UDACITY – The School of Data Science

- Data Analyst -

Project 1: Exploring Weather Trends

Created by: Zakariya Boutayeb
Email: Boutayebz@gmail.com

Project Objective:

In the project of exploring weather trends from 1750 to 2013, local and global temperature data will be analyzed and then the temperature trends of New York will be compared to the overall global temperature trends.

Data Source:

Database include 3 tables:

- city_list table contains a list of cities and countries.

- city_data table contains a list of cities and countries as well as the yearly local average temperature.

- global_data table contains the yearly average global temperature.

Tools:

The primary tools used are SQL and Excel. SQL will allow us to explore and extract the data from the database, whereas Excel will be used to import the data from SQL in order to calculate the moving average of temperature and then plot the line chart with local and global temperature trends.

Outline:

Step 1: Check if the city of physical residence is available in the database.

Step 2: Extract the global and local temperature data from the database.

Step 3: Write a SQL query that allows us to make comparison between both temperature trends.

Step 4: Import data to Excel and use the Avg function to calculate the 10-years MA.

Steps 5: Plot Line chart with local and global temperature trends.

Step 6: Observations

Step 1: To check whether New York in the database, we can write a SQL query that displays the cities beginning by "N" in USA:

```
SELECT *
FROM city_list
WHERE country LIKE 'U%' AND city LIKE 'N%';
```

Step 2: Extract the global and local temperature data from the database.

1. Write a SQL query to extract NY data for the 100-years in the city_data table in descending order:

```
SELECT city_data.*
FROM city_data
WHERE city LIKE 'New York'
ORDER BY city_data.year DESC
LIMIT 100;
```

2. Write a SQL query to extract the global data:

```
SELECT global_data.*
FROM global_data
ORDER BY global_data.year DESC
LIMIT 100;
```

3. Write a SQL query that make adjustments to both global_data and city_data tables:

```
ALTER TABLE global_data
RENAME avg_temp to global_avg_temp;
```

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

Step 3: Write a SQL query that can make comparison between global avg temp and city avg temp:

```
SELECT global_data.year, global_avg_temp, city_avg_temp
FROM global_data
JOIN city_data
ON global_data.year = city_data.year
WHERE city LIKE 'New York'
ORDER BY city_data.year DESC
LIMIT 100;
```

Step 4: Import data to **Excel** and use the **Avg function** or **Moving Average** in **Data Analysis** in **Data** menu:

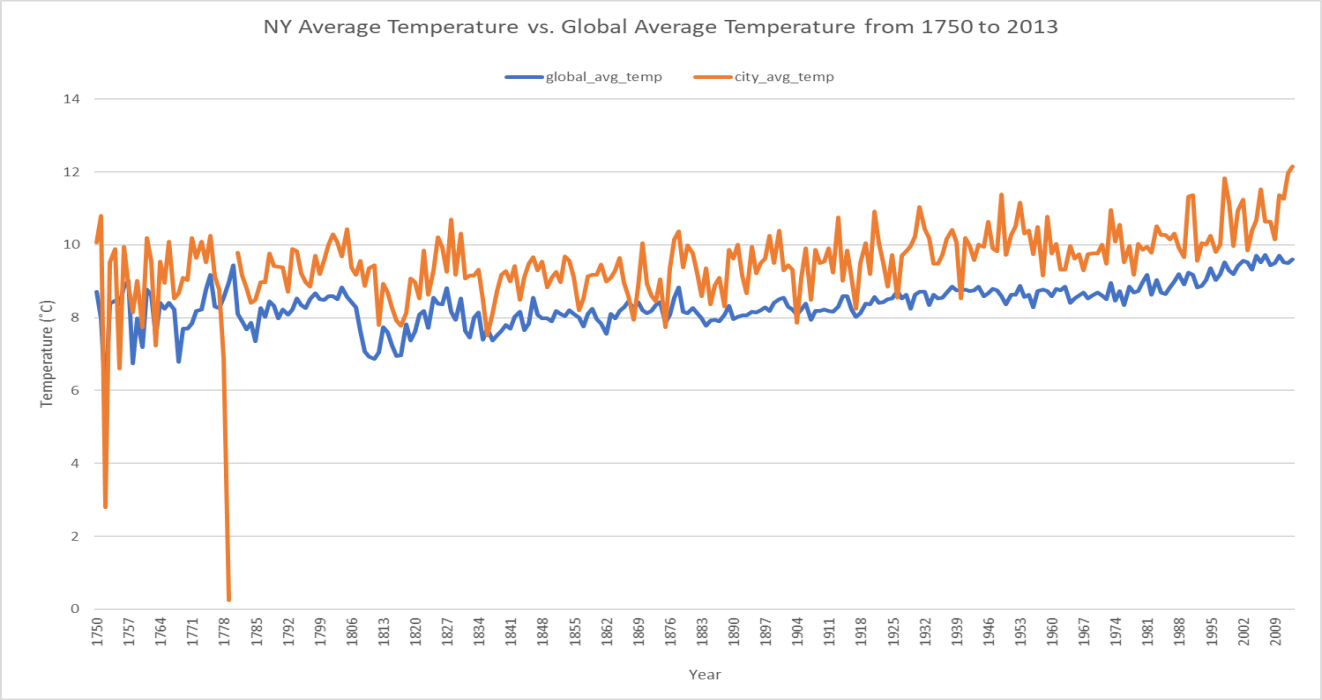
Moving averages are used to smooth out period to period fluctuations in the past values with an n-period moving average (here 10-year MA) to make it easier to observe long-term trends and also to remove short-term fluctuations and the missing data.

year	global_avg_temp	city_avg_temp	10-yrs MA global avg temp	10-yrs MA city avg temp	Avg difference
1750	8.72	10.07			1.35
1751	7.98	10.79			2.81
1752	5.78	2.81			-2.97
1753	8.39	9.52			1.13
1754	8.47	9.88			1.41
1755	8.36	6.61			-1.75
1756	8.85	9.94			1.09
1757	9.02	8.89			-0.13
1758	6.74	8.15			1.41
1759	7.99	9.01	8.03	8.57	1.02
1760	7.19	7.73	7.88	8.33	0.54
1761	8.77	10.18	7.96	8.27	1.41
1762	8.61	9.55	8.24	8.95	0.94
1763	7.5	7.23	8.15	8.72	-0.27
1764	8.4	9.55	8.14	8.68	1.15
1765	8.25	8.96	8.13	8.92	0.71
1766	8.41	10.09	8.09	8.93	1.68
1767	8.22	8.52	8.01	8.90	0.30
1768	6.78	8.67	8.01	8.95	1.89
1769	7.69	9.1	7.98	8.96	1.41
1770	7.69	9.04	8.03	9.09	1.35
1771	7.85	10.18	7.94	9.09	2.33
1772	8.19	9.64	7.90	9.10	1.45
1773	8.22	10.08	7.97	9.38	1.86
1774	8.77	9.52	8.01	9.38	0.75
1775	9.18	10.24	8.10	9.51	1.06
1776	8.3	9.13	8.09	9.41	0.83
1777	8.26	8.77	8.09	9.44	0.51
1778	8.54	6.89	8.27	9.26	-1.65
1779	8.98	0.25	8.40	8.37	-8.73
1780	9.43		8.57	8.30	-9.43
1781	8.1	9.79	8.60	8.26	1.69
1782	7.9	9.15	8.57	8.20	1.25
1783	7.68	8.81	8.51	8.06	1.13
1784	7.86	8.4	8.42	7.94	0.54
1785	7.36	8.49	8.24	7.74	1.13
1786	8.26	8.98	8.24	7.73	0.72
1787	8.03	8.97	8.21	7.75	0.94
1788	8.45	9.77	8.21	8.07	1.32
1789	8.33	9.42	8.14	9.09	1.09
1790	7.98	9.39	8.00	9.12	1.41
1791	8.23	9.37	8.01	9.08	1.14
1792	8.09	8.72	8.03	9.03	0.63
1793	8.23	9.88	8.08	9.14	1.65
1794	8.53	9.83	8.15	9.28	1.30
1795	8.35	9.21	8.25	9.35	0.86
1796	8.27	8.97	8.25	9.35	0.70
1797	8.51	8.85	8.30	9.34	0.34
1798	8.67	9.71	8.32	9.34	1.04
1799	8.51	9.19	8.34	9.31	0.68
1800	8.48	9.58	8.39	9.33	1.10

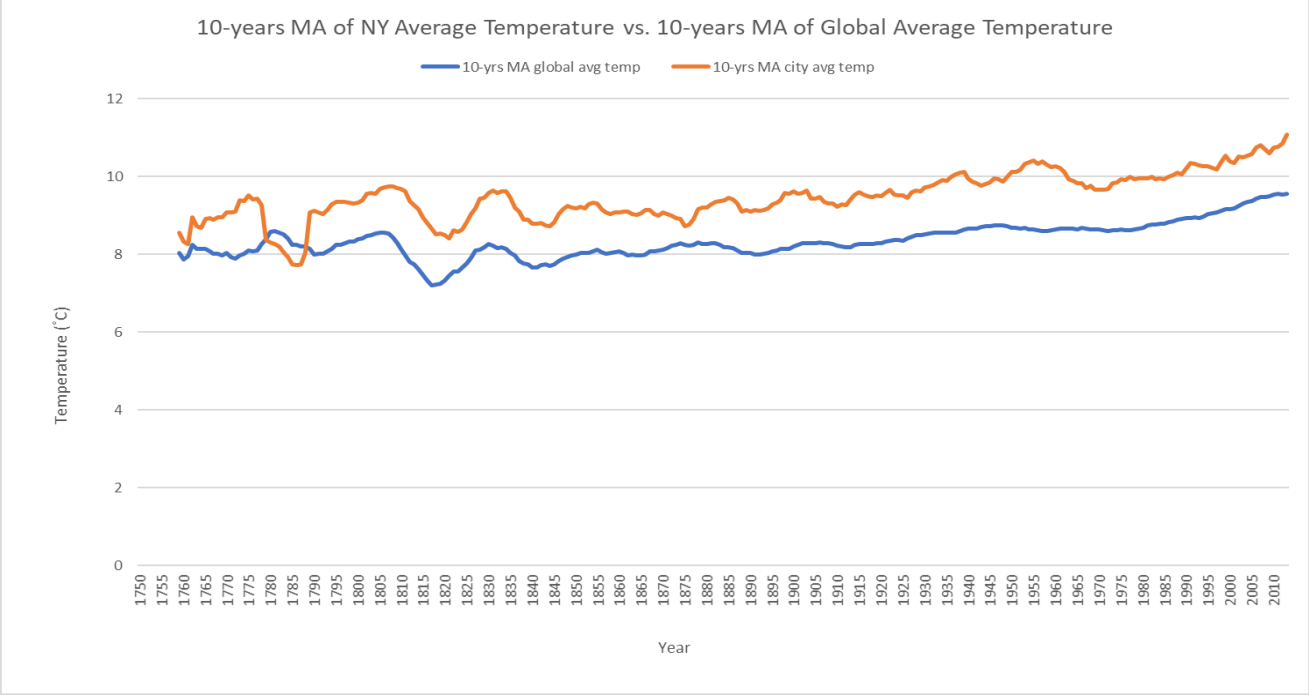
1941	8.77	10.19	8.66	9.87	1.42
1942	8.73	9.96	8.66	9.82	1.23
1943	8.76	9.58	8.70	9.76	0.82
1944	8.85	10.01	8.73	9.81	1.16
1945	8.58	9.94	8.73	9.86	1.36
1946	8.68	10.63	8.75	9.95	1.95
1947	8.8	9.91	8.76	9.92	1.11
1948	8.75	9.83	8.74	9.87	1.08
1949	8.59	11.39	8.73	10.00	2.80
1950	8.37	9.72	8.69	10.12	1.35
1951	8.63	10.27	8.67	10.12	1.64
1952	8.64	10.52	8.67	10.18	1.88
1953	8.87	11.17	8.68	10.34	2.30
1954	8.56	10.31	8.65	10.37	1.75
1955	8.63	10.39	8.65	10.41	1.76
1956	8.28	9.75	8.61	10.33	1.47
1957	8.73	10.5	8.61	10.39	1.77
1958	8.77	9.16	8.61	10.32	0.39
1959	8.73	10.77	8.62	10.26	2.04
1960	8.58	9.76	8.64	10.26	1.18
1961	8.8	10.03	8.66	10.24	1.23
1962	8.75	9.32	8.67	10.12	0.57
1963	8.86	9.32	8.67	9.93	0.46
1964	8.41	9.96	8.65	9.90	1.55
1965	8.53	9.63	8.64	9.82	1.10
1966	8.6	9.75	8.68	9.82	1.15
1967	8.7	9.3	8.67	9.70	0.60
1968	8.52	9.75	8.65	9.76	1.23
1969	8.6	9.76	8.64	9.66	1.16
1970	8.7	9.77	8.65	9.66	1.07
1971	8.6	10.01	8.63	9.66	1.41
1972	8.5	9.49	8.60	9.67	0.99
1973	8.95	10.96	8.61	9.84	2.01
1974	8.47	10.08	8.62	9.85	1.61
1975	8.74	10.56	8.64	9.94	1.82
1976	8.35	9.53	8.61	9.92	1.18
1977	8.85	9.96	8.63	9.99	1.11
1978	8.69	9.17	8.65	9.93	0.48
1979	8.73	10.02	8.66	9.96	1.29
1980	8.98	9.86	8.69	9.96	0.88
1981	9.17	9.95	8.74	9.96	0.78
1982	8.64	9.78	8.76	9.99	1.14
1983	9.03	10.51	8.77	9.94	1.48
1984	8.69	10.26	8.79	9.96	1.57
1985	8.66	10.26	8.78	9.93	1.60
1986	8.83	10.15	8.83	9.99	1.32
1987	8.99	10.32	8.84	10.03	1.33
1988	9.2	9.93	8.89	10.10	0.73
1989	8.92	9.66	8.91	10.07	0.74
1990	9.23	11.32	8.94	10.21	2.09
1991	9.18	11.36	8.94	10.36	2.18
1992	8.84	9.57	8.96	10.33	0.73
1993	8.87	10.04	8.94	10.29	1.17
1994	9.04	10	8.98	10.26	0.96
1995	9.35	10.24	9.05	10.26	0.89
1996	9.04	9.81	9.07	10.23	0.77
1997	9.2	10	9.09	10.19	0.80
1998	9.52	11.82	9.12	10.38	2.30
1999	9.29	11.16	9.16	10.53	1.87
2000	9.2	9.97	9.15	10.40	0.77
2001	9.41	10.93	9.18	10.35	1.52
2002	9.57	11.25	9.25	10.52	1.68
2003	9.53	9.84	9.32	10.50	0.31
2004	9.32	10.39	9.34	10.54	1.07
2005	9.7	10.68	9.38	10.59	0.98
2006	9.53	11.52	9.43	10.76	1.99
2007	9.73	10.63	9.48	10.82	0.90
2008	9.43	10.64	9.47	10.70	1.21
2009	9.51	10.14	9.49	10.60	0.63
2010	9.7	11.36	9.54	10.74	1.66
2011	9.52	11.27	9.55	10.77	1.75
2012	9.51	11.97	9.55	10.84	2.46
2013	9.61	12.16	9.56	11.08	2.55

Steps 5: Plot Line chart with local and global temperature trends.

The following line chart depicts a comparison between the global average temperature and the average temperature of New York from 1750 to 2013:



The chart below shows yearly temperature for the city of New York and the global temperature from 1750 to 2013, along with a 10-year moving average of the data.



Step 6: Observations

- The global temperature has been in the range of 7 °C and 10 °C in the period of 1750 to 2013. While New York has been hotter on average compared to the global average, in the range of 7 °C and 12 °C at the same period of time.

- The average difference in temperature was around 1 °C over the time period 1750 to 1950. After that, this difference increased slightly by 0.32 °C and that because New York is getting warmer compared to the earth.

- From 1750 to 1780, both global temperature and NY temperature were highly volatile. After that, the NY average temperature showed some volatility around its means while the volatility of the global average temperature was too little but constant over time. The correlation of 1.00 means that the relationship between the global temperature and New York temperature is very strong and then their temperature is increasing at the same rate, which means that the earth is getting hotter.

Time Period	1750 - 1780	1781 - 2013
Variance of 10-yrs MA of Global Avg Temp	VAR () = 0.03	0.20
Variance of 10-yrs MA of NY Avg Temp	VAR () = 0.15	0.36
Correlation between Global and NY avg temp	CORREL () = -0.27	1.00