

Spatial and Demographic Analysis of Hospital Registration Data for Optimizing Healthcare Outreach

A Mid Sem Report for the BDM capstone Project

Submitted by

Name: Yashika Sharma

Roll number: 24F2001055



IITM Online BS Degree Program,

Indian Institute of Technology, Madras, Chennai

Tamil Nadu, India, 600036

Contents

1. Executive Summary and Title	2
2. Proof Of Originality.....	3
2.1. Letter from Organization.....	3
2.2. Images.....	3
2.3. Video	3
3. Metadata	4
4. Descriptive Statistics	5
4.1. Overview of the Dataset	5
4.2. Spatial Distribution.....	5
4.3. Demographic Profile of the Patients.....	6
5. Detailed Explanation of Analysis Process	7
5.1. Data Cleaning and Processing	7
5.2. Filtering the Data	7
5.3. Analysis	8
5.4. Visualization Tools used	8
6. Results and Findings.....	9
6.1. Spatial Distribution and Trends.....	9
6.2. Demographic Patterns and Trends.....	10
6.3. Final Results	11

1. Executive Summary and Title

The BDM Capstone project on “**Spatial and Demographic Analysis of Hospital Registration Data for Optimizing Healthcare Outreach**” aims to provide support to Yatharth Super Specialty Hospital, Faridabad in enhancing patient outreach and footfall through data-driven demographic and spatial analysis. The limited insight on the distribution of patient footfall over various sub-regions and age-groups hampers the hospital’s ability to optimize outreach strategies, allocate resources effectively, and target underserved communities—ultimately limiting its potential to increase revenue and maximize the return on outreach investments.

The project utilizes the Patient Registration Data collected from the hospital to identify the patterns in patient footfall across Faridabad. Throughout this project, Patient Data Confidentiality has been maintained and its access is strictly controlled and limited to authorized personnel only.

The process of Data Cleaning and Analysis was done using Google Colab and Python libraries such as Plotly and Pandas. The insights were made by studying the trends in various data visualizations.

The insights from the analysis process revealed clear patterns in patient footfall seen for different regions and age groups. One key observation was the effect of seasons or festivals on the footfall irrespective of the geographical region or age. Additionally, the temporal trend indicated a significant drop in patient volume in some regions, possibly due to reduced marketing efforts over time. The analysis also helped identify age groups and localities with the least contribution, offering valuable input for planning future outreach programs.

In conclusion, these findings can guide the hospital in implementing targeted strategies to improve patient engagement, increase footfall in underrepresented areas, and ultimately contribute to higher revenue growth and enhanced healthcare accessibility across Faridabad.

2. Proof Of Originality

2.1.Letter from Organization

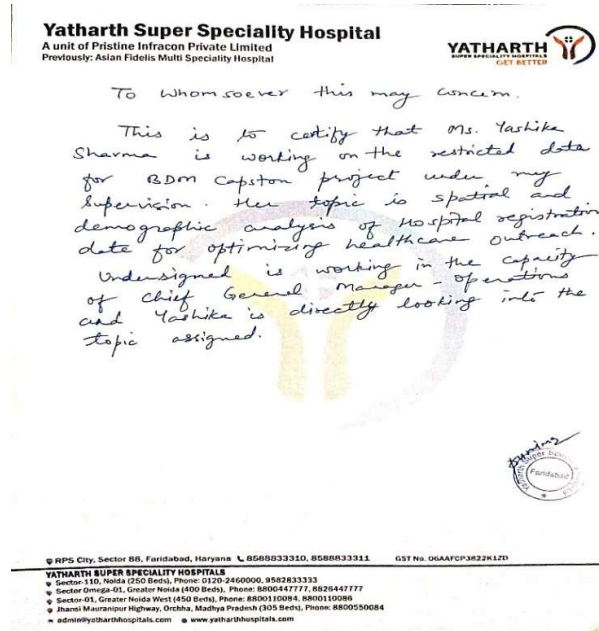


Figure 1: Letterhead from The Organization Written by the CGM-Operations

2.2.Images



Figure 2: Picture with the CGM-Operations at Reception of Yatharth Hospital

2.3.Video

The video link for the interaction with the **Chief General Manager – Operations, Ms. Purnima Rao** is : https://drive.google.com/file/d/18fQ-YlhHiRQeWxIwR3Br6I2u0r-qQxz/_view?usp=sharing

3. Metadata

The project utilizes the Patient Registration Data collected over a period of 9 months (July 2024 – April 2025). The dataset is a time-series data collected in the form of excel sheet by the organization. The original dataset would also include Patient name and personal details which are not shared by the organization due to privacy concerns.

The attributes included in the dataset are:

1. Registration no: Unique identifier assigned to the patient at the time of registration
2. Registration date: Date of Patient registration
3. Sex: Gender of the patient (Male/Female)
4. Date of Birth
5. Age
6. Local Address: Address of the patient recorded in three lines/attributes (local address1, local address2 and local address3)
7. Local pin: Pin code associated with the given local address
8. Local City: City associated with the given local address
9. Local State: State associated with the given local address
10. Permanent Address: A second address, if applicable, of the patient recorded in three lines/attributes (permanent address1, permanent address2 and permanent address3)
11. Permanent Pin: Pin code associated with the given permanent address
12. Permanent City: City associated with the given permanent address
13. Permanent State: State associated with the given permanent address

The data provided by the organization enable the segmentation of patients into groups based on key attributes such as local address, age and sex. The segmented data is further analyzed to study the footfall trends and identify high-demand areas, underserved communities and population groups with low engagement.

SAMPLE DATASET:

registration no	registration date	sex	date of birth	age	local address1	local address2	local address3	local pin	local city	local state	permanent address1	permanent address2	permanent address3	permanent pin	permanent city	permanent state
	2024-07-01	M	1976-06-20	48.0	C-405 EMERALD HEIGHTS SEC-88	NaN	NaN	NaN	FARIDABAD	HARYANA	NaN	NaN	NaN	NaN	FARIDABAD	HARYANA
	2024-07-01	F	2010-04-28	14.0	E-89, 1ST FLOOR, TO THE RETREAT	SEC-89, FBD	NaN	NaN	FARIDABAD	HARYANA	NaN	NaN	NaN	NaN	FARIDABAD	HARYANA
	2024-07-01	F	1974-05-05	50.0	9/208	DDA FLAT	MADANGIR	110062.0	NaN	DELHI	NaN	NaN	NaN	NaN	NaN	DELHI
	2024-07-01	F	1982-01-01	43.0	SHRAMIK VIHAR	SEC 30	NaN	121003.0	FARIDABAD	HARYANA	NaN	NaN	NaN	NaN	FARIDABAD	HARYANA
	2024-07-01	F	1994-09-05	29.0	PLOT NO 46	OLD BHUPANI	NaN	121002.0	FARIDABAD	HARYANA	NaN	NaN	NaN	NaN	FARIDABAD	HARYANA

Figure 3: Sample Patient Registration Data (Registration no. hidden due to privacy concerns)

4. Descriptive Statistics

4.1.Overview of the Dataset

- **Time Period:** July 2024 – April 2025
- **Geographic Scope:** The dataset was filtered to include only the patients residing in Faridabad based on the provided Local Address.
- **Number of Records in the Filtered Dataset:** 15762

4.2.Spatial Distribution

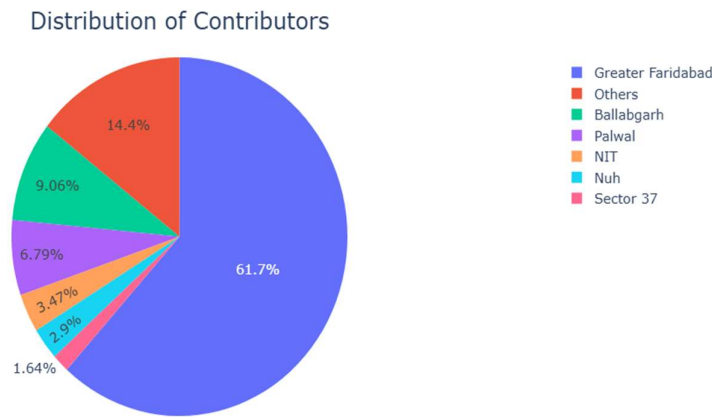


Figure 4: Spatial Distribution of Patients

- **The Filtered Dataset was divided into Sub-Regions:** Greater Faridabad, Ballabgarh, Old Faridabad, Palwal, NIT, Nuh and various sectors of Faridabad.
- The Sectors which give significant contribution to the patient population are Sector 16, Sector 29 and Sector 37.
- **Regions with Most Contribution: Greater Faridabad** (>6000 patients), primarily due to the hospital's location in Sector 88 which increases the accessibility for residents in nearby areas.
- **Other Regions with Significant Contribution: Ballabgarh** (>900 patients), **Palwal** (~700 patients), and **NIT** (~400 patients)
- In Greater Faridabad, **Sector 88** (~ 2900 patients) and **Sector 89** (~ 950 patients) has the highest number of patient footfall.

4.3.Demographic Profile of the Patients

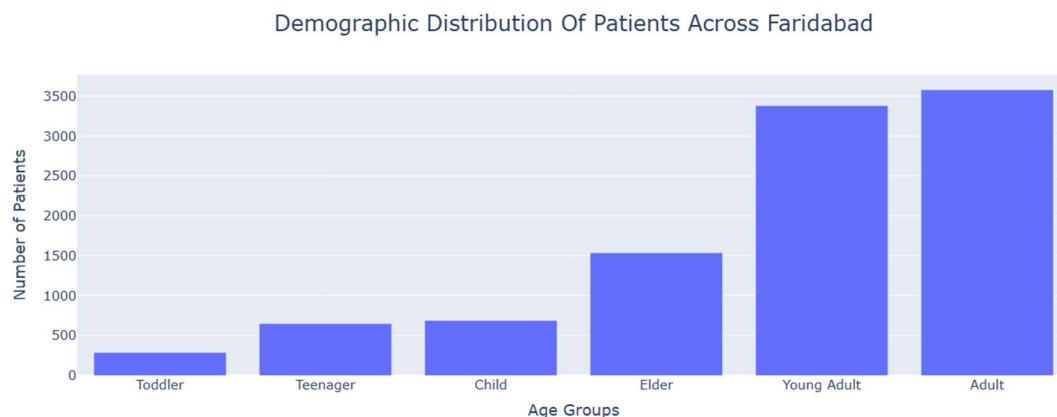


Figure 5: Demographic Distribution of Patients

1. The Filtered dataset was further divided into **6 age groups**:
 1. Toddler (age: 1 - 4 years)
 2. Child (age: 5 - 12 years)
 3. Teenager (age: 13 - 19 years)
 4. Young Adult (age: 20 - 39 years)
 5. Adult (age: 40 - 64 years)
 6. Elder (age: 65 years & above)

2. The data included number of records with empty age field which were excluded in Demographic Analysis
3. **Total number of Patients Records: 10123**
4. **Percentage of Male Population: 49.77%**
5. **Percentage of Female Population: 50.23%**
6. **Age Group with Highest Contribution: Adult** (~3500 patients) and **Young Adult** (~3300 patients)
7. **Age Group with Lowest Contribution: Toddler** (~270 patients) and **Teenagers** (~650 patients)

5. Detailed Explanation of Analysis Process

5.1. Data Cleaning and Processing

The dataset collected from the Patient Registration Data cleaned and standardized to ensure accuracy and consistency during the analysis process. This process involves identifying and dealing with duplicates, missing data and inconsistency in data formats. The steps undertaken were:

- The data was converted into a Data Frame using a Python library: **Pandas**, which provides powerful tools for efficient data cleaning, transformation, and manipulation.
- The data was checked for duplicate records, but no duplicates were found. This showed that each patient entry was unique.
- The Registration date attribute was converted into a Timestamp datatype for handling the time-series data effectively, using **pandas.to_datetime()**.
- The Local Address field was created combining the three local address lines into a single field to allow easy filtering and grouping of data by regions.

5.2. Filtering the Data

- The Data was filtered to include the records within the desired time frame (July 2024 - April 2025) and desired geographic region (Faridabad) to ensure that the analysis is focused on the relevant time period and region.

- **Geographic Segmentation:** The Filtered Data was grouped into sub-regions by searching the Local Address Field for Keywords such as ‘GR FBD’ and ‘GR FARIDABAD’ for Greater Faridabad and ‘BLB’ and ‘BALLABGARH’ for Ballabgarh. The local pincode could not be used for the same due to a lot of empty fields.
- **Demographic Segmentation:** The Filtered Data was also grouped into age groups excluding the records with empty age fields.

5.3. Analysis

- **Spatial analysis** was conducted to identify regions with the highest contribution to the patient footfall, particularly Sector 88 and Sector 89.
- **Demographic analysis** showed that the majority of the population belonged to the Young Adult and Adult age group, while Toddlers and Teenagers formed the minority.
- **Temporal Analysis** was conducted on the data to identify the trends in footfall, highlight peak periods and study the potential seasonal changes.

5.4. Visualization Tools used

The Data Visualization was done using Python libraries and Google Colab.

- **Plotly.express:** Used for making simple bar charts and line graphs to show the basic visualization of data.
- **Plotly.graph_objects:** Used to make complex graphs essential for comparing trends and customizing the graphs as required.

6. Results and Findings

6.1. Spatial Distribution and Trends

1. Patient Footfall from Regions with Moderate Contribution:

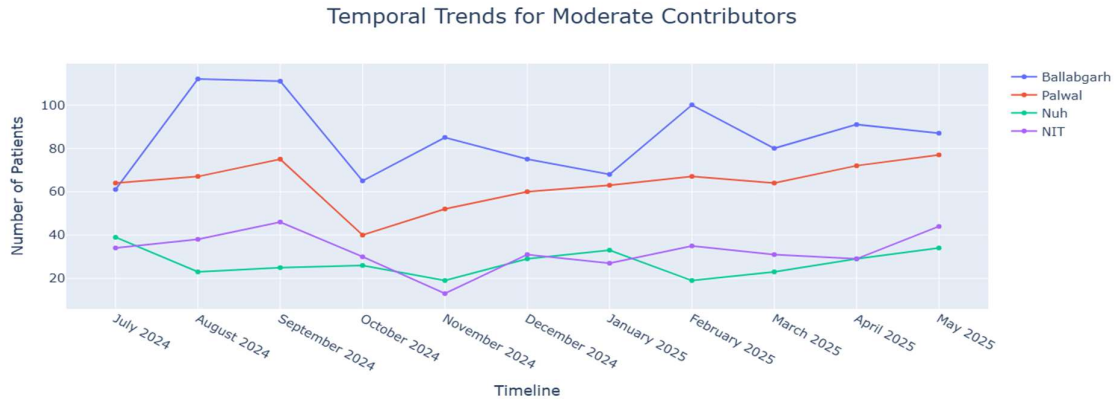


Figure 6: Patient Footfall from Regions with Moderate Contribution

- The Line Graph shows the fluctuations in the footfall of patients from Moderate Contributors: Ballabgarh, Palwal, Nuh and NIT
- The patient footfall, irrespective of the regions, has a common pattern showing a dip in the number of patients during October and March, which is due to the occurrence of festivals- Diwali and Holi.
- Significant drops in Ballabgarh are observed in October 2024 and March 2025, however unlike the other regions the number of patient footfall has not been replenish over time. This could be due to lack of consistency in patient outreach programs in Ballabgarh.

2. Patient Footfall from Low Contributors

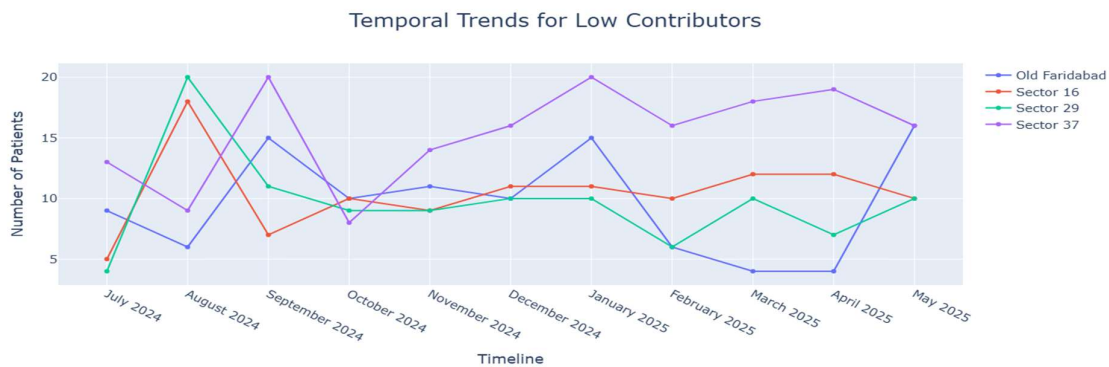


Figure 7: Temporal Trend from Regions with Low Contribution

- Due to the smaller number of patients coming from the regions, the temporal trend is highly fluctuating over the time period.
- Therefore, Old Faridabad, Sector 16, Sector 29, and Sector 37 require strategic and focused outreach initiatives to enhance the number of incoming patients.

6.2. Demographic Patterns and Trends

1. Gender-Age Distribution of the Patient Population

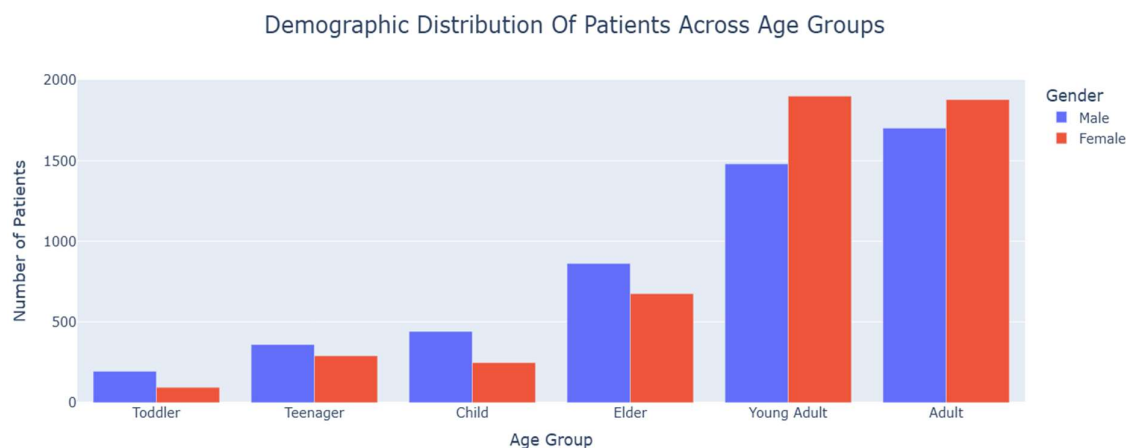


Figure 8: Age-Gender Distribution

- The population of Female patients was found to be significantly low in Age groups: Toddler and Child with respect to the Male Population.
- Overall, the Toddler, Child, and Teenager age groups contribute minimally to the total patient population relative to other age segments.
- This indicates a need for targeted outreach and awareness campaigns to improve healthcare engagement among these underrepresented age and gender groups.

6.3.Final Results

The insights drawn from the Data Analysis highlights critical areas for improvement in patient outreach and engagement for enhancing the revenue of Yatharth Super Specialty Hospital, Faridabad. Regions with moderate and low footfall require targeted and consistent outreach efforts to stabilize and increase patient inflow. Ballabgarh has shown a significant decline in patient footfall with low signs of recovery, indicating the need for renewed and consistent engagement strategies in that area. Additionally, the underrepresentation of younger age groups - particularly females - points to the need for age- and gender-specific awareness campaigns. By addressing these spatial and demographic gaps, the hospital can enhance its healthcare accessibility, improve patient coverage across underserved areas, and ultimately drive more equitable and effective community health outcomes.