

# **Data Insights from Aadhaar a Comprehensive Analysis using Qlik**

## **1.Introduction:**

### **1.1 Overview:**

This project involves extensive analysis of base data using Qlik Sense, with a focus on generating actionable insights. Activities include cleaning and modelling base datasets, creating interactive Qlik Sense dashboard reports, and extracting key visualizations such as population projections, generation/rejection, and geographic analysis.

### **1.2 Purpose:**

The primary objective of this project is to conduct a thorough analysis of Aadhaar data to enhance decision-making, policy formulation, and operational efficiency within the National Identity Authority. This analysis aims to provide valuable insights into user demographics, authentication trends, and compliance metrics for informed decision-making.

### **1.3 Technical Architecture:**

The project utilizes Qlik Sense for data visualization and dashboard creation. The Aadhaar dataset, comprising demographic information, authentication records, and geographical details, serves as the primary data source. The architecture involves data collection, preparation, visualization, and dashboard design to support strategic planning and operational improvements.

## **2.Problem Understanding:**

### **2.1 Business Problem:**

Aadhaar card is a unique 12-digit number issued by the Unique Identification Authority of India (UIDAI), the legal authority set up by the Government of India. The Aadhaar scheme aims to provide universal and robust identification to Indian residents. The plan seeks to simplify identity management procedures, enforce transparency, and facilitate the delivery of government services and benefits.

The primary data is largely analysed using Qlik Sense, with a focus on deriving actionable insights. Activities include cleaning and modelling base datasets, generating interactive Qlik Sense dashboard reports, and extracting key visualizations such as demographics, generation/rejection, and geospatial analysis. These visualizations sit with base user distributions characteristics, and also helps identify trends and patterns in data.

The primary data source is an extensive base database, which includes demographic data, certification records, and geographical descriptions. The objective of the project is to deeply analyze Aadhaar data

using Qlik Sense to extract valuable insights that can enhance decision-making, policy and operational efficiency in the National Identification Authority. This research aims to support policy development and improve service delivery, and ultimately contribute to effective governance and better resource allocation.

## 2.2 Business Requirements:

This report delves deeper into user demographics, certification trends, and compliance metrics through comprehensive data visualization. The goal is to create interactive and visually appealing dashboards that will facilitate planning and business improvement.

By examining user profiles, loyalty practices, and compliance, the report aims to:

**Informed decision making:** Using data-driven insights to guide strategic decisions in product development, marketing strategy, and user experience improvement.

**Enhanced service delivery:** Identify and address inefficiencies in the certification process to ensure a seamless and effective user experience.

**Remain compliant:** Observe and visualize compliance to demonstrate compliance and maintain trust. This report leverages the power of data visualization to transform unstructured data into actionable insights, ultimately supporting process implementation, operational efficiency and stronger compliance.

### Social Impact Analysis

- **Demographic Distribution:** Visualizations will be created to illustrate the demographic distribution of Aadhaar users, providing a clear picture of the population segments that have enrolled in the Aadhaar program.
- **Social Welfare Programs:** The analysis will examine the impact of Aadhaar on various social welfare programs, highlighting how it has enhanced the delivery and effectiveness of these initiatives.
- **Financial Inclusion:** Insights will be drawn on how Aadhaar has facilitated financial inclusion, particularly in enabling access to banking services and financial products for underserved populations.
- **Socioeconomic Indicators:** The report will explore correlations between Aadhaar usage and improvements in socioeconomic indicators such as education, health, and employment, demonstrating its broader social benefits.

### Business Impact Analysis

- **Sectoral Impact:** The analysis will focus on how Aadhaar has influenced different business sectors, particularly banking, telecommunications, and e-commerce, assessing its role in driving growth and innovation.
- **Fraud Prevention:** Evaluating the effectiveness of Aadhaar in preventing fraud, the report will highlight its contributions to reducing identity theft and ensuring secure transactions.

- **Customer Onboarding and Operational Efficiency:** The impact of Aadhaar on customer onboarding processes and operational efficiency will be assessed, showcasing improvements in speed, accuracy, and overall user experience in business operations.

### 3. Data Collection:

#### 3.1 Dataset:

The data is downloaded from the given link:

<https://drive.google.com/file/d/1dShIZsdyZKNANLyqo1DZyK66us3SJ28V/view>

**Understand the Data and features involved:**

1. **Registrar:** Entities responsible for enrollment centers and data collection.
2. **Enrollment Agency:** Agencies conducting the enrollment process.
3. **State:** Indian State.
4. **District:** Administrative division within a state.
5. **Sub-District:** Smaller administrative unit within a district.
6. **Pin Code:** Postal code of the Aadhaar card holder.
7. **Gender:** Gender of the Aadhaar card holder.
8. **Age:** Age of the Aadhaar card holder.
9. **Aadhaar generated:** Number of Aadhaar cards generated.
10. **Enrolment Rejected:** Number of enrollments rejected.
11. **Residents providing email:** Indicator of whether email is provided.
12. **Residents providing Mobile:** Indicator of whether mobile number is provided.

#### 3.2 Data Preparation:

Data preparation is an important step in business analysis, ensuring that raw data is transformed into a clean and usable form for analysis. This process ensures data accuracy, completeness, and relevance, and provides a reliable basis for actionable insights. Below is an overview of the data that is created in business analytics:

1. **Data Collection:** Integrate raw data from various sources, incl.

Databases, spreadsheets, APIs, sensors, social media and more.

2. **Data correction:** error detection and correction; inconsistent and missing values

They are right at this stage. Typical tasks include:

- Remove duplicate records.
- Filling in gaps by insertion or deletion.
- standardization of formats (e.g., date formats, measurements).
- Correcting errors (e.g. typos, outliers).

3. **Data transformation:** Transformation of data into a format that will fit

analysis: This may include:

- Normalization or standardization of data so that it is uniform.
- Aggregate or summarize data at different levels of granularity.
- Create derived variables or features that can be most useful for analysis.
- Handle categorical variables by encoding (e.g., one-hot encoding).

4. **Data Integration:** It may be necessary to combine data from different sources into one

The data set. This may require consistency in naming, data types, or otherwise issues.

5. **Feature Engineering:** Modify existing features or simply create new features

eliminating the most important factors that can improve the product performance

learning programs. Doing this is usually based on domain knowledge and.

Creative knowledge.

6. **Exploratory Data Analysis (EDA):** Analysis of data through visual displays and.

Statistical summary to get an idea of its distribution, relationships, and patterns. EDA

It is useful for interpreting what the data 'says', and in no way 'fixes.; That's the one

point out may require additional data cleansing steps.

7. **Partitioning the data:** Partitioning the dataset into training, validation, test set...

Measure the accuracy of the research models.

8. **Data Documentation:** To gain insight into how the data is prepared, and

It is necessary to write a detailed step by step procedure about the data source, cleanup steps, .  
implementation of variables, and so on.

9. **Data Validation:** After the data preparation step you have to ensure that.

the dataset is good enough to be validated against desired analysis and analysis  
requirements and business goal setting. This may involve cross-checking  
domain experts or perform sanity checks.

10. **Data governance and Security:** Ensuring data privacy and security measures  
follow through the entire data preparation process to protect sensitive information  
and comply with legal requirements.

Effective data preparation is essential to generate and build reliable insights  
informed business decisions. Lay the foundation for a successful research career  
and ensures that the research is based on high quality and reliable data.

### **3.1 Prepare the data and Preprocessing:**

Preparing the data for visualization involves cleaning the data to remove unnecessary or

By transforming data into an accessible, easy-to-manage data format, .

Analyzing data to identify patterns and trends, filtering data to focus on specific trends

Small data groups, data preparation for visualization software, and data verification

It is accurate and complete. This approach helps to make the data easier to understand and.

Ready to draw diagrams to gain insights into productivity and efficiency.

Since the data is already clean, we can move on to visualization.

In this step a few filters are added by renaming a few given field names ,

Simple data visualization:

Field name	Renamed name
Residents providing email	Email
Residents providing mobile	mobile

## Statistical Fields :

The term "statistical fields" usually refers to changes in data structures that are generated by calculation rather than derived directly from the source data These fields are created by applying statistical operations, functions, or formulas to existing data at data set internally.

1) Age group:

**Update calculated field**

Name  
Age Group

Expression

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

Preview

Kid

### Age Group:

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

### 2) Region:

Update calculated field

Name

Region

Expression

```

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')))))

```

Preview

Q

**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')))))

### 3)Email status:

Update calculated field

Name

Email Status

Expression

```

If(Email>=1,1,0)

```

Preview

Q

0
1

Update

Email status: if(Email>= 1,1,0)

### 4)Mobile Status:

**Update calculated field**

Name  
Mobile Status

Expression  
If(Mobile >= 1, 1, 0)

Preview

Q	0
	1

Update

Mobile Status: if(Mobile >=1,1,0)

### 3.2 Data Visualization:

Data visualization involves creating graphical representations of data to help understand and analyze information. Its goal is to make complex data accessible, simple, and easy to interpret. Using visual elements such as charts, diagrams and maps, data visualization allows people to quickly identify patterns, trends and trends in data.

#### Unique visualisations:

A wide variety of visualizations can be created from a given dataset to analyze the performance and efficiency of banks. Common types include bar charts, line charts, heat maps, scatter plots, pie charts, and maps. These visualizations serve various purposes, such as comparing performance across different banks or branches, tracking changes over time, understanding distribution, revealing relationships between variables, breaking down revenue sources, and displaying customer demographics and resource allocation. Additionally, maps can highlight the geographical distribution of bank branches. Utilization of data filters, which involves applying specific criteria to include or exclude certain data points, is crucial in data analysis as it allows for focusing on relevant subsets of data, eliminating noise, and irrelevant information, thus enhancing the quality of insights derived.

### 4.Dashboards:

Qlik's dynamic dashboards surpass all other forms of data visualization and business intelligence. Unlike



simplistic interactive dashboards, Qlik's immersive dashboards tell a story through various embedded visualizations. These dashboards are then shared with decision-makers, providing them with deep insights into their datasets. The two most powerful graphical approaches offered by Qlik for representing data are highly interactive, enabling a deeper and more comprehensive analysis of complex data sets.

The standout feature of Qlik's dashboards is their interactivity, which allows users to engage with multiple features on the dashboard seamlessly. This interactivity goes beyond simple point-and-click filtering, enabling users to drill down into the data for a more detailed investigation. This dynamic capability enhances the effectiveness of data analysis, allowing decision-makers to explore data from various perspectives and derive meaningful insights that drive informed decision-making.

A major advantage that sets Qlik apart from competitors is the ability to customize the dashboards' appearance and functionality. Users can tailor the layout, style, and information presentation to meet their specific needs and preferences. This includes rearranging visualizations to highlight key values, adjusting colors for optimal contrast, and modifying other visual elements. This level of customization ensures that dashboards are not only functional but also visually appealing and user-friendly.

Qlik's dashboards are more than static visualizations; they facilitate real-time exploration of each displayed metric. This capability provides decision-makers with real-time insights and actionable data, enabling quick and informed decisions. The dynamic nature of these dashboards means users can continuously engage with and analyze their data, improving the overall decision-making process.

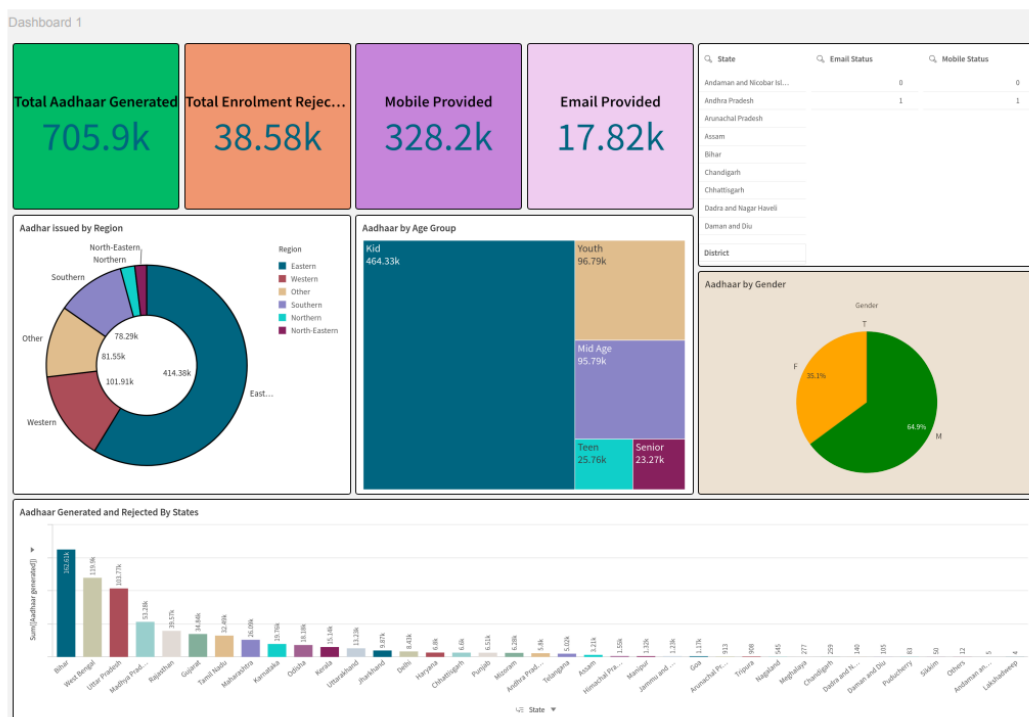
Integration is another key strength of Qlik's dashboard environment. By allowing multiple users to collaborate on a single dashboard simultaneously, Qlik transforms organizations from mere data consumers to data-driven entities. This collaborative approach helps solve problems and share solutions effectively, fostering a culture of data-centric decision-making within the organization.

Qlik has also addressed the need for accessibility across various devices, including mobile. This ensures that users can access their data anytime and anywhere, aiding decision-making even when they are not at their desks. The ability to interact with dashboards on mobile devices ensures that critical data insights are always within reach, enhancing responsiveness and flexibility.

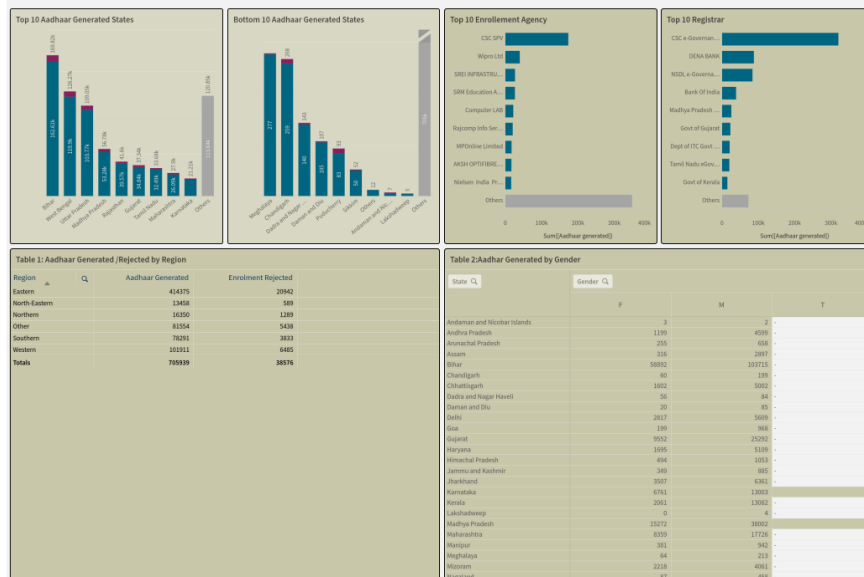
In conclusion, Qlik's dashboards are more than just information tools; they are transformative instruments that convert data into actionable insights. They empower users to make strategic decisions and fully leverage their data assets. With Qlik, the journey from raw data to actionable intelligence is seamless and enlightening, revolutionizing how organizations approach and utilize their data.

## **Dashboard 1 Overview:**

Dashboard 1 provides a detailed analysis of key performance indicators related to sales, customer demographics, and product performance. It includes visual representations such as bar charts, pie charts, and line graphs to showcase trends in sales volume, revenue growth, and customer acquisition. The dashboard highlights top-performing products, offering insights into which items drive the most revenue



Dashboard 2



## 5.Story:

The Story report provides an in-depth analysis of Aadhaar card generation across various parameters such as population, regions, countries, registrants and more for examples, highlighting data points and trends.

### Population distribution:

The report begins by detailing the total number of Aadhaar cards, broken down by the addition of mobile numbers and emails. It provides a breakdown of Aadhaar recipients by mobile number and email. Additionally, it includes a gender-wise analysis, which shows that 458.2K Aadhaar cards were issued to males and 243.73K to females, indicating higher male enrollment. The report also includes the distribution of Aadhaar cards by age species for males and females, although specific age-related data are not detailed in the subsections.

### Local research:

In a regional analysis, the report showed that the eastern region tops the list in Aadhaar card generation, with 414.4K cards issued and 20.94K rejected followed by the western, southern, northern and northeastern regions, each with different rates of supply and rejection. The "other" section covering various regions issued 81.55K Aadhaar cards of which 5.44K were rejected. These findings highlight significant regional differences in Aadhaar enrollment and rejection rates, with the highest levels of activity found in the Eastern Province

#### **State-wise base generation:**

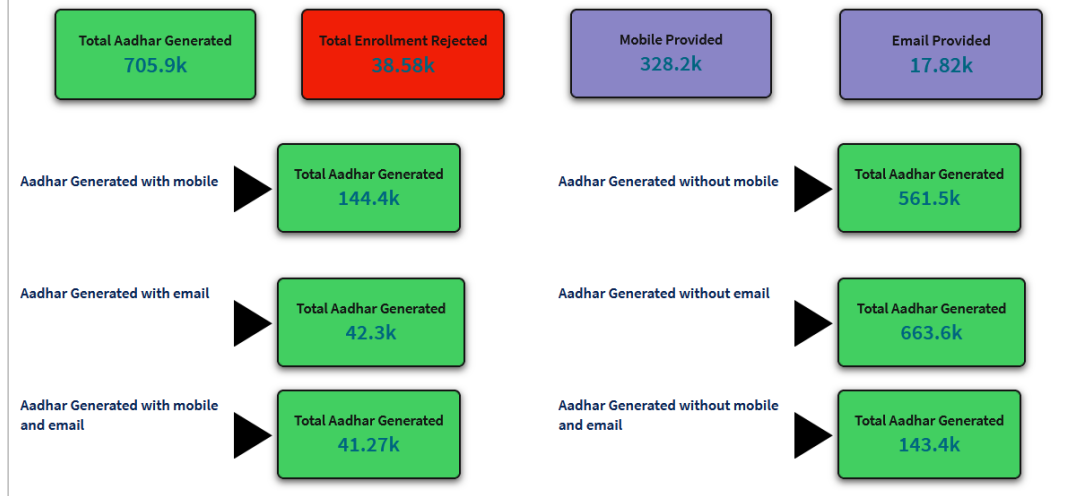
Providing a detailed analysis of the entire state, the report shows that Bihar is the state with the highest number of Aadhaar registrations at 162,607 cards. West Bengal is next with 119,901 cards and Uttar Pradesh came third with 103,767 cards. This section highlights the Aadhaar enrollment efforts in these populous countries. In contrast, Lakshadweep and Andaman and Nicobar Islands have the lowest Aadhaar generation rates, with only 5 and 7 forms, respectively, indicating minimal enrollment activities, the report said these places

#### **Aadhaar Companies:**

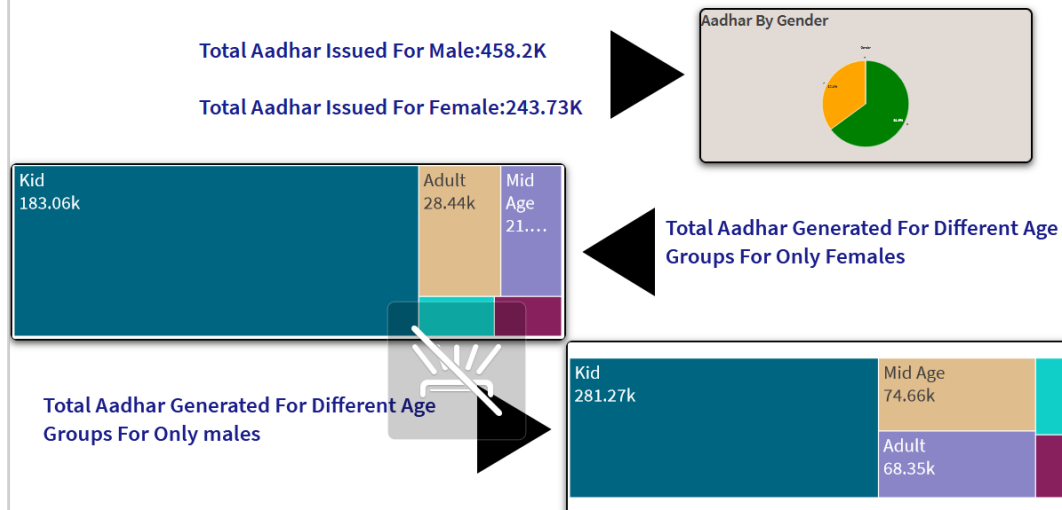
A significant part of the report looks at the performance of Aadhaar enrollment agencies. CSC SPV proves to be the top company, issuing 173.19K base cards. It is followed by Wipro Ltd with 39.62K cards and SRE Infrastructure Finances Ltd with 26.5K cards. The report also mentions CSC e-Governance Services India Ltd as the first registrant with a whopping 320.2K Aadhaar cards, surpassing Dena Bank and NSDL e-Governance Infrastructure Ltd . This study highlights the important role played by multiple agencies in providing Aadhaar enrollment.

Overall, the report provides important insights into the distribution and registration of Aadhaar cards across India. It focuses on regional and country-specific differences, the efficiency of listing companies, and gender-based differences in listing rates. These insights are important for policy makers and stakeholders to identify areas that need attention and develop strategies to improve and balance the issuance of Aadhaar cards across the country. The detailed analysis of the report is a valuable resource to understand the current state of Aadhaar enrollment and its impact on future projects.

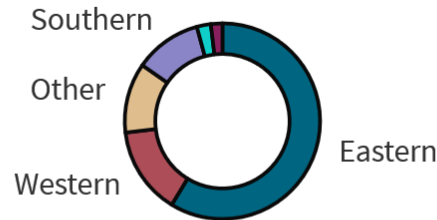
## Total Aadhar Analysis



## Aadhar Generated By Gender

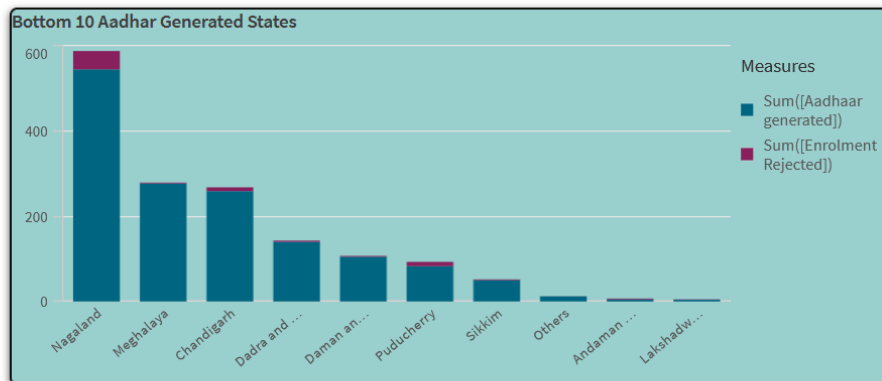


## Aadhar Generated And Rejected By Region



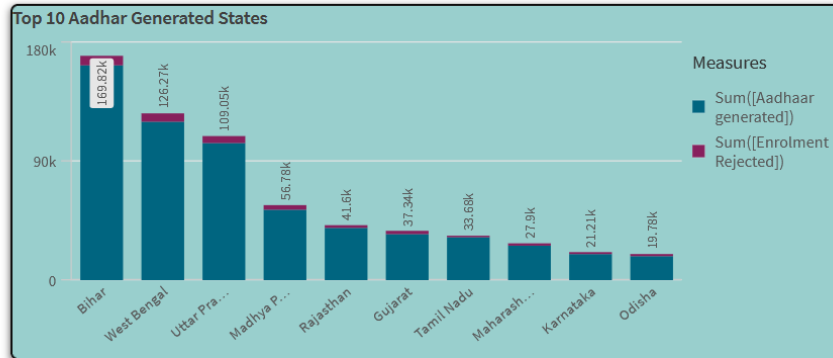
- ✓ Eastern: 414.4K Generated and 20.94K Rejected
- ✓ Northern: 16.35K Generated and 1.2K Rejected
- ✓ Western: 101.9K Generated and 6.49K Rejected
- ✓ North Eastern: 13.46K Generated and 589 Rejected
- ✓ Southern: 78.29K Generated and 3.83K Rejected
- ✓ Other: 81.55K Generated and 5.44K rejected

## Bottom 10 Aadhar Generated States



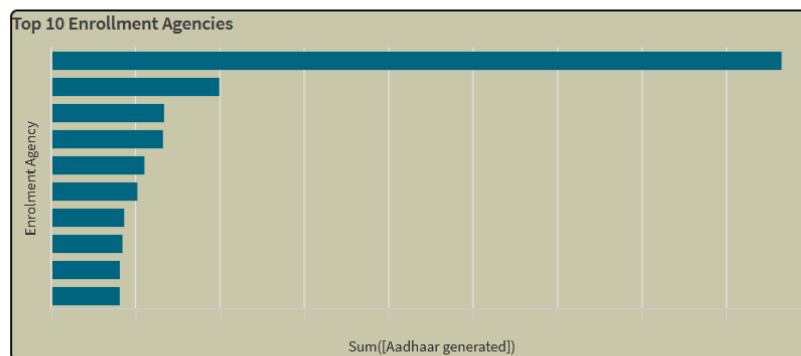
- ✓ Among the states and union territories, Lakshadweep has the lowest Aadhaar card generation count, with only 5 Aadhaar cards issued.
- ✓ Andaman and Nicobar islands ranks as the second-lowest state in terms of Aadhaar card generation, with a total count of only 7 Aadhaar cards issued

## Top 10 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 119901 Aadhaar cards.
- ✓ Uttar Pradesh is 3rd Highest Aadhaar Registration State generating total of 103767 Aadhaar cards.

## Top 10 Aadhar Enroll Agency



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

## **6. Conclusion:**

The main objective of the project is to address the critical need for efficient analytics and use of primary data through the use of Qlik Sense. This powerful analysis tool enabled base data sets to be preprocessed and visualized, and it provided interesting dashboards and graphics that made it easy to gain valuable insights

To make the best use of these resources, the project includes a comprehensive review of the existing literature on basic research. This review helped identify current research findings and recommendations from previous research, improved the project methodology and contextualised the results

The potential impact of the project is vast and multifaceted. At the social level, it can help identify the number of Aadhaar users, link it to social indicators, and assess the impact on development and economic inclusion policies, affecting social welfare programmes. It also looks at the overall effectiveness of Aadhaar in preventing fraud.

Data collection and extraction were carefully conducted, ensuring the accuracy of the data set. Through a comprehensive analysis of each aspect, the project provided valuable insights that could aid National Reference Commissions and other decision makers in policy formulation.

Overall, this project represents a major breakthrough in the use of data to positively influence decision-making and future planning.

It is well positioned for future efforts aimed at optimizing the social and economic impacts of high-security base design, including the use of Qlik Sense for advanced primary data analysis