**DATA SCIENCE TOOLBOX: PYTHON PROGRAMMING**

**PROJECT REPORT**

(Project Semester: January- April 2025)

**LANDHOLDING PATTERNS IN INDIAN AGRICULTURE**

Submitted by

Tapan Kumar Mahato

Registration No.12315468

Programme and Section K23GD

Course Code INT 375

Under the Guidance of

**Baljinder Kaur (UID:28968)**

Assistant Professor

**Discipline of CSE/IT**

**Lovely School of Computer Science**

**Lovely Professional University, Phagwara**

**CERTIFICATE**

This is to certify that Tapan Kumar Mahato, bearing Registration no. 12315468 has completed the INT375 project titled **“LANDHOLDING PATTERNS IN INDIAN AGRICULTURE”** under my guidance and supervision. To the best of my knowledge, the present work is the result of his/her original development, effort, and study.

**Baljinder Kaur (UID:28968)**

Assistant Professor

**School of Computer Science and Engineering**

Lovely Professional University

Phagwara, Punjab.

Date: 13-04-2025

**DECLARATION**

I, Tapan Kumar Mahato, student of BTech under CSE Discipline at Lovely Professional University, Punjab, hereby declare that all the information furnished in this project report is based on my own intensive work and is genuine.

Date: 13-04-2025 Signature: **Tapan Kumar Mahato**

Registration No. 12315468

**ACKNOWLEDGEMENT**

I would like to express my heartfelt gratitude to my guide, **Baljinder Kaur mam**, for their invaluable guidance, support, and encouragement throughout the completion of this project. His expertise and constructive feedback have greatly contributed to the success of this work.

I would also like to extend my sincere thanks to Lovely Professional University for providing me such a wonderful opportunity to work on this project in the subject DATA SCIENCE TOOLBOX: PYTHON PROGRAMMING with subject code INT 375, helping with the necessary resources and skills that laid the foundation for my research.

This project, titled “**Landholding pattens in Indian Agriculture**”, has been a learning experience, and I would like to acknowledge the support of my peers, family, and all others who helped me in any manner.

Thank you all for your continuous support and motivation.

**Name: Tapan Kumar Mahato  
Section:** K23GD

**Roll No. :** 02

**Reg. No. :** 12315468  
**Date:** 13-04-2025

**CONTENTS**

|  |  |  |
| --- | --- | --- |
| **S No.** | **Title** | **Page No.** |
| **1.)** | Introduction | **6** |
| **2.)** | Source of dataset | **7** |
| **3.)** | Dataset Preprocessing | **8-9** |
| **4.)** | Analysis on dataset (for each objective)   1. General Description 2. Specific Requirements 3. Analysis results 4. Visualization | **10-18** |
| **5.)** | Conclusion | **19** |
| **6.)** | Future Scope | **20** |
| **7.)** | References | **21** |

**INTRODUCTION**

Agriculture in India plays a vital role in the country’s economy and employs a significant portion of the population. The structure of agricultural landholding reveals valuable insights into ownership patterns, tenancy, and distribution across social groups and land size classes. This project focuses on analyzing the landholding patterns across Indian states using the dataset from NDAP (National Data & Analytics Platform).

With the help of Python and libraries such as Pandas, Seaborn, Matplotlib, and Squarify, the project performs data cleaning, transformation, and exploratory data analysis to derive insights from the 2015-2016 dataset on operational holdings. The findings provide a foundation for understanding agricultural inequalities and regional variations that can help inform policy-making and land reforms.

**Source of Dataset**

The dataset used in this project is sourced from the National Data & Analytics Platform (NDAP) managed by the Government of India. The file, named NDAP\_REPORT\_7171.csv, provides sub-district-level information on operational agricultural holdings for the year 2015-2016.

The dataset includes:

* Types of holdings (owned, leased, mixed)
* Area operated and number of holdings by category
* Social group classification (SC, ST, Others, Institutional)
* Land size classifications

The dataset is in CSV format, which was well-suited for analysis using Python libraries.

Dataset Link - <https://ndap.niti.gov.in/dataset/7171?tab=profile>

**DATASET PREPROCESSING**

**1. Data Loading**

The dataset named **NDAP\_REPORT\_7171.csv** was loaded using the **Pandas** library. During this step, missing values such as "NA", "NaN", "-", and blank entries were treated as null. The initial structure of the dataset was examined using basic Pandas commands to understand its shape and sample contents.

**2. Column Name Cleaning**

To ensure consistency and simplify column referencing, all column names were standardized:

* Leading and trailing whitespaces were removed.
* Column names were converted to lowercase.
* Special characters and spaces were replaced with underscores using regular expressions.

**3. Handling Missing and Duplicate Data**

Rows containing null values in critical identification columns such as state and district were removed to maintain the accuracy of region-wise analysis. Additionally, duplicate entries in the dataset were identified and dropped.

**4. Numeric Conversion**

Two key groups of columns—**holdings columns** and **area columns**—were identified for numerical analysis. The script checked if each column existed in the dataset before applying conversion. All valid columns in these groups were converted to numeric format using pd.to\_numeric() with coercion for non-numeric values.

**5. Feature Engineering**

New computed columns were created for deeper insights:

* total\_operational\_holdings: Total number of all types of operational holdings.
* total\_operated\_area: Combined operated area from all available categories.
* avg\_area\_per\_holding: Ratio of total operated area to the total number of holdings.

**6. Sample Verification**

A preview of the cleaned and processed dataset was printed, highlighting key fields such as state, district, land\_area\_size, social\_group\_type, and the newly created summary columns.

**Detailed Analysis Based on Project Objectives**

**Objective 1:** Analyze the distribution of operational holdings by land size across different states and districts.

* Visualization: Box Plot
* Insight: Small and marginal landholdings dominate across many regions.

**Objective 2:** Compare the total area and number of holdings based on tenancy status.

* Visualization: Pie Chart & Bar Chart
* Insight: Wholly owned holdings dominate, but mixed types also hold substantial area.

**Objective 3:** Study the relationship between social group type and average operational holdings.

* Visualization: Bar Chart by social group
* Insight: Institutional groups manage larger areas; SC/ST groups hold smaller plots.

**Objective 4:** Identify regions with the highest agricultural land fragmentation.

* Visualization: Line Plot & Distribution Plots
* Insight: Higher fragmentation is seen in states like Bihar and West Bengal.

**Objective 5:** Visualize and map agricultural land use trends across states and subdistricts.

* Visualization: Heatmap
* Insight: States like Punjab and Rajasthan have higher average operated areas.

**Objective 6:** Examine state-wise variation in the number of operational holdings.

* Visualization: Grouped bar or choropleth (optional)
* Insight: States with smaller average holdings have more fragmented plots.

**Objective 7:** Analyze average landholding size across tenancy categories.

* Visualization: Grouped bar chart
* Insight: Mixed tenancy often leads to larger average holding areas.

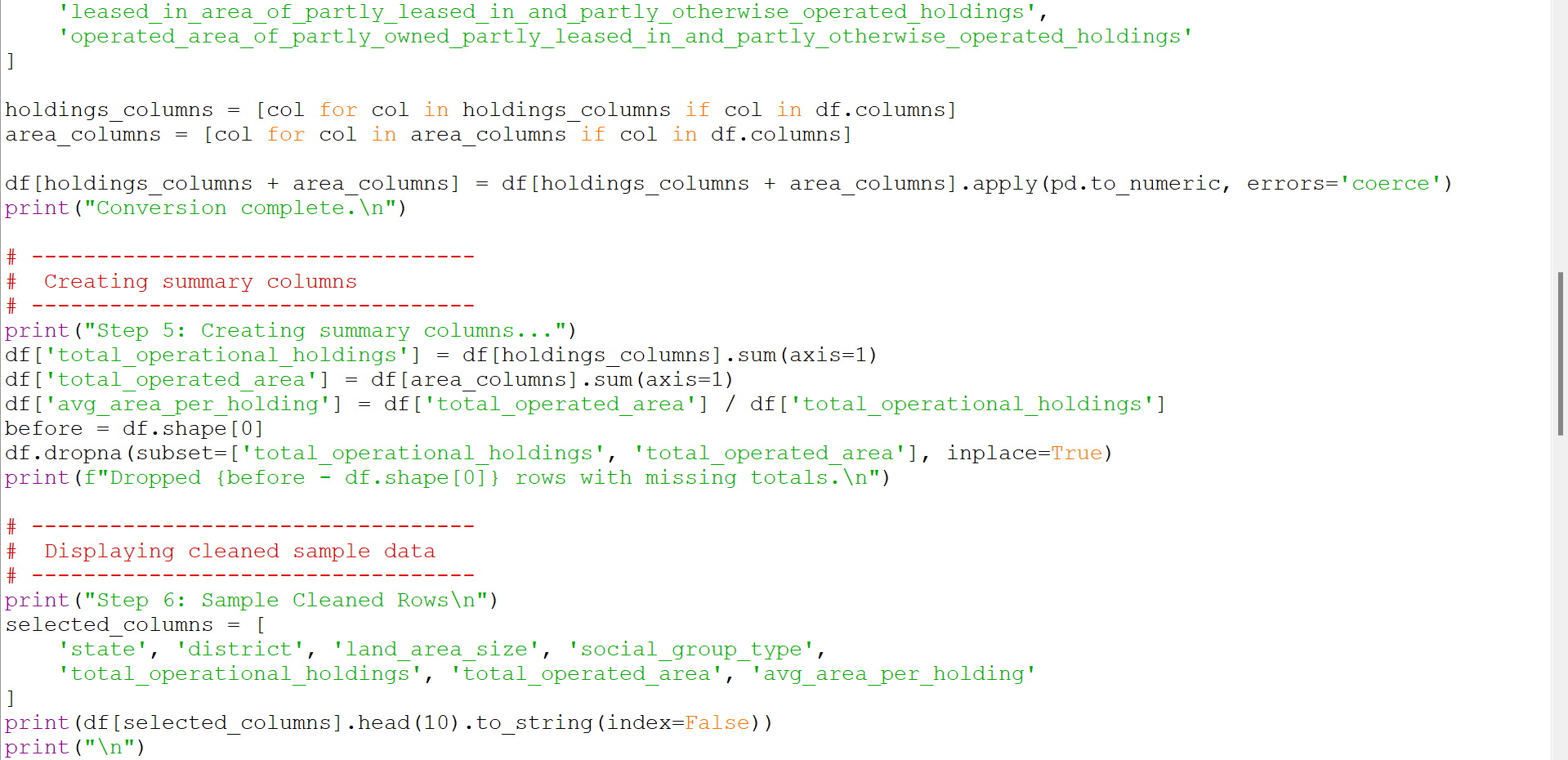
**Objective 8:** Identify trends in landholding sizes among different social groups.

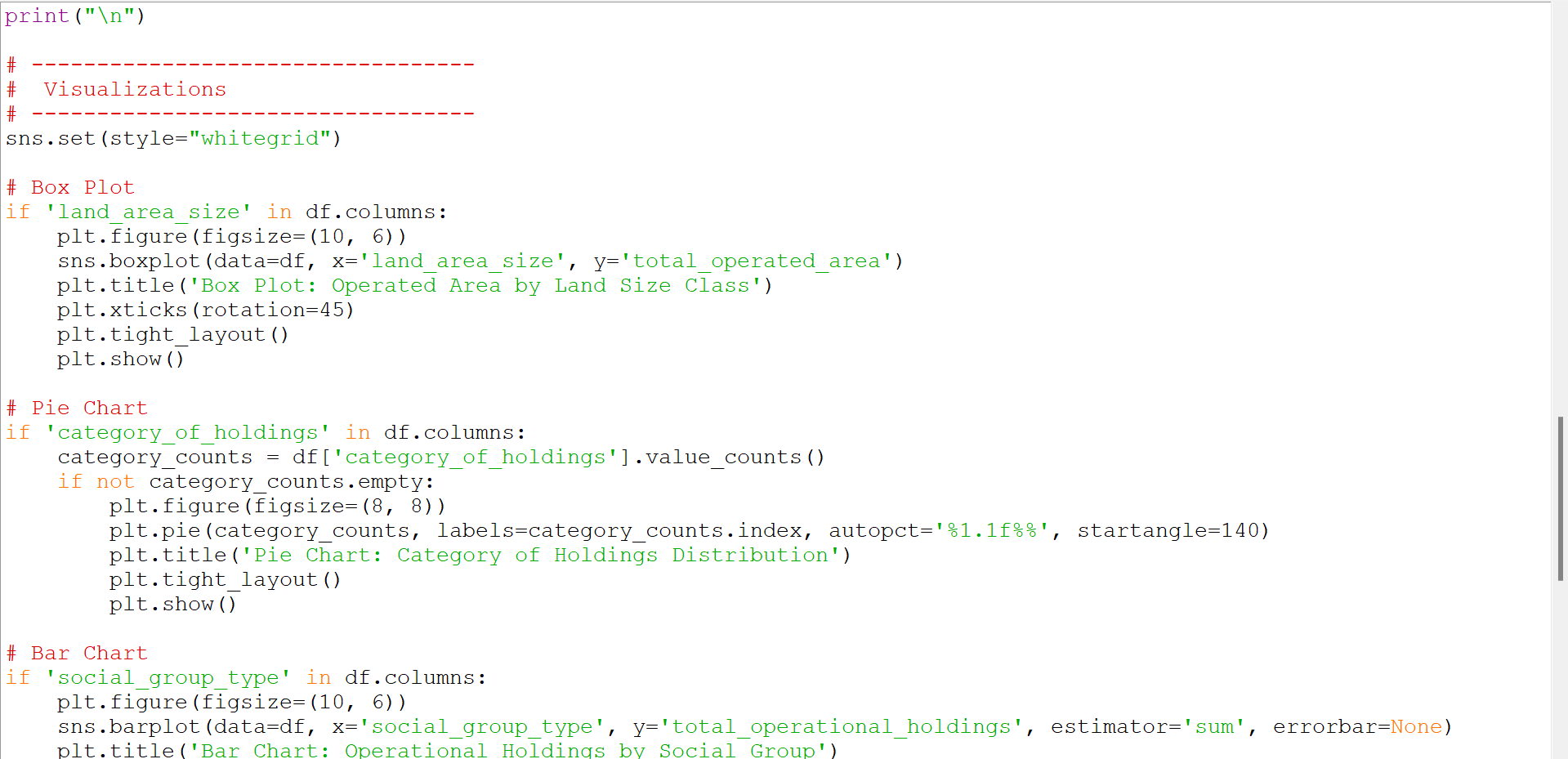
* Visualization: Treemap or bar chart
* Insight: Inequities exist, particularly for disadvantaged groups.

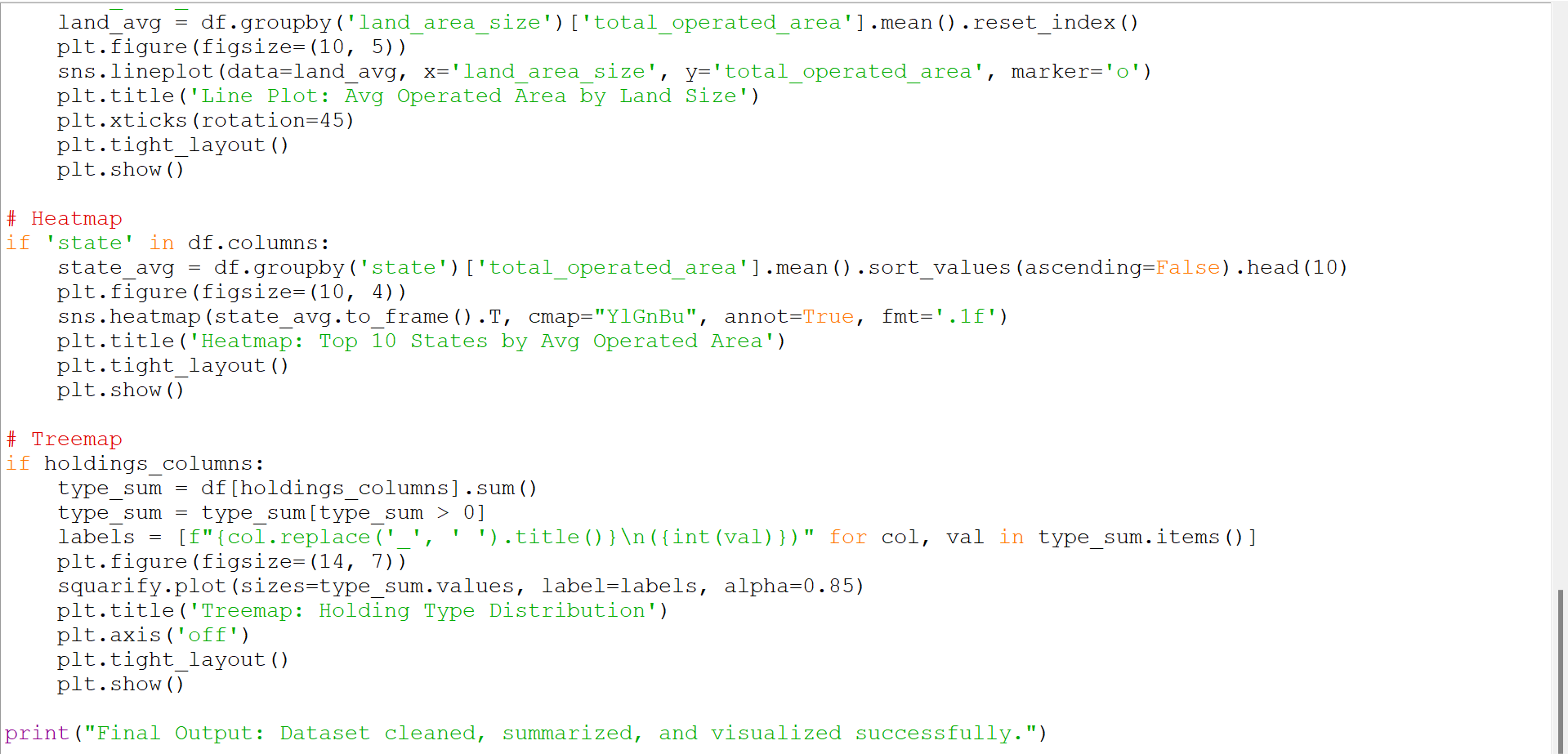
**Code**

****

****

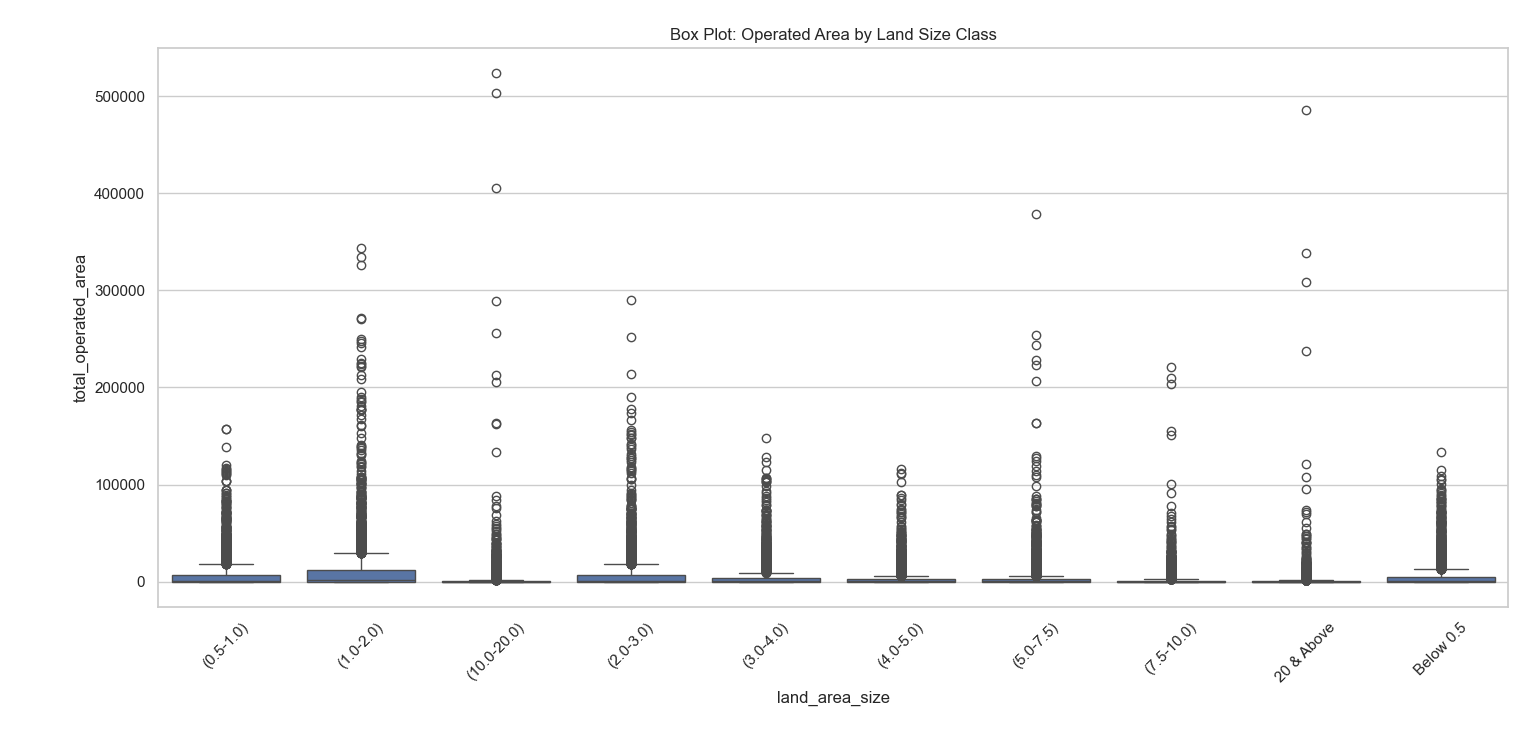
****

****

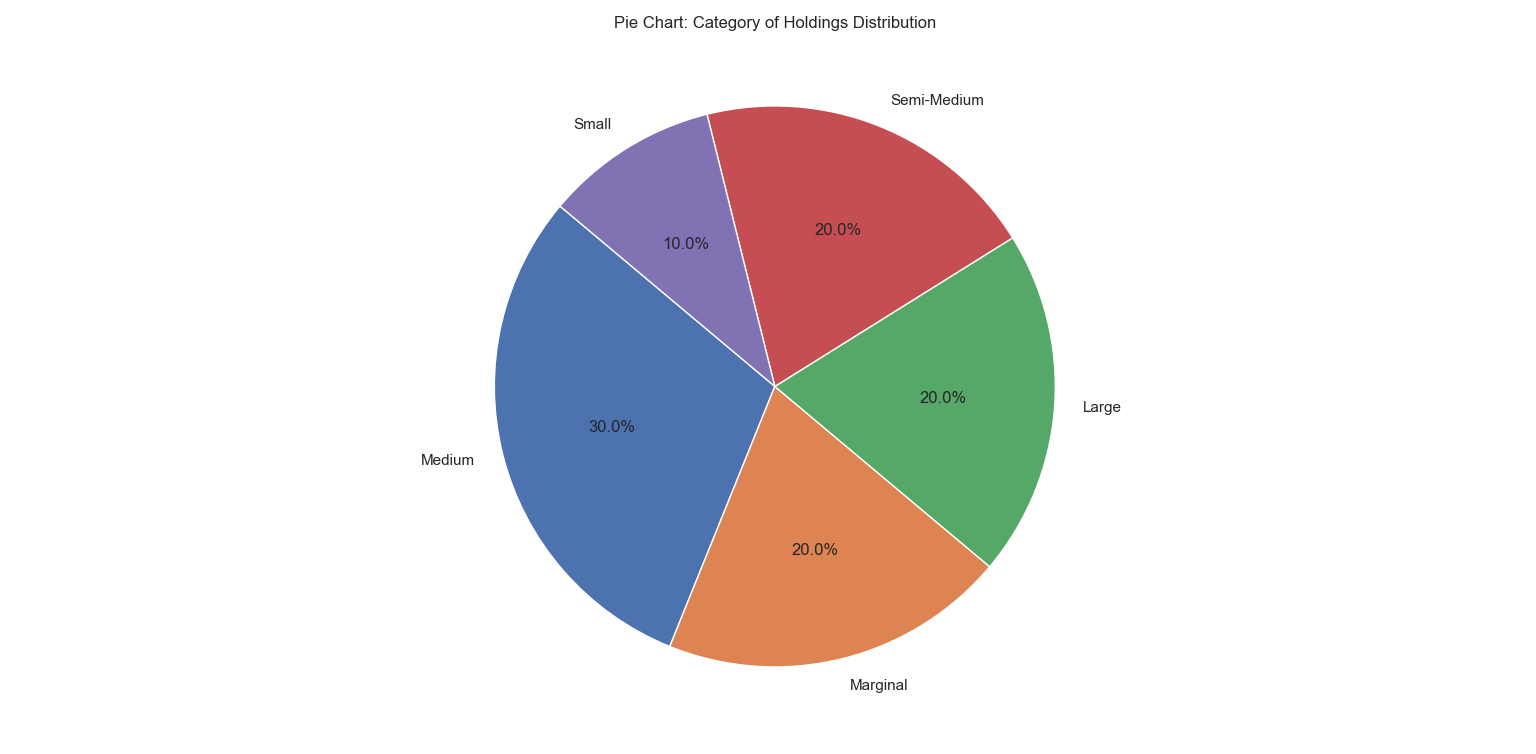
****

**Charts**

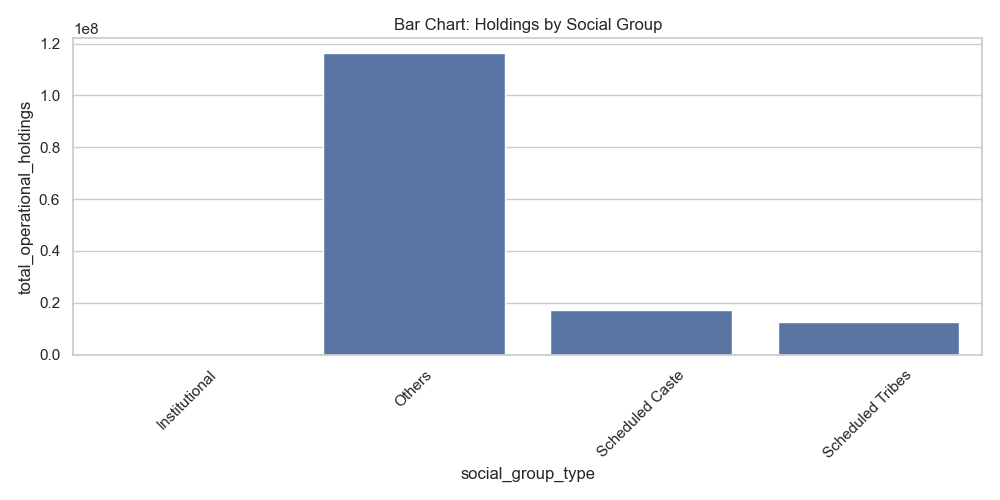
* 1. **Box Plot**

****

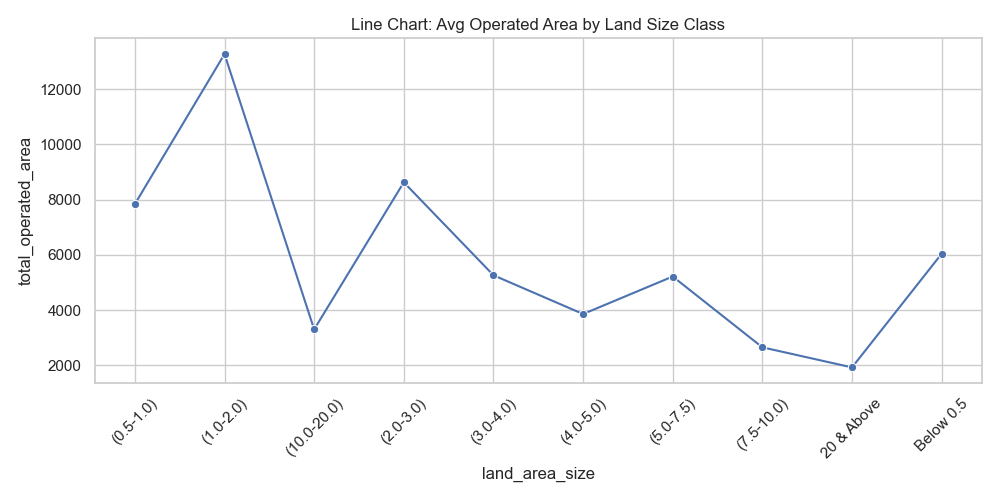
* 1. **Pie Chart**

****

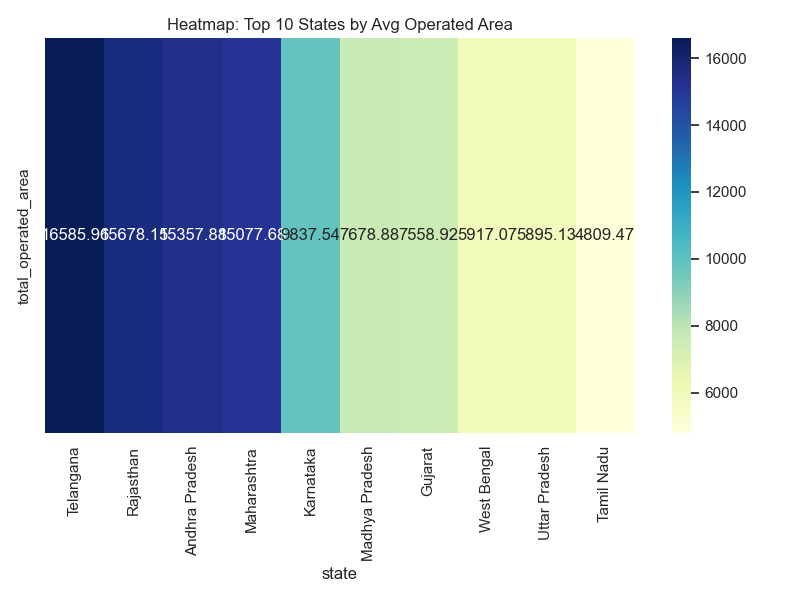
* 1. **Bar Chart**

****

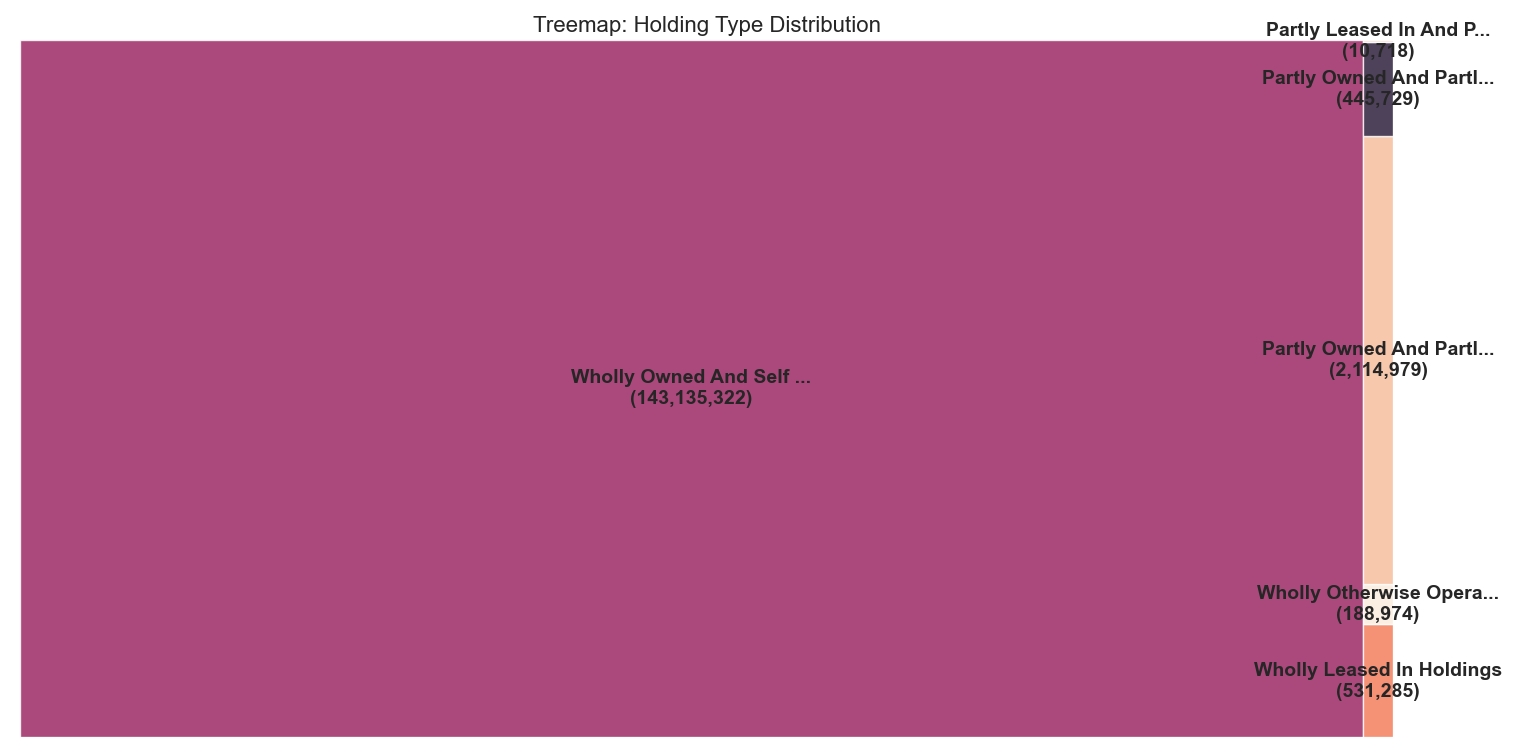
* 1. **Line Chart**

****

* 1. **Heatmap**

****

* 1. **Treemap**

****

**CONCLUSION**

This project successfully explored landholding patterns in Indian agriculture using data-driven techniques. Through comprehensive preprocessing—including column name normalization, handling of missing values, and numeric conversions—the dataset was cleaned and prepared for analysis.

Using Python libraries such as Pandas, Seaborn, Matplotlib, and Squarify, we generated multiple insightful visualizations, including box plots, pie charts, bar charts, line plots, heatmaps, and treemaps. These visualizations revealed key patterns in landholding based on tenancy status, social group, and land size categories.

The analysis uncovered critical disparities in land distribution and highlighted regions with significant land fragmentation. Notably, institutional groups operate larger land areas, while marginalized communities like SC/ST tend to manage smaller holdings. The project also emphasized regional variations, with states like Punjab and Rajasthan showing higher average operated areas.

By leveraging open government data and Python's data science toolbox, this work provides a reproducible and scalable pipeline that can support future research, policy analysis, and targeted land reform efforts.

**FUTURE SCOPE**

* **Integration of GIS-Based Maps:**  
  Incorporating GIS-based spatial visualizations can enhance regional insights by allowing district-wise mapping of operational holdings, land fragmentation, and tenancy patterns. This would make the analysis more geographically intuitive for policymakers.
* **Time-Series Analysis:**  
  Extending the current analysis to include time-series datasets across multiple years would help identify historical trends in landholding patterns, including ownership changes and agricultural shifts.
* **Predictive Modeling:**  
  Machine learning models can be developed to forecast land fragmentation trends and predict changes in ownership structures based on demographic, geographic, and economic factors.
* **Interactive Dashboards:**  
  Creating user-friendly, interactive dashboards using frameworks such as **Plotly Dash** or **Streamlit** can make the insights accessible to stakeholders, allowing dynamic filtering, visualization, and exploration of the dataset.

.

**REFERENCES**

* NDAP India – National Data & Analytics Platform (<https://ndap.niti.gov.in/>)
* Lovely Professional University Guidelines for INT 375
* Python official documentation (<https://docs.python.org>)
* Seaborn & Matplotlib visualization documentation

\*Important Links

LinkedIn - <https://www.linkedin.com/feed/update/urn:li:activity:7317145053582266368/>

GitHub - <https://github.com/Tapan-Kumar-Mahato/Landholding-Patterns-in-Indian-Agriculture>