Project Documentation Report

"Work Green: Sector responses to Climate Change as time spent on Green Jobs from 2011–2018"

Contents

INTRODUCTION:

About Group 4

Group 4 consists of three CFG-sponsored Nanodegree students, who are on the Data 2 Spring 2022 cohort. We are:
Angela YT Chan, independent artist, curator and researcher specialising in climate change
Catherine Miao, Data Analyst

Ana Diane Walia, IT Project Management Office Coordinator / Web Tester / Admin Support

Project Aims

The overall aim of this project, titled "Work Green: Sector responses to Climate Change as time spent on Green Jobs from 2011-2018", is to collaborate together as a team to explore and analyse our selected datasets to produce an informative and innovative approach to current climate change issues by implementing the data analysis tools and frameworks learned on the CFG Nanodegree.

Project Objectives

To achieve this, our objectives have been to firstly source reliable and suitable data that would form and answer a meaningful data analysis question, before using data analysis tools like Numpy and Pandas to read and clean the datasets for appropriate and efficient use in our analyses. Exploring the data, we deployed data visualisation tools like Matplotlib to answer the questions we set for our project. A further objective of our project has been on deploying our soft skills relating to collaborative working, such as project and time management, clear communication, imaginative and critical thinking, and getting to know the range of domain knowledge , technical expertise and soft skills in order to fully reach the potential of group working.

Roadmap of the report

The roadmap of this report covers the project background relating to our choice on the project concepts; the steps specifications that detail the project's technical tasks; the development process in its implementation and execution; the result reporting and finally the overall conclusion of our experience of the group project.

BACKGROUND

Project details and our chosen topic

Initial group meetings aimed to mindmap our collective interests and programming ambitions to create an outline for our final project concept and production. We shared enthusiasm in using sentiment analysis (such as using consumer reviews, sales insights, news data and social media insights like a Facebook or Twitter API) to develop a programme to identify public understandings of climate changeover time. With this, we were considering the global variances in public climate understandings, but found language barriers in the data sourcing, as well as difficulties with open data access (political news censorship). Next, we thought about framing our project on how the British media has portrayed global climate issues over the recent years. However, we reached similar challenges in sourcing the adequate data to further refine our question.
After further individual research on climate themes and further discussion in the group forum (Slack) and video call meetings, we settled on framing a data analysis project on a dataset relating to how UK sectors have been investing their labour forces and their working hours on 'green tasks' over the years. We believed this data – matched with global greenhouse gas (GHG) emissions data, UK sector specific emissions throughout these years and occupation (seniority) levels of the labour force involved in the 'green tasks' – could give a valuable insight to how different sectors are responding to climate change and how these time and green job investments were impacting GHG emissions. Although we sourced and wished to use Environmental Social Governance (ESG) ratings and corporate stock history to illustrate the economic impact of sectors 'green tasks' over time (ie. "What are the ESG

ratings of the best and worst performing companies in specific/key industries, and how does this relate to their industry's overall ESG rating?"), we agreed to eliminate this in order to focus our questions to the relations between sectors, GHG emissions, 'green tasks' and occupation levels.

Explain the nature of your analysis

Our Descriptive Analysis project explores the investments in sustainability by the sectors across the UK through Green Tasks between 2011 to 2018, taking into account the types of occupational roles involved, as well as the changing GHG emissions by sector. With this, the project aims to offer a substantial foundation from which researchers may frame and develop further investigations and theses relating to the correlation between these three variables.

Our main question was finalised as **"How does the time spent on Green Tasks by different industries correlate with their contributions to GHG emissions between 2011-2018 in England, Wales, Scotland and Northern Ireland?"**

With secondary questions to fulfil the analysis aspects we sought to answer:

Dataset: [A] Time spent on Green Tasks

- Identify the proportion of hours worked spent doing Green Tasks across the Whole Economy
- What is the mean proportionate percentage time spent by each sector on Green Tasks between 2011-2018?
- Which 10 sectors invest the highest proportionate percentage time on Green Tasks between 2011-2018?

Dataset: [B] Atmospheric emissions: greenhouse gas emissions intensity by industry

- List of Industries and their corresponded GHG Emissions
- Mean GHG per industry over the number of years
- show which industries have most and least GHG emission
- show which year has the highest GHG from the top sectors

Dataset: [C] Estimated time spent on green tasks by occupation code

- What are the top 10 occupations that spend time on doing green tasks?
- How is the time they spend on green tasks changing from 2011-2018?
- How is the Time spent doing green tasks by country from 2011-2018?

How our findings should be used (target audience)

Our data analysis could inform a range of audiences that are directly and indirectly impacted by climate and environmental issues. These include institutions, corporations and individual professionals across the UK economy and global economies that intersect by trade and relations. Particular audiences of target include policymakers, environmental movement builders (such as activists, educators and community leaders), and working people across all sectors at different levels, many of them in professions named in the analysis.
On a more public level of engagement, the results of this analysis can also inform everyday understandings of climate change (through media such as in a news article, podcast discussion and classroom textbooks, for example).

How can your analysis help?

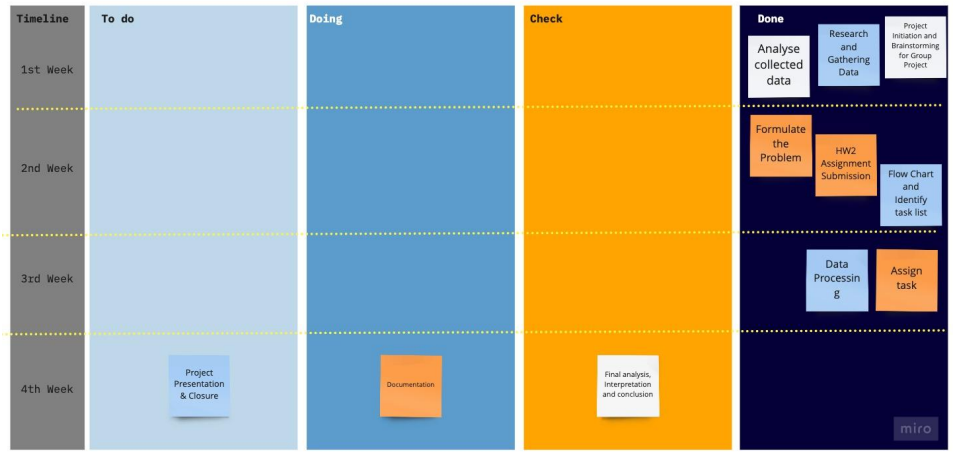
The data analysis can offer these audiences a general overview and avenues for further research on the growing impact of climate change issues on how sectors are investing in the way their employees work and if there have been impacts so far on GHG emissions as a result.

STEPS SPECIFICATIONS

Describe how your team approached each of the key steps of the data analysis: framing questions, data gathering, preprocessing, in-depth analysis etc.

- **Pre-project planning**
    1. Building the team
        - Getting to know the members
    2. Identify areas of interest options
        - Charity, Education, Digital Marketing, Environment, Business, Finance, Stock Market, Employment
    3. Brainstorming/Concept
        - Sentiment Analysis of people about Climate Change/Global Warming throughout the years
        - Descriptive Analysis of Climate Change in the actual work by industries and occupation
    4. Decide on a particular area of interest for the project
        - Climate Change: Green Task in Industries and Occupation

- **Project Planning**
    1. Formulate queries and identify problems to be solved
    2. Research and Data gathering
        - Data sources: Kaggle, Google Dataset Search, UK Office for National Statistics
    3. Assess the datasets
        - Time spent on Green Tasks by different industries https://www.ons.gov.uk/economy/environmentalaccounts/datasets/timespentongreentasks
        - Their contributions to GHG emissions between 2011-2019 https://www.ons.gov.uk/economy/environmentalaccounts/datasets/ukenvironmentalaccountsatmosphericemissionsgreenhousegasemissionsintensitybyeconomicsectorunitedkingdom
        - Occupational level and time spent on Green Tasks https://www.ons.gov.uk/economy/environmentalaccounts/datasets/estimatedtimespentongreentasksbyoccupationcode
    4. Identify Scope:
        - Occupational level involved in Green Task
        - Industries/Sectors
        - GHG Emissions
        - 2011-2019 time period
        - Location: UK
    5. Kanban board: We use Kanban Board to track our tasks and set deadlines.
        - https://miro.com/app/board/uXjVO4YuBgA=/



6. List stories/backlogs
7. Assign Tasks
    - We divide the task by processing each datasets per person and we coordinate each other.

- **Development**
    1. Actual coding
        - Import Libraries: Pandas, Numpy, Matplotlib
        - Create and read data frame
        - Cleaning data
        - Process and Finding results
        - Visualization
    2. Data Analysis and Interpretation
    3. Upload documentation report, coding scripts, datasets and presentation: https://github.com/angelaytc/Group4_Project

- **Delivery**
    1. Project Presentation
    2. Documentation
    3. Define Terminologies

Describe your data sources and how you found them.

We found our dataset on Office For National Statistics.
Datasets:

[A] Time spent on Green Tasks
[Table_6] Worksheet: Proportion of hours worked spent doing green tasks, by industry, percentage (%)
[Table_7] Worksheet: Proportion of workers doing green tasks, by industry, percentage (%)
https://www.ons.gov.uk/economy/environmentalaccounts/datasets/timespentongreentasks
[B] Atmospheric emissions: greenhouse gas emissions intensity by industry
[GHG intensity] Greenhouse gas emissions intensity 1, 2 by industry, 1990 to 2018
https://www.ons.gov.uk/economy/environmentalaccounts/datasets/ukenvironmentalaccountsatmosphericemissionsgreenhousegasemissionsintensitybyeconomicsectoruntedkingdom
[C] Estimated time spent on green tasks by occupation code

[SOC_2010] Worksheet: Estimated share of time spent on green tasks, by SOC 2010 code, percentage (%)
https://www.ons.gov.uk/economy/environmentalaccounts/datasets/estimatedtimespentongreentasksbyoccupationcode

IMPLEMENTATION AND EXECUTION

Development approach and team member roles

We distributed the individual tasks evenly to ensure we each experience and learn to a similar degree (such as sourcing research, datasets and writing small segments of code), while being open to assist each other, as we would in our collaborative tasks (such as grouping our research and code). In terms of writing the report, we gathered our individual dataset findings to this document together, with Angela specifically writing the introduction and background pages too.

Tools and libraries

In terms of coding, we used Google Colab and Jupyter Notebook to co-write our programme.

- Pandas
- Matplotlib
- Numpy
- Seaborn

Implementation process

As aforementioned, we took time to clarify our dataset choices and the data analysis question framing at the start of this project. While it may be seen as indecisive at first or creating too broad an initial question, we as a group believe it was still advantageous for us to have spent time on domain research by looking into the various potential datasets online. Our decisions to change our questions and form new subquestions, after tutor feedback (in the form of the homework) offered us clarity on the specific coding tasks at hand. By separating our responsibilities across the group by dataset, we were able to each understand our datasets well to work with them and communicate our experiments, findings and challenges with each other. As we are only three people in the group, we honed our question to explore the data with more potential questions in mind, rather than seek to achieve answers to all questions we had posed informally throughout the project.

Agile development

The success of our project was based on good communication. While we updated each other through informal messages on Slack as well as scheduled weekly team videocall meetings, we also managed and organised a scrumboard of our tasks and a shared online drive of all documents, datasets and research references.
In terms of agile elements, we reviewed each others' code and shared advice on simplifying and neatening our writing.

Each week we reflected on what we did well and what we could do better. We made sure everyone is contributing, by assigning tasks to each member and helping each other with problems that we are encountering.

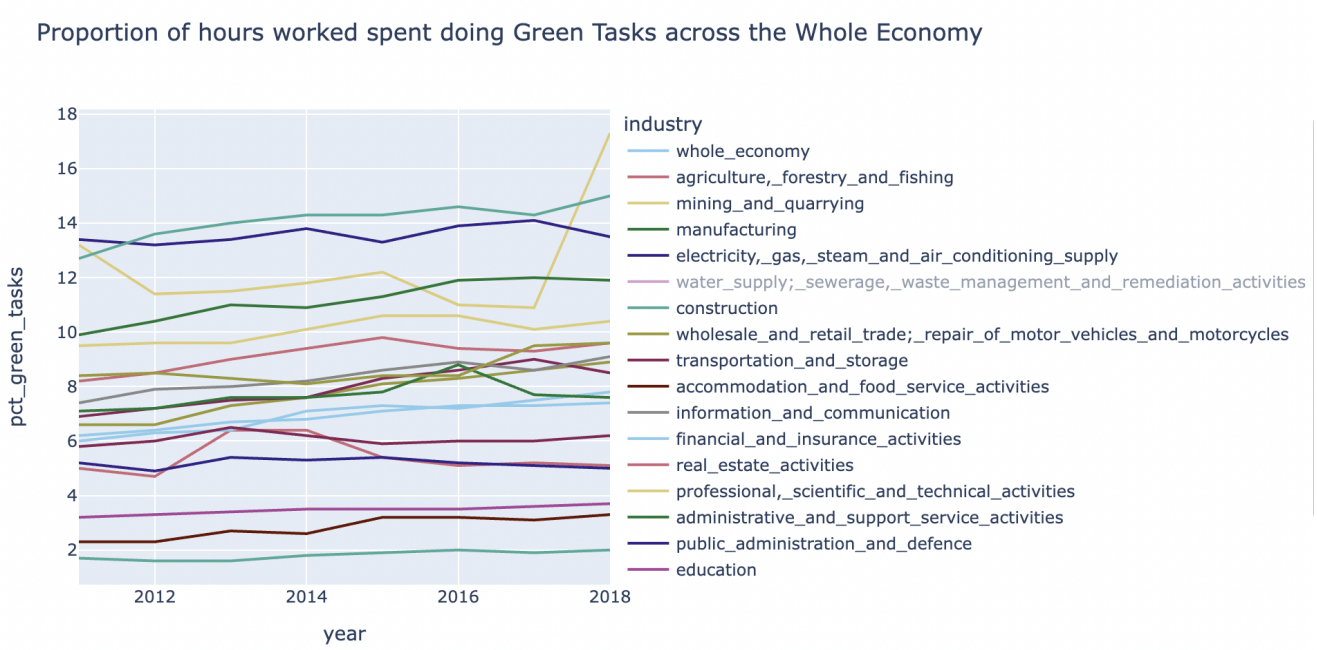
Implementation challenges

The key challenge we had is trying to ask the right question, and refine the question into a few small ones, setting clear goals and objectives.

RESULT REPORTING

What are the key findings of your project?

Identify the proportion of hours worked spent doing Green Tasks across the Whole Economy



The sector with the highest mean throughout the time range is ‘Water supply; sewerage, waste management and remediation activities’, followed by ‘Construction’ and ‘Electricity, gas, steam and air conditioning supply’, and the top ten sectors are shown below.

```
Water supply; sewerage, waste management and remediation activities    15.5
Construction                                                           15.0
Electricity, gas, steam and air conditioning supply                    14.1
Mining and quarrying                                                   17.3
Manufacturing                                                          12.0
Professional, scientific and technical activities                      10.6
Real estate activities                                                  9.8
Information and communication                                           9.1
Transportation and storage                                              9.0
Wholesale and retail trade; repair of motor vehicles and motorcycles    8.9
dtype: float64
```

GHG Emission per Industry in UK (2011-2018)

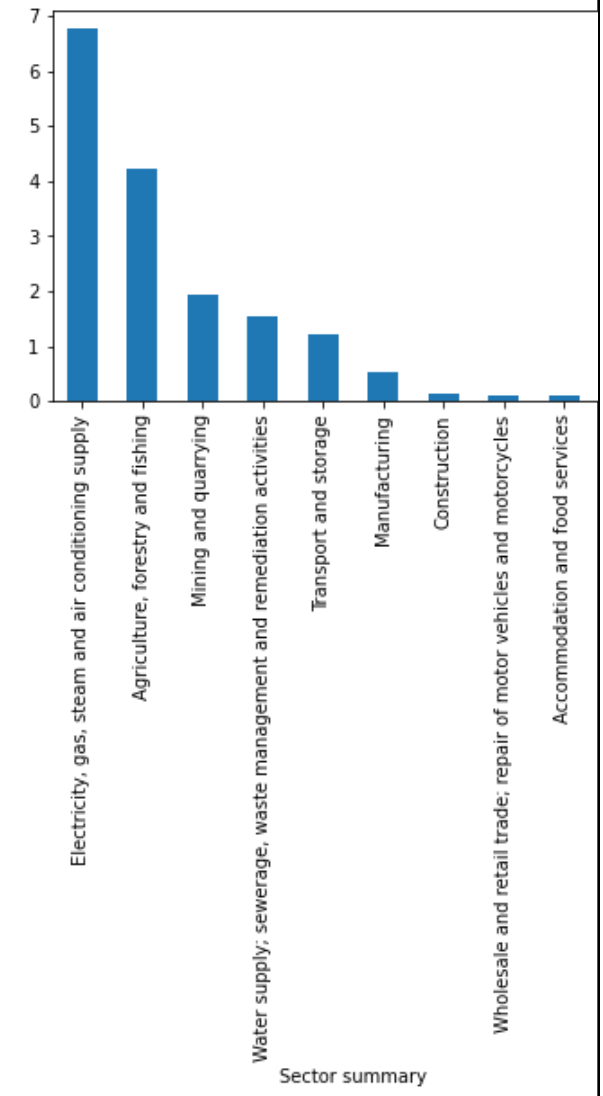
Sector summary	2011	2012	2013	2014	2015	2016	2017	2018	Mean Value
Electricity, gas, steam and air conditioning supply	6.17	6.77	6.46	5.90	5.14	4.14	3.82	3.72	5.26500
Agriculture, forestry and fishing	3.87	4.21	4.16	3.79	3.74	4.01	3.83	3.93	3.94250
Mining and quarrying	1.81	1.93	1.91	1.92	1.94	1.88	1.91	1.81	1.88875
Water supply; sewerage, waste management and remediation activities	1.55	1.48	1.28	1.18	1.12	1.04	1.06	1.07	1.22250
Transport and storage	1.23	1.16	1.10	1.11	1.13	1.15	1.07	1.10	1.13125
Manufacturing	0.52	0.51	0.53	0.51	0.49	0.46	0.46	0.45	0.49125
Construction	0.12	0.14	0.13	0.12	0.13	0.12	0.12	0.12	0.12500
Wholesale and retail trade; repair of motor vehicles and motorcycles	0.11	0.11	0.11	0.10	0.10	0.10	0.09	0.09	0.10125
Accommodation and food services	0.08	0.09	0.09	0.08	0.08	0.08	0.07	0.07	0.08000
Public administration and defence; compulsory social security	0.07	0.07	0.06	0.06	0.06	0.06	0.06	0.06	0.06250
Administrative and support service activities	0.05	0.05	0.04	0.04	0.04	0.04	0.04	0.04	0.04250
Human health and social work activities	0.04	0.05	0.05	0.04	0.04	0.04	0.04	0.04	0.04250
Arts, entertainment and recreation	0.04	0.04	0.04	0.03	0.03	0.03	0.03	0.03	0.03375
Other service activities	0.03	0.04	0.04	0.03	0.03	0.03	0.03	0.03	0.03250
Education	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03000
Professional, scientific and technical activities	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.01	0.01875
Activities of households as employers; undifferentiated goods and services-producing activities of households for own use	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01000
Information and communication	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01000
Real estate activities	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00000
Financial and insurance activities	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00000

10 Industries with the most GHG emissions:

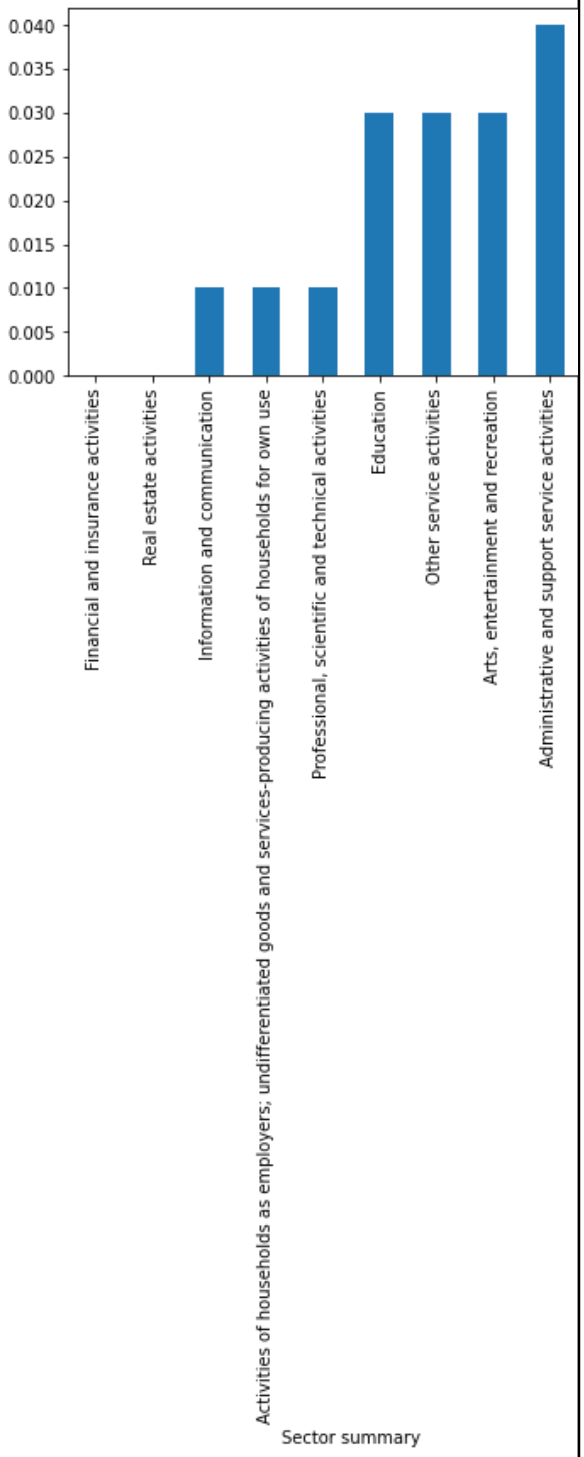
- Electricity, gas, steam and air conditioning supply
- Agriculture, forestry and fishing
- Mining and quarrying
- Water supply; sewerage, waste management and remediation activities
- Transport and storage
- Manufacturing
- Construction
- Wholesale and retail trade; repair of motor vehicles and motorcycles
- Accommodation and food services
- Public administration and defence; compulsory social security

10 Industries with the least GHG emissions:

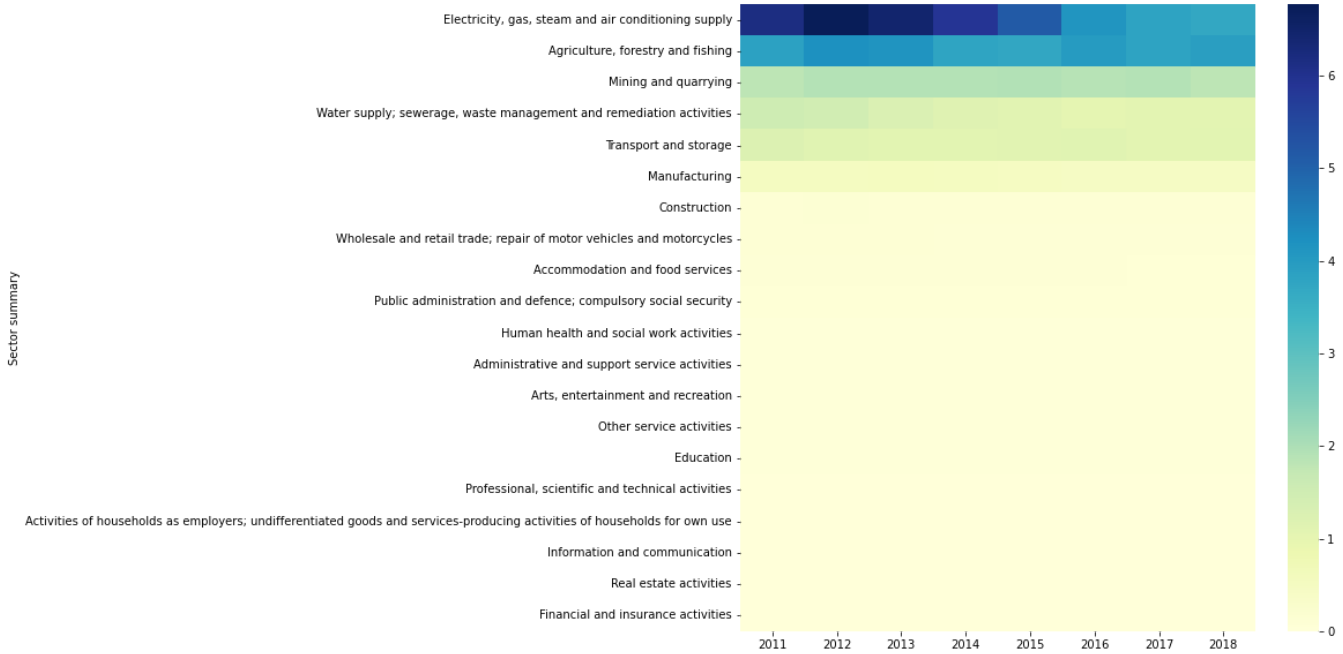
- Financial and insurance activities
- Real estate activities
- Information and communication
- Activities of households as employers; undifferentiated goods and services-producing activities of households for own use
- Professional, scientific and technical activities
- Education
- Other service activities
- Arts, entertainment and recreation
- Human health and social work activities
- Administrative and support service activities



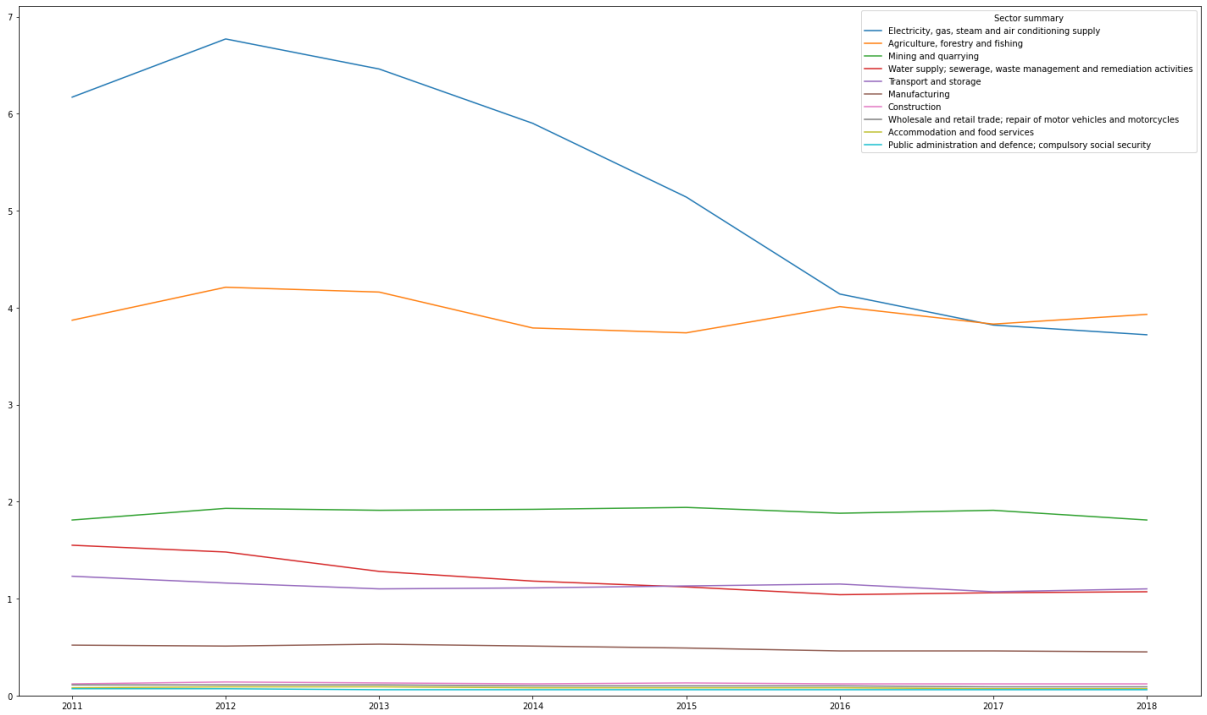
The bar graph demonstrates the industries with the highest mean values.



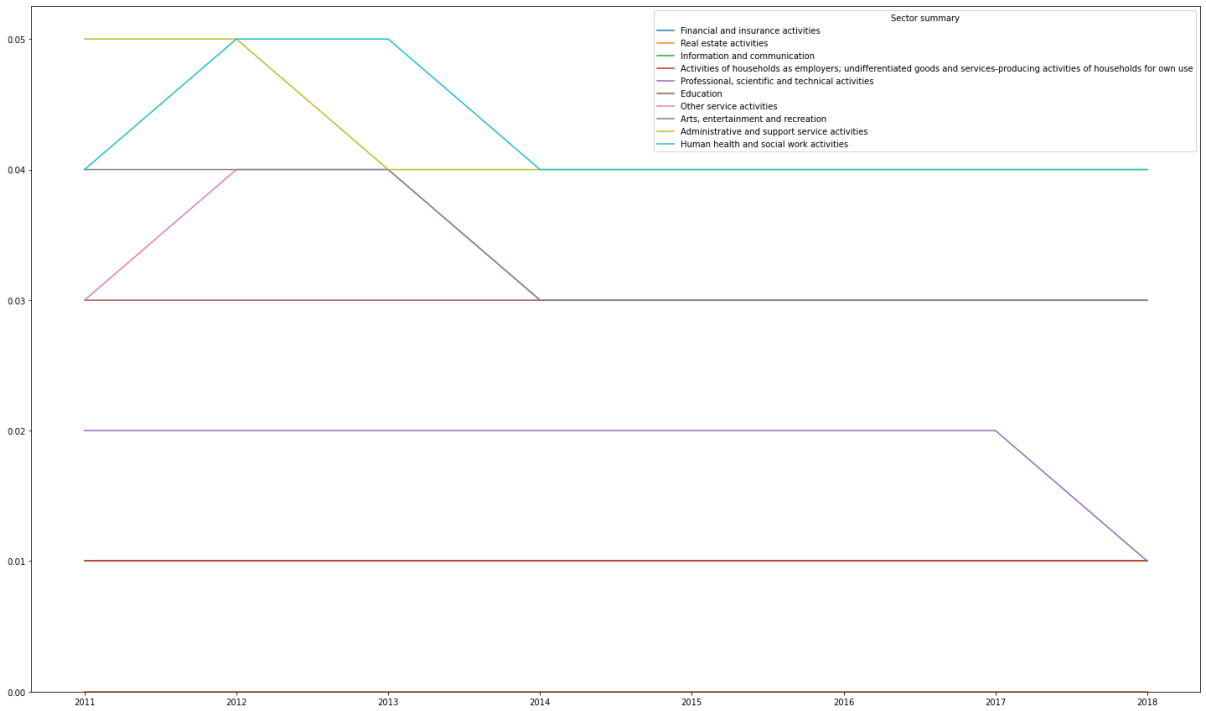
The bar graph demonstrates the industries with the lowest mean values.



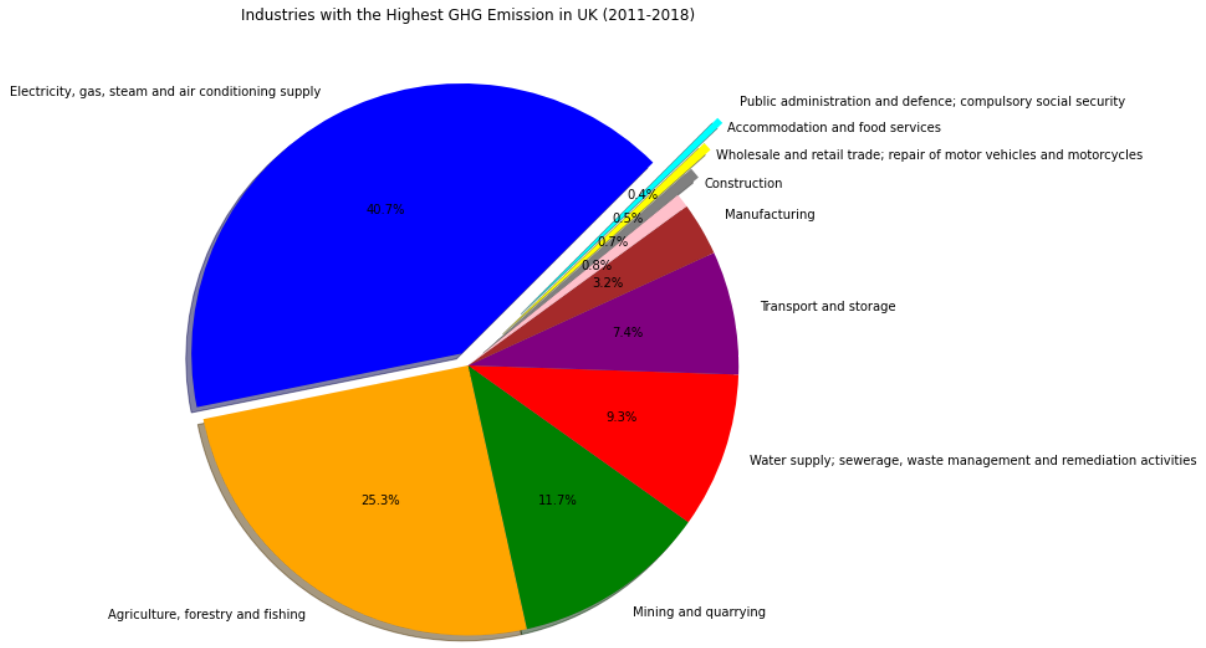
This graph provides information on GHG emissions by different industries in the UK from 2011 to 2018. The data is represented by colors, with darker colors indicating higher emission values and lighter colors representing lower emission values. The graph shows that the electricity, gas, steam and air conditioning supply industry had the highest GHG emissions in 2012. However, there has been an overall trend of reducing GHG emissions across all industries in recent years.



Plotting the industries with the highest GHG emission in the UK from 2011-2018.



Plotting the industries with the lowest GHG emission in the UK from 2011-2018.



The pie chart below reveals which industries are responsible for the most GHG emissions.

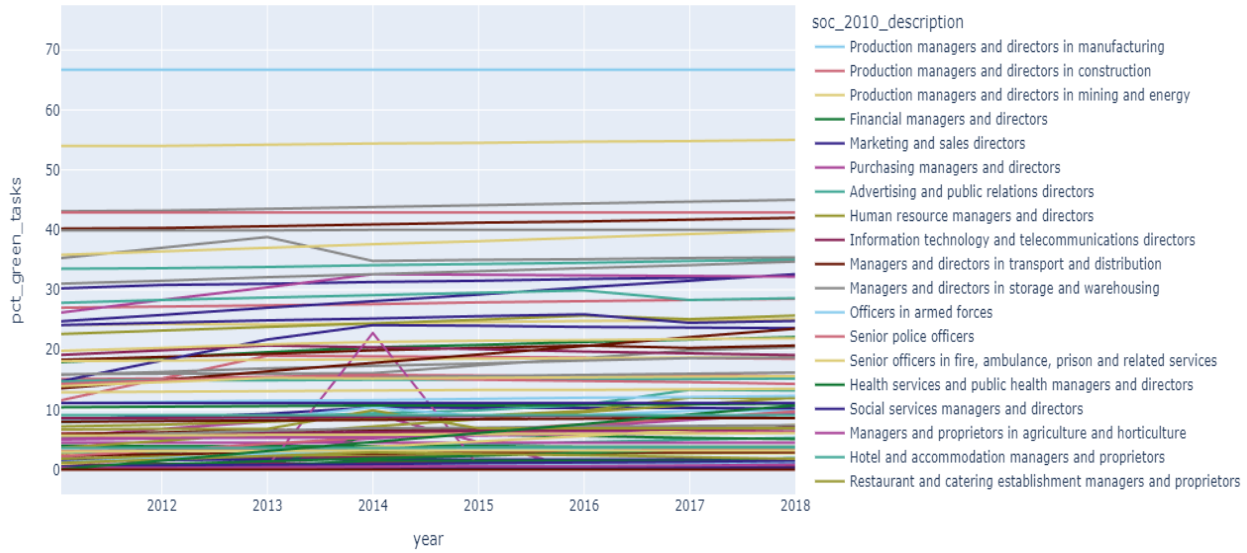
Occupation level green task from 2011-2018:

- The majority of job categories (70%) from 2011 even to 2018 have 0% of working time spent on green related activities Of those job categories (the remaining 30%) that do some green related tasks, only 6% of job categories spent more than a third of time doing green related tasks, this general pattern has held over 20 years.

	count	mean	std	min	0%	10%	20%	30%	40%	50%	60%	70%	80%	90%	92%	94%	96%	98%	100%
2011	369.0	5.733604	11.988283	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	2.06	8.48	21.50	26.604	33.30	36.764	43.1	66.7
2012	369.0	5.903252	12.103547	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	2.56	8.68	21.82	27.368	33.30	37.224	43.2	66.7
2013	369.0	6.076152	12.246598	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	2.92	8.92	22.60	27.568	33.30	38.800	43.5	66.7
2014	369.0	6.327642	12.335097	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	3.48	10.40	24.16	28.012	33.30	37.796	43.8	66.7
2015	369.0	6.261247	12.347990	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	3.26	9.52	24.12	28.380	33.30	38.240	44.1	66.7
2016	369.0	6.312195	12.422168	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	3.12	10.46	24.00	28.424	33.30	38.728	44.4	66.7
2017	369.0	6.378049	12.466692	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	3.26	10.98	23.86	28.356	33.30	39.156	44.7	66.7
2018	369.0	6.443089	12.535888	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	3.30	11.10	23.84	28.600	33.30	39.540	45.0	66.7

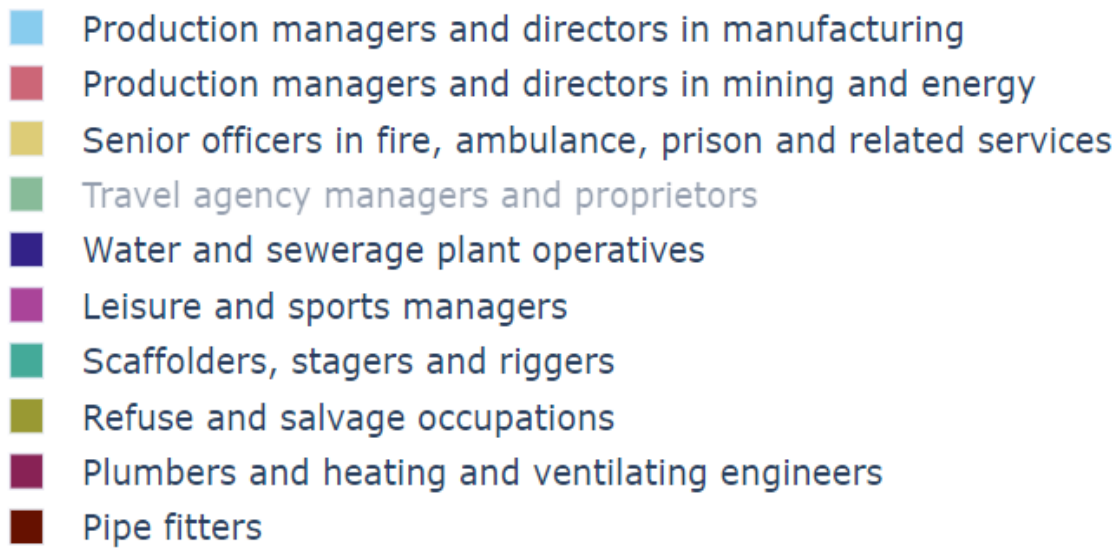
The percentage of time spending on green task changing over the year:

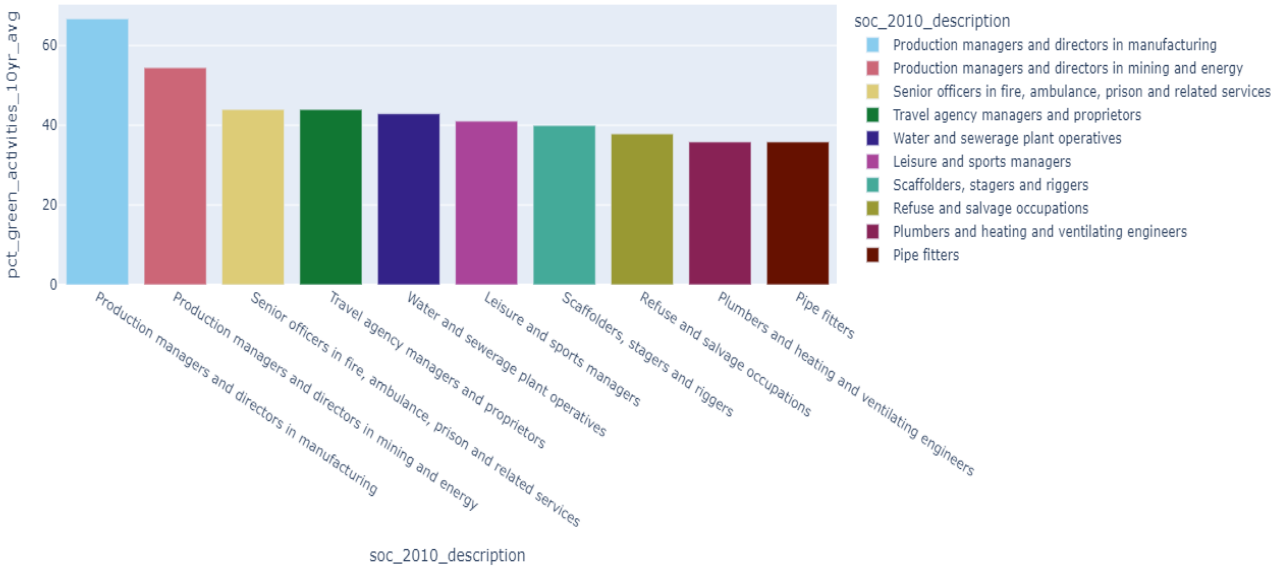
Overall the percentage is slightly increasing by each role



Top 10 workers spending time on green task:

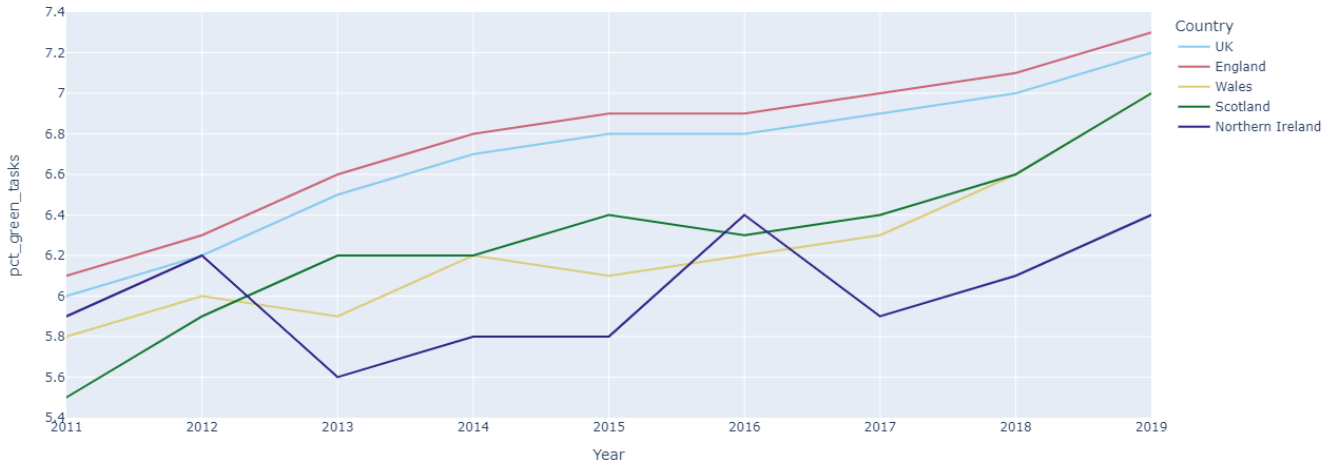
- The proportion of time spent on green tasks varies widely across industries, with production industries tending to have higher proportions.





Time spend on green task by country in the UK change over the years:

- All countries of the UK follow a similar trend in the proportion of workers and hours worked doing green tasks; England tends to have slightly higher levels of time spent on green tasks.
- England is estimated to have the greatest share of hours worked on green tasks (at 7.2%), slightly above the UK average. While estimates for Scotland, Wales and Northern Ireland are a little lower at around 7% in Scotland and Wales, and 6.4% in Northern Ireland.



CONCLUSION

In conclusion, our data analysis shows that sectors that are involved in green tasks have a positive transformation towards the green economy with an upward trend through the recent years. However, in terms of occupations 70% of job roles remain at 0% of green tasks. This may be because the data has not yet caught up on each sectors' activities on green tasks.

For occupations that are categorised as skilled jobs, there is a higher probability of there being 0% green tasks recorded. On the other hand, workers in more senior occupational roles are more likely to be working more hours in green tasks, indicating a potential for them to assist in the training and education of lower level workers. With this, sectors could be identified to have a high green potential and be advised to invest in green education for the development of more green roles.

There is still much more room for expansion on these questions to focus more on exploring the specific reasons behind these results, as well as using predictive modelling with these datasets to anticipate the whole economy's potential next steps. However, this project has brought together business, finance, environmental science and labour under the greater umbrella of the green economy. This has increasingly become an important stage for all sectors to cross, which now face the urgent costs (from financial to social) that climate issues are steering towards a greener way of sustaining businesses and their workforces (see the Green Transition Model, where sectors are turning to the green economy and normalising more sustainable practices).

For us as students building this project, we feel this has been a steep learning curve that has not only aligned our environmental and business domain interests with programming, but has also deepened our practical experiences and skills in seeing a project from initiation, production and reflection. In terms of collaborative working, this project group has been an overwhelmingly supportive environment, offering us each more confidence and self-patience to continue our learning and expansion of our potential in the field.