



Innovation & Entrepreneurship Hub for Educated Rural Youth (SURE Trust – IERY)

Comprehensive Analysis of Enrolment Trends at SURE TRUST (2022-2024)

The domain of the Project:

This project focuses on the comprehensive analysis and visualization of SURE TRUST's student enrolment data spanning the years 2022 to 2024. Utilizing Python for effective data cleaning and Power BI for creating interactive visual representations, the study aims to uncover patterns and trends in enrolment. Emphasizing data accuracy and clarity, it demonstrates how powerful visual analytics can inform strategic decision-making in educational institutions.

Team Mentors

Mr. Ravi Shankar

(Tata ProEngage volunteer from Tata Consultancy Services)

Team Members:

1. Mr. SPARSH AGARWAL B.Tech ---- Team Leader
2. Mr. PATNAM NAVEEN KUMAR B.Tech
3. Mr. KONTAM GANGADHAR B.Tech

Period of the project

2 MONTHS

25th December 2024 to 25th February 2025



Innovation & Entrepreneurship Hub for Educated Rural Youth (SURE Trust – IERY)

Declaration

The project titled “Comprehensive Analysis of Enrolment Trends at SURE TRUST (2022-2024)” has been mentored by Mr. Ravi Shankar, organized by SURE Trust, from December 2024 to February 2025, for the benefit of the educated unemployed rural youth for gaining hands-on experience in working on industry relevant projects that would take them closer to the prospective employer. I declare that to the best of my knowledge the members of the team mentioned below, have worked on it successfully and enhanced their practical knowledge in the domain.

Team Members:

1. Mr. SPARSH AGARWAL
2. Mr. KONTAM GANGADHAR
3. Mr. PATNAM NAVEEN KUMAR

Mr. Ravi Shankar
Tata Consultancy Services

Prof. Radhakumari
Executive Director & Founder
SURE Trust



Table of Contents

Executive Summary	4
1. Objective.....	4
2. Steps	4
3. Narratives	4
4. Metrics	4
5. Key Takeaways:.....	5
5.1. Problems	5
5.2. Recommendations.....	5
Introduction:.....	6
1) Background and context of the project	6
2) Problem Statement	6
3) Goals	7
4) Scope.....	7
5) Limitations	7
Project Objectives:	8
1) Objectives	8
Methodology and Results:	9
1) Methods/Technology	9
2) Tools/Software	9
3) Project Architecture	9
Learning and Reflection:	10
1) Overall Experience.....	10
2) Document all team member's new learnings and contributions	10
Future Scope:.....	11
1) Strategic Decisions	11
2) Resource Allocation:.....	11



Executive Summary

1. **Objective:** To conduct a comprehensive analysis of SURE TRUST's student enrolment data from 2022 to 2024, aiming to uncover trends and patterns that inform strategic decision-making and enhance the institution's growth.
2. **Steps:**
 - 2.1. Converting raw CSV files to XLS format for compatibility.
 - 2.2. Importing the raw XLS data into Power BI.
 - 2.3. Employing Python scripts to clean and preprocess the data without altering the original database.
 - 2.4. Applying filters to omit unnecessary information.
 - 2.5. Creating essential metrics and KPI's, including qualification-wise totals, yearly and monthly enrolment figures, gender distribution, and state-wise analysis.
 - 2.6. Utilizing Power BI's visual tools to create dashboard's that make data insights accessible.
3. **Narratives:**
 - 3.1. **Demographic Insights:** a) Gender Distribution is roughly same for both genders with **52.32%** of male students while rest are females.
b) State wise enrolment shows that **Andhra Pradesh** is the most contributing state with **75.80%** followed by **Karnataka** with **8.27%** and least contributing state is **Sikkim** with **0.01%**.
 - 3.2. **Enrolment Patterns:** a) Yearly chart shows a down trend indicating in decreased number of enrolments with **4221** students enrolling in **2022** to **3992** in **2023** to **2922** in **2024**.
b) Monthly chart shows peak enrolment in September which is due to starting of new odd semester in Indian Engineering Colleges and slow enrolment in January to April due to examinations in colleges.
 - 3.3. **Average Enrolment Patterns:** Andhra Pradesh being the top most contributing state has an average enrolment of **13 students** per college from their state with a total of **273 colleges** in **2022** and an average enrolment of **9 students** per college with **326 colleges** in **2023**.
4. **Metrics:**
 - 4.1. Gender wise distribution.
 - 4.2. State wise distribution (Map & Matrix)
 - 4.3. Monthly Trend
 - 4.4. Yearly Trend
 - 4.5. Average Enrolment Card
 - 4.6. Annual Performance for each district



Innovation & Entrepreneurship Hub for Educated Rural Youth (SURE Trust – IERY)

- 4.7. Annual Performance for each college
- 4.8. District and State wise Average enrolment
- 4.9. District and State wise number of distinct colleges
5. **Key Takeaways:** Our data collection aimed to include responses from the AICTE portal, our official website, and Google Forms. However, we received data exclusively from website, leading to a chart that reflects a downward trend not indicative of the overall engagement across all platforms.
 - 5.1. **Problems:** Data submissions from the website exhibited significant quality issues due to incorrect or inconsistent entries from students—such as mistyped state, college name, district, and gender information—which required thorough data cleaning to ensure accuracy in our analysis.
 - 5.2. **Recommendations:**
 - a) To enhance data accuracy from the website, we recommend restricting free-form typing for fields such as qualification, college name, college location (place, district, and state).
 - b) Implementing validation measures—like dropdown menus and auto-complete features—can minimize user errors. These intuitive input validations not only streamline the data entry process but also significantly reduce the time required for data cleaning, leading to more reliable and efficient data analysis.
 - c) To ensure accurate trend analysis, it's crucial to synchronize data collection across all intended sources—the AICTE portal, our website, and Google Forms. Simultaneous data gathering from these platforms will provide a complete picture, reducing inconsistencies and enhancing the reliability of our analysis.
 - d) To enhance data accuracy from the Google Form, we recommend implementing validation measures by creating drop-down lists for fields such as district, state, and gender.
 - e) Additionally, the 'WhatsApp number' field should be configured to accept only 10-digit numeric entries through automated error-checking mechanisms. These enhancements will significantly reduce user input errors, streamline data collection, and minimize the need for extensive data cleaning.



Introduction:

1) **Background and context of the project:**

SURE Trust is an educational organization committed to offering free training in emerging technologies to unemployed rural youth. By focusing on this underserved population, SURE Trust aims to bridge the gap between urban and rural opportunities, empowering young people with the skills needed to thrive in today's technologically advanced society. The organization's dedication to education not only enhances individual livelihoods but also contributes to the overall development of rural communities.

Under the leadership of Executive Director Prof. Radhakumari, SURE Trust has made significant strides in reaching out to rural youth. Prof. Radhakumari's vision and guidance have been instrumental in crafting programs that are both accessible and impactful. Her commitment to education and community development has fostered an environment where students are encouraged to learn, grow, and ultimately transform their futures.

As SURE Trust approaches its Annual Meet, the organization faces a challenge: the lack of quantitative data to effectively showcase its performance and impact over the past years. Without concrete metrics and statistical insights, it becomes difficult to illustrate the successes and areas for improvement to stakeholders, donors, and the broader community. This absence of data hampers the organization's ability to demonstrate progress, secure funding, and plan strategically for the future.

2) **Problem Statement:**

To enhance SURE Trust's performance reporting for the upcoming Annual Meet, we need to conduct a comprehensive analysis of our student enrollment data collected through our website, where students provide details about their educational backgrounds and the courses they wish to pursue. This deep dive aims to extract valuable insights—including state and region-wise student distribution, identification of the universities most students join, comparisons between enrollment numbers and successful completions, and other significant patterns. By addressing data quality issues, performing thorough analyses, identifying appropriate metrics, and creating an interactive dashboard using Power BI, we will transform raw data into actionable insights that inform strategic decisions, showcase our impact, and strengthen our mission to empower unemployed rural youth through free training in emerging technologies.



Innovation & Entrepreneurship Hub for Educated Rural Youth (SURE Trust – IERY)

3) **Goals:**

- a) Generate robust quantitative metrics to enhance SURE Trust's performance understanding.
- b) Dive deep into the student enrollment data to identify significant trends, academic fluctuations, and anomalies.
- c) Analyze the correlation between student enrollments and successful completions to pinpoint factors influencing drop-offs.
- d) Map state and region-wise distribution of students to highlight strong areas and identify regions needing increased outreach.
- e) Create engaging, interactive dashboards using Power BI to present findings in an accessible manner.
- f) Identify and rectify data quality issues during the analysis process.

4) **Scope:**

This project aims to conduct a comprehensive analysis of SURE Trust's student enrollment data from 2022 to 2024. By utilizing Python for data cleaning and Power BI for data visualization, the project seeks to transform raw enrollment data into actionable insights that will inform strategic decisions, enhance performance reporting, and support the organization's mission to empower unemployed rural youth through free training in emerging technologies.

5) **Limitations:**

One significant limitation of this project was the **insufficient amount of data** available for analysis. The limited data volume meant that the insights we derived did not fully align with the expectations and speculations of the director and committee members. This scarcity of data constrained our ability to perform a more comprehensive and in-depth analysis of student enrollment patterns at SURE Trust.

Additionally, the data provided exhibited **quality issues**, including inaccuracies, inconsistencies, and missing information. These problems required us to invest considerable time and effort into data cleaning and validation using Python scripts to enhance data integrity. While this process improved the overall quality of the dataset, it also diverted resources that could have been utilized for more analytical and interpretive tasks.



Innovation & Entrepreneurship Hub for Educated Rural Youth (SURE Trust – IERY)

Project Objectives:

1) Objectives:

- a) Generate robust quantitative metrics to enhance SURE Trust's performance understanding.
- b) Dive deep into the student enrollment data to identify significant trends, academic fluctuations, and anomalies.
- c) Analyze the correlation between student enrollments and successful completions to pinpoint factors influencing drop-offs.
- d) Map state and region-wise distribution of students to highlight strong areas and identify regions needing increased outreach.
- e) Create engaging, interactive dashboards using Power BI to present findings in an accessible manner.
- f) Identify and rectify data quality issues during the analysis process.



Methodology and Results:

1) **Methods/Technology:**

- a) Converting raw CSV files to XLS format for compatibility.
- b) Importing the raw XLS data into Power BI.
- c) Employing Python scripts to clean and preprocess the data without altering the original database.
- d) Applying filters to omit unnecessary information.
- e) Creating essential metrics and KPI's, including qualification-wise totals, yearly and monthly enrolment figures, gender distribution, and state-wise analysis.
- f) Utilizing Power BI's visual tools to create dashboard's that make data insights accessible.

2) **Tools/Software:** Power BI & Python

3) **Project Architecture:**

- a) Data Acquisition and Conversion
- b) Data Processing Layer
 - i) Data Import and ETL Operations
 - ii) Data Cleaning with Python
- c) Data Analysis Layer
 - i) Data Filtering
 - ii) Metric calculation and KPI development
- d) Data Visualization Layer
- e) Insight Generation and Reporting



Learning and Reflection:

1) Overall Experience:

- a) We mastered the effective use of Excel for data cleaning.
- b) We gained hands-on experience with Power-BI and SQL through real-time projects.
- c) We understood the importance of clear and consistent communication within the team to ensure everyone is aligned.
- d) We learned to leverage each other's strengths, distributing tasks efficiently based on individual skills.
- e) We developed proficiency in Data Analysis Expressions (DAX) for creating measures and calculated columns.
- f) We honed the skill of creating impactful visuals that tell a compelling story through data.
- g) We learned to manage deadlines effectively, ensuring the project was completed on time.
- h) We emphasized documenting processes and methodologies for future reference.

2) Document all team member's new learnings and contributions:

	Sparsh Agarwal	Gangadhar	Naveen Kumar
Experience	Learnt the collaborative aspect of the project.	Learnt to clean and organize the data presented a significant challenge.	Learnt the process of cleaning and organizing the data was a significant learning curve.
	Improved his data modeling skills and worked with DAX.	Was responsible for data transformation using Power Query and found it challenging yet fulfilling.	Learnt to document the process and create comprehensive reports.
Contribution	Converted raw CSV files to XLS format.	Worked on creating KPIs.	Made a list of final KPIs for quantitative analysis.
	Imported the raw XLS data into Power BI.	Created M-Language code for automatically cleaning the data	Completed the data cleaning for the (college_name) column.
	Employed Python scripts to clean and preprocess the data.	Helped to create an algorithm for data cleaning.	Helped in creating Python scripts.
	Applied filters to omit unnecessary information.	Utilized Power BI's visual tools to create dashboards.	Provided suggestions to the team related to metrics in Power BI.
	Created essential metrics and KPIs, including qualification-wise totals, yearly and monthly enrolment figures, gender distribution, and state-wise analysis.		



Innovation & Entrepreneurship Hub for Educated Rural Youth (SURE Trust – IERY)

Future Scope:

- 1) **Strategic Decisions:** Our project will allow Radhakumari Mam to take strategic decisions on SURE TRUST student enrollments, such as identifying where student enrollment will drop during 2021-2024.
- 2) **Resource Allocation:**
 - a) Optimizing resource allocation based on enrollment trends and predictions will enable us to effectively manage resources and make data-driven decisions.
 - b) It will help the SURE TRUST management to maintain upcoming student enrollment data in a quality manner.
 - c) By pursuing these future developments, the project can significantly enhance the effectiveness and efficiency of the student enrollment process, ultimately leading to better resource management and data-driven decision-making.