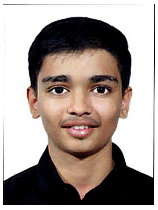
Exploratory Data Analysis on

National Rural Health Mission

by

Group 2



Kalp Shah
ID: 202103003
Course: B.Tech(MnC)

Shreyans Jain
ID: 202103009
Course: B.Tech(MnC)

Yash Garg
ID: 202101006
Course: B.Tech(ICT)

Course Code: IT 462
Semester: Winter 2023

Under the guidance of

Dr. Gopinath Panda

Dhirubhai Ambani Institute of Information and Communication Technology

April 29, 2024

Acknowledgment

I am writing this letter to express my heartfelt gratitude for your guidance and support throughout my project titled "National Rural Health Mission" Your invaluable assistance has played a pivotal role in shaping the successful completion of this endeavour.

I am incredibly fortunate to have had the opportunity to work under your mentorship. Your expertise, encouragement, and willingness to share your knowledge have been instrumental in elevating the quality and scope of my project. Your constructive feedback and insightful suggestions have helped me overcome challenges and develop a deeper understanding of the subject matter.

Furthermore, I would like to thank the entire team at Dhirubhai Ambani Institute of Information and Communication Technology for fostering an environment of collaboration and innovation. The resources and facilities provided have been crucial in conducting comprehensive research and analysis.

I would also like to express my gratitude to my peers and colleagues who have been supportive throughout this journey. Their valuable input and camaraderie have been a constant source of motivation.

Completing this project has been a tremendous learning experience. I am confident that the knowledge and skills acquired during this endeavour will be a solid foundation for my future endeavours.

Sincerely,

Members of Group 2 (202103003, 202103009, 202101006)

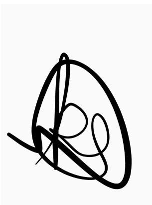
DECLARATION

We, [Kalp Shah(202103003), Shreyans Jain(202103009), Yash Garg(202101006)] now declare that the EDA project work presented in this report is our original work and has not been submitted for any other academic degree. All the sources cited in this report have been appropriately referenced.

We acknowledge that the data utilized in this project has been sourced from Data.gov.in. We affirm that we have complied with the terms and conditions specified on the website for accessing and using the dataset. We hereby confirm that the dataset employed in this project is accurate and authentic to the best of our knowledge.

We acknowledge that we have received no external help or assistance in conducting this project except for the guidance provided by our mentor, Prof. Gopinath Panda. We declare no conflict of interest in conducting this EDA project.

We have now signed the declaration statement and confirmed the submission of this report on 29th April 2024.



Kalp Shah
ID: 202103003
Course: B.Tech (MnC)

Shreyans Jain
ID: 202103009
Course: B.Tech (MnC)

Yash Garg
ID: 202101006
Course: B.Tech (ICT)

CERTIFICATE

This is to certify that Group 2 comprising Kalp Shah, Shreyans Jain and Yash Garg has completed an exploratory data analysis (EDA) project on the National Rural Health Mission, which was obtained from Data.gov.in.

The EDA project presented by Group 2 is their original work. It was completed under the guidance of the course instructor, Prof. Gopinath Panda, who provided support and guidance throughout the project. The project is based on a thorough analysis of the NRHM datasets, and the results presented in the report are based on the data obtained from the datasets.

This certificate is issued to recognize the successful completion of the EDA project on the National Rural Health Mission, which demonstrates the analytical skills and knowledge of the students of Group 2 in the field of data analysis.

Signed,
Dr. Gopinath Panda,
IT 462 Course Instructor
Dhirubhai Ambani Institute of Information and Communication Technology
Gandhinagar, Gujarat, INDIA.

April 29, 2024

Contents

Chapter 1. Introduction

1.1 Project Idea

The National Rural Health Mission (NRHM) in India has been a crucial initiative aimed at improving healthcare infrastructure and services in rural areas. In our project, we aim to conduct an exploratory data analysis (EDA) focusing on the budget allocation and infrastructure facilities under the NRHM. By analyzing relevant datasets, we seek to gain insights into the effectiveness and implementation of healthcare policies at the grassroots level.

1.2 Data Collection

For our analysis, we have collected two primary datasets: budget allocation data and information on infrastructure facilities across different rural healthcare centers. These datasets provide valuable insights into the financial resources allocated to the NRHM and the status of healthcare infrastructure in rural areas.

1.3 Dataset Description

1.3.1 Budget Allocation Data

The budget allocation dataset encompasses various aspects such as approved budgets, proposed allocations, expenditure patterns, and fund utilization under the NRHM. It offers a comprehensive overview of the financial resources allocated to different healthcare programs and initiatives aimed at improving rural healthcare services.

```python
# Displaying top 5 rows
df1.head()
```

	State_UT	State_UT_Code	Fiscal Year	Budget Proposed by the States/UTs	Budget Approved for the States/UTs	Opening Balance with the States/UTs	Release of Government of India's Fund	Total Expenditure Reported (Including States' Share)	Extent of Budget Approved Against Budget Proposed	Extent of Funds Utilised Against Budget Approved	...	Unnamed: 14	Unnamed: 15	Unnamed: 16	Unnamed: 17	Unr
0	Andhra Pradesh	1.0	2015-16	2380.02	1336.13	NaN	659.04	1105.70	56.14	82.75	...	NaN	NaN	NaN	NaN	
1	Arunachal Pradesh	2.0	2015-16	346.46	195.94	NaN	163.80	147.41	56.55	75.23	...	NaN	NaN	NaN	NaN	
2	Assam	3.0	2015-16	2275.32	1853.40	NaN	997.59	1212.25	81.46	65.41	...	NaN	NaN	NaN	NaN	
3	Bihar	4.0	2015-16	3874.98	2672.45	NaN	1269.67	1731.85	68.97	64.80	...	NaN	NaN	NaN	NaN	
4	Chhattisgarh	5.0	2015-16	1528.15	1215.09	NaN	423.31	769.33	79.51	63.31	...	NaN	NaN	NaN	NaN	

5 rows × 24 columns

Figure 1.1: First 5 rows of the Budget Allocation Data

1.3.2 Infrastructure Facilities Data

The infrastructure facilities dataset provides detailed information on the availability and quality of healthcare infrastructure across different rural healthcare facilities. It includes data on the presence of essential amenities such as medical equipment, staffing levels, availability of medicines, and the overall infrastructure condition.

	S.No.	State/UT	Number of PHCs Functioning	PHCs functoning on 24X7 basis	With Labour Room	With OT	With at least 4 beds	Without Electric Supply	Without Regular Water Supply	With Telephone	Fiscal Year
0	1	Andhra Pradesh	1142	1142	1134	1129	1142	0	0	0	2021-22
1	2	Arunachal Pradesh	126	42	67	1	46	11	18	18	2021-22
2	3	Assam	920	280	721	64	339	17	0	0	2021-22
3	4	Bihar	1492	630	465	254	612	326	374	374	2021-22
4	5	Chhattisgarh	770	477	752	63	610	17	8	8	2021-22

Figure 1.2: First 5 rows of the Infrastructure Facilities Data

1.4 Packages Required

In this project, we utilized several Python libraries to conduct our data analysis. The following packages were essential for cleaning, analyzing, and visualizing the datasets:

- **Pandas**: Pandas is a powerful library for data manipulation and analysis. We used it extensively to load, clean, and preprocess our datasets.
- **NumPy**: NumPy is a fundamental package for scientific computing in Python. It provided support for numerical operations and array manipulation, which were crucial for our data cleaning process.
- **Matplotlib**: Matplotlib is a widely-used library for creating static, interactive, and animated visualizations in Python. We utilized Matplotlib to generate various plots and charts to visualize the distribution and trends in our data.

- **Seaborn**: Seaborn is a statistical data visualization library based on Matplotlib. It offers a higher-level interface for drawing attractive and informative statistical graphics. We leveraged Seaborn to create more advanced visualizations and explore relationships between different variables in our datasets.

- **Scikit-learn (sklearn):** A comprehensive library for machine learning in Python. It provides various utilities for feature extraction, transformation, and preprocessing, such as scaling, encoding categorical variables, and generating polynomial features.

These packages provided the necessary tools and functionalities to effectively clean and analyze the datasets for our project.

Chapter 2. Data Cleaning

Before proceeding with the exploratory data analysis, it was imperative to preprocess and clean the datasets to ensure accuracy and consistency in our analysis. The data cleaning process involved several steps, including:

- **Handling Missing Values**: We identified and handled missing values in the datasets using appropriate techniques such as imputation or removal, depending on the nature of the missing data and its impact on our analysis.
- **Removing Duplicates**: We checked for and removed any duplicate entries in the datasets to avoid redundancy and ensure data integrity.
- **Standardizing Data Formats**: We standardized the formats of data fields to maintain consistency and facilitate easier analysis across different variables.
- **Correcting Data Errors**: We addressed any inconsistencies or errors in the data, such as typos or incorrect entries, to ensure the reliability of our analysis results.

By meticulously cleaning and preprocessing the datasets, we prepared them for in-depth analysis, enabling us to extract meaningful insights and draw accurate conclusions.

2.1 Missing data analysis

Missing data analysis is a fundamental step that involves identifying, understanding, and handling missing values in a dataset. It plays a crucial role in ensuring the quality and integrity of the data used for analysis and modeling.

2.1.1 Dataset-1: Financial Resources Allocated to the NRHM

First, we drop all the columns from the dataframe (in this case, 'df1') which have all values as 'NULL (NaN)'. Then, we do similar operation for the rows too. Given below is the code to implement it.

```
# Dropping the columns with all NULL Values
df1 = df1.dropna(axis=1, how='all')

# Dropping the rows with all NULL Values
df1 = df1.dropna(axis=0, how='all')
```

Figure 2.1: Dropping columns and rows with NaN values

The updated dataframe and number of missing values are shown below:

```
# Displaying top 5 rows
df1.head()
```

	State_UT	State_UT_Code	Fiscal Year	Budget Proposed by the States/UTs	Budget Approved for the States/UTs	Opening Balance with the States/UTs	Release of Government of India's Fund	Total Expenditure Reported (Including States' Share)	Extent of Budget Approved Against Budget Proposed	Extent of Funds Utilised Against Budget Approved	Extent of Funds Utilised Against Budget Proposed
0	Andhra Pradesh	1.0	2015-16	2380.02	1336.13	NaN	659.04	1105.70	56.14	82.75	46.46
1	Arunachal Pradesh	2.0	2015-16	346.46	195.94	NaN	163.80	147.41	56.55	75.23	42.55
2	Assam	3.0	2015-16	2275.32	1853.40	NaN	997.59	1212.25	81.46	65.41	53.28
3	Bihar	4.0	2015-16	3874.98	2672.45	NaN	1269.67	1731.85	68.97	64.80	44.69
4	Chhattisgarh	5.0	2015-16	1528.15	1215.09	NaN	423.31	769.33	79.51	63.31	50.34

Figure 2.2: Displaying first 5 rows of the dataset

```
df1.isnull().sum()
```

```
State_UT                                                0
State_UT_Code                                           0
Fiscal Year                                             0
Budget Proposed by the States/UTs                      14
Budget Approved for the States/UTs                     15
Opening Balance with the States/UTs                   116
Release of Government of India's Fund                   7
Total Expenditure Reported (Including States' Share)    7
Extent of Budget Approved Against Budget Proposed      17
Extent of Funds Utilised Against Budget Approved       17
Extent of Funds Utilised Against Budget Proposed       16
dtype: int64
```

Figure 2.3: Displaying number of NULL values in respective columns

Replacing rest of the 'NA' values with 'NaN', and visualizing missing values with the help of heatmap.

```
# Replace 'NA' with NaN
df1.replace('NA', np.nan, inplace=True)

# Create a heatmap to visualize null values
plt.figure(figsize=(10, 6))
sns.heatmap(df1.isnull(), cmap='viridis', cbar=False)
plt.title('Null Values in Dataset')
plt.show()
```

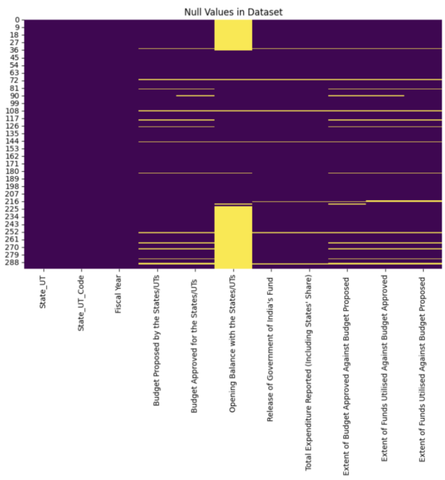


Figure 2.4:

In the heatmap, Yellow (or light color) cells represent missing/NULL values. Dark cells represent non-null values. This visualization helps in understanding the completeness of your dataset and deciding how to handle missing values during data preprocessing.

We filter and drop rows from the DataFrame where the 'State_UT' is 'Ladakh' and the 'Fiscal Year' is before 2020-21, and then reset the index of the DataFrame.

```python
# Dropping NULL values
# Filter rows where State_UT is 'Ladakh' and Fiscal Year is before 2020
mask = (df1['State_UT'] == 'Ladakh') & (df1['Fiscal Year'] < '2020-21')

# Drop rows that match the condition
df1.drop(df1[mask].index, inplace=True)

# Reset index after dropping rows
df1.reset_index(drop=True, inplace=True)


```

```python
df1[df1['State_UT'] == 'Ladakh']
```

	State_UT	State_UT_Code	Fiscal Year	Budget Proposed by the States/UTs	Budget Approved for the States/UTs	Opening Balance with the States/UTs	Release of Government of India's Fund	Total Expenditure Reported (Including States' Share)	Extent of Budget Approved Against Budget Proposed	Extent of Funds Utilised Against Budget Approved	Extent of Funds Utilised Against Budget Proposed
214	Ladakh	35.0	2020-21	239.870	131.8900	NaN	91.89	38.56	NaN	29.24	16.08
251	Ladakh	35.0	2021-22	142.840	134.3245	NaN	44.79	62.81	94.04	46.76	43.97
288	Ladakh	35.0	2022-23	217.112	196.5880	NaN	3.67	NaN	90.55	0.00	0.00

Figure 2.5: Displaying number of NULL values in respective columns

2.1.2 Dataset-2: Status of Healthcare Infrastructure in Rural Areas

As we can see, in Chandigarh, there is no data available for any year so we drop all the rows having State/UT 'Chandigarh'.

```python
df2=df2[df2['State/UT'] != 'Chandigarh']
```

Figure 2.6:

Checking for null values, we get:

```python
df2.isnull().sum()
```

```
S.No.                              0
State/UT                           0
Number of PHCs Functioning         0
PHCs functoning on 24X7 basis      0
With Labour Room                   0
With OT                            0
With at least 4 beds               0
Without Electric Supply            0
Without Regular Water Supply       0
With Telephone                     0
Fiscal Year                        0
With Electric Supply               0
With Regular Water Supply          0
dtype: int64
```

Figure 2.7: Displaying number of NULL values in respective columns

2.2 Imputation

In this section, we will try to clean the dataset as much as possible through the process of Imputation. It involves replacing missing values with estimated values based on the remaining data in the dataset.

2.2.1 Dataset-1: Financial Resources Allocated to the NRHM

We replace all the instances, where value is 0, in some particular columns to 'NaN'.

```python
# Data cleaning

# Replace 0 with NaN in specified columns
columns_to_replace = ['Budget Proposed by the States/UTs', 'Budget Approved for the States/UTs  ',"Release of Government of India's Fund ",
                      "Total Expenditure Reported (Including States' Share)"]
df1[columns_to_replace] = df1[columns_to_replace].replace(0, np.nan)
```

Figure 2.8: Displaying number of NULL values in respective columns

We now replace null values in the "Budget Proposed" column with the mean budget proposed for the same state over all years.

```python
# Replace null values in the "Budget Proposed" column with the mean budget proposed for the same state over all years
# Calculate mean budget proposed for each state
state_mean_budget = df1.groupby('State_UT')['Budget Proposed by the States/UTs'].mean()

# Function to fill missing values with state-wise mean
def fill_missing_budget(row):
    state = row['State_UT']
    budget_proposed = row['Budget Proposed by the States/UTs']
    if pd.isnull(budget_proposed):
        return state_mean_budget[state]
    else:
        return budget_proposed

# Apply the function to fill missing values in 'Budget_Proposed' column
df1['Budget Proposed by the States/UTs'] = df1.apply(fill_missing_budget, axis=1)
```

Figure 2.9:

Similarly, we impute null values with mean from all the columns in the same manner. Now, checking for null values.

```python
df1.isnull().sum()
```

```
State_UT                                                  0
State_UT_Code                                             0
Fiscal Year                                               0
Budget Proposed by the States/UTs                         0
Budget Approved for the States/UTs                        0
Opening Balance with the States/UTs                     111
Release of Government of India's Fund                     0
Total Expenditure Reported (Including States' Share)      0
Extent of Budget Approved Against Budget Proposed         0
Extent of Funds Utilised Against Budget Approved          0
Extent of Funds Utilised Against Budget Proposed          0
dtype: int64
```

Figure 2.10:

Since there are too many null values in the ”Opening Balance with the States/UTs” column, we drop that column.

```python
# Drop the column 'Opening Balance with the States/UTs'
column_to_drop = 'Opening Balance with the States/UTs'
df1.drop(column_to_drop, axis=1, inplace=True)
```

Figure 2.11:

We visualize the number of null values through heatmap.

```python
plt.figure(figsize=(10, 6))
sns.heatmap(df1.isnull(), cmap='viridis', cbar=False)
plt.title('Null Values in Dataset')
plt.show()
```

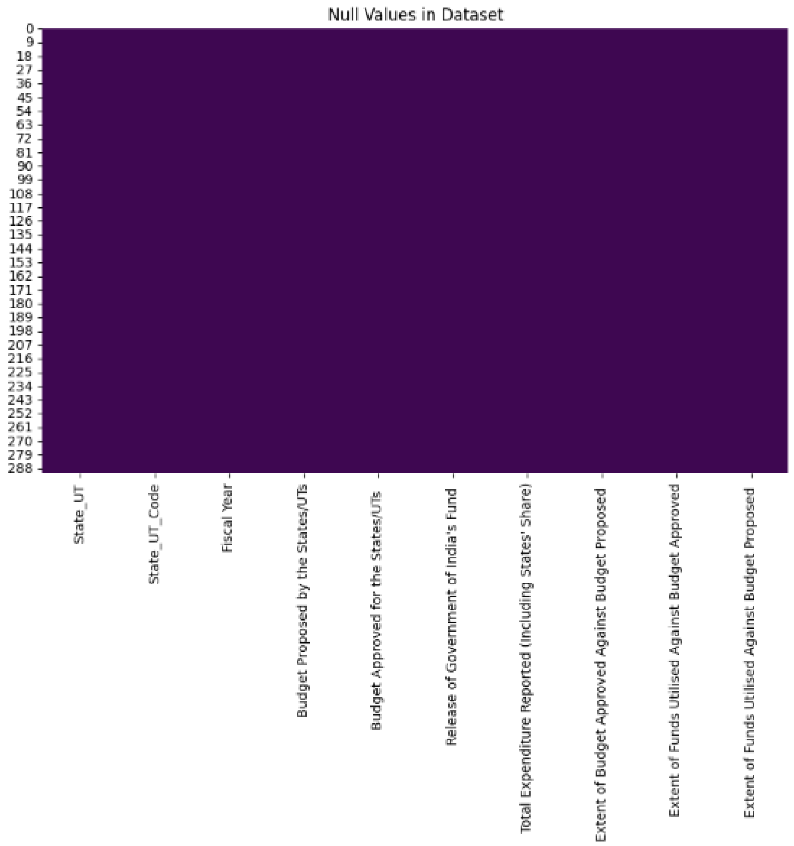


Figure 2.12:

Hence, we get our cleaned dataset.

2.2.2 Dataset-2: Status of Healthcare Infrastructure in Rural Areas

By analysing the dataset, we can see that, for some states there is a data missing for some particular year, which is filled with 0. To address the issue of missing or zero values in specific columns for certain states and fiscal years, we replace these zeros with NaN (NULL) values and impute these missing values with the mean of the corresponding data for the same state over fiscal years. This process helps to impute missing values based on the average behavior of the data within each state.

```python
# Columns of interest for zero replacement and imputation
columns_of_interest = ['Number of PHCs Functioning', 'PHCs functoning on 24X7 basis',
                       'With Labour Room', 'With OT', 'With at least 4 beds',
                       'With Electric Supply', 'With Regular Water Supply','With Telephone']

# Replace 0 values with NaN (NULL) for specified columns
df2[columns_of_interest] = df2[columns_of_interest].replace(0, np.nan)
```

Figure 2.13:

```
# Fill missing values with the mean of the same state over fiscal years for specified columns
df2[columns_of_interest] = df2.groupby('State/UT', group_keys=False)[columns_of_interest].apply(lambda group: group.fillna(group.mean()))
```

Figure 2.14:

We can see that our dataset is cleaned.

```
# visualize null values in our dataset
plt.figure(figsize=(10, 6))
sns.heatmap(df2.isnull(), cmap='viridis', cbar=False, yticklabels=False)
plt.title('Visualization of Null Values in Dataset')
plt.show()
```

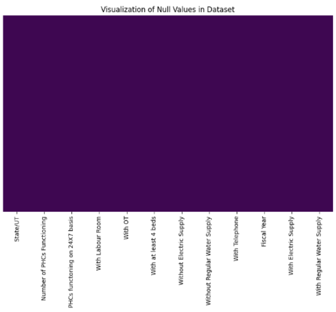


Figure 2.15:

We can observe that our dataset is cleaned.

Chapter 3. Visualization

The visualization chapter aims to present graphical representations of our analysis results to facilitate a better understanding of the data. We have used both the datasets to visualise the condition and contribution of NRHM in each states.

3.1 Univariate Analysis

In the univariate analysis section, we explore the distribution and characteristics of individual variables in the dataset. Through various graphical representations such as histograms, bar charts, and box plots, we aim to gain insights into the univariate behavior of key variables related to the National Rural Health Mission.

3.1.1 Budget Approved for States/UTs Over Financial Years

The line plot below visualizes the variation in the budget approved for different States/UTs over financial years. Each line represents a State/UT, and the y-axis represents the budget approved (in crore). The x-axis represents the fiscal year.

From the plot, we make the following observations:

1. **State/UT Budget Trends**: States/UTs like Karnataka, Kerala, Maharashtra, Tamil Nadu, and Telangana show consistent growth trends in budget allocation, indicating a sustained focus on healthcare infrastructure and services.

2. **Regional Disparities**: There are regional disparities in budget allocation, with some States/UTs consistently receiving higher budget approvals compared to others. Addressing these disparities could be crucial for achieving the objectives of the National Rural Health Mission.

3. **Consistently High Budget Allocation for Uttar Pradesh (UP)**: Uttar Pradesh consistently receives a significantly high budget allocation compared to other States/UTs across fiscal years. This observation suggests that UP receives considerable attention and resources under the National Rural Health Mission.

These observations provide valuable insights into the budgetary trends for States/UTs under the National Rural Health Mission, facilitating a better understanding of resource allocation patterns and potential areas for intervention and improvement.

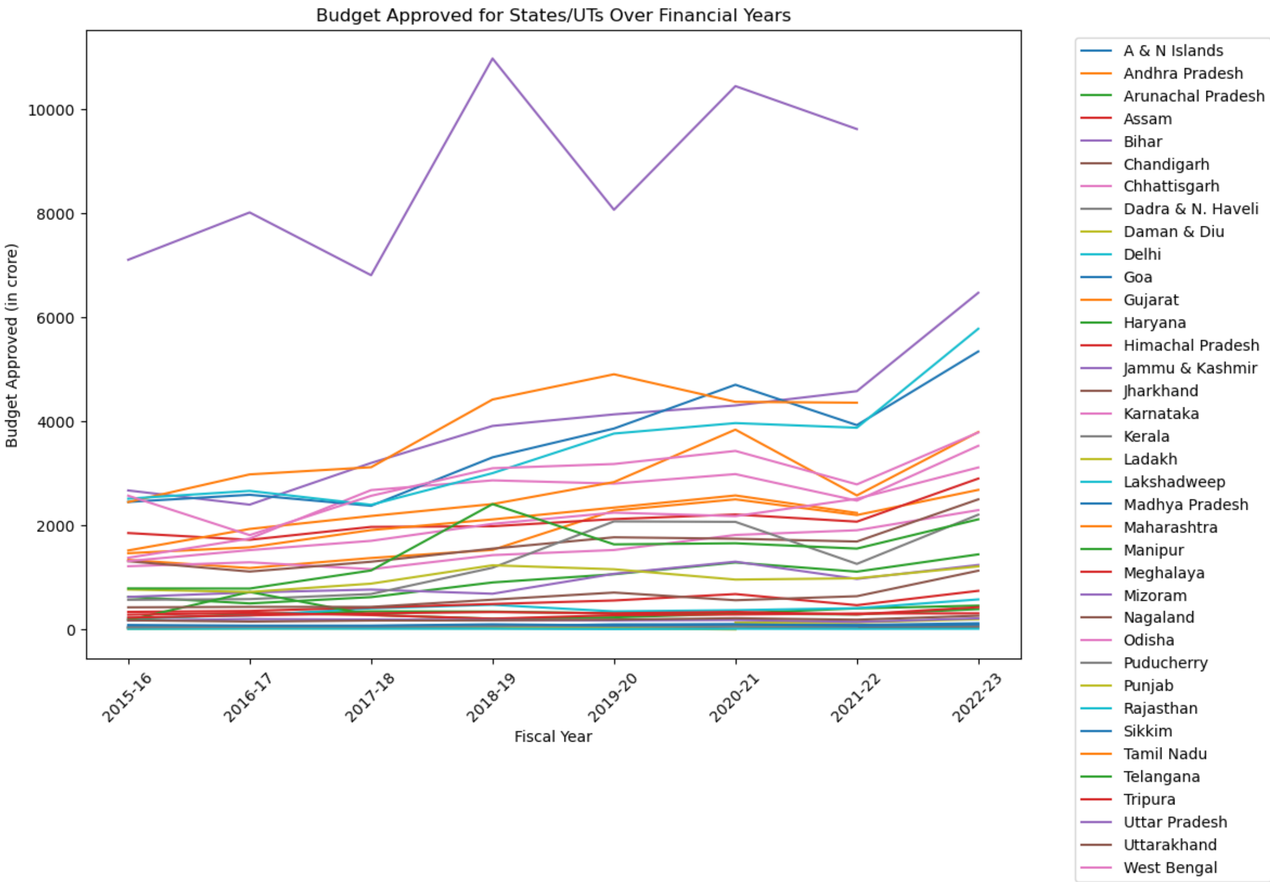


Figure 3.1: Budget Approved for States/UTs Over Financial Years

3.1.2 State/UT-wise Budget Allocation Analysis (2015-16 to 2022-23)

We conducted a detailed analysis of state/UT-wise budget allocation from 2015-16 to 2022-23. This analysis provides insights into budgetary trends over time and regional disparities in resource allocation.

Key Findings:

- **Trend Analysis:** Examined budget allocation trends could be seen on each State/UT differently. We Identified that there were a lot fluctuations for each State/UT.
- **Year-wise Comparison:**We compared budget allocations across fiscal years to understand policy impact.
- **Regional Disparities:** We also highlighted States/UTs with varying budget allocations.
- **Policy Implications:** Provides insights for policymakers to enhance healthcare infrastructure and services.

We also present graphs for Puducherry and Odisha as examples of state-specific budget allocation trends:

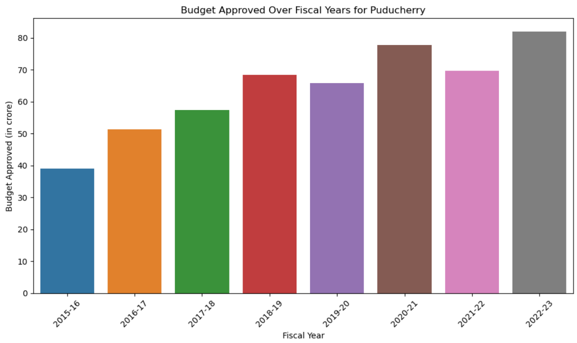


Figure 3.2: Budget Allocation Trend in Puducherry

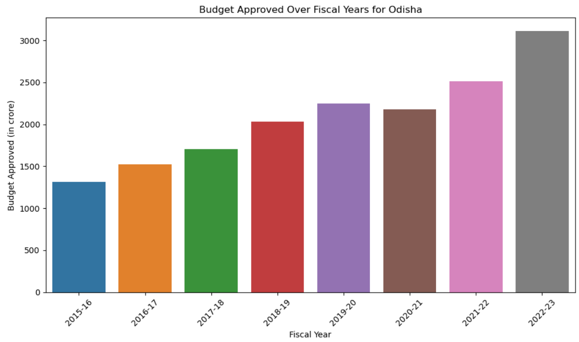


Figure 3.3: Budget Allocation Trend in Odisha

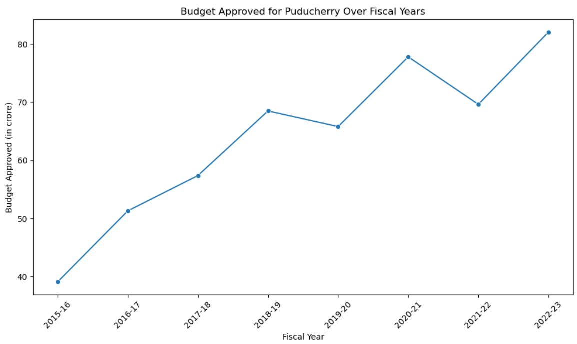


Figure 3.4: Line Graph for Budget Allocation Trend in Puducherry

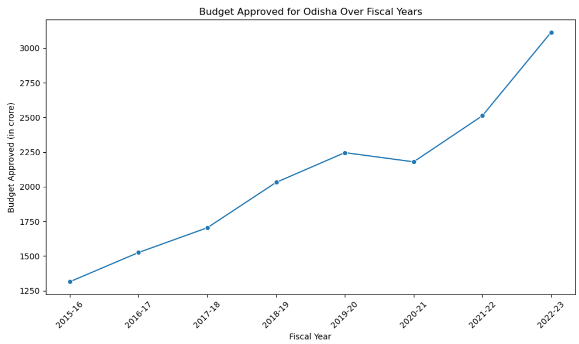


Figure 3.5: Line graph for Budget Allocation Trend in Odisha

3.1.3 Total Budget Approved per Financial Year

We analyzed the total budget approved per financial year to understand the overall trend in budget allocation over time.

Key Findings:

- **Financial Year-wise Analysis**: We examined the total budget approved for each financial year, spanning from 2015-16 to 2022-23.
- **Trend Identification**: The figure below illustrates the total budget approved per financial year. We observe a notable increase in budget allocation over the years, with a significant peak in the fiscal year 2020-21.
- **Insights**: This analysis provides valuable insights into the overall budgetary trends and priorities under the National Rural Health Mission. The observed increase in budget allocation reflects the government's commitment to enhancing healthcare infrastructure and services across the country.
- **Flatness**: We can observe that the budget approved is approximately flat since last two years showing the ignorance of nrhm from 2021-2023.

The figure below shows the total budget approved per financial year:

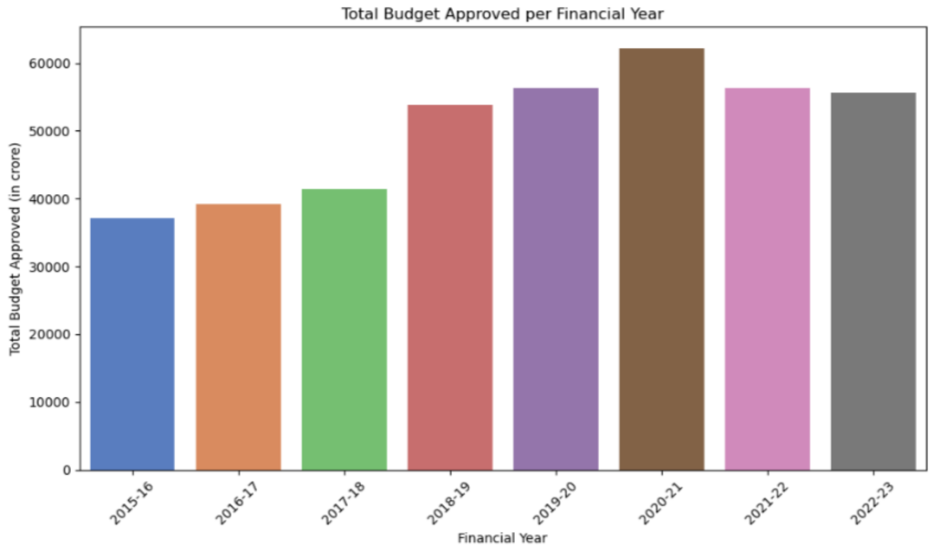


Figure 3.6: Total Budget Approved per Financial Year

3.1.4 Opening Balance Analysis for Delhi and Chhattisgarh

We analyzed the opening balance for States/UT over fiscal years to understand the financial trends and fluctuations. Taking some examples and drawing inferences.

The figures below illustrate the opening balance for Delhi and Chhattisgarh:

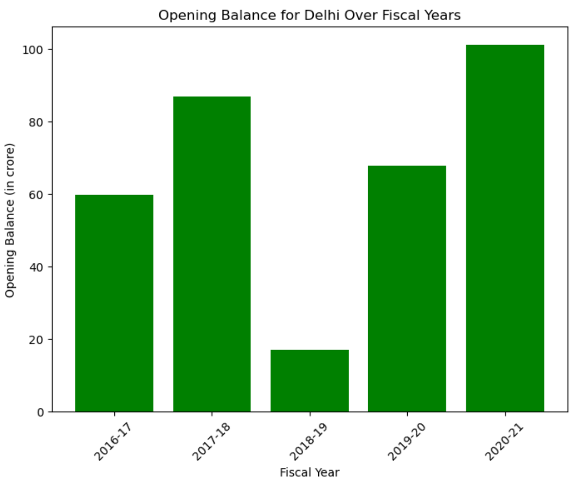


Figure 3.7: Opening Balance for Delhi Over Fiscal Years

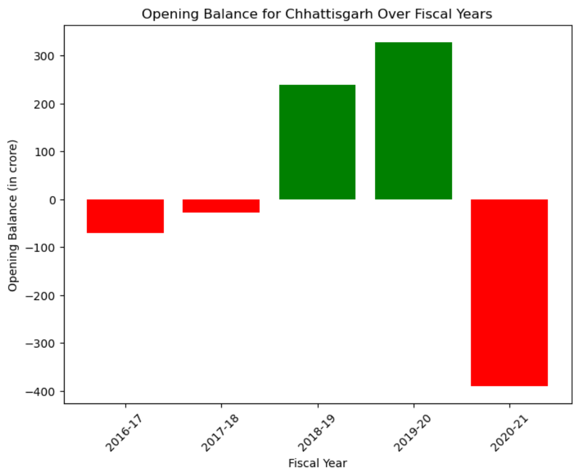


Figure 3.8: Opening Balance for Chhattisgarh Over Fiscal Years

Inferences:

- **Delhi:** The consistent positive opening balance suggests improved financial management and allocation strategies over time.

- **Chhattisgarh:** The fluctuating opening balance may indicate varying expenditure patterns or financial challenges faced by the state.

3.1.5 Average Extent of Budget Approved Against Budget Proposed by Fiscal Year

We calculated the average extent of budget approved against budget proposed across all states/UTs for each fiscal year to analyze the efficiency of budget allocation.

Key Findings:

- **Fiscal Year-wise Analysis:** The figure below illustrates the average extent of budget approved against budget proposed for each fiscal year from 2015-16 to 2022-23.
- **Approval Ratio:** The average approval ratio provides insights into the effectiveness of budget utilization, indicating the proportion of proposed budget that gets approved.
- **Trend Identification:** The analysis helps identify trends in budget approval rates over time and assess the government's response to healthcare funding requirements.

The figure below shows the average extent of budget approved against budget proposed by fiscal year:

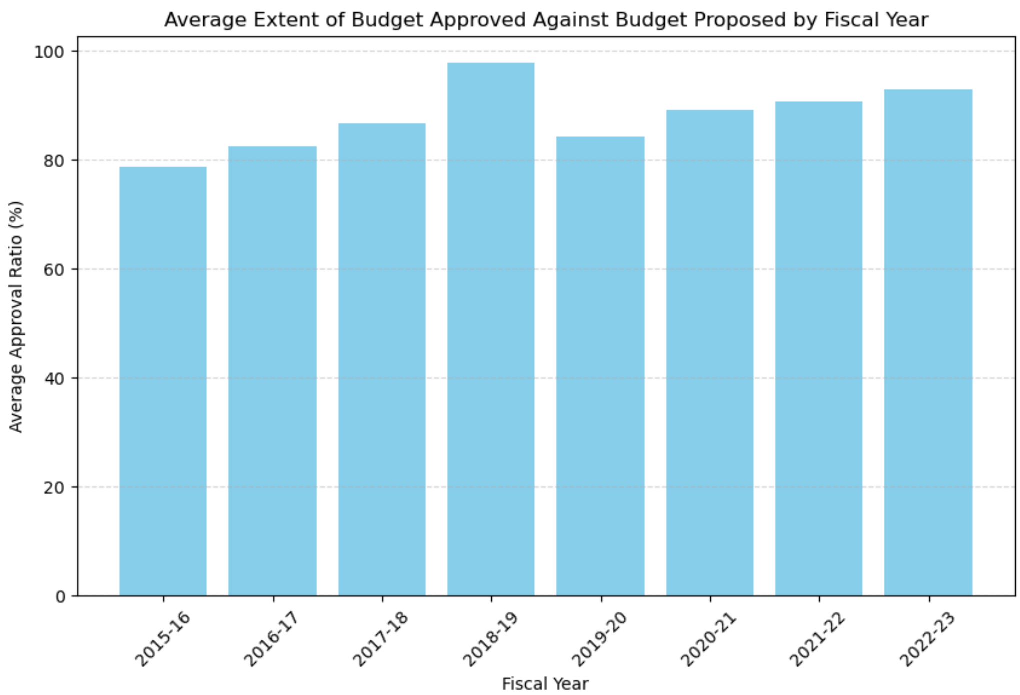


Figure 3.9: Average Extent of Budget Approved Against Budget Proposed by Fiscal Year

3.1.6 Variation of Parameters Over Fiscal Years by State

We analyzed the variation of parameters over fiscal years for different states to understand the trends and fluctuations in budget allocation, fund release, and expenditure.

Key Findings:

- **State-wise Analysis**: Separate side-by-side bar charts were created for each state, showing the variation of parameters including budget proposed, budget approved, release of funds, and total expenditure reported over fiscal years.
- **Insights**: The analysis provides insights into the budget management practices of each state, highlighting variations in budget utilization and expenditure patterns.

The figures below illustrate the variation of parameters over fiscal years for Andhra Pradesh and Bihar:

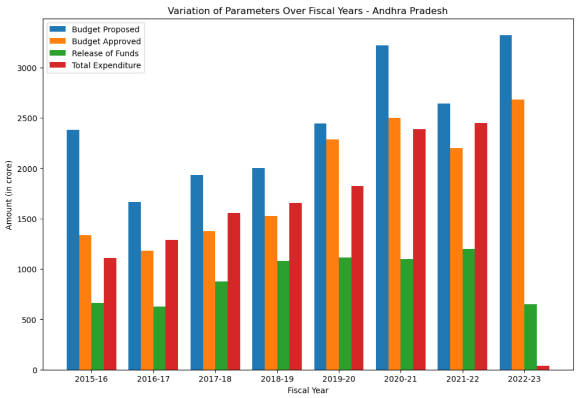


Figure 3.10: Variation of Parameters Over Fiscal Years - Andhra Pradesh

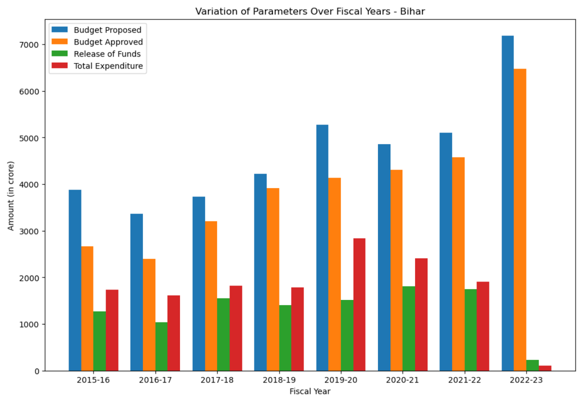


Figure 3.11: Variation of Parameters Over Fiscal Years - Bihar

3.1.7 Total Number of PHCs Functioning in India Over Fiscal Years

We analysed total number of Primary Health Centers (PHCs) functioning over each fiscal year.
Key findings:

- **Trend over Fiscal Years:** By aggregating the data by fiscal year and summing the number of PHCs functioning, the visualization provides insights into how the total number of functioning PHCs has changed over time.
- **Forecasting and Future Planning:** Understanding the historical trend of PHCs functionality can inform future planning and resource allocation decisions. Decision-makers can use this information to forecast future needs and allocate resources accordingly to ensure adequate healthcare infrastructure.

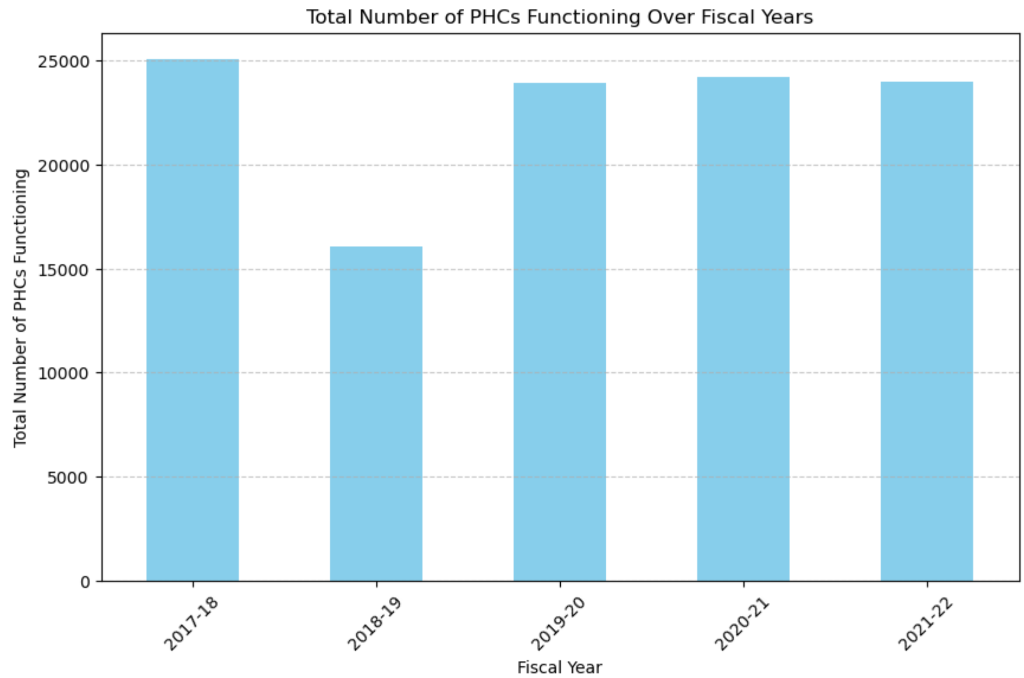


Figure 3.12: Total Number of PHCs Functioning Over Fiscal Years

3.1.8 Trend of different Infrastructure facilities over the years

We aim to display the change in infrastructure facilities across different fiscal years across the country.
Key findings:

- **Relative Importance of Facilities:** Comparing the heights of the bars across different facilities within the same fiscal year allows for an assessment of the relative importance or prevalence of each infrastructure facility.
- **Identifying Deficiencies:** Facilities with consistently low numbers or showing a declining trend across fiscal years may highlight areas of deficiency or gaps in infrastructure provision. These findings can inform targeted interventions or investment strategies to address these deficiencies.

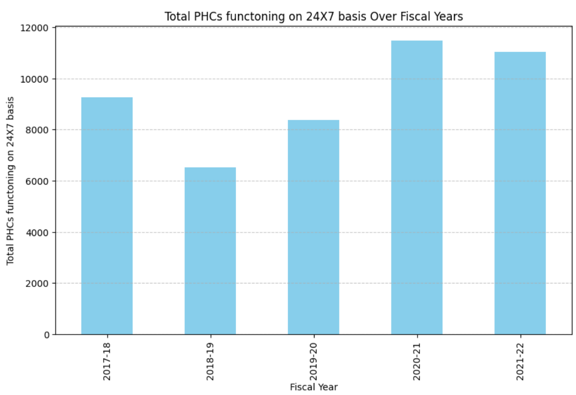


Figure 3.13: Total PHCs functioning on 24X7 basis Over Fiscal Years

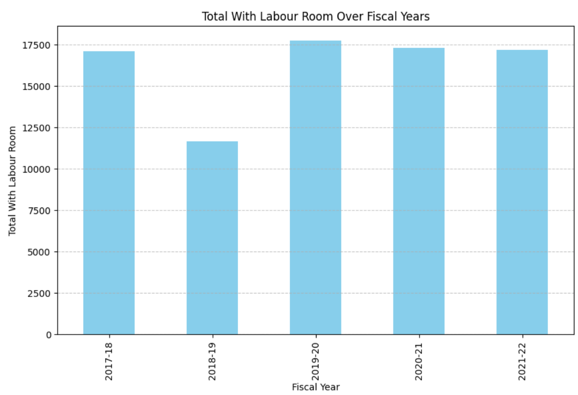


Figure 3.14: Total PHCs with Labour Room Over Fiscal Years

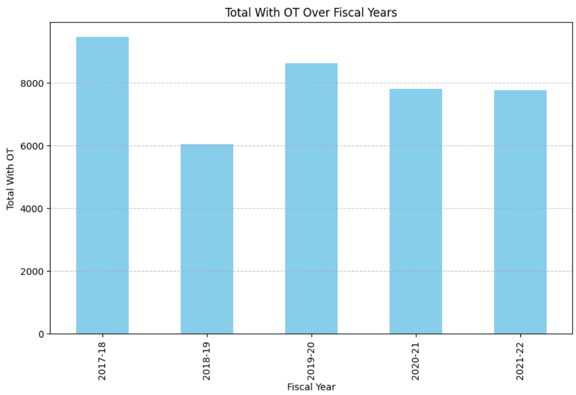


Figure 3.15: Total PHCs with OT Over Fiscal Years

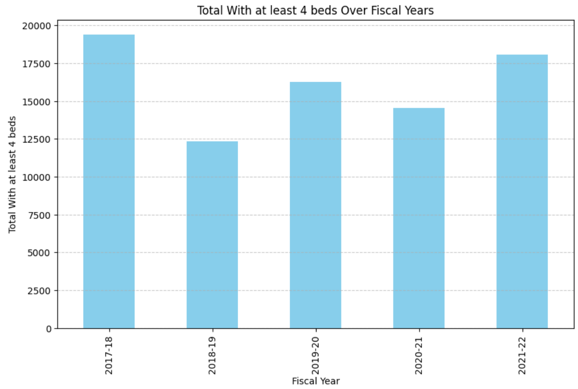


Figure 3.16: Total PHCs with atleast 4 beds Over Fiscal Years

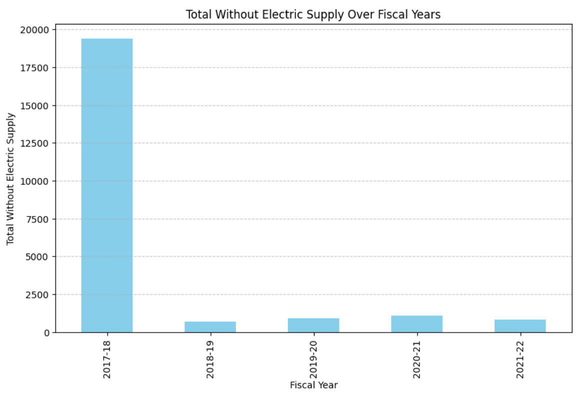


Figure 3.17: Total PHCs without Electric Supply Over Fiscal Years

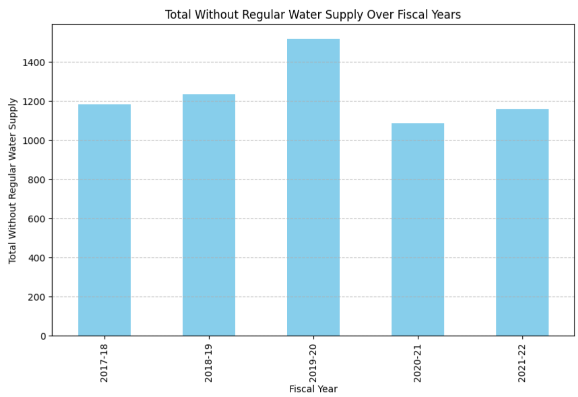


Figure 3.18: Total PHCs without Regular Supply Over Fiscal Years

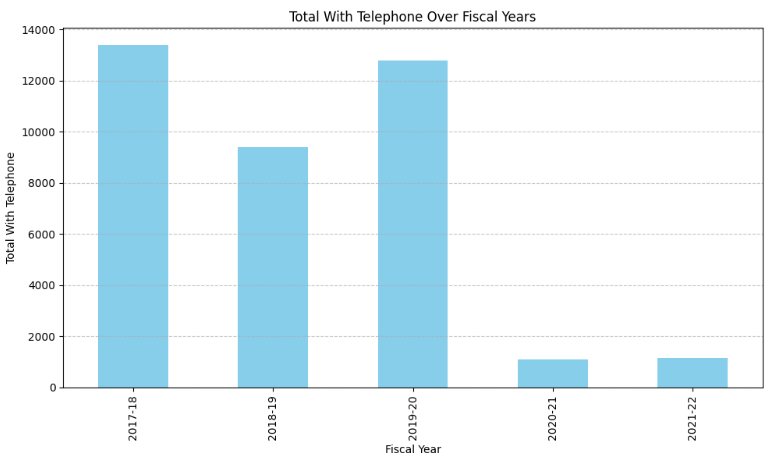


Figure 3.19: Total Number of PHCs with Telephone Over Fiscal Years

Overall, this visualization enables a comprehensive assessment of the change in infrastructure facilities over fiscal years, offering valuable insights for healthcare planning, policy evaluation, and resource allocation.

3.1.9 PHCs functioning for each State/UT over fiscal years

We analyzed the total number of PHCs functioning for each State/UT over fiscal years to understand the health-care trends and scope of improvement. Taking some examples and drawing inferences. The figures below illustrate the number of PHCs functioning for Rajasthan and Gujarat:

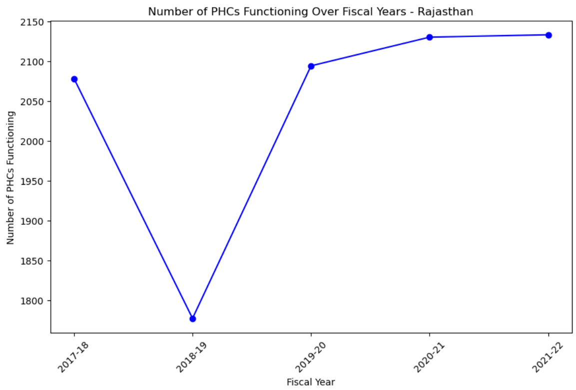


Figure 3.20: Total PHCs functioning for Rajasthan over fiscal years

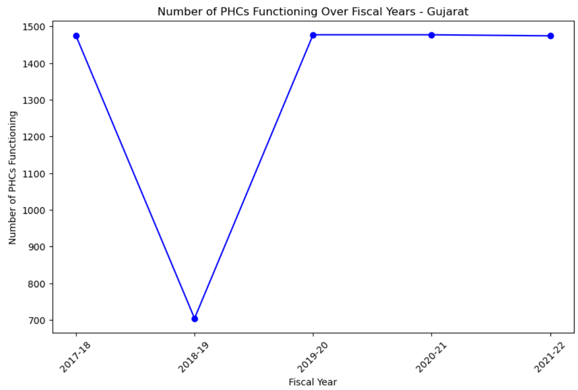


Figure 3.21: Total PHCs functioning for Gujarat over fiscal years

3.1.10 Distribution of functioning PHCs for all the states over the fiscal years

We tend to visualize the distribution of functioning Primary Health Centers (PHCs) among various states over different fiscal years by creating separate bar graphs for each fiscal year.
Key findings:

- **State-wise Distribution:** By examining these bar graphs, one can observe how the number of functioning PHCs is distributed among various states, providing insights into the geographical distribution of healthcare infrastructure.
- **Regional Disparities:** Variations in the height of bars across different states within the same fiscal year highlight regional disparities in the availability of functioning PHCs. States with higher bars indicate a larger number of functioning PHCs, while lower bars suggest fewer functioning PHCs.

Below, we have demonstrated the distribution of functioning PHCs for all the states for 2017-18 and 2021-22.

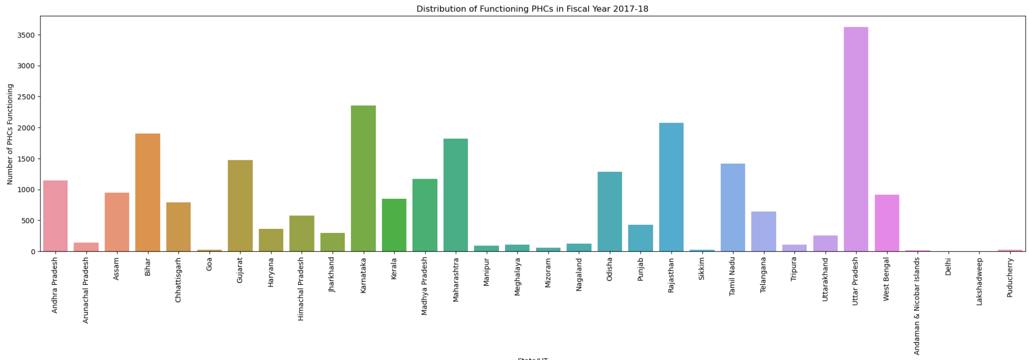


Figure 3.22: Distribution of functioning PHCs for all the states for 2017-18

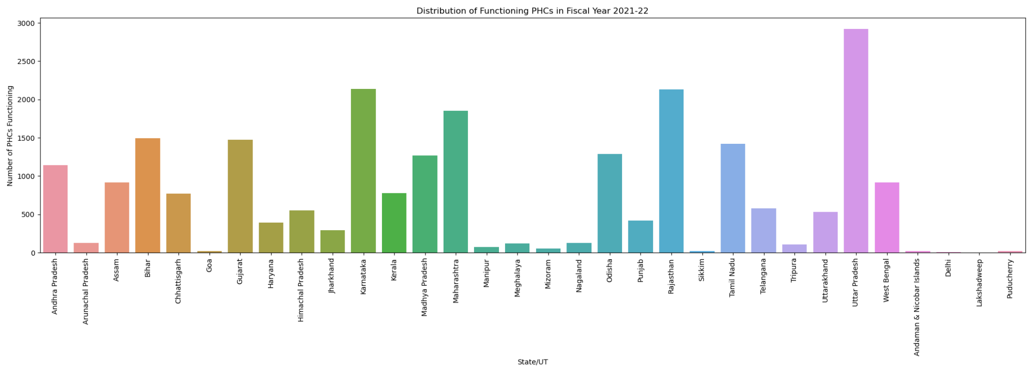


Figure 3.23: Distribution of functioning PHCs for all the states for 2021-22

3.2 Multivariate Analysis

The multivariate analysis section delves deeper into the relationships between multiple variables in the dataset. By employing techniques such as scatter plots, heatmaps, and correlation matrices, we seek to uncover patterns, dependencies, and associations among different variables under consideration. This analysis provides a holistic view of the interplay between various factors impacting the effectiveness and implementation of healthcare policies within the context of the National Rural Health Mission.

3.2.1 Comparison of Proposed vs. Approved Budgets for Each State

We compared proposed vs. approved budgets for each state to analyze discrepancies and trends in budget allocations over fiscal years.

We can see from the figures that the budget approved was even more than what was proposed in some years. This could be a result of Government Policies or schemes in the particular states.

We provide examples of the comparison for Puducherry and Odisha below:

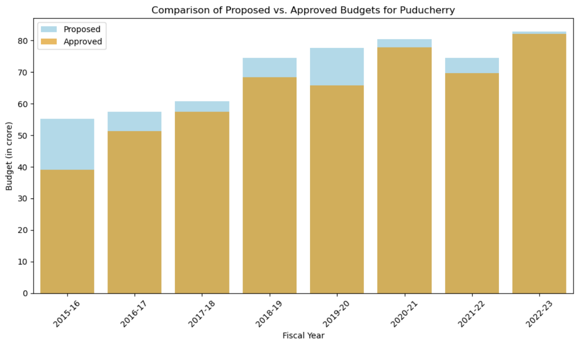


Figure 3.24: Comparison of Proposed vs. Approved Budgets for Puducherry

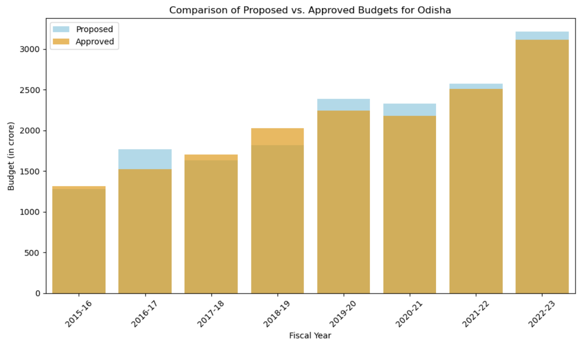


Figure 3.25: Comparison of Proposed vs. Approved Budgets for Odisha

3.2.2 Variation of Parameters Over Fiscal Years by State

We analyzed the variation of parameters over fiscal years for different states to understand the trends and fluctuations in budget allocation, fund release, and expenditure.

Key Findings:

- **State-wise Analysis**: Separate side-by-side bar charts were created for each state, showing the variation of parameters including budget proposed, budget approved, release of funds, and total expenditure reported over fiscal years.
- **Insights**: The analysis provides insights into the budget management practices of each state, highlighting variations in budget utilization and expenditure patterns.

The figures below illustrate the variation of parameters over fiscal years for Andhra Pradesh and Bihar:

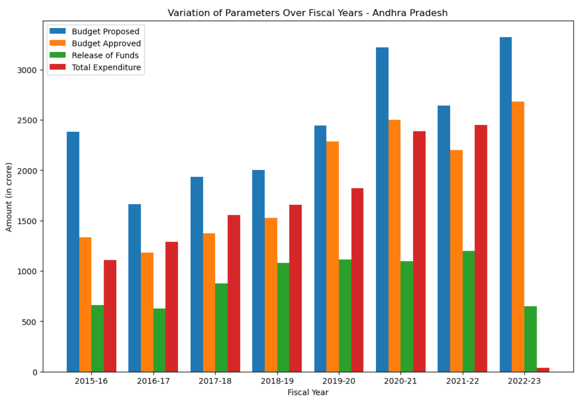


Figure 3.26: Variation of Parameters Over Fiscal Years - Andhra Pradesh

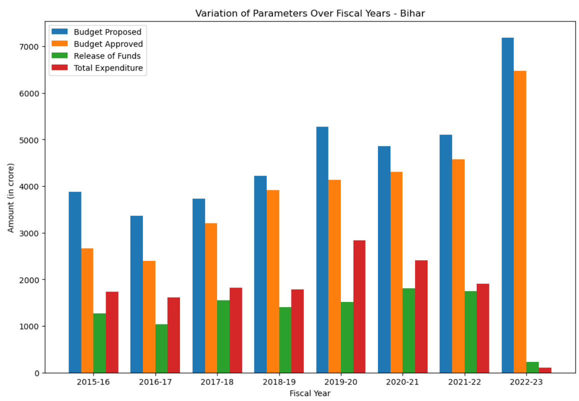


Figure 3.27: Variation of Parameters Over Fiscal Years - Bihar

3.2.3 Distribution of Budget Proposed by Different States/UTs

We visualized the distribution of budget proposed by different states/UTs for each fiscal year using pie charts to gain insights into the allocation patterns across regions.

These pie charts clearly show the proposed budget distribution. We can infer from these that Uttar Pradesh, Bihar, Maharashtra and Madhya Pradesh are constantly proposing high budgets as compared to other states.

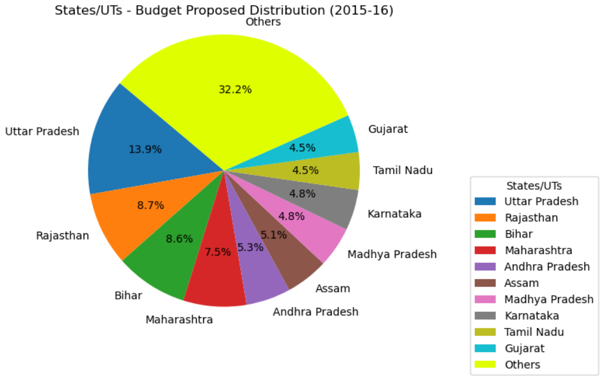


Figure 3.28: *
States/UTs - Budget Proposed Distribution (2015-16)

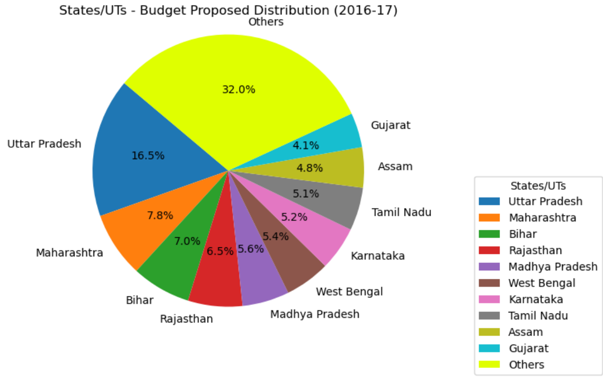


Figure 3.29: *
States/UTs - Budget Proposed Distribution (2016-17)

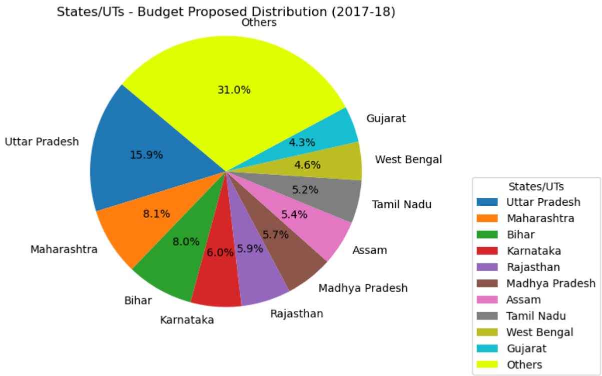


Figure 3.30: *
States/UTs - Budget Proposed Distribution (2017-18)

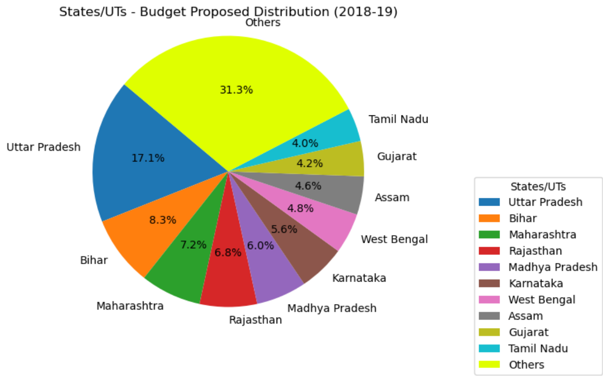


Figure 3.31: *
States/UTs - Budget Proposed Distribution (2018-19)

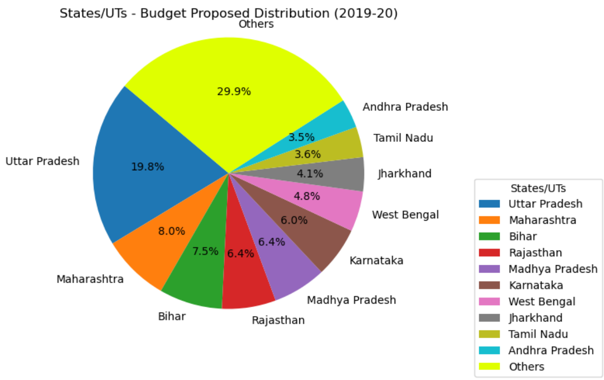


Figure 3.32: *
States/UTs - Budget Proposed Distribution (2019-20)

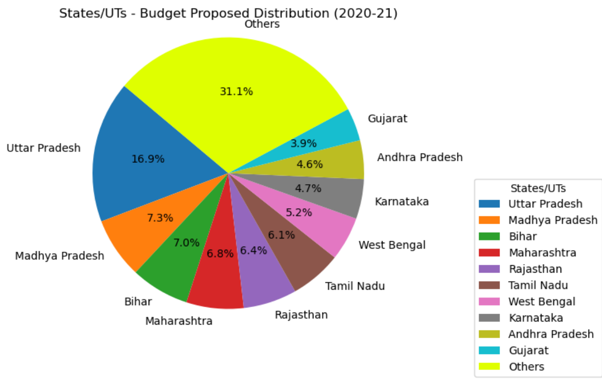


Figure 3.33: *
States/UTs - Budget Proposed Distribution (2020-21)

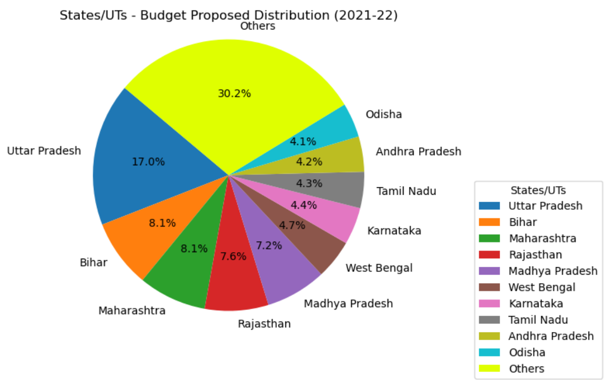


Figure 3.34: *
States/UTs - Budget Proposed Distribution (2021-22)

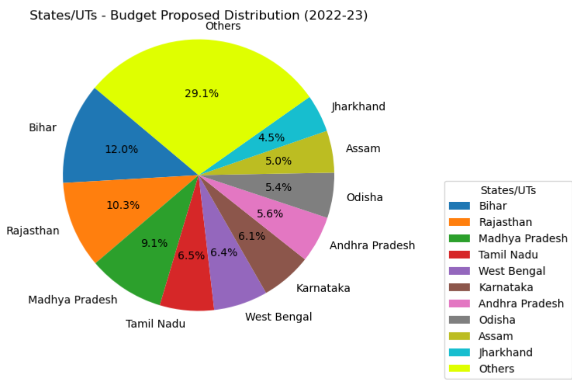


Figure 3.35: *
States/UTs - Budget Proposed Distribution (2022-23)

3.2.4 Availability of health centers without essential supplies for each state over the years

We aim to display the count of health centers without both electric supply and regular water supply for each state over fiscal years.
Key findings:

- **Comparison of Essential Supplies:** Comparing the trends of both electric and regular water supply on the same plot enables a direct comparison of their availability and any potential correlations or disparities between them.
- **State-Specific Trends:** By plotting state-specific data separately, it becomes easier to identify variations in the availability of essential supplies across different states and regions.

The figures below illustrate the variation of count of health centers without essential supplies for Himachal Pradesh and Arunachal Pradesh:

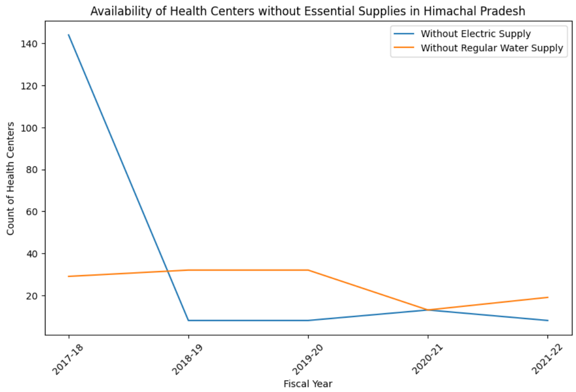


Figure 3.36: Count of Health Centers without essential supplies for Himachal Pradesh Over Fiscal Years

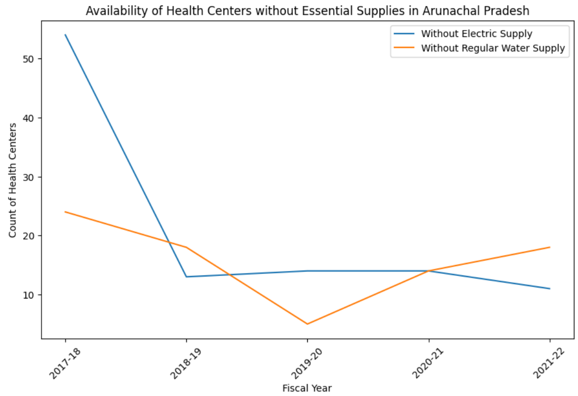


Figure 3.37: Count of Health Centers without essential supplies for Arunachal Pradesh Over Fiscal Years

3.2.5 Variation in health facilities in India over the years

We visualize the change in various health center facilities over fiscal years by plotting line plots for each facility.
Key findings:

- The visualization provides an overview of the change in health center facilities aggregated over fiscal years. By plotting line plots for each facility, it allows for an examination of the overall trend in the availability of different health center amenities over time.

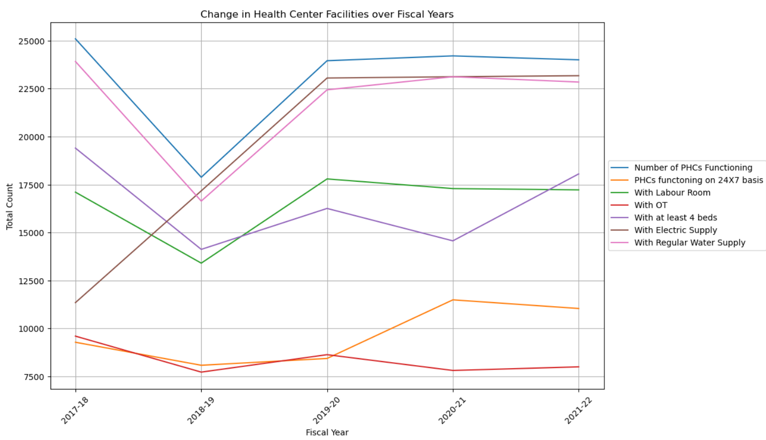


Figure 3.38: Total Number of PHCs with Telephone Over Fiscal Years

3.2.6 Changes in Infrastructure facilities over fiscal years for each state

We examine infrastructure facility trends over fiscal years for each state.

- **State-Specific Infrastructure Trends:** The visualization provides insights into the changes in infrastructure facilities for each state over fiscal years. By plotting infrastructure parameters for each state separately, it allows for a detailed examination of state-specific trends and variations in infrastructure development.

- **Identification of Priorities and Challenges:** Examining the trends in infrastructure facilities helps in identifying priorities and challenges specific to each state. States experiencing consistent improvements in infrastructure parameters may indicate successful development initiatives, while those facing stagnation or deterioration may highlight areas requiring attention and intervention.

We illustrate the variation of changes in infrastructure facilities over the years for Assam and Kerala.

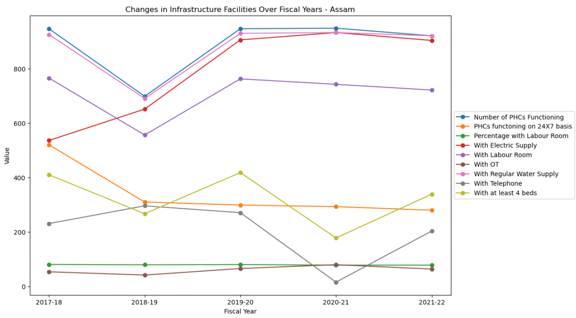


Figure 3.39: Infrastructural facilities in Assam Over Fiscal Years

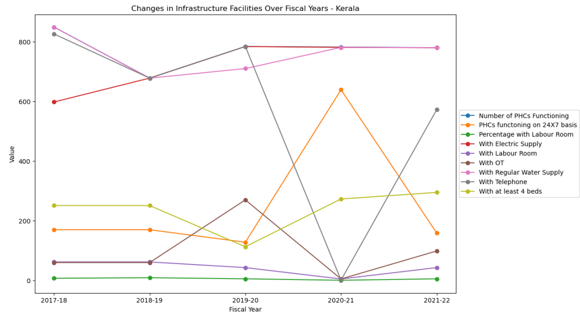


Figure 3.40: Count of Health Centers without essential supplies for Kerala Over Fiscal Years

3.2.7 Variation of each parameter over Fiscal Years for each state

We iterate over each state and creates separate side-by-side bar charts to visualize the variation of different parameters over fiscal years.
Key findings:

- Identifying Infrastructure Trends: By observing the height and distribution of bars for each parameter, we can identify trends in the availability of specific healthcare facilities over fiscal years. For example, increasing bars may indicate improvements in infrastructure, while decreasing bars may signify challenges or deficiencies.

We illustrate the variation of each parameter over the years for Punjab and Jharkhand.

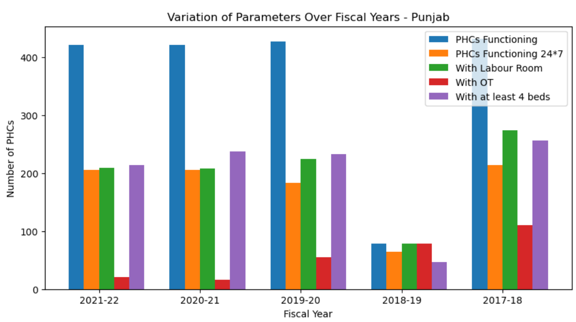


Figure 3.41: Infrastructural facilities in Assam Over Fiscal Years

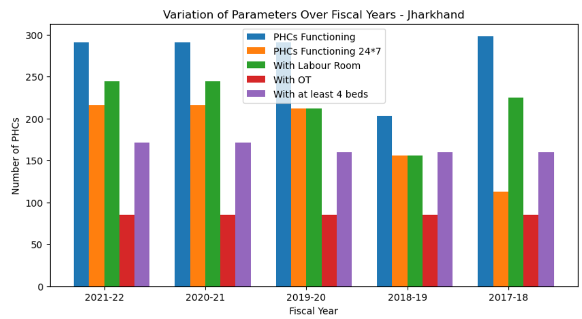


Figure 3.42: Count of Health Centers without essential supplies in Kerala Over Fiscal Years

Chapter 4. Feature Engineering

Feature Engineering is a crucial step in exploratory data analysis (EDA) that involves creating new features or modifying existing ones to improve the performance of machine learning models. It focuses on extracting relevant information from raw data and transforming it into a format that is more suitable for modeling.

4.1 Feature extraction

4.1.1 Dataset-1: Financial Resources Allocated to the NRHM

We performed one-hot encoding on categorical columns in the DataFrame, creating binary indicator columns for each category within the categorical variables. Following are the first 5 rows of the encoded Dataframe.

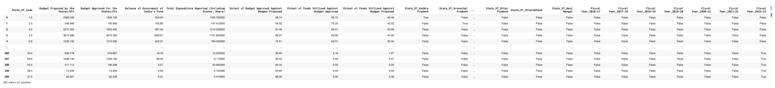


Figure 4.1:

Standardization is a data preprocessing technique used to transform numeric features in a dataset to have a mean of 0 and a standard deviation of 1. This process is also known as z-score normalization. We implemented Standardization using 'StandardScaler' class from the Scikit-learn library used for standardizing features in Python. Following are the first 5 rows of the dataframe after standardization.

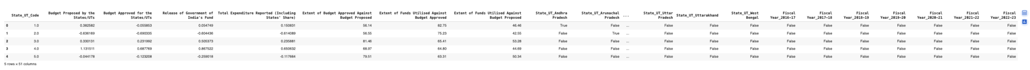


Figure 4.2:

Now, we convert a numerical feature "Budget Proposed by the States/UTs" into respective categorical bins according to their values.

```
# Dividing the budget proposed by the states into low , medium, and high
df1['Budget_Category'] = pd.cut(df1['Budget Proposed by the States/UTs'], bins=3, labels=['Low', 'Medium', 'High'])
```

Figure 4.3:

With the help of the following KDE plot, we can definitely conclude that the feature "Budget Proposed by the States/UTs" follows right-skewed distribution (positive skewness).

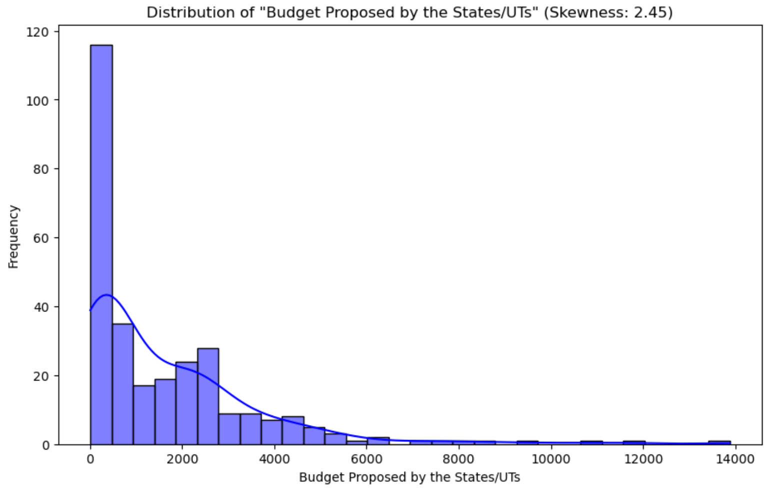


Figure 4.4:

By using feature transformation and implementing log-transformation, we get the following KDE plot which follows a left-skewed distribution (negative skewness).

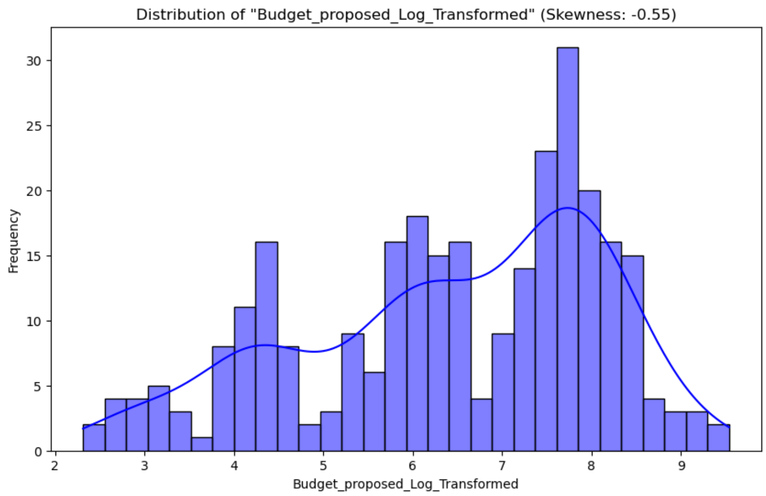


Figure 4.5:

By using feature transformation and implementing square root-transformation, we get the following KDE plot which follows a right-skewed distribution (positive skewness).

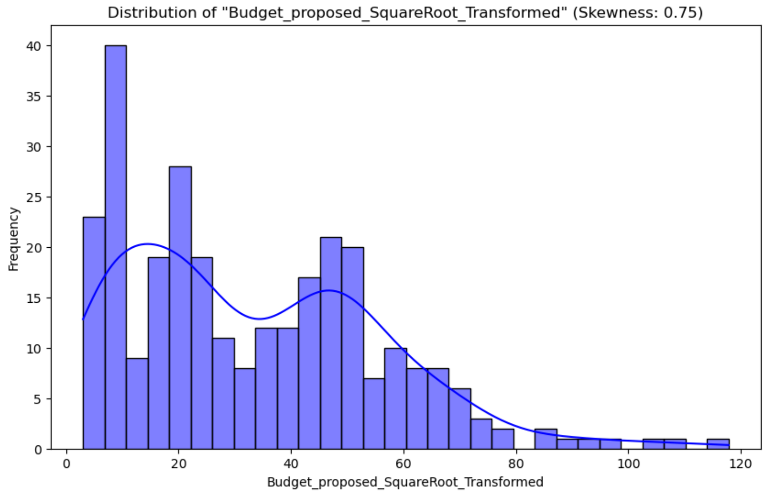


Figure 4.6:

4.1.2 Dataset-2: Status of Healthcare Infrastructure in Rural Areas

We performed one-hot encoding on categorical columns on this DataFrame, creating binary indicator columns for each category within the categorical variables.
The first 5 rows of the encoded dataframe is, as follows:

Encoded DataFrame:

	Number of PHCs Functioning	PHCs functoning on 24X7 basis	With Labour Room	With OT	With at least 4 beds	Without Electric Supply	Without Regular Water Supply	With Telephone	With Electric Supply	With Regular Water Supply	...	State/UT_Tamil Nadu	State/UT_Telangana	State/UT_Tripura	State/UT_Uttar Pradesh	State/UT_Uttarakhand	State/UT_West Bengal	Fiscal Year_2018-19	Fiscal Year_2019-20	Fiscal Year_2020-21	Fiscal Year_2021-22
0	1142.0	1142.0	1134.0	1129.0	1142.0	0	0	895.0	1142.0	1142.0	...	False	False	False	False	False	False	False	False	False	True
1	126.0	42.0	67.0	1.0	46.0	11	18	18.0	115.0	108.0	...	False	False	False	False	False	False	False	False	False	True
2	920.0	280.0	721.0	64.0	339.0	17	0	203.5	903.0	920.0	...	False	False	False	False	False	False	False	False	False	True
3	1492.0	630.0	465.0	254.0	612.0	326	374	374.0	1166.0	1118.0	...	False	False	False	False	False	False	False	False	False	True
4	770.0	477.0	752.0	63.0	610.0	17	8	8.0	753.0	762.0	...	False	False	False	False	False	False	False	False	False	True
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
159	913.0	257.0	462.0	108.0	850.0	850	21	89.0	63.0	892.0	...	False	False	False	False	False	True	False	False	False	False
160	22.0	20.0	22.0	20.0	22.0	22	0	22.0	22.0	22.0	...	False	False	False	False	False	False	False	False	False	False
162	5.0	1.0	1.0	1.0	1.0	1	0	5.0	4.0	5.0	...	False	False	False	False	False	False	False	False	False	False
163	4.0	4.0	4.0	4.0	4.0	4	0	4.0	4.0	4.0	...	False	False	False	False	False	False	False	False	False	False
164	24.0	15.0	24.0	4.0	15.0	15	0	24.0	9.0	24.0	...	False	False	False	False	False	False	False	False	False	False

160 rows × 46 columns

Figure 4.7:

In order to standardize the data, we implemented using the "StandardScaler" class from scikit-learn library. The first 5 rows of the standardized dataframe is, as follows:

	Number of PHCs Functioning	PHCs functoning on 24X7 basis	With Labour Room	With OT	With at least 4 beds	Without Electric Supply	Without Regular Water Supply	With Telephone	With Electric Supply	With Regular Water Supply	...	State/UT_Tamil Nadu	State/UT_Telangana	State/UT_Tripura	State/UT_Uttar Pradesh	State/UT_Uttarakhand	State/UT_West Bengal	Fiscal Year_2018-19	Fiscal Year_2019-20	Fiscal Year_2020-21	Fiscal Year_2021-22
0	0.557370	2.299522	1.030414	2.151284	0.928342	-0.329785	-0.500445	1.436125	0.773568	0.640943	...	0	0	0	0	0	0	0	0	0	1
1	-0.783160	-0.711813	-0.753071	-0.644421	-0.694228	-0.304451	-0.267115	-0.614661	-0.724313	-0.796383	...	0	0	0	0	0	0	0	0	0	1
2	0.264459	-0.060269	0.340086	-0.488278	-0.260457	-0.290633	-0.500445	-0.180886	0.424986	0.332349	...	0	0	0	0	0	0	0	0	0	1
3	1.019167	0.897883	-0.087816	-0.017370	0.143705	0.421015	4.347644	0.217813	0.808572	0.607582	...	0	0	0	0	0	0	0	0	0	1
4	0.066546	0.479033	0.391903	-0.490756	0.140744	-0.290633	-0.396743	-0.638045	0.206211	0.112719	...	0	0	0	0	0	0	0	0	0	1

5 rows × 46 columns

Figure 4.8:

We group States/UTs into regions based on the geographical location, such as 'NorthIndia', 'SouthIndia', 'EastIndia', 'WestIndia', 'CentralIndia', and 'NortheastIndia'. In this manner, we get an additional feature "Region", into our dataframe.

	State/UT	Number of PHCs Functioning	PHCs functoning on 24X7 basis	With Labour Room	With OT	With at least 4 beds	Without Electric Supply	Without Regular Water Supply	With Telephone	Fiscal Year	With Electric Supply	With Regular Water Supply	Percentage with Labour Room	Region	Year
0	Andhra Pradesh	1142.0	1142.0	1134.0	1129.0	1142.0	0	0	895.0	2022-01-01	1142.0	1142.0	99.299475	South India	2022
1	Arunachal Pradesh	126.0	42.0	67.0	1.0	46.0	11	18	18.0	2022-01-01	115.0	108.0	53.174603	Northeast India	2022
2	Assam	920.0	280.0	721.0	64.0	339.0	17	0	203.5	2022-01-01	903.0	920.0	78.369565	Northeast India	2022
3	Bihar	1492.0	630.0	465.0	254.0	612.0	326	374	374.0	2022-01-01	1166.0	1118.0	31.166220	East India	2022
4	Chhattisgarh	770.0	477.0	752.0	63.0	610.0	17	8	8.0	2022-01-01	753.0	762.0	97.662338	Central India	2022

Figure 4.9:

4.2 Feature selection

4.2.1 Dataset-1: Financial Resources Allocated to the NRHM

We illustrate the correlation matrix, which shows the correlation coefficients between pairs of variables in a dataset. Each cell in the matrix represents the correlation coefficient between two variables, indicating the strength and direction of their linear relationship.

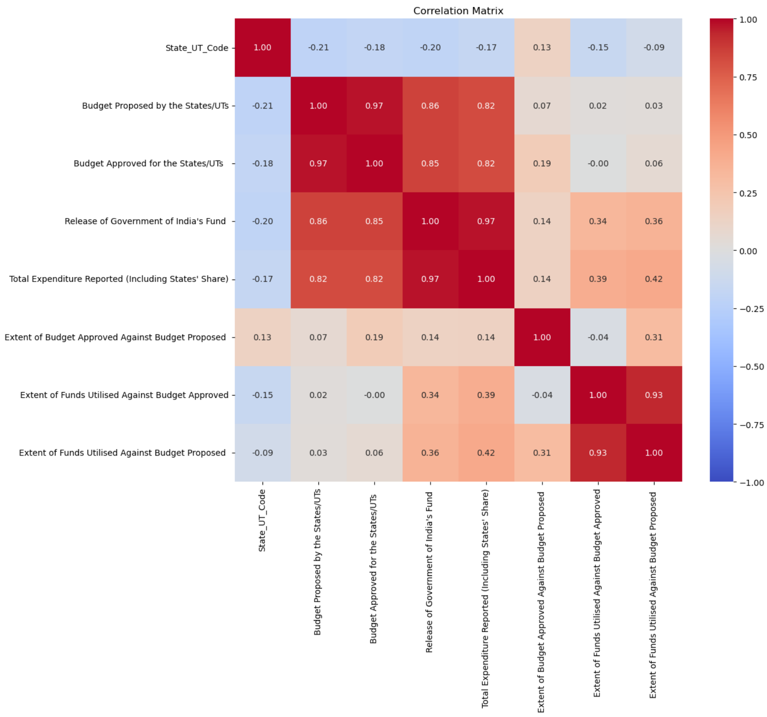


Figure 4.10:

Analysing the correlation matrix, we can conclude that the variables have correlation coefficient greater than 0.5 with the target variable, that is, "Total Expenditure Reported (Including States' Share)", can be considered as **Important Features**. Therefore, "Release of Government of India's Fund ", "Budget Proposed by the States/UTs", "Budget Approved for the States/UTs" are more relevant features with respect to the target variable.

```
# Identify features with highest absolute correlation with the target variable
target_correlation = correlation_matrix["Total Expenditure Reported (Including States' Share)"].abs().sort_values(ascending=False)
important_features = target_correlation[target_correlation >= 0.5].index.tolist()
print("Important Features based on Correlation:")
print(important_features)
```

```
Important Features based on Correlation:
["Total Expenditure Reported (Including States' Share)", "Release of Government of India's Fund ", 'Budget Proposed by the States/UTs', 'Budget Approved for the States/UTs ']
```

Figure 4.11:

4.2.2 Dataset-2: Status of Healthcare Infrastructure in Rural Areas

We set our target variable to be 'With Labour Room'. We can select important features according to the correlation matrix. We can conclude that variables that have correlation coefficient greater than 0.5 with the target variable, can be considered as **Important Features**.

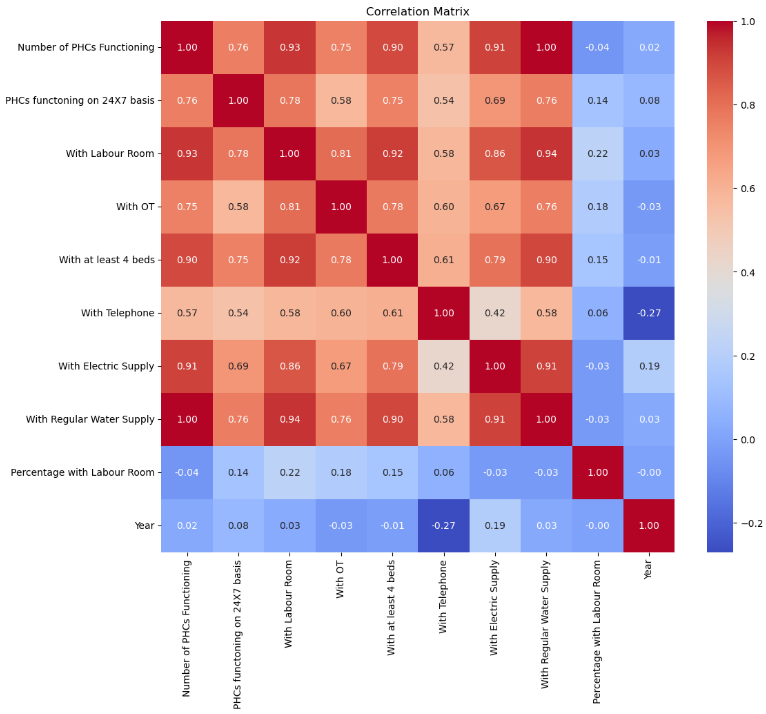


Figure 4.12:

Based on the analysis, we can observe that "With Regular Water Supply", "Number of PHCs Functioning", "With at least 4 beds", "With Electric Supply", "With OT", "PHCs functioning on 24X7 basis", "With Telephone" are more relevant features with respect to the target variable.

Chapter 5. Model fitting

Model Fitting serves as a crucial step to understand the relationships between variables and to derive insights from the data. Depending on the nature of your data and the objectives of your analysis, we'll choose an appropriate model to fit. This could range from simple linear regression to more complex models like decision trees, random forests, or neural networks. The choice of model should align with the assumptions of your data and the goals of your analysis.

Once we've selected a model, the next step is to fit it to your data. This involves estimating the parameters of the chosen model using optimization techniques like ordinary least squares (OLS) for linear regression or gradient descent for neural networks. During this process, we'll evaluate the goodness of fit using metrics like R-squared, mean squared error (MSE), or accuracy, depending on the type of model.

After fitting the model, it's crucial to assess its performance. This could involve visualizing the fitted model against the actual data points to see how well it captures the underlying patterns. We might also use cross-validation techniques to estimate the model's generalization error and ensure it's not overfitting to the training data.

Finally, the fitted models provide insights into the underlying relationships in the data. We can interpret the coefficients or feature importances to understand the factors that influence the outcome variable. This interpretation helps us draw meaningful conclusions and make informed decisions based on our analysis.

5.1 Regression

5.1.1 Dataset-1: Financial Resources Allocated to the NRHM

First, we need import all the required function in the 'scikit-learn' library.

```python
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import PolynomialFeatures, StandardScaler
from sklearn.metrics import accuracy_score
from sklearn.preprocessing import StandardScaler
from sklearn.linear_model import LinearRegression
from sklearn.metrics import mean_squared_error
```

Figure 5.1:

Linear Regression will be the most optimal regression model according to the values in the dataset. Therefore, we'll try to train our dataset according to the Linear Regression, using sk-learn library.

We set our target variable as "Total Expenditure Reported (Including States' Share)", to be predicted on the basis of rest of the columns. The following code demonstrates it.

```python
#  Define features (X) and target variable (y)
X = df_encoded.drop("Total Expenditure Reported (Including States' Share)", axis=1)
y = df_encoded["Total Expenditure Reported (Including States' Share)"]

# Split the dataset into training and testing sets (80% train, 20% test)
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)

# Initialize and train LinearRegression model
model = LinearRegression()
model.fit(X_train, y_train)

# Make predictions on the test set
y_pred = model.predict(X_test)

# Evaluate model performance (e.g., using Mean Squared Error for regression)
mse = mean_squared_error(y_test, y_pred)
print(f'Mean Squared Error: {mse:.4f}')
```

```
Mean Squared Error: 0.1021
```

Figure 5.2:

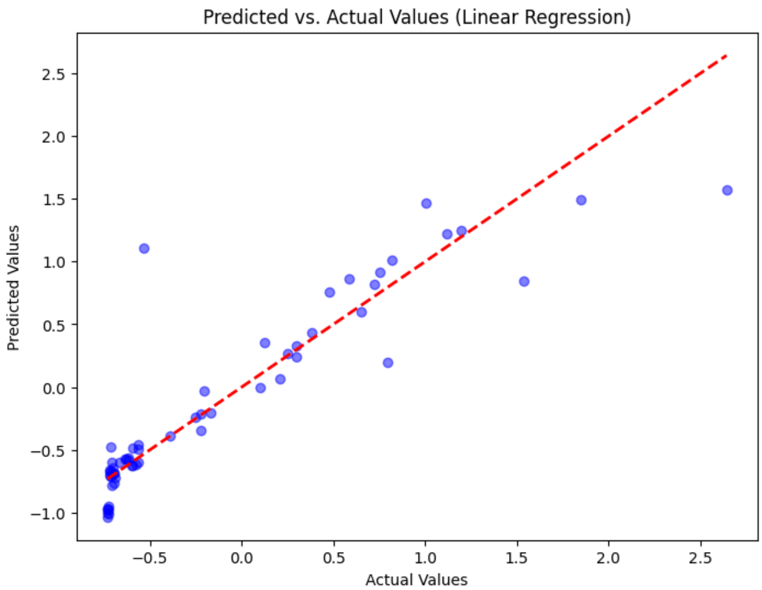


Figure 5.3: Model fitting using Linear Regression

Now, we'll implement regression model using Linear Regression Model, Random Forest Regressor, and Support Vector Regressor (SVR). This is an approach to find the best model that fits the given dataset. Given below are the illustration after implementing the regression models on the dataset.

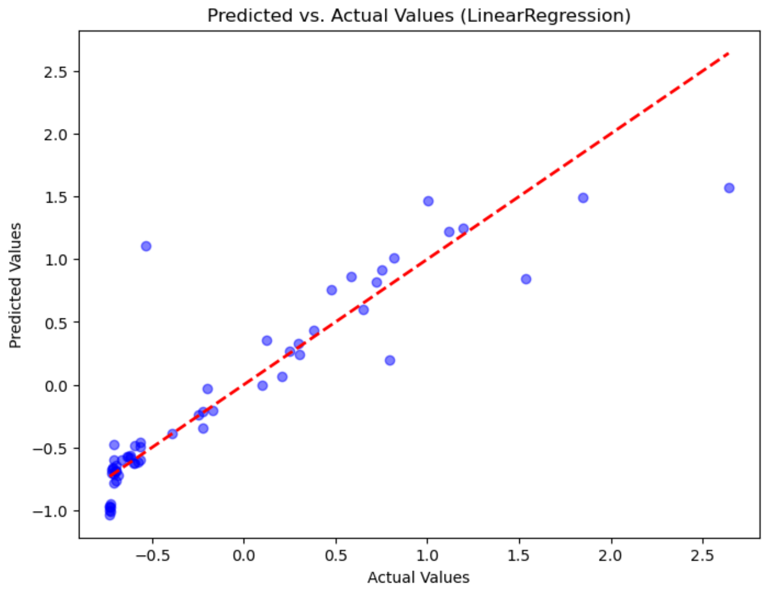


Figure 5.4: Model fitting using Linear Regression

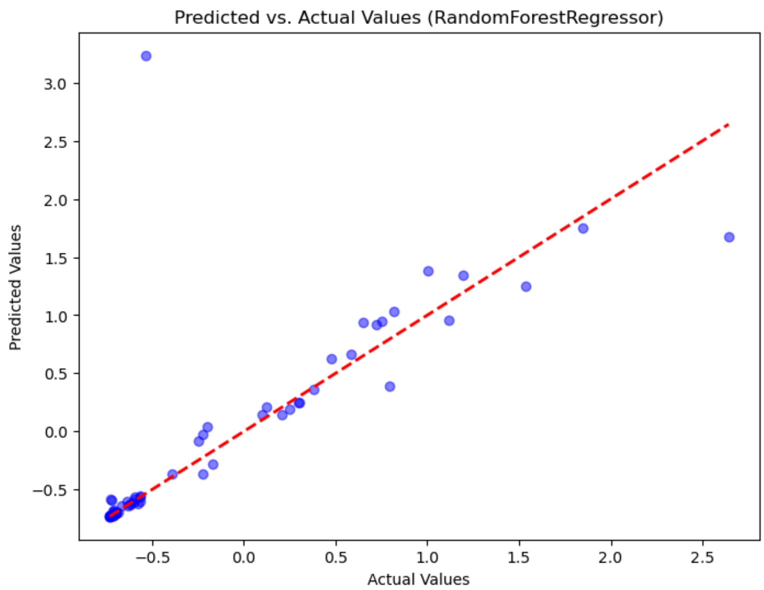


Figure 5.5: Model fitting using Random Forest Regressor

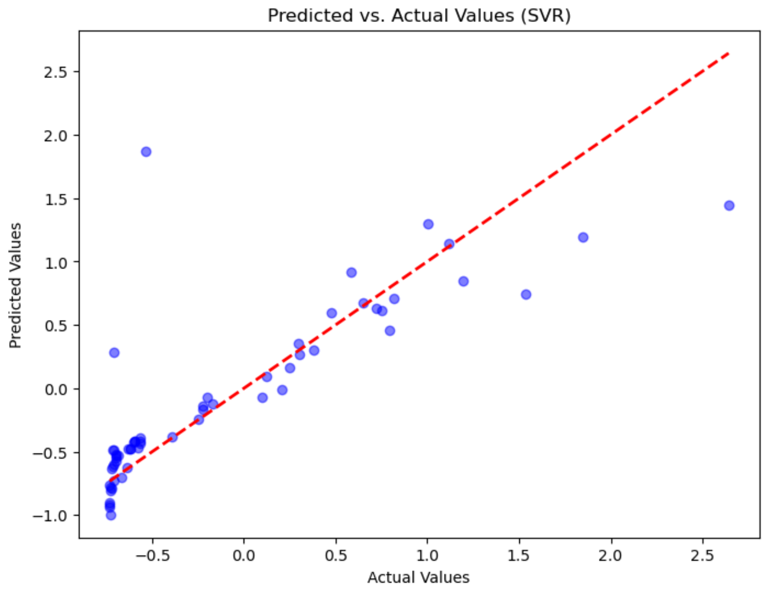


Figure 5.6: Model fitting using Support Vector Regressor

Based on the evaluation metrics (R-squared and Mean Squared Error) obtained from the different regression models (LinearRegression, RandomForestRegressor, SVR), here are the results and interpretations:

1. **Linear Regression**
   - R-squared (R^2): 0.8255, which indicates that approximately 82.55% of the variance in the target variable (Total Expenditure Reported) is explained by the linear relationship with the predictors.
   - Mean Squared Error (MSE): 0.1021, which represents the average squared difference between the predicted and actual values. Lower MSE values indicate better model performance.
2. **Random Forest Regressor**
   - R-squared (R^2): 0.5347, suggests that the Random Forest model explains about 53.47% of the variance in the target variable.
   - Mean Squared Error (MSE): 0.2723, indicates that the Random Forest model has more prediction errors on average, as compared to Linear Regression.
3. **Support Vector Regressor (SVR)**
   - R-squared (R^2): 0.6946, indiactes that the SVR model explains approximately 69.46% of the variance in the target variable.
   - Mean Squared Error (MSE): 0.1787, is lower than that of the Random Forest model, indicating potentially better performance in terms of prediction accuracy.

To conclude with Model Comparison and Selection, Linear Regression (0.8255) has the highest R-squared value (0.8255) among the three models, suggesting the best overall fit to the data. It also has the lowest MSE (0.1021), indicating more accurate predictions on average compared to the other models.

Chapter 6. Conclusion & future scope

In summary, this report offers an overview of the National Rural Health Mission (NRHM) and National Health Mission (NHM), shedding light on various governmental efforts. Through our analysis, we have noted a significant allocation of funds over the years to enhance healthcare infrastructure in both rural and urban areas. Notably, certain states have utilized a considerable portion of these funds, while others have seen a more equitable distribution across different years. Additionally, our examination reveals consistent higher utilization of funds by states like Uttar Pradesh, underscoring its prominent position in NRHM spending.

Furthermore, the report delves into various healthcare facility metrics, including the presence of Primary Health Centers (PHCs) and their nationwide distribution. Across all categories, Uttar Pradesh and Bihar consistently emerges at the forefront, whereas northeastern states typically rank lower. Hence, it is evident that Uttar Pradesh has been the primary beneficiary of the NRHM and NHM initiatives.

In conclusion, our analysis underscores the pivotal role of the NRHM and NHM initiatives in bolstering healthcare infrastructure across India. The substantial allocation of funds and the emphasis on improving healthcare access have undoubtedly yielded commendable outcomes, particularly in states like Uttar Pradesh. However, there remains room for further exploration and enhancement.

Looking ahead, future research could delve deeper into the specific impact of NRHM and NHM interventions on key healthcare indicators such as maternal and child health, disease prevalence, and healthcare accessibility in remote regions. Additionally, an examination of the effectiveness of various healthcare delivery models, including the role of community health workers and telemedicine, could provide valuable insights for optimizing resource allocation and service delivery.

Moreover, there is a pressing need to address disparities in healthcare access and quality across different regions and demographic groups. Strategies aimed at promoting equitable distribution of healthcare resources, improving healthcare infrastructure in underserved areas, and strengthening healthcare delivery systems are imperative for achieving universal health coverage and addressing healthcare inequalities.

In essence, while the NRHM and NHM have made significant strides in advancing healthcare in India, continued efforts and innovative approaches are essential to ensure that all citizens have access to quality healthcare services, irrespective of their geographical location or socioeconomic status.

6.1 Findings and Observations

1. **Significant State-Level Disparities**: Our analysis revealed notable variations in the utilization of funds allocated under the National Rural Health Mission (NRHM) and National Health Mission (NHM) across different states in India.

2. **Top Performers**: States such as Uttar Pradesh (UP), Bihar, and Madhya Pradesh (MP) consistently emerged as top performers in terms of fund utilization and healthcare infrastructure development. These states have demonstrated a proactive approach towards improving healthcare facilities and services.

3. **Prominent Beneficiaries**: Uttar Pradesh, in particular, stood out as a prime beneficiary of NRHM and NHM initiatives, consistently utilizing a significant portion of allocated funds. This underscores the state's pivotal role in driving healthcare advancements at the national level.

4. **Regional Disparities**: On the other hand, states like Assam and Odisha were observed to have lower utilization rates of NRHM and NHM funds, indicating potential challenges or limitations in healthcare infrastructure development and resource allocation in these regions.

5. **Urgent Attention Needed**: The disparities in fund utilization highlight the urgent need for targeted interventions and strategic resource allocation to address healthcare inequalities across states. Efforts should focus on bolstering healthcare infrastructure and service delivery mechanisms in underprivileged regions to ensure equitable access to healthcare services for all citizens.

6.2 Challenges Faced

6.2.1 Preparing the NRHM Report

- **Data Accessibility**: One of the primary challenges encountered in preparing the NRHM report was the accessibility and availability of comprehensive data. Gathering accurate and up-to-date information regarding fund allocations, healthcare infrastructure, and healthcare outcomes across different states posed significant hurdles.

- **Data Quality**: Ensuring the quality and reliability of the collected data presented another challenge. Inconsistent reporting standards, data discrepancies, and missing information in official reports and datasets required careful validation and verification processes.

- **Interpretation Complexity**: Interpreting the gathered data and deriving meaningful insights while considering the complex interplay of various socio-economic factors, healthcare policies, and regional dynamics demanded a nuanced approach. Analyzing the multifaceted nature of healthcare challenges and identifying underlying trends required extensive expertise and domain knowledge.

6.2.2 Conducting Exploratory Data Analysis (EDA)

- **Data Preprocessing**: Prior to conducting EDA, extensive data preprocessing was necessary to clean, format, and standardize the collected data. Dealing with missing values, handling outliers, and normalizing data from diverse sources presented initial obstacles.

- **Dimensionality Reduction**: Exploring large datasets with numerous variables posed challenges in terms of dimensionality reduction. Selecting relevant features and employing techniques such as principal component analysis (PCA) or t-distributed stochastic neighbor embedding (t-SNE) required careful consideration to retain essential information while reducing computational complexity.

- **Visualization Complexity**: Visualizing complex healthcare data in a meaningful and interpretable manner was another challenge. Choosing appropriate visualization techniques, effectively communicating insights, and ensuring clarity for diverse stakeholders necessitated creative approaches and iterative refinement.

6.3 Future Plans

6.3.1 Enhancing Data Accessibility

Efforts will be made to improve the accessibility and availability of comprehensive healthcare data, including fund allocations, healthcare infrastructure metrics, and healthcare outcomes. Collaboration with government agencies, healthcare institutions, and research organizations will be sought to streamline data collection and reporting processes.

6.3.2 Improving Data Quality

Initiatives will be undertaken to enhance the quality and reliability of healthcare data through standardized reporting protocols, data validation mechanisms, and quality assurance procedures. Regular audits and validation checks will be conducted to ensure data accuracy and integrity.

6.3.3 Advanced Analytics and Machine Learning

Integration of advanced analytics and machine learning techniques will be explored to gain deeper insights from healthcare data. Predictive modeling, sentiment analysis, and anomaly detection algorithms will be employed to identify emerging trends, predict healthcare outcomes, and optimize resource allocation strategies.

6.3.4 Stakeholder Engagement and Collaboration

Engagement with diverse stakeholders, including policymakers, healthcare professionals, academia, and community representatives, will be prioritized to foster collaboration and co-creation of healthcare solutions. Multi-disciplinary research collaborations and knowledge-sharing platforms will be established to facilitate collective action and knowledge exchange.

6.3.5 Long-term Sustainability and Impact Assessment

Long-term sustainability of healthcare interventions and their impact on population health outcomes will be assessed through rigorous evaluation frameworks and impact assessment studies. Continuous monitoring and evaluation will be conducted to track progress, identify challenges, and inform evidence-based decision-making for sustainable healthcare development.

6.4 Group Contributions

Each member of the group contributed equally to every stage of the project. The following summarizes the joint efforts of all members:

- **Data Collection and Preprocessing**: All members collaborated in gathering and cleaning the data, ensuring its accuracy and reliability for analysis.
- **Exploratory Data Analysis**: Jointly conducted thorough exploratory data analysis, identifying trends, patterns, and insights collaboratively.
- **Report Writing and Editing**: All members equally participated in drafting, reviewing, and editing various sections of the report to ensure clarity, coherence, and accuracy.
- **Future Plans Formulation**: Worked together in brainstorming and formulating future plans for advancing healthcare data analysis and improving healthcare outcomes.
- **Group Coordination and Communication**: Maintained effective communication channels throughout the project, coordinating tasks, scheduling meetings, and ensuring smooth collaboration.

Overall, the project's success is attributed to the collective efforts, dedication, and teamwork of all three members, who contributed equally at every stage of the project.

Short Bio

1. **Kalp Shah** is a dynamic software engineer interested in machine learning, exploratory data analysis (EDA), data visualization, and software development. He earned his Bachelor's degree in Computer Science from Dhirubhai Ambani University and has gained extensive experience in web development, encompassing both front-end and back-end technologies. Leveraging his expertise in programming languages like Python, JavaScript, and Java. He excels in tackling complex challenges across software development, machine learning algorithms, and data visualization techniques.

In addition to his technical pursuits, Kalp is passionate about organizing events, participating in music competitions, and engaging in musical endeavours as a member of a music band. His effective communication skills and collaborative spirit make him an invaluable team player, with a proven track record of delivering high-quality solutions within tight deadlines.

Outside of work, Kalp finds joy in diverse interests, including hiking, volunteering at local coding workshops, and exploring his musical talents. He is deeply committed to inspiring the next generation of developers and fostering innovation within the tech community.

2. **Yash Garg** Hello, my name is Yash Garg. I am currently a third-year student at DAIICT pursuing a B.Tech in information and communication technology. During my studies, I gained knowledge in C++, OOPs, DBMS, SQL, Software Engineering, Exploratory Data Analysis. My current CGPA is 8.4 I am known for my problem-solving skills, leadership abilities, quick ability to learn. I am a good team player and I believe in continuous improvement and always strive to give my best. My hobbies and interests include Traveling, problem-solving, and Sports such as swimming, badminton and yoga. Thank you that's all about myself.

3. **Shreyans Jain** is a skilled software engineer specializing in web development and data science. He is pursuing a Bachelor's degree in Mathematics and Computing from Dhirubhai Ambani University and has extensive experience in both front-end and back-end development.

He is proficient in languages like Python, JavaScript, and Java, he excels in solving complex software development challenges.

Recognized for his collaborative nature and effective communication skills, Shreyans consistently delivers high-quality solutions within tight deadlines. Outside of work, photography, and watching movies.

References

[1] data.gov.in. *URL:* https://data.gov.in

[2] Prof. Gopinath Panda. EDA slides uploaded on Moodle.

[3] Rural Health Statistics Data.