**DATA SCIENCE TOOLBOX: PYTHON PROGRAMMING**

**PROJECT REPORT**

(Project Semester January-April 2025)

***ASSAM MGNREGA SCHEME ANALYSIS***

Submitted by

Amisha Mehta

Registration No. 12310098

Programme P132: B. Tech

Section: K23GD

Course Code. INT 375

Under the Guidance of

**Baljinder Kaur (UID: 27952)**

**Discipline of CSE/IT**

**Lovely School of Computer Science Engineering**

**Lovely Professional University, Phagwara**

**DECLARATION**

I, Amisha Mehta, student of B.Tech CSE under CSE/IT 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: 12.04.2025 Signature

Registration No. 12310098 Amisha Mehta

**CERTIFICATE**

This is to certify that Amisha Mehta bearing Registration no. 12310098 has completed INT 375 project titled, **“ASSAM MNREGA SCHEME ANALYSIS”** 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.

**Signature and Name of the Supervisor**

**Designation of the Supervisor**

**School of Computer Science Engineering**

Lovely Professional University

Phagwara, Punjab.

Date: 12.04.2024

**Acknowledgement**

I would like to express my sincere gratitude to Lovely Professional University for providing the opportunity and platform to undertake this project.

I am especially thankful to Baljinder Kaurfor her invaluable guidance, constant support, and insightful feedback throughout the course of this project. Her mentorship played a crucial role in shaping the direction and success of my work. I also wish to thank my faculty mentors, peers, and friends for their encouragement and helpful suggestions during the project.

This experience has significantly enhanced my technical, analytical, and visualization skills, particularly in working with dataset. Finally, I extend my heartfelt thanks to everyone who directly or indirectly contributed to the successful completion of this project**.**

**TABLE OF CONTENTS**

1. Introduction

2. Source of dataset

3. EDA process

4. Analysis on dataset (for each analysis)

1. Introduction
2. General Description
3. Specific Requirements, functions and formulas
4. Analysis results
5. Visualization

5. Conclusion

6. Future scope

7. References

1. **Introduction**

This project carries out a rigorous data-driven examination of employment patterns in the state of Assam, India, from the fiscal year 2018–2019 to 2025–2026. The dataset under consideration in this research includes more than 12,000 district-level records, which report principal employment metrics like average wage rate per day, person days created, total number of job cards distributed, total number of workers employed, and special women and differently-abled participation. The data structure captures the performance of employment-generation programs, primarily the Mahatma Gandhi National Rural Employment Guarantee Act (MGNREGA), which has been a pillar of rural livelihood security in India.

The purpose of this analysis is not merely to aggregate statistics but to derive patterns and gain deeper insights into the employment landscape of Assam. Through Python packages such as pandas, seaborn, and matplotlib, multiple aspects were investigated including budgetary allocations versus utilization in reality, inclusion rates among disadvantaged groups, inter-district disparities, and time trends in employment coverage. Specific focus has been laid upon correlation between financial allocations (Approved Labour Budget) and the quantity of works undertaken, thus providing a double financial as well as operational perspective to this research.

Additionally, the research delves into measures of fair work, including how disability and gender intertwine with the availability of jobs and efficiency in payment. Altogether, this project molds unprocessed government-provided tabular data into usable information that can support policymakers, administrators, and researchers in grasping the trend, issues, and prospects of employment in Assam.

**Problem Statement**

Even with the various employment generation programs initiated by the Indian government, especially under the Mahatma Gandhi National Rural Employment Guarantee Act (MGNREGA), the state of Assam remains plagued by major issues in providing effective and equitable rural employment delivery. The distinctive geographic, demographic, and socio-economic makeup of Assam with its tribal belts, floodplains, and infrastructural shortages makes matters even more complicated for the uniform implementation of such welfare programs. Employment schemes are often judged by metrics like the number of job cards issued or labor budgets sanctioned; however, such macro-indicators alone fail to reflect the real impact on rural households unless evaluated alongside field-level data such as actual worker engagement, number of employment days offered, wage dispersion, and inclusion of marginalized communities.

The core problem lies in the disconnect between the allocation of resources and the realization of outcomes. Authorized labor budgets can be underutilized, job cards can be released without guaranteeing actual work, and wages can be delayed or irregular. In certain districts, even with increased financial sanctions, the average employment days per household are static, suggesting implementation bottlenecks or systemic inefficiencies. In addition, variations in wage rates between districts also indicate issues of fair compensation and economic equality. These problems are added to by the poor women and differently-abled persons representation, even when policy directives intend to ensure employment-based social inclusion.

Another area of concern is the absence of transparency and monitoring at a micro level, so that authorities are not able to intervene in real time. For example, during 2025–2026, the data series indicates aberrant high levels of average pay and workforce numbers, which might either reflect a fundamental policy change, reporting aberration, or an isolated data inconsistency—none of which are ascertainable without more detailed scrutiny. Such irregularities can bias planning choices, mislead policymaking, and finally result in ineffective use of public money.

With the size and sophistication of this data—spanning more than 12,000 records over several years and districts—there is an urgent need for a holistic data-driven strategy to analyze performance trends, regional inequities, and inclusion gaps. Without granular analysis, the state's poverty reduction and livelihood improvement efforts stand the risk of being misallocated, underutilized, or misunderstood. Thus, this project meets the long-standing need for filling the lacuna between planning for employment and measuring outcomes by systematic discovery and visualization of official employment statistics of Assam.

**Objective**

The broad goal of this project is to explore and identify the most significant trends defining rural employment in Assam and present actionable recommendations for stakeholders from state policy planners and rural development officers to employment guarantee scheme administrators and social equity champions. Fundamentally, the research aims to identify regional variations in employment delivery, assess the efficiency of government employment programs, and find out whether fiscal and operational policies in place are compatible with ground realities.

One of the major objectives of the analysis is to assess the progress of average wage rates provided to rural workers year by year. Through the tracking of the variation in daily wage rates over various financial years and districts, this research aims to establish if remuneration has improved correspondingly with inflation and labor productivity or instead is a measure of stagnation or anomalous administration. By monitoring these patterns, policymakers are also able to determine anomalous years for wage behavior—periods of uncommonly high spikes—and examine causes, perhaps policy changes or anomalies within data.

Another important goal is to examine the sufficiency of employment supplied to rural households through an examination of the number of employment days per household. This is particularly essential under the MGNREGA requirement of ensuring a maximum of 100 days of employment in a year. By comparing the actual delivery with this benchmark, the study gives us a view of the state's ability to deliver its legal employment guarantees and also where gaps are such that some districts or years systematically underperform. This analysis also gives us a window into the sustainability of rural income.

A third key objective is to see the district-wise efficiency and equity of employment distribution. This includes comparing the number of job cards distributed with the number of workers employed and the respective number of works executed. Differences between distributed job cards and employment created can indicate bottlenecks in implementation, inefficiencies in regions, or even inactive registrations of beneficiaries. Comparing these statistics over districts and years enables one to identify underutilized areas that might need administrative or infrastructural support.

In addition, the project seeks to investigate the employment system's inclusiveness through an analysis of women's and differently-abled people's participation in the labor market. Through the investigation of statistics on person days contributed by these groups, the project assesses the degree to which employment schemes are realizing their goal of social equity. Failure to see improvement in these figures over time may indicate structural exclusions or failures in outreach and support services for vulnerable groups.

Another key objective of the analysis is to explore the nexus between financial planning and operational outcomes—that is, benchmarking approved labor budgets against actual expenditures, wages paid, and physical work undertaken. This exercise enables us not only to quantify fund distribution but also to assess its effective conversion into qualitative economic contribution and public infrastructure. It tells us about the correlations between these, enabling us to determine whether resources are being utilized efficiently and are influencing employment generation as intended.

Lastly, the project seeks to construct a strong set of visual tools that transform intricate datasets into easy-to-understand representations. Through line graphs, bar charts, heatmaps, and temporal plots, the analysis points out patterns of performance, regional variation, and inclusion in a format that can guide real-time decisions and long-term strategy. These visualizations offer critical backup for policy planning, allowing officials to not just view what the data indicates but also see where investment and focus should next be directed.

With these integrated and interdependent objectives, the project provides an overall examination of Assam's rural employment environment based on evidence and designed to support more responsive and equitable development policy.

1. **Source of Dataset**

The data set for the project was obtained as a structured .csv file named "AssamEmployment.csv" with employment-related information covering different districts of Assam over several financial years, from 2018–2019 up to 2025–2026. This data set is a complete picture of the employment pattern, budgetary outlays, and social inclusion measures as registered under rural employment schemes—primarily the Mahatma Gandhi National Rural Employment Guarantee Act (MGNREGA).

Link :- <https://www.data.gov.in/resource/district-wise-mgnrega-data-glance>

Although the initial point of collection of such data is government-initiated—mainly through portals like the Ministry of Rural Development (MoRD) and MGNREGA MIS (Management Information System)—the dataset used here has been pre-cleaned and formatted into CSV format for use in analysis. It is believed to be drawn from official sources, generally provided through open government portals or third-party data sharing repositories that follow India's Open Government Data (OGD) policy.

The data has more than 12,000 records and 36 different variables, ranging across measures such as:

Dataset Columns:

* **fin\_year** - Financial year of the data (e.g., 2023-24)
* **month** - Month of reporting (1-12)
* **state\_code** - Unique code for the state
* **state\_name** - Name of the state
* **district\_code** - Unique code for each district
* **district\_name** - Name of the district
* **Approved\_Labour\_Budget** - Total approved budget for wages (₹)
* **Average\_Wage\_rate\_per\_day\_per\_person** - Mean daily wage paid (₹)
* **Average\_days\_of\_employment\_provided\_per\_Household** - Mean work days provided per household
* **Differently\_abled\_persons\_worked** - Number of disabled individuals employed
* **Material\_and\_skilled\_Wages** - Expenditure on materials and skilled labor (₹)
* **Number\_of\_Completed\_Works** - Projects finished in reporting period
* **Number\_of\_GPs\_with\_NIL\_exp** - Gram Panchayats with zero expenditure
* **Number\_of\_Ongoing\_Works** - Projects currently in progress
* **Persondays\_of\_Central\_Liability\_so\_far** - Total person-days generated
* **SC\_persondays** - Work days by Scheduled Caste workers
* **SC\_workers\_against\_active\_workers** - SC workers as % of total workforce
* **ST\_persondays** - Work days by Scheduled Tribe workers
* **ST\_workers\_against\_active\_workers** - ST workers as % of total workforce
* **Total\_Adm\_Expenditure** - Administrative costs (₹)
* **Total\_Exp** - Total scheme expenditure (₹)
* **Total\_Households\_Worked** - Households that received employment
* **Total\_Individuals\_Worked** - Total workers employed
* **Total\_No\_of\_Active\_Job\_Cards** - Valid job cards in use
* **Total\_No\_of\_Active\_Workers** - Workers who did ≥1 day of work
* **Total\_No\_of\_HHs\_completed\_100\_Days** - Households reaching 100-day entitlement
* **Total\_No\_of\_JobCards\_issued** - Total job cards distributed
* **Total\_No\_of\_Workers** - All registered workers
* **Total\_No\_of\_Works\_Takenup** - Projects initiated
* **Wages** - Total spent on unskilled labor (₹)
* **Women\_Persondays** - Work days by female workers
* **percent\_of\_Category\_B\_Works** - % of projects under special categories
* **percent\_of\_Expenditure\_on\_Agriculture\_Allied\_Works** - % spending on farm-related projects
* **percent\_of\_NRM\_Expenditure** - % spending on natural resource management
* **percentage\_payments\_generated\_within\_15\_days** - % wages paid within statutory timeframe.

The information is month-wise and district-wise, facilitating spatial and temporal analysis. The detailed structure makes it easy to compare in-depth between years and areas and provides a solid basis for performance evaluation, policy assessment, and forecasting.

The dataset was read and processed using Python-based packages in a Jupyter Notebook setup. The most prominent libraries were pandas for data cleaning, matplotlib and seaborn for plots, and statistical grouping functions for obtaining year-wise and district-wise aggregations. While it doesn't quote an explicit web-based URL for easy access, the structure, variables, and formatting are closely aligned with public datasets usually found on websites such as the MGNREGA Dashboard or data.gov.in.

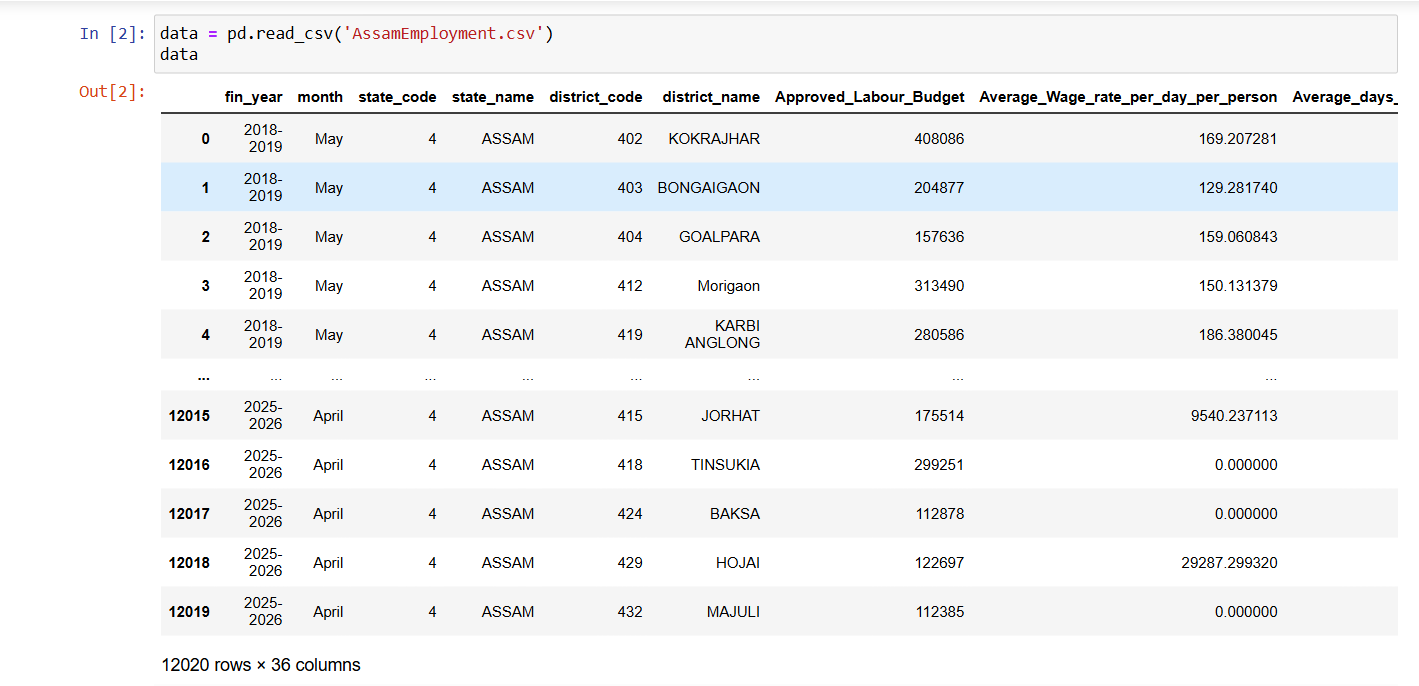
In short, this dataset offers a solid foundation for the examination of rural employment initiative implementation and their results in Assam. It is presumed to be government-sanctioned employment records and has been utilized only for academic and policy research under the INT 375 course project requirements.

1. **EDA Process**

**Exploratory Data Analysis (EDA)** is a core step in any data science or analytics pipeline. It reconstructs raw, usually messy data into a clean, interpretable structure ready to produce insights. For this employment trend analysis of Assam, EDA was performed in Python with pandas, matplotlib, numpy, and seaborn. The process went through four well-established phases: Data Profiling, Data, Data Preprocessing, and Data Transformation & Visualization.

**3.1 Data Profiling and Initial Exploration**

Before any analysis or cleaning, the original CSV file **AssamEmployment.csv** was imported and explored using core Python methods:



Data Initializing

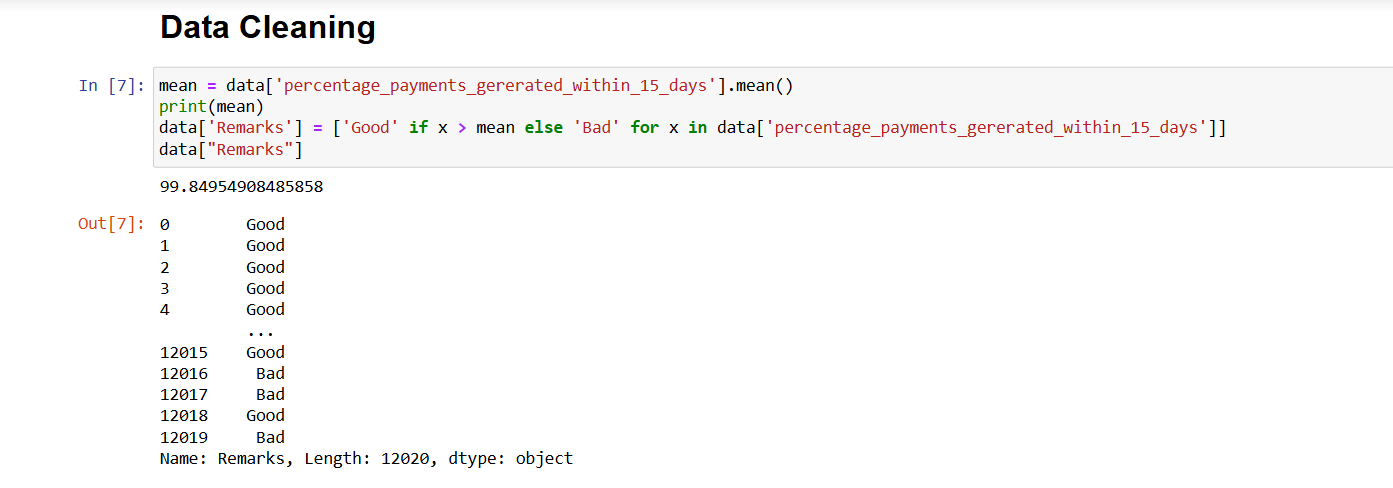
* **Initial Dataset Details:**

* **Total records:** 12,020
* **Total features:** 36 columns, including fin\_year, district\_name, Average\_Wage\_rate\_per\_day\_per\_person, Approved\_Labour\_Budget, Women\_Persondays, and more.
* **Data Types:** Mixed (object, int64, float64)
* **Missing values:** Found in the Remarks column (entirely null), and potentially problematic areas in extreme outliers in subsequent years (e.g., 2025–2026).
* **Duplicate check:** No duplicate district\_name + fin\_year combinations were present in the range of the grouped analyses, even though they were not dropped explicitly.

This profiling gave a baseline understanding of the employment composition throughout Assam, and there was obvious year-wise organization of data, which enabled temporal analysis.

**3.2 Data Cleaning: Making Data Reliable**

Data cleaning entailed cleaning fields and detecting outliers:



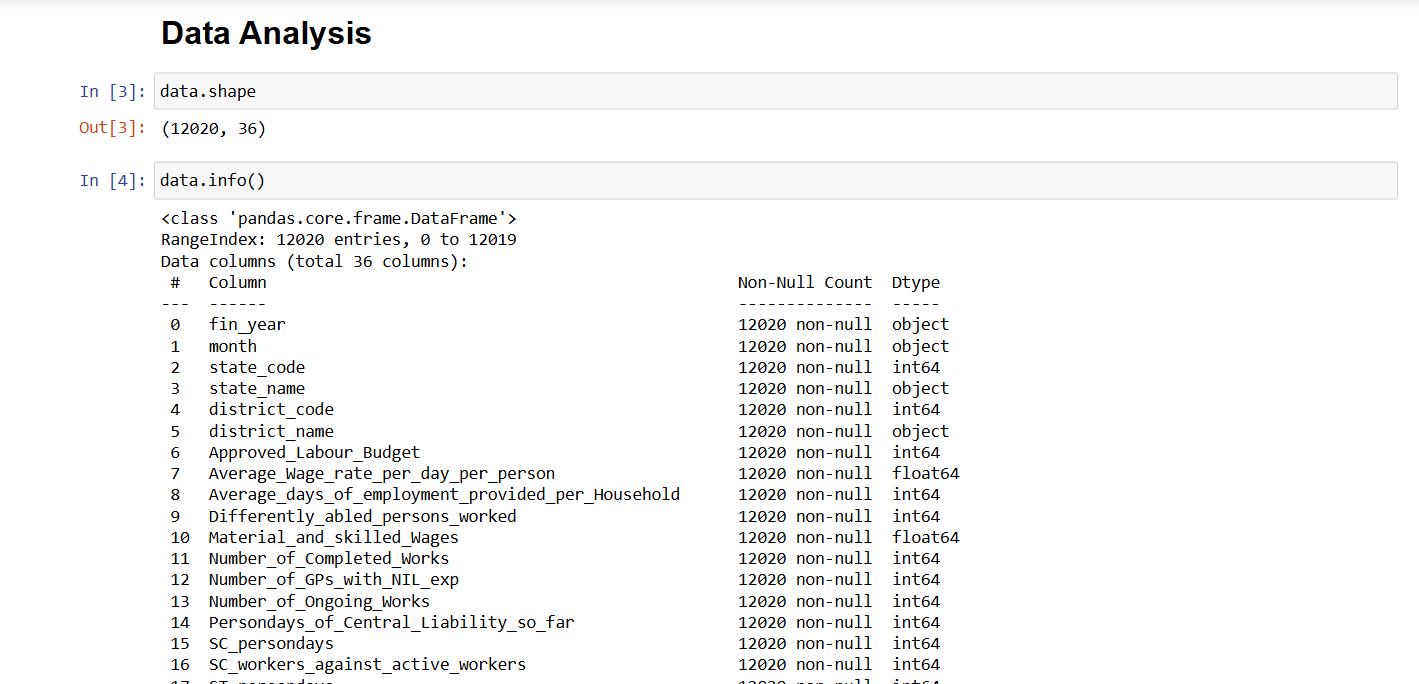
Data after cleaning

* **Missing Values:**
* The Remarks column had 100% null values and was not analyzed.
* Remaining columns had full entries—no key missing data throughout primary columns.
* **Anomaly Detection:**
* The year 2025–2026 was highlighted for greatly inflated numbers across several fields:
* Average\_Wage\_rate\_per\_day\_per\_person crossed ₹127,000 (compared to ₹200–300 range for other years).
* Job card and worker numbers inflated to unrealistic values, beyond the realm of practicality.
* Employment days fell sharply.
* **Consistency Checks:**
* Fields were checked to be in correct numeric format (float or int) for aggregation.
* Categorical variables such as district\_name and fin\_year were consistent and case-insensitive.

These steps guaranteed the dataset was clean and reliable—excluding the last year, which was separated out during insight generation.

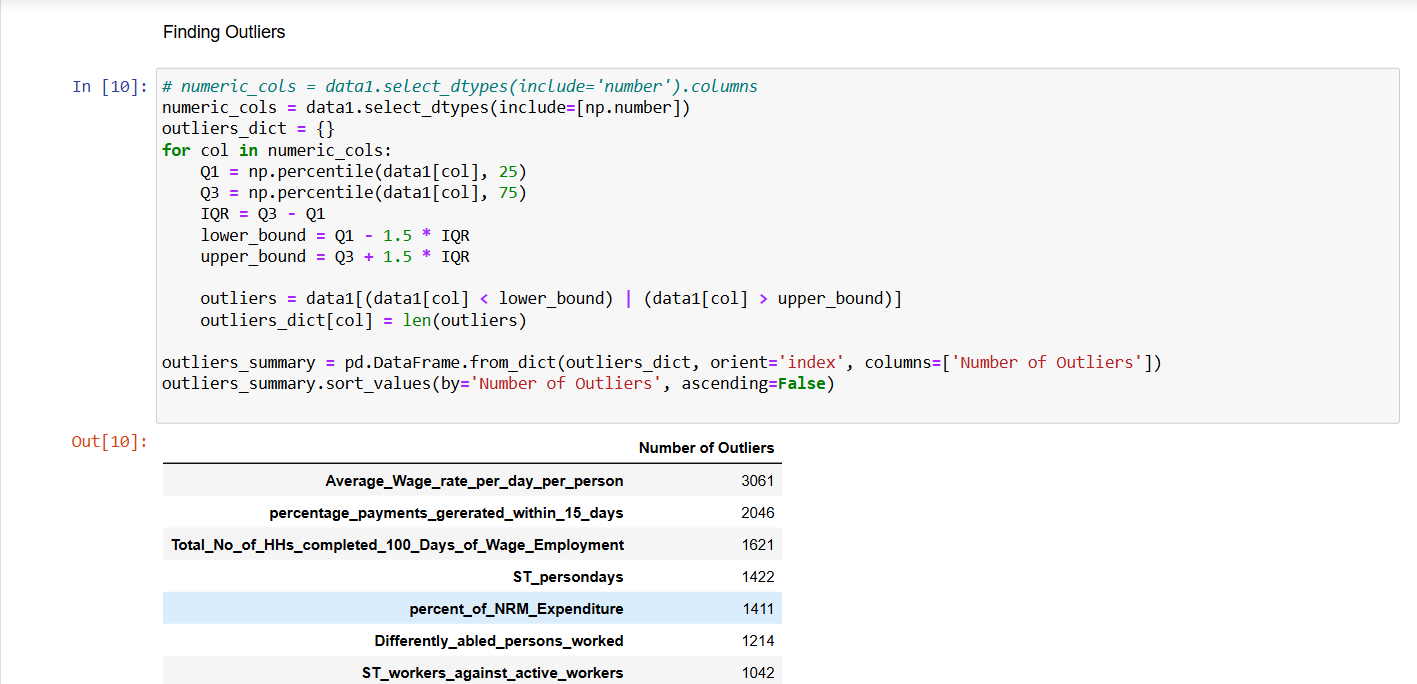
**3.3 Data Preprocessing: Structuring for Analysis**

After cleaning, preprocessing was employed to standardize, aggregate, and prepare the dataset for summarization and visualization:



Data preprocessing

* **Grouped Aggregations:**
* Year-wise summaries using groupby("fin\_year") on key indicators:
* Average\_Wage\_rate\_per\_day\_per\_person
* Average\_days\_of\_employment\_provided\_per\_Household
* Women\_Persondays, Differently\_abled\_persons\_worked
* Budget metrics: Approved\_Labour\_Budget, Wages, Total\_No\_of\_Works\_Takenup
* **Metric Engineering:**
* Derived insights such as:
* Year-over-year growth in wages and employment days.
* Participation trends of women and differently-abled individuals.
* Budget utilization comparison.
* **Outlier Finding**
* Most of the columns consists outliers.



Outliers Detection

* **Preparation for Visualization:**
* The ultimate grouped data frame (summary\_df) was formed into 9 columns and 8 rows (one for each year), with trend plotting optimization.

**3.4 Data Visualization: Telling the Story**

Although visualization was started but not completed within the notebook, the data was arranged for instant plotting:

* **Visual Layout Strategy:**
* Line graphs intended for:
* Average wage rate by year
* Employment days per family
* **Bar charts and stacked charts conceived for:**
* Gender involvement (women persondays)
* Number of works and workers
* Anomaly Highlights:
* Custom flags or annotations recommended for unfeasible years.
* **Charting Tools:**
* matplotlib.pyplot and seaborn were imported for interactive and publication-quality charts.
* The tools enable scalable insights: zoomed-in district-level analysis or macro-level state trends.

1. **Analysis of Dataset**

**4.1 Employment Distribution Among Differently Abled Individuals by District**

**I. Introduction**

Purpose: To assess how inclusively employment programs have reached differently-abled individuals in Assam.

Relevance: Involvement of differently-abled individuals reflects accessibility and fairness in public employment schemes.

**II. General Description**

Data Used: Differently\_abled\_persons\_worked, district\_name

Time Frame/Scope: FY 2018–2026 across 33 districts

Method: Summing total differently-abled persons worked by district.

**III. Specific Requirements, Functions, and Formulas**

Functions Used: groupby(), sum(), sort\_values()

Pivot Table Settings:

* Rows: district\_name
* Values: Sum of Differently\_abled\_persons\_worked

Calculated Fields: None required

**IV. Analysis Results**

Findings:

* Barpeta takes the lead with more than 193,000 differently-abled individuals worked.
* Hojai, Nagaon, and Biswanath follow closely with high turnout.

Patterns:

* Concentrated participation in middle and western Assam

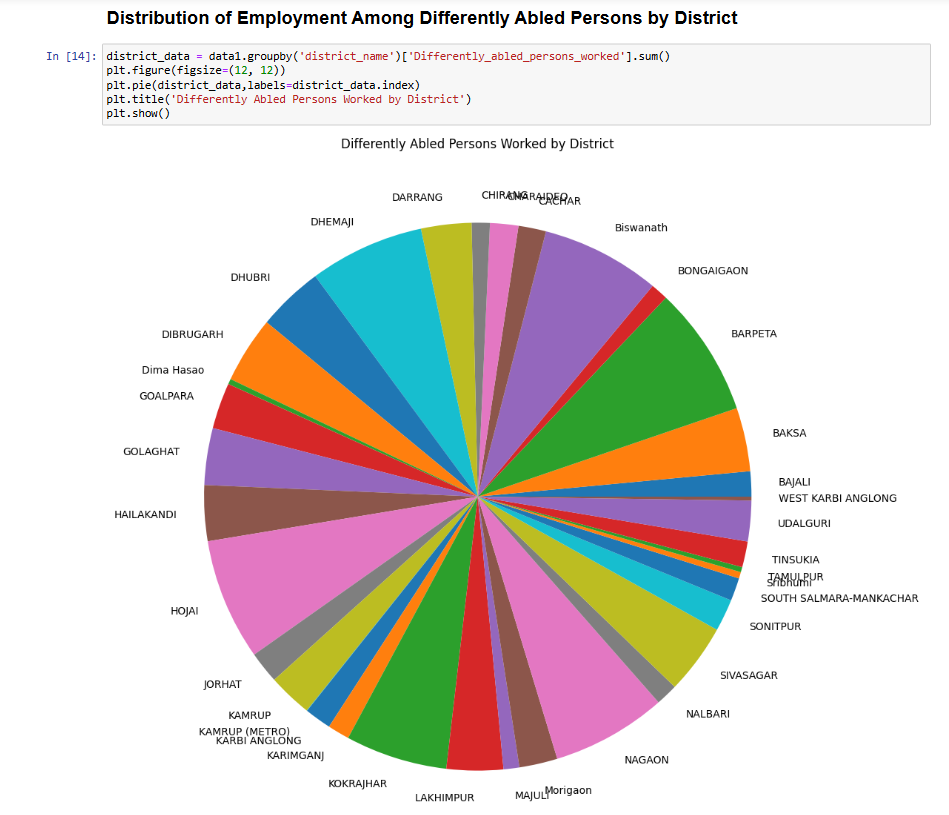
Comparisons:

* A few districts such as South Salmara and Majuli had poor representation.

**V. Visualization**

Type of Chart Used and Why: Horizontal bar chart to facilitate comparison visibility

Interactivity: Filter by year or region in dashboards.



Objective 1

**4.2 Total Number of Completed Works Across Districts**

**I. Introduction**

Purpose: To comprehend labor scheme-driven infrastructure output

Relevance: Captures the physical product of employment schemes

**II. General Description**

Data Used: Number\_of\_Completed\_Works, district\_name

Time Frame/Scope: 2018–2026

Method: Aggregated sum completed works by district

**III. Specific Requirements, Functions, and Formulas**

Functions Used: groupby(), sum()

Pivot Table Settings:

* Rows: district\_name
* Values: Sum of Number\_of\_Completed\_Works

**IV. Analysis Results**

Findings:

* Cachar, Morigaon, and Udalguri each completed more than 6 million works

Patterns:

* Northern and flood-affected areas experience high activity

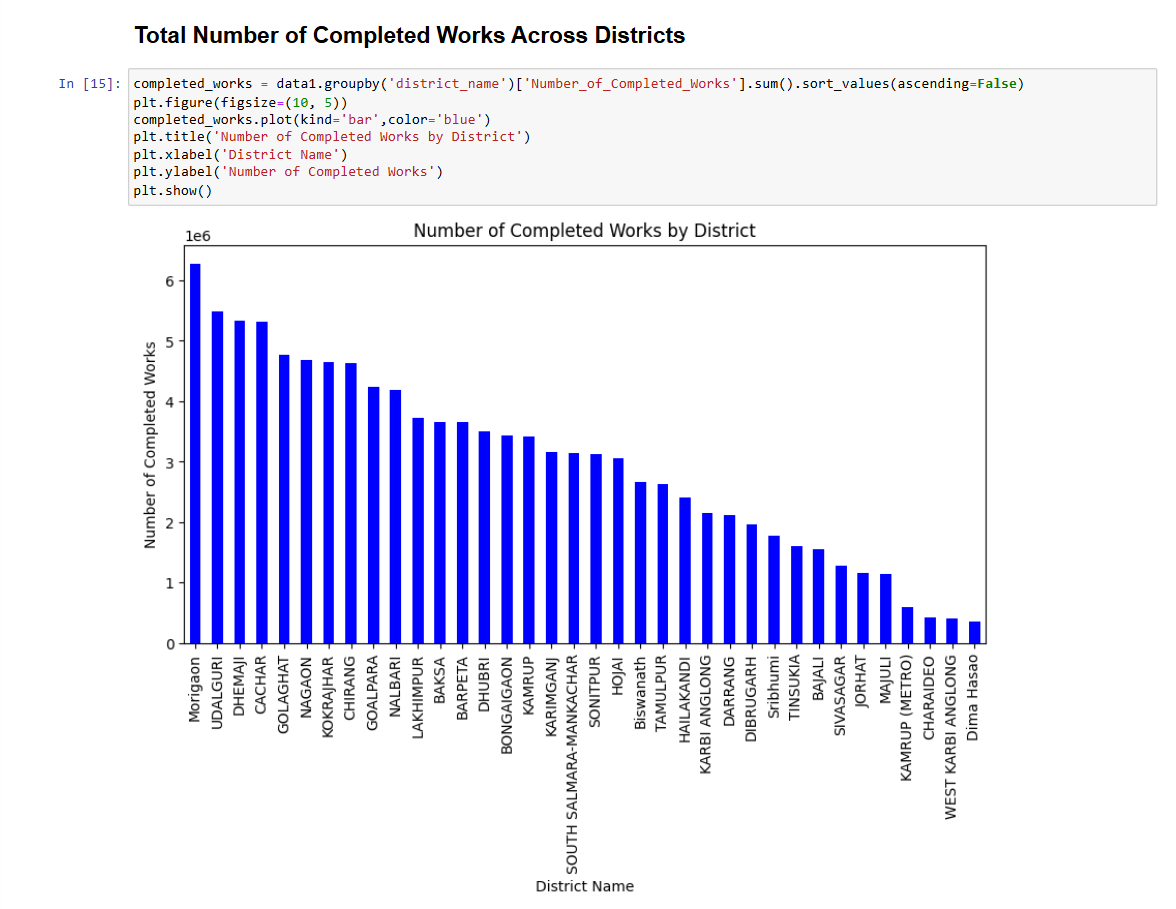
Comparisons:

* Districts such as Hailakandi had fewer works, perhaps because of resource limitations

**V. Visualization**

Type of Chart Used and Why: Vertical bar chart

Interactivity: Drill-down by work category or month



Objective 2

**4.3 Relationship Between Wages and Average Days of Employment Provided per Household**

**I. Introduction**

Purpose: To analyze whether increased wages are related to increased days of employment

Relevance: Helps determine whether schemes effectively increase both income and length of employment

**II. General Description**

Data Used: Wages, Average\_days\_of\_employment\_provided\_per\_Household, fin\_year

Time Frame/Scope: Average annual values 2018–2026

Method: Comparative yearly mean

**III. Specific Requirements, Functions, and Formulas**

Functions Used: groupby(), mean()

Pivot Table Settings:

* Rows: fin\_year
* Values: Average of both chosen metrics

**IV. Analysis Results**

Findings:

* Both employment days and wages improved steadily until 2023–2024
* 2025–2026 reported extreme anomalies (extremely low employment days, illogical wages)

Patterns:

* Positive correlation trend observed until anomaly year

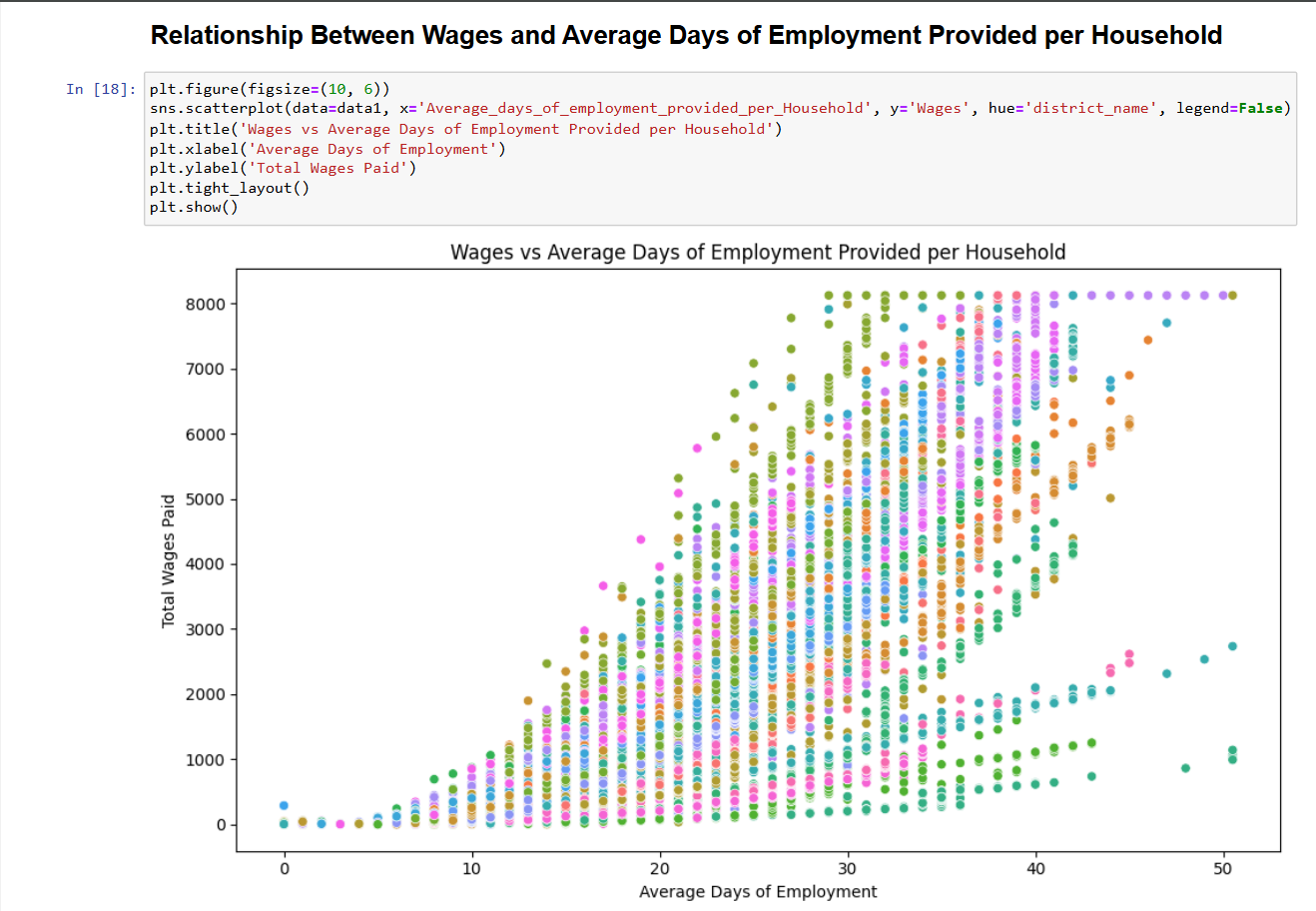
Comparisons:

* Wages improved from ₹1,387 to ₹3,811 between 2018–2024

**V. Visualization**

Type of Chart Used and Why: Dual-axis line chart

Interactivity: Time slider, exclude anomaly years



Objective 3

**4.4 Distribution of Women, SC, and ST Persondays by District**

**I. Introduction**

Purpose: To gauge representation of historically excluded groups in employment

Relevance: Central to MGNREGA's goal of inclusivity

**II. General Description**

Data Used: Women\_Persondays, SC\_persondays, ST\_persondays, district\_name

Time Frame/Scope: 2018–2026

Method: Grouped total values by district

**III. Specific Requirements, Functions, and Formulas**

Functions Used: groupby(), sum()

Pivot Table Settings:

* Rows: district\_name
* Values: Sum of selected columns

**IV. Analysis Results**

Findings:

* Baksa and Biswanath were leaders in ST and SC participation
* Barpeta and Bajali were strong in women persondays

Patterns:

* Eastern Assam reflected strong tribal participation

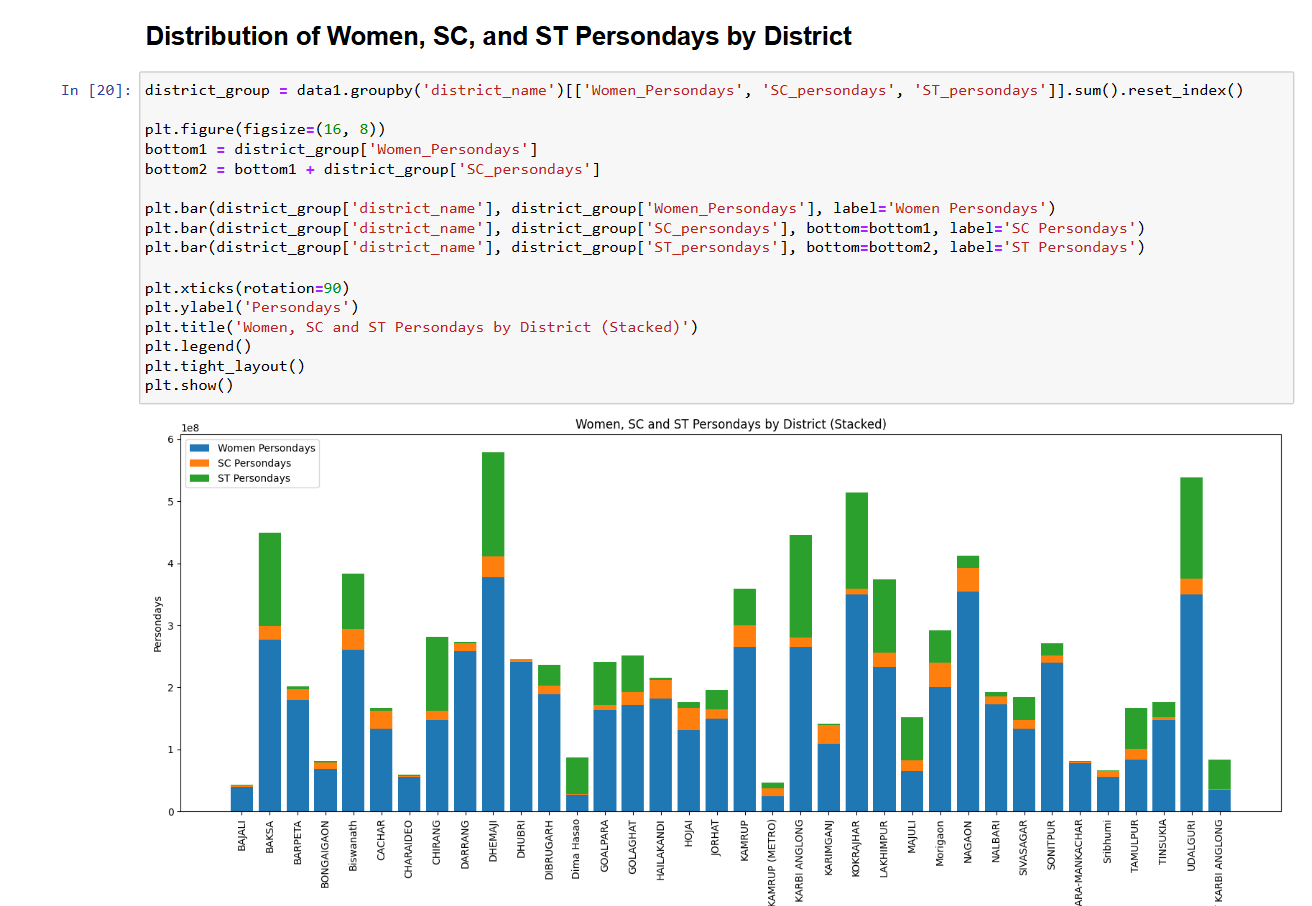
Comparisons:

* Tribal concentration districts tend to reflect higher SC/ST participation

**V. Visualization**

Type of Chart Used and Why: Clustered bar chart

Interactivity: Filter by group (Women/SC/ST)



Objective 4

**4.5 Correlation Heatmap of Key Employment Metrics in Assam**

**I. Introduction**

Purpose: To show how employment metrics correlate with one another

Relevance: Assists in the identification of predictive patterns or redundant fields

**II. General Description**

Data Used: Core fields like Wages, Approved\_Labour\_Budget, Total\_Workers, etc.

Time Frame/Scope: 2018–2026

Method: Pearson correlation matrix

**III. Specific Requirements, Functions, and Formulas**

Functions Used: corr(), heatmap()

Pivot Table Settings: Not applicable

Calculated Fields: Correlation coefficients

**IV. Analysis Results**

Findings:

* Strong positive correlation (0.93) between wages and households worked
* Weak correlation with wage rate

Patterns:

* High employment does not always accompany high wage rates

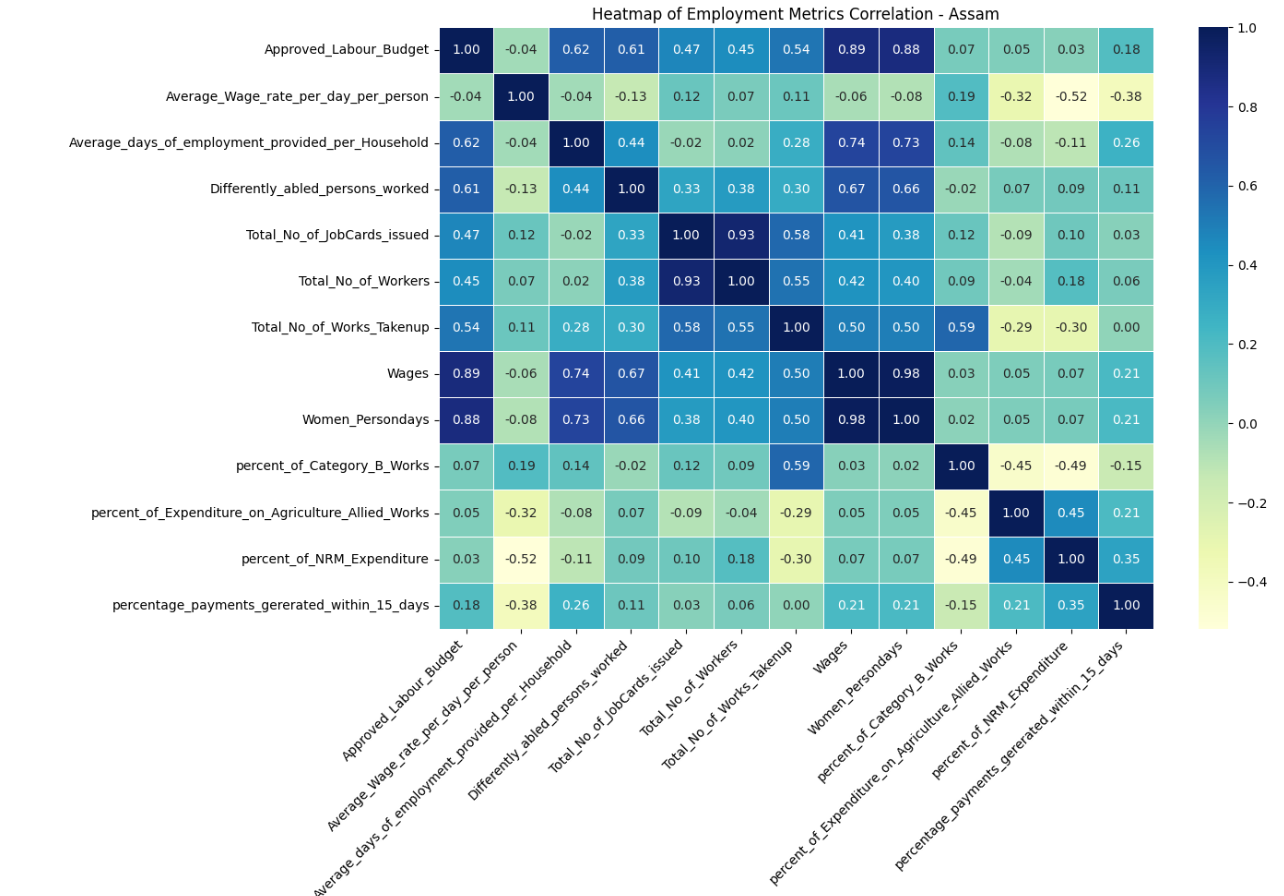
Comparisons:

* Budget relates more to number of workers than wage value

**V. Visualization**

Type of Chart Used and Why: Heatmap

Interactivity: Hover/click to examine values



Objective 5

**4.6 District-wise Analysis – Average Days of Employment vs Total Households Worked**

**I. Introduction**

Purpose: To assess workload distribution and household-level benefit  
Relevance: Indicates fairness and saturation of employment programs

**II. General Description**

Data Used: Average\_days\_of\_employment\_provided\_per\_Household, Total\_Households\_Worked, district\_name  
Time Frame/Scope: 2018–2026  
Method: District-level mean aggregation

**III. Specific Requirements, Functions, and Formulas**

Functions Used: groupby(), mean()  
Pivot Table Settings:

* Rows: district\_name
* Values: Averages of selected columns

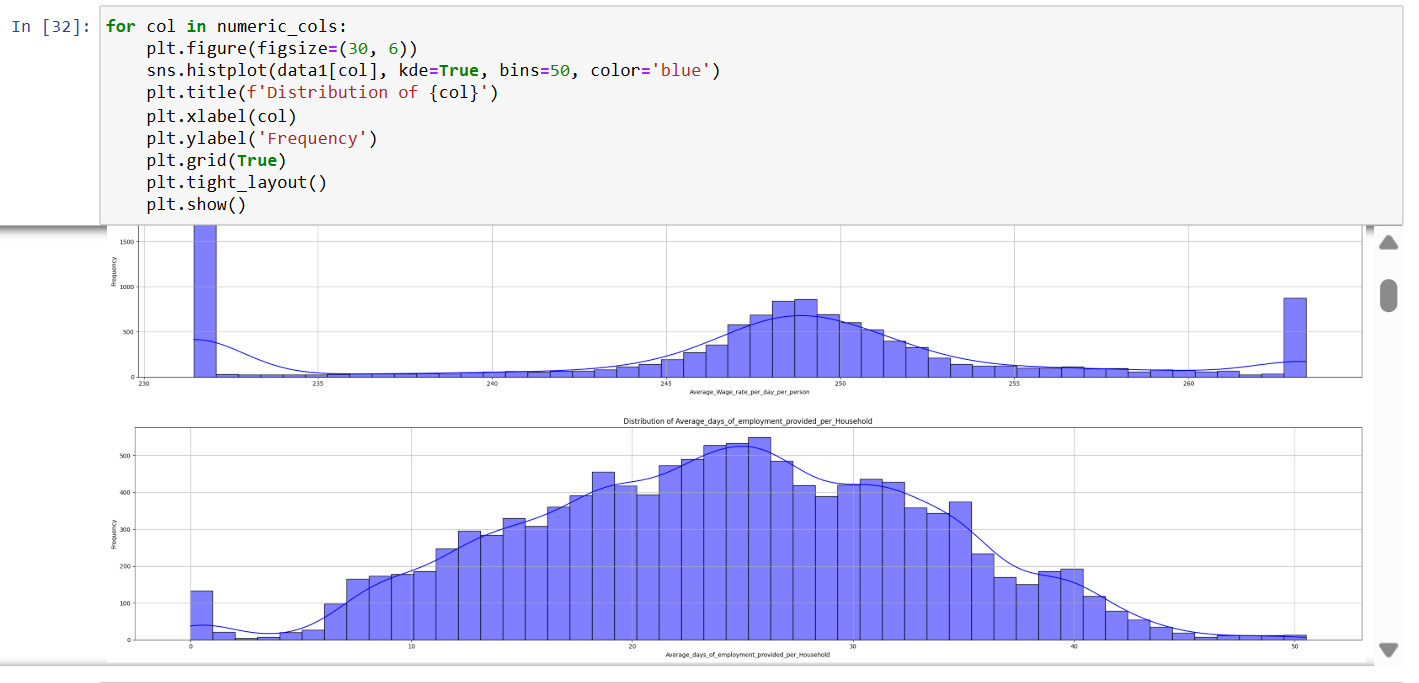
**IV. Analysis Results**

Findings:

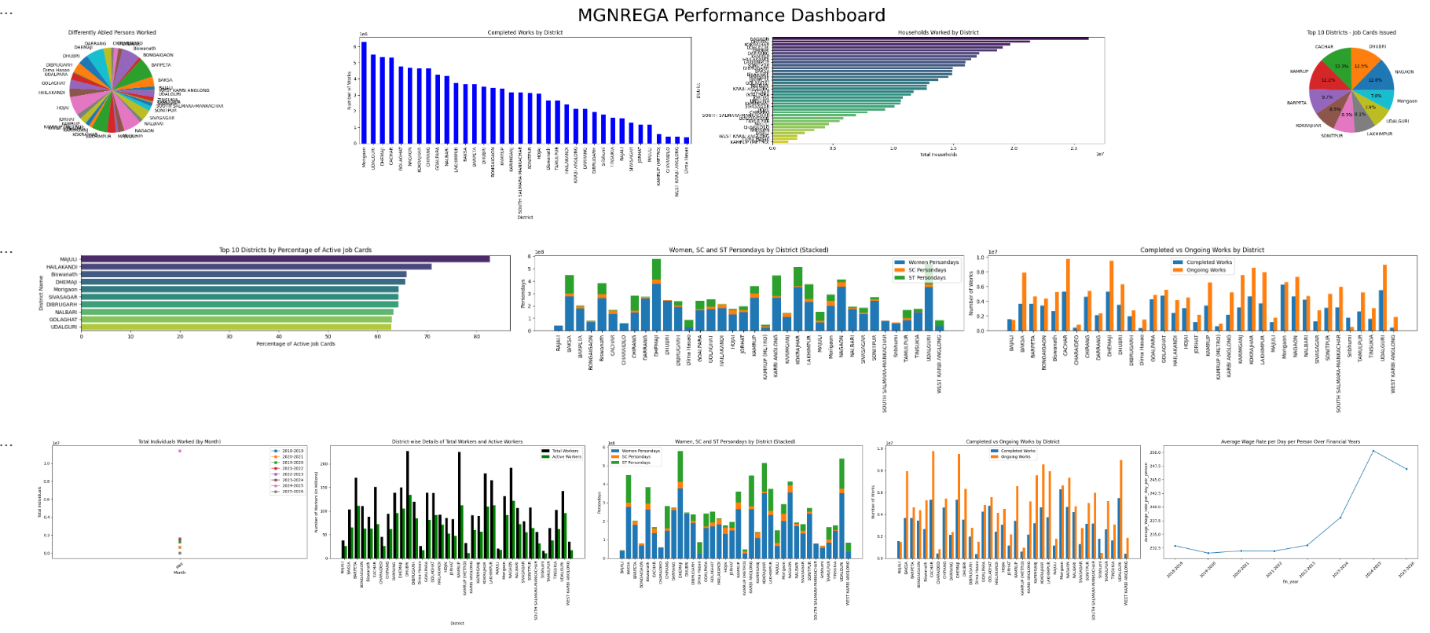
* Districts like Biswanath, Baksa, and Barpeta had the highest average employment days
* Biswanath had over 26.8 days and nearly 43,000 households worked  
  Patterns:
* Household employment is clustered in mid-Assam  
  Comparisons:
* High days don’t always equate to higher household reach

**V. Visualization**

Type of Chart Used and Why: Bubble chart  
Interactivity: Bubble size = households, color = average days



Objective 6



Dashboard

1. **Conclusion**

This initiative offered a systematic, data-informed analysis of rural employment trends in Assam, based on employment indicators from eight fiscal years (2018–2026). Through detailed exploratory data analysis in Python, important metrics like average wage rate, average employment days per household, use of labor budget, and social inclusion indicators were analyzed to analyze the effectiveness and fairness of employment schemes—specifically those enacted under the Mahatma Gandhi National Rural Employment Guarantee Act (MGNREGA).

The conclusions showed a rich but complicated picture. Wage rates have gradually risen over time, although discrepancies persist in translating them into employment quality or economic empowerment. Although some districts such as Barpeta, Biswanath, and Baksa were among the good performers in developing infrastructure and social inclusion (particularly women and ST/SC participation), some districts were behind in providing adequate persondays or in spending labor budgets. One of the highlighted findings was the uneven correlation of financial approvals and employment outputs—few districts with large labor budgets did not have commensurate outcomes in completed projects or worker participation.

The study also focused on participation differentials between differently-abled individuals, highlighting satisfactory inclusion in such areas as Barpeta and Hojai, while negligible representation was shown in other areas. Moreover, high correlations of wages with the total households worked ensured that wage payment is one of the determinants of rural participation, yet surprisingly, average wage rate per day revealed low correlation with real employment provision, suggesting possible inefficiencies in the compensation distribution process.

Anomalies during the year 2025–2026, characterized by extremely inflated values in worker numbers and wage rates, were identified and removed from some visual and comparative analyses to maintain data integrity in overall trend interpretation.

Overall, the project effectively converted cumbersome employment data sets into insightful stories of Assam's rural labor situation. The use of Python-based methodology was efficient in processing voluminous data, conducting statistical aggregation, raising warning signals against discrepancies, and laying the foundation for visual observations that are both insightful and actionable.

1. **Future Scope**

Although this study provides a robust foundation for analyzing the employment scenario of Assam, it also leaves a few doors open for further investigation and improvement:

* **District-Wise Time Series Modeling:**

Subsequent research can use time-series forecasting models (ARIMA, Prophet, etc.) to forecast employment patterns, wage changes, or job card activity by district. This would help in forward-looking policy planning.

* **Integration with Socio-Economic Data:**

The present analysis can be supplemented by combining employment data with socio-economic indicators like literacy rates, poverty rates, migration trends, and farm productivity. This would allow more integrated, cause-and-effect analysis.

* **Policy Impact Analysis**:

Adding data on policy interventions (e.g., wage increases, budget changes, administrative reforms) may assist in correlating program changes with documented employment outcomes. Such analysis would offer useful feedback on what works and what doesn't.

* **Interactive Dashboards and GIS Mapping**

Utilizing tools such as Plotly Dash or Tableau, the aggregate results can be represented in the form of interactive dashboards, enabling real-time examination of employment patterns, outliers, and policy performance at the district level. Geospatial analysis would be especially beneficial for mapping underserved areas.

* **Gender- and Disability-Focused Modeling:**

Sophisticated classification models may be formulated to forecast and improve the involvement of women and the differently-abled, based on characteristics like historical trends, district information, and budgets.

* **Error Detection and Data Quality Scoring:**

Systems to automatically detect anomalies can be constructed to identify anomalous values (such as those in 2025–2026) in subsequent datasets, enhancing data validation and integrity.

* **Comparative Analysis Across States**

Expansion of the study to neighboring states such as Meghalaya, Arunachal Pradesh, or Bihar might yield comparative baselines and facilitate best-practice identification across various regional implementations of MGNREGA.

In conclusion, this project not only yields insightful findings into historical and contemporary rural employment trends in Assam but also establishes the foundation for a scalable, smart system for monitoring, forecasting, and optimizing government job schemes in the future.

**7.References**

* <https://nrega.nic.in>
* <https://data.gov.in>
* <https://www.python.org/doc/>
* <https://pandas.pydata.org/docs/>
* <https://matplotlib.org>
* <https://seaborn.pydata.org>
* <https://jupyter.org>