**Data Insight From Aadhar A Comprehensive Analysis Using Qlik**

1. **Introduction**

**1.1.** **Overview**

This project aims to leverage the powerful data visualization and analytics capabilities of Qlik to uncover meaningful insights from Aadhaar, the world's largest biometric ID system.

**1.2.** **Purpose**

The purpose of this project is to harness the extensive dataset provided by Aadhaar to extract valuable insights that can inform and enhance policy-making and service delivery in India. Utilizing Qlik's advanced data visualization and analytics capabilities, the project aims to understand population demographics and needs, evaluate service delivery patterns, assess the effectiveness of existing policies, and discover significant trends and correlations within the data. These insights will support evidence-based decision making, equipping policymakers and stakeholders with reliable, data-driven information to guide strategic decisions and improve governance outcomes for Indian citizens.

**1.3.** **Technical Architecture**

The technical architecture of this project involves integrating Aadhaar's extensive dataset into Qlik's platform for preprocessing, data modeling, and visualization. Qlik's robust analytics capabilities allow for interactive visualization, enabling users to explore and gain insights through various dashboards, reports, and visualizations. Additionally, the architecture incorporates data governance and security measures to ensure the authenticity and integrity of Aadhaar details throughout the analysis.

The technical architecture of this project involves several key components:

1. Aadhaar Databases (Data Extraction): Extract data from Aadhaar databases for analysis.

2. Data Cleaning and Transformation: Clean and transform the data to ensure high-quality datasets.

3. SAP Transactions: Load the processed data into the Qlik platform.

4. Qlik Analytics and Visualization: Use Qlik tools to comprehensively analyze Aadhaar data and develop interactive dashboards, reports, and visualizations.

5. Security and Privacy: Implement security measures to protect confidential data and ensure compliance with privacy laws.

1. **Problem Understanding**

**2.1.** **Business Problem**

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

This project involves a comprehensive analysis of Aadhaar data using Qlik Sense, focusing on deriving actionable insights. The tasks include cleaning and modeling the Aadhaar dataset, designing an interactive Qlik Sense dashboard report, and extracting key visualizations such as demographic overviews, generation/rejection rates, and geospatial analyses.

The primary data source is the extensive Aadhaar database, which comprises 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.

**2.2.** **Business Requirements**

The research is expected to provide profound insights into customer profiles, authentication trends, and compliance metrics, which will be instrumental in supporting decision-making processes. A primary focus is on creating visually appealing and interactive dashboards to enhance strategic planning and operational efficiency. By leveraging these dashboards, organizations can gain a deeper understanding of their customer base, identify and analyze authentication patterns, and ensure adherence to regulatory requirements.

These insights will enable better decision-making by providing a clear and comprehensive view of key metrics and trends. Enhanced service delivery can be achieved by identifying areas for improvement and implementing data-driven strategies. Additionally, maintaining regulatory compliance will be facilitated through continuous monitoring and analysis of compliance metrics.

Overall, the integration of these interactive dashboards will not only improve operational performance but also drive strategic initiatives, ultimately leading to improved customer satisfaction and organizational effectiveness. The use of advanced data visualization tools will ensure that complex data sets are presented in an easily understandable format, making it easier for stakeholders to interpret and act upon the information. This holistic approach will contribute to a more agile, responsive, and customer-centric organization.

**2.3.** **Literature Survey**

Conducting a literature review on Aadhaar analysis involves examining previous studies, articles, and reports related to the topic. This research encompasses understanding the methodologies and techniques utilized in analyzing Aadhaar data, as well as the outcomes and discoveries from these studies. Academic databases such as PubMed, IEEE Xplore, Google Scholar, and institutional repositories serve as valuable sources for accessing scholarly literature. Furthermore, government reports and publications provide insights into recent developments and trends in Aadhaar analysis, offering a comprehensive overview of the current state of research in this area.

1. **Data Collection**

**3.1.** **Collect the Dataset**

Data collection is a crucial process that involves systematically gathering and measuring information about variables of interest. This structured approach enables researchers to address specific research questions, test hypotheses, evaluate outcomes, and derive meaningful insights from the gathered data. By following established methodologies and procedures, data collection ensures the reliability and validity of the information collected. Moreover, it allows for the exploration of patterns, trends, and relationships within the dataset, leading to a deeper understanding of the subject under study. Effective data collection lays the foundation for informed decision-making, strategic planning, and evidence-based research across various disciplines and industries.

**Downloading The Dataset**

The data is downloaded from the given link: https://drive.google.com/file/d/1Umb7QTOxgTZUyCXoCIniM3DHmIejKzFV/view?usp=sharing

**3.2.** **Connect Data with Qlik Sense**

The data encompasses all metadata pertaining to the columns outlined within the CSV files.

Here's a detailed description of the dataset columns:

1. Registrar: These are registrar organizations responsible for establishing registration centers, managing the enrollment process, and collecting requisite information.

2. Enrolment Agency: An Enrolment Agency refers to the entity tasked with enrolling individuals into the Aadhar system in real-time.

3. State: Denotes the Indian state corresponding to the data entry.

4. District: Represents a subdivision or administrative unit within a larger administrative area, such as a state.

5. Sub-District: Also known as Taluka or Teseel in various regions, this column signifies a smaller administrative unit within the district.

6. Pin Code: Indicates the PIN code assigned to the Aadhar cardholder.

7. Gender: Reflects the gender information as stated on the Aadhar card.

8. Age: Depicts the age as stated on the Aadhar card.

9. Aadhar Generated: Indicates the number of Aadhar cards successfully generated.

10. Enrolment Rejected: Specifies the count of enrollment attempts that were rejected.

11. Residents providing email: Indicates whether residents provided their email address during enrollment.

12. Residents providing Mobile: Indicates whether residents provided their mobile number during enrollment.

1. **Data Preparation**

**4.1.** **Prepare The Data For Visualization**

The data preparation stage holds significant importance in the realm of business analytics, serving as a foundational step where raw data undergoes transformation, cleansing, and organization to prepare it for analysis. This crucial phase ensures that the data is refined, accurate, and pertinent to the intended analytical objectives. Below is an expanded outline of the data preparation process in business analytics:

1. Data Collection: Raw data is gathered from various sources, including databases, spreadsheets, and external sources.

2. Data Cleaning: Raw data often contains errors, inconsistencies, and missing values. In this step, data is cleansed by identifying and rectifying errors, removing duplicates, and filling in missing values using appropriate techniques such as imputation.

3. Data Integration: Data from disparate sources may need to be combined to create a unified dataset. Integration involves merging data from different sources while ensuring consistency and compatibility.

4. Data Transformation: Data may need to be transformed to meet the requirements of the analysis. This could include converting data types, scaling variables, or creating new derived variables.

5. Feature Engineering: This involves creating new features or variables from existing ones to enhance the predictive power of the data. Techniques such as aggregation, binning, and one-hot encoding may be employed.

6. Data Reduction: In cases where the dataset is large or contains redundant information, data reduction techniques such as feature selection or dimensionality reduction may be applied to streamline the dataset without sacrificing analytical power.

7. Data Formatting: Data is formatted to ensure consistency and compatibility with the analytical tools and algorithms being used. This may involve standardizing date formats, encoding categorical variables, or normalizing numeric variables.

8. Data Sampling: In cases where the dataset is too large to analyze in its entirety, a representative sample may be taken to perform preliminary analysis or model training.

9. Data Splitting: The dataset may be divided into training, validation, and testing sets for model development and evaluation purposes.

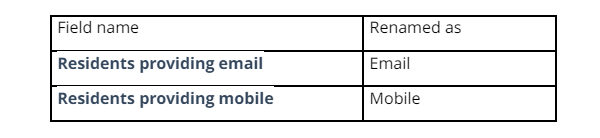
10. Data Documentation: Comprehensive documentation of the data preparation steps, including any assumptions made and transformations applied, is essential for transparency and reproducibility.

By meticulously preparing the data in accordance with these steps, businesses can ensure that their analytical endeavors are built on a solid foundation of accurate, reliable, and actionable data.

Preparing Data for Visualization

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

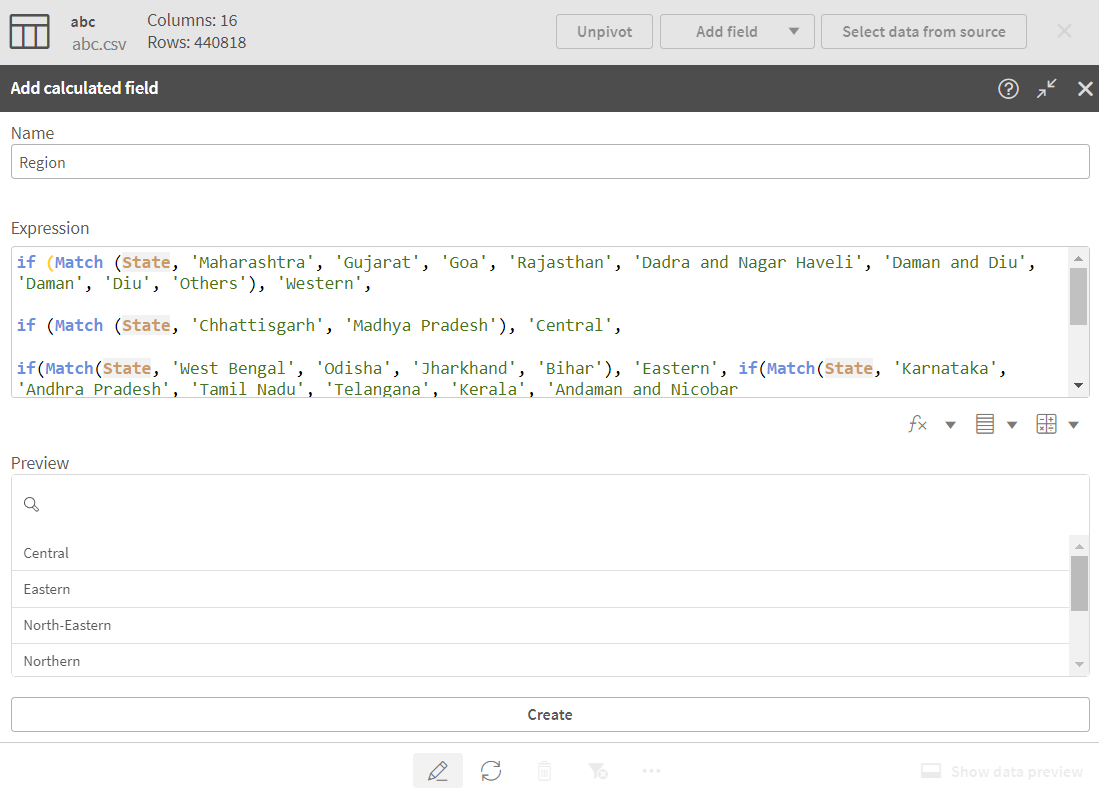
Fields that are renamed are:



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) 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')))))

2)Age Group :



Age Group :

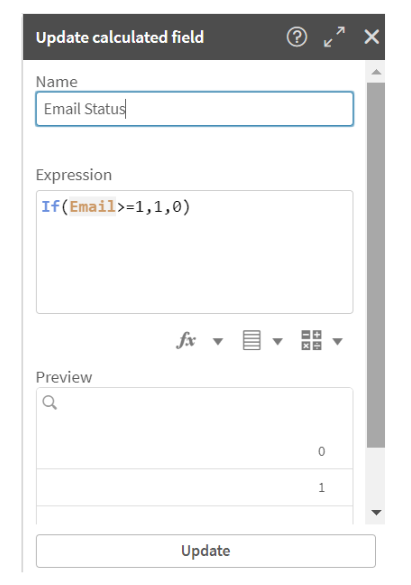
if(Age >= 60, 'Senior',

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

if(Age >= 18, 'Youth',

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

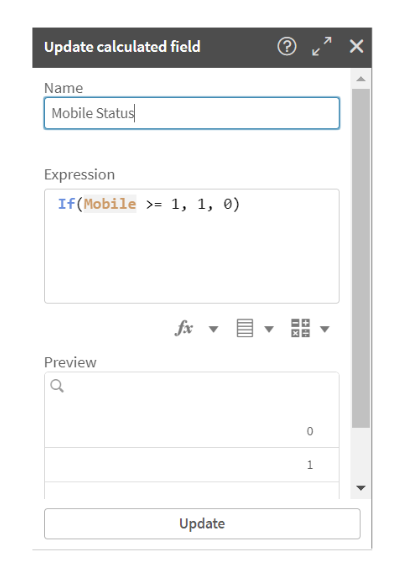
3)Email Status



Email status:

if(Email>= 1,1,0)

4)Moblie Status:



Mobile Status:

if(Mobile >=1,1,0)

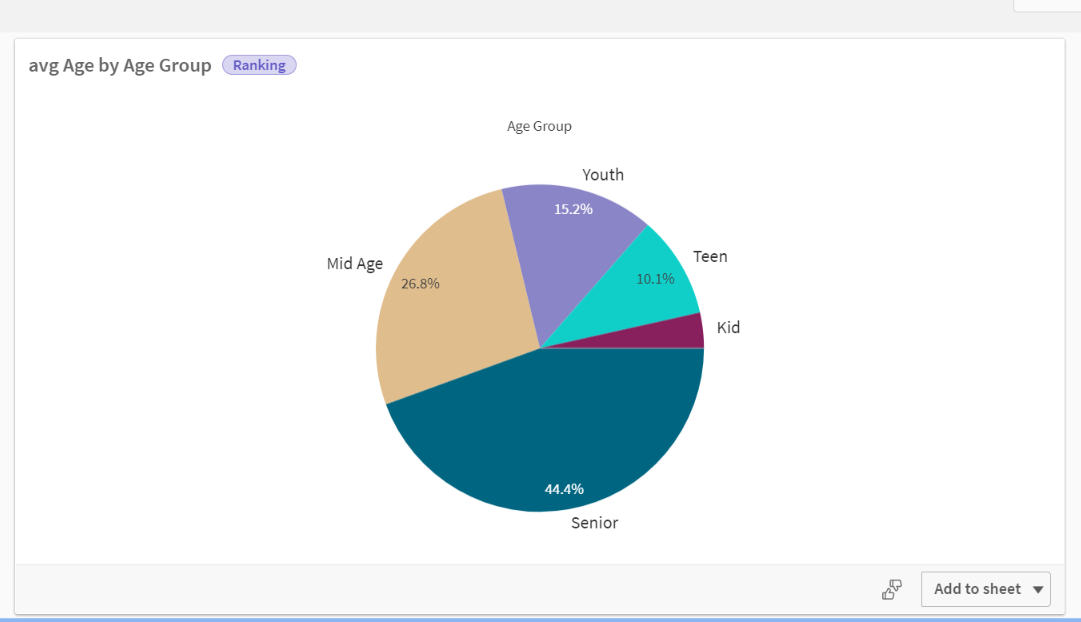
1. **Data Visualizations**

**5.1.** **Visualizations**

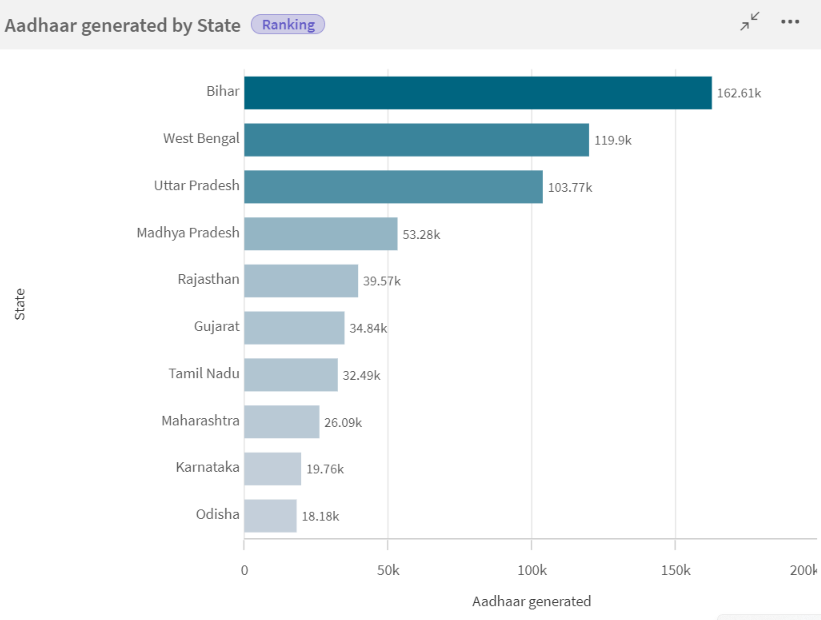
No Of Unique Visualizations:

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.

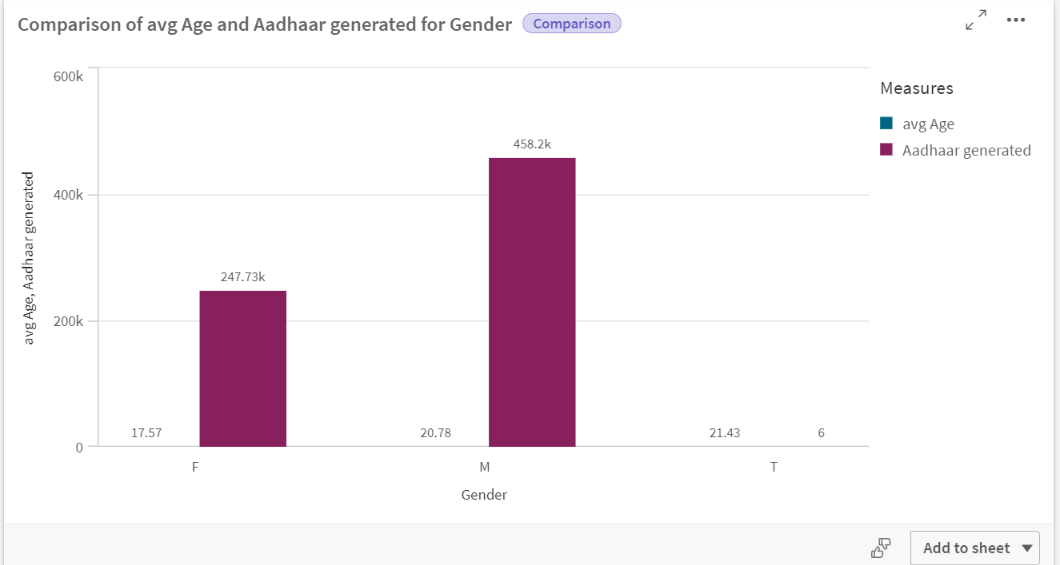
1) Aadhar Issued By Age Group :



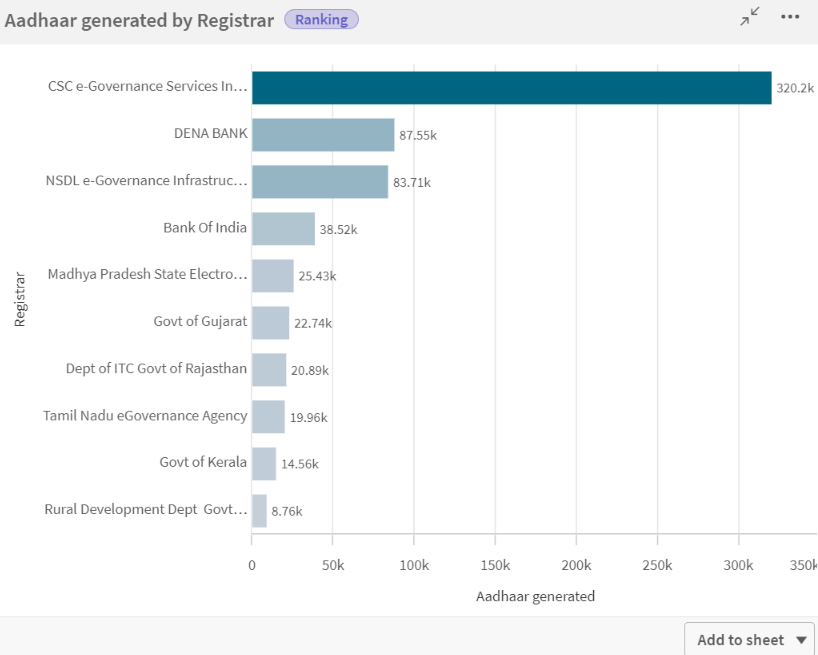
2) Top 10 Aadhar Generated States :



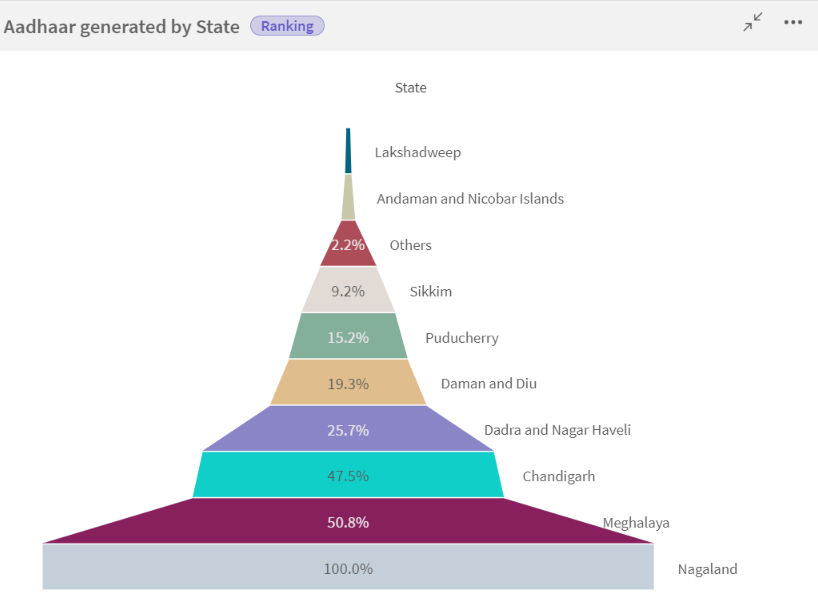
3) Aadhar Issued by Gender :



4) Aadhar Generated by Registrar :



5) Bottom 10 Aadhar Generated by State :

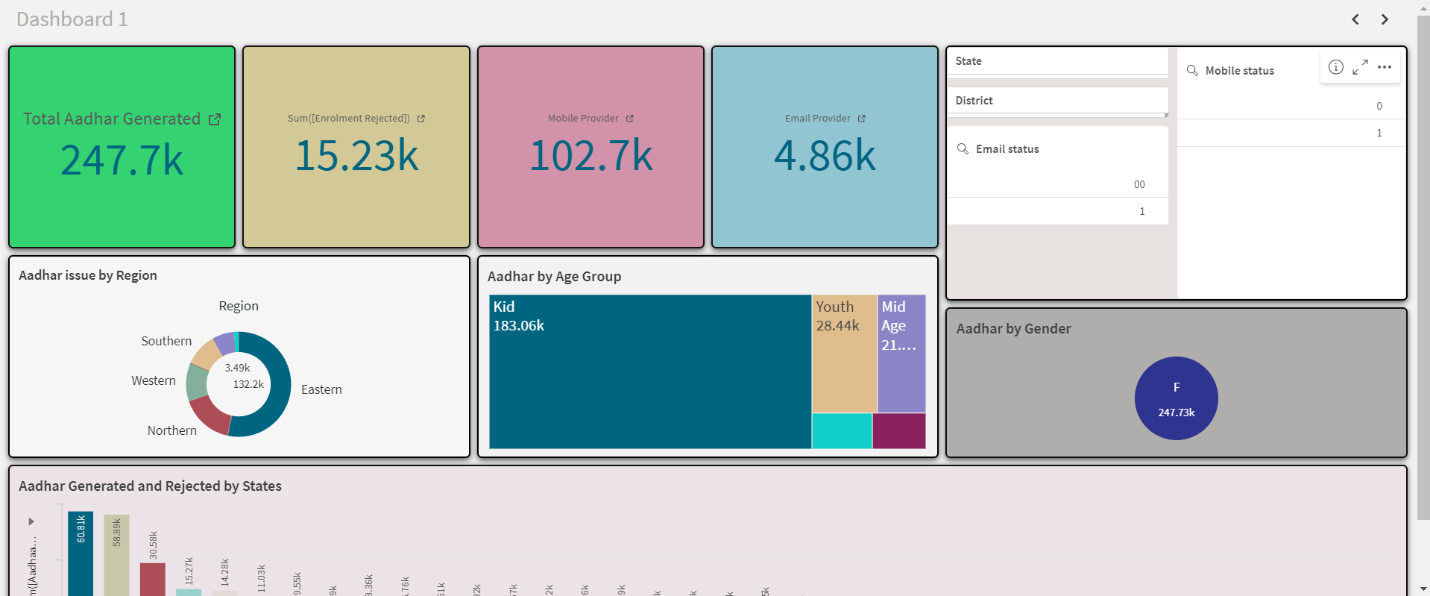


1. **Dashboard**

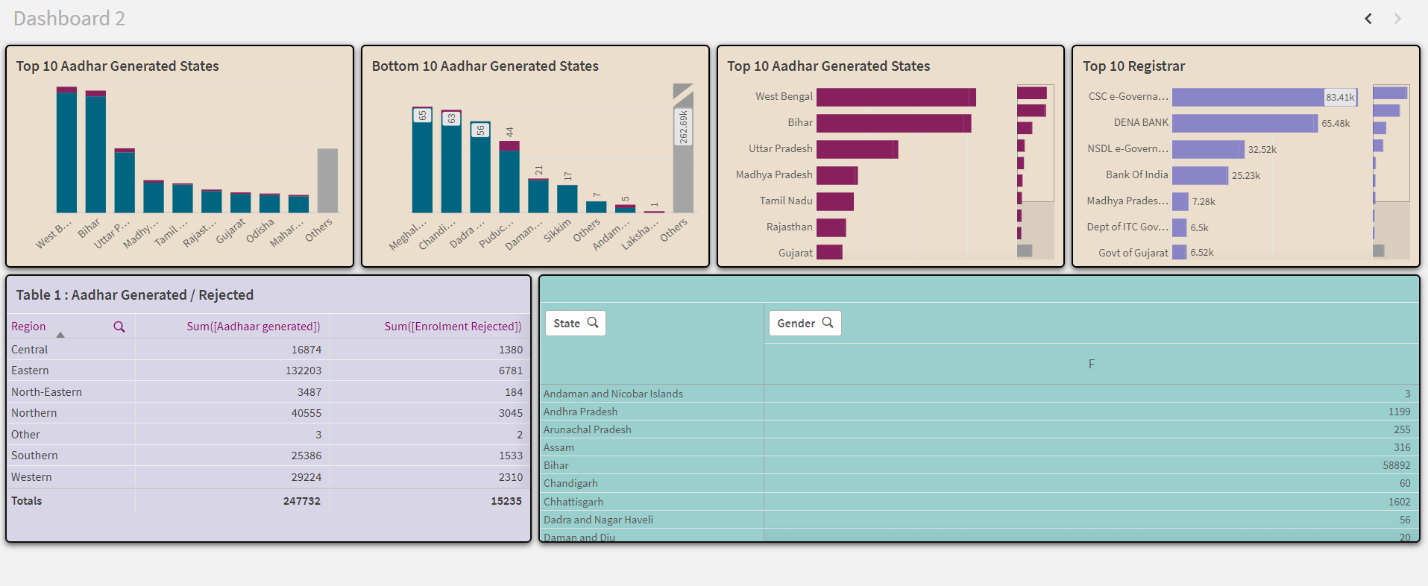
**6.1.** **Responsive and Design of Dashboard**

Qlik's dynamic dashboards stand out significantly compared to other forms of data visualization and business intelligence by offering immersive dashboards that narrate a story through various embedded visualizations, providing decision-makers with in-depth access to their datasets. These highly interactive graphical approaches allow users to filter data, drill down into details, and explore further insights, transforming static charts into dynamic tools for data exploration, facilitating effective searches within large datasets, and making it easier to detect hidden trends. The ability to personalize the look and feel of the dashboards, including layout, style, and information presentation, tailored to the user’s needs and convenience, gives Qlik an edge over the competition. Users can reposition visualizations to highlight significant values and adjust color schemes for optimized contrast and visibility. Additionally, Qlik’s dashboards allow real-time exploration of each displayed metric, providing decision-makers with actionable insights for quick and informed decisions based on current data. Integration, a fundamental strategy within Qlik’s dashboarding environment, facilitates collaborative problem-solving and solution-sharing, enabling multiple users to work on a single dashboard simultaneously and transforming organizations into data-driven institutions. Anticipating the need for accessibility across various devices, Qlik extends dashboard functionality to mobile platforms, ensuring users can access their data anytime and from any location, aiding decision-making even when not at workstations. Qlik’s dashboards are dynamic tools that turn information into insights, allowing users to leverage their data assets optimally and transforming raw data into valuable insights seamlessly. Qlik's dashboards are transformative tools that empower users to extract actionable insights, drive strategic decision-making, and unlock the full potential of their data assets, making the journey from raw data to actionable intelligence truly enlightening. Some of the dashboards created on Qlik Cloud using Aadhaar data include:

Dashboard 1 :



Dashboard 2 :



1. **Report**

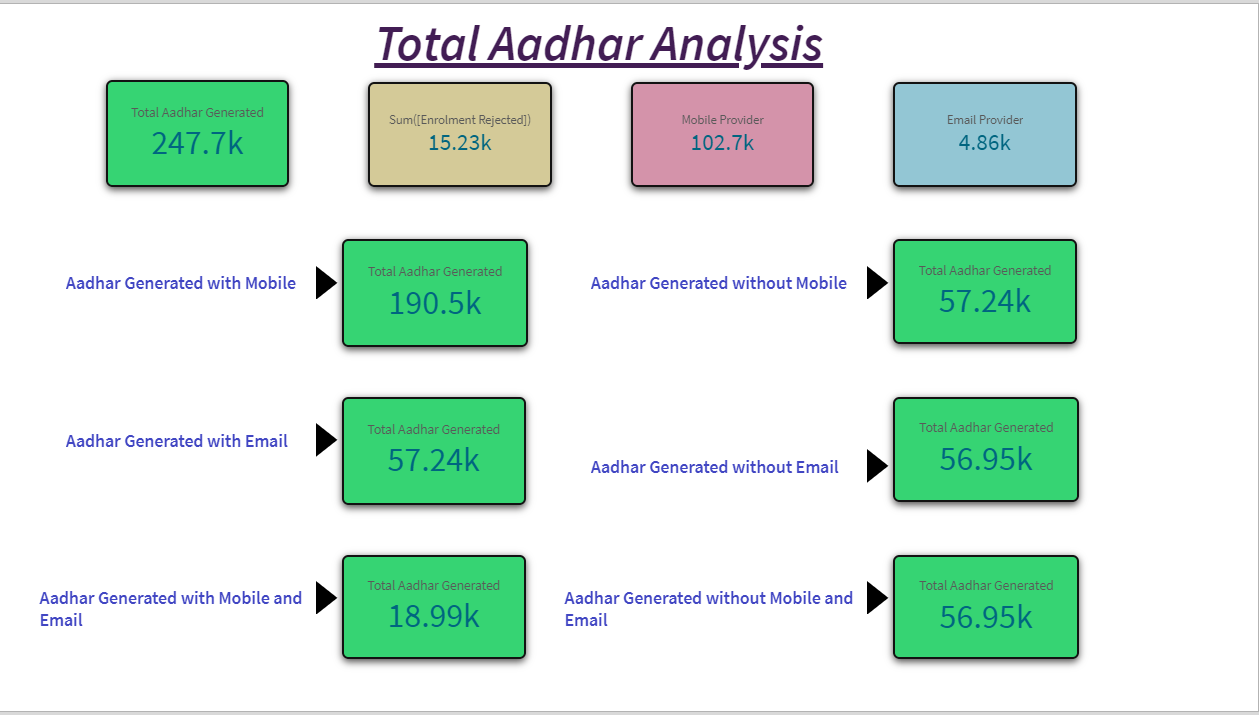
**7.1.** **Story Creation**

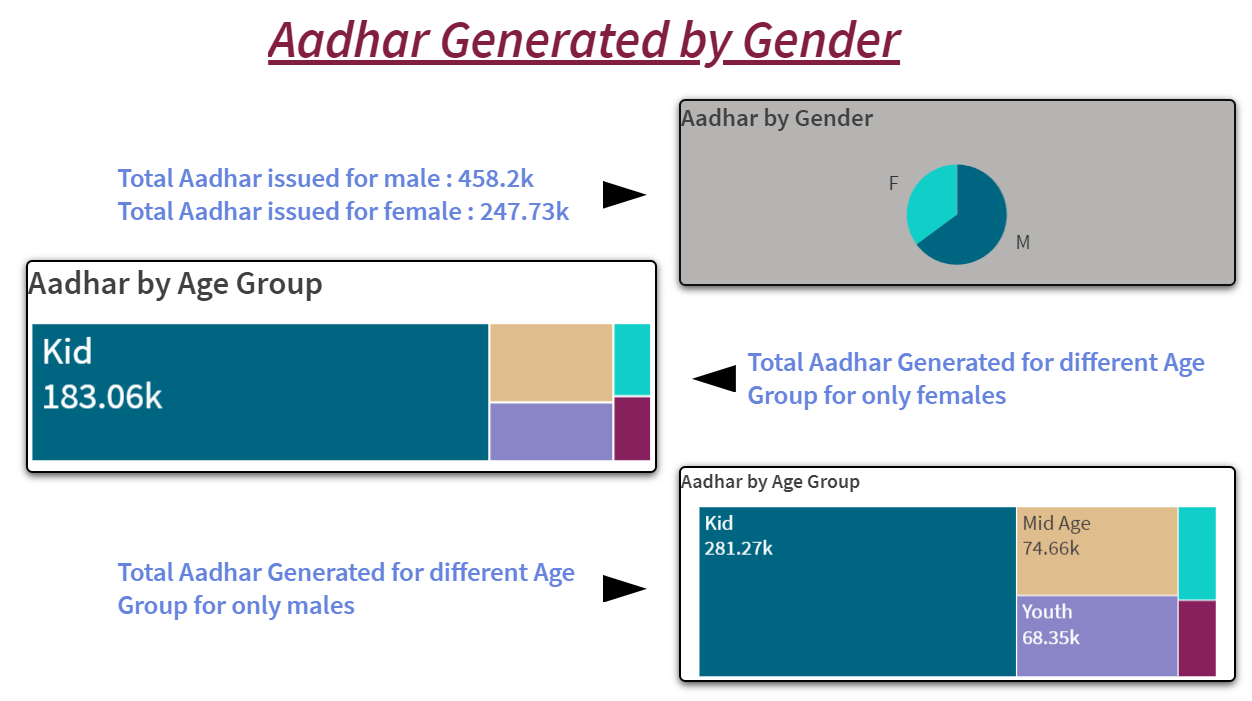
In 2009, Qlik began establishing its presence within DataWorks, marking the start of a significant transformation. Gone were the days of cumbersome, overwritten spreadsheets and basic, limited reports. In their place emerged visually stunning dashboards, vibrant with color, activity, and vitality. Sales trends flowed like a dancer’s movements, customer behavior was vividly depicted in numerous charts and graphs, and operational anomalies were laid bare for all to see.

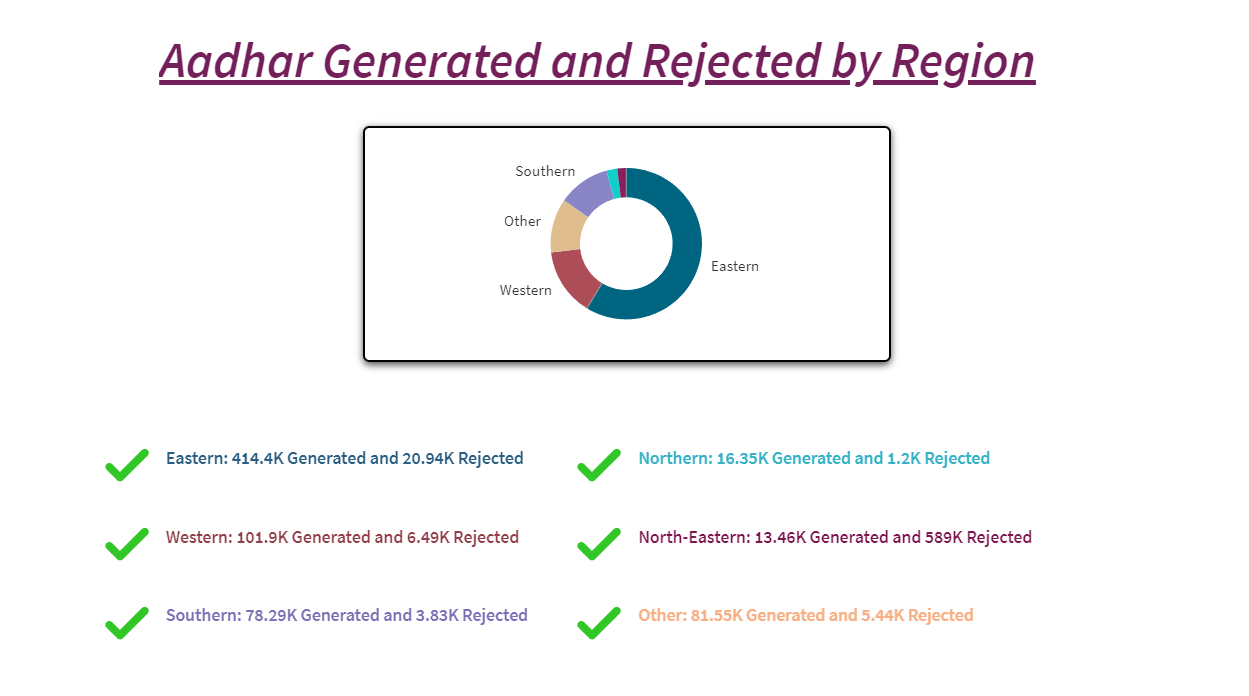
Qlik made a profound difference—not just as an enabler, but as a catalyst for cooperation and creativity. Managerial departments, once siloed and isolated, began communicating in a common language based on data. Marketing and sales teams coordinated to identify new opportunities, while operations teams optimized processes to facilitate real-time analytics. Employees and managers saved time and effort in data processing, and executives gained clearer vision and confidence in decision-making, thanks to the illuminating power of Qlik’s dashboards.

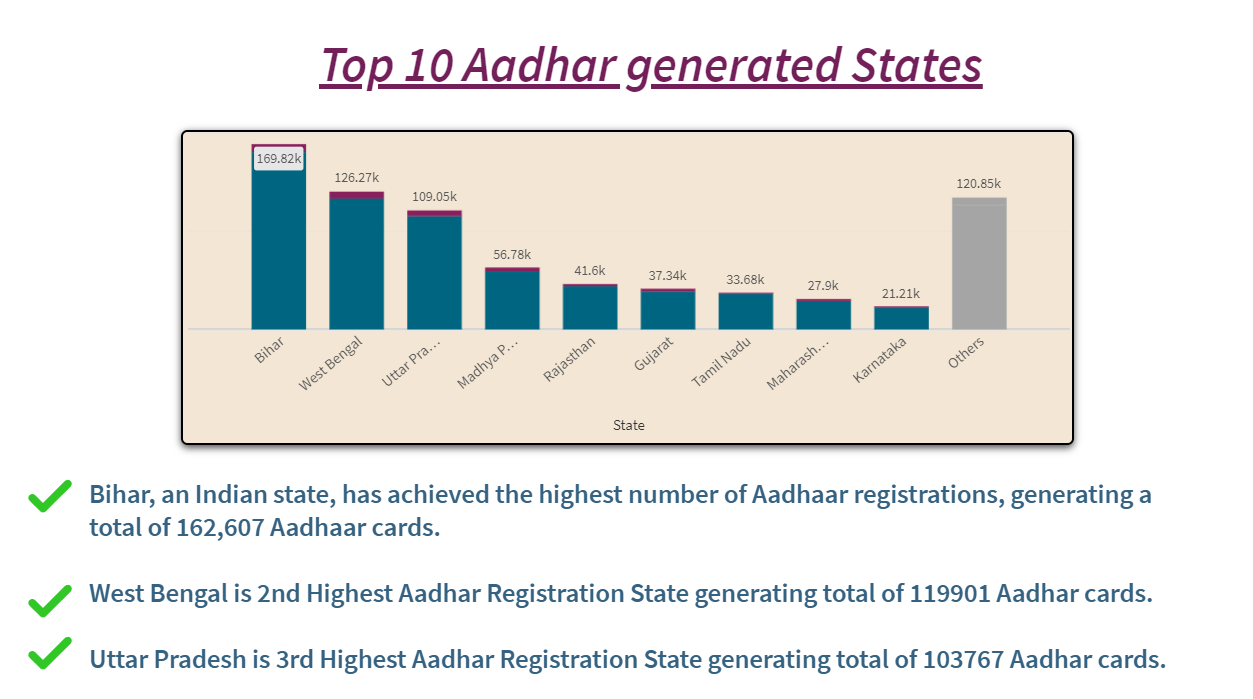
As news of DataWorks’ transformation spread, so did their reputation as frontrunners in data engineering. Businesses flocked to leverage Qlik’s innovative style for their own needs. Relationships flourished, and new opportunities beckoned on a larger horizon.

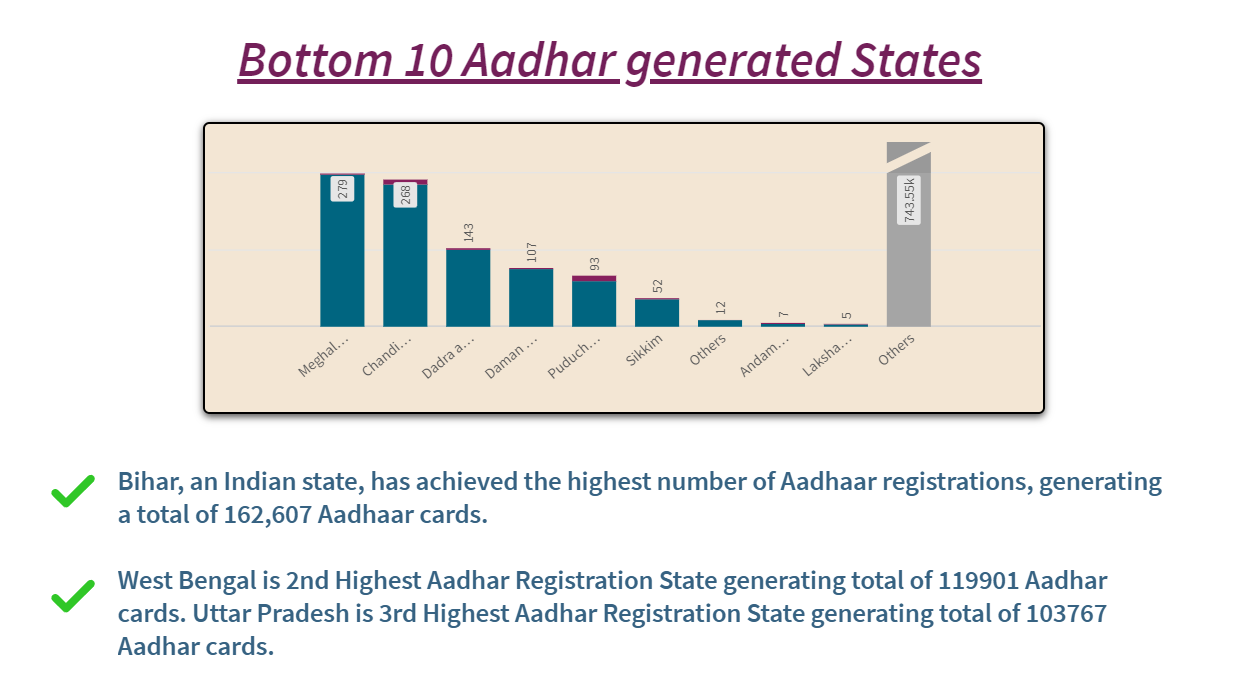
The story of DataWorks and Qlik is a testament to the power of innovative and exquisite data utilization. It demonstrates the vast opportunities that arise when experienced leadership is paired with superior technology. In today’s ever-changing world, those who embrace analytics and visualization tools not only discover untapped opportunities but also become the catalysts for change in the data-driven era.

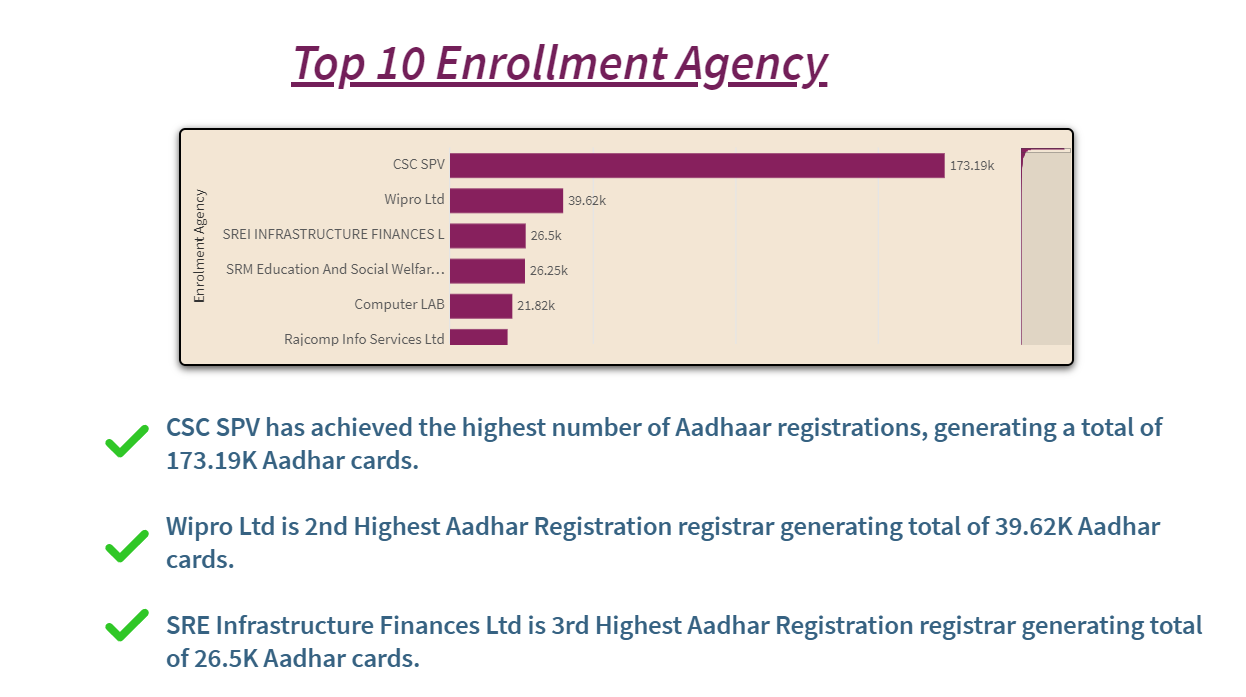


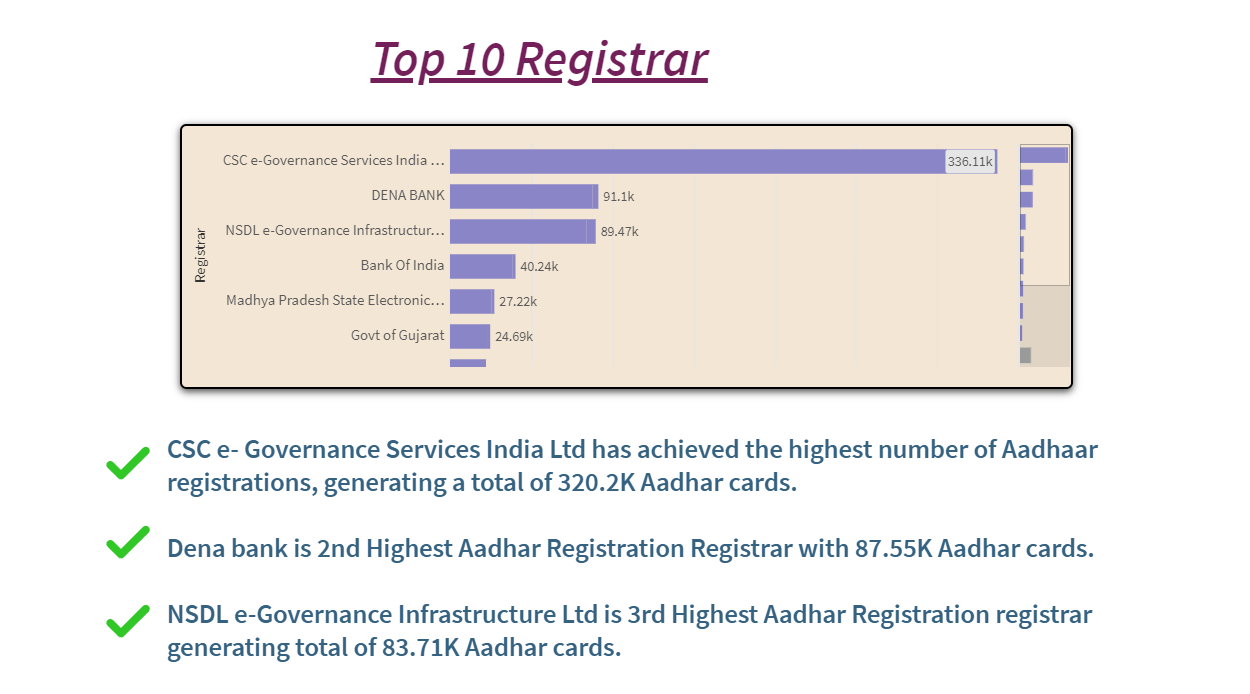












1. **Performance Testing**

**8.1.** **Amount of Data Rendered**

The "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 measures how much data has been successfully processed and made available for analysis, manipulation, or use within the system. Utilization of data filters involves applying specific criteria to a dataset to selectively include or exclude certain data points, crucial for focusing on relevant data and eliminating noise. This process allows analysts to hone in on pertinent information, leading to more accurate and meaningful insights. Filters can be applied across various dimensions such as time, geography, demographics, or transactions, enabling dynamic adjustments in real-time. In advanced tools like Qlik, filters enhance data quality and relevance in dashboards, supporting sophisticated analyses like cohort analysis, segmentation, and predictive modeling. By integrating multiple data sources, Qlik’s filtering capabilities provide a holistic view, ensuring analysts and decision-makers work with the most relevant data for accurate, actionable intelligence.

**8.2.** **Utilization of Data Filters**

Utilization of data filters involves applying specific criteria to a dataset to selectively include or exclude certain data points, crucial for focusing on relevant data and eliminating noise. This process allows analysts to hone in on pertinent information, leading to more accurate and meaningful insights. Filters can be applied across various dimensions such as time, geography, demographics, or transactions, enabling dynamic adjustments in real-time. In advanced tools like Qlik, filters enhance data quality and relevance in dashboards, supporting sophisticated analyses like cohort analysis, segmentation, and predictive modeling. By integrating multiple data sources, Qlik’s filtering capabilities provide a holistic view, ensuring analysts and decision-makers work with the most relevant data for accurate, actionable intelligence.