UDACITY

Data Analyst Nanodegree
Project 1 – Explore Weather Trends

Outline of Steps

1. **Extracting the data**
   The data was queried using SQL from the database on the project page of Udacity with the following queries and exported to CSV files.

   #Query 1 (for extracting Toronto city data)

   ```
   SELECT *
   FROM city_data
   WHERE city = 'Toronto';
   ```

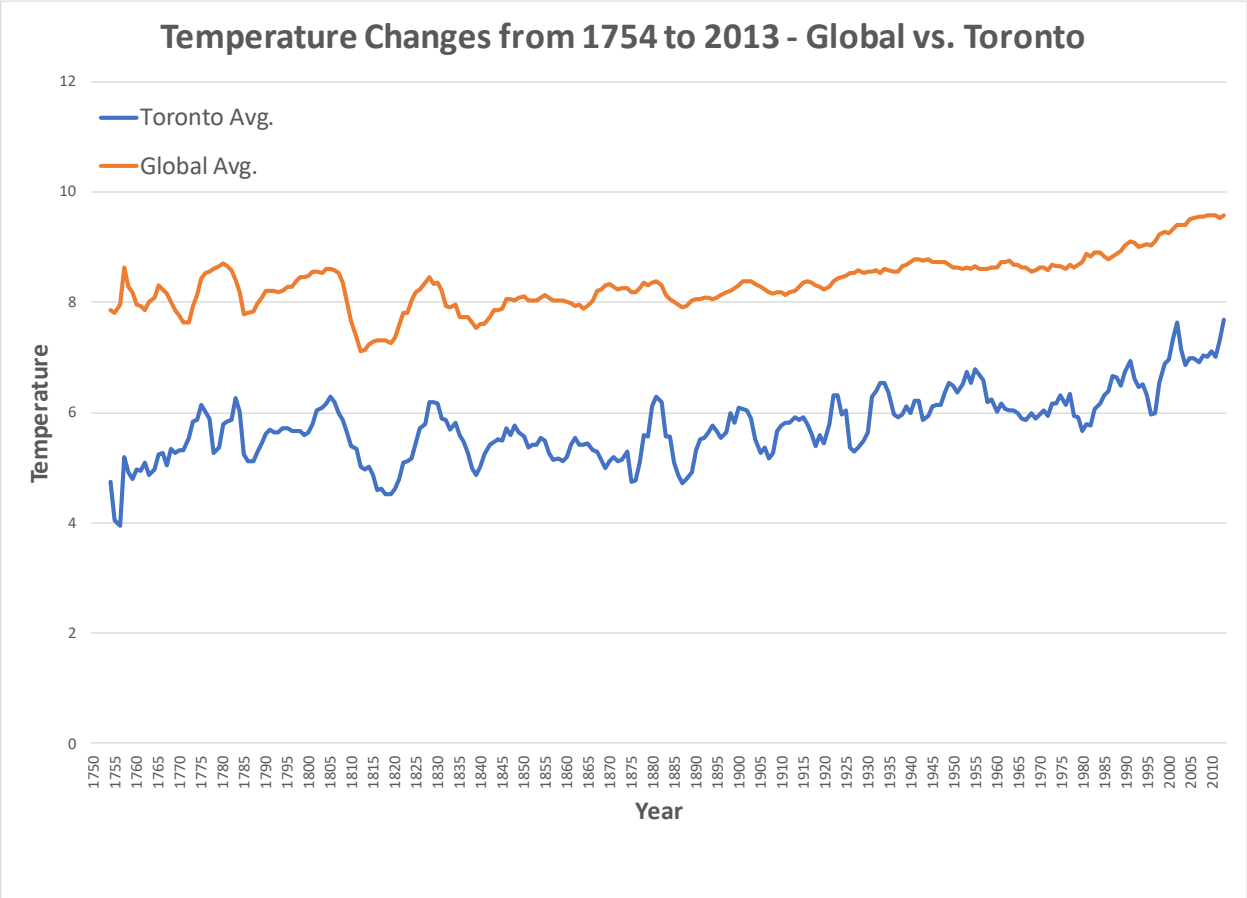
   #Query 2 (for extracting Global data)

   ```
   SELECT *
   FROM global_data;
   ```

2. **Data Combining and Cleaning**
   The tool used for this process was Microsoft Excel. The two CSV files were then combined into one Excel file. The following steps were taken to ensure data integrity:
   a. The years available in the two sets of queried data were checked against each other. It was found that in the data for Toronto city, the years 1743-1749 which was not available for the global data. Also, the years 2014 and 2015 were available in the global temperature data but not for Toronto. These were removed from the respective sets to ensure comparable sets.
   b. It was also checked whether all the remaining years were available by copy-pasting the year data side by side from 1750 and then using Excel arithmetic operator formula (subtracting the two side-by-side cells) to check that they all match. It was found that all the years from 1750 till 2013 were represented for Toronto and Global temperature datasets.
   c. These were then combined into one Excel spreadsheet and then cleaned for formatting. It was also confirmed there are all 264 data points are available for all years from 1750 till 2013 for Toronto and Global datasets. This formed the basis of the visualization.

3. **Analysis and Visualization**
   The tool used for this process was Microsoft Excel. The base data for Toronto and Global temperatures were used to calculate a 5-year moving average (which started from the year 1754 and ended in 2013). These moving average data points from 1754 to 2013 was used to create a line chart that plotted a line each for Toronto data and Global data. The plot is given below.

Line chart



Observations from the line chart

1. On average, the Toronto temperatures are about 2-2.5 degrees less than the global average temperature, meaning that it is cooler in Toronto.
2. The temperatures in Toronto show a greater fluctuation than the global average temperature. Especially, over the last 110 years or so, the global average temperature shows far lower fluctuation compared to the Toronto average temperatures.
3. There was a significant dip in Global temperatures starting in 1805 for about 10 years, leading to a fall in temperature of nearly 1.4 degrees. This led to a steeper fall in Toronto temperatures by almost 1.8 degrees, which took more time to recover from.
4. The general trend of global temperatures over the last century has been an increasing one, especially since 1890. Global temperatures have risen by nearly 1.5 degrees (8.05 in 1890 to 9.57 in 2013). Toronto has mostly followed this trend with fluctuations in between.
5. The Pearson correlation coefficient for the Global and Toronto datasets is 0.825, indicating a strong positive correlation between temperatures in Toronto and the Global average.