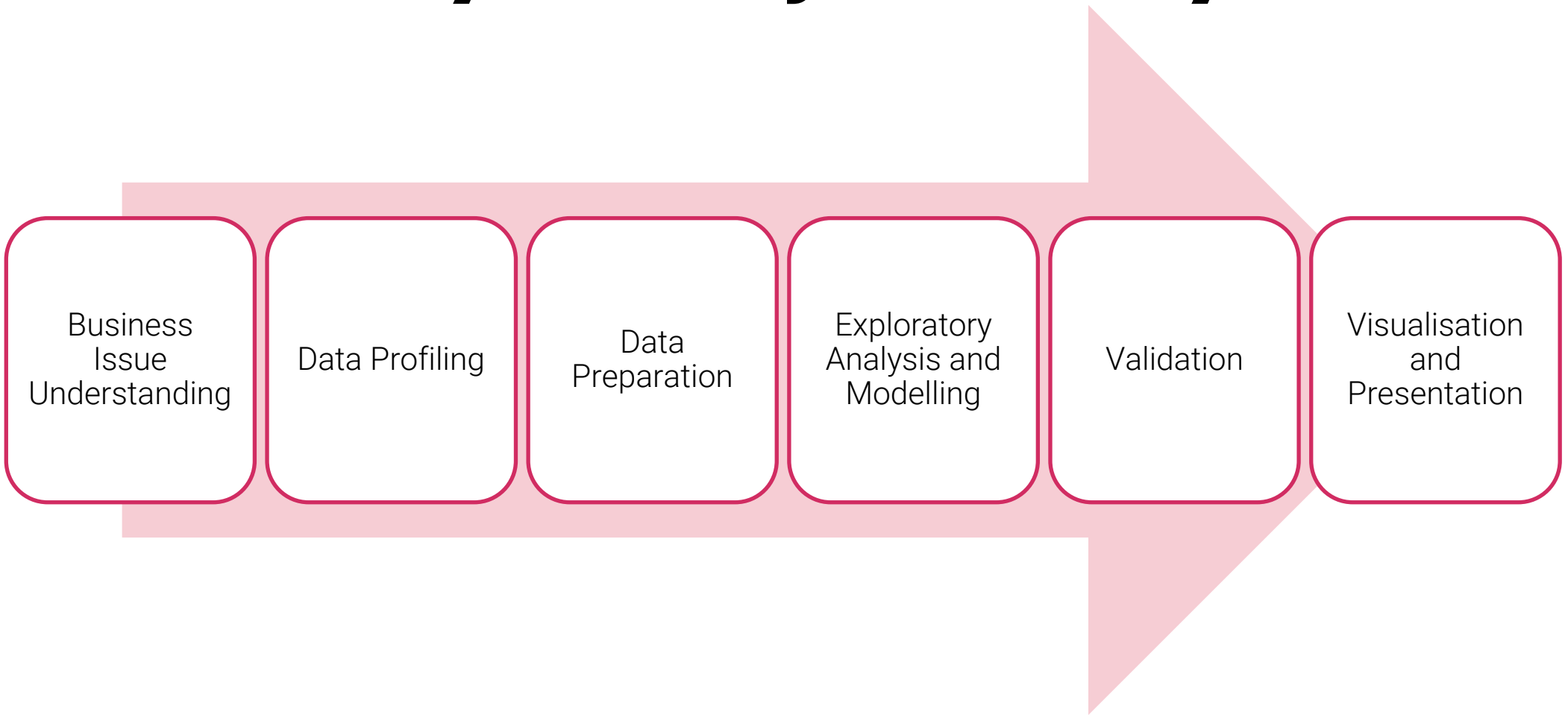
Operation Orphan

Data Science Project

CODE FIRST GIRLS

Chloé Koura
Irina Mateescu
Mayumi Hamaoka
Patricia Pedro
Oludáre Soniran

Data Analysis Project Lifecycle



Business Issue Understanding

Operation Orphan: Rubbish Science

- Operation Orphan is a charity that aims to increase scientific literacy by helping children develop problem solving strategies using freely available resources.
- Rubbish Science project:
  - increasing access to education
  - improving community health
  - raising environmental awareness
  - providing income for households
- Project currently in Vietnam and Sierra Leone. Operation Orphan aims to expand to Kenya.
- Applying for grant of £250,000

Business Issue Understanding

Key Objectives

To investigate the usefulness of expanding the Rubbish Science project to Kenya, we decided to focus on issues that may be highly impacted by the expansion:

1. **What are the waste management statistics?**
2. **What are the education continuity statistics?**
3. **Does malaria affect different Kenyan counties at different rates? Are bed net usage and health spending correlated to malaria incidence?**
4. **What are the rates of cholera mortality and clean water availability per area?**

Data Understanding

Data Sources

- Collected data via multiple APIs using python. Defined a function to apply for all data sources.
- Reviewed indicators available with potential use as variables for our analysis.
- Education
  - Enrolment by gender
  - Secondary education vs. population per county
- Health
  - % Bed net usage
  - % of fever and malaria
  - Health Spend
- Water contamination
  - Cholera rates
  - Diarrhoea mortality
  - Access to safe drinking water (disaggregated by rural/urban)

Data Understanding

Tools

Scientific computing libraries	• Pandas	• JSON/DataFrame/CSV applications
	• NumPy	• Statistical analysis • Arrays
	• Scikit-learn	• Preprocessing
Visualisations libraires	• Plotly	• Extension to other visualisation tools
	• Matplotlib	• Saving figures • Bar and line charts
	• Seaborn	• Correlation heatmaps
	• Geopandas	• Choropleth maps
Modules	• requests	• API GET requests
	• json	• json object manipulation
	• pprint	• improved readability
	• datetime	• date class manipulation
	• string	• string manipulation

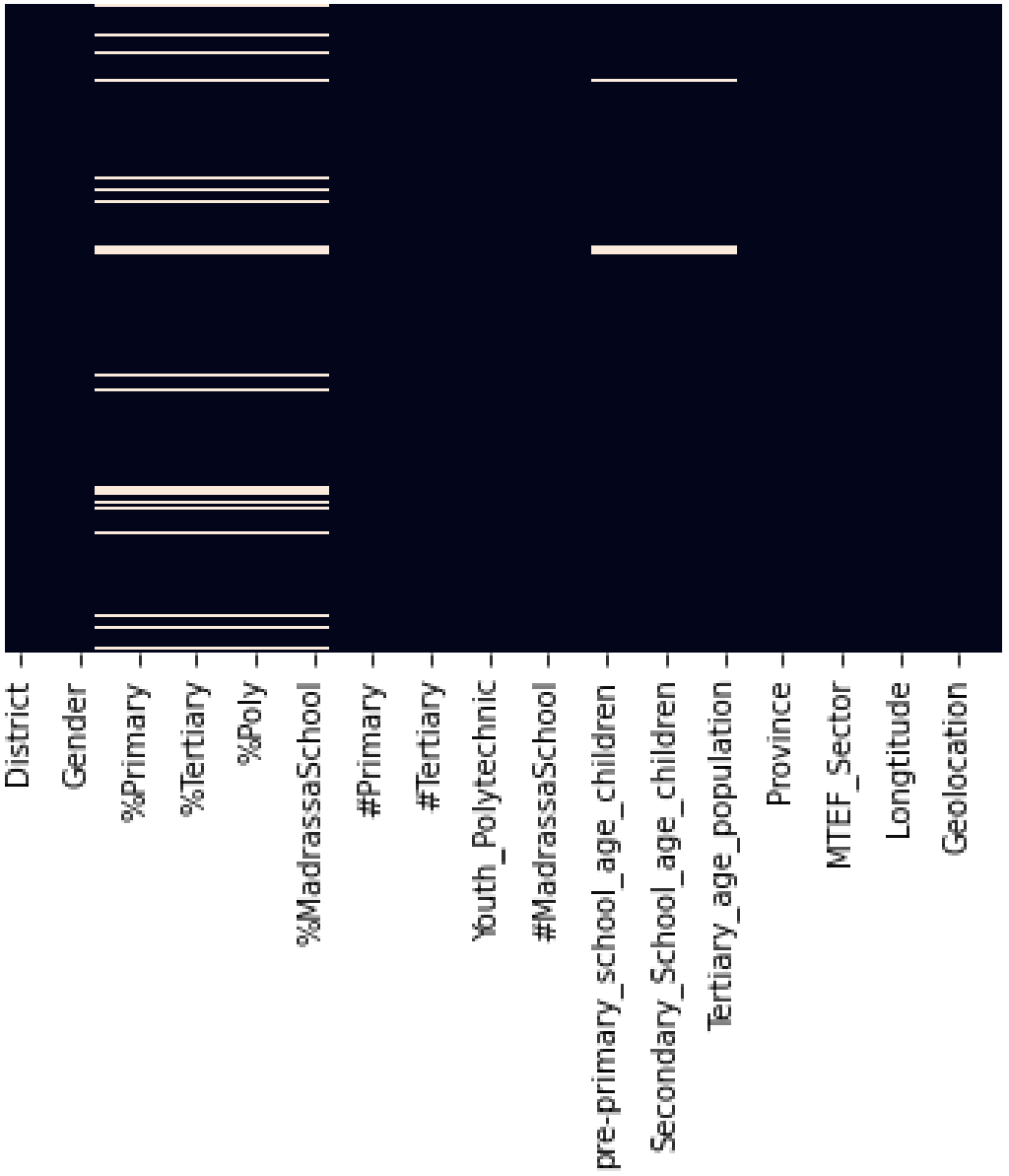
Data Preparation

Cleansing and transformation

- Education data
  - Seaborn heatmap to identify nulls
  - Dropping rows with nulls
  - Column conversion
  - New broad categories created to help categorise the data – unique values identified

- Health data
  - JSON conversion
  - Columns renaming
  - Input of averages for missing variables or eliminating nulls.

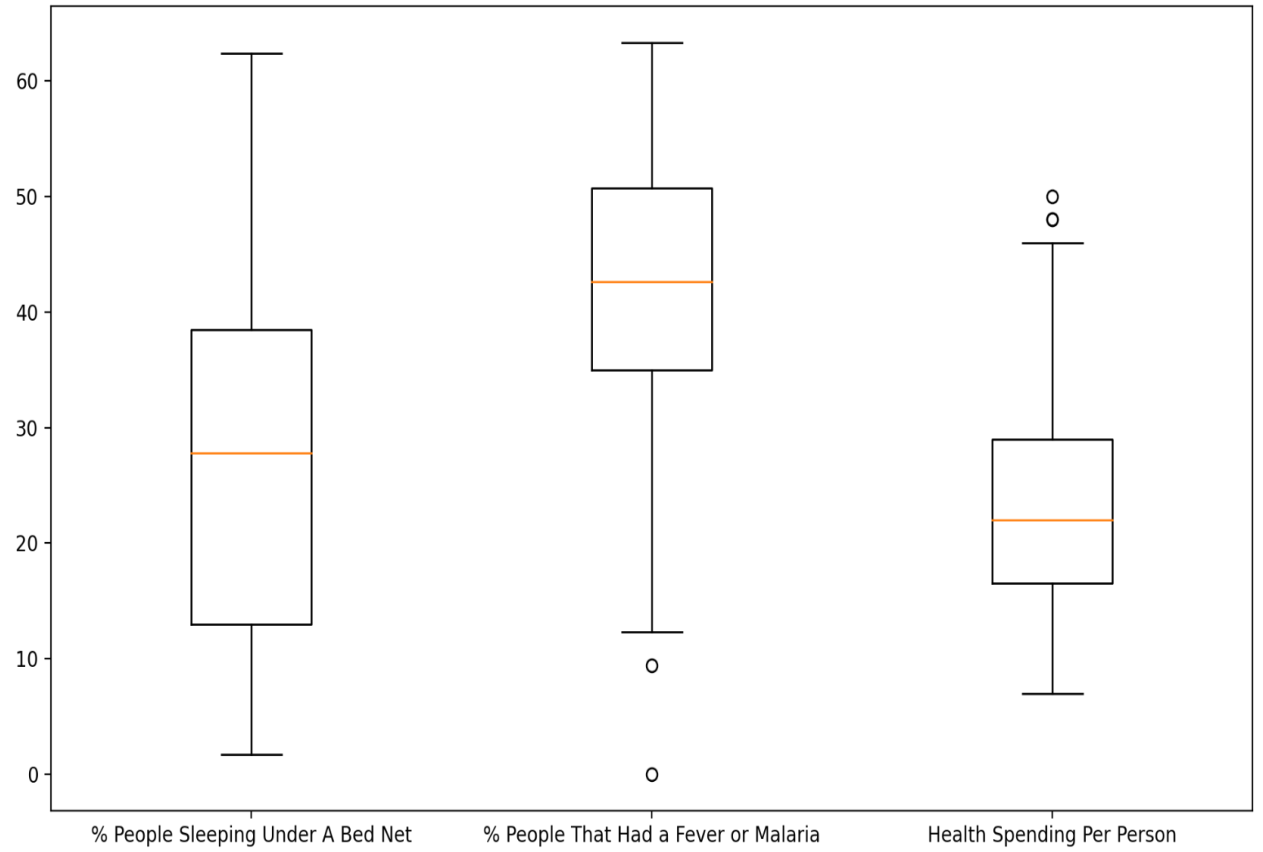
- Waster Contamination data
  - Removal of duplicates



Exploratory Data Analysis

Summary methodology

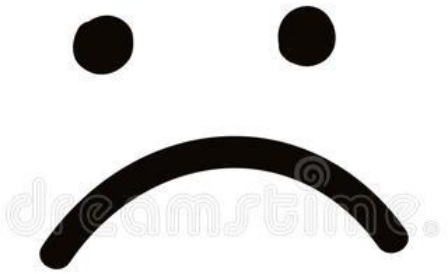
- Education data
  - df.describe() method
  - df.loc and groupby methods to extract groups of rows (separating by gender and edication levels)
  - .apply function for statistical method like sum, mean
- Health data
  - df.describe() method to review statistical summaries
- Waste Contamination data
  - Slicing to pull sections of the DataFrame
- *Heatmaps to review the correlation of the of different features*



Validation

Unapplied

- **Attempt at machine learning but none of the features were applicable, nor were the datasets robust enough.**
- **Proceeded to the visualisation stage.**
- **Had we been successful with machine learning, we would have revisited the previous steps if our results were not valid.**



Visualisations & Communication

1. What are the waste management statistics?

- Rubbish is readily available:
  - Mombasa generates approximately 2,000 tons of waste.
  - 65% collected
  - 13% access to waste management services
- Raising environmental awareness of upcycling:
  - Negative correlation found that larger households, ones closer to primary collection points or the more is spent on waste management >> the higher the tendency to haphazardly dispose of waste and increased risk of exposure to diseases.
- Aligns with the Kenyan government imitative Vision 2030:
  - The aim is to have “fully functional and compliant waste management systems, by developing strategies towards achieving sustainable waste”
  - Underway in five cities including Mombasa.
- Rubbish *Science would be a great project to aid the upcycling of recyclable material and raise environmental awareness from a young age.*

2. What are the education continuity statistics?

- **Education drops vastly after primary school, most likely due to the introduction of school fees.**
- **Rubbish Science's educational project is likely to be beneficial for communities regardless of gender.**
- **Operation Orphan could also decide to focus on the counties in the west, southeast and northeast, which have a high population (between 60,000 and 120,000) and levels of secondary education of only 20-40%.**

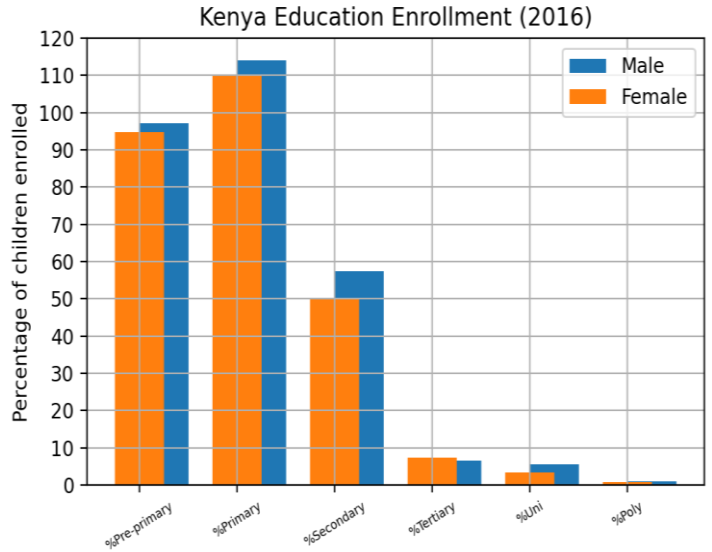


Figure 1: Education enrolment disaggregated by gender (2016)

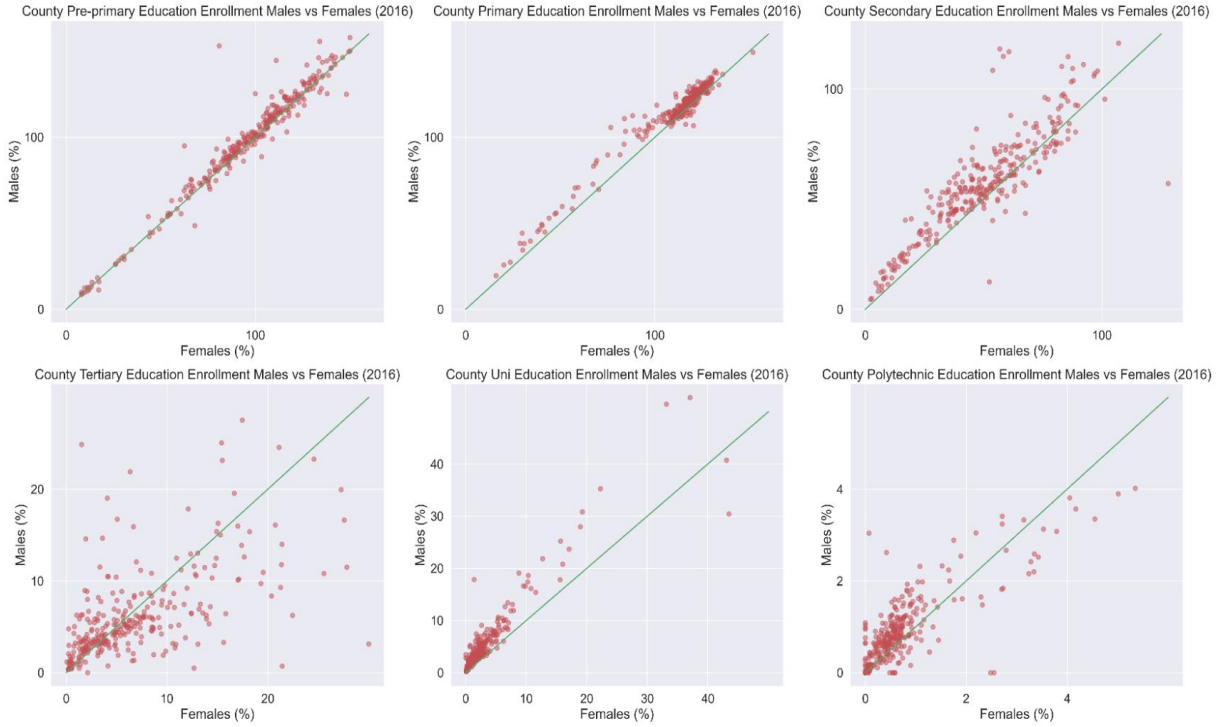


Figure 2: Subplots of gender distribution of education enrolment by county count (2016)

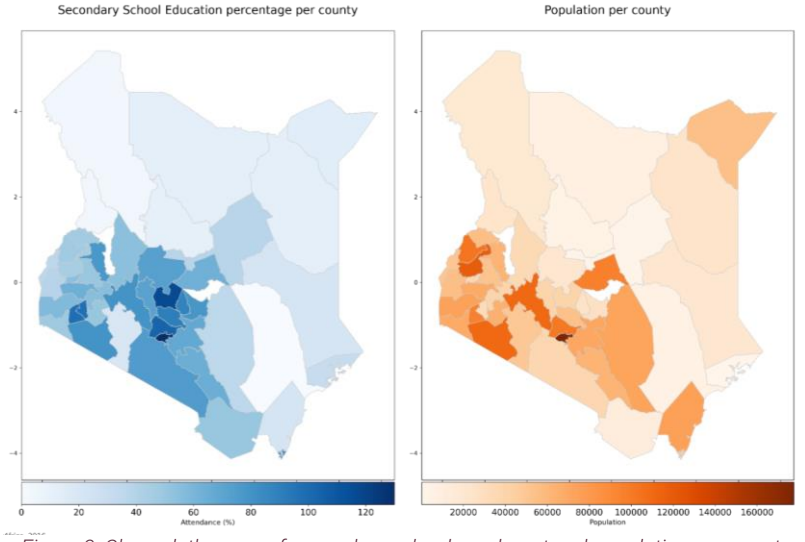
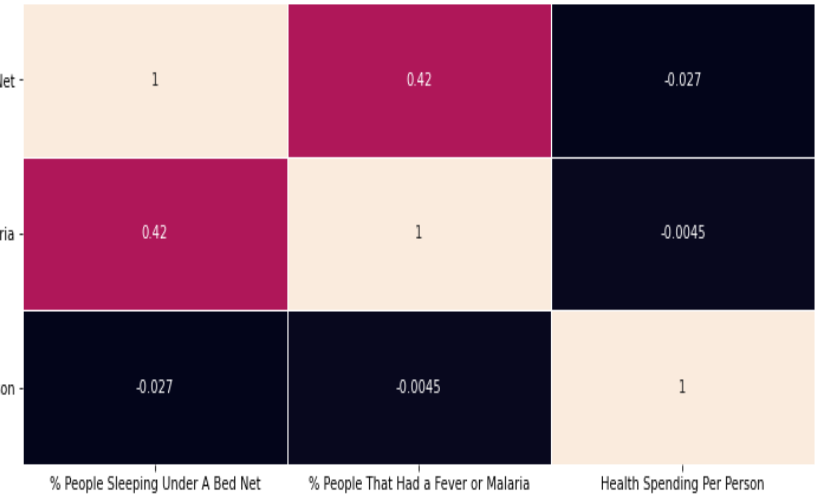


Figure 3: Choropleth maps of secondary school enrolment and population per county (2016)

3. Does malaria affect different Kenyan counties at different rates? Are bed net usage and health spending correlated to malaria incidence?

- **Majority of benchmark counties with high usage of bed nets still high rates of fever or malaria**
- **Health spend has no correlation with people sleeping under a bed net and people that had a fever or malaria.**
- **There is a weak positive correlation (0.42) between people sleeping under a bed net and people that had malaria. This should not be a deterrent to the usage of bed nets in malaria prevention.**
- **Rubbish Science should share their focus of teaching various methods of mosquito-borne disease prevention such as mosquito traps.**



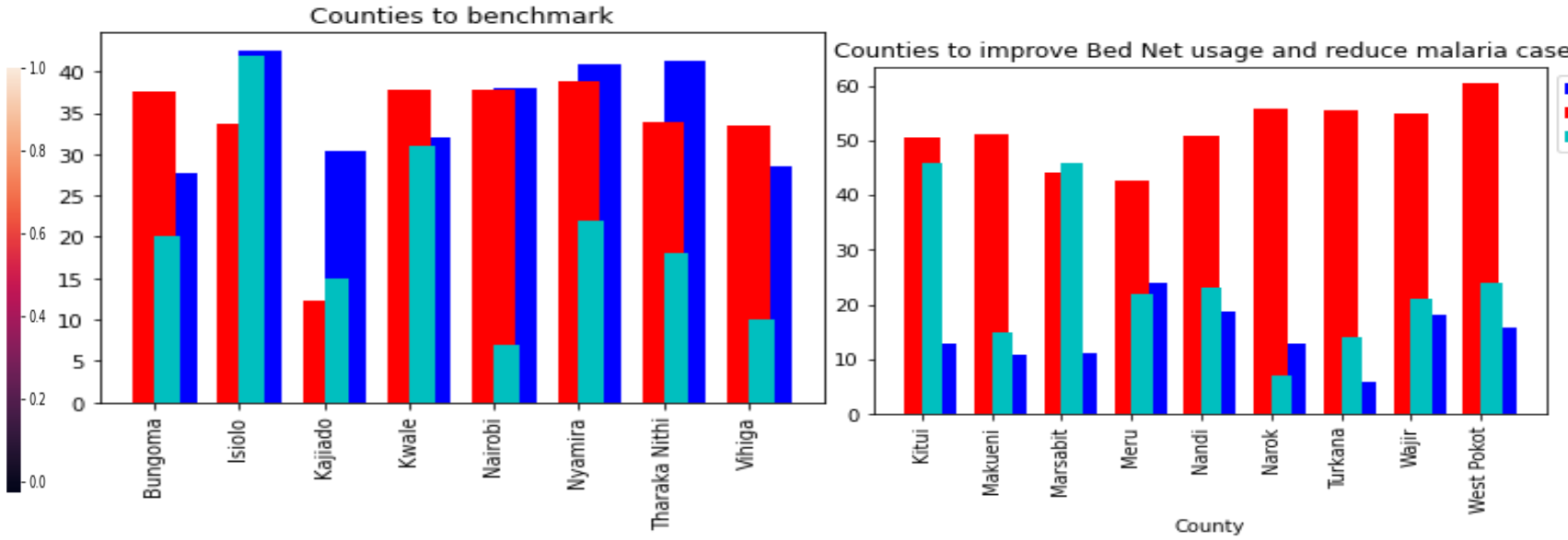


Figure 5: Graph 1: Counties to benchmark based on their high percentage in people sleeping under a bed net in comparison to the Kenyan average (26.73%) (2016). Graph 2: Counties with the lowest bed net usage in comparison to the Kenyan average (26.73%) (2016)

4. What are the rates of cholera mortality and clean water availability per area?

Solar water disinfection (SODIS) is a technique that consists of placing water into transparent plastic or glass containers (normally PET beverage bottles) which are then exposed to the sun. It has been repeatedly shown to be effective for eliminating microbial pathogens and reducing diarrhoeal morbidity including cholera.

The inexpensiveness and ease of SODIS would be an excellent lesson in Rubbish Science. It would teach children methods of creating safe drinking water to care for themselves and their families and hopefully reduce the cases of cholera.

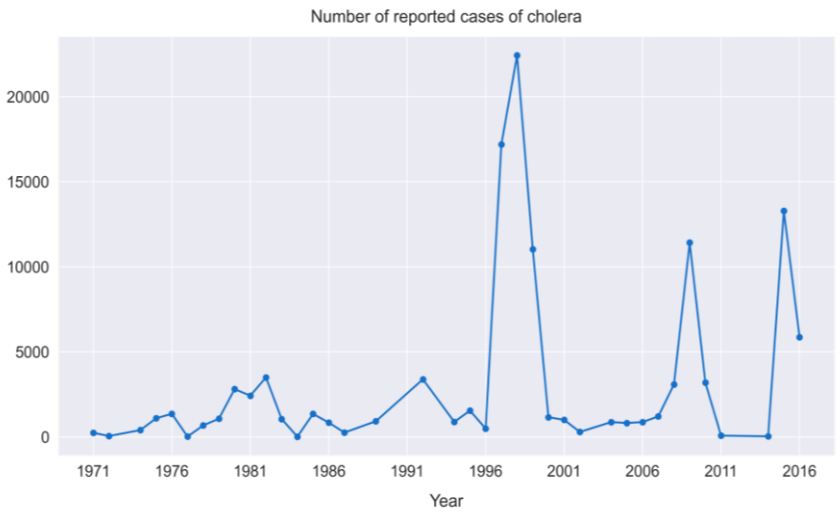


Figure 8: Number of reported cholera cases (2016)

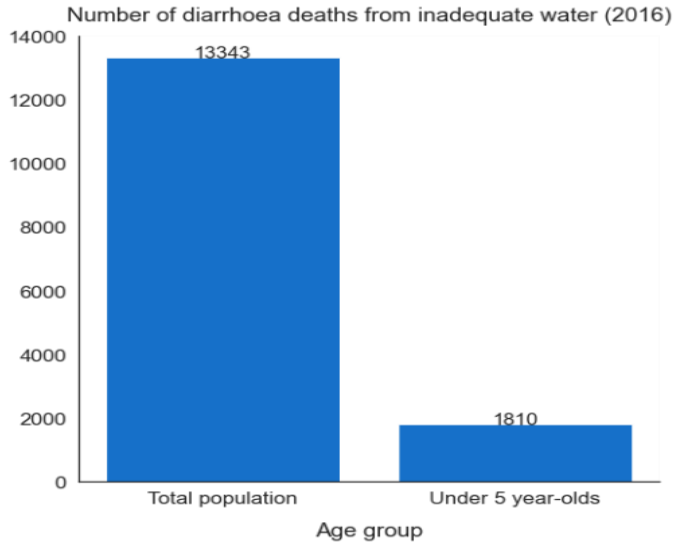


Figure 9: Number of diarrhoea deaths from inadequate water (2016)

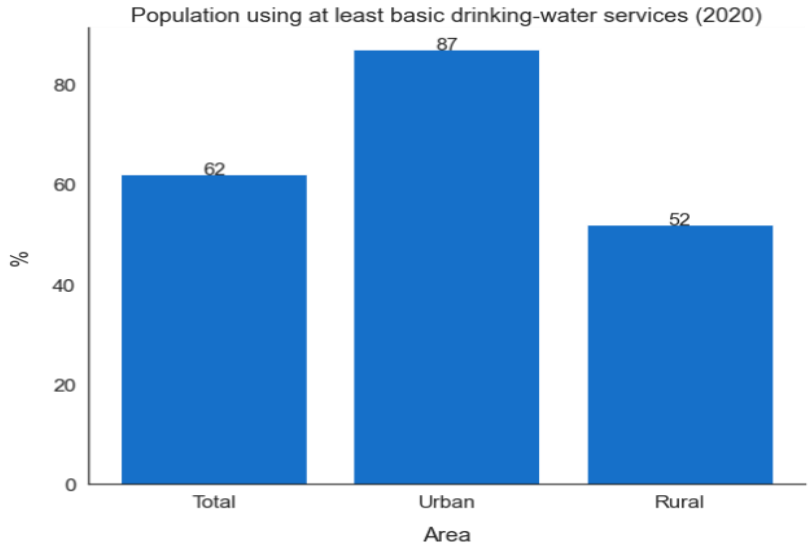


Figure 10: Population using at least basic drinking water services (2020)

Conclusion

Recommendation

We recommend the following:

1. **The counties with high population and low education levels should be the initial areas considered for Rubbish Science such as Nyeri and the counties in the west, southeast and northeast.**

NB. Operation Orphan was initially considering implementing the project in Mombasa.

2. **The target audience should be all children from underserved communities, regardless of their gender.**

3. **The emphasis on practical DIY experiments regarding mosquito nets alongside traps and water cleansing.**

SWOT Analysis

[S] Strengths

Teamwork: Effective communication, good distribution of responsibilities, passion for the topic.
Data sources: Multiple resources identified and used

[W] Weaknesses

Topic: Broadness of ideas
Data: Data security, outdated data and lack of data continuity.

[O] Opportunities

General python knowledge: Growing knowledge with opportunity to apply skills, good use of multiple libraries and modules
Documentation: Learning to read documentation

[T] Threats

Time constraints: Lots to accomplish with limited time
General python knowledge: Consolidation of continual learning

Questions?

Operation Orphan

Data Science Project

THANK YOU!

Special thanks:
Nina
Polly
Sabrina