Green Analytics

Subhabrata Chaudhuri, Saurabh Sinha, Sahil Poonatar, Piyush Jain, Suneet Taparia, Sanjay Naik

Summary

Analysis of forest cover is a complex topic, and the causes of forest cover degradation are widely debated. Green Analytics is a unique, creative and interactive platform to analyze, visualize and identify the socio-economic factors that are most relevant to the changes of Forest cover rate. It is a user-friendly visual tool that determines correlations against Forest cover changes and socio-economic factors, over the past 30 years.

Significance

Most current studies are based on a single social or economic factor. Green Analytics brings multiple factors, together under one umbrella. Users can perform meaningful analysis of the net effect of these factors on forest cover. Analysis can be done based on time, geography and income classification. This can be utilized by non-profit organizations, government agencies and independent researchers to drive policy decisions.

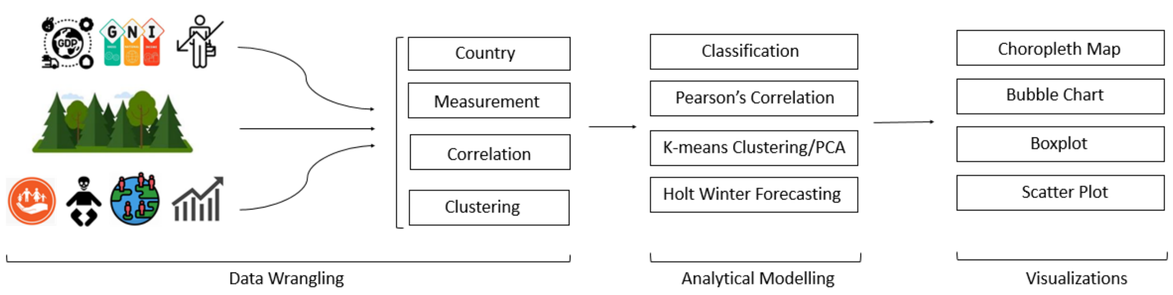
The Data

Green Analytics combines and standardizes data from multiple sources

Input Data	Source	Temporal	Approach
Forest Area	worldbank.org	Yes	Download
GDP per Capita	worldbank.org	Yes	Download
GNI per Capita	worldbank.org	Yes	Download
Inflation	worldbank.org	Yes	Download
Population	worldbank.org	Yes	Download
Human Development Index	undp.org	Yes	Download
Unemployment	worldbank.org	Yes	Download
Infant Mortality	unicef.org	Yes	Download
Income Classification	worldbank.org	No	Web Scraping
Development Status	un.org	No	Download

Output Data	Record #	Purpose
Country	207	Country name, continent, income status and development status
Measurement	103291	Country id, Year, Factor measurements, Actual vs Predicted
Relationship	1449	Correlation measure & classification
Clustering	217	Country id, Factor measurements and cluster id

The Approach



Data collection was done by downloading and web scraping.

Data cleaning and analysis was implemented using Python (pandas, sklearn, statsmodels, urllib, regex, matplotlib). Data cleaning involved eliminating countries where factor measurements were missing or the datapoints were outliers. Cleansed data was stored in csv files.

Data analysis involved -

- Calculation of correlation score between forest cover and factors using Pearson's correlation coefficient.
- Identification of top ranked factors using classification and ranking.
- Principal Component Analysis for dimensionality reduction and k-means clustering on the principal components. This was done to derive insights which would not have been available via classification.
- Prediction of Future Forest cover using Holt Winter method.

The output of this analysis is captured in multiple Tableau dashboards.

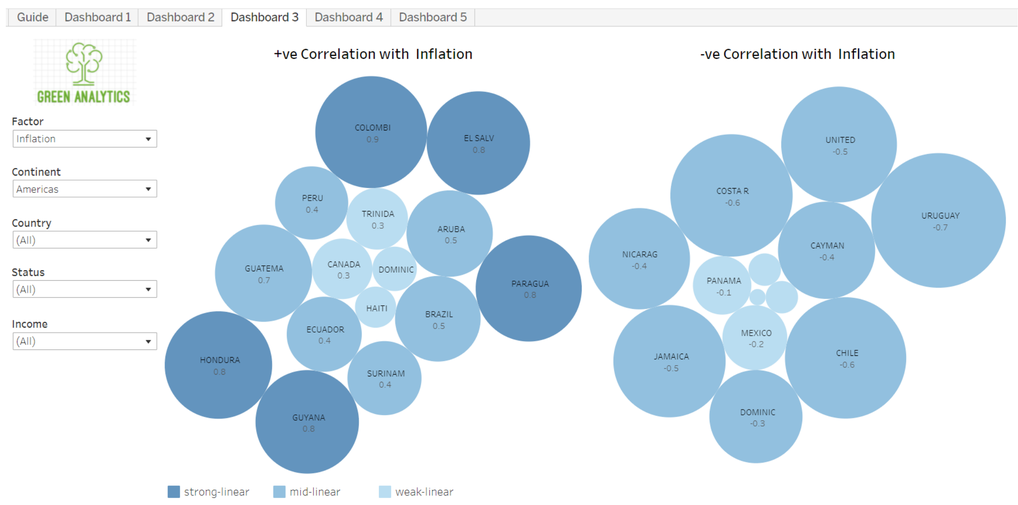
Uniqueness of our approach

By applying the above techniques, users can narrow down the potential causation factors of deforestation for a particular country or region. Knowing this would help countries strike the right balance between economic development and forest cover.

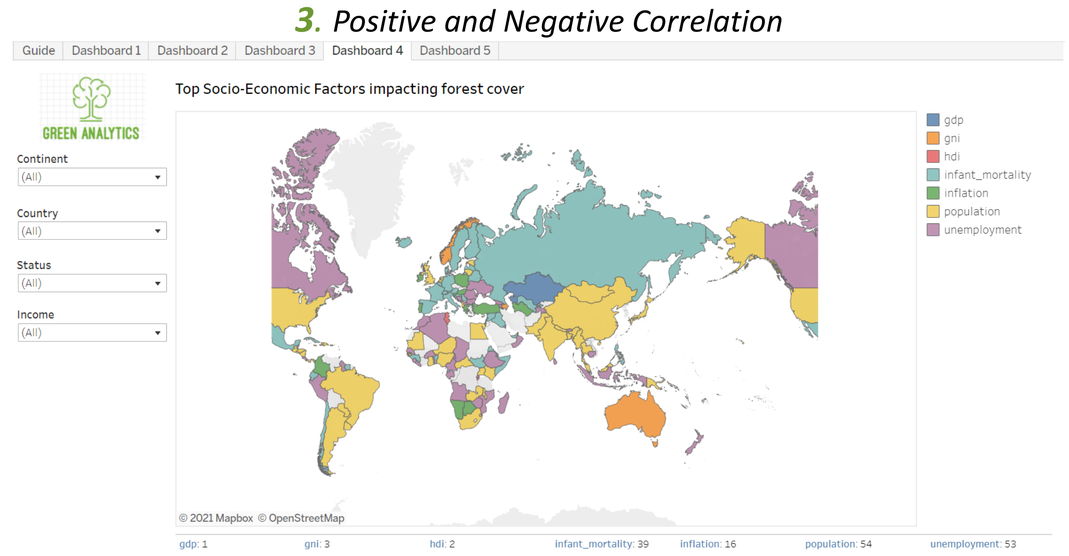
User Interfaces

The following dashboards in Tableau capture the output -

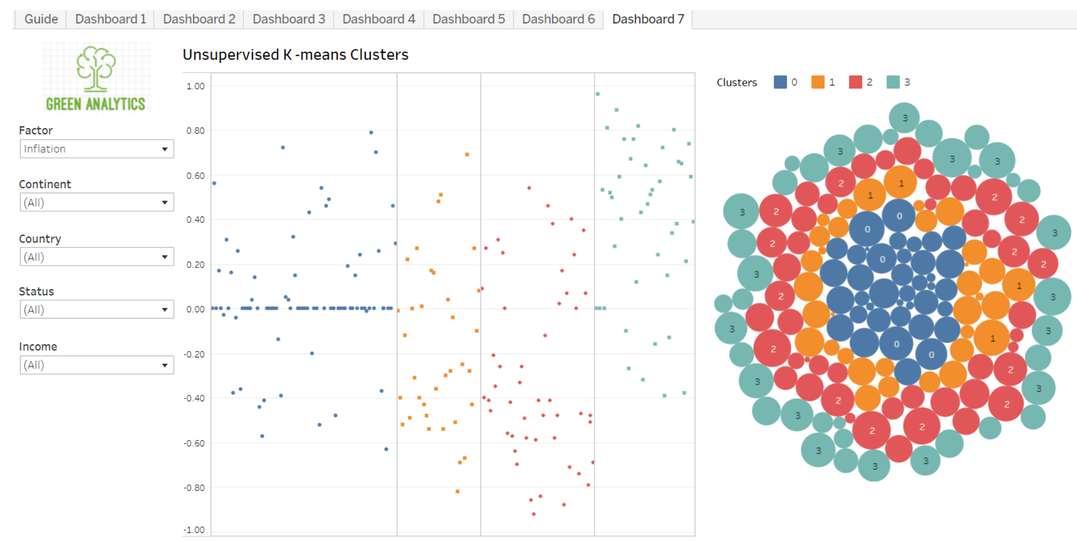
1. Choropleth World Map for every factor along with trend.
2. Choropleth World Map depicting strength of correlation for every factor.
3. Identification of countries with positive and negative correlation with forest cover for every factor.
4. Boxplot of factors and their correlation
5. Choropleth World Map depicting the top ranked factor.
6. Predicted forest cover in 2025.
7. Unsupervised k-means clusters



3. Positive and Negative Correlation



5. Top Socio-economic Factor impacting Forest cover



7. Unsupervised k-means

Experiments and Results

The dashboards help us answer the following questions -

1. Which socio-economic factor is ranked the highest overall, by continent? *For Africa, unemployment is the top factor correlating to forest cover change for 18 countries (Dashboard -5)*
2. Which countries have performed better than their neighbors? *Amongst European countries, Malta has the highest correlation (0.9) of population with forest cover. This is contrary to popular belief. (Dashboard -3)*
3. Which countries should be analyzed further to understand how deforestation can be slowed down without impacting the socio-economic factors? *Portugal has a mid linear correlation score between HDI and forest cover. This implies that the country has improved its HDI without compromising its forest cover. (Dashboard -3)*
4. What is the predicted level of deforestation? *Amongst developed countries, the United States will increase its forest cover by 0.53% by 2025 (Dashboard -6)*

Based on the inferences derived above, we can conclude that not all socio-economic factors have a negative impact on forest cover.

Conclusion

We have adopted a more holistic approach towards understanding the potential causes of change in forest cover. Based on the user's choice of region or country, similar analysis can be performed to identify factors that are not impacting the forest cover. These countries can be studied in detail and their model can be replicated in other countries.

https://public.tableau.com/app/profile/subh6914/viz/GreenAnalytics_16375909366540/Dashboard1