

Crop Production Analysis in India

Project Overview

This project analyses crop production in India, leveraging machine learning and data visualization techniques to gain insights and predict crop yields. The dataset provides extensive information about crop production trends across multiple years, seasons, and regions in India. The primary goal is to uncover key metrics influencing crop production and to use these insights for predictive modelling.

Key Objectives

1. **Explore Trends:** Analyse production patterns by year, season, and crop type.
 2. **Data Cleaning and Feature Engineering:** Handle missing values and derive new metrics, such as production per unit area.
 3. **Clustering Analysis:** Group districts with similar crop production patterns.
 4. **Predictive Modelling:** Build a machine learning model to estimate crop yields.
-

Technologies Used

- **Programming:** Python
 - **Libraries:** Pandas, NumPy, Matplotlib, Seaborn, scikit-learn
 - **Visualization Tools:** Tableau, Power BI
 - **Version Control:** Git
-

Dataset

The dataset contains information about:

- States and districts in India.
- Crop types and production volumes.
- Seasons (Kharif, Rabi, etc.).
- Cultivated area.

Dataset Source: [Download Here](#)

Project Structure

1. **crop_production.py**: Main Python script for data preprocessing, exploratory data analysis, clustering, and predictive modelling.
 2. **Clustering_Results.csv**: Output file containing district clusters.
 3. **Cleaned_Crop_Production_Data.csv**: Pre-processed dataset.
-

Methodology

Data Analysis

- **Trends Analysis**: Yearly production trends, seasonal variations, and crop-wise breakdowns.
- **Visualization**: Interactive dashboards using Tableau and static visualizations with Python.

Data Preprocessing

- Missing values in production were filled using median values grouped by crop and season.
- Derived the metric "Production per Unit Area" for better insights.

Machine Learning

- **Clustering**: Used K-Means to classify districts based on crop production patterns.
 - **Predictive Modelling**: Built a regression model for yield prediction (optional).
-

How to Use

1. Clone this repository:

`git clone <repository-url>`

2. Run the Python script:

`python crop_production.py`

3. Explore the cleaned data and clustering results saved in CSV files.
 4. Use Tableau/Power BI to create interactive dashboards.
-

Evaluation Metrics

- **Clustering Performance**: Analysed cluster characteristics to ensure meaningful groupings.

- **Data Insights:** Focused on actionable insights like identifying high-yield crops or underperforming regions.
-

Future Scope

- Add more advanced models for yield prediction.
 - Integrate weather and soil data for improved insights.
 - Deploy the solution as a web app for easy access.
-

Contributions

Contributions and suggestions are welcome! Please feel free to open issues or submit pull requests.

Acknowledgments

- The dataset and problem statement were inspired by [Project 7: Crop Production Analysis in India](#).