A Python Notebook for Analysis on Medical Data

K. H. Mallikarjuna Rao[1], G. Vishnu Vardhan[2], B. Siri Chandana[3], Dr. S. Sai Kumar[4]

B.Tech Students, Department of Information Technology[1,2,3]

Assistant Professor, Department of Information Technology[4]

Prasad V. Potluri Siddhartha Institute of Technology, Vijayawada, Andhra Pradesh, India

Abstract: *This python notebook uses the libraries present to analyze the medical data and form various data diagrams that help in the better visualization of the data which helps in better visualization that leads to better decision making. This data is primarily about availability of beds and hospitals in rural and urban areas across India. It is useful tool for the government to know about the facilities and increment them based on the data during the recent pandemic time. It will come in handy in the future also during any disasters, pandemics etc. This application visualizes different data in different forms such as pie charts, bar graphs and geographical visualization with data. It can display the sorted values and can show the state or UT that is first in the provision of the medical facilities to the people. The notebook can be used to track the major diseases and there expansion among the people in the particular disease prone areas. This application may also be used to expand beyond the limitation of beds and hospitals to medicines, equipment, specialist doctors and other facilities in the hospitals.*

Keywords: Medical Data, Data Visualization, numpy, pandas, sklearn, plotly libraries.

I. INTRODUCTION

We generally visualize or analyze the data in excel sheets or Google sheets in the past. The data is given in the excel format and applied some filters and pivoting to get the required data. It is then displayed as graphs and charts using the features present in the excel software. This worked well until the data has become the important aspect to be analyzed deeply and thoroughly to be studied to take future decisions of businesses and many organizations. After the python was released and libraries that use machine learning to analyze the data and predict the future data are introduces the data analysis and visualization took a huge leap into the future.

II. PROPOSED SYSTEM

The idea is to analyze the data using the python data interpretation libraries.This interpretation provides us better visualization of data. The application uses the numpy, matplot and sklearn libraries to implement various types of data visualization. It uses the folium library to geographical locate the data on the map. It uses the latitude and longitude of the medical facility to represent it in the map. It uses the bar graphs to represent the number of hospitals present in a state. Also, it processes that data to obtain the state having most number of medical facilities in the country. The data analysis separates the data on the number of hospitals into types of different hospitals such as primary, community and district hospitals. It also shows the number of beds present in the rural and urban areas in a bar chart. It can also show the top 10 hospitals containing the facilities and number of beds in the order required. It also displays different state data in different colours to make the data visualization easy and understandable.

III. TECHNOLOGIES USED

3.1 Python Notebook

Python notebook is used for performing the data analysis on the data. The python online notebook is used in the Kaggle notebook. It is used for exploring and running machine learning code. It is a cloud computational environment that enables reproductive and collaborative analysis. We can use either python or R to do the data analysis in the Kaggle notebook. The Kaggle has scripts and notebooks to run the code. We prefer the notebooks to perform the data analysis as it can be run on other online tools. In notebooks we provide the datasets in many formats as excel or csv which the code takes as input for visualization. The notebook and editor are easy at user interface and works faster than other tools in processing the data. We can either use the already available datasets on Kaggle or prepare our own data and upload it

into the project. We use python and machine learning algorithms to perform data analysis in the notebook.

3.2 Pyhton Libraries

There are several libraries used in the data visualization in this application. The numpy is used for array operations and matrices. Pandas is used widely used for data analysis and ML tasks. Seaborn is data visualization with matplot library. Plotly supports over 40 unique chart types for statistical analysis. Folium is a library that helps in creating and visualizing the data geographically on the several type of leaflet maps. Fbprophet library is for forecasting time series data, it generally delas with the data that is time effective. Statsmodel package explores data, statistical models and perform statistical analysis.

3.3 Data

The data is taken in the form of a single excel sheet or csv table. Then the data can be divided to various different tables to be analyzed in different aspects such as rural, urban data and number of private and public hospitals.

3.4 Dashboard

A dashboard is a collection of several notebooks and supporting information shown in a single place so you can compare and monitor a variety of data simultaneously. For example, you may have different versions of the code and application, they can be saved and viewed in the version info in the dashboard. Rather than versioning the single notebook, you can create another notebook to visualize advanced data also using the new notebook option in the dashboard.

3.5 Server

The Kaggle software has the online tools for the computing and storage of the data for analysis in the Kaggle notebook. CPU, RAM are displayed in the dashboard to show us the amount of usage by the application/program. It also keeps track of the session time and number of times each code is ran individually. The data is uploaded in the server of the Kaggle to make it accessible to the code in the notebook.

3.6 MAPS

We use the FOLIUM library for the visualization of the geographical data on the maps of different areas using the latitude and longitude data of the hospitals in the data.

IV. RESULTS

Fig: 1.1 Geographical pointing using latitude and longitude data – folium library – hospital locations.

The diagram depicts the locations of the hmis hospitals in the world map. The folium library uses the latitude and longitude information present in the dataset to point them on the world map. As, we study only data about India the dots are present only in India in fig: 1.1.
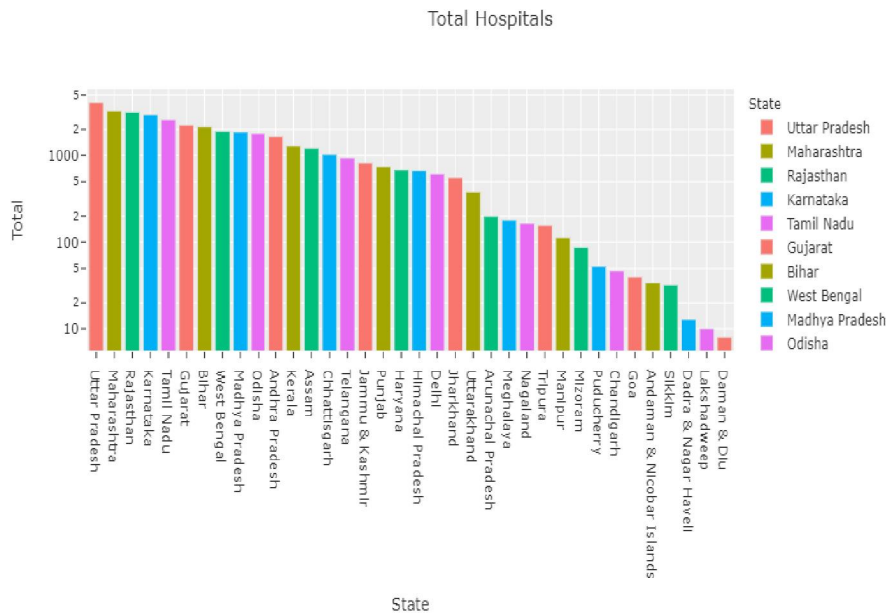

Fig: 1.2 Total number of hospitals in bar graph

Bar graph shows the different states in different colors and the number of hospitals in the state and UT. It shows them in a sorted order from highest number to the lowest number of hospitals.
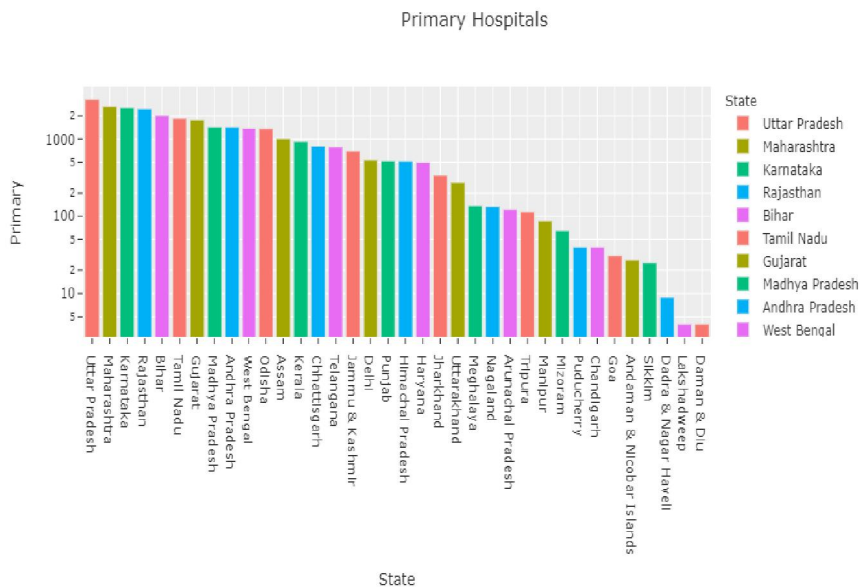

Fig: 1.3 Total number of primary hospitals

Bar chart above shows only the primary hospitals in the states and UT's in highest number to lowest number of hospitals. Also the same type of charts are obtained for the number of community hospitals, district hospitals and sub-district hospitals state wise and UT.
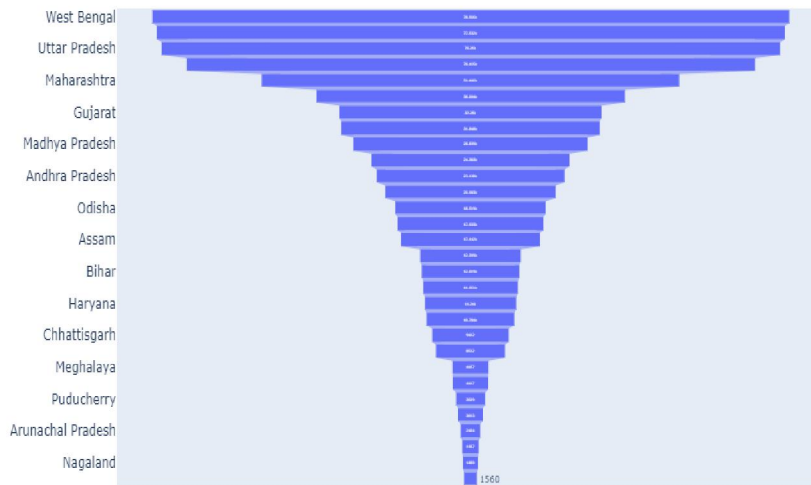
Fig: 1.4 Total number of beds state wise

This is a different type of bar chart that shows the total number of beds for patients that are available in different states and union territories in decreasing order. It is a funnel type of chart that looks like a funnel when displayed. It has the number of beds on the x-axis and the state name in y-axis.
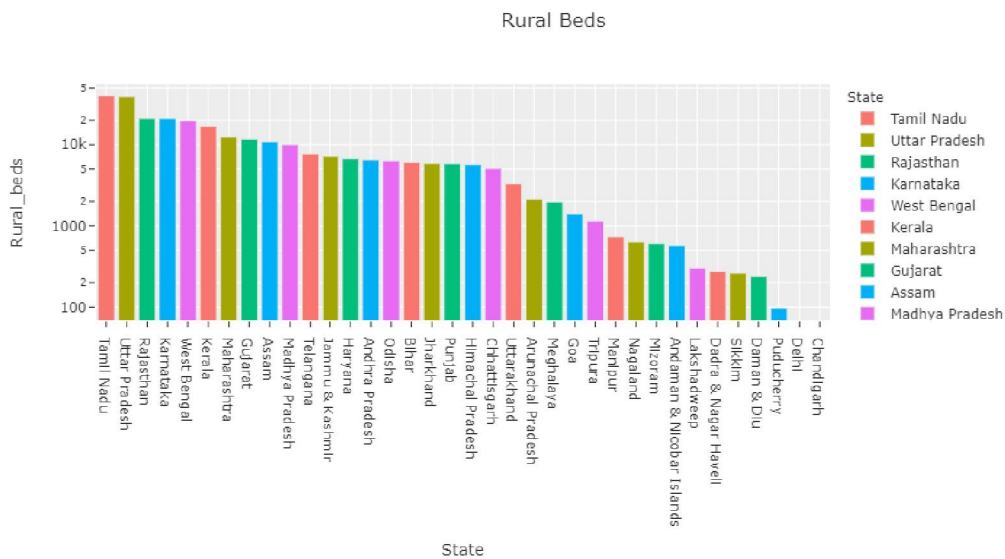

Fig: 1.5 Total number of beds in rural areas

The figure:4.1 describes the bar graph of the number of beds present in the rural hospitals in the rural areas in the different states and union territories. It also shows the number of hospitals present in the rural areas also in another bar chart.
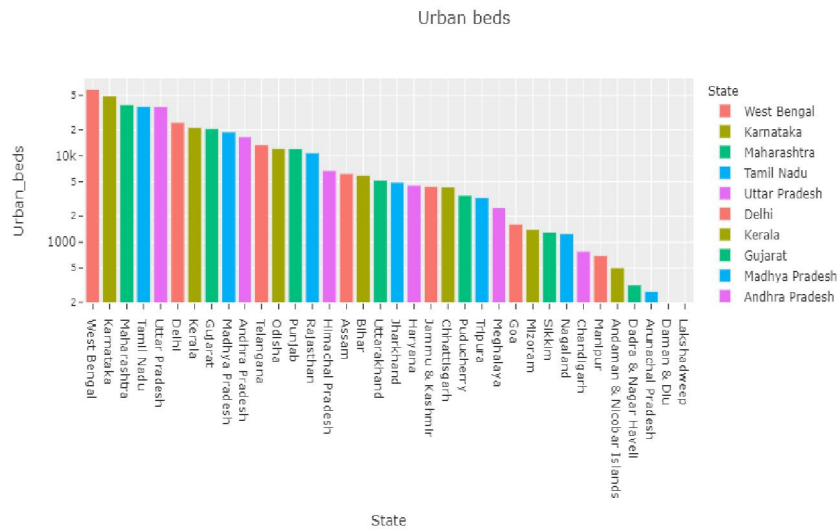
Fig: 1.6 Total number of urban hospitals – state wise

The bar graph depicts the number of hospitals present in the urban areas of different states and union territories in different colors. It also can display the number of beds present in the urban areas depending on the data in the same bar chart format.

V. SCOPE OF FUTURE USE

The application can be developed further to include with medicine data and other equipment facilities present in the hospitals. Also, the data can be analyzed in better format, if there been introduced better data analysis software or programs. The future use of this application mat exists if there is any type of pandemic or disaster that needs to be analyzed. Also, we can extend this concept to be used in the other concepts of disease spreading in an area or unemployment state wise also for other major concepts.

VI. CONCLUSION

Therefore, we have visualized the data i.e. in various scenarios such as urban, rural, public, community, district and sub-district hospitals. It also has future enhancements like, we discussed above. Finally this data helps government in building of new hospitals or introducing new equipment at the required area which increases the efficiency of providing medical support to every citizen.

REFERENCES

[1]. Datasets - https://www.kaggle.com/datasets/vishesh20/hospitalbedloc
[2]. https://www.kaggle.com/learn
[3]. Notebook - https://www.kaggle.com/docs/notebooks
[4]. Notebook link - https://www.kaggle.com/code/khmallikarjunarao/project-0/notebook