Global Temperature Data Analysis Report (last 250 years)

Iranian Cheetah one of the species in danger of extinction due to Global Warming

In this project I have chosen to work on global warming dataset since it is the biggest threat to life on earth. The data set is downloaded from Kaggle and can be accessed through this link.

To explain why I have chosen this dataset I have to say there are many reasons why it is important to care about global warming and take action to address it. Some of the key reasons include:

Human health: Rising global average temperatures can have negative impacts on human health, including increasing the risk of heat-related illness and death, particularly among vulnerable populations.

Environmental impacts: Global warming can cause a range of environmental impacts, including melting polar ice caps, rising sea levels, and changes in ecosystems. These impacts can have cascading effects on the planet and its inhabitants.

Economic impacts: Global warming can have negative impacts on economic activity, including reduced crop yields, damaged infrastructure, and other impacts, particularly in sectors such as agriculture and tourism.

Social justice: The impacts of global warming are likely to disproportionately affect vulnerable populations and communities, including those that are already marginalized or disadvantaged.

Overall, addressing global warming is important for the well-being of current and future generations, as well as for the health and stability of the planet.

In this dataset we have several CSV files including:

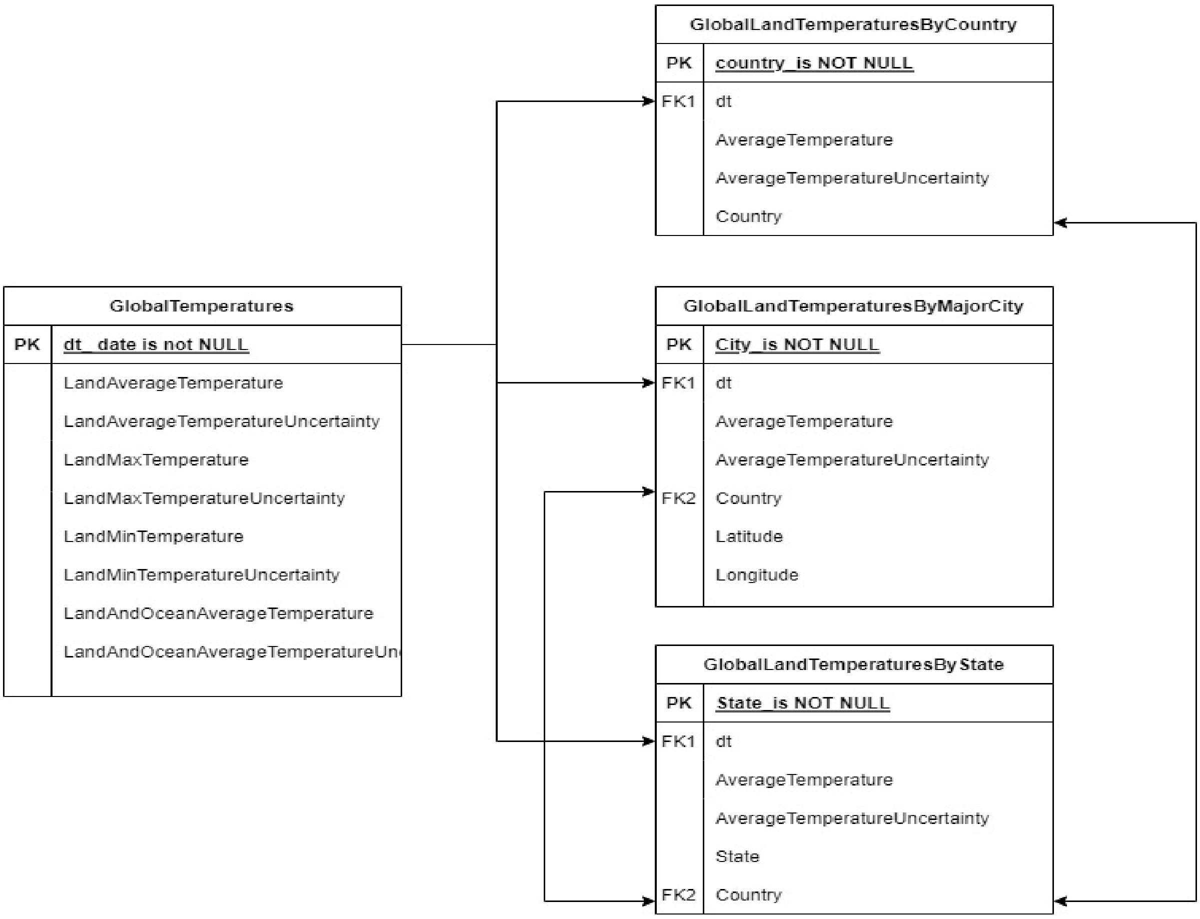
- Global Average Land Temperature by Country (GlobalLandTemperaturesByCountry.csv)
- Global Average Land Temperature by State (GlobalLandTemperaturesByState.csv)
- Global Land Temperatures By Major City (GlobalLandTemperaturesByMajorCity.csv)
- Global Land Temperatures By City (GlobalLandTemperaturesByCity.csv)
- Global Land and Ocean-and-Land Temperatures (GlobalTemperatures.csv)

I am going to ignore one file which is “City” since the file for that data is huge and other files can give me the data that I need to work.

We are trying to use these data to find out if there is any meaningful relationship between the files and finally trying to make an AI agent that tries to predict the trend of climate change and hopefully the results can be used to stop the pace of increasing the global temperature and save lives of different species including us, The homo sapiens Humans.

In Global and Country database we have data from 1750 and for States and Land we have data from 1850 to 2015 which is a long history, and it is important to be that long since temperatures change over time slightly.

To get a better understanding of the datasets I have drawn the ERD diagram which is a graphical representation of the entities and relationships within a database and it looks like this:



It is shown in above diagram that some of the columns of the datasets are marked by PK or FK which stands for Primary key (In a database, a primary key is a field (or set of fields) that uniquely identifies each row in a table. It is used to ensure the integrity of the data and to establish relationships between tables. A primary key cannot contain null values and must be unique across all rows in the table) and Foreign Key (A foreign key is a field (or set of fields) in a table that references the primary key of another table. It is used to establish a relationship between the two tables. A foreign key value must match the value of a primary key in the referenced table, or be null) and these keys can be listed just like this:

Dataset	Primary Key	Foreign key
Global Temperatures	dt	
Global Land Temperatures by country	country	dt
Global Land Temperature by City	City	dt, country
Global Land Temperature by State	State	dt, country

Here you can see the description of each dataset using the describe function:

GlobalTemperatures.csv

index	Land Average Temperature	Land Average Temperature Uncertainty	Land Max Temperature	Land Max Temperature Uncertainty	Land Min Temperature	Land Min Temperature Uncertainty	Land And Ocean Average Temperature	Land And Ocean Average Temperature Uncertainty
count	3180	3180	1992	1992	1992	1992	1992	1992
mean	8.374731132	0.938467925	14.3506	0.479782	2.743595	0.431849	15.21257	0.128532
std	4.381309771	1.096439795	4.309579	0.583203	4.155835	0.445838	1.274093	0.073587
min	-2.08	0.034	5.9	0.044	-5.407	0.045	12.475	0.042
25%	4.312	0.18675	10.212	0.142	-1.3345	0.155	14.047	0.063
50%	8.6105	0.392	14.76	0.252	2.9495	0.279	15.251	0.122
75%	12.54825	1.41925	18.4515	0.539	6.77875	0.45825	16.39625	0.151
max	19.021	7.88	21.32	4.373	9.715	3.498	17.611	0.457

It shows that the maximum Land Average Temperature of earth was 19.021 between the years of 1750 to 2015 and the minimum was -2.08 and standard deviation for that was 4.38.

The other important information that it is giving us is that the average temperature of Land and Ocean was like these Max: 17.61, Min: 12.47 and standard deviation was like 1.27.

Global Average Land Temperature by Country

index	Average Temperature	Average Temperature Uncertainty
count	544811	545550
mean	17.19335	1.0190569
std	10.95397	1.201930387
min	-37.658	0.052
25%	10.025	0.323
50%	20.901	0.571
75%	25.814	1.206
max	38.842	15.003

This information of the Countries datasets tell us that we have experienced the Max Average Temperature 38.84 and Min Temperature of -37.65 and standard deviation was like 10.95 among the countries during 1743 to 2015.

Global Land Temperatures By Major City

index	Average Temperature	Average Temperature Uncertainty
count	228175	228175
mean	18.12597	0.96934344
std	10.0248	0.97964391
min	-26.772	0.04
25%	12.71	0.34
50%	20.428	0.592
75%	25.918	1.32
max	38.283	14.037

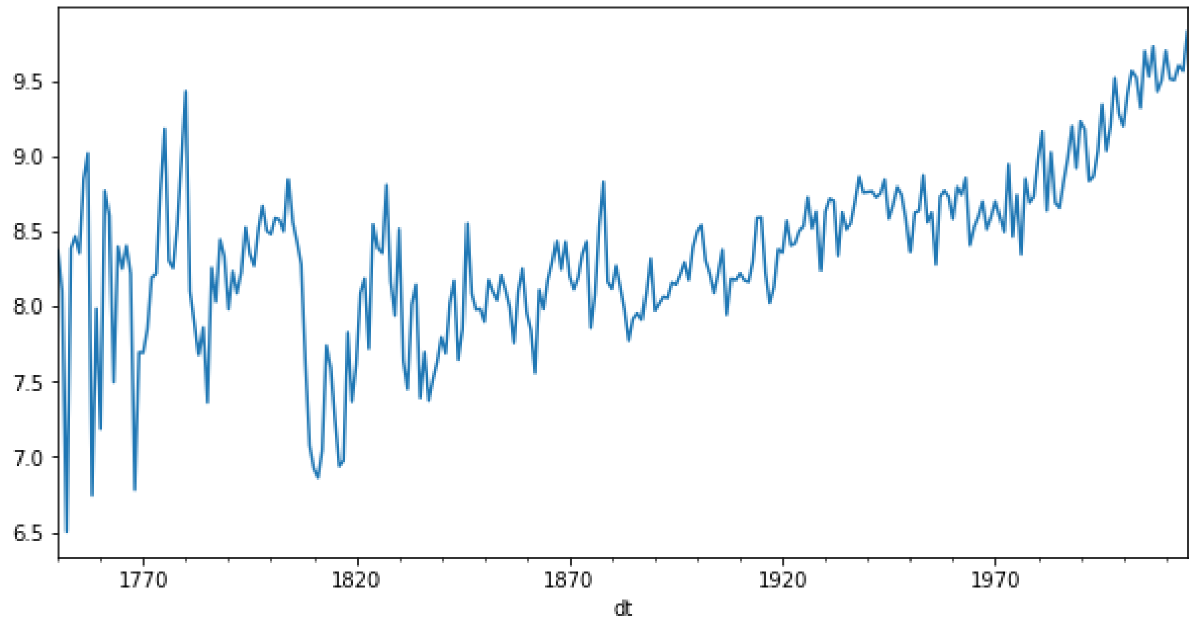
In major cities we have experienced the Max Average Temperature of 38.28 and Min Average Temperature of -26.77 and standard deviation of 10.02 during the 1850 to 2015.

Global Average Land Temperature by State

index	Average Temperature	Average Temperature Uncertainty
count	620027	620027
mean	8.99311078	1.28764715
std	13.7721503	1.36039206
min	-45.389	0.036
25%	-0.693	0.316
50%	11.199	0.656
75%	19.899	1.85
max	36.339	12.646

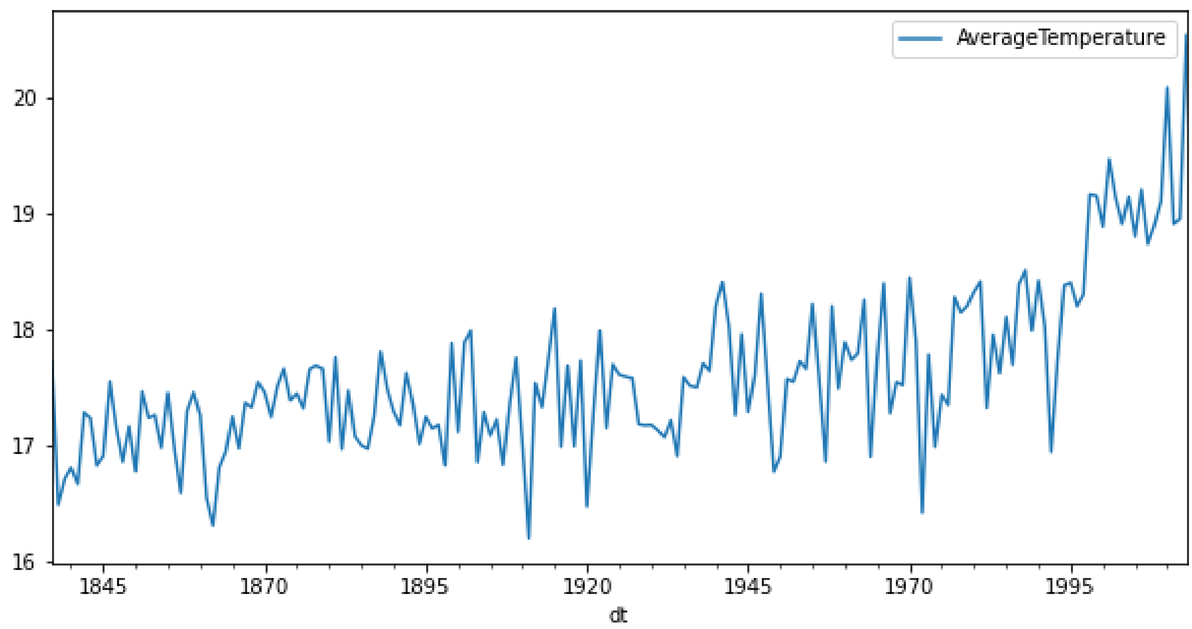
In states we have experienced the Max Average Temperature of 36.33 and Min Average Temperature of -45.38 and standard deviation of 13.77 during the 1850 to 2015 among the states of different countries.

After having a look at the statistics of our datasets I have decided to show some of the plotting based on the data to have a better understanding of how the temperature is changing during the last three decades and how it is threatening the life on earth.



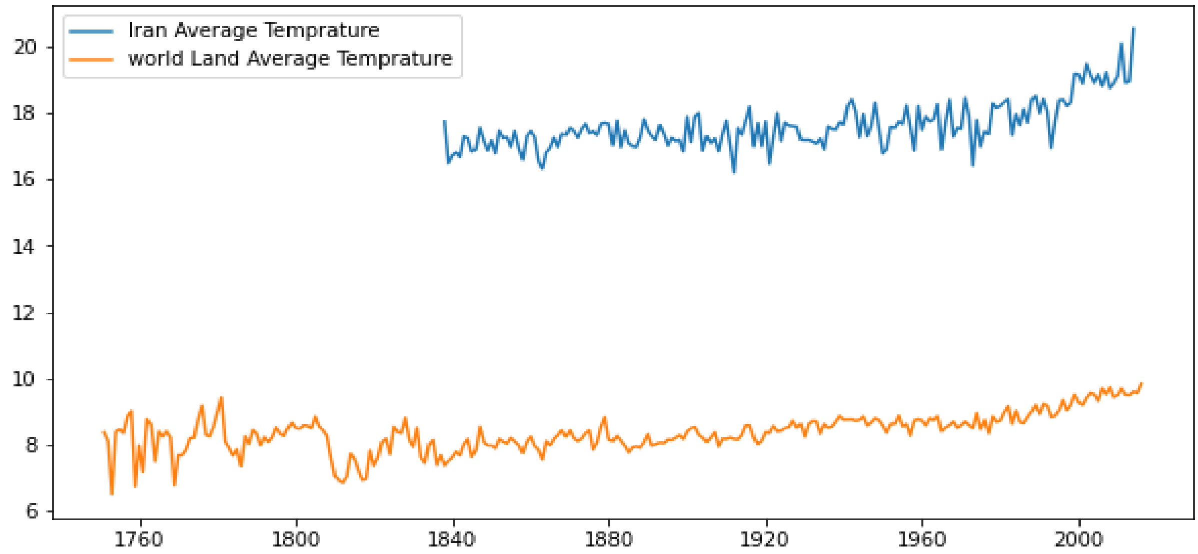
This chart is showing us how the temperature changed during the last three decades in average for Land Temperature, there might be some misinformation during the first decade (since 1750 to 1820) that can be due to the lack of technologies for assessing the temperature or also operators faults but after the 1820 the trend looks much more rational and it is growing slightly during the years.

Then it came to my mind how my country's temperature has changed during the last decades so then I have filtered the countries dataset with this term ['Country'] == 'Iran' to get the data of Iran and find stored it to a new data frame and the plot of that data looks like this:



Since for this country data are not recorded from 1750 we can see the volatality of data is much less and the chart is showing that average temperatures of Iran is also rising dramaticly just like the Global Land average temprature.

The trend of global and country looks pretty much same but I wanted too look at them in one chart to see how they are changing during all these years comparing with each other so this is how it looks when we plot Global Average Land Temprature and Iran Average Land Temprature:



Their trend is pretty much same, both are rising during these years but Iran's Temperature increasement pace jumped during the last decade which is bad. Also, we can say that the average of Iran's temperature was always much higher than the Global Land average Temperature.

For this I did some research and I can explain why Iran's average temperature is much higher than the rest of the world:

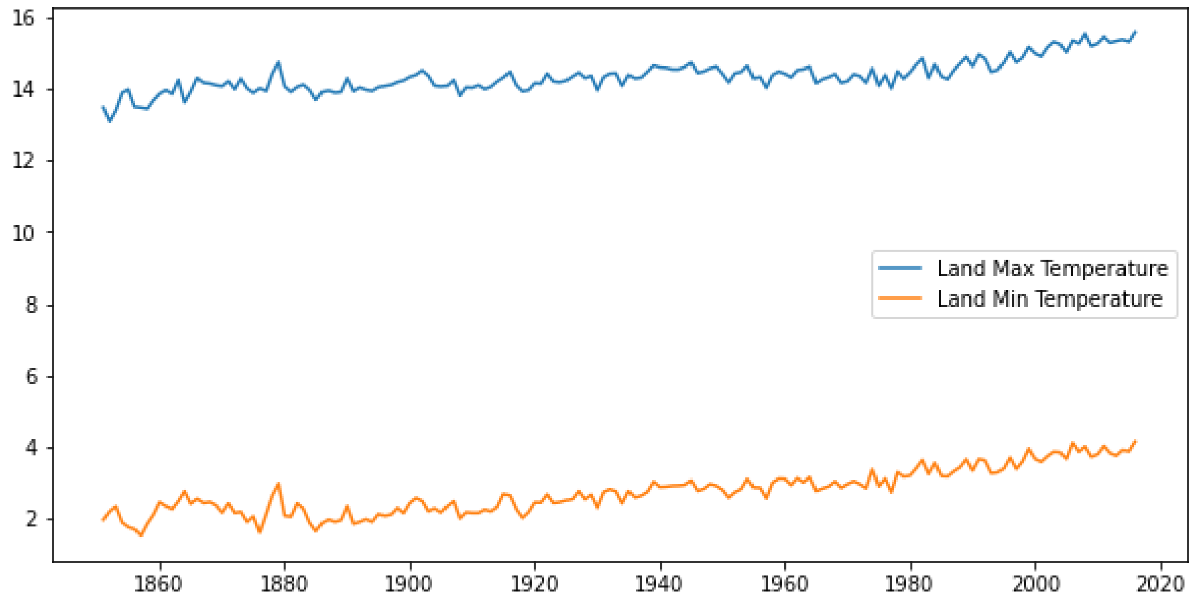
First, Iran is located in a region that is known for its high temperatures, particularly in the summer months. This region, known as the "Iranian Plateau," is located in the subtropical zone and experiences hot, dry summers and cold, wet winters. The high elevation of the plateau, which ranges from 1,000 to 3,000 meters above sea level, also contributes to the hot and dry climate.

Second, Iran is located in the middle of a large landmass, which means it is further from the moderating influence of the oceans. Coastal areas tend to have more moderate temperatures because the oceans act as a heat sink, absorbing and releasing heat more slowly than land masses. Inland areas, on the other hand, can experience more extreme temperatures because they are more exposed to the direct heating and cooling of the sun.

Finally, Iran is located in a region that is prone to hot and dry winds, known as "shamals," which can raise the temperature even further. These winds, which blow from the north and northwest, are caused by high-pressure systems over the Arabian Peninsula and can last for several days at a time.

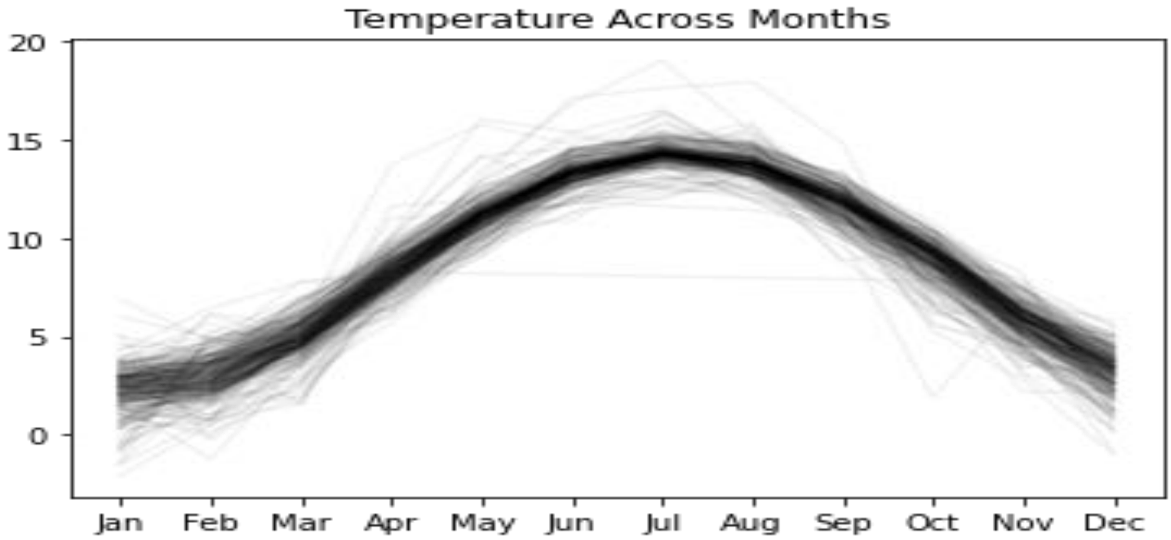
Overall, the combination of Iran's location, elevation, and climate patterns contribute to its higher average temperature compared to the global average.

Thehave decided to find out how the global Maximum and Minimum Average Temperature of Land is differentiated therefore I have plotted the following chart :



This diagram shows that the difference between Max and Min Land Temperature is maintained during the last three decades, so I was a bit curious how the average temperature of Land can vary throughout the seasons and different months.

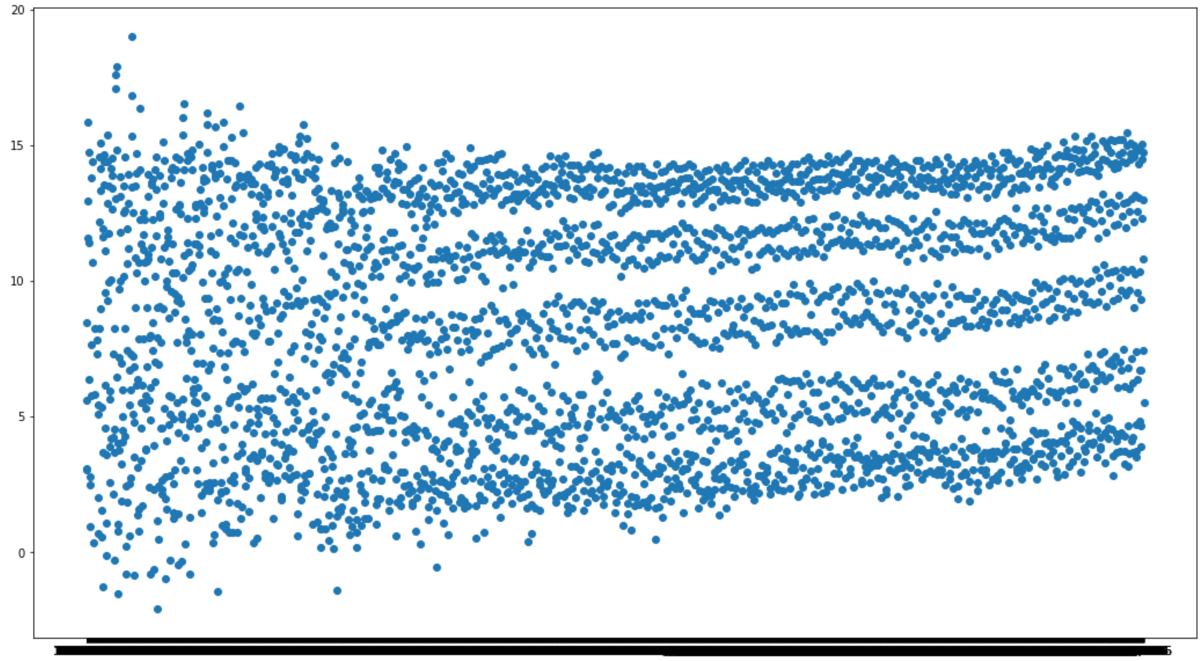
So I have listed the months into a List and then plotted the data of Global Land Temperatures based on the months and it looked like this:



It shows that the pattern of months is pretty much the same, however you can see some shades that shows the outliers but pretty much they follow the pattern of cooler weather in autumn and winter and warmer weather in summer and spring weather and the warmest month of the year was July through the last three decades.

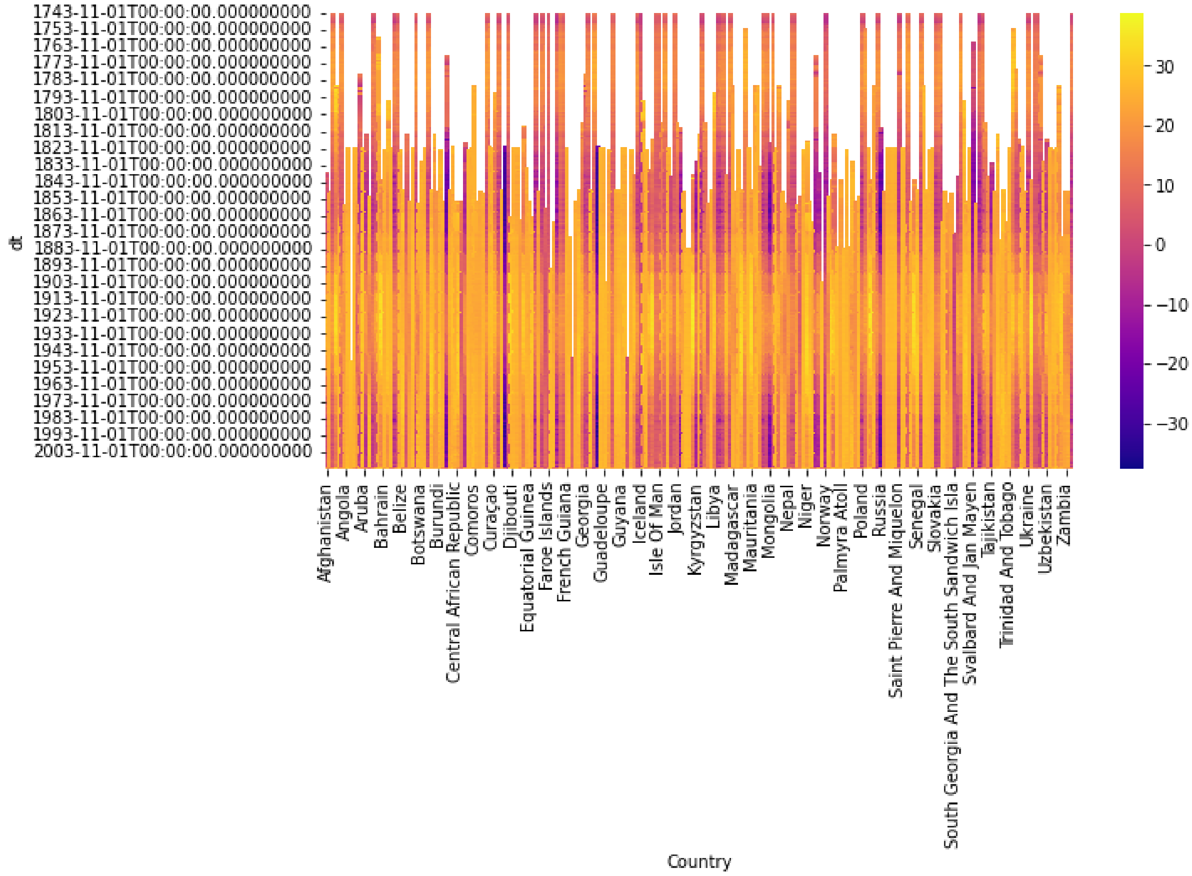
But the main reason that I have plotted the above chart was that I was looking for Season Shifting which is a phenomenon that refers to the change in the timing of seasonal events, such as the onset of spring or the arrival of migratory birds, due to changes in temperature and other factors. These changes can be caused by a variety of factors, including climate change, land use change, and natural variations in the Earth's climate system and the chart shows that season shifting did not happened that much during the last decades but still it is one of the important concerns of governments and people.

Then I have plotted the land average temperature in a scatter chart:



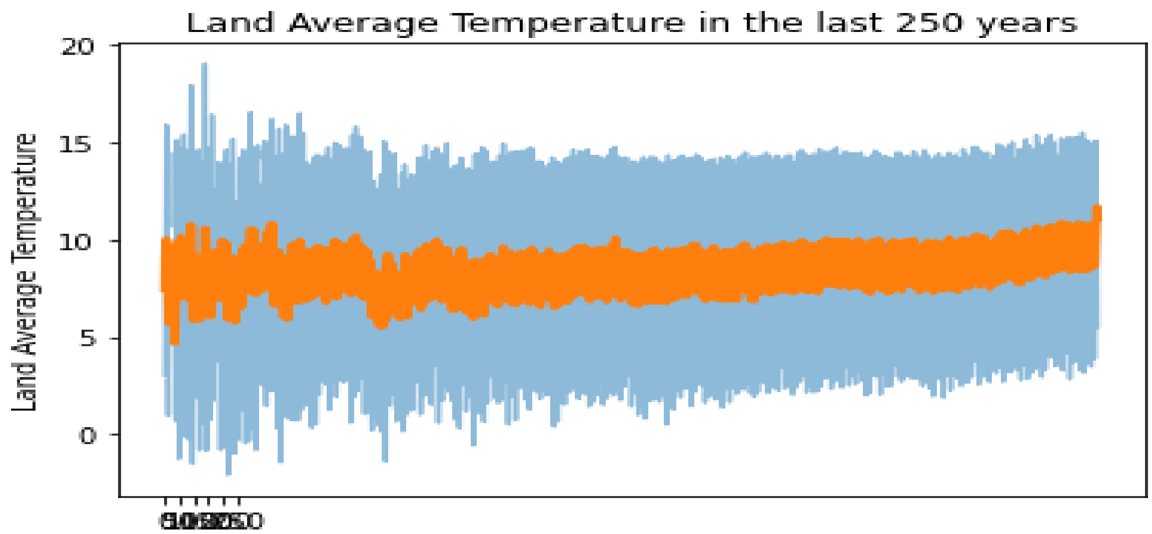
We can see how the temperature is scattered through the years and since this data is based on daily time series we can see that for each year we have different temperature and when we get closer to 2015 we can see that the dots are much closer to each other this shows we have less noise in our data (due to advancement of technology or better recording) and in the beginning of the data frame we can see some outlier temperatures.

Then I have tried to show every countries in heatmap to see how the countries differ in term of average temperature so it became like this:



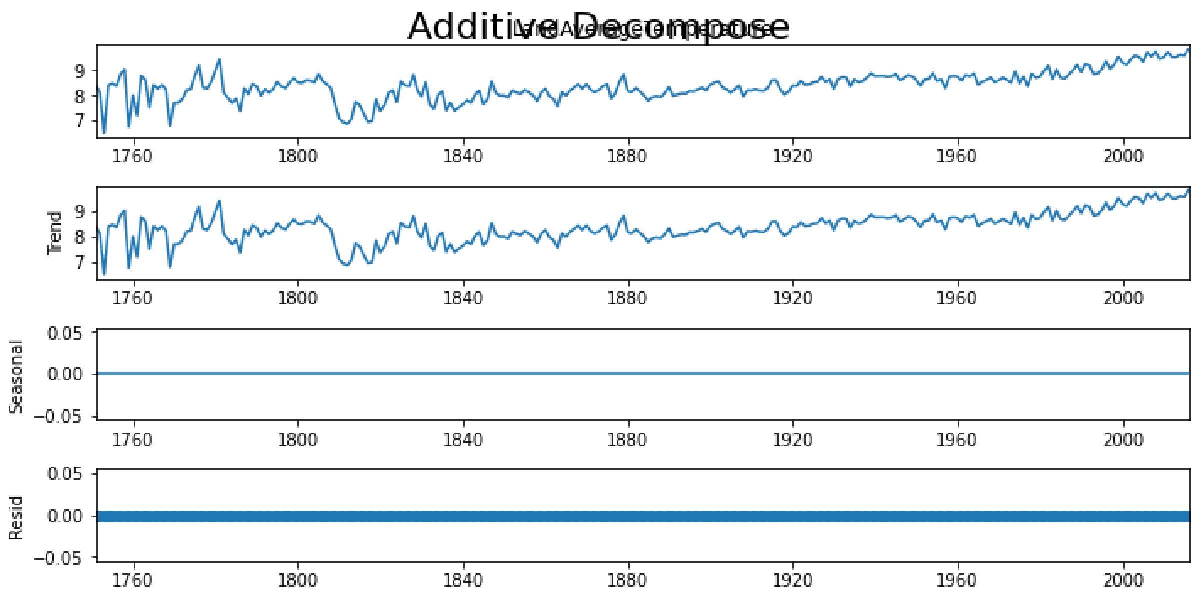
Some of the countries didn't fit the chart but it gives the sense how the countries temperature is changing during the years and due to the legend of this chart you can see that the yellow range color is mostly seen and we can assume that temperature is rising throughout the years.

In the next chart I wanted to show how the trend of max, min and average of the Land Temperature is moving so it looks like this:



The yellow color shows the average and blue shades are the Max and Min of the temperatures.

After that I have used the seasonal decompose method to show the trend over the years and also residuals and it look like this:



Then for predictions first I have used the ARIMA model to explain the ARIMA I shoud say ARIMA is an acronym that stands for "AutoRegressive Integrated Moving Average." It is a statistical model that is commonly used for time series forecasting. Time series forecasting involves using historical data to make predictions about the future.

The model is comprised of three components:

AutoRegressive (AR) terms: These terms represent the past values of the time series.

Integrated (I) terms: These terms represent the differences between the past values and the mean of the series.

Moving Average (MA) terms: These terms represent the error terms or noise in the time series.

The model is typically denoted as ARIMA(p,d,q), where p is the number of AR terms, d is the number of differences required to make the time series stationary, and q is the number of MA terms.

The model is fit to a time series by estimating the values of p, d, and q that minimize the error between the predicted values and the observed values of the time series. Once the model is fit, it can be used to make predictions about future values of the time series.

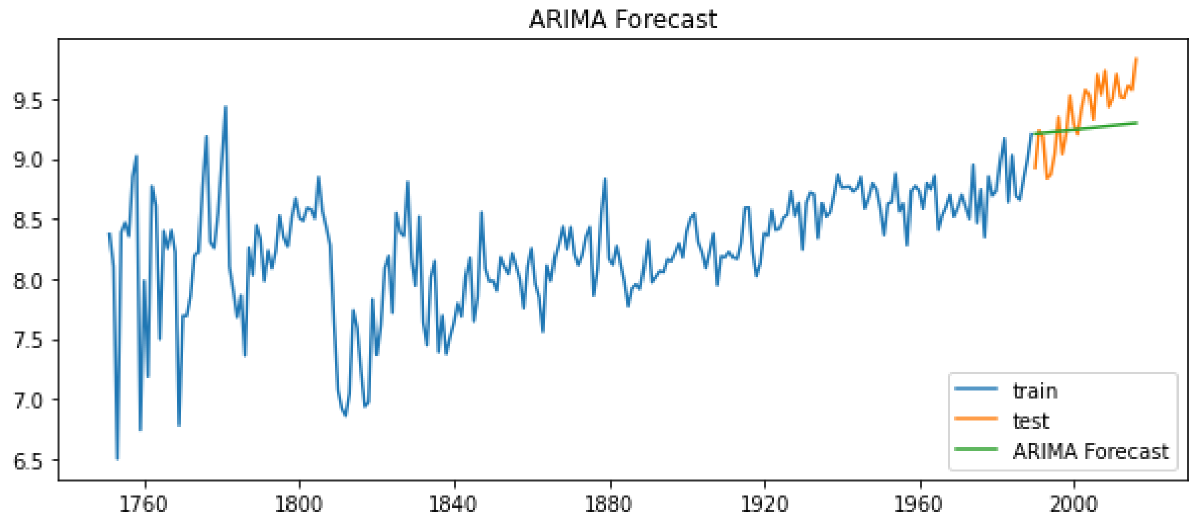
So first I made a function to calculate the best score for p, d and q since running and found out that the best score can be achived with these setting:

p_values = range(0,8)

q_values = range(0,8)

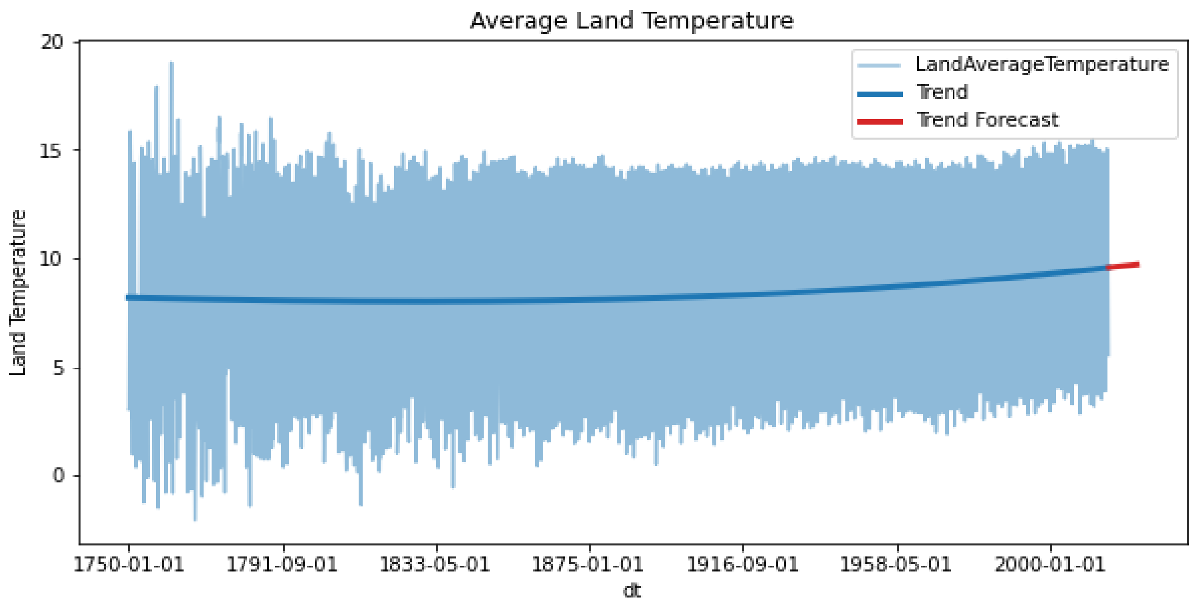
d_values = range(0,3)

and then launched the ARIMA model on my data frame to predict the future of Land average temperature and showed the prediction in a chart:



It can be seen that ARIMA forecast is close to what happened in real world and the prediction can be considered satisfying.

Then I have tried another method to predict the future of global warming and I have used Linear Regression method which is a statistical method used to model the linear relationship between a dependent variable and one or more independent variables. linear regression is a powerful and widely used statistical method that allows us to model and understand the relationships between different variables. So I have run the model on the Global Land Average Temprature and plotted the resault to see how the prediction is predicting the future of global warming.



The model was able to capture the underlying trends and patterns in the data, and the predictions it made were closely aligned with the observed values and it shows that global average land temperature is going to increase during the upcoming years.

Conclusion

Whole the data shows that the global temperature is rising dramatically and governments, NGOs and people must take action to stop it otherwise we will face such things:

Changes in weather patterns: Higher temperatures can lead to more frequent and severe heatwaves, droughts, and extreme weather events such as hurricanes and floods.

Melting of polar ice caps: Higher temperatures can cause polar ice caps to melt, which can lead to rising sea levels and coastal flooding.

Changes in ecosystems: Higher temperatures can cause shifts in the distribution of plants and animals, as well as changes in the productivity of agricultural lands.

Negative impacts on human health: Higher temperatures can increase the risk of heat-related illness and death, particularly among vulnerable populations such as the elderly and those with pre-existing health conditions.

Negative impacts on economic activity: Higher temperatures can lead to reduced crop yields, damaged infrastructure, and other economic impacts, particularly in sectors such as agriculture and tourism.

Negative impacts on biodiversity: Higher temperatures can lead to the extinction of some species and the proliferation of others, disrupting ecosystems and the services they provide.

Overall, the impacts of rising global average temperatures are likely to be widespread and varied, and they will likely disproportionately affect vulnerable populations and ecosystems and the habitants of Earth planet must take it seriously.