This project is aimed at using various SQL techniques and queries to find patterns in the Maji dogo data set as well as find inconsistencies in the data set and correct them.. The data was collected on various water sources in Maji dogo to help solve the water crises in Maji dogo. Hence this process will make our data more reliable for decision making.

The first thing that i did is to check the employee table and make sure we have the email address of all the employees. Since it was not there I used sql queries to create them.
We can determine the email address for each employee by:
- selecting the employee_name column
- replacing the space with a full stop
- make it lowercase
- and stitch it all together and add @ndogowater.gov.

The to creat the emails i did the following;
I combined this query
“SELECT
REPLACE(employee_name, ' ','.') −− Replace the space with a full stop
FROM
Employee “ which will replace the space in between the name of the employees with a full stop. And **“SELECT**
LOWER(REPLACE(employee_name, ' ','.'))
FROM
Employee” which will put everything in lowercase and lastly this
“SELECT
CONCAT(
LOWER(REPLACE(employee_name, ' ', '.')), '@ndogowater.gov') AS new_email
FROM
Employee” which will attach this “'@ndogowater.gov ” part of the email address. I put everything together in my query as indicated.

I picked up another bit we have to clean up. Often when databases are created and updated, or information is collected from different sources,

errors creep in. For example, if you look at the phone numbers in the phone_number column, the values are stored as strings.The phone numbers should be 12 characters long, consisting of the plus sign, area code (99), and the phone number digits. However, when we use
the LENGTH(column) function, it returns 13 characters, indicating there's an extra character.

```
SELECT
LENGTH(phone_number)
FROM
Employee;
```

That's because there is a space at the end of the number! If you try to send an automated SMS to that number it will fail. This happens so often
that they create a function, especially for trimming off the space, called TRIM(column).It removes any leading or trailing spaces from a string. Therefore I use the trim function to remove the spaces and then update the column.

Next I identified the top three employees with the most visits.

After that I query the total number of records collected based on the town, province , as well as rural and urban communities. The results show that Our entire country was properly canvassed, and our dataset represents the situation on the ground. Also majority of our water sources are in rural communities across Maji Ndogo.

I then check for the following;
1. How many people did we survey in total?
2. How many wells, taps and rivers are there?
3. How many people share particular types of water sources on average?
4. How many people are getting water from each type of source?

Calculating the average number of people served by a single instance of each water source type helps us understand the typical capacity or load
on a single water source. This can help us decide which sources should be repaired or upgraded, based on the average impact of each upgrade.

For example, wells don't seem to be a problem, as fewer people are sharing them. On the other hand, 2000 share a single public tap on average. Therefore we probably should focus on improving shared taps first.

Also I calculated the percentage of people that use each water source. And then rank each water source based on the number of people that use it. So that we can have a clear picture of the order in which we will start solving issues concerning each water source and which ones to channel our resources to first.

I then use the rank function to help me rank every source within each type of water source. In this case when we start to work on wells we know where to start from and when we start working on shared taps we will know which taps to start with etc.

Next i answered the following questions;
1. How long did the survey take?
2. What is the average total queue time for water?
3. What is the average queue time on different days?

For how long people have to queue on average in Maji Ndogo. Keep in mind that many sources like taps_in_home have no queues. These
are just recorded as 0 in the time_in_queue column, so when we calculate averages, we need to exclude those rows. using NULLIF().

So on average, people take two hours to fetch water if they don't have a tap in their homes.

I then use the case function to query the time in queue for each hour of each day.