Project By:
Anshika Shaw, mail.anshikshaw@gmail.com

Data Analyst Term 1
March 2018

Project
Exploring Weather Trends
Udacity Project

Version: *Project 1* Date: (19/03/2018)

Table of Contents

Introduction

1.1. Introduction

The introduction to this project involves studying in details and analyzing various angles and doing an exploratory finding in terms of mapping out the significant weather trends and fluctuations observed while working on the dataset provided. Starting with the extraction of the desired set of data, to cleaning it and keeping what exactly is needed and needs to be operated upon. Followed by a few numerical approaches on the dataset, the major one being-"Moving Average" tabulation over a period of say 30,20,10 years bar. The pictorial representation of these points on line charts- at labelling and plotting them and finally drawing conclusions out of it.

The criteria for selection of Moving Averages has been illustrated and discussed below.

A bit about Moving Average- What and why?

Moving averages are used to gauge the direction of the current trend. Moving average is a mathematical result that is calculated by averaging a number of past data points. Once determined, the resulting average is then plotted onto a chart in order to allow traders to look at smoothed data rather than focusing on the day-to-day price fluctuations that are inherent in all financial markets.

Relational Analysis to why Moving Average is needed in our project and how does it help or be of use?

Moving Average, the utility of it includes looking for smoothened data as in the data that shows less fluctuations, that is safe to play with or invest in or shows a steady motion. No high peak or drop points. This can be a crucial decision metric, noting which is the best time or the worst time based on the dataset provided.

Moving Period Selection Criteria

A close look over noisy and smoothened data, had been tactically monitored and based on it the moving average period was selected. The moving average was tabulated and noted for a period of 10, 20 and 30 years, out of which the 30 years period had the smoothest edge.

Why 10,20,30 and not 5,7 or other values?

Had this been a data related to finance or other computational worlds, I could have gone ahead conducting moving average for a more recent timeline, to mark what works best and to spot the recent trends, and to fit with the recent growing and observed trends.

Since, this is a weather data and given the more and more emphasis on smoothened data, I took up figures like 10,15 or 30. Because what worked then or what climatic condition was back 30 years before, it could have major role in establishing links to what the climatic condition we have today now. Secondly, of-course the 30 years moving average period had more smoothened data than 10,15 years.

1.2. Goal and Objectives:

1. Extraction of desired data in .csv format
2. Removal of identical values, Checking Redundancies - Cleaning data
3. Moving Average Tabulation - Application of Moving Average on the dataset
4. Plotting them in line charts
5. *Noting and dotting major weather trends in global and the city of my residence.*

1.3. My Approach at Things- How And What

STEP 1: **Extraction of data provided by Udacity**

- ➢ Initially, the data had to be extracted from the Udacity website. The data was extracted and stored in .csv format, which later was opened in Excel sheet.

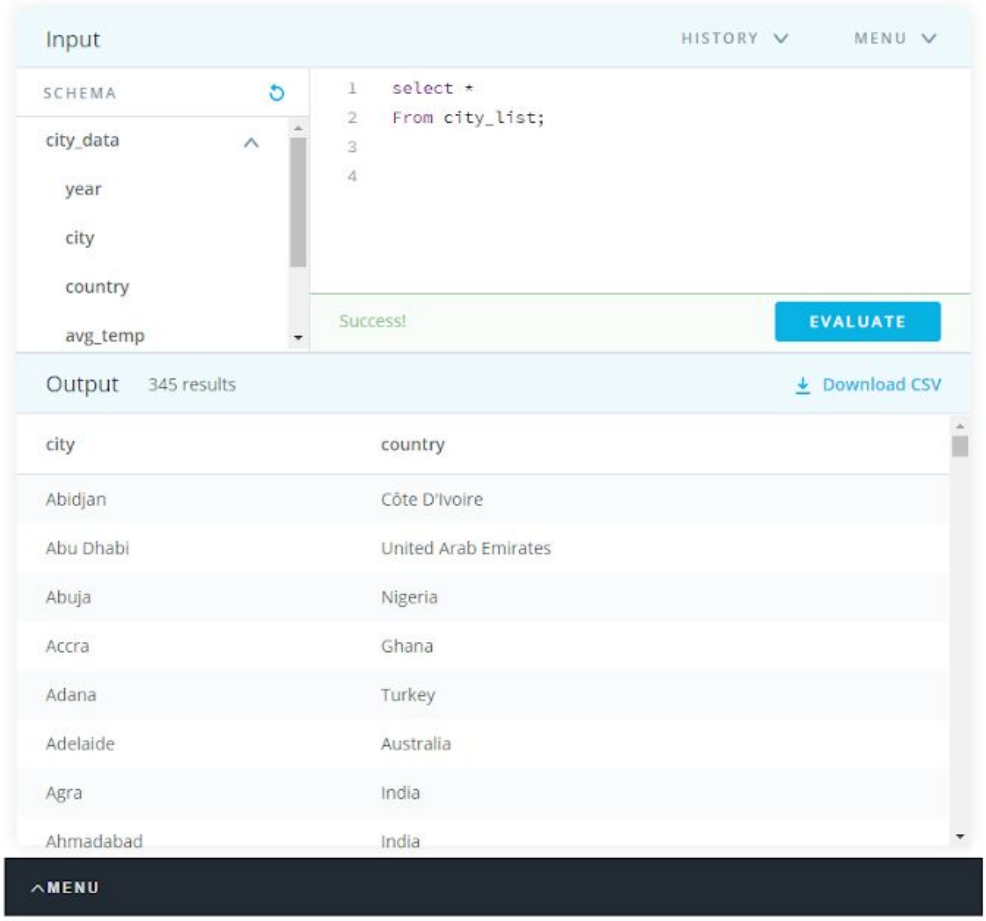
The commands that were ran to extract data:

```
SELECT *
FROM city_list;
```

This extracted the list of all the cities whose average weather temperatures was recorded.

the city nearest to you.

- city_data - This contains the average temperatures for each city by year (°C).
- global_data - This contains the average global temperatures by year (°C).

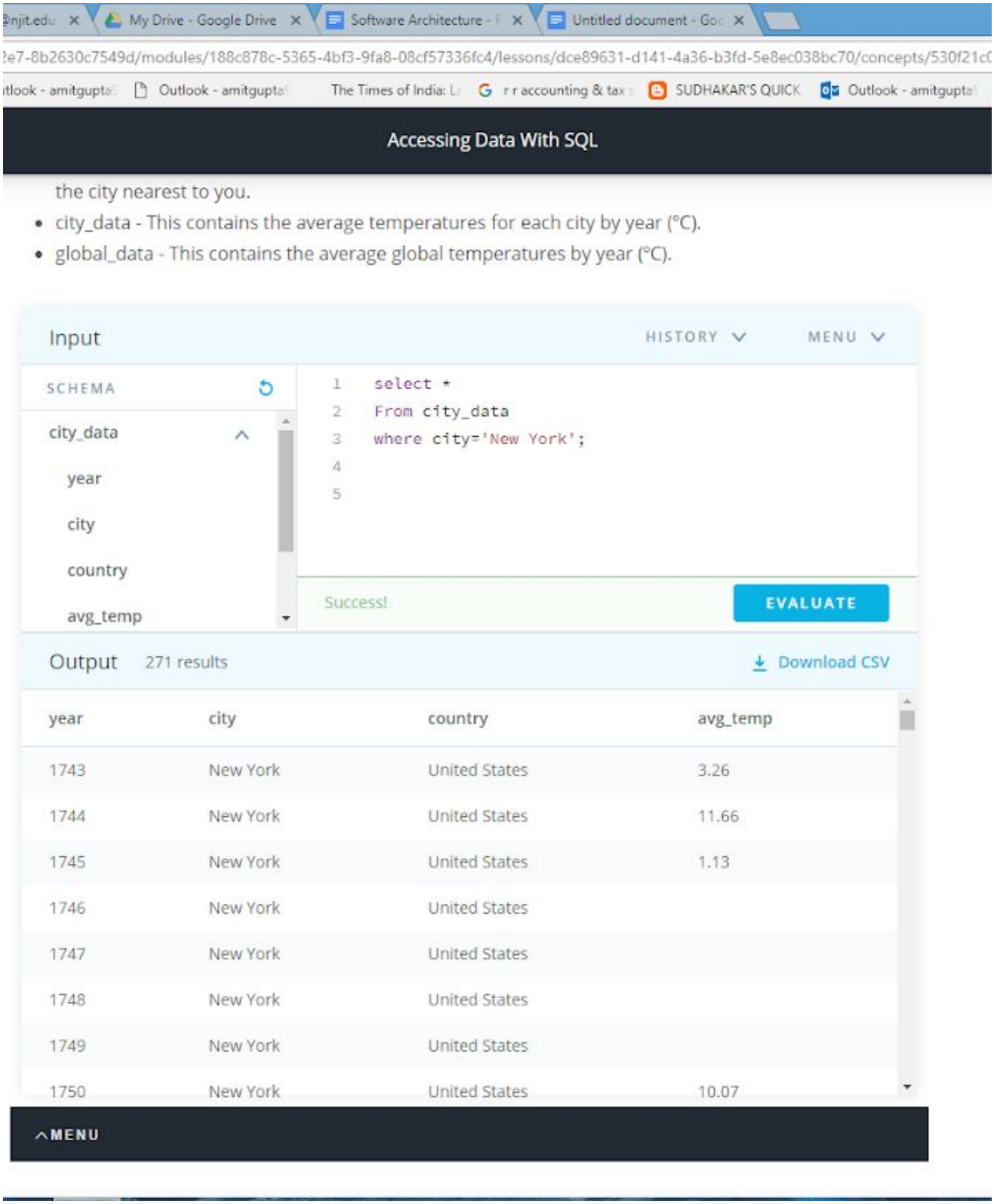


➢ Secondly, instead of using MySQL, I had used the same SQL platform provided by Udacity to get the information- average weather temperature about the city I reside in. That's New York in my case.

The commands that were ran to extract the weather information about my city 'New York':

SELECT *
FROM city_data
Where city='New York';

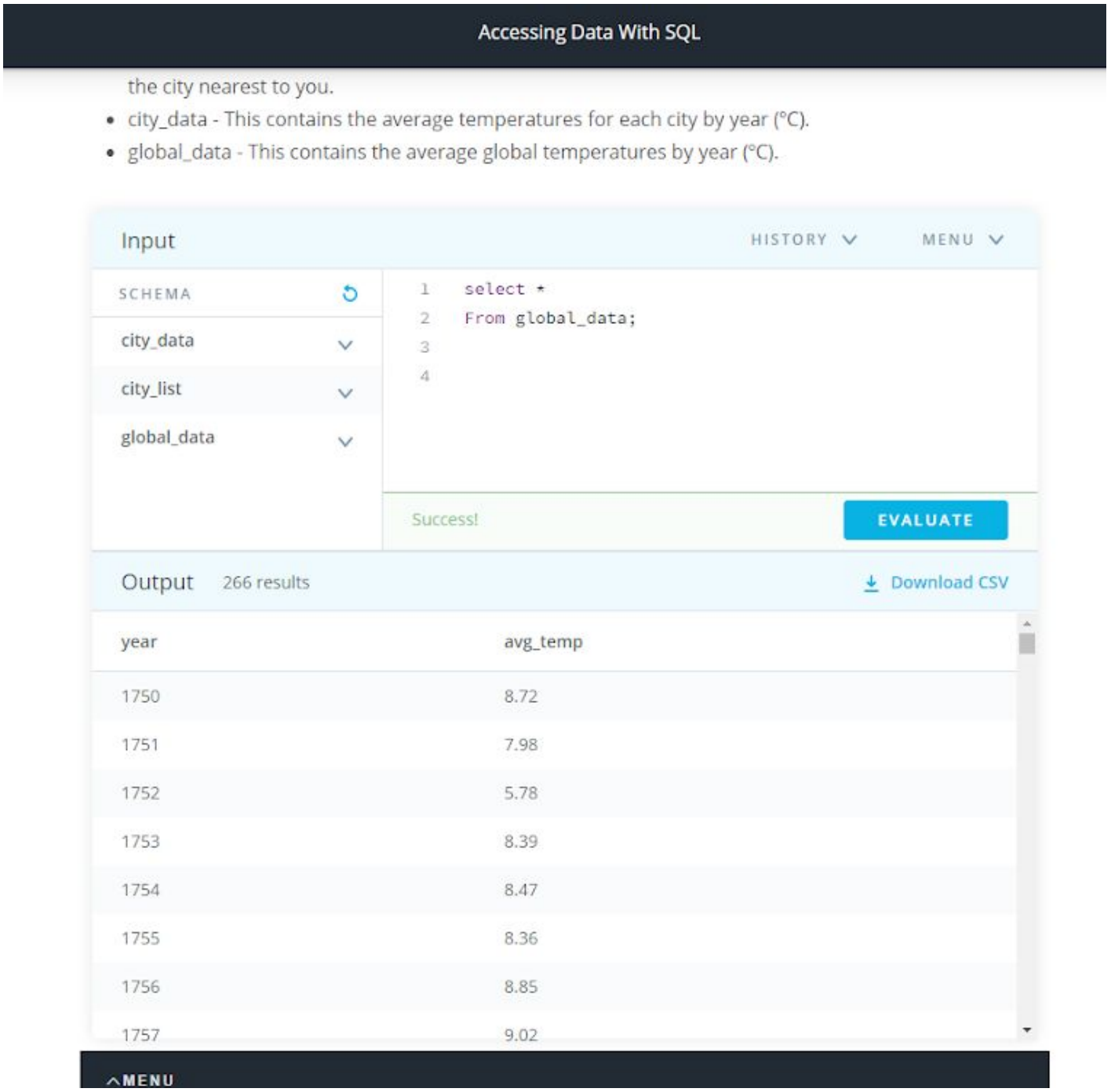
Result: This extracted a list of 271 rows, containing weather details on New York.


njit.edu
My Drive - Google Drive
Software Architecture - P
Untitled document - Goo
e7-8b2630c7549d/modules/188c878c-5365-4bf3-9fa8-08cf57336fc4/lessons/dce89631-d141-4a36-b3fd-5e8ec038bc70/concepts/530f21c0
tlook - amitgupta
Outlook - amitgupta
The Times of India: L
r r accounting & tax
SUDHAKAR'S QUICK
Outlook - amitgupta
Accessing Data With SQL
the city nearest to you.
city_data - This contains the average temperatures for each city by year (°C).
global_data - This contains the average global temperatures by year (°C).
Input
HISTORY
MENU
SCHEMA
city_data
year
city
country
avg_temp
1 select *
2 From city_data
3 where city='New York';
4
5
Success!
EVALUATE
Output
271 results
Download CSV
year
city
country
avg_temp
1743 New York United States 3.26
1744 New York United States 11.66
1745 New York United States 1.13
1746 New York United States
1747 New York United States
1748 New York United States
1749 New York United States
1750 New York United States 10.07
MENU

➢ Thirdly, the **global_data containing global average temperature was extracted.**

The command that I ran to extract:

```
Select *
From global_data;
```



These files were downloaded and stored as .csv format to be operated upon later-City_List.csv,Global_Data.csv, NewYork.csv.

STEP 2: **Cleaning dataset.**

The .csv file was opened in excel and checked for redundant values. The filter for redundancy checking was applied on the ***year column.***

Since, ***there were no identical years for both global_data and New_York dataset, we would progress without any further effort here.***

STEP 3:

Moving Average Tabulation.

I did a moving average calculation of 30 years for global_data and city_data. The reasons are listed above.

Implemented the moving average formula in excel for the desired number of rows.

Storing the moving average values in a separate column in excel file and saving the new files in .csv format. Kindly refer to the zipped folder for the moving average tabulated files. The files are **Global_Data_MovingAverage30Years.csv, NewYork_MovingAverage_30Years.csv, Global_DataVsNewYork_15Years_MovingAverage.csv.**

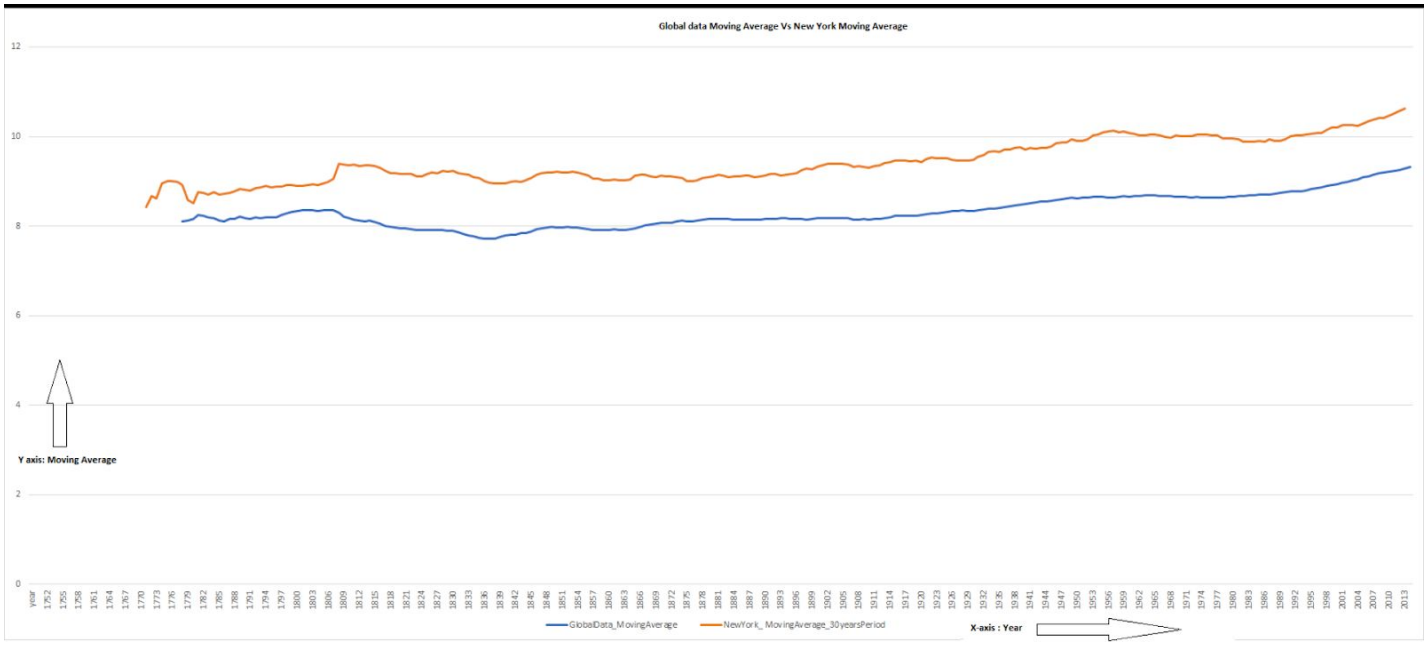
STEP 4:

Plotting the moving average values in line charts for global_data, city_data and comparative mapping between NewYork_city_data and global_data.

STEP 5:

Noting the essential and significant differences from line chart.

The key observations, that I made based on line charts.



1. **GlobalData_MovingAverage Vs NewYork_MovingAverage_30yearsPeriod**

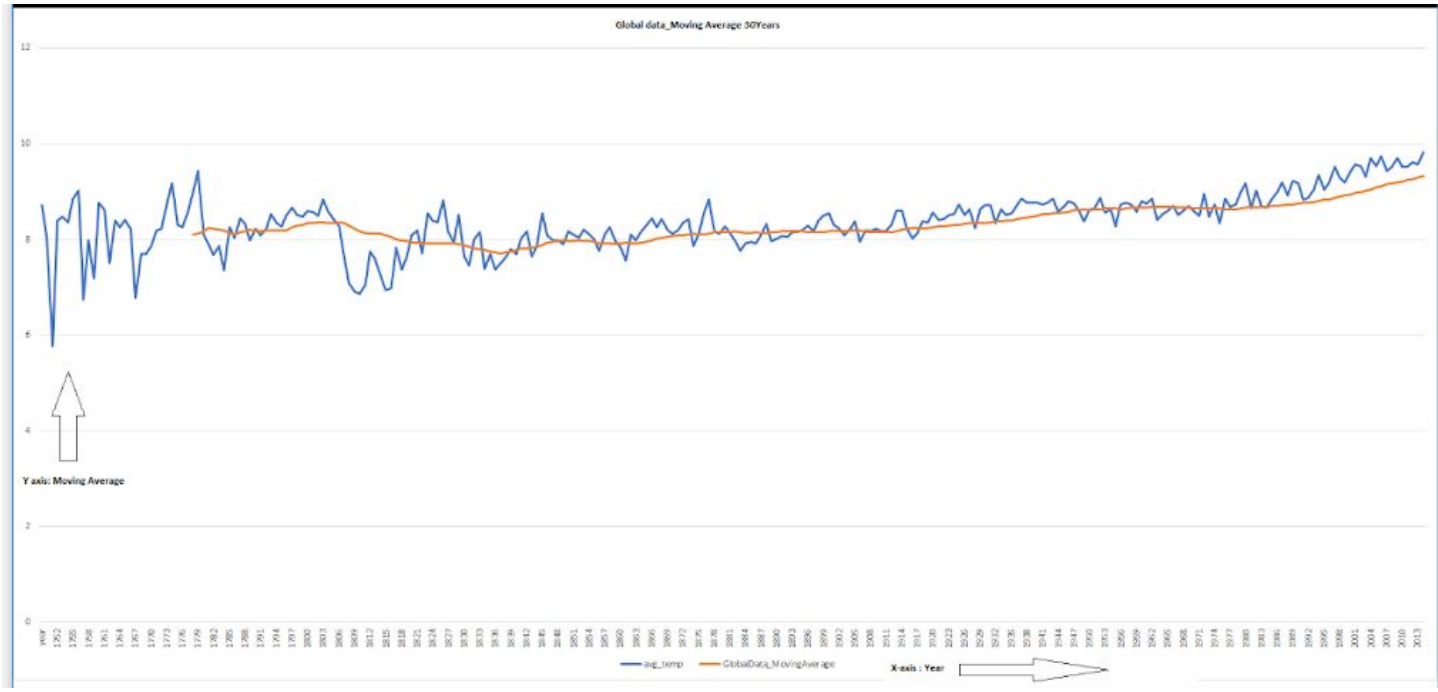
Kindly refer to image file in zipped folder for clearer picture.

Observations:

- ➢ It was in the year 1812, that the New York's average climate had elevated a little higher, not to the maximum height though but a little higher. Since, then it experienced a steady flow with minor noticeable fluctuations.
- ➢ There had been a minor rise in average temperature in 1959, with the overall average temperature being steady after that, a few here and there till 2013.
- ➢ In contrast to NewYork_MovingAverage_30yearsPeriod, the temperature rose in 1803, and dropped down in the year 1839. It lifted up a little in 1854. A little up and down here and there till year 1932. It maintained a close to steady flow after that.
- ➢ So, overall the global average temperature have a slightly better even curve and less fluctuations as compared to New York average temperature.

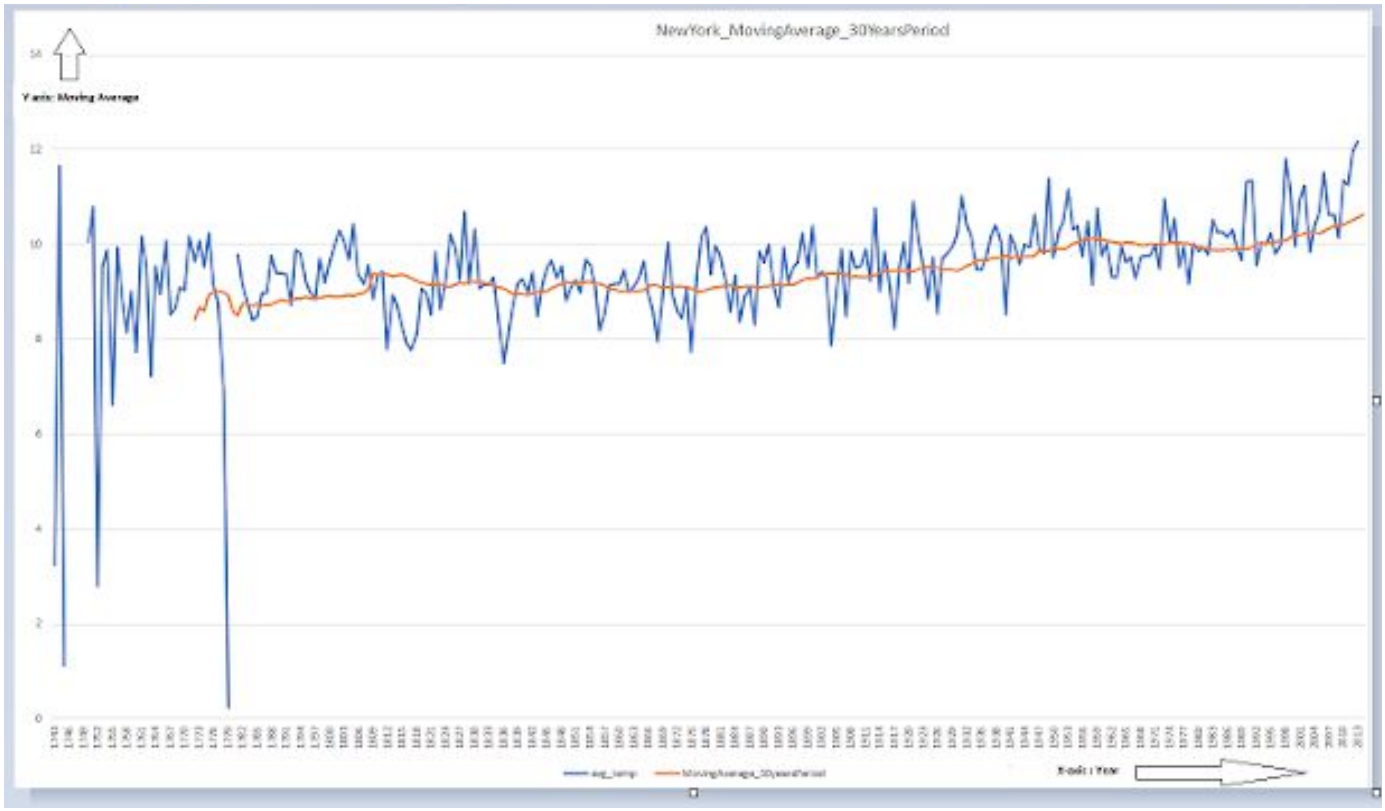
****Revision, as suggested based on review:** The reason why global temperature is less volatile and predictable as compared to New York temperature, since the moving average curve is somewhat even, not much fluctuations or changes perceived, so given the past temperature data is easier to conclude that global temperature are better off as when predictability is concerned, it's easier to gauge and estimate.

2. **GlobalData_MovingAverage30Years :**



Kindly refer to image file in zipped folder for clearer picture.

3. NewYork_MovingAverage_30YearsPeriod: Kindly refer to image file in folder for clearer picture.


NewYork_MovingAverage_30YearsPeriod
Y axis: Moving Average
X axis : Year