India Agriculture

Analysis

By: Evan Putra Gratia

BACKGROUND

AGRICULTURE IN INDIA

58% of Indians are engaged in agriculture practice.

18% of GDP is contributed by agriculture in India.

60% to overall India's agriculture GDP only by animal farming and horticulture.

11% of Indian agriculture production increased in the past 14 years.

OBJECTIVE is simple!

To maximize the potential of agriculture in India

DATASET

India Agriculture Crop Production | Kaggle

Original Dataset
State
District
Crop
Year
Season
Area
Area Units
Production
Production Units
Yield

→ → → →

Cleaned Dataset
State
District
Crop
Year
Season
Area
Area Units
Production
Production Units
Production in Tonnes
Yield

data about farm yield from different parts of India from 1998 - 2020

Cleaned dataset:
340.095 rows
11 columns

Data dictionary explanation

INDIA AGRICULTURE LANDSCAPE 1998 - 2020

Average production in Tonnes is **44.965** and yield **4.61x** from total area

Most of the crops are in **Uttar Pradesh** State and in district **Bilaspur**

Most crop yielded on **Monsoon** season

Average crop area is **11.833** Hectares

Most of crops planted with **Rice**

1024 crops has experienced crop failure

LETS DIVE DEEPER!

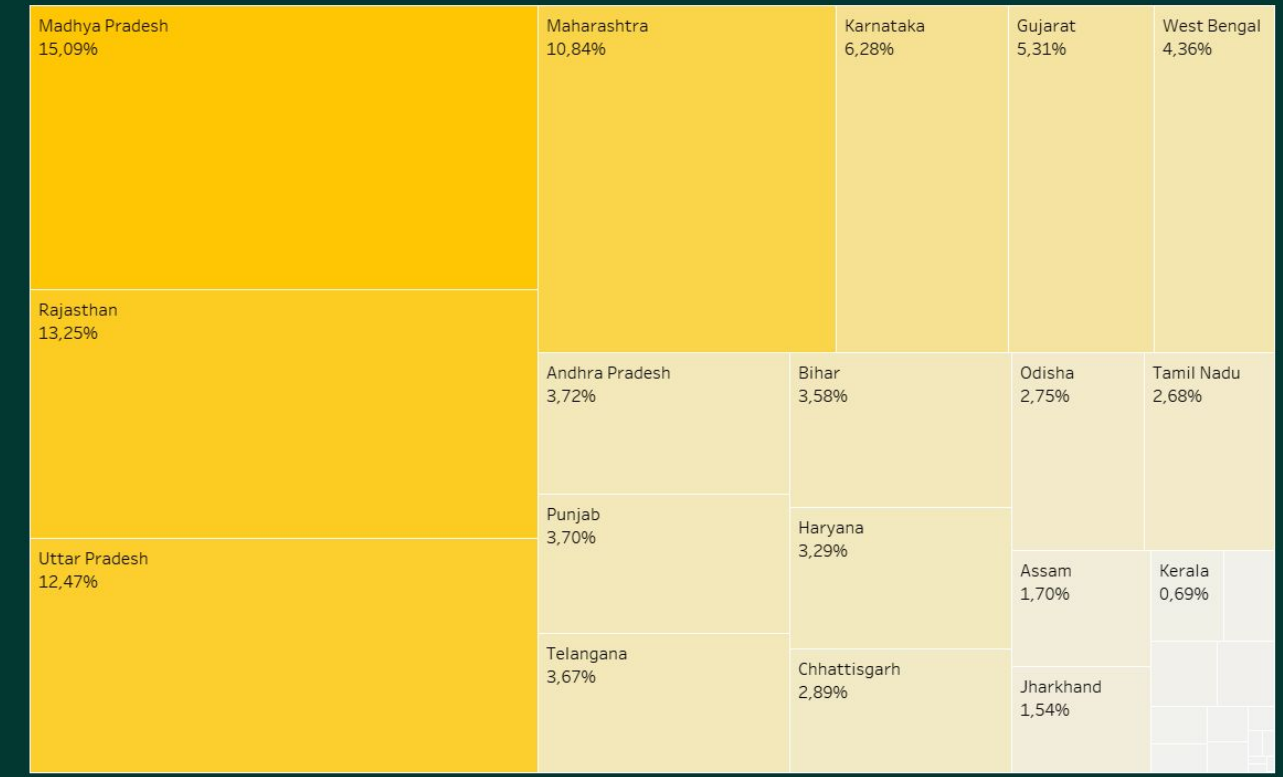
We are going to do Exploratory Data Analysis from now on.

Total Area Trendline

In the last 10 years we can see **positive trendline**, means the area for agriculture in India is always growing. In 2020 the total Area for crop is **194.956.661 ha**

Almost **60%** of the total area in India used as an agriculture crop

Area Distribution in 2020



Madhya Pradesh, Rajasthan, Uttar Pradesh, and Maharashtra dominate 51% of the total crop area (around 100mil Hectares)

What do they produce most in 2020? (Tonnes)

Madhya Pradesh

Wheat	37.507.219
Sugarcane	7.414.530
Rice	7.363.430

Rajasthan

Wheat	13.894.293
Bajra	5.086.651
Rapeseed &Mustard	4.288.797

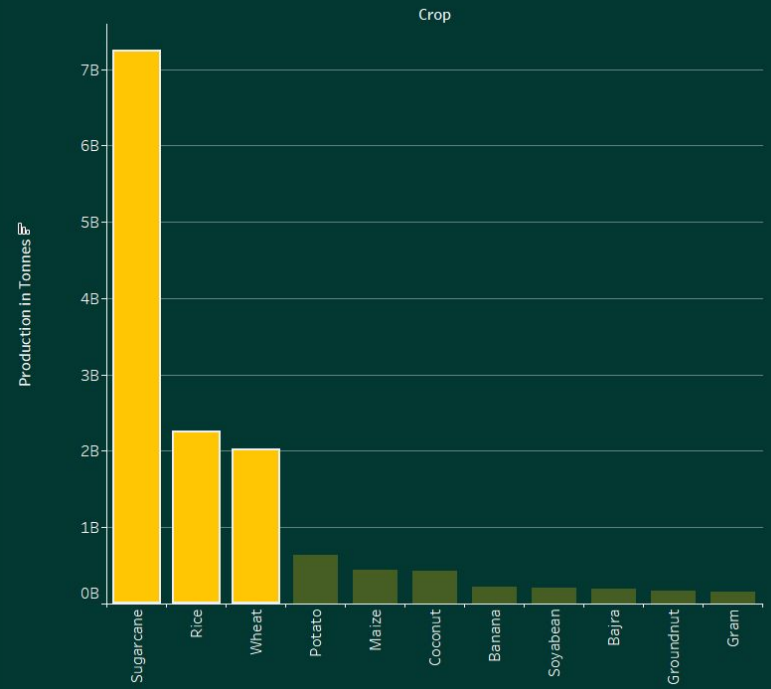
Uttar Pradesh

Sugarcane	179.567.765
Wheat	36.209.665
Rice	17.027.889

Maharashtra

Sugarcane	69.312.919
Soybean	4.825.628
Rice	2.897.433

Which crops have the most production?

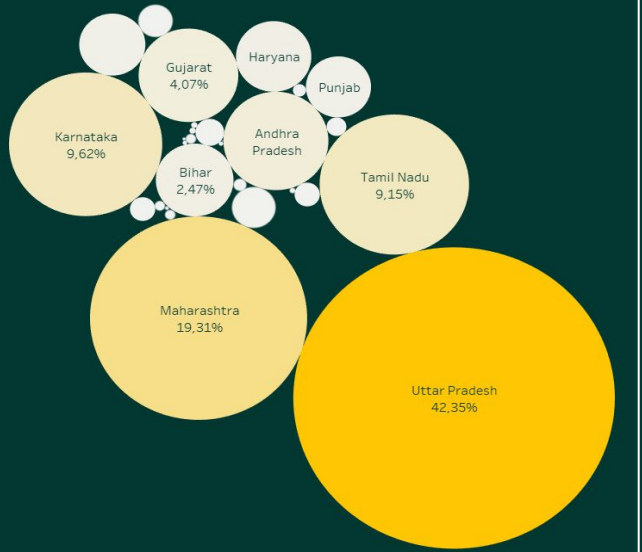


These are the top 10 crop production from total of **56 crops**

Most of the crop production is from **Sugarcane** with over than **7b tonnes** total across 1998-2020, followed by **Rice and Wheat** with **2b tonnes each**

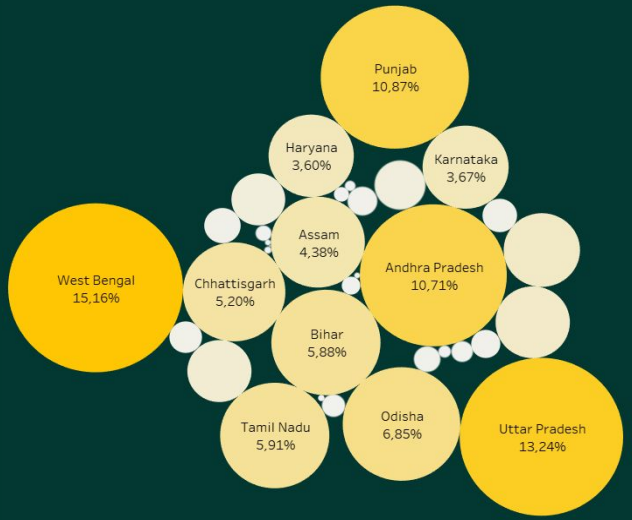
So, which State produce the most?

Sugar Cane



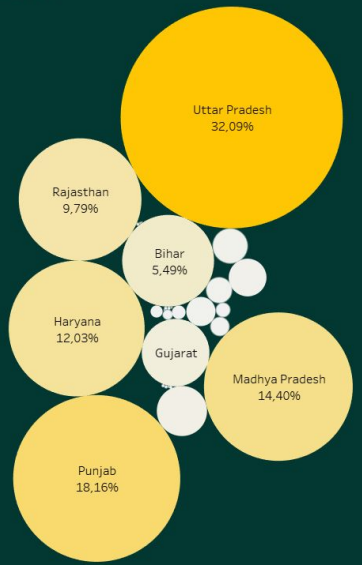
Uttar Pradesh become the biggest state that produce Sugar Cane (42%)

Rice



West Bengal, Uttar Pradesh, Andhra Pradesh, and Punjab are the top 4 State that produce Rice

Wheat



Uttar Pradesh also the excel as a Wheat producer (32%)

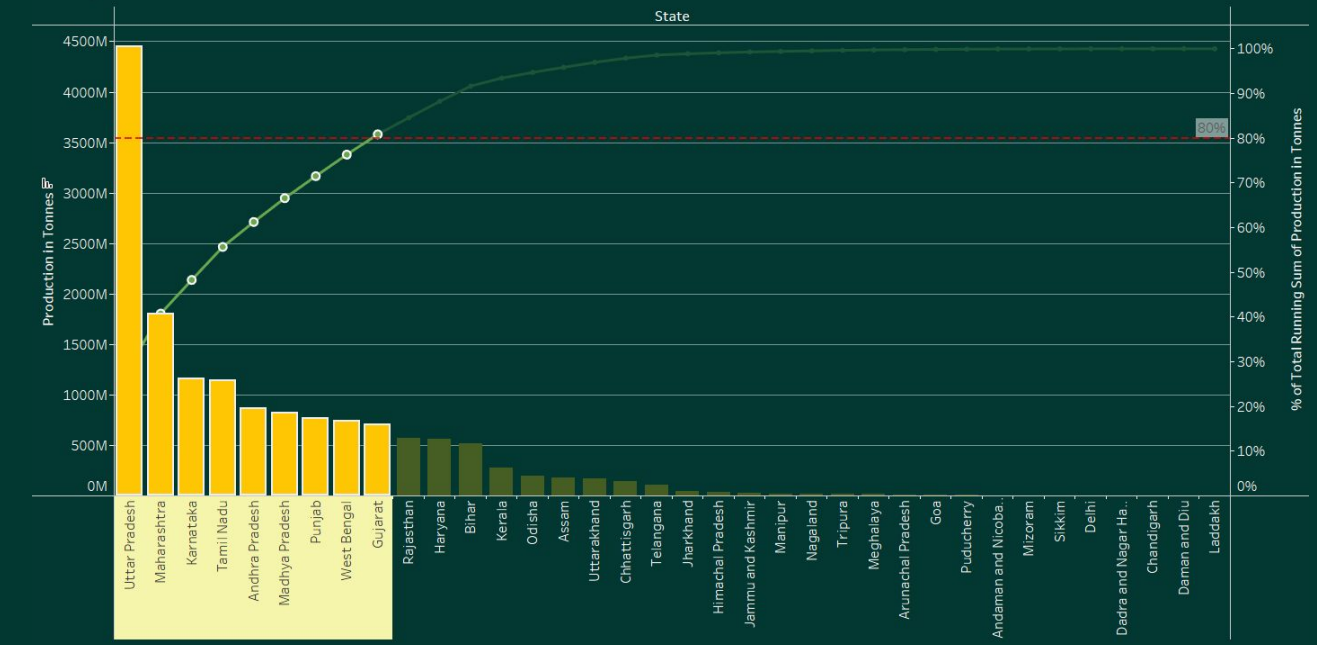
Total Production Trendline

As the Area of the crop is increasing, the **production also increases**

Average increases of the production from 1998 - 2020 is **3.3%**

Last 4 years run, production increase is averaging **5.05%**

Total



From total of 36 states , **9 states** contribute **80%** for the total production agriculture in India. These states are:

1. Uttar Pradesh
2. Maharashtra
3. Karnataka
4. Tamil Nudu
5. Andhra Pradesh
6. Madhya Pradesh
7. Punja
8. West Bengal
9. Gujarat

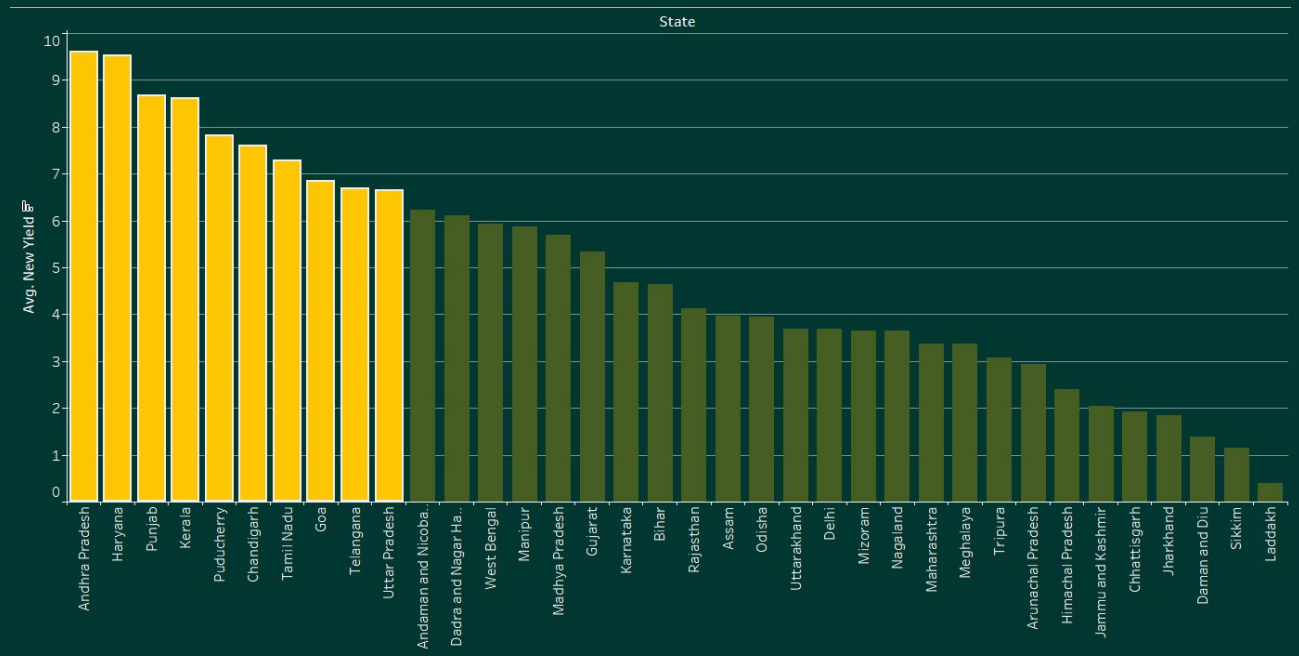
Total

Based on this graph, the most productive season is **Monsoon**

The reason **Summer** and **Autumn** can not produce much because of the irrigation issues

In short, India really rely on **yearly rain** for their agriculture industry

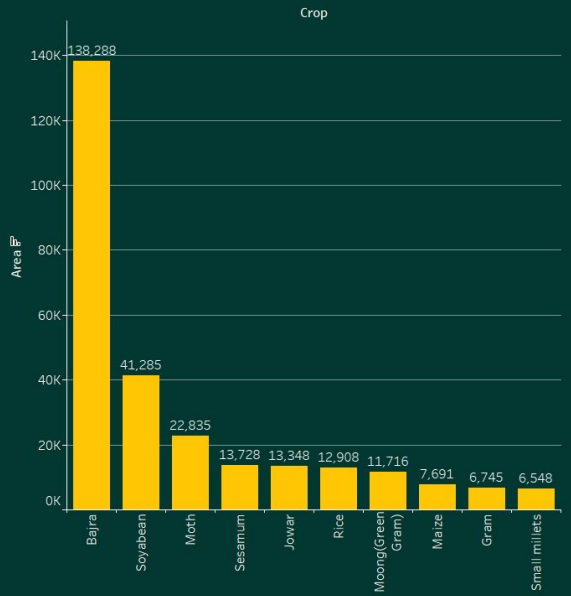
Which State Yield The Most In The Last 5 Years ?

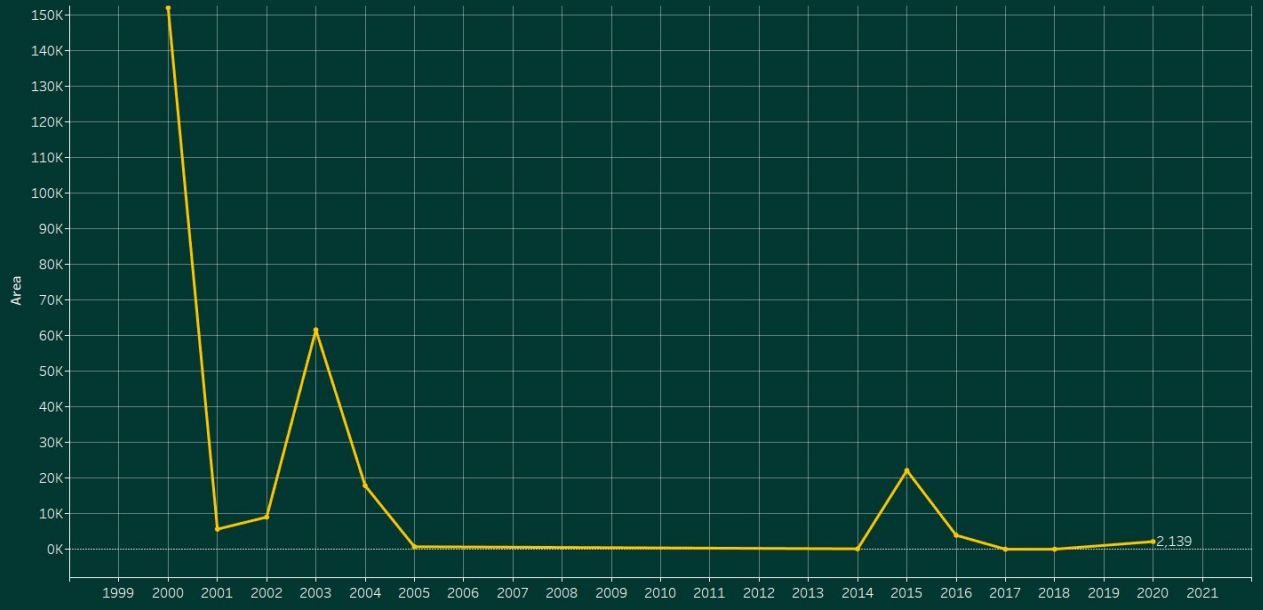


Andhra Pradesh has the biggest yield ratio between area and production with **9.6x**

Onion, Sugarcane, and Banana Yield the most from Andhra Pradesh

CROP FAILURES





There are **1024** crop failures happened across year 1998 - 2021. Total crop area that occurs crop failure is **314.246** Hectare.

The top 3 crops that failed to produce anything are **Bajra, Soybean, and Moth**

Worry not about the crop failures. In 2020 itself, only **0.0001%** crop failures occurred from the total Area (195 mil Ha)

CLUSTERING using K-means

We are trying figure out if there's any cluster among the dataset based on Area, Production in Tonnes, and Yield

METHODOLOGY

01

FEATURE STANDARDIZATION

Using StandardScaler

02

ELBOW METHOD

To know how many clusters

03

SILHOUETTE METHOD

To confirm the amount of the clusters

04

K-MEANS CLUSTERING

Assigning clusters to the dataset

K-MEANS RESULTS

Cluster	Average Area in Ha	Average Production in Tonnes	Average Yield
Normal Crop	11.597	33.281	4.45x
Outstanding Crop	199.417	8.149.305	120.94x
Failed	306	0	0x

Cluster	Count	Percentage
Normal Crop	338.582	99.56%
Outstanding Crop	489	0.14%
Failed	1024	0.3%

CLUSTER 1 INSIGHT

472 /489 from the cluster 1 is a **Sugarcane** crop

Average Sugarcane yield from cluster 1

79.83

Average Sugarcane yield from cluster 0

55.95

= **23.88**

x **x** **x**

344 outstanding Sugar Cane crops are from **Uttar Pradesh & Maharashtra**

These Sugar Cane crops sowed on **Whole Year Season & Monsoon Season** **Year 2010**

ML MODELING

We want to predict the Production in Tonnes
from State, Crop, Season , and Area

PRE-PROCESSING STEPS

1 Prepare Dataset

Features:

1. State
2. Crop
3. Season
4. Area

Target:

Production in Tonnes

2 Encode Labeling

Encode categorical features into numeric

340.095 rows & 96 columns

3 Split Dataset

80% Training Data

20% Test Data

BUSINESS SIMULATION

LINEAR REGRESSION

1

3

DECISION TREE REGRESSOR

RIDGE REGRESSION

2

4

RANDOM FOREST REGRESSOR

REGRESSION MODEL RESULTS

NO	MODEL	R2 SCORE		RMSE (Standardized)		MAPE	
		TRAIN	TEST	TRAIN	TEST	TRAIN	TEST
1	Linear Regression	0.16	0.17	0.93	0.82	307%	257%
2	Ridge Regression	0.16	0.17	0.93	0.82	307%	257%
3	Decision Tree Regressor*	0.97	0.96	0.17	0.18	52%	59%
4	Random Forest Regressor	0.99	0.95	0.07	0.18	18%	44%

*Hypertuning Parameters

CONCLUSION

- Total area and production have **positive trendline**
- **Madhya Pradesh, Rajasthan, Uttar Pradesh, and Maharashtra** posses 51% from the total crop area in 2020
- **Sugar cane, Rice, and Wheat** are the main commodities based on total production production
- **Summer and Autumn** season are not a good season for India agriculture
- Only **9 States** out of 36 contribute 80% to total production
- Crop failure is **not a main problem** anymore

HOW TO MAXIMIZE THE POTENTIAL OF AGRICULTURE IN INDIA?

- The Gov can allocate some funds to create better **irrigation system** especially in rural areas
- **Give incentives** to State that has an outstanding results
  - Ex: crowned as ‘The Agricultural State”
- Use Machine Learning Model to predict **expected outcomes** of total production (as a baseline)
- Boost average yield from “The Big 4” from **cross-study** between other States

BUSINESS SIMULATION in Madhya Pradesh (2020)

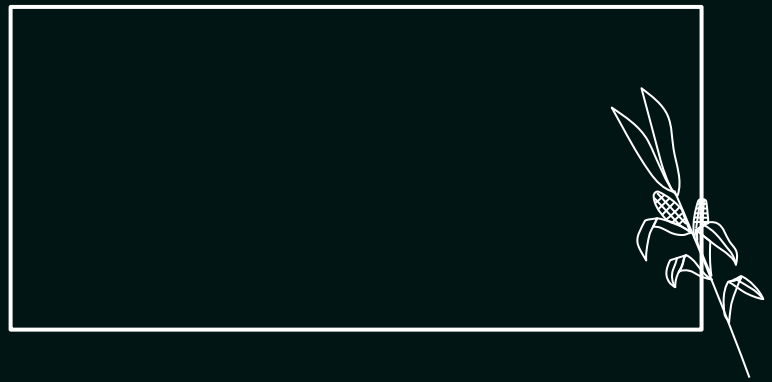
	Wheat	Sugarcane	Rice
Total Area	10.216.517 Ha	124.681 Ha	3.106.260 Ha
Average Yield	3.52x	39.05x	1.78x
Total Production	37.507.219 Tonnes	7.414.530 Tonnes	7.363.430 Tonnes

Learn From	Punjab	Uttar Pradesh	Chandigarh
Avg Yield	4.89x	70.96x	5.50x
Total Production	49.958.768 Tonnes	8.847.363 Tonnes	17.084.430 Tonnes

Total Production Difference	**12.451.549 Tonnes**	**1.432.833 Tonnes**	**9.721.00 Tonnes**

THANKS!

Do you have any questions?
evanputragratia@gmail.com
Code Sript



CREDITS: This presentation template was created by **Slidesgo**, including icons by **Flaticon**, infographics & images by **Freepik**