Analyzing Aadhaar Card Generation Data

Name: Debabrata Mondal

Team No: 19

Batch: 2023

Institute Name: Asutosh College

Project Guide: Prasun Dutta

Report submitted to Institute of Data Engineering, Analytics and Science Foundation, ISI, Kolkata

Abstract

Our internship project focuses on analyzing Aadhaar card generation data downloaded from the Kaggle website. We examine the data state-wise, age-wise and gender-wise to understand the distribution of applications. We have also analyzed the number of applications submitted, the number of rejections, and how many applicants have provided their mobile numbers and email IDs. Using Microsoft Excel 2021, we cleaned and organized the data, then performed calculations using pivot tables and formulas like UNIQUE, COUNT, SUM, and IF-ELSE. We visualize the results through bar diagrams and pie charts. This project aims to provide insights into the Aadhaar enrollment process and highlight key trends and patterns. Our analysis helps us understand regional and gender-based differences, application success rates, and the extent of digital engagement among applicants. Through this project, we contribute valuable information that can assist in improving the Aadhaar registration process.

Introduction

Our project, undertaken at IDEAS-TIH by Debabrata Mondal, focuses on the comprehensive analysis of daily aadhaar card generation data. The data sourced from Kaggle is scrutinized in this area for the state and gender-wise bifurcation, quantity of applications and rejections, and extent of the offering of mobile and email IDs by applicants, among others.

Relevance

- ➢ Aadhaar, a unique identification project in India is very vital in aspects of citizen authentication and delivery of service from the government. The processing of this data helps in the assessment of the demographic specifics and the identification of the areas requiring the intervention of policy.

Technology Involved

- ➢ Data analysis and manipulation are done using the Microsoft Excel 2021 program, which is reliable for the work. Its advanced features facilitate comprehensive data examination and presentation.

Background Material Survey

- ➢ To support our results, we mentioned other demographic investigations, official works, and statistical research that helped us deepen the meaning of the observations made.

Procedure Used

- ➢ The approach taken in this study entails pulling data from Kaggle, preprocessing it in a bid to make it more truthful, and then proceeding to do descriptive analysis and visualization. State-wise we divide the data and separately do the gender-wise analysis of the rejections and applications, and lastly, the contact information provided for the applicants.

Purpose

- ➢ The research's aim is to provide recommendations on trends in the Aadhaar card application to help policymakers and other stakeholders enhance service delivery and meet demography.

By means of this project, we will be able to make a positive significance to the presence of knowledge concerning India's Aadhaar system utilizing data analytics for social advantage.

Project Objective

Analyze State-wise Distribution:

- In this regard, analyze the distribution of the Aadhaar card generation in the different states so as to assess the differences in the enrollment process across the states.

Assess Gender-wise Distribution:

- To get the gender penetration of Aadhaar cards analyze the data according to the gender enrolled and find out the gender-wise gaps.

Evaluate Application and Rejection Rates:

- Determine the overall Aadhaar applications that have been filled and the rejected applications so as to understand the enrollment process and its problems.

Analyze Provision of Contact Information:

- Find out the number of applicants who are giving their mobile numbers and email IDs to know the extent of applicant engagement and accessibility.

Identify Patterns and Anomalies:

- Use the analytical techniques to identify trends and outliers in the data related to the Aadhaar card application, whereby the objective is to highlight peculiarities or important trends that would prompt a closer look.

Enhance Data-driven Strategies:

- To use the findings of the data analysis in identifying the likely approaches to use to enhance the chances of registration for Aadhaar, increase enrollment, and generally seek to make the service delivery process much more efficient and effective to the end consumers.

Thus, successfully obtaining these objectives our project is going to describe the real work of the Aadhaar system, and the issues connected with it and is going to give the potential to enhance the service delivery and to analyze the demographics.

Methodology

This daily information on Aadhaar card generation helps in analyzing demographic characteristics and organizational effectiveness. The project of our study will be the downloading of data from Kaggle and the subsequent analysis of the data Employing Microsoft Excel 2021. In this section, the approach, technique, and instrument by which the project objectives are to be attained are described.

Data Collection

- **Data Source**: The Aadhaar card generation data was collected from Kaggle which is a data source site for datasets.

- **Data Download**: The dataset was obtained from the Kaggle website as a CSV file since it is compatible with Excel applications.

Data Cleaning

- **Data Import**: Subsequently, the information within the CSV file was then pulled into Microsoft Excel 2021.

- **Handling Missing Values**: We have checked for the presence of missing values and dealt with them either by deletion or by replacing them by assuming them as zeros.

Data Analysis

- **Segmentation:** The data was already preprocessed and split into the enrollment agency, state, gender, district, pin code and sub-district dimensions.

- **Application and Rejection Analysis**: The applications as well as the rejections were summed up in a bid to determine the acceptances.

- **Contact Information Analysis:** Using this frequency analysis probability calculation was made to find out the number of applicants giving their mobile numbers and E-Mail IDs.

- **Unique Value Identification:** Unique values for each field (e.g., States, Districts, Pin codes, Enrollment Agencies, Gender etc) were identified to understand the diversity and range within each category.

Tools and Methods Used

- **Pivot Tables**: Pivot tables were widely used for summarizing the data and, thus, performing multi-dimensional analysis.

Formulas:

- **COUNT**: In order to count the number of entries in the different segments of a particular semester, previous years or field of study.
- **SUM**: It will be employed in the computation of the total for opinion, trial, and county.
- **IF-ELSE**: Used to carry out conditional evaluations in relation to jobs and classifications.
- **UNIQUE:** Used to find unique values in each field for a better understanding of data variety.

Distribution Tables:

- Frequency distribution tables were used to prepare the picture of enrollment state-wise or gender-wise etc.

Graphical Visualization:

- **Bar Diagrams**: Used to represent the data in the form of visuals, especially for categorical data.
- **Pie Charts**: Used to indicate the percentage of the various categories.

.

Data Analysis Steps

State-wise Analysis:

- To display the number of enrollments, rejections, and the contact details provided for every state, pivot tables were made.
- Again, for the state-wise distribution, bar diagrams were used.

Gender-wise Analysis:

- Sex-wise totals were done in pivot tables.
- As for the gender distribution, pie charts were generated.
- Unique values for gender were identified to ensure correct categorization.

District Analysis:

- District-level data was also grouped by using pivot tables.
- Unique values for districts were identified to understand the geographic spread.

Pin Code Analysis:

- The analysis was done in relation to the pin code which enabled understanding of the geographical distribution of the enrollments.
- Unique values for pin codes were identified to understand the diversity of locations.

Enrollment Agency Analysis:

- This way, the performance of different enrollment agencies was compared based on the total applications received and handled and the number of rejections.

Application and Rejection Rates:

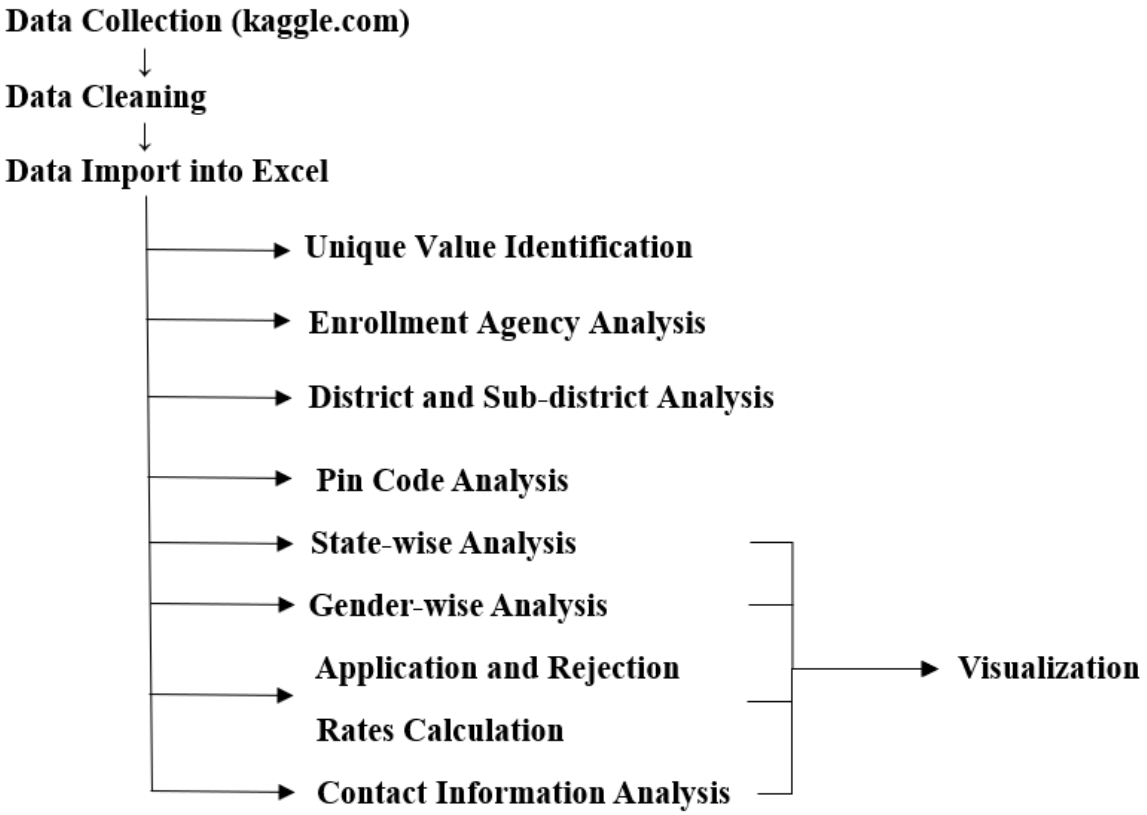
- The total count of the applications was done along with the number of rejections.
- To represent the acceptance and rejection rates bar diagrams were employed.
- Unique values for application statuses were identified to ensure accurate analysis.

Contact Information Analysis:

- The actual count of the respondents affording their mobile numbers and email IDs was found out.
- Specifically to illustrate the percentage of applicants' contact details pie charts and bar diagrams were included.
- Unique values for contact information fields were identified to understand the engagement levels.

Flow Chart of Activities

Below is a flow chart summarizing the activities undertaken during the analysis:

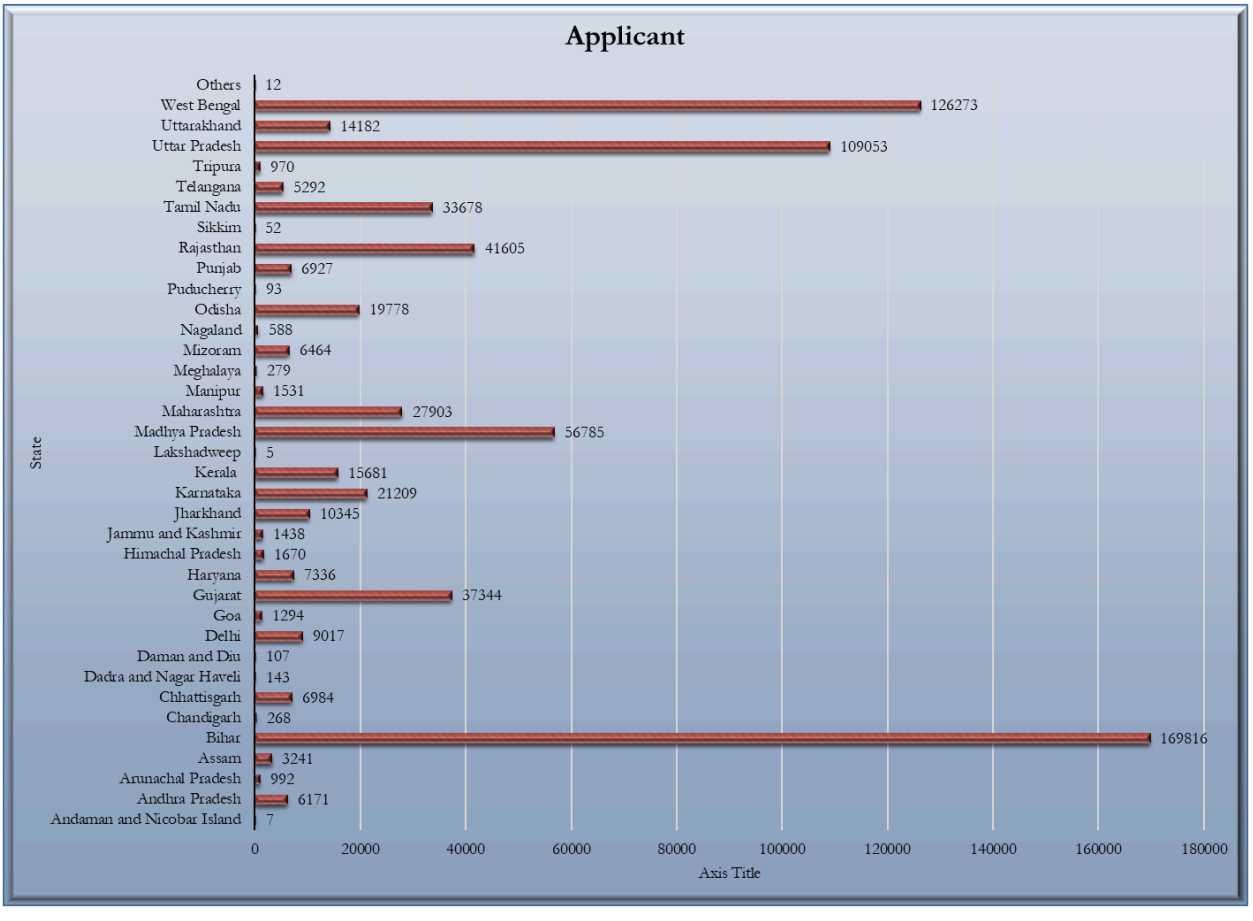


The analysis of Aadhaar card generation data was successfully conducted using Microsoft Excel 2021. By utilizing tools such as pivot tables, formulas, and built-in charts, we were able to perform a comprehensive analysis and derive valuable insights. Identifying unique values for each field allowed us to understand the diversity and range within the data. This methodology can serve as a robust framework for similar data analysis projects in the future.

Data Analysis and Results

➢ State-wise applications:

In our dataset, we analyzed the enrollment data across 37 states. The total number of applicants enrolled amounted to 744,533. This comprehensive figure highlights the extensive reach and participation in the program across the states. To provide a visual representation, we created a bar diagram illustrating the number of applicants from each state. The bar diagram clearly shows the distribution and variation in enrollment numbers across the different states. This visualization helps in identifying trends and disparities in applicant numbers among the states, facilitating a better understanding of the overall enrollment landscape.

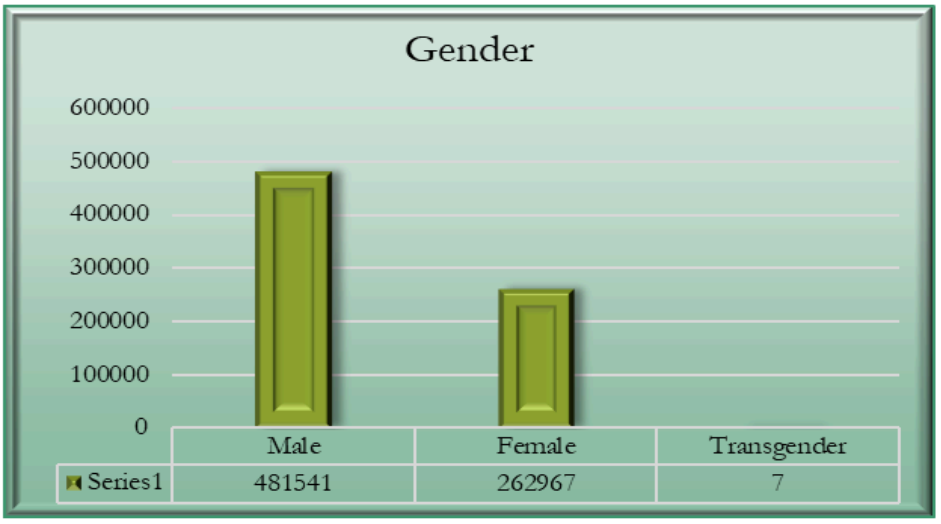


State-Wise Applicants

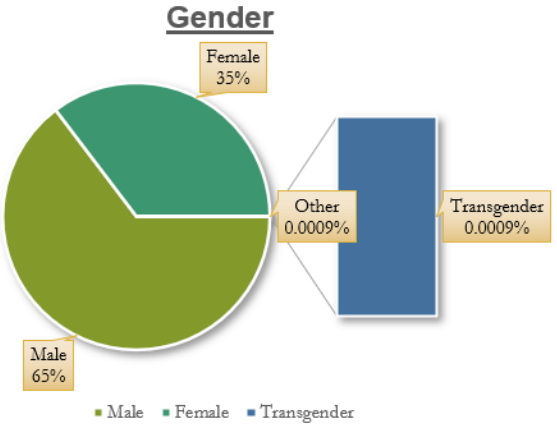
➢ Gender wise applications:

In our dataset, we analyzed the gender distribution of the applicants, categorized as Male, Female, and Transgender. The total number of male applicants is 481,541, while female applicants total 262,967. Additionally, there are 7 applicants identified as Transgender. To visualize this gender distribution, we created a bar diagram. The bar diagram effectively showcases the disparity in application numbers among the three gender categories. This visualization underscores the significant differences in application rates by gender, highlighting areas for potential focus in future outreach and support efforts.

Gender	Total
Male	481541
Female	262967
Transgender	7



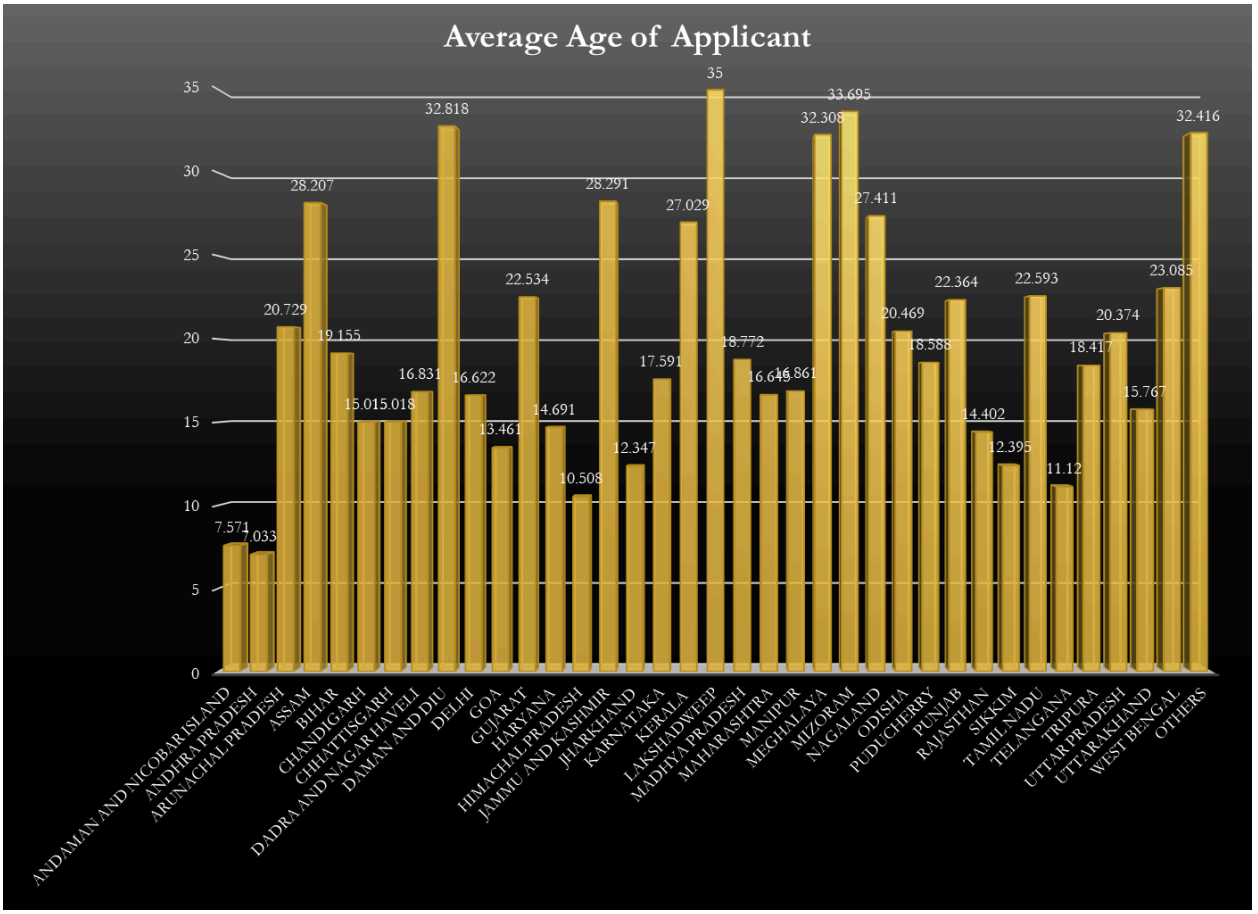
Gender Wise Application (bar diagram)



Gender-wise Application (pie chart)

➢ State wise average age of applicants:

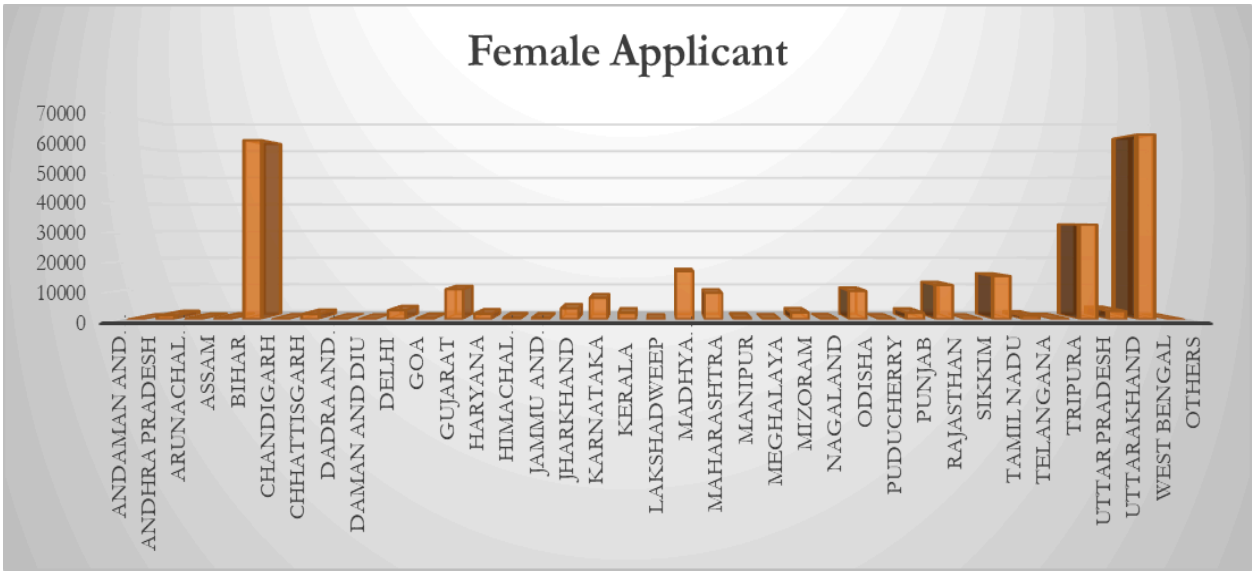
Our dataset includes applicants from various age categories across 37 states. We conducted an analysis to determine the average age of applicants on a state-wise basis. The findings from this analysis were then visualized using a bar diagram. This bar diagram provides a clear representation of the average age of applicants in each state. By examining this visualization, we can observe the variations in average age among the states, which may reflect demographic trends and characteristics unique to each state. This analysis offers valuable insights for tailoring age-specific strategies and interventions.



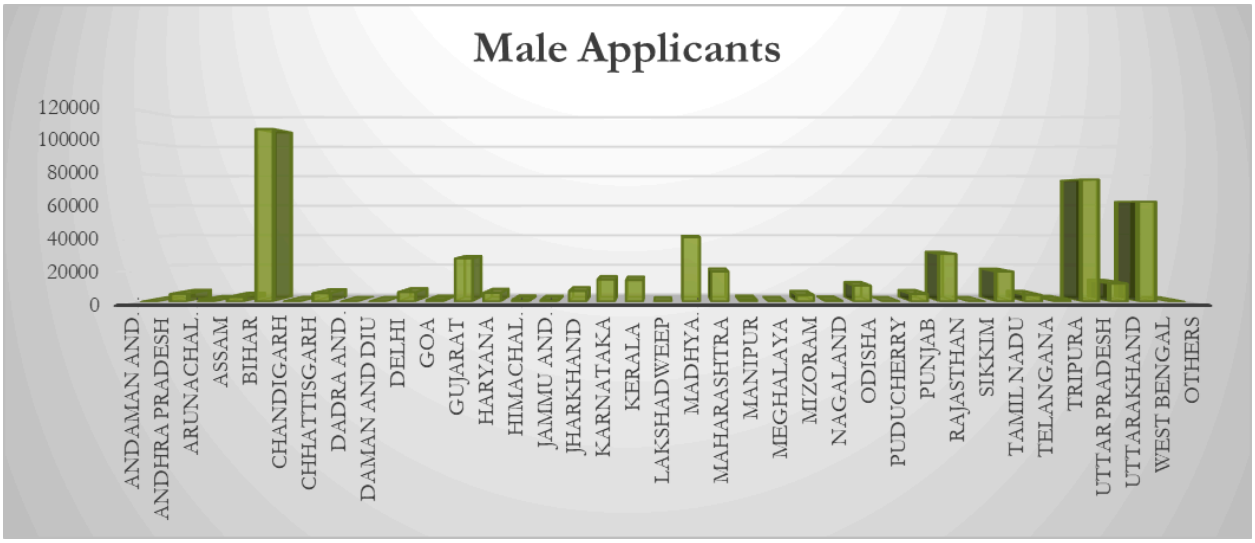
The state-wise average age of applicants

➢ State-wise female and male applicants:

We conducted an analysis of female and male applicants for Aadhaar enrollment across 37 states. The total number of female and male applicants amounts to 262,967 and 481541. Our analysis revealed that West Bengal has the highest number of female applicants, totalling 63,714 and Bihar has the highest number of male applicants, totalling 108030. To illustrate these findings, we created a bar diagram visualizing the state-wise distribution of female and male applicants. This bar diagram highlights the significant concentration of female and male applicants in West Bengal and Bihar compared to other states. The visualization aids in understanding the regional patterns of female and male Aadhaar enrollment.



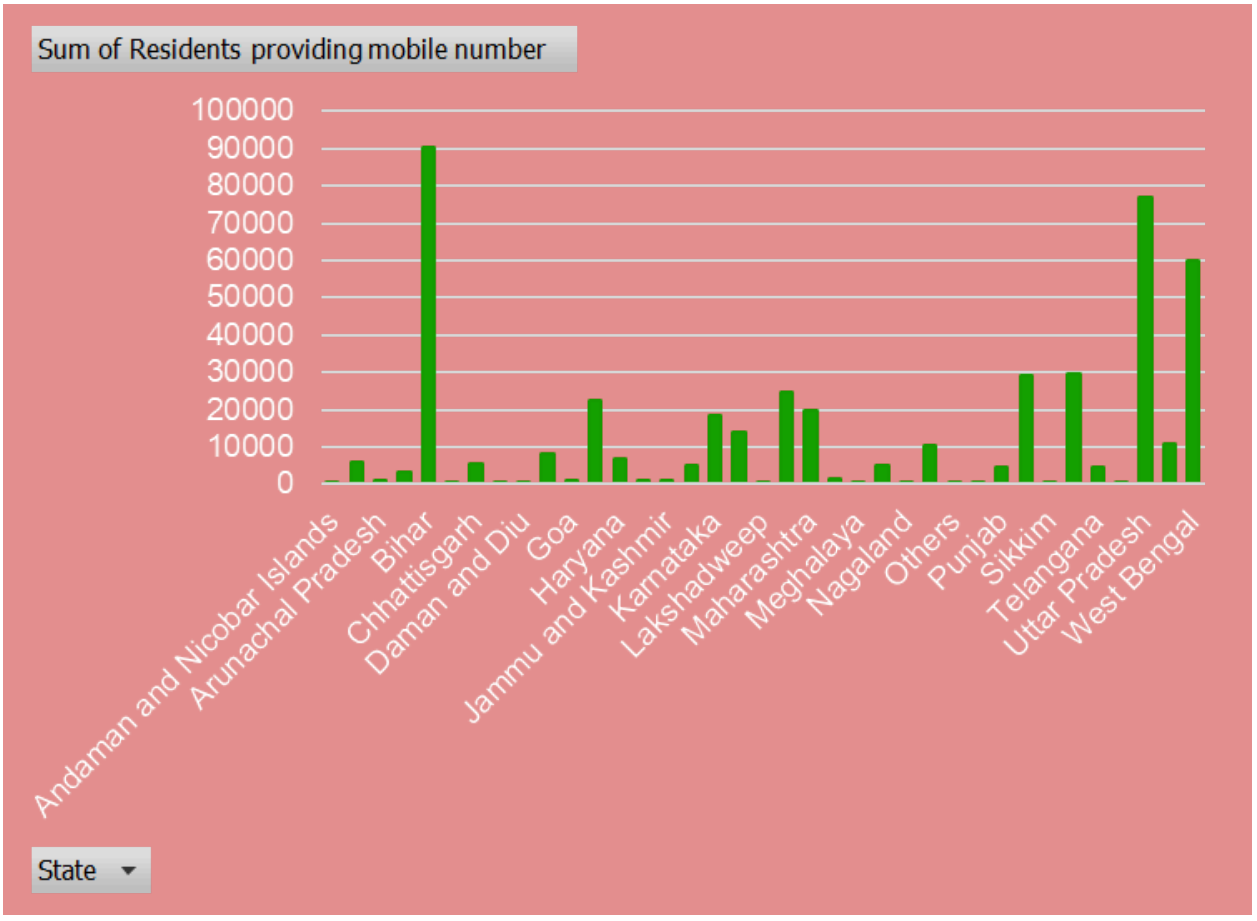
State-wise female applicants



State-wise male applicants

➢ Providing mobile numbers (State-wise):

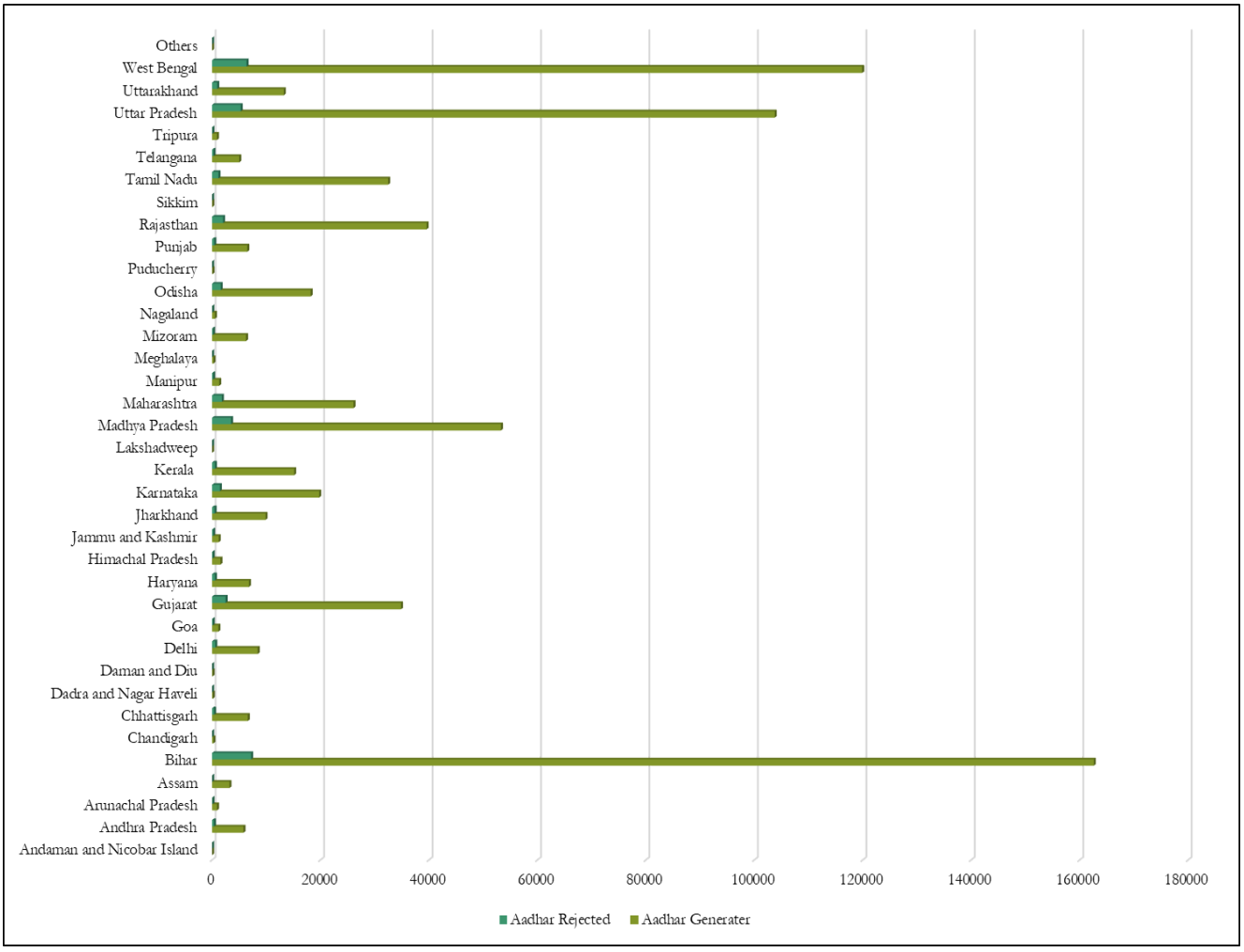
We analyzed the data to determine how many applicants provided their mobile numbers across 37 states. This analysis is crucial for understanding the extent of mobile number registration in the Aadhaar enrollment process. To visualize the findings, we created a bar diagram showing the number of applicants who provided their mobile numbers in each state. The bar diagram clearly illustrates the variation in mobile number submission among the states. This visualization helps identify patterns and trends in mobile number registration, providing insights for improving communication and outreach efforts.



The sum of residents providing mobile numbers

➢ State-wise Aadhaar generated and rejected:

Our analysis focused on Aadhaar enrollment data across 37 states, examining the outcomes of applications in terms of rejections and successful Aadhaar generation. We found that a total of 38,576 applications were rejected, while 705,939 Aadhaar numbers were successfully generated. To provide a detailed state-wise breakdown, we created a bar diagram illustrating the number of rejected and generated Aadhaar applications for each state. This bar diagram highlights the disparities in application outcomes across different states, offering a clear visual representation of the success and rejection rates. This analysis helps in understanding the efficiency of the Aadhaar enrollment process and identifying areas that may require further attention and improvement.



State-wise Aadhaar rejected and generated

Conclusion

For our project on Aadhaar card generation analysis carried out with the help of IDEAS-TIH and under the supervision of our mentor, it was insightful to get some understanding of the situation. State-wise analysis of the data followed by age-wise and gender-wise also depicted certain emerging important trends. For example, some states demonstrated the rejection frequency which is higher than the others, this can indicate some problems at the stage of registration. On analyzing gender they found that the enrollment was relatively equal in both genders and while analyzing with respect to age the younger people responded with mobile numbers and email IDs almost 100%. These were well illustrated by bar diagrams and pie charts for findings with remarkable accuracy.

Such findings may be useful in optimizing the process of getting enrolled under Aadhaar by reducing inter-regional variations and increasing requisite assistance to the users. As for the prospective research, one should extend the study in terms of demographic predictors and apply more sophisticated statistical tools for data analysis. Further, the analysis of upgraded data at various intervals helps in checking the trend and thus, improves the Aadhaar enrollment process a step further.

APPENDICES

References:

Data Source:

- The data used for this project was obtained from Kaggle. The specific dataset used is titled ”Aadhaar Card Data Analysis," which provides detailed information on Aadhaar card applications and rejections, along with other relevant metrics. The dataset can be accessed using the following link: https://www.kaggle.com/code/jindala/aadhar-card-data-analysis

Software and Tools:

Microsoft Excel:

- Used for data cleaning, analysis, and visualization. Key features utilized include pivot tables, formulas (COUNT, UNIQUE, SUM, IF-ELSE), and chart tools (bar diagrams, pie charts).

This project was conducted under the guidance of our mentor at IDEAS-TIH, who provided valuable insights and support throughout the project duration.