Project Report

Analysis of Global Land Surface Temperature Change

Xueliu Fan,Vansh Sharma, and Sally Wei, MSDA, SJSU

Course: DATA 228

Professor: Andrew Bond

Abstract

Climate change has threatened human beings with rising sea levels, severe rain and storm, and extremely hot weather. One of the significant parameters indicating climate change is the continuous rising temperature which is called global warming. As researchers have recorded the global temperature for hundreds of years, it is worth analyzing the temperature data as well as other factors that can lead to global warming such as greenhouse gas, human activity etc.

This project is launched by studying the global land surface temperature using data analytics and machine learning approaches. Moreover, demographic and geography are collected to investigate the reasons for global warming. At last, the web-based API ‘TheGreenSpartan’ is developed and deployed for users to explore the historic and predicted data of the rising temperature. As an easy-to-use interactive website, it aims at helping people track human impact on the environment via data collected from four different sources. This document presents a review of our work related to our environmental research, data collection, and use of machine learning models for prediction, motivating people to contribute to the fight against climate change.

Table of Contents

1 Introduction

1.1 Background

Climate change has been a hot topic for decades; recently, The UN called it one of the major crises of our time. The United Nations has termed climate change a “global emergency” . The science behind climate change is very dense and politics gets in the way [1]. So let’s start by defining climate change.

Climate change is a long-term change in the earth’s weather patterns, including changes in precipitation, temperature, and wind patterns. According to NASA 97% of the climate scientists believe that climate change is a result of human activity and the industrial revolution gave rise to up to 50% of the warming of the Earth’s temperature [2]. The UN has been organizing Climate Change Conferences (UNFCCC) since 1995, which aims at bringing 197 countries together to commit to tackling climate change [3]. The Paris Agreement was signed on 12 December 2015 by 196 countries and was enforced on 4 November 2016. It is a legally binding international treaty on climate change which aims to limit global warming to below 2 or preferably 1.5 degree Celsius. Its aim is that the countries should reach their goal of being carbon neutral by mid century.

“Climate change” and “global warming” are often used interchangeably but they have different meanings just like climate and weather. Weather is the local atmospheric conditions calculated over a small period i.e minutes, hours, or days like how weather stations measure snow, and rain for the month while climate is a global average of temperature, climate, and rainfall patterns over years or decades. Climate change is a long-term change in Earth’s weather patterns while Global warming is the long-term heating of Earth’s temperature which was observed since the pre-industrial period (between 1850 and 1900). It is the average increase in Earth’s global surface temperature. In our project, we have used two parameters temperature and precipitation data dating back to 1957. So to bring awareness about climate change among our SJSU community we have built a website and named it “The Green Spartan”.The motivation behind our project was that as we know various countries are experiencing heat waves in Pakistan and India rising temperatures have forced schools to shut down and forced people to stay indoors. It has damaged crops and disrupted the supply chain which is putting pressure on energy supplies. Business Insider reported that in India birds are falling from the sky amid heatwaves [4]. Vice reported that March was the hottest month India had in 122 years and Pakistan's Jackobad in arid Singh province was the hardest hit with the temperature hitting 51 degrees celsius [5].With our project we aim to inspire every student in SJSU to contribute to fight against climate change. By making small changes in our daily life we can make a huge impact, our future scope of this project is to make a customized carbon footprint app that uses AI and machine learning to provide you with tips on how one can reduce carbon emission. Our website has a simple UI where the user can select the region of choice and see the surface temperature change over the years.

1.2 Literature review

Our project is divided into two parts. The first is building a machine learning model and the second is visualizing the result on a user-friendly website. The inspiration of this project comes from NASA. They have the largest number of climate change scientists and actively monitor atmospheric conditions, global temperatures, land cover and vegetation, ice extent, ocean productivity, and several other planetary vital signs with the help of satellites and sensors [2]. We came across an interesting tool "Eyes on Earth" which is an interactive 3D visualization tool provided by NASA which lets you use the satellites' gathered data to monitor Earth's vital signs, including carbon dioxide and carbon monoxide readings as well as sea level and soil moisture levels.

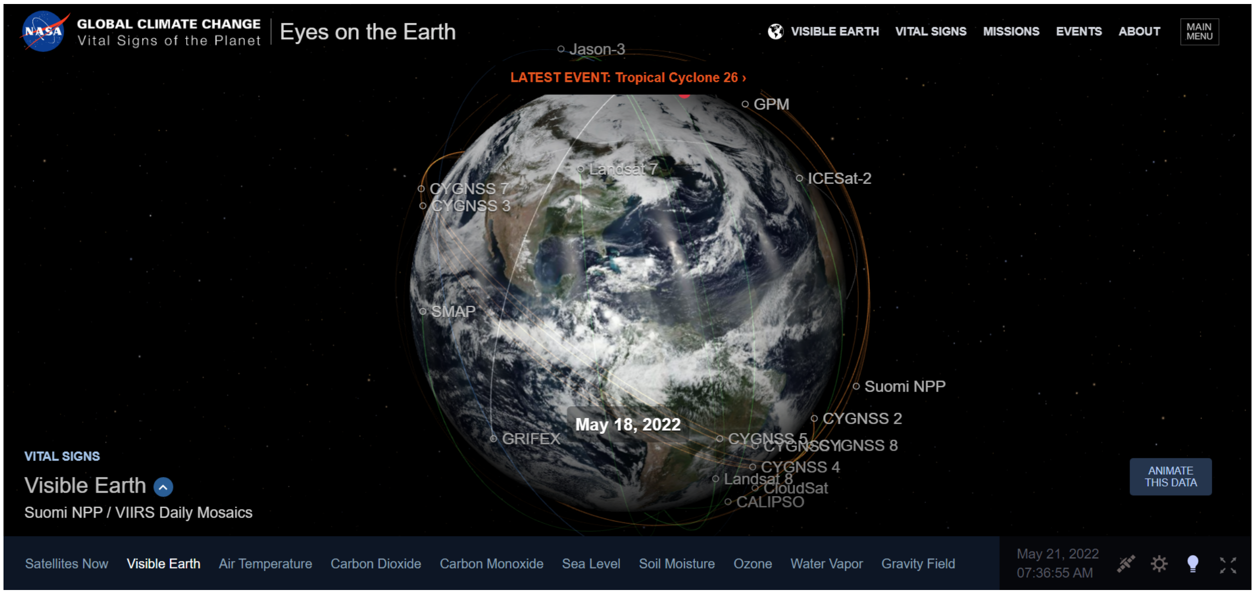


Figure 1.1 Eyes on earth [2]

1.3 Objective

The main objective of our project is to spread awareness about climate change by providing mathematical and statistical evidence of the science behind Climate Change. We aimed to provide a tool to predict the temperature based on geographic and demographic information. For visualization, we made a user-friendly website that will encourage people to take action against Climate Change at the individual level. We analyzed global land surface temperature change and the resultant precipitation increase.

2 Project description

2.1 Project design

The project workflow consists of four sections: data ingestion, data processing, data analysis, and application, as seen in figure 2.1.

The data related to the global temperature changes are collected and ingested into the Google Cloud Platform Storage. The data extraction, data transformation, and data loading pipeline are applied to integrate several data sources. The data processing includes the data cleaning and preparation. The pre-processed data is conducted via Google Cloud Platform Colab for data exploration and visualization, in which the major impacts on the temperature change are analyzed with AWS Quicksight. Since machine learning approaches are utilized to predict temperature change, the data will be prepared to train the model using Anaconda. The last section is to build a website to look up temperature for users. The website will be deployed on the Google Cloud Platform App Engine and to provide the historical data and the prediction data as well.

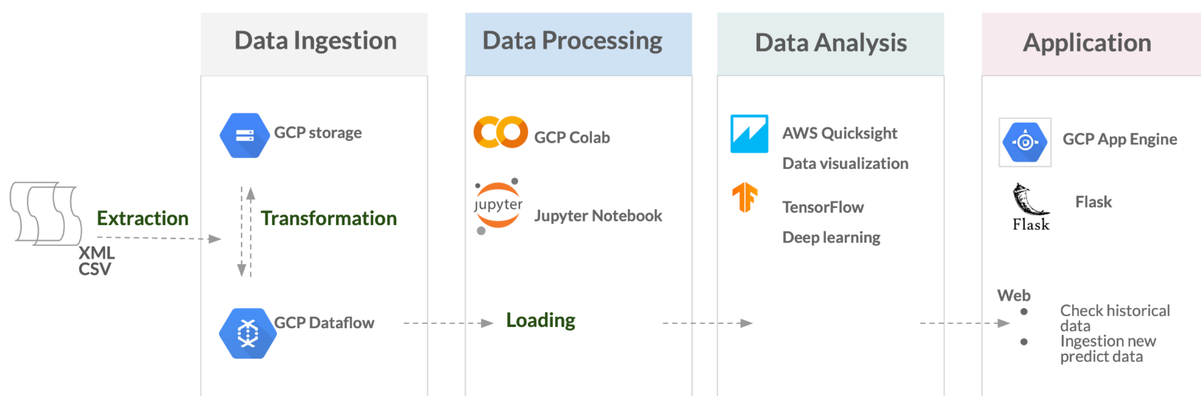


Figure 2.1 Project workflow diagram

2.2 User cases

The most significant parameters to indicate climate change are rising temperature and surface level. It is not easy to look up data of temperature changes in every country in one resource. Therefore, we design a system to let users check the temperature from any time in any country with one click operation. We also offer users the potential temperature data once the demographic information is updated.

2.3 Project resource

The cloud platform APIs are utilized including Google Cloud Platform Storage, Google Cloud Platform Colab, Google Cloud Platform App Engine, AWS Quicksight. The microservice platform Flask is applied in Python.

The python libraries are used to develop the machine learning models including Pandas, Numpy, Scipy, Matplotlib, Seaborn, and Tensorflow.

3 Data Engineering

3.1 Data collection

The data are collected from multiple resources as cited in references. They are carbon dioxide emissions, land surface temperature, country data, which also includes latitude, longitude, GDP, and demographics about each country from Google Developers [6] and Kaggle [7], and the precipitation data from Global Precipitation Climatology Center (GPCC) [8]. The temperature data are extracted from Berkeley Earth [9].

All the country datasets are provided as comma-separated files. The temperature data are scraped from the official web as a txt file, read through Python Pandas, and saved as comma-separated files.

3.2 Data processing for country data and temperature data

Data cleansing is conducted before and after the data pipeline. First, the datasets are converted from wide to long for GDP data and population data.. In detail, the fields of years are aggregated to the field of 'Year' so that the data can be joined with a tuple key ['country', 'Year'] as shown in figure 3.1.

	country	1960	1961	1962	1963	1964	1965	1966	1967	1968	...
0	Afghanistan	5.377778e+08	5.488889e+08	5.466667e+08	7.511112e+08	8.000000e+08	1.006667e+09	1.400000e+09	1.673333e+09	1.373333e+09	...
1	Albania	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...
2	Algeria	2.723593e+09	2.434727e+09	2.001428e+09	2.702960e+09	2.909293e+09	3.136259e+09	3.039835e+09	3.370843e+09	3.852116e+09	...
3	American Samoa	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...
4	Andorra	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...

(a) raw dataset of GDP

	country	Year	gdp
0	Afghanistan	1960	537777811.0
1	Afghanistan	1961	548888896.0
2	Afghanistan	1962	546666678.0
3	Afghanistan	1963	751111191.0
4	Afghanistan	1964	800000044.0

(b) converted dataset of GDP

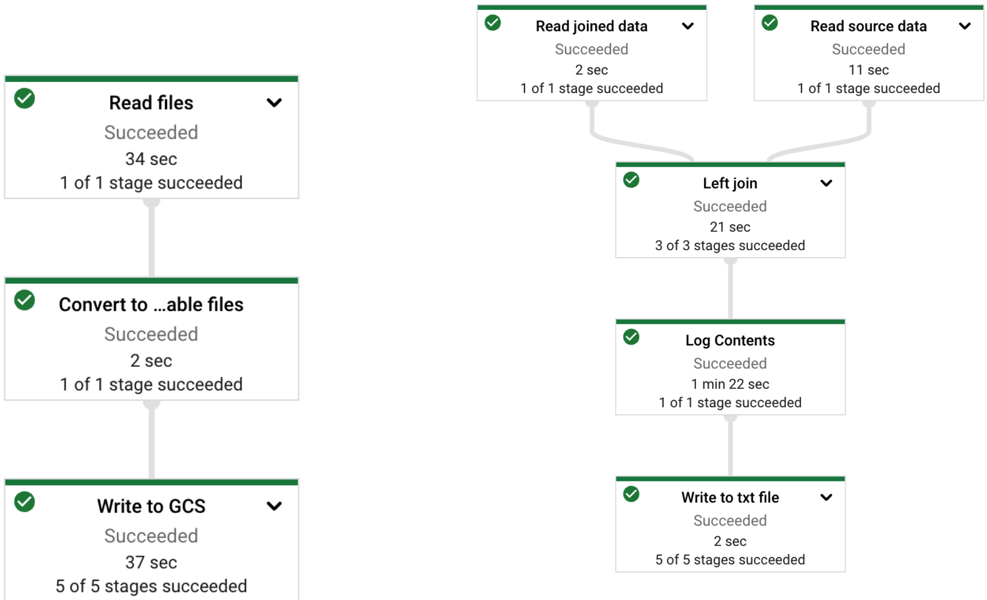
Figure 3.1 (a) The raw sample data of GDP; (b) the converted sample data.

The missing values are filled with backward shifts while the columns containing missing values more than 50% are removed because they cannot provide sufficient information. At last, outliers and abnormal data are dropped.

The data transformation is carried out by joining multiple datasets. There are 230 csv files about Global land surface temperature data by the name of each country that are concatenated as the temperature data in figure 3.2. The country information is made up of CO2 emission, GDP, population, and country area. The temperature data and the country data are transformed and left joined by the Google Cloud Platform Apache Beam. Figure 3.3 is the dataflow of temperature data concatenation (left), and joined to the entire dataset (right). The sample of the final dataset is shown in figure 3.4.

	country	Year	Annual_CO2_emissions	gdp	population	latitude	longitude	Area (sq. mi.)
0	Afghanistan	1750	0	NaN	NaN	33.93911	67.709953	647500.0
1	Afghanistan	1751	0	NaN	NaN	33.93911	67.709953	647500.0
2	Afghanistan	1752	0	NaN	NaN	33.93911	67.709953	647500.0
3	Afghanistan	1753	0	NaN	NaN	33.93911	67.709953	647500.0
4	Afghanistan	1754	0	NaN	NaN	33.93911	67.709953	647500.0

Figure 3.2 The joined dataset of country information



(a) (b)

Figure 3.3 Google Cloud Platform Dataflow (a) combination of temperature data; (b) joining country datasets

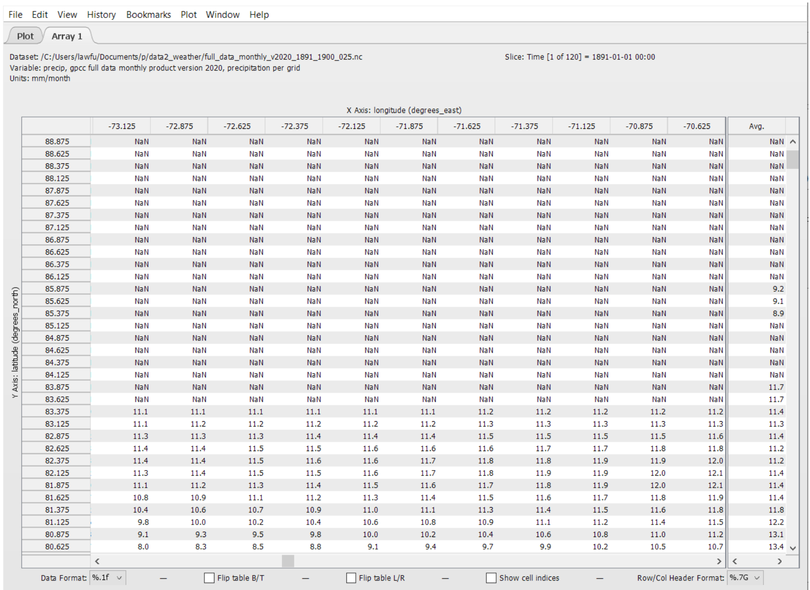
	time	Year	Month	country	countryid	latitude	longitude	population	Area (sq. mi.)	gdp	Annual_CO2_emissions	Monthly	Annual	Five-year	Ten-year
0	2017-01-01	2017	1	Afghanistan	1	33.939110	67.709953	36296111.0	647500.0	1.875347e+10	6859825.0	15.112	15.470	15.451	NaN
1	2017-02-01	2017	2	Afghanistan	1	33.939110	67.709953	36296111.0	647500.0	1.875347e+10	6859825.0	13.969	15.473	15.451	NaN
2	2017-03-01	2017	3	Afghanistan	1	33.939110	67.709953	36296111.0	647500.0	1.875347e+10	6859825.0	15.267	15.327	15.450	NaN
3	2017-04-01	2017	4	Afghanistan	1	33.939110	67.709953	36296111.0	647500.0	1.875347e+10	6859825.0	15.658	15.423	15.461	NaN
4	2017-05-01	2017	5	Afghanistan	1	33.939110	67.709953	36296111.0	647500.0	1.875347e+10	6859825.0	17.492	15.635	15.447	NaN
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...

Figure 3.4 The sample data of processed dataset

3.3 Data processing for country data and precipitation data

The precipitation data came in netCDF format, which includes multidimensional data (figure 3.5) and is typically used for geographic data. NASA's Panoply program (figure 3.6) is applied to open and view the raw netCDF files and their internal structure.

The raw precipitation data included eight sets of data, four of which also corresponded to sets of coordinates (interpolation error, gauge, precipitation, and time). We only needed the precipitation data, which referenced time, lat (latitude), and lon (longitude). When the precipitation field was viewed in Panoply, we could see the data is stored such that the precipitation values are laid out in a grid corresponding to a latitude and longitude on the globe. Areas without precipitation data collected for that year were filled with a placeholder value (NaN, “not a number”).


File Edit View History Bookmarks Plot Window Help

Plot | Array 1

Dataset: /C:/Users/lawfu/Documents/p/data2_weather/full_data_monthly_v2020_1891_1900_025.nc   Slice: Time [1 of 120] = 1891-01-01 00:00
Variable: precip, gpcc full data monthly product version 2020, precipitation per grid
Units: mm/month

X Axis: longitude (degrees_east); Y Axis: latitude (degrees_north)

	-73.125	-72.875	-72.625	-72.375	-72.125	-71.875	-71.625	-71.375	-71.125	-70.875	-70.625	Avg.
88.875	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
88.625	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
88.375	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
88.125	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
87.875	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
87.625	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
87.375	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
87.125	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
86.875	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
86.625	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
86.375	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
86.125	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
85.875	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	9.2
85.625	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	9.1
85.375	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	8.9
85.125	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
84.875	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
84.625	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
84.375	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
84.125	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
83.875	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	11.7
83.625	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	11.7
83.375	11.1	11.1	11.1	11.1	11.1	11.1	11.2	11.2	11.2	11.2	11.2	11.4
83.125	11.1	11.2	11.2	11.2	11.2	11.2	11.3	11.3	11.3	11.3	11.3	11.3
82.875	11.3	11.3	11.3	11.4	11.4	11.4	11.5	11.5	11.5	11.5	11.6	11.4
82.625	11.4	11.4	11.5	11.5	11.6	11.6	11.6	11.7	11.7	11.8	11.8	11.2
82.375	11.4	11.4	11.5	11.6	11.6	11.7	11.8	11.8	11.9	11.9	12.0	11.2
82.125	11.3	11.4	11.5	11.5	11.6	11.7	11.8	11.9	11.9	12.0	12.1	11.4
81.875	11.1	11.2	11.3	11.4	11.5	11.6	11.7	11.8	11.9	12.0	12.1	11.4
81.625	10.8	10.9	11.1	11.2	11.3	11.4	11.5	11.6	11.7	11.8	11.9	11.4
81.375	10.4	10.6	10.7	10.9	11.0	11.1	11.3	11.4	11.5	11.6	11.8	11.8
81.125	9.8	10.0	10.2	10.4	10.6	10.8	10.9	11.1	11.2	11.4	11.5	12.2
80.875	9.1	9.3	9.5	9.8	10.0	10.2	10.4	10.6	10.8	11.0	11.2	13.1
80.625	8.0	8.3	8.5	8.8	9.1	9.4	9.7	9.9	10.2	10.5	10.7	13.4

Data Format: %.1f — Flip table B/T — Flip table L/R — Show cell indices — Row/Col Header Format: %.7G

Figure 3.5 The sample data of precipitation

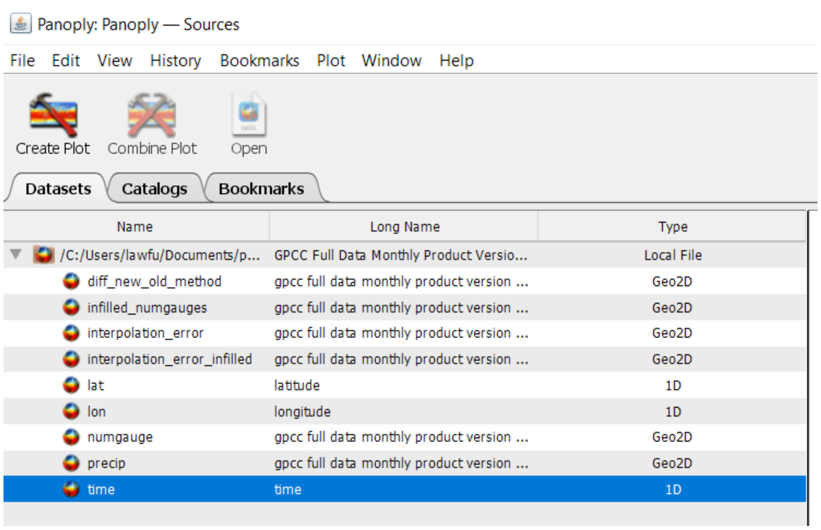


Figure 3.6 NASA’s Panoply program to view the data structure

In order to visualize and manipulate this data beyond what Panoply offers, we had to pull out each precipitation value and relate it to its year, latitude, and longitude. We wrote a Python script using netCDF4 to read from the netCDF file, and pandas and numpy to manipulate the data and serialized it into csv files (figure 3.7). We then loaded the data into Amazon S3 and Amazon Athena to combine it with country and city information based on latitude and longitude for further analysis (figure 3.8).

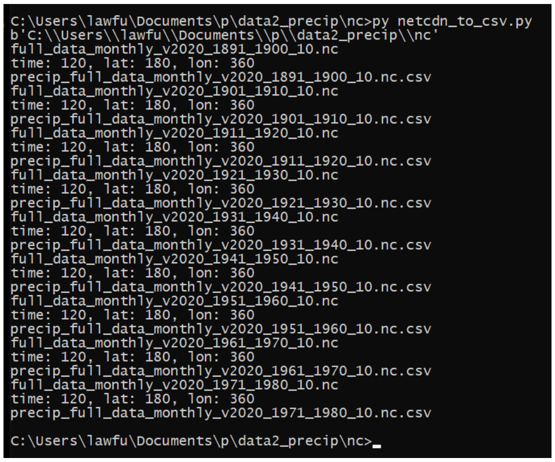


Figure 3.7 Extract data from netCDF to csv files

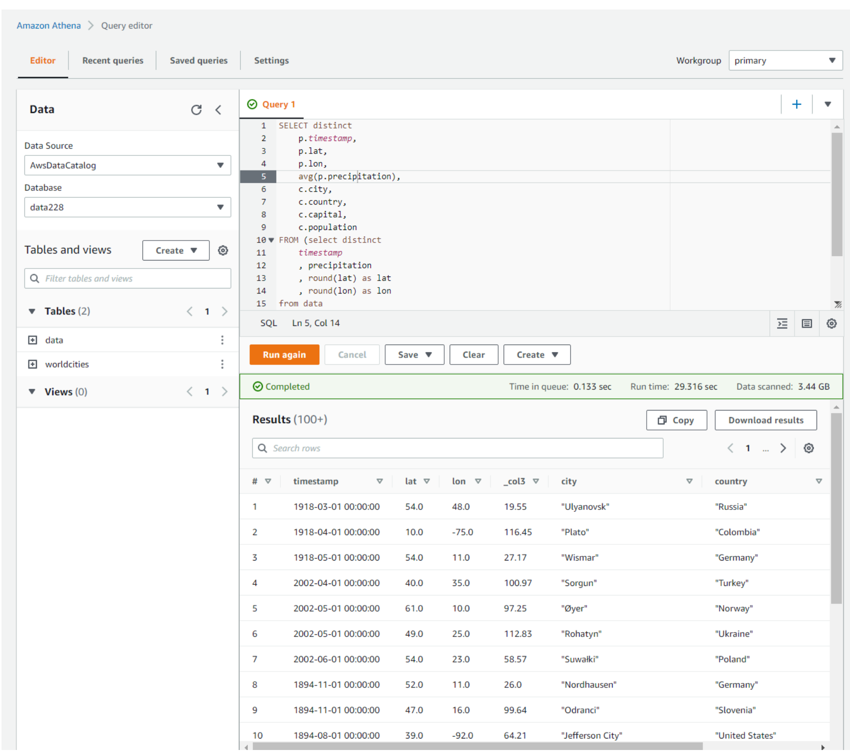


Figure 3.8 Combining precipitation data with country information

3.4 Data exploration and visualization

We visualized temperature and precipitation data in Amazon Quicksight, and discovered that precipitation levels do not correlate as highly with ground surface temperatures as do carbon dioxide emissions and land surface temperatures, so we remove it in the machine learning step.

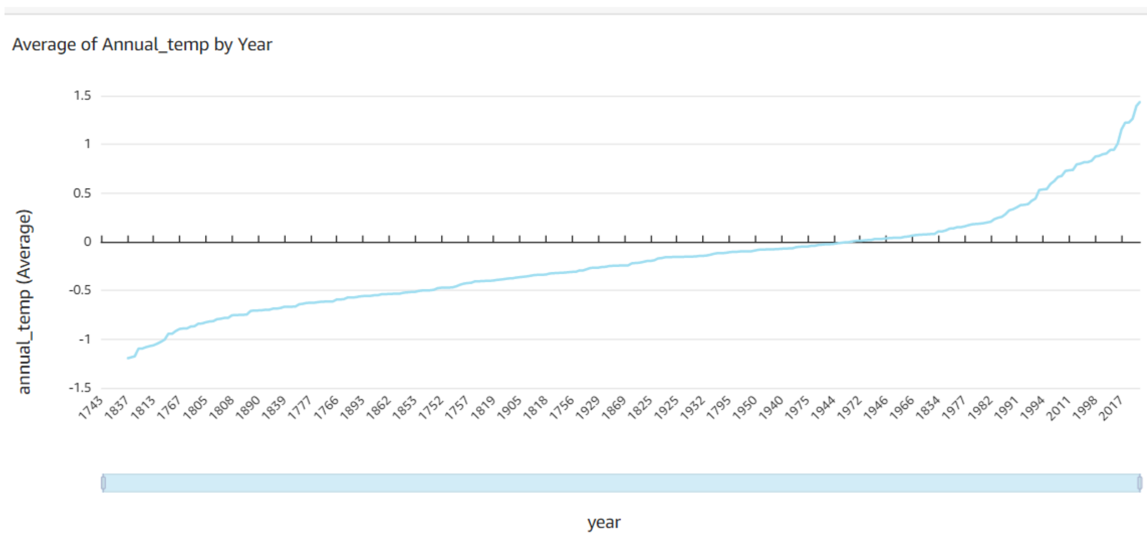


Figure 3.9 Average year-over-year change in annual temperature over time

Figure 3.9 graphs the average annual temperature over time globally. It shows a steady increase in average temperature over time. After around the 1950s, the year-over-year change in average temperature grows positive, pointing towards a rapidly heating globe.

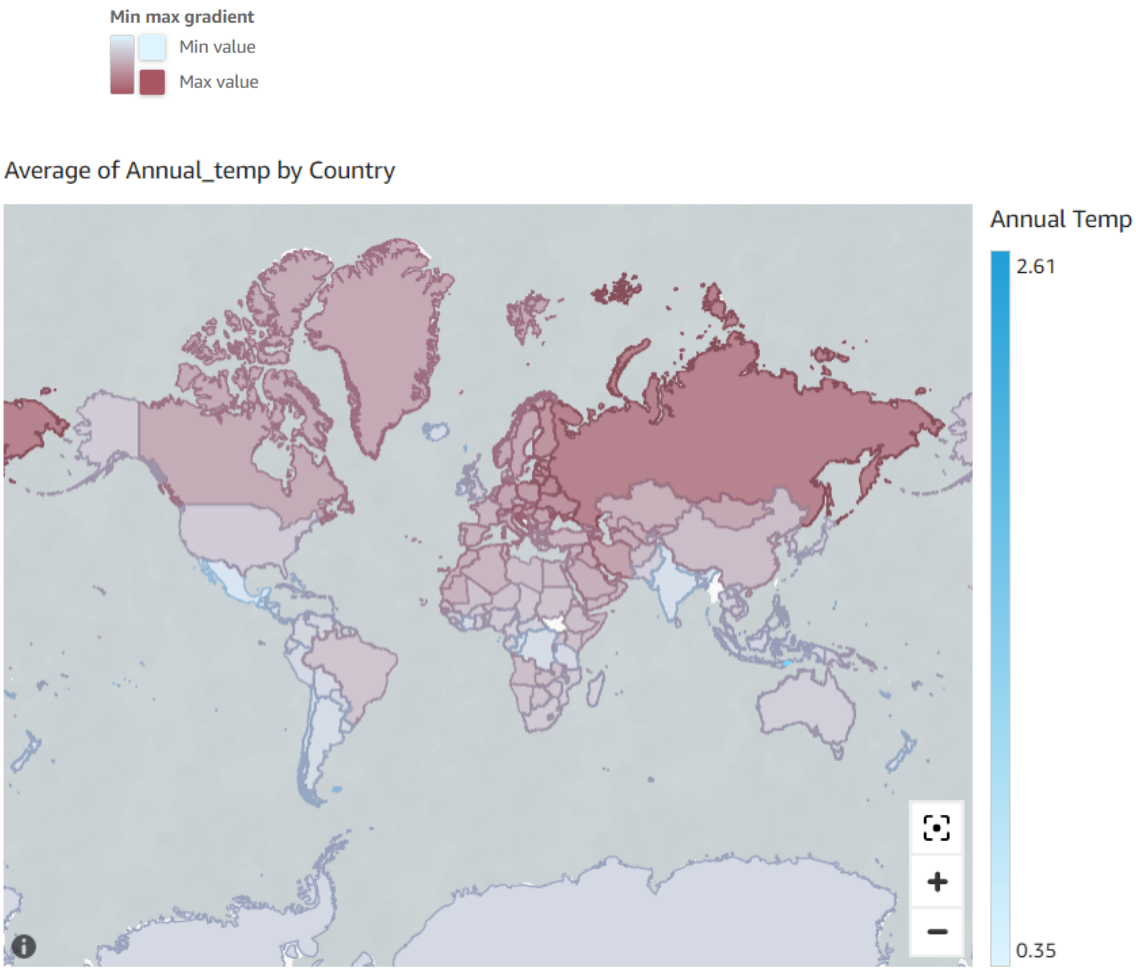


Figure 3.10 Heat map of average annual global temperature by country

Fig 3.10 shows a geographic heat map of countries across the world. The dark maroon color shows a high change in annual temperature, while light blue means a lesser rate of change in annual temperature. We can see from this map that the greatest changes in temperature occur at the northern latitudes, which are heating much faster than the southern latitudes.

Fig 3.11 displays the total carbon dioxide emissions for each country, each year. Later years are displayed on the left and earlier years on the right due to the limitations of Amazon Quicksight. We can see from the chart that carbon dioxide emissions grow rapidly after around 1945 across all groups, and mostly continue to climb ever since. Some countries have curbed their emissions slightly since the 2000s, but the overall amount of carbon dioxide emissions emitted globally continues to climb..

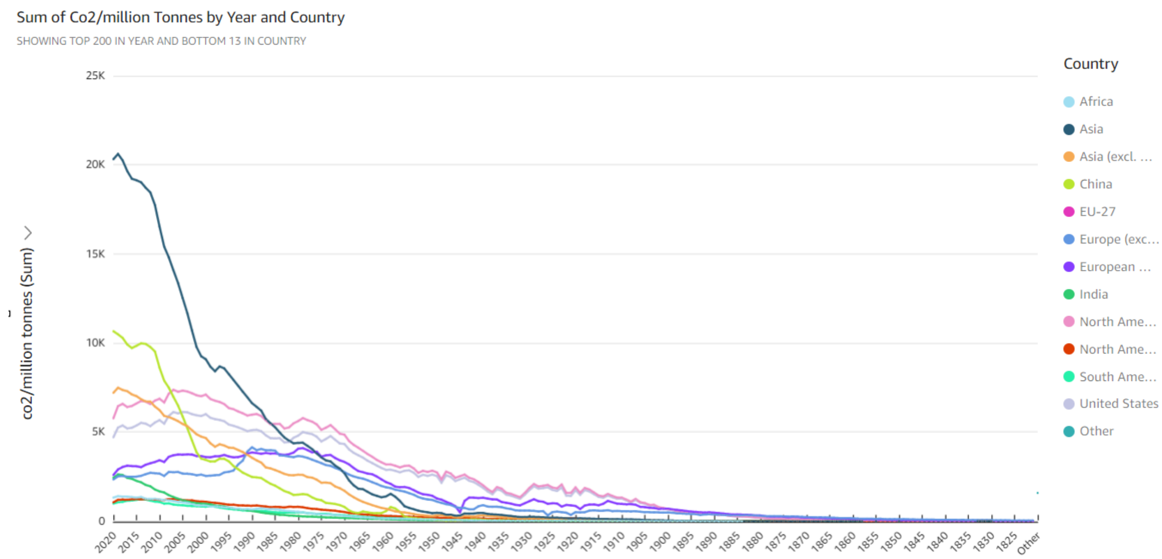


Figure 3.11 Total carbon dioxide emissions by year and country

The visualization of precipitation data is shown in the following figures by AWS Quicksight. Figure 3.12 and figure 3.13 are the overall average temperature on a map which is consistent with the map produced earlier in Panoply from the raw NetCDF data. The southern areas show more precipitation than the northern areas with dark blue marks. The trends are both shown in the visualized results from the converted data (figure 3.12) and the raw data (figure 3.13).

Splitting the precipitation dataset by latitude, we can see how average precipitation at each latitude has changed over time. The trends in figure 3.14 and figure 3.15 demonstrate the vibration of precipitation in the recent years with increased amplitude. That can indicate the effect of rising temperature on global precipitation.

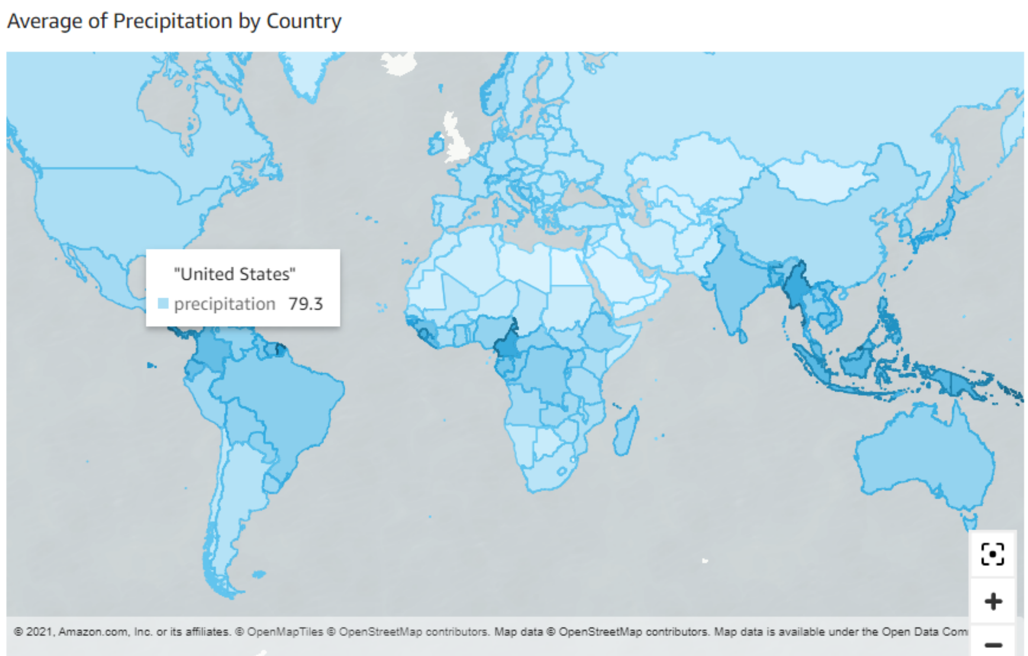


Figure 3.12 Visualization of precipitation from AWS Quicksight

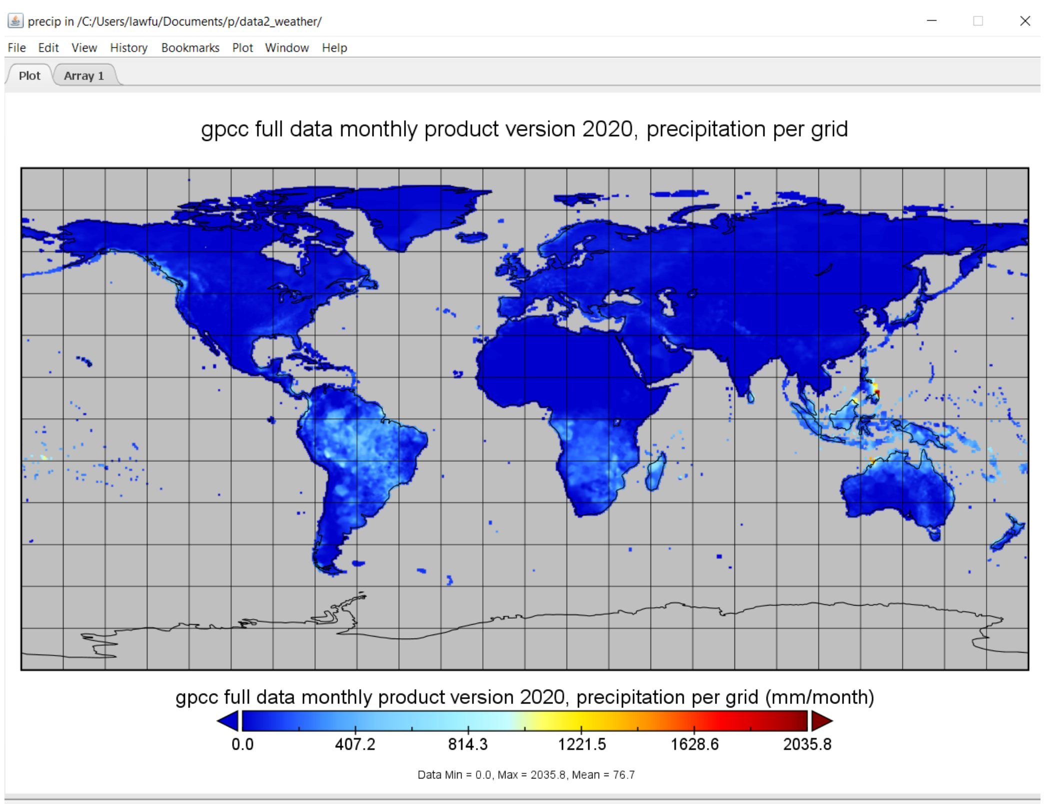


Figure 3.13 Visualization of precipitation from NASA Panoply

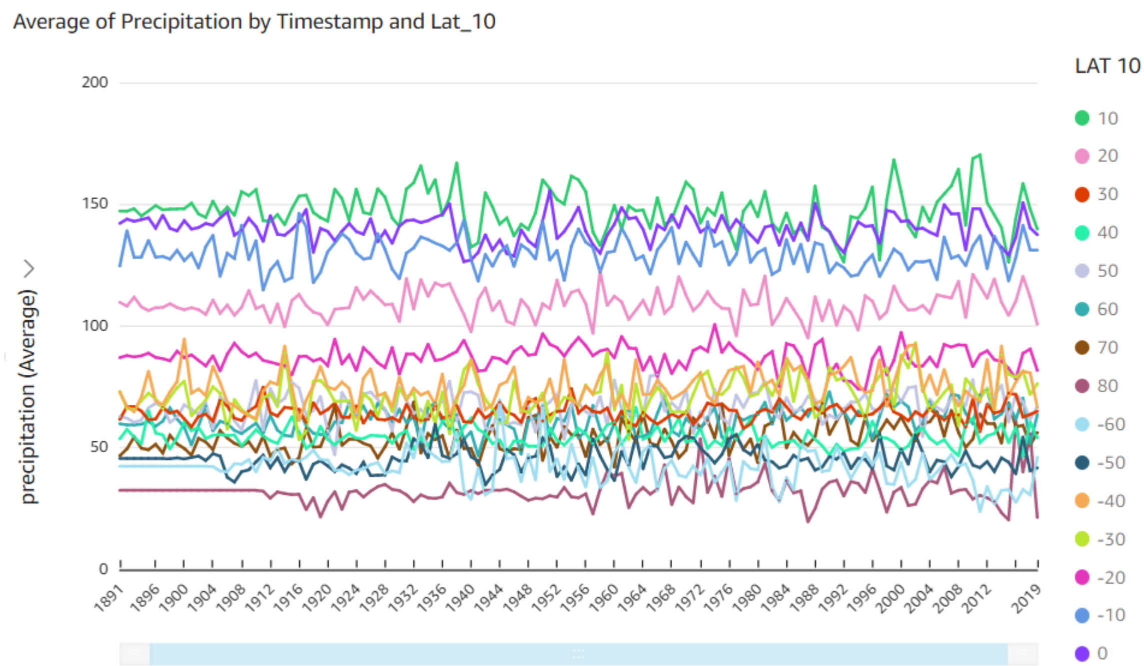


Figure 3.14 Average precipitation over time by latitude

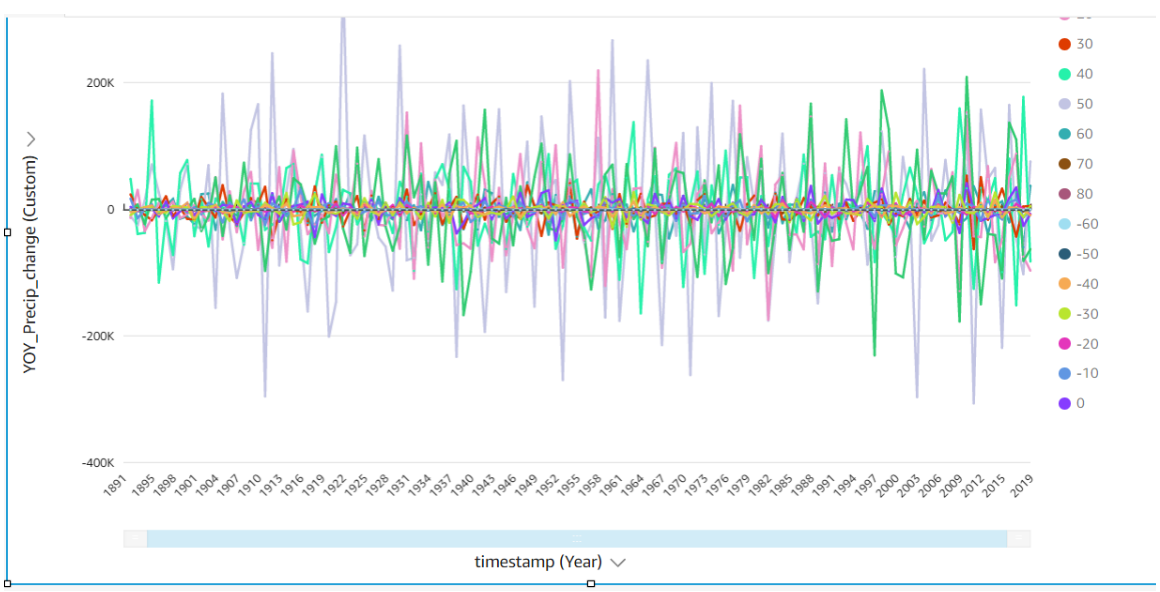


Figure 3.15 Year-over-year change in precipitation over time by latitude

4 Machine learning to predict temperature

This project aims to forecast the land surface temperature by countries. Human activities have released huge amounts of greenhouse gasses and contributed to global warming. The historical data of geographic and demographic information, the gross domestic product, annual CO_2 emission are considered for temperature prediction from different resources. The target feature is the monthly temperature in each country from U.S. non-profit organization Berkeley Earth.

The processed data have 13 fields as listed in figure 4.1. The data types are numerical except 'country' as country name. The feature 'countryid' is created to represent the country name from 1 to 234. The features of 'Monthly', 'Annual', 'Five-year', and 'Ten-year' are the average land surface temperature selected from observation data. The features of 'latitude', 'longitude', 'Area (sq. mi.)' are constant in each country respectively, and the features of 'gdp', 'Annual_CO2_emissions', and 'population' are changing over time.

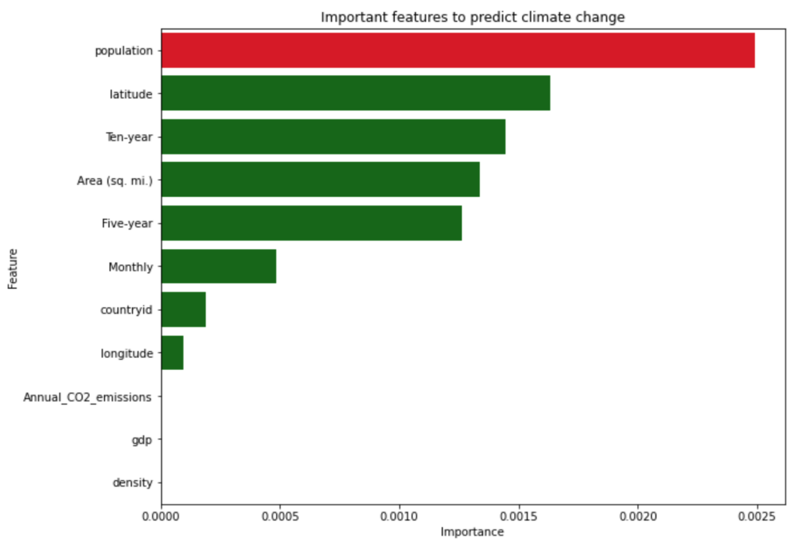
```
DatetimeIndex: 524505 entries, 1848-05-01 to 2020-12-01
Data columns (total 14 columns):
 #   Column                Non-Null Count   Dtype
---  ------                --------------   -----
 0   Year                  524505 non-null  int64
 1   Month                 524505 non-null  int64
 2   country               524505 non-null  object
 3   countryid             524505 non-null  int64
 4   latitude              435727 non-null  float64
 5   longitude             435727 non-null  float64
 6   population            403725 non-null  float64
 7   Area (sq. mi.)        423838 non-null  float64
 8   gdp                   398998 non-null  float64
 9   Annual_CO2_emissions  438480 non-null  float64
 10  Monthly               498290 non-null  float64
 11  Annual                493150 non-null  float64
 12  Five-year             479096 non-null  float64
 13  Ten-year              467106 non-null  float64
dtypes: float64(10), int64(3), object(1)
```

Figure 4.1 The summary of data fields.

4.1 Feature engineering

The feature importance is evaluated by the Random Forest Regression method. The scores of the feature significance are generated and shown in ranking in figure 4.1. Since the features of ‘Annual’, ‘Five-year’, and ‘Ten-year’ are linear to ‘Monthly’, it has demonstrated a higher score. The population, location, and area demonstrate an important impact on the temperature while the GDP and population/area have the least influence. Therefore, the features of ‘latitude’, ‘longitude’, ‘Area (sq. mi.)’ and ‘populations’ are treated as part of input data.

	Feature	Importance
0	Annual	9.910668e-01
1	population	2.492058e-03
2	latitude	1.633316e-03
3	Ten-year	1.444151e-03
4	Area (sq. mi.)	1.336974e-03
5	Five-year	1.259992e-03
6	Monthly	4.817932e-04
7	countryid	1.894761e-04
8	longitude	9.541575e-05
9	Annual_CO2_emissions	1.168016e-09
10	gdp	7.246401e-10
11	density	1.403420e-10



(a) (b)

Figure 4.2 (a) feature score (b) ranking of feature importances

4.2 Machine learning model

Machine learning is the technique to learn the patterns from big data and make predictions and classifications on the unseen data. It has been widely used in time-series forecasting. In this project, long short-term memory (LSTM) is designed to train the historical data.

The neural network cannot accept the negative values. The data are prepared by standardizing into (0, 1) , converting longitude and latitude into x, y, and z coordinates, transforming data into time-series with 1 lag backward shift. The statistics of finalized input data is shown in figure 4.3.

	countryid	population	Area (sq. mi.)	x	y	z	Annual	Monthly
count	330827.000000	330827.000000	3.308270e+05	330827.000000	330827.000000	330827.000000	330827.000000	330827.000000
mean	0.495946	0.002218	3.100132e-03	0.510568	0.443441	0.512518	0.672239	0.672240
std	0.288563	0.008244	7.942838e-03	0.252343	0.245094	0.350408	0.135340	0.136199
min	0.004274	0.000001	7.768868e-07	0.003897	0.015841	0.001294	0.108814	0.014586
25%	0.247863	0.000118	1.459735e-04	0.349804	0.235511	0.162401	0.562986	0.565657
50%	0.500000	0.000461	6.016049e-04	0.504356	0.495257	0.517109	0.710329	0.711171
75%	0.743590	0.001250	2.284778e-03	0.690049	0.592647	0.877759	0.794314	0.794014
max	0.978632	0.159828	5.069907e-02	0.995303	0.985343	0.999909	0.857386	0.881771

Figure 4.3 The statistics of normalized datasets

The model has 72 batches, 1 hidden layer, and 50 epoches. The callback is set up with an early stopping that the model will terminate as the loss doesn't decrease. As a result, the loss function is decreasing as the model is optimized, as seen in figure 4.4. The model achieved a high performance with 0.0195 room mean squared error, and 0.0107 mean absolute percentage error. The comparison of prediction vs observation data is shown in figure 4.5.

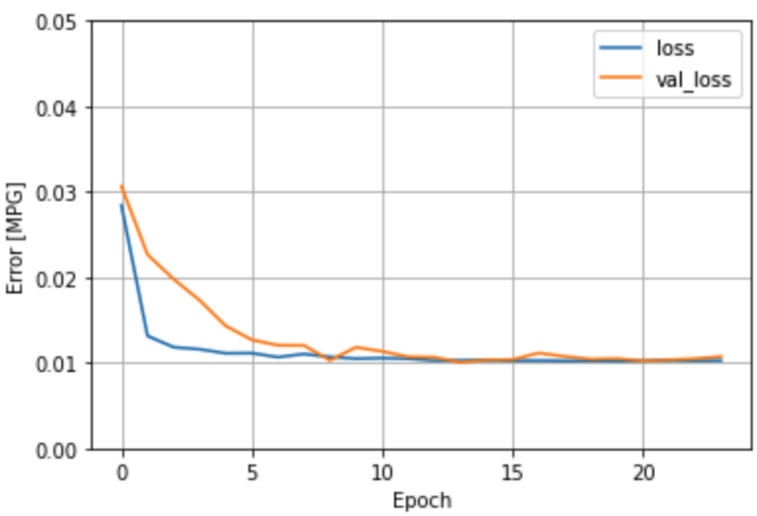


Figure 4.4 The loss function over the epochs

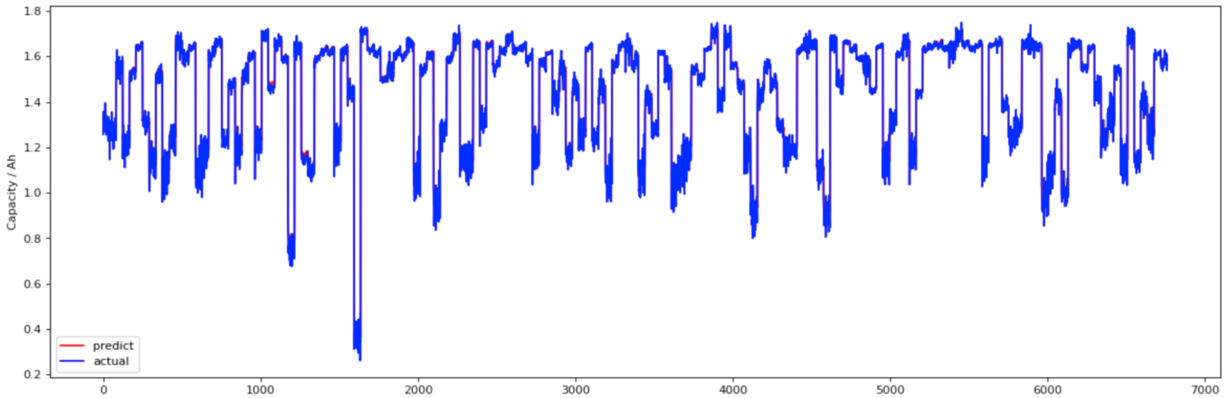


Figure 4.5 Comparison between the prediction results and the observation data

The model can make predictions for each country from 2016 to 2021 using this historical data from 1800 to 2015. Figure 4.6 is the prediction results with the historical data for Cuba. The temperature is rising slowly. The model can detect the temperature range by geographic and demographic information. The figure 4.7 and 4.8 shows the differences of prediction results in temperature range and the rising slope as well.

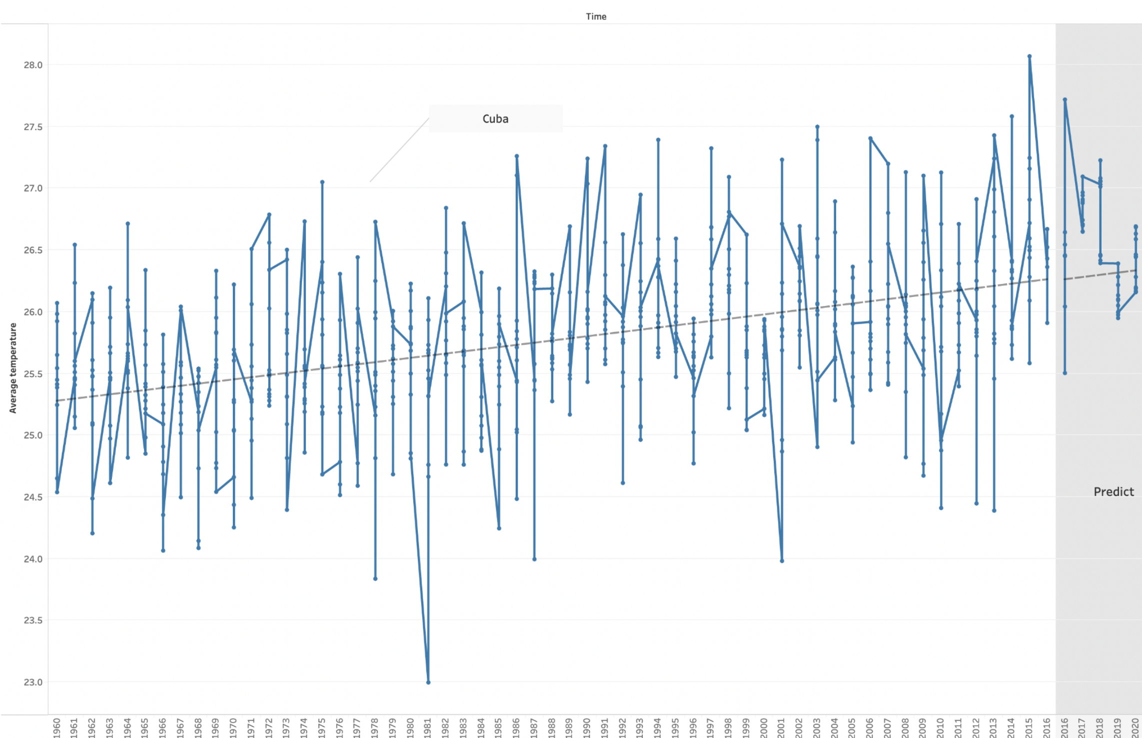


Figure 4.6 The rising temperature in Cuba

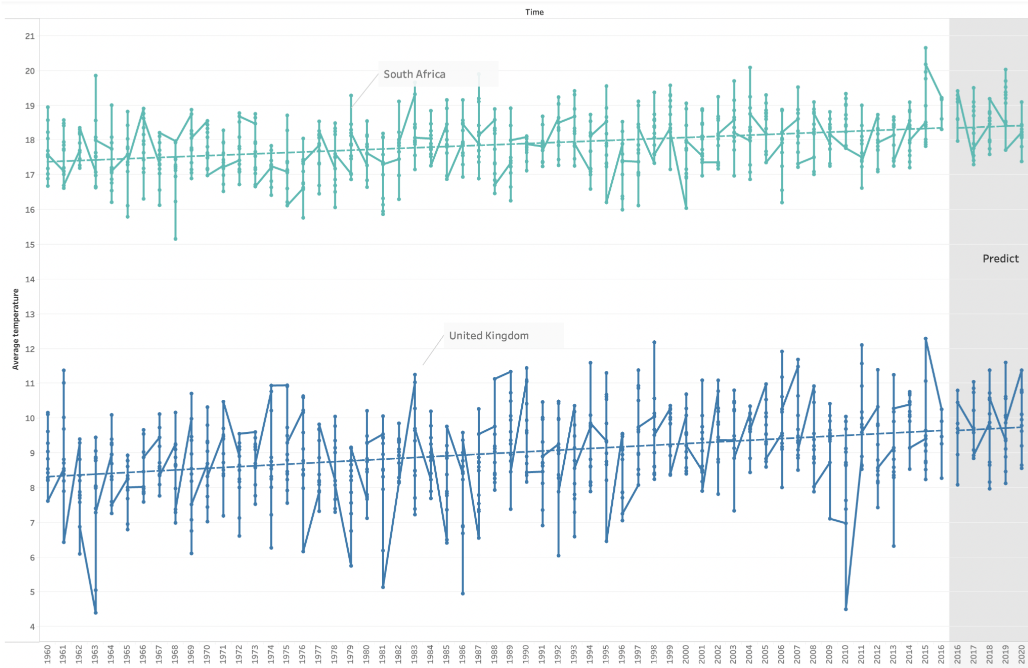


Figure 4.7 The rising temperature in South Africa and United Kingdom

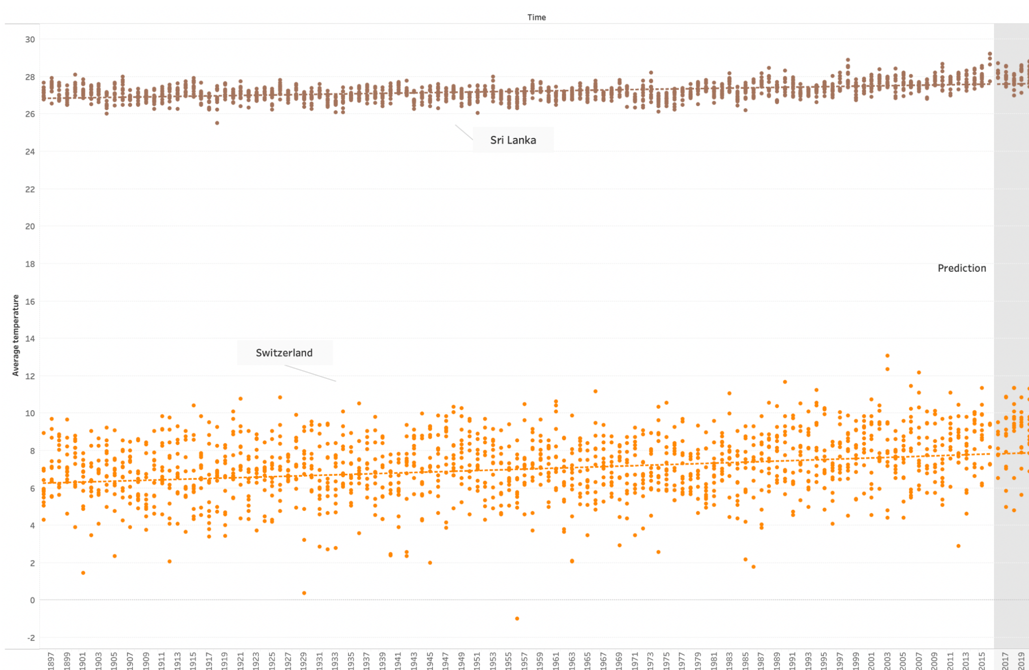


Figure 4.8 The rising temperature in SriLanka and Switzerland

Overall, this model can predict the future temperature once the geographic and demographic information are available.

5 Web-based Application

We used the web framework Flask to build an interactive website to display the results of our model. The website displays a graph of average annual temperature for a selected decade and country, shown in Fig 5.1. A list of countries and decades are shown at the top of the page and can be selected to refresh the data in the graph, allowing users to easily see changes in data. When a country or decade is selected, the flask app sends the new data and the page is refreshed.

The data is split by country and by decade to give the user an intuitive view of the scope of the data, and gives a time scale that is neither too large nor too small to see changes in temperature over time. Keeping the data set small also decreases the amount of time it takes to display results when a new set of data is queried.

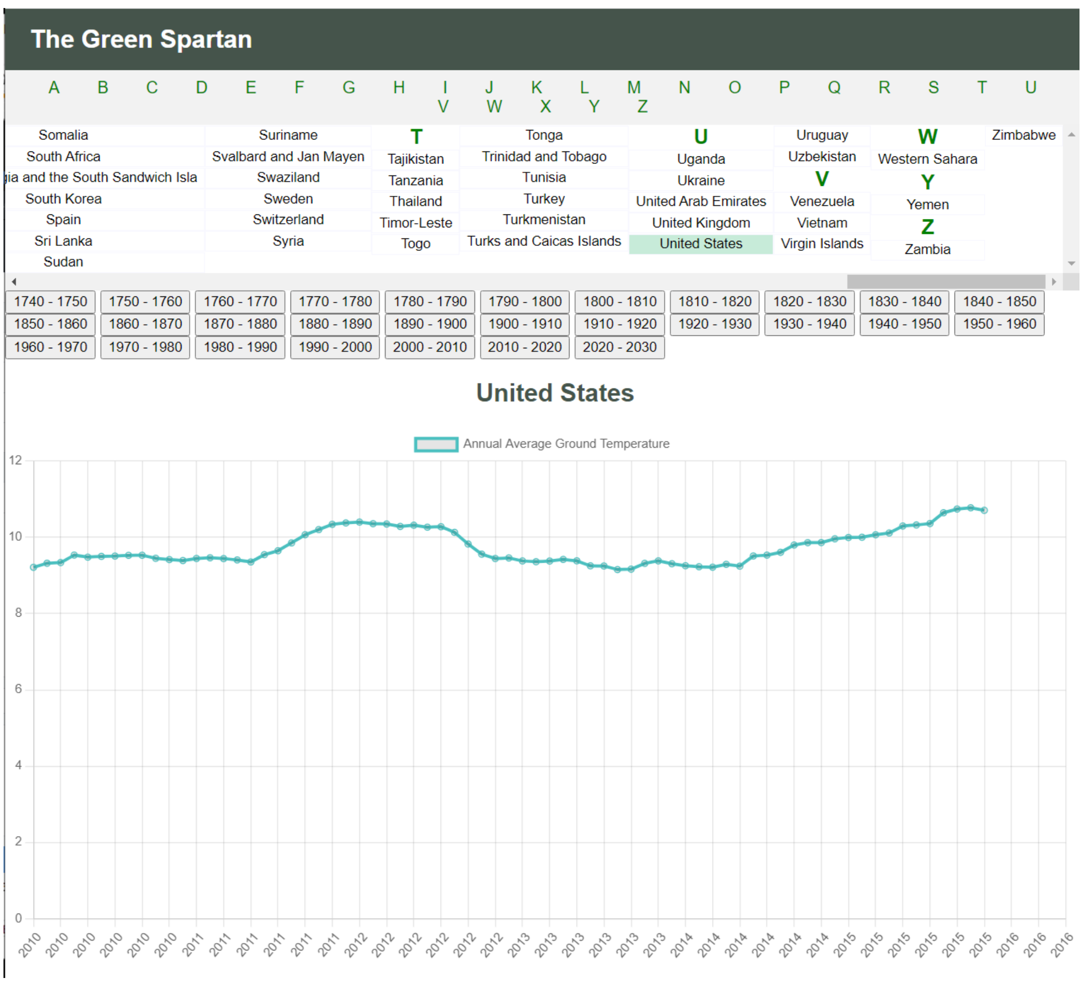


Figure 5.1 The Green Spartan API

The list of countries is generated from the available countries, and the decades from the available years, in the data. We used Flask's templating engine, Jinja, to create jumps in the country list by inserting letters when the leading letter of the country changes, as shown in Fig 5.2.

```
{% extends 'index.html' %}

{% block countrySelect %}
  {% set ns = namespace(letter='') %}
  {% for c in countries %}
    {% if ns.letter|string() != c[:1]|string() %}
      {% set ns.letter = c[:1] %}
      <div class="alphabet-header">
        {{ ns.letter }}
        <div class="alphabet-dest" id="{{ ns.letter }}"></div>
      </div>
    {% endif %}
    <div class='country-button {% if country|string() == c|string() %}selected{% endif %}'>
      <form method="post">
        <button type="submit" name='country' value ="{{ c }}">{{ c }}</button>
      </form>
    </div>
  {% endfor %}
</div>
{% endblock %}
```

Figure 5.2 Client interface design using the Jinja templating language

The website is hosted in Google Cloud App Engine through cloud CLI. The deployment is configured by using the flexible environment with 2 instance virtual machines. The virtual machine configuration takes 3-4 minutes and the deployment for 15 minutes. The website can run without latency and match the correctness expectation.

Filter Filter versions

Status	Traffic Allocation	Instances	Runtime	Environment
Serving	100%	2	python	Flexible
Stopped	0%	0	python	Flexible

Figure 5.3 The deployment of web API on Google Cloud Platform App Engine

6 Conclusion

In this project, the global land surface temperature is analyzed . We collected data from multiple sources and built the data pipeline for data transformation. The global land surface temperature data and the resultant precipitation data are visualized to analyze the effect of CO2 emission, geographic and demographic data on climate change. The long short-term memory model is designed for temperature prediction. The results demonstrate the high prediction accuracy. We built a website for users to look up historical data and the prediction data.

References

[1] United Nations, “Climate change is a ‘global emergency’, people say in biggest ever climate poll” UN News, 2021 [Online]. Available: https://news.un.org/en/story/2021/01/1083062

[2] H. Shaftel, S. Callery, R. Jackson, D. Bailey, S. Callery, “Scientific Consensus: Earth's Climate Is Warming” NASA Global Climate Change: Vital Signs of the Planet, 2022 [Online]. Available: https://climate.nasa.gov/scientific-consensus/

[3] “What is the United Nations Framework Convention on Climate Change?” United Nations Climate Change [Online]. Available: https://unfccc.int/process-and-meetings/the-convention/what-is-the-united-nations-framework-convention-on-climate-change

[4] J Mann, “Birds are falling from the sky in India as a record heatwave dries up water sources” Business Insider Africa 2022 [Online]. Available: https://africa.businessinsider.com/science/birds-are-falling-from-the-sky-in-india-as-a-record-heatwave-dries-up-water-sources/6zerw12#:~:text=BI%20Africa%20%3E%20science-,Birds%20are%20falling%20from%20the%20sky%20in%20India%20as,heatwave%20dries%20up%20water%20sources&text=A%20caretaker%20feeds%20water%20mixed,heatwave%20dries%20up%20water%20sources.

[5] https://www.vice.com/en/article/qjbyk5/birds-heat-stroke-deadly-heat-wave-india

[1] Google Developers. “Countries” Google Data Publishing Language, 2015 [Online]. Available: https://developers.google.com/public-data/docs/canonical/countries_csv

[2] F. Lasso. ”Countries of the World” Kaggle, 2018 [Online]. Available: https://www.kaggle.com/datasets/fernandol/countries-of-the-world

[3] U. Schneider, A. Becker, P. Finger, E. Rustemeier, and M. Ziese, “GPCC Full Data Monthly Version 2020 at 1.0°: Monthly Land-Surface Precipitation from Rain-Gauges built on GTS-based and Historic Data.” Global Precipitation Climatology Centre (GPCC) at Deutscher Wetterdienst, 2020 [Online]. doi: 10.5676/DWD_GPCC/FD_M_V2020_100. Available: https://opendata.dwd.de/climate_environment/GPCC/html/fulldata-monthly_v2020_doi_download.html

[4] “2021 Temperature Report: List of Countries” Berkeley Earth, 2021 [Online].
http://berkeleyearth.lbl.gov/country-list/