

RELATIONAL DATABASE DATA ANALYSIS PROJECT

Jimoh, Sekinat Eniola

March 15, 2025

Tables of Content

1. Introduction
2. Database and Table Overview
3. SQL Queries
4. Pictures
5. Conclusion
6. Link to the Query

INTRODUCTION

The Project aims to analyze water supply and sanitation data across various communities and countries in Africa. The objective is to assess water availability, functionality of water points, sanitation maintenance costs, and the impact on public health. This report summarizes key insights derived from the dataset through SQL queries.

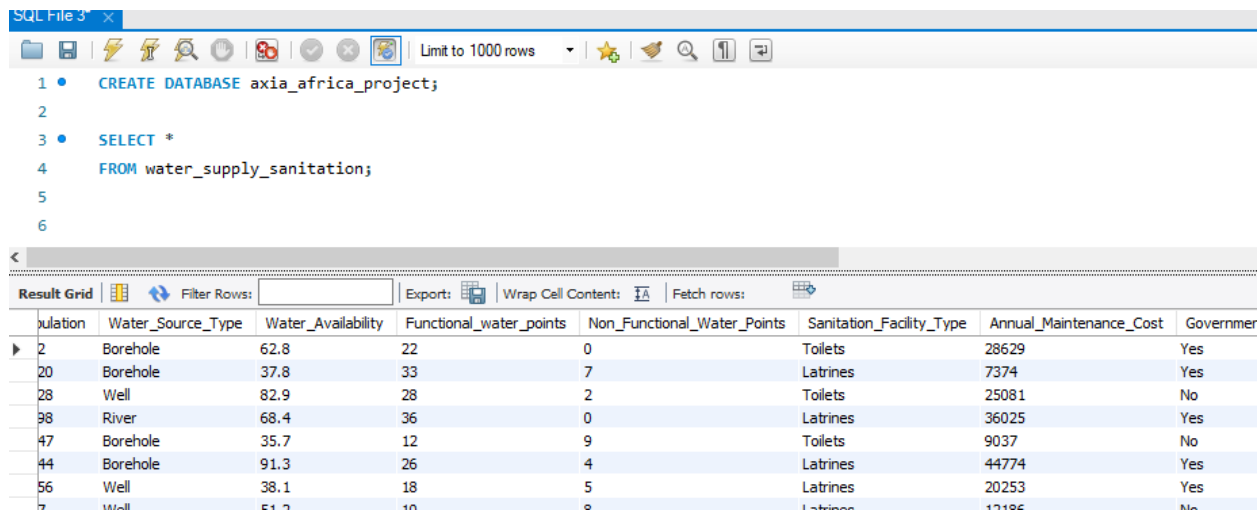
DATABASE AND TABLE OVERVIEW

The data used in this analysis is stored in the **axia_africa_project database**, which was created and structured to house water supply and sanitation records across various regions.

The primary table, **water_supply_sanitation**, was imported into the database using the *Table Data Import Wizard*.

To standardize and improve data accessibility, several columns were renamed as follows:

- Water Availability (liters per capita per day) changed to `water_availability`
- Number of Functional Water Points changed to `Functional_water_points`
- Number of Non-Functional Water Points changed to `Non_Functional_Water_Points`
- Sanitation Facility Type changed to `Sanitation_Facility_Type`
- Annual Maintenance Cost (USD) changed to `Annual_Maintenance_Cost`
- Government Support changed to `Government_Support`
- NGO Support changed to `NGO_Support`
- Average Distance to Water Source (km) changed to `Avg_Distance_to_Water_Source`
- Waterborne Diseases Incidence Rate (%) changed to `Waterborne_Disease_Incidence_Rate`
- Community Satisfaction Rate (%) changed to `Community_Satisfaction_Rate`



The screenshot displays a MySQL SQL File editor window titled "SQL File 3". The editor contains the following SQL code:

```
1 • CREATE DATABASE axia_africa_project;
2
3 • SELECT *
4 FROM water_supply_sanitation;
5
6
```

Below the SQL editor, a "Result Grid" is shown, displaying the data imported from the `water_supply_sanitation` table. The grid has 8 columns: `Location`, `Water_Source_Type`, `Water_Availability`, `Functional_water_points`, `Non_Functional_Water_Points`, `Sanitation_Facility_Type`, `Annual_Maintenance_Cost`, and `Govermmer` (likely a typo for `Government_Support`). The data is as follows:

	Location	Water_Source_Type	Water_Availability	Functional_water_points	Non_Functional_Water_Points	Sanitation_Facility_Type	Annual_Maintenance_Cost	Govermmer
2	Borehole	Borehole	62.8	22	0	Toilets	28629	Yes
20	Borehole	Borehole	37.8	33	7	Latrines	7374	Yes
28	Well	Well	82.9	28	2	Toilets	25081	No
38	River	River	68.4	36	0	Latrines	36025	Yes
47	Borehole	Borehole	35.7	12	9	Toilets	9037	No
44	Borehole	Borehole	91.3	26	4	Latrines	44774	Yes
56	Well	Well	38.1	18	5	Latrines	20253	Yes
7	Well	Well	51.7	10	0	Latrines	12186	No

Dataset imported into Mysql

KEY FINDINGS

1. Write an SQL query to calculate the average water availability (liters per capita per day) for each country.

SQL Query

```
3 -- the average water availability (liters per capita per day) for each country
3 • SELECT country, AVG(water_availability) AS avg_water_availability
3 FROM water_supply_sanitation
1 GROUP BY country
2 ORDER BY avg_water_availability DESC;
3
```

Result

Result Grid			Filter Rows:	Export:
	country	avg_water_availability		
▶	Nigeria	58.60197044334978		
	Senegal	57.89		
	Rwanda	56.66857142857141		
	Uganda	56.138942307692304		
	Tanzania	55.205357142857096		
	Ghana	55.180232558139515		
	Ethiopia	54.987500000000004		
	Kenya	54.746086956521715		
	Malawi	54.418269230769205		
	Zambia	53.61970443349754		

The above query and result shows insights on the average water availability (liters per capita per day) for each country, it is sorted in descending order. So as to know which country has the highest and which has the lowest. Nigeria has the highest at 58.60 liters per capita per day and Zambia being the lowest with 53.62 liters per capita per day.

- Write an SQL query to retrieve details of communities where at least one water point is non-functional.

SQL Query

```
11 -- retrieving the details of communities where at least one water point is non-functional.
12 • SELECT *
13 FROM water_supply_sanitation
14 WHERE Non_Functional_Water_Points >0;
15
16
```

Result

Region	Country	Community_Name	Population	Water_Source_Type	Water_Availability	Functional_water_points	Non_Functional_Water_Points	Sanitation_Facility_Type	Annu
Northern Malawi	Malawi	Northern Community 2	14820	Borehole	37.8	33	7	Latrines	7374
Dar es Salaam	Tanzania	Dar Community 3	14128	Well	82.9	28	2	Toilets	25081
Eastern Nigeria	Nigeria	Eastern Community 5	14147	Borehole	35.7	12	9	Toilets	9037
Northern Malawi	Malawi	Northern Community 6	14944	Borehole	91.3	26	4	Latrines	44774
Eastern Nigeria	Nigeria	Eastern Community 7	14456	Well	38.1	18	5	Latrines	20253
Northern Malawi	Malawi	Northern Community 8	7467	Well	51.2	10	8	Latrines	12186
Northern Malawi	Malawi	Northern Community 9	4851	Borehole	49.7	19	2	Toilets	4546
Eastern Rwanda	Rwanda	Eastern Community 10	12970	Borehole	74.1	24	3	Latrines	40207
Eastern Ethiopia	Ethiopia	Eastern Community 11	14945	Well	94.3	14	3	Toilets	24134
Eastern Nigeria	Nigeria	Eastern Community 12	9757	Well	16.3	2	6	Latrines	23275
Eastern Rwanda	Rwanda	Eastern Community 14	3677	Borehole	65.3	25	9	Latrines	31746
Western Kenya	Kenya	Western Community 16	12950	River	33	31	4	Latrines	1341

In this query, we retrieved the information where at least 1 water point is non-functional, from the result we can see communities like Northern community 2 which have 7 non-functional water points, Dar community3 which has 2, Eastern community 5 which has 9 and so on.

3. Please retrieve the information for the top five communities with the highest annual sanitation maintenance costs

SQL Query

```
16  -- top 5 communities with the highest annual sanitation maintenance costs.
17  •  SELECT *
18      FROM water_supply_sanitation
19      ORDER BY Annual_Maintenance_Cost DESC
20      LIMIT 5;
```

Result

Region	Country	Community_Name	Population	Water_Source_Type	Water_Availability	Functional_water_points	Non_Functional_Water_Points	Sanitation_Facility_Type	Annual_Maintenance_Cost
Dar es Salaam	Tanzania	Dar Community 329	982	River	31.1	20	9	Toilets	49959
Eastern Nigeria	Nigeria	Eastern Community 1658	3378	Well	86.9	46	8	Toilets	49957
Eastern Nigeria	Nigeria	Eastern Community 1021	3871	River	38.2	48	9	Toilets	49949
Western Uganda	Uganda	Western Community 1259	11088	River	10.9	32	1	Toilets	49904
Northern Ghana	Ghana	Northern Community 638	5675	Borehole	52	28	0	Latrines	49894

```
19  -- top 5 communities with the highest annual sanitation maintenance costs.
20  •  SELECT region, country, community_name, Annual_Maintenance_Cost
21      FROM water_supply_sanitation
22      ORDER BY Annual_Maintenance_Cost DESC
23      LIMIT 5;
```

region	country	community_name	Annual_Maintenance_Cost
Dar es Salaam	Tanzania	Dar Community 329	49959
Eastern Nigeria	Nigeria	Eastern Community 1658	49957
Eastern Nigeria	Nigeria	Eastern Community 1021	49949
Western Uganda	Uganda	Western Community 1259	49904
Northern Ghana	Ghana	Northern Community 638	49894

The query shows insights on the top 5 communities with the highest sanitation maintenance costs. The top 5 communities are shown in the above picture with Daar community 329 in Tanzania having the highest maintenance cost of \$49,959 and Northern community 638 in Ghana \$49,894.

4. Calculate the total number of functional and non-functional water points per country

SQL Query

For total functional water points

```
23
24  -- total number of functional and non-functional water points per country
25 •  SELECT country,
26      SUM(functional_water_points) AS total_functional_points
27  FROM water_supply_sanitation
28  GROUP BY country
29  ORDER BY total_functional_points DESC;
30
31
```

Result



country	total_functional_points
Kenya	6049
Rwanda	5491
Ethiopia	5296
Uganda	5287
Nigeria	5271
Zambia	5253
Malawi	5172
Senegal	5039
Ghana	4371
Tanzania	4354

SQL Query

For non-functional points

```
30
31      -- total number of non-functional water points per country
32 •    SELECT country,
33          SUM(Non_Functional_Water_Points) AS total_non_functional_points
34      FROM water_supply_sanitation
35      GROUP BY country
36      ORDER BY total_non_functional_points DESC;
37
38
```

Results

<		
Result Grid		
Filter Rows: <input type="text"/>		
Export:  Wrap Cell Content: 		
	country	total_non_functional_points
▶	Kenya	1116
	Ethiopia	1088
	Malawi	1029
	Zambia	1020
	Nigeria	985
	Rwanda	966
	Uganda	959
	Senegal	946
	Ghana	886
	Tanzania	879

The queries and results above show the total of both functional and non-functional water points, showcasing which country had the highest water points and which had the lowest.

5. Identify communities with a high incidence of waterborne diseases (>20%)

SQL Query

```
37
38 -- communities with high incidence of water borne diseases
39 • SELECT *
40 FROM water_supply_sanitation
41 WHERE Waterborne_Diseases_Incidence_Rate >20;
42
43
44
```

Results

```
37
38 -- communities with high incidence of water borne diseases
39 • SELECT *
```

Region	Country	Community_Name	Population	Water_Source_Type	Water_Availability	Functional_water_points	Non_Functional_Water_Points	Sanitation_Facility_Type	Annual_Maintenan
Southern Zambia	Zambia	Southern Community 1	1142	Borehole	62.8	22	0	Toilets	28629
Eastern Nigeria	Nigeria	Eastern Community 4	13398	River	68.4	36	0	Latrines	36025
Northern Malawi	Malawi	Northern Community 6	14944	Borehole	91.3	26	4	Latrines	44774
Eastern Ethiopia	Ethiopia	Eastern Community 11	14945	Well	94.3	14	3	Toilets	24134
Eastern Nigeria	Nigeria	Eastern Community 12	9757	Well	16.3	2	6	Latrines	23275
Southern Zambia	Zambia	Southern Community 13	4514	Well	50	11	0	Toilets	41917
Western Uganda	Uganda	Western Community 18	1712	Well	14	38	10	Latrines	42220
Western Kenya	Kenya	Western Community 21	13322	Borehole	22.9	6	1	Toilets	42469
Eastern Nigeria	Nigeria	Eastern Community 23	12872	Borehole	12.7	37	2	Latrines	22265
Eastern Ethiopia	Ethiopia	Eastern Community 24	10756	Borehole	31.5	41	9	Toilets	40984
Dar es Salaam	Tanzania	Dar Community 29	9266	Well	91.3	2	7	Latrines	27671
Eastern Nigeria	Nigeria	Eastern Community 30	2716	River	62.2	6	4	Toilets	42125
Eastern Nigeria	Nigeria	Eastern Community 31	9535	Well	57.2	41	7	Latrines	43918
Central Senegal	Senegal	Central Community 35	672	River	64.7	44	3	Latrines	36637
Central Senegal	Senegal	Central Community 36	2672	Well	44	9	2	Toilets	4170
Southern Zambia	Zambia	Southern Community 44	9747	River	20.3	4	10	Latrines	5758
Eastern Ethiopia	Ethiopia	Eastern Community 45	1589	Borehole	87.6	39	2	Latrines	10003
Western Kenya	Kenya	Western Community 48	11660	Well	00.2	47	2	Toilets	12470

This shows the communities that waterborne diseases incidence rate is greater than 20%.

A clearer view of the diagram is shown below, selecting fewer columns

```
-- communities with high incidence of water borne diseases
• SELECT region, country, Community_Name, Waterborne_Diseases_Incidence_Rate
FROM water_supply_sanitation
WHERE Waterborne_Diseases_Incidence_Rate >20;
```

region	country	Community_Name	Waterborne_Diseases_Incidence_Rate
Northern Malawi	Malawi	Northern Community 96	26.4
Northern Malawi	Malawi	Northern Community 99	26.9
Western Uganda	Uganda	Western Community 102	23.1
Dar es Salaam	Tanzania	Dar Community 107	27
Northern Malawi	Malawi	Northern Community 111	27.4
Northern Malawi	Malawi	Northern Community 112	27
Dar es Salaam	Tanzania	Dar Community 119	27
Central Senegal	Senegal	Central Community 120	22.3
Eastern Rwanda	Rwanda	Eastern Community 127	27.5
Central Senegal	Senegal	Central Community 128	26
Northern Malawi	Malawi	Northern Community 131	21.1
Dar es Salaam	Tanzania	Dar Community 137	24.5
Northern Malawi	Malawi	Northern Community 139	27.1

6. Find the average distance to the water source per region

SQL Query

```
42
43 -- average distance to the water source per region
44 • SELECT region,
45     AVG(avg_distance_to_water_source) AS avg_distance
46     FROM water_supply_sanitation
47     GROUP BY region
48     ORDER BY avg_distance DESC;
49
```

Result

49

<

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

region	avg_distance
Western Kenya	5.539565217391304
Eastern Nigeria	5.477832512315274
Northern Malawi	5.414423076923076
Central Senegal	5.255789473684208
Southern Zambia	5.221674876847291
Eastern Rwanda	5.208571428571427
Northern Ghana	5.187209302325583
Dar es Salaam	5.15059523809524
Western Uganda	5.0375
Eastern Ethiopia	5.00480769230769

Result 22 x

7. List the communities that receive both government and NGO support

SQL Query

```
--  
50 -- communities that receive both government and NGO support  
51 • SELECT Community_Name  
52 FROM water_supply_sanitation  
53 WHERE Government_Support = 'yes'  
54 AND NGO_Support='yes';  
55  
56
```

Result

49

50 -- communities that receive both government and NGO support

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

Community_Name
Northern Community 2
Northern Community 9
Eastern Community 10
Western Community 19
Eastern Community 23
Eastern Community 24
Dar Community 29
Eastern Community 30
Eastern Community 31
Southern Community 33
Southern Community 38
Southern Community 39
Northern Community 41
Eastern Community 45
Western Community 55
Northern Community 59
Southern Community 69

water_supply_sanitation26 x

Output

The result of this query shows the communities that has both NGO and government support

8. Identify the community with the highest population per country?

SQL Query

```
56  -- communities with the highest population per county
57  •  SELECT *
58  FROM water_supply_sanitation AS WSS
59  WHERE Population=(
60      SELECT MAX(Population)
61      FROM water_supply_sanitation
62      WHERE Country= WSS.Country
63  );
64
```

Result

59 WHERE Population=(

Region	Country	Community_Name	Population	Water_Source_Type	Water_Availability	Functional_water_points	Non_Functional_Water_Points	Sanitation_Facilit
Northern Malawi	Malawi	Northern Community 6	14944	Borehole	91.3	26	4	Latrines
Western Uganda	Uganda	Western Community 121	14948	Well	55.9	3	8	Latrines
Southern Zambia	Zambia	Southern Community 748	14952	Borehole	95.3	11	8	Latrines
Eastern Nigeria	Nigeria	Eastern Community 932	14997	Borehole	29.9	36	6	Toilets
Western Kenya	Kenya	Western Community 1036	14976	River	40	49	3	Latrines
Dar es Salaam	Tanzania	Dar Community 1199	14888	Well	27.3	22	6	Latrines
Northern Ghana	Ghana	Northern Community 1393	14992	River	95	46	0	Latrines
Central Senegal	Senegal	Central Community 1395	14886	River	98.1	43	3	Latrines
Eastern Rwanda	Rwanda	Eastern Community 1440	14890	River	80.5	6	4	Latrines
Eastern Ethiopia	Ethiopia	Eastern Community 1877	14993	Borehole	20.2	5	1	Toilets



water_supply_sanitation27 x

OR

```

65
66      -- communities with the highest population per county
67 •   SELECT Community_Name, Country, Population
68      FROM water_supply_sanitation AS WSS
69      WHERE Population=(
70          SELECT MAX(Population)
71          FROM water_supply_sanitation
72          WHERE Country= WSS.Country
73      )

```

Result Grid			
Filter Rows: <input type="text"/>			
Export:  Wrap Cell Content: 			
	Community_Name	Country	Population
▶	Northern Community 6	Malawi	14944
	Western Community 121	Uganda	14948
	Southern Community 748	Zambia	14952
	Eastern Community 932	Nigeria	14997
	Western Community 1036	Kenya	14976
	Dar Community 1199	Tanzania	14888
	Northern Community 1393	Ghana	14992
	Central Community 1395	Senegal	14886
	Eastern Community 1440	Rwanda	14890
	Eastern Community 1877	Ethiopia	14993

CONCLUSION

The analysis of the water_supply_sanitation dataset provides critical insights into the status of water and sanitation infrastructure across different regions. By answering key business questions, stakeholders can make informed decisions to improve water accessibility, enhance sanitation facilities, and mitigate waterborne diseases. Continued monitoring and data-driven interventions will be crucial in achieving sustainable water and sanitation solutions in Africa.

Link to The Query

https://drive.google.com/file/d/1b-u906Iy_4gW-qyRIBvsG1d6KLnSbaU3/view?usp=sharing