Global Air Pollution 2010-2017

Introduction –

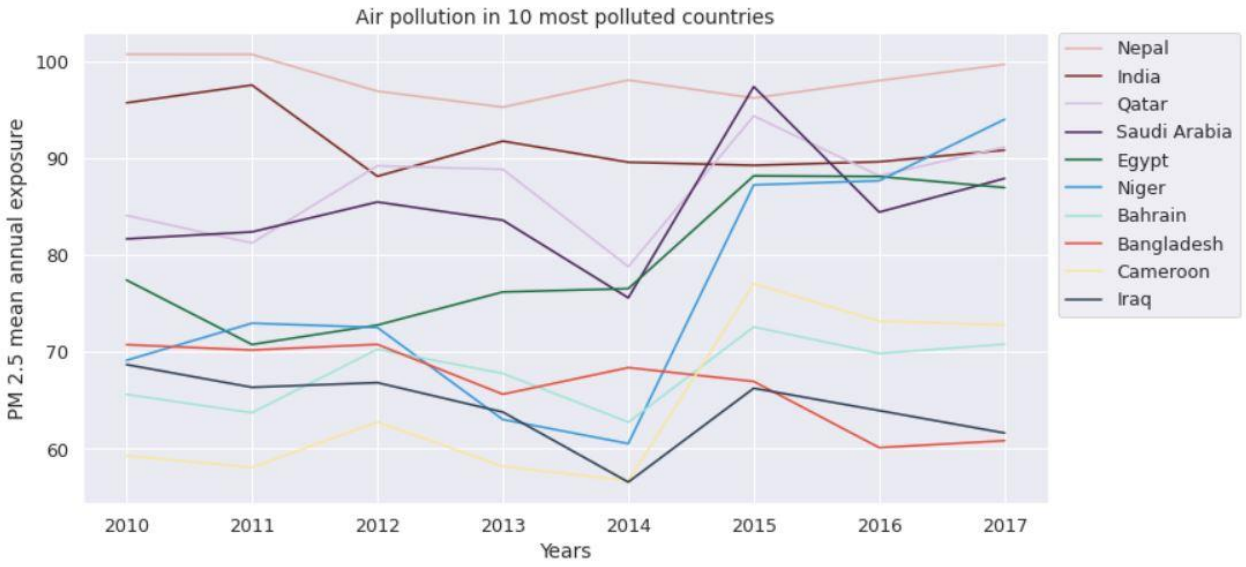
Air pollution is the need of the hour. According to WHO, air pollution kills about 13 people every minute in the world, more than 3 times the combined amount of Malaria, Tuberculosis and AIDS. It has also caused a tremendous increase in the global warming levels across the world, but the worst is yet to come. According to scientists, the global health is said to go down by 3% by 2025. Thus, it is a very important time to analyse and stop the unnecessary increment of air pollution. This project performs an analysis on the global air pollution levels from 2010-2017.

Data sources and description –

Credits to the dataset goes to - "Karl Weinmeister". Please find the dataset attached at - https://www.kaggle.com/kweinmeister/pm25-global-air-pollution-20102017. The data has been edited and modified to suit the purpose of the project. This data is a measure of mean annual exposure (micrograms per cubic meter) of PM2.5 across several countries from the year 2010 to 2017. PM2.5 also called as Particulate Matter 2.5, is a mixture of solid, liquid and aerosol particles that are suspended in the air. The 2.5 refers to their diameter of 2.5 micrometer. PM 2.5 is probably the most important factor in air pollution detection. For more information regarding PM 2.5 please visit - https://www.airveda.com/blog/what-is-pm2-5-and-why-is-it-important.

Methodology and Results –

The data contains a total of 197 countries with their annual mean PM 2.5 exposure values from 2010 – 2017. The first step is the plot the data across the timespan. However, plotting 197 data lines in a single plot won't make any sense. Hence, the data has been sorted according to the mean annual exposure of PM 2.5 and the top 10 most polluted countries have been plotted across the timespan of 2010 – 2017. According to WHO, the safe limit for mean annual PM 2.5 exposure is 35 micrograms per cubic meter. Following is the plot of the 10 most polluted countries in the timespan of 2010-2017 –



By observing the line plot, following points can be deduced -

- **Nepal** is the most polluted country across the timespan of 2010-2017.
- In 2015, **Saudi Arabia** caused more pollution than Nepal. However, the pollution level of Saudi Arabia decreased after 2015.
- **Niger** and **Cameroon** faced the highest rise in pollution level from 2010 to 2017. The pollution of **Niger** rapidly increased to become the 2nd most polluted country in 2017.
- **Bangladesh** and **Iraq** were able to reduce their annual pollution level across the timespan of 2010 to 2017.
- By observing the 10 countries, it can be seen that the pollution level varies from **60 micrograms** to **100 micrograms** per cubic meter. This hints at the fact that an overwhelming number of countries are doing good to moderately well in countering air pollution.

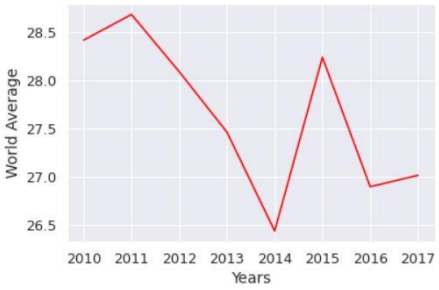
Since, mean annual exposure doesn't give an idea of how good or bad a country performed in the timespan; the data was sorted according to the change in the pollution level from 2010 to 2017. The data was sorted and adjusted, where high positive value means better performance in countering air pollution while low negative values hints towards failure in combating air pollution. Following is the sorted list of top 5 and bottom 5 countries and their performance in combating air pollution.

	Country Name	Country Code	2010	2011	2012	2013	2014	2015	2016	2017	mean_val	Change in Pollution
0	Sri Lanka	LKA	31.033781	31.232075	30.083973	28.898228	26.922777	25.337344	14.025445	11.099617	24.829155	19.934164
1	China	CHN	69.479570	70.542029	63.827017	65.514550	59.767368	59.063422	52.211322	52.664596	61.633734	16.814975
2	Bangladesh	BGD	70.764996	70.210549	70.793909	65.645083	68.395377	66.965749	60.126275	60.845785	66.718465	9.919211
3	Vietnam	VNM	39.365397	38.486975	37.227452	36.835906	34.896015	32.466444	30.241312	29.626728	34.893279	9.738669
4	Pakistan	PAK	68.005190	68.544936	62.376646	61.022163	59.518363	60.085893	58.632764	58.282419	62.058547	9.722772

	Country Name	Country Code	2010	2011	2012	2013	2014	2015	2016	2017	mean_val	Change in Pollution
192	Chad	TCD	56.586627	55.195160	58.675388	53.056407	51.247169	68.682394	64.770769	66.029206	59.280390	-9.442578
193	Egypt, Arab Rep.	EGY	77.446619	70.786356	72.790569	76.213654	76.559601	88.213388	88.148500	86.999452	79.644767	-9.552833
194	Cameroon	CMR	59.293403	58.076901	62.776697	58.186338	56.715306	77.051743	73.205291	72.793096	64.762347	-13.499693
195	Nigeria	NGA	52.609064	50.973089	56.129638	49.740689	48.633374	75.398992	71.369601	71.798174	59.581578	-19.189110
196	Niger	NER	69.150177	72.982767	72.544645	63.031717	60.541357	87.281282	87.708472	94.053818	75.911779	-24.903641

As it can be seen from the lists, **Sri Lanka** and **China** did extremely well in combating air pollution across the years, while countries like **Cameroon**, **Niger** and **Nigeria** have failed extremely in abating air pollution.

There have been mixed performances from the countries. However from the following plot, we can safely infer that the world average of air quality has increased by a small 0.4%, after suffering a blow in 2015.



Further, unsupervised learning has been used to cluster the countries according to set pattern. The clustering has been done using K Means and it can be used to group the data into 3 clusters, namely –

- Lowly Polluted Countries (119).
- Highly Polluted Countries (16).
- Averagely Polluted Countries (62).

The biggest cluster is of 119 countries which are the ones with mean annual exposure of PM 2.5, well below the safety level of **35 µg/m^3** set by WHO. The second biggest cluster is of averagely polluted countries whose mean exposure varies from **27 – 53 µg/m^3**. The third cluster, which is the smallest cluster and comprises of only 16 countries are the actual pollution hotspots of the world and contribute to 37.31% of the world mean air pollution.

After the clustering has been performed, a classification model is trained on the data. The classification model used is Logistic Regression, with an average accuracy of 81%. For any new data of a country not present in the dataset, the model can predict the cluster label for the new data. This can help the country in taking necessary steps in abating further air pollution.

Conclusion –

Through the project, a comprehensive analysis of the world air pollution levels were done and machine learning algorithms were used to cluster the data into specific groups, and for any new data, the classification model could predict which group it belongs to. This can help the new country take proper organized steps in combating air pollution.

The plots and tables mentioned above brought to light several new inferences. However, sadly the average air pollution level has only decreased by 0.4% in a span of 8 years. This is due to the fact that although many countries were successful in controlling their air pollution levels, several other countries failed disastrously to combat air pollution. There are multiple reasons for not being able to maintain the pollution levels, but the primary one, is the absence or shortage of funds that these under-developed and developing countries face. To counter that some new policies can be undertaken by the United Nations, which help these countries grow and control their levels of air pollution. Following are a few suggestions that could curb the severe cases of air pollution –

- United Nations and respective world organizations could take it upon themselves to identify 10 most polluted countries and help with annual funds. These funds could be used for developing and installing air filtration plants across the countries.
- Electric vehicles and the stakeholders could use these countries to not only boost their sales but it will also help to minimize the world pollution level. New companies could enter the market; new factories could be made to help in the growth of electric vehicle culture.

These small steps could help the countries progress towards a more cleaner and pollution free future. This would also in turn reduce the world pollution levels in the years to come. Finally, it is about working together for a better and cleaner Earth and not about individual gains and profit.

Developed by – Sourin Roy

Project link - https://bit.ly/3gFkB9i

Kernel link - https://tinyurl.com/y7nu3otc

Find me at - https://tinyurl.com/y9rm346p