

BRIDGING THE INDIAN AGRICULTURAL PROTECTION GAP

By Resha Dirga, Jeanne Dubois and Anniek Brink
MSc Big Data and Business Analytics ESCP Europe

TABLE OF CONTENT

- Problem Statement
- Tools Used and Analyses Performed
- Clusters Presentation
- Main Cluster Drivers
- Defining the Risks of the Clusters
- Risk Matrix State Level
- Risk Matrix District Level
- Summary Findings
- Appendices



PROBLEM STATEMENT

- Agricultural production in India is very dependent on the weather conditions: e.g. **55 %** of India's arable land depends on precipitation and with **1°C** rise in temperature, wheat yields are predicted to fall by **5-10 %**.
- In order to profile the agricultural risks to better advise Scor's clients and to estimate the right premiums, **clustering** will be performed to segment the different Indian agricultural zones based on similar factors and behaviors for losses.



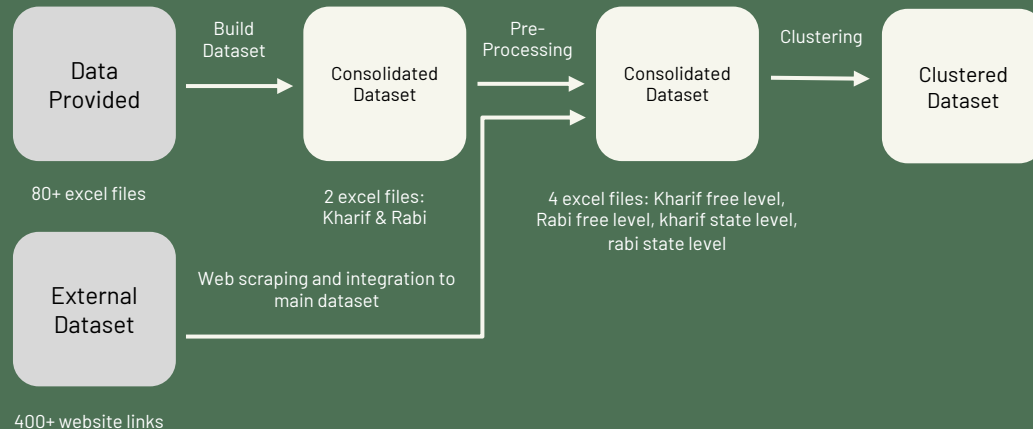
TOOLS USED AND ANALYSES PERFORMED

CLUSTERING

Tools used: Numpy, Pandas, Excel, Python, Jupyter Notebook

Tools used: Selenium, BeautifulSoup, Geckodriver

Tools used: Sklearn, Faiss



ANALYSIS

Tools used: SHAP, Tableau, Excel, Python, Jupyter Notebook

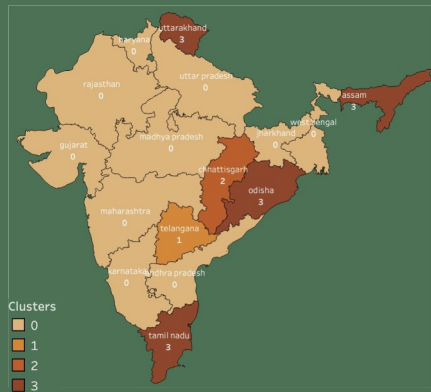
Main Cluster Drivers

Factors Affecting Risks

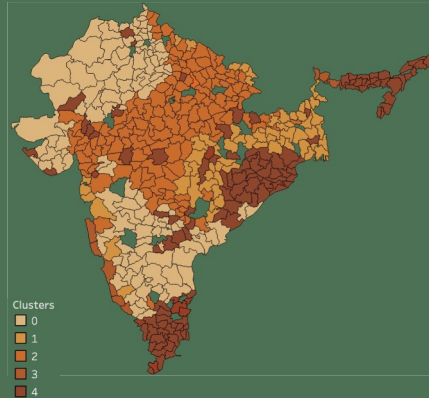
Risk Matrix

CLUSTERS PRESENTATION

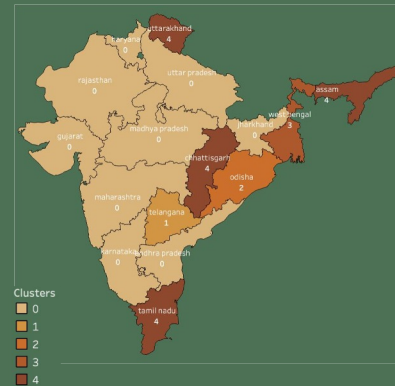
RABI STATE LEVEL



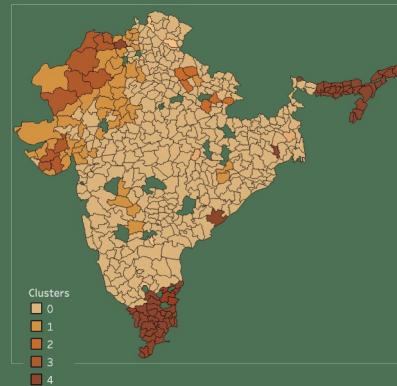
RABI DISTRICT LEVEL



KHARIF STATE LEVEL



KHARIF DISTRICT LEVEL



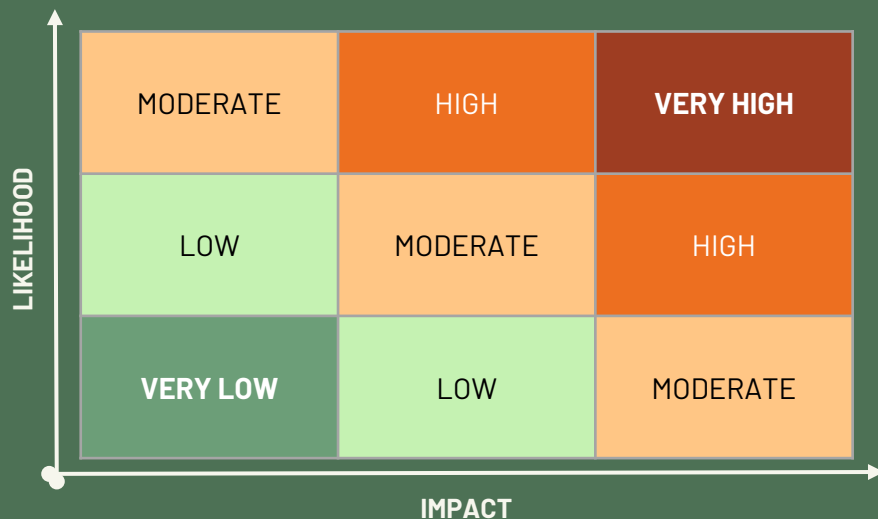
MAIN CLUSTER DRIVERS

- To show how much each predictor contributed to the clusters, we used the Python module **SHAP** (*Beeswarm analysis*).

	RABI STATE LEVEL	RABI DISTRICT LEVEL	KHARIF STATE LEVEL	KHARIF DISTRICT LEVEL
CLUSTER 0	HIGH PRECIPITATION FEBRUARY	HIGH AVERAGE TEMPERATURE NOVEMBER	HIGH PRECIPITATION MAY	HIGH LOSS CALCULATION
CLUSTER 1	OUTLIER	LOW RAINY DAYS DECEMBER	OUTLIER	LARGE AREA SOWN BY CROP
CLUSTER 2	OUTLIER	HIGH PRECIPITATION MARCH	OUTLIER	LOW LOSS CALCULATION
CLUSTER 3	LOW PRECIPITATION FEBRUARY	HIGH RAINY DAYS NOVEMBER	OUTLIER	LARGE AREA SOWN BY CROP
CLUSTER 4			LOW MINIMUM TEMPERATURE JULY	

DEFINING THE RISKS OF THE CLUSTERS

RISK MATRIX FRAMEWORK



LIKELIHOOD

Defined as the **probability of the risk to happen**. In assessing the likelihood, we (1) built a probability-based model that predicts whether a region will have a loss (loss calculation > 0) and (2) assigned the prediction probability as the likelihood of the risk to happen.



IMPACT

Defined as **the severity when the risk occurs**. In assessing the impact, we used the definition of **the amount of loss for every yield that goes below the threshold**, which can be denoted as:
$$Impact = Loss / \max(0, threshold\ yield - actual\ yield).$$

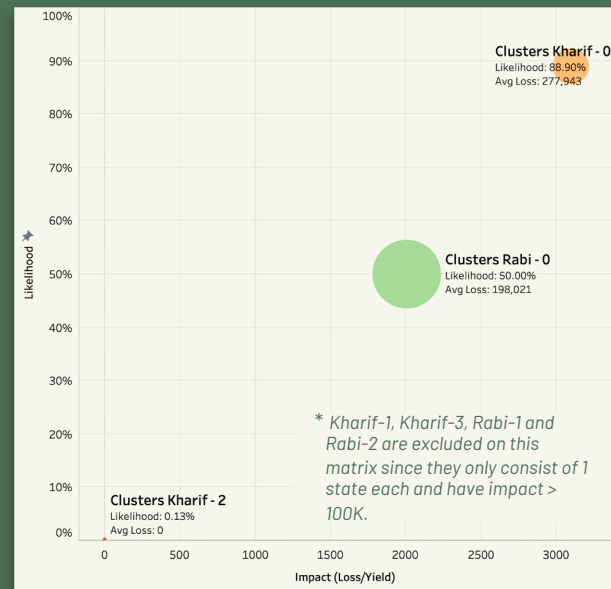
RISK MATRIX STATE LEVEL

- Generally, the risk of loss during Rabi season is **lower than** the risk of loss during Kharif season.
- There are two states during Kharif season and one state during Rabi season that have **significantly higher risk** compared to the other states.
- There is always one state that does not produce any crops during each season.

	# OF STATES	LIKELIHOOD	LOSS/YIELD	AVG. LOSS
Kharif - 0	9	88.90%	3,098.80	277,942.55
Kharif - 1	1	100%	215,333.75	210,257.57
Kharif - 2	1	0.13%	-	-
Kharif - 3	1	100%	6,631,939.87	215,299.73
Rabi - 0	10	50.00%	2,009.07	198,020.56
Rabi - 1	1	100%	105,079.13	185,686.87
Rabi - 2	1	0%	-	-



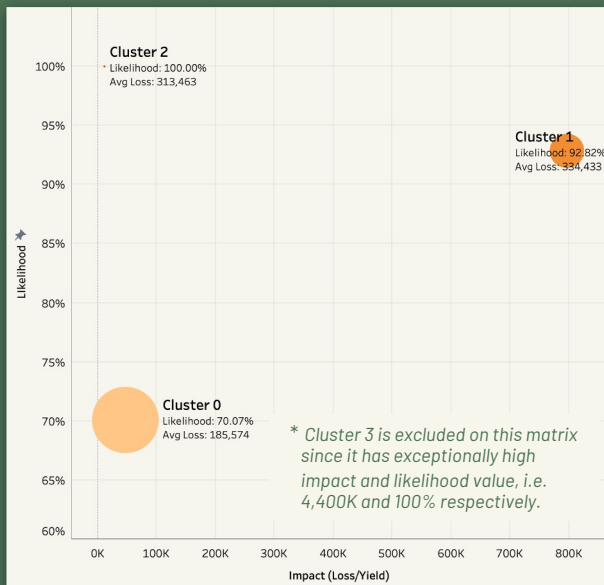
STATE RISK MATRIX *



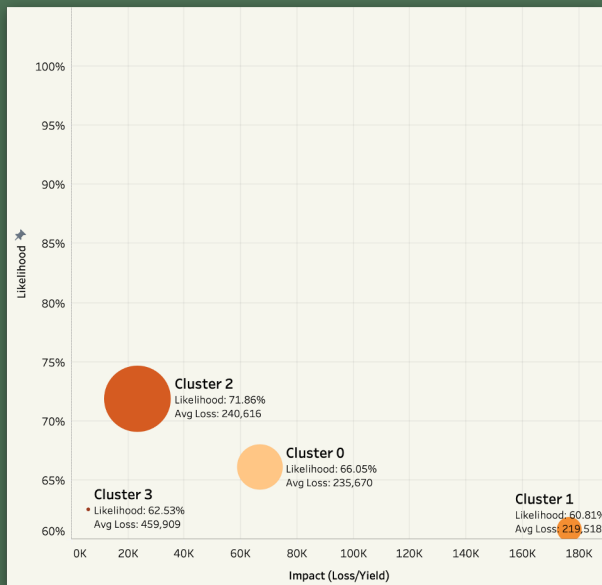
Legend:  Number of regions in cluster

RISK MATRIX ON DISTRICT LEVEL

KHARIF RISK MATRIX *



RABI RISK MATRIX



Legend: Number of regions in cluster

- There is a **clear distinction of risk** for Kharif clusters: Cluster 0 has significantly lower risk compared to the other cluster.
- Risks between Rabi clusters are **relatively harder to distinguish**, since clusters with low impact have high probability, and vice versa.

SUMMARY FINDINGS

- Indian agriculture loss behavior **heavily relies on weather**, thus change in weather behaviour and variability leads to an increase in agricultural risk.
- Looking into **unusual weather behaviour** that will cause agricultural loss can help to precisely segment different agricultural zones and place them on the risk matrix and thus better advise clients.
- Rabi district clustering ('winter' season): Cluster with highest average loss (3) = lowest average temperature = highest level of precipitations. [DASHBOARD](#)
- Kharif district clustering ('Monsoon' season): Cluster with highest average loss (3) = lowest precipitation level = lower number of rainy days. [DASHBOARD](#)



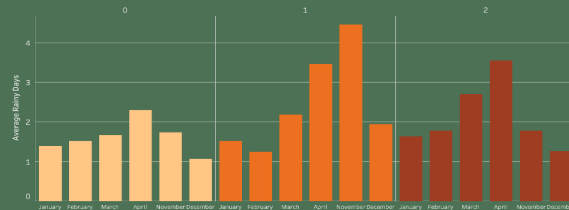
APPENDIX

APPENDIX 1:

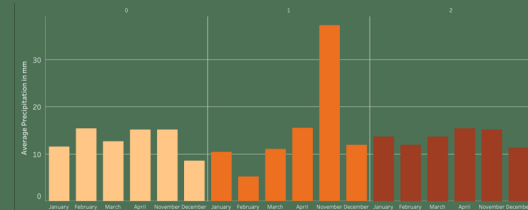
Cluster Characteristics

RABI STATE LEVEL

Average Rainy Days per Month

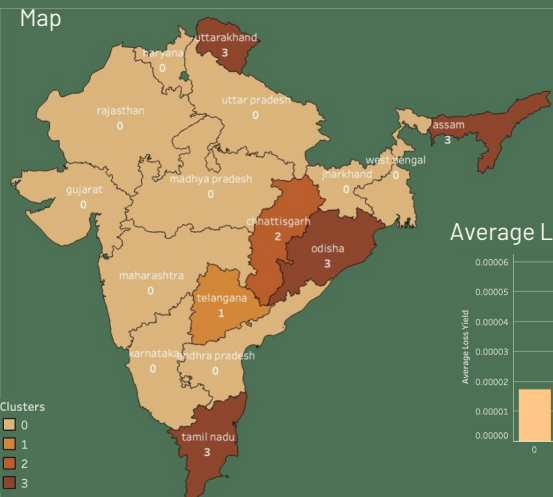


Average Precipitation in mm per Month

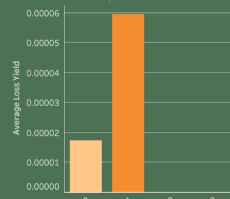


Rabi
0
1
2

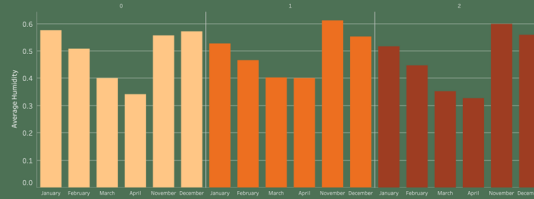
Map



Average Loss Yield



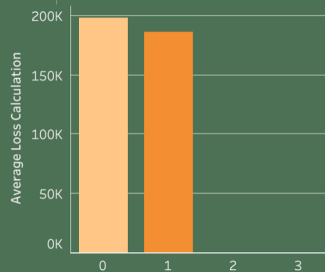
Average Humidity



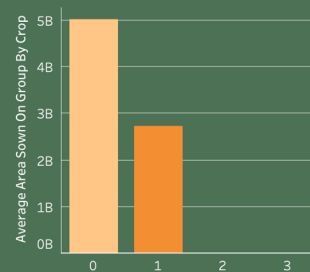
Average Max. Temperature in Celsius



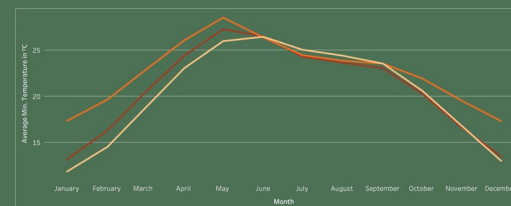
Average Loss Calculation



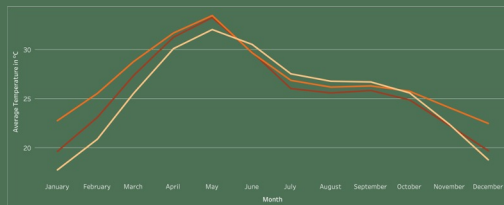
Average Area Sown on Group by Crop



Average Min. Temperature in Celsius

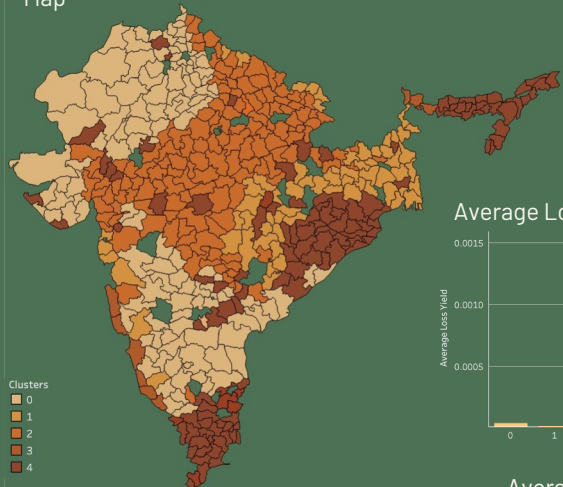


Average Temperature in Celsius

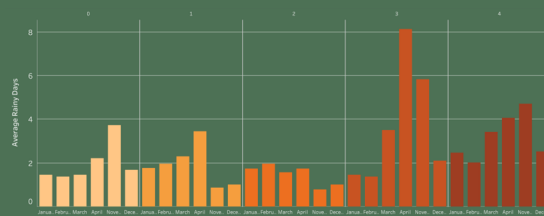


RABI DISTRICT LEVEL

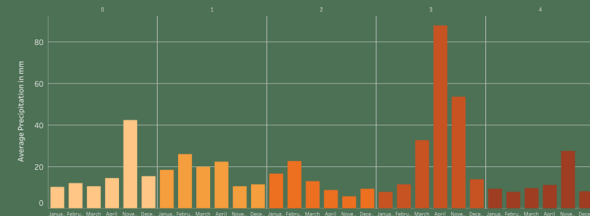
Map



Average Rainy Days per Month



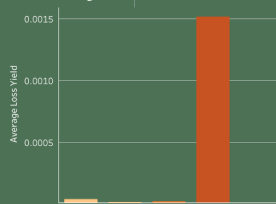
Average Precipitation in mm per Month



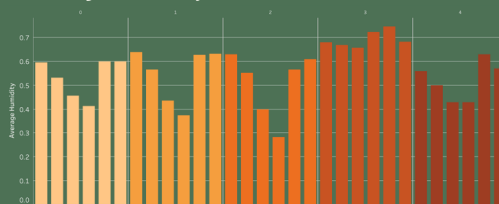
Kharif



Average Loss Yield



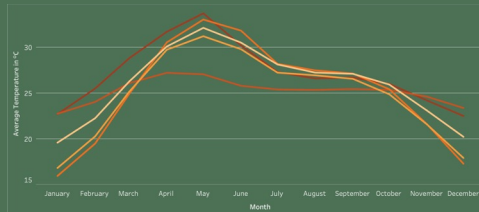
Average Humidity



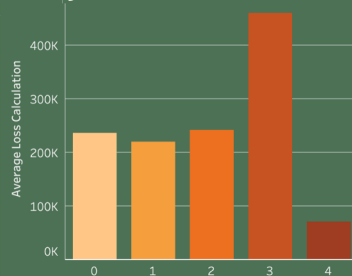
Average Max. Temperature in Celsius



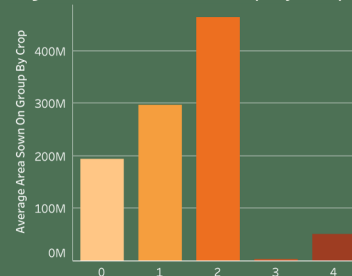
Average Temperature in Celsius



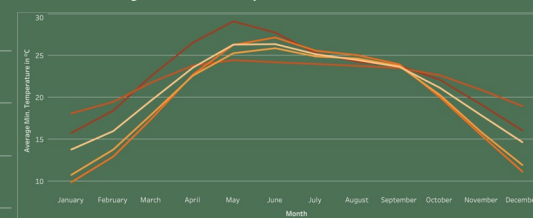
Average Loss Calculation



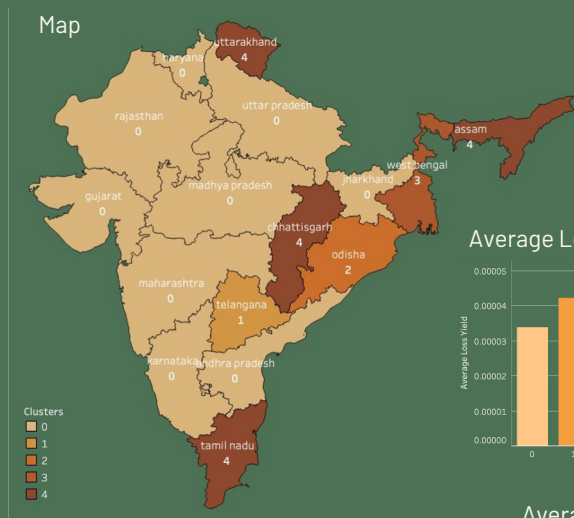
Average Area Sown on Group by Crop



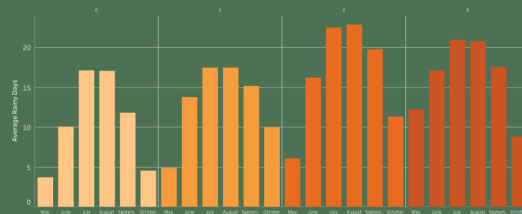
Average Min. Temperature in Celsius



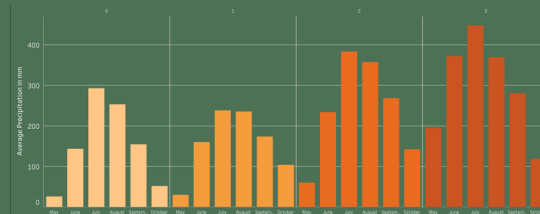
KHARIF STATE LEVEL



Average Rainy Days per Month



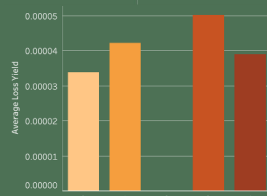
Average Precipitation in mm per Month



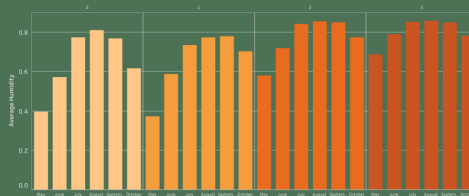
Kharif



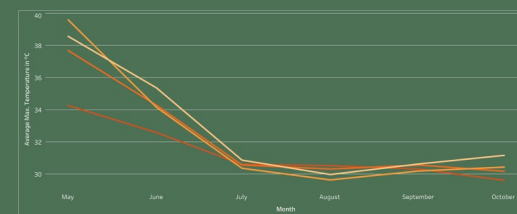
Average Loss Yield



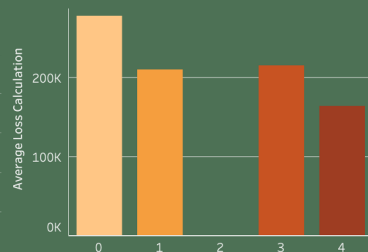
Average Humidity



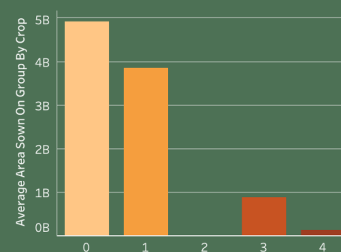
Average Max. Temperature in Celsius



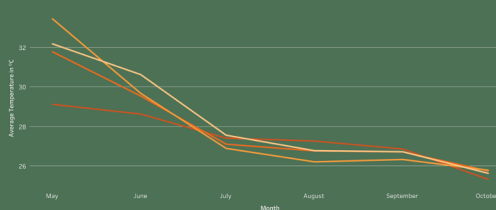
Average Loss Calculation



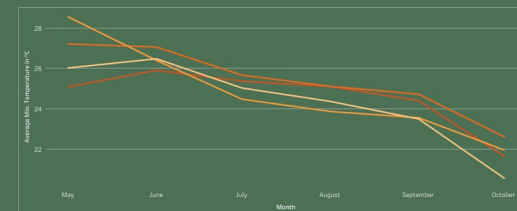
Average Area Sown on Group by Crop



Average Temperature in Celsius

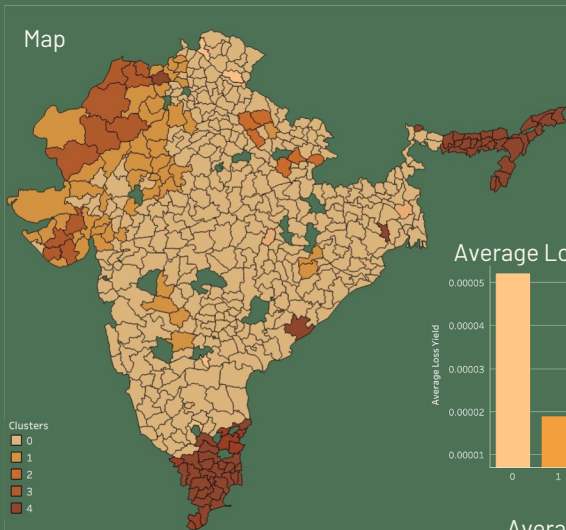


Average Min. Temperature in Celsius

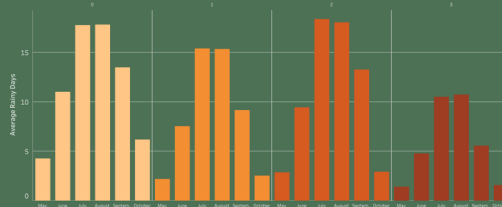


KHARIF DISTRICT LEVEL

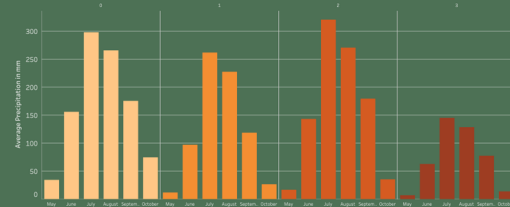
Map



Average Rainy Days per Month



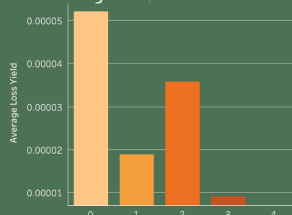
Average Precipitation in mm per Month



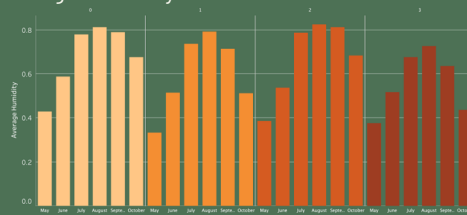
Kharif



Average Loss Yield



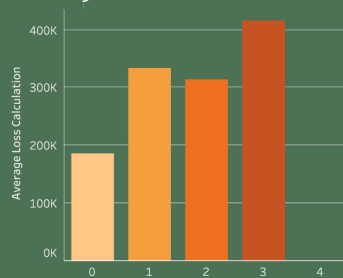
Average Humidity



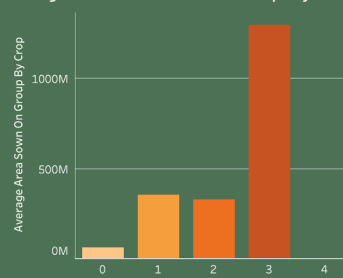
Average Max. Temperature in Celsius



Average Loss Calculation



Average Area Sown on Group by Crop



Average Temperature in Celsius



Average Min. Temperature in Celsius

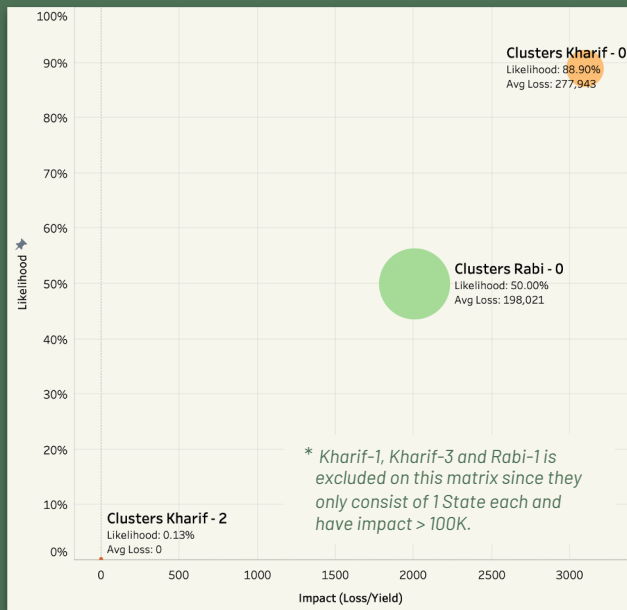


APPENDIX 2:

Risk Matrix Values

RISK MATRIX STATE LEVEL

STATE RISK MATRIX *

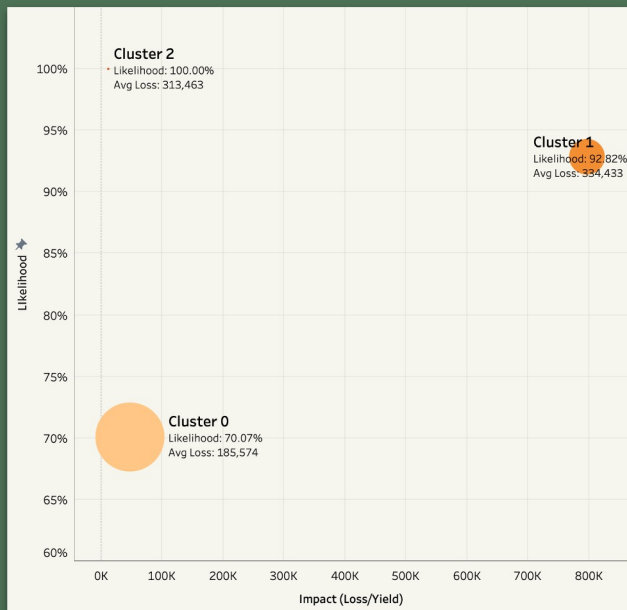


	# OF STATES	LIKELIHOOD	LOSS/YIELD	AVG. LOSS
Kharif - 0	9	88.90%	3,098.80	277,942.55
Kharif - 1	1	100%	215,333.75	210,257.57
Kharif - 2	1	0.13%	-	-
Kharif - 3	1	100%	6,631,939.87	215,299.73
Rabi - 0	10	50.00%	2,009.07	198,020.56
Rabi - 1	1	100%	105,079.13	185,686.87
Rabi - 2	1	0%	-	-

Legend:  Number of regions in cluster

RISK MATRIX OF KHARIF ON DISTRICT LEVEL

KHARIF RISK MATRIX *



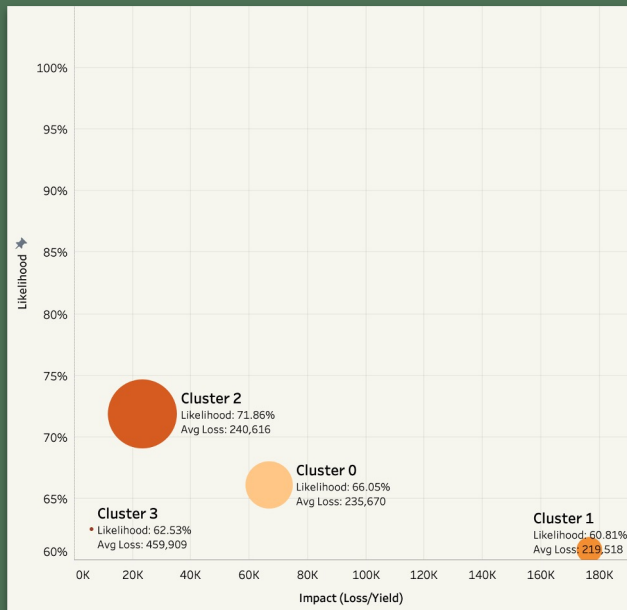
Legend: Number of regions in cluster

	# OF STATES	LIKELIHOOD	LOSS/YIELD	AVG. LOSS
Cluster 0	401,671	70.07%	47,233.11	185,574.11
Cluster 1	16,510	92.82%	797,165.61	334,432.98
Cluster 2	15,545	100%	12,136.99	313,462.89
Cluster 3	4,680	99.99%	4,428,228.54	416,405.86

* Cluster 3 is excluded on this matrix since it has exceptionally high impact and likelihood value.

RISK MATRIX OF RABI ON DISTRICT LEVEL

RABI RISK MATRIX



Legend: Number of regions in cluster

	# OF STATES	LIKELIHOOD	LOSS/YIELD	AVG. LOSS
Cluster 0	49,107	66.04%	66,741.45	235,669.83
Cluster 1	19,924	60.81%	176,780.23	219,518.04
Cluster 2	127,870	71.86%	23,324.24	240,616.15
Cluster 3	1,105	62.53%	5,897.47	459,909.13