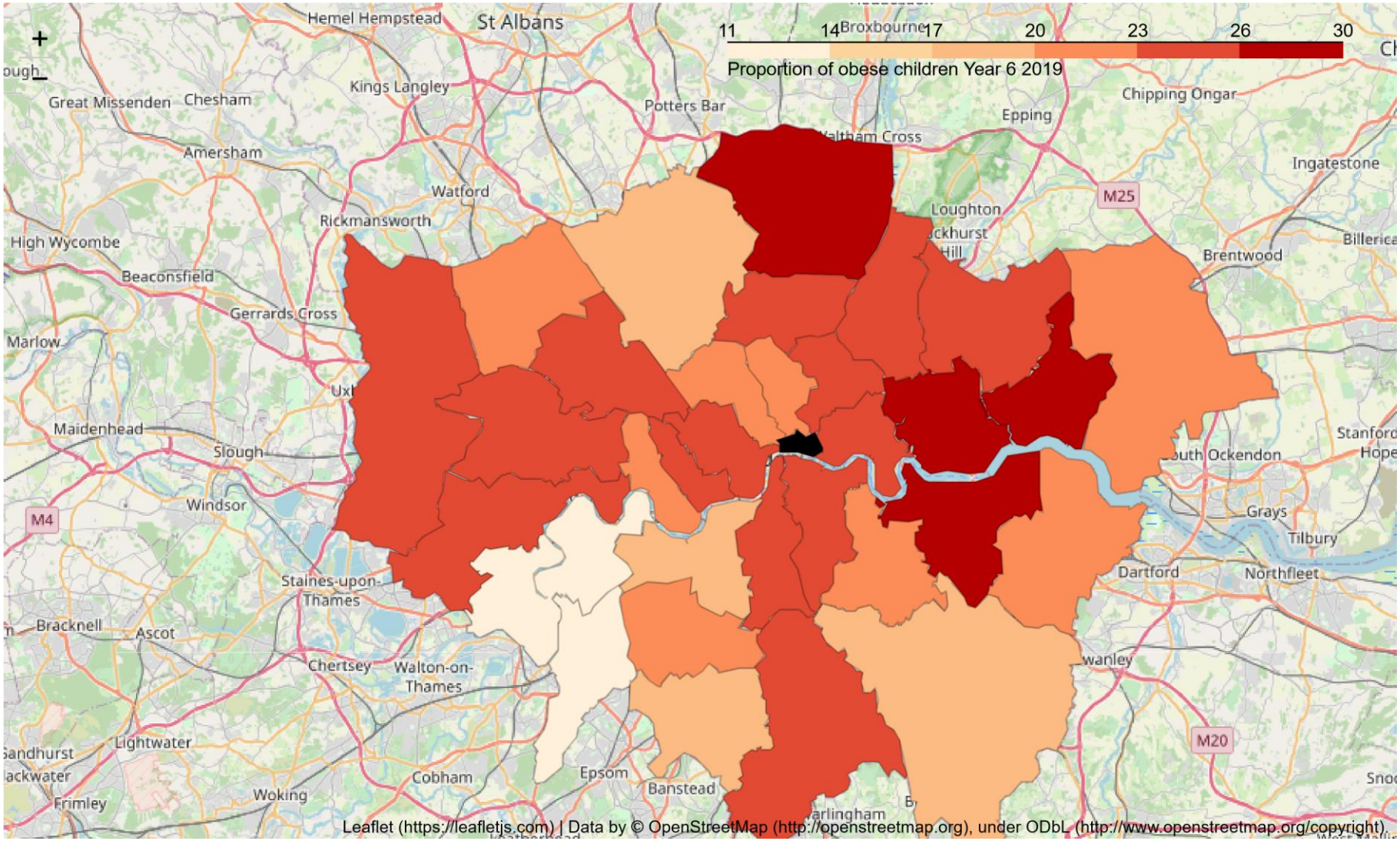
I got the geoJSON data for the boundary coordinates for the Greater London boroughs from https://skgrange.github.io/www/data/london_boroughs.json (https://skgrange.github.io/www/data/london_boroughs.json) and downloaded it to local storage.

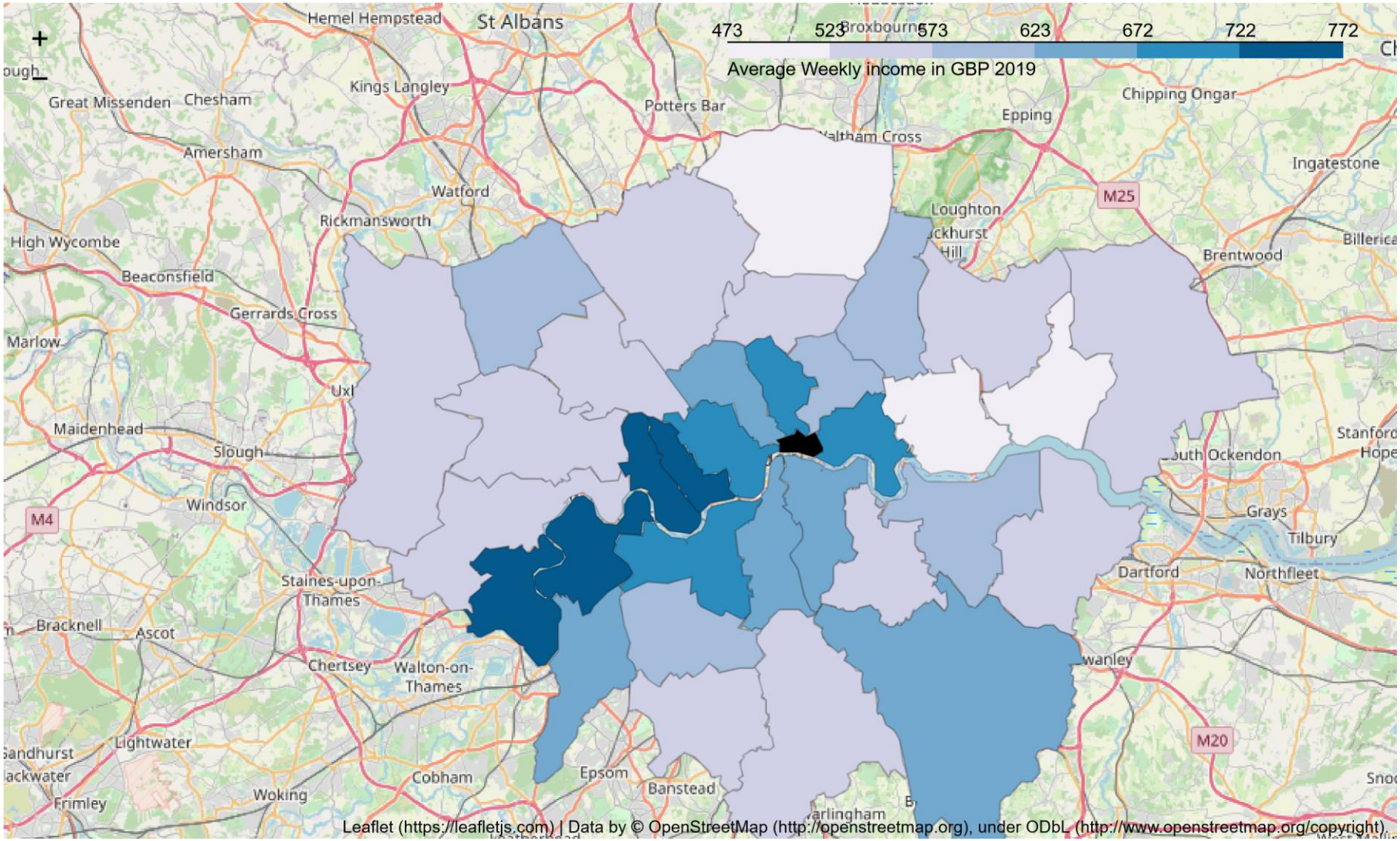
```
The geographical coordinates of London are 51.5073219, -0.1276474.
```

Now I will create 2 maps.

- **lon_ob_map** will be a Choropleth map of the level of childhood (year 6) obesity for each greater London borough.
- **lon_in_map** will be a Choropleth map of the level of average weekly income for each greater London borough.

Now let’s display the maps


11
14
17
20
23
26
30
Proportion of obese children Year 6 2019
Hemel Hempstead
St Albans
Broxbourne
Kings Langley
Potters Bar
Great Missenden
Chesham
Amersham
Watford
Rickmansworth
Chipping Ongar
Epping
Ingatestone
M25
Loughton
Brentwood
High Wycombe
Beaconsfield
Gerrards Cross
Marlow
Maidenhead
Slough
Windsor
M4
Staines-upon-Thames
Bracknell
Ascot
Chertsey
Walton-on-Thames
Grays
Tilbury
Dartford
Northfleet
M20
Lightwater
Woking
Cobham
Epsom
Banstead
Frimley
Leaflet (https://leafletjs.com) | Data by © OpenStreetMap (http://openstreetmap.org), under ODbL (http://www.openstreetmap.org/copyright).


473
523
573
623
672
722
772
Average Weekly income in GBP 2019
Hemel Hempstead
St Albans
Broxbourne
Great Missenden
Chesham
Kings Langley
Potters Bar
Chipping Ongar
Epping
Amersham
Ingatestone
Watford
M25
Rickmansworth
Loughton
High Wycombe
Brentwood
Beaconsfield
Gerrards Cross
Marlow
Uxl
Maidenhead
Slough
Windsor
M4
Grays
Tilbury
Staines-upon-Thames
Dartford
Northfleet
Bracknell
Ascot
Chertsey
Walton-on-Thames
M20
Lightwater
Cobham
Epsom
Banstead
Woking
Frimley
Leaflet (https://leafletjs.com) | Data by © OpenStreetMap (http://openstreetmap.org), under ODbL (http://www.openstreetmap.org/copyright).

Analysis of Choropleth maps

As you can see from the above maps there seems to be some correlation between income and childhood obesity.

If you notice the 2 boroughs with the lowest proportion of year 6 children that are obese **(Kingston upon Thames and Richmond upon Thames)** are 2 of the wealthiest boroughs based on average weekly income.

Also notice that the 3 poorest boroughs **(Enfield, Barking, Newham and Barking and Dagenham)** are also the 3 boroughs with the highest proportion of obese year 6 children. (see table below)

NB: There is no data available for the 'city of London' as is not a London borough (the black region at the centre of the maps)

```
Mean Proportion of obese year 6 children in Greater london boroughs is 22.71 %
Mean Average Weekly Income in Greater london boroughs is £ 598.11
```

```
The 2 wealthiest boroughs by average weekly income
```

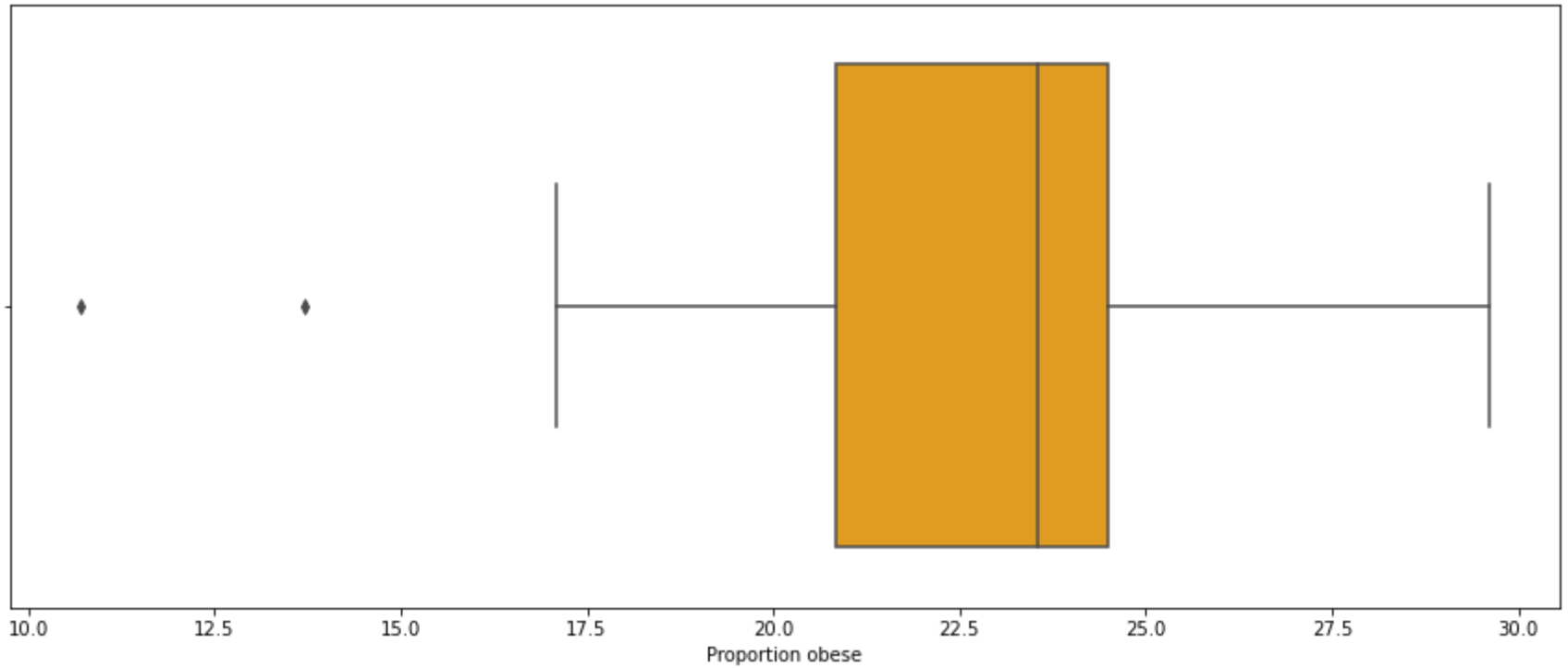
	Borough	Proportion obese	Average Weekly Income
25	Richmond upon Thames	10.7	734.4
19	Kingston upon Thames	13.7	623.0

These 2 boroughs have significantly lower than average Proportion of obese year 6 children in Greater london boroughs and a significantly higher than average weekly income.

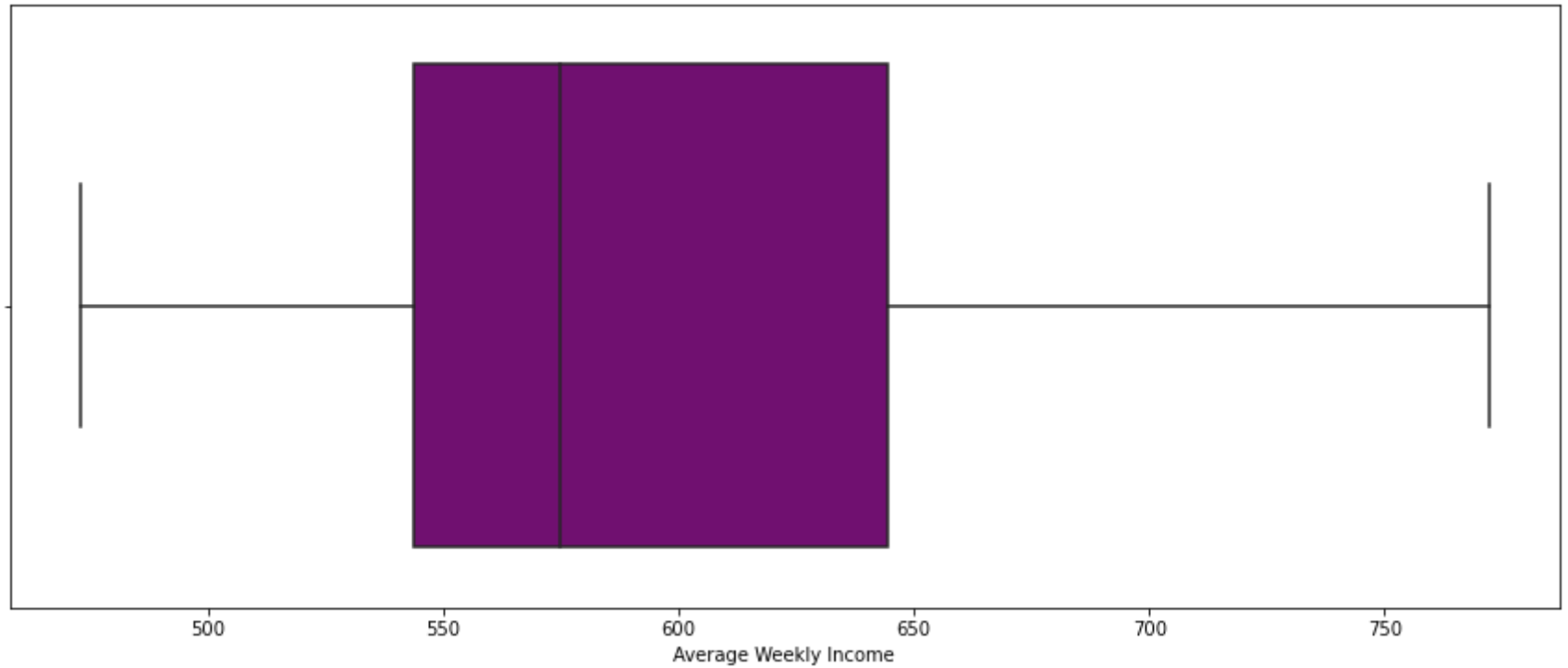
```
The 3 poorest boroughs by Average Weekly Income
```

	Borough	Proportion obese	Average Weekly Income
0	Barking and Dagenham	29.6	472.9
8	Enfield	27.2	482.5
23	Newham	27.7	517.1

Whereas these 2 boroughs have a higher than average Proportion of obese year 6 children in Greater London boroughs and a significantly lower than average weekly income.

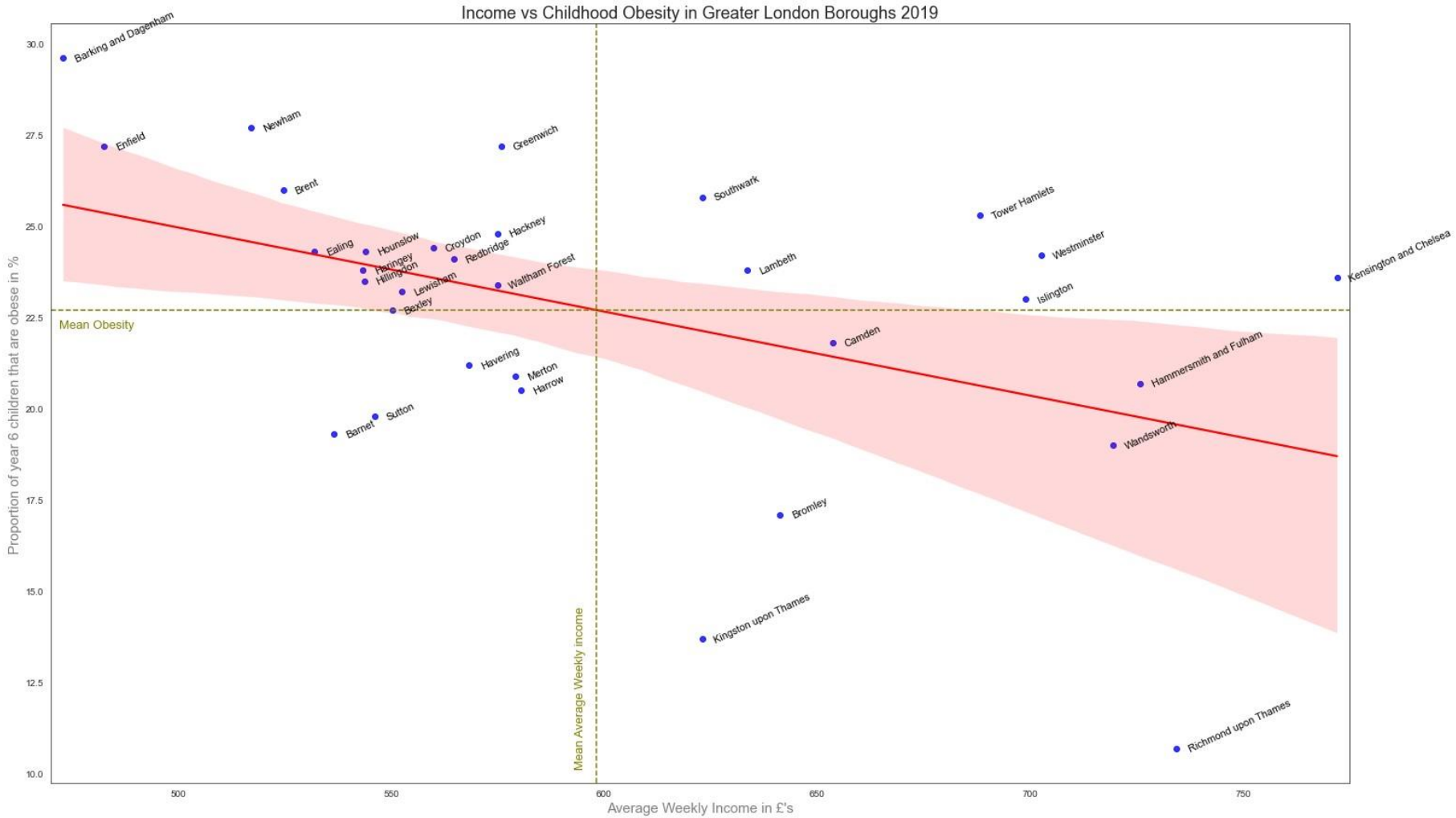


From the boxplot you can see that there are 2 outliers, namely **Kingston upon Thames** and **Richmond upon Thames** which are more than 1.5 times below the interquartile range. 50% of the boroughs have between 21% and 24.5% proportion of year 6 children that are obese in their boroughs.



There are no outliers in the Average Weekly Income. There is quite a big spread between the minimum and maximum Average Weekly Income, with 50% of boroughs have an average weekly income of between **£545** and **£645**.

I will use a scatter plot with linear regression to see how strong the correlation is.



Analysis

From the plot above we can make the following deductions:

- **15** boroughs are **lower** than mean average weekly income and **higher** than the mean proportion of obese children *(top left quarter)*
- **6** boroughs are **higher** than mean average weekly income and **higher** than the mean proportion of obese children *(top right quarter)*
- **5** boroughs are **lower** than mean average weekly income and **lower** than the mean proportion of obese children (bottom left quarter)
- **6** boroughs are **higher** than mean average weekly income and **lower** than the mean proportion of obese children *(bottom right quarter)*

This means that if a child lives in a borough which has lower than the mean average weekly income, they are **15/5 = 3** times more likely to be obese than not.

If the child lives in a borough which has higher than the mean average weekly income, they are **6/6 = 1** times more likely to be obese than not, i.e. the child is just as likely to be obese as they are not.

20 out of 32 boroughs have an average income **less** than the mean average. Out of these, **15** have higher proportions of obese children than the mean average, that is to say **75%** of them.

12 out of 32 boroughs have an average income higher than the mean average. Out of these, **6** have higher proportions of obese children than the mean average, that is to say **50%** of them.

Notice that there seems to be a quite good correlation between income and childhood obesity when the income is lower than the mean average. Here **10** boroughs fall within the **95%** zone of probability and fit quite closely to the regression line. Whereas when the average income is higher than the mean average, only **3** boroughs are within the **95%** zone.

Looking at the spread of points, it is unlikely that I will be able to find a polynomial regression model that wouldn't overfit the data.

Conclusion

From the data we can see that there is some correlation between the average weekly income of a borough and the proportion of obese year 6 children in that borough. This correlation is more obvious in the boroughs that have a lower average weekly income. In the boroughs that have a higher than average weekly income, there seems to be little correlation, with the same number of boroughs that have a higher than mean proportion of obese children as there are boroughs that have a lower than mean proportion of obese children.

At the extremes there is correlation, **Barking and Dagenham** which is the poorest borough and has the highest proportion of childhood obesity.

At the other extreme there is **Richmond upon Thames**, which is the second wealthiest borough and the lowest proportion of childhood obesity.

But then you have the borough of **Kensington and Chelsea** which has the highest average weekly income, but also one of the highest proportion of childhood obesity.

There doesn’t seem to very much correlation when we look at the boroughs that have an average weekly income of between **£525** and **£725**, which is where most of the boroughs are. Here for example the boroughs of **Wandsworth** and **Barnet** have significantly different average weekly income but similar proportions of childhood obesity.

In conclusion, the data shows that there is some correlation between average income and childhood obesity but it isn't strong enough to say that this correlation proves a definitive link between them, i.e. we cannot say for certain that the wealthier the borough, is the lower the proportion of year 6 children will be obese.

Is there any correlation between Average Income in a borough and the levels of childhood obesity and the venues in the borough?

Now lets examine if there is link between the venues and the proportion of childhood obesity in the borough. For this I will be using the Foursquare API to pull venue data from greater London postcode coordinates.

UK Post Code Information.

Postcodes in the UK are comprised of **Postcode Area** + **Postcode District**

Postcode Area – this is the largest geographical unit of the postcode. Each one comprises one or two alpha characters generally chosen to be a mnemonic of the area eg MK for Milton Keynes, SO for Southampton. There are currently 124 Postcode areas including Guernsey (GY) Jersey (JE) and the Isle of Man (IM)

Postcode District – Each postcode area is divided into a number of districts which are represented by the numerical portion of each part of the postcode. These numbers range from 0 to 99 eg MK42. In London a further alpha character is used to divide some districts into sub divisions eg EC1A.

There are 20 Post Code Area's in Greater London

First lets the list of Postcode Area codes for Greater London. I will get this data from https://www.robertsharp.co.uk/2017/08/09/a-table-that-shows-the-uk-regionfor-all-postcode-districts/ (https://www.robertsharp.co.uk/2017/08/09/a-table-that-shows-the-uk-region-for-all-postcode-districts/)

	Postcode prefix	Postcode district	UK region
0	AB	Aberdeen	Scotland
1	AL	St. Albans	East of England
2	B	Birmingham	West Midlands
3	BA	Bath	South West
4	BB	Blackburn	North West

Let’s clean up the data and obtain the data we need.

We are only interested in the Postcode prefix for Greater London post codes.

	Postcode prefix	Postcode district	UK region
0	BR	Bromley	Greater London
1	CR	Croydon	Greater London
2	DA	Dartford	Greater London
3	E	London	Greater London
4	EC	London	Greater London

To confirm we have all 20 Greater London Postcode prefixes: The number of rows in dataframe is 20 **Now**

Now we have to add **Postcode District** to the Postcode Area prefixes. All UK post Codes have prefixes in the range from **0 to 99**. lets generate them.

N.B Not all postcodes will have 99 Postcode District's this is just a dataframe of all possible Greater London postcodes

	Post Code
1995	WD95
1996	WD96

1997	WD97
1998	WD98
1999	WD99

Now lets get the geographical coordinated for uk postcodes. I have got them from https://www.freemaptools.com/download/full-postcodes/ukpostcodes.zip (https://www.freemaptools.com/download/full-postcodes/ukpostcodes.zip) and downloaded to local storage for convenience.

	Post Code	latitude	longitude
0	AB10	57.13514	-2.11731
1	AB11	57.13875	-2.09089
2	AB12	57.10100	-2.11060
3	AB13	57.10801	-2.23776
4	AB14	57.10076	-2.27073

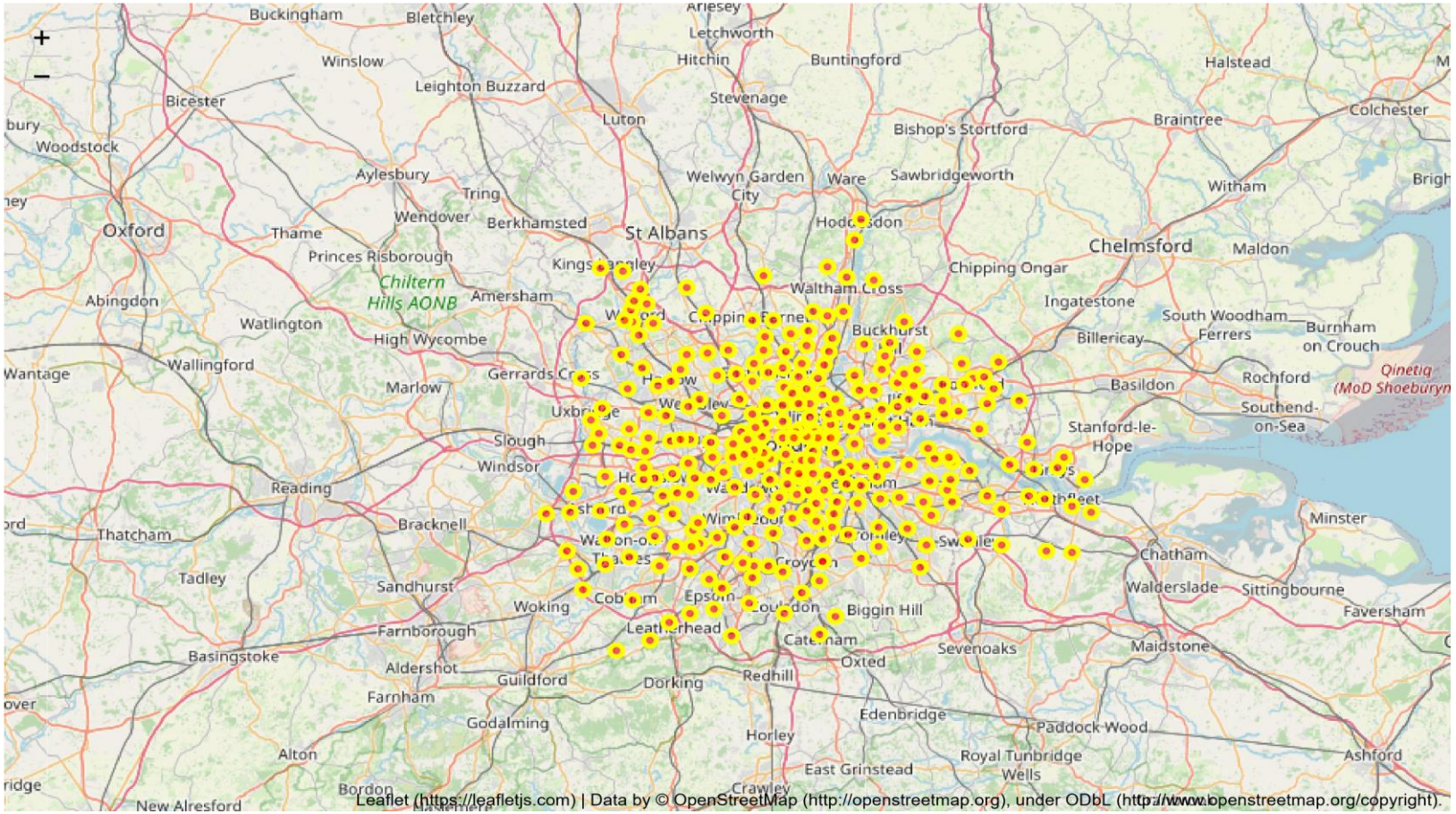
```
The total number of UK Postcode Area + Postcode Districts is: 2975
```

Now let's find the postcodes that are Greater London postcodes

	Post Code	latitude	longitude
0	BR1	51.41107	0.02192
1	BR2	51.38858	0.02237
2	BR3	51.40297	-0.03020
3	BR4	51.37559	-0.00695
4	BR5	51.38983	0.10436

```
The number of Greater London Post Codes is: 287
```

Now I will plot the postcodes on a map to verify that they are all within Greater London.



Let’s get some data from Foursquare

First I will examine if the number **'Food** and **'Athletics & Sports'** venues in an area have any correlation to the levels of childhood obesity in the borough.

I will use the Foursquare API to get 50 **'Food** and **'Athletics & Sports'** venues within a radius of 1000m for every postcode in the Greater London Area.

In the **Food** category, I will only be looking for the sub categories that are more likely to be linked to childhood obesity.

The sub categories I will be looking at are: **'Bakery', 'Burger Joint', 'Dessert Shop', 'Donut Shop', 'Fast Food Restaurant', 'Fish & Chips Shop', 'Fried Chicken Joint', 'Pizza Place', 'Snack Place'** and **'Wings Joint'**

In the **'Athletics & Sports'** category, I will only be looking for the sub categories that are more likely to be used by children.

The sub categories I will be looking at are: **'Badminton Court'**, **'Basketball Court'**, **'Boxing Gym'**, **'Gym Pool'**, **'Gymnastics Gym'**, **'Martial Arts Dojo'**, **'Track'**, **'Skate Park'**, **'Soccer Field'**, **'Tennis Court'**, **'Volleyball Court'**, **'Indoor Play Area'**, **'Park, ‘Playground and Recreation Center'**

I will use the Foursquare API and category codes from the Foursquare website to only search for venues that I believe have the biggest influence on childhood obesity.

Show the first 5 rows of the dataframe "london_venues" which contains the data we got from the Foursquare request

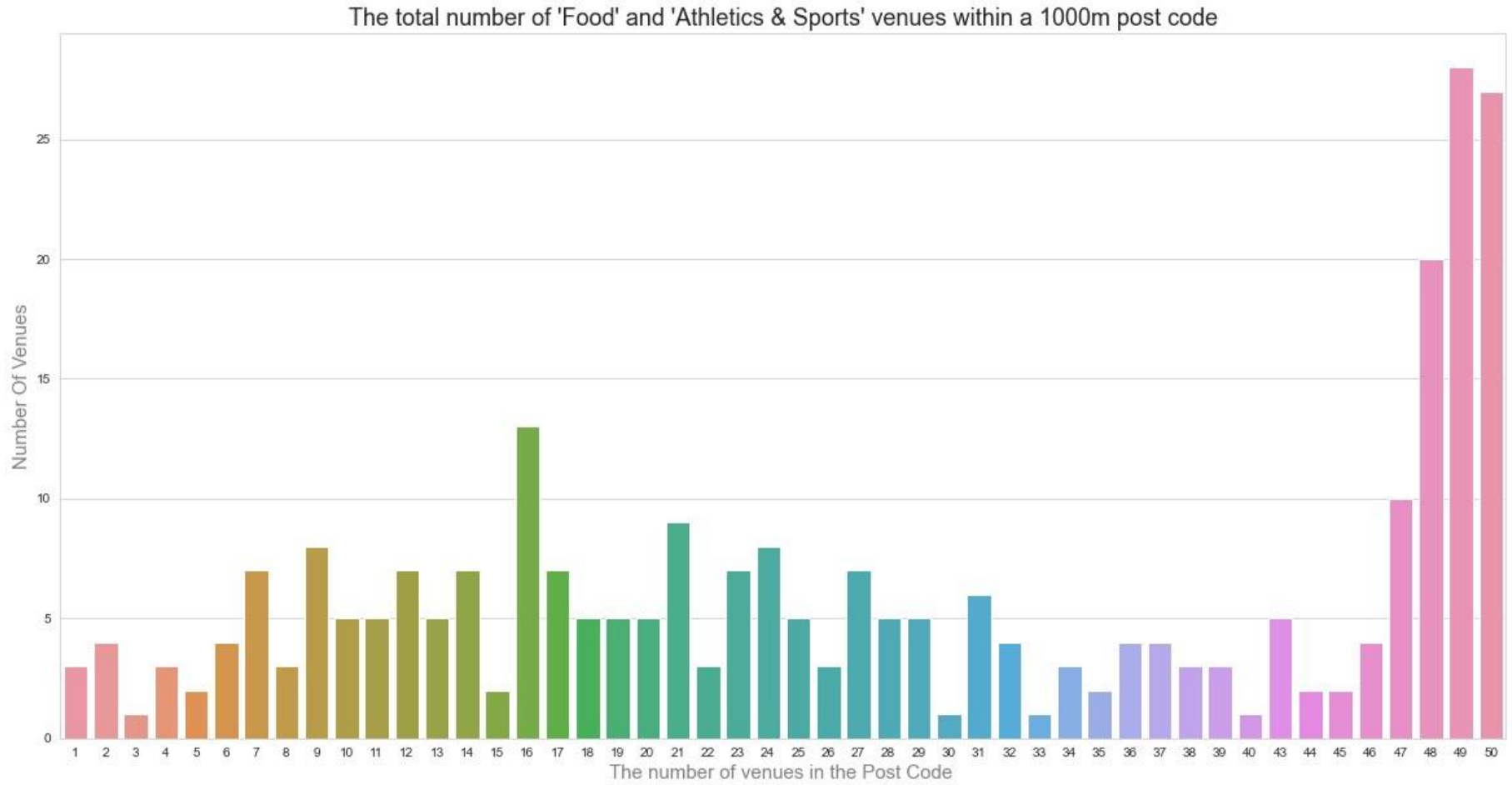
	Post Code	Post Code Latitude	Post Code Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	BR1	51.41107	0.02192	The Pantry	51.414253	0.020361	Bakery
1	BR1	51.41107	0.02192	Franco Manca	51.405992	0.016181	Pizza Place
2	BR1	51.41107	0.02192	McDonald's	51.403054	0.016430	Fast Food Restaurant
3	BR1	51.41107	0.02192	Five Guys	51.405580	0.015490	Fast Food Restaurant
4	BR1	51.41107	0.02192	KFC	51.402165	0.015907	Fast Food Restaurant

```
8325 venues retrived from Foursquare
For 283 Post Codes
```

Notice that we retrieved venues for 283 post codes not 287 post codes that we expected. This is because 4 post codes had no 'Unhealthy Food' **and** 'Athletics & Sports'** venues that we are interested in within a radius of 1000m

Notice also that we retrieved only 8325 venues and not 50 venues for every post code This is because some of the post codes are in area's with very few local **'Food'** and **'Athletics & Sports'** venues such as residential areas or industrial and business areas.

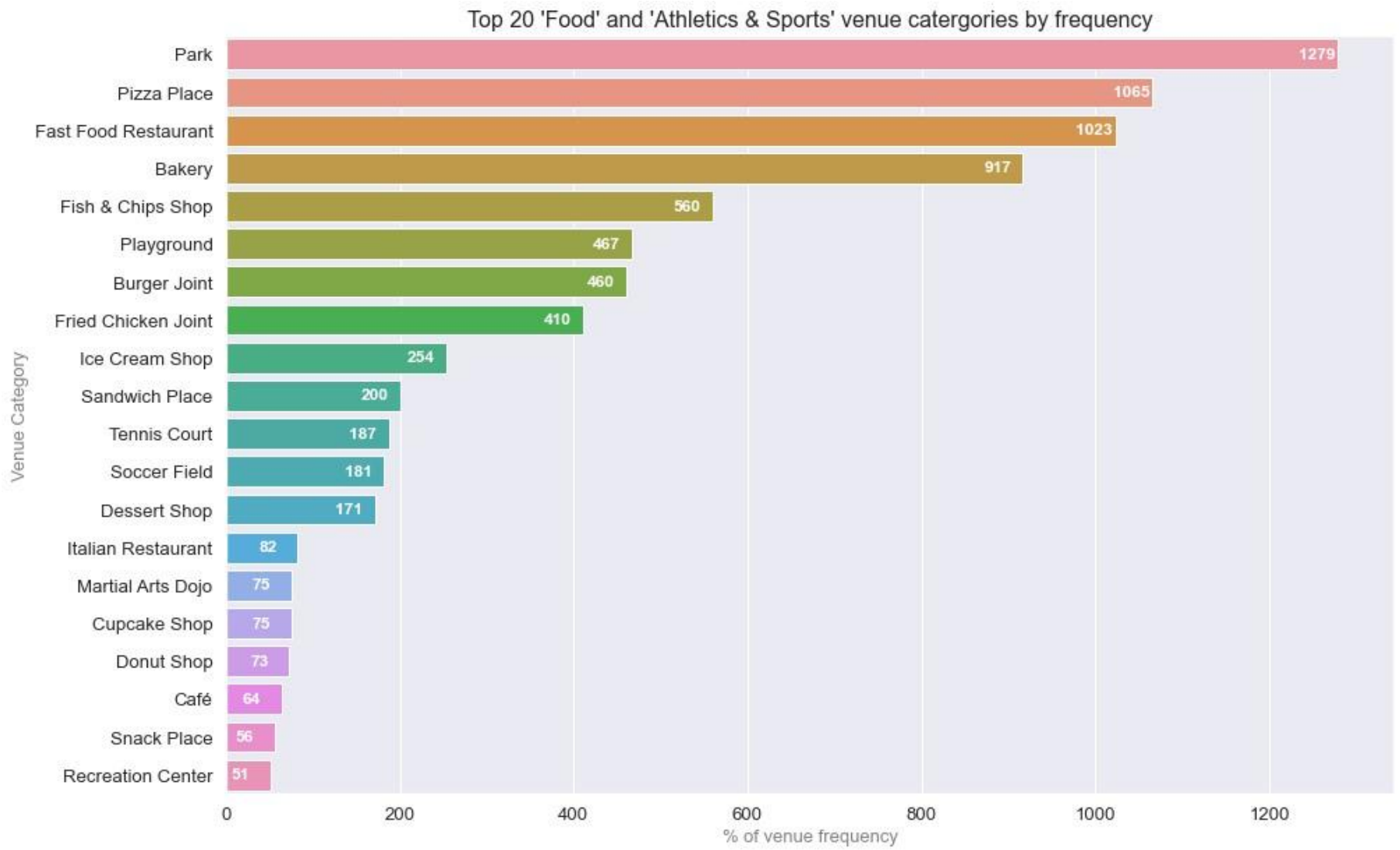
See bar graph below.



As you can see from the table above, not all post codes have 50 'Unhealthy Food or 'Athletics & Sports' within a 1000m radius.

Now let’s have a look at the frequency of venue categories from the data we retrieved from Foursquare.

I will get the top 20 venue categories by frequency and plot a bar graph of the data



As you can see, **Parks** are the most frequent venue in followed by **Pizza Places** and **Fast Food Restaurants**. **Playgrounds**, **Tennis Courts** and **Soccer Fields** are the 6th, 10th and 11th most frequent venues.

In the **top 20 there are 7650 venues**, out of these **2240** are **Athletics & Sports** venues **(29.28%)** and **5410** are **Unhealthy food** venues **(70.71)**

So just **over 2 thirds** venues are **Unhealthy food** venues and just **under 1 third** are **Athletics & Sports** venues.

Greater London doesn't seem like a very healthy city for Children.

Now I will use a Kmeans pipeline from sklearn to cluster the postcodes to see if I can find any insight.

I will group the post codes in to 3 clusters

```
283 labels generated
```

	Post Code	Cluster longitude	1st Most Common Labels	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	latitude Venue
0	BR1	51.41107	0.02192	0	Fast Food Restaurant	Park	Playground	Pizza Place	Bakery
1	BR2	51.38858	0.02237	2	Pizza Place	Park	Fish & Chips Shop	Sports Club	Tennis Court
2	BR3	51.40297	-0.03020	2	Park	Fish & Chips Shop	Fast Food Restaurant	Pizza Place	Bakery
3	BR4	51.37559	-0.00695	2	Park	Pizza Place	Bakery	Fish & Chips Shop	Soccer Field
4	BR5	51.38983	0.10436	2	Pizza Place	Bakery	Park	Fish & Chips Shop	Playground

```
Cluster 0 has 93 Post codes
Cluster 1 has 58 Post codes
Cluster 2 has 132 Post codes
```

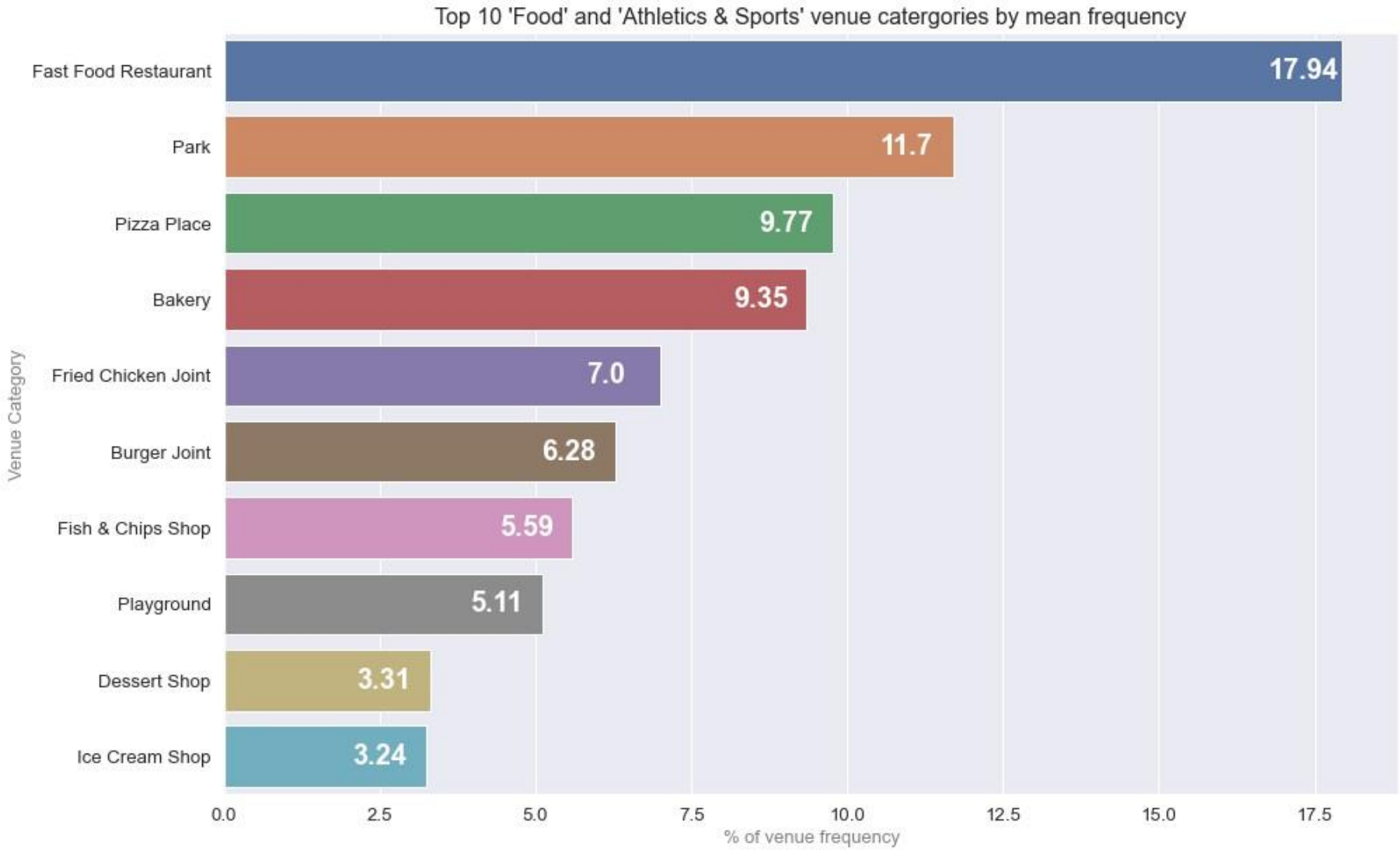
You can see that:

- **Cluster 0** has **93** Post codes which is **32.86%** of the post codes.
- **Cluster 1** has **58** Post codes which is **24.49%** of the post codes.
- **Cluster 2** has **132** Post codes which is **46.80%** of the post codes.

Let’s look at the Clusters in more detail

Cluster 0

	Post Code	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	BR1	Fast Food Restaurant	Park	Playground	Pizza Place	Bakery
19	DA3	Basketball Court	Soccer Field	Yoga Studio	Food & Drink Shop	Cupcake Shop
20	DA5	Tennis Court	Fast Food Restaurant	Cupcake Shop	Bakery	Park
21	DA6	Bakery	Fast Food Restaurant	Fish & Chips Shop	Dessert Shop	Pizza Place
24	DA9	Fast Food Restaurant	Playground	Cupcake Shop	Park	Sandwich Place


Top 10 'Food' and 'Athletics & Sports' venue catergories by mean frequency
Fast Food Restaurant
Park
Pizza Place
Bakery
Fried Chicken Joint
Burger Joint
Fish & Chips Shop
Playground
Dessert Shop
Ice Cream Shop
17.94
11.7
9.77
9.35
7.0
6.28
5.59
5.11
3.31
3.24
Venue Category
0.0
2.5
5.0
7.5
10.0
12.5
15.0
17.5
% of venue frequency

2 out of the top 10Athletic & sports related and **8 out of the top 10** are **Unhealthy food related**

Fast Food Restaurants are most frequent venue in this cluster accounting for **17.94%** of the venues in this cluster.

If we add up all the **Unhealthy Food venues** we get **62.48%**, so we can deduce **that at least 62.48%** of the venues within 1000m of the post codes are **Unhealthy Food venues** as we are only looking at the top 10 and not the whole cluster.

If we add all the **Athletic & sports related** venues up we get **16.81%**, so we can deduce **that at least 16.81%** of the venues within 1000m of the post codes are **Athletic & sports venues** that are suitable for children as we are only looking at the top 10 and not the whole cluster.

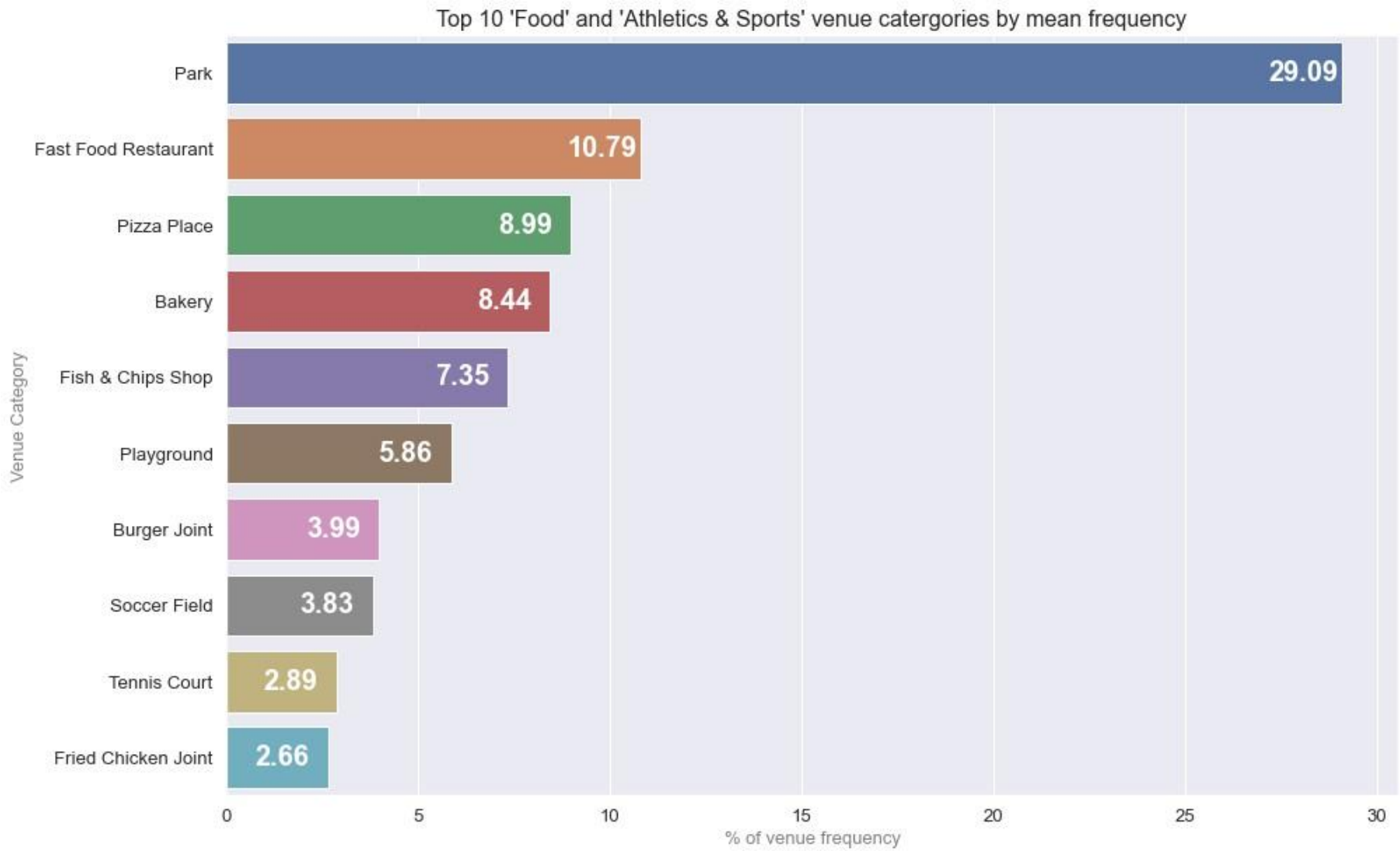
There seems to be less than a third of the number of sports venues and there does seem to be a lot of Unhealthy food venues in this cluster.

These seem like a quite healthy post codes for children so I will categorise this as the 'Unhealthy Cluster'

Cluster 1

	Post Code	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
8	CR0	Park	Fish & Chips Shop	Bakery	Pizza Place	Fast Food Restaurant
9	CR2	Park	Fish & Chips Shop	Bakery	Fast Food Restaurant	Boxing Gym
10	CR3	Park	Pizza Place	Fast Food Restaurant	Bakery	Snack Place
11	CR4	Park	Fast Food Restaurant	Fish & Chips Shop	Playground	Tennis Court
13	CR6	Soccer Field	Park	Fish & Chips Shop	Bakery	Yoga Studio

In this cluster we can see that most frequent venues are



6 out of the top 10Unhealthy Food related and **4 out of the top 10** are **Athletic & sports related**

Parks are the most frequent venue in this cluster accounting for **29.09%** of the venues in this cluster.

If we add up all the unhealthy **Unhealthy Food venues** we get **42.22%**, so we can deduce **that at least 42.22%** of the venues within 1000m of the post codes are **Unhealthy Food venues** as we are only looking at the top 10 and not the whole cluster.

If we add all the **Athletic & sports related** venues up we get **41.67%**, so we can deduce **that at least 41.67%** of the venues within 1000m of the post codes are **Athletic & sports venues** that are suitable for children as we are only looking at the top 10 and not the whole cluster.

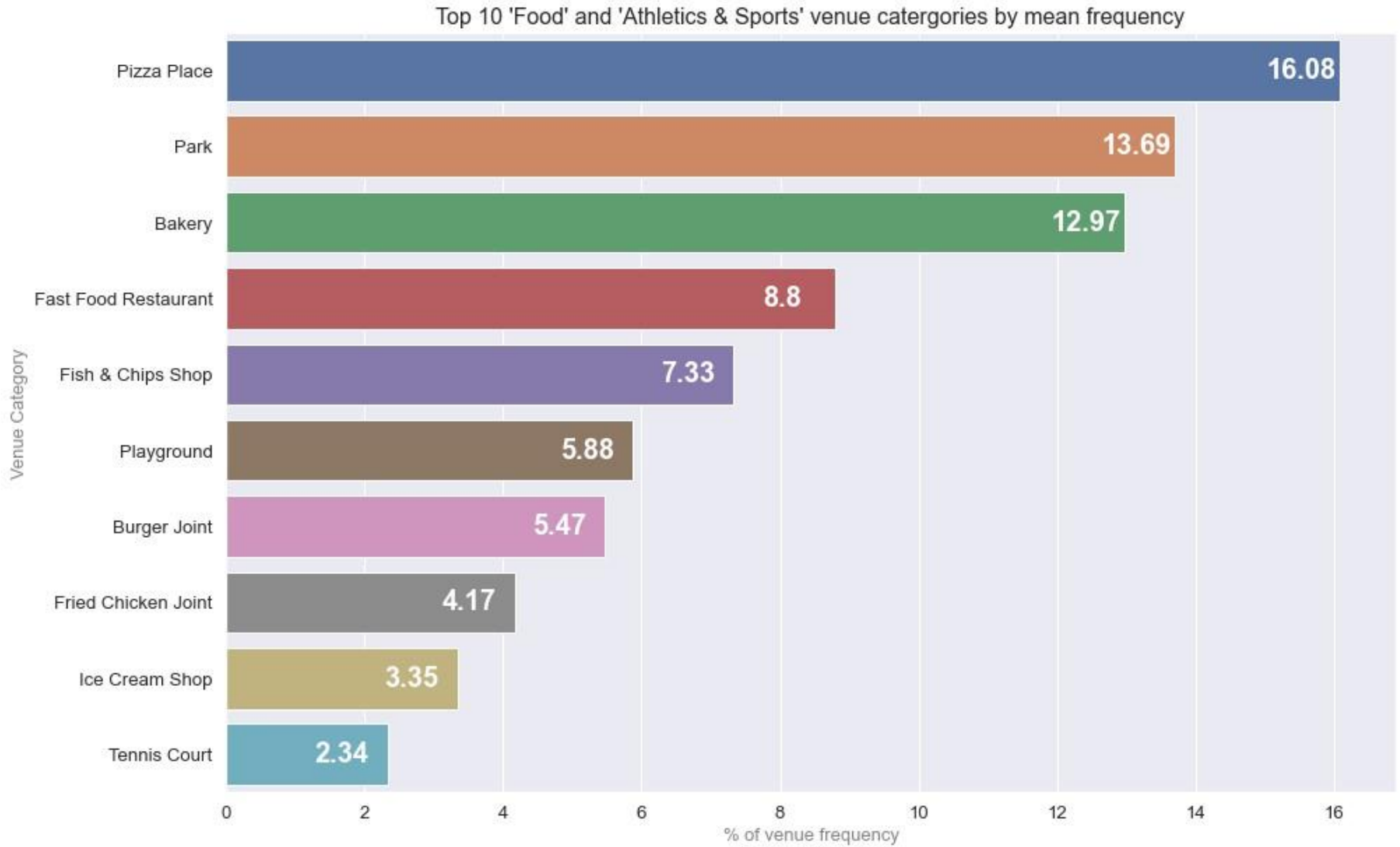
There seems to be roughly as many sport venues for children as there are unhealthy food venues in these postcode which.

These seem like quite healthy post codes so I will categorise this as the 'Healthy Cluster'

Cluster 2

	Post Code	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
1	BR2	Pizza Place	Park	Fish & Chips Shop	Sports Club	Tennis Court
2	BR3	Park	Fish & Chips Shop	Fast Food Restaurant	Pizza Place	Bakery
3	BR4	Park	Pizza Place	Bakery	Fish & Chips Shop	Soccer Field
4	BR5	Pizza Place	Bakery	Park	Fish & Chips Shop	Playground
5	BR6	Bakery	Fast Food Restaurant	Pizza Place	Sandwich Place	Café

In this cluster we can see that most frequent venues are



7 out of the top 10 Unhealthy food related and **3 out of the top 10** are **Athletic & sports related**.

Pizza Places most frequent venues in this cluster accounting for **16.08%** of all the venues in this cluster.

If we add up all the **Unhealthy Food venues** we get **58.17%**, so we can deduce **that at least 58.17%** of the venues within 1000m of the post codes are **Unhealthy food venues** as we are only looking at the top 10 and not the whole cluster.

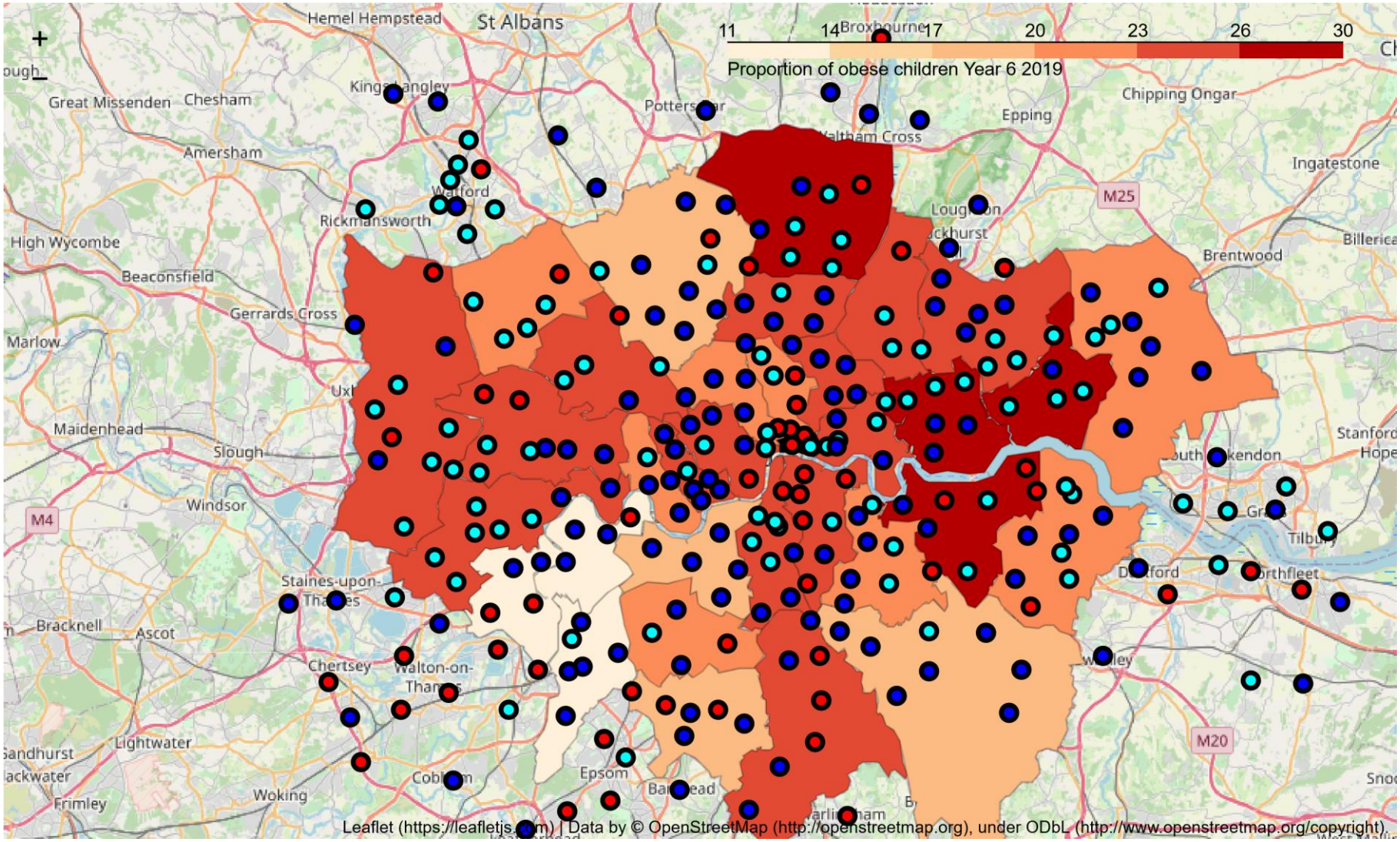
If we add all the **Athletic & sports related** venues up we get **21.97%**, so we can deduce **that at least 21.97%** of the venues within 1000m of the post codes are **Athletic & sports venues** that are suitable for children as we are only looking at the top 10 and not the whole cluster.

There seems to be less than half of the number of sports venues and there does Unhealty food venues in this cluster.

These seem like quite Unhealthy post codes so I will categorise this as the 'Moderately Unhealthy Cluster'

In this cluster we can see that most frequent venues are

Now I will plot the Clusters onto the Choropleth map for the proportion of obese children that I created earlier *(lon_ob_map)*



```
Cluster 0 are cyan Coloured Markers, the 'UNHEALTHY' Cluster

Cluster 1 are red  Coloured Markers, the 'HEALTHY' Cluster

Cluster 2 are blue Coloured Markers, the 'MODERATELY UNHEALTHY' Cluster
```

Analysis

We would have expected the 2 boroughs with the lowest proportion of childhood obesity **(Kingston upon Thames and Richmond upon Thames)** to have the most post codes that are in the **'Healthy'** Cluster (**Red Marker**), but the 2 boroughs contain between them:

- 1 **'Unhealthy'** marker (Cyan),
- only 3 **'Healthy'** markers (Red),
- 10 **'Moderately Unhealthy'** markers (Blue).

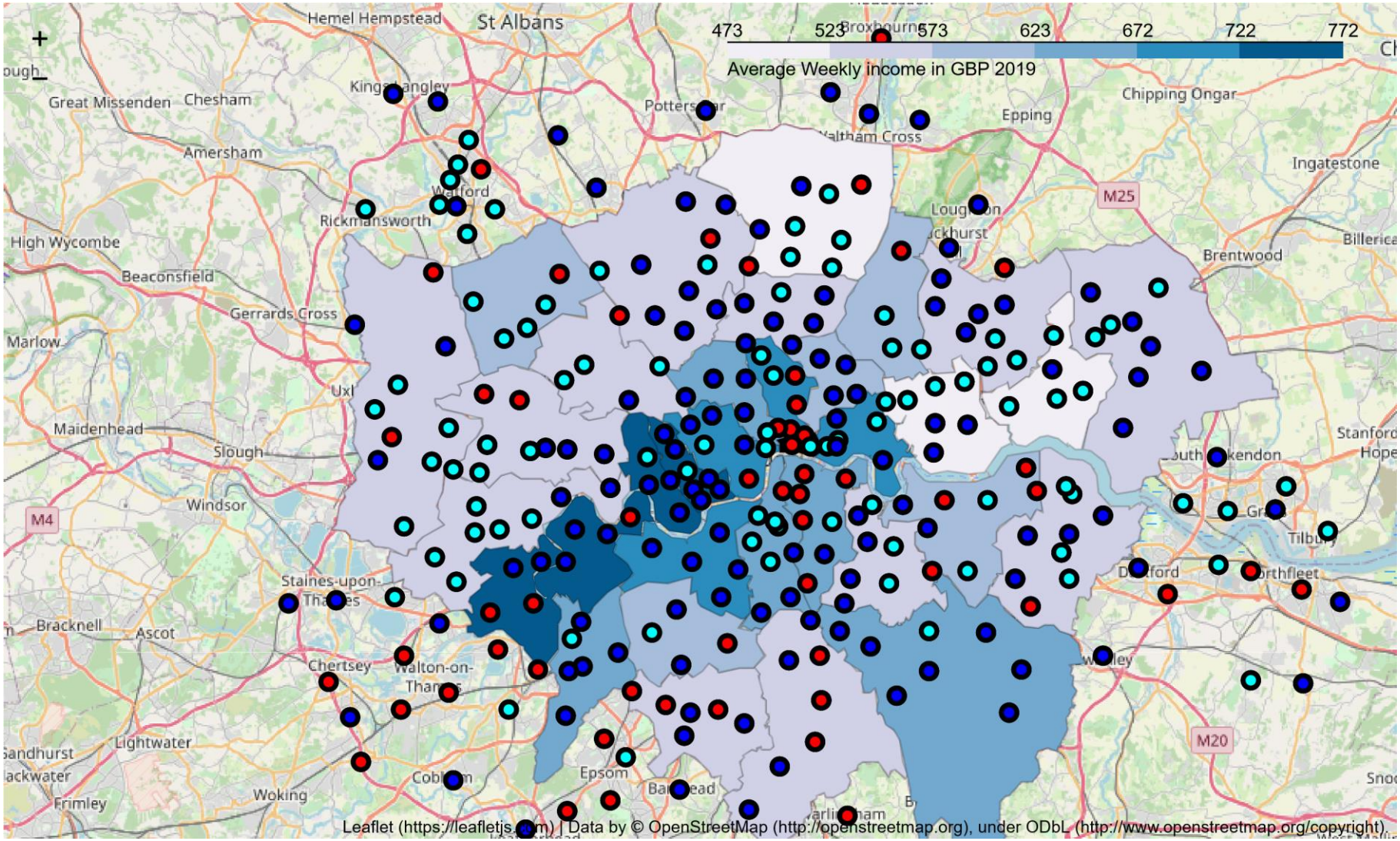
And we would have expected that the 3 boroughs with the highest proportion of childhood obesity **(Enfield, Barking, Newham and Barking and Dagenham)** would be contain mainly **'Unhealthy'** and **'Moderately Unhealthy'** postcodes (**Cyan Markers**), we find between them they contain: 16 **'Unhealthy'** markers

- (Cyan),
- 5 **'Healthy'** markers (Red),
- 7 **'Moderately Unhealthy'** markers (Blue).

As you can see there are indeed more ‘Unhealthy’ markers in these boroughs, but surprisingly there are more 'Healthy' markers than in the 2 boroughs with the least obesity levels.

You can also see that in the **south London boroughs** and the **central London boroughs**, there are a high concentration of red markers, but not a significantly lower proportion of obese children. In fact, these boroughs have similar obesity levels as the boroughs in **North West London** which contain very few 'healthy' Markers.

Now I will plot the Clusters onto the Choropleth map for the Average Weekly Income that I created earlier *(lon_in_map)*



Cluster 0 are cyan Coloured Markers, the 'UNHEALTHY' Cluster

Cluster 1 are red Coloured Markers, the 'HEALTHY' Cluster

```
Cluster 2 are blue Coloured Markers, the 'MODERATELY UNHEALTHY' Cluster
```

Analysis

Here again we see very little correlation between the average weekly income of a borough and the number of 'Healthy' markers.

We can see that 2 boroughs with some of the highest weekly income **(Kensington & Chelsea** and **Hammersmith & Fulham**) have no healthy markers.

Whereas some of the boroughs in south and south west London (**Croydon** and **Sutton**) have relatively low average income, but have quite a high proportion of 'Healthy' markers.

Conclusion

Form this study we can draw a few conclusions:

1. The average weekly income has a significant impact on childhood obesity when the average weekly income is low.

1. This impact is not proportionally inverse, the higher than average weekly income boroughs do not see a significant drop in childhood obesity, with some of the wealthiest borough having some of the highest proportion of obesity.

1. The number of unhealthy food venues and Athletic & Sports venues in a borough does not correlate very well with the weekly income or childhood obesity rates in the borough. We would have expected there to be more sports facilities in the richer boroughs and unhealthier food venues in the less wealthy boroughs.

1. We would have expected the boroughs with the highest childhood obesity levels to have the highest proportion of unhealthy food venues. This didn’t turn out to be the case, with the unhealthy markers quite evenly spread.
2. Conversely, we would have expected the boroughs with the highest proportion of childhood obesity to have the highest proportion of Unhealthy markers, but this wasn't the case.

So there must be some other factors which increase the likely hood of childhood obesity in the poorer boroughs, not just the food and sports venues in these boroughs. This could be many things like, housing, the parent's educational levels, ethnic makeup of the boroughs or the standard of schools in the borough.

I can conclude that solving the childhood obesity problem in Greater London is not a simple solution and requires looking at the problem from many different angles and that there are many factors involved, not just income.