Tampere University

Erick Rico Esparza

COMPARATIVE ANALYSIS OF DRINKING WATER TREATMENT PLANTS IN MEXICO CITY

YEB.250 Working Life Project in Environmental Engineering
Supervisor: Chunyan Tang
24.04.2025

CONTENTS

.

1. INTRODUCTION

Having access to safe drinking water is crucial for public health and is widely recognized as a basic human right[1]. This issue is especially important in big cities where lots of people live closely together, and the systems that clean water must work hard to make sure everyone gets enough clean water to drink. In such crowded places, having reliable water treatment systems is essential not only for keeping people healthy but also for helping the city deal with challenges like pollution and water shortages. Managing these water systems well is very important because any problems can quickly affect a lot of people. In Mexico City, the water that comes from the tap is directly not drinkable but it still exists the clear need for an efficient water treatment system as it tries to adjust to the needs of its rising population while at the same time dealing with environmental and logistical challenges.

Mexico City, which is one of the largest cities in the world, showcases the complex issues involved in managing water in a sustainable way for a huge urban population. With a growing population of almost 10 million residents—and over 20 million including the metropolitan area—, the city heavily relies on its many water treatment plants[2][3]. These plants are crucial for providing the city's water needs and making sure that the water supply is clean and safe. Recent news articles from 'El Universal' and 'El Economista' have pointed out big government efforts to improve the city's water treatment systems. These efforts are part of a bigger plan to fix the city's long-term water shortage problems by upgrading old systems to better meet the needs of its people. The data for this study offers detailed information on the status of 41 water treatment facilities in the city, including their processes, how much water they can handle, and how much they actually process, among other important details[4][5].

The present study aims to conduct a detailed comparative analysis of the operational capacities and technological efficacy of water treatment plants across Mexico City, utilizing governmental open data from 2023, to particularly focus on:

- Evaluating the operational efficiency of the plants by comparing installed capacities and actual flow rates, thus assessing how well each one meets the demands;

- Analyzing the effectiveness of several treatment processes employed by these plants in terms of their operational performance, to ultimately identify technological strengths and weaknesses;
- And detecting strategic opportunities for infrastructure improvement and the optimal distribution of resources to improve the efficiency and sustainability of the city's water treatment system.

To achieve these goals, this study is guided by some questions about the distinctions of water treatment efficacy in the city, that will be answered throughout this study:

1. What are the main operational differences between Mexico City's water treatment plants, and how do these differences affect their efficiency?
2. Which water treatment processes are the most efficient in terms of capacity utilization and flow rate treated, and what does this suggest about the current state of water management technology in use?
3. How can the findings from this study advise about improvements in infrastructure and resource allocation to develop the functionality of water treatment plants in Mexico City?

The point of this study is based on the clear need for clean water, which is basic to prevent diseases and support economic and social growth. Therefore, to have clean water easily accessible is key for healthy communities, particularly in developing areas where it can significantly improve living conditions. Enhancing water quality can lead to better education and more job opportunities, which helps whole communities thrive. As mentioned, this work points out where changes in water management can be most effective, contributing to worldwide efforts to improve health and development[6].

Finally, it is important to remark that this study combines international health standards with detailed local data, creating a good approach that mixes data analysis with a detailed examination of current water management practices. It not only follows but seeks to apply the WHO's guidelines, which call for detailed strategies to ensure water is safe from the source to the tap[1]. Ultimately, this work will try to identify key areas for improvement, ensuring that all city residents have access to clean and safe water.

2. REVIEW OF DRINKING WATER TREATMENT PROCESSES AND COMPARATIVE STUDIES

This chapter reviews the scientific and technical literature on drinking water treatment. It discusses the definition and function of water treatment plants, the main treatment processes and technologies used to secure safe water, relevant studies from both global and local perspectives, and comparative analyses from similar urban settings. The review sets the foundation for the methodological approach in Chapter 3.

2.1 Definition and Function of Drinking Water Treatment Plants

These sorts of treatment facilities are carefully designed to remove contaminants from water so that it meets health and safety standards for human consumption. According to the CONAGUA's website, a water treatment plant ("planta potabilizadora" in Spanish) processes raw water from various sources by applying a series of treatment steps to make sure that this vital liquid is harmless for public use[7]. Similarly, the Centers for Disease Control and Prevention (CDC) describe the treatment process as a series of steps—including filtration, disinfection, and storage—that work together to deliver clean drinking water[8]. Figure 1 presents a simplified overview of the key steps in a typical drinking water treatment process, from the initial extraction of water to its final distribution for human consumption.

Figure 1. A basic illustration of the water supply process, showing major treatment steps[9].

These plants serve as a crucial link between raw water sources and the communities that rely on them. They are essential not only for preventing waterborne diseases but also for supporting long-term environmental sustainability. By effectively removing pathogens, chemicals, and other contaminants, water treatment plants protect public health, reduce the risk of epidemics, and ensure that water remains a reliable resource for agriculture, industry, and daily domestic use.

2.2 Water Treatment Processes and Technologies

The technologies employed in water treatment are varied, each with its own strengths and limitations. Some treatment processes that exist are adsorption, direct filtration, and reverse osmosis. For example, adsorption involves the removal of dissolved substances through adhesion to a solid surface, while direct filtration physically removes suspended particles. Reverse osmosis, on the other hand, uses a semi-permeable membrane to eliminate both dissolved and particulate impurities[6].

Advances in disinfection technology have been highlighted in recent research. Ngwenya et. al. (2013) explain how modern disinfection techniques have evolved to address both traditional pathogens and emerging contaminants[10]. These advances include improvements in chemical disinfectants, as well as non-chemical methods like ultraviolet light, which are increasingly being adopted due to their efficiency and lower formation of harmful by-products.

In addition, many studies emphasize the role of international guidelines—such as those from the World Health Organization (WHO)—in shaping water treatment practices[1]. These guidelines set out recommended limits for disinfectant residuals and contaminant levels, ensuring that treatment technologies maintain a consistent quality standard worldwide. So, by offering a global standard, this framework also inspires national and local authorities to develop policies that align with scientific recommendations, which inevitably would help minimize regional disparities in water quality, given that utilities across different countries adopt comparable targets for pathogen removal, chemical usage, and distribution network maintenance. In practice, these guidelines can serve as a mechanism for technological upgrades, operator training, and collaborations among different levels of governments, aimed at improving water treatment infrastructure.

2.3 Synthesis of Empirical Studies on Drinking Water Treatment and Management

2.3.1 International and Regional Trends in Water Treatment Technologies and Management

A wide range of research has focused on the effectiveness of water treatment processes and the challenges faced by water utilities around the world. Several global studies have examined the successes and limitations of, for example, various methods of disinfection. For instance, data-driven analyses have become more popular as researchers aim to predict water quality based on historical data. The survey that was made by Kang et. al. (2017) discusses how big data approaches and how machine learning models are used to evaluate water quality more accurately[11].

Another review, (Aliashrafi et. al., 2021) provides insights into how data-driven models help operators adjust treatment parameters to improve water quality[12]. These studies point out the importance of merging old and traditional water treatment processes with state-of-the-art data analytics to optimize performance and predict potential issues before they occur. Furthermore, research on drinking water infrastructure and environmental disparities has revealed that differences in infrastructure can lead to uneven water quality outcomes, particularly affecting low-income and minority communities[13]. This body of work demonstrates that while technological advances are critical, the management and equitable distribution of resources also play a key role in water quality.

2.3.2 Investigations Specific to Mexico City

Research focusing specifically on Mexico City highlights the unique challenges that this megacity faces. Silva (2024), for instance, proposes a model designed to address the issues of aging infrastructure, limited resources, and coordination challenges among municipal authorities[14]. This study shows that even though Mexico City has a robust economy, its water supply system struggles to meet demand efficiently due to various operational setbacks.

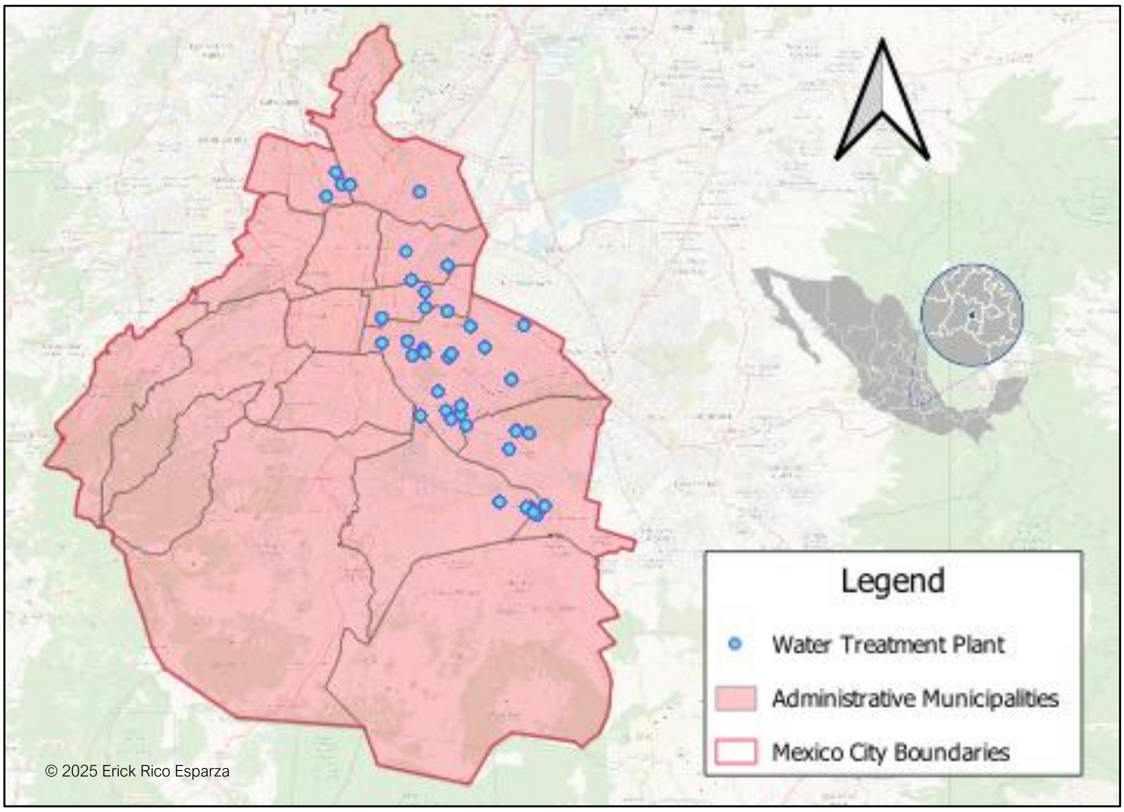

Figure 2. Approximate location of water treatment plants in Mexico City (blue dots), shown alongside municipality boundaries. Own data adaptation in QGIS from SHP files[15][16].

Additionally, Silva et. al (2022) offers insights into how different cities handle similar water challenges. This research points out key differences in governance, financial frameworks, and resource distribution, indicating that some of Singapore's strategies—such as integrated planning and strong public–private partnerships—could be adapted to improve Mexico City's water sector[17]. The authors also highlight how Singapore's consistent investment in research and infrastructure has inevitably led to more efficient operations and higher water quality standards. In contrast, Mexico City's fragmented management structure, corruption, budget limitations, and political interests often slow down reforms and technological upgrades. Although numerous studies have examined water treatment in Mexico City, there is still a noticeable gap in comparative work that looks specifically at ways to combine local priorities with global best practices. Bridging this gap, so to speak, may help local policymakers refine strategies, ensuring that lessons from cities like Singapore are applied in a way that respects Mexico City's unique social, and economical landscape.

2.4 Comparative Analysis in Similar Contexts

Comparative studies are valuable because they provide a broader perspective on water management practices. For instance, the already mentioned comparative research between Mexico City and Singapore reveals how different governance models, technological implementations, and policy frameworks impact the overall efficiency of water treatment systems[17]. Singapore's success in achieving a high standard of water safety and reliability—even with limited natural water resources—offers lessons that could be useful for Mexico City.

These studies discuss various methodologies, such as qualitative assessments and data-driven evaluations, to compare the efficiency of water treatment processes in different urban settings. They highlight that while both cities face challenges, the strategies used to address these challenges differ significantly. Singapore's approach, which emphasizes coordinated decision-making and robust infrastructure investment, contrasts with the more fragmented management observed in Mexico City. This difference points to the need for a more integrated and data-informed management system in Mexico City.

Comparative analyses also draw attention to environmental disparities and the need for equitable water distribution. The research introduced earlier on water infrastructure and disparities shows that gaps in technology and resource allocation can lead to uneven water quality outcomes among different socioeconomic groups[13]. These findings underscore the importance of not only improving technology but also ensuring that policy and management practices are aligned to provide equal access to safe water for all residents.

In summary, the literature reveals a complex landscape where advanced treatment technologies and data-driven models are emerging as key tools to improve water quality. However, significant challenges remain in managing infrastructure and ensuring equitable water distribution—especially in large, diverse urban centers like Mexico City. This review has highlighted both global trends and local challenges, establishing a strong foundation for the next section. In Chapter 3, the methodology will build on these insights by describing a data-driven approach using Python to analyze the operational efficiency of Mexico City's water treatment plants, compare treatment processes, and identify opportunities for improvement.

3. METHODOLOGY

3.1 Dataset Description and Geographic Context

The primary dataset used in this study comes from official open-data platforms corresponding to the year 2023[18]. It includes information about 41 water treatment plants located in various municipalities of Mexico City. Each record in the dataset provides details such as the name of the plant, its installed capacity, the type of purification processes it employs, and the corresponding hydrological administrative region.

Mexico City is divided into 16 municipalities—also known as “alcaldías” or “delegaciones” (see *Figure 3*). These municipalities are administered under the 13th Hydrological Administrative Region, referred to as “Waters of the Valley of Mexico,” which covers the city’s wider metropolitan area (see *Figure 4*). Additionally, an important thing to note for context is that since Mexico City is the nation’s capital, the management of its water resources falls under both local and federal authorities.

Figure 3. Mexico City’s 16 municipalities[19].

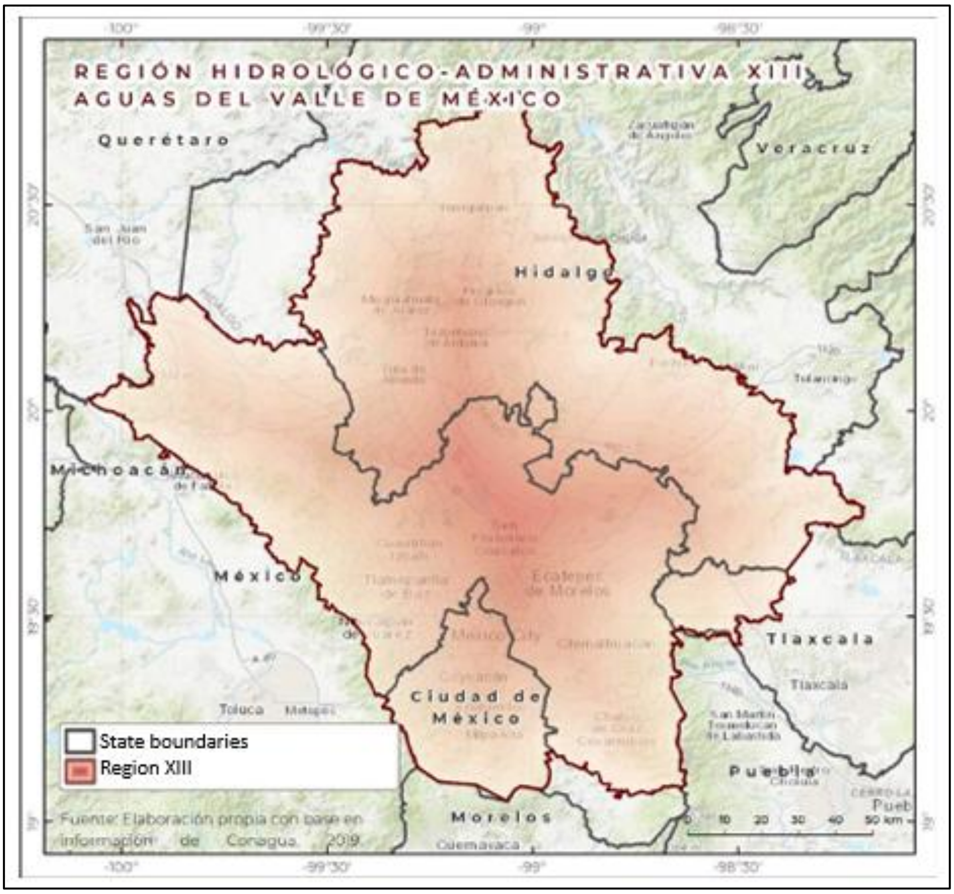

Figure 4. Hydrological Administrative Region XIII: Mexico Valley. Own data adaptation from CONAGUA[20].

These maps provide a clear visualization of Mexico City's municipal boundaries and its placement within the larger Hydrological Administrative Region XIII. They offer valuable geographical context for understanding the city's water management challenges and illustrate the broader regional interconnections influencing local water treatment and resource allocation efforts in the city.

3.2 Data Dictionary

Table 1 presents the data dictionary for the variables used in this study, offering a concise guide to the dataset's structure. The first column lists the original Spanish variable names as they appear in the dataset, while the following columns clarify each field's meaning and, where applicable, specify the expected data type or unit of measurement. This data dictionary serves as a crucial reference point throughout the analysis, helping maintain consistency, accuracy, and a clear understanding of all variables involved.

Table 1. Data Dictionary[18].

Column Name	Description	Additional Notes
id	Key identifier	Unique numeric or text identifier for each record.
anio	Year of the purification process	This dataset corresponds to 2023. Data type: integer (e.g., 2023).
planta	Water Treatment Plant	Name of the water treatment plant. Mexico City has 41 plants.
rha	Hydrological Administrative Region	Mexico has 13 hydrological regions, with Region XIII ("Waters of the Valley of Mexico") covering Mexico City. Data type: text or number (e.g., XIII).
estado	State	Mexico has 32 states; for this dataset, the state is "Ciudad de México."
municipio	Municipality	Mexico City is divided into 16 municipalities (e.g., Álvaro Obregón, Iztapalapa, Xochimilco, etc.). Data type: text.
proceso	Process carried out in the water treatment plant	In Mexico City, the main processes are Adsorption, Direct Filtration, and Reverse Osmosis. Data type: categorical (text).
capacidad	Installed water treatment capacity	Measured in liters per second (L/s).
caudal	Drinking water flow rate	Also in L/s. Reflects the actual volume of treated water.

3.3 Software and Tools

For this analysis, Python 3 running on Google Colaboratory was used[21]. The environment provided an interactive workspace with a user-friendly interface and the ability to easily share code and results. Only two libraries were used for data manipulation and visualization: *pandas*, and *matplotlib.pyplot*. These tools allowed for efficient handling of the dataset and the creation of clear and informative graphs that illustrate trends and relationships among the water treatment plants.

Figure B1 in *Appendix B* shows a code snippet that demonstrates how the libraries were imported in the Google Colab environment. In this snippet, it is possible to observe that each library has been imported with an alias—*pandas* as *pd,* and *matplotlib.pyplot* as *plt*. Using aliases helps simplify the code, making it easier to call functions from these libraries without having to write out their full names each time.

3.4 Data Import, Exploration, and Preprocessing

Firstly, the data file obtained from the Mexican government open data site[18] was manually adjusted in Excel to remove unnecessary columns and rows. Next, the cleaned file was uploaded to Google Colab using the file uploader feature. Once in the environment, it was imported into a *pandas DataFrame* with the *pd.read_csv()* function. *Figure B2* in *Appendix B* shows the code used in this process.

Now, in such figure, *df* represents the *DataFrame* containing the dataset. As abovementioned, this dataset is for 2023 and originally covered all water treatment plants in the country. However, after preprocessing, only the important data was retained: the 41 rows corresponding to the drinking water treatment plants located within Mexico City's limits.

After importing the dataset, the command *df.head()* was used to display the first few rows (refer to *Figure B3* in *Appendix B*). This step was performed to verify that the data was correctly loaded and that all columns—such as those described in the Data Dictionary (*Section 3.2*)—appear as expected. Furthermore, to ensure data integrity, missing or null values were checked using the command *df.isnull()*. Because the dataset contains 41 rows, the first 5 and the last 5 rows using *pd.concat([df.isnull().head(5), df.isnull().tail(5)])* were combined to create a concise visual summary (see *Figure B4* in *Appendix B*). Although only these 10 rows are shown, to confirm that all cells returned "*False*," the complete output was manually reviewed ensuring there were no missing values.

3.5 Analytical Approach

This analysis was structured around three main objectives: evaluating the operational efficiency of water treatment plants in Mexico City, comparing the performance of different treatment processes, and identifying areas where infrastructure and resource distribution could be improved.

To begin the analysis, basic statistical measures such as average (mean), minimum, maximum, and percentage efficiency, were used. These were calculated for key variables including: the installed capacity (liters per second), the actual flow rate (liters per second), and the operational efficiency (flow rate divided by capacity, expressed as a percentage). These indicators helped to understand how well each plant was performing. For example, some plants are working close to their maximum capacity, while others are significantly underused. A comparison was also made using the values by municipality and by treatment process. This allowed to better understand the main operational differences across the city and to begin answering the initial research questions.

Now, to identify trends and patterns, several visualizations were created: a bar chart showing operational efficiency for each plant, another one comparing average efficiency by treatment process, and a final one comparing average efficiency by municipality. These graphs helped to visualize which plants are operating efficiently and which ones may need attention. After this, it is possible to see how different technologies perform and how plant performance varies across the city.

Additionally, as an extra step, since it already exists a comparison of mean efficiency across three groups (treatment processes—a categorical variable), and the operating efficiency is also existing (in percentage—a dependent variable), a basic statistical analysis was used. In this case, the one used to enrich the results was ANOVA (Analysis of Variance). This sort of test was quite useful for determining whether there are statistically significant differences between the means of more than two groups. This analysis complemented the one for the second objective, but now not only descriptively (averages), but also by examining whether the observed differences between processes are statistically significant or could simply be due to chance.

Moreover, since the dataset does not include geographic data like coordinates or shapefiles, mapping tools such as *geopandas* or *folium* were not utilized; however, online GIS tools were used to interpret the results, enrich the discussion, and ultimately support the final conclusions.

On a side note, it is important to mention that in this project, other and more advanced statistical models were not used. The dataset used is relatively small (41 plants), and the relationships between variables are quite straightforward. For example, efficiency is directly calculated from flow rate and capacity. The methods used were appropriate because the research questions focused more on understanding and comparing performance. Also, the combination of implementing a basic ANOVA test, descriptive statistics and clear visualizations already provided useful insights.

3.6 Study Limitations

This study has a few limitations that should be considered when interpreting the results. Like many data-driven studies, the quality and scope of the findings depend directly on the availability and completeness of the data.

First, as mentioned, the original dataset included all water treatment plants in Mexico. However, due to the large amount of information, it was necessary to manually reduce the dataset in Excel. Only the plants located in Mexico City were kept, and several columns that were not relevant to this analysis were removed. While this helped make the dataset easier to manage and focus on the local context, it also means that comparisons with other states or regions could not be made.

Second, the data used is from 2023, which is the most recent version publicly available. Although almost two years have passed since this was last updated, the dataset still provides a valuable snapshot of the situation and can help identify patterns and trends that may still be relevant today. However, it is important to note that some conditions may have changed, and future research should include more recent data when it becomes available.

Finally, the dataset does not include certain variables that could have enriched the analysis, such as geographic coordinates, real-time operational data, or financial indicators. These types of data could support more advanced analyses, like geospatial mapping or cost-efficiency evaluations. Their absence limits the scope of this study, but the findings remain useful for understanding the current state of water treatment infrastructure in the city.

4. RESULTS AND DISCUSSION

4.1 Data Analysis and Results

This section presents the findings from the dataset described in *Chapter 3*, linking each subsection to one of the three objectives defined in *Section 1* of this study, alongside a fourth subsection related to the ANOVA test.

4.1.1 Evaluating Operational Efficiency of Plants

To begin the analysis, the operational efficiency of each water treatment plant in Mexico City was examined by comparing its installed capacity and actual flow rate. A new variable was calculated, showing how much of each plant's potential is currently being used. This helped identify which plants are working close to their maximum capacity and which are underutilized. The analysis was also useful to detect potential imbalances in the way resources are distributed across the city.

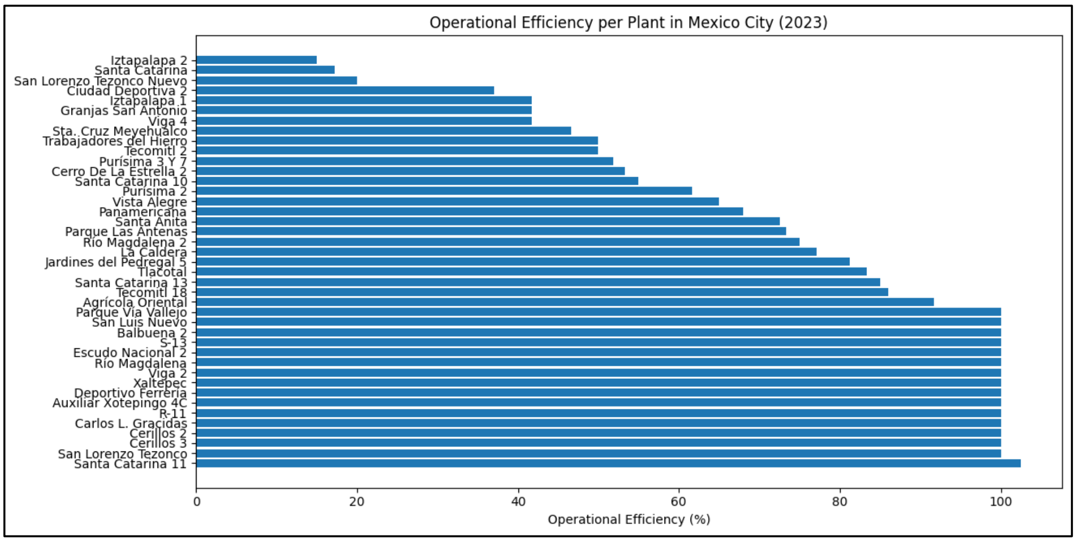

Figure 5. Operational Efficiency per Plant in Mexico City (2023).

The results in the chart above (refer to *Table A1* in *Appendix A* to see more details) showed that some plants operate at or near 100% of their capacity, while many others work at less than 50%. This indicates a clear variation in plant performance. Some municipalities seem to manage their water treatment resources more efficiently, while others may face operational challenges. These differences can help inform decisions about maintenance, reallocation of resources, or future infrastructure planning.

4.1.2 Analyzing Efficiency of Treatment Processes

To understand how each water treatment process performs, three groups of plants were created by type of treatment: Direct Filtration, Reverse Osmosis, and Adsorption. We then calculated the average operational efficiency for each group. This allowed to compare how different technologies are being used in practice and whether some processes are more efficient than others in real-life conditions.

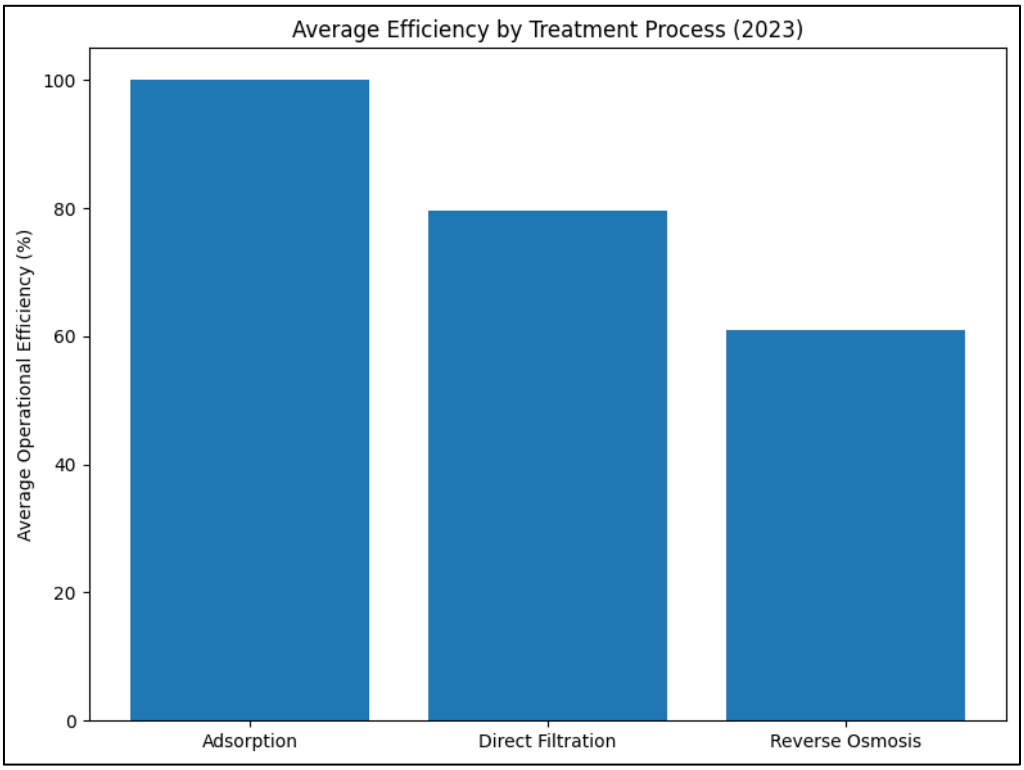

Figure 6. Average Efficiency by Treatment Process (2023).

The bar chart shows the average operational efficiency of water treatment plants in Mexico City, grouped by treatment process. *Adsorption* has the highest average efficiency (100%), but it is only used in one plant, so this result cannot be generalized. *Direct Filtration*, used in 27 plants, has an average efficiency of around 80%, making it the most commonly used and generally reliable process. *Reverse Osmosis*, used in 13 plants, shows the lowest average efficiency, approximately 61% (see *Table A2* in *Appendix A* for more details). This may be due to its technical complexity, higher energy consumption, or maintenance requirements. Although Direct Filtration and Adsorption seem more efficient, it's important to consider that factors such as water quality, location, and infrastructure conditions may affect performance. These results suggest that Direct Filtration is the most stable and scalable process in the current system, while Reverse Osmosis may require operational improvements.

4.1.3 Identifying Strategic Opportunities for Improvement

To explore where infrastructure improvements or better resource distribution might be needed, groups of plants were created by municipality. Thus, it was possible to visualize the number of plants in each area, their total capacity and flow rate, and the average operational efficiency. This helped identify which municipalities may need more support or where there is already good performance that could serve as a model for others.

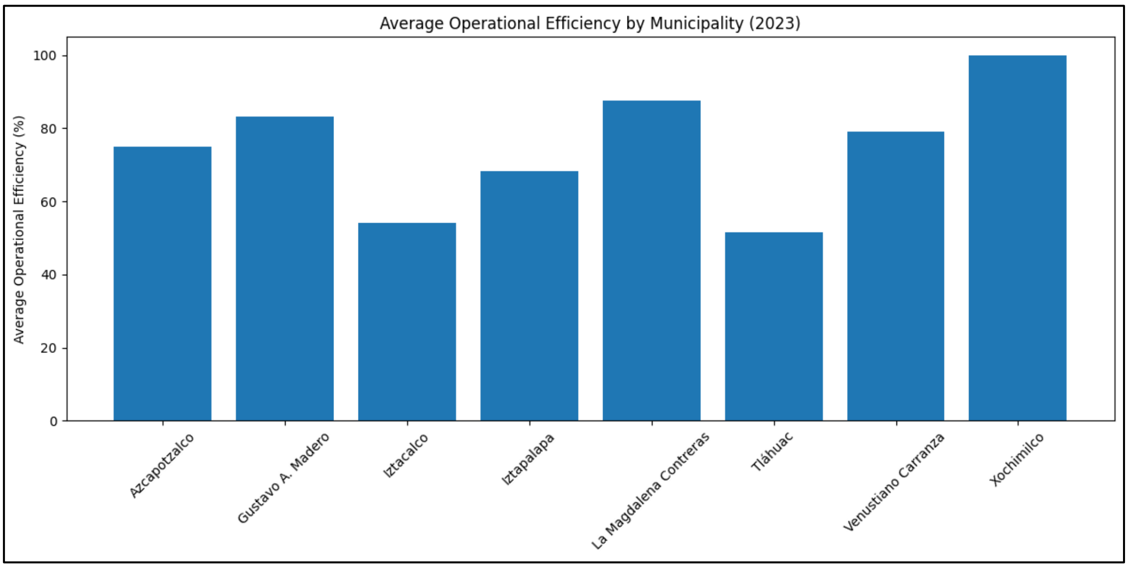

Figure 7. Average Operational Efficiency by Municipality (2023).

The data showed both in *Figure 7* and *Table A3* in *Appendix A* that some municipalities, like Xochimilco and La Magdalena Contreras, have high average efficiency despite having fewer plants. Others, like Iztapalapa, have way more plants but lower efficiency. These results suggest that some areas may benefit from operational adjustments or additional investment. On the other hand, municipalities with high efficiency could share their practices or maybe even be used as benchmarks.

4.1.4 Significance Analysis for Treatment Process Efficiency

Finally, to better understand the differences in efficiency between treatment processes, a one-way ANOVA test was conducted along with a boxplot. The goal was to explore whether the average operational efficiency varies significantly depending on the treatment technology used. The ANOVA test helps identify if the differences in group averages are statistically meaningful, while the boxplot provides a visual way to compare how efficiency values are distributed within each group.

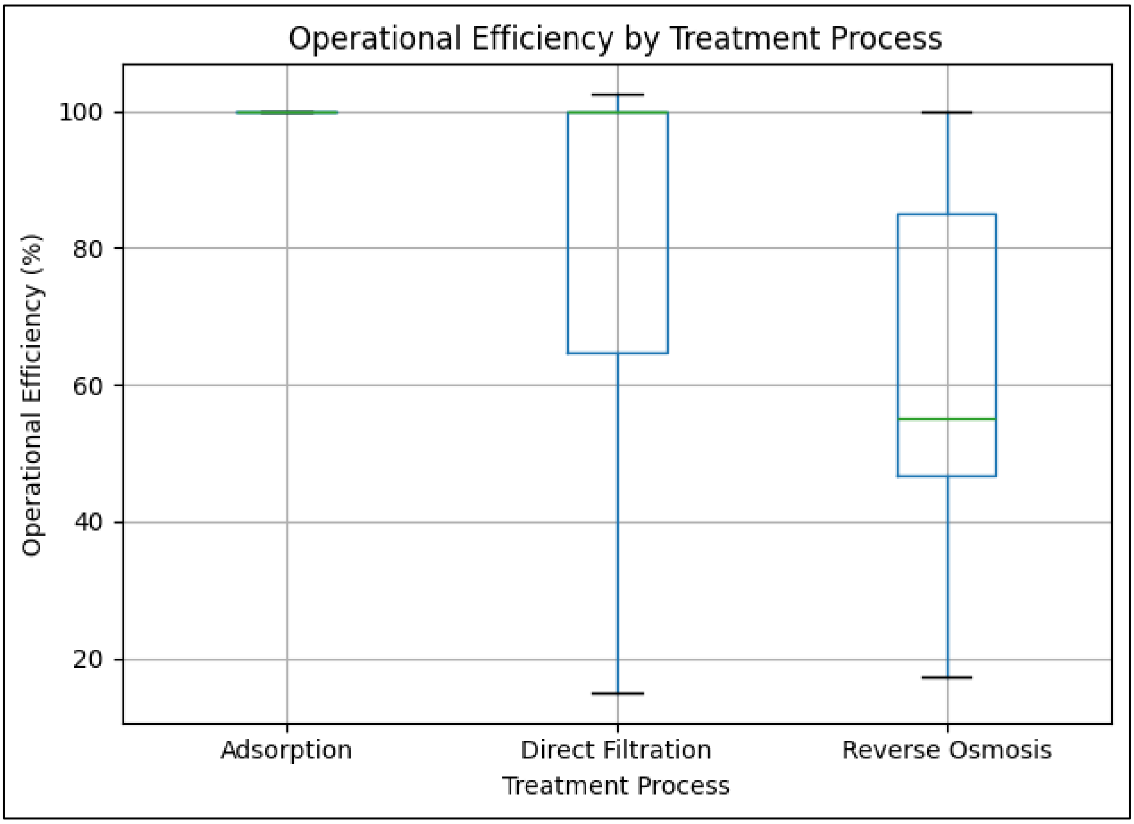

Figure 8. Operational Efficiency by Treatment Process in Boxplot Format.

The results of the ANOVA test showed an F-value of 2.77 and a p-value of 0.076 (see more in *Table A4* in *Appendix A*). Since the p-value is slightly above the standard threshold of 0.05, it is not possible to conclude that the differences between treatment processes are statistically significant. However, the boxplot (*Figure 8*) revealed interesting patterns. Plants using *Adsorption* reached 100% efficiency but had very few data points. *Direct Filtration* had a high median efficiency close to 100%, but also showed high variability, with some plants operating far below their potential. *Reverse Osmosis* had the lowest median efficiency and the widest spread, suggesting inconsistency (refer to *Table A5* in *Appendix A*). These visual differences, although not statistically confirmed, suggest that further investigation with a larger dataset is advisable, which ultimately may reveal meaningful distinctions in performance.

4.2 Discussion

The results obtained in Section 4.1 shed light, so to speak, on important trends in both the capacity utilization and technological choices across the 41 water treatment plants in Mexico City. First, regarding the research question 1—the main operational differences and their impact on efficiency—it is quite evident that plants employing Direct Filtration generally register higher operational efficiencies (around 80%) while Reverse Osmosis (RO) plants average roughly 61%. However, despite that apparent difference, the ANOVA test that was carried out did not show statistically significant p-values ($p = 0.076$), indicating that we cannot confirm strong mean differences by process. This statistical distinction aligns with the fact that while descriptive data is able to reveal patterns, problems like overfitting in ML models may arise with small datasets, so larger ones or more advanced modeling are needed often to achieve irrefutable conclusions and thus provide increased and better knowledge[12]. Still, the contrast in average efficiency suggests that operational, or even maintenance variables could be playing a major role.

Additionally, some municipalities—like Xochimilco—consistently reach full capacity utilization. In contrast, others such as Iztapalapa have more plants but are less efficient. The presence of older infrastructure, different local maintenance policies, and varying raw water quality are plausible explanations as VanDerslice (2011) indicates[13]. Particularly, the data also reveals that certain plants using Reverse Osmosis are operating at under 50% of their installed capacity. This could create an interrogation related to cost-effectiveness and resource management, given that Wimalawansa (2013) mentions that RO is an excellent method for dealing with high contaminant loads and generating potable water but still may demand more energy to maintain its advantages[22].

Now, considering research question 2—which water treatment processes appear most efficient in terms of capacity utilization and flow rate—the top performers remain those employing Direct Filtration. This aligns with the suggestion that simpler and well-established methods can show greater consistency across different raw water scenarios if turbidity and other parameters are controlled. On the other hand, the few plants adopting Adsorption (in fact one single plant in the dataset) reached a perfect 100% usage, but the small sample size makes it difficult to generalize. Interestingly, Reverse Osmosis, despite showing variable performance, is highlighted in the literature (Wimalawansa 2013) for its high removal efficiency of dissolved solids and heavy metals[22]. Thus, the lower observed operational efficiency in *Figure 6* might not reflect a failure in technology per se, but a not optimal managerial or maintenance practices—like insufficient staff training or more energy requirements that may cause partial shutdowns. This perspective is also supported by Silva et. al. (2022) indicating that advanced treatment often excels technologically but demands robust financial and institutional frameworks[17]. Another element to interpret is the variation in infrastructure by municipality. Some localities are characterized by older networks or insufficient budgets, factors that definitely impede maximizing each plant expected output in terms of design flowrate. As abovementioned, VanDerslice (2011) focused his research on disparities in infrastructure emphasized the importance of equitable resource distribution to enhance performance[13], which takes on relevance in this case. The data from the Mexican government open data site[18] used in this analysis partially supports that argument, as high-efficiency municipalities might be receiving either more targeted investment or benefit from geographical advantages in raw water supply, but more research would still be needed.

Finally, regarding research question 3—how these findings can guide improvements in infrastructure—two main opportunities arise. Firstly, it becomes essential to reevaluate the role of advanced processes like RO if they remain underutilized. Considering the difficulties of RO plants, authorities might need to improve capacity, budgeting, and operator training. As mentioned in comparative assessments between Mexico City and Singapore, integrated governance and strong public–private partnerships can adopt better maintenance and more stable operations[17]. Secondly, the data reveals that differences in performance might be better tackled by municipality-specific strategies rather than a one-size-fits-all approach, so to speak. Ultimately, bridging these efficiency gaps requires a combination of improvements in local infrastructure, capacity building, and institutional coordination, which is definitely not easy.

In sum, the discussion around these results shows that while there are no statistically conclusive differences across treatment processes in general, there are patterns that emerge from the descriptive analysis. The findings emphasize both the crucial role of proven technologies (like Direct Filtration) and the need for more targeted, and well-funded strategies to unlock the full potential of advanced methods. Furthermore, references from both global[22] and regional[13] studies reinforce that behind the question of efficiency rests an extensive socio-technical system, including strong policy frameworks, staff expertise, and sustainable resource allocation.

5. CONCLUSIONS AND RECOMMENDATIONS

The results from this study underline many messages that, taken together, respond to the goals of evaluating operational efficiency, exploring treatment process effectiveness, and identifying improvement pathways for water treatment plants in Mexico City. From the start, the evaluation of each plant's capacity against actual flow rates allowed to see that many facilities remain underused, while just a few get close or even reach full operational capacity. This diversity of performance levels suggests that neither technology alone nor municipality-specific variables can fully explain the outcomes obtained above. Instead, a closer look at management practices, resource allocation, and even operational protocols is likely to become more relevant for future improvements.

Additionally, in terms of treatment processes, Direct Filtration emerged as the most consistently efficient approach, aligning maybe with a relatively straightforward design. Plants using Reverse Osmosis—often praised for high contaminant removal rates—did exhibit considerable potential, yet their real-world performance remained inconsistent, indicating that plants with these kinds of treatments may require specialized operator training, stable funding for ongoing maintenance, and a clear framework for energy usage. These findings emphasize the link between technical sophistication and the practical realities of implementing such technologies on a certain scale.

Moreover, this analysis suggests that operational differences can be put into context by looking at municipal resource distribution and decision-making. In some areas, historically higher investment in water infrastructure coincides with better performance metrics, while in others, limited budgets or older pipelines impede progress. Yet, the data does not imply that advanced technologies cannot thrive in a less privileged municipality. On the contrary, with adequate institutional support and well-designed programs, even communities with budget restrictions can benefit from improved water quality solutions.

Another cornerstone emerging from this investigation is the significant gap that exists when it comes to connecting data-driven knowledge to everyday decision-making. While the descriptive statistics and basic ANOVA test provided a wide view, more integrated data-driven methods—from real-time sensors to predictive analytics—could further inform operators about optimal coagulant doses, advanced process parameters, or scheduling of maintenance tasks, just to give a few examples. Such steps would need cross-institutional coordination, but the benefits could be important by reducing downtime, preventing system overload, and maintaining high-quality effluent.

Speaking of larger water management strategies, it looks important that any improvement plan incorporate not only technical upgrades but also training programs that build local capacity. Field operators, for example, play such a vital role in adjusting day-to-day controls that affect overall plant efficiency. Thus, establishing or refining standard operating procedures, and ensuring adequate professional development can inevitably lead to more sustainable plant outcomes. Also, it is important to mention that consistent monitoring and transparent reporting may help municipalities adjust their investments where they are most needed, create responsibility, and therefore raise public trust.

From an organizational point of view, good synergy among local government agencies, private partners, and academic institutions can improve the adoption of best practices. Greater and better collaborations can encourage knowledge sharing across different plants, ensuring that smaller or lower-performing facilities benefit from the lessons learned from the most successful ones. Such networks may also help articulate the case for additional funding, focusing on improvements that yield measurable efficiency gains and better water quality for the population.

Finally, the results encourage to reflect on long-term needs: population growth, urban development, and climate variability all place increased demands on water treatment plants. Thus, sustainable solutions will require not only immediate operational improvements but also forward-looking policies that integrate water resources management, land-use planning, and community engagement. So, by proactively investing in both technology and institutional capacity, Mexico City can strengthen the resilience of its water systems, ensuring that all residents have consistent access to safe drinking water. In conclusion, although each plant's context differs, the main messages remain consistent: balance proven processes with well-managed advanced methods, support operators through specific training, and invest in reliable data-driven decision support. Together, these efforts can address current challenges while positioning the city’s water sector to adapt and thrive in the face of future demands.

REFERENCES

[1] World Health Organization. (2011). Guidelines for drinking-water quality [Book]. In *WHO Library Cataloguing-in-Publication Data* (FOURTH EDITION). Available at: https://iris.who.int/bitstream/handle/10665/44584/9789241548151_eng.pdf

[2] Comisión Nacional del Agua. (2024, March). *Estadísticas del Agua en México 2023.* Available at: https://sinav30.conagua.gob.mx:8080/Descargas/pdf/EAM2023_f.pdf

[3] INEGI. (n.d.). Número de Habitantes. *Cuéntame de México.* Available at: https://cuentame.inegi.org.mx/monografias/informacion/df/poblacion/

[4] Sánchez, F. (2025, January 13). Rehabilitan planta potabilizadora “Río Magdalena 1”; opera desde los años 70 siendo la más antigua de CDMX. *El Universal*. Available at: https://www.eluniversal.com.mx/metropoli/rehabilitan-planta-potabilizadora-rio-magdalena-1-opera-desde-los-anos-70-siendo-la-mas-antigua-de-cdmx/

[5] Lázaro, E. (2024, December 30). Gobierno de la CDMX traza metas a 2027 para resolver la escasez de agua. *El Economista*. Available at: https://www.eleconomista.com.mx/estados/gobierno-cdmx-traza-metas-2027-resolver-escasez-agua-20241230-740084.html

[6] Parsai A., Rokade V., Faniran G., Ojo D., Kumwenda S., Kurui E., Ogendi G., Moturi W., Nyawanga D., Makaudze E., & Treacy J. (2019). *The relevance of hygiene to health in developing countries* (N. Potgieter & A. N. Traore Hoffman, Eds.). Available at: https://mts.intechopen.com/storage/books/6682/authors_book/authors_book.pdf

[7] Comisión Nacional del Agua. (2017, April 4). *¿Qué es una planta potabilizadora?* Available at: https://www.gob.mx/conagua/articulos/que-es-una-planta-potabilizadora?idiom=es

[8] U.S. Centers for Disease, Control and Prevention. (2024, February 6). *How water treatment works*. Available at: https://www.cdc.gov/drinking-water/about/how-water-treatment-works.html

[9] Environmental Protection Agency. (2018). *Drinking Water Report for Public Supplies 2017*. Environmental Protection Agency, Ireland. Retrieved from: https://www.epa.ie/publications/compliance--enforcement/drinking-water/annual-drinking-water-reports/2017-DW-Report_web_Final.pdf

[10] Ngwenya, N., Ncube, E.J., & Parsons, J. (2013). Recent Advances in Drinking Water Disinfection: Successes and Challenges. *Reviews of environmental contamination and toxicology*, *222*, 111-70. https://doi.org/10.1007/978-1-4614-4717-7_4

[11] Kang G., Gao J. Z. & Xie G. (2017). Data-Driven Water Quality Analysis and Prediction: A Survey. *2017 IEEE Third International Conference on Big Data Computing Service and Applications (BigDataService)*, Redwood City, CA, USA, 224-232. https://doi.org/10.1109/BigDataService.2017.40

[12] Aliashrafi, A., Zhang, Y., Groenewegen, H. & Peleato N. M. (2021) A review of data-driven modelling in drinking water treatment. *Rev Environ Sci Biotechnol, 20*, 985–1009. https://doi.org/10.1007/s11157-021-09592-y

[13] VanDerslice J. (2011). Drinking Water Infrastructure and Environmental Disparities: Evidence and Methodological Considerations. *American Journal of Public Health, 101*, S109_S114. https://doi.org/10.2105/AJPH.2011.300189

[14] Silva, J. (2024). Model to improve the management of drinking water supply for the inhabitants of Mexico City. *Tecnología y Ciencia del Agua, 15*(3), 391-421. https://doi.org/10.24850/j-tyca-2024-03-09

[15] CONAGUA-Sistema Nacional de Información Del Agua (n.d.). Available at: https://sinav30.conagua.gob.mx:8080/

[16] Gobierno de la Ciudad de México. (2023). *Alcaldías (shapefile)*. Portal de Datos Abiertos. Sistema Ajolote. Retrieved from: https://datos.cdmx.gob.mx/dataset/bae265a8-d1f6-4614-b399-4184bc93e027/resource/8648431b-4f34-4f1a-a4b1-19142f944300?inner_span=True&activity_id=ad2844be-2e5b-4d80-af3f-31dd1dd8e1b8

[17] Silva, J.A. and Martínez Omaña, M.C. (2023), Comparative study of drinking water management in Mexico City and Singapore, *Management of Environmental Quality*, *34*(2), 308-330. https://doi.org/10.1108/MEQ-04-2022-0108

[18] CONAGUA-Sistema Nacional de Información del Agua (n.d.). *Plantas Potabilizadoras.* Infraestructura Hidráulica. Retrieved from: https://sinav30.conagua.gob.mx:8080/SINA/?opcion=potabilizadoras

[19] *DELEGACIONES DEL DISTRITO FEDERAL.* Retrieved from: https://deleg16.weebly.com/

[20] CONAGUA. (n.d.). *Programa Hídrico Regional 2021-2024 - Región Hidrológico-Administrativa XIII Aguas del Valle de México*. Gobierno de México. Retrieved from: https://files.conagua.gob.mx/conagua/generico/PNH/PHR_2021-2024_RHA_XIII_OCAVM.pdf

[21] *Google Colab*. (n.d.). Available at: https://colab.research.google.com/

[22] Wimalawansa, S. J. (2013). Purification of contaminated water with reverse osmosis: effective solution of providing clean water for human needs in developing countries. *International journal of emerging technology and advanced engineering*, *3*(12), 75-89. Available at: https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=14777e52b5efa8ff94e529051b24fad15808df6a

APPENDIX A: RESULTS IN TABLE FORMAT

Table A1. Operational Efficiency.

Plant	Municipality	Treatment Process	Capacity	Flow Rate	Operational Efficiency (Flow Rate / Capacity)
Santa Catarina 11	Iztapalapa	Direct Filtration	40	41	102.500000
San Lorenzo Tezonco	Iztapalapa	Reverse Osmosis	6	6	100.000000
Cerillos 3	Xochimilco	Direct Filtration	40	40	100.000000
Cerillos 2	Xochimilco	Direct Filtration	40	40	100.000000
Carlos L. Gracidas	Iztapalapa	Direct Filtration	38	38	100.000000
R-11	Xochimilco	Direct Filtration	40	40	100.000000
Auxiliar Xotepingo 4C	Iztapalapa	Direct Filtration	50	50	100.000000
Deportivo Ferrería	Azcapotzalco	Direct Filtration	50	50	100.000000
Xaltepec	Iztapalapa	Direct Filtration	500	500	100.000000
Viga 2	Iztapalapa	Direct Filtration	40	40	100.000000
Río Magdalena	La Magdalena Contreras	Direct Filtration	210	210	100.000000
Escudo Nacional 2	Xochimilco	Direct Filtration	40	40	100.000000
S-13	Xochimilco	Adsorption	40	40	100.000000
Balbuena 2	Venustiano Carranza	Direct Filtration	40	40	100.000000
San Luis Nuevo	Xochimilco	Direct Filtration	60	60	100.000000
Parque Vía Vallejo	Gustavo A. Madero	Direct Filtration	40	40	100.000000
Agrícola Oriental	Iztapalapa	Reverse Osmosis	240	220	91.666667
Tecomitl 18	Tláhuac	Reverse Osmosis	50	43	86.000000
Santa Catarina 13	Iztapalapa	Reverse Osmosis	60	51	85.000000
Tlacotal	Iztacalco	Reverse Osmosis	60	50	83.333333
Jardines del Pedregal 5	Gustavo A. Madero	Direct Filtration	80	65	81.250000
La Caldera	Iztapalapa	Direct Filtration	700	540	77.142857
Río Magdalena 2	La Magdalena Contreras	Direct Filtration	200	150	75.000000
Parque Las Antenas	Iztapalapa	Direct Filtration	30	22	73.333333
Santa Anita	Venustiano Carranza	Direct Filtration	40	29	72.500000
Panamericana	Gustavo A. Madero	Direct Filtration	50	34	68.000000
Vista Alegre	Venustiano Carranza	Reverse Osmosis	40	26	65.000000
Purísima 2	Iztapalapa	Direct Filtration	60	37	61.666667
Santa Catarina 10	Iztapalapa	Reverse Osmosis	60	33	55.000000
Cerro De La Estrella 2	Iztapalapa	Direct Filtration	60	32	53.333333
Purísima 3 Y 7	Iztapalapa	Reverse Osmosis	135	70	51.851852
Tecomitl 2	Iztapalapa	Reverse Osmosis	60	30	50.000000
Trabajadores del Hierro	Azcapotzalco	Direct Filtration	50	25	50.000000
Sta. Cruz Meyehualco	Iztapalapa	Reverse Osmosis	120	56	46.666667
Viga 4	Iztacalco	Direct Filtration	60	25	41.666667
Granjas San Antonio	Iztapalapa	Direct Filtration	60	25	41.666667
Iztapalapa 1	Iztapalapa	Reverse Osmosis	60	25	41.666667
Ciudad Deportiva 2	Iztacalco	Direct Filtration	100	37	37.000000
San Lorenzo Tezonco Nuevo	Iztapalapa	Reverse Osmosis	60	12	20.000000
Santa Catarina	Tláhuac	Reverse Osmosis	500	86	17.200000
Iztapalapa 2	Iztapalapa	Direct Filtration	60	9	15.000000

Table A2. Treatment Process Efficiency Summary.

Treatment Process	Total Capacity (LPS)	Total Flow Rate (LPS)	Avg. Operational Efficiency (%)	Number of Plants
Adsorption	40	40	100	1
Direct Filtration	2778	2259	79.631834	27
Reverse Osmosis	1451	708	61.02963	13

Table A3. Municipality Water Infrastructure Summary.

Municipality	Number of Plants	Total Capacity (LPS)	Total Flow Rate (LPS)	Avg. Operational Efficiency (%)
Azcapotzalco	2	100	75	75
Gustavo A. Madero	3	170	139	83.083333
Iztacalco	3	220	112	54
Iztapalapa	20	2439	1837	68.324735
La Magdalena Contreras	2	410	360	87.5
Tláhuac	2	550	129	51.6
Venustiano Carranza	3	120	95	79.166667
Xochimilco	6	260	260	100

Table A4. ANOVA Test Results.

Test Type	Dependent Variable	Grouping Variable	F-value	p-value	Significant?
ANOVA	Operational Efficiency (%)	Treatment Process	2.77	0.076	No ($p > 0.05$)

Table A5. ANOVA Boxplot Results.

Treatment Process	Median Efficiency	Variability	Interpretation
Adsorption	1	None	Very efficient, but small sample
Direct Filtration	~100%	High	Generally efficient, but with some underperformers
Reverse Osmosis	~55%	High	Less efficient overall, inconsistent

APPENDIX B: PYTHON CODES AND OUTPUTS

Figure B1. Libraries importing code snippet.

```python
import pandas as pd            # For handling DataFrames and data manipulation
import matplotlib.pyplot as plt  # For creating various types of graphs
```

Figure B2. File import into Google Colab.

```python
from google.colab import files

df = pd.read_csv('/content/Plantas-potabilizadoras-CDMX-2023.csv')
```

Figure B3. *df.head()* output.

```python
df.head()
```

	id	anio	planta	rha	estado	municipio	proceso	capacidad	caudal
0	1901	2023	Río Magdalena 2	Aguas del Valle de México	Ciudad de México	La Magdalena Contreras	Filtración Directa	200	150
1	1908	2023	Xaltepec	Aguas del Valle de México	Ciudad de México	Iztapalapa	Filtración Directa	500	500
2	1918	2023	Jardines del Pedregal 5	Aguas del Valle de México	Ciudad de México	Gustavo A. Madero	Filtración Directa	80	65
3	1925	2023	Santa Catarina 10	Aguas del Valle de México	Ciudad de México	Iztapalapa	Ósmosis Inversa	60	33
4	1932	2023	Santa Catarina 13	Aguas del Valle de México	Ciudad de México	Iztapalapa	Ósmosis Inversa	60	51

Figure B4. *df.isnull()* output.

```
pd.concat([df.isnull().head(5), df.isnull().tail(5)])
```

	id	anio	planta	rha	estado	municipio	proceso	capacidad	caudal
0	False	False	False	False	False	False	False	False	False
1	False	False	False	False	False	False	False	False	False
2	False	False	False	False	False	False	False	False	False
3	False	False	False	False	False	False	False	False	False
4	False	False	False	False	False	False	False	False	False
36	False	False	False	False	False	False	False	False	False
37	False	False	False	False	False	False	False	False	False
38	False	False	False	False	False	False	False	False	False
39	False	False	False	False	False	False	False	False	False
40	False	False	False	False	False	False	False	False	False

Figure B5. Python code for *Section 4.1.1*

```
# Renaming and translating columns
df.rename(columns={
    'planta': 'plant',
    'municipio': 'municipality',
    'proceso': 'treatment_process',
    'capacidad': 'capacity',
    'caudal': 'flow_rate'
}, inplace=True)

# Translating treatment process values
df['treatment_process'] = df['treatment_process'].replace({
    'Filtración Directa': 'Direct Filtration',
    'Ósmosis Inversa': 'Reverse Osmosis',
    'Adsorción': 'Adsorption'
})

# Calculating operational efficiency
df['operational_efficiency'] = (df['flow_rate'] / df['capacity']) * 100

# Plot
plt.figure(figsize=(12,6))
plt.barh(eff_table['plant'], eff_table['operational_efficiency'])
plt.xlabel('Operational Efficiency (%)')
plt.title('Operational Efficiency per Plant in Mexico City (2023)')
plt.tight_layout()
plt.show()

# Sorting and displaying efficiency table
eff_table = df[['plant', 'municipality', 'treatment_process', 'capacity', 'flow_rate', 'operational_efficiency']]
eff_table = eff_table.sort_values(by='operational_efficiency', ascending=False)
display(eff_table)
```

Figure B6. Python code for *Section 4.1.2*

```
# Grouping by treatment process
process_summary = df.groupby('treatment_process').agg({
    'capacity': 'sum',
    'flow_rate': 'sum',
    'operational_efficiency': 'mean',
    'plant': 'count'
}).rename(columns={
    'capacity': 'Total Capacity (LPS)',
    'flow_rate': 'Total Flow Rate (LPS)',
    'operational_efficiency': 'Avg. Operational Efficiency (%)',
    'plant': 'Number of Plants'
}).reset_index()

# Plot
plt.figure(figsize=(8, 6))
plt.bar(process_summary['treatment_process'], process_summary['Avg. Operational Efficiency (%)'])
plt.ylabel('Average Operational Efficiency (%)')
plt.title('Average Efficiency by Treatment Process (2023)')
plt.tight_layout()
plt.show()

# Displaying table
display(process_summary)
```

Figure B7. Python code for *Section 4.1.3*

```
# Grouping by municipality
municipal_summary = df.groupby('municipality').agg({
    'plant': 'count',
    'capacity': 'sum',
    'flow_rate': 'sum',
    'operational_efficiency': 'mean'
}).rename(columns={
    'plant': 'Number of Plants',
    'capacity': 'Total Capacity (LPS)',
    'flow_rate': 'Total Flow Rate (LPS)',
    'operational_efficiency': 'Avg. Operational Efficiency (%)'
}).reset_index()

# Plot
plt.figure(figsize=(12, 6))
plt.bar(municipal_summary['municipality'], municipal_summary['Avg. Operational Efficiency (%)'])
plt.ylabel('Average Operational Efficiency (%)')
plt.title('Average Operational Efficiency by Municipality (2023)')
plt.xticks(rotation=45)
plt.tight_layout()
plt.show()

# Displaying table
display(municipal_summary)
```

Figure B8. Python code for *Section 4.1.4*

```
from scipy import stats

# Grouping by treatment process
df['treatment_process'] = df['treatment_process'].replace({
    'Filtración Directa': 'Direct Filtration',
    'Ósmosis Inversa': 'Reverse Osmosis',
    'Adsorción': 'Adsorption'
})

# Creating groups
df['operational_efficiency'] = (df['flow_rate'] / df['capacity']) * 100
df_direct = df[df['treatment_process'] == 'Direct Filtration']['operational_efficiency']
df_reverse = df[df['treatment_process'] == 'Reverse Osmosis']['operational_efficiency']
df_adsorp = df[df['treatment_process'] == 'Adsorption']['operational_efficiency']

# Running ANOVA
f_stat, p_val = stats.f_oneway(df_direct, df_reverse, df_adsorp)
print("F-statistic:", round(f_stat, 2))
print("p-value:", round(p_val, 3))

# Boxplot
plt.figure(figsize=(8, 6))
df.boxplot(column='operational_efficiency', by='treatment_process')
plt.title('Operational Efficiency by Treatment Process')
plt.suptitle('')
plt.xlabel('Treatment Process')
plt.ylabel('Operational Efficiency (%)')
plt.tight_layout()
plt.show()
```