**Forest fires analysis on Portugal dataset**

Forest fires have been major environmental concerns. It adversely affects wildlife, nature and impacts people too. This paper is to analyze the burn area with weather conditions and location. The data used for this analysis is from Montesinho natural park in Portugal.

Data & its limits

The data for this project is from Montesinho natural park, from the Tr´as-os-Montes northeast region of Portugal (Ref: [LINK](https://archive.ics.uci.edu/ml/datasets/forest+fires)). Data is pulled from 2 separate databases, one from fire inspector data and 2nd every 30min metrological data. The time period is Jan 2000 to Dec 2003. This data set includes spatial location, burned area and weather observation i.e temperature, rain, humidity, and wind. The vegetation data is removed due to bad quality data. It does not include time, date, or year. So, trend analysis cannot be done for this set. Canadian system of rating fire danger is used. Each index is derived based on vegetation or weather conditions like rain, temp, humidity. It is not clear how certain indices are calculated when vegetation data is removed from original set. Also, the data doesn’t include any man-made incidents. So, we cannot predict any accidental fires like campfires, equipment malfunction etc. caused by negligence.

**Analysis**

EDA is done to understand what parameters may impact the burn area. For ease of analysis, the columns coordinates X, Y, months, days are converted to as.factors and data is summarized. Also added levels on weekdays to plot based on weekdays from Monday to Sunday. See plot below. The plot shows that the months of August and September has more burn area reported especially during weekends. This may be as more people visit parks during weekends.

A screenshot of a calendar

Description automatically generated with low confidence

*Plot 1: Shows that August and September months have more burn areas reported, which is predominantly on weekends*

Further EDA was done with weather factors vs burn area. Plots 2,3,4,5 show the effect of temp, RH, wind and rain wrt burn area.

|  |  |
| --- | --- |
| *Plot 2* | *Plot 3* |
| *Plot 4* | *Plot 5* |

Further EDA is done to understand the influence of the indices. FFMC – Fine Fuel Moisture Code and ISI – Initial Spread Index have an impact on the burn area. See Plot 6 and 7 for overall data with FMC and ISI. Other indices did not have stronger correlation to impact burn area. This can be further be seen on subset of weekend data, where FFMC and ISI are higher in the months of August and September.

|  |  |
| --- | --- |
| *Plot 6* | *Plot 7* |
| *Plot 8* | *Plot 9* |