

Capstone 1 Proposal – Predicting Biocapacity

Springboard Data Science Career Track

Melissa Schmitz

Background

A nation's ecological footprint is a measure of the ecological assets that its population requires in order to produce the natural resources it consumes and to absorb the waste generated. Consumables include plant-based food and fiber products, livestock and fish products, timber and forest products, and space for urban infrastructure. A type of waste would include carbon emissions.

The aggregate total footprint for each country includes six categories of productive surface areas (cropland, grazing land, fishing grounds, urban land, forest area, and carbon demand on land) whereas a country's biocapacity represents the productivity of these ecological assets. These areas can also absorb generated waste, especially carbon emissions.

If a population's ecological footprint exceeds the biocapacity of the region, that nation creates an ecological deficit. That is, its demand for resources that land and water can provide has exceeded what the region's ecosystems can renew. Such a deficit prompts a nation to meet those demands through importing, draining its own ecological assets (e.g., overfishing), and/or emitting carbon dioxide into the atmosphere. Alternatively, if a region's biocapacity can exceed its ecological footprint, it then has an ecological reserve (producing more than it consumes).

While the ultimate goal of environmentalists is to reduce all types of pollution and improve sustainability holistically, governing bodies do not necessarily share the same priorities. Therefore, analysis of these factors could provide a means of prioritization in order to create more tangible goals for government regulations or independent environmental programs.

Dataset

The ecological footprint measure was invented by Mathis Wackernagel and William Rees at the University of British Columbia and this ecological footprint data was provided by the Global Footprint Network. The variables contained in this data set include:

- **Country** – The country where data in the other columns originated.
- **Region** – The major part of the world where the Country resides.
- **Population** – The number of people residing in that country, in millions.
- **HDI** – [The Human Development Index](#), an assessment of the capability of a country's inhabitants, shifting focus away from economics as a measure of success. It is measured with indicators such as life expectancy, education levels, and standard of living (GNI index).

- **GDP per capita** – The Gross Domestic Product, a monetary measure of the market value of all final goods and services which does not take into account the cost of living or inflation rates.
- **Cropland Footprint** – The amount of cropland required to meet the crop demands of a country. Measured in global hectares.
- **Grazing Footprint** – The amount of land required within a country to meet the demands of animal grazing. Measured in global hectares.
- **Forest Footprint** – The amount of land within a country where trees are cut down for lumber. Measured in global hectares.
- **Carbon Footprint** – The measure of CO₂ emissions associated with fossil fuel use. Measured in units of tonnes per year.
- **Fish Footprint** – The area within a country where fish are collected for food. Measured in global hectares.
- **Total Ecological Footprint** – A sum of the demand on biocapacity measured from the aforementioned footprint data. Measured in global hectares.
- **Cropland** – The land area within a country available for harvesting crops. Measured in global hectares.
- **Grazing Land** – The land area within a country available for animal grazing. Measured in global hectares.
- **Forest Land** – The land area within a country available for forests. Measured in global hectares.
- **Fishing Water** – The land area within a country available for fishing. Measured in global hectares.
- **Urban Land** – The land area within a country that is built-up with man-made structures. Measured in global hectares.
- **Total Biocapacity** – The capacity of a physical area in an ecosystem to produce resources, regenerating what is removed by demand, and absorb waste material. Measured in standardized hectares.
- **Biocapacity Deficit** – The difference between total biocapacity and the total ecological footprint.
- **Earths Required** – The number of Earths required to sustain the consumption of a country.
- **Countries Required** – The number of countries required to sustain the consumption of a single country.

Note: There is a productivity scaling factor (the equivalence factor) that converts a specific land type into a universal unit of biologically productive area, a global hectare. In a given year, the equivalence factors are the same for all countries.

[\[link to full glossary of relevant terms\]](#)

[\[link to dataset\]](#)

Project Approach

Data Cleaning: Check for missing data and determine the appropriate substitute for missing values. Examine any major outliers to see if they have significance to the data. Search for other abnormalities in the data set.

Initial Analysis: Exploratory data analysis with data visualization techniques to find interesting patterns. Look at potential correlations between different variables to determine how they might impact prediction results.

Core Analysis: TBD. Suspected regression to predict the biocapacity deficit or reserve based on the types of footprints and available resources. May also examine whether population is the only major impact on resources by looking at data per capita rather than gross data.

Writeup

The deliverables will include a PowerPoint presentation and a Github repository with the main Python code displayed in Jupyter notebooks with appropriate comments and notes.