MySQL_Project_By; Buyanzi Charity Purity

Project Name: Water Crisis Insights for Maji Ndogo

Objective: Analyze a dataset of 60,000 records to extract meaningful insights regarding the water crisis in Maji Ndogo across five key focus areas.

Introduction

Maji Ndogo is facing a severe water crisis, affecting livelihoods, health, and sustainability. To understand the scope of the problem, an extensive survey gathered **60,000 records** detailing water access, quality, and usage. The goal is to analyze this data, uncover key insights, and identify actionable solutions that will drive effective water management and long-term relief for the community.

---

1. Understanding the Dataset

Before diving into the analysis, **explore the database structure** and understand the key tables.

Tasks:

✅ Load and explore the data

```
– - Query sample records from each table:
SELECT * FROM md_water_services.location
LIMIT 5
;

SELECT *
type_of_water_source
FROM md_water_services.water_source;
```

✅ Understand relationships between tables

```
SELECT *

FROM md_water_services.water_source

WHERE source_id

IN ("AkKi00881224", "SoRu37635224", "SoRu36096224", "AkRu05234224", "HaZa21742224",
"AkLu02211224");
```

---

2. Key Areas of Analysis & Insights

1. Water Source Availability & Functionality

Goal: Identify different water source types and assess their operational status.

Task:

✅ **List unique water sources**

```
SELECT DISTINCT
type_of_water_source
FROM md_water_services.water_source;
```

---

2. Water Accessibility

Goal Find out which sources have the most alarming queue time

✅ **Task** Analyze extremely long queue times

```
SELECT *
FROM md_water_services.visits
WHERE time_in_queue > 500;
```

---

3. Water Quality & Contamination Levels

Goal: Assess water quality and identify contamination risks.

Tasks:

✅ **Evaluate water quality scores where the quality was good that is quality score is 10 but the source was visited a second time. In essence the field surveyors did not need to visit a source the second time when the water source was already excellent.**

NOTE: This should be highlighted in the report since we might have corrupt employees or mismanaged resources.

```
SELECT *
FROM md_water_services.water_quality
WHERE subjective_quality_score = 10 AND visit_count =2
;
```

✅ **Identify water sources with well pollution issues**

Anything above 0.01 should be considered polluted so we check if the results is Clean but the biological column is > 0.01.

```
SELECT *

FROM md_water_services.well_pollution

WHERE ( results  = "Clean" AND biological > 0.01 );
```

Some data entry professionals wrote Clean: then Contaminated when we water can either be clean or contaminated, not both. Lets correct this

```
SET SQL_SAFE_UPDATES = 0;
UPDATE md_water_services.well_pollution
SET description = 'Bacteria: E. coli'
WHERE description = 'Clean Bacteria: E. coli';
```

Update these changes so that the well pollution table can be analysed correctly

```
CREATE TABLE md_water_services.well_pollution_copy
AS (
      SELECT *
      FROM md_water_services.well_pollution
    );

UPDATE
well_pollution_copy
SET
description = 'Bacteria :E. coli'
WHERE description = 'Clean Bacteria: E. coli';

UPDATE
well_pollution_copy
SET
description = 'Bacteria: Giardia Lamblia'
WHERE description = 'Clean Bacteria: Giardia Lamblia';

UPDATE
well_pollution_copy
SET
results = 'Contaminated : Biological'
WHERE biological > 0.01 AND results = 'Clean';
```

---

5. Portfolio Documentation

📌 **Find SQL scripts in this GitHub repository**
