I'd be happy to discuss the project with you. Please find attached the csv files containing the data.

Once you have understood what needs to be done, you can tell me how much time you would need to do the project. If you are well versed with R, this shouldn't take you more than a few hours.

The basic idea for the project is to determine which distribution GLM model fits the data best for each of the two cases (i.e.number of fires and area burnt).

The task is pretty simple. The model that needs to be fitted is of the form Fire~factor(Month)+factor(Year). Since the data is not in that form, you will need to modify it accordingly, such that the year and month are independent variables and the data is a dependent variable.

Once you have that sorted, you will take the first file called FRA which contains the area burnt by fires in a region for each month for a 26 year period. You will need to fit GLM models for appropriate distributions to the model to the data.

You may start with power law form f(x)= C x^a, Pareto or zero-inflated models here and try any other distributions that you feel will fit this type of data better.

The second file is called FRN and it contains the number of fires in each month for the same 26 year period. You may start with distributions such as Poisson and Negative binomial, which are the most common for count data. You will use the functions glm and glm.nb to fit the model for the data for these two distributions.

While you're doing this, you will also need to perform chi square goodness of fit tests for each distribution tested.

Once you have done all this, you will also need to provide plots for validation including qq and fitted residual plots.

You may take the missing values as zero.

Do you think you'll be able to do this?