

Introduction - Motivation

There is a debate among economists and environmentalists that economic development comes at the cost of deforestation. Over the last 10000 years, we have lost one-third of the world's forest cover, half of which happened in the last 100 years. However, not all kind of economic advancements may be linked with deforestation. ^{[3],[13],[16]}.

Problem Definition

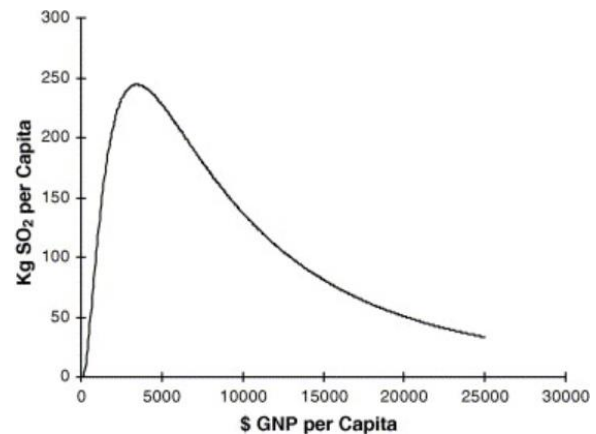
The objective is to determine what correlations exist between deforestation and socio-economic factors like per capita GDP, inflation, infant mortality etc. By bringing multiple such factors together, we can do meaningful analysis of the net effect of these factors on deforestation ^[12] across countries.

Our aim is to orchestrate a data pipeline, train analytical models and build interactive visualizations that would help users garner critical information on this important topic. We will make use of data from reliable sources like worldbank.org, undp.org and unicef.org for this project.

Survey

Research has linked deforestation ^[11] with trade-liberalization ^[1], unemployment ^[2], pollution and carbon emission ^[5], infant mortality ^[6], colonization ^[18], population ^{[7],[14]}, GDP ^[8] and corruption ^{[15],[19]}.

Most of the current studies are based on a single factor and have not been inclusive of other factors. For example - the Environmental Kuznets curve (EKC) hypothesizes the relationship between various indicators of environmental degradation like deforestation and income per capita ^[9].



Even though the existence and relevance of the first half of the EKC curve is well established ^[10], we feel more analysis needs to be done with multiple factors.

Proposed Method

Intuition

Deforestation is a complex topic and there could be many localized factors at play. Factors relevant for one country may not be applicable to another.

We analyzed many research papers on this topic and realized the need for

- Analyzing different factors collectively to gain insights on the causes of deforestation
- An interactive visualization that would allow the user to slice and dice the information and derive insights from it.

Our project aims to address these two requirements in an innovative way. We intend to -

- Analyze each of these factors against deforestation data
- Identify the most relevant factors by ranking them by correlation score.
- Predict future deforestation rate for a country / region
- Identify cluster of countries based on all relevant data points.

The intuition behind our approach is to provide a unique and creative platform to analyze and identify the socio-economic factors that are most relevant to the deforestation rate globally / country / region. This platform can be utilized by government agencies and independent researchers to derive policy decisions. Once the most relevant factors are identified, a deeper study can be done on the shortlisted factors. This will help us derive inferences about the strength of the relationship of these factors with deforestation.

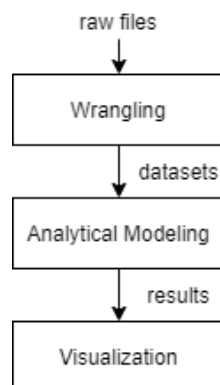
Additionally, we will rank the factors in order of their importance and bring all meaningful comparisons under one hood. User can interact with the data and compare various factors against deforestation statistics across individual countries or collectively.

Approach

We will use clustering and time series exponential smoothing algorithms for studying deforestation rate against the factors and for forecasting.

The project is divided into three sub-systems:

1. Wrangling – cleanse, transform, and integrate the disparate data files
2. Analytical Modeling – prep and train the models with the data to generate the results
3. Visualization – Dashboards to expose the results to user



Data

Raw data was obtained from eight different websites - data.worldbank.org, hdr.undp.org, ourworldindata.org, data.unicef.org, stats.oecd.org, risk-indexes.com, transparency.org, and blogs.worldbank.org as flat files.

The factors that we have considered are –

- a) Forest area (% of land area)
- b) GDP per capita
- c) GNI per capita (Atlas method)
- d) Inflation % of consumer prices
- e) Population density (people per sq. km of land area)
- f) Human Development Index
- g) Unemployment %
- h) Infant mortality %
- i) Global Corruption Index

A representative example of the integrated dataset is as mentioned below -

Field	Example Value
Continent	Asia
Country	India
Country income level as defined by latest World Bank country classification	Lower-middle income
Year	2020
Socio-economic factor	GDP per capita (current US\$)
Measurement of socio-economic factor	1900.70

Compute

The datasets are cleansed, transformed, aggregated to obtain a yearly measure by each country for the last 30 years, and passed through a compute layer that calculates the target measures and obtains the analytical models results. The compute layer performs the following –

- Existence of a relationship between annual % change in Forest area (% of land area) and annual % change in socio-economic factor
- Strength of the relationship between the factors and deforestation using Pearson's correlation coefficient
- For each country, rank the factors by their impact on deforestation
- Segment the data points into unsupervised clusters using k-means algorithm based on their similarity and closeness to each other

Visualization

The visualization subsystem consists of four interactive Tableau dashboards –

Dashboard #1

Filters: Geography (Continent, Income level, Country), Socio-economic factor, Time range
Layout: Heatmap/choropleth with average correlation coefficient, trendlines depicting correlation by year

Dashboard #2

Filters: Geography (Continent, Income level, Country), Socio-economic factor, Time range
Layout: Table of Country and their ranking of factors

Dashboard #3

Filters: Geography (Continent, income level, country), future year
Layout: Predicted heatmap/choropleth with Country and Deforestation % in tabular format

Dashboard #4

Filters: no filters
Layout: K column table mentioning cluster of countries as obtained via K-means clustering

Design of Experiments/ Evaluation

Testbed consists of following components:
OS: Windows 10
Browser: Google Chrome

Python: Python 3, Pandas, Numpy

Tableau: Tableau Desktop 2021

Database: SQLite 3

IDE: Visual Studio Code, Jupyter notebook

Version Control: GitHub

The questions the Tableau dashboards will be designed to answer are as follows

Dashboard #1

- Which countries have seen an increase/decrease in the relationship strength?
- Which countries are worst affected?
- Which countries have performed the best?
- What is the average as well as yearly changes to the correlation coefficient between the socio-economic factors and deforestation?

Dashboard #2

- Which socio-economic factors are the most important in understanding deforestation?

Dashboard #3

- What would be the predicted deforestation % in the future?

Dashboard #4

- Which countries are group together based on similarity and closeness to each other?

In the short to mid-term, we are measuring the relevance and relationship of the factors with deforestation rate. In the long term, we hope these results would help in conducting deeper studies on the higher ranked factors. Also, we would like to showcase the results to non-government agencies, which will help them draft policies for the future.

Observations

Once the data pipeline is created and the analytical model is trained, we would analyze

the strength of the relationships between deforestation and the factors. Then, we would review the Tableau visuals and make observations on the information and the patterns detected.

The observations will be listed against following questions after the data pipeline and computations are performed:

1. Which socio-economic factor is ranked the highest overall, by continent, and by income group level?
2. Which countries have performed better than their neighbors?
3. Which countries should be analyzed further to understand how deforestation can be slowed down without impacting the socio-economic factors?
4. What is the predicted level of deforestation?

Conclusions and discussion

This section will be populated once the conclusions are drawn based on the analysis.

Distribution of team member efforts

The project team is a mix of Project Management, Functional and Technical experts.

Thus far, all team members have contributed similar amount of effort in the following activities:

- exploring project problem statements
- conducting literature survey
- identifying possible data sources
- discussing data transformation and high level design
- project deliverables planning and coordination

Future workload has been equally divided amongst the team members as follows.

Responsibility	Champion(s)
Project Lead	Subhabrata Chaudhuri
Wrangling	Piyush Jain, Sahil Poonatar, Suneet Taparia, Sanjay Naik
Analytical Modeling	Suneet Taparia, Sanjay Naik, Saurabh Sinha
Visualization	Subhabrata Chaudhuri, Piyush Jain, Saurabh Sinha, Sahil Poonatar

References:

1. Juan Robalino and Luis Diego Herrera, 2009. Trade and Deforestation. World Trade Organization, p. 30
2. Muhammad Tariq, 2015. An Overview of Deforestation Causes and Its Environmental Hazards in Khyber Pukhtunkhwa. Journal of Natural Sciences Research, p. 4
3. Robert Walker, 1993. Deforestation and economic development. Canadian Journal of Regional Science, p.483
4. EK Yiridoe, DM Nanang, 2001. An econometric analysis of the causes of tropical deforestation: Ghana. American Agricultural Economics Association Conference
5. Bo Pieter Johannes Andrée, Andres Chamorro, Phoebe Spencer, Eric Koomen, Harun Dogo, 2019, Revisiting the relation between economic growth and the environment; a global assessment of deforestation, pollution and carbon emission, Renewable and Sustainable Energy Reviews, Elsevier
6. A Chakrabarti, 2021, Deforestation and infant mortality: Evidence from Indonesia, Economics & Human Biology, Elsevier
7. M Bhattarai and M Hamming, 2001. Institutions and the Environmental Kuznets Curve for deforestation, World Development Vol. 29, p. 5
8. Bolarinwa A. Ajanaku and Alan R. Collins, 2020. Economic growth and deforestation in African countries: Is the Environmental Kuznets Curve hypothesis still applicable? Agricultural & Applied Economics Association Annual Meeting, Kansas City, p. 26
9. David I. Stern, June 2003. The Environmental Kuznets Curve
10. Jeremy Irvin, Hao Sheng, Neel Ramachandran, Sonja Johnson-Yu, Sharon Zhou, Kyle Story, Rose Rustowicz, Cooper Elsworth, Kemen Austin, Andrew Y. Ng, Nov 2020. ForestNet: Classifying Drivers of Deforestation in Indonesia using Deep learning on Satellite Imagery
11. Jesús Crespo Cuaresma, Olha Danylo, Steffen Fritz, Ian McCallum, Michael Obersteiner, Linda See & Brian Walsh, 2017. Economic development and forest cover: Evidence from satellite data
12. Shichao Gao, 2019. Deforestation prediction using Time Series and LSTM
13. G. Cornelis van Kooten and Sen Wang, 2003. Institutional, social and economic factors behind deforestation: a cross-country examination
14. K Pahari, S Murai, 1999, Modelling for prediction of global deforestation based on the growth of human population, ISPRS journal of photogrammetry and remote sensing, Elsevier, p.320
15. Cuneyt Koyuncu and Rasim Yilmaz, 2009, The Impact of Corruption on Deforestation: A Cross-Country Evidence, The Journal of Developing Areas, Vol. 42, No. 2, pp. 213-222
16. Lykke E. Andersen, 1997, Modelling the Relationship between Government Policy, Economic Growth and Deforestation in the Brazilian Amazon
17. Huirong Feng, C. W. Lim, Liqun Chen, Xinnian Zhou, Chengjun Zhou, Yi Lin, "Sustainable Deforestation Evaluation Model and System Dynamics Analysis", The Scientific World Journal, vol. 2014, Article ID 106209, 14 pages, 2014.
<https://doi.org/10.1155/2014/106209>

18. Haeuber, Richard. "Development and Deforestation: Indian Forestry in Perspective." *The Journal of Developing Areas* 27, no. 4 (1993): 485–514.
<http://www.jstor.org/stable/4192258> .
19. Atrayee Banerjee and Chowdhury Madhurima, Vol. 5(8), pp. 122-129, September 2013, 'Forest degradation and livelihood of local communities in India: A human rights approach'
<https://academicjournals.org/journal/JHF/article-full-text-pdf/488F2773349>