

Maji Ndogo: Visualizing the currents of change in Maji Ndogo

# Moulding data into visual stories in Maji Ndogo















Setting the stage for our data exploration journey.



## **Buiding a data model**

Unleashing PowerBI with data models



#### **National scale**

Bird's eye of our national survey results



#### **Visualising queues**

Recreating queue related visuals



#### Cleaning up wells

Recreating a pollution map



## Connecting crime

Linking crime data to find new insight into water data 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

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.



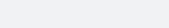




















Setting the stage for our data exploration journey.



# **Buiding a data model**

Unleashing PowerBI with data models



#### **National scale**

Bird's eye of our national survey results



## **Visualising queues**

Recreating queue related visuals



## **Cleaning up wells**

Recreating a pollution map

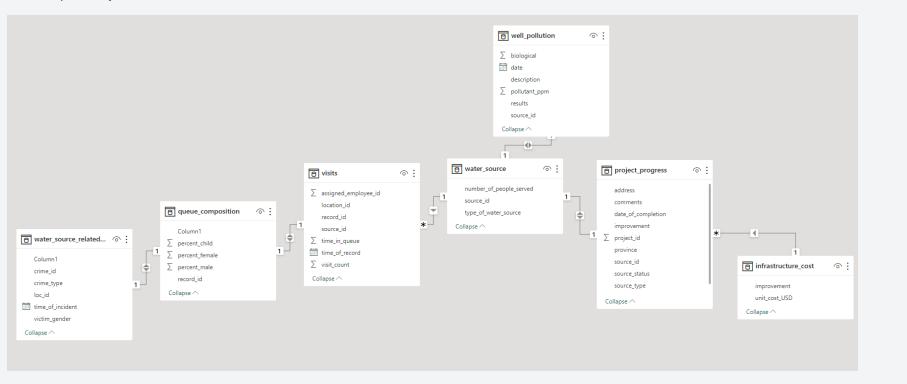


# **Connecting crime**

Linking crime data to find new insight into water data



**Dalila Lesedi** 







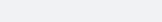




















Setting the stage for our data exploration journey.



## **Buiding a data model**

Unleashing PowerBI with data models



#### **National scale**

Bird's eye of our national survey results



#### **Visualising queues**

Recreating queue related visuals



#### Cleaning up wells

Recreating a pollution map



# **Connecting crime**

Linking crime data to find new insight into water data



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.































Setting the stage for our data exploration journey.



#### **Buiding a data model**

Unleashing PowerBI with data models



#### **National scale**

Bird's eye of our national survey results



#### **Visualising queues**

Recreating queue related visuals



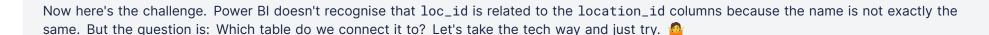
#### Cleaning up wells

Recreating a pollution map



## **Connecting crime**

Linking crime data to find new insight into water data



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!

























Setting the stage for our data exploration journey.



## **Buiding a data model**

Unleashing PowerBI with data models



#### **National scale**

Bird's eye of our national survey results



#### Visualising queues

Recreating queue related visuals



#### Cleaning up wells

Recreating a pollution map



# **Connecting crime**

Linking crime data to find new insight into water data

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.





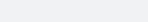




















Setting the stage for our data exploration journey.



## **Buiding a data model**

Unleashing PowerBI with data models



#### **National scale**

Bird's eye of our national survey results



#### **Visualising queues**

Recreating queue related visuals



#### Cleaning up wells

Recreating a pollution map



#### **Connecting crime**

Linking crime data to find new insight into water data



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.



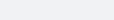












**Dalila Lesedi** 









## Introduction

Setting the stage for our data exploration journey.



## **Buiding a data model**

Unleashing PowerBI with data models



#### **National scale**

Bird's eye of our national survey results



# **Visualising queues**

Recreating queue related visuals



# Cleaning up wells

Recreating a pollution map



# **Connecting crime**

Linking crime data to find new insight into water data

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.



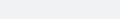




















Setting the stage for our data exploration journey.



## **Buiding a data model**

Unleashing PowerBI with data models



#### **National scale**

Bird's eye of our national survey results



#### Visualising queues

Recreating queue related visuals



#### Cleaning up wells

Recreating a pollution map



## Connecting crime

Linking crime data to find new insight into water data



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.



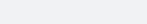




















Setting the stage for our data exploration journey.



## **Buiding a data model**

Unleashing PowerBI with data models



#### **National scale**

Bird's eye of our national survey results



#### **Visualising queues**

Recreating queue related visuals



## Cleaning up wells

Recreating a pollution map

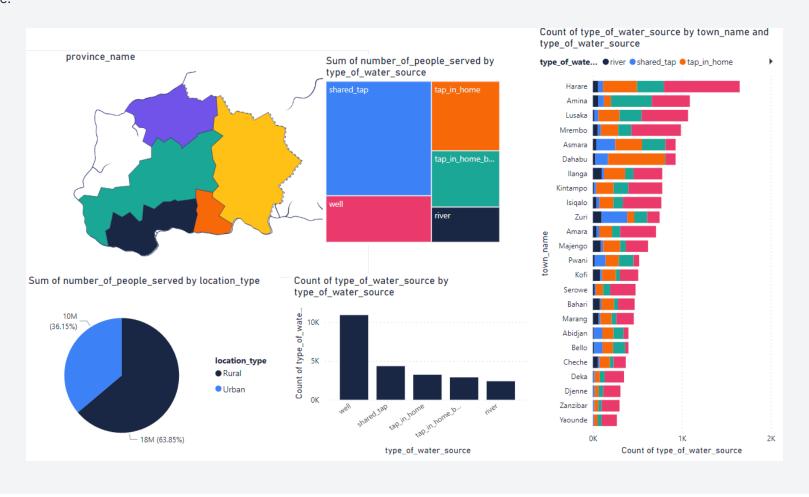


#### **Connecting crime**

Linking crime data to find new insight into water data

# **Dalila Lesedi**

#### This is mine:



10:17





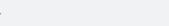






10













Setting the stage for our data exploration journey.



## **Buiding a data model**

Unleashing PowerBI with data models



#### **National scale**

Bird's eye of our national survey results



#### **Visualising queues**

Recreating queue related visuals



## Cleaning up wells

Recreating a pollution map



# **Connecting crime**

Linking crime data to find new insight into water data



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







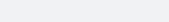






**Dalila Lesedi** 















Setting the stage for our data exploration journey.



# **Buiding a data model**

Unleashing PowerBI with data models



#### **National scale**

Bird's eye of our national survey results



## **Visualising queues**

Recreating queue related visuals



## **Cleaning up wells**

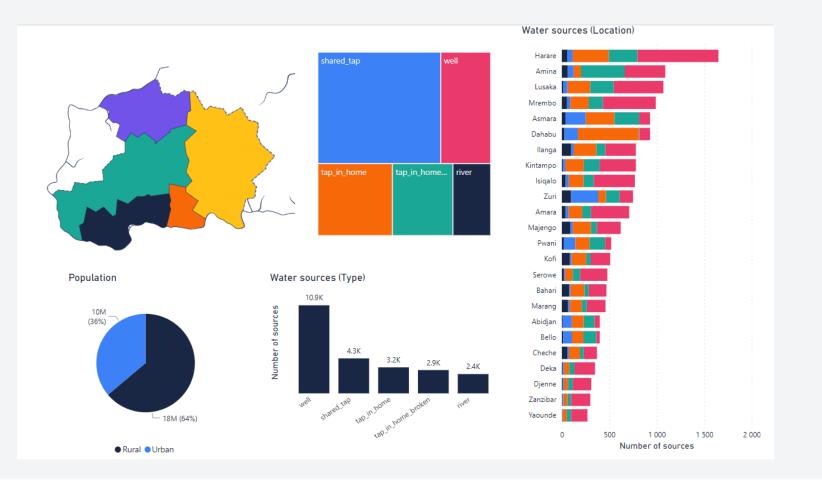
Recreating a pollution map



# **Connecting crime**

Linking crime data to find new insight into water data

# And once we clean it up a bit we get:





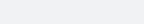




















Setting the stage for our data exploration journey.



## **Buiding a data model**

Unleashing PowerBI with data models



#### **National scale**

Bird's eye of our national survey results



#### Visualising queues

Recreating queue related visuals



## Cleaning up wells

Recreating a pollution map



# **Connecting crime**

Linking crime data to find new insight into water data



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.





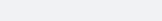




















Setting the stage for our data exploration journey.



# **Buiding a data model**

Unleashing PowerBI with data models



#### **National scale**

Bird's eye of our national survey results



#### **Visualising queues**

Recreating queue related visuals



#### Cleaning up wells

Recreating a pollution map



# Connecting crime

Linking crime data to find new insight into water data



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.

























Setting the stage for our data exploration journey.



# **Buiding a data model**

Unleashing PowerBI with data models

#### **National scale**

Bird's eye of our national survey results



## **Visualising queues**

Recreating queue related visuals



#### **Cleaning up wells**

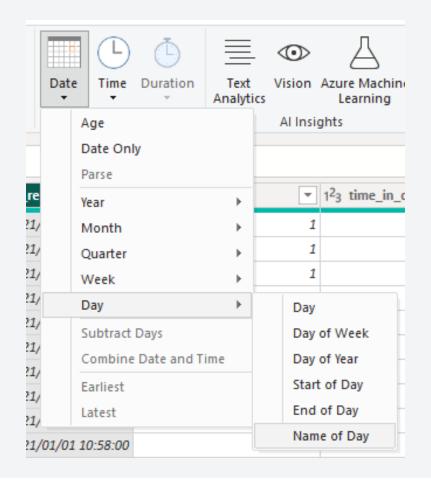
Recreating a pollution map



# **Connecting crime**

Linking crime data to find new insight into water data

This is the menu you should look out for:









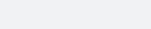






15











11:13





#### Introduction

Setting the stage for our data exploration journey.



## **Buiding a data model**

Unleashing PowerBI with data models

2

#### **National scale**

Bird's eye of our national survey results



## **Visualising queues**

Recreating queue related visuals



## **Cleaning up wells**

Recreating a pollution map



#### **Connecting crime**

Linking crime data to find new insight into water data Select the time\_of\_record column again, and create a new column that extracts the hour.

You should now have two columns:

-	A <sup>B</sup> C day_name ▼	1 <sup>2</sup> 3 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.

























Setting the stage for our data exploration journey.



# **Buiding a data model**

Unleashing PowerBI with data models



#### **National scale**

Bird's eye of our national survey results



## **Visualising queues**

Recreating queue related visuals



## **Cleaning up wells**

Recreating a pollution map



# **Connecting crime**

Linking crime data to find new insight into water data 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.



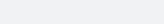




















Setting the stage for our data exploration journey.



# **Buiding a data model**

Unleashing PowerBI with data models



#### **National scale**

Bird's eye of our national survey results



## **Visualising queues**

Recreating queue related visuals



## **Cleaning up wells**

Recreating a pollution map



# **Connecting crime**

Linking crime data to find new insight into water data

# **Dalila Lesedi**

# This is what I came up with:







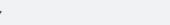




















Setting the stage for our data exploration journey.



## **Buiding a data model**

Unleashing PowerBI with data models



#### **National scale**

Bird's eye of our national survey results



#### **Visualising queues**

Recreating queue related visuals



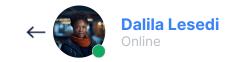
#### Cleaning up wells

Recreating a pollution map



#### **Connecting crime**

Linking crime data to find new insight into water data



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?!



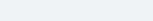




















Setting the stage for our data exploration journey.



## **Buiding a data model**

Unleashing PowerBI with data models



#### **National scale**

Bird's eye of our national survey results



## **Visualising queues**

Recreating queue related visuals



#### **Cleaning up wells**

Recreating a pollution map



# **Connecting crime**

Linking crime data to find new insight into water data



11:49

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

11:50





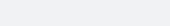






**Dalila Lesedi** 













Setting the stage for our data exploration journey.



# **Buiding a data model**

Unleashing PowerBI with data models



#### **National scale**

Bird's eye of our national survey results



## **Visualising queues**

Recreating queue related visuals



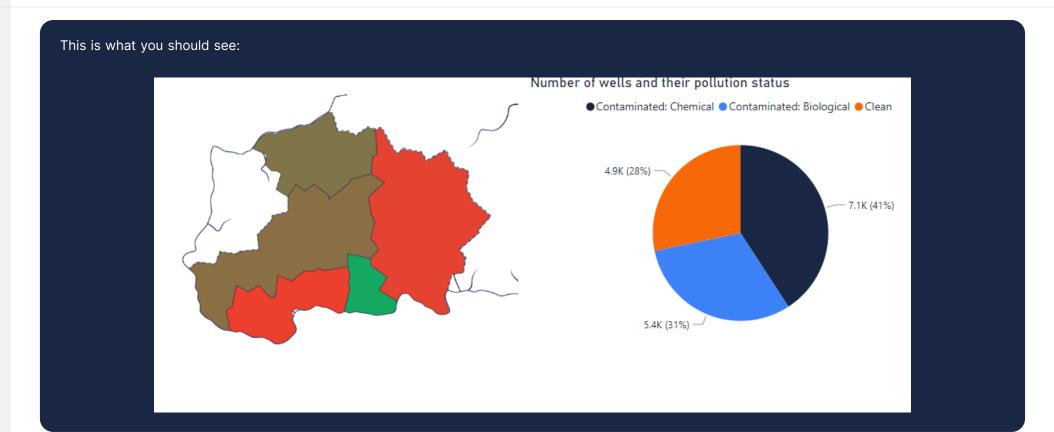
## **Cleaning up wells**

Recreating a pollution map



# **Connecting crime**

Linking crime data to find new insight into water data



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







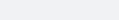






**Dalila Lesedi** 





**Dalila Lesedi** 









## Introduction

Setting the stage for our data exploration journey.



# **Buiding a data model**

Unleashing PowerBI with data models



#### **National scale**

Bird's eye of our national survey results



## **Visualising queues**

Recreating queue related visuals



#### **Cleaning up wells**

Recreating a pollution map



# **Connecting crime**

Linking crime data to find new insight into water data

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.



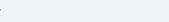






















Setting the stage for our data exploration journey.



# **Buiding a data model**

Unleashing PowerBI with data models



#### **National scale**

Bird's eye of our national survey results



#### **Visualising queues**

Recreating queue related visuals



#### Cleaning up wells

Recreating a pollution map



#### **Connecting crime**

Linking crime data to find new insight into water data 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!

























Setting the stage for our data exploration journey.





#### **National scale**

Bird's eye of our national survey results



#### **Visualising queues**

Recreating queue related visuals



#### **Cleaning up wells**

Recreating a pollution map



# **Connecting crime**

Linking crime data to find new insight into water data 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?



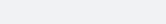




















Setting the stage for our data exploration journey.



# **Buiding a data model**

Unleashing PowerBI with data models



## **National scale**

Bird's eye of our national survey results



## **Visualising queues**

Recreating queue related visuals



## **Cleaning up wells**

Recreating a pollution map



#### **Connecting crime**

Linking crime data to find new insight into water data



## Here's mine:















25















Setting the stage for our data exploration journey.

data models

## **Buiding a data model**

Unleashing PowerBI with

#### National scale

Bird's eye of our national survey results

## **Visualising queues**

Recreating queue related visuals



#### **Cleaning up wells**

Recreating a pollution map



#### **Connecting crime**

Linking crime data to find new insight into water data

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?





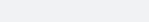




















Setting the stage for our data exploration journey.



## **Buiding a data model**

Unleashing PowerBI with data models



#### **National scale**

Bird's eye of our national survey results



## **Visualising queues**

Recreating queue related visuals



## **Cleaning up wells**

Recreating a pollution map



# **Connecting crime**

Linking crime data to find new insight into water data

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





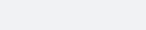






**Dalila Lesedi** 













Setting the stage for our data exploration journey. **Dalila Lesedi** 

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

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

Visualising queues Recreating queue related visuals

Cleaning up wells Recreating a pollution map

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

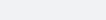
Side note: This is my idea for the crime data page. It is still a work in progress, but this is the type of stuff we're going to do to make the report come alive. What do you think?











**Dalila Lesedi** 









## Introduction

Setting the stage for our data exploration journey.



# **Buiding a data model**

Unleashing PowerBI with data models



#### **National scale**

Bird's eye of our national survey results



## **Visualising queues**

Recreating queue related visuals



## **Cleaning up wells**

Recreating a pollution map



# **Connecting crime**

Linking crime data to find new insight into water data

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!









