**Project 4**

**Problem Statement**: Show the correlation between pollution and lung cancer and predict the long term affects if not changed.

**Objective**: The objective is the show which pollutions maybe affecting the health of people the most and how we maybe able to change this.

**Key Technologies**: Python Pandas, Sklearn, Tableau, possible HTMl

**Data set**: Harvard Dataverse. Air Quality-Lung Cancer Data

**Data size:** There are over 2000 data points including the level of lung cancer, the various different air pollution (PM10, SO2, NO2, O3, CO, CN, CS2) plus several other pollutants.