

**Data Analysis using Python**

**Project title: Global temperature analysis of top economies**

By:

**Jupalli Nitesh: 2021PGP208, Contact No. 7680057572**

**Rakshita Singh: 2021PGP228**

**Guduri Anupam Reddy: 2021PGP143**

**Suraj Agarwal: 2021PGP239**

**Narendra Penumarti: 2021PGP159**

**Problem Statement:**

Climate change is one of the pressing issues which the world is facing and the most important factor to track climate change is change in temperatures across time.

Our problem statement is to analyse the data and find out the cities of top 8 countries which are more vulnerable to global warming effect due to temperature change beyond the permitted threshold. This will include detailed analysis (Numerical and Graphical representation) of temperatures over past 150 years.

**Dataset Description:**

Some say climate change is the biggest threat of our age while others say it’s a myth based on dodgy science. We are turning some of the data over to you so you can form your own view.

We have repackaged the data from a newer compilation put together by the Berkeley Earth, which is affiliated with Lawrence Berkeley National Laboratory. The Berkeley Earth Surface Temperature Study combines 1.6 billion temperature reports from 16 pre-existing archives. It is nicely packaged and allows for slicing into interesting subsets (for example by country). They publish the source data and the code for the transformations they applied. They also use methods that allow weather observations from shorter time series to be included, meaning fewer observations need to be thrown away.

**The dataset contains following columns:**

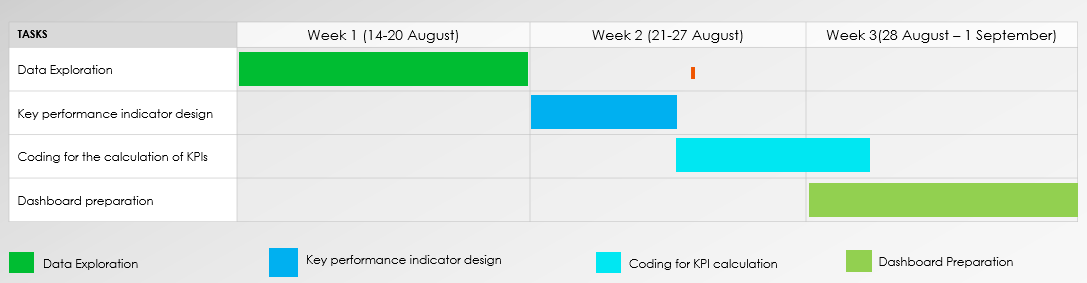
* **Date:** starts in 1750 for average land temperature and 1850 for max and min land temperatures and global ocean and land temperatures
* **Land Average Temperature:** Global average land temperature in Celsius
* **Land Average Temperature Uncertainty:** The 95% confidence interval around the average
* **Land Max Temperature:** Global average maximum land temperature in Celsius
* **Land Max Temperature Uncertainty:** The 95% confidence interval around the maximum land temperature
* **Land Min Temperature:** Global average minimum land temperature in Celsius
* **Land Min Temperature Uncertainty:** The 95% confidence interval around the minimum land temperature
* **Land And Ocean Average Temperature:** Global average land and ocean temperature in Celsius
* **Land And Ocean Average Temperature Uncertainty:** The 95% confidence interval around the global average land and ocean temperature

**Advantages Of the Proposed Solution:**

This data would help us in gauging the temperature change in top economies of the world. What has been the trend since 1700s to till date. It would help the countries to know the trend and predict the future rise in temperatures. This analysis also provides the cities where maximum or minimum temperature is recorded. Which further opens up the ways for countries to improve the sustainability in the particular region.

**Timeline:**

1. Data exploration
2. Key performance indicator design
3. Coding for the calculation of KPIs
4. Dashboard preparation



# 

**Implementation on Flask dashboard**

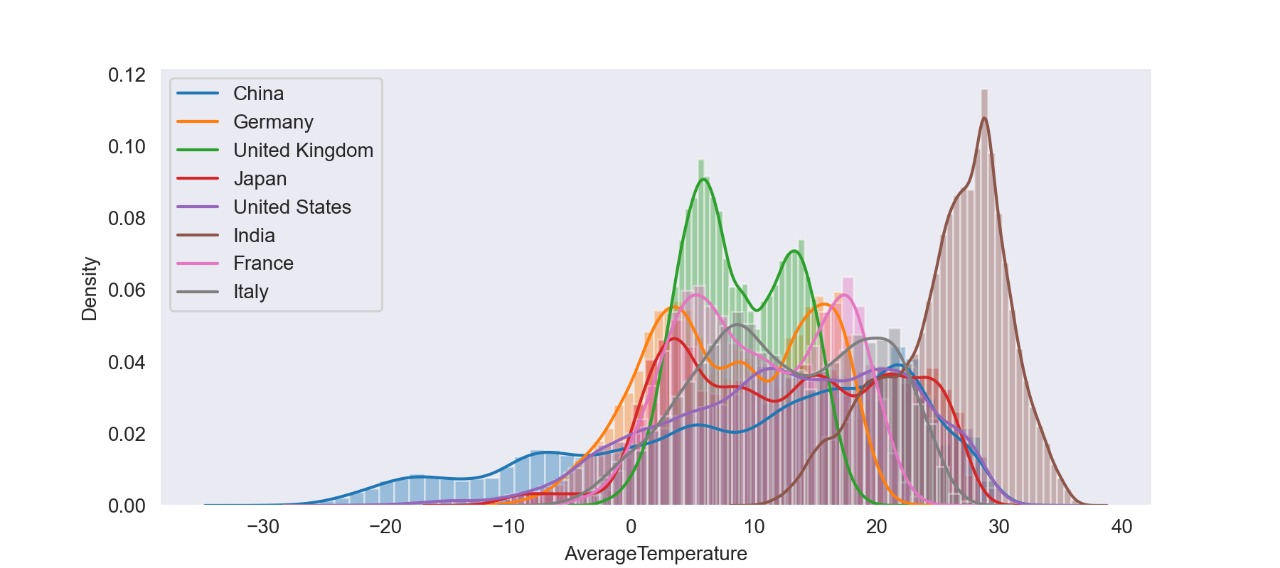
Following process is implemented to host the pictures onto the flask dashboard.

1. Initial data analysis is done on the Jupyter notebook. Cleaning the data and dealing with the missing values has been done.
2. After the data analysis and visual interpretation representation, we have decided to host the data on web dashboard with the help of Pycharm application.
3. After installing the Pycharm professional application, we hosted the code on the app.py file and render template using index.html.
4. Then the code was run to host it on the web dashboard.

**Analysis**

## Bivariate analysis of Mean Temperatures across nations

Bivariate analysis is study of analysis of concurrent relation between two variates. Here in this scenario, we analysed how the temperatures are changed across time in mentioned 8 countries and the frequency of their occurrence(density). After doing analysis, we’re able to generate the below graph and from which we inferred some key points.



### Observations

* China has the longest range of temperature range, starting from -35 degrees to 35 degrees centigrade.
* China also didn’t have any high values in density, which indicates that their temperatures were distributed uniformly and didn’t have much frequency of a particular temperature point
* India on the other hand have its temperatures range from just 10 degrees to 40 degrees centigrade
* India is also recorded the highest temperatures among these 8 countries i.e., 40 deg
* India has the highest peak in the distribution axis at 30 degrees, which implies that many times India temperatures were at the same temperature point
* Japan has their temperature range from -15 degrees to 30 degrees and their curve is pretty uniformly distributed with a few curves going up and down, and we can understand from this that their temperatures were recorded uniformly across years
* United Kingdom, France, Italy and Germany have a similarity with two peak temperatures recorded over years and this leaves us with an understanding that the winters and summers in these countries were recording a similar peak temperature most of the times
* The United States graph is similar to China’s graph except the range of the temperatures is low when compared to China and this helps us to infer that US also have a uniform distribution curve with different temperature points occurring frequently

## Univariate Analysis of mean temperatures across India

As the name depicts, Univariate analysis is analysing data with a single variable. In this case, we took the mean temperatures of cities in India and plotted the top 13 on the map using latitude and longitudes.

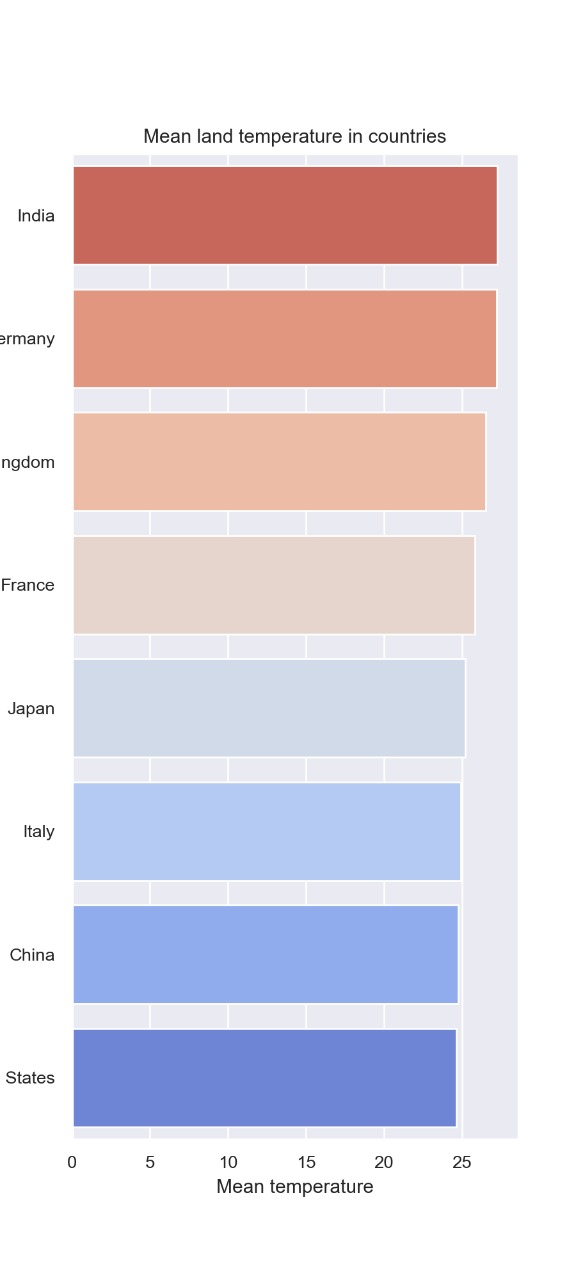


### Observations

* The top 13 cities with mean temperatures were spread uniformly across India from North to South
* The city with highest mean temperature is:

|  |  |
| --- | --- |
| **City Names** | **MeanTemperature** |
| Alandur | 28.41785845 |
| Ambattur | 28.41785845 |
| Avadi | 28.41785845 |
| Bhimavaram | 27.4098748 |
| Ambur | 27.2815941 |
| Adoni | 27.25905463 |
| Anantapur | 27.25905463 |
| Bellary | 27.25905463 |
| Adilabad | 27.21464912 |
| Bidar | 26.84797368 |
| Bhavnagar | 26.76565862 |
| Ambarnath | 26.63145215 |
| Badlapur | 26.63145215 |

## Bivariate Analysis on Mean Land Temperature in different countries

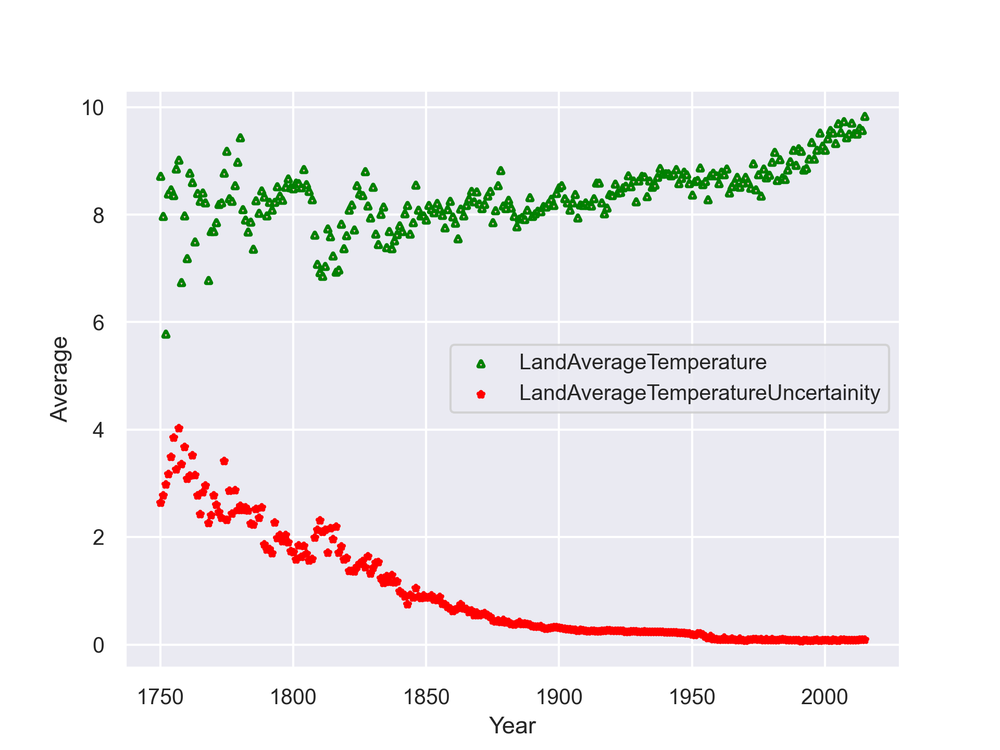
In this scenario, we tried to compare the countries based on the mean land temperature and tried to understand the climatic situation in these countries.

Observations

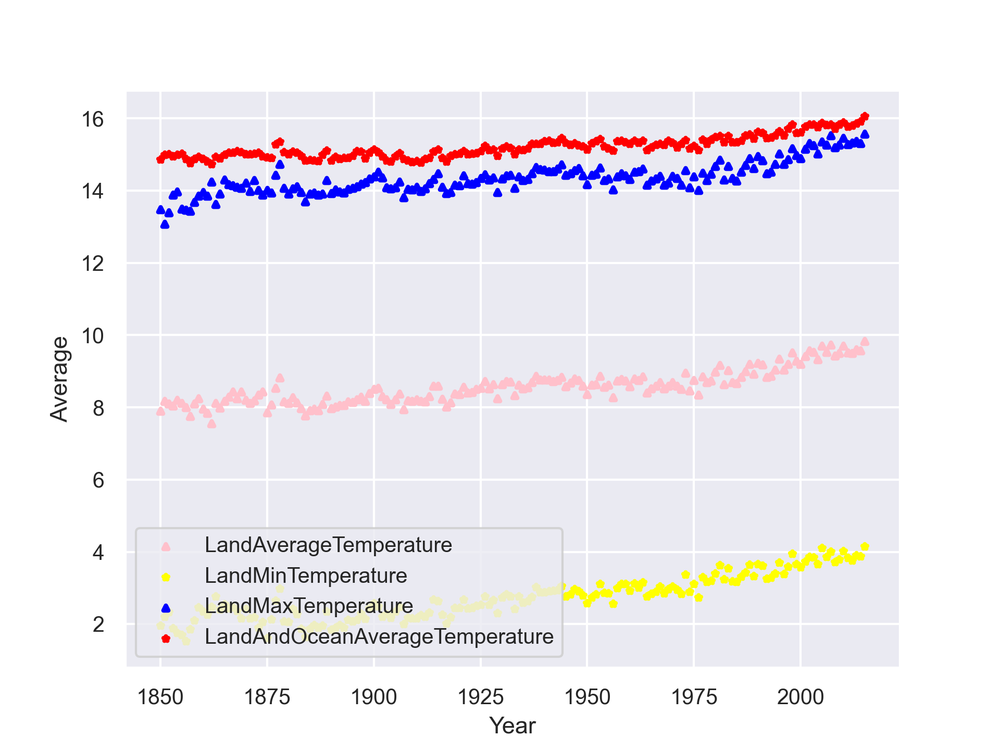
* India has the highest mean temperatures when compared to the other 7 nations under discussion and one of the possible reasons could be due to the presence of the Himalayan Mountains operating as a barrier to winds coming from Central Asia and China
* India is followed by Germany, France, UK, Japan and Italy. Out of these except Japan all other belongs to Europe and have geographically similar conditions and this can explain the reason why all of them have similar temperatures
* USA and China were above the tropic of cancer and due to winds from high altitudes, these two countries have lower mean temperature when compared to other countries

## Multivariate analysis of Mean temperature and Mean uncertainty:

Multivariate analysis is where we’ll try to analyse more than two variables to understand the relation among them. Here we compared the change of temperatures of these countries and the uncertainty across years. We created a scatter plot to analyse this scenario. The results were as shown below.



### Observations:

* We can see that the mean land temperature and uncertainty values of the scatter plot were initially scattered away but towards the 20th century they got closer and this shows that the relation initially was moderate but coming towards it became strong.
* The temperatures were having a positive slope and that means that with time temperatures were increasing and on the opposite side uncertainty kept on decreasing which implies that uncertainty kept on decreasing and thanks to the new oil of this era DATA
* We can observe that around the year 1975, the plot took a sudden spike and this shows that the rapid industrialization had actually impacted the Global warming
* From 1975 we can also observe that the slope of temperatures has actually raised and this implies that Global warming started raising the temperatures rapidly
* As mentioned by an old Confucius statement, “The best time to plant a tree was 20 years ago, the second-best time is now”. It’s time to realize that GLOBAL WARMING IS REAL and act now as it’s better late than never

## Multivariate analysis of average max, min and mean land temperature and mean ocean & land temperature

In this scenario, we tried to compare the average of maximum, minimum and mean land temperatures and mean land & ocean temperature across the 8 nations plotted against the timeline. We again plotted this on a scatter plot for analysis.

### Observations

* We can clearly observe that all the curves were almost parallel and this clearly depicts that the temperatures maximum and minimum temperatures grown uniformly across the years
* Here also we can observe the little spike in all curves around 1975 and there is slight increase in the slope post that which indicates impact of Global warming
* We can also observe that, initially the maximum land temperature and mean land & ocean temperature curves have some gap between them but since 1975 it kept reducing and currently both of them were at almost same level which indicates that the land temperature is growing steeply when compared to the ocean temperatures due the global warming effect
* We can observe a spike around 1875 and while trying to find the reason for this we went through an article which discussed below points
  + Between 1875 and 1878, severe droughts ravaged India, China and parts of Africa and South America and lead to deaths of 50 million people
  + The result was a famine that struck three continents and lasted three years
  + The most obvious was a [big El Niño in 1877-78](http://dx.doi.org/10.1007/s10584-008-9470-5). During an [El Niño](https://www.newscientist.com/article/dn28595-massive-el-nino-sweeping-globe-is-now-the-biggest-ever-recorded/), warm water spreads over the Pacific, releasing heat into the air
  + For more details regarding this you can find it here: <https://www.newscientist.com/article/2183901-a-freak-1870s-climate-event-caused-drought-across-three-continents/>

**Conclusion**

After observing the above-mentioned aspects from the analysis, we understood few points:

* Global Warming is real and a raise in the temperatures from 1975 is the solid proof
* If we don’t act immediately, the consequence would be an even steeper raise in the temperature followed by melting of ice which can destruct the globe
* Based on the geographical location of a country, the range of temperatures is varying, for example, India has the range from 10 degrees to 40 degrees whereas China have it from -35 degrees to 35 degrees and Japan have it from -15 degrees to 30 degrees
* From the Multivariate analysis of Mean temperature and Mean uncertainty, we can see that the density of scattered points was increasing towards the end and this implies that the relationship between had increased with time
* Another key observation here is the temperature spike in 1875 and when deeply analysed the cause was found out to be **El Niño.** It is a climate pattern that describes the unusual warming of surface waters in the eastern equatorial Pacific Ocean and 1975 **El Niño** is described as the one of the deadliest environment disasters as the drought caused by it killed almost 50 million people in between 1875-78 period
* The mean temperature of India is highest due to the presence of Himalayan range
* European countries have similar mean temperatures due to geographical similarities
* US and China have the least temperatures due to winds from higher altitudes