**REPORT  
TASK 1  
Introduction**

Water is a vital resource essential for human survival, and its efficient management is crucial for the well-being of individuals and communities alike. As a responsible entity, as a Data Analyst I have been tasked with the duty of ensuring that this essential resource is distributed and utilized in a fair and sustainable manner.   
One key aspect of this responsibility is the accurate billing of water consumption to the Billibg Classes.  
  
 This report provides an analysis of the water billing data from January 2019 to October 2022, with a focus on the trends, patterns, and insights that can inform our efforts to improve the efficiency and effectiveness of our water billing system. Specifically, we will examine the total amount of water billed, the average monthly consumption, and the locations consuming the most versus the billing in each period, as well as any notable changes or fluctuations that may have occurred over this four-year period.

**BODY**

**Commonalities throughout the data**The data is comprised of 574 locations, all situated in South Africa.  
The water billing data is managed by Division 3, with a single division responsible for water billing.

The locations are categorized into six distinct billing classes:

* Domestic
* Commercial
* Other
* Not assigned
* Industrial
* Departmental (internal changes)

**Year 2019**

**1. 2019 January (WATER BILLING DASHBOARD FOR JANUARY 2019)**  
  
The data for January 2019 reveals significant discrepancies in water billing. The total quantity of water consumed was 8,564,046,191 kilolitres, while the total billing amount was R397,177,791.25. This translates to a rate of water consumption per billing amount of 4.64%. This means that only 4.64% of the total water consumption was actually billed, leaving a significant portion of 95.36% unbilled.

This stark reality suggests that there are significant shortfalls in water billing, with a substantial portion of the total water consumption remaining unaccounted for. This highlights the need for a thorough review of the billing system and potential leaks or inefficiencies in the process to ensure accurate and timely billing.  
  
  
  
**FAULTS IN THE DATA SOURCE AND ITS DATA COLLECTION**  
  
The data presented appears to show that the majority of water consumption is from Cape Town, as depicted on the map provided in the first visualization of the story, with a few isolated areas scattered across the country. However, this data is not reliable due to its limited representation of Cape Town and potential distortions. Upon further research, it becomes clear that the original source website www.capetown.gov.za classifies the data as "Cape Town" specific, which raises concerns about its accuracy and generalizability as a trusted source.

This classification suggests that the data may not be representative of the broader national picture and may be biased towards Cape Town's unique characteristics. As a result, this data may be spreading false and skewed information, highlighting the need for improvement in data collection and representation to ensure more accurate and reliable findings.

Moreover, the lack of context and transparency regarding the data collection methodology and sampling process further erodes confidence in the data's reliability. It is essential to have a clear understanding of how the data was collected and what biases or limitations may be present in order to make informed decisions.

Therefore, it is crucial to exercise caution when using this data and to consider alternative sources that may provide a more comprehensive and accurate picture of water consumption patterns across the country. Only through rigorous and transparent data collection and analysis can we hope to achieve a more accurate understanding of this critical issue. The reason for bringing this up is because the same mistake has been repeated for all the years upon research in their set of data collection

**WATER CONSUMPTION BY LOCATION**The top 3 locations that consume the most water are; Phillipi, Nyanga and Strand.  
The high water consumption in Philippi, Nyanga, and Strand can be attributed to several factors:

**Population Density and Growth:**

Philippi and Nyanga are among the most densely populated areas in Cape Town. High population density leads to increased overall water consumption as more people use water for daily activities such as drinking, cooking, bathing, and sanitation. Rapid urbanization and population growth in these areas further exacerbate the demand for water.

**Infrastructure and Leakages:**

In older and densely populated areas like Nyanga and parts of Philippi, aging infrastructure can lead to significant water losses due to leaks and inefficiencies. These losses contribute to higher recorded water consumption as the water lost through leaks is accounted for in the overall usage metrics​​.

**Social and Cultural Factors:**

In communities with lower socio-economic status, there may be a higher incidence of shared water facilities and communal taps. This can lead to inefficient water use and wastage, driving up overall consumption. Additionally, public facilities such as schools and community centres in these areas also contribute to higher water use​.

The Billing Class consuming the most water is the Department, and the Billing Class consumes less water is the Domestic Customers. **WATER BILLING BY LOCATION AND BILLING CLASS**

Top 3 places with the most billing are Philippi, Observatory and Cape Town City Centre. Here we notice that although Philippi is still the first, Observatory and Cape Town City Centre were not in the top 3.   
The top three areas with the highest water bills in Cape Town each face distinct challenges driving their elevated billing. In Philippi, high water consumption is due to its dense population and significant agricultural activity, which requires extensive water usage.

Observatory, with its mix of residential and commercial properties, experiences high water bills driven by economic factors and higher living standards, which lead to increased water consumption.

The Cape Town City Centre, being a major commercial and tourist hub, sees elevated water usage from businesses, hotels, and public facilities, resulting in substantial water bills. The combination of economic rates, situational demands, and varied usage patterns across these areas contributes to their higher water billing​.

The Billing class that pays more water bills is the Domestic customers and the Billing class that pays more bills is the Domestic customers.

**Key Take-away.**Despite consuming less water overall compared to billing classes like the Department and commercial sectors, domestic users often end up paying more due to the billing structure based on usage. The tiered pricing system charges domestic customers increasingly higher rates as their water consumption rises. This means that even though their overall usage may be lower, the incremental increase in rates with higher consumption results in disproportionately higher bills for domestic users. Additionally, factors such as aging infrastructure in residential areas leading to water loss through leaks contribute to inflated bills for domestic customers. Consequently, while they may use less water compared to other sectors, the direct billing based on consumption results in domestic users bearing a heavier financial burden for their water usage.  
  
In contrast, the Department class utilizes the most water overall, encompassing municipal services such as parks, public buildings, and city maintenance. These entities require substantial water resources for activities such as watering parks, cleaning public facilities, and sustaining municipal operations. Despite their high-water consumption, departmental costs are often covered by different budget arrangements, shielding them from the same high bills faced by domestic users. This discrepancy highlights a disparity in the financial burdens associated with water usage in Cape Town, with domestic customers paying more for their consumption despite using less water overall. Addressing this imbalance may require reevaluating budget allocations and exploring strategies to ensure fair and sustainable water management practices across all sectors of the city.

**2. 2019 December (WATER BILLING DASHBOARD FOR DECEMBER 2019)**

The data for December 2019 reveals significant discrepancies in water billing. The total quantity of water consumed was 9 277 044 950 kilolitres, while the total billing amount was   
R883 532 311,14. This translates to a rate of water consumption per billing amount of 9.52%. There has been an increase of 4,88% in the water billing rate per consumption from January 2019.

The comparison between December 2019 and January 2019 reveals a significant disparity in water billing dynamics. In January 2019, the rate of water consumption per billing amount stood at 4.64%, indicating that only a fraction of the total water consumed was billed, leaving a substantial portion unbilled. However, by December 2019, this rate had increased to 9.52%, representing a considerable rise in the cost of water relative to consumption. This 4.88% increase suggests that customers are now paying a higher price for the same quantity of water consumed compared to January 2019.

Various factors could contribute to this shift, including adjustments in water tariffs, changes in infrastructure maintenance costs, or fluctuations in resource availability and demand. Analysing these figures alongside January 2019 data provides valuable insights into long-term trends and variations in water billing practices, highlighting the need for stakeholders to closely monitor and assess water management strategies. This ensures that pricing remains fair and sustainable while upholding the quality and accessibility of water services for all consumers.

The billing rate is the highest in Paranoma where Departmental rates are dominant, Eppind Industria 2 lead by Not Assigned Billing class then discrete commercial and industrial billing class, and Far Field Estate department (largely by) then small commercial and very small domestic .

In Paranoma, the highest billing rate is driven by significant water consumption from municipal services like parks and public buildings, categorized under Departmental rates. Paranoma's appeal as a beautiful residential area also contributes to higher costs, as residents use more water for maintaining gardens and large properties. This increased demand, combined with Departmental billing rates, results in higher overall water bills for the area.

In Epping Industria 2, the "Not Assigned" billing class indicates a diverse mix of water users, including industrial and commercial entities. This area is characterized by numerous factories and warehouses, which have high water demands for manufacturing processes, cooling systems, and sanitation. The concentration of industrial and commercial activities leads to elevated water consumption, reflected in the higher billing rates for Epping Industria 2.

Far Field Estate's billing rate is influenced by the predominant Departmental billing class, showing significant water usage by municipal services. The presence of small commercial and very small domestic billing classes suggests a mixed-use area with varying water consumption levels. Additionally, as a primarily renting area, Far Field Estate's transient population may have different water usage habits compared to homeowners, affecting overall consumption and billing rates. This highlights the need to consider demographic factors and residential patterns when analysing water billing data to understand the underlying drivers of consumption and pricing disparities across different neighbourhoods.

**3. 2021 January (WATER BILLING DASHBOARD FOR JANUARY 2021)**  
The data for January 2021 reveals the total quantity of water consumed was 540 092 823,23 kilolitres, while the total billing amount was   
R480 917 164,87 This translates to a rate of water consumption per billing amount of 89%. There has been an increase of 79,52% in the water billing rate per consumption from December 2019.  
  
Concrete evidence from WHO shows that by January 2021, South Africa had reported a substantial number of COVID-19 deaths. According to the World Health Organization (WHO), by the end of January 2021, South Africa had recorded approximately 44,164 COVID-19 related deaths. This decrease in population due to the pandemic likely contributed to the reduced water consumption while maintaining high billing amounts, thus increasing the billing rate per consumption significantly. The lower population density would have led to decreased overall water usage, but fixed costs and tariff structures would have kept the billing amounts relatively high, resulting in a higher rate of water consumption per billing amount.  
  
For the first time in a very long time, domestic consumption was the greatest. This surge in domestic water usage is because of preventive measures recommended during the pandemic. Health guidelines advised people to drink 8-9 glasses of hot water daily to help combat the virus. Since there was no lockdown yet, workers had to bathe with hot water every time they returned home to kill the virus and immediately wash their clothing. Additionally, frequent hand washing and disinfecting household surfaces significantly increased water usage. People also used more water for cleaning groceries and packages brought into the home, cooking more meals at home due to restaurant closures, and maintaining higher standards of cleanliness overall to reduce the risk of infection. Increased usage of water for gardening and home improvement projects, as people spent more time at home, further contributed to the unprecedented rise in domestic water consumption. These factors collectively led to the highest domestic water usage seen in a long time.  
  
In contrast, the industrial and commercial billing classes saw a significant decline in water consumption. This decrease was primarily due to the pandemic's severe impact on the workforce, with many workers falling ill or dying, significantly reducing operational capacity and performance. As businesses struggled with diminished productivity, their demand for water decreased accordingly. Thus, while domestic consumption surged and departmental usage remained relatively high, the industrial and commercial sectors experienced a substantial reduction in water usage and billing due to reduced activity and workforce constraints.  
  
Water consumption shifted primarily towards domestic use, while usage in other sectors, including government facilities, decreased. Recognizing this trend, government initiatives may have aimed to align with broader conservation efforts by reducing water consumption to a lesser extent. As domestic usage surged due to increased hygiene practices and more time spent at home, government agencies may have proactively implemented measures to optimize water usage within their facilities. This could involve adjusting operational practices, promoting water-saving behaviours among employees, and prioritizing essential water needs while minimizing non-essential usage. By aligning with conservation initiatives during a period of heightened demand for domestic water, government entities could contribute to overall resource efficiency and resilience during the pandemic.

**4. 2021 December (WATER BILLING DASHBOARD FOR DECEMBER 2021)**  
  
The data for December 2021 shows a notable decrease in the water billing rate per consumption compared to January 2021. Despite a total quantity of water consumed amounting to 313,679,795.49 kilolitres, the total billing amount was R841,899,822.32. This translates to a rate of water consumption per billing amount of 37.26%. The decrease in the billing rate per consumption from January 2021 is significant, marked at 51.74%.

This reduction in the billing rate per consumption could stem from various factors. One potential reason could be the stabilization of water usage patterns following the initial surge in domestic consumption during the peak of the COVID-19 pandemic. As the situation evolved and preventive measures became more routine, domestic water usage may have gradually normalized, leading to a decrease in overall consumption. Additionally, government initiatives and public awareness campaigns promoting water conservation and efficiency could have contributed to the decline in water usage and subsequent billing rates.

Furthermore, economic considerations likely played a significant role in reducing water billing rates per consumption. The economic impact of the COVID-19 pandemic led to widespread financial strain, with many individuals experiencing job losses, income reductions, and economic uncertainty. To address these challenges, government agencies and water utilities may have implemented measures to reduce water billing rates, ensuring that essential water services remain accessible to all residents, regardless of their financial circumstances. Overall, the decrease in water billing rates per consumption reflects a multifaceted approach aimed at addressing both economic challenges and sustainability goals, promoting equitable access to this vital resource while supporting affordability for all.  
  
During this time, the COVID-19 pandemic wrought significant loss, with a notable number of lives had succumbed to the virus. The pandemic prompted widespread lockdown measures, disrupting daily life and economic activities. Concurrently, the economic fallout led to a surge in job losses, exacerbating financial hardships for countless individuals and families.

Moreover, the pandemic-induced disruptions also resulted in the loss of reliable sources of data for tracking water consumption and billing. Lockdown measures impeded the functioning of businesses and institutions, including water utilities and government agencies responsible for data collection. The closure of offices, reduced fieldwork, and limitations on in-person interactions hindered the timely gathering and analysis of data. As a result, there emerged a discrepancy in the patterns of water consumption and billing, with traditional data collection methods proving inadequate in capturing the nuances of evolving usage patterns during the pandemic hence even from the visualisation there is no clear pattern and correlation  
  
However it does look like the billing rate is high due to high cost. This could have been due to the fact the government was managing water wastage as too much water was being consumed.  
  
  
  
  
**5. 2022 January (WATER BILLING DASHBOARD FOR JANUARY 2022)**

The data for January 2022 reveals significant discrepancies in water billing. The total quantity of water consumed was 1 285 347 206,65 kilolitres, while the total billing amount was   
R455 497 473. This translates to a rate of water consumption per billing amount of 35,44%. There has been an decrease of 1,82% in the water billing rate per consumption from December 2021.

Despite domestic consumers leading in water consumption, the billing hierarchy presents a different scenario. Domestic users top the billing charts, followed by commercial billing classes, and then government departments. This shift in billing priorities reflects the impact of COVID-19 on consumption patterns. During the pandemic, domestic water usage surged as people spent more time at home, following hygiene protocols and addressing health concerns. As a result, domestic consumers emerged as the primary users of water.

Meanwhile, government departments maintained stable consumption levels, benefitting from consistent budget allocations and water subsidies. However, commercial activities gradually resumed with the easing of restrictions, leading to increased water usage in this sector. Despite domestic consumers leading in water consumption, their billing reflects the intricate dynamics of pandemic-induced behaviour shifts and the subsequent return to normalcy in commercial activities.  
  
  
**6. 2022 October (WATER BILLING DASHBOARD FOR OCTOBER 2022)**

The data for January 2022 highlights notable disparities in water billing. Despite a total consumption of 528,477,751.74 kiloliters, the total billing amount decreased to R362,669,590.25. This results in a rate of water consumption per billing amount of 68.63%. The decrease in both quantity consumption and billing is significant, indicating a shift in water usage patterns or billing practices.

The reduction in both water consumption and billing suggests several possible factors at play. Firstly, there may have been improvements in water efficiency or conservation efforts, leading to a decrease in overall consumption. Additionally, changes in weather patterns or seasonal variations could influence water usage, resulting in fluctuations in consumption levels. Moreover, adjustments in billing structures or tariff rates could contribute to the observed decrease in billing amounts despite stable or reduced consumption.

Overall, while the decrease in both quantity consumption and billing may reflect positive trends in water management or conservation, further analysis is needed to understand the underlying factors driving these changes. Monitoring trends over time and assessing the effectiveness of conservation measures will be crucial for ensuring sustainable water usage and billing practices in the future.  
  
There is a significant change in water consumption trends, with "other" and departmental categories now leading, while domestic usage has dropped to fifth position. This shift reflects positive efforts in water conservation among households. However, despite using less water, domestic users still face high bills, followed by "other," commercial, and departmental users. This discrepancy underscores the need for fairer billing practices, ensuring affordability for all while promoting water conservation.

**7. Task 2: (HISTORY OF WATER CONSUMPTION DASHBOARD FOR APRIL 2018)**  
I chose this dataset to examine household water consumption across various suburbs and assess if the consumption trends observed in April 2018 align with those seen in subsequent years, specifically from 2019 to 2022. By comparing consumption patterns over multiple years, we can identify any consistent trends or deviations, providing insights into long-term water usage behaviours in different suburban areas. This analysis aims to determine whether the high-water consumption observed in certain suburbs in April 2018 persists over time or if there are notable changes in consumption patterns in the years following 2019. Understanding these trends can inform water management strategies and resource allocation decisions to ensure sustainable and equitable access to water resources across suburban communities.  
  
In April 2018, household water consumption data reveals significant disparities across suburbs, with Philipi, Observatory, Parklands, Strand, and Gugulethu emerging as the top consumers. Philippi stands out for its high population density, primarily comprising informal settlements. Observatory, on the other hand, accommodates a substantial number of working individuals, often from the middle class, drawn to its proximity to Cape Town city and educational institutions. Parklands is notable for its rental properties, attracting residents seeking accommodation options. Meanwhile, Strand's appeal lies in its scenic beauty, particularly its beachfront, attracting both residents and tourists. Gugulethu, characterized by densely populated informal settlements, also features prominently in water consumption data.

The high-water consumption observed in Philipi and Observatory can be linked to their demographic and socioeconomic characteristics. The dense population in Philipi, coupled with inadequate access to basic amenities, likely contributes to higher water usage for daily necessities. Similarly, Observatory's residential composition, largely comprising working individuals with middle-class incomes, may result in increased water consumption for household activities and amenities. These consumption patterns persist across subsequent years, underscoring the enduring impact of demographic factors on water usage trends.

The persistence of demographic and socioeconomic influences on water consumption underscores the need for targeted interventions to address underlying challenges and promote sustainable water management practices across suburban communities. For this reason, efforts to improve access to basic amenities, enhance water efficiency measures, and address socioeconomic disparities are crucial in mitigating the high-water consumption observed in areas like Philipi and Observatory. Because these places are largely occupied by domestic customers, addressing the root causes of high-water usage is essential to align water billing with consumption patterns and ensure equitable access to water resources for all residents.

**CONCLUSION AND RECOMMENDATIONS**

* The report highlights the need for improved data collection and representation to ensure accurate and reliable findings.
* The analysis provides valuable insights into long-term trends and variations in water billing practices, highlighting the need for stakeholders to closely monitor and assess water management strategies.
* The report also underscores the importance of considering demographic factors and residential patterns when analysing water billing data.
* The COVID-19 pandemic may have led to increased water consumption in households, particularly in areas with high population density or aged infrastructure in year 2021.
* The pandemic may have also led to increased water billing rates, particularly in areas with high water consumption.
* The data may provide insights into how households responded to the pandemic, such as increased use of hot water or frequent hand washing.

However, it's important to note I ambasing these points and may not be representative of the entire population. Further analysis of a larger scale data would be needed to draw more definitive and strong conclusion . Also regarding that the Data source has some loophones in its data collection as I have proven so in the beginning of this Report.  
  
  
**ASSUMPTION**

Covid-19 had an impact on the consumption of water and water billing changes.

**Task 3**

A. Pre-processing and data cleaning:

For the Tableau assignment on Capetown water billing, I applied the following pre-processing and data cleaning steps:

1. **Data Cleaning:**

* Removed duplicate entries, missing values, and outliers to ensure data accuracy.
* Removed inconsistent data
* Removed unnecessary columns and aggregated data for easier visualization
* Handled missing values by imputing them with median values
* Added a Date column for time-based analysis.
* Created a new column for quantity, converting liters to kiloliters to maintain consistency.
* Removed negative values from the Quantity column.
* Replaced '#' values in the Location column with 'UNKNOWN' to standardize entries.

1. **Data Aggregation:**

* Calculated aggregates such as the rate of billing per consumption to analyse cost-efficiency.
* Summarized quantity to understand total water consumption.
* Determined maximum values to identify peak consumption periods.

These steps helped to ensure that the data was clean and ready for analysis.

**B. Non-trivial questions that can be answered by visualizations:**

The visualizations presented can answer non-trivial questions such as:

\* How does the total water usage vary by billing class? (Pie chart in showing total water usage by billing class -This is in all the dashboards, represented differently eg dashboard 2 is in a form of a table)

\* How does the total quantity of billing by location fluctuate over time, and are there any noticeable trends or patterns in billing behaviour? (Represented by map in dashboard 1, a bar graph in dashboard 2)

\* What is the trend in water consumption by Location (Represented by a tree map in dashboard 1, a tree map in dashboard 3)

These visualizations allow for more in-depth analysis and exploration of the data.

**C. Explanations on the non-trivial questions:**  
1. The first question explores how water usage patterns change across different billing classes over time and whether there are significant shifts in these patterns. This involves comparing total water usage for each billing class over multiple years using charts. The aim is to identify fluctuations in water consumption distribution among billing classes over time and understand the reasons behind these changes.

2. The second question focuses on identifying specific locations or suburbs with consistent trends in water consumption over the years and understanding the factors contributing to these trends. This involves visualizing the trend in water consumption by location over time, often depicted using a tree map. Stakeholders analyse these visualizations to pinpoint areas showing persistent consumption patterns or anomalies and investigate factors influencing these trends.

**D. Guidelines/principles applied in constructing the dashboard:**

guidelines/principles that I applied in constructing the dashboard are:

\* Keep it simple: I tried to avoid clutter and focus on the most important information, making it easy for users to understand the data.

\* Tell a story: I used a combination of visualizations to tell a story about the data, starting with a high-level overview and then drilling down into more detailed information.

\* Clarity and Accessibility: I ensured that the visualizations were easy to interpret and navigate, with clear labels and tooltips to provide context and insights to the user.

\* Consistency in Design: I maintained a consistent design theme throughout the dashboard, using standardized color schemes, fonts, and layout structures to enhance visual coherence and user experience.

**E. Recommended visualizations:**

1. **Scatter Plot:** While scatter plots are effective in displaying the relationship between two variables, such as water consumption and billing rates, they did not fit the data format or characteristics. In some cases, the dataset lacked the necessary variables or exhibited data points that did not lend themselves well to a clear scatter plot representation. Nonetheless, if applicable, scatter plots could still be useful for exploring correlations between different aspects of water usage and identifying potential outliers or anomalies in the data.
2. **Gantt Chart:** Gantt charts are excellent for visualizing project timelines or schedules, which could be relevant for tracking infrastructure development or maintenance activities related to water management. However, the dataset did not include detailed information on such projects or timelines, making it challenging to create a meaningful Gantt chart. Nevertheless, if available, a Gantt chart could provide stakeholders with a clear overview of planned or ongoing initiatives in water infrastructure management and their respective timelines.
3. **Heat Map:** Heat maps are useful for visually representing data density or concentration across different regions or categories. While heat maps could offer insights into spatial variations in water consumption or billing patterns, the dataset did not contain geospatial information at a granular level necessary for creating informative heat maps. Nevertheless, if spatial data were available, heat maps could help identify hotspots of water usage or billing activity, guiding targeted interventions or resource allocation efforts.

These visualizations would provide additional insights into the data and help users better understand complex relationships between different variables.