Project – 1

Explore Weather Trends

April – 21

Udacity – Data Analyst Nanodegree program

Naman jain

Project-1 , Explore Weather Trends

Agra, INDIA

Overview of Project

This Project is all about Analyzing the temperature of different city or country according to the global Temperature. The all data regarding to temperature of different cities will be provided by the Udacity itself. So, in this project I have just analyze that data.

Goals of project

- **Extraction of data from the database by performing Some SQL Queries and to export CSV file.**
- **Visualization and making charts of extracted data.**
- **And , finally Observe that data.**

Tools and programming language Used

- **SQL – For Extraction of Data from Database.**
- **Python- For writing Code of Calculating moving Average and for plotting Different graphs.**
- **Anaconda - Jupyter notebook- Interface used for writing Python code and Visualize the Graphs.**
- **MS -Excel – To open Csv files and Have a look of Data.**
- **MS-Word – For writing Project report file.**

STEP – 1 EXTRACTION OF DATA FROM DATABASE

To extract the Data from provided Database on Udacity Portal itself. i write some SQL Queries to extract the data from Database. I will learn Join Basics relational database Basics from A online platform And used the learned concepts in our project.

1. **CHECKING THE CITIES OF INDIA WHOSE DATA IS AVAILABLE ON DATABASE.**

```
SELECT *
        FROM city_list
  WHERE country LIKE 'India'
```

2. ***Now I have to join the tables to find relevant data , but when joining the tables I got to know that from the Schema that the column 'avg_temp'* is same in both *city_data* and *global_data.* So , I change the Name of columns and give them a new Name.**

```
ALTER TABLE city_data RENAME COLUMN avg_temp to cityAT;
    -- cityAT = city Average Temp.
ALTER TABLE global_data RENAME COLUMN avg_temp to GlobalAT;
    -- GlobalAT = global Average temperature
```

3. Now I Join the Tables by performing the SQL Query given below

```
SELECT global_data.year,  global_data.GlobalAT,  city_data.cityAT
FROM global_data JOIN city_data
ON global_data.year = city_data.year
WHERE city LIKE 'Agra';
```

4. After Performing all queries I download the CSV file with three columns .

Step – 2 Writing Code of Python for Making Line Chart

In [2]:
```
# Importing the Libraries

import numpy as np
import pandas as pd                    # Loading data on Notebook
from matplotlib import pyplot as plt   # For Drawing chart
```

In [4]:
```
# importing the  data which i download in form of csv file by performing some sql queries
data = pd.read_csv("results_main.csv")
```

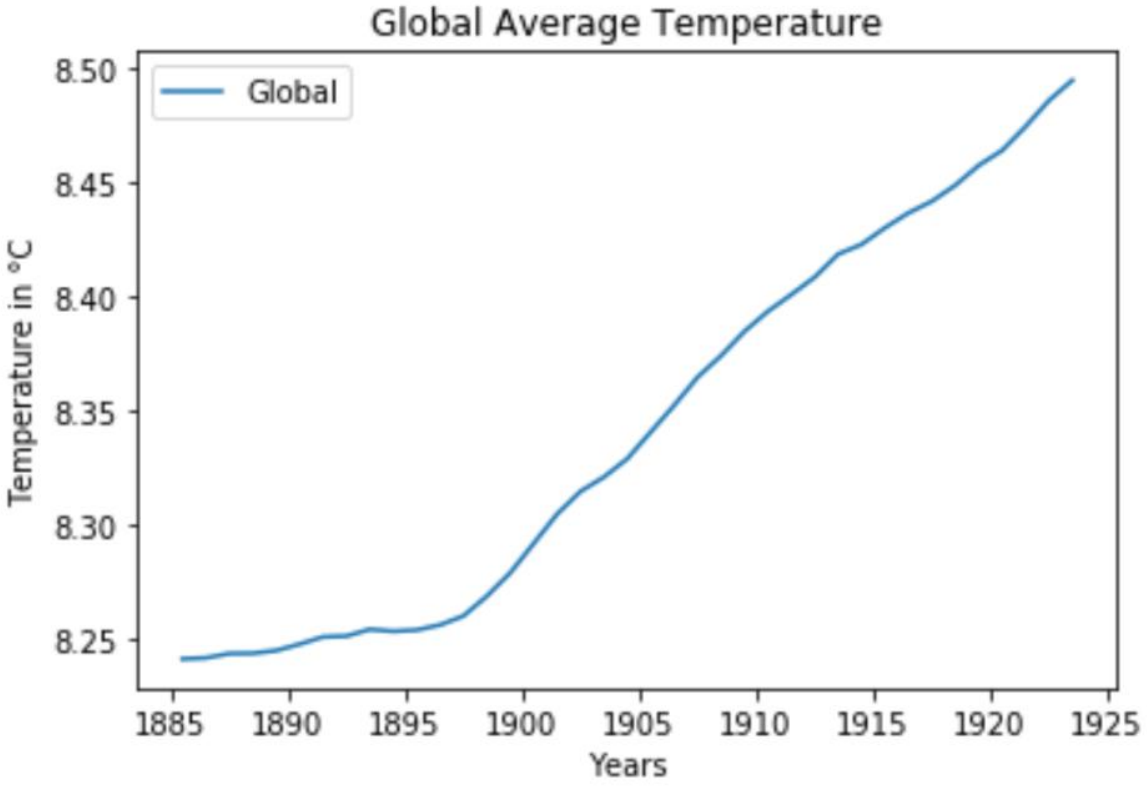
CALCULATION OF MOVING AVERAGE (PYTHON FUNCTION)

In [7]:
```
# Function that claculates moving Average

def moving_avgerage(movingAverage_range, data_input):   # with local variables
    output = data_input.rolling(window = movingAverage_range, center = False, on = "cityat").mean().dropna()
    return output

#Calling the function with the range of Moving Average

movingAverage_value = 180

chartOf_moving_avg = moving_avgerage(movingAverage_value, data)     # with global variables
```

DRAWING GRAPH OF GLOBAL TEMPERATURE

In [8]:
```
# Drawing the graph of  Global Temperature

plt.plot(chartOf_moving_avg['year'], chartOf_moving_avg['globalat'], label='Global')
plt.legend()
plt.xlabel("Years")
plt.ylabel("Temperature in °C")
plt.title("Global Average Temperature")
plt.show()
```

OUTPUT : GLOBAL TEMPERATURE CHART

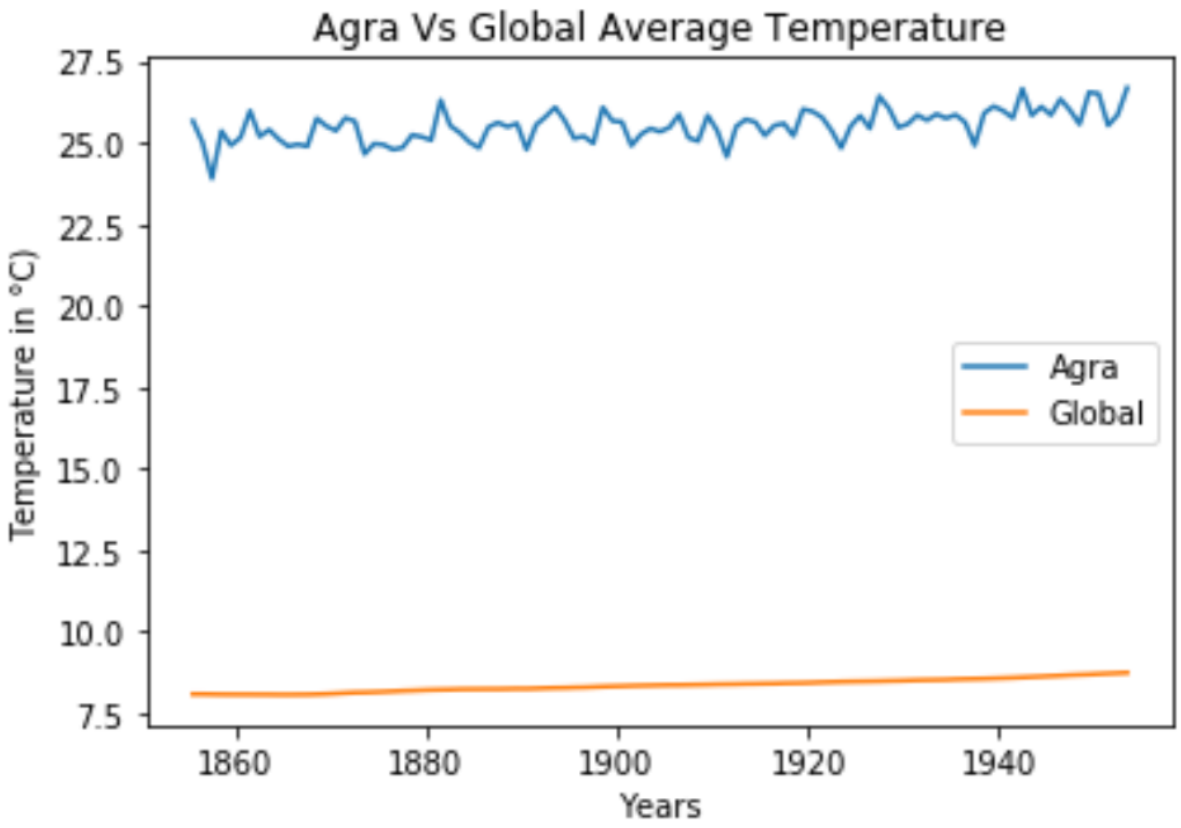


THIS IS THE SEPARATE GRAPH OF GLOBAL AVERAGE TEMPERATURE RISES ALONG WITH THE YEARS.

PYTHON CODE FOR DRAW GRAPH B'W AGRA AND GLOBAL AVERAGE TEMPERATURE

In [9]:
```
# Drawing the graph Temperature of Agra and Global temperature

plt.plot(chartOf_moving_avg['year'], chartOf_moving_avg['cityat'], label='Agra')
plt.plot(chartOf_moving_avg['year'], chartOf_moving_avg['globalat'], label='Global')
plt.legend()
plt.xlabel("Years")
plt.ylabel("Temperature in °C)")
plt.title("Agra Vs Global Average Temperature")
plt.show()
```

NOW ANALYZING WITH THE GRAPH B'W AGRA AND GLOBAL AVERAGE TEMPERATURE



Note : All code are related to previous code shell So, please execute these line in order from top to bottom.

Result : Observations

I observe that if I take moving average range value small(like 10,20,..50) then my graph is not that much clear it his little bit hard to analyze that graph so I take large moving range that is 120. So by giving this moving range value my graph will now relatively smooth.

Observation from the Charts

1. So if u see the global average temperature graph from the year 1900 its almost increasing constantly by 0.1 degree Centigrade that's why it form a inclined graph that's why to show these things I make a separate graph for global average temperature .

2. From the year 1883 the global average temperature suddenly change and starts decreasing that's why the graph is declined till just before year 1900 after that temperature increase constantly.

3. Then I analyze Second graph that is graph between Agra and Global average temperature .then I got to know that there is huge difference in temperature between agra and Global Average Temperature.

4. By Analyzing second Graph , I can say that Agra is Seems to be hotter than any other cool place in the world. if u further deep study about the temperature of Agra and find the reason by it is hotter , it is because Agra is surrounded by industrial area like "Sikandra footwear industrial Area" this place have many industries who made footwear in large amount.Agra is known to be the biggest industrial hubs in India. This also may be the reason of higher temperatures locally.

FINDING COEFFICIENT OF CORRELATION BETWEEN AGRA AND GLOBAL AVERAGE TEMPERATURE :

SO FOR FINDING COEFFICIENT OF CORRELATION I TAKE LAST 10 YEARS DATA USING PYTHON CODE GIVEN BELOW:

```
In [4]: data.tail(10)
```

Out[4]:

	year	globalat	cityat
208	2004	9.32	26.11
209	2005	9.70	25.85
210	2006	9.53	26.34
211	2007	9.73	26.00
212	2008	9.43	25.57
213	2009	9.51	26.55
214	2010	9.70	26.51
215	2011	9.52	25.53
216	2012	9.51	25.86
217	2013	9.61	26.69

LET ,

X = GLOBAL AVERAGE TEMPERATURE

Y = CITY(AGRA) AVERAGE TEMPERATURE

Let X = Globalat = Global Average Temperature
Y = cityat = city Average Temperature

	X	Y	$U = X - \bar{X}$	$V = Y - \bar{Y}$	U^2	V^2	UV
1	9.32	26.11	-0.23	0.01	0.05	0	0
2	9.70	25.85	0.15	-0.25	0.02	0.06	-0.03
3	9.53	26.34	-0.02	0.23	0	0.05	0
4	9.73	26.00	0.18	-0.1	0.03	0.01	-0.01
5	9.43	25.57	-0.12	-0.53	0.01	0.28	0.06
6	9.51	26.55	-0.04	0.45	0	0.20	-0.01
7	9.70	26.51	0.15	0.41	0.02	0.16	0.06
8	9.52	25.53	-0.03	-0.57	0	0.32	0.01
9	9.51	25.86	-0.04	-0.24	0	0.05	0
10	9.61	26.69	0.06	0.59	0	0.34	0.03
Total	95.56	261.01	0	0	0.13	1.75	0.11

Now $\bar{X} = \frac{1}{n}\Sigma X = \frac{95.56}{10} = 9.556$

$\bar{Y} = \frac{1}{n}\Sigma Y = \frac{261.01}{10} = 26.101$

Now $\bar{U} = \frac{1}{n}\Sigma U = 0 \quad [\because \Sigma U = 0]$

$\bar{V} = \frac{1}{n}\Sigma V = 0 \quad [\because \Sigma V = 0]$

$Cov(U,V) = \frac{1}{n}\Sigma UV - \bar{U}\bar{V}$

$= \frac{1}{10} \times 0.11$

$Cov(U,V) = 0.01$

Now $\sigma u^2 = \frac{1}{n}\sum u^2 - (\bar{u})^2$

$= \frac{1}{10} \times 0.13 - 0$

$= 0.013$

$\sigma v^2 = \frac{1}{n}\sum v^2 - (\bar{v})^2$

$= \frac{1}{10} \times 1.35 - 0$

$= 0.135$

Now correlation coefficient b/w u and v

$r(u,v) = \frac{cov(u,v)}{\sigma u \, \sigma v} = \frac{0.01}{\sqrt{0.013 \times 0.135}}$

$= \frac{0.01}{0.04} = 0.25$

∴ correlation coefficient is change of scale of origin

So

$\boxed{r(x,y) = 0.25}$

NOW,

COV(X,Y) = 0.01

AND FINALLY COEFFICIENT OF CORRELATION B'W X AND Y IS

r(X,Y) = 0.25

NOW ,

WE ESTIMATE THE AVERAGE TEMPERATURE IN AGRA BASED ON THE AVERAGE GLOBAL TEMPERATURE.

USING LINE OF REGRESSION OF Y ON X :

$$Y - \bar{Y} = r(X,Y)\frac{\sigma Y}{\sigma X}(X - \bar{X})$$

NOW ,

$\hat{Y}$ is Estimate value of Y

SO WE NOW SOLVE LINE OF REGRESSION OF Y ON X MANUALLY

Now

Eqn of line of regression of y on x is

$$y - \bar{y} = r(x,y) \frac{\sigma_y}{\sigma_x}(x - \bar{x})$$

$$= y - 26.101 = 0.25 \times \frac{0.41}{0.11}(x - 9.556)$$

$$= \boxed{y = 0.93x + 17.19}$$

$\hat{y}$ is Estimate value of y

So, To estimate average temperature of your city in my case (Agra) base on the Avege global temperature we use this eqn.

for example = Take year 2013.

$$\hat{y} = 0.93x + 17.19$$

Now put x value from global average temperature column of year 2013.

$$\boxed{x = 9.61}$$

$$\hat{y} = 0.93 \times 9.61 + 17.19$$

$$\boxed{\hat{y} = 26.12}$$ → Estimate value of y

So ,

$$\hat{Y} = 0.93X + 17.19$$

NOW PUTTING THE AVERAGE GLOBAL TEMPERATURE VALUE AT THE PLACE OF X , YOU CAN ESTIMATE THE AVERAGE TEMPERATURE OF CITY BASED ON GLOBAL TEMPERATURE .

********** THIS ALL IS MY OBSERVATION **********

KEY CONSIDERATION :

- YEARS IS VISIBLE ON X-AXIS
- TEMPERATURE IS SHOWN ON Y-AXIS
- FOR VISUALIZATION USE OF MATPLOTLIB
- DEFINE A FUNCTION TO MOVING AVERAGE FUNCTION
- USE OF JUPYTER NOTEBOOK FOR CODING
- FINDING OF COEFFICIENT OF CORRELATION BETWEEN CITY AVERAGE TEMPERATURE AND GLOBAL AVERAGE TEMPERATURE
- ESTIMATE THE AVERAGE TEMPERATURE OF CITY BASED ON GLOBAL AVERAGE TEMPERATURE

REFERENCES :

- **How to calculate moving average in python :** https://www.learndatasci.com/tutorials/python-finance-part-3-moving-average-trading-strategy/

- **Joining Tables in Sql :** https://www.w3schools.com/SQL/sql_join.asp

- **Rolling() Parameters :** https://pandas.pydata.org/pandas-docs/stable/reference/api/pandas.DataFrame.rolling.html

- **Front Page Image :** https://www.gettyimages.in/photos/taj-mahal?mediatype=photography&phrase=taj%20mahal&sort=mostpopular