CPE 031

Exploratory Data Analysis of Healthcare Distribution by Region

Submitted by:

Esteban, Prince Wally G.
Fernandez, Don Eleazar T.
Sanchez, Christan Ray R.
Tandayu, Leoj Jeam B.
Valleser, Kenn Jie L.

Bachelor of Science in Computer Engineering
2nd year, CPE21S4

Submitted to:

ENGR. MARIA RIZETTE H. SAYO

Table of Contents

I. Introduction and Purpose of the Analysis

INTRODUCTION

Healthcare encompasses the entire system and industry dedicated to providing medical services to individuals and communities. It includes not only the medical services themselves but also the infrastructure, policies, and entities that support the delivery of these services. Understanding healthcare is crucial for patients and caregivers to navigate the system effectively and advocate for high-quality care. This article will explore the components of healthcare, explain the difference between healthcare and health care, provide examples, and highlight how patients and caregivers can optimize their interactions with the healthcare system.

Healthcare refers to the comprehensive system and industry involved in the provision of medical services. It includes hospitals, clinics, healthcare professionals, insurance companies, pharmaceutical companies, public health organizations, and the regulatory bodies that govern health services. Healthcare encompasses everything from policy and infrastructure to the direct care provided to patients (*What Is Healthcare?*, n.d.).

In the 21st Century, health care systems across the world are focusing policy efforts on improving the quality of healthcare delivered to their population. In contrast, healthcare quality improvement in earlier time periods arose from a series of seemingly unrelated incidents and developments. In this paper, we sequentially review key international historical events that improved health care quality during the years 1860–1960, including innovation in health care financing, care delivery and workforce diversity. The modern nursing workforce of today continues to encounter many of these same challenges across the globe (Sheingold & Hahn, 2014).

SIGNIFICANCE OF THE STUDY

Health care is the sum of preventive services and measures provided by the Directorate of Basic Health Care and its affiliated institutions to all members of the society at large in order to raise the health level of the community and prevent diseases and spread, including the provision of services that help to improve the general health level. Keep the environment clean and the water safe from pollution. Caring for mothers during pregnancy and childbirth, taking care of feeding children and vaccinating them with the necessary vaccines on time (Importance of Health Care, 2019).

Access to healthcare is a fundamental necessity, and pharmacies play a vital role in delivering essential medications to each community. This analysis aims to evaluate the distribution of pharmacies across different regions, identify potential disparities, and provide insights for improving regional healthcare infrastructure.

STATEMENT OF THE PROBLEM

The primary objective of this research is to analyze healthcare data to determine:

1. Which regions have the highest and lowest number of healthcare services and professionals, such as for hospitals, clinics, and pharmacies.
2. Identify disparities in healthcare services and professionals distribution across the regions.
3. How this analysis can guide decisions for equitable healthcare access.

By addressing these questions, this study contributes to identifying underserved areas and informing policies for equitable distribution of healthcare resources.

II. Data Dictionary

Column Name	Description
Unnamed: 0	Index or identifier column, likely redundant for analysis
adm1_en	Name of the first-level administrative division (e.g., region)
adm1_pcode	Code for the first-level administrative division
adm2_en	Name of the second-level administrative division (e.g., region)
adm2_pcode	Code for the second-level administrative division
amd3_en	Name of the third-level administrative division (e.g., region)
adm3_pcode	Code for the third-level administrative division
amd4_en	Name of the fourth-level administrative division (e.g., region)
adm4_pcode	Code for the fourth-level administrative division
brgy_total_area	Total area of the barangay

Table 1.1 : Location Dataset

Table 1.1, the Location Dataset, has 879 rows and 10 columns, it contains information about different administrative divisions and their geographic details. Each row represents a specific location, starting from the first-level administrative division (like regions) down to the fourth level (barangays). The dataset includes columns such as “adm1_en” and “adm1_pcode” for the name and code of the first-level division, all the way to “adm4_en” and “adm4_pcode” for barangay names and codes. It also has a column called “brgy_total_area”, which provides the total area of each barangay, probably in square kilometers. There's also an “Unnamed: 0” column that seems to be an extra index or placeholder.

Column Name	Description
uuid	Unique identifier for each data record.
adm4_pcode	Code for the fourth-level administrative division, matching the Location.
date	Date when the data was collected or entered.
freq	Frequency of visits or occurrences (context-dependent).
clinic_count	Number of clinics in the area.
clinic_nearest	Distance to the nearest clinic (unit not specified).
dentist_count	Number of dentists in the area.
dentist_nearest	Distance to the nearest dentist.
doctors_count	Number of doctors in the area.
doctors_nearest	Distance to the nearest doctor.
hospital_count	Number of hospitals in the area.
hospital_nearest	Distance to the nearest hospital.
optician_count	Number of opticians in the area.
optician_nearest	Distance to the nearest optician.
pharmacy_count	Number of pharmacies in the area.
pharmacy_nearest	Distance to the nearest pharmacy.

Table 1.2 : OSM POI Health Dataset

Table 1.2, the OSM POI Health Dataset, has 7,911 rows and 16 columns, focusing on health-related information for various locations. It includes a uuid column as a unique identifier and an “adm4_pcode” column to match the data with the barangays in the Location Dataset. The dataset also has a date column for when the data was collected, and a freq column, which might indicate the frequency of visits or records. It provides details about healthcare facilities, including the number of clinics, dentists, doctors, hospitals, opticians, and pharmacies, along with how far away the nearest one is (e.g., “clinic_count” and “clinic_nearest”). This dataset helps in understanding the availability and accessibility of health services across different areas. Both datasets are useful for studying spatial and healthcare-related patterns.

III. Analysis Process

METHODOLOGY

Each datasets utilized were assessed to determine its relevance to the research objectives. The necessary datasets were imported into Python Colab for analysis. The chosen datasets were examined to identify their structure, including the number of rows, columns, and variable types.

The two datasets, “Location Dataset” and “OSM POI Health Dataset”, were merged with the “adm4_pcode” column. This integration allowed pharmacy counts and other health metrics from the health dataset to be associated with corresponding regional attributes in the location dataset. After ensuring the data consistency, the merged dataset was grouped by “adm1_en” (regional names) to calculate the total pharmacy count for each region.

For visualization, the bar charts was utilized to display the pharmacy distribution across regions. And it was chosen for their effectiveness in representing categorical data and their ability to highlight disparities visually. The chart included features such as rotated x-axis labels for readability, gridlines to enhance interpretability, and color differentiation to improve visual appeal.

STATISTICAL TREATMENT

The study involved the use of descriptive statistics to analyze the number of professionals and facilities in each region. The aim of descriptive statistics is to summarize categorical and numerical data in an informative way, which is used to help describe key features or characteristics of data (Green et al., 2022).

IV. Analysis and Insights

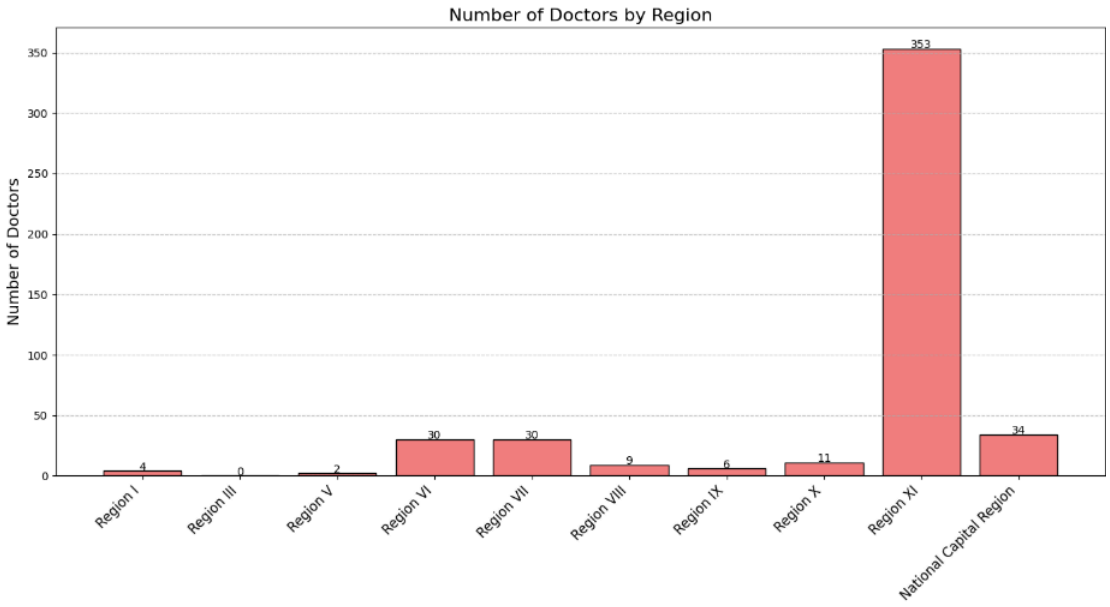

Figure 1.1: Number of Doctors by Region

Figure 1.1 displays the number of doctors by region with Region XI has the highest number of doctors with 353, followed by NCR with 34, Region VI and VII both have 30, Region X with 11, Region VIII has 9, Region IX has 6, Region I has 4, and Region V has 2, while Region III has none.

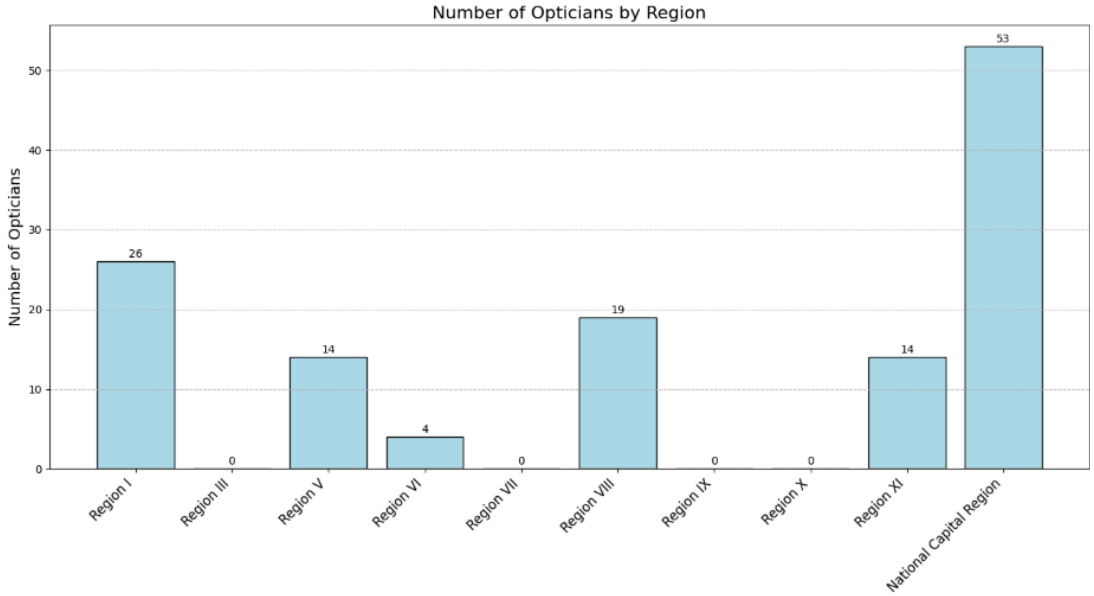

Figure 1.2: Number of Opticians by Region

Figure 1.2 displays the number of opticians by region with NCR leads in the number of opticians with 53, followed by Region I with 26, Region VIII has 19, Region V and XI both have 14, and Region VI has 4, while Regions III, VII, VIII, IX, and X have none.

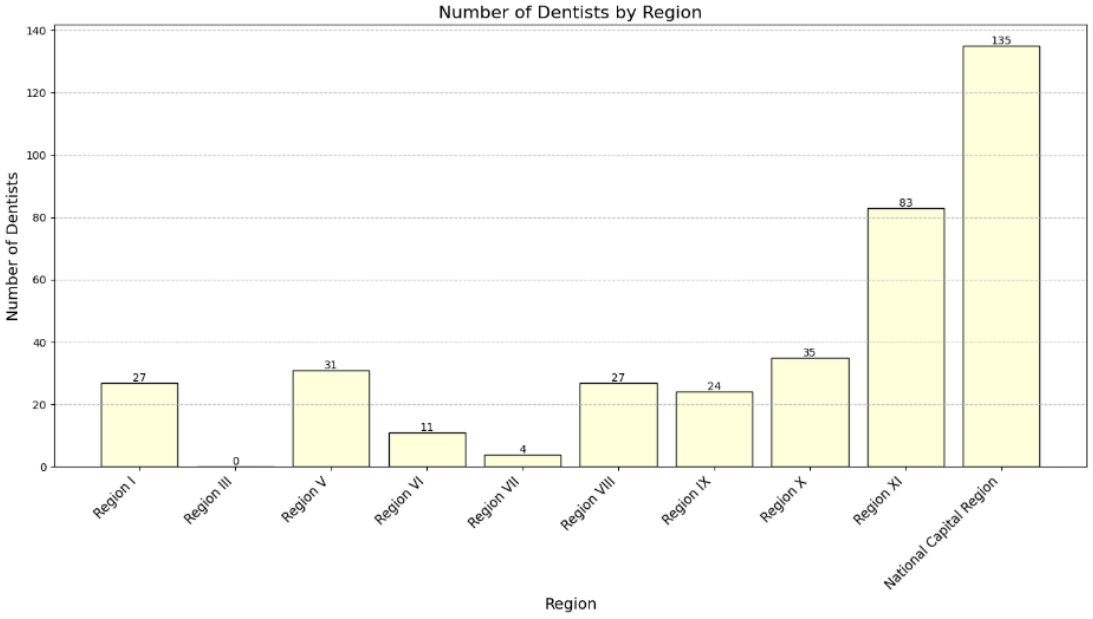

Figure 1.3: Number of Dentists by Region

Figure 1.3 displays the number of dentists by region with NCR has the most dentists with 135, followed by Region XI with 83, Region X with 35, Region V has 31, Region I and Region VIII both have 27, Region IX has 24, and Region VII has 4, while Regions III has none.

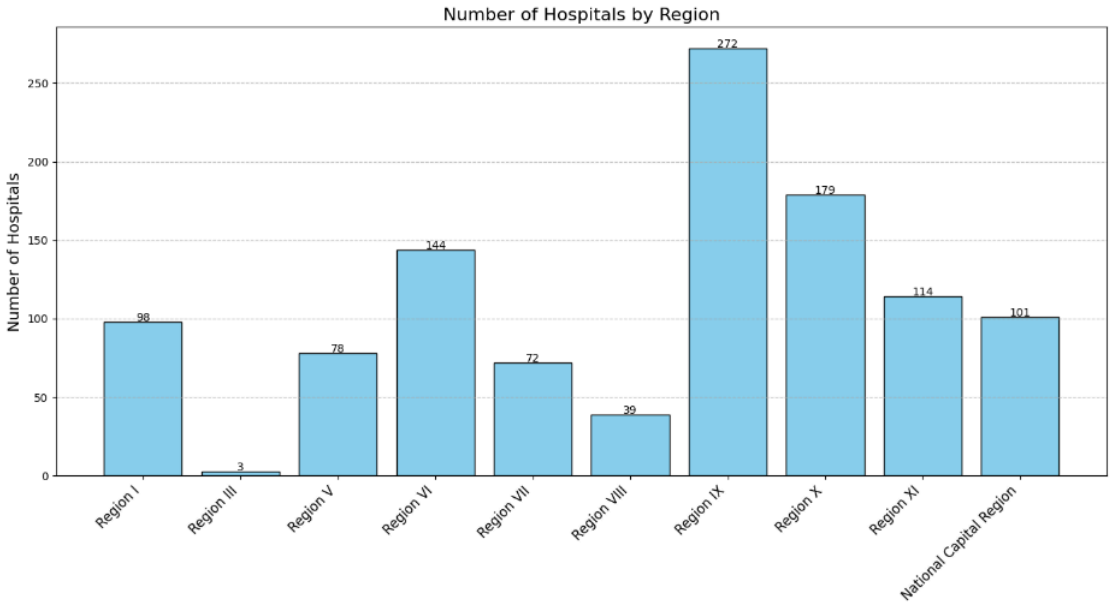

Figure 1.4: Number of Hospitals by Region

Figure 1.4 displays the number of hospitals by region with Region IX has the highest number of hospitals with 272, followed by Region X with 179, Region VI has 144, Region XI has 114, NCR has 101, Region I has 98, Region V has 78, Region VII has 77, Region VIII has 39, and Region III has 3.

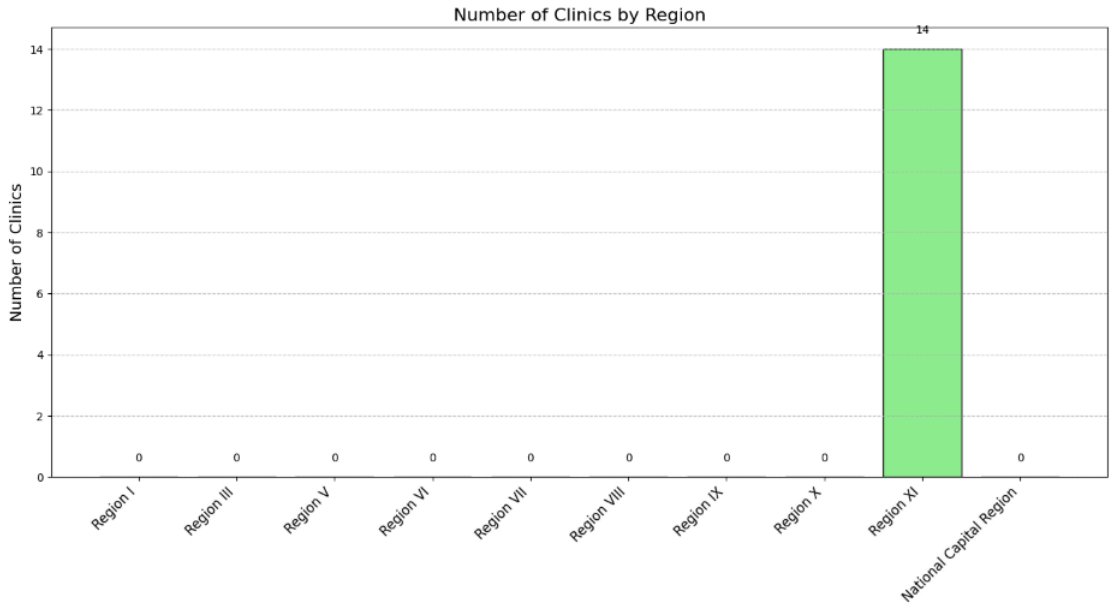

Figure 1.5: Number of Clinics by Region

Figure 1.5 displays the number of clinics by region with Region XI has the highest number of clinics with 14, while all other regions, Region I, Region II, Region III, Region IV, Region V, Region VI, Region VII, Region VIII, Region IX, Region X, Region XI, and NCR have none.

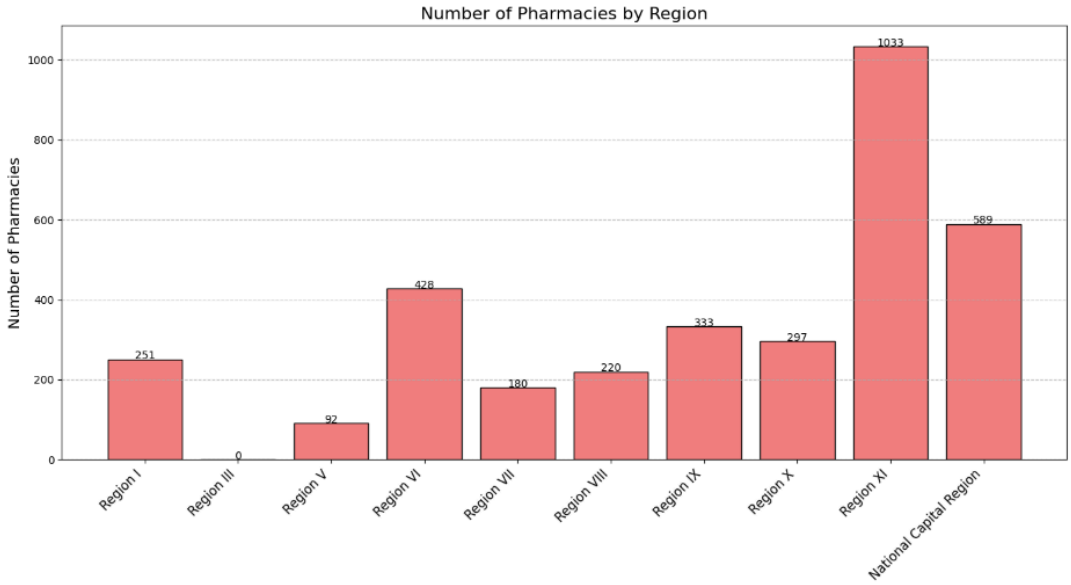

Figure 1.6: Number of Pharmacies by Region

Figure 1.6 displays the number of pharmacies by region with Region XI has the highest number of pharmacies with 1,033, followed by NCR with 589, Region VI has 428, Region IX has 333, Region X has 297, Region I has 251, Region VIII has 220, Region VII has 180, and Region V has 92, while Region III has none.

Research Question: Which regions have the highest and lowest number of healthcare services and professionals, such as for hospitals, clinics, and pharmacies.

Figure 1.1 shows that Region XI has the highest number of doctors with 353, while Region III has none. Figure 1.2 reveals that NCR leads in the number of opticians with 53, while Regions III, VII, VIII, IX, and X have none. In Figure 1.3, NCR has the most dentists with 135, while Region III has none. According to Figure 1.4, Region IX has the highest number of hospitals with 272, while Region III has the lowest with 3. Figure 1.5 shows that Region XI has the highest number of clinics with 14, while all other regions (Region I, Region II, Region III, Region IV, Region V, Region VI, Region VII, Region VIII, Region IX, Region X, and NCR) have none. Lastly, Figure 1.6 demonstrates that Region XI has the highest number of pharmacies with 1,033, while Region III has none. Region XI stands out across multiple healthcare categories, with a high count of doctors, dentists, clinics, and pharmacies, suggesting strong healthcare infrastructure and professional availability. In contrast, Region III faces significant

healthcare gaps, with no doctors, opticians, dentists, clinics, or pharmacies, highlighting critical deficiencies in both manpower and infrastructure. NCR leads in opticians and dentists, reflecting urbanized access to specialized care, while several regions, including III, VII, VIII, IX, and X, lack opticians, showing underserved areas. The abundance of hospitals in Region IX points to robust healthcare infrastructure, whereas limited hospital presence in Region III highlights the severe healthcare limitations. Overall, the disparities in these regions emphasize the need for targeted interventions to address these healthcare gaps and improve access to services.

Research Question: Identify disparities in healthcare services and professionals distribution across the regions.

The analysis identified disparities in pharmacy distribution across regions. Region XI emerged as the region with the highest number of pharmacies, having approximately 1,000 facilities. It was followed by the National Capital Region (NCR) with around 600 pharmacies. These regions benefit from higher levels of urbanization and larger population densities, leading to better access to healthcare resources. On the other hand, Region VIII had the lowest number of pharmacies, indicating potential healthcare access challenges. Other regions, such as Region V, also demonstrated relatively low pharmacy counts. These findings highlight a clear divide between urban and rural areas, with urbanized regions having significantly better access to healthcare infrastructure.

The observed disparities across regions highlight the significant imbalances in healthcare infrastructure and the availability of professionals, this reflect on a need for targeted interventions in underserved areas. The organizations should focus on building more healthcare facilities, encouraging professionals to work in these regions, and making sure medicines and equipment are available.

Research Question: How can this analysis guide decisions for equitable healthcare access?

Based on Figures 1.1, 1.2, and 1.3, it highlights that some regions lack sufficient healthcare manpower, while Figures 1.4, 1.5, and 1.6 highlights the lack of infrastructure in certain areas. These two issues are closely connected, as the lack of healthcare facilities directly impacts the availability of job opportunities in the sector. This emphasizes the need for government intervention to address these problems by investing in both infrastructure and manpower development to ensure accessible healthcare across all regions.

V. Conclusion and Recommendation

CONCLUSION

The study demonstrated an uneven distribution of healthcare resources distribution across regions, with urban areas such as Region XI and NCR having abundant resources, while rural regions like Regions III and VIII face severe shortages. Region XI stands out with the highest number of doctors (353), clinics (14), and pharmacies (1,033), while Region III lacks any doctors, clinics, or pharmacies. Similarly, NCR leads in opticians (53) and dentists (135), but several regions, including III, VII, VIII, IX, and X, have no opticians. Region IX has the highest

number of hospitals (272), whereas Region III has only three. These results reveal not just a shortage of physical healthcare infrastructure in certain areas but also significant gaps in healthcare professionals, which intensifies the disparities. In particular, the lack of healthcare infrastructure in rural areas limits employment opportunities for healthcare professionals and creates a cycle of limited service provision, making it difficult for disadvantaged regions to attract healthcare workers. This problem is especially evident in Region III, which faces the most critical gaps in both healthcare services and personnel.

To address these disparities, the study suggests prioritization of investments in healthcare infrastructure and personnel in disadvantaged regions like Region III. The policymakers are encouraged to develop programs that motivate healthcare professionals to serve in rural areas through financial benefits or career development opportunities. Additionally, addressing population density and geographical challenges is crucial to ensuring that healthcare services are accessible to all. By targeting these gaps, healthcare access and outcomes can be significantly improved, promoting equity across all regions.

RECOMMENDATION

For Government

- It is recommended that policymakers focus on improving healthcare infrastructure in underserved regions. Initiatives could include providing incentives for pharmacies to establish operations in rural areas and increasing investments in healthcare facilities. Additionally, further study should examine underlying factors contributing to these disparities, such as socioeconomic conditions, population density, and transportation accessibility. By addressing these challenges, stakeholders can make meaningful strides toward achieving equitable healthcare access nationwide.

For Future Researchers

- Future researchers are encouraged to address the limitations of this study by gathering more comprehensive and accurate data. It is essential to include a broader range of respondents, as this analysis reveals discrepancies, such as mismatches between infrastructure data and personnel counts. For instance, some regions may have infrastructure but lack reported personnel because not all doctors, clinics, or facilities contributed data. By ensuring a more inclusive dataset, future studies can provide a clearer and more accurate picture of healthcare distribution and its challenges.

VI. References

Green, J. L., Manski, S. E., Hansen, T. A., & Broatch, J. E. (2022, November 18). Descriptive statistics. *International Encyclopedia of Education (Fourth Edition)*, pp. 723-733. https://www.sciencedirect.com/science/article/abs/pii/B9780128186305100831

Importance of health care. (2019, October 21). Wellbeing Magazine. Retrieved December 2, 2024, from https://wellbeingmagazine.com/importance-healthcare/

Improving Access to Healthcare in Underserved Communities. (2024, September 9). Medical Missions. Retrieved December 6, 2024, from https://www.medicalmissions.com/resources/78682/access-to-healthcare

Project CCHAIN. (2024, September 30). Kaggle. Retrieved December 2, 2024, from https://www.kaggle.com/datasets/thinkdatasci/project-cchain?resource=download&select=location.csv

Sheingold, B. H., & Hahn, J. A. (2014). The history of healthcare quality: The first 100 years 1860–1960. *International Journal of Africa Nursing Sciences*, *1*, pp. 18-22.

What Is Healthcare? (n.d.). Patient Better. Retrieved December 2, 2024, from https://patientbetter.com/glossary/healthcare/#:~:text=Healthcare%3A%20Refers%20to%20the%20entire,pharmaceuticals%2C%20and%20public%20health%20initiatives.

VII. Annex

1. Scripts and Codes

Python Code for Merging and Visualization:

```
import pandas as pd
import matplotlib.pyplot as plt

# Load the datasets
location_data = pd.read_csv('location.csv')
```

```
osm_poi_health_data = pd.read_csv('osm_poi_health.csv')

# Ensure required columns are present
if 'adm4_pcode' in location_data.columns and 'adm4_pcode' in
osm_poi_health_data.columns and \
   'adm1_en' in location_data.columns and \
   'clinic_count' in osm_poi_health_data.columns and \
   'hospital_count' in osm_poi_health_data.columns and \
   'pharmacy_count' in osm_poi_health_data.columns:

    # Merge datasets using 'adm4_pcode'
    merged_data = pd.merge(osm_poi_health_data, location_data,
on='adm4_pcode', how='inner')

    # Group by region (adm1_en) and sum the counts for each
healthcare category
    healthcare_counts =
merged_data.groupby('adm1_en')[['clinic_count', 'hospital_count',
'pharmacy_count']].sum().reset_index()

    # Handle regions without data (fill missing regions with zero
counts)
    all_regions = location_data['adm1_en'].unique()
    healthcare_counts = (
        healthcare_counts.set_index('adm1_en')
        .reindex(all_regions, fill_value=0)
        .reset_index()
    )

    # Function to plot individual healthcare count categories
    def plot_category(category, title, ylabel, color):
        plt.figure(figsize=(14, 8))
        bars = plt.bar(
            healthcare_counts['adm1_en'],
            healthcare_counts[category],
            color=color,
            edgecolor='black'
        )
```

```
        plt.title(title, fontsize=16)
        plt.xlabel('Region', fontsize=14)
        plt.ylabel(ylabel, fontsize=14)
        plt.xticks(rotation=45, ha='right', fontsize=12)
        plt.grid(axis='y', linestyle='--', alpha=0.7)

        # Add labels to show exact values on each bar
        for bar, count in zip(bars, healthcare_counts[category]):
            plt.text(
                bar.get_x() + bar.get_width() / 2,  # Position at the
center of the bar
                bar.get_height() + 0.5,  # Slightly above the bar
                str(int(count)),  # Convert count to integer for
display
                ha='center',
                fontsize=10
            )

        plt.tight_layout()
        plt.show()

    # Plot each healthcare service in separate charts
    plot_category('hospital_count', 'Number of Hospitals by Region',
'Number of Hospitals', 'skyblue')
    plot_category('clinic_count', 'Number of Clinics by Region',
'Number of Clinics', 'lightgreen')
    plot_category('pharmacy_count', 'Number of Pharmacies by Region',
'Number of Pharmacies', 'lightcoral')

else:
    print("Required columns ('adm4_pcode', 'adm1_en',
'hospital_count', 'clinic_count', 'pharmacy_count') are not found in
the datasets.")
```

```
import pandas as pd
```

```
import matplotlib.pyplot as plt

# Load the datasets
location_data = pd.read_csv('location.csv')
osm_poi_health_data = pd.read_csv('osm_poi_health.csv')

# Ensure required columns are present
if 'adm4_pcode' in location_data.columns and 'adm4_pcode' in
osm_poi_health_data.columns and \
   'adm1_en' in location_data.columns and \
   'doctors_count' in osm_poi_health_data.columns and \
   'optician_count' in osm_poi_health_data.columns and \
   'dentist_count' in osm_poi_health_data.columns:

    # Merge datasets using 'adm4_pcode'
    merged_data = pd.merge(osm_poi_health_data, location_data,
on='adm4_pcode', how='inner')

    # Group by region (adm1_en) and sum the counts for each
healthcare category
    healthcare_counts =
merged_data.groupby('adm1_en')[['doctors_count', 'optician_count',
'dentist_count']].sum().reset_index()

    # Handle regions without data (fill missing regions with zero
counts)
    all_regions = location_data['adm1_en'].unique()
    healthcare_counts = (
        healthcare_counts.set_index('adm1_en')
        .reindex(all_regions, fill_value=0)
        .reset_index()
    )

    # Function to plot individual healthcare count categories
    def plot_category(category, title, ylabel, color):
        plt.figure(figsize=(14, 8))
        bars = plt.bar(
            healthcare_counts['adm1_en'],
```

```
            healthcare_counts[category],
            color=color,
            edgecolor='black'
        )
        plt.title(title, fontsize=16)
        plt.xlabel('Region', fontsize=14)
        plt.ylabel(ylabel, fontsize=14)
        plt.xticks(rotation=45, ha='right', fontsize=12)
        plt.grid(axis='y', linestyle='--', alpha=0.7)

        # Add labels to show exact values on each bar
        for bar, count in zip(bars, healthcare_counts[category]):
            plt.text(
                bar.get_x() + bar.get_width() / 2,  # Position at the
center of the bar
                bar.get_height() + 0.5,  # Slightly above the bar
                str(int(count)),  # Convert count to integer for
display
                ha='center',
                fontsize=10
            )

        plt.tight_layout()
        plt.show()

    # Plot each healthcare service in separate charts
    plot_category('doctors_count', 'Number of Doctors by Region',
'Number of Doctors', 'lightcoral')
    plot_category('optician_count', 'Number of Opticians by Region',
'Number of Opticians', 'lightblue')
    plot_category('dentist_count', 'Number of Dentists by Region',
'Number of Dentists', 'lightyellow')

else:
    print("Required columns ('adm4_pcode', 'adm1_en',
'doctors_count', 'optician_count', 'dentist_count') are not found in
the datasets.")
```