Project Extension Plan
Ananya Bajaj

Motivation / Problem Statement:

This project involved an analysis of wildfires in a specified region and their repercussions on the surrounding areas, with a specific focus on fires within a 1250-mile radius of Pueblo West County, Colorado. In the initial phase, I conducted a comprehensive analysis of wildfire extent, accompanied by the computation of a smoke estimate designed quantitatively. My subsequent plan is to broaden the scope by incorporating socio-economic factors, providing a human-centric perspective that could directly affect Pueblo West County.

Understanding the spread and occurrence of wildfire smoke is crucial for grasping the dynamics of these natural disasters, often exacerbated by climate change. The smoke emitted contributes to air pollution, and my examination of socio-economic impacts offers insights into the broader implications of climate change on local environments and communities. My study of wildfire smoke spread enhances comprehension of its health ramifications, especially respiratory diseases issues such as Asthma and potential long-term effects on the affected population.

My analysis aids in constructing resilient communities by providing data-driven insights into socio-economic vulnerabilities arising from exposure to wildfire smoke. It serves as a tool for increasing public awareness about potential socio-economic impacts, fostering a sense of responsibility among residents and promoting proactive measures to reduce wildfire risk and impact.

Impact focus:

I have chosen to analyze healthcare data, specifically focusing on Asthma and heart attack data in Pueblo West County, CO, as it provides a human-centered approach to quantify the impacts of smoke. Asthma and heart attacks are conditions directly linked to air quality, and their prevalence can serve as tangible indicators of the consequences of smoke exposure on the community's health. By examining the occurrences of these respiratory and cardiovascular issues, I aim to establish a clear connection between the

incidents and periods of increased smoke, providing a quantitative measure of the health impacts on residents. This analysis not only enhances our understanding of the immediate effects of smoke but also contributes valuable insights for developing targeted healthcare interventions and public health strategies to mitigate the impact of environmental factors on community well-being.

Questions I aim to answer:

1. What percentage of adults in Pueblo West have been diagnosed with Asthma by a healthcare professional, and how many still have the condition? I plan to correlate this data with the previously obtained smoke estimates, with a particular emphasis on the years 2014-2017. To provide a comprehensive understanding, I intend to construct a correlation matrix, visually representing the relationships between these variables. Through this analysis, I aim to uncover potential trends and connections between Asthma prevalence and smoke exposure, offering valuable insights into the health dynamics of the community during the specified period.

2. What percentage of adults in Pueblo West are diagnosed with Angina or Coronary Heart Disease by healthcare professionals, determining those who still have the condition? By correlating this heart disease data with previously derived smoke estimates from 2014-2017, I aim to create a correlation matrix to visually represent the relationships between these variables? Through this analysis, I will check if this unveils patterns and associations between heart disease prevalence and smoke exposure, providing valuable insights into the health landscape of the community during this specific timeframe.

Prediction using time series model: In the previous part of the project I employed exponential smoothing to predict future smoke estimates, This time I aim to identify patterns and correlations between smoke exposure and health outcomes. The model can then be trained on this historical data to make predictions for future scenarios. For example, we could forecast the potential impact of varying levels of smoke exposure on asthma or heart disease prevalence in Pueblo West. This predictive capability can assist healthcare professionals, city planners, and policymakers in preparing for and mitigating potential health risks associated with future smoke events.

Datasets and Modelling:

Link to the dataset:

https://www.cohealthmaps.dphe.state.co.us/cdphe_community_level_estimates/

The Colorado Department of Public Health and Environment has CDPHE Community Level Estimates from 2014-2017.

Brief Information about the Data Source: Community Level Estimates, as defined by CDPHE, refer to health condition and risk behavior estimates generated for smaller geographical areas, specifically census tracts. These estimates are derived from statistical models, addressing the limitations of traditional public health surveillance systems that often cannot provide data for census tracts due to sample size constraints. The models used incorporate individual responses from the Colorado Behavioral Risk Factor Surveillance System (BRFSS) survey, along with socio-demographic and contextual information from the U.S. Census (American Community Survey). The multilevel model considers both individual and sociodemographic characteristics, allowing the estimation of health conditions or risk behaviors at the census tract level.

Census Data:
https://www.census.gov/quickfacts/fact/table/pueblowestcdpcolorado/POP060210

Terms of Use:

Unknowns / Dependencies:

1. **Timeframe 2014-2017:** The analysis is limited to the years 2014-2017. Any developments, changes, or trends that occurred before or after this timeframe may not be considered in the assessment. Events or factors that occurred outside this period could have potential impacts on the correlation between respiratory diseases and smoke estimates.

2. **No correlation between respiratory disease and Smoke estimates:** The assumption or current understanding is that there is correlation between respiratory diseases and the smoke estimates. However, this statement also acknowledges the possibility of not having any correlations at all. The analysis is based on the available data and information up to the present, and it does not definitively rule out the existence of a correlation.

3. **Another crucial dependency involves the necessity for data filtration and cleaning.** The dataset encompasses all counties in Colorado, presenting a broader scope than the specific focus on Pueblo West. Therefore, to ensure the accuracy and relevance of the analysis, a meticulous data cleaning process is imperative. This involves filtering the dataset to isolate and retain only the pertinent information related to Pueblo West. The data cleaning procedure will likely include removing irrelevant entries, refining data formats, and addressing any inconsistencies or discrepancies. The success of the analysis hinges on the meticulous execution of this data cleaning process to tailor the dataset to the targeted geographical area, Pueblo West.'

4. **An additional dependency is the reliance on open-sourced data**. While the current dataset is openly accessible, there exists a potential vulnerability if the data source decides to restrict or close access in the future. This could be due to policy changes, privacy concerns, or other factors beyond the control of the analysis. Consequently, it underscores the importance of regularly monitoring the data source's policies, terms of use, and any notifications regarding potential changes. Should the data source limit accessibility, alternative data repositories or updated agreements would need to be explored to maintain the continuity and reliability of the analysis.

Completion timeline

Data Collection: Filtering and cleaning the dataset to extract Pueblo West County info.

Model Selection and Development: Choosing an appropriate model to establish correlations between smoke estimates and other predictor variables. Constructing and refining the model based on the dataset.

Forecasting and Visualization: Testing the model's effectiveness and creating visualizations to illustrate trends in asthma cases over the years. Comparing these trends with smoke estimates. Utilizing the entire population dataset along with smoke estimates to predict future trends, such as the number of lung cancer cases.

Generate Insights: Conducting a semantic check to ensure the analysis is coherent and logical. Developing insights in a format that is easily comprehensible for local government authorities, providing valuable information for their future reference and decision-making.

Timeline

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