Data Insights from Aadhar a Comprehensive Analysis using Qlik

Project Report Titles

1. **Introduction**
   - 1.1. Overview: A brief description about your project
   - 1.2. Purpose: The use of this project. What can be achieved using this.
   - 1.3. Technical Architecture
2. **Define Problem / Problem Understanding**
   - 2.1. Specify the business problem
   - 2.2. Business requirements
   - 2.3. Literature Survey
   - 2.4. Social & Business Impact
3. **Data Collection**
   - 3.1. Collect the dataset
   - 3.2. Understand the Dataset
   - 3.3. Connect Data with Qlik Sense
4. **Data Preparation**
   - 4.1. Prepare the Data for Visualization
5. **Data Visualisation**
   - 5.1. Visualizations
6. **Dashboard**
   - 6.1. Responsive and Design of Dashboard
7. **Story**
   - 7.1. Design of Story
8. **Performance Testing**
   - 8.1. Amount of Data Rendered
   - 8.2. Utilization of Data Filters
   - 8.3. Number of calculation fields
   - 8.4. Number of visualisations / graphs

Define Problem / Problem Understanding

Specify the Problem

The Unique Identification Authority of India (UIDAI), a statutory body set up by the Indian government, issues the 12-digit Aadhaar Card unique identifying numbers. The goal of the Aadhaar initiative is to give Indian citizens access to a strong and comprehensive identity infrastructure.

Using Qlik Sense, an extensive analysis of Aadhaar data is carried out with an emphasis on obtaining actionable insights. The project entails preparing an interactive Qlik Sense dashboard report, cleaning and modelling the Aadhaar dataset, and extracting important visualisations, including geographic analyses, energization/rejections, and demographic overviews.

The main source of data is the massive Aadhaar database, which includes geographic information, demographic data, and records of authentication. The project's goal is to use Qlik Sense to analyse Aadhaar data in depth in order to gain insightful information that will improve the National Identity Authority's capacity for policymaking, decision-making, and operational effectiveness. The Aadhaar database contains sensitive personal information that needs to be managed carefully, so it's necessary to think about the privacy and security consequences of accessing it, even though employing Qlik Sense for data analysis can yield insightful results. Furthermore, depending only on one data source could reduce the analysis's thoroughness and accuracy because different information sources could offer more context and insights.

Business requirements

In order to support wise decision-making, the analysis attempts to offer insightful data on user demographics, authentication trends, and compliance measures. The creation of visually stimulating and interactive dashboards to assist with operational and strategic planning is the main goal. The investigation yielded valuable insights that would facilitate informed decision-making, improve service delivery, and guarantee regulatory compliance.

1. Look more closely at the particular user demographics under analysis and how those affect organisational decision-making.
2. Take a closer look at the trends in authentication and discuss any implications for the next strategies and security measures.
3. Take a closer look at the tracking and measurement of compliance metrics, emphasising important performance indicators that are essential for regulatory conformity.
4. Talk about any obstacles or constraints that might have arisen throughout the data analysis process and how they were resolved to guarantee that reliable insights were obtained
5. Take into account using case studies or actual situations to illustrate how the knowledge

gained from this research can be used to improve compliance and service performance.

Literature Survey

In order to do a literature survey for the Aadhar analysis, it would be necessary to look up and evaluate earlier publications, articles, and studies on the subject. This could contain details on the procedures and methods employed to analyse the Aadhar analysis, in addition to the findings and recommendations of these investigations. Investigating scholarly databases like PubMed, IEEE Xplore, Google Scholar, and institutional repositories is advised. Government papers and reports might offer insights into the most recent advancements in Aadhaar analysis.

Researchers can find gaps that need to be filled in their own study and obtain a better understanding of the state of knowledge on the topic by reviewing the existing literature. This can assist in providing guidance for the Aadhar analysis's research questions, methodology, and data analysis strategies. Reviewing earlier research can also be helpful in identifying potential restrictions and difficulties that might occur when performing the analysis, enabling researchers to foresee and proactively address these problems. All things considered, a comprehensive literature review is necessary to establish the framework for an effective Aadhar analytical study.

Social & Business Impact

Social Impact Analysis:
- Create visualisations to showcase the demographic distribution of Aadhaar users.
- Analyse how Aadhaar has impacted social welfare programmes, financial inclusion, and other key areas.
- Explore any correlations between Aadhaar usage and improvements in socioeconomic indicators. Analyse the potential privacy and security risks associated with Aadhaar implementation.
- Examine any studies or reports that have already been conducted on the correlation between Aadhaar usage and improvements in socioeconomic indicators like poverty levels or access to education.
- Discuss any potential challenges or criticisms surrounding the implementation of Aadhaar and its impact on privacy concerns or the exclusion of marginalised communities.
- Consider conducting interviews with experts in the field of social impact analysis to gather insights on the long-term effects of Aadhaar on India's society and economy.

Business Impact Analysis:
- Analyse 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 visualisations to represent the growth in Aadhaar-based services.
- Explore the ethical implications of Aadhaar implementation, considering how it balances privacy concerns with the potential benefits for marginalised communities.
- Examine the legal challenges and controversies surrounding Aadhaar, such as concerns about data security and potential misuse of personal information.
- Investigate the role of technology in driving government initiatives like Aadhaar, including the use of biometrics and digital identification systems in other countries.
- Assess the effectiveness of Aadhaar in achieving its intended goals, such as reducing fraud, improving access to government services, and promoting financial inclusion.
- Discuss the public perception of Aadhaar in India, examining how attitudes towards this unique identification system have evolved over time and whether there is broad support for its continued implementation.

Data Collection & Extraction from Database

Collect the dataset

Refer the link to see the collected Database - [Data-Insights-from-Aadhaar-A-Comprehensive-Analysis-using-Qlik/Data Collection and Extraction from Database at main · vaibhhaav/Data-Insights-from-Aadhaar-A-Comprehensive-Analysis-using-Qlik (github.com)](#)

Understand the dataset

The data includes all of the meta data pertaining to the columns that are provided in the CSV files.

Dataset's Column Description:

1. Registrar: organisations in charge of establishing enrollment centres, overseeing the enrollment procedure, and gathering required information.

2. Enrollment Agency: An enrollment agency is in charge of carrying out the actual process of enrolling people in the Aadhaar system.

3. States: The Indian States

4. District: A district is an administrative division or unit that is typically part of a larger administrative territory, such as a state.

5. Sub-District: A sub-district is a smaller administrative division that is a part of a district. It is also referred to as a taluka or tehsil in various locations.

6. Pin Code: The Aadhar card holder's PIN code

7. Gender: The Aadhar card holder's gender

8. Age: The Aadhar card holder's age

9. Aadhaar produced: Number of Aadhars created

10. Rejected enrollment: Total number of rejected enrollments

11. Residents supplying email: Whether or not a resident provides email

Connect Data with Qlik Sense

Data Preparation

Prepare the Data for Visualization

In order to prepare data for visualisation, it must first be cleaned to remove irrelevant or missing data, then transformed into a format that makes it easy to visualise, explored to find patterns and trends, filtered to concentrate on particular subsets of data, prepared for visualisation software, and verified to be accurate and complete. Through this process, the data is made more comprehensible and prepared for the creation of visualisations that will provide insights into efficiency and performance. We can proceed with visualisation now that the data has been cleaned. Putting the data into a visual format—charts, graphs, dashboards, etc.—makes it simpler to understand and derive conclusions. Investigating patterns and trends in the data might yield insightful information for formulating strategies and making decisions. We can find linkages and correlations that were previously hidden by narrowing the data to concentrate on particular subgroups. All things considered, the process of getting data ready for visualisation is essential to seeing the full value of the information and using it to inform decisions and advancements.

In this process, we have added the following columns in our dataset:

1. Age Group
2. Region
3. Mobile Status
4. Email Status

Data Visualization

The technique of producing graphical representations of data to aid in understanding and information exploration is known as data visualisation. Making complex data sets more approachable, intuitive, and interpretable is the aim of data visualisation. Data visualisations make patterns, trends, and outliers in the data easier to see by using visual components like charts, graphs, and maps. This may result in a deeper comprehension of the underlying data and better decision-making. Good data visualisation is a useful tool for academics, analysts, and decision-makers alike since it may aid in disseminating important ideas and discoveries to a larger audience. Data visualisation is becoming more and more crucial in helping us make sense of the vast amounts of data that are being created and gathered.

Visualizations
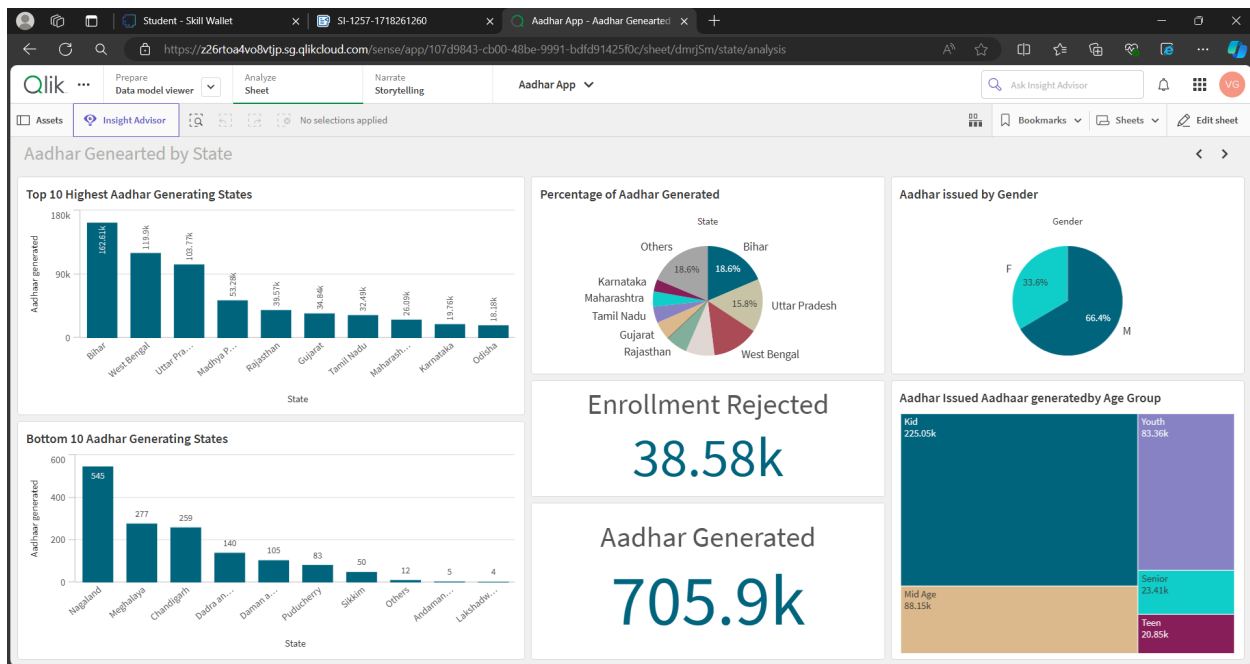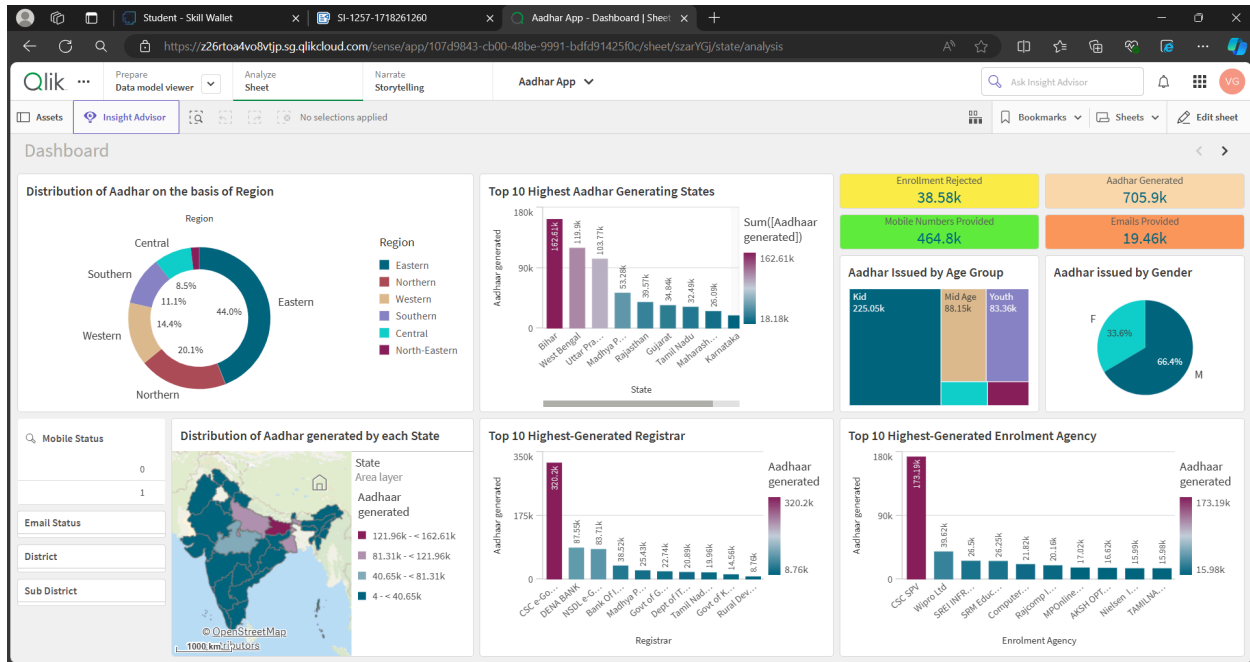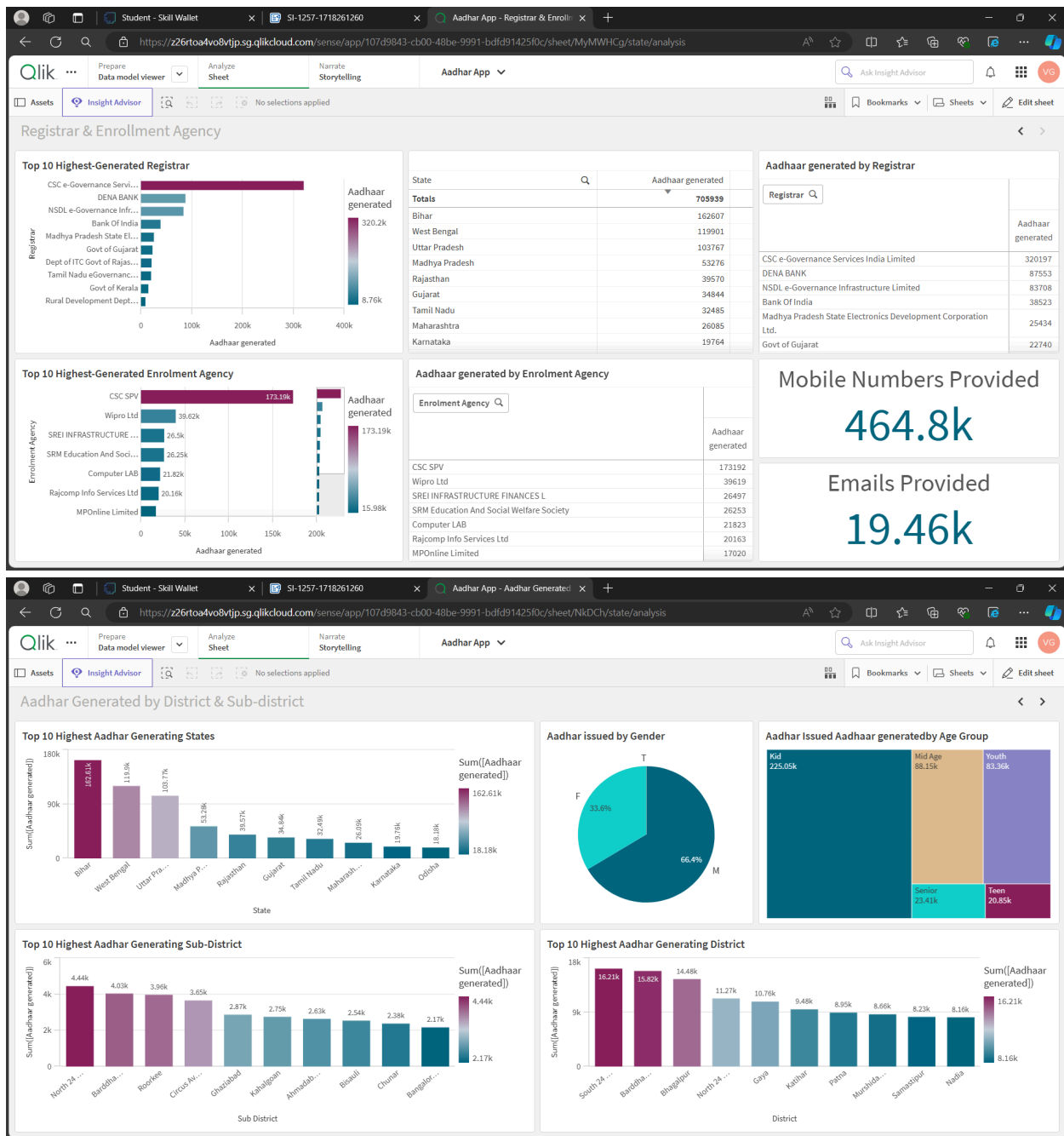
Dashboard

A dashboard is a type of graphical user interface (GUI) that presents data and information in a logical, readable manner. Dashboards are usually made for a particular use case or purpose and are frequently used to provide real-time data monitoring and analysis. Dashboards can be used in a variety of settings, such as business, finance, manufacturing, healthcare, and many other industries. They can be used to measure key performance indicators (KPIs), monitor performance metrics, and show data in the form of charts, graphs, and tables. Overall, dashboards are tremendously useful tools for decision-makers to rapidly and easily access and analyse key data. Dashboards offer data to customers in an aesthetically pleasing and easily

comprehensible format, making it easier for them to spot patterns, trends, and possible problems. Whether used for monitoring sales performance, tracking patient outcomes, or measuring production efficiency, dashboards serve a critical role in helping organisations make informed decisions and drive success. With the option to create dashboards to suit specific needs and goals, organisations may acquire important information and stay ahead of the competition.

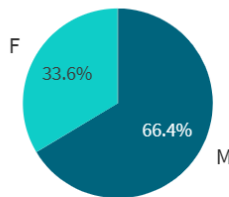## Responsive and Design of Dashboard
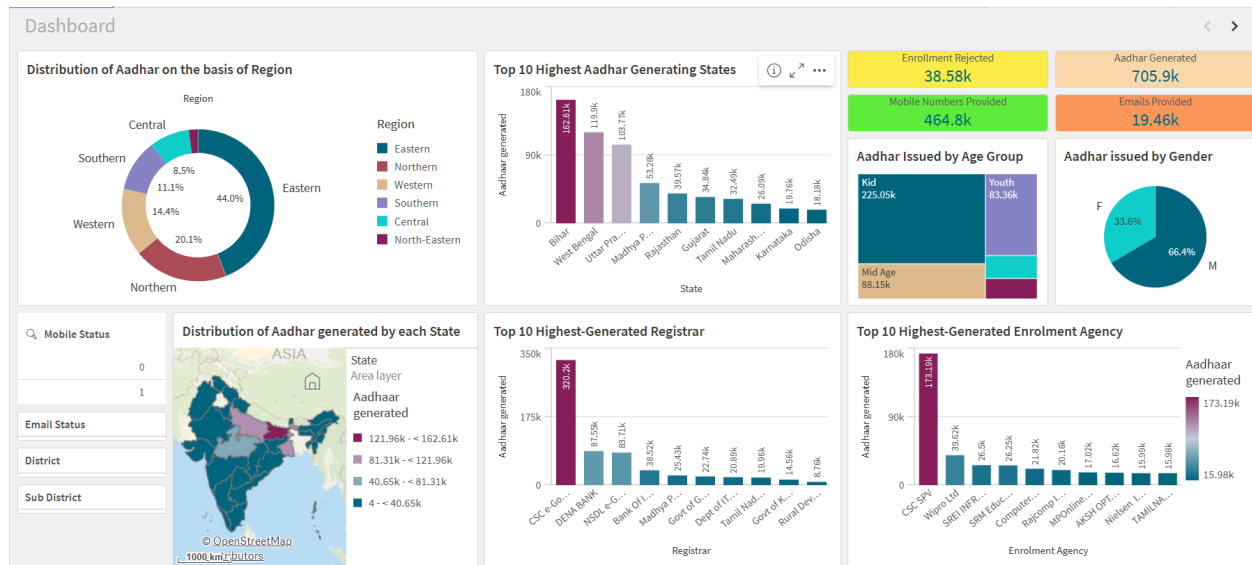
Story

A data story is a narrative structure used to display data and analysis with the intention of improving the information's readability and engagement. A data story usually consists of three parts: an explanation of the background and context of the data in a clear and concise introduction; a logical and systematic presentation of the data and analysis in the body; and a summary of the main conclusions and their implications in the conclusion. A range of media, including reports, presentations, interactive visualisations, and films, can be used to tell data stories. Using visual aids like graphs, charts, and infographics can help to clarify important ideas and increase the data's accessibility to a larger audience. Furthermore, adding case studies or real-world examples might help to humanise the data and increase its relatability. All things considered, data stories are a powerful tool for conveying complicated information in a way that is both captivating and impactful, which encourages thoughtful decision-making and action.
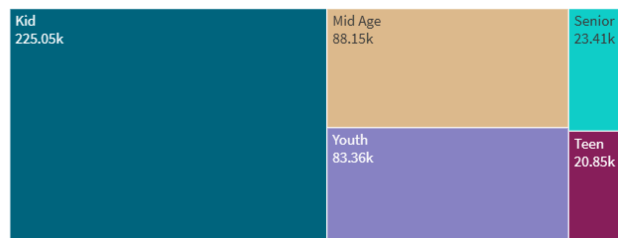
Design of Story

## Overview of Aadhar Analysis

Aadhar Generated	Enrollment Rejected	Mobile Numbers Provided	Emails Provided
705.9k	38.58k	464.8k	19.46k

Percentage of Male and Female Enrolled

F 33.6%
M 66.4%

Age Group Distribution of Generated Aadhar

Kid 225.05k
Mid Age 88.15k
Senior 23.41k
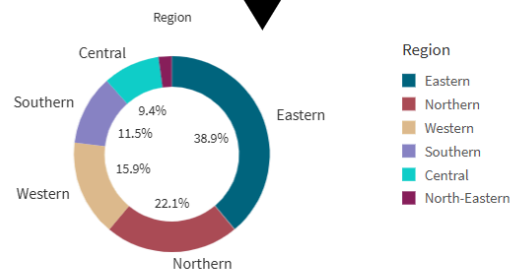Youth 83.36k
Teen 20.85k

Aadhar generated by Gender

Females
- Total = 247.7k
- Eastern Region = 132.2k
- Northern Region = 40.55k
- Western Region = 29.22k
- Southern Region = 25.39k
- Central Region = 16.87k
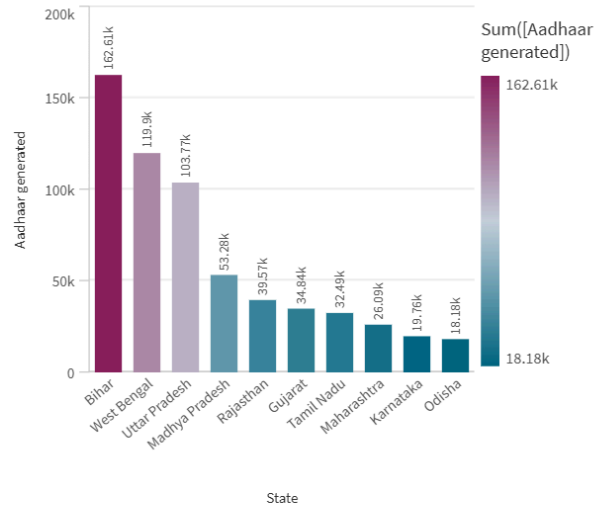- North-Eastern Region = 3.49k

F 33.6%
M 66.4%
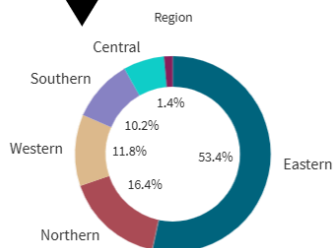
Males
- Total = 458.2k
- Eastern Region = 178.4k
- Northern Region = 101.2k
- Western Region = 72.7k
- Southern Region = 52.91k
- Central Region = 43k
- North-Eastern Region = 10.02k

Females donut chart:
Region
- Eastern
- Northern
- Western
- Southern
- Central
- North-Eastern

Central 1.4%
Southern 10.2%
Western 11.8%
Northern 16.4%
Eastern 53.4%

Males donut chart:
Region
- Eastern
- Northern
- Western
- Southern
- Central
- North-Eastern

Central 9.4%
Southern 11.5%
Western 15.9%
Northern 22.1%
Eastern 38.9%

Highest Aadhar Producing States
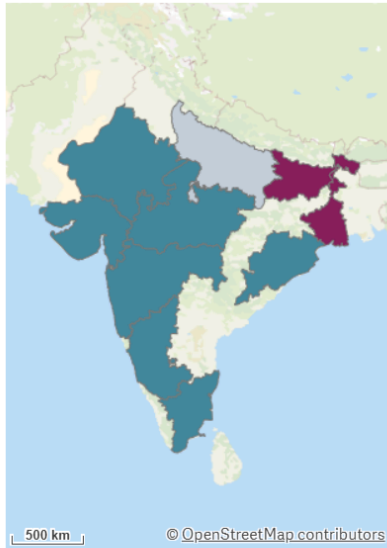
Dashboard

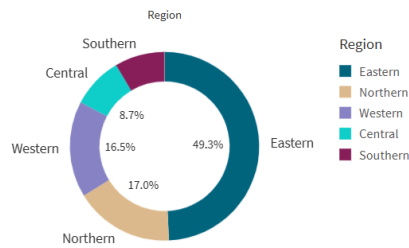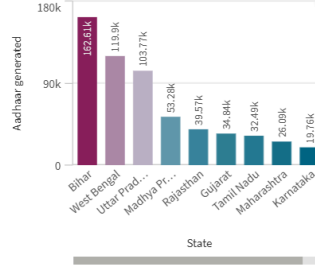Click sheet to make selections   Reset selections   Go to sheet

Distribution of Aadhar on the basis of Region

Top 10 Highest Aadhar Generating States
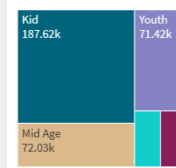

Enrollment Rejected	Aadhar Generated
32.96k	610.5k

Mobile Numbers Provided	Emails Provided
382.6k	14.33k

Aadhar Issued by Age...

Kid 187.62k
Youth 71.42k
Mid Age 72.03k

Aadhar issued by Ge...
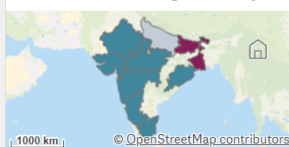

F 34.7%   M 65.3%
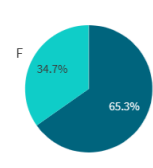
Mobile Status
0
1

Email Status
0
1

District

Sub District

Distribution of Aadhar generated by e...

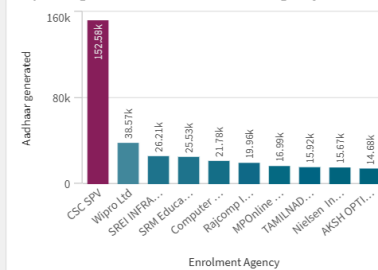

State
Area layer

Aadhaar generated
18.18k - < 66.32k   114...
66.32k - < 114.47k

Top 10 Highest-Generated Registrar

Top 10 Highest-Generated Enrolment Agency

Data last loaded: Jun 16, 2024, 12:20 AM

Performance Testing

Amount of Data Rendered

The term "Amount of Data Loaded" describes the amount or volume of data that has been loaded into a database, software programme, system, or other data processing or storage environment. It is an indicator of the volume of data that has been effectively processed and made accessible for usage, manipulation, and analysis within the system. One crucial statistic that organisations should monitor to make sure their systems are operating well is the volume of data that is loaded. Businesses can see any possible bottlenecks or problems that might be compromising the functionality of their systems by keeping an eye on the volume of data being loaded. Organisations can use this information to enhance overall data management procedures and optimise their workflows for processing data.

abc
Registrar
Enrolment Agency
State
District
Sub District
Pin Code
Gender
Age
Aadhaar generated
Enrolment Rejected
Email
Mobile
abc.District_GeoInfo
Age Group
Region
Mobile Status
Email Status

Utilisation of Data Filters

The process of adding particular criteria or conditions to a dataset in order to include or exclude specific data points is referred to as "using data filters." In data analysis, this filtering technique is essential since it enables the emphasis on pertinent data subsets while removing noise and unimportant information. Additionally, data filters aid in spotting patterns, trends, and outliers in the dataset, which facilitates the extraction of insightful conclusions and the formulation of defensible judgements. Analysts are able to extract relevant information quickly and efficiently by applying data filters properly. Furthermore, data filters ensure that only the most pertinent data is taken into account throughout the analysis, which can increase the accuracy and dependability of the results.
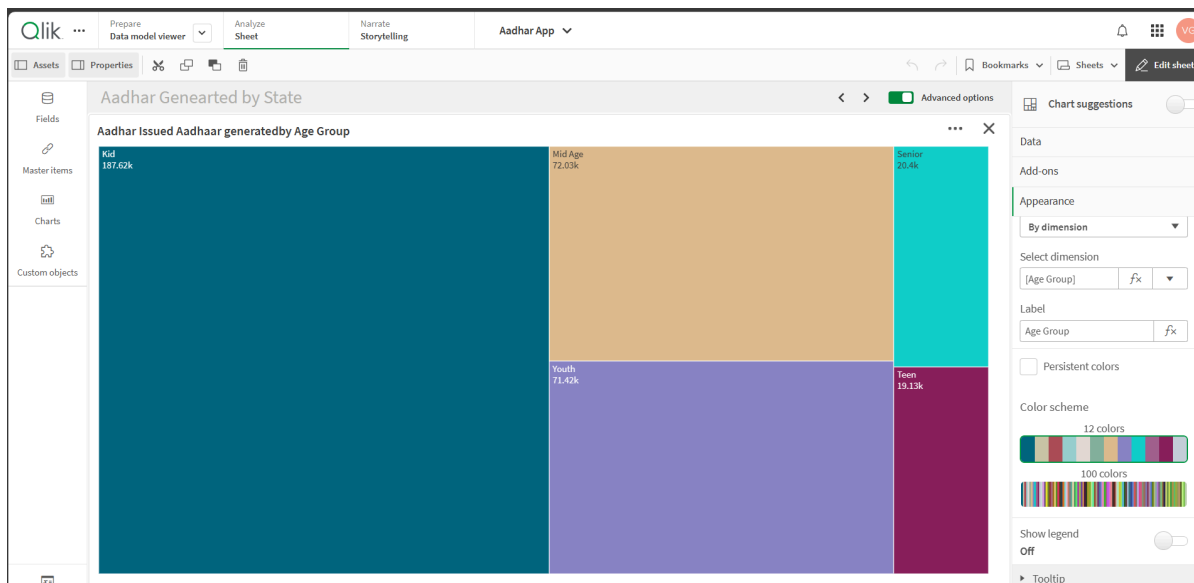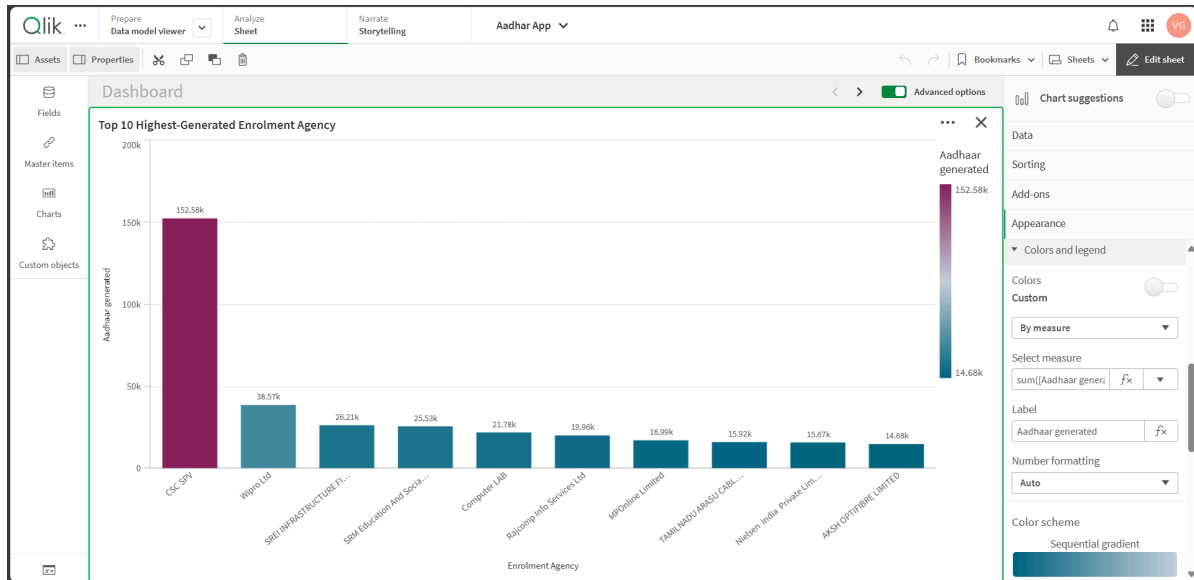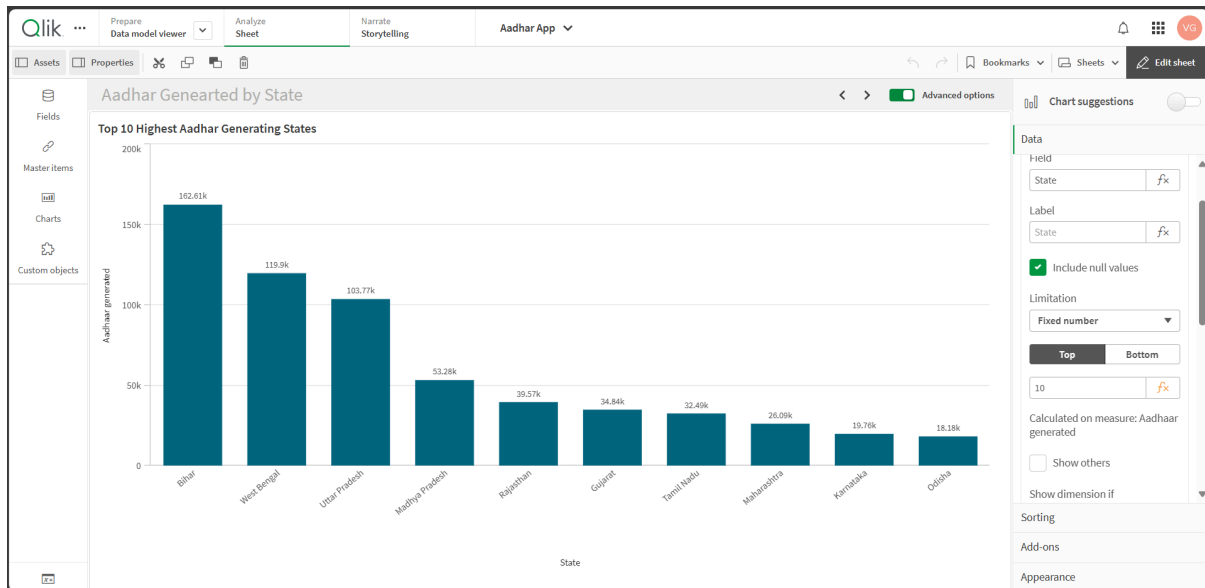
Number of Calculation Fields

The variables in a dataset that are created by computations instead of being taken straight from the source data are commonly referred to as "calculation fields". Applying mathematical functions, processes, or formulas to the dataset's preexisting data yields these fields. Calculation fields can be utilised to extract fresh insights from the data and carry out complex studies. They are frequently used to make interactive graphs and charts that show the computation results in data visualisation tools. Researchers and analysts can find links, patterns, and trends in the dataset by using calculation fields; these findings may not be immediately evident from the raw data on its own.

We have added the following calculation fields in our dataset:

1. Age Group:
Expression:-
if(Age >= 60, 'Senior',

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

if(Age >= 18, 'Youth',

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

))))

Expression:-
if(Match(State, 'Maharashtra', 'Gujarat', 'Goa', 'Rajasthan' ,'Dadra and Nagar Haveli', 'Daman and Diu', 'Daman', 'Diu', 'Others'), 'Western',
    if(Match(State, 'Chhattisgarh', 'Madhya Pradesh'), 'Central',
        if(Match(State, 'West Bengal', 'Odisha', 'Jharkhand', 'Bihar'), 'Eastern',
            if(Match(State, 'Karnataka', 'Andhra Pradesh', 'Tamil Nadu', 'Telangana', 'Kerala','Andaman and Nicobar Islands', 'Puducherry', 'Lakshadweep'), 'Southern',
                if(Match(State, 'Jammu and Kashmir', 'Punjab', 'Himachal Pradesh', 'Uttarakhand', 'Haryana', 'Uttar Pradesh', 'Delhi', 'Chandigarh'), 'Northern',
                    if(Match(State, 'Assam', 'Meghalaya', 'Tripura', 'Nagaland', 'Arunachal Pradesh', 'Sikkim', 'Mizoram', 'Manipur'),
 'North-Eastern', 'Other'))))))

3. Mobile Status

Expression:-
If(Mobile >= 1, 1, 0)

4. Email Status

Expression:-
If(Email >= 1, 1, 0)

Number of Visualizations / Graph

1. Enrollment Rejected
2. Aadhar Generated
3. Mobile Numbers Provided
4. Emails Provided
5. Distribution of Aadhar on the basis of Region
6. Top 10 Highest Aadhar Generating States
7. Top 10 Highest-Generated Registrar
8. Top 10 Highest-Generated Enrollment Agency
9. Aadhar Issued by Age Group
10. Aadhar issued by Gender
11. Bottom 10 Aadhar Generating States
12. Percentage of Aadhar Generated
13. Aadhaar generated by Age Group
14. Top 10 Highest Aadhar Generating Sub-District
15. Top 10 Highest Aadhar Generating District
16. Aadhar generated by States
17. Aadhaar generated by Enrolment Agency
18. Aadhaar generated by Registrar