# LONDON GCSE RESULTS

# BACKGROUND

- GCSEs, short for General Certificate of Secondary Education, are academic qualifications which are generally taken by pupils in secondary education (aged 15-16) in England, Wales and Northern Ireland. Young students usually gain a number of GCSEs in subjects such as English, mathematics and the sciences, which are marked from 9 to 1, with 9 being the highest grade and 1 being the lowest. The grades achieved by pupils can have a major influence on their lives, as they are seen as foundational education one must attain. The impact of good/bad GCSE results can be substantial, as these can effect which further education institutions pupils get accepted into, what universities they can attend and what jobs they can land. All these factors play an important role in the future life of pupils.

# PROBLEM

- This project will focus analysing what areas of London achieve high GCSE grades and which areas achieve low grades. We will use machine learning tools to try and find patterns to see which areas parents should send their kinds.

- Our goal is see if areas of high achievers can areas of low achievers have clear disparities when looking at the Foursquare location data. Is the presence of certain facilities/venues in an area an important factor for the average GCSE grade in that area?

# DATA

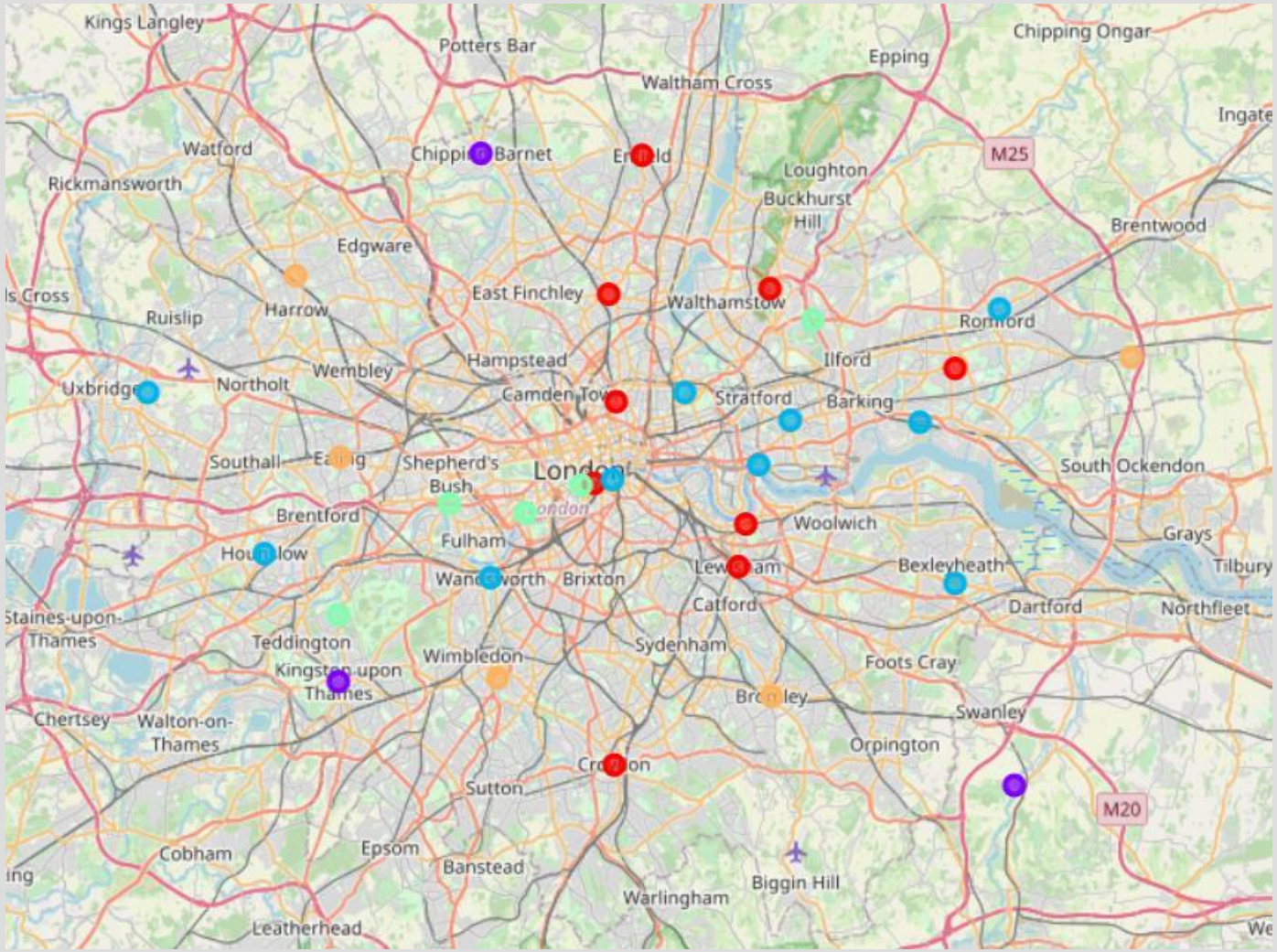
- The data we will be using is taken from the UK department for education, a link to this data is available here: https://data.london.gov.uk/download/gcse-results-by-borough/12a95356-81d3-49d6-8a13-e41b62f5e5c4/gcse-results.csv.
- This data gives us information about the "Attainment" of pupils. This variable gives us the average pupils grade across eight subjects for each London Borough (Area). We will use the most recent data entries (2018/2019) for the different London Boroughs to see how these compare. We will also limit our analysis to coeducational schools only.

# METHODOLOGY

- Firstly, we need to acquire the GCSE results data and load it to a pandas data frame. Next we do basic data cleaning processes such as removing irrelevant data entries/columns. After which we use folium to visualise where the different London Boroughs are located on a map of London. Then we can use the foursquare API to attain geospatial data on the different venues located in the area. Using this data we proceed with cluster analysis to find out how different clusters of London Boroughs compare with respect to GCSE results and surrounding venues. The last step is to summarise our findings in the concluding section.

# RESULTS

- Created 5 clusters for different groups of GCSE results, as shown on map which was created using folium:



# RESULTS:

- Clusters 1 (purple) and 3 (green) have got the highest average attainments (GCSE grades), these locations are mostly located on the outskirts of London, but the ones located nearer the centre are in the western part of the city. Clusters 0 (red) and 2 (blue) have got the lowest GCSE results, these are mostly located in east London and closer to the centre of London.

# DISCUSSION

- If we look at data entry 28 (Sutton) which has the highest average attainment score, we will find that the most common venues at that location are places which we would associate with high culture and prestige, such as historic sights, tennis courts and yoga studios. We also find that Sutton is located on the outskirts of the city. Comparing this with Lewisham (data entry 22) which has the lowest grades, we find that the venues surrounding that location are retailers, coffee shops/cafes and public transport facilities, places with one can associate with a busy city. Additionally Lewisham is located near the centre of the city, this further adds to the idea that the further from the centre the higher the grades.

# CONCLUSION

- A final comment would be to use the data to access the level of urbanization of the area to access how well a student would perform in their GCSEs. Highly urbanized areas seem to have a negative impact on the GCSE grades. Of course there are several other factors at play which effect the GCSE results of pupils that this basic analysis has not taken into account, and it would be ludicrous to think otherwise.