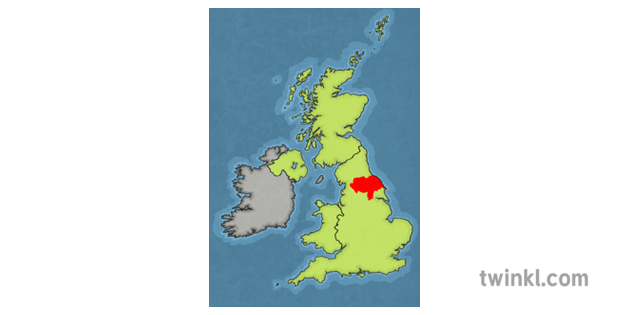
**Determining the Best Location for a New High School in North Yorkshire, UK**

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

The population of the UK is growing steadily, as is the population of the North Yorkshire county region (Fig. 1). The growth trend in the region is likely to continue and possibly accelerate due to the movement from urban to rural living stimulated by Covid-19 and the increased acceptance of home working. The prohibitively high cost of living in many southern areas is also causing many people to consider a move further north.



**Fig. 1 – UK map with North Yorkshire county highlighted in red.1**

Although the population of the UK (and North Yorkshire) is ageing, the overall rate of population growth is significantly higher, and therefore the number of children in the region is increasing. This results in an increased requirement for education and high schools. As most high schools in the region are already near capacity, it is unlikely that current education provision will be proficient. It is therefore necessary to determine where a new school should be located.

As much of the North Yorkshire county is very rural, many children are already required to travel significant distance to reach their school. This may have a negative impact on a child’s education due to increased tiredness and reduced learning effectiveness. This is also a disadvantage for the local authority as they are required to fund transport for all children to their nearest school.

* 1. **Problem**

North Yorkshire county council needs to consider the increased requirement for high school places in the next few years, and determine the best location to site a new high school.

Data that is likely to influence the decision includes population distribution within the region, the locations of existing high schools, the current capacity situation of existing high schools, and the average distance children must travel to their nearest high school. This project aims to use these data to propose the location of a new high school in North Yorkshire.

* 1. **Interest**

North Yorkshire county council would be the key interested party in this analysis, as they would be able to use the analysis to locate a new school in the area with the most benefit to the local population, and reduced requirements to fund transport.

Other interested parties may include families looking to move to the North Yorkshire region, and considering the proximity to local high schools as part of their decision.

1. **Data Acquisition**
   1. **Data Sources**

Geographic data for the North Yorkshire region was taken from [Doogal](https://www.doogal.co.uk/PostcodeDownloads.php)2, which compiles a UK wide postcode directory and hierarchical assignments (e.g. assignments of postcodes to constituency, local authority city, county) using data from Ordnance Survey, Royal Mail and the Office for National Statistics.

Relevant data on schools including name, location (latitude/longitude), postcode and school type (primary school, high school etc.) can be accessed using Foursquare API .

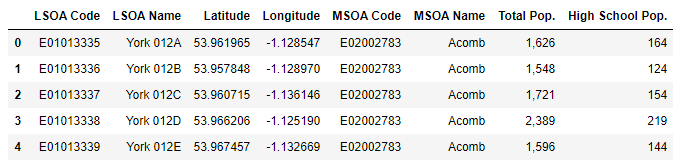
Data on population distribution was sourced from the [North Yorkshire County Council Data Hub](https://hub.datanorthyorkshire.org/dataset/population-estimates/resource/d95b8528-c4fc-444f-a691-b5d8b90c365d)3. This contains the most accurate and localised data on local population distribution as it is compiled from UK census information. The lowest level data available at the time of analysis is by ‘Lower Super Output Area’ (LSOA). The UK is divided into LSOAs, each of which is a geographic region with a minimum population of 1000 and a mean of 15004.

In order to perform a more detailed analysis, supplementary data on each school in the region of interest was sourced from the [British Government Department of Education](https://www.get-information-schools.service.gov.uk/Downloads)5, including the classification of each school as rural/urban, and the current school capacity information.

* 1. **Data Cleaning**

Data on population distribution was first cleaned to contain only the North Yorkshire region, and all unneeded columns dropped. The data is supplied at postcode level, but as the population distribution data is only available by LSOA, the data was aggregated to LSOA, and the mean latitude and longitude calculated for each LSOA.

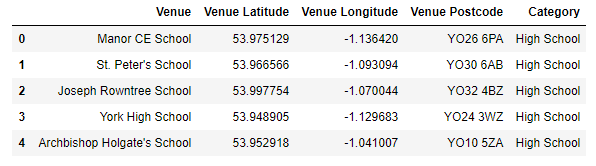
The population data was reduced to contain only the total population for each LSOA, and the high school age population calculated for each by summing the data for ages 11-18.



**Fig. 3 – Sample of dataframe containing the geographical coordinates, total population and estimated high school population for each LSOA in the North Yorkshire.**

School data was obtained from foursquare using a Venue Search using search term ‘school’. This was chosen in place of a category search as schools may be assigned to many categories within Foursquare (e.g. School, High School, Community College). Data was requested for the geographical coordinates of each LSOA, within a radius of 10km and using a limit of 300. The high limit and radius were used to maximize the probability of returning all schools within the region.

The data was cleaned by removing duplicates and dataframe slicing based on category to remove anomalies (e.g. driving school) and restrict the data set to only high schools, resulting in a list of 130 high schools with their geographical coordinates, and postcodes where available.



**Fig. 2 – Sample of High School data obtained from Foursquare.**

The Department of Education data on school capacity was cleaned to retain only those high schools within the LSOAs in the North Yorkshire region, and the Urban/Rural status and Capacity % obtained for each. It is noted that there is missing capacity data, and this will be dealt with in the next section. This data will then be joined on the Fourpoint data in the exploratory data analysis stage**.**



**Fig. 3 – Sample of dataframe containing school capacity and rural/urban classification for schools in North Yorkshire.**

**References:**

1. <https://www.twinkl.co.za/illustration/united-kingdom-map-with-north-yorkshire>, accessed 23 January 2021.
2. <https://www.doogal.co.uk/PostcodeDownloads.php>, accessed 23 January 2021
3. <https://hub.datanorthyorkshire.org/dataset/population-estimates/resource/d95b8528-c4fc-444f-a691-b5d8b90c365d>, accessed 23 January 2021.
4. <https://datadictionary.nhs.uk/nhs_business_definitions/lower_layer_super_output_area.html>, accessed 24 January 2021.
5. <https://www.get-information-schools.service.gov.uk/Downloads>, accessed 23 January 2021.