Crop Production Data Visualisation

GROUP-8
Arya Pinaki(200194)
Bhavya Garg(200270)
Gauri G Menon(200382)
Kaushikraj V Nadar(200499)
S Pradeep(200826)

CS661A Project Report

1 Acknowledgement

We would like to extend our sincere gratitude to our course instructor, Prof. Soumya Dutta, for guiding us in this endeavour and equipping us with relevant skills. We also thank the institute IIT Kanpur for giving us the opportunity to work on this enriching and challenging project.

2 Introduction

India is an agrarian country, where agriculture has been the backbone of its economy for centuries. Owing to the massive population and declining agricultural resources, we face challenges regarding the adequate and nutritional supply of food, with an added emphasis on sustainability and self-sufficiency. The agriculture sector in India has faced many challenges in recent years, such as climate change, water scarcity, and decreasing soil fertility, which have affected crop yields and farmers' incomes.

In this context, our project focuses on highlighting the cropping patterns and rainfall trends of the last 20 years. It aims to analyse the historical crop production data, rainfall patterns, and crop choices made by farmers to identify the relevant correlation between them. This analysis can help in predicting the impact of rainfall on crop production, identifying the most suitable crops for a particular region based on the same and lay emphasis on unsustainable aspects of agricultural practices across the country. Eventually, this can assist policymakers in formulating crop-specific policies for suitable regions.

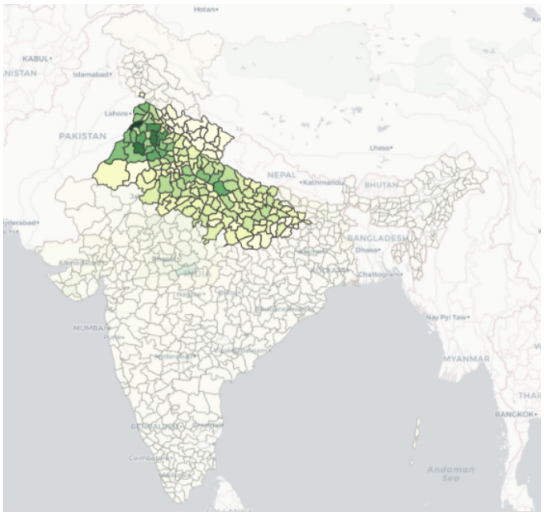
3 Dataset description

Crop Production Statistics - India: This is our primary dataset and it contains comprehensive data on crop production statistics for India, categorized by state and district. The dataset covers four major crop seasons, namely Kharif, Rabi, Summer, and Autumn, from the year 1997 to 2017. The data provides information on the annual production and yield(per hectare) of crops grown in different parts of the country for each crop.

Rainfall Statistics - India: Additionally, we plan on using a secondary rainfall data in India over the same time period, to allow for a correlational visualization of rainfall and crop production over time. We can also use this to estimate the stress on other methods of irrigation.

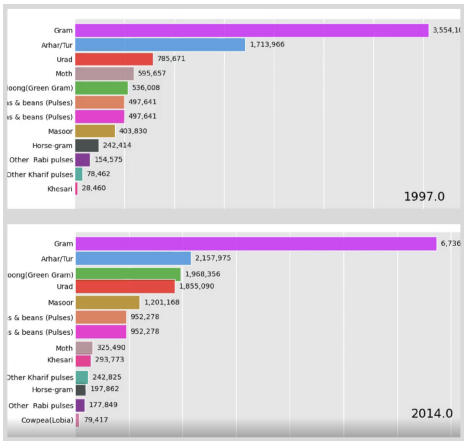
4 Task description

District-wise production analysis: To visualize the crop production statistics across the country, district wise, to provide a relatable graphical representation of the data. This helps in providing geographical insights.



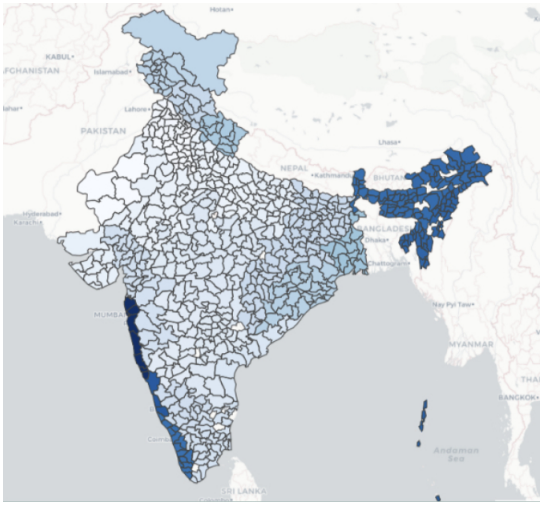
Wheat Production in India in the Indo-Gangetic Plains

Year-wise production analysis: To visualize the rate of growth(or decline) of total production of each crop(belonging to a particular crop group) to highlight the differences in cropping pattern and overall crop diversity over time.



Pulses comparison

Rainfall correlation: To understand the influence of climate and whether patterns over crop production(particularly relevant due to recent climate change trends). In addition, it can help estimate the stress on other water resources(ie. Rivers, groundwater etc) especially when resource intensive crops are grown on unsuitable soils.



Rainfall in India

5 Data Processing

1. Two of our tasks involved mapping tabular data to geographical shapes on the map. GeoJSON files encode geographical data structures in JavaScript Object Notation (JSON). It represents points, lines, polygons, and multi-part groupings of these kinds. We build the current GeoJSON with all districts from scratch.

2. The Survey of India (SOI)'s National Topographical Database (NTDB), under the Department of Science and Technology, provided the newest Indian administrative borders (updated Feb. 21). Administrative borders were SHPs (Shapefiles). SHP is a common spatial data format.

3. To create a GeoJSON from these files, we used QGIS software. The problem with the GeoJSON was that the NTDB only provides spatial data with UTM projection on the WGS84 datum. Knowing about Coordinate Reference System (CRS) is essential to understand this. A CRS consists of a coordinate system, which specifies how points in space are represented numerically, and a reference datum, which provides a standard for delineating the shape and orientation of the Earth. The UTM(Universal Transverse Mercator) projection splits the Earth into 60 zones, every 6 degrees of longitude wide, and uses a transverse Mercator projection to turn the three-dimensional Earth into a two-dimensional map.

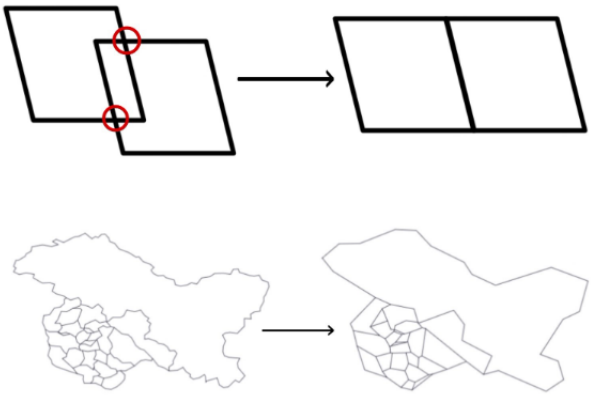
4. A location's coordinates are measured meters from the equator and Prime Meridian using a metric Cartesian coordinate system and the UTM projection on the WGS-84 datum. Geographic data is centred on Earth using the WGS-84 datum.

5. In EPSG:4326, a point's coordinates are expressed in decimal degrees, with latitude values between -90 and +90 degrees and longitude values between -180 and +180 degrees. EPSG:4326 is a non-metric system, so distances between map points cannot be accurately measured without considering the Earth's curvature. As per our proposal, we decided to use Plotly Choropleth Mapbox to plot choropleth maps which also, by default, only finds EPSG:4326. The QGIS software came in handy again to change the reference system.

6. After fixing the reference frame, it was essential to reduce the size of our data to process the tasks efficiently. We used an online tool, MAPSHAPER, to decrease the quality of shapes while also maintaining the structural integrity of the nation. This gave rise to the problem of intersecting polygons. We used MAPSHAPER again to resolve this issue, reducing the size from 150 MB to nearly 10 MB.

7. The agricultural dataset had data for 1998-2017, while our GeoJSON had districts from 2021. Five pairs of districts shared the same name across dif-

ferent states. This led to wrong mapping and no representation on the map. To sort these problems, we had to change the names by manual inspection with the help of the internet while also enhancing our general knowledge in the process.



8. The rainfall dataset used subregions instead of districts, so we had to add a subregion section to the geoJSON.

6 Expected Results

Task 1: We expect the visualization of the first task to enable us to infer the different crop production patterns along with their diversity and yield over time. Hopefully, this will allow us to draw useful insights into the geographical spread of different crops around the country.

Task 2: This running bar graph will help us understand how agricultural crop diversity and agricultural crop yield have changed over the years in India.

Task 3: This visualization will help us understand the dependence of Indian agricultural output on the monsoon. It will be interesting to note this dependence with respect to time and the subsequent improvement in irrigational practices. By querying the average amount of water required for crop production(depending on the type of crop), and the water available due to rainfall, we can estimate the stress on external and limited sources of irrigation(canals, rivers, groundwater etc).

7 Methods

Task 1

The Task 1 of our data visualisation enables us to visualise the crop production patterns in our country over the years. This helps us gain valuable insight on the geography and the history of cropping patterns in our country. We have incorporated Task 1 into our web app with the help of the Streamlit library. In the map, the user can choose a crop and view a map that shows how much of that crop was produced in various districts over time. This is done through the use of a choropleth map. The code employs the Plotly library for generating choropleth maps and the Streamlit library for showcasing the maps and the crop selection dropdown. The interactive maps enable users to hover over a district and view its name alongside the corresponding production value. The dropdown for selecting crops has been implemented through a Streamlit form. The program imports multiple datasets, which include a GeoJSON file that encompasses the geometries of Indian districts, and CSV files that contain production data for different crops. The source code additionally incorporates CSS styling elements to modify the typography and colour of the textual content within the web-based software.

Task 2

To visualize year-wise production analysis, racing bar plots from raceplotly.plots were used. We also used barchartrace library to create static(non-interactive) videos. All the crops were divided into specific crop groups, which included cereals, pulses, spices, cash crops, nuts and seeds and fruits and vegetables, to make appropriate comparisons. An interactive element was added to select one of these crop groups, or to customize and choose desired crops for a comparative analysis.

Task 3

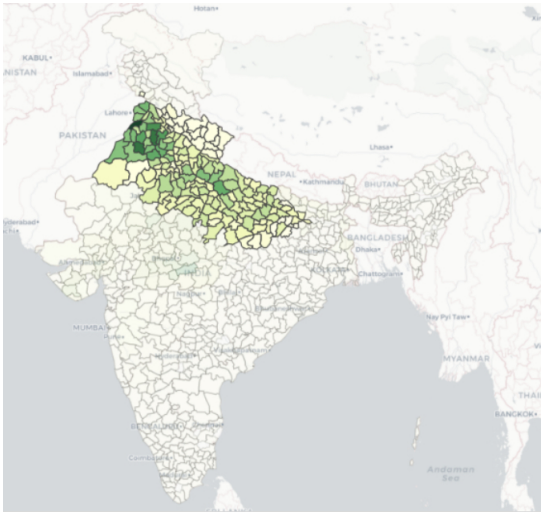
For task 3, we've created an innovative tool that shows how Crop Production and Rainfall Datasets are related. The tool showcases the annual rainfall and crop production plots of India in a side-by-side manner, with the rainfall plot on top and the crop production plot at the bottom. Just like in Task 1, we've gone ahead and incorporated Task 3 into our web app with the help of the Streamlit library. The first plot showcases a choropleth map of India, with varying

hues indicating the annual precipitation in distinct districts. In the bottom map, the user can choose a crop and view a map that shows how much of that crop was produced in various districts over time. This is done through the use of a choropleth map. The code employs the Plotly library for generating choropleth maps and the Streamlit library for showcasing the maps and the crop selection dropdown. The interactive maps enable users to hover over a district and view its name alongside the corresponding production value or rainfall. The dropdown for selecting crops has been implemented through a Streamlit form. The program imports multiple datasets, which include a GeoJSON file that encompasses the geometries of Indian districts, a CSV file that contains yearly precipitation data for various districts, and CSV files that contain production data for different crops. The source code additionally incorporates CSS styling elements to modify the typography and colour of the textual content within the web-based software.

8 Analysis

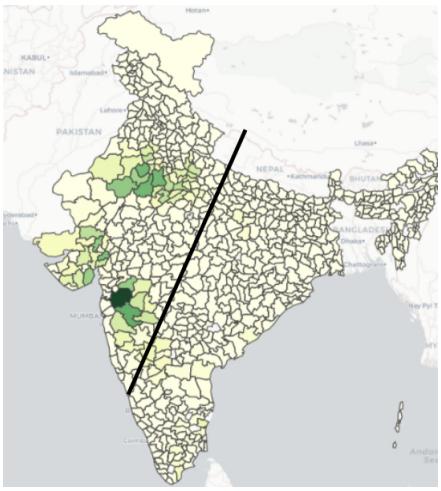
Task 1

1. Wheat Cultivation in India - Changing Patterns: Traditionally, the Northern Plains(Indo-Gangetic Plains) have been the primary cultivators of wheat. These included states like Punjab, Haryana, and Uttar Pradesh which can be seen on the wheat map in the 1990s. With time, this cultivation pattern has been extended further south to encompass areas of Rajasthan and Madhya Pradesh. Today, Madhya Pradesh and Rajasthan have joined the list of the largest wheat-producing states. Govt. policies and missions like NFSM, RKVY, PMFBY, subsidies, and farmer knowledge are likely causes for this change.[8]

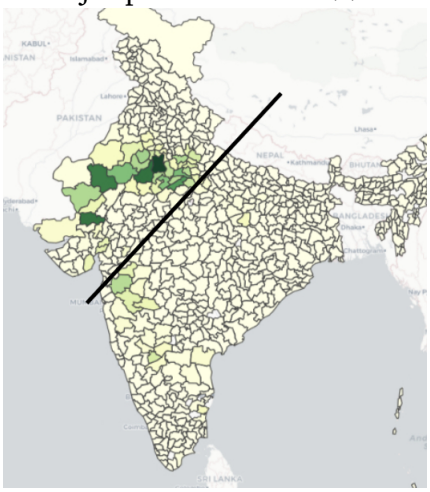


Wheat Production in India in the Indo-Gangetic Plains

2. Bajra Cultivation change: Bajra was produced the most in Maharashtra during the early years with Rajasthan next in line. In recent years, production has dropped significantly in Maharashtra, while it has primarily concentrated in Rajasthan.



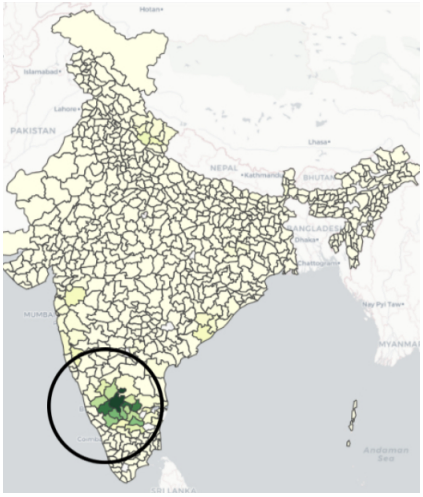
Bajra production in 1998



Bajra production in 2017

3. Other observations: The cultural culinary significance of certain crops are also seen.

- For example, Ragi cultivation is concentrated in the South Karnataka region where it is consumed regularly.
- It is also seen that Jowar is primarily cultivated in Maharashtra, because of the black regur soil which is rich in lime, potash-magnesium, carbonates, allumina and calcium.[3]
- Arhar/Tur dal cultivation is concentrated in the East Maharashtra(Marathwada) region. It produces 28% of the national production with the primary districts being Latur and Hingoli districts in Marathwada and Akola district in Vidarbha. [6]
- We also note that Jalgaon district in Maharashtra has extremely high banana production and is known as the "Banana Town" of India. It even is on the globabl list of highest banana-producing districts.[5]



Ragi Production in India

Task 2

Cereals:

1. Rice and Wheat were cultivated most, over all the years.

2. Bajra production peaked in 2012 and decreased henceforth.

3. Jowar production decreased overall, and consistently after 2008.

4. Ragi production has remained stable, decreasing slightly since 2013

Conclusion:

1. Millet production has decreased over the years, especially since 2010. The proportion of other crops has decreased.[1]

Pulses:

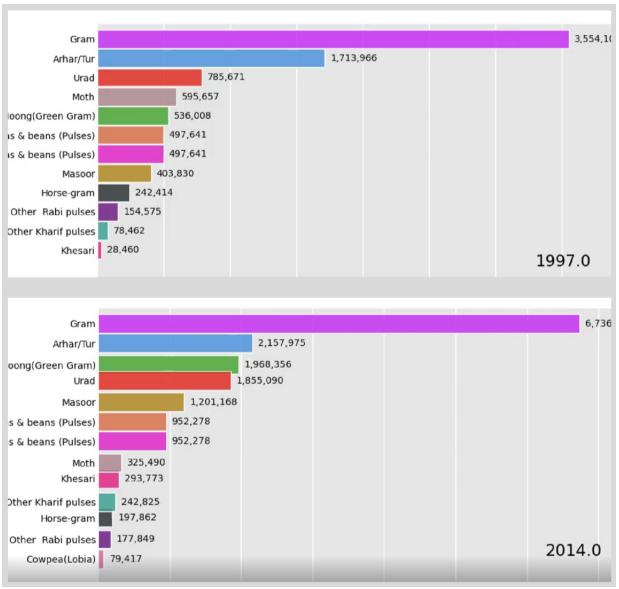
1. Proportions of Gram, Arhar and Urad have remained mostly consistent, and the production has increased.

2. Proportion of moong has decreased in comparison.

3. Moth and Horse-gram production has decreased.

Conclusion:

1. Crop diversity has largely been maintained.

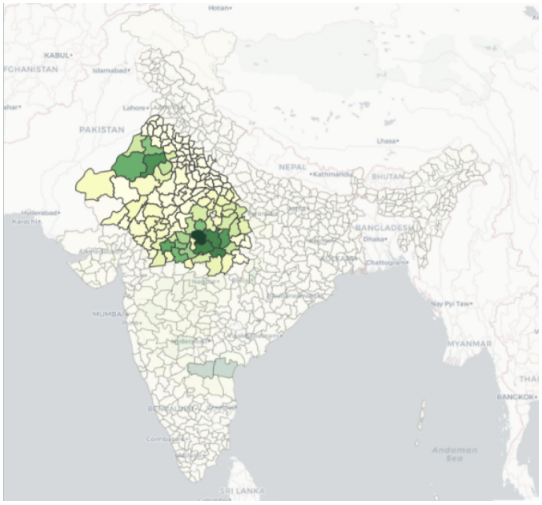


Racing Bar Plot

Task 3: Analysis

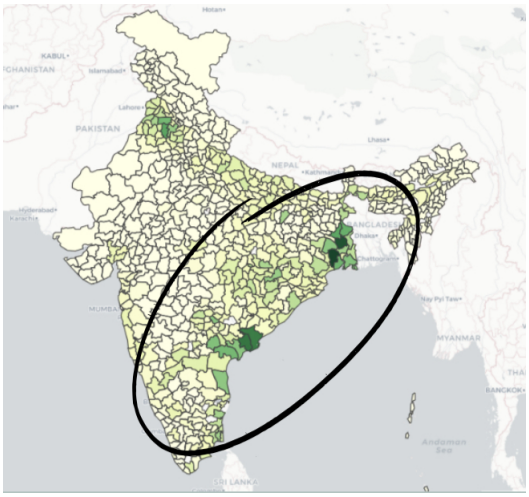
Based on the visual analysis of the plots, the following are some observations:

1. The dry land crops such as Bajra and Gram are mostly grown in Rajasthan and Madhya Pradesh, which have low rainfall. These crops are adapted to the dry and arid conditions and can survive with less water.



Gram Production in India

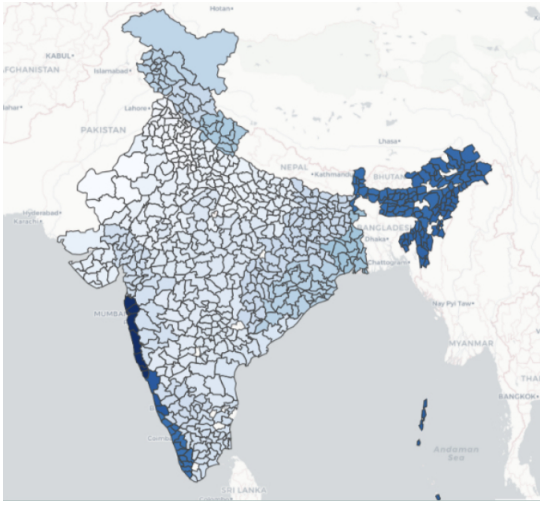
2. The eastern regions of India receive rainfall in Nov-Dec, which is due to retreating monsoons. This explains why rice is cultivated as a Rabi crop in South/East India. The retreating monsoon period provides the required water for rice cultivation during this time of the year. Wheat, on the other hand, requires water from rivers and is grown in the Indo-Gangetic plains. This region is well-irrigated due to the presence of several rivers, which makes it an ideal location for wheat cultivation.



Rice Production in India

3. Soybean cultivation is concentrated in Madhya Pradesh and Maharashtra, which have black soil and can withstand waterlogging caused by monsoon patterns. The crop is adapted to the specific soil type and weather conditions in these regions, which makes it a viable option for cultivation.[7]

4. It was also observed that the Punjab/Haryana region of the country despite having mediocre rainfall is responsible for the production of a water-intensive crop like rice in large quantities. This should put lot of pressure on the river and groundwater resources from our analysis. This is consistent with the geological surveys indicating a depleting water table. This is also true for the Bundelkhand region of UP/MP where sugarcane(another water intensive crop) production is concentrated.[4][2]

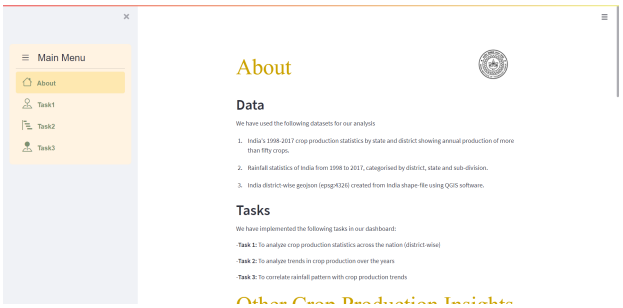


Average Rainfall in India

Based on the above observations, it can be concluded that crop production in India is largely influenced by the distribution of rainfall and soil types in different regions. The cultivation of specific crops is linked to the availability of water and soil conditions required for their growth. This highlights the importance of understanding the local environmental conditions and adopting suitable agricultural practices to maximise crop yields.

9 Streamlit

We used the Streamlit library in python to make a standalone dashboard. We used this library as it was easy to use and also provided support for all kinds of plotly charts. We used `streamlit.plotly_chart` to embed it into the dashboard. Additionally, we used the widgets offered by streamlit for making the plots interactive. The temporal information was coded using an inbuilt slider which plotly adds with choropleth and barplot charts. We also added a navigation sidebar for switching between the task visualisations.



Streamlit dashboard

10 Challenges

We faced many challenges in this project. Some of these include:

- Rainfall data was given region-wise, while the crop data was available district wise. In this case, we had to map each district to a corresponding rainfall region on the geojson file.
- Some of the district names were changed overtime and new districts were created. Additionally, there were districts with the same name and also some differences in spellings of districts between the crop data and the geojson from the National Topographical Database (NTDB). This created a challenge in creating an appropriate geojson file.
- We faced problems in cleaning the crop data as it had mistypes and also Null values. Some of the crop data was missing for certain years and so these years were dropped(2018-2023).
- There were some challenges in making our initial code standalone using streamlit as we have used `bar_chart_race` library to make the initial racing bar plots for the crop data. This wasn't interactive and wasn't embedaable as a plotly chart in streamlit.

11 Conclusion

In conclusion, the project has successfully achieved its objectives and provided valuable insights into the crop production patterns in India. Through Task 1 for crop production analysis, we were able to gain geographical and historical insights on the cropping patterns and the effect of new policies and initiatives.Through Task 2 for cumulative crop production analysis, we were able to gain historical insights on the crop diversity and the changing trends in different crop productions especially in the context of specific crop types. Through Task3 for rainfall correlation, we were able to gain geographical insights on types of crops grown in different parts of the country(water-intensive, dry crops etc). We were also able to predict the strain on ground and river water resources in certain regions. Overall, the project has contributed to a better understanding of crop patterns in India and can serve as a useful resource for future research in this area.

Moving forward, there are several potential avenues for further exploration and improvement. These include adding data for more crops and more, including soil quality and type correlation, map of the difference between the rainfall required and the received rainfall for each crop and region, and a crop suggestion mechanism which suggests crops to be chosen to cultivate based on season, rainfall and soil type. By addressing these areas, we can continue to build upon the success of this project and further expand our understanding of the cropping patterns in India.

Overall, we are confident that this project will have a positive impact on agricultural study, and we look forward to how its findings and insights could be used to drive progress and innovation in the future.

12 Code Link

Crop-data Visualization

13 Contribution

GROUP-8

Member	Contribution
Data Preprocessing	Bhavya, Arya
Task 1	Arya, Bhavya
Task 2	Gauri, Pradeep
Task 3	Arya, Bhavya, Kaushik
Sreamlit	Pradeep, Gauri, Kaushik
Documentation	Arya, Gauri, Bhavya, Pradeep

References

[1] *As india tries to popularise millets it should learn from the cautionary tale of quinoa*. URL: https://scroll.in/article/1045442/as-india-tries-to-popularise-millets-it-should-learn-from-the-cautionary-tale-of-quinoa#:~:text=Millets%20once%20accounted%20for%20about,and%2077%25%20in%20urban%20areas. (visited on 04/25/2023).

[2] L Chourasia and Dalchand Jhariya. "Water Crisis in the Bundelkhand Region: An Observation". In: *IOP Conference Series: Earth and Environmental Science* 597 (Dec. 2020), p. 012024. DOI: 10.1088/1755-1315/597/1/012024.

[3] *Give a geographical reason for each of the following i jowar*. URL: https://www.knowledgeboat.com/question/give-a-geographical-reason-for-each-of-the-following-i-jowar--11963080123057402O (visited on 04/25/2023).

[4] *Groundwater depletion in haryana a cause of serious concern*. URL: https://india.mongabay.com/2021/07/groundwater-depletion-in-haryana-a-cause-of-serious-concern/#:~:text=The%20annual%20groundwater%20withdrawal%20in,challenges%20with%20respect%20to%20groundwater. (visited on 04/25/2023).

[5] *How jalgaon banana city of india became 7th biggest banana producer in world*. URL: https://krishijagran.com/agriculture-world/how-jalgaon-banana-city-of-india-became-7th-biggest-banana-producer-in-world/ (visited on 04/25/2023).

[6] *Maharashtra may see shift in cultivation towards foodgrain*. URL: https://www.livemint.com/Politics/S2r7kPCMgEfQchUS011oOO/Maharashtra-may-see-shift-in-cultivation-towards-foodgrain.html (visited on 04/25/2023).

[7] *Soybean the most cultivated crop in madhya pradesh*. URL: https://bansaloilandfoods.com/soybean-the-most-cultivated-crop-in-madhya-pradesh/#:~:text=The%20Malwa%20region%20of%20Madhya,%2C%20Neemuch%2C%20Shajapur%20and%20Rajgarh. (visited on 04/25/2023).

[8] *Wheat in madhya pradesh india*. URL: https://www.prepdata.org/dashboards/wheat-in-madhya-pradesh-india (visited on 04/25/2023).