Optimizing a Healthcare Network for Improved Service Delivery

Zohair Hashmi
School of Electrical Engineering & Computer Science
National University of Sciencc & Technology
H-12 Campus, Islamabad, Pakistan
hashmi.zohair@gmail.com

***Abstract* – Overstaffing and understaffing in health care facilities is a major concern that affects both, the healthcare workers and the patients, in terms of time and cost. The state of Washington has five such facilities and the healthcare ministry would like to solve the problem by making necessary relocations or upgradations of staff members across the five facilities. The problem is addressed with a data-driven approach where data of census, location and travel time is extracted from different sources to present valuable suggestions to the Washington State Healthcare Ministry. The final solution is a suggestion to the ministry of readjusting the staff counts in each facility, i.e. increase or decrease the count based on the requirements of the area.**

I. Introduction

In the state of Washington, there are 5 healthcare facilities (A-E) in the cities of Bellevue, Seattle, Snohomish, Snoqualmie and Wenatchee. Each of these facilities tends to provide healthcare to the population residing nearby. These facilities are either overstaffed or understaffed to serve the population. This tends to raise a concern for people living in areas nearby as they often have to travel to a far-off facility to get their healthcare done, as the facility nearby is understaffed to provide services to each patient.

Therefore, to facilitate the patients some of the staff needs to be relocated or upgraded. There are two things that need to be considered before relocation.

1. Staff to patient's ratio should be approximately 1/2808.
2. Staff count changes need to be done in a way that people travel to the facility closest to them in terms of travelling duration.

We are only provided with the details of the facilities in a tabulated form as shown below, that include the zip code of the area, Facility ID and the staff count allocated to the facility.

TABLE I

1.3.1 Facilities

Facility ID	Facility Area-Zipcode	Facility Staff Count
Facility A	98007	21
Facility B	98290	52
Facility C	98065	43
Facility D	98801	9
Facility E	98104	64

II. Assumptions

In order to simplify this problem, it is assumed that more than half of the population living nearby a facility would opt for healthcare in a year. This assumption is based on the fact that not everyone gets sick in a year. Therefore, an assumption of 60% population is taken into account.

Moreover, it is assumed that by 'areas nearby' it is meant that only the population living in the five cities with healthcare facilities is considered. This is because the problem is concerned to the Washington Health Ministry only, which eliminates considering the population of neighboring states. Also, other cities of Washington that are situated away from the five cities provided will have their own facilities and it will be inconvenient for them to consider travelling to any of these five facilities as they are too far away for them to drive to.

When considering the duration of travelling to a facility from a specific location, travelling is considered via car only. This assumption is based on the fact that majority of people living in the cities will opt for their own transport, which is a car in the majority of case, to travel to a healthcare facility.

Another important assumption made regarding travelling to the facility is that people opt for the fastest route possible. The fastest route data is used from the data extracted from a routing API. People often use online maps to navigate to their desired locations and the maps always recommend a fastest route to the users.

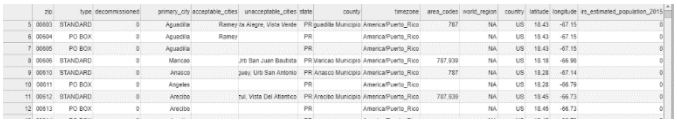
III. Data Analysis

A. *Data Collection*

1) Zip Codes and Location Information: The zip codes of the facilities are already provided to us, but in order to cover the entire neighborhood we need to ensure that we have a complete database of the zip codes in the United States of America. The website www.unitedstateszipcodes.org is the most convenient source of information. The website provides a complete database of all the zip codes in the country and their respective locations. As we must deal with population as well, the database also provides an estimated population of each locality as recorded for the year 2015.

The collected data is as shown in the figure below.

Figure 1



	zip	type	decommissioned	primary_city	acceptable_cities	unacceptable_cities	state	county	timezone	area_codes	world_region	country	latitude	longitude	irs_estimated_population_2015
5	00603	STANDARD	0	Aguadilla	Ramey	la Alegria, Vista Verde	PR	guadilla Municipio	America/Puerto_Rico	787	NA	US	18.43	-67.15	0
6	00604	PO BOX	0	Aguadilla	Ramey		PR		America/Puerto_Rico		NA	US	18.43	-67.15	0
7	00605	PO BOX	0	Aguadilla			PR		America/Puerto_Rico		NA	US	18.43	-67.15	0
8	00606	STANDARD	0	Maricao		Urb San Juan Bautista	PR	Maricao Municipio	America/Puerto_Rico	787,939	NA	US	18.18	-66.98	0
9	00610	STANDARD	0	Anasco		Joaq. Urb San Antonio	PR	Anasco Municipio	America/Puerto_Rico	787	NA	US	18.28	-67.14	0
10	00611	PO BOX	0	Angeles			PR		America/Puerto_Rico		NA	US	18.28	-66.79	0
11	00612	STANDARD	0	Arecibo		hal. Vista Del Atlantico	PR	Arecibo Municipio	America/Puerto_Rico	787,939	NA	US	18.45	-66.73	0
12	00613	PO BOX	0	Arecibo			PR		America/Puerto_Rico		NA	US	18.45	-66.73	0

2) Travel time Data: HERE API is one of the data sources available for gathering data related to traffic, routes and maps. HERE's Routing API provides data about the travel time it takes to move from an origin to a desired destination.
Prepare your paper in full-size format on US letter size paper (8.5 by 11 inches).

B. Data Handling

Data retrieved from the above sources is stored in '.csv' format. This is to ensure that requests are not sent over and over again when testing the source code. Since, the source code was written on IBM Developer Skills online jupyter notebook, the data is stored on cloud in the cognitive class AI cloud storage available to my account.

In the source code the data is converted into a dataframe using pandas. This allows us to use the data in the cleaning and processing stage.

C. Data Cleaning

The data obtained from the sources is cleaned before entering the processing stage so that only the useful information is kept while the rest is removed. Therefore, we only keep the columns: zip, primary_city, state, latitude, longitude and irs_estimated_population_2015.

Also, we remove all the rows where state is not equal to 'WA', as we are not conerned with other states. Using the zip codes given in Table I, we obtain the rows corresponding to those zip codes and extract the city names. Now, these names are sued to obtain all the zip codes in the database of the cities extracted.

Now, we are left with a cleaned and trimmed data that contains only the information important to us.

D. Data Processing

1) Clustering:

Resources should be adjusted in such a way that people travel to the closest facility. Therefore, we need to cluster each zip coded location (nearby areas) according to the facility closest to them.

Using the cleaned dataframe constructed from the database, a list of coordinates (latitudes and longitudes) of each distinct zip code location is made and used to retrieve travel time to each facility through HERE api. This returns the duration it takes to travel to each facility from each location point.

The fact that people will always prefer to go to the clsoest facility makes it important for us to claasify the locations according to the minimum duration it takes to travel. The facility with the minimum distance is assigned as a label to the location point being assessed.

E. Data Visulaization

Data is visualized using folium maps and matplotlib. Folium maps are sued to observe the cluster formed after each location point is assigned its closest facility.

Matplotlib on the other hand is used to plot a scatter plot of staff count to population, demonstrating the variation amongst facilities of the allocated resources. If the scatter points lied on a straight line, this would have demonstrated a fair allocation of resources.

IV. SOLUTION

A. Optimization

The clustering of our data set allows us to have a straightforward approach towards optimization. The clustering is carried out on basis of the minimum time it takes to approach any facility. Therefore, now it is required that the staff member count is readjusted in such a way that overstaffing or understaffing is minimized.

Considering the rule of thumbs of a staff to patient ratio of 1/2808, we started it off by dividing our estimated patient count for each facility by 2808. This returned us the expected staff count for each facility to maintain the rule of thumb. Dividing by 2808 returns us the figures displayed in the following table.

TABLE II

Facility ID	Actual staff count	Suggested staff count
A	21	31
B	52	13
C	43	3
D	9	9
E	64	177

Since, the total of the suggested staff count (233) exceeds the available staff number (189), we tweak the ratio slightly and raise it from 2808 by an amount of 10, until it reaches a point where the total staff count remains 189. This needs to be done only if the ministry is not looking to add new staff members but is willing to relocate current staff only.

When this is achieved an equal distribution of workload is ensured and staff relocations are facilitated without the need of hiring more staff members. The following staff count suggestions are concluded.

TABLE III

Facility ID	Actual staff count	Suggested staff count
A	21	25
B	52	10
C	43	3
D	9	7
E	64	144

SOURCE CODE(PYTHON)

The source code is attached in the following github link: https://gist.github.com/zohairhashmi/4a26a8dd9fa9faaf339fd5c5a5892db1

REFERENCES

[1] https://pandas.pydata.org/pandas-docs/stable/
[2] https://numpy.org/doc/
[3] https://matplotlib.org/3.1.1/api/_as_gen/matplotlib.pyplot.html
[4] https://python-visualization.github.io/folium/
[5] https://requests.readthedocs.io/en/master/
[6] https://www.unitedstateszipcodes.org/zip-code-database/
[7] https://developer.here.com/documentation