# Project Report

# 1.INTRODUCTION

## 1.1 Overview

"Data Insights From Aadhaar: A Comprehensive Analysis Using Qlik" likely refers to a project that utilizes Qlik, a data analytics platform, to uncover insights from Aadhaar, India's unique identification system. This analysis could explore demographics linked to Aadhaar enrollment, identify trends in Aadhaar usage, or even investigate potential areas for improvement within the system.

## 1.2 Purpose

The purpose of "Data Insights From Aadhaar: A Comprehensive Analysis" is to unlock valuable knowledge (data insights) from Aadhaar, India's unique ID system. This in-depth examination aims to identify patterns and trends within the Aadhaar data.

## 1.3 Technical Architecture

- **Data Source:** Aadhaar data would be the primary data source. This could be stored in a centralized government database or another secure location.
- **Data Extraction, Transformation, and Loading (ETL):** A process to extract data from Aadhaar's source, transform it into a format usable by Qlik, and load it into the Qlik platform.
- **Qlik Sense or QlikView:** The specific Qlik product used for data analysis. Qlik would handle data visualization, manipulation, and creating insights.
- **Security Measures:** The architecture would likely incorporate robust security measures to protect sensitive Aadhaar data throughout the process.

## 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 with regulations.

## 2.3 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.

## 2.4 Social Or Business Impact.

### 2.4.1 Social Impact Analysis:

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

### 2.4.2 Business Impact Analysis:

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

## 3.Data Collection

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

The Dataset is given in the Comma Seperated Values(".csv") file format the dataset can be downloaded using the following link and can be loaded into the Qlik Cloud

**Dataset Link:**

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

### 3.2 Connect Data with Qlik Sense:

The Downloaded Dataset is Saved in downloads in the name of "abc.csv" .

**Then login into MyQlik The cloud service provided by qlik sense then follow the instructions to load the data into cloud:**

**Step 1**:Click Add New

**Step 2**:Select new analytics app

**Step 3**:In Create new app dialogue box enter the name of the application in the name field

**Step 4**:Then Click Create the application loads in datamanager tab

**Step 5**: Click the files and others source options to load the dataset from the the local devices and other cloud storages

Step 6: Drag and drop the file or click upload button to load the file then select the file to upload

eg:"abc.csv"

Step 7:Then the file opens and displays its contents,Click next and wait for fewminutes until the data is loaded

Step 8: After the data is loaded Start the data preparation and visualization process.

## 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.

**Steps To be Done Before Moving To Visualization**

**Step1: Add Some Field To the dataset from the given dataset to make the visualization more undterstanble**

**Step 2: Field To be added as columns**

1. **Age Group -> Kid,Teen,Youth,Mid Age,Senior**
2. **Region -> Western,Eastern,Southern,Northern,North-Eastern**
3. **Mobile Status -> 0,1(generated with/without Mobile Provided)**
4. **Email Status -> 0,1(generated with/without Email Provided)**

## 5.Data Visualizations

### 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.

Here We Are creating following

1. Aadhar Issued By Age Group
2. Top 10 Aadhar-Generated States
3. Aadhar Issued By Region
4. Issue By Gender
5. Top 10 Highest Generated Enrollment Agency
6. Top 10 Highest-Generated Registrar
7. Bottom 10 Aadhar-Generated State

## 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.

## 6.1 Responsive and Design of Dashboard

Here,We Are creating two dashboards and giving name to the dashboards

**Ex:**

1. **Dashboard 1**
2. **Dashboard 2**

**The Dashboard 1 Contains the following**

**componenets/elements:**

1. **Total Aadhar Generated**
2. **Total Enrolment Rejected**
3. **Mobile Provided**
4. **Email Provided**
5. **Aadhar Issued By Region**
6. **Aadhar Issued By Age Group**
7. **Aadhar Issued By Gender**
8. **Aadhar Generated and Rejected by state**

**The Dashboard 2 Contains the following**

**componenets/elements:**

1. **Top 10 Aadhar-Generated States**
2. **Bottom 10 Aadhar-Generated State**
3. **Top 10 Highest Generated Enrollment Agency**
4. **Top 10 Highest-Generated Registrar**
5. **Aadhar Generated and Rejected by Region**
6. **Aadhar Generated by Gender**

## 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.

**This Story Contains the following**

1. **Total Aadhar Analysis**
2. **Aadhar Generated by Gender**
3. **Aadhar Generated and Rejected by state**
4. **Top 10 Aadhar-Generated States**
5. **Bottom 10 Aadhar-Generated State**
6. **Top 10 Highest Generated Enrollment Agency**
7. **Top 10 Highest-Generated Registrar**

## 8.Performance Testing

### 8.1 Amount of Data Rendered to DB

"**Amount of Data Rendered to DB**" 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

Here, the following containing theor respective data values are imported

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

## 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.

**The following are the places where data filters are used**

1. **Top 10 Aadhar-Generated States**
2. **Bottom 10 Aadhar-Generated State**
3. **Top 10 Highest Generated Enrollment Agency**
4. **Top 10 Highest-Generated Registrar**

## Conclusion:

**Here By, I Conclude that I Have Completed the Project "Data Insights From Aadhaar A Comprehensive Analysis Using Qlik" with the help the Qlik Sense product (i.e.,) Qlik Cloud with true to the best of my knowledge.**