Determining the Best Location for a New High School

North Yorkshire, UK

Applied Data Science Capstone by IBM/Coursera

The Background, The Problem…

- UK population is growing
- Existing high school provision in the North Yorkshire region is becoming inadequate
- New schools will be required to cope with the increased demand
- Predicting the best location for a new school helps the local authority
  - Where will a new school have the most impact
  - Reduction in school transport costs
  - Encouraging families to move to the local area therefore increasing development
- Can we use information on population distribution and existing schools to identify optimum areas?

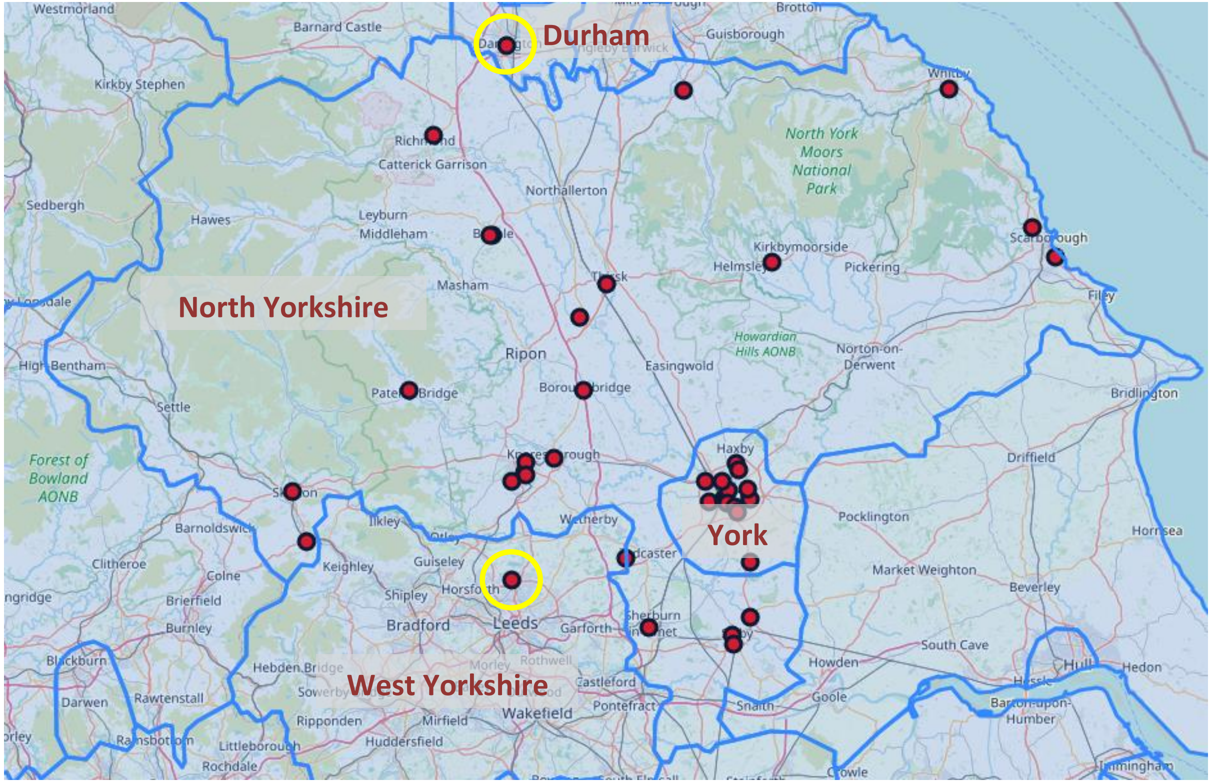
Data Acquisition & Cleaning

- School location data obtained using Foursquare API
- School capacity data obtained from the British Government, Department of Education
- Population distribution data obtained from the North Yorkshire County Council Data Hub
- Geographic & administrative area data obtained from Doogal
- Duplicate data, data from outside the region of interest, and non-high schools were dropped
- Cleaned data contained 38 schools and 493 population areas within the region

Visualising Existing Schools

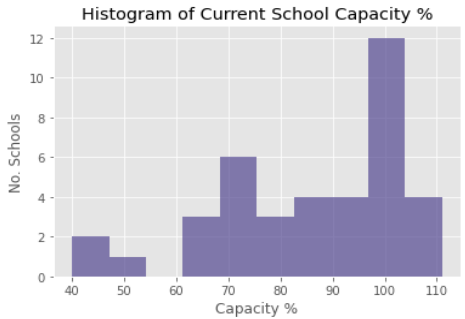
- North Yorkshire county is made up of 2 administrative areas
  - York
  - North Yorkshire

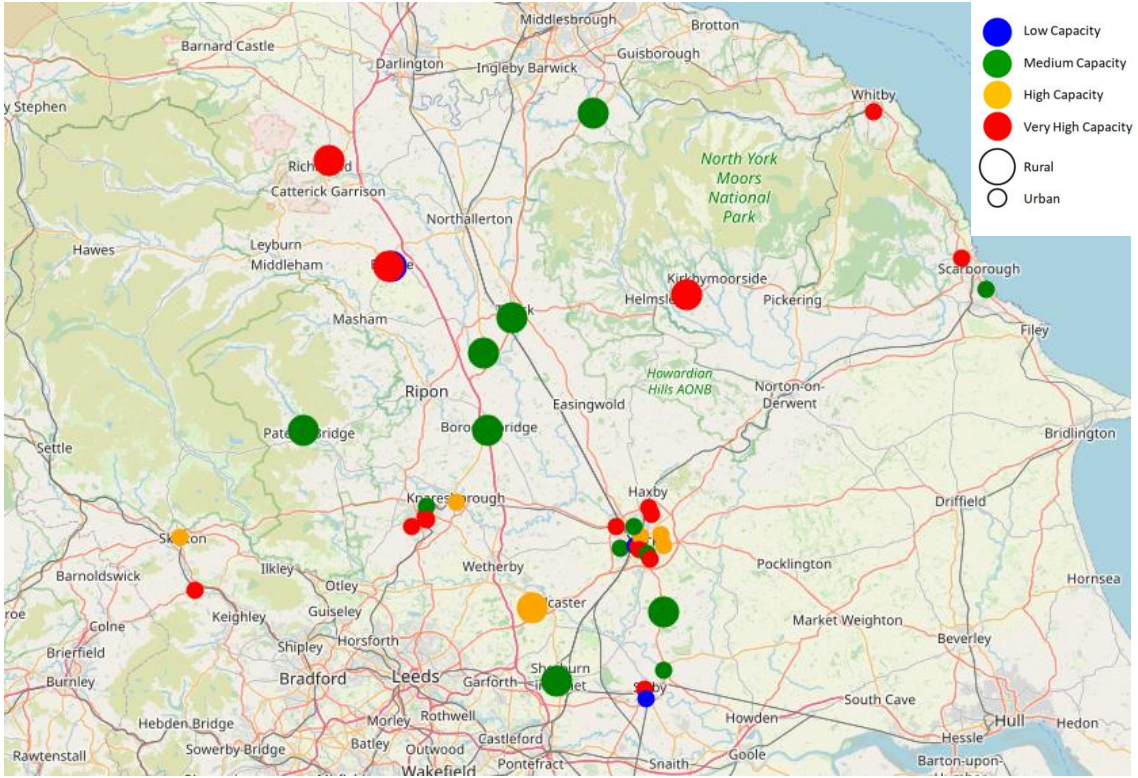
- Schools circled in yellow dropped as outside region of interest



Capacity Status of Existing Schools

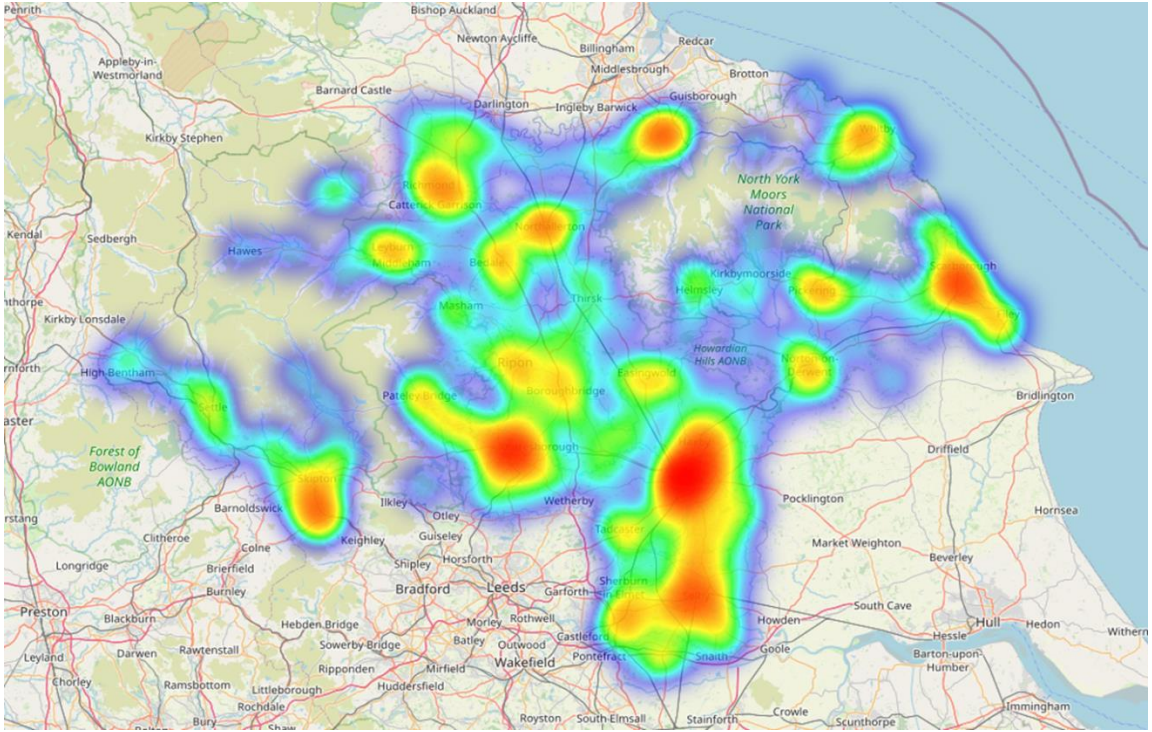
- Significant number of schools with high or very high capacity
- No obvious pattern of capacity distribution by location
- Rural schools more likely to have lower capacity utilisation





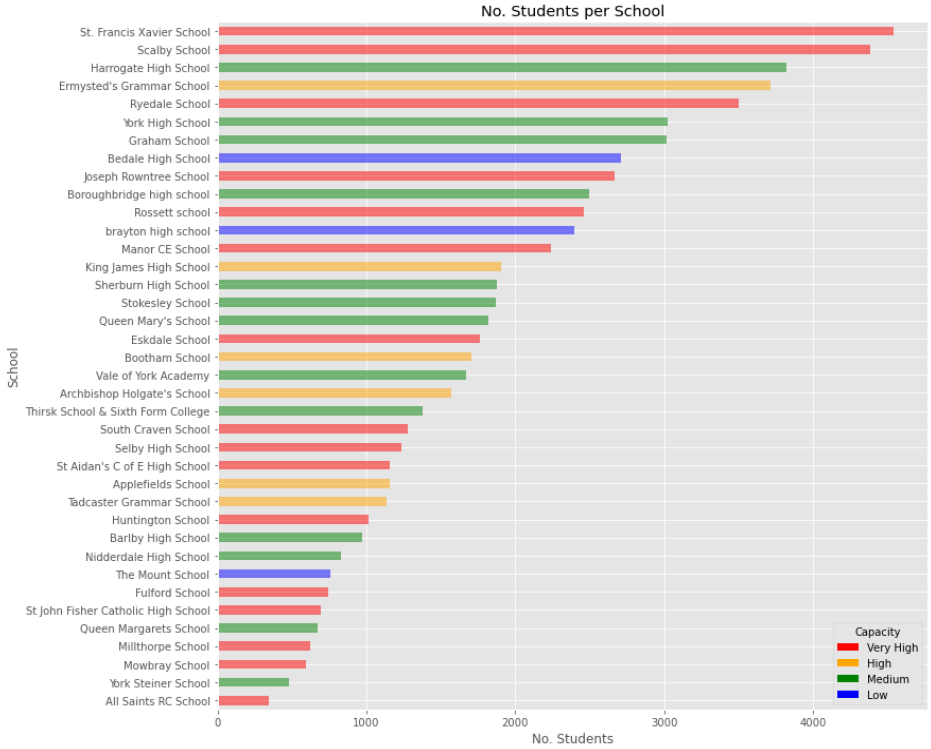
Population Distribution

- High school age (11-18) population only
- Aligns well with school distribution
  - High population density is mirrored in higher density of school provision



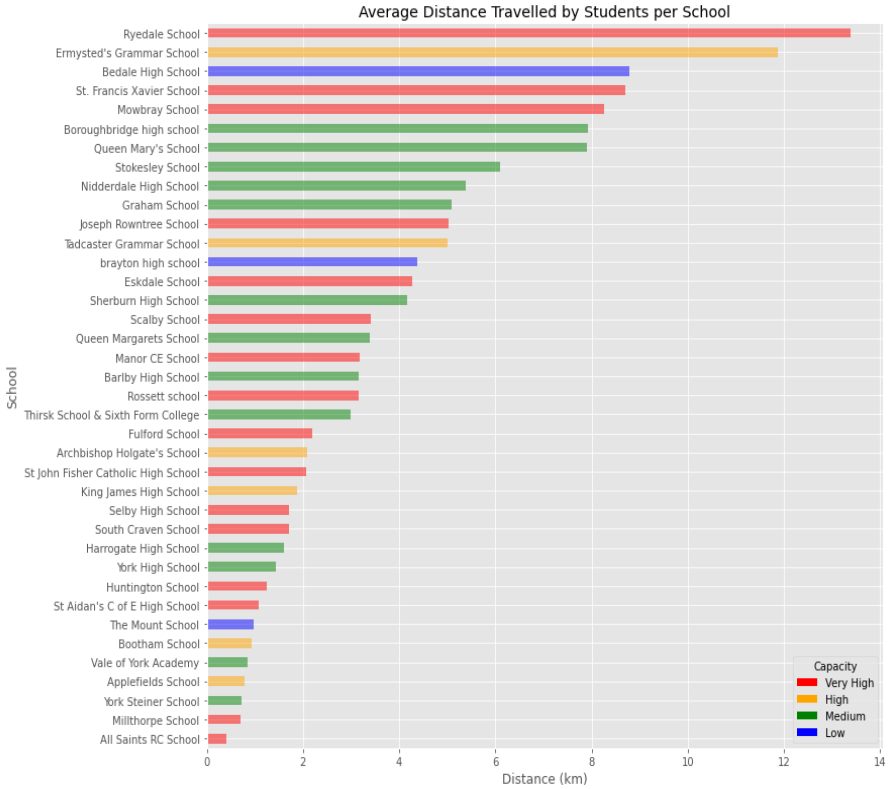
No. Students Allocated per School

- Large variation ranging from < 500 to > 4000
- Significant number of schools with very high student allocation
- No obvious correlation with capacity status
  - But 4 of 5 schools with highest number of students allocated also under significant capacity strain



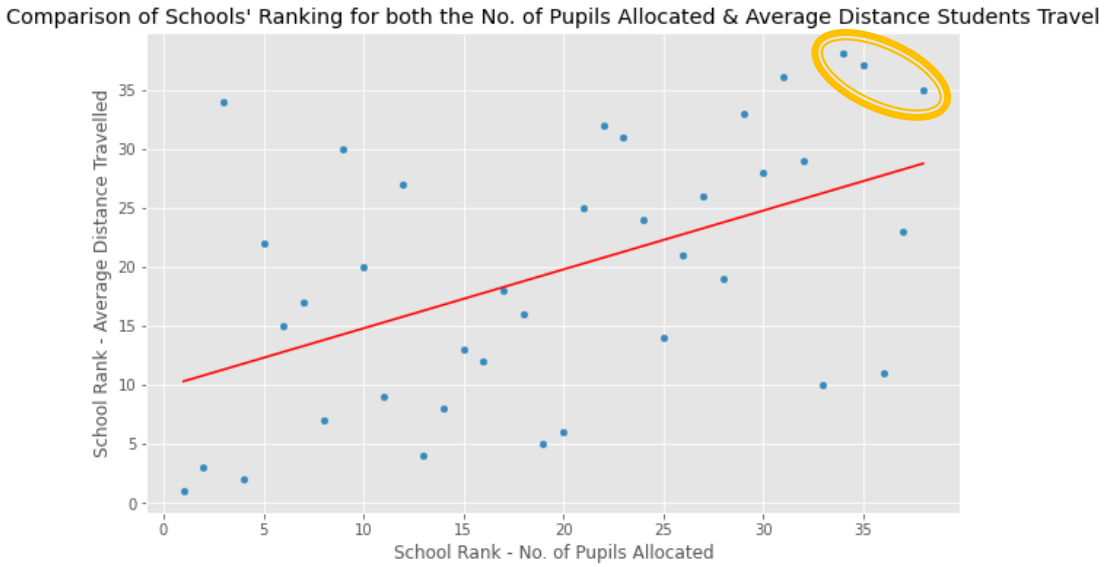
Ave. Distance Travelled by Students per School

- Overall average distance travelled is relatively low across schools
- Schools with higher average distance indicate a larger geographical catchment area
- No obvious correlation with capacity status
  - But 4 of 5 schools with highest number of students allocated also under significant capacity strain



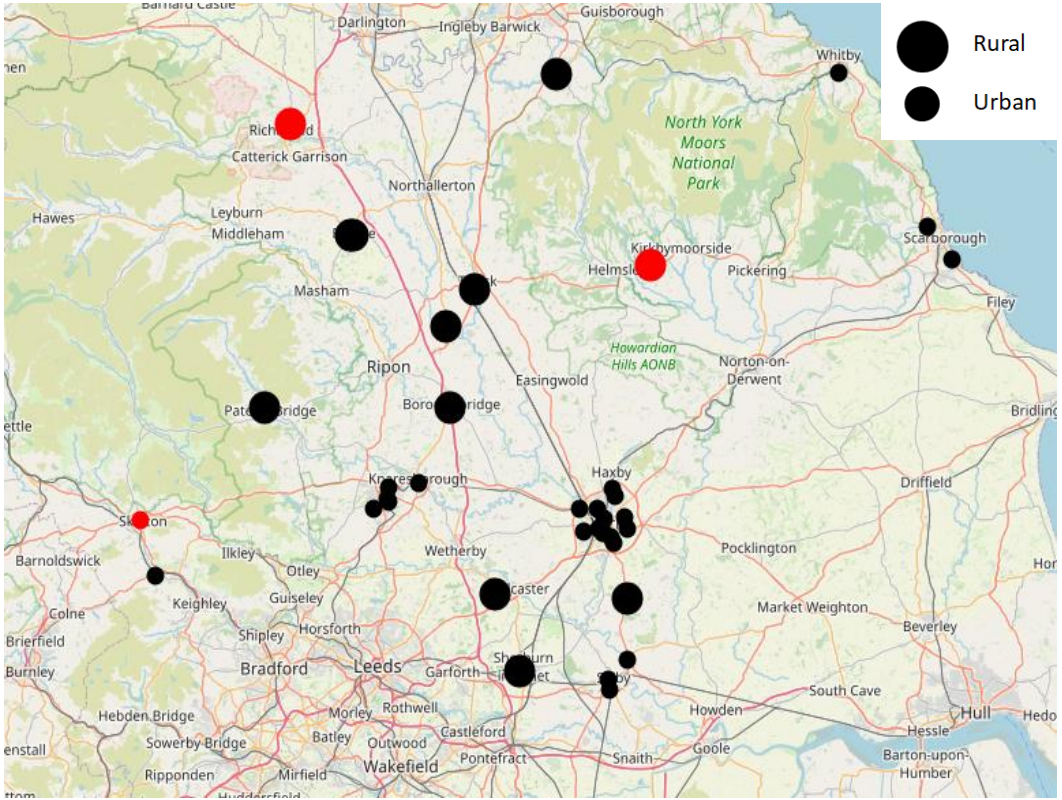
No. Students Allocated vs. Average Distance

- Schools with a high number of students allocated also have a higher average distance travelled
- Schools performing poorly in both analyses (lower rank) are indicative of insufficient education provision in the area
- Three schools performing in the bottom 5 for both identified



Proposed Areas for a New School

- Three schools previously identified marked in **red**
- No other schools in the vicinity – limited alternative options for students
- All have high / very high current capacity utilisation
- Represent three starting points for more detailed investigation



Conclusion & Future Direction

- Three potential areas for siting a new school have been identified
- Follow-up analysis is necessary to narrow down the locations further
- Accuracy could be improved by testing error arising from assumptions and including additional data features
- Future ideas to include:
  - Average travelling time (in addition to distance)
  - School OFSTED rating
  - School type – State/Independent
  - Localised rates of population growth