Capstone Project #1 –
Identifying counties that are the most fire-prone and predicting the cause of a fire wildfire.

Problem:

Forest fires are a major environmental issue, creating economic and ecological damage while endangering human lives. Fast detection is a key element for controlling such a phenomenon. Despite an increase in state expenses to control this disaster, each year millions of forest hectares are destroyed all around the world. In California alone, we witnessed some of the deadliest and most destructive fires in state history. 98 civilians and 6 firefighters lost their lives. In total, a staggering 1.9 million acres were burnt, with the loss of over 10,300 structures, and incurred more than $3.5 billion in damages. A further $1.8 billion was spent by the various agencies to control and extinguish the fires. Wildfires are costly events, in so many ways. The objective of this potential project would be to identify What counties are the most fire-prone and to predict the cause of a fire wildfire.

Clients:

The clients would be agencies responsible for controlling and extinguish the fires including Fire department. This model enables the appropriate organizations to take preventative action, such as cutting firebreaks, as well as informing planning and preparedness activities, such as where to store fire retardant. It can be used in ensuring that the front-line firefights are deployed to the right locations to have the maximum impact, while simultaneously minimizing the risk to their safety.

Data: 1.88 Million US Wildfires (Kaggle1)

This data publication contains a spatial database of wildfires that occurred in the United States from 1992 to 2015. It is the third update of a publication originally generated to support the national Fire Program Analysis (FPA) system. The wildfire records were acquired from the reporting systems of federal, state, and local fire organizations. The following core data elements were required for records to be included in this data publication: discovery date, final fire size, and a point location at least as precise as Public Land Survey System (PLSS) section (1-square mile grid). The data were transformed to conform, when possible, to the data standards of the National Wildfire Coordinating Group (NWCG). Basic error-checking was performed, and redundant records were identified and removed, to the degree possible. The resulting product, referred to as the Fire Program Analysis fire-occurrence database (FPA FOD), includes 1.88 million geo-referenced wildfire records, representing a total of 140 million acres burned during the 24 years.

Approach:

The approach would be to Conduct regression analysis considering factors such as COUNTY, Geographic Area, size, location and date

Deliverables:

The project deliverables will include Jupyter notebooks containing methods used and code to support the analysis. In addition, a final report will be delivered.

1.   https://www.kaggle.com/rtatman/188-million-us-wildfires