EXPLORING WEATHER TRENDS

Project name

1st of Dec 2020

Data Analyst Nanodegree

Student Name:
Mohammed Meshal Alotaibi

Abstract:

I would to say thank you to everyone work at **Udacity**, this project helped me a lot to think in different way and utilize my skills that gained through my journey in **Data science** career in the recently couple of years.

Project steps:

1- Using SQL to extract data stored in DBs.
2- Export to CSV file.
3- Exploratory Data Analysis (EDA).
4- Manipulate missing values.
5- Applied statistics methods.
6- Data Visualisation.

Documentation:

First, every Data science project need to know about the dataset that we want to be analysis and extract useful, actionable, and meaningful insights, and the Professional Data scientist need to understand the business or understanding datasets before go through the analysis as first step in the project.

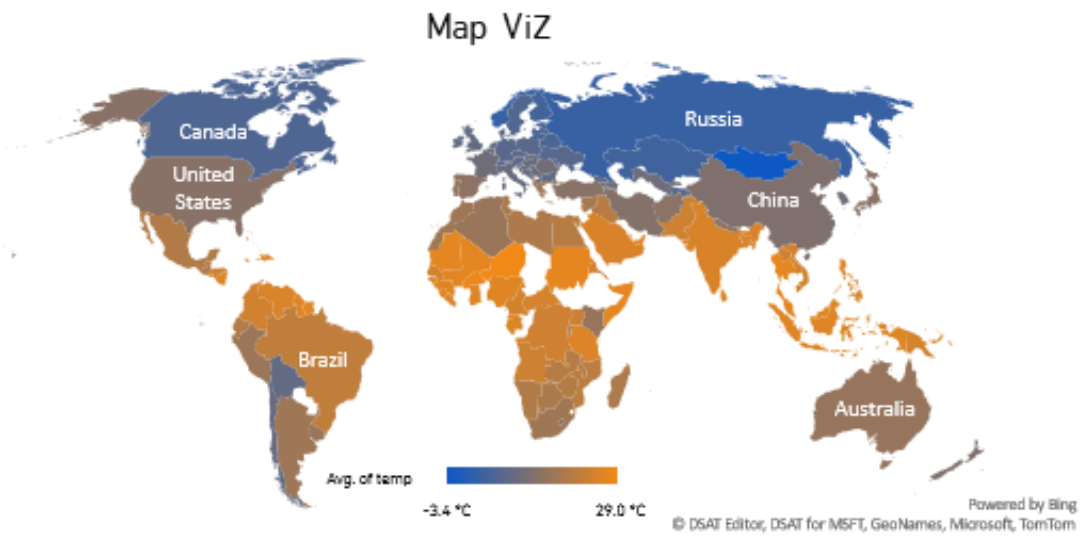
In this project I read about Temperatures issue worldwide and I take some notes that may be help me to understand more about the real problem and how Data science can solve this issue!

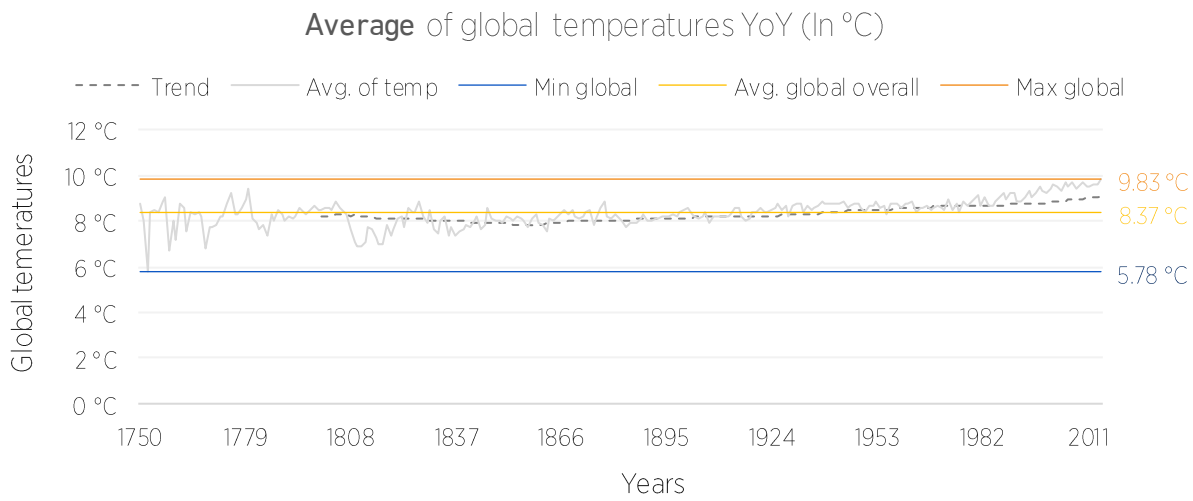
Then I went extract the dataset stored in Udacity workspace and make important steps before the final report as following:

1- Extract data stored in DBs as RAW data to CSV.
2- Make sure if any missing values exist in dataset that belong to the city that I live into it.
3- Thinking how I can Manipulate that missing values using Statistics methods.
4- Manipulated the missing values such as (NULL, NaN) in that dataset by Avg. of Riyadh overall and pass it to the rows with NULL values into it.
5- Create Data visualisation to represent the issue that world suffer it.

Tools / Techniques used in this report:

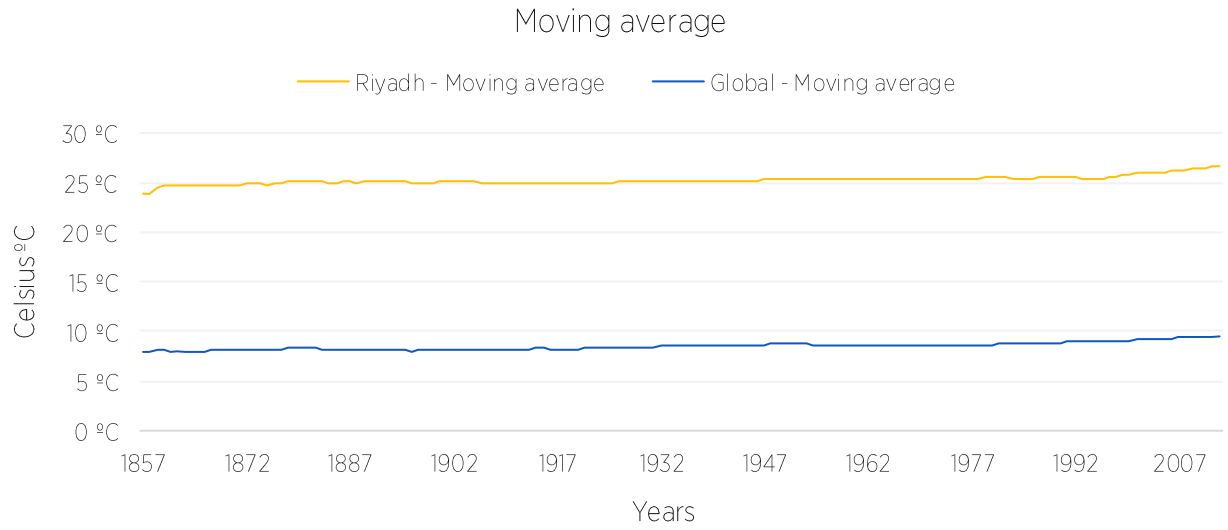
1- SQL.
2- MS Excel.
3- Statistics methods.

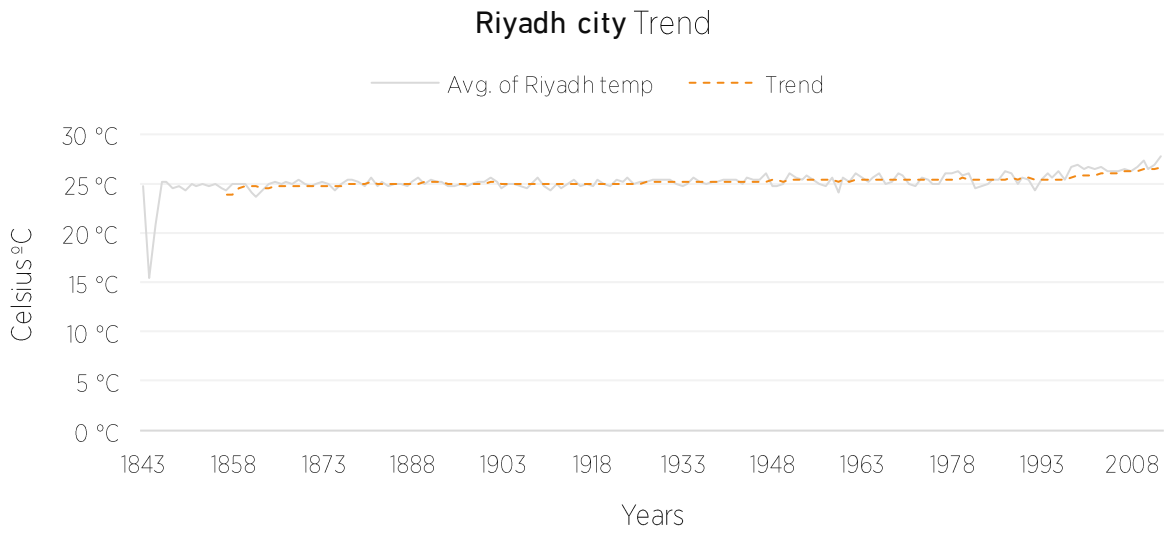




The above **line chart indicated** to the Average of global temperatures as yearly basis started by 1750 year up to 2015 (2 centuries and half)

1- Min average of temp amongst of the last 266 years is (5.78 °C).
2- The overall average of global temperatures is (8.37°C), No. of years was above the average is (134 yrs.) and below the average is (132 yrs.).
3- Max average of temp amongst of the last 266 years is (9.83 °C).
4- Map Viz represent average of temperatures worldwide.





Capital of **Saudi Arabia Riyadh city** on average of yearly changes Year-Over-Year is (+0.018 °C), started by 1843 year up to 2013 (2 centuries and half)

The above **line chart indicated** to the Average of Riyadh temperature as yearly basis started by 1843 year up to 2013, and Moving average **15 yrs.**

1- Min average of temp amongst of the last 171 years is (15.45 °C).
2- The overall average of Riyadh temperature is (25.21 °C), No. of years was above the average is (79 yrs.) and below the average is (81 yrs.).
3- Max average of temp amongst of the last 171 years is (27.78 °C).
4- Riyadh city is hotter compared to the global average.
5- Riyadh city trend to hotter Year-Over-Year.
6- Moving average used in this page is based on 15 yrs. For both locations (Riyadh, and global).

*moving average: I used MS Excel to calculate moving average by made this formula (SUM(avg_temp)/COUNT(avg_temp)) or we can use Average() function to calculate it!, and also there is feature in plot is Trend line and choose period for moving average.

Thank you!

Appendix

Query used in this project

Extract data from Global data:

```
SELECT

g.*,

(SELECT MIN(gg.avg_temp) FROM global_data gg) AS Min_global,

(SELECT AVG(gg.avg_temp) FROM global_data gg) AS Avg_global,

(SELECT MAX(gg.avg_temp) FROM global_data gg) AS MAX_global

FROM global_data g
```

Riyadh city on average compared to Average of global temperatures as yearly basis, +Handling or Replacement / Manipulation of missing values (NULL, NaN) by General average for Riyadh city:

*I can handle the missing values in Excel using IF formula, but I preferred SQL.

```
SELECT

*,

CASE WHEN t1."Manipulate" > t1."Global_temp" THEN

'hotter compared to the global average'

ELSE 'Cooler on average compared to the global average' END AS "Classification"

FROM

(SELECT

*,

CASE WHEN c.avg_temp IS NULL THEN

(SELECT AVG(c1.avg_temp) FROM city_data c1 WHERE c1.city = 'Riyadh')

ELSE c.avg_temp END

AS "Manipulate",

(SELECT g.avg_temp FROM global_data g WHERE c.year = g.year) AS "Global_temp"

FROM city_data c


WHERE city = 'Riyadh' -- My city

) AS t1
```

Average of changes YoY of Riyadh city:

```
SELECT
AVG(t2."YoY")
FROM
(
SELECT
*,
t1."Manipulate"
-
LAG(t1."Manipulate") OVER (ORDER BY t1.year)
AS "YoY"
FROM
(SELECT
*,
CASE WHEN c.avg_temp IS NULL THEN
(SELECT AVG(c1.avg_temp) FROM city_data c1 WHERE c1.city = 'Riyadh')
ELSE c.avg_temp END
AS "Manipulate"
FROM city_data c

WHERE city = 'Riyadh' -- My city
) AS t1) AS t2
```