TECHNICAL EXERCISE

DATA SCIENCE PROJECT

16 JANUARY 2025

JESUBUKADE EMMANUEL AJAKAYE
ajakaye.bukade@gmail.com | 07884489099

Table of Contents

Contents

List of Figures

Executive Summary

Countries.seeks.to.achieve.sustainable.environments.and.reduce.the.impact.of.climate.change.in.our.society¡.Therefore?this.technical.exercise.seeks.to.analyse.the.trends.in.greenhouse.gas.emissions.by.industries¡.Kendall.Tau.and.Linear.Regression.models.were.used¡.Findings.revealed.a.general.decreasing.trends.of.greenhouse.gas.emissions.in.the.UK¡. Decreasing. trends. in. industries. such. as. Manufacturing? Public. administration. ™. security. and. Mining. and. quarrying. while. increasing. trends. was. found. in. Real. estate. activities?Construction.and.Accommodation.™ .food.services¡.The.UK.government.must.continue. to. implement. policies. and. brings. reductions. in. the. emission. of. greenhouse. gases¡

Report Word Count: 873

Introduction

Countries seeks to achieve sustainable environments and reduce the impact of climate change in our society. Human activities are the most contributors to global warming which has sometimes made controlling of natural disasters more difficult. The greenhouse gas emissions are gases released into the atmosphere which traps heat and contributes to climate change

Paris Agreement

The Paris Agreement which is legally binding on all participating countries covers seven gases which are known as direct greenhouse gases, and they are the largest contributors to global warming [1]. See [2] for list of greenhouse gases. The Paris Agreement develops an accountability framework which helps participating countries to communicate actions to be taken to reduce greenhouse gas emissions which underscores the importance of data science process [1]. It is therefore imperative for governments to track and analyse the emissions of these gases to develop policies towards a safer and sustainable environment.

Technical Implementation

Data Information

The dataset used for this technical exercise was the Atmospheric emissions: greenhouse gases by industry and gas. This dataset has 10 sheets, but the technical exercise focused on the "GHG Total" sheet and extracted the first table for analysis. The extracted dataset has 26 columns (Year, Industries label A to T and 3 Consumer expenditure information) and 34 rows (Years starting from 1990 - 2023).

Data Preparation and Visualization

The aggregate atmospheric emissions of all greenhouse gases by industry were extracted. This table was then imported into a Jupyter notebook (Python environment). Datetime was made index for a timeseries visualization and Explorative Data analysis was carried out. (See Code book for visuals).

Model Development

Kendall Tau Statistics was favoured for model development because it can effectively reveal the direction (increasing or decreasing) and strength of association between time trends and variable been measured [3]. For a further investigation and forecasting, a linear regression model was fitted as it works well on small dataset. Functions were written in a way that are reuseable for all the tables in the dataset.

Technical Discussion

Findings

The tau model revealed significant trends in the greenhouse gas emissions except in industries such as Transport & storage and Human health and social work activities ($p > 0.05$). Generally, in the UK, the total greenhouse gas emissions ($tau = -0.93$, $p < 0.01$) revealed a decreasing trend over time.

Manufacturing, Public administration & defence and Mining & quarrying ($tau = -0.95, -0.94, -0.91$ respectively $p < 0.01$) are the top industries with decreasing trends overtime. While Real estate activities, Construction and Accommodation & food services ($tau = 0.66, 0.65, 0.55$ respectively $p < 0.01$) are top industries with increasing trends in greenhouse gas emissions overtime.

	tau	pvalue
Agriculture, forestry and fishing	-0.668449	0.000000
Mining and quarrying	-0.910873	0.000000
Manufacturing	-0.946524	0.000000
Electricity, gas, steam and air conditioning supply	-0.604278	0.000001
Water supply; sewerage, waste management and remediation activities	-0.850267	0.000000
Construction	0.654189	0.000000
Wholesale and retail trade; repair of motor vehicles and motorcycles	0.454545	0.000157
Transport and storage	-0.005348	0.964527
Accommodation and food services	0.547237	0.000005
Information and communication	-0.704100	0.000000
Financial and insurance activities	-0.542373	0.000007
Real estate activities	0.661319	0.000000
Professional, scientific and technical activities	-0.782531	0.000000
Administrative and support service activities	0.308378	0.010329
Public administration and defence; compulsory social security	-0.935829	0.000000
Education	-0.809991	0.000000
Human health and social work activities	-0.058824	0.624695
Arts, entertainment and recreation	-0.529412	0.000011
Other service activities	-0.390374	0.001168
Activities of households as employers	0.178412	0.138179
Consumer expenditure [note 4]	-0.429590	0.000353
Consumer expenditure - Not travel	-0.593583	0.000001
Consumer expenditure - Travel	0.269162	0.025189
Total greenhouse gas emissions	-0.925134	0.000000

Figure 1: Tau Model Metrics

Further investigation with the linear regression model confirmed the results from the tau model. Total greenhouse gas emissions (r2 = 0.93, mape = 0.04) showed high confidence in the predictions. Also, Public administration & defence (r2 = 0.97, mape = 0.06), Mining and quarrying (r2 = 0.95, mape = 0.08) and Manufacturing (r2 = 0.95, mape = 0.06) are industries showing high confidence in prediction which is a decreasing trends of greenhouse gas emissions.

	mean	mape	r2_score	gradient	intercept	mae	mse	rmse
Agriculture, forestry and fishing	52,441.43	0.017516	0.744045	-196.45	446,616.10	916.15	1,277,805.52	1,130.40
Mining and quarrying	32,031.39	0.080842	0.950072	-1,120.27	2,279,848.23	2,182.24	6,347,963.04	2,519.52
Manufacturing	117,467.68	0.061816	0.948561	-3,482.43	7,104,958.74	6,379.49	63,298,572.71	7,956.04
Electricity, gas, steam and air conditioning supply	159,317.21	0.151760	0.686669	-3,577.78	7,338,124.02	21,264.53	562,189,565.65	23,710.54
Water supply; sewerage, waste management and remediation activities	50,961.53	0.129252	0.922886	-2,124.64	4,314,055.74	5,297.35	36,304,089.24	6,025.29
Construction	9,685.13	0.043473	0.659502	73.55	-137,888.67	420.05	268,806.14	518.47
Wholesale and retail trade; repair of motor vehicles and motorcycles	14,238.31	0.078156	0.289343	90.40	-167,139.73	1,115.66	1,931,699.37	1,389.86
Transport and storage	83,954.83	0.108699	0.007022	91.56	-99,762.31	8,487.44	114,100,815.82	10,681.80
Accommodation and food services	3,797.34	0.060569	0.596028	34.42	-65,259.06	230.60	77,270.81	277.98
Information and communication	1,063.76	0.058340	0.777860	-16.39	33,944.21	63.22	7,381.13	85.91
Financial and insurance activities	297.61	0.038493	0.490491	-1.56	3,429.62	11.35	243.61	15.61
Real estate activities	889.17	0.048801	0.717345	9.04	-17,248.57	43.15	3,098.98	55.67
Professional, scientific and technical activities	2,195.06	0.042804	0.889207	-36.19	74,809.58	95.26	15,706.44	125.33
Administrative and support service activities	3,118.69	0.038475	0.223712	8.15	-13,225.41	118.62	22,160.42	148.86
Public administration and defence; compulsory social security	7,943.95	0.060637	0.968298	-268.56	546,816.94	394.91	227,283.91	476.74
Education	4,396.93	0.091167	0.888661	-133.32	271,896.15	391.47	214,327.37	462.96
Human health and social work activities	5,877.02	0.086847	0.010989	-6.52	18,952.90	484.50	367,891.91	606.54
Arts, entertainment and recreation	1,525.00	0.113619	0.543176	-26.79	55,288.35	187.31	58,117.05	241.07
Other service activities	993.73	0.077065	0.321852	-6.70	14,429.62	78.70	9,093.32	95.36
Activities of households as employers	90.36	0.065701	0.074511	0.22	-349.32	5.86	57.41	7.58
Consumer expenditure [note 4]	148,051.28	0.047261	0.469641	-782.55	1,718,230.20	6,877.69	66,561,811.35	8,158.54
Consumer expenditure - Not travel	82,855.46	0.055364	0.696000	-847.02	1,782,399.90	4,522.37	30,161,364.52	5,491.94
Consumer expenditure - Travel	65,195.81	0.050999	0.024012	64.48	-64,174.36	3,229.77	16,263,092.13	4,032.75
Total greenhouse gas emissions	700,337.41	0.036075	0.932908	-11,472.81	23,720,529.19	25,031.92	911,108,081.98	30,184.57

Figure 2: Linear Regression Model Metrics

The linear regression forecasted decreasing values for total greenhouse gas emissions over the next five years.

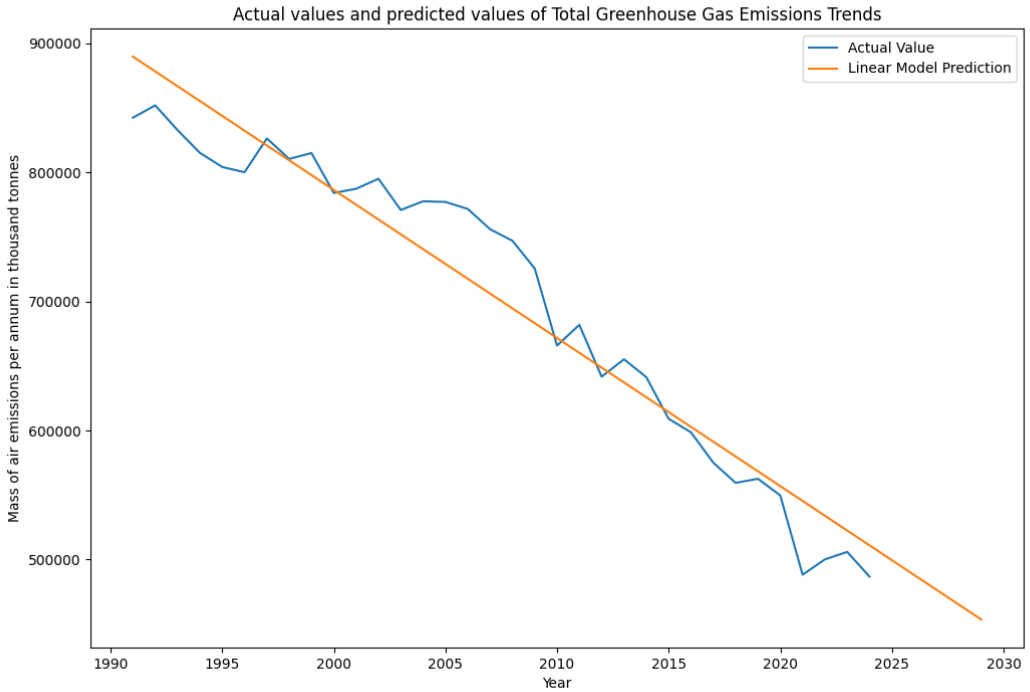

Figure 3: Chart of Actual and Predicted Total Greenhouse Gas Emissions

The values forecasted in Figure.0 are expected values of total greenhouse gas emissions at the end of each year with a +/- 4% upper and lower bound values (mape = 0.04)

	Total greenhouse gas emissions
2024-12-31	499563.249733
2025-12-31	488090.440474
2026-12-31	476617.631215
2027-12-31	465144.821956
2028-12-31	453672.012697

Figure 4: Forecasted Total Greenhouse Gas Emissions

Implications and Recommendations for UK Government

The UK government is making progress in achieving a net zero greenhouse gas emissions by 2050 but still need to take considerable steps to actualize this as the projections showed decreasing trends but not steep to achieve net zero in the next 25 years, hence, the following recommendations:

- Policies should be put in place for industries showing increasing trends.
- Incentivise industries achieving decreasing trends to continue.
- Quarterly monitoring of progressing (greenhouse gas emissions) to include seasonality in the dataset which will improve model performance and generate new insights.

Drawbacks on Model

The linear regression model used for forecasting the greenhouse gas emissions returned values of a linear slope. This can be improved by adding other factors to the model to improve and predict values. Also models such as support vector regression (SVR) can also be used for model development.

Areas for Further Research/Analysis

My technical exercise modelling and reporting focused and concentrated on the Total greenhouse gas emissions in all industries and overall, in the UK. In-depth analysis can be carried out on sub-industry groups to understand which sub-industry contributes more to greenhouse gas emissions. Furthermore, analysis can be carried out on each of the seven gases to understand the trends in each for better policy implementation around reduction of greenhouse gas emissions in the UK.

Conclusion

Achieving a sustainable environment in our ecosystem is paramount and activities of all stakeholders must be towards ensuring a safer climate. Data science ensures that relevant tools are applied to predict and develop insights in the near future. This technical exercise is a glimpse of what we can achieve using data analytics in various industries.

References

[x] United Nations Climate Change. The.Paris.Agreement. unfccc.int [Online]. Available: https://unfccc.int/process-and-meetings/the-paris-agreement [Accessed: 16 January 2025]

[y] National Atmospheric Emissions Inventory. Greenhouse. Gases. naei.energysecurity.gov.uk [Online]. Available: https://naei.energysecurity.gov.uk/greenhouse-gases [Accessed: 16 January 2025]

[z] A. Ostwal (2020, Dec. 22). Test.for.existence.of.a.Trend.in.a.Time.Series. Medium [Online]. Available: https://towardsdatascience.com/test-for-existence-of-a-trend-in-a-time-series-3a44f242c329 [Accessed: 16 January 2025]