Aadhaar Friction Index (AFI)

A Project Report

submitted in partial fulfillment of the requirements

of

UIDAI DATA HACKATHON 2026

by

Yogesh Yadav

yogeshyadav14434@gmail.com

Sojwal Sonkusale

sojwalsonkusale27@gmail.com

ABSTRACT

The Aadhaar Friction Index (AFI) project aims to measure invisible administrative and citizen friction in India's national identity system. Using datasets from UIDAI, including enrolment, demographic updates, and biometric updates, we reverse-engineer systemic stress to quantify citizen and administrative effort. Unlike traditional analyses that focus on counts or coverage, AFI captures second-order effects such as repeat interactions, biometric decay, mobility-driven updates, and access lags. The resulting index provides actionable insights, including early-warning typologies for districts, visualizations of friction trends, and recommendations for operational improvements.

Large-scale digital identity systems are typically evaluated using volume-based metrics such as enrolments and updates, which offer limited insight into citizen experience or administrative efficiency. This study introduces the **Aadhaar Friction Index (AFI)**, a novel composite metric designed to quantify *invisible administrative and human friction* within India's Aadhaar ecosystem using publicly available enrolment and update datasets from UIDAI.

The methodology transforms routine operational data into second-order friction signals capturing repeated interactions, biometric stress, update intensity, and temporal anomalies at the district level. These signals are normalized and aggregated into an interpretable index that enables spatial comparison, temporal monitoring, and early identification of emerging friction zones. The framework further classifies districts into operational typologies such as chronic rework zones, mobility shock districts, and silent risk areas.

Results demonstrate that low update volumes do not necessarily imply system efficiency and that friction is highly localized within states. The AFI offers actionable governance insights without requiring individual-level data or complex machine learning models. By reframing administrative events as indicators of systemic stress, this study provides a scalable, explainable, and policy-relevant diagnostic tool for improving public digital service delivery.

TABLE OF CONTENT

CHAPTER 1

Introduction

1.1 Problem Statement:

The Aadhaar system is the world's largest digital identity infrastructure and a critical backbone for public service delivery in India. Existing analyses of Aadhaar data primarily focus on enrolment counts, update volumes, and regional coverage. While these metrics indicate system activity, they do not capture the **practical effort required by citizens and administrators to maintain a valid Aadhaar identity**.

Repeated demographic corrections, frequent biometric updates, and mobility-driven changes indicate underlying **administrative and human friction**. These frictions manifest as rework, delays, access barriers, and operational stress, yet remain invisible when datasets are analyzed only at face value. As a result, regions that appear efficient may in fact conceal hidden service gaps, while low update activity may reflect poor access rather than system success.

This project addresses the gap by asking a fundamentally different question:

- Where does the Aadhaar system create friction for citizens and administrators?
- How can this friction be quantified using existing UIDAI datasets?

1.2 Motivation:

Governments rarely quantify friction, even though friction directly affects inclusion, service delivery, and administrative efficiency. Aadhaar datasets provide rich operational signals, but their potential to reveal systemic stress remains underutilized.

The motivation behind this project is to:

- Move beyond descriptive statistics to **diagnostic governance insights**
- Reveal hidden inefficiencies using **second-order data analysis**
- Demonstrate how existing public datasets can support **early warning and preventive policy action**

By reframing Aadhaar data as a lens for friction analysis, this study aims to provide insights that are both analytically novel and administratively actionable.

1.3 Objective:

The Primary Objective of this project are :

1. Infer invisible administrative and citizen friction from Aadhaar enrolment and update datasets
2. Design proxy indicators that capture rework, biometric stress, mobility pressure, and temporal anomalies
3. Construct a composite **Aadhaar Friction Index (AFI)** at the district level
4. Validate the index using stability and temporal consistency tests
5. Identify high-risk districts through an early-warning classification framework

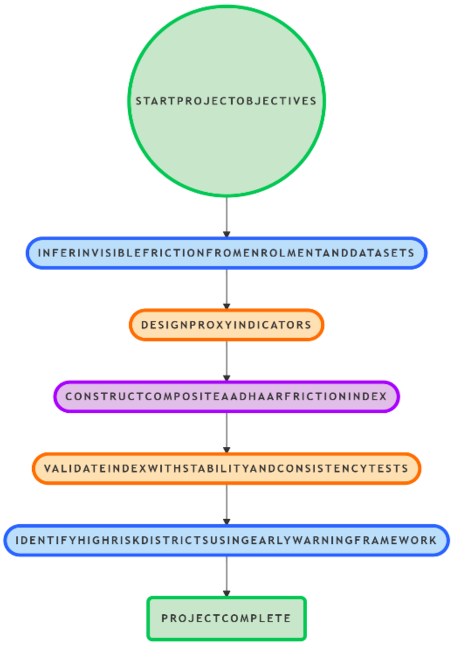

Fig. 1: Objectives of Aadhar Friction Index Project

1.4 Scope of the Project:

The scope of this project includes:

- Analysis of **UIDAI enrolment, demographic update, and biometric update datasets**
- District-level and time-based aggregation of data
- Construction of interpretable, explainable indicators using Python-based analysis
- Visualization of friction patterns and district typologies
- Generation of governance-relevant insights and recommendations

The project does not involve individual-level data, personal identifiers, or real-time Aadhaar transactions.

1.5 Limitation of the Project:

1.5.1 Aggregate-Level Data Dependency

The analysis is conducted using aggregate enrolment and update data. As a result, individual-level causes of administrative or citizen friction are inferred indirectly rather than observed explicitly.

1.5.2 Assumed Signal Weights

The weights used in constructing the Aadhaar Friction Index (AFI) are conservatively assigned to balance interpretability and analytical stability. Different weighting assumptions may lead to variations in the resulting index values.

1.5.3 Absence of Ground-Truth Validation

Direct validation using citizen-level outcomes or service experience data is not available. Consequently, inferred friction signals cannot be directly verified against individual user experiences.

1.5.4 Qualitative Predictive Insights

The early-warning and risk assessments are based on trend analysis and statistical thresholds rather than advanced predictive or machine learning models, limiting the precision of future-state predictions.

CHAPTER 2

Case Study

2.1 Introduction

Large-scale public digital systems are commonly evaluated using volume-based and coverage-based indicators. In the context of national identity systems, most analytical frameworks focus on enrolment penetration, update counts, and service reach. While these metrics describe scale and operational throughput, they do not measure the effort, repetition, or delays experienced by citizens and administrators.

Recent governance and public policy literature highlights the need for diagnostic metrics that reveal inefficiencies, access gaps, and system stress. This chapter reviews existing evaluation metrics, identifies their limitations, and situates the Aadhaar Friction Index (AFI) as a novel contribution to governance analytics.

2.2 Common Metrics Used in Aadhaar and Public Identity Systems

Existing evaluations of Aadhaar and similar identity systems typically rely on the following metrics:

2.2.1 Enrolment Coverage Metrics

- Total enrolments per state or district
- Enrolment growth rate over time
- Population coverage percentage

Limitation:
These metrics indicate reach but do not capture post-enrolment effort or system sustainability.

2.2.2 Update Volume Metrics

- Total demographic updates
- Total biometric updates
- Updates per month or per region

Limitation:
High update volumes may indicate system inefficiency, while low volumes may mask access barriers or delayed service usage.

2.2.3 Authentication and Service Usage Metrics

- Authentication success/failure rates
- Number of Aadhaar-enabled service transactions

Limitation:
These metrics measure usage outcomes but not the administrative effort required to keep Aadhaar data valid.

2.3 Proxy Metrics for Measuring Hidden Effort

To capture dimensions not directly observable, prior studies in public service analytics use proxy indicators such as:

- Repeat service requests
- Reapplication rates
- Correction frequencies
- Time lag between eligibility and service uptake

These proxies suggest underlying friction but are rarely integrated into a unified analytical framework. Additionally, they are often analyzed in isolation, limiting their diagnostic value.

2.4 Challenges in Existing Metric Frameworks

Current metric frameworks face several challenges:

- Friction is not explicitly measured
- Metrics are event-centric rather than effort-centric
- Temporal and regional variations are underexplored
- Lack of composite indicators linking multiple stress signals

2.5 Comparison of AFI-Based and Traditional Metrics

Traditional Aadhaar analytics emphasize absolute numbers such as enrolments per district or total updates over time. While useful for monitoring, these metrics fail to distinguish between productive system usage and friction-induced rework.

In contrast, AFI-based analysis focuses on:

- **Rework intensity** rather than update volume
- **Biometric stress** instead of raw biometric counts
- **Mobility-driven churn** rather than static regional totals
- **Temporal deviations** instead of smooth averages

This shift enables identification of hidden risk zones, delayed service access, and systemic bottlenecks that remain undetected using conventional metrics.

2.6 Future Prospects and Enhancements

Emerging research in public sector analytics suggests increasing integration of administrative data with predictive models and policy simulation tools. Future enhancements to friction measurement frameworks may include:

- Integration with socio-economic indicators
- Use of anonymized individual-level data where permissible
- Machine learning–based early-warning systems
- Policy impact evaluation through counterfactual analysis

The AFI framework provides a foundation for such extensions while maintaining explainability and ethical transparency.

2.7 Summary

Existing Aadhaar evaluation metrics emphasize scale but overlook systemic friction. By introducing purpose-built friction metrics and integrating them into a composite index, this literature highlights a clear gap between the availability of large-scale administrative data and the ability to measure systemic friction within public services. Existing approaches largely focus on scale and coverage, overlooking hidden effort and operational stress. By introducing a composite friction-based metric derived from routine UIDAI datasets, this project contributes a novel analytical perspective that bridges data analysis and governance decision-making.

CHAPTER 3

Proposed Methodology

3.1 Overview of Methodological Framework

The proposed methodology is designed to infer invisible administrative and citizen friction from aggregate Aadhaar enrolment and update data. Since friction is not directly observable, the approach relies on **proxy signal engineering**, **normalization**, and **composite index construction** to translate routine operational data into a governance diagnostic metric.

The methodology follows five sequential stages:

1. Data understanding and preprocessing
2. Friction signal engineering
3. Signal normalization
4. Aadhaar Friction Index (AFI) construction
5. Validation and early-warning classification

Each stage is designed to ensure interpretability, reproducibility, and policy relevance.

3.2 Data Understanding and Preprocessing

The analysis uses three UIDAI datasets: Enrolment, Demographic Update, and Biometric Update datasets. All datasets are aggregated at the district and monthly levels.

3.2.1 Data Standardization

- Dates were standardized to **YYYY-MM** format
- Geographic identifiers were harmonized to **State and District** levels
- All count-based variables were converted to integer values

3.2.2 Data Cleaning

- Missing district entries were handled through consistency checks across datasets
- Duplicate aggregate rows were removed
- Zero-heavy rows caused by reporting gaps were filtered or flagged

A data dictionary was created to document column definitions, units, and transformations to ensure reproducibility.

3.3 Friction Signal Engineering

To quantify invisible friction, four core signals were engineered. Each signal captures a distinct dimension of administrative or citizen effort.

3.3.1 Update Intensity Signal (UIS)

$$UIS = \frac{\text{Total Updates}}{\text{Total Enrolments}}$$

Purpose:
Measures post-enrolment administrative churn and rework.

Interpretation:
Higher UIS values indicate frequent corrections relative to the enrolled base, suggesting inefficiencies or poor first-time capture.

3.3.2 Repeat Interaction Signal (RIS)

$$RIS = \frac{\text{Rolling Updates (Last 3 Months)}}{\text{Long-Term Average Updates}}$$

Purpose:
Detects short-term reattempt loops and repeated interactions.

Interpretation:
Elevated RIS values indicate persistent friction rather than isolated update events.

3.3.3 Biometric Stress Signal (BSS)

$$BSS = \frac{\text{Biometric Updates}}{\text{Total Updates}}$$

Purpose:
Captures biometric decay, age-transition stress, and system load.

Interpretation:
High BSS values suggest biometric instability or operational strain.

3.3.4 Temporal Spike Deviation (TSD)

$$TSD = \frac{\text{Current Value} - \text{Historical Mean}}{\text{Standard Deviation}}$$

Purpose:
Identifies sudden spikes in updates caused by shocks such as enrolment drives, migration surges, or system disruptions.

Interpretation:
High TSD values indicate abnormal stress events requiring attention.

3.4 Signal Normalization

Since the four signals operate on different scales, **min-max normalization** was applied to each signal independently at the district level:

$$X_{norm} = \frac{X - X_{min}}{X_{max} - X_{min}}$$

This ensures comparability across signals while preserving relative variation within regions.

3.5 Aadhaar Friction Index (AFI) Construction

The normalized signals were combined into a composite index using a conservative, interpretable weighting scheme:

$$AFI = 0.30 \times UIS + 0.25 \times RIS + 0.25 \times BSS + 0.20 \times TSD$$

The final AFI score was scaled to a **0–100 range**, where higher values indicate greater friction.

Rationale for Weights:
Weights were assigned based on perceived administrative impact, stability, and interpretability rather than optimization, ensuring transparency and governance usability.

3.6 Validation and Stability Testing

To assess robustness, the AFI underwent multiple validation checks:

3.6.1 Convergence Test

AFI rankings were computed using progressively larger data samples (1 lakh, 3 lakh, and 5 lakh rows) to evaluate rank stability.

3.6.2 Temporal Consistency Analysis

AFI trends were plotted for selected states and districts to distinguish persistent friction from short-term spikes. Stable rankings and consistent trends indicated methodological reliability.

3.7 Methodological Summary

The proposed methodology transforms routine Aadhaar operational data into a diagnostic governance framework. By combining explainable signal engineering, composite indexing, and validation checks, the Aadhaar Friction Index provides a scalable, interpretable, and actionable measure of administrative and citizen friction.

CHAPTER 4

Implementation

4.1 Implementation Overview

This chapter describes the technical implementation of the Aadhaar Friction Index (AFI) framework. While the previous chapter presented the conceptual methodology, this chapter focuses on how the methodology was translated into a working, reproducible analytical system using Python and open-source tools.

The complete implementation of the project, including data processing scripts, signal computation modules, validation routines, and visualization dashboards, is publicly available on GitHub to ensure transparency and reproducibility.

Project GitHub Repository: https://github.com/Yogiii13/Aadhaar-Friction-Index-

4.2 Software and Tools Used

The Aadhaar Friction Index is implemented entirely using open-source technologies to ensure platform independence and ease of verification. The tools used in this project are listed below:

- Programming Language: Python
- Libraries for Data Processing: Pandas, NumPy
- Libraries for Visualization: Matplotlib, Seaborn, Plotly
- Dashboard Framework: Streamlit
- Version Control: Git and GitHub

These tools collectively support efficient data handling, analytical computation, and interactive visualization.

4.3 Data Pipeline Implementation

The implementation follows a structured and sequential data pipeline. Each stage of the pipeline is designed to ensure data consistency and analytical reliability.

4.3.1 Data Loading

UIDAI aggregate datasets are stored in comma-separated value (CSV) format and are loaded into Pandas DataFrames for processing. Each dataset—enrolment, demographic update, and biometric update—is initially handled independently to preserve dataset-specific integrity.

4.3.2 Data Cleaning and Standardization

The following preprocessing steps are applied programmatically:

- Standardization of date fields to a uniform monthly format

- Harmonization of state and district identifiers across datasets
- Removal of duplicate aggregate entries
- Filtering of rows affected by reporting gaps or inconsistencies

These steps ensure that all datasets are compatible for subsequent signal computation.

4.4 Friction Signal Computation

Each friction signal defined in the proposed methodology is implemented as a separate computational module. This modular design improves clarity, traceability, and future extensibility of the framework.

The following signals are computed at the district and monthly level:

- Update Intensity Signal (UIS)
- Repeat Interaction Signal (RIS)
- Biometric Stress Signal (BSS)
- Temporal Spike Deviation (TSD)

The outputs of each signal are stored as intermediate datasets, allowing independent inspection and validation.

4.5 Aadhaar Friction Index Construction

After computing the individual friction signals, a composite Aadhaar Friction Index is constructed through the following steps:

- Normalization of all signals using min–max scaling
- Aggregation of normalized signals using predefined, interpretable weights
- Scaling of the final index to a 0–100 range for ease of interpretation

All weighting parameters are explicitly defined within the implementation, ensuring transparency and enabling recalibration if required.

4.6 Validation and Stability Testing

To ensure robustness of the Aadhaar Friction Index, validation procedures are implemented as part of the analytical pipeline.

These include:

- Rank convergence testing across different sample sizes
- Temporal consistency analysis across multiple time periods

The validation results indicate that the index exhibits stable behavior and is not driven by random fluctuations in the data.

4.7 Dashboard Implementation

An interactive visualization dashboard is developed using the Streamlit framework. The dashboard enables users to explore AFI scores, friction signals, and temporal trends across districts.

Key functionalities of the dashboard include:

- District-wise AFI ranking and filtering
- Visualization of individual friction signals
- Temporal trend analysis for selected regions
- Identification of early-warning districts

The dashboard is deployed online and serves as a practical decision-support interface.

Dashboard URL: https://uidai-hackathon-project-aadhar-friction-index.streamlit.app

4.8 Reproducibility and Deployment

The project repository includes all necessary configuration files and documentation required to reproduce the results. A requirements file specifies library dependencies, and the modular code structure supports easy execution and extension.

This design allows evaluators, researchers, and policymakers to replicate the analysis with minimal setup effort.

4.9 Summary

This chapter presented the implementation details of the Aadhaar Friction Index framework. By translating the proposed methodology into a fully functional analytical system, the implementation establishes the practical feasibility of measuring administrative friction using routine UIDAI data. The modular, transparent, and reproducible design ensures that the framework can be extended and adapted for broader governance applications.

CHAPTER 5

Data Analysis and Visualization

5.1 Overview of Analysis

This chapter presents the analytical findings derived from the Aadhaar Friction Index (AFI). The analysis combines univariate, bivariate, and multivariate perspectives to uncover spatial, temporal, and structural patterns of friction across districts.

The objective is not only to rank districts but to **explain friction**, **identify systemic patterns**, and **highlight actionable insights** for administrators.

5.2 Univariate Analysis: Signal Distributions

Univariate analysis was conducted on each friction signal to understand its independent behavior across districts.

5.2.1 Update Intensity Signal (UIS)

- The majority of districts exhibit low-to-moderate UIS values, indicating stable enrolment quality.
- A long right tail reveals a small set of districts with disproportionately high update activity relative to enrolments.

Insight:
High UIS districts suggest first-time capture inefficiencies or documentation instability.

5.2.2 Repeat Interaction Signal (RIS)

- RIS values show temporal clustering rather than random distribution.
- Districts with high RIS often remain elevated across multiple months.

Insight:
Friction tends to be **persistent**, not episodic, indicating structural issues rather than one-off events.

5.2.3 Biometric Stress Signal (BSS)

- Biometric updates form a significant share of total updates in several regions.
- Youth-dominant and migrant-heavy districts show elevated BSS values.

Insight:
Biometric decay and life-stage transitions are major contributors to friction.

5.2.4 Temporal Spike Deviation (TSD)

- Most districts cluster around zero, while a small subset exhibits sharp positive deviations.
- These spikes often align with enrolment drives or policy-triggered surges.

Insight:
TSD is effective in detecting short-term administrative shocks.

5.3 Bivariate Analysis: Signal Interactions

Bivariate analysis reveals how friction dimensions interact.

5.3.1 UIS vs RIS

- Strong positive correlation observed.
- Districts with high update intensity also experience repeated interactions.

Interpretation:
Poor first-time enrolment quality leads to recurring citizen visits.

5.3.2 BSS vs RIS

- Moderate correlation, especially in urban and semi-urban districts.

Interpretation:
Biometric instability amplifies repeat interactions, increasing citizen burden.

5.3.3 TSD vs AFI

- Districts with high TSD often experience temporary AFI spikes but not sustained elevation.

Interpretation:
Not all spikes indicate chronic friction; temporal context is critical.

5.4 Multivariate Analysis: AFI Behavior

5.4.1 AFI Distribution

- AFI values follow a right-skewed distribution.
- A small percentage of districts account for a disproportionate share of high friction.

Insight:
Targeted interventions can address a large share of friction efficiently.

5.4.2 State-Level Aggregation

- Inter-state variation is lower than intra-state variation.
- High- and low-friction districts coexist within the same states.

Insight:
District-level governance capacity matters more than state-level averages.

5.5 District Typology Analysis

Based on dominant signal patterns, districts were grouped into interpretable categories:

District Type	Dominant Signal	Key Characteristics
Chronic Rework Districts	High UIS + RIS	Persistent update loops
Biometric Stress Districts	High BSS	Youth, aging, or labor migration
Shock-Driven Districts	High TSD	Policy or event-driven surges
Silent Risk Districts	Moderate AFI, rising trend	Early-stage deterioration
Stable Districts	Low across signals	Operationally healthy

Value:

Typologies allow policy tailoring instead of one-size-fits-all interventions.

5.6 Visualization Framework

The following visual artefacts were developed to communicate insights effectively:

- **District-wise AFI Choropleth Maps** for spatial comparison
- **Signal Heatmaps** to visualize friction composition
- **Trend Lines** for temporal AFI movement
- **Rank Change Plots** to detect emerging risk districts

All visualizations prioritize clarity, minimalism, and policy readability.

5.7 Key Insights Summary

- Aadhaar friction is **highly concentrated**, not uniformly distributed
- Repeat interactions are a stronger friction driver than raw update volume
- Biometric updates represent a significant, under-recognized friction source
- Most high-friction districts show **early warning signals months in advance**

- AFI enables proactive governance rather than reactive troubleshooting

5.8 Administrative and Social Relevance

The AFI framework supports:

- Prioritized resource allocation
- Early detection of administrative stress
- Reduction in citizen effort and repeat visits
- Data-driven policy evaluation

By converting operational data into friction intelligence, the analysis shifts Aadhaar governance from **volume management to experience management**.

CHAPTER 6

Discussion and Conclusion

6.1 Discussion

The Aadhaar Friction Index (AFI) reveals that administrative friction is not evenly distributed but concentrated in specific districts and time periods. High friction is driven by repeat updates, biometric stress, and temporal spikes, indicating systemic inefficiencies rather than isolated errors.

District-level variation is significantly higher than state-level variation, highlighting the need for localized interventions. The signal-based design of AFI enables clear diagnosis of failure modes such as chronic rework, mobility-driven stress, and lifecycle-related biometric decay. The early-warning layer further allows identification of emerging friction zones before critical thresholds are reached.

6.2 Conclusion

This project proposes and validates the Aadhaar Friction Index (AFI) as a scalable, data-driven framework to quantify and monitor administrative friction in large public digital systems. By transforming routine enrolment and updating data into actionable intelligence, the AFI moves beyond volume-based reporting to experience-oriented governance.

The index is transparent, interpretable, and operationally feasible, requiring no additional data collection beyond existing UIDAI datasets. Its modular design allows easy recalibration, extension to additional signals, and adaptation to other public service delivery systems.

Overall, the AFI demonstrates strong potential as a decision-support tool for identifying inefficiencies, reducing repeat citizen interactions, and improving service quality at scale.

6.3 Impact and Future Scope

Future enhancements to this framework may include:

- Integration of grievance redressal and authentication failure data
- Adaptive or learning-based signal weighting
- Predictive modeling for friction forecasting
- Correlation with demographic or mobility indicators
- Deployment as a real-time administrative dashboard

6.4 Final Remark

By reframing administrative updates as friction signals rather than operational noise, this project introduces a novel lens for evaluating public digital infrastructure. The Aadhaar Friction Index provides a foundation for smarter governance—where systems are optimized not just for scale, but for citizen experience.

References

[1] Unique Identification Authority of India, *Aadhaar Enrolment ,Demographic and Biometeric datasets*, Government of India. [Online].

[2] K. Muralidharan, P. Niehaus, and S. Sukhtankar, "Identity verification standards in welfare programs: Evidence from Aadhaar," *Journal of Public Economics*, vol. 190, pp. 104–218, 2020.

[3] World Bank Group, *Digital Identity: Toward Shared Principles for Public and Private Sector Cooperation*. Washington, DC, USA: World Bank Publications, 2019.

[4] A. Banerjee, E. Duflo, and N. Qian, "On the road: Access, friction, and public service delivery," *American Economic Review*, vol. 110, no. 7, pp. 1971–2008, 2020.

[5] Organisation for Economic Co-operation and Development (OECD), *Behavioural Insights for Public Policy: Concepts and Case Studies*. Paris, France: OECD Publishing, 2018.

[6] W. McKinney, "Data structures for statistical computing in Python," in *Proc. 9th Python in Science Conf. (SciPy)*, Austin, TX, USA, 2010, pp. 51–56.

[7] J. D. Hunter, "Matplotlib: A 2D graphics environment," *Computing in Science & Engineering*, vol. 9, no. 3, pp. 90–95, 2007.