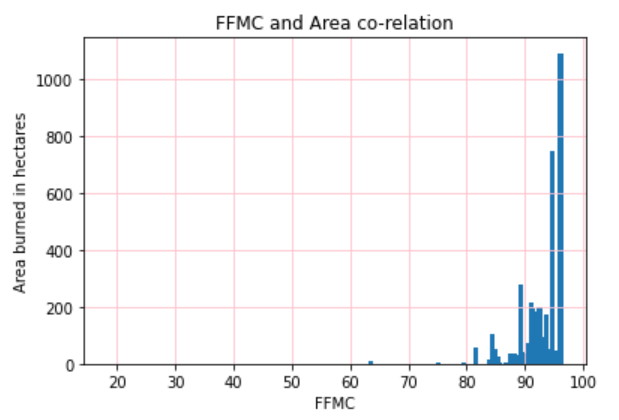
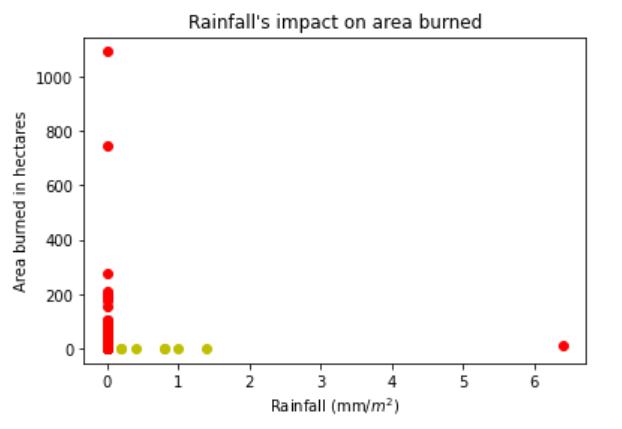
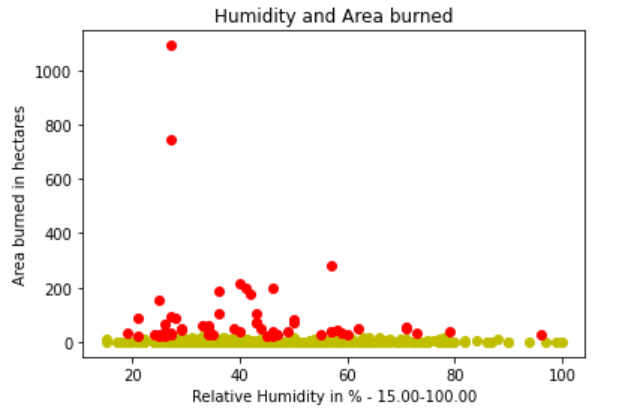
**Forest Fires Final Project**

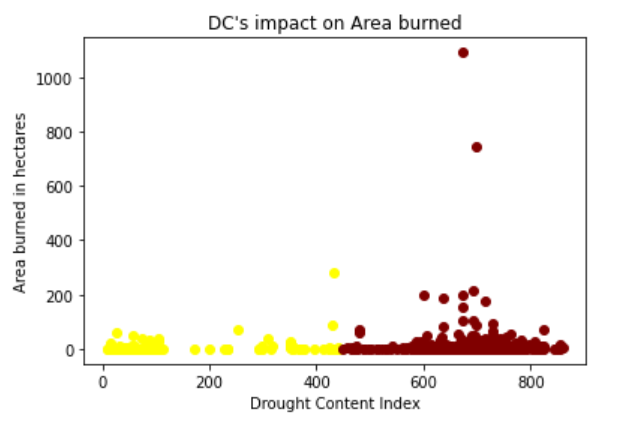
This is a multivariate dataset that contains data regarding the forest fires in Montesinho National Park in Parâmio, northeastern Portugal. It has a total of 517 row entries and 13 column entries, otherwise called attributes. Before the collection, the part was divided into zones/spatial coordinates on a 9x9 square scale. They were in 2 directions on the 2-D map, X and Y. The first column in the dataset contains the X-coordinate of the zone/part of the park. The values can range from 1 to 9. The second column in the dataset has the Y-coordinate of the zone/part of the park. These values can range from 2 to 9, as the value of 1 would mean a point is being considered as opposed to an area. The third attribute is the month of the year, abbreviated to “mon.” This contains the first 3 letters of the month in lower case. Possible values are: “jan”, “feb”, “mar”, ”apr”, “may”, “jun”, “jul”, “aug”, “sep”, “oct”, “nov”, or “dec”. The next column contains the day of the week. This contains the first 3 letters of the day in lower case. Possible values are: “mon”, “tue”, “wed”, “thu”, “fri”, “sat”, or “sun”.

FFMC (Fine Fuel Moisture Content) is the moisture content of litter and cured fine fuels. This is a positive number value, with possible values ranging from 18.7 to 96.2. DMC (Duff Moisture Content) is a numeric rating of the average moisture content of loosely compacted organic layers of moderate depth. This is similar to FFMC, but is not the same. The values can range from 1.1 to 291.3. DC is the drought count index, which measures the amount of dryness in the soil and helps to spread fires faster. The values for this range from 7.9 to 860.6. The ISI is the Initial Spread Index, which measures how fast the fire spreads once it is caught on. It can have values ranging from 0 to 56.1. The temp is the measure of the temperature in degree Celsius, it has values ranging from 2.2 degree Celsius to 33.30 degree Celsius. A higher temperature helps fires spread faster. Relative humidity is a percent measure of water vapor in the air. It has values ranging from 15.0 to 100.00. Wind speed is the speed of wind in kilometers per hour. It can have values ranging from 0.4 to 9.40 kmph. Rain contains the amount of rainfall in the area, it is measured in mm/m^2. A higher rainfall would lead to lower area burned by the fire. The area is the amount of area (in hectares) that was destroyed by the fire. It is measured in hectares and contains values from 0 to 1090.84.

I ignored the month, day, X, Y, DMC, ISI, wind, and temp, as they were not giving me any substantial results. The area was my dependent variable and FFMC, DC, Relative Humidity, and Rain made up my independent variables. I read the data into a DataFrame by using Pandas and then extracted the columns as lists. Then, I plotted each of the blow graphs by using the lists are axes. However, I used an index array to find whether values were above certain limits to color code it. Later, I added the titles and converted them to grids.







Temperature, FFMC, DC, DMC, ISI, and wind are directly proportional to the area burned. There were some outliers in the data, due to which there was higher or lower area burned at times. This is a multivariate dataset, a combination of factors will need to work together to produce the most devastating or least devastating fire. Conditions like FFMC, DMC, and DC are inter-related, while conditions like ISI and low temperatures are extremes – cannot co-exist. There was an inverse relationship between rainfall and relative humidity. There was a strong co-relation between FFMC and area burned – just like DC and area burned - as they were factors that helped fuel the fire. High humidity does not let the fire to last, just like high rainfall. I attempted to implement the regression model on this, but since the output values were eschewed towards 0, I was unsuccessful. Later, I realized that I had to either convert these to a smaller unit (square meters) or use a logarithmic transform. I could not figure out the logarithmic transform in time.

Most of the data was widely spread, as evidenced by the standard deviations of all the columns.

