

## **What is the main cause for wildfires in the USA changed over years?**

First, start spark to import the downloaded data. The csv file from NIFC(Open Data National Interagency Fire Center).

For this data set, I just keep some meaningful attributes, such as code, identifier attributes to remove, because they are artificial increase, help record data and management, to analyze the cause of the forest fire does not help.

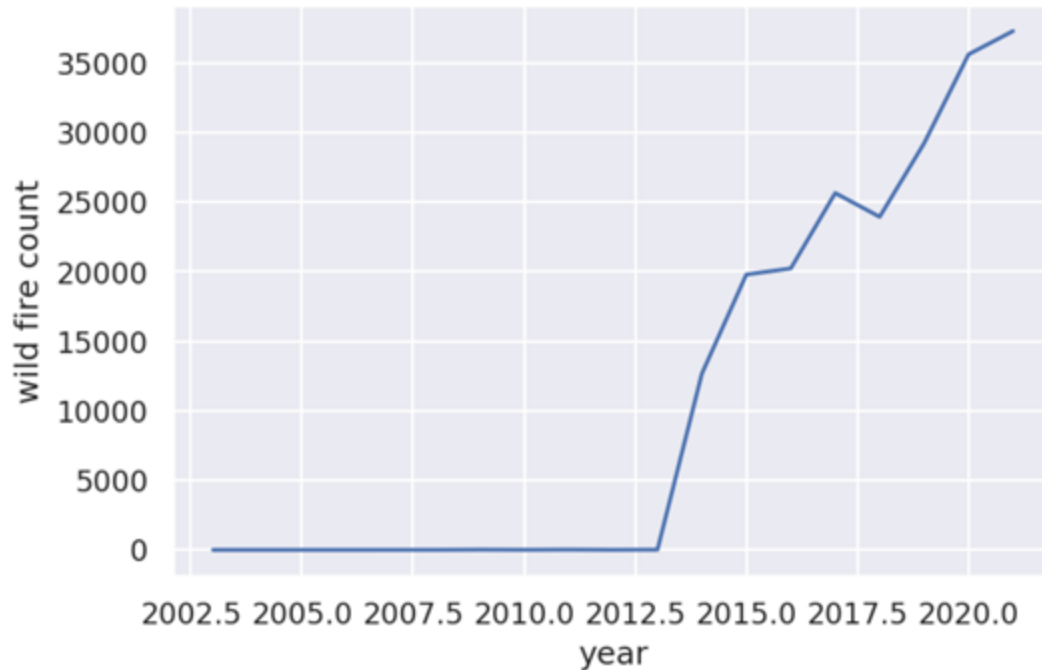
I just keep the following properties:

```
a = ['PrimaryFuelModel', 'Calculated Acres', 'Fire Discovery Date Time', 'Fire Out Date Time',  
'Discovery Acres', 'Fire Behavior General1', 'Fire Behavior General2', 'Fire Behavior General3',  
'Fire Behavior General', 'FireCause', 'FireCauseGeneral', 'FireCauseSpecific', 'Initial Latitude ',  
'Initial Longitude', 'Predominant Fuel Group', 'Total Incident Personnel', 'Incident Type  
Category', 'Estimated Cost To Date', 'POO State']
```

Then check the null value: There are too many null values! So, in the analysis of the need to pay extra attention to the impact of the null value, should increase the necessary judgment conditions, and to find a way to get useful data in the presence of null values.

Since wildfire was the only type to be analyzed, we will remove types that don't belong to “wild fire” type[1].

Now we can see the number of wildfires each year and add the year to attributes; At this point, we can draw the number of wildfire change trend every year:



It can be seen that the statistics before 2012 were insufficient and therefore very low in plot; Since 2013, the number of forest fires has increased rapidly, not only because of the increase in statistical capacity, there must be other reasons.

There are too many null values in “FireCause” , “FireCauseGeneral” and “FireCauseSpecific”, especially the latter 2. We can’t fill the null with appropriate value; If all the null values are removed, there is insufficient analysis data, so the FireCause attribute is analyzed. This attribute has a smaller portion of missing values. For null values in FireCause, classify them directly as Unknown. Next, look at the number of different fire causes:

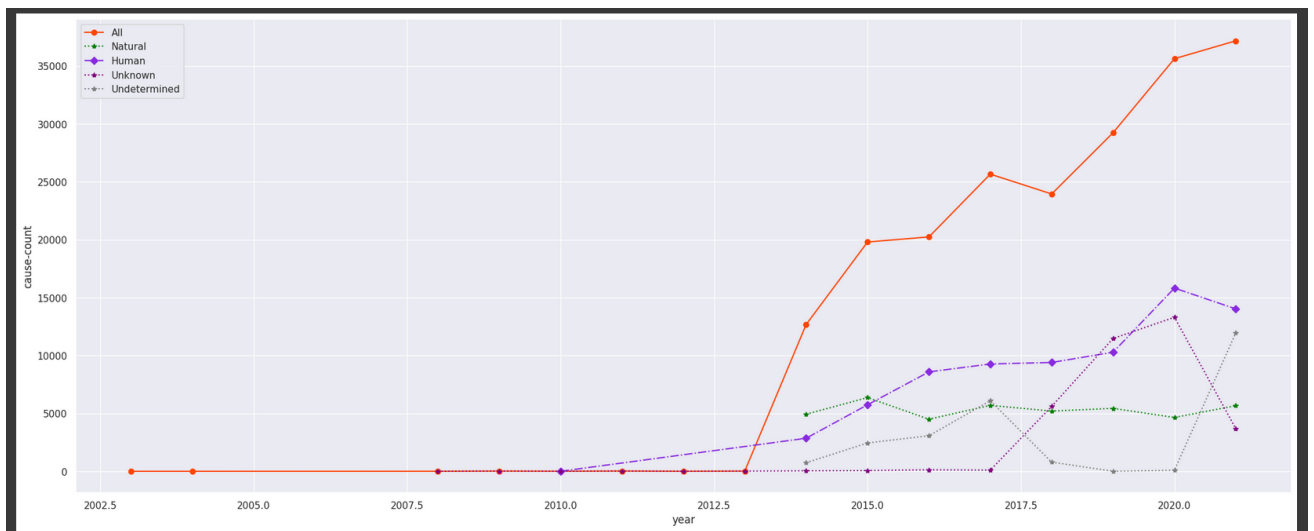
```

filtered_df.groupBy("FireCause").count().show()
✓ 1.7s
+-----+-----+
| FireCause|count|
+-----+-----+
| Unknown|60840|
|Undetermined|25199|
| Human|76055|
| Natural|42423|
+-----+-----+

```

Then we will try to find which kind of cause is primary. The figures below show

1. Change in the number of all wildfires with the year
2. Change of the number of wildfires with different causes with the year

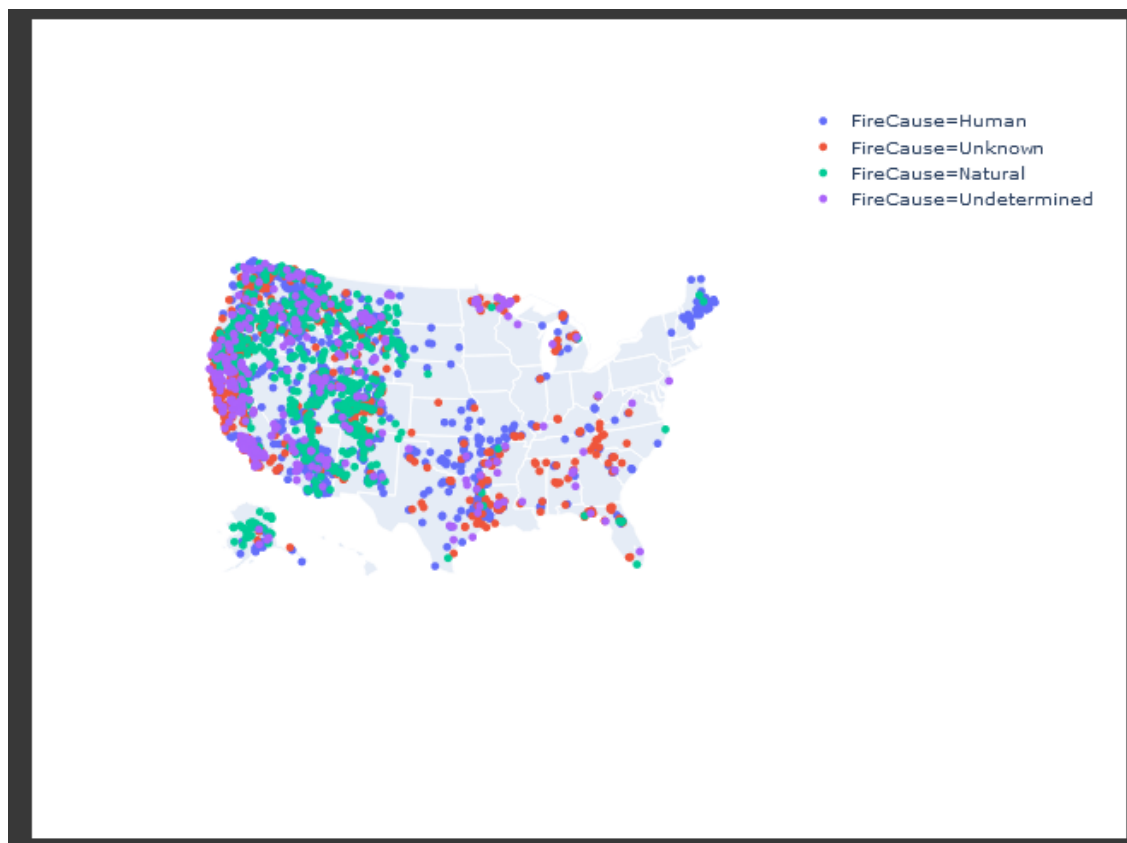


We can see from it that in the above picture, the purple broken line is an artificial category, and the wildfire caused by artificial factors increases obviously every year. The green line is a natural cause and it can be seen that there is not much change from year to year.

There are two undetermined fire causes, unknown and undetermined, which also increase over time as a whole, but we are not sure whether they are man-made or natural. The only certainty is that in the known data, the number of man-made fires increases over time and the number of natural fires remains essentially unchanged.

Of course, these unknown causes may also be caused by climate. According to the researchers' findings [2], climate change is also an important reason for the frequent occurrence of forest fires.

In addition, we can also analyze the distribution of wildfires. It can be seen that the current data set mainly records wildfire data from the United States. We can see that the natural and undetermined causes are mostly located on the West coast. In addition, the middle of the USA's WildFire is mostly caused by humans.



## REFERENCES

[1][https://data-nifc.opendata.arcgis.com/datasets/nifc::wfigs-wildland-fire-locations-full-history/  
about](https://data-nifc.opendata.arcgis.com/datasets/nifc::wfigs-wildland-fire-locations-full-history/about)

[2]<https://www.worldweatherattribution.org/western-north-american-extreme-heat-virtually-impossible-without-human-caused-climate-change/>