**DEPARTMENT OF EPIDEMIOLOGY STATISTICS & INFORMATICS**

1st KEMRI Data Re-use Hackathon Challenge 2022

**Theme: *Deriving new insights from archived biomedical research data to inform universal health coverage (UHC) national agenda in Kenya***

|  |
| --- |
| **HACKATHON 2022 REPORT FEEDBACK TEMPLATE** |

|  |  |
| --- | --- |
| To: | Head, DESI, Kenya Medical Research Institute |
| From: | Alex Kakai |
| Name of Your Group: | N/A |
| Date of Submission: | 10/12/2022 |
| Subject: | Optimizing health facilities’ location, installed capacity, functionality, expansion space and workflow insights for better service delivery and management. |

**Executive Summary**

* **Introduction**

Every election cycle, Kenyan electorate receive hefty promises of building new health facilities and bettering health care systems generally from aspiring politicians seeking elective office. At the same time, counties directors of health services are getting it hard to maintain existing health facilities not mentioning the increased pressure to open new facilities.

* **Objective**

The main objective of this project was to optimize the hospitals' location, installed capacity, functionality, expansion space and workflow insights to bring out new message to speak to politicians and counties' directors of health.

* **Key Findings**

Majority of the facilities (82%) are in good physical status i.e., the building structure, floors, furniture, and plants, also almost the facilities (93%) have sufficient space for expansion. While these facilities have good physical condition, majority lack capacity to offer prescribed health services with 74% having minimum capacity and 20% have no capacity at all. The dispensaries lead in terms of facilities with least installed capacity (all dispensaries have either minimum installed capacity or no capacity at all) and have the least average visitation/workload across the three years. Lastly, only 7.58% of the facilities were rated as good functionality wise and are offering basic minimum health services. This shows that majority of the facilities as much as they have good physical structures, they luck the required installed capacity to offer basic minimum health services. Physical land elevation was found to have a strong positive correlation (0.7 for 2011, 0.7 for 2012 & 0.6 for 2013) with workload/visitations for patients visiting sub district health facility with majority of them preferring to visit facilities which are located less than 1200 meters above sea level.

* **Recommendations**

From the findings, we observe that building of new facilities is not needed rather expansion to be done to the existing facilities. However, the major issue is installed capacity for the facilities to offer basic minimum health services, majority of the facilities lack this, and more effort should go towards installing more capacity in the existing facilities than building new ones.

**Context**

The objective of the analysis was to optimize the hospitals' location, installed capacity, functionality, expansion space and workflow insights. The following research questions helped with the analysis:

1. What is the distribution of functionality, expansion space and installed capacity across the health facilities?
2. Does facility location have an impact on workload/visitation?
3. What was the distribution of the workflow across the years and health facilities?

**Data Analytics & Visualization:**

Python and Excel were the main tool used for data preparation, analysis, and visualization. The project used basic statistic techniques to make inferences with descriptive statistics being the main technique used. In the descriptive statistics, count was the most used technique to get the distribution in the different categories. Also mean and median which are part of central measure of tendency were used. On visualization, pie charts and grouped bar graphs were used to display the different distributions. Box plots were used to display relationship and distribution of workload which is a continuous variable and level of health facility and land elevations which are categorical variables. Lastly, correlation was used to check if there exists a relationship between workload and land elevation. The point biserial correlation coefficient, a special case of Pearson’s correlation coefficient was used to check for the relationship between workload which is a continuous variable and elevation which is a binary categorical variable.

**Recommendations**

* A freeze notice should be put in place for building new health facilities, unless the place really requires a new facility, efforts should go towards upgrading and expanding the current available facilities.
* A policy planners should set a bare minimum installed capacity to offer basic minimum health services for each health facility level, this will help County Health directors to be able to monitor and manage performance of the various health facilities in their jurisdictions.
* Location of larger health facilities matters to patients with majority preferring to visit facilities whose location’s land elevation is not above 1200 meters above sea level.

|  |
| --- |
| Technical Appendix |
| Python was used for data analysis. Below is part of the code used in the analysis    *#importing necessary packages for the analysis*  import numpy as np  import pandas as pd  import matplotlib.pyplot as plt  import seaborn as sns  import scipy.stats as stats  *#importing dataset to jupyter notebook*  kemri = pd.read\_csv(r'D:\Downloads\Copy of Muranga health facilities capacity physical(987).csv', encoding='latin-1')  *#Previewing the first five rows of the data*  kemri.head()  *#checking size and shape of dataset*  kemri.shape  *#checking the data types of the various variables*  kemri.info()  *#checking descriptive statistics of numeric variables*  kemri.describe()  *#checking distribution of the different health institutions*  kemri.LEVEL.value\_counts()  *#ploting the health institutions on a pie chart*  level = kemri.LEVEL.value\_counts()  plt.figure(figsize = [5,5])  plt.pie(level,labels =level.index , autopct='%1.2f%%', startangle = 90,  wedgeprops= {"edgecolor":"white",  'linewidth': 1,  'antialiased': True})  plt.title('Medical Facility Distribution');  *#distribution of whether the facilities are offering the basic minimum health services by Level*  kemri.groupby(by = ['LEVEL','FUNCTIONALITY2'])['FACILITY\_CODED\_IDENTIFY'].count()  *#plot of distribution of whether the facilities are offering the basic minimum health services*  ff = kemri.groupby(by = ['LEVEL','FUNCTIONALITY2'])['FACILITY\_CODED\_IDENTIFY'].count().reset\_index()  sns.barplot(data = ff, x = 'LEVEL', y = 'FACILITY\_CODED\_IDENTIFY', hue = 'FUNCTIONALITY2')  plt.xlabel('Level')  plt.ylabel('Count')  plt.title('Medical Facility Level and Funcionality')  plt.legend(loc='upper right')  plt.show();  *#Boxplot showing distribution of workload across the different levels*  plt.figure(figsize = [20,5])  plt.subplot(1,3,1)  sns.boxplot(data = kemri, x ='YEAR\_2011' , y = 'LEVEL', hue ='Land\_elevation' )  plt.title('Year 2011')  plt.yticks(rotation = 90)  plt.xlabel('Workload')  plt.subplot(1,3,2)  sns.boxplot(data = kemri, x ='YEAR\_2012' , y = 'LEVEL', hue ='Land\_elevation' )  plt.title('Year 2012')  plt.yticks(rotation = 90)  plt.xlabel('Workload')  plt.subplot(1,3,3)  sns.boxplot(data = kemri, x ='YEAR\_2013' , y = 'LEVEL', hue ='Land\_elevation' )  plt.title('Year 2013')  plt.yticks(rotation = 90)  plt.xlabel('Workload');  *## Biserial correlation*  stats.pointbiserialr(kemri.land\_elevation,kemri.YEAR\_2013)  stats.pointbiserialr(kemri.land\_elevation,kemri.YEAR\_2012)  stats.pointbiserialr(kemri.land\_elevation,kemri.YEAR\_2011) |