Capstone 2 Proposal

Project Name
Wildfires in The U.S. - Understanding the Annual Correlation Between the Number of Fires, Acres Impacted and the Associated Suppression Costs

Problem Statement
Wildfires are a significant environmental issue that also endanger human lives. While they have become less prominent over the past two decades, they have become more devastating with acres affected and suppression costs in mind. Almost 148 million acres have been burnt over the past two decades compared to around 48 million acres the prior 15 years. Further, suppression costs have exceeded $34 billion over the past two decades compared to over $6 Billion over the prior 15 years. The objective of this project is to predict the future state of U.S. wildfires from an acres affected and suppression costs focus.

Clients
Federal and state agencies that focus on land conservation and are responsible either directly or indirectly for overseeing, supporting, or extinguishing wildfires (fire departments included). Some states will stand to benefit from this information more (think California) than others (think Minnesota). Nonetheless, knowledge is key so it is important to leverage historical data to better understand the future wildfire outlook for the United States. These agencies can leverage the insight derived from this project to either implement superior regulations to prevent future wildfires or to gain a superior understanding on financial resources needed to combat future wildfires.

Data Sources
I obtained my Wildfire focused dataset from Kaggle which links the dataset to The National Interagency Fire Center (NIFC). The NIFC works alongside other U.S. federal agencies in developing environmental focused regulations and by providing support and information to wildland fire personal and other emergency services employees, among other focuses. The dataset covers wildfires from 1985 to 2020.
https://www.kaggle.com/datasets/kkhandekar/total-wildfires-acres-affected-1983-2020?resource=download

Methodology
My approach will be to conduct regression analysis via supervised learning that will focus on the number of fires, acres affected, and total suppression costs each year.

Deliverables
The deliverables will include Jupyter notebooks containing code leveraging methods used (data wrangling, exploratory data analysis, machine learning) to conduct desired analysis as well a final report containing a PowerPoint presentation.