Exploring Weather Trends

WITH PYTHON

Saurabh Tyagi | Udacity Data Analyst Nanodegree| August-17. 2018

OVERVIEW

In this project, I have analyzed and compared b/w local yearly temperatures of Patna, India & Global Average temperatures. The data provided by Udacity portal from where I extracted, manipulated & visualized it by using the SQL, NUMPY AND PANDA.

GOALS

1. Extraction of data from the database and export to CSV file
2. Plotting charts based on extracted data
3. Deducing meaningful observations based upon the charts

TOOLS USED

1. SQL : To extract & manipulate data from the provided database
2. Python : For writing the source code used for calculating moving averages & plotting charts
3. Spyder : Cross platform IDE used for writing & compiling the code

STEPS INVOLVED-

1. Extraction of data - Having learned from basic sql earlier, I was able to extract, alter, and manipulate data from the provided database.

    - To see which cities are available for field “India” in the city_list table

      SELECT * FROM city_list WHERE country LIKE ‘India’

    - Altering The Tables- On a close examination of the given database schema, I found that both global_data & city_data had a common column “avg_temp”. I have changed the names of this column into two distinct names, which would help me in “JOINING” these two tables in the next step.

```
ALTER TABLE city_data RENAME COLUMN avg_temp to
city_avg_temp
```

```
ALTER TABLE global_data RENAME COLUMN avg_temp to
glob_avg_temp
```

- Joining the two tables- In order to have a single table I used the "JOIN" command.

```
SELECT global_data.year, global_data.global_avg_temp,
city_data.city_avg_temp

FROM global_data JOIN city_data

ON global_data.year = city_data.year
WHERE city LIKE 'Patna'
```

After this, I was able to download and save the results under the name of "joined_global_city_tables."

2. Source code for plotting the data – I had taken some free classes from Udacity & Datacamp in python (Numpy and Panda specifically). These libraries (along with Matplotlib) proved to be of immense help in the visualization of project data.

```
# IMPORTING THE NEEDED LIBRARIES

import numpy as py

import panda as pd

from matplotlib import pyplot as plt
```

#IMPORTING THE EXTRACTED DATA

```
data = pd.read_csv("joined_global_city_tables.csv")
```

#I HAVE DEFINED A FUNCTION CALLED MOVING_AVG THAT CALCULATES MOVING/ROLLING AVERAGES FOR A SMOOTHER GRAPH

```
def moving_avg(mA_range, data_input):

output = data_input.rolling(window = mA_range, center = false, on = "city_avg_temp").mean().dropna()

return output
```

#CALLING THE FUNCTION WITH A SUITABLE RANGE

```
mA_value = 190

chart_moving_avg = moving_avg(mA_value, data)
```

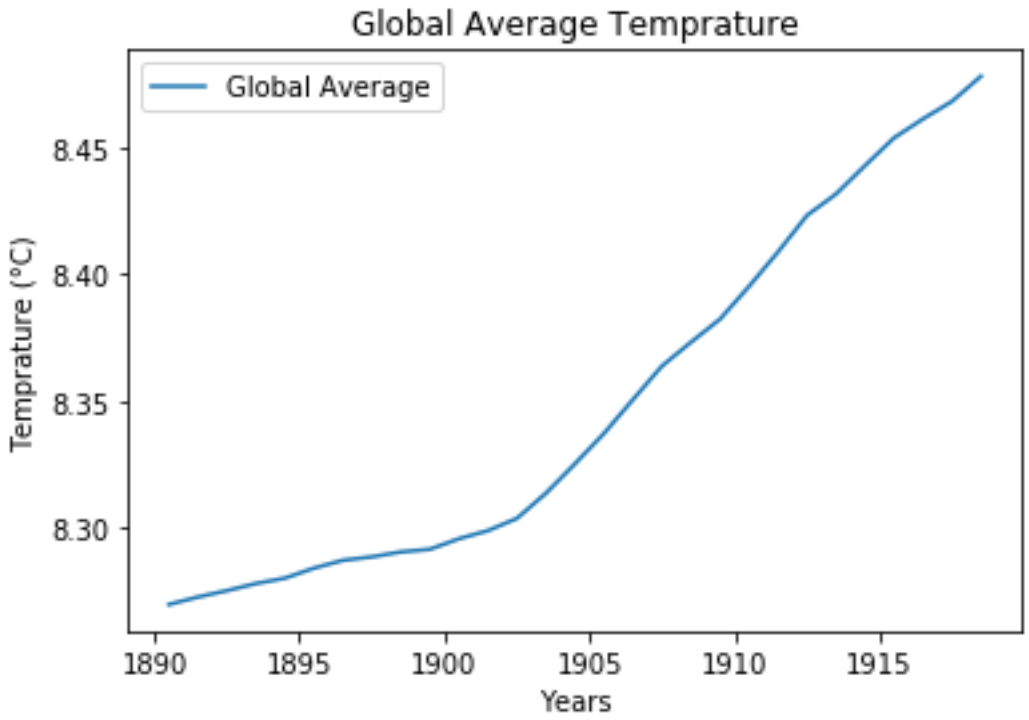
#PLOTTING THE GRAPH: GLOBAL TEMPRATURES

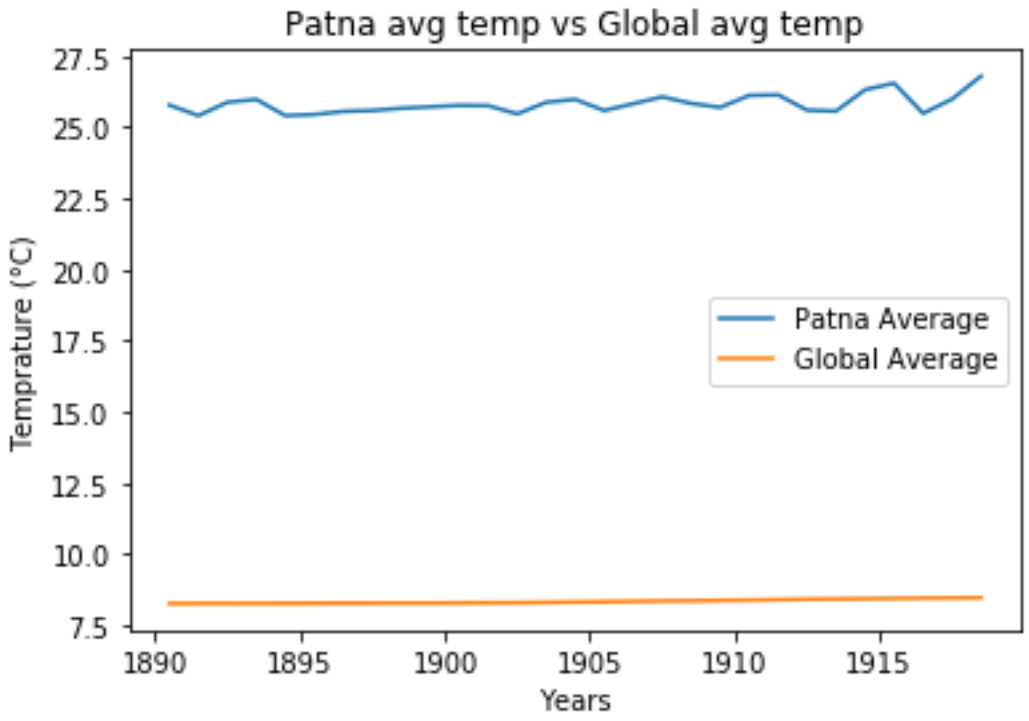
```
 plt.plot(chart_moving_avg ['year'], chart_moving_avg ['global_avg_temp'], label = 'Global')

 plt.legend()

 plt.xlabel ("Years")

 plt.ylabel ("Temperature (°C)")

plt.title ("Global Average Temperature")

 plt.show ()
```

#PLOTTING THE GRAPH: GLOBAL TEMPRATURES VS LOCAL TEMPRATURES

```
plt.plot(chart_moving_avg['year'], chart_moving_avg['city_avg_temp'], label = 'Patna Average')

plt.plot(chart_moving_avg['year'], chart_moving_avg['glob_avg_temp'], label='Global Average')

plt.legend()

plt.xlabel("Years")

plt.ylabel("Temprature (°C)")

plt.title("Patna avg temp vs Global avg temp" )

plt.show()
```

3. RESULTS AND OBSERVATIONS





#OBSERVATIONS

I have observed that if use 7 years or 50 years as the range of moving average, a lot of "noise" creeps into my graphs. Therefore, I have used a range of 190, which enabled me in getting much more meaningful and smoother graphs.

Although the "Industrial revolution" started somewhere around the 1840's, It did not picked up any meaningful improvements in machine technologies until the mid-1890's.

AND THAT'S EXACTLY WHERE THE FIRST GRAPH STARTS RISING

In addition, the rise in the graph is approximately 0.1 degree Celsius. This results in a total rise of 1.5 degree Celsius over the total period.

Reasons may include-

1- **Rise in greenhouse gases (emitted by industries & vehicles) which trap heat inside the earth's atmosphere**.

The second graph shows that local temperatures-

1- Have remained mostly consistent in the given range.

2- Started from an average of about 25.5 degree Celsius and have risen to an average of 26.8

3- The local temperatures are much hotter then global temperatures.

Reasons for the above are-

1- **Being a capital city of Bihar state, besides being a rapidly developing commercial hub, it is also home to several medium and large-scale industries. These industries started developing around 1920's and thus pollution started rising.**
This explains a rise of 1.3 degree Celsius over the years.

2- **Patna is enjoys a tropical climate zone and this gives rise to temperatures much hotter on an average then the global average temperature. This is because only a small part of continents resides between the two tropics and the rest of the world resides further away. This gives much lower average global temperatures overall.**

#CONCLUSIONS

From the above data, we can see that the average world temperatures are rising and the earth is getting hotter. This is also affecting the ocean currents, carbon dioxide concentrations in the ocean water, and weather patterns all around the world.

Local average yearly temperatures have also been affected by this phenomenon and a steep rise is seen in Patna's temperature in recent years.

#REFRENCES

Python tutorials – Courtesy of Udacity

Sql tutorials – Courtesy of Datacamp.com

Special thanks to – Udacity team who have always provided quality content to people all over the world.