

PROJECT REPORT

Date	17 November, 2022
Team ID	PNT2022TMID21477
Project Name	Project - Estimate the crop yield during Data Analytics

ESTIMATE THE CROP YIELD USING DATA ANALYTICS

1. INTRODUCTION

1.1 Project Overview:

Crop production in India is one of the most important sources of income and India is one of the top countries to produce crops. Crop production may vary due to various factors such as season, temperature, place, year etc. As per this project we will be analyzing some important visualization, creating a dashboard and by going through these we will get most of the insights of Crop production in India. We are going to analyse the crop yield production using different attributes like year, area, season and so on. Using those attributes, we help to make a good crop yield production in future from the past data. It is quite helpful for the farmers and the people to get right crops to cultivate in a particular season

1.2 Purpose:

Technology is used to make the people life better. It is to make and get a better crop yield production using data analytics in python from the past crop productions. It is to get insights from the past data with important parameters and help to make a good decisions for the crop yield production.

2. LITERATURE SURVEY:

2.1 Existing problem:

Crop Yield Prediction Using Machine Learning Algorithm [1] Machine learning (ML) is significant since it offers a decision-support tool for Crop Yield Prediction (CYP), which may help with decisions like which crops to cultivate and what to do during the crop's growing season. Crop yield estimation's major purpose is to boost agricultural crop production, and it does so using a variety of well-established models. To assist farmers in maximizing agricultural yield, timely instructions to forecast future crop output and analysis are required

Estimation of Crop Yield From Combined Optical and SAR Imagery Using Gaussian Kernel Regression [2] The synthetic aperture radar (SAR) interferometric coherence can complement optical data for the estimation of crop growth parameters, but it has not been yet investigated for predicting crop yield. Many studies have used machine-learning methods, such as neural networks, random forest, and Gaussian process regression, to estimate crop yield from remotely sensed data. The main objective was to investigate the synergetic use of Sentinel-2 vegetation indices and Sentinel-1 interferometric coherence data through Gaussian kernel regression for estimating rice grain yield. The

prediction accuracy was assessed using in situ measured yield data collected in 2019 and 2020 over Xinghua county in Jiangsu Province, China. In all cases, Gaussian kernel regression outperformed the probabilistic Gaussian regression and Bayesian linear inference.

Crop Yield Estimation in India Using Machine Learning [3] The main aim of this paper is to predict crop yield using area, yield, production, and area under irrigation. Four machine learning techniques Decision Tree, Linear Regression, Lasso regression, and Ridge Regression have been applied to estimate the crop yield. Cross validations methods, for validation, mean absolute error, mean squared error, and root mean squared error, were used to validate. The Decision tree outperforms other machine learning techniques.

2.2 References:

- [1] R. J, V. K. G. Kalaiselvi, A. Sheela, D. S. D and J. G, "Crop Yield Prediction Using Machine Learning Algorithm," 2021 4th International Conference on Computing and Communications Technologies (ICCCT), 2021, pp. 611-616, doi: 10.1109/ICCCT53315.2021.9711853.
- [2] Y. Alebele et al., "Estimation of Crop Yield From Combined Optical and SAR Imagery Using Gaussian Kernel Regression," in IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 14, pp. 10520-10534, 2021, doi: 10.1109/JSTARS.2021.3118707.
- [3] M. Kavita and P. Mathur, "Crop Yield Estimation in India Using Machine Learning," 2020 IEEE 5th International Conference on Computing Communication and Automation (ICCCA), 2020, pp. 220-224, doi: 10.1109/ICCCA49541.2020.9250915.

2.3 Problem Statement Definition:

Crop production in India is one of the most important sources of income and India is one of the top countries to produce crops. Using python, we will be analyzing some important visualization and getting interpretation and going through these we will get most of the insights of Crop production in India and it is used to help the people to make a better choice for cultivation depends on a season and location.

3. IDEATION & PROPOSED SOLUTION:

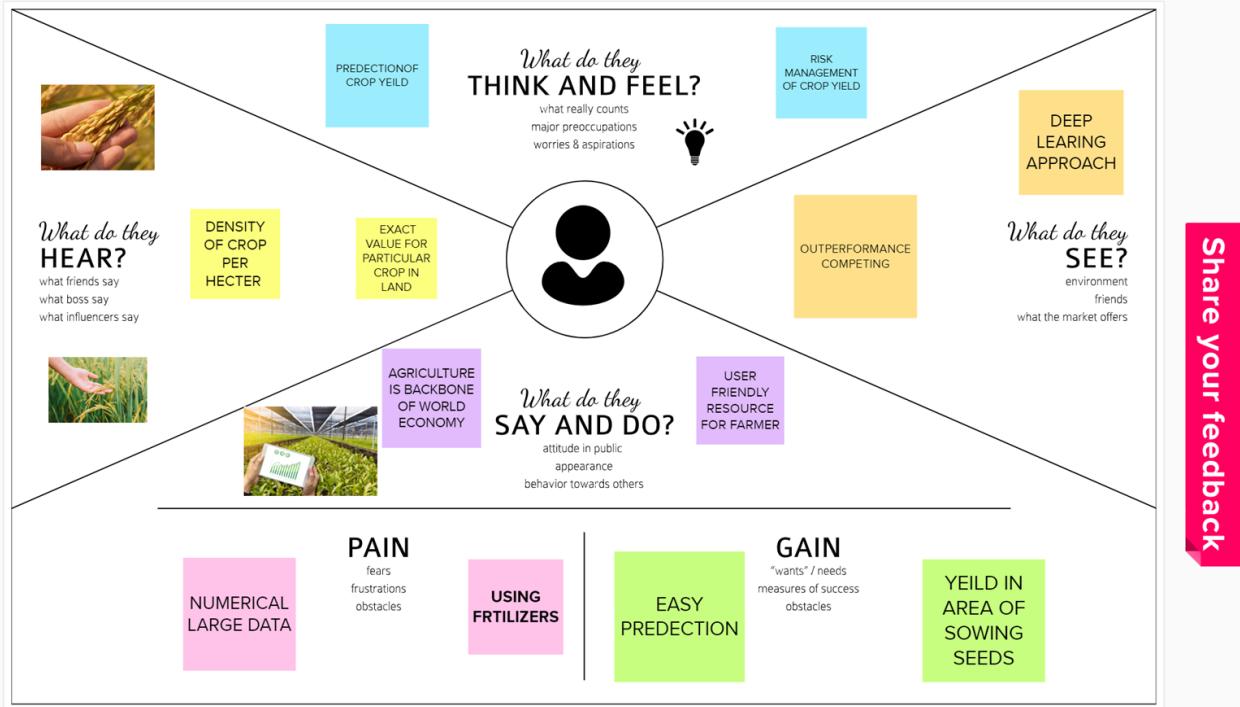
3.1 Empathy Map Canvas:

Empathy Map Canvas

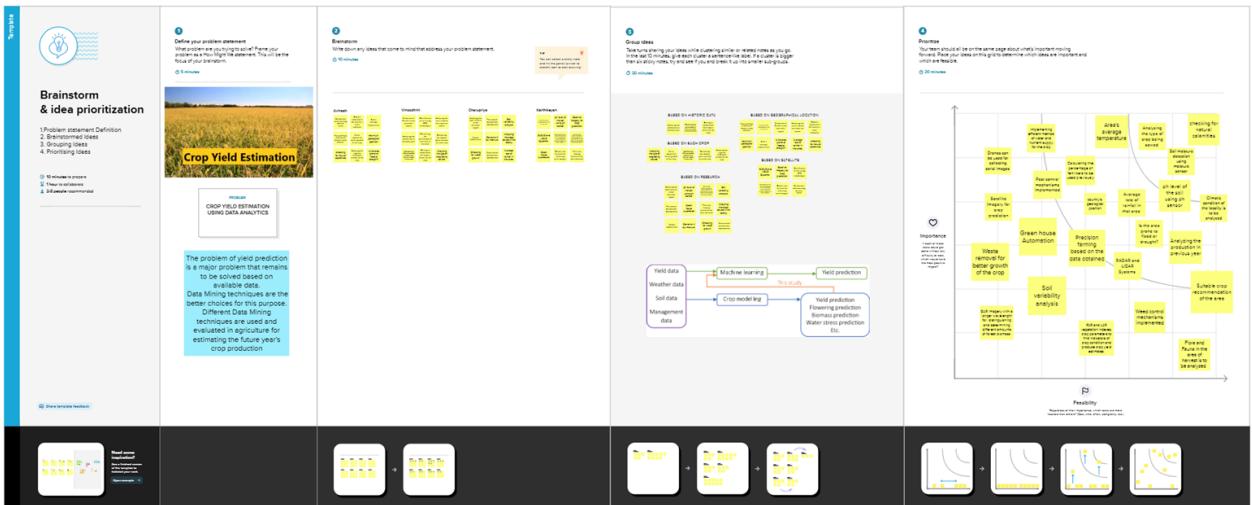
Gain insight and understanding on solving customer problems.

1

Build empathy and keep your focus on the user by putting yourself in their shoes.



3.2 Ideation & Brainstorming:



3.3 Proposed Solution:

Proposed Solution

S.No.	Parameter	Description
1.	Problem Statement (Problem to be solved)	<p>CROP YIELD PREDICTION ANALYSIS:</p> <p>soil quality analysis to achieve high crop yield through out technology solution. To increase quality and yields, it is crucial to understand the current nutrient levels of the soil. Analyze pH to check for soil acidity and alkalinity, Conductivity to determine optimised fertilizer usage as well as Sodium, Potassium, Nitrate and Calcium levels.</p> <p>Fertilizer manufactures understand the importance of particle size. It directly affects certain aspects including release rates, fertilizer potency and also hazardous dust generation. To ensure quality and consistency, a minimum frequency of measurements must be made and our Particle Analyzers are ideal for this task.</p>
2.	Idea/Solution description	<p>ICP-OES and LaquaTwin</p> <p>ICP-OES is an analytical technique that is widely utilized throughout the agricultural industry and within research and development institutions. It is an ideal agriculture analysis technique to determine major and minor elements in soils and plants as well as detecting heavy metal contents. Our Ultima has been used in many crop science institutions to undertake such important content detection.</p>

3.	Novelty/Uniqueness	<p>How ICP-OES is used:</p> <ul style="list-style-type: none"> • Drinking water quality and safety • Soil analysis • Environmental impact assessments • Food safety • Pharmaceutical analysis <p>LquaTwin Detection systems used with ICP-OES.</p>
4.	Social Impact/Customer satisfaction	<p>The main objectives of this technique in prediction of crop-yield which can be extremely useful to farmers in planning for harvest and sale of grain harvest. Increasing crop yields is a high priority for growers.</p>
5.	Business Model (Revenue Model)	<p>The introduction of technology into the agricultural sector has led to a major rise in productivity. Technology improvements have made new concepts like precision agriculture and have observed and analysed the several crops that are grown, as well as their area and production rates in various states and districts.</p>
6.	Scalability of the Solution	<p>Production rate averaging seasons. Experence a rise in productivity. Yield average in seasons. Boots the crop productivity.</p>

S.No.	Parameter	Description
1.	Problem Statement (Problem to be solved)	Variations in annual rainfall, average temperature, global increase of atmospheric CO ₂ , and fluctuations in sea levels are some of the major manifestations of climate change, which negatively impact crop yields.
2.	Idea/Solution description	The regression analysis model between historical climatic data and yield data for food crops over the last 30 years in Nepal showed an increase in temperature of approximately 0.02–0.07°C per year in different seasons and a mixed trend in precipitation.
3.	Novelty/Uniqueness	No significant impact of climate variables on yields of all crops the regression analysis revealed negative relationships between maize yield and summer precipitation, between wheat yield and winter minimum temperature, and finally positive relationship was observed between millet yield and summer maximum temperature.

4.	Social Impact/Customer satisfaction	<p>Though the climate changes and global warming, there will be an more yield in crops and grains.</p> <p>The nutrition also be increased by using the estimation of crop yield techniques and analysis.</p>
5.	Business Model (Revenue Model)	<p>Agriculture production and vulnerability.</p> <p>The accurate prediction of crop yield certainly benefits the farmers in choosing the right method to reduce the crop damage and it is used to produce the sustainability of the crops in any kind of environmental temperature.</p> <p>The aspiration of the planned method is to afford transparent, easily accessible, reproducible and for predicting the yield.</p>
6.	Scalability of the Solution	<ul style="list-style-type: none"> • Increased productivity from warmer temperature. • Decreased moisture stress. • Possibility of growing new crops. • Accelerated. • Maturation rates.

S.No.	Parameter	Description
1.	Problem Statement (Problem to be solved)	Lack of mechanisation Some parts of country most of agricultural operations in lager parts are carried on human hands using simple and conventional tools.
2.	Idea/Solution description	The progress has been made for mechanizing agriculture. Strategies and programs in farming have been directed towards the replacement of traditional and inefficient implements by Improved ones, enabling the farmers to own tractors, power tillers, harvesters and other machines.
3.	Novelty/Uniqueness	The agriculture machines as also be developed for large industrial base. Strenuous efforts are made to encourage the farmers to adopt technically Advanced with agricultural equipment.
4.	Social Impact/Customer satisfaction	Time consumption in yield analysis. Cover crop and integrated pest management(IPM) have shown excellent result in increasing the profitability of farming in long-term.

		<p>Yield production encourages the farmers.</p> <p>Economize the agricultural production process.</p>
5.	Business Model (Revenue Model)	<p>Without risk management Correct practises and plan for the long term gain can be achieved easily and reduces the capital investment.</p>
6.	Scalability of the Solution	<ul style="list-style-type: none"> • Livestock management system. • Productivity of soil and water. • Complies with community norms and meet social expectations. • It will be a profitability of the sustainable farming

3.4 Proposed Solution Fit:

Project Title: Estimate the crop yield using data analytics Project Design Phase-I - Solution Fit Template

Define CS, fit into CC	<p>1. CUSTOMER SEGMENT(S) Who is your customer? i.e. working parents of 0-5 y.o. kids</p> <p>CS</p> <p>This project is mainly for the farmers who are looking for the best and profitable yield.</p>	<p>6. CUSTOMER CONSTRAINTS What constraints prevent your customers from taking action or limit their choices of solutions? i.e. spending power, budget, no cash, network connection, available devices</p> <p>CC</p> <p>The estimation may be wrong at rare cases.</p>	<p>5. AVAILABLE SOLUTIONS Which solutions are available to the customers when they face the problem? What do they need to get the job done? What have they tried in the past? What pros & cons do these solutions have? i.e. pen and paper is an alternative to digital notetaking</p> <p>AS</p> <p>This project will aims at accurate estimation,using natural resources,GIVING appropriate alerts,etc.,</p>
Focus on J&P, tap into BE, understand RC	<p>2. JOBS-TO-BE-DONE / PROBLEMS Which jobs-to-be-done (or problems) do you address for your customers? There could be more than one; explore different sides.</p> <p>J&P</p> <p>As there is no high yield for farmers these days, this project will help them in a great way.</p>	<p>9. PROBLEM ROOT CAUSE What is the real reason that this problem exists? What is the back story behind the need to do this? i.e. customers have to do it because of the change in regulations.</p> <p>RC</p> <p>Crop production is the important source of income of farmers and india.</p>	<p>7. BEHAVIOUR What does your customer do to address the problem and get the job done? Directly related: find the right solar panel installer; calculate usage and benefits; Indirectly associated: customers spend free time on volunteering work (i.e. Greenpeace)</p> <p>BE</p> <p>Existing solutions will estimate with low accuracy at sometimes but also helpful for farmers.</p>
Identify strong TR & EM	<p>3. TRIGGERS What triggers customers to act? i.e. seeing their neighbour installing solar panels; reading about a more efficient solution in the news. For yielding good crops and high income, correct estimation of crop is needful.</p> <p>TR</p> <p>Agriculture with new perspective; agriculture with difference.</p>	<p>10. YOUR SOLUTION If you are working on an existing business, write down your current solution first, fill in the canvas, and check how much it fits really. If you are working on a new business proposition, then keep it blank until you fill in the canvas and come up with a solution that fits within customer limitations, solves a problem and matches customer behaviour.</p> <p>SL</p> <p>Estimating the crop yield in a field for farmers to increase their yield and revenue.</p>	<p>8. CHANNELS OF BEHAVIOUR 8.1 ONLINE What kind of actions do customers take online? Extract online channels from #7</p> <p>CH</p> <p>8.2 OFFLINE What kind of actions do customers take offline? Extract offline channels from #7 and use them for customer development.</p> <p>OF</p> <p>Farmers are the customers.</p>

4. REQUIREMENT ANALYSIS:

4.1 Functional Requirement:

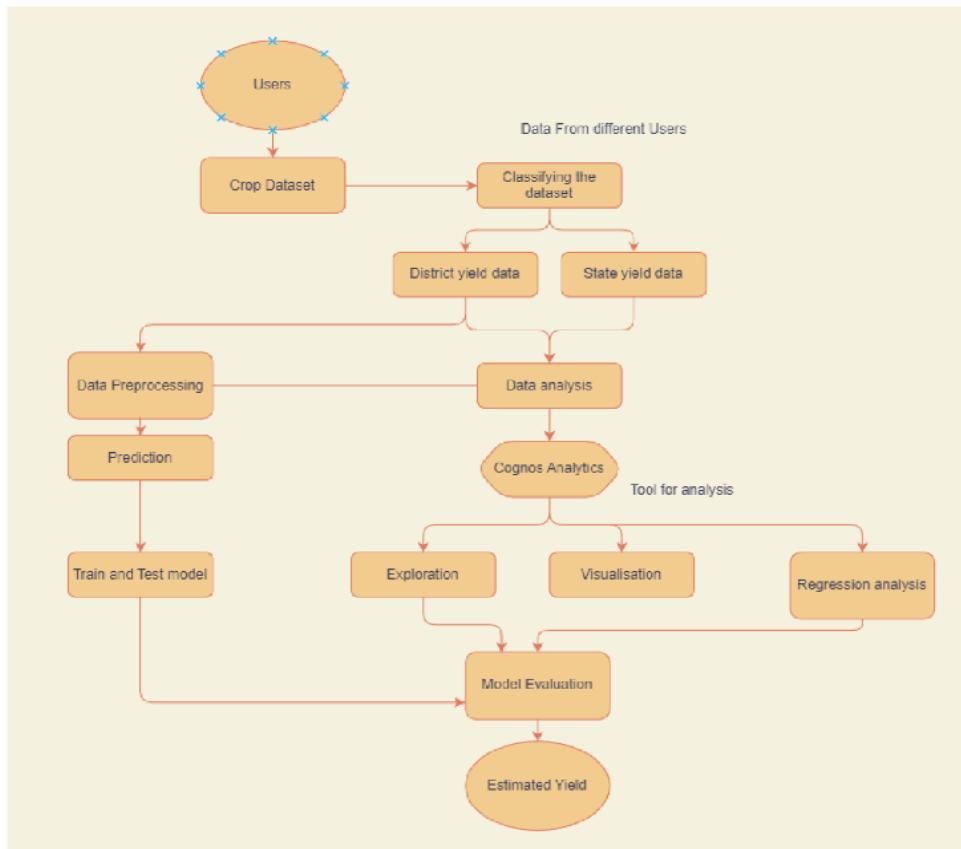
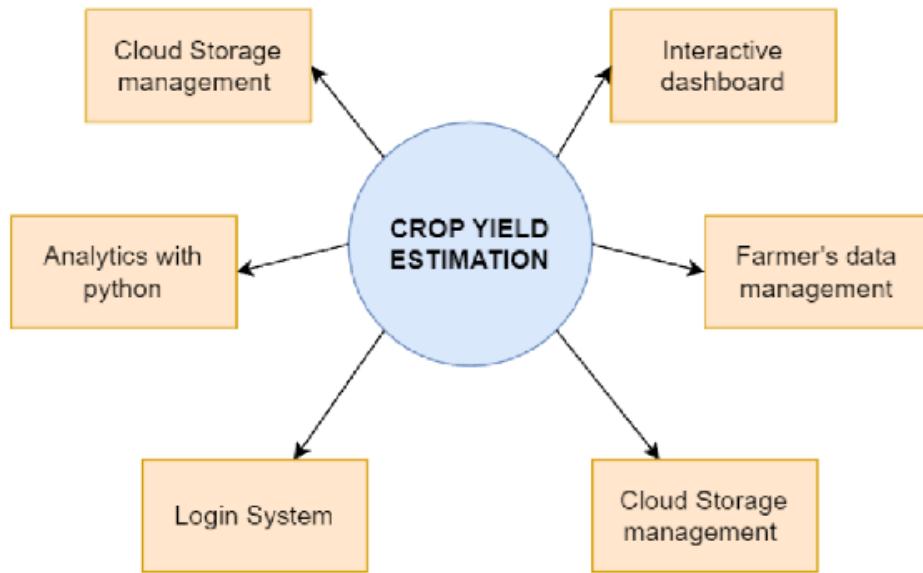
FR No.	Functional Requirement (Epic)	Sub Requirement (Story / Sub-Task)
FR-1	User Requirement	Knowledge of seeds ,crops ,mechanism ,soil ,climate & agriculture science. Right use of resources like soil and water. Time management .Market demand drive production.
FR-2	User Business rules	Three laws - the farmers produce trade and commerce, the farmers agreement of price assurance and farm services act and the essential commodities act
FR-3	User Factors	Crop prediction is highly sensitive to climate. It is affected by long-term trend in average rainfall and temperature, interannual climate variability, shocks during specific phenological stages and extreme weather events.
FR-4	Registration	Registering a new user through registration Crop yield estimates constitute a particular important productivity metric, both an aggregate level as well as in plot-level productivity analysis and impact evaluations of new technologies and policy interventions.
FR-5	User Objectives	Formulation and implementation of policies and programmes aimed at achieving rapid agricultural growth through optimum utilization of land, water, soil and plant resources of the state.
FR-6	User Improvement	It becomes necessary to increase the crop variety to produce disease-resistance off springs of the crops. It also helps in providing better and superior varieties based on the quality and quantity of the yield.
FR-7	Estimation and Analysis	Prediction of crop yield is made in estimation and the data is analysed to gain useful insights

4.2 Non-Functional requirements:

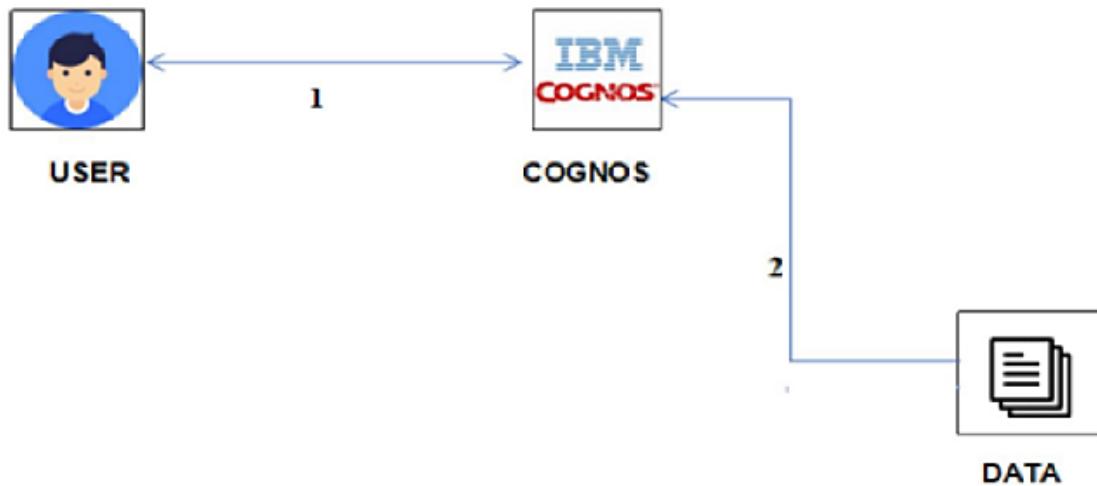
FR No.	Non-Functional Requirement	Description
NFR-1	Usability	To empower farmers and to increase the productivity there is need to provide the best dissemination tool for their farming activities.
NFR-2	Security	The developed ICT agriculture tools focus on very important agricultural services such as crop detection ,crop predictor will help farmers to make decision in future.
NFR-3	Reliability	This will remove multilingual issues and bridge the gap between farmers and technology. Effective tool that all farmers can use for management of all kind of crops
NFR-4	Performance	Multiple technologies and services that will Improve the usability in agricultural activities.
NFR-5	Availability	Both website and mobile application Interface and developed in local language and the content is available in localized language
NFR-6	Scalability	i)Increased productivity from warm temperature ii)Decreased moisture stress iii)Possibility of growing new crops iv)Productivity of soil and water

5. PROJECT DESIGN:

5.1 Data Flow Diagrams:



5.2 Solution & Technical Architecture:



5.3 User Stories:

User Stories

Use the below template to list all the user stories for the product.

User Type	Functional Requirement (Epic)	User Story Number	User Story / Task	Acceptance criteria	Priority	Release
Customer (Mobile user and Laptop users)	Registration	USN-1	As a user, I can register for the application by entering my email, password, and confirming my password.	I can access my account / dashboard	High	Sprint-1
		USN-2	As a user, I will receive confirmation email once I have registered for the application	I can receive confirmation email & click confirm	High	Sprint-1
		USN-3	As a user, I can register for the application through Facebook	I can register & access the dashboard with Facebook Login	Low	Sprint-2
		USN-4	As a user, I can register for the application through Gmail		Medium	Sprint-1
	Login	USN-5	As a user, I can log into the application by entering email & password		High	Sprint-1
	Dashboard	USN-6	Can use the methods provided in the dashboard		Medium	Sprint-1
	Invest	USN-7	With help of desired results obtained from application ,making profit or loss	Gain or Loss	High	Sprint-2
Administrator	Updating data		Collecting the data and storing it	Checking and updating dataset	High	Sprint-1
Customer (Web User)	Accessing the resources	USN -8	Using my own credentials for accessing the data	These resources cannot be accessed by others but only me	High	Sprint -1
	Satellite Visioning	USN -9	Having a view with geographic data		Medium	Sprint-2
Customer tools	Tools	USN -10	Analysis is performed by tools like cognos analytics	Ease of accessing the tools	High	Sprint 2

6. PROJECT PLANNING & SCHEDULING:

6.1 Sprint Planning & Estimation:

Product Backlog, Sprint Schedule, and Estimation (4 Marks)

Sprint	Functional Requirement (Epic)	User Story Number	User Story / Task	Story Points	Priority	Team Members
Sprint-1	Registration	USN-1	As a user, I can register for the application by entering my email, password, and confirming my password.	2	High	Avinash R
Sprint-1		USN-2	As a user, I will receive confirmation email once I have registered for the application	1	High	Charupriya S
Sprint-2		USN-3	As a user, I can register for the application through Facebook	2	High	Karthikeyan B
Sprint-1		USN-4	As a user, I can register for the application through Gmail	1	Medium	Vinoodhini D
Sprint-1	Login	USN-5	As a user, I can log into the application by entering email & password	1	High	Avinash R

Sprint-2	Dashboard	USN-6 Can use the methods provided in the dashboard	2	Medium	Karthikeyan B
Sprint-2		USN-7 With help of desired results obtained from application,making profit or loss and Collecting the data and storing it	1	High	Vinoodhini
Sprint 3		USN-8 Using my own credentials for accessing the data	1	High	Karthikeya
Sprint-3	Visualization	USN-9 Having a view with geographic data	2	High	Avinash R
Sprint 4		USN 10 Analysis is performed by tools like cognos analytics	1	High	Charupriya

6.2 Sprint Delivery Schedule:

Project Tracker, Velocity & Burndown Chart: (4 Marks)

Sprint	Total Story Points	Duration Sprint Start Date Sprint End Date (Planned)	Story Points Completed (as on Planned End Date)	Sprint Release Date (Actual)
Sprint-1	20	6 Days 24 Oct 2022 29 Oct 2022	20	29 Oct 2022
Sprint-2	20	6 Days 31 Oct 2022 05 Nov 2022	20	05 Nov 2022
Sprint-3	20	6 Days 07 Nov 2022 12 Nov 2022	20	12 Nov 2022
Sprint-4	20	6 Days 14 Nov 2022 19 Nov 2022	20	19 Nov 2022

6.3 Reports from JIRA



Roadmap

The Roadmap view shows a Gantt chart for the project. It includes columns for Oct, Nov, Dec, Jan '22, and Feb '22. The chart displays tasks grouped by sprint, with specific issue keys and descriptions. A legend indicates task status: Done (orange), In Review (green), To Do (blue), and In Progress (purple).

Sprint	Oct	Nov	Dec	Jan '22	Feb '22
ECYUDA-15 Registration	ECYUDA-1 As a user, I can register for the application through Facebook	ECYUDA-2 As a user, I will receive confirmation email once I have registered for the application	ECYUDA-3 As a user, I will receive confirmation email once I have registered for the application	ECYUDA-4 As a user, I can register for the application through Gmail	
ECYUDA-16 Login	ECYUDA-5 As a user, I can log into the application by entering email & password	ECYUDA-6 As a user, I can log into the application by entering email & password	ECYUDA-7 As a user, I can log into the application by entering email & password	ECYUDA-8 As a user, I can log into the application by entering email & password	
ECYUDA-17 Dashboard	ECYUDA-9 Can use the methods provided in the dashboard for visualization	ECYUDA-10 Can use the methods provided in the dashboard for visualization	ECYUDA-11 With help of desired results obtained from application making profit or loss. Collecting the data and storing it	ECYUDA-12 Using my own credentials for accessing the data for an application	
ECYUDA-18 Visualization	ECYUDA-13 Having a view with geographical data	ECYUDA-14 Analysis in the dashboard			

All sprints

The All sprints view provides a detailed look at the backlog across multiple sprints. It shows four columns: To Do (2 issues), In Progress (1 issue), In Review (2 issues), and Done (1 issue). Each card contains a brief description and the corresponding Jira issue key.

Sprint	To Do	In Progress	In Review	Done
ECYUDA Sprint 1 24 Oct - 29 Oct (4 issues)	ECYUDA-6 As a user, I can register for the application through Facebook	ECYUDA-9 As a user, I can log into the application by entering email & password	ECYUDA-2 As a user, I will receive confirmation email once I have registered for the application	ECYUDA-1 As a user, I can register for the application through Gmail
ECYUDA Sprint 2 31 Oct - 5 Nov (3 issues)	ECYUDA-10 Can use the methods provided in the dashboard for visualization	ECYUDA-11 With help of desired results obtained from application making profit or loss. Collecting the data and storing it		
ECYUDA Sprint 3 7 Nov - 12 Nov (2 issues)	ECYUDA-12 Using my own credentials for accessing the data for an application	ECYUDA-13 Having a view with geographical data		

Backlog

The Backlog view lists issues grouped by sprint. It includes a 'Create Issue' button and a 'Complete sprint' button for each sprint section.

Sprint	Issues
ECYUDA Sprint 1 24 Oct - 29 Oct (4 issues)	ECYUDA-6 As a user, I can register for the application through Facebook ECYUDA-9 As a user, I can log into the application by entering email & password ECYUDA-2 As a user, I will receive confirmation email once I have registered for the application ECYUDA-1 As a user, I can register for the application through Gmail
ECYUDA Sprint 2 31 Oct - 5 Nov (3 issues)	ECYUDA-10 Can use the methods provided in the dashboard for visualization ECYUDA-11 With help of desired results obtained from application making profit or loss. Collecting the data and storing it
ECYUDA Sprint 3 7 Nov - 12 Nov (2 issues)	ECYUDA-12 Using my own credentials for accessing the data for an application ECYUDA-13 Having a view with geographical data

7. CODING & SOLUTIONING

7.1 Feature 1:

Exploratory Data Analysis:

```
df
```

	State_Name	District_Name	Crop_Year	Season	Crop	Area	Production
0	Andaman and Nicobar Islands	NICOBARS	2000	Kharif	Arecanut	1254.0	2000.0
1	Andaman and Nicobar Islands	NICOBARS	2000	Kharif	Other Kharif pulses	2.0	1.0
2	Andaman and Nicobar Islands	NICOBARS	2000	Kharif	Rice	102.0	321.0
3	Andaman and Nicobar Islands	NICOBARS	2000	Whole Year	Banana	176.0	641.0
4	Andaman and Nicobar Islands	NICOBARS	2000	Whole Year	Cashewnut	720.0	165.0
...
246086	West Bengal	PURULIA	2014	Summer	Rice	306.0	801.0
246087	West Bengal	PURULIA	2014	Summer	Sesamum	627.0	463.0
246088	West Bengal	PURULIA	2014	Whole Year	Sugarcane	324.0	16250.0
246089	West Bengal	PURULIA	2014	Winter	Rice	279151.0	597899.0
246090	West Bengal	PURULIA	2014	Winter	Sesamum	175.0	88.0

246091 rows × 7 columns

```
[ ] df.isnull().sum()
```

```
State_Name      0
District_Name   0
Crop_Year       0
Season          0
Crop            0
Area            0
Production      3730
dtype: int64
```

```
[ ] df.dropna(inplace=True)
```

```
▶ df.isnull().sum()
```

```
State_Name      0
District_Name   0
Crop_Year       0
Season          0
Crop            0
Area            0
Production      0
dtype: int64
```

```
df = df[df['Production'] != 0]
```

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 238838 entries, 0 to 246090
Data columns (total 7 columns):
 #   Column           Non-Null Count   Dtype  
--- 
 0   State_Name       238838 non-null    object  
 1   District_Name    238838 non-null    object  
 2   Crop_Year        238838 non-null    int64   
 3   Season           238838 non-null    object  
 4   Crop              238838 non-null    object  
 5   Area              238838 non-null    float64 
 6   Production        238838 non-null    float64 
dtypes: float64(2), int64(1), object(4)
memory usage: 14.6+ MB
```

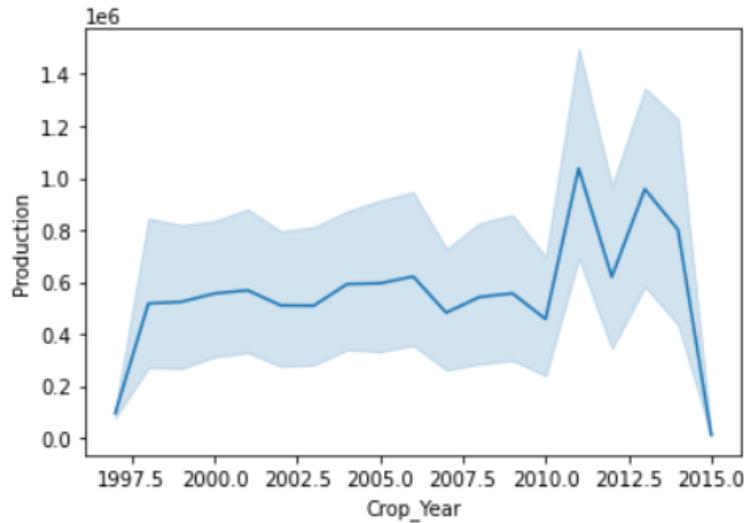
Data Visualisation

```
sns.lineplot(data["Crop_Year"], data["Production"])
```

```
/usr/local/lib/python3.7/dist-packages/seaborn/_decorators.py:43:
```

```
Pass the following variables as keyword args: x, y. From version 0
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7f43aeeeb950>
```

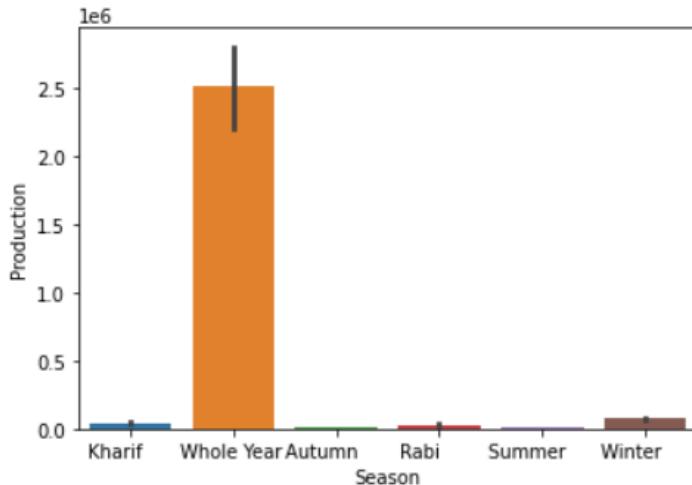


```
sns.barplot(data["Season"],data["Production"])
```

```
/usr/local/lib/python3.7/dist-packages/seaborn/_decorators.py:4
```

```
Pass the following variables as keyword args: x, y. From versio
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7f43ae7fe750>
```



```
data.groupby("Season",axis=0).agg({"Production":np.sum})
```

Season	Production
Autumn	6.441377e+07
Kharif	4.029970e+09
Rabi	2.051688e+09
Summer	1.706579e+08
Whole Year	1.344248e+11
Winter	4.345498e+08

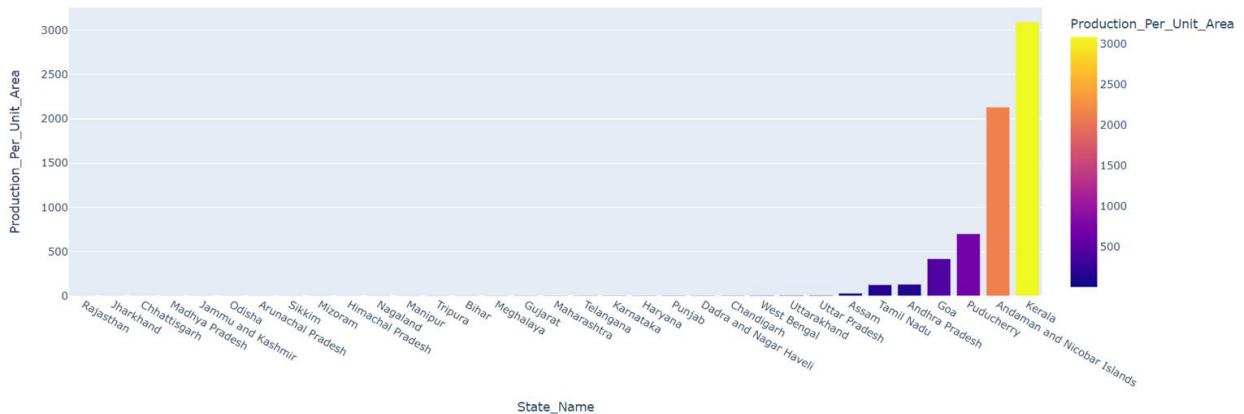
```
data["Crop"].value_counts()[:5]
```

Rice	15078
Maize	13723
Moong(Green Gram)	10037
Urad	9669
Groundnut	8732

Name: Crop, dtype: int64

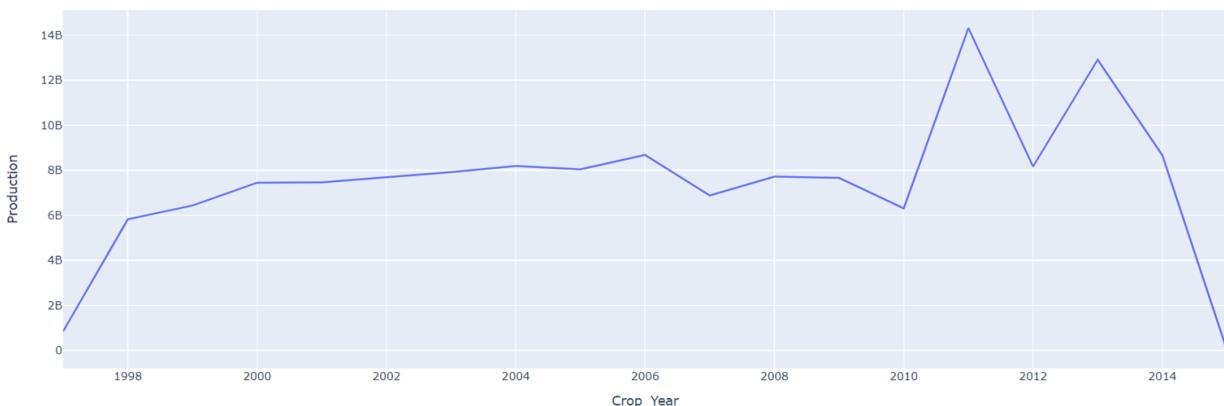
#PRODUCTIVITY OF DIFFERENT STATES

```
temp = data.groupby('State_Name')[ 'Area', 'Production'].sum().reset_index()
temp['Production_Per_Unit_Area'] = temp['Production']/temp['Area']
temp = temp.sort_values(by='Production_Per_Unit_Area')
px.bar(temp, 'State_Name', 'Production_Per_Unit_Area', color='Production_Per_Unit_Area', )
```

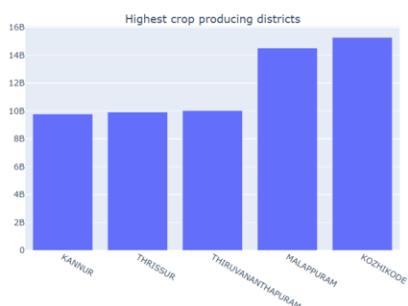


#OVERALL PRODUCTION THROUGH YEARS

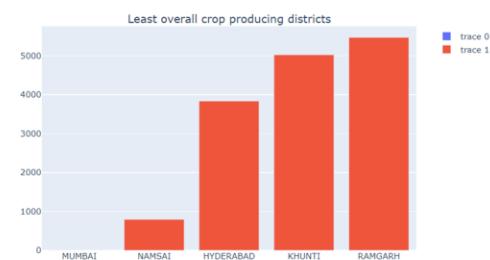
```
temp = data.groupby(by='Crop_Year')[ 'Production'].sum().reset_index()
px.line(temp, 'Crop_Year', 'Production')
```



Highest crop producing districts



Least overall crop producing districts



7.2 Feature 2:

Algorithm & Metric Evaluation:

• SPLIT DATA INTO TRAINING AND TESTING SET

```
[ ] x = f_data.drop("Production",axis=1)
y = f_data[["Production"]]

[ ] from sklearn.model_selection import train_test_split
x_train,x_test,y_train,y_test = train_test_split(x,y,test_size=0.33, random_state=42)

[ ] print("x_train :",x_train.shape)
print("x_test :",x_test.shape)
print("y_train :",y_train.shape)
print("y_test :",y_test.shape)

x_train : (160021, 146)
x_test : (78817, 146)
y_train : (160021, 1)
y_test : (78817, 1)
```

MODEL

```
[ ] from sklearn.linear_model import LinearRegression
regression = LinearRegression()
regression.fit(x_train, y_train)

y_pred = regression.predict(x_test)

[ ] from sklearn.metrics import mean_squared_error, r2_score
mean_squared_error(y_test,y_pred)
r2_score(y_test,y_pred)
```

1.0

MODEL PERFORMANCE METRICS

```
[ ] from sklearn.metrics import confusion_matrix, accuracy_score, precision_score, recall_score

[ ] cutoff = 0.7                      # decide on a cutoff limit
y_pred_classes = np.zeros_like(y_pred)    # initialise a matrix full with zeros
y_pred_classes[y_pred > cutoff] = 1

[ ] y_test_classes = np.zeros_like(y_pred)
y_test_classes[y_test > cutoff] = 1

▶ # CONFUSION MATRIX
conf_matrix=confusion_matrix(y_test_classes, y_pred_classes)
conf_matrix

⇒ array([[ 186,    10],
       [     0, 78621]])

[ ] # RECALL
```

1.0

```

TP = conf_matrix[1][1]
TN = conf_matrix[0][0]
FP = conf_matrix[0][1]
FN = conf_matrix[1][0]
print('True Positives:', TP)
print('True Negatives:', TN)
print('False Positives:', FP)
print('False Negatives:', FN)

# calculate accuracy
conf_accuracy = (float(TP+TN) / float(TP + TN + FP + FN))

# calculate the sensitivity
conf_sensitivity = (TP / float(TP + FN))

# calculate the specificity
conf_specificity = (TN / float(TN + FP))

# calculate precision
conf_precision = (TN / float(TN + FP))

# calculate recall
recall = recall_score(y_test_classes, y_pred_classes)

# calculate f_1 score
conf_f1 = 2 * ((conf_precision * conf_sensitivity) / (conf_precision + conf_sensitivity))

print('*'*50)
print(f'Accuracy: {round(conf_accuracy,2)}')
print(f'Sensitivity: {round(conf_sensitivity,2)}')
print(f'Specificity: {round(conf_specificity,2)}')
print(f'Precision: {round(conf_precision,2)}')
print(f'Recall: {recall_score(y_test_classes, y_pred_classes)}')
print(f'f_1 Score: {round(conf_f1,2)}')

```

```

True Positives: 78621
True Negatives: 186
False Positives: 10
False Negatives: 0
-----
Accuracy: 1.0
Sensitivity: 1.0
Specificity: 0.95
Precision: 0.95
Recall: 1.0
f_1 Score: 0.97

```

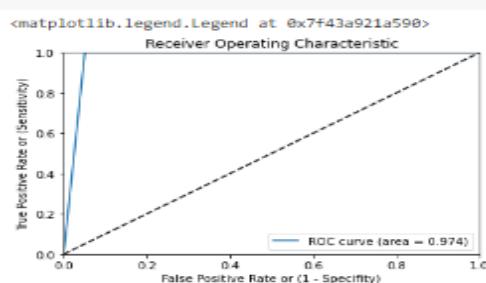
```

from sklearn.metrics import roc_curve
from sklearn.metrics import auc

fpr, tpr, thresholds = roc_curve(y_test_classes, y_pred_classes)
roc_auc = auc(fpr, tpr)

plt.plot(fpr, tpr, label='ROC curve (area = %0.3f)' % roc_auc)
plt.plot([0, 1], [0, 1], 'k--') # random predictions curve
plt.xlim([0.0, 1.0])
plt.ylim([0.0, 1.0])
plt.xlabel('False Positive Rate or (1 - Specificity)')
plt.ylabel('True Positive Rate or (Sensitivity)')
plt.title('Receiver Operating Characteristic')
plt.legend(loc="lower right")

```



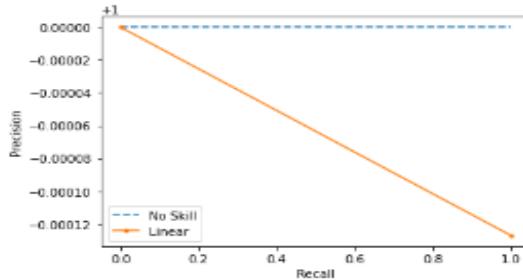
```

# precision-recall curve and f1
from sklearn.datasets import make_classification
from sklearn.linear_model import LogisticRegression
from sklearn.model_selection import train_test_split
from sklearn.metrics import precision_recall_curve
from sklearn.metrics import f1_score
from sklearn.metrics import auc
from matplotlib import pyplot

lr_precision, lr_recall, _ = precision_recall_curve(y_test_classes, y_pred_classes)
lr_f1, lr_auc = f1_score(y_test_classes, y_pred_classes), auc(lr_recall, lr_precision)
# summarize scores
print('Linear: f1=%f auc=%f' % (lr_f1, lr_auc))
# plot the precision-recall curves
no_skill = len(y_test[y_test==1]) / len(y_test)
pyplot.plot([0, 1], [no_skill, no_skill], linestyle='--', label='No Skill')
pyplot.plot(lr_recall, lr_precision, marker='.', label='Linear')
# axis labels
pyplot.xlabel('Recall')
pyplot.ylabel('Precision')
# show the legend
pyplot.legend()
# show the plot
pyplot.show()

```

Linear: f1=1.000 auc=1.000



8. TESTING

8.1 Test Cases:

Test Case Details													
Test Case ID	Feature Type	Component	Test Scenario	Pre-Requisite	Steps To Execute	Test Data	Expected Result	Actual Result	Status	Comments	TC for Automation(Y/N)	BU G ID	Executed By
HomePage_TC_001	Functional	Home Page	Verify user is able to see the Login/Signup popup when user clicked on Login Button in the Homepage	Account to login or data to signup	1.Enter URL and click go 2.Click on Login Button 3.Verification of login/Signup popup displayed or not	http://localhost:3000	Login/Signup popup should display	Working as expected	Pass		Y		Avinash.R
LoginPage_TC_002		Login Page	Verify the UI elements in Login/Signup popup	Credentials to login	1.Enter URL and click go 2.Click on My Account dropdown button 3.Verification of login/Signup popup with below UI elements: a.email text box b.password text box c.Login button d.New customer? Create account link	http://localhost:3000	Application should show below UI elements: a.email text box b.password text box c.Login button with orange colour d.New customer? Create account link e.Last password? Recovery password link	Working as expected	Pass		Y		Karthikyan
LoginPage_TC_003	Functional	Login page	Verify user is able to log into application with Valid credentials	Account to login or data to signup	1.Enter URL(http://localhost:3000/) and click go 2.Click on My Account dropdown button 3.Enter Valid username/email in Email text box 4.Enter valid password in password text box 5.Click on login button	Username: avinash@gmail.com password: Testing123	User should navigate to user account homepage	Working as expected	Pass		Y		Vinoosini D
Test Case Details												4 marks	
Test Case ID	Feature Type	Component	Test Scenario	Pre-Requisite	Steps To Execute	Test Data	Expected Result	Actual Result	Status	Comments	TC for Automation(Y/N)	BU G ID	Executed By
LoginPage_TC_004	Functional	Login page	Verify user is able to log into application with Invalid credentials	Login credentials	1.Enter URL(http://localhost:3000/) and click go 2.Click on My Account dropdown button 3.Enter Invalid username/email in Email text box 4.Enter Invalid password in password text box 5.Click on login button	Username: avinash@gmail.com password: Testing1236786867868676876	Application should show 'Incorrect email or password' validation message.	Working as expected	Pass		Y		Avinash.R
LoginPage_TC_005	Functional	Login page	Verify user is able to log into application with Invalid credentials	Account to login or data to signup	1.Enter URL(http://localhost:3000/) and click go 2.Click on My Account dropdown button 3.Enter Invalid username/email in Email text box 4.Enter Invalid password in password text box 5.Click on login button	Username: karthi@gmail.com password: Testing1236786867868676876	Application should show 'Incorrect email or password' validation message.	Working as expected	Pass		Y		Karthikyan

8.2 User Acceptance Testing:

2. Defect Analysis

This report shows the number of resolved or closed bugs at each severity level, and how they were resolved

Resolution	Severity 1	Severity 2	Severity 3	Severity 4	Subtotal
By Design	9	3	2	3	18
Duplicate	1	0	3	0	4
External	2	3	0	1	6
Fixed	10	2	4	20	36
Not Reproduced	0	0	1	0	1
Skipped	0	0	1	1	2
Won't Fix	0	5	2	1	8
Totals	22	13	13	26	72

3. Test Case Analysis

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

Section	Total Cases	Not Tested	Fail	Pass
Print Engine	7	0	0	7
Client Application	51	0	0	51
Security	2	0	0	2
Outsource Shipping	3	0	0	3
Exception Reporting	9	0	0	9
Final Report Output	4	0	0	4
Version Control	2	0	0	2

9. RESULTS

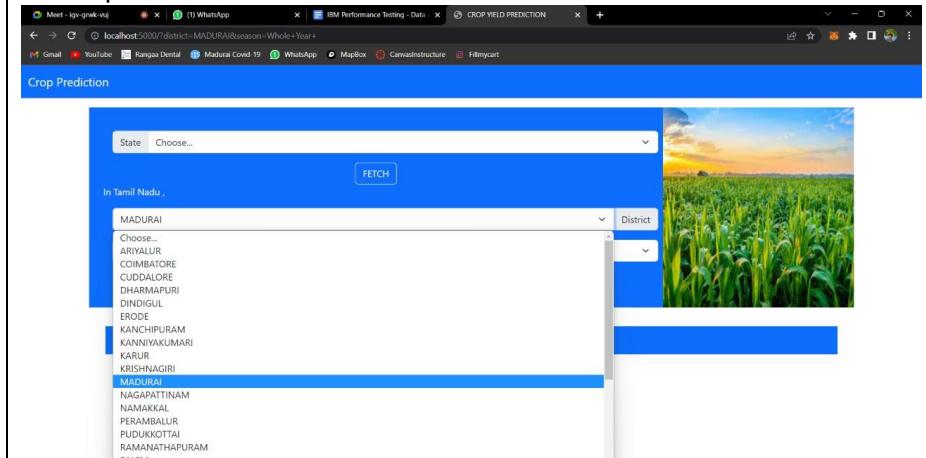
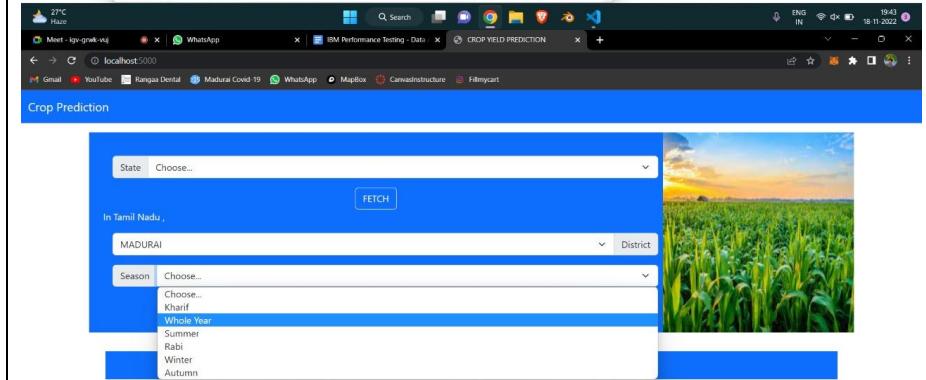
9.1 Performance Test:

Date	10 November 2022
Team ID	PNT2022TMID21477
Project Name	Estimate The Crop Yield Using Data Analytics
Maximum Marks	10 Marks

Model Performance Testing:

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

S.No.	Parameter	Screenshot / Values
1.	Dashboard design	No of Visualizations / Graphs – 13
2.	Data Responsiveness	<p>Yes, the website is responsive completely, that is by resizing the browser window size as per the test scenario.</p> <p>CROP PRODUCTION DATASET</p> <p>The dataset contains 7 rows and 246091 record and dataset contains different state name, different district name, crop year ,crop, area, season and production</p>
3.	Amount Data to Rendered (DB2 Metrics)	<p>To connect IBM Db2 database cloud with cognos analytics:</p> <p>By using IBM Db2 to create Dashboard, Report, Story, Visualization and Exploratory data analytics(EDA) , We have optimized the required data , and it fetches data based on the requirement.</p>
4.	Utilization of Data Filters	Utilization of data filters - 12
5.	Effective User Story	<p>No of Scene Added – 8</p> <p>To create the Registration page of the Website</p> <p>To create the Login page of the Website</p> <p>To create the Dashboard page of the Website</p> <p>To work on the given dataset, Understand the Dataset</p> <p>Load the dataset to Cloud platform then Build the required Visualizations</p> <p>Using the Crop production in Indian dataset, create various graphs and charts to highlight the insights and visualizations.</p> <p>Build a Visualizations to showcase</p> <p>Average Crop Production by Seasons</p> <p>Showcase the Yearly usage of Area in Crop Production</p>

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

VISUALISATIONS:

Crop_Year colored by State_Name

State_Name

Andaman and Nicobar Islands

Assam

Chhattisgarh

Gujarat

Andhra Pradesh

Bihar

Dadra and Nagar Haveli

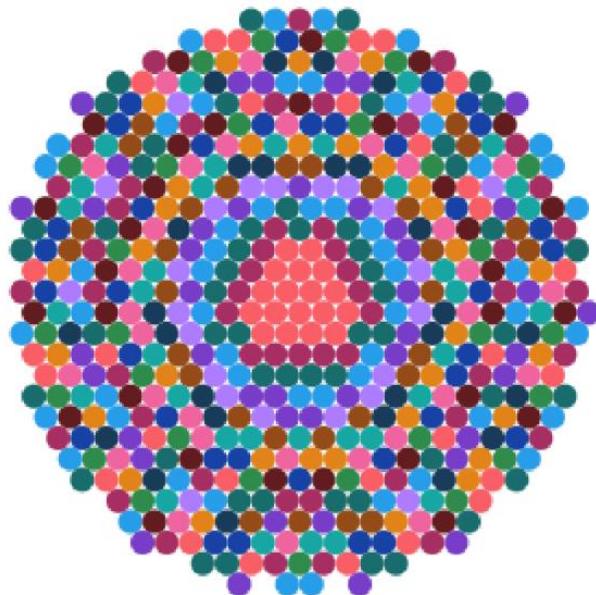
Haryana

Arunachal Pradesh

Chandigarh

Goa

Himachal Pradesh



District_Name for State_Name and Crop_Year



District_Name	Andaman and Ni...	Andhra Pradesh	Arunachal Pradesh	Assam
1997	(no value)	13	13	
1998	(no value)	13	13	
1999	(no value)	13	13	
2000	2	13	13	
2001	2	13	13	
2002	2	13	15	
2003	2	13	15	
2004	2	13	16	
2005	2	13	16	
2006	2	13	16	

District_Name colored by Season

Season

Autumn

Kharif

Rabi

Summer

Whole Year

Winter



Season and Crop

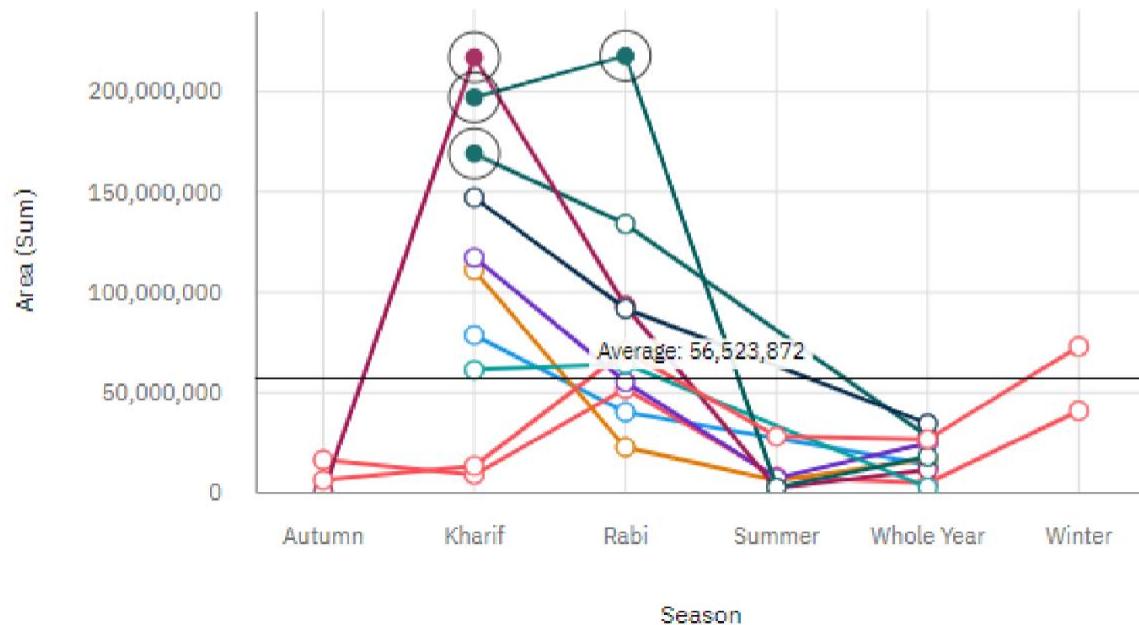
Crop	Season
Apple	Whole Year
Arcanut (Processed)	Whole Year
	Kharif
Arecanut	Rabi
	Whole Year
	Autumn
Arhar/Tur	Kharif
	Rabi
	Summer
	Whole Year
	Winter

Area by Season colored by State_Name

▼ ▲ ⚡

State_Name

- West Bengal
- Uttar Pradesh
- Rajasthan
- Punjab
- Maharashtra
- Madhya Pradesh
- Karnataka
- Gujarat
- Bihar
- Andhra Pradesh



Area and Production for State_Name regions

Area (Sum)



12,522

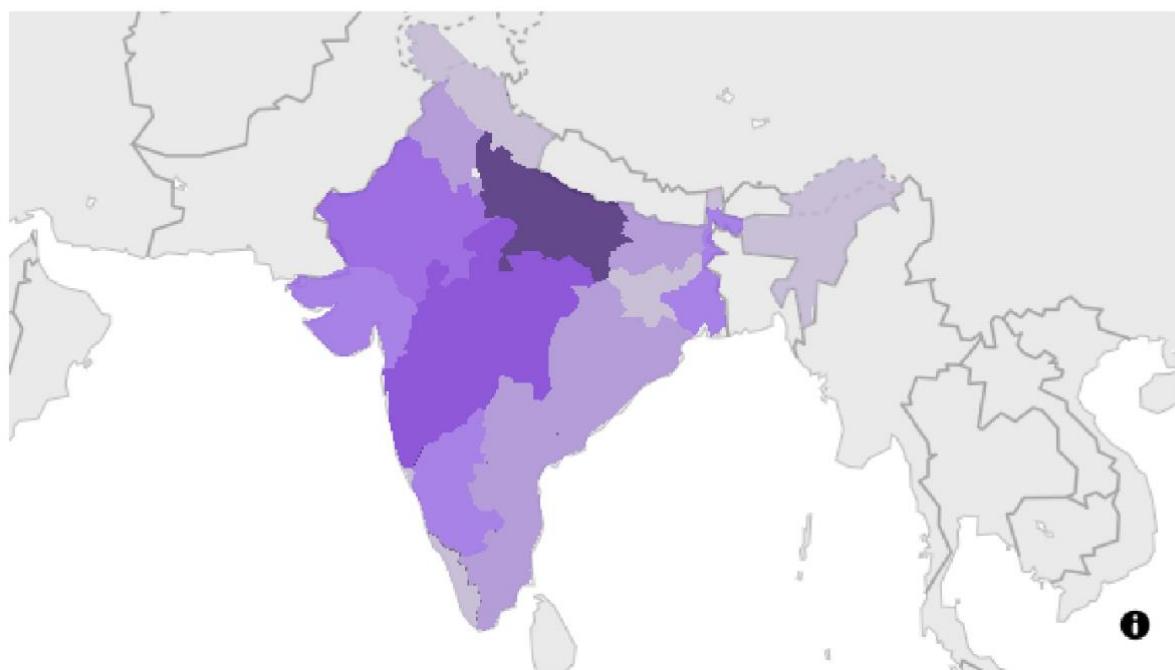
433,631,630

Production (Sum)



63,956.5

97,880,046,376.7



Area and Production by Season

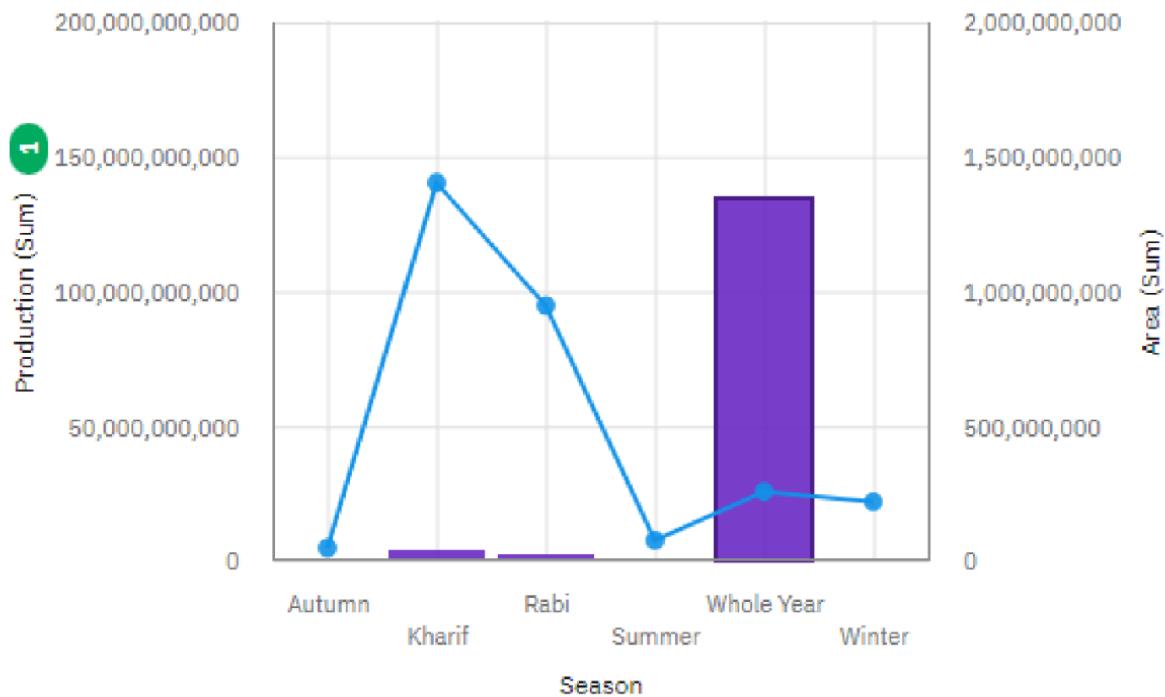


Column

● Production (Sum)

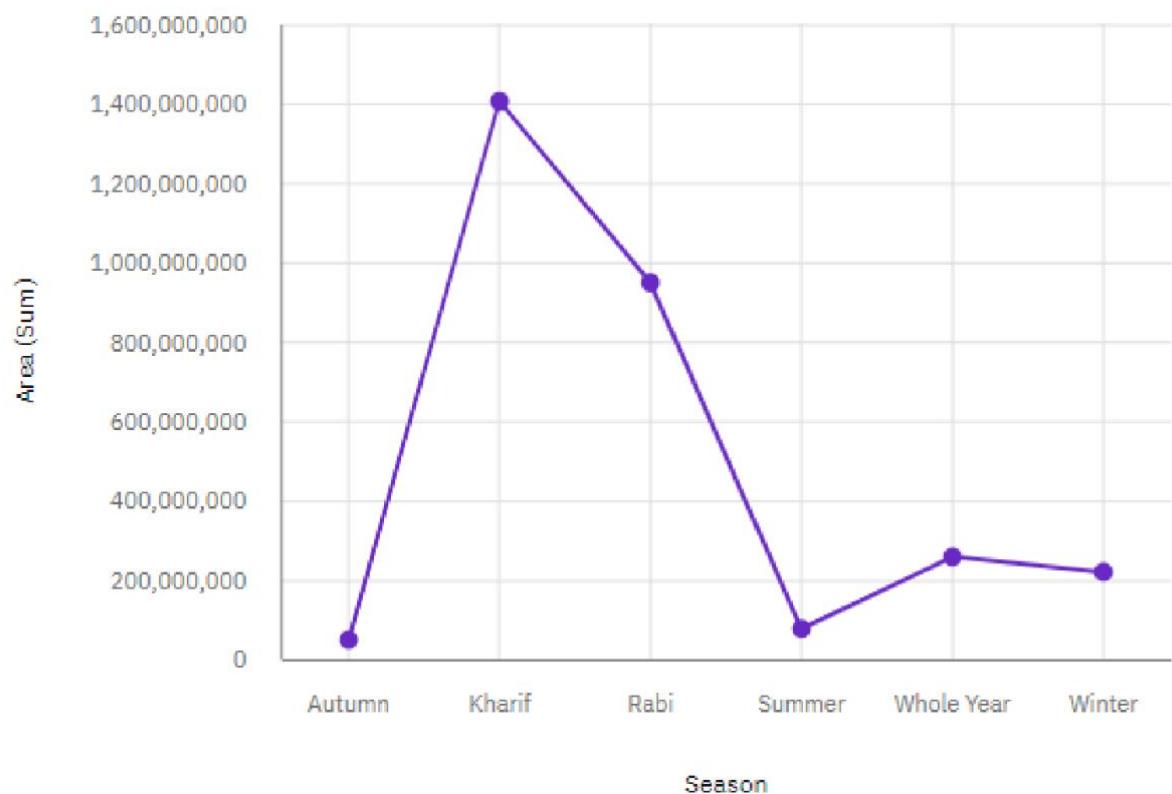
Line

● Area (Sum)



Area by season

Legend



Production for State_Name regions 1

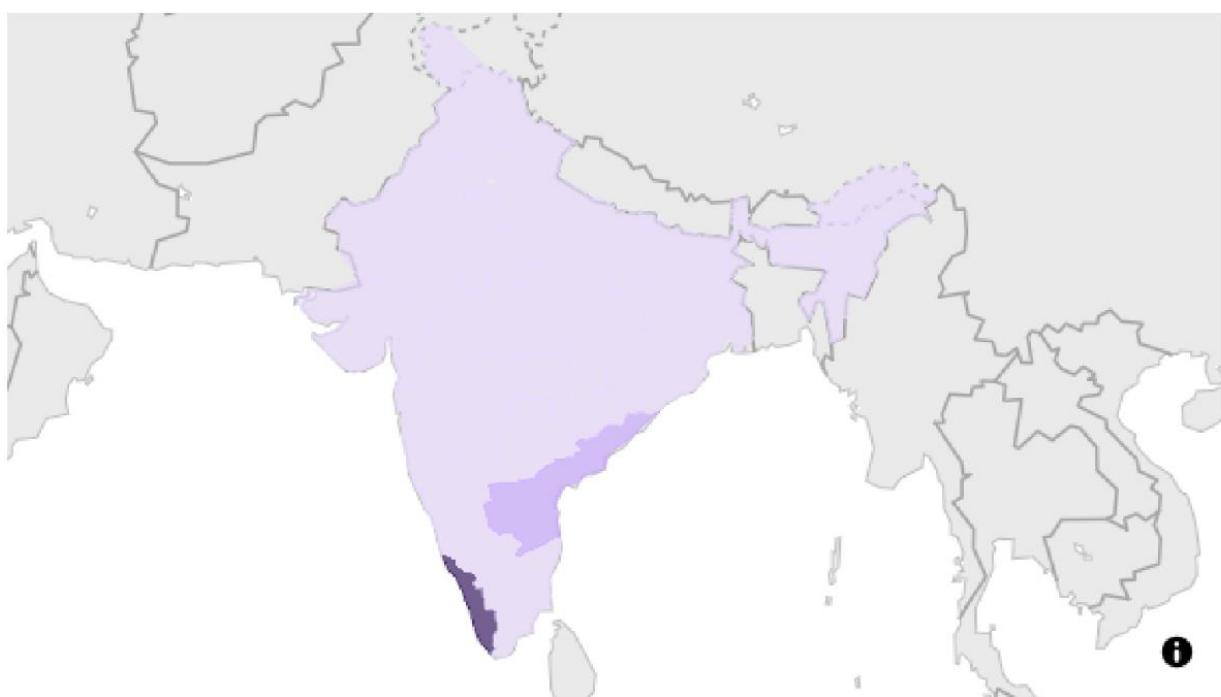
Q

Production (Sum)



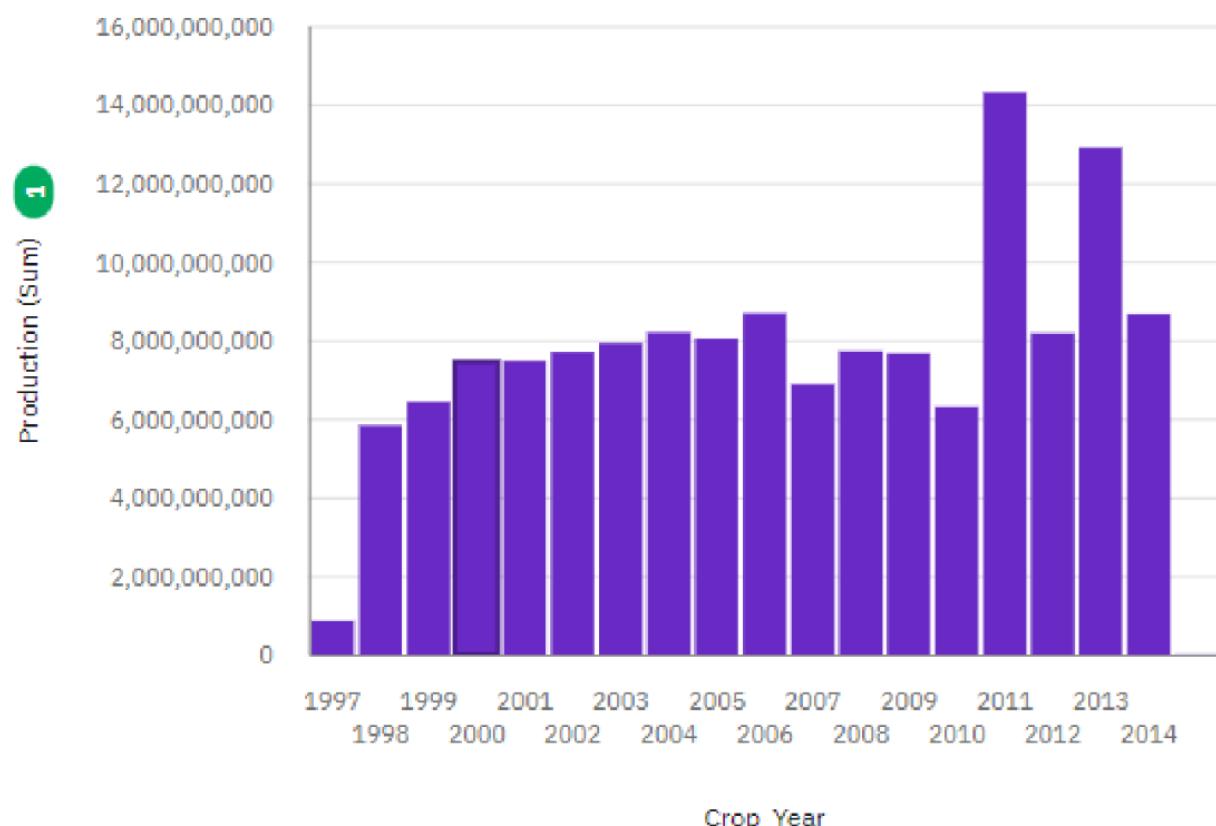
63,956.5

97,880,046,376.7



Production by Crop_Year

Line Q



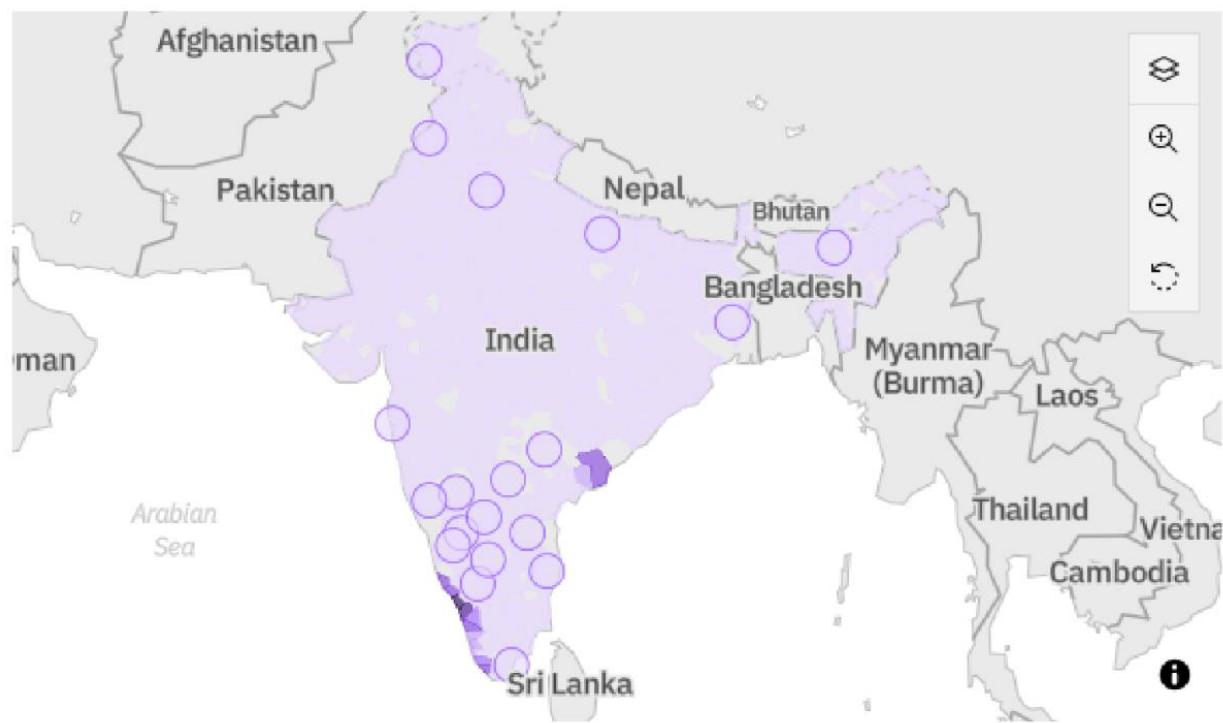
Production for District_Name regions



Production (Sum)



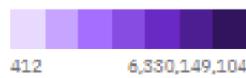
2 16,280,739,862.96



Production by State_Name and Crop_Year 1

Q

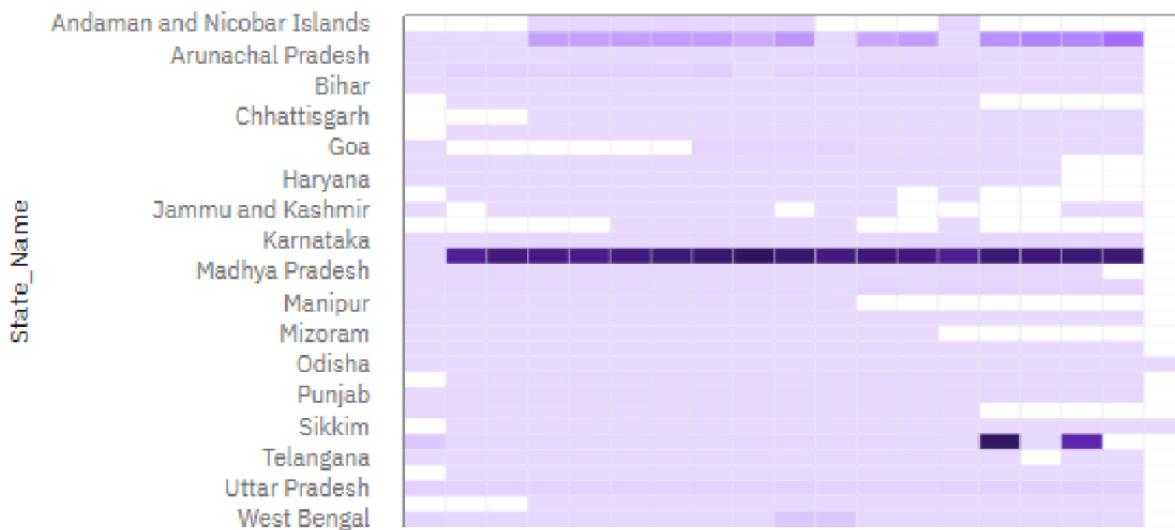
Production (Sum)



412 6,330,149,104

Crop_Year

1998 2000 2002 2004 2006 2008 2010 2012 2014
1997 1999 2001 2003 2005 2007 2009 2011 2013



Production by District_Name, Crop_Year and Area

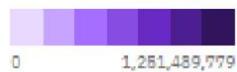
1



Area (Sum)



Production (Sum)

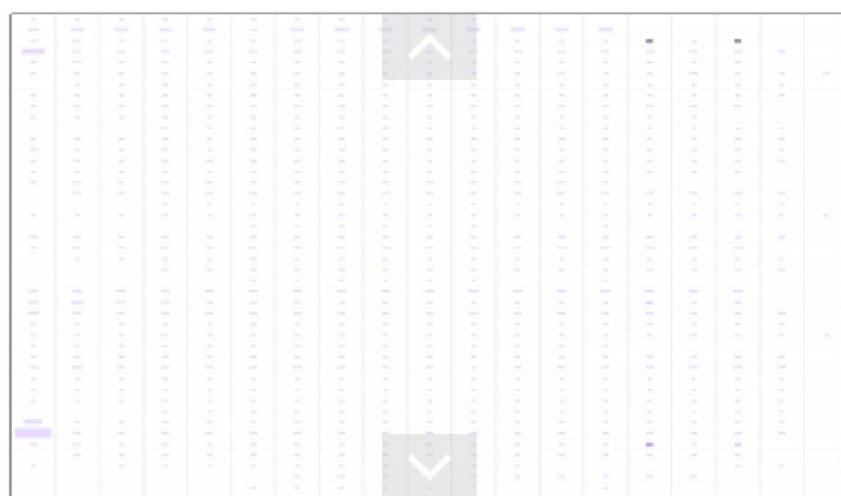


Crop_Year

1998 2000 2002 2004 2006 2008 2010 2012 2014
1997 1999 2001 2003 2005 2007 2009 2011 2013

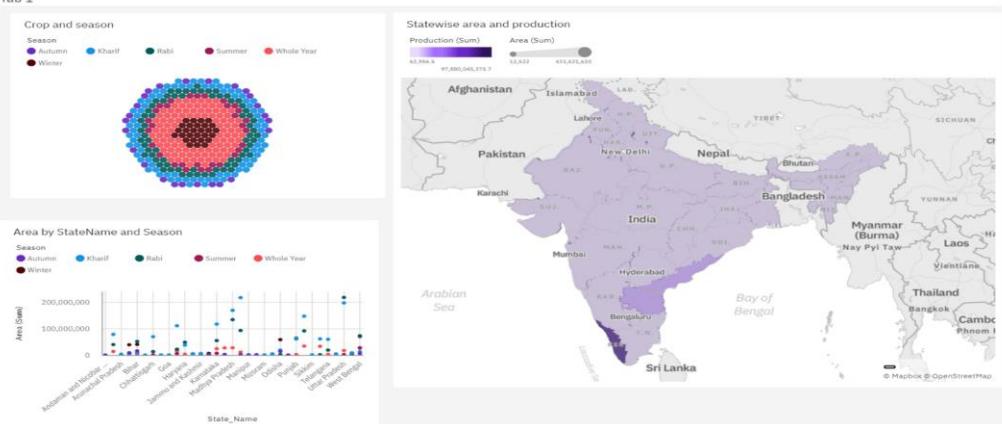
District_Name

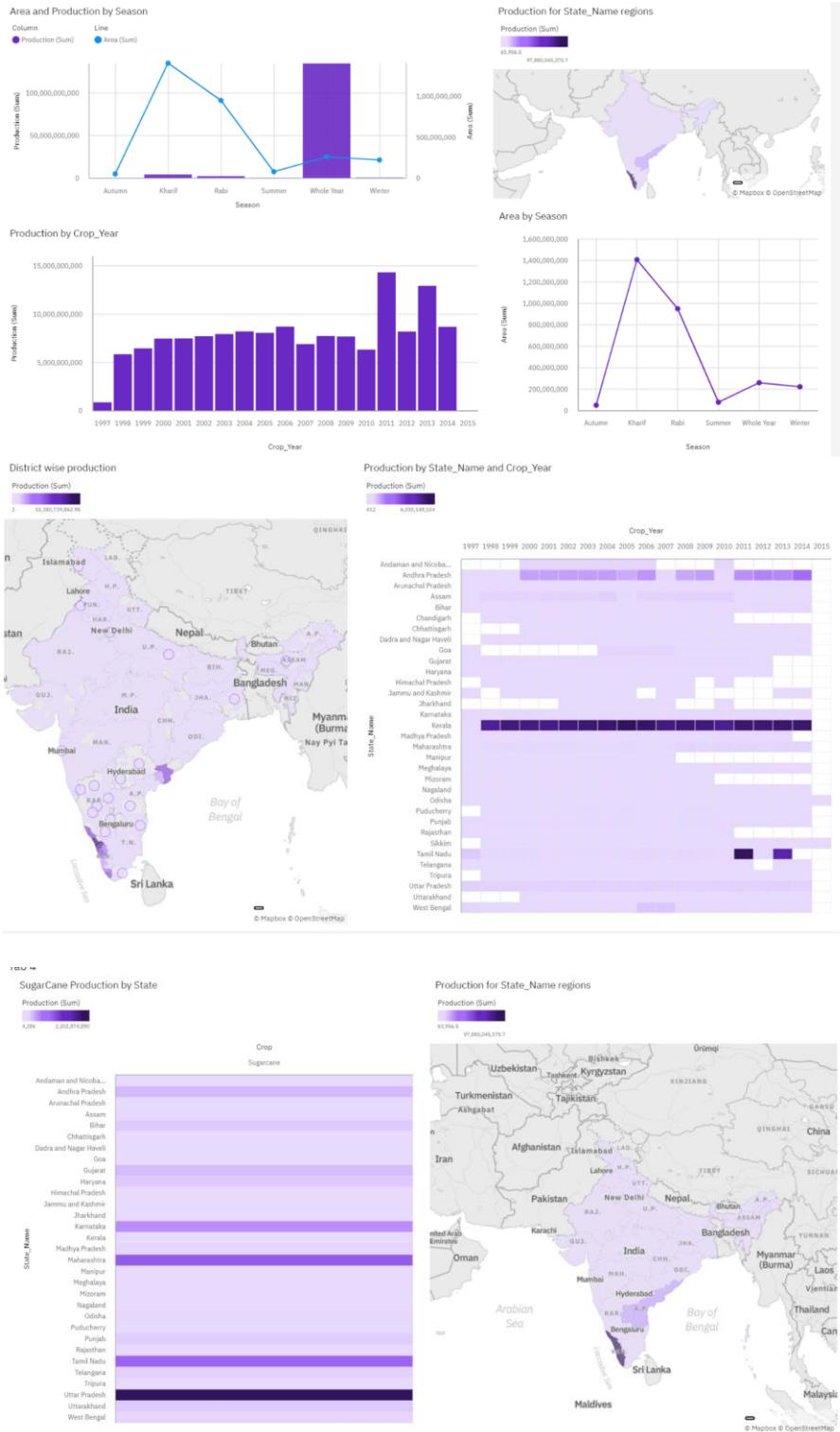
CHURACHANDPUR
CUDDALORE
DAMOH
DARJEELING
DAVANGERE
DEORIA
DHANBAD
DHEMAJI
DHULE
DIMAPUR
DINDORI

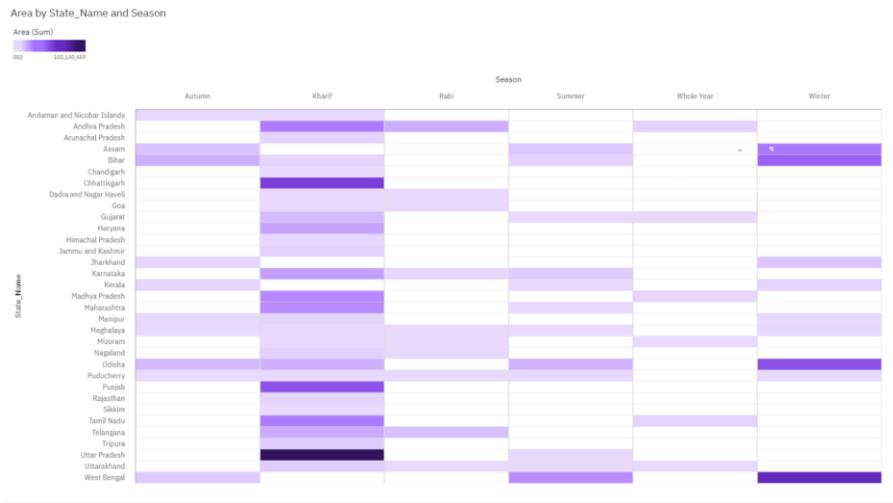


DASHBOARD:

Tab 1







STATES WITH CROP PRODUCTION ALONG WITH THE SEASON:

The screenshot shows the IBM Cognos Analytics interface with two reports open:

- Top Report (Crop Blackgram):** This report shows the top 10 states with the most area for the crop 'Blackgram'. The results are as follows:

State_Name	Area (Sum)
Assam	102,140,469
Bihar	102,140,469
Kerala	102,140,469
Nagaland	102,140,469
Others	102,140,469

 A filter dialog for 'Crop' is open, showing 'Blackgram' selected.
- Bottom Report (Season):** This report shows the state with crop production for different seasons. The results are as follows:

Season	State
Kharif	Assam
Rabi	Bihar
Whole Year	Kerala

 A filter dialog for 'Season' is open, showing 'Kharif', 'Rabi', and 'Whole Year' selected.

10. ADVANTAGES & DISADVANTAGES

Advantages

- People can easily identifies the crop choice for cultivation depends on the location, season and so on
- Crop production increases due to right choice of crop during cultivation
- It gives profit for the people who chose the seasonal crop in a particular location

Disadvantages

- Accuracy may be differ due to some unavoidable conditions
- Not at all the features that are impact the production be taken

11. CONCLUSION

With the help of technology, people can cultivate and harvest seasonal crops and get better yield with more accuracy. Hence the crop yield production can be increased and better understand of the nature of the soil for crop cultivation. The following steps can be done for the estimation of crop yield production. Previous year data can be collected and get exploratory data analysis. After removing the null values and outliers, data visualisation is analysed and get insights for the future work

12. FUTURE SCOPE

In the future work, can implement with IOT devices as edge layer like temperature sensor, humidity sensor and so on. It is used to collect the real-time data and understand the real condition of the soil and moisture level. So, people can easily find the better crop choice depends on conditions and make better accuracy

13. APPENDIX

- Sourcecode:<https://github.com/IBM-EPBL/IBM-Project-33119-1660214826/tree/main/Final%20Deliverables>
- Github link -<https://github.com/IBM-EPBL/IBM-Project-33119-1660214826>
- Project demo link - <https://github.com/IBM-EPBL/IBM-Project-33119-1660214826/tree/main/Final%20Deliverables/Project%20Demo>