Varied Access to NHS Dentists in England: A Clustering Project

Capstone Project for IBM Data Science Professional Certificate

Peter Lueckert
February 2021

Contents

List of Tables and Figures

1. Introduction

Unfortunately, oral health and access to NHS dental practices is not the same across England. According to Public Health England, the percentage of 5-year olds with visually obvious dental decay varies dramatically from district to district. For example, in Blackburn with Darwen 50.9% of 5-year olds showed obvious tooth decay, compared to Hastings with only 1.1% in 2018/19 (Public Health England, 2021).

There are also stark differences when it comes to the number of dental practices per population. In Ashfield for instance are only 0.7 dental practices per 10,000 population whereas Westminster, London has 4.1 dental practices per 10,000 population.

A report by QualityWatch, a research programme by Nuffield Trust and Health Foundation, also found that people living in deprived areas are more likely to suffer from poor dental health (Appleby & Reed, 2017). Therefore, having access to affordable NHS dentists is especially vital in areas with high deprivation levels.

According to a 2020 Report from the National Audit Office on Dentistry in England, the total funding for the NHS dental sector was 2.9 billion in 2018-19. In real terms, funding decreased by 4% from 2014-15 to 2018-19. Furthermore, the number of working NHS dentists per 10,000 population in the UK was 5.3 in 2018, which is fewer than in Germany, France and Italy (National Audit Office, 2020). This clearly highlights a lack of access to affordable dental healthcare in the UK.

In this Capstone project of the IBM Data Science Course, a clustering algorithm will be used combining factors such as tooth decay in children, deprivation levels and number of NHS Dental practices by inhabitants and surface area for each Local Authority District in England. The aim is to highlight the geographic differences in dental health and access to affordable dental care across England, allowing the identification of Local Authority Districts that require serious attention.

The findings in this report should be considered by the British Dental Association as well as the different Local Authorities, in an attempt to provide fairer and more affordable dental health in all areas of England. Improving dental health is improving health in general.

2. Data

The following data was used for the clustering project to highlight the geographic differences of NHS dental care across England:

Geographic Data

This data from December 2019 contains the Longitude, Latitude and surface area in square kilometre for every Local Authority District in England from the Office for National Statistics (Office for National Statistics, 2021). The coordinates will be used to plot the different clusters and Local Authority Districts onto a visual map of England. The surface area will be used to calculate the number of NHS dental practices per square kilometre to give an indication of geographic access to NHS dental practices in an area. The assumption is that the fewer dental practices a District has per square kilometre, the further the patients have to travel to been seen by a dentist. How the practices are geographically distributed within a District has not been taken into consideration in the project.

Population Data

Population data by Local Authority District in England, dated from December 2019, was taken from Office for National Statistics (Office for National Statistics, 2021). It was used to calculate the number of dental practices per 10,000 population for each Local Authority District, which allowed for a like for like

comparison among the District with regards to access to NHS dental practices. The higher the number of practices per population the better the access to affordable dental care.

Tooth Decay Data

The percentage of 5-year olds with visually obvious dental decay in 2018-19 from Public Health England was used as an indicator for the level of dental health in each District (Public Health England, 2021). Tooth decay in children is an important factor as it forecasts the future impact on the dental health system. Children that have poor dental health are more likely to have dental problems later in life, putting a burden on the health system. Since the values are expressed as percentages, the level of tooth decay in children can be easily compared among the Districts.

Deprivation Data

The Index for Multiple Deprivation (IMD) scores in 2019 obtained from the government statistics on gov.uk were used to indicate the level of deprivation in each Local Authority District. The IMD combines seven different domains to create the overall IMD score and includes Income, Employment, Education, Health, Crime, Barriers to Housing & services and Living Environment. The higher the IMD score in a District, the more deprived the District is relative to other Districts (Ministry of Housing, Communities & Local Government, 2019). The level of deprivation in an area is important to consider as it is more vital to ensure access to affordable dental services in more deprived areas. People on lower income are more reliant on NHS dental services that are reasonably priced (and in some cases even free) compared to private sector services.

Dental Practice Data

The number of NHS dental practices per District in 2020 was taken from Digital NHS (Digital NHS, 2021). Divided by population and surface area, it served as a measure for access to NHS dental services. The more practices per population and square kilometre the better the access to affordable dental care. Occupancy of the dental practices has not been taken into consideration in this project. It was assumed that all practices are fully occupied. Another assumption was that the number of dentists per practice is roughly the same for all practices.

Data from Foursquare was not used in this project as it did not differentiate between NHS and private Dental Practices. Establishing which practices are NHS was essential to ensure that access to affordable dental care could be measured.

3. Methodology

The methodology section describes exploratory analysis of the variables, a principal component analysis and a K-Means clustering model that was carried out in Python. A feature set, consisting of the variables, tooth decay in children, deprivation levels and number of NHS Dental practices by population and square kilometre by District, was fed into the clustering algorithm to form distinct clusters. K-Means is a very popular clustering technique and belongs to the category of unsupervised Machine Learning. This technique partitions the data into clusters using an optimisation algorithm to find the largest distance between clusters and smallest distance within clusters. The number of clusters, K, will need to be defined by the user.

3.1. Exploratory Data Analysis

In this section the input variables, i.e. Features, are visually explored to get a sense for the variations within the data, which Districts in the country performed best and worst when it comes to dental health,

deprivation etc. Furthermore, the variables are assessed on multicollinearity to see if any of the variables correlate with each other.

Percentage of Obvious Tooth Decay in 5-Year-Old

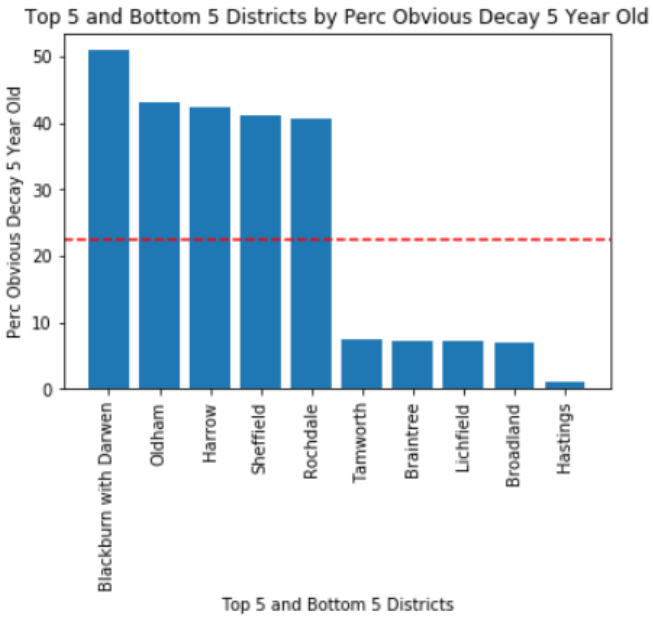

Figure 1 Top 5 and Bottom 5 Districts by Percentage of Obvious Tooth Decay in 5-Year-Old

The visual shows how large the gap is between the worst and best district when it comes to cavities in children. Blackburn with Darwen shows a percentage of visual decay in 5 year olds of 50.9%, whilst Hastings only has 1.06%. The percentage of 5-year-old with visual tooth decay for England overall is 22.5%.

Number of Dental Practices per 10,000 Population

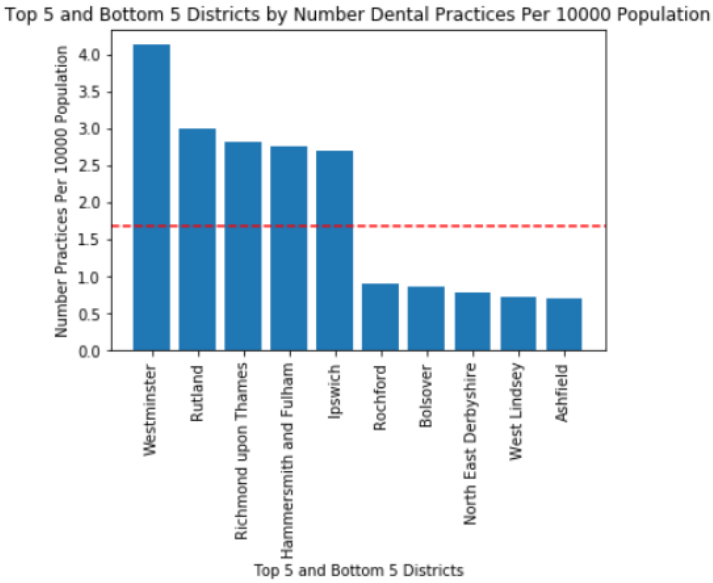

Figure 2 Top 5 and Bottom 5 Districts by Number of Dental Practices per 10,000 Population

The visual above shows again a very large gap between the worst and best district with regards to access to NHS dental care within the Districts. Westminster is clearly leading here with 4.1 NHS dental practices per 10,000 population, whilst Ashfield only has 0.7 practices for the same number of people. The average for England overall is 1.69.

Number of Dental Practices per Square Kilometre

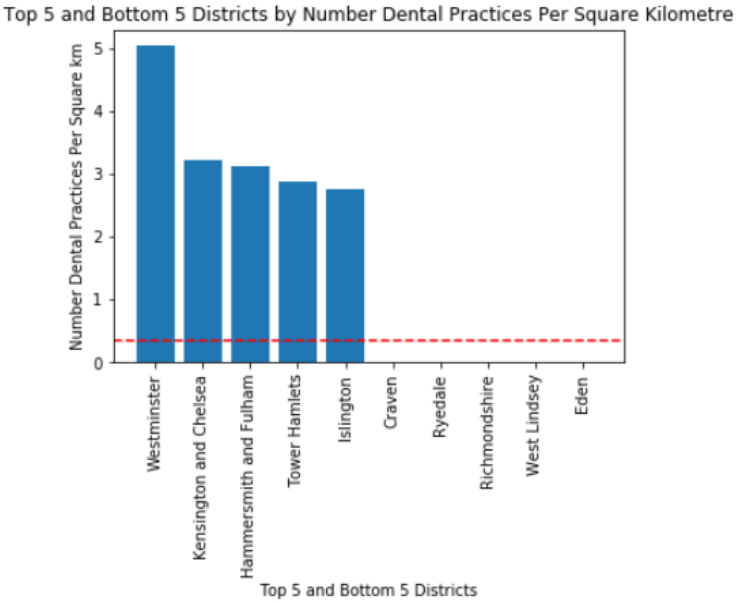

Figure 3 Top 5 and Bottom 5 Districts by Number of Dental Practices per Square Kilometre

The chart reveals once again a large gap between the worst and best district regarding geographical access to NHS dental care within the Districts. Westminster came again top with 5 NHS dental practices per square kilometre. Eden on the other end has 0.005 surgeries per square kilometre. This seems intuitive since Westminster has a much greater population density than Eden. The average number of NHS dental practices per square kilometre is 0.35.

Level of Deprivation

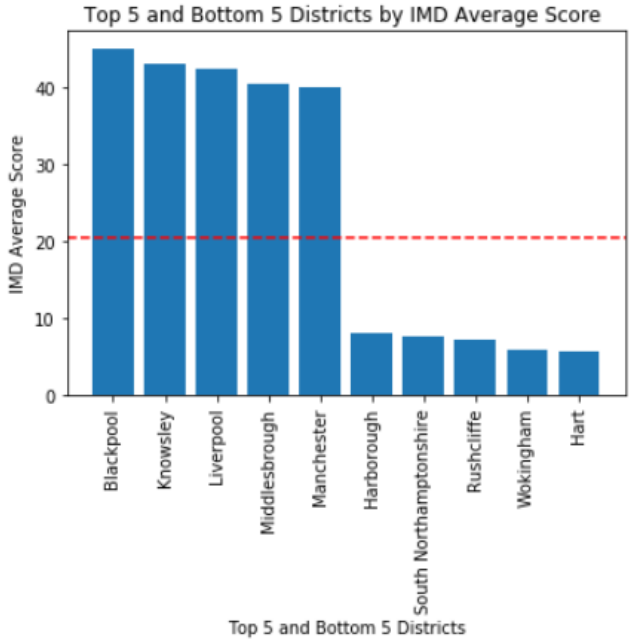

Figure 4 Top 5 and Bottom 5 Districts by IMD Average Score

The column chart above highlights how much the Districts vary within England with regards to deprivation. Blackpool appears to be the most deprived District with an average IMD score of 45, whilst Hart is the least deprived with 5.5. The average score for England is 20.4.

Correlations

If two or more variables (or features) are strongly correlated, these features would then have a higher impact on the clustering model which may lead to unintended results (i.e. clusters are skewed towards those variables).

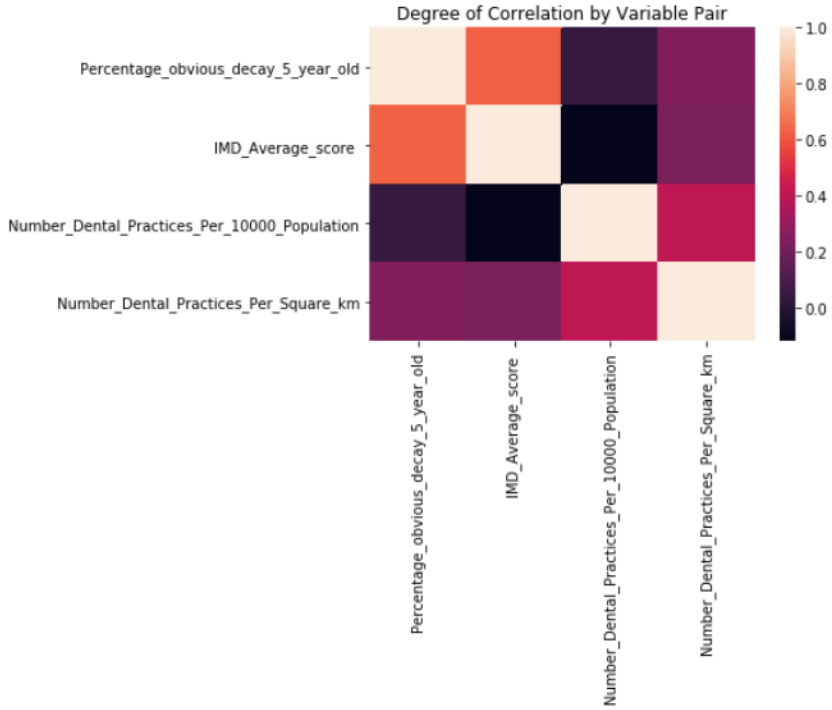

Figure 5 Correlation Heatmap

In the correlation heatmap above, The Pearson Correlation is calculated for each pair of variables. A correlation coefficient of 1 means those values are perfectly correlated and a coefficient of 0 means no correlation at all. The heatmap above shows that Percentage of decay in 5-year-old and IMD Score have strong correlation. The number of dental practices per 10,000 population and number of practices by km^2 seem to exhibit a certain level of correlation too, but much weaker than the other pair.

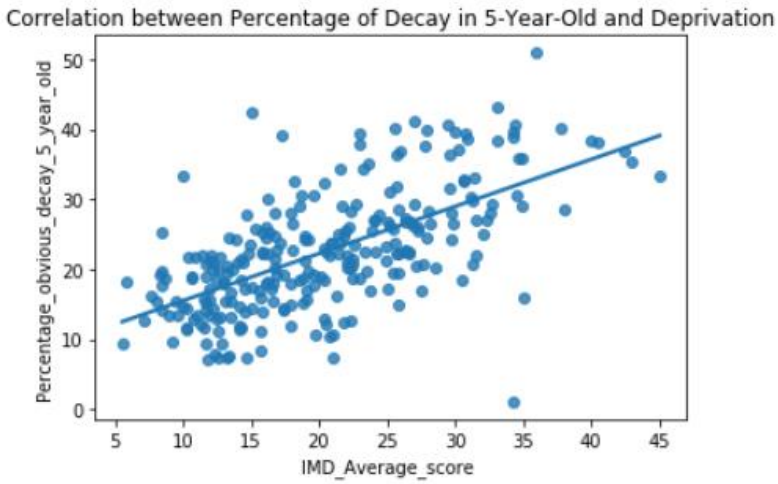

Figure 6 Scatter Plot: Percentage Tooth Decay in 5-Year-Old and IMD Score

The scatter plot above suggests a strong positive correlation between dental health in children and deprivation. The Pearson Correlation Coefficient is 0.627 indicating a strong correlation.

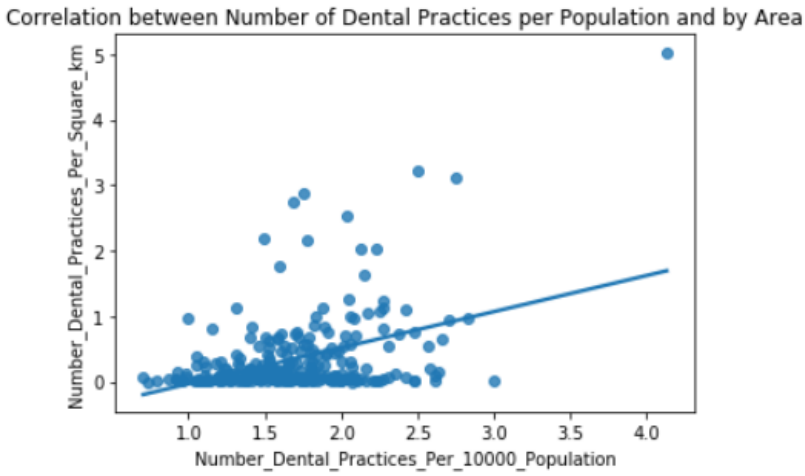

Figure 7 Scatter Plot: Number of Dental Practices per Population and Area

The trendline in the scatter plot above suggests a positive correlation between the number of dental practices per 10,0000 population and the number of practices per square kilometre. Looking at distribution of datapoints in scatter plot, it can been seen that whilst some data follows roughly the trendline, another part of the data concentrates around the lowest number of dental surgeries per population but shows a great variation by surface area. The Pearson Correlation Coefficient is 0.410 indicating a weaker correlation.

3.2. Principal Component Analysis

Since multicollinearity has been observed within the feature set (especially tooth decay in children and deprivation), either one of the variables are dropped from the feature set or a Principal Component Analysis (PCA) is performed. Dropping one variable might be easy but could lead to a loss of some important information. PCA allows us to reduce the dimensionality without losing the information. It puts similar features into groups so that each group is equally weighted or treated by the algorithm when finding the best clusters.

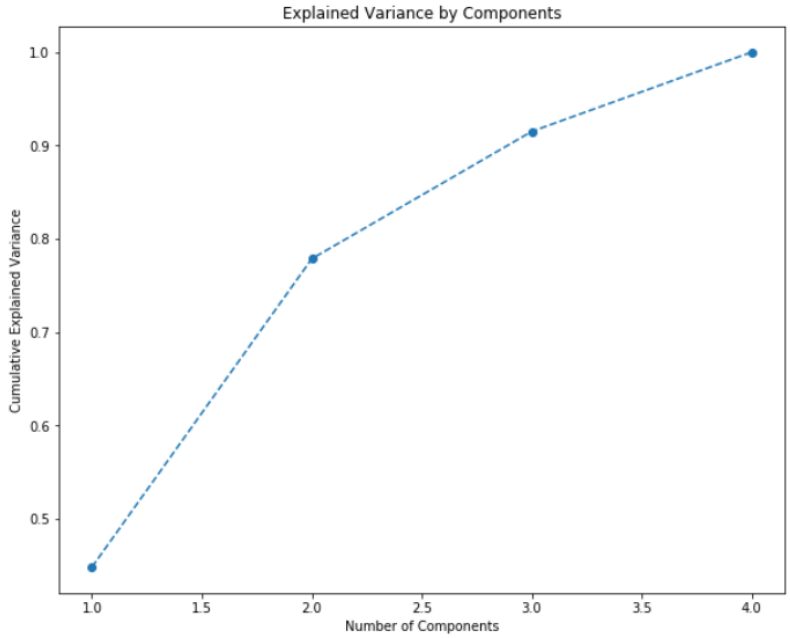

Figure 8 Explained Variance Plot for PCA

The above chart shows the cumulative explained variance by number of components. A good rule of thumb is to choose the number of components that preserves 80% of the variance. Based on the above graph two Components were chosen. The feature set was then normalised and PCA transformation into two components performed. The PCA-transformed dataset was then fed into the K-Means Clustering model.

3.3. K-Means Clustering

The number of clusters, K, has to be set by the Analyst. Using the scree plot method, also called Elbow-Method, can help identify the best K.

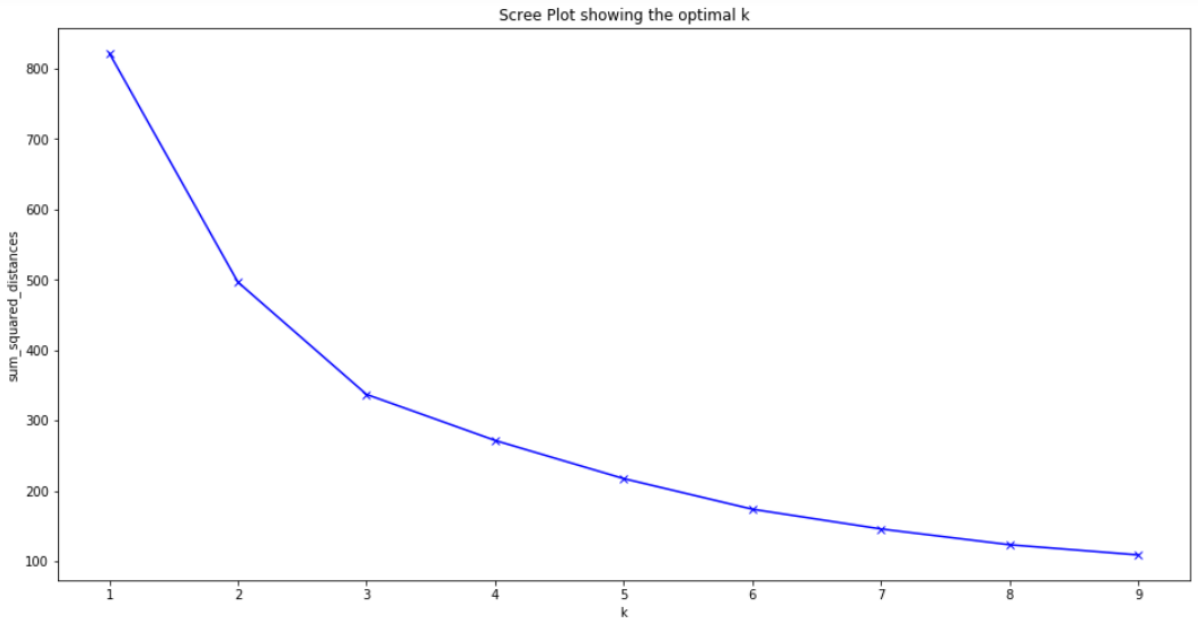

Figure 9 Scree Plot for K-Means Clustering

From the Scree Plot above we can see that 3 clusters is probably the best K, i.e. number of clusters, as the improvement to K=4, K=5 and so on is only marginal and therefore would not add further information to the model.

```
# set number of clusters
kclusters = 3

# run k-means clustering
kmeans = KMeans(n_clusters=kclusters, random_state=0).fit(scores_pca)

# check cluster labels generated for each row in the dataframe
labels = kmeans.labels_
labels = labels + 1 #so that the first cluster has the label 1 and not 0
print(labels)
```

Figure 10 Python Code K-Means Clustering

The above code was used to perform the K-Means model, dividing the data of the two PCA components into 3 distinct clusters.

4. Results

The results of the K-Means clustering can be found in the summary table below, showing the average values for each of the clusters and variables or features.

Average Values by Cluster

Cluster	Number of Districts	Perc Number of District of Total	Percentage Obvious Decay 5-Year-Old	IMD Average Score	Number Dental Practices Per 10,000 Population	Number Dental Practices Per Square km
1	140	53.0	17.2	15.1	1.66	0.11
2	35	13.3	25.2	21.1	2.26	1.44
3	89	33.7	29.9	28.6	1.51	0.3

Figure 11 Average Values by Cluster

In addition to the averages presented in the above table, the clusters need to be evaluated further by looking at the distributions of the data within each cluster. The averages themselves may not tell the whole story. Boxplots will help shed more light into the clusters.

Boxplots

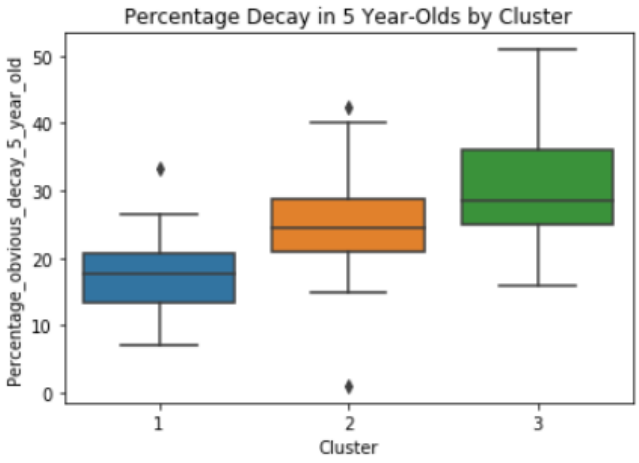

Figure 12 Boxplot Percentage of Tooth Decay in 5-Year-Old by Cluster

The above boxplot shows the distribution of data by cluster and where most data points are concentrated (i.e. coloured boxes). It shows that all clusters seem to be heterogeneous and homogeneous with regards to decay in children, meaning that the coloured boxes do not overlap too a great extent. For example, the 75 th percentile (i.e. upper black line of the box) for cluster 1 is lower than the 25th percentile of cluster 2.

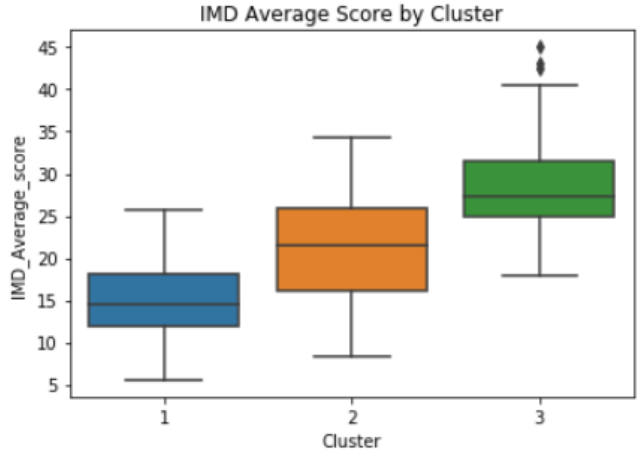

Figure 13 Boxplot IMD Average Score by Cluster

With regards to deprivation levels all clusters appear to be fairly heterogeneous and homogeneous. For example, the mean IMD score for cluster 1 (middle black line of the blue box) is lower than the 25 th percentile of cluster 2.

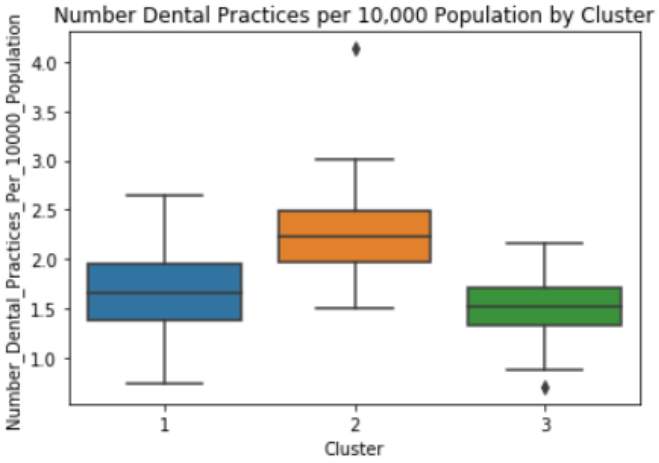

Figure 14 Boxplot Number of Dental Practices per 10,000 Population by Cluster

In terms of Number of Dental Practices per 10,000 Population, Cluster 2 is very distinct from the other two clusters. However, cluster 1 and 3 do overlap to some extent. Cluster 3 is a bit more compressed which means that the data points are less spread. Cluster 1 shows a bit more variation.

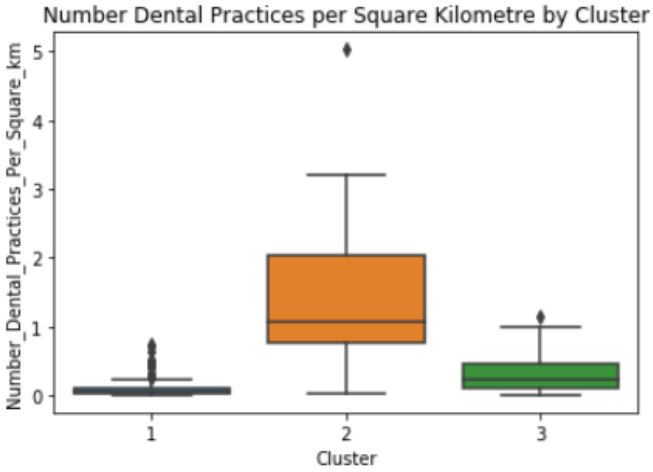

Figure 15 Boxplot Number of Dental Practices per Square Kilometre by Cluster

The boxplot for the number of dental surgeries per square kilometre demonstrates that the k-mean algorithm has done a good job in putting the data into very distinct clusters. The difference between cluster 1 and 3 is not significant, but cluster 1 is extremely compressed whilst cluster 3 shows a bit more variation. Cluster 2 sits apart from the other two clusters and is more significantly more spread.

Map of England with Clusters

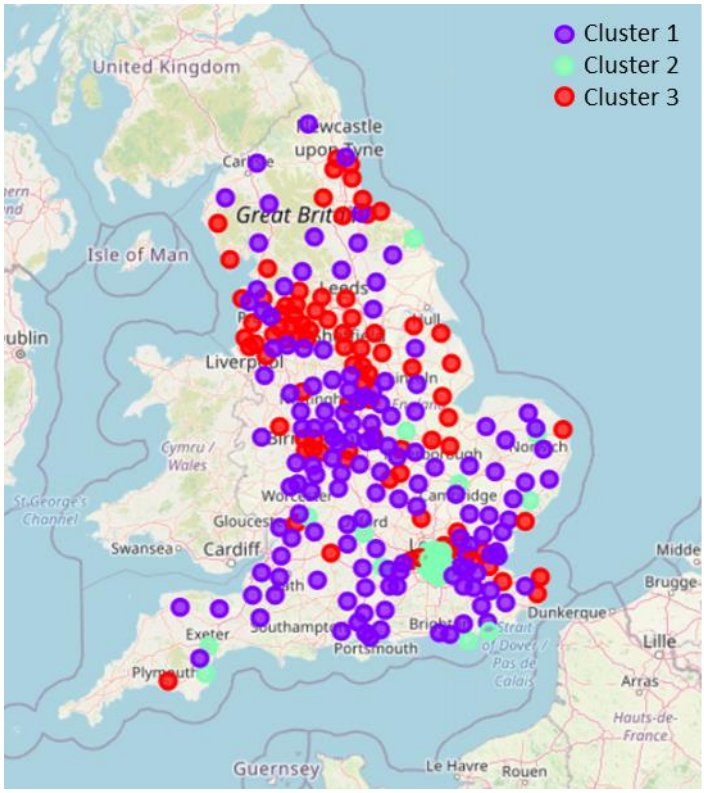

Figure 16 Clusters on Map of England

A visual representation of the clusters on the map of England reveal a formation of distinct geographic groups. Cluster 3 is heavily concentrated around the North West of England. Concentrations of cluster 3 can also be found in North East of the country, along the Mid-East coast and in the East of London. Cluster 2 is largely concentrated in London, where concentration of NHS dental practices is the highest. The majority of the country belongs to cluster 1.

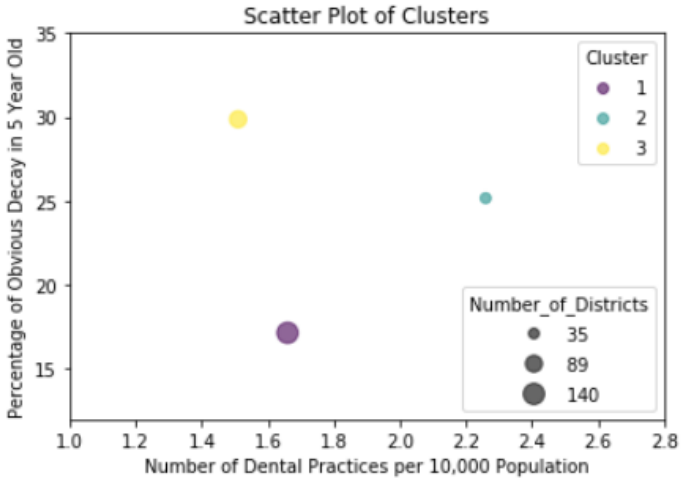

Figure 17 Bubble Chart of Clusters

The above Bubble Chart is another display of the differences and similarities of the final clusters. It shows that cluster 3 (yellow bubble) is the least healthy when it comes to dental health whilst having the fewest numbers of affordable dental practices per population.

Summary of Results

Cluster 1 is the largest cluster with 140 Districts and contains areas with the greatest levels of dental health and lowest levels of deprivation. The mean percentage of obvious tooth decay in 5-year-olds in this cluster is 17.5% which is much lower than the national average of 22.5%. The number of NHS dental practices per 10,000 population is only slightly below 1.69 which is the average in England. The number of practices per square kilometre however appears to be quite low with 0.11, compared to the national average of 0.35. This suggests that inhabitants in those districts may have to travel large distances to be seen by the next NHS dentist.

Cluster 2 is the smallest cluster containing 35 Districts. This cluster stands out because of its high concentration of dental surgeries per population and surface area. It contains Districts such as Westminster or Fulham in London where population density is among the highest in country. The local authorities in those Districts provide their residents with great access to affordable NHS dental care.

Cluster 3 is the second largest cluster with 89 Districts. It is the cluster with the highest mean deprivation scores and percentage of tooth decay in children. The spread of the data also confirms that the clusters are very distinct from each other. On average in this cluster, there are 1.5 of dental practices per 10,000 population, which is much lower than the national average of 1.69. Furthermore, the boxplot revealed that the data for this variable is fairly compressed, meaning not many districts in this cluster have more than 1.69 surgeries per 10,000 population.

In summary the clustering algorithm showed that cluster 3 requires greater access to affordable dental practices as inhabitants in those districts are the most deprived, had lowest levels of dental health in children and had fewest numbers of surgeries by population. More NHS dental practices in those areas would help deprived people to get access to much needed and affordable dental care in an attempt to improve oral health in children and therefore for the future population of England.

5. Discussion

A few limitations of the methodology were identified. One is the fact that outliers were not removed in this project since the author was especially interested in areas with stark differences to the national average as well as to retain integrity of all the Districts in England where data was available. Another limitation of this project was that data was not available for all 316 District across England. In the end only data for 264 District and 4 variables constituted the feature space for K-Means clustering model.

This clustering exercise was performed at a fairly global level, i.e. District level. In order to detect more local variations with regards to dental health within the Districts more granular data should be used when running the clustering. However, this project provides a starting point and may spark an interest for national as well as local governments to run more focused or local clustering projects or reviews. Furthermore, the clustering project carried out in this project could further enhanced by using more variables which may lead to a much richer picture and potentially more distinct clusters.

The clustering exercise carried out in this project resulted in the formation of three distinct clusters. The objective is now to device a tailored treatment plan for these 3 clusters. At a high level the author suggests the following strategies.

Cluster	Characteristics	Recommended Strategy
1	• Above average dental health • Below average levels of deprivation (IMD score) • Average number of NHS dental practices per population • Below average number of NHS dental practices per square kilometre	• Maintain the level of dental care in these Districts • Assess the geographic reachability to NHS dentists for residents
2	• Average dental health • Average levels of deprivation (IMD score) • Above average number of NHS dental practices per population • Above average number of NHS dental practices per square kilometre	• Improve education and awareness regarding dental hygiene and set incentives for people to make regular visits to their local dentists (e.g. remind patients of their regular check-ups via SMS or phone)
3	• Below average dental health • Above average levels of deprivation (IMD score) • Below average number of NHS dental practices per population • Average number of NHS dental practices per square kilometre	• Improve education and awareness regarding dental hygiene and set incentives for people to make regular visits to their local dentists (e.g. remind patients of their regular check-ups via SMS or phone) • Increase the number of NHS dental practices to provide more choice for inhabitants • Carry out further analyses to investigate more local differences in terms of dental care and health

Figure 18 Characteristics and Strategies by Cluster

To define a specific set of measures to improve the state of dental care and health, the underlying causes of the problems especially evident in cluster 3 need to be identified first. Root causes can be: a shortage of NHS dentists as many of them move into the more lucrative private sector or even abroad where working conditions and pay are better, lack of incentives (e.g. monetary compensation) for Dentists to move to less affluent, rural areas to join or open a practice and difficulties associated with the current UDA (units of dental activity) target system (Chhaya, 2021). In addition, funding for NHS dental care in England, after adjusting for inflation, has decreased from 2014-15 to 2018-19 (National Audit Office, 2020). However, this is outside the scope of this project.

6. Conclusion

A combination of PCA and K-Means Clustering, unsupervised Machine Learning algorithm, was performed to divide 264 Local Authority Districts into 3 distinct clusters. The features, tooth decay in children, deprivation levels and number of NHS Dental practices by inhabitants and surface area for each Local Authority District in England, were used as input variables for the clustering model. It clearly highlighted the geographic differences in dental health and access to affordable dental care across England, allowing the identification of Local Authority Districts where dental services and care need improving. Cluster 3 exhibited the worst characteristics, with below average dental health, above average levels of deprivation and below average number of NHS dental practices per population. Cluster 3 is heavily concentrated around the North West of England. Further areas include Districts along the Mid-East coast and in the East of London.

Ultimately, to improve dental health in cluster 3 areas three things would need to be put in place. One, is to increase availability of dental services, two is incentivise people to use these services and three is educate people about oral health. However, it would be naïve to say that it is that simple. The root causes for the problematic dental care situations in these Districts need to be properly understood so that specific and concrete measures can be devised to address these underlying issues.

The findings in this report should be taken into consideration by the British Dental Association and the different Local Authorities to provide more equal access to affordable dental care across England as well as improving dental health as a whole.

References

Appleby, J., & Reed, R. (2017). *Root causes - Quality and Inequality in Dental Health.* London: QualityWatch.

Chhaya, A. (2021). *Interview with a Dentist in London.* London.

Digital NHS. (2021, January 31). *Miscellaneous*. Retrieved from Miscellaneous: https://digital.nhs.uk/services/organisation-data-service/data-downloads/miscellaneous

Ministry of Housing, Communities & Local Government. (2019, September 26). English Indices of Deprivation 2019. UK. Retrieved January 31, 2021, from https://www.gov.uk/government/statistics/english-indices-of-deprivation-2019

National Audit Office. (2020). *Dentistry in England.* National Audit Office.

Office for National Statistics. (2021, January 31). *Geoportal Statistics*. Retrieved from Local Authority Districts (December 2019) Boundaries UK BFC: https://geoportal.statistics.gov.uk/datasets/1d78d47c87df4212b79fe2323aae8e08_0/data

Office for National Statistics. (2021, January 31). *Nomis - Official Labour Market Statistics*. Retrieved from Population estimates - local authority based by single year of age: https://www.nomisweb.co.uk/query/construct/summary.asp?reset=yes&mode=construct&dataset=2002&version=0&anal=1&initsel=

Public Health England. (2021, January 31). *Public Health Profiles*. Retrieved from fingertips.phe: https://fingertips.phe.org.uk/search/dental#page/0/gid/1/pat/6/par/E12000004/ati/102/are/E06000015/cid/4/tbm/1/page-options/ovw-do-0