EXPLORE AI ACADEMY

Maji Ndogo: Visualizing the currents of change in Maji Ndogo

Visualising Maji Ndogo's past

Aziza Naledi
Online

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

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

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

Dalila Lesedi
Online

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.

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

Dalila Lesedi
Online

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

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

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`.

08:15

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

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:

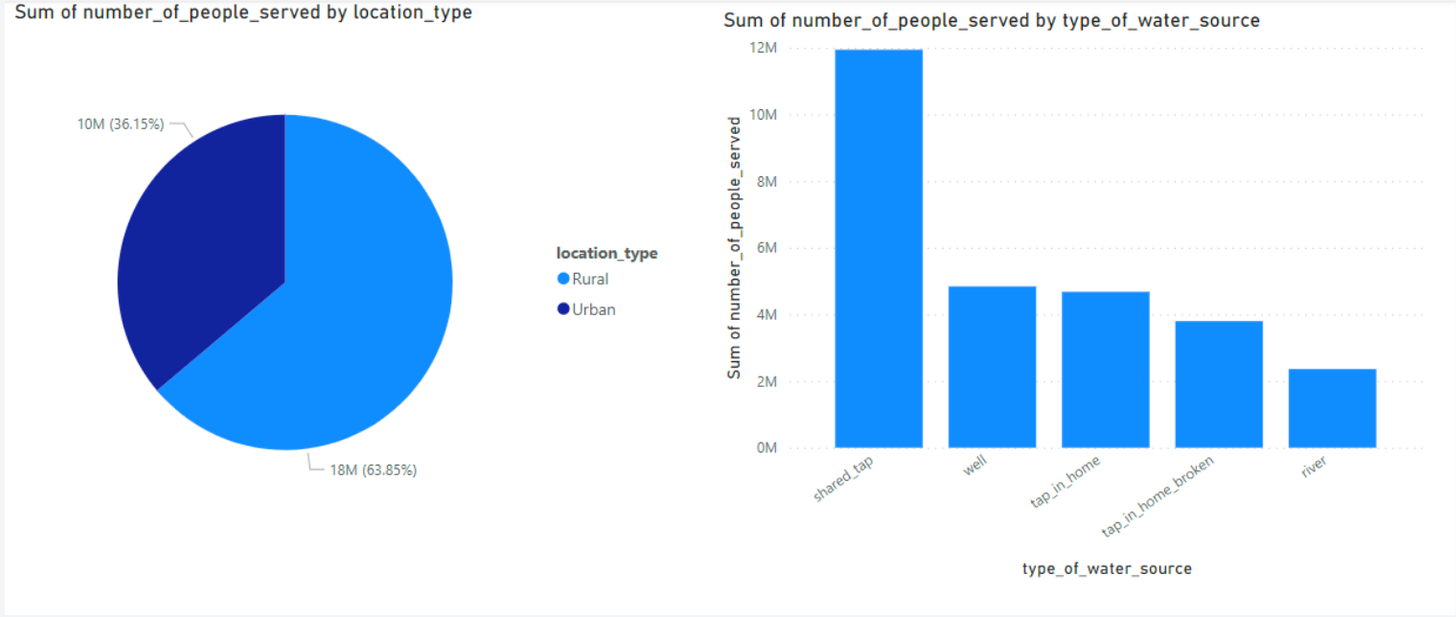
1. Create a pie chart to illustrate the total `Urban` and `Rural` population split in Maji Ndogo. Use `SUM(number_of_people_served)`, split by `location_type`
2. Create a bar (column) chart showing the total population using the various water sources. Use `SUM(number_of_people_served)`, split by `type_of_water_source`

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

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:



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

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

Dalila Lesedi
Online

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.

09:09

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

09:16

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.

09:24

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

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

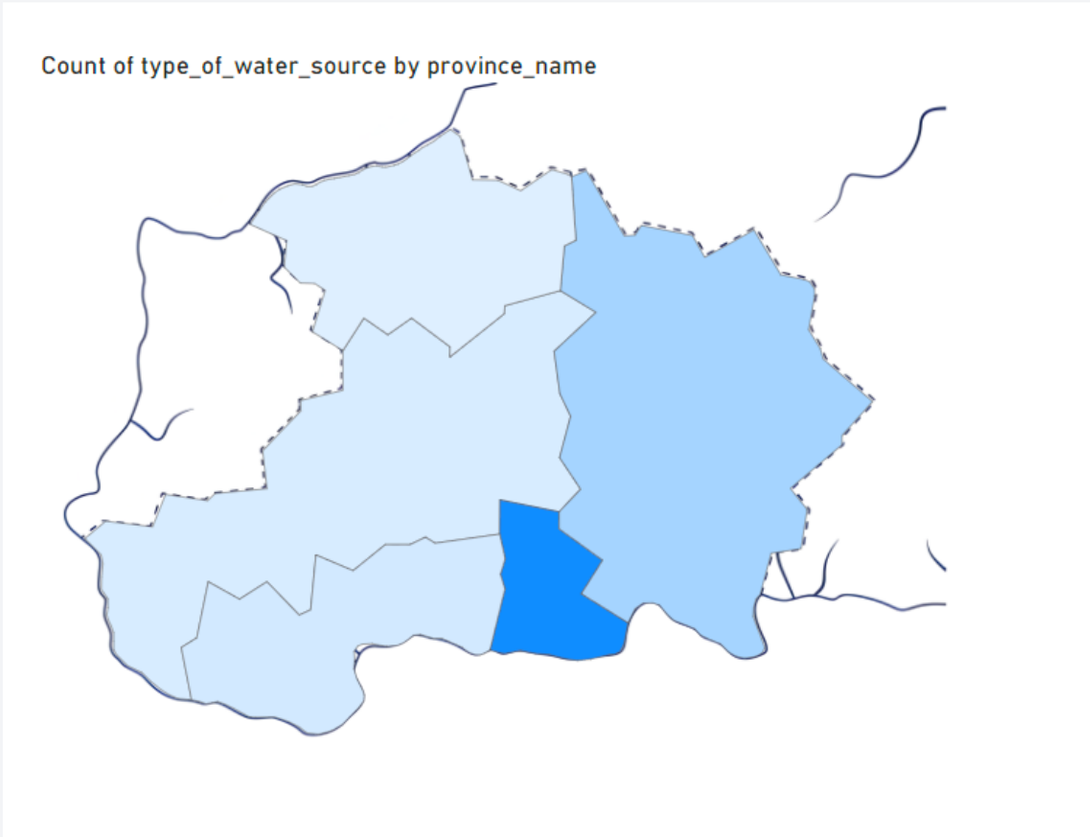
Then add the `type_of_water_source` column to the filter pane and select `tap_in_home_broken`, to only show data where `tap_in_home_broken` is shown.

10:40

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.

10:40

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:



10:47

Dalila Lesedi
Online

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

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

11:13

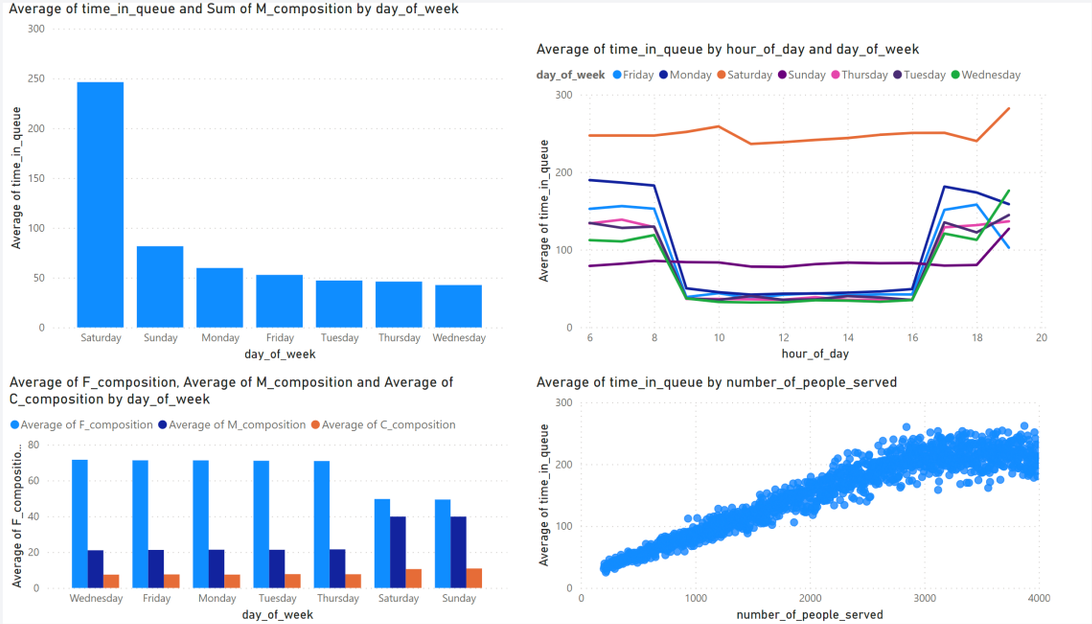
Create a new Power BI page and add the following visuals:

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

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

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

Dalila Lesedi
Online

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.

12:11

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.

12:17

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

12:23

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.

12:24

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

Dalila Lesedi
Online

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

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

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

Dalila Lesedi
Online

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

13:59

EXPLORE AI ACADEMY

Maji Ndogo: Visualizing the currents of change in Maji Ndogo

Moulding data into visual stories in Maji Ndogo

Dalila Lesedi
Online

Hello there! I hope you're doing great!
It's time to level up from our single-table approach and explore how Power BI allows us to integrate and visualise data from various sources seamlessly. Let's get started and see how we can make our data tell a more compelling story. Are you with me?

07:25

We're not going to connect to the SQL database we made but see the link below. I've sent you over an Excel workbook with the tables we'll need. That should do the trick.

Md_water_services_data.xlsx

07:30

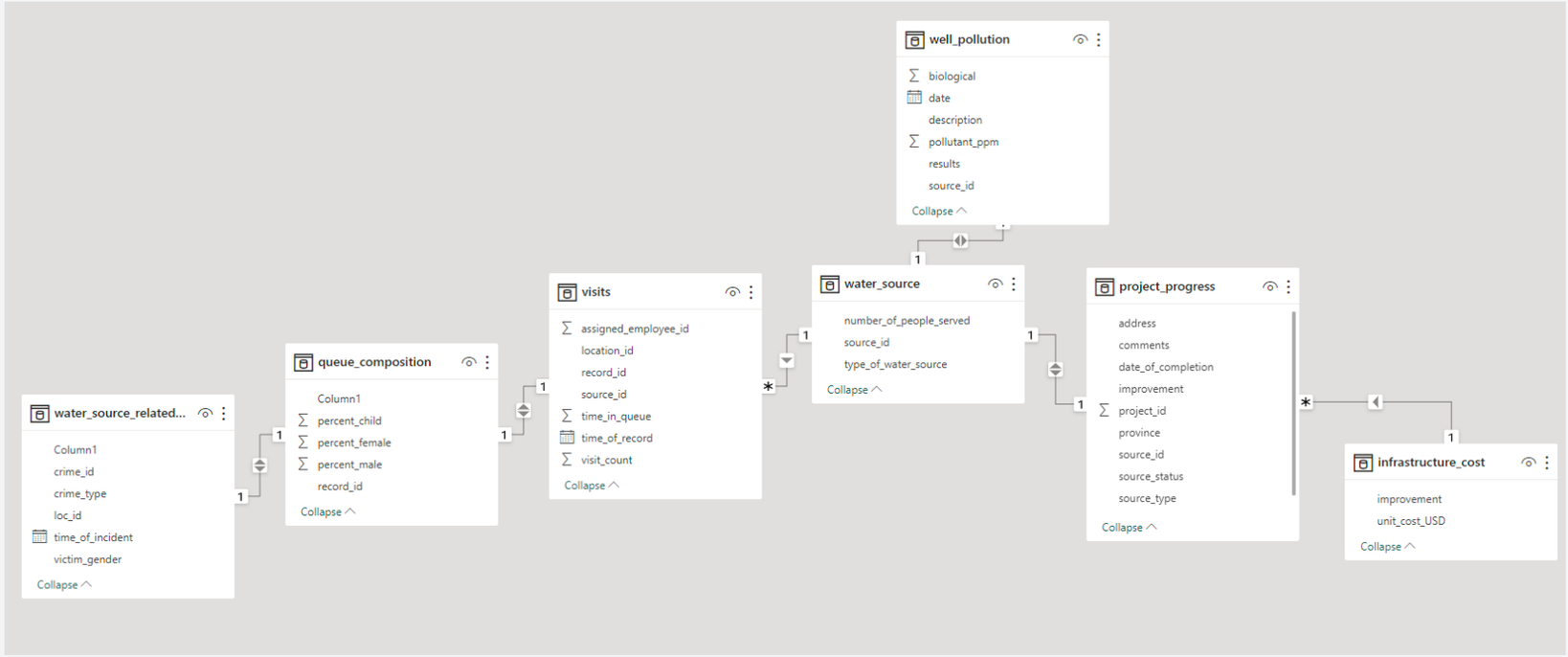
With the Excel file, it's a breeze — one file, multiple tables. Just pick what we need, and we're good to go.

07:37

Create a new Power BI file, and choose the Excel file as the data source. Add all of the tables in the Excel file to the data model.
This might take a while due to the size of the dataset we are importing. If you are using a Virtual Machine (VM) to run Power BI and struggling to import the data, it might help to copy the dataset to the VM rather than using it from your local machine.

07:42

Your model will probably look like this:



07:46

Dalila Lesedi
Online

If you do a quick scan of the model, you should notice several 1-to-1 and 1-to-many relationships, along with directional indicators for each of these connections.

07:49

Have a closer look at `visits`.
Remember from SQL, we used a similar table quite often to JOIN tables together, because it was a "bridging table". Two of our tables are in the model, but they are not correctly imported, and the relations are not established, or the relationship is incorrect. In order to use the data in those tables, we have to fix this first.

07:57

Let's start with `location`. The first issue is that none of the column names (or headers) were imported correctly. So open up Power Query Editor and fix that. As soon as you do this, `location` should update its relations automatically.

08:05

Next up, check out `queue_composition`. You can fix that problem by removing `Column 1` from the model. It was an identifying column from Sanaa's data that we don't need.

08:11

Finally, we need to connect `water_source_related_crime` to the rest of our model.

08:18

First, remove `Column 1` from the table.

08:26

Dalila Lesedi
Online

Now here's the challenge. Power BI doesn't recognise that `loc_id` is related to the `location_id` columns because the name is not exactly the same. But the question is: Which table do we connect it to? Let's take the tech way and just try. 🤷

08:33

So open up the relationship manager and link `water_source_related_crime` to `visits` using `location_id / loc_id`. If you try, Power BI warns you that there is a many-to-many relationship. As you may have seen, many-to-many relationships often cause problems with filtering, so we should avoid this.

08:39

So let's unpack the problem a little. `water_source_related_crime` contains crimes over the last 10 years that happened either at or near the water source `loc_id`. More than one crime could have been committed at a specific location, so `loc_id` is not unique. Similarly, `visits` have multiple visits to some locations, so `location_id` is not unique either.

08:47

This is why there is a many-to-many relationship here. Power BI cannot connect `location_ids` because it cannot figure out which duplicated `location_id` in `visits` is linked to a `loc_id` in `water_source_related_crime`.

08:50

So, how do we solve this? We have two options, one is to create a bridging table that will have unique `location_ids` linking to `visits`, or in our case, we're lucky because one of our tables already has a unique set of `location_ids`!

08:54

Dalila Lesedi
Online

Go sniffing around a bit and figure out which table has a unique set of `location_ids` and link `water_source_related_crime` to the data model.

08:59

Next up, let's look at the relationships.
Many of the relationships in our data model have bi-directional filtering, but most of them are 1-to-1 relationships which is perfectly fine. Remember that 1-to-many relationships with bi-directional filtering can cause problems. So it is best to start off with as many one-directional filters as possible, and if we run into an issue, we will change it.

09:11

Make sure that all 1-to-many relationships have single directional filtering enabled.

09:14

Okay, now we need to make sure all of the data is loaded as we expected in the tables (numbered columns are loaded as numbers, etc.). Spend a couple of minutes going through the tables in the `Table view` tab to make sure the data is correct.

09:18

Because we did a lot of work to clean up the data in SQL, our data looks ready to go.

09:26

Dalila Lesedi
Online

Let's pause a second and unpack the model a bit before we get to visuals.
Ideally, we want a fact table, with details about those facts in various dimension tables. `visits` is our fact table because it recorded thousands of inspections of water sources across Maji Ndogo.

09:33

`location` has details about where the visit happened, `water_source` has details about what they found at those locations, `well_pollution` has more details about some of the sources, `queue_composition` has data related to who we "saw" queue when we visited the location.

09:38

But tables like `project_progress` and `water_source_related_crime` have nothing to do with `visits`. These are fact tables on their own.

09:46

Star schemas in Power BI make filtering and drilling simple, so we always aim for that, but that comes at a cost. All of the data has to be at the same granularity so that the relationships are 1-to-1 or 1-to-many.

09:50

In our case, we have more than one fact table with data describing different things. `water_source_related_crime` links crimes from the past 10 years to a source location, which is unrelated to `visits`. This complex relationship prevents us from making a simple star schema.

09:55

Dalila Lesedi
Online

We could describe our model as a multi-star schema. This is common when we connect data from different parts of our "business". Connecting data that isn't directly related can help us get deeper insight into our data, but it comes at the cost of efficiency.

10:02

Dalila Lesedi
Online

Why did we just do all of this? Well, now our data model is in good shape to help us visualise data effectively in PowerBI. So let's rebuild the visuals we need for the report.

10:08

Create a `National` page and add the following visuals.

1. Map of Provinces in Maji Ndogo.
2. A pie/doughnut chart of the population split between urban and rural.
3. A tree map of the total number of people per source type.
4. Column chart showing the total number of each source of water type, for every town.
5. Column chart that counts the different sources by type.

Remember to set the appropriate filter for `visit_count`, and choose provinces and towns from `location`. Try to use `town` from `project_progress` to see what happens. Also, remember that `Rural` is a "town" in our dataset, so for 3, we need to remove the rural data.

10:14

Introduction
Setting the stage for our data exploration journey.

1 **Buiding a data model**
Unleashing PowerBI with data models

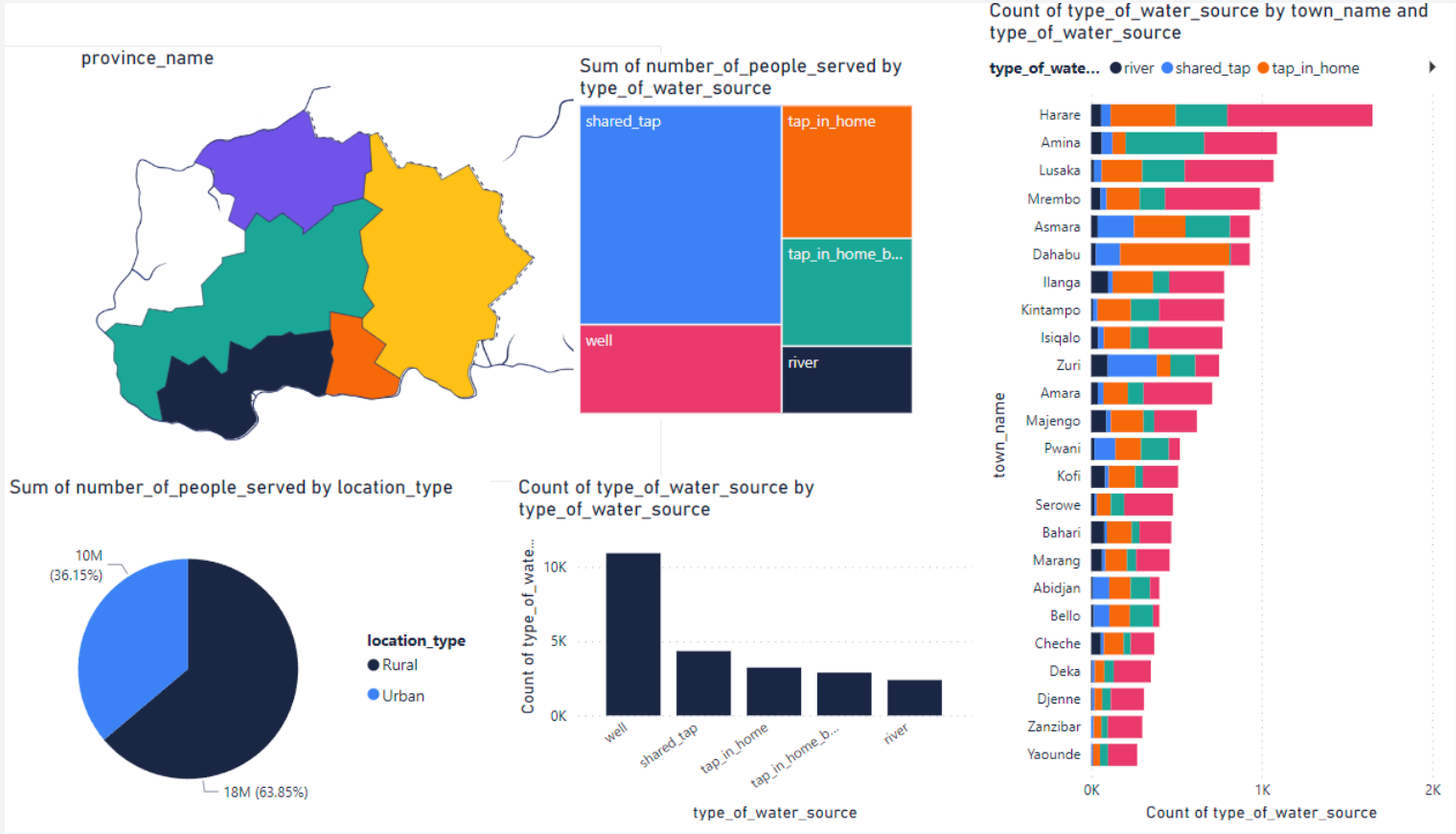
2 **National scale**
Bird's eye of our national survey results

3 **Visualising queues**
Recreating queue related visuals

4 **Cleaning up wells**
Recreating a pollution map

5 **Connecting crime**
Linking crime data to find new insight into water data

This is mine:



10:17

Dalila Lesedi
Online

Now I know... you are yelling ... "MY WATER SOURCE VISUAL DOESN'T LOOK LIKE YOURS!" I know... This is a problem with one of our filters! The location and source tables are linked by lines, but the arrow between `water_source` and `visits` points toward `visits`. That means the filter can only work in that direction.

10:23

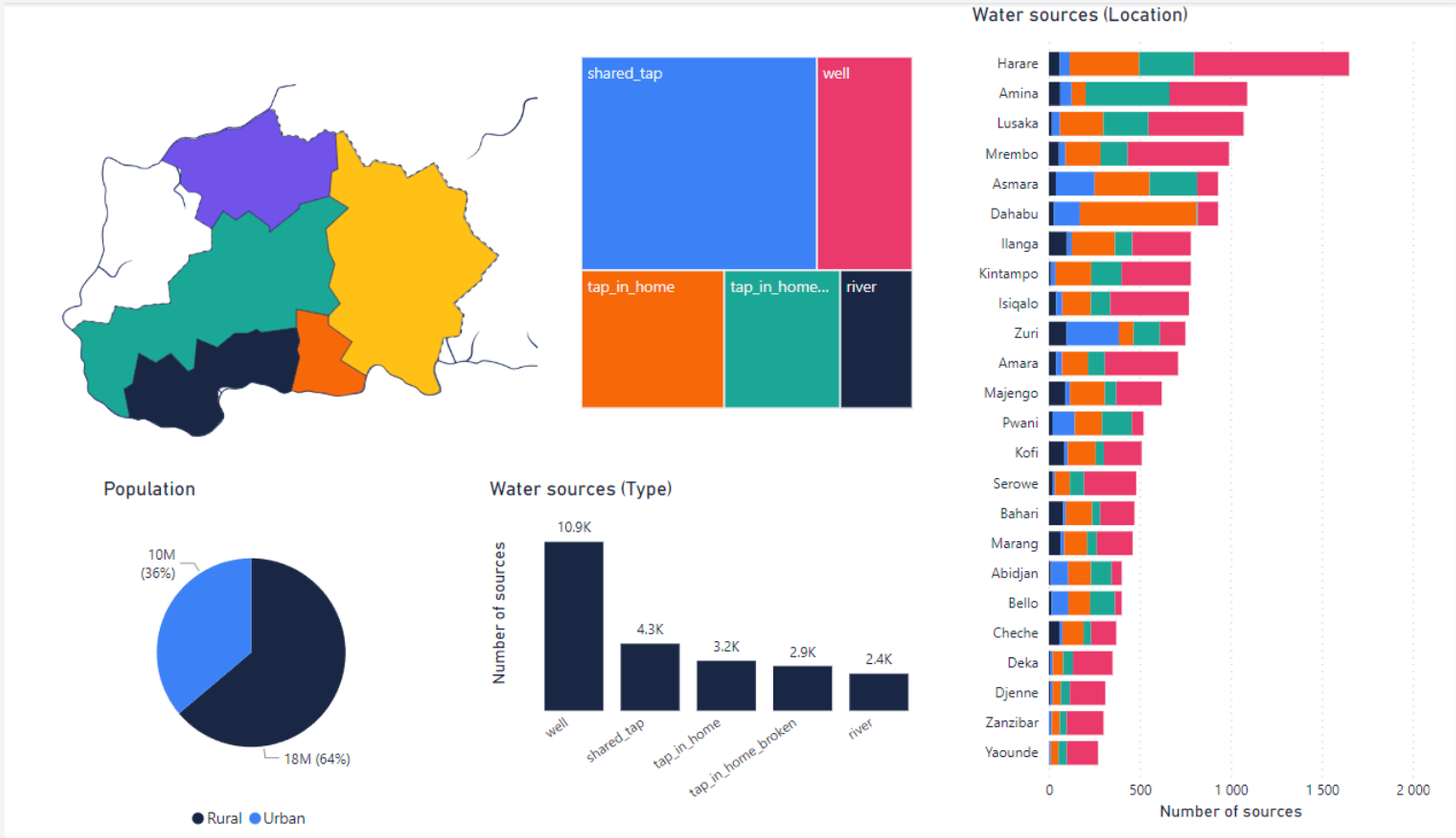
When we create the visual that has `type_of_water_source` as the values and we want to split up that data per province, the filter has to come from `location` (where `province_name` is) to `visits` (which the arrow allows) but then from `visits` to `water_source`, the filter cannot be applied.

10:25

So the fix? We allow filters to pass in both directions. Set the directionality on the `visits-water_source` relationship to bi-directional, and it should work now.

10:23

And once we clean it up a bit we get:



10:23

Dalila Lesedi
Online

Ok, now let's explore what we can do on this page.
With nothing selected, all of the metrics we show are at the National level. When we select one of the provinces on the map, the location type, people served, and water sources all filter.

10:30

When a decision maker looks at this page, they can look at a specific province, and see all of the different results in the context of all the national data.

10:36

The treemap shows us how many people use the water sources per type of source, but its function is actually to select each source type, similar to the function of the map. Filtering on sources can help engineers who are responsible for shared taps, for example, to see what they need to do Nationally, Provincially, rurally, and in each town. Cool right? We're going to keep on working on this later.

10:43

Dalila Lesedi
Online

Create a page where we can focus on the queues and copy the map to this page.

10:51

The visuals we want to include here are the ones related to queue times. If you try one, you're going to run into a problem. In the table we used last time, I calculated the hour of the day, and day of the week for you, but now we have a timestamp. If you use that in the queue time for days of the week, it, gives us the option to look at the data per year, month and day, but we need something different.

10:57

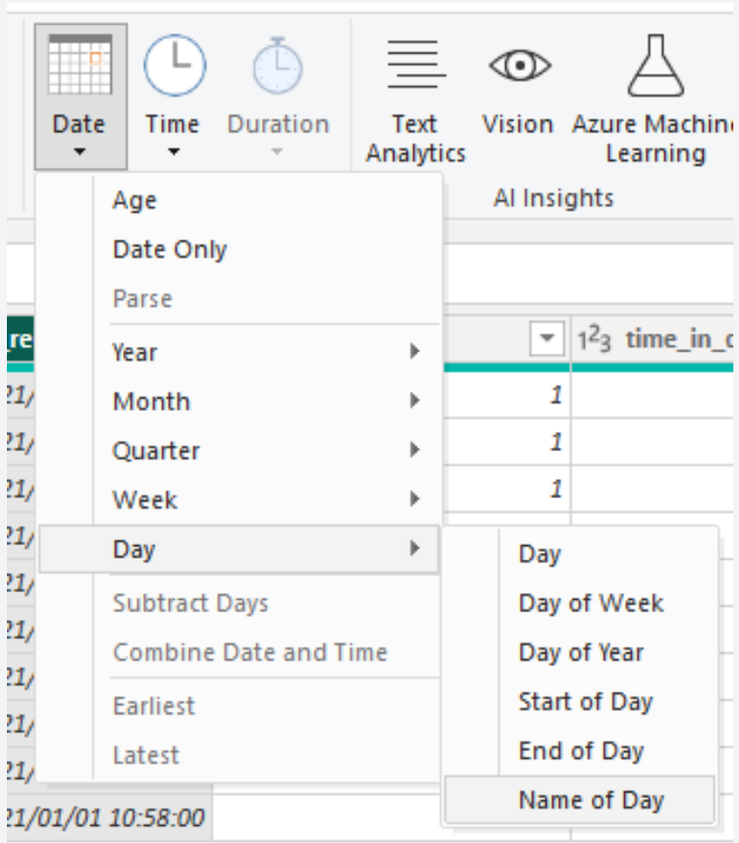
Head back to the data model. We need to extract some information into new columns. I know you've never done this before, but I'll show you how to do it! 😉

11:01

Open Power Query Editor and select the `visits` table. Select the `time_of_record` column and select the `Add Column` tab on the ribbon at the top, then on the `Date` dropdown menu, navigate to `Day`, and select `Name of Day`. That should create a new column called `Day Name`. Rename it to match our naming format.

11:05

This is the menu you should look out for:



11:09

Select the `time_of_record` column again, and create a new column that extracts the hour.

11:13

You should now have two columns:

	day_name	hour_of_day
2	Friday	9
6	Friday	9
0	Friday	9
1	Friday	9
4	Friday	10
0	Friday	10
0	Friday	10
4	Friday	10

11:18

Great! Now Close and Apply, and head back to the visuals page. You should see two new columns under `visits` in the data pane.

11:23

Dalila Lesedi
Online

Now, let's add these plots to the Queues page:

1. Average queue time per hour of the day as a line plot.
2. Average queue time for each day.
3. Average queue composition.
4. Total time queued per province.

Make sure to include only shared_taps in all visualisations on the queue page.

11:29

Introduction
Setting the stage for our data exploration journey.

1 **Buiding a data model**
Unleashing PowerBI with data models

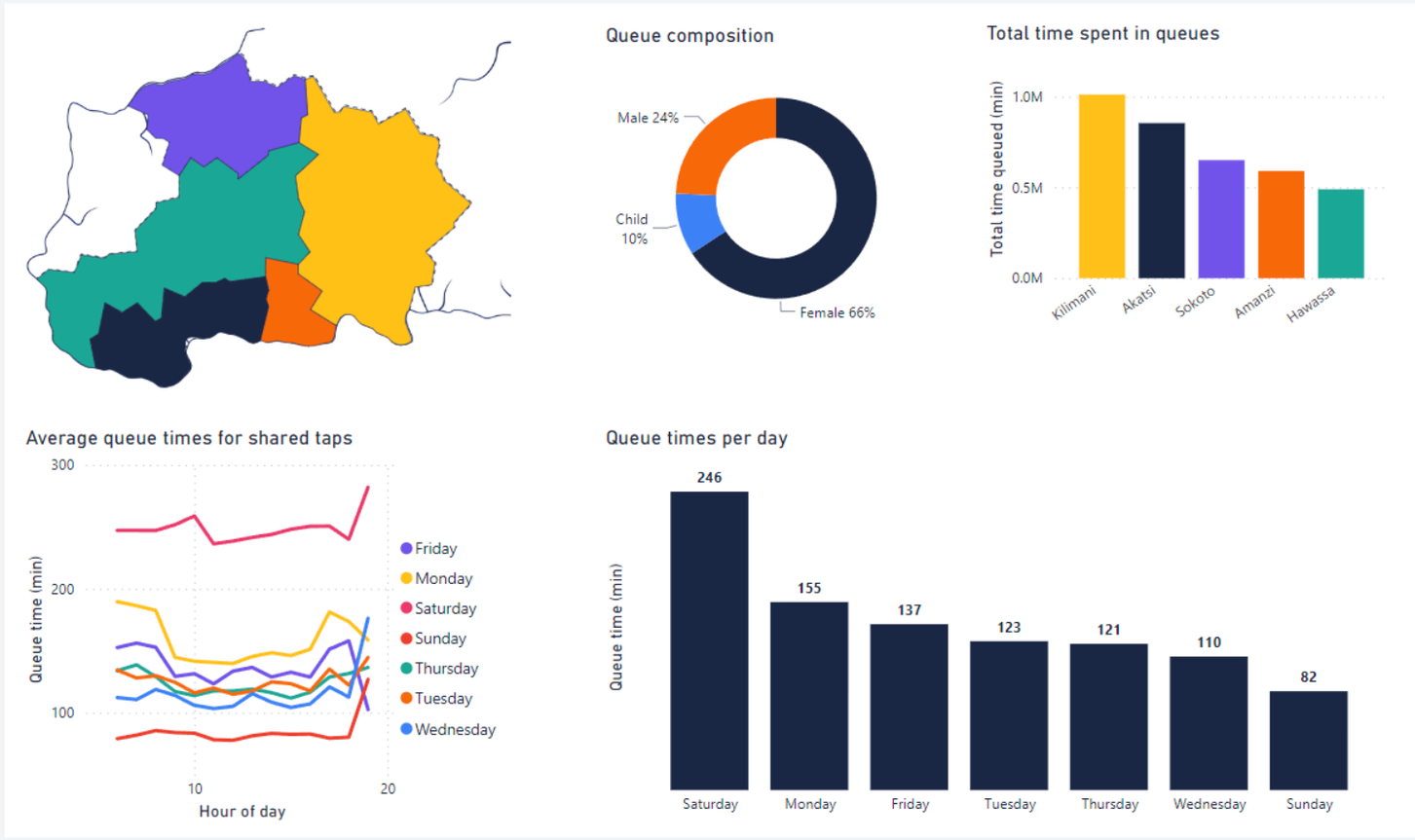
2 **National scale**
Bird's eye of our national survey results

3 **Visualising queues**
Recreating queue related visuals

4 **Cleaning up wells**
Recreating a pollution map

5 **Connecting crime**
Linking crime data to find new insight into water data

This is what I came up with:



11:34

Dalila Lesedi
Online

Now, let's interact with it as a decision-maker would. Selecting one of the provinces now filters all of the data by province. We can see that in Amanzi, the average queue is mostly made up of men, while on a national level, it is mostly women. On Saturdays, queues across Maji Ndogo are 40% men, and on Mondays, 71% women. We can even zoom in on 15:00 on a Saturday and see that the queue has only 3% children, while an hour later, the queue has 29% children!

11:39

We can even combine these filters by CTRL-clicking (SHIFT-clicking on Mac) on more than one. For example, select Sokoto on the map, Monday on the column chart, and 16:00 from the hourly line plot. You should see that 42% of the queue is made up of children!

11:47

Spreadsheets can't do this without a LOT of work, but Power BI is built for this. By modelling the data well, this all happens automatically! Isn't that amazing?! 🤩

11:47

Dalila Lesedi
Online

Next page; Pollution data. Copy over the map again, and edit the map by adding the count of pollution results to the `Color Saturation setting`.

11:49

Add a composition chart that shows the number of results by result; Clean, Chemically contaminated or Biologically contaminated

11:50

Introduction
Setting the stage for our data exploration journey.

1 **Buiding a data model**
Unleashing PowerBI with data models

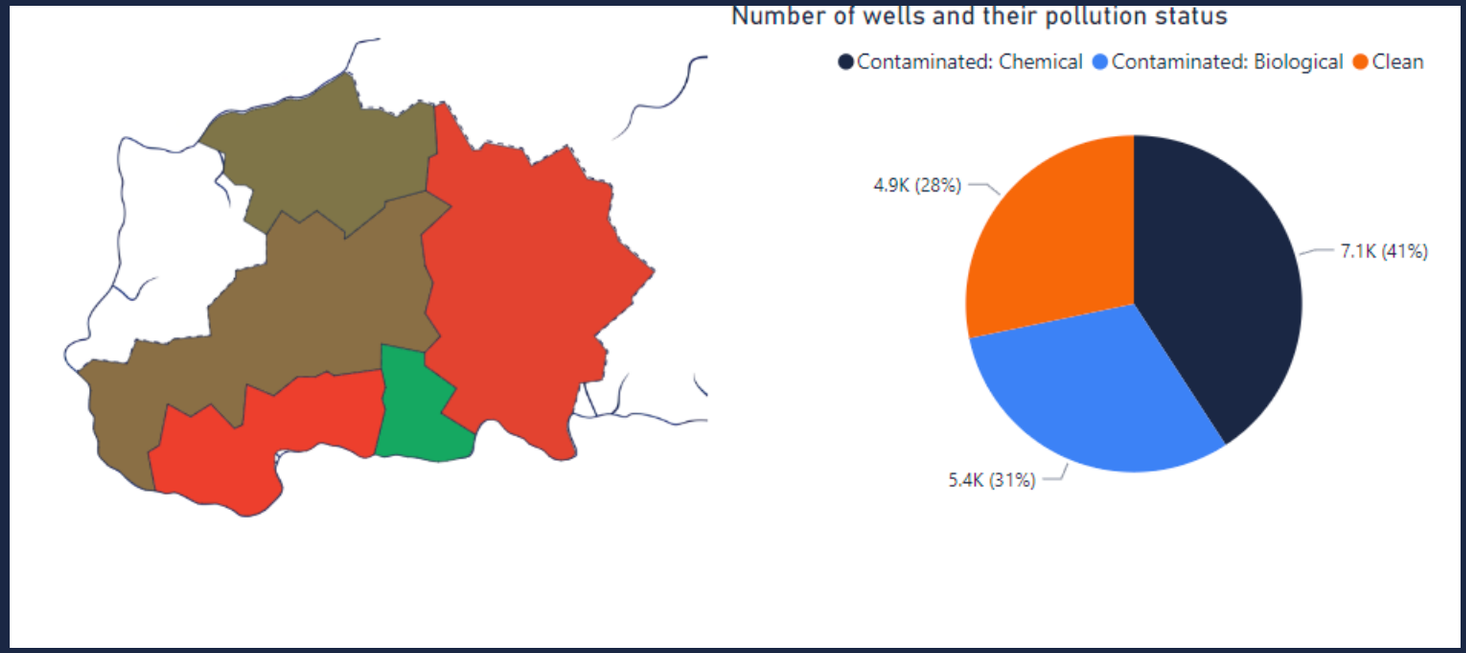
2 **National scale**
Bird's eye of our national survey results

3 **Visualising queues**
Recreating queue related visuals

4 **Cleaning up wells**
Recreating a pollution map

5 **Connecting crime**
Linking crime data to find new insight into water data

This is what you should see:



11:50

Now the map is actually useful. If you select `Clean` we can see that most provinces look bad (green means few wells are clean). It is the opposite way we would think it works, right? That's because green is normally seen as positive, but here green has a negative meaning (few clean wells).

11:52

Dalila Lesedi
Online

If we select biologically contaminated wells, we can see that Akatsi is not too bad (green), but if we select chemically contaminated wells, we can see Akatsi is bright red. So Akatsi has a lot of clean wells, and most of the polluted wells are chemically polluted. If you select Akatsi on the map, the composition chart should confirm that.

11:58

Dalila Lesedi
Online

Just a reminder again; Sanaa pulled data from Maji Ndogo's crime database that were recorded at the locations where water sources are, or where victims reported they were travelling to, or from, water sources when the crimes occurred.

12:05

You may wonder, how does this relate to water access? Well, imagine you are collecting water from a shared source at night. If you do not feel safe, will you go? This is the grim reality faced by many people in Maji Ndogo and in Africa in general. Women and children are often victims at these places, especially in the mornings and at night.

12:08

So, we should check to see if the data in Maji Ndogo also tells the same story.

12:16

Create a new page for the crime-related data. Oh, and pause a second to see if you have named all of your pages so far!

12:24

Create visuals that will show the following:

1. Plot the different crime types to show the gender disparity that women face. Try to identify the crimes that are affecting women in Maji Ndogo.
2. Plot the total number of crimes affecting men, women and children. These are crimes related to people collecting water. What are the patterns you notice?
3. Plot the number of crimes for the time of day and create another visual that shows the number of crimes per day of the week for the different types of victims. Do you note any patterns?
    - You will have to make `hour_of_day` and `day_of_week` columns in `water_source_related_crime` in order to plot these, just like we did for `visits`.
4. Analyse the number of crimes per province. Which provinces have high crime rates, and which provinces have low crime rates, specifically focusing on women?

12:35

Introduction
Setting the stage for our data exploration journey.

1 **Buiding a data model**
Unleashing PowerBI with data models

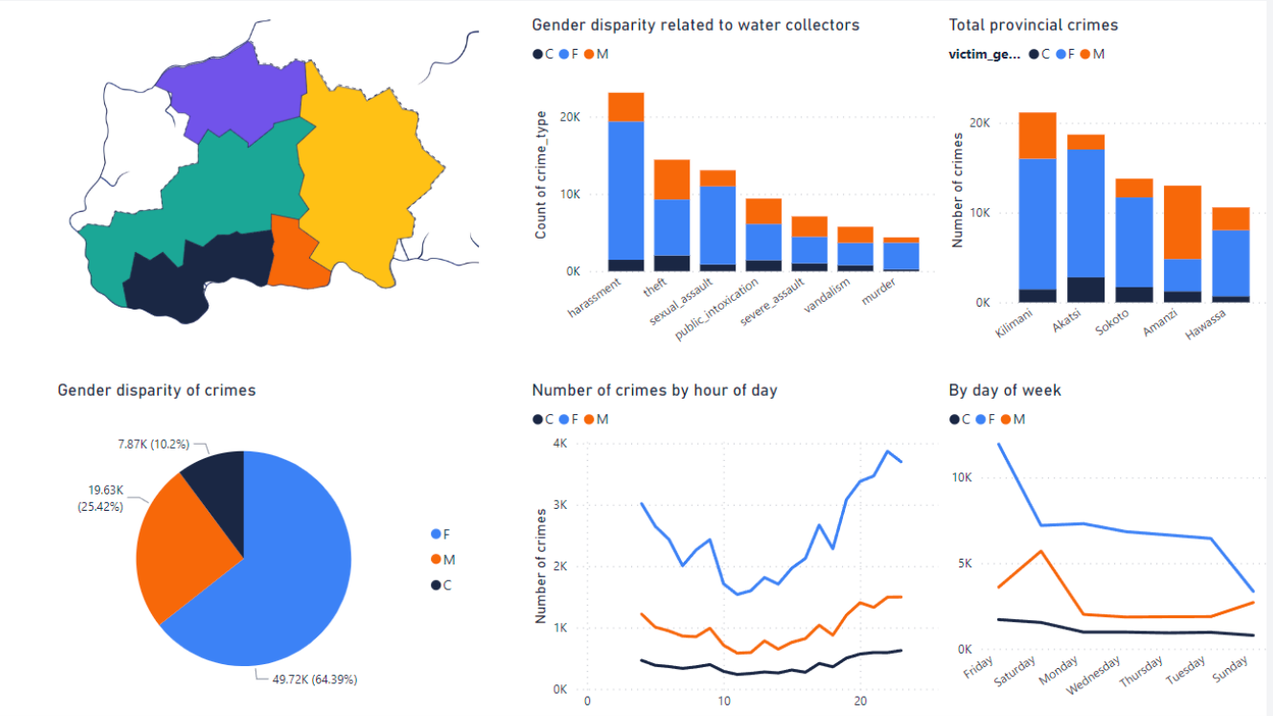
2 **National scale**
Bird's eye of our national survey results

3 **Visualising queues**
Recreating queue related visuals

4 **Cleaning up wells**
Recreating a pollution map

5 **Connecting crime**
Linking crime data to find new insight into water data

Here's mine:



12:40

Dalila Lesedi
Online

See if you can spot these trends I found.

1. As water collectors, women are twice as likely to be a victim of crime than men.
2. Women are most likely to be victims of harassment, followed by sexual assault.
3. Crime spikes over weekends, and almost twice as many crimes are committed early in the mornings or at night with women again facing the greatest threat.
4. In Amanzi, women face significantly less risk of being crime victims. Can you think why this is the case in Amanzi?

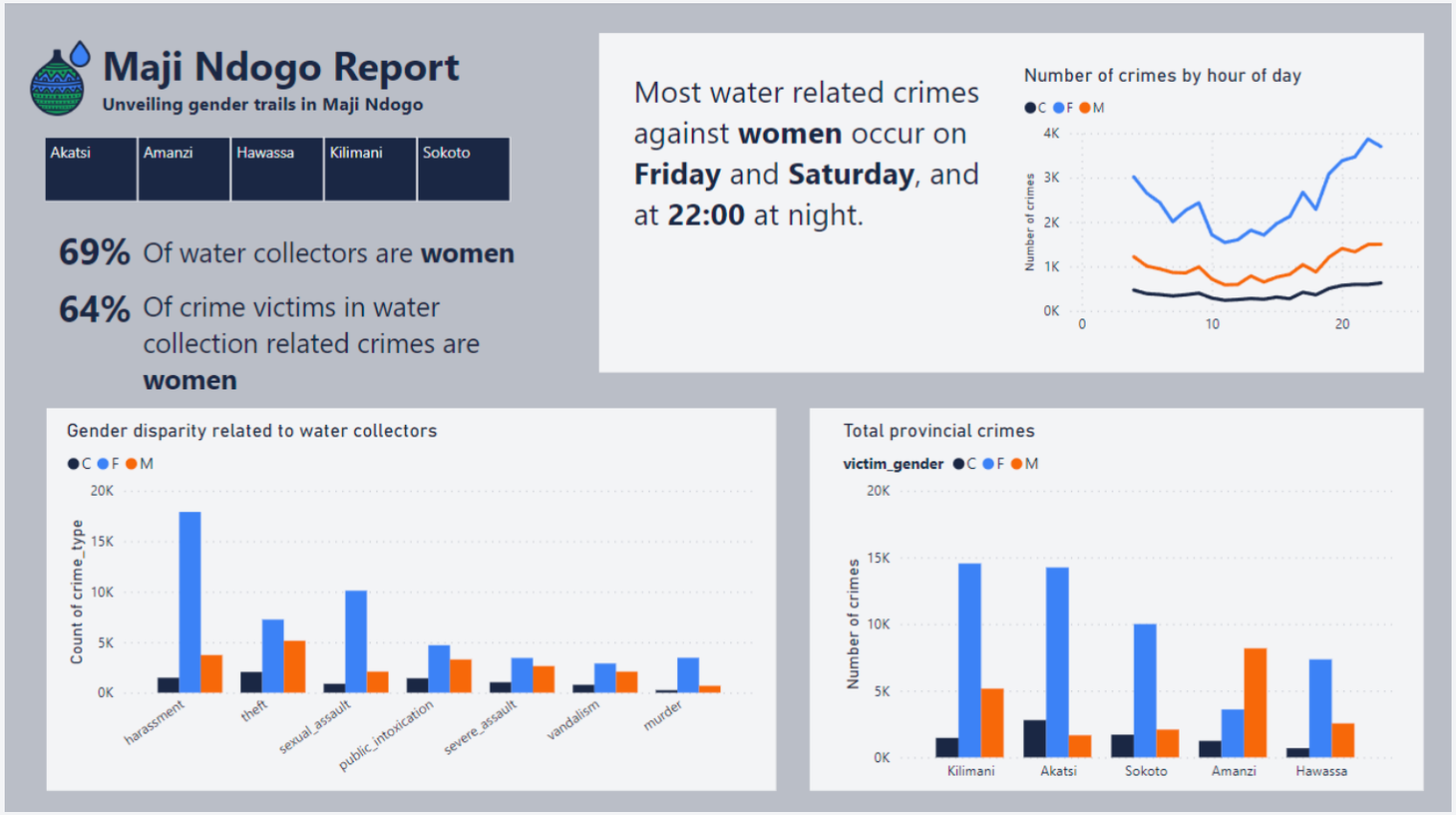
12:44

Dalila Lesedi
Online

Okay, so we're getting close to creating the report, but it still looks messy, and it is difficult for someone new to interpret this information. So, we need to simplify and summarise some of the facts we really want to showcase. For example, our map is cool but for what it does, it takes up a LOT of space. And some facts may be easier if we just report a number that changes when we select a province, instead of plotting the data for each province, and just highlighting the correct column. We're going to need a few more tools, but next time we're going to wrap this project up!

12:50

Side note: This is my idea for the crime data page. It is still a work in progress, but this is the type of stuff we're going to do to make the report come alive. What do you think?



12:55

28

Dalila Lesedi
Online

Until next time, make sure to help out at least one person this week. You may be the one who gets all of this quickly, but remember we are only as good as our teams, so rise by helping others! If you're struggling, your team will help you out, so make sure you reach out!

12:58

EXPLORE AI ACADEMY

Maji Ndogo: Visualizing the currents of change in Maji Ndogo

Communicating our findings in Maji Ndogo

Introduction
Setting the stage for our data exploration journey.

1 **User focus**
Tailoring report design for decision-makers

2 **Data and visuals**
Identifying needs and crafting visual solutions

3 **Data Insights**
Transforming analysis into strategic actions

4 **Provincial Perspective**
Customizing reports for localized decision-making

5 **Enriching data**
Delving into data processing and analysis

6 **Report Creation**
Crafting data narratives for impactful reporting

Aziza Naledi
Online

Dear Dalila and team

How is the little one doing?

It has been a tough week for me. I have been trying to go through the report Chidi's team sent a couple of weeks ago, but it has been hard to understand how all of this fits together. I am struggling to understand the status at the national and provincial levels. How are people getting water? How many people in Maji Ndogo are without basic water? How much capital do we need to make these upgrades and what is it spent on?

Then we really need to see the problem at the provincial level. The provincial leaders I appointed are in the best position to make decisions about their province, so we should have some way for them to understand where they will need to spend their funds.

Would you mind focussing on creating a report that will help us see this data so we can get some of these projects started?

Best regards,
Aziza

07:25

I'm sure you saw that Pres. Naledi gave us some clearer instructions for the report. It often happens that the people we're supporting have changing needs, and we have to adapt.

07:32

We're creating this report to support decision-makers by providing them with accurate and understandable data relevant to their decisions. But what do they want? The best way of knowing is by putting yourself in their shoes. If we were the users, how would we want to interact with the dashboard? This is called a user story, which answers the question: "As a user, I want to..." and designing the report to meet that need.

07:39

So, let's put ourselves in the shoes of one of our users. Imagine you are **Pres. Naledi** for a moment.
As Aziza, I want to:

1. See the key points of the survey results so I understand the overall status of water access in Maji Ndogo.
2. Know how many people are affected by water access challenges in Maji Ndogo and what those challenges are.
3. Know how much money we will need to complete the upgrades, and where that money needs to be spent.
4. Understand this data on a national level and a provincial level.

07:47

If we use this user story as a blueprint to create a page in the report for Pres. Naledi, we can create visuals for each of her concerns. Doing this automatically answers the "Why do we need this visual?" question.

07:51

Next, for each of these user stories, we have to think about **what data** the user will need to answer that question.

07:58

So let's apply the **what**.

1. I want to see the key points of the survey results so I understand the overall status of water access in Maji Ndogo.
    - We should summarise population-related access to water on a national and provincial level.
    - A lot of people in Maji Ndogo live in rural and urban areas, so we should summarise their results and the challenges they face separately.
2. I want to know how many people are affected by water access challenges in Maji Ndogo and what those challenges are.
    - We need to communicate the number of sources, and also the amount of people affected by these sources.
    - Perhaps we should also consider making the less important sources like `tap_in_homes` less visible since we're not upgrading those.
3. I want to know how much money we will need to complete the upgrades, and where that money needs to be spent.
    - This one will need a bit of work from us. There is some financial data in our data model called `infrastructure_cost`, but we will need to do some calculations to include this type of data.
    - Pres. Naledi will need to see this data at a national, provincial, and rural/urban level, and even per improvement type.
4. I want to understand this data on a national level and a provincial level.
    - We have to take this into account, as well as at a rural/urban split.

08:04

Then **how** will we visualise that data? We may change this as we go, but let's start with a set of visualisations, and we'll add more, or remove some as we go along:

1. I want to see the key points of the survey results so I understand the overall status of water access in Maji Ndogo.
    - We can split the page into rural vs. urban, and then show the population of Maji Ndogo, and highlight key metrics.
2. I want to know how many people are affected by water access challenges in Maji Ndogo and what those challenges are.
    - Here we show the distribution of source types in rural vs. urban populations
    - A sorted visual of the number of source types.
3. I want to know how much money we will need to complete the upgrades, and where that money needs to be spent.
    - We need to show a single number representing the data nationally, and per province.
    - We should also break down costs in various ways:
        - per province
        - rural/urban
        - source type
4. I want to understand this data on a national level and a provincial level.
    - If we build the report well, we can drill down into different categories.

08:11

Dalila Lesedi
Online

Our second user is a **provincial leader**. Their user story is: As a leader of a particular province, I want to see data only relevant to my province. I need to understand the state of water access in my province, the scope of work we have to complete, and understand the financial aspects related to the improvements.

08:18

We will need to show the following data:

1. Total people served for each water source type in a province.
2. Number of water sources, their type, and whether it is rural or urban.
3. Show the relevant stats for towns in that province.
4. Add relevant provincial data. Queues, gender compositions and crime, broken taps by town, etc., in provinces where it is relevant.
5. Summary of improvements and costs.

08:26

So we're going to build the first page together, and you have the honour of wrapping up the rest of the pages.

08:33

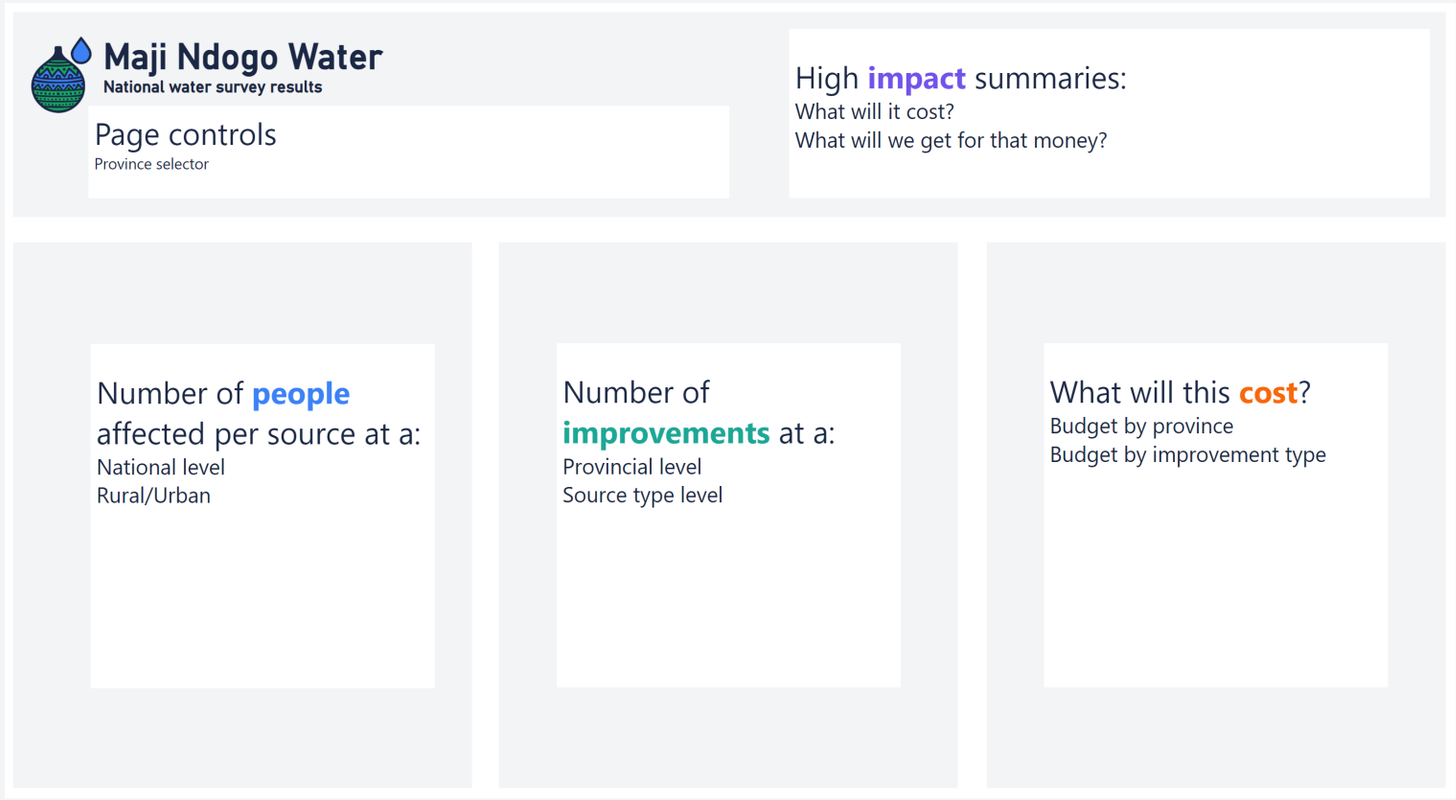
Let's start at the top with Pres. Naledi's page. If we look back at the story, we see 3 components; Population/water breakdown, "What do we need to do?", and "How much will it cost?". Let's split our dashboard into three components, but also reserve some space for slicers somewhere, and reserve some space for the big idea results that are linked to the purpose: "How much do we need to spend in Maji Ndogo, and what will that money do?"

08:38

1. **Introduction**
   Setting the stage for our data exploration journey.
2. **User focus**
   Tailoring report design for decision-makers
3. **Data and visuals**
   Identifying needs and crafting visual solutions
4. **Data Insights**
   Transforming analysis into strategic actions
5. **Provincial Perspective**
   Customizing reports for localized decision-making
6. **Enriching data**
   Delving into data processing and analysis
7. **Report Creation**
   Crafting data narratives for impactful reporting

This is my thinking. You are welcome to build your's like mine, but try to do something different if you're comfortable.

08:45



08:53

Introduction
Setting the stage for our data exploration journey.

1 **User focus**
Tailoring report design for decision-makers

2 **Data and visuals**
Identifying needs and crafting visual solutions

3 **Data Insights**
Transforming analysis into strategic actions

4 **Provincial Perspective**
Customizing reports for localized decision-making

5 **Enriching data**
Delving into data processing and analysis

6 **Report Creation**
Crafting data narratives for impactful reporting

I grouped similar things in boxes, and the three different boxes tell this story:

- Who are we doing this for and where are they?
- What do we need to do, and where do we need to do it?
- What will this cost?

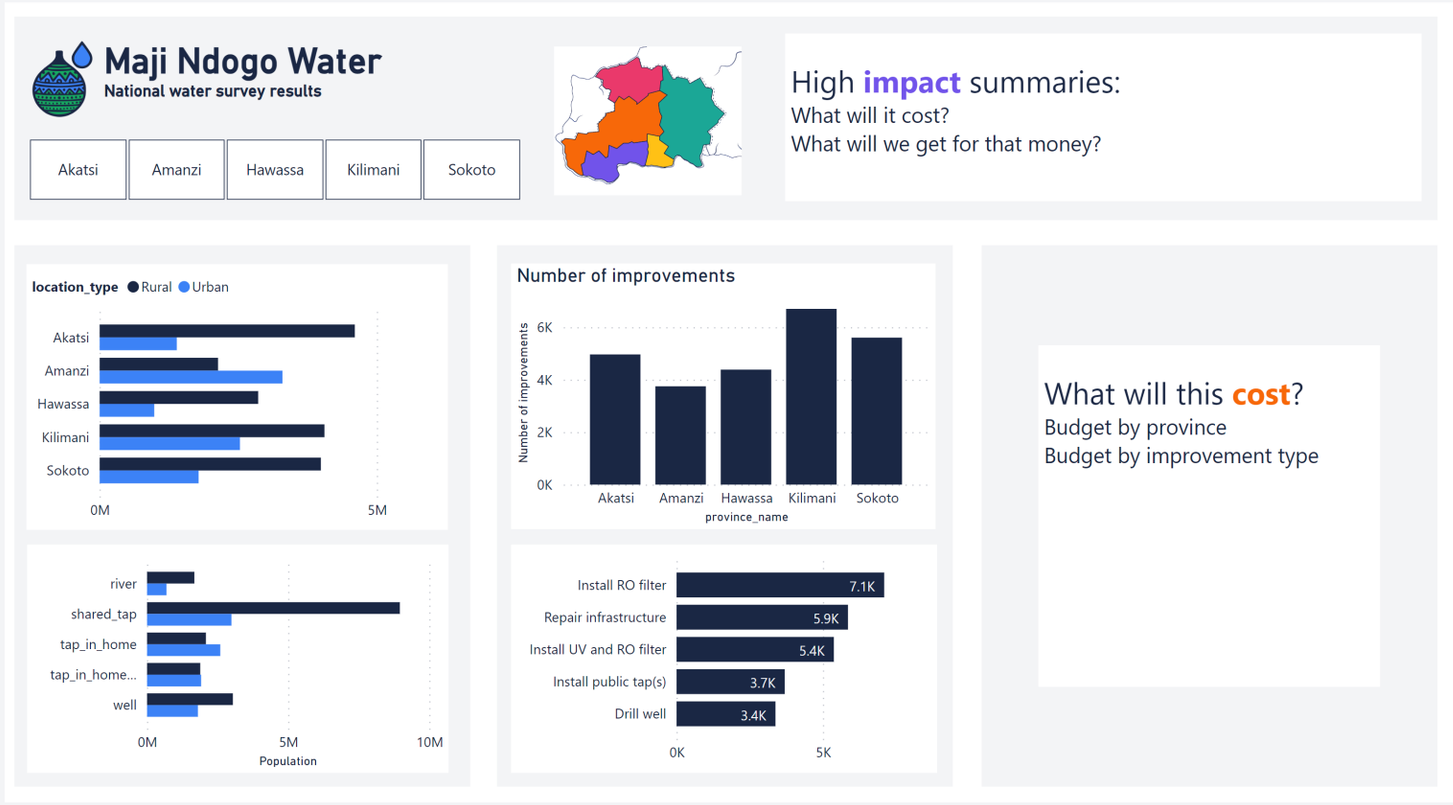
The data story is natural in this order. We don't start with cost, then who we're improving water for, and end with what we want to do. Make sense?

09:01

So... go ahead!! Create a new visuals page, `National`, split it up like mine if you want, create a title, and start adding visuals! 📊 Also, add a slicer to the controls section so that we can select provinces.

09:05

I also added a map. So now, when we select a province in the slicer, the map highlights where that province is.



09:09

Do you notice that your improvements column chart has more categories than mine? I used a bit of DAX to aggregate all of the Install 1-8 taps values into one called `Install public tap(s)*`, and shortened Diagnose local infrastructure too. Do you see the * in "Install public tap(s)*"? It is because the meaning of this category isn't quite the same now that we've aggregated it. What does a value of 3696 for Install public tap(s) mean? You should add a short note explaining it somewhere nearby, or use a tooltip! 💬

09:12

Here is the DAX I used in `project_progress`. I added a <BLANK> value in there so you can figure out what you need to use:

```
Aggregated_improvements =
IF(
    CONTAINSSTRING(
        'project_progress'[improvement], "<BLANK>"
    ),
    "Install public tap(s)*",
    IF(
        'project_progress'[improvement] == "Diagnose local infrastructure",
        "Repair infrastructure",
        'project_progress'[improvement]
    )
)
```

09:17

Up to this point, we only used our data as-is. To make more impactful visuals, we need to do some calculations and summarise data.

09:29

Dalila Lesedi
Online

Let's start off with the budget. No surprise; these improvements will cost a lot of money. To make sure we are spending it wisely, we will create a budget.

09:33

The project managers added an `infrastructure_cost` table that has an estimate of what it will cost to make that repair. For each `improvement` in `Infrastructure_cost`, there is a cost of improving a water source in US Dollars. For example, to improve one river source, we have to drill a well, which will cost $ 8500.

09:36

Our project planners thought through the costs a bit more and realised that rural water sources are much harder to improve than urban ones. So they suggested that we add 50% to the cost of any rural source's improvement budget. So we need to increase the `Budgeted_improvement_cost` by 50%. We'll call that `Rural_adjusted_cost`.

09:43

Create a new column called `Rural_adjusted_cost` and calculate the increased budget for each of the improvements.

09:49

We will start off slow today... Here's my calculation:

```
Rural_adjusted cost =
    infrastructure_cost[unit_cost_USD] * 1.5
```

09:56

Dalila Lesedi
Online

Next up, `Budgeted_improvement_cost`. Now that we know the cost of each improvement, let's update the `project_progress` table with a new column that calculates the cost of each improvement, based on the type of improvement, and whether it is rural or not.

10:00

Create a DAX formula that looks up the improvement in the `project_progress` table, and either fetches the `unit_cost_USD` value, if the source is in a town, or fetches `Rural_adjusted_cost` if the source is in a rural area.

10:06

Dalila is letting go of your hand now! This is what I used, but I added some <BLANK> values for you to fill in:

```
Budgeted_improvement_cost =
    IF(
        'project_progress'[<BLANK>] == "<BLANK>",
        <BLANK>('infrastructure_cost'[Rural_adjusted cost]), <BLANK>('infrastructure_cost'[unit_cost_USD])
    )
```

10:12

The next column we will need to create is `Basic_water_access`. How will we know if we succeeded in the project? "Dalila, come on..." I hear you say, "When people have water!" Yes... But how do we measure that?

10:21

We can classify water sources according to the UN requirements:

Service level	Definition
Safely managed	Drinking water from an improved source that is **accessible on premises**, available when needed, and **free from** faecal and priority chemical **contamination**.
Basic	Drinking water from an improved source, provided **collection time is not more than 30 minutes** for a round trip, including queue time.
Limited	Drinking water from an improved source, for which **collection time exceeds 30 minutes** for a round trip, including queue time.
Unimproved	Drinking water from an **unprotected** dug **well** or unprotected **spring**.
Surface water	Drinking water **directly from** a river, dam, lake, pond, stream, canal, or irrigation canal.

10:27

In order for a water source to be classified as **basic**, it must satisfy these requirements:

- Rivers are not improved sources, so they are not included.
- Wells are improved, **only** if they are clean. So all contaminated wells are excluded.
- Public taps are improved sources, but **only** if the queue time is less than 30 min.
- Broken infrastructure "taps" are not basic, because they do not work.
- All taps installed in homes across Maji Ndogo are at least basic. They are actually safely managed.

10:34

Dalila Lesedi
Online

Once we know the water access level, we can estimate how many people's lives we will improve by upgrading all of the water sources.

10:41

So we need to first calculate where our access level is, and as we make repairs, we will hopefully get that number close to 100% access. This is our goal! So if we spend money on this project, we can measure how water access improves, to make sure what we're doing is making a difference.

10:48

So the idea is to calculate the number of clean wells, the number of public taps with queue times < 30 min, and taps in homes, then use that to calculate how many people have access to basic water. We could do that directly, but including all the conditional flow logic will complicate things a little too much.

10:55

So we break it into three steps:

1. Classify sources as `Basic Access` or `Below Basic Access` in a new column. Which table should we use?
    - For wells, we have to check if it is polluted from the `well_pollution` table, which is just a lookup.
    - But to get the queue time of a shared tap, we have to remember that shared taps were visited multiple times, so it is not as simple as fetching the queue time for each row, because there are multiple entries for each `source_id`.
2. Once we have that, we can sum up all of the people using a basic water source, divided by the total population.

11:00

Step 1 is complex, so let's break it down together a bit more. Step A, we have to calculate the average queue time for each source. This means we need to aggregate per source.
Step B, we need to create the control flow logic to check wells and taps to classify them as being basic or not.

11:03

Step A: Let's calculate the average queue time for each water source first in the `water_source` table. Remember, we want to aggregate queue times - using the average - for each `source_id` so that the average queue times are calculated for each source.

11:08

This is what I used, but I added some <BLANK> values for you to fill in:

```
Average_queue_time =
    CALCULATE(
        <BLANK>('visits'[<BLANK>]), // Calculate the average time in queue
        FILTER(
            'visits',
            'visits'[<BLANK>] = 'water_source'[<BLANK>] // Filter the visits for each specific water source
        )
    )
```

11:12

Ok, now we should have a new column with average queue times for each water source!

11:22

Step B: We classify each source as being basic or below basic.

11:26

This is mine, but I added some <BLANK> values for you to fill in:

```
Basic_water_access =
    IF(
        AND( // Checks if the water source is a well and if its clean
            <BLANK>,
            <BLANK>
        ),
        "Basic Access", // Assigns 'Basic Access' if both conditions are true (clean well)
        IF( // Checks if the water source is a tap inside the home
            <BLANK>,
            "Basic Access", // Assigns 'Basic Access' if true (tap in home)
            IF(
                AND( // Checks if the water source is a shared tap and the average queue time is less than 30 minutes
                    <BLANK>,
                    <BLANK> < 30
                    ),
                "Basic Access", // Assigns 'Basic Access' if both conditions are true (shared tap with queue < 30 min)
                "Below Basic Access" // Assigns 'Below Basic Access' if none of the above conditions are met
            )
        )
    )
```

11:30

Ok, now the final step is to create a new measure that calculates the basic water access level as the number of people who have basic access to water divided by the total number of people in Maji Ndogo. Remember to format this as a % and show no decimals.

11:38

Dalila Lesedi
Online

And a last-last step! The basic access percentage tells us how many people have access to basic water services, but we might also need the percentage of people who don't have access to basic water too. If you think about this, we want that number to be 0 once we're all done, right?

11:47

Ok, so now we have calculated two things; how much will all of these upgrades cost, and what will be the impact? I think we're ready to complete the report now!

11:52

Now in the third block we want to give the President all of the financial information she will need to make decisions. She will want to understand the costs at a provincial level, and also understand what the different improvements will cost us.

11:59

Since Pres. Aziza will have to send the local provinces some funds, we really have to give her numbers too! So we will create a budget table. The downside is that tables struggle to show a part-of-a-whole story, let's visualise the budget too.

12:04

Add budget tables that have the breakdown of costs per province and per improvement type. Add a visual that will show the percentages of the budget allocated to each province.

12:08

Doesn't fit into that block well does it? 🙈
Instead of trying to cram all of the information into the third block, let's make a bookmark! We can make a button that will toggle between the `Province` and `Improvements` tables. This way the user can choose which data to focus on.

12:12

First, here are some images I made for the buttons:

Buttons.zip

12:20

We will create two bookmarks, `Province` and `Improvements`. For the `Province` bookmark, we will hide the `Improvements_table`, and for `Improvements` we will hide the `Province_table`, and we will use the buttons to toggle between these two bookmarks.

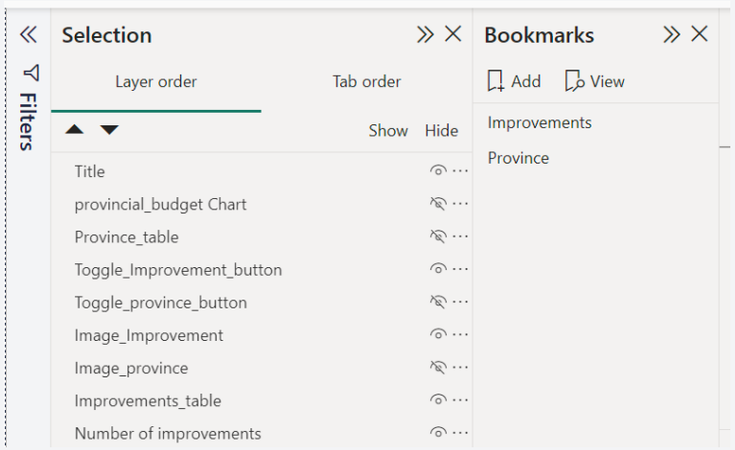
12:26

Use the selection and bookmarks panes to add two buttons that toggle between `Province` and `Improvements` bookmarks. Hints:

1. Remember to name the visuals, buttons, and images properly.
2. Add two buttons and add the two images as images.
3. Use the selection pane to display and hide the elements we don't want to see in a bookmark. See the image below for what I display in my `Improvements` bookmark.
4. Before connecting the functions to the buttons, make sure that everything works when you toggle between bookmarks.
5. The button showing in the `Province` bookmark should open the `Improvements` bookmark and vice-versa. Use this to set the actions.
6. Make sure the buttons are in front of your images. Move elements up and down the selection pane to do this.
7. Test your buttons to see if they work.

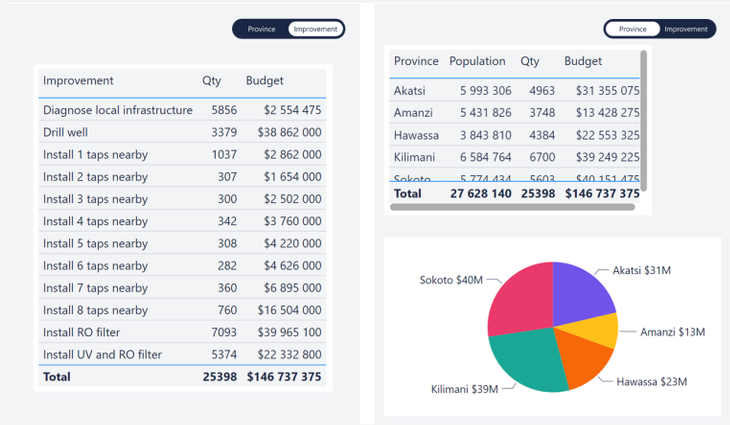
12:29

This is my `Improvements` bookmark: (Note which of the visuals are displayed.)



12:33

When it is all done, your two panels will look something like this:



Province | Improvement

Improvement	Qty	Budget
Diagnose local infrastructure	5856	$2 554 475
Drill well	3379	$38 862 000
Install 1 taps nearby	1037	$2 862 000
Install 2 taps nearby	307	$1 654 000
Install 3 taps nearby	300	$2 502 000
Install 4 taps nearby	342	$3 760 000
Install 5 taps nearby	308	$4 220 000
Install 6 taps nearby	282	$4 626 000
Install 7 taps nearby	360	$6 895 000
Install 8 taps nearby	760	$16 504 000
Install RO filter	7093	$39 965 100
Install UV and RO filter	5374	$22 332 800
Total	**25398**	**$146 737 375**

Province | Improvement

Province	Population	Qty	Budget
Akatsi	5 993 306	4963	$31 355 075
Amanzi	5 431 826	3748	$13 428 275
Hawassa	3 843 810	4384	$22 553 325
Kilimani	6 584 764	6700	$39 249 225
Sokoto	5 774 434	5603	$40 151 475
Total	**27 628 140**	**25398**	**$146 737 375**

12:41

Next to our heading, we add a placeholder for some high-impact statistics. This is why we calculated the water access and no-access measures. If you think about it, the point of this report is to help Pres. Naledi make data-driven decisions. She needs to know how much money we will need, and what problem the money will solve. So the total cost and total improvement numbers are the key metrics we want to display.

12:46

Add a card with the total cost of upgrades, the current percentage of basic access to water, and what percentage of people's lives will be improved.

12:52

Dalila Lesedi
Online

You can use this DAX code to add the plus sign (+) to the Improvement % card if you want:

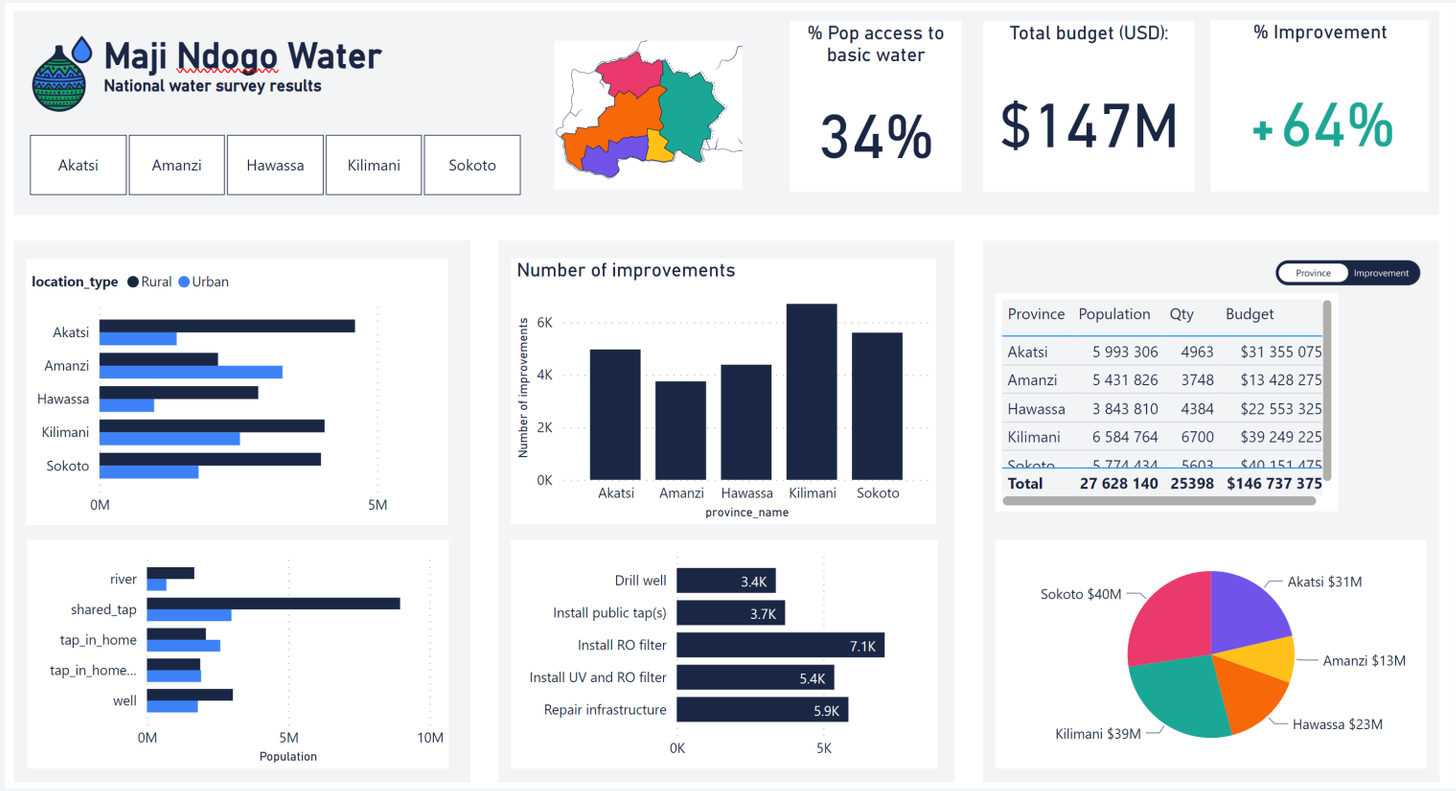
```
Formatted_change =
    VAR
        ChangePct =
        [Access once_complete] * 100
    RETURN

    IF(
        ChangePct > 0,
            "+" & FORMAT(ChangePct, "0") & "%",
        FORMAT(ChangePct, "0") & "%"
    )
```

12:55

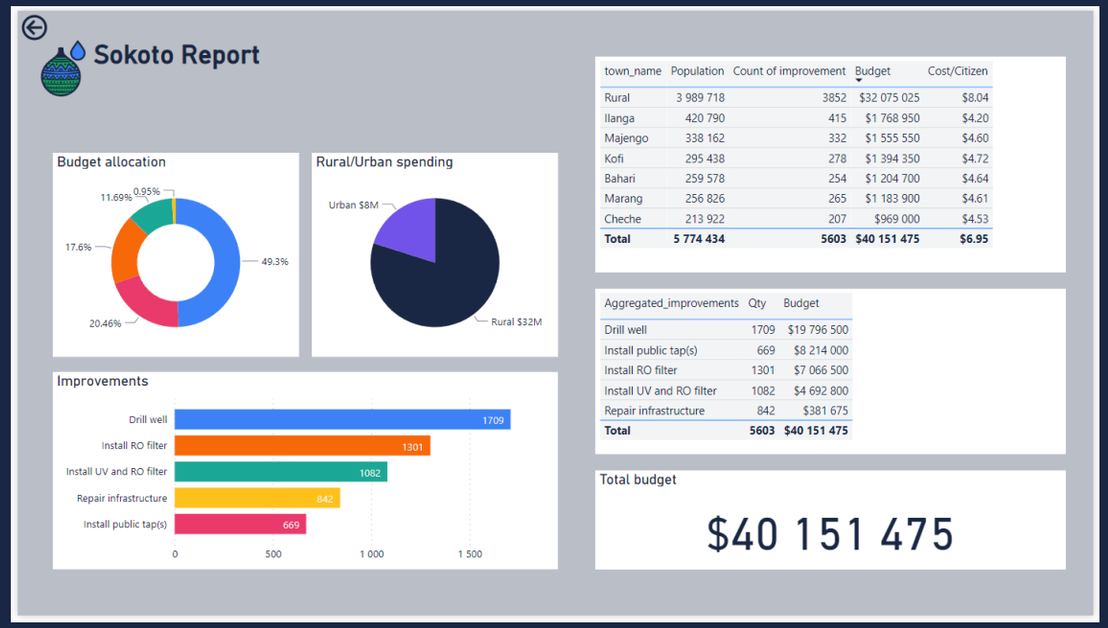
1. **Introduction**
   Setting the stage for our data exploration journey.
2. **User focus**
   Tailoring report design for decision-makers
3. **Data and visuals**
   Identifying needs and crafting visual solutions
4. **Data Insights**
   Transforming analysis into strategic actions
5. **Provincial Perspective**
   Customizing reports for localized decision-making
6. **Enriching data**
   Delving into data processing and analysis
7. **Report Creation**
   Crafting data narratives for impactful reporting

And this is mine! It's ok if yours looks different, in fact, I prefer that!



13:02

Now we have built the National level report, but we still need to make one for each province. Use my example for Sokoto as guidance to make new pages for each province. Once you made it for one, it is simple to duplicate the pages, and tweak them to tell the story of each province. For example, what would the provincial leader of Amanzi need to know? At a minimum, you will need a breakdown of the budget for each town (quantity and cost), the two visuals showing the urban/rural split and the budget allocation visual for each province.



town_name	Population	Count of improvement	Budget	Cost/Citizen
Rural	3 989 718	3852	$32 075 025	$8.04
Ilanga	420 790	415	$1 768 950	$4.20
Majengo	338 162	332	$1 555 550	$4.60
Kofi	295 438	278	$1 394 350	$4.72
Bahari	259 578	254	$1 204 700	$4.64
Marang	256 826	265	$1 183 900	$4.61
Cheche	213 922	207	$969 000	$4.53
Total	**5 774 434**	**5603**	**$40 151 475**	**$6.95**

Aggregated_improvements	Qty	Budget
Drill well	1709	$19 796 500
Install public tap(s)	669	$8 214 000
Install RO filter	1301	$7 066 500
Install UV and RO filter	1082	$4 692 800
Repair infrastructure	842	$381 675
Total	**5603**	**$40 151 475**

13:06

Then make sure to add drill-through options so you can drill through to the provincial pages from the main page.

13:09

Dalila Lesedi
Online

Now this report is far from perfect. It is better to get a report like this into the hands of the user, to see if they find it useful before we spend hours optimising and perfecting it. It is better to iterate and improve the dashboard over time. For now, improve it like you want, but keep the principles in mind!

13:12

So this is it... I'll publish the dashboard once you are done, and let Pres. Naledi know. They will be using this report to get teams on the ground to solve the water crisis using our support. Isn't that exciting? Everything we did up to this point was to help our team make data-driven decisions on how to solve the water crisis, and now it is actually happening! It makes me feel so good at the end of a project to look back at where I was before, and compare it to where I am now.

13:15

Anyway, I need to go pick up my daughter, Ziza, from pre-school, so we'll chat soon! Look after yourself until then!

13:22

EXPLORE AI ACADEMY

Maji Ndogo: Visualizing the currents of change in Maji Ndogo

Transparency in tracking Maji Ndogo's water funds

Aziza Naledi
Online

Dear Dalila and team

Firstly, thank you for the detailed report the team sent me! We could use it to plan out some budgets and even show it to an international funder, who is helping us fund the solution to this crisis.
I wanted to update you on the Maji Ndogo water source improvement project. We've decided to engage local vendors for this initiative, following a thorough vetting process. Each team has been operational for over five years, has clean and updated financial records, and has undergone extensive training. This is to ensure they fully understand the responsibility of their role and the consequences of any dishonesty.

To streamline our monitoring and decision-making process, I am requesting a dashboard to be created. The dashboard should provide clear insights into several key aspects of the project:

- Current progress status of the project.
- Total expenditure to date.
- Breakdown of spending by location.
- Specifics on what the funds have been used for.
- Forecasting whether the allocated budget will suffice for project completion.
- Potential areas for cost reduction.
- Comprehensive data representation at national, provincial, and town levels.
- Provide some insight into what the teams are doing on the ground.

This dashboard will be instrumental in guiding our next steps and ensuring efficient use of resources.

All the best,
Aziza

07:25]

Dalila Lesedi
Online

Hello! Well, here we are... The project just kicked off, and we have our first few water sources completed! 🚀 🍾

07:28

Can you believe it is finally happening? We have one more thing to help out with... We need to build a dashboard for the public so they can see where the money went. We want to show them where the money is being spent, and what the effect is. That dashboard will also be useful for the project decision-makers, to make sure our project is on track, and our resources are put to good use.

07:36

I noticed my PowerBI is getting a bit slow, so if yours is too, these steps will help optimize our `pbix` file a bit:

- Make a copy of the pbix file from last time and put it into a new folder and save this one as `Public_dashboard` using `"Save As..."`. Then delete all of the visual pages. These will slow down Power BI quite a bit, so let's start fresh.
- I removed some of the data we don't need anymore, so if you are seeing empty columns it's ok.
- You are welcome to remove the following columns:
  - `address`, `province`, and `comments` columns from `project_progress` table
  - `date`, `biological`, `description`, and `pollutant_ppm` columns from `well_pollution`
  - `address` column from `location`
- Remove the `queue_composition` table, `water_source_related_crime` tables.

07:42

Next up, we have to update the dataset. I sent you an Excel file with some of the new data. If you replace the Excel file you have, it will load the new data in our model. But, remember, this might break our report from last time, so to avoid that, you can store the Excel file somewhere else, and then edit the `Data source settings` in the power query editor to link to the new location.

07:46

Here is the new data:

Md_water_services_data.xlsx

07:50

Ok, make sure your model loads correctly, and that there are no errors, and look at `project_progress`. You will notice some new columns. `province` is gone, and now `town` has town names, but now Amina and Harare have the first letter of the province at the end so we know which Harare we're talking about and rural areas are also now appended with the name of the province! This is updated in the `location` too.

07:58

Dalila Lesedi
Online

This update will break the `budgeted_improvement_cost` column in the `project_progress` table we calculated last time. Here is the DAX to fix it. Since rural areas now have a province name attached, we need to search for a string rather than having to match it exactly:

```
budgeted_improvement_cost =
    IF(
        CONTAINSSTRING(
            'project_progress'[town],
            "Rural*"
        ),
        RELATED('infrastructure_cost'[Rural_adjusted_cost]), RELATED('infrastructure_cost'[unit_cost_USD])
)
```

08:03

Your calculated `[Rural_adjusted_cost]` column in `infrastructure` may have a different name than mine.

08:07

There are also new columns:

- `date_started`, recording the date an improvement started
- `date_of_completion`, showing when the project was completed
- `cost`, the actual cost of the project in USD($). This cost is the amount we paid to make the upgrade. It includes travel costs, materials and labour. In an ideal world, this would match the `Budgeted_improvement_cost column` we created to budget for the project.
- `assigned vendor`, an ID of the vendor that made the upgrade.

08:13

Make sure that all of the data types are correct, and the dates are set to dates in this format (YYYY/MM/DD). If the date filters we're going to make don't work, check this step again.

08:19

Finally, check the relationships and directionalities. What should the `water_source-to-project_progress` relationship be? These may have changed when we import data, so make sure these are correct, and what we expect.

08:25

Next up, `vendor_ID`'s are a little impersonal, so if you import the Excel file again, you will see a table called `vendors`. We won't use it, but here you can see who the heroes are on the ground, improving the lives of our people!

08:29

Choose `vendors` only, and import it into the data model. Make sure it imports correctly and you will see information like the company name, who owns it, and what they do:

- Groundwater Extraction teams drill wells.
- Water Distribution System Installation teams install taps.
- Water Purification System Installation teams install the UV and RO well filters.
- Civil Infrastructure Assessment teams are working on the broken infrastructure.

08:36

Dalila Lesedi
Online

Let's check the data in `project_progress`. For each project, we now know if it is `Complete`, or still in the `Backlog`, and once it is complete, we can see which team worked on it when the team started working on the project, when they completed it, and how much we paid them for the improvements. We have all of the other data we had before, so we are going to make some amazing things!

08:44

Let's think through the stories a bit before we make the dashboard. We have two requests from Aziza. The dashboard has to make the work we're doing visible to the public and help decision-makers monitor the progress of the project.

08:49

Dalila Lesedi
Online

So, as a member of the **public**, I want to see:

1. How the project is going. Very simply: Which sources have been completed?
2. How much money has been spent so far?
3. Where was the money spent?
4. What the money was spent on?
5. Details about everything in my town.

08:55

As a **decision-maker**, I want to see:

1. How far is the project?
2. How much money has been spent so far?
3. Where was the money spent?
4. What the money was spent on?
5. Will we have enough money to complete the project?
6. Where can we cut costs?
7. I want to see data at the national, provincial and town level.

08:58

Dalila Lesedi
Online

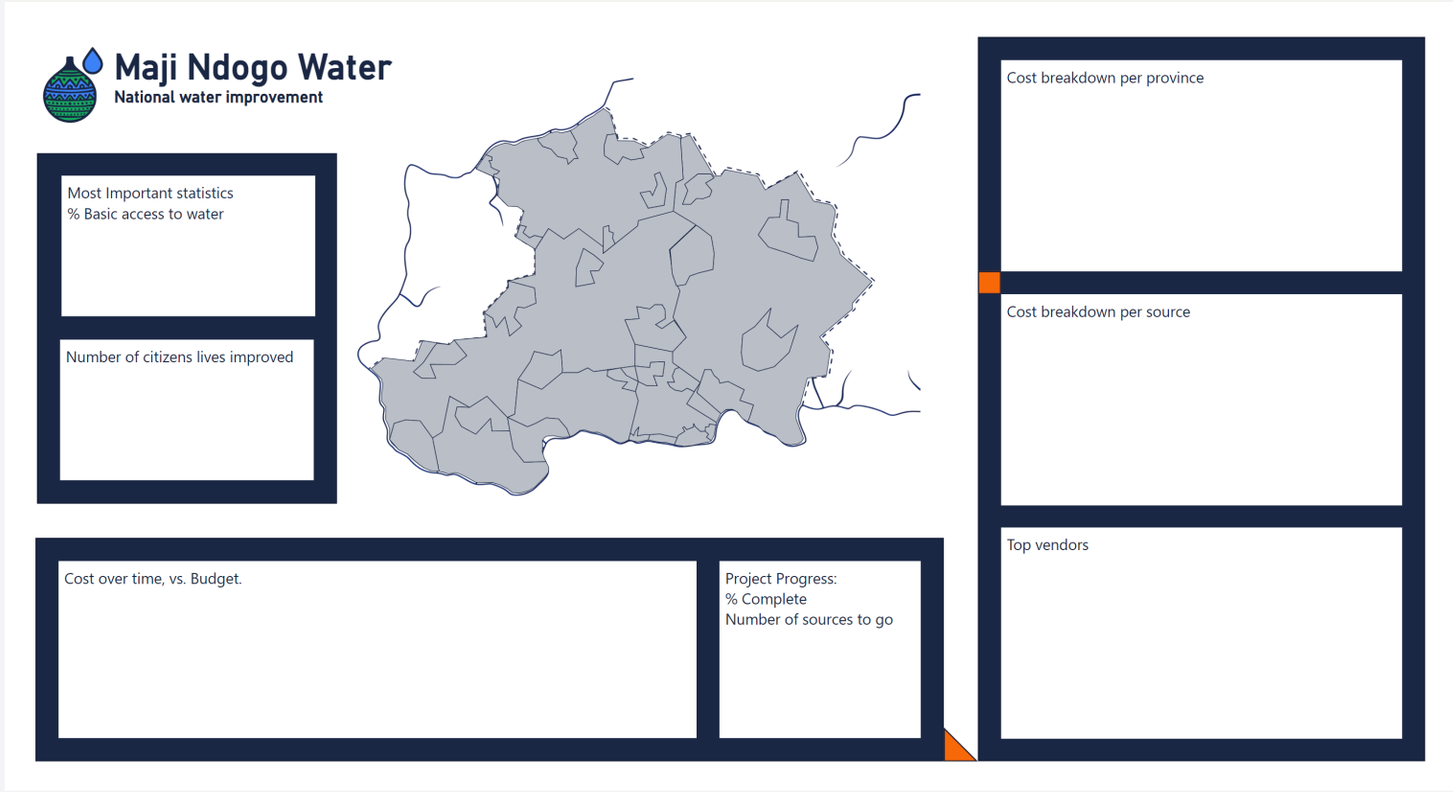
Their stories are quite similar, so we can build one dashboard for both purposes and create detailed drill-throughs for more specific content when we need it.

09:03

So let's create a basic layout of our dashboard, then dive into what we need to calculate.

09:06

This is how I laid out mine:



09:10

Dalila Lesedi
Online

I hope you did something else! But did you notice my little orange rulers? I use them to make sure the spacing of my visuals is consistent across the page, so there is a little "nugget" for you, as I used to tell Chidi. I am also still unsure about the "Top vendors" section. Maybe it will be better to put page controls there.

09:16

So the main visuals I want to build, are:

- A map that will show us how far along the project is. Users should be able to see visually how far the project is. Users should also be able to select any town, or rural area, and know immediately how the project is doing.
- A cost tracking visual. A KPI plot is a nice option that will show us how much we thought it would cost, vs. how much it is really costing us.
- Break down the cost of the project so people can see where the money was spent.
- Summaries of key metrics:
  - Project progress (%)
  - Basic access to water (%), updated now that the new water sources are installed.
  - How many people are affected by the changes.? Remember, we're doing this for the citizens of Maji Ndogo!
  - How far to go before the project is complete?

We may add more, but this is my plan for now. Each of these visuals needs to be interactive at the national, provincial, and local levels but also has to interact with the data over time.

09:19

To make the new and improved map, you are going to need this shape map JSON file as our Map:

MD_Full_map.json

09:23

The `town` column in `project_progress` is the column we use in the location well. If you drag `source_status` into the `Color saturation` well, you should see it light up! This now just shows us how many water sources there are in each location, not their actual status, or how many are complete. So, we need to create a measure to make this visual effective. We need the same data for the `Project progress (%)` metric too, so let's make a measure for it.

09:31

The problem is that the measures we create this time need to interact with the location and date filters properly. The calculations are simple, but the filters are going to complicate things.

09:38

To calculate the `Project progress (%)`, we need to calculate the number of complete projects divided by all of the projects. At a national level, it includes all of the projects in Maji Ndogo, but for Djenne, it will be the number of projects completed in Djenne, divided by the number of projects in **Djenne** in total. So we need the measure to always take into account the `town` filter.

09:44

The total number of improvements is:

```
total_improvements =
    CALCULATE(
        COUNTROWS('project_progress'),
        ALLEXCEPT('project_progress','project_progress'[town]))
```

09:50

Let's unpack it together.

09:53

First up, we have CALCULATE. It lets us modify how we calculate a measure based on specific conditions, which in this case are set by two conditions `ALLEXCEPT(project_progress, project_progress[town])` and `ALL(project_progress[date_of_completion])`.

09:59

Inside CALCULATE, there's COUNTROWS(project_progress). This part is straightforward – it's just counting how many rows are in our project_progress table.

10:03

Now, the interesting part: `ALLEXCEPT(project_progress, project_progress[town])`. This bit is telling Power BI to ignore all filters except for the ones in the town column. So, no matter what other filters are applied elsewhere, the only one that will apply is the town column. Adding a date slicer as a filter will create problems. When we change the date, the rows are filtered by the date, which means we don't get the total number of projects.

10:08

What we get is the total number of projects in the project_progress table, but if we select a town, it calculates all of the improvements in that town.

10:14

Next up, the `number_completed_projects`. What is the most reliable way to know whether a project is complete? Well, we can count the number of rows that satisfy the condition for a project being complete. You should get a number of 29.

10:22

Now that we have that, go ahead and calculate the percentage (`pct_project_complete`). Remember to format it so it will show up as a % with no decimals.

10:31

Your value should be 0% at the start on a national level, but Serowe, for example, is 1% done. Check that your values are the same as mine, and add them to the dashboard as cards.

10:43

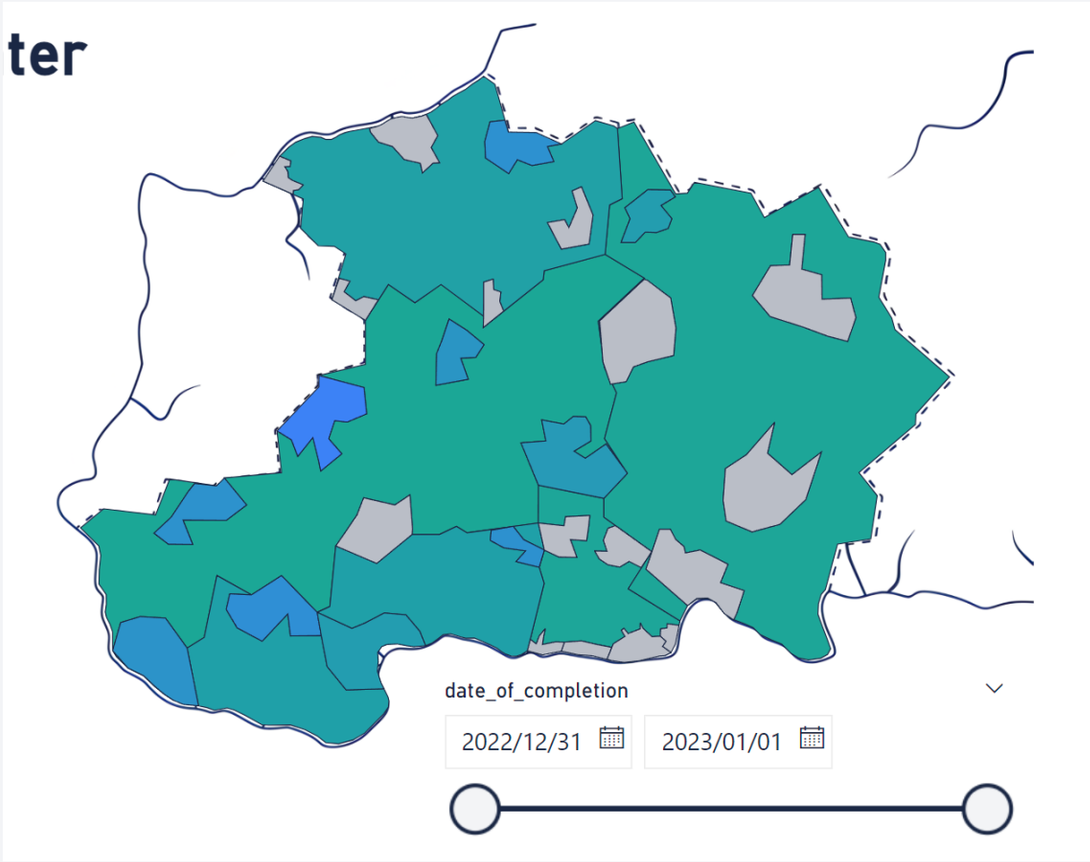
I added another metric there; "More sources to go". You can calculate that one on your own.

10:47

Now we also have a good metric to show progress on the map, so use `pct_project_complete` as the colour saturation variable. Since we used percentages, the visual will scale better, so the colours are now more informative than before.

10:51

This is what my map looks like:



10:54

Dalila Lesedi
Online

Towns like Zuri, Mrembo, and Isiqalo are grey in mine. What does that represent? Then there are coloured towns in mine. What do these colours represent? Hover your mouse over them. Also add some information like `number_completed_projects` in the tooltips to make it more informative.

11:02

Dalila Lesedi
Online

Next up, we calculate the Basic access to water numbers. We did something like this last week, but now we need to do it in a way that includes the new water sources. Rivers were replaced by wells, wells are clean once they are complete, so these citizens now have basic access to water from an improved water source, so: `Basic access to water = (population_with_basic_access + population_improved)/total_population`

11:09

The `total_population` is:

```
total_population =
    CALCULATE(
        <BLANK>('water_source'[<BLANK>]), -- Sum of people
        ALLEXCEPT(
            'project_progress',
            'project_progress'[<BLANK>]
        ) --Only keep [town] as a filter.
    )
```

11:17

Fill in the blanks to calculate the `total_population`.

11:21

You should get 28M people for the population of Maji Ndogo, and 3M people for the population of Kilimani. Make sure it updates correctly! Do you see that the total population calculation works just like the `total_improvements` one did, but we just aggregate using a different method?

11:31

Now we calculate `population_with_basic_access`. Last time we used a column to classify sources as Basic or not, then added the rows. Since our dashboard is slowing down, let's use some DAX magic to put it into a measure all in one go. This way, Power BI doesn't have to store all of those strings in the table, so it calculates faster and takes up less space. Since we did it last time, I'll give you all of the code you need to do it.

11:35

The `population_with_basic_access` is:

```
population_with_basic_access =
    CALCULATE( -- Calculate with some conditions
        SUM('water_source'[number_of_people_served]),-- Population when....
        FILTER(
            ALL(water_source),
            OR(-- Nested or to have well, OR tap_in_home OR shared tap
                OR(
                    AND( -- When it is a well, it must be clean too
                        'water_source'[type_of_water_source] = "well",
                        RELATED(well_pollution[results]) = "Clean"
                    ),
                    'water_source'[type_of_water_source] = "tap_in_home"
                ),
                AND( -- When it is a shared tap, it must have a short queue time
                    'water_source'[type_of_water_source] = "shared_tap",
                    'water_source'[Average_queue_time] < 30
                )
            )
        )
    )
```

11:39

Much more complex than last time, right? But this is a bit more efficient. Power BI does tend to slow down as a project progresses, so we need to clean up now and again, and this measure avoids creating a column full of data. Here we basically calculate the sum of people who use sources that fulfill the conditions of basic water.

11:45

Next up, we need to calculate the number of people who now have basic access, once their source is improved. How would we calculate it? The worst thing is how nested the conditions become. 😢

11:49

Calculate the number of people who will have access to basic water once the water source is improved as `population_now_basic_access`. Make sure that the column you use to filter with will accurately show if a source is improved or not. You should get 10760 people who now have access to basic water (excluding everyone who already had basic access).

11:53

Calculate the number of people that will have access to basic water once all the improvements are complete (including those that had access already), `pct_population_now_basic_access`, and remember to format it to display a percentage. It should be 34% at the start, and if you select the town of Serowe in Hawassa, you should see 57% access. `Basic access to water = (population_with_basic_access + population_improved)/total_population`

12:00

Dalila Lesedi
Online

Great! Now, let's jump into costs! The main thing we need to know about our finances is how much we have spent so far, and whether we have spent more than we wanted to. Our budget from last time is our target, and ideally, we want to spend less than the budget. The KPI visual is an excellent tool to visualise this since it will track our progress over time.

12:11

For this visual to work well, we need a date slicer so we can select data between some dates. Add a date slicer to the visual.

12:18

For this visual, we need to calculate two measures; the running cost (`cumulative_cost`) and running budget (`cumulative_budget`). These numbers should sum up all of the budgeted costs and real costs, up to a date, respectively. For example, if we paid vendors $100 000 on day one and $150 000 on day two, the running cost is $250 000 on day two.

12:24

Our calculation for `cumulative_budget` should sum up all of the budget costs for projects with completion dates <= the current date. Blanks are going to be a pain when we compare dates, so we need to remove them in the DAX. Try to create this measure on your own first. You should get a number of $128 450 for the data we currently have.

12:27

Dalila Lesedi
Online

This is how I calculated `cumulative_budget`. Some of the fields missing <BLANK>:

```
cumulative_budget =
    CALCULATE(
        SUM('<BLANK>'[<BLANK>]),
        FILTER(
            ALL('project_progress'[<BLANK>]),
            'project_progress'[<BLANK>] <= MAX('project_progress'[date_of_completion]) &&
            <BLANK>(ISBLANK('project_progress'[<BLANK>])) --This line removes blank\null values.
        )
    )
```

12:31

To calculate `cumulative_cost`, we can simply copy-paste the budget and change the columns to use the correct data. You should get an amount of $131 914.91.

12:39

To set it up, create a KPI visual, and drag the `cumulative_cost` into the `Value` well, the `cumulative_budget` into the `Target` well, and `'project_progress'[date_of_completion]` into the `Trend axis` well. Once these are set up, you should see a visual like this:

12:51

This format is not quite correct, so we have to change some of the settings. First, you will notice that the Goal is $98.70K, and we spent $131.91K. So we have spent more than we budgeted for, but the KPI is interpreting this as a good thing with a correct tick mark. To change this, change the `Trend axis` to `Low is good`. Next change the `Target label` to `Budget`

12:54

Now this visual communicates a bunch of things:

- Trend of the overall cost over time;
- How much money we have spent;
- How much we wanted to spend; and
- With colours and icons, clearly show whether we have over-spent (Budget deficit) or if we spent less (Budget surplus).

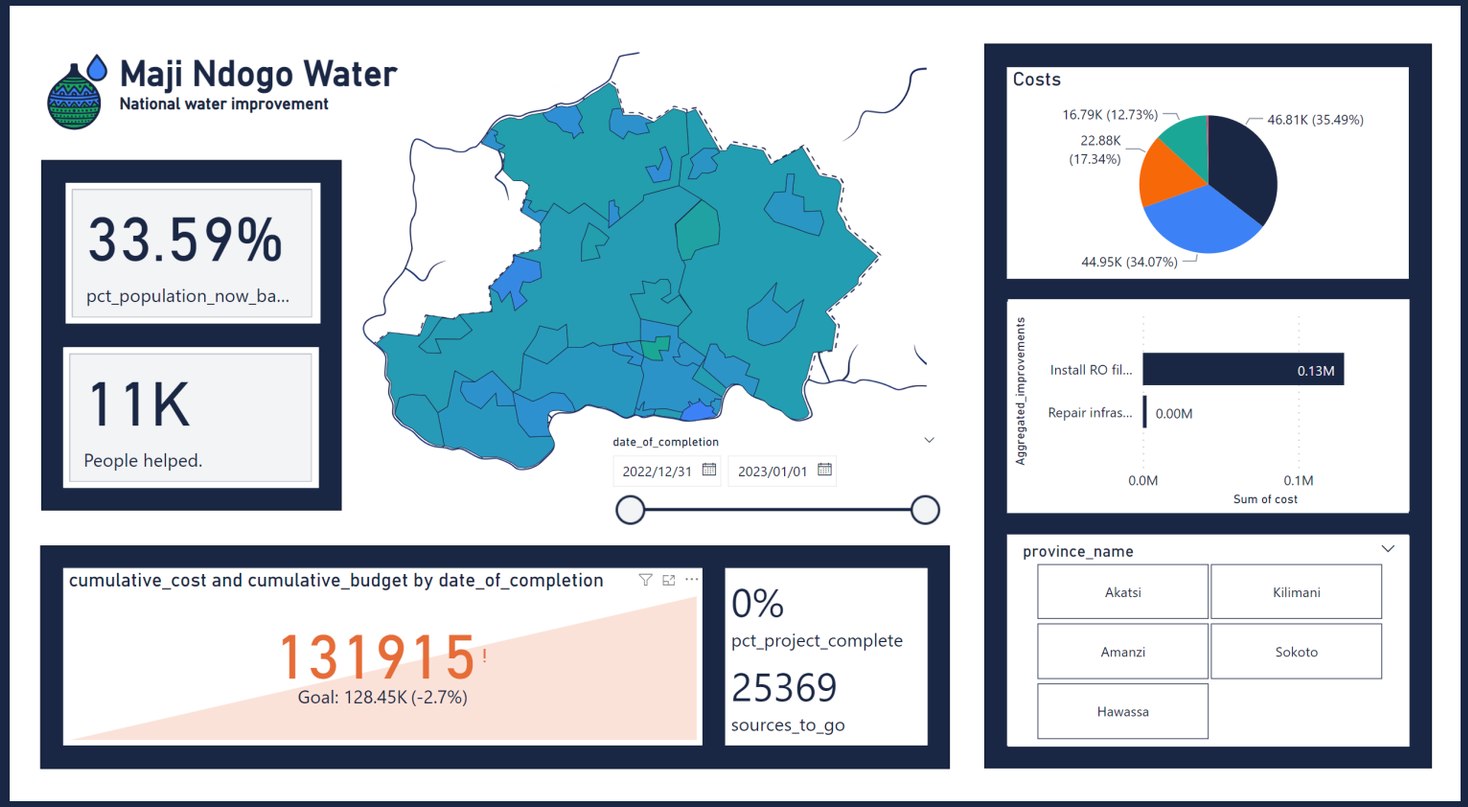
13:00

Dalila Lesedi
Online

Add metrics like People helped (`population_now_basic_access`), number of sources to still improve (which is `total_improvements - number_of_completed_projects`), and use the `Aggregated_improvements` column from last time to break down costs per improvement type.

13:03

Update your dashboard to have all of the visuals and slicers I made. Make sure to format all values, titles and labels that are displayed, unlike mine:



13:06

Dalila Lesedi
Online

Let's recap if we achieved the user stories we started with:

1. How far is the project? ✅
2. How much money has been spent so far? ✅
3. Where was the money spent? ✅
4. What the money was spent on? ✅
5. Will we have enough money to complete the project? ✅
6. Where can we cut costs? ❌
7. I want to see data at the national, provincial and town level. ✅

It's ok that we could not finish 6. We don't know where we can cut costs before the project starts, so let's wait for some data to come in, and we can try to help decision-makers with this once we know more.

13:11

Then, take a few minutes to clean up all of the titles, text, and tooltips, make sure your spacing is correct, remove your rulers, and get ready to launch this dashboard to the public! 🚀🚀🚀

13:18

4 January 2024

Hello! It has been a while! After a year of upgrades, we have an update on our data, and now you will see the power of a dashboard. So let's get it going!!

13:22

Here is the new data up to 2024-01-03:

Md_water_services_data.xlsx

13:25

Here is a visual I made to show how the sources are being upgraded across Maji Ndogo:

Maji Ndogo flowing again

13:33

Overwrite the last Excel file with this one, or link it as the new source.

13:39

Dalila Lesedi
Online

Once you're all done, head back to your visual, and be prepared! Your date slider expanded, and now we can select data up to 2024! With no extra modelling and no copy-pasting results into spreadsheets, we have an update of our story! If we accessed a SQL database, it would even happen automatically!

13:47

Everything is updated, so let's see what happened last year. We started at 34% access at the start, up to 48% now! That is massive, right? The project is 22% complete, so we still have a long while to go. If we got this far in one year, it will probably take 5 years in total! Ok.

13:52

The bad news is money! If you look at the KPI plot, we can see we're about 10% over budget at the moment. Let's look around, and try to understand why.

13:56

If we look at the different provinces, we can see that Sokoto has taken a large part of the budget, but we knew it would be expensive since we had to drill many wells.

14:01

If you select Sokoto, you will see it is almost 40% over budget already. Sokoto is a rough place with few roads, so it makes sense that it is so expensive for teams to work there. We didn't take that into account.

14:07

Dalila Lesedi
Online

Open up a new visual page and create a key indicators visual. Think about what we want to analyse and what could explain the high costs. Examples could include the town, province, location type, or even the time it takes to complete a project (calculate the time difference with a column).

14:14

What I could see from the key influencers:

- When we have to drill wells, it is expensive. We took this into account in our budget, so this is not the reason for the high costs.
- When sources are in rural areas, costs are higher. We took this into account in our budget, so this is not the reason for the high costs.
- When jobs take long, costs go up. This has two components if you think about it. One is, that certain jobs take longer. Installing 8 taps at a location takes much longer than installing a filter. But the second component of why a job can take so long is the time it takes to travel to a new location. These travel days can add up a lot over the course of a project, and add costs.

14:18

Create a new visuals page and make several charts examining the effect on the average cost of an improvement and try to show the following results:

- It is almost twice as expensive to improve a source in a rural area, compared to an urban area.
- Sokoto has a very high average cost of improvement, both rurally and in urban areas.
- We are over budget in every province.
- We underestimated the cost of rural improvements in Sokoto.

14:22

Dalila Lesedi
Online

While this data is useful to know, we can't do anything about it. Rural areas need larger vehicles that use more fuel. Sources are far apart, and there are many of them. But we can't change that. So let's focus on finding a change we **can** make to make sure we stay on budget.

14:30

Let's look at the vendors. If we see vendors that are asking too much, we stop using them. Simple!

14:37

Create a new page showing the average cost vendors' are asking for improvements, and sort the table from most expensive to the cheapest teams.

14:44

So some teams are way more expensive than others, right? But we need to compare apples to apples. Teams that drill wells **will** naturally charge more than teams installing filters, and we also saw that it also costs more to improve sources in Sokoto and Kilimani, so that's another factor.

14:49

Create a slicer for a date, improvement type (aggregated_improvements), and rural/urban split and add our map to this page.

14:52

So now we can filter by "Drill well", and compare the prices of the vendors.

14:57

Dalila Lesedi
Online

If you select the most expensive vendor, `MBS605`, what do you notice? Where are they upgrading sources?

15:02

`MBS605` is drilling wells across the rural parts of Sokoto only! So they are working in the harshest conditions. Can we blame them if it costs more? The four most expensive vendors are working either in Sokoto, or Kilimani.

15:09

The cheapest four drilling teams are in urban areas and outside of Sokoto, where it is easier to access. So which team is actually expensive, and who isn't?

15:15

Isn't that beautiful? As people, we often make choices simple for ourselves, but here, the data tells us if we fired the expensive teams, it would have been a mistake.

15:19

Don't believe me yet? Select all of the teams installing filters, RO and RO + UV, and we're going to look at two vendors, `ERI893`, **the best water purification installers in the whole of Maji Ndogo**, Entebbe RO Installers, and `OW290` a "not so good" company called Ouagadougou Waterworks.

15:25

Did you notice `ERI893` is the most expensive filter installer in Maji Ndogo? They are asking almost 50% more than the cheapest team! "How can Dalila then say they are the best team in Maji Ndogo?" 🤑

15:30

Dalila Lesedi
Online

Haha, where did they work?

15:35

Why do I say they're the best team? Well, first, add the count of completed projects to the table. Who finished the most projects? I won't laugh again. 🤭

15:40

You are still doubting me, right? "Dalila, they are expensive!!"

15:45

Let's compare Entebbe RO Installers to companies who work in a similar environment. Try this before you go on.

15:48

Filter the table to only include the data for rural areas and the Sokoto province.

15:53

Hmm...🤓

15:56

That's right...

16:03

Dalila Lesedi
Online

Entebbe RO Installers is the best team in Maji Ndogo! Look how cheap they are when we compare them to the other teams in the right context.

16:09

Now I am being playful, but there is a very hard truth in this exercise. Complex real-world data is difficult to work with! Sometimes we may think we have an answer, but often we are wrong. So we keep exploring the data, and finding the truth! In the next 5 years of your career, this is the skill that you will develop and one day, you will sniff around in the data like I did to find this nugget.

16:17

But **why** is Entebbe RO Installers so cheap?

16:22

Look at how Entebbe RO Installers (**Blue**) moved throughout Maji Ndogo, completing jobs, compared to Ouagadougou Waterworks (**Orange**):

ERI893 and OW290 journey

16:30

Entebbe RO Installers stayed in one place, picking up new projects near them, keeping their travel costs low, while Ouagadougou Waterworks is moving from town to town, between rural areas travelling a lot between the projects. Perhaps they are not considering that less travel, means they can start new projects quicker.

16:37

Dalila Lesedi
Online

Entebbe RO Installers has made the most out of all the purification teams, not because they charged a lot, but because their choices mean they get more projects done! Cool right?

16:41

So what can we do? Well, I'll get our designers and animators to make a video, and create some digital infographics we can send to the teams, explaining how they can improve, and how to search for jobs closer to them on the system.

16:49

If one team follows our advice, we make a small change, but hopefully all of them take note, and pick jobs that are more optimal. If this worked, we might even get our budget under control.

16:54

Last thing. Isn't it amazing that all of this insight is hidden in the data? We could spend another year just learning from it. By being curious and determined, we can uncover the truth hidden within.

16:59

We'll check in again once we have more data. Until then, take care!

17:06

Dalila Lesedi
Online

3 December 2027

Hello! Can you believe it has been almost 4 years? It feels like yesterday!

17:10

Here is the final data:

Md_water_services_data.xlsx

17:13

Introduction
Setting the stage for our data exploration journey.

1 **Updating our data**
The start of the final stretch

2 **Prioritizing users**
Crafting user-focused visuals

3 **Visualising the change**
Creating the canvas of hope

4 **Improving lives**
Upgrading Maji Ndogo one source at a time

5 **The cost of change**
Where will the money go

6 **Water flows**
One year update

7 **The final update**
Journey's end

Dalila Lesedi
Online

Did it work? 🙃

17:19