

**Exploring Weather Trends**

**Jeffrey Smith**

***05/06/2020***

**Project Overview**

The goal of this project is to analyze local and global temperature data and compare the temperature trends from my local area, Columbus, Ohio, to the overall global temperature trends. This analysis will first identify specific project objectives and discuss the various tools used throughout. Next, this analysis will explore the various global and city specific weather data made available for this project. We will then analyze annual moving averages of both the local and global temperatures. This project will conclude with a discussion of observations, highlighting differences between the localized and global analysis.

**Project Objectives**

* Export temperature data from Udacity's database
* Create visualizations comparing local and global temperatures
* Make observations of similarities and differences in local and global temperature averages

**Tools Used:**

* **Python 3.7.4 64-bit**: Calculations and visualizations
* **SQL**: Data extraction

**Data Acquisition**

The data acquired for this project was provided by Udacity through a SQL Workspace. This database provides three distinct tables.

* **city\_list** – Table containing a list of cities and countries in the database.
* **city\_data** - Table contains the average temperatures for each city by year (C°).
* **global\_data** – Table contains the average global temperatures by year (C°).

Using the temperatures database and the three tables provided, we executed the following queries to obtain the data required for this project.

**Global Data Query**

**SELECT** \*

**FROM** global\_data

**City Data Query**

**SELECT** \*

**FROM** city\_data

**WHERE** country = 'United States' **AND** city= 'Columbus';

**Data Exploration**

Data exploration began with the ingestion of both the global and local data sets. This exploration first focused on the global data. Table 1 provides a snippet of the global dataset, Table 2 presents information about the full dataset, and Table 3 denotes summary information.

**Table 1: Global Average Temperature; 5 rows**

|  |  |  |
| --- | --- | --- |
| Index | year | Global Average Temperature |
| 0 | 1750 | 8.72 |
| 1 | 1751 | 7.98 |
| 2 | 1752 | 5.78 |
| 3 | 1753 | 8.39 |
| 4 | 1754 | 8.47 |

**Table 2: Global Average Temperature Data Information**

|  |  |  |  |
| --- | --- | --- | --- |
| <class 'pandas.core.frame.DataFrame'> | | | |
| RangeIndex: 266 entries, 0 to 265 | | | |
| Data columns (total 2 columns): | | | |
| # | Column | Non-Null Count | Dtype |
| 0 | year | 266 non-null | int64 |
| 1 | GAT | 266 non-null | float64 |
| dtypes: float64(1), int64(1) | | | |
| memory usage: 4.3 KB | | | |

**Table 3: Global Data Description**

|  |  |  |
| --- | --- | --- |
| Global Data Description: | | |
|  | year | GAT |
| count | 266.000000 | 266.000000 |
| mean | 1882.500000 | 8.369474 |
| std | 76.931788 | 0.584747 |
| min | 1750.000000 | 5.780000 |
| 25% | 1816.250000 | 8.082500 |
| 50% | 1882.500000 | 8.375000 |
| 75% | 1948.750000 | 8.707500 |
| max | 2015.000000 | 9.830000 |

Examining the global data set with these descriptive statistics, we find that there are 266 temperatures listed between the years of 1750 and 2015. Additionally, this dataset contains no missing values.

Table 4, Table 5, and Table 6 provide the same information and descriptive statistics for the local dataset.

**Table 4: First Five Rows, Local Avg Temp**

|  |  |  |
| --- | --- | --- |
| Index | year | Local Average Temperature |
| 0 | 1743 | 7.46 |
| 1 | 1744 | 15.73 |
| 2 | 1745 | 6.91 |
| 3 | 1746 | NaN |
| 4 | 1747 | NaN |

**Table 5: Local Average Temperature Data Information**

|  |  |  |  |
| --- | --- | --- | --- |
| <class 'pandas.core.frame.DataFrame'> | | | |
| RangeIndex: 271 entries, 0 to 270 | | | |
| Data columns (total 2 columns): | | | |
| # | Column | Non-Null Count | Dtype |
| 0 | year | 271 non-null | int64 |
| 1 | GAT | 266 non-null | float64 |
| dtypes: float64(1), int64(1) | | | |
| memory usage: 4.4 KB | | | |

**Table 6: Local Data Description**

|  |  |  |
| --- | --- | --- |
| Global Data Description: | | |
|  | year | LAT |
| count | 271.000000 | 266.000000 |
| mean | 1878.000000 | 13.941090 |
| std | 78.375166 | 1.107914 |
| min | 1743.000000 | 6.060000 |
| 25% | 1810.500000 | 13.582500 |
| 50% | 1878.000000 | 14.070000 |
| 75% | 1945.500000 | 14.450000 |
| max | 2013.000000 | 16.050000 |

Examining the local data, we find that there are 271 entries of which 266 are non-null. We also note that the data set spans the years of 1743 to 2013 which differs from the global data set’s year span of 1750 to 2015. To preprocess the local data, records with null average temperature were removed. The remainder of the analysis utilizes a statistical analysis termed moving average.

**Moving Averages Analysis**

Moving average is a trend-following technical analysis method based on past data. This method of averaging helps to smooth out data allowing for a more intuitive interpretation, especially in terms of time-series data. Equation 1 denotes the Moving Average equation.

*where n is the moving average step size*

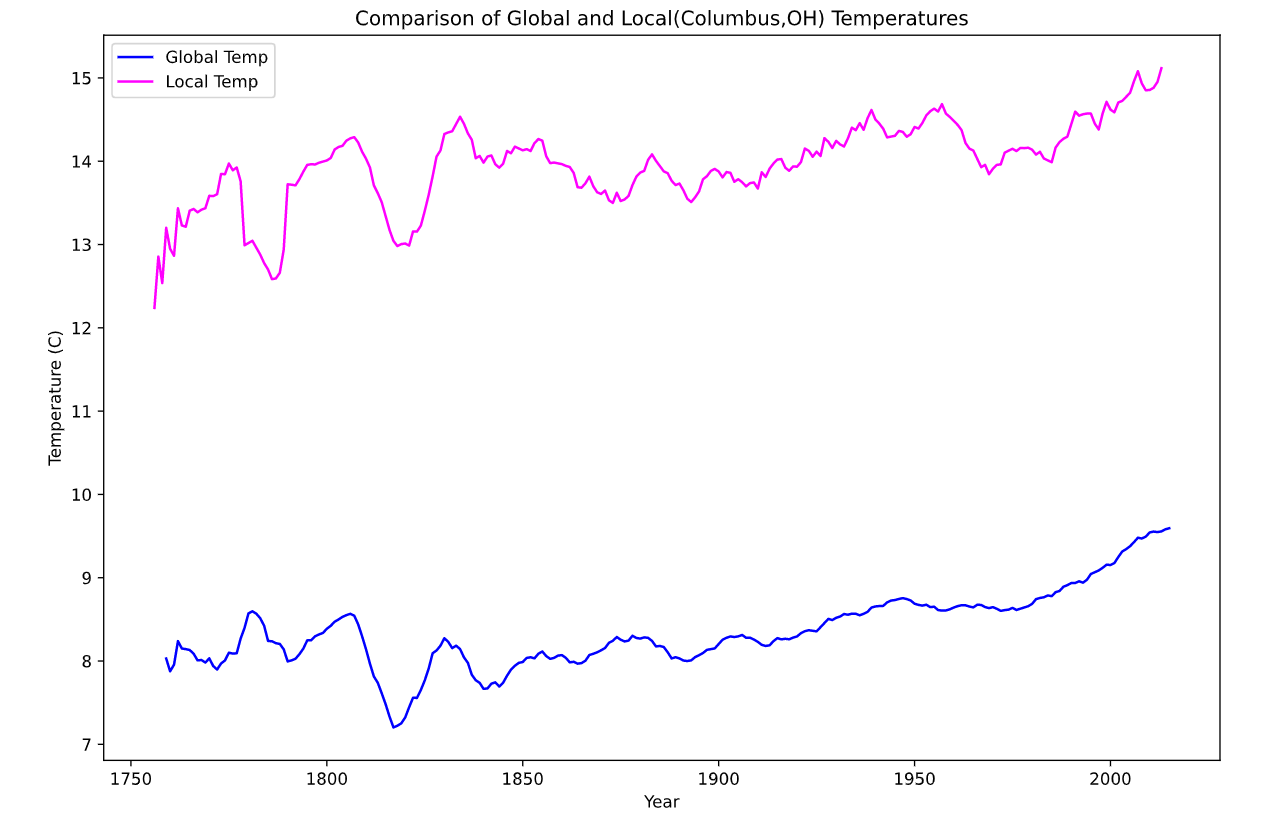
For this analysis we will implement a 10-year moving average on both the global and local data. To implement this average, we will create a new column in both datasets for the moving average and implementing pandas rolling method and adding the results as a new column to both datasets. Table 7 illustrates the first 12 rows of the global data with the inclusion of the moving average data.

**Table 7: Global Moving Average Temperature; 12 rows**

|  |  |  |  |
| --- | --- | --- | --- |
| index | year | Global Average Temperature | Global Moving Avgerage |
| 0 | 1750 | 8.72 | NaN |
| 1 | 1751 | 7.98 | NaN |
| 2 | 1752 | 5.78 | NaN |
| 3 | 1753 | 8.39 | NaN |
| 4 | 1754 | 8.47 | NaN |
| 5 | 1755 | 8.36 | NaN |
| 6 | 1756 | 8.85 | NaN |
| 7 | 1757 | 9.02 | NaN |
| 8 | 1758 | 6.74 | NaN |
| 9 | 1759 | 7.99 | 8.030 |
| 10 | 1760 | 7.19 | 7.877 |
| 11 | 1761 | 8.77 | 7.956 |

**Data Visualization**

Figure 1, depicts a line plot of the moving average of both the global and local data.



**Figure 1: Global and Local 10-year Moving Averages**

These moving averages indicate a significant visual difference between the global and average temperature that Columbus, OH has realized over the centuries.

**Observations**

Findings that can be observed in Figure 1 include the following:

* On average, the temperature in Columbus, OH is higher than the global temperature.
* Both the global and local temperatures follow a similar movement pattern over time.
* Although there is a visual difference in the temperatures between the global and local temperatures, they have both followed a similar trend including a steady increase of the past decade.
* It appears that in the early 1800s there was a global decrease in the temperature.
  + Further investigation into this decline in temperature indicates that in 1816 severe weather abnormalities caused a decrease in global temperatures. More information can be found [here](https://en.wikipedia.org/wiki/Year_Without_a_Summer).

**Conclusion**

Analyzing the moving average plot, we find that the temperatures in Columbus, OH are significantly warmer than the global temperature average. We were also able to pinpoint a specific temperature anomaly and conduct further research to uncover more information on its cause. Additionally, we were able to identify a gradual increase in temperatures across both the global and local data from around 1820 and continuing into today.