UIDAI DATA HACKATHON 2026

TEAM LEADER – SHLOK JHA
TEAM LEADER APAAR ID - 717529688040
TEAM ID – UIDAI_7349

TEAM MEMBERS :	**APAAR ID:**
1.ATUL KUMAR	159051633810
2.AKSHAT BEDI	967078791066
3.HARSH KUMAR	455909414197

TEAM LEAD EMAIL ID : shlokjha2321@gmail.com

PROBLEM STATEMENT –

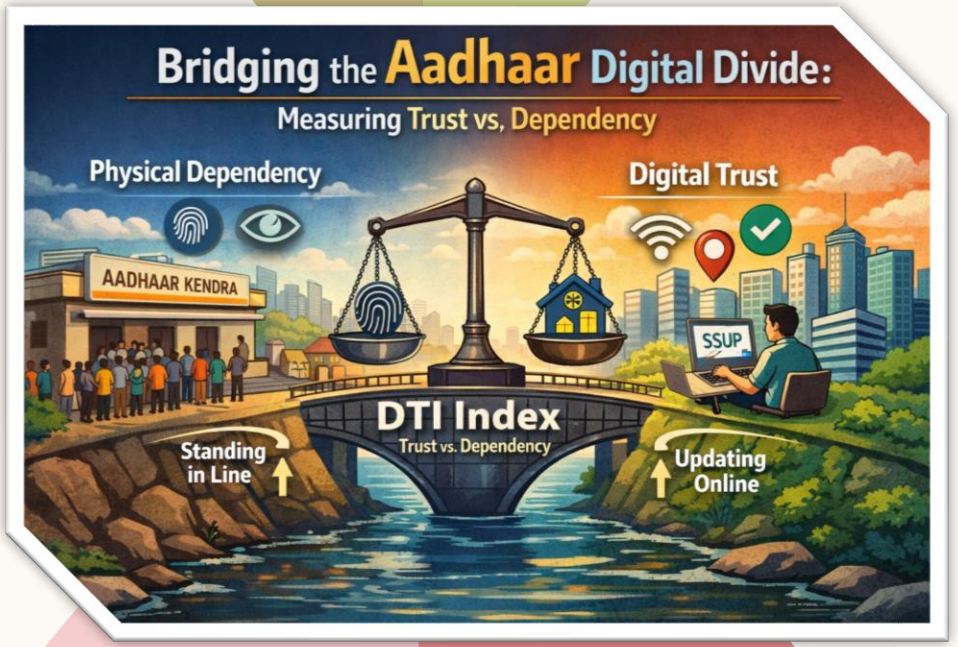
“BRIDGING THE AADHAAR DIGITAL DIVIDE: MEASURING TRUST VS. DEPENDENCY”

(1.) - Problem Statement and Approach:

- **Theme:** Persistent reliance on physical Aadhaar centers for demographic updates, despite the availability of the SSUP, highlights gaps in digital awareness, trust, and access.

- **Core Metric:** A **Digital Trust vs. Physical Dependency Index (DTI)** compares demographic (online-capable) updates with biometric (mandatory physical) updates at the pincode level.

- **Objective:** To identify physical-dependent regions and drive **targeted SSUP awareness and digital literacy interventions** that reduce center load and improve service efficiency.

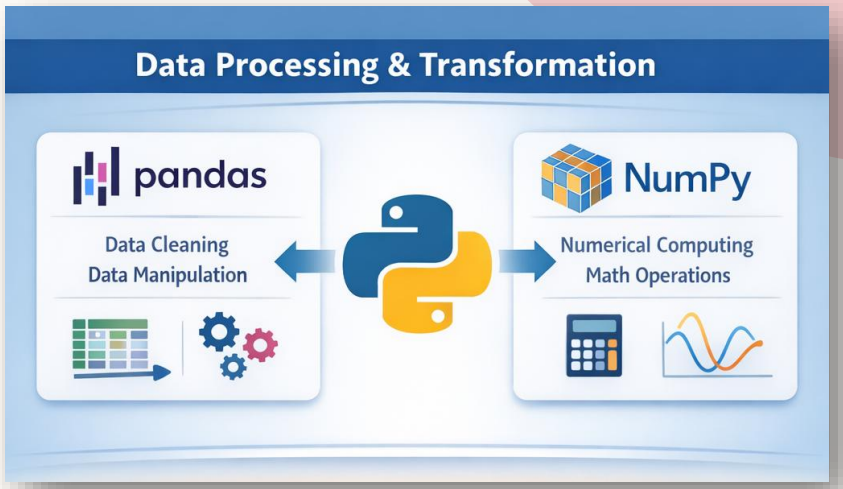


- **Proposed Framework :**

-> **Incentivizing Digital Updates:** Propose a "Digital First" campaign in these specific districts where online updates are processed with higher priority or lower fees (if applicable) to shift the load away from physical centers.

TECHNICAL APPROACH

- **Primary Language: Python**
  - ❖ Chosen for its robust ecosystem and scalability in data science tasks.
- **Data Processing & Transformation**
  - ❖ **Pandas:** Utilized for high-performance data cleaning, manipulation, and structural transformation.
  - ❖ **NumPy:** Employed for advanced mathematical operations and efficient numerical computing.
- **Visualization & Statistical Analysis**
  - ❖ **Matplotlib & Seaborn:** a Used to generate static and interactive graphs for trend analysis and visual storytelling.
  - ❖ **SciPy:** Applied for scientific computing and extending visualization capabilities with statistical algorithms.

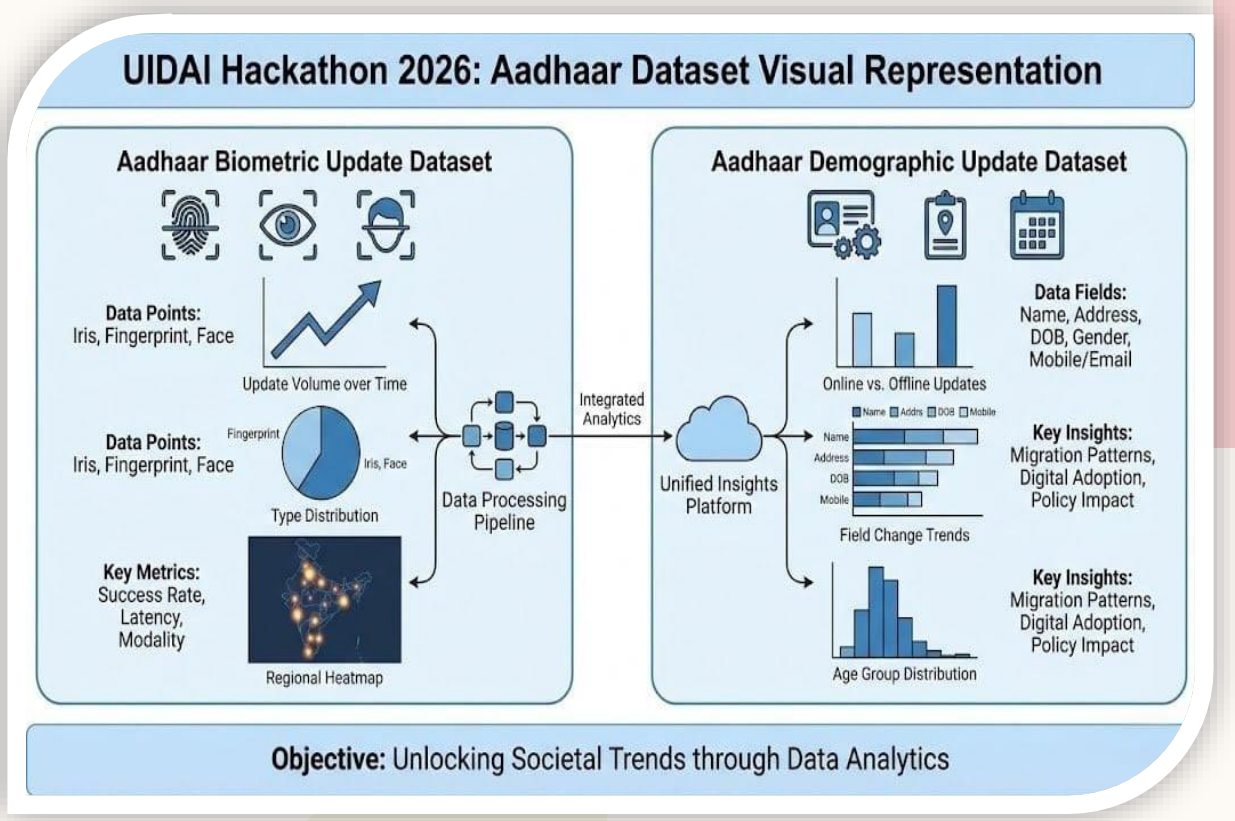


❑ Aadhaar Demographic Update dataset

- This dataset captures aggregated information related to updates made to residents' demographic data linked to Aadhaar, such as name, address, date of birth, gender, and mobile number.
- It provides insights into the frequency and distribution of demographic changes across different time periods and geographic levels (state, district, and PIN code).

❑ Aadhaar Biometric Update dataset

- This dataset contains aggregated information on biometric updates (modalities such as fingerprints, iris, and face).
- It reflects the periodic revalidation or correction of biometric details, especially for children transitioning into adulthood.



TRANSFORMATION

Data Cleaning

- Removed the 'Date' column to eliminate redundancy and streamline the dataset.

District Aggregation

- Consolidated district-level entries by merging duplicates and summing their respective age counts.

Biometric Standardization

- Applied identical cleaning and aggregation protocols to the biometric dataset for consistency

State-level Extraction

- Segregated and extracted state-wise data subsets for both Biometric and Demographic categories

DATA CLEANING

OBJECTIVE :

Ensure **consistency**, **accuracy**, and **reliability** across Aadhaar datasets before analysis.

Noise Removal : (Dropped non-essential columns such as date fields)

Duplicate Consolidation: (Merged repeated district and pincode entries with age-wise summation)

Missing Value Handling: (Removed records with incomplete or NaN values)

Standardization : (Corrected inconsistent state and district naming formats)

	state	bio_age_5_17	bio_age_17_	demo_age_5_17	demo_age_17_
2	Andaman and Nicobar Islands	405	1979	13	1046
3	Andaman and Nicobar Islands	11401	6913	604	5583
4	andhra pradesh	2241448	1473144	321143	1974362
5	andhra pradesh	19	22	5	72
6	Arunachal Pradesh	42894	29500	5783	30660
7	Assam	596624	386098	84480	928098
8	Bihar	2208141	2689446	380023	4434327
9	Chhatisgarh	2	3	0	4
10	Chhatisgarh	884553	1764176	165207	1840227
11	Dadra and Nagar Haveli and Daman and Diu	72	481	32	203
12	Dadra and Nagar Haveli and Daman and Diu	10791	16444	1310	4723
13	Dadra and Nagar Haveli and Daman and Diu	753	1779	306	3434
14	Dadra and Nagar Haveli and Daman and Diu	531	1654	41	603
15	Dadra and Nagar Haveli and Daman and Diu	4215	2548	234	1318
16	Delhi	556237	748125	175535	1263399
17	Goa	34244	34153	3275	31845
18	Gujarat	1460655	1735859	208474	1615853
19	Haryana	693667	941787	139314	1026826
20	Himachal Pradesh	191566	204668	19424	129790
21	Jammu and Kashmir	114	315	23	403
22	Jammu and Kashmir	420235	370983	57850	348926
23	Jharkhand	876613	1149684	99376	1301813
24	Karnataka	1244999	1390955	264981	1430304
25	Kerala	647939	961791	61064	683888
26	Ladakh	2809	2954	1375	4360
27	Lakshadweep	2226	2594	170	1006
28	Madhya Pradesh	3200117	2723654	407098	2505840
29	Maharashtra	3512712	5713427	273322	4781280
30	Manipur	166465	116122	41464	260085
31	Meghalaya	36531	51095	8426	78952
32	Mizoram	85531	34798	5302	36604
33	Nagaland	32561	77032	4314	32477

Raw Data

	state	state_ratio	state_score
2	Andaman and Nicoba	0.350082134	8.224674103
3	Arunachal Pradesh	0.503398072	20.10958181
4	Assam	1.030380922	60.96079798
5	Bihar	0.983004488	57.28822127
6	Chhatisgarh	0.757130765	39.77870218
7	Dadra and Nagar Hav	0.31078741	5.178583964
8	Delhi	1.103170746	66.60339686
9	Goa	0.513472813	20.89056631
10	Gujarat	0.57072392	25.3286184
11	Haryana	0.713037481	36.36063214
12	Himachal Pradesh	0.376580505	10.27880304
13	Jammu and Kashmir	0.514373199	20.96036339
14	Jharkhand	0.691502282	34.69124375
15	Karnataka	0.643139068	30.94217279
16	Kerala	0.462780715	16.9609624
17	Ladakh	0.995141419	58.22906483
18	Madhya Pradesh	0.491737105	19.20563458
19	Maharashtra	0.547856693	23.55597241
20	Manipur	1.067101459	63.8073396
21	Meghalaya	0.99716979	58.3863022
22	Mizoram	0.348261849	8.083567335
23	Nagaland	0.335705748	7.110230153
24	Odisha	0.451149309	16.05930671
25	Pondicherry	0.468658809	17.41662673
26	Punjab	0.532069787	22.33218627
27	Rajasthan	0.705293301	35.7603106
28	Sikkim	0.891323401	50.1811894
29	Tamil Nadu	0.470875459	17.5884594
30	Telangana	0.93799341	53.79900471

Cleaned Data

PREPROCESSING

1. Merged biometric and demographic data at district level

- Shows granularity and that you are capturing the digital divide locally.

2. Created a separate state-level dataset

- Shows multi-level analysis (district vs state), which adds depth.

3. Cleaned and standardized the data

- Signals data reliability and seriousness of preprocessing.

4. Distinguished Aadhaar dependency vs trust indicators

- This directly connects preprocessing to your core idea “Trust vs Dependency”.

DIGITAL TRUST INDEX

This formula calculates how much "extra" load offline centers are carrying beyond what is mandatory (biometric updates).

$$DTI_{Gravity} = \frac{Demographic\ Updates}{Biometric\ Updates}$$

Interpretation:

If the ratio is > 1.0:
The district has **low digital trust**.
People are visiting physical centers for services that could have been completed online *(Demographic updates > Biometric updates)*.

If the ratio is < 1.0:
The district has **high digital trust**.
People are likely using the online portal for demographic changes and visiting centers only when **biometrics are mandatory**.

Z-SCORE

The Statistical "Z-Score" Method (Recommended for Ranking)

Since different districts have different population sizes, a simple ratio might be skewed.
Use a Relative Trust Score to compare districts against the state average.
Calculate the District Ratio (R_d):

1. Demographic / Biometric
   Calculate the State Average Ratio (R_{avg}):

2. Total State Demographic / Total State Biometrics.

3. Index Score:

$$DTI_{Index} = \frac{R_d}{R_{avg}}$$

Score > 1: “Center Reliant” (Lower Digital Trust than average).
Score < 1: “Digital Adopters” (Higher Digital Trust than average).

A Description of Key Findings and Insights

Decoding the Data – The Digital Trust Index (DTI)

1. The Urban Paradox (Metro Anomaly)

(**i**)Data: Najafgarh (Delhi) scored a DTI of 0.20, lagging 81% behind the State Average (1.10).

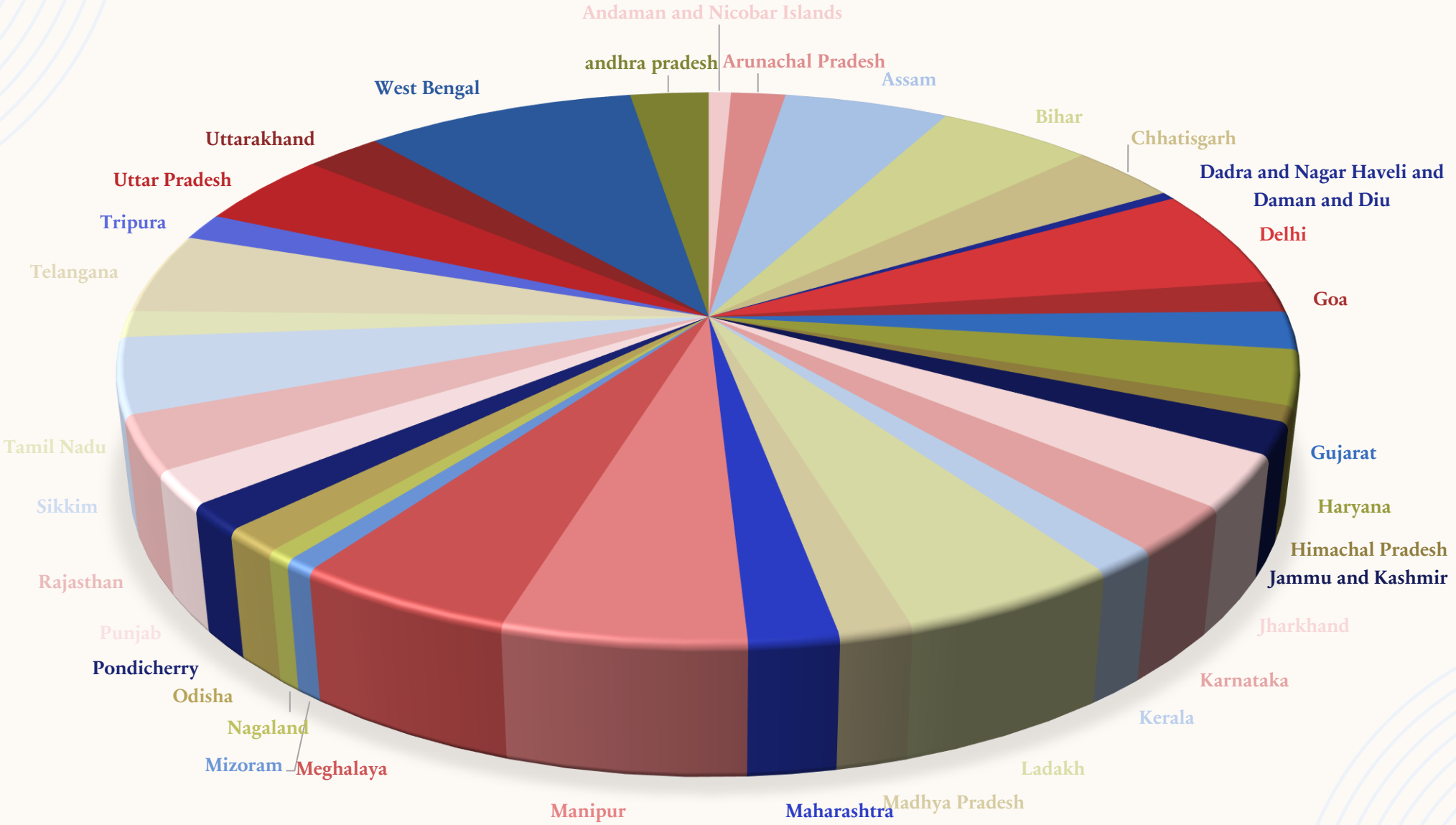
(**ii**)Insight: Proximity to a city ≠ Digital Adoption. Even metro pockets remain "Physically Dependent."

2. The Habit Gap (Rural Insight)

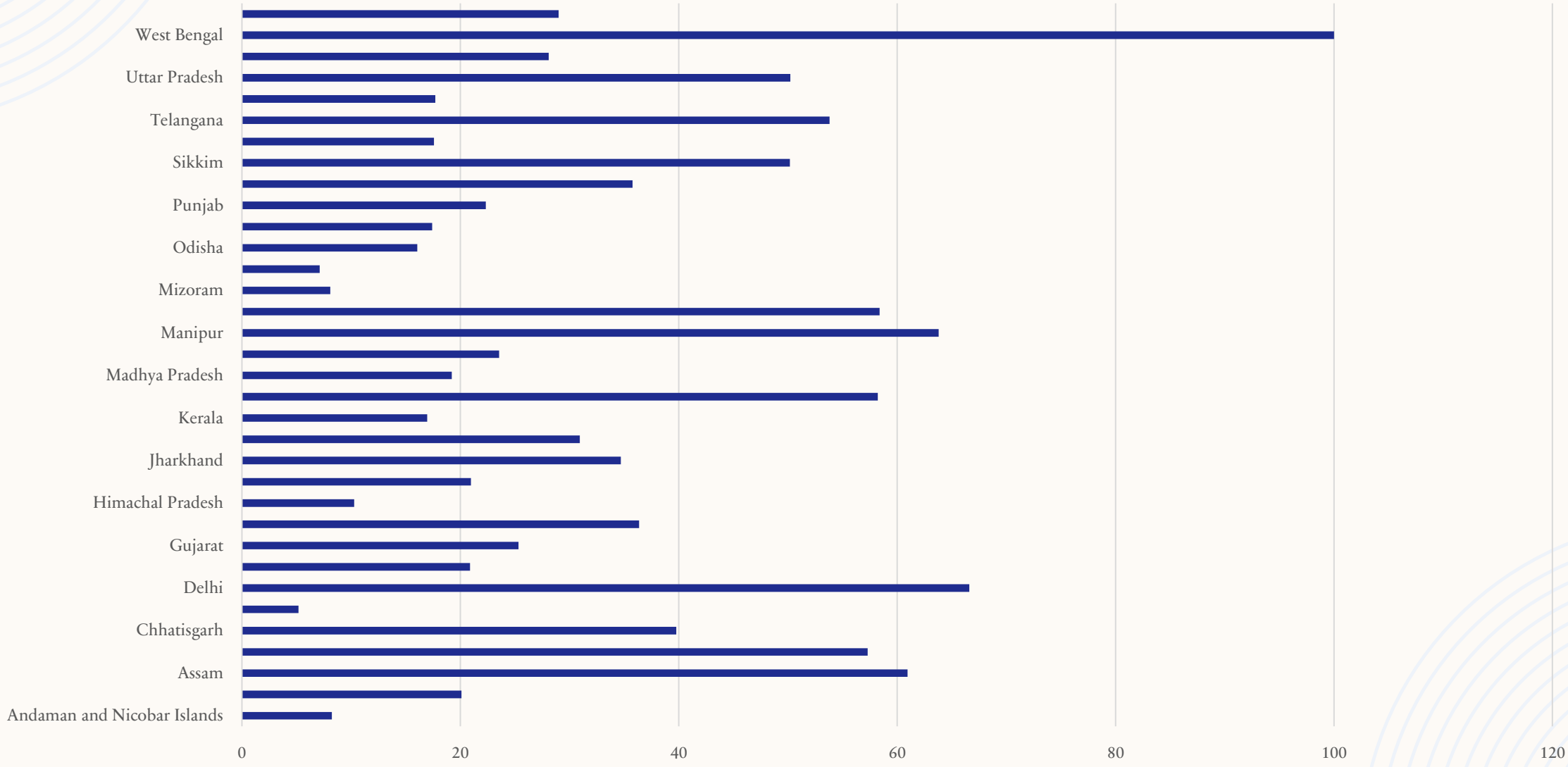
(**i**)Data: Purbi Singhbhum shows high Biometric volume but near-zero SSUP usage.

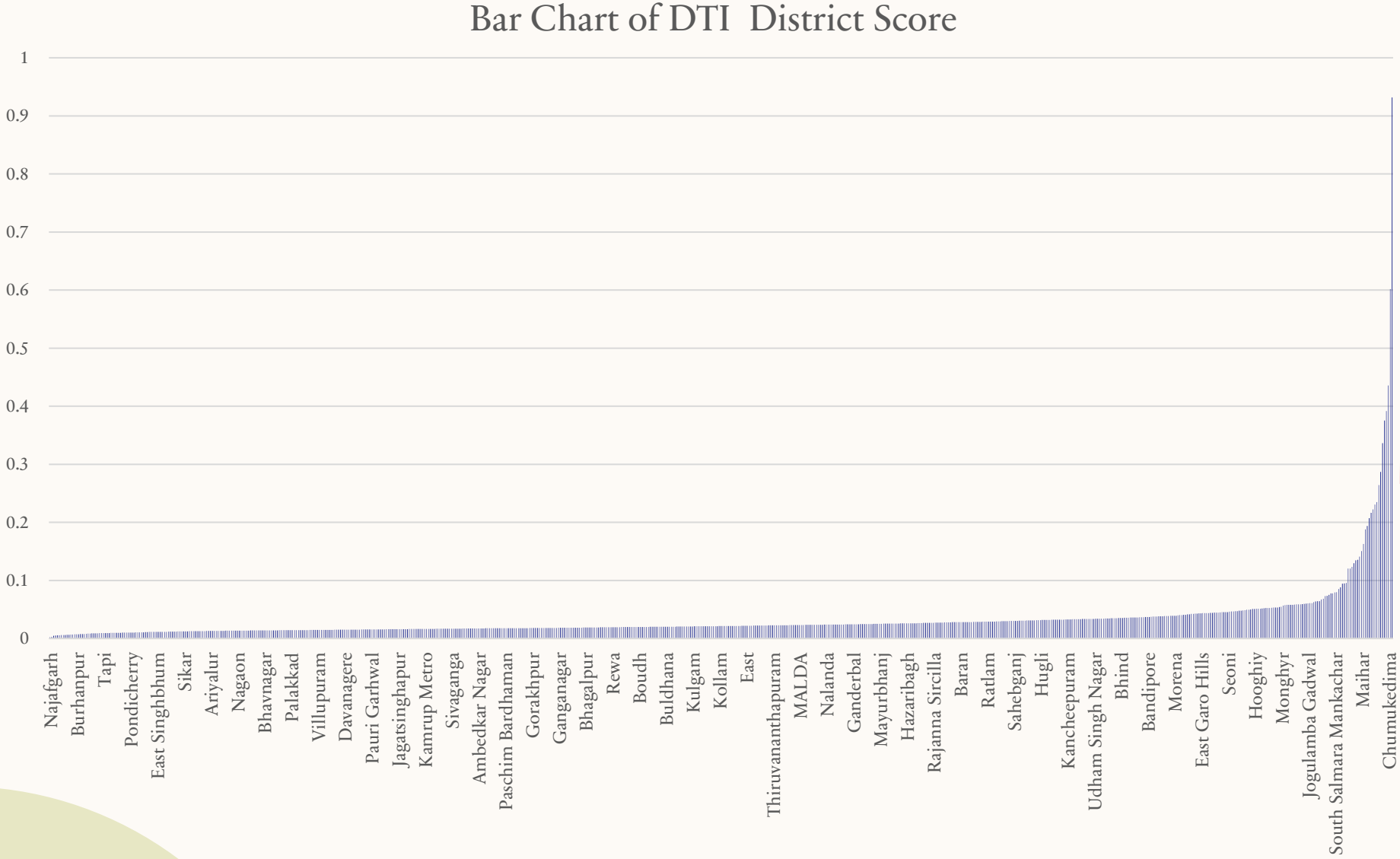
(**ii**)Insight: The issue is Awareness, not Access. Citizens visit centers for tasks they could do online.

PIE CHART OF DTI STATE SCORE



Bar Chart of DTI State score





Code File - [GitHub Repository]