Capstone Project 2: Project Proposal

Project Title

"Time Series Analysis of Global Temperature Trends"

Project Proposed by

Purushottam Malgi

Project Mentor

Debarya Dutta

What is the problem to be solved?

Earth is a indigenous planet with home to many species. Humans, animals, and plants are all the living species that share the space on Earth. But the most necessary and important thing for all these species to survive is the environment surrounding them. Every living thing on planet Earth requires fresh air to breathe, good water to drink and food for survival.

And all this is dependent on the climate and its effect. It's necessary that it rains because it is a source of our drinking water and is necessary for vegetation to habitat on the surface.

Recently climate change and global warming are the most critical problems faced by the mankind. Through this project we can see how the climate has changed over time and what can be the effects on the future climate change.

Who is the client or interested body for this problem to be solved?

Global Temperature change trends are helpful to every individual surviving on this planet. But more importantly this problem will give certain solutions to all the country's governments and international bodies who are responsible to keep climate change in limits. Organizations such as UNESCO will be most benefited by the solution of this problem of climate change. Also as humans we can see the Global Temperature changes and change our habits in order that the Earth survives for a longer time in the future.

Looking at the trends and the temperature rise all the governments can take valuable measure to reduce Global Temperature rise. Measures such as Stop plastic use, less use of harmful gas emitting devices and vehicles.

What data I will be using and where did I acquire the data?

The dataset is received from a kaggle competition and it is a data of Climate Change: Earth Surface Temperature Data.

It contains files which are divided into Global Temperature of the world, Global Average Land Temperature by Country, Global Average Land Temperature by State and Global Land Temperature by Major City.

The link to the dataset is:

https://www.kaggle.com/berkeleyearth/climate-change-earth-surface-temperature-data

Techniques to be used in solving the problem:

1. Data Wrangling - I will observe the data for all the missing values, incorrect entries and based on the observation I will clean the data for further use.

2. Exploratory Data Analysis - I would plot the relation between different parameters in the dataset and based on that I can find the trends for storytelling.
3. Inferential Statistics - This is another important aspect based on which we can find out some statistical relations in the dataset. I will be applying different testing methods in Inferential Statistics.
4. Time Series Modelling - I will perform Time Series Analysis to find the future trends in the temperature change by using models such as Autoregressive model, Moving Average model and ARIMA model.

Deliverables:

I will be using the following tools to work on the project :-

- Jupyter Notebook for the major part of the project.
- I will be using Python version 3.x for writing the code.
- I will use Google doc for submission on the in process project reports.
- Powerpoint for Data Storytelling presentation.

References:

- Springboard Curriculum material
- DataCamp courses
- Google, Stackoverflow, Github, etc.
- And any other Open Resources.

Data Wrangling:

The dataset consists of the following files:
1. Global Land Temperature by Major City
2. Global Land Temperature by Country
3. Global Land Temperature by States
4. Global Temperature

We did some work on Global Temperature dataset and some preliminary features of that data file were as follows:

```
In [5]: temp_global.shape
Out[5]: (3192, 8)

In [6]: print (temp_global.info())

        <class 'pandas.core.frame.DataFrame'>
        DatetimeIndex: 3192 entries, 1750-01-01 to 2015-12-01
        Data columns (total 8 columns):
        LandAverageTemperature                   3180 non-null float64
        LandAverageTemperatureUncertainty        3180 non-null float64
        LandMaxTemperature                       1992 non-null float64
        LandMaxTemperatureUncertainty            1992 non-null float64
        LandMinTemperature                       1992 non-null float64
        LandMinTemperatureUncertainty            1992 non-null float64
        LandAndOceanAverageTemperature           1992 non-null float64
        LandAndOceanAverageTemperatureUncertainty 1992 non-null float64
        dtypes: float64(8)
        memory usage: 224.4 KB
        None
```

The data file consisted of 3192 data points and 8 columns as features.

| dt	LandAverageTemperature	LandAverageTemperatureUncertainty	LandMaxTemperature	LandMaxTemperatureUncertainty	LandMinTemperature	LandMinTempe
1750-01-01	3.034	3.574	NaN	NaN	NaN	
1750-02-01	3.083	3.702	NaN	NaN	NaN	
1750-03-01	5.626	3.076	NaN	NaN	NaN	
1750-04-01	8.490	2.451	NaN	NaN	NaN	
1750-05-01	11.573	2.072	NaN	NaN	NaN	

This is the first five rows of the data file. It shows that the data is from 1st january 1750.

The columns of the data file are as follows:

```
In [30]: temp_global.columns
```

```
Out[30]: Index(['LandAverageTemperature', 'LandAverageTemperatureUncertainty',
                'LandMaxTemperature', 'LandMaxTemperatureUncertainty',
                'LandMinTemperature', 'LandMinTemperatureUncertainty',
                'LandAndOceanAverageTemperature',
                'LandAndOceanAverageTemperatureUncertainty'],
               dtype='object')
```

Land Average Temperature, Land Average Temperature Uncertainty, Land Maximum Temperature, Land Maximum Temperature Uncertainty, Land Minimum Temperature, Land Minimum Temperature Uncertainty, Land and Ocean Average Temperature, Land and Ocean Average Temperature Uncertainty. These are the columns in the data file.

Exploratory Data Analysis:
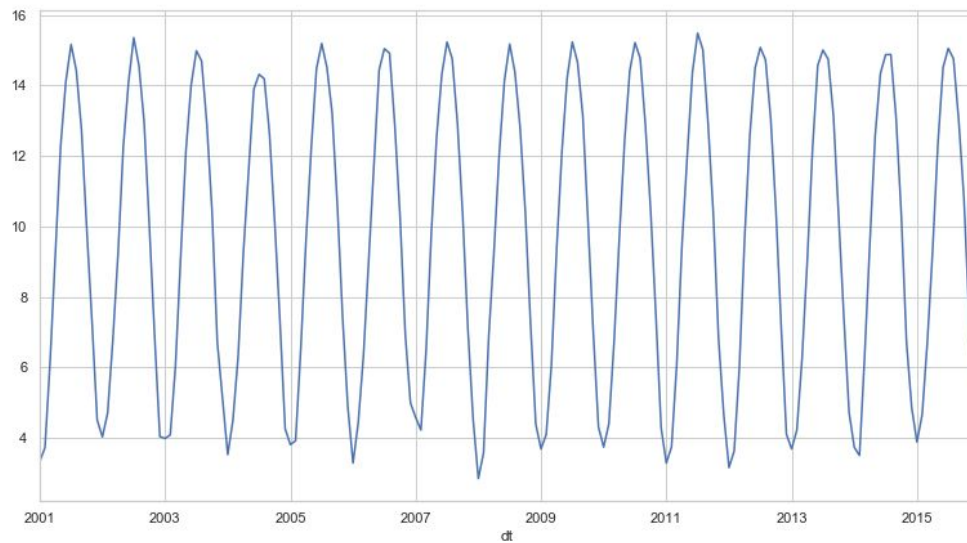
It can be seen from the above image that temperature fluctuation is equally distributed every which suggests it is a Time Series occurring with equal time period.
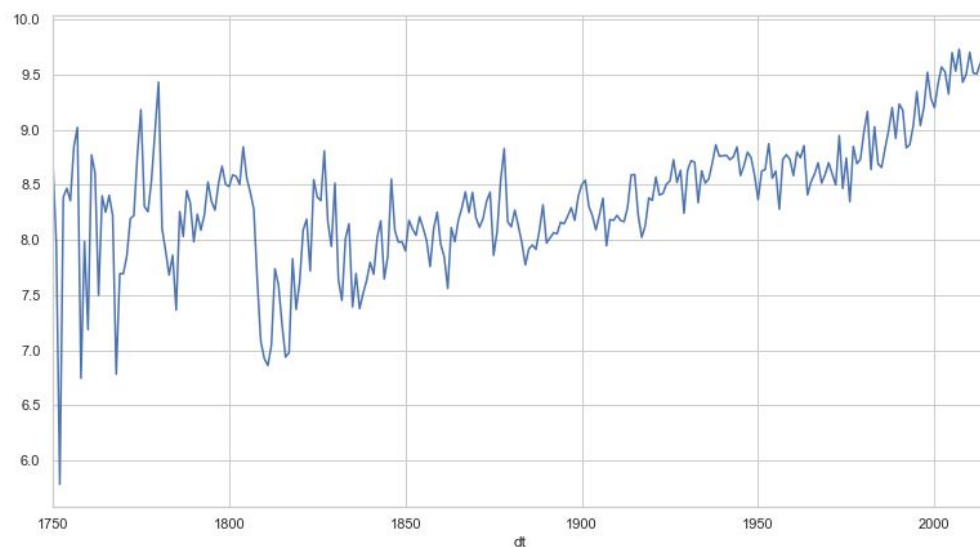

This graph shows the upward trend of the Temperature starting from 1750 and going beyond 2000. It shows an upward trend of temperature and with respect to this we can suggest there is Global Warming that's taking place in the world.
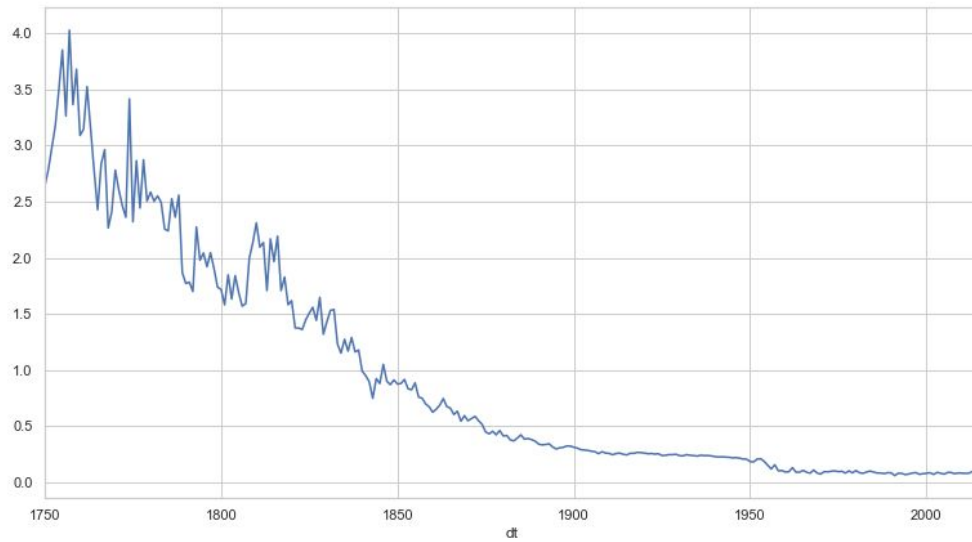
The mean of the temperature of every year shows a negative trend and after 1950s it is rather constant.

Average temperature between 1750 and 2003
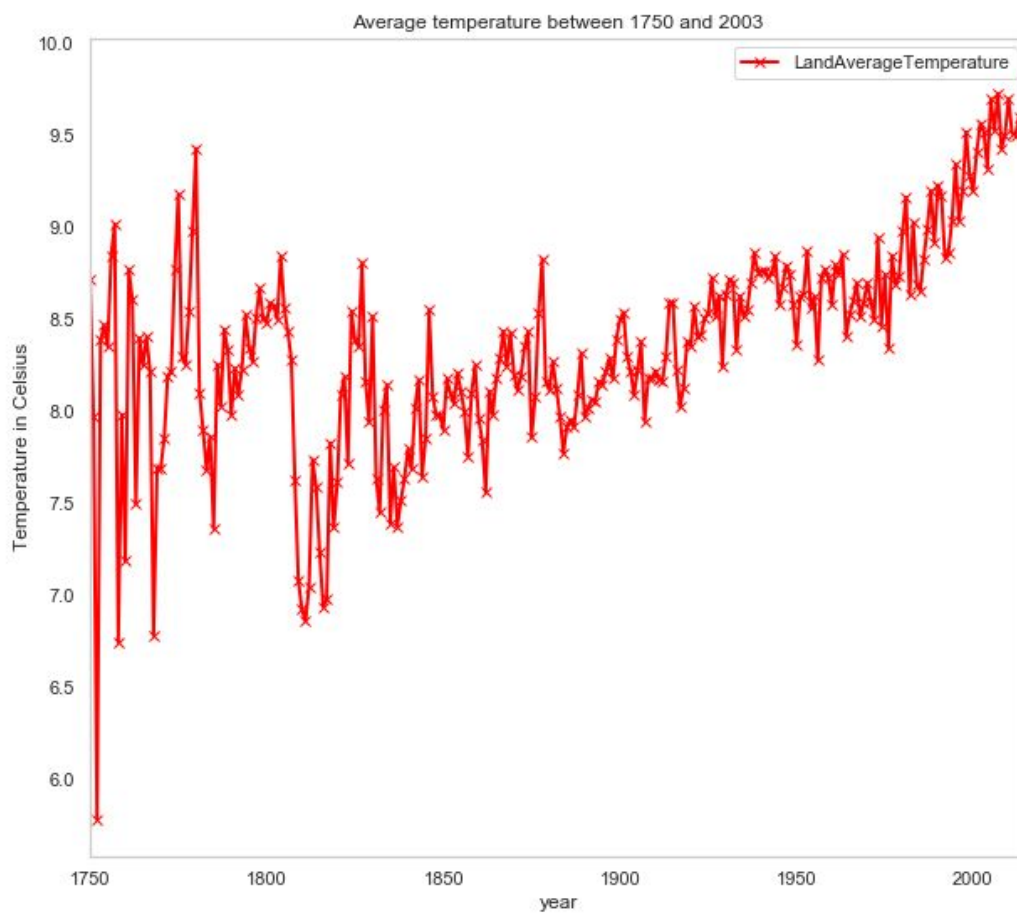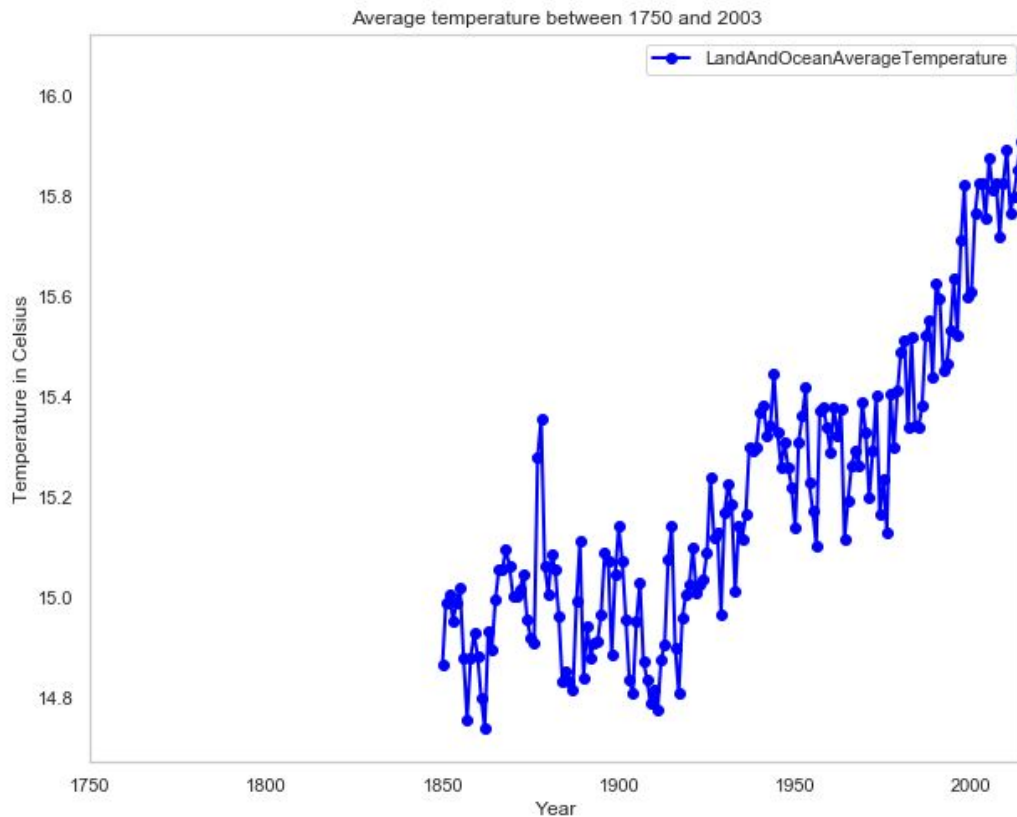
From the above two plots it can be seems that the Land Average Temperature and Land and Ocean Average Temperature has an increasing trend.

This suggests that the temperatures are increasingly constantly over the years and it shows signs of Global Warming. The ice caps are melting and the levels of oceans are increasing.

The threat to humanity is the rising temperature with which the existence of humans will be in danger in future if precautionary measures are not taken by the people and various governments of countries to tackle the temperature rise due to Global Warming.

Let's find out some trends in temperature in major Indian cities with some following visuals plots,

## Cities with the highest temperature differences
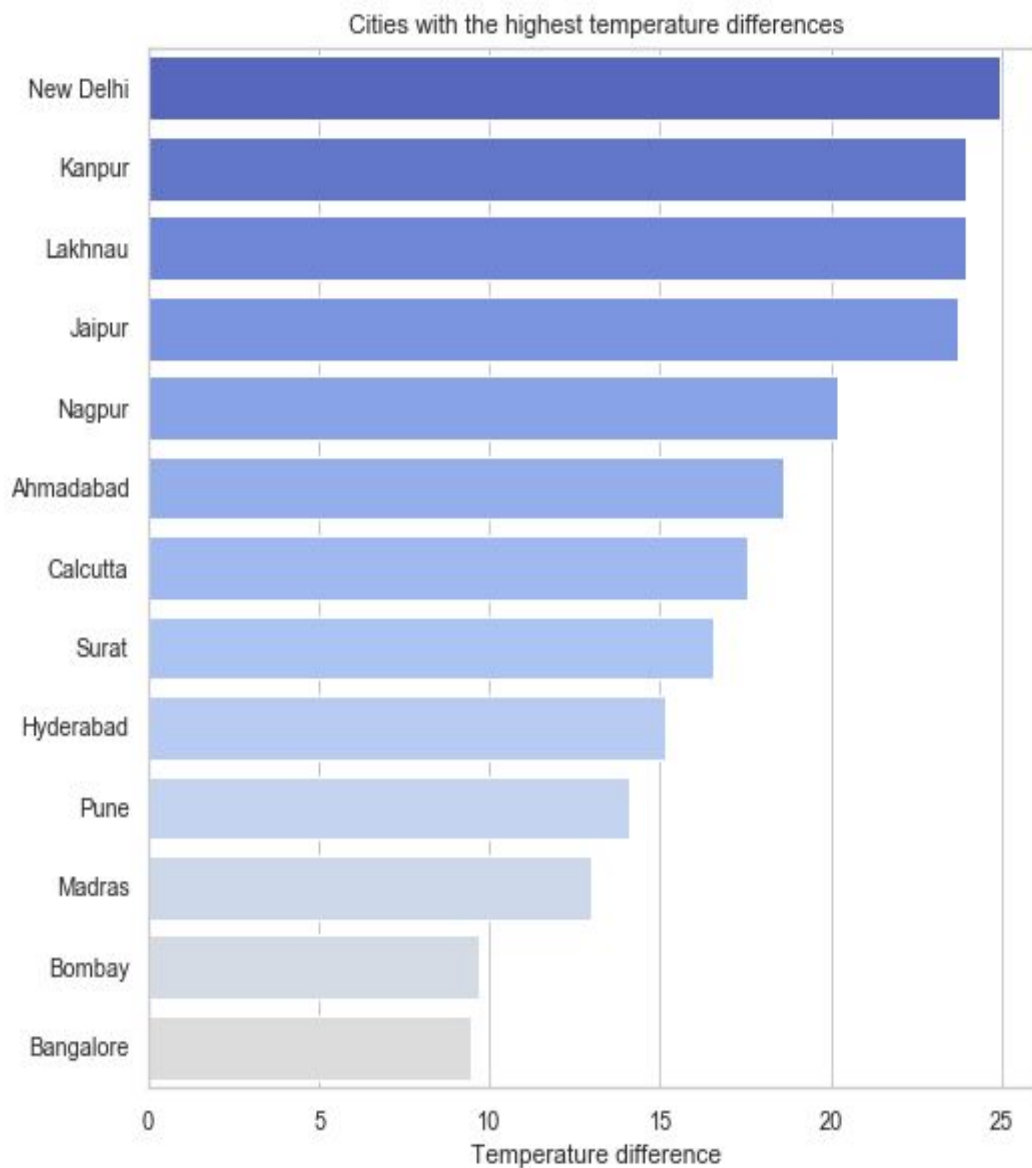
The above plot shows the temperature differences in major Indian city. Temperature difference is the difference of Maximum and Minimum Temperatures occurring in the cities.

It can be seems that New Delhi has the maximum temperature difference and Bangalore the minimum.

Cities maximum temperatures
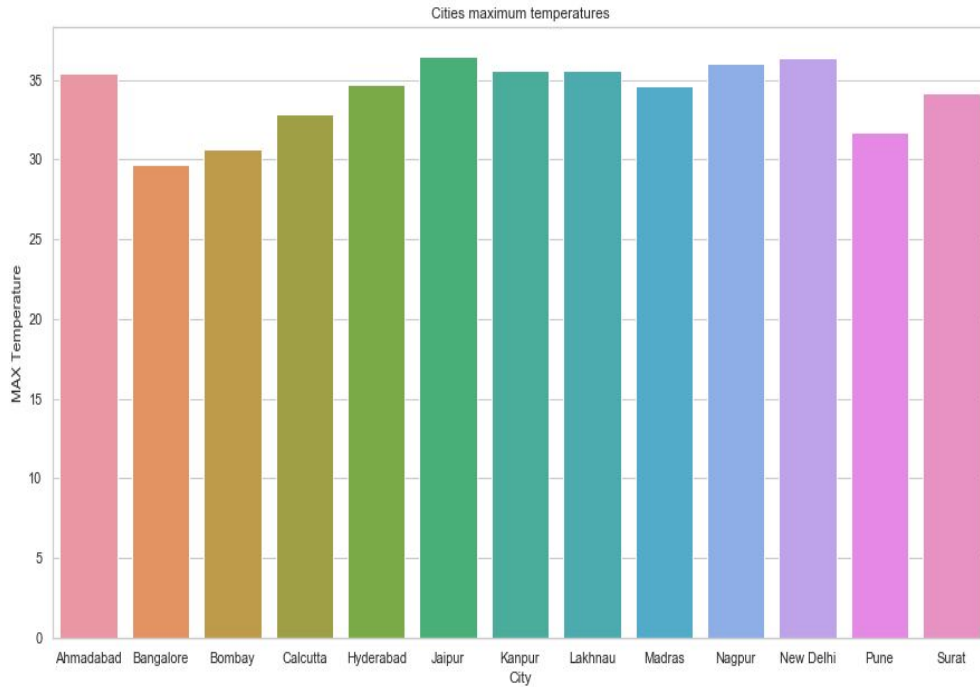
This bar plot shows the Maximum Temperature of different cities of India.
Ahmedabad, Jaipur, Nagpur and New Delhi topped the list with an average maximum temperature of 36 °C.
Bangalore and Pune were on the lowest side of 29 °C and 32 °C respectively.

Cities minimum temperatures
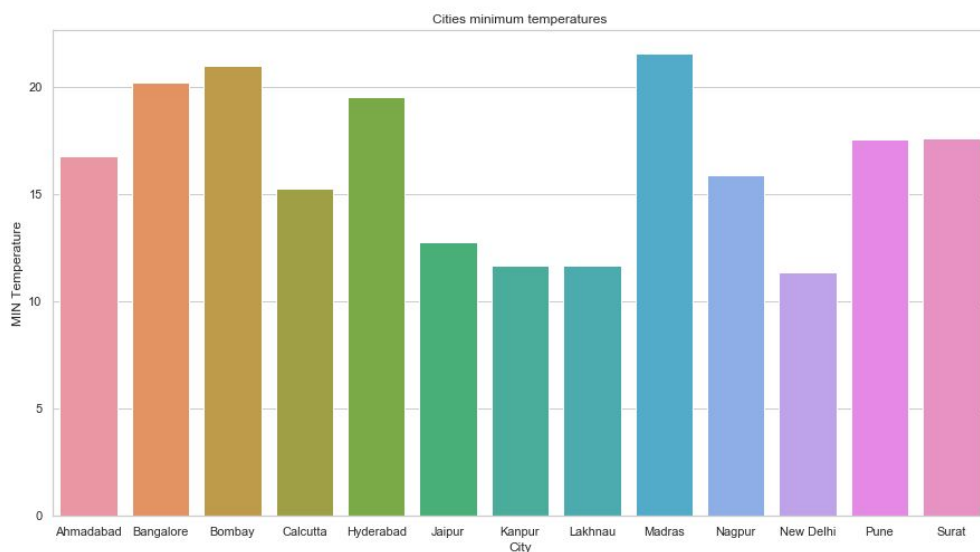
The above plot shows cities with minimum temperatures.

Cities Min MAX temperatures