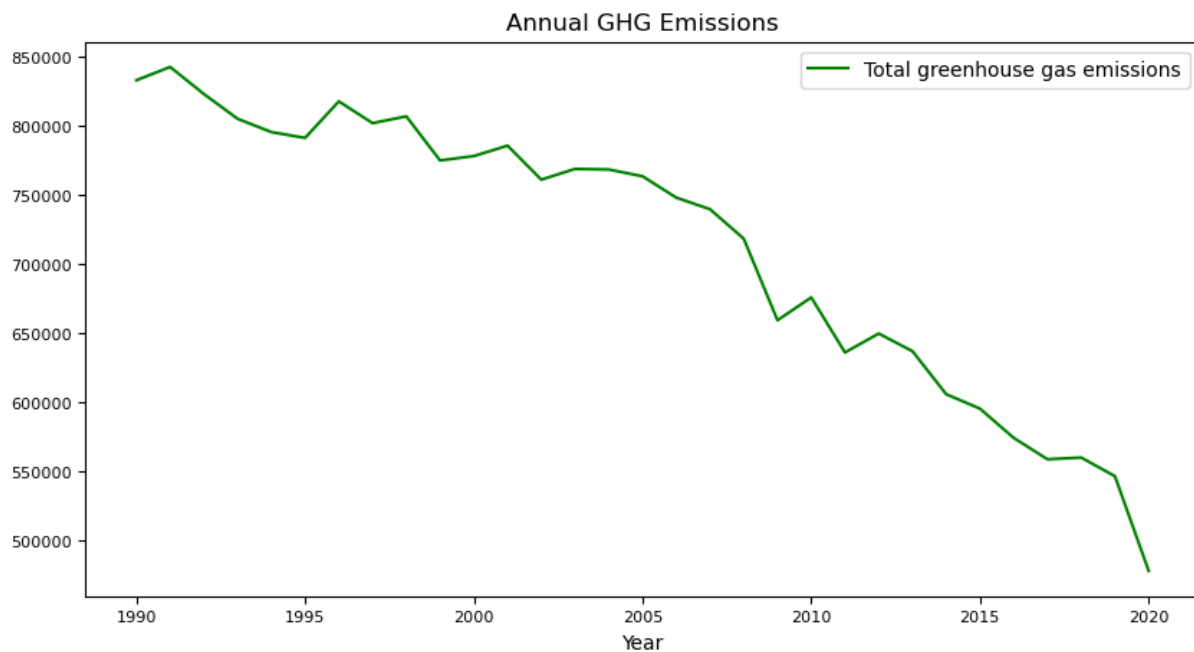
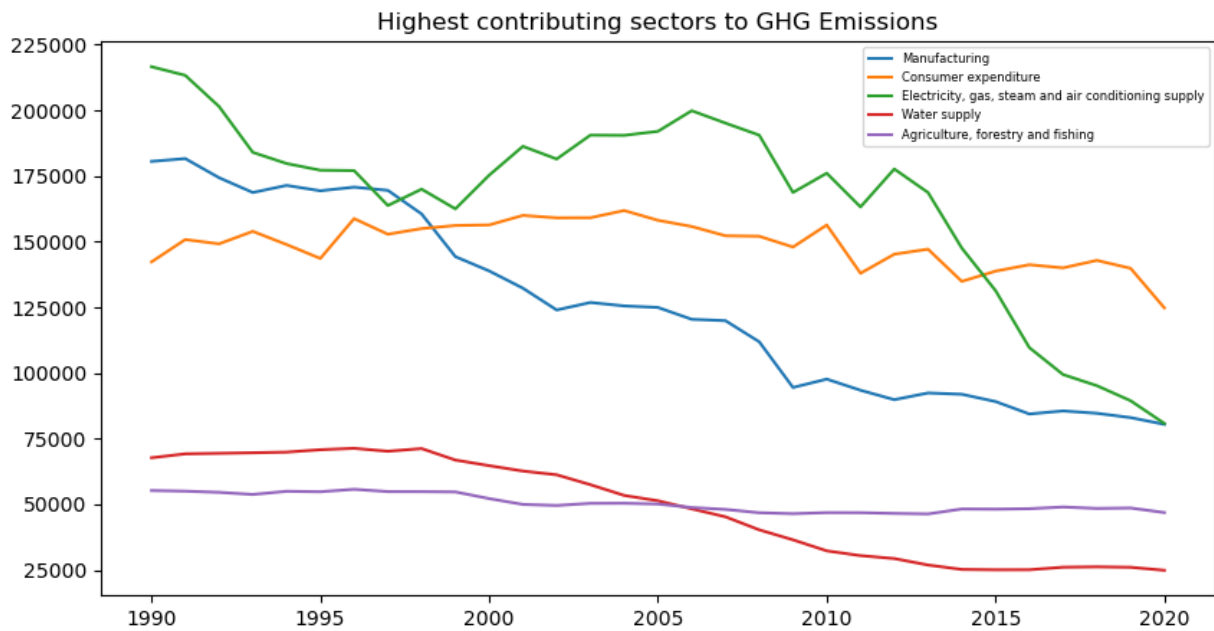


The data used was the current data set on the greenhouse emissions gases by industry from the Office of national statistics. The data accounted for the total greenhouse gas (GHG) emissions by all sectors and industries defined under the UK Standard Industrial Classification from 1990 to 2020.

This data came in an excel workbook with several worksheets. I worked with the total greenhouse gas (GHG) emissions sheet. I began by loading the GHG total sheet and then selecting the needed data from the sheet. Then I transposed the data to make it easier to work with. I confirmed the data I loaded by viewing it. Then, to be sure, I double-checked the shape and the description. I looked for null values and renamed a few very long-named columns to make them easier to visualize. The year column was changed to a DateTime format, and I also determined the average percentage change for each sector between 1990 and 2020.



The graph above displays the total yearly emissions of greenhouse gases from 1990 to 2020. The highest reduction since 1990, measured in percentage terms, occurred in 2020, with a 12.5% decrease, followed by 2009 with an 8.2% decline from 2008. The UK has been able to reduce its overall GHG emission by roughly 47% between 1990 and 2020. This signifies that there is a good chance for the net zero target to be achieved in the 30 years between 2020 and 2050.

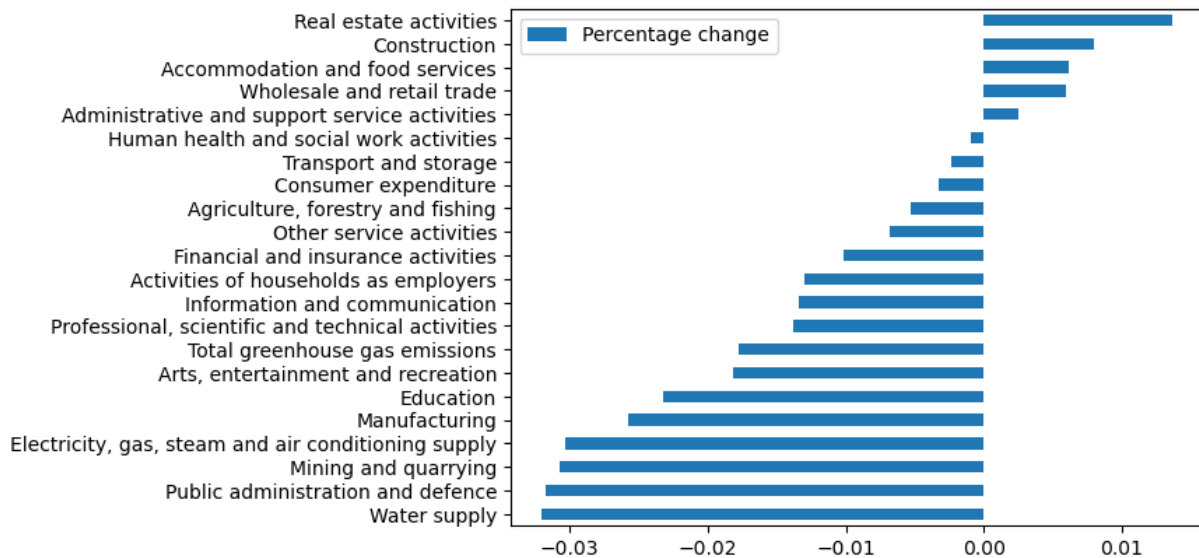


The sectors that have contributed the most to GHG emissions over time are shown in the graph above. We can see that overall, there has been a reduction in GHG emissions. While electricity and gas emissions have been significantly declining, consumer expenditure, which represents emissions from households, has been the highest in recent years.

I applied analytical methods to carry out some descriptive analyses. I looked at the average, the standard deviation, and the annual percentage change of the data set. This was done to examine the industry sector's trend over the years.

The analysis of the mean revealed that the electricity supply industry has, on average, produced higher greenhouse gas emissions than other industry sectors. Consumer expenditure came next and then manufacturing, transportation, and agriculture, in that order. The least amount of GHG emissions was created by activities of households as employers, financial and insurance operations, real estate activities, and information and communication activities respectively.

From the standard deviation, the electricity supply industry was found to be less consistent in their GHG emission reduction, whereas activities of households as employers were the most consistent sector in reducing GHG emission.

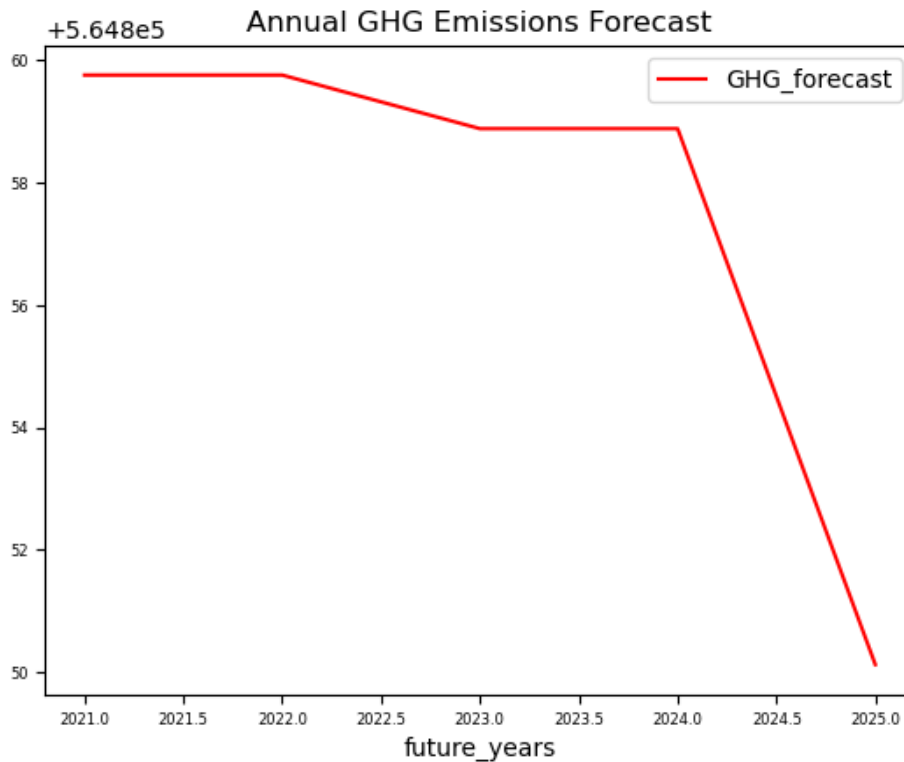


The above graph illustrates how industry sectors have reduced their GHG emissions over time. I examined the percentage change in each industry sector over the years to determine which sector had performed the best and which had performed the least, as well as to determine the overall average change. The sector of the water supply has the lowest percentage of change overall. This is followed at close range by the public administration and defense industry sector. The least-performing industry over time has been real estate activities. The real estate industry sector's overall growth over time has been 1.3%. The construction sector came in second, with an overall growth of around 0.7% throughout the 30 years we have studied.

I chose the time series linear regression model to investigate how GHG emissions in the UK have changed over time. Time series regression is a statistical method for predicting future results based on historical data. I chose this linear regression machine learning method because it can be applied to time series data to produce future predictions. Additionally, since the data I am working with are numerical and not categorical variables, linear regression seemed appropriate. Lastly predicting the value of a dependent variable based on the values of independent variables over time is best done with Linear regression time series forecasting.

An assumption made by this model is that the industry sector will cut its GHG emissions by 2%, which is a rounded-up value from the overall total GHG emission percentage reduction throughout the years (1.7%). The model also makes the following additional assumptions: that the dependent and independent variables are linear; that the data is normally distributed; and that there is no link between any two observations.

I was able to predict the yearly GHG emission for the following five years using the time series linear regression model with Xgboost. Below are the model's forecasted values.



According to my model, greenhouse gas emissions in the UK would be reduced overall. This suggests that the UK is gradually transitioning to a low-carbon economy. However, the model shows that the movement is sluggish and inconsistent. The implication for the UK government is that achieving its net zero goal of greenhouse gas (GHG) emissions by 2050 will be quite difficult. This is because underlying factors affect the overall rate at which most companies are reducing their GHG emissions. Therefore, to achieve the net zero goal, the UK government must increase attempts to develop more practical and efficient ways to significantly reduce GHG emissions by increasing energy efficiency.

One way to improve my model, will be finding and eliminating outliers from the data set. Additionally, adding more data, most likely a monthly batch of GHG emission data, to the model. More data will help the model learn faster and operate more effectively. Lastly, enhancing the hyperparameters—such as learning rate, number of estimators, etc.—can increase the model's accuracy.