Forest Fire Exploratory Data Analysis and Prediction

-By Aman Kumar

The Amazon Forest if said to be the life inducer of the Earth won't be wrong. Amazon forests account for 20% oxygen fixing in the atmosphere, further it is known for its medicinal herbs, and amongst those lush green tropical evergreen forests, one may find a new diversity of organisms. Despite being so rich in flora and fauna, Amazon forests still face crises when it comes to forest fires. This issue has been the concern of global delegates.

Through this project I aim to spot the specific months the forest fire hits more profoundly than other months.

Throughout this project I have used python to calibrate, categories and compare the dataset hosted on Kaggle (Which is saved locally) using NumPy and pandas, moreover I have made some significant changes while calibrating data i.e., omitting rows with missing values and replacing invalid data. There were some aspects of the data which had to be replaced for the convenience of the reader because they were in different languages, for example: Month names. For better and optimized visualization, I have used Matplotlib and Seaborn.
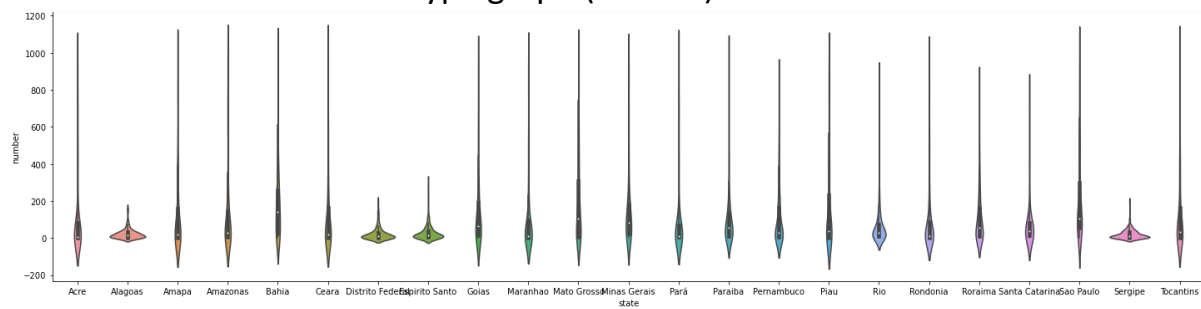
Inference

Worst hit states are:
Mato Grosso
Paraiba
Sao Paulo

As can be seen from violin type graph (Line 15)

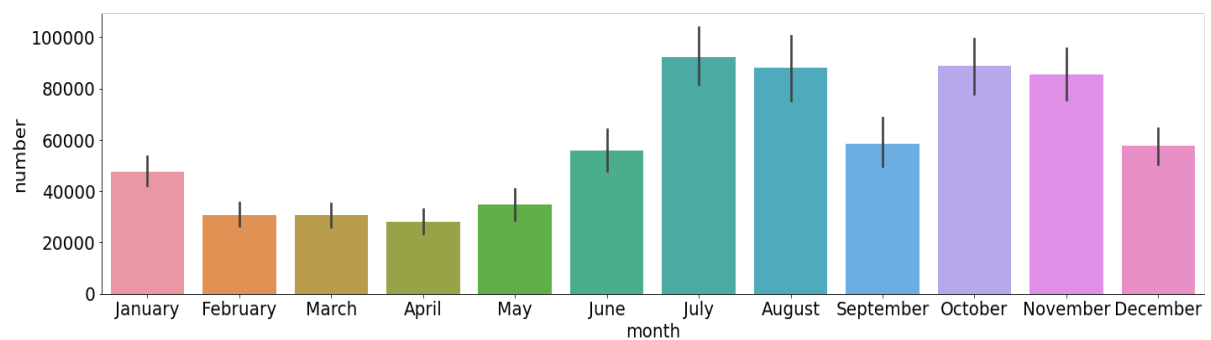Worst hit years are:
2003
2015
2016
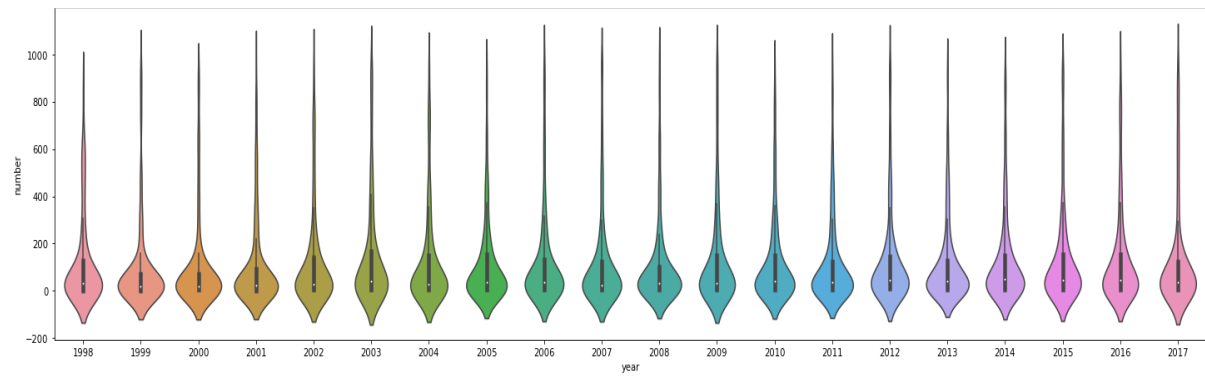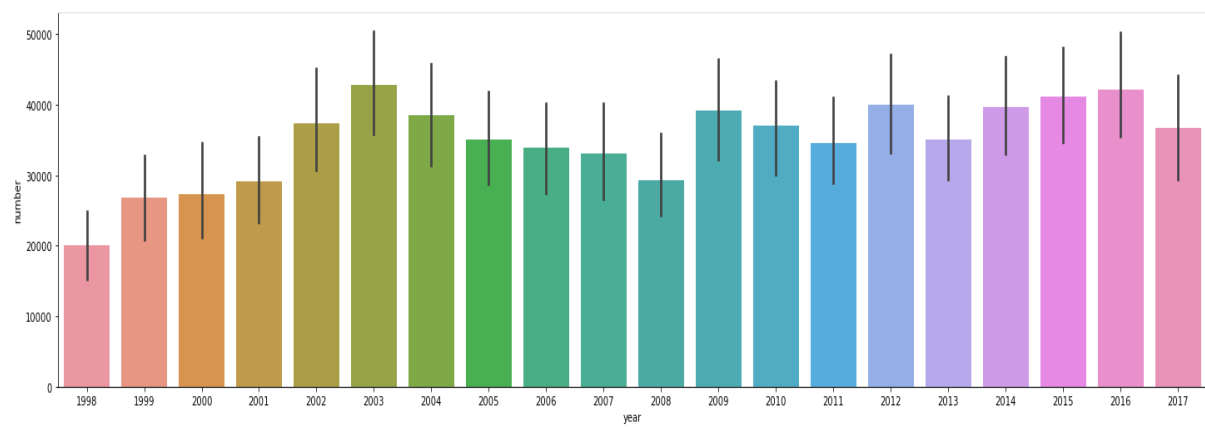
Worst hit months are:
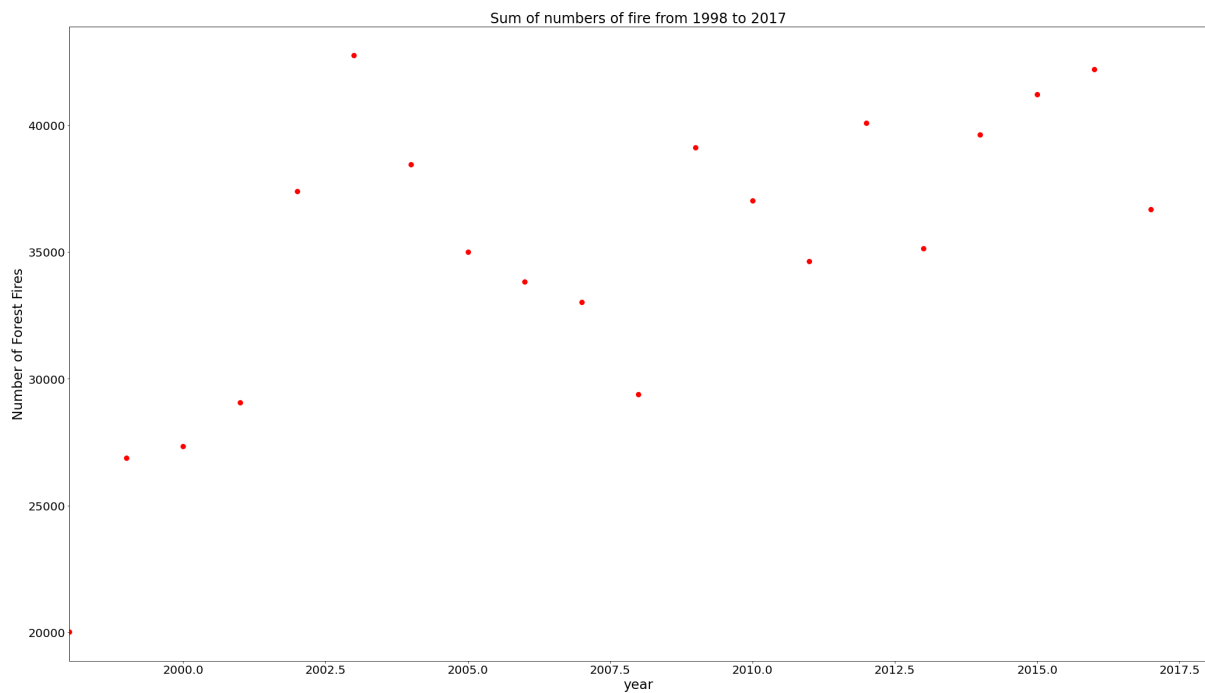August
July
November
October

As can be seen from violin type graph, bar type graph and scatter plot graph.
(Line 17, 18, 19 and 20)

Bar graph for clear analysis
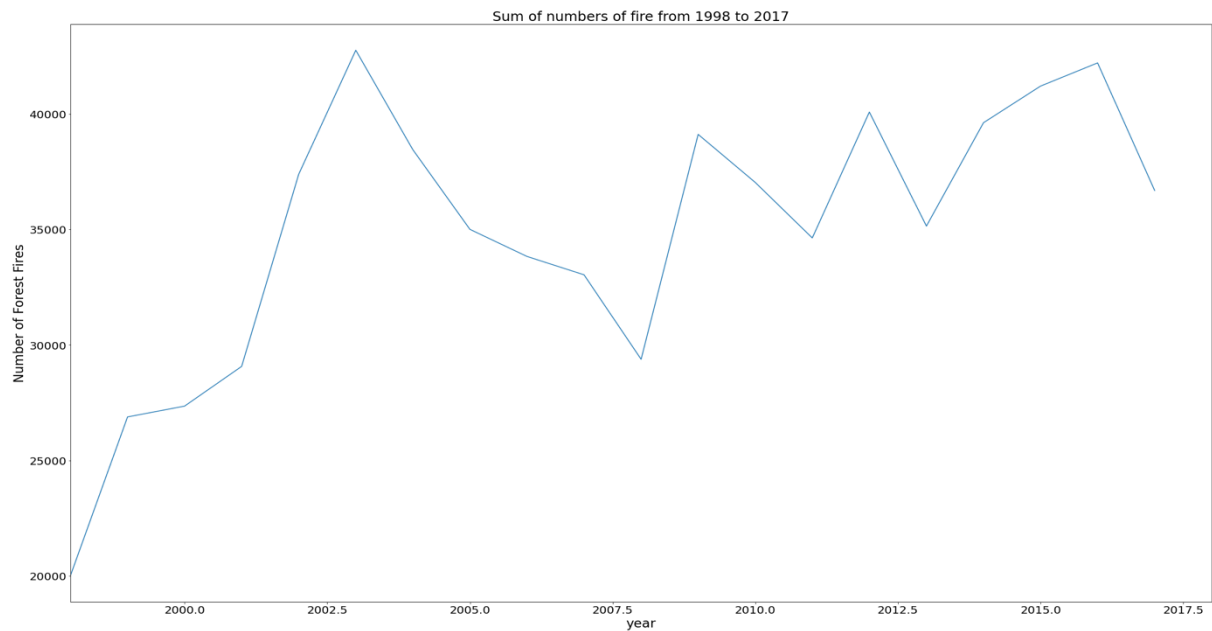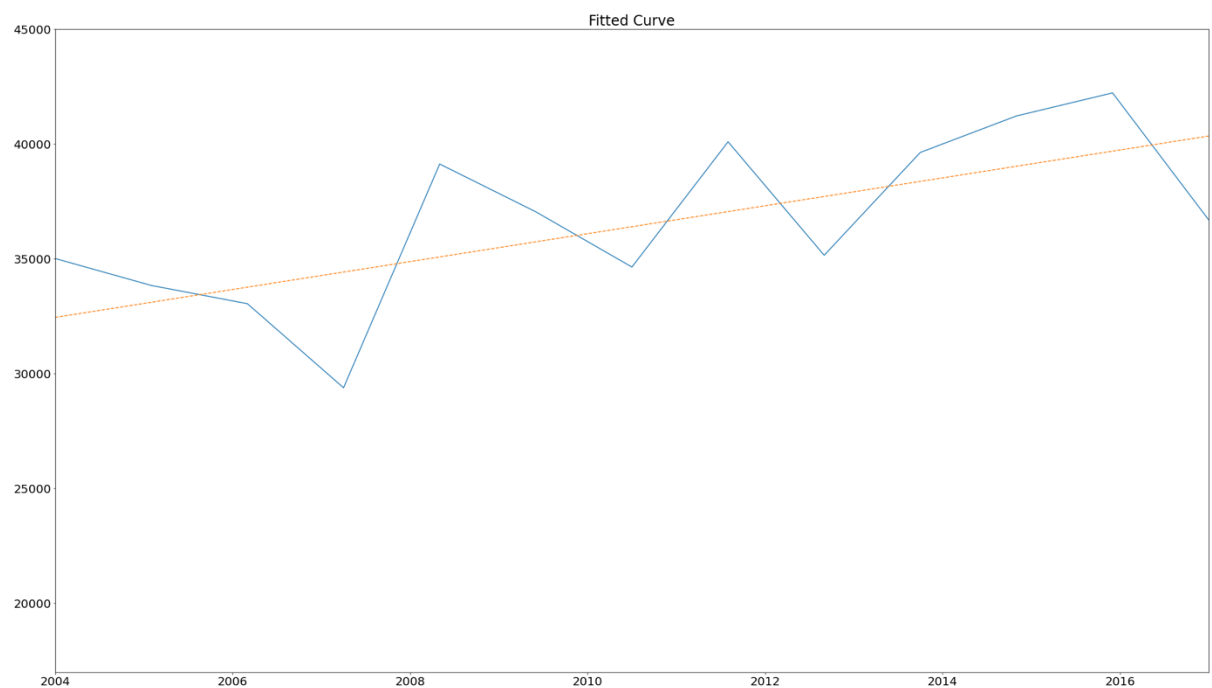

Scatter and spike type graph for visualizing trends and numbers as well

Sum of numbers of fire from 1998 to 2017

Predicting the trends in cases with 1st degree polynomial

Fitted Curve

Final prediction results:

```
2018 - 40942
2019 - 41549
2020 - 42156
2021 - 42763
2022 - 43370
2023 - 43978
2024 - 44585
2025 - 45192
```