**Maxim Rohit**

**Data Science**

**Understanding Pollution Pattern of Maharashtra, India**

**Summary:**

Indian Government on 8th December 2018 released few disturbing facts about how the increase in air Pollution had a degrading impact on Indian Living conditions.

Air pollution can be attributed to of 12.5% of deaths in India[1].

12.4 lakh deaths in India attributed to air pollution in 2017, of which 50+% were in individuals younger than 70 years.[1]

Life expectancy in India has gone down by 1.7 years due to health loss caused by higher (than accepted) air pollution level. [1]

We will attempt to analyse localize impact on region of Maharashtra in detail.

1. Analyse & Categorize the rate at which air pollution is increasing in Various cities & Towns of Maharashtra using clustering techniques like hierarchical clustering and K-mean.
   1. Validate how city progression maps to its Air pollution levels?
2. Understand how various pollutants and other features co-relates to each other using PCA
3. Is their time based (seasonal & otherwise) variation in the spread of various pollutant, via EDA.
4. Use timeseries analysis to project the pollution growth pattern in the coming year and predict when the air pollutant level goes beyond breathable standards.

**Research Question: -**

1. Categorize various regions within Maharashtra, India based on Air pollution trend.
   1. Separating out Healthy, unhealthy and risky cities into categories, might enable us to put similar preventive measure.
2. Analyse the difference between these categories of regions w.r.t
   1. Concentration Type
   2. Rate of Change (Yearly, Quarterly etc)
   3. Population
   4. Elevation from Sea level
   5. Seasonal Time of Year
   6. Buying trends of automobiles/Impact of Additional Vehicles
   7. Rainfall levels
   8. Longitude
   9. Latitude
   10. Others
   11. Understanding if any one particular variable has a major say in the pollution trends.
3. How does cities like Pune (An IT Hub in India) & Mumbai (An Overall commercial Hub) fair against Neighbouring upcoming Towns (Nasik, Nagpur) versus the less progressive one like Ahmednagar, Solapur over the last couple of decades?
   1. Understanding the extent of impact of a city progression, might help us plan out city development (may be put better restrains) better.
4. Understating if One pollutant impacts the presence of another.
   1. Understanding pollutants correlation and impact of other variables may help us modulate them better?
5. Predicting by when the Air pollution levels will go beyond acceptable?
   1. Gives us a time frame to put-up a plan & to act on it before thing become irreversible.
6. Understanding the seasonal behaviour of the pollutants?
   1. Different seasons might call of different actions, only when we know how they differ can we effectively counter them.

Details:-

Aims(general, not objective) and Objectives(quantifiable):-

Work leading upto the project

Data source

Processing data and analysis: cleaning management, meathod used to avoid bais due to missing data or confounders

What data analysis methdos will eb used

Wat software will be used