Interactive Webpage to Communicate Greenhouse Gases Emission Reduction Momentum

Linh Ngo[1], **Selin Taskin**[2], **Teemu Saha**[3], **Daniel Aaltonen**[4], **Hieu Pham**[5], **Dung Nguyen Anh**[6]

[1] Data Science, 906502, linh.l.ngo@aalto.fi
[2]Data Science, 943413, selin.taskin@aalto.fi
[3] Information and Service Management, 713216, teemu.saha@aalto.fi
[4] Data Science, 885610, daniel.aaltonen@aalto.fi
[5] Data Science, 895118, hieu.pham@aalto.fi
[6] Data Science, 906476, dung.nguyenanh@aalto.fi

Abstract

Climate change currently looks like one of humanity's biggest challenges of the century. With the development of industrial technologies along with population growth, greenhouse gases emitted into the atmosphere have increased significantly. This paper discusses the results of our work for a client who was interest in creating an interactive web page to communicate the emission reduction momentum to a predominantly Nordic audience. In this project, the emission of greenhouse gases (such as carbon dioxide) has been analyzed throughout the years, with a predominant focus on Nordic countries. Furthermore, a prediction model for the global carbon dioxide emission for the next 5 years has been built. In light of the historical data analysis, it was found that a lot of countries are taking measures by implementing policies regarding the emission of greenhouse gases, specifically, in the European Union countries, where the emission trend seems to be downwards. Even though there is a consistent increase in the total amount of greenhouse gasses emitted, the strike of the Covid pandemic inflicted a reduction in emissions, primarily in the transportation sector. Lastly, actions we can take as individuals has been discussed.

1. Introduction

This project is a part of the course CS-C3250 Data Science Project 2022 under the supervision of Jorma Laaksonen at Aalto University and Saku Suuriniemi, Reaktor's Data Scientist.

The main goal of this project is to communicate the climate change reduction momentum, with Nordic audiences as a predominant target. However, as climate change affects everyone on a global scale, we also include information about other regions in the world. Moreover, this project also demonstrates the current greenhouse gas (GHG) emission situations on different scales and shows the effects of possible actions (or lack of changes) on the reduction of greenhouse gas emissions. In order to achieve our goal, the process includes finding related data from public sources, exploring, modeling, and visualizing data in an interactive way. The result consists of various visualizations that provide insights into climate change from different perspectives and scales. Our visualizations are showcased on a public website made by our group.

The remaining of this report is structured as follows: Section 2 presents our data sources and formats as well as details necessary preprocessing methods. Section 3 outlines the construction of visualizations and the ARIMA model. Section 4 discusses in detail the findings and functionality of each website component. Section 5 discusses ethical issues regarding our project. Section 6 discusses teamwork practices and possible shortcomings/issues encountered. Section 7 discusses how we divided work and outlines specific work performed by each student. Finally, Section 8 concludes our findings and discusses future prospects.

2. Data

The following section discusses data sources used in this project, the ways data is stored, formatted, and preprocessed.

2.1. Data Sources

The data used in this project is mostly from public sources. Data about greenhouse gas (mainly CO_2) emissions of all countries in the world is gathered from EDGAR—Emissions Database for Global Atmospheric Research, which is an official website of the European Union [1]. We utilize this dataset to see how the fight against climate change is developing in different regions by getting a closer look at the amount of CO_2 emissions by capita and CO_2 emissions by sector from 1970-2021.

Another database used is European Environment Agency (EEA) database on greenhouse gas policies and measures (PaM) in Europe [2]. This database consists of a number of policies and measures implemented, adopted, or planned by European countries to reduce greenhouse gas emissions. These PaMs have been reported by European countries under the Governance of the Energy Union and Climate Action Regulation in 2021 and two countries (Germany and Iceland) updated their submissions in 2022.

To visualize the impact of different carbon pricing policy designs, we also utilize data from Resources for the Future: Carbon Pricing Calculator [3]. Resources for the Future (RFF) describes itself as an independent, non-profit research institution situated in Washington, DC. Their goal is to improve the decision-making process around environmental policy, via research and policy action.

Finally, to provide sector specific emissions scaled by GDP, a year-by-year population data set from 1960 to 2021 provided by the World Bank was used [4]. The World Bank describes itself as providing financing, policy advice, and technical assistance to governments of developing countries. They also offer a comprehensive data bank with various economic development indicators, such as population data.

2.2. Data Formats And Access

Data downloaded from online public sources is mostly in CSV and XLSX format. After downloading the data, we utilized the *pandas* library in Python and store data in the dataframe forms to better process them later on.

2.3. Data Preprocessing

As our data is mostly downloaded from trusted public sources, it comes quite clean. However, to better prepare data for analysis and visualization, we perform the following steps on our data:

- Removing unnecessary columns and adding necessary columns
- Pivoting the data from wide-format to long-format

  This step is performed to facilitate the visualizations that will be created later on. Data downloaded from EDGAR is usually stored in wide-format, in which each row in the dataset represents the variable measure of each country and each column represents the year of the variable measured. Even though in *plotly*, charts should facilitate wide formatted data, we encountered problems with our analysis. Therefore, we decided to perform this pivoting step for convenience. Figure 1 and Figure 2 show an example of the transformation of dataframe from wide-format to long-format.

	Country	1970	1971	1972	1973	1974	1975	1976	1977	1978	...	2012	2013	2014	2015
0	Aruba	1.386249	1.377678	1.594510	1.653806	1.471866	1.733329	1.585203	1.744450	1.774656	...	15.777813	15.396213	15.566530	15.409848
1	Afghanistan	0.156348	0.152203	0.146371	0.144899	0.178432	0.161841	0.147963	0.175156	0.146609	...	0.333520	0.267372	0.244712	0.247858
2	Angola	1.321143	1.232565	1.464238	1.562567	1.583252	1.422773	0.926134	1.482940	1.699974	...	0.963765	1.070159	1.129542	1.169955
3	Anguilla	0.338637	0.336122	0.348216	0.321684	0.356316	0.389807	0.365537	0.379403	0.432496	...	1.841831	1.954019	1.929322	1.916736
4	International Aviation	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN	NaN

Figure 1. Wide-format dataframe

	Country	Year	CO2_per_capita
0	Aruba	1970	1.386249
1	Afghanistan	1970	0.156348
2	Angola	1970	1.321143
3	Anguilla	1970	0.338637
4	International Aviation	1970	NaN

Figure 2. Long-format dataframe

- Categorizing countries into regions

  To understand the difference in the trends of emissions among continents in the world, we also group the countries by region. We utilized the Python library *PyCountry* to complete our country data for easy matching or aggregation with other data sets.

3. Methods

3.1. Visualization And Deployment

Our visualizations are created utilizing *plotly*—a Python plotting library for interactive visualizations. Moreover, we also utilize Dash—a low-code framework for building data apps in Python with customized user interfaces. Afterward, our website is deployed using Heroku. However, as Heroku's free plan ends at the end of November, we are considering other options: either subscribe to Heroku's basic plan or switch to other deployment websites.

3.2. Auto Regressive Integrated Moving Average Model

For our predictive model, we utilized an Auto Regressive Integrated Moving Average (ARIMA) model. ARIMA models are suitable for time series data, as is our case. In essence, an ARIMA model uses the linear combination of a realized number of past observations and random errors [5]. The preprocessing and construction of the model was conducted as presented in the lecture slides of the Aalto course MS-C2128 - *Prediction and Time Series Analysis*. The model was created using the Python library *statsmodels*.

In general, ARIMA models are presented in the format ARIMA(p, d, q), where p refers to the lag of the autoregressive (AR) part, d refers to the degree of differencing performed on the data to be modeled, while q refers to the order of the moving average (MA) part. Without seasonal elements, an ARIMA(p, d, q) model can be expressed as the following function:

$$Y_t = \phi_0 + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \cdots + \phi_p Y_{t-p} + \epsilon_t - \theta_1 \epsilon_{t-1} - \theta_2 \epsilon_{t-2} - \cdots - \theta_q \epsilon_{t-q} \tag{1}$$

where Y_t are the datapoints t of the timeseries, ϵ_t are the error terms, while p and q are the parameters of the AR and MA parts respectively.

Our chosen model was ARIMA(1,1,0). We ended up with this model via the following process: first, we stationarized our realized time series data. That is, we require our underlying data to be modeled by a stochastic random process - the mean and variance should be constant over time. By differencing the time series once, the modified time series seemed visually stationary. Next, the plots of the autocorrelation function (ACF) and partial autocorrelation function (PACF) were observed. Only the PACF plot indicated any significant spike at a lag of 1. All other lags did not significantly deviate from zero. As such, we added an AR(1) component to the model. The next observation of the ACF and PACF plots indicated that the correlations at all lags did not deviate significantly from zero. As such, we concluded that the ARIMA(1,1,0) model was satisfactory.

As we have a once-differenced model, we included the addition of a constant term to our model (drift term). This is justified as the realized time series seems to have a near linear trend. The final equation that represents our model is provided in (2) and the visualization of both, the realized time series and the forecasted timeseries, is presented in Figure 8.

4. Results

This section discusses in detail visualizations and key findings from them.

4.1. CO_2 Emissions Per Capita

To allow viewers to have a general sense of the current situation with GHGs emissions in the world, we created a choropleth map depicting the CO_2 emission per capita for each country and region in the world from 1970 to 2021 [1]. CO_2 emission per capita is the average amount of CO_2 emitted by each citizen of each country. CO_2 is a good representation of all GHGs since it takes up a large portion of the total volume, and is familiar to most readers. Per capita data was selected as it offers an important perspective on the global CO_2 challenge and gives a clear point of how much individuals can reduce emissions to together tackle this climate change. Moreover, as total CO_2 emission is skewed towards crowded areas such as Russia, China, and the US, CO_2 emissions per capita is a suitable metric to remove the biases against these regions.

In general, there are very large inequalities in per capita emissions in different regions across the world. More developed countries such as China, the US, and Canada tend to produce more CO_2 per capita. By looking at the graph from past to present, readers can notice the general increasing trend of CO_2 emission in the whole world. This effect is most significant in fast-developing countries such as India and Vietnam. However, a slowing increase or even decrease has been witnessed for the CO_2 emission per capita in Europe for the last decade. For Nordic countries, the CO_2 emission per capita has been on the rise since 1970, but is now slowing down.

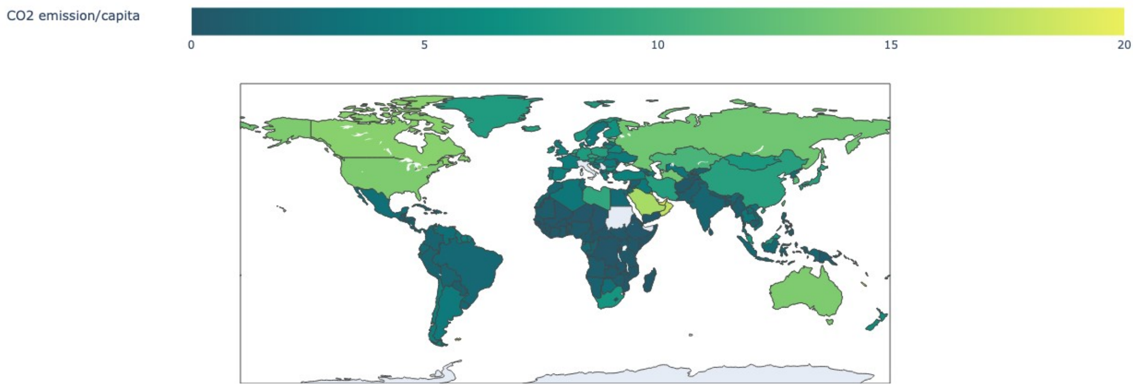

Figure 3. CO_2 emissions per capita

4.2. CO_2 Emissions Per Sector

CO_2 emitted into the atmosphere comes from various types of activities, with some sectors producing far more carbon emissions than others. Moreover, the proportions of sector-specific emissions is obviously not consistent between countries or years. As such, a stacked line chart representing the yearly division of emissions into 5 high-level sectors provides a quick understanding of recent trends and the development of emissions over time.

The CO_2 emissions by sector graph allows a user to select a specific country (or an aggregated global scope), and choose between a display of total emissions or emissions per capita. The sectors are as follows: *buildings*, *transportation*, *power industry*, *other industrial combustion*, and the general *other* category. These were the categorizations provided in our EDGAR dataset.

With the issue of climate change far from resolved, understanding the determinants behind the emission levels can help us prioritize appropriate actions and improve policymaking. For example, rather than targeting policies for emissions as a whole, the graph we provide may help to prioritize specific sectors that carry more weight than others.

Figure 4 below presents global CO_2 emissions per capita from 1970 to 2021 on a global scale. As we notice, per capita emissions have not drastically increased compared to absolute figures. Compared to 1970, the increase to 2021 is approximately 0.5 tonnes or roughly a 10% increase. Compared to the total absolute emissions which have greatly increased, we have indication that general population growth is a strong explanatory factor of increased CO_2 emissions.

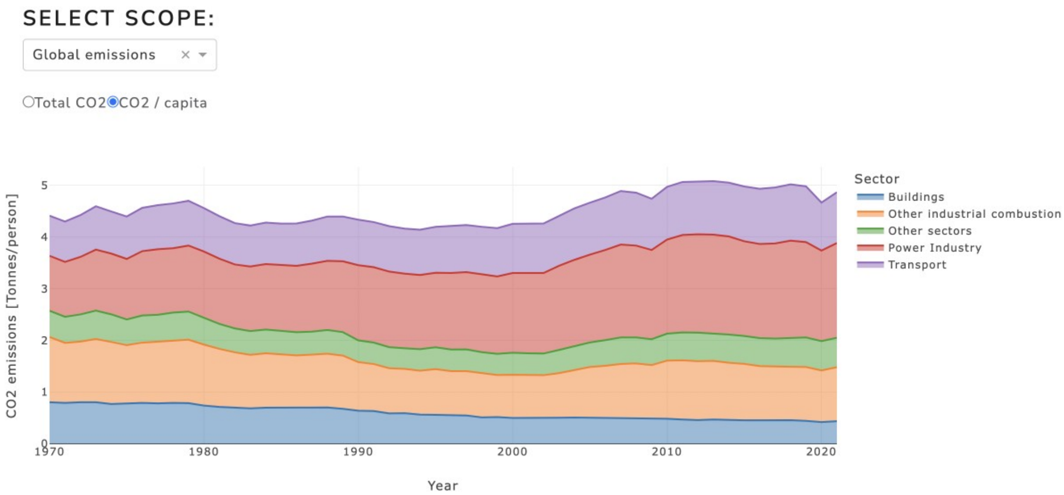

Figure 4. CO_2 emissions per sector

4.3. GDP and CO_2 Emissions Per Capita by Country and Region

From Figure 3, it can be seen that CO_2 emissions per capita varies greatly among countries and regions. It is historically known that the CO_2 emissions are strongly correlated with the health of a country's economy and welfare. Hence, one way to explain this variation is to look at the relation between GDP per capita and CO_2 per capita of all countries in the world. Gross Domestic Product (GDP) is a suitable metric in this case, as it represents the value of all goods and services produced over a specific time period within a country's borders [6].

In general, there is a strong correlation between the amount of CO_2 emitted and the standards of living. A brief look at the graph shows that CO_2 emissions are rising as the economy grows, except for a decline in both GDP per capita and CO_2 in 2019 as a result of a global pandemic. However, many countries have managed to achieve economic growth while reducing emissions. Nordic countries are examples of this opposite trend, in which the GDP per capita is rising and CO_2 emissions per capita are still decreasing. This trend clearly shows that we together can make an effort to reduce the CO_2 emissions while still facilitating economic growth.

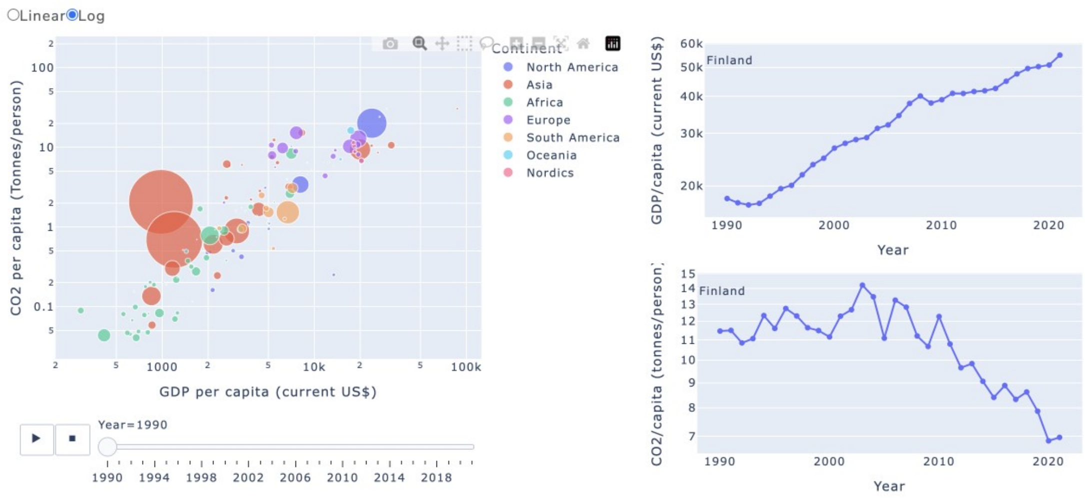

Figure 5. GDP per capita and CO_2 emissions per capita by country and region

4.4. Number of policies and measures by sector

In order to prevent the negative effect of climate change and the emission of greenhouse gases into the atmosphere, countries have started creating new policies and taking measures to minimize their impact on global warming. These policies are related to different sectors to ensure the reduction in the emissions of greenhouse gases. Current policies to reduce, or at least slow down growth, in CO_2 and other greenhouse gas emissions will have a significant impact on tackling this global issue. The European Council has adopted the European climate law in 2021. With this law, EU countries are legally obliged to cut at least 55% of emissions by 2030 and achieve climate neutrality to reach a net-zero emissions balance by 2050. In order to meet this agreement, different EU countries in Europe have adopted policies with different objectives [7]. It is necessary to understand the number of policies and measures per sector, as this allows us to see which sector has the biggest portion of the emitted greenhouse gases. By looking at the number of policies created for each sector, we can make important interpretations. If there is a high number of policies for a sector made by a country, this may imply that the sector is one of the critical reasons for the greenhouse emissions of the country. In Figure 6, we can explore the number of policies adopted in different countries in Europe by sector.

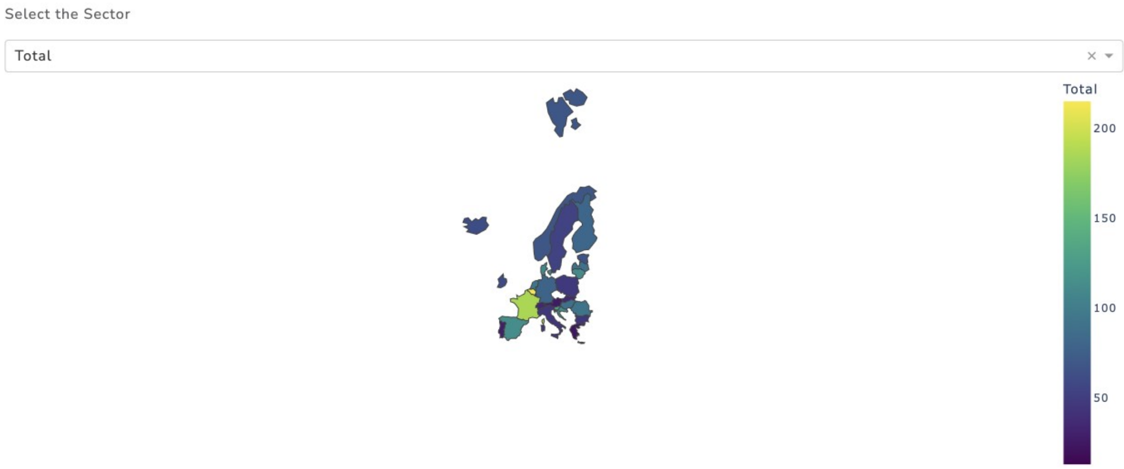

Figure 6. Number of policies and measures by sector

Based on the visualization, it can be seen that Belgium has significantly more policies and measures regarding the emissions of greenhouse gases compared to other EU countries. This makes sense as the European Commission is headquartered in Brussels. When we look at the number of policies and measures for all countries, we find that the energy sector has the highest number of policies and measures made by the European Union Countries. This means that the energy sector has the largest impact on the emission of greenhouse gases into the atmosphere. This phenomenon can also be seen in Figure 4.

4.5. Emissions by Country

European Commission has a great database called EDGAR—Emissions Database for Global Atmospheric Research [1]. It allows us to see how the fight against climate change is developing in different regions. We used this data for a simple line graph with radio items to give users the freedom to see different metrics. These include total CO_2 and greenhouse gas emissions. However, the raw emission data can be misleading, as it does not account for changes in population or the changes in GDP. To address this, we also added the option to use these metrics. It is also quite common to focus only on the CO_2 emissions, although other greenhouse gases also contribute significantly to the problem. A good example is methane which contributes 30% to global warming [8]. The total greenhouse gas emissions here are carbon equivalent emissions, meaning that it includes methane as well. Finally, the country-specific data does not account for emissions resulting from imported or exported goods and services. Thus, it has a bias towards economies that have outsourced polluting industries.

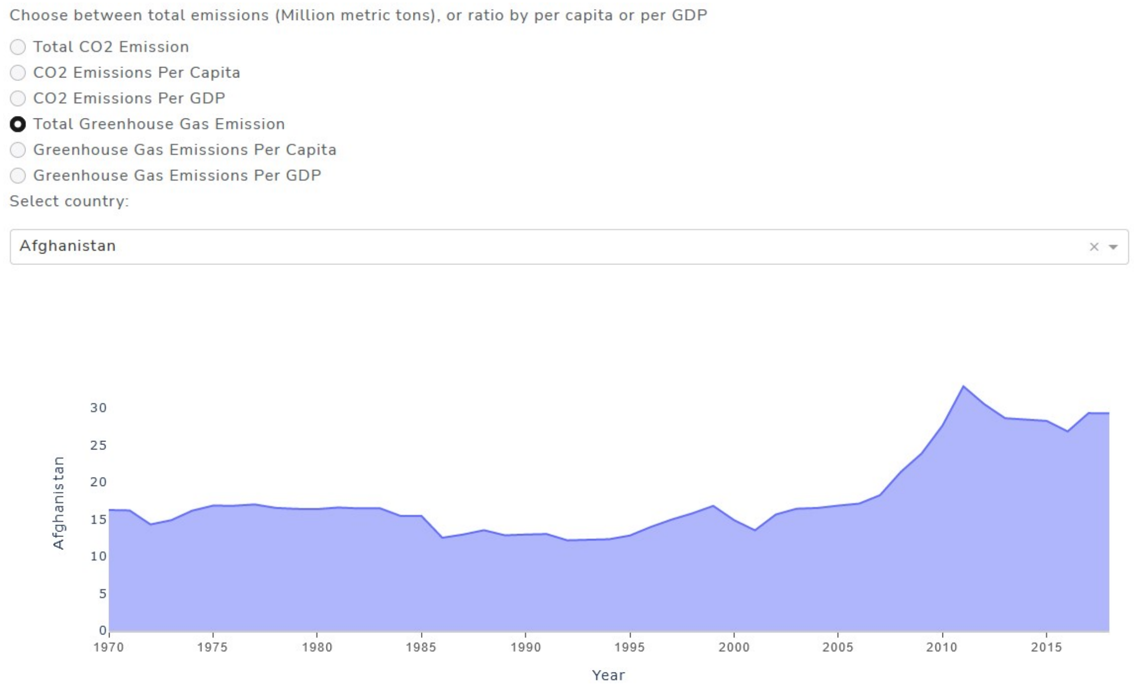

Figure 7. Emissions by country

4.6. Global CO_2 Emission Forecast for 5 Years

To provide a simple predictive model for the future CO_2 emissions, we implemented a suitable time-series model. This proved to be an ARIMA(1,1,0) model. A detailed procedure for how the model was constructed is provided in Section 3.2. The final equation representing our model is given:

$$y_t = y_{t-1} + a(y_{t-1} - y_{t-2}) + b \tag{2}$$

where y_t are datapoints t of the timeseries and a and b are fitted model parameters. With this model, we predict the global CO_2 emissions for the next 5 years.

The resulting plot is shown below in Figure 8. As typical for a simple ARIMA forecast, the future predicted values quickly regress to the mean after a few time units. Although this is a crude and standard time series model, the graph serves its purpose which is to show the general increasing trend of the emission in the world: it illustrates the likely future CO_2 emissions if no actions are taken to address the problem.

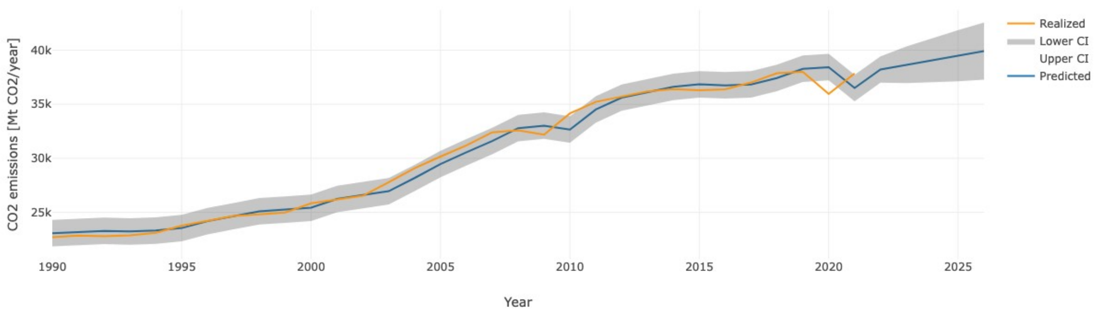

Figure 8. Global CO_2 emissions 5-year forecast

4.7. Global GHG Emissions by Scenario

The United Nations has many sub-organizations, and one of them is The Intergovernmental Panel on Climate Change (IPCC). IPCC is the scientific hub for evaluating climate change, and it has many working groups focused on different aspects of climate change. Their Working Group III is focused on climate change mitigation and produces useful information on the different possible scenarios [9]. The Global GHG Emissions by Scenario graph aims to communicate the difference between each temperature goal in terms of their respective emissions' trajectory. Simply put, the path emissions need to take to reach a certain goal. As global warming is a multifaceted problem, there are a lot of uncertainties. To communicate these uncertainties, IPCC gave 95% confidence intervals that we also included in our version of the graph. Like our projection in Global CO_2 Emission Forecast for 5 Years, IPCC expects Global greenhouse gas emissions in 2030 associated with the implementation of Nationally Determined Contributions (NDCs) announced before the 26^{th} UN Climate Change Conference to result in global warming exceeding 1.5 °C [9].

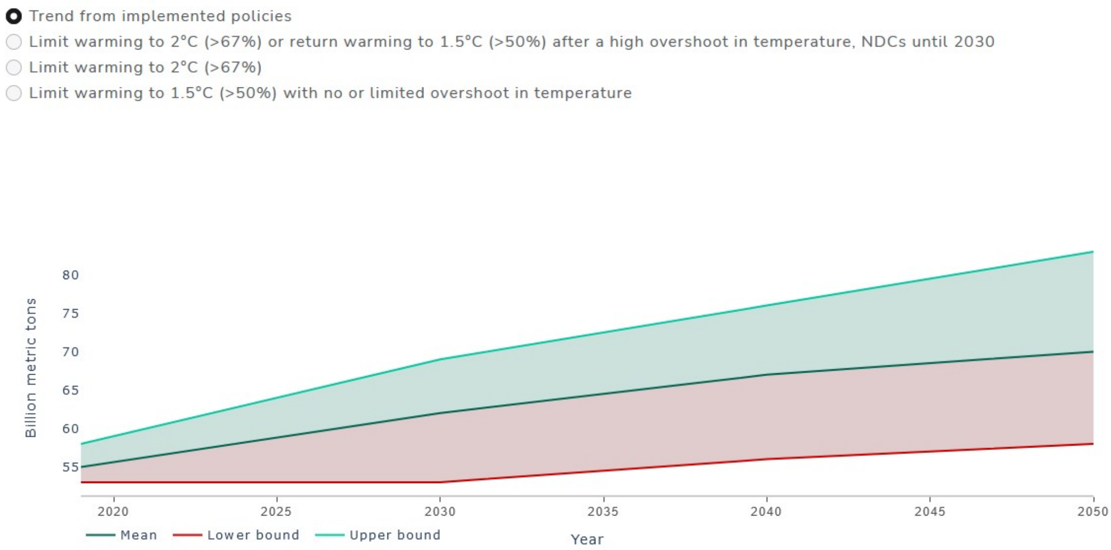

Figure 9. Global greenhouse gas emissions by scenario

4.8. Scenario Calculator

At the end of our webpage, we have created a calculator to help the user of our website understand how their actions can reduce greenhouse gas emissions. In our calculator, the user can choose which action they can take. These three actions have different challenge levels, and they are suitable for people with different habits. The first action one can take is driving their car for a lesser amount of time. The second action is cutting down electricity usage and the third, taking quicker showers. The user can select one or more of these actions. Different online resources were used for calculating the average carbon dioxide emission that is caused by different activities [10] [11] [12]. Then based on the information that was found on the internet, the carbon dioxide emission reduction is calculated. The reduction is the difference between the average person's carbon dioxide emission and our website's users' desired carbon dioxide emissions. Furthermore, the users can see how much carbon dioxide emission can be decreased by selecting the number of people that can implement these desired actions. This calculator is essential in understanding how small actions can create a tremendous impact on a greater scale.

4.9. Website Structures

At the start of the project when we didn't know much about what we wanted to do, we decided to go with a simple approach that goes from a worldwide view to a smaller, more specific (Nordic countries) view of GHGs emissions. Then, when we have agreed on which graphs to implement and the gap minder-inspired quiz at the start of the website, the structure at that point was the user will do the quiz and get some interesting facts about the questions, after that, the user will start exploring our interactive graphs in the same page. We continued with this structure until the code was too long which made it harder to navigate and reduced the implementation speed, so, Dung suggested splitting graphs about historical data and the prediction into two different pages. This new structure not only helped us with the shorter code, but also gives the user a more reasonable flow.

As our website was divided into different pages, it needed a navigation system. Teemu implemented a simple navigation bar that appears at the top and bottom of the website. This was done using Dash Bootstrap Components library [13], which expands on the Dash by Plotly. This also improved the overall appearance of the website, as the dark grey navigation bar gave contrast to the otherwise bright layout. We had also planned a quiz at the start to gain users' attention and retain commitment with interaction, but this was later moved to a separate section. We considered the pros and cons of forcing users to take the quiz before accessing the rest of the site and decided it to be too demanding for the users, especially the returning ones. The quiz was inspired by the one made by Gapminder [14].

5. Ethical Issues

The project mainly concerns communicating the past and current momentum in reducing greenhouse gas emissions to targeted customers through the means of a website, thus in the process of conducting this project, we had to make some ethical choices. Controlling greenhouse gas emissions itself is a complex matter, as it cannot be an independent field of its own, but requires a combination of policies in multiple fields such as economics and politics. This characteristic results in our project's liability to contain opinions and biases. With our roles which consist of collecting emission and policies data, and visualizing and analyzing those data, we tried our best not to be opinionated throughout different stages of the project.

In the process of collecting data, we made sure that all sources are trustworthy and neutral. For preprocessing data, we mainly manipulate their dimensions and format without changing anything. Missing values were left missing, so as to not create misinformation. For visualizing data, we first used aggregate data for all regions and sectors. Later, we found out that this metric of total greenhouse gas emission can create large skewness towards populated areas such as China, Russia, and The US. Per capita data was made as a response to this issue. For analyzing data, we mostly let data speak for itself without incorporating too much explanation and analysis. In this way, we avoided shifting the viewer's focus onto specific aspects and regions. Those were our conscious decisions regarding the ethical aspects of the project as of now. For future purposes, more ethical decisions and evaluations must be considered.

6. Project Work Assessment

We divide our assessment of our work and practices of this project into successes and challenges. Finally, we go over the key takeaways for future projects. As the first data science project for most of the team members, the project offered new challenges on top of the typical intricacies of group work.

Our team found that our shared experience with Python offered a common ground for the foundation of the project. Deciding the communication channels, cloud services and regular meeting times was easy, as platforms like telegram and Google Drive were familiar to all participants. We also found a good spot for additional weekly meetings from our calendars to be right before the Monday lecture. Having two meetings per week helped to balance the internal working cycle and increased coordination. At the start of the meetings each of us told what progress was made, and at the end of the meetings, we decided what to do for the next occasion. This practice helped us to divide the workload equally and keep up with the schedule. Although the development of the website slipped past the soft deadline originally intended, we had given a longer runway to account for any delays. This foresight in scheduling allowed us to be ready for the final presentation at Reactor. Preparations for the presentation were also proven adequate, as we had no problems with the live demo, and there were no issues during the presentation. We had practiced the presentation in our Monday meeting before, giving us the confidence needed for a pleasant experience on the day of the presentation. Despite so many positive remarks, there were also challenges and room to improve in our practices.

One of the typical difficulties associated with group work is changes to previously agreed timetables. We faced this issue first when the study period changed and the previous time for our usual Monday meetings was no longer fitting for all the team members. For the second half of the project, the meeting time for our additional meeting was irregular. It was difficult to find the time that suited each of us, but we wanted to maintain the face-to-face meetings as our group practice. We used questionnaire-style meeting schedulers to combat this challenge.

The second obstacle that we faced was the practices of working with a collaborative code base. We all had used GitHub for our own work before, but things like version and branch management were new things in practice. At first, we had our own folders as we developed parallel visuals for the website, but after combining them into one, we encountered challenges. Different practices in writing and documenting resulted in bugs a few times. These were often very difficult to track down, as Python is dynamically typed, and the same variable might be used on two separate occasions for completely different purposes. One notable example of this was a problem where some visuals were not rendering properly, despite the program raising no errors. This issue kept the whole team puzzled for two weeks, after which Daniel found the root of the problem and informed the team. This bug was a direct result of poor practices and the lack of communication during the development. Although we were successful in setting many common practices for the team, the ones concerning actual development lacked the unification needed in this type of project. On a positive note, the issues were resolved, and we learned about the pitfalls that group work might entail.

The main takeaway can be derived from the previous paragraph. Good practices in collaborative work are imperative. Although good practices are important when working on your own, it is still easy to sidestep them. If you are the only person working on the code, making the code readable for others is one of the good practices that might be easy to forget. There are no immediate ramifications for this when working on your own, although later on, it might be difficult to return to older projects. In collaborative work, however, this is not the case. If team members can't read each other's code, it is impossible for them to fix issues that involve code written by other team members. Although we were able to solve the problem we had, we all agreed that our most important lesson was to set common practices for our future projects.

The second practice of collaborative work that we learned to be important was version and branch management. After the initial start, we only used one branch for the project. This meant that our development code was the same as the live version. We had not automated the updates to our server in Heroku, but it would have been detrimental to break the main branch that we used. It also made the updates of multiple people messy, as the changes could sometimes overwrite development made by others if not handled correctly. After the project, we all shared the concern that collaborative practices might need to be highlighted more in the course. This would help groups made of students to avoid these vulnerabilities and take on good practices from the beginning.

7. Student's Roles

At the beginning of the project, we decided on a non-hierarchical structure. With the project having concluded successfully, we can announce that this was a good decision. Despite not choosing a dedicated group lead or manager, the allocation of work was efficient, and tasks were quite naturally divided based on personal skills and interests. There were no evident clashes or necessity for a tie-breaker; decisions were unanimous and satisfactory to all.

We decided on a rolling secretary for our meetings at the beginning of the project—everyone would take turns documenting the meeting agenda and main points. This started off wonderfully, but slightly fell off towards the end. One reason is of course the reduced need for planning and brainstorming work as the project progressed. Towards the end, it was clear what everyone's tasks were, and the final implementation had been decided. Meetings were more concise and less elaborate, focusing on bug fixes and cosmetic changes to the website, while considering the feedback from the company representative, Saku.

Despite no hierarchy-focused roles, each group member had a main task they worked on. These roles largely emerged based on previous experience and skills. We highlight these roles below.

Teemu had worked with Dash and Heroku before, so he took the responsibility for getting our website hosted and laying the initial example of how to utilize Dash. In addition to that, Teemu handled the general country specific CO_2 emissions graph and the policy-based prediction graph.

Daniel implemented the CO_2 emissions by sector graph as well as the time-series model and its graph. This too was a natural choice, as Daniel had taken a course on time-series forecasting and had some preliminary knowledge of the process.

Linh worked on Figure 5. Furthermore, Linh was also the provider of the LaTeX file which was used for writing this report.

Selin worked on the policy graph, demonstrated in Figure 6. Importantly, Selin found our trusted EDGAR data set, off of which the majority of our visualizations and analysis were conducted. Selin also worked on implementing our CO_2 reduction calculator, together with Dung.

As mentioned, Dung and Selin worked on the CO_2 reduction calculator. Dung did a large portion of the "compilation" work, putting together everyone's components and fitting them into the layout of our website. Dung also worked on implementing the quiz section available on our website.

Hieu implemented the stand-out country graph highlighting historical CO_2 emissions visible immediately as the website is launched. Furthermore, Hieu spent a considerable amount of time polishing the work, something at least the author of this section benefitted off of by shamelessly taking a peak at his code.

Moreover, each individual student contributed to the report by writing their own respective sections in the final product.

8. Conclusions and Future Prospects

8.1. Conclusions

This project was part of the course CS-C3250—Data Science Project 2022 and was conducted under the supervision of Jorma Laaksonen, Teaching Assistant Hanne Sauer, and in collaboration with Saku Suuriniemi, the representative from Reaktor Technology company. The project's main objective was to create an interactive web page to communicate the emission reduction momentum to a predominantly Nordic audience.
From the Greenhouse Gasses emissions of all world countries dataset from the Emission Database for Global Atmospheric Research—EDGAR, we have created **four** interactive graphs and **one** ARIMA model to predict the global carbon emissions in the next five years. Moreover, with two interesting datasets from the European Economic Area (EEA) database and Resources for the Future organization (RRF), we are able to give the user an interesting view of the number of policies and measures by sector in Europe and how different carbon pricing will affect the US's carbon emissions. In addition to those graphs, we created five quizzes about the current situation and a carbon reduction calculator that each user can use to explore the effect of their actions. To conclude, we believe our interactive web page can help the user to better understand global emissions and guide them to a more sustainable lifestyle with our carbon-reducing calculator.

8.2. Future Prospects

Regarding the future, our projects can be improved in three ways. Firstly, as the majority of users will access our web page with their mobile phones, element sizing and alignment can be improved in the future. This can be achieved through CSS media rules or with Dash Bootstrap Components. Secondly, a script to retrieve new datasets when they are published can be created so that an end-to-end data pipeline can be created. Lastly, the codebase can be refactored in order to modularize all components on our web page. With this new structure, the development velocity can be increased as there will be fewer bugs, components can be reused, and it will be much easier to collaborate.

References

[1] European Commission. Co2 emissions of all world countries — EDGAR - Emissions Database for Global Atmospheric Research, 2022. [Online; accessed 19-September-2022].

[2] European Environment Agency. Ea database on greenhouse gas policies and measures in europe, 2022. [Online; accessed 15-October-2022].

[3] Resources for the Future. Carbon pricing calculator, 2022. [Online; accessed 15-November-2022].

[4] World Bank. Population, total | data, 2021. [Online; accessed 27-October-2022].

[5] Durdu Ömer Faruk. A hybrid neural network and arima model for water quality time series prediction. *Engineering Applications of Artificial Intelligence*, 23(4):586–594, 2010.

[6] OCED. Gross domestic product (gdp) (indicator), 2022. [Online; accessed 15-November-2022].

[7] Council of the European Union. Climate change: what the eu is doing, 2022. [Online; accessed 20-November-2022].

[8] International Energy Agency. Methane and climate change — International Energy Agency, 2022. [Online; accessed 10-October-2022].

[9] IPCC Working Group III. Climate change 2022 mitigation of climate change summary for policymakers — Intergovernmental Panel on Climate Change, 2022. [Online; accessed 23-October-2022].

[10] Vaasan Sahko. Electricity consumption at your home, 2021. [Online; accessed 16-November-2022].

[11] Andrew Gross. Americans spend an average of 17,600 minutes driving each year, 2020. [Online; accessed 20-November-2022].

[12] Wikimedia Foundation. List of countries by number of households, 2022. [Online; accessed 16-November-2022].

[13] Dash Bootstrap Components contributors. Navbar — Dash Bootstrap Components, 2022. [Online; accessed 15-November-2022].

[14] Gapminder team. Climate action — Gapminder Upgrader, 2022. [Online; accessed 12-October-2022].