REGRESSION ANALYSIS PROJECT OUTLINE (GROUP- 4)

- **Project Title:** Air Pollution to Mortality: A Case Study Using Regression

- **Data Description:**

The name of the dataset is "air pollution". There are **16 variables** in this dataset with 60 records with no missing values. The description of variables is given in the table below:

Variable No.	Variable	Description
1.	PREC	Average annual precipitation in inches
2.	JANT	Average January temperature in degrees F
3.	JULT	Average July temperature in degrees F
4.	OVR65	% of 1960 SMSA population aged 65 or older
5.	POPN	Average household size
6.	EDUC	Median school years completed by those over 22
7.	HOUS	% of housing units which are sound & with all facilities
8.	DENS	Population per sq. mile in urbanized areas, 1960
9.	NONW	% non-white population in urbanized areas, 1960
10.	WWDRK	% employed in white collar occupations
11.	POOR	% of families with income < $3000
12.	HC	Relative hydrocarbon pollution potential
13.	NOX	Relative nitric oxides pollution potential
14.	SO.	Relative sulphur dioxide pollution potential
15.	HUMID	Annual average % relative humidity at 1pm
16.	MORT	Total age-adjusted mortality rate per 100,000

- **Aim of Study:**

  Here we want to **regress variable 'MORT' (Mortality)** by **all other variables as regressors using as less number of regressors as possible.**

  Here after getting data, we have checked the correlation between MORT and other variables. Except for two or three variables, other correlations are quite high. So here linear model can be fitted.

  Here we have done some preliminary analysis on our data and we have found that our data suffers from multicolinearity issue for some variables. Our intention will be to make a model by taking necessary steps.

  If in our further analysis we get any other issue, we shall take necessary steps to remove that issue and new model will be made.

- **Data Source:**

  The actual source of data is:

  McDonald, G.C. and Schwing, R.C. (1973) 'Instabilities of regression estimates relating air pollution to mortality', Technometrics, vol.15, 463- 482.

  But the data can be found on this given link:
  http://lib.stat.cmu.edu/datasets/pollution *(Please click on this link)*

- **Data File:**

  https://drive.google.com/file/d/1c8W_J1hDVjWwqCLyey7ZQ2qo7Efdzh0y/view?usp=sharing

  (Please click on this link)

- **Group Details:**
  - ✓ Members:
    1. Abir Naha (201257)
    2. Anjan Kumar Kayal (201271)
    3. Arkonil Dhar (201279)
    4. Koyel Pramanick (201333)
    5. Suchismita Roy (201440)