**NATIONAL RURAL HEALTH MISSION**

A picture containing logo

Description automatically generated

IT 495 EXPLORATORY DATA ANALYSIS

SEMESTER II

INSTRUCTOR: DR. GOPINATH PANDA

|  |  |  |  |
| --- | --- | --- | --- |
| NAKUL TOMAR  202218026 | SHASHVA MACHCHHAR  202218028 | AYUSH JAIN  202218039 | HARSHIT SHAH  202218040 |

GROUP NO: 2Declaration

We, the undersigned, declare that this project titled **“National Rural Health Mission”** is our original work and that all information contained herein has been properly cited and referenced. We affirm that we have not engaged in any form of academic dishonesty, including but not limited to plagiarism or falsification of data. This project has not been submitted, in part or in whole, for any other academic purpose and all sources of information used in this project have been identified in the references section. We understand that any breach of academic integrity may result in severe Penalties, including revocation of our degrees.

Additionally, we confirm that each member of the group contributed equally to the completion of the project. Each member participated in this project's planning, research, analysis, and writing. We also communicated and collaborated effectively throughout the project, ensuring that each member's contributions were properly integrated.

|  |  |  |  |
| --- | --- | --- | --- |
| A picture containing plant  Description automatically generated  NAKUL TOMAR  202218026 | Text, letter  Description automatically generated  SHASHVA MACHCHHAR  202218028 | AYUSH JAIN  202218039 | Text  Description automatically generated  HARSHIT SHAH  202218040 |

Date: 02-May-2023

Certificate

This is to certify that the group project report titled **“National Rural Health Mission”** submitted by **Group II** members **NAKUL TOMAR, SHASHVA MACHCHHAR, AYUSH JAIN** and **HARSHIT SHAH** was original work and was completed under our supervision.

We confirm that this project report is the result of the original work of this group and that all sources of information users have been properly acknowledged with appropriate citations and references.

Every member of the team participated equally in the planning, research, analysis, and writing of this report and we recommend it for evaluation and examination with confidence as it meets the highest standards of academic excellence.

Signature of project supervisor

Date: 02-May-2023

# Contents

[List of Figures V](#_Toc133956016)

[Introduction: 1](#_Toc133956017)

[Tools and Technologies: 1](#_Toc133956018)

[Methodology: 2](#_Toc133956019)

[Exploratory Data Analysis (EDA): 2](#_Toc133956020)

[1. National Health Mission Budget/Funds: 2015-16 to 2022-23 Dataset : 2](#_Toc133956021)

[Applying regression model for data that seems to follow linear relationship: 14](#_Toc133956022)

[Regression on Release of Government of India's Fund and Total Expenditure Reported ..17](#_Toc133956023)

[2. Mobile Medical Units (MMUs) data from September 2016: 18](#_Toc133956024)

[3. Operational Ambulance Under NRHM: 20](#_Toc133956025)

[4. Established Special New-born Care Units (SNCU) under NRHM: 22](#_Toc133956026)

[5. Accredited Social Health Activists (ASHAs) Selected under NRHM: 25](#_Toc133956027)

[6. Health Centres Under NRHM Dataset: 27](#_Toc133956028)

[Conclusion: 3](#_Toc133956029)6

[References:](#_Toc133956030) 37

[For definitions:](#_Toc133956031) 37

[For data collection:](#_Toc133956032) 37

[For plots:](#_Toc133956033) 37

# List of Figures:

[1 Heatmap of null values in data frame 4](#_Toc133954137)

[2 Heatmap after Dropping Null Values 4](#_Toc133954138)

[3 Correlation Matrix of Features 5](#_Toc133954139)

[4 Scattered Pair Plot 6](#_Toc133954140)

[5 Stacked Bar Chart Of Budget Proposed And Budget Released over the years 7](#_Toc133954141)

[6 Bar Chart of Extent of Funds Utilised Against Budget Poposed 10](#_Toc133954142)

[7Bar Chart of Extent of Funds Utilised Against Budget Approved 10](#_Toc133954143)

[8 Bar Charts of Government Funds Released and Opening Balance over the years 11](#_Toc133954144)

[9 Bar Chart of Extent of budget approved over budget Proposed and Total Expenditure Over the years 11](#_Toc133954145)

[10 Bar Chart of Budget Approved and Budget Proposed By state 12](#_Toc133954146)

[11 Pie Charts Showing Budget proposed By every State Over the Years 12](#_Toc133954147)

[12 Scatter Plot to see if two features are linearly related 15](#_Toc133954148)

[13 Regression Line after Doing Predictions On Budget Proposed and Budget Approved 15](#_Toc133954149)

[14 Scatter Plot to see If Release of gvt Funds And Expenditure are Related 16](#_Toc133954150)

[15 Regression Line after finding linearity in data 17](#_Toc133954151)

[16 Top 10 State-wise Operational MMUs 19](#_Toc133954152)

[17 Top 10 State-wise Operational Ambulances 21](#_Toc133954153)

[18 Choropleth Map of State-wise Established SCNU in 2020 24](#_Toc133954154)

[19 Tree Map of State-wise Accreditted Social HEalth Activists in 2018 26](#_Toc133954155)

[20 Heatmap for checking null values 27](#_Toc133954156)

[21 Bar chart of Distribution of Health Centers 28](#_Toc133954157)

[22 Distribution Of Rural Health Centres Across Sates 29](#_Toc133954158)

[23 Distribution Of Urban Health Centers Across Sates 20](#_Toc133954159)

[24 Horizontal Bar Charts of Facilities State-wise 31](#_Toc133954160)

[25 Bubble Chart of States by Number of Facilities per state 35](#_Toc133954161)

Introduction:

The National Rural Health Mission (NRHM) was launched by the Hon’ble Prime Minister on 12th April 2005, to provide accessible, affordable, and quality health care to the rural population, especially the vulnerable groups. The Union Cabinet vide its decision dated 1st May 2013, has approved the launch of National Urban Health Mission (NUHM) as a Sub-mission of an over-arching National Health Mission (NHM), with National Rural Health Mission (NRHM) being the other Sub-mission of National Health Mission.

NRHM seeks to provide equitable, affordable, and quality health care to the rural population, especially the vulnerable groups. Under the NRHM, the Empowered Action Group (EAG) States, as well as the North-eastern States, Jammu and Kashmir, and Himachal Pradesh, have been given special focus. The thrust of the mission is on establishing a fully functional, community-owned, decentralized health delivery system with inter-sectoral convergence at all levels, to ensure simultaneous action on a wide range of determinants of health such as water, sanitation, education, nutrition, social and gender equality. Institutional integration within the fragmented health sector was expected to provide a focus on outcomes, measured against Indian Public Health Standards for all health facilities.

In the case of the National Rural Health Mission (NRHM), EDA can help us to gain insights into the health infrastructure and services in rural areas and the impact of the mission on the population. It can also provide insights into the demographic and socio-economic characteristics of the rural population and their impact on health outcomes. In this project we have tried to perform EDA on NRHM for this we have taken from official government website [data.gov.in](https://data.gov.in/) and perform EDA on various sectors of health sector which are part of NHRM.

Tools and Technologies:

For our analysis we have used Python and following python libraries:

* Pandas
* Matplotlib
* Seaborn
* NumPy
* Plotly Express
* Ipywidgets
* Scikit Learn
* JSON

Methodology:

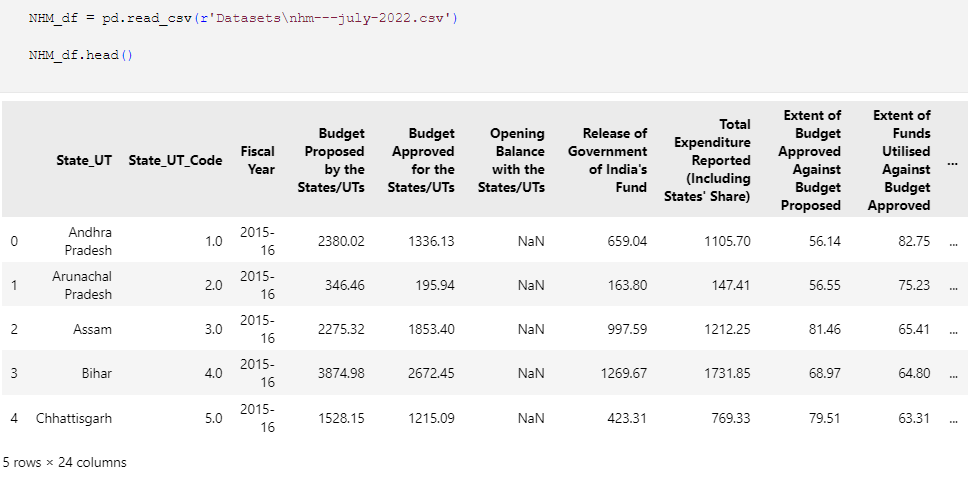
For our analysis we couldn’t find a single dataset large enough to perform EDA therefore we decided to perform EDA on several small datasets. We collected datasets from [data.gov.in](https://data.gov.in/) website. Then we followed step by step procedure for each data set:

* Collect and load the data.
* Data Description.
* Data Cleaning.
* Data Pre-processing.
* Summarizing the Data.
* Visualization.
* Predictions (if possible).
* Interpretations

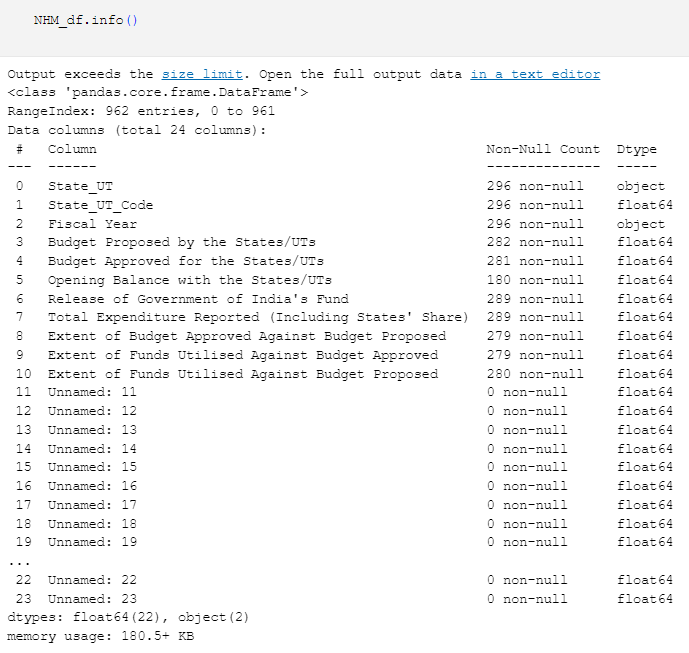
Exploratory Data Analysis (EDA):

1. National Health Mission Budget/Funds: 2015-16 to 2022-23 Dataset :

We start by loading the csv file.



Then, to get insights about the data we used .info() method which gives information about total number of rows and columns, name of each column, and datatype of each column.



From the above output, we see that our dataset contains 962 rows and 24 columns, but we also see that there are many columns which have only null values. Therefore, we decided to drop these columns.

A picture containing text

Description automatically generated

Now, to visualize how many null values are there in each row we plot a heatmap.

A picture containing graphical user interface

Description automatically generated

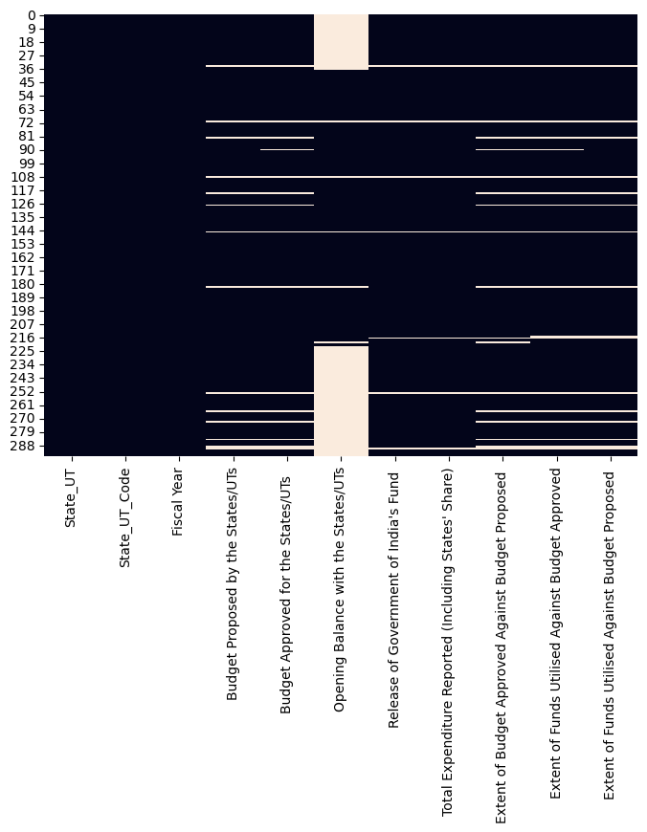
A picture containing table

Description automatically generated

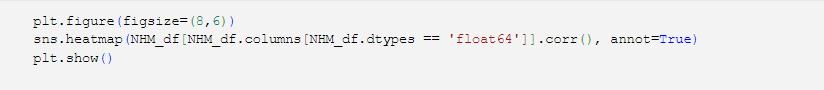
Since there are null values in all the columns, we can discard those values without any data loss. Opening Balance with the States/UTs is not available for the fiscal year 2015-16 , 2021-22 and 2022-23 but values for other years between these are available, thus we skip dropping this Null values.

Text

Description automatically generated with low confidence



After removing the null values from the rows and columns we try find the correlation between all the columns.

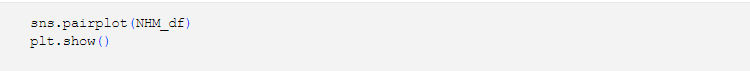


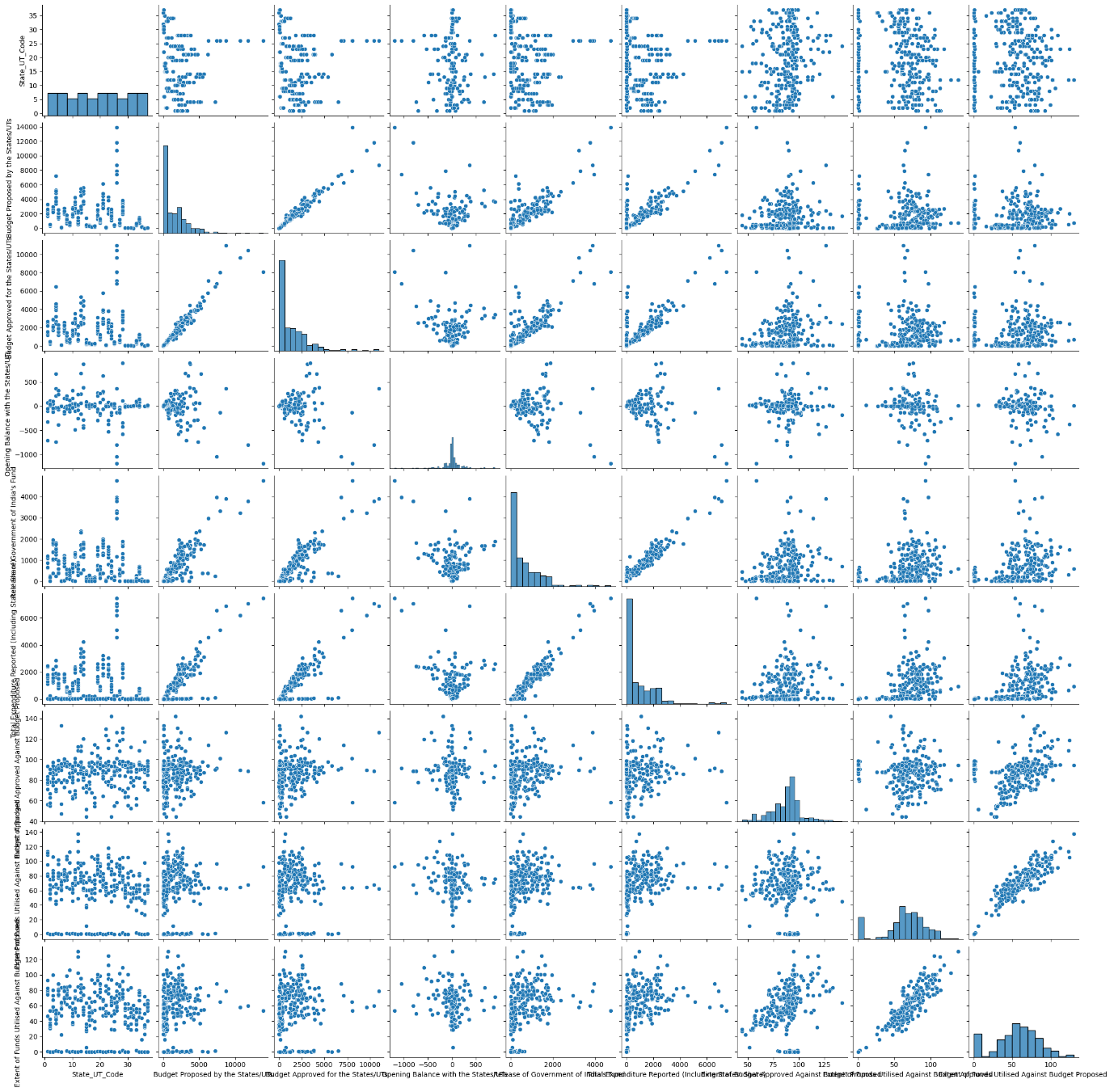
Timeline

Description automatically generated with low confidence

Correlation heatmap between the features show that how each column is related to each other. From the graph it is evident that features like Budget Proposed and Budget Allocated are related to each other. Similarly Release of government of India fund is highly correlated to the expenditure.

Also, Extent of funds utilised against Budget approved is highly related to Extent of funds utilised against Budget Proposed.





Scatter Pair plot gives an idea about each features relationship with other features.

Trend that is followed by a feature by change in value other features are captured through it.

We have made this just to find relationship between features so that we can further apply Linear Regression Model to the ones that share linear relationship.

We plotted some stacked bar chart using Plotly library to visualize how much budget was approved and how much was utilized in different years. Below is the code for doing the same:



Chart, bar chart

Description automatically generated

Chart, bar chart

Description automatically generatedChart, bar chart

Description automatically generatedChart, bar chart

Description automatically generated

Chart, bar chart

Description automatically generated

Chart, bar chart

Description automatically generated

Chart, bar chart

Description automatically generated

Chart, bar chart

Description automatically generated

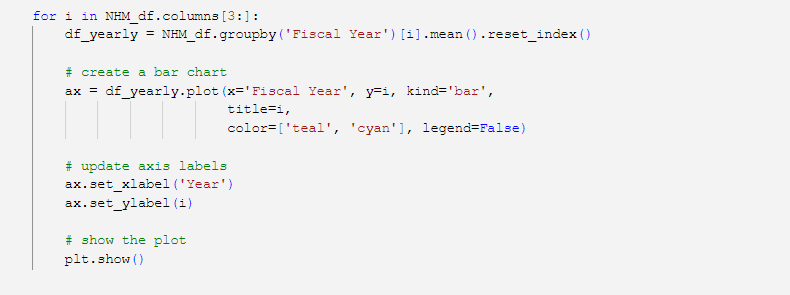
From the chart we can see that the amount of funding that states have spent on NRHM has generally increased over the years. In some years, the amount of funds used by certain states was significantly higher, while in other years the increased funds were distributed more evenly among the different states. We can also observe that some states consistently use more funds for NRHM than others, such as Uttar Pradesh, which has the highest threshold in most years. In most states, the total budget proposed/allocated by the government increases year by year. This can be seen in the upward movement of the bars.

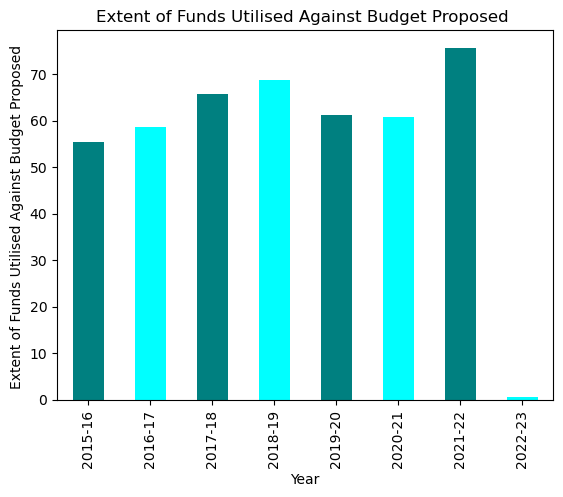
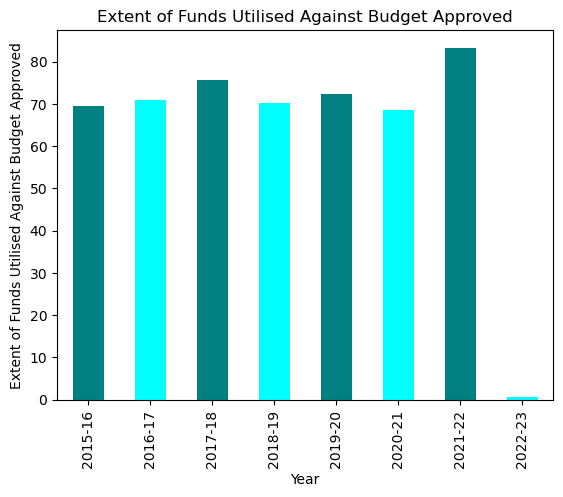
Over the years, Uttar Pradesh, Maharashtra, Karnataka, and Gujarat have had the highest combined budgets. Sikkim, Nagaland, and Mizoram have had the lowest combined budgets for many years. Many state governments have proposed substantial increases in total budgets for 2019-20. This may be due to various factors, such as the implementation of new policies or the allocation of funds for specific projects.

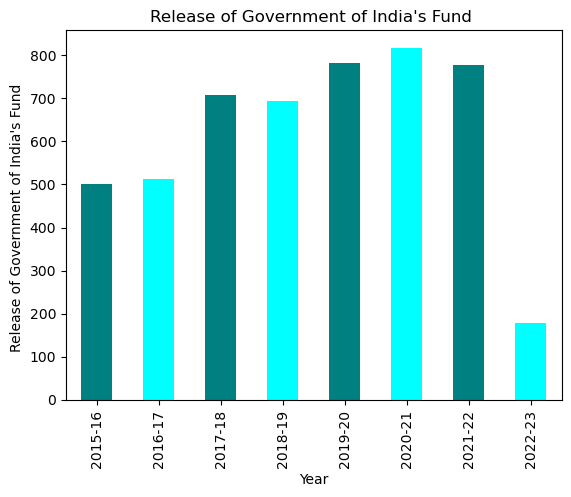
The total budget proposed by Andhra Pradesh from 2016-17 to 2018-19 has been drastically reduced. This could be due to factors such as a change of government or a change in policy.

The bar graph shows that the total budget proposed by the government has generally increased over the years, although there has been some fluctuation and variation between states and years.

Now to find the change in different stages of Budget such as Proposal, approving etc. we will plot the moving average based on different years.



After the year 2016-17, there was a significant spike in funding which continued to increase in the following years as well. However, there seems to be a slight dip in funding after the year 2020-21. Overall, the trend seems to be positive and indicative of the government's increasing focus on improving rural health infrastructure.

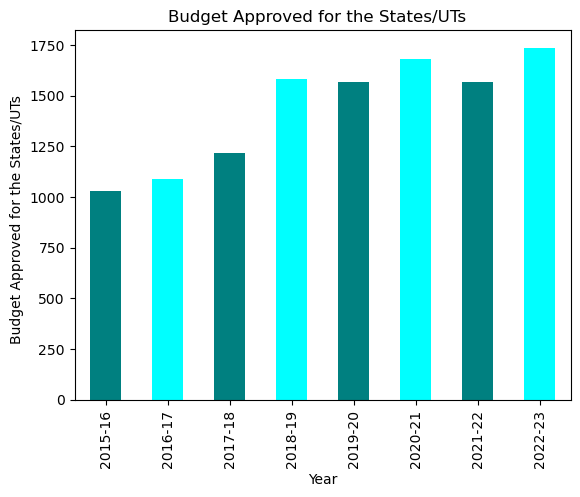
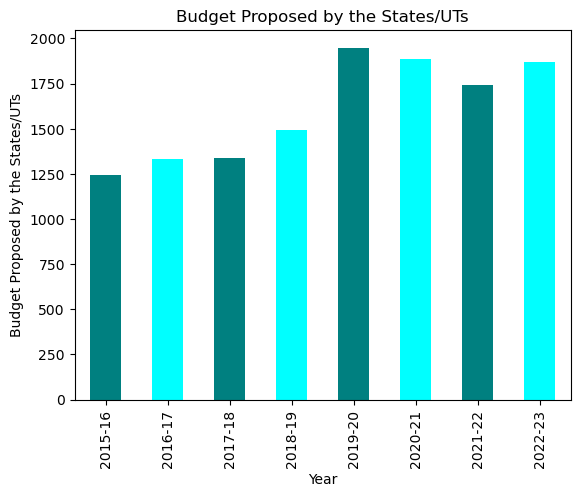
The state government often began the year with a budget deficit, as they received less funding than what they had proposed for in most years.

Chart, bar chart, histogram

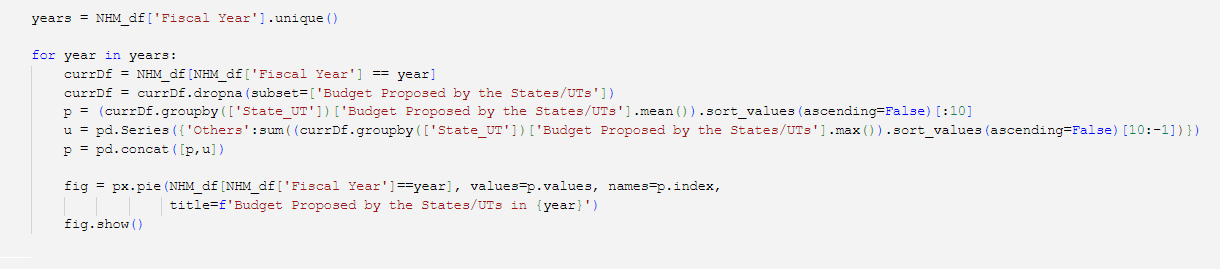
Description automatically generated Chart, bar chart

Description automatically generated

Total expenditure by the government is increasing on yearly basis as they are focusing more on health sector.

To find out which states/UTs have the most share in the Budget throughout the years we used pie chart for visualization.



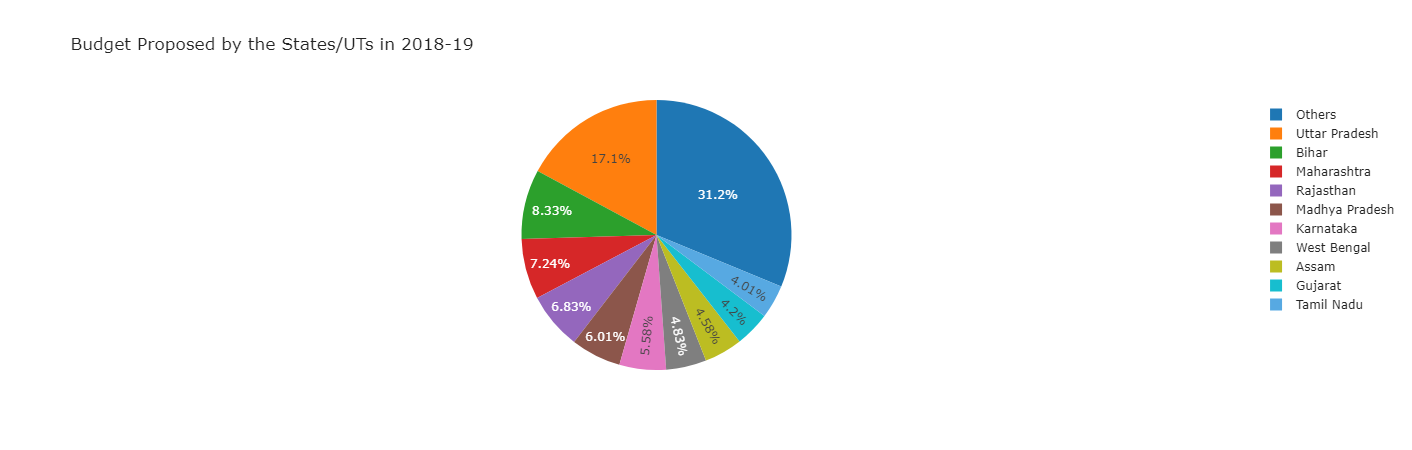
Chart, pie chart

Description automatically generated

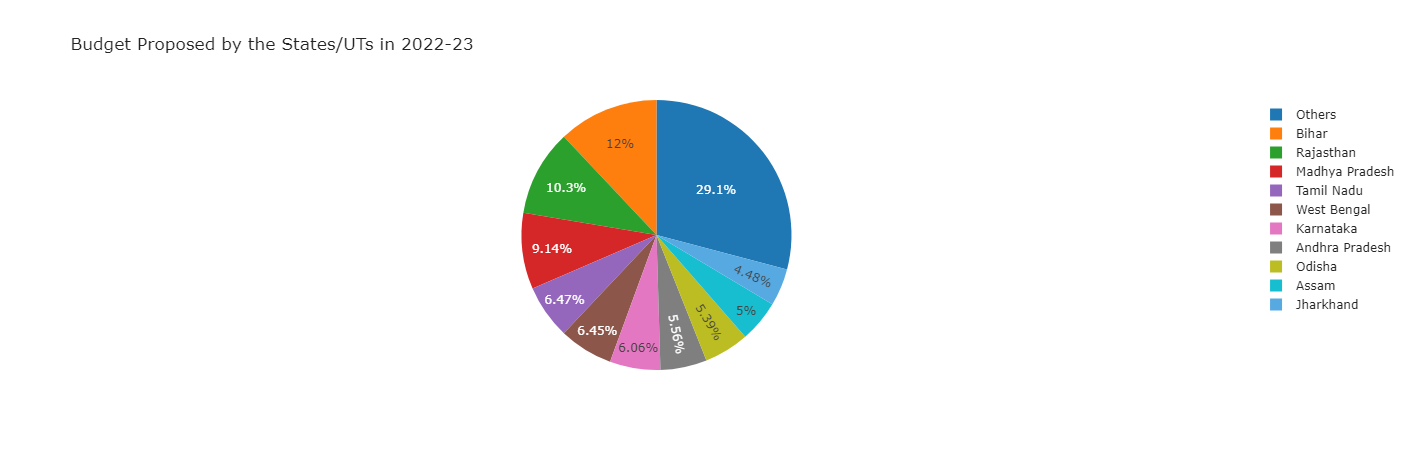
Chart, pie chart

Description automatically generatedChart, pie chart

Description automatically generated

Chart, pie chart

Description automatically generatedChart, pie chart

Description automatically generated

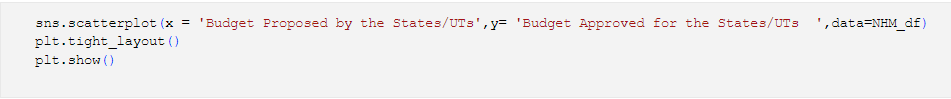
A review of the budgets put up by several states over the years reveals some intriguing trends. It appears that only a few states—Uttar Pradesh, Bihar, Rajasthan, Maharashtra, Karnataka, Assam, Gujarat, Madhya Pradesh, and West Bengal—propose over 70% of the nation's total budget. This implies that the states have more influence over how funds is distributed at the federal level.

Additionally, it can be seen that Uttar Pradesh (UP) ranked among the top states for budget proposals across all years considered. This demonstrates that UP has fought for a bigger chunk of the state budget.

The remaining states account for another 30% of the total share and play a crucial role in the allocation of funds. The distribution of this budget proposal across states is critical because it allows us to understand each state's relative priorities, including their focus on the health sector. By analyzing trends in budget proposals, we can also gain insight into which states are investing more in the health sector and which states may need more attention in this area.

Applying regression model for data that seems to follow linear relationship:

Regression on Budget Proposed by the States/UTs and Budget Approved for the States/UTs

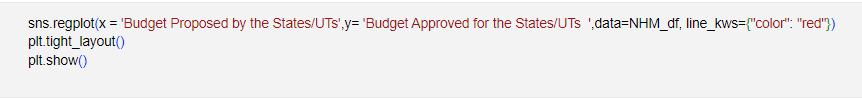


Chart, scatter chart

Description automatically generated

Graphical user interface, text, application

Description automatically generated



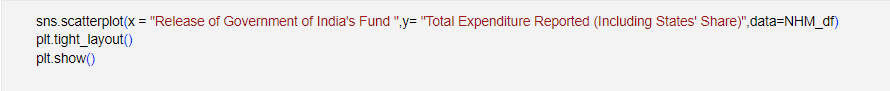
Chart, scatter chart

Description automatically generated

Looking at the scatterplots it could be seen that Budget Proposed by the States/UTs and Budget Approved for the States/UTs has some kind of linear relationship thus we modelled it with linear regression and found that we can predict the Expenditure based on Release of funds.

Our model had R2 score of 0.98 and it did great job predicting the budget approved.

Regression on Release of Government of India's Fund and Total Expenditure Reported



Chart, scatter chart

Description automatically generated

Graphical user interface, text

Description automatically generated



Chart, scatter chart

Description automatically generated

Looking at the scatterplots it could be seen that Release of Government of India's Fund and Total Expenditure Reported has some kind of linear relationship thus we modelled it with linear regression and found that we can predict the Expenditure based on Release of funds.

Our model had R2 score of 0.96 and it did great job predicting the budget approved.

2. Mobile Medical Units (MMUs) data from September 2016:

This dataset tells us about the operational Mobile Medical Units (MMUs) in each state in year 2016. So, first we load our dataset and apply .info() method to initial insights.

Graphical user interface, text, application, email

Description automatically generated

We observed that dataframe has 37 row and only 3 columns and there are no null values. Since there is no need for S1. No. column we will drop it.

Graphical user interface, text

Description automatically generated with medium confidence

Table

Description automatically generated

For ease of analysing further we renamed the columns and dropped Total row form the df. After that we ranked the states based on the number of operational MMUs and created a separate df for top 10 states.

Graphical user interface, text, application

Description automatically generated

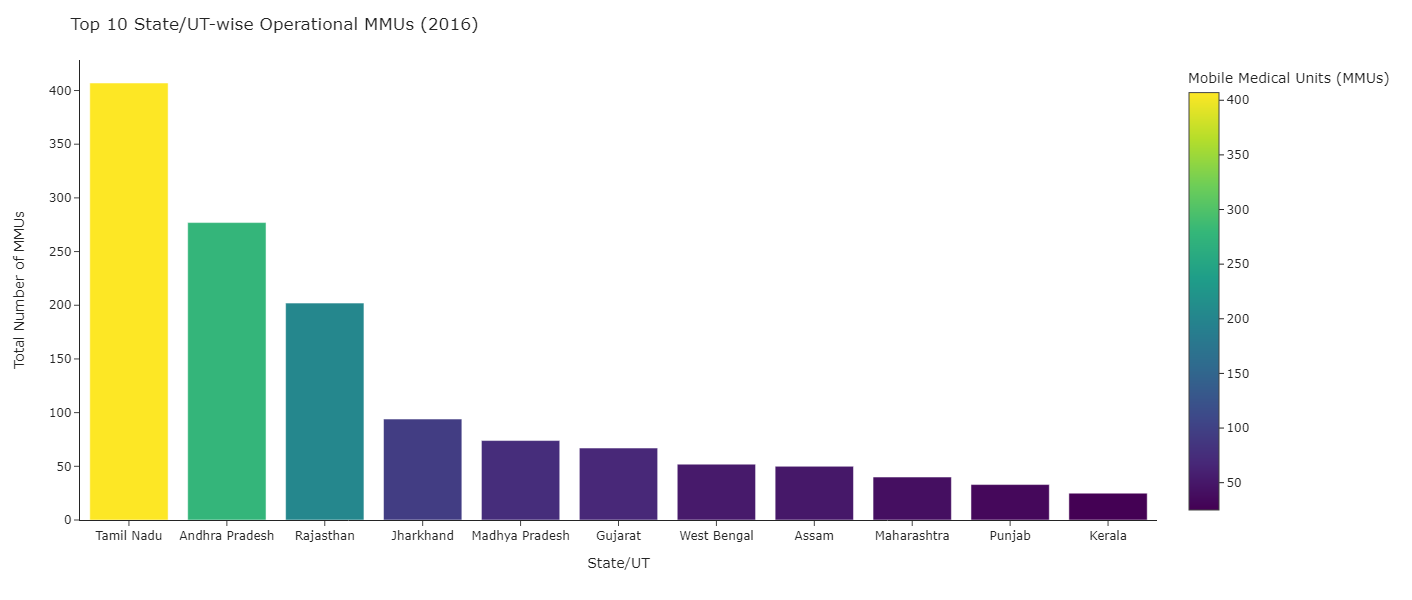
Table

Description automatically generated

To plot the Top 10 States, we used simple bar chart.

Graphical user interface, text, application

Description automatically generated



From the above plot, we observed that Tamil Nadu has the maximum number of operational MMUs (407) according to the data which then followed by Andhra Pradesh and Rajasthan which have 277 and 202 operational units.3. Operational Ambulances Under NRHM:

This dataset has the information about operational ambulances under NHM during 2018. Load the data and apply .info().

Text

Description automatically generated

Dataframe contains 37 rows and 6 columns, drop unnecessary rows. And rank the states based on total number of ambulances.

Graphical user interface, text, application, email

Description automatically generated

Table

Description automatically generated

We used stacked barplot to represent diversity in type of ambulances Top 10 states posess.



Chart, bar chart

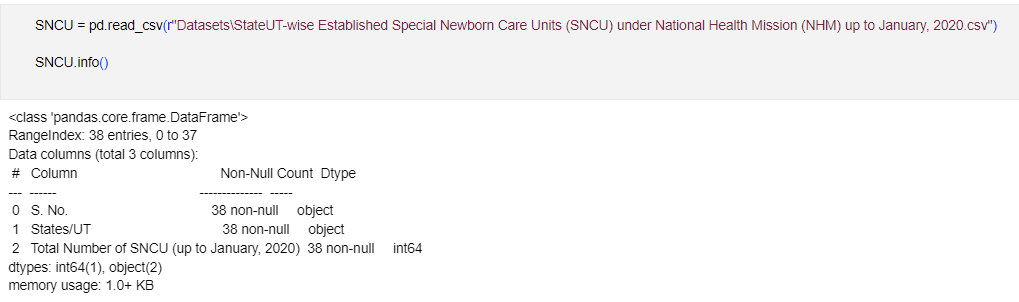
Description automatically generated

From the above plot, we can deduce that Uttar Pradesh has the greatest number of Operational Ambulances.

Also UP has the greatest number of Dial 108 ambulances whereas Maharashtra and West Bengal has the greatest number of Dial 102/104 and other ambulances respectively.

4. Established Special New-born Care Units (SNCU) under NRHM:

SNCU dataset tells us about the established Special New-born Care Units in a State/UT up to January 2020. Load the data and apply .info() method.



SNCU df has 38 rows and 3 columns, then we dropped unwanted columns and rows.

Text

Description automatically generated

A picture containing graphical user interface

Description automatically generated

Graphical user interface, text, application

Description automatically generated

For visualization part, we used Plotly library to plot Map of India and shade states based on the number of SNCUs.

So, for plotting the map we need a GeoJSON file which contains boundary of each state. After loading the JSON file we created a dictionary with state id and state name as key value pairs.

Text

Description automatically generated

There was difference in spellings in the state name, so we altered the names for solving the problem. And added state id column in our SNCU df.

A picture containing text

Description automatically generated

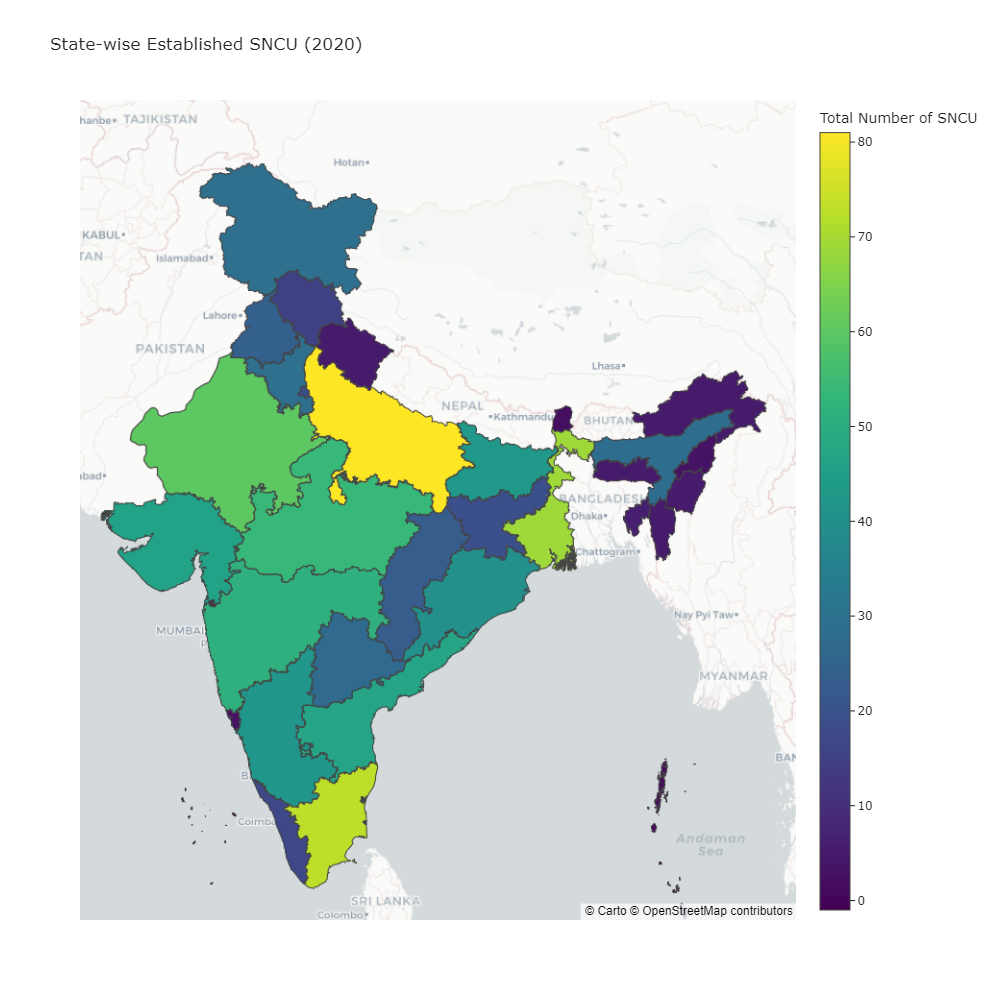
Table

Description automatically generated

Now, to plot the map we used choropleth\_mapbox() function from Plotly express, and changed the color based on the total number of SNCU.

Graphical user interface, text, application

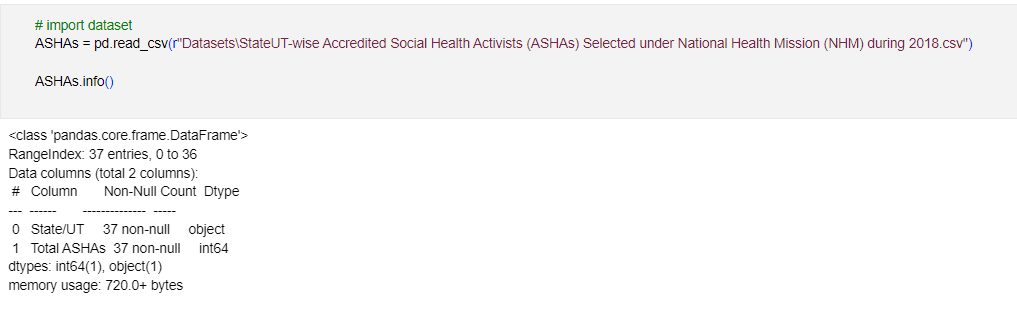
Description automatically generated



From the above map we can interpret that UP has the greatest number of established SNCU while north-eastern states have very a smaller number of SNCU. States like MP, Gujarat, Rajasthan, Maharashtra etc are in between.

5. Accredited Social Health Activists (ASHAs) Selected under NRHM:

This dataset tells us about the number of accredited Social health activists during 2008.



Dataframe has 37 rows and 2 columns. We removed the unwanted row which contains total of all the states.

Graphical user interface, application, Word

Description automatically generated

Table

Description automatically generated

For visualization part, we used tree-map which easily shows which state/UT rank higher and which state/UT ranks lower.

Graphical user interface, text, application, Word

Description automatically generated

Chart, treemap chart

Description automatically generated

From the above Tree-map, we can see that again UP has the most number of ASHAs while Bihar and MP take second and third place.

6. Health Centres Under NRHM Dataset:

Health Centres dataset gives insight about several characteristics of health centres throughout the country. We tried to show these characteristic as follows:

Graphical user interface, text, application, email

Description automatically generated

Dataset has 200438 rows and 11 columns and datatype of all the columns is set to object. To visualize null values in each row we plot a heatmap.

Graphical user interface, text

Description automatically generated

Chart, bar chart

Description automatically generated

From the heatmap we can saw that Facility Address column has a lot of null values, therefore we dropped that column.

A picture containing chart

Description automatically generated

Graphical user interface

Description automatically generated with medium confidence

Then we typecasted latitude and longitude columns into numeric values from objects.

Text

Description automatically generated with medium confidence

After pre-processing our data, we started plotting the graphs. The first plot tells us about the number of Rural vs Urban healthcare centres using a countplot.

Company name

Description automatically generated with medium confidence

Chart, bar chart

Description automatically generated

We can clearly see that Rural healthcare centres are much more in numbers than Urban which tells us government focused on Rural Healthcare more.

Then to find out which state has most number of rural HCs we used countplot.

A picture containing text

Description automatically generated

Chart, bar chart

Description automatically generated

Uttar Pradesh, Bihar, Rajasthan, Maharashtra, Karnataka, Assam, Gujarat, Madhya Pradesh, and West Bengal are the leading proposers of the rural health sector budget in the country . This suggests that these states have a significant influence on the allocation of funds at the national level, which can play a crucial role in their rural health management. By having a greater say in the allocation of funds, these states can prioritize their specific health needs and ensure necessary resources for their rural populations. This can lead to better health outcomes, improved access to healthcare and ultimately a healthier population.

By having a greater say in the allocation of funds, these states can prioritize areas that need the most attention and address specific health challenges facing rural populations. This may include improving health care infrastructure, providing essential medicines, promoting preventive health care measures, and raising health care awareness among rural populations.

An increased focus on rural health management could significantly improve health outcomes in these states. It can also help reduce the burden of preventable disease, increase life expectancy, and improve the quality of life for millions of people in rural areas.

To find how many urban HCs are there in each State/UT we plotted countplot.

A picture containing text

Description automatically generated

Chart, bar chart

Description automatically generated

The top states that seem to care more about urban health than rural health are Bihar, Maharashtra, Tamil Nadu, Delhi, Uttar Pradesh, and Karnataka. Numerous factors, such as the greater concentration of infrastructure and resources in urban areas, as well as the higher population densities there, could account for this.

The fact that rural inhabitants frequently have poor access to healthcare and are more susceptible to a range of health issues makes it important to note that neglecting rural health can have major repercussions. A country's current health inequalities across its various areas and inhabitants can also be made worse by concentrating on urban health at the expense of rural health.

Therefore, states and governments must give rural and urban health care systems top priority and funding to guarantee that everyone has access to high-quality healthcare, regardless of where they live or their socioeconomic status. This can be accomplished in several ways, including by boosting funding for rural health programmes, enhancing the facilities and resources available there, and encouraging medical personnel to practise in rural regions.

Among rural and urban HCs there are different types of facilities dedicated for specific purposes. So, we plotted multiple count-plot to get State/UT-wise count of different type of facilities.

A picture containing text

Description automatically generated

Chart, bar chart

Description automatically generated

UP and Rajasthan has a greater number of Community Health Centres. Maybe their geographical size is the reason for this.

Chart, bar chart

Description automatically generated

Most number of district hospitals are in UP, since UP has highest number of districts, but other states like MP, which has more districts than Maharashtra, has lower number of District hospitals.

Chart, bar chart

Description automatically generated

UP once again has highest number of primary health centres as well.

Chart, bar chart

Description automatically generated

Chart, bar chart

Description automatically generated

From the above plots we can conclude that State of UP has highest number of Health Centres in each category. While North-Eastern States have lower number.

To summarize, the total number of State/UT Healthcare Centres we plotted a map. For that, we grouped rows based on the name of State/UT. Then we created a new df which contains pin-point location of each state using latitude and longitude.

A picture containing chart

Description automatically generated

Table

Description automatically generated

Join both the grouped df and new longLatdata into one dataframe.

Text

Description automatically generated with low confidence

*Table

Description automatically generated*

We used scatter\_mapbox() function from plotly express library to plot a scatter plot on a map. In this plot we used the number of facilities column to vary the size of bubble and latitude and longitude columns to assign position to each bubble based on their state/UT name.

Text

Description automatically generated

Map

Description automatically generated

Conclusion:

In conclusion, this report has provided an overview on the National Rural Health Mission (NRHM) and National Health Mission (NHM) by highlighting different aspects in which government has worked. From our analysis we found out that significant amount of Funds has been allocated throughout the years for improving healthcare facilities in rural as well as urban areas. We also observed that, the amount of funds used by certain states was significantly higher, while in other years the increased funds were distributed more evenly among the different states. We can also observe that some states consistently use more funds for NRHM than others, such as Uttar Pradesh, which has the highest threshold in most years.

Other aspects such as number of Ambulances, MMUs, SNCUs and ASHAs were also explored in this report. We also analyzed different types of healthcare facilities and their distribution throughout the nation.

In all these categories UP has been consistently on the Top of the list while north-eastern states were bottom of the list. So, we can conclude that UP has been the prime beneficiary from the NRHM & NHM.

References:

For definitions:

<https://nhm.gov.in/index1.php?lang=1&level=1&lid=49&sublinkid=969>

For data collection:

<https://data.gov.in/>

<https://openbudgetsindia.org/dataset/national-health-mission>

For plots:

<https://plotly.com/python/>

<https://seaborn.pydata.org/examples/index.html>

<https://matplotlib.org/stable/gallery/color/named_colors.html>

Text, table

Description automatically generated

Text

Description automatically generatedTable

Description automatically generated

Table

Description automatically generated Text, table, letter

Description automatically generated