Crop Production Data Analysis and Trends Visualization of India

Nilotpal Kapri[1]

1. nilotpal623401@gmail.com

Abstract: *This report presents the analysis and visualization of different types of crop production in India. Now days due to technology revolution, it is becoming the best solution for all the problems in agricultural product processing and distribution. To fulfill this demand data science and different ML algorithms are very much important now. In this paper analyzing crop production data and visualizing yearly trends of the same is focused. All the analysis are presented in very basic level, any advance level analysis is out of scope for this project.*

Keywords: Data Science, Agriculture, Crop Production, Data Processing, Data Visualization, Jupyter Lab.

1. Introduction: Agriculture plays a vital role in the Indian economy. It is the backbone of the country. Most of it's population (about 60 per cent) is depends on Agriculture. Nearly 16 per cent of overall GDP of India (approx. Rs 19.48 lakh crore or US$ 276.37 billion) is contributed by agriculture, forestry and fisheries[1]. The weather conditions and soils conditions allow for crops in India is perfectly suited for it.

The scope of this project is to mainly proposing a business problem and finding the possible solution. So, here is the proposal for which I've worked for the entire project- 'To provide the necessary information visualization, I'll analyse the crop production datas of india in different region and map them using Geocoder and Folium library.'

So, extension of this project will be very useful for Startup Business in the field of agriculture. From this project one can identify the cluster of production of any crops that are produced during 1997 to 2015. Also in addition to this at the end of this project one can observe the trends of production of any crops for any state of India. Although discussion is limited for this project still many problem can be illuminate using the results.

2. Background: The Indian food industry is poised for huge growth, increasing its contribution to world food trade every year due to its immense potential for value addition, particularly within the food processing industry. Indian food and grocery market is the world's 6th largest, with retail contributing 70 per cent of the sales. The Indian food processing industry accounts for 32 per cent of the country's total food market, one of the largest industries in India and is ranked 5th in terms of production, consumption, export and expected growth.

India's production of food grains has been increasing every year, and India is among the top producers of several crops such as wheat, rice, pulses, sugarcane and cotton. It is the highest producer of milk and second highest producer of fruits and vegetables.

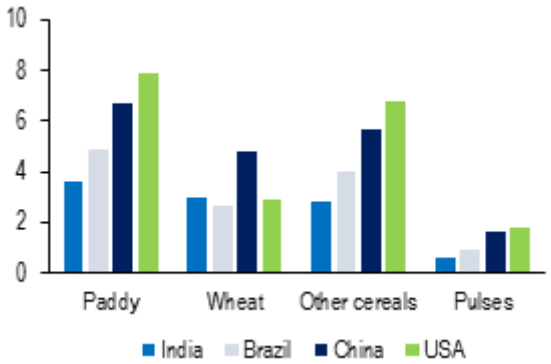

Figure 1: Yield in different countries (tonne/ha)[2]

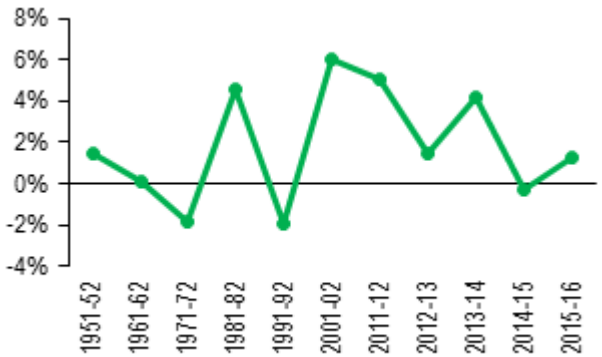

Figure 2: Agricultural growth of India(in %)[2]

Key issues affecting agricultural productivity include the decreasing sizes of agricultural land holdings, continued dependence on the monsoon, inadequate access to irrigation, imbalanced use of soil nutrients resulting in loss of fertility of soil, uneven access to modern technology in different parts of the country, lack of access to formal agricultural credit, limited procurement of food grains by government agencies, and failure to provide remunerative prices to farmers.

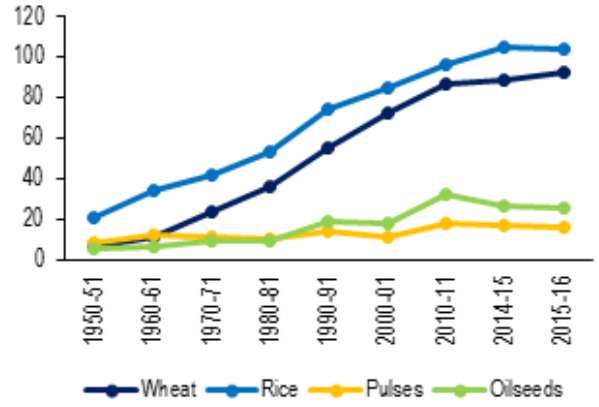

Figure 3: Agricultural production (million tonnes)[2]

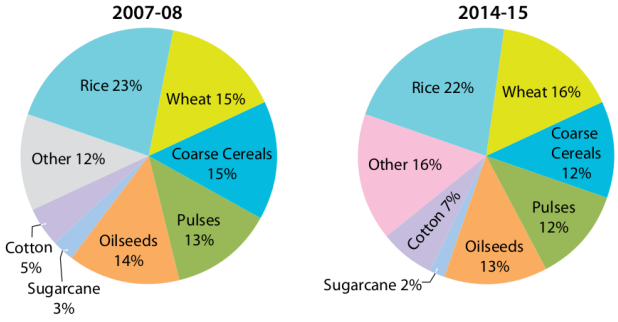

Fig 4: Changes in Share of Area Under Major Crops[4]

3. Data acquisition and Processing: For this project I'll use Kaggle crop production dataset, which contains more than 2,00,000 data rows with detailed information of crop production in India from 1997 to 2015 with respect to defferent districts and States. Source url- https://www.kaggle.com/divyosmi2009/crop-production-in-india-statevise?select=crop_production.csv

This dataset contains 7 unique columns- 1. State Name, 2. District Name, 3. Crop Year, 4. Season, 5. Crop Name, 6. Area of Production, 7. Quantity of Production. From this I used District and State Name to fetch the grographical coordinates (latitude, longitude) of all the locations using **geocoders library**.

4. Methodology: As the aim of this work is to map all the required datas so, at first the project will collect some inputs to process with customization. It will take Years of data, Seasons and Name of the crops to be processed. Based on these inputs it will create a new dataframe which will include all the location datas along with the base datas from the main dataframe. After preparing this new dataset it will proceed with the Python **Folium library** to point out India's map using location coordinates and mark all the datas on it as per user choice, to visualize seperately for each year.

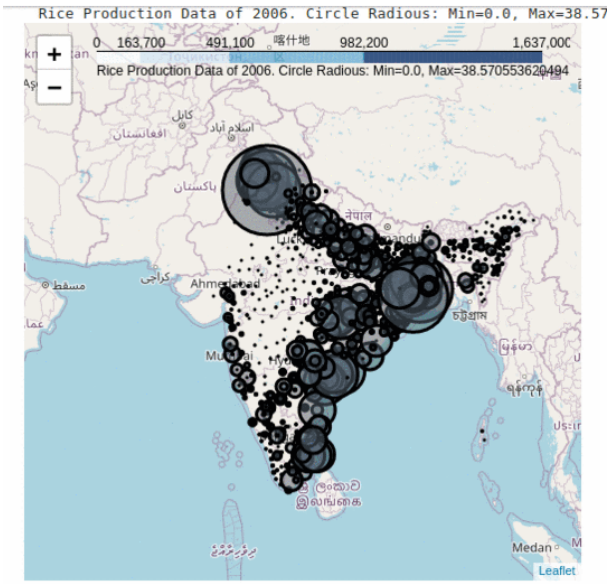

Figure 5: Rice Production Data of 2006 to 2014.

Here, in this report only Rice production data visualization is given from the year 2006 to 2014 for all the Seasons(summed) in GIF format below.

Although I took only one crop for presenting in the report, you can choose as many as you want but as the map will become very much conjusted, it is limited to show upto five different crops with different color, after that all other crops will mark as red. To avoid this confusion It is recomended to choose maximum five crops. As per the observation very few datas are available for 2015, so in most of the cases it is recomended not to take 2015 results in consideration.

Now in the second part of this project again it will take some user input like the state and crops for which trends to be visualized. As no location data is required for this work It will directly select required datas from the main dtataframe and make all the temporary required dataframe by looping through the user choices. And finaly all the district datas will be summed up for the selected state seperately for each year and processed for plotting these datas using python **pyplot** and **seaborn library**.

As like the previous in this case also I'll only show rice production trends of the maximum producer state during 1997 to 2014.

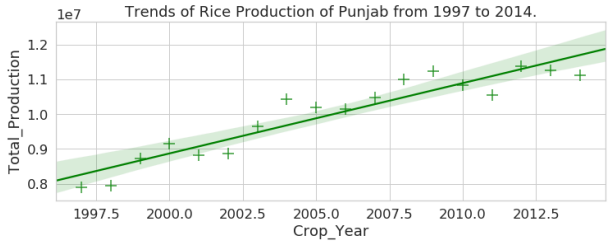

Figure 6: Rice Production Trends of Punjab from 1997 to 2014.

The above plot shows the quantity if rice produced in respective year from 1997 to2014. Next we will see area upon which rice produced during this season.

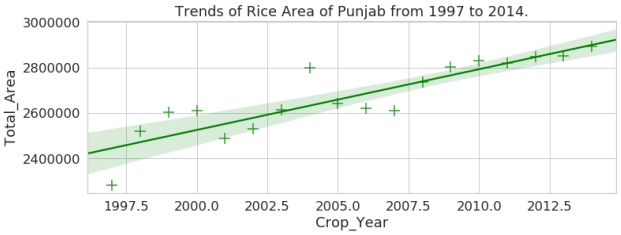

Figure 7: Rice Production Area Trends of Punjab from 1997 to 2014.

Now we can devide this two to see how production per unit area is varing with time.

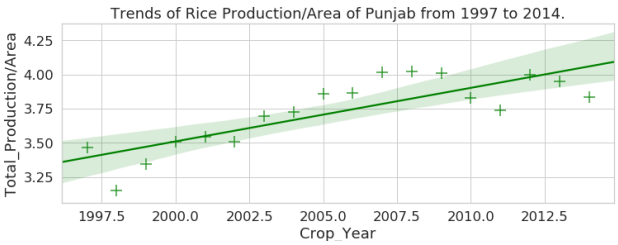

Figure 8: Rice Production per unit Area Trends of Punjab from 1997 to 2014.

And finally as conclution we will compare these two using bar chart to observe the effeciency of production.

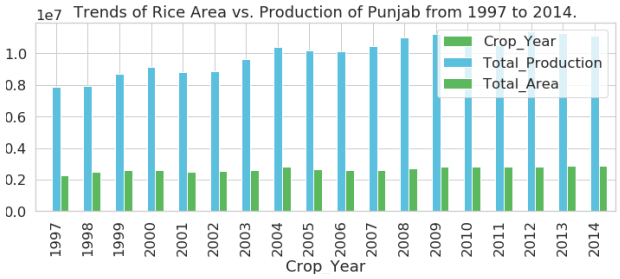

Figure 9: Trends Comparision of Rice with Rice Production Area of Punjab from 1997 to 2014.

Note: Next we can cluster the mapped datas autometically for beeter visualization using K-means clustering algorithm. One more benefits of this will be you can observe how clusters are moving year to year so that you can decide the best location for managing your business and for long run you can also predict this trends for future scope.

5. Results: Although most of the results are already presented with the the methodology discussion, Let's get some inshits from these results.

It is not possible to insert the intractive map here (otherwise you can click on the circles to

get details), but still you can see in the figure 5, Punjub, West Bengal and Andhra Pradesh are the three main rice production cluster for all the time. Gujrat and Madhya Pradesh are on the way of extinction for rice production with time.

From the second observation you can that figure 6 datas are flutuating mostly depended on figure 7. The area is flutuating periodically, so something must be happening with the farmers for which they are loosing their hope or interest towards farming rice for certain years. Here either weather or the pricing of the previous year might be a factor for this behavious. So for getting more insight we need to combine wheather dataset and product pricing dataset along with this which is out of scope for this project.

6. Discussion: Now coming to the discussion let's understand the trends, what is going through behind all the happenings in your business with the flow of time.

1st Part: Ovserve the figure 5 gif carefully, you can see how clusters are moving from different region. One more important observation is that in some of the states rice production is becoming zero gradually, so for business management it can be a very big factor as transportation charge can increase a lot after some years due to the unavailability of products in these regions.

2nd part: Ovserve figure 6, 7 and 8, you can see that overall production of Punjub (maximum rice producer state) is increasing drastically. And one big cluster is always present there, so for rice processing it can be the most preferable location in this region. Also from figure 9 you can see that efficiency i.e, production per unit area is decreasing gradually, so price of the rice in this region shall increase with time naturally. So, all this factors can take a major role in your future business outcome.

7. Conclusion: One of the best advantage of this project is that you can easily interact with the data directly and observe the trends of as many product and region as you wish. So, as discussed earlier this can play a major role for choosing the sight and cost estimation for your business. Although as far I analyzed is not sufficient enough for a big business, but a slight extension with this project can make it a perfect efficiency booster of your business in this field.

8. References:

[1] 'Agriculture in India: Information About Indian Agriculture & Its Importance', India Brand Equity Foundation, Ministry of Commerce & Industry, Govt. of India. url- *https://www.ibef.org/industry/agriculture-india.aspx*

[2] 'State of Agriculture in India', PRS Legislative Research ("PRS"). url- *https://www.prsindia.org/policy/discussion-papers/state-agriculture-india*

[3] 'Agricultural Marketing', Department of Agriculture, Co-operation and Farmers Welfare, Ministry of Agriculture and Farmers Welfare. *url- http://www.agricoop.nic.in/divisiontype/agricultural-marketing*.

[4] 'Pocket Book of Agricultural Statistics, 2015', Ministry of Agriculture and Farmers Welfare, Government of India. url- *http://eands.dacnet.nic.in/PDF/Pocket-Book2015.pdf*.