Significant Paper Report

Article Title: PM2.5 air pollution and cause-specific cardiovascular disease mortality

Source: International Journal of Epidemiology

The article dives into the relationship between ambient PM2.5 exposure with defined cardiovascular disease mortality. In the study they had 565 men and 477 women, who were between the age of 50 and 71. The participants came from six states: California, Florida, Louisiana, North Carolina and Pennsylvania. The following urban cities were also selected for the study, Atlanta, Georgia and Detroit, Michigan. The cities and states were selected based on data that was analyzed from the National Institute of Health (NIH-AARP) Diet and Health Study. This dataset had members primarily active the states and cities, and they also represented many cultures.

To analyze the data the researches first used a hybrid land use regression geological statistical model to estimate the PM2.5 amount. Then they use a multivariate cox regression model to estimate the risk of mortality in patients with a 95% confidence interval. Annually the social security master death file was looked into to determine the mortality rate.

The study also included questionnaires, which collected the linkage to national databases for mortality follow ups, along with census data. They were asked about age, health problems, and industry that they worked in.

What they found was that for each 10 ug/m3 increase in PM2.5, there was a 14% increase in mortality from a stroke and a 16% increase in mortality from ischemic heart disease. They concluded that long-term exposure to PM2.5 can increase the chance of mortality for cardiovascular diseases, and that we should spend more time looking at ways that we can improve the air quality around us.

In class we spend great amount of time talking about the applications of Data Science and how it can be used to further advancements in society. For example, how researchers are attempting to cure cancer with the help of machine learning. In a similar fashion, a smaller application of machine learning can be to look at data points such as those in the article and make changes that can impact multiple people across the globe. The article shows the effects of air pollution on the health’s of the general population. We can then go further and dive into what things we can do to prevent the air pollution which would improve patient health and create a more sustainable future for everyone.

As well as talking about the use-cases of data science, we also learned about selecting different models in class, or how to select different weights for categories, and how to get a line that best fits our data. These similar points can be found in the article on how they tried different models and what different weights to the attributes worked best for them.

We paper is in-line with our project. In the paper, they discuss PM2.5’s effect on cardiovascular disease and in our paper, we discuss the effect of PM2.5 on airborne diseases. The model helped us formulate ideas on what factors we should take into consideration, such as mortality rate for different illnesses, location, age, type of population. These factors helped us select a model that would best fit our problem statement. We also tried to utilize the model suggestions that were provided in the paper, but those did not fare as well for our problem statement.

Overall the paper provided us with great insight about our project, along with reinforcing the material that we had learned in class. The findings in the paper helped push all of us to work towards understanding our problem statement and answering our problem as best we could in the timeframe that we had. Along with that, we were shown how Data Science can be used to help inform people into making decision that can benefit themselves in the future.