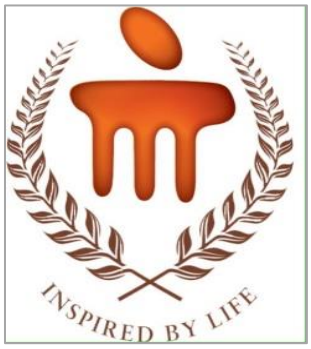
School of Information Sciences

(A Constituent Institute of Manipal University)



Analysis and Visualization of Indian Agriculture Crop Statistics

A Project Report Submitted By

Carol Steffi Dcunha **171046001**
Sushmitha Shetty **171046037**

Under the Guidance of

Mr. Deepak Rao

Assistant Professor

First Semester Master of Engineering
in
Big Data and Data Analytics

Index

Abstract

The importance of carrying out effective and sustainable agriculture is getting increasingly obvious. Modern computers and the internet have made it much easier to make graphics out of tabular data and give these graphics the qualities of interactivity through data visualization. Agriculture yield data is used to analyze and improve the crop yield and represent in the form of a graphs through data visualization technique. The visualization methods facilitates the display of large volumes of data and allows data users to extract information from the statistics.

Chapter 1

Introduction

The Indian economy is heavily dependent on agriculture and its intensive subset, horticulture. However, the usage of latest technologies to these sectors is limited. In last few years, both agriculture and horticulture have been affected by various factors like lack of knowledge of crops, cost estimate, yield and production. The project is based on Big Data analytics of agricultural and horticultural data that includes average yield, cost estimate, land utilization and production of various principal crops. Big Data provides an efficient methodology that enables to gain insights from data and make efficient analytical decisions. The project deals with analysis, prediction and visualization of different categories of data for the better understanding of agricultural statistics in India.

1.1 Scope

The project will be a boon to the Indian Agricultural sector in understanding the various aspects that plays role in agricultural productivity. It can be used for planning agricultural policies. Moreover, it will help the farmers such that they can do agriculture more smartly in a much better calculated way.

1.2 Definitions and Abbreviations

Analytics:

The systematic computational analysis of data or statistics.

Visualization:

The representation of an object, situation or set of information as a chart or other image.

Data extraction:

The act or process of retrieving data out of (usually unstructured or poorly structured) data sources for further data processing or data storage.

1.3 Objective

- To analyze agriculture data like average yield, land utilization and production of principal crops in agriculture and horticulture.
- To analyze area and production of horticulture data.
- To visualize data in graphical format and infer relationships among the data.

Chapter 2

Specifications

2.1 Specifications of the Project

The project uses the dataset related to following attributes:

- Agriculture
    - Area
    - Land utilization
    - Yield
    - Production
    - Cost

- Horticulture
    - Area
    - Production

Chapter 3

Design

3.1 Operating Environment

- Ubuntu 16.04

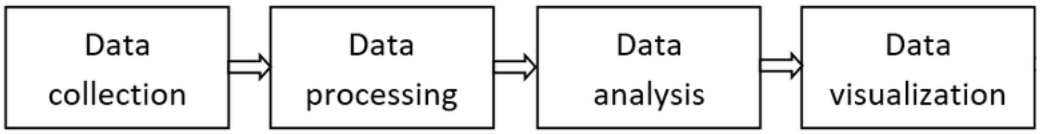
3.2 Software

- Python
- Pandas
- Numpy
- Matplotlib
- Basemap

3.3 Hardware

- 64-bit machine
- 2GB RAM or above
- Hard Disk Capacity: 1GB
- Processor Intel core i5/i7

3.4 Block Diagram



Dataset collection:

This phase deals with collecting the required datasets based on the factors that need to be analyzed in the project. The dataset collected with respect to agriculture and horticulture includes land utilization, area and production, yield and cost estimate. Data is obtained from the government website of Ministry of Statistics and Programme Implementation

Data processing:

The datasets collected may be in different formats. In this phase, the required attributes are extracted from the datasets into structured formatted files. The pre-processed data helps in more accurate prediction of the data. Data is extracted using Python and Dataframes of Pandas.

Data analysis:

This phase explores the data in a number of different ways in order to determine useful information, find correlations, suggest conclusions, and support decision-making.

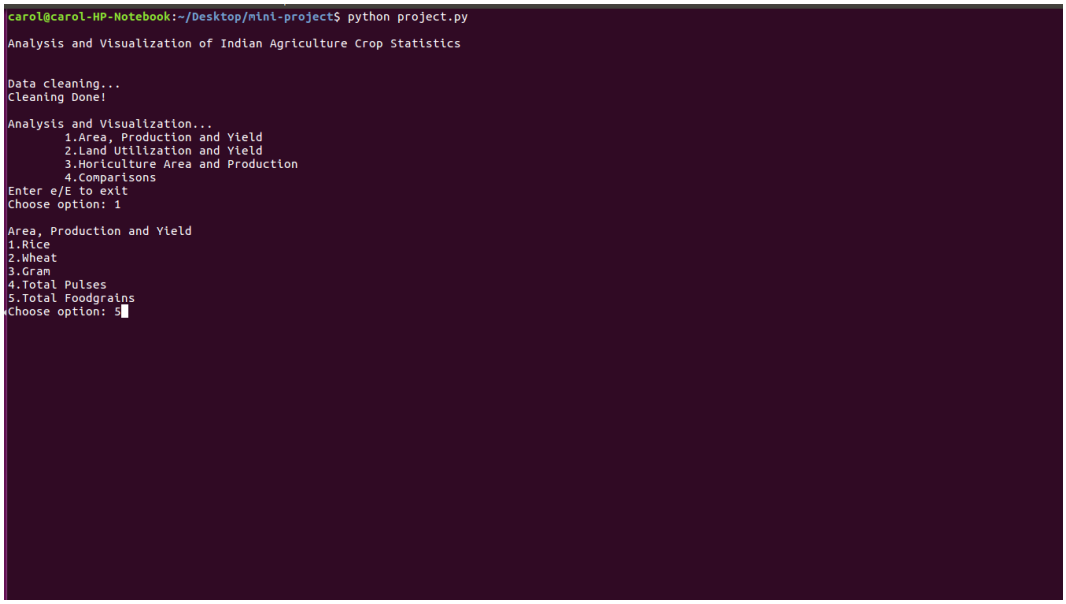
Data visualization:

The next phase in the project is visually representing the classified data in terms of charts and graphs. Data is visualized using matplotlib package in Python.

Chapter 4

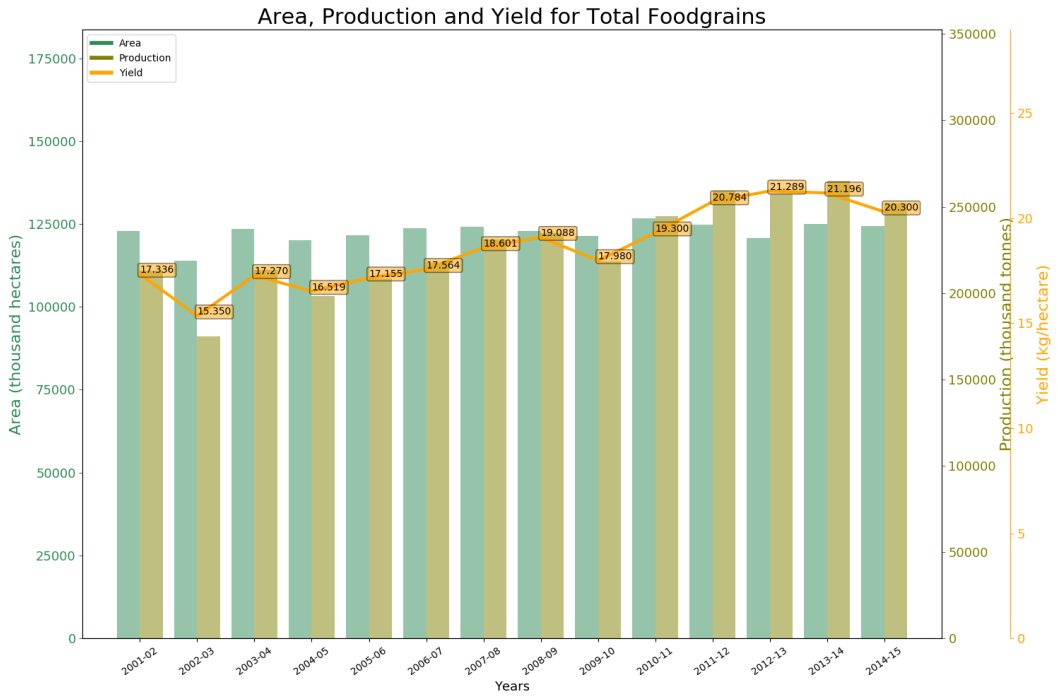
Results

The following figure shows the user interface of the project. Here the user is given the option to select a particular visualization.



Choice 1: Area, Production and Yield

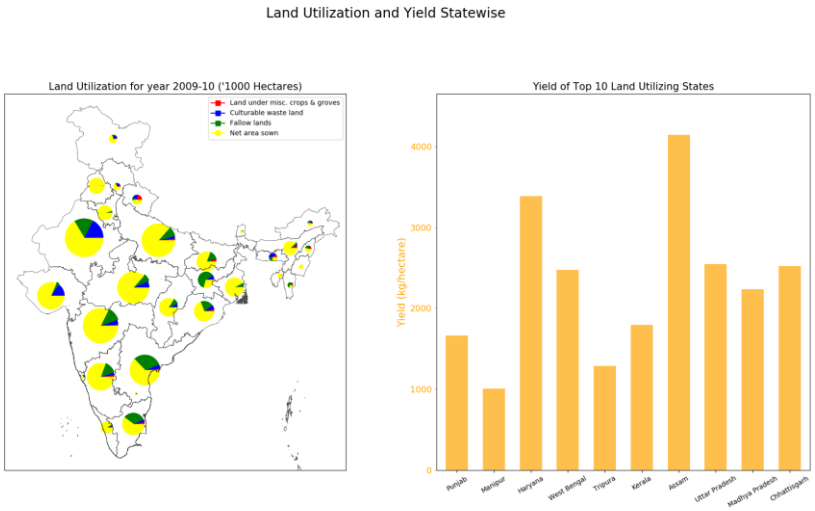
This visualization shows the time series plot of variation of yield with respect to area and production for a particular crop.



Conclusion: With the drop in production in the years 2002 and 2004, there is a drop in yield. Subsequently, from year 2010, there is a steady increase in production as well as yield. Hence, it can be inferred that the yield depends on the production. Also, the variation of production suggest that there are factors other than area that influence it.

Choice 2: Land Utilization and Yield

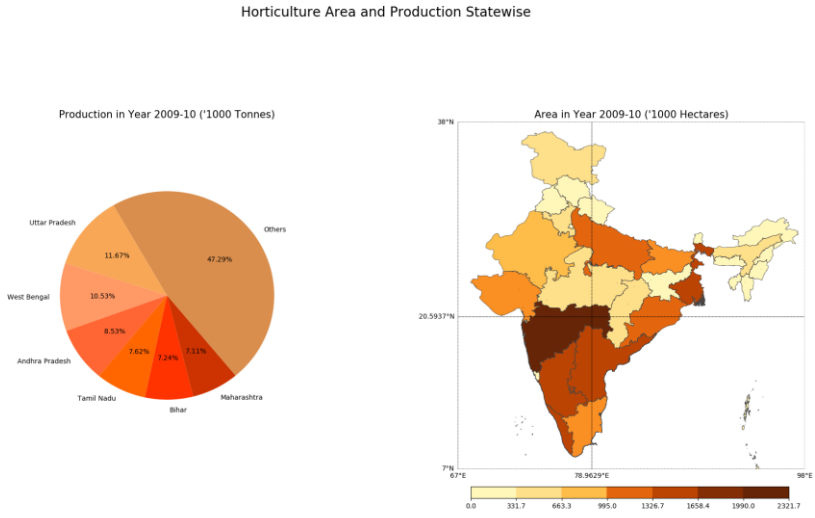
This visualization shows the graduated symbol map for land utilization in Indian states for a particular year. Besides, it displays the yield of top ten land utilizing states in terms of the net sown area.



Conclusion: The high land utilization in Assam and Haryana has resulted in high yield for the year 2009. Whereas, the yield for Manipur is very low. This suggests that land utilization has an effect on yield along with various other factors.

Choice 3: Horticulture Area and Production

This is a choropleth map visualization of horticulture area in Indian states for a particular year. It also includes a pie chart of top six horticulture producing states.

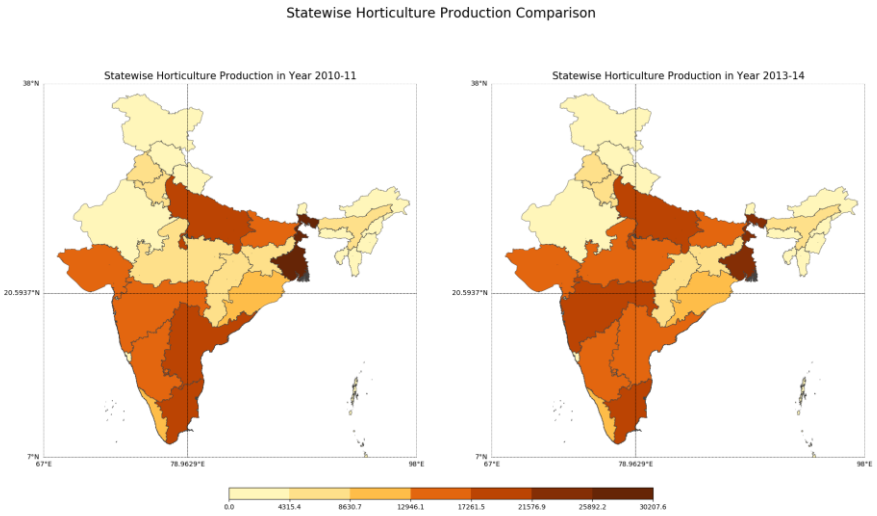


Conclusion: Here Maharashtra provided a large area for horticulture in year 2009. On the same hand, it was among the top six horticulture producing states. This even applies to other states like

Tamil Nadu and Andhra Pradesh. Whereas, even though Bihar provided less area, it had a higher produce. This suggests an indirect relationship between area and production.

Choice 4: Comparisons

These visualizations provide a comparison for two years, between horticulture areas, productions as well as agriculture land utilization.



Conclusion: The transition in colors for the states depict the increase or decrease in horticulture production in the years 2010 and 2013.

4.1 Cumulative Conclusion

The visualizations obtained suggest that:

- Area and production are indirectly dependent
- Yield is dependent on area, production and land utilization, but some trends indicate the influence of other factors

Chapter 5

Scope for future work

There are other factors that affect the productivity of agriculture and horticulture. Some of them include production cost, irrigation techniques, weather conditions, soil quality and even industrialization. Hence the relationship of these factors with respect to productivity can be analyzed through visualizations. Subsequently, predictions can be made to determine future trends.

Chapter 6

References

6.1 Bibliography

[1] K. Ravisankar, K. Sidhardha, Prabadevi B,"Analysis of Agricultural Data Using Big Data Analytics",Journal of Chemical and Pharmaceutical Sciences,July - September 2017

[2] Purva Grover, Rahul Johari, "PAID:Predictive Agriculture Analysis of Data Integration in India", 2016 International Conference on Computing for Sustainable Global Development (INDIACom),IEEE, 2016

[3] Niketa Gandhi, Leisa J. Armstrong, "A review of the application of data mining techniques for decision making in agriculture",2016 2nd International Conference on Contemporary Computing and Informatics (ic3i),IEEE,2016

6.2 Webliography

- http://www.mospi.gov.in/statistical-year- book-india/2017/177
- http://www.mospi.gov.in/statistical-year- book-india/2017/178
- http://aims.fao.org/activity/blog/big-data- unlocking-future- agriculture