Development Of An Integrated Dashboard For Public Health Systems

A Research and Development project

by

Akshaan Khan (22B2226)

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.

Analysis was initiated by exploring various data analysis tools and websites to source relevant data. Subsequently, several online datasets were leveraged to conduct preliminary data analysis. Later, district-specific data was acquired from hospitals in Maharashtra and analyzed.

Analysis was expanded 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
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, online datasets were utilized and experimented with various data analysis tools for the analysis.

3.1 Longitudinal Aging Study in India (LASI) data

3.1.1 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

4.1 Cancer Patient load Analysis

4.1.1 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 cancer data. Cancer cases are categorized into four types: Oral, Cervical, Breast, and Others. Here, each unit represents a district.

4.1.2 Data Preprocessing

1. Nine individual datasets, each spanning nine quarters, were consolidated into a single dataset to enable straightforward comparison and analysis.
2. Any missing values within the dataset were uniformly filled with zeros, reflecting the assumption of zero patient count in instances where data was unavailable.
3. Cancer data across various types were aggregated for each unit, allowing for an assessment of the cumulative cancer patient load per unit.

4.1.3 Methodology

The objective of this study was to assess and rank healthcare units (districts) based on their patient load for various types of cancer, thereby inferring the availability of adequate diagnostic facilities within these units. Several methodologies were employed in the analysis process:

1. Quarter-wise Patient Load Analysis: Quarter-wise increases in the number of cancer patients were plotted for each unit.
2. Direct Ranking Based on Cancer Patient Count: Units were ranked directly based on the number of diagnosed patients for each type of cancer within them. The patient load distribution was visualized on a geographical map.
3. Classification based on Overall Cancer Load: Units were classified based on their overall cancer load across all 9 quarters:
    i. High Cancer Load: Units: with consistently high patient load across all types of cancer.
    ii. Medium Cancer Load: Units with varying patient loads, some high and some low, across different types of cancer, or an overall medium load.
    iii. Low Cancer Load: Units where most types of cancer exhibit low patient load.
    iv. Very Low Cancer Load: Units where patient load for all types of cancer is very low.

4.1.4 Classification based on Overall Cancer Load:

The overall cancer load is determined by the total number of diagnosed patients across all nine quarters and for every type of cancer. This classification system allows us to ascertain whether the surrounding population of a given unit experiences a higher prevalence of cancer or if the unit possesses the necessary medical resources for diagnosing various cancer types, as opposed to specializing in just one. This classification was carried out across four distinct classes, as previously outlined.
The steps involved in this analysis are:
Correlation Analysis: Correlation between the patient loads of the four types of cancers was assessed to understand if the load of one cancer type affects that of others.

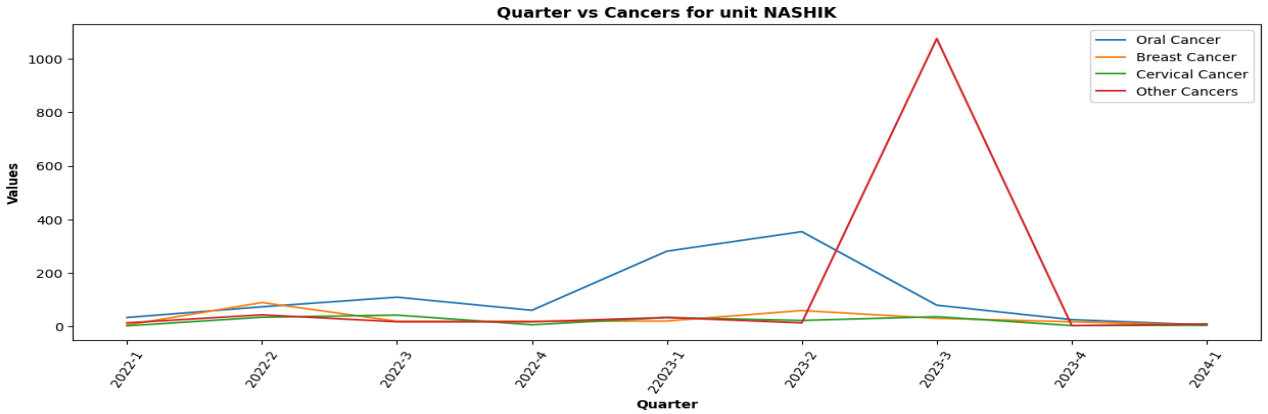
- Principal Component Analysis (PCA): PCA was performed to reduce the dimensionality from 4 to 2 and eliminate the correlation between cancer types.
- K-means Clustering: K-means clustering was applied to classify the units into four categories based on their overall cancer load across the 9 quarters.
- Visualization: The clusters derived from k-means clustering were visualized: Initially on a 2D plot of the principal components. Subsequently, on a geographical map of Maharashtra to provide spatial context to the classification results.

4.1.5 Results:

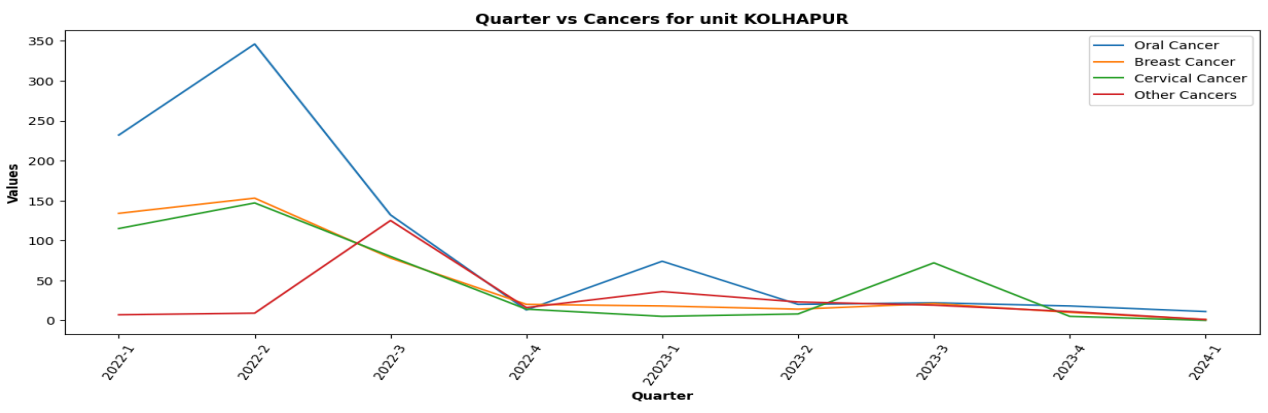
4.1.5.1 Quarter-wise patient load analysis

Plots for districts having the highest total cancer over 9 quarters are (top 3):

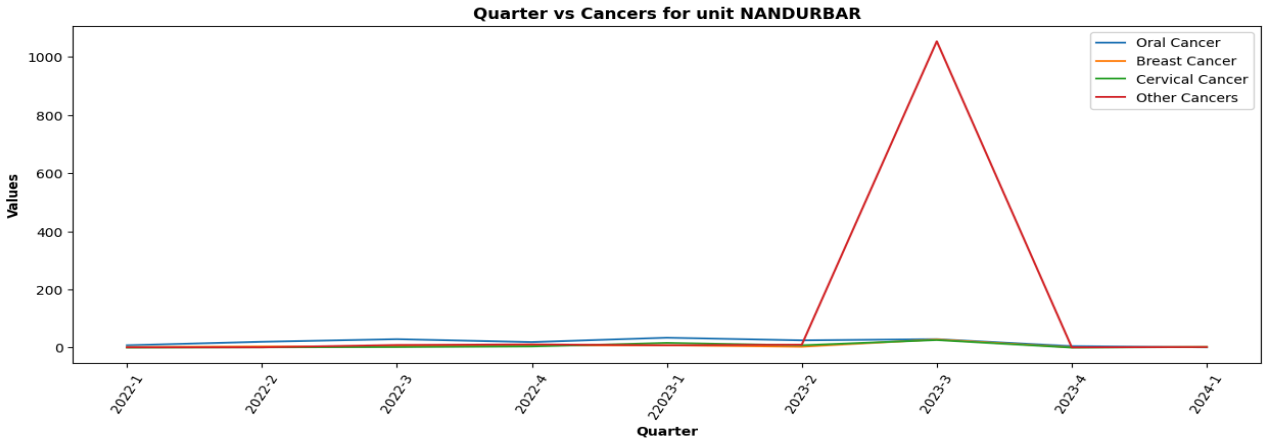
1. Nashik: Total Cancer diagnosed over 9 quarters: 2688



2. Kolhapur: Total Cancer diagnosed over 9 quarters: 2010



3. Nandurbar: Total Cancer diagnosed over 9 quarters: 1385



These results show the distribution of different cancers over all the quarters units-wise.

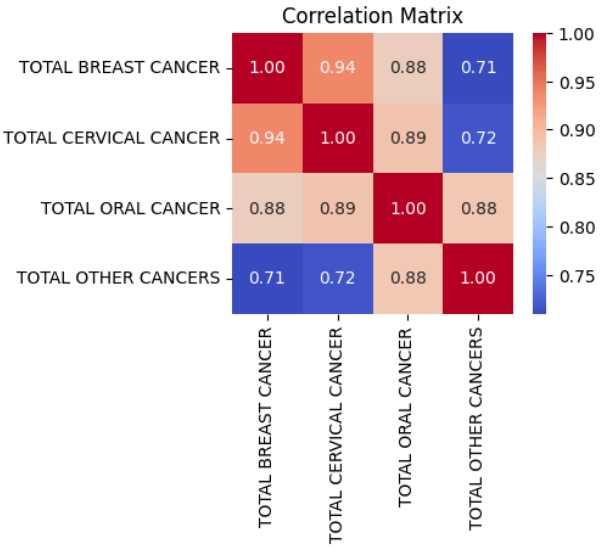
4.1.5.2 Direct Ranking based on Cancer patient count

The table below shows the rankings of districts based on total cancer patients overall in the 9 quarters.

Rank	Unit Name	TOTAL CANCER ALL
1	NASHIK	2688.0
2	KOLHAPUR	2010.0
3	NANDURBAR	1385.0
4	PUNE	1263.0
5	AKOLA	1219.0
6	NAGPUR	998.0
7	AMARAVATI	788.0
8	SATARA	569.0
9	BHANDARA	411.0
10	WASHIM	359.0
11	GADCHIROLI	294.0
12	LATUR	293.0
13	SINDHUDURG	246.0
14	AHAMADNAGAR	223.0
15	AURANGABAD	167.0
16	RATNAGIRI	166.0
17	CHANDRAPUR	163.0
18	NANDED	106.0
19	MUMBAI CITY	105.0
20	WARDHA	87.0
21	THANE	87.0
22	USMANABAD	71.0
23	JALNA	51.0
24	BEED	44.0
25	SOLAPUR	42.0
26	GONDIA	37.0
27	BULDHANA	27.0
28	PARBHANI	23.0
29	DHULE	19.0
30	RAYGAD	17.0
31	JALGAON	5.0
32	YAVATMAL	5.0
33	SANGLI	4.0
34	PALGHAR	1.0
35	HINGOLI	1.0

4.1.5.3 Classification based on Overall cancer load

For the classification, first, the correlation was calculated for all the types of cancers with each other. The correlation matrix is as follows:

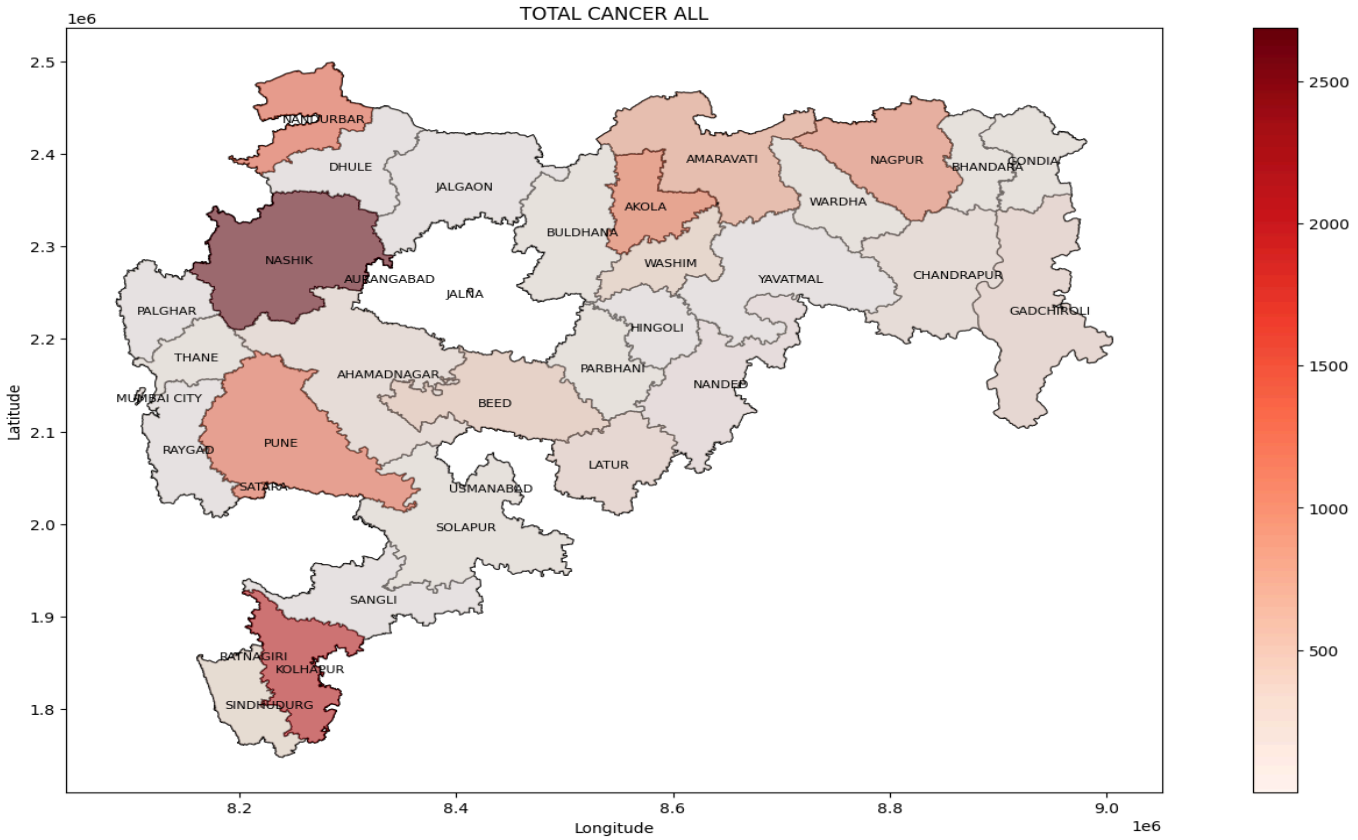


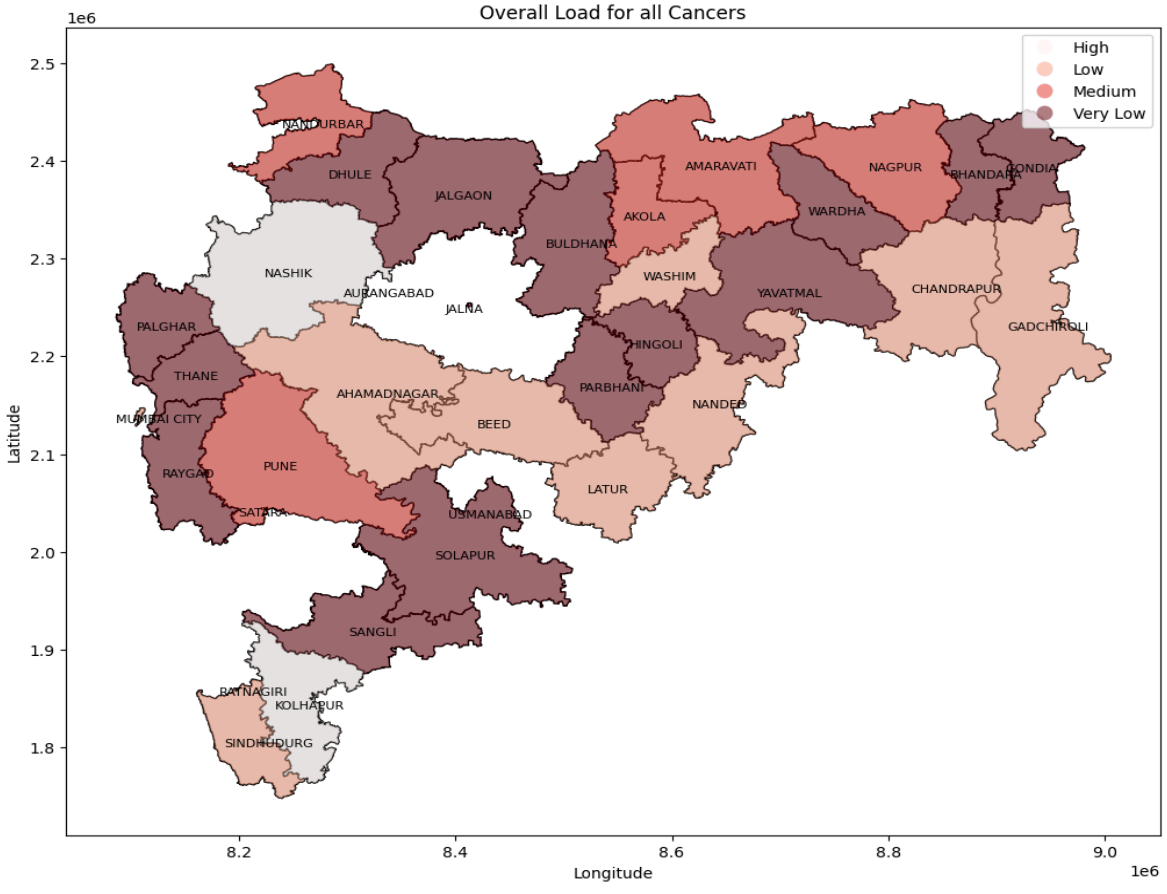
Principal Component Analysis (PCA) was utilized to mitigate correlation and reduce the dimensionality of the dataset from four to two dimensions, achieving a variance capture of 97.1325%. Following dimension reduction, the classification process involved employing k-means clustering with k = 3 to delineate three load classes: High, Medium, and Low. Additionally, a fourth class denoted "Very low load" was manually assigned to units exhibiting a collective patient count of fewer than 100 across all nine quarters. Notably, outliers were identified among the units, with Nashik and Nandurbar standing out due to their notably high patient loads. While Nandurbar, primarily burdened with "Other cancers," was designated as Medium load, Nashik, grappling with high patient volumes across all cancer types, was manually classified as High load. The remaining units underwent classification via k-means clustering, ensuring a comprehensive categorization approach that accounted for both automated and manual determinations based on load characteristics.

Unit Name	TOTAL BREAST CANCER	TOTAL CERVICAL CANCER	TOTAL ORAL CANCER	TOTAL OTHER CANCERS	Overall Load for all Cancers	TOTAL CANCER ALL
NASHIK	263.0	182.0	1019.0	1224.0	High	2688.0
KOLHAPUR	449.0	446.0	868.0	247.0	High	2010.0
AHAMADNAGAR	80.0	30.0	73.0	40.0	Low	223.0
SINDHUDURG	74.0	30.0	68.0	74.0	Low	246.0
SATARA	96.0	65.0	331.0	77.0	Low	569.0
AURANGABAD	36.0	12.0	101.0	18.0	Low	167.0
RATNAGIRI	49.0	7.0	53.0	57.0	Low	166.0
BHANDARA	61.0	30.0	220.0	100.0	Low	411.0
CHANDRAPUR	26.0	16.0	88.0	33.0	Low	163.0
GADCHIROLI	46.0	22.0	171.0	55.0	Low	294.0
NANDED	36.0	6.0	42.0	22.0	Low	106.0
WASHIM	99.0	84.0	115.0	61.0	Low	359.0
MUMBAI CITY	20.0	13.0	47.0	25.0	Low	105.0
LATUR	94.0	58.0	86.0	55.0	Low	293.0
PUNE	340.0	175.0	525.0	223.0	Medium	1263.0
NANDURBAR	59.0	59.0	170.0	1097.0	Medium	1385.0
NAGPUR	160.0	82.0	550.0	206.0	Medium	998.0
AKOLA	390.0	242.0	484.0	103.0	Medium	1219.0
AMARAVATI	284.0	149.0	326.0	29.0	Medium	788.0
JALGAON	0.0	1.0	1.0	3.0	Very Low	5.0
WARDHA	24.0	10.0	37.0	16.0	Very Low	87.0
THANE	18.0	12.0	33.0	24.0	Very Low	87.0
SOLAPUR	4.0	4.0	26.0	8.0	Very Low	42.0
SANGLI	0.0	1.0	3.0	0.0	Very Low	4.0
BEED	18.0	5.0	9.0	12.0	Very Low	44.0
RAYGAD	2.0	1.0	13.0	1.0	Very Low	17.0
BULDHANA	7.0	9.0	2.0	9.0	Very Low	27.0
PARBHANI	3.0	0.0	18.0	2.0	Very Low	23.0
PALGHAR	0.0	0.0	1.0	0.0	Very Low	1.0
USMANABAD	10.0	21.0	21.0	19.0	Very Low	71.0
DHULE	4.0	2.0	9.0	4.0	Very Low	19.0
GONDIA	3.0	4.0	30.0	0.0	Very Low	37.0
HINGOLI	1.0	0.0	0.0	0.0	Very Low	1.0
JALNA	20.0	10.0	9.0	12.0	Very Low	51.0
YAVATMAL	0.0	0.0	5.0	0.0	Very Low	5.0

4.1.5.4 Plotting on the map of Maharashtra

Following are the map plots for Total cancer and Overall cancer loads. Units are represented by regions in which they are located. Note that data for some of the regions are not marked on the map, such as Aurangabad, Jalna etc.





4.1.6 Conclusion:

The analysis reveals several key insights regarding the distribution of patient loads across different districts (units) and their implications for medical facilities and cancer diagnosis:

1. Districts with high patient loads are indicative of robust medical facilities capable of diagnosing and treating various types of cancers effectively. This is evidenced by units such as Nashik and Kolhapur, which exhibit consistently high patient loads across all cancer types, suggesting the presence of accessible and well-equipped medical infrastructure.
2. Conversely, districts with low or very low patient loads may not necessarily possess adequate medical facilities for cancer diagnosis and treatment. Alternatively, patients in these areas may seek care from other private hospitals perceived to offer superior medical services, as observed in major districts like Mumbai City and Thane.
3. Correlation analysis underscores a strong interrelation between different cancer types, indicating that hospitals specializing in the diagnosis and treatment of one cancer type are likely equipped to handle others as well. High correlations, such as those between breast cancer and cervical cancer (0.94) and cervical cancer and oral cancer (0.89), emphasize the interconnectedness of cancer diagnoses and the likelihood of shared medical resources and expertise across multiple cancer types.
4. Nandurbar unit have very high load for Other cancer types and low load for cervical, breast and oral cancers is classified as medium load unit, and it is difficult to say anything about its facilities and alternatives from the data used.

These insights collectively suggest that the distribution of patient loads across districts serves as a proxy for the availability and quality of medical facilities for cancer care. Units experiencing consistently high patient loads likely offer comprehensive cancer diagnosis and treatment services, while lower-load districts may face challenges in providing comparable medical care, potentially necessitating patient reliance on private units with more robust healthcare infrastructure.

4.1.7 Code:

1. Cancer patient load analysis