

Amazon Rainforest Degradation

Introduction:

The aim of this project is to analyze the effect the rise in temperature has had in the degradation of Brazil's Amazon Rainforest over time. This includes analyzing the temperature changes in the country of Brazil and the degradation area of the forest, to observe if these events are related to one another in some way. For this project, two datasets were used.

Data Sources:

Data Source 1: FAO Temperature Change

This dataset provides the mean surface temperature change for a large number of countries, but for this project, only the country of Brazil will be taken into account. The data is from FAO of the UN, thus the quality can be assured. This data source is of data type CVS, and structured with columns such as country name, value of the mean temperature, etc. The license is CC BY 3.0 IGO, which gives the user the right to redistribute the material and transform and build upon it, even commercially.

Data Source 2: INPE Deforestation Area

This dataset provides the deforestation area of the Amazon Rainforest. The data is from INPE program PRODES, which is dedicated to monitoring the rainforest via satellite. This data source is also of data type CVS, and structured with columns such as occurrence year and different states. The license is CC0 Public Domain, which gives the user the right to modify and redistribute the work, all without asking for permission.

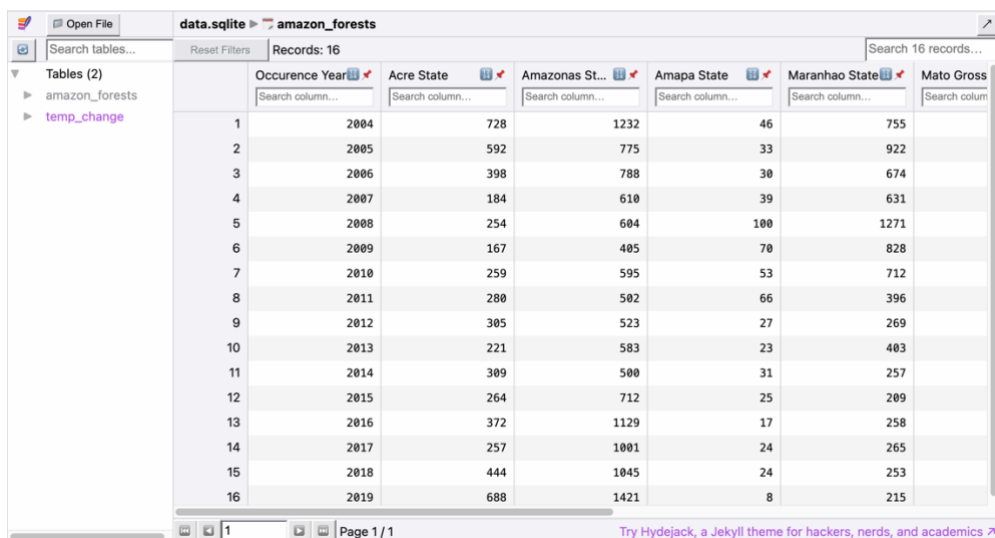
Data Pipeline:

The data pipeline is automated using the ETL (extract, transform, load) process. For the pipeline, various Python libraries were used, such as Pandas (for modifying the datasets), Numpy (for mathematical purposes; in this project it is used to identify empty rows), and SQLite (for storing the modified datasets).

Because the datasets were found on Kaggle, and the website has its own API to extract the datasets, what was done was that the appropriate datasets linked above were put into Google Drive, and extracted from there. On the datasets, some basic transformation was made, where only certain columns were kept

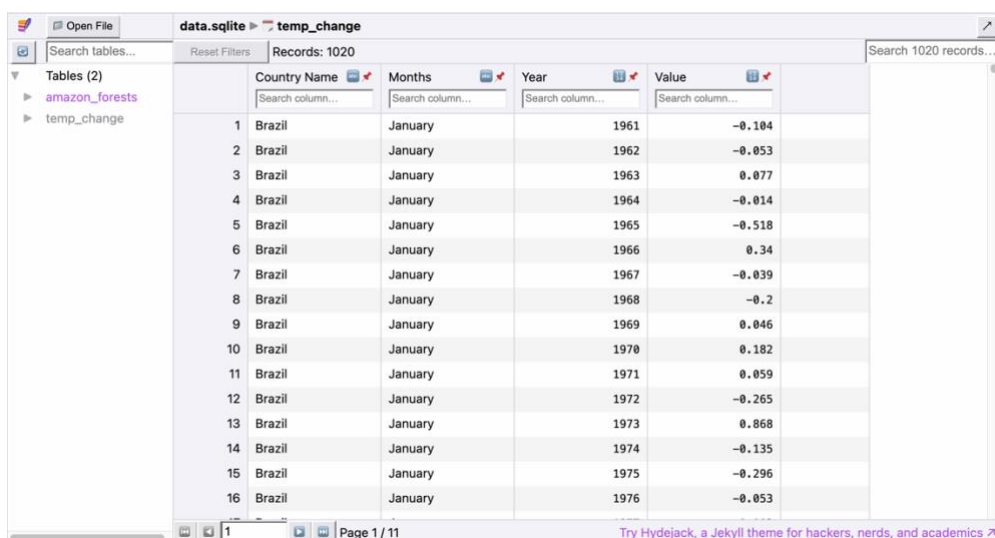
and empty rows and columns were deleted. For the rainforest dataset, a filter was applied so that only the results for the country of Brazil were visible. Then the datasets were saved into an SQLite database as two tables: `amazon_forests` and `temp_change`.

One problem that was encountered had to do with the extraction of the datasets from Kaggle. A solution provided from many users online was with using the requests library, but unfortunately, that did not work for me, so to combat this problem, I simply downloaded the appropriate files and uploaded them to Google Drive, so that they could be extracted from there. Although it is not ideal, the use of APIs was prohibited, thus this was the only other solution I could think of.



The screenshot shows an SQLite database viewer interface. The title bar indicates the database is 'data.sqlite' and the current table is 'amazon_forests'. The interface includes a search bar, a 'Reset Filters' button, and a 'Records: 16' indicator. The table structure is as follows:

	Occurrence Year	Acre State	Amazonas St...	Amapa State	Maranhao State	Mato Gross
1	2004	728	1232	46	755	
2	2005	592	775	33	922	
3	2006	398	788	30	674	
4	2007	184	610	39	631	
5	2008	254	604	100	1271	
6	2009	167	405	70	828	
7	2010	259	595	53	712	
8	2011	280	502	66	396	
9	2012	305	523	27	269	
10	2013	221	583	23	403	
11	2014	309	500	31	257	
12	2015	264	712	25	209	
13	2016	372	1129	17	258	
14	2017	257	1001	24	265	
15	2018	444	1045	24	253	
16	2019	688	1421	8	215	



The screenshot shows an SQLite database viewer interface. The title bar indicates the database is 'data.sqlite' and the current table is 'temp_change'. The interface includes a search bar, a 'Reset Filters' button, and a 'Records: 1020' indicator. The table structure is as follows:

	Country Name	Months	Year	Value
1	Brazil	January	1961	-0.104
2	Brazil	January	1962	-0.053
3	Brazil	January	1963	0.077
4	Brazil	January	1964	-0.014
5	Brazil	January	1965	-0.518
6	Brazil	January	1966	0.34
7	Brazil	January	1967	-0.039
8	Brazil	January	1968	-0.2
9	Brazil	January	1969	0.046
10	Brazil	January	1970	0.182
11	Brazil	January	1971	0.059
12	Brazil	January	1972	-0.265
13	Brazil	January	1973	0.868
14	Brazil	January	1974	-0.135
15	Brazil	January	1975	-0.296
16	Brazil	January	1976	-0.053

These are the resulting tables after the transformation.

Results and Limitations:

The output of the following data pipeline is an SQLite database with two tables, that will be used throughout the course of this project to answer the primary question of the effects the temperature has had on Amazonian rainforests. Through certain steps of cleaning the data and eliminating unnecessary rows and columns, the data has been transformed and normalized into appropriate SQL tables that will help in analysis. Because the original datasets were from government sources, the quality and reliability is high.

One potential problem that may arise is the compatibility of these two datasets, due to them having different year range. This might cause inaccuracies when analyzing them and can impact how rigorous the overall analysis may be, considering the documentation of the Amazon rainforest deforestation does not span across many years.