

Data Analyst Nanodegree #Project 1

Explore Weather Trends

NAME: AYUSH

ayush.nda7@gmail.com

DATE: 02/05/2020

COUNTRY: INDIA

Introduction

In this project, I will analyse local and global temperature data and compare the temperature trends where I live to overall global temperature trends.

What Software Do I need?

To complete this project, I used the following:

- SQL
- MS Excel

Extracting Data

To start I wrote an SQL query to retrieve all the temperature data from:

- Selecting Nearby City (Patna)

```
SELECT*FROM city_list
```

```
WHERE country='India';
```

- City Database

```
SELECT*FROM city_data
```

```
WHERE country='India' AND city='Patna';
```

- Global Database

```
SELECT*FROM global_data;
```

Upon retrieving the data, it was extracted on a CSV file for further evaluation.

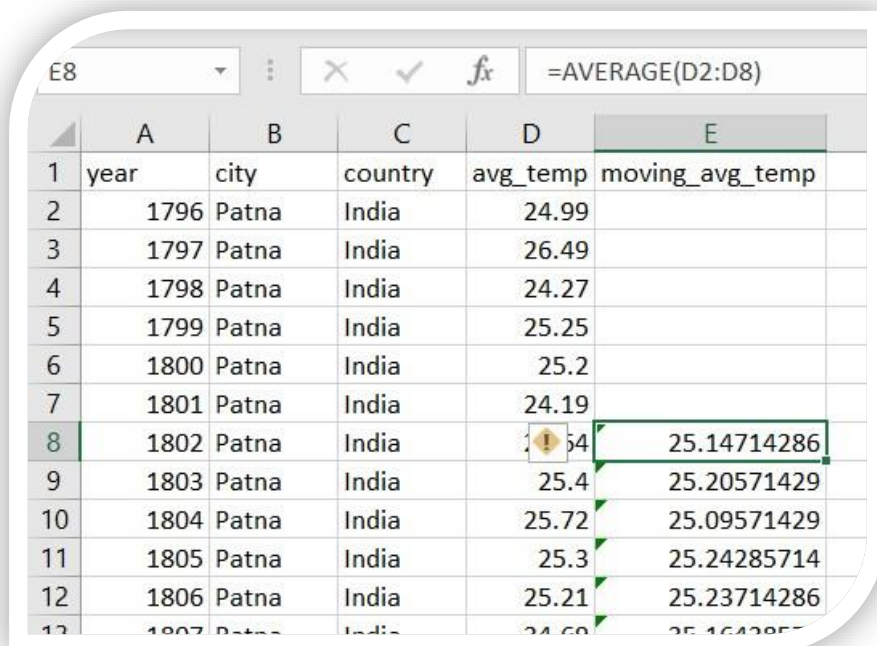
Data Selection & Manipulation

Once the data has been extracted on CSV, it was possible to further evaluate the data on Microsoft Excel. To provide a more accurate and useful dataset for comparison a range of the common years only was chosen.

Moreover, due to fluctuations in yearly averages, the data was evaluated considering moving averages (7-years) to provide smoother results during data visualization.

Steps to get moving averages :

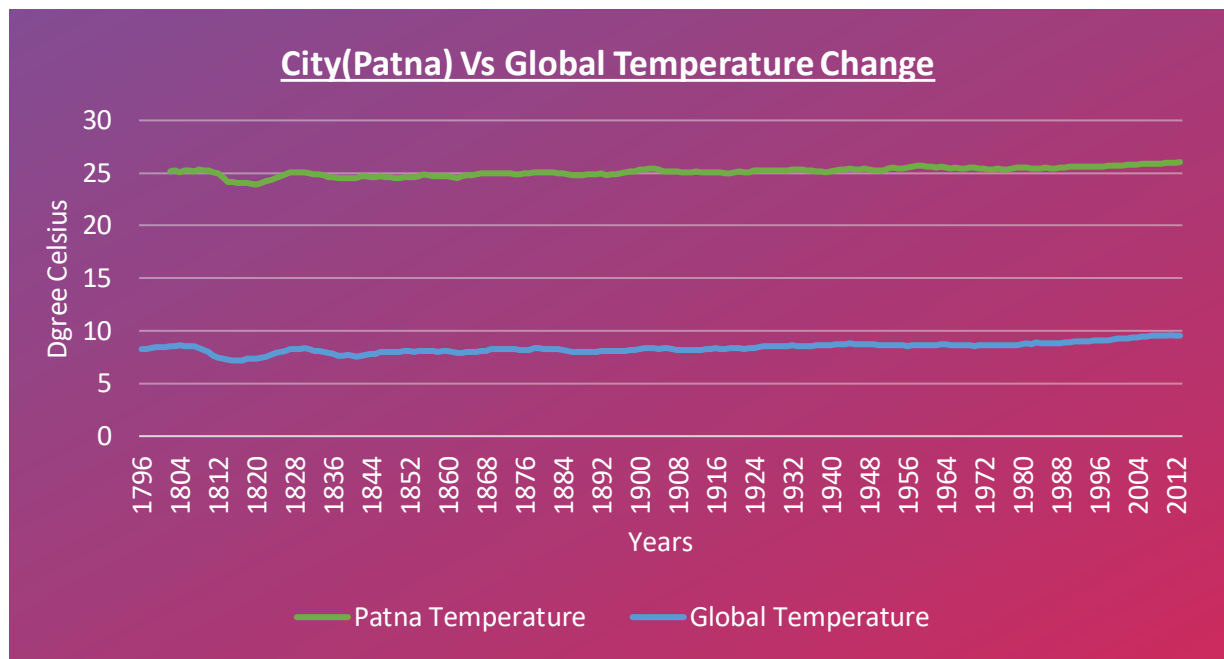
- Select the location where moving data is to be stored(E8).
- Apply average function in it
- Select the range D2:D8 (for 7 years moving avg.)
- Press enter (Moving_avg for 7 years will be generated)
- Now drag the function from cell(E8) down.



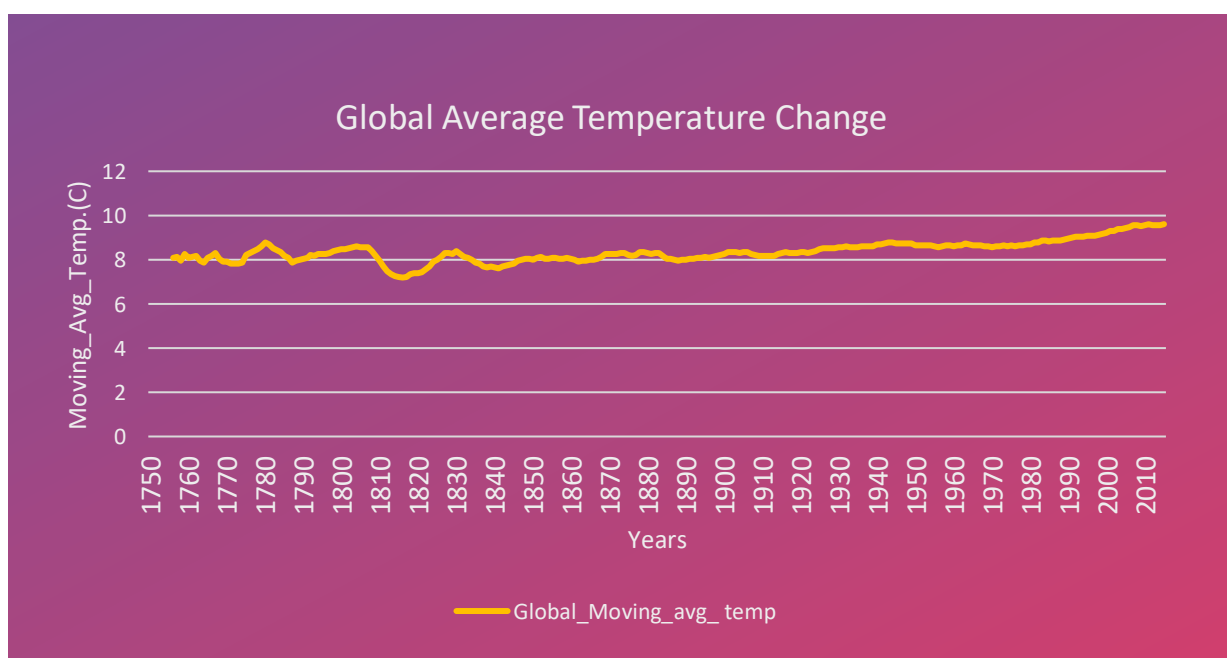
	A	B	C	D	E
1	year	city	country	avg_temp	moving_avg_temp
2	1796	Patna	India	24.99	
3	1797	Patna	India	26.49	
4	1798	Patna	India	24.27	
5	1799	Patna	India	25.25	
6	1800	Patna	India	25.2	
7	1801	Patna	India	24.19	
8	1802	Patna	India	25.4	25.14714286
9	1803	Patna	India	25.4	25.20571429
10	1804	Patna	India	25.72	25.09571429
11	1805	Patna	India	25.3	25.24285714
12	1806	Patna	India	25.21	25.23714286
13	1807	Patna	India	24.68	25.16428571

Data Visualization

Once the dataset was ready, then it was all about visuals. The first chart shows a comparison between the global and local temperature set across the years.



The second plot shows the global average temperature range across all the years to view a trend for the entire dataset.



What is correlation coefficient?

Correlation coefficients are used in statistics to measure how strong a relationship is between two variables.

There are several types of correlation coefficient formulas. One of the most commonly used formulas is Pearson's correlation coefficient formula:

$$r = \frac{n(\sum xy) - (\sum x)(\sum y)}{\sqrt{[n\sum x^2 - (\sum x)^2][n\sum y^2 - (\sum y)^2]}}$$

Observations

Similarities:-

- ✚ On the short term, both lines are volatile, but on the long term, both display a slow increase trend.
- ✚ Both graphs show increase in average temperature with time, which means earth is getting hotter.

Differences:-

- ✚ Local average temperature is observed to be hotter than the global average temperature.
- ✚ Global moving average temperature is increasing at faster rate in comparison to local moving average temperature.

Overall:-

- ✚ Patna's weather has been warmer than the global average in the past couple hundred years.
- ✚ A significant rise in the yearly average temperature can be observed in Patna in the past couple of decades.
- ✚ The yearly average temperature seems to be increasing abnormally on a global scale in the last 2-3 decades.

CONCLUSION:

✚ **World is getting hotter!!**