

<DATA ANALYTICS WITH POWER BI>

PROJECT REPORT

August 2025-January 2026

NFHS India Health Analytics and Visualization with Power BI

Submitted by

Anup Pandey

Registration No. 12310363

Bachelor of Technology
Computer Science and Engineering
Course Code INT374

Under the Guidance of

B Monica
UID: 32352

Discipline of CSE/IT

Lovely School of Computer Science

Lovely Professional University, Phagwara

CERTIFICATE

This is to certify that **Anup Pandey** bearing Registration no. **12310363** has completed **INT374** project titled, "**NFHS India Health Analytics and Visualization with Power BI**" under my guidance and supervision. To the best of my knowledge, the present work is the result of his/her original development, effort and study.

Signature and Name of the Supervisor

Designation of the Supervisor

School of Computer Science and Engineering

Lovely Professional University

Phagwara, Punjab.

Date: 19/12/2025

DECLARATION

I, Anup Pandey, student of Bachelor of Technology under CSE/IT Discipline at, Lovely Professional University, Punjab, hereby declare that all the information furnished in this project report is based on my own intensive work and is genuine.

Date: 19/12/2025

Signature

Registration No. 12310363

Anup Pandey

ACKNOWLEDGEMENT

I would like to express my sincere gratitude to everyone who contributed to the successful completion of this project titled "**NFHS India Health Analytics and Visualization using Power BI.**"

I am deeply thankful to my faculty guide for their continuous guidance, valuable suggestions, and encouragement throughout the course of this project. Their insights helped me understand the practical aspects of data analysis and dashboard design.

I also extend my gratitude to my institution for providing the necessary resources and learning environment to carry out this project effectively. Special thanks to the data providers whose publicly available datasets made this analysis possible.

Lastly, I would like to thank my friends and peers for their support, constructive feedback, and motivation during the development of this project.

Contents

ACKNOWLEDGEMENT	4
Introduction	6
Source of dataset.....	7
Data Preprocessing	8
Data Modeling	11
Dashboard Design & Pages	12
Conclusion.....	19
Future Enhancements	20
References	21

Introduction

Public health data plays a critical role in shaping national policies, healthcare programs, and social interventions. In India, the National Family Health Survey (NFHS) is one of the most comprehensive and authoritative datasets capturing health, nutrition, and demographic indicators across states and districts.

However, NFHS data is:

- Extremely **wide** (100+ indicators),
- Spread across **multiple granularities** (national, state, district, time),
- Difficult to analyze without **robust preprocessing and modeling**.

This project focuses on transforming raw NFHS datasets into a **scalable, interactive, and insight-driven Power BI dashboard**, emphasizing **data preprocessing, dimensional modeling, and analytical storytelling** rather than simple visualization.

The primary objectives of this project are:

- To preprocess and clean large-scale public health datasets.
- To design a **proper star-schema data model** for analytical use.
- To create a **dynamic Power BI dashboard** enabling multi-level analysis.
- To extract insights related to:
 - Maternal & child health
 - Nutrition and malnutrition
 - Non-communicable diseases (NCDs)
 - Household amenities and digital access
- To demonstrate **end-to-end analytics skills**, including Power Query, data modeling, and DAX.

Source of dataset

All datasets used in this project are **official government sources** obtained from data.gov.in.

Dataset 1: All-India & State/UT Factsheets (NFHS-5)

- Granularity: State / UT
- Rows: ~36
- Columns: ~134 indicators

🔗 <https://www.data.gov.in/resource/all-india-and-stateut-wise-factsheets-national-family-health-survey-nfhs-5-2019-2021>

Dataset 2: District-wise Factsheets (NFHS-5)

- Granularity: District
- Rows: ~13,900+
- Columns: ~134 indicators

🔗 <https://www.data.gov.in/resource/india-districts-factsheets-national-family-health-survey-nfhs-5-2019-2021-provisional>

Dataset 3: Year-wise Child Health Indicators (NFHS-1 to NFHS-5)

- Granularity: National, Time-series
- Covers surveys from **1992–93 to 2019–21**
- Indicators: Stunting, Wasting, Underweight

🔗 <https://www.data.gov.in/resource/year-wise-details-indicators-children-national-family-health-survey-nfhs-1-nfhs-5-1992-93>

Data Preprocessing

Data preprocessing was the **core foundation** of this project.

Structural Challenges

- Extremely wide tables (100+ indicator columns)
 - Inconsistent indicator naming
 - Duplicate indicator definitions across datasets
 - Non-standard column formatting
 - Mixed metadata and data columns
-

Key Preprocessing Steps

a. Unpivoting

- Converted wide indicator columns into a **long-format fact table**
 - Created:
 - indicator
 - value
 - Enabled flexible filtering and analysis
-

b. Creation of DimIndicator

A dedicated **Indicator Dimension Table** was created to standardize and manage indicators. Fields included:

- indicator_id
- indicator_full (original NFHS indicator name)
- indicator_short (clean, dashboard-friendly name)
- category (logical grouping)

This avoided:

- Repetitive renaming in fact tables
 - Hard-coded logic in visuals
 - Model complexity
-

c. Indicator Categorization

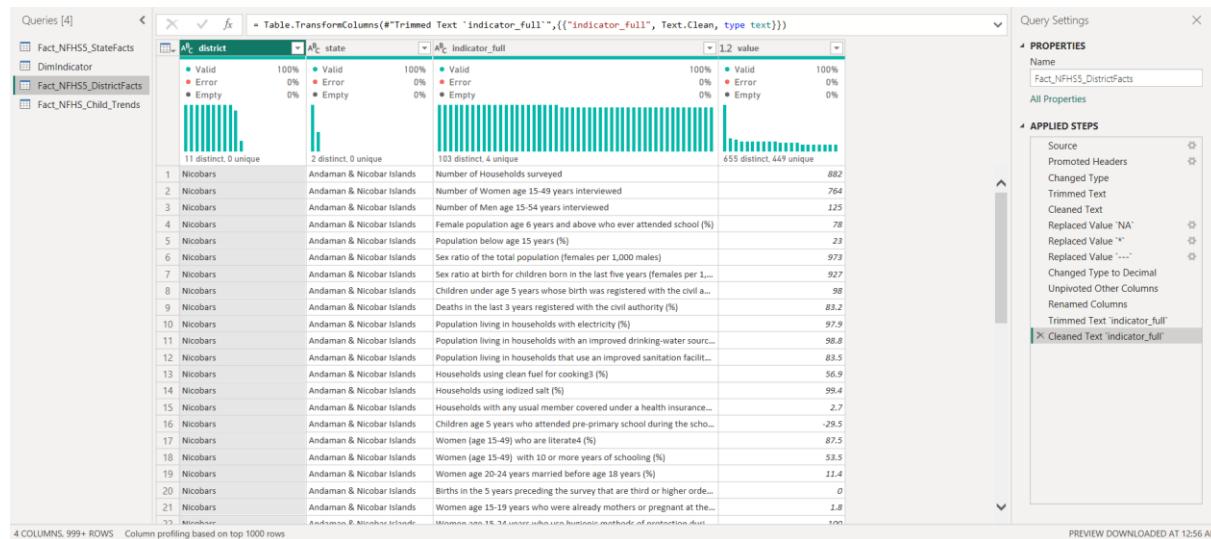
Indicators were grouped into meaningful analytical categories such as:

- Demographics & Basic Info
 - Education & Awareness
 - Household Amenities
 - Family Planning
 - Maternal Health
 - Child Health & Nutrition
 - Immunization
 - Non-Communicable Diseases
 - Social Indicators

This enabled **category-level slicing** across the entire dashboard.

d. Data Cleaning

- Removed extra spaces and hidden characters
 - Fixed encoding issues ($\%$, \geq , \leq symbols)
 - Ensured numeric columns were correctly typed
 - Handled missing and blank values gracefully



Queries [4] < ✓ fx - Table.TransformColumnTypes(#"Replaced Value Stunting %",{{"value", type number}}) >

Query Settings

PROPERTIES

- Name: Fact_NFHS_Child_Trends
- All Properties

APPLIED STEPS

- Source
- Promoted Headers
- Changed Type
- Renamed Columns
- Trimmed Text
- Cleaned Text
- Unpivoted Columns
- Renamed Columns "indicator_..."
- Replaced Value Underweight %
- Replaced Value Wasting %
- Replaced Value Stunting %

Fact_NFHS_StateFacts
DimIndicator
Fact_NFHS_DistrictFacts
Fact_NFHS_Child_Trends

	A _c NFHS_Round	A _c Year	A _c indicator_full	1.2 value
• Valid	100%	• Valid	100%	• Valid
• Error	0%	• Error	0%	• Error
• Empty	0%	• Empty	0%	• Empty
5 distinct, 0 unique	5 distinct, 0 unique	3 distinct, 0 unique	15 distinct, 15 unique	
1 NFHS-1	1992-93	Children under 5 years who are underweight (height-for-age)<18 (%)		53.4
2 NFHS-1	1992-93	Children under 5 years who are wasted (weight-for-height)<18 (%)		17.5
3 NFHS-1	1992-93	Children under 5 years who are stunted (height-for-age)<18 (%)		52
4 NFHS-2	1998-99	Children under 5 years who are underweight (height-for-age)<18 (%)		47
5 NFHS-2	1998-99	Children under 5 years who are wasted (weight-for-height)<18 (%)		15.5
6 NFHS-2	1998-99	Children under 5 years who are stunted (height-for-age)<18 (%)		45.5
7 NFHS-3	2005-06	Children under 5 years who are underweight (height-for-age)<18 (%)		42.5
8 NFHS-3	2005-06	Children under 5 years who are wasted (weight-for-height)<18 (%)		19.8
9 NFHS-3	2005-06	Children under 5 years who are stunted (height-for-age)<18 (%)		48
10 NFHS-4	2015-16	Children under 5 years who are underweight (height-for-age)<18 (%)		35.8
11 NFHS-4	2015-16	Children under 5 years who are wasted (weight-for-height)<18 (%)		21
12 NFHS-4	2015-16	Children under 5 years who are stunted (height-for-age)<18 (%)		38.4
13 NFHS-5	2019-21	Children under 5 years who are underweight (height-for-age)<18 (%)		32.1
14 NFHS-5	2019-21	Children under 5 years who are wasted (weight-for-height)<18 (%)		19.3
15 NFHS-5	2019-21	Children under 5 years who are stunted (height-for-age)<18 (%)		35.5

Queries [4]

Fact_NHHS_StateFacts
DimIndicator
Fact_NHHS_DistrictFacts
Fact_NHHS_Child_Trends

Table.SelectColumns(WithID, {"indicator_id", "indicator_full", "indicator_short", "category"})

indicator_id	indicator_full	indicator_short	category
1	Valid	Valid	Valid
2	Error	Error	Error
3	Empty	Empty	Empty
134 distinct, 134 unique	134 distinct, 134 unique	134 distinct, 134 unique	13 distinct, 0 unique
1	Number of Households surveyed	Households Surveyed	Demographics & Basic Info
2	Number of Women age 15-49 years interviewed	Women Interviewed (15-49)	Demographics & Basic Info
3	Number of Men age 15-54 years interviewed	Men Interviewed (15-54)	Demographics & Basic Info
4	Female population age 6 years and above who ever attended school (%)	Female Ever Attended School (%)	Education & Awareness
5	Population below age 15 years (%)	Child Population (%)	Demographics & Basic Info
6	Sex ratio at the total population (females per 1,000 males)	Sex Ratio of the total population (females per 1,000 males)	Other
7	Sex ratio at birth for children born in the last five years (females per 1...)	Sex Ratio at Birth	Demographics & Basic Info
8	Children under age 5 years whose birth was registered with the civil authority (%)	Birth Registration (>5 yrs) (%)	Other
9	Deaths in the last 3 years registered with the civil authority (%)	Deaths Registered (3 yrs) (%)	Other
10	Population living in households with electricity (%)	Households with Electricity (%)	Household Amenities
11	Population living in households with an improved drinking-water source (%)	Improved Drinking Water (%)	Household Amenities
12	Population living in households that use an improved sanitation facility (%)	Improved Sanitation (%)	Household Amenities
13	Households using clean fuel for cooking (%)	Clean Fuel for Cooking (%)	Household Amenities
14	Households using iodized salt (%)	Households using Iodized Salt (%)	Household Amenities
15	Households with any usual member covered under a health insurance plan (%)	Households with Health Insurance (%)	Household Amenities
16	Women (age 15-49) who are literate (%)	Women Literacy (%)	Education & Awareness
17	Men (age 15-49) who are literate (%)	Men Literacy (%)	Education & Awareness
18	Women (age 15-49) with 10 or more years of schooling (%)	Women Schooling 10+ (%)	Education & Awareness
19	Men (age 15-49) with 10 or more years of schooling (%)	Men Schooling 10+ (%)	Education & Awareness
20	Women (age 15-49) who have ever used the Internet (%)	Women Internet Use (%)	Digital Access
21	Men (age 15-49) who have ever used the Internet (%)	Men Internet Use (%)	Digital Access
22	Minors (age 0-17) who received basic health care (%)	Excl. Education (Minors (%)	Healthcare & Enrollment (%)

4 COLUMNS, 134 ROWS Column profiling based on top 1000 rows

Query Settings

PROPERTIES Name DimIndicator All Properties

APPLIED STEPS Source OnlyIndicator DistinctIndicators Cleaned AddedShort AddedCategory WithID Reordered

PREVIEW DOWNLOADED AT 12:56 AM

Data Modeling

A **star schema** was implemented for performance and scalability.

Fact Tables

- Fact_NFHS5_StateFacts
- Fact_NFHS5_DistrictFacts
- Fact_NFHS_TrendFacts

Dimension Tables

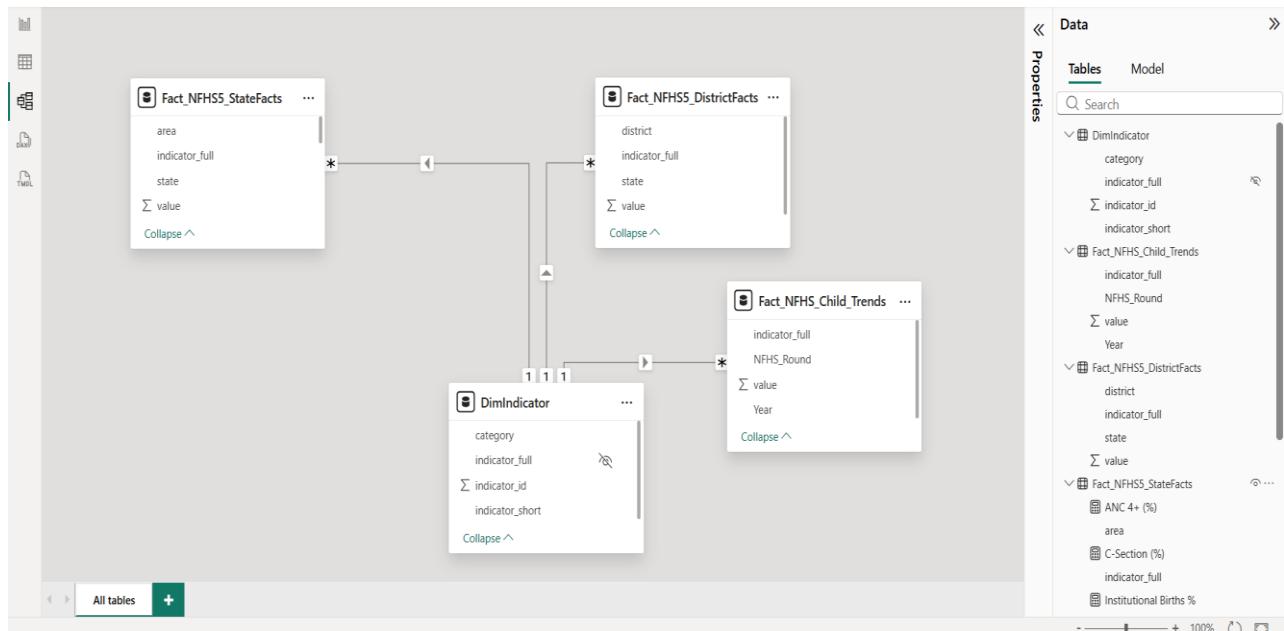
- DimIndicator
- DimState
- DimDistrict
- DimSurveyYear

Model Highlights

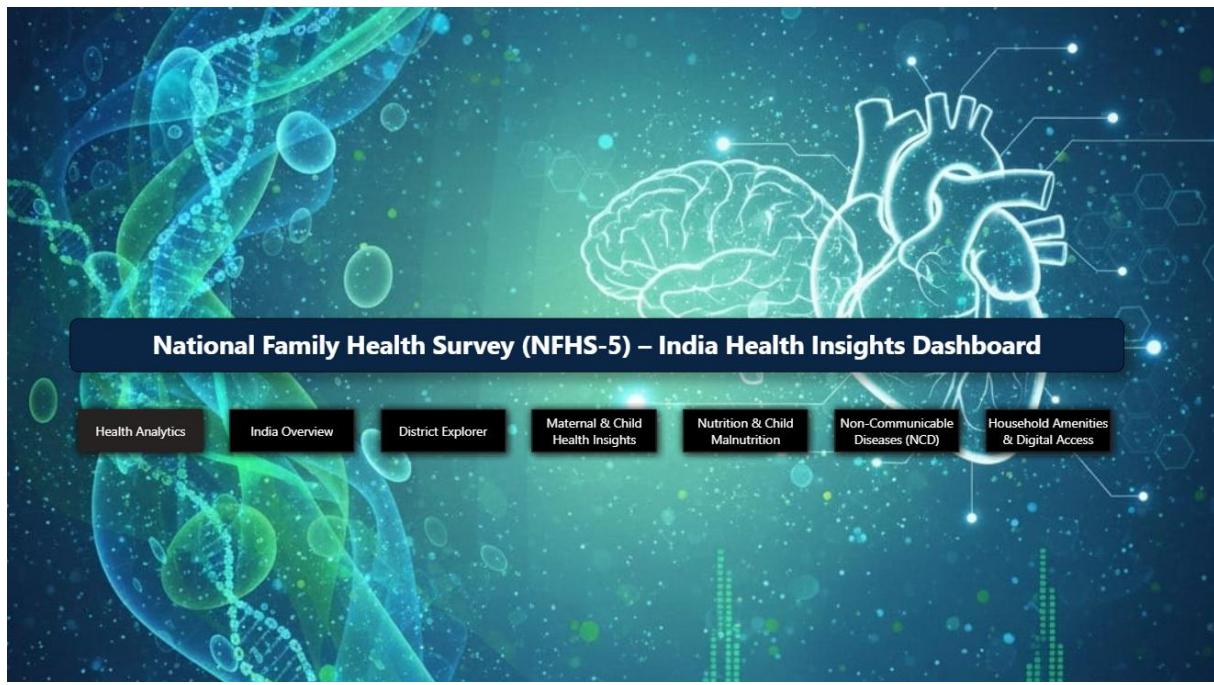
- One-to-many relationships
- Single-direction filtering
- Shared dimensions across multiple fact tables
- No circular dependencies

This model supports:

- Cross-filtering
- Drill-downs
- Multi-dataset integration

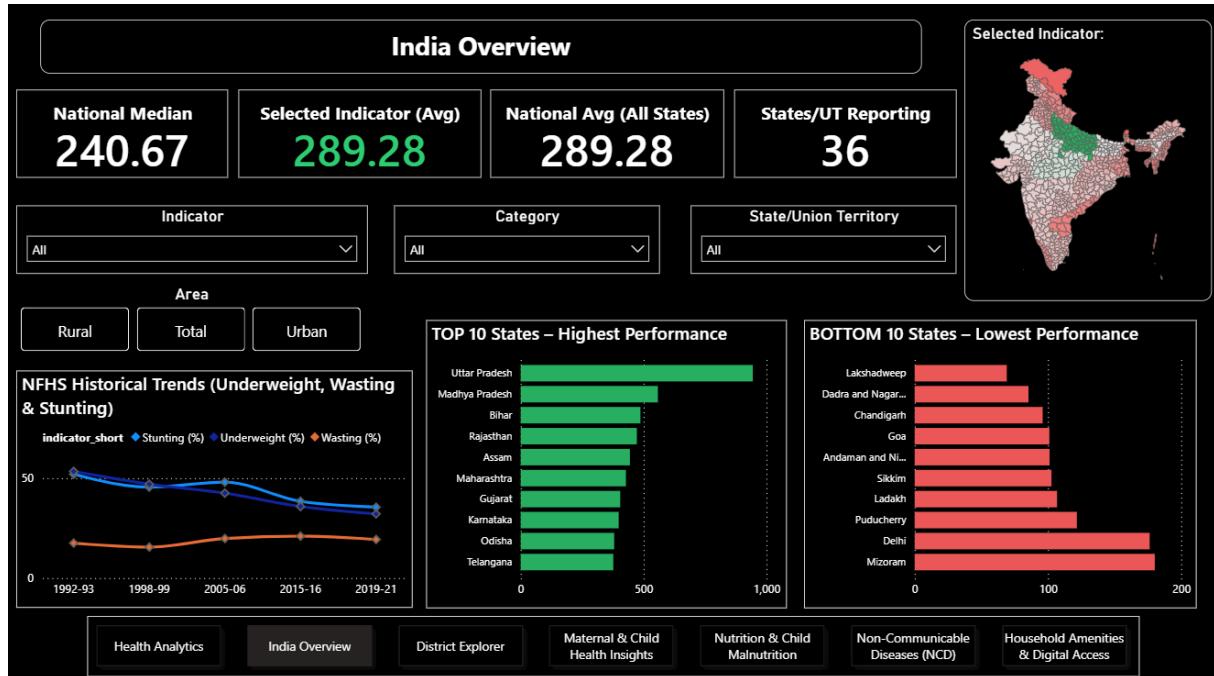


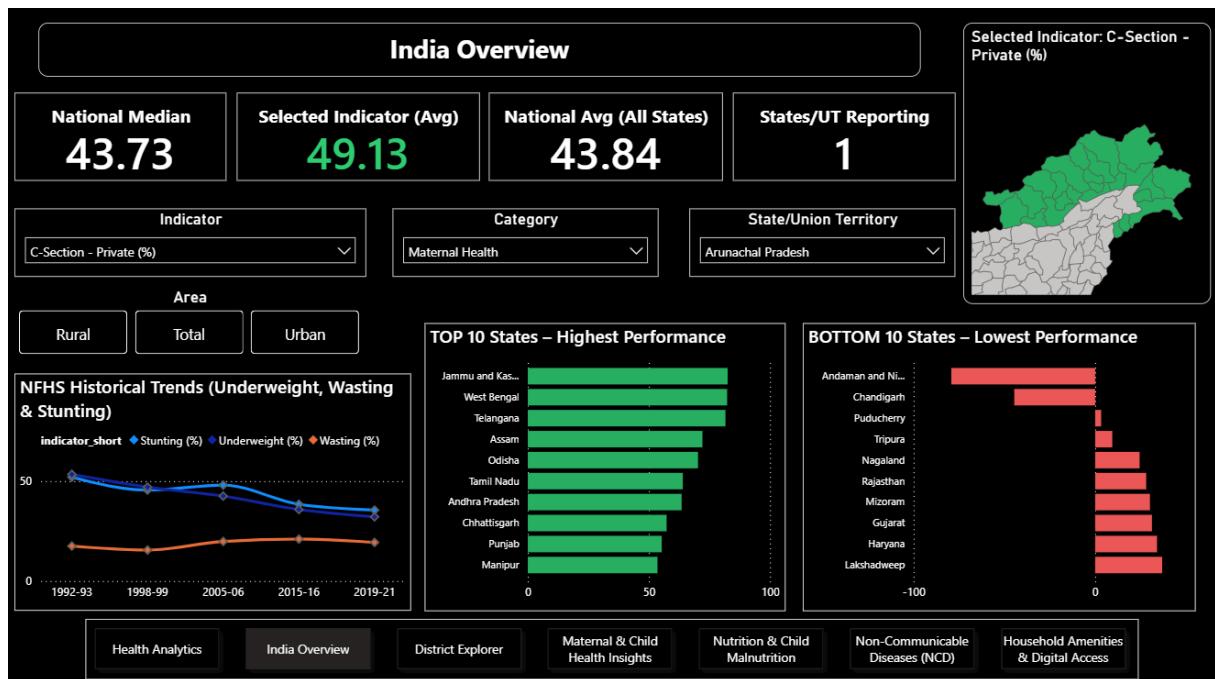
Dashboard Design & Pages



Page 1: National & State Overview

- **KPI cards (key health indicators)**
- **Top N states comparison**
- **Interactive state slicer**

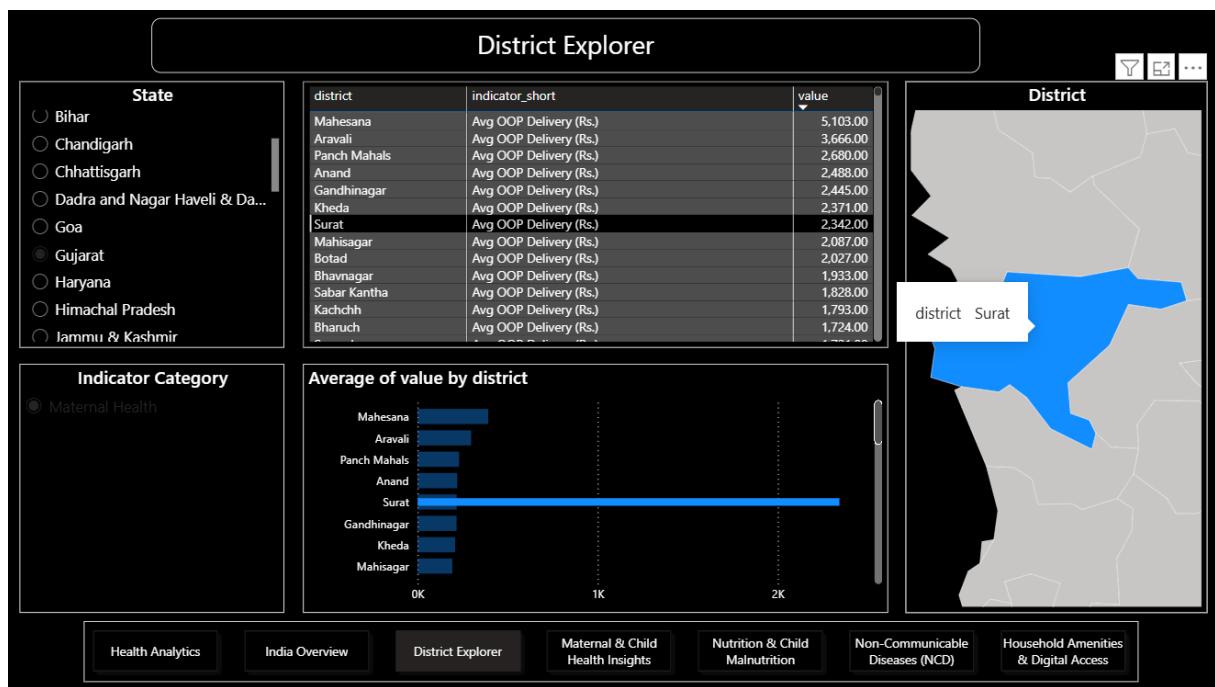




Page 2: District Explorer

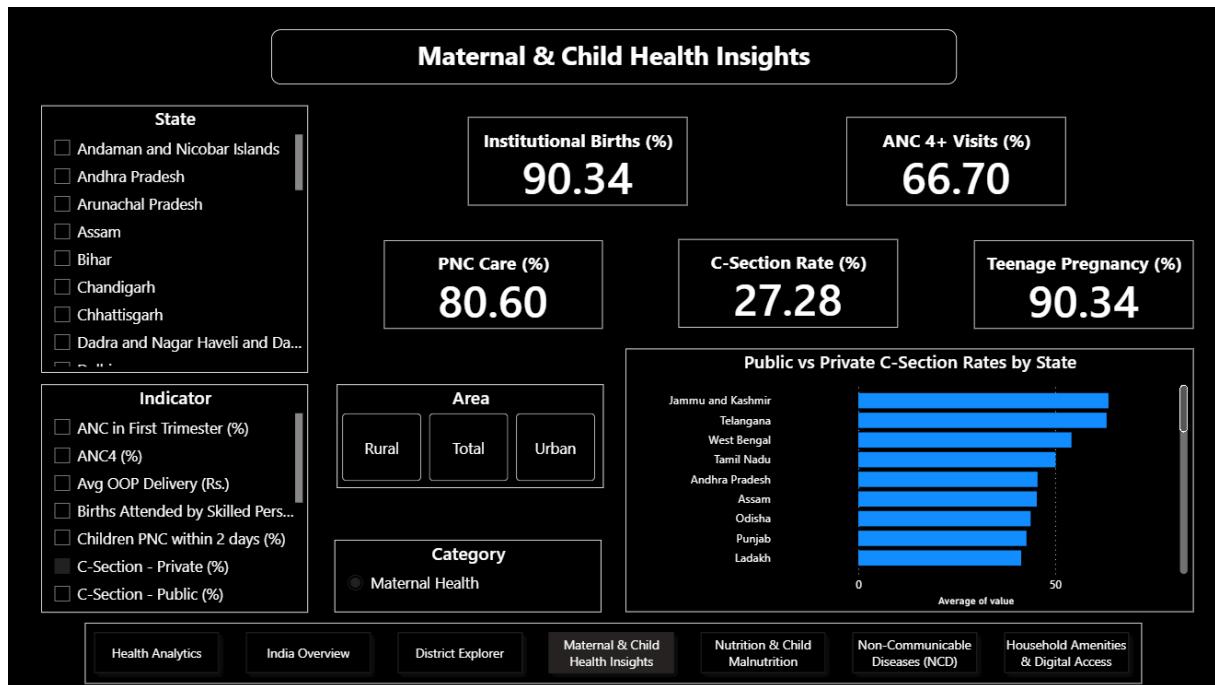
- State-based district filtering
- District-level table view
- District heat maps
- Category-based bar charts

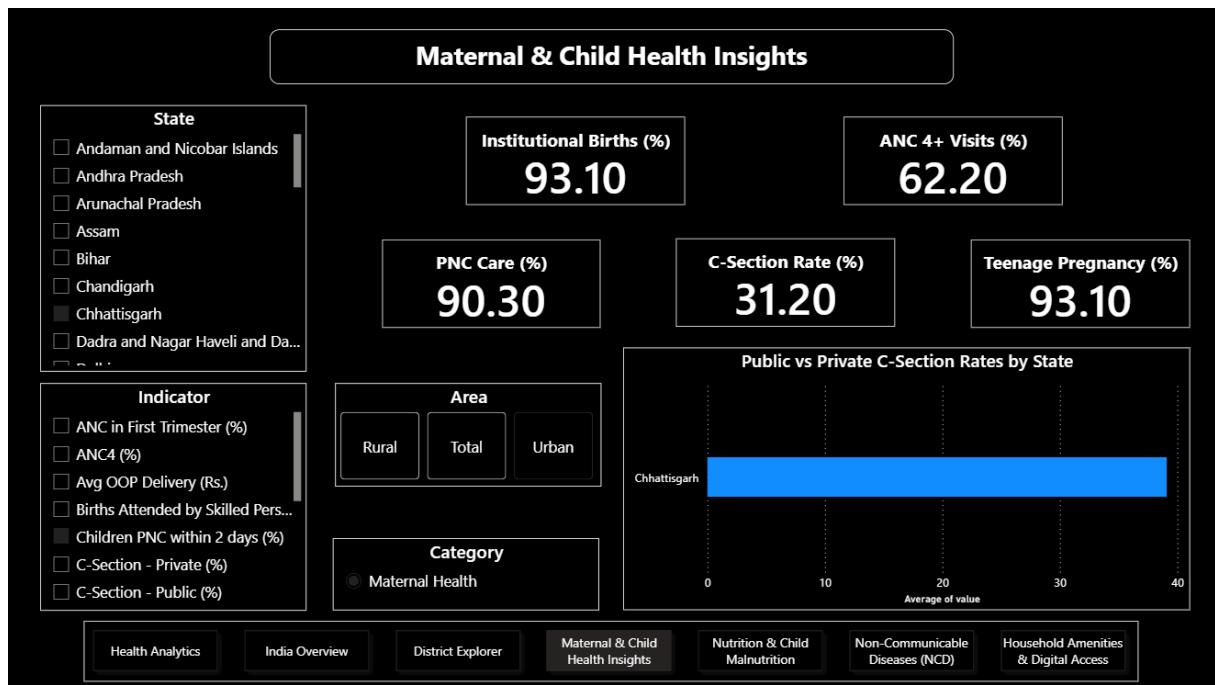




Page 3: Maternal & Child Health

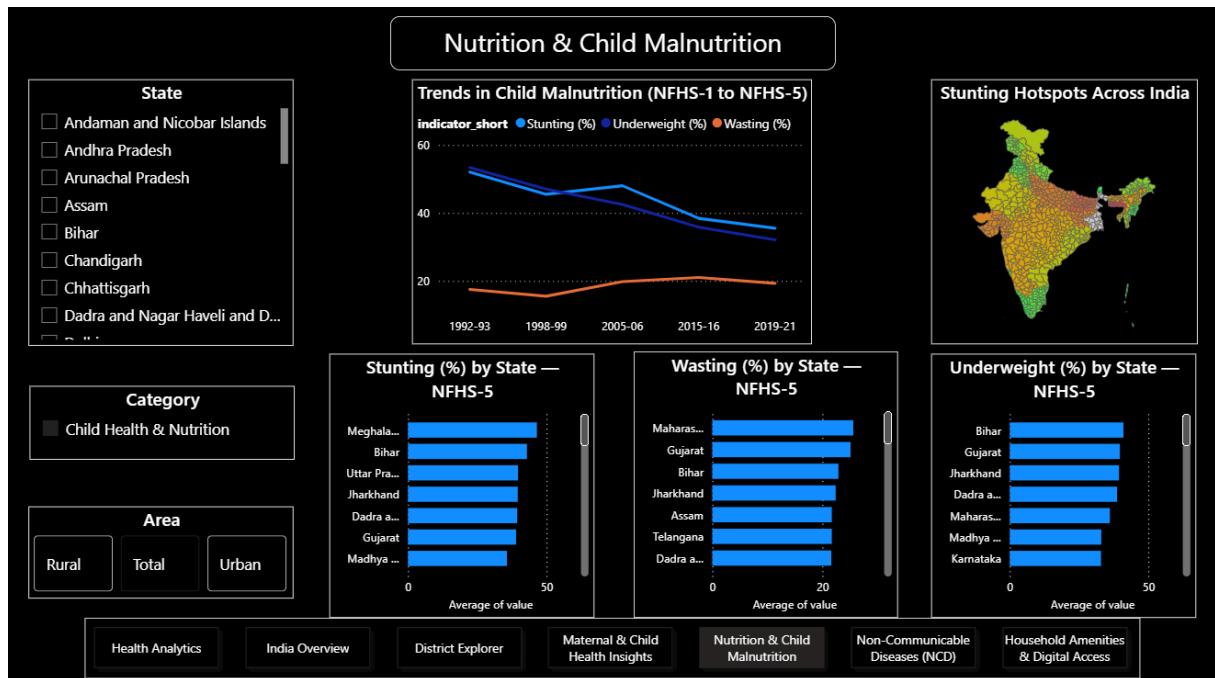
- **Institutional births**
- **ANC 4+ visits**
- **PNC coverage**
- **Public vs Private C-section comparison**
- **Teenage pregnancy indicators**

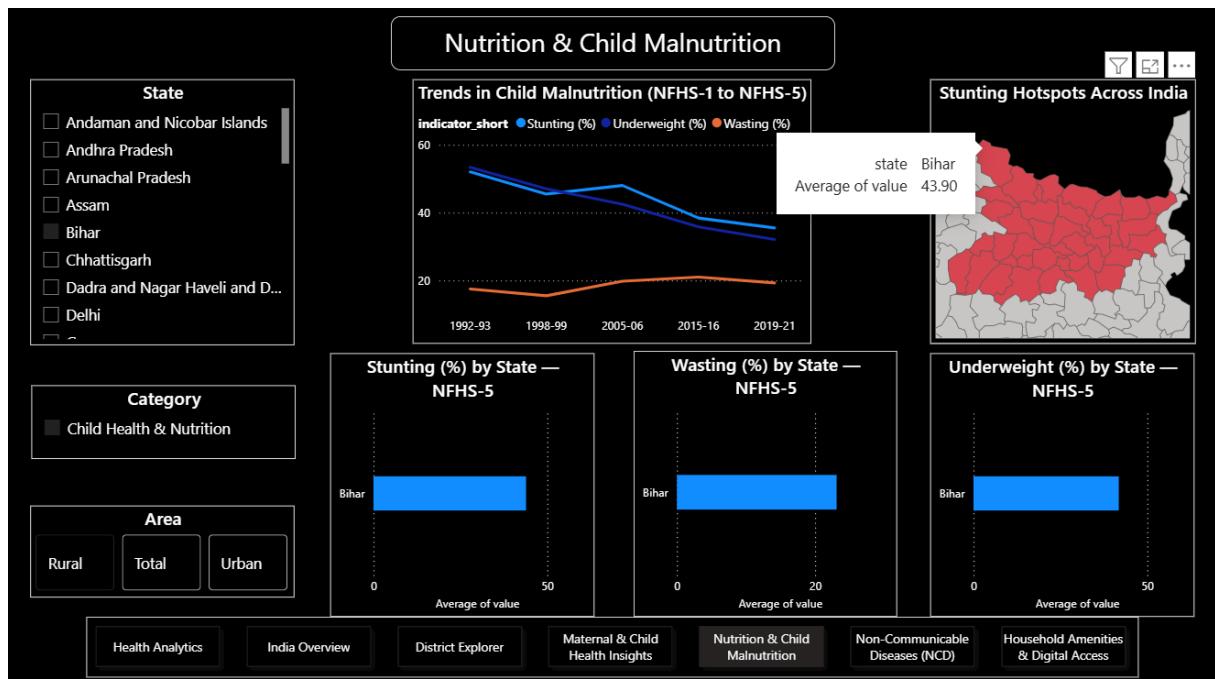




Page 4: Nutrition & Child Malnutrition

- Stunting, wasting, underweight by state
- Malnutrition hotspots (map)
- NFHS-1 to NFHS-5 trend analysis





Page 5: Non-Communicable Diseases (NCD)

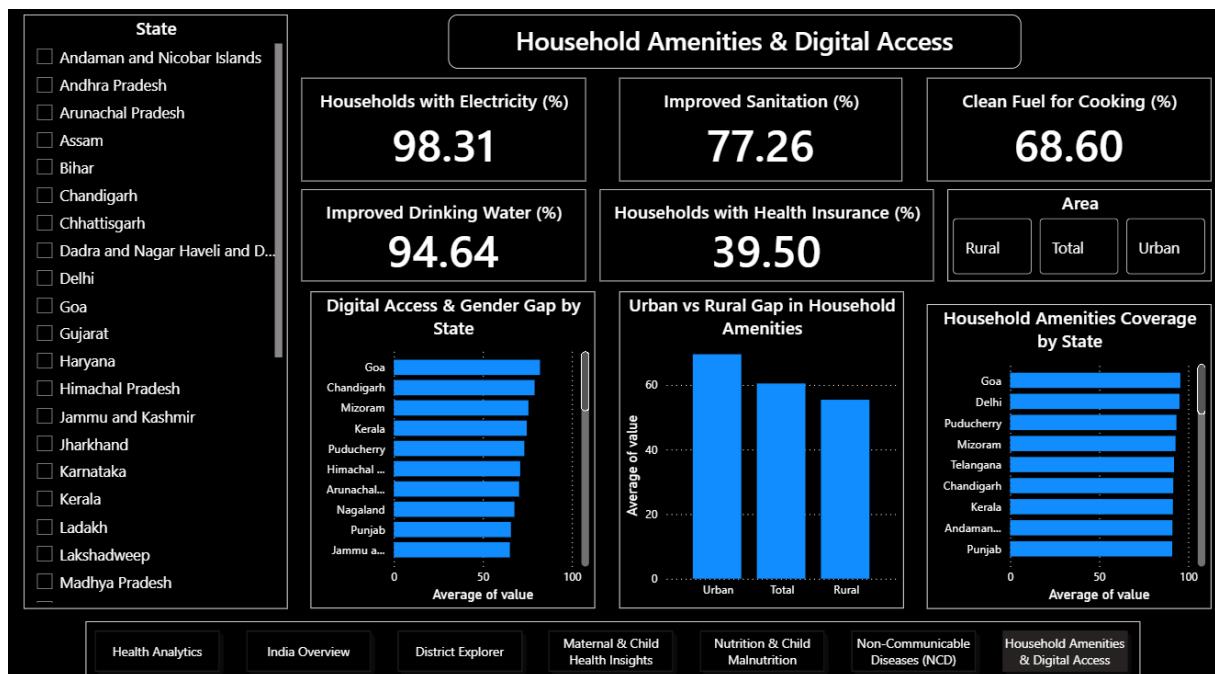
- **Anaemia (women, men, children)**
- **BMI indicators**
- **Blood pressure and blood sugar levels**
- **State-level heatmaps**

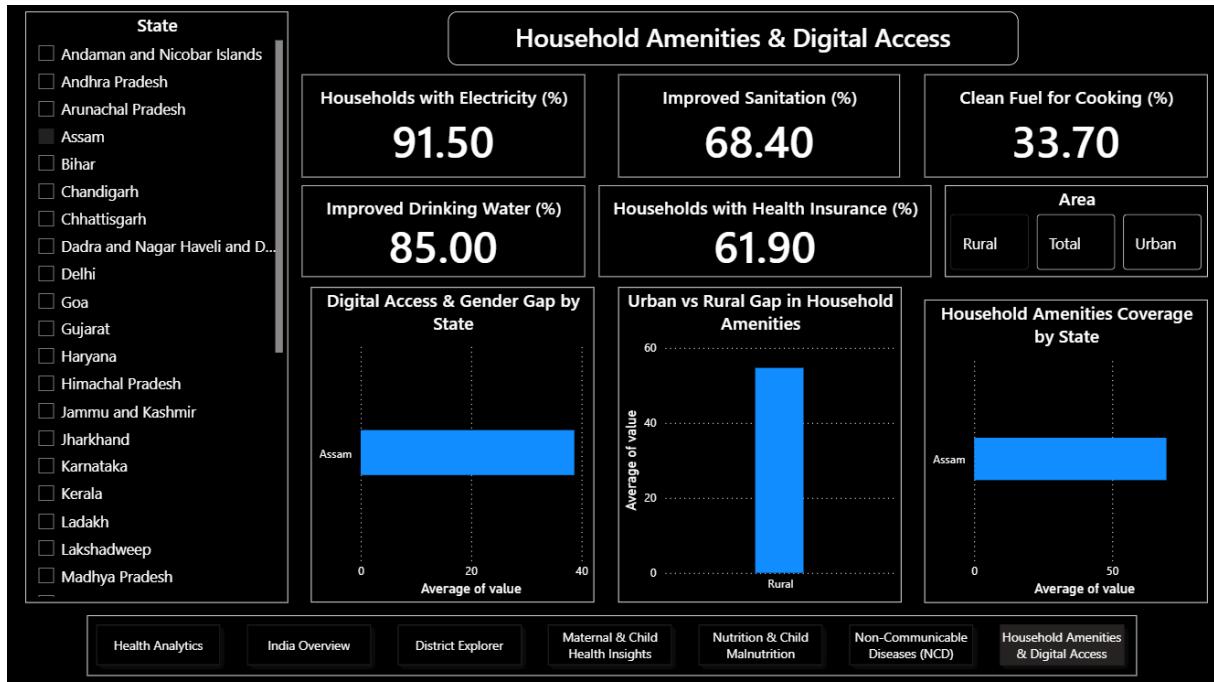




Page 6: Household Amenities & Digital Access

- Electricity, sanitation, drinking water**
- Clean fuel usage**
- Health insurance coverage**
- Internet & mobile usage (gender gap)**





DAX & Measures

- Aggregations using AVERAGE, SUM, and CALCULATE
- Top N logic using visual-level filters
- Context-aware measures responding to slicers
- Time-series aggregation for NFHS trends

Key Insights

- Significant inter-state and inter-district disparities in health indicators
- Persistent child malnutrition, though improving over survey rounds
- Wide variation in C-section rates between public and private facilities
- Growing digital access, but with gender gaps
- Anaemia remains a major public health concern

Tools & Technologies Used

- Power BI Desktop
- Power Query (M Language)
- DAX
- Government Open Data (data.gov.in)

Conclusion

This project demonstrates how **strong data preprocessing, clean modeling, and thoughtful visualization** can transform complex public health datasets into meaningful insights. Beyond dashboard creation, the project highlights:

- Analytical thinking
- Data engineering fundamentals
- Real-world problem solving
- Scalable BI design

The NFHS India Health Analytics Dashboard can support:

- Policy analysis
- Academic research
- Public health planning
- Data-driven decision-making

Future Enhancements

- Integration with NFHS-6 (when released)
- Urban vs Rural breakdown (if data available)
- Predictive modeling on health outcomes
- Automated data refresh pipelines
- Advanced statistical correlation analysis

References

- Ministry of Health and Family Welfare (MoHFW), Government of India.
National Family Health Survey (NFHS-5), 2019–2021: All-India and State/UT Factsheets.
[Data.gov.in](https://www.data.gov.in/resource/all-india-and-stateut-wise-factsheets-national-family-health-survey-nfhs-5-2019-2021)
Available at:
<https://www.data.gov.in/resource/all-india-and-stateut-wise-factsheets-national-family-health-survey-nfhs-5-2019-2021>
- Ministry of Health and Family Welfare (MoHFW), Government of India.
National Family Health Survey (NFHS-5), 2019–2021: District-wise Factsheets (Provisional).
[Data.gov.in](https://www.data.gov.in/resource/india-districts-factsheets-national-family-health-survey-nfhs-5-2019-2021-provisional)
Available at:
<https://www.data.gov.in/resource/india-districts-factsheets-national-family-health-survey-nfhs-5-2019-2021-provisional>
- Ministry of Health and Family Welfare (MoHFW), Government of India.
Year-wise Details of Indicators for Children from NFHS-1 to NFHS-5 (1992–93 to 2019–21).
[Data.gov.in](https://www.data.gov.in/resource/year-wise-details-indicators-children-national-family-health-survey-nfhs-1-nfhs-5-1992-93)
Available at:
<https://www.data.gov.in/resource/year-wise-details-indicators-children-national-family-health-survey-nfhs-1-nfhs-5-1992-93>
- Udit-001 (GitHub Repository).
India Maps Data – TopoJSON Files for Indian States and Districts.
GitHub.
Available at:
<https://github.com/udit-001/india-maps-data>
- Microsoft Corporation.
Power BI Documentation: Data Modeling, DAX, and Visualization.
Available at:
<https://learn.microsoft.com/power-bi>

LinkedIn: https://www.linkedin.com/posts/anup-pandey-powerbi-dataanalytics-datapreprocessing-activity-7405827960710107136-G7TO?utm_source=share&utm_medium=member_desktop&rcm=ACoAAEe82pUBIjI7B29DRyyNYg2YtUvvXHSsh-0

Anup Pandey + You
MERN Stack Developer @ Gokboru Tech | Building Full-Stack Applications | DSA + ...
[View my portfolio](#)
4d •

NFHS-5 India Health Analytics Dashboard | Power BI Project IN
Built something I've been wanting to do for a while — an end-to-end Power BI analytics project powered by the National Family Health Survey (NFHS), one of India's richest and most complex public health datasets.
This project goes beyond visualization. It focuses heavily on data preprocessing, modeling, and analytical storytelling to turn complex health data into clear, meaningful insights.

- What the dashboard delivers:
 - National & State-level health overview
 - District-level deep dives & comparisons
 - Maternal & Child Health insights (ANC, PNC, Institutional births, C-sections)
 - Nutrition & child malnutrition trends (NFHS-1 → NFHS-5)
 - Non-Communicable Diseases (Anaemia)
 - Household amenities & digital access (Urban vs Rural gaps)
- Datasets used (Official Government Sources):
 - All-India & State/UT Factsheets (NFHS-5)
<https://lnkd.in/gwxRJi3h>
 - District-wise Factsheets (NFHS-5)
<https://lnkd.in/gVUbwWSB>
 - Year-wise Child Health Indicators (NFHS-1 to NFHS-5)
<https://lnkd.in/gxsHszgE>
- Why this project:
To demonstrate how strong preprocessing + clean data modeling + thoughtful visualization can transform public health data into actionable insights for policy, research, and decision-making.
Would love feedback, discussions, or ideas for further enhancement

#PowerBI #DataAnalytics #DataPreprocessing #DataModeling #DAX #PublicHealth #NFHS #IndiaData #BusinessIntelligence #DashboardDesign #AnalyticsProject



You and 133 others 100 comments • 2 reposts

Reactions

Love Comment Repost Send

1,716 impressions [View analytics](#)

Tell them what you loved...

Most relevant

Pawan Singh + 1st
Aspiring Software Developer | C++, Java, Python | DSA Enthusiast | SQL, Excel | N...
interesting project
Like - 1 | Reply | 1 reply

Anup Pandey Author
MERN Stack Developer @ Gokboru Tech | Building Full-Stack Applications | ...
Pawan Singh Thanks
Like | Reply | 127 impressions

Sanchi Chauhan + 1st
Attended Lovely Professional University
Good