Monitoring the Water Utility Performance in Drinking Water Quality Compliance using Data Mining Approaches

Janvier Mwitirehe (janvier.mwitirehe@gmail.com)  
University of Rwanda

W. K. Cheruiyot  
University of Rwanda

Charles Ruranga  
University of Rwanda

Research Article

Keywords: Drinking water quality index, key performance indicators, performance discrepancy scores, performance monitoring, quality requirement, Twitter sentiment scores

Posted Date: June 27th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1772248/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License. 
Read Full License
Monitoring the Water Utility Performance in Drinking Water Quality Compliance using Data Mining Approaches

Janvier Mwitirehe1*, W.K. Cheruiyot2 and Charles Ruranga3

1 PhD Candidate, African Centre of Excellence in Data Science, University of Rwanda
2 Professor, Jomo Kenyatta University of Agriculture and Technology (JKUAT)
3 Professor, African Centre of Excellence in Data Science, University of Rwanda
* Corresponding author: janvier.mwitirehe@gmail.com

Abstract

Monitoring the quality of water ensures it is safe for life. Nowadays, its effectiveness requires embracing data science. This was done in the context of the Water and Sanitation Corporation (WASAC) of Rwanda, which manages 18 water treatment plants (WTPs). This research aims at monitoring the drinking water quality key performance indicators (KPIs) using an interactive dashboard that can map interventions, perform calculations, assess the performance, and provide near real-time information. A four-step approach—KPI identification, data collection, monitoring tools, and progress tracking—was used. Step 1 used the literature review to identify KPIs in terms of compliance with clean water regulatory and reliability requirements, and the existence of a functional system to address customer needs. Then, primary and secondary data sets, including Twitter data, were collected in step 2. The third step consisted of computing the drinking water quality index (DWQI), performance discrepancy scores (PDS), and Twitter sentiment scores, as well as an interactive dashboard. Finally, results were analyzed to determine WASAC’s performance towards quality KPIs. Compliance with clean water requirements has been generally consistent, and the first five WTPs were Kadahokwa, Cyunyu, Mutobo, Muhazi, and Kimisagara. However, the water reliability could not be confirmed due to missing data. Twitter data analysis confirmed the existence of a system to address customer needs and to notify customers of planned rationing or interruptions of water supply. Therefore, the suitability of water for human consumption was generally established, but more efforts were recommended to ensure full and consistent compliance.

Key words: Drinking water quality index; key performance indicators; performance discrepancy scores; performance monitoring; quality requirement; Twitter sentiment scores.
1. Introduction

Despite the fact that water is essential for human life, it may contain harmful substances (Sargen, 2019). National legislation and international standards are therefore in place to guide the production of safe water. They require the creation of water safety strategies, adjustments of chemical and microbiological parameter limitations in response to identified risks, and the implementation of risk-based monitoring techniques (World Health Organization, 2018).

There are two ways to monitor the suitability of the drinking water: performance and operational monitoring. The topic of this article is performance monitoring, which entails regular sampling and testing to demonstrate compliance with water guideline values and other regulatory requirements. Operational monitoring, on the other hand, involves establishing and maintaining a system that identifies, prevents, and corrects faulty and inefficient operation of operational processes in order to ensure the safety of drinking water (Chow, 2005).

The quality of drinking water in sub-Saharan Africa, where Rwanda is located, is still an issue. In fact, only 30.2% of sub-Saharan people have access to improved water sources within their premises, available when needed, and free from microbiological and priority chemical contamination, compared to 75.4% in Latin America & Caribbean, 78.5% in the Middle East & North Africa, 97.8% in Europe, 97.3% in the United States, and 99% in Canada (The World Bank, 2022). In Rwanda, the same source indicates that only 12.1% of the Rwandan population benefits from those services.

In sub-Saharan Africa, the quality of water may be influenced not only by climate change, rapid population growth, urbanization, and agriculture practices (van de Berg, et al., 2019), but also by insufficient efforts and resources put in place by the country's water services to monitor the quality of drinking water (Souza, Ramba, Wensley, & Delport, 2006). For instance, the evaluation of testing programs for fecal contamination done over 72 institutions (water suppliers and public health agencies) across 10 countries in sub-Saharan Africa revealed that only 85% of institutions had conducted some microbial water testing in a one-year period (Peletz, Kumpel, Bonham, Rahman, & Khush, 2016).

In Rwanda, Water and Sanitation Corporation (WASAC) Limited, the largest and government-supported water supplier, was given the mission to contribute to the "sustainable, equitable, reliable, and affordable access to safe drinking water for all Rwandans as a contribution to improving public health and socio-economic development" (Ministry of Infrastructure, 2018). Water quality data reports from each water treatment plant are collected daily using excel spreadsheets and consolidated.
every Thursday by the Corporate Planning Department, which has monitoring and evaluation functions in its attributions. There is no computerized system in place to process and analyze that data, whose volume increases every day. In addition, there are other sources of data, such as social media and call center data, that would be used to track key performance indicators (KPIs) associated with the achievement of the mission above.

Nevertheless, a performance monitoring approach can be explored to verify the compliance with clean water regulatory and reliability requirements and the existence of a functional system to address customer needs. Being an integral part of business intelligence systems, Sultan, Khedr, Idrees, & Kholeif (2017) indicated that the choice of key performance indicators (KPIs) is critical to success. This process may include eight KPIs, such as the degree of drinking water quality compliance; urban access to improved drinking water sources; rural access to improved drinking water sources; achievement of target drinking water flow and pressure; water supply reliability; a system for addressing customer complaints; and well-functioning water supply interruption and rationing programs (Rwanda Utilities Regulatory Authority, 2015; The Water Research Foundation, 2014). Failure to analyze daily data reports on a regular and consistent basis represents a missed opportunity for daily monitoring of drinking water quality KPIs. For instance, through ranking WTPs based on quality performance, for instance, this could create a competition for producing excellent water and meeting customer needs.

Three metrics are considered for quality control and compliance at the water treatment plant level: turbidity, power of hydrogen (pH), and residual free chlorine. According to the Rwanda Bureau of Standards (RBS), potable water should have a turbidity not exceeding 5 Nephelometric Turbidity Units (NTU) and a pH of 6.5-8.5 (Rwanda Bureau of Standards, 2011). Turbidity is the amount of cloudiness in the water and can indicate the presence of hazardous chemical and microbial contaminants and has significant implications for water quality (World Health Organization, 2017). As for pH, a measure opposing the levels of acidic (hydrogen) and alkalinity (base) substances in the treated water, is expressed on a scale of 0 to 14, with pH of 7 being the ideal level where water is neutral (Nancy & John, 2005). In addition, WTPs measure the level of residual free chlorine in the treated water and water distribution networks. Residual free chlorine is a disinfectant that kills bacteria and similar microorganisms, and according to RBS guidance, it should be between 0.2 and 0.5 mg/liter in the absence of epidemic diseases; however, it may increase in the case of faecal contamination or a potential outbreak of waterborne disease (Rwanda Bureau of Standards, 2011).

While WASAC collects huge amounts of data from different sources, including but not limited to social media, its strategic plan 2018–2024 includes only two measurable quality KPIs out of eight
provided in this article, and those are rural and urban access to drinking water (Ministry of Infrastructure, 2018). Therefore, this article went beyond measurable KPIs scope to leverage big data to get more insights on drinking water quality compliance. It adopted the use of a performance dashboard in Microsoft Power BI as a visual tool for drinking water quality compliance monitoring. This is a contribution towards reversing the traditional reporting of having 80% text and 20% graphics (Emery, 2015). Available data on water quality KPIs was used to develop an interactive performance dashboard that can be used by a water utility to internally and externally track drinking water quality compliance.

The quality KPIs used are described in Table 1 and 2 below. Apart from the "achievement of the target drinking water flow and pressure", for which the data was not available, the remaining KPIs are the subject matter of this article.

Table 1. KPIs for Regulatory Compliance

<table>
<thead>
<tr>
<th>KPI</th>
<th>Targets</th>
</tr>
</thead>
<tbody>
<tr>
<td>The Degree of Drinking Water Quality Compliance</td>
<td>A water utility consistently meets all drinking water standards at 100% each year.</td>
</tr>
<tr>
<td>Urban access to improved sources of drinking water</td>
<td>100% of households have access to clean drinking water within 200 m in urban areas.</td>
</tr>
<tr>
<td>Rural access to improved sources of drinking water</td>
<td>100% of households have access to clean drinking water within 500 m in rural areas.</td>
</tr>
</tbody>
</table>

Table 2. KPIs for Reliability requirements and Addressing Customer Needs

<table>
<thead>
<tr>
<th>KPI</th>
<th>Targets</th>
</tr>
</thead>
<tbody>
<tr>
<td>Achievement of target drinking water flow and pressure</td>
<td>The pressure of water supplied to customers in the range between 0.6 and 1.5 bars, or from 6 to 15 m static head, is recorded daily and consistently reported annually at 100%.</td>
</tr>
<tr>
<td>Water supply dependability (reliability)</td>
<td>The average number of days the water was not supplied per a week is equal to or less than 3 days</td>
</tr>
<tr>
<td>A system for addressing customer complaints</td>
<td>The utility maintains a reliable, timely, and responsive customer service center.</td>
</tr>
<tr>
<td>Well-functioning water supply interruption programs</td>
<td>100% of annual water supply interruptions with at least 48 hours' notice.</td>
</tr>
<tr>
<td>Well-functioning water supply rationing programs</td>
<td>100% of annual rationing programs that respected the requirements of at least 72-hour prior notice and 95% adherence to.</td>
</tr>
</tbody>
</table>

It is in that framework that this article aims at monitoring the drinking water quality using an interactive dashboard in Microsoft Power Business Intelligence (BI) that can map interventions, perform calculations, assess the quality performance, identify trends, and provide near real-time information. Specifically, the article strives to achieve the following specific objectives:

- Create a performance discrepancy score to measure the extent to which the treated water complies with the quality requirements in terms of the standard values for turbidity, pH, and residual free chlorine.
- Carry out Twitter text mining and sentiment analysis to monitor the process of addressing customer needs.
- Track and document the progress on the key quality performance indicators in line with the performance benchmarks for effectively managed utilities.

2. Materials and Methods

2.1. Datasets

Datasets used include raw numerical and Twitter datasets as well as secondary data. In addition, Demographic and Health Survey (DHS) reports maintained on the website of the National Institute of Statistics of Rwanda were also used.

Numerical dataset:

This was a 24-month dataset presented in 13,158 rows and 32 columns with daily data recorded by all 18 WTPs from 1st July 2019 to 30th June 2021. Its variables are divided into the following categories:

a. Each WTP's name, location (province and district)

b. Water data: raw, treated, used, and supplied water in cubic meters (m3)
c. Quality of data: turbidity of raw and treated water, pH and residual free chlorine. It was noted that Shyogwe-Mayaga WTP had not provided quality data.

d. Chemical products: daily quantities used by each WTP

Twitter data:

931 sampled tweets tagged "@wasac_rwanda" were collected. The sample included tweets that appeared on the WASAC Twitter account from February 3rd, 2020, to June 18th, 2021. This required an application for a Twitter Developer account that was approved on July 5th, 2020. Upon approval of the application, customer keys, consumer secret keys, access tokens, and access token secrets were granted to access and use Twitter data.

DHS secondary data:

Information from DHS reports for the years 2010, 2014/2015, and 219/2020 was used to track and compare three consecutive levels of urban and rural access to improved sources of drinking water.

2.2. Methodology

The numerical dataset was received from the WASAC Cooperate Planning Department. Therefore, a compiled excel sheet (final dataset) with additional computed columns was created, and the latter included relevant variables such as drinking water quality index (DWQI), and performance discrepancy score (PerfScore, or PDS). Finally, energy and time use were also computed as proportions of actual values to the corresponding daily treated water volumes and daily available hours, respectively.

As for Twitter data, tweets were extracted from the WASAC Twitter account and then converted into a dataset using R/RStudio. In addition, tweets that were expressed in Kinyarwanda were translated into English to facilitate their analysis using Google translation. Before doing this, the dataset was exported in Excel to clean up all abbreviations and inconsistent records in Kinyarwanda and then imported back into R/RStudio for data processing and data mining.

1) DWQI Computation
The use of DWQI was motivated and fully explained in the other article "Operational monitoring of drinking water quality compliance". In fact, DWQI was calculated on chemicals used by WTPs using the following formula (Tyagi, Sharma, Singh, & Dobhal, 2013; Akter, et al., 2016):

$$\text{DWQI} = \frac{\sum qi wi}{\sum wi}$$  \hspace{1cm} \text{Equation 1}

Where:

i. The quality rating scale $qi = \frac{Ci}{Si} \times 100$ \hspace{1cm} \text{Equation 2}

$Ci$, and $Si$ indicate respectively the concentration of $i$ parameter, and standard value of $i$ parameter.

ii. The relative weight $wi = \frac{1}{Si}$ \hspace{1cm} \text{Equation 3}

That is the standard value of the $i$ parameter is inversely proportional to the relative weight.

2) PDS computation

PDSs measure the proximity of the actual values to the critical values. They convey the idea that, the more PDSs increase in absolute values, the more WTP fails to comply with water quality requirements. The compliance should be ascertained with minimal scores. They were computed using three quality variables, such as treated water turbidity, pH, and residual free chlorine, through the process explained here below.

i. The critical or expected value is selected in such a way that the PDS will be minimal when the actual value and the critical value are close to one another, and vice versa. The required performance is to have a median or lower performance score.

<table>
<thead>
<tr>
<th>Table 3. Quality Performance Critical Values</th>
</tr>
</thead>
<tbody>
<tr>
<td>No</td>
</tr>
<tr>
<td>-----</td>
</tr>
<tr>
<td>1</td>
</tr>
<tr>
<td>2</td>
</tr>
<tr>
<td>3</td>
</tr>
</tbody>
</table>

Sources: (World Health Organization, 2018; Rwanda Bureau of Standards, 2011)

ii. Calculate the difference between the actual ($x_i$) and and the critical values ($\mu_i$) for each quality indicator

iii. Using the Z-score formula, standardize the computed differences per each value to make them comparable by dividing by the standard deviation ($\sigma_i$):
\[ Z - \text{Score} = \frac{(x_i - \mu_i)}{\sigma_i} \]  

Equation 4

iv. Compute the sum of the squared computed \( z \)-scores.

v. Calculating the square root of the results from step iv

vi. Obtain performance discrepancy scores by dividing the new results as per step v by three.

3) **Twitter sentiment scores**

Twitter text analytics and sentiment scores were computed using the R packages “TwitterR” and “SentimentAnalysis” with other support packages. They were calculated using tweets without punctuation and other non-informative words.

4) **Creating an interactive dashboard**

Processed data, including clean Twitter data and sentiment scores, was then uploaded to Power BI desktop to create a water quality performance monitoring dashboard. Power BI is a self-service tool that completes the entire data management process of connecting data sources, importing the desired data, applying and tracking transformations to data to get it in the reporting format, building interactive data visualizations, and then publishing and sharing them. The interactive dashboard was created by means of dragging and dropping Power BI visualization tools (tables, charts, maps, etc.) in the appropriate areas or running the interconnected R and Python script visuals. Using the dashboard, all 18 WTPs were described, and the progress on the KPIs was documented and analyzed.

When exploring data in Power BI, queries are generated dynamically and sent back to the source. It is a live connection; any field selection or filter sends a query back to the source, and the visual is updated with the new results. After saving the report, any of the visuals can be pinned to a customized dashboard. The data in the dashboard can be refreshed at any time, and the dashboard can be shared within the organization to keep the entire team up-to-date (Theresa, 2015).
3. Results

The dashboard offers an opportunity to visualize global and specific water quality performances by WASAC managed water treatment plants per day, month, quarter, semester, and year.

3.1. Performance discrepancy scores

The dashboard conveys the information related to PDS (PerfScore) together with its input variables such as pH, turbidity of treated water, and residual free chlorine. The PerfScore distribution is skewed due to outliers in variables used to compute the scores; and so, the PerfScore average of 0.60 with a standard deviation of 6.66 is different from its median of 0.44. Under such conditions, median values were considered in the analysis.

With PDS minimum and maximum median values of 0.2 and 1.0, respectively, the scores of 16 of 17 WTPs (94%) ranged between 0.2 and 0.7 median values. The analysis indicated that the most potable water production was found in Kadahokwa, Cyunyu, and Mutobo with a score of 0.2 each, followed by Muhazi and Kimisagara with a score of 0.3. Then, Mpanga, Karenge, Nyagatare-Gatsibo, and Kanyonyomba follow them with a score of 0.4 each. The next treatment plant was Ngenda, with a score of 0.5, followed by Gisuma and Kanyabusage, with a score of 0.6 each. The next bunch includes Nyamabuye, Gihira, Gihuma, and Nzove, with a score of 0.7 each. Finally, Rwasaburo comes in last place with a score of 1.0. Due to missing data, Shyogwe-Mayaga was not classified.

The dashboard provided a means for playing around with every plant’s score to see the associated values of input variables. For example, Kadahokwa, which produced the most potable water according to this analysis, had a pH of 7.00, which was a desirable level where water is neutral in acid and base, turbidity of 0.49 NTU, which was in the recommended range of 0.2 and 0.5 NTU, and residual free chlorine of 0.70 mg/liter, which seemed a bit outside the range of 0.2 and 0.5 as provided by RBS but was still recommended in case of faecal. However, in the case of Rwasaburo, which came in last place in terms of potable water, it appeared that the actual values in input variables were mostly outside the recommended ranges or critical values. For example, its pH was 6.5, while the recommended range is between 6.5 and 8.5, and it had a median turbidity of 3.38 NTU (the highest) and residual free chlorine of 0.9 mg/liter, both of which were outside of the recommended ranges.
As per Figure 2, a relationship between DWQI and PerfScore was suspected but it was not proven statistically significant (correlation =-0.005, t =-0.55, df = 12127, p > 0.05). Let’s compare the results.
of both metrics in the following table.

**Table 4. Comparison between Performance Discrepancy Score and DWQI**

<table>
<thead>
<tr>
<th>PerfScore Rank</th>
<th>WTP</th>
<th>Median PerfScore</th>
<th>Rating per DWQI</th>
<th>Median DWQI</th>
<th>DWQI Rank</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Kadahokwa</td>
<td>0.15</td>
<td>A</td>
<td>3.71</td>
<td>8</td>
</tr>
<tr>
<td>2</td>
<td>Cyunyu</td>
<td>0.17</td>
<td>A</td>
<td>0.52</td>
<td>2</td>
</tr>
<tr>
<td>3</td>
<td>Mutobo</td>
<td>0.22</td>
<td>A</td>
<td>1.00</td>
<td>3</td>
</tr>
<tr>
<td>4</td>
<td>Muhazi</td>
<td>0.26</td>
<td>A</td>
<td>8.84</td>
<td>13</td>
</tr>
<tr>
<td>5</td>
<td>Kimisagara</td>
<td>0.28</td>
<td>A</td>
<td>3.88</td>
<td>9</td>
</tr>
<tr>
<td>6</td>
<td>Mpanga</td>
<td>0.36</td>
<td>A</td>
<td>12.21</td>
<td>16</td>
</tr>
<tr>
<td>7</td>
<td>Karenge</td>
<td>0.39</td>
<td>A</td>
<td>6.89</td>
<td>11</td>
</tr>
<tr>
<td>8</td>
<td>Nyagatere-Gatsibo</td>
<td>0.39</td>
<td>A</td>
<td>1.94</td>
<td>5</td>
</tr>
<tr>
<td>9</td>
<td>Kanyonyomba</td>
<td>0.41</td>
<td>A</td>
<td>10.52</td>
<td>14</td>
</tr>
<tr>
<td>10</td>
<td>Ngenda</td>
<td>0.52</td>
<td>A</td>
<td>15.86</td>
<td>18</td>
</tr>
<tr>
<td>11</td>
<td>Gisuma</td>
<td>0.62</td>
<td>A</td>
<td>2.10</td>
<td>6</td>
</tr>
<tr>
<td>12</td>
<td>Kanyabusage</td>
<td>0.64</td>
<td>A</td>
<td>0.51</td>
<td>1</td>
</tr>
<tr>
<td>13</td>
<td>Nyamabuye</td>
<td>0.66</td>
<td>A</td>
<td>1.80</td>
<td>4</td>
</tr>
<tr>
<td>14</td>
<td>Gihira</td>
<td>0.69</td>
<td>A</td>
<td>7.55</td>
<td>12</td>
</tr>
<tr>
<td>15</td>
<td>Gihuma</td>
<td>0.73</td>
<td>A</td>
<td>11.46</td>
<td>15</td>
</tr>
<tr>
<td>16</td>
<td>Nzove</td>
<td>0.74</td>
<td>A</td>
<td>3.92</td>
<td>10</td>
</tr>
<tr>
<td>17</td>
<td>Rwasaburo</td>
<td>1.04</td>
<td>A</td>
<td>2.73</td>
<td>7</td>
</tr>
<tr>
<td>18</td>
<td>Shyogwe-Mayaga</td>
<td></td>
<td></td>
<td>13.61</td>
<td>17</td>
</tr>
</tbody>
</table>

In light of Table 4, a consistency in terms of quality performance under both metrics was only noticed for Cyunyu and Mutobo WTPs. This appeared not to be the case for the remaining plants. For instance, Kadahokwa, the highest performer under PerfScore, came in 8th place under DWQI, while Kanyabusage, with the highest performance in terms of DWQI, was ranked 12th under PerfScore. Furthermore, Rwasaburo ranked 17th and 7th in PerfScore and DWQI, respectively, while Ngenda ranked 10th in PerfScore despite having the lowest DWQI performance.
3.2. Twitter Data Insights

Out of 931 tweets posted during 16.5 months (3rd February 2020 to 18th June 2021), 325 were retweeted and 854 were replied to. The dashboard (Figure 3) can help select tweets that occurred on any given date, month, quarter, and year. The word cloud indicates key words used most frequently. Those are water, problem, help, follow, call, number, going, and give in the sentences (Table 5).

Table 5. Examples of Tweets Containing Keywords

<table>
<thead>
<tr>
<th>Key word</th>
<th>Frequency</th>
<th>First text</th>
<th>Last text</th>
</tr>
</thead>
<tbody>
<tr>
<td>water</td>
<td>150</td>
<td>A few days ago, we informed you about the problem of water shortage</td>
<td>You usually call and he comes to provide you with water. You can find his number on past invoices.</td>
</tr>
<tr>
<td>call</td>
<td>129</td>
<td>Give us your help numbers, or give us a call</td>
<td>You usually call and he comes to provide you with water. You can find his number on past invoices.</td>
</tr>
<tr>
<td>help</td>
<td>121</td>
<td>Give us your help number, or give us a call</td>
<td>You can call us on 0788775248 and we will help you</td>
</tr>
<tr>
<td>follow</td>
<td>107</td>
<td>Where do you have this problem? It is good that we know it so that we can follow it</td>
<td>You can send us your number to direct message and we will follow up with you</td>
</tr>
<tr>
<td>know</td>
<td>97</td>
<td>Where do you have this problem?</td>
<td>We didn’t know. We are going to follow it</td>
</tr>
<tr>
<td>problem</td>
<td>93</td>
<td>Where did you have this problem?</td>
<td>You can call us at the number we have given you and we will come and address the issue</td>
</tr>
<tr>
<td>going</td>
<td>74</td>
<td>Thank you for the direction, we are going to come and see what is going on</td>
<td>We didn’t know. We are going to follow it up. Thank you</td>
</tr>
<tr>
<td>give</td>
<td>70</td>
<td>Give us your help number, or give us a call</td>
<td>You can give us your numbers or call us on 0737715768 immediately</td>
</tr>
<tr>
<td>number</td>
<td>55</td>
<td>Where is it? You can give us a POC number and we will help you. Thank you</td>
<td>You usually call and he comes to provide you with water. You can find his number on past invoices.</td>
</tr>
</tbody>
</table>

While the WASAC Twitter account is a forum where water-related issues are discussed, key words are indicated in general positive dialogues between WASAC representatives and customers. This can be confirmed with the positive average mean-sentiment scores throughout the year (Figure 5). However, the lowest scores were generally found during the dry season due to frequent water shortage issues. The high values were found during the rainy seasons, i.e., in April, September, October, November, and December.

### 3.3. Monitoring the progress on KPIs

This section includes drinking water quality indicators grouped into compliance with regulatory, reliability, and customer requirements or needs.

#### 3.3.1. Consistency in Drinking Water Quality Performance

On a quarterly and yearly perspective, except for Nyamabuye, Gihira, and Ngenda WTPs that had slightly larger performance fluctuations, the remaining WTPs had relatively consistent performance. In terms of yearly performance discrepancy scores, two distinct clusters were noted among WTPs:

a. **Cluster 1-Most performers**: nine WTPs that scored 0.44 (median PerfScore) or below.
   i. WTPs that consistently scored 0.2 or below (the most potable) were: Kadahokwa, Cyunyu and Mutobo
ii. Those whose consistent scores are in the range of 0.2 and 0.3 inclusively included Muhazi and Kimisagara

iii. Those whose consistent scores are in the range of 0.3 and 0.4 inclusive were Mpanga, Karenge, Nyagatare-Gatsibo and Kanyonyomba

Figure 4. Dashboard-DWQI KPI 2019/2020

b. **Cluster 2-Less performers:** six WTPs that scored above 0.44 (median PDS).

i. Those that made a significant improvement to reach the level of 0.5 were: Gihira, Ngenda, Nyamabuye

ii. Those that consistently scored between 0.6 and 0.8 inclusive were: Gisuma, Kanyabusage, Gihuma, and Nzove

iii. A WTP whose PerfScore is above 0.8 (least potable water): Rwasaburo

3.3.2. **Reliability of water supply**

Water has been supplied each day. However, the figure of median daily water supply indicated rises and falls (Fig. 5), which raised a question about the reliability of water supply. The dashboard selection
options helped confirm the same issue per each WTP and per each month and fiscal year. However, without household water supply survey data, it was not possible to estimate the actual number of days water was not supplied to WASAC customers per WTP.

Figure 5. Daily median water supply from 1 July 2019 to 30 June 2021

3.3.3. Access to improved sources of drinking water

DHS 2019/2020 stated that 80% of the households in Rwanda had access to improved sources of drinking water. This marked an increase of 7% from 2014, according to the DHS 214/2015 report (Figure 6).

In urban areas, access to clean data has been increasing and reached 96% in 2019/2020, up from 90% in 2010 and 91% in 2014/2015. The rural areas seemed to be left behind with 71%, 69%, and 77%, respectively, in 2010, 2014/2015, and 219/2020. Unlike urban areas, access to clean water decreased by 2% from 2010 to 2015, though there has been a big increase of 8% thereafter. Though DHS reports do not state the distance of the improved sources of drinking water from the households, as stated in both KPIs, it appeared that 100% of households with access to clean drinking water within 200 m in urban areas and 100% of households with access to clean drinking water within 500 m in rural areas had not yet been achieved.
3.3.4. WASAC’s system for addressing customer complaints and water supply interruption and rationing notification

Customer care is under the responsibility of the WASAC commercial unit, which operates in 20 branches countrywide, of which six are in Kigali. This unit is also responsible for billing and recovery, revenue collection, marketing, and a call center with a toll-free number 3535. This number is used by customers in need of any support. Suggestion boxes were placed at headquarters and at all branches for customers to file complaints. In addition, WASAC holds weekly Friday meetings that deal with rationing the water supply. Those meetings accommodate the senior management, including all WTP managers.

Using Twitter's data dashboard, it was realized that WASAC was used to informing the public of any anticipated water supply interruption and/or rationing plan. For instance, selecting "supplying", "supply", "works", "construction", and "rationing" words in the word cloud, it was noted that the water supply interruption and rationing plans were communicated to customers ahead of time as per the figure below.
It was also noted that WASAC also provided shared updates with its customers, as per the below examples: "Our client, the problem that prevented Nyakabanda from getting water has been solved"; "This issue has been resolved." It was caused by the water pipe we washed"; we resumed the supply today. The repair of a big pipe from Kimisagara WTP, which supplies Kacyiru, is almost done. Thanks", etc.

4. Discussions

In terms of drinking water quality compliance, nine first-quality performing WTPs (i.e., 52.94%) have had performance discrepancy scores equal to the median scores (0.4) or less. Those included Kadahokwa, Cyunyu, Mutobo, Muhazi, Kimisagara, Mpanga, Karenge, Nyagatare-Gatsibo, and Kanyonyomba, respectively. They were followed by Gihira, Ngenda, Nyamabuye, Kanyabusage, Gihuma, and Nzove with scores between 0.5 and 0.8. Finally, the lowest quality water production was found in Rwasaburo WTP with a score of 1.0. Furthermore, the results above did not outline any significant correlation with DWQI. This section explored this further and discussed the validity of the results.
4.1. Relationship between DWQI and PDS

As described in the introduction, there are two approaches to monitoring the suitability of the drinking water: performance and operational monitoring. From the perspective of operational monitoring, DWQI was computed to detect whether the dosage of the chemical inputs used during the treatment process respects acceptable limits set by WHO, RBS, or product manufacturers. On the other hand, PDS was computed as a performance monitoring tool aimed at measuring the extent to which the treated water complies with the quality requirements. Under such conditions, a significant correlation was expected between both quality metrics.

Though the PDS and DWQI relationships were not confirmed statistically, both had minimal values in the months of the dry season (short and extended). In the short dry season (January-first half of March), normally the sunny days predominate the rainy ones, while there is almost no rainfall occurring in the extended dry season (second half of May to end of August). Therefore, high values were seen in the months of April for DWQI and September, October, November, and December for DWQI and PDS when there was heavy rainfall. Furthermore, using a logistical model, the DWQI-based analysis identified significant relations between the quality of drinking water with turbidity and the geographical location (province) of the WTP. This appeared to be the case even for PDS, as depicted in Fig. 8.

Accordingly, the WTPs in the Northern, Southern, and Eastern provinces had lower performance discrepancy scores. Their high-quality water production might be associated with lower levels of turbidity, pH, and residual free chlorine. Western province and Kigali City, in contrast, had high levels
of turbidity and residual free chlorine, which might be associated with their reduced quality of water produced in their WTPs.

Nevertheless, the Southern and Easter provinces that showed relatively lower performances (with a little bit higher DWQI than the other provinces), have been good performers in terms of PDS (Fig. 9). It is worth noting that all WTPs with DWQI below 25 or PDS of 0.4 or lower were qualified as excellent water producers.

To understand this situation, let’s revert to the methodology used to select the water quality testing parameters. Indeed, water quality is divided into three different areas, such as microbial contamination, chemical contamination, and physical contamination (Li, 2021). The operational monitoring was done on chemical inputs during the water treatment process, while the performance monitoring was on final water outputs in water tanks and distribution pipes to test for physical contamination, leaving out microbial contamination due to a lack of data. By nature, operational monitoring catalyzes preventive measures to reduce toxic and hazardous substances in the water (Chow, 2005). Although quality water outputs heavily depend on operational monitoring, the latter cannot replace performance monitoring. For instance, water may still get contaminated due to harmful substances that may be present where it is stored or flows. In short, the more successful operational monitoring is done, the more performance monitoring will outline quality water production.
4.2. Quality of the used data and validity of results

Finally, the numerical data used in the analysis was skewed and was characterized by missing data and outliers. Data was collected on a daily basis and compiled by WASAC's corporate planning department. In the presence of many data records, there was an inherent risk of errors that could go undetected. This was the reason why missing data and outliers were encountered, and as a result, data distributions were skewed. Under such conditions, this article considered median values, which were less affected by outliers than in Kaur, Stoltzfus, & Yellapu (2018) to preserve the central tendency.

As for Twitter data, tweets in Kinyarwanda were translated into English to take advantage of the natural language processing offered by the R/RStudio package. Therefore, Google translations might not have been accurate due to abbreviations and inconsistent expressions found in tweets expressed in Kinyarwanda. Therefore, very important key words around the water supply interruption and rationing, such as "rationing", "supply", and others were verified and rewritten in a consistent manner as much as possible. In addition, in the interpretation of the results, the consideration was based on keywords together with associated words.

5. Conclusion

At the end of this article, it is worth emphasizing the importance of monitoring the drinking water quality compliance to ensure the suitability of the water for human consumption. The water treatment process involves the use of drugs that can be harmful to human health if the standards for their use are not respected. After the treatment, the drinking water can still be contaminated with toxic and hazardous substances from where it is stored or flows.

Safe drinking water here refers to the compliance with clean water regulatory and reliability requirements and the existence of a functional system to address customer needs. Therefore, seven KPIs were monitored to confirm the suitability of the water. Those were: degree of drinking water quality compliance, urban access to improved sources of drinking water, rural access to improved sources of drinking water, reliability of water supply, a system for addressing customer complaints, and well-functioning water supply interruption and rationing programs.

Primary and secondary data sets, including Twitter data, were used to compute the monitoring tools, such as DWQI, PDS, and Twitter sentiment scores. Both the DWQI and PDS results indicated water turbidity and WTP geographical location as key quality factors determining compliance. However, comparing quarterly and yearly compliance results (2019 to 2021), the quality compliance of the most
WTPs has been consistent, and the first nine quality water productions were found in Kadahokwa, Cyunyu, Mutobo, Muhazi, Kimisagara, Mpanga, Karenge, Nyagatare-Gatsibo, and Kanyonyomba. In accordance with the regulation of water supply services in Rwanda (Rwanda Utilities Regulatory Authority, 2015) about water reliability, it was noted that water was supplied every day, but the trend analysis indicated some ups and downs in the daily water supply. This lack of consistency raised doubts about the reliability of the water supply. Without household water supply data, it was not possible to estimate the actual number of days water is not supplied to WASAC customers.

As for access to clean water, the used DHS reports did not state the distance to drinking water sources as provided in the KPI definition. That is, 200 meters and 500 meters in urban and rural areas, respectively. They only indicated that 96% and 77% of urban and rural households, respectively, have access to clean water. Moreover, the discussion with the head of the WASAC Commercial Unit and the Twitter data analysis indicated the existence of a system to address the customer's needs and to notify customers of planned rationing or interruption of water supply. Therefore, the suitability of Rwanda's treated and distributed water for human consumption was generally established, but more efforts are recommended to rigorously monitor and ensure full and consistent compliance.

**Acknowledgements**

This work could not have been possible without the financial support of the African Center of Excellence in Data Science (ACE-DS) of the University of Rwanda. The same gratitude also goes to WASAC, which accepted to be the case study and, therefore, to provide the data used in this paper.

**Competing interests**

The authors declare that they have no financial or personal relationships that may have inappropriately influenced them in writing this article.

**Author’s contributions**

JM: Conceptualization, Data curation, Formal analysis, Methodology, Project administration, Resources, Investigation, Visualization, Writing - original draft, writing - review & editing. CWK: Project administration, Supervision, Validation, writing - review & editing. CR: Project administration, Supervision, Validation, writing - review & editing.

**Funding information**
This research was funded by the University of Rwanda, African Center of Excellence in Data Science.

Data availability
https://drive.google.com/drive/folders/1meF7XdRfDhyEGnsY3OBovbx7ULJuuMlE?usp=sharing

Disclaimer

The views and opinions expressed in this article are those of the authors, and do not necessarily reflect the official policy or position of any affiliated agency of the authors.

References


Sargen, M. (2019, September 26). *Biological Roles of Water: Why is water necessary for life?* Retrieved from Science in the news:


