Part 2 - An Extension Plan

The project focuses on the impact of wildfires on the socio-economic impact of major cities in the US. The smoke caused by these wildfires has affected major fields like agriculture, health, tourism, property, and other societal aspects. The project's goal is to inform policy-makers, city managers, city councils, or other civic institutions, to make an informed plan for how they could or whether they should make plans to mitigate future impacts from wildfires. Split into 4 different stages we are trying to propose to the authorities that bring out the human-centered aspect of the otherwise technical project. The project is a perfect example of the application of Data Science for social change.

Progress so far

The first task (Part 1 - Common Analysis) set the stage for the project. It involved creating a composite numeric estimate by using factors related to wildfires. We were each assigned a city in the US, and tasked with extracting data about all the wildfires in and around the city. We will evaluate these wildfires for their impact on the socio-economic ramifications. Wildfire data had a trove of information documented about each of the fire. To name a few, the area it burned, the shape it took, the distance from the city, etc. The variety of information captured provided a holistic insight into the phenomenon.

Common analysis can be briefly summarized under the following main tasks:

1. Extract Wildfire data - Data is read from the GeoJSON and pre-processed to obtain Wildfire attributes like shape, area of the area burnt, days the fire lasted, distance of the fire from the city, etc. Data is filtered for the city assigned - Bismarck, North Dakota.
2. Filtering for relevant Wildfires - We calculated the closest distance of each wildfire to the city, forming a new attribute. Filtering fires within 1250 miles of the city, we focused on these proximity-related incidents. Additionally, we analyzed fires active since 1963, narrowing our examination to long-standing fire occurrences.
3. Estimate the Smoke Impact - To estimate the smoke impact from wildfires, I've crafted an approach considering crucial factors like fire size, intensity, proximity to the city, and other relevant attributes. After meticulously selecting these variables, a formulated function combines them, emphasizing larger, closer, and more intense fires to predict smoke impact. This estimation spans the fire season, accommodating the lack of specific fire duration data. Applying this formula to fires within 1250 miles of the city, I'll aggregate annual estimates, validating them against available Air Quality Index (AQI) data to ascertain their reliability.
4. Compare the smoke estimate to AQI - In my evaluation process, I needed to compare my smoke estimate with available Air Quality Index (AQI) data from the US EPA. These constraints necessitated comparing my estimate and the EPA's approximated data, considering these nuances in station coverage and estimation methods.
5. Developed a predictive model - Lastly, I developed a predictive model based on the fire data and smoke estimate for my assigned city. The model predicted smoke estimates for every year for the next 25 years (i.e., 2024-2049).

Extension Plan

Building upon the groundwork laid in our initial analysis, this project delves deeper into understanding the far-reaching socio-economic consequences of wildfires, especially in major US cities. With a multidimensional perspective encompassing fields like agriculture, health, tourism, and property, we aim to equip decision-makers with informed insights. By presenting a human-centered narrative within this technically driven project, our proposed extension plan is designed to spotlight the tangible impact of wildfires on society. Through the amalgamation of data science and societal concerns, this endeavor stands as a testament to the practical application of data analytics for societal betterment.

My intention for this extension plan is to hone in on the economic repercussions, specifically within the realm of agriculture, stemming from the aftermath of wildfires. Recognizing the pivotal role of agriculture in local economies, I aim to study the nuanced effects that wildfires exert on crop production and overall agricultural sustainability. By scrutinizing these economic dimensions, I aspire to shed light on the impact of wildfires and the economic stability of our communities. The agricultural realm makes a huge impact within the state of North Dakota and my assigned city which is the capital.

Motivation

Understanding the economic impact of wildfires on agriculture is paramount owing to its multifaceted implications for both the agricultural sector and the broader community. This analysis is not merely helpful in quantifying the relationship between wildfires and agriculture but also is inclined to provide and assess the damage to the economic stability of communities reliant on agriculture. Wildfires, with their influence, pose a substantial threat to agricultural productivity, soil health, and the livelihoods of farmers, engendering a ripple effect on food security, local economies, and the overall socio-economic fabric. By diving deep to perceive the effect, the analysis seeks to address a pressing concern: how these natural disasters reverberate through the agricultural landscape and the subsequent ramifications for communities reliant on agriculture.

Delving into the economic aftermath intends to unravel not just the direct losses in crop yield, livestock, and agricultural infrastructure but also the cascading impacts on market stability, employment in the agricultural sector, and the region's overall economic well-being. By quantifying and delineating these impacts, the analysis endeavors to furnish decision-makers, policymakers, and agricultural stakeholders with actionable insights to fortify resilience, develop effective mitigation strategies, and aid in post-wildfire recovery efforts.

Moreover, the significance of this analysis extends beyond the economic domain, transcending into the realms of societal welfare and community sustainability. Agricultural communities, often the bedrock of many regions, rely not only on economic stability but also on the cultural and social fabric fostered by agrarian practices.

Hence the problem statement I am inclined to make some progress on using the wildfire data is "*Investigation of the direct and indirect economic effects of wildfires on agriculture. I aim to discern their*

impact on crop yields, soil health, and the overall economic resilience of agrarian communities, crucial for formulating targeted mitigation strategies."

Impact Focus

My impact focus from the options presented to us is going to be economic-centric. I intend to use this particular focus owing to its ability to show us a very human ingestible impact of wildfires. While other focus areas do have a direct relationship with the factors I am investigating, the economic and monetary ramifications leave a lasting impact as they need very little understanding. The way I am going to represent the economic impacts of wildfires is through the agricultural aspects. One of the primary revenue streams of the state of North Dakota happens to be agriculture. Hence, drawing parallels between agriculture and the subsequent revenue it rakes in is within the reach of the scope and the objective of this project. My main focus will revolve around the yield of the crops and the factors responsible for the yield of the mentioned crops interwoven with the smoke impact that I formulated in the first part of the project.

The yield seems like an appropriate socio-economic indicator for the effect of wildfire on the city of Bismarck since it is directly related to the revenue of the state. The focus area is simple to grasp and quantify for people with minimal understanding of the nuances of the analysis and the underlying technicalities. Since this analysis acts as a bridge between the authorities and the citizens of the city, I used this factor to help take a strong stance and quantify the damage done. The analysis will also yield tangible and intangible solutions by tagging a dollar impact on every solution and policy it triggers.

Data or Model Used

In my quest to find data that supported my overarching idea to quantify the relationship between wildfire and smoke estimate, I stumbled upon the data published by the United States Department of Agriculture - National Agricultural Statistics Service. They have hosted a range of datasets that span from crops to animals and environmental data. Since my focus area is agriculture, I will be using crop data. License and terms of use are linked on this page.

The dataset encapsulates a comprehensive array of agricultural data, primarily focusing on crop-related statistics and practices within specific geographical regions and temporal scopes. It presents a rich reservoir of information encompassing various agricultural commodities such as sunflower, barley, and wheat, each delineated by distinct production and utilization practices. The recorded statistics range from crucial metrics like harvested acreage to economic indicators such as prices received, providing a multifaceted view of the agricultural landscape. Geographically, the data spans diverse locations, including counties and states within the United States, shedding light on localized agricultural trends and variations. With details cataloged across different years and frequency intervals, this dataset serves as a valuable resource for understanding agricultural dynamics, production patterns, and economic aspects within specific regions and crop categories.

They encompass a spectrum of essential information, including:

- Production Practices: Detailed descriptions of the methods and practices employed in crop production, covering aspects such as planting techniques, irrigation, fertilization, and pest control strategies.
- Utilization Practices: Information regarding how the harvested crops are utilized, whether for specific industries, processing methods, or end-user consumption.
- Harvested Area: Data on the acreage or area of land utilized for cultivating specific crops, providing an understanding of cultivation scales and trends over time.
- Yield and Production Volume: Metrics related to the number of crops yielded per unit area, offering insights into productivity trends and agricultural output.
- Price Received: Economic indicators denote the prices farmers receive for their produce, indicating market conditions, economic viability, and profitability within the agricultural sector.
- Geographical Specifics: Details encompassing the geographical domain, including counties, states, regions, and watershed areas, delineating localized agricultural trends and variations.
- Temporal Context: Records spanning multiple years and frequency intervals, enabling the observation of crop-related patterns, changes, and fluctuations over time.

These statistics collectively form a comprehensive portrait of crop cultivation, economic aspects, and agricultural dynamics, empowering stakeholders, policymakers, and researchers with detailed insights into the multifaceted dimensions of agricultural practices and their economic implications within specified regions and timeframes.

In the initial phase of the project, the SARIMA model served as the predictive tool for estimating smoke impact over forthcoming years. Building upon this foundation, the subsequent phase entails an enhancement of the predictive model by incorporating additional features. Specifically, the intention is to augment the existing SARIMA model by integrating other pertinent variables to bolster prediction accuracy. One crucial augmentation involves conducting rigorous statistical tests to ascertain a statistically significant relationship between the estimated smoke impact and crop yield. This strategic integration aims to elucidate and validate potential correlations or dependencies between smoke estimates derived from the SARIMA model and the resultant effects on agricultural yield. Ultimately, this iterative refinement seeks to fortify the predictive capacity of the model and provide deeper insights into the interplay between smoke impact and agricultural productivity.

Unknowns and dependencies

When we try to predict how much crops will grow, there are some things we often forget to think about. Little changes in weather in specific areas, like small changes in temperature or how much rain falls, can really affect how well crops grow. The soil where plants grow is also super important. Different soils with different nutrients and things like that can change how much crops grow a lot. Some bugs and diseases can harm plants and lower the amount of crops we get. These things can be a bit hard to predict.

Besides natural factors, things like money, new technology, and what people want to eat can also change how many crops we get. If prices for crops change or there are new machines that help farmers, they might change how much they grow. Even what people like to eat or new diets can change what farmers decide to grow, and that affects the number of crops we get. Considering all these things when guessing how much crops will grow helps us make better predictions and plans for farming. One additional information that came across while researching the chief source of income of North Dakota, I came across the prevalent oil and natural gas industry. These industries have a profound impact on the agriculture industry. These were some important factors or dependencies that could be potentially useful in yield prediction.

Timeline to completion

The tentative timeline I am hoping to follow to implement the project is listed below:

Data Collection (2 days - 11/17 to 11/19): Gather additional data relevant to crop yields, including historical yield data, soil health metrics, microclimatic variations, and pest dynamics.

Data Preprocessing (1 day - 11/19 to 11/20): Cleanse and prepare the collected data for analysis, ensuring consistency and compatibility across datasets.

Feature Engineering (2 days - 11/20 to 11/22): Identify and incorporate relevant features for the enhanced predictive model, considering factors like soil health, microclimatic variations, and pest dynamics into the existing SARIMA model.

Model Enhancement (2 day - 11/22 to 11/24): Augment the SARIMA model by integrating additional features, perform statistical tests to establish correlations between smoke estimates and crop yield, and refine the model for improved accuracy.

Model Testing and Analysis (2 days- 11/24 to 11/26): Validate the enhanced model, conduct rigorous testing, and analyze the predictive power of the augmented features in predicting crop yield.

Visualization of Results (1 day - 11/27 to 11/28): Create visual representations (charts, graphs) to illustrate model outcomes and relationships between smoke estimates and crop yield.

Documentation and Reporting (3 days - 12/4-12/7): Prepare comprehensive documentation detailing the methodology, findings, and insights derived from the analysis.

Preparation for Presentation (1 week - 11/23 to 11/30): Craft a presentation summarizing the project's key aspects, findings, and conclusions.

In conclusion, this plan charts a detailed course to enhance the model, aiming to uncover the relationship between smoke impact and crop growth. By refining data analysis and incorporating new factors, this approach seeks to gain deeper insights into how smoke influences agricultural yields.