Our Challenge/Problem That We Want To Solve

Air Quality Monitoring

Introduction:
The quality of air can heavily affect people in their daily life. And for some people with chronic conditions, polluted and dry air can even be close to deadly. Air is all around us, and we need to monitor its quality to see if it could be potentially dangerous for people. Our project will organise the collected data over the years, to show what strategies were used over the years, and which ones worked in favour of a positive outcome. We will collect data and organise it into data frames and graphs, to better understand and visualise it.

Problem statement:
What are we trying to achieve in this project?
Our main goal in this project is to create visual representations of how the air quality has been affected, and its changes during the years, and further predict how it will be, to take that into account and work with it to improve the situation. We are going to also use factors that contribute to air quality directly, to see which ones are beneficial and which are not, and then predict future data with the information that we have collected, for future use and example.

Expected outcome:
Out of this project at the end we expect to have some visualisation of how and which factors have affected air quality throughout the years, like a graph, and a dataframe, and also a report stating how we achieved this outcome, and an analysis of the data that will be presented. Also, we will have a prediction of how the positive changes will affect the quality of air, for future reference.

Python tools:
What python tools are we going to use?
To visualise the data that we will collect, first, we will work with it using Pandas. After the data that we have will be ready for use, we will then proceed to use MatPlotLib, the library that we were recently introduced to, to visualise this data and make a very well organised graph, that will have a title, labels, and other components to make it not only a visual representation but also coherent and easy to understand. We will also use NumPy and Pandas to calculate the further data.

Strategy:

- Start first by understanding fully what problem we are looking to solve, and what it has been by researching information on it.
- Start looking into where we can get the data necessary for our project
- Collect all of the data, and put it in one place for easier access and a better understanding
- Start processing the data, getting rid of missing values, structuring it properly, and etc.

- After all of the data is collected and is in order, start visualising it and make a linear regression showcasing what we have collected.
- Write a report, in which we state what we had done, and an analysis on the data, as well as what could be done to improve the problem.
- Calculate future data (predict)

What tools and techniques will you use?

- **MatPlotLib for a visualisation of the data**
- **Pandas for working with and structuring the data**
- **Google documents for report and notes**
- **Github for tracking our project.**
- **Numpy (and Pandas) for predicting the future Data.**

Data Strategy Updates:

For predictions we will use

'Name' column of our data set.:
We will be using this column, because it specifies the type of factor that is being measured, which we will then use to see how to improve it.

Measure column of our dataset:
We need this column to that we can see what units the factor is being measured in

'Data Value' column of our data set:
It shows the measured value of a specific type of factor we are trying to improve.

Geo place Name column:
We need it because it is also a factor that heavily contributes to Air quality, since the climate and conditions can be very different from place to place.