# DATA SCIENCE TOOLBOX: PYTHON PROGRAMMING

PROJECT REPORT

(Project Semester January–April 2025)

**6th Minor Irrigation Census-Surface Lift- Rajasthan Using Python**

Submitted by

Devesh Kumar

Registration No. 12315695

Programme and Section: B.Tech in Computer Science & Engineering, Section K23EP

Course Code: INT375

Under the Guidance of

**Dr. Tanima Thakur**

**U.Id 23532**

**Designation: Assistant Professor**

**Discipline of CSE/IT**

**Lovely School of Computer Science & Engineering**

**Lovely Professional University, Phagwara**

# Declaration

I, Devesh Kumar, student of B.Tech in Computer Science & Engineering, Section K23EP, with Registration No. 12315695, hereby declare that this project titled "6th Minor Irrigation Census-Surface Lift-Rajasthan Using Python" is the result of my own work. The project has been carried out under the guidance of Dr. Tanima Thakur, Assistant Professor in the Discipline of Computer Science & Engineering, Lovely Professional University, Phagwara.

The data used in this project is sourced from the 6th Minor Irrigation Census-Surface Lift-Rajasthan dataset, and I have adhered to academic integrity guidelines throughout the project. I confirm that the report has not been submitted previously for any other course or program and is a true representation of my understanding and work.  
  
Date: April 12, 2025  
Place: Phagwara

**CERTIFICATE**

This is to certify that **Devesh Kumar** bearing Registration no. **12315695** has completed the Data Science Toolbox: Python Programming project titled, “**6th Minor Irrigation Census-Surface Lift-Rajasthan using Python**” under my guidance and supervision. To the best of my knowledge, the present work is the result of his original development, effort and study.  
 **Name of the Supervisor**  
Dr. Tanima Thakur  
 **Designation of the Supervisor**  
Assistant Professor **School of Computer Science & Engineering**  
Lovely Professional University  
  
Phagwara, Punjab.  
  
Date: April 12, 2025

# Acknowledgement

I would like to express my sincere gratitude to all those who have supported me throughout the completion of this project.  
  
Firstly, I would like to thank **Dr. Tanima Thakur**, Assistant Professor, Discipline of Computer Science & Engineering, Lovely Professional University, Phagwara, for her valuable guidance and supervision during this project. Her insights and encouragement helped me immensely in achieving the objectives.  
  
I also appreciate the support of my professors and classmates who contributed to my learning and motivation.

Special thanks to my family for their unwavering support, patience, and encouragement, which motivated me to complete this project successfully.  
  
Lastly, I acknowledge the importance of the dataset used in this project, which enabled meaningful analysis and learning.

**Table of Contents**

**1. Introduction ......................................................... 1**

**2. Source of Dataset ................................................. 2**

**3. EDA Process ............................................................ 3**

**4. Analysis on Dataset .............................................. 4**

**4.1 Top 10 Villages By Number of Schemes**

**4.2 Schemes By Owner Type**

**4.3 Average Pump Operating Days By Season**

**4.4 Correlation Between Scheme Costs, Usage & Horsepower**

**4.5 Villages With Both On-Stream & Off-Streams**

**5. Conclusion .............................................................. 5**

**6. Future Scope .......................................................... 6**

**7. References .............................................................. 7**

**1. Introduction**

Government-led infrastructure projects, particularly in rural and agricultural regions, play a crucial role in ensuring sustainable water access, improving irrigation, and supporting local economies. By analyzing data related to these schemes, policymakers and planners can uncover trends, evaluate performance, and optimize resource allocation.

This project focuses on analyzing a dataset of **irrigation and water schemes**, covering details like village-wise scheme distribution, ownership types, seasonal pump operations, and cost components. The dataset serves as a valuable source for understanding how water resources are managed and deployed across different regions.

Using **Python** and standard data science libraries like **Pandas**, **Matplotlib**, and **Seaborn**, this project applies **Exploratory Data Analysis (EDA)** to extract insights and visualize key patterns. The key objectives of the analysis are:

* To identify villages with the highest number of water schemes
* To assess ownership patterns in scheme management
* To evaluate average pump operating days across different agricultural seasons
* To study cost correlations with horsepower and usage
* To analyze the distribution of villages having both surface flow and lift irrigation schemes

The insights drawn from this analysis can help guide improvements in infrastructure planning, maintenance strategies, and equitable distribution of water resources.

# 2. Source of Dataset

The dataset used for this project is an Excel file containing comprehensive records of government-initiated water and irrigation schemes across various villages. It includes information such as village names, scheme ownership types, construction and maintenance costs, horsepower of lifting devices, seasonal pump operating days, and scheme classifications like surface flow and surface lift.

The data was imported into the Python environment using the **Pandas** library for efficient handling and manipulation. Preliminary data cleaning steps were applied to handle missing values and ensure consistency in column names. Essential Python libraries such as **Matplotlib** and **Seaborn** were utilized to create informative visualizations that support meaningful insights throughout the analysis.

It is taken from data.gov.in : **https://www.data.gov.in/resource/6th-minor-irrigation-census-surface-lift-rajasthan**

# 3. EDA Process

Exploratory Data Analysis (EDA) is a crucial step in understanding the underlying structure, completeness, and characteristics of the dataset. In this project, EDA was performed to explore trends in scheme distribution, seasonal operations, and cost correlations across villages. The following steps were undertaken:

1. **Data Loading**: The dataset was imported from an Excel file using the Pandas library, enabling efficient manipulation and processing of the records.
2. **Basic Inspection**: The structure of the dataset was examined, including the number of rows and columns, data types of each attribute, and the presence of missing or null values.
3. **Data Cleaning**: Column names were standardized by removing extra spaces. Missing or incomplete entries, particularly in key fields like seasonal pump operating days and scheme classifications, were addressed by either filtering or imputation.
4. **Feature Selection**: Relevant columns such as village\_name, ref\_scheme\_owner\_name, pump\_operating\_days\_\*, scheme\_construction\_cost, scheme\_machinery\_cost, ipu\_total, ipc\_total, and horse\_power\_of\_lifting\_device were selected for focused analysis.
5. **Visualization Setup**: The Matplotlib and Seaborn libraries were configured to produce insightful and aesthetically pleasing visualizations, aiding in the interpretation of patterns and relationships within the data.

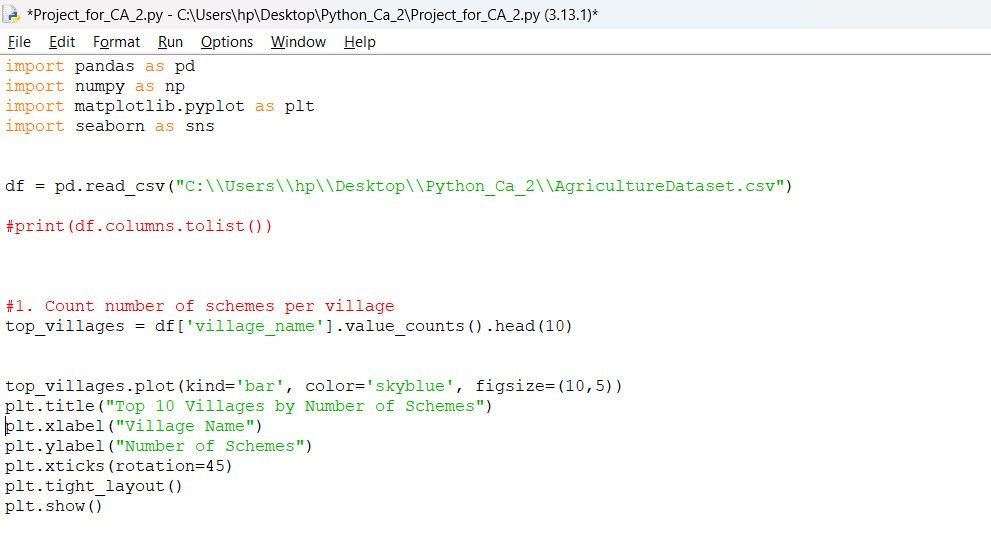
**4. Analysis on Dataset**

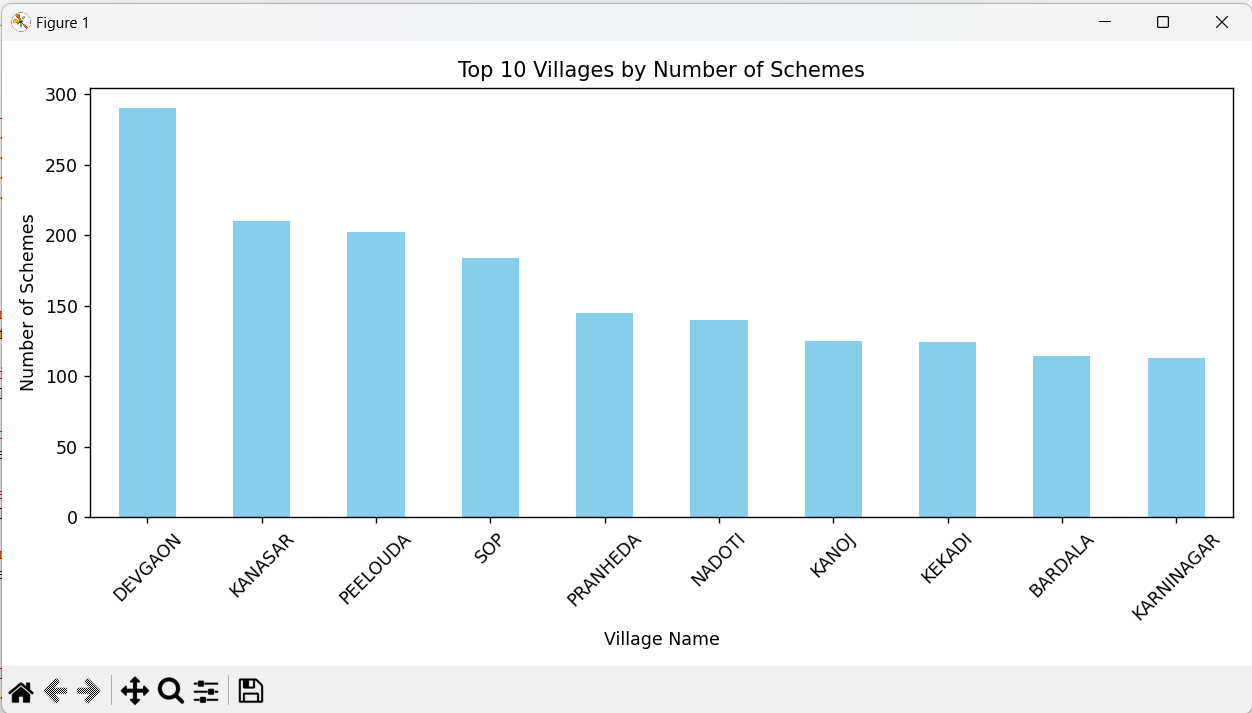
## 4.1 Top 10 Villages By Number of Schemes

This analysis identifies the villages with the highest number of implemented schemes. By counting how many schemes are associated with each village, we can determine which areas have received the most attention in terms of water resource development.

The data was grouped by village name and sorted to find the top 10 villages with the highest number of schemes. A bar chart was used to visually represent the results for easier comparison and interpretation.

**Insights**: Villages with the highest number of schemes may indicate either higher water infrastructure demand or a strategic focus on development in those areas.



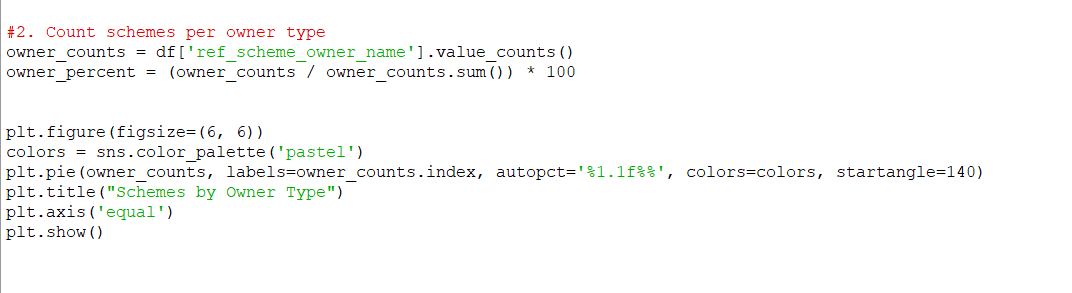


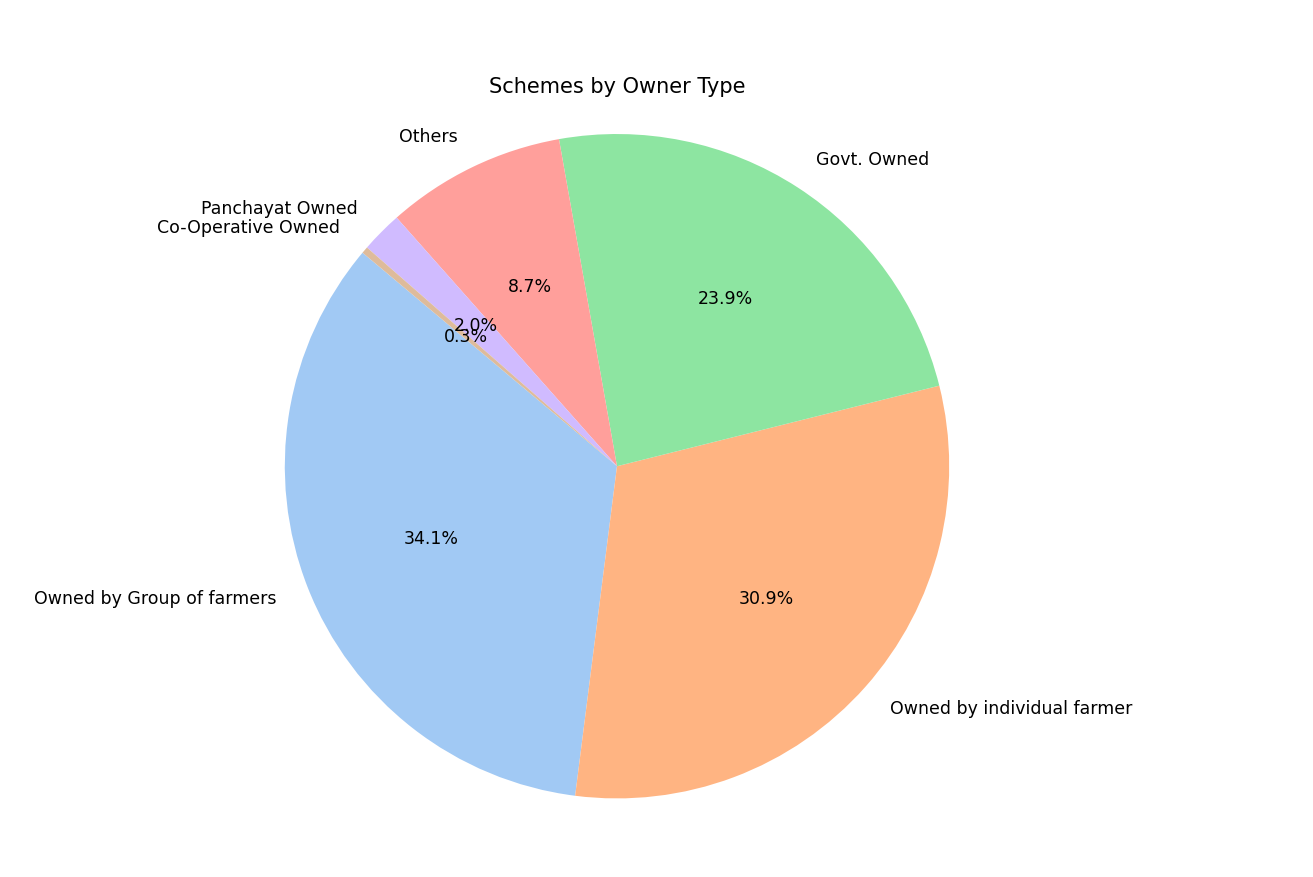
## 4.2 Schemes By Owner Type

This analysis explores the distribution of water schemes based on ownership type. It categorizes the schemes by their associated owners using the ref\_scheme\_owner\_name attribute and visualizes the proportions through a pie chart.

Understanding which organizations or entities own the majority of the schemes provides insight into the role of government bodies, private entities, or community-led initiatives in water resource management.

**Insights**: Certain owner types dominate scheme ownership, highlighting either centralized control or region-specific trends in infrastructure development.



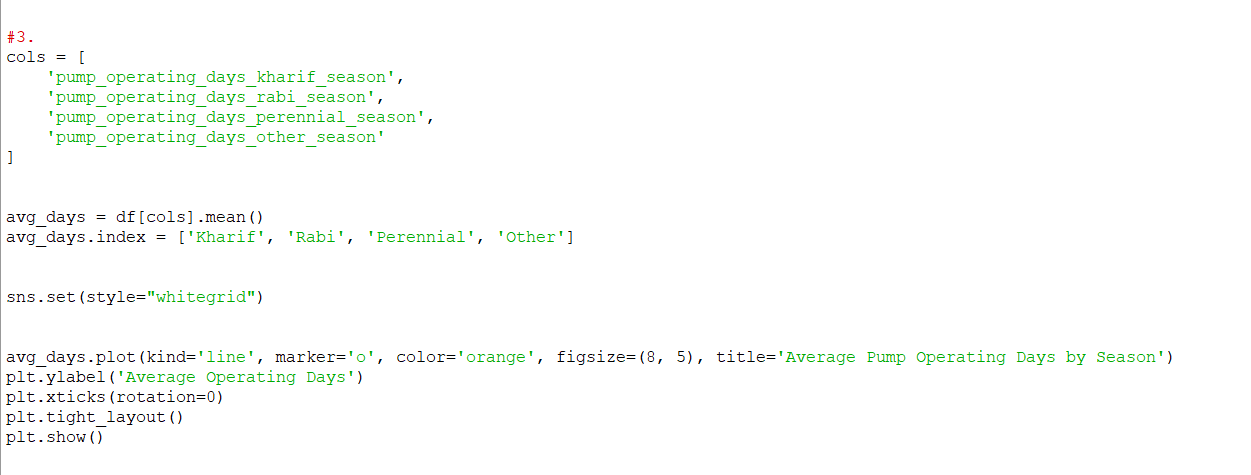


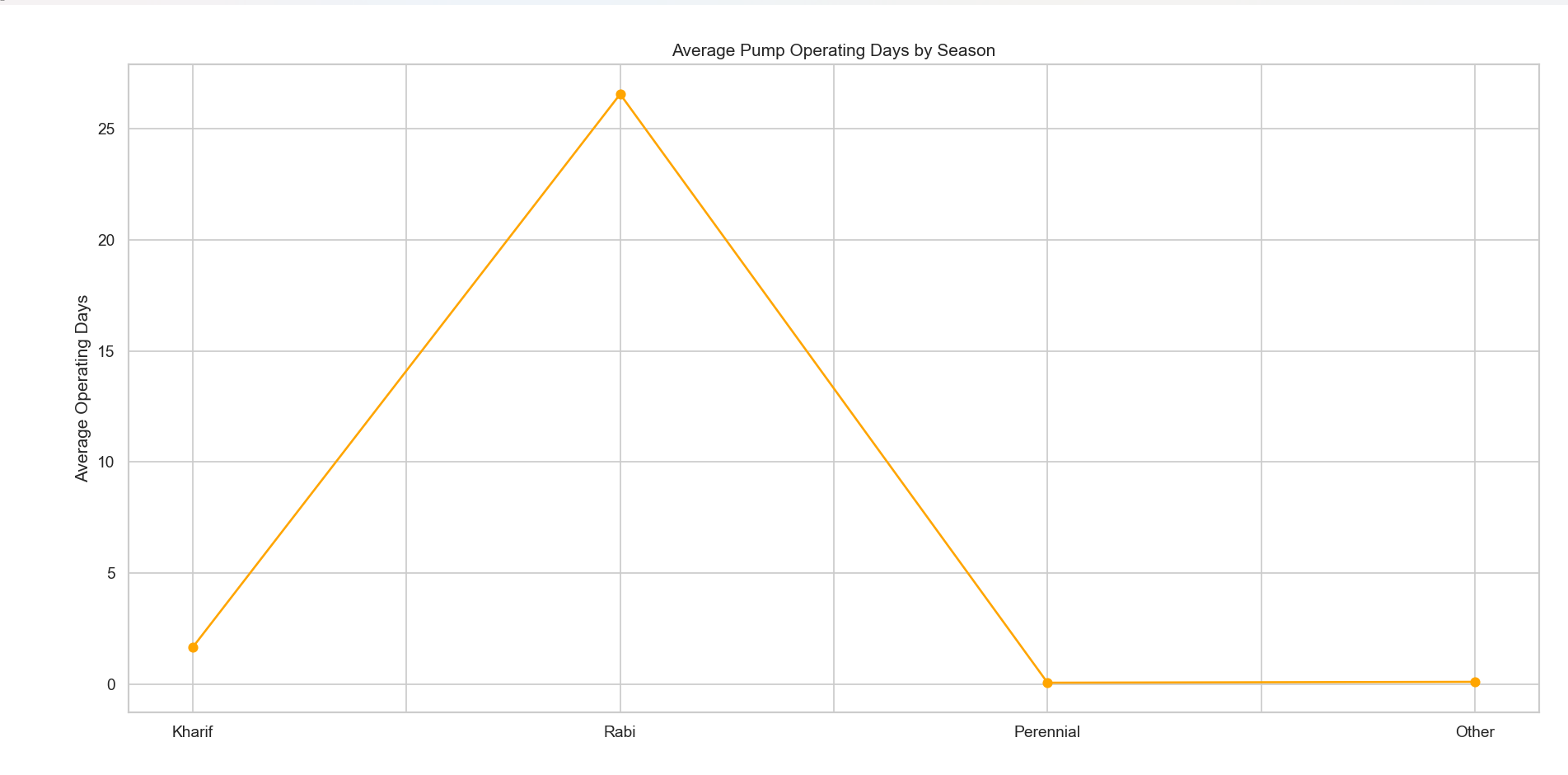
## 4.3 Average Pump Operating Days BY Season

This objective focuses on evaluating the average number of days pumps are operational across different agricultural seasons: Kharif, Rabi, Perennial, and Other.

The dataset includes specific columns for each season’s operating days. By calculating the average for each, a line chart is plotted to visualize operational patterns throughout the year.

**Insights**: The analysis reveals seasonal dependencies on water pumps, with peak usage typically aligning with major crop-growing seasons, indicating the importance of irrigation during those periods.



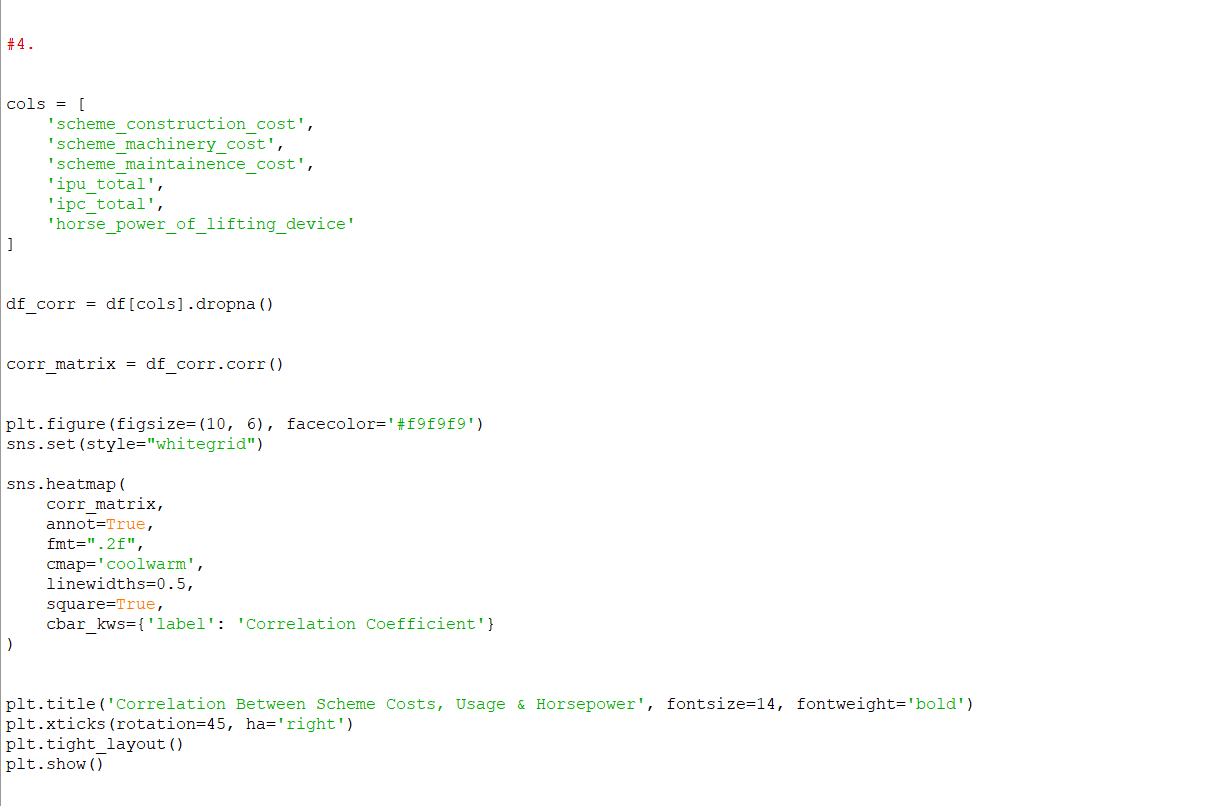


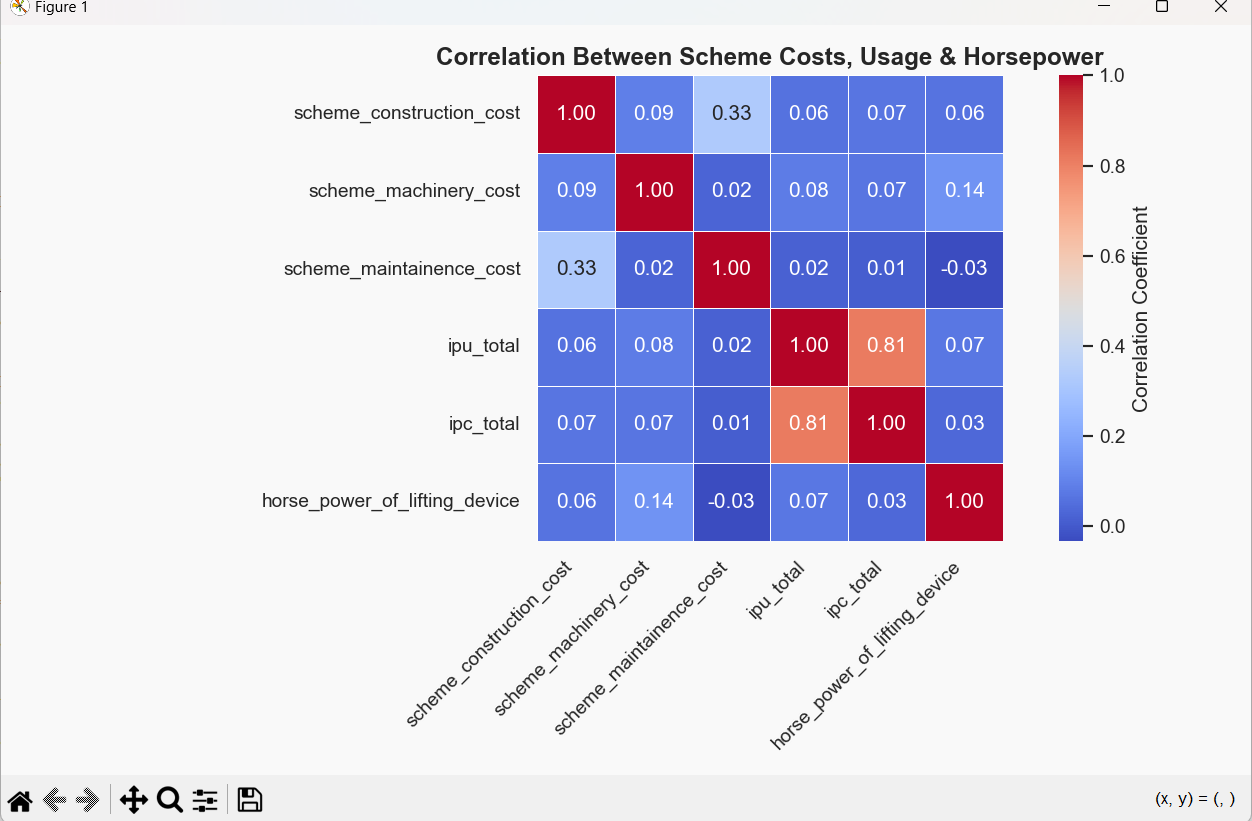
## 4.4 Correlation Between Scheme Costs, Usage & Horsepower

This analysis investigates the relationships among various financial and operational parameters of the schemes, including construction cost, machinery cost, maintenance cost, IPU (irrigation potential utilized), IPC (irrigation potential created), and the horsepower of lifting devices.

A correlation heatmap is generated to identify positive or negative associations between these variables.

**Insights**: Strong correlations between certain costs and performance indicators can guide efficient resource allocation and highlight potential areas for cost optimization or performance improvement.





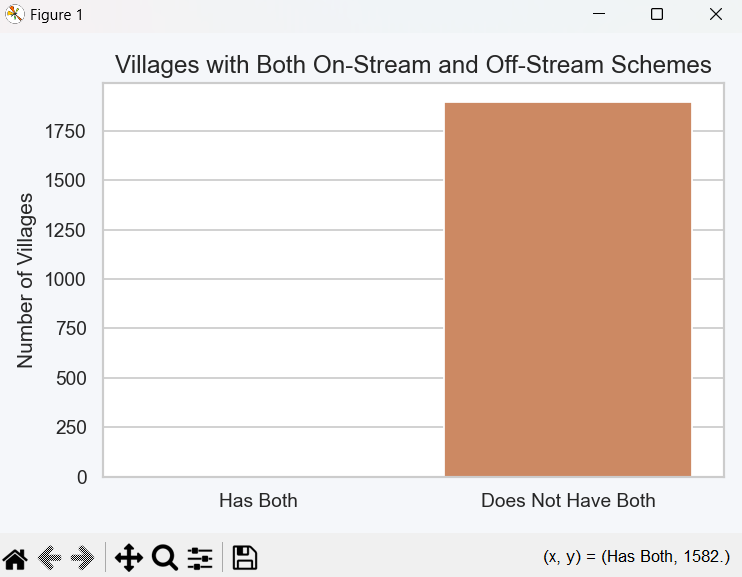
**4.5 Villages With Both On-Stream & Off-Streams**

This analysis compares the number of villages that have implemented both surface flow and surface lift irrigation schemes with those that have not.

Using scheme type references, the dataset is filtered to count how many unique villages have both types of systems. A bar chart visually represents this comparison.

**Insights**: A limited number of villages possess both surface flow and lift schemes, suggesting that integrated irrigation approaches are not widespread and may represent an opportunity for more comprehensive infrastructure planning.



****

**5. Conclusion**

This project analyzed water scheme data across various villages using Python’s data analysis libraries to uncover meaningful insights. The study focused on identifying the top villages by scheme count, understanding ownership patterns, examining seasonal pump usage, evaluating cost and performance correlations, and comparing scheme types across villages.

Through exploratory data analysis and visualization, the project highlighted key trends such as village-level scheme distribution, seasonal demands, and the financial interplay within scheme infrastructure. These insights support informed decision-making for water resource planning and policy formulation.

The analysis underscores the power of data-driven approaches in infrastructure monitoring and demonstrates how Python can streamline complex evaluations for real-world applications in rural development and resource management.

# 6. Future Scope

1. \*\*Predictive Modeling Leveraging machine learning techniques to forecast future water demand, scheme utilization, or seasonal pump requirements at the village level.

2. \*\* Infrastructure Prioritization \*\*: Developing models to identify high-priority villages for future scheme implementation based on current usage patterns and coverage gaps.

3. \*\* Seasonal Trend Analysis \*\*: Applying time series analysis to better understand seasonal irrigation needs and optimize pump operation schedules accordingly.

4. \*\*Interactive Dashboards\*\*: Creating dynamic dashboards using Power BI, Tableau, or Python-based tools (e.g., Dash or Streamlit) for real-time monitoring and stakeholder reporting.

5. \*\* Geospatial Integration \*\*: Incorporating GIS data to visualize scheme distribution and assess proximity to water bodies or agricultural zones for better planning and allocation.

# 7. References

[1] Python Software Foundation, Python Language Reference, https://www.python.org  
[2] Wes McKinney, Python for Data Analysis, O’Reilly Media  
[3] https://pandas.pydata.org/  
[4] https://seaborn.pydata.org/  
[5] https://matplotlib.org/