

Data Insight From Aadhar A Comprehensive Analysis Using Qlik

Introduction:

Overview:

This project focuses on conducting a comprehensive analysis of data insights derived from Aadhar, India's unique identification system.

Purpose:

This project aimed to use the gigabytes of data given through Aadhar to gain insights in many directions: the attitudes and needs of much of the population, service delivery patterns, and how effective policies are or always have been. By applying Qlik services to Aadhar data, quantitative input will help discovery probable trends, patterns, and correlations. When analysed together such information can be decisive when considering a particular direction for policy making based on evidence rather than feelings alone.

Technical Architecture:

The technical architecture of this particular project, as shown below Dhar- The secret data is integrated into Qlik's platform and preprocessing – Data modelling visualisation. Qlik's robust analytics can visualise the data in an interactive way for users to explore and gain insights across dashboards/reports/visualisations. Moreover, the architecture might consist of data governance and security to provide authenticity and integrity to the Aadhar details at all times of analysis. It involves:

Aadhar databases (Data extraction)

Cleaning up and transforming the data for proper quality checking of the datasets

- SAP Transactions: Load the processed data in Qlik platform
- Use Qlik tools to analyse the Aadhar comprehensively and develop

dashboards, reports, and visualisations • Implementing security functions to secure and protect confidential data as well as being in line with the privacy laws.

Problem Understanding:

Business Problem :

Aadhaar Card is a 12-digit unique identification number issued by the Unique Identification Authority of India (UIDAI), a statutory authority established by the Indian government. The Aadhaar initiative aims to provide a universal and robust identity infrastructure for residents of India.

A comprehensive analysis of Aadhaar data is conducted using Qlik Sense, with a focus on deriving actionable insights. The project involves cleaning and modelling the Aadhaar dataset, designing an interactive Qlik Sense dashboard Report , and extracting key visualisations such as demographic overviews, Generation/Rejections and geospatial analyses.

The primary data source is the extensive Aadhaar database, comprising demographic information, authentication records, and geographical details. The objective of the project is to conduct a thorough analysis of Aadhaar data using Qlik Sense, with the aim of extracting valuable insights to enhance decision-making, policy formulation, and operational efficiency within the National Identity Authority.

Business Requirements :

It is anticipated that the research will offer significant knowledge on customer profile, authentication trends, and compliance metrics to assist in decision-making. The major emphasis here is on developing interactive dashboards that are visually appealing to aid strategic planning as well as operational improvement. This interpretation will help in better decision-making, improved service delivery, and regulatory compliance.

Literature Survey :

A literature survey on Aadhar analysis would involve researching previous studies, articles, and reports related to the subject. This can have information on how the method and techniques used in analysing Aadhar Analysis as well as the results and findings from these studies. Explore academic databases such as PubMed, IEEE Xplore, Google Scholar and Institutional repositories. Moreover, government reports and publications can give an idea of what's happening recently with respect to Aadhar analysis.

Social Or Business Impact :

Social Impact Analysis:

- Create visualisations to showcase the demographic distribution of Aadhar users.
- Analyse how Aadhar has impacted social welfare programs, financial inclusion, and other key areas.
- Explore any correlations between Aadhar usage and improvements in socioeconomic indicators.

Business Impact Analysis:

- Analyse how Aadhar has affected businesses, especially in sectors like banking, telecommunications, and e-commerce.
- Evaluate the impact of Aadhar on fraud prevention, customer onboarding, and operational efficiency.
- Create visualisations to represent the growth in Aadhaar-based services.

Data Collection & Extraction From Database :

Data collection is the process of gathering and measuring information on variables of interest in an established, systematic fashion that enables one to answer stated research questions, test hypotheses, evaluate outcomes, and generate insights from the data.

Downloading The Dataset

The data is downloaded from the given link:

<https://drive.google.com/file/d/1Umb7QTOxgTZUyCXoCIniM3DHmIejKzFV/view?usp=sharing>

Understand The Data

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

Column Description of the Dataset:

- 1. Registrar:** The registrars are registrar organizations that set up registration centres, manage the enrolment procedure and gather the necessary information.
- 2. Enrolment Agency:** An Agency of Enrolment is that entity responsible for doing real time of enrolling individuals into Aadhar system
- 3. State:** Indian state
- 4. District:** A district is a subdivision or unit of administration within a larger administrative area, such as state
- 5. Sub-District:** A sub-district, also known as Taluka or Teseel in different parts, is a smaller administration unit in the district.
- 6. Pin Code:** The PIN code issued to the Aadhar cardholder
- 7. Gender:** Gender on the Aadhar card
- 8. Age:** Age on the Aadhar card
- 9. Aadhar generated :** Number of Aadhar Generated
- 10. Enrolment Rejected :** Number of Enrolments Rejected
- 11. Residents providing email :** Email supplied by residents or not
- 12. Residents providing Mobile :** Mobile no supplied by residents or not

Data Preparation:

The stage of data preparation is a critical aspect in business analytics, whereby raw data is altered, scrubbed, and set in order to make it ready for analysis. This ensures that the data is accurate, complete, and relevant for the intended analysis. Here's a general overview of the data preparation process in business analytics:

1. Data Collection: Get together raw information from different sources, including databases, spreadsheets, APIs, sensors, social media, among others.

2. Data Cleaning: Identifying and correcting mistakes; inconsistent and missing values are corrected during this step. Common tasks include:

- Removing duplicate records.
- Filling in gaps with imputation or deletion.
- Standardizing formats (e.g., date formats, units of measurement).
- Correcting errors (e.g., typos, outliers).

3. Data Transformation: Alteration of data into a form that will be appropriate for analysis; this may involve:

- Normalizing or standardising data to ensure uniformity.
- Aggregating or summarising data at different levels of granularity.
- Creating derived variables or features that might be more useful for analysis.
- Handling categorical variables through encoding (e.g., one-hot encoding).

4. Data Integration: There may be a need to merge data from various sources into one dataset. This may include harmonising naming conventions, data types, or other issues.

5. Feature Engineering: Creating new features by converting existing ones or just extracting the very important features that can boost the performance of machine learning algorithms. Doing this is usually dependent on domain knowledge and creativity.

6. Exploratory Data Analysis (EDA): Searching the data via visual displays and statistical summaries to get an idea of its distributions, relationships, and patterns. EDA is useful for deciphering what the data 'tells', nor make it 'ready' in any way.; it only points out may require further data cleaning steps.

7. Splitting Data: Dividing the data set into training, validation, and test sets to accurately measure the performance of analytical models. 8. Data Documentation: To achieve transparency on how the data was prepared, it is necessary to write a detailed step-by-step process on data sources, cleaning steps, types of transformations used, etc.

9. Data Validation: After the data preparation step, you should make sure that the dataset is good enough for desired analysis by validating it against analysis requirements and business objectives system. This may involve cross-checking with domain experts or performing sanity checks.

10. Data Governance and Security: Ensuring that data privacy and security protocols are followed throughout the data preparation process to protect sensitive information and comply with regulatory requirements.

Effective data preparation is essential for obtaining reliable insights and making informed business decisions. It lays the foundation for successful analytics projects and ensures that the analysis is based on high-quality, trustworthy data.

Prepare The Data For Visualization

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

In this step, a few of the given field names are renamed and added a few filters for easy visualisation of the data.

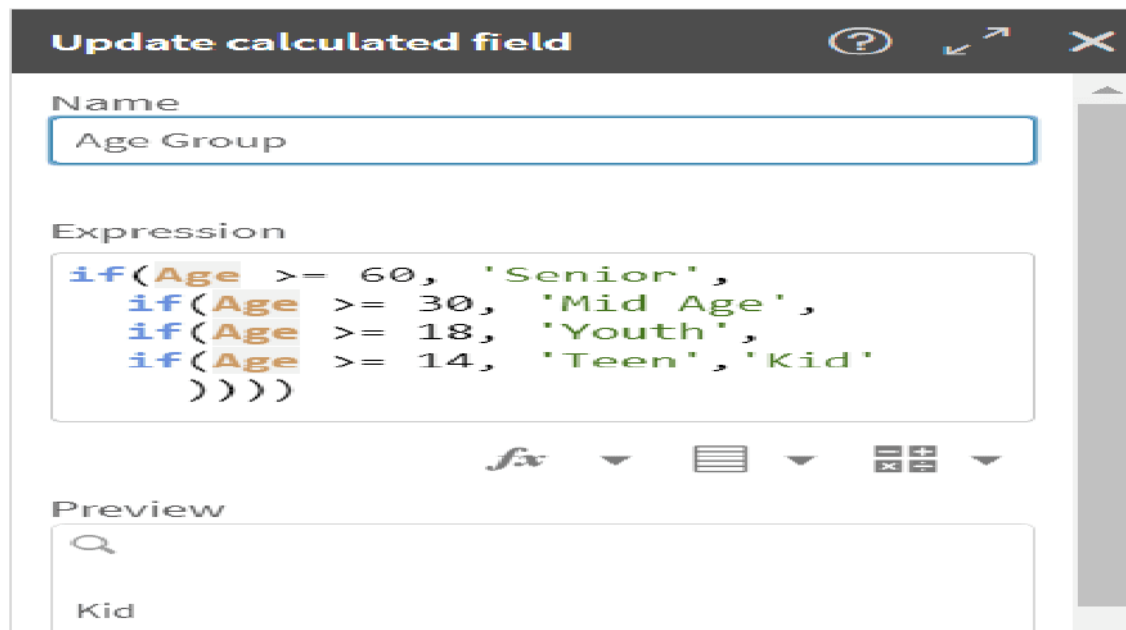
Fields that are renamed are:

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

No Of 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.

1) Age group:



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

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:

The screenshot shows a dialog box titled "Update calculated field". It has a "Name" field containing "Email Status" and an "Expression" field containing "If(Email)>=1,1,0)". Below the expression field is a "Preview" section with a search icon and a table showing two rows: one with "0" and one with "1". At the bottom is an "Update" button.

Preview
0
1

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

4)Mobile Status:

The screenshot shows a dialog box titled "Update calculated field". It has a "Name" field containing "Mobile Status" and an "Expression" field containing "If(Mobile >= 1, 1, 0)". Below the expression field is a "Preview" section with a search icon and a table showing two rows: one with "0" and one with "1". At the bottom is an "Update" button.

Preview
0
1

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

Data Visualization:

Data visualisation is the process of creating graphical representations of data to help people understand and explore the information. The goal of data visualisation is to make complex data sets more accessible, intuitive, and easier to interpret. By using visual elements such as charts, graphs, and maps, data visualisations can help people quickly identify patterns, trends, and outliers in the data.

No Of Unique Visualizations

The number of unique visualisations that can be created with a given dataset. Some common types of visualisations that can be used to analyse the performance and efficiency of banks include bar charts, line charts, heat maps, scatter plots, pie charts, Maps, etc. These visualisations can be used to compare performance, track changes over time, show distribution, and relationships between variables, breakdown of revenue and customer demographics, workload, resource allocation, and location of banks. 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 us to focus on relevant subsets of data, eliminating noise and irrelevant information.

Dashboards:

Compared to all other forms of data visualisation and business intelligence, Qlik's dynamic dashboards stand above and beyond. Instead of such types of dashboards which provide a simplistic yet interactive display for the data, immersive dashboards are created which tell a story with the help of the various visualisations embedded in it. These dashboards are then circulated among decision makers who, through this way, can have access to the depth of their datasets. The two most powerful graphical approaches offered by Qlik to represent your gamut of data are interactive.

The salient characteristic of Qlik's dashboard is their interactivity, which lets users to interact with several features on the dashboard. There is an obvious difference between pointing-and-clicking at charts to filter type or investigate further details through drill-downs applications. The dynamism that develops allows you to effectively

search for insights from voluminous data sets; thus, making it easier for you to detect obscure trends.

However, the one point that gives Qlik an edge over the competition is the ability to personalise the look and feel of the dashboards which includes its layout, style and information presented to suit the user's needs and convenience. From shifting certain visualisations into other locations in order to bring the most significant values to the foreground to changing tones of colours to optimize contrast or other aspects that affect the perceived target's visibility, users have complete control over the functional and, especially, the visual components of the dashboards.

In addition to being simply read-only visualisations of data, Qlik's dashboard can also explore the dynamics of each displayed metric, in real time. This capability focuses on the real-time capacity to offer the decision-maker insights and actionable data to facilitate quick and informed decisions.

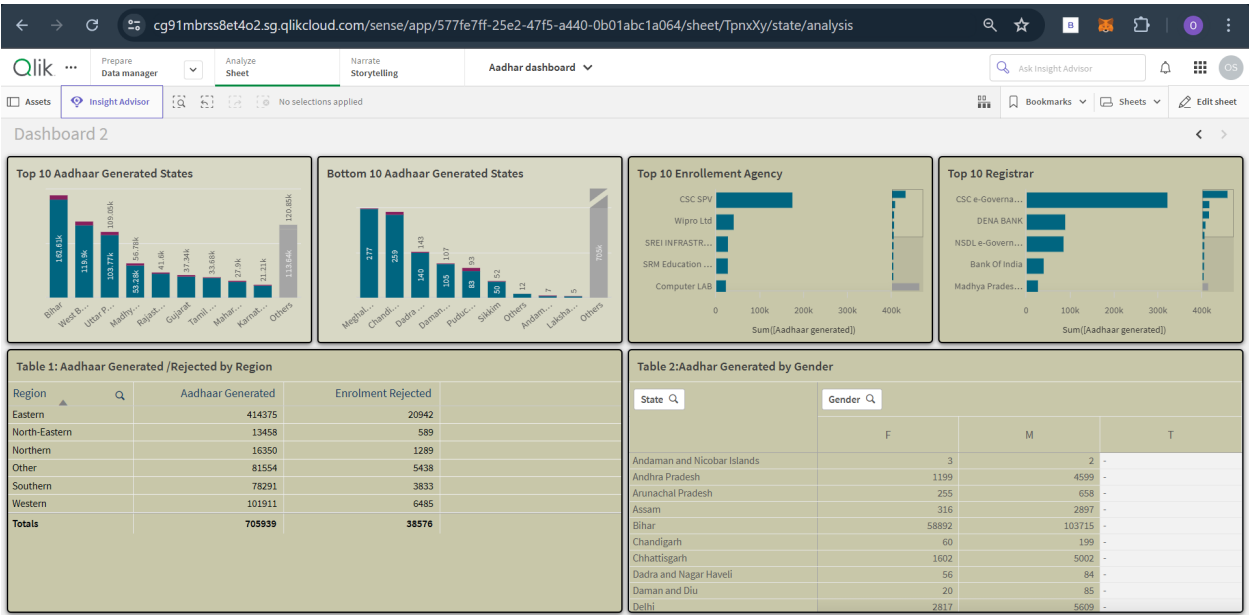
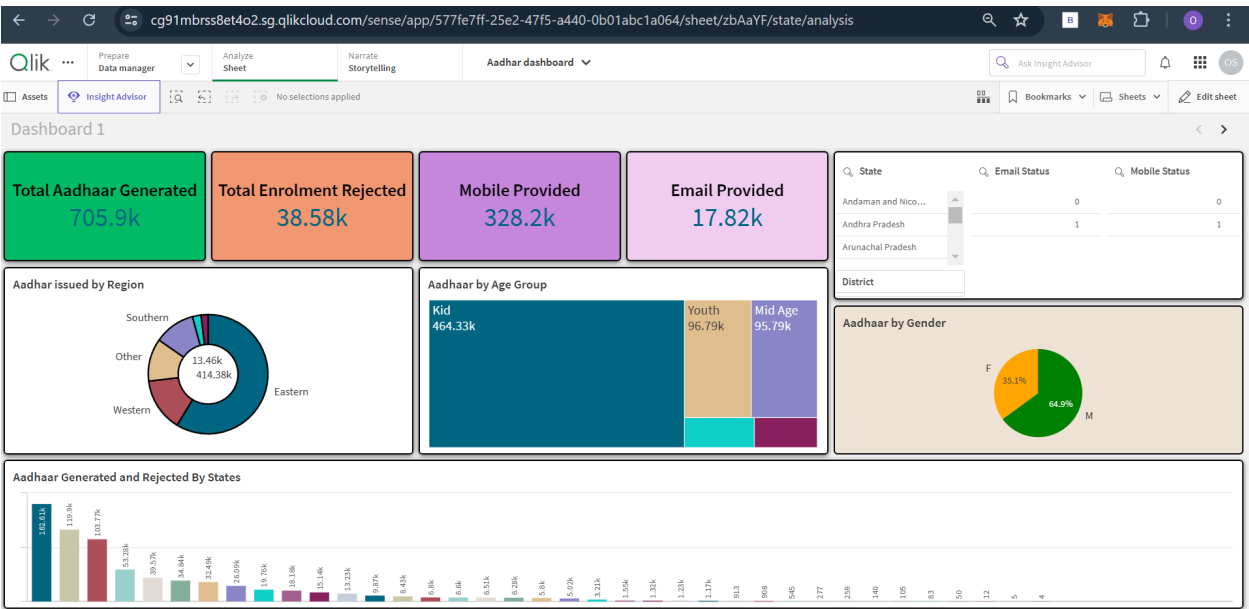
Another one of the key strategies of Qlik's dash boarding environment, integration is a pre-eminent concept. Through problems and sharing solutions that pass through the dashboard, several users can work on a single dashboard at an agreed time, and hence transform organizations from embodying mere data consumer institutions to data-driven institutions.

Moreover, Qlik have been anticipating the accessibility of the dashboards not only for fixed but, for the mobile family of devices. This mobility ensures that the users are able to access their data at any time period and from any location, hence assisting in decision-making even when not physically at the working station.

In this journey, dashboards are not simply static sheets that convey information; rather, they are weapons of change that enable users to turn information into insights and gain the necessary power to optimally utilize their information assets. As for Qlik, turning raw data into valuable insights is not just an unbroken line proposition, but a revelation.

Qlik's dashboards represent more than just visual representations of data—they are transformative tools that empower users to extract actionable insights, drive strategic decision-making, and unlock the full potential of their data assets. With Qlik, the journey from raw data to actionable intelligence becomes not just seamless but truly enlightening.

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



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

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.

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.

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

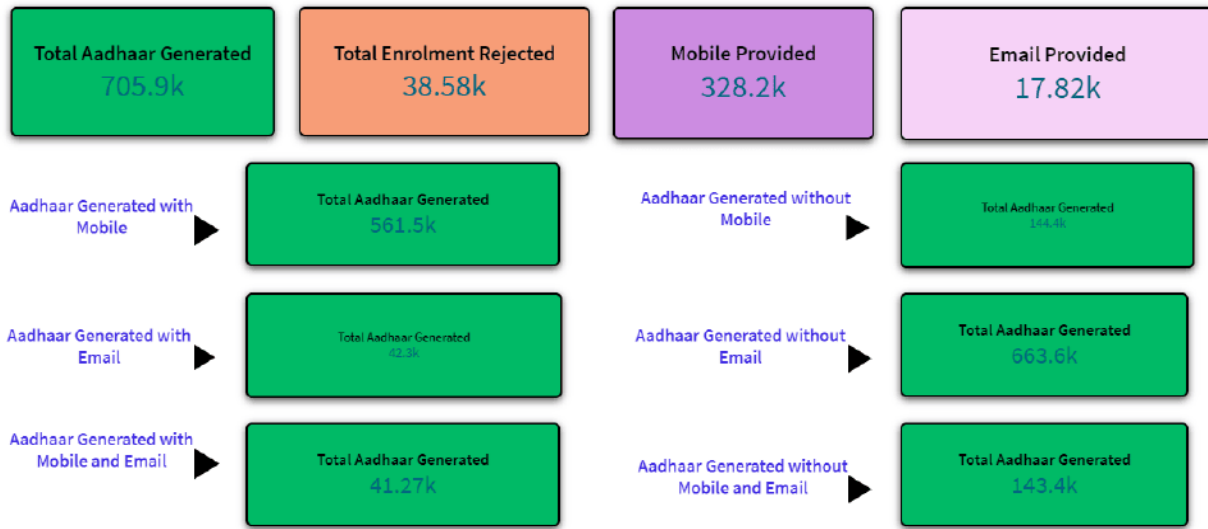
- Bihar leads with the highest Aadhar registrations at 2,162,067.
- West Bengal follows with 1,199,011 Aadhar cards generated.
- Uttar Pradesh has a total of 1,037,67 Aadhar registrations.
- The chart includes seven other states, each with varying numbers of registrations, contributing to the overall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

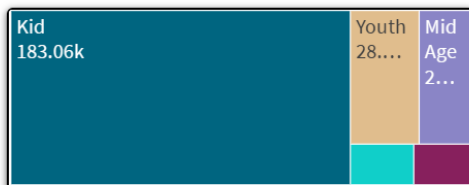
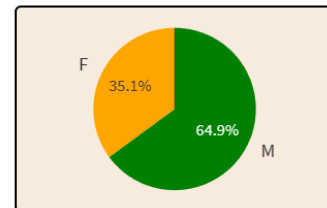
The "Top 10 Registrar" chart delineates the quantity of Aadhar registrations overseen by different organizations in India. CSC e-Governance Services India Ltd emerges as the foremost registrar, commanding a substantial lead with 320.32K Aadhar cards registered. Notably, NSDL e-Governance Infrastructure 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.

Total Aadhaar Analysis



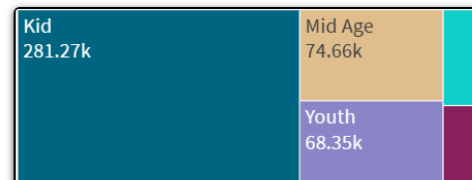
Aadhaar Generated by Gender

Total Aadhaar issue for Male: 458.2 k
Total Aadhaar issue for Female: 243.73k

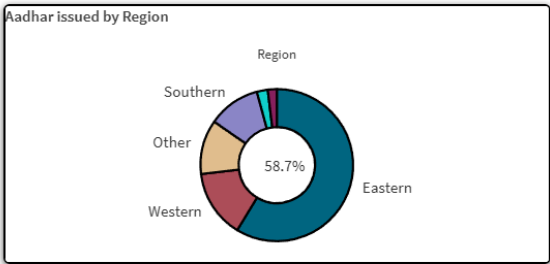


Total Aadhaar Generated for different Age Group for only Females

Total Aadhaar Generated for different Age Group for only Males



Total Aadhaar Analysis



Eastern: 414.4K Generated and 20.94K Rejected

Northern:16.35K Generated and 1.2K Rejected

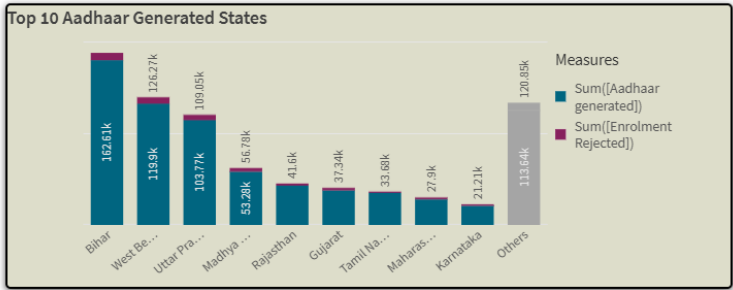
Western: 101.9K Generated and 6.49K Rejected

NorthEastern: 13.46K Generated and 589 Rejected

Southern: 78.29K Generated and 3.83K Rejected

Other: 81.55K Generated and 5.44K Rejected

Top 10 Aadhaar Generated States

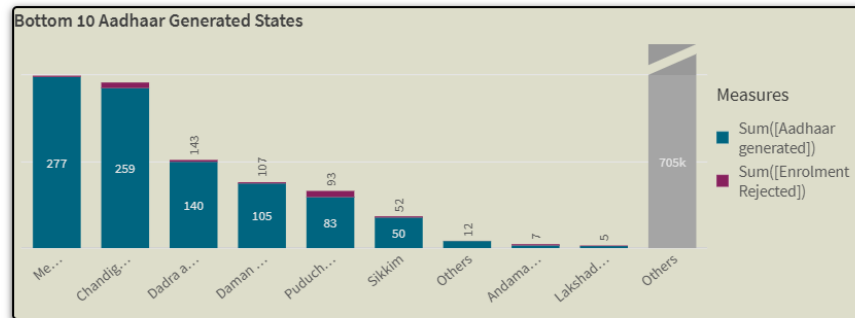


Bihar, an Indian state, has achieved the highest number of Aadhaar registrations, generating a total of 162,687 Aadhaar cards.

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 183767 Aadhar cards.

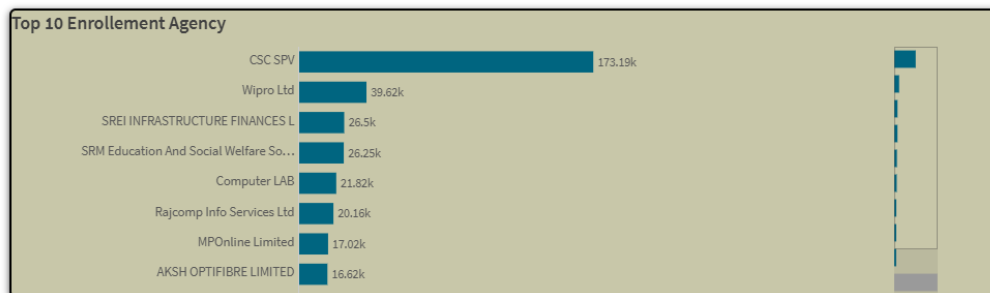
Bottom 10 Aadhaar Generated States



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 Aadhaar Enrolment Agency

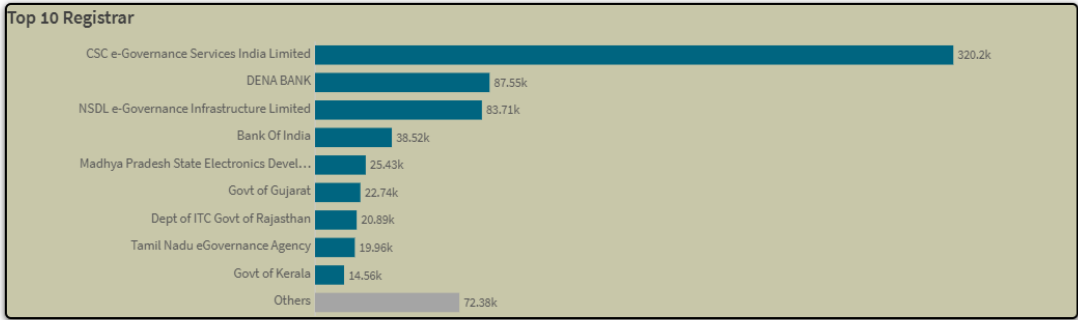


CSC SPV has achieved the highest number of Aadhaar registrations, generating a total of 173.19K Aadhar cards.

Wipro Ltd is 2nd Highest 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 Aadhaar Registrars



CSC e-Governance Services India Ltd has achieved the highest number of Aadhaar registrations, generating a total of 328.2K 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.

Conclusion:

The project's purpose was to respond to the critical gap in the efficient analysis and utilization of Aadhar data by applying Qlik Sense. With the help of this great analytical tool, the work of the project was able to preprocess and visualise the Aadhar dataset, and create dashboards and visually appealing figures which are good for extraction of insightful information.

In order to ensure that these components were carried out effectively, the project involved a subsequent survey of literature in relation to the Aadhar analysis to establish state-of-the-art research findings as well as recommendations of prior related research. This not only enriched and expanded the possibilities for the project's methodology but also helped in the contextualization of the outcomes.

Nonetheless, it is possible to note that the social and business potential of the given project is rather large-scale and complex. From a social viewpoint, the project may help determine the demographics of Aadhar users, find a relation to social indicators, and evaluate the effect on development and financial inclusion schemes that affect social welfare programs. From a business perspective, the analysis can facilitate the understanding of strengths, weaknesses and other implications that Aadhar has on various sectors such as banking and telecom sector, e-commerce, along with assessing its effectiveness in the prevention of frauds and in overall business operations.

The approach of data gathering and extraction was carefully performed, thereby contributing to the validity of the dataset. Overall, given that with each attribute, efforts were made to understand the elements of data comprehensively, it was possible to create insights out of the project, which may be useful to the National Identity Authority and other decision-makers in formulating policies.

This project marks a major improvement in the use of such data to positively impact decision-making and future planning. The project is well-positioned for the future endeavours that seek to optimise the social and economic effects of the Highly Secured Aadhar architecture, by using Qlik Sense to perform an extensive analysis of the Aadhar data.