# Correlation between Co2 emissions and temperature in India

**Name – Aniruddha Maiti**
**Matriculation number – 23195586**
**Username - uq82umir**
**AI Software Applications**

# Introduction

**Analyze the relationship between Co2 emissions and temperature in India**

- In 2019, 21 out of 30 most polluted cities are in India, according to a worldwide survey.
- 1.1 to 5.1 degrees temperature rise is expected in India by 2100.

**Objective:**

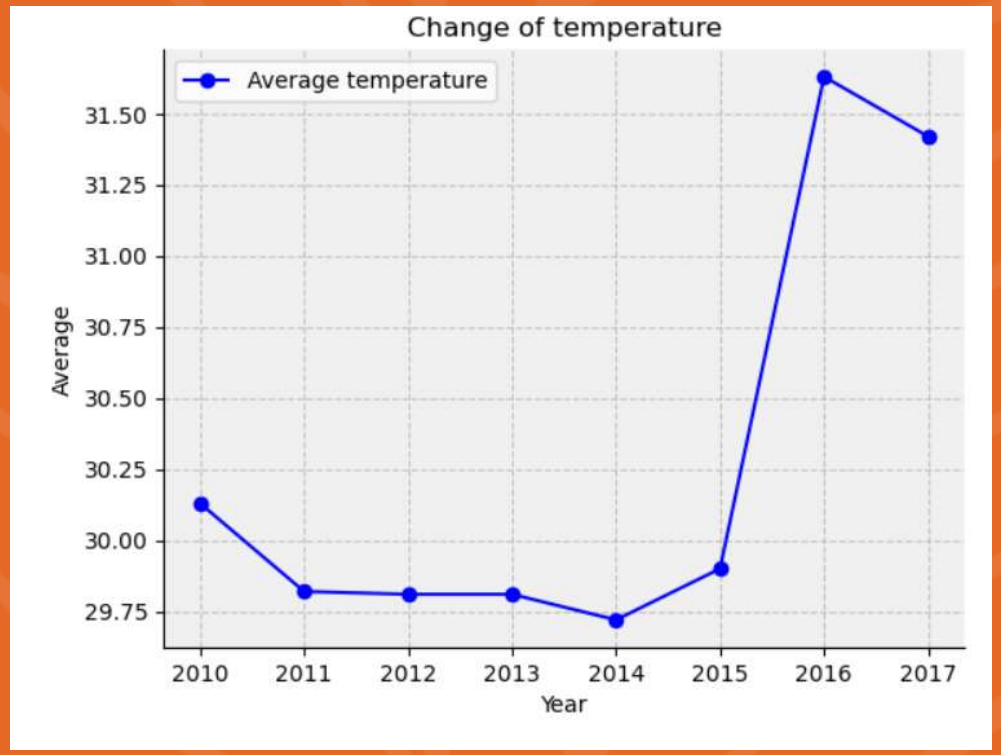
- Check the behavior of temperature and Co2 emissions between 2010 to 2017.
- Use statistical metrics to prove or disprove correlation between the two factors.
- Is there a change in both temperature and Co2 emissions between 2010 and 2017 ?

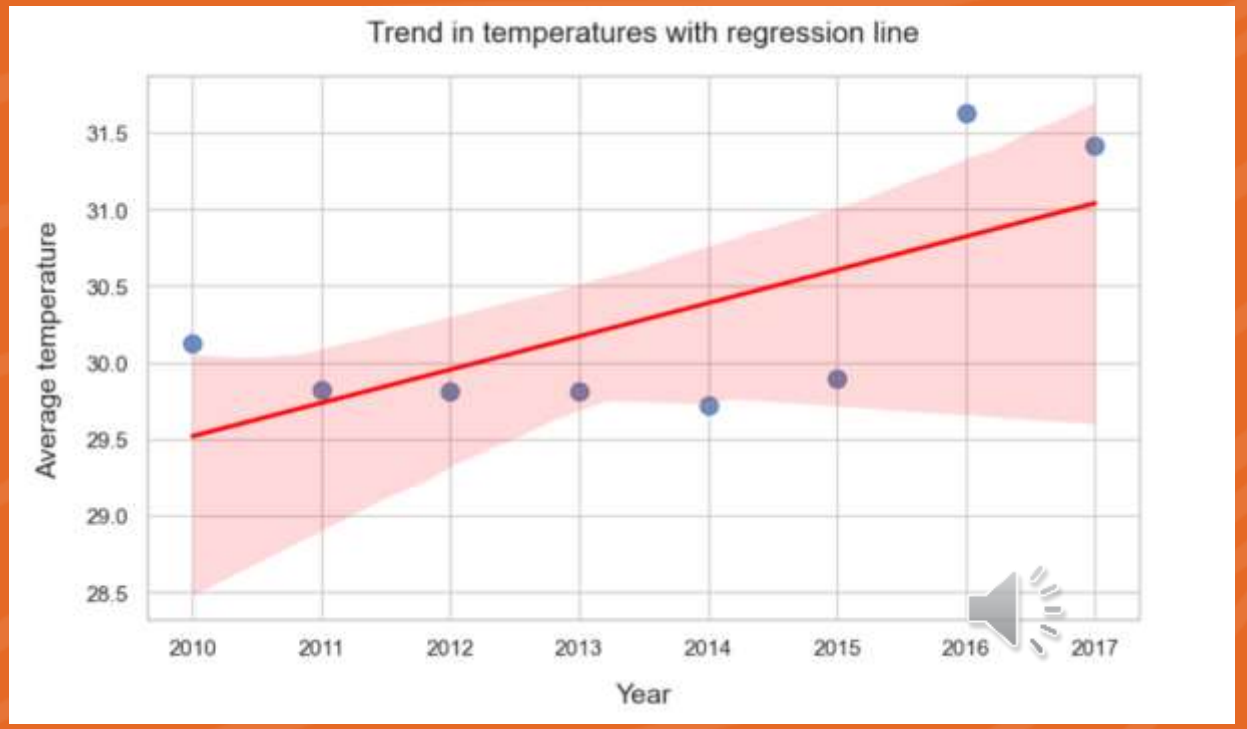
# Methods and materials

- Collect data of temperature and Co2 emissions in India from Kaggle: open source platform
- Clean the data – remove and rename unwanted columns to ease of use
- Arrange and sort data based on a common base – Year (time)
- Create pipeline, test files and CI

# Question 1: Is there a change in temperature between 2010 and 2017 ?

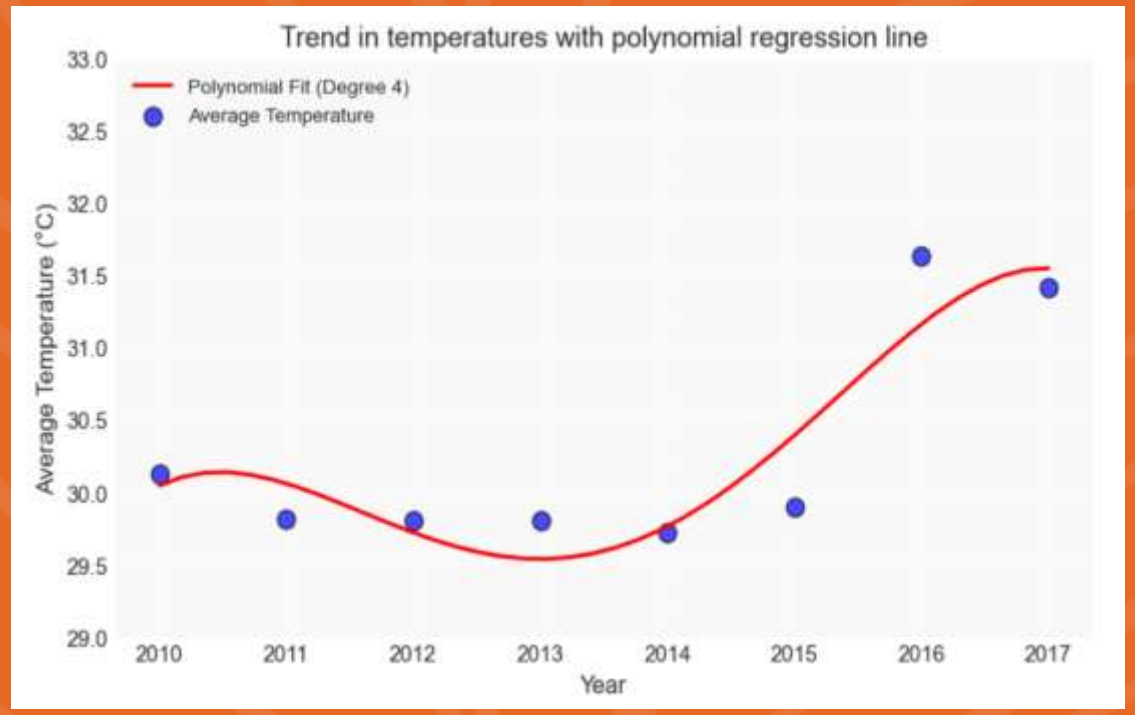
- Between 2011 and 2014 there is a directional behavior
- After 2014, there is a constant increase in temperature
- A regression graph is used to analyze the behavior but it cannot capture the non-linear data



# Question 1: Continuation

- In order to capture and analyse the non-linear: we use a polynomial regression of degree 4.
- Statistical metrics demonstrate a high correlation : more than 0.5 and covariance matrix implies a positive relation

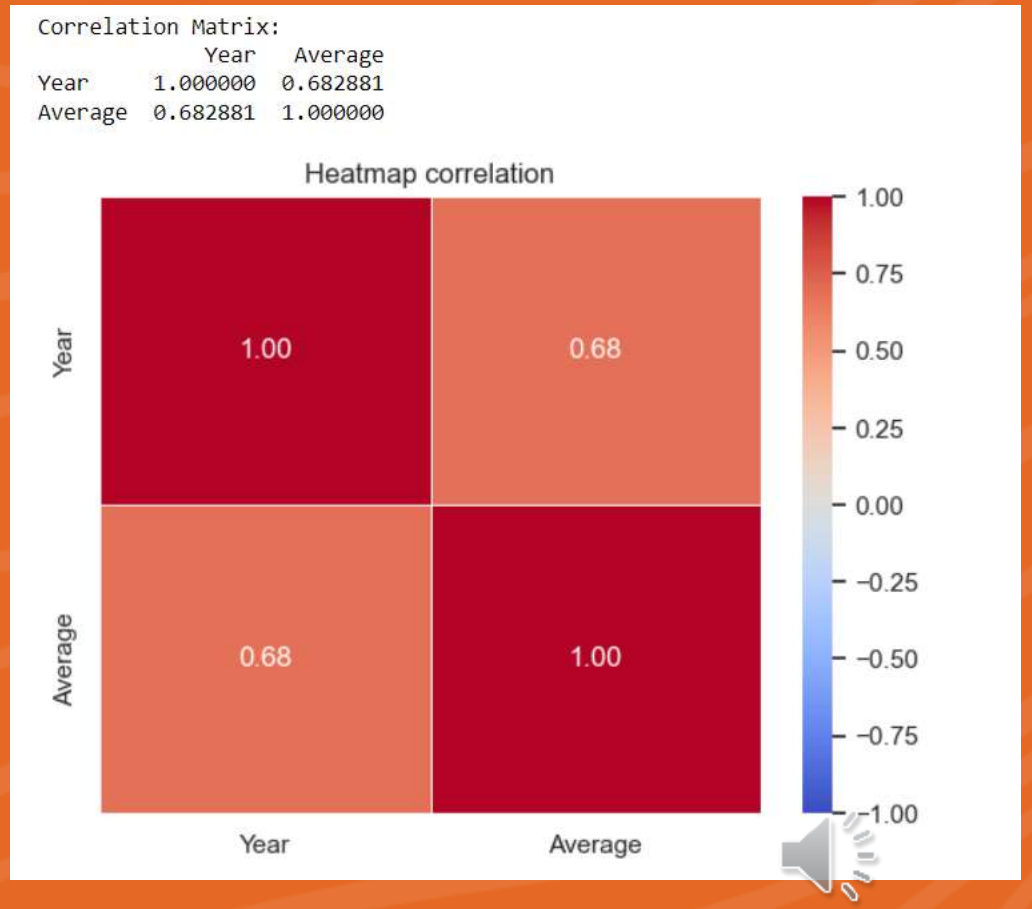



```
Covariance Matrix:
              Year   Average
Year      6.000000  1.304286
Average   1.304286  0.608000
```

```
Correlation coefficient between Year and Average temperature: 0.6828814611969964
Pearsons correlation: 0.683
R-squared score: 0.8525143486692721
```

# Question 1: Continuation

- The correlation matrix indicates a moderately positive relation.
- The heatmap illustrates the coefficient and correlation visually.



# Result
## - Inference 1

**After considering the visual graphs and statistical metrics, we attempt to prove or disprove correlation:**

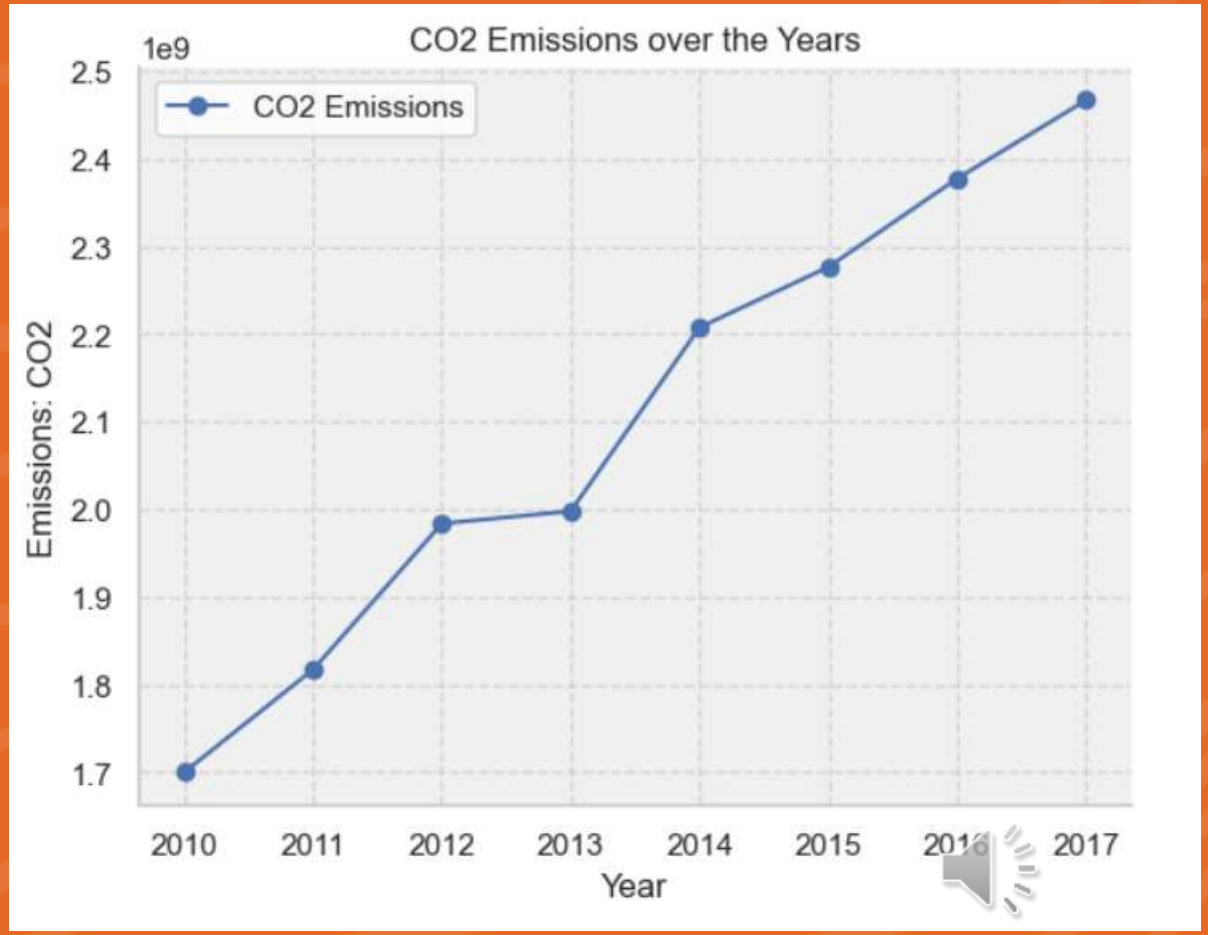
- In the initial time period between 2011 and 2014: we can see a directional relation as it was increasing and decreasing simultaneously.
- After the year 2014, we can see an upward trend of temperature.
- Statistical metrics like correlation coefficient and scatter plots imply a positive correlation mathematically.
- Result: Temperature is positively correlated with time.
- Expectation: There is a chance of further increase of temperature.

# Question 2: Is there a change in Co2 emissions between 2010 and 2017 ?

- With respect to time, we can see an upward trend in Co2 emissions with respect to time.
- The covariance matrix describes a strong relationship between Year and Co2 emissions.
- The high magnitude of the covariance suggests a very strong relation in absolute terms

Covariance Matrix:

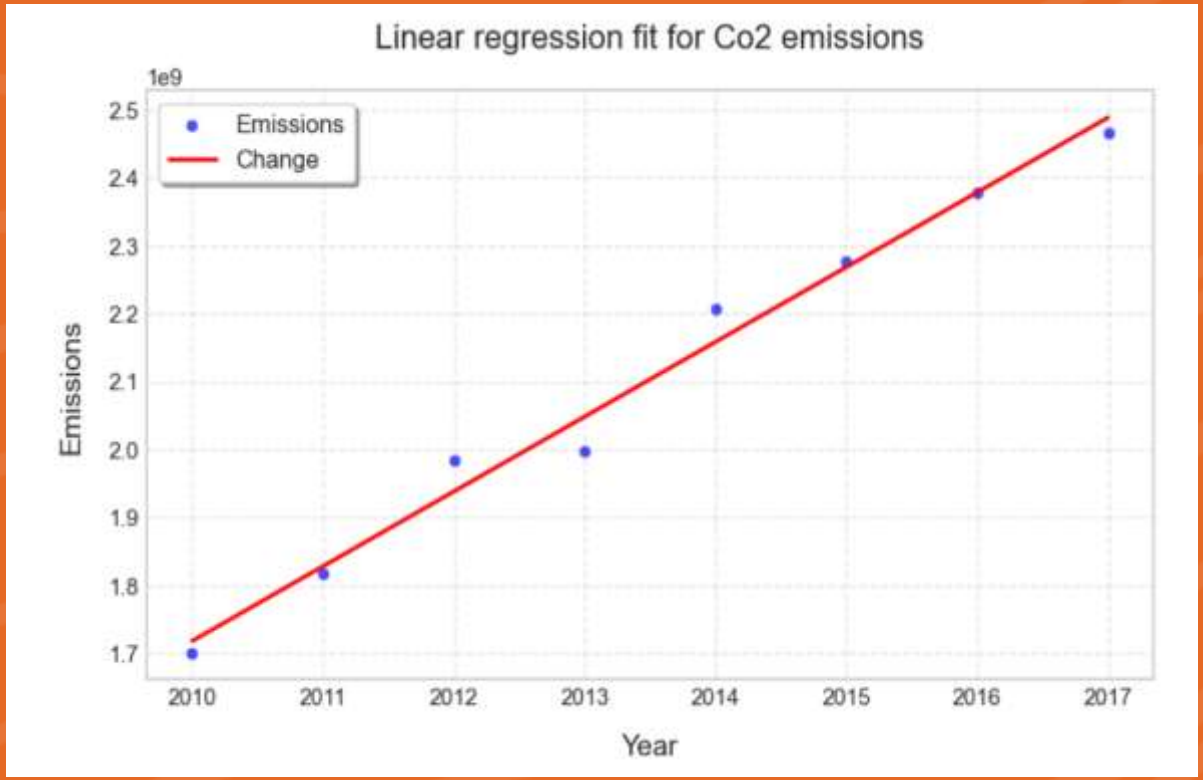
| | Year | Emissions |
|---|---|---|
| Year | 6.0 | 6.611534e+08 |
| Emissions | 661153402.0 | 7.399956e+16 |



# Question 2: Continuation

- The linear regression graph for Co2 emissions is useful for making predictions and forecasts.

- It can also be used to understand the trend and behaviour of Co2 emissions and the government can take important policy decisions.

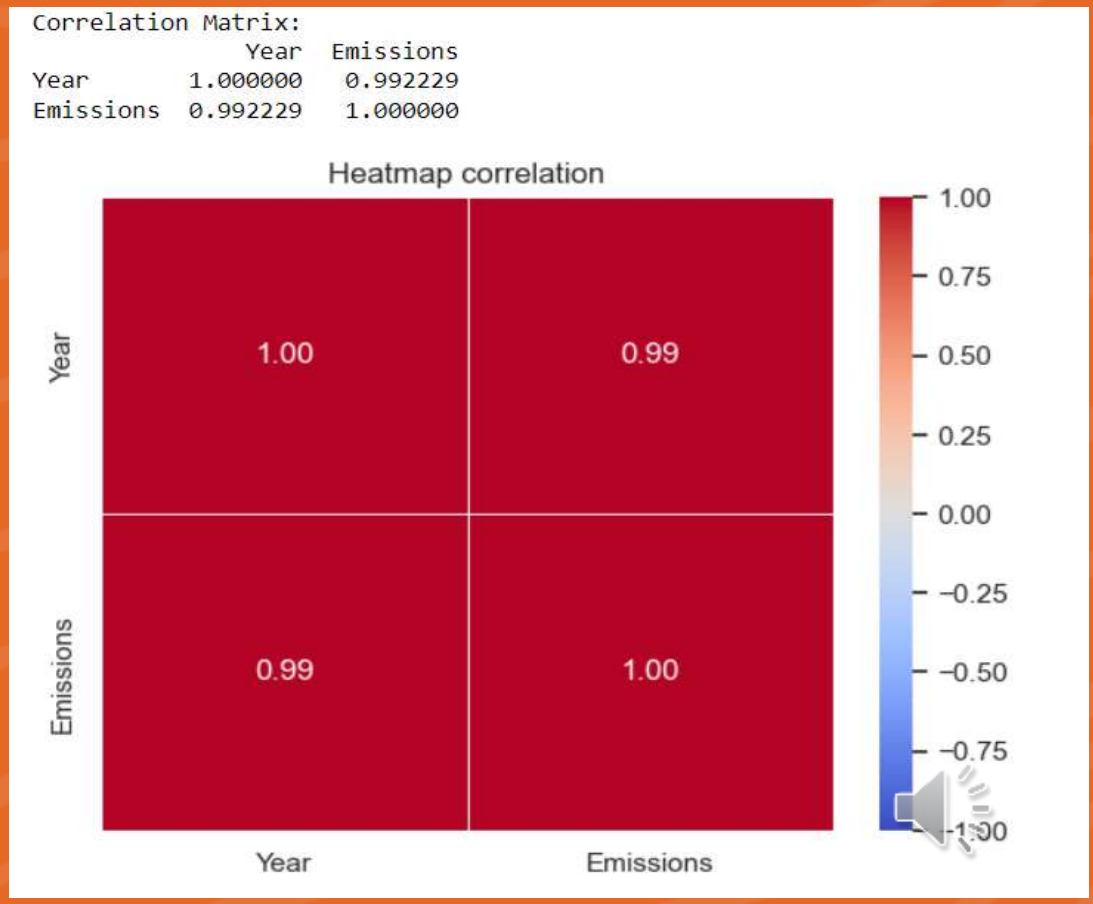


# Question 2: Is there a change in Co2 emissions between 2010 and 2017 ?

- The correlation coefficient and R-square score indicate a strong positive relation between year and Co2 emissions.
- The high correlation implies a strong and consistent association between the two variables, indicating that they tend to increase in the same direction.

```
Correlation coefficient between Year and Co2 emissions: 0.9922292926400624
Pearsons correlation: 0.992
R-squared score: 0.984518969173002
```

```
Correlation Matrix:
                Year  Emissions
Year        1.000000   0.992229
Emissions   0.992229   1.000000
```



# Result
## - Inference 2

**After considering the visual graphs and statistical metrics, we attempt to prove or disprove correlation:**

- With respect to time, we can see a constant upward increase of Co2 emissions.
- With the help of visualization graphs: we can report a positive correlation of Co2 emissions with respect to time.
- Statistical metrics like correlation coefficient support the positive correlation mathematically.
- Result: Co2 emissions are positively correlated with time.
- Expectation: There is a chance of further increase in Co2 emissions

# Question 3

- Is there a correlation between Co2 emissions and temperature between 2010 and 2017 ?

# Result

## - Inference 3

**After considering the visual graphs and statistical metrics, we attempt to prove or disprove correlation:**

- Considering the common base of time - Year. We can conclude that both Co2 emissions and temperature are correlated to each other.
- Using the scatter plots (and heatmaps) of both temperature and Co2 emissions: we can see an increase in both the factors.
- The statistical metrics such as Pearson's coefficient, R-squared score and covariance matrices provide enough evidence and support that both the factors are increasing during the time period of 2010 - 2017. Conclusively, they are correlated to each other.

# Conclusion

**Based on the previous two questions and all the information gathered so far:**

- The visual graphical plots: scatter plot and heatmaps on a common base of time - Year, relations have been described by covariance matrices, correlation matrices and results derived from Pearson correlation coefficient, R-square score, we can arrive at the conclusion that both temperature and Co2 emissions are related to each other.
- In summary, the data regarding both temperature and Co2 emissions indicate a moderate positive correlation between both temperature and Co2 emissions with respect to year, suggesting a constant increment and positive correlation. These findings suggest a robust association between the variables, emphasizing the strong influence of time on CO2 emissions.

# Further research

Although we could draw a conclusion from the gathered information and insights, certain questions are still prevailing:

1. If temperature and Co2 emissions are correlated to each other, how come there is a constant rise in Co2 emissions between 2010 and 2014 despite a stagnant directional behavior of temperature in the same phase of time?

2. Are either of the factors causing a change in the other factor ?

# Future endeavors

In order to overcome the limitations stated above, the following ideas can be pursued to prepare more extensive data engineering projects:

- Which factors are exactly causing an increase in temperature ? Is it air pollution, deforestation, waste management or natural factors ?
- Is there a particular location where Co2 emissions and temperature correlation is more ?
- As the graphs demonstrated high correlation of Co2 emissions: what are the sources of Co2 emissions ?

# Thank you