AN INTERNSHIP FINAL REPORT ON

SUBMITTED TO THE

SAVITRIBAI PHULE PUNE UNIVERSITY, PUNE

BACHELOR OF ENGINEERING (COMPUTER ENGINEERING)

SUBMITTED BY

OMKAR BABANRAO LATPATE
University PRN :72292591D

UNDER SUPERVISION OF

EDUNET FOUNDATION, IN COLLABORATION WITH EY GDS-AICTE
DURATION: 6 WEEKS

DEPARTMENT OF COMPUTER ENGINEERING

MODERN EDUCATION SOCIETY'S

WADIA COLLEGE OF ENGINEERING, PUNE-411001

Approved by AICTE,

Affiliated to **SAVITRIBAI PHULE PUNE UNIVERSITY**

A. Y. 2023-24 SEM-II

DEPARTMENT OF COMPUTER ENGINEERING

MODERN EDUCATION SOCIETY'S

WADIA COLLEGE OF ENGINEERING, PUNE-411001

CERTIFICATE

This is to certify that the "**Internship report"** submitted by **Omkar Babanrao Latpate (University PRN:72292591D)** is work done by him and submitted during 2023-24 academic year, in partial fulfilment of the requirements for the award of the degree of **BACHELOR OF ENGINEERING in COMPUTER ENGINEERING,** at **MES Wadia College of Engineering.**

College Internship Coordinator,
MES WCoE, Pune

Department Internship Coordinator
Department of Computer Engineering

Head of Department,
Department of Computer Engineering, MES WCoE Pune

Place: Pune
Date: 20-04-202

INTERNSHIP OFFER LETTER

Date: 5th March 2024

Dear Mr./Ms. OMKAR BABANRAO LATPATE

AICTE Student ID: STU657dc039e78e81702740025
AICTE Internship ID: INTERNSHIP_170003939765548ae545859

We extend our warmest congratulations on your selection for an **internship presented by the Edunet Foundation, in collaboration with EY GDS-AICTE**, focusing on Advanced Data Analytics.

The EY GDS - AICTE Internship is structured to support individuals in developing essential foundational skills needed for productive careers in the IT sector. Participants have the opportunity to earn credentials and benefit from guidance provided by industry experts, all at no cost throughout the duration of the program.

EY GDS (Ernst & Young – Global Delivery Services), in partnership with AICTE, provides a unique learning experience through a 6-week Internship, commencing from **28th February 2024 to 16th April 2024.** Throughout this period, you will have the opportunity to work independently on a project, with guidance from a mentor who will assist you in identifying solutions and developing them into a tangible project.

Benefits:

- Personalized mentorship sessions and collaborative group learning.
- Opportunities to expedite learning through project-based internships.
- A holistic learning experience provided by industry experts through knowledge-sharing sessions.
- Showcase your skills by creating prototypes to solve real-world challenges.
- Earn certifications from EY, AICTE, and Edunet, boosting your confidence and value to potential future employers.
- Opportunity to present your project prototypes to a panel of industry experts at a regional showcase event.

Timeline and the Project:

Week	Agenda	Student Deliverables
Week 0	Orientation of Internship, Project allocation	• Selection of Projects through Google form • Commence Project Related Tasks
Week 1	Master session on Resume Building, Mentorship sessions, and Ask Me Anything sessions	• Complete the Weekly Milestone and feedback form • Project related tasks • Self-paced learning
Week 2	Master session on Design Thinking in Real World Applications, Mentorship sessions, and Ask Me Anything sessions	• Complete the Weekly Milestone and feedback form • Project related tasks Self-paced learning
Week 3	Master session on Managing codebase with Git Hub, Mentorship sessions, and Ask Me Anything sessions	• Complete the Weekly Milestone and feedback form • Project related tasks • Self-paced learning

INTERNSHIP OFFER LETTER

Week 4	Master session on DevOps: Code to Production, Mentorship sessions, Ask Me Anything sessions, and Final Project Submission	• Complete Weekly Milestone • Self-Paced learning • Final project and PPT Submission
Week 5	Mock project presentations	• Presentation of Project in front of Subject Matter Experts • Incorporate changes in the project mentioned by Mentors
Week 6	Final Project Presentations (PPT)	• Final Presentation of Project before EY Industry Experts

Criteria for certification:

- Participation in master sessions with EY & EF experts is mandatory
- Participation in mentorship sessions with EF experts is mandatory
- Completion of tasks/milestones every week
- Submission of a project presentation required
- Commitment of 6-8hrs/week throughout internship

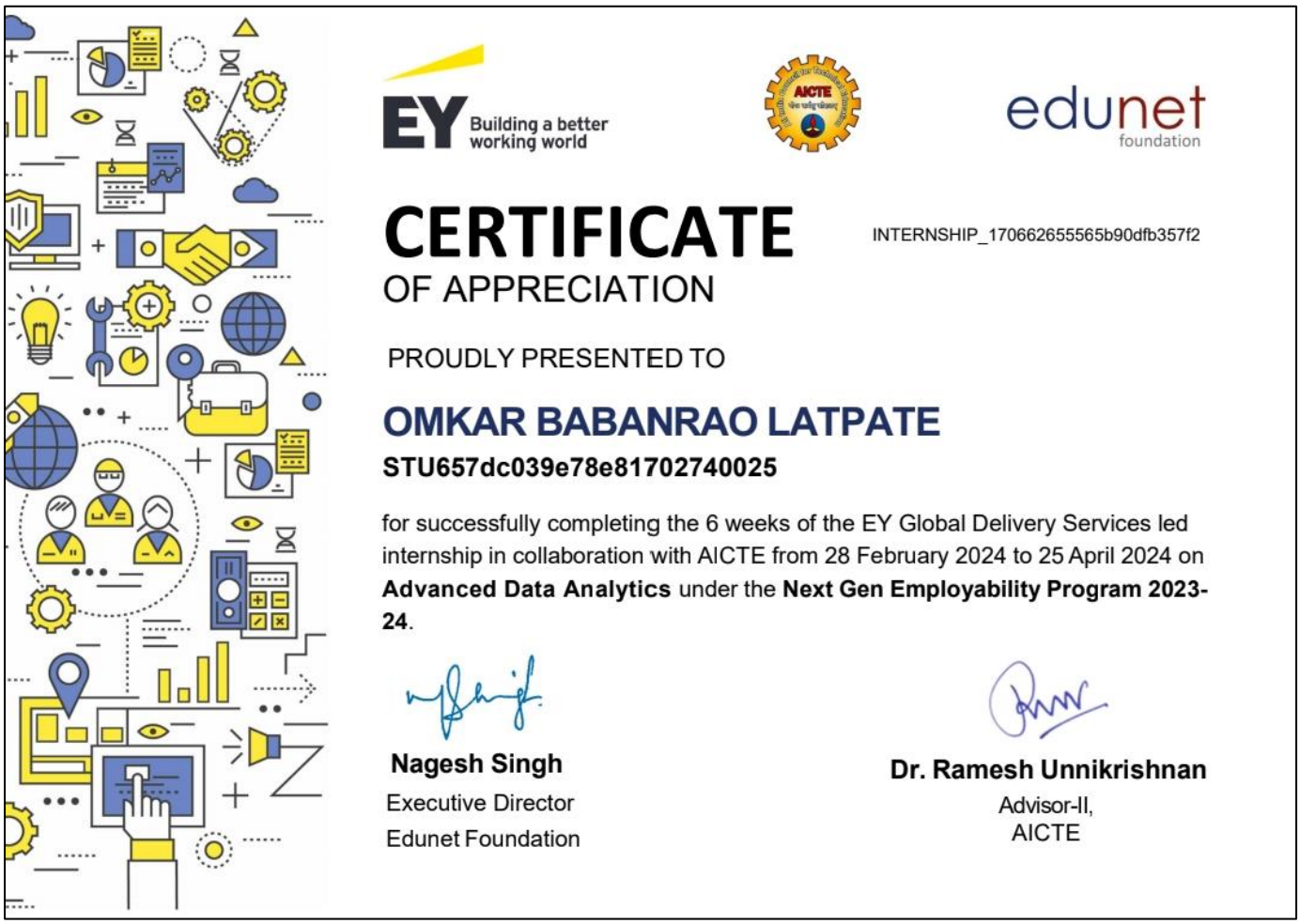
Stipend:

There will be NO stipend for this internship.

We wish you a great learning experience during the internship. Thank you!

Sincerely,
Nagesh Singh
Executive Director – Edunet Foundation

INTERNSHIP COMPLETION LETTER



Building a better working world

edunet foundation

CERTIFICATE
OF APPRECIATION

INTERNSHIP_170662655565b90dfb357f2

PROUDLY PRESENTED TO

OMKAR BABANRAO LATPATE

STU657dc039e78e81702740025

for successfully completing the 6 weeks of the EY Global Delivery Services led internship in collaboration with AICTE from 28 February 2024 to 25 April 2024 on **Advanced Data Analytics** under the **Next Gen Employability Program 2023-24**.

Nagesh Singh
Executive Director
Edunet Foundation

Dr. Ramesh Unnikrishnan
Advisor-II,
AICTE

ACKNOWLEDGEMENT

I would like to extend my heartfelt gratitude to Edunet Foundation for providing me with the opportunity to undertake my internship in collaboration with EY-GDS and AICTE in the domain of Advanced Data Analytics. This experience has been invaluable in honing my skills and gaining practical insights into the field.

I am deeply indebted to Prof. Gajanan Aochar and Dr. (Mrs.) Shalaka Deore for their unwavering support, guidance, and mentorship throughout this journey. Their expertise and encouragement have played a pivotal role in shaping my learning experience.

I would also like to express my sincere appreciation to Dr. (Mrs.) N. F. Shaikh for their constant encouragement and support, which has motivated me to strive for excellence.

Lastly, I extend my thanks to all the faculty members and staff of **Modern Education Society's Wadia College of Engineering, Pune** for their support and encouragement.

Omkar Babanrao Latpate
University PRN: 72292591D

ABSTRACT

- **Company Information**:

EY GDS (Ernst & Young Global Delivery Services) is a renowned global leader in professional services. Committed to corporate social responsibility (CSR), EY extends its dedication to skill development through various initiatives. One such initiative is the collaboration with Edunet & AICTE for the Next Gen Employability Program, aimed at enhancing students' employability in the technical education ecosystem.

- **Programs and opportunities:**

The EY GDS-AICTE Internship offers a unique opportunity to dive into the world of data analytics and full-stack web development. Through a 6-week virtual internship camp, participants gain hands-on experience and mentorship from industry experts. They work on real-world challenges, developing project prototypes using their preferred technology track.

- **Methodologies:**

Participants undergo intensive training in data analytics, Power BI, and related concepts. They analyze 19 years of Indian agriculture data, exploring various parameters of crop production, prediction based on historical data, and determining Minimum Support Price for crops.

- **Key parts of the report**:
  - Understanding the basics of data analytics and Power BI.
  - Analyzing Indian agriculture data: identifying top crops, states, and districts by production.
  - Exploring production trends across different seasons and years.
  - Predicting future production and assessing Minimum Support Price for crops.

- **Benefits of the Company/Institution through our report:**

The report showcases the effectiveness of the EY GDS-AICTE Internship program in enhancing participants' technical skills and problem-solving abilities. It highlights the collaboration between EY GDS, Edunet, and AICTE in fostering employability among students. Additionally, it demonstrates the practical application of data analytics in addressing real-world challenges, contributing to the participants' professional growth and the advancement of the agriculture sector.

Table of Contents

WEEKLY OVERVIEW OF INTERNSHIP ACTIVITIES

	DATE	DAY	NAME OF THE TOPIC/MODULE COMPLETED
ZERO WEEK			
	28-02-24	Wednesday	Orientation of Internship, Project allocation
	29-02-24	Thursday	
	01-03-24	Friday	Selection of Projects through Google form
	02-03-24	Saturday	Allocation of project

	DATE	DAY	NAME OF THE TOPIC/MODULE COMPLETED
FIRST WEEK	04-03-24	Monday	Commence Project Related Tasks
	05-03-24	Tuesday	Mentorship sessions- Data Preprocessing
	06-03-24	Wednesday	Project related tasks
	07-03-24	Thursday	Master session on Design Thinking in Real World Applications
	08-03-24	Friday	Self-paced learning
	09-03-24	Saturday	Ask Me Anything sessions

	DATE	DAY	NAME OF THE TOPIC/MODULE COMPLETED
SECOND WEEK	11-03-24	Monday	Complete Weekly Milestone and Submit
	12-03-24	Tuesday	Mentorship sessions-Visualization in Power BI
	13-03-24	Wednesday	Project related tasks
	14-03-24	Thursday	Master session on Managing codebase with Git Hub
	15-03-24	Friday	Self-paced learning
	16-03-24	Saturday	Ask Me Anything sessions

	DATE	DAY	NAME OF THE TOPIC/MODULE COMPLETED
THIRD WEEK	18-03-24	Monday	Complete Weekly Milestone and Submit
	19-03-24	Tuesday	Mentorship sessions-DAX Functions in Power BI
	20-03-24	Wednesday	Project related tasks
	21-03-24	Thursday	Master session on DevOps: Code to Production.
	22-03-24	Friday	Self-paced learning
	23-03-24	Saturday	Ask Me Anything sessions

	DATE	DAY	NAME OF THE TOPIC/MODULE COMPLETED
FOURTH WEEK	25-03-24	Monday	Complete Weekly Milestone and Submit
	26-03-24	Tuesday	Mentorship sessions-MSP & Formatting visuals and canvas background.
	27-03-24	Wednesday	Project related tasks
	28-03-24	Thursday	Master session on DevOps: Code to Production.
	29-03-24	Friday	Self-paced learning
	30-03-24	Saturday	Ask Me Anything sessions

	DATE	DAY	NAME OF THE TOPIC/MODULE COMPLETED
FIFTH WEEK	01-04-24	Monday	▪ Mock project presentations ▪ Presentation of Project in front of Subject Matter Experts ▪ Incorporate changes in the project mentioned by Mentors
	02-04-24	Tuesday	
	03-04-24	Wednesday	
	04-04-24	Thursday	
	05-04-24	Friday	
	06-04-24	Saturday	

	DATE	DAY	NAME OF THE TOPIC/MODULE COMPLETED
SIXTH WEEK	08-04-24	Monday	▪ Final Project Presentations (PPT) ▪ Final Presentation of Project before EY Industry Experts
	09-04-24	Tuesday	
	10-04-24	Wednesday	
	11-04-24	Thursday	
	12-04-24	Friday	
	13-04-24	Saturday	

1. INTRODUCTION

In today's dynamic business landscape, organizations are continuously exploring innovative technologies to enhance their efficiency, scalability, and cost-effectiveness. One such transformative technology is Power BI, which is revolutionizing the way businesses harness data to drive insights and decision-making.

My endeavor was to delve into the realm of Power BI and its application in the exhaustive analysis of the Indian Agriculture Sector (DA). With a focus on leveraging data analytics to gain comprehensive insights, my aim was to uncover the vast potential of Power BI in transforming agricultural practices and policies.

Throughout this internship, I embarked on a journey to understand the intricacies of Power BI-driven analysis in the context of the Indian agriculture sector. By diving into 19 years of agricultural data, my exploration sought to unravel critical aspects such as crop production dynamics, predictive modeling, and minimum support price (MSP) determination.

Power BI is reshaping the way organizations leverage data to drive informed decisions. By harnessing the capabilities of Power BI, businesses in the agriculture sector can address key challenges and seize opportunities for optimization and innovation.

Through real-world case studies and practical applications, I aimed to uncover the advantages and challenges associated with Power BI-driven analysis in agriculture. By the end of this internship, my goal is to equip myself with the knowledge and skills necessary to make informed decisions about leveraging Power BI for agricultural data analysis. These learnings will not only benefit me personally but also provide valuable insights for businesses seeking to enhance their operations and drive growth through data-driven strategies in the agriculture sector.

2. PROBLEM STATEMENT AND OBJECTIVES

2.1 PROBLEM STATEMENT

To optimize decision-making in the Indian agriculture sector, this internship project aims to leverage Power BI for exhaustive analysis. By harnessing Power BI, the project seeks to address key challenges in the sector, including the complexity of agricultural data, the need for predictive analysis, determining Minimum Support Prices (MSP), scalability, accessibility, and security concerns. The objective is to empower stakeholders with actionable insights derived from comprehensive analysis, thereby enhancing decision-making processes and driving sustainable growth in the Indian agriculture sector.

2.2 OBJECTIVES

By the conclusion of this activity, students will demonstrate mastery in the following areas:

- Comprehensive grasp of fundamental principles in data analytics, Power BI, and associated concepts, fostering proficiency in handling complex datasets and deriving actionable insights.

- Proficient execution of a data analytics case study focused on the Indian agriculture sector, showcasing adeptness in:
    - Identifying the top 5 crops by production, enabling informed decision-making regarding agricultural resource allocation.
    - Determining the top 3 states by production, facilitating targeted interventions for agricultural development.
    - Identifying the top 3 districts by production, enabling localized strategies for crop management and optimization.
    - Analyzing production patterns across various seasons, facilitating season-specific agricultural planning and management.
    - Evaluating year-wise production trends, enabling long-term strategic planning and trend forecasting.
    - Employing predictive analytics to forecast production for a future year, empowering proactive decision-making and risk mitigation strategies.
    - Assessing the feasibility of Minimum Support Price for a given crop in a specific year, aiding in policy formulation and farmer welfare initiatives.

3. MOTIVATION & RATIONAL STUDY

- Empowering Analytical Proficiency: The project's core motivation lies in fostering a comprehensive understanding of data analytics, Power BI, and related concepts among participants. By embarking on this journey, individuals delve into the intricacies of analyzing extensive datasets spanning 19 years of Indian agriculture data. This endeavor serves as a springboard for enhancing analytical proficiency, equipping participants with the skills to dissect complex data sets, derive meaningful insights, and make informed decisions.

- Driving Impactful Insights in Agriculture: Beyond mere skill acquisition, the project's rationale extends to driving tangible impacts within the agricultural domain. Through meticulous analysis, participants uncover critical insights regarding crop production trends, geographical variations, and seasonal fluctuations. By identifying the top 5 crops by production, the top 3 states and districts contributing to agricultural output, and delineating production trends across different seasons and years, participants are empowered to inform strategic agricultural interventions, optimize resource allocation, and bolster productivity across various regions.

- Fostering Predictive Capabilities and Policy Relevance: Furthermore, the project underscores the significance of predictive analytics in shaping future agricultural landscapes. By extrapolating insights gleaned from historical data, participants gain the foresight to predict production levels for future years, thereby facilitating proactive decision-making and resource planning. Additionally, the exploration of Minimum Support Price (MSP) for specific crops in given years adds a policy-relevant dimension to the study, enabling participants to evaluate the economic implications of agricultural policies and advocate for equitable support mechanisms for farmers.

4. SYSTEM ANALYSIS

➢ Existing System:

- **Limited Data Utilization:** The existing system lacks a comprehensive framework for harnessing the vast repository of Indian agriculture data spanning 19 years effectively.
- **Manual Analysis:** Data analysis processes are predominantly manual, time-consuming, and prone to errors, hindering the extraction of meaningful insights**.**
- **Fragmented Insights:** Analysis efforts are fragmented, lacking a cohesive approach to derive actionable insights from the data, resulting in suboptimal decision-making.
- **Reactive Decision-Making**: Decision-making in the agricultural sector is primarily reactive, lacking predictive capabilities to anticipate future trends and challenges**.**
- **Inadequate Policy Evaluation:** Evaluation of agricultural policies, such as Minimum Support Price (MSP), is limited, with insufficient data-driven insights to assess their efficacy and impact accurately.

➢ Proposed System:

- **Comprehensive Data Analysis:** The proposed system integrates advanced data analytics techniques and Power BI to analyze the extensive Indian agriculture dataset comprehensively.
- **Automated Analysis:** Leveraging automation, the proposed system streamlines data analysis processes, reducing manual effort and minimizing the likelihood of errors.
- **Cohesive Insights:** Through a structured approach, the proposed system consolidates analysis efforts, facilitating the derivation of cohesive insights into crop production, regional variations, and seasonal trends.
- **Proactive Decision-Making:** With predictive analytics capabilities, the proposed system enables proactive decision-making by forecasting production for future years, empowering stakeholders to anticipate and address challenges preemptively.
- **Enhanced Policy Evaluation**: By evaluating parameters such as MSP for specific crops in given years, the proposed system provides policymakers with data-driven insights to assess policy effectiveness accurately and make informed adjustments as necessary.

5. SOFTWARE REQUIREMENTS SPECIFICATIONS

5.1System Configurations

For each of the software solutions developed during the internship at Edunet Foundation, the following system configurations are required:

- **Software Requirements:**
  - **Operating System:** The software solutions are compatible with major operating systems such as Windows, macOS, and Linux.
  - **Programming Language:**
    - Python, R
  - **Technology (Tool):**
    - Power BI
  - **Database:**
    - Database used is csv file of Indian Agriculture from 2005 to 2023.

- **Hardware Requirements:**
  - **System:**
    - A system with a modern processor and adequate processing power is recommended for running the software solutions smoothly.
    - Compatibility with both desktop and server environments ensures flexibility in deployment.
  - **Hard Disk:**
    - Sufficient hard disk space is required to accommodate the software installation files and data storage.
  - **RAM:**
    - A minimum of 2GB RAM is recommended for optimal performance of the software solutions.

6. METHODOLOGICAL DETAILS

1. **Data Collection:**
    a. Identify and access relevant sources of Indian agriculture data, including government databases, agricultural research institutes, and industry reports.
    b. Gather data spanning 19 years to ensure a comprehensive understanding of long-term trends and patterns in crop production, regional variations, and policy impacts.
    c. Validate the integrity and reliability of the collected data to ensure its suitability for analysis.

2. **Data Preparation:**
    a. Cleanse the raw data by removing inconsistencies, outliers, and duplicate entries to ensure data integrity.
    b. Handle missing values through imputation techniques or exclusion, maintaining the accuracy and completeness of the dataset.
    c. Standardize data formats and units to facilitate uniform analysis across different variables and time periods**.**

3. **Exploratory Data Analysis (EDA):**
    a. Conduct descriptive statistical analysis to summarize the main characteristics of the dataset, including measures of central tendency, dispersion, and distribution.
    b. Visualize data using charts, graphs, and heatmaps to identify patterns, trends, and relationships among variables.
    c. Perform correlation analysis to explore the strength and direction of relationships between crop production, geographical factors, and other relevant variables.

4. **Advanced Analytics:**
    a. Utilize statistical methods such as regression analysis and time series modeling to analyze production trends and forecast future yields.
    b. Apply machine learning algorithms, such as decision trees or random forests, to identify factors influencing crop production and predict production outcomes.

c. Implement the proposed solution:
    i. Calculate the average production of the selected crop for the past three years.

    ii. Gather data on the production of the crop for the current year.
    iii. Compare the current year's production with the average production of the past three years.
    iv. Determine if the current year's production qualifies for Minimum Support Price based on the comparison results.

5. Power BI Implementation:

a. Develop interactive dashboards and reports using Power BI to visualize and communicate key insights derived from the analysis.
b. Incorporate the results of the proposed solution into the dashboards to provide stakeholders with real-time information on crop production eligibility for Minimum Support Price.

6. Interpretation and Validation:

a. Interpret the findings, including the results of the proposed solution, in the context of the agricultural sector's dynamics and stakeholders' needs.
b. Validate the results against domain knowledge, expert opinions, and existing literature to ensure accuracy and reliability.

7. Documentation and Reporting:

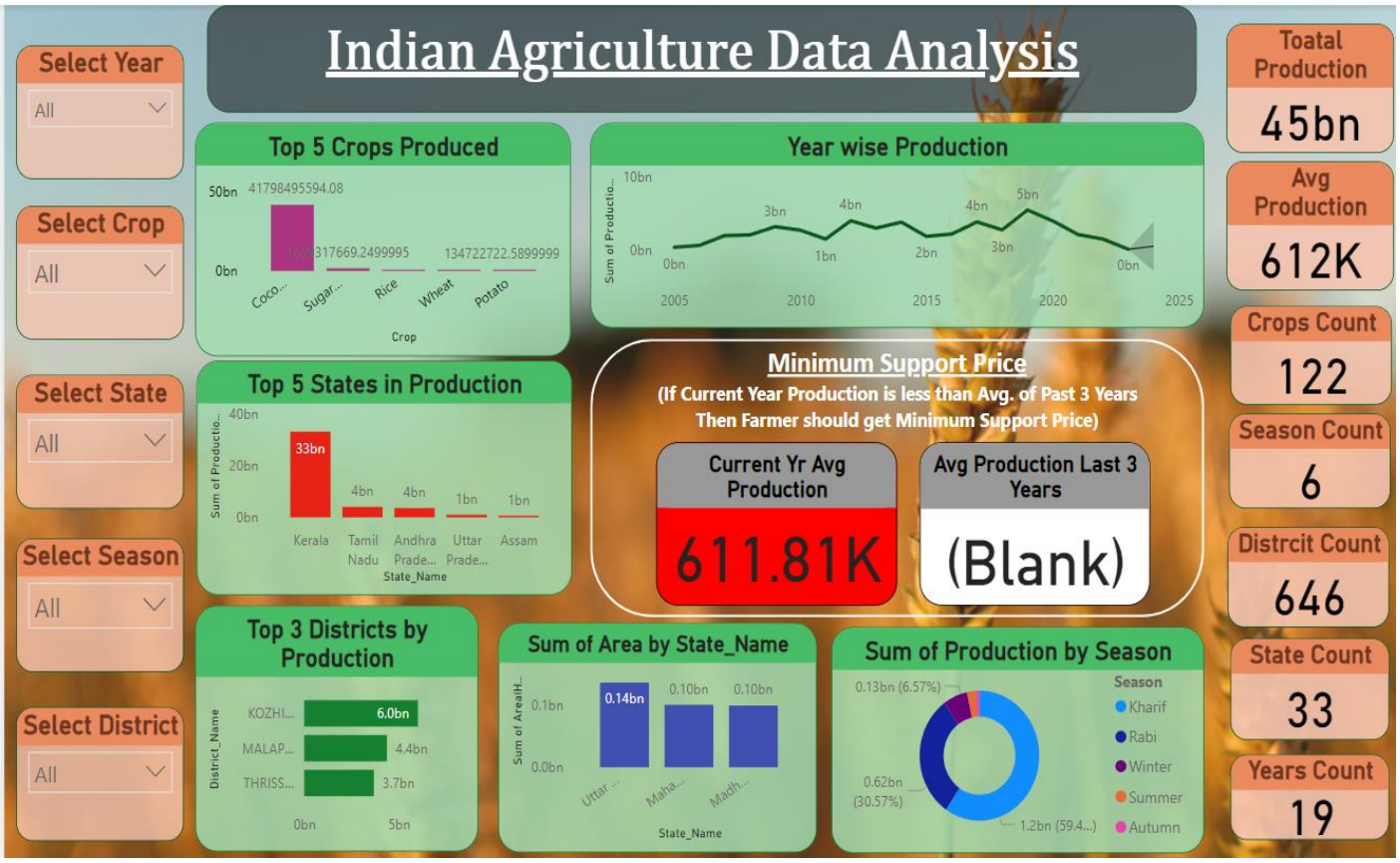
a. Prepare a comprehensive report documenting the entire analytical process, including data collection, preparation, analysis, and interpretation.
b. Summarize the research objectives, methodologies, key findings, and proposed solutions, emphasizing actionable insights and recommendations for stakeholders.

8. By integrating the proposed solution into the methodology, the project aims to provide stakeholders with enhanced insights into crop production dynamics and eligibility for Minimum Support Price, enabling more informed decision-making and policy formulation in the agricultural sector.

7. RESULT ANALYSIS

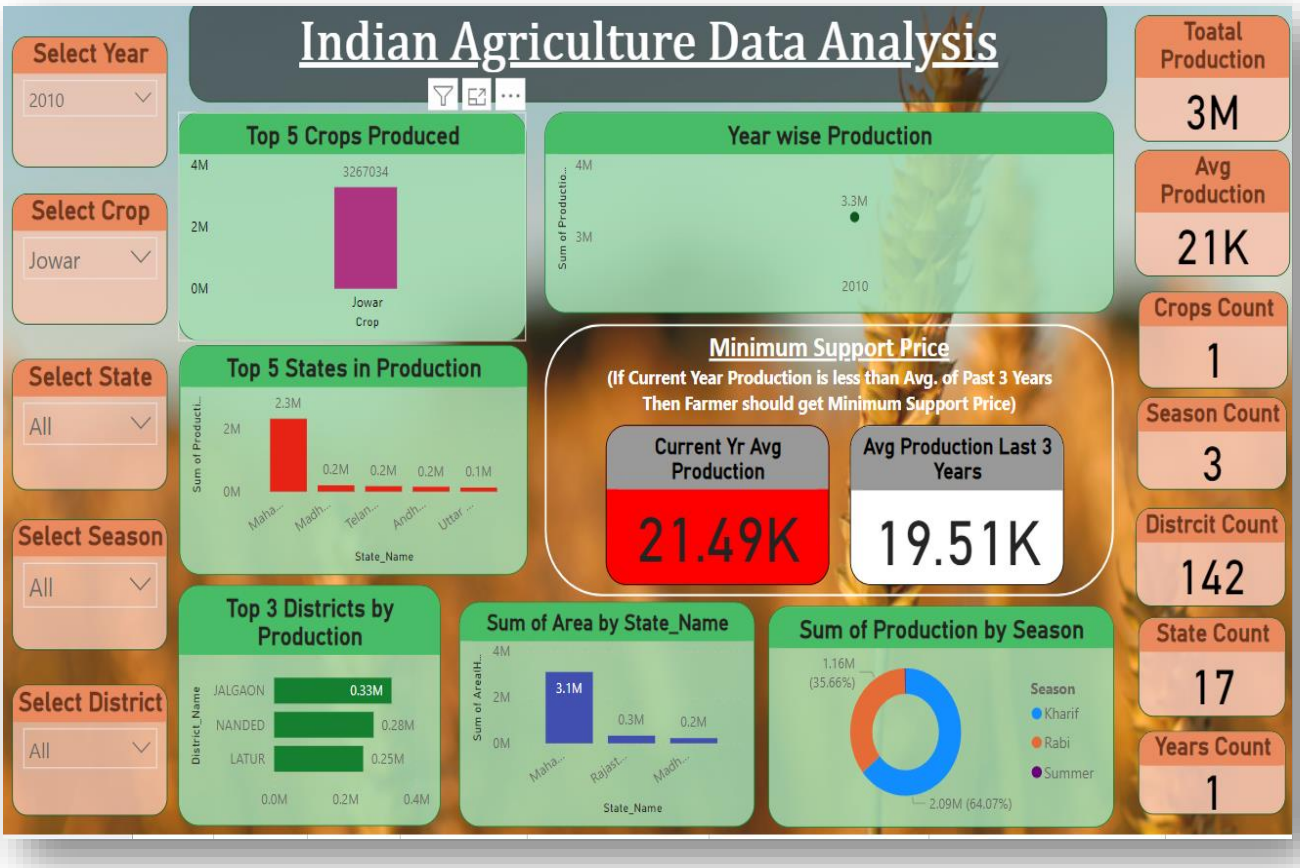
HOME PAGE OF DASHBOARD

- In the home page I have added the slicer or filter which have been created for filtering.
- Then the card has added to summarize the total production, avg production, distinct count of crops, districts and states, seasons.
- Visualization for top 5 crops by production, top 5 states by production, year wise production, minimum support price, production by season, top 3 districts by production, area wise top 3 states etc.



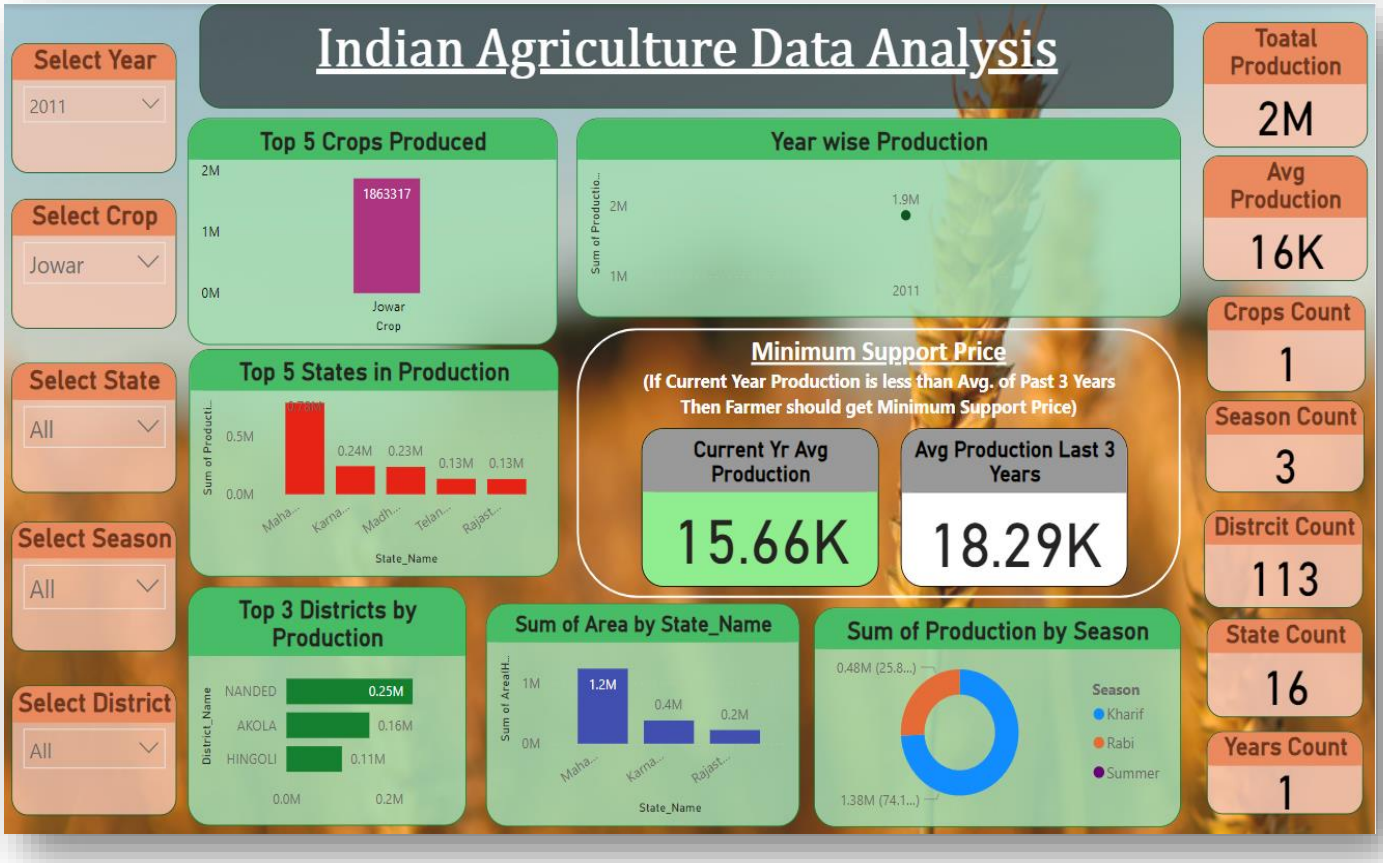
- **MSP**
  - Logic for giving MSP used is If Current Year Production is less than Avg. of Past 3 Years Then Farmer should get Minimum Support Price.
  - **Formula:**

    MSP = SWITCH(TRUE(),
    AVERAGE(apy_1[Production(Tons)])<=CALCULATE([AverageProductionLastThreeears]),
    "LightGreen",
    "Red"
)
  - As we can see in the year 2010 above picture the average of current year is not less than previous three years average so MSP will not be given so red color.



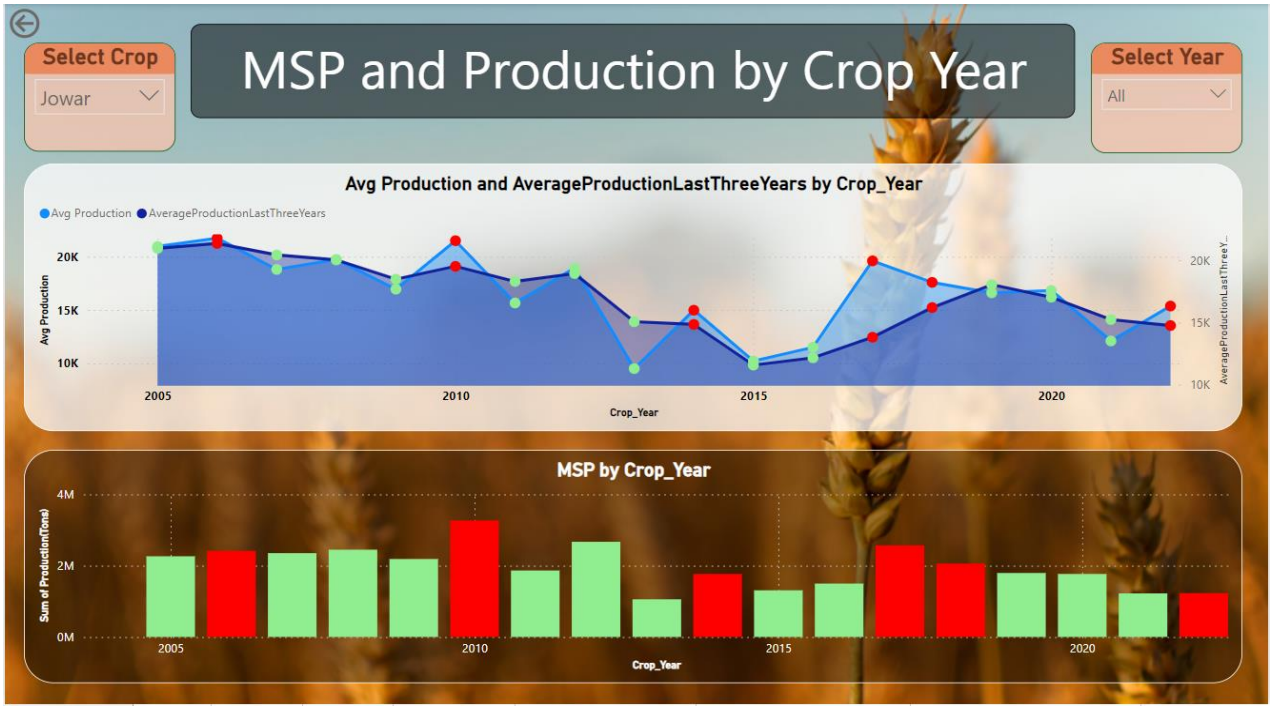
- But if we change the year in slicer/filter to 2011 then the average production of that year is less than the average of previous three years average so MSP will be given to farmer, therefore color changed to Green.

MSP CHECK OF JOWAR CROP



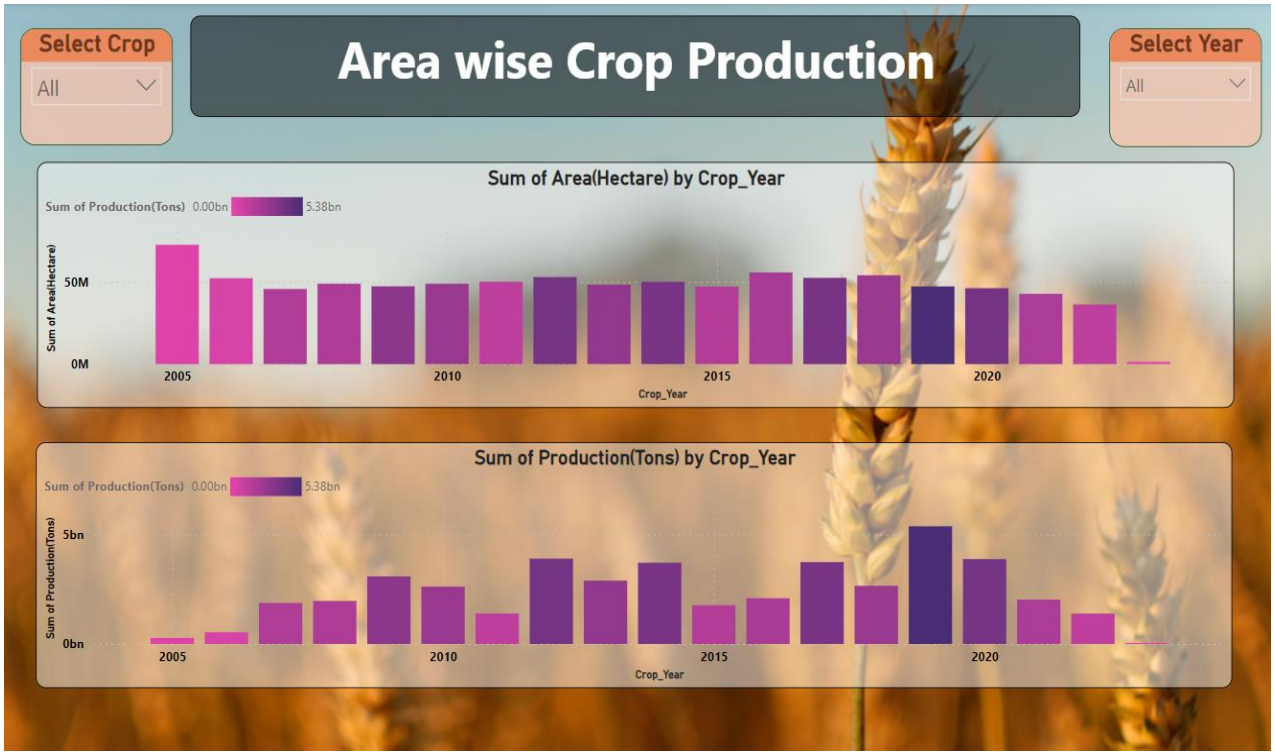
YEAR WISE TREND OF MSP FOR JOWAR CROP

Here we can see that in overall 19 years most of times Jowar crop had got MSP this means there is lesser production of jowar in recent year and farmers need to produce it more.



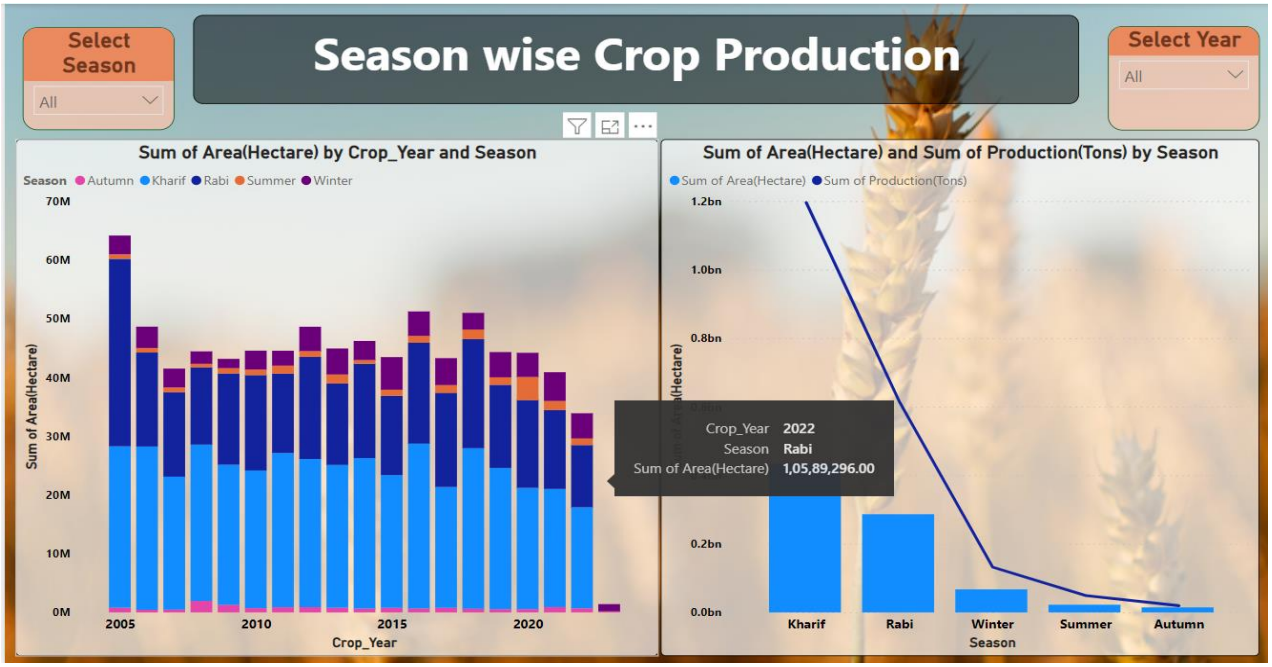
AREA WISE CROP PRODUCTION THROUGHOUT 19 YEARS

Production is highest in the year 2019 though area of production is less.



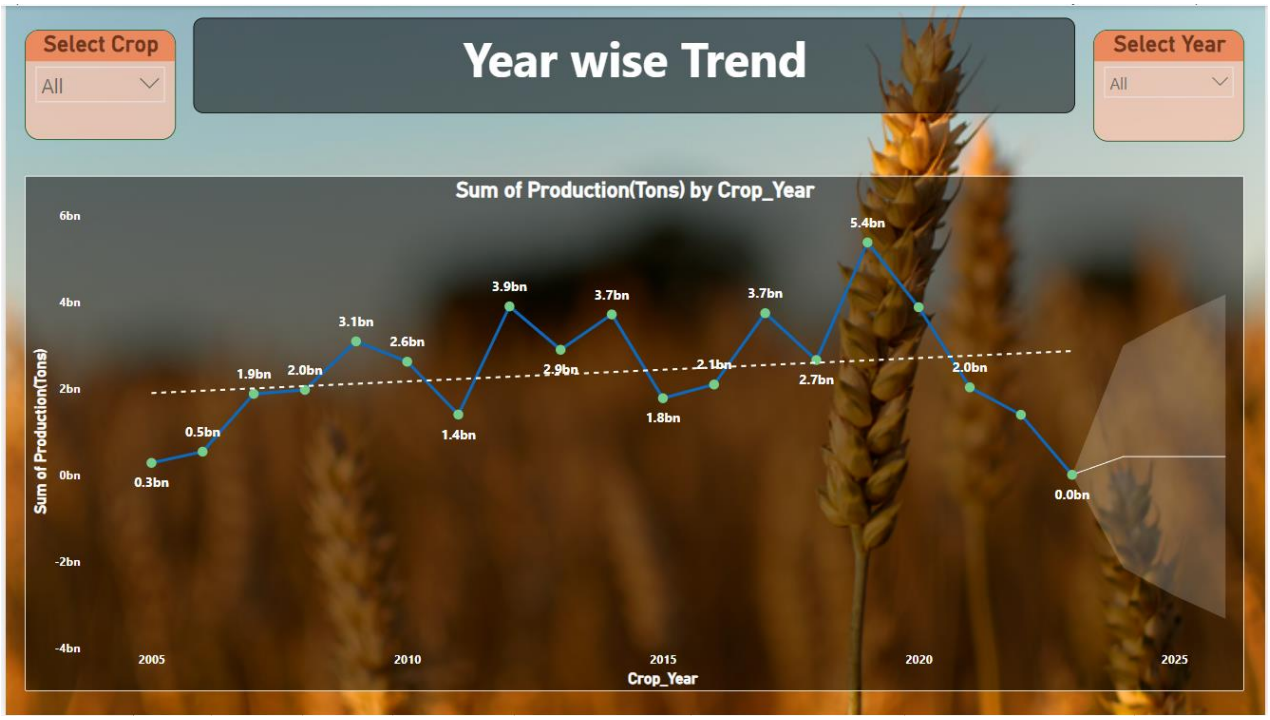
SEASON WISE CROP PRODUCTION OF 19 YEARS

Most of production happened in Kharif season followed by Rabi and Autumn season has lowest production throughout the year.



YEAR WISE TREND OF PRODUCTION AND PREDICTION OF NEXT 2 YEAR

In this graph we can see the trend line for production of 19 years and using all the 19 years data the prediction is done for the next 2 years with 95% accuracy.



8. CONCLUSION

Power BI's robust visualization capabilities have significantly contributed to the development of interactive dashboards and reports, particularly in the agricultural sector. By utilizing these tools, farmers can efficiently access and comprehend crucial information regarding crop Minimum Support Price (MSP) eligibility status. This streamlined approach not only simplifies complex data but also enhances decision-making processes for farmers, empowering them to make informed choices about their crops and resources.

Moreover, this case underscores the importance of leveraging technology to address the challenges faced in agriculture. By embracing innovative solutions like Power BI, the agricultural sector can overcome obstacles more effectively and improve overall efficiency. The success of this implementation highlights the potential for technology to revolutionize agricultural practices, from policy formulation to implementation.

Looking ahead, integrating such technologies into agricultural systems presents promising opportunities for further advancements. By continuing to harness the power of data visualization and analytics, stakeholders can drive meaningful progress in agricultural policy, productivity, and sustainability. Embracing these technologies is not just a step forward; it's a leap towards a more resilient and prosperous agricultural future.

REFERENCES

1. Power Query Editor
https://learn.microsoft.com/en-us/power-query/power-query-ui
https://learn.microsoft.com/en-us/power-query/

2. Power BI Desktop
https://learn.microsoft.com/en-us/power-bi/fundamentals/desktop-what-is-desktop
https://learn.microsoft.com/en-us/power-bi/fundamentals/
https://www.youtube.com/watch?v=-_DJPRrFQXI

3. Data Pre-Processing
https://www.youtube.com/watch?v=zv6RWIP9rpg
https://www.youtube.com/watch?v=9MMj7NsBM_U

4. Data Visualization
https://www.youtube.com/watch?v=_1w9w7tjSys

5. DAX
https://www.youtube.com/watch?v=Ar6hSITP-w4

6. Formatting
https://www.youtube.com/watch?v=giWb_rpTGT0&t=171s

7. Project Preparation
https://www.youtube.com/watch?v=9tF1IrfLflg

8. Saving the Project
https://www.youtube.com/watch?v=6IfrAyTBzYk