



Data Science in Agriculture

Wipro Project [Batch 1] : Group 18

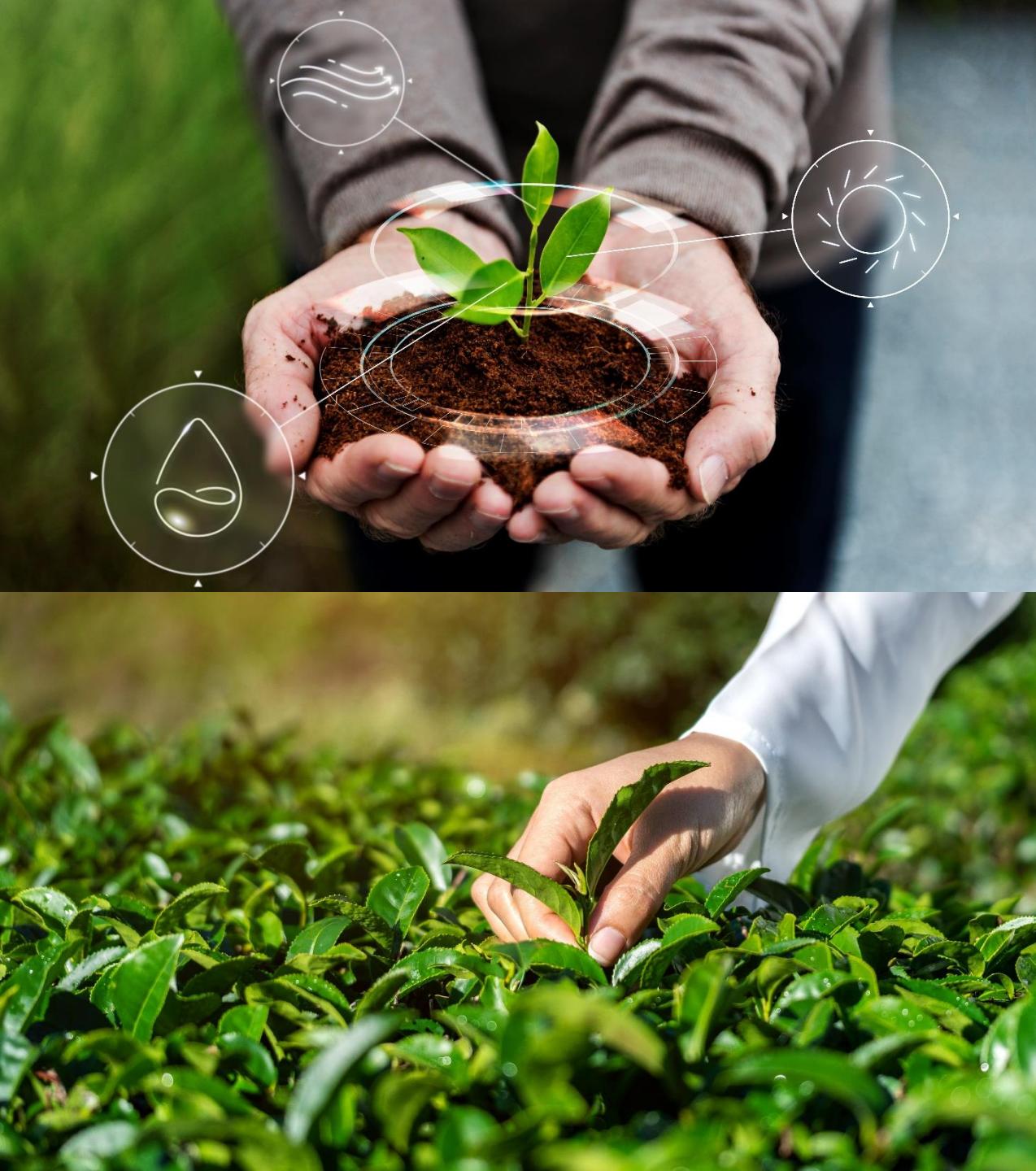
Team members :

Yash Namdeo , Gutha Venkata Sai Yashwanth , Pragati Raizada ,
Mondi Kanaka Girija Rani , Deepansh Pahuja , Mopuru Shanmuga Sai
, Neelam Sai Praneeth , Sirisha Duddukunta , Rushikesh Rupnar ,
Pranali Mahendra Nalawade

Contents

- Objective
- Introduction
- Working
- Result
- Advantages & Applications
- Conclusion



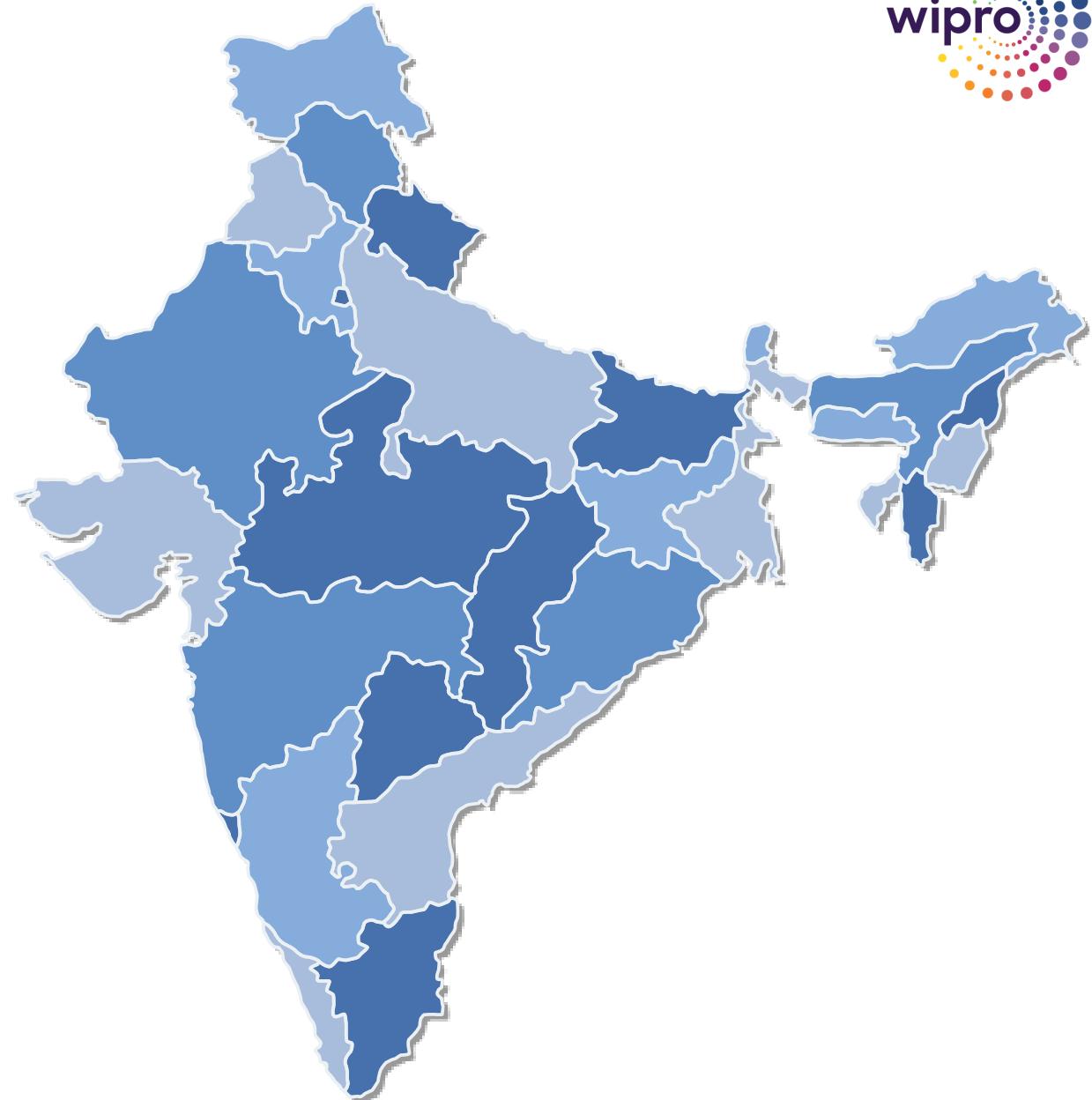


Objective

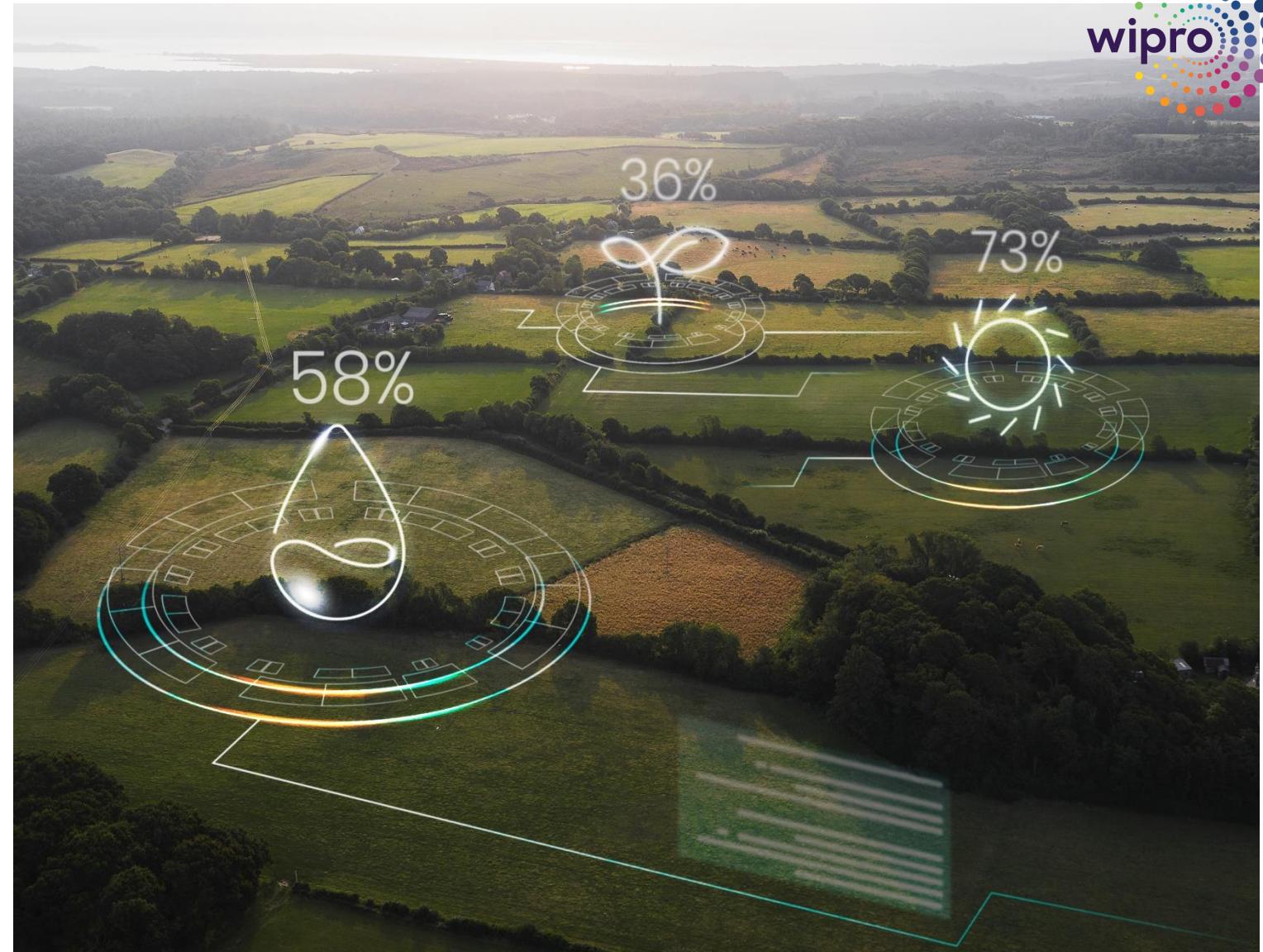
“To bring the data for analysis and prepare for smart farming. The analysis should help the farmers in yielding more profits and providing the food grains in equal distributed manner across India”

Introduction

- While agriculture's share in India's economy has progressively declined to less than 15% due to the high growth rates of the industrial and services sectors, the sector's importance in India's economic and social fabric goes well beyond this indicator.
- First, nearly three-quarters of India's families depend on rural incomes. Second, the majority of India's poor (some 770 million people or about 70 percent) are found in rural areas. And third, India's food security depends on producing cereal crops, as well as increasing its production of fruits, vegetables and milk to meet the demands of a growing population with rising incomes.
- To do so, a productive, competitive, diversified and sustainable agricultural sector will need to emerge at an accelerated pace.



- India is a global agricultural powerhouse. It is the world's largest producer of milk, pulses, and spices, and has the world's largest cattle herd (buffaloes), as well as the largest area under wheat, rice and cotton.
- It is the second largest producer of rice, wheat, cotton, sugarcane, farmed fish, sheep & goat meat, fruit, vegetables and tea.
- The country has some 195 m ha under cultivation of which some 63 percent are rainfed (roughly 125m ha) while 37 percent are irrigated (70m ha). In addition, forests cover some 65m ha of India's land.
- Agriculture in India is livelihood for a majority of the population and can never be underestimated .





Although its contribution in the gross domestic product (GDP) has reduced to less than 20 per cent and contribution of other sectors increased at a faster rate, agricultural production has grown. This has made us self-sufficient and taken us from being a begging bowl for food after independence to a net exporter of agriculture and allied products.

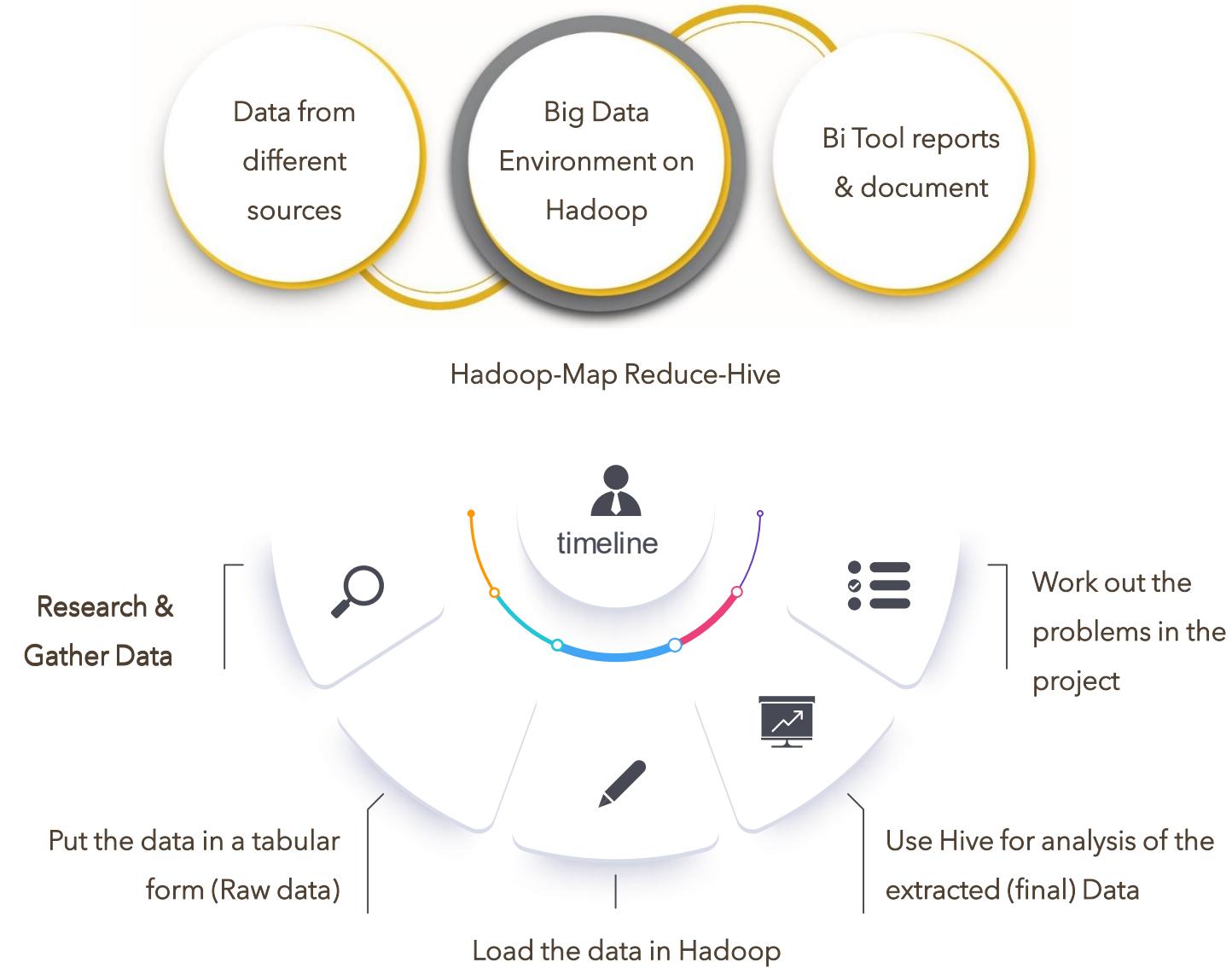
In spite of all these facts, the average productivity of many crops in India is quite low. The country's population in the next decade is expected to become the largest in the world and providing food for them will be a very prime issue. Farmers are still not able to earn respectable earnings

Working

For this Hadoop-Map Reduce-Hive & BI Tool reports would be used for the working of the project. The project was distributed into 5 parts so that the execution would be seamless.

- Research & Gather Data**

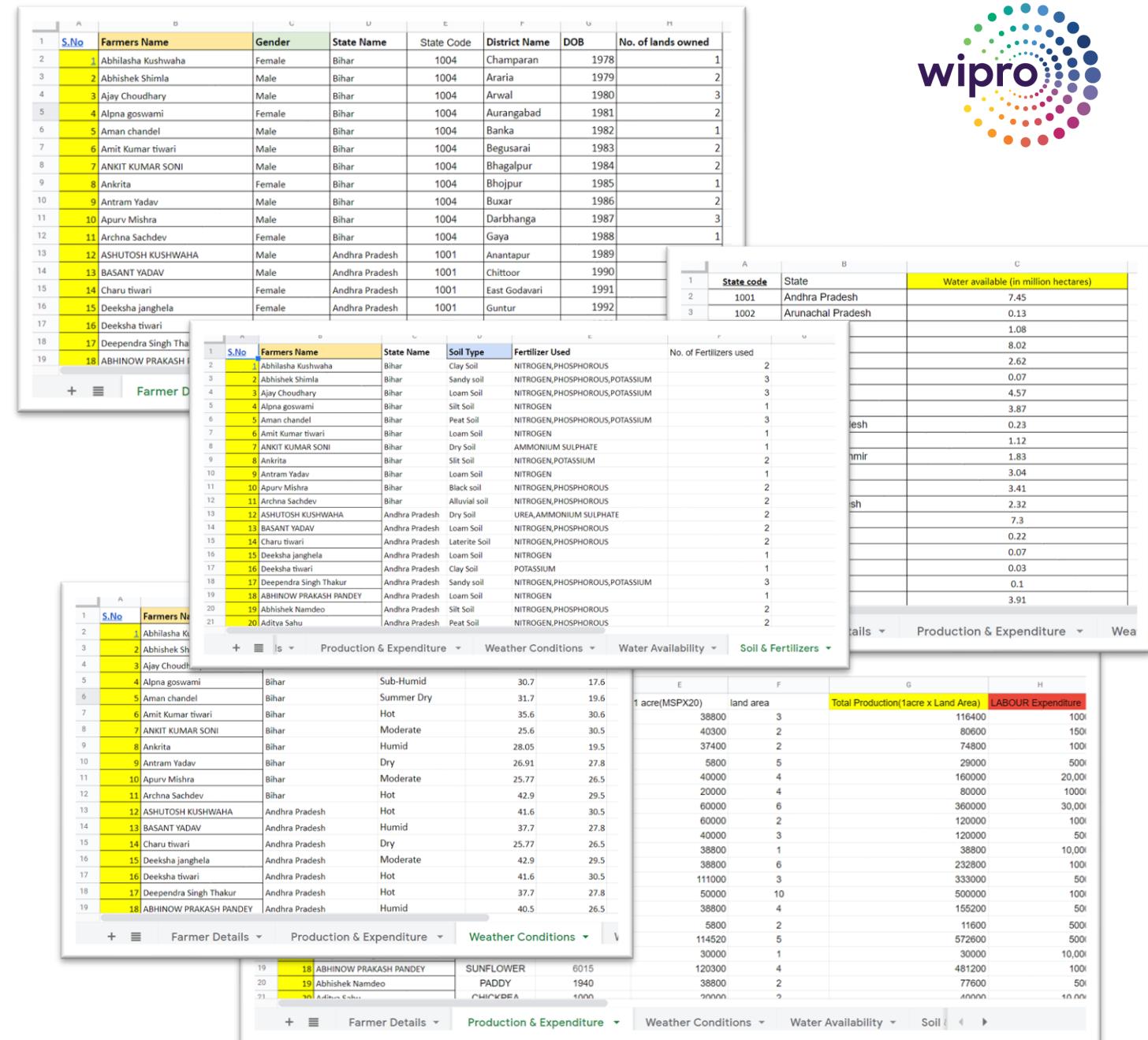
Exploration regarding the project was done and the basic data was gathered from numerous government sites which were pertinent to agriculture. Agricultural research can be considerably defined as any research activity intended for enhancing the productivity and quality of crops improvement, better plant protection, irrigation, storage methods, farm mechanization, efficient marketing, and better management of resources.



- The Data was gathered and was put in a tabular manner.

All the columns were related to agriculture in data Science like farmer's names, state, district, weather, expenditure, production and many more columns were added.

One of the simplest methods used to analyze the data and display the data is in tabular form. In the tabular form, we got a systematic arrangement of rows and columns. It is very accurate as well as an easy method to display the data.



Farmer Details:

S.No	Farmers Name	Gender	State Name	State Code	District Name	DOB	No. of lands owned
1	Abhilasha Kushwaha	Female	Bihar	1004	Champaran	1978	1
2	Abhishek Shimla	Male	Bihar	1004	Araria	1979	2
3	Ajay Choudhary	Male	Bihar	1004	Arwal	1980	3
4	Alpna goswami	Female	Bihar	1004	Aurangabad	1981	2
5	Aman chandel	Male	Bihar	1004	Banka	1982	1
6	Amit Kumar tiwari	Male	Bihar	1004	Begusarai	1983	2
7	ANKIT KUMAR SONI	Male	Bihar	1004	Bhagalpur	1984	2
8	Ankrita	Female	Bihar	1004	Bhojpur	1985	1
9	Antram Yadav	Male	Bihar	1004	Buxar	1986	2
10	Apurv Mishra	Male	Bihar	1004	Darbhanga	1987	3
11	Archna Sachdev	Female	Bihar	1004	Gaya	1988	1
12	ASHUTOSH KUSHWAHA	Male	Andhra Pradesh	1001	Anantapur	1989	
13	BASANT YADAV	Male	Andhra Pradesh	1001	Chittoor	1990	
14	Charu tiwari	Female	Andhra Pradesh	1001	East Godavari	1991	
15	Deeksha janghela	Female	Andhra Pradesh	1001	Guntur	1992	
16	Deeksha tiwari						
17	Deependra Singh Thakur						
18	ABHINOW PRAKASH PANDEY						

Production & Expenditure:

State code	State	Water available (in million hectares)
1001	Andhra Pradesh	7.45
1002	Arunachal Pradesh	0.13
		1.08
		8.02
		2.62
		0.07
		4.57
		3.87
		0.23
		1.12
		1.83
		3.04
		3.41
		2.32
		7.3
		0.22
		0.07
		0.03
		0.1
		3.91

Weather Conditions:

S.No	Farmers Name	State Name	Soil Type	Fertilizer Used	No. of Fertilizers used
1	Abhilasha Kushwaha	Bihar	Clay Soil	NITROGEN,PHOSPHOROUS	2
2	Abhishek Shimla	Bihar	Sandy soil	NITROGEN,PHOSPHOROUS,POTASSIUM	3
3	Ajay Choudhary	Bihar	Loam Soil	NITROGEN,PHOSPHOROUS,POTASSIUM	3
4	Alpna goswami	Bihar	Silt Soil	NITROGEN	1
5	Aman chandel	Bihar	Peat Soil	NITROGEN,PHOSPHOROUS,POTASSIUM	3
6	Amit Kumar tiwari	Bihar	Loam Soil	NITROGEN	1
7	ANKIT KUMAR SONI	Bihar	Dry Soil	AMMONIUM SULPHATE	1
8	Ankrita	Bihar	Slit Soil	NITROGEN,POTASSIUM	2
9	Antram Yadav	Bihar	Loam Soil	NITROGEN	1
10	Apurv Mishra	Bihar	Black soil	NITROGEN,PHOSPHOROUS	2
11	Archna Sachdev	Bihar	Alluvial soil	NITROGEN,PHOSPHOROUS	2
12	ASHUTOSH KUSHWAHA	Andhra Pradesh	Dry Soil	UREA,AMMONIUM SULPHATE	2
13	BASANT YADAV	Andhra Pradesh	Loam Soil	NITROGEN,PHOSPHOROUS	2
14	Charu tiwari	Andhra Pradesh	Laterite Soil	NITROGEN,PHOSPHOROUS	2
15	Deeksha janghela	Andhra Pradesh	Loam Soil	NITROGEN	1
16	Deeksha tiwari	Andhra Pradesh	Clay Soil	POTASSIUM	1
17	Deependra Singh Thakur	Andhra Pradesh	Sandy soil	NITROGEN,PHOSPHOROUS,POTASSIUM	3
18	ABHINOW PRAKASH PANDEY	Andhra Pradesh	Loam Soil	NITROGEN	1
19	Abhishek Namdeo	Andhra Pradesh	Silt Soil	NITROGEN,PHOSPHOROUS	2
20	Aditya Sahu	Andhra Pradesh	Peat Soil	NITROGEN,PHOSPHOROUS	2

Soil & Fertilizers:

E	F	G	H
1 acre(MSPX20)	land area	Total Production(1acre x Land Area)	LABOUR Expenditure
38800	3	116400	1000
40300	2	80600	1500
37400	2	74800	1000
5800	5	29000	500
40000	4	160000	20000
60000	6	360000	30000
60000	2	120000	1000
40000	3	120000	500
38800	1	38800	10000
38800	6	232800	1000
111000	3	333000	500
50000	10	500000	1000
38800	4	155200	500
5800	2	11600	500
114520	5	572600	500
30000	1	30000	10000
120300	4	481200	1000
38800	2	77600	500
20000	2	40000	10000

Farmer Details:

S.No	Farmers Name	State Name	Soil Type	Fertilizer Used	No. of Fertilizers used
1	Abhilasha Kushwaha	Bihar	Sub-Humid	30.7	17.6
2	Abhishek Shimla	Bihar	Summer Dry	31.7	19.6
3	Ajay Choudhary	Bihar	Moderate	25.6	30.5
4	Alpna goswami	Bihar	Humid	28.05	19.5
5	Aman chandel	Bihar	Hot	35.6	30.6
6	Amit Kumar tiwari	Bihar	Moderate	25.6	30.5
7	ANKIT KUMAR SONI	Bihar	Humid	28.05	19.5
8	Ankrita	Bihar	Hot	35.6	30.6
9	Antram Yadav	Bihar	Dry	26.91	27.8
10	Apurv Mishra	Bihar	Moderate	25.77	26.5
11	Archna Sachdev	Bihar	Hot	42.9	29.5
12	ASHUTOSH KUSHWAHA	Andhra Pradesh	Hot	41.6	30.5
13	BASANT YADAV	Andhra Pradesh	Humid	37.7	27.8
14	Charu tiwari	Andhra Pradesh	Dry	25.77	26.5
15	Deeksha janghela	Andhra Pradesh	Moderate	42.9	29.5
16	Deeksha tiwari	Andhra Pradesh	Hot	41.6	30.5
17	Deependra Singh Thakur	Andhra Pradesh	Hot	37.7	27.8
18	ABHINOW PRAKASH PANDEY	Andhra Pradesh	Humid	40.5	26.5

Production & Expenditure:

E	F	G	H
1 acre(MSPX20)	land area	Total Production(1acre x Land Area)	LABOUR Expenditure
38800	3	116400	1000
40300	2	80600	1500
37400	2	74800	1000
5800	5	29000	500
40000	4	160000	20000
60000	6	360000	30000
60000	2	120000	1000
40000	3	120000	500
38800	1	38800	10000
38800	6	232800	1000
111000	3	333000	500
50000	10	500000	1000
38800	4	155200	500
5800	2	11600	500
114520	5	572600	500
30000	1	30000	10000
120300	4	481200	1000
38800	2	77600	500
20000	2	40000	10000

Weather Conditions:

E	F	G	H
1 acre(MSPX20)	land area	Total Production(1acre x Land Area)	LABOUR Expenditure
38800	3	116400	1000
40300	2	80600	1500
37400	2	74800	1000
5800	5	29000	500
40000	4	160000	20000
60000	6	360000	30000
60000	2	120000	1000
40000	3	120000	500
38800	1	38800	10000
38800	6	232800	1000
111000	3	333000	500
50000	10	500000	1000
38800	4	155200	500
5800	2	11600	500
114520	5	572600	500
30000	1	30000	10000
120300	4	481200	1000
38800	2	77600	500
20000	2	40000	10000

Water Availability:

E	F	G	H
1 acre(MSPX20)	land area	Total Production(1acre x Land Area)	LABOUR Expenditure
38800	3	116400	1000
40300	2	80600	1500
37400	2	74800	1000
5800	5	29000	500
40000	4	160000	20000
60000	6	360000	30000
60000	2	120000	1000
40000	3	120000	500
38800	1	38800	10000
38800	6	232800	1000
111000	3	333000	500
50000	10	500000	1000
38800	4	155200	500
5800	2	11600	500
114520	5	572600	500
30000	1	30000	10000
120300	4	481200	1000
38800	2	77600	500
20000	2	40000	10000

Soil & Fertilizers:

E	F	G	H
1 acre(MSPX20)	land area	Total Production(1acre x Land Area)	LABOUR Expenditure
38800	3	116400	1000
40300	2	80600	1500
37400	2	74800	1000
5800	5	29000	500
40000	4	160000	20000
60000	6	360000	30000
60000	2	120000	1000
40000	3	120000	500
38800	1	38800	10000
38800	6	232800	1000
111000	3	333000	500
50000	10	500000	1000
38800	4	155200	500
5800	2	11600	500
114520	5	572600	500
30000	1	30000	10000
120300	4	481200	1000
38800	2	77600	500
20000	2	40000	10000

Video of the Raw Data



Raw Data Sheet

DOB

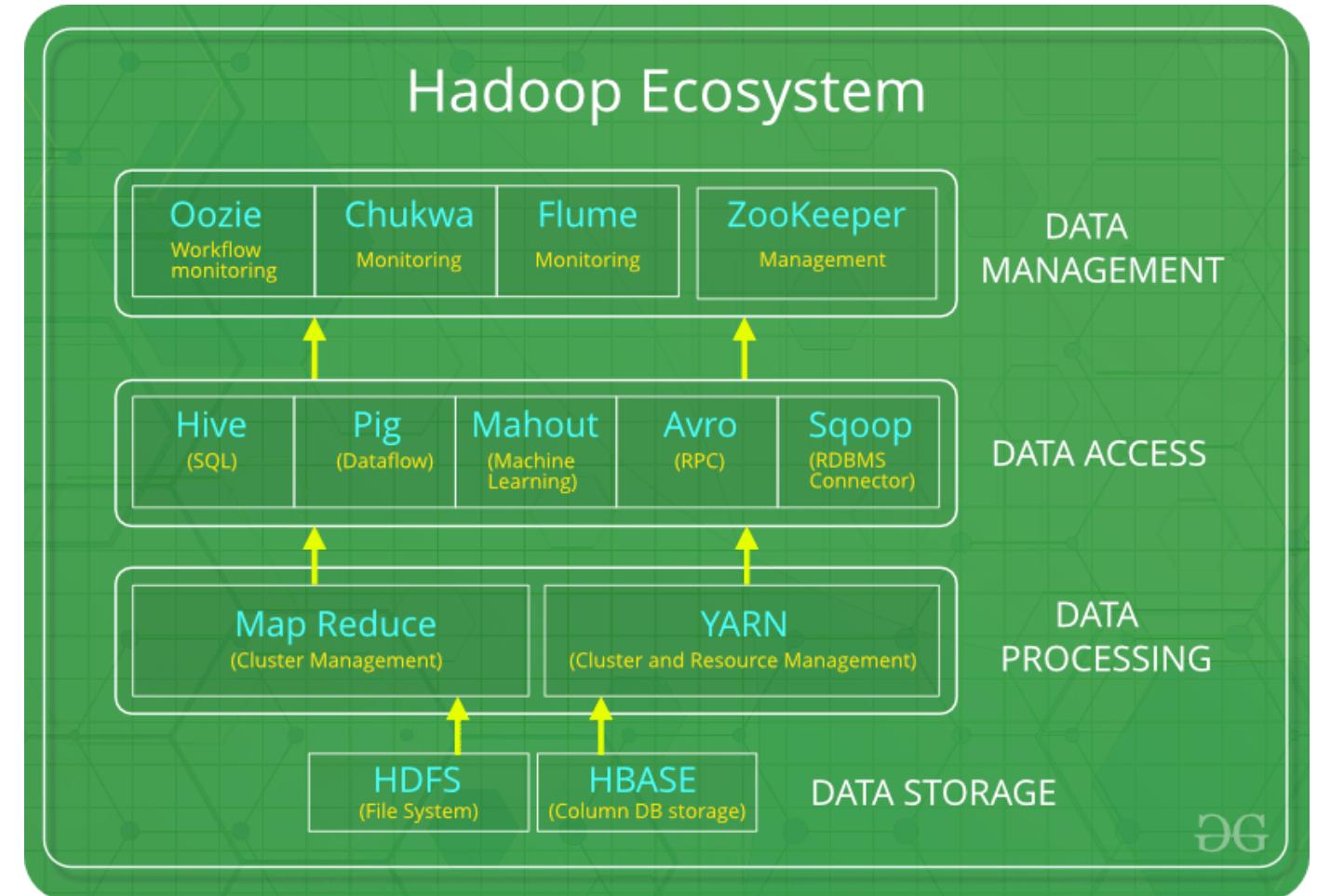
	A	B	C	D	E	F	G	H	I	J	K	L	M
1	S.No	Farmers Name	Gender	State Name	State Code	District Name	DOB	No. of lands owned					
2	1	Abhilasha Kushwaha	Female	Bihar	1004	Champaran	1978	1					
3	2	Abhishek Shimla	Male	Bihar	1004	Araria	1979	2					
4	3	Ajay Choudhary	Male	Bihar	1004	Arwal	1980	3					
5	4	Alpna goswami	Female	Bihar	1004	Aurangabad	1981	2					
6	5	Aman chandel	Male	Bihar	1004	Banka	1982	1					
7	6	Amit Kumar tiwari	Male	Bihar	1004	Begusarai	1983	2					
8	7	ANKIT KUMAR SONI	Male	Bihar	1004	Bhagalpur	1984	2					
9	8	Ankrita	Female	Bihar	1004	Bhojpur	1985	1					
10	9	Antram Yadav	Male	Bihar	1004	Buxar	1986	2					
11	10	Apurv Mishra	Male	Bihar	1004	Darbhanga	1987	3					
12	11	Archna Sachdev	Female	Bihar	1004	Gaya	1988	1					
13	12	ASHUTOSH KUSHWAHA	Male	Andhra Pradesh	1001	Anantapur	1989	1					
14	13	BASANT YADAV	Male	Andhra Pradesh	1001	Chittoor	1990	1					
15	14	Charu tiwari	Female	Andhra Pradesh	1001	East Godavari	1991	2					
16	15	Deeksha janghela	Female	Andhra Pradesh	1001	Guntur	1992	3					
17	16	Deeksha tiwari	Female	Andhra Pradesh	1001	Kadapa	1993	2					
18	17	Deependra Singh Thakur	Male	Andhra Pradesh	1001	Krishna	1994	1					
19	18	ABHINOW PRAKASH PANDEY	Male	Andhra Pradesh	1001	Kurnool	1997	2					
20	19	Abhishek Namdeo	Male	Andhra Pradesh	1001	Nellore	1998	2					
21	20	Aditya Sahu	Male	Andhra Pradesh	1001	Prakasam	1999	1					
22	21	AKASH MEHRA	Male	Andhra Pradesh	1001	Srikakulam	1992	2					
23	22	Aman Chourasiya	Male	Arunachal Pradesh	1002	Anjaw	1999	3					
24	23	Anam Taha Khan	Female	Arunachal Pradesh	1002	Changlang	1998	1					
25	24	Priyanka Tiwari	Female	Gujarat	1007	Bharuch	1996	1					

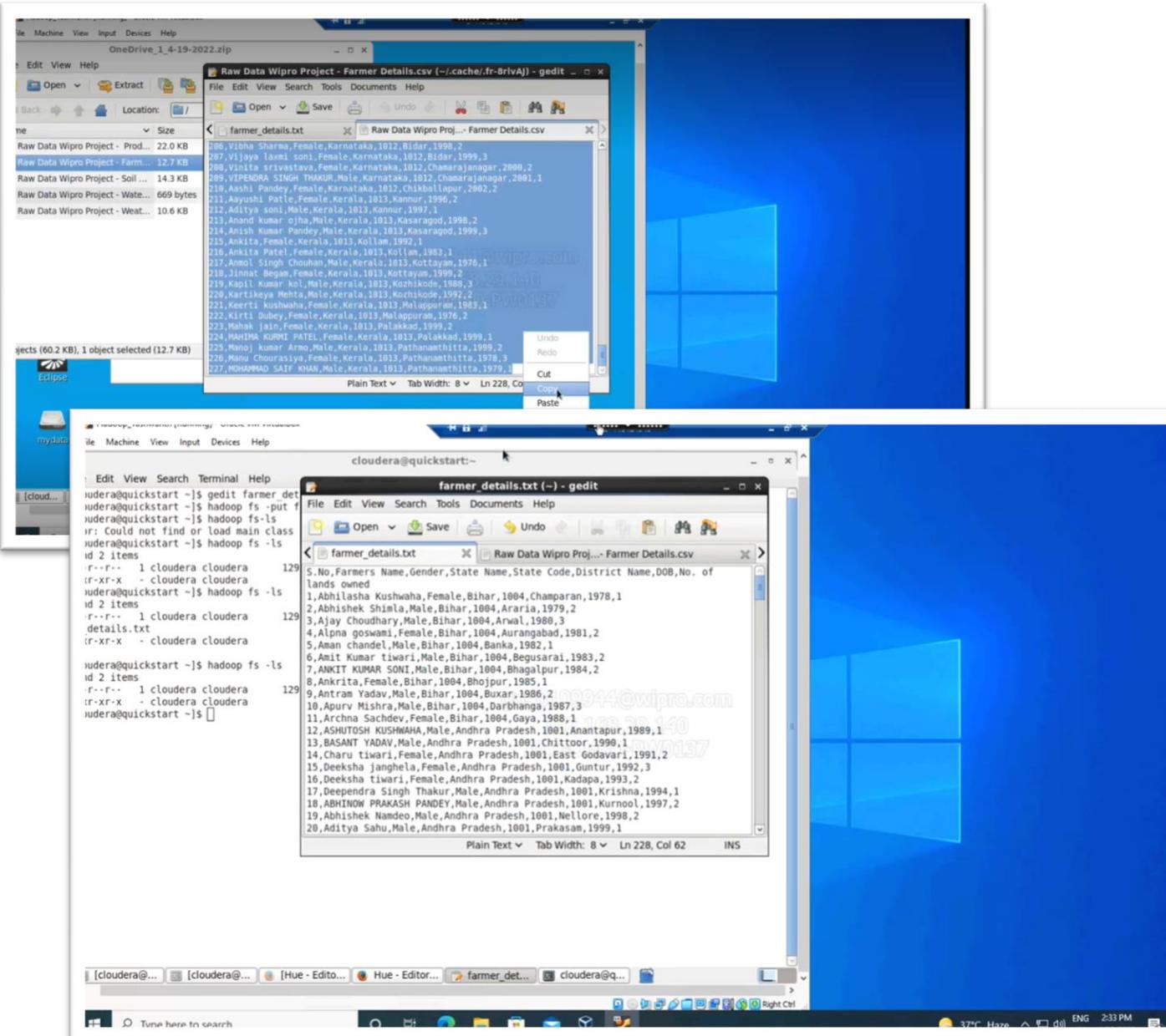
Farmer Details Production & Expenditure Weather Conditions Water Availability Soil & Fertilizers +

What is Hadoop

Apache Hadoop is an open source framework that is used to efficiently store and process large datasets ranging in size from gigabytes to petabytes of data. Instead of using one large computer to store and process the data, Hadoop allows clustering multiple computers to analyze massive datasets in parallel more quickly. Hadoop consists of four main modules:

- Hadoop Distributed File System (HDFS)
- Yet Another Resource Negotiator (YARN)
- MapReduce
- Hadoop





Why is Hadoop important?

- Ability to store and process huge amounts of any kind of data.
- Computing power.
- Fault tolerance.
- Flexibility.
- Low cost.
- Scalability.

Load the data in Hadoop

Firstly Hadoop was set up by carefully following steps in our RDP(Remote Desktop). The loading of CSV files in Hadoop (it is a comma-separated values file is a delimited text file that uses a comma to separate values. Each line of the file is a data record. Each record consists of one or more fields, separated by commas.) ,

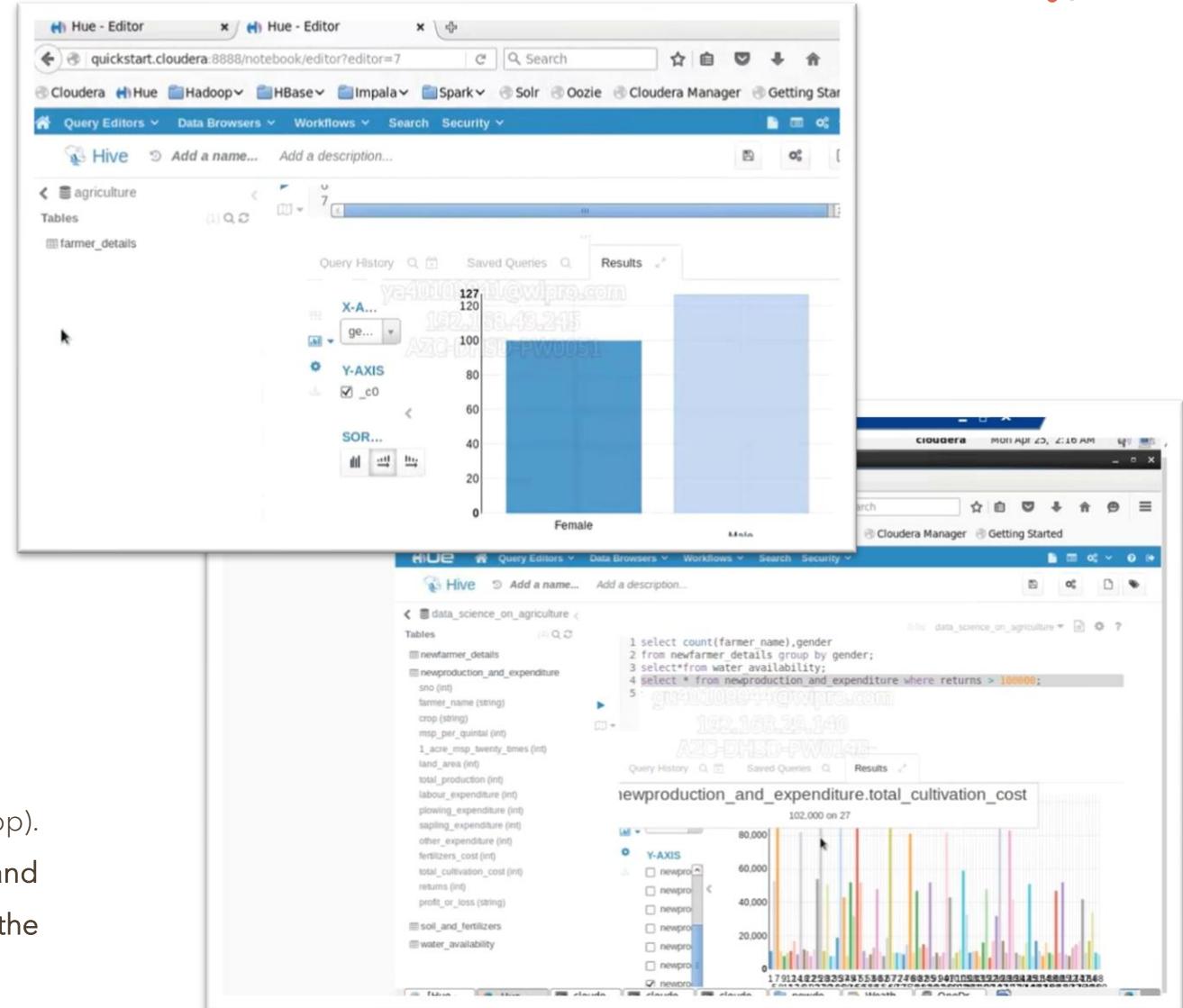
How Hadoop Works

Hadoop makes it easier to use all the storage and processing capacity in cluster servers, and to execute distributed processes against huge amounts of data. Some of the most popular applications are:

- [Spark](#) - An open source, distributed processing system commonly used for big data workloads.
- [Hive](#) - Allows users to leverage Hadoop MapReduce using a SQL interface, enabling analytics at a massive scale.
- [HBase](#) - An open source, non-relational, versioned database that runs on top of Amazon S3 (using EMRFS) or HDFS.

• Use Hive for analysis of the extracted (final) Data

Firstly Hive was set up by carefully following steps in our RDP(Remote Desktop). The loading of CSV files in Hadoop Hive, Raw data has been analysed and visually represented In the form of bar graphs by selecting the columns which the user wants to display.



Result



With the assistance of Hadoop and Hive, Raw data has been analysed and visually represented In the form of bar graphs by selecting the columns which the user wants to display.

- **Analytics:** to predict the success of a product or crop or the ill effects of a natural event on crops
- **Data Analysis:** Process vast volumes of data regarding crops, geography, production, fertilizers etc.
- **Real-Time Analytics:** to provide farmers with real-time assistance by analyzing information provided by them in real-time.
- **Increased Agricultural Productivity through Agriculture Data**
- **The great success of Fertilizing Products** across a variety of geographic conditions for Agriculture companies.



Visual Representation of the data

Result in RDP Hadoop & Hive

Hue Query Editors Data Browsers Workflows Search Security

Hive Add a name... Add a description...

Tables

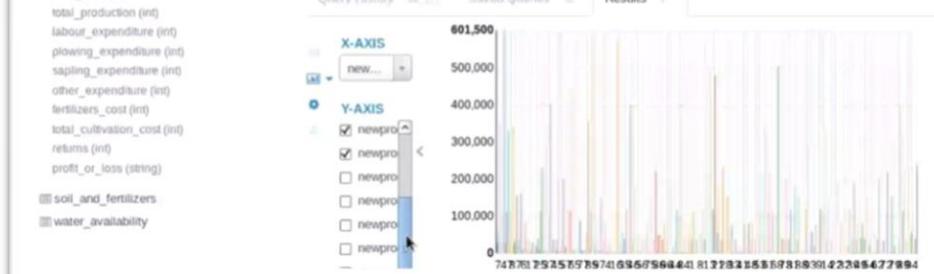
- newfarmer_details
- newproduction_and_expenditure
- sno (int)
- farmer_name (string)
- crop (string)
- mp_per_quantal (int)
- 1_acre_msp_twenty_times (int)
- land_area (int)
- total_production (int)
- labour_expenditure (int)
- plowing_expenditure (int)
- sapling_expenditure (int)
- other_expenditure (int)
- fertilizers_cost (int)
- total_cultivation_cost (int)
- returns (int)
- profit_or_loss (string)
- soil_and_fertilizers
- water_availability

```

1 select count(farmer_name),gender
2 from newfarmer_details group by gender;
3 select*from water_availability;
4 select * from newproduction_and_expenditure;
5
    
```

Query History Saved Queries Results

X-AXIS: newproduct... Y-AXIS: newproduct...



Cloudera Hue Hadoop HBase Impala Spark Solr Oozie Cloudera Manager Getting Started

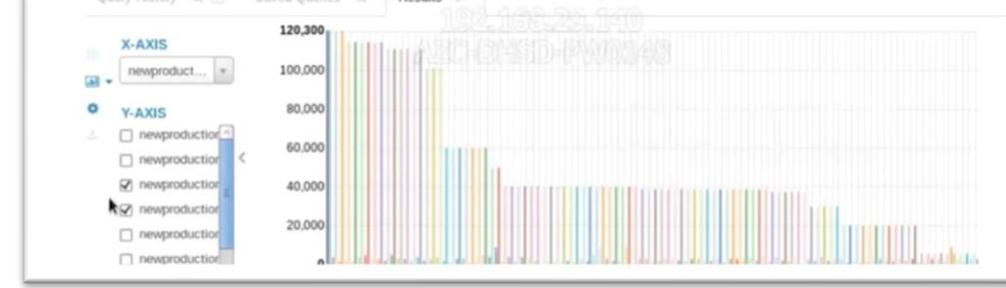
File data_science_on_agriculture

```

1 select count(farmer_name),gender
2 from newfarmer_details group by gender;
3 select*from water_availability;
4 select * from newproduction_and_expenditure;
5
    
```

Query History Saved Queries Results

X-AXIS: newproduct... Y-AXIS: newproduct...



Hue - Editor Mozilla Firefox

quickstart.cloudera:8888/notebook/editor?editor=22

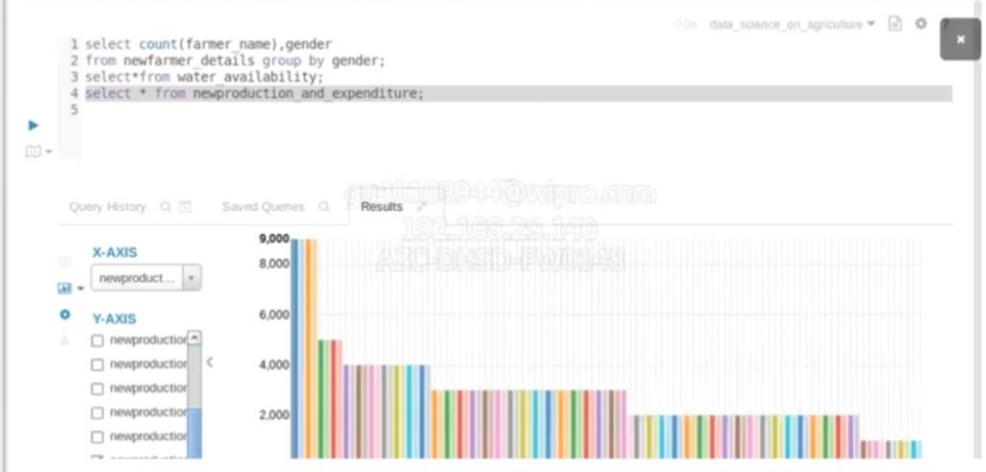
Cloudera Hue Hadoop HBase Impala Spark Solr Oozie Cloudera Manager Getting Started

```

1 select count(farmer_name),gender
2 from newfarmer_details group by gender;
3 select*from water_availability;
4 select * from newproduction_and_expenditure;
5
    
```

Query History Saved Queries Results

X-AXIS: newproduct... Y-AXIS: newproduct...



Hue - Editor Mozilla Firefox

quickstart.cloudera:8888/notebook/editor?editor=31

Cloudera Hue Hadoop HBase Impala Spark Solr Oozie Cloudera Manager Getting Started

Hive Add a name... Add a description...

Tables

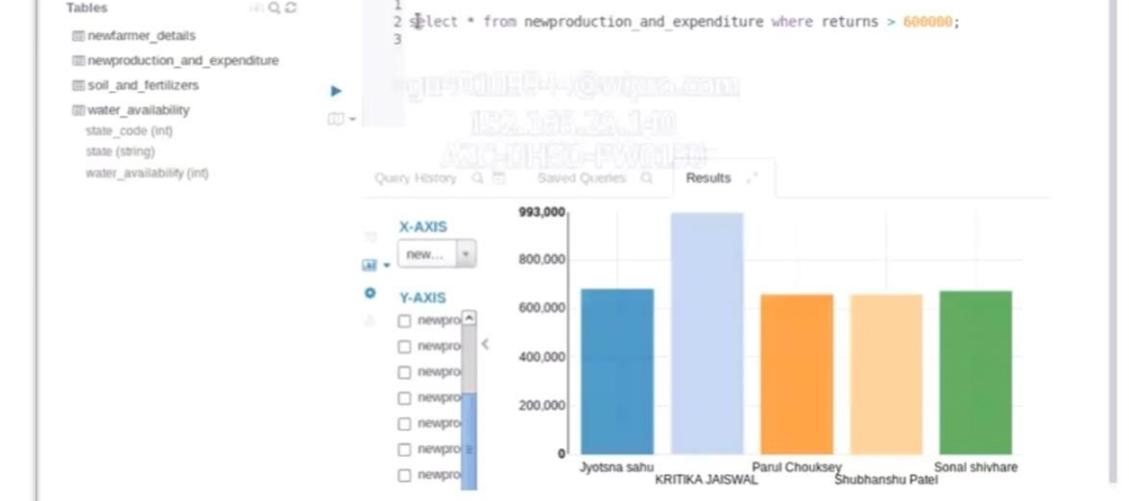
- newfarmer_details
- newproduction_and_expenditure
- soil_and_fertilizers
- water_availability
- state_code (int)
- state (string)
- water_availability (int)

```

1
2 select * from newproduction_and_expenditure where returns > 600000;
3
    
```

Query History Saved Queries Results

X-AXIS: newproduct... Y-AXIS: newproduct...



Analysis of the extracted (final) Data

Applications & Advantages

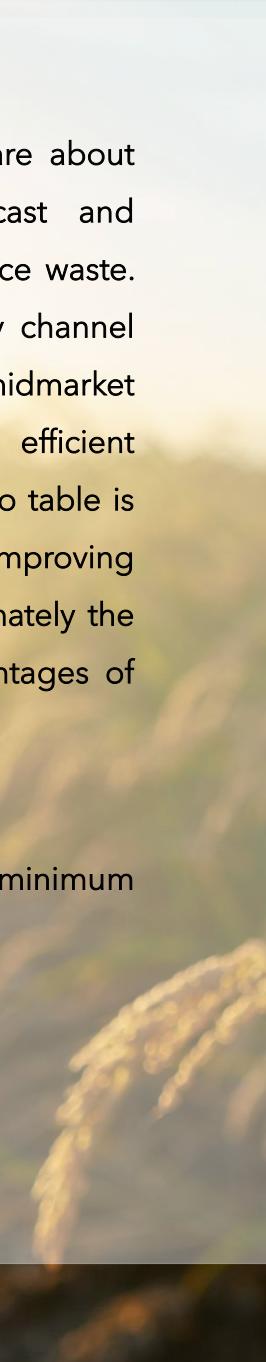


The features are included by keeping in mind the different aspects of agriculture . The goal of the study is to benefit society as much as it can.

Smart farming systems reduce waste, improve productivity and enable management of a greater number of resources. In traditional farming methods, it was a mainstay for the farmer to be out in the field, constantly monitoring the land and condition of crops. But with larger and larger farms, it has become more challenging for farmers to monitor everything everywhere.

On the demand side, smart farming systems are about careful management of the demand forecast and delivering goods to market just in time to reduce waste. Furthermore, it is about managing the delivery channel and ensuring that the transfer of product to the midmarket handler reduces waste through gentle and efficient handling. Overall, the entire process from farm to table is software-managed , reducing overall costs, improving overall yield and quality of the supply, and ultimately the experience for the consumer. Benefits or advantages of Smart Agriculture

- It allows farmers to maximize yields using minimum resources such as water, fertilizers, seeds etc.
- It is cost effective method.
- It delivers high quality crop production.



Conclusion



- Data preparation :
 - loading into Hadoop
 - Quality data preparation
 - Identifying tool for BI
- Identified Dimensions and Measures.
- State and District wise crops every year by quantity.
- Max and Min crop with the profits gained.
- Crops that need to be transported because of lack of demand in the same state.
- Comparing the same soul and climate condition Vs the crops that are produced.

Thank You

