

# **Data Insights From Aadhaar: A Comprehensive Analysis Using Qlik**

## **1. Introduction**

### **1.1 Overview**

This project aims to leverage Qlik's powerful data visualization tools to analyze Aadhar data, the world's largest biometric ID system. By acquiring, preparing, and loading data into Qlik Sense or QlikView, the project will create interactive dashboards to uncover trends and insights related to demographics, enrollment, authentication, and update patterns. The analysis will support data-driven decision-making for policymakers, improve operational efficiency, and enhance public awareness through clear visualizations. Key challenges include ensuring data privacy, maintaining data quality, and managing scalability.

### **1.2 Purpose**

The purpose of this project is to utilize Qlik's advanced data visualization and analytical capabilities to derive meaningful insights from the extensive Aadhar dataset. By systematically analyzing this data, the project aims to support informed decision-making, enhance operational efficiencies, and provide valuable information to stakeholders and the public.

### **Achievements**

#### **1. Enhanced Policy Making**

- Provide data-driven insights to policymakers for better governance and strategic planning.
- Identify demographic trends and patterns to inform targeted social and economic policies.

#### **2. Operational Efficiency**

- Pinpoint inefficiencies in enrollment and authentication processes.
  - Offer recommendations for streamlining operations based on data-driven analysis.
- 3. Improved Public Services**
- Ensure effective distribution of public services by understanding demographic needs.
  - Highlight areas needing infrastructure improvement based on Aadhar data insights.
- 4. Public Awareness**
- Create easily understandable visualizations to inform the public about the Aadhar system's reach and impact.
  - Increase transparency by making data insights accessible to the general populace.
- 5. Research and Development**
- Provide a robust data foundation for academic and institutional research.
  - Facilitate the development of new technologies and applications leveraging Aadhar data.
- 6. Security and Fraud Detection**
- Analyze authentication and transaction data to identify and mitigate fraudulent activities.
  - Enhance security measures by understanding and addressing vulnerabilities in the system.
- 7. Resource Allocation**
- Aid in optimal resource allocation by understanding population distribution and needs.
  - Support efficient budgeting and resource planning for government and organizations.

## **1.3 Technical Architecture :-**

### **Data Acquisition and Storage**

The data comes from publicly available Aadhaar datasets provided by UIDAI, encompassing

demographic, enrollment, authentication, and update information. ETL tools like Talend, Apache NiFi, or custom scripts in Python or SQL extract, transform, and load the data into data warehouses such as Amazon Redshift, Google BigQuery, or Snowflake. We clean, transform, and store the data in structured formats, creating subject-specific data marts for focused analysis

## **Integration with Qlik**

We use Qlik Sense/QlikView to connect to the data warehouse, designing an efficient data model in Qlik for optimal performance. Interactive dashboards and advanced analytics are developed in Qlik, with potential integration of R or Python for deeper analysis.

## **Reporting and Distribution**

We generate and share reports and insights with stakeholders through various distribution methods, such as Qlik's web portal, email distribution, or embedding in other applications.

## **Security, Governance, and Performance**

Robust security measures, role-based access controls, and data governance policies ensure data privacy and integrity. We maintain scalability and performance by utilizing scalable cloud infrastructure and regularly optimizing ETL processes, data models, and Qlik applications. The technical stack includes ETL tools (Talend, Apache NiFi), data storage solutions (Amazon S3, Redshift, BigQuery, Snowflake), and the Qlik platform (Qlik Sense, QlikView), along with additional tools for advanced analytics (R, Python).

## 2. Problem Statement

### 2.1 Business Problem

The Aadhar system, being the world's largest biometric ID system, generates vast amounts of data related to demographic information, enrollment processes, authentication transactions, and data updates. However, extracting meaningful insights from this extensive dataset is challenging due to the sheer volume and complexity of the data. Policymakers, public service providers, and other stakeholders often lack the tools and methodologies needed to analyze and interpret this data to make informed decisions.

#### Key Issues:

1. **Data Overload:** The volume of data generated by the Aadhar system is immense, making it difficult to manage, process, and analyze effectively.
2. **Lack of Insightful Analysis:** Without advanced analytics, the raw data does not translate into actionable insights that can guide policy and operational decisions.
3. **Operational Inefficiencies:** Identifying inefficiencies in the enrollment and authentication processes is challenging without detailed, data-driven analysis.
4. **Informed Policy Making:** Policymakers need reliable insights derived from data to formulate effective policies and allocate resources efficiently.
5. **Public Awareness:** There is a need for clear and understandable visualizations to inform the public about the system's reach and impact.
6. **Security and Fraud Detection:** Ensuring the security of the system and detecting fraudulent activities requires sophisticated analysis of authentication and transaction data.

This project aims to address these issues by leveraging Qlik's data visualization and analytics capabilities to transform Aadhar data into actionable insights, thereby supporting informed decision-making, enhancing operational efficiencies, and increasing public awareness.

### 2.2 Business Requirements

## 1. Data Management and Integration

- **Data Acquisition:** Extract data from Aadhar datasets including demographic, enrollment, authentication, and updates.
- **ETL Processes:** Implement robust ETL processes for data extraction, cleaning, transformation, and loading.

## 2. Analytical Capabilities

- **Dashboard Creation:** Develop interactive dashboards in Qlik for visualizing demographic distributions, enrollment trends, authentication insights, and update patterns.
- **Advanced Analytics:** Integrate advanced analytics for deeper statistical analysis, trend forecasting, and correlation.

## 3. Security and Compliance

- **Data Privacy:** Implement security measures to ensure data privacy and compliance.
- **Access Control:** Establish role-based access controls for secure data access.

## 4. Performance and Scalability

- **Scalability:** Ensure system can handle large data volumes and concurrent users.
- **Performance Optimization:** Regularly optimize ETL processes, data models, and applications.

These requirements aim to transform Aadhar data into actionable insights, ensuring security, scalability, and compliance while providing advanced analytical capabilities.

## 2.3 Literature Survey

A literature survey for the Aadhar analysis would involve researching and reviewing previous studies, articles, and reports on the topic. This could include information on the methods and techniques used for analyzing Aadhar Analysis, as well as the results and conclusions of these studies. We recommend exploring academic databases such as

PubMed, IEEE Xplore, Google Scholar, and institutional repositories. Additionally, government reports and publications can provide insights into the latest developments in Aadhar analysis.

### 1. Demographic Trends and Enrollment Patterns

- **Research by UIDAI and Government Reports:** Various reports by UIDAI provide comprehensive data on demographic trends and enrollment patterns, which highlight the coverage and reach of the Aadhar system across different states and regions in India.
- **Scholarly Articles:** Studies such as those by Banerjee and Raju (2016) explore the demographic distribution of Aadhar holders, focusing on age, gender, and regional disparities.

### 2. Authentication Behaviors and Effectiveness

- **Case Studies:** Research conducted by institutions like NIPFP (National Institute of Public Finance and Policy) analyze the effectiveness of Aadhar authentication in reducing identity fraud and enhancing the delivery of public services.
- **Journal Articles:** Articles in journals like the "Journal of Public Affairs" detail the impact of Aadhar on improving access to government services, highlighting success stories and areas needing improvement.

## Business Intelligence and Data Visualization

### a. Title: "Effective Dashboard Design: A Literature Review and Framework"

- **Authors:** Few, S.
- **Publication:** International Journal of Human-Computer Interaction, 2013.
- **Summary:** Reviews literature on effective dashboard design principles and provides a framework for designing interactive dashboards that facilitate data-driven decision-making.

### b. Title: "The Role of Visualization and Visual Analytics in Sense making"

- **Authors:** Yi, J. S., Kang, Y. A., Stasko, J. T., & Jacko, J. A.
- **Publication:** Information Visualization, 2007.
- **Summary:** Discusses the importance of visualization and visual analytics in understanding complex datasets and extracting actionable insights.

### 3 Data Collection

Data contains all the meta information regarding the columns described in the CSV files

Column Description of the Dataset:

1. Registrar: Registrar entities that are responsible for setting up enrollment centers, managing the enrollment process, and collecting necessary data.
2. Enrollment Agency: An Enrollment Agency is responsible for conducting the actual process of enrolling individuals into the Aadhar system
3. State: Indian State
4. District: A district is an administrative division or unit that is usually part of a larger administrative region, such as a state
5. Sub-District: A sub-district, also known as taluka or tehsil in different regions, is a smaller administrative unit that is part of a district.
6. Pin Code: PIN code of the Aadhar card holder
- 7 Gender: Gender of the Aadhar card holder
8. Age: Age of the Aadhar card holder
9. Aadhar generated: No of Aadhar Generated
10. Enrolment Rejected: No of Enrolment Rejected
11. Residents providing email: Whether Email is provided or not
12. Residents providing Mobile: Whether Mobile provided or not

### **3.2 Connet Data with Qlik Sence**

Connect Aadhar data with Qlik Sense, begin with data preparation, ensuring compatibility and accuracy. In the Data Manager, add the data source and configure connection settings, then optimize the generated data model. Customize the data load script to import Aadhar data, using Qlik Script to specify transformation steps. Once loaded, create visualizations in Qlik Sense to analyze demographic trends and enrollment behaviors. Develop interactive dashboards, combining multiple visualizations for dynamic exploration. Schedule regular data refreshes to maintain up-to-date insights. This seamless integration empowers users to extract actionable insights from Aadhar data, aiding informed decision-making and strategic planning.

## **4. Data Preparation**

Preparing the data for visualization involves cleaning the data to remove irrelevant or missing data, transforming the data into a format that can be easily visualized, exploring the data to identify patterns and trends, filtering the data to focus on specific subsets of data, preparing the data for visualization software, and ensuring the data is accurate and complete. This process helps to make the data easily understandable and ready for creating visualizations to gain insights into the performance and efficiency. Since the data is already cleaned, we can move to visualization.

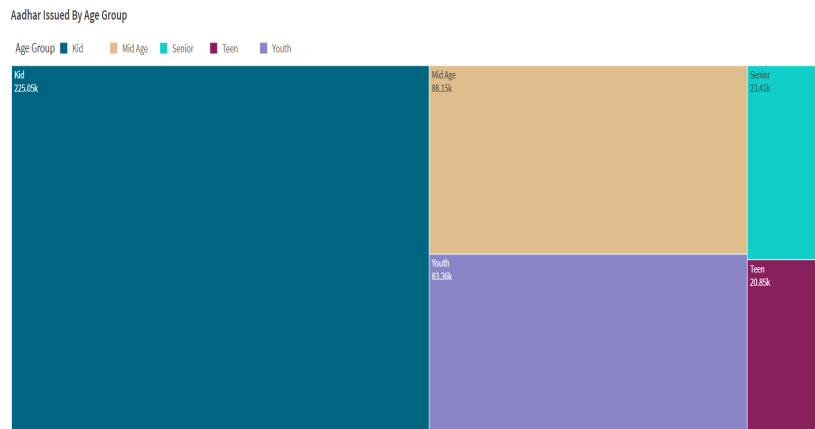


## 5. Data Visualization

### 1. Tree Map

"This Tree Map represents the total Aadhar numbers generated, categorized by age group. The dimension is Age Group, and the measure is the sum of Aadhar numbers generated."

In Qlik Sense, we use dimensions to categorize the data and measures for numerical calculations or aggregations.

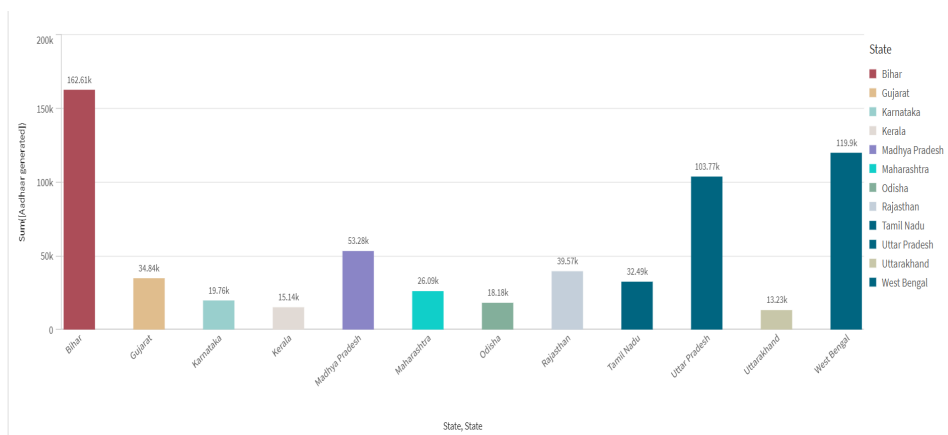


### 2. Bar Chart

"This bar chart represents the total Aadhar generated and rejected by states. To achieve this, we first create a hierarchy dimension consisting of State, District, and Sub District. The measure for the bar chart is the sum of Aadhar generated and rejected."

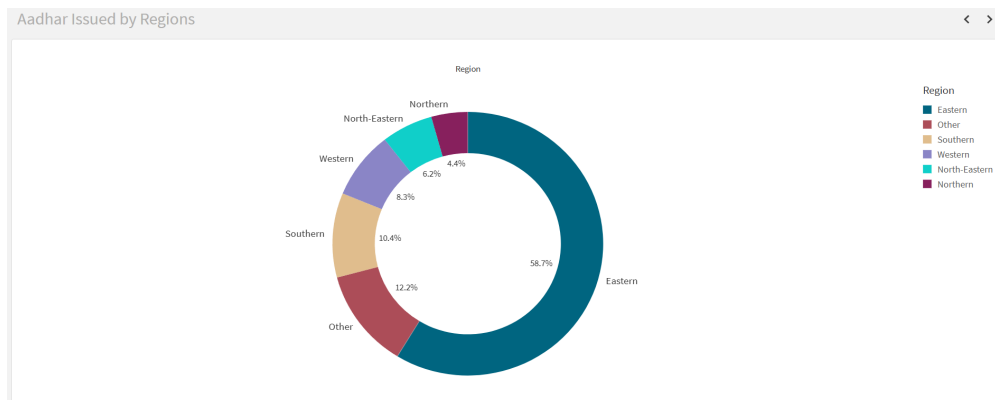
To be clear:

- **Hierarchy Dimension:** State > District > Sub District
- **Measures:** Sum of Aadhar generated and Sum of Aadhar rejected



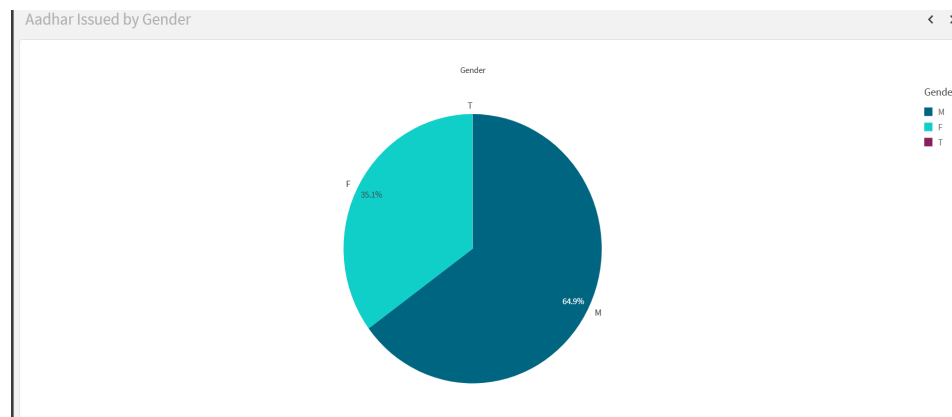
### 3. Donut Chart

"This donut chart represents the Aadhar numbers issued by region. The dimension used is Region, and the measure is the sum of Aadhar generated."



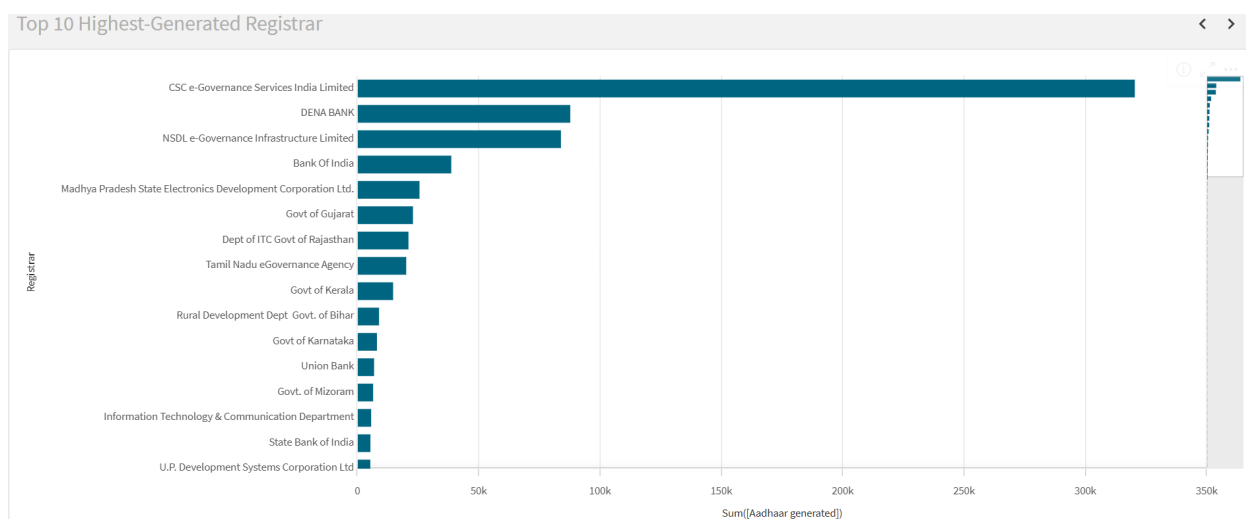
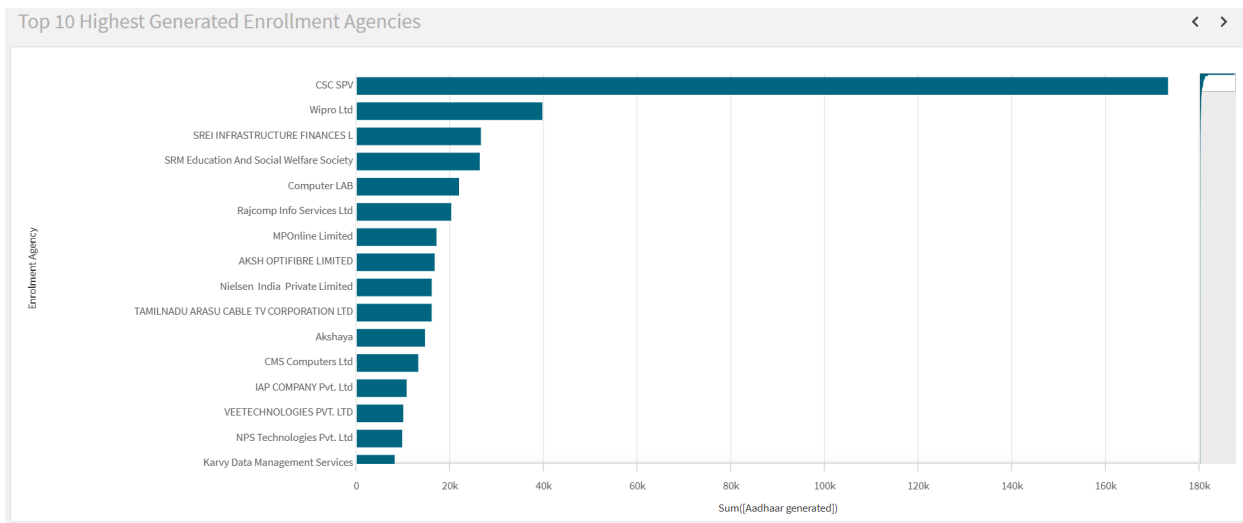
### 4. Pie Chart

"This pie chart represents the total number of Aadhar generated by male and female (Gender). The dimension is Gender, and the measure is the sum of Aadhar generated."



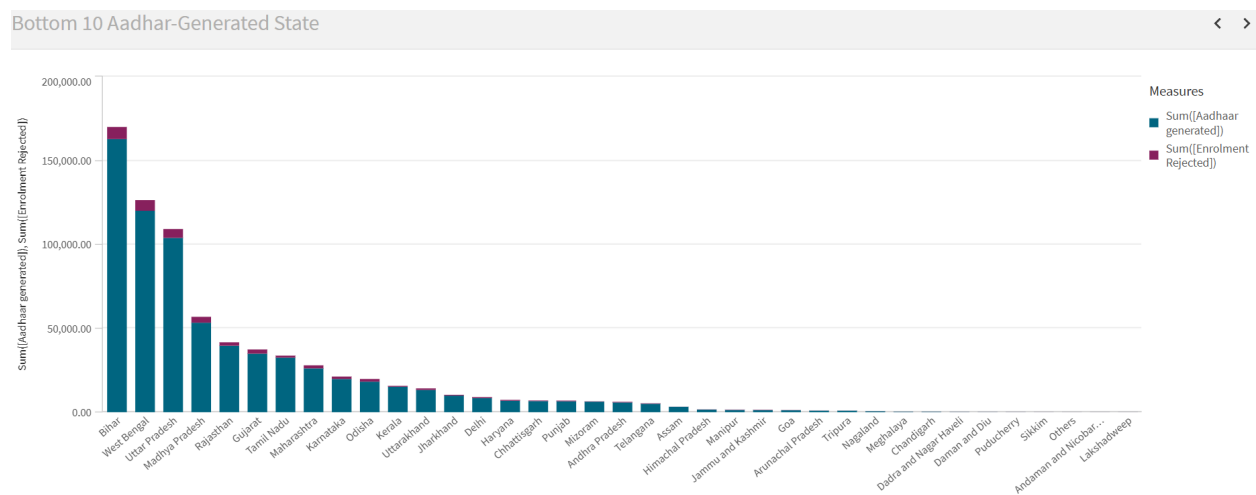
### 5. Horizontal Bar Charts

This horizontal bar chart represents the total number of Aadhar generated by the top 10 enrollment agencies. The dimension is Enrollment Agency, and the measure is the sum of Aadhar generated," and the total number of Aadhar generated by the top 10 Registrars. The dimension is Registrar, and the measure is the sum of Aadhar generated."



## 6. Stacked Bar Chart

"This stacked bar chart represents the total number of Aadhar generated and rejected by states. The dimension is State, and the measures are the sum of Aadhar generated and the sum of Aadhar rejected."



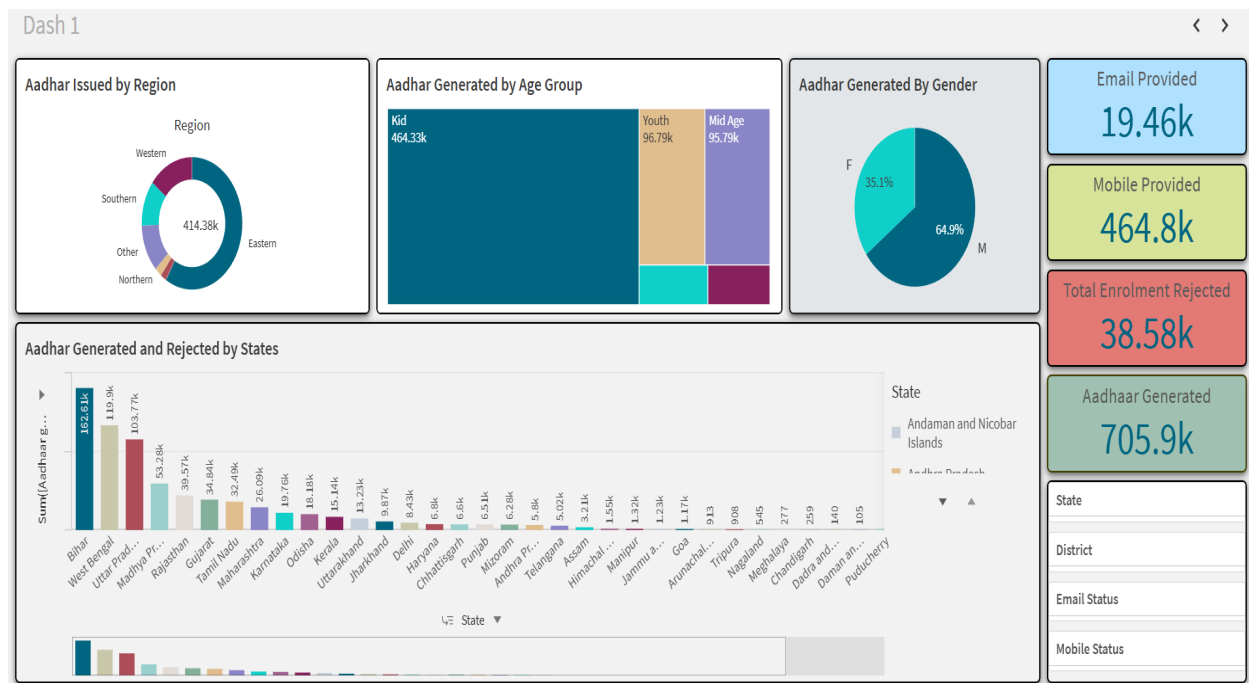
## 6. Dashboard

A dashboard is a graphical user interface (GUI) that displays information and data in an organized, easy-to-read format. Dashboards are often used to provide real-time monitoring and analysis of data and are typically designed for a specific purpose or use case. Dashboards can be utilized in a variety of settings, including business, finance, manufacturing, healthcare, and many other industries. They can be used to track key performance indicators (KPIs), monitor performance metrics, and display data in the form of charts, graphs, and tables..

### Dashboard 1

This dashboard provides a comprehensive view of Aadhar numbers generated and rejected across various categories:

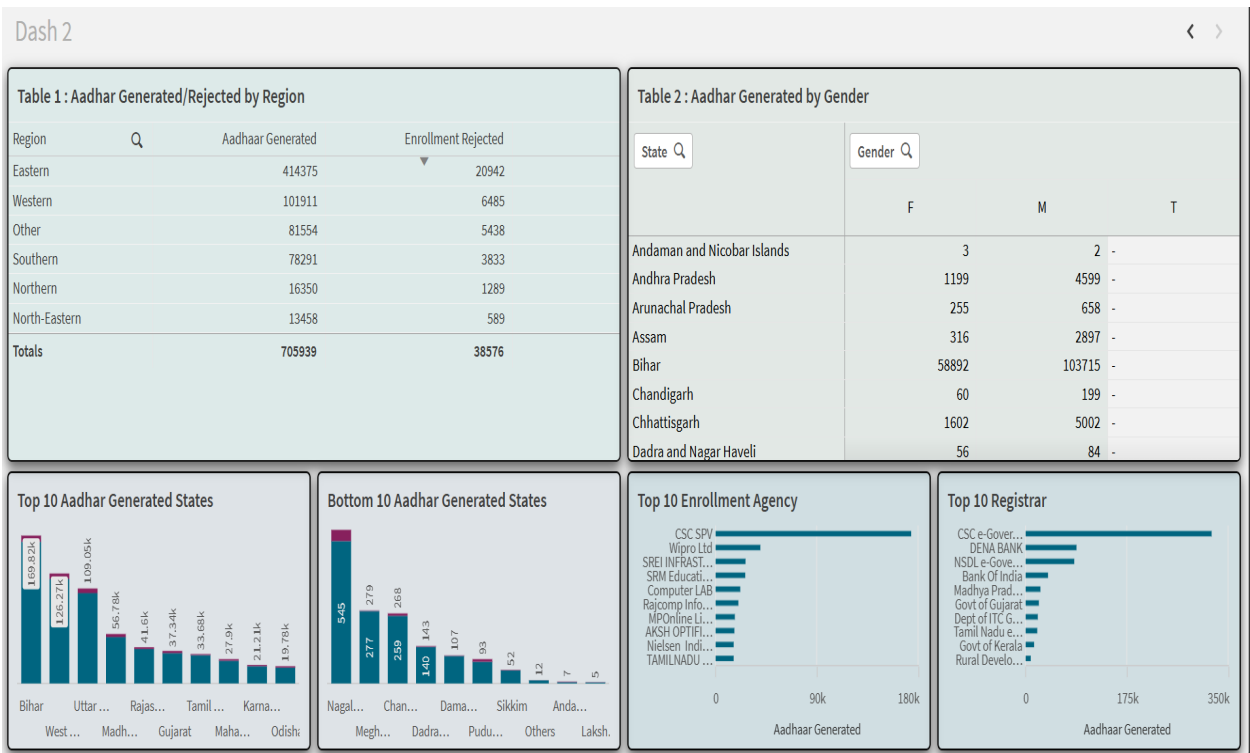
- **States and Regions:** Visualizes the distribution of Aadhar generation and rejections across different states and regions.
- **Sub districts:** Breaks down the data further to sub district levels for more granular insights.
- **Mobile and Email:** Displays statistics on Aadhar numbers linked to mobile numbers and email addresses, highlighting trends and patterns in digital association.



Dashboard 2

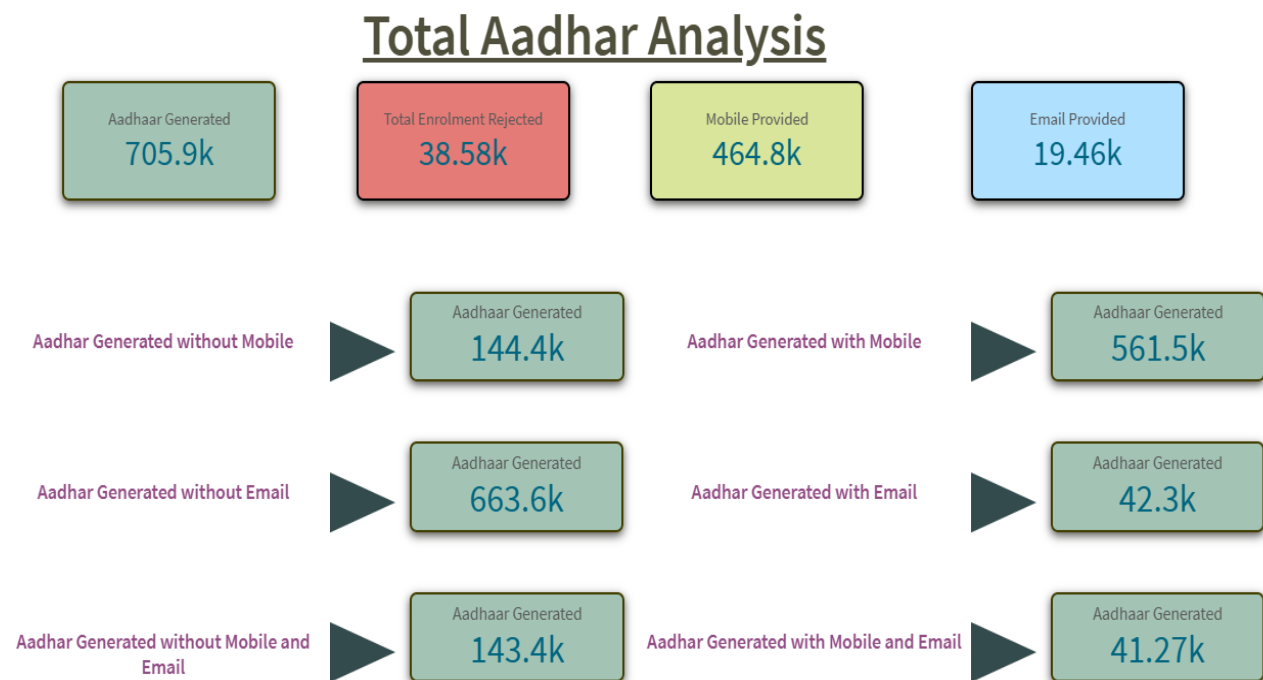
This dashboard focuses on detailed tabular and graphical representations:

- **Table 1 and Table 2:** Present the total Aadhar numbers generated and rejected by regions, offering a clear comparison and analysis.
- **Pivot Table by Gender:** Shows the total Aadhar numbers generated, segmented by gender, providing insights into gender-based distribution.
- **Horizontal Bar Charts:**
  - **Top 10 and Bottom 10 States:** Displays the states with the highest and lowest Aadhar generation, facilitating a quick performance review.
  - **Top 10 Enrollment Agencies and Registrars:** Highlights the most active enrollment agencies and registrars, showcasing their contribution to Aadhar generation.



## 7. Story

"Total Aadhar Analysis" provides a comprehensive overview of Aadhar generation and enrollment rejection statistics. It shows that a total of 705.9k Aadhar numbers have been generated, with 38.58k enrollments rejected. Out of the total generated, 464.8k Aadhar numbers are associated with provided mobile numbers, and 19.46k with provided email addresses. Additionally, the data is further broken down into specific categories: 144.4k Aadhar numbers were generated without a mobile number, whereas 561.5k included a mobile number; 663.6k were generated without an email, while 42.3k included an email. Finally, 143.4k Aadhar numbers were generated without both mobile and email, and 41.27k included both mobile and email. This detailed breakdown helps in understanding the distribution and linkage of Aadhar numbers with mobile and email data.



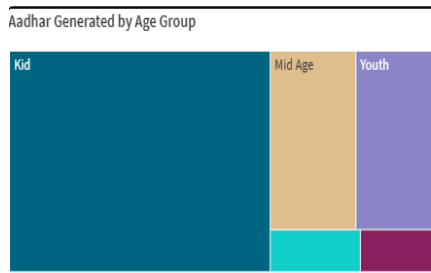
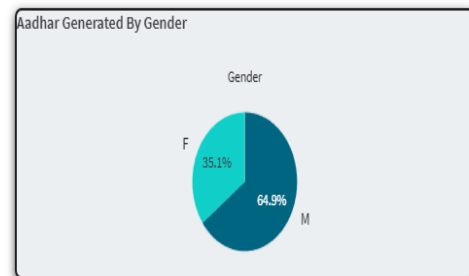
The statistical data on Aadhar card issuance in India, broken down by gender and age groups. It reveals a significant gender disparity, with 65.19% of Aadhar cards being issued to males and 34.81% to females. The total numbers stand at 458.2k for males and 243.73k for females.

Additionally, the data is categorized by age groups, indicating the distribution of Aadhar cards among kids, adults, and seniors. A specific focus is given to the female demographic, with a separate bar graph detailing the total Aadhar cards generated for different female age groups. This visual data representation underscores the gender and age-related trends in Aadhar card issuance across the country.

## Aadhar Generated by Gender

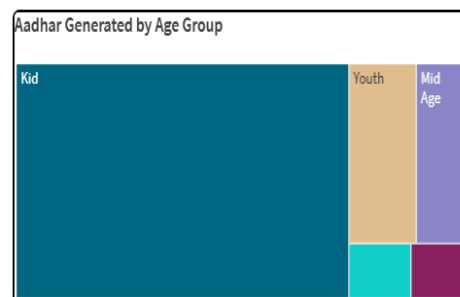
Total Aadhar Issued Male: 458.2k

Total Aadhar Issued Female: 243.73k



Total Aadhar Generated for the Different Age Group for Female Only

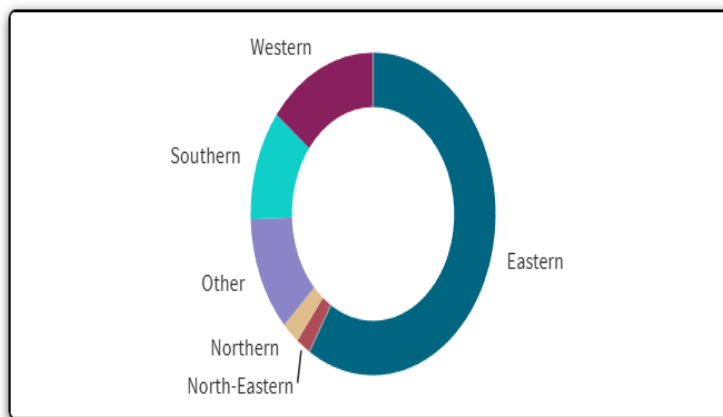
Total Aadhar Generated for the Different Age Group for Male Only



The donut chart with data on Aadhar card generation and rejection across various regions in India. The chart is segmented to reflect different regions, each with two color codes representing the number of Aadhar cards generated and rejected. For instance, the Northern region shows 16.35K generated and 1.2K rejected, while the Eastern region has a significantly higher number of 414.4K generated and 20.94K rejected. The Northeastern region has 13.46K generated and 589 rejected, the Western region has 101.9K generated and 6.49K rejected, and the Southern region has 78.29K generated and 3.83K rejected. This visual data highlights the

efficiency and challenges in the Aadhar card generation process across India, pinpointing regions with higher or lower rejection rates, which may indicate the success or issues within the system.

## Aadhar Generated and Rejected by Region



✓ Northern: 16.35K Generated and 1.2K Rejected

✓ Eastern: 414.4K Generated and 20.94K Rejected

✓ Northeastern: 13.46K Generated and 589 Rejected

✓ Western: 101.9K Generated and 6.49K Rejected

✓ Other: 81.55K Generated and 5.44K Rejected

✓ Southern: 78.29K Generated and 3.83K Rejected

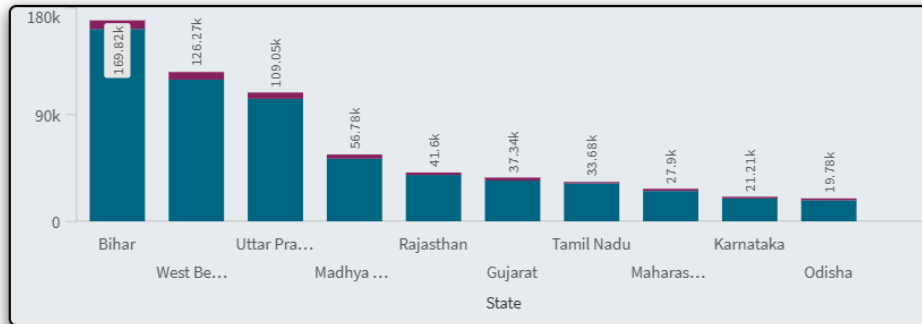
"Top 10 Aadhar Generated States" showcases the distribution of Aadhar registrations across various states in India.

- Bihar leads with the highest Aadhar registrations at 2,162,067.
- West Bengal follows with 1,199,011 Aadhar cards generated.
- Uttar Pradesh has a total of 1,037,67 Aadhar registrations.
- The chart includes seven other states, each with varying numbers of registrations, contributing to the overall statistics of Aadhar generation in India.

This data paints a clear picture of how Aadhar registrations are spread across different states, with Bihar emerging as the front runner in terms of numbers.



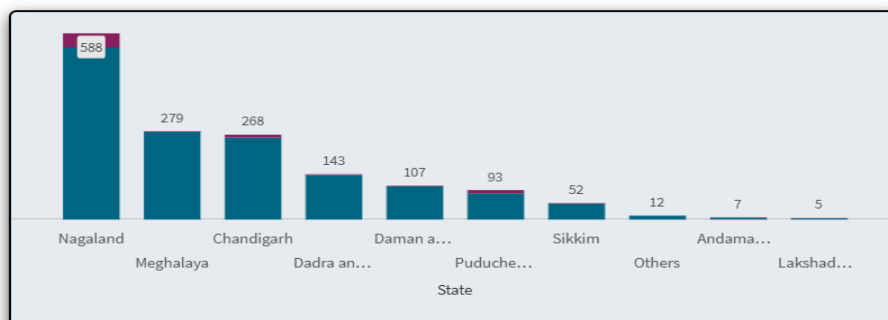
## Top 10 Aadhar Generated States



- ✓ Bihar, an Indian state, has achieved the highest number of Aadhaar registrations, generating a total of 162,607 Aadhaar cards.
- ✓ West Bengal is 2nd Highest Aadhaar Registration State generating total of 119,901 Aadhaar cards.
- ✓ Uttar Pradesh is 3rd Highest Aadhaar Registration State generating total of 103,767 Aadhaar cards.

"Bottom 10 Aadhar Generated States," outlining the states in India with the lowest number of Aadhar card registrations. Lakshadweep records the lowest count with only 3 Aadhar cards issued, followed by Arunachal Pradesh with 5, and Andaman and Nicobar Islands with 12. Additionally, the list includes Nagaland, Meghalaya, Chandigarh, Daman and Diu, Puducherry, and Sikkim, all exhibiting relatively low numbers of Aadhar cards generated. This dataset sheds light on the varying degrees of Aadhar card adoption across different regions in India, showcasing these ten states with the smallest uptake.

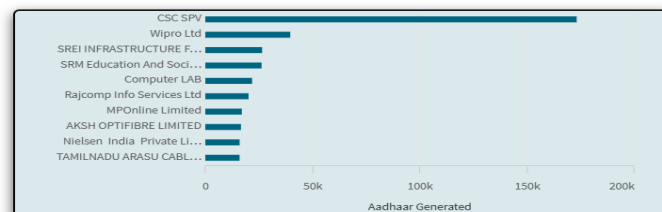
## Bottom 10 Aadhar Generated States



- ✓ Among the states and union territories, Lakshadweep has the lowest Aadhaar card generation count, with only 5 Aadhaar cards issued.
- ✓ Andaman and Nicobar islands ranks as the second-lowest state in terms of Aadhaar card generation, with a total count of only 7 Aadhaar cards issued.

"Top 10 Enrolment Agency," which illustrates the percentage of Aadhar registrations handled by different agencies in India. The leading agency is CSC SPV, responsible for 50% of the registrations, followed by Wipro Ltd. and SREI Infrastructure Finance Ltd. with 39.62K and 26.5K Aadhar cards generated, respectively. This chart is a visual representation of the distribution of Aadhar card registrations among various agencies, highlighting the significant role of CSC SPV in the process.

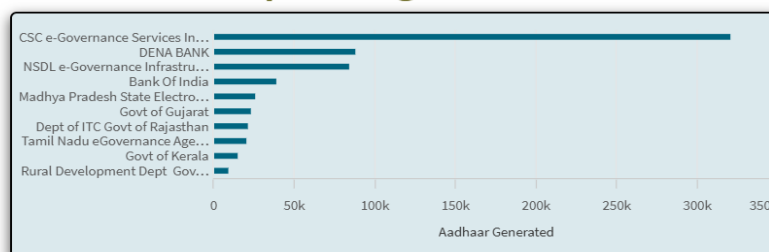
### Top 10 Enrolment Agency



- ✓ CSC SPV has achieved the highest number of Aadhaar registrations, generating a total of 173.19K Aadhar cards.
- ✓ Wipro Ltd is 2nd Highest Aadhaar Registration registrar generating total of 39.62K Aadhar cards.
- ✓ SRE Infrastructure Finances Ltd is 3rd Highest Aadhaar Registration registrar generating total of 26.5K Aadhar cards.

The "Top 10 Registrar" chart delineates the quantity of Aadhar registrations overseen by different organizations in India. CSC e-Governance Services India Ltd emerges as the foremost registrar, commanding a substantial lead with 320.32K Aadhar cards registered. Notably, NSDL e-Governance Infrastructure Ltd secures the third position with 87.55K registrations. This graphical portrayal accentuates the proactive involvement of these entities in the Aadhar registration endeavor, underscoring their pivotal contributions to this nationwide initiative.

### Top 10 Registrar



- ✓ CSC e- Governance Services India Ltd has achieved the highest number of Aadhaar registrations, generating a total of 320.2K Aadhar cards.
- ✓ Dena bank is 2nd Highest Aadhaar Registration Registrar with 87.55K Aadhar cards
- ✓ NSDL e-Governance Infrastructure Ltd is 3rd Highest Aadhaar Registration registrar generating total of 83.71K Aadhar cards.

## 8. Performance Testing

### 8.1 Amount of Data Rendered

"Amount of Data Loaded" refers to the quantity or volume of data that has been imported, retrieved, or loaded into a system, software application, database, or any other data storage or processing environment. It is a measure of how much data has been successfully processed and made available for analysis, manipulation, or use within the system.

### 8.2 Utilization of Data Filters

Utilization of data filters refers to the process of applying specific criteria or conditions to a dataset in order to include selectively or exclude certain data points. This filtering process is crucial in data analysis as it allows to focus on relevant subsets of data, eliminating noise and irrelevant information.

#### Calculation Fields

The term "calculation fields" typically refers to the variables in a dataset that have been generated through calculations rather than being directly obtained from the source data. These fields are derived by applying mathematical operations, functions, or formulas to existing data within the dataset.

#### Age Group:

```
if(Age >= 60, 'Senior',  
  
    if(Age >= 30, 'Mid Age',  
  
        if(Age >= 18, 'Youth',  
  
            if(Age >= 14, 'Teen','Kid'  
            )))
```

**Region:**

if(Match(State, 'Maharashtra', 'Gujarat', 'Rajasthan', 'Goa', 'Daman and Diu', 'Dadra and Nagar Haveli'), 'Western',

if(Match(State, 'Uttar Pradesh', 'Bihar', 'Jharkhand', 'Odisha', 'West Bengal', 'Sikkim'), 'Eastern',

if(Match(State, 'Karnataka', 'Andhra Pradesh', 'Telangana', 'Tamil Nadu', 'Kerala', 'Puducherry'), 'Southern',

if(Match(State, 'Punjab', 'Haryana', 'Himachal Pradesh', 'Jammu and Kashmir', 'Chandigarh'), 'Northern',

if(Match(State, 'Assam', 'Arunachal Pradesh', 'Nagaland', 'Manipur', 'Mizoram', 'Tripura', 'Meghalaya', 'Sikkim'), 'North-Eastern', 'Other'))))

**Project Demo Link**

[https://drive.google.com/file/d/1DyL1RqDR-UIVWw\\_J8ELUoOnGp4pFDeuC/view?usp=drive\\_link](https://drive.google.com/file/d/1DyL1RqDR-UIVWw_J8ELUoOnGp4pFDeuC/view?usp=drive_link)