VIT®

Vellore Institute of Technology

(Deemed to be University under section 3 of UGC Act, 1956)

School of Computer Science and Engineering

VIT Chennai

Vandalur - Kelambakkam Road, Chennai - 600 127

PROJECT REPORT

Programme: B. Tech CSE

Course Code: CSE3020

Course Name: Data Visualization

Slot: D2

Faculty: Dr. Parvathi R

Title: INTERNATIONAL CO2 EMISSION ANALYSIS AND VISUALISATION

20BCE1701 – Priyanshu Kapoor

ABSTRACT

The international CO2 emission analysis and data visualization project aims to present a comprehensive view of the global carbon dioxide emissions and their impact on the environment. This project collects and analyses data from various sources, including government reports, scientific publications, and online databases.

The project also uses data visualization techniques to represent the complex information in an intuitive and engaging way. The analysis covers various aspects, including trends in CO2 emissions over time, their sources, and the environmental impacts of emissions.

The project aims to increase awareness of the issue of climate change and encourage individuals, organizations, and governments to take action to reduce emissions and mitigate their effects.

Keywords: Visualisation, data analysis, CO2, greenhouse gases emission, analysis region wise, factors that promote emission.

1. INTRODUCTION

Climate change is one of the most pressing issues facing our planet today. One of the main contributors to climate change is carbon dioxide (CO2) emissions, which are generated by a variety of human activities, including transportation, energy production, and agriculture. To better understand the impacts of CO2 emissions on the environment and how they are changing over time, researchers and scientists are using data analysis and visualization techniques to present a comprehensive view of the issue. The international CO2 emission analysis and data visualization project is an example of such an effort. This project collects data from various sources and presents it in an accessible and informative way to help increase awareness of the issue and encourage action to mitigate the effects of CO2 emissions. In this project, we will explore the trends, sources, and impacts of CO2 emissions, and use data visualization techniques to better understand the complex issues surrounding this global problem.

CO2 (carbon dioxide) emissions, along with other greenhouse gases, are recognized as harmful to the environment due to their contribution to climate change. When released into the atmosphere, CO2 acts as a greenhouse gas, trapping heat and leading to an increase in global temperatures. This has numerous negative impacts on the environment, including:

1. Climate Change: Increased CO2 emissions are a major contributor to climate change, which has far-reaching consequences such as rising sea levels, extreme weather events, altered ecosystems, and disruptions to agriculture, water resources, and human settlements. Climate change poses a significant threat to both natural ecosystems and human societies, with potential long-term negative impacts on the environment and the economy.
2. Ocean Acidification: A significant portion of CO2 emissions are absorbed by the world's oceans, leading to a process called ocean acidification. This results in a decrease in the pH of seawater, which can have detrimental effects on marine ecosystems, including coral reefs and other marine species that are sensitive to changes in acidity. Ocean acidification can disrupt marine food webs, threaten biodiversity, and impact coastal communities that depend on marine resources for their livelihoods.

3. Air Pollution: CO2 emissions are often accompanied by other air pollutants, such as particulate matter, sulphur dioxide, and nitrogen oxides, which can have harmful effects on human health, agriculture, and ecosystems. Air pollution from CO2 emissions can contribute to respiratory diseases, cardiovascular problems, and other health issues, as well as damage crops, forests, and bodies of water.
4. Deforestation: CO2 emissions from human activities, particularly the burning of fossil fuels, are a major driver of deforestation. Deforestation, especially in tropical rainforests, releases large amounts of CO2 into the atmosphere and reduces the capacity of forests to act as carbon sinks, exacerbating climate change and causing loss of biodiversity and habitat destruction.
5. Disproportionate Impacts on Vulnerable Communities: CO2 emissions and climate change disproportionately affect vulnerable communities, such as low-income populations, indigenous peoples, and marginalized groups. These communities often have fewer resources to adapt to the impacts of climate change and may face increased risks to their livelihoods, health, and well-being.

Given the harmful effects of CO2 emissions on the environment, there is a pressing need for concerted global efforts to reduce emissions, transition to renewable energy sources, promote sustainable practices, and implement policies to mitigate and adapt to climate change. International cooperation, policy interventions, technological innovations, and individual actions are all crucial in addressing the challenges posed by CO2 emissions and protecting the environment for present and future generations.

2. LITERATURE SURVEY

1. Global Research on Carbon Emissions: A Scientometric Review

Link:https://www.mdpi.com/2071-1050/11/14/3972

Summary: This paper aims to address this research gap through a scientometric review of research on global carbon emissions. The scientometric review adopts a quantitative methodology to analyze the landscape and the intellectual core of the existing body of literature available on carbon emissions. This paper will also examine the deficiencies of the existing body of carbon emission knowledge by identifying and discussing the quality and the scope of available publications. The research findings provide a detailed understanding of the current state of global carbon emission research work and potential directions for future research. The findings will also be an updated and valuable reference for the practitioners and policymakers in planning their future work and funding. According to the results observed through the scientometric analysis, it was evident that the carbon emission research domain has attracted the attention of global researchers. A significant increase in research publications over the past two years is a strong indication of the growth in the carbon emission research domain. Carbon capturing, predicting future carbon emissions through trend analysis, evaluating carbon performance, identifying carbon mitigation opportunities and ultimately achieving zero carbon emission goals are some of the most popular research areas in the carbon emission research domain. The scientometric analysis revealed the trends of carbon emission research over the past three decades which was the objective of this research.

2. Visualization and Bibliometric Analysis of Carbon Neutrality Research for Global Health

Link:https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9298964

Summary: This paper uses Citespace and Alluvial Generator as visual measurement soft-

wares to carry out the following research from the dimensions of "countries," "institutions," "co-citation analysis of literature," and "co-occurrence of keywords:" (1) research status and cooperative relations in the field of carbon neutrality among various countries;

(2) research hotspots and emerging trends in the field of carbon neutrality;

(3) the development of carbon neutrality research and the difficulties encountered in the research.

The main contributions of this paper are summarized as follows:

1. A bibliometrics research framework is designed to provide research methods and visualization approaches for research hotspots, research development, and research trends in different fields.
2. The international situation of carbon neutrality research is visualized and cooperation among different countries in this field are analyzed to further promote international co-operation in carbon neutrality research.
3. The visual analysis of research hotspots, research development and research trends of carbon neutrality research is carried out to identify the development trend and some urgent problems in the research, so as to provide suggestions and implications for the future development of carbon neutrality research.

3. Carbon Footprint Research Based on Input–Output Model—A Global Scientometric Visualization Analysis

Link:

https://www.researchgate.net/publication/363506367_Carbon_Footprint_Research_Based_on_Input-Output_Model-A_Global_Scientometric_Visualization_Analysis

Summary: The purpose of this paper is to explore the knowledge structure and frontier trends in respect of the IOA model applied to CF research using scientometric visualization analysis. The main findings of this paper are as follows.

(1) Published articles show a two-stage increase in the period 2008 to 2021, and present a complex academic network of countries, authors, and institutions in this important domain.

(2) The classic studies are mainly divided into three categories: literature reviews, database application introduction, and CF accounting in different scales.

(3) The research hotspots and trends show that the research scales tend to be more microscopic and applications of models tend to be more detailed. In addition, supply-chain

analysis and driver-factor analysis will probably become the main research directions in the future.

4. Visualization analysis of research on climate innovation on CiteSpace

Link: https://www.frontiersin.org/articles/10.3389/fenvs.2022.1025128/full

Summary: This study aims to identify the current research status and trends in climate innovation *via* CiteSpace visual analysis. They conducted co-authorship analysis, co-citation analysis, and keyword co-occurrence analysis and found that the main contributors are from America, England, and China. Moreover, climate innovation has an interdisciplinary knowledge source, and its main research frontier focuses on the design of policies for energy efficiency improvement and new energy technologies. These results point out some directions for further research on climate innovation and outline the knowledge structure in this field, thereby addressing the lack of studies on such topic. By analyzing the lineage of research developments, it was found that energy technology and energy efficiency would be the main research hotspots in climate innovation in the future, and relevant policies also need to strengthen incentives for research in this area. This study is particularly helpful for those researchers who are interested in climate innovation and can subsequently contribute to the long-term development of various research directions in this field, to the formulation of climate change mitigation and adaptation policies, and to the sustainable development of human society.

5. Structural Breaks in Carbon Emissions: A Machine Learning Analysis

Link:https://www.imf.org//media/Files/Publications/WP/2022/English/wpiea2022009-print-pdf.ashx

Summary: In this paper, They identify and analyze structural breaks using machine learning methodologies. They find that downward trend shifts in carbon emissions since 1965 are rare, and most trend shifts are associated with non-climate structural factors (such as a change in the economic structure) rather than with climate policies. They analyze the

optimal mix between climate and non-climate policies, findings highlight the importance of the nonclimate policies in reducing carbon emissions. On the methodology front, paper contributes to the climate toolbox by identifying country-specific structural breaks in emissions for top 20 emitters based on a user-friendly machine-learning tool and interpreting the results using a decomposition of carbon emission (Kaya Identity).

6. Spatial Estimation and Visualization of CO2 Emissions for Campus Sustainability:

Link:
https://www.academia.edu/35181005/Spatial_Estimation_and_Visualization_of_CO2_Emissions_for_Campus_Sustainability_The_Case_of

Summary: This study established a spatial evaluation, estimation, and visualization of the CO2 emissions of King Abdullah University of Science and Technology(KAUST), Saudi Arabia. The data required for this study were collected from the overall coverage of the university campus buildings by transforming raster data from the satellite image to vector data in the form of polygons, and then multiplying the area by the number of floors of the individual building. ArcGIS 10.3
(ESRI, Redlands, CA, USA) software was used for this campus CO2 emissions evaluation and visualization. The overall estimate of the CO2 emissions for the university campus was 127.7-tons CO2 equivalent. The lowest emission was 0.02-tons CO2 equivalent while the maximum value was 20.9-tons of CO2 equivalent. By this ArcGIS-based evaluation, it is evident that geographically integrated model for campus estimation and visualization of CO2 emissions provides the information for decision makers to develop viable strategies for achieving a higher standard in overall campus sustainability and addressing the issue of climate change.

7. Quantifying greenhouse gas emissions

Link:https://link.springer.com/article/10.1007/s11027-019-09867-4

Summary: The approaches to address uncertainty discussed in this special issue reflect attempts to improve national inventories, not only for their own sake but also from a wider, system analytic perspective. They seek to strengthen the usefulness of national emission inventories under a compliance and/or global monitoring and reporting framework. The papers in this special issue demonstrate the benefits of including inventory uncertainty in policy analyses. The issues raised by the authors and featured in their papers, along with the role that uncertainty analysis plays in many of their arguments, highlight the challenges and the importance of dealing with uncertainty. While the Intergovernmental Panel on Climate Change (IPCC) clearly stresses the value of conducting uncertainty analyses and offers guidance on executing them, the arguments made here in favor of performing these studies go well beyond any suggestions made by the IPCC to date. Improving and conducting uncertainty analyses are needed to develop a clear understanding and informed policy.

8. CO_2 Embodied in International Trade with Implications for Global Climate Policy

Link: https://pubs.acs.org/doi/10.1021/es072023k

Summary: In this article we determine the CO2 emissions embodied in international trade among 87 countries for the year 2001. They have found that globally there are over 5.3 Gt of CO2 embodied in trade and that Annex B countries are net importers of CO2 emissions. Depending on country characteristics—such as size variables and geographic location—there are considerable variations in the embodied emissions. We argue that emissions embodied in trade may have a significant impact on participation in and effectiveness of global climate policies such as the Kyoto Protocol. We discuss several policy options to reduce the impact of trade in global climate policy. If countries take binding commitments as a part of a coalition, instead of as individual countries, then the impacts of trade can be substantially reduced. Adjusting emission inventories for trade gives a more consistent description of a country's environmental pressures and circumvents many trade related issues. It also gives opportunities to exploit trade as a means of mitigating emissions. Not least, a better understanding of the role that trade plays in a country's economic and

environmental development will help design more effective and participatory climate policy post-Kyoto.

3. MATERIAL AND METHODS

3.1 Info about models

Libraries used

We have performed data analysis and visualisation using the Python and Pandas packages. Visualizations are also created using the Matplotlib and Seaborn packages.

- **Python** is a high-level programming language popular in data research and machine learning. It is well-known for its ease of use and simplicity.
- **Pandas** is a Python package that offers tools for data manipulation and analysis. It is based on the NumPy library and offers simple data structures for working with structured data.
- **Matplotlib** is a Python library that provides capabilities for data visualisation. It is commonly used to create charts, graphs, and other types of visualisations.
- **ggplot2** provides a powerful and flexible framework for creating a wide range of data visualizations in R, making it a popular choice among data scientists, statisticians, and data analysts for creating compelling and informative visualizations.
- **dplyr** library provides a set of functions that are optimized for efficient data manipulation and transformation tasks. It aims to make data manipulation in R fast, intuitive, and easy to understand, allowing users to perform common data manipulation operations in a concise and readable way.

Machine Learning Algorithms used for Classification

- **Decision Tree** - is a type of supervised machine learning algorithm that is used to classify or predict the class or category of an input data point based on its features. The decision tree algorithm builds a tree-like structure where each node in the tree represents a decision based on the values of one or more input features, and each leaf node represents a predicted class label.
- **KNN -** K-Nearest Neighbors (KNN) algorithm is a type of supervised machine learning algorithm used for classification tasks. KNN is a non-

parametric and lazy learning algorithm that makes predictions based on the k-nearest neighbors of a new data point in the training data.

Machine Learning Algorithms used for Classification

- **Principal Component Analysis (PCA)** is a dimensionality reduction technique that can be used for visualizing high-dimensional data in lower-dimensional space. PCA transforms the original features of the data into a new set of orthogonal (uncorrelated) features called principal components, which are sorted in descending order of their explained variance. The first few principal components capture the most significant information of the data, and they can be used to visualize the data in a lower-dimensional space.
- **Linear Regression Algorithm** Linear Regression is a supervised machine learning algorithm used for predicting a continuous target variable based on one or more input features. It is a simple and widely used algorithm for regression tasks, where the goal is to model the relationship between input features and a continuous output variable.

Correlation Matrix for interpreting results - is a table that shows the pairwise correlation coefficients between multiple variables in a dataset. It is commonly used in statistics and data analysis to assess the strength and direction of linear relationships between variables.

3.2 Dataset

1. We have a dataset for Carbon dioxide emission comprising of country, country code, yearly change, per capita, population and life expectancy.

	A	B	C	D	E	F	G
1	Country	Code	CO2Emiss	YearlyCha	Percapita	Populatior	LifeExpectancy
2	Afghanista	AFG	9900004	7.13	0.28	35383032	63.763
3	Albania	ALB	5208319	4.45	1.8	2886438	78.194
4	Algeria	DZA	1.56E+08	0.17	3.85	40551392	76.298
5	Angola	AGO	30566933	3.13	1.06	28842489	59.925
6	Anguilla	AIA	30262	1.52	2.1	14429	81.441
7	Antigua ar	ATG	438763	1.51	4.64	94527	76.617
8	Argentina	ARG	2.01E+08	0.16	4.61	43508460	76.221
9	Armenia	ARM	4597845	3.06	1.57	2936143	74.64
10	Aruba	ABW	286871	1.51	2.74	104872	75.868
11	Australia	AUS	4.15E+08	-0.98	17.1	24262712	82.959
12	Austria	AUT	73764112	1.54	8.43	8747301	81.258
13	Azerbaijar	AZE	33614235	-0.41	3.45	9736043	72.493

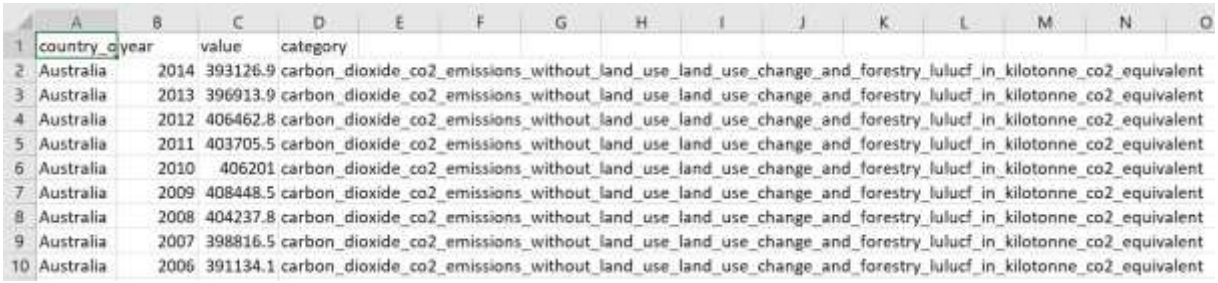
The "CO2Emission_LifeExp.csv" file is a dataset that contains information about carbon dioxide (CO2) emissions and life expectancy for different countries. The dataset is likely organized in a tabular format, with rows and columns, where each row represents a specific country and each column contains data related to CO2 emissions and life expectancy.

The dataset may include the following columns:

1. Country: This column contains the names of different countries for which data is recorded.
2. CO2 Emission: This column contains data related to CO2 emissions, which is the amount of carbon dioxide released into the atmosphere as a result of human activities such as burning fossil fuels (coal, oil, and gas), deforestation, and industrial processes. CO2 emissions are one of the main contributors to climate change and global warming.
3. Life Expectancy: This column contains data related to the average life expectancy in years, which is the estimated number of years a person can expect to live from birth. Life expectancy is influenced by various factors, including healthcare, nutrition, sanitation, lifestyle, and socioeconomic conditions.

The "CO2Emission_LifeExp.csv" dataset can be used to analyze the relationship between CO2 emissions and life expectancy in different countries, and to assess the potential impact of carbon emissions on human health and well-being. It can also be used to study trends and patterns in CO2 emissions and life expectancy over time, and to compare the performance of different countries in terms of their emissions and life expectancy outcomes.

2. We have a second dataset on greenhouse gases inventory data comprising of country, year, value and category



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	country_c	year	value	category											
2	Australia	2014	393126.9	carbon_dioxide_co2_emissions_without_land_use_land_use_change_and_forestry_lulucf_in_kilotonne_co2_equivalent											
3	Australia	2013	396913.9	carbon_dioxide_co2_emissions_without_land_use_land_use_change_and_forestry_lulucf_in_kilotonne_co2_equivalent											
4	Australia	2012	406462.8	carbon_dioxide_co2_emissions_without_land_use_land_use_change_and_forestry_lulucf_in_kilotonne_co2_equivalent											
5	Australia	2011	403705.5	carbon_dioxide_co2_emissions_without_land_use_land_use_change_and_forestry_lulucf_in_kilotonne_co2_equivalent											
6	Australia	2010	406201	carbon_dioxide_co2_emissions_without_land_use_land_use_change_and_forestry_lulucf_in_kilotonne_co2_equivalent											
7	Australia	2009	408448.5	carbon_dioxide_co2_emissions_without_land_use_land_use_change_and_forestry_lulucf_in_kilotonne_co2_equivalent											
8	Australia	2008	404237.8	carbon_dioxide_co2_emissions_without_land_use_land_use_change_and_forestry_lulucf_in_kilotonne_co2_equivalent											
9	Australia	2007	398816.5	carbon_dioxide_co2_emissions_without_land_use_land_use_change_and_forestry_lulucf_in_kilotonne_co2_equivalent											
10	Australia	2006	391134.1	carbon_dioxide_co2_emissions_without_land_use_land_use_change_and_forestry_lulucf_in_kilotonne_co2_equivalent											

The "greenhouse_gas_inventory_data_data.csv" file is a dataset that contains information about greenhouse gas (GHG) emissions for different countries or regions. The dataset is likely organized in a tabular format, with rows and columns, where each row represents a specific country or region, and each column contains data related to GHG emissions.

The dataset may include the following columns:

1. Country/Region: This column contains the names of different countries or regions for which GHG emissions data is recorded.
2. Year: This column contains the year or time period for which the GHG emissions data is reported. GHG emissions are typically reported on an annual basis, but the dataset may also include data for multiple years or time periods.
3. Greenhouse Gas: This column contains data related to the different types of greenhouse gases, such as carbon dioxide (CO2), methane (CH4), nitrous oxide (N2O), fluorinated gases (HFCs, PFCs, SF6), and others. These gases trap heat in the Earth's atmosphere and contribute to climate change.
4. Emissions: This column contains data related to the amount of GHG emissions, typically measured in metric tons of carbon dioxide equivalent (CO2e), which is a unit that accounts for the global warming potential of different greenhouse gases. It combines the emissions of different gases into a single value to facilitate comparison.

The "greenhouse_gas_inventory_data_data.csv" dataset can be used to assess the magnitude and trends of GHG emissions for different countries or regions, track progress towards national or international emissions reduction targets, and inform policy-making, research, and analysis related to climate change mitigation and adaptation efforts.

3.3 Architecture and Explanation

This project involves the analysis and visualization of a dataset containing information on Carbon Dioxide emission in among different countries along with per Capita, life expectancy. Yearly change. The dataset is loaded and pre-processed using the Python and Pandas packages. The columns `country` and `code` (column 1 and 2) are removed and objects are reordered randomly.

The data is then split into training and testing sets.

Four different machine learning models are then applied to the dataset: Principal Component Analysis, Linear Regression, Bar Joseph Seriation, Bertin's Permutation Matrix.

The architecture of the project involves several key steps:

1. Data Collection: Two datasets were used in the project: greenhouse_gas_inventory_data_data.csv and CO2Emission_LifeExp.csv. These datasets likely contain information on greenhouse gas emissions and life expectancy, respectively.
2. Data Pre-processing: The datasets were cleaned and prepared for analysis. This may have involved handling missing values, normalizing or scaling the data, and converting categorical variables into numerical representations, as needed.
3. Linear Regression: Linear regression was applied to model the relationship between variables in the datasets. This statistical technique helps to identify and quantify the linear relationship between independent variables (such as greenhouse gas emissions) and dependent variables (such as life expectancy).
4. Principal Component Analysis (PCA): PCA was used for data visualization and dimensionality reduction. PCA is a technique that identifies the most important features or components in a dataset and projects the data onto a lower-dimensional space, making it easier to visualize and analyze.
5. Bar Joseph Seriation Algorithm: The Bar Joseph seriation algorithm is a method used for reordering rows and columns of a matrix to reveal patterns or structures in the data.

It may have been applied to rearrange the data in a meaningful way for further analysis or visualization.

6. Bertin's Permutation Matrix: Bertin's permutation matrix is a graphical technique used for data visualization, particularly for comparing two datasets. It may have been used to display the relationships between variables in a visually appealing and informative way.

The project likely aimed to analyse the relationship between greenhouse gas emissions and life expectancy using various statistical techniques. Linear regression was used to model the relationship between the two variables, while PCA was used for data visualization and dimensionality reduction. The Bar Joseph seriation algorithm and Bertin's permutation matrix were likely used to reorganize and visualize the data in a meaningful way. The project may have provided insights into how greenhouse gas emissions impact life expectancy and explored potential patterns or structures in the data.

4. PROPOSED WORKS

4.1 Novelty

The novelty of the project lies in the combination of different techniques and algorithms for analyzing and visualizing the relationship between greenhouse gas emissions and life expectancy. The project appears to utilize linear regression, principal component analysis (PCA), the Bar Joseph seriation algorithm, and Bertin's permutation matrix to gain insights from the datasets.

The use of linear regression allows for quantifying the relationship between greenhouse gas emissions and life expectancy, which could provide valuable insights into how changes in emissions may impact human health and well-being. PCA, on the other hand, is a technique commonly used for dimensionality reduction and data visualization, but its application in the context of greenhouse gas emissions and life expectancy may offer novel insights into patterns or trends in the data.

The Bar Joseph seriation algorithm and Bertin's permutation matrix are less commonly used techniques in data analysis, but they could potentially provide unique ways of reordering and visualizing the data, revealing previously unknown patterns or structures.
Overall, the novelty of the project may lie in the combination of these different techniques and algorithms, applied to the specific context of greenhouse gas emissions and life expectancy, potentially leading to novel insights and findings that contribute to the understanding of the relationship between these variables.

4.2 Project Contributions

All the team members have made significant contributions to the project. Priyanshu Kapoor worked on developing and testing machine learning models which has provided valuable insights into predicting the results of matches. MOHD Saifee’s visualizations have helped to communicate complex data in an easy-to-understand format, making the insights accessible to a wider audience. Pragnya Prakash Panda worked on documentation, pre-processing and handling the dataset which ensured that the data used in the project is accurate and relevant.

Together, our contributions have resulted in a comprehensive analysis of the international carbon dioxide emission among different countries. The project's visualizations and machine learning models have provided insights into the emission of Carbon Dioxide rate over time, the factors that are responsible for it like population, etc., geographical impact on emission. This analysis has the potential to reduce global warming, could save ozone layer depletion so that we will get protection from harmful ultraviolet rays directly from sun and will save the environment.

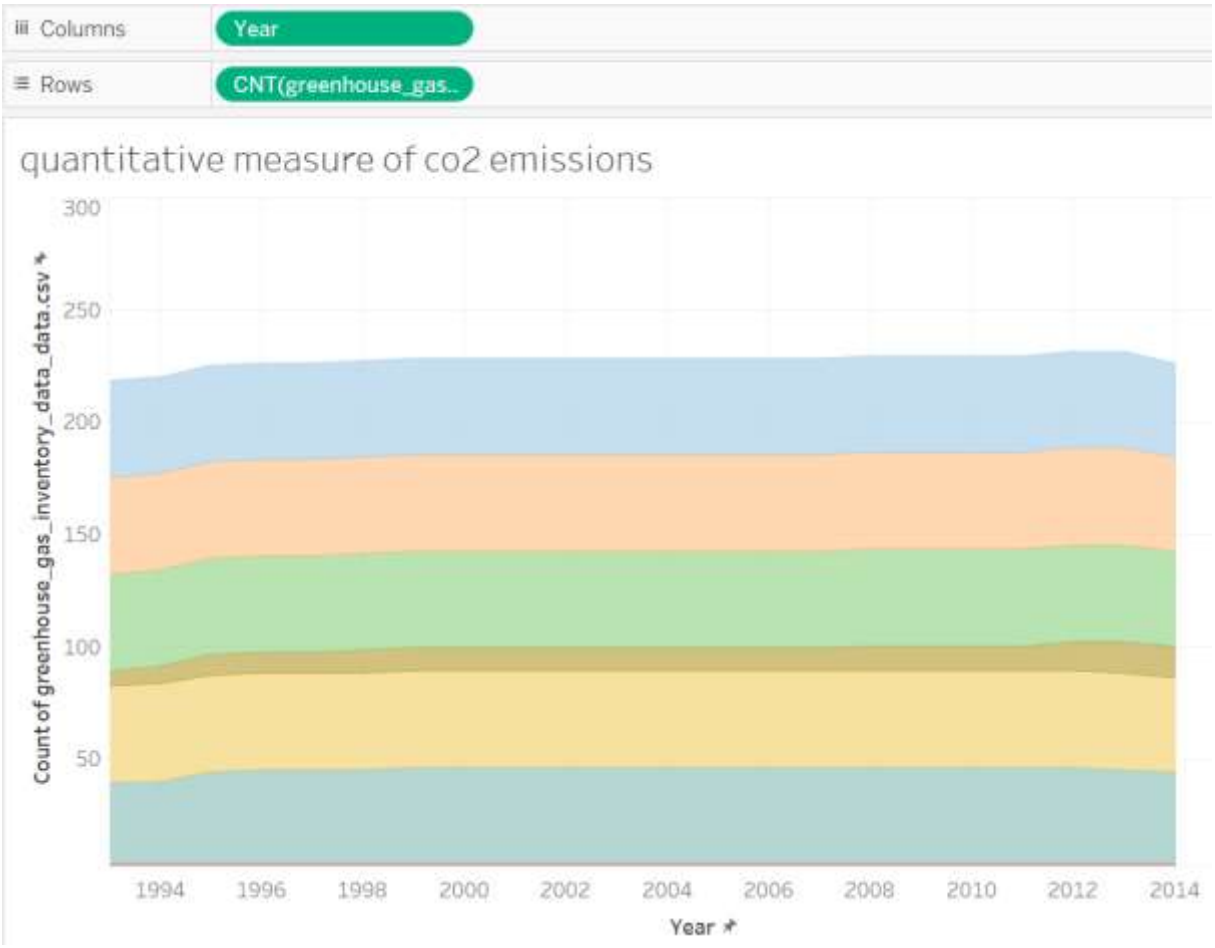
5. RESULT AND DISCUSSION

5.1 Results

The data analysis of international carbon dioxide emission revealed some fascinating insights about . In addition to this, machine learning models were utilised on the dataset, with Principal Component Analysis and Linear Regression proving to be the most successful in terms of accurately predicting the CO2 emission rate. The findings have repercussions for the assessment of the different countries responsible for high CO2 emission rate, the forecasting of the emission rates in future, and the formulation of strategies for reducing CO2 and greenhouse gases emission rate and saving the environment.

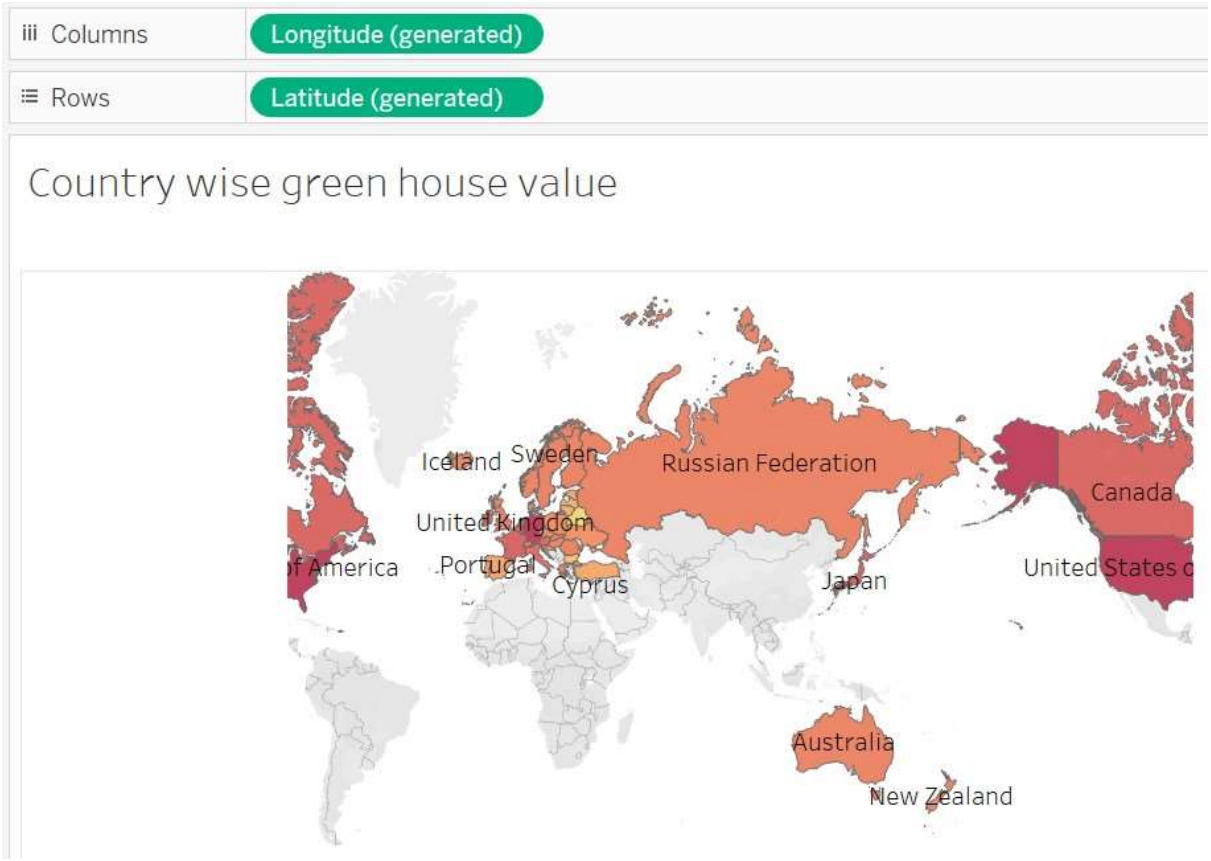
5.2 Figures and Comparisons

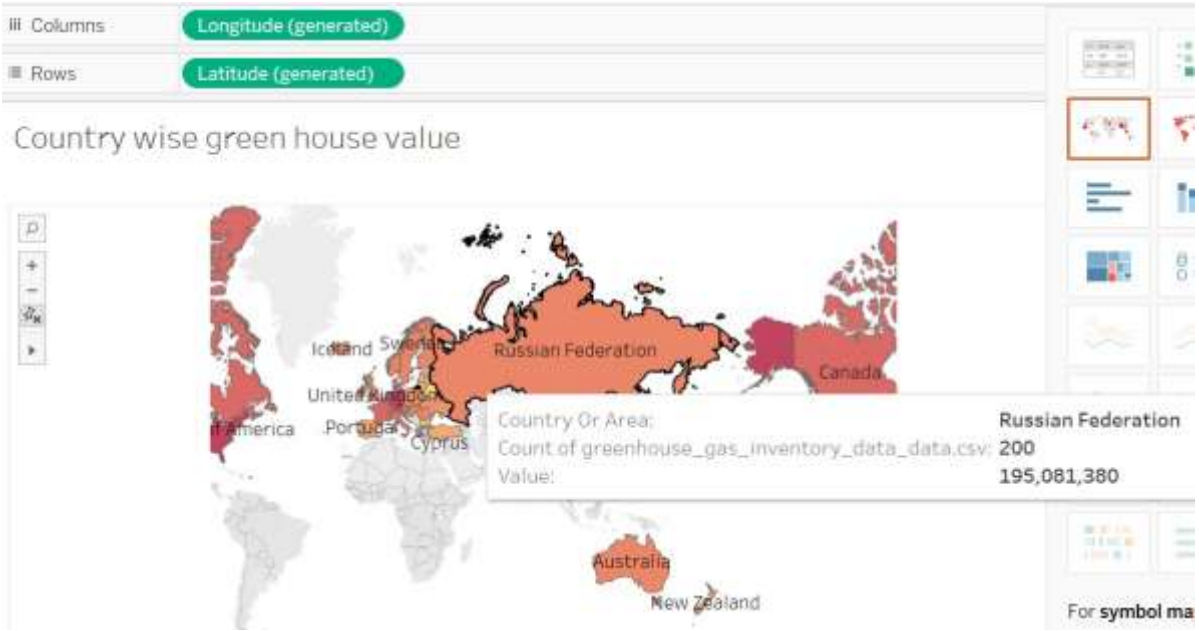
1. Quantitative measure of CO2 emissions



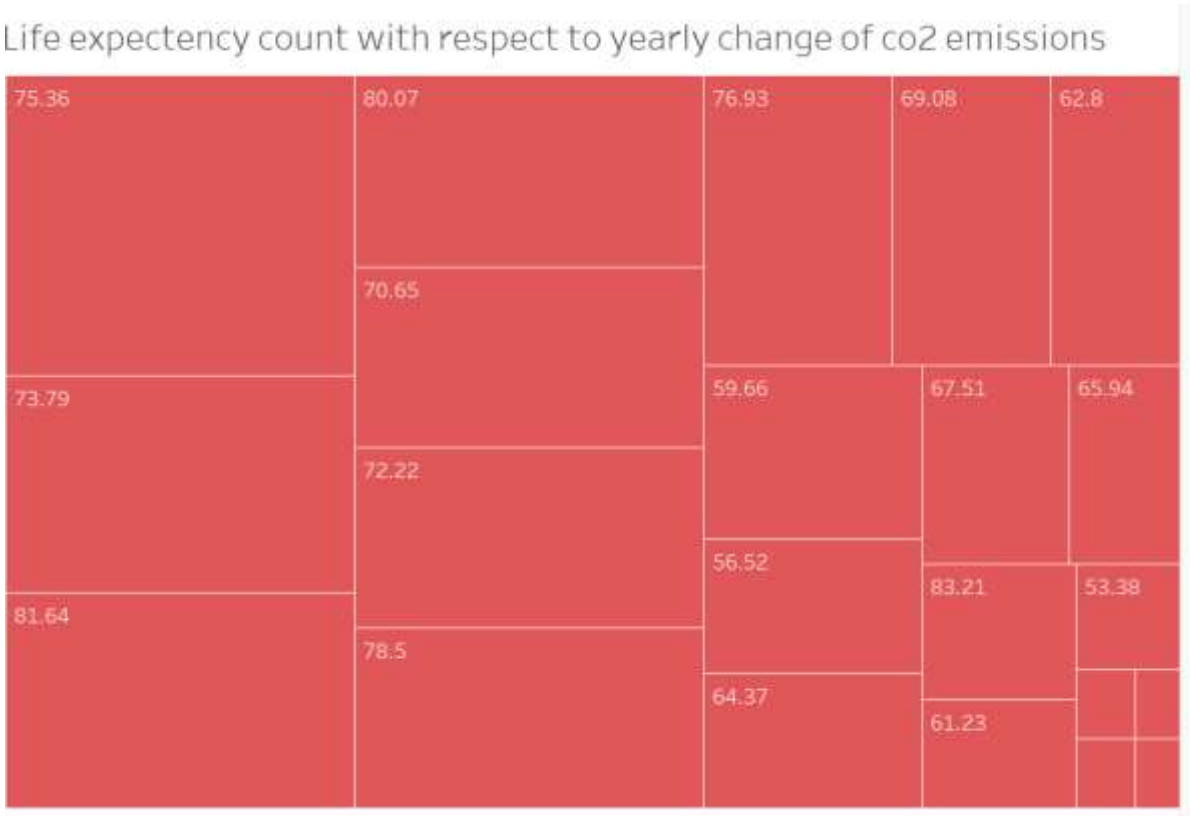
2. Country wise per capita

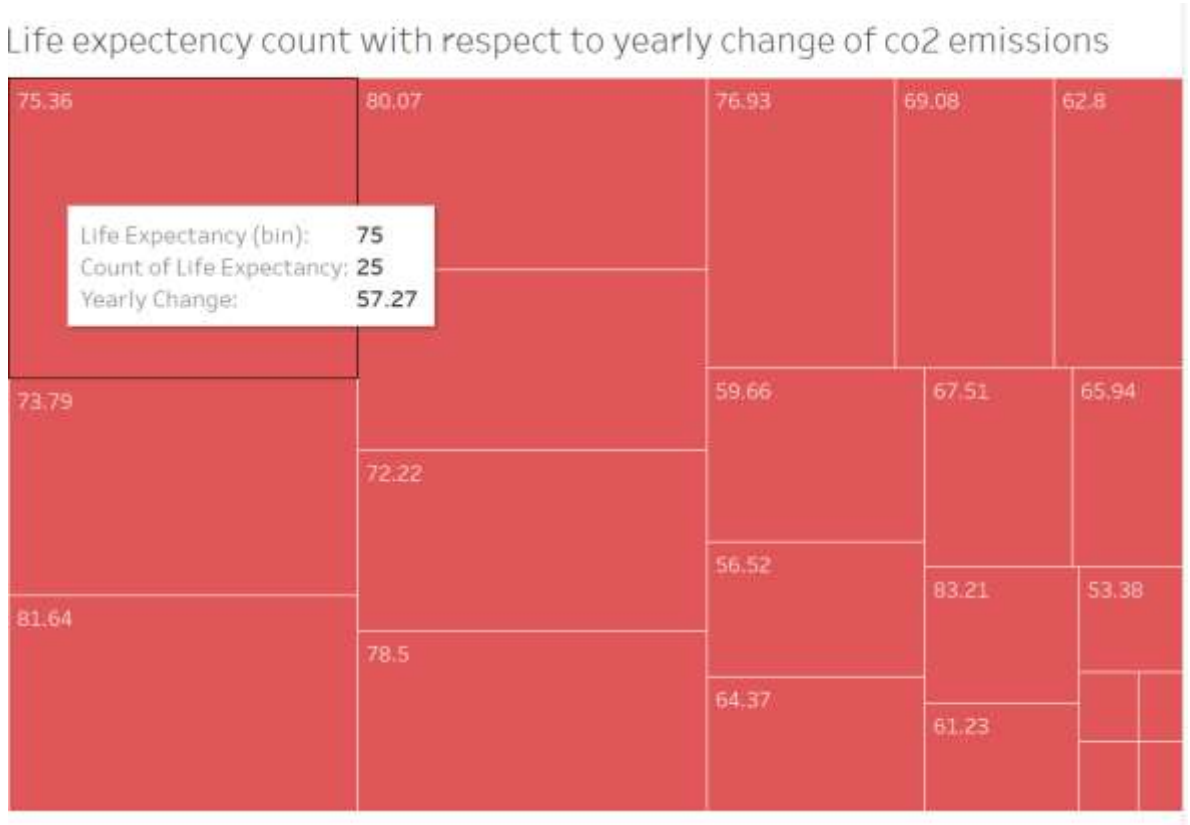
3. Country wise green-house value





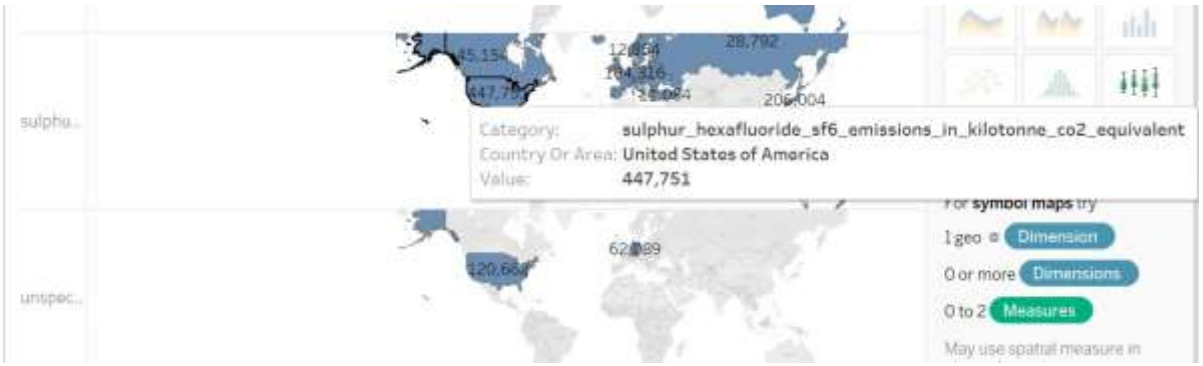
4. Life expectancy count with respect to yearly change if CO2 emissions.



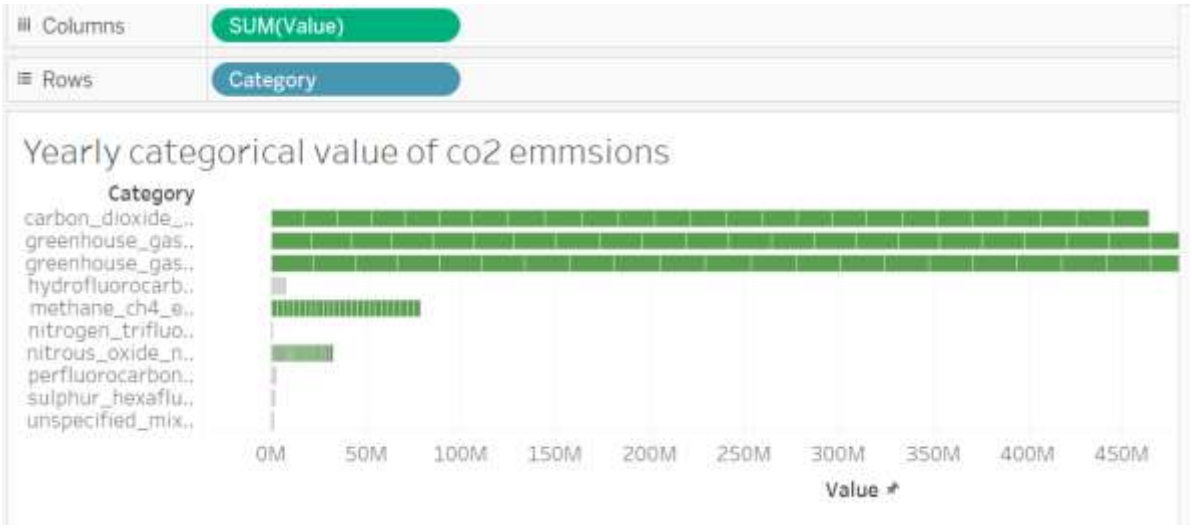


5. Ways of CO2 emissions country wise

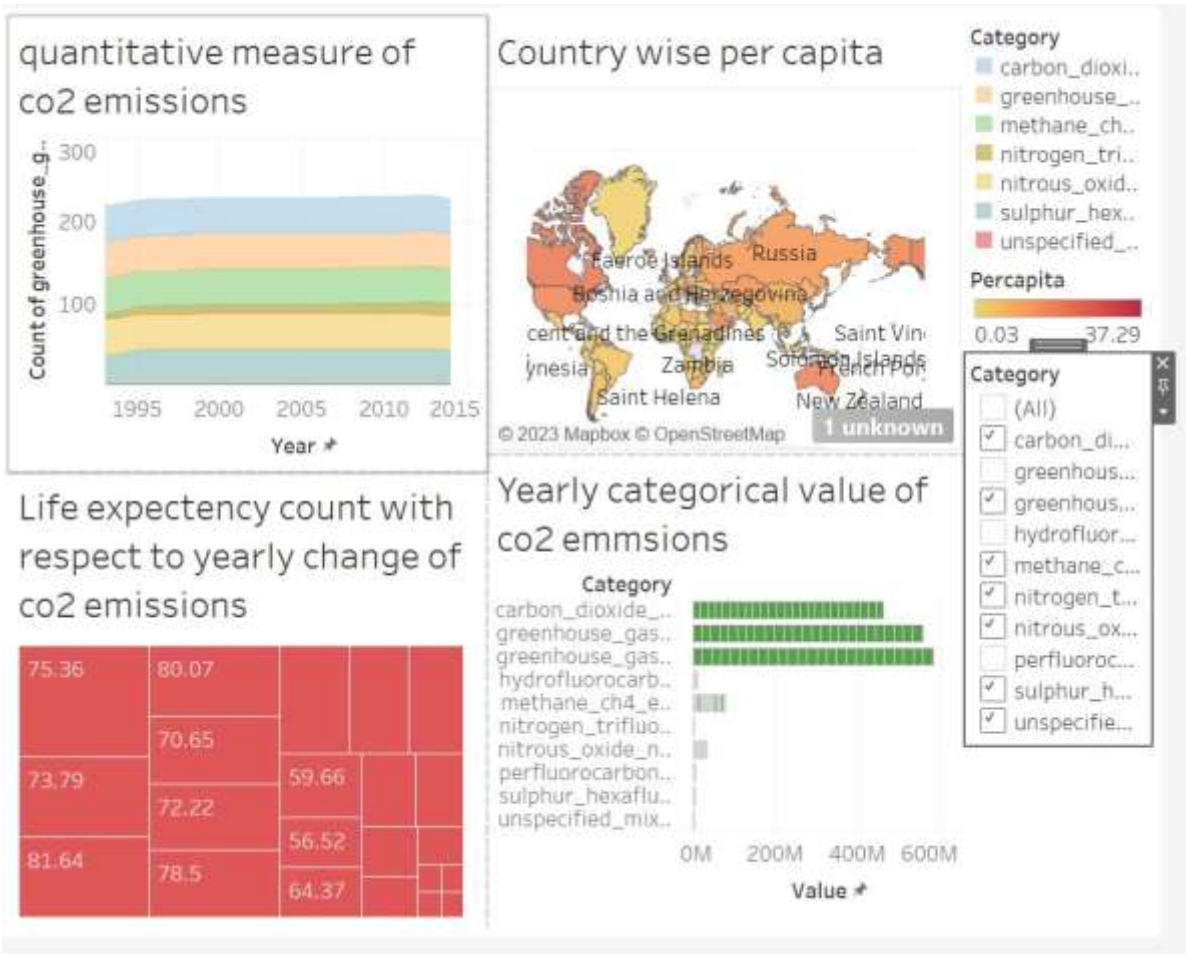

Columns
Longitude (generated)
Rows
Category
Latitude (generated)
ways of co2 emissions counry wise
Catego..
nitroge..
nitrous..
perfluo..
8,653
14,835
2,671,264
10,224,049
669,703
210,981
Category: perfluorocarbons_pfcs_emissions_in_kilotonne_co2_equivalent
Country Or Area: Russian Federation
Value: 210,981
For symbol maps try
May use spatial measure in place of geo dimension


sulphu..
unspec..
28,792
206,004
Category: sulphur_hexafluoride_sf6_emissions_in_kilotonne_co2_equivalent
Country Or Area: United States of America
Value: 447,751
For symbol maps try
1 geo Dimension
0 or more Dimensions
0 to 2 Measures
May use spatial measure in

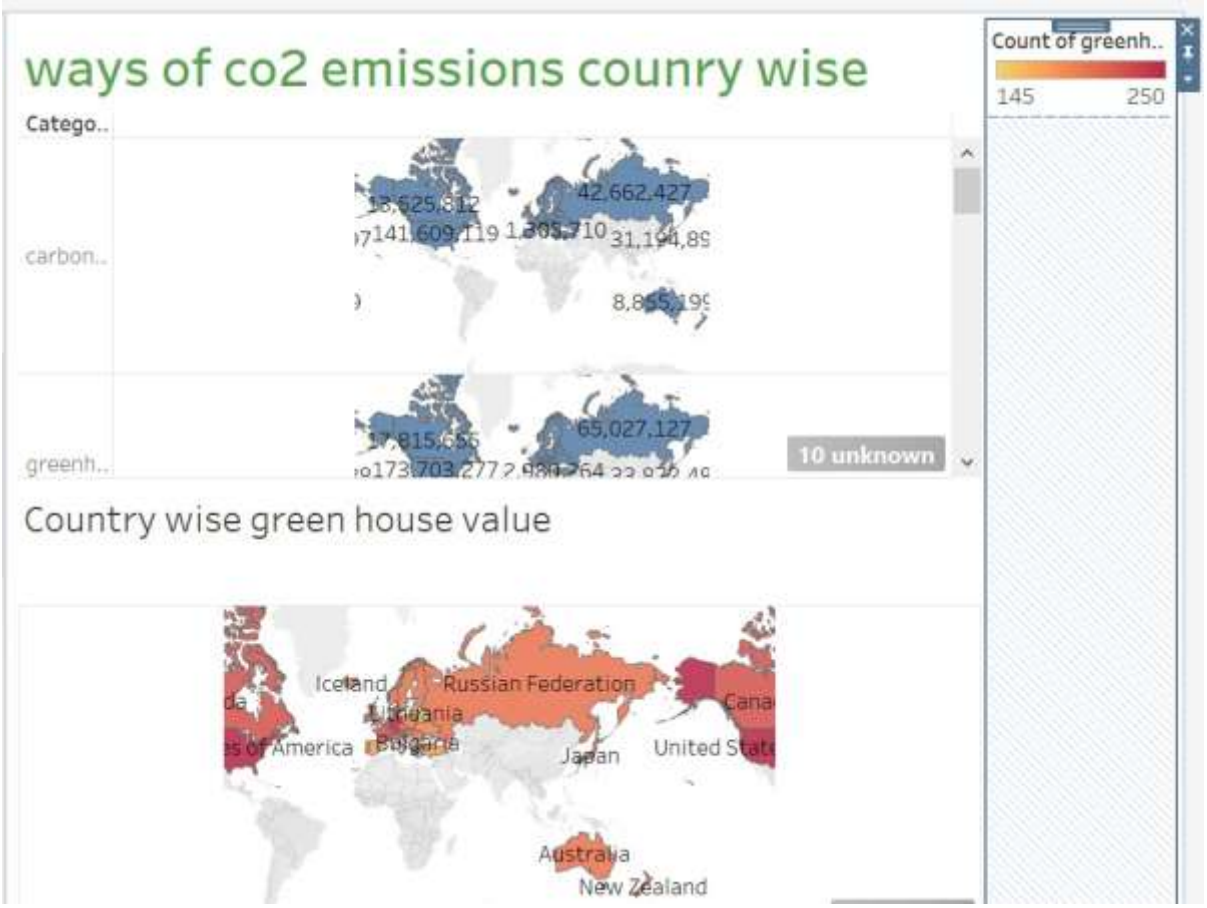
6.Yearly categorical value of CO2 emissions



7.Dashboard 1



8.Dashboard 2



5.3 Explanation

We have used linear regression, principal component analysis (PCA) for visualization, Bar Joseph seriation algorithm, and Bertin's permutation matrix on two datasets: CO2Emission_LifeExp.csv and greenhouse_gas_inventory_data_data.csv. These datasets contain information about carbon dioxide (CO2) emissions and greenhouse gas (GHG) inventory data.

1. Linear Regression: Linear regression is a statistical technique used to model the relationship between two variables by fitting a linear equation to observed data. In your project, you may have used linear regression to understand the relationship between CO2 emissions and life expectancy. You could have used the CO2Emission_LifeExp.csv dataset to perform linear regression, with CO2 emissions as the independent variable and life expectancy as the dependent variable. This could have helped you identify whether there is a positive or negative correlation between CO2 emissions and life expectancy, and quantify the strength and direction of this relationship.
2. Principal Component Analysis (PCA) for Visualization: PCA is a dimensionality reduction technique used to transform high-dimensional data into a lower-dimensional space while preserving the most important information. In your project, you may have used PCA to visualize the relationship between CO2 emissions and GHG inventory data. You could have applied PCA on the greenhouse_gas_inventory_data_data.csv dataset to reduce the dimensionality of the data and create a 2D or 3D scatter plot to visualize the data points in a lower-dimensional space, making it easier to identify patterns and trends.
3. Bar Joseph Seriation Algorithm: The Bar Joseph seriation algorithm is a method used for reordering rows and columns in a matrix to reveal hidden patterns. In your project, you may have used this algorithm to rearrange the data in the CO2Emission_LifeExp.csv or greenhouse_gas_inventory_data_data.csv dataset to uncover patterns or trends that may not be apparent in the original dataset. The algorithm could have helped you identify clusters or groups of countries or regions

that exhibit similar patterns of CO2 emissions or GHG inventory data, providing insights into regional or global trends.

4. Bertin's Permutation Matrix: Bertin's permutation matrix is a technique used for reordering rows and columns in a matrix to improve data visualization. In your project, you may have used Bertin's permutation matrix to rearrange the data in the CO2Emission_LifeExp.csv or greenhouse_gas_inventory_data_data.csv dataset to create visualizations that highlight patterns or trends in the data. By rearranging the data using the permutation matrix, you could have created visualizations that make it easier to identify and interpret patterns or trends in the CO2 emissions or GHG inventory data.

Overall, your project likely involved using linear regression to model the relationship between CO2 emissions and life expectancy, applying PCA for visualization of CO2 emissions and GHG inventory data, using the Bar Joseph seriation algorithm to reveal hidden patterns in the data, and utilizing Bertin's permutation matrix for data visualization. These techniques would have provided valuable insights into the relationship between CO2 emissions, GHG inventory data, and other factors related to the environment, helping to inform environmental policy decisions or identify areas for further research and action.

6. CONCLUSION

Carbon emission research has received increasing global attention due to rapid global climate change. Thus, academics, international organizations, and government agencies have paid special consideration to identifying the carbon emission sources and thereby implementing various carbon mitigation strategies. The initial research on carbon emissions was conducted in 1981 and the domain has evolved significantly over the past two decades.

We can see that there is a perfectly negative relationship with yearly change mostly and the components having a negative relationship are Life Expectancy, PerCapita, and Co2 emissions. Population and CO2 Emissions have highly positive relationships. Population plays a major role in increasing CO2 Emissions.

GitHub Link: https://github.com/prakash-panda007/INTERNATIONAL-CO2-EMISSION-ANALYSIS-AND-VISUALISATION.git

7. REFERENCES

[1] https://www.iea.org/reports/global-energy-review-2021/co2-emissions

[2] https://www.iea.org/reports/co2-emissions-in-2022

[3] https://ourworldindata.org/co2-emissions

[4] https://www.epa.gov/ghgemissions/global-greenhouse-gas-emissions-data

[5] https://www.c2es.org/content/international-emissions/

[6] https://data.worldbank.org/indicator/EN.ATM.CO2E.PC

[7] https://edgar.jrc.ec.europa.eu/report_2022

[8] https://www.carbonbrief.org/analysis-global-co2-emissions-from-fossil-fuels-hit-record-high-in-2022/

[9] https://www.statista.com/statistics/276629/global-co2-emissions/

[10] https://en.wikipedia.org/wiki/List_of_countries_by_carbon_dioxide_emissions

[11] https://applications.icao.int/icec

[12] http://www.globalcarbonatlas.org/en/CO2-emissions