# Data Insights from Aadhaar a Comprehensive Analysis using Qlik

## 1.Introduction:

### 1.1 Overview:

This project entails a thorough examination of basic data using Qlik Sense, with an emphasis on delivering actionable insights. Base datasets are cleaned and modelled, interactive Qlik Sense dashboard reports are created, and essential visualisations such as population predictions, generation/rejection, and geography analysis are extracted

### 1.2 Purpose:

The major goal of this project is to conduct a detailed study of Aadhaar data to improve decision-making, policy formulation, and operational efficiency at the National Identity Authority. This research attempts to provide significant insights into user demographics, authentication trends, and compliance data to help you make informed decisions

### 1.3 Technical Architecture:

The project makes use of Qlik Sense to visualise data and create dashboards. The key data source is the Aadhaar dataset, which includes demographics, authentication records, and geographical information. The architecture includes data gathering, preparation, visualisation, and dashboard design to aid strategic planning

## 2.Problem Understanding:

### 2.1 Business Problem:

The Unique Identification organisation of India (UIDAI), a legal organisation established by the Indian government, issues Aadhaar cards with unique 12-digit numbers. The Aadhaar project intends to offer Indian inhabitants with universal and robust identification. The initiative aims to streamline identity management procedures, increase transparency, and improve the delivery of government services and benefits.

The primary data is mostly examined with Qlik Sense, with an emphasis on gaining actionable insights. Activities include cleaning and modelling base datasets, providing interactive Qlik Sense dashboard reports, and extracting essential visualisations such as demographics, generation/rejection, and geospatial analysis. These visualisations assist detect trends and patterns in data and complement the characteristics of the basic user distribution.

The primary data source is a large base database that contains demographic information, certification records, and geographic descriptors. The project's goal is to thoroughly analyse Aadhaar data using Qlik Sense in order to extract important insights that will help the National

Identification Authority improve decision-making, policy, and operational efficiency. This research intends to aid policy creation and service delivery, ultimately contributing to better governance and resource allocation.

## 2.2 Business Requirements:

This research looks deeper into user demographics, certification patterns, and compliance indicators using extensive data visualisation. The goal is to create dynamic and visually appealing dashboards that will help with planning and business improvement.
 The report attempts to investigate user profiles, loyalty practices, and compliance.

**Informed decision making**:Using data-driven insights to make strategic decisions about product development, marketing strategy, and user experience enhancement.
 Improved service delivery: Identify and eliminate inefficiencies in the certification process to ensure a smooth and effective user experience.
**Remain compliant**: Observe and visualise compliance to demonstrate it and keep trust.
 This research uses data visualisation to transform unstructured data into meaningful insights, resulting in improved process implementation, operational efficiency, and compliance.

**Social Impact Analysis**

- **Demographic Distribution:** Visualisations will be made to depict the demographic distribution of Aadhaar users, offering a clear image of the population groups that have enrolled in the Aadhaar programme.

- **Social Welfare Programs:** The report will look at the impact of Aadhaar on several social assistance programmes, noting how it has improved their delivery and efficacy.

- **Financial Inclusion:**Insights will be gained into how Aadhaar has facilitated financial inclusion, particularly in terms of providing underprivileged groups with access to banking services and financial goods.

- **Socioeconomic Indicators:** The analysis will look at how Aadhaar adoption correlates with advances in socioeconomic indices including education, health, and employment, indicating the program's larger social advantages.

**Business Impact Analysis**

- **Sectoral Impact:** The investigation will look at how Aadhaar has influenced several business sectors, including banking, telecoms, and e-commerce, and how it has helped to drive growth and innovation.

- **Fraud Prevention:**The report will evaluate Aadhaar's efficacy in combating fraud, focusing on its contributions to minimising identity theft and guaranteeing secure transactions.

- **Customer Onboarding and Operational Efficiency:** The impact of Aadhaar on customer onboarding processes and operational efficiency will be evaluated, with results demonstrating improvements in speed, accuracy, and overall user experience in

corporate 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 a key stage in business analysis since it ensures that raw data is cleaned and ready for analysis. This approach guarantees that data is accurate, thorough, and relevant, as well as providing a solid foundation for actionable insights. The following is an overview of the data generated 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. **Feature Engineering:** Create new features or modify existing ones to improve product performance and learning programmes. Typically, this is done using domain knowledge. Creative knowledge.

4. **Exploratory Data Analysis (EDA):** Data analysis using visual displays.
Statistical summary helps gain an understanding of the distribution, correlations, and patterns. EDA
It is valuable for analysing what the data'says', but it does not 'fix' anything. Note that extra data purification steps may be necessary.

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

## 3.3 Prepare the data and Preprocessing:

Preparing data for visualisation entails cleaning data to remove superfluous or
 By converting data into an accessible, manageable data format.
 Analyse data to detect patterns and trends, filter data to focus on certain trends.
 Small data groups, data preparation for visualisation software, and data validation
 It is factual and comprehensive. This strategy makes the data easier to grasp.
 Prepare to create diagrams to obtain insight into productivity and efficiency.
 We can proceed with visualisation now that the data has been cleaned up.
 This phase involves altering a few field names to add filters and simplify data visualisation.

| 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:

if(Age >= 60, 'Senior',

if(Age >= 30, 'Mid Age',

if(Age >= 18, 'Youth',

if(Age >= 14, 'Teen','Kid' ))))

## 2) 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:

if(Email>= 1,1,0)

## 4)Mobile Status:

if(Mobile >=1,1,0)

## 3.4 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 outperform all other types of data visualisation and business intelligence. Unlike simple interactive dashboards, Qlik's immersive dashboards convey a story using a variety of integrated visualisations. These dashboards are subsequently shared with decision-makers, giving them detailed insights into their datasets. Qlik's two most powerful graphical techniques to data representation are highly interactive, allowing for more in-depth and thorough examination of complicated data sets.

The interactivity of Qlik dashboards is a unique feature, allowing users to easily interact with multiple components on the dashboard. This interaction extends beyond simple point-and-click filtering, allowing users to delve down into the data for more in-depth analysis. This dynamic capacity improves the effectiveness of data analysis by allowing decision-makers to examine data from several viewpoints and extract valuable insights that drive informed decision-making.

Qlik's flexibility to customise the appearance and functionality of its dashboards is a significant edge over competitors. Users can customise the layout, design, and information display to suit their unique requirements and tastes. This includes rearranging visualisations to emphasise critical values, tweaking colours for maximum contrast, and changing other visual aspects. This amount of customisation guarantees that dashboards are not only functional, but also visually beautiful and easy to use.

Qlik dashboards are more than just static visualisations; they allow for real-time study of any shown metric. This capacity provides decision-makers with real-time insights and actionable data, allowing for quick and informed decisions. Because these dashboards are dynamic, users can interact with and analyse their data in real time, enhancing overall decision-making efficiency.

Integration is another significant advantage of Qlik's dashboard environment. By allowing several users to collaborate on a single dashboard at the same time, Qlik transforms organisations from data consumers to data-driven entities. This collaborative approach facilitates problem solving and solution sharing, promoting an organization-wide culture of data-centric decision-making.

Qlik has also addressed the need for accessibility on multiple devices, including mobile. This ensures that users may access their data at any time and from any location, allowing them to make decisions even when they are not at their workstation. The ability to interact with dashboards from mobile devices means that essential data insights are always available,

increasing responsiveness and flexibility.

Finally, Qlik dashboards are transforming tools that translate data into meaningful insights. They enable consumers to make strategic decisions and maximise their data assets. With Qlik, the transition from raw data to actionable intelligence is smooth and enlightening, revolutionising how businesses approach and use data.

**Dashboard 1 Overview:**

Dashboard 1 contains a detailed examination of important performance metrics such as sales, customer demographics, and product performance. It uses visuals like bar charts, pie charts, and line graphs to highlight trends in sales volume, revenue growth, and customer acquisition. The dashboard showcases the best-performing products and provides insights into which things generate the most revenue and consumer interaction. It also categorises clients based on demographics, giving for a better knowledge of purchase behaviour and market segmentation. This information is critical for making strategic decisions since it assists stakeholders in identifying market strengths and opportunities.

**Dashboard 2 Overview:**

Dashboard 2 emphasises operational efficiency and internal performance metrics. It provides extensive breakdowns of production data like as throughput times, defect rates, and overall equipment effectiveness (OEE). The dashboard displays real-time performance statistics against predefined targets via gauges, scatter graphs, and histograms. It also monitors personnel productivity and utilisation rates, providing insights into operational health and efficiency. This information is critical for operational managers and team leaders, allowing them to optimise processes, minimise waste, and increase overall organisational productivity.

The dashboards that have been created on Qlik Cloud using Aadhar data are:

# Dashboard 1

| Total Aadhar Generated | Total Enrollment Rejected | Mobile Provided | Email Provided |
|---|---|---|---|
| 705.9k | 38.58k | 328.2k | 17.82k |

**State**
- Andaman and Nicobar Islands
- Andhra Pradesh
- Arunachal Pradesh
- Assam
- Bihar

**District**

**Email Status**
- 0
- 1

**Mobile Status**
- 0
- 1

### Aadhar Issued By Regions

Regions: Eastern, Northern, Western, Southern, Central, North-Eastern

- Eastern 44.0%
- Northern 20.1%
- Western 14.4%
- Southern 11.1%
- Central 8.5%

### Aadhar By Age Group

- Kid 464.33k
- Youth 96.75k
- Mid Age 95.75k
- Teens 25.76k
- Senior 23.27k

### Aadhar By Gender

Gender: M, F, T
- M 64.9%
- F 35.1%

### Aadhar Generated and Rejected By States

# Dashboard 2

### Top 10 Aadhar Generated States

### Bottom 10 Aadhar Generated States

### Top 10 Enrollment Agencies

### Top 10 Registrars

### Table 1: Aadhar Generated/Rejected By Region

| Regions | Sum([Aadhaar generated]) | Sum([Enrollment Rejected]) |
|---|---|---|
| Central | 59880 | 3889 |
| Eastern | 310558 | 15654 |
| North-Eastern | 13508 | 591 |
| Northern | 141770 | 8121 |
| Southern | 78300 | 3836 |
| Western | 101923 | 6485 |
| Totals | 705939 | 38576 |

### Table 2: Aadhar Generated By Gender

| State | F | M | T |
|---|---|---|---|
| Andaman and Nicobar Islands | 3 | 2 | - |
| Andhra Pradesh | 1199 | 4599 | - |
| Arunachal Pradesh | 255 | 658 | - |
| Assam | 316 | 2897 | - |
| Bihar | 58892 | 103715 | - |
| Chandigarh | 60 | 199 | - |
| Chhattisgarh | 1602 | 5002 | - |
| Dadra and Nagar Haveli | 56 | 84 | - |
| Daman and Diu | 20 | 85 | - |
| Delhi | 2817 | 5609 | - |
| Goa | 199 | 968 | - |
| Gujarat | 9552 | 25292 | - |
| Haryana | 1695 | 5109 | - |
| Himachal Pradesh | 494 | 1053 | - |
| Jammu and Kashmir | 349 | 885 | - |
| Jharkhand | 3507 | 6361 | - |
| Karnataka | 6761 | 13003 | 0 |
| Kerala | 2061 | 13082 | - |
| Lakshadweep | 0 | 4 | - |
| Madhya Pradesh | 15272 | 38002 | 2 |
| Maharashtra | 8359 | 17726 | - |
| Manipur | 381 | 942 | - |
| Meghalaya | 64 | 213 | - |

# 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 cardsand 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
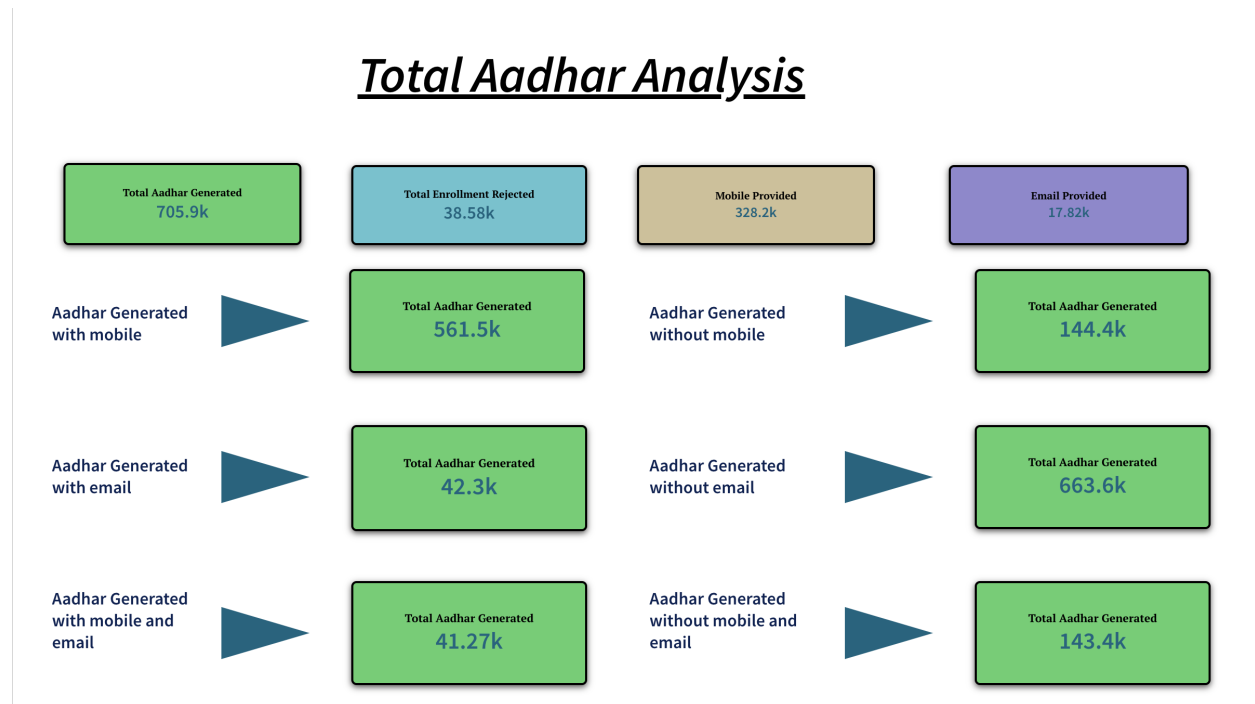
**Aadhaar Companies:**

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.
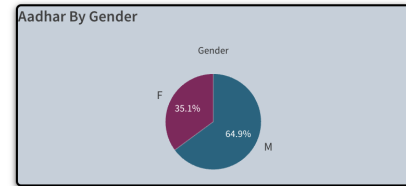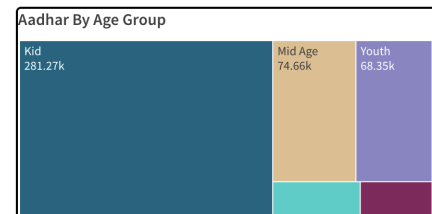
## Total Aadhar Analysis

| | | | |
|---|---|---|---|
| **Total Aadhar Generated** 705.9k | **Total Enrollment Rejected** 38.58k | **Mobile Provided** 328.2k | **Email Provided** 17.82k |

**Aadhar Generated with mobile** ▶ **Total Aadhar Generated** 561.5k

**Aadhar Generated without mobile** ▶ **Total Aadhar Generated** 144.4k

**Aadhar Generated with email** ▶ **Total Aadhar Generated** 42.3k

**Aadhar Generated without email** ▶ **Total Aadhar Generated** 663.6k

**Aadhar Generated with mobile and email** ▶ **Total Aadhar Generated** 41.27k

**Aadhar Generated without mobile and email** ▶ **Total Aadhar Generated** 143.4k

# _Aadhar Generated By Gender_

**Total Aadhar Issued For Male:458.2K**

**Total Aadhar Issued For Female:243.73K**



Aadhar By Gender

Gender

F 35.1%

64.9% M

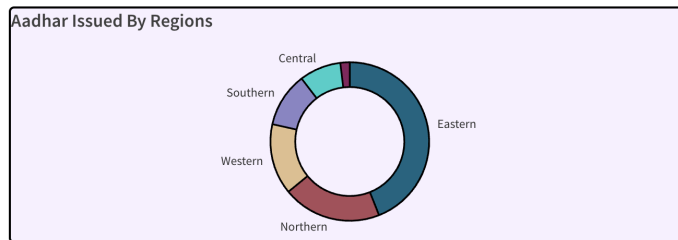Aadhar By Age Group

Kid
183.06k

Youth
28.44k

Mid
Age
21....

**Total Aadhar Generated For Different Age Groups For Only Females**

**Total Aadhar Generated For Different Age Groups For Only males**

Aadhar By Age Group

Kid
281.27k

Mid Age
74.66k

Youth
68.35k

# Aadhar Generated And Rejected By Region

Aadhar Issued By Regions

Central

Southern

Western

Eastern

Northern

**>> Eastern: 414.4K Generated and 20.94K Rejected**

_>>Western: 101.9K Generated and 6.49K Rejected_

_>>Southern: 78.29K Generated and 3.83K Rejected_

_>>Northern:16.35K Generated and 1.2K Rejected_

_>>NorthEastern: 13.46K Generated and 589 Rejected_

_>>Other: 81.55K Generated and 5.44K rejected_

# Bottom 10 Aadhar Generated States

**Bottom 10 Aadhar Generated States**
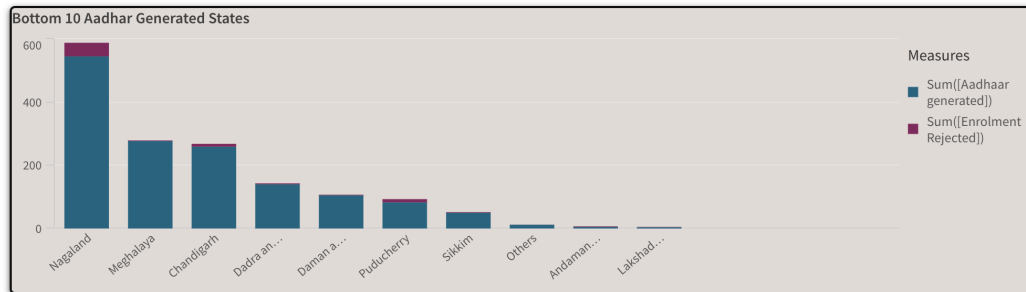


✓ *Among the states and union territories, Lakshadweep has the lowest Aadhaar card generation count, with only 5 Aadhaar cards issued.*

✓ *Andaman and Nicobar islands ranks as the second-lowest state in terms of Aadhaar card generation, with a total count of only 7 Aadhar cards issued*

# Top 10 Aadhar Generated States

**Top 10 Aadhar Generated States**
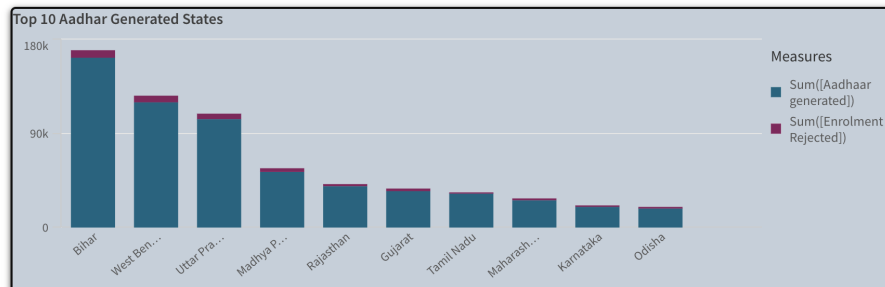


✓ *Bihar, an Indian state, has achieved the highest number of Aadhaar registrations, generating a total of 162,607 Aadhaar cards.*

✓ *West Bengal is 2nd Highest Aadhar Registration State generating total of 119901 Aadhar cards.*

✓ *Uttar Pradesh is 3rd Highest Aadhar Registration State generating total of 103767 Aadhar cards.Uttar Pradesh is 3rd Highest Aadhar Registration State generating total of 103767 Aadhar cards.*

# *Top 10 Aadhar Enroll Agency*



Top 10 Enrollment Agencies

Enrolment Agency

Sum([Aadhaar generated])
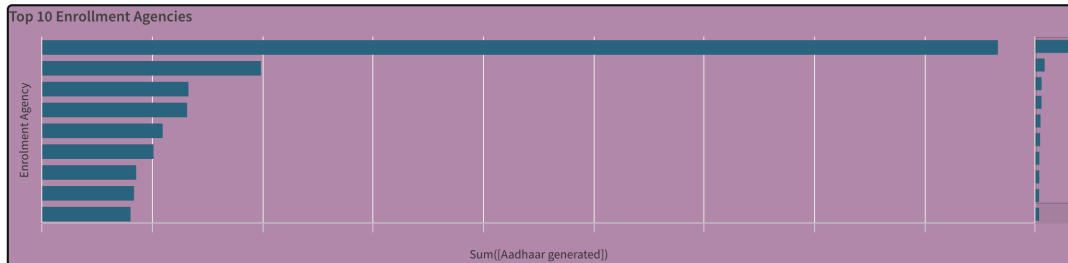
✓ *CSC SPV has achieved the highest number of Aadhaar registrations, generating a total of 173.19K Aadhar cards.*

✓ *Wipro Ltd is 2nd Highest Aadhar Registration registrar generating total of 39.62K Aadhar cards.*

✓ *SRE Infrastructure Finances Ltd is 3rd Highest Aadhar Registration registrar generating total of 26.5K Aadhar cards*

# *Top 10 Aadhar Registrars*



Top 10 Registrars

Registrar

Sum([Aadhaar generated])
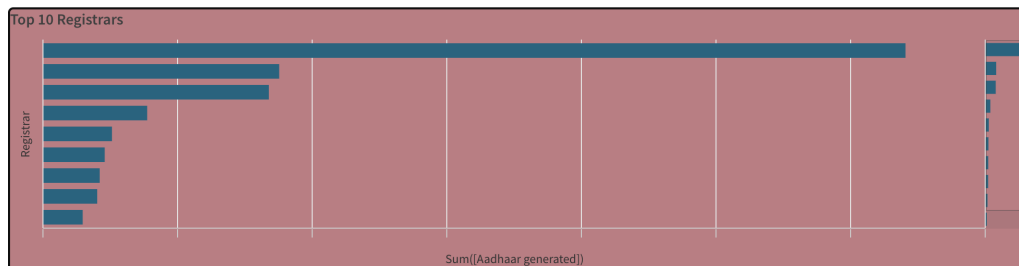
✓ *CSC e- Governance Services India Ltd has achieved the highest number of Aadhaar registrations, generating a total of 320.2K Aadhar cards*

✓ *Dena bank is 2ndHighest Aadhar Registration Registrar with 87.55K Aadhar cards.*

✓ *NSDL e-Governance Infrastructure Ltd is 3rd Highest Aadhar Registration registrar generating total of 83.71K Aadhar cards*

## Amount Of Data Loaded:

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's a measure of how much data has been successfully processed and made available for analysis,

manipulation, or use within the system



## Utilization Of Data Filters:

Utilization of data filters refers to the process of applying specific criteria or conditions to a dataset in order to selectively include 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.