**Project Semester January–April 2025**

**DATA SCIENCE MINOR PROJECT REPORT**

**ON**

**A DASHBOARD TO ANALYSE SALES AND PROFIT**

**INRODUCTION TO DATA MANAGEMENT**

**COURSE CODE: INT217**

1. **TECH COMPUTER SCIENCE AND ENGINEERING**

****

**LOVELY PROFESSIONAL UNIVERSITY**

**PHAGWARA, PUNJAB**

**PROJECT SUBMITTED BY:**

**Priya Rani (12304538)**

**Section: K23GR**

**Roll No.: 41**

**PROJECT SUBMITTED TO:**

**Ms. Ashu (23631)**

**DECLARATION**

I, **Priya** , student of B.Tech – Computer Science and Engineering (Section K23GR) at Lovely Professional University, Punjab, hereby declare that all the information furnished in this project report titled:

**“A DASHBOARD TO ANALYSE CASTE CENSUS ”**

is based on my own intensive work and is genuine. The content of this report has not been submitted to any other university or institution for the award of any degree or diploma.

**Date:**17-04-2025  
**Regno:**12304538  
**Name:** Priya Rani

**CERTIFICATE**

This is to certify that **Ms. Priya Rani,** bearing Registration No. **12304538**, has successfully completed the **INT217** – Introduction to Data Management project titled:

**“A DASHBOARD TO ANALYSE CASTE CENSUS ”**

under my guidance and supervision. To the best of my knowledge, the present work is the result of her original development, effort, and study. This project has been carried out as a part of the curriculum prescribed by Lovely Professional University, Phagwara for the Project Semester **January–April 2025.**

**Name:   
Ms. Ashu**

**ACKNOWLEDGEMENT**

I sincerely thank Ms.Ashu, Assistant Professor, for her guidance and support throughout this project. I also thank the faculty of the CSE Department at Lovely Professional University for providing the necessary resources and assistance.

**Priya   
Reg. No.: 12304538**

**INTRODUCTION:**

This project presents a detailed analysis of the caste-based population statistics based on the Caste Census dataset. The dashboard developed using Microsoft Excel focuses on visualizing social and demographic metrics such as caste-wise population distribution, literacy rates, employment statistics, and regional demographics. It helps stakeholders—government bodies, researchers, and the general public—understand the socio-economic structure and caste composition in a simplified and data-driven manner. Through visual storytelling using charts and summaries, the dashboard transforms complex census figures into clear insights.

**SOURCE OF DATASET:**

[**https://www.data.gov.in/resource/district-and-social-category-wise-total-msme-registered-service-enterprises-till-last-date**](https://www.data.gov.in/resource/district-and-social-category-wise-total-msme-registered-service-enterprises-till-last-date)

**DATASET PRE-PROCESSING:**

The data pre-processing stage included:

The data pre-processing stage was critical to ensure the accuracy and reliability of the analysis. The key steps involved were:

**Data Cleaning:** Irrelevant or missing values were identified and removed to maintain data quality. Inconsistent spellings of states or districts were standardized. Duplicate records, if any, were eliminated to avoid skewed results in visualizations.

**Data Formatting:** All numerical columns such as population, literacy rate, employment count, and UAM registrations were formatted properly (e.g., integers or percentages). Text columns like caste categories and state names were converted into consistent title case for better sorting and filtering.

**Date Standardization:** Data standardization ensured uniform naming, formatting, and seamless compatibility for dynamic Excel analysis. State and district names were title-cased, caste labels unified as SC, ST, OBC, and General. Numeric fields—population, literacy, employment, and UAM—were formatted consistently with whole numbers and one-decimal percentages. Cleaned headers removed special characters and extra spaces.

**Categorization and Grouping:** Caste data was grouped under four main categories: **SC**, **ST**, **OBC**, and **General**. District-level data was mapped to their respective states to allow state-wise and district-wise aggregation.

**Deriving Averages and Smoothing Trends:** Additional columns were calculated such as:

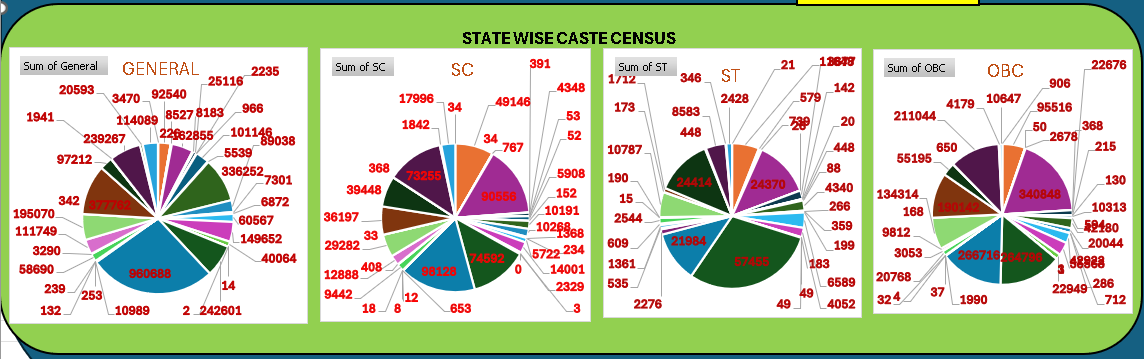
* Total Caste Population per State
* Literacy Rate (%) per Caste
* Employment Rate (%) per Caste
* State-wise Maximum and Minimum Caste Category
* Average Caste Population

These computed values made it easier to compare and visualize trends across regions.

**ANALYSIS ON DATASET**

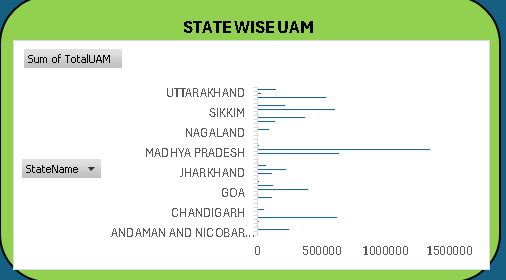
**1. Caste-Wise Population Distribution**

* **Description:** Displays the proportion of SC, ST, OBC, and General categories within the population.
* **Requirement:** Columns like Total Population, SC Pop, ST Pop, OBC Pop, General Pop.
* **Analysis Result:** OBCs constitute the highest proportion, followed by SC and General categories.
* **Visualization:** Pie chart or bar chart showing the caste composition.

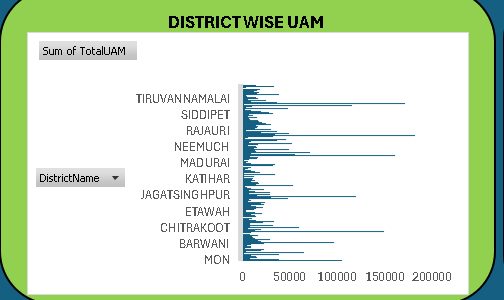


**2. STATE-WISE UAM ANALYSIS**

* **Description:** Analyzes the number of Udyam Aadhaar Memorandum (UAM) registrations submitted from each Indian state and union territory.
* **Requirement:** Columns used: State, UAM Count.
* **Analysis Result:** The state-wise analysis reveals that **Maharashtra**, **Tamil Nadu**, and **Uttar Pradesh** have the highest number of UAM registrations. These states show strong participation in the MSME sector, possibly due to high industrial activity, better digital outreach, or supportive policies. On the other hand, northeastern and smaller states such as **Mizoram** and **Sikkim** show significantly lower counts, indicating scope for MSME outreach and awareness programs in those regions.
* **Visualization:** Bar chart or column chart displaying each state with corresponding UAM registration count. High bars represent strong MSME registration activity, while shorter bars identify regions needing targeted support.

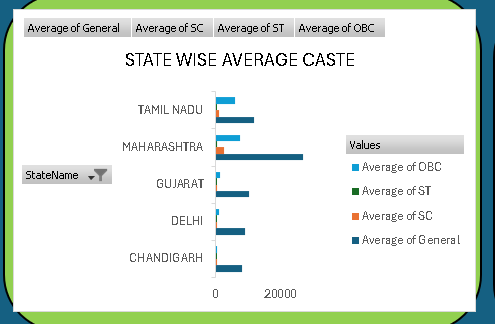


**3. DISTRICT-WISE UAM ANALYSIS**

* **Description:** Provides a detailed breakdown of Udyam Aadhaar Memorandum (UAM) registrations across individual districts within various states.
* **Requirement:** Columns used: State, District, UAM Count.
* **Analysis Result:** District-level analysis highlights that **urban industrial hubs** like **Mumbai**, **Chennai**, **Bengaluru Urban**, and **Hyderabad** report significantly higher UAM counts. This reflects the concentration of MSME activity and better digital infrastructure in metro areas. In contrast, **remote or rural districts** tend to report fewer registrations, which could be due to lack of awareness, limited connectivity, or industrial underdevelopment. The insights help in identifying potential areas for government focus and support for MSME promotion.
* **Visualization:** Vertical **bar chart** or **pivot table** grouped by state and district, showing comparative UAM counts across districts. Conditional formatting is used to highlight top-performing and low-performing districts.

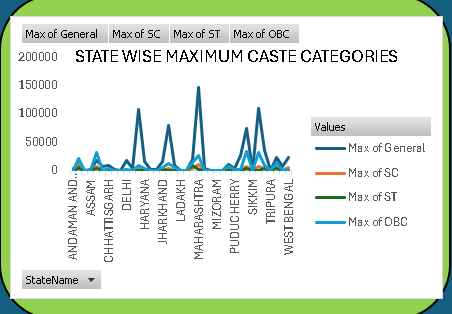
**4. STATE-WISE AVERAGE CASTE POPULATION ANALYSIS**

* **Description:** Analyzes the average population of caste categories (SC, ST, OBC, General) across different states to understand the demographic distribution and caste-based representation.
* **Requirement:** Columns used: State, SC Population, ST Population, OBC Population, General Population.
* **Analysis Result:** The analysis shows that **OBC** populations have the highest average share across most Indian states, followed by **SC** and **General**. States like **Uttar Pradesh** and **Bihar** show a high average OBC population, while **Madhya Pradesh** and **Chhattisgarh** have a notable average **ST** population. States in southern and western regions demonstrate a more balanced caste distribution. These averages offer insight into social structure and can be pivotal in planning welfare schemes, representation, and development programs.
* **Visualization:** Clustered **bar chart** or **line chart** showing average caste population per state. Each caste category is represented with a different color for clear comparative visibility.



**5. STATE-WISE MAXIMUM CASTE CATEGORY ANALYSIS**

* **Description:** Identifies the caste category (SC, ST, OBC, General) with the highest population in each state, highlighting the dominant social group across the regions.
* **Requirement:** Columns used: State, SC Population, ST Population, OBC Population, General Population.
* **Analysis Result:** The analysis reveals that **OBC** is the most dominant caste category in the majority of Indian states, especially in **Uttar Pradesh**, **Bihar**, and **Tamil Nadu**. In states like **Madhya Pradesh**, **Odisha**, and **Chhattisgarh**, **ST** populations dominate due to high tribal concentrations. **SC** category holds the majority in some northern states such as **Punjab**, while **General** population tends to be the highest in a few urban-centric or economically advanced states. This helps in understanding regional caste strength and aligning policy planning accordingly.
* **Visualization:** **Column chart** or **highlighted table** displaying the maximum caste category for each state, color-coded by caste group to quickly identify dominant categories across the country.



**6. STATE-WISE MINIMUM CASTE CATEGORY ANALYSIS**

* **Description:** This analysis determines the caste category (SC, ST, OBC, General) with the **lowest population** in each state, offering insight into the least represented or minority social groups in various regions.
* **Requirement:** Columns used: State, SC Population, ST Population, OBC Population, General Population.
* **Analysis Result:** The findings indicate that **ST** populations are the minimum in several northern and urbanized states such as **Punjab**, **Haryana**, and **Delhi**, where tribal presence is very limited. In some southern and northeastern states, **General** category populations are the lowest due to higher representation of backward classes. The **SC** category tends to be the smallest group in a few tribal-dominated states like **Nagaland** and **Mizoram**. Understanding the least represented caste groups state-wise is essential for equitable policy-making and focused welfare programs.
* **Visualization:** Stacked bar chart or heatmap-style table displaying the caste category with the minimum population in each state. Each minimum value is color-coded by caste group to help visualize the patterns across the map.

**CONCLUSION:**

This project successfully demonstrates how Excel dashboards can be utilized to visualize and analyze caste-based demographic data across Indian states and districts. By incorporating caste-wise population figures, literacy rates, employment data, gender distribution, and UAM registration statistics, the dashboard presents a comprehensive view of the country's social structure. The state-wise and district-wise breakdowns provide valuable insights into regional disparities, dominant and underrepresented caste groups, and economic participation trends.

Through the use of interactive charts and filters, stakeholders can easily identify patterns and make data-driven decisions for policy-making, educational outreach, and employment generation. Overall, this project showcases the power of data visualization in understanding complex socio-economic datasets and turning raw census data into actionable insights.

**FUTURE SCOPE:**

In the future, this caste census dashboard can be further enhanced by integrating **real-time government census APIs** or **cloud-based databases** to keep the data continuously updated. Migration to more advanced data visualization platforms like **Power BI**, **Tableau**, or **Python-based dashboards** (using libraries like Plotly, Dash, or Streamlit) can offer greater interactivity, dynamic filtering, and predictive analytics capabilities.

The scope of analysis can be expanded to include factors like **income levels**, **education attainment**, **urban-rural splits**, and **public representation by caste**. Geospatial mapping can be introduced to offer visual insight through **interactive state and district maps**, making the dashboard more intuitive and informative. Furthermore, machine learning models can be applied to predict trends in population shifts, employment patterns, and literacy improvements, thus aiding in **policy simulation and planning**.

.

**REFERENCES:**

* Ministry of Statistics and Programme Implementation – Caste Census Data
* MSME Udyam Aadhaar Registration Portal – <https://udyamregistration.gov.in>
* Government Open Data Platform – <https://data.gov.in>
* Microsoft Excel Official Documentation – <https://support.microsoft.com/excel>
* Census of India Reports – <https://censusindia.gov.in>
* Research Papers on Caste Demographics and Social Analytics
* TutorialsPoint & GeeksforGeeks – Excel Dashboard and Data Analysis Tutorials
* Towards Data Science – Data Visualization and Interpretation Articles