Amazon Fires

Mid-project

DA Bootcamp Ironhack

OCT, 2022 | Adela & Bruno

Index

1/ Amazon Forest Context

 # 6.7 million square kilometers among 9 countries

or 28 times the size of the UK

 # 80K species of flora

and an estimated 400 billion trees standing

 # 6% of the world's oxygen production

act as a carbon sink, absorbing large amounts of carbon dioxide from the atmosphere (Leman J., 2021),

 # home for 30 million people

including almost 3 million indigenous people

2/ Problem

The rate at which the Amazon is deforested is a global concern, as it plays a key role in the environment and the atmosphere, not only in South America but also at a global scale.

Different studies indicate that deforestation has meant the **loss of 20% of the forest**, and if it continued at this rate by 2050, 40% of it would have disappeared (H Ter Steege, 2019). Between 36% and 57% of the Amazonian tree species could be in danger.

3/ Dataset information

Characteristics of the fires in the Amazon forest between 1998 and 2013,

Shape 1513059 rows × 15 columns,

Some of the main features:

- **C:** Carbon emissions in g C/m2,
- **NPP:** Net primary production, carbon retained in g C/m2,
- **Fire spots:** number of spot fires ignited outside the perimeter of the main fire.
- **Burned area:** area affected by each fire in ha.
- **Latitude/Longitude:** geographical coordinates of the fires

4/ Analysis

Did the number of fires increased?

Do carbon emissions correlate with the size of the burned areas?

What is the capacity of the affected areas to retain Carbon?

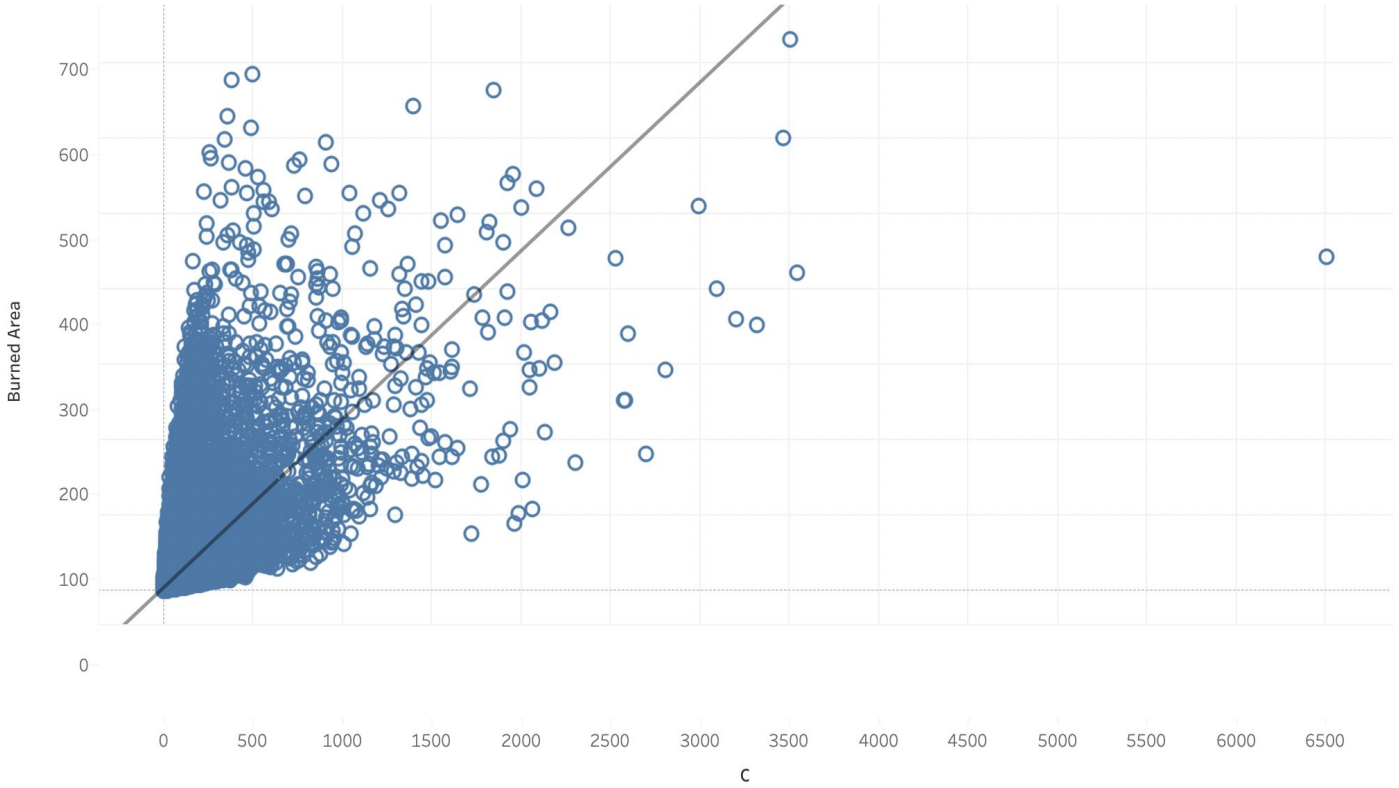
What are the most affected areas and how are they distributed in the territory?

How the Amazon fires behave throughout the months of the years?

Number of fires rapidly decreased and then stabilized

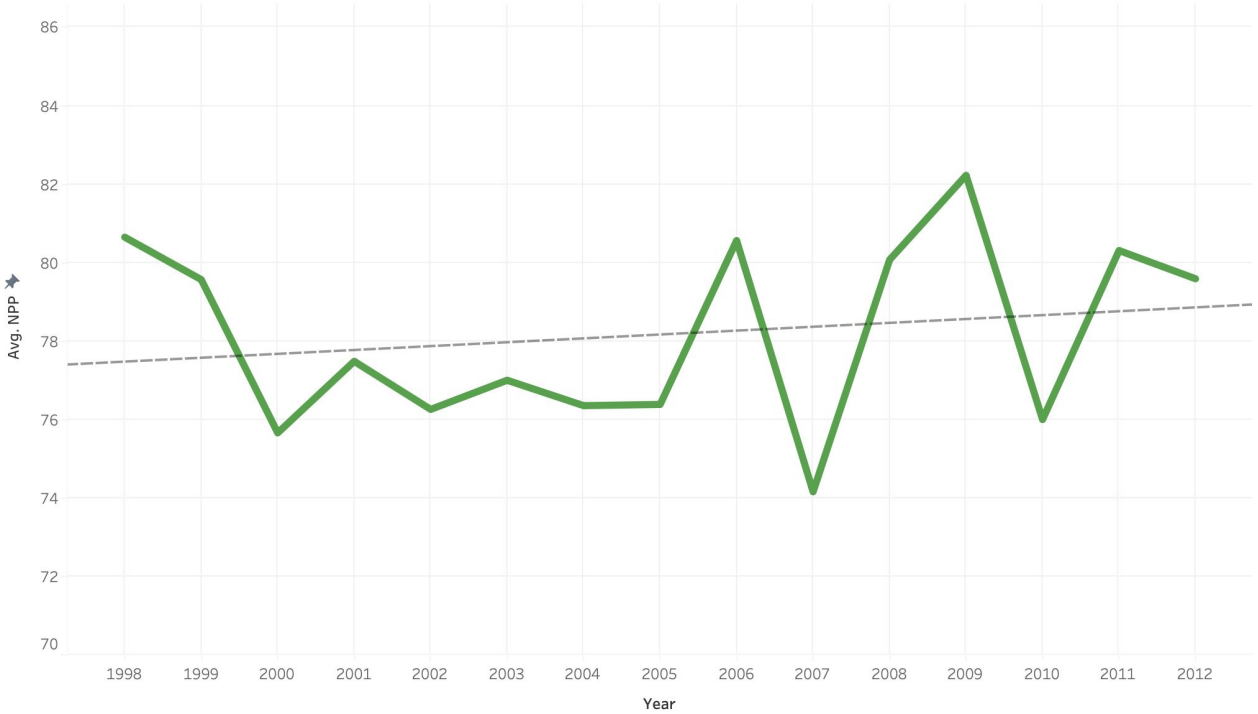
- Peak with more than 40K fires/year in 1998/99
- From 2000 to 2012, average of to 18.5K fires/year

Moderate correlation between Carbon Emission and Burned Area



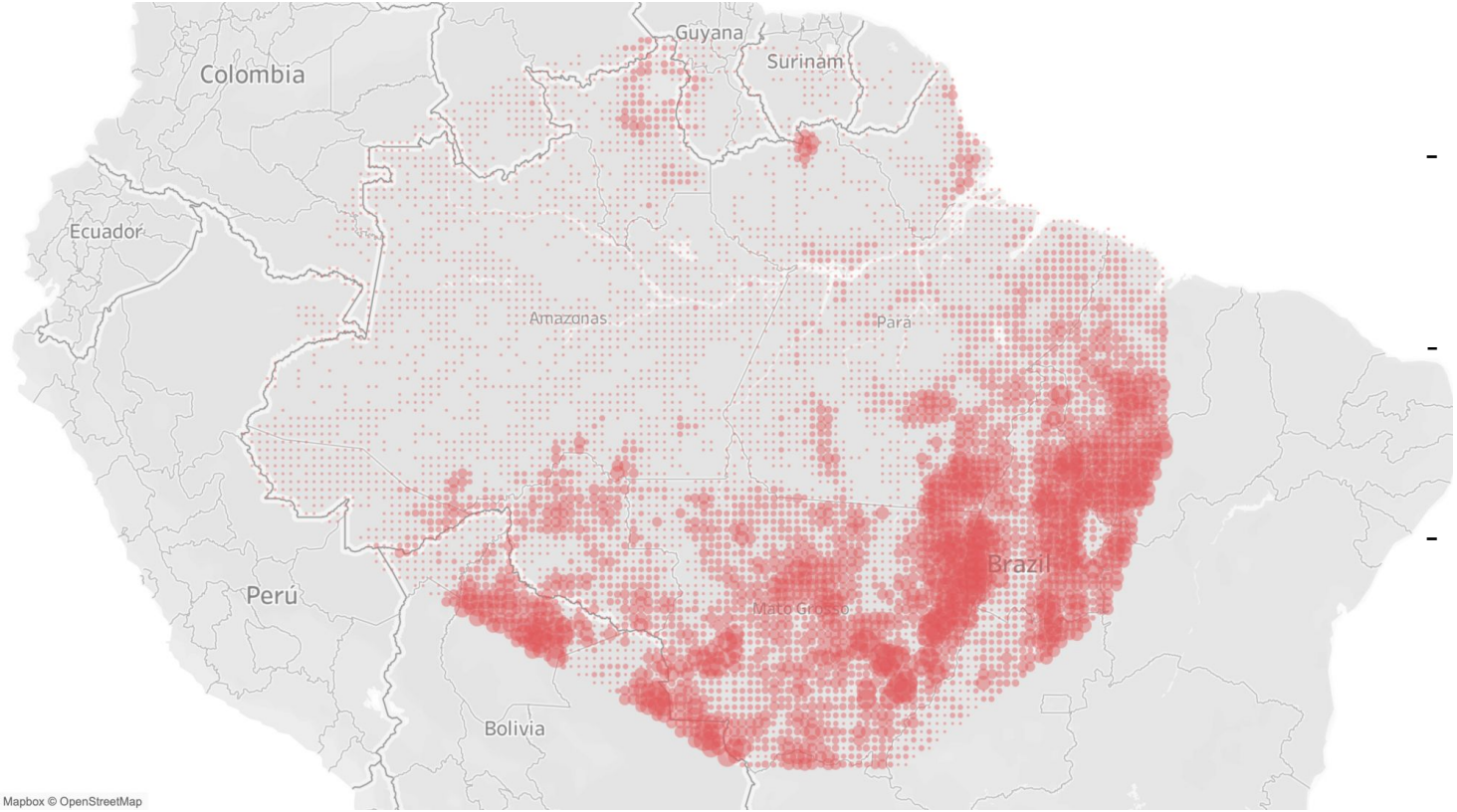
- Correlation of 0.6 between both features
- Carbon emissions don't necessarily increase in the same rate if burned area is bigger

Average capacity to retain carbon slightly increased



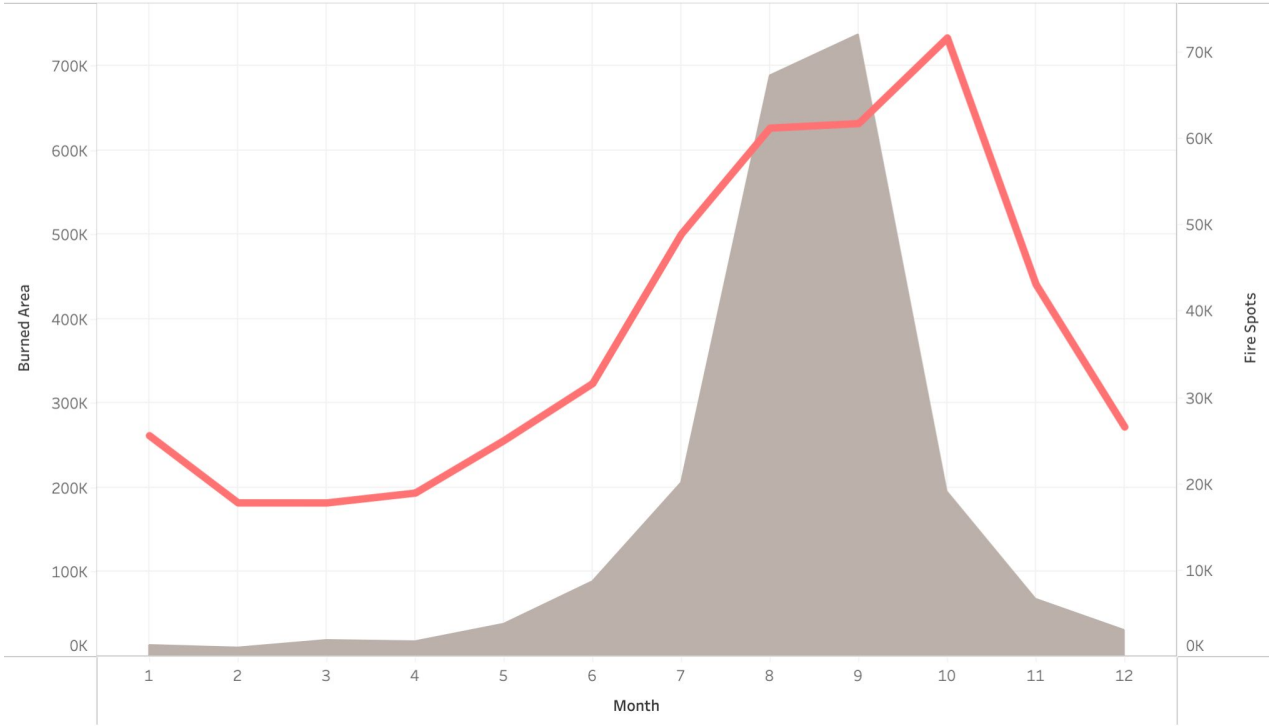
- Average NPP increased among the years reaching a peak of 82k g/C m2 in 2009
- Capacity of the affected areas to retain carbon from the atmosphere slightly improved

Burned Areas: mostly concentrated in the surroundings of the forest



- Southern and eastern regions of the amazon have the greater burned areas
- Most affected areas are in the states of Tocantins and Mato Grosso (Brazil)
- 2.115.433ha of area burned according to this dataset (0.3% of the whole territory)

Evidence of “fire season”, where both spot fires and burned area increase

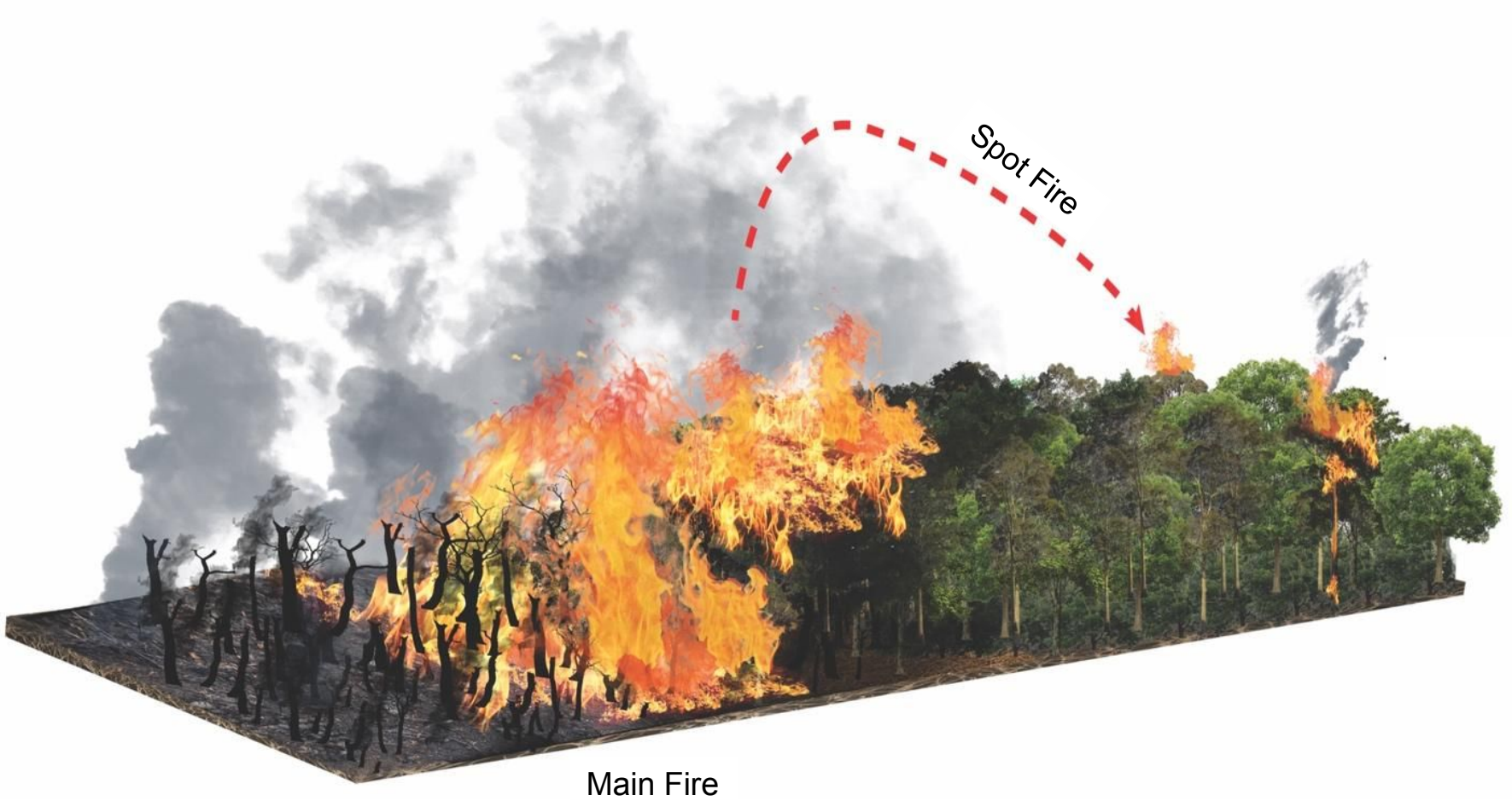


- Season of fires in the amazon occurs mainly between july and october
- Amount of spot fires increase significantly during fire season
- Both burned area and spot fires follow a similar pattern

5/ Predictive Model

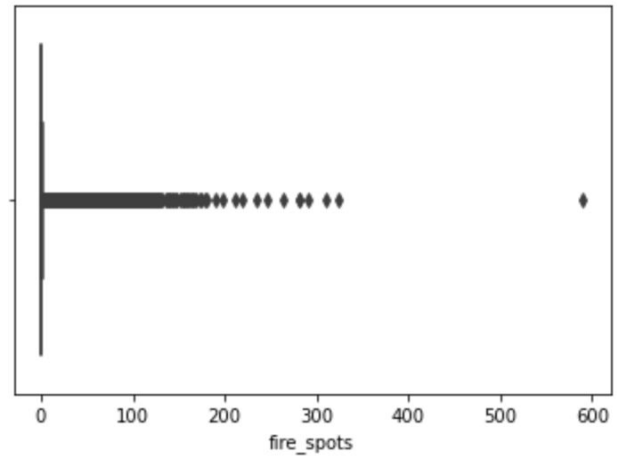
Spot Fires in the Amazon Forest

Target feature:
fire_spot_ranges


Spot Fire
Main Fire

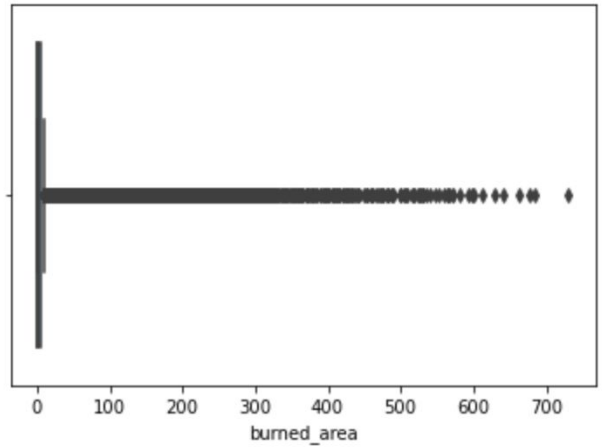
Data Cleaning

Features with high presence of outliers, such as fire spots and burned area, were kept



Fire Spots feature

Most of the fires had 0 spot fires, but 24% of them had 1 or more



Burned Areas feature

Burned areas varies from less than 1 ha to more than 700 ha

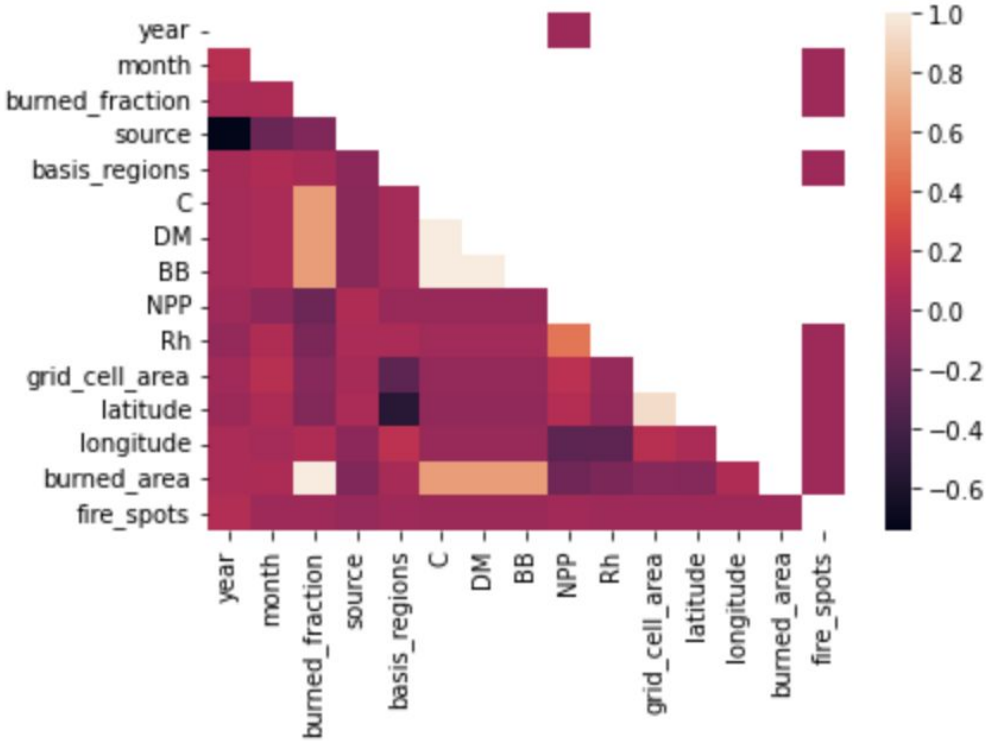
Feature Selection

Features with correlations higher than 0.9 were dropped

Multicollinearity analysis

Explanatory features dropped:

- burned_fraction
- DM (Dry matter)
- BB (Biomass burning)
- grilled_cel_area



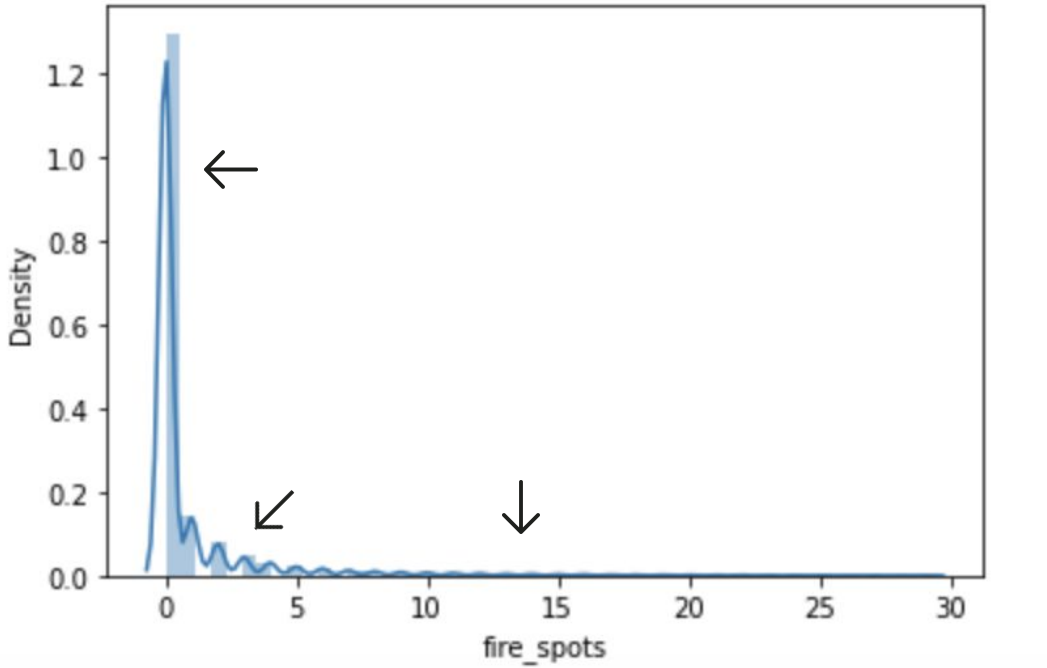
Multi Classification: KNN Model

Predicting different ranges of spot fires in the Amazon

Classification decision based on fire_spots distribution

Multi class decision:

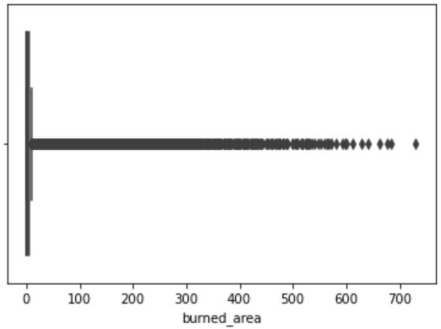
- 0: no spot fires detected
- 1: from 1 to 5 spot fires detected
- 2: more than 6 spot fires detected



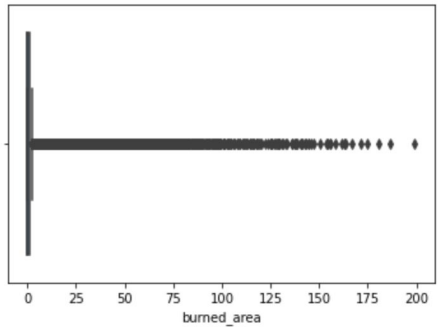
Feature scaling: scaler selected to deal with outliers

Scaler

- Robust Scaler
  - Removes the median and scales the data according to the quantile range



Original feature burned_area: data distributed between 0 and 700 ha

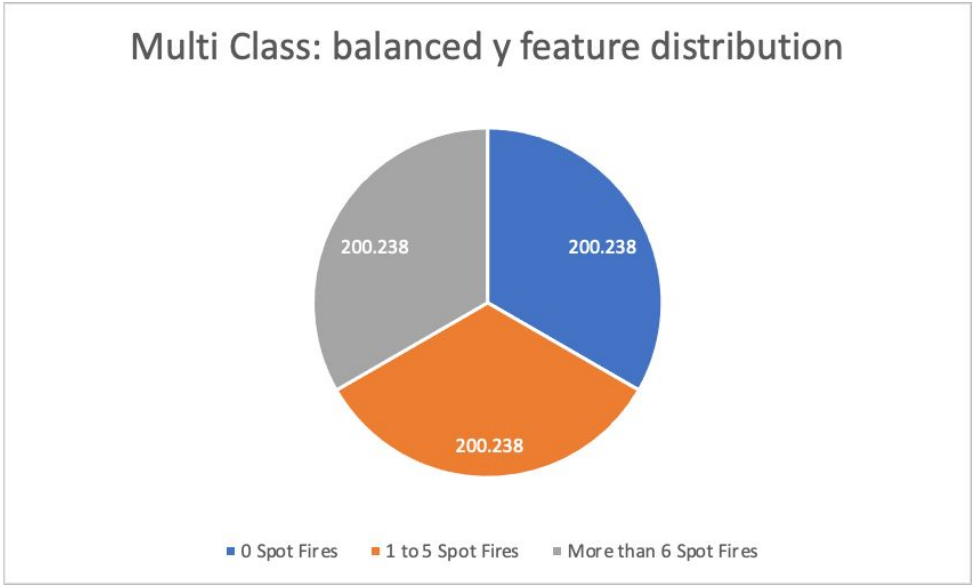


Robust Scaled feature burned_area: data distributed between 0 and 200 ha

Balancing the data

SMOTE oversampling method

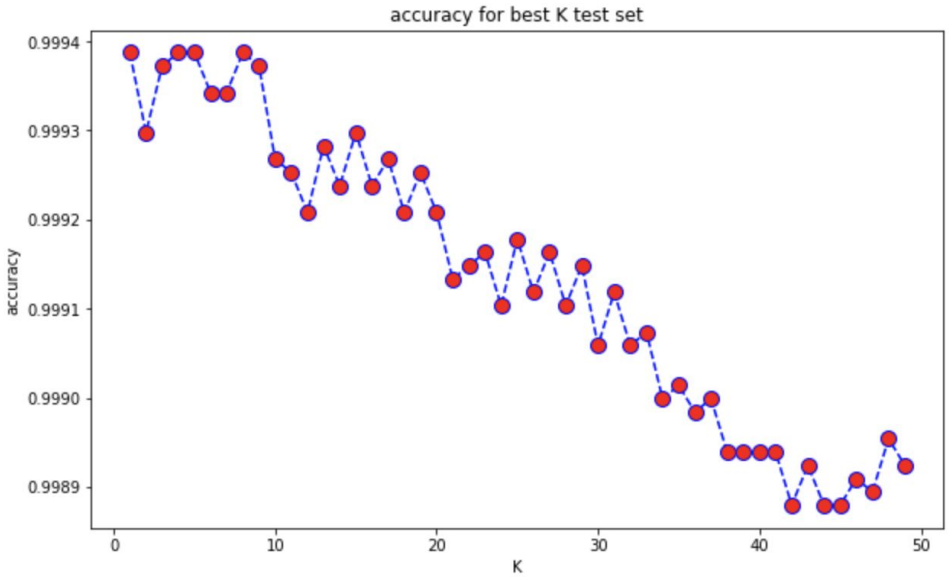
- Class 0: 74.8% -> 33.3%
- Class 1: 19.9% -> 33.3%
- Class 2: 5.2% -> 33.3%



Running the KNN model

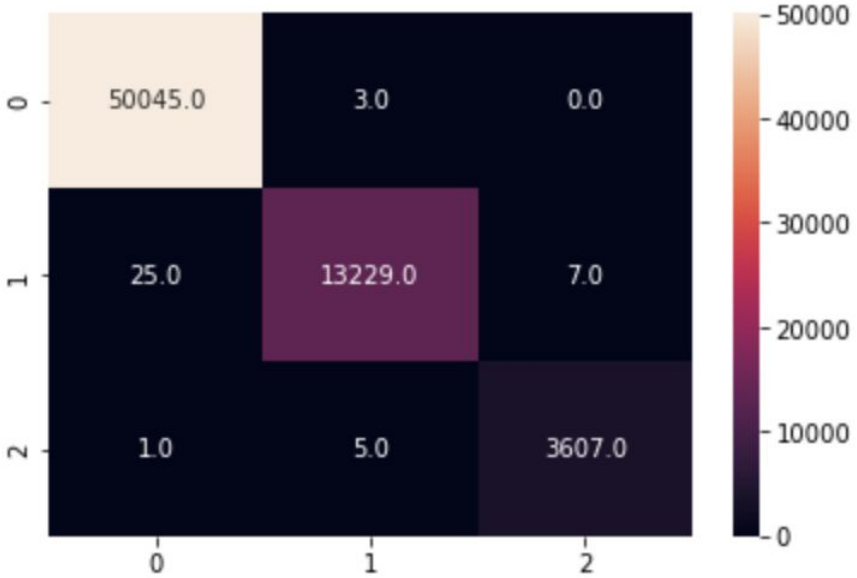
KNN Parameters

- weight = ‘distance’
  - greater influence from closer neighbors slightly improved the model
- K = 1
  - best number of neighbours according to the plot, showing accuracy result for each K value between 0 and 50



Results: high accuracy prediction for the classes of spot fires

Confusion Matrix
on the test set



99.6%
test set accuracy

100%
train set accuracy

0.3032 - 0.3064
conf interval for fire_spots
(98% conf level)

6/ Model Conclusions

- The model is able to predict the range of spot fires in the Amazon with an accuracy of more than 99% on the test set

- Overfits when the data is imbalanced and depending on the scalar adopted

- Highly sensitive to outliers, as the results improve when applying a scalar that reduces significantly the influence of them.

6/ General Conclusions

- From 1998 to 2000 the number of the fires in the Amazon rapidly decreased, but stabilized for the following 13 years, showing no clear trend for the future,

- There is an evident season of fires in the Amazon and the spot fires increase significantly in that period, a similar pattern followed by the burned areas,

- Predicting the spot fires could be relevant to control the spread of the fire, specially during the fire season,

- The multi class model high-performs with a 99% of accuracy on the test set to predict the 3 different ranges of spot fires in the Amazon forest,

7/ Recommendations for further studies

- Increase the number of classes to understand if the model is capable to predict with high accuracy more ranges of spot fires

- Explore the types of fire and fire methods to try to explain the presence of several outliers in the dataset (such as the relation between low carbon emissions in big burned areas)

- Study the impact of agriculture and livestock in the Amazon to better understand the presence of fires in the outskirts of the forest


Thank you

Annexes

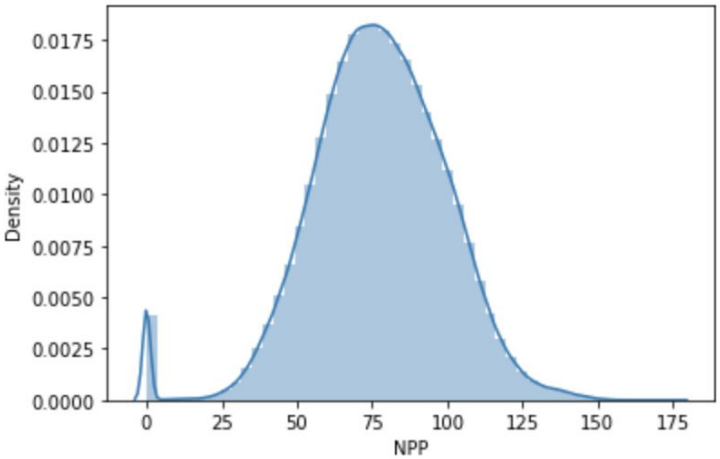
Prediction of spot fires in the Amazon

The main goal of this study is to understand some of the main characteristics of the fires in the Amazon and develop a model to predict the amount of spot fires for each main fire to address the problem of fire spreads in the forest.

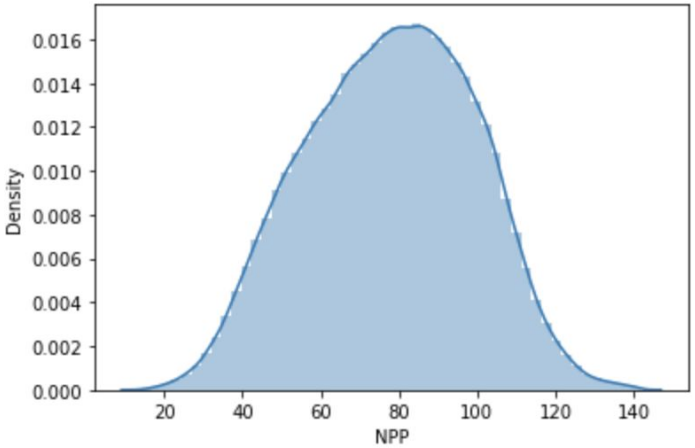
We carried out an analysis of a database with information of fires in the Amazon between the years of 1998 and 2013, following all the steps of the data science to analyze the data and develop the prediction model.

Data Cleaning

Features closer to a gaussian distribution (like Net Primary Production) had its outliers removed



Original NPP distribution



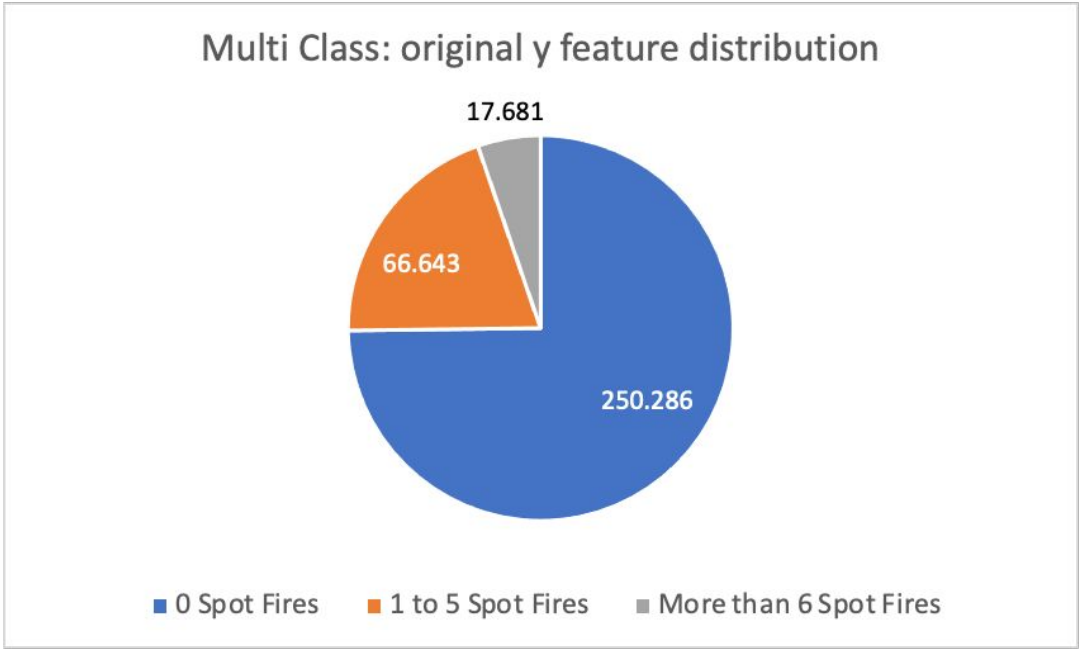
NPP range from 15 to 140 g of C per m2 after removing outliers

Classified data is imbalanced, mostly with 0 values

Imbalanced data

Original feature distribution

- 0: 74.8%
- 1: 19.9%
- 2: 5.3%



KNN Model: classification report (train and test sets)

	precision	recall	f1-score	support
0	1.00	1.00	1.00	200238
1	1.00	1.00	1.00	200238
2	1.00	1.00	1.00	200238
accuracy			1.00	600714
macro avg	1.00	1.00	1.00	600714
weighted avg	1.00	1.00	1.00	600714

	precision	recall	f1-score	support
0	1.00	1.00	1.00	50048
1	0.99	0.99	0.99	13261
2	1.00	0.99	1.00	3613
accuracy			1.00	66922
macro avg	1.00	0.99	0.99	66922
weighted avg	1.00	1.00	1.00	66922