WORLD FAMINE:

GLOBAL CROP YIELD ANALYSIS

DEC 2021

Currently there are nearly 1 in 8 people who do not have enough food to lead a healthy life.

870 million people do not have enough food to eat.

Global population expected to reach 8.5 billion by 2030

Data Science Team

Alessandro Chiari

Alessandro is a data scientist from **Italy**, whose agriculture is dominated by **wheat**, **rice**, and **potatoes**.

Adam M. Klesc

Adam is a data obsessive from **Indonesia**, whose agriculture is dominated by **palm oil** but are better known for their **coffee**!

Manu Mulaveesala

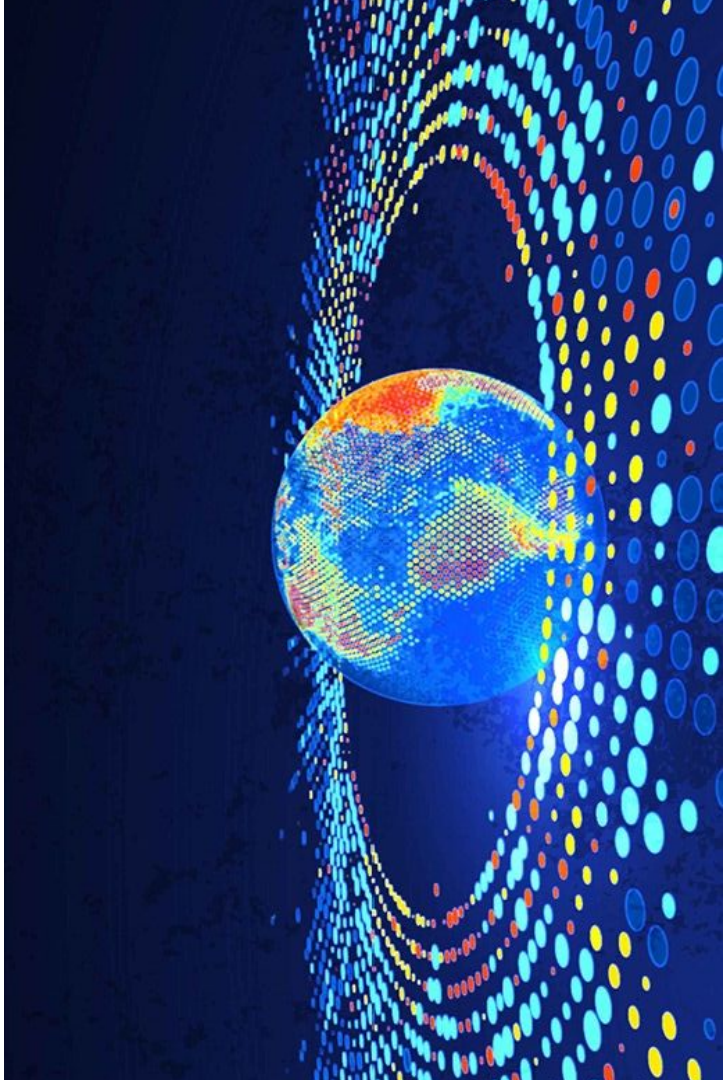
Manu is a data scientist from **India**, whose agriculture is dominated by **rice**, **wheat**, and **cotton**.

Mona Kalika

Mona is a data scientist from **Guyana**, whose agriculture is dominated by **sugar** and **rice** production. Most famous for Demerara Brown Sugar!

Data Science and World Hunger

- ➔ You can’t feed the world with data – or can you?
- ➔ If Data has the power to transform business, it has the power to transform global initiatives at scale
- ➔ A key challenge of global hunger is *scale*



The Data Science Process

1) Gather Data
2) Exploratory Data Analysis (EDA)
3) Data Cleaning and Processing
4) Building Machine Learning models
5) Evaluating our Models
6) Making Conclusions about our Findings

The scope: Create a model that successfully predicts crop yield

Data:

Food and Agriculture Organization (FAO) of the United Nations

Food Crops:

Maize, Potato, Rice, Wheat, Sorghum, Cassava, Barley, Soybeans, Yams

Crop Yield:

Crop yield measured as hectograms / hectare (hg/ha)

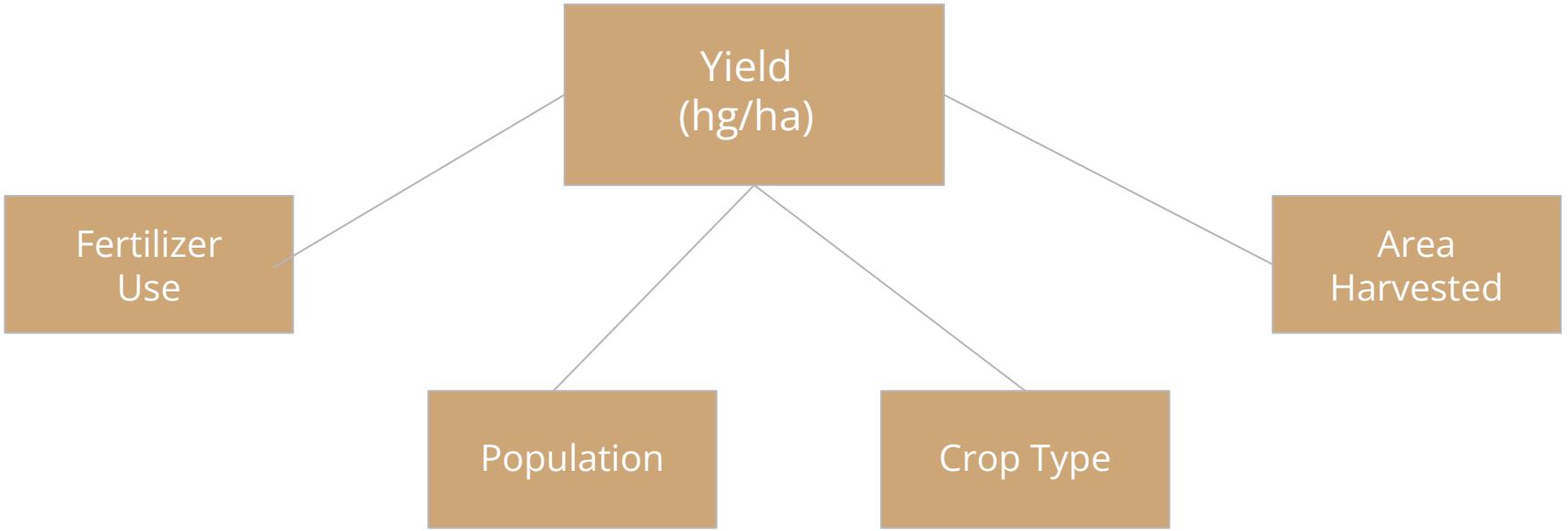
Models:

7 Regression Models Tested
2 Feature Sets Tested

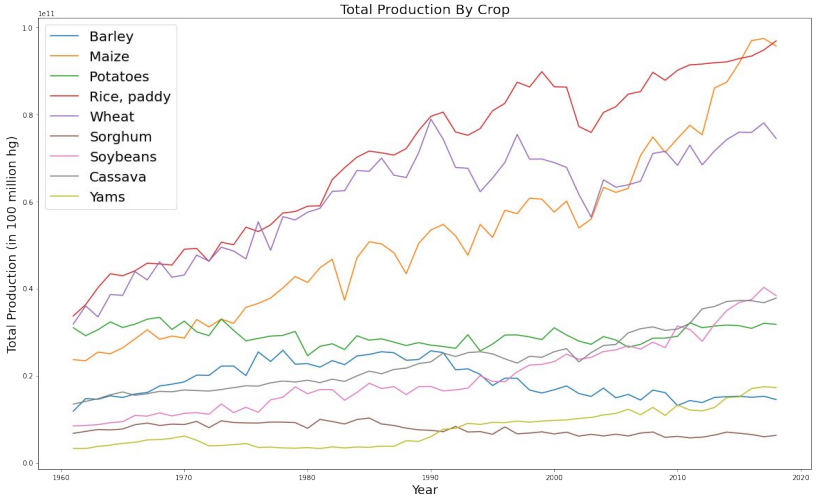
Goal for 2030:

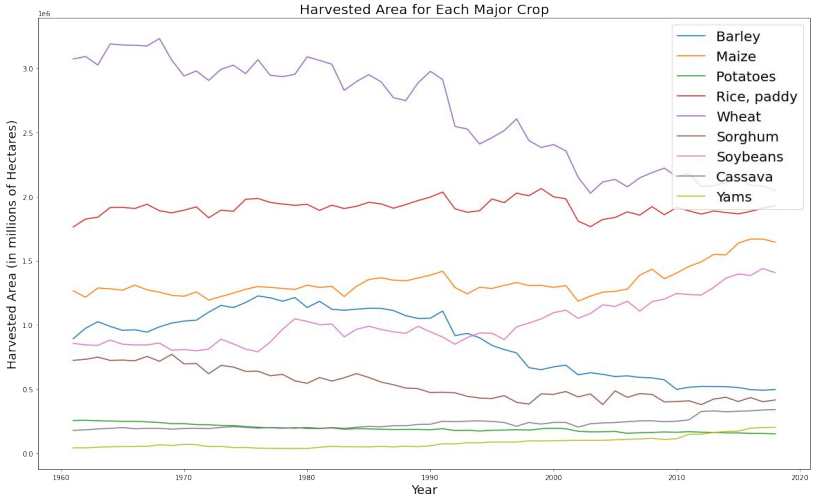
In order to end world hunger by 2030, the UN needs to come up with a strategy for nearly 940 million at the current rate of population growth.

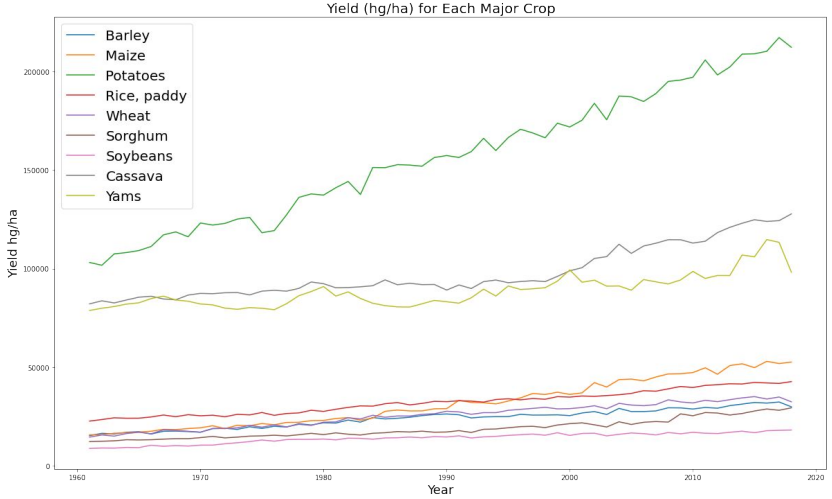
EDA BRIEFING

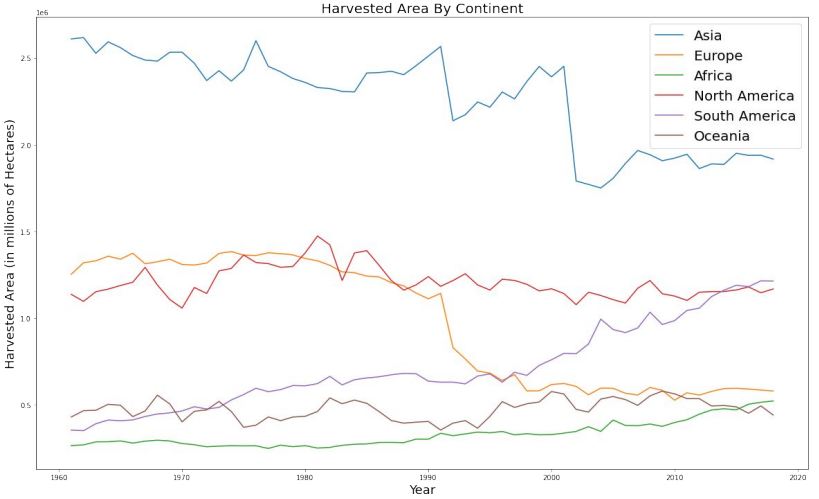


Global Production, Yield and Area

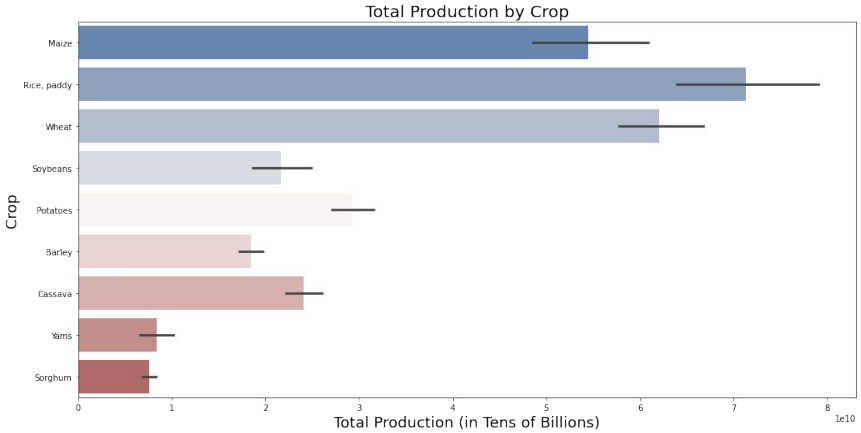


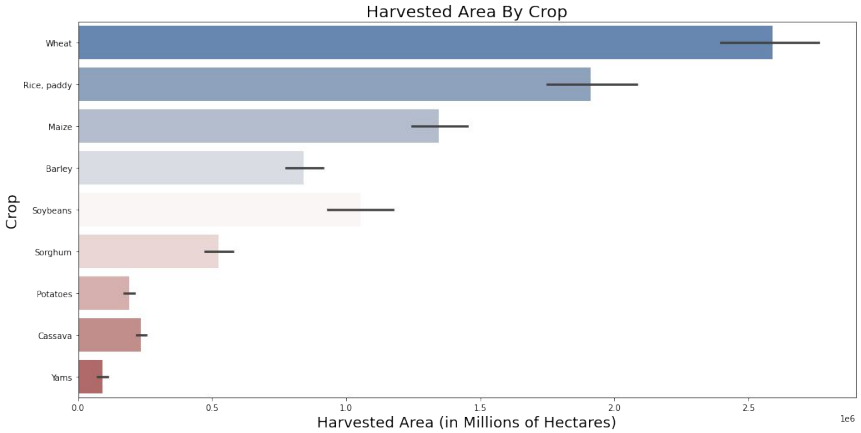


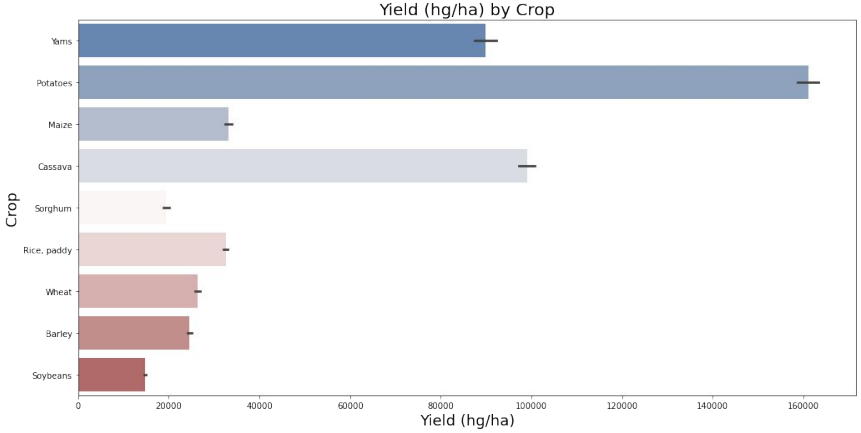


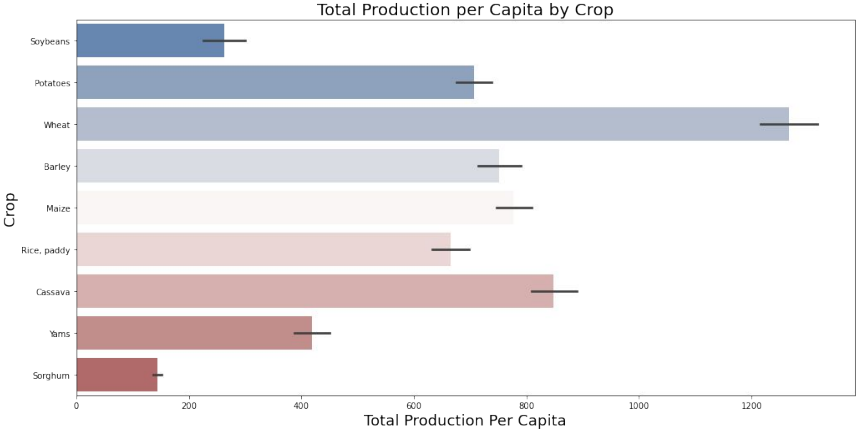


Global Production, Yield and Area

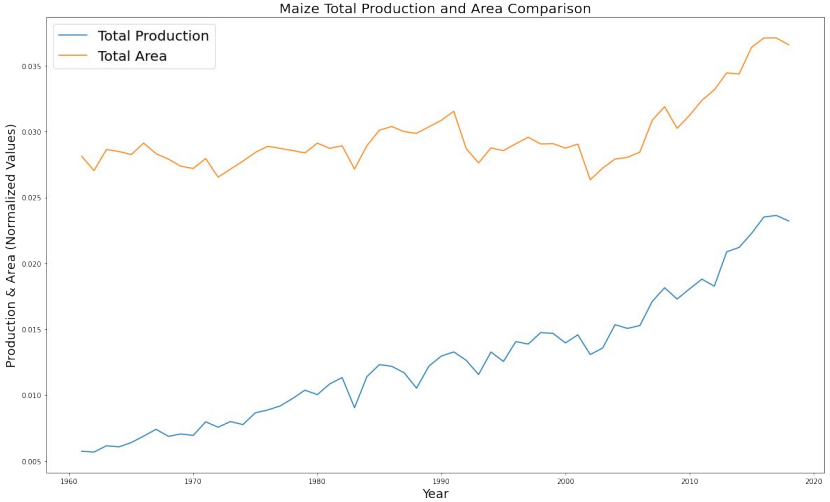


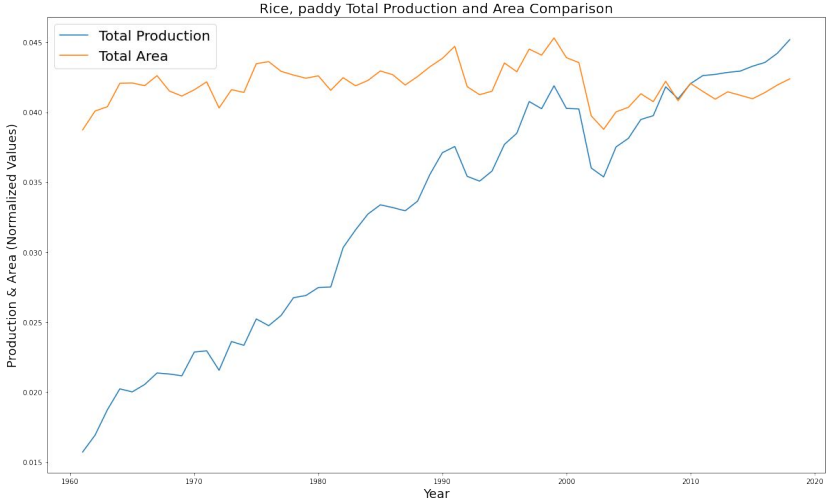


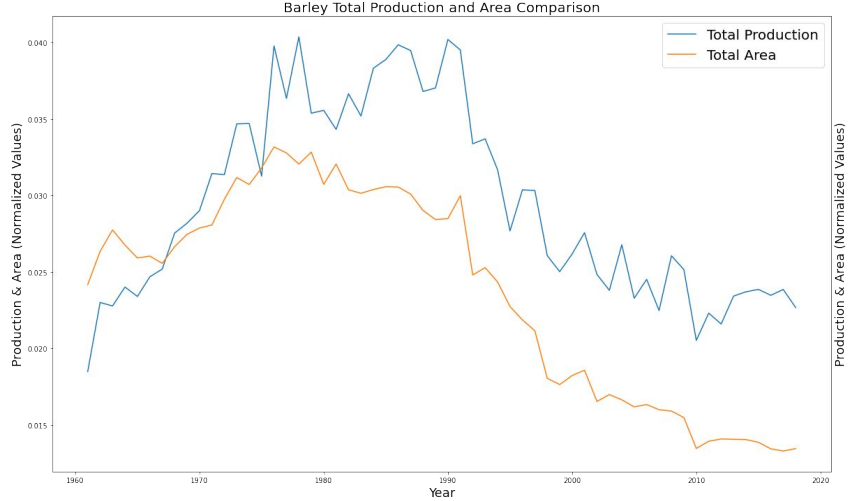


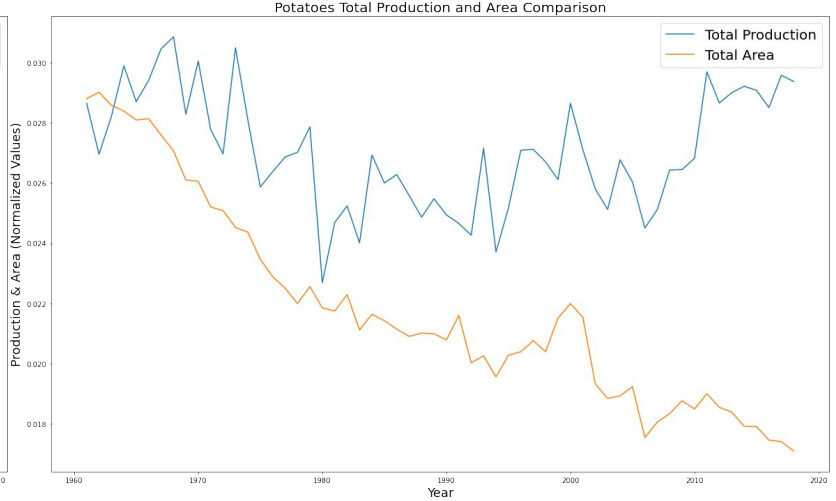


Global Production, Yield and Area

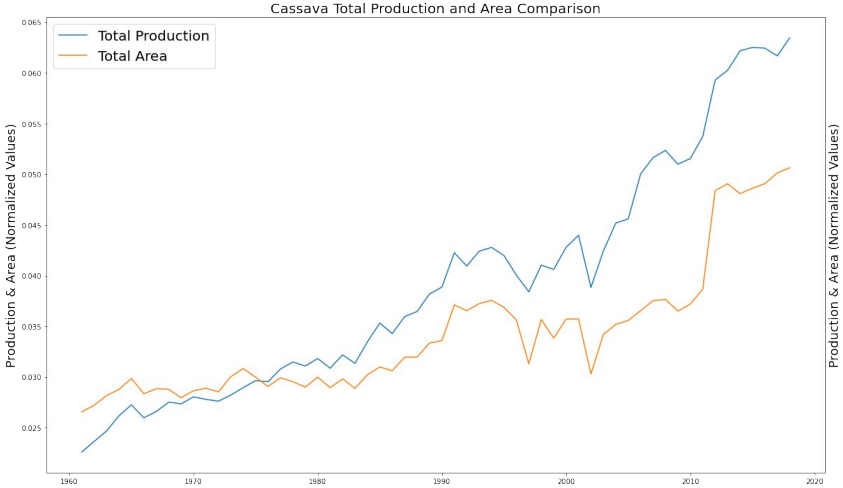


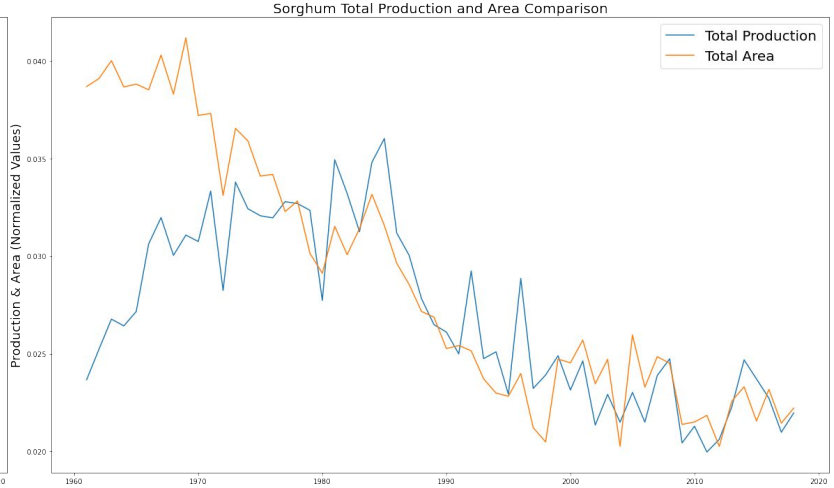


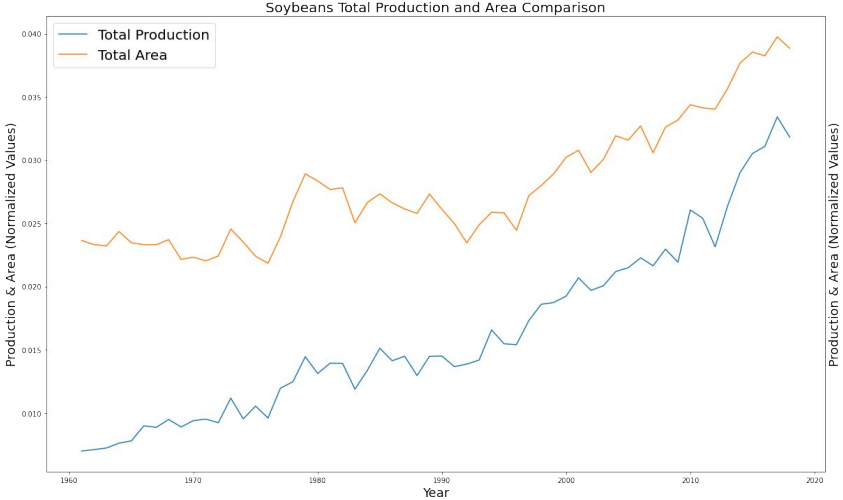


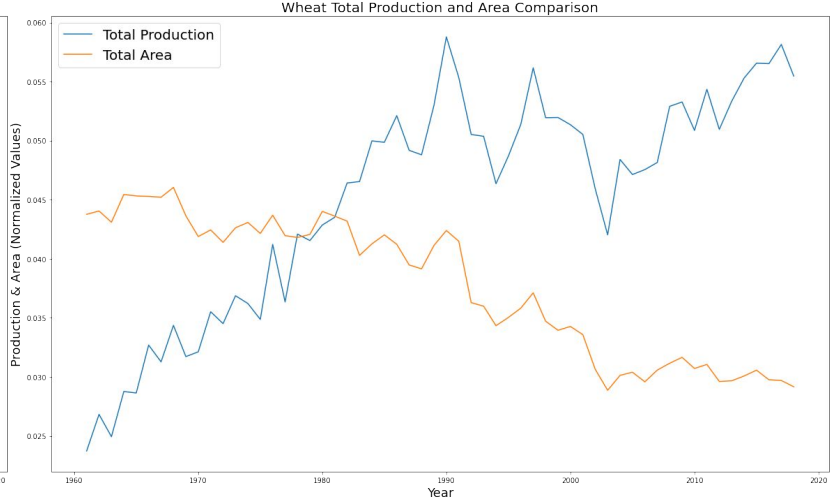


Global Production, Yield and Area

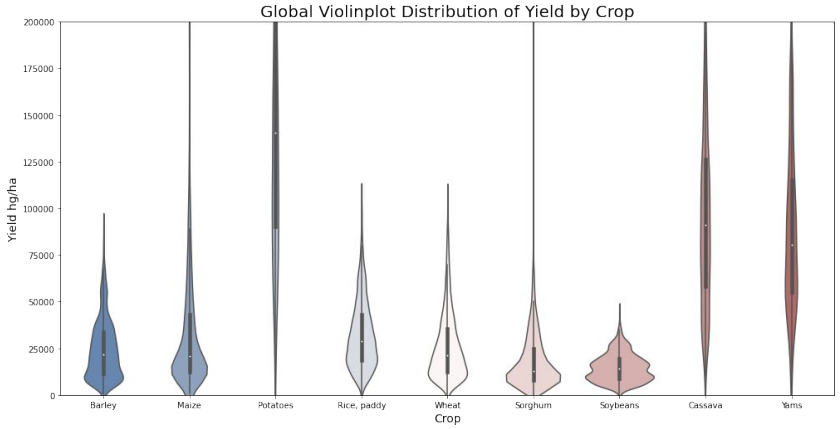


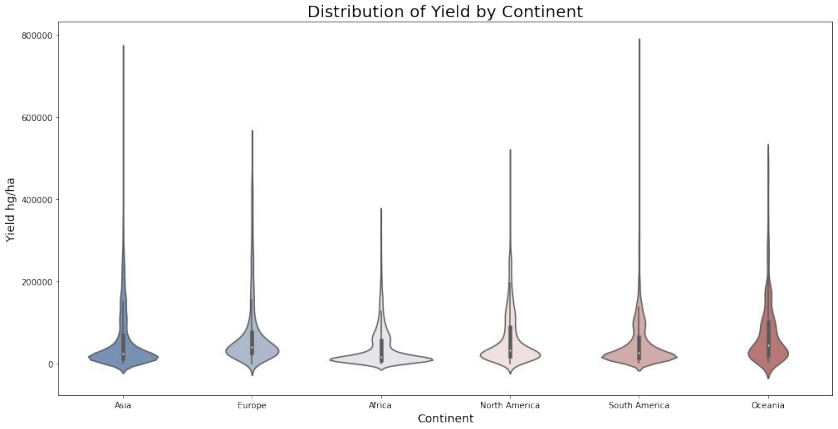


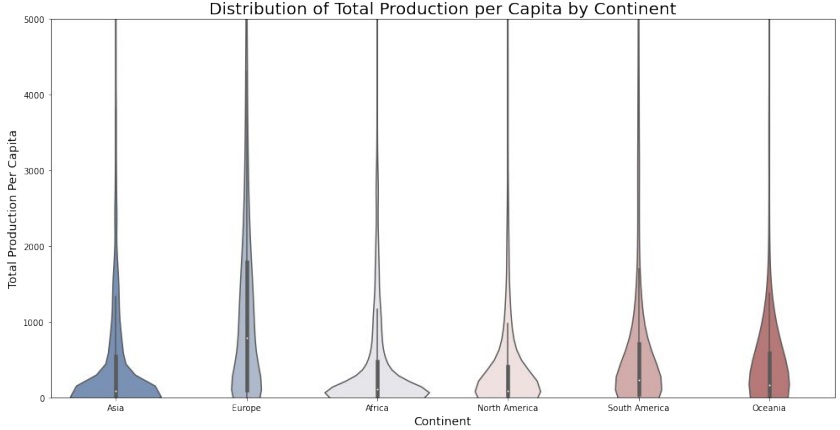




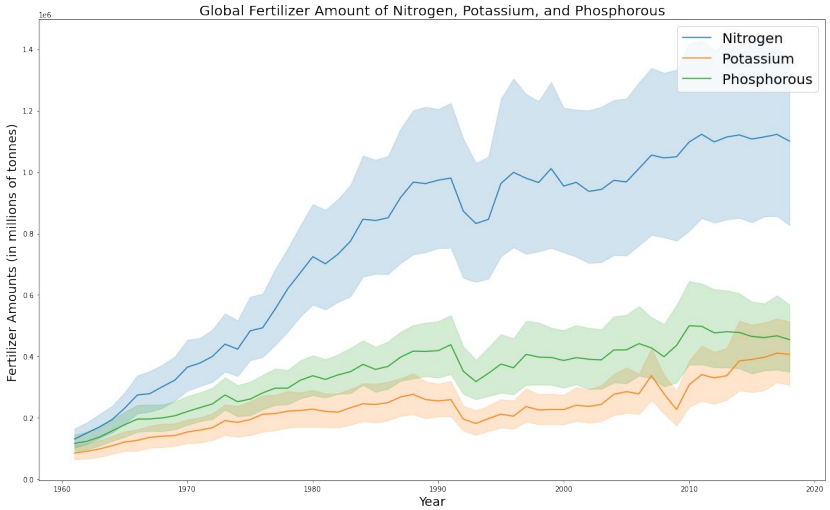
Global Production, Yield and Area







Global Production, Yield and Area



Model Type	Train Score	Test Score	Train Score	Test Score
Multiple Linear Regression	0.6127	0.6105	0.7149	0.7149
KNN Regressors	0.8569	0.7759	0.8569	0.7759
Decision Tree	1.0	0.8610	1.0	0.8987
Bagging	0.9854	0.9229	0.9898	0.9428
Random Forest	0.9899	0.9305	0.9928	0.9517
Ada Boost	**0.9998**	**0.9364**	**0.9999**	**0.9600**
Gradient Boost	0.7672	0.7560	0.7953	0.7883

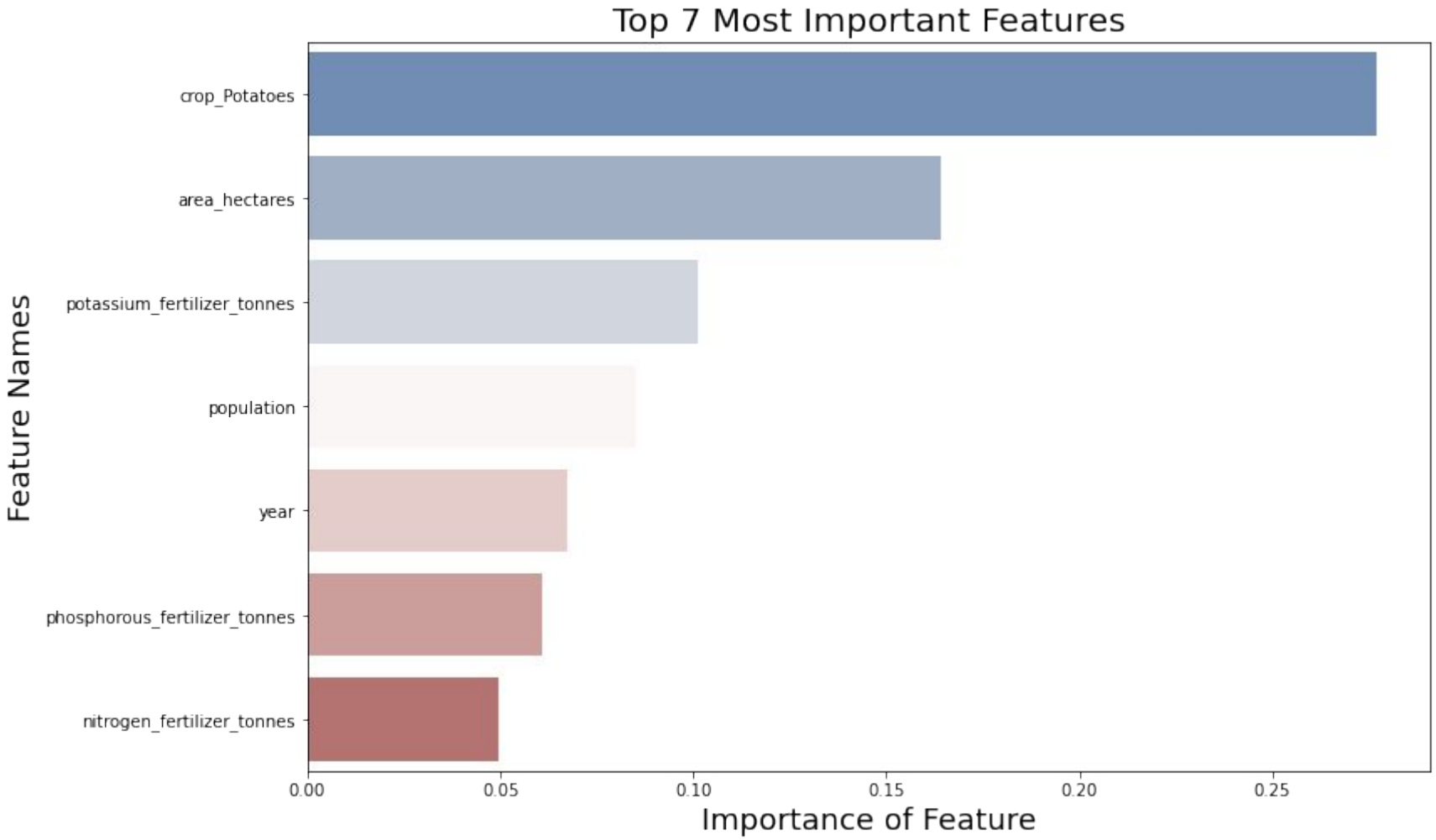
Models Tested

Set 1: Individualized Crop and Continent (19 Features)

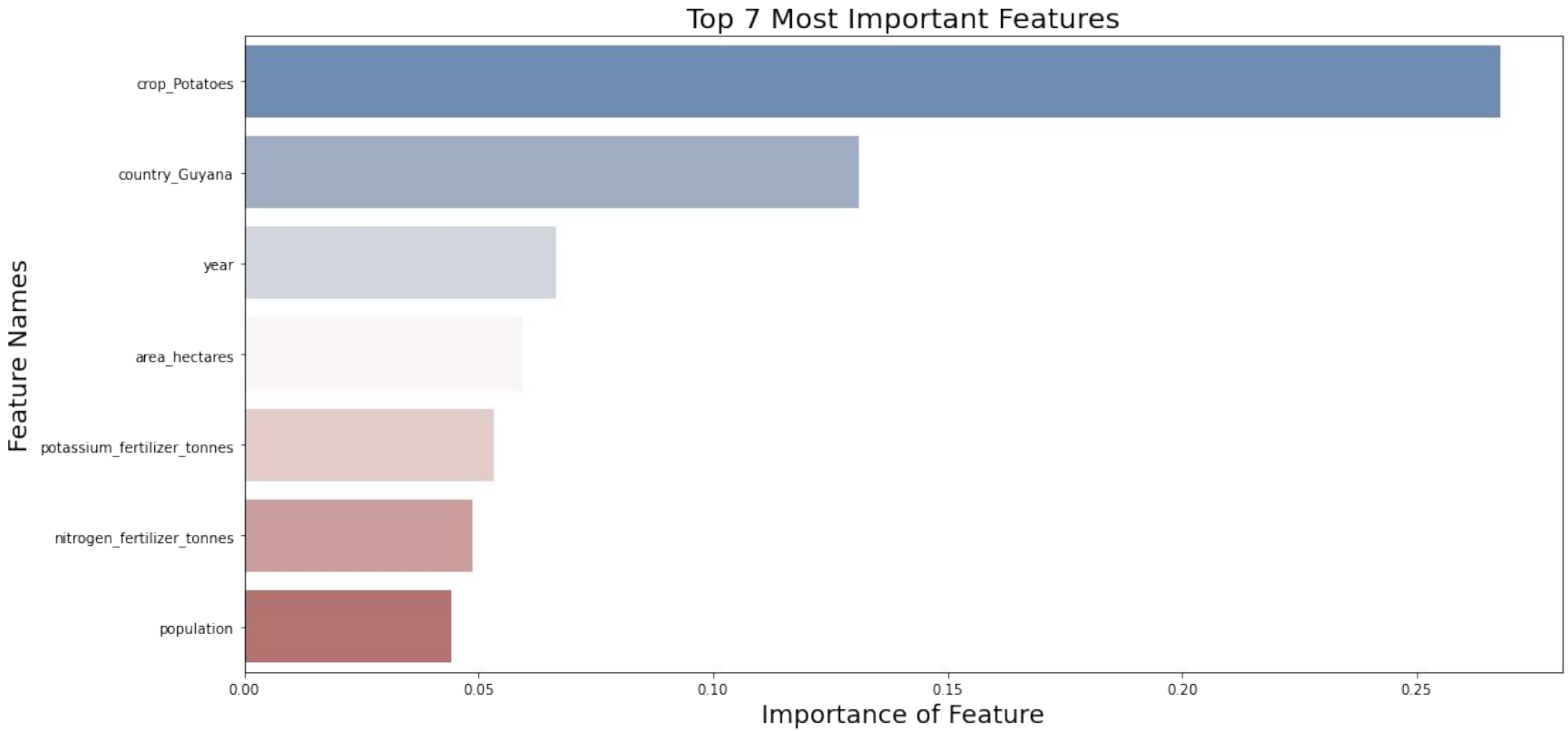
Set 2: Individualized Crop, Country, and Continent (189 Features)

Ada-Boost Model was the winner!

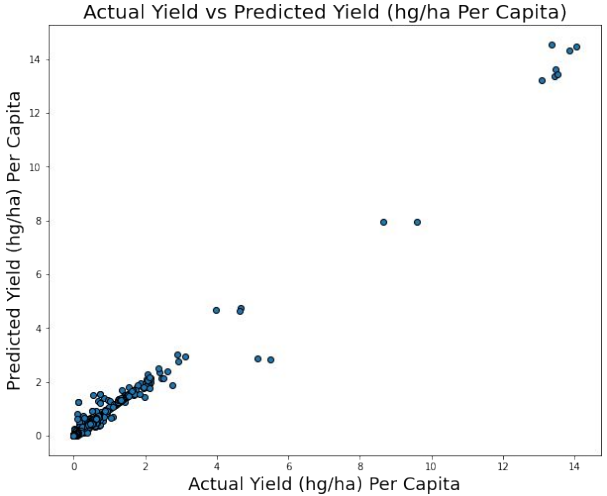
Top Features for Set 1

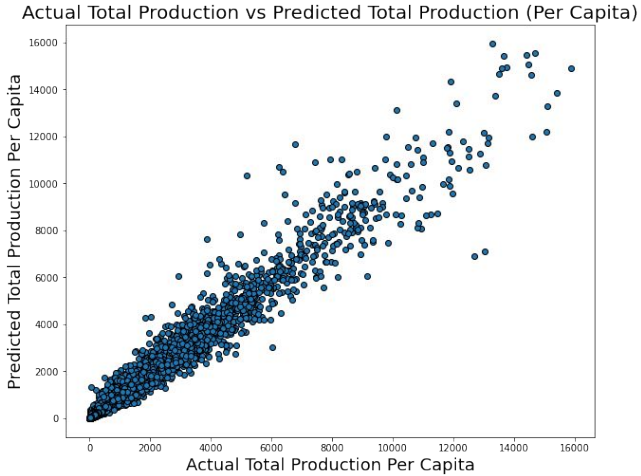


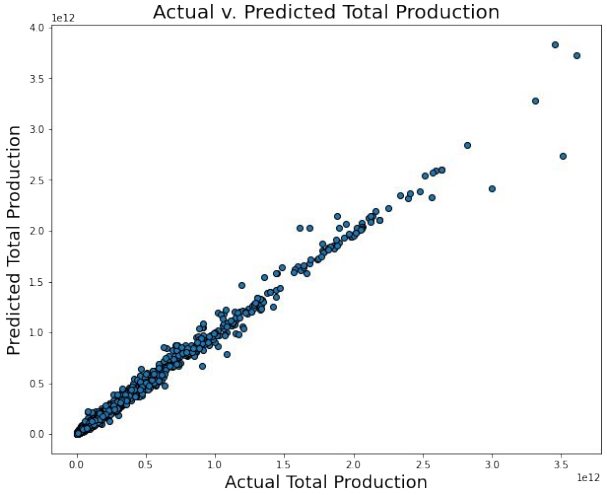
Top Features for Set 2

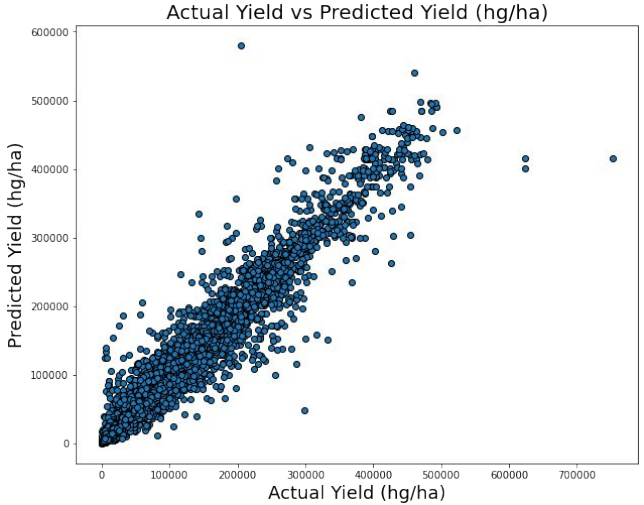


Top Features for Set 1









You Reap What You Sow, Is That So?

A machine learning model has value in predicting crop yield and total production

Our models can successfully isolate the most important factors for predicting crop yield

Crop Yield is generally increasing for all major crops, even while harvested area decreases

Crop yield will need to be considered with other types of metrics (crop yield / capita, total production, total production per capita) to get a fuller picture of the global hunger crisis

More agronomical data will be necessary to correctly predict each single crop locally

sources:

- World Food Program Statistics: https://www.wfpusa.org/
- Food & Agriculture Organization (UN): https://www.fao.org/faostat/