Development Of An Integrated Dashboard For Public Health Systems

A Research and Development project

by

Hrithik Mhatre (210040092)

Under the Supervision of

Dr. Raghavendran Lakshminarayanan

Koita Centre for Digital Health

INDIAN INSTITUTE OF TECHNOLOGY BOMBAY

Mumbai - 400076, India

May, 2024

Contents

Chapter 1

Introduction

The Koita Centre for Digital Health, IITB and the Public Health Department of Maharashtra propose to develop an integrated dashboard for public health systems that consolidates and visualizes data from multiple sources.

We initiated our analysis by exploring various data analysis tools and websites to source relevant data. Subsequently, we leveraged several online datasets to conduct preliminary data analysis. Later, we acquired district-specific data from hospitals in Maharashtra and proceeded to analyze it.

We took a close look at mental health data from every district in Maharashtra to see how COVID-19 has affected people's mental well-being. We carefully analyzed the information to understand exactly how the pandemic has changed the mental health situation across the state. Come along with us as we explore the details in the data and try to show just how much COVID-19 has impacted mental health in Maharashtra.

Chapter 2

Researching various data analysis software and health-related websites to gather data.

Software	Description of Data	How to Collect
Mera Aspatal	You can view all your recent visits to the government hospitals for the last 7 days. The Hospital name, patient name* and the date of visit to the hospital can be viewed [Not able to receive OTP to login] How to operate	-
RCH Portal	RCH and ABHA dashboards can be accessed [Login required to view reports and to perform data entry]	Data scraping
Biometric Attendance System	Registered organization can be viewed [Login required for employee specific data]	-
Nikshay	[Login required]	-
District Health Information System (DHIS2)	Not a Data source DHIS2 is a free and open-source software platform for the collection, reporting, analysis and dissemination of aggregate and individual-level data	-
Maharashtra Emergency Medical System (108 call centre)	Ambulance Data	Data scraping
SNCU	Not able to find the website	-
Bio medical equipments dashboard	Not a Data source	-

Chapter 3

Researching publicly available datasets online.

Due to the unavailability of data for Maharashtra districts until the end of February, we utilized online datasets and experimented with various data analysis tools for our analysis.

3.1 Health Management Information System data Analysis

3.1.1 Dataset

The HIMS dataset primarily encompasses data related to pregnancy and associated factors.

3.1.2 Data Preprocessing

1. The dataset had several null values across different columns, necessitating the removal of those rows
2. Duplicate values were also eliminated from the dataset
3. Given our focus on health data, any non-health-related statistics were excluded
4. Initially comprising 206 columns, the cleaned dataset was narrowed down to 129 columns

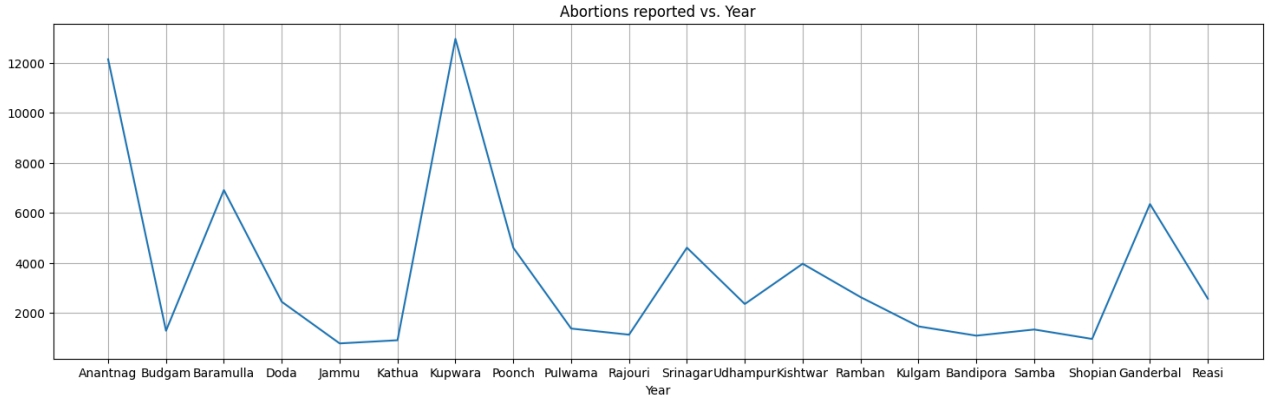
3.1.3 Plotting data specific to a district

This preliminary analysis was conducted to lay the groundwork for upcoming tasks. As we anticipate receiving data for the state of Maharashtra, it will be imperative to generate statistical plots for each district.

3.1.3.1 Plotting data for Jammu and Kashmir

A fundamental approach was employed to plot district-specific data by selecting the Jammu and Kashmir state from the dataframe and utilizing the df.groupby() function.

Example : Abortions reported vs Year for Jammu and Kashmir



Since district-specific data for Maharashtra state became available during this period, we opted to skip this step and transition to working on Maharashtra state data.

Chapter 4

Utilizing data analysis techniques to analyze hospital data collected from districts in Maharashtra.

4.1 Mental Health Data Analysis

4.1.1 Dataset

The analysis utilized the 'Mental Health Clinic Monthly Report,' which encompasses mental health statistics spanning from January 2022 to March 2024 across all districts of Maharashtra. The dataset was organized chronologically, segmented into yearly intervals, with each year further subdivided into four sectors, each comprising three months of data.

4.1.2 Data Preprocessing

1. The dataset was refined to include only columns deemed essential for modeling the mental health status of each district. These columns retained are:
    a. Organisation Unit Name
    b. Common Mental Disorders out of Total OPD-PHC
    c. Severe Mental Disorders out of Total OPD-PHC
    d. Suicidal Risks out of Outreach OPD-PHC_NMH
    e. PHCs with Functional DMHP ClinicOPD_NMHP
    f. PHCs with Availability of Psychotropic Drugs_NMHP
    g. Year
    h. Sector
2. Null values within the column were addressed by replacing them with zeros.

4.1.3 Health Index

1. To facilitate the mapping and comparative analysis of health conditions across districts, a composite parameter termed the 'Health Index' was developed. The health index was derived as a weighted average, combining several key metrics with varying weights:
    a. Common mental disorders out of total OPD-PHC: -0.1
    b. Severe mental disorders out of total OPD-PHC: -0.5
    c. Suicidal risks out of outreach OPD-PHC_NMHP: -3
    d. PHCs with functional DMHP clinic OPD_NMHP: 2
    e. PHCs with availability of psychotropic drugs_NMHP: 3
2. The determination of these weights was informed by the correlation of each health parameter with its impact on mental health within the respective district.
3. A correlation matrix was employed to validate the appropriateness of these weight assignments, ensuring they accurately reflect the influence of each parameter on mental health outcomes.

4.1.4 Plotting the data on map of Maharashtra

4.1.4.1 Function

1. parse_kml
    a. The function, `parse_kml`, is designed to extract geographical data from a Keyhole Markup Language (KML) file, typically used to represent geographic features and their associated attributes. Here's a brief overview of what the function does:
    b. Input: The function takes a KML file path (`kml_file`) as input
    c. Parsing the KML File: It parses the KML file using the `ElementTree` module from Python's standard library (`xml.etree.ElementTree`). This allows the function to navigate through the XML structure of the KML file
    d. Extracting District Data: It iterates through each `<Placemark>` element in the KML file, which typically represents a geographic feature, such as a district in this case
    e. Extracting Attributes: For each `<Placemark>`, the function extracts relevant attributes such as `district_name`, `district_id`, and `state` from the `<ExtendedData>` section of the KML file. These attributes are typically stored as `<SimpleData>` elements within the `<ExtendedData>`
    f. Extracting Geometry: The function also extracts the coordinates of the district boundary from the `<coordinates>` element within each `<Placemark>`. It parses these coordinates and constructs a polygon using the `shapely` library, representing the geometry of the district
    g. Constructing DataFrame: All the extracted data, including district attributes and geometry, are appended to a list of dictionaries
    h. Returning DataFrame: Finally, the function returns a Pandas DataFrame containing the extracted district data, where each row represents a district with its associated attributes and geometry
    i. This function is useful for processing geographic data stored in KML format and converting it into a structured DataFrame, which can be further analyzed or visualized using various tools and libraries available in Python

2. plot_map
    a. The function `plot_map` is designed to visualize geographic data overlaid with additional information, focusing on districts within the state of Maharashtra
    b. Filtering Data: The function filters the input DataFrame (`df`) based on the specified `year` and `sector`, selecting only the relevant data for visualization
    c. Data Preparation: It converts the `district_id` column in both the filtered DataFrame (`df_new`) and another DataFrame called `maharashtra_districts` to integers, assuming both contain this column. Then, it merges these DataFrames based on the `district_id` column to combine the geographic information with the data to be visualized
    d. Creating GeoDataFrame: It constructs a GeoDataFrame (`gdf`) from the merged data, associating each row with its corresponding geometry (district boundary)
    e. Plotting: The function generates a plot using Matplotlib, specifying the size of the plot (`figsize`) and the colormap (`cmap`) to use for visualizing the specified column (`column`). It adds black edges to the district polygons for clarity and includes a legend for interpreting the colormap
    f. Annotations: Annotations are added for district names using the centroid of each district polygon. Additionally, specific districts specified in the `districts_to_annotate` list are annotated with their names and corresponding values

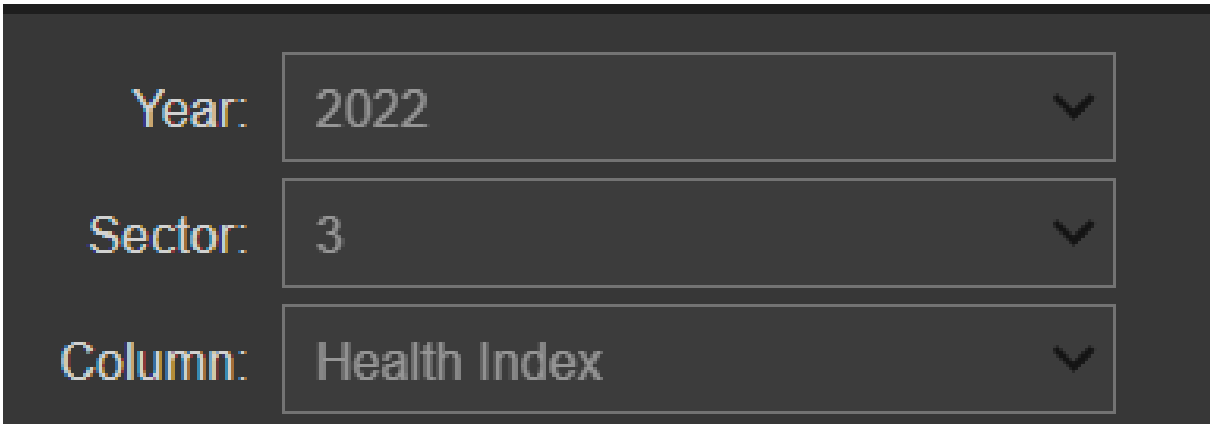
from the specified `column`. These annotations help in highlighting specific districts of interest and their associated data

g. Display: Finally, the function displays the plot using `plt.show()`
h. Overall, this function provides a convenient way to visualize geographic data along with additional information, aiding in the exploration and interpretation of spatial patterns and trends within the state of Maharashtra

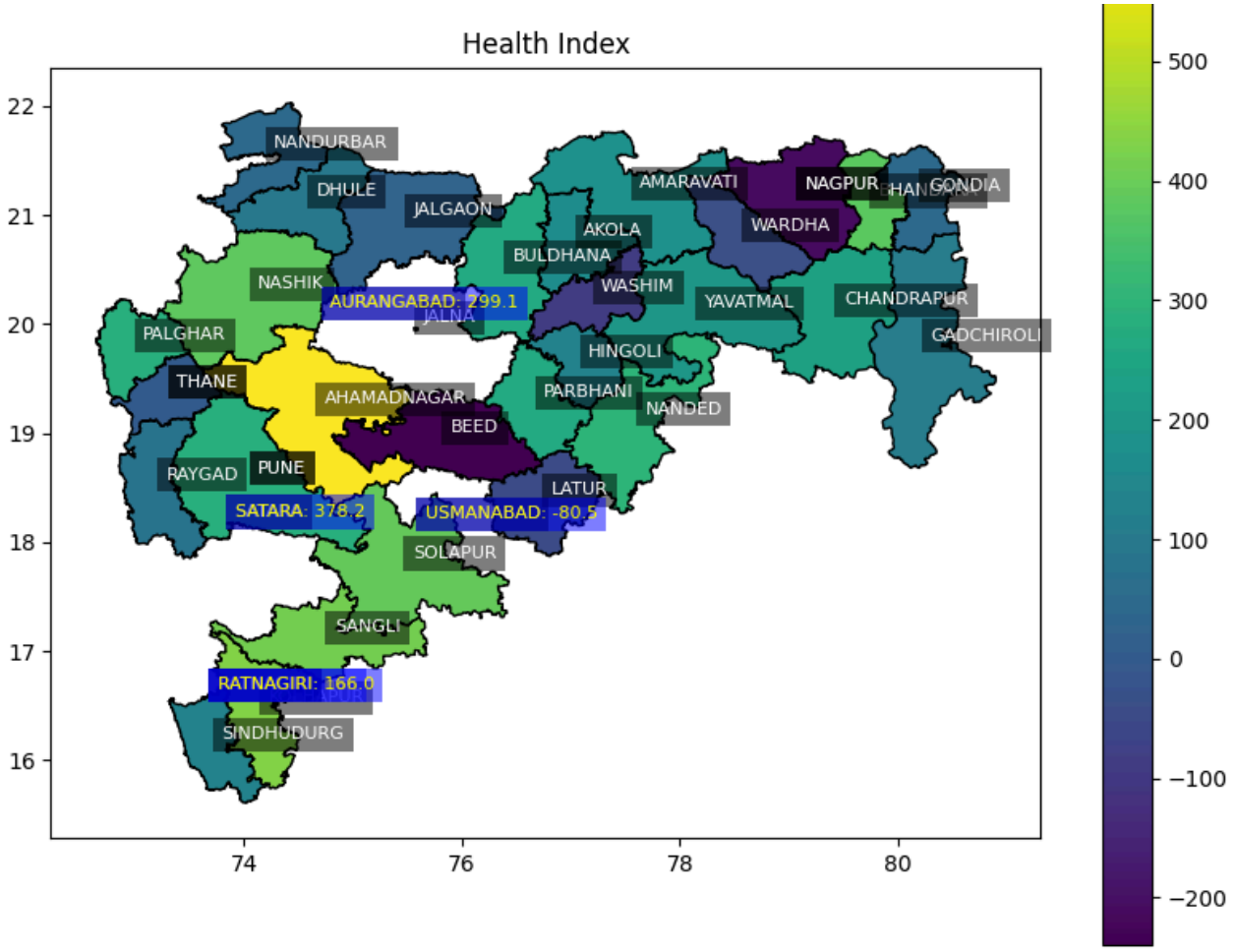
3. interactive_plot_map
    a. The `interactive_plot_map` function enables interactive visualization of geographic data overlaid with additional information
    b. It utilizes `ipywidgets` to create dropdown menus for selecting parameters such as `year`, `sector`, and `column`
    c. Users can dynamically adjust these parameters to explore different aspects of the data. Upon selection, the function calls `plot_map` with the chosen parameters, generating a plot that reflects the selected data subset
    d. This interactive functionality enhances user engagement and facilitates exploration of spatial patterns and trends within the dataset

4.1.4.2 Results

The interactive graph functionality empowers users to generate plots for various years and sectors, across different columns.



The following graph represents the Health Index for year 2022 and sector July-Sept

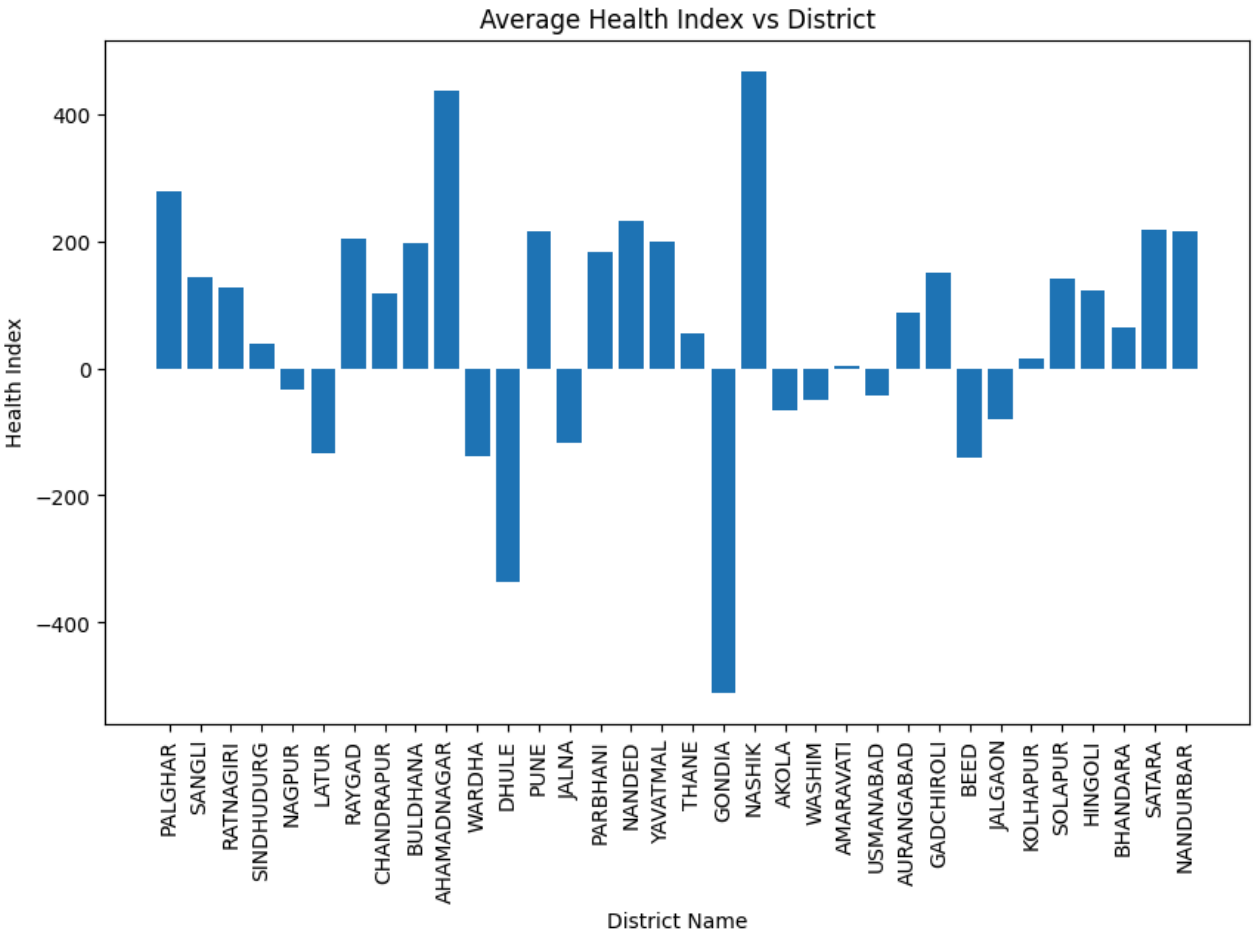


4.1.5 Feature vs Districts

Note : Data availability for districts such as Ratnagiri, Pune, and Thane was limited. It could imply that residents in these districts either do not seek treatment for mental health issues at government hospitals or that recording practices within these hospitals faced challenges, resulting in incomplete data collection.

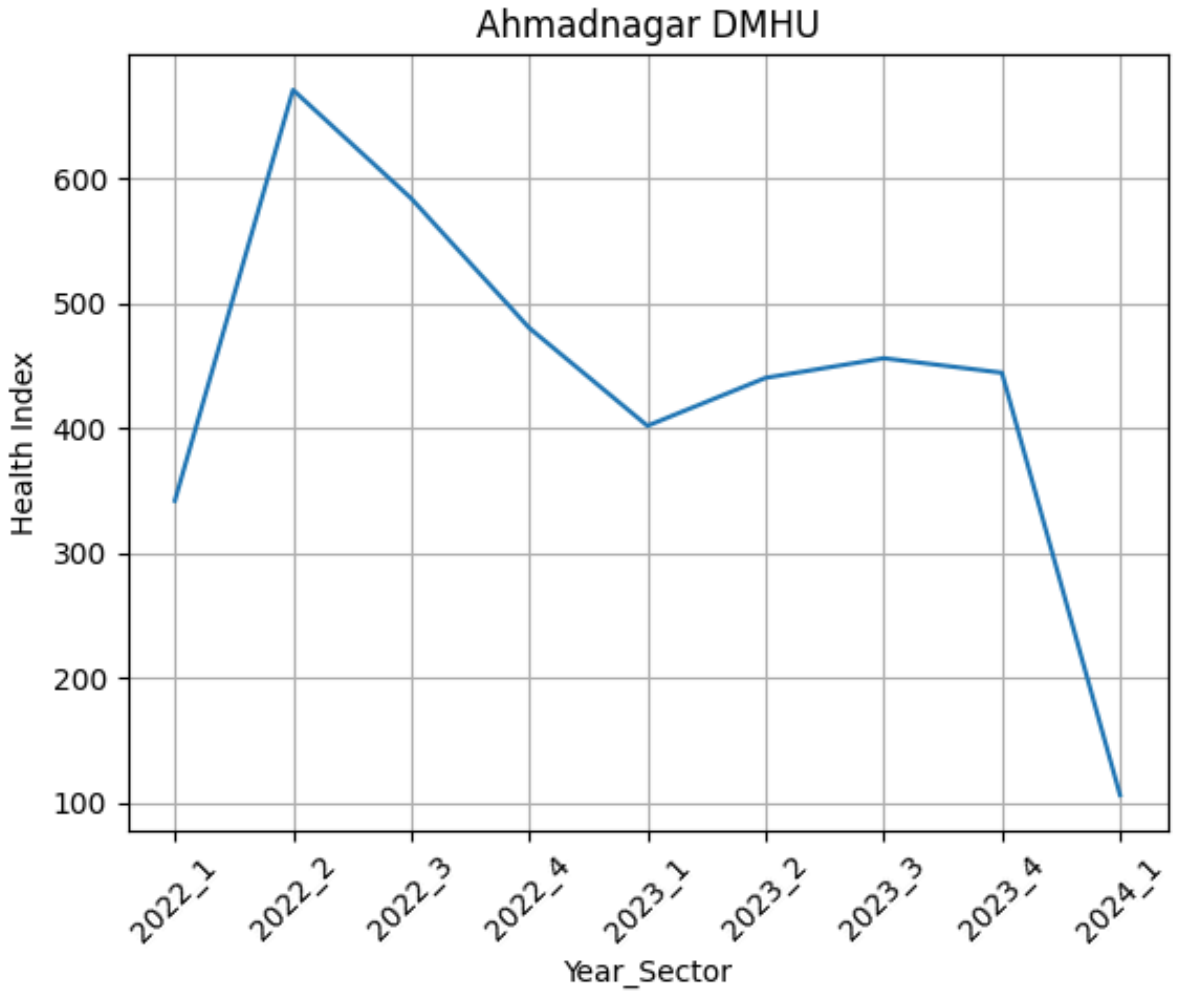
4.1.5.1 Health Index vs Districts

1. The figure below displays the average health index of all districts over a four-year span from 2022 to 2024.
2. Negative values of the health index observed in some districts can be attributed to a disproportionate ratio between mental health cases and the availability of hospitals equipped to address them adequately.
3. The reliability of the results for Ratnagiri, Pune, and Thane districts may be compromised due to significantly limited data availability.



1. When plotting the Health Index against individual districts, no consistent pattern was observed across all observations
2. However, a common trend emerged around the beginning of 2022, where a peak in the health index was noticeable
3. This phenomenon coincided with the onset of the COVID-19 era, during which fewer individuals sought mental health checkups, leading to a decrease in the number of reported cases. Consequently, the health index spiked as the relative scarcity of patients skewed the data
4. A distinct decrease in the health index was observed around 2024 for few districts, coinciding with reductions in the availability of primary health care centers equipped with psychotropic drugs and clinicopds.

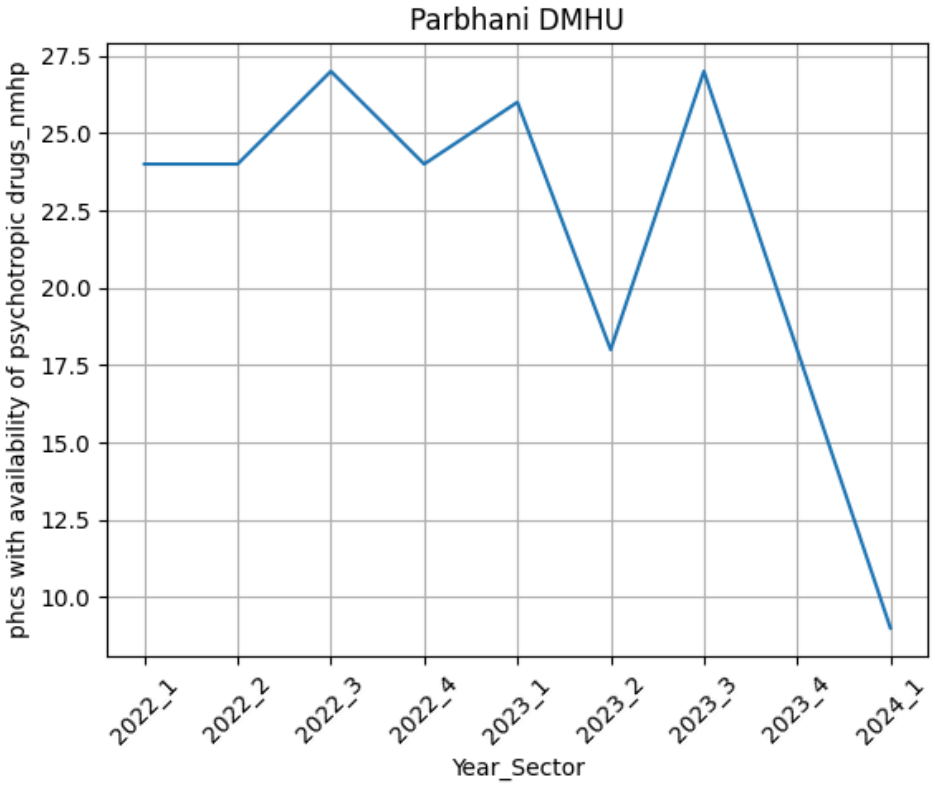
Health Index for Ahmadnagar DMHU

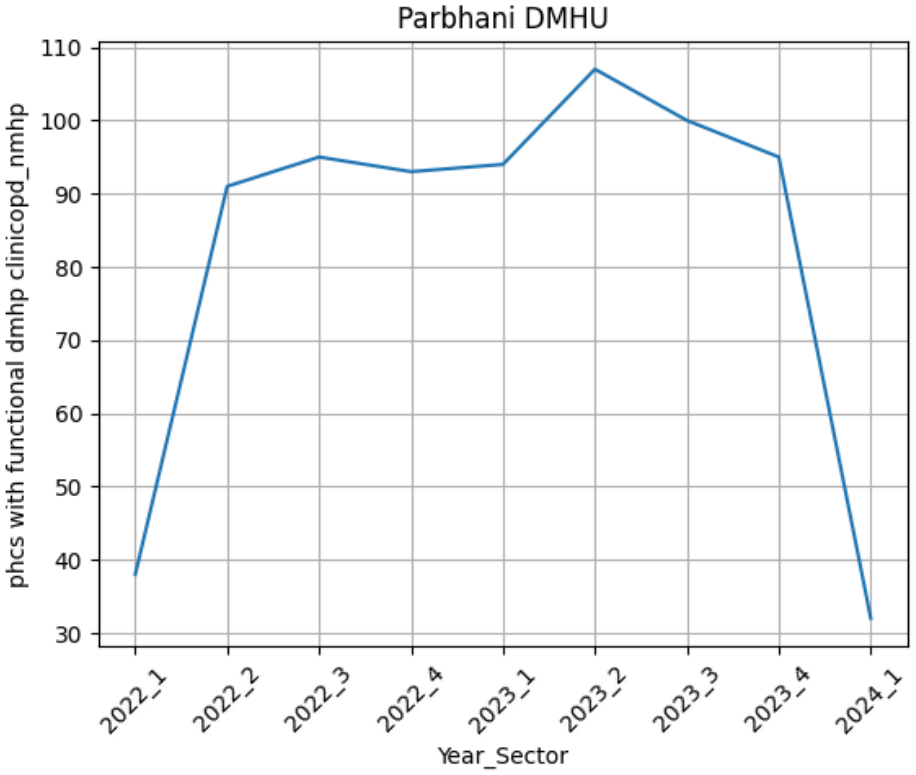


4.1.5.2 PHCs with availability of psychotropic drugs and PHCs with functional clinicopd vs Districts

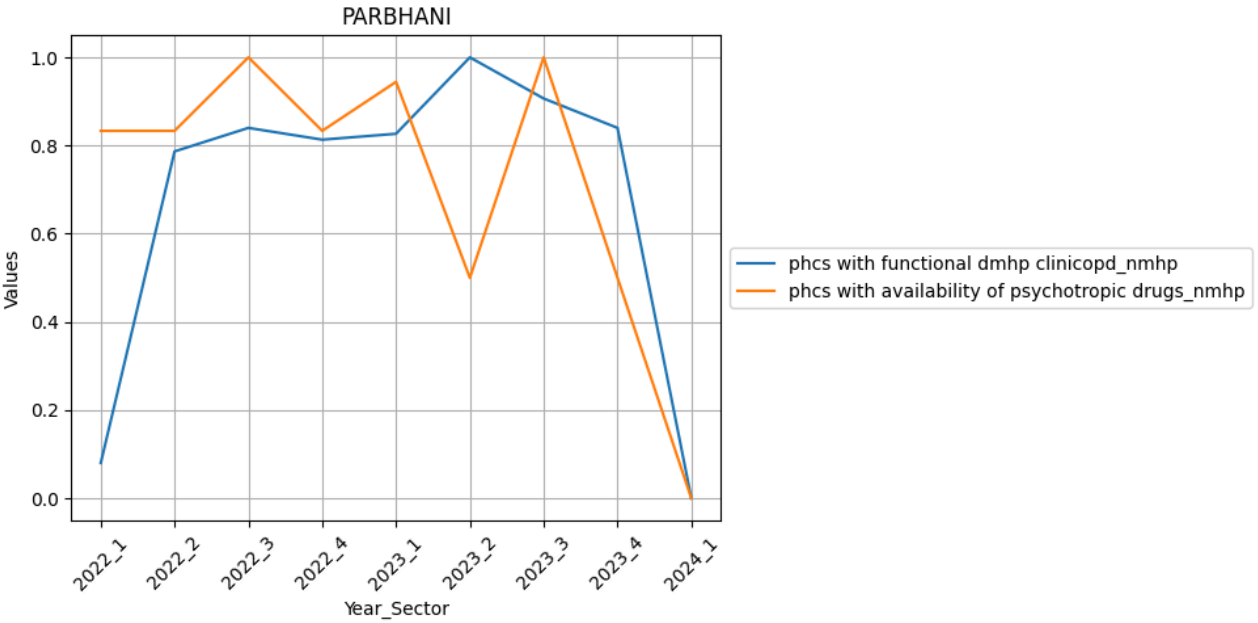
1. The general trend was spikes around 2022 but decreased around 2024
2. It's plausible that during the COVID pandemic, an increase in the number of active hospitals occurred, primarily to address COVID-related concerns rather than specifically for mental health monitoring. This surge in hospital activity could potentially account for the spikes observed in the data during that period

PHCs with availability of psychotropic drugs and PHCs with functional clinicopd vs Prabhani DMHU District



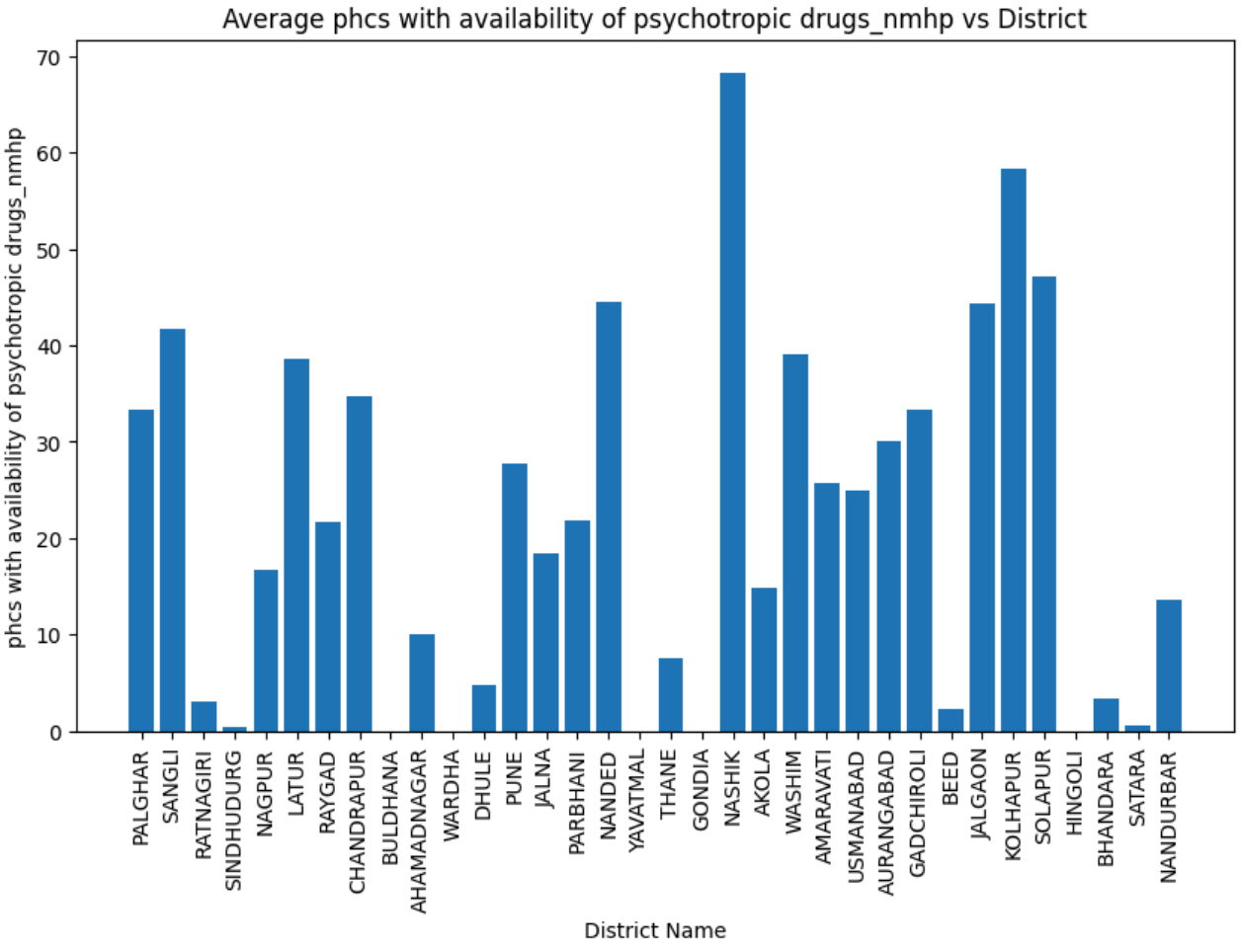


Normalized Graph (values normalized between 0 and 1 for each parameter)

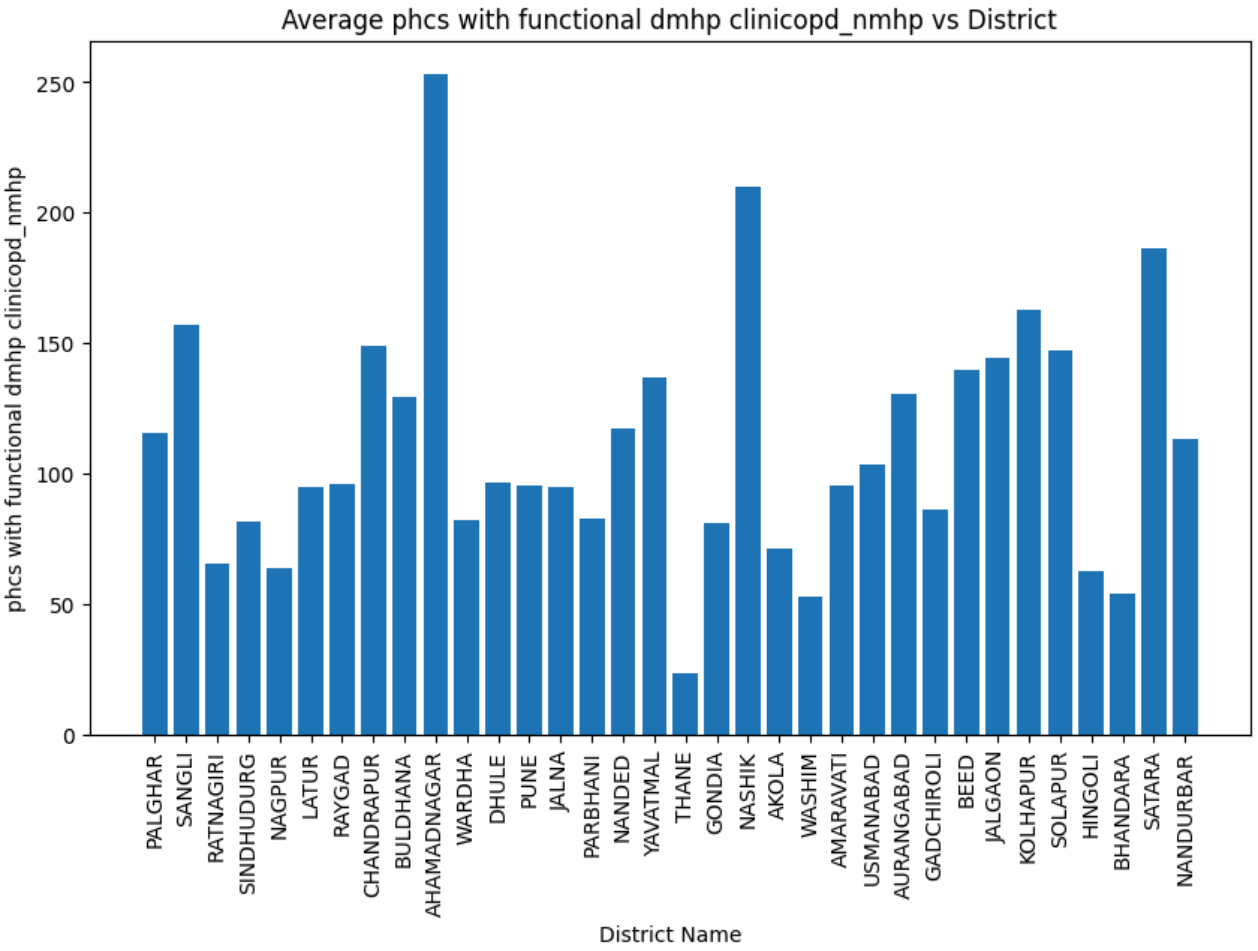


Similar relationships were found between PHCs with availability of psychotropic drugs and PHCs with functional clinicopd in almost all districts.

The mean availability of psychotherapeutic drugs in primary health care centers across all districts, averaged over the period from 2022 to 2024 is plotted below. The mean is for a span of three months.



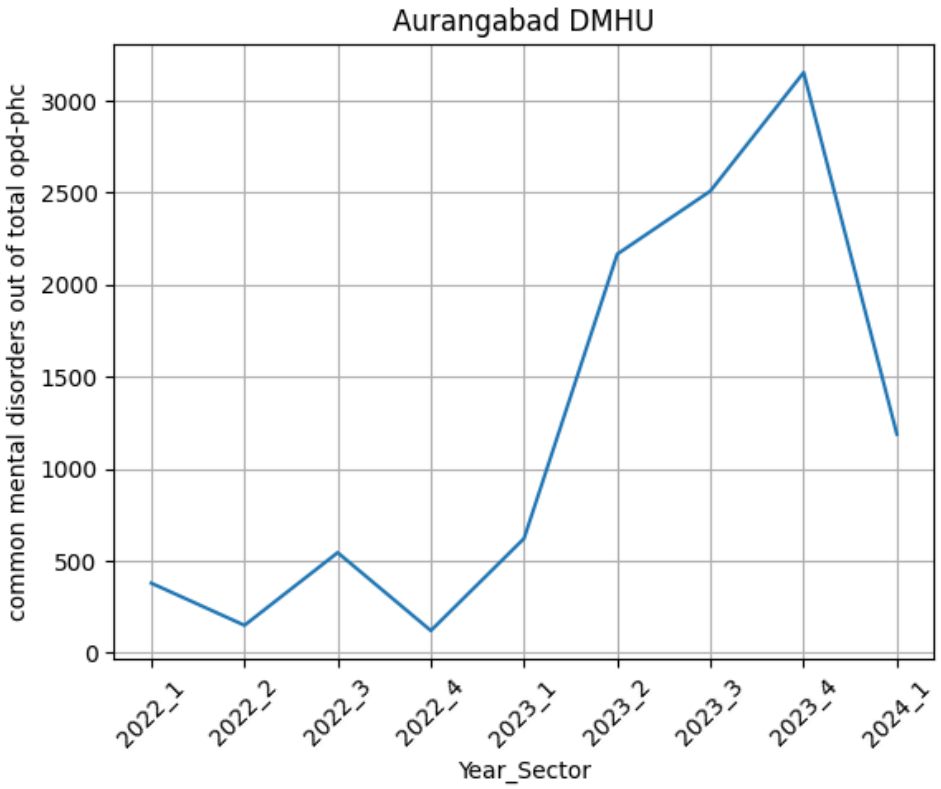
The mean availability of clinicpod in primary health care centers across all districts, averaged over the period from 2022 to 2024 is plotted below. The mean is for a span of three months.

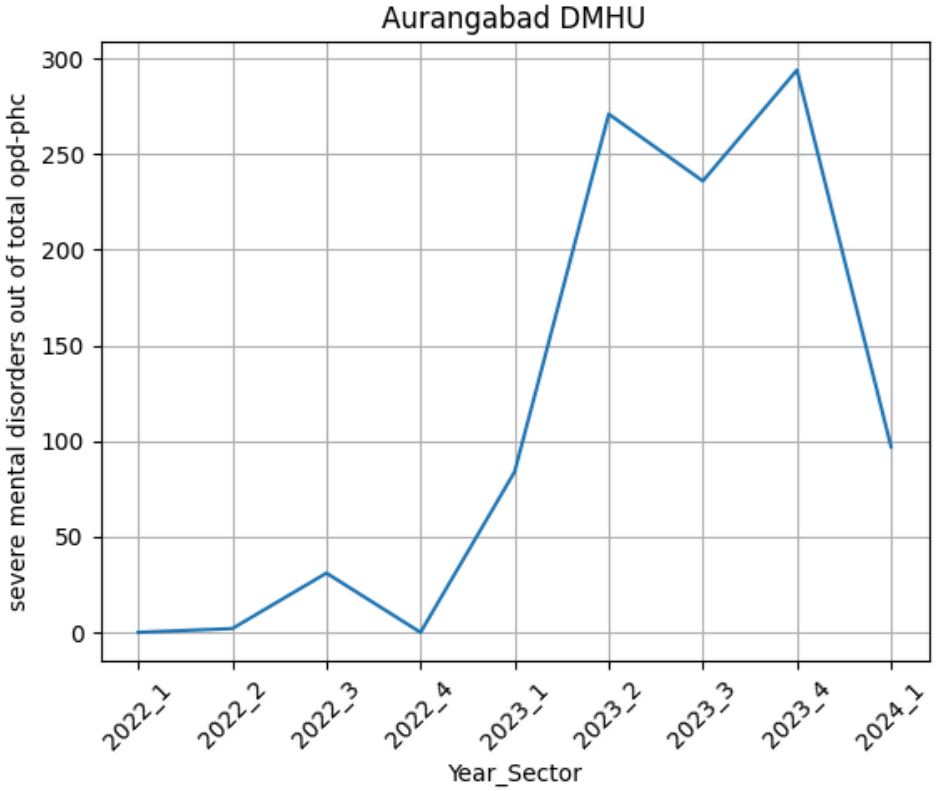


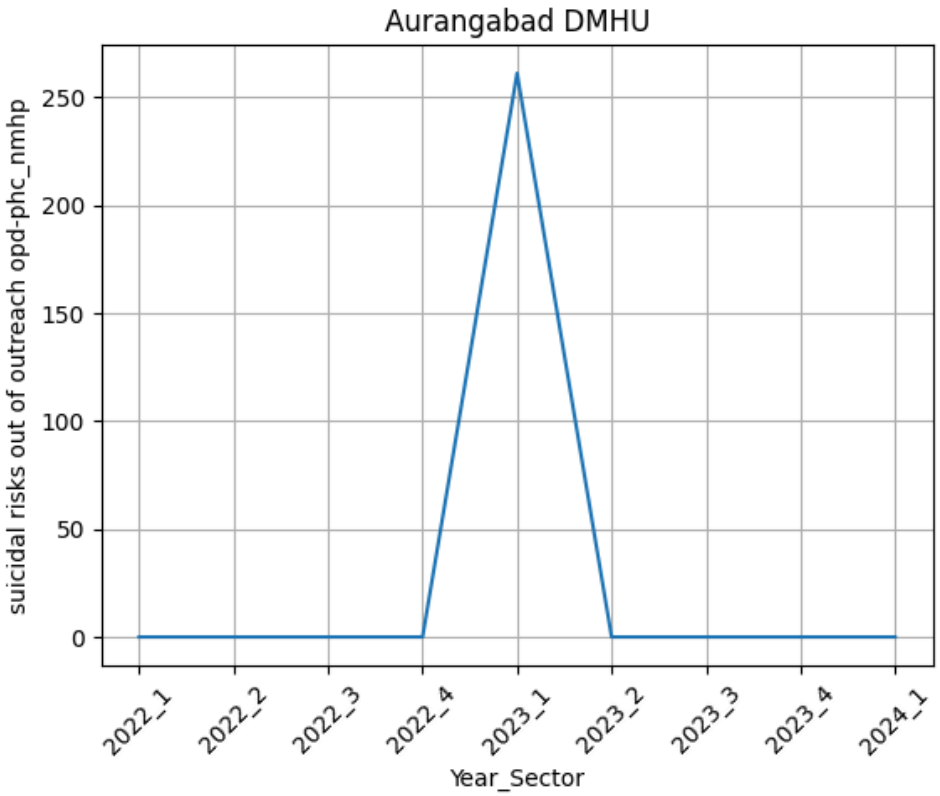
4.1.5.3 Patients with Common mental disorder, Severe mental disorder and Suicidal risk vs Districts

1. The overall trend indicated a decrease in the number of patients during the early months of 2022, coinciding with the onset of the COVID-19 pandemic, where fewer individuals sought mental health checkups. However, towards late 2022 and 2023, there was a notable increase in the number of patients seeking care
2. There was another decline in the number of patients observed around 2024, particularly noticeable in certain districts

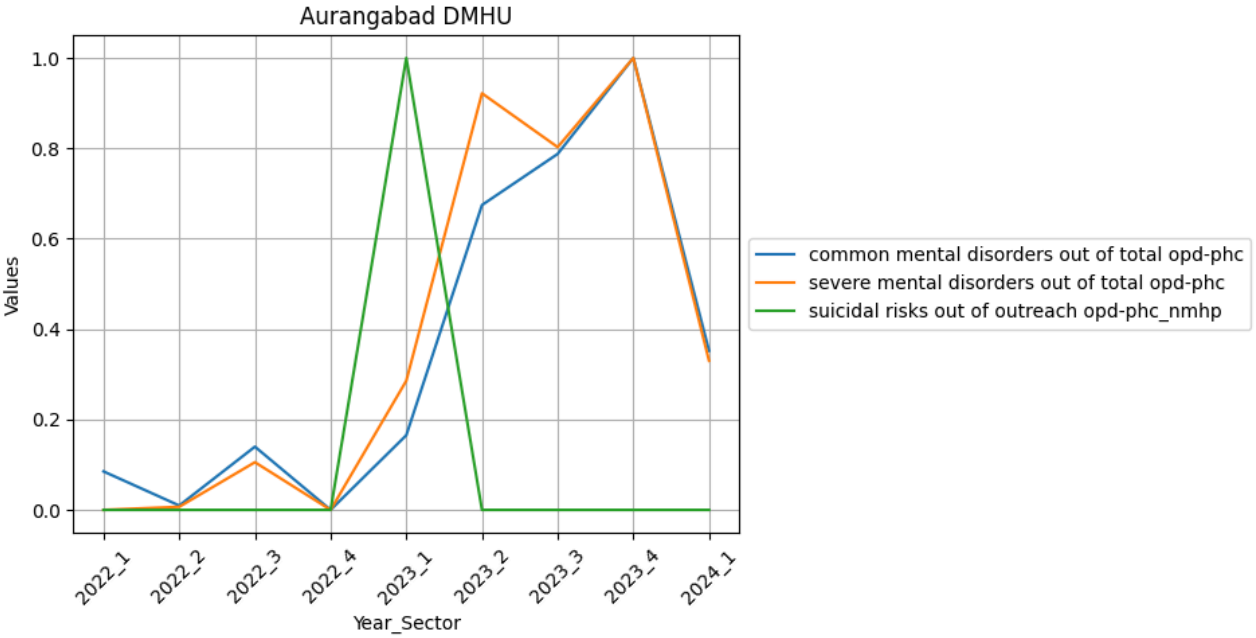
Patients with Common mental disorder, Severe mental disorder and Suicidal risk vs Aurangabad District


Aurangabad DMHU
common mental disorders out of total opd-phc
3000
2500
2000
1500
1000
500
0
2022_1
2022_2
2022_3
2022_4
2023_1
2023_2
2023_3
2023_4
2024_1
Year_Sector


Aurangabad DMHU
severe mental disorders out of total opd-phc
300
250
200
150
100
50
0
2022_1
2022_2
2022_3
2022_4
2023_1
2023_2
2023_3
2023_4
2024_1
Year_Sector

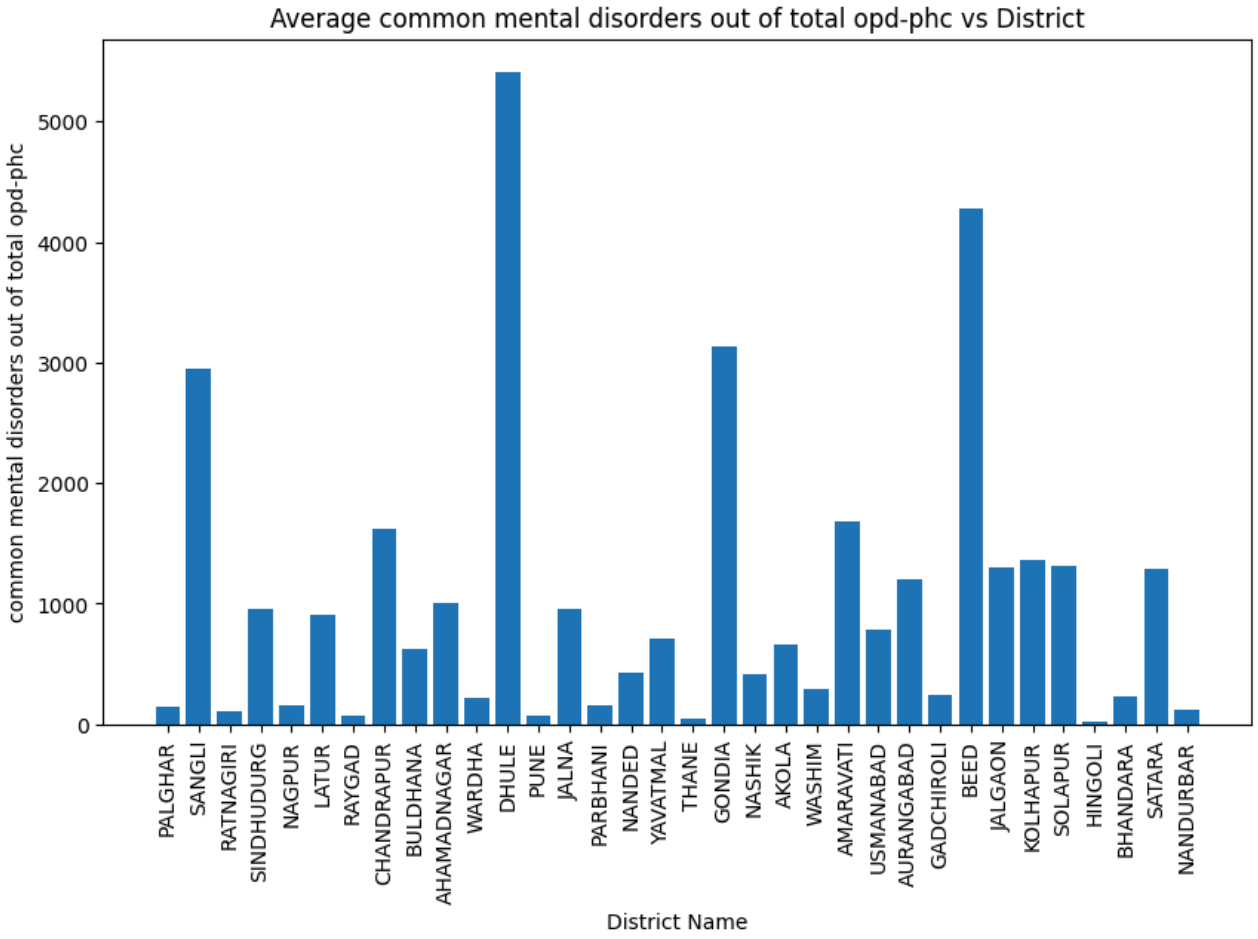


Normalized Graph (values normalized between 0 and 1 for each parameter)

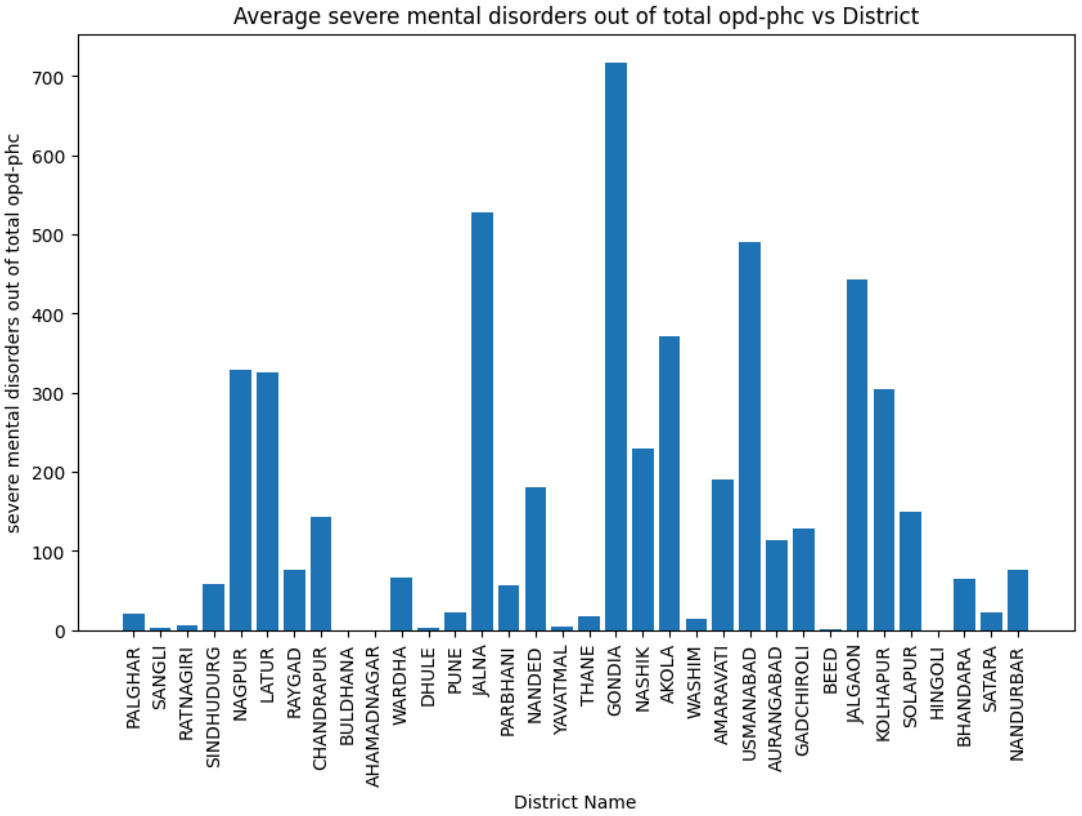


Similar trends were shown between Patients with Common mental disorder, Severe mental disorder and Suicidal risk in almost all districts

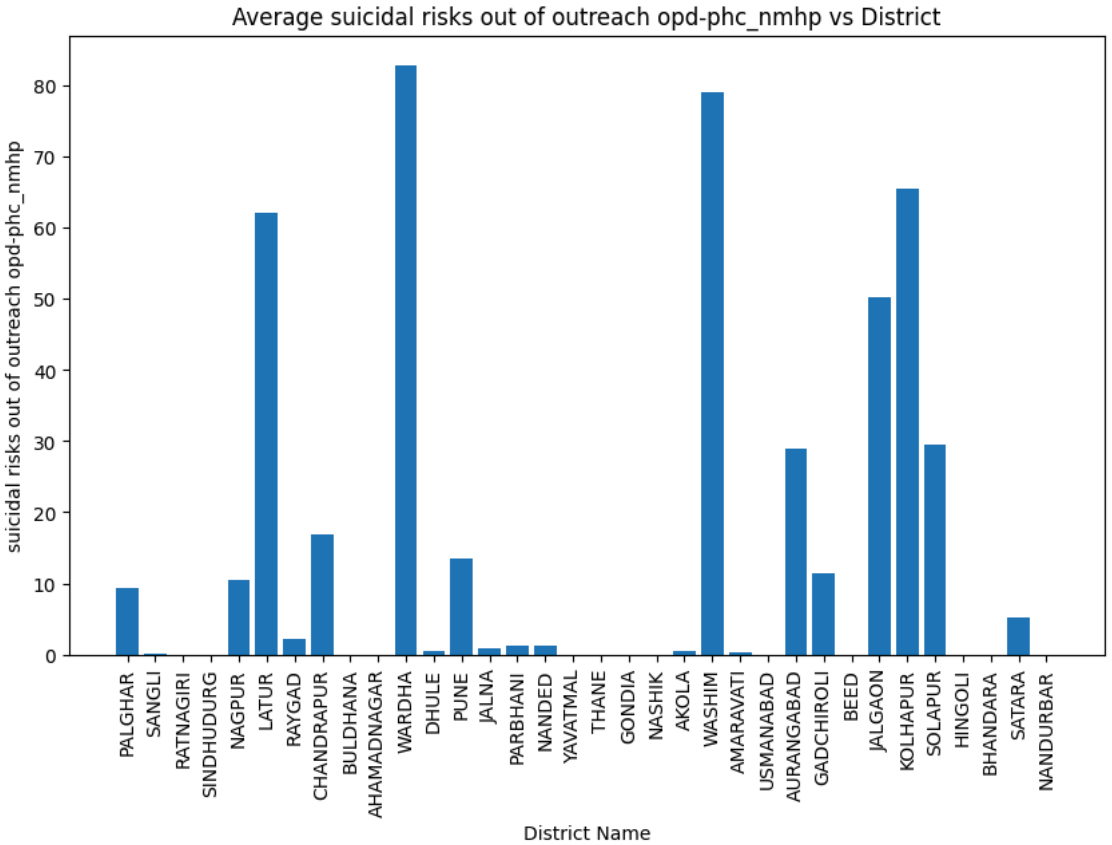
The mean number of common mental disorder patients across all districts, averaged over the period from 2022 to 2024 is plotted below. The mean is for a span of three months.

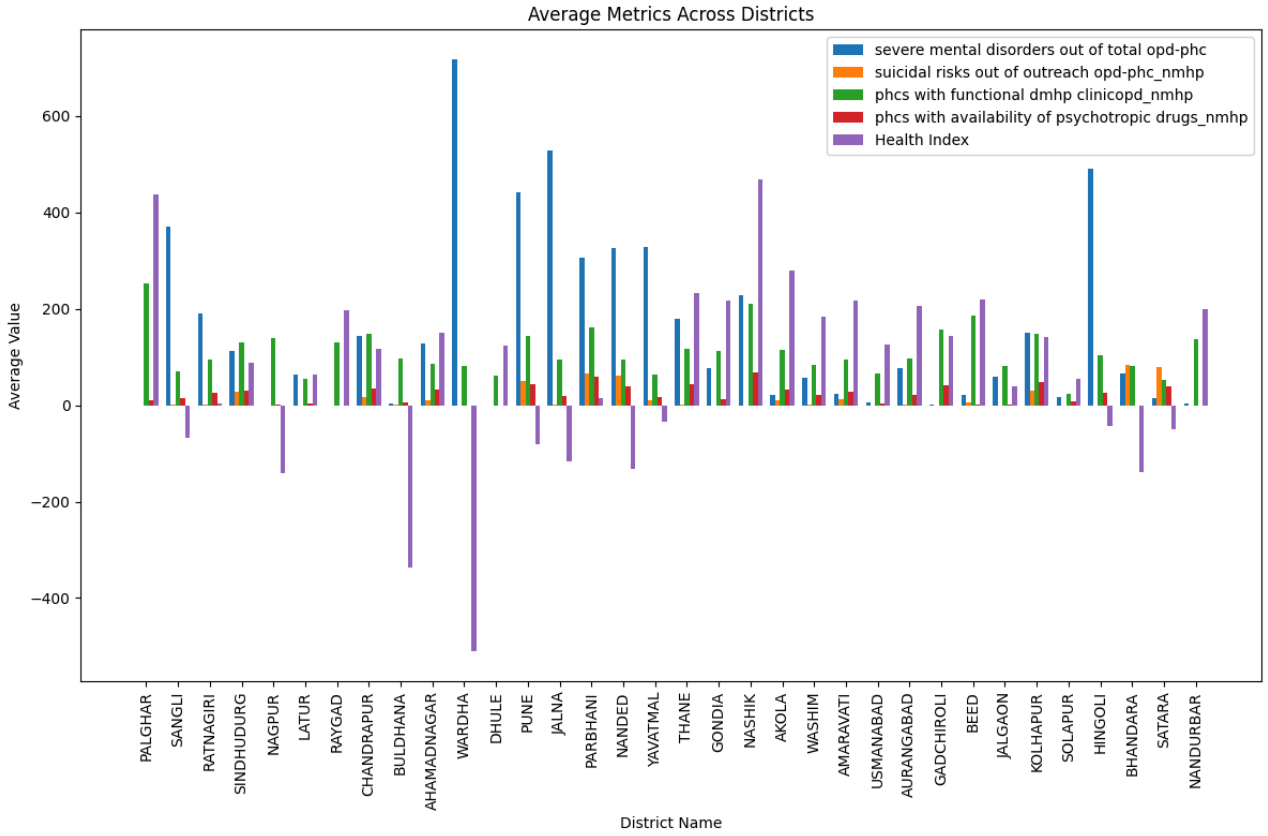


The mean number of severe mental disorder patients across all districts, averaged over the period from 2022 to 2024 is plotted below. The mean is for a span of three months.



The mean number of suicidal risk patients across all districts, averaged over the period from 2022 to 2024 is plotted below. The mean is for a span of three months.





4.1.6 Conclusion

The impact of COVID-19 on mental health appears to be significant. Analysis indicates a notable decrease in the number of mental health patients during the pandemic, likely attributed to the challenges individuals faced in accessing hospitals amidst the health crisis. Consequently, records reflected a decrease in patient numbers, leading to a surge in the Health Index. This trend can be attributed to the inverse relationship between mental health patient numbers and the Health Index, wherein a lower count of mental health patients correlates with a higher Health Index. However, as COVID-19 restrictions eased, hospital visits for treatment increased, causing a subsequent rise in patient numbers and a subsequent decline in the Health Index post-pandemic.

As can be observed, there are a few districts with a negative Health Index. It is evident that these districts have very few resources available to take care of the number of patients. Special attention should be paid to these districts.

4.1.6 Codes

1. Mental Health
2. HMIS Dataset 1
3. HMIS Dataset 2
4. Drive link for datasets