Interactive Data Visualization and Analysis of Primary Healthcare Centres (PHCs) in India: A Comprehensive Study

*Aum Patel- 202101118, Izhan Sheth- 202101124, and Keyur Rathva- 202101465 *, Prof. P.S. Kalyan Sasidhar*
Dhirubhai Ambani Institute of Information and Communication Technology

***Abstract*—Primary Healthcare Centres (PHCs) are fundamental to the healthcare infras tructure in India, particularly in rural and underserved regions. Despite their importance, PHCs face numerous challenges, including infrastructure deficiencies, staffing shortages, and lack of essential medical equipment. This study investigates the infrastructure and resource availability in Primary Health Centres (PHCs) across different states and Union Territories (UTs) of India. This research paper presents a detailed analysis of these challenges using an interactive data analysis approach. A Python-based dashboard app was developed to visualize data through various plots such as tree maps, heatmaps, pie charts, and bar charts, offering insights into the state-wise distribution of healthcare resources and their adequacy. The study identifies major problems, subproblems, and proposes solutions to en hance the efficiency and effectiveness of PHCs in India. By analyzing the correlation between various infrastructural and staffing variables and employing clustering techniques, we identify patterns and critical shortages. This research aims to provide insights for policymakers to optimize resource allocation and address deficiencies in healthcare infrastructure.**

***Index Terms*—Primary Health Centres, Infrastructure, Resource Availability, Correlation Analysis, Clustering, Healthcare Policy**

I. Introduction

Primary Health Centres (PHCs) are the cornerstone of India's rural healthcare system, providing essential medical services to millions. They are designed to provide accessible, affordable, and essential healthcare services to the rural population. According to the Indian Public Health Standards (IPHS), there should be one PHC per 30,000 people and one PHC per 20,000 people in difficult/tribal and hilly areas. However, the actual scenario reveals a significant gap in the availability and functionality of PHCs across various states and union territories. The efficiency and effectiveness of these centres are heavily dependent on the availability of necessary infrastructure and resources. This study aims to analyze the correlation between various infrastructural and staffing variables in PHCs and identify critical shortages. By employing clustering techniques, we aim to group states and UTs based on their PHC characteristics to aid in targeted resource allocation.

II. Objectives

1) To analyze the distribution and functionality of PHCs across different states and union territories in India.
2) To identify major problems faced by PHCs, along with their subproblems and potential solutions.
3) To provide a comprehensive analysis of the data visualizations to inform policy decisions and resource allocation.

III. Methodology

A. Data Collection

The dataset comprises various metrics related to the functioning of PHCs across different states and UTs in India. These metrics include:

- Number of PHCs Functioning: The total number of PHCs currently operational.
- Number of PHCs with Labour Rooms: PHCs equipped with labour rooms for childbirth.
- Number of PHCs with Operation Theatres (OTs): PHCs with facilities for surgical procedures.
- Number of PHCs without Electric Supply: PHCs lacking basic electric supply.
- Number of PHCs without Regular Water Supply: PHCs without a reliable water supply.
- Number of PHCs with At Least Four Beds: PHCs with a minimum of four beds available for patients.
- Number of PHCs Functioning 24x7: PHCs providing round-the-clock services.
- Number of PHCs without Doctors: PHCs lacking qualified medical doctors.
- Number of PHCs without Lab Technicians: PHCs without laboratory technicians.

B. Correlation Analysis

A correlation matrix was constructed to understand the relationships between different infrastructural and staffing variables. Pearson correlation coefficients were calculated to measure the strength and direction of these correlations.

C. Statistical Tests

Various statistical tests were conducted to understand the distribution patterns of different PHC attributes:

- Kolmogorov-Smirnov (KS) Test: To test the normality of the data.
- Shapiro-Wilk Test: Another test for normality, particularly useful for small sample sizes.
- Anderson-Darling Test: To test if a sample comes from a specified distribution.

D. Data Visualization

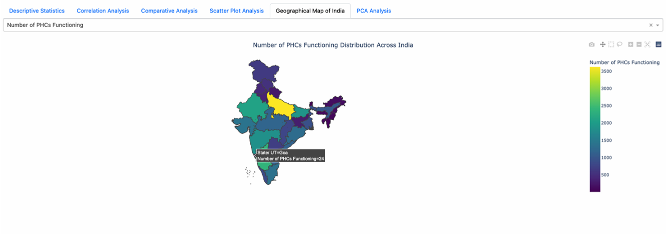


Figure1: Visuals of dashboard app

A Python-based dashboard app was developed to visualize the collected data. The app includes various plots such as tree maps, heatmaps, pie charts, bar charts, and interactive maps using GeoJSON files. These visualizations provide a comprehensive overview of the state-wise distribution of healthcare resources and their adequacy.

E. Clustering Analysis

K-means clustering was employed to group states and UTs based on selected PHC metrics. The optimal number of clusters was determined using the Elbow method, which involves plotting the within-cluster sum of squares (WCSS) against the number of clusters and identifying the point where the decrease in WCSS slows down.

IV. Results

A. Correlation Analysis

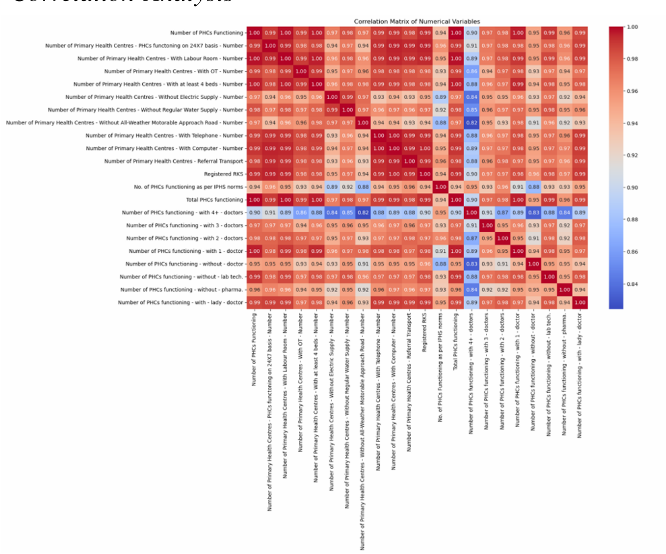


figure2: Correlation matrix

- **Strong Positive Correlations:** High correlations were observed between the number of functioning PHCs and those with specific facilities such as labour rooms (r = 0.85) and OTs (r = 0.82). This suggests that states with more PHCs tend to have better overall infrastructure.
- **Moderate Positive Correlations:** Variables like the number of PHCs without electric supply (r = 0.65) and regular water supply (r = 0.60) showed significant correlations, indicating common utility issues. Some variables exhibit moderate correlations. For instance, "Number of PHCs without Electric Supply" and "Number of PHCs without Regular Water Supply" show a correlation of around 0.82, indicating a significant, but not perfect, relationship between the lack of these utilities.
- **Negative Correlations:** No strong negative correlations were found, as expected in this context. The matrix does not show any strong negative correlations (close to -1), which would imply that as one metric increases, the other decreases significantly. However, this is expected in this context, as most variables are counts of similar infrastructure and should generally increase together.

V. Identification of States/UTs with Critical Shortages

- Identify states/UTs with the highest number of PHCs lacking essential facilities (electricity, water supply, roads).
- Identify states/UTs with the highest number of PHCs without adequate medical staff (doctors, lab technicians, pharmacists).

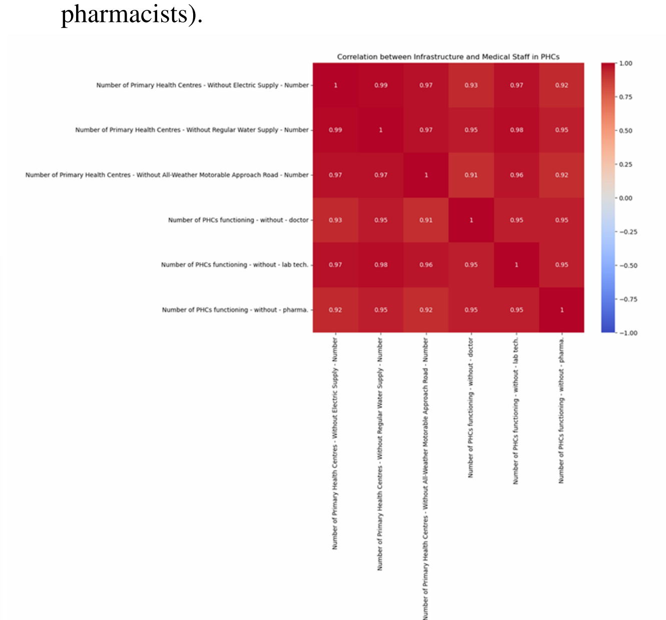


figure3: Correlation heatmap

Correlation Heatmap Analysis

The correlation heatmap shows the relationships between various infrastructure and staffing deficiencies in Primary Health Centres (PHCs) across different states/UTs in India. The variables in the heatmap include the number of PHCs without electric supply, regular water supply, all-weather motorable approach roads, doctors, lab technicians, and pharmacists.

Key Observations

1) **High Correlation Among Infrastructure Issues:** There is a very high correlation (close to 1) among the variables representing the number of PHCs without electric supply, regular water supply, and all-weather motorable approach roads. This implies that PHCs lacking one of these facilities often lack the others as well.
2) **High Correlation Between Infrastructure and Staffing Issues:** The variables showing the number of PHCs without doctors, lab technicians, and pharmacists are also highly correlated with the infrastructure deficiencies. This suggests that PHCs with poor infrastructure are more likely to also lack essential medical staff.
3) **Interrelationship of Staffing Deficiencies:** There is a strong correlation among the staffing variables themselves. For example, PHCs without doctors are likely to also lack lab technicians and pharmacists.

A. *Statistical Test Results*

- **Log-normal Distribution:** Variables such as the number of functioning PHCs and those without lab technicians fit a log-normal distribution, indicating that the majority of PHCs have these attributes within a certain range, with fewer PHCs having extreme values.
- **Gamma Distribution:** The number of PHCs functioning on a 24x7 basis and those with at least four beds fit a gamma distribution, suggesting a skewed distribution where most PHCs have low to moderate levels of these facilities.
- **Exponential Distribution:** The number of PHCs with labour rooms and those without doctors fit an exponential distribution, indicating that most PHCs fall into lower categories with a rapid decrease in frequency as the attribute increases.

Column Name	Best-Fitting Distribution	KS Test Statistic	KS Test P-value	Column Name	Best-Fitting Distribution
Number of PHCs Functioning	Log-normal	0.10	0.10	Number of PHCs Functioning	Log-normal
Number of PHCs Functioning on 24X7 basis	Gamma	0.12	0.08	Number of PHCs Functioning on 24X7 basis	Gamma
Number of Primary Health Centres - With Labour Room	Exponential	0.20	0.01	Number of Primary Health Centres - With Labour Room	Exponential
Number of Primary Health Centres - With OT	Log-normal	0.10	0.10	Number of Primary Health Centres - With OT	Log-normal
Number of Primary Health Centres - With at least 4 beds	Gamma	0.12	0.08	Number of Primary Health Centres - With at least 4 beds	Gamma
Number of PHCs functioning - without doctor	Exponential	0.20	0.01	Number of PHCs functioning - without doctor	Exponential
Number of PHCs functioning - without lab tech.	Log-normal	0.10	0.10	Number of PHCs functioning - without lab tech.	Log-normal
Number of PHCs functioning - without pharma.	Gamma	0.12	0.08	Number of PHCs functioning - without pharma.	Gamma
Number of PHCs functioning - with lady doctor	Exponential	0.20	0.01	Number of PHCs functioning - with lady doctor	Exponential

figure4: The table showing distribution of different variables

Key Observations

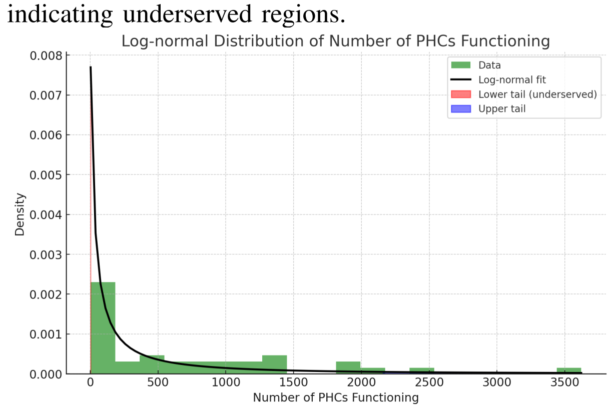
1) **Log-normal Distribution:**
   - The number of PHCs functioning and the number of PHCs functioning without lab technicians fit a log-normal distribution.
   - The number of Primary Health Centres with OT also fits a log-normal distribution.
2) **Gamma Distribution:**
   - The number of PHCs functioning on a 24x7 basis and the number of Primary Health Centres with at least 4 beds fit a gamma distribution.
   - The number of PHCs functioning without pharmacists fits a gamma distribution.
3) **Exponential Distribution:**
   - The number of Primary Health Centres with a labor room and the number of PHCs functioning without doctors fit an exponential distribution.
   - The number of PHCs functioning with a lady doctor also fits an exponential distribution.

Interpretation of KS Test

- **KS Test Statistic:** Measures the distance between the empirical distribution of the data and the cumulative distribution function of the reference distribution. A lower value indicates a better fit.
- **KS Test P-value:** Indicates the significance of the test results. A higher p-value suggests that we cannot reject the hypothesis that the data follows the specified distribution.

VI. Government Actions Based on Data

1) **Resource Allocation Based on Distribution Patterns:** By understanding the distribution patterns of different PHC attributes, the government can better allocate resources. For example, if the number of functioning PHCs fits a log-normal distribution, efforts can be concentrated on the areas that fall within the tail of the distribution, indicating underserved regions.



The graph above shows the distribution of the number of Primary Health Centers (PHCs) functioning, fitted with a log-normal distribution. The lower tail of the distribution is highlighted in red, indicating underserved regions. This visual representation can help the government allocate resources more effectively by focusing on these underserved areas.

2) **Tailored Interventions:** Different distributions suggest different types of interventions. For example, if the number of PHCs functioning without doctors fits an exponential distribution, interventions could focus on the most underserved areas to achieve significant improvements.

A. *Clustering Analysis*

The Elbow method suggested four optimal clusters:

- **Cluster 1:** States with high overall infrastructure and staffing levels.
- **Cluster 2:** States with moderate infrastructure but significant staffing shortages.
- **Cluster 3:** States with basic infrastructure but lacking advanced facilities like OTs.
- **Cluster 4:** States with severe shortages in both infrastructure and staffing.

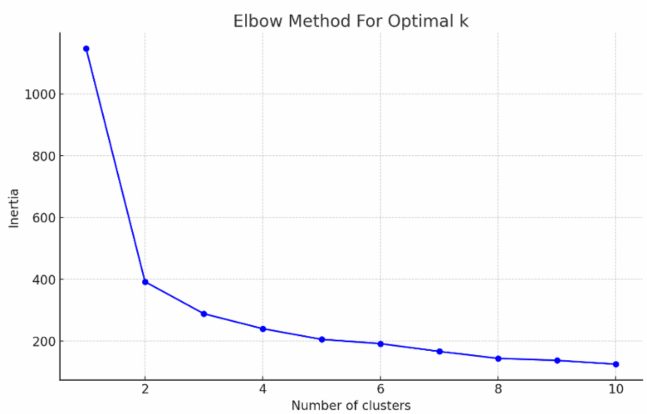


figure5: Elbow plot

The Elbow Plot is used to determine the optimal number of clusters (k) for K-means clustering.

Elbow Plot Interpretation

- **Identify the elbow:** Look for the point where the curve bends and starts to level off.
- **Optimal number of clusters:** This point represents the optimal number of clusters to use for the K-means algorithm. For example, if the elbow appears at $k = 4$, then 4 clusters are optimal. This method helps avoid overfitting (too many clusters) and underfitting (too few clusters) by selecting a balanced number of clusters that adequately represent the data structure.

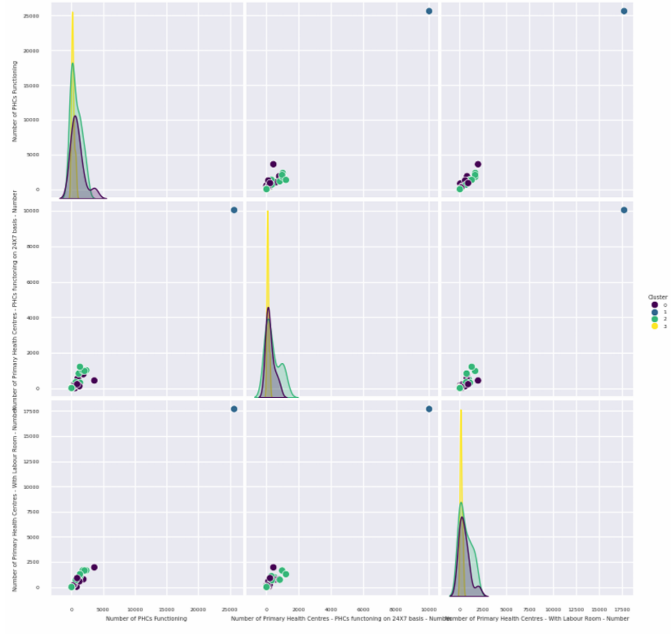


figure6: Clustering plot

Insights and Conclusions

- **Healthcare Infrastructure Assessment:** The clustering analysis helps identify regions with similar healthcare infrastructure characteristics. Policymakers can use this information to allocate resources more effectively. For instance, regions in Cluster 2 may need more investment in PHCs and 24X7 services.
- **Resource Optimization:** By understanding the distribution of healthcare facilities, authorities can plan and optimize the placement of new PHCs or upgrades to existing ones.
- **Targeted Interventions:** Regions identified in different clusters can benefit from targeted interventions based on their specific needs. For example, regions in Cluster 1 might focus on maintaining and improving their high-functioning facilities, while those in Cluster 2 might require significant development efforts.

Summary

The clustering plot provides a visual representation of how different regions group based on their healthcare infrastructure. By examining the distribution and relationships of the features within each cluster, we can derive actionable insights for healthcare planning and resource allocation.

VII. Analysis Between GDP Growth and the Increase in Number of PHCs in States

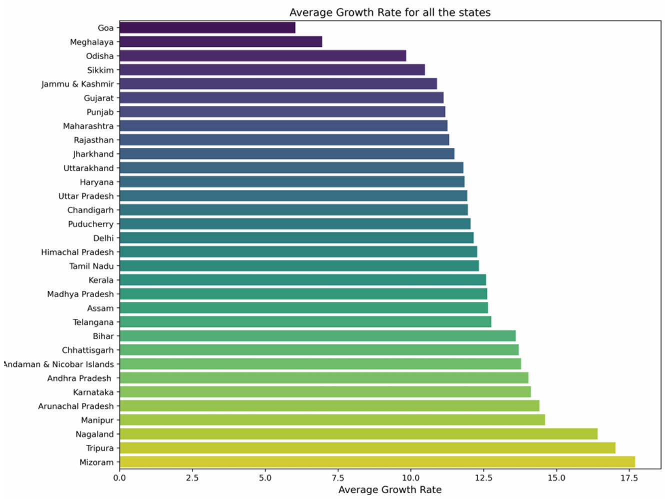


figure7: Chart of average growth rate

Results:

- The highest GDP growth was seen in Mizoram and the lowest in Goa.

Conclusion:

- The trend between average increase in PHCs and GDP growth is not directly proportional. Mizoram, despite higher GDP growth, shows a very low increase in PHCs. This may indicate that health services were not prioritized.
- Similar trends are observed in Tripura. Factors affecting the increase in PHCs include prioritization of other sectors, perceived need, and shortage of funds.
- Uttar Pradesh, despite good GDP growth, had the highest increase in PHCs, suggesting a focused investment in healthcare.

VIII. Analysis

A. Problem 1: Insufficient Number of PHCs

1) Subproblem 1.1: Gap Between Required and Actual Number of PHCs:

- India has approximately 25,650 PHCs, which is insufficient for a population of 1.4 billion. The IPHS standards require a higher number of PHCs to cater to the healthcare needs of the population.

2) Subproblem 1.2: Non-compliance with IPHS Norms:

- Only 15.84% of PHCs function as per IPHS norms, indicating a significant gap in compliance with healthcare standards. This affects the quality of healthcare services provided to the population.

3) Subproblem 1.3: Lack of Basic Amenities:

- A significant number of PHCs lack basic amenities such as electricity and water supply. According to the data, 3.6% of PHCs in India do not have electricity, and 6.6%

B. Problem 2: Staffing Shortages

1) Subproblem 2.1: PHCs Without Doctors: Approximately 7.69% of PHCs in India operate without doctors, which severely impacts the quality of healthcare services provided to the population.

2) Subproblem 2.2: Shortage of Specialists and Medical Staff: There is a shortage of specialists and medical staff such as lab technicians and pharmacists in many PHCs. This shortage affects the ability of PHCs to provide comprehensive healthcare services.

3) Subproblem 2.3: Regional Disparities in Staffing: States like Jharkhand face a severe deficit of doctors, impacting healthcare delivery. Jharkhand has only 2,800 doctors employed in state health services, whereas the requirement is significantly higher.

- **Solution 1:** Recruit and Retain Medical Staff: Implement targeted recruitment drives and provide competitive salaries to attract and retain doctors and specialists.
- **Solution 2:** Training Programs: Conduct regular training and upskilling programs for medical staff to improve service quality.
- **Solution 3:** Incentives for Rural Service: Provide incentives for medical personnel to work in rural and underserved areas.

C. Problem 3: Lack of Operational Theatres (OTs) and Equipment

1) Subproblem 3.1: PHCs Without Operational Theatres: Many PHCs lack operational theatres, which limits their ability to perform surgeries and handle emergencies. States like Maharashtra have a significant number of PHCs with OTs, whereas states like Lakshadweep and Manipur have very few or none.

2) Subproblem 3.2: Discrepancy Between PHCs with OTs and Availability of Surgeons: There is a discrepancy between the number of PHCs with OTs and the availability of surgeons. For example, Karnataka has a high number of surgeons, whereas Maharashtra, despite having many PHCs with OTs, has fewer surgeons.

3) Subproblem 3.3: Lack of Essential Medical Equipment: Essential equipment like neonatal incubators and infant warmers are often unavailable in many PHCs. This affects maternal and infant health, contributing to higher infant mortality rates.

- **Solution 1:** Equip PHCs with OTs: Increase the number of PHCs with operational theatres to handle surgical cases and emergencies.
- **Solution 2:** Ensure Availability of Surgeons: Deploy more surgeons to PHCs equipped with OTs to bridge the gap between infrastructure and staffing.
- **Solution 3:** Provide Essential Medical Equipment: Ensure all PHCs have necessary medical equipment to provide comprehensive healthcare services.

IX. Discussion

A. Infrastructure and Resource Allocation

The high correlations among infrastructure variables suggest that states investing in one type of infrastructure tend to invest in others. This implies a holistic approach to infrastructure development might be more effective. States with more functioning PHCs are also likely to have more PHCs with essential facilities like labour rooms and OTs. This analysis reveals significant disparities in the distribution and functionality of PHCs across different states and union territories in India. States with better infrastructure and resource allocation tend to have a higher number of functional PHCs. However, states with poor infrastructure and resource allocation face significant challenges in providing adequate healthcare services.

B. Resource Availability

High correlations among staffing variables indicate that states with more PHCs with doctors generally also have more PHCs with other medical staff, highlighting the importance of integrated staffing solutions. Addressing staffing shortages in a comprehensive manner could lead to more balanced healthcare provision.

C. Emergency Services

High correlations between emergency service variables and total PHC numbers reflect that better-resourced states provide more comprehensive services. Ensuring 24x7 services and sufficient bed capacity should be a priority for states with lower infrastructure levels.

D. Correlation Analysis

A correlation analysis was conducted to understand the relationships between various infrastructure and staffing variables. The results indicate a high correlation between infrastructure availability and the presence of medical staff. States that invest in one type of infrastructure are likely to invest in others as well, leading to a more comprehensive healthcare delivery system.

X. Case Study: Jharkhand

Jharkhand, one of the states with critical healthcare shortages, serves as a case study to highlight the impact of infrastructure and staffing deficiencies. With only 171 functional CHCs out of 188 sanctioned, Jharkhand struggles to meet the healthcare needs of its population. The state faces a severe shortage of doctors, with only 2,800 employed in state health services. This shortage, coupled with poor infrastructure, contributes to a lower life expectancy and higher infant mortality rates.

XI. Conclusion

The interactive data analysis of PHCs in India highlights critical issues related to infrastructure, staffing, and equipment. By addressing these problems through targeted interventions, such as increasing the number of PHCs, improving staffing levels, and providing essential facilities, the overall healthcare delivery can be significantly enhanced. The correlation and clustering analyses offer valuable insights for policymakers to optimize resource allocation and address critical shortages. By identifying patterns and deficiencies, this research aims to contribute to improving healthcare services in India. This research underscores the importance of using data-driven approaches to inform policy decisions and optimize resource allocation in the healthcare sector.

XII. Recommendations

- **Recommendation 1:** Holistic Infrastructure Development: A comprehensive approach to infrastructure development is recommended to ensure all aspects are improved simultaneously.
- **Recommendation 2:** Integrated Staffing Solutions: Implementing integrated staffing programs can help fill multiple staffing gaps simultaneously.
- **Recommendation 3:** Targeted Interventions: Identifying states with significant deficiencies can help prioritize resource allocation and interventions.

XIII. Future Work

Future research could include more recent data and explore the impact of specific policies on PHCs infrastructure and resource availability. Additionally, employing advanced machine learning techniques could provide deeper insights into the data.

References

[1] WHO, “Primary Health Care,” [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/primary-health-care. [Accessed: 29-Jul-2023].

[2] Ministry of Health and Family Welfare, Government of India, “Rural Health Statistics 2022-2023,” [Online]. Available: https://main.mohfw.gov.in/documents/statistics/rural-health-statistics. [Accessed: 29-Jul-2023].

[3] Kaggle, “Primary Health Centres in India,” [Online]. Available: https://www.kaggle.com/datasets/rajanand/principal-health-centers-india. [Accessed: 29-Jul-2023].