

**PROJECT TITLE**

**Data Insights From Aadhaar A Comprehensive**

**Analysis Using Qlik**

**PROJECT DONE BY**

ISRAEL OSIPILLI(B-TECH)

**Data Insights From Aadhaar A Comprehensive**

**Analysis Using Qlik**

**Project Flow**

To accomplish this, we have to complete all the activities listed below,

* Define Problem / Problem Understanding
  + Specify the business problem
  + Business requirements
  + Literature Survey
  + Social or Business Impact.
* Data Collection
  + Collect the dataset,
  + Connect Data with Qlik Sense
* Data Preparation
* Prepare the Data for Visualization
* Data Visualizations
  + No of Unique Visualizations
* Dashboard
  + Responsive and Design of Dashboard
* Story
  + Story Creation
* Performance Testing
  + Amount of Data Rendered to DB ‘
  + Utilization of Data Filters
  + No of Calculation Fields
  + No of Visualizations/ Graphs
* Project Demonstration & Documentation
  + Record explanation Video for project end to end solution
  + Project Documentation-Step by step project development procedure

**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

**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 analyzing 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 Aadhaar analysis.

**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.

**Downloading The Dataset**

Please use the link to download the dataset

abc.csv - Google Drive..

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

For Using This Data Set We Can See the OutPut of the Adhar users.

**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 a 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

**Data Preparation**

**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.

Data Preparation.mp4 - Google Drive..

<https://drive.google.com/file/d/1FvImqVlGyBNED8MlG6LS9TXJ-ZjueMGO/view?usp=sharing>

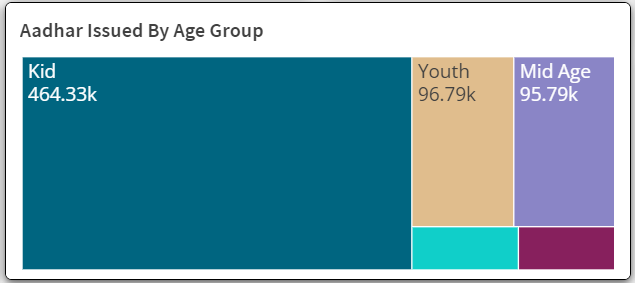
**Data Visualization**

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.

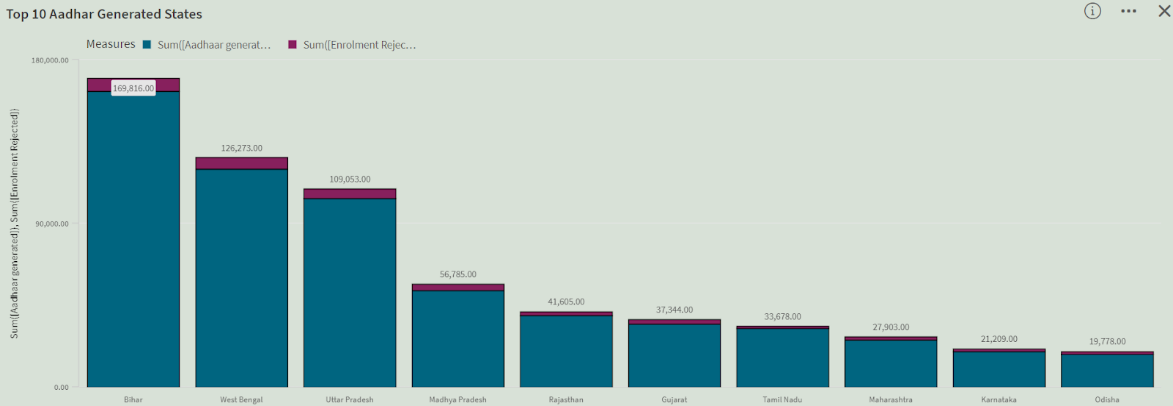
### **No Of Unique Visualizations**

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.

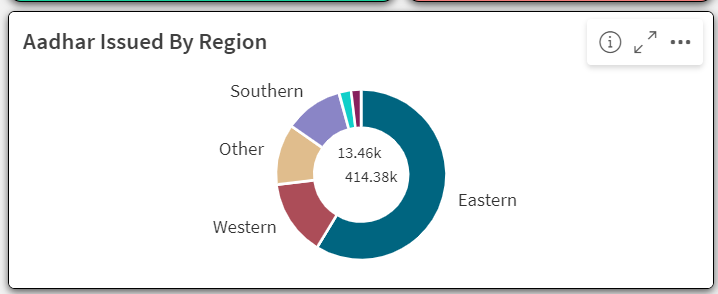
                                Aadhar Issued By Age Group



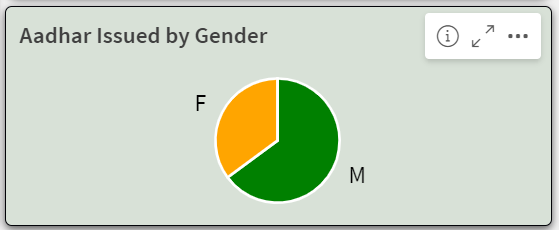
Top 10 Aadhar-Generated States



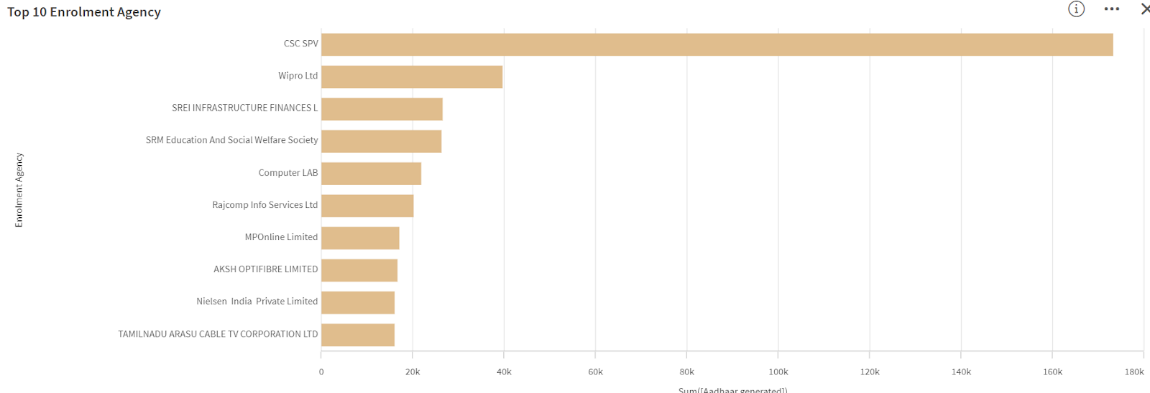
 Aadhar Issued By Region



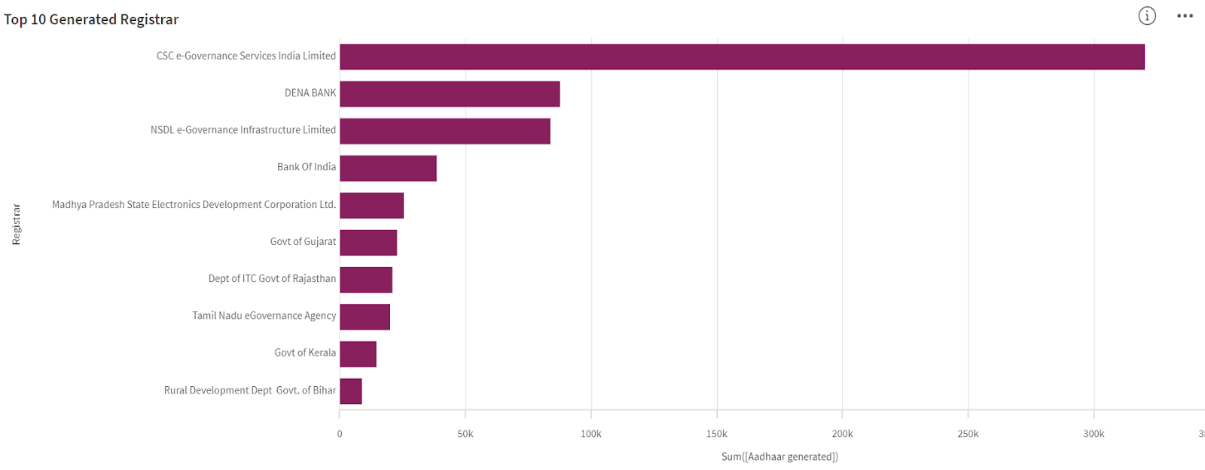
 Issue By Gender



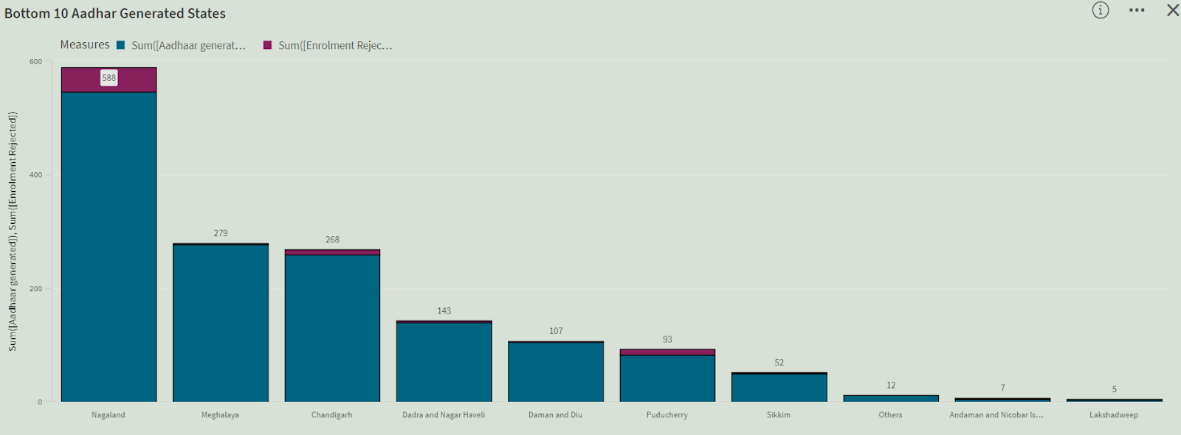
 Top 10 Highest Generated Enrollment Agency



 Top 10 Highest-Generated Registrar



 Bottom 10 Aadhar-Generated State



**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.

**Responsive And Design Of Dashboard**

Explanation video link Part 1:

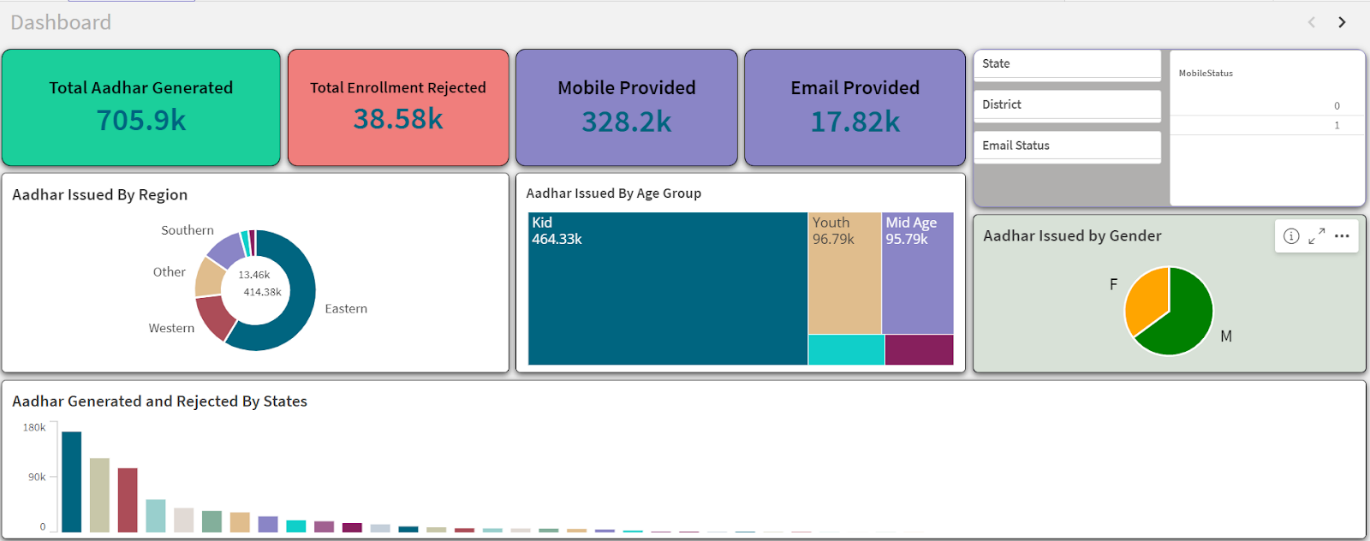
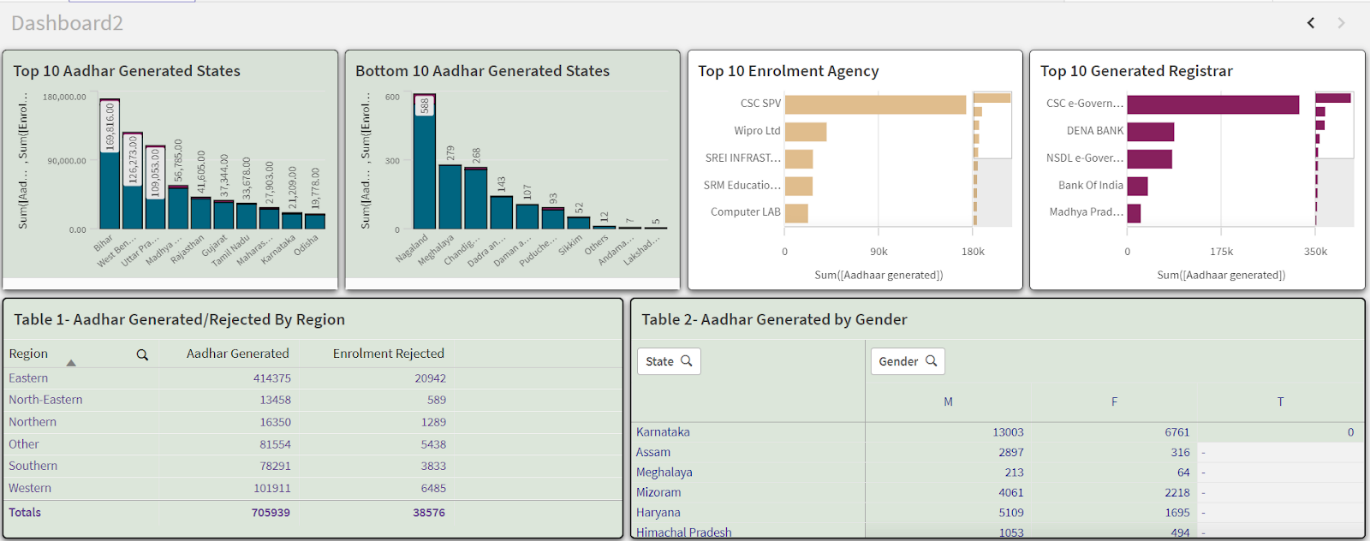
<https://drive.google.com/file/d/1wxDMZvuL0O2rOaNaya_4kMgb5zCKQ-v0/view?usp=sharing>

Explanation video link Part 2:

<https://drive.google.com/file/d/1WBgvK156MgfoH47z8elOCrG13zF0MTOf/view?usp=sharing>

Explanation video link Part 3:

<https://drive.google.com/file/d/1cSEIZ3l18tKsnaZUTFAHjuS4z0FN2dQt/view?usp=sharing>

Explanation video link Part 4: <https://drive.google.com/file/d/1TvH02rAcAoh0EAQXUgjGJ0evvwAPm-X8/view?usp=sharing> 

**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.

**Design Of Story**

Explanation video link Part 1:

<https://drive.google.com/file/d/1LF5mOTnecMIyEmUlbSVa321m7Rw5N8RN/view?usp=sharing>

Explanation video link Part 2

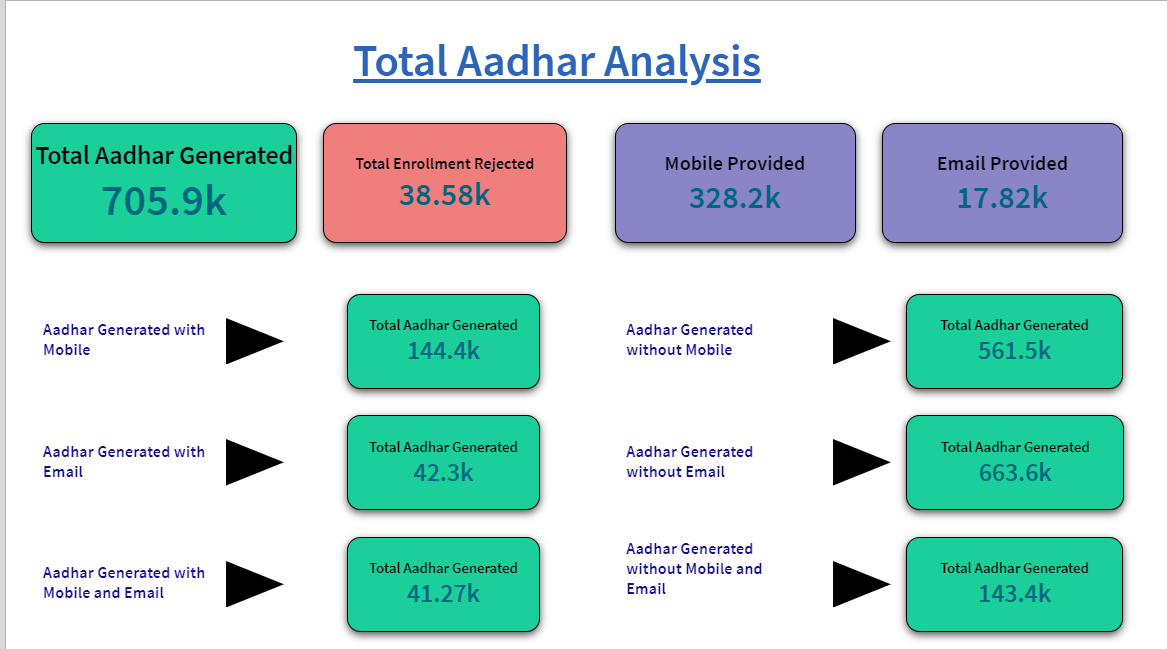
<https://drive.google.com/file/d/1oFZtHTl7JeA4xFLxjLQYhr0oWvXHyvUl/view?usp=sharing>

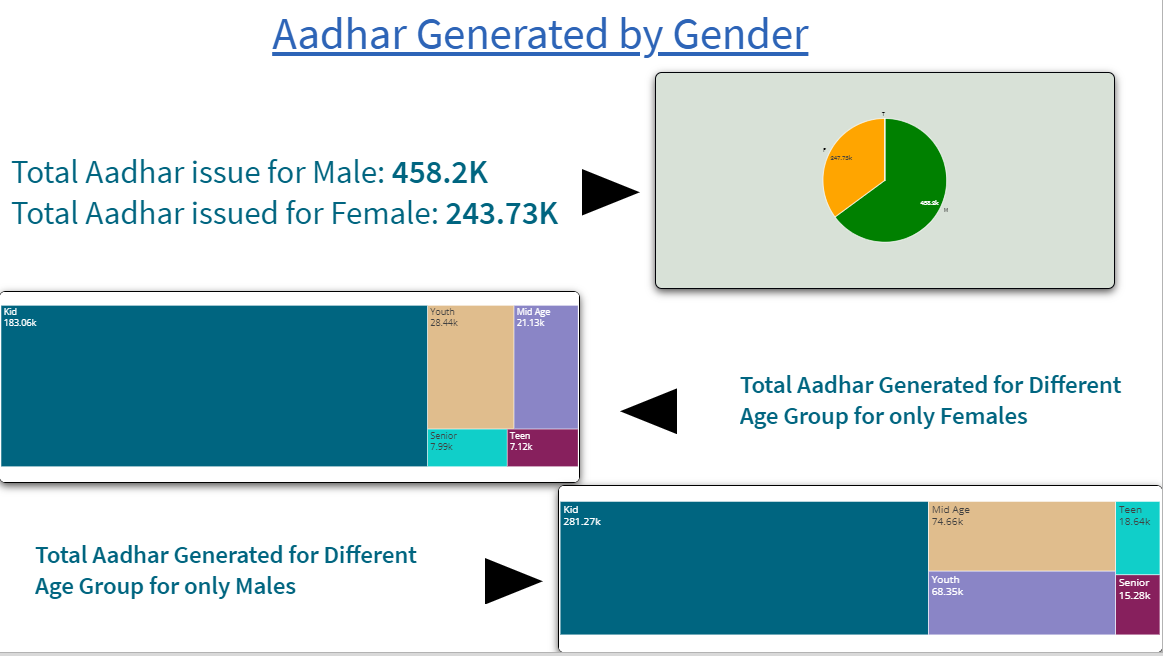
Explanation video link Part 3

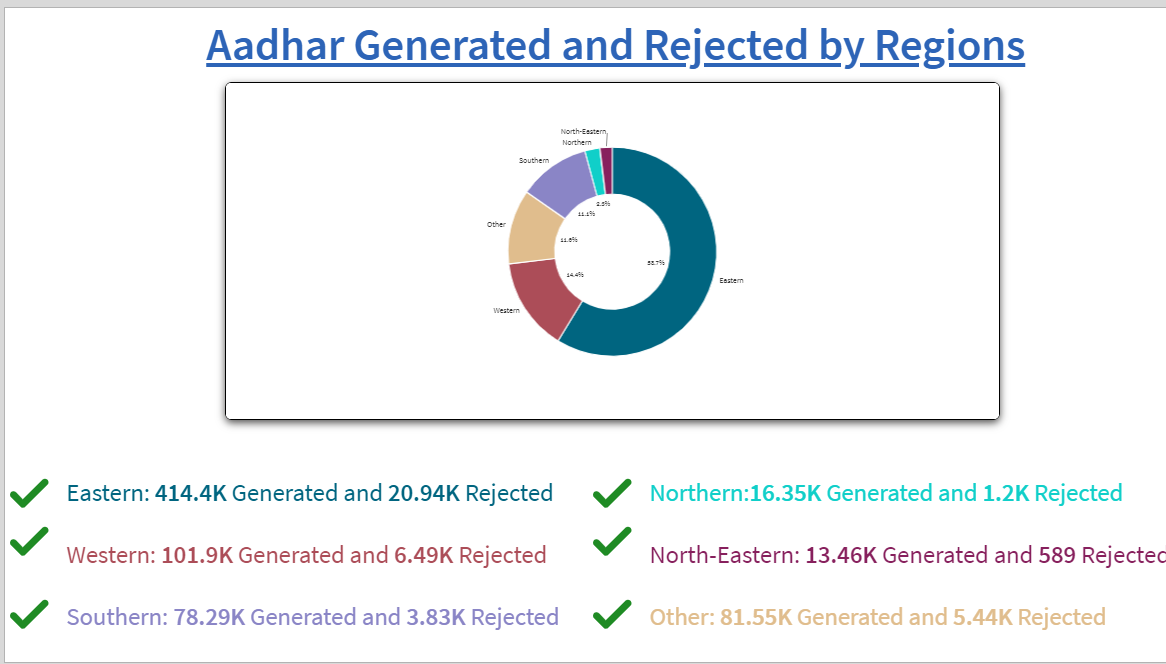
<https://drive.google.com/file/d/1gzDMHxj67RHfFSsyIwZZVXki_1odYtRf/view?usp=sharing>

Explanation video link Part 4

<https://drive.google.com/file/d/1zG0fCutG0DWn2GOOtmL-gWqjhpYYkeRr/view?usp=sharing>



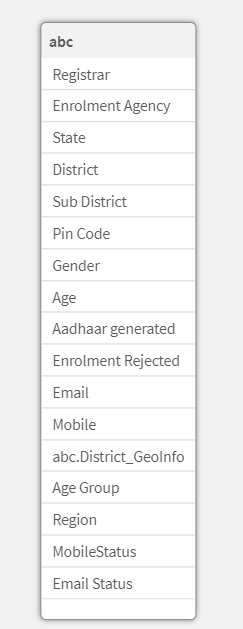




### **Performance Testing**

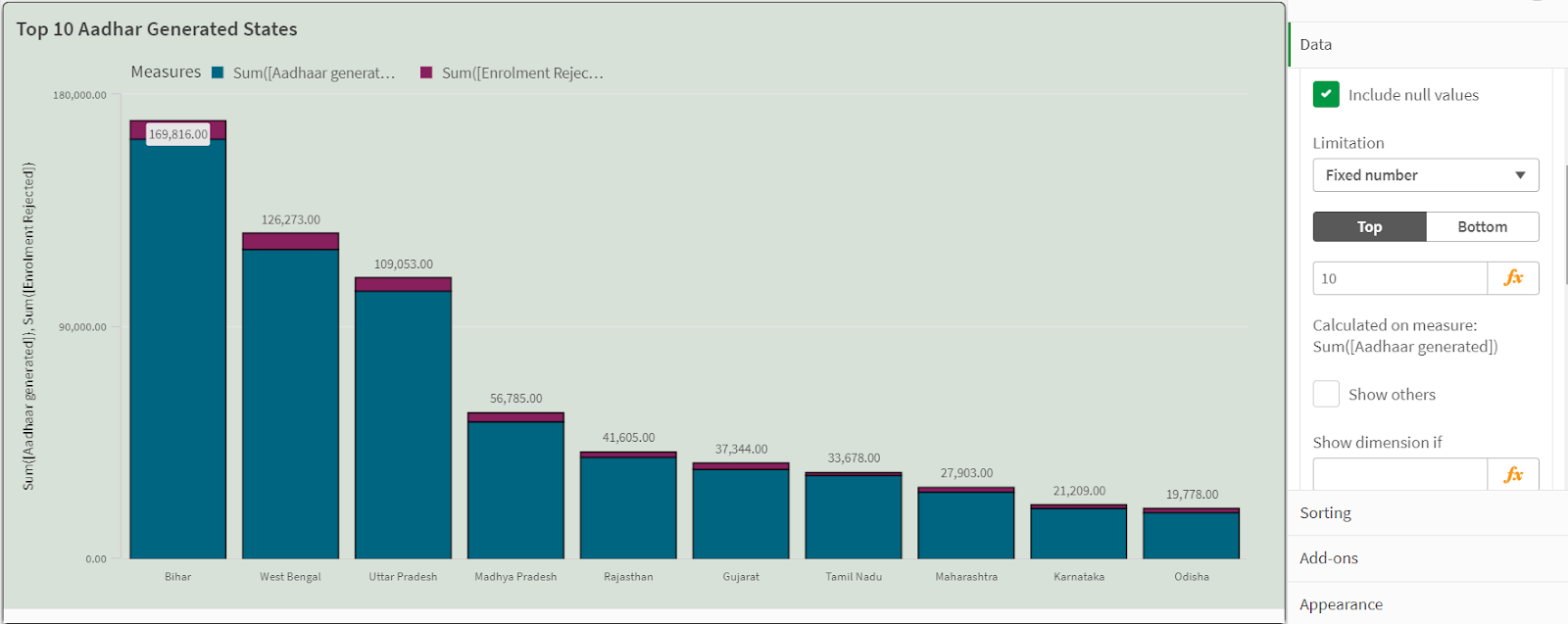
**Amount Of Data Loaded**

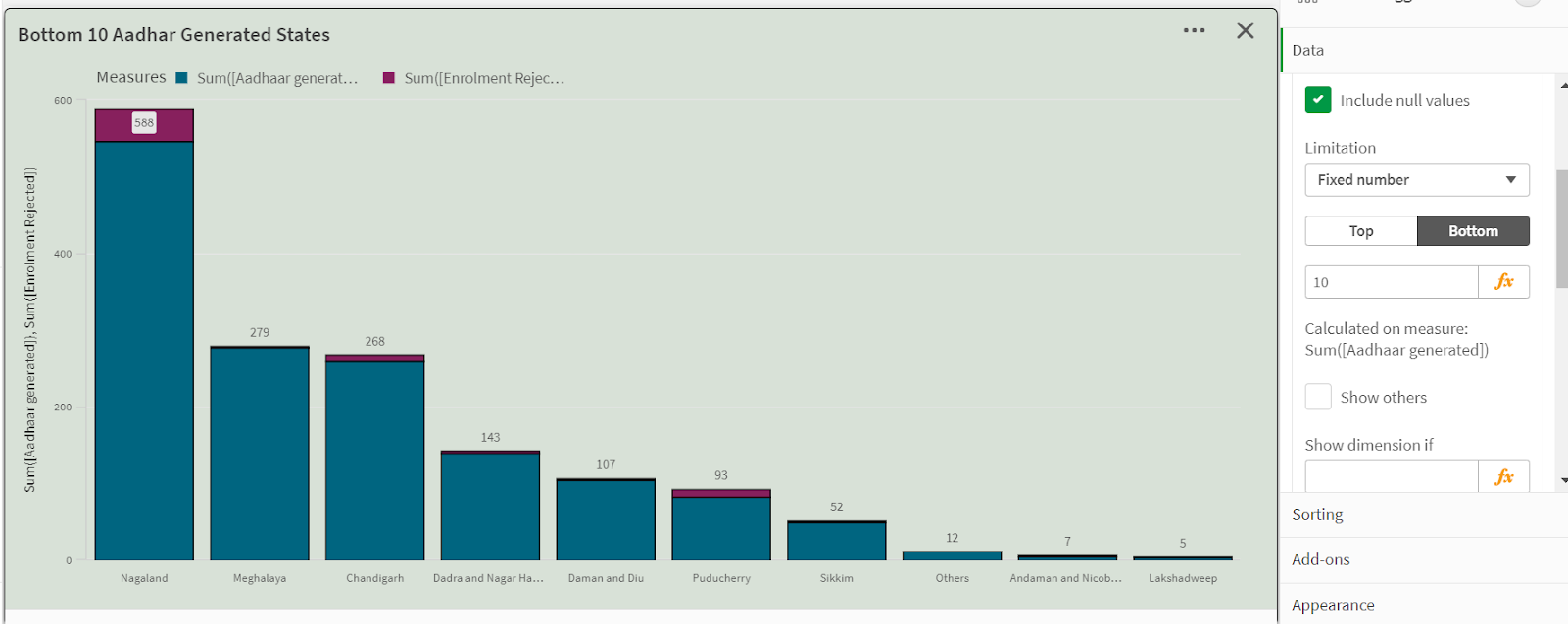
"Amount of Data Loaded" refers to the quantity or volume of data that has been imported, retrieved, or loaded into a system, software application, database, or any other data storage or processing environment. It's a measure of how much data has been successfully processed and made available for analysis, manipulation, or use within the system



**Utilization Of Data Filters**

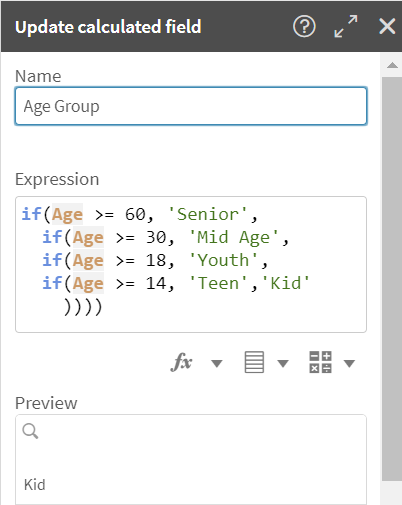
Utilization of data filters refers to the process of applying specific criteria or conditions to a dataset in order to selectively include or exclude certain data points. This filtering process is crucial in data analysis as it allows to focus on relevant subsets of data, eliminating noise and irrelevant information.





**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.



Age Group:

if(Age >= 60, 'Senior',

  if(Age >= 30, 'Mid Age',

  if(Age >= 18, 'Youth',

  if(Age >= 14, 'Teen','Kid'

    ))))



Region:

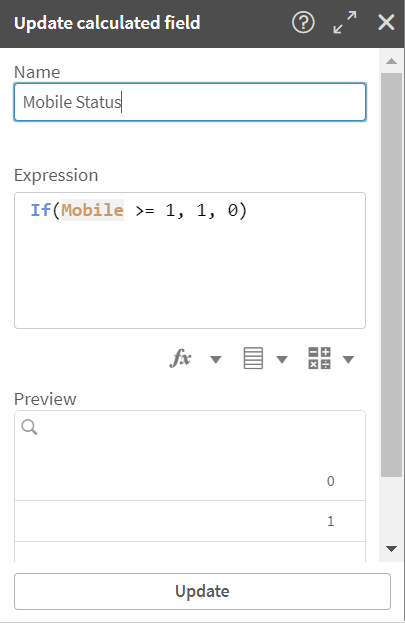
if(Match(State, 'Maharashtra', 'Gujarat', 'Rajasthan', 'Goa', 'Daman and Diu', 'Dadra and Nagar Haveli'), 'Western',

if(Match(State, 'Uttar Pradesh', 'Bihar', 'Jharkhand', 'Odisha', 'West Bengal', 'Sikkim'), 'Eastern',

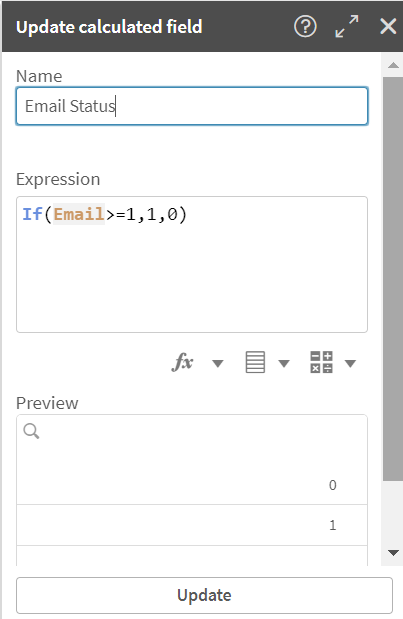
if(Match(State, 'Karnataka', 'Andhra Pradesh', 'Telangana', 'Tamil Nadu', 'Kerala', 'Puducherry'), 'Southern',

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')))))



+



### 

### **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