



**Maji Ndogo: Visualizing the currents of change in Maji Ndogo**

# **Visualising Maji Ndogo's past**

**Introduction**

Setting the stage for our data exploration journey.

**Importing our data**

Updating our dataset

**Visual analysis**

Seeing the reality of the crisis

**Mapping our data**

Understanding spatial relationship

**Who collects our water**

Gender equality with water access

**Crime and water**

Linking data to find new insight

**Congratulations and Next Steps for Addressing Gender Inequality in Water Access**

Dear Team,

I hope this message finds you all well. First and foremost, I want to extend my heartfelt gratitude for your exemplary work on the water access project in Maji Ndogo. Your commitment and the leadership shown by Chidi have been instrumental in driving this project forward and shaping our understanding of the challenges we face.

Our recent project has shed light on several crucial aspects, including the gender dynamics in water collection. The innovative use of photographs by our data scientists to analyze queues has provided us with meaningful insights into this facet of the crisis. This approach exemplifies the power of data in revealing underlying societal issues and is a testament to the team's ingenuity.

As we transition to the next phase, our focus must be on effectively communicating these findings to our key stakeholders: the national and local governments, as well as our funders. The report we prepare needs to be comprehensive, clear, and impactful, conveying the full scope of our research and its implications.

Dalila, with your expertise in data visualization and Power BI, I am entrusting you with the leadership of this crucial task. Your skill in presenting complex data in an accessible manner will be pivotal in crafting a report that not only informs but also engages our stakeholders, helping them understand the depth and breadth of the water access issue in Maji Ndogo.

The aim is to create a narrative that is informative and compelling, yet balanced and objective. We need to ensure our report is a tool for decision-making, providing a clear picture of the current situation and guiding future strategies and interventions.

I am confident in your ability to lead this initiative and am looking forward to your approach to this important task. Your work will play a key role in shaping our efforts to tackle the water crisis effectively.

Thank you again for your hard work and dedication. I eagerly await your plan for moving forward with the report and am excited about the progress we will make together.

Best regards,  
Aziza Naledi

07:13

2



**Introduction**

Setting the stage for our data exploration journey.

**Importing our data**

Updating our dataset

**Visual analysis**

Seeing the reality of the crisis

**Mapping our data**

Understanding spatial relationship

**Who collects our water**

Gender equality with water access

**Crime and water**

Linking data to find new insight

**Re: Congratulations and Next Steps for Addressing Gender Inequality in Water Access**

Dear President Naledi,

Thank you for your message and for entrusting me with this significant responsibility. I am both humbled and honoured to lead this initiative, especially under your leadership, which consistently emphasizes the importance of data-driven approaches.

In alignment with your vision, I fully recognize the importance of effectively conveying our findings to our key stakeholders, including the national and local governments, and our funders. While the gender dynamics in water collection are a part of our findings, I understand the need for a balanced and comprehensive report that addresses the wider spectrum of issues revealed by our data.

To this end, I propose developing a series of interactive, visually engaging reports using Power BI. These reports will aim to present our data in a manner that is not only accessible but also compelling, ensuring that the key messages resonate with our stakeholders. Our focus will be on clarity and impact, ensuring that the information provided drives informed decision-making and strategy formulation.

In our visual narrative, we will utilize clean and intuitive designs, with a colour scheme that subtly underscores the water crisis theme. Our goal is to make complex data comprehensible, engaging, and actionable. I believe that by presenting our data in this manner, we will not only inform but also empower our stakeholders to understand and address the challenges we face in Maji Ndogo.

I will be coordinating with the team to initiate this project promptly. We are committed to delivering a report that meets the high standards set by our mission and your leadership.

Thank you again for your guidance and support. I look forward to updating you on our progress and contributing to our collective goal of addressing the water crisis in Maji Ndogo.

All the best,

Dalila

07:13





## Introduction

Setting the stage for our data exploration journey.



## Importing our data

Updating out dataset



## Visual analysis

Seeing the reality of the crisis



## Mapping our data

Understanding spacial relationship



## Who collects our water

Gender equality with water access



## Crime and water

Linking data to find new insight



Dalila Lesedi  
Online



Hello Team,

I'm excited to work with each of you on this pivotal project to highlight the findings from the survey data you worked on, and to add a dimension to our analysis, looking at the gender inequality in water access in Maji Ndogo.

07:25

You may wonder, why are we doing this? The answer lies in the power of data to illuminate truths that are often overlooked. Our goal is to make these truths not just visible, but visceral, compelling people to act. I'll explain some of the data we have in a minute, but we have a lot of thinking to do before we start creating anything.

07:28

You've all seen the principles of design, and I am entrusting these decisions to you. Whether it's choosing the right colour palette or selecting impactful visuals, the choices will be yours to make, but I will be showing you my thinking. My role here is to guide you through the intricate world of Power BI and to help transform our survey's findings into visual stories that resonate with whoever sees them.

07:31

Let's pause for a second so I can update you on everything that happened in the last couple of months.

07:36

A lot of my teams have started using ChatGPT to help them out. One note on that... ChatGPT was trained on a version of Power BI that has changed a lot, so its advice is not very good with Power BI-related tasks. Some of the more programmatic things in Power BI still work fine, but be careful if you ask about how to do something.

07:38

4





Dalila Lesedi  
Online



## Introduction

Setting the stage for our data exploration journey.



## Importing our data

Updating out dataset



## Visual analysis

Seeing the reality of the crisis



## Mapping our data

Understanding spacial relationship



## Who collects our water

Gender equality with water access



## Crime and water

Linking data to find new insight



When the field surveyors visited the different water sites, they took a lot of photos. Sanaa, our incredible data scientist looked at the photos and realised we could learn a lot from them if we use AI.

07:45

So, here's the brilliant part. Sanaa used something called Machine Learning, a type of Artificial Intelligence, to analyze these photos. She always impresses me when we talk! Think of it as teaching a computer to recognize patterns, almost like how a child learns to identify shapes or colours. In our case, Sanaa trained the computer to detect people in the pictures, but even better, she figured out a way to see if the people in the queues were men, women or children. So for each visit, we have an idea of the % men, women and children in each queue.

07:51



**Introduction**

Setting the stage for our data exploration journey.

**Importing our data**

Updating out dataset

**Visual analysis**

Seeing the reality of the crisis

**Mapping our data**

Understanding spacial relationship

**Who collects our water**

Gender equality with water access

**Crime and water**

Linking data to find new insight

Anyway, here is an example of what her team did:



42 Amani Loop, Sokoto, (record\_id = ID:17399)

07:55

Here the AI identifies the different people in the image, classifies them, and then counts each type.

Sanaa sent us a table with the record\_id (linked to the one from the visits table), and % composition as M, F or child. So for each time a queue time was recorded, we now know how many men, women and children were in the queue at that time.

08:00





## Introduction

Setting the stage for our data exploration journey.



## Importing our data

Updating out dataset



## Visual analysis

Seeing the reality of the crisis



## Mapping our data

Understanding spacial relationship



## Who collects our water

Gender equality with water access



## Crime and water

Linking data to find new insight



Dalila Lesedi  
Online



So here is a link to the latest version of the data:



[Md\\_summary.csv](#)

08:03

If you are using a virtual machine (VM) - mac users - I have some advice... Before you import data into Power BI, make sure the file you are using is on the VM. Copy it from the shared drive into the Downloads folder on the machine. If you don't, Power BI will take VERY long to load your data.

08:06

The data is now in one table to keep it simple. Since the new data is going to guide us, we have to first analyse it a bit so we know what story the data has to tell.

08:09

First things first, go ahead and import the CSV file into Power BI.

Just one thing to look out for; Power BI tries its best to identify the data types of data, but it fails sometimes. It may identify these columns as text: (percent\_male, percent\_female, percent\_child), but they're actually numbers. Make sure you check that the data has the correct format in the Table view.

08:15





## Introduction

Setting the stage for our data exploration journey.



Dalila Lesedi  
Online



## Importing our data

Updating our dataset



## Visual analysis

Seeing the reality of the crisis



## Mapping our data

Understanding spatial relationship

08:15



## Who collects our water

Gender equality with water access



## Crime and water

Linking data to find new insight

This is a short explanation of the new data:

1. hour\_of\_day - The hour of day at the time\_of\_record. This will help us aggregate by the hour of the day.
2. day\_of\_week - The day of the week from the time\_of\_record. This will help us aggregate by the day of the week.
3. percent\_male - The percentage of men in the queue at the time\_of\_record.
4. percent\_female - The percentage of men in the queue at the time\_of\_record.
5. percent\_child - The percentage of men in the queue at the time\_of\_record.

Now that the data is in, let's recap the main points from our SQL analysis and visually represent these.

1. Most water sources are rural in Maji Ndogo.
2. 43% of our people are using shared taps. 2000 people often share one tap.
3. 31% of our population has water infrastructure in their homes but within that group
4. 45% face non-functional systems due to issues with pipes, pumps, and reservoirs. Towns like Amina, the rural parts of Amanzi, and a couple of towns across Akatsi and Hawassa have broken infrastructure.
5. 18% of our people are using wells, but within that, only 28% are clean. These are mostly in Hawassa, Kilimani and Akatsi.
6. Our citizens often face long wait times for water, averaging more than 120 minutes:
  - Queues are very long on Saturdays.
  - Queues are longer in the mornings and evenings.
  - Wednesdays and Sundays have the shortest queues.

08:22



**Introduction**

Setting the stage for our data exploration journey.

**Importing our data**

Updating out dataset

**Visual analysis**

Seeing the reality of the crisis

**Mapping our data**

Understanding spacial relationship

**Who collects our water**

Gender equality with water access

**Crime and water**

Linking data to find new insight

Let's start with our national-level results. We're going to make a couple of visuals on the same page that communicate the results of the survey at a high level. These are the insights we want to show on the page:

1. Rural vs. urban population
2. 43% of our people are using shared taps.
3. 31% of our population has water infrastructure in their homes.
4. 18% of our people are using wells.

08:30

To do this, we create the following visuals on a single page:

1. Create a pie chart to illustrate the total Urban and Rural population split in Maji Ndogo. Use `SUM(number_of_people_served)`, split by `location_type`
2. Create a bar (column) chart showing the total population using the various water sources. Use `SUM(number_of_people_served)`, split by `type_of_water_source`

HINT: Remember the population was ≈28 Million. If you get more than that, remember that `visit_count = 1` prevented duplicate records, so use it as a filter condition by adding `visit_count` to the filter pane, and selecting 1.

08:36





## Introduction

Setting the stage for our data exploration journey.



1

## Importing our data

Updating out dataset



2

## Visual analysis

Seeing the reality of the crisis



3

## Mapping our data

Understanding spacial relationship



4

## Who collects our water

Gender equality with water access



5

## Crime and water

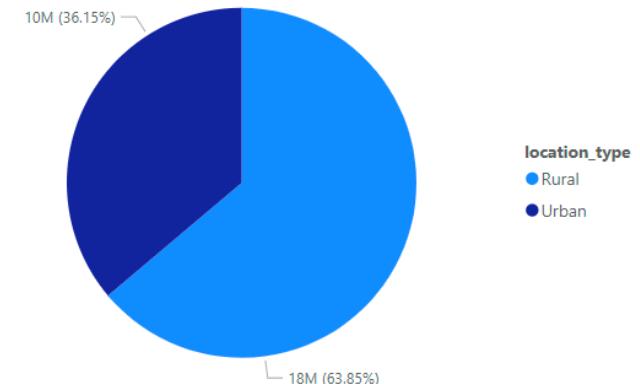
Linking data to find new insight



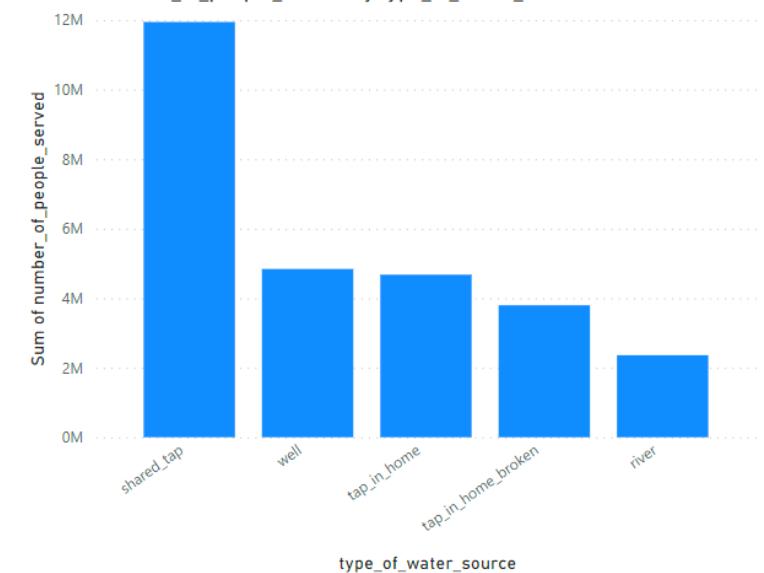
I am a bit mean here... I chose a terrible colour palette, and I am going to make a lot more mistakes in the other visualisations so we can see what good and bad visuals look like later on.

This is what I got:

Sum of number\_of\_people\_served by location\_type



Sum of number\_of\_people\_served by type\_of\_water\_source



08:44

I heard this once: "The only worse design than a pie chart is several of them". Pie charts are hard to read for humans, but, when there are few categories, and the differences between them are clear, pie charts really help us tell the "part of a whole" story well. For example, you can see that a lot more people live in rural areas compared to urban ones very easily.

08:47

10



**Introduction**

Setting the stage for our data exploration journey.

1

**Importing our data**

Updating our dataset

2

**Visual analysis**

Seeing the reality of the crisis

3

**Mapping our data**

Understanding spatial relationship

4

**Who collects our water**

Gender equality with water access

5

**Crime and water**

Linking data to find new insight

Because we have a couple of water source types, the bar chart makes it simpler to understand the breakdown of the source data. The most obvious part of the bar chart is just how much more people use shared taps compared to the others! 43% doesn't seem that big, but since the next largest source is only 18%, it is clear that shared taps make up a large portion of Maji Ndogo's water supply. However, the bar chart makes it a bit harder to understand the relative proportion of shared taps to the total population - 41%.

08:51

Duplicate the page we just made by right-clicking on the Page 1 tab in the bottom left corner. Switch the data sources so that the pie chart shows the water source type data, and the bar chart the population breakdown. Once you have made them, compare them to the first page:

1. How do these compare to the first set we made?
2. Which visuals are easier to interpret?
3. Do you learn anything more by switching these visuals?
4. What story do these visuals tell compared to the other set?

Once you're done, delete the duplicate page and rename the original page to something like National, so we can keep track of the different pages in our project going forward.

08:56

Since we made these two visuals on the National page, Power BI connects the underlying relationships in the data in the two visuals. We can select urban on the pie chart, and only the urban data is shown on the bar chart. The National bars are now a bit lighter, while the urban data is now highlighted on the bars.

09:04



**Introduction**

Setting the stage for our data exploration journey.

**Importing our data**

Updating out dataset

09:09

**Visual analysis**

Seeing the reality of the crisis

**Mapping our data**

Understanding spacial relationship

09:16

**Who collects our water**

Gender equality with water access

09:24

**Crime and water**

Linking data to find new insight

Clicking on the shared\_tap bar will filter the data for the pie chart, showing how many people share taps in urban and rural areas. Can you also see how the rural highlight in the pie chart is a bit larger than the urban one (larger radius)? This tells us that there are proportionally more shared taps in rural areas compared to urban ones.

Just a note, if you would like to see the national data again, without any filters, just click on the highlighted bar again.

It is important to remember that I had to explain the interpretation of the pie chart highlighting to you. It is always better to create visuals that don't need explanations and require little effort to understand.

Now select the tap\_in\_home bar. What do you see, and how can you interpret that?

09:27

So with one page and using only two visuals, we could illustrate insights 1, 2 and 3, and learn something more about taps installed in urban vs. rural houses we didn't know before. This is why I love visualising data! ❤️

09:34





## Introduction

Setting the stage for our data exploration journey.



## Importing our data

Updating out dataset



## Visual analysis

Seeing the reality of the crisis



## Mapping our data

Understanding spacial relationship



## Who collects our water

Gender equality with water access



## Crime and water

Linking data to find new insight

Dalila Lesedi  
Online



Next up, let's take on the second part of insight 4: Where is the infrastructure broken?

10:15

One of the visuals with the most impact is a map, so this is one of the crucial visualisations we'll use a lot!

10:20

Since Maji Ndogo isn't well known, we have to use a custom map to visualise the data. Here's the link to download the JSON map file and an image of Maji Ndogo's map we're going to need.



[MD\\_Provinces.json](#)



[MD\\_Map.png](#)

10:24

Add a page by clicking on the plus tab at the bottom of the Report view and then create a new Shape map visual from the Visualisations tab. Add province\_name in the location field, and the type\_of\_water\_source column into the Color saturation field.

10:28

Then navigate to the Format your visual tab on the visualisations page. Select Map settings and change the Map type to Custom map. Just below the Map type field, click on the Add a map type field and select the MD\_Provinces.json you downloaded. Now, on the format tab, select General, then Effects, and turn the Background off using the toggle.

10:35

13





Dalila Lesedi  
Online



## Introduction

Setting the stage for our data exploration journey.



## Importing our data

Updating out dataset

10:40



## Visual analysis

Seeing the reality of the crisis



## Mapping our data

Understanding spacial relationship

10:40



## Who collects our water

Gender equality with water access



## Crime and water

Linking data to find new insight

Then add the type\_of\_water\_source column to the filter pane and select tap\_in\_home\_broken, to only show data where tap\_in\_home\_broken is shown.

You should see a map of all the provinces in Maji Ndogo! It is very hard to know what we're looking at, so let's improve it by placing an image behind the map. Navigate to the Insert tab on the ribbon, and to the right you can select Image. Find the MD\_map.png on your computer, and select it.





### Introduction

Setting the stage for our data exploration journey.



### Importing our data

Updating our dataset



### Visual analysis

Seeing the reality of the crisis



### Mapping our data

Understanding spatial relationship



### Who collects our water

Gender equality with water access



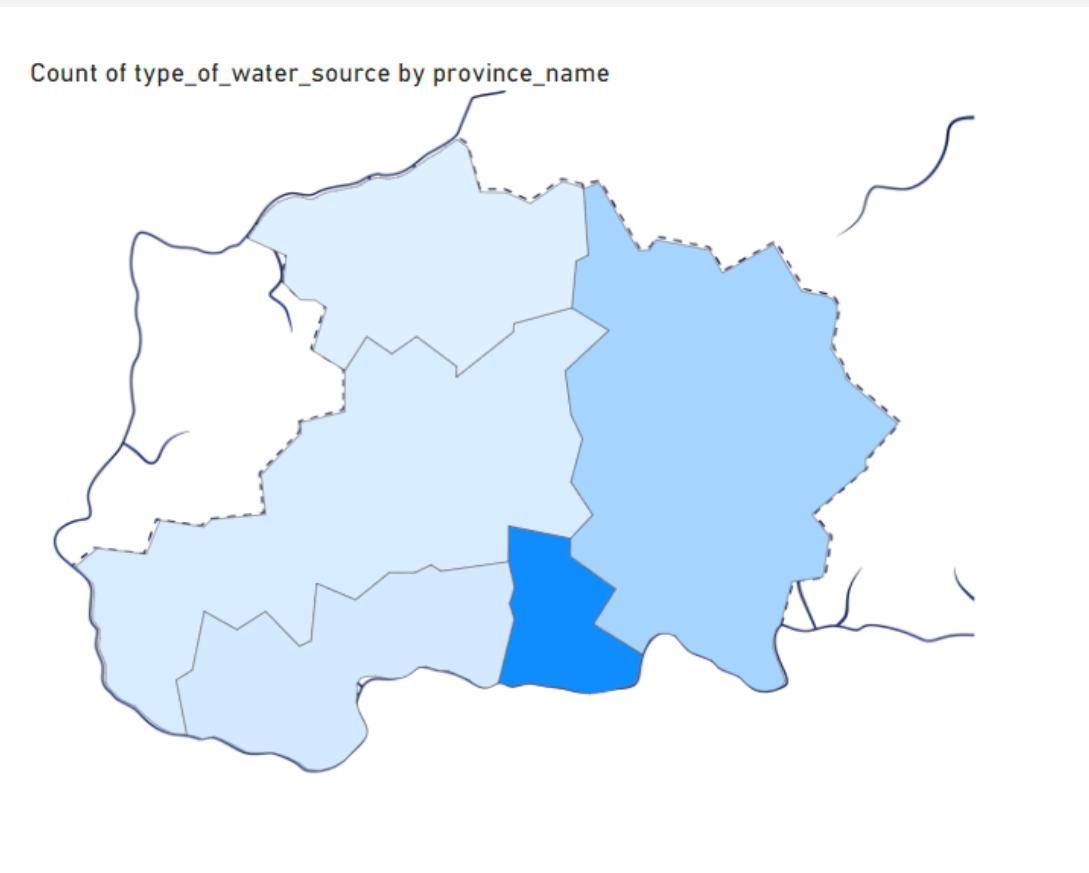
### Crime and water

Linking data to find new insight

The image is displayed over our map, so we need to send it to the background. To do this we select the image, click on Format in the ribbon, select Send backward, and then Send to back.

The final task is to overlay them on top of each other.

This is what it should look like:





Dalila Lesedi  
Online



### Introduction

Setting the stage for our data exploration journey.



### Importing our data

Updating out dataset



### Visual analysis

Seeing the reality of the crisis



### Mapping our data

Understanding spacial relationship



### Who collects our water

Gender equality with water access



### Crime and water

Linking data to find new insight

Phew! 😱 There were a lot of things we had to do to make this map, but now we can see, province by province where the broken infrastructure is!

10:50

Amanzi is at the top of our priority, then Kilimani. Another cool feature is that you can select a province on the map, and it will filter all of the results to only include that province. This is a very intuitive way of adding a filter on a page, without using the filter pane.

11:03

Make sure to name this page too, so you can keep track of it.

11:09



16

**Introduction**

Setting the stage for our data exploration journey.

**Importing our data**

Updating out dataset

11:13

**Visual analysis**

Seeing the reality of the crisis

**Mapping our data**

Understanding spacial relationship

**Who collects our water**

Gender equality with water access

**Crime and water**

Linking data to find new insight

Ok, so let's look at insight 6 on queue times for a minute. Create another new page.

11:13

Create a new Power BI page and add the following visuals:

1. Add a bar chart visual with average queue times on the one axis, and the day of the week on the other axis.
2. Create a line chart that plots the average queue times for the hour of the day, and add the days of the week as different coloured lines.
3. Create a bar chart that shows the average composition of queues for each day of the week. HINT: Plot the average percent\_female, average percent\_male and average percent\_child on the Y-axis and day\_of\_week on the X-axis.
4. Create a scatter plot with number\_of\_people\_served vs. Average of time\_in\_queue. Make sure to add a filter to this plot to include only data where the visit\_count > 1. To use it as a filter condition, add visit\_count to the filter pane, and use an advanced filter.

Dalila, why visit\_count > 1? Well... I had a look and I noticed a data anomaly. If you're curious, before you added the visit\_count > 1 filter, there is a line of values right at the bottom where the queue is 3000 people long but still only 30 mins. We have to look into what happened there. The data is little enough that we can ignore it for now. But it shows you...although we used this data for a while now, by visualising it we could see a problem.

11:18





## Introduction

Setting the stage for our data exploration journey.



## Importing our data

Updating our dataset



## Visual analysis

Seeing the reality of the crisis



## Mapping our data

Understanding spatial relationship



## Who collects our water

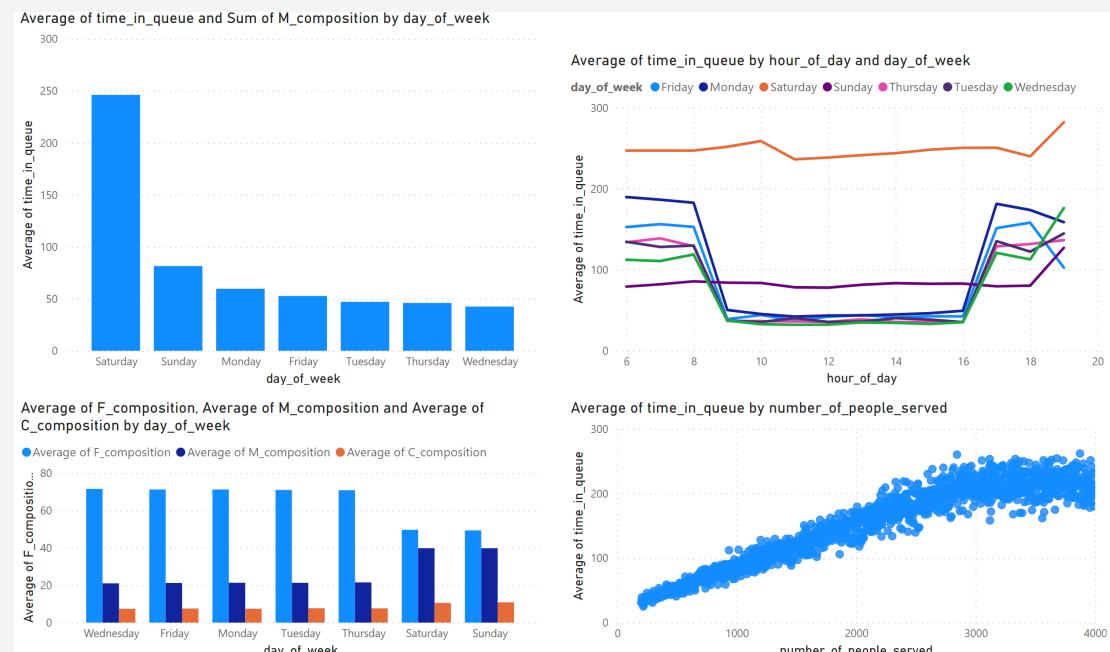
Gender equality with water access



## Crime and water

Linking data to find new insight

Your page should look something like this:



11:26

OK, so once again, these plots are not formatted well, and not laid out well, so we will have to live with a bit of chaos for now. But let's analyse a bit...

11:33





## Introduction

Setting the stage for our data exploration journey.



## Importing our data

Updating out dataset



## Visual analysis

Seeing the reality of the crisis



## Mapping our data

Understanding spacial relationship



## Who collects our water

Gender equality with water access



## Crime and water

Linking data to find new insight



Dalila Lesedi  
Online



So, the queue times per day bar chart really brings the point across that Saturdays have VERY long queues. If you were a citizen, would you get that from this visual? Do we need to change anything to make it simpler or clearer? Would you choose another day to collect water if you saw this? If not, what do you need to see on this visual to get that message?

11:41

Do you see how we need to think about visualisation? It is not just a graph for the sake of a graph. Think through these questions, and apply all of the knowledge from the academy to really critically think about it.

11:48

Anyway... Look at the queue times per hour graph now. Can you see that people queue quite similarly during the week, but weekends are different? During the week, people queue early in the morning, or late in the afternoon, before or after work. On the weekends, people queue throughout the day more consistently.

11:53

Remember back to your time with Chidi. SQL was very good at accessing and assembling data, but Power BI brings it to life! It is much simpler to see trends in PowerBI so it is a great tool for analysis.

11:56

So you may be asking, Dalila... Come on... I can do this in a spreadsheet... You sure could, but select Saturday for a second. Instantly the hourly queue time plots show us the hourly data for Saturday, blown up, and we can get even better insight. If people have to queue on a Saturday, 11 AM to 18:00 AM has 15% shorter queues. You can't do that in a spreadsheet...

12:04

19





← +

Dalila Lesedi  
Online



## Introduction

Setting the stage for our data exploration journey.



## Importing our data

Updating out dataset

12:11



## Visual analysis

Seeing the reality of the crisis



## Mapping our data

Understanding spacial relationship



## Who collects our water

Gender equality with water access

12:17



## Crime and water

Linking data to find new insight

12:23

Next up, let's look at the clustered bar visual that shows the queue composition for the different days. Do you see the gender disparity here? Women queue for water a lot more than men, but we can see over the weekends, men do a lot more. As an analyst that makes me curious... Why do men queue more on weekends? Maybe men are more likely to work full-time, so women are forced to take this responsibility.

I read a report recently that showed that women in Africa are the primary water collectors in Africa. They are more prone to injury because they carry these large containers. Women are often attacked when they collect water early in the morning and late at night. Women are the primary child carers, so children often have to go with their mothers and queue for water. This was my mother, and probably yours too.

And look at the queues... Women queue with their children in Maji Ndogo too. This is why we need to change things.

12:24

Lastly, let's look at the scatter plot. Do you remember that the insights from the survey were that we should install an additional tap if the queue is more than 30 min? Well, this is the visual Chidi used to make that decision. Without a visual, it wasn't easy to explain, but now you can see our thinking.

So we can see the relationship between the queue length and the time it takes is mostly linear. So if you add a tap close to a shared tap serving 3000 people, 1500 of them can use the new one, and 1500 can use the old one. The queue time drops from 200 min to 100 min with just one tap... installing 10 taps means that 300 people use each tap, and we get a queue time of 40 min.

12:26

20



**Introduction**

Setting the stage for our data exploration journey.

**Importing our data**

Updating out dataset

**Visual analysis**

Seeing the reality of the crisis

**Mapping our data**

Understanding spacial relationship

**Who collects our water**

Gender equality with water access

**Crime and water**

Linking data to find new insight

While we looked at this, we noticed that the data actually flattens off a bit (we call that plateauing) around 3000 people served. It puzzled us for a while, but Sanaa came up with a pretty good explanation. If queues are too long, people give up. The queue times don't increase much past 3000 because people see long queues, and go look for water in other places.

12:30



**Introduction**

Setting the stage for our data exploration journey.

**Importing our data**

Updating out dataset

**Visual analysis**

Seeing the reality of the crisis

**Mapping our data**

Understanding spacial relationship

**Who collects our water**

Gender equality with water access

**Crime and water**

Linking data to find new insight

So, remember that report about water access for women I mentioned? Sanaa had a look at some old crime records in Maji Ndogo and managed to link thousands of crimes in the past ten years to specific water source locations. These crimes happened at the locations in our current data set, or as people were travelling to and from that location. Each crime has a unique crime\_id, the type of crime that was committed, the gender of the victim and the location\_id. To make it simpler this week, I added the day of the week, and the hour of the day that the crime occurred.

13:06

Here is the link:

[Md\\_queue\\_related\\_crime.csv](#)

13:03

Next time we will dive into this data a bit deeper, but you should explore it. Visualise the number of crimes committed by type, province and type, split by victim gender. See if you can find a relationship between queues and crimes happening. For example, are there days when the queues are long, and there are a lot of crimes happening as well?

13:44





Dalila Lesedi  
Online



### Introduction

Setting the stage for our data exploration journey.



### Importing our data

Updating out dataset

13 : 59



### Visual analysis

Seeing the reality of the crisis



### Mapping our data

Understanding spacial relationship



### Who collects our water

Gender equality with water access



### Crime and water

Linking data to find new insight

Good luck! I look forward to seeing you next time!  
Remember to collaborate with your team! Together we are more!

