Project Report

GitHub URL
https://github.com/EoinDunne98/UCDPA_EoinDunne

Abstract

Analysis of the degradation of Brazilian rainforests throughout the years of 1990 – 2020.

We will look that the causes of the degradation in these beautiful rainforests, and the regression throughout the past 3 decades.

Introduction
Brazil is home to the Amazon Rainforest. A moist tropical rainforest that covers an area of 5.5m km2. Comprising roughly 40% of Brazil's total land area. (***The Editors Encyclopaedia Brittanica [1])***

The Amazon Rainforest plays a key part in trapping billions of tonnes of CO2 from our atmosphere each year. It also benefits local indigenous communities, as well homing 2.5M insect species, and thousands of species of plant, fishes, birds, mammals, amphibians, and reptiles.

I chose this project as, unfortunately, mainly due to human intervention, we see the impacts of deforestation, and forest fires. Deforestation mainly occurs due to cattle ranching, mining, logging, and climate change.

In this project, I will mainly focus on the degradation impact of forest fires in the Brazilian Rainforest. I will analyse how it impacts deforestation levels, and how climate change can have an impact of the rate of forest fires.

Dataset
Brazil Forest Fires 1999 – 2019
https://queimadas.dgi.inpe.br/queimadas/bdqueimadas/

- This dataset pulls a list of all reported forest fire outbreaks in Amazonia.
- This dataset covers the period of 1999 to 2019.
- I chose this dataset as it contains hundreds of recordings of accurate data, that is backed by the Brazilian Government (***INPE - National Institute for Space Research, 2020***). ***[2]***

Brazil Deforested Areas 2004 – 2019
http://www.obt.inpe.br/OBT/assuntos/programas/amazonia/prodes

- This dataset pulls a list of deforested areas by km2 in the Brazilian Rainforest by state and year.
- This dataset covers the period of 2004 to 2019.
- I chose this dataset as it is an incredibly reliable source, obtained by the PRODES project. The PRODES project performs satellite monitories of deforestation in the Amazon Rainforest since 1988. This data is used by the Brazilian Government to enact public policies. ***(Data from the PRODES Deforestation Monitoring System in the Legal Amazon) [3]***

Brazil Precipitation 1990 – 2020
https://climateknowledgeportal.worldbank.org/country/brazil/climate-data-historical

- This dataset pulls a dataset of Brazil's climate by mean precipitation levels per month.
- The dataset covers the period of 1990 – 2020.
- I chose this dataset as the data comes from a reputable source and is historically recorded data from the Climatic Research Unit (CRC) of University of East Anglia. [4]

Implementation Process

1 Importing, Cleaning, and Understanding the Data

1.1 Importing reported fires in Brazil Amazon Regions

Using pandas, I read the CSV file "**Brazil_Fires_1999-2019.csv**" and assign it to the data frame "**brazilfires**". Then I analyse the head, shape, and description of the dataset to understand the information.

1.1.1 Cleansing Brazil Fires File

I begin cleansing the files by dropping duplicates, sorting by year and month, and checking for missing data.

I use the pandas ".fillna" function to fill the missing data with the mean of "firespots".

Since I had sorted by year and month using the ".sort_values" function, the index is now out of order. I amend this by using the "reset_index" function. I drop the index column using "drop = True".

1.2 Importing Deforested Regions in Brazil

Using pandas, I read the CSV file "def_area_2004_2019.csv" and assign it to the data frame "pivoted_deforest_brazil". I use shape and head functions to understand the data

1.2.1 Cleansing the Data (Deforested Regions Brazil)

I use the drop function to remove the "AMZ LEGAL" column.

I use the ".drop_duplicates" function to clear any duplicate reports from my dataset.

Using ".sort_values" function, I sort by "Ano/Estados". This is the year column and named in Portuguese since I pulled this from a Brazilian source. I check if there are any duplicates in my dataset using the ".isnull() and .sum()" function. There is one missing data in the "AP" column. I use the ".fillna" and ".mean()" functions to fill the missing data with the mean integer of the "AP" column.

To match the state names to the "brazilfires" data frame, I rename my columns accordingly. I do this my first indicating what the columns will be, then assigning them to my dataset using the ".columns" function on "pivoted_deforest_brazil".

Finally, I reset the index and unpivot my data frame using ".reset_index" and ".melt" functions. This makes my dataset easier to read in seaborn and matplotlib.

I assign this to the unpivoted data frame, "**deforest_brazil**".

1.3 Merging the datasets

Using the pandas function ".merge()", I merge the two datasets that I have created ("**brazilfires**" and "**deforest_brazil**"). This will make it easier for seaborn and matplotlib to read my data.

2 Analysing our Data

2.1 Analysis of Brazilian Fires

2.1.1 Per Year

Using matplotlib (imported as plt) and seaborn (imported as sns), I firstly visualise the number of fire spottings per year in the Brazilian rainforest.

I set the figure size and the range (assigning to a year list). Then I use seaborn to generate a line plot from "brazilfires. Using matplotlib, I plot the title, X and Y labels, and the ticks (from the year range that I assigned earlier).

2.1.2 Per Month

Currently, the month dataset is assigned to integers. 0 being January, 1 being February, and so on.

I first need to assign the month integers to strings. I create a list for the month label. I set the figure size and use seaborn to create a bar plot. I then set the title, and labels of the bar plot.

Using the list for the month labels that I created earlier, I set the X ticks. The ticks must read only unique values, as there are over 2,000 entries, but there are only 12 unique values, which matches the entries on my list. I use the ".unique()" function to achieve this.

2.1.3 Per State

Finally, I will visualise the states where the fires occur most frequently. I set the figure size, use seaborn to create a bar plot. I then set the title and labels.

Unfortunately, this bar plot is not ordered. Thankfully, I can use ".groupby" pandas functions to fix this. I create a new list called "sorted_states", which orders the states by number of fire spottings (highest to lowest).

When re-creating my bar plot in seaborn, I use the "order" command and assign it to the list I just created ("sorted_states").

2.2 Analysis of Brazilian Deforestation

2.2.1 Per Year Analysis

As this dataset covers a shorter period, I set the range from 2004 to 2020.

Using seaborn, I create a scatterplot from the "**degradation_brazil**" dataset. I use the hue command to visualise the states where the deforestation was recorded.

2.2.2 Per State and Year Analysis

Similar to the above visualization, I create a bar plot instead. This time visualizing the year as the hue.

2.2.3 Per State and Year Analysis

Using ax subplots, I visualise the area of deforestation per year by state of occurrence. Ensuring that the X-axis is shared (sharex = True).

I use seaborn to visualise line plots, and python logical conditions to ensure that the Y labels are not duplicated for each line plot.

3 Comparison

3.1 Fire spottings vs Deforestation

The most effective way to visualise the impact that forest fires have on deforestation levels, is to overlay the graphics. I achieve this by using subplots in matplotlib and seaborn.

Firstly, I create a bar subplot in matplotlib. To indicate on the figure which is the deforested area, I set a label and a legend.

Secondly, I create a line plot using seaborn. I use the matplotlib ".twinx()" function to allow the two figures to share the X axis, and the "ax" command to allow the figures to read the function and overlay each other.

3.2 The impact of the Brazilian Dry Season

My third dataset will import file called "brazil_precipitation_1990_2020.csv". Unfortunately, this file isn't very clean.

3.2.1 Cleaning the data using SQL

Importing sqlite3 and csv, I create a database called "brazil_precipitation13.db". This allows me to connect to this database, a create a new table called "brazil_precip8".

I import my CSV file into this table using "INSERT INTO" SQL commands, and CSV reader commands ("csv.reader").

Once the contents have been imported into this new table, I delete unnecessary rows and columns. (DELETE state "name unknown" row) and (ALTER TABLE, DROP "Annual" column).

I export these changes to a new CSV file using CSV commands and save it as "**brazil_precipitation_clean.csv**". I then close the SQL connection.

3.2.2 Mean Precipitation in all of Brazil 1990 – 2020

To visualise the wettest/driest months in Brazil, I just need to look at the first row of my dataset. I place this into a new data frame ("brazil_mean_precip2").

Using seaborn and matplotlib, I create a bar plot to visualise this data.

For my next analysis, I need to clean and transpose this data frame. I create new columns called "Month" and "Precipitation" and reset the index.

3.2.3 Brazilian Dry Seasons vs Number of Fire Spottings

I create another list to indicate the order of months and use matplotlib to first create a bar plot for mean precipitation levels in Brazil.

Using seaborn, I then create a line plot to visualise mean fire spottings in Brazil per month. I don't want to see the confidence interval, so I remove this. I use ax commands to overlay the graphics and share the X-axis.

4 Machine Learning

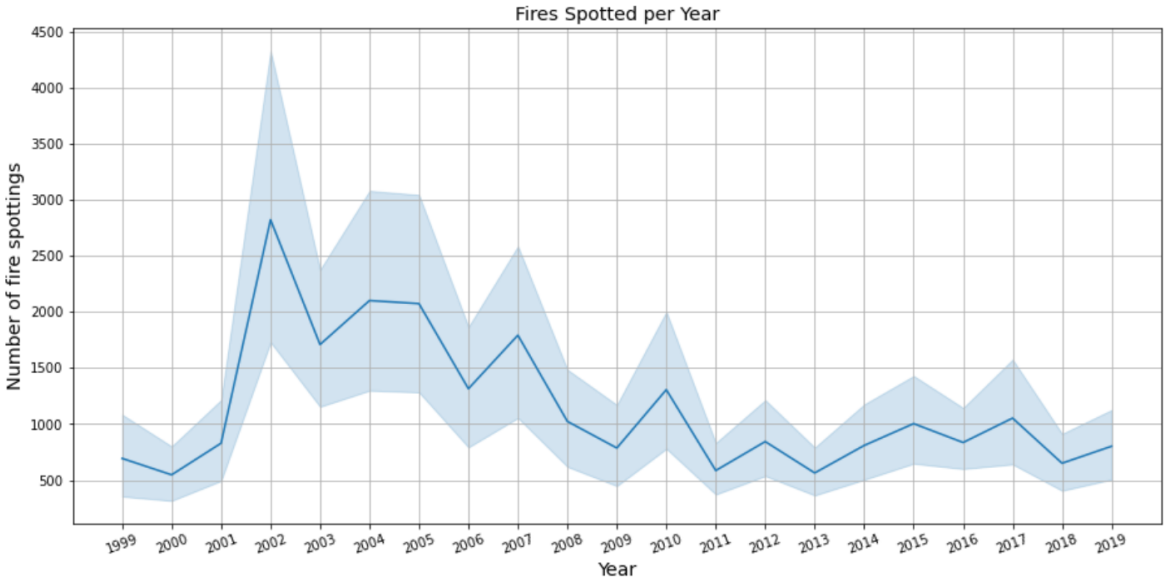
4.1 Deforestation in Brazilian Rainforests

Using machine learning techniques, I analyse the "degradation_brazil" data frame. I wish to just analyse the "Year" and "Deforested Area" so I create a new data frame ("deforest").

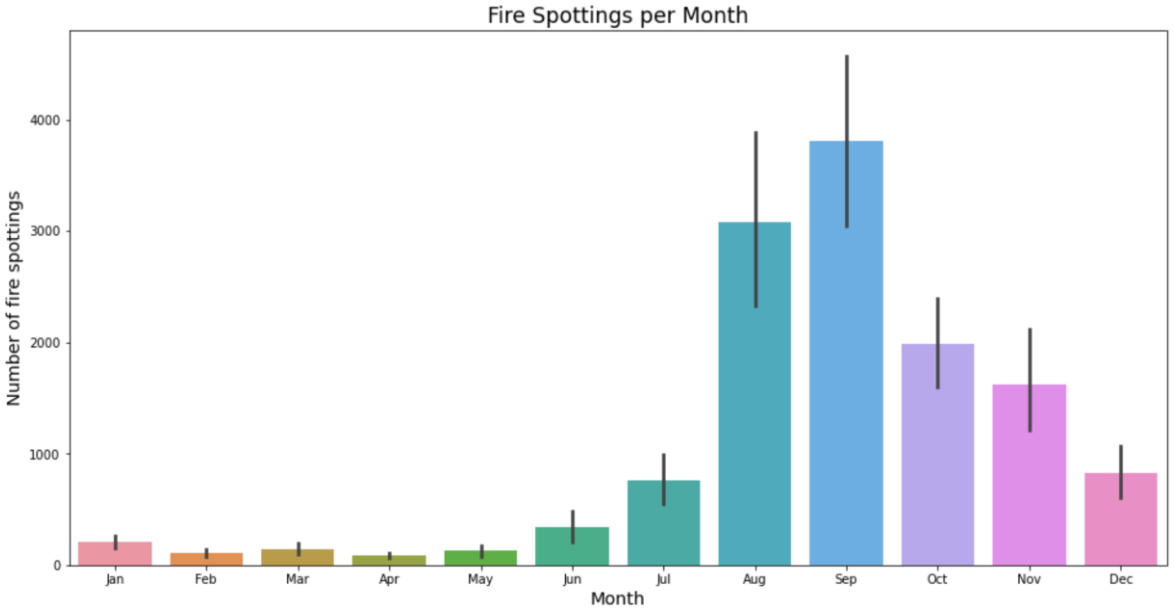
Using Linear Regression commands, I create my model to predict X (deforested_area). I check my prediction score using ".score(X , Y)", and test my prediction of X. I also test the prediction for 2022 using (**.predict([[2022]])**)

Using matplotlib and seaborn, I visualise a regression model by creating a scatter plot of X and Y (Deforestation and Year). I plot a line in my scatter plot to better visualise the regression.

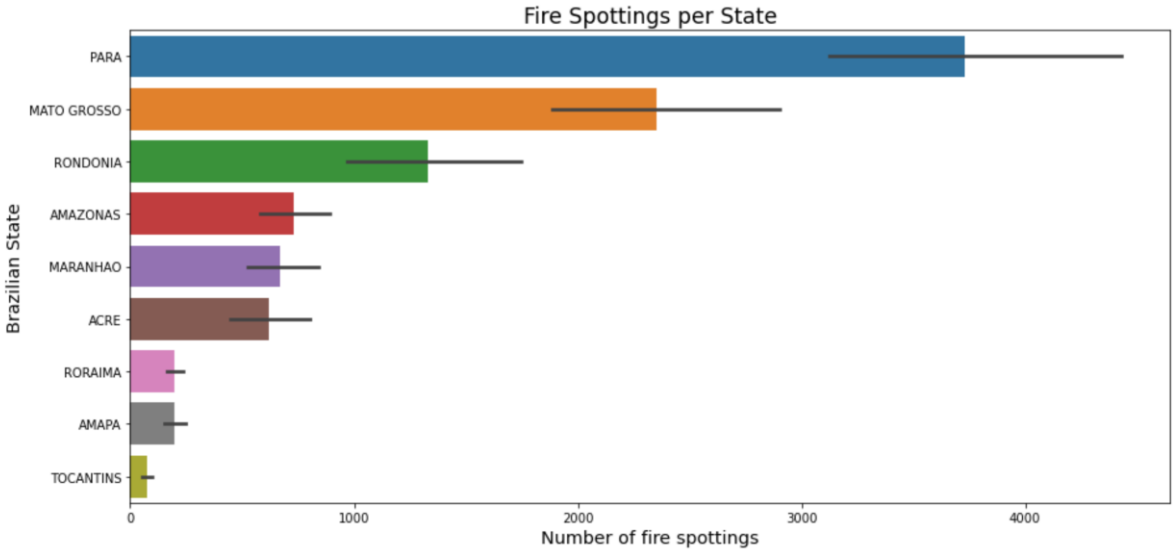
Results



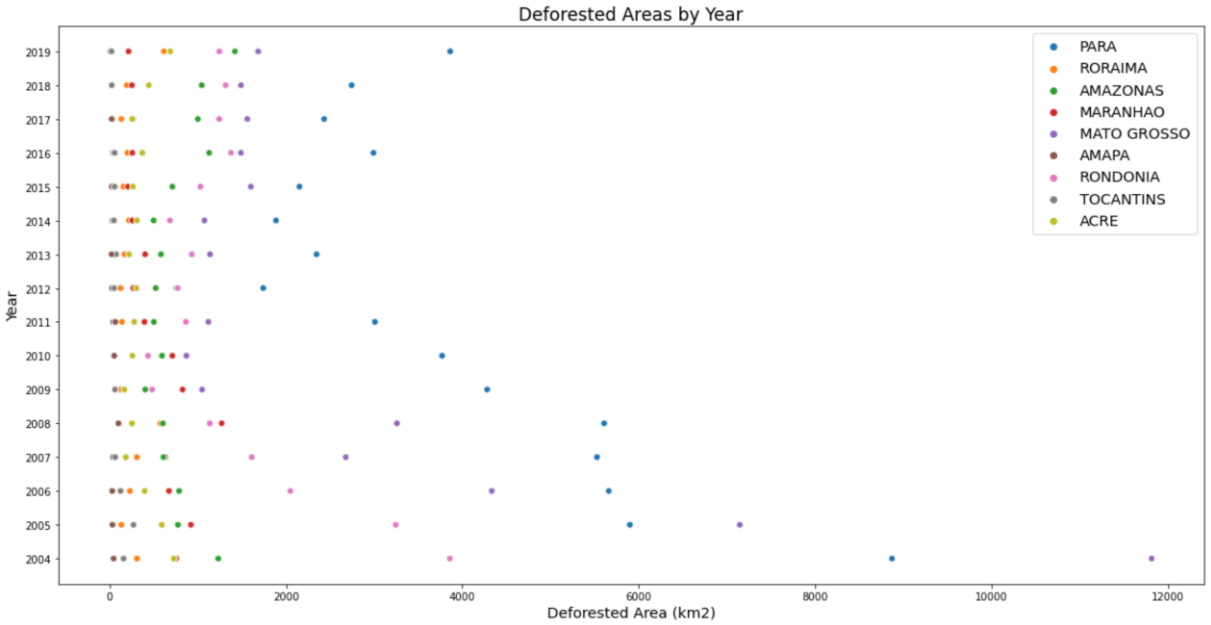
- This line plot visualises the frequency of fire spottings per year in the Brazilian Rainforest. From this graphic, 2002 was the most abundant year for fire spottings (nearly 3,000). The number of recorded fire spottings is reducing each year.



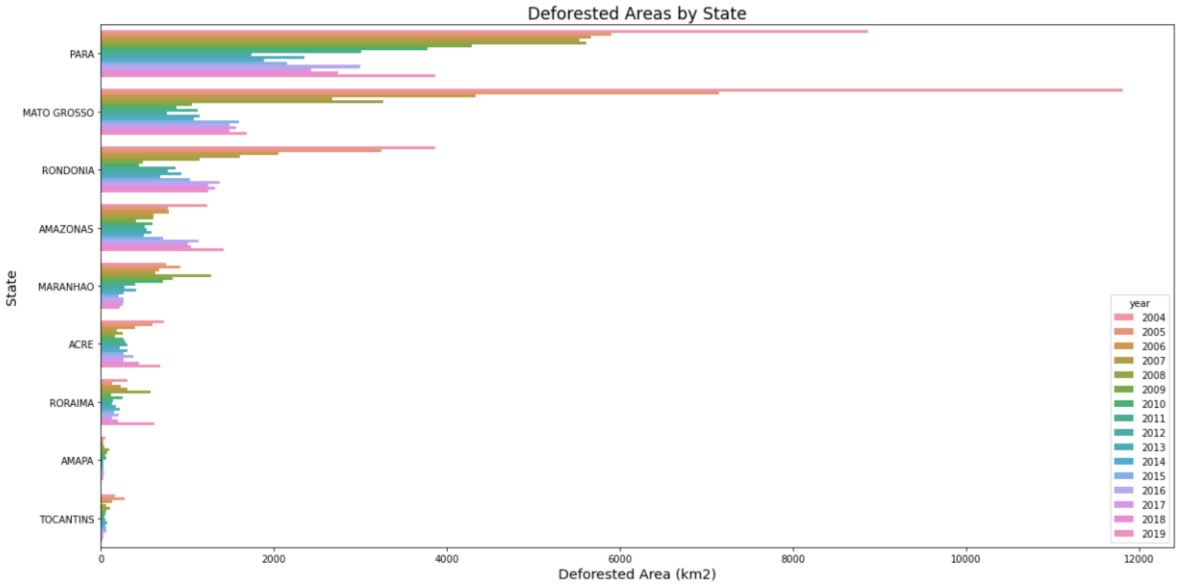
- This bar plot visualises the number of fire spottings per month from 1999 – 2019. From this graphic, we can identify that September and August see the highest number of fire spotting reports. January to July do not see many reports.



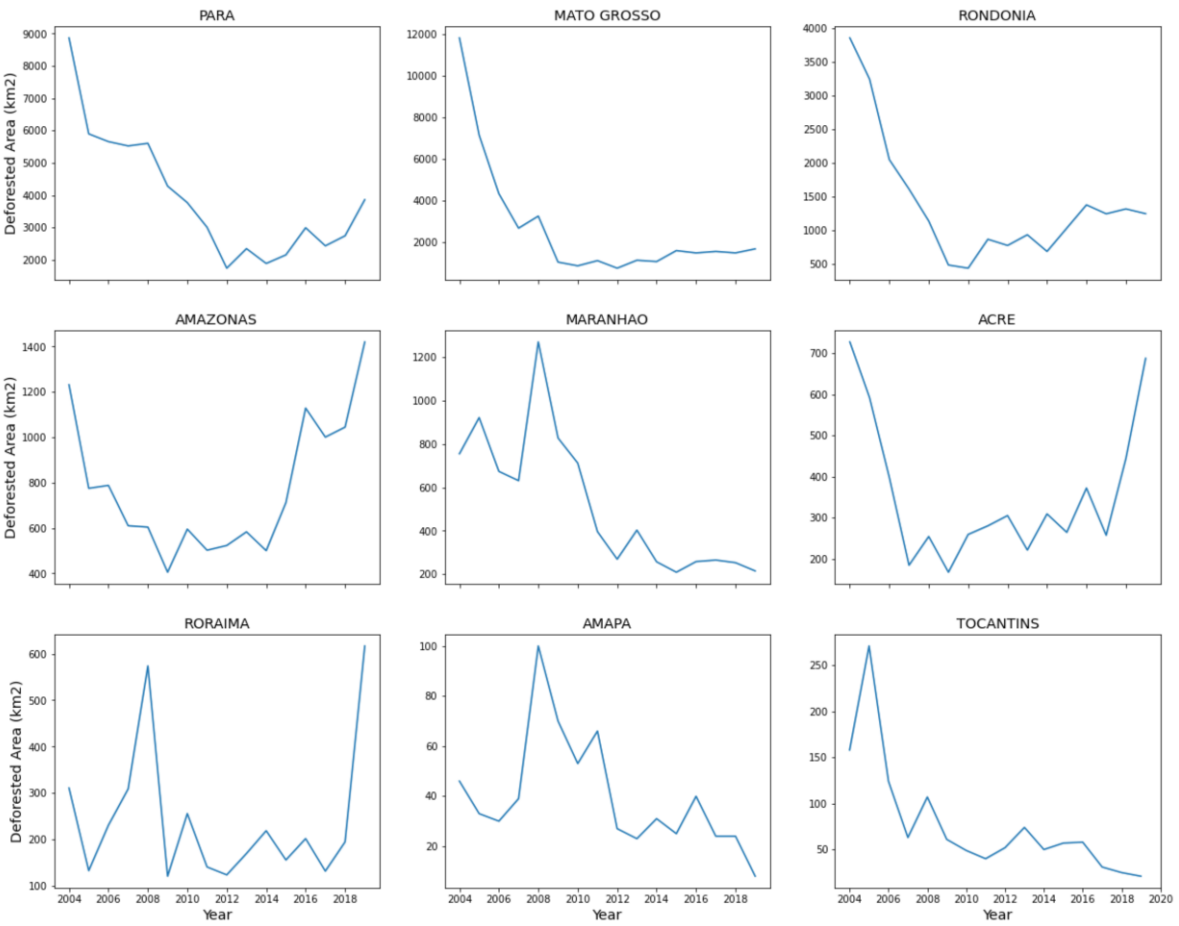
- This bar plot visualises the number of fire spottings per state from 1999 – 2019. The state of “PARA” sees the highest number of reports, whilst the state of “TOCANTINS” sees the fewest.



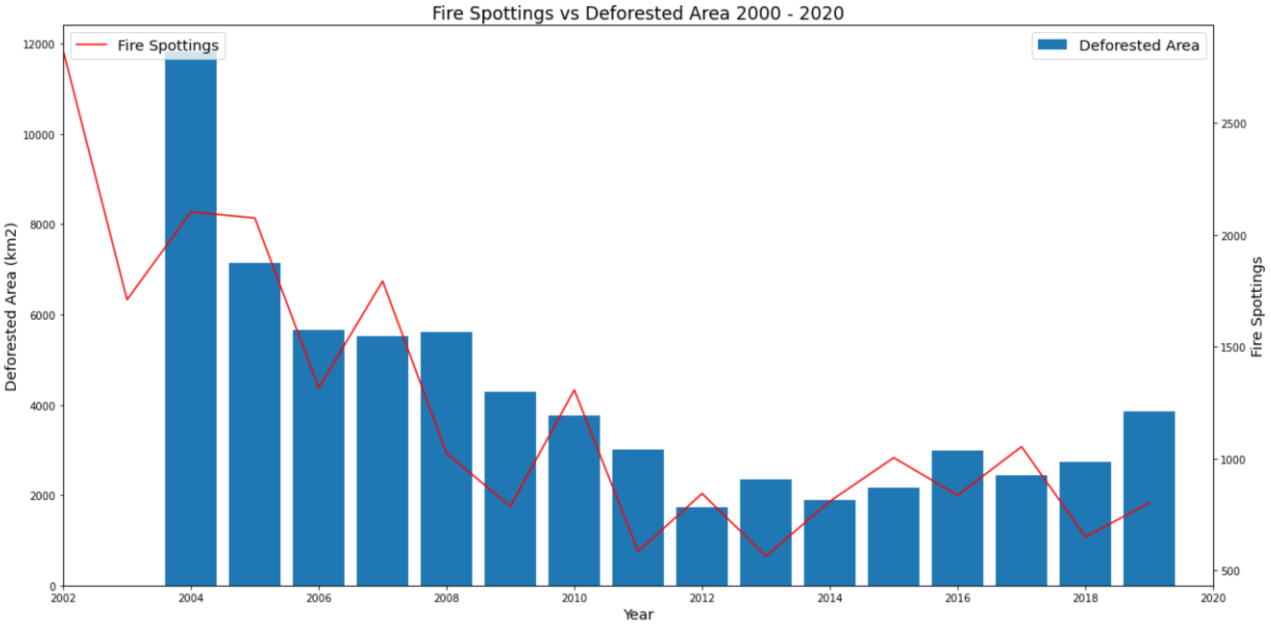
- This scatter plot visualises the deforested area (km2) per year from 2004 – 2019. A hue enables the visualisation by state of data recorded.



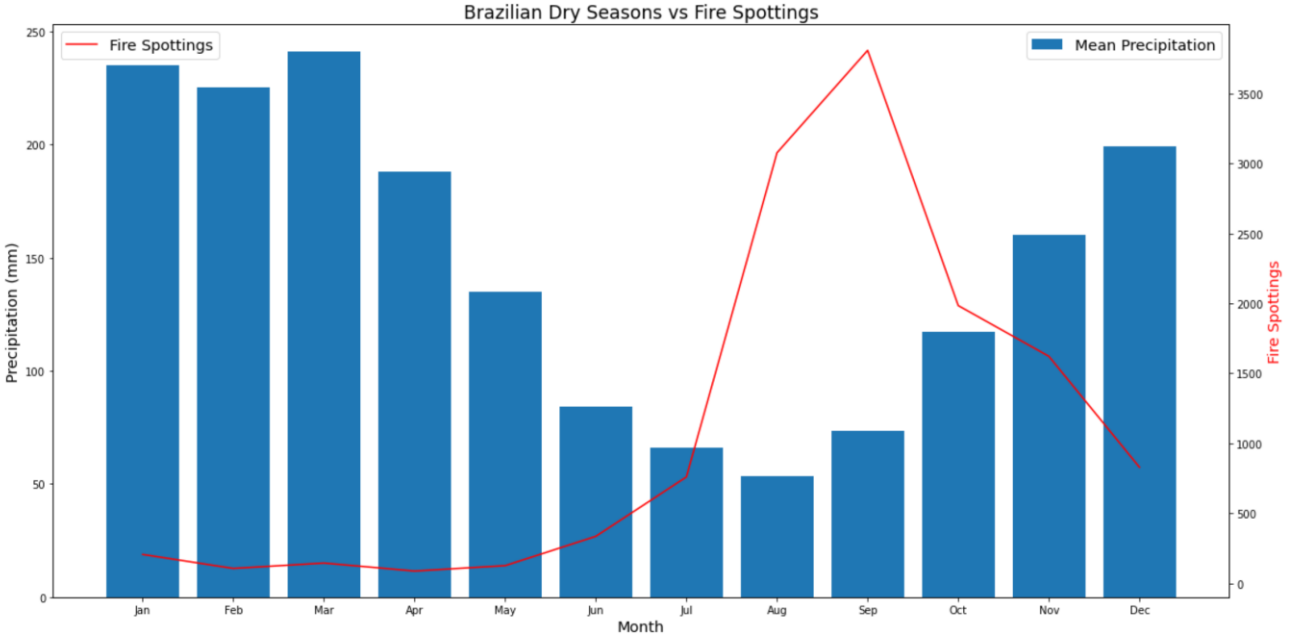
- This bar plot visualises the deforested area (km2) per state from 2004 – 2019. A hue enables the visualization by year of data recorded.



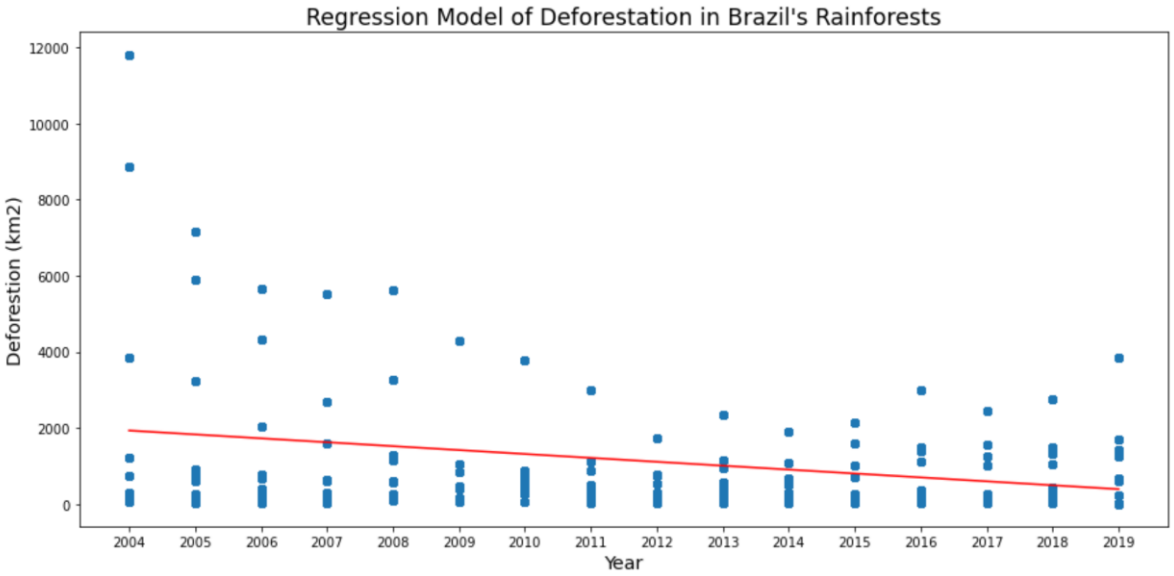
- These subplots visualise the deforested area (km2) per year by state from 2004 – 2019. Each line plot analyses the data recorded in each state. In Amazonas, Acre, and Roraima, we can see that deforestation is increasing at an alarming rate. Whereas in Mato Grosso, Tocantins, Amapa, and Maranhao, deforestation is reducing.



- This line plot and bar chart overlay each other to visualise the frequency of fire spottings against reported deforestation levels. They both share the X-axis (Year) which covers 2002 – 2019. A legend indicates that the blue chart visualises the “Deforested Area” and the red chart visualises fire spottings. (The Y-axis label for Fire Spottings is also in red to better visualise the findings)



- This line plot and bar chart overlay each other to visualise how the dry seasons in Brazil impact the number of fires spotted. The data covers the years from 1990 – 2020 for precipitation levels, and 1999 – 2019 for fire spottings.



- This regression model visualises the trend of deforestation from 2004 – 2019. It is mapped on a scatter plot to visualise the deforested areas (km2) per year and includes a regression line to see the trend.

Insights

1. Fires Spotted Per Year

- The frequency of fire spottings per year in Brazilian rainforests has been reducing since the height of 2002. This is great as this leads to less lost rainforest due to the destruction of forest fires.

2. Deforested Areas by State

- The area of deforested lands is reducing heavily in several states of Brazil. The states of Mato Grosso, Rondonia, Maranhao, Amapa, and Tocatins are seeing massive reductions of deforestation. However, we can see the rise of deforestation in Para, Amazonas, Acre, and Roraima. This is very alarming.

3. Fire Spottings vs Deforested Area 2000 – 2020

- As fewer fires are spotted each year, we can also see that the deforestation levels are reducing. There is a clear correlation between the two. Many Amazonian farmers cause man-made forest fires to clear land for agriculture. As fewer fires are reported, we can also see that deforestation reduces.

4. Brazilian Dry Season vs Fires Spotted

- As Brazil enters the dry season, where precipitation levels are at their lowest, fire spottings increase. When the climate become drier; fires become more prominent as plants are more flammable when dry. Farmers also take advantage of this dry season to create man-made forest fires to clear land for their agriculture.

5. Regression Model of Deforestation in the Brazilian Rainforest

- The degradation of the Brazilian Rainforest reduces each year. According to the regression model, we should expect to see the area of deforestation to reduce. This is incredible news, as the Amazon rainforest is a crucial biome to our planet's survival. However, as we are affected by climate change, the climate will become more unpredictable, and the seasons drier. As we have analysed previously, when the seasons become drier, forest fires become more prominent.

References

[1] https://www.britannica.com/place/Amazon-Rainforest - The Editors Encyclopaedia Brittanica, Oct 2022.

[2] PE - National Institute for Space Research, 2020. Burning and Forest Fire Monitoring Portal. Available at http://www.inpe.br/queimadas , November 2022

[3] http://www.obt.inpe.br/OBT/assuntos/programas/amazonia/prodes, Data from the PRODES Deforestation Monitoring System in the Legal Amazon was licensed under a Creative Commons License - Attribution - ShareIgual 3.0 Not Adapted, November 2022.

[4] https://climateknowledgeportal.worldbank.org/country/brazil/climate-data-historical , Climatic Research Unit (CRU) of University of East Anglia, November 2022.