

UDACITY Data Analysis Nanodegree

Project 01

Exploring Weather Trends

Goals:

- Extract data from the database
- Export to CSV file
- Chart visualization based on extracted data
- Observation-based on chart

Tools Used:

- SQL: To extract the data from the database
- Python: For reading the CSV file and plotting a line chart
- Microsoft Excel: For plotting the line chart
- Google Sheets: For preparing the report along with the observations

STEP 1 - Extraction of Data from provided Database:

SQL Query was used for downloading CSV Files.

Data about global temperature has been imported by using the query:-

1. global_data.csv

```
SELECT * FROM global_data;
```

Similarly for the city of Agra, following SQL Query was used.

A CSV file with year, city_name = 'Agra', country_name = 'India', avg_temp was imported.

2. city_data.csv

```
SELECT * FROM city_data  
WHERE city = 'Agra' AND country = 'India';
```

STEP 2 - Moving Averages

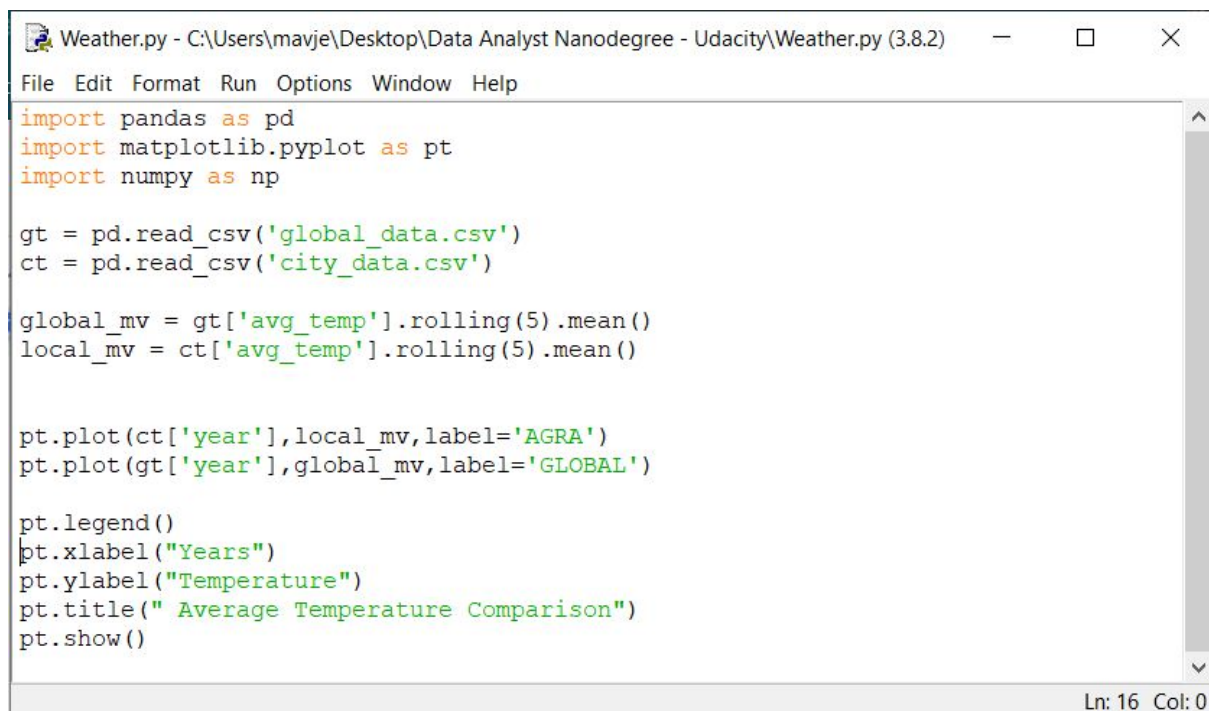
The data obtained was smoothened out using Rolling Average and to make it easier to observe the trends when it is shown in Charts.

Python was used for calculating the Moving Average Using built-in Functions such as 'rolling and 'mean')

```
global_mv = gt['avg_temp'].rolling(5).mean()
local_mv = ct['avg_temp'].rolling(5).mean()
```

STEP 3 - Reading and Plotting the Chart in Python

Following python code was used to read the CSV file using the Python Pandas library and Matplotlib was used to plot the following chart of Average Temperature Comparison.

A screenshot of a Python script in a text editor window. The window title is "Weather.py - C:\Users\mavje\Desktop\Data Analyst Nanodegree - Udacity\Weather.py (3.8.2)". The menu bar includes "File", "Edit", "Format", "Run", "Options", "Window", and "Help". The code is as follows:

```
import pandas as pd
import matplotlib.pyplot as plt
import numpy as np

gt = pd.read_csv('global_data.csv')
ct = pd.read_csv('city_data.csv')

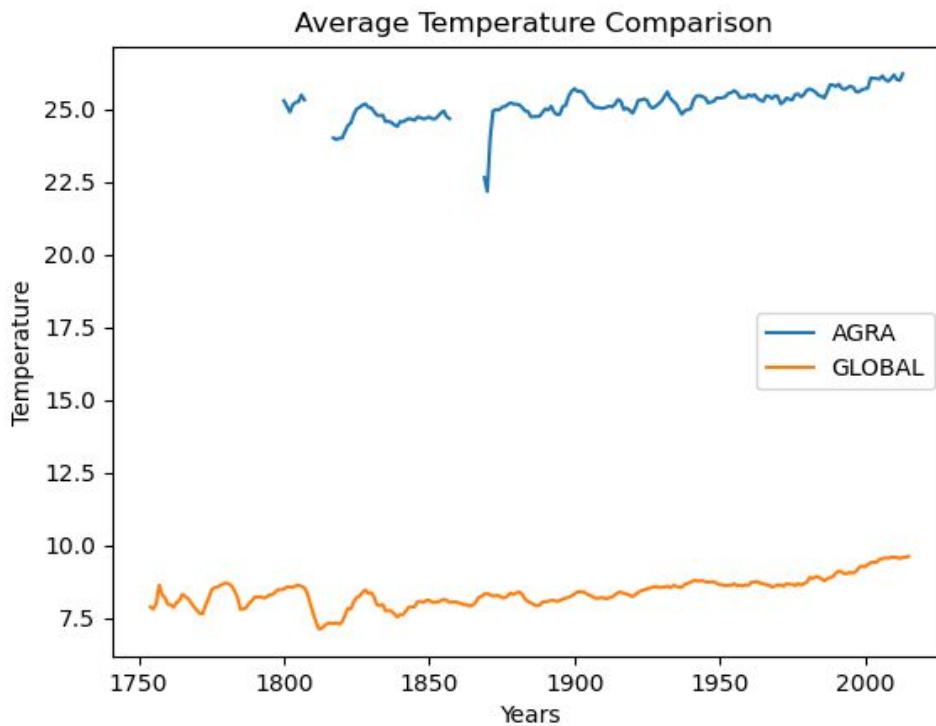
global_mv = gt['avg_temp'].rolling(5).mean()
local_mv = ct['avg_temp'].rolling(5).mean()

plt.plot(ct['year'], local_mv, label='AGRA')
plt.plot(gt['year'], global_mv, label='GLOBAL')

plt.legend()
plt.xlabel("Years")
plt.ylabel("Temperature")
plt.title(" Average Temperature Comparison")
plt.show()
```

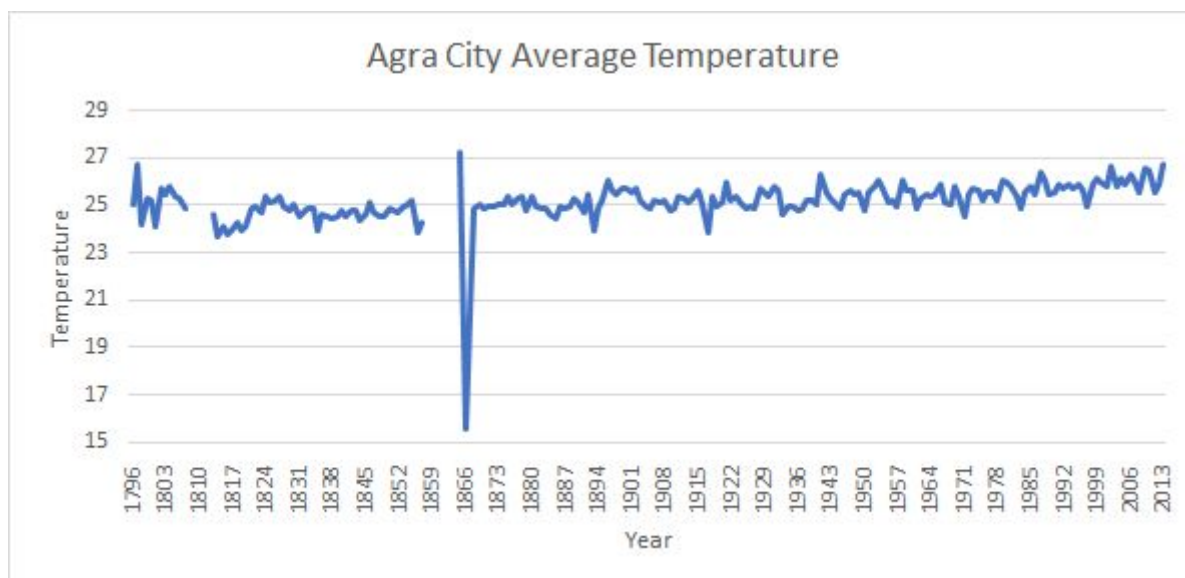
The status bar at the bottom right shows "Ln: 16 Col: 0".

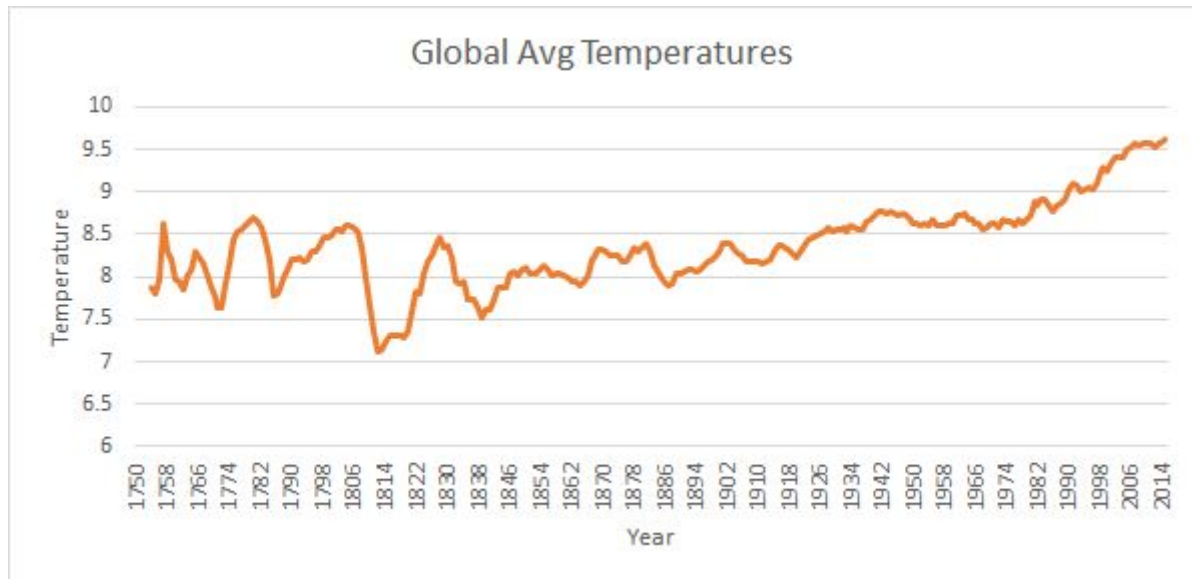
Chart of the above data:



STEP 4 - Plotting the datasets imported

The following charts of the city_data.csv and global_data.csv using Microsoft Excel





Observations:

1. From the above charts, we can notice that the Global Average Temperature is constantly rising since the 1800s.
2. The difference in temperature between the Globe and Agra city is drastic.
3. In the city of Agra, the average temperature has increased by approximately 2°C whereas globally it has increased by approximately 1.5°C.
4. The average temperature for Agra has constantly increased since 1866.

Conclusion:

All the Observations and Charts suggest accurately that the global temperature is rising over years leading to global warming.