

ZAALIMA DEVELOPMENT

DATA ANALYTICS REPORT



NAME: SOUMIT MAL

BRANCH: B.TECH CSE (DATA SCIENCE)

INTERNSHIP DURATION: 8 WEEKS

ORGANIZATION: ZAALIMA DEVELOPMENT

DECLARATION

I, Soumit Mal, declare that this internship report titled “Data Analytics Internship at Zaalima Development” is my original work and has been prepared based on the tasks, learning, and experiences gained during my 8-week internship. The content of this report truly reflects the work I carried out, including data analysis, visualization, and dashboard creation. I confirm that all information presented is accurate to the best of my knowledge and has not been submitted elsewhere for any academic or professional purpose. Any external references or resources used have been properly acknowledged.

ACKNOWLEDGMENT

I would like to express my heartfelt gratitude to Zaalima Development for providing me with the opportunity to work as a Data Analytics Intern. This internship has been a transformative learning experience, allowing me to apply the theoretical knowledge from my B.Tech (CSE – Data Science) program to real-world scenarios. I am especially thankful to the Data Analytics team for their consistent guidance, constructive feedback, and technical support throughout the duration of the project.

Their mentorship helped me understand the practical aspects of data wrangling, exploratory analysis, and visualization. The inclusive and collaborative work environment at Zaalima Development greatly enriched my learning journey and encouraged me to take initiative and solve problems independently. I also appreciate the opportunity to work on a meaningful project that sharpened my analytical and reporting skills. This experience has significantly contributed to my academic and professional development, and I am truly thankful to everyone who supported me during this internship.

CONTENTS

S.NO	CONTENTS	PG. NO
1	INTRODUCTION	5
2	OBJECTIVE	6
3	TOOLS AND TECHNOLOGIES USED	7
4	DATASET DESCRIPTION	8
5	AGRICULTURAL EDA SUMMARY	9
6	GRAPHS AND CHARTS	10
7	POWER BI ANALYSIS	11
8	CHALLENGES FACED	12
9	FUTURE TECHNOLOGIES	13
10	CONCLUSION	14

INTRODUCTION

This internship was undertaken as part of the academic requirements for the B.Tech program in Computer Science and Engineering (Data Science). The primary objective was to gain practical exposure to the field of data analytics, develop technical skills in tools like Python, Power BI, and DAX, and apply theoretical knowledge to real-world datasets.

The project focused on analyzing agricultural data to uncover patterns in crop yield, rainfall, fertilizer usage, and regional productivity across India. Through tasks such as data cleaning, exploratory data analysis (EDA), feature engineering, and dashboard creation, the internship provided a complete view of the data analytics lifecycle.

Working with structured datasets and visualization tools enhanced problem-solving abilities, critical thinking, and technical fluency in handling large volumes of data. This experience not only bridged the gap between academics and industry but also laid the groundwork for a future career in data science and analytics.

OBJECTIVE

The primary objectives were:

- To gain hands-on experience in the field of Data Analytics using real-world datasets.
- To apply concepts of data cleaning, preprocessing, and transformation for accurate analysis.
- To enhance technical skills in tools such as Python, Pandas, Power BI, and DAX.
- To develop interactive dashboards and visual reports for meaningful data storytelling.
- To explore data-driven decision-making from both technical and business perspectives.
- To identify patterns, trends, and correlations through exploratory data analysis (EDA).
- To understand the workflow of end-to-end analytics projects, from raw data to insights.
- To strengthen analytical thinking and problem-solving abilities through practical challenges.

This internship was aligned with the academic goals of my course and enriched my understanding of how data science is implemented in the industry.

TOOLS AND TECHNOLOGIES USED

TECHNOLOGY	PURPOSE
PYTHON (PANDAS, MATPLOTLIB, SEABORN)	DATA CLEANING & EDA
JUPYTER NOTEBOOK	SCRIPT EXECUTION & VISUALIZATION
POWER BI	DASHBOARD DESIGN & VISUALIZATION
MICROSOFT EXCEL	DATA MANIPULATION & REPORTING
SQL	QUERYING DATASETS
DAX	DATA ANALYSIS EXPRESSIONS IN POWER BI

DATASET DESCRIPTION

The dataset consisted of 19,689 records with 13 columns, focusing on agricultural crop data across various Indian states over multiple years. It contained essential details such as crop type, cultivation area, yield, rainfall, and geographical indicators. This data was used to analyze trends, develop insights, and support data-driven decisions in agriculture.

COLUMN-WISE DESCRIPTION

- **Crop:** Name of the cultivated crop (e.g., Rice, Wheat, Cotton).
- **Crop_Year:** Year in which the crop was grown or harvested.
- **Season:** Agricultural season (e.g., Kharif, Rabi, Whole Year).
- **State:** Indian state where the crop was cultivated.
- **Area:** Total cultivated area (in hectares).
- **Production:** Crop production in tonnes.
- **Annual_Rainfall:** Average rainfall (in mm) for the respective year and region.
- **Fertilizer:** Quantity of fertilizers applied (in kg/ha).
- **Pesticide:** Quantity of pesticides used (in kg/ha or liters).
- **Yield:** Calculated as $\text{Production} / \text{Area}$ (tonnes per hectare).
- **Region:** Geographical region (North, South, East, West, Central, Northeast).
- **Latitude & Longitude:** Geographical coordinates of the state, used for mapping and spatial analysis.

This multidimensional dataset enabled correlation analysis, regional yield comparison, and trend identification. It was cleaned, processed, and visualized using Python and Power BI for deriving meaningful agricultural insights.

AGRICULTURAL EDA SUMMARY

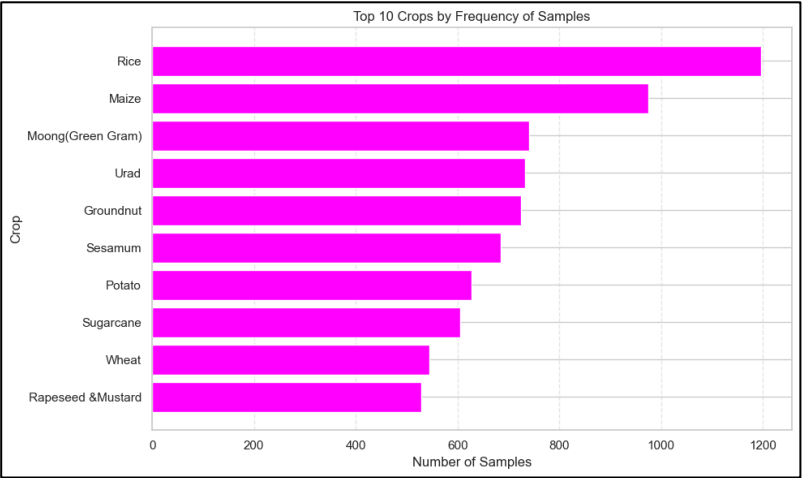
The file `Regional_Crop_Yield_Analysis.ipynb` is a Jupyter Notebook developed in VS Code for analyzing agricultural crop data. It focuses on uncovering patterns in crop production, rainfall, fertilizer use, and yield across various Indian states and regions.

CORE COMPONENTS

- **Data Preparation:** Loaded and cleaned a structured dataset containing crop-wise records. Handled missing values and standardized columns for consistent analysis.
- **Feature Engineering:** Calculated yield as the ratio of production to area, enabling productivity analysis across regions and seasons.
- **Exploratory Data Analysis (EDA):** Utilized statistical summaries and visualizations to explore trends in rainfall, pesticide and fertilizer use, and crop output. Plots included line charts, bar graphs, histograms, and heatmaps.
- **Geospatial Mapping:** Used latitude and longitude coordinates to create region-wise yield maps, enabling spatial analysis and hotspot identification.
- **Correlation Analysis:** Assessed relationships between rainfall, fertilizer usage, and crop yield using correlation matrices and scatter plots.

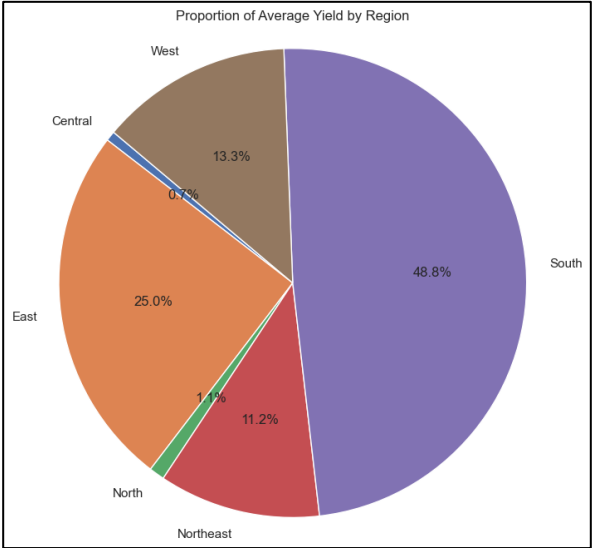
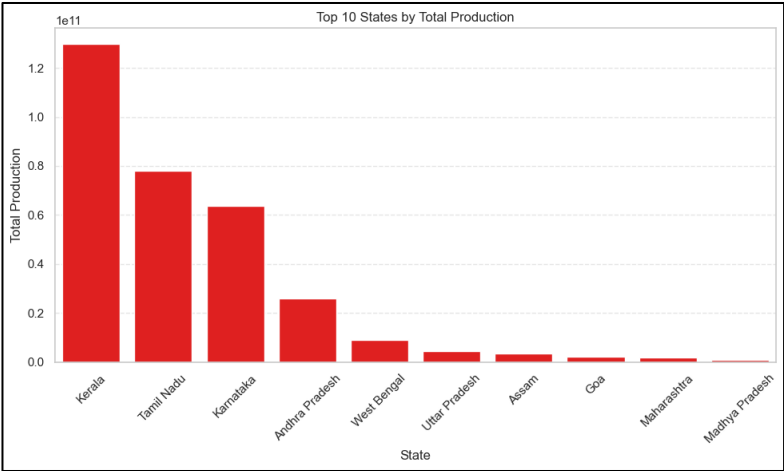
This notebook served as a foundational step in analyzing real-world agricultural data, supporting visualization and dashboard creation in further stages.

GRAPHS AND CHARTS

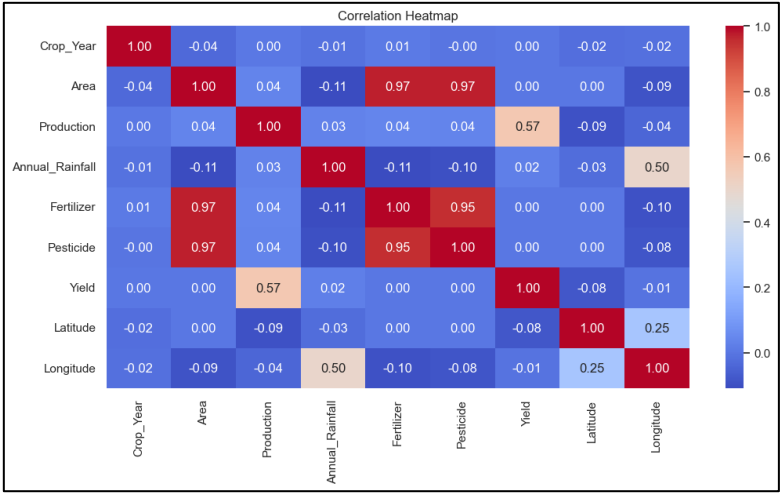


FREQUENCY OF SAMPLES

STATE-WISE TOTAL PRODUCTION

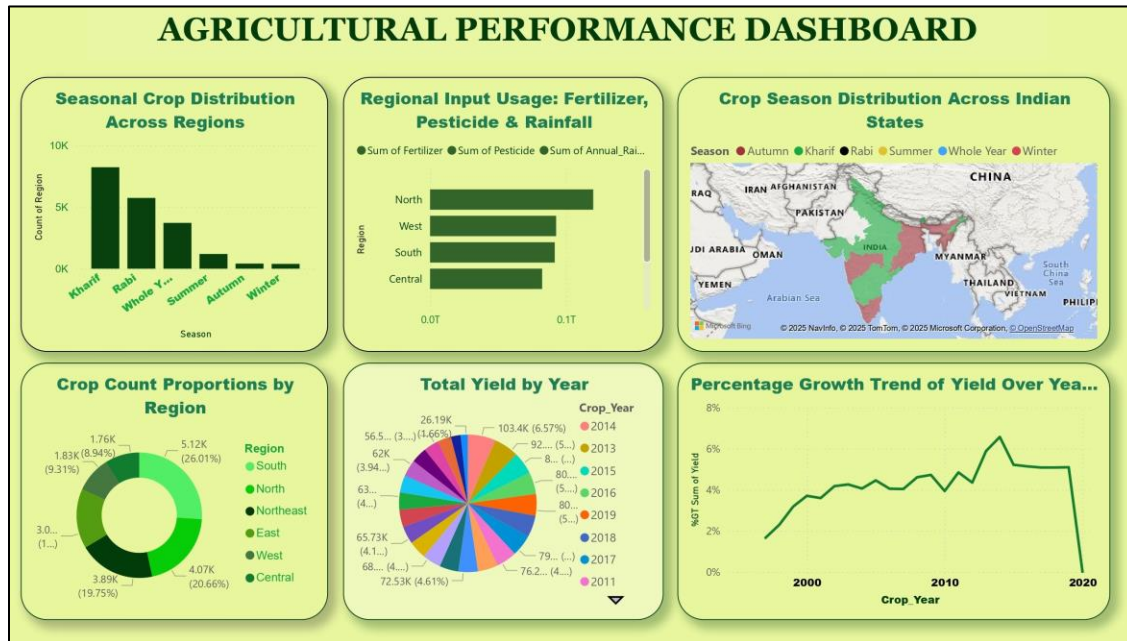


REGIONAL CONTRIBUTION TO AVERAGE CROP YIELD



CORRELATION HEATMAP

POWER BI ANALYSIS



The Power BI dashboard provides a professional and interactive analysis of agricultural data, covering crop yield, rainfall, fertilizer and pesticide usage across Indian states and regions. Built using a star schema with proper relationships between fact and dimension tables (Crop, State, Region, Season, and Year), the model supports efficient filtering and drill-down. Advanced DAX measures such as Total Yield, YoY Growth, Fertilizer Efficiency, and Region-wise Contribution enhance decision-making insights. Visuals include line charts, maps, bar graphs, and KPI cards, making the data easily interpretable for both academic and professional use.

CHALLENGES FACED

Several challenges were encountered that contributed to valuable learning:

- **Data Quality Issues:** Missing values and inconsistencies in crop production, rainfall, and input data required extensive cleaning and validation.
- **Complex Relationships:** Establishing meaningful links between diverse fields like rainfall, fertilizer usage, and yield demanded careful analysis and logical structuring.
- **DAX Learning Curve:** Writing optimized and nested DAX measures for performance tracking and dynamic visualizations was initially difficult.
- **Visualization Design:** Selecting the most impactful chart types and layout to make the dashboard user-friendly and insightful required multiple iterations.
- **Time Management:** Balancing academic work with internship deadlines was a key challenge that improved time prioritization skills.
- **Performance and Responsiveness:** With nearly 20,000 records, performance optimization became necessary, especially for slicers and map visuals. Best practices like summarization and column indexing were applied.

FUTURE TECHNOLOGIES

The field of data analytics in agriculture is rapidly evolving, and this project opens doors to several future possibilities:

- **Machine Learning Models:** Predictive modeling for yield forecasting and rainfall impact using regression, decision trees, or XGBoost.
- **NLP for Reports:** Automated insight generation using natural language processing to generate textual summaries from dashboards.
- **Cloud-Based Dashboards:** Deploying Power BI dashboards to the cloud for real-time data updates and accessibility across devices.
- **IoT Integration:** Connecting real-time sensor data (e.g., soil moisture, weather) for dynamic dashboards and smarter farming insights.
- **Advanced GeoAnalytics:** Integrating GIS and satellite data for spatial yield analysis and precision agriculture planning.

CONCLUSION

This project was a valuable opportunity to apply data analytics in a real-world domain. It involved a complete cycle—from data cleaning and feature engineering to dashboard design and insight generation. Core skills in Python, Power BI, DAX, and data visualization were not only developed but applied in solving a practical problem.

The experience improved understanding of structured data modeling, analytical logic, and effective storytelling through visuals. It also emphasized the importance of clean data, meaningful relationships, and performance-aware design. Overall, this project provided a solid foundation for future work in data science, analytics, and applied technology.