IBM Data Capstone Assignment

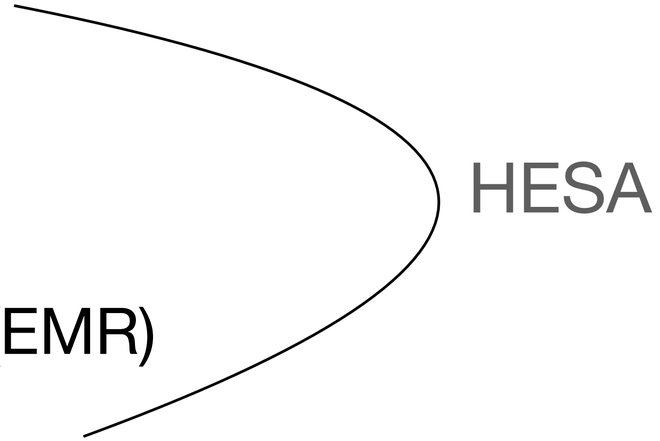
Predicting Carbon emissions outputs for UK Universities

Paddy Marshall

Data

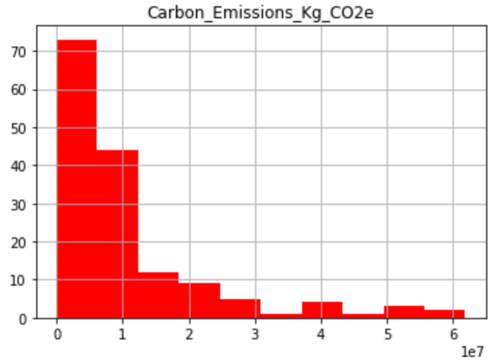
Data Collection and Sources

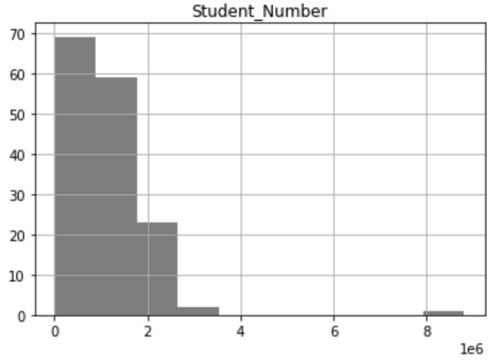
- Student Record
- Staff Record
- Estates Maintenance Record (EMR)

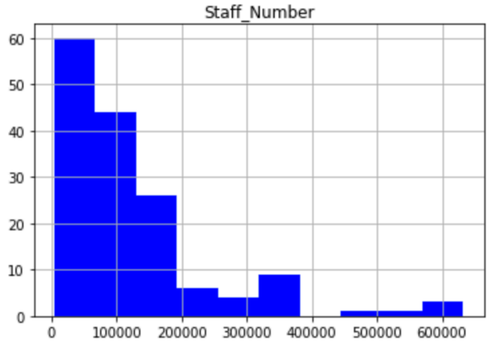


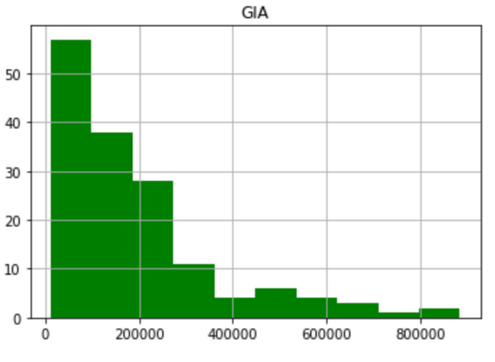
- Foursquare API
- Location data

Data Preparation and Understanding

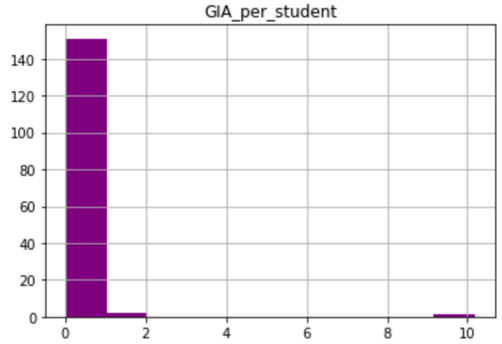


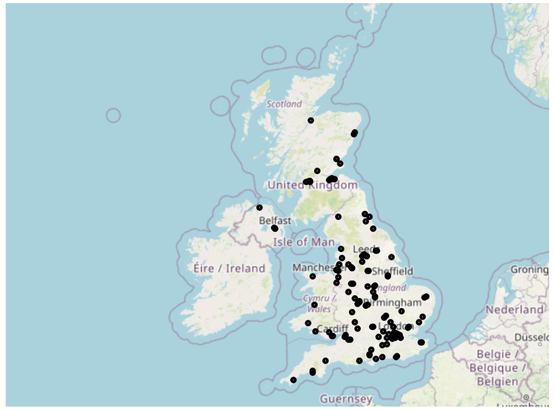






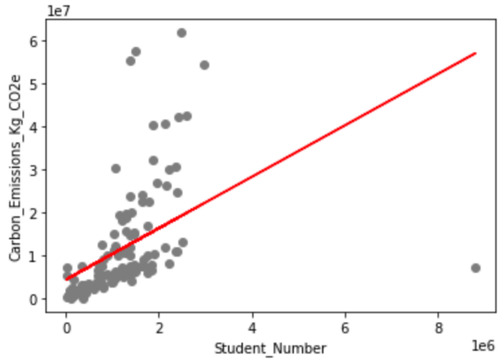

```
UKPRN                      float64
Provider                    object
Student_Number               int64
Staff_Number                 int64
GIA                        float64
Carbon_Emissions_Kg_CO2e   float64
Longitude                  float64
Latitude                   float64
dtype: object
```

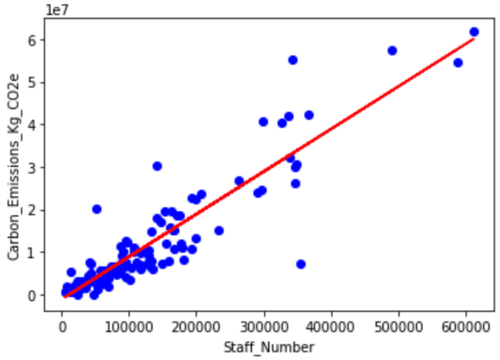


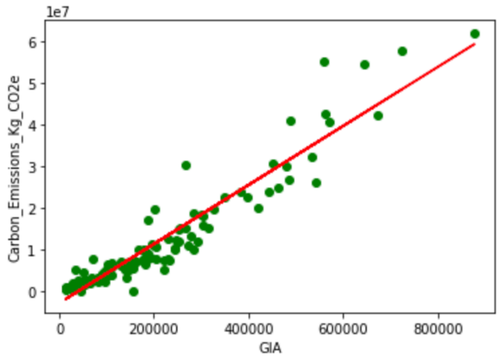


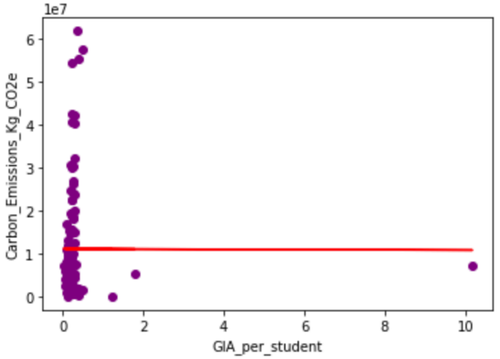
Methodology

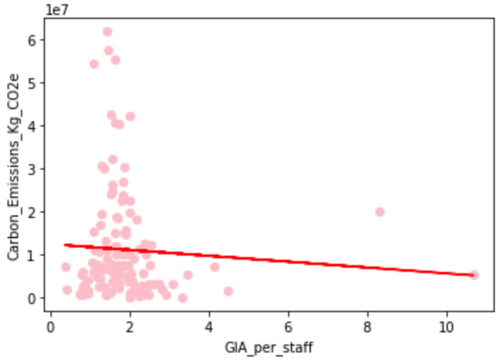
Linear Regression



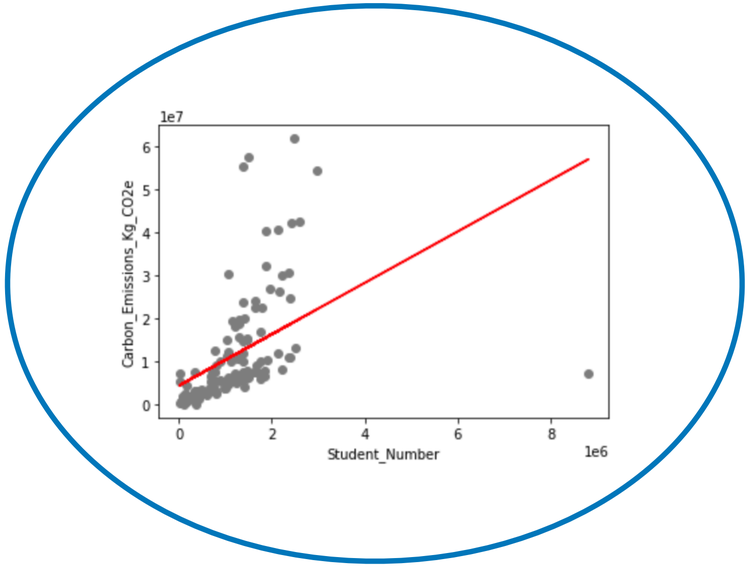


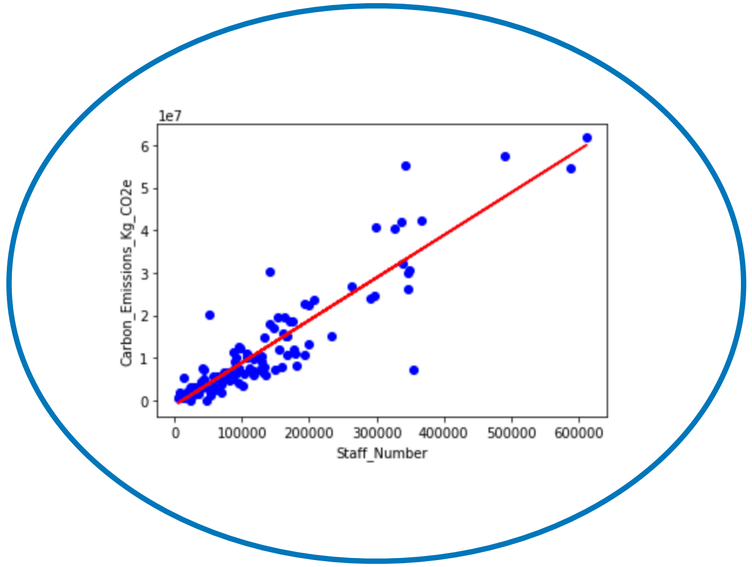


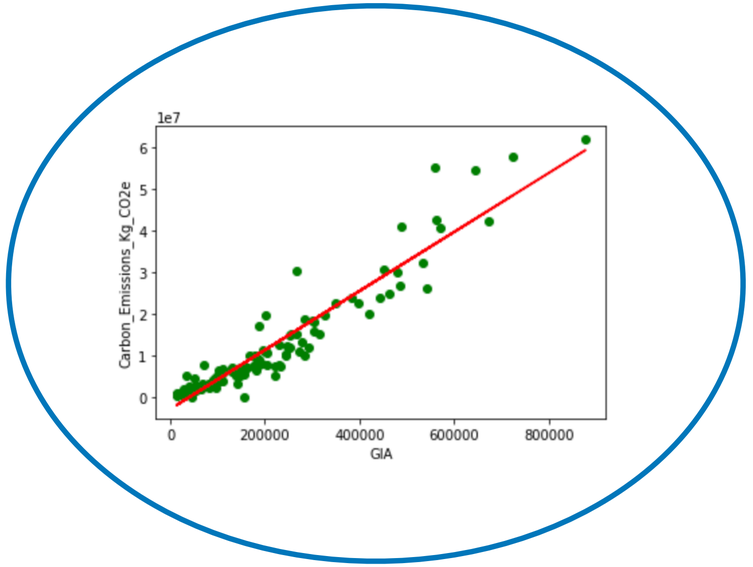


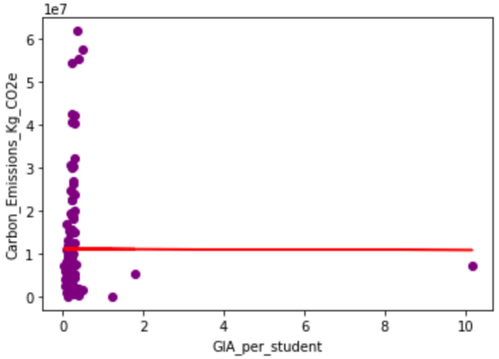


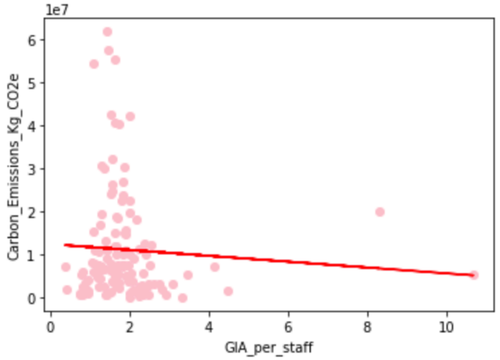
Linear Regression



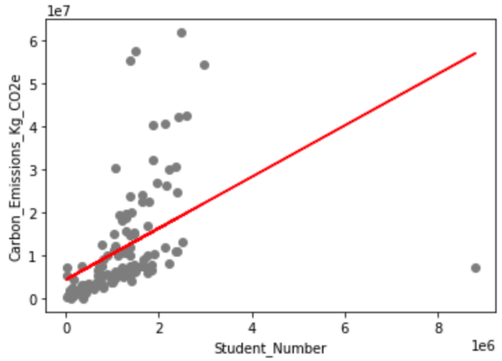


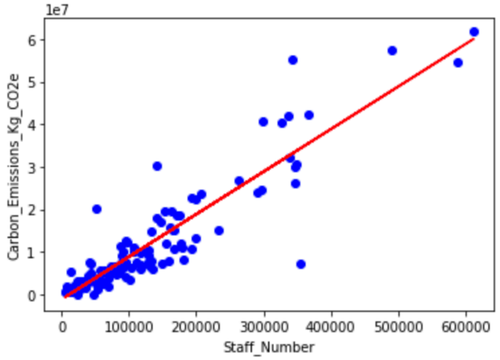


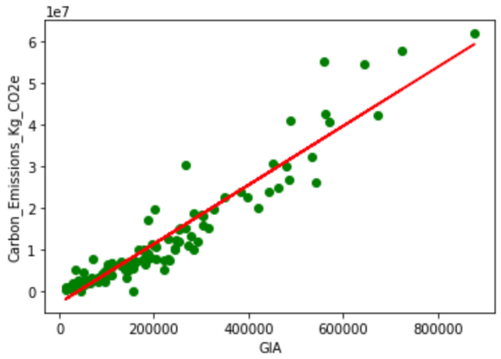




Linear Regression







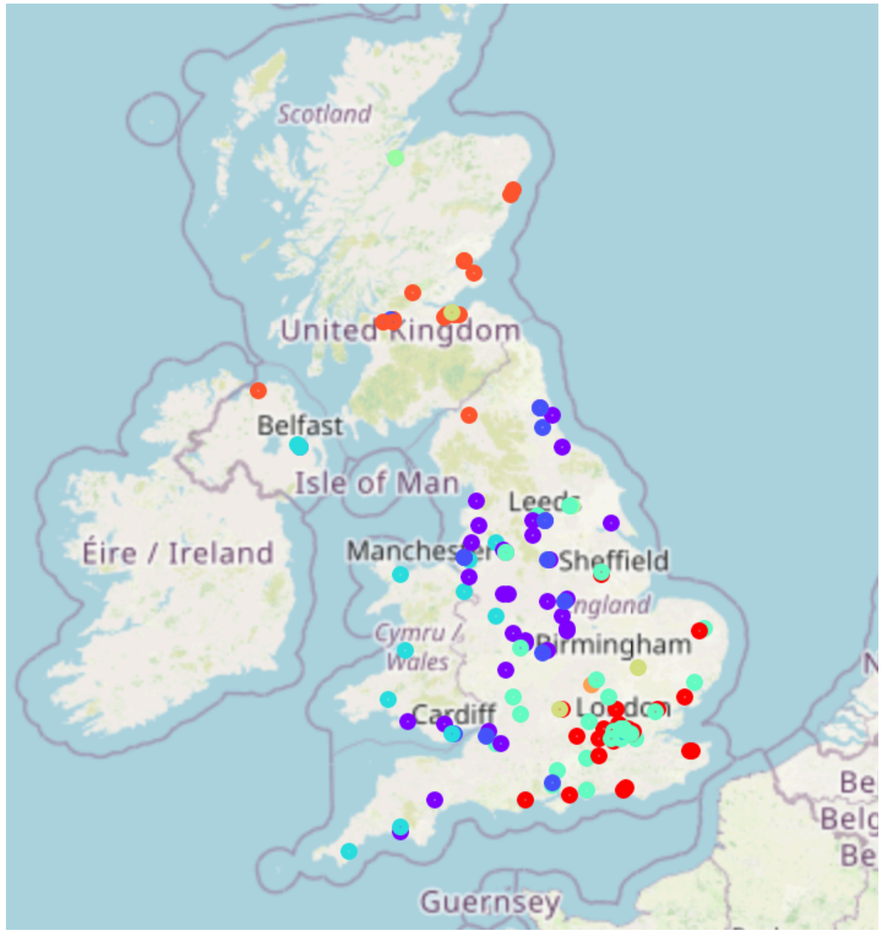
0.36
R2

0.87
R2

0.94
R2

Variance Score = 0.93

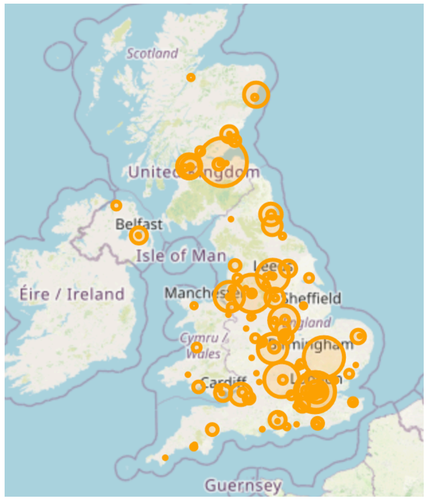
Cluster groups



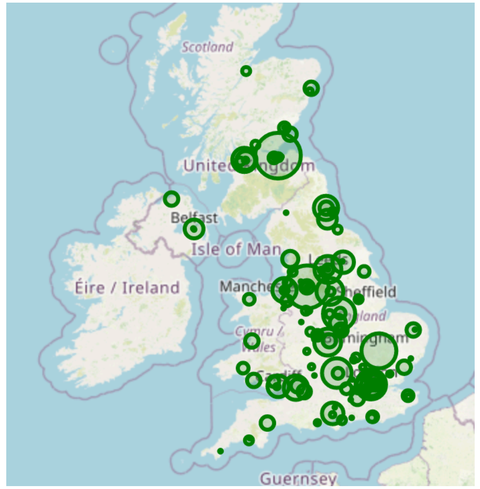
0	Teaching Universities 1
1	Holding Group University
2	Research Giants
3	Teaching Universities 2
4	Oddities
5	Teaching Universities 3
6	Scottish Universities
7	The Open University
8	The Russell Group
9	Specialists and Conservatoires

Mapping and Regional Variations

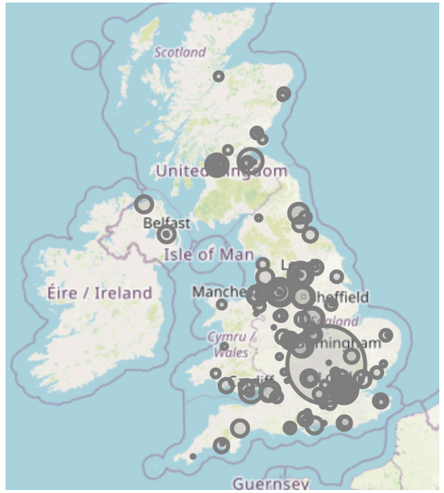
Carbon Emissions



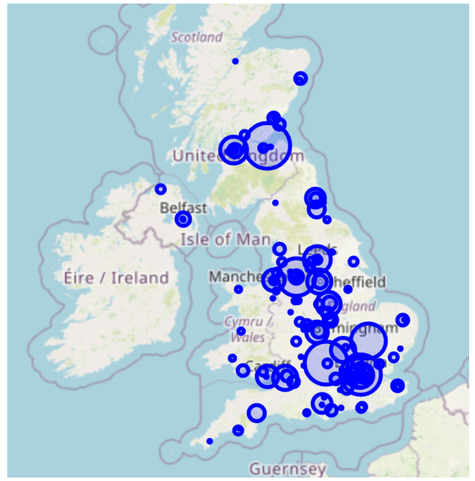
GIA



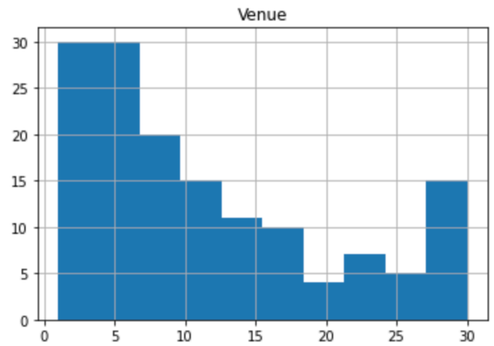
Student Number

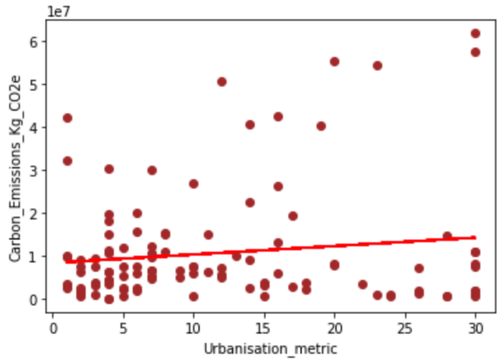


Staff Number

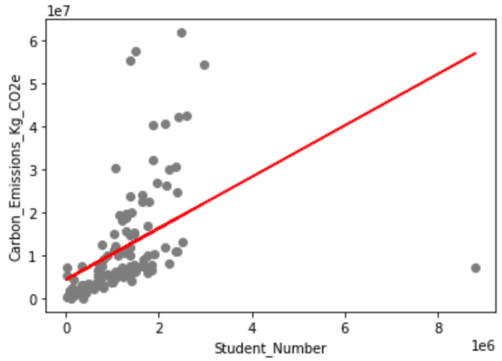


Urbanisation Metric

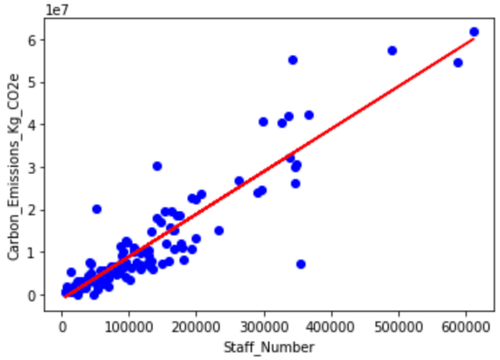




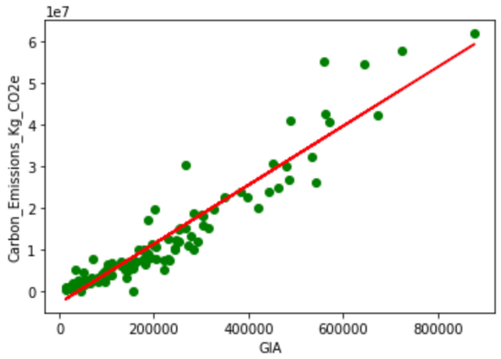
Linear Regression



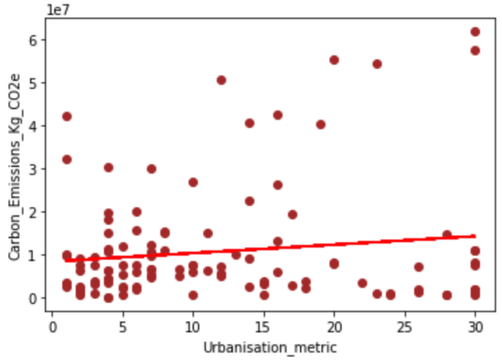
0.36
R2



0.87
R2



0.94
R2

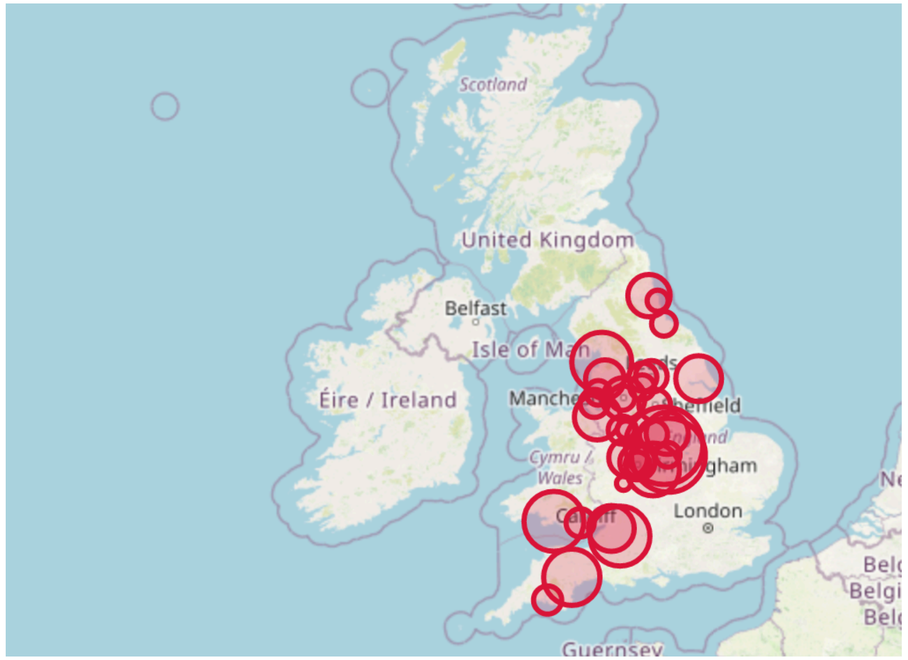


Variance Score = 0.84

Results

University 1 - Poppleton University

- The model predicts that the University would produce 10413210.43849994 Kg/CO2e Carbon Emissions
- This falls within cluster 1 of our groups - which ranges from 3229459.5110000004 to 19674461.337
- These Universities generally are within metropolitan non-coastal areas, which is representative of our chosen case study.



University 2 - the University of Life

- The model predicts that the University would produce 595366.28194308 Kg/CO2e Carbon Emissions
- This falls within cluster 5 of our groups - which ranges from 456340.662 to 7656865.655 Kg/CO2e
- These Universities are generally more rural areas some of which are coastal.

Discussion

Benefits

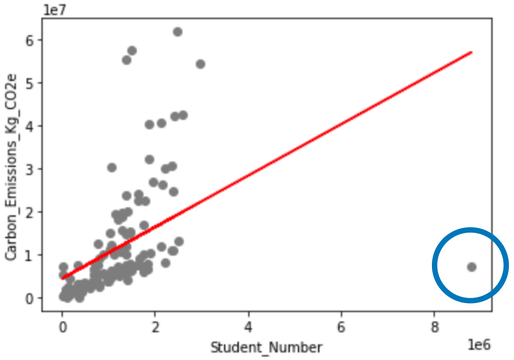
- A high variance score means we can be confident in the accuracy of the model.
- Clustered groups allows us to provide a range to stakeholders for target setting.
- Recommendations can be made to steer direction within this range, based on regional variances identified through our mapping
- The visualisation of this data helps stakeholders who may not be as data literate.

Limitations

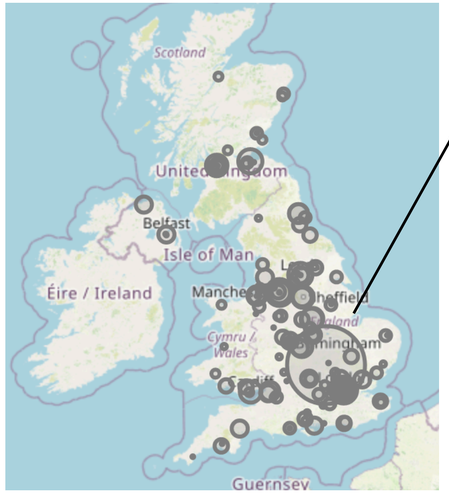
- This report is based on 2018/19 data due to a data publishing lag from HESA.
    - At the time of writing this is 2 years out of date, we could create a trend metric to try and predict change over these 2 years and then build from this.
- We used the total scope 1 and 2 emissions as one of our variables, to increase the accuracy of our model, we could have looked at contributions to this (e.g. carbon emissions from transport) to create a base variable that we know effects our target.
- We allocated clusters to our predictions visually, this could have been strengthened by looking at classification strategies such as K nearest Neighbour.

Barriers

- The Open University influenced the strength of some results, particularly on student numbers, due to its unique operating model of very high student numbers, and very low GIA.
- The foursquare API can only call information within 100,000 metres, and has a maximum count of 30.
- We only counted for scope 1 and 2 emissions. There are difficulties across the sector in recording scope 3, so we can't trust the accuracy of available data.



The Open University



Conclusion

Conclusion

Our original aim was to provide two case study universities with recommendations of how to set carbon emission targets for the coming year, based on benchmarking of other Universities carbon emissions, across the UK.

To do this we created a model which predicts carbon emissions output, based on publicly available sector data, such as staff numbers, student numbers and urbanisation, on historical carbon outputs, and regional variances which may influence this, such as urbanisation.

We can be confident in the accuracy of our model, and that any supplementary recommendations we made to stakeholders were robust and defensible.