Data Analyst Nanodegree

Exploring Weather Trends Project

May 1, 2021

Index:

Project Summary:

In this project, I compared between my city "Mecca" average temperatures with global average temperatures. Also, predicted the next 61 average temperatures. In addition to compared average temperatures of 3 capital cities in Asia with global average temperatures. I extracted the data using SQL from database given by Udacity and analyzed it using Excel.

Comparison Between Mecca and Global Average Temperatures
SQL queries

My city is Mecca and I used SQL to extract data from the database provided by Udacity to compare between my city average temperature to global average temperature.
At first I wrote a query to explore the database.

```
SELECT *
FROM city_data
```

```
SELECT *
FROM city_list
```

```
SELECT *
FROM global_data
```

Then I looked for my city "Mecca" by this query:

```
SELECT *
From city_data
WHERE city LIKE '%Mecca%'
```

After I found my city I wrote a query to join two tables (city_data table and global_data) to extract the year, Mecca average temperature, and global average temperature columns to be in one single result:

```
SELECT c.year, city, c.avg_temp Mecca_avg_temp, g.avg_temp Global_avg_temp
From city_data c
JOIN global_data g
ON c.year = g.year
WHERE c.city = 'Mecca'
```

The result of this query contained NULLs values in Mecca average temperature column. so, I decided to delete this values from my query and kept only the years from 1861 to 2013 to be fair compression between both database:

```
SELECT c.year, city, c.avg_temp Mecca_avg_temp, g.avg_temp Global_avg_temp
From city_data c
JOIN global_data g
ON c.year = g.year
WHERE c.city = 'Mecca' AND g.year Between 1861 AND 2013
```

Finally, I exported the result to CSV file and opened it in Excel in order to analyze it.

10 Years Moving Average

I calculated 10 years moving average (MA) to smooth out the line chart by using AVERAGE function in new column for both Mecca average temperature and global average temperature.
Then I plotted a line chart from the MA columns.

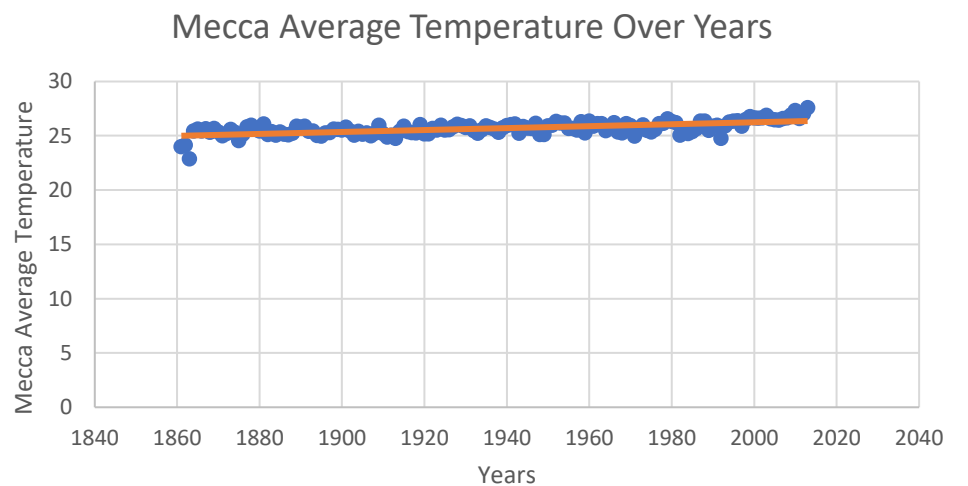10 Years Moving Average Temperature (Mecca and Globally)
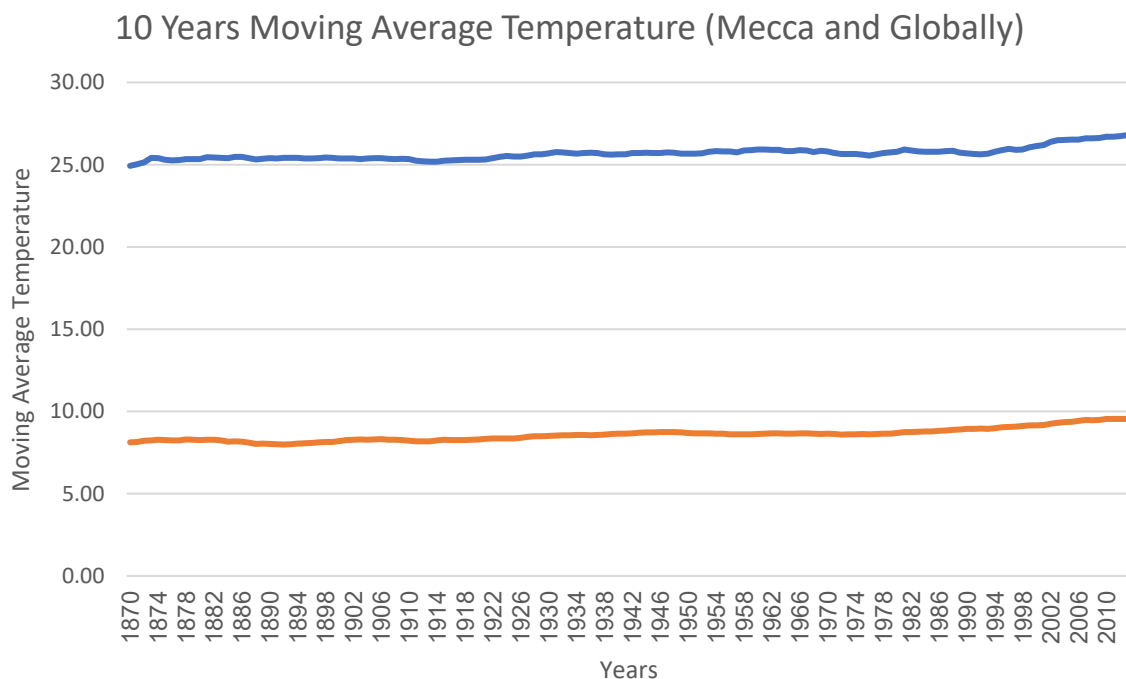
Observations:

1- The MA of Mecca temperature is varies between 24.94 and 26.80
2- The MA of global temperature is varies between 8.00 and 9.61
3- The maximum MA of Mecca temperature is 26.80 and the minimum is 24.94
4- The maximum MA of global temperature is 9.61 and the minimum is 8.00
5- The MA of Mecca temperature is 3 times higher than global which is mean Mecca temperature is hotter 3 times than the global temperature.
6- The MA of both Mecca and global temperature are increasing over years and tend to be hotter..

Comparison Between Mecca and Global Predicted Average Temperatures

Linear Regression Prediction

I used the linear regression equation ($\hat{y}=b_0+b_1x_1$) to predict the average temperature for the next 61 years.

First: I plotted scatter plot for Mecca average temperature in y axis and years in x axis and added a straight line to see if the relationship between variables are linear. Then did a regression module from data analysis tool in Excel to calculate correlation coefficient, R squared, intercept, and slope coefficient.

Mecca Average Temperature Over Years

Mecca

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.641361413
R Square	0.411344462
Adjusted R Square	0.40744608
Standard Error	0.471894213
Observations	153

ANOVA

	df	SS	MS	F	ignificance F
Regression	1	23.49690544	23.49690544	105.5167405	4.22E-19
Residual	151	33.62530632	0.222684148		
Total	152	57.12221176			

	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%	ower 95.0%	Jpper 95.0%
Intercept	8.488387848	1.673593157	5.071954202	1.13946E-06	5.181704	11.79507	5.181704	11.79507
year	0.008872951	0.000863788	10.27213417	4.21872E-19	0.007166	0.01058	0.007166	0.01058

From the scatter plot and correlation coefficient in regression statistics (0.6) we can see that there are a positive moderate linear relationship between the average temperature of Mecca and years.

Second: I wanted to predict average temperature of Mecca city based on years. So, I plugged in the equation:
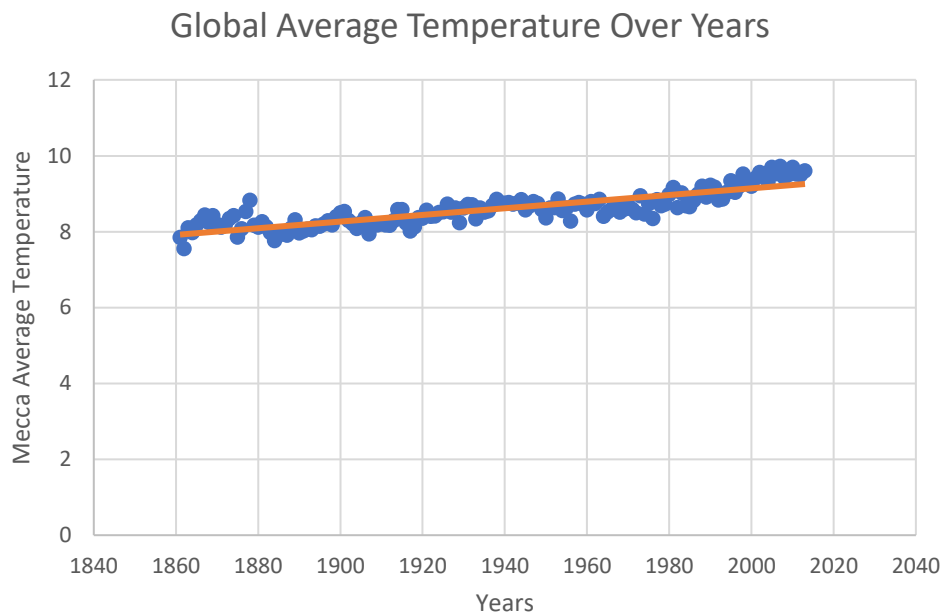$\hat{y}=b_0+b_1x_1$
Where:
b_0 = 8.488387848
b_1 = 0.008872951
I calculated it in Excel by making the intercept and slope "years" cells as a absolute cell references and years as a relative cell reference. Then plugged in the equation and filled all the rest of cells.

Third: After predicting average temperature, I calculated 10 years MA and plotted a line chart for the result.

I did the same for the average temperature of Global.
First: Plot the scatter plot and calculate correlation coefficient.

Global Average Temperature Over Years

Global

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.855814494
R Square	0.732418448
Adjusted R Square	0.730646385
Standard Error	0.235167666
Observations	153

ANOVA

	df	SS	MS	F	ignificance F
Regression	1	22.85784438	22.85784438	413.3139412	4.48E-45
Residual	151	8.350878492	0.055303831		
Total	152	31.20872288			

	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%	ower 95.0%	Jpper 95.0%
Intercept	-8.357390211	0.83403226	-10.02046397	1.95336E-18	-10.0053	-6.70951	-10.0053	-6.70951
year	0.008751458	0.000430467	20.33012398	4.47984E-45	0.007901	0.009602	0.007901	0.009602

From the scatter plot and correlation coefficient in regression statistics ($r = 0.9$) we can see that there are a positive strong linear relationship between the average temperature of global and years.

Second: I predicted average temperature of global based on years by plugging in the equation:
$\hat{y} = b_0 + b_1 x_1$
Where:
b_0 = -8.357390211
b_1 = 0.008751458

I calculated it in Excel by making the intercept and slope "years" cells as a absolute cell references and years as a relative cell reference. Then plugged in the equation and filled all the rest of cells.

Third: After predicting average temperature, I calculated 10 years MA and plotted a line chart for the result.

10 Years Moving Average

After I predicted the average temperature and calculated 10 years MA for both Mecca and global. I plotted a line chart to compares Mecca city temperatures with the global temperatures.

10 Years Predicted Moving Average Temperature (Mecca and Globally)
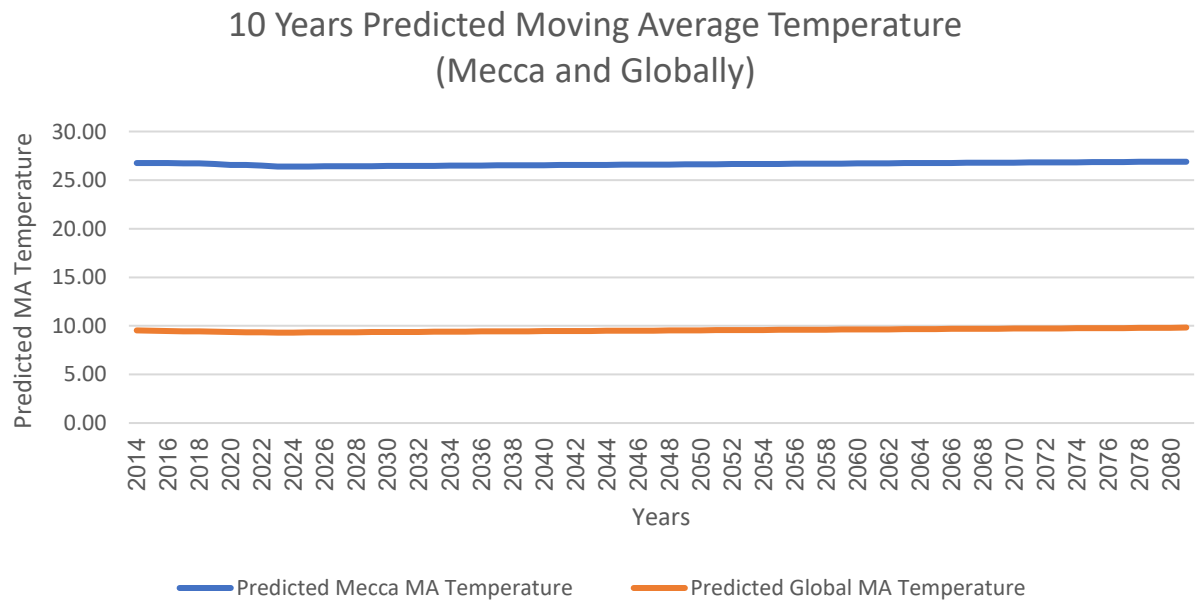
Observations:

1- The predicted MA of Mecca temperature is varies between 26.40 and 26.91

2- The predicted MA of global temperature is varies between 9.31 and 9.82

3- The maximum predicted MA of Mecca temperature is 26.91

4- The maximum predicted MA of global temperature is 9.82

5- The predicted MA of Mecca temperature is decreasing from 26.78 to 26.49 between 2014 and 2033 then increasing from 26.50 to 26.91 between 2034 to 2081.

 6- The predicted MA of global temperature is decreasing from 9.55 to 9.32 between 2014 to 2025 then increasing from 9.33 to 9.82 between 2026 and 2081.

7- from 2021, expected to decreasing average temperature in the next 9 years in Mecca city while for global expected to decreasing for the next 4 years.

Comparison Between 3 Capital Cities and Global Average Temperatures

SQL queries

I looked for the capital city for 3 countries in Asia to compare them to global. Riyadh for Saudi Arabia, New Delhi for India, and Seoul for Korea. So, I wrote a query to find out these cites in the database:
For New Delhi

```
SELECT *
From city_data c
WHERE city LIKE 'New%'
```

For Riyadh

```
SELECT *
From city_data c
WHERE city LIKE 'Riya%'
```

For Seoul

```
SELECT *
From city_data c
WHERE city LIKE 'Seo%'
```

Then I joined two tables (city_data and global_data) to have the average temperature for all 3 cities that I choosed and globally.

```
SELECT city, c.year, c.avg_temp Capital_city_temp, g.avg_temp Global_temp
From city_data c
JOIN global_data g
ON c.year = g.year
WHERE c.city IN ('New Delhi', 'Riyadh', 'Seoul')
```

But there was a NULLs values and different period of time so, I decided to delete NULLs and limited the year between 1870 to 2013 to matched them and be fair in comparison

```
SELECT city, c.year, c.avg_temp Capital_city_temp, g.avg_temp Global_temp
From city_data c
JOIN global_data g
ON c.year = g.year
WHERE c.city IN ('New Delhi', 'Riyadh', 'Seoul') AND c.year Between 1870 AND 2013
```

3 Years Moving Average

I imported the csv file to Excel and calculated 3 years MA average temperature for the 3 capital cities in addition to global by using AVERAGE function. Then I plotted the result in one line chart to compares the 3 capital cities temperatures with the global temperatures.

3 Years MA Temperatures
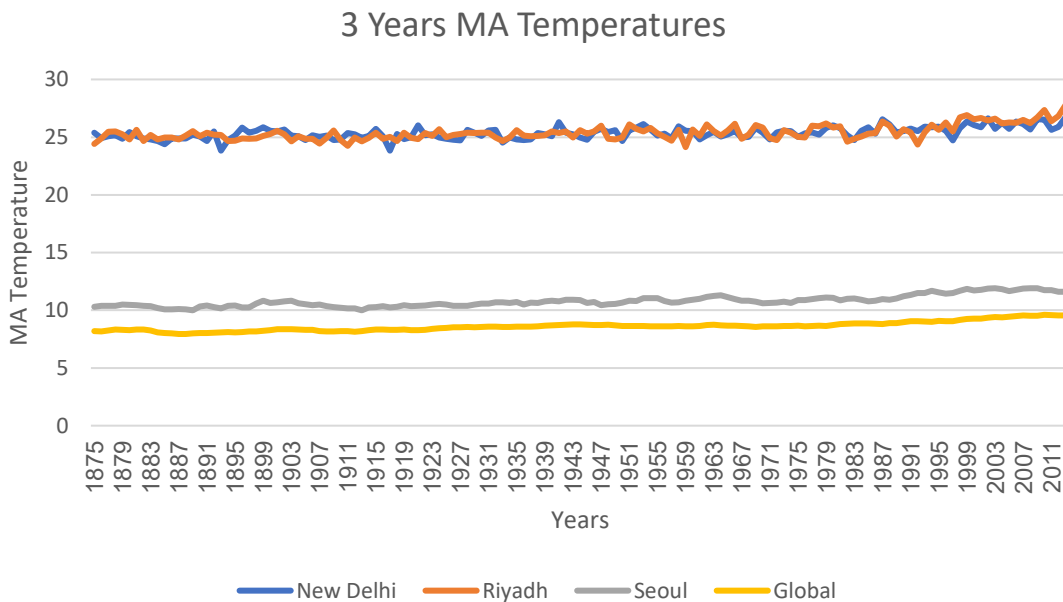
Observations:
1- The MA of Riyadh city temperatures is varies between 24.80 and 26.88
2- The MA of New Delhi city temperatures is varies between 24.76 and 26.17
3- The Ma average for both New Delhi and Riyadh are almost in the same range and more hotter than global MA temperatures.
4- The MA of Seoul city temperatures is varies between 10.00 and 11.91
5- The MA global temperatures is varies between 7.94 and 9.60
6- The MA of Seoul temperatures is a little higher than global but still colder than New Delhi and Riyadh.

Conclusion:
After analyzed the data we can note that the global average temperatures is the most colder and the average temperatures is increasing over years.