

# **PROJECT EXTENSION PLAN**

**ANISH DIXIT**

## **MOTIVATION**

This project undertakes an analysis of wildfires in a certain area and assessing their impacts on the surroundings. Specifically, I am looking at fires within a 1250 mile radius of the city of **North Platte, Nebraska**. In the first part of the project, I conducted a common analysis of wildfire extent and calculated a smoke estimate which aims to explain the spread of the fire to North Platte, in a quantitative manner. Now, I plan to extend this work by trying to apply it to certain socio-economic factors, giving it a human centered angle, which can actually impact the community of North Platte as well.

Investigating the spread and occurrence of wildfire smoke is essential for comprehending the dynamics of these natural disasters. Wildfires are often exacerbated by climate change, and their smoke contributes to air pollution. Analyzing the socio-economic impact provides insights into the broader implications of climate change on local environments and communities. Studying the spread of wildfire smoke allows for a better understanding of its health implications. This includes respiratory problems, cardiovascular issues, and the potential long-term health effects on the affected population. By identifying the socio-economic impact of wildfire smoke, the analysis can contribute to more effective risk mitigation strategies. This could involve improved emergency response planning, public health interventions, and land management practices to reduce the frequency and severity of wildfires. Understanding the economic consequences on nearby cities helps in allocating resources efficiently. This information is vital for local governments, emergency services, and healthcare providers to allocate funds and personnel appropriately.

The analysis can contribute to building resilient communities by providing data-driven insights into the socio-economic vulnerabilities that arise from exposure to wildfire smoke. It can serve as a tool for raising public awareness about the potential socio-economic impact of wildfires. This information is crucial for fostering a sense of responsibility among residents and encouraging proactive measures to reduce the risk and impact of wildfires. The findings can inform the development of policies and regulations aimed at mitigating the socio-economic impact of wildfire smoke. This includes zoning regulations, building codes, and land-use planning to minimize the vulnerability of communities to wildfires.

## **IMPACT FOCUS**

The focus area I am choosing to base my analysis on is **Healthcare**. I chose this domain because I believe the smoke estimate calculated in the common analysis would have the maximum impact and correlation with the health of the residents of North Platte. As mentioned in the motivation, standalone wildfires can affect the health and well being of the individuals residing in the city, and can have detrimental effects on the long term fitness as well. Of course, wildfires are not the only contributors to this wane of health, it is a much larger systemic problem, but it can still serve as one indicator.

### **Problem Statement**

- 1. I intend to study occurrences of bronchial cancer and tuberculosis in the North Platte area across years, try to map and correlate it to the yearly smoke estimate in part 1, and predict how future frequency of cancer cases/mortality rates would look.*
- 2. I also intend to correlate cancer occurrences with demographic trends and gauge what percentage of the population of the area was being afflicted.*

I decided to expand upon these diseases specifically as these are **respiratory illnesses**. They can be directly brought about, cause or if inherently present, intensified by excessive smoke in the region, which my previous work can help with. I also will try to

A future prediction can help medical services be prepared for sudden spikes in cases in a cyclic manner if any seasonality exists, or if there is simply a generic upward trend in the number of cases found. This work could help the health services in the North Platte city and Lincoln county better understand how wildfire occurrences could be a potential harbinger of respiratory diseases and how the population is being affected by the same. Efforts can be made to better handle these humanitarian problems.

## **DATASETS AND MODELING**

To carry out this analysis, I will use health data provided by the state of Nebraska. The Nebraska government has very well documented data reports relating to healthcare, disease occurrence for historical data, and with a granularity across counties, gender and age demographics. The department of revenue has well maintained census data for each city in the state which we will use.

### **Links to Data**

1. [Public Health Reports : Nebraska Department of Health and Human Services](#)
  - This is an inventory of Nebraska's public health data. The website contains a lot of statistics and reports relating to different health parameters such as abortions, dental visits, cancer, oral health, pregnancy risk, motor vehicle accidents, tuberculosis, rural health, organized by year.
  - A sample [cancer data report](#) contains a plethora of data, including the number of cancer cases in the state for a stipulated time period (5-6 years), number of mortalities, and then a breakdown of the occurrences of each type of cancer. We will pick the lung and bronchial cancer occurrences to specifically analyze from this. Picking up 2-3 such reports, we will get yearly data from around 1990-2021.
  - A sample [tuberculosis data report](#) contains information relating to the number of active TB cases in Nebraska by year, county wise spread of cases, mortality rates and gender/age/ethnicity wise splits of the cases for a cluster of years. Picking up 2-3 such reports, we will get yearly numbers from around 1990-2021.
2. [Census Reports : Nebraska Department of Revenue](#)
  - The [Certified Municipal Population](#) table shows certified populations for municipalities. These tables are presented on yearly tabs. We will simply pull the population counts for the city of North Platte and get the yearly census from 1990-2022.

### **Terms of Use:**

The [Nebraska Department of Health and Human Services Newsroom](#) says:

*'Per state statute, any requests for data not on the Department's public domain are to be submitted via public records request.'*

This makes it clear that the data publicly available can be used for academic research purposes.

## **Modeling Technique**

We will have yearly data for the number of cancer cases, mortality rate, number of TB cases, and population counts. In addition to this, we will append the smoke model estimate and determine if a correlation exists between smoke and respiratory diseases, by constructing charts/graphs of yearly trends. Further, we can create a prediction model to forecast the number of cancer cases expected for the next 5 years by potentially utilizing the smoke estimate as a predictor along with population and other historical data.

## **UNKNOWNNS AND DEPENDENCIES**

There are a lot of factors that could impede this analysis plan. I have already identified and am aware of a few of these, such as:

- The health data is all available via reports. These are PDF files with the data being contained in tables and graphs. This cannot be parsed by a system directly into a Python notebook. I will have to manually read each report and enter the data into an excel sheet, recreating the entire data set from scratch, just in the exact proper format I need. This could be a time consuming task as manual data entry requires careful attention and is an inhibitor to the project finishing on time.
- The health data, while public right now, could be made private at any time and taken off the website. If this happens, I would be unable to use the data and this would hurt the chances of completing the project.
- A bigger concern is that the health data is available largely at a state level granularity. There is no specific data for North Platte or Lincoln county. I am simply having to make the assumption that the distribution of the state mirrors those of the city and can assume it can be extrapolated across geographical regions. This is the only way to proceed because there is no local city level data available for public use. A request for this data was made by me, but I received no reply.
- Another crucial point to keep in mind is that at the end of the day, there simply may not be any trend or correlation between smoke estimate and respiratory illness occurrence. This makes sense because the smoke estimate calculation was based on wildfires from up to 1250 miles away, which would barely impact a small city such as North Platte. Also, as mentioned earlier, wildfires are a very small cause, and respiratory diseases can be caused by a plethora of reasons, which cannot be mirrored in this analysis.

## **COMPLETION TIMELINE**

### **Milestones and Significant Tasks**

1. Data Collection : Scraping the data out of the reports and manually entering it into an excel/csv. Collating multiple such reports and making a combined dataset for cancer, tuberculosis and population.
2. Model Choice and Development: Choosing what model to utilize to correlate smoke estimate with other predictor variables. Build out the model on the dataset.
3. Forecasting and Visualization: Test out the model and generate graphs showing trends in cancer cases/mortality/tuberculosis cases across years and compare with smoke estimates. Then use all population + smoke estimates together to forecast future trends of a particular predictor, for e.g. number of lung cancer cases.
4. Generate Insights: Semantic check if the analysis done makes sense and is logical. Develop reasoning in a format that local government authorities can easily understand and utilize for their reference later on.

### **Timeline**

<b>TASK</b>	<b>TIMESPAN</b>
Smoke Estimate Calculation	11/1 - 11/15
Data Collection/Augmentation	11/18 - 11/20
Model Selection and Development	11/21 - 11/22
Result Visualization & Forecasting	11/23 - 11/24
Insight Generation	11/25 - 11/26
Project Presentation (PechaKucha)	11/27 - 11/30
Project Report Preparation	12/1 - 12/8