Development Of An Integrated Dashboard For Public Health Systems

A Research and Development project

by

Darshan Bitla (210020034)

Under the Supervision of

Dr. Raghavendran Lakshminarayanan

Koita Centre for Digital Health

INDIAN INSTITUTE OF TECHNOLOGY BOMBAY

Mumbai - 400076, India

May, 2024

Contents

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.

We then expanded our analysis to data made available by NPCDCS for number of patients diagnosed for various non-communicable diseases across the hospital units 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	-

Software	Description of Data	How to Collect
eAushadhi	This system is used by hospitals and procurement cells to manage drug inventory. It provides information on drugs offered under government programs and the location of warehouses along with contact details across various states. In Maharashtra, a dashboard is available on the website, providing Essential Drugs List (EDL) stock details compared to circle, store type, and store name. [Login required for additional data]	CSV and PDF available, additional data can be web scraped
eSanjeevani	This system serves patients, doctors, and Community Health Officers (CHOs) to enhance the availability of quality healthcare services in remote regions. It provides information on OPD timings. Chief Health Officers, doctors, and patients can access additional data upon logging in. This data includes details about patients visiting, hospital information, and doctor availability.	Web scraping
eHospital	You can track hospital records and statuses, including patient registrations, prescriptions, billing, and more through the e-hospital system. Additionally, state-wise data for the E-blood bank is available. For the ORS dashboard, visit https://ors.gov.in/dashboard/ to monitor hospital registrations, teleconsultations, and appointment numbers.	Some information in excel and pdf format, and some to be web scraped
eRaktkosh	This system is utilized by blood donors, individuals in need of blood donations, hospitals, and other relevant parties. It offers data such as the number of camps held, blood units collected, registered donors, and blood availability searches district-wise. The information includes blood bank locations, contact details, and available blood types. Additionally, a blood directory with similar details is accessible, catering to both donors and seekers.	Web scraping
HwC Portal	The available data includes the count of Ayushman Arogya Mandirs established to offer primary healthcare services. Access to Covid19 financial reporting and health worker campaign data requires login credentials. [Login required for access to Covid19 financial reporting and health worker campain data]	Reports from previous years are accessible in PDF format.
NCD portal	Utilized for screening men and women over 30 for non-communicable diseases. [Login required]	-
Covid19 Dashboard	The system offers various datasets including the number of deaths, active cases, and discharges, along with state-wise COVID-19 data and vaccination statistics. Additionally, district-wise weekly positivity rates data is available. However, the availability of this data is limited to a specific timeframe (13th June to 19th June).	xlsx format

IHIP Integrated Health Information Platform	Used to manage data of various health indicators. [Login required]	-

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

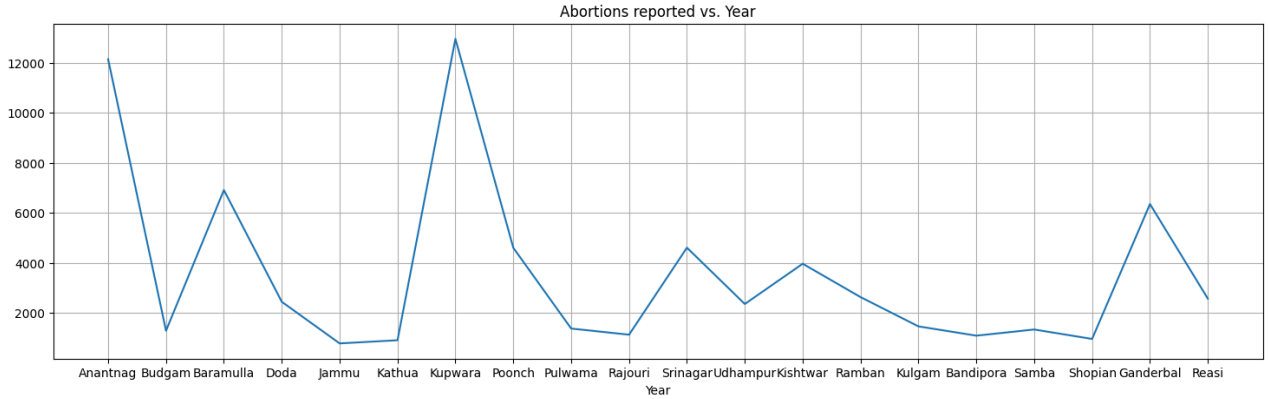
3.1.3 Plotting data specific to a district

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.

3.2 Longitudinal Aging Study in India (LASI) data

3.2.1 Dataset classification:

There are many LASI datasets available, some of them concerned with the elderly people are classified below:

1 Based on type of Healthcare received to elder people:

State/UT wise Percentage of households with any member hospitalized in the past 12 months and households incurring catastrophic health expenditure by place of residence Under Longitudinal Ageing Study in India (LASI) Wave 1
States/UT wise Percent distribution of older adults age 45 and above by perceived quality of care for healthcare facilities during inpatient and out-patient care in the most recent visit prior to survey Under LASI,Wave 1 , 2017-18
Percentage of physically active and inactive older adults age 45 and above by sex Under Longitudinal Ageing Study in India (LASI) Wave 1
States/UT wise Percentage of older adults who utilized out-patient care in one month and inpatient care in one year prior to survey Under Longitudinal Ageing Study in India (LASI) Wave 1, 2017-18
States/UT wise Percent distribution of older adults who received out-patient care in one year prior to survey by type of health facilities Under Longitudinal Ageing Study in India (LASI) Wave 1, 2017-18
States/UT wise Percent distribution of older adults who received inpatient care in one year prior to survey by type of health facilities Under Longitudinal Ageing Study in India (LASI) Wave 1, 2017-18

2. Based on Condition and Situation of living of elderly people such as their daily activities and requirements:

State/UT wise Percentage of elderly age 60 and above with ADL and IADL Under Longitudinal Ageing Study in India (LASI) Wave 1
States/UT wise Percentage of older adults who experienced food insufficiency at household Under Longitudinal Ageing Study in India (LASI) Wave 1
States/UT wise Percentage of older adults age 45 (Including spouse irrespective of age) and above who are frequent non-heavy drinker and heavy episodic drinker Under Longitudinal Ageing Study in India (LASI) Wave 1
State/UT wise Percentage of older adults having family members who are unable to carry out basic daily activities Under Longitudinal Ageing Study in India (LASI) Wave 1
States/UT wise Percentage of older adults by use of smokeless tobacco status and sex Under Longitudinal Ageing Study in India (LASI) Wave 1
States/UT wise Percentage of older adults using aids or supportive devices1 Under Longitudinal Ageing Study in India (LASI) Wave 1
States/UT wise Percentage of older adults with ADL or IADL limitations who needed helper Under Longitudinal Ageing Study in India (LASI) Wave 1

Chapter 4

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

4.3 Diabetes and Hypertension Patient Data Analysis

4.3.1 Dataset

The analysis utilized data titled "Newly Diagnosed Patients under NPCDCS". The NPCDCS initiative aims to facilitate early diagnosis of common non-communicable diseases through dedicated units. The dataset comprises unit-wise distribution of the number of patients for different cancer types and covers the period from January 2021 to March 2024, segmented into nine quarters, each year divided into four quarters. Our focus within this analysis pertains to Diabetes and Hypertension data.

4.3.2 Data Preprocessing

The columns of interest in the given dataset are-

- organizationalunitname
- newly diagnosed patients-diabetes only
- newly diagnosed patients-hypertension only

newly diagnosed patients-dm & htn (both).
All the other columns were dropped and all files were merged together keeping track of the period in a separate column.
The missing values are all associated with the Mumbai District across various periods. Each missing entry corresponds to a different period, therefore-

- Data might consistently be missing OR
- Not reported for Mumbai District across these periods.

4.3.3 Exploratory Data Analysis

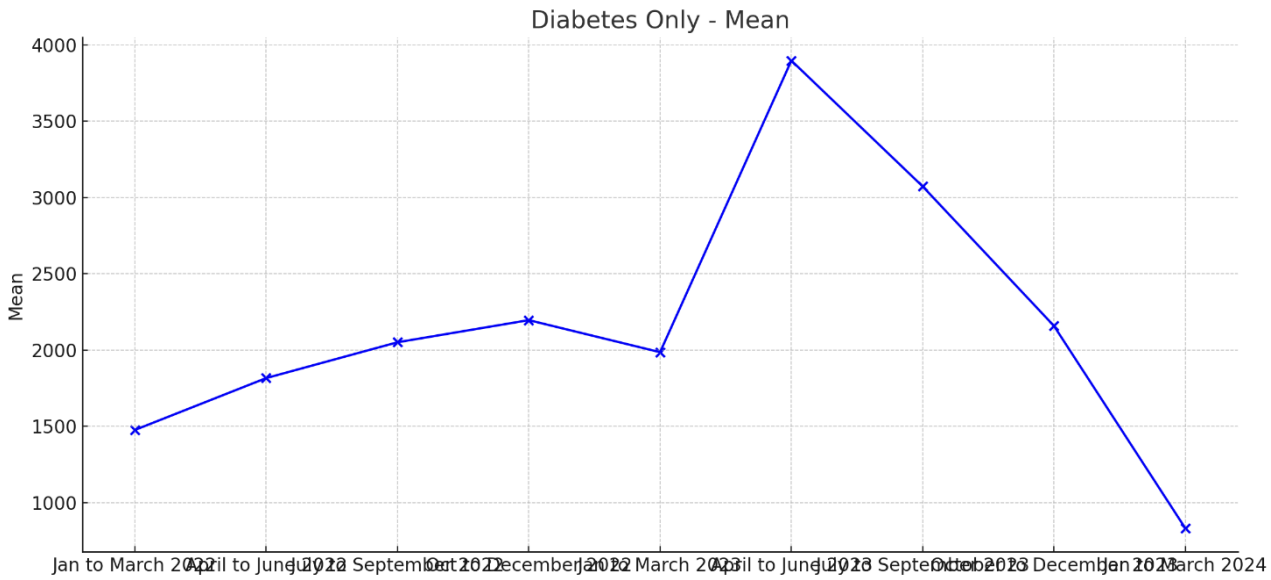
Basic statistics such as mean, median, standard deviation, and range for each condition in different districts were calculated to get an overview of the data distribution.

4.3.3.1 Exploratory Data Analysis Results

- **FOR DIABETES ONLY-**

Period	Mean	Median	STD	Range

Jan to March 2022	1475.52	502.5	1910.89	8469
April to June 2022	1815.57	950.5	2071.07	7263
July to September 2022	2050.48	1156.5	2365.29	9404
October to December 2022	2195.63	1166.0	2480.63	9891
Jan to March 2023	1985.53	1133.0	2330.71	8722
April to June 2023	3897.29	1407.5	4917.78	18160
July to September 2023	3072.33	1454.0	3679.94	15290
October to December 2023	2157.71	1086.0	2451.47	9360
Jan to March 2024	831.33	392.5	994.66	3727



Observations-

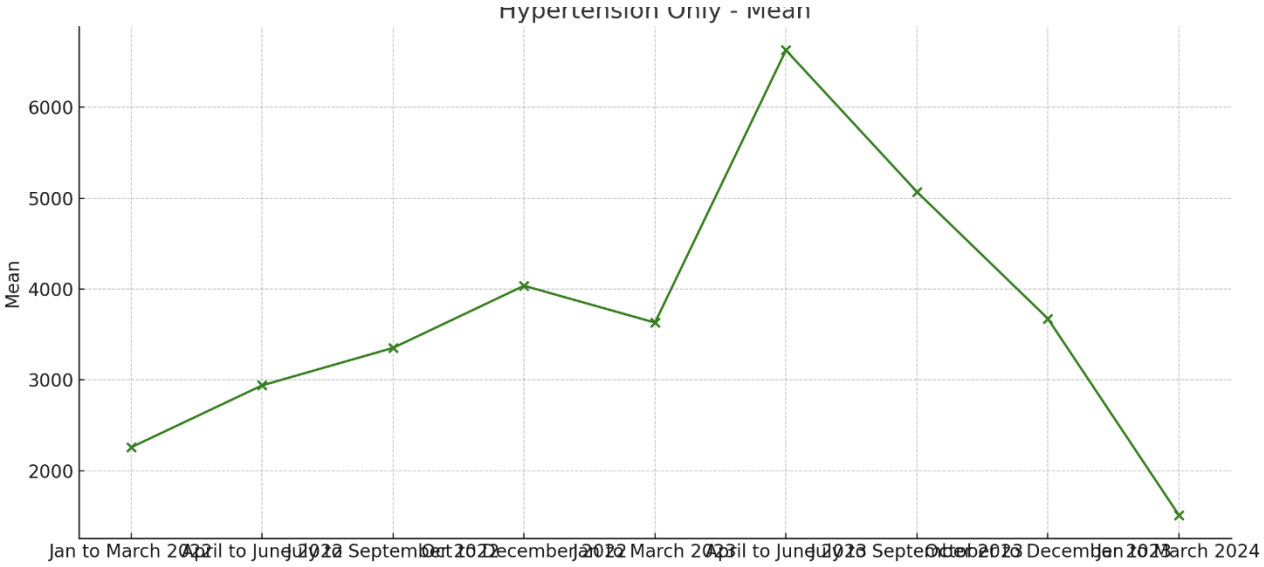
- Mean and Median: The mean and median diagnosis rates for diabetes show significant variation over the periods, peaking in April to June 2023. This suggests a potential increase in diagnosis or reporting during this period.

- Standard Deviation: The standard deviation is high in periods with higher means, indicating a wide dispersion of data and possibly outlier districts with very high diagnosis rates.

- Range: The range is notably high in periods like April to June 2023, reflecting extreme values between the lowest and highest diagnosing districts.

- **FOR HYPERTENSION ONLY-**

Period	**Mean**	**Median**	**STD**	**Range**
Jan to March 2022	**2259.33**	**791.0**	**3098.23**	**12918**
April to June 2022	**2940.57**	**1551.0**	**3731.27**	**14618**
July to September 2022	**3353.10**	**1861.0**	**4183.25**	**16525**
October to December 2022	**4036.47**	**1947.0**	**4998.65**	**18853**
Jan to March 2023	**3631.02**	**1834.0**	**4550.36**	**17184**
April to June 2023	**6628.00**	**1990.5**	**10079.86**	**47382**
July to September 2023	**5070.29**	**2051.0**	**7125.49**	**35315**
October to December 2023	**3674.90**	**1745.0**	**4287.44**	**16092**



Observations-

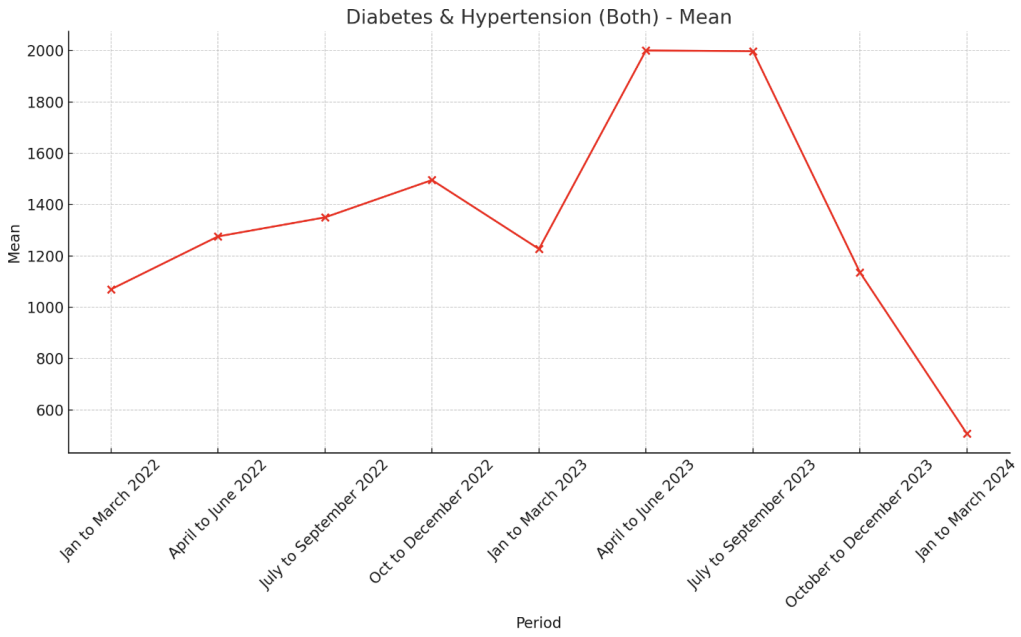
- Mean and Median: The mean and median values are generally higher than those for diabetes, with the peak again in April to June 2023. This might suggest more widespread screening or higher prevalence of hypertension.
- Standard Deviation: Very high standard deviations, especially in periods like April to June 2023, indicate significant variability across districts, suggesting that while some areas might have robust screening programs, others might lag.

- Range: The extremely high ranges, particularly in April to June 2023 and July to September 2023, highlight that some districts report exceptionally high diagnosis rates, which could skew the overall data.

- **FOR DIABETES+HYPERTENSION (BOTH)ONLY-**

Period	Mean	Median	STD	Range
Jan to March 2022	1069.90	365.5	1601.01	5873
April to June 2022	1275.71	557.5	1698.01	6677
July to September 2022	1349.95	730.5	1738.64	6842
October to December 2022	1495.30	710.0	1864.56	7651
Jan to March 2023	1226.79	640.0	1548.34	5741
April to June 2023	2000.10	887.0	2626.37	9854

July to September 2023	1997.24	856.0	2767.92	13683
October to December 2023	1136.00	444.0	1388.51	5387
Jan to March 2024	507.29	258.0	614.95	2325



Observations-

- Mean and Median: These values generally follow the trends seen in individual conditions but are lower, as expected since both conditions need to be present for a dual diagnosis.
- Standard Deviation: Consistently high standard deviation values indicate that the number of dual diagnoses also varies widely across districts.
- Range: The range is significantly high in periods like July to September 2023, which could point to specific public health campaigns or increased healthcare access during these times.

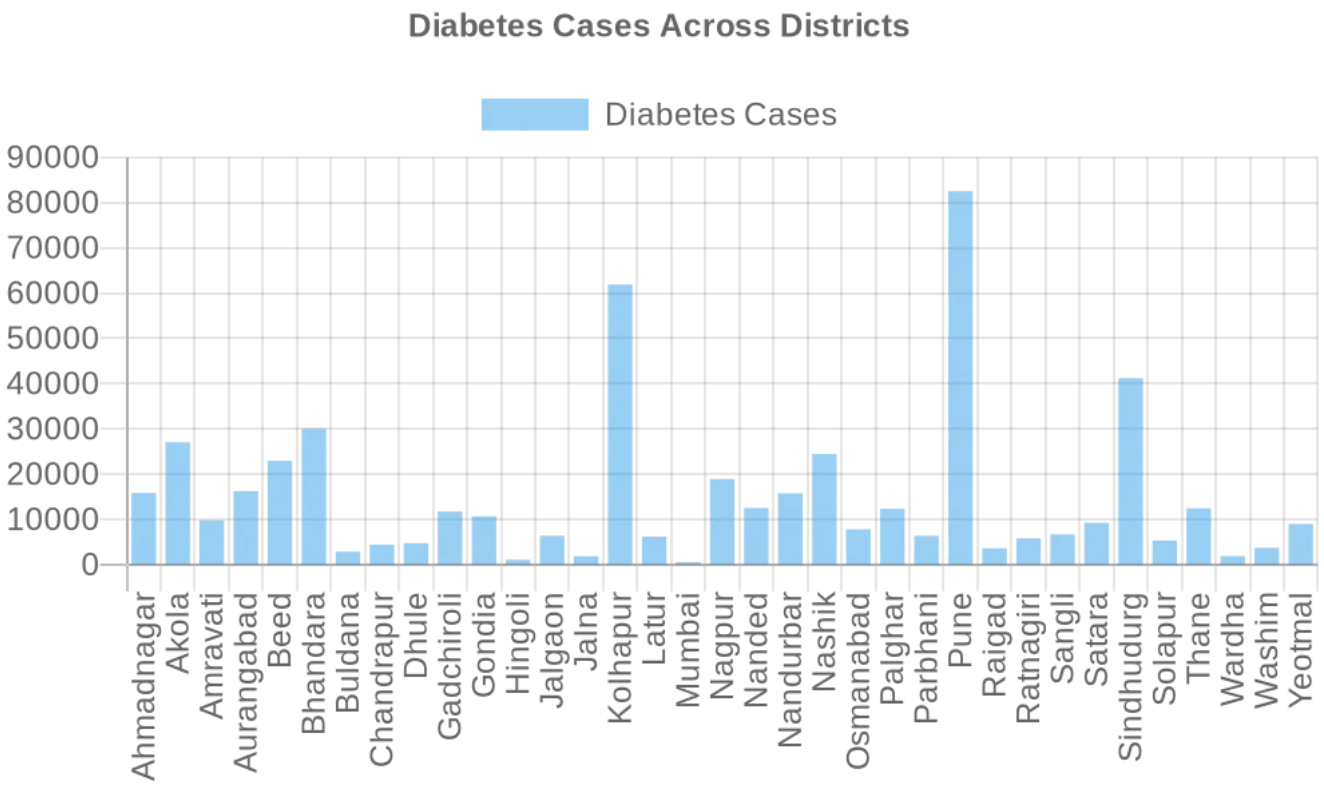
4.3.3.2 Exploratory Data Analysis Insights

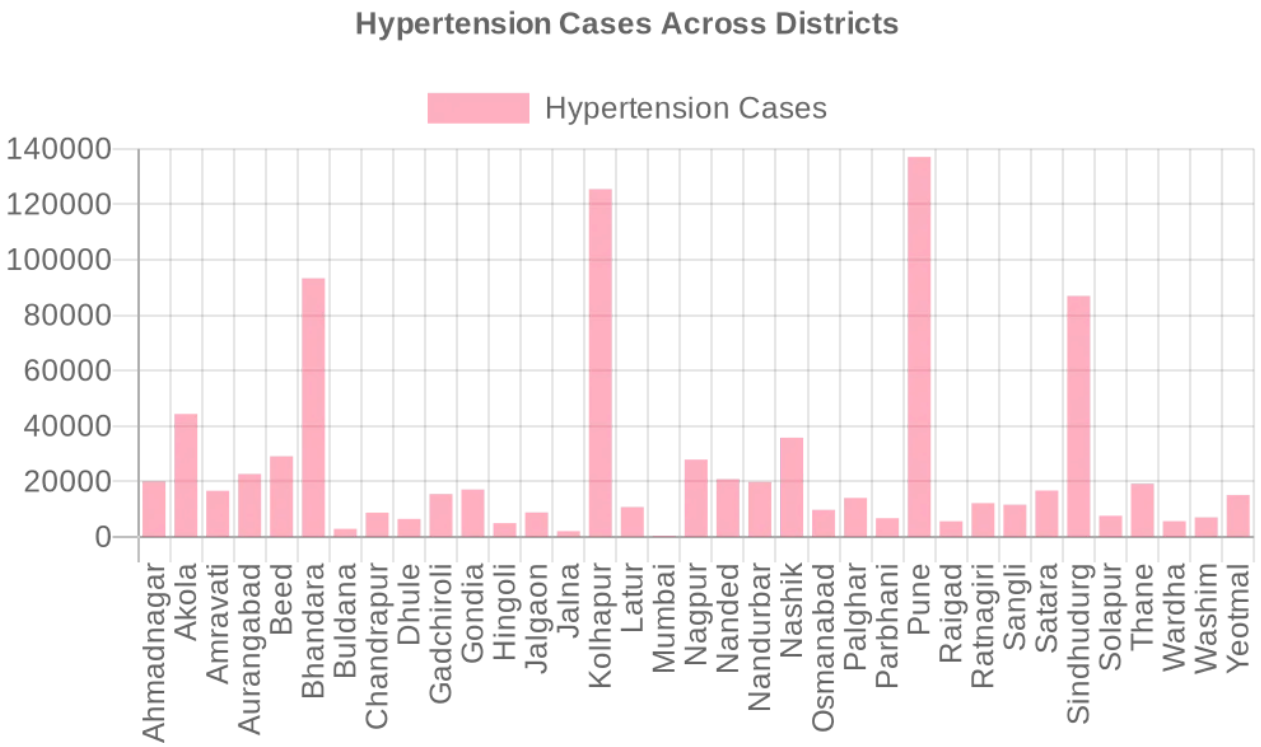
- Periodic Peaks in Diagnosis: Certain periods, like April to June 2023, show significant peaks across all conditions, which could be due to public health campaigns, seasonal variations in health-seeking behavior, or changes in healthcare policy or funding.

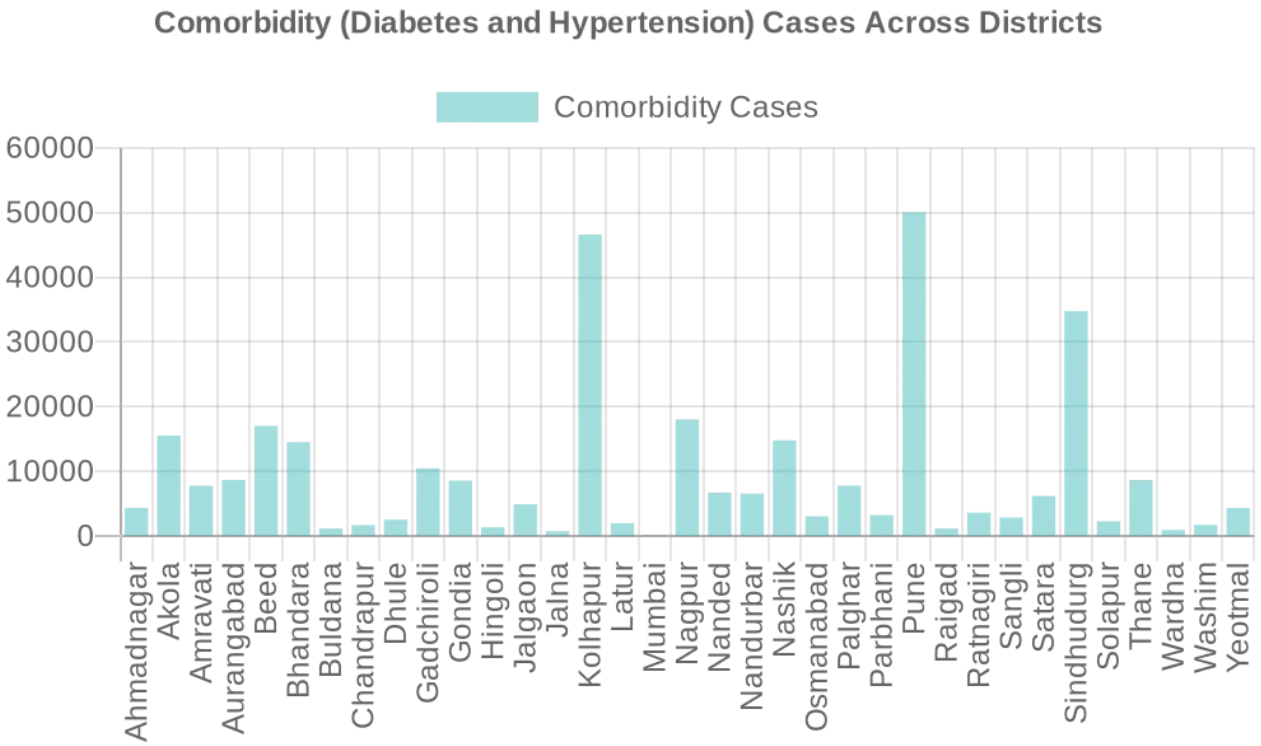
- Discrepancies Across Districts: High ranges and standard deviations across all tables highlight discrepancies in how different districts diagnose these conditions, which may necessitate targeted interventions to ensure more uniform healthcare standards.

- Potential for Over or Under-Diagnosis: The extreme values seen in ranges suggest that some districts might be over-diagnosing or under-diagnosing conditions, which could affect healthcare planning and resource allocation.

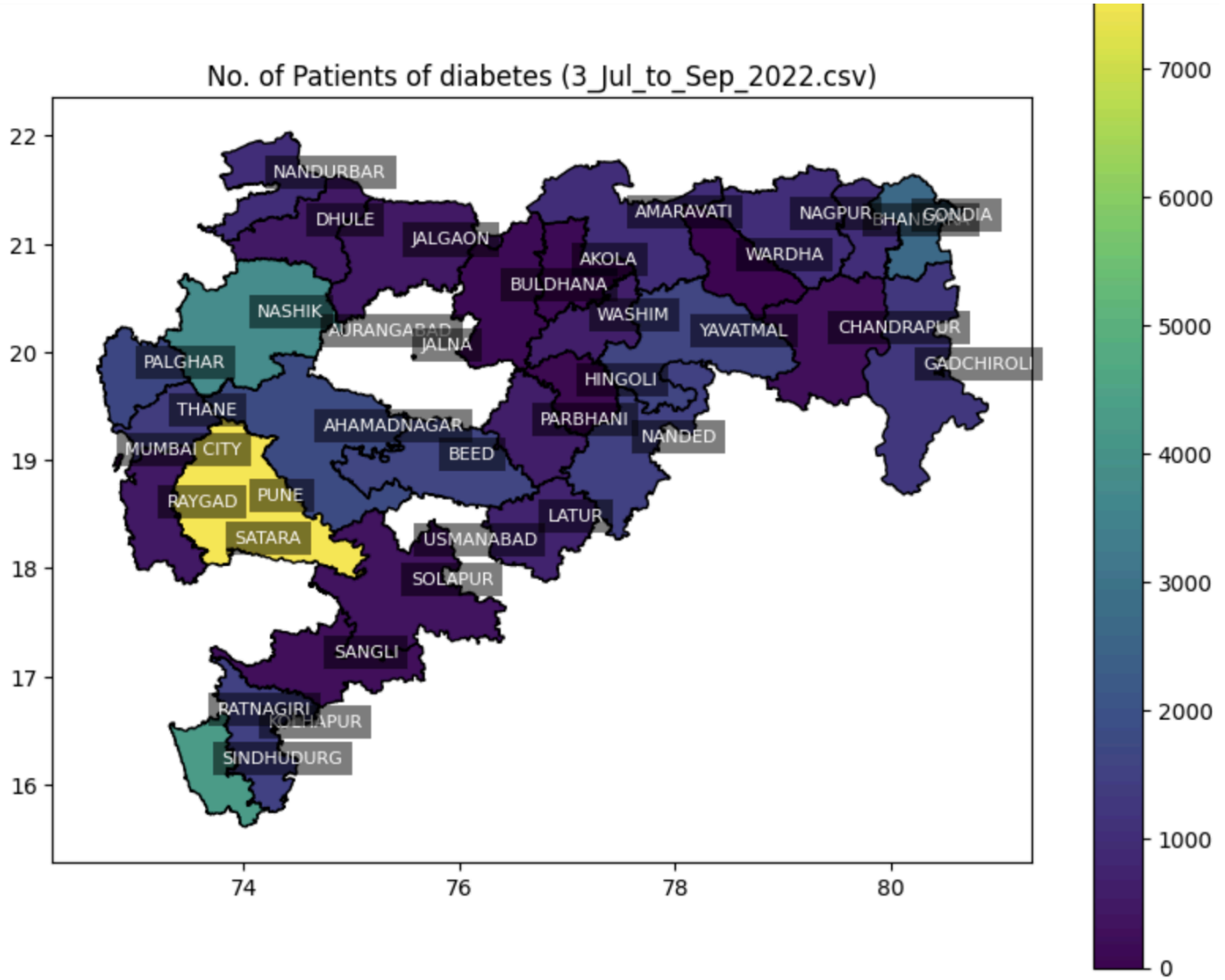
4.3.4 Data Visualization

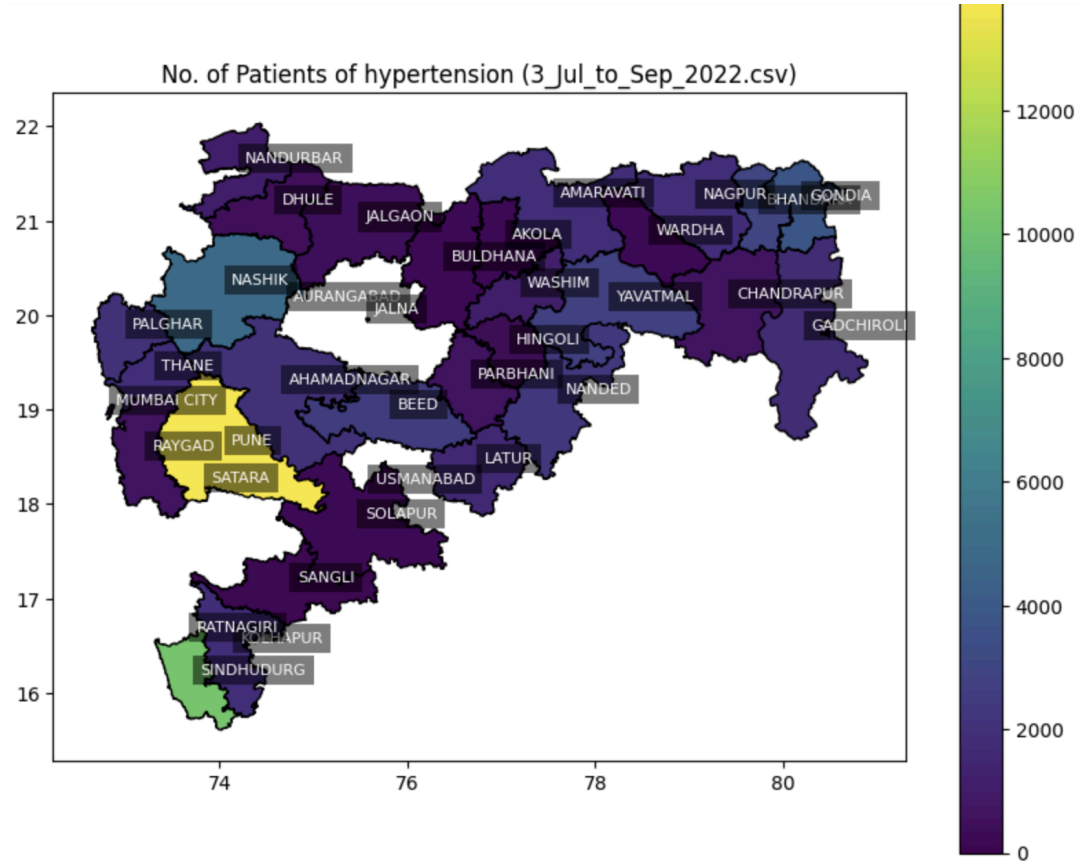


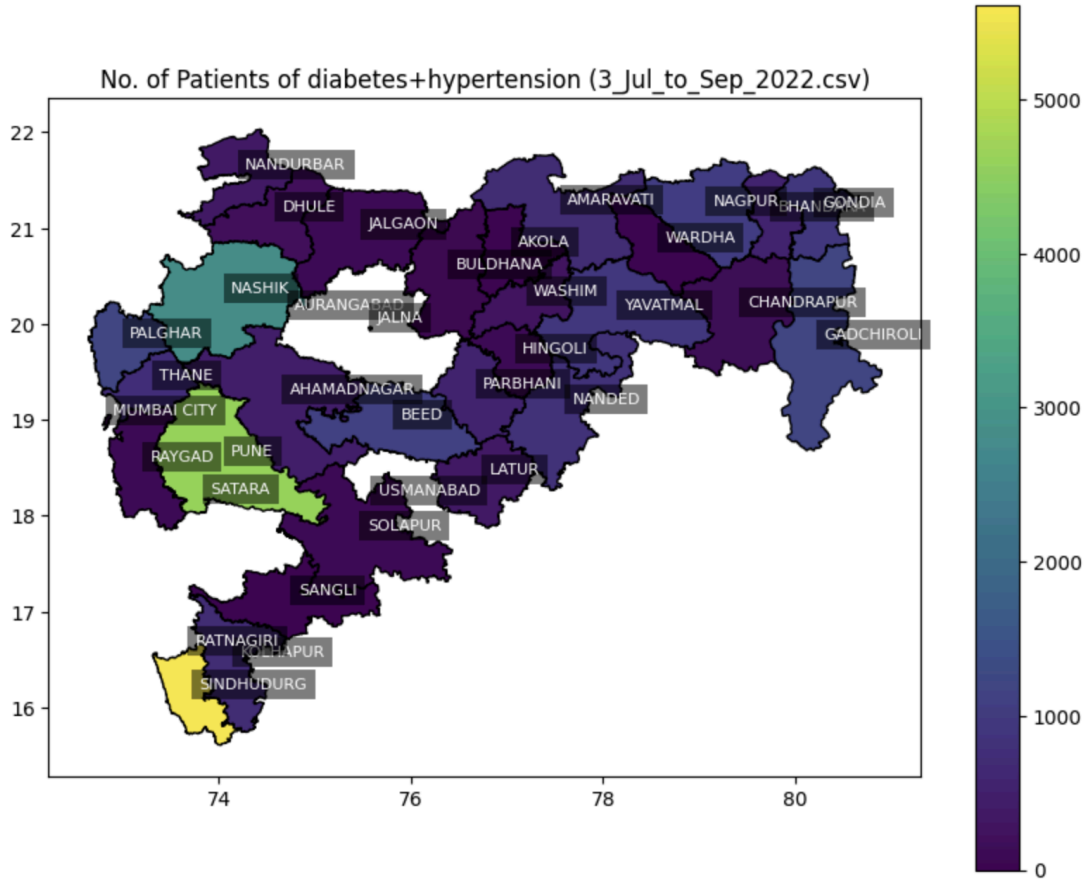

Hypertension Cases Across Districts
Hypertension Cases
140000
120000
100000
80000
60000
40000
20000
0
Ahmadnagar
Akola
Amravati
Aurangabad
Beed
Bhandara
Buldana
Chandrapur
Dhule
Gadchiroli
Gondia
Hingoli
Jalgaon
Jalna
Kolhapur
Latur
Mumbai
Nagpur
Nanded
Nandurbar
Nashik
Osmanabad
Palghar
Parbhani
Pune
Raigad
Ratnagiri
Sangli
Satara
Sindhudurg
Solapur
Thane
Wardha
Washim
Yeotmal


Comorbidity (Diabetes and Hypertension) Cases Across Districts
Comorbidity Cases
60000
50000
40000
30000
20000
10000
0
Ahmadnagar
Akola
Amravati
Aurangabad
Beed
Bhandara
Buldana
Chandrapur
Dhule
Gadchiroli
Gondia
Hingoli
Jalgaon
Jalna
Kolhapur
Latur
Mumbai
Nagpur
Nanded
Nandurbar
Nashik
Osmanabad
Palghar
Parbhani
Pune
Raigad
Ratnagiri
Sangli
Satara
Sindhudurg
Solapur
Thane
Wardha
Washim
Yeotmal

4.3.5 Geospatial Analysis




No. of Patients of hypertension (3_Jul_to_Sep_2022.csv)
NANDURBAR
DHULE
JALGAON
AMARAVATI
NAGPUR
GONDIA
AKOLA
WARDHA
BULDHANA
NASHIK
WASHIM
YAVATMAL
CHANDRAPUR
AURANGABAD
JALNA
PALGHAR
GADCHIROLI
HINGOLI
THANE
AHAMADNAGAR
PARBHANI
NANDED
MUMBAI CITY
BEED
RAYGAD
PUNE
LATUR
SATARA
USMANABAD
SOLAPUR
SANGLI
RATNAGIRI
KOLHAPUR
SINDHUDURG
22
21
20
19
18
17
16
74
76
78
80
12000
10000
8000
6000
4000
2000
0


No. of Patients of diabetes+hypertension (3_Jul_to_Sep_2022.csv)
NANDURBAR
DHULE
JALGAON
AMARAVATI
NAGPUR
GONDIA
AKOLA
WARDHA
BULDHANA
NASHIK
WASHIM
YAVATMAL
CHANDRAPUR
AURANGABAD
JALNA
PALGHAR
GADCHIROLI
HINGOLI
THANE
AHAMADNAGAR
PARBHANI
NANDED
MUMBAI CITY
BEED
RAYGAD
PUNE
LATUR
SATARA
USMANABAD
SOLAPUR
SANGLI
RATNAGIRI
KOLHAPUR
SINDHUDURG
22
21
20
19
18
17
16
74
76
78
80
5000
4000
3000
2000
1000
0

4.3.6 District Ranking

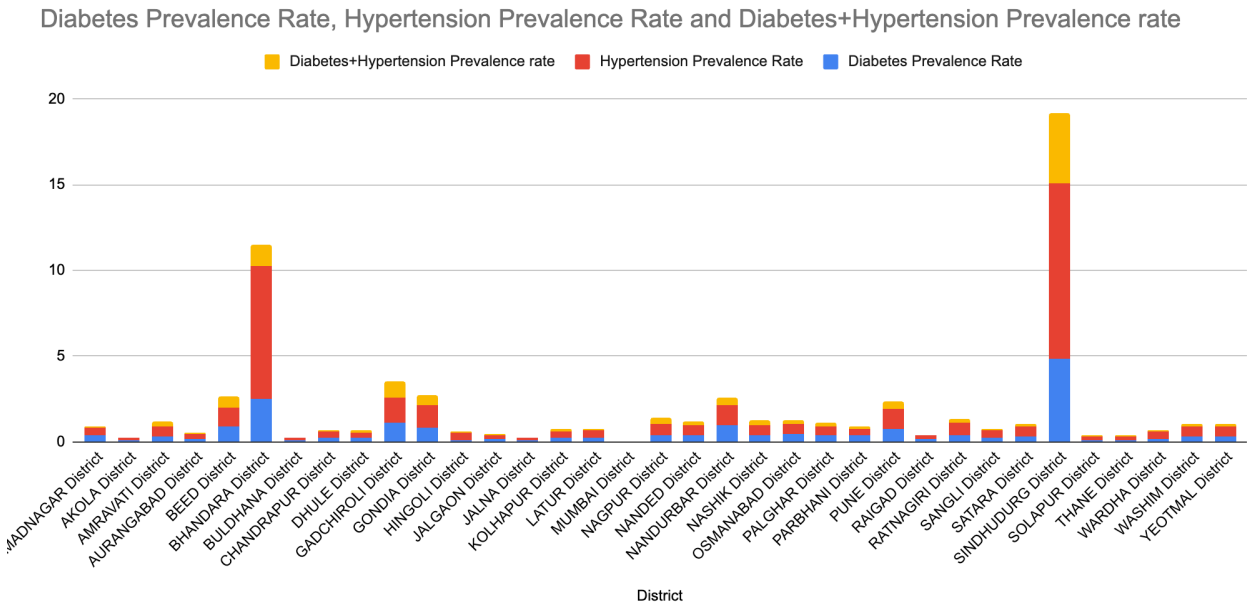
Rank	Diabetes	Hypertension	Diabetes+Hypertension
1	PUNE District	PUNE District	PUNE District
2	SINDHUDURG District	BHANDARA District	SINDHUDURG District
3	BHANDARA District	SINDHUDURG District	NAGPUR District
4	NASHIK District	NASHIK District	BEED District
5	BEED District	BEED District	NASHIK District
6	NAGPUR District	NAGPUR District	BHANDARA District
7	AHMADNAGAR District	NANDED District	GADCHIROLI District
8	NANDURBAR District	AHMADNAGAR District	THANE District
9	NANDED District	NANDURBAR District	GONDIA District
10	THANE District	THANE District	PALGHAR District
11	PALGHAR District	GONDIA District	AMRAVATI District
12	GADCHIROLI District	SATARA District	NANDED District
13	GONDIA District	AMRAVATI District	NANDURBAR District
14	AMRAVATI District	GADCHIROLI District	SATARA District
15	SATARA District	YEOTMAL District	KOLHAPUR District
16	YEOTMAL District	KOLHAPUR District	JALGAON District
17	KOLHAPUR District	PALGHAR District	AHMADNAGAR District
18	OSMANABAD District	RATNAGIRI District	YEOTMAL District
19	AURANGABAD District	SANGLI District	RATNAGIRI District
20	SANGLI District	LATUR District	AURANGABAD District
21	JALGAON District	OSMANABAD District	PARBHANI District
22	PARBHANI District	AURANGABAD District	OSMANABAD District
23	LATUR District	JALGAON District	SANGLI District
24	RATNAGIRI District	CHANDRAPUR District	DHULE District
25	SOLAPUR District	SOLAPUR District	SOLAPUR District
26	DHULE District	WASHIM District	LATUR District
27	CHANDRAPUR District	PARBHANI District	WASHIM District
28	WASHIM District	DHULE District	CHANDRAPUR District
29	RAIGAD District	WARDHA District	HINGOLI District
30	BULDHANA District	RAIGAD District	RAIGAD District
31	WARDHA District	HINGOLI District	BULDHANA District
32	JALNA District	BULDHANA District	WARDHA District
33	AKOLA District	AKOLA District	JALNA District
34	HINGOLI District	JALNA District	AKOLA District
35	MUMBAI District	MUMBAI District	MUMBAI District

4.3.7 District Ranking normalized with Population

Population data was loaded from the 2011 Population Consensus.
Prevalence rate(expressed as %) was calculated as (total cases)/(total population)*100

District	Diabetes Prevalence Rate	Hypertension Prevalence Rate	Diabetes+Hypertension Prevalence rate
AHMADNAGAR District	0.3489202117	0.4414989658	0.0957043326
AKOLA District	0.08964080829	0.1385959361	0.03070721416
AMRAVATI District	0.3388328322	0.5770925186	0.2693490788
AURANGABAD District	0.1868541765	0.239916872	0.09015795068
BEED District	0.8879909046	1.12686452	0.6585948661
BHANDARA District	2.498054708	7.775835726	1.209080139
BULDHANA District	0.1124404448	0.1133684265	0.04446578802
CHANDRAPUR District	0.2004711685	0.3956345464	0.07644125796
DHULE District	0.2326826476	0.3175250212	0.1225338419
GADCHIROLI District	1.096331395	1.445371698	0.9754488127
GONDIA District	0.8077840042	1.295569702	0.6476336231
HINGOLI District	0.09300587338	0.424259669	0.1145798385
JALGAON District	0.1515868988	0.2094367336	0.1158887988
JALNA District	0.09504626231	0.1071950327	0.03823289499
KOLHAPUR District	0.210577861	0.3779926785	0.1389060529
LATUR District	0.2533619972	0.4405516104	0.08096337864
MUMBAI District	0.006113095255	0.00421076841	0.0008229166689
NAGPUR District	0.4065695799	0.599926508	0.3874444781
NANDED District	0.3729815797	0.6215764652	0.1990008604
NANDURBAR District	0.9577169135	1.204032045	0.3976229983
NASHIK District	0.4004953508	0.5864074573	0.242157314
OSMANABAD District	0.4724971887	0.590802473	0.1838226422

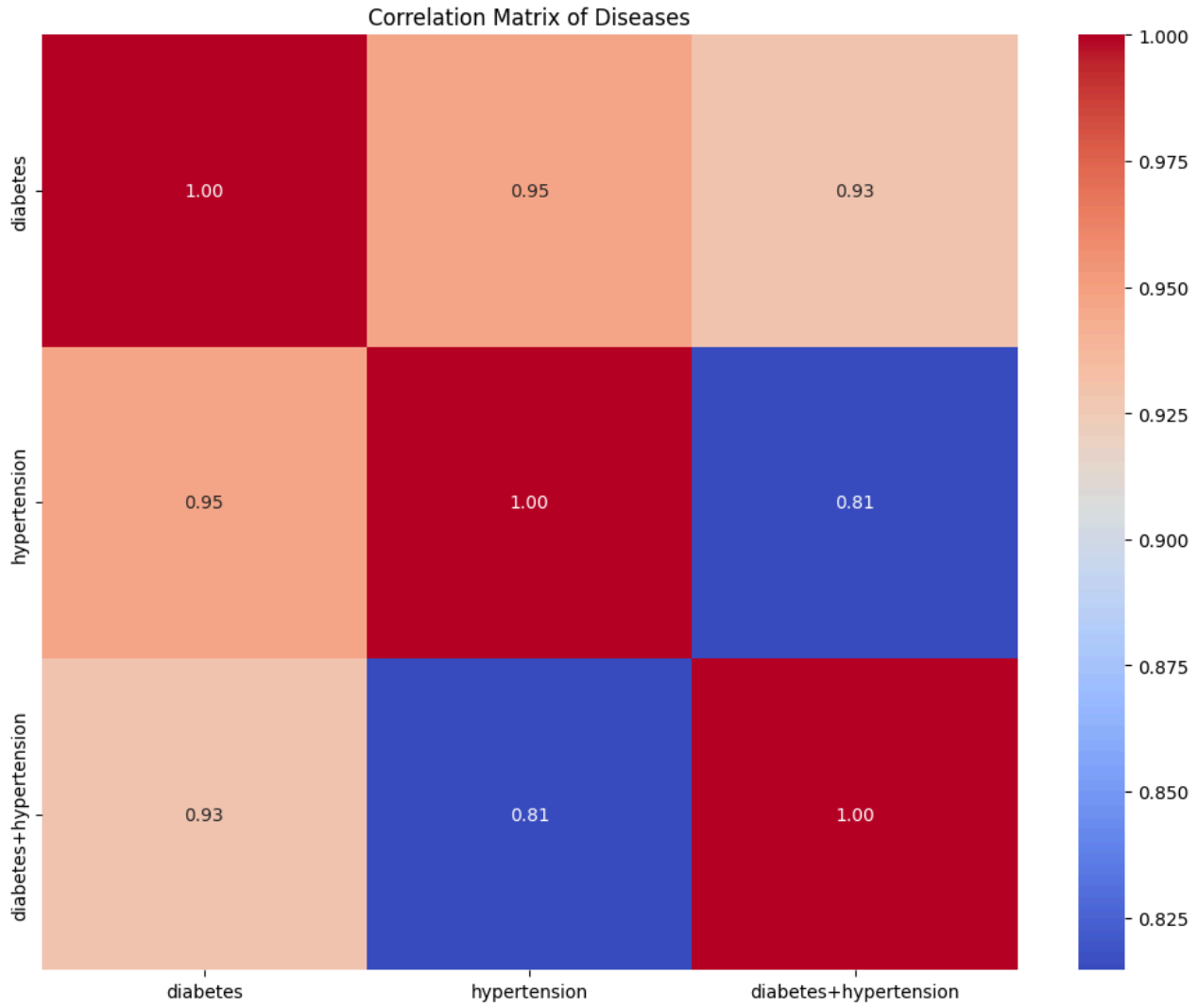
PALGHAR District	0.4134956637	0.4710854027	0.2606587838
PARBHANI District	0.348186305	0.3686101849	0.1759721495
PUNE District	0.7200027828	1.194889435	0.4409078491
RAIGAD District	0.1378027485	0.2155113507	0.04452964847
RATNAGIRI District	0.3605418716	0.7560667687	0.2235817789
SANGLI District	0.2375854094	0.4134801107	0.1010225208
SATARA District	0.3082489469	0.5578377097	0.2060097725
SINDHUDURG District	4.847519746	10.23584978	4.090738433
SOLAPUR District	0.1237448341	0.1771985263	0.05287468768
THANE District	0.1126386374	0.173858433	0.07823584278
WARDHA District	0.1462206348	0.4380468859	0.07211091243
WASHIM District	0.3119048415	0.5925690802	0.1423368639
YEOTMAL District	0.3242377941	0.5469371089	0.156221369



As we can see, **Sindhudurg** and **Bhandara** districts show exceptionally higher prevalence rate in comparison to other districts.

Rank	Diabetes	Hypertension	Diabetes+Hypertension
1	SINDHUDURG District	SINDHUDURG District	SINDHUDURG District
2	BHANDARA District	BHANDARA District	BHANDARA District
3	GADCHIROLI District	GADCHIROLI District	GADCHIROLI District
4	NANDURBAR District	GONDIA District	BEED District
5	BEED District	NANDURBAR District	GONDIA District
6	GONDIA District	PUNE District	PUNE District
7	PUNE District	BEED District	NANDURBAR District
8	OSMANABAD District	RATNAGIRI District	NAGPUR District
9	PALGHAR District	NANDED District	AMRAVATI District
10	NAGPUR District	NAGPUR District	PALGHAR District
11	NASHIK District	WASHIM District	NASHIK District
12	NANDED District	OSMANABAD District	RATNAGIRI District
13	RATNAGIRI District	NASHIK District	SATARA District
14	AHMADNAGAR District	AMRAVATI District	NANDED District
15	PARBHANI District	SATARA District	OSMANABAD District
16	AMRAVATI District	YEOTMAL District	PARBHANI District
17	YEOTMAL District	PALGHAR District	YEOTMAL District
18	WASHIM District	AHMADNAGAR District	WASHIM District
19	SATARA District	LATUR District	KOLHAPUR District
20	LATUR District	WARDHA District	DHULE District
21	SANGLI District	HINGOLI District	JALGAON District
22	DHULE District	SANGLI District	HINGOLI District
23	KOLHAPUR District	CHANDRAPUR District	SANGLI District
24	CHANDRAPUR District	KOLHAPUR District	AHMADNAGAR District
25	AURANGABAD District	PARBHANI District	AURANGABAD District
26	JALGAON District	DHULE District	LATUR District
27	WARDHA District	AURANGABAD District	THANE District
28	RAIGAD District	RAIGAD District	CHANDRAPUR District
29	SOLAPUR District	JALGAON District	WARDHA District
30	THANE District	SOLAPUR District	SOLAPUR District
31	BULDHANA District	THANE District	RAIGAD District
32	JALNA District	AKOLA District	BULDHANA District
33	HINGOLI District	BULDHANA District	JALNA District
34	AKOLA District	JALNA District	AKOLA District
35	MUMBAI District	MUMBAI District	MUMBAI District

4.3.8 Correlation Analysis



In summary, all three conditions - diabetes, hypertension, and comorbidity - show strong positive correlations with each other. Areas with high prevalence rates of one condition are likely to have high prevalence rates of the other conditions as well. This information is valuable for healthcare planning and resource allocation in regions with a high burden of these chronic diseases.

4.3.9 Conclusion

- Trends Over Time: The analysis revealed significant variations in the diagnosis rates of diabetes and hypertension over different periods. Peaks in diagnosis rates were observed, particularly in April to June 2023, indicating potential factors such as public health campaigns, seasonal variations, or changes in healthcare policies.

- Discrepancies Across Districts: High ranges and standard deviations across districts highlighted discrepancies in how different regions diagnose these conditions. This suggests the need for targeted interventions to ensure more uniform healthcare standards and resource allocation.

- Prevalence Rates: Certain districts, such as Sindhudurg and Bhandara, exhibited exceptionally high prevalence rates compared to others, indicating potential areas of focus for healthcare interventions and resource allocation.

- Correlation Analysis: The correlation analysis revealed strong positive correlations between diabetes, hypertension, and comorbidity, suggesting that areas with high prevalence rates of one condition are likely to have high rates of the other conditions as well.

In conclusion, this analysis provides valuable insights into the distribution and prevalence of diabetes and hypertension across districts in Maharashtra. It underscores the importance of targeted interventions, uniform healthcare standards, and resource allocation strategies to address the burden of these chronic diseases effectively.