ESTIMATE THE CROP YIELD USING DATA ANALYTICS

TEAMID: PNT2022TMID20305

TEAM LEADER - MARYS AXSEELIAS FELCITAS.P

TEAM MEMBER 1 - ARULSELVI.L

TEAM MEMBER 2 - ARUN.N

TEAM MEMBER 3 - MATHAN ABISHEEK.R

ABSTRACT

Agriculture is important for human survival because it serves the basic need. A well-known fact that the majority of population (≥55%) in India is into agriculture. Due to variations in climatic conditions, there exist bottlenecks for increasing the crop production in India. It has become challenging task to achieve desired targets in Agri based crop yield. Various factors are to be considered which have direct impact on the production, productivity of the crops. Crop yield prediction is one of the important factors in agriculture practices.

Farmers need information regarding crop yield before sowing seeds in their fields to achieve enhanced crop yield. The use of technology in agriculture has increased in recent year and data analytics is one such trend that has penetrated into the agriculture field. The main challenge in using big data in agriculture is identification of effectiveness of big data analytics. Efforts are going on to understand how big data analytics can agriculture productivity.ant factors in agriculture practices. In this project, we use a dashboard to view, analyze, and extract the majority of the findings.

1.INTRODUCTION :

Agriculture forms the basis for food security and hence it is important. In India, majority of the population i.e., above 55% is dependent on agriculture as per the recent information. Agriculture is the field that enables the farmers to grow ideal crops in accordance with the environmental balance. In India, wheat and rice are the major grown crops along with sugarcane, potatoes, oil seeds etc. Farmers also grow non-food items like rubber, cotton, jute etc. More than 70% of the household in the rural area depend on agriculture.

This domain provides employment to more than 60% of the total population and has a contribution to GDP also (about 17%). In the farm output, India ranks second considering the world wide scenario. This is the widest economic sector and has an important role regarding the framework of socioeconomic fabric of India. Farming depends on various factors like climate and economic factors like temperature, irrigation, cultivation, soil, rain fall, pesticide and fertilizers. Historical information regarding crop yield provides major input for companies engaged in this domain. These companies make use of agriculture products as raw

materials, animal feed, paper production and so on. The estimation of production of crop helps these companies in planning supply chain decision like production scheduling. The industries such as fertilizers, seed, agrochemicals and agricultural machinery plan production and activities like marketing based on the estimates of crop yield. Farmers experience was the only way for prediction of crop yield in the past days. Technology penetration into agriculture field has led to automation of the activities like yield estimation, crop health monitoring etc. Crop yield prediction helps the farmers in various ways by providing the record of previous crop yield. This is helpful to government in framing policies related to crops such as crop insurance policies, supply chain operation policies. Knowing what crops has been grown, and how much area of it had been shown historically, combined with the prices at which it could have been sold at the nearest market-place provides the income-growth profile of the farmer.Agriculture sector is struggling to increase the productivity of crop in India.

Monsoon rainfall is the main source of water for more than 60 percent of the crops. Smart agriculture driven by Information Technology is the emerging trend in the research in this area in recent days. One of the areas being explored is the

problem of yield prediction which is a major concern. Data mining techniques are being widely used as a part of solution for crop yield prediction. Various data mining techniques are under evaluation for estimation of crop production of the future years. Data mining is the process in which the hidden patterns are discovered using analysis of large data sets. The data mining and data analytics techniques use artificial intelligence, statistics, machine learning and database system. In data mining, unsupervised and supervised methods are being used.

BRAIN STORMING AND IDEA PRIOTIRIZATION:

1

Define your problem statement

What problem are you trying to solve? Frame your problem as a How Might We statement. This will be the focus of your brainstorm.

⏱ **5 minutes**

PROBLEM

Estimation of crop yield using data analytics

Key rules of brainstorming

To run an smooth and productive session

😊 Stay in topic.　　💡 Encourage wild ideas.

😐 Defer judgment.　　👂 Listen to others.

🗂 Go for volume.　　👁 If possible, be visual.

2

Brainstorm

Write down any ideas that come to mind that address your problem statement.

⏱ 10 minutes

Predicts the crop estimation	Improves the farmers idealogy	Resource management decisions
Understanding of environmental challenges	Causes of yield gaps	Detecting weeds
Monitoring health of crops	Transformation of agriculture technologies	Improving Profits
Data Evaluation of crops	Detecting Weeds	Estimating of enviromental conditions

Improvement of productivity	Estimates the climatic conditions	Soil Quality Measurements
Predicts analytics to future yields	Digital Agriculture	Selection of suitable seeds
Productivity Improment	Analysis of data can be stored	Prediction of crop yield
Causes of yield gaps	Providing the crop resources	Granular data on water cycles

3

Group ideas

Take turns sharing your ideas while clustering similar or related notes as you go. In the last 10 minutes, give each cluster a sentence-like label. If a cluster is bigger than six sticky notes, try and see if you and break it up into smaller sub-groups.

⏱ 20 minutes

Predicts the crop estimation

Productivity Improment

Estimates the climatic conditions

Transformation of agriculture technologies

4

Prioritize

Your team should all be on the same page about what's important moving
forward. Place your ideas on this grid to determine which ideas are important and
which are feasible.

🕐 20 minutes

Importance

If each of these
tasks could get
done without any
difficulty or cost,
which would have
the most positive
impact?

Causes of
yield gaps

Estimates
the climatic
conditions

Predicts the
crop
estimation

Soil Quality
Measurements

Selection of
suitable
seeds

Productivity
Improment

Resource
management
decisions

TIP

Participants can use their
cursors to point at where
sticky notes should go on
the grid. The facilitator can
confirm the spot by using
the laser pointer holding the
H key on the keyboard.

Improving
Profits

Improves
the farmers
idealogy

Feasibility

Regardless of their importance, which tasks are more
feasible than others? (Cost, time, effort, complexity, etc.)

DATASET:

"Estimate The Crop Yield Using Data Analytics" project is based on understanding the crop production of India.Download the dataset from the below link. It has 2,46,092 data points (rows) and 6 features (columns) describing each crop production related details.

Dataset Link : [Dataset](#)

Let's understand the data we're working with and give a brief overview of what each feature represents or should represent

1. State Name - All the Indian Statenames.

2. District Name - Different Districtnames.

3. Crop Year- contains the crop years.

4. Season - Different seasons for cropproduction.

5. Area - Total number of areas covered.

6. Production - production of crops.

UNDERSTAND THE DATASET:

1.   Download the dataset from the given dataset link.
2.   Data Given in csv file is understood.
3.   The data format is as shown in the below image:

LITERATURE SURVEY:

SL NO:	TITLE OF THE PAPER	AUTHOR	YEAR OF PUBLICATION	HIGHLIGHTS
1.	Agriculture Data Analytics in Crop Yield Estimation: A Critical Review	Sagar Bm	2018	The use of technology in agriculture has increased in recent year and data analytics is one such trend that has penetrated into the agriculture field. The present study gives insights on various data analytics methods applied to crop yield prediction and also signifies the important points in the proposed area of research.
2.	The Impact of Data Analytics in Crop Management based on Weather Conditions	Swarupa Rani A	2017	Discussed the application of mathematical model like fuzzy logic designs in optimization of the crop yield, artificial neural networks in validation studies, genetic algorithms designs in accessing the fitness of the model applied, decision trees, and support vector machines to study soil, climate conditions and water regimes related to crop growth and pest management in agriculture.
3.	A Study on Crop Yield Forecasting Using Classification Techniques	R.Sujatha, Dr.P.Isakki Devi	2016	Discuss the importance of comparing previous agricultural data with present to identify optimum condition favor enhanced crop yield. Envisaged the importance of best crop selection depending on the season and the climatic factors which supports enhanced crop yield.
4.	Prediction of Crop Yield using Regression Analysis	V. Sellamand E.Poovammal	2016	Regression analysis was carried out to find the relationship among the parameters i.e Area under Cultivation (AUC), Annual Rainfall (AR) and Food Price

			Index (FPI) which influences the final crop yield and
5.	Data requirements for reliable crop yield simulations and yield - gap analysis	Patricio Grassinia, Lenny G.J. van Bussel, Justin Van Warta, Joost Wolf,Lieven Claessens, d, Haishun Yanga, Hendrik Boogaarde, Hugo de Groote,Martin K. van Ittersumb, Kenneth G. Cassman	2015	Presented a case study (Nebraska - USA and at a national scale for Argentina and Kenya) on the application of an explicit rationale design approach in identifying the data sources which simulates Crop (maize) yield and also helps in quantifying the maize yield gaps. Suggested the robust guidelines for analyzing the crop yield gaps, accessing the climate and land use changes at global level to address the issues of crop yield.
6.	A Survey on Crop Yield Prediction based on Agricultural Data	Dhivya B H, Manjula R, Siva Bharathi S, Madhumathi R	2017	Presented a survey on the different algorithms applied in the assessment and prediction of crop yield. Discussed about the mechanism of knowledge the discovery in Agricultural data estimation
7.	The use of satellite data for crop yield gap analysis	David B. Lobell	2013	Discussed the use of remote sensing technology to identify and measure the causes of yield gaps and the assess the impact on the overall crop yield. Reported very simple methodologies to measure the yield difference with respect to season, environment and the land use.

3. IDEATION & PROPOSED SOLUTION

SOLUTION 1:

The accurate prediction of crop yield certainly benefits the farmers in choosing the right method to reduce the crop damage and gets best prices for their crops. A research group conducted a work with an objective of accurate prediction of crop yield through big data analytics to assess various crop yield influencing factors such as Area under Cultivation (AUC) interims of hectors, Annual Rainfall (AR) rates and Food Price Index (FPI) and to develop relationship among these parameters. Regression Analysis (RA) methodology was applied to examine the selected factors and their impact on crop prediction and final yield.

RA methodology is a multivariable investigation practice which can categorize the factors in to groups such as explanatory and response variables and helps to assess their interaction to obtain a resolution. All the selected factors of the present study design known as AR, AUC and FPI were measured for a period of 10 years between the years 1990-2000.

A novel method called Linear Regression (LR) is applied to analyze the relationship between explanatory variables (AR, AUC, FPI) and the crop yield considered as response variable. Study reported that the value for the studied factors clearly indicate that crop yield is principally

depends on AR. Study also reported that the other two factors (AUC and FPI) screened were also found to have significant impact after the AR. Study shall be continued to analyze the impact of for other substantial factors like Minimum Support Price (MSP), Cost Price Index (CPI), Wholesale Price Index (WPI) etc. and their relationship on the yields of different crops.

SOLUTION 2:

Agriculture is important for human survival because it serves the basic need. A well-known fact that the majority of population (≥55%) in India is into agriculture. Due to variations in climatic conditions, there exist bottlenecks for increasing the crop production in India. It has become challenging task to achieve desired targets in Agri based crop yield. Factors like climate, geographical conditions, economic and political conditions are to be considered which have direct impact on the production, productivity of the crops. Crop yield prediction is one of the important factors in agriculture practices.

Farmers need information regarding crop yield before sowing seeds in their fields to achieve enhanced crop yield. The use of technology in agriculture has increased in recent year and data analytics is one such trend that has penetrated into the agriculture field being used for management of crop yield and monitoring crop health. The recent trends in the domain of agriculture have made the people to understand the significance of Big data. The main challenge using big data in agriculture is identification of impact and effectiveness of big data analytics. Efforts are going on to understand how big data analytics can be used to improve the productivity in agricultural practices.

In order to achieve the higher crop yield, farmers must need to tackle the influencing factors such as influence of change in climate conditions on the prospects of crop yields.

The analysis of data related to agriculture helps in crop yield prediction, crop health monitoring and other such related activities. There exist several studies

related to the use of data analytics in the agriculture domain. There are various data analytics methods applied to crop yield prediction.

SOLUTION 3:

Analyzing the yields of crop is necessary to update the policies to ensure food security. A research group conducted a study with the aim in suggesting a novel data mining method to predict the yields of crop depends on agricultural big data analytics methodologies, which were progressively contrast with conventional data mining methodologies in the process of handling data and modeling designs.

Nearest neighbors modeling is one such novel data mining technique which works on the results collected based on data processing structures form the farmers and suggest a well unbiased result on the base of accuracy and prediction time in advance.

Simulation models based on field experiment are valuable technologies for studying and understanding crop yield gaps, but one of the critical challenge remain with these methods is scaling up of these approach to assess the data collated between different time intervals from the broader geographical regions. Satellite retrieved data have frequently been revealed to present data sets that, by itself or in grouping with other information and model designs, can precisely determine the yields of crop in agricultural lands.

The yield maps developed shall provide an unique opportunity to overcome both spatial and temporal based scaling up challenges and thus improve the ideology of crop yield gaps prediction.

First method works closely with the constructive maps representing the average

crop yields, it can be used directly to accesses specific crop yield influencing factors for further studies whereas the second method use the remote sensing technology to retrieve the data for providing the useful information regarding the crop yield prediction and estimation.

SOLUTION 4:

The main objective is collecting agricultural dataset which can be used to analyzed for useful crop yield forecasting. To predict the crop yield with the help of data mining technique, advanced methods can be introduced to predict crop yield and it is also helps the farmer to choose the most suitable crop, thereby improving the value and gain of the farming area. Initially the raw data set was collected and it is subjected to preprocess for noise removing (replacement of missing values) and computational methods.

From that dataset, it is subjected to Feature selection for make a predictive modeling. In this proposed approach it is mainly focused on Regression Techniques. Various regression analysis should be performed and it was compared and tested.

Regression analysis is a form of predictive modeling technique which investigates the association between a dependent (target) and independent variable(s) (predictor). This technique is used for forecasting, time series modeling and discovers the causal effect relationship between the variables.

Regression analysis indicates the significant relationships between dependent variable and independent variable and it indicates the strength of impact of multiple independent variables on a dependent variable.

Yield gaps can be defined as the difference between the expected crop yields with respect to the actual crop yield and accurate, spatially unambiguous awareness and information about the yield gaps is necessary to achieve sustainable amplification of agricultural yields.

PROPOSED SOLUTIONS:-

S.No.	Parameter	Description
1.	Problem Statement (Problem to be solved)	The foundation of the Indian economy is agriculture, which meets a fundamental requirement for human survival. Due to weather variations and other variables, many farmers in India are not receiving the anticipated crop output. In this situation, timely advise must be given to farmers in order to forecast future agricultural productivity, and analysis must be done in order to help farmers to increase crop production.
2.	Idea / Solution description	Utilizing statistical techniques to forecast crop output across multiple regions allows for the estimation of the best crop production, allowing farmers to benefit from the forecast. Additionally, to increase agricultural productivity and decrease waste.
3.	Novelty / Uniqueness	Identifying appropriate meteorological conditions and soil types for each crop by

		visually showing the relationship between numerous parameters. • Enhanced and precise visualization.
4.	Social Impact / Customer Satisfaction	Farmers will be able to recognize the ideal growing conditions so they may make plans to boost crop yield and income.
5.	Business Model (Revenue Model)	Numerous factors, including soil fertility, climate, rainfall, and others, will have an impact on crop output, which will have an impact on revenue. Crop yields and income are increased through examination of each factor's ideal conditions.
6.	Scalability of the Solution	Crop yield can be boosted by using the knowledge gained through the examination of many parameters, such as weather patterns and soil types.

4. PROJECT DESIGN

Project Flow

- Users create multiple analysis graphs/charts.

- Using the analyzed chart create the dashboard.

- Saving and Visualizing the final dashboard in the IBM Cognos Analytics. To accomplish this,we have to complete all the activities and tasks listed below

- IBM Cloud Account

- Login to Cognos Analytics

- Working with the Dataset

- Understand the Dataset

- Loading the Dataset

- Data visualization charts

- Seasons with average productions

- With years usage of Area and Production

- Top 10 States with most area

- State with crop production

- States with the crop production along with season (Text Table)

- Dashboard Creation

- Export the Analytics

Solution Architecture:

A complicated process with numerous sub-processes, solution architecture connects business issues with technological solutions. Its objectives are to

- Track down the most effective technological remedy for current business issues. Explain to project stakeholders the structure, traits, behaviour, and other features of the software.
- Specify the features, stages of development, and requirements for the solution.
- Offer guidelines for how the solution is created, managed, and delivered.

5.TESTING :

Model Performance Testing:

Project team shall fill the following information in model performance testing template.

S.No	Parameter	Screenshot / Values
1.	Dashboard design	No of Visualizations / Graphs – 5
2.	Data Responsiveness	Yes, the website is responsive completely,that ia by resizing the browser window size as per the test scenario.
3.	Amount Data to Rendered (DB2 Metrics)	Totally there are **246092** records in the dataset.
4.	Utilization of Data Filters	Data Filter used in Estimate The Crop Yield is the Crop attribute.
5.	Effective User Story	No of Scene Added – 8 ● To create the Registration page of the Website ● To create the Log in page of the Website ● To create the Dashboard page of theWebsite ● To work on the given dataset,Understand the Dataset ● Load the dataset to Cloud platform then Build the required Visualizations ● Using the Crop production in Indian dataset,create various graphs and charts to highlight the insights andvisualizations. ● Build a Visualizations to showcase Average Crop Production by Seasons ● Showcase the Yearly usage of Area inCrop Production

6.	Descriptive Reports	No of Visualizations / Graphs – 5
		- Visualization1 - Average Crop Production by Seasons - Visualization2 - Yearly usage of area in crop production - Visualization3 - Top 10 States in Crop Yield Production by Area - Visualization 4 - Crop Production byState - Visualization5 - Represent the States with Seasonal Crop Production using a Text representation

User Acceptance Testing:

User Acceptance Testing is a critical phase of any project and requires significant participation by the end user. It also ensures that the system meets the functional requirements.

The purpose of this document is to briefly explain the test coverage and open issues of the [Estimation of crop yield analysis] project at the time of the release to
User Acceptance Testing (UAT).

Test Case Analysis:

This report shows the number of test cases that have passed, failed, and untested.

Section	Total Cases	Not Tested	Fail	Pass
User Acceptance	4	0	0	4
Client Application	45	0	2	43
Security	3	0	0	3
Outsource Data	5	0	0	5
Exception Reporting	9	0	0	9
Final Report Output	4	0	0	4
Version Control	2	0	0	2

RESULTS:

Using the Crop production in Indian dataset, we plan to create various graphs and charts to highlight the insights and visualizations.

* Build a Visualization to showcase Average Crop Production by Seasons.
* Showcase the Yearly usage of Area in Crop Production.
* Build a visualization to show case top 10 States in Crop Yeild Production by Area.
* Build the required Visualization to showcase the Crop Production by State.
* Build Visual analytics to represent the Sates with Seasonal Crop Production using a Text representation.

different seasons with average production

Production by Season colored by Season

Season
● Autumn   ● Kharif   ● Rabi   ● Summer   ● Whole Year   ● Winter
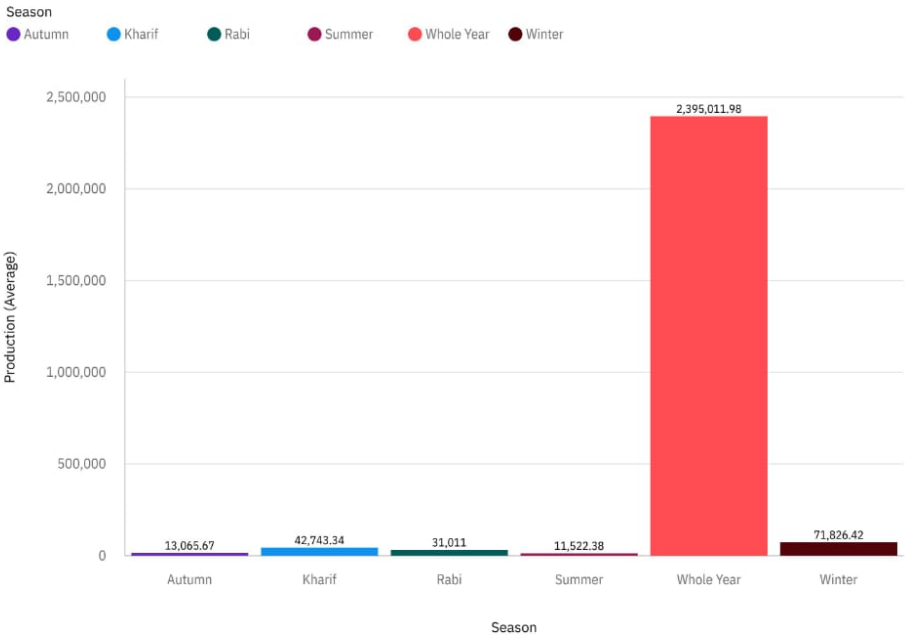
with years usage of area and production

Production by Crop_Year
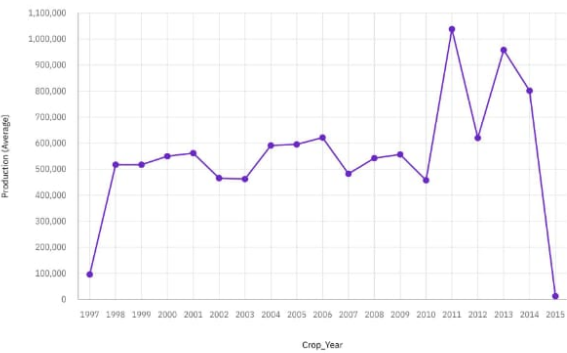
Area by Crop_Year
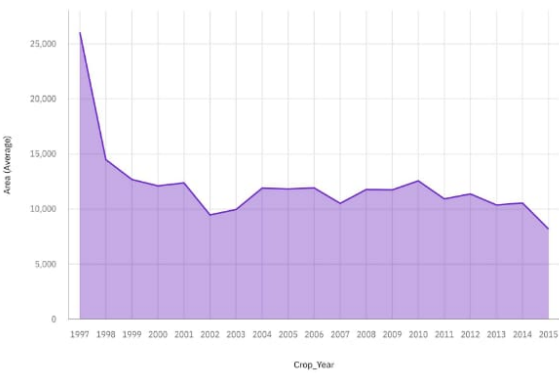
top 10 states with most area

Area by State_Name colored by State_Name

State_Name
● Andhra Pradesh ● Gujarat ● Haryana ● Madhya Pradesh ● Maharashtra ● Punjab ● Rajasthan
● Telangana ● Uttar Pradesh ● West Bengal
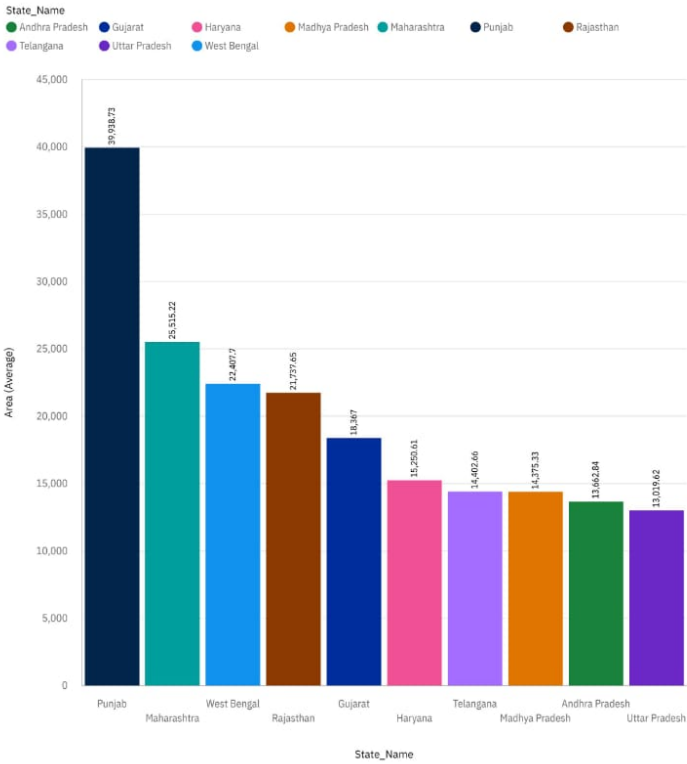
states with crop production

State_Name for State_Name regions

State_Name

● Arunachal Pradesh	● Assam	● Gujarat
● Himachal Pradesh	● Maharashtra	● Nagaland
● Telangana	● Tripura	● West Bengal
● Bihar	● Chandigarh	● Chhattisgarh
● Jharkhand	● Karnataka	● Manipur
● Haryana	● Jammu and Kashmir	● Madhya Pradesh
● Meghalaya	● Mizoram	● Odisha
● Punjab	● Sikkim	● Tamil Nadu
● Uttar Pradesh	● Uttarakhand	● Andaman and Nicobar Islands
● Andhra Pradesh	● Dadra and Nagar Haveli	● Goa

© Mapbox © OpenStreetMap

table

State_Name and Crop

Crop	State_Name
Apple	Tamil Nadu

Season and Crop

Crop	Season
Apple	Whole Year

Tab 6

Production by Season colored by Season

Season
● Autumn   ● Kharif   ● Rabi   ● Summer
● Whole Year   ● Winter

Autumn   Kharif   Rabi   Summer   Whole Year   Winter

Production by Crop_Year

1,000,000

500,000

0

1997   2001   2005   2009   2013   2...
   1999   2003   2007   2011

Crop_Year

Area by Crop_Year

20,000

10,000

0

1997 1999 2001 2003 2005 2007 2009 2011 2013 2015

Crop_Year

State_Name for State_Name regions

State_Name
● Arunachal Pradesh
● Assam
● Gujarat
● Himachal Pradesh

© Mapbox © OpenStreetMap

Area by State_Name colored by State_Name

State_Name
● Andhra Pradesh   ● Gujarat   ● Haryana

Punjab   West Bengal   Gujarat   Telangana   Andhra Pradesh
   Maharashtra   Rajasthan   Haryana   Madhya Pradesh   Uttar...

Season and Crop

Crop	Season
Apple	Whole Year
Arcanut (Processed)	Whole Year
	Kharif
Arecanut	Rabi
	Whole Year
	Autumn
	Kharif

State_Name and Crop

Crop	State_Name
Apple	Tamil Nadu
Arcanut (Processed)	Karnataka
	Andaman and Nicobar Isla...
	Andhra Pradesh
	Assam
	Goa
Arecanut	Karnataka