A green letter and black text

Description automatically generated

**AGRIFOOD CO2 EMISSION**

**Introduction:**

The Agri-Food CO2 Emission Dataset Forecasting with Machine Learning project is an ambitious endeavor that aims to analyze and predict carbon dioxide (CO2) emissions associated with the agricultural and food sectors. The dataset used in this project, available on Kaggle and curated by Alessandro Lobello, encompasses a comprehensive collection of relevant variables, offering valuable insights into the environmental impact of various agri-food activities.

The dataset comprises a diverse set of columns that cover a wide range of factors influencing CO2 emissions in the agri-food sector. These columns include information on the geographic area, year of observation, emissions from savanna fires and forest fires, crop residues, rice cultivation, drained organic soils, pesticides manufacturing, food transport, forestland, net forest conversion, food household consumption, food retail, on-farm electricity use, food packaging, agrifood systems waste disposal, food processing, fertilizers manufacturing, industrial processes and product use (IPPU), manure management, fires in organic soils, fires in humid tropical forests, on-farm energy use, rural population, urban population, total population by gender, total emissions, and average temperature.

The project's primary objective is to gain a deeper understanding of the interplay between these diverse factors and their cumulative impact on CO2 emissions. By employing machine learning techniques, the project endeavors to forecast and predict future emissions trends, providing valuable insights for policymakers, researchers, and environmentalists to devise effective strategies for mitigating climate change and promoting sustainable practices within the agri-food industry.

**Dataset Overview:**

The dataset is collected from Kaggle. This dataset contains 6965 rows where the attributes represent the Savanna fires, Forest fires, Crop residues, Rice cultivation etc.

The agricultural CO2 emission dataset was meticulously compiled by integrating data from multiple reputable sources such as the Food and Agriculture Organization (FAO) and the Intergovernmental Panel on Climate Change (IPCC). Through careful cleaning, preprocessing, and merging of these datasets, a comprehensive and coherent dataset was generated, serving the purpose of analysis and forecasting.

**Dataset Source:**

<https://www.kaggle.com/datasets/alessandrolobello/agri-food-co2-emission-dataset-forecasting-ml>

**Attributes:**

A total of 29 attributes are used to describe Agrifood CO2 Emission. Below is the list.

1. **Savanna fires**: Emissions from fires in savanna ecosystems.
2. **Forest fires**: Emissions from fires in forested areas.
3. **Crop Residues**: Emissions from burning or decomposing leftover plant material after crop harvesting.
4. **Rice Cultivation**: Emissions from methane released during rice cultivation.
5. **Drained organic soils (CO2)**: Emissions from carbon dioxide released when draining organic soils.
6. **Pesticides Manufacturing**: Emissions from the production of pesticides.
7. **Food Transport**: Emissions from transporting food products.
8. **Forestland**: Land covered by forests.
9. **Net Forest conversion**: Change in forest area due to deforestation and afforestation.
10. **Food Household Consumption**: Emissions from food consumption at the household level.
11. **Food Retail**: Emissions from the operation of retail establishments selling food.
12. **On-farm Electricity Use**: Electricity consumption on farms.
13. **Food Packaging**: Emissions from the production and disposal of food packaging materials.
14. **Agrifood Systems Waste Disposal**: Emissions from waste disposal in the agrifood system.
15. **Food Processing**: Emissions from processing food products.
16. **Fertilizers Manufacturing**: Emissions from the production of fertilizers.
17. **IPPU**: Emissions from industrial processes and product use.
18. **Manure applied to Soils**: Emissions from applying animal manure to agricultural soils.
19. **Manure left on Pasture**: Emissions from animal manure on pasture or grazing land.
20. **Manure Management**: Emissions from managing and treating animal manure.
21. **Fires in organic soils**: Emissions from fires in organic soils.
22. **Fires in humid tropical forests**: Emissions from fires in humid tropical forests.
23. **On-farm energy use**: Energy consumption on farms.
24. **Rural population**: Number of people living in rural areas.
25. **Urban population**: Number of people living in urban areas.
26. **Total Population - Male**: Total number of male individuals in the population.
27. **Total Population - Female**: Total number of female individuals in the population.
28. **total emission**: Total greenhouse gas emissions from various sources.
29. **Average Temperature °C**: The average increasing of temperature (by year) in degrees Celsius.

**Tools:**

1. Python (Jupyter notebook)
2. Tableau

**Exploratory Data Analysis:**

**A screenshot of a computer

Description automatically generated**

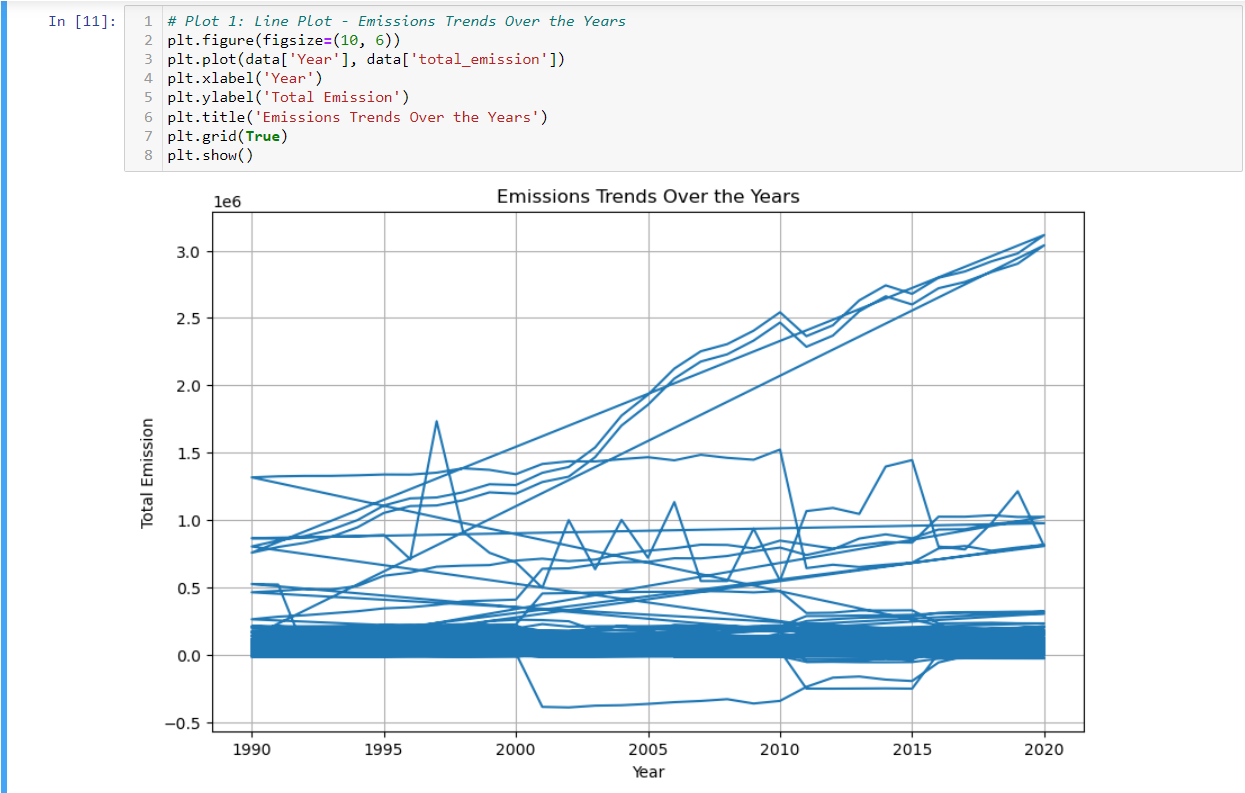
* These Columns from dataset Crop Residues, Forestland, Net Forest conversion, Food Household Consumption, IPPU, Manure applied to Soils, Manure Management, On-farm energy use has been removed. Columns are not useful for further analysis.

**A screenshot of a computer

Description automatically generated**

Above is used to explore the missing values in the dataset. Below are the visualizations used for analyzing, identifying the relations between the attributes.

**1. Emissions Trends Over the Years**



Line plot shows a graph of global emissions trends over the years. The graph shows that emissions have been increasing steadily since 1990, with a slight dip in 2008 during the financial crisis. The total emissions in 2020 were about 2.5 times higher than they were in 1990. This increase in emissions is a major contributor to climate change.

**2. Comparison of Emissions from Different Sources**

The bar plot shows a graph comparing emissions from different sources. The largest source of emissions is Rice Cultivation followed by Drained organic soils and then Savanna Fires, crop residues and Forest fires.

A screenshot of a graph

Description automatically generated

**3. Relationship Between Temperature and Emissions**

A screen shot of a graph

Description automatically generated

The above plot shows a scatter plot of the relationship between temperature and emissions. The scatter plot shows that there is a positive correlation between temperature and emissions, meaning that as temperature increases, emissions also tend to increase. The correlation is not perfect, however, as there are some data points that do not follow the trend. This may be due to other factors that influence emissions, such as economic activity or population growth.

**4. Population Distribution in Urban and Rural Areas**

A screen shot of a computer

Description automatically generatedA graph with orange and blue lines

Description automatically generated

The image shows a graph of the population distribution in urban and rural areas in the Democratic People's Republic of Korea (DPRK). The graph shows that the percentage of the population living in urban areas has been increasing steadily over the past few decades, while the percentage of the population living in rural areas has been decreasing. In 1990, only about 20% of the DPRK population lived in urban areas, but by 2020, this number had increased to over 40%. This trend is likely due to a few factors, including economic development, urbanization, and rural-to-urban migration.

**5. Contribution of Different Sources to Total Emissions**

A screen shot of a graph

Description automatically generated

The image shows a graph of the contribution of different sources to total emissions over time. The graph shows that the contribution of fossil fuels to total emissions has been decreasing, while the contribution of other sources, such as agriculture and deforestation, has been increasing. In 1990, fossil fuels accounted for about 70% of total emissions, but by 2020, this number had decreased to about 50%. This trend is likely due to a few factors, including the increasing use of renewable energy sources, the improvement of energy efficiency, and the expansion of deforestation.

**6. Distribution of Average Temperature**

The image shows a box plot of the distribution of average temperature. The box plot shows that the average temperature is typically between 10 and 20 degrees Celsius, with a median of 15 degrees Celsius. There are a few outliers, with temperatures as low as 5 degrees Celsius and as high as 25 degrees Celsius. This suggests that the distribution of temperature is somewhat skewed, with a longer tail on the high end.

A screen shot of a graph

Description automatically generated

**Hypothesis:**

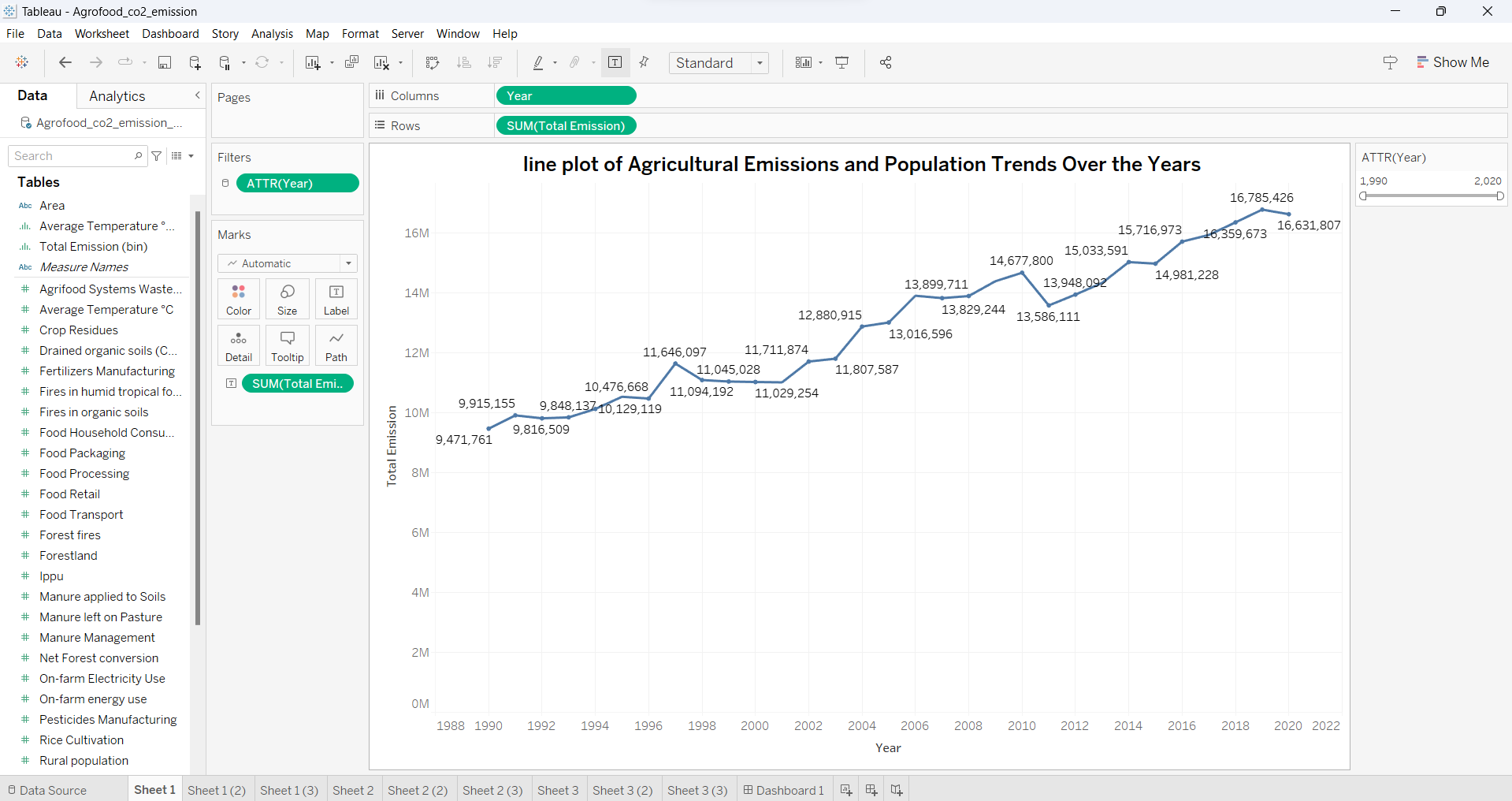
1. Agricultural Emissions and Population Trends Over the Years.

2. Impact of Different Emission Sources on Total Emissions.

3. Relationship Between Average Temperature and Emissions Based on year.

**1. Agricultural Emissions and Population Trends Over the Years.**

For this hypothesis I have taken Line, Histogram and Tree maps.

****

The image shows a line graph of agricultural income and population trends over the years. The graph shows that agricultural income has been increasing steadily over the years, while the population has also been increasing. However, the rate of increase in agricultural income has been slower than the rate of increase in population, which has led to a decrease in agricultural income per capita.

**A screenshot of a computer

Description automatically generated**

The image shows a histogram of agricultural emissions and population trends over the years. The histogram shows that agricultural emissions have been increasing steadily over the years, while the population has also been increasing. However, the rate of increase in agricultural emissions has been slower than the rate of increase in population, which has led to a decrease in agricultural emissions per capita.

**A screenshot of a computer

Description automatically generated**

The image shows a tree map of agricultural emissions and population trends over the years. The tree map shows that agricultural emissions have been increasing steadily over the years, with the largest increase in emissions coming from the livestock sector. The population has also been increasing steadily over the years, but the rate of increase in population has been slower than the rate of increase in agricultural emissions.

**2. Impact of Different Emission Sources on Total Emissions.**

For this hypothesis I have taken Area, Scatter and Bubble plot.

The image shows an area plot of the impact of different emission sources on total emissions. The area of each segment is proportional to the total emissions from that source.

**A graph on a computer screen

Description automatically generated**

The scatter plot shows that there is a positive correlation between the emission sources and total emissions, meaning that as the emission sources increase, so do total emissions. However, the correlation is not perfect, as there are some data points that do not follow the trend.

**A screen shot of a graph

Description automatically generated**

The bubble plot shows the impact of different emission sources on total emissions. The bubbles on the plot are proportional to the total emissions from each source. The size of the bubble also indicates the relative importance of the emission source.

**A screenshot of a computer screen

Description automatically generated**

**3. Relationship Between Average Temperature and Emissions Based on year.**

For this hypothesis I have taken Combination, Histogram and Dual lines plot.

A graph on a computer screen

Description automatically generated

The image shows a combination plot of the relationship between average temperature and emissions based on year. The combination plot consists of a line graph showing the trend of average temperature over time, and a scatter plot showing the relationship between average temperature and emissions. The line graph shows that average temperature has been increasing steadily over the years, while the scatter plot shows that there is a positive correlation between average temperature and emissions, meaning that as average temperature increases, emissions also tend to increase.

A screenshot of a graph

Description automatically generated

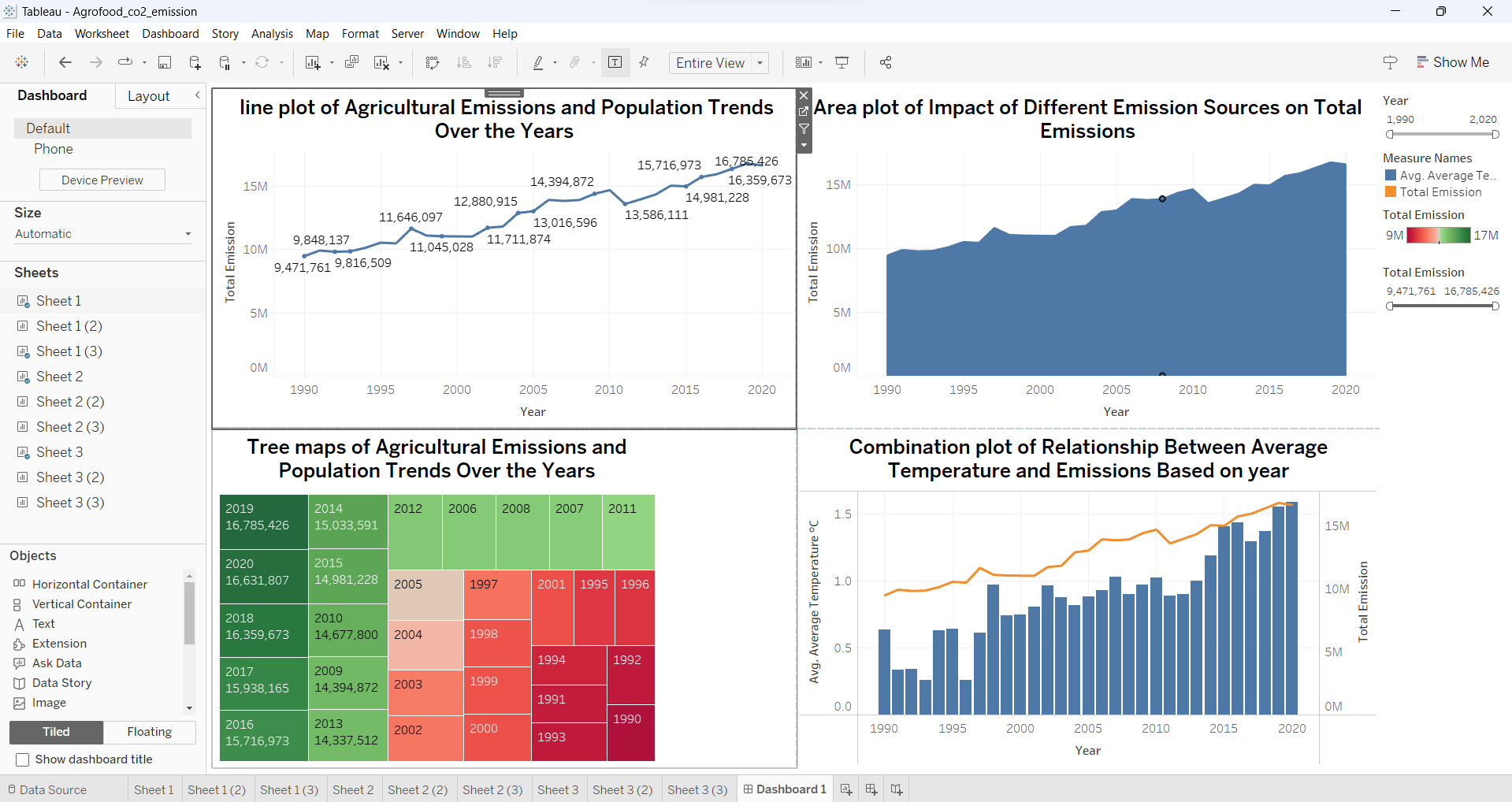
The histogram shows the relationship between average temperature and emissions rates for a year. There is a positive correlation between average temperature and emissions rates. This means that as the average temperature increases, emissions rates also tend to increase. The histogram also shows that there is a wide range of emissions rates for a given average temperature.

A graph on a computer screen

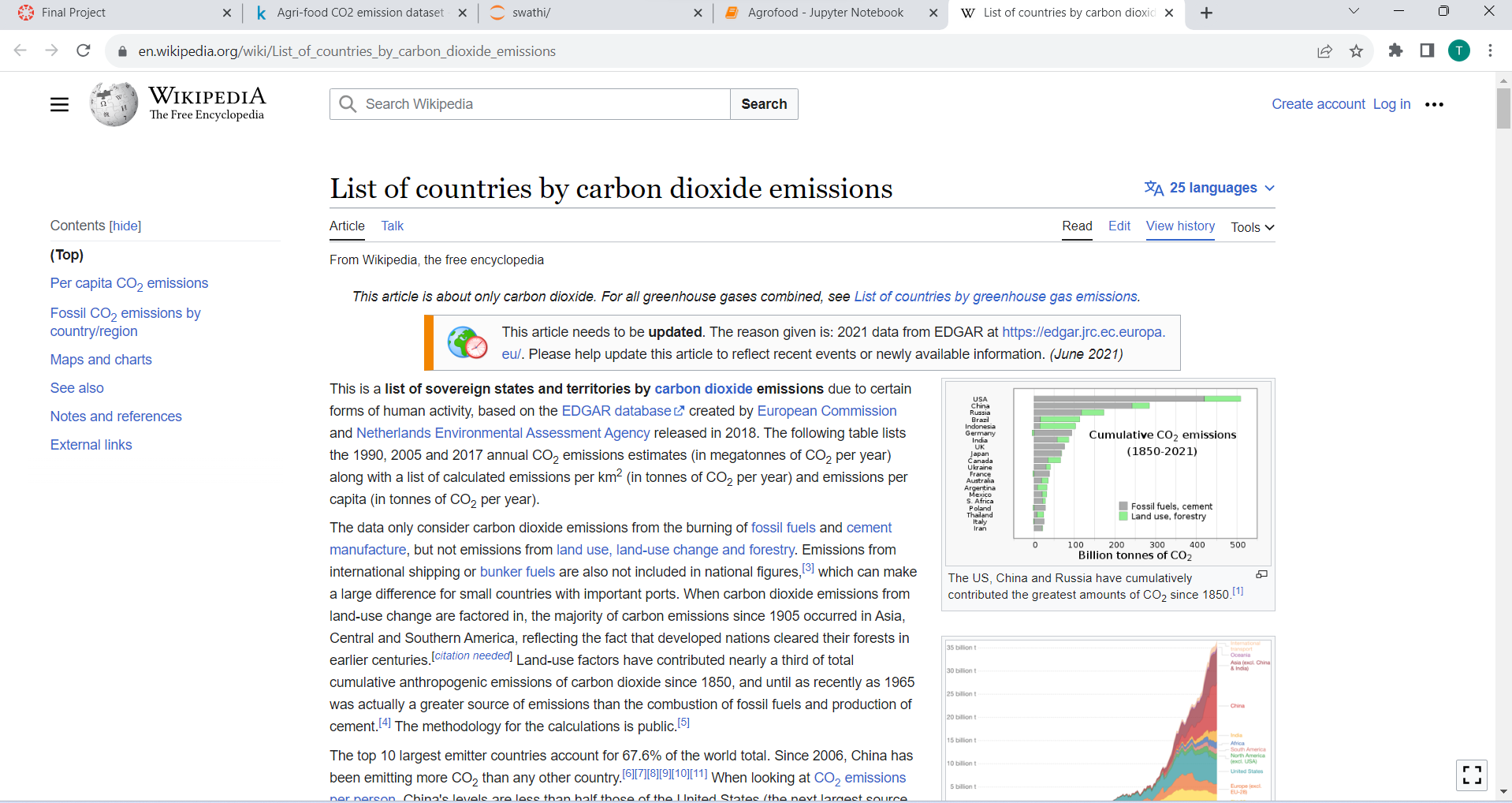
Description automatically generated

The two lines show the trend of average temperature and emissions over time. The line for average temperature shows that it has been increasing steadily over the years. The line for emissions shows that it has also been increasing, but at a slower rate.

**Dashboard:**

****

In dashboard the four charts together show that agricultural emissions have been increasing steadily over the years, with the largest increase in emissions coming from the livestock sector. The population has also been increasing steadily over the years, but the rate of increase in population has been slower than the rate of increase in agricultural emissions. This suggests that the agricultural sector is becoming more emission-intensive, as the overall demand for agricultural products is also increasing.



**Conclusion:**

In Conclusion, the analysis conducted in this project underscores the crucial role of the agri-food sector in contributing to global CO2 emissions and its significant impact on climate change. Gaining a comprehensive understanding of the diverse factors influencing emissions is essential for shaping sustainable policies and practices aimed at mitigating these effects. The utilization of machine learning and forecasting techniques has proven valuable in projecting future emission trends, enabling the formulation of effective strategies to combat climate change.

By prioritizing sustainable approaches, managing population growth, preserving forests, and adopting responsible agricultural practices, we can collectively strive towards a more sustainable future, benefiting the agri-food industry and the entire planet.

**References:**

[1]. Co2 Emissions by Hannah Ritchie and Max Roser

website: <https://ourworldindata.org/co2-emissions>

[2]. Co2 Emissions in 2022 by International Energy Agency

website: <https://www.iea.org/reports/co2-emissions-in-2022>

[3]. Emissions data are sourced from Climate Watch Historical GHG Emissions(2020)

website: <https://data.worldbank.org/indicator/EN.ATM.CO2E.PC>