# **Water Access and Sanitation Conditions in Kakuma: Student Capstone Brief**

## **Project Title: Water Access and Sanitation Conditions in Kakuma Refugee Camp**

## **1. Project Overview**

This capstone project focuses on analyzing water access and sanitation conditions within the Kakuma Refugee Camp. Students will apply data analytics techniques to assess the availability, accessibility, and sufficiency of water resources, as well as the state of sanitation facilities. The goal is to derive actionable insights that can inform interventions to improve public health and living conditions for refugees. The project will involve data cleaning, transformation, and visualization, culminating in a comprehensive Power BI dashboard and a professional presentation of findings.

## **2. Dataset**

The primary dataset for this project is kakuma\_water\_sanitation.csv. This dataset contains information related to water sources, consumption, and sanitation status, with the following key fields:

* **Zone**: The geographical zone within Kakuma Refugee Camp.
* **Household\_ID**: Unique identifier for each household.
* **Household\_Size**: The number of individuals in the household.
* **Water\_Source**: The type of water source used by the household (e.g., borehole, tap stand, river).
* **Distance\_km**: The distance in kilometers from the household to the primary water source.
* **Liters\_Per\_Person\_Per\_Day**: The average liters of water collected per person per day by the household.
* **Collection\_Times\_Daily**: The number of times per day water is collected by the household.
* **Sanitation\_Status**: The overall sanitation condition or status for the household/zone.
* **Sanitation\_Facility\_Type**: The type of sanitation facility used by the household (e.g., communal latrine, household latrine, open defecation).

## **3. Project Objectives**

The student will be required to achieve the following objectives:

### **3.1. Data Preparation**

* **Data Cleaning:**
  + Identify and handle missing values appropriately for critical fields such as water source, distance, liters per person per day, household size, and sanitation status.
  + Ensure consistent data types and formats across all relevant columns (e.g., numerical values for distance and liters, categorical for water source and sanitation types).
  + Address any inconsistencies or errors in textual data (e.g., zone names, water source descriptions, sanitation facility types).
* **Data Transformation:**
  + Create "distance bins" (e.g., 0-0.5km, 0.5-1km, >1km) to categorize and analyze water access based on proximity.
  + Calculate or verify "per capita water access" based on Liters\_Per\_Person\_Per\_Day and Household\_Size.
  + Group data by Zone to enable zone-specific analysis of water access and sanitation.
  + Implement any other necessary transformations, such as categorizing water source types or standardizing sanitation indicators.

### **3.2. Data Analysis & Insights**

The analysis should address the following key questions:

1. **✅ PREREQUISITES**
2. **Ensure you have:**
3. **✅ Cleaned file (kakuma\_cleaned.csv) — from previous Python step**
4. **✅ Power BI Desktop installed**
5. **📊 DASHBOARD OVERVIEW**
6. **🔷 Page 1: Water Access Overview**
7. **🔷 Page 2: Water Sufficiency & Sanitation Insights**
8. **🔹 PAGE 1: WATER ACCESS OVERVIEW**
9. **🔸 Layout Suggestions:**

| 1. **Visual Type** | 1. **Description** | 1. **Fields** |
| --- | --- | --- |
| 1. **Card** | 1. **Avg Distance (km)** | 1. **Distance\_km (Average)** |
| 1. **Card** | 1. **Avg Liters per Person per Day** | 1. **Liters\_Per\_Person\_Per\_Day (Average)** |
| 1. **Bar Chart** | 1. **Avg Distance per Zone** | 1. **Axis: Zone, Value: Distance\_km** |
| 1. **Stacked Column** | 1. **Households per Distance Bin** | 1. **Axis: Distance\_Bin, Value: Count** |
| 1. **Pie Chart** | 1. **Water Source Distribution** | 1. **Legend: Water\_Source, Values: Count** |

1. **🔸 How-To Instructions (in Power BI Desktop):**
2. **Import Data**
3. **Click Get Data → Text/CSV → Select kakuma\_cleaned.csv**
4. **Create Cards**
5. **Drag Distance\_km into a Card and set to Average**
6. **Do the same for Liters\_Per\_Person\_Per\_Day**
7. **Bar Chart: Avg Distance per Zone**
8. **Axis: Zone**
9. **Value: Distance\_km (set to Average)**
10. **Stacked Column: Distance Bin**
11. **Axis: Distance\_Bin**
12. **Values: Household\_ID (set to Count)**
13. **Pie Chart: Water Source**
14. **Legend: Water\_Source**
15. **Values: Count of Household\_ID**
16. **🔹 PAGE 2: WATER SUFFICIENCY & SANITATION INSIGHTS**
17. **🔸 Layout Suggestions:**

| 1. **Visual Type** | 1. **Description** | 1. **Fields** |
| --- | --- | --- |
| 1. **Donut Chart** | 1. **% Meeting 20L Standard** | 1. **Meets\_Water\_Standard (Count of True/False)** |
| 1. **Bar Chart** | 1. **% Households Meeting Standard by Zone** | 1. **Axis: Zone, Value: % True in Meets\_Water\_Standard** |
| 1. **Clustered Bar** | 1. **Sanitation Status by Zone** | 1. **Axis: Zone, Legend: Sanitation\_Status** |
| 1. **Table** | 1. **Detailed Records by Zone** | 1. **Columns: Zone, Sanitation\_Status, Water\_Source, Distance\_km** |

1. **🔸 Instructions:**
2. **Donut Chart**
3. **Field: Meets\_Water\_Standard**
4. **Count values (True vs False)**
5. **Bar Chart: Sufficiency by Zone**
6. **Create a measure (optional) or use visual filter:**
7. **Axis: Zone**
8. **Values: Count of Household\_ID**
9. **Filters: Meets\_Water\_Standard = True**
10. **Sanitation Status by Zone**
11. **Axis: Zone**
12. **Legend: Sanitation\_Status**
13. **Values: Count of Household\_ID**
14. **Table View**
15. **Add table with columns:**
16. **Zone, Water\_Source, Distance\_km, Sanitation\_Status, Liters\_Per\_Person\_Per\_Day**
17. **📁 OPTIONAL: Page 3 — Water vs. Sanitation (Advanced)**
18. **Scatter Plot: Distance\_km vs Liters\_Per\_Person\_Per\_Day, color-coded by Sanitation\_Status**
19. **Matrix or Heatmap: Zone vs Water\_Source, values = Average Liters**
20. **🛠 INTERACTIVITY TIPS**
21. **Add Slicers for:**
22. **Zone**
23. **Water\_Source**
24. **Distance\_Bin**
25. **Use Filters on this page to isolate specific analysis areas.**
26. **💾 Export & Deliver**
27. **Save the .pbix file and publish to Power BI service (optional)**
28. **Export to PDF if needed for printing or slides**
29. **Would you like me to generate:**
30. **A sample PBIX file for you (you'll upload cleaned CSV to complete it)**
31. **Python scripts to generate visual PNG charts to embed**
32. **Or help with a specific Power BI visual setup?**
33. **Let me know how hands-on you want the help to be!**
34. **Ask ChatGPT**
35. **Top of Form**
36. **Bottom of Form**
37. **🎯 Choose a Visual to Set Up:**
38. **📊 Bar Chart – Average Water Collection Distance by Zone**
39. **🧼 Stacked Column – Sanitation Facility Type by Zone**
40. **💧 Donut Chart – % of Households Meeting Water Sufficiency (≥ 20L)**
41. **📍 Map – Geographic Distribution of Sanitation Status (if you have coordinates)**
42. **📉 Scatter Plot – Distance vs Liters per Person, colored by Sanitation**
43. **📋 Interactive Table – Filterable by Zone & Water Source**
44. **Please reply with the number or name of the one you'd like help with (or describe it).  
    I’ll then give you step-by-step Power BI instructions to create that exact visual.**
45. **Ask ChatGPT**
46. **You said:**
47. **all please**
48. **ChatGPT said:**
49. **Perfect! Here's a step-by-step Power BI visual setup guide for all the key visuals your project needs, based on your cleaned dataset kakuma\_cleaned.csv. Each instruction assumes you've already imported the CSV into Power BI Desktop.**
50. **✅ 1. 📊 Bar Chart – Average Water Collection Distance by Zone**
51. **Goal:**
52. **Show which zones walk farther to collect water.**
53. **How to Create:**
54. **Go to the “Report” view.**
55. **Select the Clustered Bar Chart visual.**
56. **Drag Zone to the Axis.**
57. **Drag Distance\_km to Values → change aggregation to “Average”.**
58. **Add a data label (Format pane → Data labels → On).**
59. **✅ 2. 🧼 Stacked Column – Sanitation Facility Type by Zone**
60. **Goal:**
61. **Compare the types of sanitation facilities available in each zone.**
62. **How to Create:**
63. **Select Stacked Column Chart.**
64. **Drag Zone to Axis.**
65. **Drag Household\_ID to Values (it will count households).**
66. **Drag Sanitation\_Facility\_Type to Legend.**
67. **You can sort zones by total households or any specific facility type.**
68. **✅ 3. 💧 Donut Chart – % of Households Meeting 20L Sufficiency Standard**
69. **Goal:**
70. **Visualize water access sufficiency (based on Meets\_Water\_Standard column).**
71. **How to Create:**
72. **Select Donut Chart.**
73. **Drag Meets\_Water\_Standard to Legend.**
74. **Drag Household\_ID to Values (count of households).**
75. **Format the chart with:**
76. **Data labels (Format pane → Detail labels → Percent of total)**
77. **Title: “% of Households Meeting 20L/Person/Day”**
78. **✅ 4. 📍 Map – Geographic Distribution of Sanitation Status *(Optional)***
79. **Note:**
80. **Only do this if your dataset has latitude/longitude or location zones mapped to coordinates.**
81. **How to Create (if geodata exists):**
82. **Use the Map visual.**
83. **Drag Latitude and Longitude to Location.**
84. **Drag Sanitation\_Status to Legend.**
85. **Drag Household\_ID to Size (count of records).**
86. **❗ If you don’t have GPS data, skip the map or create a Choropleth-style substitute using bar charts.**
87. **✅ 5. 📉 Scatter Plot – Distance vs Liters/Person, Colored by Sanitation**
88. **Goal:**
89. **Visualize the relationship between distance walked and water volume collected.**
90. **How to Create:**
91. **Select Scatter Chart.**
92. **Drag Distance\_km to X-Axis.**
93. **Drag Liters\_Per\_Person\_Per\_Day to Y-Axis.**
94. **Drag Sanitation\_Status to Legend.**
95. **Drag Household\_ID to Details or Values (Size).**
96. **✅ 6. 📋 Table – Interactive Zone Summary Table**
97. **Goal:**
98. **Let viewers explore key values per household or per zone.**
99. **How to Create:**
100. **Select Table visual.**
101. **Add columns:**
102. **Zone**
103. **Water\_Source**
104. **Distance\_km**
105. **Liters\_Per\_Person\_Per\_Day**
106. **Sanitation\_Status**
107. **Sanitation\_Facility\_Type**
108. **Add a Slicer to control it:**
109. **Insert a Slicer → add field Zone**
110. **Another slicer for Water\_Source (optional)**
111. **🛠 Bonus Tips (Polish Your Dashboard)**
112. **Turn on Data Labels for all charts**
113. **Add Page Titles using text boxes (e.g., “Water Access Overview”)**
114. **Use consistent color themes (blue for water, brown for sanitation)**
115. **Use Tooltips: Hover to reveal more info (e.g., liters per person, zone name)**
116. **🔄 Dashboard Page Layout**
117. **🔷 Page 1: *Water Access Overview***
118. **Cards: Avg Distance, Avg Liters**
119. **Bar Chart: Distance by Zone**
120. **Donut Chart: Water Source Distribution**
121. **Stacked Column: Households by Distance Bin**
122. **🔷 Page 2: *Sanitation & Sufficiency Insights***
123. **Donut Chart: Water Sufficiency %**
124. **Bar Chart: Water Sufficiency by Zone**
125. **Stacked Column: Sanitation Facility Type by Zone**
126. **Table: Household-level details**
127. **🔷 Optional Page 3: *Advanced Correlation***
128. **Scatter: Distance vs Liters, color by Sanitation\_Status**
129. **Matrix or heatmap: Zone vs Water\_Source**
130. **Would you like me to prepare:**
131. **A PBIX template file structure you can populate?**
132. **Python visual exports (PNG) you can upload into Power BI or PowerPoint?**
133. **Let me know how you'd like to proceed!**
134. **Top of Form**
135. **Tools**
136. **Bottom of Form**
137. **Sanitation Conditions by Zone:** Identify which zones within Kakuma Refugee Camp exhibit the worst (or best) sanitation conditions, based on Sanitation\_Status and Sanitation\_Facility\_Type.
138. **Average Water Collection Distance:** Calculate and analyze the average distance walked by households to collect water across the camp and within different zones. Identify areas with significant access challenges.
139. **Water Sufficiency Assessment:** Assess whether households are accessing the recommended minimum of 20 liters of water per person per day, as per humanitarian standards. Identify the proportion of households meeting this standard and those falling short.
140. **Water Source Reliance and Efficiency:** Analyze the distribution of different Water\_Source types across zones and their correlation with Liters\_Per\_Person\_Per\_Day and Collection\_Times\_Daily.
141. **Correlation between Water Access and Sanitation (Optional but Recommended):** Explore if there's a relationship between better water access (e.g., shorter distances, higher per capita liters) and improved sanitation conditions in specific zones.

## **4. Tools and Deliverables**

### **4.1. Tools**

Students have the flexibility to choose their preferred data processing environment:

* **Option 1:** Microsoft Excel for data cleaning and transformation, followed by Power BI for visualization.
* **Option 2:** Python (using libraries like Pandas for data manipulation) for data cleaning and transformation, with the processed data then imported into Power BI for visualization.

### **4.2. Deliverables**

* **Power BI Dashboard (Minimum 2 Pages):**
  + **Page 1: Water Access Overview:** Provide a high-level summary of water access metrics (e.g., average distance, average liters per person per day) across the camp and broken down by zones. Include visualizations showing water source distribution and distance bins.
  + **Page 2: Water Sufficiency & Sanitation Insights:** Focus on the assessment of water sufficiency against the 20 liters/person/day standard and present findings related to sanitation conditions by zone. Include interactive charts and filters.
  + **Additional Pages (Optional):** Students are encouraged to create additional pages for deeper dives into specific insights or interactive explorations.
  + The dashboard should be interactive, clearly labeled, and visually appealing, adhering to best practices in data visualization.
* **PowerPoint Presentation (10-15 slides):**
  + **Introduction:** Briefly introduce the project, its objectives, and the importance of analyzing water access and sanitation in a refugee context.
  + **Methodology:** Describe the data cleaning and transformation steps taken, including challenges encountered and solutions applied.
  + **Key Findings & Insights:** Present the answers to the analysis objectives (Section 3.2) using compelling visualizations from the Power BI dashboard. Clearly articulate the insights derived from the data.
  + **Recommendations:** Based on the findings, propose actionable recommendations for improving water access and sanitation conditions in Kakuma Refugee Camp.
  + **Limitations & Future Work:** Discuss any limitations of the analysis and suggest areas for future research or data collection.
  + **Conclusion:** Summarize the main takeaways from the project.

## **5. Evaluation Criteria**

Projects will be evaluated based on:

* Completeness and accuracy of data cleaning and transformation.
* Depth and relevance of the analysis, addressing all stated objectives.
* Clarity, interactivity, and design quality of the Power BI dashboards.
* Effectiveness of the PowerPoint presentation in conveying insights and recommendations.
* Demonstration of critical thinking and problem-solving skills.

This project offers a valuable opportunity to apply data analytics skills to a humanitarian context, providing insights that can genuinely contribute to improving public health and well-being.