Relational Database Project

(Water Access and Sanitation In Africa).

My Axia Africa Project Submission.

March 13,2025

Oke Tolulope.

Outline

1. Introduction.
2. Data Cleaning.
3. Data Import.
4. Key Business Questions.

As part of the project, steps like data cleaning and data import was carried out on a structured dataset with 2000 rows and 15 columns using MS Excel and MySQL to get data driven insights from the datasets.

1. INTRODUCTION

This dataset contains information on water access and sanitation in Africa from various country. It contains 2000 rows and 15 columns which are;

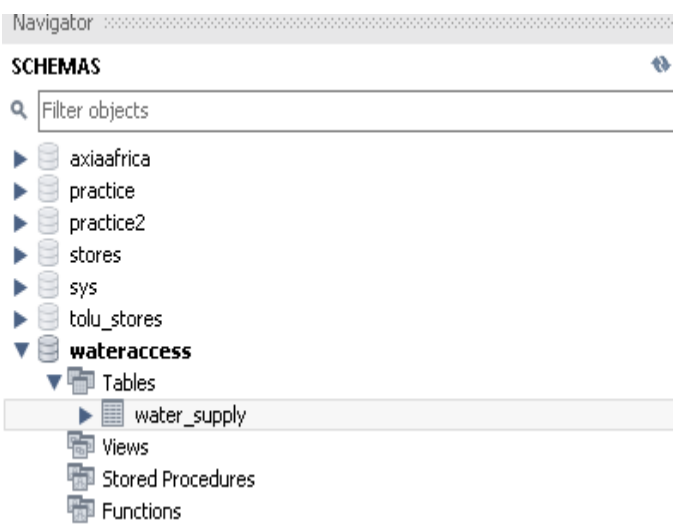
- Region.
- Country.
- Community name.
- Population.
- Water source type.
- Water availability(Liter per capita per day).
- Number of functional water points.
- Number of non functional water points.
- Sanitation facility type.
- Annual maintenance cost(USD).
- Government support.
- NGO support.
- Average distance to water source(KM).
- Waterborne disease incidence rate(%).
- Community satisfaction rate(%).

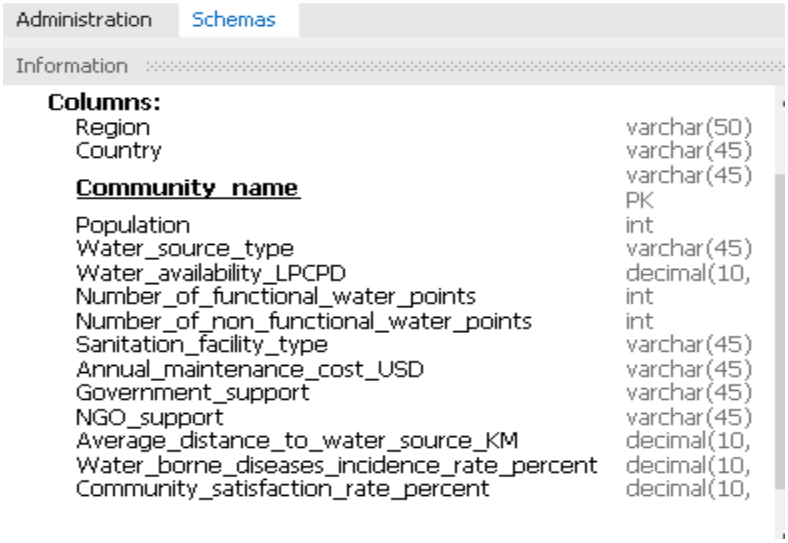
2. DATA CLEANING

I used MS Excel to check through the dataset, formatting each columns to its format and checked for nulls and duplicates and other inconsistencies and I discovered that there were none.

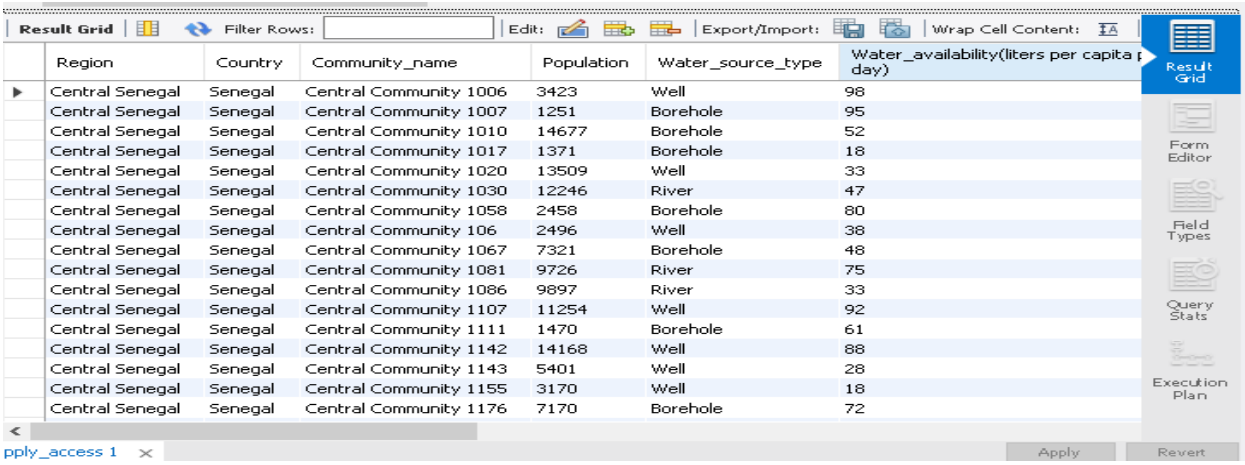
3.DATA IMPORTATION

- A schema named wateraccess was created and a table named water_supply was created and then the data was imported into the table.





A quick preview of the imported dataset.



3. KEY BUSINESS QUESTIONS

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

```
2
3   -- 1 Write an SQL query to calculate the average water availability (liters per capita per day) for each cou
4   SELECT Country,AVG(Water_availability_LPCPD) AS Avg_water_availability
5   FROM water_supply
6   GROUP BY Country;
```

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

Country	Avg_water_availability
Senegal	57.9368
Tanzania	55.2798
Rwanda	56.7190
Ethiopia	55.0240
Nigeria	58.6650
Malawi	54.4663
Ghana	55.2442
Zambia	53.6798
Kenya	54.8000
Uganda	56.1587

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

```
8   -- 2 Write an SQL query to retrieve details of communities where at least one water point is non-functional
9   SELECT * FROM water_supply
10  WHERE Number_of_non_functional_water_points>0;
11
```

Result Grid | Filter Rows: | Edit: | Export/Import: | Wrap Cell Content:

Region	Country	Community_name	Population	Water_source_type	Water_availability_LPCPD	Number_of_functional_water_
Central Senegal	Senegal	Central Community 1006	3423	Well	98	10
Central Senegal	Senegal	Central Community 1007	1251	Borehole	95	43
Central Senegal	Senegal	Central Community 1010	14677	Borehole	52	33
Central Senegal	Senegal	Central Community 1017	1371	Borehole	18	1
Central Senegal	Senegal	Central Community 1030	12246	River	47	28
Central Senegal	Senegal	Central Community 1058	2458	Borehole	80	10
Central Senegal	Senegal	Central Community 106	2496	Well	38	22
Central Senegal	Senegal	Central Community 1067	7321	Borehole	48	16
Central Senegal	Senegal	Central Community 1081	9726	River	75	39
Central Senegal	Senegal	Central Community 1086	9897	River	33	19
Central Senegal	Senegal	Central Community 1107	11254	Well	92	24
Central Senegal	Senegal	Central Community 1142	14168	Well	88	46
Central Senegal	Senegal	Central Community 1143	5401	Well	28	28
Central Senegal	Senegal	Central Community 1155	3170	Well	18	33
Central Senegal	Senegal	Central Community 1176	7170	Borehole	72	35
Central Senegal	Senegal	Central Community 1179	11753	River	47	19
Central Senegal	Senegal	Central Community 1183	4861	Well	95	28
Central Senegal	Senegal	Central Community 1185	11188	Borehole	78	48

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

```
12  -- 3 Please retrieve the information for the top five communities with the highest annual sanitation maintenan
13  SELECT * FROM water_supply
14  ORDER BY Annual_maintenance_cost_USD DESC
15  LIMIT 5;
```

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

Region	Country	Community_name	Population	Water_source_type	Water_availability_LPCPD	Number_of_functional_water_
Southern Zambia	Zambia	Southern Community 1979	11440	Well	77	8
Eastern Ethiopia	Ethiopia	Eastern Community 903	3274	Well	74	7
Dar es Salaam	Tanzania	Dar Community 412	6447	River	44	26
Eastern Rwanda	Rwanda	Eastern Community 682	2010	Well	43	23
Western Uganda	Uganda	Western Community 1784	1352	Borehole	76	45
NULL	NULL	NULL	NULL	NULL	NULL	NULL

Result Grid | Form Editor

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

```
17  -- 4 Calculate the total number of functional and non-functional water points per country
18  SELECT Country, SUM(Number_of_functional_water_points) AS Total_functional_water_points,
19  SUM(Number_of_non_functional_water_points) AS Total_non_functional_water_points
20  FROM water_supply
21  GROUP BY Country;
```

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

Country	Total_functional_water_points	Total_non_functional_water_points
Senegal	5039	946
Tanzania	4354	879
Rwanda	5491	966
Ethiopia	5296	1088
Nigeria	5271	985
Malawi	5172	1029
Ghana	4371	886
Zambia	5253	1020
Kenya	6049	1116
Uganda	5287	959

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

```
23  -- 5 Identify communities with a high incidence of waterborne diseases (>20%)
24  SELECT Community_name,Water_borne_diseases_incidence_rate_percent
25  FROM water_supply
26  WHERE Water_borne_diseases_incidence_rate_percent>20;
```

Result Grid | Filter Rows: | Edit: | Export/Import: | Wrap Cell Content:

Community_name	Water_borne_diseases_incidence_rate_percent
Central Community 1010	28
Central Community 1058	27
Central Community 1081	25
Central Community 1183	28
Central Community 1185	25
Central Community 120	22
Central Community 1226	22
Central Community 1228	30
Central Community 1256	24
Central Community 128	26
Central Community 1298	26
Central Community 1331	22
Central Community 1348	21
Central Community 1398	26
Central Community 1414	26
Central Community 1433	22
Central Community 1482	29
Central Community 1492	29
Central Community 1539	29

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

```
28      -- 6 Find the average distance to the water source per region
29 •    SELECT Region,AVG(Average_distance_to_water_source_KM) AS AVG_distance
30      FROM water_supply
31      GROUP BY Region;
```

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

Region	AVG_distance
Central Senegal	5.3105
Dar es Salaam	5.2262
Eastern Rwanda	5.3143
Eastern Ethiopia	5.0769
Eastern Nigeria	5.5320
Northern Malawi	5.4808
Northern Ghana	5.2442
Southern Zambia	5.2562
Western Kenya	5.6043
Western Uganda	5.1202

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

```
33      -- 7 List the communities that receive both government and NGO support-
34 •    SELECT Community_name,Government_support,NGO_support FROM water_supply
35      WHERE Government_support="Yes" AND NGO_support="Yes";
```

Result Grid | Filter Rows: | Edit: | Export/Import: | Wrap Cell Content:

Community_name	Government_support	NGO_support
Central Community 1017	Yes	Yes
Central Community 1020	Yes	Yes
Central Community 1111	Yes	Yes
Central Community 1143	Yes	Yes
Central Community 1176	Yes	Yes
Central Community 1183	Yes	Yes
Central Community 1226	Yes	Yes
Central Community 1256	Yes	Yes
Central Community 1288	Yes	Yes
Central Community 1348	Yes	Yes
Central Community 1433	Yes	Yes
Central Community 1451	Yes	Yes
Central Community 1503	Yes	Yes
Central Community 1526	Yes	Yes
Central Community 1530	Yes	Yes
Central Community 1628	Yes	Yes
Central Community 1705	Yes	Yes
Central Community 1736	Yes	Yes
Central Community 1818	Yes	Yes

8. Identify the community with the highest population per country

```
38      -- 8 Identify the community with the highest population per country
39 •    SELECT Community_name,Country,Population
40      FROM water_supply
41      ORDER BY Population DESC
42      LIMIT 1;
```

Result Grid | Filter Rows: | Edit: | Export/Import: | Wrap Cell Content:

Community_name	Country	Population
Eastern Community 932	Nigeria	14997
NULL	NULL	NULL

The link to the query :
https://drive.google.com/file/d/14ccLCsKq4YRkw250yfwlyisn41ymMArm/view?usp=drivesdk