# WPDX | Water Points in Zambia Data Analysis

## Project Scope

This project aims to analyze data related to water points in Zambia. The data regarding the water points was compiled by the **Water Point Data Exchange** (WPDX). The WPDX is a platform to compile crowdsourced data focused on rural water points (wells, springs, tap stands, etc.) with contributions from governments, NGOs, and researchers.

This analysis focused on factors such as;

- Number of water points in a district
- Status (functional & non-functional) of the water points
- Type of water source
- Technology used to access water
- Donors of water access technology
- Water points installed in a particular year

## Project Approach

The dataset for this analysis was obtained from the **Humanitarian Data Exchange** (HDX) website. The preparation and cleaning of the data in the dataset was done using **Microsoft Excel**. Thereafter, the dataset was imported into **MySQL** Workbench for analysis using **SQL**. Finally, visualizations of the analyzed data were created using **Tableau**.

## Part 1: Data Preparation and Cleaning

Before the dataset can be imported into MySQL Workbench for analysis, it has to be prepared and cleaned. Good data preparation and cleaning allow for efficient data analysis and limit errors and inaccuracies that can occur to data during analysis. The data preparation was done using Microsoft Excel.

### Step 1: Download the dataset and open the file

Below is a snapshot of the dataset.

| | A | B | C | D | E | F | G | H | I | J | K |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | lat_deg | lon_deg | status_id | report_date | source | water_source_clean | water_tech_clean | clean_cou | clean_country_name | clean_adn | clean_adm2 |
| 2 | #geo+lat | #geo+lon | #indicator | #date+report | #meta+source | #indicator+water_source | #indicator+water_tech | #country+ | #country+name | #adm1+na | #adm2+name |
| 3 | -17.6098 | 26.14396 | Yes | 14/02/2020 | WaterAid UK | Piped Water | | ZMB | Zambia | Southern | Kazungula |
| 4 | -15.4419 | 28.3804 | Yes | 13/06/2017 | Water and Sanitation for the Urban Poor | Piped Water | Public Tapstand | ZMB | Zambia | Lusaka | Lusaka |
| 5 | -16.3778 | 27.86315 | Yes | 05/04/2021 | WaterAid | Piped Water | Public Tapstand | ZMB | Zambia | Southern | Monze |
| 6 | -11.1164 | 29.64958 | Yes | 05/04/2023 | Water4 | Piped Water | Public Tapstand | ZMB | Zambia | Luapula | Chifunabuli |
| 7 | -16.2188 | 27.71373 | Yes | 16/12/2019 | WaterAid UK | Borehole/Tubewell | Hand Pump - Rope | ZMB | Zambia | Southern | Monze |
| 8 | -16.1203 | 28.50933 | Yes | 10/10/2014 | Living Water International | Borehole/Tubewell | Hand Pump - India Mark | ZMB | Zambia | Lusaka | Chirundu |
| 9 | -16.0243 | 28.81783 | Yes | 17/08/2014 | Living Water International | Borehole/Tubewell | Hand Pump - India Mark | ZMB | Zambia | Lusaka | Chirundu |
| 10 | -15.1695 | 22.66039 | Yes | 05/02/2014 | Village Water Zambia | Borehole/Tubewell | Hand Pump | ZMB | Zambia | Western | Kalabo |
| 11 | -15.1207 | 22.71119 | Yes | 19/08/2018 | Village Water Zambia | Borehole/Tubewell | Hand Pump | ZMB | Zambia | Western | Kalabo |
| 12 | -16.9924 | 24.69335 | Yes | 25/02/2020 | WaterAid UK | Piped Water | Public Tapstand | ZMB | Zambia | Western | Mwandi |
| 13 | -15.6805 | 23.26918 | Yes | 14/03/2017 | Village Water Zambia | Borehole/Tubewell | Hand Pump | ZMB | Zambia | Western | Nalolo |
| 14 | -15.631 | 28.11635 | Yes | 19/09/2014 | SNV-Zambia | Protected Well | Hand Pump - India Mark | ZMB | Zambia | Lusaka | Kafue |
| 15 | -15.2123 | 28.78044 | Yes | 16/09/2014 | SNV-Zambia | Protected Spring | Rope and Bucket | ZMB | Zambia | Lusaka | Chongwe |
| 16 | -15.0597 | 23.35711 | Yes | 19/03/2018 | District Joint Monitoring Team (DJMT) | Borehole/Tubewell | Hand Pump | ZMB | Zambia | Western | Limulunga |
| 17 | -15.2573 | 29.1905 | Yes | 18/08/2016 | Living Water International | Borehole/Tubewell | Hand Pump - India Mark | ZMB | Zambia | Lusaka | Rufunsa |
| 18 | -14.5067 | 24.89886 | Yes | 25/02/2015 | Village Water Zambia | Borehole/Tubewell | Hand Pump | ZMB | Zambia | Western | Kaoma |
| 19 | -15.4183 | 28.98783 | Yes | 20/09/2016 | Living Water International | Borehole/Tubewell | | ZMB | Zambia | Lusaka | Rufunsa |
| 20 | -14.8531 | 25.1466 | Yes | 22/09/2014 | Village Water Zambia | Borehole/Tubewell | Hand Pump | ZMB | Zambia | Western | Nkeyema |
| 21 | -14.7505 | 27.86366 | Yes | 14/09/2022 | Living Water International | Borehole/Tubewell | Hand Pump | ZMB | Zambia | Central | Chibombo |
| 22 | -15.2857 | 28.66212 | Yes | 05/10/2014 | SNV-Zambia | Borehole/Tubewell | Hand Pump - India Mark | ZMB | Zambia | Lusaka | Chongwe |
| 23 | -14.6337 | 27.59867 | Yes | 22/12/2020 | Living Water International | Borehole/Tubewell | | ZMB | Zambia | Central | Chibombo |

The dataset was comprised of **6, 127** records and **54** fields.

## Step 2: Expand Dataset Columns as Necessary

## Step 3: Review the Data

This step is of utmost importance. There's a need to make sure the data is correct and in the required format. This is done by formatting the dataset as a table. To do this, the keyboard shortcut, **Ctrl + A > Ctrl + T > Ok**, is used and the dataset should look something like this:

| | A | B | C | D | E | F | G | H | I | J |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | lat_deg | lon_deg | status_id | report_date | source | water_source_clean | water_tech_clean | clean_country_id | clean_country_name | clean_adm1 |
| 2 | #geo+lat | #geo+lon | #indicator+v | #date+reporte | #meta+source | #indicator+water_source | #indicator+water_tech | #country+code+v_is | #country+name | #adm1+name |
| 3 | -17.60979 | 26.143956 | Yes | 14/02/2020 | WaterAid UK | Piped Water | | ZMB | Zambia | Southern |
| 4 | -15.4419 | 28.3804 | Yes | 13/06/2017 | Water and Sanitation for the Urban Poor | Piped Water | Public Tapstand | ZMB | Zambia | Lusaka |
| 5 | -16.37783 | 27.863147 | Yes | 05/04/2021 | WaterAid | Piped Water | Public Tapstand | ZMB | Zambia | Southern |
| 6 | -11.11638 | 29.649575 | Yes | 05/04/2023 | Water4 | Piped Water | Public Tapstand | ZMB | Zambia | Luapula |
| 7 | -16.21877 | 27.713727 | Yes | 16/12/2019 | WaterAid UK | Borehole/Tubewell | Hand Pump - Rope | ZMB | Zambia | Southern |
| 8 | -16.12033 | 28.509333 | Yes | 10/10/2014 | Living Water International | Borehole/Tubewell | Hand Pump - India Mark | ZMB | Zambia | Lusaka |
| 9 | -16.02433 | 28.817833 | Yes | 17/08/2014 | Living Water International | Borehole/Tubewell | Hand Pump - India Mark | ZMB | Zambia | Lusaka |
| 10 | -15.16947 | 22.660389 | Yes | 05/02/2014 | Village Water Zambia | Borehole/Tubewell | Hand Pump | ZMB | Zambia | Western |
| 11 | -15.12069 | 22.711194 | Yes | 19/08/2018 | Village Water Zambia | Borehole/Tubewell | Hand Pump | ZMB | Zambia | Western |
| 12 | -16.9924 | 24.693349 | Yes | 25/02/2020 | WaterAid UK | Piped Water | Public Tapstand | ZMB | Zambia | Western |
| 13 | -15.6805 | 23.26918 | Yes | 14/03/2017 | Village Water Zambia | Borehole/Tubewell | Hand Pump | ZMB | Zambia | Western |
| 14 | -15.63101 | 28.11635 | Yes | 19/09/2014 | SNV-Zambia | Protected Well | Hand Pump - India Mark | ZMB | Zambia | Lusaka |
| 15 | -15.21231 | 28.780438 | Yes | 16/09/2014 | SNV-Zambia | Protected Spring | Rope and Bucket | ZMB | Zambia | Lusaka |
| 16 | -15.05965 | 23.35711 | Yes | 19/03/2018 | District Joint Monitoring Team (DJMT) | Borehole/Tubewell | Hand Pump | ZMB | Zambia | Western |
| 17 | -15.25733 | 29.1905 | Yes | 18/08/2016 | Living Water International | Borehole/Tubewell | Hand Pump - India Mark | ZMB | Zambia | Lusaka |
| 18 | -14.50671 | 24.89886 | Yes | 25/02/2015 | Village Water Zambia | Borehole/Tubewell | Hand Pump | ZMB | Zambia | Western |
| 19 | -15.41833 | 28.987833 | Yes | 20/09/2016 | Living Water International | Borehole/Tubewell | | ZMB | Zambia | Lusaka |
| 20 | -14.85305 | 25.1466 | Yes | 22/09/2014 | Village Water Zambia | Borehole/Tubewell | Hand Pump | ZMB | Zambia | Western |
| 21 | -14.75052 | 27.863662 | Yes | 14/09/2022 | Living Water International | Borehole/Tubewell | Hand Pump | ZMB | Zambia | Central |
| 22 | -15.28569 | 28.66212 | Yes | 05/10/2014 | SNV-Zambia | Borehole/Tubewell | Hand Pump - India Mark | ZMB | Zambia | Lusaka |
| 23 | -14.63372 | 27.59867 | Yes | 22/12/2020 | Living Water International | Borehole/Tubewell | | ZMB | Zambia | Central |

### Step 3.1: Delete the second row

The second row had to be deleted as it was a duplicate of the first row. For example, cell **A2** contained the data **#geo+lat,** whereas cell **A1** had **lat_deg,** the two cells were both referring to the **latitude** in degrees. This was the case with all the cells in row 2. As such, there was no need to keep the data in the second row.

| | A | B | C | D | E | F | G | H | I |
|---|---|---|---|---|---|---|---|---|---|
| 1 | lat_deg | lon_deg | status_id | report_date | source | water_source_clean | water_tech_clean | clean_country_id | clean_country_name |
| 2 | -17.60979 | 26.143956 | Yes | 14/02/2020 | WaterAid UK | Piped Water | | ZMB | Zambia |
| 3 | -15.4419 | 28.3804 | Yes | 13/06/2017 | Water and Sanitation for the Urban Poor | Piped Water | Public Tapstand | ZMB | Zambia |
| 4 | -16.37783 | 27.863147 | Yes | 05/04/2021 | WaterAid | Piped Water | Public Tapstand | ZMB | Zambia |
| 5 | -11.11638 | 29.649575 | Yes | 05/04/2023 | Water4 | Piped Water | Public Tapstand | ZMB | Zambia |
| 6 | -16.21877 | 27.713727 | Yes | 16/12/2019 | WaterAid UK | Borehole/Tubewell | Hand Pump - Rope | ZMB | Zambia |

### Step 3.2: Rename some field (column) names

Some columns had vague names, for example, **clean_adm1** instead of **Province** or **clean_adm2** instead of **District**. Such fields and many others, had to be renamed so that the field names could be more specific.

| | G | H | I | J | K | L | M |
|---|---|---|---|---|---|---|---|
| 1 | Water_Tech_Clean | Clean_Country_Id | Clean_Country_Name | Clean_Provinc | Clean_Distri | clean_adm3 | clean_adm4 |
| 2 | | ZMB | Zambia | Southern | Kazungula | | |
| 3 | Public Tapstand | ZMB | Zambia | Lusaka | Lusaka | | |
| 4 | Public Tapstand | ZMB | Zambia | Southern | Monze | | |
| 5 | Public Tapstand | ZMB | Zambia | Luapula | Chifunabuli | | |
| 6 | Hand Pump - Rope | ZMB | Zambia | Southern | Monze | | |
| 7 | Hand Pump - India Mark | ZMB | Zambia | Lusaka | Chirundu | | |
| 8 | Hand Pump - India Mark | ZMB | Zambia | Lusaka | Chirundu | | |

### Step 3.3: Delete columns that are not required or are blank

There were columns such as **adm3** and **adm4** which were **blank**, such cells were deleted as they did not contain any data. Additionally, columns containing data not required for the analysis were deleted. For example, **Fecal_Coliform_Value**, **Prediction_no_2y**, **Predicted_status_0y**, etc. The purpose of the analysis was centred on water availability, source, location as well as the state of the water infrastructure. As such, columns which contained non-relevant data were deleted.

| | E | F | G | H | I | J | K | L | M |
|---|---|---|---|---|---|---|---|---|---|
| 1 | NGO_Source | Water_Source | Water_Tech | Country | Province | District | WPDX_Id | Install_Year | Installer |
| 2 | WaterAid UK | Piped Water | | Zambia | Southern | Kazungula | 5GJ894RV+2G9 | 2019 | WaterAid |
| 3 | Water and Sanitation for the Urban Poor | Piped Water | Public Tapstand | Zambia | Lusaka | Lusaka | 5GPCH95J+65R | | |
| 4 | WaterAid | Piped Water | Public Tapstand | Zambia | Southern | Monze | 5GM9JVC7+W69 | 2020 | WaterAid |
| 5 | Water4 | Piped Water | Public Tapstand | Zambia | Luapula | Chifunabuli | 5GWFVJMX+CRW | 2021 | |
| 6 | WaterAid UK | Borehole/Tubewell | Hand Pump - Rope | Zambia | Southern | Monze | 5GM9QPJ7+FFX | 2019 | WaterAid |
| 7 | Living Water International | Borehole/Tubewell | Hand Pump - India Mark | Zambia | Lusaka | Chirundu | 5GMCVGH5+VP8 | 2014 | |
| 8 | Living Water International | Borehole/Tubewell | Hand Pump - India Mark | Zambia | Lusaka | Chirundu | 5GMCXRG9+748 | 2014 | |

After deleting the unnecessary columns, the remaining columns amounted to **26**. This implied that **28** irrelevant columns had been deleted.

### Step 3.4: Change the date format from dd/mm/yyyy to yyyy/mm/dd

The dataset contained a field named **Report_Date**, the data in this field had the **DD/MM/YYYY** format. This had to be changed to **YYYY/MM/DD**, as this is the accepted MySQL format.

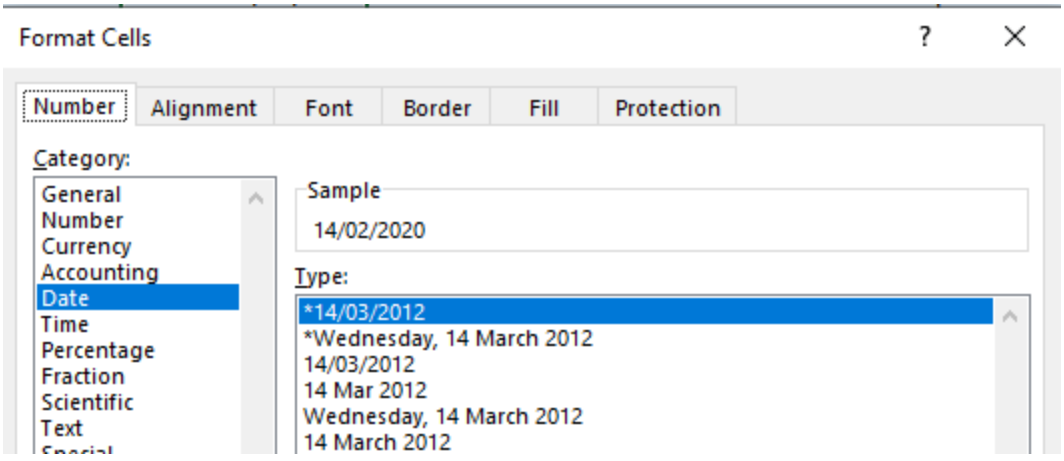
### Step 3.4.1

Select the date cells in the cell range **D2:D6128** using the keyboard shortcut **Shift + Ctrl + down arrow**.

| Report_Date |
|---|
| 14/02/2020 |
| 13/06/2017 |
| 05/04/2021 |
| 05/04/2023 |
| 16/12/2019 |
| 10/10/2014 |
| 17/08/2014 |

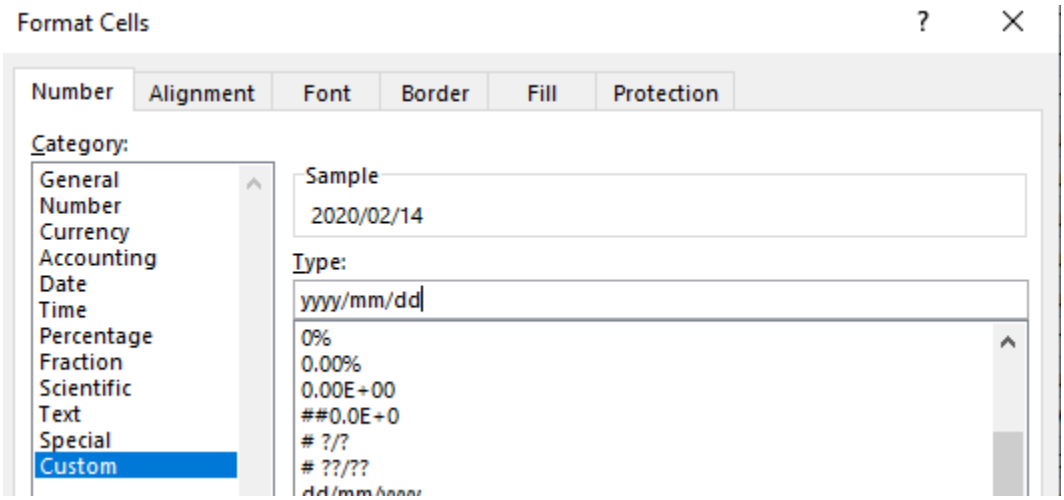
### Step 3.4.2

Activate the **Format Cells** dialogue box using the keyboard shortcut **Ctrl + 1**.

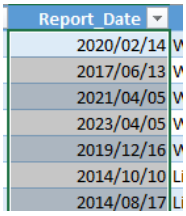


**Step 3.4.3**

The **Custom** Category was selected and the **yyyy/mm/dd** format was entered in the **Type**: combo box and the **OK** button was clicked.

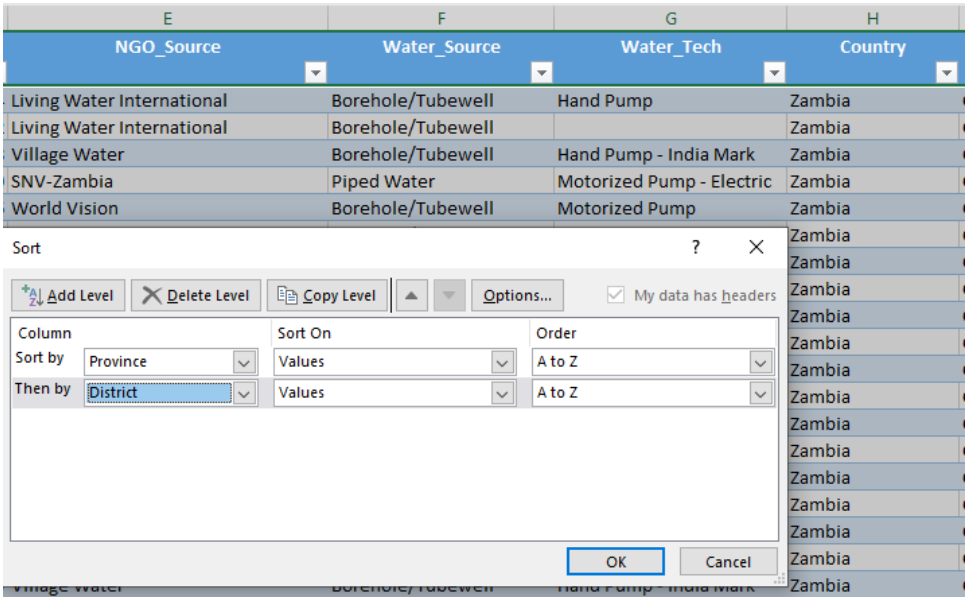



The date format was successfully converted to 'yyyy/mm/dd' from 'dd/mm/yyyy'.




**Step 3.5: Sort the dataset in ascending order according to provinces**

The dataset had to be sorted according to provinces so that it could be segmented as such. This was done in the following manner; **Ctrl + A** > **Data tab** > **Sort** > **Sort by** > **Province** > **Add Level** > **Then by** > **District** > **OK.**



Below is a snapshot of the sorted dataset.

| F | G | H | I | J |
|---|---|---|---|---|
| Water_Source | Water_Tech | Country | Province | District |
| Borehole/Tubewell | Hand Pump | Zambia | Central | Chibombo |
| Borehole/Tubewell | | Zambia | Central | Chibombo |
| Borehole/Tubewell | Hand Pump - India Mark | Zambia | Central | Chibombo |
| Borehole/Tubewell | Hand Pump - India Mark | Zambia | Central | Chibombo |
| Borehole/Tubewell | Hand Pump - India Mark | Zambia | Central | Chibombo |
| Borehole/Tubewell | Hand Pump - India Mark | Zambia | Central | Chibombo |
| Protected Well | Hand Pump - Rope | Zambia | Central | Chibombo |
| Protected Well | Hand Pump - Rope | Zambia | Central | Chibombo |
| Borehole/Tubewell | Hand Pump - India Mark | Zambia | Central | Chibombo |
| Protected Well | Hand Pump - Rope | Zambia | Central | Chibombo |
| Borehole/Tubewell | Hand Pump - India Mark | Zambia | Central | Chibombo |
| Borehole/Tubewell | Hand Pump - India Mark | Zambia | Central | Chibombo |
| Borehole/Tubewell | Hand Pump - India Mark | Zambia | Central | Chibombo |

**Step 3.6: Convert the data in columns V (Distance_to_Primary) to Z (Distance_to_Town) from meters to kilometres**.

The dataset did not specify whether the distance values in columns V to Z were in meters or kilometres. Looking at the screenshot below, it seems likely that the distance of **2064** meters (2km) from the water point to a **primary** school as being likely possible. I live in Lusaka and that distance in a rural part of Lusaka seems accurate.

As **1000** meters are equivalent to **1** kilometre, the values in the range **V2:Z6127** will be divided by **1000** to convert the values.

**Step 3.6.1**

Change the column headings for columns V to Z to, for example, Distance_to_Primary to Distance_to_Primary_Km.

| V | W | X | Y | Z |
|---|---|---|---|---|
| Distance_to_Primary_Km | Distance_to_Secondary_Km | Distance_to_Tertiary_Km | Distance_to_City_Km | Distance_to_Town_Km |
| 73.69890483 | 17.51308831 | 0.996617963 | 86.33483808 | 60.70482017 |
| 101.4347034 | 26.32648796 | 0.778420217 | 113.5945755 | 69.6851049 |

**Step 3.6.2**

Enter a value of 1000 in any of the blank cells adjacent to the section of the dataset containing data, for example, cell AC2.

| AC |
|---|
| Column1 |
| 1000 |

**Step 3.6.3**

Copy the value (1000) in the cell AC2.

**Step 3.6.4**

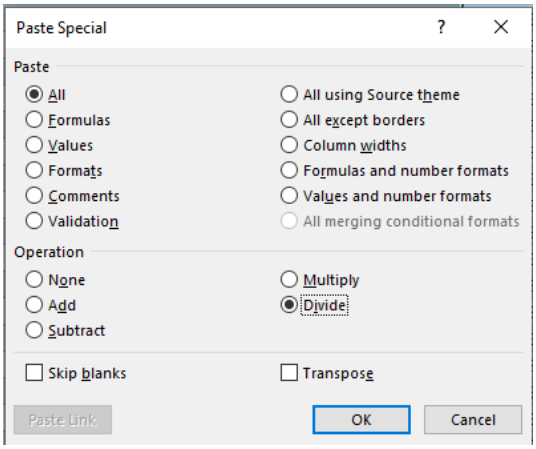
Select the range V2:Z6127.

Using the mouse pointer to select cells in the range V2:Z2 and then using the keyboard shortcut **Ctrl + Shift + down arrow** to select the entire range V2:Z6127.

| V | W | X | Y | Z |
|---|---|---|---|---|
| Distance_to_Primai | Distance_to_Seconda | Distance_to_Tertia | Distance_to_City | Distance_to_Tov |
| 73698.90483 | 17513.08831 | 996.6179633 | 86334.83808 | 60704.82017 |
| 101434.7034 | 26326.48796 | 778.4202172 | 113594.5755 | 69685.1049 |
| 133482.0122 | 35659.06467 | 214.0770569 | 119833.3292 | 23693.24457 |
| 74907.34946 | 29387.76646 | 7203.077275 | 83943.99447 | 55966.6152 |
| 194899.7041 | 49527.62202 | 23083.96049 | 185161.9576 | 55945.98174 |
| 194214.1823 | 50345.96786 | 19035.01048 | 184451.835 | 57473.52128 |
| 213197.8849 | 64220.75068 | 61142.40861 | 235415.6933 | 64327.94001 |
| 211330.9995 | 84562.94846 | 65776.13386 | 248987.9511 | 84599.41035 |
| 239413.2135 | 59988.34555 | 56154.63418 | 263449.4604 | 60049.41786 |
| 217815.8121 | 82977.60573 | 55683.63769 | 261044.5556 | 84209.19052 |
| 229576.7648 | 71542.74764 | 60444.33983 | 264928.6852 | 73026.40815 |
| 234892.975 | 67564.15356 | 58882.45074 | 270620.7194 | 71200.39344 |

**Step 3.6.4**

Use the keyboard shortcut **Ctrl + Alt + V** to activate **Paste Special** > **Operation** > **Divide** > **OK**.

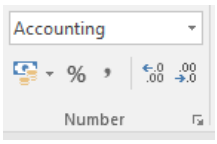



The values in the range are converted to kilometres and the 1000 in cell AC2 can be deleted as it is of no use anymore.

| V | W | X | Y | Z |
|---|---|---|---|---|
| Distance_to_Primai | Distance_to_Seconda | Distance_to_Tertia | Distance_to_City | Distance_to_Tov |
| 73.69890483 | 17.51308831 | 0.996617963 | 86.33483808 | 60.70482017 |
| 101.4347034 | 26.32648796 | 0.778420217 | 113.5945755 | 69.6851049 |
| 133.4820122 | 35.65906467 | 0.214077057 | 119.8333292 | 23.69324457 |
| 74.90734946 | 29.38776646 | 7.203077275 | 83.94399447 | 55.9666152 |
| 194.8997041 | 49.52762202 | 23.08396049 | 185.1619576 | 55.94598174 |
| 194.2141823 | 50.34596786 | 19.03501048 | 184.451835 | 57.47352128 |

**Step 3.6.5**

Reduce the number of decimal places to two (2). This was done using the '**,**' **Comma Style** button located on the right – side of the **%** button.



Below is a screenshot of the values with two decimal places.

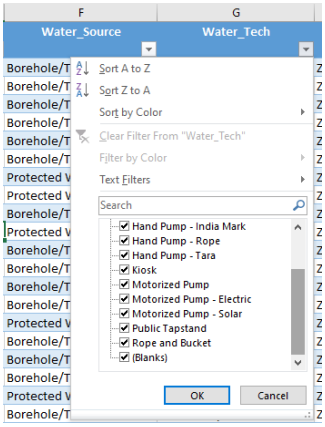
| V | W | X | Y | Z |
|---|---|---|---|---|
| Distance_to_Primary _Km | Distance_to_Secondary _Km | Distance_to_Tertiary _Km | Distance_to_City_Km | Distance_to_Town _Km |
| 73.70 | 17.51 | 1.00 | 86.33 | 60.70 |
| 101.43 | 26.33 | 0.78 | 113.59 | 69.69 |
| 133.48 | 35.66 | 0.21 | 119.83 | 23.69 |
| 74.91 | 29.39 | 7.20 | 83.94 | 55.97 |
| 194.90 | 49.53 | 23.08 | 185.16 | 55.95 |
| 194.21 | 50.35 | 19.04 | 184.45 | 57.47 |

### Step 3.7: Fill up all blank cells

All the blank cells had to be filled with data. This was necessary as importing the data into MySQL would raise errors and the process would not be successful.

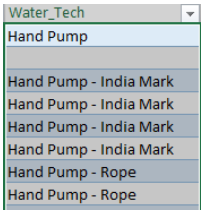
### Step 3.7.1

Check each of the columns for blank cells using the drop-down button next to the header name.



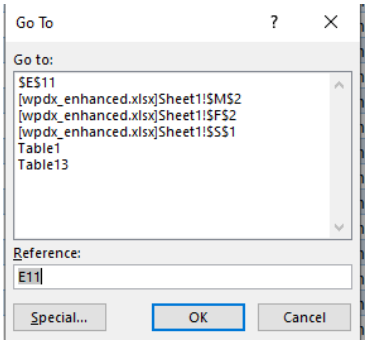

### Step 3.7.2

Having identified each of the columns containing blank cells, select the section of the column containing data, except the column headers. This was done by typing the range in the name box, for example, **G2:G6127 + Enter.**



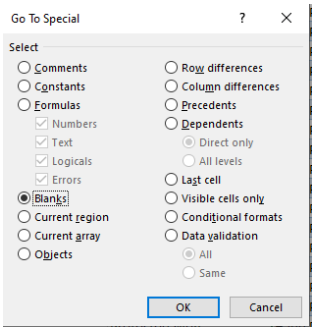
**Step 3.7.3**

Use the keyboard shortcut **Ctrl + G** is to activate the **Go To** dialog box. **Go To > Special**.




**Step 3.7.4**

After the **Go To Special** dialog box appears, select **Blanks > OK.**

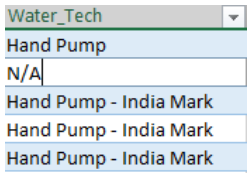



**Step 3.7.5**

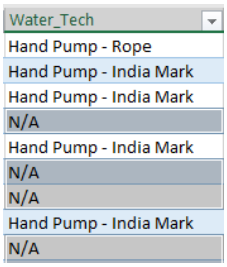
After checking the data source, HDX, the missing data could not be found. Additionally, the blank cell containing, for example, population figures could not be simply guessed. Deleting the blank cells would result in the cells being shifted upwards, thereby causing most of the records in the dataset to have inaccurate data.

As such, I (the data analyst) decided to fill blank cells in columns containing **numeric** data, with **0s**. On the other hand, **N/A** would be entered in the blank cells of columns consisting of **text/string** data.

To enter N/A in a blank cell such as G3, press **F2** to turn cell G3 into an active cell.



**Step 3.7.6**
Having typed N/A in cell G3, use to keyboard shortcut **Ctrl + Enter**, to fill all the other blank cells with N/A in the cell range G2:G6127.




Filling up all the blank cells with the appropriate data, marked the end of the data preparation and cleaning phase.

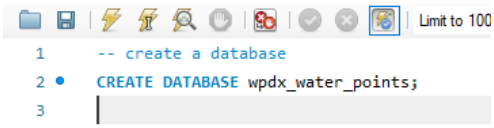
The Format Worksheet as a Table was turned off by **Right–clicking** in any cell in the table > **menu** > **Table** > **Convert to Range**.

# Part 2: Import the Data

After the dataset has been had been cleaned, it had to be imported into **MySQL Workbench**. For Part 2, all the tasks were carried out using **SQL**.

## Step 1: Create a database

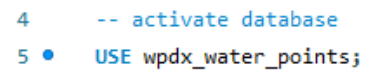
First off, a database was created as this is where the data from the dataset will be kept. The dataset can be stored in a table from an existing database or a new one can be created. For this project, a new database will be created using MySQL Workbench.




A new database named **wpdx_water_points** has been created. The two **- -** hyphens at the beginning of line 1 indicate that it is a comment and it doesn't affect the SQL statements.

## Step 2: Activate the database

The database that had been created had to be activated for it to be used.




## Step 3: Create a table

A table which would hold the data from the dataset had to be created. The columns within the table must match the columns in the dataset in name and data type. The table would be named **zambian_water_points**. The screenshot below is a portion of the SQL statement.

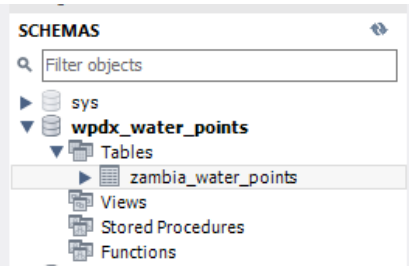
```
7     -- create a table
8  ●  CREATE TABLE zambia_water_points (
9                                   Latitude float,
10                                  Longitude float,
11                                  Status_Id char (5),
12                                  Report_Date date,
13                                  NGO_Source char (50),
14                                  Water_Source char (30),
15                                  Water_Tech char (30),
```

Here, a table called zambia_water_points was created, it was designed in a way to fit the dataset by matching the columns and data types.

The columns were created in this manner (as in the example above): first off, specify the column **name**, then add the **data type** of that column and in the parenthesis, the **size** of the variable was entered.

It was important to check the data in the dataset file before specifying the field sizes in the database table. For example, specifying that 'Province' or 'Water_Tech', or whatever field it was, was of a max size of 20 would imply that rows which contain fields that have more than 20 characters would not be imported.

To confirm if the table has been created, go to the left panel of the MySQL Workbench; double click on the database (zambian_water_points) name in the left panel and then right-click on the table recently created: If the created table does not appear in the left panel, click the refresh button next to **SCHEMAS**.

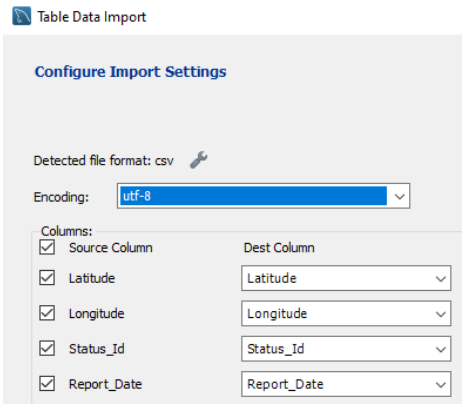



## Step 4: Import the dataset file

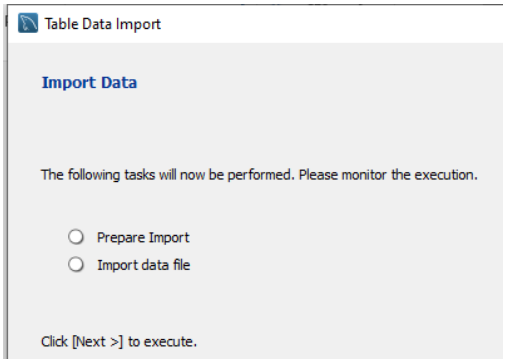
To import the data from the dataset file, place the cursor over the table name > right-click. From the drop-down menu, select the 'Table Data Import Wizard' option. After selecting the needed option, a 'Table Data Import' window will open pop up. This window makes it possible to browse for the dataset file that has to be imported. Look it up and when the file is located click on the next button.

After clicking ‘Next >’, MySQL will ask for the destination table where the data from the CSV file will go and then click on the ‘Next >’ button.

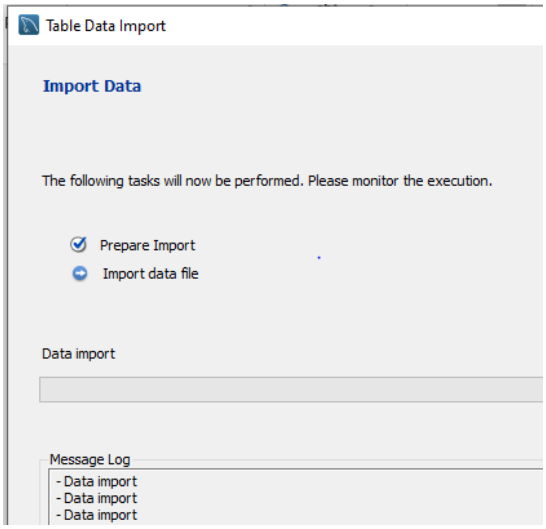

Ensure the source and destination columns match. Where necessary make changes using the drop-down arrow. When the columns match, click on the ‘Next >’ button.

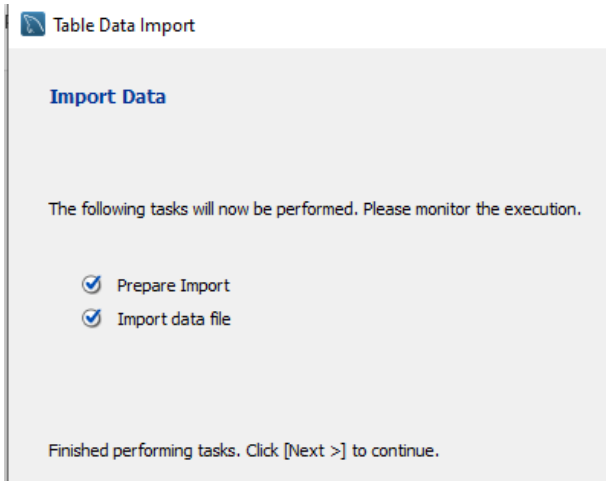



After clicking the ‘Next >’ button, the window shown below appears.

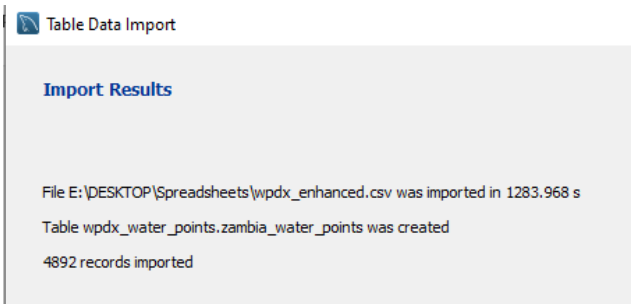


The data importation process begins after clicking the ‘Next >’ button.




After the import data process is complete, click on the ‘Next >’ button.




Finally, click on the ‘Finish’ button to complete the data import process.

```
37    -- show the imported dataset
38 •  SELECT * FROM zambia_water_points;
39
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content: | Fetch rows:

| Latitude | Longitude | Status_Id | Report_Date | NGO_Source | Water_Source | Water_Tech | Country | Province |
|---|---|---|---|---|---|---|---|---|
| -14.75 | 27.86 | Yes | 2022-09-14 | Living Water International | Borehole/Tubewell | Hand Pump | Zambia | Central |
| -14.63 | 27.6 | Yes | 2020-12-22 | Living Water International | Borehole/Tubewell | N/A | Zambia | Central |
| -15.27 | 27.75 | Yes | 2014-11-04 | SNV-Zambia | Borehole/Tubewell | Hand Pump - India Mark | Zambia | Central |
| -15.26 | 27.74 | Yes | 2014-11-04 | SNV-Zambia | Borehole/Tubewell | Hand Pump - India Mark | Zambia | Central |
| -15.26 | 27.75 | Yes | 2014-11-03 | SNV-Zambia | Borehole/Tubewell | Hand Pump - India Mark | Zambia | Central |
| -15.26 | 27.74 | Yes | 2014-11-04 | SNV-Zambia | Borehole/Tubewell | Hand Pump - India Mark | Zambia | Central |
| -15.25 | 27.74 | Yes | 2014-11-04 | SNV-Zambia | Protected Well | Hand Pump - Rope | Zambia | Central |
| -15.25 | 27.75 | Yes | 2014-11-03 | SNV-Zambia | Protected Well | Hand Pump - Rope | Zambia | Central |

Result Grid | Form Editor

The screenshot above shows a portion of the imported dataset.

After going through the imported dataset, it was noted that one of the field headers was incorrect. It was named ‘Date_Since_Report’ instead of ‘Days_Since_Report’. The SQL statement in line **41** was used to rectify this issue.

```
40      -- renaming a column header
41 •    ALTER TABLE zambia_water_points RENAME COLUMN Date_Since_Report TO Days_Since_Report;
42 •    SELECT * FROM zambia_water_points;
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content: | Fetch rows:

| mary_Km | Distance_to_Secondary_Km | Distance_to_Tertiary_Km | Distance_to_City_Km | Distance_to_Town_Km | Is_Urban | Days_Since_Report |
|---|---|---|---|---|---|---|
| | 17.51 | 1 | 86.33 | 60.7 | FALSE | 445 |
| | 26.33 | 0.78 | 113.59 | 69.69 | FALSE | 1076 |

# Part 4: Exploratory Data Analysis (EDA)

## Step 1: Number of columns and rows in the ‘zambia_water_points’ table

What is the shape of the database table, i.e., the number of columns and rows?

```
45      -- number of rows in the dataset
46 •    SELECT COUNT(*) AS rows_num FROM zambia_water_points;
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

| rows_num |
|---|
| 4892 |

```
48      -- check the number of columns
49 •    SELECT COUNT(*) AS cols_num FROM information_schema.columns WHERE TABLE_NAME = "zambia_water_points";
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

| cols_num |
|---|
| 26 |

The CSV dataset had 6,127 records, but for some reason only 4,892 records were imported. The data import was attempted again, it was discovered that for some records, an import error occurred. Some of the errors which occurred are in the parenthesis for each of the fields below;

- Distance_to_City_Km (“Data truncated ‘Distance_to_City_Km’ at row 1”).
- Installer (“Data too long for column ‘Installer’ at row 1”, 1406)
- Installation_Year (“Incorrect integer value for column ‘Installation_Year’ at row 1”, 1366)
- Status_Id (“Data too long for column ‘Status_Id’ at row 1”,1406)
- Province (“Data too long for column ‘Province’ at row 1”, 1406)

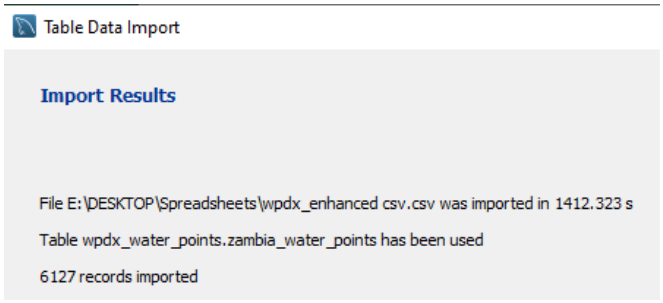
The following measures were implemented to resolve the errors above;

```
-- modify table to successfully import the data for the csv dataset
-- drop distance to city column
ALTER TABLE zambia_water_points DROP COLUMN Distance_to_City_Km;

-- modify the length of the data, lines 50 - 52, modify datatype line 53
ALTER TABLE zambia_water_points
MODIFY COLUMN Installer char (25),
MODIFY COLUMN Status_Id char (10),
MODIFY COLUMN Province char (15),
MODIFY COLUMN Installation_Year char (10);
```

The Distance_to_City_Km column was dropped as it was missing a lot of data.
**Note**: Installer field length was further modified from char (25) to (40).

Thereafter, the data was imported using the method mentioned on pages 10 to 12.




All the 6127 records were imported successfully.

To complete step 1: Number of columns and rows in the 'zambia_water_points' table.

```
-- exploratory data analysis
-- check the number of rows in the dataset
SELECT COUNT(*) AS rows_num FROM zambia_water_points;
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

| rows_num |
|---|
| 6115 |

```
-- check the number of columns
SELECT COUNT(*) AS cols_num FROM information_schema.columns WHERE TABLE_NAME = "zambia_water_points";
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

| cols_num |
|---|
| 25 |

There are **6115** records and **25** columns/fields which have been imported.

## Step 2: Distinct Values

The SQL statements below were used to look up distinct values in some of the fields.

### 2.1: NGO Source

```
64    -- look up distinct values of some of the fields
65 •  SELECT DISTINCT NGO_Source FROM zambia_water_points;
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

| NGO_Source |
|---|
| Living Water International |
| SNV-Zambia |
| Village Water |
| World Vision |
| Village Water Zambia |
| District Joint Monitoring Team (DJMT) |
| Water4 |
| WaterAid |
| WaterAid UK |
| UNHCR |
| Water and Sanitation for the Urban ... |
| Rain Water for Africa |

The were were twelve (**12**) different Non-Governmental Organisations (NGOs) which were providing water supply solutions to the communities.

### 2.2: Water Source

```
66 •  SELECT DISTINCT Water_Source FROM zambia_water_points;
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

| Water_Source |
|---|
| Borehole/Tubewell |
| Protected Well |
| Protected Spring |
| Piped Water |
| N/A |
| Undefined Well |

There were five (5) water sources which were listed in the dataset.

### 2.3: Water Technology

```
67 •  SELECT DISTINCT Water_Tech FROM zambia_water_points;
```

Result Grid | Filter Rows: | Export: | Wrap Cell Conter

| Water_Tech |
|---|
| Hand Pump |
| N/A |
| Hand Pump - India Mark |
| Hand Pump - Rope |
| Hand Pump - Afridev |
| Rope and Bucket |
| Motorized Pump |
| Motorized Pump - Electric |
| Hand Pump - Blue |
| Public Tapstand |
| Kiosk |
| Hand Pump - Tara |
| Motorized Pump - Solar |

There were twelve (**12**) water supply technologies which were used.

## 2.4: Provinces

```
68 •  SELECT DISTINCT Province FROM zambia water points;
```

Result Grid | Filter Rows: | Export: | Wrap Cell C

| Province |
|---|
| Central |
| Copperbelt |
| Eastern |
| Luapula |
| Lusaka |
| Muchinga |
| Northern |
| North-Western |
| Southern |
| Western |

There were ten (**10**) provinces which were covered.

## 2.5: Districts

```
69 •  SELECT DISTINCT District FROM zambia_water_points;
```

Result Grid | Filter Rows: | Export: | Wrap Cell Conte

| District |
|---|
| Chibombo |
| Chisamba |
| Itezhi-tezhi |
| Kabwe |
| Kapiri Mposhi |
| Mumbwa |
| Chingola |
| Kitwe |
| Lufwanyama |
| Mpongwe |

| | | | |
|---|---|---|---|
| 34 | 14:28:56 | SELECT DISTINCT District FROM zambia_water_points LIMIT 0, 1000 | 78 row(s) returned |

Seventy-eight (**78**) districts across the ten provinces were covered.

## 2.6: Installation Year

```
70 •  SELECT DISTINCT Installation_Year FROM zambia_water_points ORDER BY 1 DESC;
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

| Installation_Year |
|---|
| N/A |
| 2023 |
| 2022 |
| 2021 |
| 2020 |
| 2019 |
| 2018 |
| 2017 |
| 2016 |
| 2015 |

| | | | |
|---|---|---|---|
| 37 | 14:44:31 | SELECT DISTINCT Installation_Year FROM zambia_water_points ORDER BY 1 DESC LI... | 72 row(s) returned |

There were seventy-one (**71**) different years in which water points were installed.

### 2.7: Installer

```
71 •  SELECT DISTINCT Installer FROM zambia_water_points ORDER BY 1 ASC;
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

| Installer |
|---|
| Abantu. zambia |
| Abesu |
| ADB |
| ADB\|JICA |
| ADB\|personal |
| africare |
| Amali |
| AMIT |
| Annamillar |
| ARISH AID |

39 14:51:01 SELECT DISTINCT Installer FROM zambia_water_points ORDER BY 1 ASC LIMIT 0, 1000 194 row(s) returned

There were one hundred ninety-one (**191**) installers of water points across the country.

## Step 3: Analytical Questions

### 3.1: How many water points exist in each province?

```
74    -- How many water points exist in each province?
75 •  SELECT Province, COUNT(DISTINCT District) AS Number_of_Districts, COUNT(DISTINCT WPDX_Id) AS Number_of_Water_Points,
76    ROUND((COUNT(DISTINCT WPDX_Id) / (SELECT COUNT(DISTINCT WPDX_Id) FROM zambia_water_points)) * 100, 1)
77    AS Percentage FROM zambia_water_points GROUP BY Province ORDER BY 3 DESC;
78
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

| Province | Number_of_Districts | Number_of_Water_Points | Percentage |
|---|---|---|---|
| Lusaka | 8 | 1643 | 26.9 |
| Luapula | 9 | 1062 | 17.4 |
| Central | 6 | 966 | 15.8 |
| Western | 12 | 801 | 13.1 |
| Northern | 7 | 523 | 8.6 |
| Southern | 13 | 476 | 7.8 |
| Muchinga | 5 | 357 | 5.8 |
| North-Western | 9 | 184 | 3.0 |
| Eastern | 4 | 86 | 1.4 |
| Copperbelt | 5 | 17 | 0.3 |

**Lusaka** had the largest number of water points amounting to **1643** which represented **26.9%.** The **Copperbelt** province had the least number of water points at **17**, representing **0.3%.**

### 3.2: How many water points exist in each district?

```
79    -- how many water points exist in each district?
80    SELECT District, COUNT(DISTINCT WPDX_Id) AS Number_of_Water_Points,
81    ROUND((COUNT(DISTINCT WPDX_Id) / (SELECT COUNT(DISTINCT WPDX_Id) FROM zambia_water_points)) * 100, 3)
82    AS Percentage FROM zambia_water_points GROUP BY District ORDER BY 2 DESC;
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

| District | Number_of_Water_Points | Percentage |
|---|---|---|
| Chibombo | 712 | 11.643 |
| Chifunabuli | 554 | 9.060 |
| Chongwe | 474 | 7.751 |
| Kasama | 335 | 5.478 |
| Shibuyunji | 327 | 5.348 |
| Rufunsa | 308 | 5.037 |
| Samfya | 293 | 4.791 |
| Kafue | 241 | 3.941 |
| Kaoma | 221 | 3.614 |
| Mumbwa | 158 | 2.584 |

There were a total of **78** districts in the dataset. From these 78 districts, **Chibombo** had the largest number of water points standing at **712** which translated into **11.6**%. **Lunga**, **Mpongwe**, **Chadiza**, **Nyimba**, **Luangwa**, **Sioma** and **Shiwang'andu** each had **1** water point translating into **0.016%.**

### 3.3: How many water points were donated by each NGO?

```
84    -- how many water points were supplied by each NGO in each district?
85    SELECT NGO_Source, COUNT(DISTINCT WPDX_Id) AS Number_of_Water_Points,
86    COUNT(DISTINCT District) AS Number_of_Districts FROM zambia_water_points
87    GROUP BY NGO_Source ORDER BY 2 DESC;
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

| NGO_Source | Number_of_Water_Points | Number_of_Districts |
|---|---|---|
| SNV-Zambia | 2089 | 17 |
| Water4 | 963 | 6 |
| World Vision | 859 | 33 |
| Living Water International | 851 | 26 |
| Village Water | 429 | 21 |
| Village Water Zambia | 353 | 14 |
| UNHCR | 221 | 5 |
| District Joint Monitoring Team (DJMT) | 91 | 10 |
| WaterAid UK | 87 | 6 |
| Water and Sanitation for the Urban Poor | 71 | 2 |
| WaterAid | 71 | 10 |
| Rain Water for Africa | 30 | 1 |

**SNV – Zambia** supplied the **2089** number of water points which were distributed in **17** districts. On the bottom end of the list, **Rain Water for Africa** supplied **30** water points in **1** district. Despite having donated only **856** water points, World Vision distributed these water points to **33** districts which was the highest distribution number to districts.

### 3.4: What are the various water sources in the districts?

```
89    -- what is the distribution of water sources among districts?
90 •  SELECT Water_Source, COUNT(DISTINCT WPDX_Id) AS Number_of_Water_Points,
91    ROUND((COUNT(DISTINCT WPDX_Id) / (SELECT COUNT(DISTINCT WPDX_Id) FROM zambia_water_points)) * 100, 2)
92    AS Percentage, COUNT(DISTINCT District) AS Number_of_Districts
93    FROM zambia_water_points GROUP BY Water_Source ORDER BY 2 DESC;
94
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

| Water_Source | Number_of_Water_Points | Percentage | Number_of_Districts |
|---|---|---|---|
| Borehole/Tubewell | 4693 | 76.75 | 77 |
| Piped Water | 925 | 15.13 | 20 |
| Protected Well | 448 | 7.33 | 17 |
| N/A | 24 | 0.39 | 3 |
| Protected Spring | 24 | 0.39 | 10 |
| Undefined Well | 1 | 0.02 | 1 |

**Borehole/Tubewell** was the most common water source amounting to **76.75%** and being used in **77** districts. **Undefined Wells** were the least common water source representing **0.02%** and only used in **1** district.

### 3.5: What is the distribution of water technology among districts?

```
95    -- what is the distribution of water technology among districts?
96 •  SELECT Water_Tech, COUNT(DISTINCT WPDX_Id) AS Number_of_Water_Points,
97    ROUND((COUNT(DISTINCT WPDX_Id) / (SELECT COUNT(DISTINCT WPDX_Id) FROM zambia_water_points)) * 100, 2)
98    AS Percentage, COUNT(DISTINCT District) AS Number_of_Districts
99    FROM zambia_water_points GROUP BY Water_Tech ORDER BY 2 DESC;
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

| Water_Tech | Number_of_Water_Points | Percentage | Number_of_Districts |
|---|---|---|---|
| Hand Pump - India Mark | 2785 | 45.54 | 48 |
| Public Tapstand | 889 | 14.54 | 17 |
| Motorized Pump | 854 | 13.97 | 34 |
| N/A | 682 | 11.15 | 25 |
| Hand Pump | 524 | 8.57 | 26 |
| Hand Pump - Rope | 180 | 2.94 | 17 |
| Rope and Bucket | 76 | 1.24 | 12 |
| Hand Pump - Afridev | 58 | 0.95 | 19 |
| Kiosk | 30 | 0.49 | 3 |
| Motorized Pump - Electric | 19 | 0.31 | 8 |
| Motorized Pump - Solar | 11 | 0.18 | 8 |
| Hand Pump - Tara | 6 | 0.10 | 2 |
| Hand Pump - Blue | 1 | 0.02 | 1 |

The most commonly used water access technology across the country is the **Hand Pump – India Mark** which accounts for **45.54%** and is used in **48** districts. The least used water access technology is the **Hand Pump – Blue**, making up **0.02%** and used in only **1** district.

### 3.6: When was the oldest water point installed?

```
101     -- when was the oldest water point installed?
102 •   SELECT DISTINCT Installation_Year, District, Water_Source, Water_Tech, NGO_Source
103     FROM zambia_water_points ORDER BY 1 ASC LIMIT 10;
104
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content: | Fetch rows:

| Installation_Year | District | Water_Source | Water_Tech | NGO_Source |
|---|---|---|---|---|
| 1902 | Chinsali | Protected Spring | Rope and Bucket | SNV-Zambia |
| 1902 | Shibuyunji | Borehole/Tubewell | Hand Pump - India Mark | SNV-Zambia |
| 1903 | Kasama | Protected Spring | N/A | SNV-Zambia |
| 1904 | Shibuyunji | Borehole/Tubewell | Hand Pump - India Mark | SNV-Zambia |
| 1922 | Kasama | Protected Spring | Rope and Bucket | SNV-Zambia |
| 1941 | Shibuyunji | Protected Well | Hand Pump - Rope | SNV-Zambia |
| 1941 | Shibuyunji | Protected Well | Rope and Bucket | SNV-Zambia |
| 1954 | Lavushimanda | Protected Spring | Rope and Bucket | SNV-Zambia |
| 1956 | Chongwe | Borehole/Tubewell | Hand Pump - India Mark | SNV-Zambia |
| 1956 | Shibuyunji | Protected Well | Hand Pump - Rope | SNV-Zambia |

The oldest water point in the country was installed in **1902** in **Chinsali** and **Shibuyunji** districts. The water access technology was **Rope and Bucket** and **Hand Pump – India Mark** for Chinsali and Shibuyunji respectively. The NGO which supplied the water access technology for the two districts was **SNV – Zambia**.

### 3.7: What is the status of the water points across the country?

```
105     -- What is the status of the water points across the country?
106 •   SELECT DISTINCT Status_of_Water_Point, COUNT(DISTINCT WPDX_Id) AS Number_of_Water_Points,
107     ROUND((COUNT(DISTINCT WPDX_Id) / (SELECT COUNT(DISTINCT WPDX_Id) FROM zambia_water_points)) * 100, 1)
108     AS Percentage, COUNT(DISTINCT District) AS Number_of_Districts
109     FROM zambia_water_points GROUP BY 1 ORDER BY 2 DESC;
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

| Status_of_Water_Point | Number_of_Water_Points | Percentage | Number_of_Districts |
|---|---|---|---|
| Non-Functional | 4380 | 71.6 | 76 |
| Functional | 1735 | 28.4 | 17 |

A total of **4380** water points which represented **71.6%** were non-functional. Only **1735** which translated into **28.4%** were functional. In relation to the districts, **76** districts had non – functional waterpoints and a meagre **17** districts had functional ones.

### 3.8: How many water points are in urban areas?

```
111     -- How many water points are in urban areas?
112 •   SELECT DISTINCT Is_Urban AS Urban_Area, COUNT(DISTINCT WPDX_Id) AS Number_of_Water_Points,
113     ROUND((COUNT(DISTINCT WPDX_Id) / (SELECT COUNT(DISTINCT WPDX_Id) FROM zambia_water_points)) * 100, 1)
114     AS Percentage, COUNT(DISTINCT District) AS Disrict
115     FROM zambia_water_points GROUP BY 1 ORDER BY 2 DESC;
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

| Urban_Area | Number_of_Water_Points | Percentage | Disrict |
|---|---|---|---|
| FALSE | 5953 | 97.4 | 77 |
| TRUE | 162 | 2.6 | 17 |

There were **5953** water points translating into **97.4%** were in **rural** areas, whereas **162** water points making up **2.6%** were in **urban** areas.

### 3.9: Which water point capacity match the local population?

```
117     -- Did any of the water points' capacity match the assigned population?
118 •   SELECT DISTINCT District, Water_Source, Water_Tech, Assigned_Population, Local_Population
119     FROM zambia_water_points
120     WHERE Assigned_Population >= Local_Population ORDER BY 4 DESC;
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

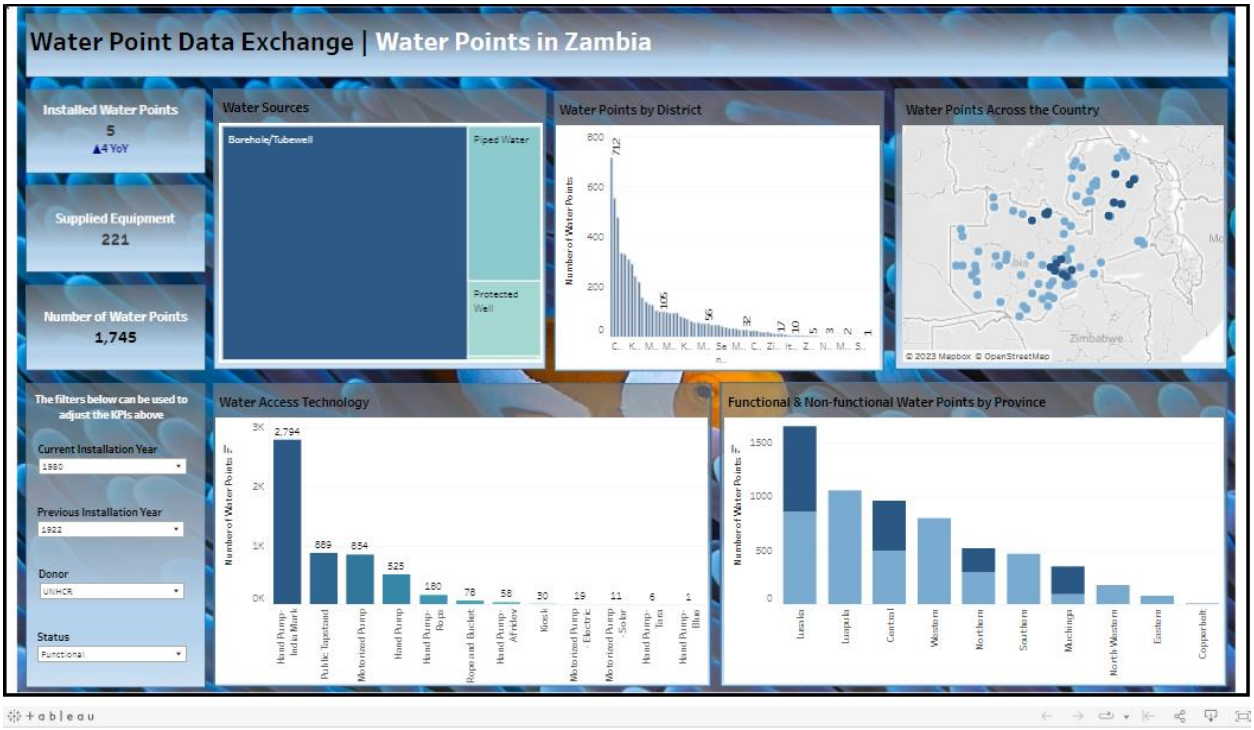
| District | Water_Source | Water_Tech | Assigned_Population | Local_Population |
|---|---|---|---|---|
| Lusaka | Borehole/Tubewell | N/A | 38511 | 38511 |
| Lusaka | Borehole/Tubewell | N/A | 19241 | 19241 |
| Lusaka | Borehole/Tubewell | Hand Pump - India Mark | 18962 | 18962 |
| Ndola | Borehole/Tubewell | Hand Pump - India Mark | 18850 | 18850 |
| Kitwe | Borehole/Tubewell | Hand Pump - India Mark | 17365 | 17365 |
| Lusaka | Borehole/Tubewell | N/A | 16072 | 16072 |
| Mpika | Borehole/Tubewell | Hand Pump - India Mark | 11349 | 11349 |
| Mongu | Borehole/Tubewell | Hand Pump | 10854 | 10854 |
| Ndola | Borehole/Tubewell | Hand Pump - India Mark | 9721 | 9721 |

160 20:26:56 SELECT DISTINCT District, Water_Source, Water_Tech, Assigned_Population, Local_P... 1258 row(s) returned

A total of **1258** water points matched the needs of the communities they were set up in.

# Part 4: Visualizations

Visualizations for this project were designed and developed using [Tableau]



## Recommendations for stakeholders

1. **Repair or replace non – functional water points:** A greater number of water points are non – functional, only **1735** out of a total of **6127** water points are functional.
2. **Install more water points:** Most of the water points do not meet the needs of the local population. Out of **6127** water points, only **1258** water points match the needs of the communities where they have been set up.
3. **Replace old water installations:** Districts such as **Chinsali** and **Shibuyinji** have water points which were installed as far back as **1902**. Such water points need to be replaced or undergo a significant overhaul.
4. **Install more water points on the Copperbelt:** The Copperbelt province only has 17 water points, in comparison, Lusaka province has 1643 water points.

## Tools and Technology Used

Throughout this project, I utilized

1. Microsoft Excel for data organization, preparation and cleaning.
2. MySQL Workbench for data import and analysis.
3. Tableau for visualizations.

thereby enhancing my skills and knowledge in these essential data analytical tools.

Dataset link
Tableau dashboard link
Github link
LinkedIn link