July 26, 2022

Wildfire Risk Due to Climate Change

CS-442-BSCS

USAMA-RAHEEM-002

19-A-19011519-002

University of Gujrat- Hafiz Hayat Campus-Gujrat

Project-Idea-Name-Assignment-3

Data Science
BS Computer Science, University of Gujrat.
Due Date: 27 July 2022

Name: Usama Raheem
Roll No: 19011519-002

Topic Title: *DS Project*

1. Project Name

"Wildfire Risk Due to Climate Change"

2. Description

Wildfires are some of the most dangerous effects humanity is facing right now due to climate change. Every year, there are hundreds and thousands of wildfires reported across the world, causing huge economical losses. Resulting in an overall economic imbalance in world markets and trading.

3. Objectives

- My project aim is to provide insights on wildfires, their causes, and their behaviors.
- Aim is to provide useful recommendations to increase public safety and minimize the economic damage from potential future wildfires
- I will analyze the geographical information on where and when the wildfires occurred which could provide useful information to identify high-risk areas that are vulnerable to wildfires.
- My project findings will help make better strategies on where firefighters' departments are located and how to respond on time.
- I will find/predict the causes and risks of wildfires that can help plan preventive measures in future
- Increasing public understanding of wildfires can help prevent future losses and save billions of dollars

4. Problem

- Wildfires are costly, and firefighting is also costly. There are approx.. 75000+ wildfires per annum in the US on average
- 400 billion of total economic loss in 2018 from wildfires in California.
- 16.5 billion is lost from the costliest wildfires in 2018.

- Most Wildfires in the US are caused by humans.
- Most Destructive Wildfires are caused by Nature.

5. Solution

- Interactive maps to visualize the historical fire data.
- There should be Statistical Analysis to reveal frequencies, trends, and correlations
- I will apply some Machine learning models to predict the causes and risks for wildfires.

6. Dataset

The data set is from USDA forest services 1992-2015

7. Techniques

- Data Collection and Data Cleaning: by using NOAA API, wildfire dataset from US forecast service.
- EDA and Data visualization: using SciPy, SKlearn, Matplotlib, Plotly, folium, Seaborn
- Machine Learning Models: Random Forest, creating Gaussian Classifier.

8. Is your Problem/Project Multi Class or Binary Class

- My problem which is to analyze and predict the cause factor of occurring major wildfires, after the observation, my problem is Binary class, because I classify my problem into 0 or 1 as a resultant factor so far so good

```
In [74]: ca_wf_missing['natural_cause_predict'].value_counts(normalize = True)

Out[74]: 0    0.968404
         1    0.031596
         Name: natural_cause_predict, dtype: float64
```

Based on the above result, we can see among all the wildfires that have missing or unidentified reasons, 3% maybe caused by natural reasons, and 97% were probably caused by human-related reasons.

9. What Features are involved in your Dataset

How many times fire errupts due to which cause?

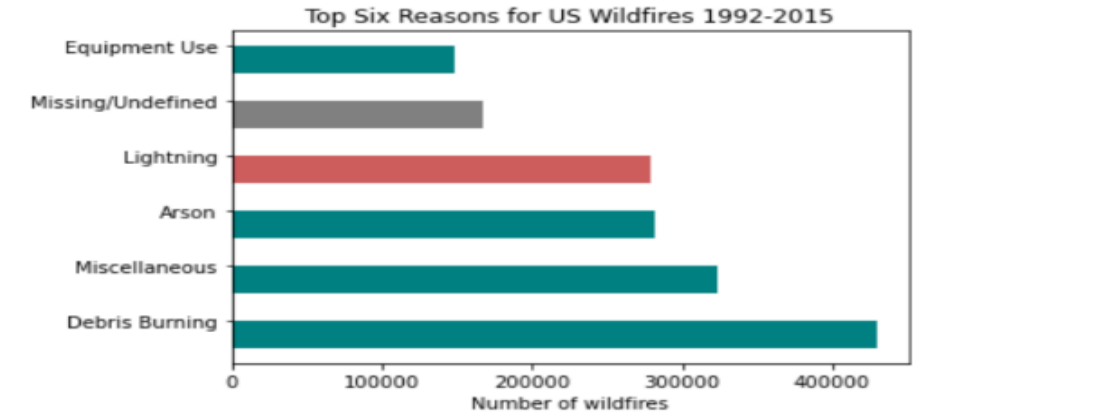
```
In [13]: wf_cause_num = wildfire.stat_cause_descr.value_counts()
         wf_cause_num

Out[13]: Debris Burning       429028
         Miscellaneous        323805
         Arson                281455
         Lightning            278468
         Missing/Undefined    166723
         Equipment Use        147612
         Campfire              76139
         Children              61167
         Smoking               52869
         Railroad              33455
         Powerline             14448
         Fireworks             11500
         Structure              3796
         Name: stat_cause_descr, dtype: int64
```

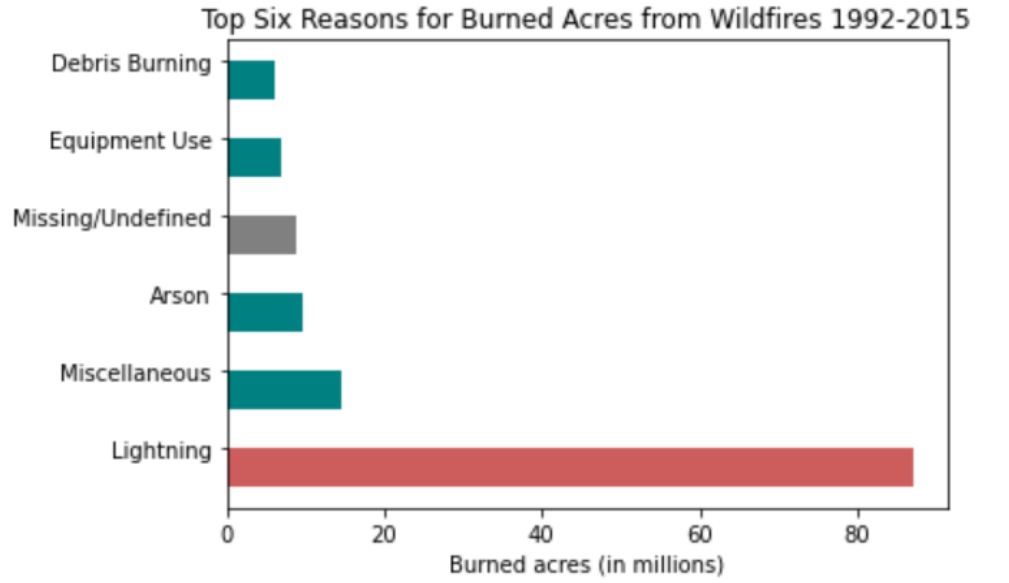
What are the most important features in my DataSet

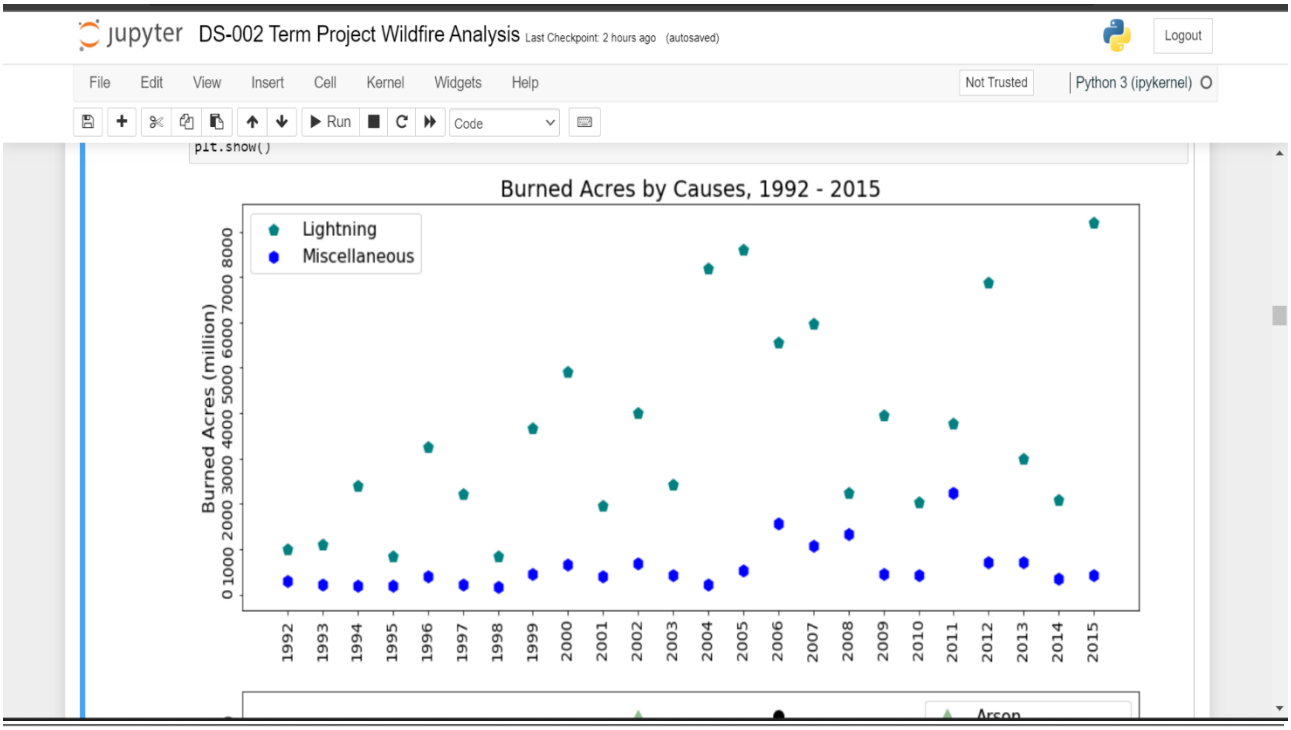
In [72]:
```
# Get numerical feature importances
importances = list(clf.feature_importances_)
# List of tuples with variable and importance
feature_importances = [(feature, round(importance, 2)) for feature, importance in zip(ca_wf_md1_list, importances)]
# Sort the feature importances by most important first
feature_importances = sorted(feature_importances, key = lambda x: x[1], reverse = True)
feature_importances
```

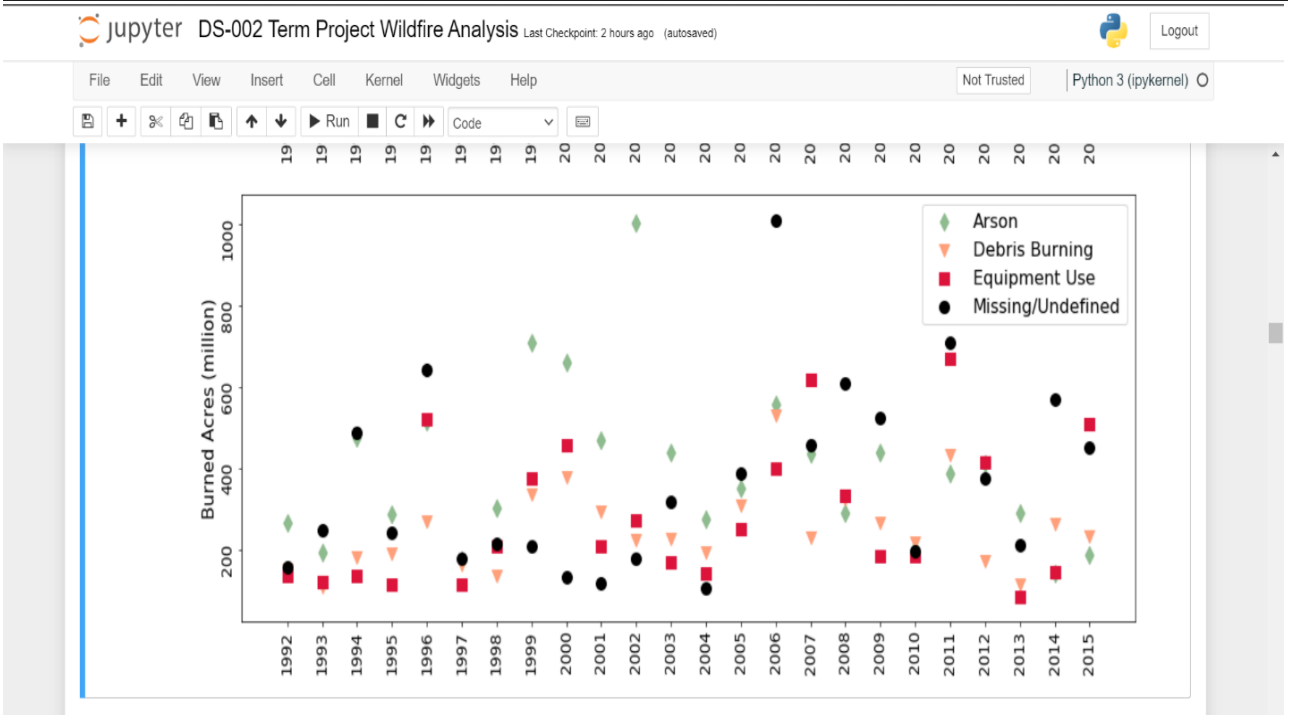
Out[72]:
```
[('discovery_doy', 0.26),
 ('latitude', 0.25),
 ('longitude', 0.24),
 ('burning_days', 0.15),
 ('TAVG', 0.1)]
```

From above, we can see that the most important variables for predicting the causes of wildfire is longitude, discovery date, and latitude.






jupyter DS-002 Term Project Wildfire Analysis Last Checkpoint: 2 hours ago (autosaved)
Logout
File Edit View Insert Cell Kernel Widgets Help
Not Trusted
Python 3 (ipykernel)
Run
Code
plt.show()
Burned Acres by Causes, 1992 - 2015
Lightning
Miscellaneous
Burned Acres (million)
0 1000 2000 3000 4000 5000 6000 7000 8000
1992 1993 1994 1995 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010 2011 2012 2013 2014 2015


jupyter DS-002 Term Project Wildfire Analysis Last Checkpoint: 2 hours ago (autosaved)
Logout
File Edit View Insert Cell Kernel Widgets Help
Not Trusted
Python 3 (ipykernel)
Run
Code
Arson
Debris Burning
Equipment Use
Missing/Undefined
Burned Acres (million)
200 400 600 800 1000
1992 1993 1994 1995 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010 2011 2012 2013 2014 2015

jupyter DS-002 Term Project Wildfire Analysis Last Checkpoint: 2 hours ago (autosaved)

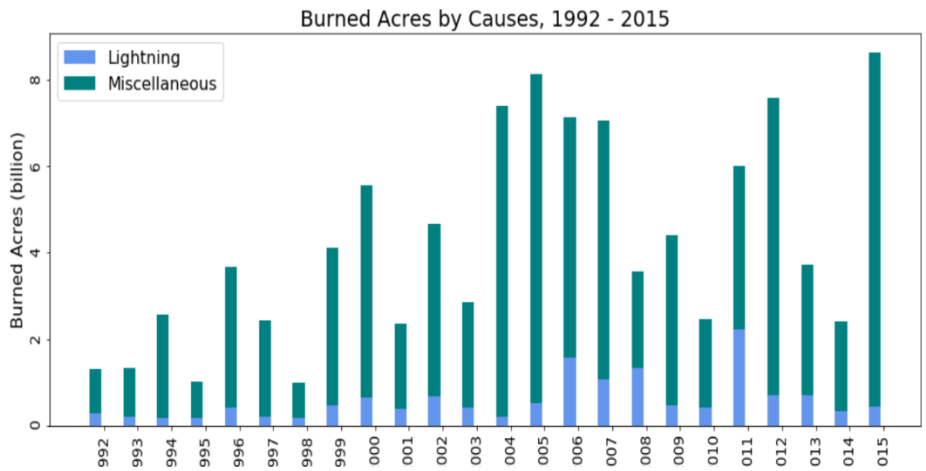
Logout

File Edit View Insert Cell Kernel Widgets Help

Not Trusted | Python 3 (ipykernel)

Run Code

```
ax2.set_xticks(tick_positions)
ax2.set_xticklabels(y)
ax2.tick_params(labelsize=14,rotation = 90)
plt.show()
```



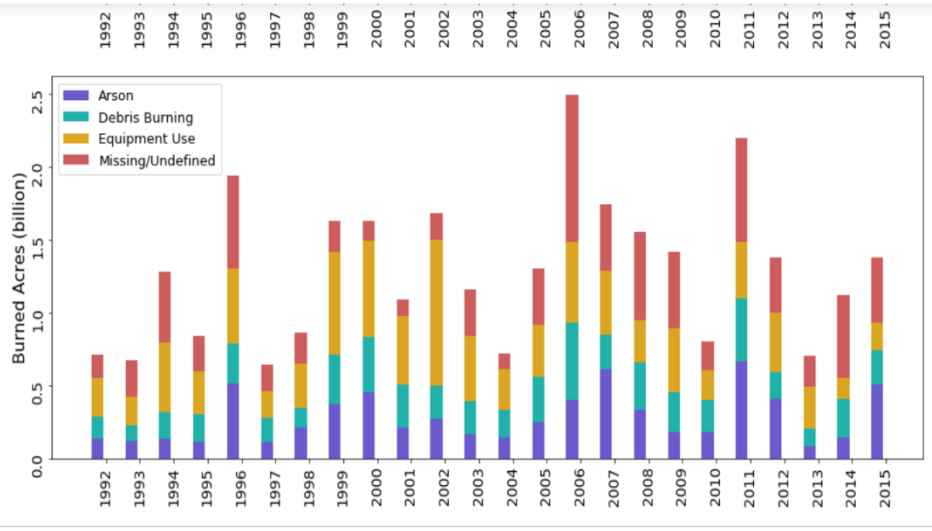
jupyter DS-002 Term Project Wildfire Analysis Last Checkpoint: 2 hours ago (autosaved)

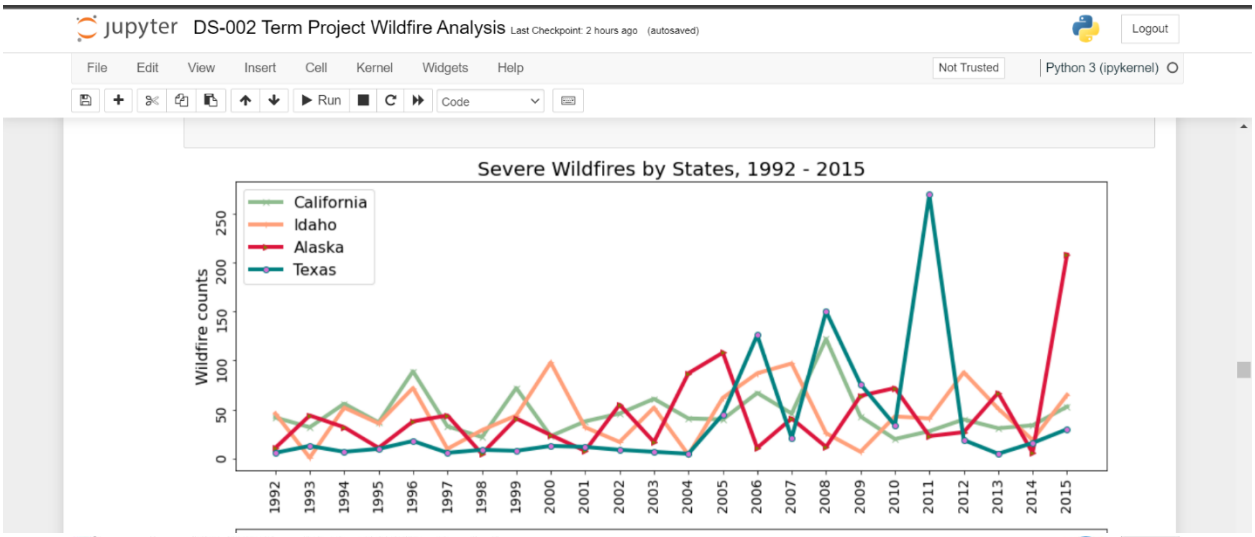
Logout

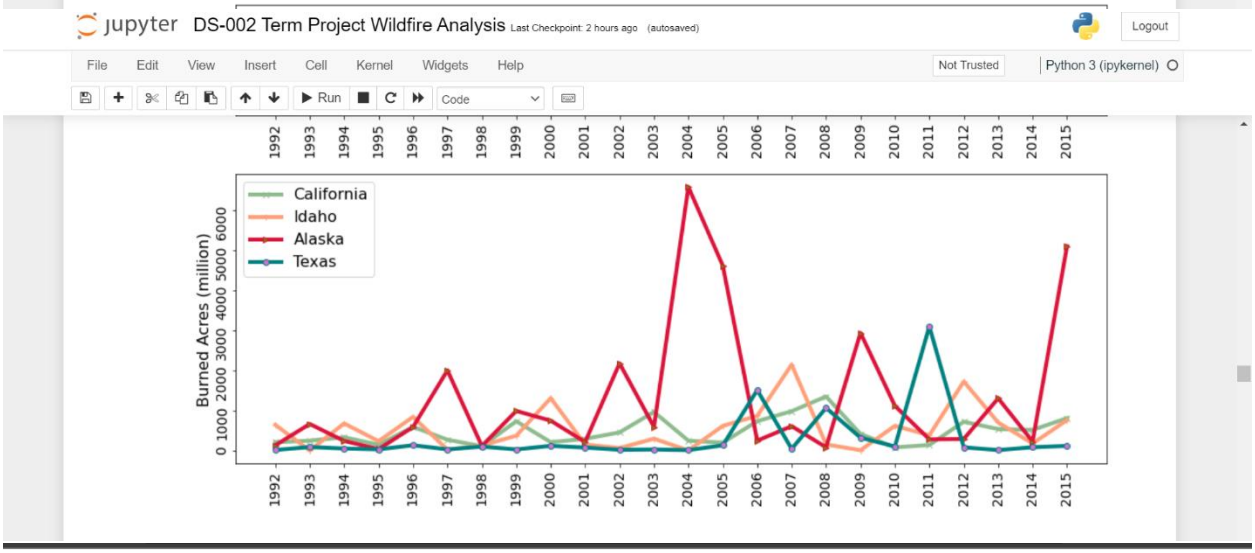
File Edit View Insert Cell Kernel Widgets Help

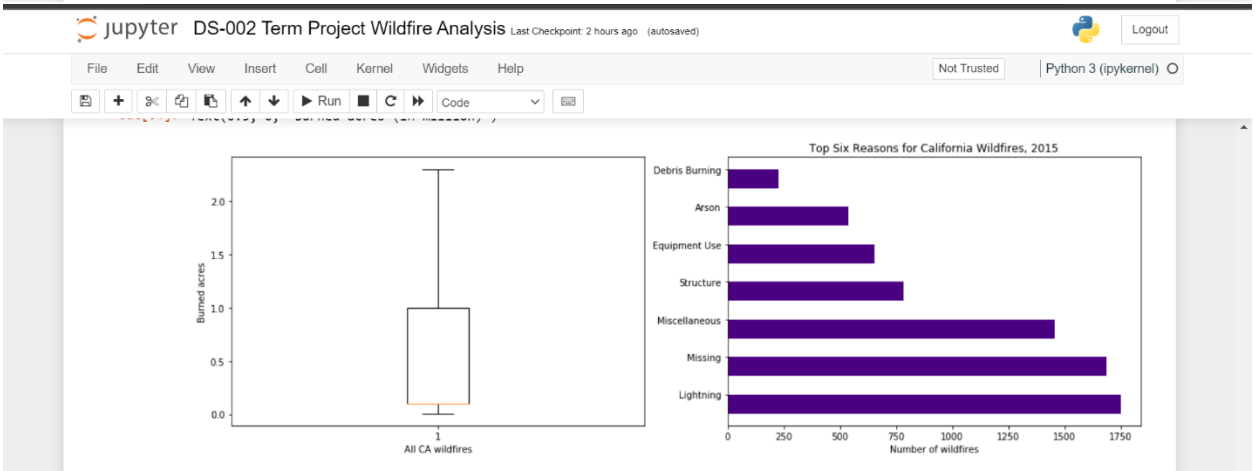
Not Trusted | Python 3 (ipykernel)

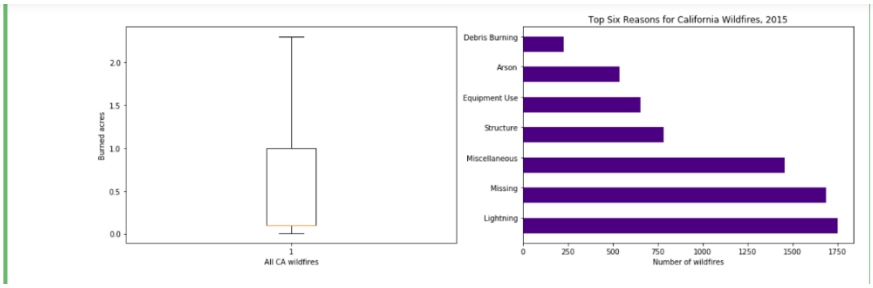
Run Code

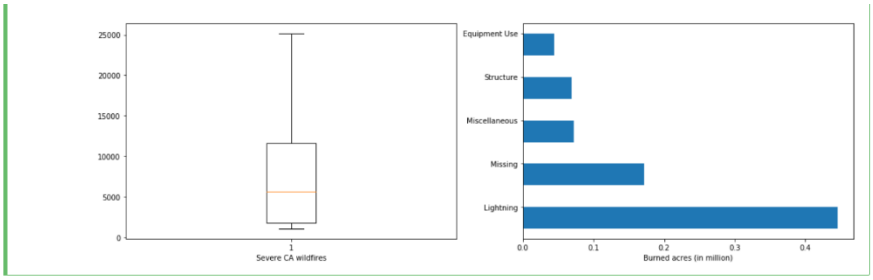


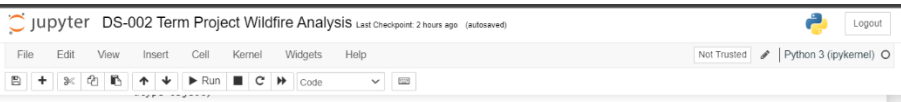

jupyter DS-002 Term Project Wildfire Analysis Last Checkpoint: 2 hours ago (autosaved)
Logout
File Edit View Insert Cell Kernel Widgets Help
Not Trusted
Python 3 (ipykernel)
Run
Code
Severe Wildfires by States, 1992 - 2015
California
Idaho
Alaska
Texas
Wildfire counts
0 50 100 150 200 250
1992 1993 1994 1995 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010 2011 2012 2013 2014 2015

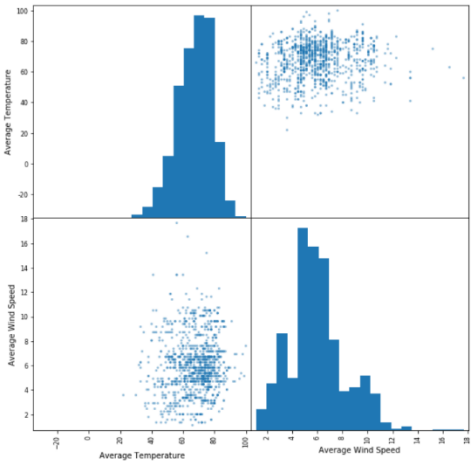

jupyter DS-002 Term Project Wildfire Analysis Last Checkpoint: 2 hours ago (autosaved)
Logout
File Edit View Insert Cell Kernel Widgets Help
Not Trusted
Python 3 (ipykernel)
Run
Code
1992 1993 1994 1995 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010 2011 2012 2013 2014 2015
California
Idaho
Alaska
Texas
Burned Acres (million)
0 1000 2000 3000 4000 5000 6000
1992 1993 1994 1995 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010 2011 2012 2013 2014 2015

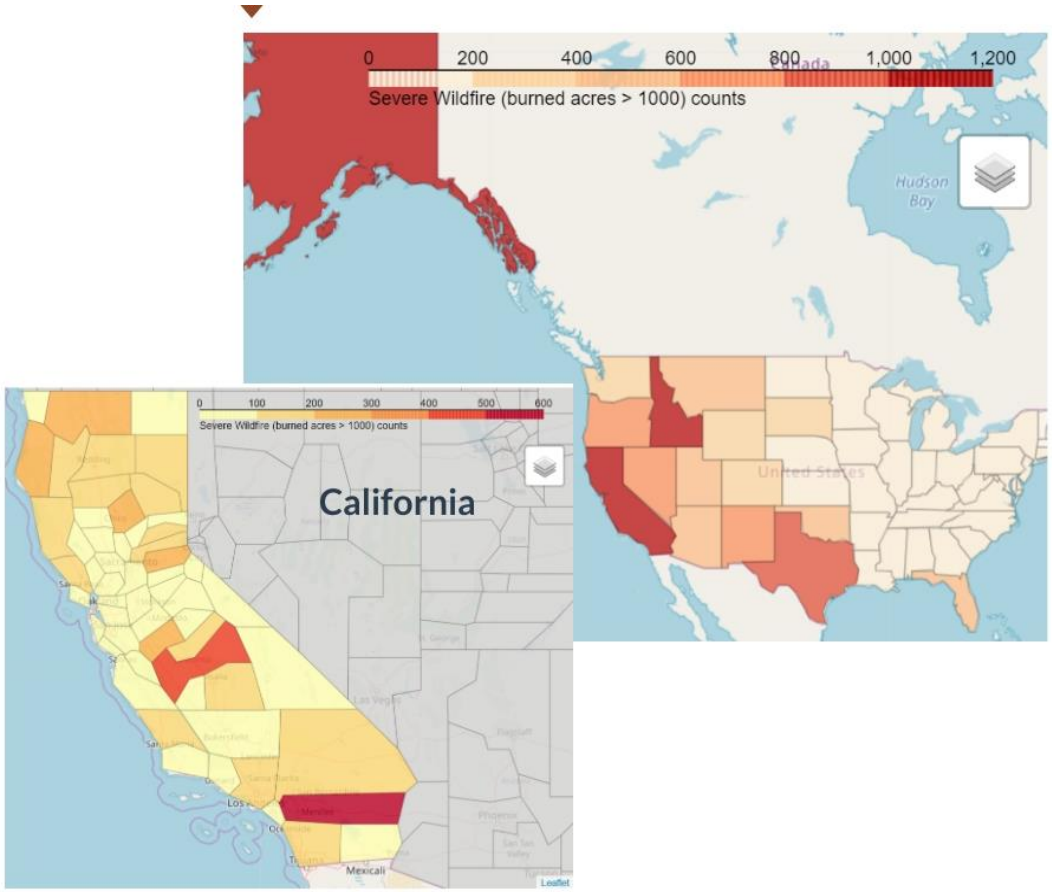

jupyter DS-002 Term Project Wildfire Analysis Last Checkpoint: 2 hours ago (autosaved)
Logout
File Edit View Insert Cell Kernel Widgets Help
Not Trusted
Python 3 (ipykernel)
Run
Code
Burned acres
0.0 0.5 1.0 1.5 2.0
1
All CA wildfires
Top Six Reasons for California Wildfires, 2015
Debris Burning
Arson
Equipment Use
Structure
Miscellaneous
Missing
Lightning
0 250 500 750 1000 1250 1500 1750
Number of wildfires


Burned acres
All CA wildfires
Top Six Reasons for California Wildfires, 2015
Debris Burning
Arson
Equipment Use
Structure
Miscellaneous
Missing
Lightning
Number of wildfires


Severe CA wildfires
Equipment Use
Structure
Miscellaneous
Missing
Lightning
Burned acres (in million)


jupyter
DS-002 Term Project Wildfire Analysis
Last Checkpoint: 2 hours ago (autosaved)
Logout
File Edit View Insert Cell Kernel Widgets Help
Not Trusted
Python 3 (ipykernel)
Run
Code


Average Temperature
Average Wind Speed
Average Temperature
Average Wind Speed



- **These highlighted areas that are shown Wildfire severity across Californian Counties using color.**