2022

PUBLIC DATA ANALYTICS:
Data-Driven Community Problem Solving

A COMPILATION of USE CASES FROM THE OCDEX NETWORK IN 2022

VOLUME 2 OF THE OCDEX HANDBOOK SERIES

LAYER _TECH

OCDex

ATTRIBUTION

Neither the publisher nor the authors nor the editors give a warranty, express or implied, with respect to the material contained herein or for any errors or omissions that may have been made.

For questions, comments, and suggestions, please reach out to: learning@layertechlab.com.

ACKNOWLEDGMENT

This handbook is a compilation of public data science and analytics use cases generated by the fellows and partners of Project OCDex in 2022. Team OCDex would like to acknowledge the following authors, reviewers, partner institutions, and contributors to this handbook:

Authors and Contributors

Frei Sangil
Layertech Labs

Rex Bringula
University of the East

Roselle Basa
University of the East

John Paul Miranda
Don Honorio Ventura State University

Sittie Pasandalan
Mindanao State University - Iligan Institute of Technology

Rabby Laviles
Mindanao State University - Iligan Institute of Technology

Mia Amor Tinam-Isan
Mindanao State University - Iligan Institute of Technology

Marben Ramos
Bicol University

Jennifer Montemayor
Mindanao State University - Iligan Institute of Technology

Shehab Ibrahim
Mindanao State University - Iligan Institute of Technology

Nico Aspra
Bicol University

John Raymond Barajas
Bicol University

Pee Jay Gealone
Bicol University

Arpon Lucero Jr.
Bicol University

Oliver Padua
Bicol University

Vianca Jasmin Anglo
Women in Technology Tarlac

OCDex 2022 Fellows

Ramon Rodriguez
National University

Mia Amor C. Tinam-isan
Mindanao State University - Iligan Institute of Technology

Roben Juanatas
National University

Gabriel Avelino Sampedro
National University

Manolito Octaviano
National University

Jimson Ornido
National University

Mark Emmanuel Malimban
National University

Rabby Q. Lavilles
Mindanao State University - Iligan Institute of Technology

Sittie NB Pasandalan
Mindanao State University - Iligan Institute of Technology

Shehab D. Ibrahim
Mindanao State University - Iligan Institute of Technology

Jennifer Joyce M. Montemayor
Mindanao State University - Iligan Institute of Technology

Lany Maceda
Bicol University

Jennifer Llovido
Bicol University

Mary Joy Canon
Bicol University

Christian Sy
Bicol University

Pee Jay N. Gealone
Bicol University

Ramon Gian A. Bron
Bicol University

Nico O. Aspra
Bicol University

The OCDex team would like to acknowledge the support of the International Republican Institute (IRI) to the 10-month OCDex 2022 project, the creation and the publication of this handbook. The team would also like to thank the participation of experts from the Department of Information Communications Technology Region 5, representatives from the city government of Legazpi, and the provincial government of Albay in the OCDex 2022 project activities. The team would like to acknowledge the partnership with the Department of Trade and Industry Albay during the OCDex 2022 Public Data Analytics conference which served as the venue for the authors of this handbook to present and discuss their use case reports to stakeholders.

The insights and opinions in this book are solely attributed to the respective contributors and authors of the use cases. They do not, in any way, reflect the opinions and position of Layertech and/or our partners as a whole.

Compiled and designed by:

Frei Sangil and Liselle Custodio

PREFACE

Data science and analytics has demonstrated its power in informing decision-making and problem-solving. Data can reveal trends and insights that would have otherwise been obscured. It can give decision-makers key information needed to craft effective and optimal solutions to organizational problems. It can help predict potential bottlenecks and challenges, so that organizations may come prepared when it happens. Data science and analytics is a sought-out skill in the digital age.

The Covid-19 pandemic and its resulting limitations on mobility has forced many transactions and communications to migrate from the physical space to the digital space. This sudden global digitalization resulted in an increase in data produced, and a subsequent increase in the potential game-changing insights that these data may be hiding.

While many in the private sector have been seen leveraging the power of data for business insights and maximization of revenue, the public sector is yet to catch up in terms of digitalization and data utilization, especially in developing countries. The power of data would especially help communities and local governments in coming up with efficient, effective, and inclusive policies and solutions to problems.

The aim of the 2022 OCDex project run is to bring data scientists and analysts together, and demonstrate how analysis of government data can be used to help solve problems in local communities. The project aims to demonstrate how it can help inform local policymaking and project planning, and how citizens and researchers can participate and help their respective local government units in overcoming community challenges hand-in-hand. This handbook hopes to convince local governments and authorities to invest in good data housekeeping and integrate data science and analytics into their decision-making.

This handbook features how academics and data enthusiasts used public data to help inform solutions to various community problems such as healthcare, inclusivity and accessibility for persons with disabilities, fairness and transparency in public procurement and ensuring enough supply of utilities. Lastly, this handbook presents a replicable model of cooperation between local governments and their local researchers and data enthusiasts towards the effective use of data science and analytics for community building.

TABLE OF CONTENTS

PART 3

DEFINITION OF TERMS

AI
Artificial Intelligence or AI refers to the various efforts to get machines to perform tasks in the same way that humans can.

Machine Learning
Machine Learning is a branch of AI that aims to mimic the way humans learn by 'feeding' machines with data. A machine uses the data to 'learn' and gradually increase its accuracy.

Similar to how a child is given books so that they can learn to do tasks on their own, machine learning gives data to a computer to 'learn from.'

Dashboard *(Digital)*
A dashboard is a type of digital interface that shows a graphical report of data and is organized in a way that is easy to understand and relevant to a certain group or topic.

E.g. *This dashboard shows a report of all the PWDs and their respective disabilities in Metro Manila.*

Machine Readable Data
Machine-readable data refers to data formatted in a way that can be processed by a computer.

For example, if I scan a hard-copy contract document and save it as an image file, it is not machine-readable because the text and relevant information in the contract cannot be easily processed by a computer as an image file. On the other hand, if the contract is saved as a document file or a CSV file, the computer can process the contents easier compared to an image. We can say that the CSV file is machine-readable.[1]

[1] Machine readability and prescribed data formats are still changing definition, depending on the current capabilities of computers, as well as industry standards published.

Data *(Digital)*
Data refers to any information or 'footprints' left on a digital device or platform.

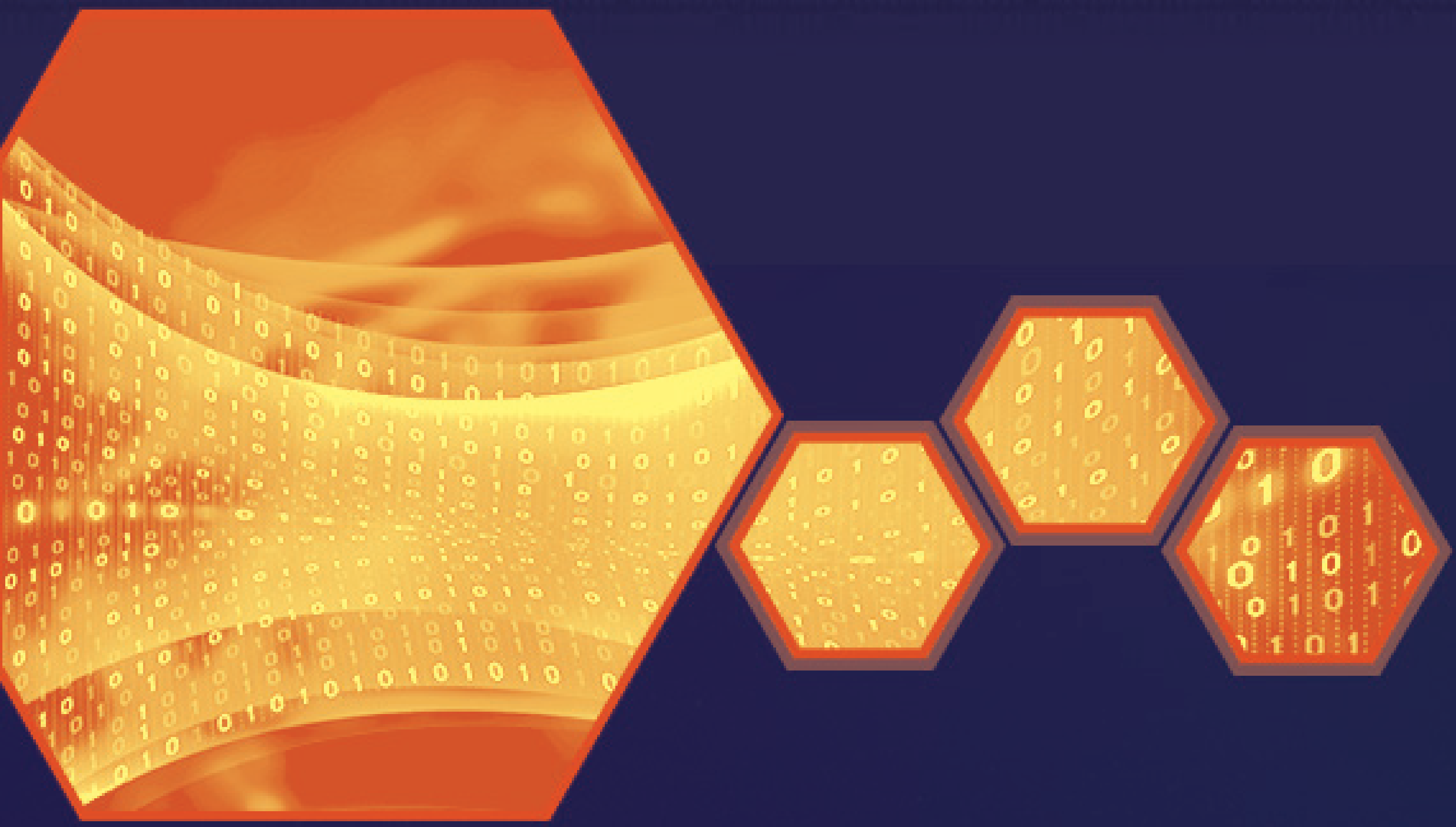
Public Data
Public Data refers to datasets, information that can be used, modified, and redistributed by anyone, without legal restrictions. However, public data may not necessarily be machine readable or easily accessible.

Data Analytics
Data Analytics refers to the process of analyzing data to draw out useful insights relating to a certain topic or question.

Open Data
When a dataset is considered 'open data', it means that it can be used, modified, and re-distributed freely without any constraints. At the same time, the dataset is machine readable and easily accessible.

Data Science
Data Science also aims to draw out useful insights from raw data. However, Data Science focuses more on the data and how it can be transformed, modeled, and programmed to perform a certain task.

Visualization *(Data)*
Data VIsualization refers to the practice of creating graphical representations of data. Visualization is a very important tool in making decision-makers understand insights from data.



Part 1: INTRODUCTION

A Review of Concepts and Tools

1.1 About the handbook

1.1.1 Objective of the handbook

This handbook provides sample use cases and respective Data Science and Analytics methodologies in Philippine public and government datasets. The first part of this handbook is written by Layertech, and it provides a quick recap of basic concepts related to Data Science and Analytics for general readers. The first part also presents a quick overview of the tools used by the researchers who produced the use cases.

The second part of this handbook presents the actual use cases and their results as submitted by the authors. The references and additional notes for each use case are provided at the end of each section.

Some of the information contained in this handbook are highly dependent on specialized knowledge. However, efforts to simplify explanations were undertaken by the contributors and editors.

1.1.2 Learning Objectives

This handbook can be used as a self-paced or group resource. Readers will explore the following:

- Basic concepts and terms related to Data Science and Analytics;
- Tools used by the authors to generate insights from data;
- Source of datasets;
- Basic data mining frameworks; and,
- Example cases wherein public data was used and analyzed to inform decision-making and community problem-solving.

1.1.3 Delimitations

This handbook does not attempt to address all possible procedures or methods of analytics or imply that it is limited to the contents of this handbook. Readers are urged to view this handbook as a beginning resource; to supplement their knowledge of public data analytics and methods as part of their ongoing personal or professional development. In addition, the procedures and methods introduced do not provide assurance that the reader's own application will be successful. Finally, this handbook is not meant to replace domestic policies and procedures.

1.2 The OCDex Project 2022 Run and the Public Data Science and Analytics Conference

Project OCDex started in 2018 as a portal that hosts public procurement-related datasets and resources. In 2020, the team started receiving requests for other datasets and resources such as data on public health, data on electricity and utilities, and other datasets that various civil society organizations and researchers find the most relevant to the current issues in their local communities.

In 2022, with the support of the International Republican Institute's 10-month grant, the OCDex team reached out to three universities (National University, Bicol University, and Mindanao State University-Iligan Institute of Technology) and kicked off a year-long fellowship program for 18 faculty members. The fellows underwent a series of training sessions on data science and analytics, current government programs related to the topic, and related policies to consider such as the data privacy law when using public and government data. The goal of the fellowship program is to capacitate and encourage members of the academe to work with their local government units and use actual government data to come up with studies and innovations that would help their respective communities.

The 18 fellows became part of the OCDex network of analysts.

Figure 1.2 Photo A

OCDex Public Data Analytics Conference Keynote Speaker Department of Information Communications Technology Undersecretary Atty. Jocelle Batapa-Sigue Poses With OCDex Fellows and Researchers

On May 11, 2022, the OCDex team launched a call for proposals for data use-cases in the Philippine context. Four teams were selected and supported by OCDex team and its network, to build the use cases. The use cases were presented during the OCDex Public Data Analytics Conference on August 23, 2022, held in Legazpi city Albay, where the researchers networked and presented their reports in front of multiple stakeholders from government agencies, the private sector, and fellow researchers.

Figure 1.2 Photo B

Researchers from Bicol University, Don Honorio Ventura State University, and Women in Technology-Tarlac Accepts Their Plaques for Presenting Their Use-cases at the OCDex Public Data Science Conference

1.3 A review of Concepts – Data Analysis

What is Data and How is it being used?

Data refers to any information or 'footprints' left on a digital device or platform. The Covid-19 pandemic has become one of the strongest drivers for digitalization as the limitations in mobility forced many physical operations into the digital space. For example, classes were suddenly held online. Companies had to interact with their employees and clients through video conferencing and use digital platforms to carry on with their tasks. This surge in the use of digital tools left even more 'footprints' or 'data' over the past couple of years.

Data in itself doesn't tell us much. However, if data is given context, it turns into 'information.' And information is powerful. The multimedia company, Netflix, for example, has become a staple example for many beginning data science courses. Netflix's use of data science and analytics was documented to have significantly contributed to its revenue. A popular example is how Netflix uses their subscribers' data (movies they watch, favorite genre, etc.) to predict the kind of shows that they will like. This specific application of data is called a 'recommendation engine.' Because Netflix's recommendation engine could accurately predict which shows the viewers will most likely watch, the viewers tend to spend more hours on Netflix, thereby increasing their user retention.

Another example would be popular mobile navigation applications such as grab and uber. The software uses data to calculate the estimated fare and the best possible route for the driver to pick up a booked passenger. This is only possible because of a good dataset and a good data model or algorithm.

Given this, the use of data in the public sector will greatly help decision-makers come up with more effective, resource-efficient, and inclusive policies and programs for the public.

Data Analytics, Data Science, and Data Mining– What is the Difference?

You might hear these terms used interchangeably, but there are differences between these terms.

Data Analytics refers to the process of analyzing data to draw out useful insights relating to a certain topic or question. Data Analytics focuses more on how the insights generated from the data can answer questions or can inform solutions to a certain problem.

On the other hand, Data Science focuses more on the data and how it can be transformed, modeled, and programmed to perform a certain task. Data Scientists explore the data, find correlations, create models from data, run algorithms on data, and perform statistical tests.

Data Mining refers specifically to the process of extracting useful information or insights from data.

The CRISP-DM Data Mining Framework

The cross-industry process for data mining or CRISP-DM is a widely-used data mining standard in the industry.

The CRISP-DM model consists of six phases:

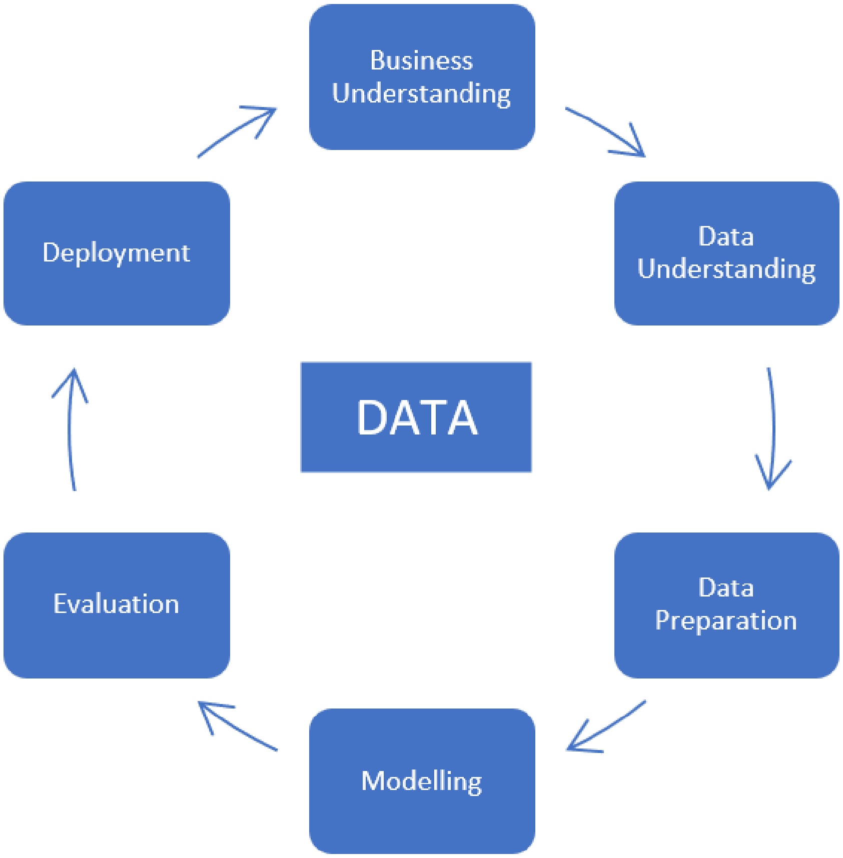


Figure 1.3 Diagram A

Business Understanding - The first step is to have a clear understanding of the context, the business goals, or the questions that need to be answered. A clear understanding of the data mining goals and plan is a crucial first step.

Data Understanding - Know the data. This phase is about collecting initial data, describing the data that the analyst is working with, exploring it, and verifying its quality and fitness to be used in the data mining task.

Remember GI-GO!

GI-GO: Garbage In - Garbage Out.

If the data quality is not good and not appropriate for the task, do not expect to get useful and accurate insights. Assessment of the quality and integrity of your data is a crucial step!

Figure 1.3 Diagram B

Data Preparation - Many times, datasets that will be used for data mining need to be 'cleaned' or a more formal term is, 'pre-processed.' For example, if the dataset contains missing values, or inconsistent formats (e.g. different date formats), it is often necessary to standardize these to ensure that the dataset will have a uniform format fit for processing.

Modeling - In this phase, modeling techniques and algorithms are used on the data (e.g. neural networks, regression, etc.). This phase is usually done in multiple iterations, fine-tuning parameters until it yields satisfactory results.

Evaluation - In this phase, the data model and the results are evaluated if it meets the success criteria of the data mining objectives and tasks. In other words, this phase should tell whether or not the results answer the business questions and satisfy the requirements.

Deployment - Finally, the results and insights from the data mining, if proven appropriate, effective, and feasible, needs to be integrated into the organization's operations or decision-making. Then, it needs to be monitored if the insights are indeed creating value for the organization. The results of the monitoring after deployment can be used as input for another iteration of the CRISP-DM process.

The ‘OPEN DATA’ Movement

When a dataset is considered ‘open data’, it means that it can be used, modified, and re-distributed freely without any constraints. According to the World Bank, having open data is essential as its use can improve government services, open new economic opportunities, improve public safety, reduce poverty, drive innovation and promote participation and transparency.

The two dimensions of open data are:

1. The data must be placed in the public domain with minimal restrictions. In other words, it is easily accessible to the public.

2. The data must be published in a format that is machine-readable. Meaning, the data is formatted in a way that can be processed by a computer and by most commercial data processing software. For example, if I scan a hard-copy contract document and save it as an image file, it is not machine-readable because the text and relevant information in the contract cannot be easily processed by a computer as an image file. On the other hand, if the contract is saved as a document file or a CSV file, the computer can process the contents easier compared to an image. We can say that the CSV file is machine-readable.

The open data movement is gaining popularity in recent years. There are several open data repositories online which can be explored.

1.4 A review of Tools Used

Apache OpenOffice Calc

OpenOffice Calc is a free and open-source spreadsheet software by Apache. Calc is used to organize and manage data in a tabular format and can be programmed to perform calculations and generate visualizations.

LINK: https://www.openoffice.org/product/calc.html

Excel

Microsoft Excel is a proprietary spreadsheet software by Microsoft. Excel is used to organize and manage data in a tabular format and can be programmed to perform calculations. Excel also has visualization capabilities.

LINK: https://www.microsoft.com/en-us/microsoft-365/excel

Google Data Studio

Google Data Studio allows the conversion of data into visual reports and dashboards. The tool can be accessed for free by individuals and small teams with a registered Google account.

LINK: https://datastudio.google.com/

Python and IDLE

Python is a flexible, popular programming language with many applications. Python has many libraries that allow software written in python to be integrated into a wide variety of platforms and protocols. There are many libraries in Python built to process and visualize data.

To create and run python codes, Integrated Development Environments (IDE) are usually used. The default IDE for Python is IDLE or Integrated Development and Learning Environment.

LINK: https://www.python.org/

Google Colab

Google Colab or Google Colaboratory is an online tool that allows users to write and execute Python codes in the browser. Google Colab is accessible using a registered Google account.

LINK: https://colab.research.google.com/

R and R Studio

R is a programming language used for statistical computing and visualization. R has many libraries that allow the processing, modeling, and visualization of data. R studio is a free, open-sourced IDE for R.

LINK: https://www.r-project.org/about.html and https://www.rstudio.com/

1.5 A model for cooperation – LGU/GOVT

One of the critical challenges faced by local governments is the lack of an in-house data analyst or data scientist that would help them generate value from their data. The 2022 run of Project OCDex aims to create linkages between local researchers, data analysts and scientists, and data enthusiasts and their respective local governments and other government agencies to co-create innovations, solutions, and data analysis projects that are valuable to the local context.

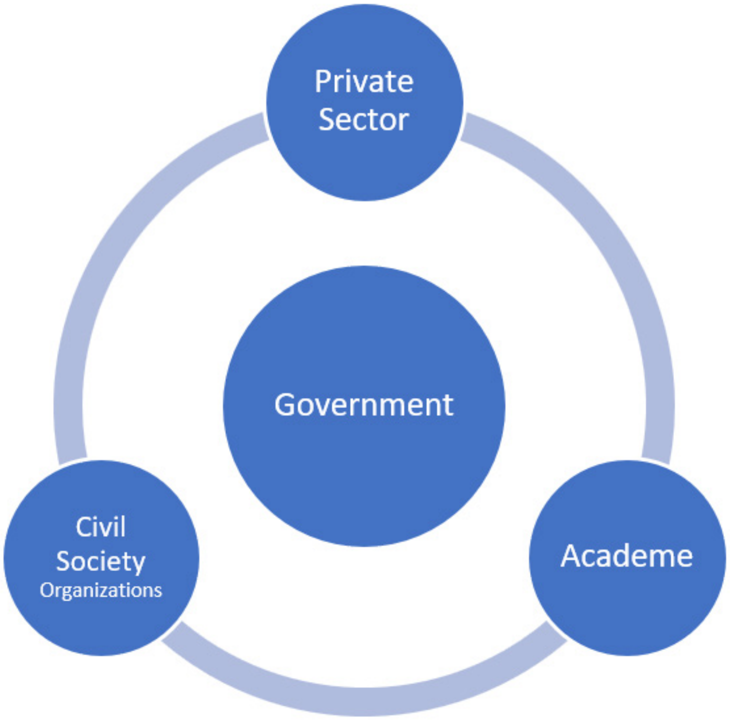
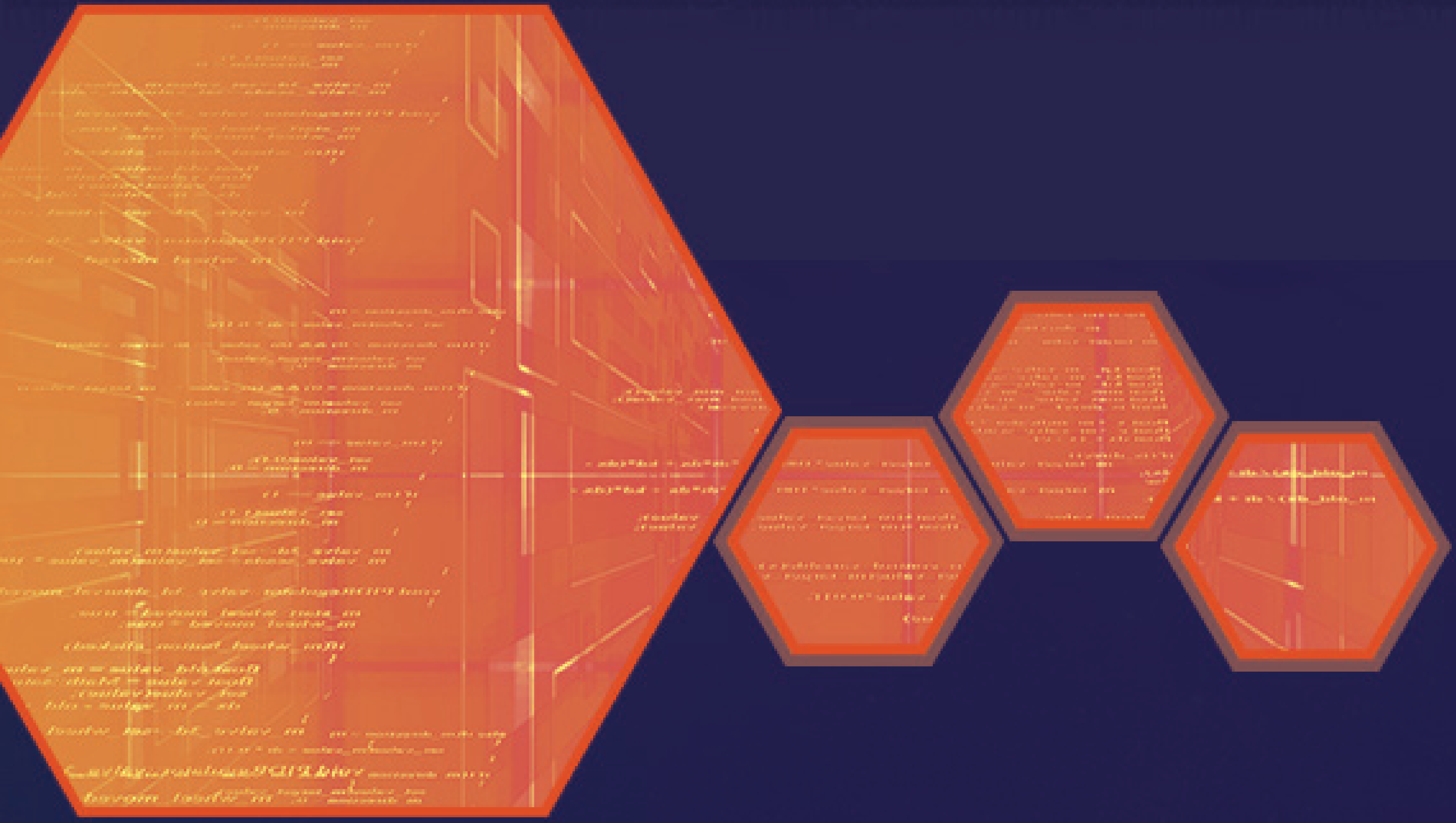


Figure 1.5 Diagram A

In this model, the local government serves as the main data source and implementor of the proposed policies and programs. The academe, given their natural interest in research and development and technical expertise in data mining, is a valuable ally in generating value and innovation from local government data. Civil Society Organizations and the Private sector can also contribute in terms of building context and ensuring that the results are translated into actionable recommendations that are inclusive and effective. There are also private institutions and CSOs with data and capabilities to analyze data that can contribute to the local government and academe partners.



Part 2: INTRODUCTION

Use Cases Of Public Data Science And Analytics In The Philippine Context

2.1 USE CASE 1: Philippine PWD Dashboard

VIANCA JASMIN ANGLO

2.1.1 About the author

VIANCA JASMIN ANGLO

Vianca Jasmin Anglo is the Chief Executive Officer of Women in Technology Tarlac chapter (WiTech Tarlac). Vianca graduated Magna Cum Laude at the University of the Philippines, Diliman and she works as a data analyst at the ASEAN Business Youth Association. She is a speaker and trainer for various organizations such as USAID and UN General Assembly for Youth.

2.1.2 Executive Summary/Abstract

Rationale

A national dashboard is created to unify registered Filipino PWDs to provide aid in the development of policies and initiatives that are appropriate for the needs of PWDs. The dashboard aims to be a tool used by local chief executives and local leaders in cities to inform, direct, plan, and educate its community, most especially its vulnerable groups like the PWD community, in creating inclusive and equitable city-level programs.

Objectives

The objectives of this dashboard are to provide insights into the number of PWDs in cities in the Philippines, to provide insights into equal opportunity & access to public services for PWDs, and to give a general overview of a regional number of PWD from 2011 - 2019.

Methodology

After the compilation of relevant documents and data, all of the acquired data were cleaned in correspondence that the Google Data Studio (GDS) system would be able to read. Academic journal articles, gray literature of National Government Agencies (NGA), and local news were also analyzed to comprehend the extent of inclusiveness, issues faced, and efforts being made by NGAs in consideration of the Philippine PWD community. Subsequently, the dashboard has undergone different rounds of polishing on its metrics, design, and data quality. The PWD Statistics Dashboard was then presented to different intergovernmental organizations for vetting, potential for usability, and recommendations on key metrics to be added to the dashboard as new data is presented.

Recommendations

Many PWDs have underlying health and face socio-economic issues that make them more prone to diseases, discriminatory practices, and prejudices resulting in higher rates of infection-related mortality, dropping out of school, and unemployment. Given the number of certain categorical PWDs per city, there is baseline data on where local leaders can create targeted programs, create inclusive public systems, and health benefit allocation that can be addressed. Recommendations and proposed next steps for local government units and local leaders allay some of these issues faced by PWDs by i) conducting micro-planning sessions with local advocacy groups that champion PWD communities, ii) providing affordable assistive technology and tools serving persons with disabilities in local pharmacies, clinics, and public hospitals, iii) delivering traction and clear guidelines on the registration of a PWD card and providing information on its benefits to the cities constituents, iv) creating more inclusive and accessible public spaces, v) disaggregating the categorical data of PWDs for more localized and targeted programs, and vi) mapping of all Persons with Disabilities Affairs Office (PDAO) offices in cities and municipalities.

2.1.3 Use-Case Body

Introduction

The COVID-19 pandemic has extensively placed great constraints on the Philippine healthcare, education, as well as, economic systems. Filipino People with Disabilities (PWD) are one of the communities that are heavily impacted by the policy placements to regulate the spread of infection and programs imposed as a response to the pandemic (UNHCR, 2020). The Philippine PWD community has remained to be one of the most vulnerable and unheard voices in our country even before the onset of the COVID-19 pandemic. They experience inaccessibility, inequality, prejudices, and limited autonomy to public spaces, transport, healthcare accessibility, and education (Velasco, et al, 2021).

In the Philippine National Deployment and Vaccination Plan (NDVP), the PWD community shall be given access to COVID-19 health insurance coverage through PhilHealth, the government should also provide Filipino sign language interpreters on broadcast media, and accessible information through public media and physical spaces. However, there are challenges by some local government units (LGUs) in the implementation and recognition of the type of support to be given to the PWD community depending on which category of their disabilities. There is a tendency for policies and advisories to overgeneralize PWDs from a national level which in turn is hard to localize and appropriate the necessary support to be given to PWD. It is crucial to identify current gaps and obstacles to overcome specific categories of PWDs to address unequal access and adjust existing policies to accommodate emerging needs.

In this dashboard, a national database is created to unify registered Filipino PWD in each city to provide aid in the development of policies and initiatives that are appropriate for the needs of PWDs. On the side of LGUs, there is recognition that cities should be inclusive in informing, directing, and educating their community, most especially its vulnerable groups like the PWD community of their legal rights and of the LGU's current initiatives or programs. PWD representatives should be included in the decision-making procedures or micro-planning sessions of the LGUs, most especially for their welfare and right to inclusive public spaces by ensuring it is appropriate, accessible, equitable, and inclusive.

Objectives

In creating a national dashboard of registered PWDs local chief executives (LCE) and local leaders are aware of the different categories of PWD present in cities which can provide data-driven, appropriate, and localized responses that can allay the disproportionate inequities that are experienced PWDs to existing policies and plans for future community programs.

Overall, this dashboard aims to:

- Provide insights on the number of PWDs across cities in the Philippines
- Provide insights into equal opportunity & access to public services for PWDs
- Give a general overview of the regional number of registered PWDs from 2011 - 2019

Materials Used

Data used for the dashboard came from relevant documents that were retrieved through an extensive search for government websites through the National Council on Disability Affairs, Philippine Statistics Authority, and Department of Health.

In terms of the platform of the data, the data visualization process is created in Google Data Studio (GDS). GDS is an open-source and online tool of Google that can convert any data into customizable dashboards and reports. It provides easy access for downloading pages and data from the dashboard, as well as, ease of sharing and collaborative work.

Methodology

After the compilation of relevant documents and data from the relevant stakeholders, all of the acquired data were cleaned in correspondence that the GDS system would be able to read. Other data visualization platforms were explored such as PowerBI, Tableau, and ArcGIS, but ultimately Google Data Studio was decoded to be the best platform to host and visualize data due to its open-source and ease of collaborative features. Academic journal articles, gray literature of national government agencies (NGA), and local news were also analyzed to comprehend the extent of inclusiveness, issues faced, and efforts being made by NGAs in consideration of the Philippine PWD community. Subsequently, the dashboard has undergone different rounds of polishing on its metrics, design, and data quality.

The PWD Statistics Dashboard was then presented to different intergovernmental organizations for vetting, the potential for usability, and recommendations on key metrics to be added to the dashboard as new data is presented.

Recommendations and Proposed Next Steps

All members of a community have been placed at great risk from the COVID-19 pandemic and its effects, but PWDs is one of the communities that are most deeply affected because of existing institutional systems and inaccessibility issues that are further intensified during the pandemic. Many PWDs have underlying health and face socio-economic issues that make them more prone to diseases, discriminatory practices, and prejudices resulting in higher rates of infection-related mortality, dropping out of school, and unemployment. The PWD community continues to encounter discrimination and other obstacles while trying to get financial assistance for their means of subsistence, take part in online learning opportunities, or seek protection from violent crime.

Given the number of certain categorical PWDs per city, there is baseline data on where local leaders can create targeted programs, create inclusive public systems, and health benefit allocation that can be addressed. As the data moves forward, here are some recommendations and proposed next steps for local government units and local leaders allay some of these issues faced by PWDs:

- Conducting micro-planning sessions with local advocacy groups that champion PWD communities with local leaders on planning and implementing strategies of existing and/or future programs can be more equitable and inclusive of PWDs present in their cities;
    - COVID-19 materials and news;
    - Open Educational Education materials;
    - Public facilities;
    - Among others
- Providing affordable assistive technology and tools serving persons with disabilities in local pharmacies, clinics, and public hospitals as based on the identified number and category of PWD present in the given city;
    - Affordable assistive devices and tools (wheelchair, hearing aids, magnifier)
    - Medication (mental health, etc)
- Creating more inclusive and accessible public spaces (ie crosswalks that cater to the visually impaired), public transportation, and public architecture institutions should also design access for PWDs (i.e. ramps, lever handles, wide doorways, etc);
- Delivering better traction and clear guidelines on the registration of a PWD card and providing information on its benefits to the cities constituents;

- Creating more inclusive and accessible public spaces (ie crosswalks that cater to the visually impaired), public transportation, and public architecture institutions should also design access for PWDs (i.e. ramps, lever handles, wide doorways, etc);
    - Some existing accessible structures are just made for minimum compliance and are not constructed in accordance with international standards.
        - Braille markings on public signs, pedestrian crossings, and elevator buttons.
- Disaggregating the categorical data of PWDs for more localized and targeted programs. The needed disaggregated data for each city is based on:
    - Type of Mental Health Issue
    - Type of Visual Impairment
    - Type of Cancer
    - Multiple Disabilities
- Mapping of all Persons with Disabilities Affairs Office (PDAO) offices in cities and municipalities.

As the nation moves forward in a post-pandemic society, there is a need for more inclusive and equitable community engagement, programs, and calls for more partnership from local communities and national government agencies to ensure that no one is left behind.

The Philippine PWD Statistics Dashboard
LINK: bit.ly/PHPWDDashboard

Regional View

In the first tab of the dashboard, the Regional Data View, is able to visualize the number of registered PWDs annually from 2010 until 2019 across municipalities and cities in the Philippines. This dashboard view can be filtered by annual and regional data available.

The data view can also visualize the top regions with the highest number of registered PWD per category as seen in Table III. The PWD categories listed are the following: cancer, Deaf, intellectual, learning, mental disability, physical, psychosocial, and speech impairment.

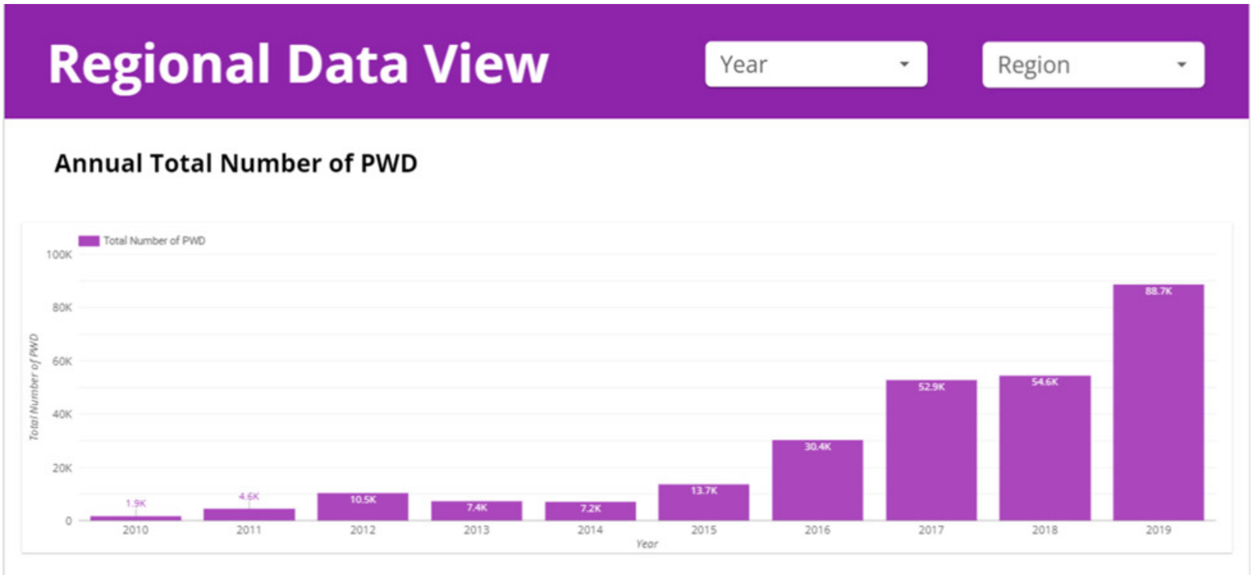


Figure 2.1 Table I: **Annual Total Number of PWDs in the Philippines**

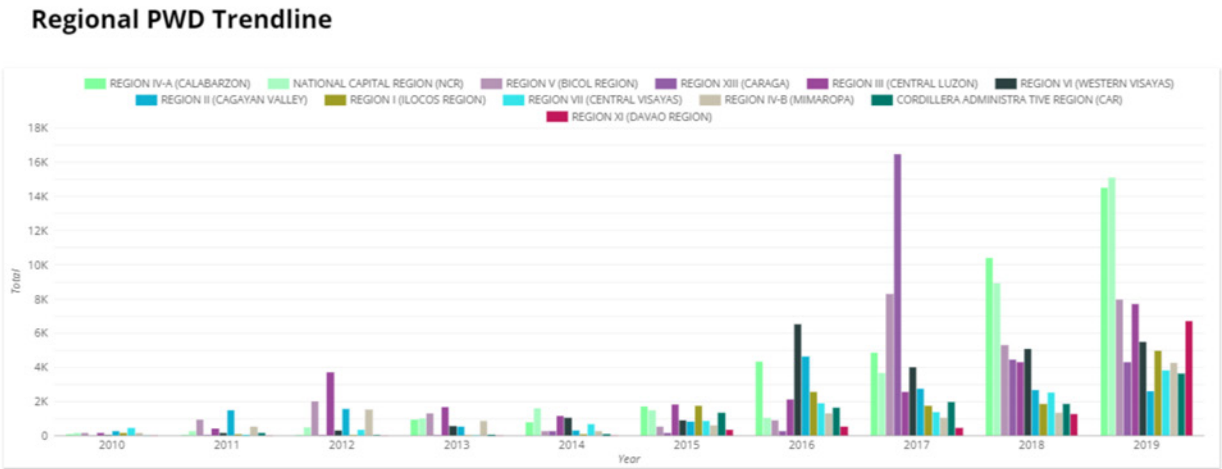


Figure 2.1 Table II: **Annual Regional PWD Trendline**

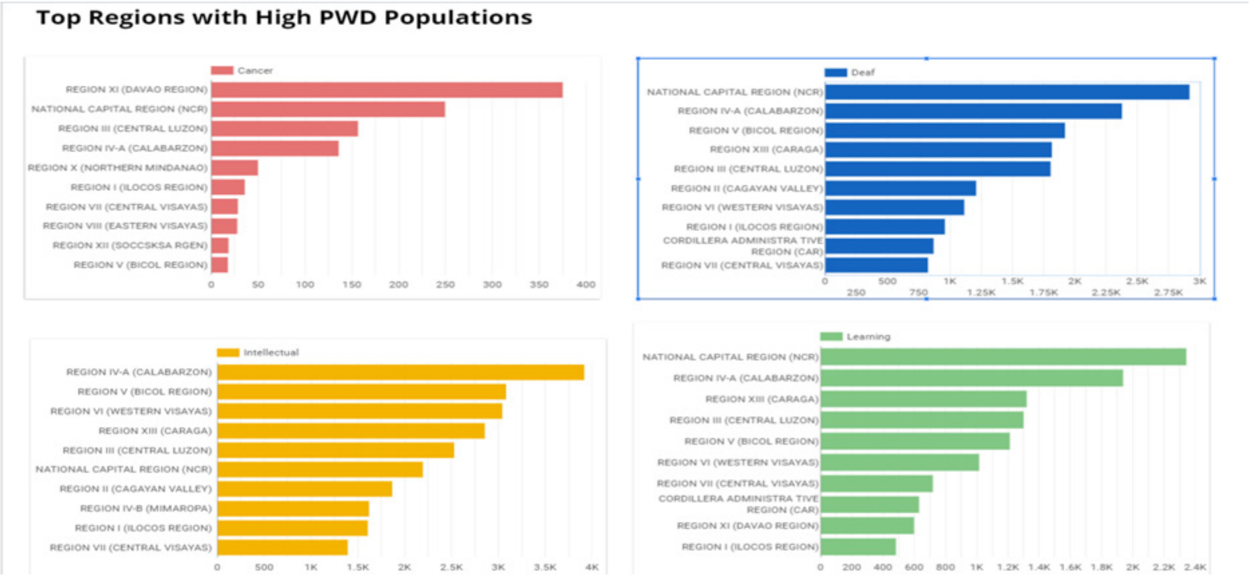


Figure 2.1 Table III. **Top Regions with High PWD Population**

Regional Trendline Per Category

In the second tab, the Regional Trendline Per Category view, it is able to visualize the total number of registered PWD per category in each region of the Philippines. Categories listed are similar to the first tab, Regional View.

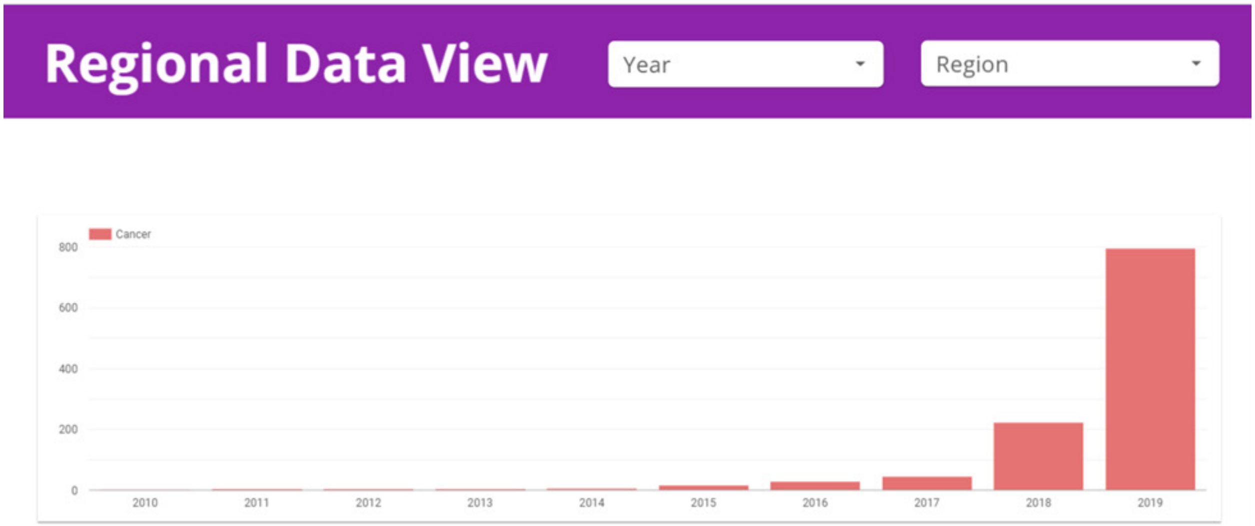
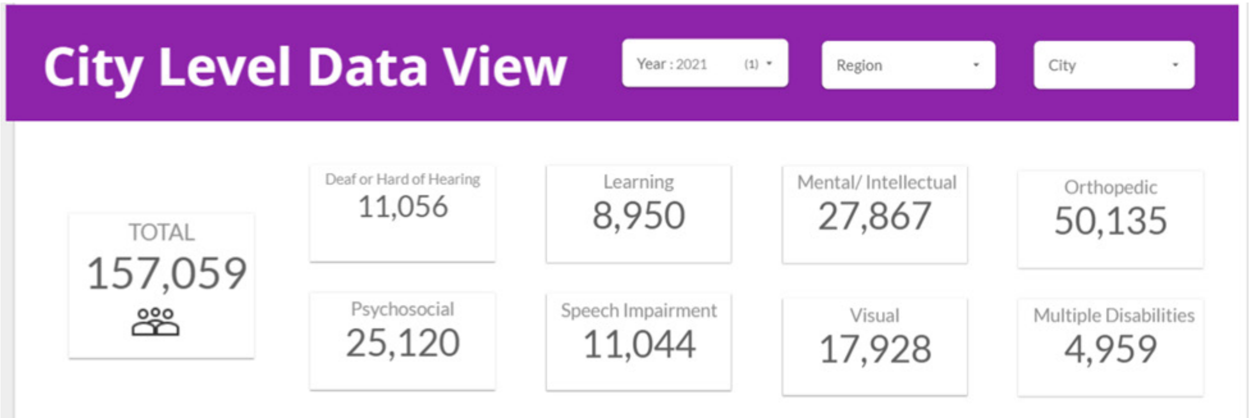


Figure 2.1 Table IV. **Regional Trendline Per Category**

City Data on PWDs

The third tab, City Data on PWDs, visualizes the annual (2020 to 2021), regional, and city-level data on each PWD category. This allows LCEs and local leaders to review the presence of the number of registered PWDs in their communities. In Table VI, the city-level view can be filtered by 100,000 population or the raw number of PWD present in the city. This gives a contextual number for each city.



City Level Data View

Year : 2021 (1) ▾ | Region ▾ | City ▾

TOTAL 157,059

Deaf or Hard of Hearing 11,056

Learning 8,950

Mental/ Intellectual 27,867

Orthopedic 50,135

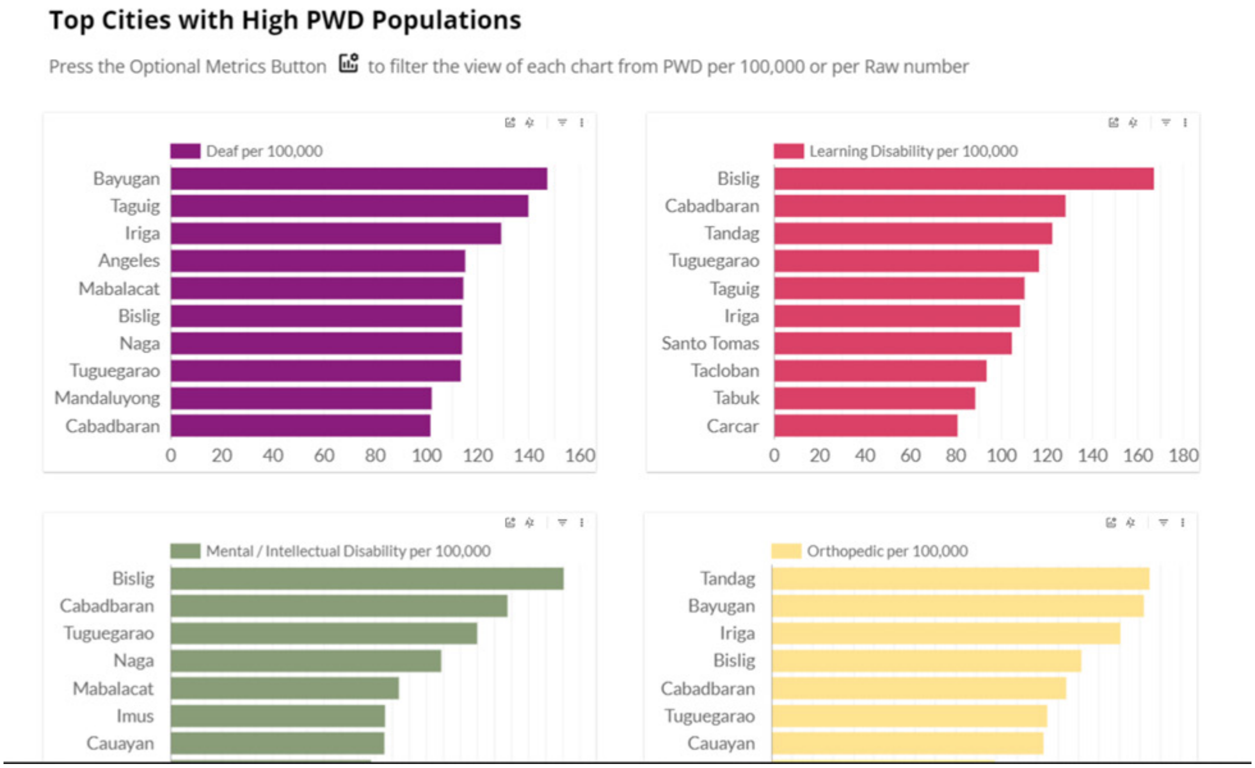
Psychosocial 25,120

Speech Impairment 11,044

Visual 17,928

Multiple Disabilities 4,959

Figure 2.1 Table V. **Total Number of Registered PWDs in Philippine Cities**



Top Cities with High PWD Populations

Press the Optional Metrics Button to filter the view of each chart from PWD per 100,000 or per Raw number

Figure 2.1 Table VI. **Top Cities with High PWD Population**

City Trendline per Category

In the last tab, the City Trendline Per Category view, is able to visualize the total number of registered PWD per category in each city of the Philippines. It can also be filtered down per 100,000 population or its raw number. Regional comparisons for each category may also be studied and dissected. The categories listed are similar to the third tab, City Data on PWDs.

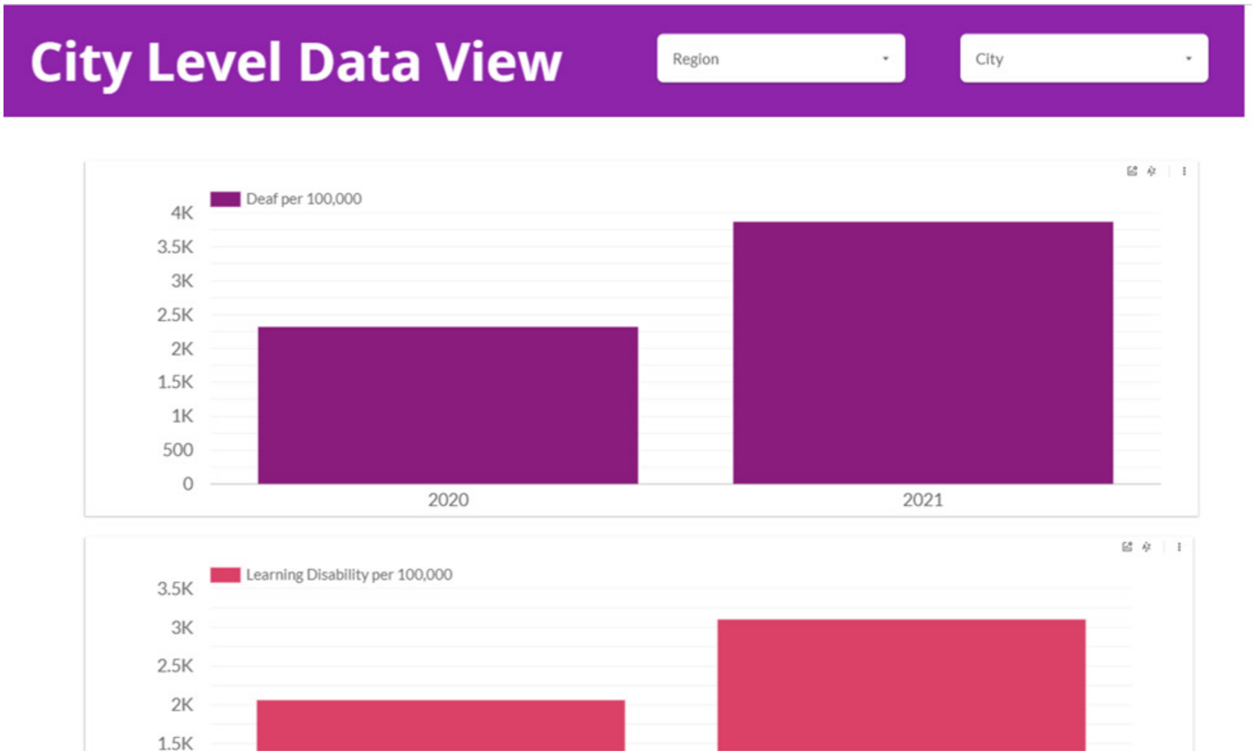


Figure 2.1 Table IV. **City Trendline Per PWD Category**

2.1.4 References

COVID-19 and the Rights of Persons with Disabilities. Office of the High Commissioner for Human Rights. Retrieved from: https://www.ohchr.org/en/covid-19-and-persons-disabilities

Luna F. (2020), Lawmakers want review of disability benefits law amid 'fake PWD' allegations. The Philippine Star. Retrieved from: https://www.philstar.com/headlines/2020/06/21/2022479/lawmakers-want-review-disability-benefits-law-amid-fake-pwd-allegations

National Council on Disability Affairs (NCDA). RA 7277 - Magna Carta of Disabled Persons. Retrieved from: https://www.ncda.gov.ph/disability-laws/implementing-rules-and-regulations-irr/irr-of-republic-act-no-10070/#:~:text=It%20is%20declared%20policy%20of%20RA%20No.%207277,people%20to%20take%20their%20proper%20place%20in%20society.

National Council on Disability Affairs (NCDA). RA 11106- Filipino Sign Language Act. Retrieved from: https://www.ncda.gov.ph/disability-laws/republic-acts/ra-11106/#:~:text=RA%2011106%20%E2%80%93%20An%20Act%20Declaring%20The%20Filipino,And%20Workplaces%20%3A%20National%20Council%20on%20Disability%20Affairs

The Philippine National Deployment and Vaccination Plan for COVID-19 Vaccines. Republic of the Philippines: Inter-Agency Task Force for the Management of Emerging Infectious Diseases (2021). Retrieved from: https://ro7.doh.gov.ph/official-report/870-philippine-national-deployment-and-vaccination-plan-ndvp-for-covid-19-vaccines-interim-plan

Velasco, J. ,Obnial, J.,Pastrana, A., Ang, H., Viacrusis, P., & Lucero-Prisno, D. (2021). COVID-19 and persons with disabilities in the Philippines: A policy analysis. Health Promotion Perspectives. doi: 10.34172/hpp.2021.38. Retrieved from:https://hpp.tbzmed.ac.ir

World Report on Disability. World Health Organization (2011). Retrieved from: https://www.who.int/disabilities/world_report/2011/report.pdf

REX P. BRINGULA

2.2
USE CASE 2
Towards the Creation of a Barangay Health Vulnerability Index

2.2.1 About the authors - UE and DHVSU

REX P. BRINGULA

Rex P. Bringula is a full professor in the College of Computer Studies and Systems at the University of the East (UE). As a Department of Science and Technology (DOST) scholar, he earned a BS in Computer Science from UE. He earned his master's degree in Information Technology and his doctorate in Technology Management from the Technological University of the Philippines. He is completing his PhD in Computer Science dissertation at Ateneo de Manila University. He has participated in numerous school- and government-funded research projects, as well as local and international conferences. He has published more than 100 research articles indexed in major databases. He has received numerous national and international awards for his research.

JOHN PAUL P. MIRANDA

John Paul P. Miranda is an associate professor at the Don Honorio Ventura State University (DHVSU) Mexico Campus. Mr. Miranda is concurrently the International Linkages Coordinator and Secretariat for Scholarship Grants of DHVSU. He obtained his Master of Information Technology thru Commission on Higher Education's K to 12 Transition Program Scholarship program at Systems Plus College Foundation in Angeles City, Philippines. He is currently finishing his dissertation for his degree, Doctor of Information Technology at the University of the East in Manila, Philippines. Mr. Miranda also completed the Digital Transformation Training for Higher Education Institutions in the Philippines: An Executive Education Program sponsored by the Commission on Higher Education in partnership with Mapua University in the Philippines and Carnegie Mellon University in Australia. He is an active and proud member of National Research Council of the Philippines and the Philippine Society of Information Technology Educators. Mr. Miranda also serves as a reviewer for multiple international journals. Most of his publications are indexed in Scopus and Clarivate Analytics within the areas of data science, data mining, computer science/IT education, and software development. He also presented multiple papers in both national and international fora.

JOHN PAUL P. MIRANDA

ROSELLE S. BASA

Roselle S. Basa is an Associate Professor in the College of Computer Studies and Systems at the University of the East (UE-CCSS). She earned her BS in Computer Science in UE. She completed her Master's Degree in Information Technology and all academic units of Ph.D. in Technology Management from the Technological University of the Philippines.

ROSELLE S. BASA

Concurrently, she is also the Program Coordinator for Instructional Technology in the Office of Curriculum Development and Instruction (OCDI) and the Data Protection Officer of UE.

2.2.2 Executive Summary/Abstract

Abstract

This case study developed the Barangay Health Vulnerability Index (BHeVI) and its associated survey instrument. The BHeVI instrument is composed of 20 questions that determined the demographic profile of the respondents and the household health vulnerability in times of disaster. The instrument was developed through a series of refinements. The BHeVI formula was derived to determine the health vulnerability of a household. Initially, the researchers employed ranking of the 3 identified diseases that might need to be prioritized in the event of a disaster. However, despite two rounds of ranking the diseases, the researchers did not reach a logical consensus. The instrument was also sent to six health professionals but a similar problem arose. This led to rating the diseases instead of ranking them. Consequently, a high correlation of ratings from the six health professionals was achieved. Thus, the rating was used in determining the priority of the diseases. The BHeVI formula was once again derived and was used to determine the health vulnerability of 138 respondents from one community (i.e., barangay) in Orani, Bataan, Philippines. More than a quarter of the households in the dataset are vulnerable to a moderate extent. Five cases of household vulnerability indicate that there is one household that is highly vulnerable during disasters considering their health conditions. These findings suggest that the BHeVI can be utilized to determine the health vulnerability of the household. Multi-sectoral cooperation and collaboration are needed to minimize and manage household health vulnerability in a community. Specifically, the barangay that participated in this study and the Department of Social Welfare and Development are the direct beneficiaries of the findings of this study. Experts' pieces of opinions from the Department of Health may also be solicited to further enhance the BHeVI instrument.

2.2.3 Use-Case Body

Introduction

Vulnerability is the "inability to resist a hazard or to respond when a disaster has occurred" (p. 8, United Nations Office for Disaster Risk Reduction, 2004). During a health crisis or natural disaster, health-vulnerable households are disproportionately affected. Despite this pressing concern, there is almost no health vulnerability index at the household level. One local study formulated a health index at the barangay level (Caalim et al., 2021). However, the study determined the health vulnerability of different communities and not the households within each community.

The gap in the literature may result in unsuitable health and disaster prevention, mitigation, and response planning for the households of the community. The goal of this project is to create a BHeVI instrument and use it to assess the household health vulnerability index. The outcome of the project is expected to serve as a foundation for the development of governmental policies and decision-making based on publicly available datasets and information.

Methodology

Development of Barangay Health Vulnerability Index Instrument

The study employed two stages of development of the Barangay Health Vulnerability Index Instrument (BHeVII). In the first stage, the researchers listed 10 leading causes of morbidity in the Philippines (Department of Health, 2022) plus three additional diseases added by the authors. The authors deliberated on how to measure the 13 diseases. (Arthritis, Asthma/TB / lung-related diseases, Cancer, Coronary heart problems, Dementia / Alzheimer's, Diabetes, Epilepsy, Hyperacidity/Ulcer, Hypertension/High blood, Kidney disease, Liver-related diseases, Skin allergy, and Thyroid problems).

The second stage involved the pilot-testing of the instrument on 21 residents of the barangay. The research instrument was revised based on the comments or clarifications of the respondents (e.g., unclear instruction, vague words, missing items, etc.). One of the revisions entailed the inclusion of three more diseases in the survey form. The three diseases included were hydrocephalus, Polycystic Ovary Syndrome (PCOS), and cerebral palsy. The final instrument was achieved with 16 diseases in determining the number of family members who are vulnerable in each of the diseases. The instrument can be answered using a scale from 0 (none of the family members) to 5 (at least 5 family members).

Development of the BHeVI Formula

Originally, the researchers ranked the 13 diseases from 1 to 13, where 1 is the highest priority. They were guided by the instruction "In the event of a disaster, prioritize the emergency response for the following health conditions, with 1 being the highest priority". Afterward, six health professionals were requested to rank the 13 diseases. Inter-rater reliability (IRR) test was utilized to compute the level of agreement. In the event of disagreement (i.e., IRR < 0.70), the evaluation of the instrument was revised. With the low IRR, the researchers decided to change how diseases were measured. Instead of ranking, the health professionals were again requested to rate the new set of diseases (i.e., 16 diseases) from 1 (highest priority) to 10 (least priority). They were guided by the instruction "Considering the mobility of a person in the event of a disaster, rate each health condition based on the level of priority for medicine and medical care (starting with 1 for the highest priority)". Kruskal-Wallis one-way ANOVA was employed to determine if the responses were significantly different from one another. The weighted mean was used to determine the weight of each disease.

After establishing whether the responses of the six professionals did not significantly differ from one another, the final set of diseases was sent to five barangay health workers (BHWs) who are part of the local disaster response unit of the community. The five BHWs were also requested to rate the diseases. They were involved in rating the diseases since they also have first-hand experience in responding to the citizens' needs in times of disaster. The final weights for each disease were calculated by taking the average of the health professionals' and BHWs' ratings.

Study 2: Implementation of the BHeVI Formula

Research Setting, Household Population, Determination of Sample Size, and Sample Size

The study was conducted in Tugatog, Orani, Bataan, Philippines. Tugatog has a network of esteros leading to a nearby creek. The creek overflows during rainy seasons or typhoons, flooding some of the houses near the creek. Meanwhile, the town of Orani is a first-class municipality (GovPH, 2022). It ranked 418th in terms of resiliency. The descriptive location of Tugatog (i.e., the case study site) is given in Table 1. There are 1,694 households on the case study site with an average of 4.05 people per household. However, based on the interview with the barangay chairman, there are now about 2,500 households. Using Slovin's formula where n = 0.10, a sample size of 97 households was computed.

Indicators	Data
Location	Tugatog is located on the island of Luzon at approximately 14.8007, 120.5209 (Figure 1). At these coordinates, the elevation is estimated to be 16.0 meters (52.5 feet) above mean sea level.
No. of Households	1,694 households
Average Household Size	4.05
Population (2020 Census)	8,138
Population Relative to the Municipality	11.57%

Figure 2.2 Table 1. **Case Study Site (source: PhilAtlas, 2022)**

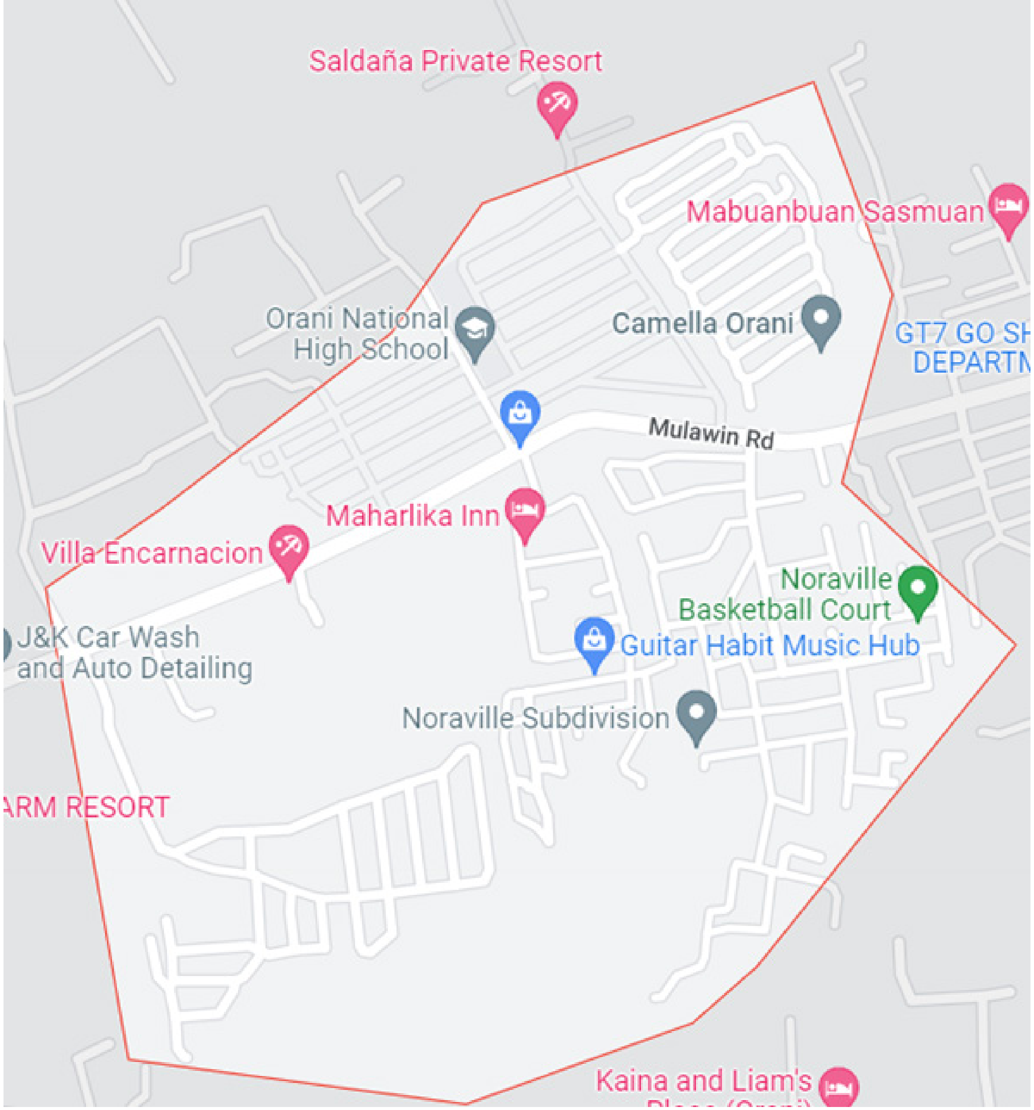


Figure 2.2 1. **Study Site Map (Google Map)**

Data Gathering Procedure

A permit was secured from the Office of the Barangay Chairman to distribute 120 printed survey form. To encourage participation, a raffle was facilitated to award PHP 500 cash to five lucky winners. Each of the four enumerators who facilitated the survey distributed 30 printed survey forms. The household was selected by counting three houses starting from the barangay hall. Three printed survey forms were not returned. Thus, 117 printed survey forms were retrieved.

An online version of the survey form was deployed using Google Forms via the Facebook page of the barangay. Twenty-one participants answered the online survey form forming the dataset of the study with a total of 138 records.

Data Analysis

Descriptive statistics such as frequency counts, percentages, mean (M), and standard deviations (SD) were used to present the data. The BHeVI was also employed to determine the household's health vulnerability. The household vulnerability index ranges and categories were based on Table 2 (Caalim et al., 2021).

Range Values	Category
0.0000 to 0.2000	Very Low Vulnerability
0.2001 to 0.4000	Low Vulnerability
0.4001 to 0.6000	Moderate Vulnerability
0.6001 to 0.8000	High Vulnerability
0.8001 to 1.000	Very High Vulnerability

Figure 2.2 Table 2. **Vulnerability Index Category**

Results and Findings

Study 1: BHeVI Instrument and Formula Derivation,

To begin, the researchers and six health professionals ranked the 13 diseases. However, ranking the diseases posed methodological challenges. The first challenge involved low inter-rater reliability (IRR) even after two rounds of ranking (less than 40% inter-reliability score) from both sets of raters. The second challenge was that even if an acceptable IRR score was reached, the weight (w) of an item with different ranking scores could not be established.

The first formula derived was given in Equation 1.

$$H = \frac{\sum_{i=1}^{n} w_i f_i}{5n}$$

Equation 1

where:

H = household health vulnerability index

i = index

n = number of diseases in the questionnaire

w = weight of the disease based on the ranking of the health professionals

f = frequency of people in a household having the disease w

Figure 2.2 Equation 1

H stands for the household health vulnerability index. This can be calculated by dividing the sum of the product of the weight ranked w of the disease and the number of people f in a household having the disease w by the product of constant 5 and n number of diseases in the questionnaire. The constant 5 represents the maximum number of people in a household who have the disease. However, as mentioned, this initial formula cannot handle the different rankings of the raters.

The researchers decided to change the evaluation of the diseases. Instead of a ranking, the diseases were rated from 1 to 10, where 1 is the highest priority. Furthermore, the researchers no longer rated the diseases. Instead, six health professionals (a nurse, a radio technician, and four doctors) were asked to rate the diseases. The responses were then evaluated using the Kruskal-Wallis H test. The Kruskal-Wallis H test indicated that there is a non-significant difference in the ratings between the different groups, $\chi 2(5) = 3.89$, $p = 0.565$, with a mean rank score of 55.66 for Group 1, 47.72 for Group 2, 39.06 for Group 3, 44.31 for Group 4, 53 for Group 5, and 51.25 for Group 6. These findings implied that the ratings of the health professionals are very similar to one another. Thus, the ratings were eligible for analysis.

The ratings of the health professionals for each disease were computed to determine the weight of each disease. The ratings were reverse-coded. The weights of each disease are given in Table 3. The BHeVI formula was also revised (Equation 2). The revised formula used the same variables except that w is now a rating. The numerator denotes the sum of the product of the weight of the disease and the number of people in the household who suffer from a disease. This is then divided by the sum of the product of 5 (the maximum rating on the scale) and the weight of the disease. The constant 5 represents a household with at least 5 family members suffering from a specific disease. Both the numerator and denominator are normalized by finding their logarithm values.

$$H = \frac{\log\left(\sum_{i=1}^{d} w_i f_i\right)}{\log\left(\sum_{i=1}^{d} 5w_i\right)}$$

Equation 2

where: H = household health vulnerability index

i = index

d = number of diseases in the instrument

w = weight of each disease

f = frequency of people in a household having the disease d

Figure 2.2 Equation 2

As shown in Table 3, people with coronary heart disease had the highest mean rating. Patients with hydrocephalus had the second highest mean rating. Cancer patients also have the highest average weight on the list. Another high priority is a cerebral palsy patient. Epileptics are ranked fifth in terms of disease priority. Skin allergies had the lowest mean rating of any disease.

The weights (w) of the diseases are in increasing order. The higher the mean rating of the disease, the higher the weight it will receive. The disease with the lowest mean rating received one point (i.e., lowest priority), while the disease with the highest mean rating received the highest points (i.e., highest priority). Since there are 16 diseases, coronary heart disease will have 16 points.

Disease	Health Professionals	BHWs	Mean Rating	Weight *w*
Coronary health problems	9.5000	4.600	7.0500	16
Hydrocephalus	5.8333	8.200	7.0167	15
Cancer	8.6667	4.600	6.6333	14
Cerebral palsy	6.0000	7.200	6.6000	13
Epilepsy	5.1667	8.000	6.5833	12
Asthma/ TB/ Lung-related diseases	8.0000	5.000	6.5000	11
Liver-related diseases (e.g., hepatitis)	6.0000	6.200	6.1000	10
Kidney diseases	8.0000	3.400	5.7000	9
Hypertension/High blood	8.6667	1.000	4.8333	8
Thyroid problems	3.0000	6.000	4.5000	7
Diabetes	7.0000	1.800	4.4000	6
Dementia/ Alzheimer	4.8333	3.200	4.0167	5
PCOS	2.1667	5.400	3.7833	4
Skin allergy	1.8333	3.800	2.8167	3
Arthritis	2.3333	2.600	2.4667	2
Hyperacidity/ Ulcer	2.8333	1.600	2.2167	1

Figure 2.2 Table 3. **Weight of Each Disease**

Study 2. Results of the Implementation of the BHeVI Formula

Table 4 shows that almost three-quarters (73.2%) of the households belong to the low and very low vulnerability groups. More than a quarter (26.1%) of the household population belongs to the moderate vulnerability group. Only one household belonged to the highly vulnerable group. This means that generally, the households in this community are healthy. Nevertheless, the high and moderate vulnerable group may need assistance from the LGU in times of disaster or health crisis.

Category	*f*	*%*
Very High Vulnerability	0	0
High Vulnerability	1	0.7
Moderate Vulnerability	36	26.1
Low Vulnerability	51	37.0
Very Low Vulnerability	50	36.2

Figure 2.24 Table 4. **Frequency Distribution of Health Vulnerability**

Five sample cases in a dataset were selected to further describe the results (Table 5). The first case is about a family member with 3–4 family members. Three of the family members are elderly (i.e., senior citizens). The youngest is 42 years old. The household has a vulnerability index of 0.6514. They have a total of 14 diseases in their household. All of them suffer from diabetes. The three elderly people suffer from hypertension and arthritis. Thus, their household must be prepared for disasters and the local government unit must be informed about their vulnerabilities. The household itself must develop a disaster plan so that it can remain mobile in time of disaster.

Case 2 is of moderate vulnerability. This household consists of 5–6 family members living in their family-owned house. The youngest member is 22 years old, while the oldest is 57 years old. Only one member of the family works as a tricycle driver. The household is highly vulnerable because one member is suffering from asthma and another two members have hypertension.

The third case has a household low vulnerability index of 0.3810. The household is composed of at least ten members, where the youngest is 6 and the oldest is 63. They live in a family-owned house. One member of the family works as a government employee, and they have a "sari-sari" store as an additional source of income. Two of the family members have diabetes.

The fourth case has a very low vulnerability index of 0.1684. The household is composed of 3–4 members. They are renting a house. The household is relatively young: the youngest is 2 years old, while the oldest is 33 years old. The primary source of income is from working as an employee of a private company. Arthritis and hyperacidity are the illnesses reported in this household.

The last case is about a family of 7–8 members. The youngest member of the family is 5 months old and the oldest is 62 years old. One member of the family works as a tricycle driver. Only one member of the family suffers from arthritis. The household has a very low vulnerable index.

Cases	Description	BHeVI	Category
Case 1	A family with 3-4 members living in their family-owned house. The youngest is 42 years old while the oldest is 73 years old. Two members have a source of income (i.e., one pensioner, one a private company employee). All have diabetes. Three are senior citizens. All three senior citizens have hypertension and arthritis.	0.6514	High Vulnerability
Case 2	A family with 5-6 members living in their family-owned house. The youngest is 22 years old while the oldest is 57 years old. Only one member has a job (i.e., a tricycle driver). One member of the family has asthma, and two others have hypertension.	0.5053	Moderate Vulnerability
Case 3	A family with at least ten members. The youngest is 6 while the oldest is 63. They are living family-owned house. One member of the family is a government employee. They also have a sari-sari store. Two members of the family are diabetic.	0.3810	Low Vulnerability
Case 4	A family with 3-4 members. They rent a house. The youngest is 2 years old while the oldest is 33. One member of the family is an employee of a private company. A member of the family has arthritis and another has hyperacidity.	0.1684	Very Low Vulnerability
Case 5	A family with 7-8 members. The youngest is 5 months old while the oldest is 62 years old. Their house is family-owned. Only one member of the family works as a tricycle driver. One member of the family has arthritis.	0.1063	Very Low Vulnerability

Figure 2.2 Table 5. **Sample Cases in a Dataset**

Recommendations for Use and Future Work

The findings of the study can be beneficial in four ways. First, the findings of the study can be utilized by the LGU where this study was conducted, i.e., the barangay level. With these findings, the barangay may deploy health programs, information campaigns, and disaster preparedness activities suitable to the current health conditions of the community. For example, a nutrition information campaign and an exercise program for people with hypertension can be instituted by the LGU.

Second, the study can be replicated in other barangays. In particular, the BHeVI instrument can be used to determine the level of household health vulnerability. A comparative study can be initiated to determine if there are similarities or differences in the health conditions of different communities. LGU officials could use the results of BHeVI to develop policies that would facilitate disaster and health crisis response and reduction, as well as fully optimize their relief and emergency activities during health crises.

Third, the Department of Social Welfare and Development (DSWD) may utilize the data collected and the BHeVI formula. The dataset and results of the BHeVI may provide the DSWD with a data-driven action plan for policies, programs, and activities in disseminating relief goods in the community. The expertise of the Department of Health may also be tapped to further enhance the BHeVI formula.

Finally, a web-based information system could be developed to expand the reach of information dissemination. People can use the website to calculate the vulnerability of their household. The purpose of the BHeVI results is not to instill fear, but to provide information useful for disaster preparedness. The information could inform the household to make necessary preparations so that it could minimize the impact of disasters on their households.

Acknowledgments

The researchers are grateful to LayerTech for funding this research through Project OCDEx. They are also grateful to Barangay Chairman Mauro Dela Rosa, enumerators, and study participants for their invaluable assistance. Finally, the authors would like to express their gratitude to the school administrators of the authors’ affiliations for their ongoing assistance.

2.2.4 References

Caalim, A. N. A., Aquino, C. A., Bongabong, R. C. V., Osia, S. A., Ong, A. K. S., & German, J. D. (2021). Health Vulnerability to COVID-19: A Barangay Level Assessment for Bocaue, Bulacan. In Proceedings of the Second Asia Pacific International Conference on Industrial Engineering and Operations Management Surakarta, Indonesia (pp. 1806-185). Retrieved from http://ieomsociety.org/proceedings/2021indonesia/347.pdf

Department of Health. (2022). What are the leading causes of mortality in the Philippines? Retrieved from https://doh.gov.ph/node/1058

Google Developers. (2022). Normalization. Retrieved from https://developers.google.com/machine-learning/data-prep/transform/normalization

GovPH. (2022). Cities and Municipalities Competitive Index. Retrieved from https://cmci.dti.gov.ph/prov-profile.php?prov=Bataan&year=2020

PhilAtlas. (2022). Tugatog, Municipality of Orani, Province of Bataan. Retrieved from https://www.philatlas.com/luzon/r03/bataan/orani/tugatog.html

United Nations Office for Disaster Risk Reduction. (2004). Learning from today's disasters for tomorrow's hazards: 2004 World Disaster Reduction Campaign. Retrieved from https://www.unisdr.org/2004/campaign/booklet-eng/Pagina8ing.pdf

JOHN RAYMOND BARAJAS

PEE JAY GEALONE

NICO ASPRA

2.3 USE CASE 2 What is an anomalous awarded tender?: Identification of anomalies through outlier detection machine learning algorithms

2.3.1 About the authors - BU

ENGR. JOHN RAYMOND BARAJAS

Engr. John Raymond Barajas is a licensed chemical engineer and a data scientist. He is graduate of Master of Science in Chemical Engineering at De La Salle University - Manila and a graduate of Master of Science in Data Science at Asian Institute of Management. He is currently a full-time faculty in Bicol University College of Engineering.

ENGR. PEE JAY GEALONE

Engr. Pee Jay Gealone is a registered electrical engineer. He is a graduate of Master in Engineering Technology at Camarines Sur Polytechnic Colleges. He is currently a full-time faculty in Bicol University College of Engineering.

ENGR. NICO ASPRA

Engr. Nico Aspra is a registered mechanical engineer. He is currently a full-time faculty in Bicol University College of Industrial Technology.

ENGR. ARPON LUCERO, JR.

Engr. Arpon Lucero, Jr. is a licensed chemical engineer. He is a graduate of Master of Science in Environmental Science at Daegu Catholic University. He is currently a full-time faculty in Bicol University College of Engineering.

ARPON LUCERO, JR.

ENGR. OLIVER PADUA

Engr. Oliver Padua is a licensed civil engineer. He is currently a full-time faculty in Bicol University College of Engineering.

OLIVER PADUA

ENGR. MARBEN RAMOS

Engr. Marben Ramos is a licensed electrical engineer, He currently serves as the Dean of the School of Engineering and Computer Studies Electrical Engineering Department.

MARBEN RAMOS

2.3.2 Abstract/ Executive Summary

Abstract

This work developed a data science pipeline for the identification of potentially anomalous tenders posted in the Philippine Government Procurement System (PhilGEPS) from January 01, 2020 to June 30, 2021. This pipeline is divided into five phases, namely, (1) extraction of procurement data, (2) cleaning of data and feature engineering, (3) selection of features, (4) training and development of anomaly detection model, and (5) interpretation of the trained anomaly detection model. Through the combination of decision scores derived from five trained outlier detection algorithms, 1,672 out of 113,585 awarded tenders have been flagged as anomalous. These anomalous tenders are equivalent to potential losses totaling Php68.9 billion. Interpretation of how the trained model predicts anomalous tenders revealed that the causes of such anomalies are likely due to (1) excessively large budget allocation per line-item stipulated in the tender and (2) tenders exceeding the prescribed timeline either on the issuance of notice of award or notice to proceed documents.

Highlights

- 1,672 tenders totaling to Php68.9 billion have been identified to be potentially anomalous.
- Tenders given "Notice to Proceed" issuances prior to the release of "Notice of Award" issuance are likely to be anomalous.
- Tenders given "Notice to Proceed" issuances 47 calendar days after the release of "Notice of Award" issuance are likely to be anomalous.

2.3.3 Use Case Body

Introduction

The Philippines annually enacts into law the General Appropriations Act (GAA) which details the allocated budget of the country for the succeeding fiscal year. In just a matter of four years (from 2019-2022), the national budget of the country has reached Php5.024 trillion in 2022, which represents an average annual increase in the enacted budget of 11%. With this trend, it is thus projected that the Philippines 2023 would enact a budget totaling Php5.5 trillion.

Because of the recently reported scandals (e.g. Pharmally), there has already been public unrest with regard to the utilization of funds for public procurement. As of this writing, the Commission on Audit (COA) further fuels this unrest by flagging the laptops procured by the Department of Education amounting to Php2.4 billion as anomalous, and hence, the call for effective utilization of the Philippine budget has never been this louder.

While COA releases results of audits annually, post-audits generally last at least a year due to the voluminous number of tenders to be audited and the lack of digitalized documents for ease of evaluation. With such challenges at hand, it would be near impossible for auditors to scrutinize tenders posted by a government agency. In an attempt to address this concern, the overall objective then of this work is to provide a means for COA auditors to systematically prioritize the evaluation of tenders based on its potential to contain suspicious characteristics. Through the use of the PhilGEPS dataset, this work finally implemented a data science pipeline that would develop a machine learning model capable of identifying anomalous tenders.

Data

Dataset Description

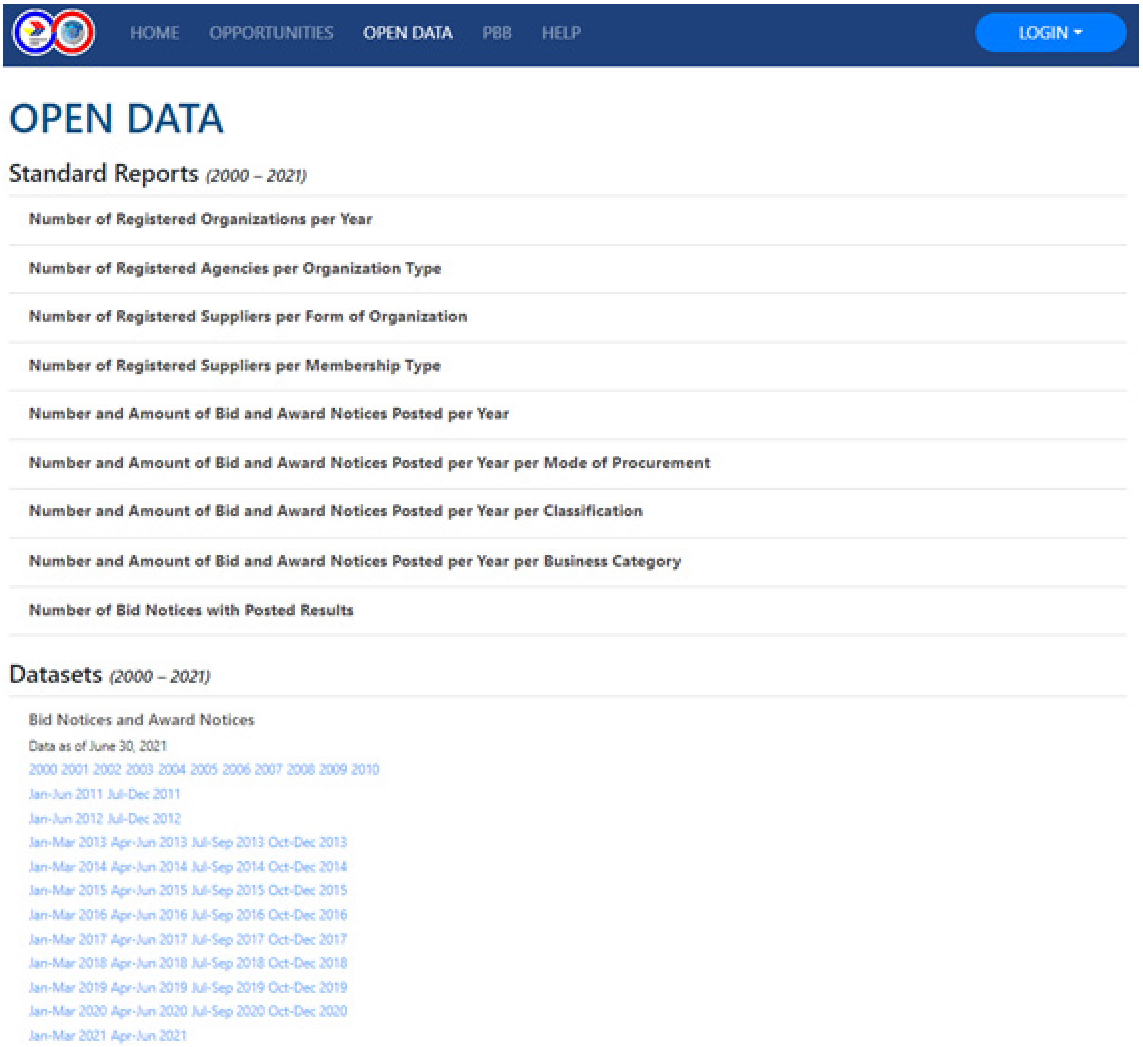
The dataset used for this use-case was taken from the official website of PhilGEPS [1]. The period covered considered in this work was from the year 2020 to the year 2021. This comprised 1.38 million records which represented about 0.95 million unique tenders. As summarized in Table 1, the PhilGEPS dataset contains forty (40) features that provide details on the tenders posted between 2020 and 2021.

Feature Name	*Feature Description*	*Data Type*
Organization Name	Name of the procuring entity	string
Reference ID	Unique identifier of the bid	int
Solicitation No.	Identifier for the document	string
Notice Title	Title of the bid	string
Publish Date	Date bid was published	date
Classification	Class of the bid	string
Notice Type	Type of the bid	string
Business Category	Business category of the bid	string
Funding Source	Where the funding source for procurement will come from	string
Funding Instrument	The legal basis for the use of funds	string
Procurement Mode	Type of procurement mode	string
Trade Agreement	Rules and regulations to be followed for procurement	string
Approved Budget of the Contract	Total approved budget of the contract	float
Area of Delivery	Location of the delivery	string
Contract Duration	Duration of the contract	float
Calendar Type	Type of duration (e.g. days, months, years)	string
Line Item No.	Unique identification number of line items in a bid	int
Item Name	Name of items to be procured	string
Item Desc	Description of items to be procured	string
Quantity	Quantity of the line item	int
UOM	Unit of measure for the quantity of line item	string

Item Budget	Budget allocation for the line item	float
PreBid Date	Date the prebid was set	date
Closing Date	Date the bids will close	date
Notice Status	Status of bid whether it is closed, awarded, or cancelled	string
Award No.	Identifier for awarded bid	int
Award Title	Title of the award document	string
Award Type	Type of the award document issued	string
UNSPSC code	UNSPSC code for the line item	string
UNSPSC Description	Description following the UNSPSC	string
Awardee Corporate Title	Corporate title of the awardee	string
Contract Amount	Amount of the contract	float
Contract No	Identification number for the contract	int
Publish Date (Award)	Date when the award was published in PhilGEPS	date
Award Date	Date when the bid was awarded as shown in award document	date
Notice to Proceed Date	Date when the Notice to Proceed document was issued	date
Contract Effectivity Date	Date when the contract starts	date
Contract End Date	Date when the contract ends	date
Reason for Award	Reason why the award was given to the winning bidder (e.g. lowest calculated responsive bid)	string
Award Status	Status of award whether is updated, posted, or cancelled	string

Figure 2.3 Table 1. **Features of the PhilGEPS Dataset**

However, as of this writing, available data for bid notices and award notices were only up to the 2nd quarter of 2021 (see Figure 1 for a snapshot of the PhilGEPS website where the dataset was taken from). This is possibly due to the ongoing upgrade and migration of uploaded data from the old PhilGEPS website to the new modernized website. The analysis therefore of this dataset was limited only to tenders posted until June of 2021.



OPEN DATA

Standard Reports *(2000 – 2021)*

Number of Registered Organizations per Year

Number of Registered Agencies per Organization Type

Number of Registered Suppliers per Form of Organization

Number of Registered Suppliers per Membership Type

Number and Amount of Bid and Award Notices Posted per Year

Number and Amount of Bid and Award Notices Posted per Year per Mode of Procurement

Number and Amount of Bid and Award Notices Posted per Year per Classification

Number and Amount of Bid and Award Notices Posted per Year per Business Category

Number of Bid Notices with Posted Results

Datasets *(2000 – 2021)*

Bid Notices and Award Notices
Data as of June 30, 2021
2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
Jan-Jun 2011 Jul-Dec 2011
Jan-Jun 2012 Jul-Dec 2012
Jan-Mar 2013 Apr-Jun 2013 Jul-Sep 2013 Oct-Dec 2013
Jan-Mar 2014 Apr-Jun 2014 Jul-Sep 2014 Oct-Dec 2014
Jan-Mar 2015 Apr-Jun 2015 Jul-Sep 2015 Oct-Dec 2015
Jan-Mar 2016 Apr-Jun 2016 Jul-Sep 2016 Oct-Dec 2016
Jan-Mar 2017 Apr-Jun 2017 Jul-Sep 2017 Oct-Dec 2017
Jan-Mar 2018 Apr-Jun 2018 Jul-Sep 2018 Oct-Dec 2018
Jan-Mar 2019 Apr-Jun 2019 Jul-Sep 2019 Oct-Dec 2019
Jan-Mar 2020 Apr-Jun 2020 Jul-Sep 2020 Oct-Dec 2020
Jan-Mar 2021 Apr-Jun 2021

Figure 2.3 - 1. **PhilGEPS website for open data**

Figure 1. Snapshot of the PhilGEPS website for open data. This is the repository of open procurement data for bid notices and award notices posted from the year 2000 until the year 2021. The datasets are made available on a per-quarter basis.

Derived Features

To extract more data from the PhilGEPS dataset that would be relevant to achieving the objective of this project, additional features were also derived. Since the timeline when a tender is awarded is not directly reflected in the base features of the dataset, the time (in days) it took between each phase of the procurement process was also calculated. As based on the revised implementing rules and regulations of the Philippine Procurement Law (i.e. Republic Act 9184) [2], features summarized in Table 2 were hence derived.

Feature Name	*Feature Description*	*Data Type*
Contract Duration (Days)	Total duration of contract in days	int
Publish-PreBid (Days)	Difference in days between the PreBid Date and the Publish Date	int
PreBid-Closing (Days)	Difference in days between the Closing Date and the PreBid Date	int
Closing-Award (Days)	Difference in days between the Award Date and the Closing Date	int
Award-NTP (Days)	Difference in days between the Notice to Proceed Date and the Awarded Date	int
Quarter	Quarter when the bid was published	int

Figure 2.3 Table 2. **Derived Features from the PhilGEPS Dataset**

Tender Selection

To allow the selected machine learning models to find anomalies at any particular point in time of procurement, records that have an award status of "Updated" were the only ones considered for analysis since these are the tenders that were validated to have completed the public bidding process (i.e. up to the issuance of the "Notice to Proceed" document). While additional records could be likely acquired through data imputation, this approach was highly impractical to do so since the percentage of missing values under the "Notice to Proceed Date" feature was 88%. Hence, the number of records used for model training was about 136,000 completed tenders.

Feature Selection

As summarized in Table 1, the dataset (with derived features) used generally contained numeric and categoric features, of which 13 are numeric, 25 are categoric, and 8 are time-stamps. Exploration of the data revealed, however, that irrelevant features were included in the dataset. These features identified were features that contained unique identifiers and text (e.g. "Notice Title", "Item Desc"). With these irrelevant features identified, the number of features to be used for model training was reduced to 19 from 46 (see Table 3 for the list of the retained features).

Feature Name	Feature Description	Feature Type
Classification	Class of the bid	categoric
Notice Type	Type of the bid	categoric
Business Category	Business category of the bid	categoric
Funding Source	Where the funding source for procurement will come from	categoric
Funding Instrument	The legal basis for the use of funds	categoric
Procurement Mode	Type of procurement mode	categoric
Trade Agreement	Rules and regulations to be followed for procurement	categoric
Approved Budget of the Contract	Total approved budget of the contract	numeric
Area of Delivery	Location of the delivery	categoric
Quantity	Quantity of the line item	numeric
Item Budget	Budget allocation for the line item	numeric
Award Type	Type of the award document issued	categoric
Contract Amount	Amount of the contract	numeric
Reason for Award	Reason why the award was given to the winning bidder (e.g. lowest calculated responsive bid)	categoric
Contract Duration (Days)	Total duration of contract in days	numeric
Publish-PreBid (Days)	Difference in days between the PreBid Date and the Publish Date	numeric
PreBid-Closing (Days)	Difference in days between the Closing Date and the PreBid Date	numeric
Closing-Award (Days)	Difference in days between the Award Date and the Closing Date	numeric
Award-NTP (Days)	Difference in days between the Notice to Proceed Date and the Awarded Date	numeric
Quarter	Quarter when the bid was published	numeric

Figure 2.3 Table 3. **Features Retained in the PhilGEPS Dataset**

Summary Statistics

To have an overview of feature selected data, Table 4 provides the descriptive statistics of the numeric features in the dataset. On average, it was noted that an awarded tender would be given a budget of Php3 million, net a potential savings of approximately 13% of this allocation, and last for about 34.3 days. Additionally, it could be derived that it would usually take about 35.7 days for a tender to be awarded to a winning bidder which is well within the prescribed timeline stipulated under the revised implementing rules and regulations of the Philippine Procurement Law [2]-[3]. Coincidentally, erroneous values (e.g. negative values under the "Award-NTP (Days)" which imply that the "Notice to Proceed" document was issued prior to the awarding date) indicate the presence of anomalies that are already reflected under Table 4. Details on these anomalies are discussed further in the succeeding sections.

Statistic	Approved Budget of the Contract (million Php)	Quantity	Item Budget (million Php)	Contract Amount (million Php)	Contract Duration (Days)	Publish-PreBid (Days)	PreBid-Closing (Days)	Closing-Award (Days)	Award-NTP (Days)
mean	3.0	6676.7	2.8	2.6	34.3	1.9	3.1	12.1	18.6
std	31.5	382219.5	22.5	22.0	68.8	3.5	6.0	17.3	23.9
min	0.0	0.0	0.0	0.0	0.0	-264.0	0.0	-426.0	-2184.0
25%	0.1	1.0	0.1	0.1	5.0	0.0	0.0	1.0	4.0
50%	0.3	1.0	0.3	0.3	15.0	0.0	0.0	7.0	13.0
75%	1.0	1.0	1.0	1.0	30.0	1.0	1.0	16.0	27.0
max	4486.8	83,703,690.0	3800.8	3800.8	7200.0	35.0	275.0	731.0	170.0

Figure 2.3 Table 4. **Descriptive Statistics of the PhilGEPS Dataset**

Exploratory Data Analysis

As a primer to machine learning model training and development, it is necessary to first conduct an exploration of the data to have a better grasp of the patterns that could possibly be contained within the dataset. The succeeding sections provide then preliminary insights that are seen to be beneficial in building up a model capable of detecting anomalous tenders.

Awarded Tenders

Combining the data from the year 2020 to the year 2021 revealed that for 18 months about 2 out of 10 tenders have only been awarded to winning bidders. This finding potentially reflects the drastic consequences of mobility restrictions rolled out during the height of the COVID-19 pandemic. As seen in Figure 2A, the majority of the awarded tenders belonged to the "Goods" classification which indicate the spike in the procurement of consumables (e.g. disinfectant, face masks) to mitigate the spread of the COVID-19 pandemic. Surprisingly, it was seen in Figure 2B that most of the awarded tenders were for purchases of materials related to water service connection. Relative to this observation, it could be derived that COVID-19 related tenders were mostly emergency procurements as enacted from the Bayanihan to Heal as One Act (e.g. Republic Act 11469) [4]. It was then concluded from this initial exploration that COVID-19 related tenders would not be entirely captured under the investigated dataset since emergency procurements are not included in the data. Analysis of these emergency procurements, however, is beyond the scope of this project.

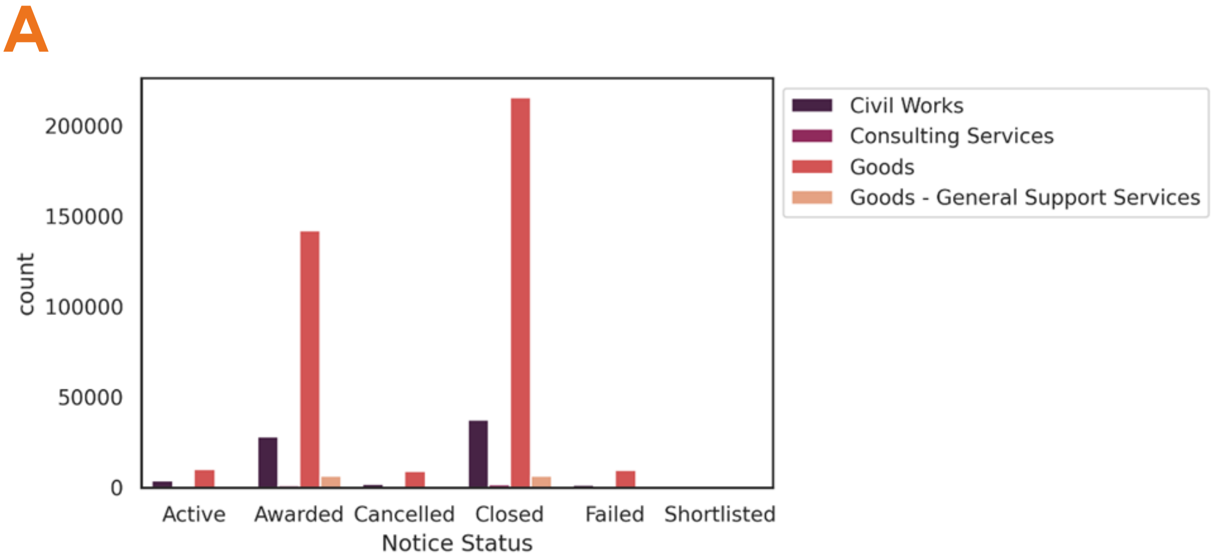


Figure 2.3 - 2 A **Barplot of awarded tenders per Classification**

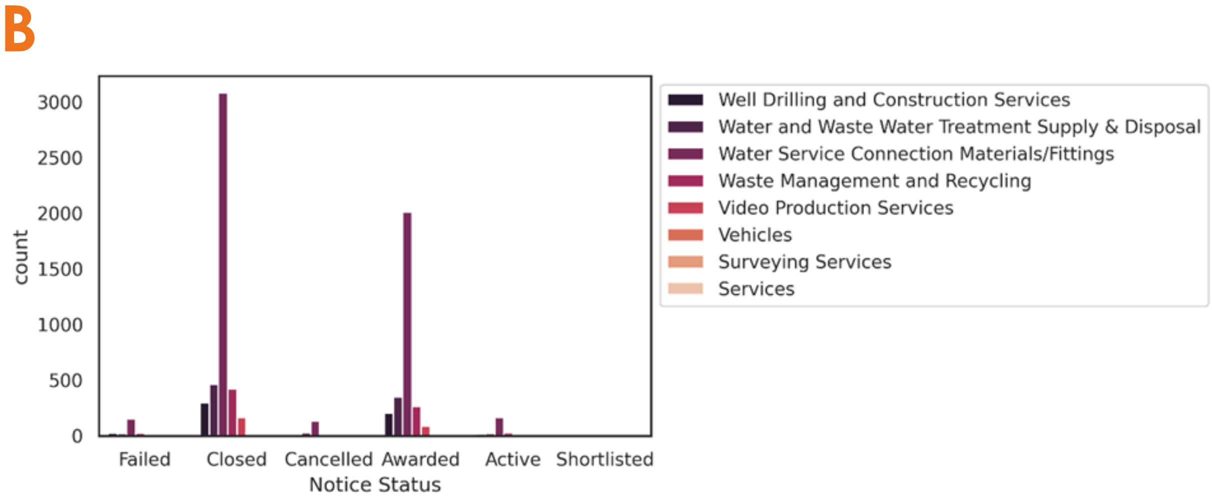


Figure 2.3 - 2 B **Barplot of awarded tenders categorized per (A) "Classification" and (B) "Business Category". About 150,000 of awarded tenders belonged to the "Goods" classification and further delineating this revealed that most of the awarded contracts were procurement of materials for water service connection.**

Empirical Cumulative Distribution Function of Awarded Tenders

To better understand the implications of ranges of the data for the awarded tenders in the PhilGEPS dataset, the empirical cumulative distribution function of a number of features have also been explored. These investigated features are "Approved Budget of the Contract", "Item Budget", "Contract Duration (Days)", and "Award-NTP (Days)" – selected features that provide critical information on an awarded tender. As revealed in Figure 3A and Figure 3B, it was observed that about 0.1% of the awarded tenders would garner approved budgets as large as Php1 Billion. The feature "Contract Duration (Days)", as shown in Figure 3C, was also noted to exhibit a similar pattern to that of the "Approved Budget of the Contract" and "Item Budget" features. Additionally, relative to the observations seen under Table 4, Figure 3D also validates that there is indeed a small portion of awarded tenders that have been issued "Notice to Proceed" documents prior to the issuance of notice of award to winning bidders – a clear violation of the revised implementing rules and regulations of the Philippine Procurement Law [2]. These observations then further substantiate the insight that anomalous tenders would likely be the proportion of records that fall on the extreme ranges of such features.

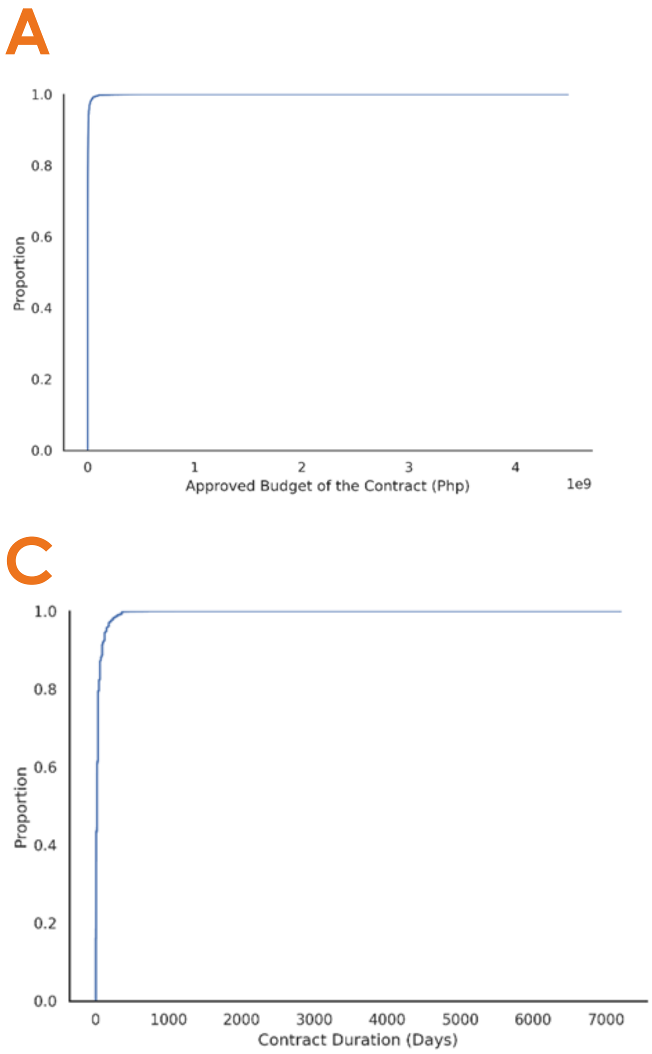


Figure 2.3 - 3. **Empirical cumulative distribution function plots of selected features of the PhilGEPS dataset. Almost 99.9% of the proportion of the tenders have (A) "Approved Budget of the Contract" that would fall between the thousandth and millionth mark. Similar observations were noted in the (B) "Item Budget" and (C) "Contract Duration (Days)" features. In contrast, a very small proportion of records were observed to have negative values for the (D) "Award-NTP (Days)" feature.**

Correlated Features

As the final exploratory analysis implemented in this project, the correlation of the features in the data were also investigated. Using the Pearson correlation coefficient (i.e. Pearson's r) as a measure of correlation, it was found at a threshold of 0.8 that the features "Contract

Amount" and "PreBid-Closing (Days)" had a strong positive correlation to other features in the dataset (see Figure 4 for the visualization of these correlations). Relative to this observation, it was then imperative to exclude these features in the training and development of machine learning models for the detection of anomalous tenders in the data.

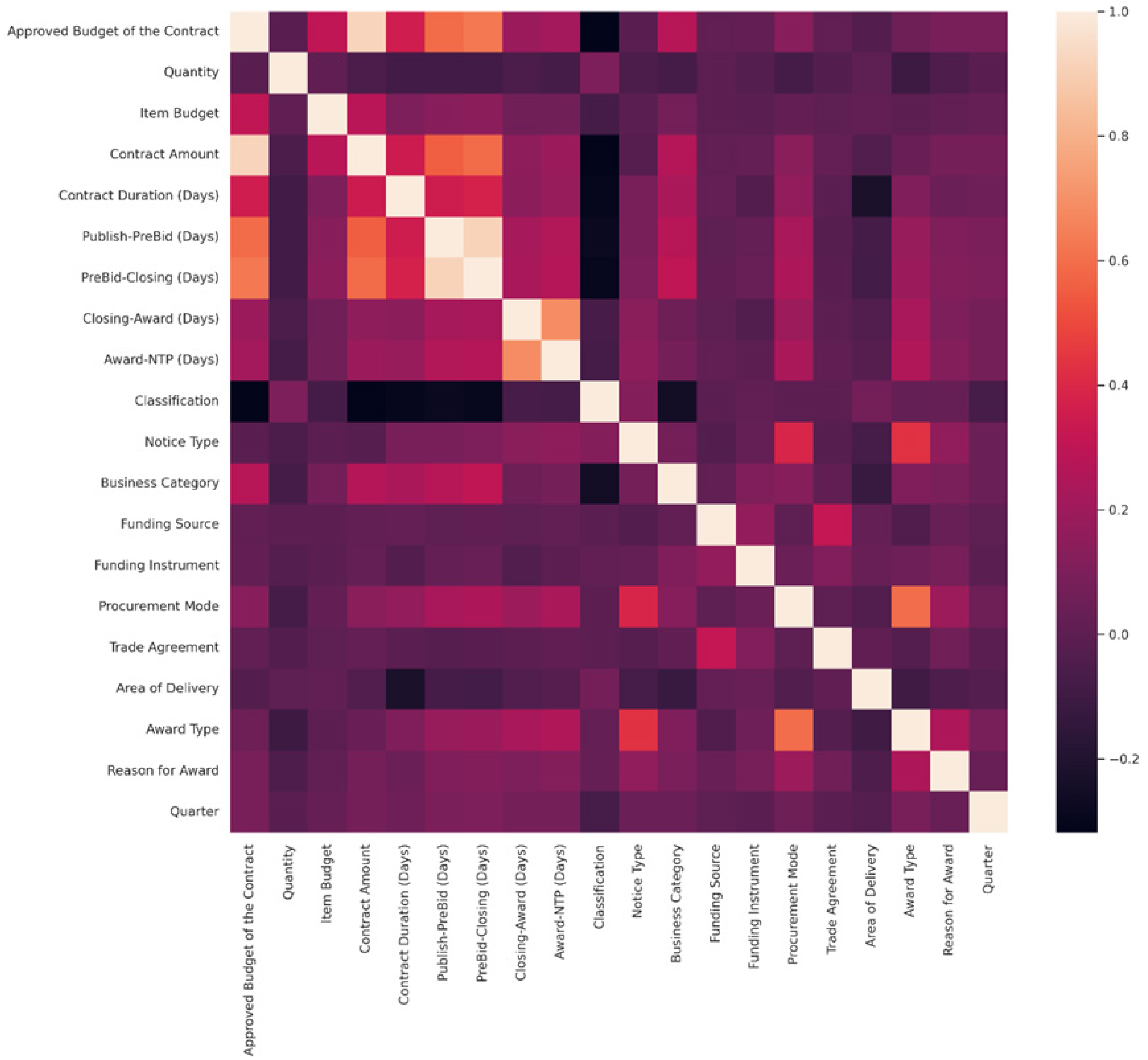


Figure 2.3 - 4. **Pearson correlation coefficient matrix of selected features in the PhilGEPS dataset. At a threshold of 0.8, the features "Contract Amount" and "Pre-Bid-Closing (Days)" were found to be highly correlated to other features in the PhilGEPS dataset.**

Methodology

In order to achieve the overall objective of this work, a pipeline divided into five-phases was adopted to train a machine learning model capable of identifying anomalous tenders. These five-phases are (1) Data Extraction, (2) Data Cleaning and Pre-Processing, (3) Feature Selection, (4) Machine Learning Model Training, and (5) Machine Learning Model Interpretation (see Figure 5). Briefly, procurement data was first extracted from the PhilGEPS website in xls format and then later converted to CSV format for ease of use. Implemented data cleaning was minimal since the extracted data was provided in a structured format. In contrast, implemented data pre-processing was extensive since feature engineering was necessary to deal with the inherent skewness of the data. After cleaning and pre-processing, irrelevant features were identified and then excluded through exploratory data analysis and the use of the Pearson correlation coefficient. Once the cleaned and processed data were found suitable for model training, five outlier detection algorithms (Isolation Forest, Clustering Based Local Outlier Factor, Principal Component Analysis, k-Nearest Neighbors, and Histogram-Based Outlier Detection) were explored for the detection of anomalous tenders [5]. To further enhance the consistency of the anomaly detection results, anomaly scores derived for each tender were standardized through four combination techniques, namely (1) average, (2) maximization, (3) average of maximum, and (4) maximum of average [5]-[6]. Finally, the characteristics that define an anomalous contract were established by interpretation of the features from the developed outlier detection model [5]-[8]. The details on how each phase of this implemented pipeline was conducted are discussed further in the succeeding sections.

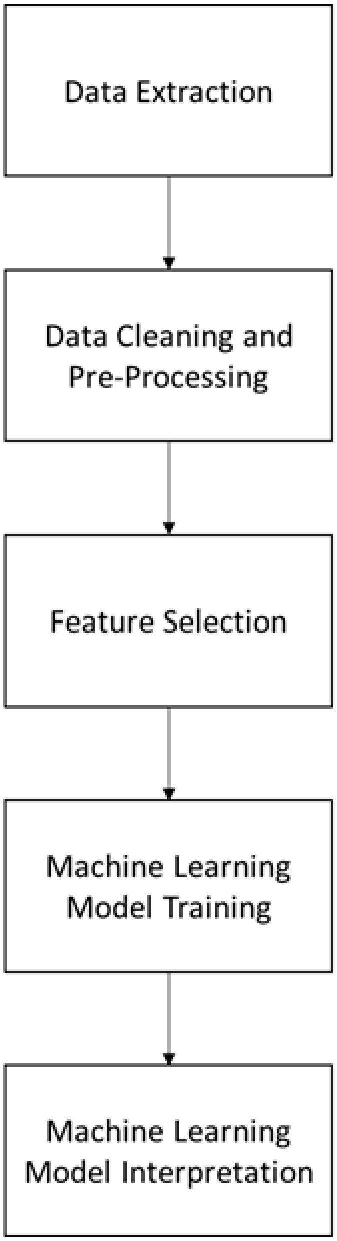

Data Extraction
Data Cleaning and Pre-Processing
Feature Selection
Machine Learning Model Training
Machine Learning Model Interpretation

Figure 2.3 - 5. **A high-level overview of the implemented methodology. This is the summary of the data science pipeline implemented in this project.**

Data Extraction

The PhilGEPS dataset was taken from the PhilGEPS official website [1] and is made available in xls format and uploaded in a shareable google drive link. This dataset is released on a per quarter basis as multiple xls files and hence, necessitated that it be merged into a single dataframe (e.g. stored in CSV format). Only the 2020 and 2021 fiscal years were considered for analysis. Fiscal years below 2020 were not considered since tenders are not reflective of the consequences brought by the COVID-19 pandemic.

Data Cleaning and Pre-Processing

Since records detailing tender information was limited only to reflect tenders that completed the public bidding process (i.e. records with "Award Status" of "Updated"), data cleaning implemented was minimal since the uncleaned data only necessitated the removal of duplicates. Implemented data pre-processing, in contrast, was extensive. Since the data was highly skewed, the data was log transformed to allow for faster convergence of the outlier detection algorithms [9]-[10]. Categorical variables in the dataset were also found to be highly cardinal (i.e. features having unique categories of at least 1,000). One-hot encoding of these highly cardinal features would definitely increase the dimensionality of the dataset multiple-folds, increasing the computing resources needed for model training [11]. To address this issue of increasing dimensionality, count-encoding, which replaces the names of the unique categories with their counts, was used to transform categoric variables into their numeric equivalent. Finally, to allow the features in the data to be comparable with each other, the data was scaled using standard scaling.

Feature Selection

Initial feature selection was conducted through exploratory data analysis. As previously discussed in Section 2.4, irrelevant features pertaining to unique identifiers and as well as features containing text were excluded from the data used for model training. The final step applied to select the final set of features for model training was the use of the Pearson correlation coefficient at a threshold of 0.8 (either positive or negative correlation). Use of the Pearson correlation coefficient revealed that "Contract Amount" and "PreBid-Closing (Days)" features were highly correlated to other features in the data and hence necessitated that these features be removed in the training of the outlier detection algorithms. The final number of features used for model training was thus reduced to 17 features.

Machine Learning Model Training

For this specific use-case, five outlier detection algorithms were explored for the detection of anomalous tenders. These are (1) Isolation Forest (IF), (2) Clustering Based Local Outlier Factor (CBLOS), (3) Principal Component Analysis (PCA), (4) k-Nearest Neighbors (kNN), and Histogram-Based Outlier Detection (HBOS) [5]. To achieve consistency in the results of these applied algorithms, the decision scores returned were standardized using four combination techniques, namely, (1) average, (2) maximization, (3) average of maximum (aom), and (4) maximum of average (moa) [5]-[6]. These combination techniques, however, did not directly return labels for the records identified as anomalies. In this case, the cut-off for deciding whether a record is anomalous or not was based on decision scores that are outside of 3 standard deviations from the mean of the standardized decision scores. Finally, a tender was then classified as anomalous if two or more from these applied combination techniques classified a record as an outlier.

Machine Learning Model Interpretation

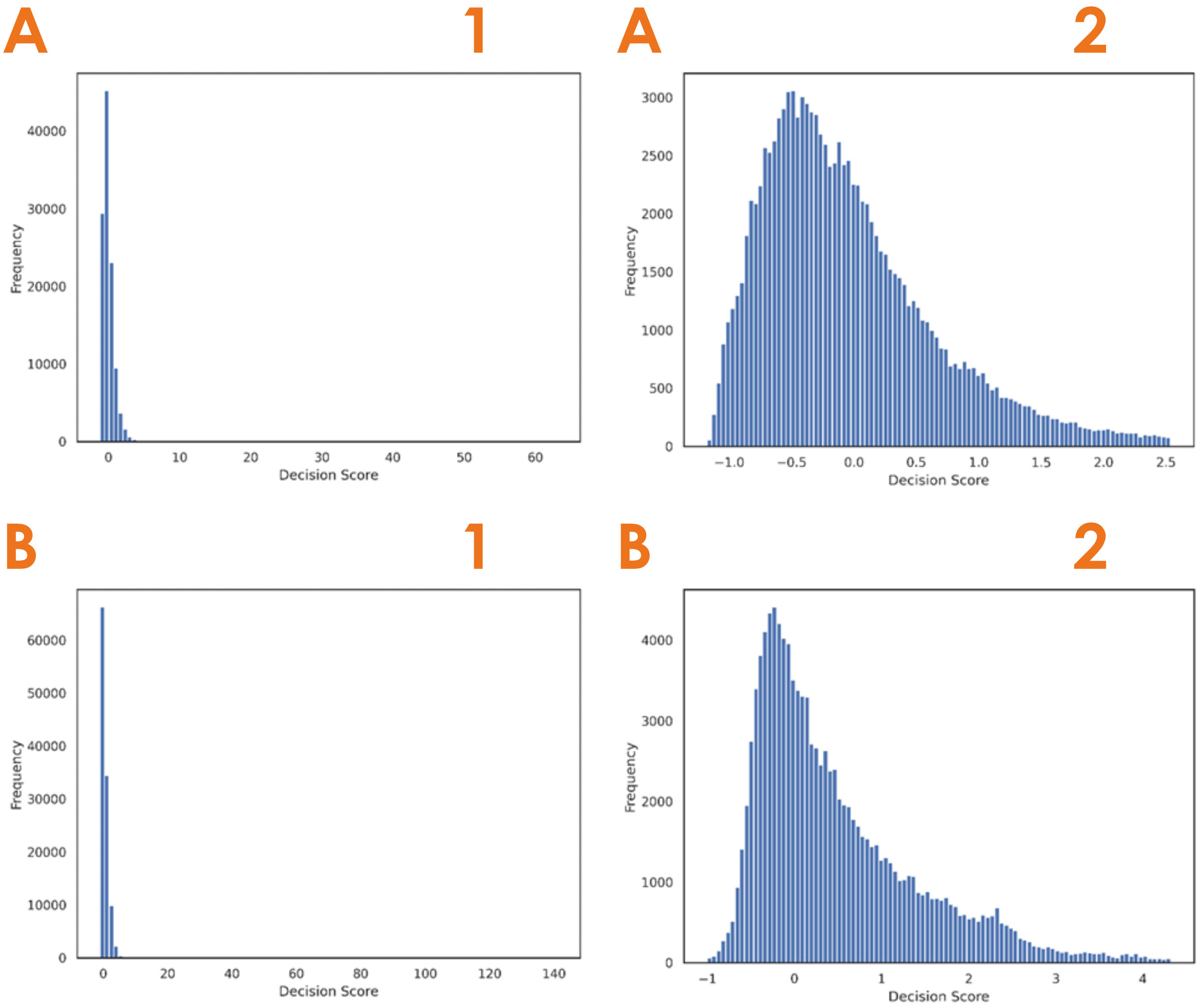
Lastly, in accordance to the features of the dataset used, the way the trained model classifies a record either an anomaly or a normal tender was interpreted. To derive an interpretation out of these features, feature importance, which would estimate the likely impact of a feature on why it is being classified as anomalous, was derived in accordance to reviewed literature [5]-[8].

Results and Findings

This section provides an in-depth discussion on the crucial insights derived from the results of the implemented data science pipeline in this work. By developing a machine learning model capable of detecting anomalous tender, the characteristics of a suspicious contract in terms of the features in the PhilGEPS dataset was then established.

Distribution of Decision Scores

Visually, it could be observed that the decision scores for each record plotted as histograms confirmed the presence of anomalies in the data. As revealed in Figure 6, it was noted that the distributions of the decision scores across all combination techniques applied are highly skewed to the right. This is to be expected since as revealed previously in Figure 3 and Table 4, the maximum value for each of the features in the data is excessively high when compared to the majority of the proportion of records. In this case, the implemented detection algorithms would flag records further to the right of the mean as outliers or anomalies. In an attempt to balance this, the cut-off for outliers was determined to be outside of 3 standard deviations (3σ) from the mean. Collectively, by majority voting (i.e. if two or more combination techniques flag a record as anomalies), about 1672 tenders (1.5% of the total number of tenders) amounting to Php68.9 billion were found to be potentially anomalous.



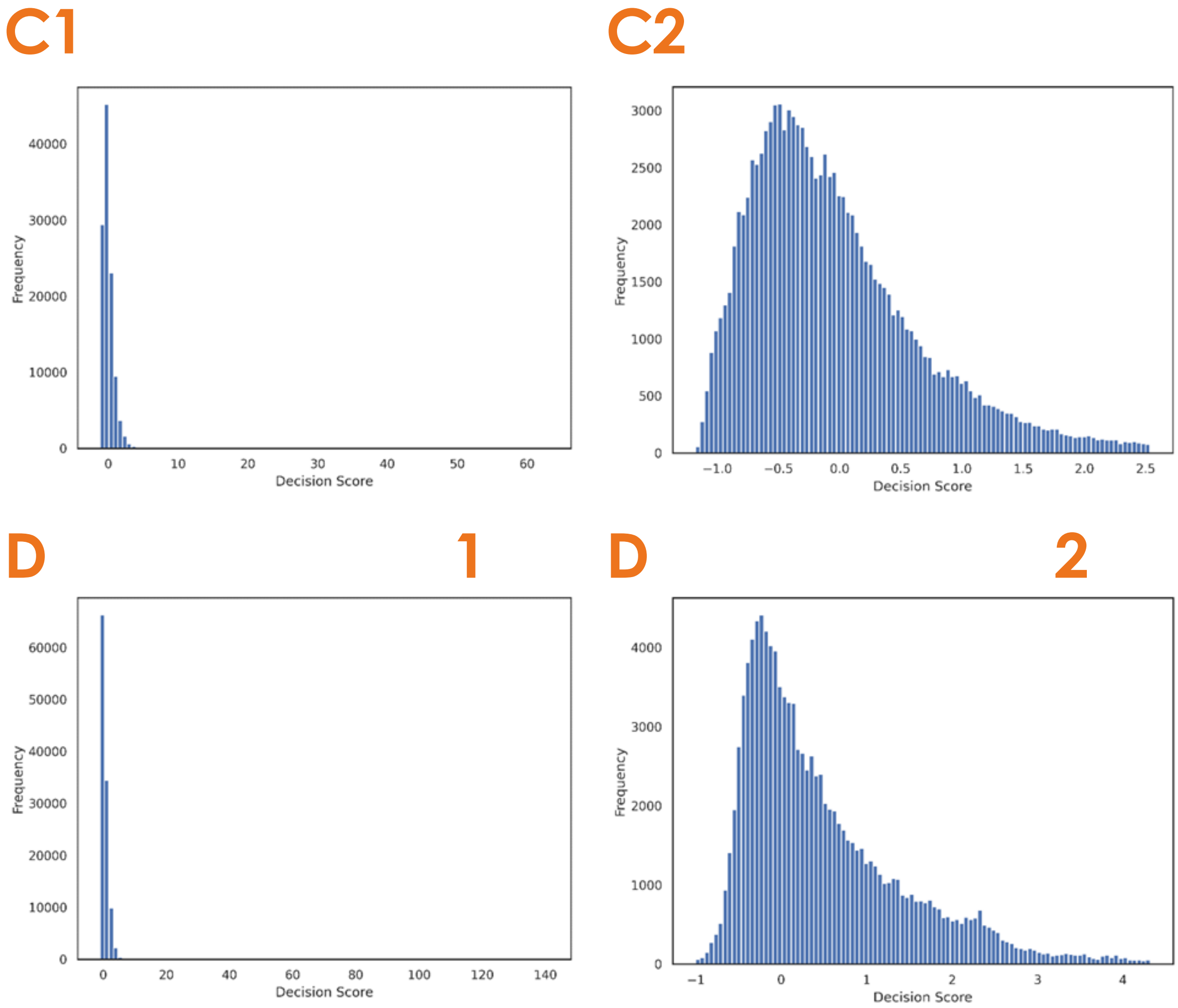


Figure 2.3 - 6. **Histogram plot of decision scores for applied outlier combination techniques. The histogram plots inclusive of the outliers (i.e. anomalous tenders) are denoted by (1) while the ones without are denoted by (2). It was observed that (A) average and (C) average of maximum techniques resulted to the similar distribution of decision scores while (B) maximization and (D) maximum of average were another pair with similar decision score distribution.**

Interpretation of Developed Model

Plotting the anomalies in a pairplot, as shown in Figure 7, revealed that anomalies in a number of combination of features may not be that obvious when compared to other pairings. For instance, the pairing "Area of Delivery" and "Business Category" would not be good candidates for directly identifying anomalous tenders since the anomalies are substantially mixed with normal tenders. The same could also be said for the pairing "Area of Delivery" and "Contract Duration (Days)". In contrast, in the pairing "Item Budget" and "Award-NTP (Days)", the separation between the normal and anomalous tenders are very much obvious, with the normal tenders centering near the origin. The crucial insight to be derived out of these observations is the fact that there are features in the dataset that would strongly push the classification of a record as an anomaly and likewise there are also features in the dataset that would strongly classify a record as a normal tender. Collectively, the results shown in Figure 7 imply that the features used in the PhilGEPS dataset could be amply used to separate normal tenders from abnormal ones.

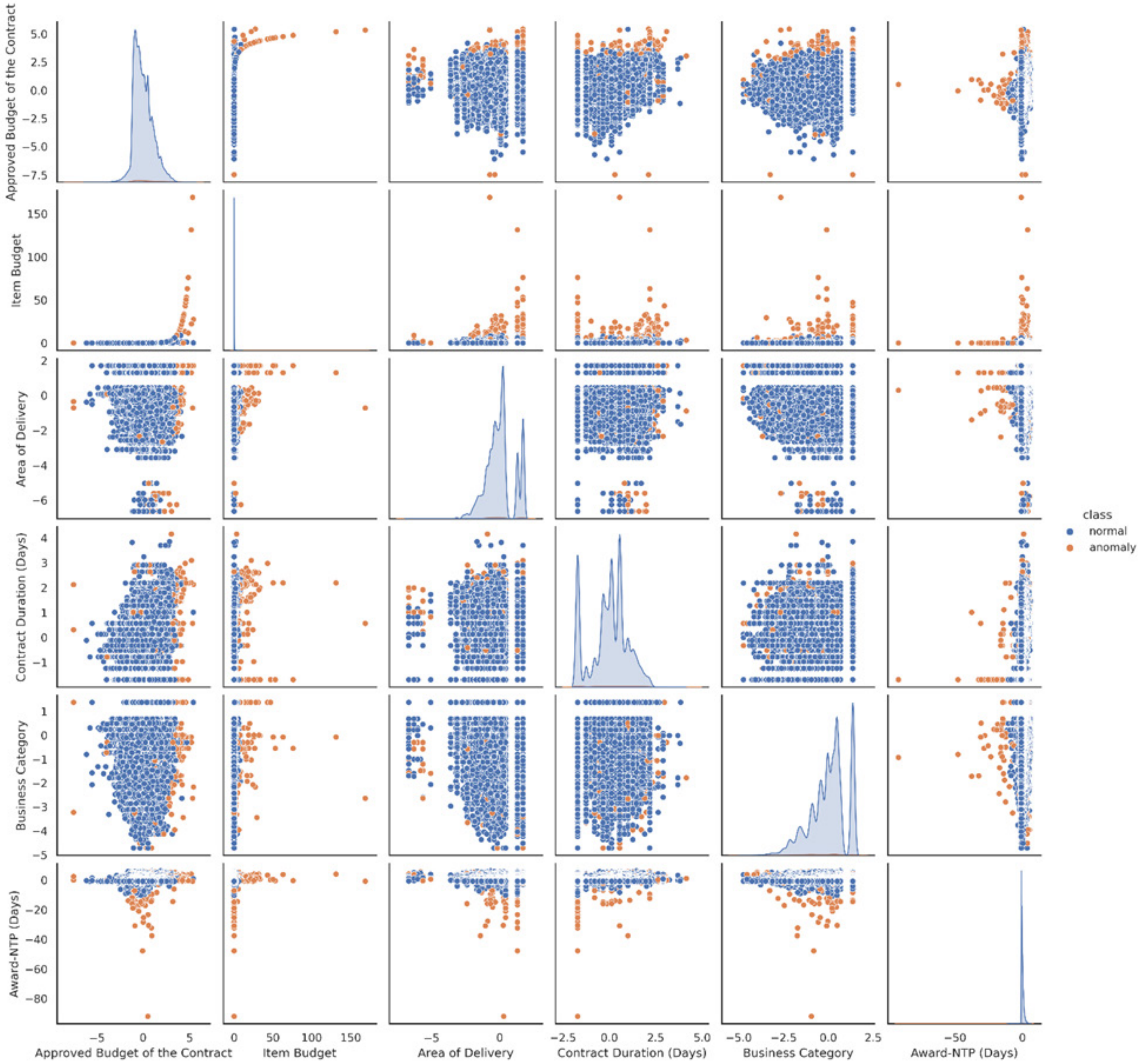


Figure 2.3 - 7. **Pairplot of features with scaled values. Visualizing the identified anomalies on a pairplot revealed that such anomalies could be easily spotted if these were to be plotted on a "Item Budget" vs "Award-NTP (Days)" plot.**

Building up on the previous discussion, Figure 8 enumerates the features in the PhilGEPS dataset that are seen by the developed model to strongly push for the classification of tenders to be anomalous. The top three (3) identified features are "Item Budget", "Approved Budget of the Contract", and "Contract Duration (Days)". However, as seen in Figure 7, interpreting these features individually may not generate much insight since in some feature combinations, anomalous tenders are substantially mixed with normal tenders which could likely result into misclassification. In this sense, caution was further exercised in extracting insights from the anomalous tenders.

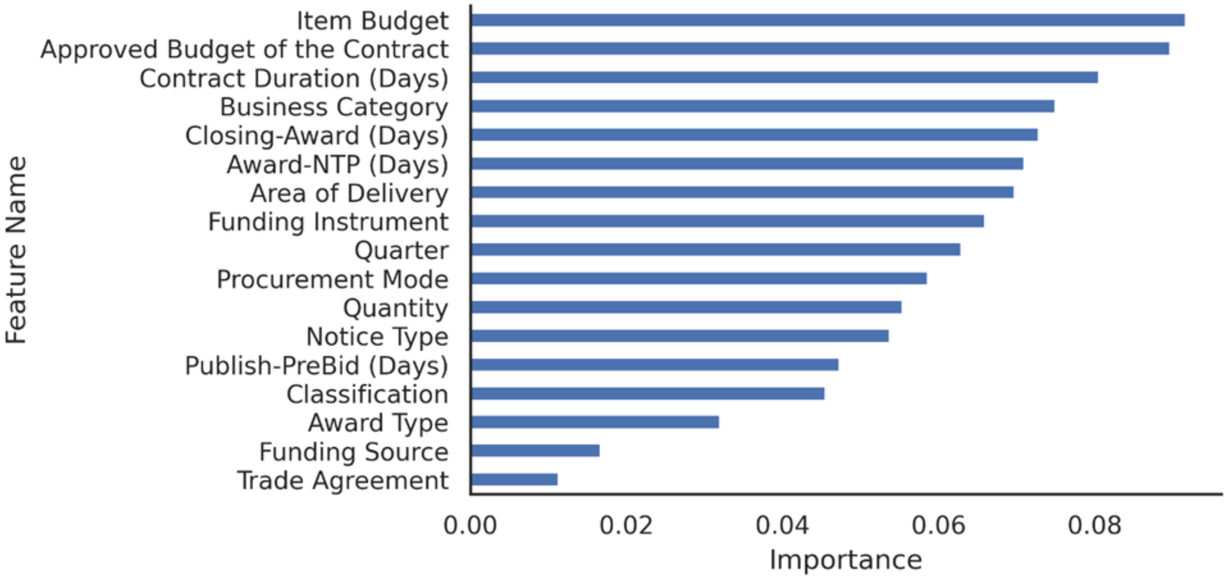


Figure 2.3 - 8. **Feature importance plot of the developed outlier detection model. The feature "Item Budget" was identified as the top feature that pushes a record to be identified as anomalous.**

Clustering of Anomalous Tenders

To exercise more caution in the interpretation of the characteristics of an anomalous tender, the identified anomalous contracts were clustered using hierarchical clustering via Ward's method [13]. As shown in Figure 9, it was observed that the anomalies identified could be categorized further into two big groups namely:

- Hard-to-Recognize Anomalous Contracts (Cluster 1): These are potentially anomalous tenders that would necessitate the use of machine learning for these tenders to be identified. As shown in Figure 10A and Figure 10B, there is no obvious difference in the boxplots of the features of this cluster when compared to the boxplots of the features of normal tenders.
- Easy-to-Recognize Anomalous Contracts (Cluster 2): These are potentially anomalous tenders that would not necessitate the use of machine learning for these tenders to be identified. As shown in Figure 10A and Figure 10B, there is an obvious difference in the boxplots of the features of this cluster when compared to the boxplots of the features of normal tenders.

Indeed, these results confirm that it is very likely that either normal or cluster 1 anomalous contracts could be interchangeably classified by the trained model. However, in the absence of ground truth labels, the rate at which the model may misclassify tenders cannot be quantified and is an area of future work being considered as an off-shoot to this project.

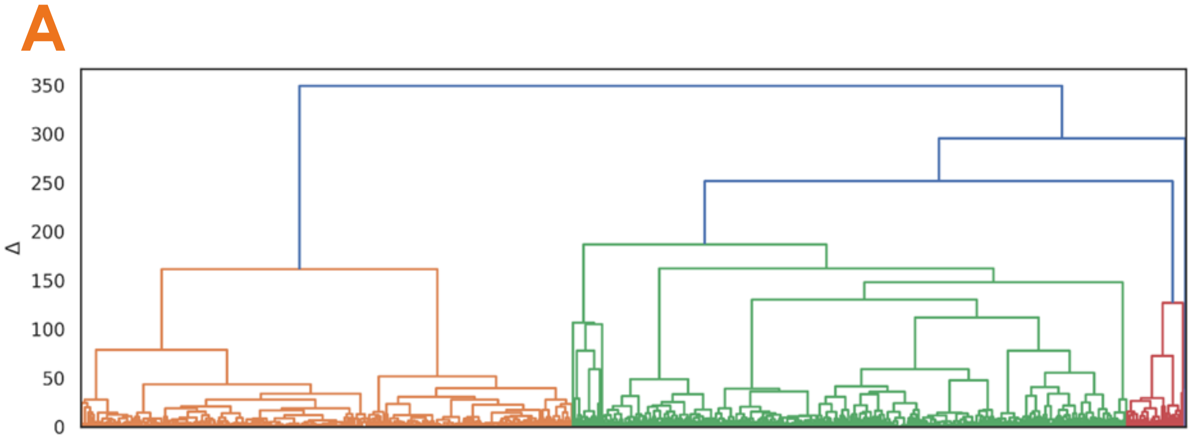


Figure 2.3 - 9 A. **Hierarchical clustering of anomalous tenders through Ward's method. At a delta of 320 (A), two clusters within the anomalous tenders were identified (B).**

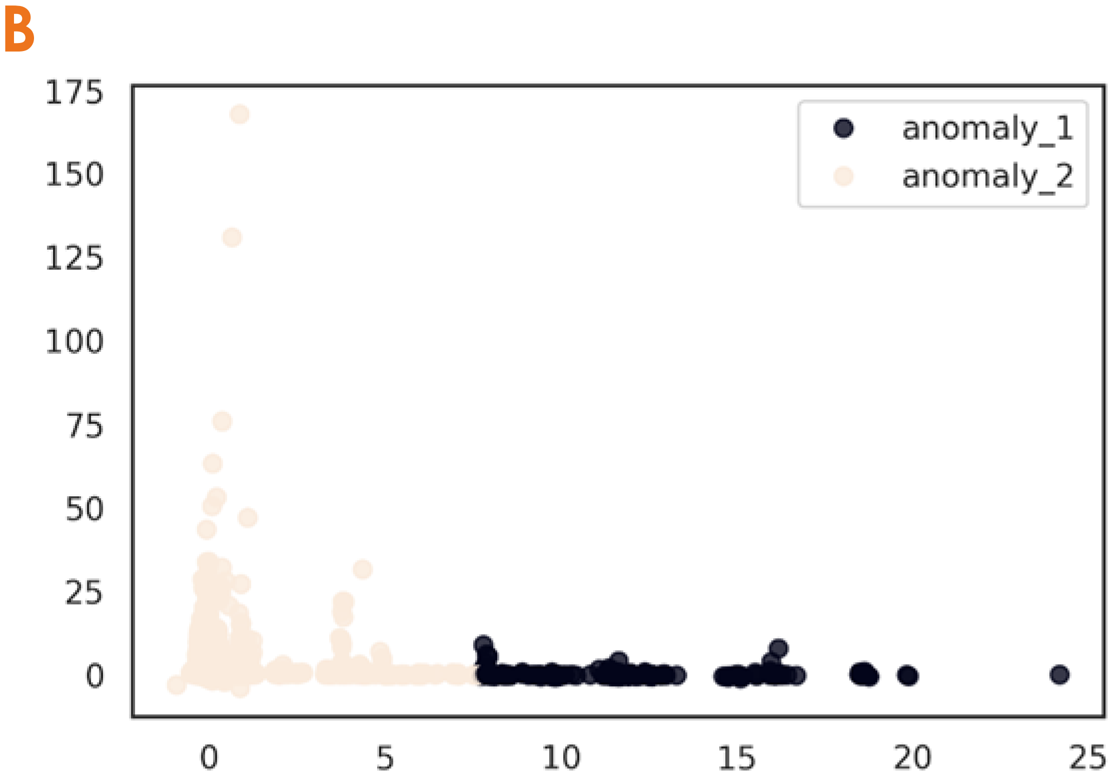
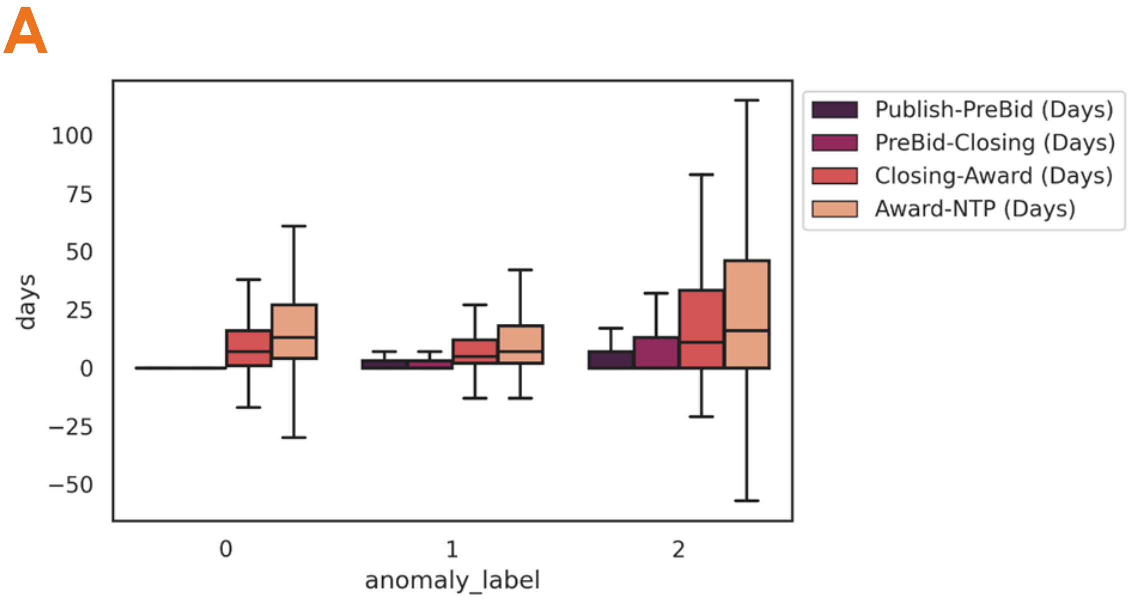


Figure 2.3 - 9 B. **Hierarchical clustering of anomalous tenders through Ward's method. At a delta of 320 (A), two clusters within the anomalous tenders were identified (B).**



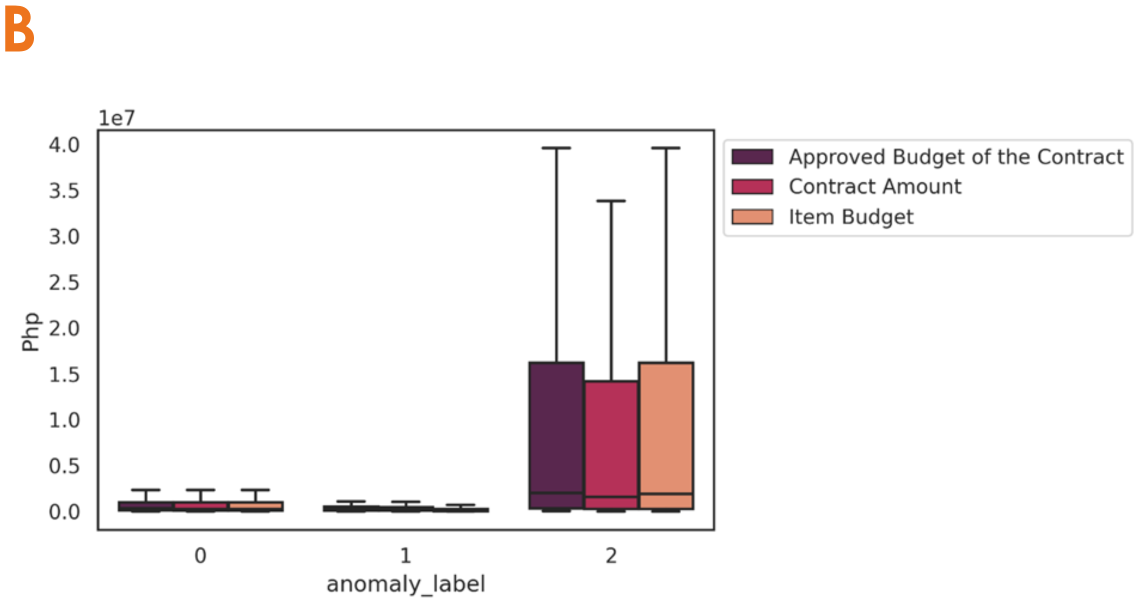


Figure 2.3 - 10. **Boxplot of clustered anomalies (plotted without outliers for cleaner visualization). It was identified that cluster 2 (anomaly_label "2") is the most erroneous and easily identifiable anomalous tenders among the identified outliers while cluster 1 (anomaly_label "1") are relatively difficult to recognize since these have characteristics that are similar to that of normal tenders (anomaly_label "0").**

Proposed Model Deployment Plan for Pilot Testing

In the expectation of strong resistance towards the integration of analytics in the detection of anomalous tenders, this project proposes to initially target Commission on Audit (COA) auditors as end-users of the developed trained anomaly detection model. The primary motivation for this initial move is the fact that COA auditors would be more open to such changes given the reality that it is the job of an auditor to post-audit completed tenders. For instance, the automation of the identification of tenders to be prioritized for post-audit activities is seen to positively aid COA auditors in recommending legal actions against the winning bidders since such recommendations could now be formulated earlier than usual (about 1 year and is usually coincidental with the release of annual audit reports). As shown in Figure 11, the realization of this vision would require the integration of a model inferencing server in the PhilGEPS system. The flow on how predictions of the trained model would be made available to COA auditors is as follows:

· Step 1: Relevant details of awarded tender (e.g. "Notice to Proceed Date") is uploaded in PhilGEPS.

· Step 2: The model inferencing server hosted in the cloud would extract this uploaded data from PhilGEPS on a set frequency.

· Step 3: The extracted data would then be pre-processed in accordance with features necessitated by the trained model.

· Step 4: Decision scores for each uploaded tender will then be calculated in the model prediction layer.

· Step 5: From a pre-defined threshold, each uploaded record will either be classified as normal or anomalous tender.

· Step 6: Details of the tenders classified as anomalous will be returned as an entry in a smart sheet (see Figure 12 for an example) that will be made accessible either in the cloud (or locally if necessary)

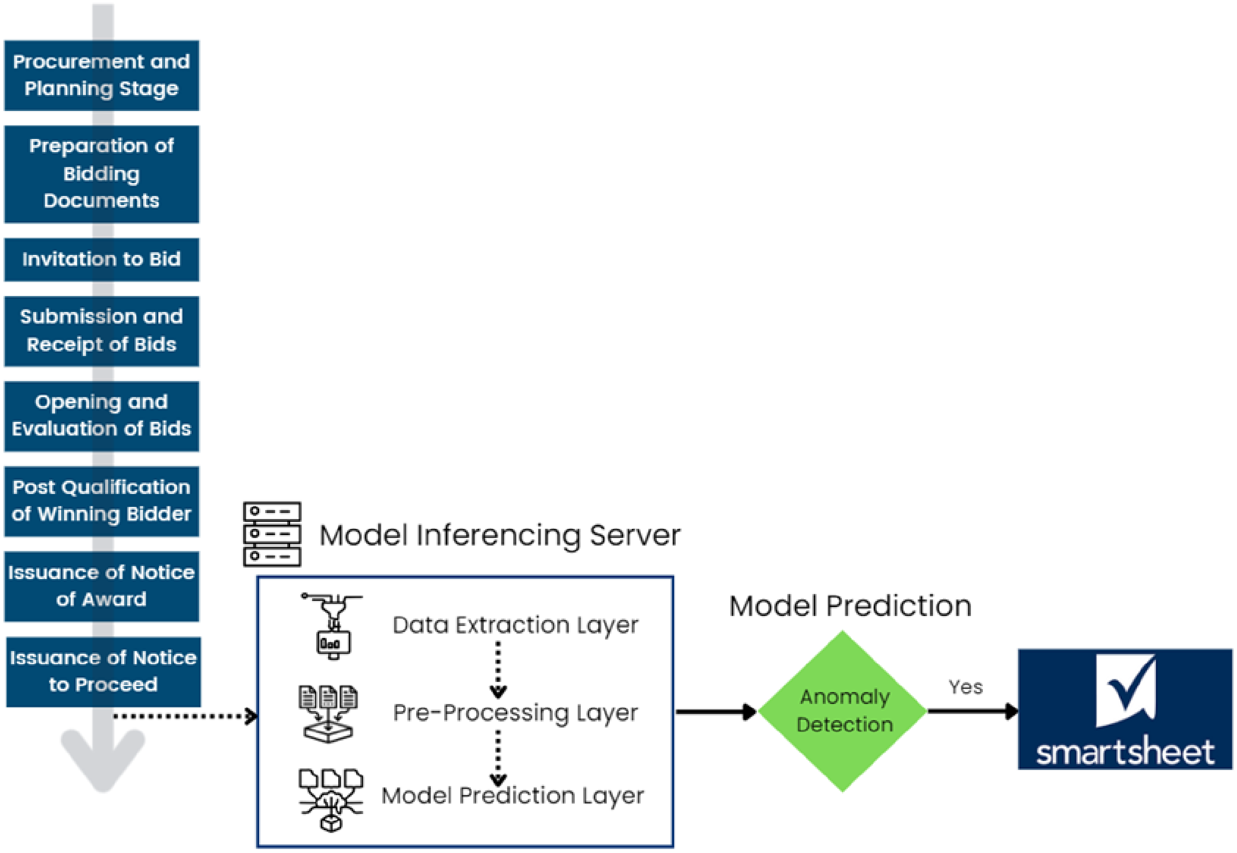
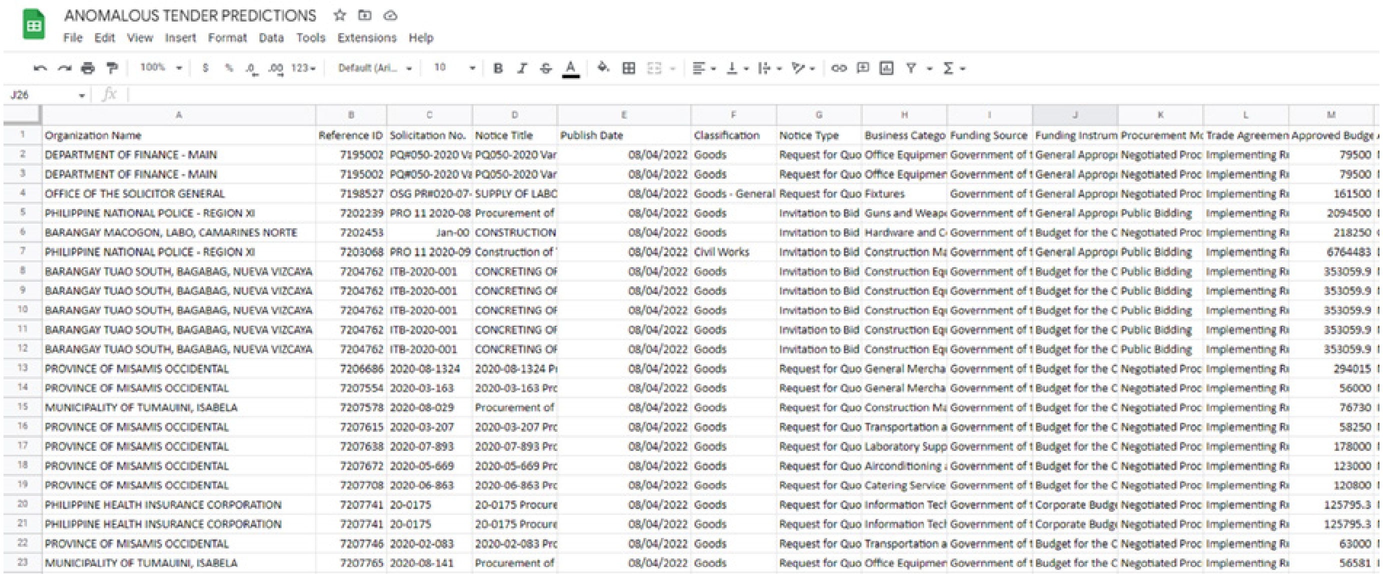

Procurement and Planning Stage
Preparation of Bidding Documents
Invitation to Bid
Submission and Receipt of Bids
Opening and Evaluation of Bids
Post Qualification of Winning Bidder
Issuance of Notice of Award
Issuance of Notice to Proceed
Model Inferencing Server
Data Extraction Layer
Pre-Processing Layer
Model Prediction Layer
Model Prediction
Anomaly Detection
Yes
smartsheet

Figure 2.3 - 11. **Proposed deployment plan for pilot testing of anomaly detection model. With Commission on Audit (COA) auditors as the initial target users of the trained detection model, a model inferencing server is proposed to be deployed right after the issuance of the "Notice to Proceed" phase of the public bidding process.**


ANOMALOUS TENDER PREDICTIONS
File Edit View Insert Format Data Tools Extensions Help
100%
Default (Ari...
10
J26

	Organization Name	Reference ID	Solicitation No.	Notice Title	Publish Date	Classification	Notice Type	Business Catego	Funding Source	Funding Instrum	Procurement Mc	Trade Agreemen	Approved Budge
2	DEPARTMENT OF FINANCE - MAIN	7195002	PQ#050-2020 Va	PQ050-2020 Var	08/04/2022	Goods	Request for Quo	Office Equipmen	Government of t	General Appropi	Negotiated Proc	Implementing Ru	79500
3	DEPARTMENT OF FINANCE - MAIN	7195002	PQ#050-2020 Va	PQ050-2020 Var	08/04/2022	Goods	Request for Quo	Office Equipmen	Government of t	General Appropi	Negotiated Proc	Implementing Ru	79500
4	OFFICE OF THE SOLICITOR GENERAL	7198527	OSG PR#020-07-	SUPPLY OF LABC	08/04/2022	Goods - General	Request for Quo	Fixtures	Government of t	General Appropi	Negotiated Proc	Implementing Ru	161500
5	PHILIPPINE NATIONAL POLICE - REGION XI	7202239	PRO 11 2020-08	Procurement of	08/04/2022	Goods	Invitation to Bid	Guns and Weap	Government of t	General Appropi	Public Bidding	Implementing Ru	2094500
6	BARANGAY MACOGON, LABO, CAMARINES NORTE	7202453	Jan-00	CONSTRUCTION	08/04/2022	Goods	Invitation to Bid	Hardware and C	Government of t	Budget for the C	Negotiated Proc	Implementing Ru	218250
7	PHILIPPINE NATIONAL POLICE - REGION XI	7203068	PRO 11 2020-09	Construction of	08/04/2022	Civil Works	Invitation to Bid	Construction Ma	Government of t	General Appropi	Public Bidding	Implementing Ru	6764483
8	BARANGAY TUAO SOUTH, BAGABAG, NUEVA VIZCAYA	7204762	ITB-2020-001	CONCRETING OF	08/04/2022	Goods	Invitation to Bid	Construction Eq	Government of t	Budget for the C	Public Bidding	Implementing Ru	353059.9
9	BARANGAY TUAO SOUTH, BAGABAG, NUEVA VIZCAYA	7204762	ITB-2020-001	CONCRETING OF	08/04/2022	Goods	Invitation to Bid	Construction Eq	Government of t	Budget for the C	Public Bidding	Implementing Ru	353059.9
10	BARANGAY TUAO SOUTH, BAGABAG, NUEVA VIZCAYA	7204762	ITB-2020-001	CONCRETING OF	08/04/2022	Goods	Invitation to Bid	Construction Eq	Government of t	Budget for the C	Public Bidding	Implementing Ru	353059.9
11	BARANGAY TUAO SOUTH, BAGABAG, NUEVA VIZCAYA	7204762	ITB-2020-001	CONCRETING OF	08/04/2022	Goods	Invitation to Bid	Construction Eq	Government of t	Budget for the C	Public Bidding	Implementing Ru	353059.9
12	BARANGAY TUAO SOUTH, BAGABAG, NUEVA VIZCAYA	7204762	ITB-2020-001	CONCRETING OF	08/04/2022	Goods	Invitation to Bid	Construction Eq	Government of t	Budget for the C	Public Bidding	Implementing Ru	353059.9
13	PROVINCE OF MISAMIS OCCIDENTAL	7206686	2020-08-1324	2020-08-1324 Pr	08/04/2022	Goods	Request for Quo	General Mercha	Government of t	Budget for the C	Negotiated Proc	Implementing Ru	294015
14	PROVINCE OF MISAMIS OCCIDENTAL	7207554	2020-03-163	2020-03-163 Prc	08/04/2022	Goods	Request for Quo	General Mercha	Government of t	Budget for the C	Negotiated Proc	Implementing Ru	56000
15	MUNICIPALITY OF TUMAUINI, ISABELA	7207578	2020-08-029	Procurement of	08/04/2022	Goods	Request for Quo	Construction Ma	Government of t	Budget for the C	Negotiated Proc	Implementing Ru	76730
16	PROVINCE OF MISAMIS OCCIDENTAL	7207615	2020-03-207	2020-03-207 Prc	08/04/2022	Goods	Request for Quo	Transportation a	Government of t	Budget for the C	Negotiated Proc	Implementing Ru	58250
17	PROVINCE OF MISAMIS OCCIDENTAL	7207638	2020-07-893	2020-07-893 Prc	08/04/2022	Goods	Request for Quo	Laboratory Supp	Government of t	Budget for the C	Negotiated Proc	Implementing Ru	178000
18	PROVINCE OF MISAMIS OCCIDENTAL	7207672	2020-05-669	2020-05-669 Prc	08/04/2022	Goods	Request for Quo	Airconditioning a	Government of t	Budget for the C	Negotiated Proc	Implementing Ru	123000
19	PROVINCE OF MISAMIS OCCIDENTAL	7207708	2020-06-863	2020-06-863 Prc	08/04/2022	Goods	Request for Quo	Catering Service	Government of t	Budget for the C	Negotiated Proc	Implementing Ru	120800
20	PHILIPPINE HEALTH INSURANCE CORPORATION	7207741	20-0175	20-0175 Procure	08/04/2022	Goods	Request for Quo	Information Tech	Government of t	Corporate Budge	Negotiated Proc	Implementing Ru	125795.3
21	PHILIPPINE HEALTH INSURANCE CORPORATION	7207741	20-0175	20-0175 Procure	08/04/2022	Goods	Request for Quo	Information Tech	Government of t	Corporate Budge	Negotiated Proc	Implementing Ru	125795.3
22	PROVINCE OF MISAMIS OCCIDENTAL	7207746	2020-02-083	2020-02-083 Prc	08/04/2022	Goods	Request for Quo	Transportation a	Government of t	Budget for the C	Negotiated Proc	Implementing Ru	63000
23	MUNICIPALITY OF TUMAUINI, ISABELA	7207765	2020-08-141	Procurement of	08/04/2022	Goods	Request for Quo	Office Equipmen	Government of t	Budget for the C	Negotiated Proc	Implementing Ru	56581

Figure 2.3 - 12. **Sample output of smartsheet. Entries outputted in the smartsheet are only those tenders that have been classified as anomalous. In consideration of storage space, tenders predicted as normal will not be returned as an entry in the smartsheet.**

Preferred Deployment Location of Trained Anomaly Detection Model

The model deployment plan proposed in Figure 11, however, is not ideal if the objective now would be to prevent the awarding of compromised tenders. This is due to the fact that the money disbursed to a winning bidder under an anomalous tender would either be ill-gotten money or lost opportunity for the betterment of the Filipino people. With this in mind, it is preferred that the anomaly detection model be deployed between the "Invitation to Bid" and "Submission and Receipt of Bids" phases of the public bidding process (see Figure 13). This is the ideal deployment location since this is where a procuring government agency decides whom to award the posted tender. The primary challenge, however, with this framework is the fact that data between these two phases is not readily available since this stage of the public bidding process is treated with utmost confidentiality to minimize bias and unwanted intervention from either the procuring entity or the bidders. Hence, it may be nearly impossible to train an anomaly detection model at this stage of the public bidding process in the absence of data.

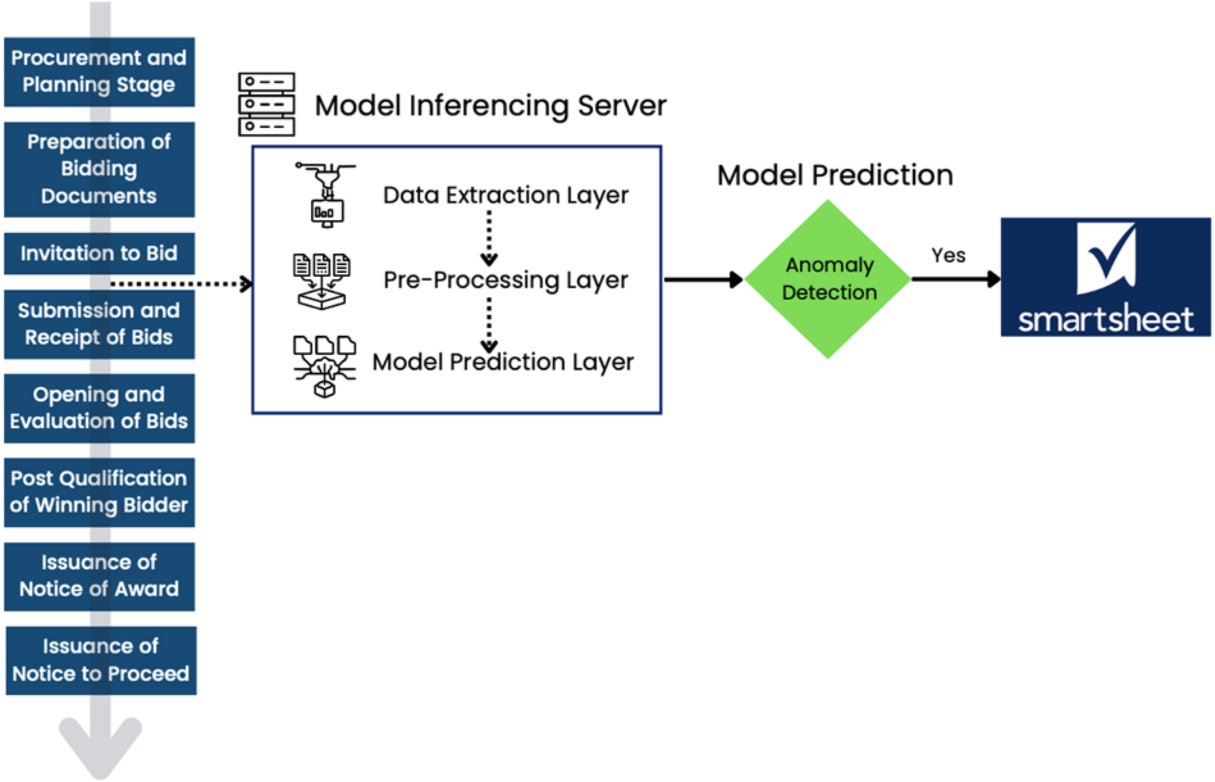

Procurement and Planning Stage
Preparation of Bidding Documents
Invitation to Bid
Submission and Receipt of Bids
Opening and Evaluation of Bids
Post Qualification of Winning Bidder
Issuance of Notice of Award
Issuance of Notice to Proceed
Model Inferencing Server
Data Extraction Layer
Pre-Processing Layer
Model Prediction Layer
Model Prediction
Anomaly Detection
Yes
smartsheet

Figure 2.3 - 13. **Ideal deployment location of trained anomaly detection model. Deployment of the trained model between the "Invitation to Bid" and "Submission and Receipt of Bids" is seen to detect early the occurrence of anomalies in the public bidding process.**

Key Takeaways

This project has successfully achieved its overall objective. In developing a machine learning model capable of identifying anomalous tenders, this project was also able to showcase the power of leveraging data to safeguard the interest of the general public, specifically in the public procurement process. In conclusion, the findings of this project could be summarized into three substantial insights:

- Salient Insight 1: As derived from the anomalies identified by the developed model, tenders having at least Php2.5 million of the approved budget of the contract should be meticulously scrutinized. This essentially means that a phase-by-phase review of the public bidding process (e.g. scrutiny of PreBid phase, scrutiny of Award Phase) that these tenders went through should be mandated and possibly be included as part of the rules and regulations governing such procurements.

· Salient Insight 2: As derived from the anomalies identified by the developed model, tenders having a difference between the "Notice to Proceed Date" and "Award Date" that is less than 3 calendar days (this is the minimum total number of days mandated for the release of the "Notice to Proceed" document) or a difference between the "Notice to Proceed Date" and "Award Date" that is greater than 47 calendar days (this is the maximum total number of days mandated for the release of the "Notice to Proceed" document) could already be flagged in the PhilGEPS system as "anomalous". These are clear violations of the revised implementing rules and regulations under the Philippine Procurement Law [2].

· Salient Insight 3: Similar to that of the previous insight, negative differences in the time-stamps of tenders for the "Publish-PreBid (Days)", "PreBid-Closing (Days)", and "Closing-Award (Days)" features could also be flagged "anomalous" in the PhilGEPS system since the public bidding process should strictly follow the mandated sequence of phases in the bidding [2].

Tools Used

Python was the primary tool used to realize the pipeline implemented in this work. Cloud computing services such as GoogleColab was utilized as the virtual machines for the initial training of the machine learning models. Finally, the final model training was conducted on a rented machine that has 4-cores, 16 GB of RAM and 8 GB of GPU.

Acknowledgement

The authors of this work would like to acknowledge Layertech Software Labs, Inc. for the funding given to this project. Moreover, the authors would also like to express their utmost gratitude to Ms. Frei Sangil and the whole of the Layertech Team for the commendable technical and administrative support given to the authors.

2.3.4 References

[1] "PhilGEPS", Notices.philgeps.gov.ph, 2022. [Online]. Available: https://notices.philgeps.gov.ph/opendataSRD.html. [Accessed: 03- Aug- 2022].

[2] Gppb.gov.ph, 2022. [Online]. Available: https://www.gppb.gov.ph/assets/pdfs/Updated%202016%20IRR_31%20March%202021.pdf. [Accessed: 03- Aug- 2022].

[3] Ocdex.tech, 2022. [Online]. Available: https://www.ocdex.tech/wp-content/uploads/2020/09/Procurement_Analytics_DIGITAL_v1.1-3.pdf. [Accessed: 03- Aug- 2022].

[4] Congress of the Philippines, "Republic Act No. 11469", Metro Manila, 2020.

[5] Y. Zhao, Z. Nasrullah, and Z. Li, "PyOD: A Python Toolbox for Scalable Outlier Detection," arXiv, 2019, doi: 10.48550/ARXIV.1901.01588.

[6] Y. Zhao, X. Wang, C. Cheng, and X. Ding, "Combining Machine Learning Models using combo Library," arXiv, 2019, doi: 10.48550/ARXIV.1910.07988.

[7] F. T. Liu, K. M. Ting, and Z.-H. Zhou, "Isolation Forest," 2008 Eighth IEEE International Conference on Data Mining. IEEE, Dec. 2008. doi: 10.1109/icdm.2008.17.

[8] M. Carletti, C. Masiero, A. Beghi, and G. A. Susto, "Explainable Machine Learning in Industry 4.0: Evaluating Feature Importance in Anomaly Detection to Enable Root Cause Analysis," 2019 IEEE International Conference on Systems, Man and Cybernetics (SMC). IEEE, Oct. 2019. doi: 10.1109/smc.2019.8913901.

[9] Y. Kim, I. Ohn, and D. Kim, "Fast convergence rates of deep neural networks for classification," Neural Networks, vol. 138. Elsevier BV, pp. 179–197, Jun. 2021. doi: 10.1016/j.neunet.2021.02.012.

[10] X. Wan, "Influence of feature scaling on convergence of gradient iterative algorithm," Journal of Physics: Conference Series, vol. 1213, no. 3. IOP Publishing, p. 032021, Jun. 01, 2019. doi: 10.1088/1742-6596/1213/3/032021.

[11] P. Cerda and G. Varoquaux, "Encoding high-cardinality string categorical variables," arXiv, 2019, doi: 10.48550/ARXIV.1907.01860.

[12] W. D McGinnis, C. Siu, A. S, and H. Huang, "Category Encoders: a scikit-learn-contrib package of transformers for encoding categorical data," The Journal of Open Source Software, vol. 3, no. 21. The Open Journal, p. 501, Jan. 22, 2018. doi: 10.21105/joss.00501.

[13] A. Großwendt, H. Röglin, and M. Schmidt, "Analysis of Ward's Method." arXiv, 2019. doi: 10.48550/ARXIV.1907.05094.

SHEHAB D. IBRAHIM

RABBY Q. LAVILLES

2.4
USE CASE 4
Water Distribution Analysis in Iligan City

2.4.1 About the authors

Shehab D. Ibrahim

Shehab D. Ibrahim is a graduate of Master of Science in Data Science at the Asian Institute of Technology in the year 2018 and was also a graduate in Master of Science in Information Technology at MSU-IIT. His field of expertise is in the field of Data Science and Networking. He joined the Information Technology Department as Assistant Professor in the year 2016. He is currently pursuing his Doctor in Information Technology at Cebu Institute of Technology-University.

Rabby Q. Lavilles

Rabby Q. Lavilles is the current Dean of the College of Computer Studies. He has been in the position for 4 years. He is also the Chairman of the University Academic Scholarship Panel.

He finished his Doctorate Degree in Information Technology at De La Salle University – Manila last 2018 and gained his Masters Degree in Information Technology at the same university last 2010. His field of expertise is in Grounded Theory and Information Systems.

Mia Amor C. Tinam-isan

Mia Amor C. Tinam-isan is a graduate of Masters in Information Technology at MSU-Iligan Institute of Technology. She has been in the teaching profession for 14 years; 9 years in MSU-Marawi and 5 years in MSU-IIT. Her publications were focused on social computing and ICT4D.

MIA AMOR C. TINAM-ISAN

Jennifer Joyce M. Montemayor

Jennifer Joyce M. Montemayor received her BS and MS in Computer Science from MSU-Iligan Institute of Technology where she currently works as a faculty for the Department of Computer Science in the College of Computer Studies. Her research interests include Machine Learning and Evolutionary Computation.

JENNIFER JOYCE M. MONTEMAYOR

Sittie Noffaisah B. Pasandalan

Sittie Noffaisah B. Pasandalan finished her MA in English Language Studies and BA in English at MSU-Iligan Institute of Technology. She teaches writing, language, and literature classes at the Department of English and currently holds the post of Assistant Dean of the College of Arts and Social Sciences. Her research interest is on Mindanao culture and language towards peace and development in the region. She is currently pursuing a Doctor of Philosophy in Organizational Development and Planning at SAIDI.

SITTIE NOFFAISAH B. PASANDALAN

2.4.2 Abstract

Water consumption in Iligan City is increasing as a result of population growth and economic development. According to the 2015 Philippine Census of Population and Housing, the city has a population of 342,618 and 44 administrative units. However, the Iligan City Waterworks System (ICWS) only serves 68% of its administrative units. The supplied areas experience periods of water interruption, intermittent water supply, and low-quality of water. Management of water resources is critical in order to achieve sustainable goals for socio-economic development. This study aims to analyze the water resources supply and demand management of the Iligan City Waterworks System. A methodical examination of collected data from various sources was performed and generated insights are presented through different visualizations. Study shows that the eight watersheds of the city are capable of providing more than the daily target water supply for each member of a household served by ICWS. It also highlights that Non-Revenue Water (NRW) caused by old and leaking pipes significantly affects the quality of service provided by ICWS and is the root of most major problems in water supply in the city.

2.4.3 Use-case

INTRODUCTION

In the Philippines, the main sources of water are rivers, lakes, river basins, and groundwater reservoirs. Undeniably, getting water from these sources to households needs proper planning and implementation. Water is served and provided for by either the local government units, water districts, large-scale private operators, or small-scale independent providers. As reported by the World Health Organization (WHO), there is a water shortage in the Philippines since one out of ten people still does not have access to improved water sources. This resulted in acute water diarrhea as one of the top ten leading causes of death in the Philippines.

Water supply and distribution problems are a major concern of Philippine cities. Population growth, lack of investment in water infrastructure, and the upper limit of the available water supply are three interrelated factors that may contribute to this issue (Van der Bruggen, Borghgraef & Vinckier, 2010). The equitable and efficient distribution of water supply to the residents mandates that policymakers adapt to changes in socio-economic, environmental, and geographic contexts.

From 2009 to 2013, the City Government of Iligan took a loan from the World Bank amounting to Php 365.23 million for the improvement of the water system including the installation of 31.74 km. of transmission and distribution lines, development of source, and construction of treatment and filtration facilities and installation of valves and appurtenances (Implementation Completion and Results Report, World Bank. 2017). But despite the said initiative, water distribution in the city is still a major problem as attested by the Iligan City Waterworks System (ICWS) announcements and posted for dissemination on social media pages.

According to Schouten et al., (2003), a system that reliably and sustainably meets the needs of 80% of the population but leaves the poorest 20% un-served cannot be counted as a success. Iligan City has eight watersheds, and only 30 out of its 44 barangays are supplied with water by the ICWS; this roughly converts to 164,306 out of 342,618 persons (based on the 2015 PSA census) having access to clean water. In addition to non-supply of water in some areas, water interruptions due to leaking (or damaged) pipes are common. Thus, numerous barangays and a great number of the city's population are facing intermittent water source flow. Aside from this, there are several problems identified due to inadequate access to drinking water such as water-related diseases (cholera, diarrhea, dysentery among others), and economic burdens that can have a significant impact on the poor and vulnerable groups (Asian Development Bank, 2015).

With the United Nations Sustainable Development Goals including clean water and sanitation (SDG 6), it is imperative to conduct an initiative that can guide the City Government of Iligan in creating policies and decisions toward the attainment of SDG 6 and addressing the issue of the city's water distribution. This research utilizes data from various sources and aims to share insights through a visualization system that will show the relationship between the water consumption of each household and the actual need of each barangay in Iligan City. Specifically, this study aims to address the intermittent water supplies in every household and sufficient water supply capable to cater household needs per barangay.

OBJECTIVE

To provide insights to policymakers in Iligan City, through a visualization system, on the relationship between the water distribution, consumption, and population of every barangay in the city.

PROBLEM

- Determine the inadequacy of water supply in Iligan City; can be translated to: Why is there an inadequacy of water in Iligan City?
- Determine the water supply provided by water sources (pumping stations)
- Determine the water supply needed based on population

METHODOLOGY

DATA GATHERING

The main source of the data comes from the Waterworks System Inc. of Iligan City (ICWS). There is already an existing MOU between the City Government of Iligan and the Association of Barangay Captains of the City of Iligan, and a few of the clauses stipulated the right to conduct research projects and community service of the University with the assistance of the partners. A letter of request was sent to Iligan City Waterworks System Inc to ask for any available water-related information, especially the water consumption per household per barangay, number of households connected, the list of pumping stations and their location, list of barangays connected to the pumping station, and the discharge rate per pumping station. Table 1 lists the data provided by ICWS. Constant monitoring and follow-ups were done to gather all relevant data.

List of Barangays
Number of Households Connected on every Barangay
List of Pumping Stations
List of Barangays Connected to every Pumping Station
Discharge (Liters Per Second) of every Pumping Stations
Location of every Pumping Station in an Image Format (Figure 1)

Table 1: ICWS Data Provided

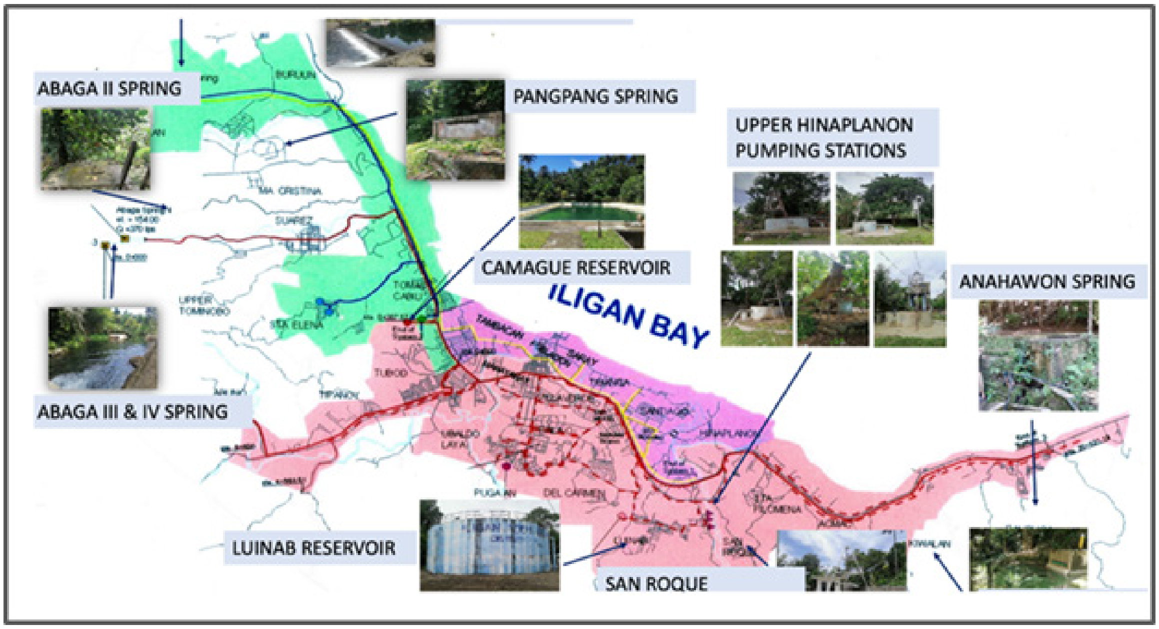


Figure 2.4 - 1: **Map of the Location of the Pumping Stations**

Since we need to determine the total number of residents connected to every barangay and the data provided to ICWS is just the number of households connected to every barangay, we gather the data for the average household size of every barangay in Iligan City (Figure 2) for the year 2015 in the absence of 2022 data from the Philippine Statistics Authority Website (psa.gov.ph).

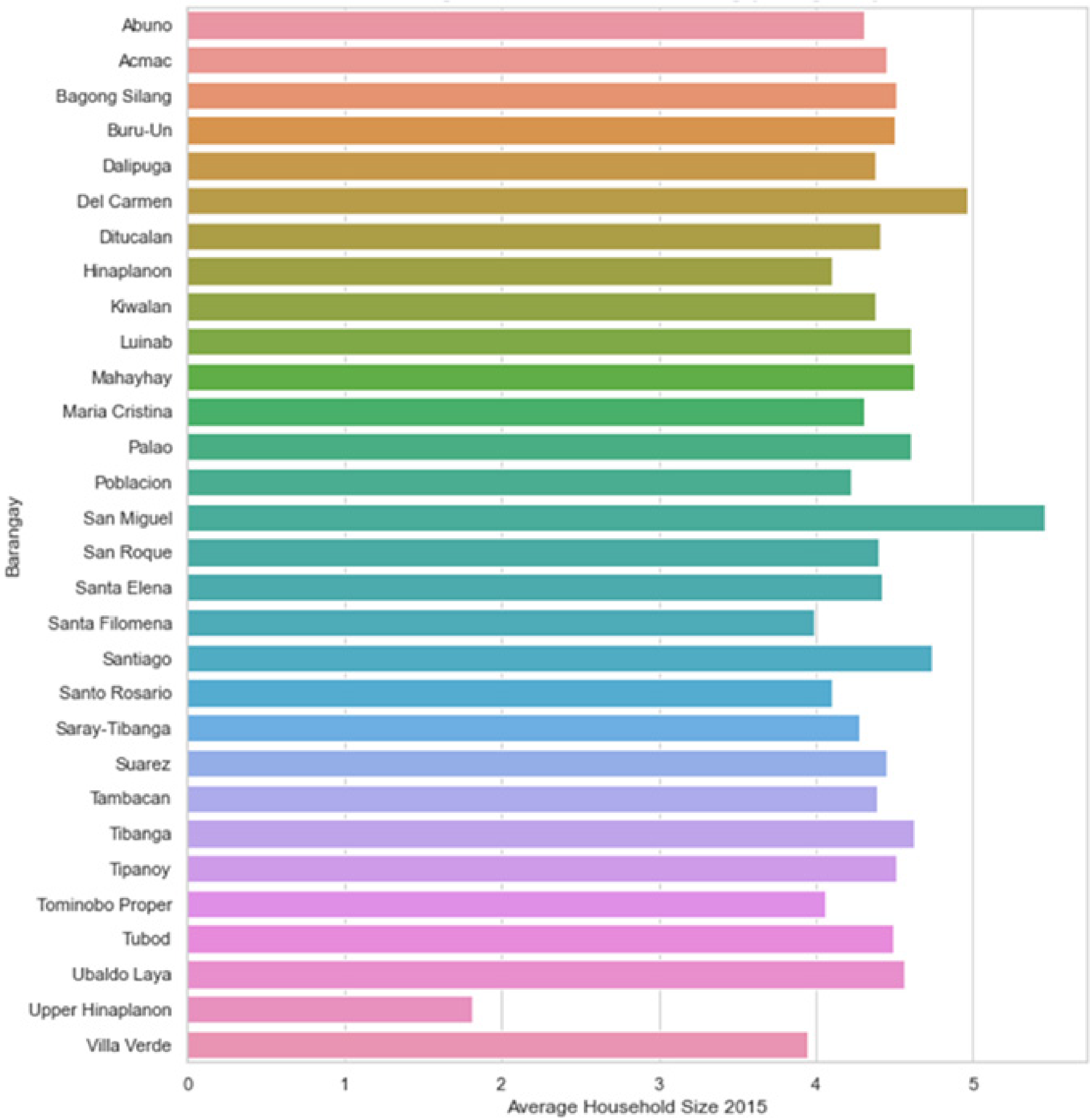


Figure 2.4 - 4: **Ave rage Household Size of Barangays in Iligan City**

EXPLORATORY DATA ANALYSIS

Iligan, officially known as the City of Iligan is an urban city in the region of Northern Mindanao, Philippines (Census of Population, 2020). Iligan is politically subdivided into 44 barangays. Based on the data from ICWS, only 30 out of 44 barangays are connected to ICWS (Figure 3). Surprisingly, only 20% of the Iligan area size is connected to the ICWS and only 44% of Iligan City's total population is connected to the ICWS.

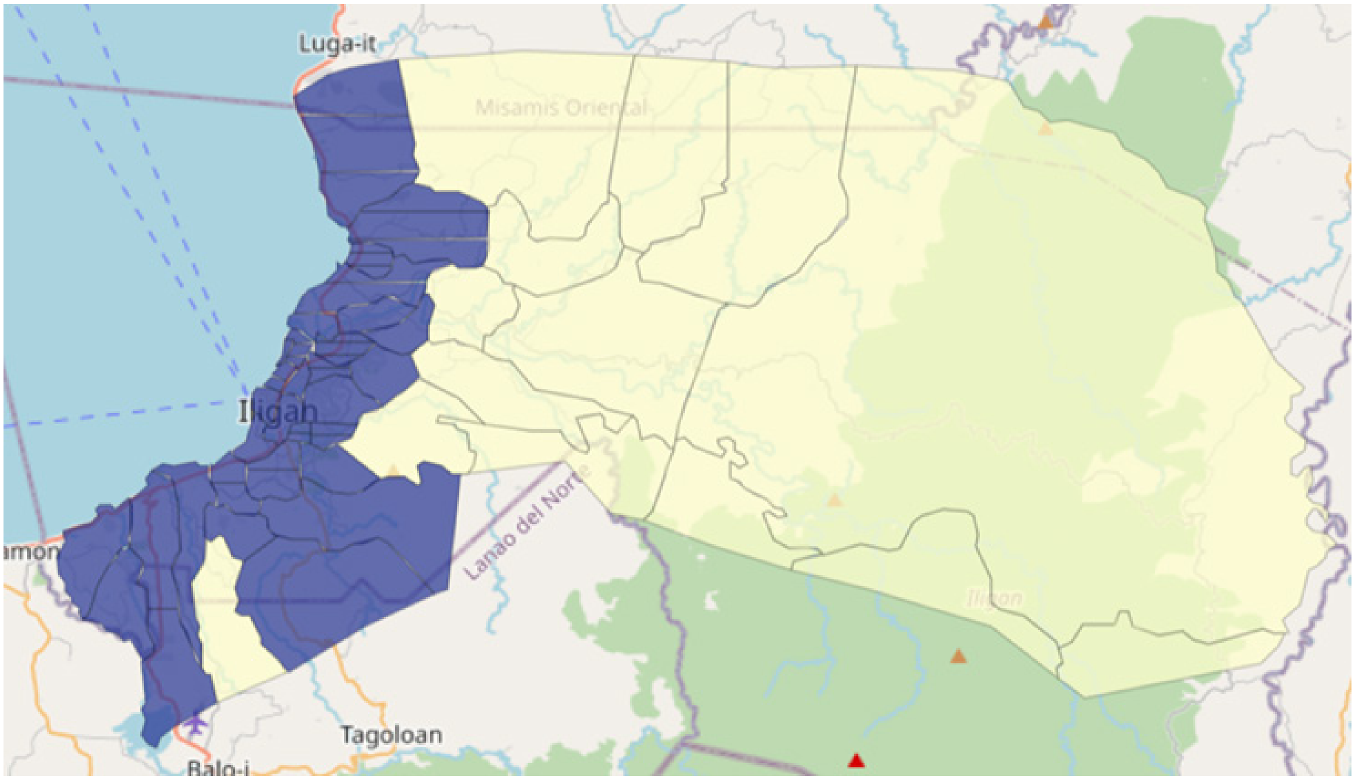


Figure 2.4 - 3: **Map of Iligan City. Blue means connected to ICWS while light yellow means not connected.**

Figure 4, shows the number of connected residents of every barangay. Barangay Tubod has the highest number of connected residents with 23,954 while Abuno has the lowest with 91. It is important to note that not every resident in a barangay is connected to ICWS. As shown in Figure 5, only a portion of households in every barangay are connected to ICWS.

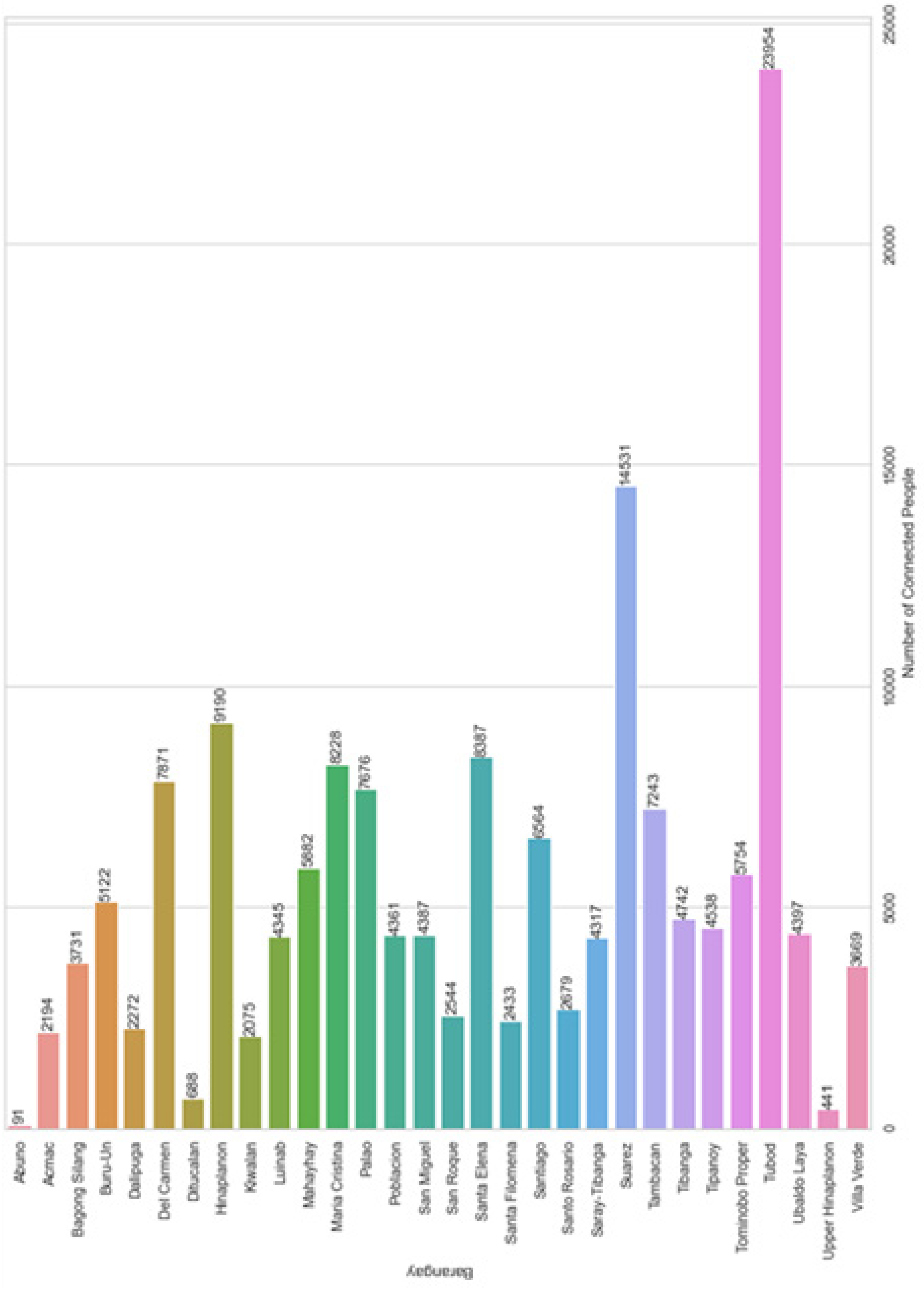


Figure 2.4 - 4: **Number of Connected Residents of Every Barangay**

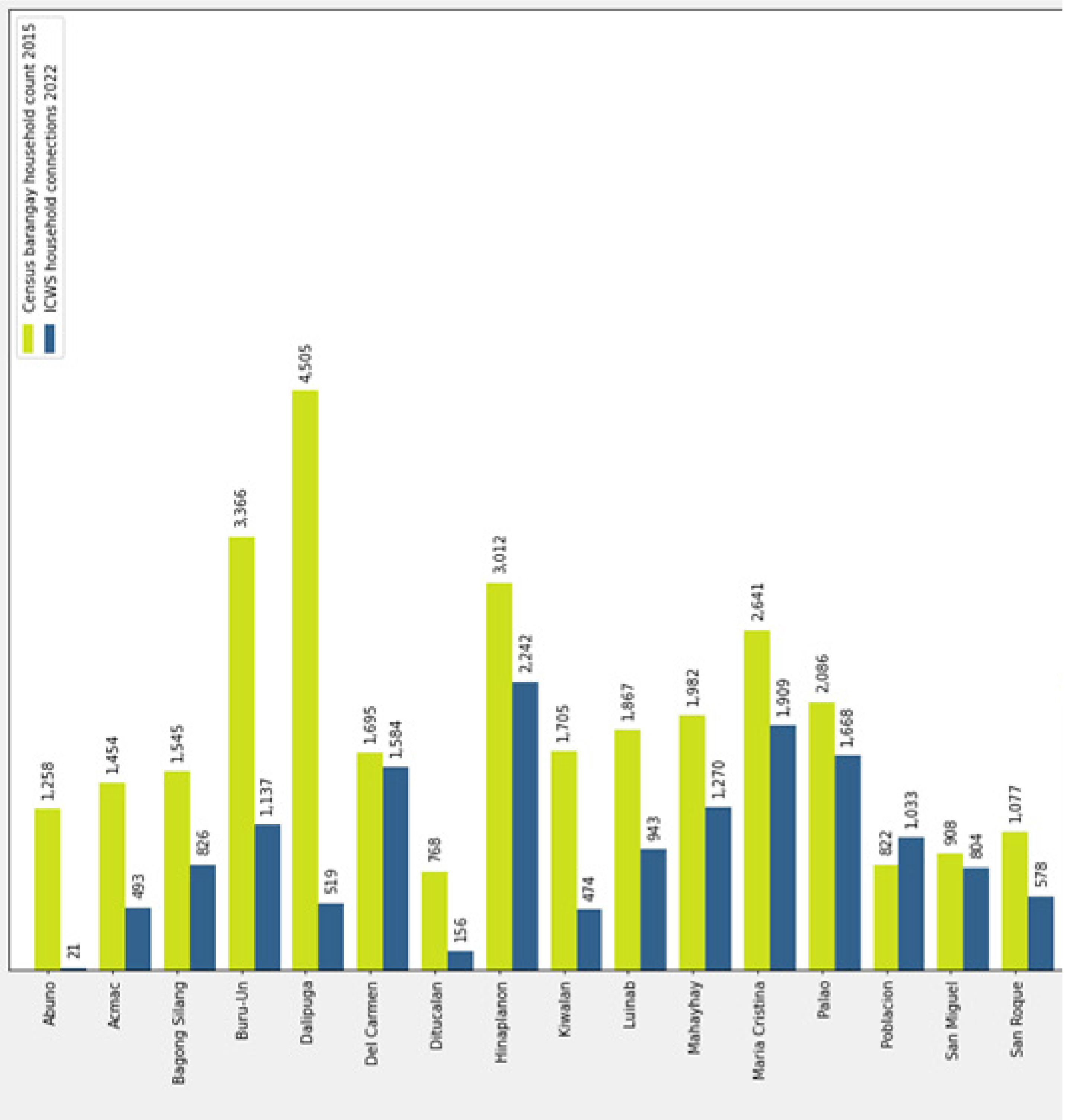


Figure 2.4 - 5: **Comparison of Connected to Households to the total Households of every Barangay**

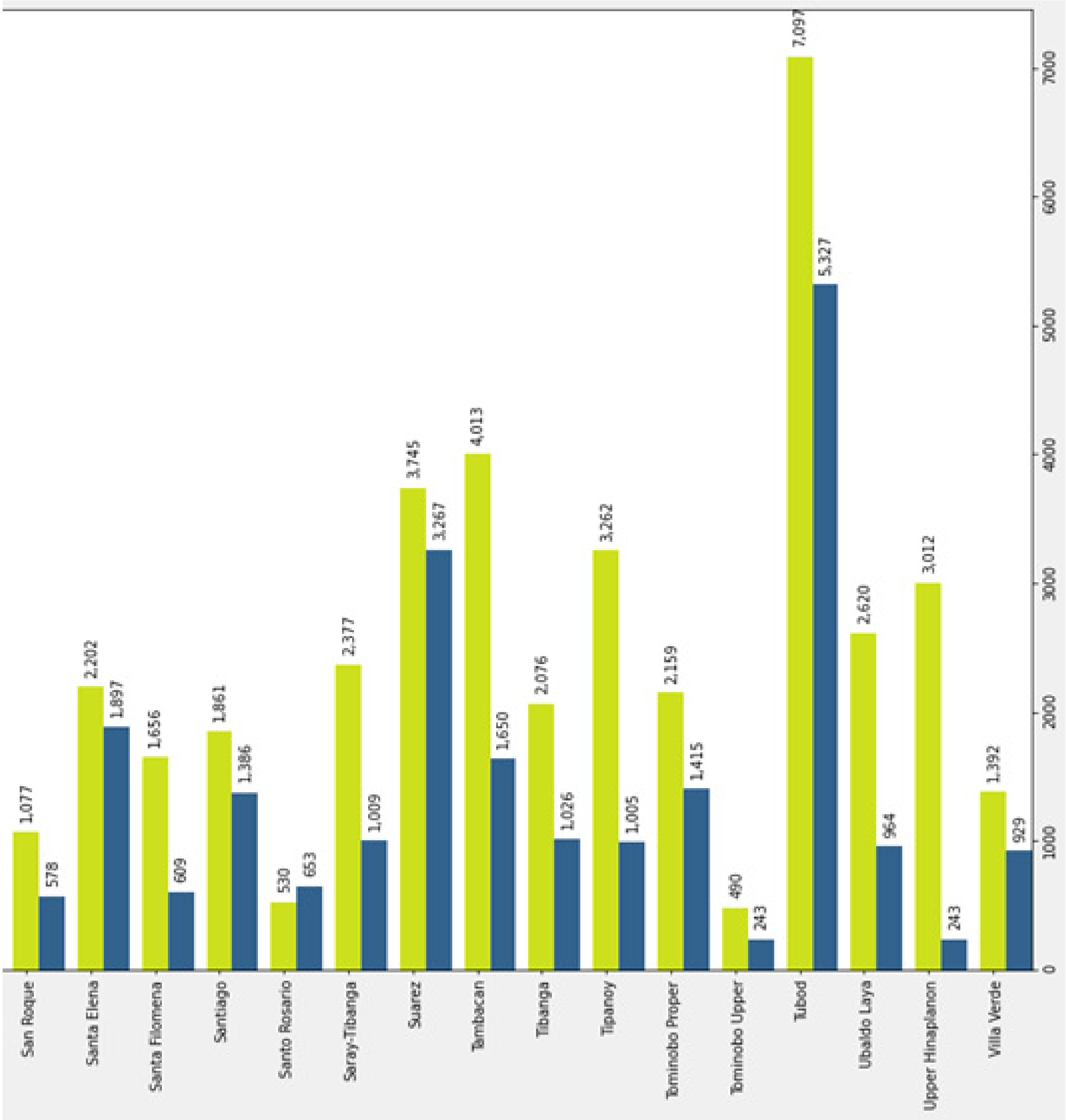


Figure 2.4 - 5: **Comparison of Connected to Households to the total Households of every Barangay**

Currently, there are ten(10) working pumping stations installed and strategically positioned throughout Iligan City (Figure 1). From this figure, we were able to estimate the latitude and longitude of every station that serves as geolocation data needed for the Geographical Information System that we develop. These pumping stations are managed and maintained by Iligan City Water Works System. All of these stations are pumping water for 24 hours a day to their connected barangays (Figure 6) with corresponding discharges (Figure 7).

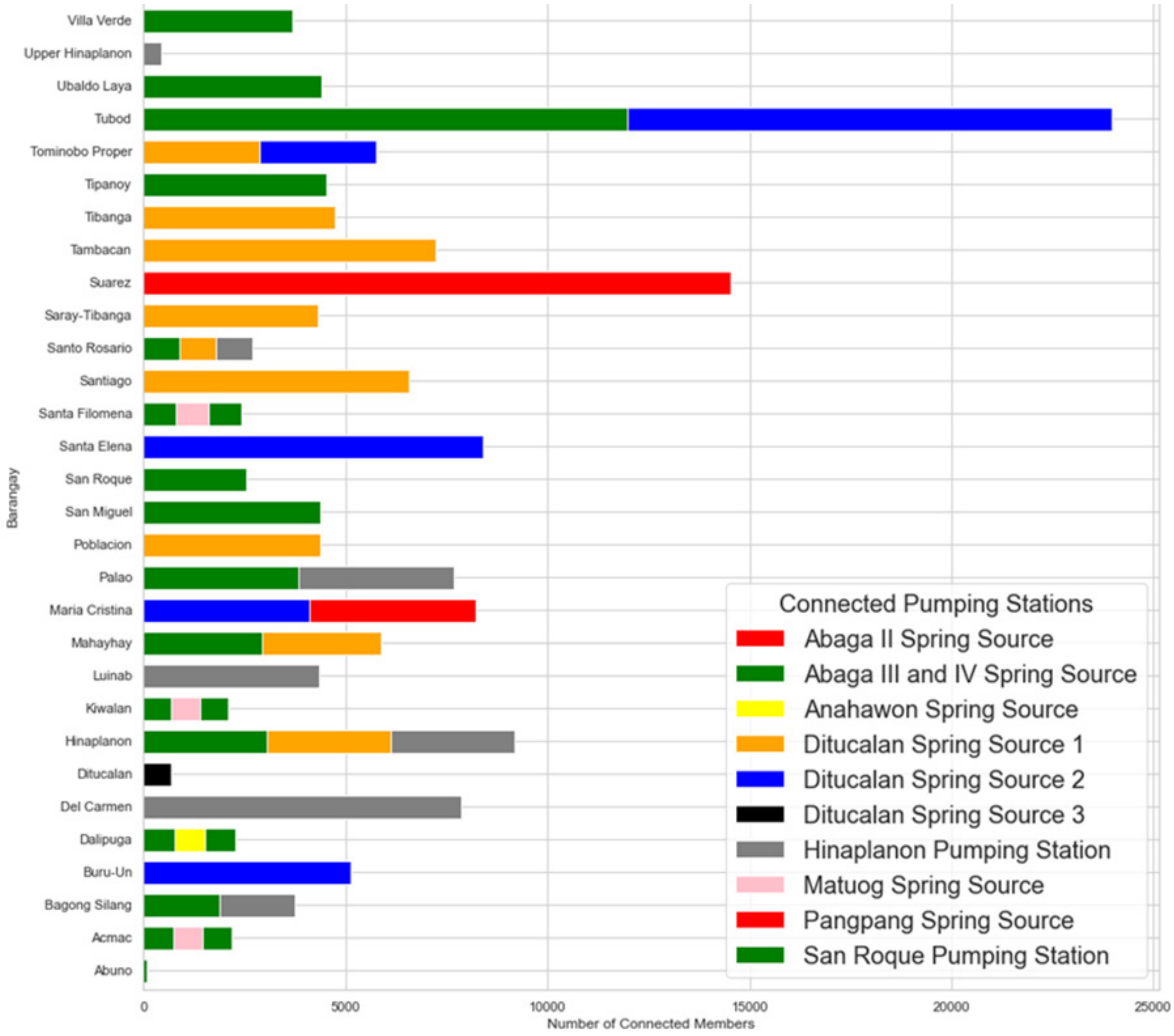


Figure 2.4 - 6: **Visualization of Barangays with their Connected Pumping Stations. For instance, there are two pumping stations connected to Tubod Barangay, namely Abaga III and IV Spring Source and Ditucalan Spring Source 2.**

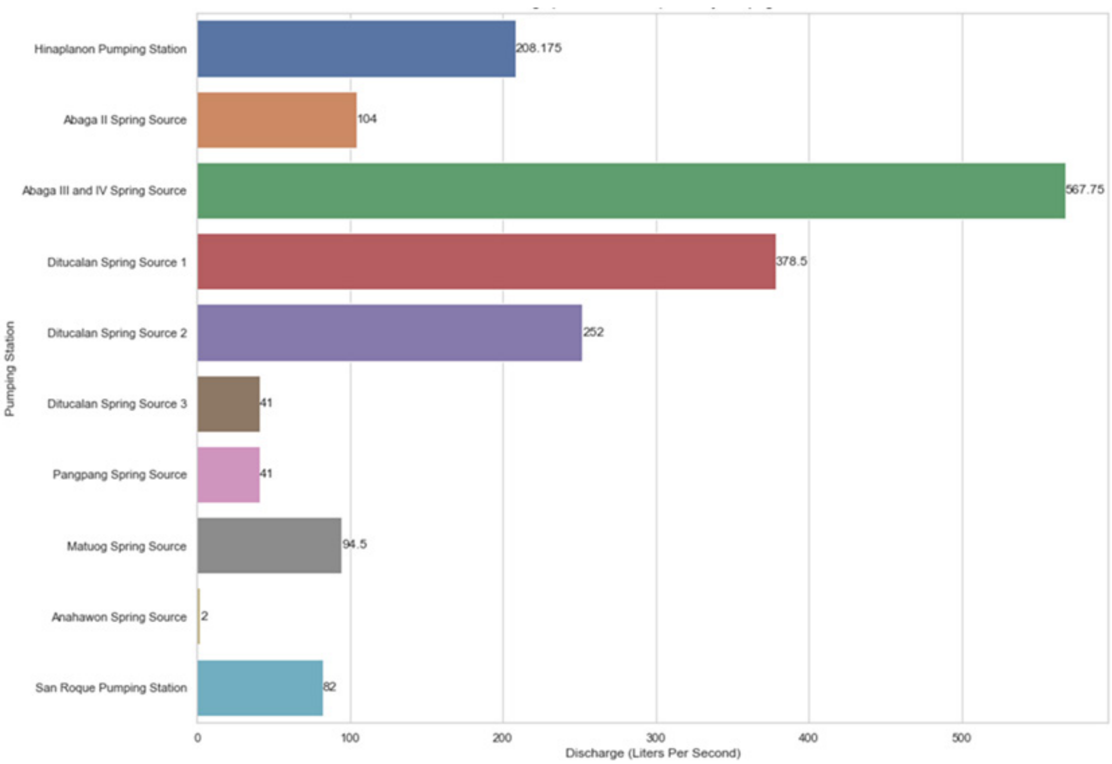
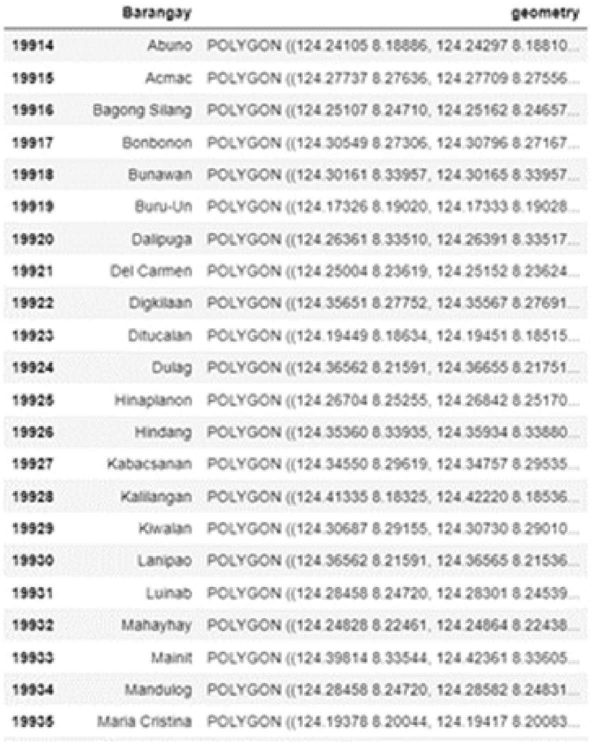


Figure 2.4 - 7: **Discharge (Liters Per Second) for every Pumping Station**

Data Preprocessing for Geographic Data

We were able to download the geographical data in shapefile of every barangay in Iligan City (Figure 8), from GADM, the Database of Global Administrative Areas (gadm.org). This is where you can download high-resolution geographical data of all countries at all levels and on any given time period.

As mentioned above, there are barangays connected to multiple stations and there are pumping stations connected to multiple barangays. For the "Barangays with multiple stations connected", we spatially and equally divided the geographical area of a barangay by its number of stations connected. Correspondingly, we also divided its total connected members by the number of stations connected. We named the divided areas of Barangay to "Name of Barangay" Area "Number". For instance, Barangay Hinaplanon is divided into Hinaplanon Area 1, Hinaplanon Area



	Barangay	geometry
19914	Abuno	POLYGON ((124.24105 8.18886, 124.24297 8.18810...
19915	Acmac	POLYGON ((124.27737 8.27636, 124.27709 8.27556...
19916	Bagong Silang	POLYGON ((124.25107 8.24710, 124.25162 8.24657...
19917	Bonbonon	POLYGON ((124.30549 8.27306, 124.30796 8.27167...
19918	Bunawan	POLYGON ((124.30161 8.33957, 124.30165 8.33957...
19919	Buru-Un	POLYGON ((124.17326 8.19020, 124.17333 8.19028...
19920	Dalipuga	POLYGON ((124.26361 8.33510, 124.26391 8.33517...
19921	Del Carmen	POLYGON ((124.25004 8.23619, 124.25152 8.23624...
19922	Digkilaan	POLYGON ((124.35651 8.27752, 124.35567 8.27691...
19923	Ditucalan	POLYGON ((124.19449 8.18634, 124.19451 8.18515...
19924	Dulag	POLYGON ((124.36562 8.21591, 124.36655 8.21751...
19925	Hinaplanon	POLYGON ((124.26704 8.25255, 124.26842 8.25170...
19926	Hindang	POLYGON ((124.35360 8.33935, 124.35934 8.33880...
19927	Kabacsanan	POLYGON ((124.34550 8.29619, 124.34757 8.29535...
19928	Kalilangan	POLYGON ((124.41335 8.18325, 124.42220 8.18536...
19929	Kiwalan	POLYGON ((124.30687 8.29155, 124.30730 8.29010...
19930	Lanipao	POLYGON ((124.36562 8.21591, 124.36565 8.21536...
19931	Luinab	POLYGON ((124.28458 8.24720, 124.28301 8.24539...
19932	Mahayhay	POLYGON ((124.24828 8.22461, 124.24864 8.22438...
19933	Mainit	POLYGON ((124.39814 8.33544, 124.42361 8.33605...
19934	Mandulog	POLYGON ((124.28458 8.24720, 124.28582 8.24831...
19935	Maria Cristina	POLYGON ((124.19378 8.20044, 124.19417 8.20083...

Figure 2.4 - 8: **Geographical data (geometry column) of the barangays in Iligan City**

2, and Hinaplanon Area 3. (Figure 9). As for the "Pumping stations connected to multiple barangays", we divided the discharge volume of the pumping station by the number of connected barangays (Figure 10).

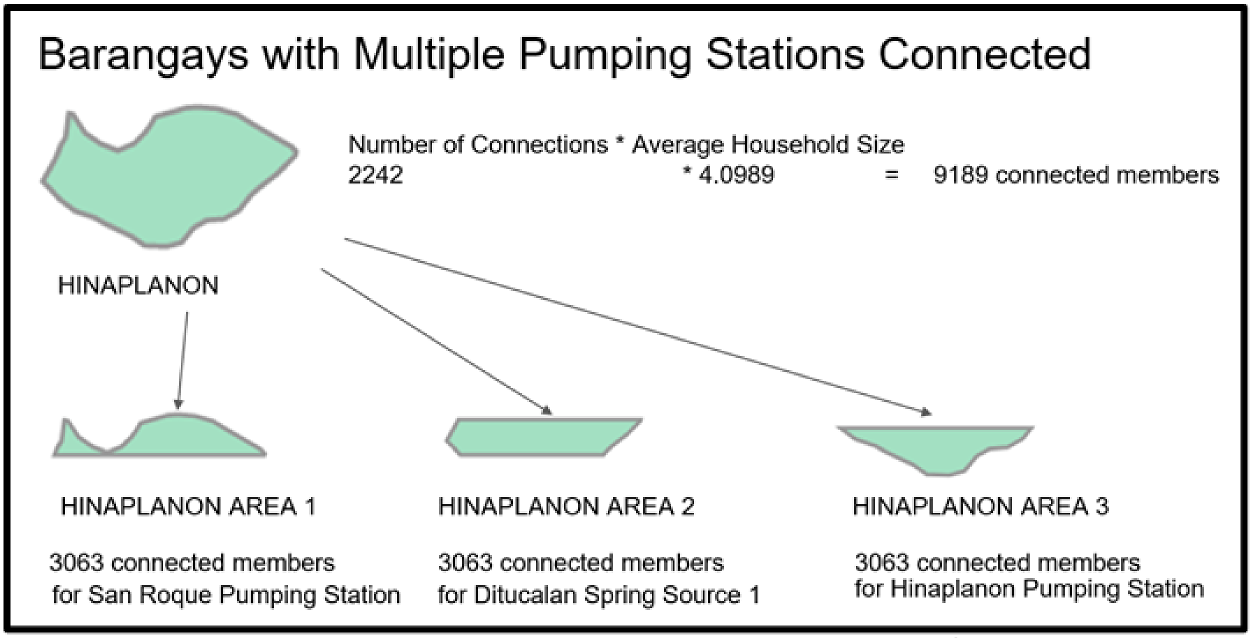

Barangays with Multiple Pumping Stations Connected
Number of Connections * Average Household Size
2242 * 4.0989 = 9189 connected members
HINAPLANON
HINAPLANON AREA 1
3063 connected members
for San Roque Pumping Station
HINAPLANON AREA 2
3063 connected members
for Ditucalan Spring Source 1
HINAPLANON AREA 3
3063 connected members
for Hinaplanon Pumping Station

Figure 2.4 - 9. **Dealing with Barangays with Multiple Pumping Stations**

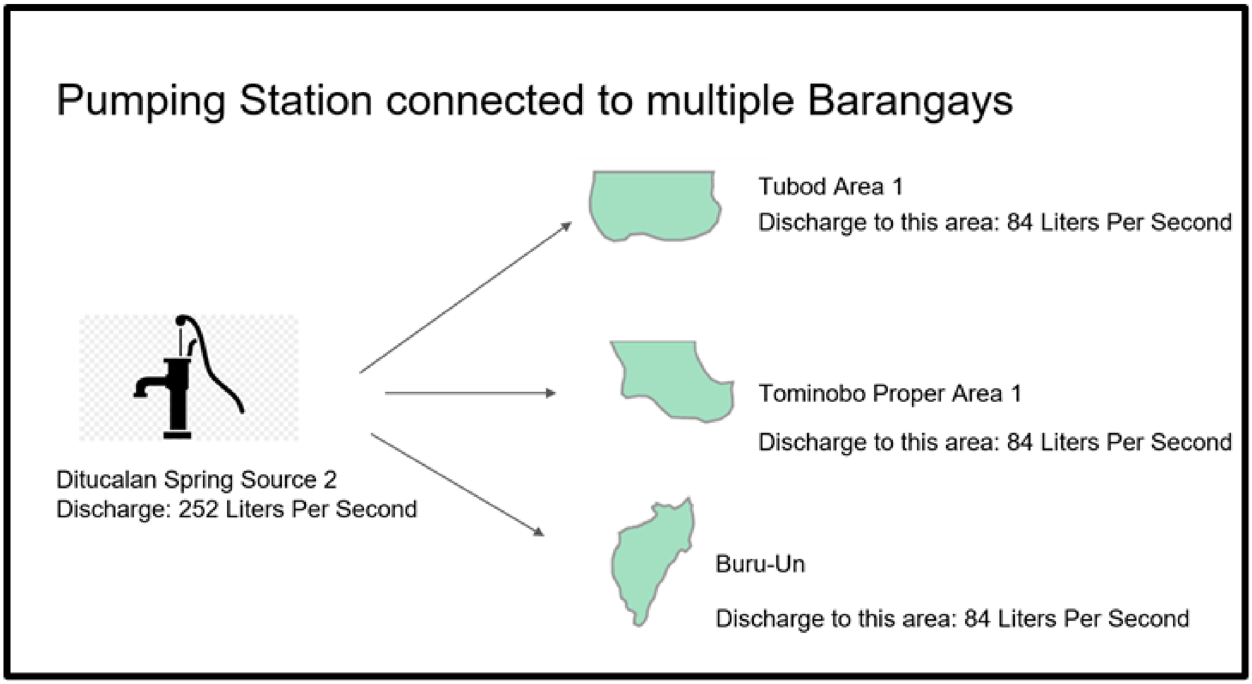

Pumping Station connected to multiple Barangays
Tubod Area 1
Discharge to this area: 84 Liters Per Second
Tominobo Proper Area 1
Discharge to this area: 84 Liters Per Second
Ditucalan Spring Source 2
Discharge: 252 Liters Per Second
Buru-Un
Discharge to this area: 84 Liters Per Second

Figure 2.4 -10. **Dealing with Pumping Stations connected to multiple barangays**

Non-Revenue Water

According to ICWS, there is a 64% unaccounted water supply called "Non-Revenue Water" (NRW). Based on the data, the total water supply is 168,709 m3/day but the total billed volume is only 59,048 m3/day. This means that there is 97,091 m3/day unbilled or unaccounted, constituting to 64% of the total water supply. Hence, only 36% of the total water supply is accounted for.

Getting the Average Water Supply of every Barangay

The main objective of this study is to determine if the barangay is undersupplied with water or not. This is represented by the value "Average Water (Liters Per Day) of the Barangay. We devise a formula for getting the mentioned value through deliberation among us and careful analysis of the available data. The formula is as follows:

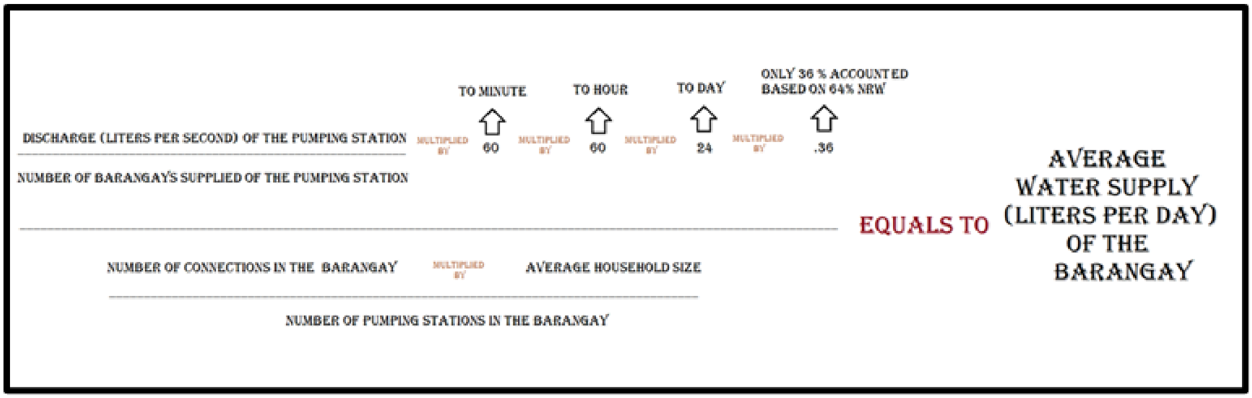

TO MINUTE
TO HOUR
TO DAY
ONLY 36 % ACCOUNTED BASED ON 64% NRW
DISCHARGE (LITERS PER SECOND) OF THE PUMPING STATION
MULTIPLIED BY
60
MULTIPLIED BY
60
MULTIPLIED BY
24
MULTIPLIED BY
.36
NUMBER OF BARANGAY'S SUPPLIED OF THE PUMPING STATION
EQUALS TO
AVERAGE WATER SUPPLY (LITERS PER DAY) OF THE BARANGAY
NUMBER OF CONNECTIONS IN THE BARANGAY
MULTIPLIED BY
AVERAGE HOUSEHOLD SIZE
NUMBER OF PUMPING STATIONS IN THE BARANGAY

Figure 2.4 - 11. **The formula for getting the Average Water Supply of the Barangay**

The water discharge per pumping station to every barangay was originally provided in liters per. The original value of the water discharge is multiplied by 60 (hence it was in seconds) for it to be liters per minute, then, multiplied by 60 again to get the discharged water in liters per hour. Then, multiplied it by 24 to convert it into liters per day. Finally, the value in liters per day was multiplied by 0.36 (see Section 4.4) to represent the accounted water supply. The derived value is divided by the number of connected members per barangay which gives you the value of the "Average Water Supply (Liters Per Day) of the Barangay.

Getting the route from a pumping station to a barangay

In visualizing the route of water tubes from a pumping station to a barangay, we used the library osmnx specifically its method djikstra_path. This gets the shortest path from the pumping station to the centroid of the barangay following the street layout of Iligan City (Figure 12).

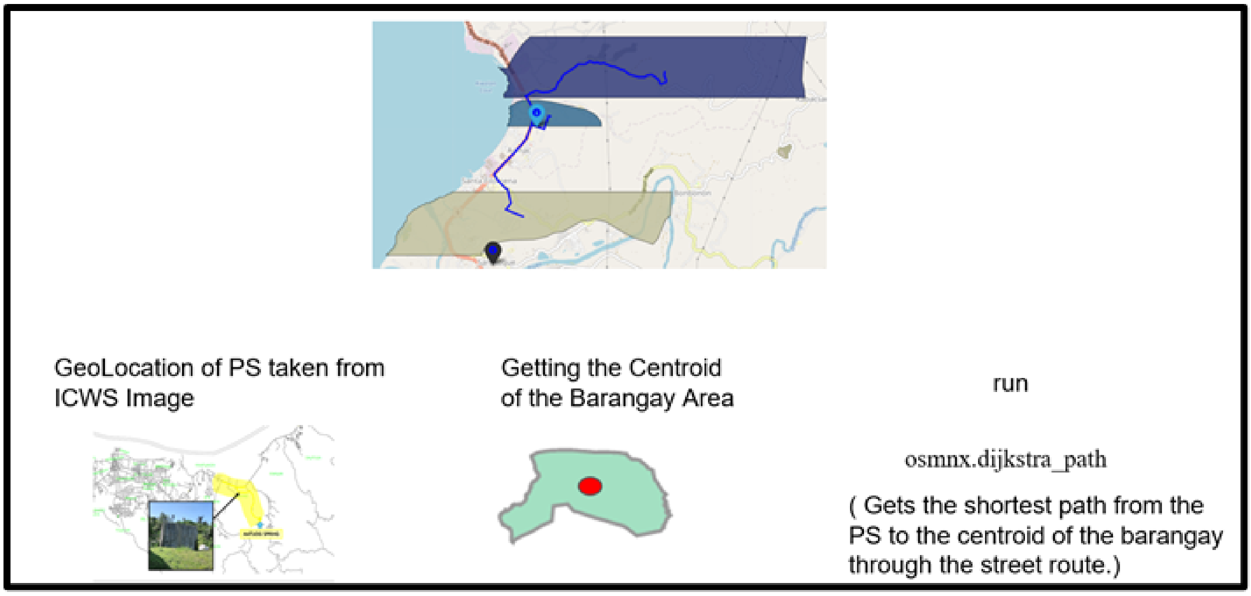

GeoLocation of PS taken from ICWS Image
Getting the Centroid of the Barangay Area
run
osmnx.dijkstra_path
(Gets the shortest path from the PS to the centroid of the barangay through the street route.)

Figure 2.4 - 12. **Getting the Route from a Pumping Station to a Barangay**

Results and Discussion

Getting the Undersupplied Barangays

Using the formula in deriving the average water supply (section 4.5), we were able to determine barangays that are undersupplied (Figure 12). According to ICWS, the average consumption in liters per day for every person is 260. This means that any value lesser than 260 is considered undersupplied. As shown in Figure 12, 26 barangays are deemed to be undersupplied.

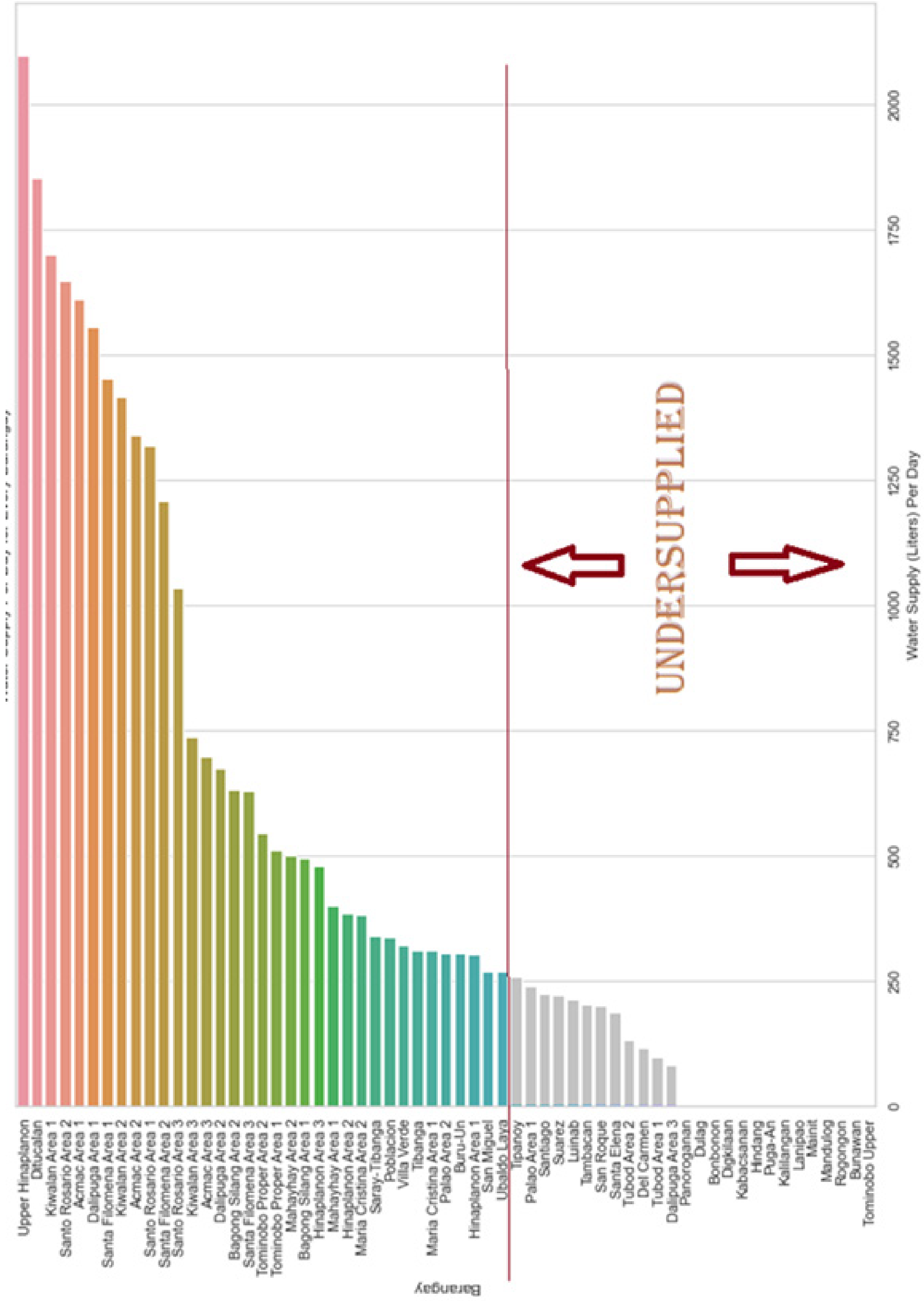


Figure 2.4 -12 B. **Water Supply Per Day for Every Barangay**

Further, Figure 13 shows an overview of the relationship between the Barangay, Daily Water supply, and Pumping Station.

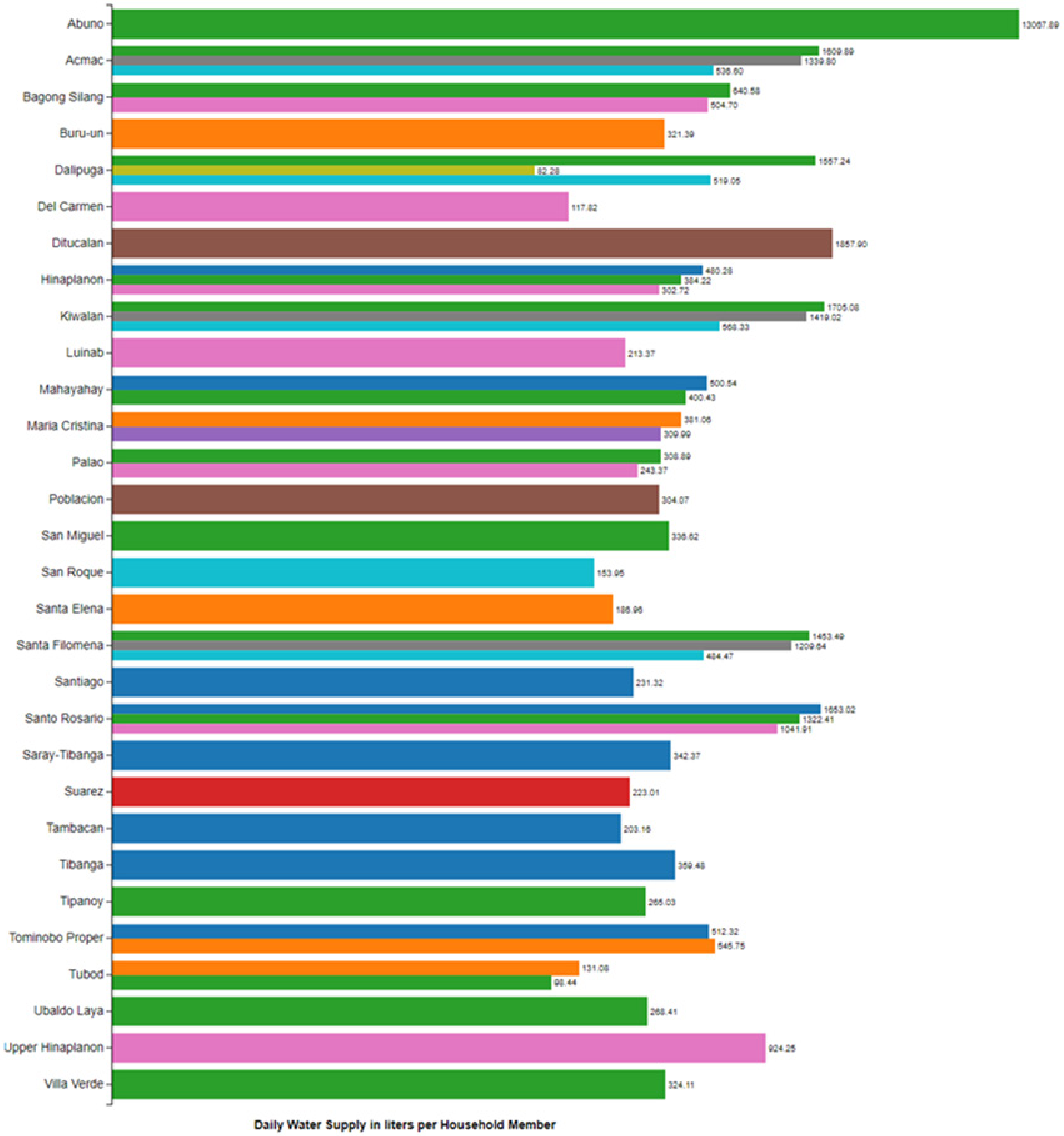


Figure 2.4 - 13. **Daily Water Supply in Liter Per Household Member for every Barangay**

We were also able to visualize the barangays that are undersupplied and which pumping station is supplying it (Figure 14). For instance, the Hinaplanon Pumping Station and Ditucalan Spring Source fall short in supplying water to almost half of its connected Barangay Areas, while San Roque Pumping Station, Abaga III, and IV Spring Source, and Matuog Spring Source are performing well in terms of providing enough water to their connected barangay areas.

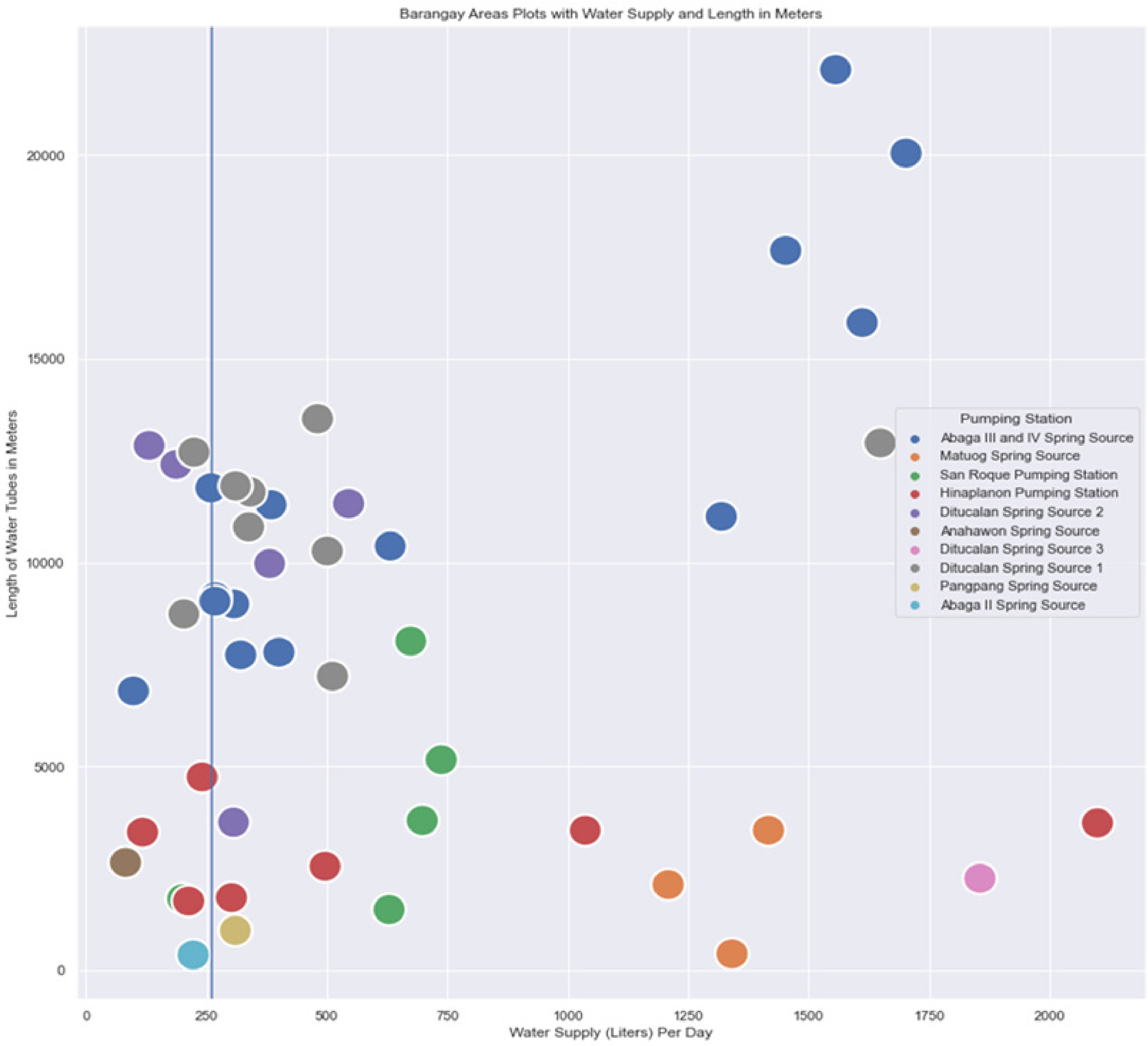


Figure 2.4 - 14. **Barangay Area Plots with Water Supply and Length**

We also look into the estimated Water Supply once the total population of every barangay is connected to ICWS. As expected, the water supply value decreases, and the number of barangays with shortage of water supply increases (Figure 15).

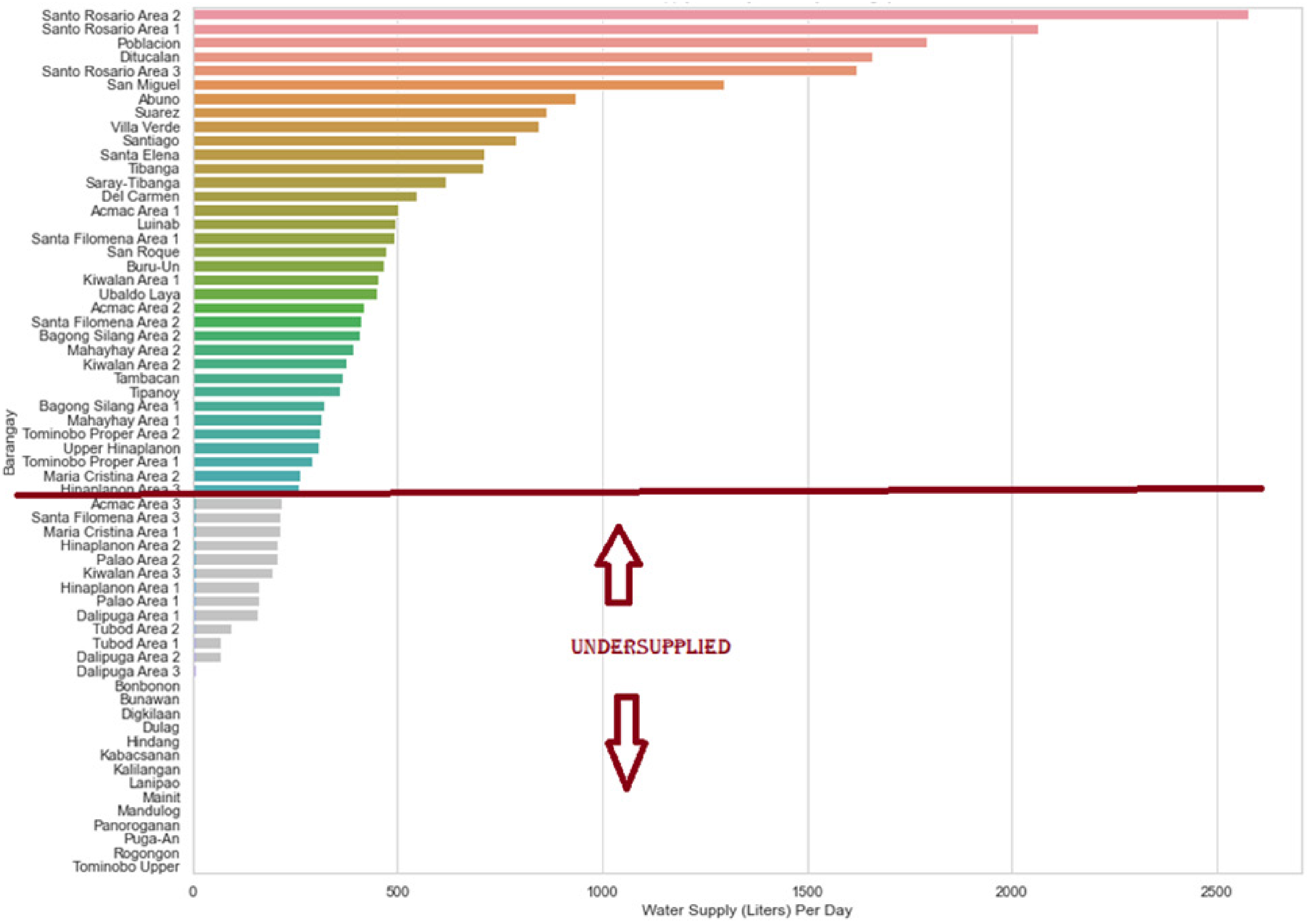


Figure 2.4 - 15. **Water Supply if the Total Population of Every Barangay Connects to ICWS**

To meet the demand for water once every household is connected to ICWS some pumping stations need to increase their discharges. Estimates of discharges are computed to answer the question of how much water supply is needed to prevent undersupply of all barangays. These are as follows:

- Abaga III and IV Spring Source

From 567 to 2,120 liters per second (lps)

- Anahawon Spring Source

From 2 to 62 lps

- Ditucalan Spring Source 2

From 252 to 706 lps

- Hinaplanon Pumping Station

From 208 to 331 lps

- Pangpang Spring Source

From 41 to 50 lps

- San Roque Pumping Station

From 82 to 310 lps

Pumping Station Analysis

Figure 16 shows the geographic location of the pumping stations in Iligan City. Sources are either through spring or deep well.

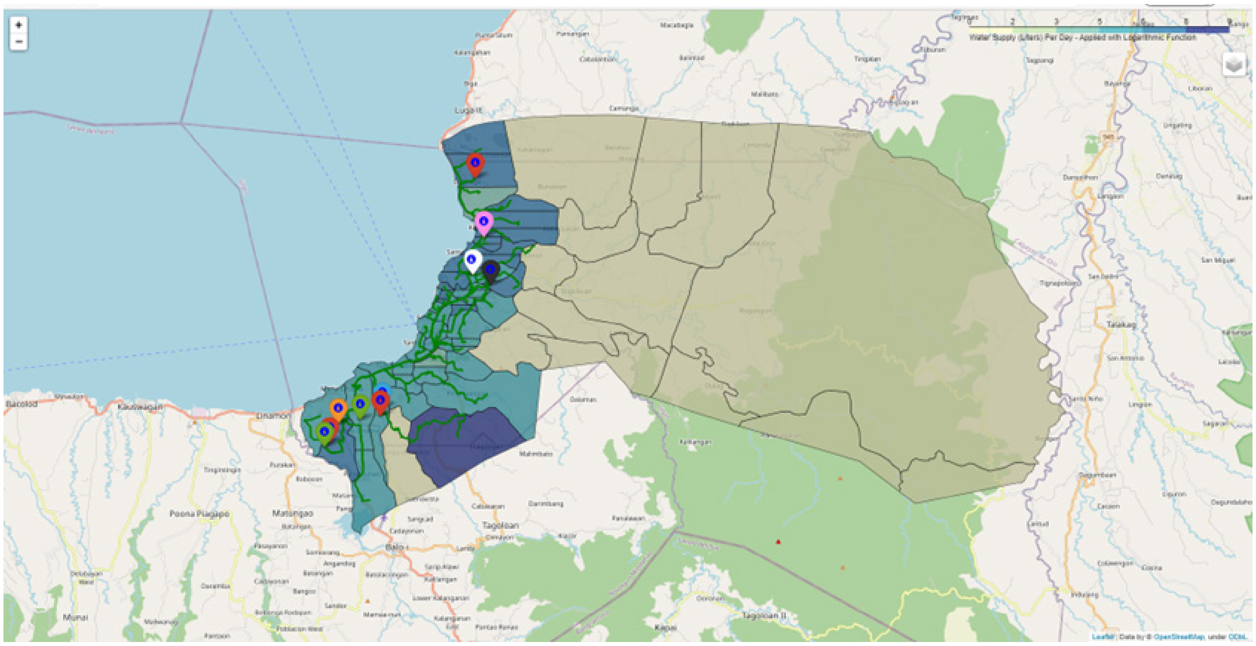

2
3
5
6
8
Water Supply (Liters) Per Day - Applied with Logarithmic Function
Leaflet | Data by © OpenStreetMap, under ODbL.

Figure 2.4 - 16. **Iligan City Map showing the location of the Pumping Stations**

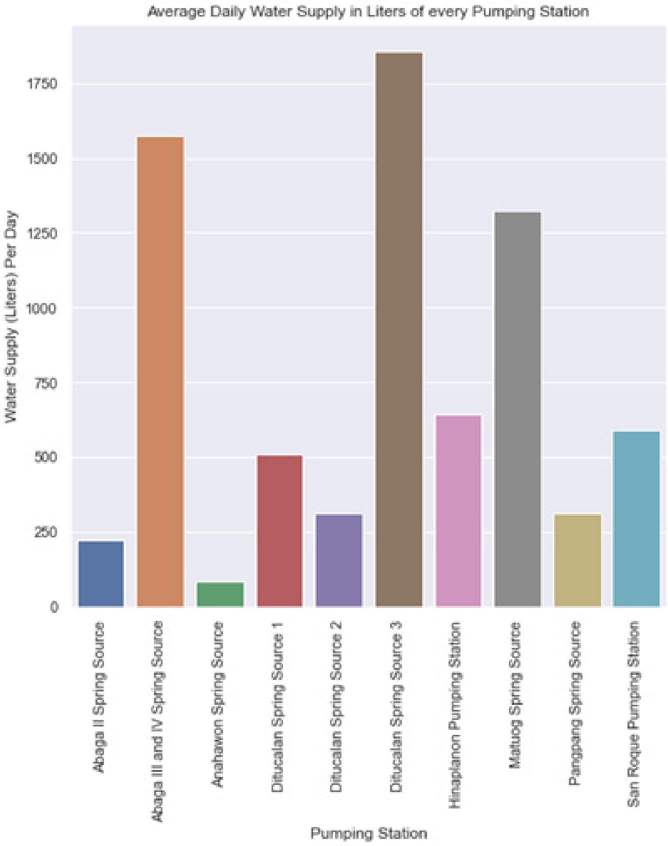


Figure 2.4 - 17. **Average Daily Water Supply in Liters of Every Pumping Station.**

As depicted in Figure 15, each pumping station can produce different amounts of water. For instance, Ditucalan Spring Source 3 has been producing the highest average daily water supply with more than 1750 liters per day to every connected member while Anahawon has the lowest with less than 100 liters per day.

There are also vague results on some pumping stations. As for the San Roque pumping station (Figure 18), it is providing a shortage in the water supply to the areas where the station is located but providing a high volume of water to its other connected areas. On the other hand, the Ditucalan Spring Source 1 (Figure 19) has been providing a high volume of water supply to Santo Rosario Area 2 while providing a decent supply to other barangays except for Santiago and Tambacan where they are receiving under-supply of water.

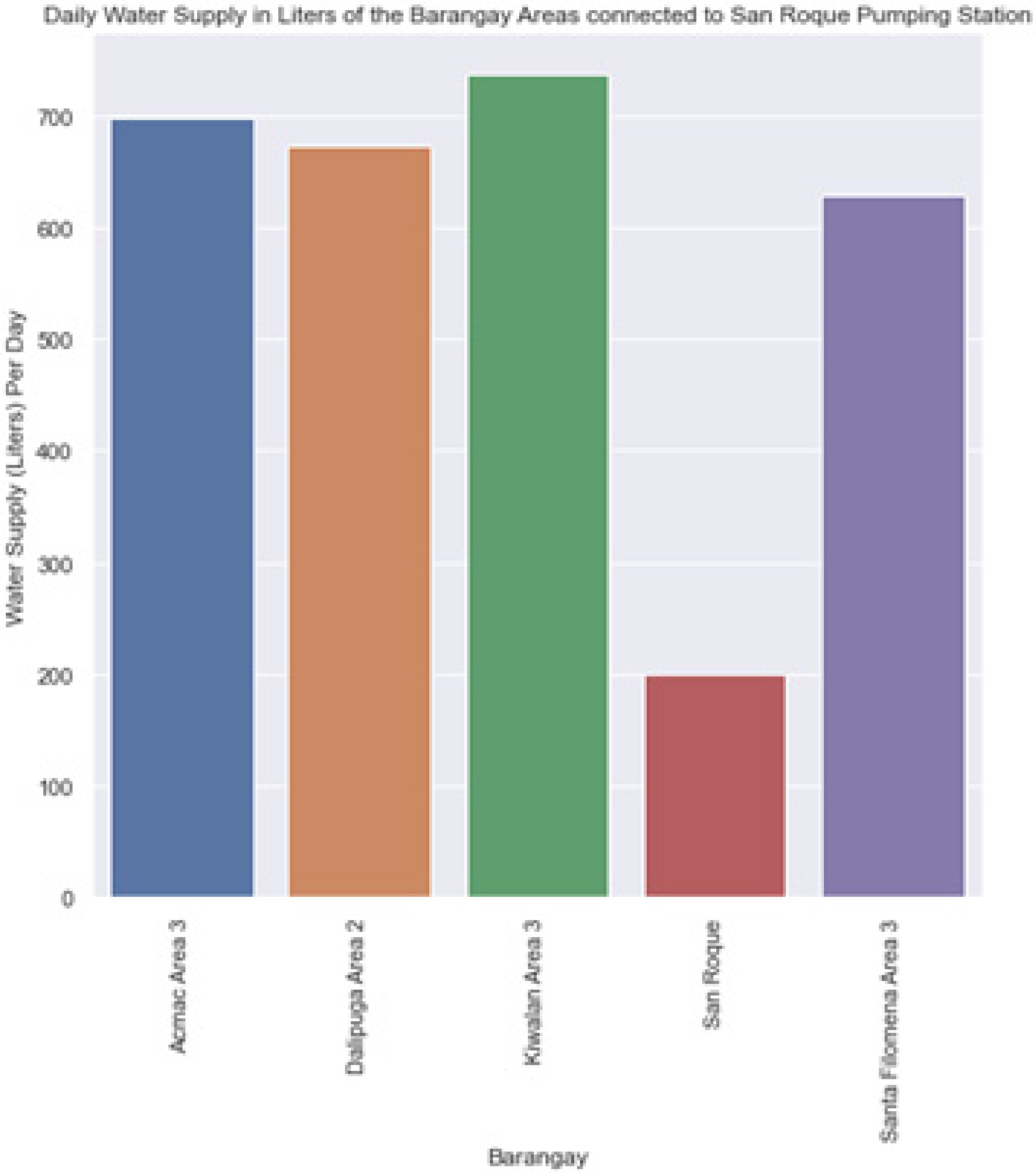


Figure 2.4 - 18. **Daily Water Supply of the Barangay Areas connected to San Roque Pumping Station.**

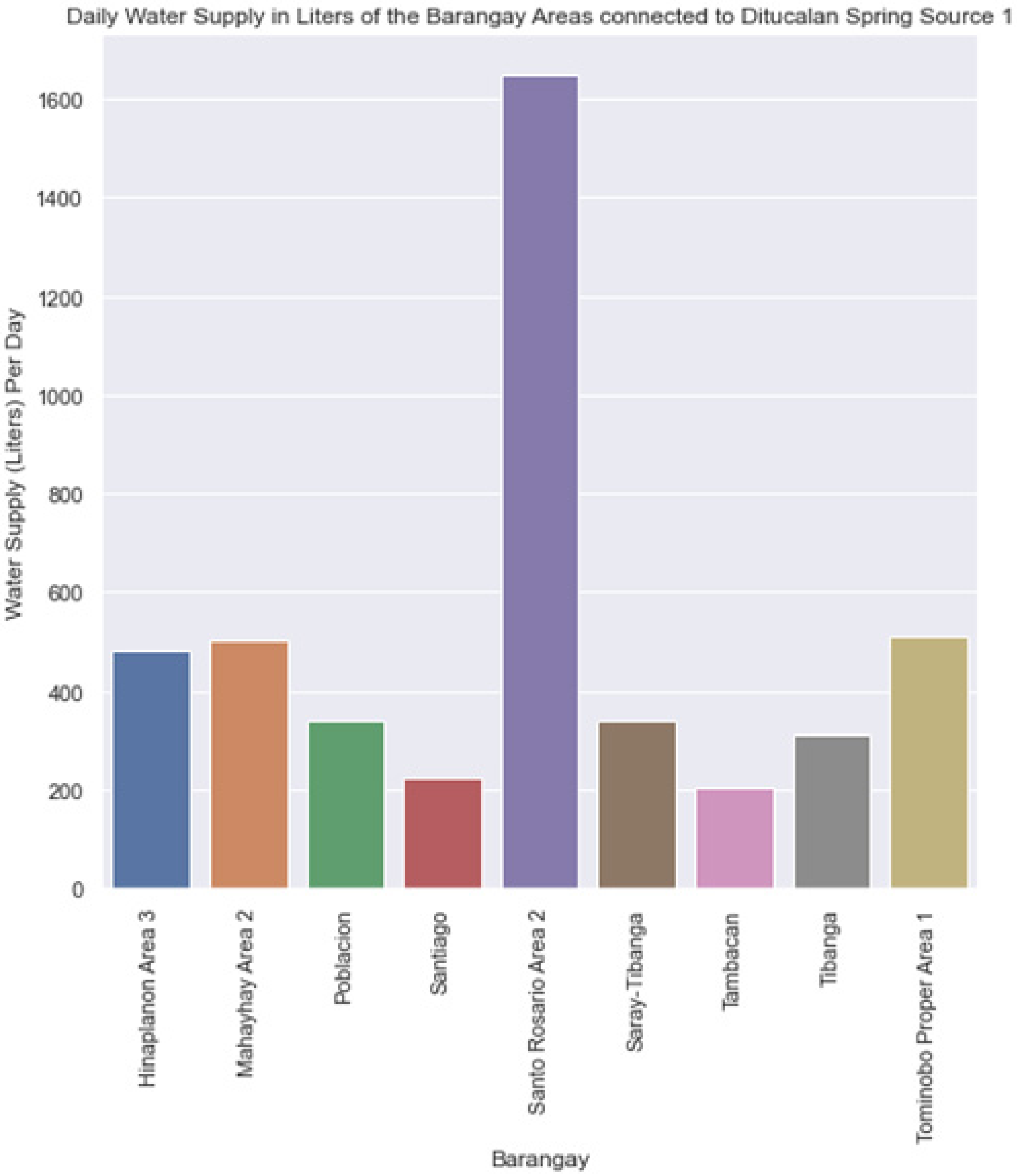


Figure 2.4 - 19. **Daily Water Supply of the Barangay Areas connected to Ditucalan Spring Source 1.**

The Ditucalan Spring Source 2 has been providing undersupply of water to two of its connected barangay areas (Figure 19) while in Hinaplanon pumping station, most of the barangay areas connected to this station are undersupplied (Figure 19).

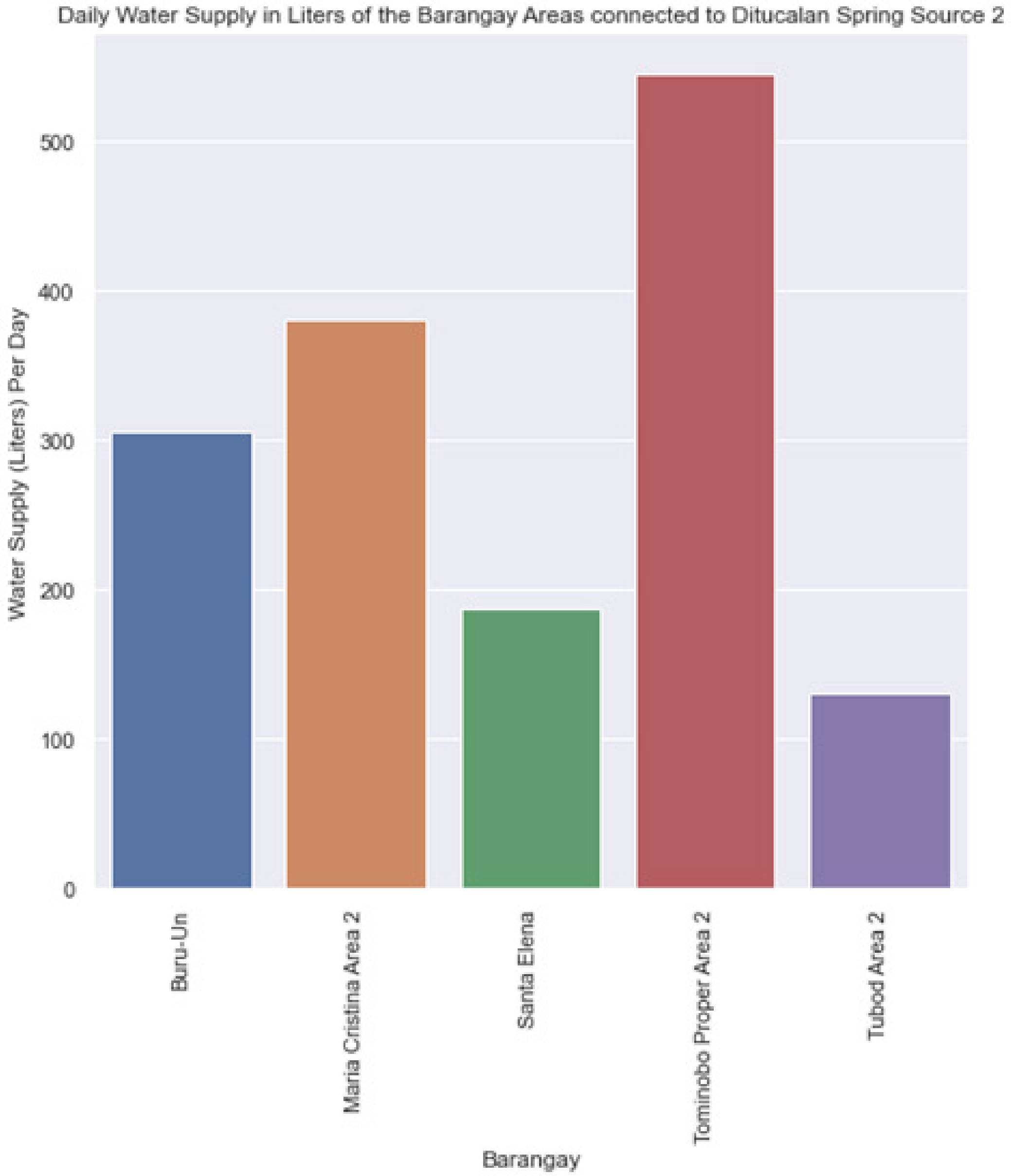


Figure 2.4 - 20. **Daily Water Supply of the Barangay Areas connected to Ditucalan Spring Source 2.**

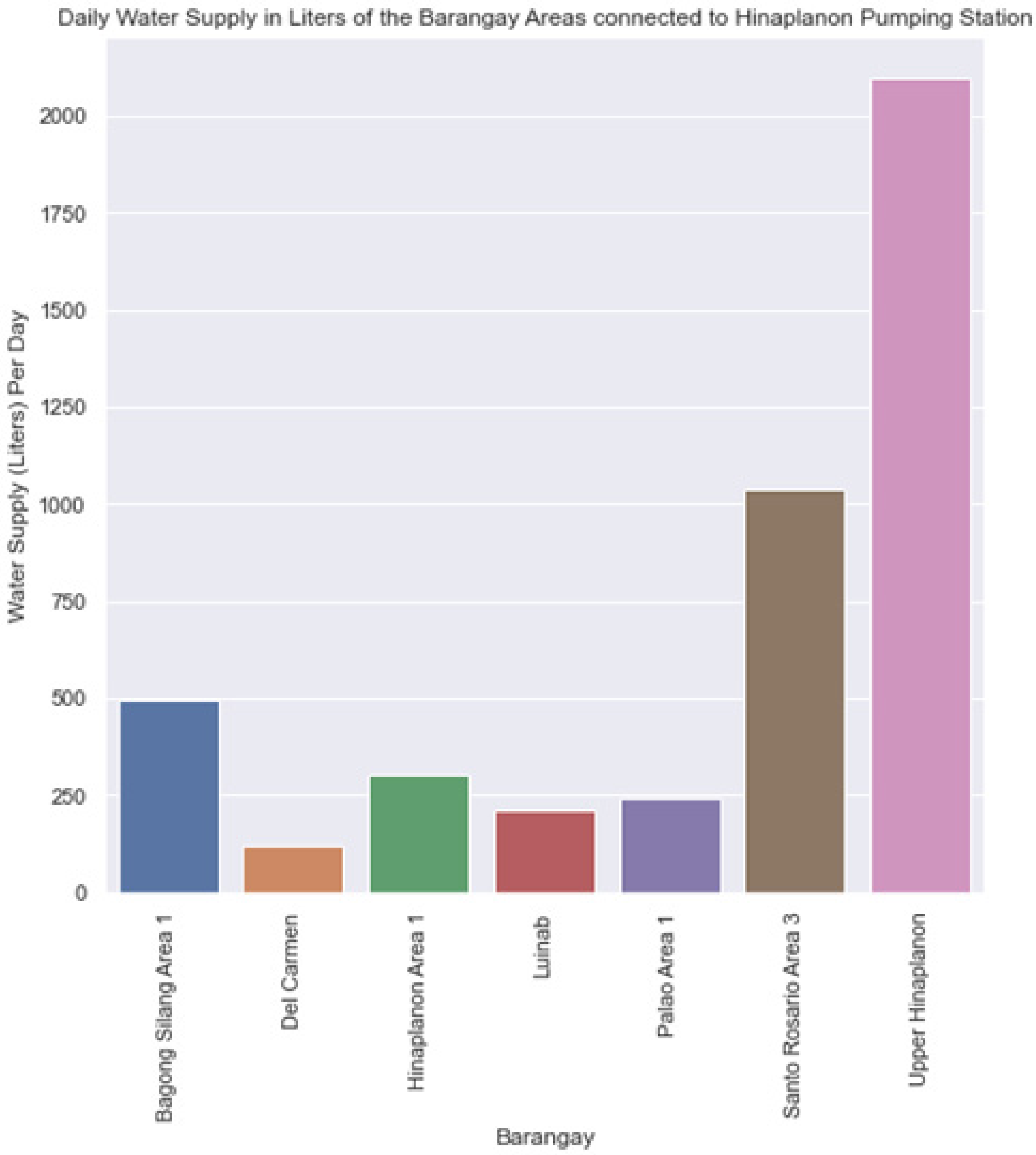


Figure 2.4 - 21. Daily Water Supply of the Barangay Areas connected to Hinaplanon Pumping Station.

The Abaga III and IV Spring Source, consequently, is providing a massive volume of water to Barangay Abuno but inadequate supply to Tubod Area 1 (Figure 22).

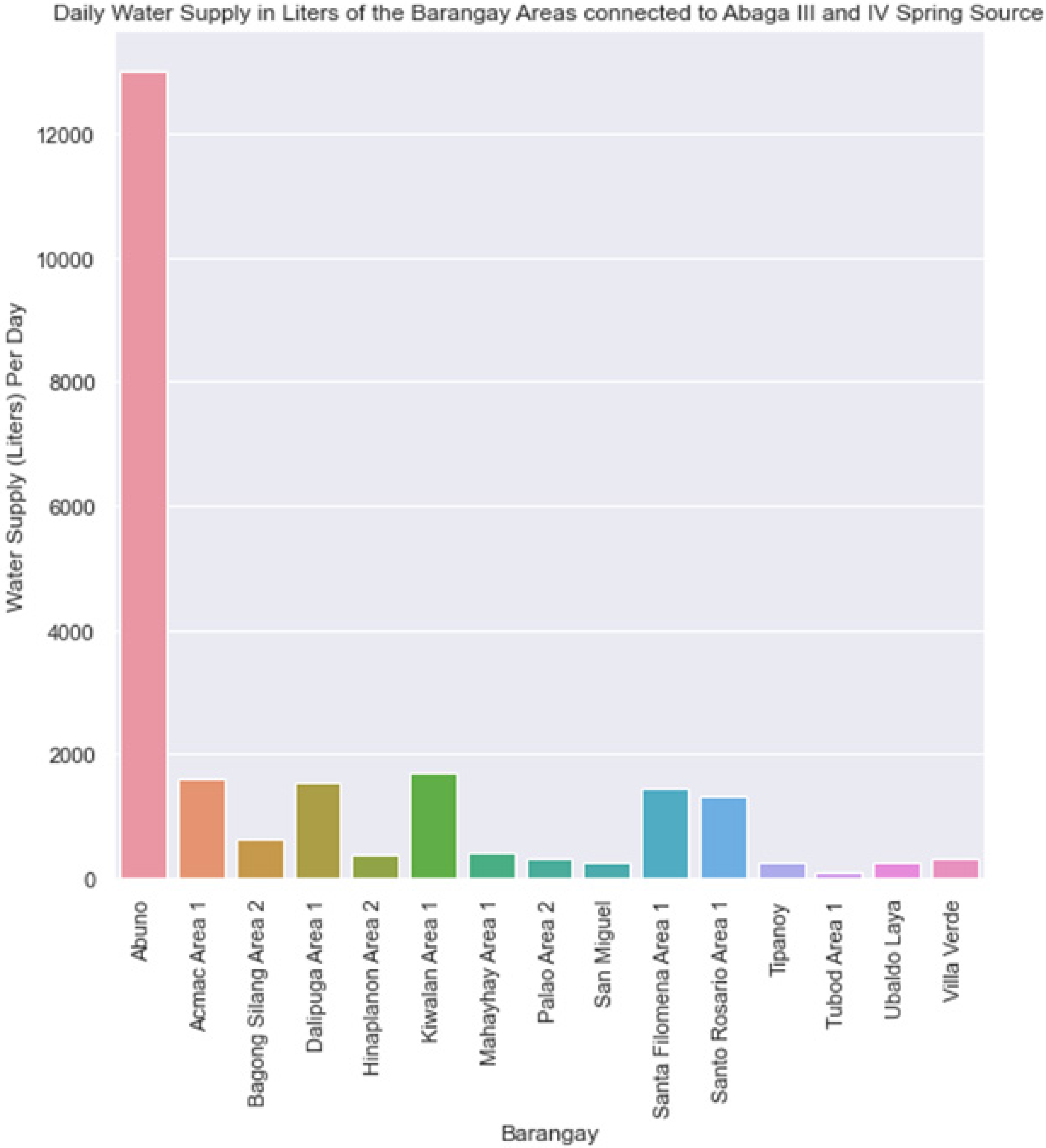


Figure 2.4 - 22. **Daily Water Supply of the Barangay Areas connected to Abaga III and IV Spring Source.**

As shown from the figures above, some of the Pumping Stations are providing a disproportionate water supply to their connected barangay areas. Some barangay areas are receiving a high volume of water while some areas are undersupplied from the same pumping station.

Mapping of Barangay Areas and Pumping Station in terms of Water Supply

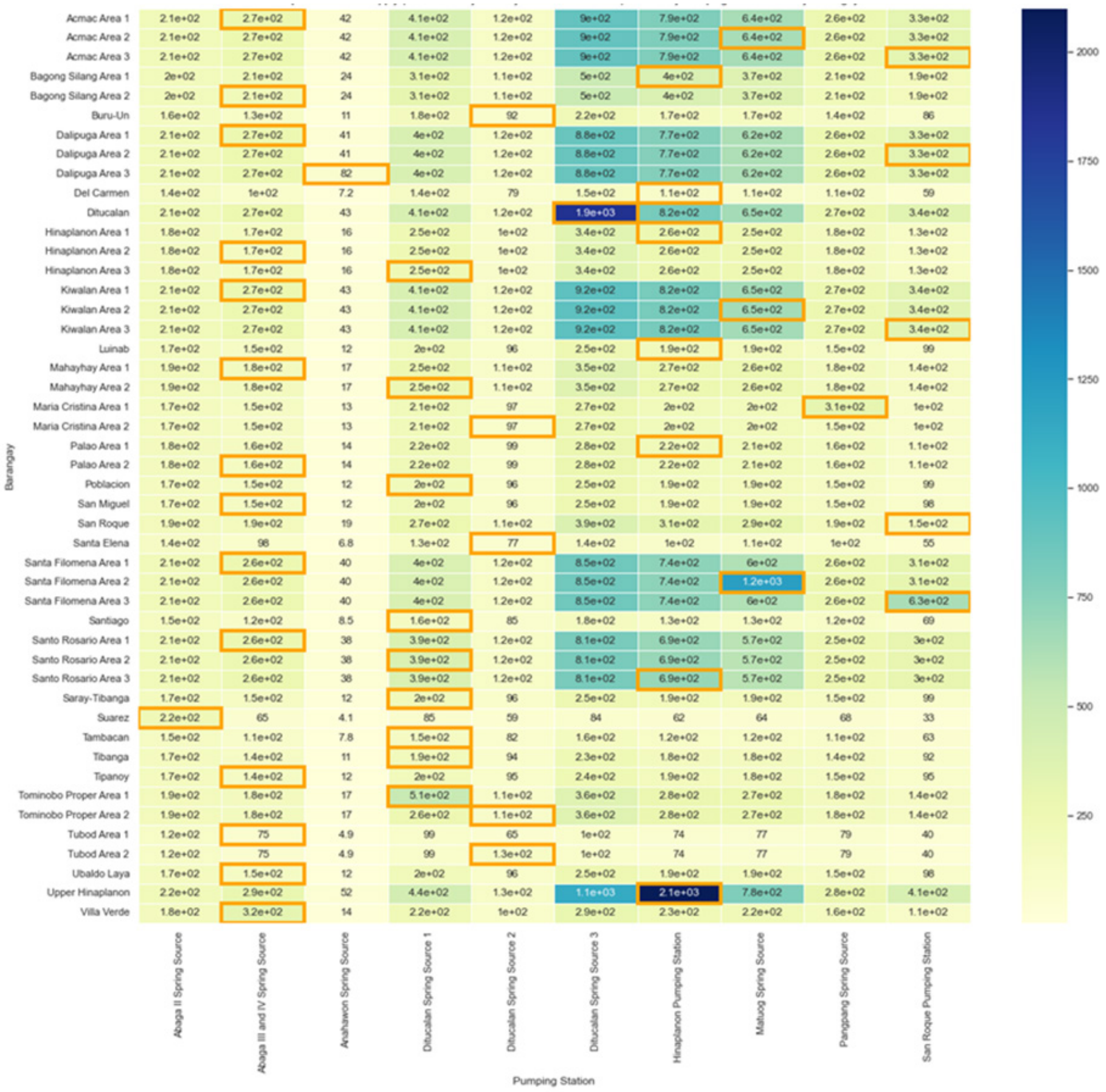
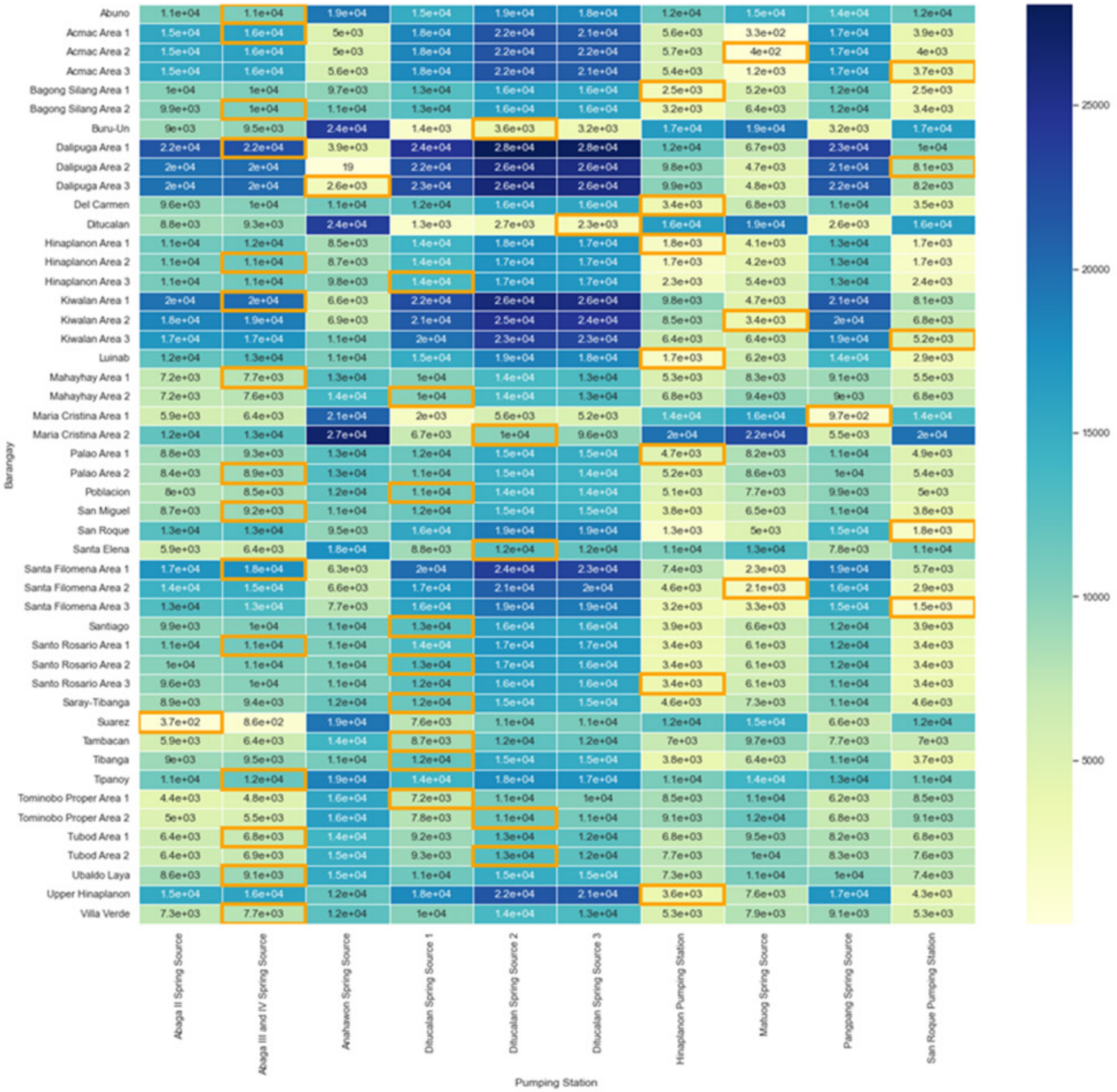


Figure 2.4 - 23. **Water Supply(lpd) of Heat Map from every Pumping Station to every Barangay. The cell of the intersection of these two axes is the water supply value. This visual shows where the ideal pumping station should a barangay area be connected.**

The Barangay Areas: Ditucalan, Santa Filomena Area 2, Suarez, Tominbo Proper Area 1, and Upper Hinaplanon have the ideal connected pumping station while the rest could have a higher water supply if connected to their ideal pumping station. For instance, Acmac Area 1 will have a higher water supply if connected to Ditucalan Spring Source 3.

Mapping of Barangay Areas and Pumping Station in terms of Distance of Water Tubes



Barangay	Abaga II Spring Source	Abaga III and IV Spring Source	Anahawon Spring Source	Ditucalan Spring Source 1	Ditucalan Spring Source 2	Ditucalan Spring Source 3	Hinaplanon Pumping Station	Matuog Spring Source	Pangpang Spring Source	San Roque Pumping Station
Abuno	1.1e+04	1.1e+04	1.9e+04	1.5e+04	1.9e+04	1.8e+04	1.2e+04	1.5e+04	1.4e+04	1.2e+04
Acmac Area 1	1.5e+04	1.6e+04	5e+03	1.8e+04	2.2e+04	2.1e+04	5.6e+03	3.3e+02	1.7e+04	3.9e+03
Acmac Area 2	1.5e+04	1.6e+04	5e+03	1.8e+04	2.2e+04	2.2e+04	5.7e+03	4e+02	1.7e+04	4e+03
Acmac Area 3	1.5e+04	1.6e+04	5.6e+03	1.8e+04	2.2e+04	2.1e+04	5.4e+03	1.2e+03	1.7e+04	3.7e+03
Bagong Silang Area 1	1e+04	1e+04	9.7e+03	1.3e+04	1.6e+04	1.6e+04	2.5e+03	5.2e+03	1.2e+04	2.5e+03
Bagong Silang Area 2	9.9e+03	1e+04	1.1e+04	1.3e+04	1.6e+04	1.6e+04	3.2e+03	6.4e+03	1.2e+04	3.4e+03
Buru-Un	9e+03	9.5e+03	2.4e+04	1.4e+03	3.6e+03	3.2e+03	1.7e+04	1.9e+04	3.2e+03	1.7e+04
Dalipuga Area 1	2.2e+04	2.2e+04	3.9e+03	2.4e+04	2.8e+04	2.8e+04	1.2e+04	6.7e+03	2.3e+04	1e+04
Dalipuga Area 2	2e+04	2e+04	19	2.2e+04	2.6e+04	2.6e+04	9.8e+03	4.7e+03	2.1e+04	8.1e+03
Dalipuga Area 3	2e+04	2e+04	2.6e+03	2.3e+04	2.6e+04	2.6e+04	9.9e+03	4.8e+03	2.2e+04	8.2e+03
Del Carmen	9.6e+03	1e+04	1.1e+04	1.2e+04	1.6e+04	1.6e+04	3.4e+03	6.8e+03	1.1e+04	3.5e+03
Ditucalan	8.8e+03	9.3e+03	2.4e+04	1.3e+03	2.7e+03	2.3e+03	1.6e+04	1.9e+04	2.6e+03	1.6e+04
Hinaplanon Area 1	1.1e+04	1.2e+04	8.5e+03	1.4e+04	1.8e+04	1.7e+04	1.8e+03	4.1e+03	1.3e+04	1.7e+03
Hinaplanon Area 2	1.1e+04	1.1e+04	8.7e+03	1.4e+04	1.7e+04	1.7e+04	1.7e+03	4.2e+03	1.3e+04	1.7e+03
Hinaplanon Area 3	1.1e+04	1.1e+04	9.8e+03	1.4e+04	1.7e+04	1.7e+04	2.3e+03	5.4e+03	1.3e+04	2.4e+03
Kiwalan Area 1	2e+04	2e+04	6.6e+03	2.2e+04	2.6e+04	2.6e+04	9.8e+03	4.7e+03	2.1e+04	8.1e+03
Kiwalan Area 2	1.8e+04	1.9e+04	6.9e+03	2.1e+04	2.5e+04	2.4e+04	8.5e+03	3.4e+03	2e+04	6.8e+03
Kiwalan Area 3	1.7e+04	1.7e+04	1.1e+04	2e+04	2.3e+04	2.3e+04	6.4e+03	6.4e+03	1.9e+04	5.2e+03
Luinab	1.2e+04	1.3e+04	1.1e+04	1.5e+04	1.9e+04	1.8e+04	1.7e+03	6.2e+03	1.4e+04	2.9e+03
Mahayhay Area 1	7.2e+03	7.7e+03	1.3e+04	1e+04	1.4e+04	1.3e+04	5.3e+03	8.3e+03	9.1e+03	5.5e+03
Mahayhay Area 2	7.2e+03	7.6e+03	1.4e+04	1e+04	1.4e+04	1.3e+04	6.8e+03	9.4e+03	9e+03	6.8e+03
Maria Cristina Area 1	5.9e+03	6.4e+03	2.1e+04	2e+03	5.6e+03	5.2e+03	1.4e+04	1.6e+04	9.7e+02	1.4e+04
Maria Cristina Area 2	1.2e+04	1.3e+04	2.7e+04	6.7e+03	1e+04	9.6e+03	2e+04	2.2e+04	5.5e+03	2e+04
Palao Area 1	8.8e+03	9.3e+03	1.3e+04	1.2e+04	1.5e+04	1.5e+04	4.7e+03	8.2e+03	1.1e+04	4.9e+03
Palao Area 2	8.4e+03	8.9e+03	1.3e+04	1.1e+04	1.5e+04	1.4e+04	5.2e+03	8.6e+03	1e+04	5.4e+03
Poblacion	8e+03	8.5e+03	1.2e+04	1.1e+04	1.4e+04	1.4e+04	5.1e+03	7.7e+03	9.9e+03	5e+03
San Miguel	8.7e+03	9.2e+03	1.1e+04	1.2e+04	1.5e+04	1.5e+04	3.8e+03	6.5e+03	1.1e+04	3.8e+03
San Roque	1.3e+04	1.3e+04	9.5e+03	1.6e+04	1.9e+04	1.9e+04	1.3e+03	5e+03	1.5e+04	1.8e+03
Santa Elena	5.9e+03	6.4e+03	1.8e+04	8.8e+03	1.2e+04	1.2e+04	1.1e+04	1.3e+04	7.8e+03	1.1e+04
Santa Filomena Area 1	1.7e+04	1.8e+04	6.3e+03	2e+04	2.4e+04	2.3e+04	7.4e+03	2.3e+03	1.9e+04	5.7e+03
Santa Filomena Area 2	1.4e+04	1.5e+04	6.6e+03	1.7e+04	2.1e+04	2e+04	4.6e+03	2.1e+03	1.6e+04	2.9e+03
Santa Filomena Area 3	1.3e+04	1.3e+04	7.7e+03	1.6e+04	1.9e+04	1.9e+04	3.2e+03	3.3e+03	1.5e+04	1.5e+03
Santiago	9.9e+03	1e+04	1.1e+04	1.3e+04	1.6e+04	1.6e+04	3.9e+03	6.6e+03	1.2e+04	3.9e+03
Santo Rosario Area 1	1.1e+04	1.1e+04	1.1e+04	1.4e+04	1.7e+04	1.7e+04	3.4e+03	6.1e+03	1.2e+04	3.4e+03
Santo Rosario Area 2	1e+04	1.1e+04	1.1e+04	1.3e+04	1.7e+04	1.6e+04	3.4e+03	6.1e+03	1.2e+04	3.4e+03
Santo Rosario Area 3	9.6e+03	1e+04	1.1e+04	1.2e+04	1.6e+04	1.6e+04	3.4e+03	6.1e+03	1.1e+04	3.4e+03
Saray-Tibanga	8.9e+03	9.4e+03	1.2e+04	1.2e+04	1.5e+04	1.5e+04	4.6e+03	7.3e+03	1.1e+04	4.6e+03
Suarez	3.7e+02	8.6e+02	1.9e+04	7.6e+03	1.1e+04	1.1e+04	1.2e+04	1.5e+04	6.6e+03	1.2e+04
Tambacan	5.9e+03	6.4e+03	1.4e+04	8.7e+03	1.2e+04	1.2e+04	7e+03	9.7e+03	7.7e+03	7e+03
Tibanga	9e+03	9.5e+03	1.1e+04	1.2e+04	1.5e+04	1.5e+04	3.8e+03	6.4e+03	1.1e+04	3.7e+03
Tipanoy	1.1e+04	1.2e+04	1.9e+04	1.4e+04	1.8e+04	1.7e+04	1.1e+04	1.4e+04	1.3e+04	1.1e+04
Tominobo Proper Area 1	4.4e+03	4.8e+03	1.6e+04	7.2e+03	1.1e+04	1e+04	8.5e+03	1.1e+04	6.2e+03	8.5e+03
Tominobo Proper Area 2	5e+03	5.5e+03	1.6e+04	7.8e+03	1.1e+04	1.1e+04	9.1e+03	1.2e+04	6.8e+03	9.1e+03
Tubod Area 1	6.4e+03	6.8e+03	1.4e+04	9.2e+03	1.3e+04	1.2e+04	6.8e+03	9.5e+03	8.2e+03	6.8e+03
Tubod Area 2	6.4e+03	6.9e+03	1.5e+04	9.3e+03	1.3e+04	1.2e+04	7.7e+03	1e+04	8.3e+03	7.6e+03
Ubaldo Laya	8.6e+03	9.1e+03	1.5e+04	1.1e+04	1.5e+04	1.5e+04	7.3e+03	1.1e+04	1e+04	7.4e+03
Upper Hinaplanon	1.5e+04	1.6e+04	1.2e+04	1.8e+04	2.2e+04	2.1e+04	3.6e+03	7.6e+03	1.7e+04	4.3e+03
Villa Verde	7.3e+03	7.7e+03	1.2e+04	1e+04	1.4e+04	1.3e+04	5.3e+03	7.9e+03	9.1e+03	5.3e+03

Figure 2.4 - 24. **Distance Heat Map from every Pumping Station to every Barangay. The cell of the intersection of these two axes is the distance of the water tube value. This visual shows where the ideal pumping station should a barangay area be connected in terms of distance. Ideally, the barangay should be connected to the nearest pumping station. Longer distance water tubes means more prone to leakages.**

Only around 25% of the Barangay areas are connected to their ideal pumping station in terms of distance. For Instance, Dalipuga Area 2 is currently connected to San Roque Pumping Station through an 8 kilometers water tube while it is closer to Anahawon Pumping Station by an estimated 19 meters only.

Conclusions and Recommendations

With eight watersheds in Iligan City, the demand of water to connected constituents is expected to be sufficient. In fact, the water supplied by the ICWS is more than the daily needs of some of the connected barangays.

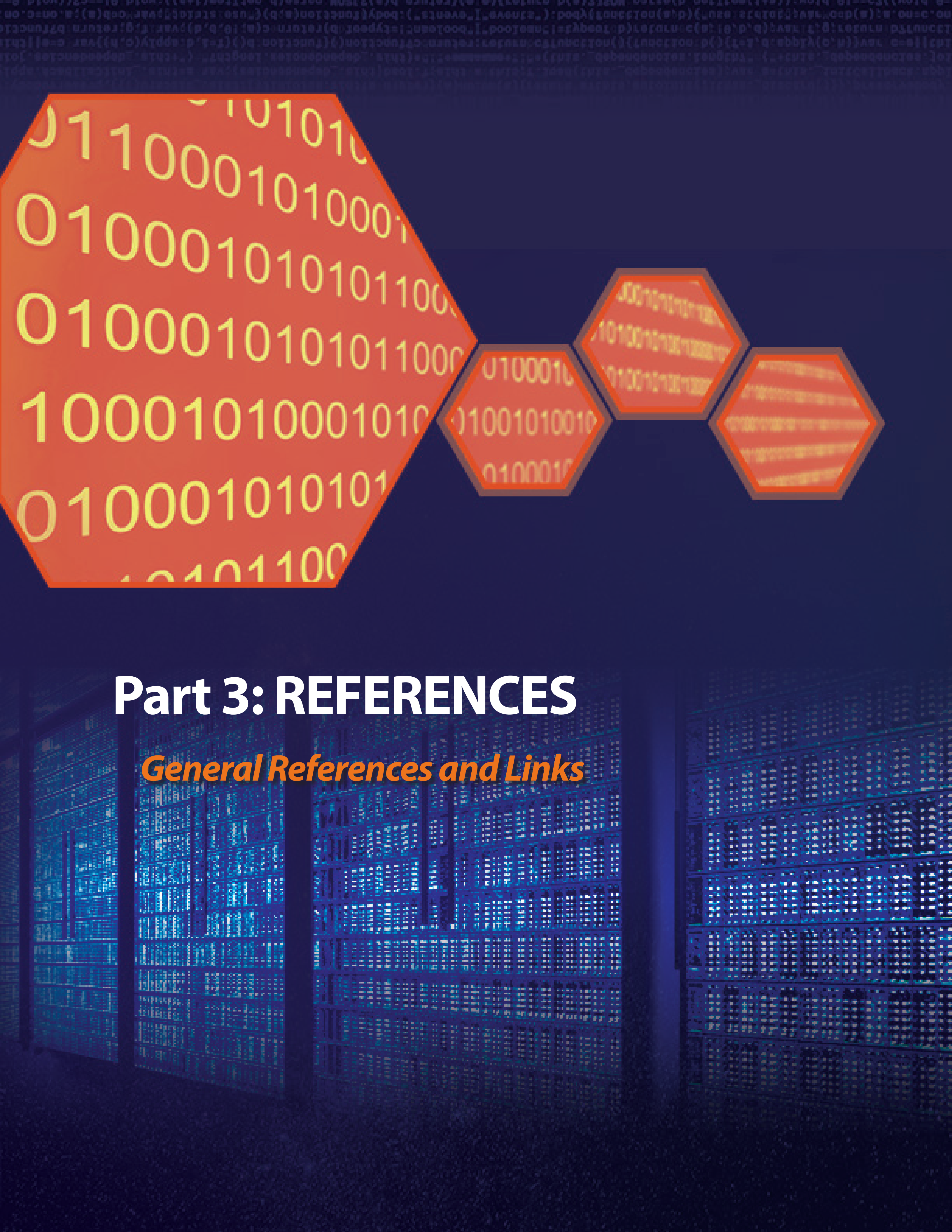
However, one cause of the shortage of water in other areas is accounted due to non-revenue water, caused by old pipes and leaking pipes. There are eleven barangays that are not connected to and supplied by ICWS while there are two barangays that have more ICWS connections than the number of households in those barangays. Further, ICWS may have not considered the proximity of barangays to that of the source of water or the location of pumping sites.

To address the concerns about the shortage of water in Iligan City, a discreet analysis of the demand and supply of water must be considered. Open discussions and forums must be conducted between ICWS and constituents to evaluate whether water demands on the ground are met. In addition, repair and maintenance of connections (pipes) should be addressed to lessen non-revenue water while improving the connections is also highly recommended.

The Local Government Unit of Iligan City may also consider crafting policies that will open the distribution of water to private entities. The LGU can also venture into a public-private partnership to strengthen ICWS. Further, the city should craft a new water master plan given the growth of population and business industries, and the increase in water demand. As water is a basic need, hinterland barangays should also have access to clean potable water.

A further study on Iligan City's water should be conducted to validate the data. LGU can partner with institutions such as the academe for data analysis and investigations.

However, water security is "the capacity of a population to safeguard sustainable access to adequate quantities of and acceptable quality water for sustaining livelihoods, human well-being, and socio-economic development, for ensuring protection against water-borne pollution and water-related disasters, and for preserving ecosystems in a climate of peace and political stability" (UN Task Force on Water Security, 2013), every Iliganon should be involved in ensuring the city's water security that leads to having access to acceptable quality water. This can be done through multi-sectoral approaches and engagements. Recently, the Local Government Unit of Iligan City in partnership with MSU-IIT organized a Water Dialogue with the theme "Sustaining Iligan City's Water, Improving Lives of Iliganon." This is a good start and should continue as the LGU recognized that they alone cannot solve the issue of water. They have to involve as many people/sectors as they can because the issue of water is too big to ignore.



Part 3: REFERENCES

General References and Links

3.0 References

Copeland, B.J. (2022) "Artificial Intelligence", Britannica. Available at: https://www.britannica.com/technology/artificial-intelligence (Accessed: 18 Sep 2022).

IBM (2021) "CRISP-DM Help Overview", IBM Documentations. Available at: https://www.ibm.com/docs/en/spss-modeler/saas?topic=dm-crisp-help-overview (Accessed: 18 Sep 2022).

IBM Cloud Education (2020) "Machine Learning", IBM Cloud Learning Hub. Available at: https://www.ibm.com/cloud/learn/machine-learning (Accessed: 18 Sep 2022).

Merriam Webster Dictionary (no date) Data. Available at: https://www.merriam-webster.com/dictionary/data (Accessed: 18 Sep 2022).

Merriam Webster Dictionary (no date) Dashboard. Available at: https://www.merriam-webster.com/dictionary/dashboard (Accessed: 18 Sep 2022).

Stobierski, T. (2021) "What's the difference between data analytics & data science?", Harvard Business School Online. Available at: https://online.hbs.edu/blog/post/data-analytics-vs-data-science (Accessed: 18 Sep 2022).

The World Bank (no date) "Open Data Essentials", Open Data Toolkit. Available at: http://opendatatoolkit.worldbank.org/en/essentials.html (Accessed: 18 Sep 2022).

The World Bank (no date) "Open Data in 60 Seconds", Open Data Toolkit. Available at: http://opendatatoolkit.worldbank.org/en/open-data-in-60-seconds.html (Accessed: 18 Sep 2022).

email: support@layertechlab.com

http://www.layertechlab.com

LAYER
_TECH

OCDex