2020 Brazil`s Wildfire Analysis

I want to thank Alice Adativa for presenting this amazing dataset for us and teaching some really nice techniques about plotting maps. Muito Obrigado Alice! Bons códigos!

If you want to check her notebook on this dataset, check the URL below:

https://www.kaggle.com/aliceadativa/an-lise-de-queimadas-no-brasil-2020

In Brail, and in the South America as well, almost all the wildfires are man made, and they have a wide range os reasons: cleaning huge filds for large crops, prepare of the fields, deforestation, vandalism, São João Balloons (https://i.ytimg.com/vi/1Gb8I8Xae1o/maxresdefault.jpg) and others. With more than 300.000 wildfires annualy, Brazil is on the 5th position on countries that emit more pollutants

Brief description of the data set and a summary of its attributes

This data set was downloaded from the Brazilian National Institute of Aerospacial Research website, and the URL is provided below.

URL: https://queimadas.dgi.inpe.br/queimadas/bdqueimadas

On the website, you can see all the wildfires that are going on in South America live, and you can also download all the data from any time period. All of the information provided, was collected by the Satelite "AQUA_M-T".

Data Descripction

Tha data set that is going to be studied have a shape of 220094 rows and 13 columns. Each row is a different wildfire entry and is described by the following features:

- datahora - date and time of the wildfire
- satelite - name of the satelite that collected the information
- pais - wildfire`s city
- estato - wildfire`s state
- municipio - wildfire`s city
- bioma - biome (collection of plants and animals that have common characteristics for the environment they exist in)
- dias sem chuva - days without raining when the wildfire was registered
- precipitacao - milimiter of rain from all the past days until the wildfire break
- risco fogo - predicted risk of breaking a wildfire for that specific day(0 = no chance; 1 - high chance)
- latitude - latitude
- longitude - longitude
- frp - Fire Radiative Power, MW (megawatts)

Initial plan for data exploration

First of all, we are going to take a look at the distribution of the predicted risk of breaking a wildfire on a map when the fire happened. Then we are going to see witch of the brazilian biomes are the worst ones when we are talking about the risk od breaking a wildfire.

After that, we are going to take a look at all the states and see witch one has more risk of beaking a wildfire. We are going to analyse deeply the top 2 states with the higher number of critical risk of fire and see if they fit on the biomes studies before. We are going to take a look as well at the top 1 city from each state to see if we can extrat some more info

Actions taken for data cleaning and feature engineering

In this section, I will give a brief explanation about what was the actions taken to clean the data for a better understanding ans for better vizualization.

- Installing missing packages, as geopandas and contextily. Both of them are for map vizualization
- Transform the features Latitute and Longitude to CRS(Coordinate Reference Sytstem)
- Get the riscofogo(risk of breaking out a wildfire) into categories, such as 'Cities/Rivers', 'Minimum', 'Low', 'medium', 'high', 'critical', so we can plot all the informationa and see exatly where the critical points were.
- Transform the column "datahora" into 3 different columns, year, month, and day, so we can divide the information between years, and in the future, make possible predictioins based on the past year
- Drop all the features that we already transformed and are not going to be used again, for example: "datahota", "latitude" and "longitude". Also drop the columns that have single values, for example: "country", and "satelite"
- Divide the dataset in two parts. The first have the 2020's brazilian wildfire data. The second one have the 2021's brazilian wildfire data

More detailed description are given through out the analysis below.

Key Findings and Insights

All the Key Findings and Insights are given through out the analysis. You can read the final conclusion paragraph at the end of the analysis.

Formulating Hypotesis about the data that we are going to analyse.

- The majority of brazilian wildfires happen in the central area of the country where the temperatures are high and the air humidity is very low during the whole year

- The state with less wildfires in Brazil is probably Rio Grande do Sul. I think that this is a plausable hypotesis, bacause this is the most southern state of Brazil and have the cooler temperature
- The rain forest (Amazonia) biome is one of the most humid biome in brazil, but it is also the one with more deforestation in Brazil. My hypotesis is, even being the most humid biome, it may have one of the higher numbers of wildfires

Next Steps in Analysing this data:

The next step in my point of view is to analyse where the bigger wildfires are breaking in Brazil, and study as well the number of wildfires per month per year and cross it with the weather stations to see if they are related some how, even knowing that the major number os wildfires in Brazil are caused by humans.

It is possible as well to cross the wildfires with the number of population to see if it is related some way. For example: If the city have more population, does it make the wildfires number drop or rise? For that we need a dataset with all the brazilian cities and population.

It is also possible to study the ammount of wildfires in places that have almost no population, but has some huge fileds of crops. My hypotesis for this one is that the number are going to skyrocket. Maybe we could create a new feature called wildfires/citizen/m^2. Therefore we can study more precisely where most wildfires happen.

We could also study the time feature and see what time of the day wildfires break more often, and therefore, create some plans to control it

Quality of this data set and a request for additional data

The quality of this data set was very good. It provided the exact place where the wildfire broke(state and city), the date and time that the fire broke and the power of the wildfire.

For this analysis be considered more complete, it needs more data. The number os citizens per city/state is a very important feature to study, knowing that most wildfires are man made. Also, we could get more data about the previous years, untill 2003(in 2003 the satelite that gathered all the data was replaced by a more accurate one, so there is a drastical change in number os wildfires between 2002 and 2003) and see how the wildfires are evolving through the years.

Another very interesting data to analyse as well, would be the government budget to fight against illegal wildfires. If the number of wildfires increased every year, and the budget as well, where all that money are going to? Brazil is a known country by its corruption, so who knows what

Enjoy the analysis below.

If you have any questions about the analysis, the tought process or even about some information that is in portuguese on the website provided, you can email me on my personal email:

rafaelgherrero@gmail.com

Google Drive link and installing Packages

```
pip install geopandas
```

```
Collecting geopandas
  Downloading https://files.pythonhosted.org/packages/d7/bf/e9cefb69d39155d122b6ddca53893b61535fa6ffdad70bf5ef708977f53f/geopan
     |████████████████████████████████| 1.0MB 4.0MB/s
Collecting fiona>=1.8
  Downloading https://files.pythonhosted.org/packages/9c/fc/9807326c37a6bfb2393ae3e1cca147aa74844562c4d5daa782d6e97ad2bc/Fiona-
     |████████████████████████████████| 15.4MB 234kB/s
Requirement already satisfied: shapely>=1.6 in /usr/local/lib/python3.7/dist-packages (from geopandas) (1.7.1)
Collecting pyproj>=2.2.0
  Downloading https://files.pythonhosted.org/packages/11/1d/1c54c672c2faf08d28fe78e15d664c048f786225bef95ad87b6c435cf69e/pyproj
     |████████████████████████████████| 6.6MB 27.3MB/s
Requirement already satisfied: pandas>=0.24.0 in /usr/local/lib/python3.7/dist-packages (from geopandas) (1.1.5)
Requirement already satisfied: certifi in /usr/local/lib/python3.7/dist-packages (from fiona>=1.8->geopandas) (2020.12.5)
Collecting munch
  Downloading https://files.pythonhosted.org/packages/cc/ab/85d8da5c9a45e072301beb37ad7f833cd344e04c817d97e0cc75681d248f/munch-
Requirement already satisfied: click>=4.0 in /usr/local/lib/python3.7/dist-packages (from fiona>=1.8->geopandas) (7.1.2)
Requirement already satisfied: setuptools in /usr/local/lib/python3.7/dist-packages (from fiona>=1.8->geopandas) (57.0.0)
Requirement already satisfied: six>=1.7 in /usr/local/lib/python3.7/dist-packages (from fiona>=1.8->geopandas) (1.15.0)
Collecting cligj>=0.5
  Downloading https://files.pythonhosted.org/packages/73/86/43fa9f15c5b9fb6e82620428827cd3c284aa933431405d1bcf5231ae3d3e/cligj-
```

```
Collecting click-plugins>=1.0
  Downloading https://files.pythonhosted.org/packages/e9/da/824b92d9942f4e472702488857914bdd50f73021efea15b4cad9aca8ecef/click_
Requirement already satisfied: attrs>=17 in /usr/local/lib/python3.7/dist-packages (from fiona>=1.8->geopandas) (21.2.0)
Requirement already satisfied: numpy>=1.15.4 in /usr/local/lib/python3.7/dist-packages (from pandas>=0.24.0->geopandas) (1.19.5
Requirement already satisfied: pytz>=2017.2 in /usr/local/lib/python3.7/dist-packages (from pandas>=0.24.0->geopandas) (2018.9)
Requirement already satisfied: python-dateutil>=2.7.3 in /usr/local/lib/python3.7/dist-packages (from pandas>=0.24.0->geopandas
Installing collected packages: munch, cligj, click-plugins, fiona, pyproj, geopandas
Successfully installed click-plugins-1.1.1 cligj-0.7.2 fiona-1.8.20 geopandas-0.9.0 munch-2.5.0 pyproj-3.1.0
```

```
pip install contextily
```

```
Requirement already satisfied: contextily in /opt/conda/lib/python3.7/site-packages (1.1.0)
Requirement already satisfied: rasterio in /opt/conda/lib/python3.7/site-packages (from contextily) (1.2.1)
Requirement already satisfied: mercantile in /opt/conda/lib/python3.7/site-packages (from contextily) (1.1.6)
Requirement already satisfied: joblib in /opt/conda/lib/python3.7/site-packages (from contextily) (1.0.1)
Requirement already satisfied: requests in /opt/conda/lib/python3.7/site-packages (from contextily) (2.25.1)
Requirement already satisfied: pillow in /opt/conda/lib/python3.7/site-packages (from contextily) (7.2.0)
Requirement already satisfied: matplotlib in /opt/conda/lib/python3.7/site-packages (from contextily) (3.4.0)
Requirement already satisfied: geopy in /opt/conda/lib/python3.7/site-packages (from contextily) (2.1.0)
Requirement already satisfied: geographiclib<2,>=1.49 in /opt/conda/lib/python3.7/site-packages (from geopy->contextily) (1.50)
Requirement already satisfied: kiwisolver>=1.0.1 in /opt/conda/lib/python3.7/site-packages (from matplotlib->contextily) (1.3.1
Requirement already satisfied: pyparsing>=2.2.1 in /opt/conda/lib/python3.7/site-packages (from matplotlib->contextily) (2.4.7)
Requirement already satisfied: python-dateutil>=2.7 in /opt/conda/lib/python3.7/site-packages (from matplotlib->contextily) (2.
Requirement already satisfied: cycler>=0.10 in /opt/conda/lib/python3.7/site-packages (from matplotlib->contextily) (0.10.0)
Requirement already satisfied: numpy>=1.16 in /opt/conda/lib/python3.7/site-packages (from matplotlib->contextily) (1.19.5)
Requirement already satisfied: six in /opt/conda/lib/python3.7/site-packages (from cycler>=0.10->matplotlib->contextily) (1.15.
Requirement already satisfied: click>=3.0 in /opt/conda/lib/python3.7/site-packages (from mercantile->contextily) (7.1.2)
Requirement already satisfied: attrs in /opt/conda/lib/python3.7/site-packages (from rasterio->contextily) (20.3.0)
Requirement already satisfied: certifi in /opt/conda/lib/python3.7/site-packages (from rasterio->contextily) (2020.12.5)
Requirement already satisfied: click-plugins in /opt/conda/lib/python3.7/site-packages (from rasterio->contextily) (1.1.1)
Requirement already satisfied: affine in /opt/conda/lib/python3.7/site-packages (from rasterio->contextily) (2.3.0)
Requirement already satisfied: snuggs>=1.4.1 in /opt/conda/lib/python3.7/site-packages (from rasterio->contextily) (1.4.7)
Requirement already satisfied: cligj>=0.5 in /opt/conda/lib/python3.7/site-packages (from rasterio->contextily) (0.7.1)
Requirement already satisfied: idna<3,>=2.5 in /opt/conda/lib/python3.7/site-packages (from requests->contextily) (2.10)
Requirement already satisfied: urllib3<1.27,>=1.21.1 in /opt/conda/lib/python3.7/site-packages (from requests->contextily) (1.2
Requirement already satisfied: chardet<5,>=3.0.2 in /opt/conda/lib/python3.7/site-packages (from requests->contextily) (3.0.4)
Note: you may need to restart the kernel to use updated packages.
```

Imports and CRS transformation for map vizualization

```
from shapely.geometry import Point
import pandas as pd
import geopandas as gpd
import seaborn as sns
import os
import numpy as np
import contextily
import matplotlib.pyplot as plt
import warnings
from scipy import stats
import math

%matplotlib inline
warnings.filterwarnings('ignore')
```

```
df = pd.read_csv('../input/queimadas-brasil-2020/Focos_2020-01-01_2020-12-31.csv')
df.head()
```

	datahora	satelite	pais	estado	municipio	bioma	diasemchuva	precipitacao	riscofogo	latitude	longitude	frp
0	2020/07/11 16:50:00	AQUA_M-T	Brasil	GOIAS	CORUMBA DE GOIAS	Cerrado	43.0	0.0	1.0	-15.914	-48.868	11.0
1	2020/07/11 16:50:00	AQUA_M-T	Brasil	GOIAS	CORUMBA DE GOIAS	Cerrado	42.0	0.0	0.9	-15.911	-48.864	9.5
2	2020/07/11 16:45:00	AQUA_M-T	Brasil	RIO DE JANEIRO	CANTAGALO	Mata Atlantica	38.0	0.0	0.8	-21.897	-42.340	13.0
[illegible]	2020/07/11	AQUA_M-	[illegible]	MINAS	MATIAS	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]

Criando um ponto de GPS através das colunas de latitude e longitude fornecidas

Creating a CRS point through the "latitude" and "longitude" columns

```
geometry = [Point(xy) for xy in zip(df['longitude'], df['latitude'])]
```

```
wildfire_brazil = gpd.GeoDataFrame(df, crs='EPSG:4326', geometry=geometry)
```

```
wildfire_brazil.head()
```

	datahora	satelite	pais	estado	municipio	bioma	diasemchuva	precipitacao	riscofogo	latitude	longitude	frp
0	2020/07/11 16:50:00	AQUA_M-T	Brasil	GOIAS	CORUMBA DE GOIAS	Cerrado	43.0	0.0	1.0	-15.914	-48.868	11.0
1	2020/07/11 16:50:00	AQUA_M-T	Brasil	GOIAS	CORUMBA DE GOIAS	Cerrado	42.0	0.0	0.9	-15.911	-48.864	9.5
2	2020/07/11 16:45:00	AQUA_M-T	Brasil	RIO DE JANEIRO	CANTAGALO	Mata Atlantica	38.0	0.0	0.8	-21.897	-42.340	13.0
3	2020/07/11 16:50:00	AQUA_M-T	Brasil	MINAS GERAIS	MATIAS CARDOSO	Caatinga	46.0	0.0	1.0	-14.838	-43.881	17.6

Transforming the dataset into Coordinate Reference System (CRS) of the GeoDataFrame

```
wildfire_brazil = wildfire_brazil.to_crs(crs='EPSG:3857')
```

Data manipulation

RiscoFogo

First of all, we need to understand what this feature mean.

The feature riscofogo, is a value between 0 and 1 to forecast the chance os a wildfire break in that specific area in the same day that it happened. This numberd is calculaed through a method shown in the URL below. Here is the abstract of the paper in english for a brief explanation

ABSTRACT

This document describes the method used to calculate the vegetation Fire Risk (FR) products at INPE´s Wildfire Monitoring Program. The meteorological principle used is that the higher the number of consecutive days without rain at a place, the higher the risk for burning its vegetation. Additionally considered are local effects of the vegetation type, maximum daily air temperature and minimum relative humidity, topographic elevation and latitude, and also the occurrence of fire in the area. The number of days without rain prior to the day of interest is calculated, up to the limit of 120 days; when rain occurs in the period the amount of precipitation and the interval to the day of the estimate are weighted in the estimates and a hypothetical number of consecutive dry days is obtained. Maximum air temperature above 30ºC and minimum Relative Humidity below 40% increase in a linear mode the FR for the day of interest; below and above those thresholds, respectively, FR is reduced. The local detection of active fires in satellite images, the topographical elevation and the latitude also increase the FR directly. The categories of FR are five in the scale of 0 to 1: Minimum, below 0.15; Low, from 0.15 to 0.4; Average, from 0.4 to 0.7; High, from 0.7 to 0.95 and; Critic, above 0.95. The daily FR forecasts up to five days and from one to four weeks follow the same principles of the observed FR, and are calculated data from numerical weather forecasts. The FR maps and the "Firegrams" produced are presented at INPE´s Wildfire Monitoring Program Portal and also automatically sent on and individual basis to registered users. Skill analysis of the FR for the five regions of Brazil are automatic using active fires detected by satellite monitoring, and the performance is above 95% for the FR classes of Average, High and Critical. This document replaces the previous 20130910_RF_V9 and all other issued before.

Keywords: Fire Risk. Fire Danger. Wildfires. Forest Fires. Vegetation.

http://queimadas.cptec.inpe.br/~rqueimadas/documentos/RiscoFogo_Sucinto.pdf

If you want to understand more about the article, and you cant read in portuguese, please email me so I can answer all your questions about the paper

Now that we understood what this number represents, we need to divide the column "riscofogo"(risk of fire) into categories, such as 'Cities/Rivers', 'Minimum', 'Low', 'medium', 'high', 'critical', so we can plot all the informationa and see exatly where the critical points were.

For that, we are going to set some ranges called below by categorias, and change the value of each row for that category.

But before we do that, we need to replace all the missing values and the values that are out of the range between 0 and 1.

```
wildfire_brazil.info()
```

```
<class 'geopandas.geodataframe.GeoDataFrame'>
RangeIndex: 222797 entries, 0 to 222796
Data columns (total 13 columns):
 #   Column        Non-Null Count   Dtype
---  ------        --------------   -----
 0   datahora      222797 non-null  object
 1   satelite      222797 non-null  object
 2   pais          222797 non-null  object
 3   estado        222797 non-null  object
 4   municipio     222797 non-null  object
 5   bioma         222797 non-null  object
 6   diasemchuva   217727 non-null  float64
 7   precipitacao  217727 non-null  float64
 8   riscofogo     217727 non-null  float64
 9   latitude      222797 non-null  float64
 10  longitude     222797 non-null  float64
 11  frp           220164 non-null  float64
 12  geometry      222797 non-null  geometry
dtypes: float64(6), geometry(1), object(6)
memory usage: 22.1+ MB
```

```
wildfire_brazil.shape
```

```
(222797, 13)
```

```
wildfire_brazil.describe()
```

	diasemchuva	precipitacao	riscofogo	latitude	longitude	frp
count	217727.000000	217727.000000	217727.000000	222797.000000	222797.000000	220164.000000
mean	20.614338	0.706719	-5.033625	-11.170909	-53.032945	66.164066
std	82.961591	3.034363	75.652691	6.460917	7.454275	146.748061
min	-999.000000	0.000000	-999.000000	-33.557000	-73.669000	0.000000
25%	3.000000	0.000000	0.400000	-15.914000	-57.393000	15.700000
50%	10.000000	0.000000	1.000000	-9.805000	-52.925000	29.700000
75%	36.000000	0.000000	1.000000	-6.647000	-47.386000	63.200000
max	120.000000	125.500000	1.000000	5.149000	-34.823000	8589.800000

If we take a closer look at the "riscofogo" feature, it shows us that the minimun value is -999.

After some research on the website privided above, I found a FAQ page, that answered what this value means.

URL https://queimadas.dgi.inpe.br/queimadas/portal/informacoes/perguntas-frequentes

The number -999 represents an invalid value, that can be connected to some urban areas or rivers/lakes

```
wildfire_brazil.riscofogo.describe()
```

```
count    217727.000000
mean         -5.033625
std          75.652691
min        -999.000000
25%           0.400000
50%           1.000000
```

```
75%           1.000000
max           1.000000
Name: riscofogo, dtype: float64
```

```
num = wildfire_brazil['riscofogo']._get_numeric_data()
```

Replacing all the values that are below 0 for -0.1

```
num[num < 0] = -0.1
```

```
wildfire_brazil.head()
```

	datahora	satelite	pais	estado	municipio	bioma	diasemchuva	precipitacao	riscofogo	latitude	longitude	frp
0	2020/07/11 16:50:00	AQUA_M-T	Brasil	GOIAS	CORUMBA DE GOIAS	Cerrado	43.0	0.0	1.0	-15.914	-48.868	11.0
1	2020/07/11 16:50:00	AQUA_M-T	Brasil	GOIAS	CORUMBA DE GOIAS	Cerrado	42.0	0.0	0.9	-15.911	-48.864	9.5
2	2020/07/11 16:45:00	AQUA_M-T	Brasil	RIO DE JANEIRO	CANTAGALO	Mata Atlantica	38.0	0.0	0.8	-21.897	-42.340	13.0
3	2020/07/11 16:50:00	AQUA_M-T	Brasil	MINAS GERAIS	MATIAS CARDOSO	Caatinga	46.0	0.0	1.0	-14.838	-43.881	17.6
4	2020/07/11 16:50:00	AQUA_M-T	Brasil	PARA	OBIDOS	Amazonia	0.0	0.4	0.2	-1.823	-55.207	18.7

```
wildfire_brazil.info()
```

```
<class 'geopandas.geodataframe.GeoDataFrame'>
RangeIndex: 222797 entries, 0 to 222796
Data columns (total 13 columns):
 #   Column        Non-Null Count   Dtype   
---  ------        --------------   -----   
 0   datahora      222797 non-null  object  
 1   satelite      222797 non-null  object  
 2   pais          222797 non-null  object  
 3   estado        222797 non-null  object  
 4   municipio     222797 non-null  object  
 5   bioma         222797 non-null  object  
 6   diasemchuva   217727 non-null  float64 
 7   precipitacao  217727 non-null  float64 
 8   riscofogo     217727 non-null  float64 
 9   latitude      222797 non-null  float64 
 10  longitude     222797 non-null  float64 
 11  frp           220164 non-null  float64 
 12  geometry      222797 non-null  geometry
dtypes: float64(6), geometry(1), object(6)
memory usage: 22.1+ MB
```

We still have some missing values. So lets replace it by the mean value. Usually this replace method is not the best one, but, since there are just 6200 missing values (~3%), I think that this replace is not going to change the results

```
wildfire_brazil['riscofogo'] = wildfire_brazil['riscofogo'].replace(np.nan, 0.5)
```

```
wildfire_brazil.head()
```

	datahora	satelite	pais	estado	municipio	bioma	diasemchuva	precipitacao	riscofogo	latitude	longitude	frp
0	2020/07/11 16:50:00	AQUA_M-T	Brasil	GOIAS	CORUMBA DE GOIAS	Cerrado	43.0	0.0	1.0	-15.914	-48.868	11.0
1	2020/07/11 16:50:00	AQUA_M-T	Brasil	GOIAS	CORUMBA DE GOIAS	Cerrado	42.0	0.0	0.9	-15.911	-48.864	9.5
2	2020/07/11	AQUA_M-	Brasil	RIO DE	CANTAGALO	Mata	38.0	0.0	0.8	-21.897	-42.340	13.0

```
semlat = wildfire_brazil.drop(columns=['latitude', 'longitude'])
```

```
correlations = semlat.corr()
print(correlations)
```

```
              diasemchuva  precipitacao  riscofogo       frp
diasemchuva      1.000000     -0.054046   0.220762  0.055845
precipitacao    -0.054046      1.000000  -0.370108 -0.034247
riscofogo        0.220762     -0.370108   1.000000  0.118800
frp              0.055845     -0.034247   0.118800  1.000000
```

Dividing all the values into categories and replacing it by names. We also have to replace all the negative and missing values as cities/rivers

```
categorias = [-0.1, 0, 0.15, 0.4, 0.7, 0.95, np.inf]

new_riscofogo = ['cities/rivers', 'minimum', 'low', 'medium', 'high', 'critical']
wildfire_brazil["riscofogo"] = pd.cut(wildfire_brazil["riscofogo"], categorias, labels = new_riscofogo)
```

```
wildfire_brazil.head()
```

	datahora	satelite	pais	estado	municipio	bioma	diasemchuva	precipitacao	riscofogo	latitude	longitude	frp
0	2020/07/11 16:50:00	AQUA_M-T	Brasil	GOIAS	CORUMBA DE GOIAS	Cerrado	43.0	0.0	critical	-15.914	-48.868	11.0
1	2020/07/11 16:50:00	AQUA_M-T	Brasil	GOIAS	CORUMBA DE GOIAS	Cerrado	42.0	0.0	high	-15.911	-48.864	9.5
2	2020/07/11 16:45:00	AQUA_M-T	Brasil	RIO DE JANEIRO	CANTAGALO	Mata Atlantica	38.0	0.0	high	-21.897	-42.340	13.0
3	2020/07/11 16:50:00	AQUA_M-T	Brasil	MINAS GERAIS	MATIAS CARDOSO	Caatinga	46.0	0.0	critical	-14.838	-43.881	17.6

```
wildfire_brazil['riscofogo'].unique()
```

```
['critical', 'high', 'low', 'cities/rivers', 'medium', 'minimum', NaN]
Categories (6, object): ['cities/rivers' < 'minimum' < 'low' < 'medium' < 'high' < 'critical']
```

```
wildfire_brazil['riscofogo'].value_counts()
```

```
critical         109554
medium            31242
low               26693
high              25955
cities/rivers     15539
minimum           12560
Name: riscofogo, dtype: int64
```

```
wildfire_brazil['riscofogo'] = wildfire_brazil['riscofogo'].fillna('cities/rivers')
```

```
wildfire_brazil['riscofogo'].unique()
```

```
['critical', 'high', 'low', 'cities/rivers', 'medium', 'minimum']
Categories (6, object): ['cities/rivers' < 'minimum' < 'low' < 'medium' < 'high' < 'critical']
```

```
wildfire_brazil['riscofogo'].value_counts()
```

```
critical         109554
medium            31242
low               26693
high              25955
cities/rivers     16793
minimum           12560
Name: riscofogo, dtype: int64
```

DataHora

Now that we have tha feature riscofogo all set, we need to replace this datahora by 3 different features, that are going to allow ur to divide the dataset into 2 (2020's brazilian wildfire and 2021's brazilians wildfire)

We are going to use the function str.split(expand=True) twice, once to create date and time, and the second time to split the date into year, month and day

```
wildfire_brazil[['date', 'time']] = wildfire_brazil.datahora.str.split(expand=True)
```

```
wildfire_brazil.head()
```

	datahora	satelite	pais	estado	municipio	bioma	diasemchuva	precipitacao	riscofogo	latitude	longitude	frp
0	2020/07/11 16:50:00	AQUA_M-T	Brasil	GOIAS	CORUMBA DE GOIAS	Cerrado	43.0	0.0	critical	-15.914	-48.868	11.0
1	2020/07/11 16:50:00	AQUA_M-T	Brasil	GOIAS	CORUMBA DE GOIAS	Cerrado	42.0	0.0	high	-15.911	-48.864	9.5
2	2020/07/11 16:45:00	AQUA_M-T	Brasil	RIO DE JANEIRO	CANTAGALO	Mata Atlantica	38.0	0.0	high	-21.897	-42.340	13.0

```
wildfire_brazil[['year', 'month', 'day']] = wildfire_brazil.date.str.split("/",expand=True)
wildfire_brazil.head()
```

	datahora	satelite	pais	estado	municipio	bioma	diasemchuva	precipitacao	riscofogo	latitude	longitude	frp
0	2020/07/11 16:50:00	AQUA_M-T	Brasil	GOIAS	CORUMBA DE GOIAS	Cerrado	43.0	0.0	critical	-15.914	-48.868	11.0
1	2020/07/11 16:50:00	AQUA_M-T	Brasil	GOIAS	CORUMBA DE GOIAS	Cerrado	42.0	0.0	high	-15.911	-48.864	9.5
2	2020/07/11 16:45:00	AQUA_M-T	Brasil	RIO DE JANEIRO	CANTAGALO	Mata Atlantica	38.0	0.0	high	-21.897	-42.340	13.0
3	2020/07/11 16:50:00	AQUA_M-T	Brasil	MINAS GERAIS	MATIAS CARDOSO	Caatinga	46.0	0.0	critical	-14.838	-43.881	17.6
4	2020/07/11 16:50:00	AQUA_M-T	Brasil	PARA	OBIDOS	Amazonia	0.0	0.4	low	-1.823	-55.207	18.7

```
wildfire_brazil['year'].unique()
```

```
array(['2020'], dtype=object)
```

Now we are going to drop the features that are not going to be used again.

The column datahora was already divided in 3 different features, therefore, we can drop it. Satelite and Pais(Country), can also be droped because they only have one unique value.

The feature "date" is also going to be droped bacause it is not going to give the insights that we need.

```
wildfire_brazil = wildfire_brazil.drop(columns=[ 'datahora', 'satelite', 'pais', 'date',])
```

```
wildfire_brazil_2020 = wildfire_brazil.loc[(wildfire_brazil['year']=='2020')]
wildfire_brazil_2020.head()
```

	estado	municipio	bioma	diasemchuva	precipitacao	riscofogo	latitude	longitude	frp	geometry	time	year
0	GOIAS	CORUMBA DE GOIAS	Cerrado	43.0	0.0	critical	-15.914	-48.868	11.0	POINT (-5439960.876 -1794765.623)	16:50:00	2020
1	GOIAS	CORUMBA DE GOIAS	Cerrado	42.0	0.0	high	-15.911	-48.864	9.5	POINT (-5439515.598 -1794418.358)	16:50:00	2020
2	RIO DE JANEIRO	CANTAGALO	Mata Atlantica	38.0	0.0	high	-21.897	-42.340	13.0	POINT (-4713267.240 -2499163.337)	16:45:00	2020
3	MINAS GERAIS	MATIAS CARDOSO	Caatinga	46.0	0.0	critical	-14.838	-43.881	17.6	POINT (-4884810.575 -1670537.264)	16:50:00	2020

```
wildfire_brazil_2020.info()
```

```
<class 'geopandas.geodataframe.GeoDataFrame'>
Int64Index: 222797 entries, 0 to 222796
Data columns (total 14 columns):
```

```
 #   Column        Non-Null Count   Dtype
---  ------        --------------   -----
 0   estado        222797 non-null  object
 1   municipio     222797 non-null  object
 2   bioma         222797 non-null  object
 3   diasemchuva   217727 non-null  float64
 4   precipitacao  217727 non-null  float64
 5   riscofogo     222797 non-null  category
 6   latitude      222797 non-null  float64
 7   longitude     222797 non-null  float64
 8   frp           220164 non-null  float64
 9   geometry      222797 non-null  geometry
 10  time          222797 non-null  object
 11  year          222797 non-null  object
 12  month         222797 non-null  object
 13  day           222797 non-null  object
dtypes: category(1), float64(5), geometry(1), object(7)
memory usage: 24.0+ MB
```

Data Analysis

In this analysis, we are going to look only at the wildfires that happened in 2020

From the data cleaning, we already know some information about the dataset:

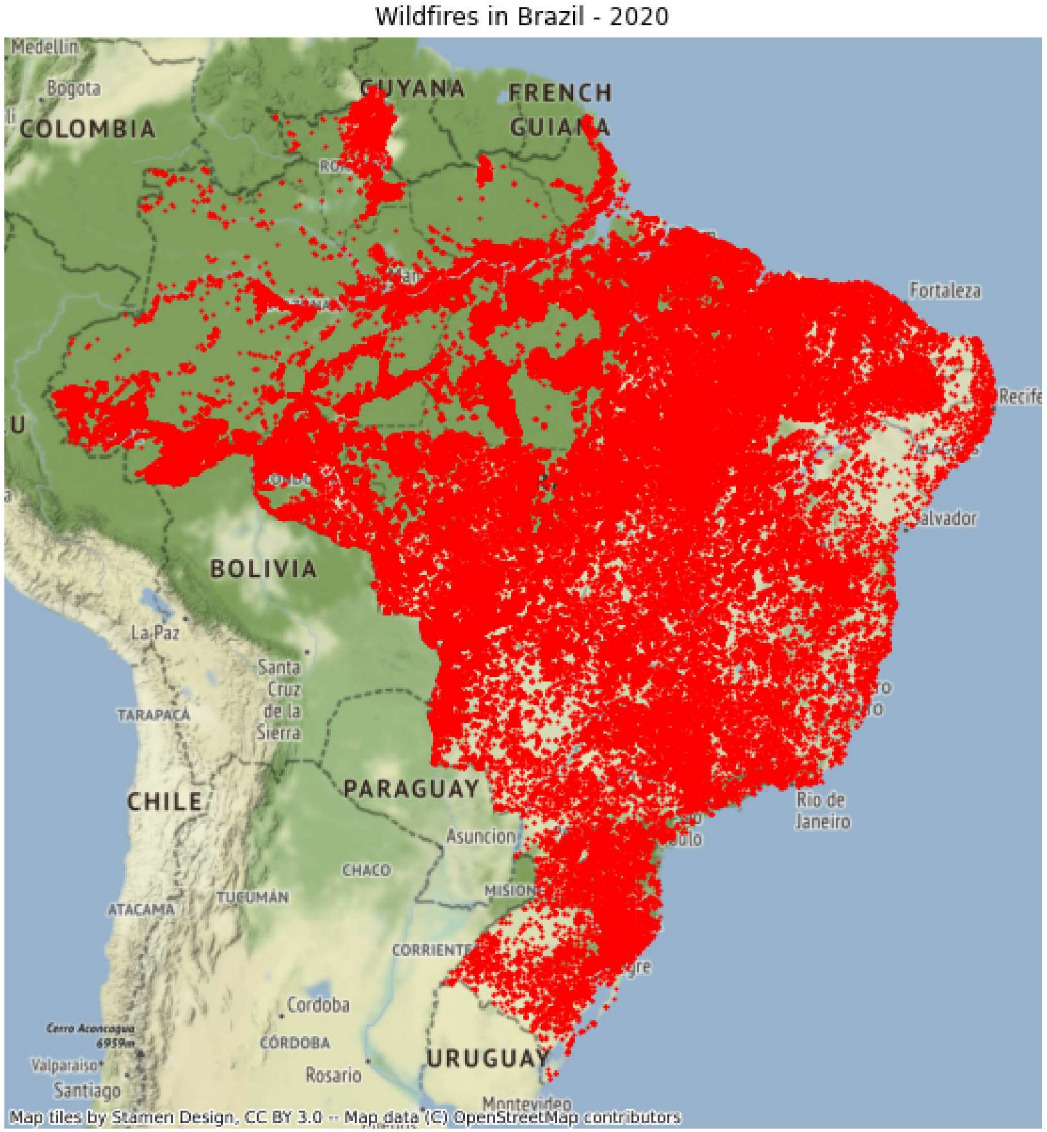
- Number of Fires registred in 2020 = 222.797
- This means that in brazil, there were aprox 610 wildfires per day

The map shown below represents all the wildfires in brazil in the year of 2020

```
ax = wildfire_brazil_2020.plot(figsize=(15,10), markersize=1, color='red')
contextily.add_basemap(ax)
```

```
ax.set_axis_off()

plt.title('Wildfires in Brazil - 2020')
plt.show()
```



Risco de Fogo

We divides tha riscofogo in 6 differents categorys between 0 and 1

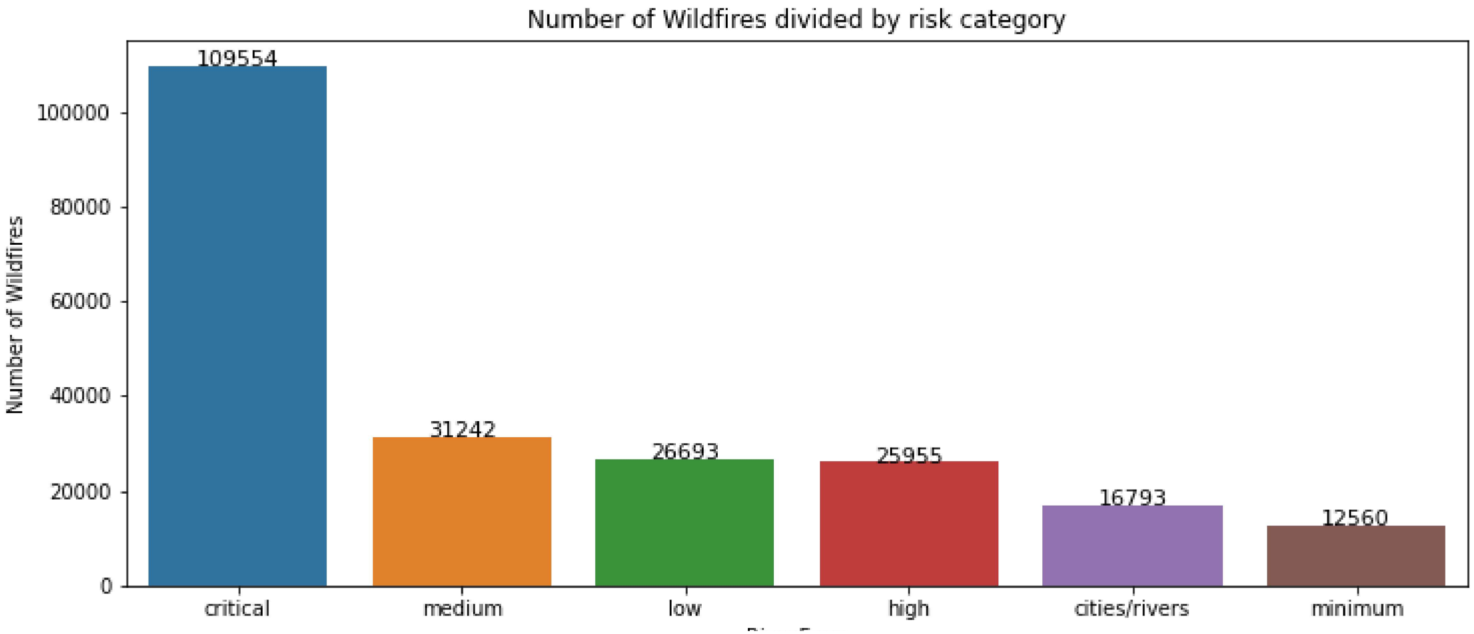
'Cities/Rivers', 'Minimum', 'Low', 'medium', 'high', 'critical'

1. Cities/Rivers = 0;
2. Minimum, below 0,15
3. Low, between 0,15 and 0,4;
4. Medium, between 0,4 and 0,7;
5. High, between 0,7 and 0,95;
6. critical, above 0.95

```
wildfire_brazil_2020['riscofogo'].value_counts()
```

```
critical         109554
medium            31242
low               26693
high              25955
cities/rivers     16793
minimum           12560
Name: riscofogo, dtype: int64
```

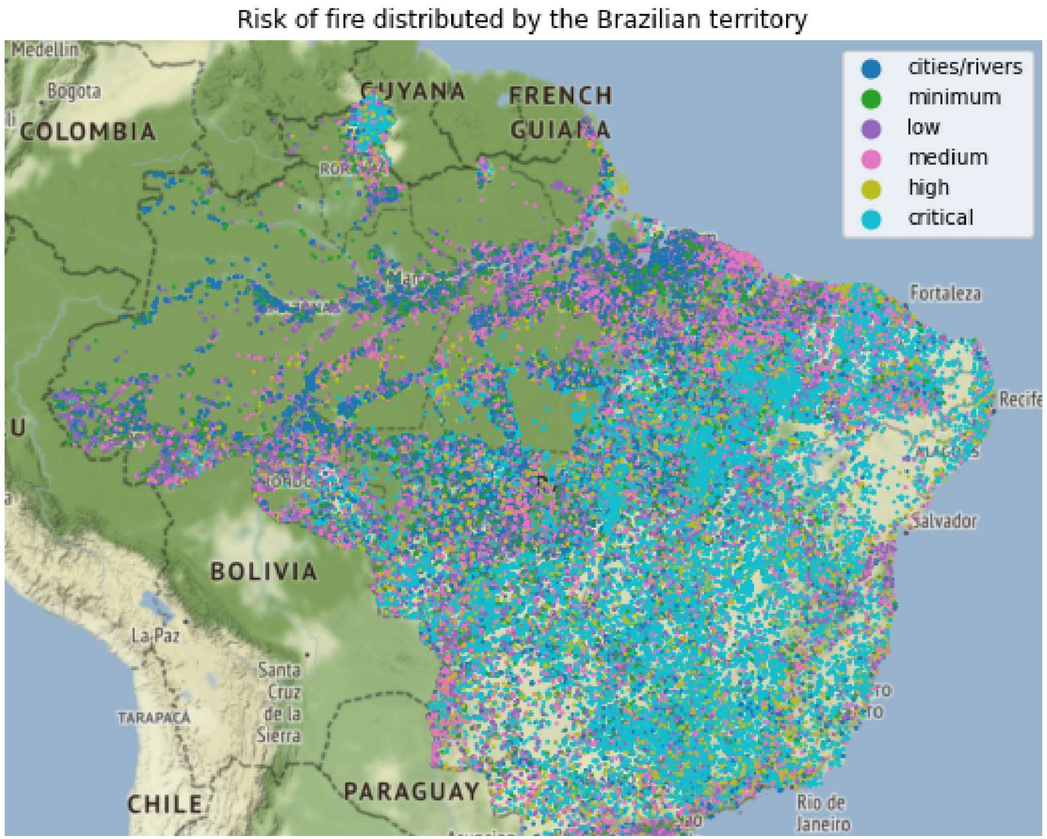
```
plt.figure(figsize=(12,5))
ax = sns.countplot(x="riscofogo", data=wildfire_brazil_2020, order = wildfire_brazil_2020['riscofogo'].value_counts().index)
ax.set_xlabel('RiscoFogo')
ax.set_ylabel('Number of Wildfires')
plt.title('Number of Wildfires divided by risk category')
for rect in ax.patches:
    ax.text (rect.get_x() + rect.get_width()  / 2,rect.get_height()+ 0.75,rect.get_height(),horizontalalignment='center', fontsize =
```



Most of the wildfires that happen in brazil, have a critical prediction (52%). That means that we had the information that the wildfire was going to happen in that regions

```
ax = wildfire_brazil_2020.plot(figsize=(15, 10), column='riscofogo', markersize=1, legend=True)
contextily.add_basemap(ax)
ax.set_axis_off()

plt.title('Risk of fire distributed by the Brazilian territory')
plt.show()
```



As we predicted before in our hypotesis, the central area of brazil is the region that has the most number of critical risk of fire

Hypotesis 1 - The majority of brazilian wildfires happen in the central area of the country where the temperatures are high and the air humidity is very low during the whole year

Brazilian Biomes

Now that we know where all the critical regions are, lets cross that information with the Brazilian Biomes, and see witch one prones to burn more

Now, we will discuss the different biomes of Brazil, and analyse wildfires in it. But what is a biome?

It is a group of ecosystem with a common history and climate and therefore being characterized by the same animals and plants. Biome concept includes all living beings of a community but in practice biomes are defined by the vegetation general appearance. Is a unit of biological classification used to classify major geographic regions of the world

Nowadays, six different types of biomes are defined in Brazil: Amazon, Atlantic Forest, Cerrado, Caatinga, Pampa and Pantanal. Lets undertand a little bit what are the mais caractheristics of each one of them before diving to the analysis.

- Amazon - The Amazon basin area is the world's largest forest and the most biodiverse biome in Brazil. It occupies almost 50% of the country and is seriously threatened due to the deforestation caused by logging industries and soybean crops. Currently it is estimated that 16% of the amazon rainforest is under anthropic pressure.
- Atlantic forest - Is a tropical forest covering the coastal region of Brazil and therefore it is characterized by humid winds coming from the sea and steep reliefs. It is composed of a variety of ecosystems because a high variety of altitudes, latitudes and therefore, climates ranging from semideciduous seasonal forests to open mountain fields and Araucaria's forests in the south.
- Cerrado - It is the second largest biome of South American covering 22% of Brazil. It is considered the richest savannah in the world in terms of species number. It contains a high level of endemic species and it is considered one of the global hotspots in terms of biodiversity. Containing 11,627 species of plants (of which 40% are endemic) and 200 animal species, 137 of which are threatened to extinction.

- Caatinga - It is the only exclusively Brazilian biome and occupies 11% of the country. Its name comes from a native language of Brazil, the Tupi-Guarani and means white forest. However, this biome is the most undervalued and little known because of its aridity. The climate of the caatinga is semi arid and soils are stony. The vegetation is steppe and savannah like and is characterized by a great adaptation to aridity (xerophyte vegetation) often prickly. The caatinga trees lose their leaves during dry season, leaving a landscape full of whitish trunks.
- Pampa - Pampa is a biome that occupies a single state in Brazil, Rio Grande do Sul covering only 2% of the country. Pampa biome is also very well represented in Uruguay and northern Argentina. It includes a large diversity of landscapes, ranging from plains, mountains and rocky outcrops, but the more typical are grass fields with hills and isolated trees nearby water courses.

```
wildfire_brazil_2020['bioma'].unique()
```

```
array(['Cerrado', 'Mata Atlantica', 'Caatinga', 'Amazonia', 'Pantanal',
       'Pampa'], dtype=object)
```

```
wildfire_brazil_2020['bioma'].value_counts()
```

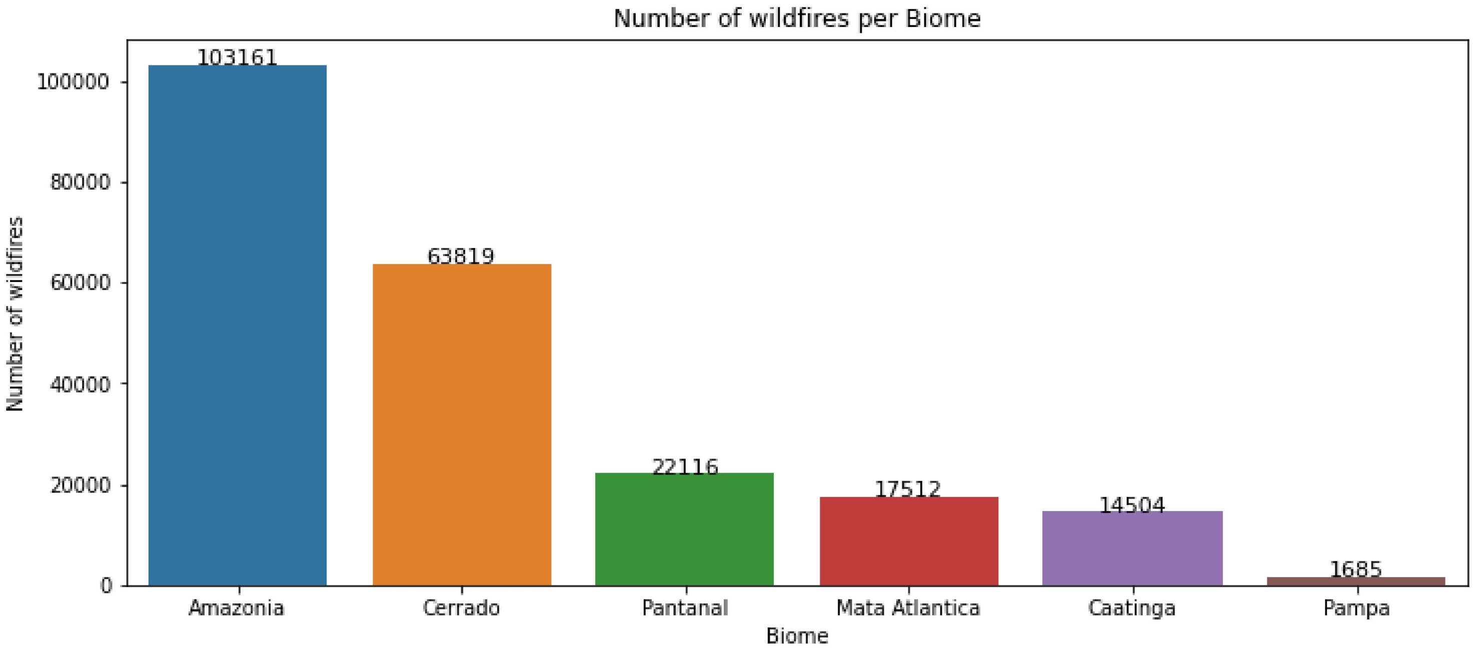
```
Amazonia          103161
Cerrado            63819
Pantanal           22116
Mata Atlantica     17512
Caatinga           14504
Pampa               1685
Name: bioma, dtype: int64
```

Now we know that amazonia is the biome with the largest number os wildfires in Brazil, and the biome with less wildfires is Pampa, located in the south region of Brazil

This statement proves my hypotesys right. Even being a really humid biome, Amazonia still suffers a lot with wildfires that are generated by deforestation

Hypotesis - The rain forest (Amazonia) biome is one of the most humid biome in brazil, but it is also the one with more deforestation in Brazil. My hypotesis is, even being the most humid biome, it may have one of the higher numbers of wildfires

```
plt.figure(figsize=(12,5))
ax = sns.countplot(x="bioma", data=wildfire_brazil_2020, order = wildfire_brazil_2020['bioma'].value_counts().index)
ax.set_xlabel('Biome')
ax.set_ylabel('Number of wildfires')
plt.title('Number of wildfires per Biome')
for rect in ax.patches:
    ax.text (rect.get_x() + rect.get_width()  / 2,rect.get_height()+ 0.75,rect.get_height(),horizontalalignment='center', fontsize =
```

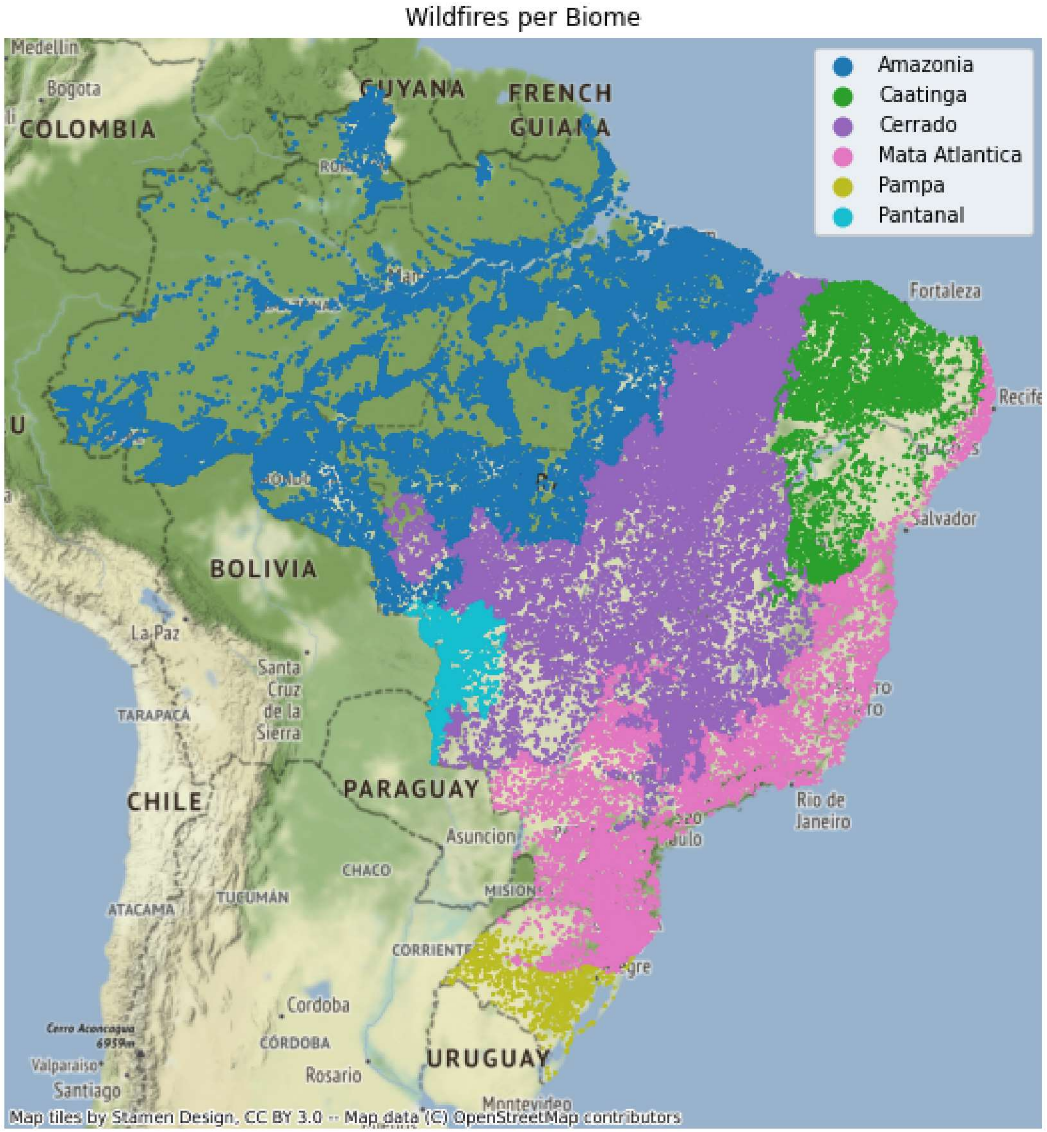


Now, lets take a look at the map, and see where all of the wildfires are located. All of biomes are divides by colors. And each color represents a biome

```
ax = wildfire_brazil_2020.plot(figsize=(15, 10), column='bioma', markersize=1, legend=True)
contextily.add_basemap(ax)
ax.set_axis_off()

plt.title('Wildfires per Biome')
```

```
plt.show()
```



Now that we have a top view of the problem, lets dive deeper. Now we are going to study the top 2 worst biomes, Amazonia and Cerrado

Amazônia

From the biome study, we already know that Amazonia has 47% of all registered wildfires in 2020. But one question still stands, are those wildfires in the critical risk? Or they couldn't be predicted?

Lets take a look at the distribution od the risk of fire in this biome

```
amazonia_2020 = wildfire_brazil_2020.query('bioma == "Amazonia"')
amazonia_2020.head()
```

	estado	municipio	bioma	diasemchuva	precipitacao	riscofogo	latitude	longitude	frp	geometry	time	yea
4	PARA	OBIDOS	Amazonia	0.0	0.4	low	-1.823	-55.207	18.7	POINT (-6145615.128 -202969.680)	16:50:00	202
5	PARA	BARCARENA	Amazonia	1.0	2.8	cities/rivers	-1.605	-48.674	10.1	POINT (-5418364.895 -178691.154)	16:50:00	202
7	PARA	NOVO PROGRESSO	Amazonia	36.0	0.0	medium	-7.681	-56.080	172.6	POINT (-6242797.044 -857617.685)	16:50:00	202
9	PARA	NOVO PROGRESSO	Amazonia	36.0	0.0	medium	-7.678	-56.054	220.4	POINT (-6239902.737 -857280.704)	16:50:00	202

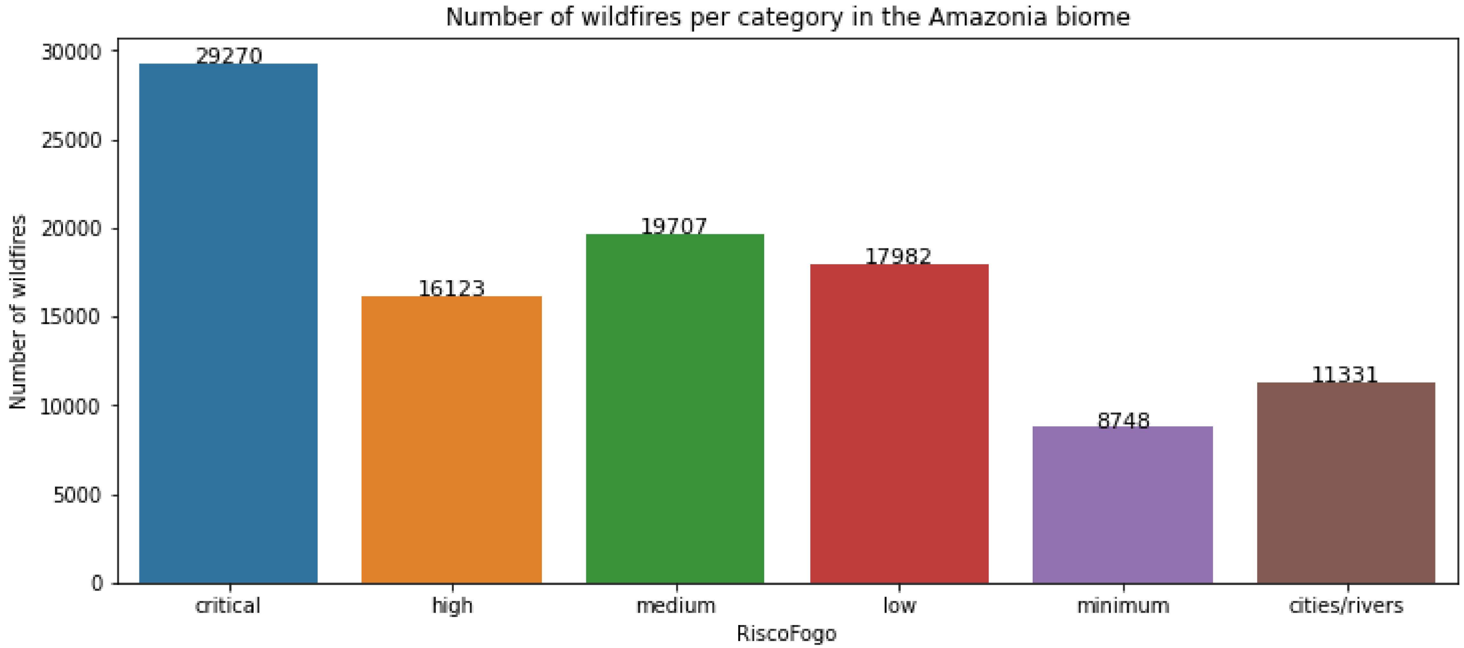
```
amazonia_2020.shape
```

```
(103161, 14)
```

```
amazonia_2020['riscofogo'].value_counts().sort_index(ascending=True)
```

```
cities/rivers    11331
minimum           8748
low              17982
medium           19707
high             16123
critical         29270
Name: riscofogo, dtype: int64
```

```
risco_ordem = ['critical', 'high', 'medium', 'low', 'minimum', 'cities/rivers']
plt.figure(figsize=(12,5))
ax = sns.countplot(x="riscofogo", data=amazonia_2020, order=risco_ordem)
ax.set_xlabel('RiscoFogo')
ax.set_ylabel('Number of wildfires')
plt.title('Number of wildfires per category in the Amazonia biome')
for rect in ax.patches:
    ax.text (rect.get_x() + rect.get_width()  / 2,rect.get_height()+ 0.75,rect.get_height(),horizontalalignment='center', fontsize =
```

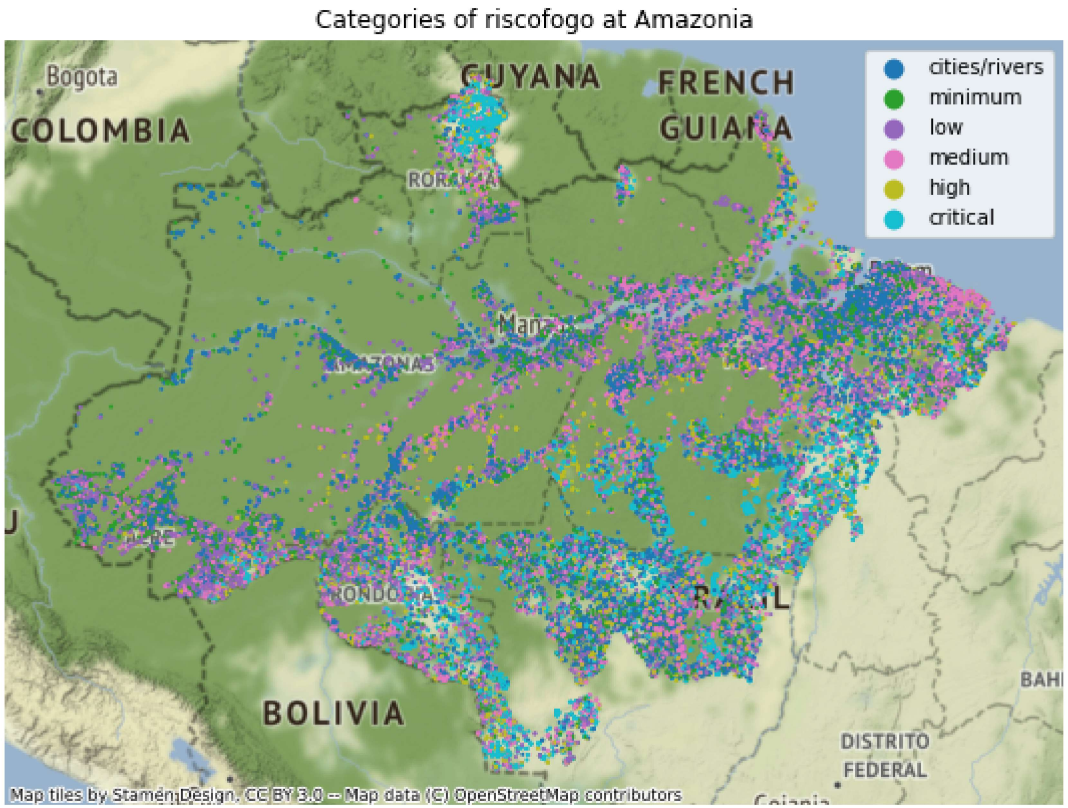


I was expecting a much higher number of critical days in Amazonia. But this number is actually a lot plausable, knowing that the most of the wildfires are human made in this Biome

Only 27% of all Brazilian critical cases belong to the Amazonia biome

```
ax = amazonia_2020.plot(figsize=(12, 7), column='riscofogo', markersize=1, legend=True)
contextily.add_basemap(ax)
ax.set_axis_off()

plt.title('Categories of riscofogo at Amazonia')
plt.show()
```




```
amazonia_2020['estado'].value_counts()
```

```
amazonia_2020['estado'].value_counts()
```

```
PARA            38603
MATO GROSSO     20648
AMAZONAS        16729
RONDONIA        11140
ACRE             9193
MARANHAO         3589
RORAIMA          1930
AMAPA             750
TOCANTINS         579
Name: estado, dtype: int64
```

The worst state in the Amazonia biome, is Para. After we analyse the second worst biome, we will take a closer look at the worst cities as well

We can see that the majority of the wildfires are located at the southest part of the biome, making border with Cerrado, the second most afected biome.

Lets take a closer look to at Cerrado's biome

Cerrado

Carrado is the second worst biome when we are talking about wildfires. Lets take a deeper look at this biome

```
cerrado_2020 = wildfire_brazil_2020.query('bioma == "Cerrado"')
cerrado_2020.head()
```

	estado	municipio	bioma	diasemchuva	precipitacao	riscofogo	latitude	longitude	frp	geometry	time
0	GOIAS	CORUMBA DE GOIAS	Cerrado	43.0	0.0	critical	-15.914	-48.868	11.0	POINT (-5439960.876 -1794765.623)	16:50:00
1	GOIAS	CORUMBA DE	Cerrado	42.0	0.0	high	-15.911	-48.864	9.5	POINT (-5439515.598	16:50:00

-871885.401)

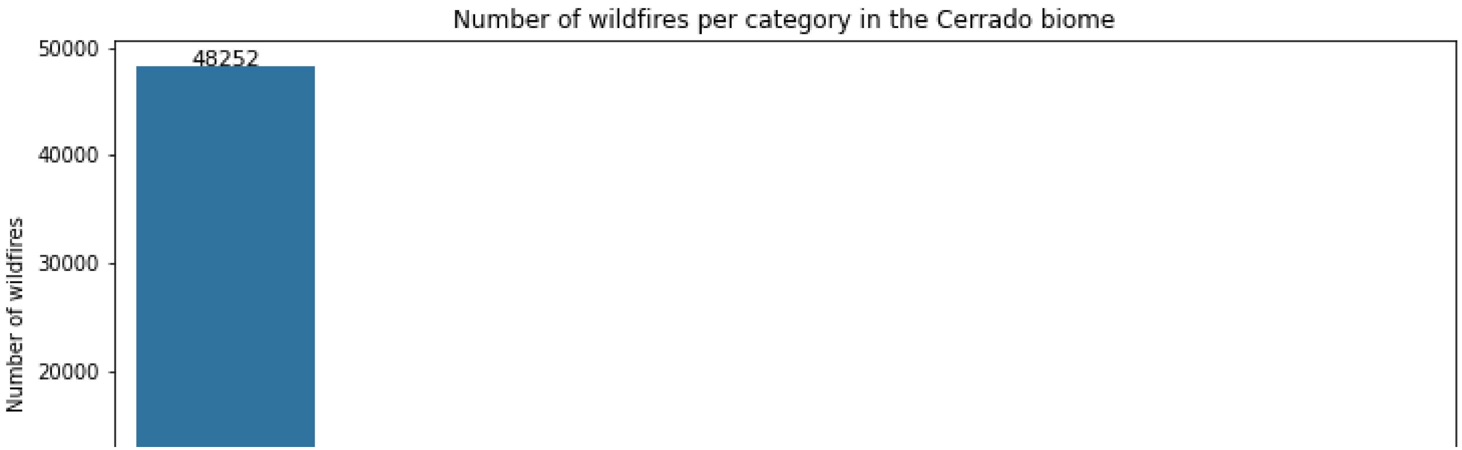
```
cerrado_2020.shape
```

```
(63819, 14)
```

```
cerrado_2020['riscofogo'].value_counts().sort_index(ascending=True)
```

```
cities/rivers     2117
minimum           1528
low               3244
medium            4625
high              4053
critical         48252
Name: riscofogo, dtype: int64
```

```
risco_ordem = ['critical', 'high', 'medium', 'low', 'minimum', 'cities/rivers']
plt.figure(figsize=(12,5))
ax = sns.countplot(x="riscofogo", data=cerrado_2020, order=risco_ordem)
ax.set_xlabel('RiscoFogo')
ax.set_ylabel('Number of wildfires')
plt.title('Number of wildfires per category in the Cerrado biome')
for rect in ax.patches:
    ax.text (rect.get_x() + rect.get_width()  / 2,rect.get_height()+ 0.75,rect.get_height(),horizontalalignment='center', fontsize =
```



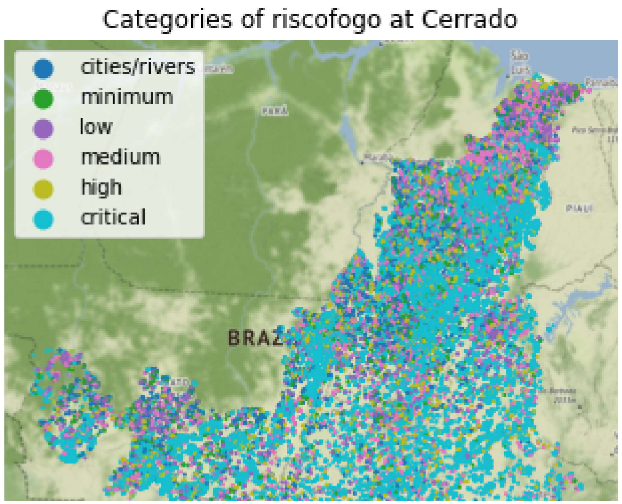
After tha analysis, we can conclude that 45% of all critical cases of risk fire in Brazil, are located in Cerrado. Witch means that this state is the one with more natural fires caused by the dry weather and elevated temperatures.

But in the other hand, it contains just 28% of all Brazilian wildfires

Now lets take a look at the map and find out witch states are included in this biome, and where in the continent is locates this biome.

```
ax = cerrado_2020.plot(figsize=(12, 7), column='riscofogo', markersize=1, legend=True)
contextily.add_basemap(ax)
ax.set_axis_off()

plt.title('Categories of riscofogo at Cerrado')
plt.show()
```




```
cerrado_2020['estado'].value_counts()
```

```
MATO GROSSO            13853
MARANHAO               13073
TOCANTINS              11514
GOIAS                   5730
MINAS GERAIS            5595
PIAUI                   4153
BAHIA                   4135
MATO GROSSO DO SUL      2775
SAO PAULO               2729
DISTRITO FEDERAL         196
PARANA                    61
RONDONIA                   5
Name: estado, dtype: int64
```

States Analysis

Intro States

Number of wildfires per state in 2020

```
wildfire_brazil_2020['estado'].value_counts()
```

```
MATO GROSSO            47708
PARA                   38603
MARANHAO               16817
AMAZONAS               16729
TOCANTINS              12093
MATO GROSSO DO SUL     12080
RONDONIA               11145
PIAUI                   9317
ACRE                    9193
MINAS GERAIS            8737
BAHIA                   7912
SAO PAULO               6123
GOIAS                   6008
CEARA                   3979
RIO GRANDE DO SUL       3612
PARANA                  3519
SANTA CATARINA          2425
RORAIMA                 1930
PERNAMBUCO              1017
PARAIBA                  910
AMAPA                    750
RIO DE JANEIRO           693
RIO GRANDE DO NORTE      664
ESPIRITO SANTO           401
DISTRITO FEDERAL         196
ALAGOAS                  159
SERGIPE                   77
Name: estado, dtype: int64
```

Now we are going to analyse deeper the top 2 worst states. The chosen ones are Mato Grosso and Para. and then we will try to extract some usefull information about the worst cities in this 2 states

Mato Grosso - Worst state in Brazil 's critical category and the state that has 21% of all registers in 2020 (47.708 wildfires)

Para - Worst state in Cerrado's critical category and the state that has 44% of all critical registers in Brazil in the year of 2020 (38.603 wildfires)

Mato Grosso

```
mato_grosso_state = wildfire_brazil_2020.query('estado == "MATO GROSSO"')
mato_grosso_state.head()
```

	estado	municipio	bioma	diasemchuva	precipitacao	riscofogo	latitude	longitude	frp	geometry	time
8	MATO GROSSO	CAMPINAPOLIS	Cerrado	47.0	0.0	critical	-14.347	-53.372	65.6	POINT (-5941343.863 -1614057.282)	16:50:00
15	MATO GROSSO	SANTA CARMEM	Amazonia	48.0	0.0	high	-12.077	-54.750	40.5	POINT (-6094742.121 -1354472.758)	16:50:00
20	MATO GROSSO	SANTA CARMEM	Amazonia	49.0	0.0	high	-12.073	-54.756	47.4	POINT (-6095410.038 -1354017.406)	16:50:00
57	MATO GROSSO	SANTA CARMEM	Amazonia	58.0	0.0	high	-12.137	-55.048	30.5	POINT (-6127915.329 -1361303.868)	16:50:00

```
mato_grosso_state['estado'].value_counts()
```

```
MATO GROSSO    47708
Name: estado, dtype: int64
```

```
mato_grosso_state['riscofogo'].value_counts()
```

```
critical        32435
medium           3717
low              3409
cities/rivers    3248
high             2915
minimum          1984
Name: riscofogo, dtype: int64
```

```
mato_grosso_state['bioma'].value_counts()
```

```
Amazonia    20648
Cerrado     13853
Pantanal    13207
Name: bioma, dtype: int64
```

```
mato_grosso_state_amazonia = mato_grosso_state.query('bioma == "Amazonia"')
mato_grosso_state_amazonia['riscofogo'].value_counts()
```

```
critical         11177
cities/rivers     2448
medium            2087
low               2048
high              1527
minimum           1361
Name: riscofogo, dtype: int64
```

```
mato_grosso_state_pantanal = mato_grosso_state.query('bioma == "Pantanal"')
mato_grosso_state_pantanal['riscofogo'].value_counts()
```

```
critical         10516
high               788
medium             787
low                686
cities/rivers      224
minimum            206
Name: riscofogo, dtype: int64
```

```
mato_grosso_state_cerrado = mato_grosso_state.query('bioma == "Cerrado"')
mato_grosso_state_cerrado['riscofogo'].value_counts()
```

```
critical         10742
medium             843
low                675
high               600
```

```
cities/rivers      576
minimum            417
Name: riscofogo, dtype: int64
```

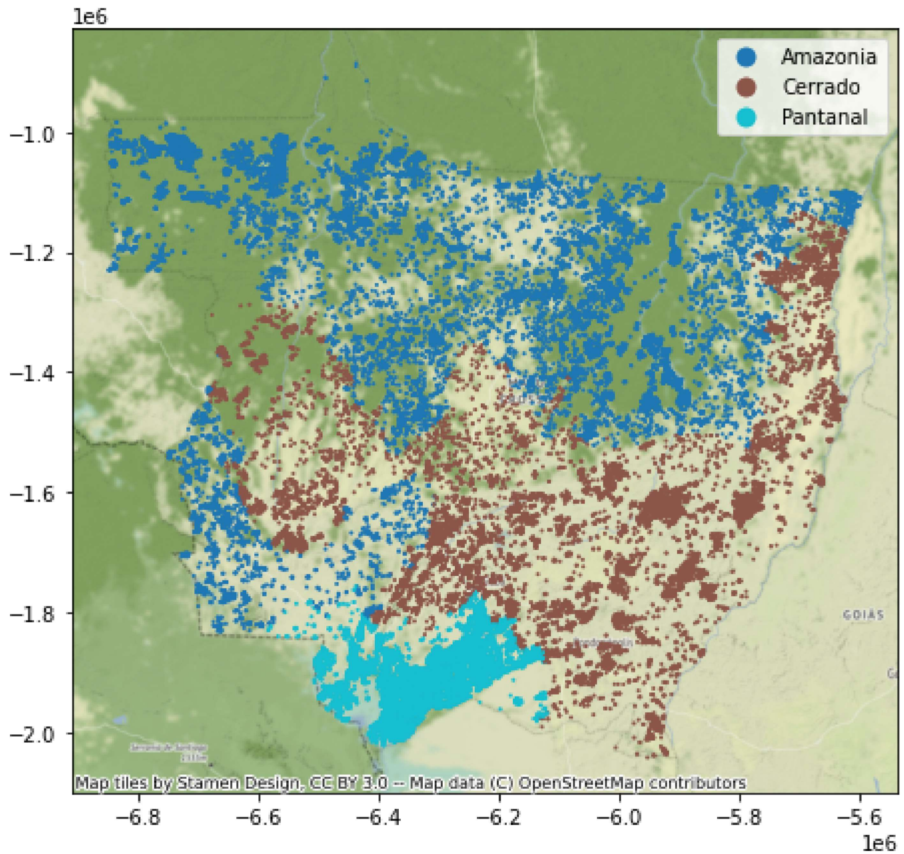
Mato Grosso, has 32.435 (75%) critical registers, and all of them are divided almost equaly between three different biomes, two of them were analysed above and proven to be the worst biomes in terms of numbers of wildfires in 2020:

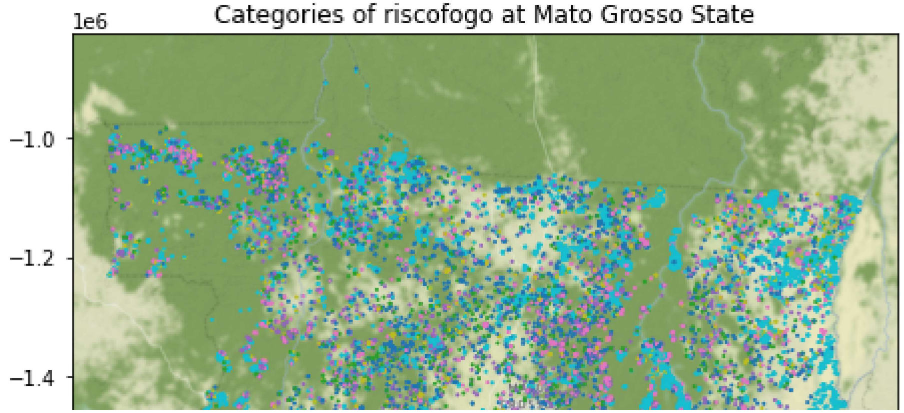
- Amazonia - 11.117 critical
- Pantanal - 10.516 critical
- Cerrado - 10.742 critical

Now let's take a look at the map to undestand how the distribution of wildfires are in Mato Grosso:

```
ax1 = mato_grosso_state.plot(figsize=(12, 7), column='bioma', markersize=1, legend=True)
ax2 = mato_grosso_state.plot(figsize=(12, 7), column='riscofogo', markersize=1, legend=True)
contextily.add_basemap(ax1)
contextily.add_basemap(ax2)
ax.set_axis_off()

plt.title('Categories of riscofogo at Mato Grosso State')
plt.show()
```





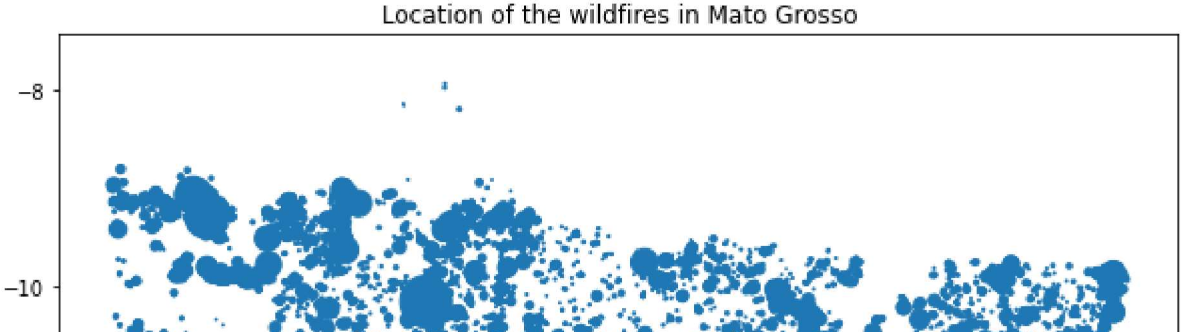

```
#remove zero values
mask1 = mato_grosso_state['longitude'] != 0
mask2 = mato_grosso_state['latitude'] != 0
```

```
x = mato_grosso_state[mask1&mask2]['longitude']
y = mato_grosso_state[mask1&mask2]['latitude']
z = mato_grosso_state[mask1&mask2]['frp']

#use the scatter function
plt.figure(figsize=(10,10))
plt.title('Location of the wildfires in Mato Grosso')
plt.xlabel('Longitude')
plt.ylabel('Latitude')
ax = plt.scatter(x, y, s=z/10, alpha=1)

plt.show
```

```
<function matplotlib.pyplot.show(close=None, block=None)>
```



Most of the fires are well distributed arround all the state, and the intense ones are in the southern part of the state in the Pantanal biome.

Now lets take a look at the city with the higher number of wildfires in Mato Grosso and find out in what biome it is contained

```
mato_grosso_state['municipio'].value_counts()
```

```
POCONE                       5641
BARAO DE MELGACO             3737
CACERES                      2755
COLNIZA                      1873
SANTO ANTONIO DO LEVERGER    1397
                             ...
ARENAPOLIS                     10
SAO JOSE DOS QUATRO MARCOS      9
RIO BRANCO                      8
INDIAVAI                        6
FIGUEIROPOLIS D'OESTE           2
Name: municipio, Length: 140, dtype: int64
```

Poconé

Longitude

Poconé is a small city in the state of Mato Grosso, and it is located close to the Bolivian boarder right by the Mato Grosso do Sul boarder as well. It had 5.259 registered wildfires in 2020

```
pocone_municipio = mato_grosso_state.query('municipio == "POCONE"')
pocone_municipio.head()
```

	estado	municipio	bioma	diasemchuva	precipitacao	riscofogo	latitude	longitude	frp	geometry	time	yea
1161	MATO GROSSO	POCONE	Pantanal	17.0	0.0	critical	-16.987	-57.282	48.0	POINT (-6376603.072 -1919311.816)	18:00:00	202
1162	MATO GROSSO	POCONE	Pantanal	27.0	0.0	critical	-16.991	-57.305	41.5	POINT (-6379163.420 -1919777.413)	18:00:00	202
1164	MATO GROSSO	POCONE	Pantanal	24.0	0.0	high	-16.997	-57.299	71.3	POINT (-6378495.503 -1920475.826)	18:00:00	202
1895	MATO GROSSO	POCONE	Pantanal	20.0	0.0	high	-16.996	-56.609	71.5	POINT (-6301685.054 -1920359.422)	17:05:00	202

Clique duas vezes (ou pressione "Enter") para editar

```
pocone_municipio.municipio.value_counts()
```

```
POCONE    5641
Name: municipio, dtype: int64
```

```
pocone_municipio['bioma'].value_counts()
```

```
Pantanal    5572
Cerrado       69
Name: bioma, dtype: int64
```

```
pocone_municipio_pantanal = pocone_municipio.query('bioma == "Pantanal"')
pocone_municipio_pantanal['riscofogo'].value_counts()
```

```
critical         3922
```

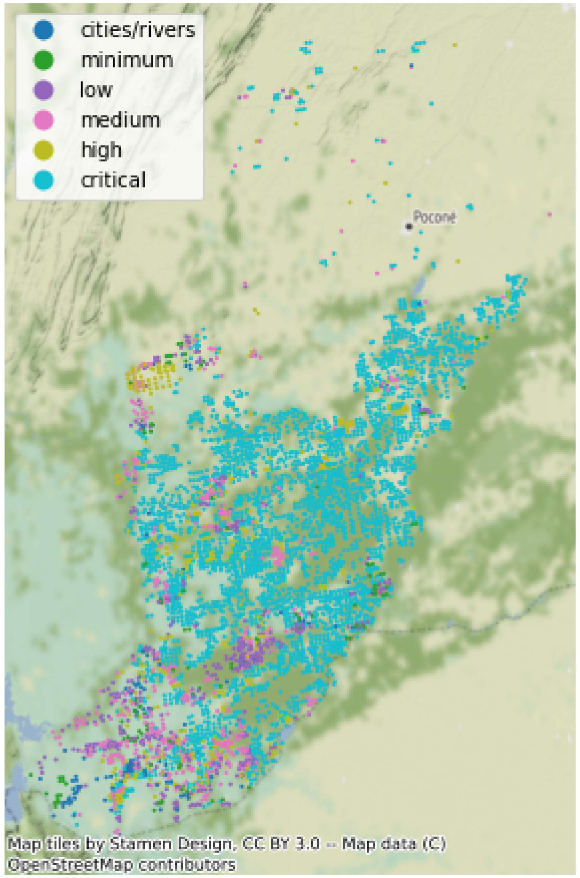
```
medium            491
high              487
low               435
minimum           125
cities/rivers     112
Name: riscofogo, dtype: int64
```

Now it is possible to say that 41% of all Pantanal critical wildfires inside Mato Grosso state. And 37% of the critical wildfires in Mato Grosso happen in the city of Poconé

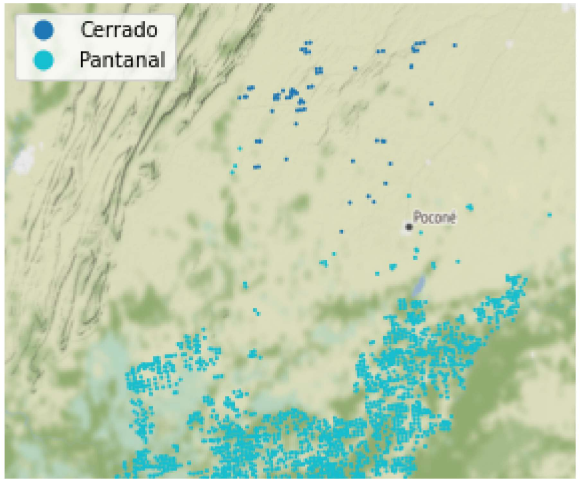
Pará

```
ax1 = pocone_municipio.plot(figsize=(12, 8), column='riscofogo', markersize=1, legend=True)
ax2 = pocone_municipio.plot(figsize=(12, 8), column='bioma', markersize=1, legend=True)
contextily.add_basemap(ax1)
contextily.add_basemap(ax2)
ax1.set_axis_off()
ax2.set_axis_off()
```

```
plt.title('Categories of riscofogo at Poconé city')
plt.show()
```



Categories of riscofogo at Poconé city



```
para_state = wildfire_brazil_2020.query('estado == "PARA"')
para_state.head()
```

	estado	municipio	bioma	diasemchuva	precipitacao	riscofogo	latitude	longitude	frp	geometry	time	yea
4	PARA	OBIDOS	Amazonia	0.0	0.4	low	-1.823	-55.207	18.7	POINT (-6145615.128 -202969.680)	16:50:00	202
5	PARA	BARCARENA	Amazonia	1.0	2.8	cities/rivers	-1.605	-48.674	10.1	POINT (-5418364.895 -178691.154)	16:50:00	202
7	PARA	NOVO PROGRESSO	Amazonia	36.0	0.0	medium	-7.681	-56.080	172.6	POINT (-6242797.044 -857617.685)	16:50:00	202
9	PARA	NOVO PROGRESSO	Amazonia	36.0	0.0	medium	-7.678	-56.054	220.4	POINT (-6239902.737 -857280.704)	16:50:00	202

```
para_state['estado'].value_counts()
```

```
PARA    38603
Name: estado, dtype: int64
```

```
para_state['riscofogo'].value_counts()
```

```
critical         11901
medium            7142
high              6798
low               5656
cities/rivers     4287
minimum           2819
Name: riscofogo, dtype: int64
```

```
para_state['bioma'].value_counts()
```

```
Amazonia    38603
Name: bioma, dtype: int64
```

```
para_amazonia = mato_grosso_state.query('bioma == "Amazonia"')
para_amazonia['riscofogo'].value_counts()
```

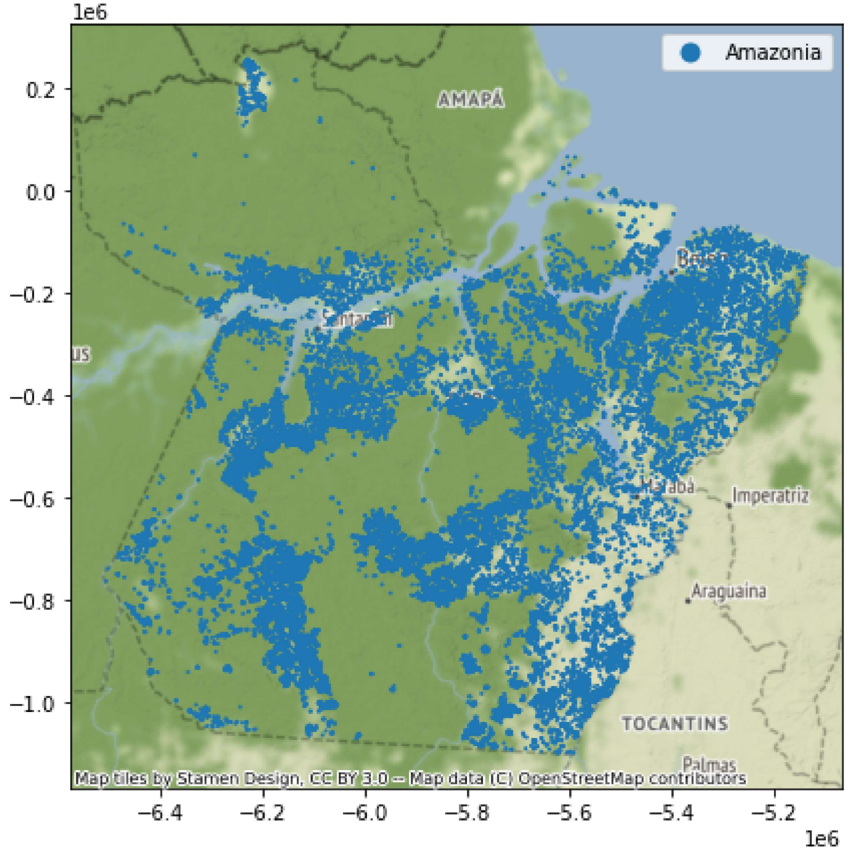
```
critical         11177
cities/rivers     2448
medium            2087
low               2048
high              1527
minimum           1361
Name: riscofogo, dtype: int64
```

Pará, has 11.891 (11%) critical registers, and all of them are located at one bioime only, Amazonia.

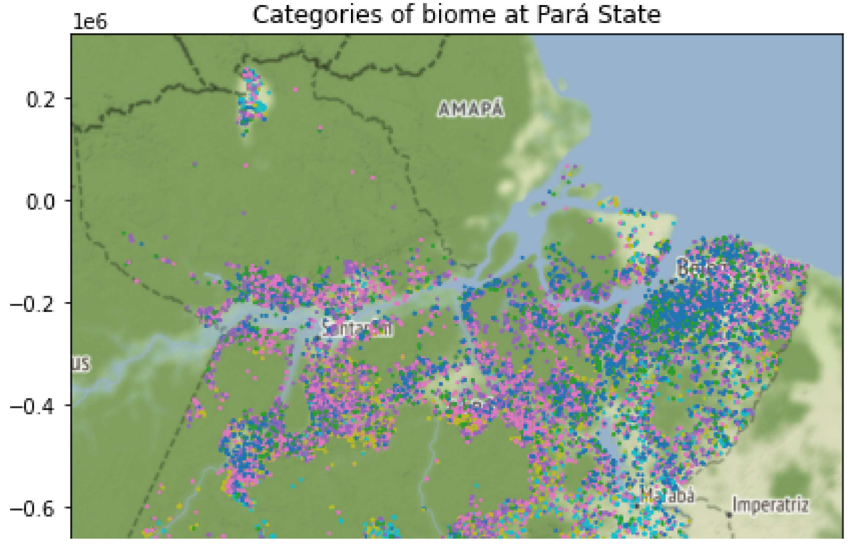
Now let's take a look at the map to undestand how the distribution of wildfires are in Pará:

```
ax1 = para_state.plot(figsize=(12, 7), column='bioma', markersize=1, legend=True)
ax2 = para_state.plot(figsize=(12, 7), column='riscofogo', markersize=1, legend=True)
contextily.add_basemap(ax1)
contextily.add_basemap(ax2)

plt.title('Categories of biome at Pará State')
plt.show()
```



Categories of biome at Pará State



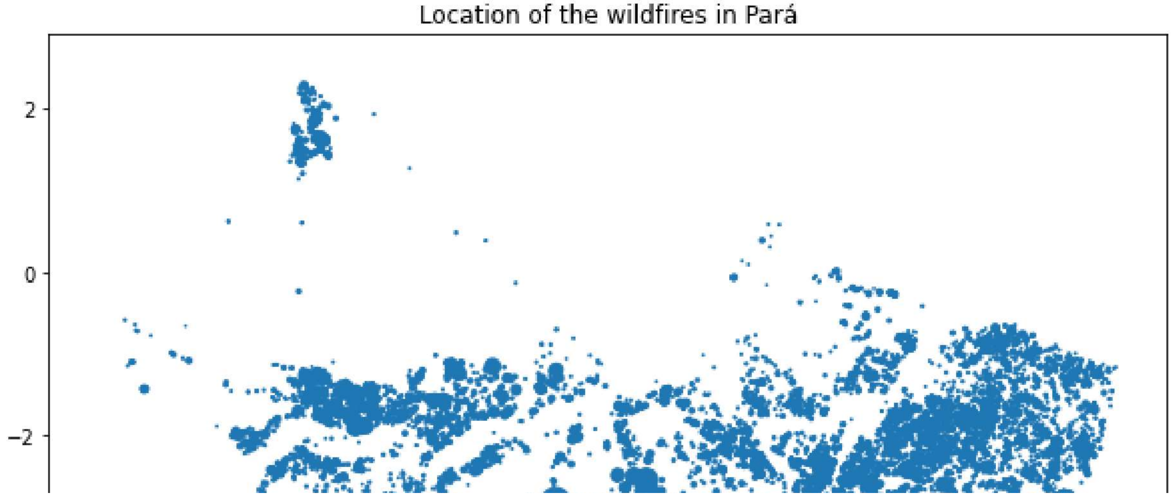
```
#remove zero values
mask1 = para_state['longitude'] != 0
mask2 = para_state['latitude'] != 0

x = para_state[mask1&mask2]['longitude']
y = para_state[mask1&mask2]['latitude']
z = para_state[mask1&mask2]['frp']

#use the scatter function
plt.figure(figsize=(10,10))
plt.title('Location of the wildfires in Pará')
plt.xlabel('Longitude')
plt.ylabel('Latitude')
ax = plt.scatter(x, y, s=z/10, alpha=1)

plt.show
```

```
<function matplotlib.pyplot.show(close=None, block=None)>
```



Most of the critical fires are located in the southern part of the state

Now lets take a look at the city with the higher number of wildfires in Pará and find out in what biome it contains

```
para_state['municipio'].value_counts()
```

```
SAO FELIX DO XINGU       5290
ALTAMIRA                 4865
NOVO PROGRESSO           2376
ITAITUBA                 1799
PORTEL                   1162
                         ...
BELEM                       2
ANANINDEUA                  1
SANTO ANTONIO DO TAUA       1
SANTA BARBARA DO PARA       1
BENEVIDES                   1
Name: municipio, Length: 143, dtype: int64
```

São Félix do Xingú

São Félix do Xingú is a small city in the state of Pará. It had 5.285 registered wildfires in 2020

```
sfx_municipio = para_state.query('municipio == "SAO FELIX DO XINGU"')
sfx_municipio.head()
```

	estado	municipio	bioma	diasemchuva	precipitacao	riscofogo	latitude	longitude	frp	geometry	time	year
279	PARA	SAO FELIX DO XINGU	Amazonia	32.0	0.0	critical	-5.514	-51.062	102.3	POINT (-5684195.839 -614765.363)	17:35:00	2020
400	PARA	SAO FELIX DO XINGU	Amazonia	33.0	0.0	medium	-5.499	-51.064	73.0	POINT (-5684418.478 -613087.829)	17:35:00	2020
401	PARA	SAO FELIX DO XINGU	Amazonia	32.0	0.0	critical	-5.512	-51.055	282.9	POINT (-5683416.602 -614541.689)	17:35:00	2020
402	PARA	SAO FELIX DO XINGU	Amazonia	32.0	0.1	medium	-5.510	-51.034	237.0	POINT (-5681078.893 -614318.016)	17:35:00	2020

```
sfx_municipio.municipio.value_counts()
```

```
SAO FELIX DO XINGU    5290
Name: municipio, dtype: int64
```

```
sfx_municipio['bioma'].value_counts()
```

```
Amazonia    5290
Name: bioma, dtype: int64
```

```
sfx_municipio_amazonia = sfx_municipio.query('bioma == "Amazonia"')
sfx_municipio_amazonia['riscofogo'].value_counts()
```

```
critical        2558
```

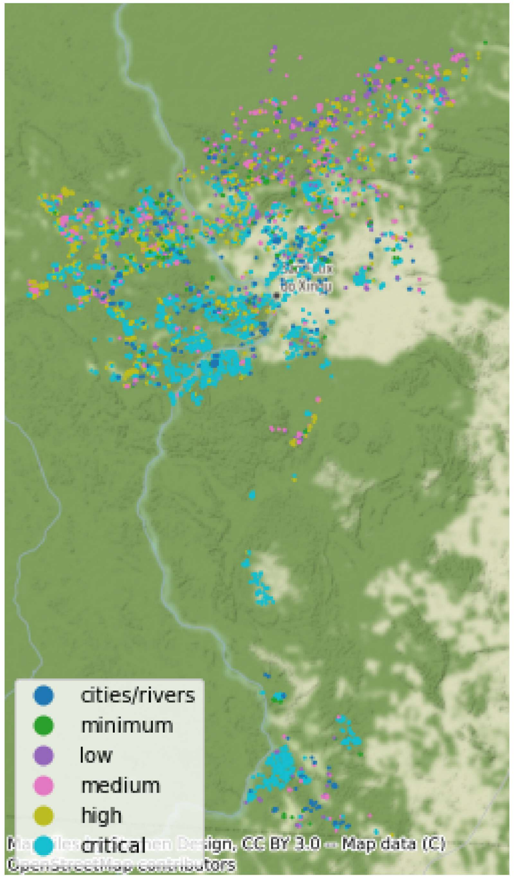
```
high             1378
medium            601
low               322
cities/rivers     290
minimum           141
Name: riscofogo, dtype: int64
```

Now it is possible to say that 23% of all Amazonia critical wildfires inside Pará state happened in the city of São Felix do Xingú. And 21% of the critical wildfires in Mato Grosso happen in the city of Poconé

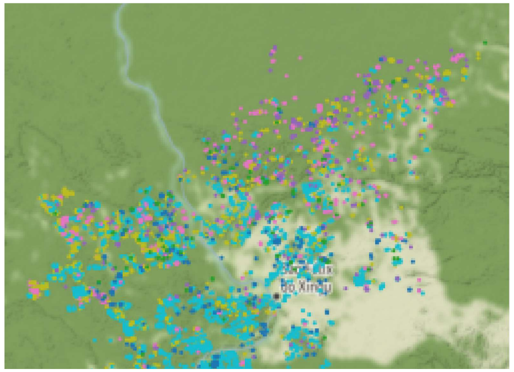
```
ax1 = sfx_municipio.plot(figsize=(12, 8), column='riscofogo', markersize=1, legend=True)
ax2 = sfx_municipio.plot(figsize=(12, 8), column='riscofogo', markersize=1, legend=True)
contextily.add_basemap(ax1)
contextily.add_basemap(ax2)
ax1.set_axis_off()
ax2.set_axis_off()



plt.title('Categories of riscofogo at São Félix do Xingú city')
plt.show()
```



Categories of riscofogo at São Félix do Xingú city



Key Findings and Insights

We were able to map all of the wildfires that happened in 2020 in Brazil and come to the final conclusion:

In the year of 2020, there were 222.797 different wildfires registered in the whole national territory of Brazil. This means that we have approximately 610 wildfires per day. Inside the 222.797 fires registered, more then half (52%) are categorized as critical by the parameter RisgoFogo that predicts whether it is going to break out a fire in that especific place or not in the day that the fire happened.

After looking at the dataset as a whole, we started to analyse it from the perspective of the biomes in Brazil.

We discovered that 47% of the whole country 's wildfires are located in Amazonia, but only 27% of critical ones, meaning that probably, most fires in that biome are man made. The worse state inside that biome is Pará, with 38.603 wildfires inside this biome.

Cerrado is another Brazilian biome that had inumerous wildfires during the year of 2020. only 28% of Brazil 's wildfire are located in this biome. 45% of all critical locations in Brazil are inside this biome. The worst state in this biome is Mato Grosso, with 10.742 critical cases

After reviewing the critical biomes, we started to analyse the worst two states in Brazil when we are talking about wildfires, Mato Grosso and Pará

Mato Grosso had 47.708 wildfires, 30% of all critical wildfires in Brazil, and inside the state, the numbers are almost divides equaly between 3 biomes, Amazonia(20.648 wildfires, 11.177 critical), Pantanal(13.207 wildfires, 10.516 critical) and Cerrado(13.853 wildfires, 10.742 critical).

The worse city in Mato Grosso is Poconé, with 5.572 cases, and located 100% inside the Pantanal biome

Pará, in other hand, had 38.603 wildfires in the year of 2020, making 36% of the whole country. Amazonia is the only biome in Pará, with 11.117 critical locations

We also looked at the worse city in Pará, São Felix do Xingú, with just 5.290 wildfires, and 2.558 critical ones inside Amazonia.

Therefore, we can assume that in the hottest places in Brazil, located at the north of the country, have more chances of breaking a wildfire. We can say that the two most sensible biomes in Brazil are Amazonia and Cerrado, the two states are Mato Grosso and Pará, and the respectice cities are Paconé and São Félix do Xingú.