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:

The purpose of this project is to harness the vast amount of data collected through Aadhar to gain valuable insights into various aspects of society, including demographics, service delivery, and policy effectiveness. By analysing Aadhar data using Qlik, we aim to uncover trends, patterns, and correlations that can inform decision-making processes and drive evidence-based policies.

Technical Architecture:

The technical architecture of this project involves integrating Aadhar data into the Qlik platform, where it undergoes preprocessing, data modelling, and visualization. Qlik's powerful analytics capabilities allow for interactive exploration of the data, enabling users to generate insights through dashboards, reports, and visualisations. Additionally, the architecture may include data governance and security measures to ensure the confidentiality and integrity of the Aadhar data throughout the analysis process.

It involves:

Data extraction from Aadhar database.

Data cleaning and transformation to ensure data quality.

Loading the processed data into the Qlik platform.

Developing dashboards, reports, and visualisations using Qlik tools to analyse the Aadhar data comprehensively.

Implementing security measures to protect sensitive information and comply with privacy regulations.

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 Aadhar initiative aims to provide a universal and robust identity infrastructure for residents of India.

A comprehensive analysis of Aadhar data is conducted using Qlik Sense, with a focus on deriving actionable insights. The project involves cleaning and modelling the Aadhar 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 Aadhar database, comprising demographic information, authentication records, and geographical details. The objective of the project is to conduct a thorough analysis of Aadhar 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 :

The analysis aims to provide valuable insights into user demographics, authentication trends, and compliance metrics for informed decision-making. The primary focus is on creating interactive and visually compelling dashboards to support strategic planning and operational improvements.The insights derived from this analysis will be instrumental in making informed decisions,enhancing service delivery, and ensuring compliance with regulations.

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 analysing Aadhar Analysis, as well as the results

and conclusions of these studies. It is recommended to explore 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.

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: Registrar entities that are responsible for setting up enrolment centres, managing the enrolment process, and collecting necessary data.

2. Enrolment Agency: An Enrolment 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 tensile 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 is provided or not

Data Preparation:

Data preparation is a critical stage in business analytics where raw data is transformed, cleaned, and organized to make it suitable for analysis. It involves several steps to ensure 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: Gather raw data from various sources such as databases, spreadsheets, APIs, sensors, social media, etc.

2. Data Cleaning: This step involves identifying and correcting errors, inconsistencies, and missing values in the data. Common tasks include:

   - Removing duplicate records.

   - Handling missing data through imputation or deletion.

   - Standardizing formats (e.g., date formats, units of measurement).

   - Correcting errors (e.g., typos, outliers).

3. Data Transformation: Transform the data into a format suitable for analysis. This may include:

   - Normalization or standardisation of 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: Combining data from multiple sources into a single dataset, if necessary. This may involve resolving inconsistencies in naming conventions, data types, or other discrepancies.

5. Feature Engineering: Creating new features from existing ones or extracting relevant features that can improve the performance of analytical models. This often requires domain knowledge and creativity.

6. Exploratory Data Analysis (EDA): Exploring the data visually and statistically to gain insights into its distribution, relationships, and patterns. EDA helps in understanding the characteristics of the data and identifying further data preparation needs.

7. Data Splitting: Splitting the dataset into training, validation, and test sets to evaluate the performance of analytical models accurately.

8. Data Documentation: Documenting the data preparation process thoroughly, including details about data sources, cleaning steps, transformations applied, etc. This documentation ensures transparency and reproducibility of the analysis.

9. Data Validation: Validating the prepared dataset to ensure that it meets the requirements of the analysis and business objectives. 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

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:

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

4)Mobile Status:

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:

Qlik's dynamic dashboards represent a pinnacle in the realm of data visualisation and business intelligence. These dashboards serve as immersive windows into the data universe, offering users a profound understanding of their datasets through interactive visualisations. With Qlik, users can seamlessly integrate data from diverse sources, spanning databases, spreadsheets, and cloud services, creating holistic dashboards that encapsulate the entirety of their data landscape.

One of the most striking features of Qlik's dashboards lies in their interactivity. Users can engage with various elements within the dashboard, employing intuitive interactions such as clicking on charts to filter data or delving deeper into specific insights through drill-down capabilities. This interactivity fosters a dynamic exploration

of data, empowering users to unearth valuable insights and uncover hidden trends with unprecedented ease.

Moreover, Qlik's dashboards offer unparalleled customization options, allowing users to tailor the layout, design, and content to suit their unique requirements and preferences. Whether it's rearranging visualisations to prioritise key metrics or fine-tuning colour schemes for enhanced clarity, users have full control over the aesthetic and functional aspects of their dashboards.

Beyond mere static representations of data, Qlik's dashboards enable real-time analysis, enabling users to monitor critical metrics and trends as they unfold. This real-time capability equips decision-makers with timely insights, empowering them to respond swiftly to changing circumstances and make informed choices with confidence.

Collaboration is another cornerstone of Qlik's dashboard ecosystem. Through seamless sharing and collaboration features, multiple users can access and interact with the same dashboard simultaneously, fostering a culture of data-driven decision-making and collective intelligence within organizations.
Furthermore, Qlik's commitment to accessibility ensures that dashboards are not confined to desktop environments but are also readily accessible on mobile devices. This mobility empowers users to stay connected with their data anytime,anywhere,facilitating informed decision-making even on the go.

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:

Dashboard 1

Total Aadhaar Generated
705.9k

Total Enrolment Rejec…
38.58k

Mobile Provided
328.2k

Email Provided
17.82k

State	Email Status	Mobile Status
Andaman and Nicobar Isl…	0	0
Andhra Pradesh	1	1
Arunachal Pradesh		
Assam		
Bihar		
Chandigarh		
Chhattisgarh		
Dadra and Nagar Haveli		
Daman and Diu		
District		

Aadhar issued by Region

Region: Eastern, Western, Other, Southern, Northern, North-Eastern

78.29k
81.55k
101.91k
414.38k

Aadhaar by Age Group

Kid 464.33k
Youth 96.79k
Mid Age 95.79k
Teen 25.76k
Senior 23.27k

Aadhaar by Gender

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

Aadhaar Generated and Rejected By States

Sum([Aadhaar generated])

States: Bihar, West Bengal, Uttar Pradesh, Madhya Prad…, Rajasthan, Gujarat, Tamil Nadu, Maharashtra, Karnataka, Odisha, Uttarakhand, Jharkhand, Delhi, Haryana, Chhattisgarh, Punjab, Mizoram, Andhra Prad…, Telangana, Assam, Himachal Pra…, Manipur, Jammu and …, Goa, Arunachal Pr…, Tripura, Nagaland, Meghalaya, Chandigarh, Dadra and Di…, Daman and Diu, Puducherry, Sikkim, Others, Andaman an…, Lakshadweep

State

Dashboard 2

Top 10 Aadhaar Generated States

Bihar, West Bengal, Uttar Pradesh, Madhya Pradesh, Rajasthan, Gujarat, Tamil Nadu, Maharashtra, Karnataka, Others

Bottom 10 Aadhaar Generated States

Meghalaya, Chandigarh, Dadra and Nag…, Daman and Diu, Puducherry, Sikkim, Andaman and Nic…, Lakshadweep, Others

Top 10 Enrollement Agency

CSC SPV
Wipro Ltd
SREI INFRASTRU…
SRM Education A…
Computer LAB
Rajcomp Info Ser…
MPOnline Limited
AKSH OPTIFIBRE…
Nielsen India Pr…
Others

Sum([Aadhaar generated])

Top 10 Registrar

CSC e-Governan…
DENA BANK
NSDL e-Governa…
Bank Of India
Madhya Pradesh …
Govt of Gujarat
Dept of ITC Govt …
Tamil Nadu eGov…
Govt of Kerala
Others

Sum([Aadhaar generated])

Table 1: Aadhaar Generated /Rejected by Region

Region	Aadhaar Generated	Enrolment Rejected
Eastern	414375	20942
North-Eastern	13458	589
Northern	16350	1289
Other	81554	5418
Southern	78291	3833
Western	101911	6485
Totals	705939	38576

Table 2: Aadhaar Generated by Gender

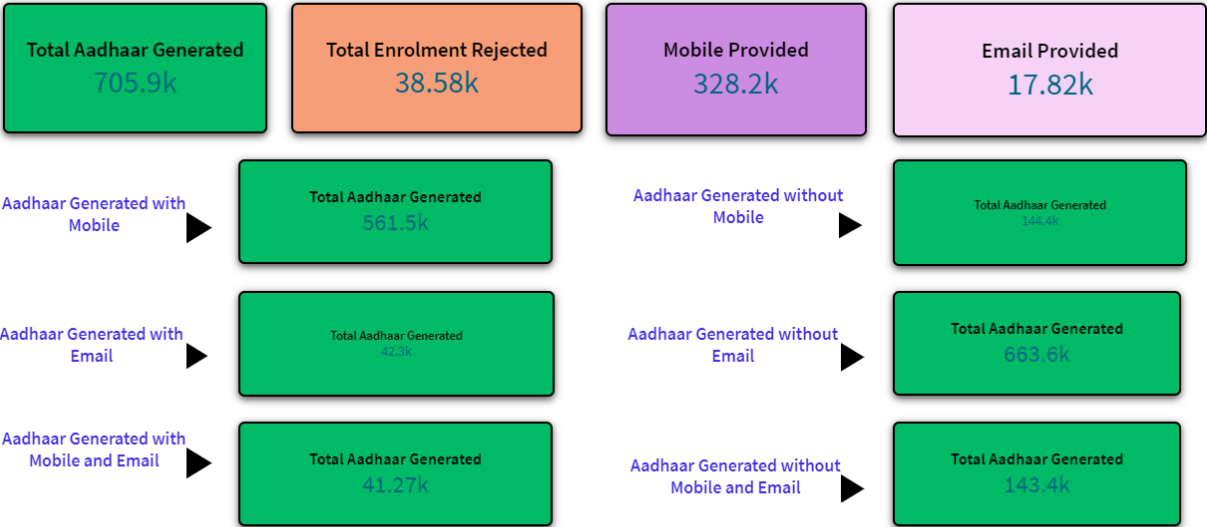| State	F	M	T
Andaman and Nicobar Islands	3	2	
Andhra Pradesh	1199	4599	
Arunachal Pradesh	255	658	
Assam	316	2897	
Bihar	59892	103715	
Chandigarh	60	199	
Chhattisgarh	1602	5002	
Dadra and Nagar Haveli	56	84	
Daman and Diu	20	85	
Delhi	2817	5609	
Goa	199	968	
Gujarat	9552	25292	
Haryana	1695	5109	
Himachal Pradesh	494	1053	
Jammu and Kashmir	349	885	
Jharkhand	3507	6301	
Karnataka	6761	13093	8
Kerala	2061	13082	
Lakshadweep	0	4	
Madhya Pradesh	19272	38002	2
Maharashtra	8359	17726	
Manipur	381	942	
Meghalaya	64	213	
Mizoram	2218	4081	
Nagaland	87	408	

Story:

As Qlik took root within DataWorks, a metamorphosis began to unfold. Gone were the days of wrestling with spreadsheets and static reports. In their place emerged dynamic dashboards, alive with colour, movement, and insight. Sales trends danced across the screen, customer behaviour revealed itself in vibrant charts and graphs, and operational inefficiencies were laid bare for all to see.

But Qlik was more than just a tool—it was a catalyst for collaboration and innovation. Departments that once operated in silos now found themselves united by a common language of data. Sales teams worked hand in hand with marketing to identify new opportunities. Operations teams optimised processes with insights gleaned from real-time analytics. And executives gained newfound clarity and confidence in their decision-making, guided by the illuminating glow of Qlik's dashboards.

As word of DataWorks' transformation spread, so too did their reputation as pioneers of data-driven innovation. Clients flocked to them, eager to harness the power of Qlik for their own businesses. Partnerships blossomed, and new horizons beckoned on the horizon.

And so, the story of DataWorks and Qlik serves as a testament to the transformative power of data and the boundless potential unleashed when visionary leadership meets cutting-edge technology. In the ever-evolving landscape of business, those who dare to embrace the tools of analytics and visualisation find themselves not just keeping pace with change, but leading the charge into a future where data reigns supreme.
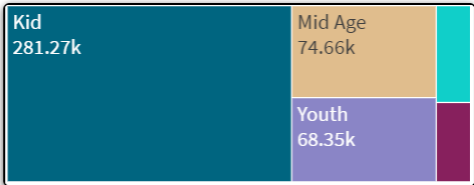
Total Aadhaar Analysis

Total Aadhaar Generated	Total Enrolment Rejected	Mobile Provided	Email Provided
705.9k	38.58k	328.2k	17.82k

Aadhaar Generated with Mobile ▶
Total Aadhaar Generated
561.5k

Aadhaar Generated without Mobile ▶
Total Aadhaar Generated
144.4k

Aadhaar Generated with Email ▶
Total Aadhaar Generated
42.3k

Aadhaar Generated without Email ▶
Total Aadhaar Generated
663.6k

Aadhaar Generated with Mobile and Email ▶
Total Aadhaar Generated
41.27k

Aadhaar Generated without Mobile and Email ▶
Total Aadhaar Generated
143.4k

Aadhaar Generated by Gender

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

Kid
183.06k | Youth 28.... | Mid Age 2...
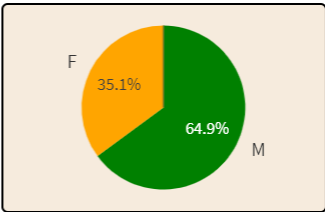
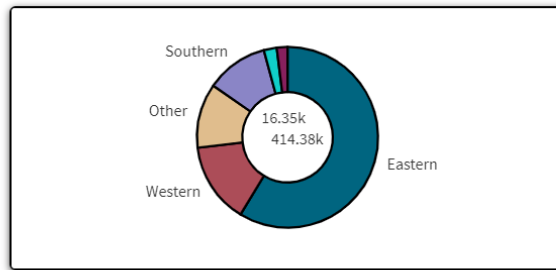◀ Total Aadhaar Generated for different Age Group for only Females

Total Aadhaar Generated for different Age Group for only Males ▶

Kid
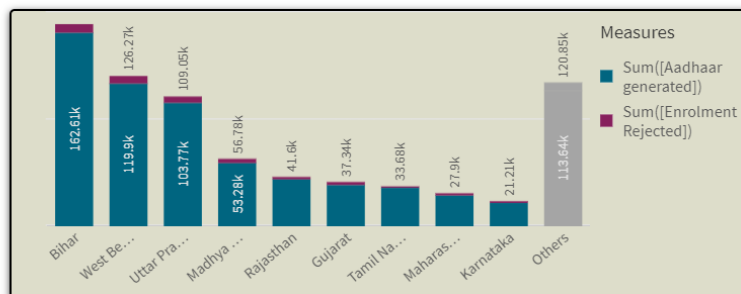**281.27k** | **Mid Age** **74.66k**
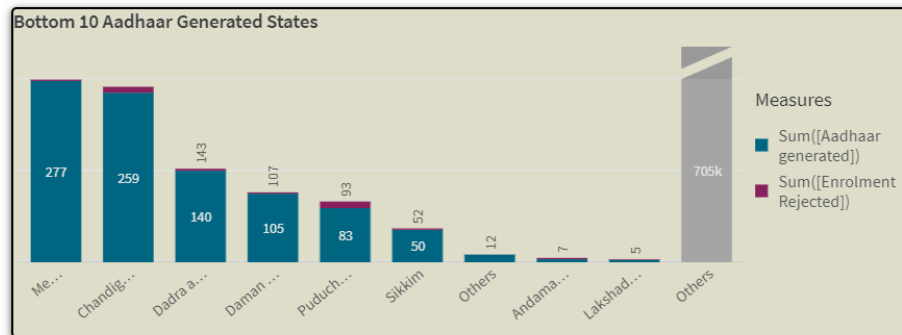| **Youth** **68.35k**

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**
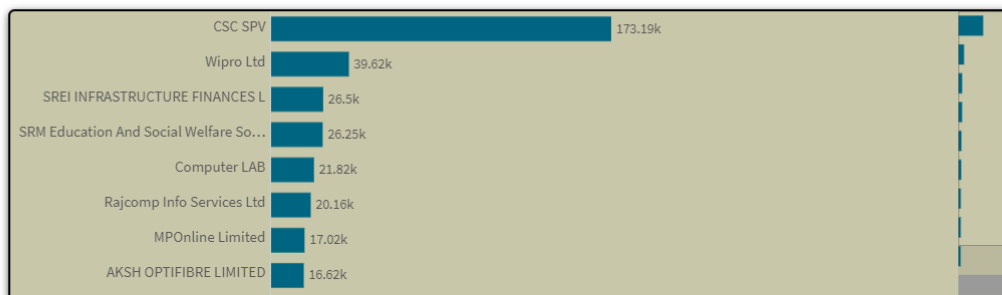
Top 10 Aadhaar Generated States

✔ **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

Bottom 10 Aadhaar Generated States

Measures
- Sum([Aadhaar generated])
- Sum([Enrolment Rejected])

277 259 143 140 107 105 93 83 52 50 12 7 5 705k

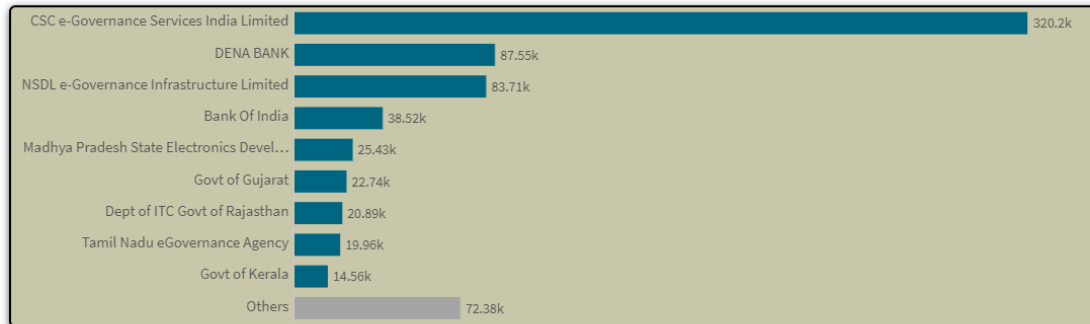Me... Chandig... Dadra a... Daman ... Puduch... Sikkim Others Andama... Lakshad... Others

✓ 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

Agency	Value
CSC SPV	173.19k
Wipro Ltd	39.62k
SREI INFRASTRUCTURE FINANCES L	26.5k
SRM Education And Social Welfare So…	26.25k
Computer LAB	21.82k
Rajcomp Info Services Ltd	20.16k
MPOnline Limited	17.02k
AKSH OPTIFIBRE LIMITED	16.62k

✓ 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 Limited	320.2k
DENA BANK	87.55k
NSDL e-Governance Infrastructure Limited	83.71k
Bank Of India	38.52k
Madhya Pradesh State Electronics Devel…	25.43k
Govt of Gujarat	22.74k
Dept of ITC Govt of Rajasthan	20.89k
Tamil Nadu eGovernance Agency	19.96k
Govt of Kerala	14.56k
Others	72.38k

✓ 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:

In conclusion, the project aimed to address the pressing need for comprehensive analysis and utilization of Aadhar data through the implementation of Qlik Sense. By leveraging this powerful analytics platform, the project successfully cleaned and modelled the Aadhar dataset, designed interactive dashboards, and extracted key visualisations to derive actionable insights.

Through a systematic literature survey, the project ensured that its analysis was informed by existing research and best practices in Aadhar analysis. This not only enriched the project's methodology but also provided valuable context for interpreting the findings.

The social and business impact of the project is vast and multifaceted. Socially, the project has the potential to shed light on the demographic distribution of Aadhar users, explore correlations with socioeconomic indicators, and assess the impact on social welfare programs and financial inclusion initiatives. From a business perspective, the analysis can provide insights into Aadhaar's influence on sectors such as banking, telecommunications, and e-commerce, as well as its role in fraud prevention and operational efficiency.

The process of data collection and extraction was conducted meticulously, ensuring the integrity and reliability of the dataset. By understanding the nuances of each data attribute, the project was able to generate meaningful insights that can inform decision-making and policy formulation within the National Identity Authority and beyond.

This project represents a significant step forward in harnessing the power of data to drive positive change and inform strategic decisions. By leveraging Qlik Sense to analyse Aadhar data comprehensively, the project has laid the groundwork for future initiatives aimed at maximising the societal and economic benefits of this unique identification system.