Visualization Project on Agricultural Crop Production in India

Name: Meghna Mondal Roll No.: MDS202123

10/01/2022

Introduction:

India's agriculture is composed of many crops, with the foremost food staples being rice and wheat. Indian farmers also grow pulses, potatoes, sugarcane, oilseeds, and such non-food items as cotton, tea, coffee, rubber, and jute. Despite the overwhelming size of the agricultural sector, however, yields per hectare of crops in India are generally low compared to international standards.

So, here we wish to analyze more about the crop cultivation, cost invested in cultivation in different parts of our country. Now we consider 2 datasets from the Kaggle dataset on Agricultural Crop Production in India (source: https://www.kaggle.com/srinivas1/agricuture-crops-production-in-india/version/1?select=datafile+%282%29.csv) namely datafile (1).csv and datafile (2).csv containing different variables.
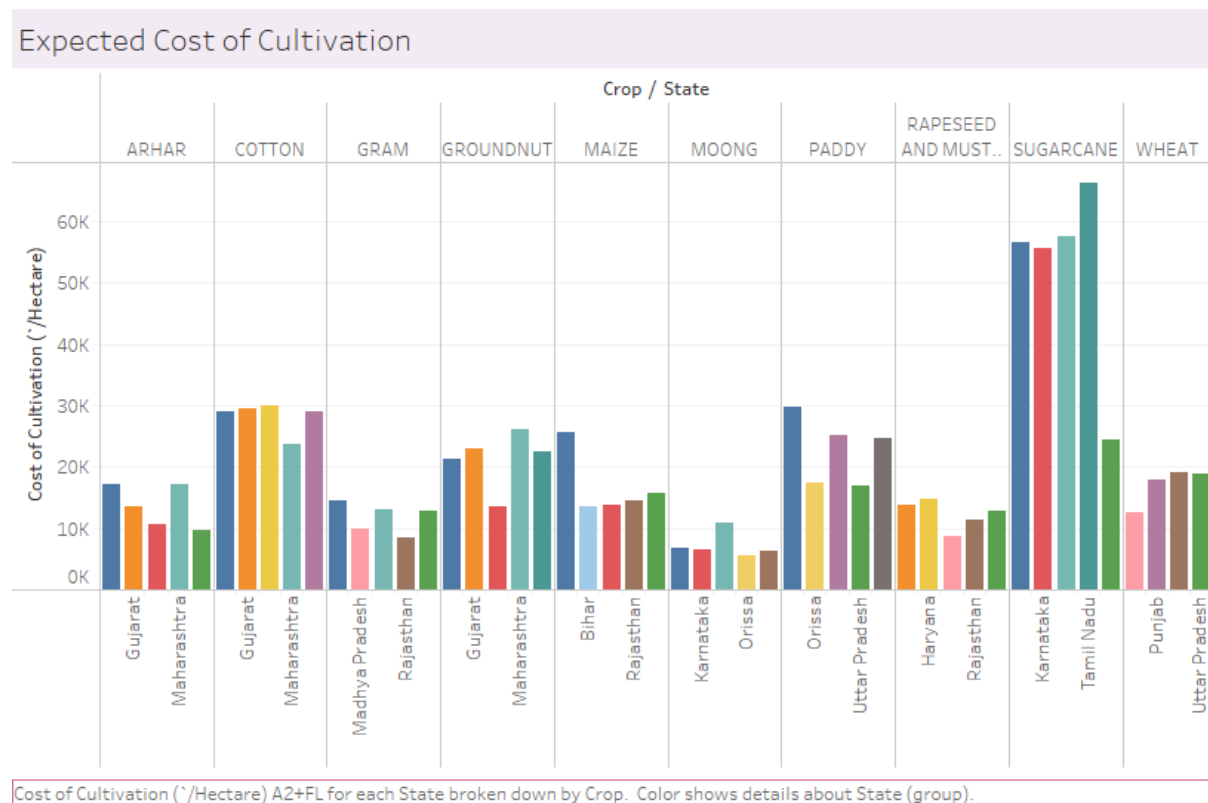
Methodology:

In datafile(1) we have production cost, cultivation cost and yield of some specified crops in particular states. We have a series of data for a nominal attribute. So, we will use Bar Diagrams for analyzing those Cross-Sectional Data.

In datafile(2) we have amount of production, area of cultivation and yield of some specified crops for five years. We have a series of data for a nominal attribute for varying time. So, we will use Bar Diagrams for analyzing those Panel Data.
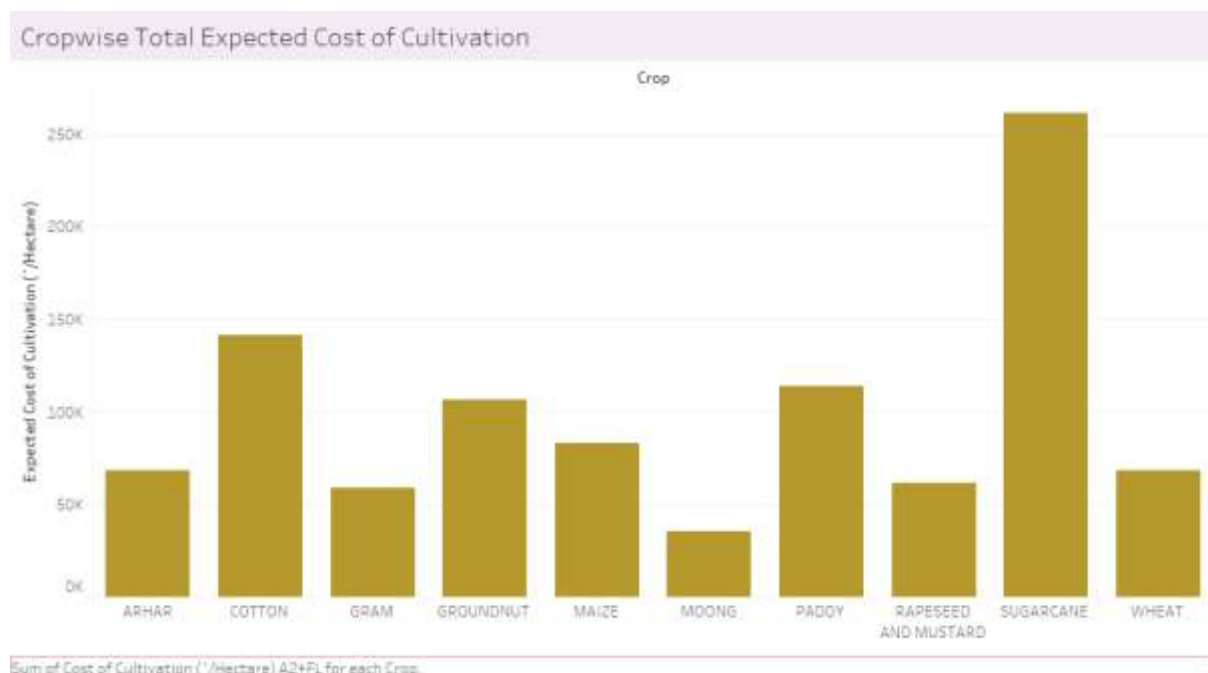
For the first dataset we have four continuous variables (Expected Cost of Cultivation, Cost of Cultivation, Cost of Production, Total Yield) with respect to two attributes. So, for each variable we get two Bar plots one with state wise total another with crop wise total. For e.g., the above graph is Bar Plot with Crop wise total of Expected Cost of Cultivation. With this, we can get an idea of the behaviors of the concerned variable over the attributes states and crops.

Now for analysis of the variables taking account of both the attributes – Crops and States has been done also.

Cost of Cultivation (`/Hectare) A2+FL for each State broken down by Crop. Color shows details about State (group).

The above graph is the analysis of Expected cost of Cultivation with respect to both Crops and States. It gives us an idea about which state is expected to invest how much in each crop cultivation. This same analysis can be done with the other three variables.

For the second dataset, we have three (Production, Area, Yield) sets of variables for five timepoints (2006-07, 2007-08, 2008-09, 2009-10, 2010-11) i.e., 15 continuous variables for 55 types of Crops.

For comparative analysis of crops, we can have 15 bar plots for each variable to see the behavioral pattern of the concerned variable against different crops.

For e.g., the following graph is the bar plot of Area used for cultivation for the year 2008-09 for different crops. Here, we can observe for the crop Soybean maximum area has been used in the cultivation year 2008-09.

Cropwise Area 2008-09

Area 2008-09 for each Crop. Color shows details about Crop.
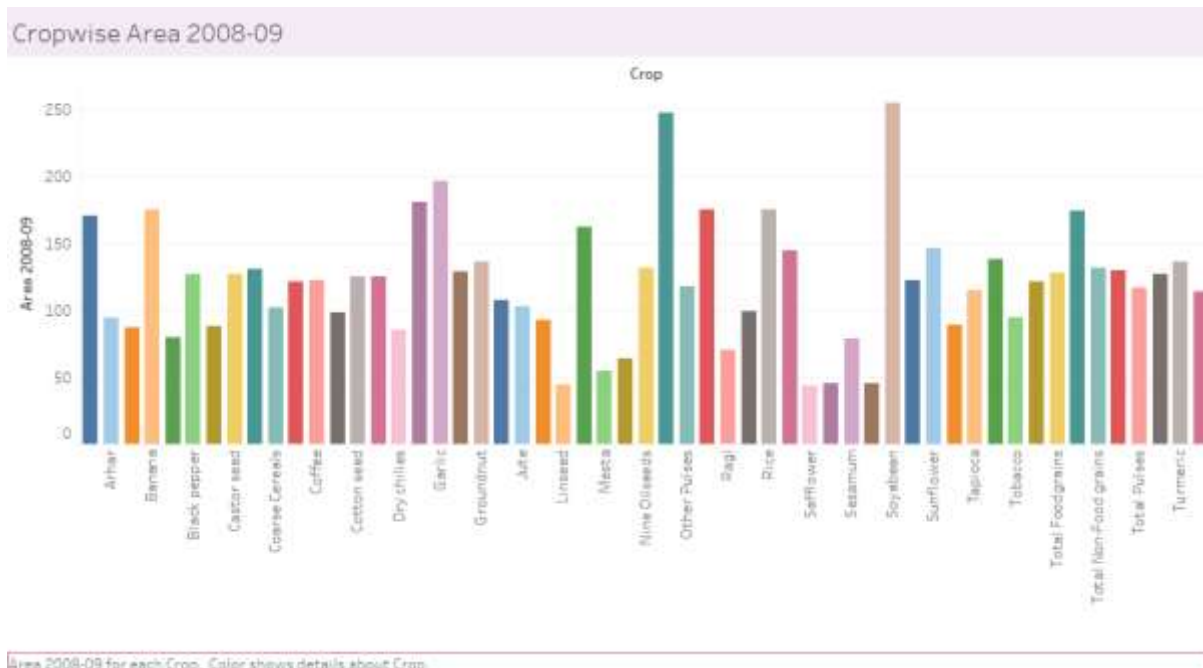
Now, for the analysis of the three variables (Production, Area, Yield) against both time and crops together we get a graph like this where we get a side- by- side bar diagram with set of five bars for each crop. So, for each crop we get an idea of the behavioral pattern of the concerned variable for the five-year span.

For e.g., following is the five-year area analysis for each crop. Here, we observe the area used for cultivation for Banana, Coffee, Cotton Seed, Cotton, Gram, Sesamum etc. is consistently increasing over the five-year span. Eventually we can infer this type of information from this graph.

Conclusion:

The project has been done in three components. First, we have used the ggplot2 package in R for analysis of this variable in R Markdown file and got a pdf output out of it. Secondly, we have created a dashboard using the R-Shiny App which shows the graphs generated in the R-Markdown file. Next, we generate those graphs using Tableau Software. The links to these files are given in the References.

From the univariate and bivariate analysis of the above datasets we get an idea about

in which state, how much amount of crop are cultivated in how much area.

For e.g., we get to know the yield and cost of cultivation per Hectare are comparatively higher for sugarcane i.e., it is being produced in a large amount.

Again, in Andhra Pradesh, Maharashtra, Karnataka, Tamil Nadu, Uttar Pradesh yield produced are much higher than the other states.

In the context of cost of cultivation Andhra Pradesh shows much higher values than the other states.

We observe that Production and Yield of Total Spices is way too higher comparative to other crops in India.

Eventually we can infer this type of information from each graph.

We also have computed the correlation matrix for the variables considered. If the correlation coefficient comes out to be near 1 two variables said to be positively correlated, or if the correlation coefficient comes out to be near -1 two variables said to be negatively correlated and if it comes out to be near 0, we can expect the two variables to be independent on each other.

For e.g., we observe for all the five years the crop production and yield are highly positively correlated but the area and yield are area moderately related.

In R-Shiny Dashboard we have three tabs. In first tab titled Data Description contains the data description of both datasets. In second and third tab we have the univariate and multivariate analysis of the first and second datasets respectively.

In Tableau, first we need to import the datasets and then just by selecting the columns we get the plots needed for visualising the data. The figures shown here in the report are Tableau graph screenshots. These graphs have been also generated in the R-Markdown file and R-Shiny Dashboard.

References:

Link to rmd file and pdf output:

https://drive.google.com/file/d/1XTYPs9X4N0Lmswf928YnfSySDaF_ewrw/view?usp=sharing

Link to the Dashboard:

https://meghna.shinyapps.io/meghna/?_ga=2.103048328.948707756.1641293955-183824976.1637960154

Link to Presentation Video:

https://youtu.be/fDAhDOqN340

Link to Tableau File:

https://drive.google.com/file/d/10YuBggjDf5dkY50Evu8NK_e1suUZvN2a/view?usp=sharing