Data Insights From Aadhaar A Comprehensive Analysis Using Qlik

Category: Data Analytics
Skill : Business Intelligence

1. **Project Description:**

This project focuses on harnessing the power of Qlik, a leading data analytics and visualization platform, to perform an in-depth analysis of the Aadhaar dataset. Aadhaar is the world's largest biometric ID system, operated by the Government of India, providing unique identification to over a billion residents. This analysis aims to uncover valuable insights from this vast dataset to enhance understanding and drive data-driven decision-making.

2. Define Problem / Problem Understanding

2.1 Specify The 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 modeling the Aadhaar dataset, designing an interactive Qlik Sense dashboard Report , and extracting key visualizations 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

2.2 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 withregulations.

2.3 Literature Survey:

Skill Tags: Research, Data Analysis, Aadhaar, Literature Review.

The literature survey on Aadhaar analysis reveals a broad spectrum of research focusing on biometric authentication, data privacy, integration with public services, financial inclusion, and legal frameworks. Exploring these areas provides a comprehensive understanding of the

successes, challenges, and future prospects of Aadhaar. Future research should continue to address emerging concerns related to privacy, security, and ethical use of biometric data.

2.4   Social Or Business Impact :

Social Impact Analysis:

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

Business Impact Analysis:

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

3. 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.

3.1 Downloading The Dataset

Please use the link to download the dataset

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

3.2 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 enrollment centers, managing the enrollment process, and collecting necessary data.

2. Enrollment Agency: An Enrolment Agency is responsible for conducting the actual process of enrolling individuals into the Aadhaar 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 state

5. Sub-District: A sub-district, also known as taluka or tehsil 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. Aadhaar 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 provided or not

4. Data Preparation

4.1 Prepare The Data For Visualization

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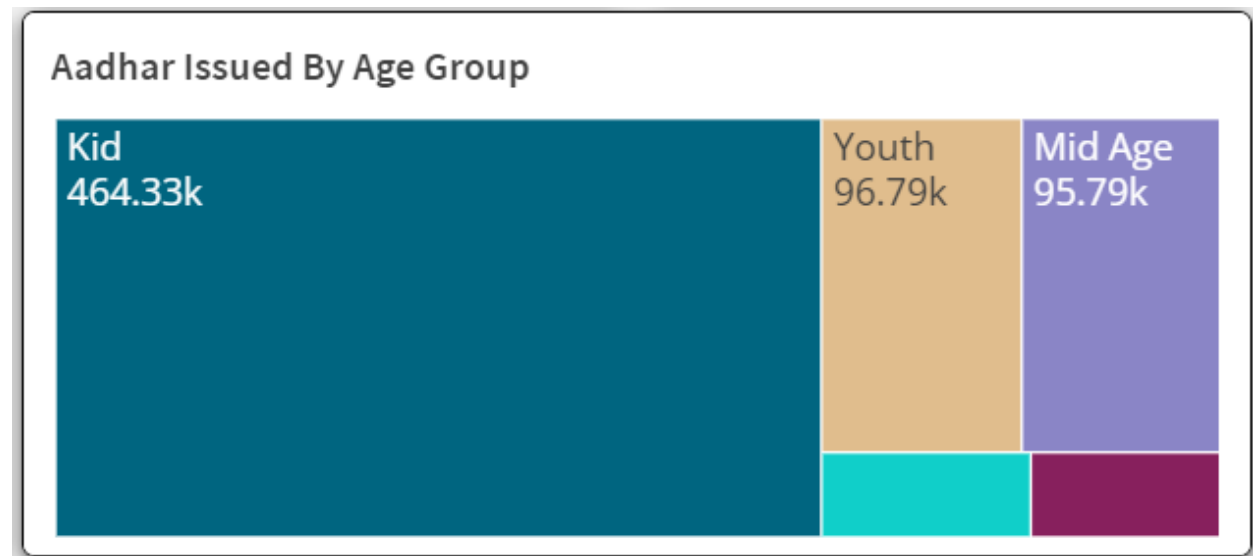
5 .Data Visualization

Data visualization is the process of creating graphical representations of data

to help people understand and explore the information. The goal of data visualization is to make complex data sets more accessible, intuitive, and easier to interpret. By using visual elements such as charts, graphs, and maps, data visualizations can help people quickly identify patterns, trends, and outliers in the data.
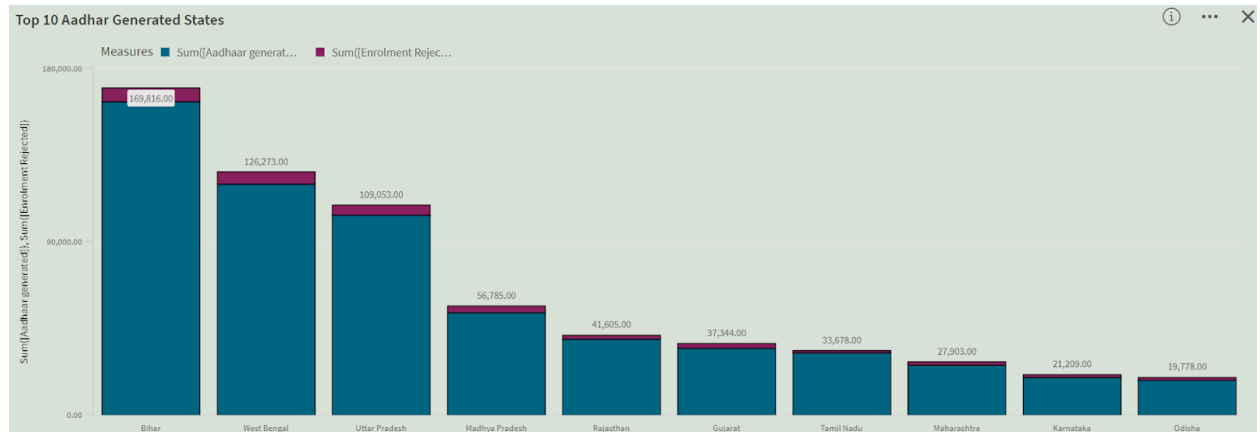
5.1. No Of Unique Visualizations

The number of unique visualizations that can be created with a given dataset. Some common types of visualizations that can be used to analyze the performance and efficiency of banks include bar charts, line charts, heat maps, scatter plots, pie charts, Maps, etc. These visualizations 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.
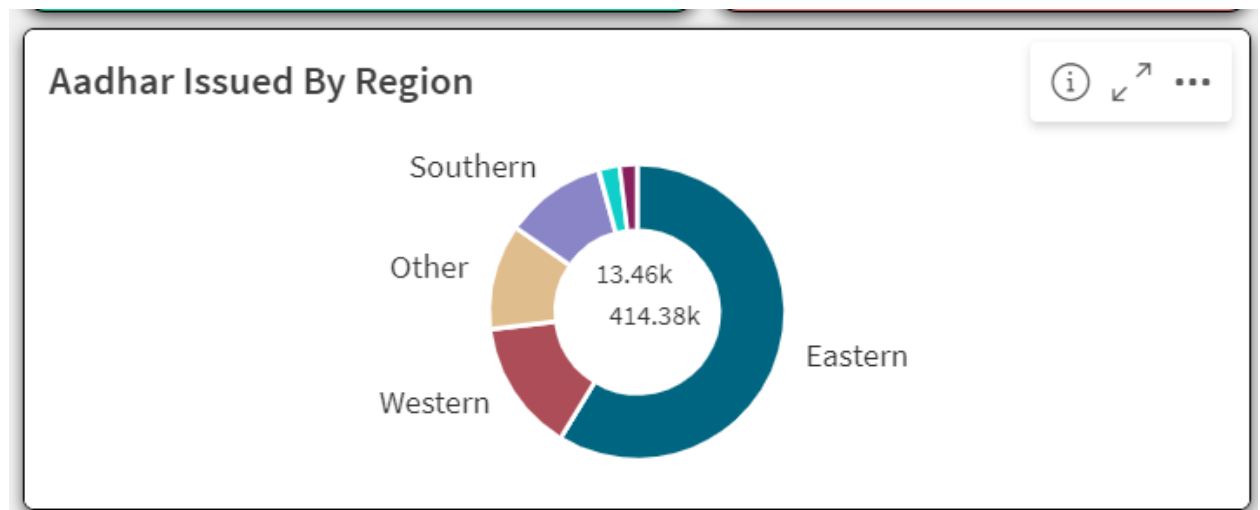
Aadhar Issued By Age Group

Top 10 Aadhar-Generated State

Top 10 Aadhar Generated States

Measures ■ Sum([Aadhaar generat... ■ Sum([Enrolment Rejec...

State	Value
Bihar	169,816.00
West Bengal	126,273.00
Uttar Pradesh	109,053.00
Madhya Pradesh	56,785.00
Rajasthan	41,605.00
Gujarat	37,344.00
Tamil Nadu	33,678.00
Maharashtra	27,903.00
Karnataka	21,209.00
Odisha	19,778.00

Aadhar Issued By Region

Aadhar Issued By Region

Southern

Other

Western

Eastern

13.46k
414.38k

Issue By Gender
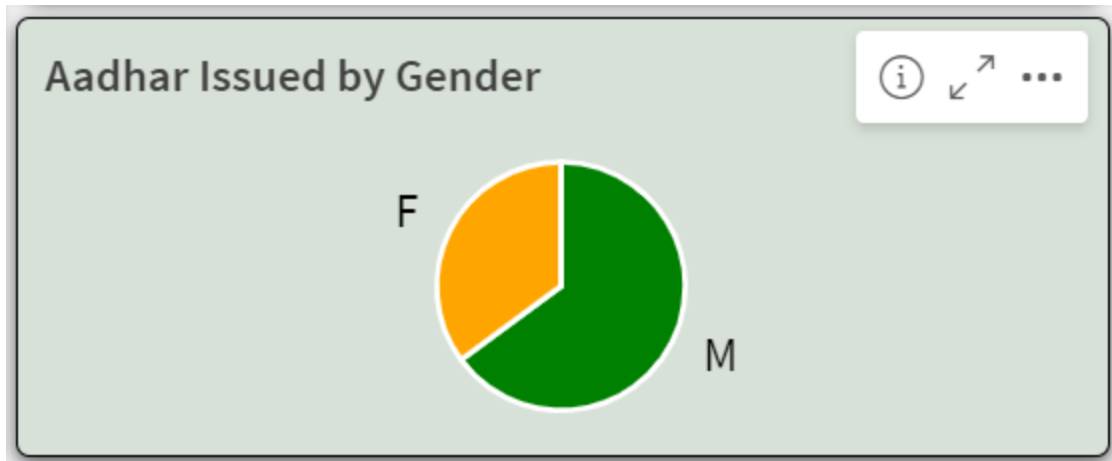
Top 10 Highest Generated Enrollment Agency

Top 10 Highest-Generated Registra

Bottom 10 Aadhar-Generated State
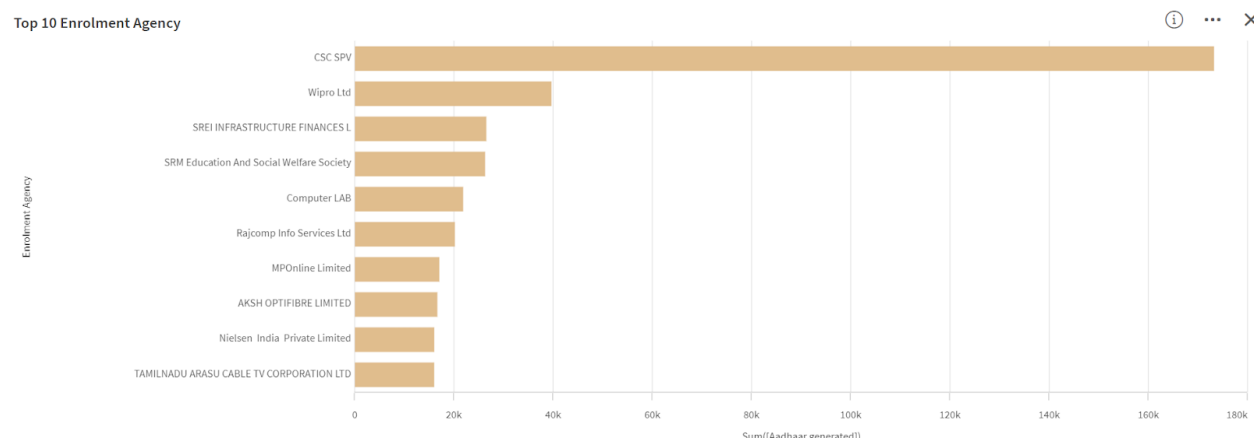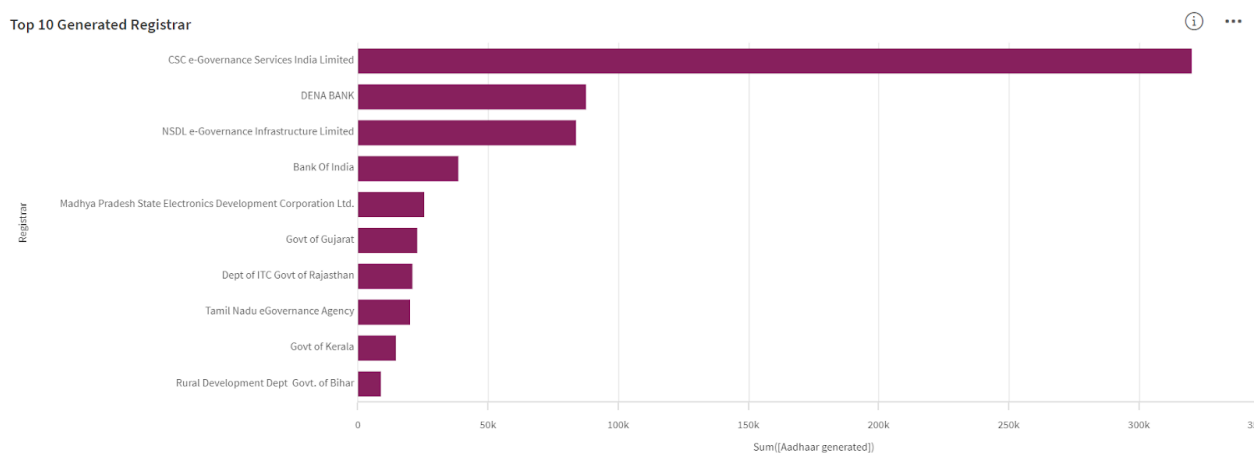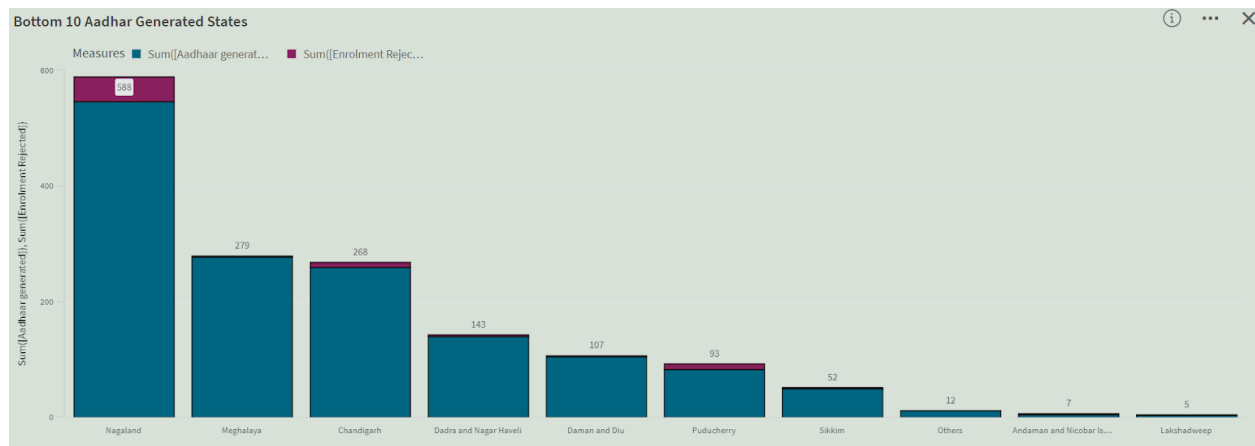
Bottom 10 Aadhar Generated States

6. Dashboard

A dashboard is a graphical user interface (GUI) that displays information and data in an organized, easy-to-read format. Dashboards are often used to provide real-time monitoring and analysis of data and are typically designed for a specific purpose or use case. Dashboards can be used in a variety of settings, such as business, finance, manufacturing, healthcare, and many other industries. They can be used to track key performance indicators (KPIs), monitor performance metrics, and display data in the form of charts, graphs, and tables.

Dashboard

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

State
District
Email Status

MobileStatus
0
1

Aadhar Issued By Region

Southern
Other
13.46k
414.38k
Western
Eastern

Aadhar Issued By Age Group

Kid
464.33k

Youth
96.79k

Mid Age
95.79k

Aadhar Issued by Gender

F
M

Aadhar Generated and Rejected By States

180k
90k
0

Dashboard2

Top 10 Aadhar Generated States

180,000.00
90,000.00
0.00

169,816.00
126,273.00
109,053.00
56,785.00
41,606.00
37,344.00
33,678.00
27,963.00
24,209.00
19,778.00

Bihar
West Ben...
Uttar Pra...
Madhya ...
Rajasthan
Gujarat
Tamil Nadu
Maharas...
Karnataka
Odisha

Bottom 10 Aadhar Generated States

600
300
0

568
279
268
143
107
93
52
12
7
5

Nagaland
Meghalaya
Chandig...
Dadra an...
Daman a...
Puduche...
Sikkim
Other
Andama...
Lakshad...

Top 10 Enrolment Agency

CSC SPV
Wipro Ltd
SREI INFRAST...
SRM Educatio...
Computer LAB

0   90k   180k
Sum([Aadhaar generated])

Top 10 Generated Registrar

CSC e-Govern...
DENA BANK
NSDL e-Gover...
Bank Of India
Madhya Prad...

0   175k   350k
Sum([Aadhaar generated])

Table 1- Aadhar Generated/Rejected By Region

Region	Aadhar Generated	Enrolment Rejected
Eastern	414375	20942
North-Eastern	13458	589
Northern	16350	1289
Other	81554	5438
Southern	78291	3833
Western	101911	6485
Totals	705939	38576

Table 2- Aadhar Generated by Gender

State   Gender

|	M	F	T
Karnataka	13003	6761	0
Assam	2897	316	-
Meghalaya	213	64	-
Mizoram	4061	2218	-
Haryana	5109	1695	-
Himachal Pradesh	1053	494	-

7. Story

A data story is a way of presenting data and analysis in a narrative format, with the goal of making the information more engaging and easier to understand. A data story typically includes a clear introduction that sets the stage and explains the context for the data, a body that presents the data and analysis in a logical and systematic way, and a conclusion that summarizes the key findings and highlights their implications. Data stories can be told using a variety of mediums, such as reports, presentations, interactive visualizations, and videos.

Total Aadhar Analysis

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

Aadhar Generated with Mobile ▶ Total Aadhar Generated **144.4k**

Aadhar Generated without Mobile ▶ Total Aadhar Generated **561.5k**

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 issue for Male: **458.2K**
Total Aadhar issued for Female: **243.73K** ▶
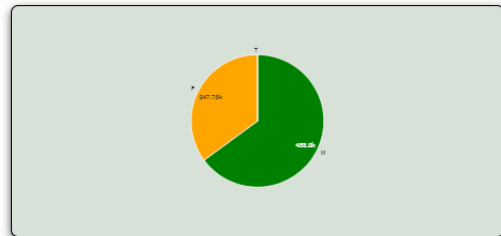
◀ Total Aadhar Generated for Different Age Group for only Females

Kid 183.06k	Youth 28.44k	Mid Age 21.13k
	Senior 7.99k	Teen 7.12k

Total Aadhar Generated for Different Age Group for only Males ▶

Kid 281.27k	Mid Age 74.66k	Teen 18.64k
	Youth 68.35k	Senior 15.28k

Aadhar Generated and Rejected by Regions
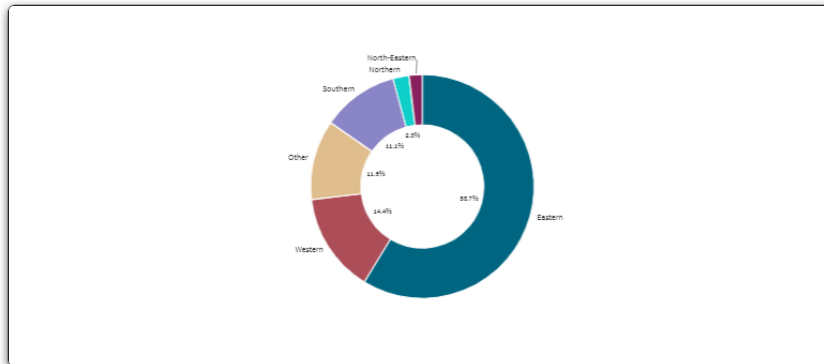
✔ Eastern: **414.4K** Generated and **20.94K** Rejected   ✔ Northern:**16.35K** Generated and **1.2K** Rejected

✔ Western: **101.9K** Generated and **6.49K** Rejected   ✔ North-Eastern: **13.46K** Generated and **589** Rejected

✔ Southern: **78.29K** Generated and **3.83K** Rejected   ✔ Other: **81.55K** Generated and **5.44K** Rejected

8 Performance Testing

8.1 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

8.2 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.
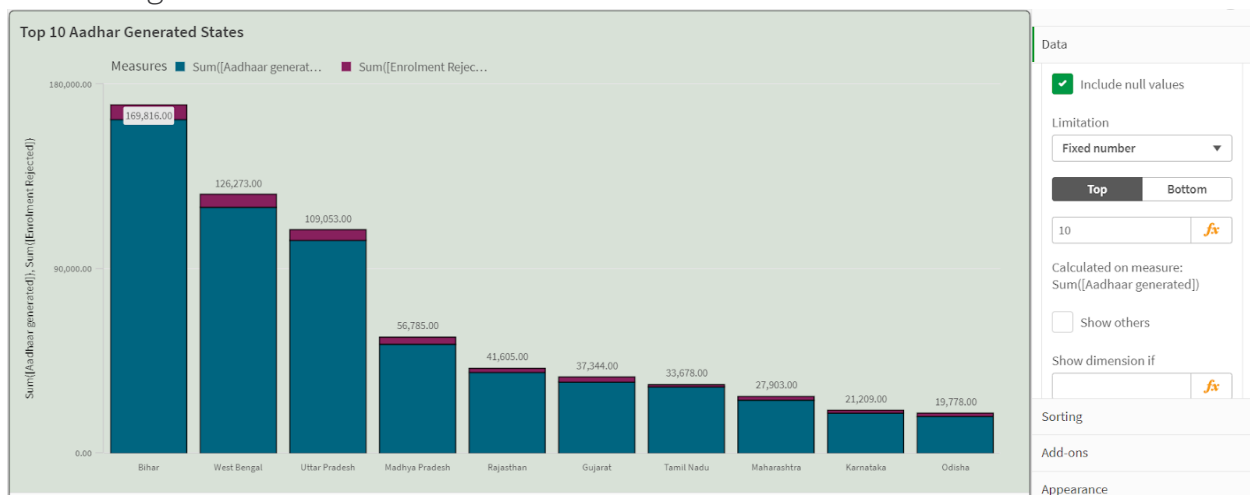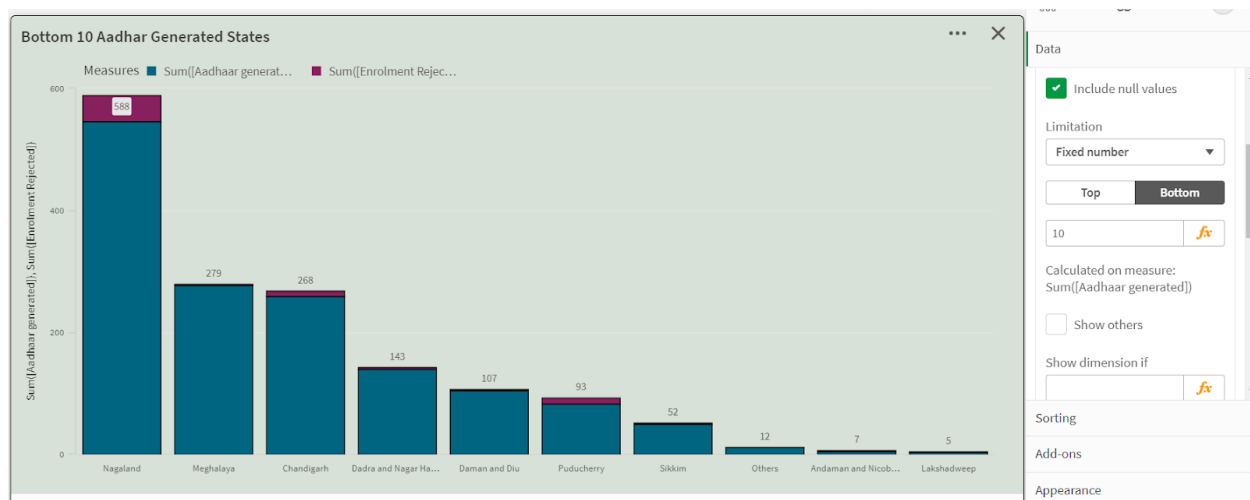
8.3 No Of Calculation Fields

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.

Update calculated field

Name

Age Group

Expression

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

fx ▼   ▦ ▼   ▦ ▼

Preview

🔍

Kid

Age Group:

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

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')))))

Update calculated field

Name

Mobile Status

Expression

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

fx ▼   ▤ ▼   🔢 ▼

Preview

🔍

	0
	1

Update

Update calculated field

Name

Email Status

Expression

```
If(Email>=1,1,0)
```

fx ▾   ▾   ▾

Preview

🔍

	0
	1

Update

8.3 No Of Visualizations/ Graphs

1. Aadhar Issued By Age Group
2. Top 10 Aadhar generated States
3. Aadhar Generated By region
4. Aadhar Generated By gender
5. Aadhar issued By Mobile
6. Top 10 Highest Aadhar Generated Registrar
7. Bottom10 Aadhar generated States