

Solving a Healthcare Problem for Improved Service Delivery

1 INTRODUCTION

The Washington State Health Ministry would like to optimally upgrade staff or resources in existing facilities across Washington so that they can allocate resources to where they are most needed, based on current and future demand for different types of services in different geographic regions/areas.

To address this problem, the health ministry needs to:

- Understand the demographic make-up of each area and nearby areas of the facilities to determine demand for types of health services.
- Predict future demand for types of services in each area and nearby areas of the facilities. This involves looking at population increase/decrease forecasts, changing demographics, future infrastructure/job investments or population migration.

1.1 SAMPLE PROBLEM

Majority of the population living in Area A need access to a certain type of service (e.g. maternity), but would need to travel on average 2 hours per return trip to get their needs met in Facility B, instead of travelling 0.5 hours to their nearby Facility A. This is because Facility A has a limited capacity for that service as opposed to a much greater capacity in Facility B. People living in Area A cannot travel this far, so they need to get their needs met by Facility A, otherwise they go without healthcare.

1.2 DESIRED SOLUTION

To group zipcode areas and nearby areas into health services according to their demographics so that the ministry can see if the current capacity of a facility meets the needs of its nearby population.

The health services groups are:

- Maternity (basic care)
- Elderly (serious care)

The solution also involves predicting future demand for a type of service in populations nearby each facility.

The below table is a sample of what the current and future population analysis might look like:

Area ID	Nearby Facility ID	Area Population	Maternity Population within Area	Predicted future additional increase/decrease maternity population	Nearby facility type	Current capacity for maternity in facility
2	6	10,000	100 (1%)	0	Primary	36000
3	7	3,000	300 (10%)	0	Secondary	0
4	8	2,000	600 (30%)	0	Secondary	0
5	9	5,000	2,500 (50%)	100 (so 2,600 in total)	Tertiary	0

1.3 DATA SOURCES

1.3.1 Facilities

Facility ID	Facility Area-Zipcode	Facility Capacity Primary	Facility Capacity Secondary	Facility Capacity Tertiary
Facility A	98007	200	44200	0
Facility B	98290	36000	0	0
Facility C	98065	12000	24336	0
Facility D	98801	0	10150	22100
Facility E	98104	58000	32700	8050

1.3.2 Population

Example sources: US Census Bureau, World Population Review, etc.

These sources can be used to grab population estimates in facility zipcode areas and nearby areas, and demographics in each area. Demographics include age groups, child dependencies, household type, employment status, income, etc.

Note: Not the entire population living nearby a facility would need healthcare services, but we assume at least more than half would seek services for a variety of problems.

1.3.3 Understanding services offered in each facility type

Primary: Handles basic care, maternity and child care, prevention treatment.

Secondary: Handles specialized care such as respiratory, acute and other serious illnesses.

Tertiary: Handles advanced medical investigation and treatment such as cancer, neuro disease, cardiac and other serious illness.

2 DELIVERABLE

Write a short technical document addressing the following:

- **Introduction:** What is your understanding of the problem? Can you write the introduction and the problem statement in formal conference paper-like format?
- **Assumptions:** What assumptions would you make to simplify the problem?
- **Current population counts in each area:** How would you acquire the data for maternity and elderly populations in each area? Would you take any steps to process and refine the data?
- **Future population counts in each area:** Share your prediction model details or how you would predict future populations for maternity populations in each area.
- **References:** Cite any libraries, APIs, publications. We believe in standing on the shoulders of giants. Please reuse any existing research papers, source code, libraries but make sure to cite them.
- **Source code:** Share your R or Python source code.

2.1 CURRENT POPULATION COUNTS

- Describe how you would use the population data sources to extract current population counts of maternity people and elderly people in each area.
- Give a short outline or coded example demonstrating the concept of how you would extract this information from this data source.
- Are there any other freely, publicly available data sources you would consider using and how would you extract information from these sources?

2.2 FUTURE POPULATION COUNTS

- Describe how you would use the population data sources to predict future additional population/increase or decrease in population counts of maternity people and elderly people in each area.
- Give a short outline or coded example demonstrating the concept of how you would predict this from the population data sources.
- Demographics alone give a limited picture of possible future populations. For example, if job investments are planned for some locations this could result in more new families moving to these locations. Are there any other freely, publicly available data sources you would consider using and how would you extract information from these sources?

3 EVALUATION

Your submission will be evaluated based on the following aspects.

Note: Don't worry if you think any of the aspects is missing. We are not looking for a top-notch PhD dissertation. Rather, we are looking for a deliverables that demonstrates genuine interest in data science; motivation and curiosity; good writing and communication skills; coding skills; dealing with ambiguity; and ability to work under tight deadlines.

- **Understanding of the problem:** Data Science Dojo consultants often work on cutting-edge data science, machine learning, optimization, and mathematical modelling problems. Your ability to understand problem is critical to your success at Data Science Dojo.
- **Rigor:** How detailed your research was and how hard you tried matters more than the actual results.
- **Attention to detail:** Make sure that your document is not missing any details. Typos are never fun.
- **Demonstration of skills:** We are not looking for a unicorn. We don't expect you to be outstanding in all the areas of the problem. We do, however, want you to be good at a few of the skills. Your coding skills; data extraction and manipulation; investigate and research data sources and publications; ability to model problems; and ability to communicate complex concepts.