Transparency in tracfiing 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. Isent 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

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

Dalila LeSedi
Online

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-mafier**, 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

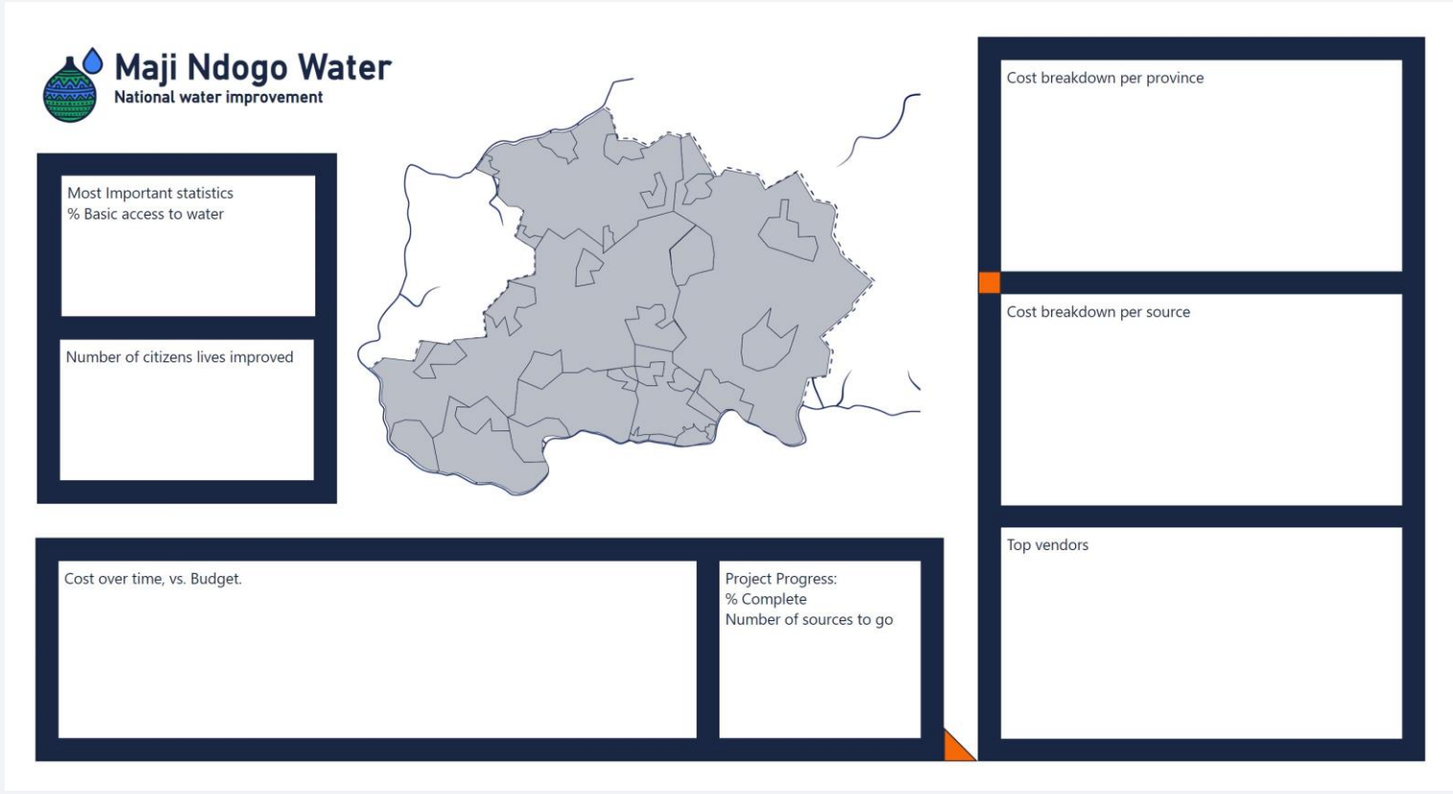
Their stories are quite similar, so we can build one dashboard for both purposes and create detailed drill-throughs for more specific content when we need it.

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

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

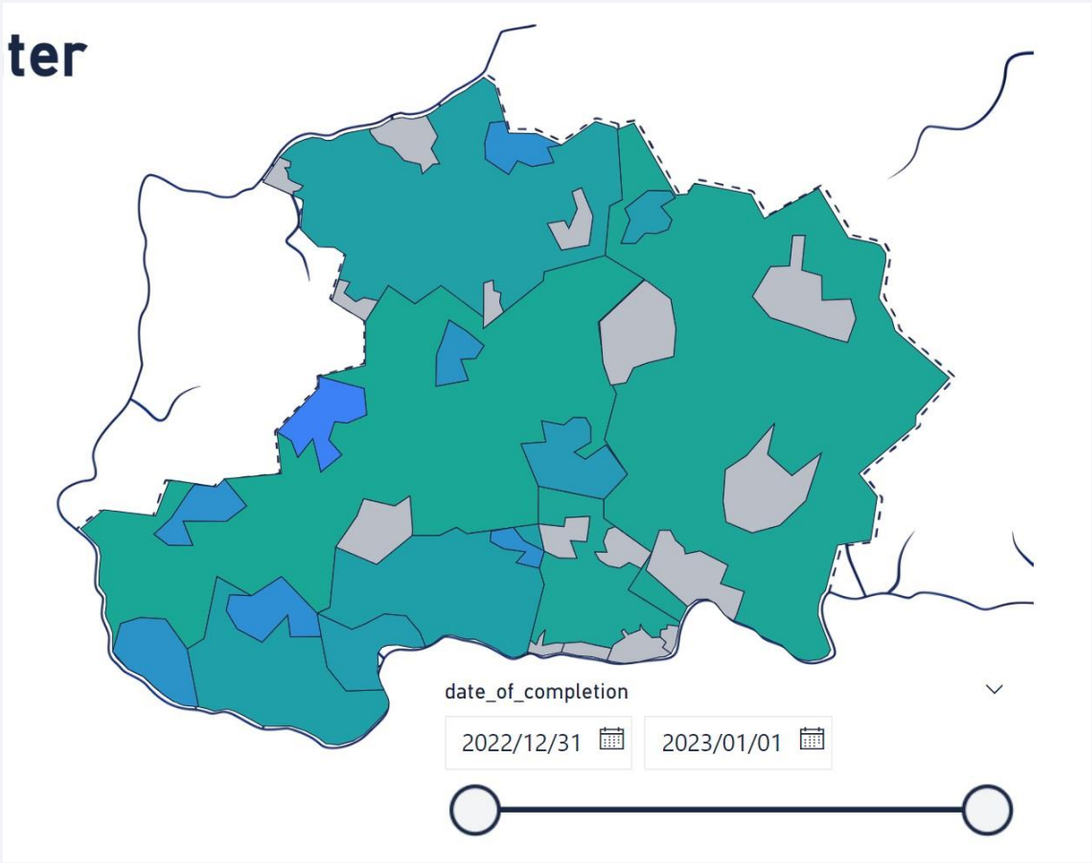
I added another metric there; "More sources to go". You can calculate that one on your own.

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

Dalila LeSedi
Online

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

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:

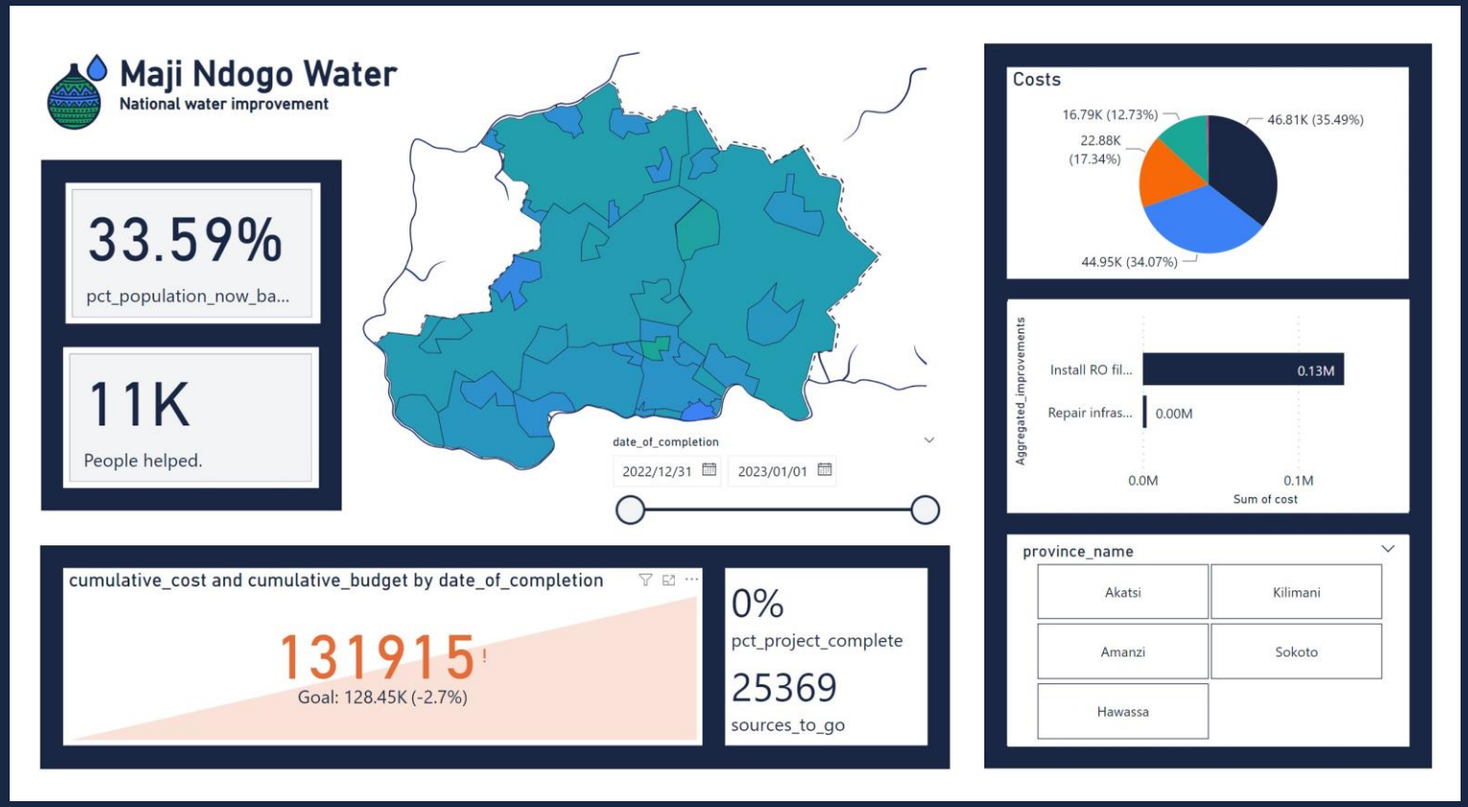
- Trend of the overall cost over time;
- How much money we have spent;
- How much we wanted to spend; and
- With colours and icons, clearly show whether we have over-spent (Budget deficit) or if we spent less (Budget surplus).

13:00

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

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 Isay they're the best team? Well, first, add the count of completed projects to the table. Who finished the most projects? Iwon'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

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 Iam 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

Dalila LeSedi

Online

Did it work? 🙃

17:19

1. **Introduction**
   Setting the Stage for our data eXploration journey.
2. **Updating our data**
   The Start of the final Stretch
3. **Prioritizing uSerS**
   Crafting uSer-focuSed ViSualS
4. **ViSualiSing the change**
   Creating the canVaS of hope
5. **Improving liveS**
   Upgrading Maji Ndogo one Source at a time
6. **The coSt of change**
   Where will the money go
7. **Water flowS**
   One year update
8. **The final update**
   Journey'S end