

Project A9: Crop Production Analysis

Katarina Kuusk, Tsienza Smith Steven, Ulrich Ojoh Dioni

Introduction

Successful managerial prospect is based on reliable information validation for decision making. In this prospect, the UN FAO have been recording worldwide data on primary cash crops production by country by year to ensure sustainable food security for all. It is in our prime objective with the example of apple production in this project to reveal the world's major producers, correlate the dependence between overall yield and producers' market price and try to predict on future production and price variation. These analysis could shape our appreciation on the world's leadership strategies on tackling production challenges.

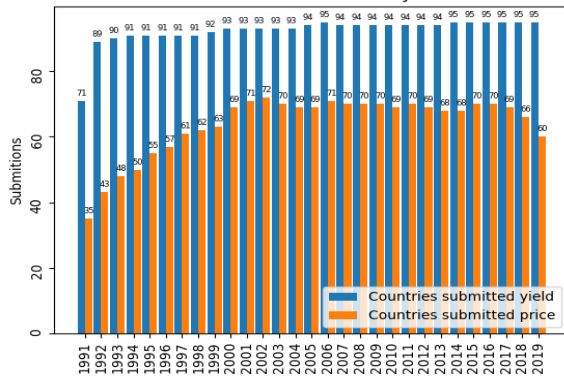
Getting the Data

Our dataset is downloaded from the FAOSTAT website that contains an extensive database of crop production submitted by the UNO member states. Due to the availability of entries for production per country per yield, and to producers' market price for all target features, we had to restraint our dataset collection from the years 1991 to 2019.

Link to the FAOSTAT database <https://www.fao.org/faostat/en/>

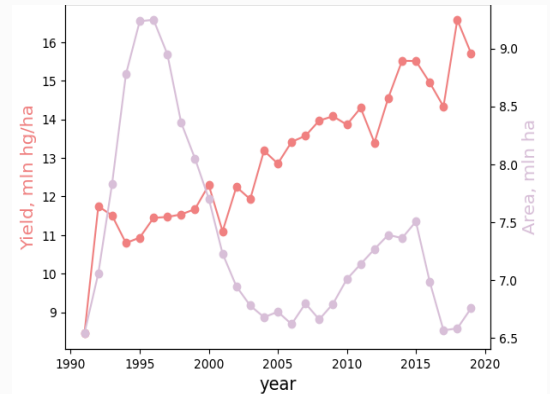
Data Exploration and Visualization for Countries Record from 1991 - 2019

Countries Submitted Data by Years



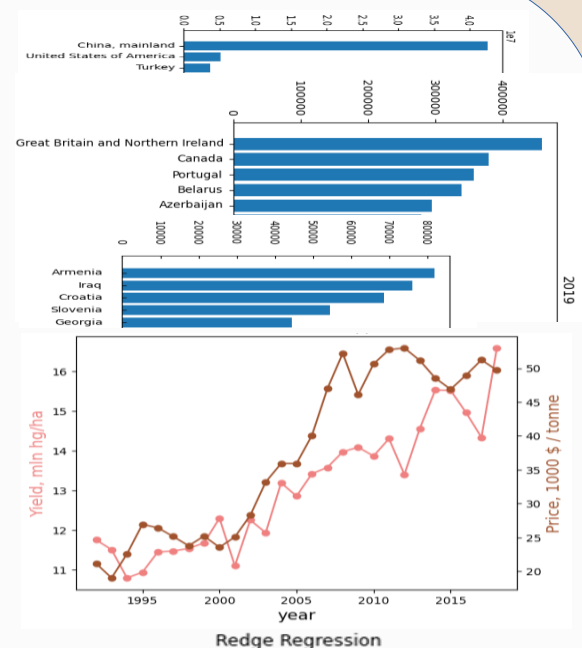
Left: we highlighted on a first approach the viability of the data by countries in search for frequent patterns to depict correlations with various plots and dropping values.

Right: we found that it took less than a decade starting 1997 for countries to critically overthrow the relative link between yields to harvested area. This could be due to the high integration of new technologies in farming practices.



Key Analysis

- We sorted three groups of apple producing countries lead respectively by China, the UK and Armenia in 2019.
- Hypothetically, yield is evaluated from area under the crop harvested area and thus some further analysis showed that the producer price was increasing with yield over the years regardless of missing values.
- We tried different sets of features and performed feature selection by lasso regression with a coefficient of determination of 0.74. We also did simple linear regression. Yet, we are dealing with complex phenomena that may include environmental and economic variables in the process making models' coefficients for countries predicted more significant than yield and harvested area in determination of producer price values.
- The overall yield and harvested area under the crop data is regular whereas producer price is submitted by less number of countries. We interpolated data by linear filling in forward direction if values before and after were present. From the number of submissions by countries we've visualized, the USSR had a fairest share in 1991 and saw its position overtaken by China in 1992.



Conclusion

We couldn't find strong correlation between prices per yield suggesting that would involve more complex features which are yet to determine with further research.

