Group_Assignment_INFS3603_T18A_G3.pdf

Quick Submit

Quick Submit

İstanbul Teknik Üniversitesi

Document Details

Submission ID
trn:oid:::1:2886195296

Submission Date
Apr 15, 2024, 8:53 AM GMT+3

Download Date
Apr 15, 2024, 8:55 AM GMT+3

File Name
Group_Assignment_INFS3603_T18A_G3.pdf

File Size
1.6 MB

27 Pages

4,326 Words

26,242 Characters

How much of this submission has been generated by AI?

***1%**

of qualifying text in this submission has been determined to be generated by AI.

* Low scores have a higher likelihood of false positives.

Caution: Percentage may not indicate academic misconduct. Review required.

It is essential to understand the limitations of AI detection before making decisions about a student's work. We encourage you to learn more about Turnitin's AI detection capabilities before using the tool.

Frequently Asked Questions

What does the percentage mean?
The percentage shown in the AI writing detection indicator and in the AI writing report is the amount of qualifying text within the submission that Turnitin's AI writing detection model determines was generated by AI.

Our testing has found that there is a higher incidence of false positives when the percentage is less than 20. In order to reduce the likelihood of misinterpretation, the AI indicator will display an asterisk for percentages less than 20 to call attention to the fact that the score is less reliable.

However, the final decision on whether any misconduct has occurred rests with the reviewer/instructor. They should use the percentage as a means to start a formative conversation with their student and/or use it to examine the submitted assignment in greater detail according to their school's policies.

How does Turnitin's indicator address false positives?
Our model only processes qualifying text in the form of long-form writing. Long-form writing means individual sentences contained in paragraphs that make up a longer piece of written work, such as an essay, a dissertation, or an article, etc. Qualifying text that has been determined to be AI-generated will be highlighted blue on the submission text.

Non-qualifying text, such as bullet points, annotated bibliographies, etc., will not be processed and can create disparity between the submission highlights and the percentage shown.

What does 'qualifying text' mean?
Sometimes false positives (incorrectly flagging human-written text as AI-generated), can include lists without a lot of structural variation, text that literally repeats itself, or text that has been paraphrased without developing new ideas. If our indicator shows a higher amount of AI writing in such text, we advise you to take that into consideration when looking at the percentage indicated.

In a longer document with a mix of authentic writing and AI generated text, it can be difficult to exactly determine where the AI writing begins and original writing ends, but our model should give you a reliable guide to start conversations with the submitting student.

INFS3603 Written Report

Alyssa Fasiha Azhar

Meryl Lingham

Wan Nur Alissa Binti Wan Mohd Zakirudin

Muhammad Faris Nurhafiz Abdul Khalid

Ukjin An

Table of Content

Introduction

Gender-based violence (GBV) is a global issue affecting vulnerable communities worldwide. KenyaPulse's commitment to uplifting impoverished communities within Nairobi compelled its strategic goals to combat GBV. By identifying hotspots, understanding vulnerabilities, evaluating program impact, and advocating for policy changes, this dashboard aims to fulfil KenyaPulse's mission.

Following are KenyaPulse's strategic goals and its following impacts:

1. To raise GBV awareness within the urban slums community

This effort is crucial as it encourages societal dialogue, which in turn spreads awareness about the prevalence and consequences of GBV. The National Crime Research Centre's (2014) studies indicate that Kenyans have a general understanding of GBV but with notable gaps in awareness concerning certain types of violence and child victimization. The prevalent perception among both genders was that GBV involves physical harm by men against women. However, other forms of GBV such as abuse of children or psychological harm, were less recognized.

2. To be an advocate of policy changes.

Kenyan authorities fail to ensure that GBV survivors have access to quality medical treatment, including mental health, protection services, and financial aid, and to thoroughly investigate and prosecute cases (Human Rights Watch 2021). Despite Kenya adopting treaties, such as the UN Convention on the Elimination of All Forms of Discrimination against Women (CEDAW) and enacting national laws like the Sexual Offences Act to criminalize violence against women, activists argue that government policies remain largely ineffective. There are no government-backed national campaigns addressing GBV or its consequences, and the justice system is criticized for its slow and inefficient prosecution of offenders (Lawal, S 2024).

3. To facilitate access to support services

This ensures that individuals affected by gender-based violence have access to essential support services such as healthcare and education. This initiative aims to bridge the gap in service delivery, providing tangible services that directly benefit survivors and contribute to their recovery and empowerment.

4. To empower affected communities

KenyaPulse is committed to providing education initiatives and leadership development programs as it seeks to offer opportunities for skill-building, mentorship, and other forms of support. The organization would like to enable individuals to amplify their voices, fostering a culture of solidarity and empowerment within the community.

Ethical considerations and privacy concerns are critical in data analytics through data collection, storage, and usage. Following its values as a non-profit organisation specialising in marginalised communities, it is especially essential that KenyaPulse complies with moral standards and the legal obligations surrounding data. Globally, the ethical standard for organisations is to be socially conscious, considerate throughout their data operations and comply with UNSDG principles on data (United Nations Development Group 2017). In data collection, it is critical that primary data is obtained

through informed consent and that secondary data sources are verifiable and follow ethical collection (People in Need 2022). Furthermore, as the data was sourced primarily and on a sensitive topic like GBV, it is important that KenyaPulse prepared analysts surveying to be emotionally considerate, provide referral services and monitor safety (World Health Organization 2007). Teams must respect participants' right to privacy through the removal or encryption of identifiable information when storing data and creating databases. Encryption also aids in cybersecurity to protect sensitive data from unrestricted access. The ethical practice of fair data usage involves transparency to data owners on its usage, companies monitoring future usage and ensuring employees cannot use the data maliciously. KenyaPulse must also comply with Kenya's legal requirements such as those outlined in the Data Protection Act 2019. As a non-profit organization dedicated to empowering communities, KenyaPulse must always consider ethical values and always prioritize community concerns.

Power BI Dashboard Overview, High-Level Design and Functionality

High-level design principles adopted in creating the dashboard.

- **User Centric Design**
  The image of activists fighting for GBV used for the dashboard header sets the context for the dashboard's focus. User interface elements, such as slicers, filters, and text boxes presented consistently throughout the pages of the report, ensure a smooth user experience across all pages. A text box with essential background information on Kenya, GBV, and the purpose of the dashboard follows, thus aiding the interpretation of data visualisations. Visual interaction, including filters and the highlight feature, serves to let the user concentrate on exclusive arrays of data and for their data-driven decision-making process. Tooltips facilitate navigation for beginners. Geometric shapes, for instance, rectangles or lines, provide a function as a sectional divider, helping the content arrangement and user navigation. Across, visualisations are fuelled by the cross-filtering and cross-highlighting functionalities that enable users to dig deeper into the data.

- **Data Visualisation Best Practices**

  The KenyaPulse dashboard prioritises clear communication through carefully selected chart types and consistent visual style with cohesive colour palettes and fonts. To avoid clutter, we have minimised "chart junk" and incorporated filtering features like "Top N" in Power BI which allow users to focus on the most significant data points by filtering out less relevant information, ensuring clean and centred visualisations. Techniques such as ordering bars by value and employing muted backgrounds ensure the data speaks clearly while keeping the interface clean and focused.

- **Accessibility Considerations**

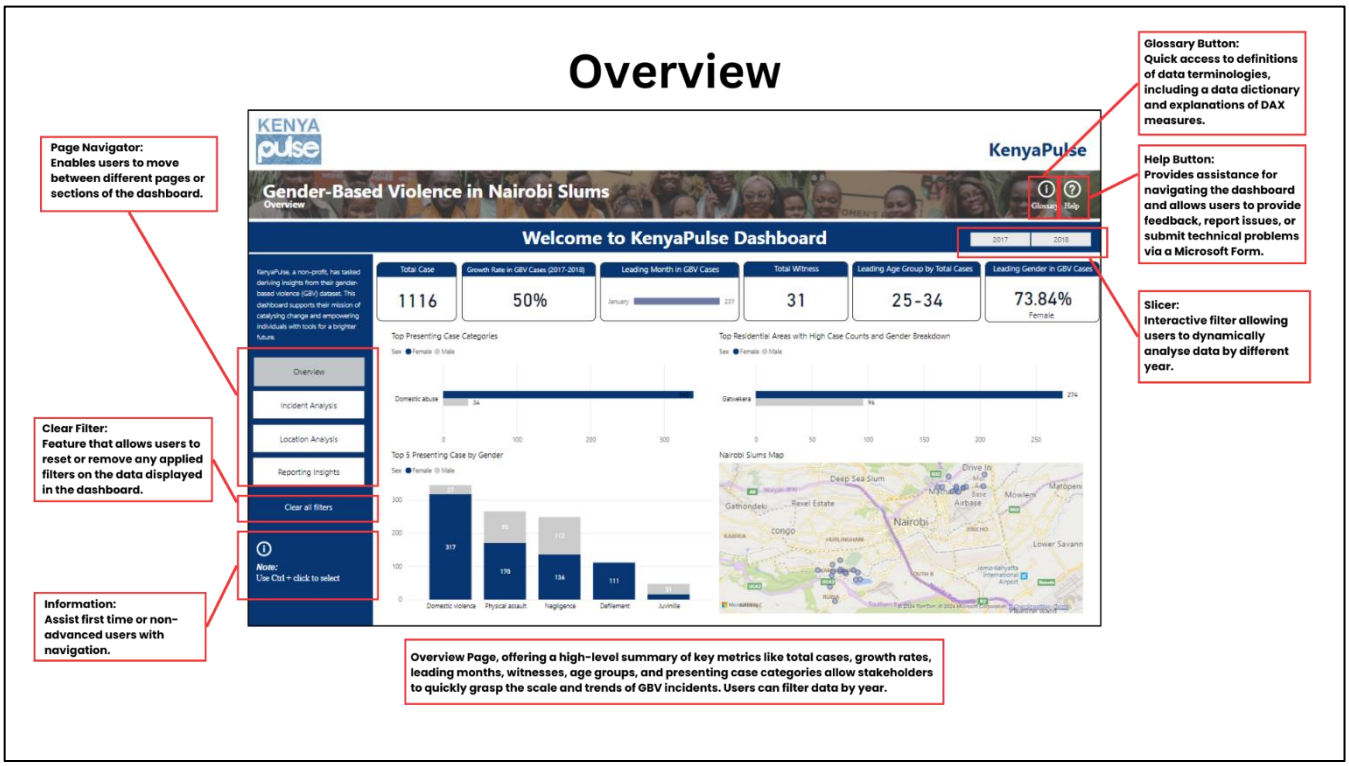
  The KenyaPulse dashboard prioritises accessibility with tested screen user compatibility, descriptive visual titles, and other visual elements across all pages. Utilising an accessible report theme ensures consistent formatting and colour contrasts for a cohesive and inclusive user experience.

Functionality of the dashboard and key visualisations.

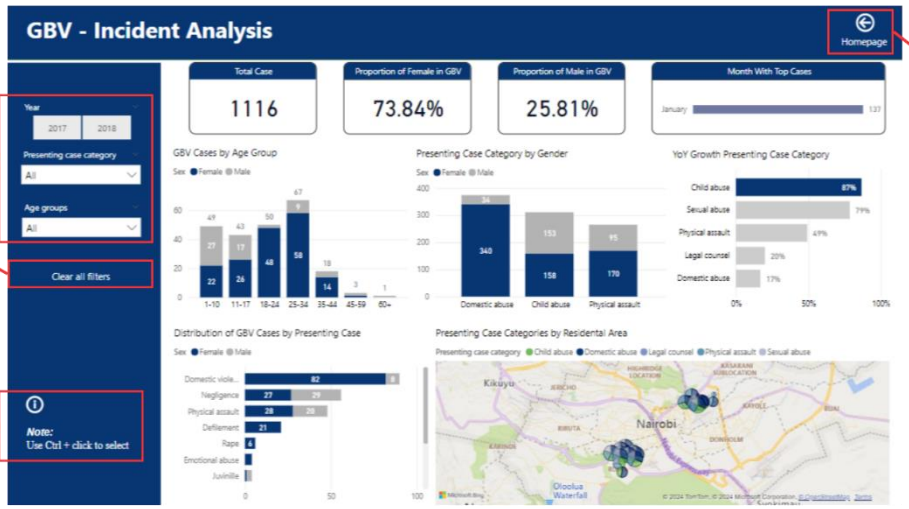
- **DAX Measures**

  We used DAX measures to enhance our dashboard visualisations as it offers custom calculations and dynamic aggregations, enabling deeper analysis and uncovering valuable insights within our data (*See Appendix A for detailed DAX measure*).

Table Name	Display Folder	Measure Name	Description
Case_DT		Case Percentage	
		Current Year (2018)	
		Previous Year	
		Total Case	
		YoY Growth by Category	
		YoY_Growth_Rate	
Victim_DT		LeadingAgeGroupByTotalCases	
		TotalVictims	



Incident Analysis



Slicer:
Interactive filter allowing users to dynamically analyse data by different category.

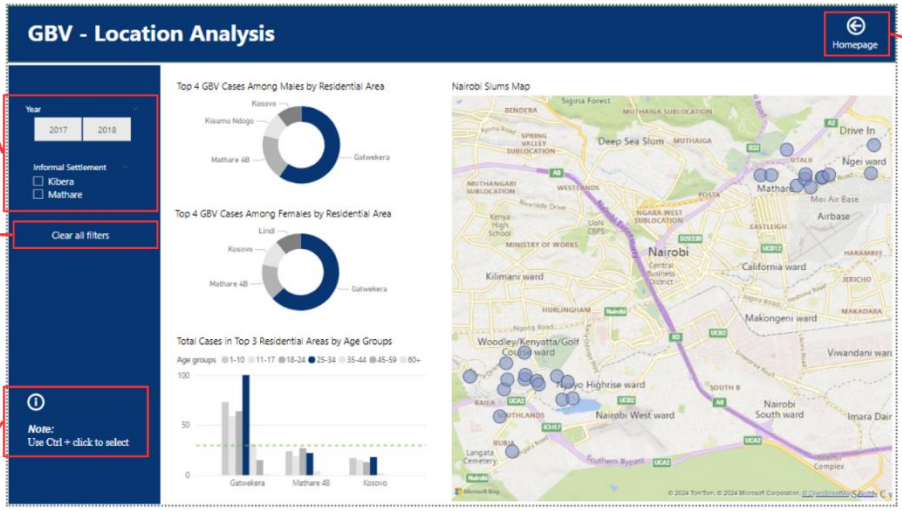
Clear Filter:
Feature that allows users to reset or remove any applied filters on the data displayed in the dashboard.

Information:
Assist first time or non-advanced users with navigation.

Homepage Button:
Serves as a quick link to return users to the main or starting page of the dashboard

The Incident Analysis Page allows for deeper exploration, enabling users to filter by year, case categories, and age groups. Visualisations include age group distribution and gender breakdown of presenting case categories provide nuanced insights into the characteristics of GBV incidents, aiding in the development of targeted interventions and resources

Location Analysis



Slicer:
Interactive filter allowing users to dynamically analyse data by different category.

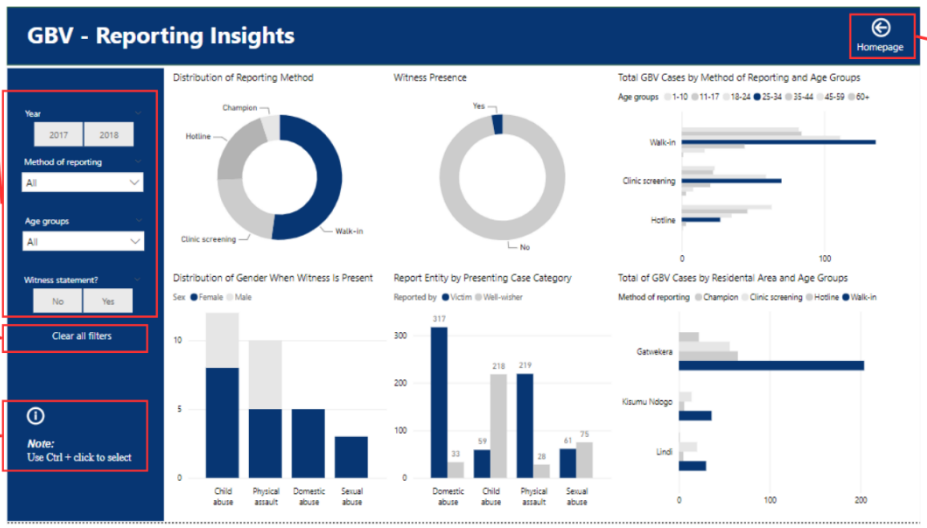
Clear Filter:
Feature that allows users to reset or remove any applied filters on the data displayed in the dashboard.

Information:
Assist first time or non-advanced users with navigation.

Homepage Button:
Serves as a quick link to return users to the main or starting page of the dashboard

The Location Analysis Page focuses on understand GBV patterns across residential areas and informal settlements crucial for effective resource allocation and intervention strategies. The map visualisation on the Location Analysis Page helps identify hotspots and prioritise areas for targeted interventions. Users can filter by year and location, with a map visualisation showing presenting case categories.

Reporting Insights



Slicer:
Interactive filter allowing users to dynamically analyse data by different category.

Clear Filter:
Feature that allows users to reset or remove any applied filters on the data displayed in the dashboard.

Information:
Assist first time or non-advanced users with navigation.

Homepage Button:
Serves as a quick link to return users to the main or starting page of the dashboard

The Reporting Analysis Page examines reporting methods and witness involvement, with filters for year, reporting method, age group, and witness presence, where stakeholder can gain insights into the barriers victims face in reporting GBV incidents and the effectiveness of current reporting mechanisms. This information is invaluable for improving reporting systems and ensuring victims have access to support and justice.

Glossary

Case id: Unique identifier for each incident
Intake Date: Date when the incident was reported
Presenting case: Description of the presenting incident
Presenting case category: Category of the presenting incident
Reported by: Entity reporting the incident
Method of reporting: Method used to report the incident
Witness: Indicates if there was a witness
Witness statement: Indicates if there was a witness statement
Residential Area: Residential area where the incident was reported
Latitude: Latitude coordinates of the incident
Longitude: Longitude coordinates of the incident
Informal Settlement: Indicates the informal settlement where the incident took place
Age: Age of the individual involved in the case
Sex: Gender of the individual

DAX Measures: Click Here

Click here button:
Allow users to navigate to an Excel file containing detailed records and DAX measures.

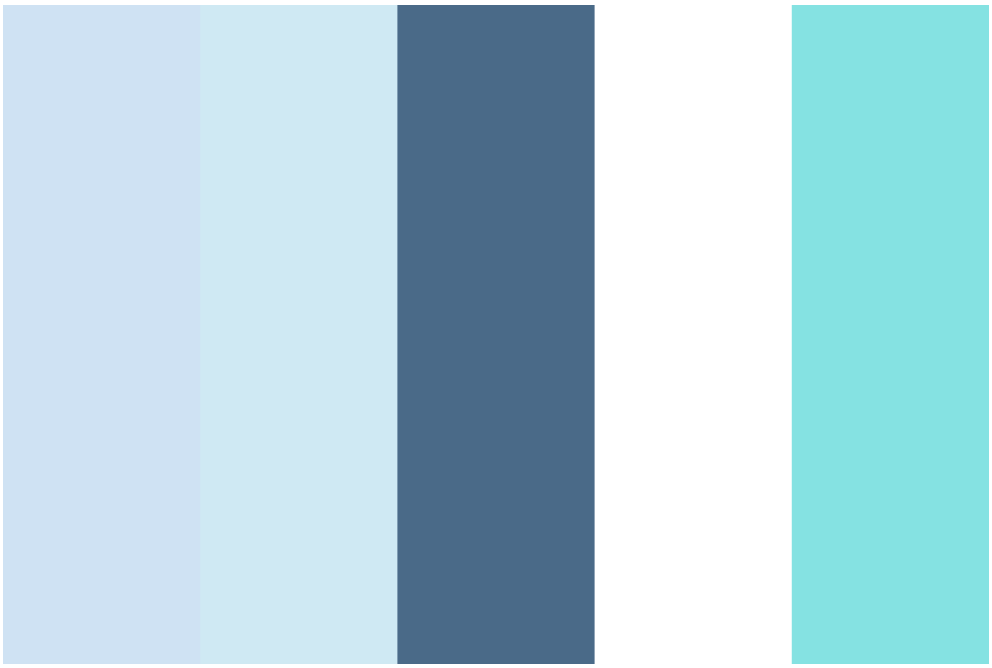
The Glossary Page is a centralised reference within our Power BI dashboard. It includes a Data Dictionary with clear definitions of data terminologies and explanations of DAX measures, ensuring users have the necessary clarity to explore and analyse the data effectively.

UX/UI wireframes and design decisions

This section of the report incorporates the design proposal and ideas which were used to create the prototype of the dashboard.

Design Proposal

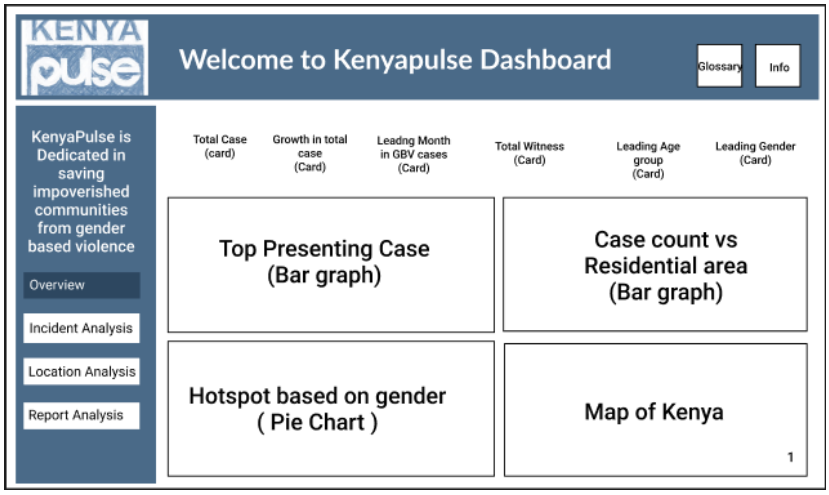
Initially, our team had a dispute over the colour choice between orange and blue. Blue has eventually been chosen as the primary colour of the dashboard due to its many advantages. Blue is often associated with trust and stability, widely recognized for its calming and soothing properties. By creating a calming interface, users are likely to feel more at ease while interacting with the dashboard which can improve focus on it.



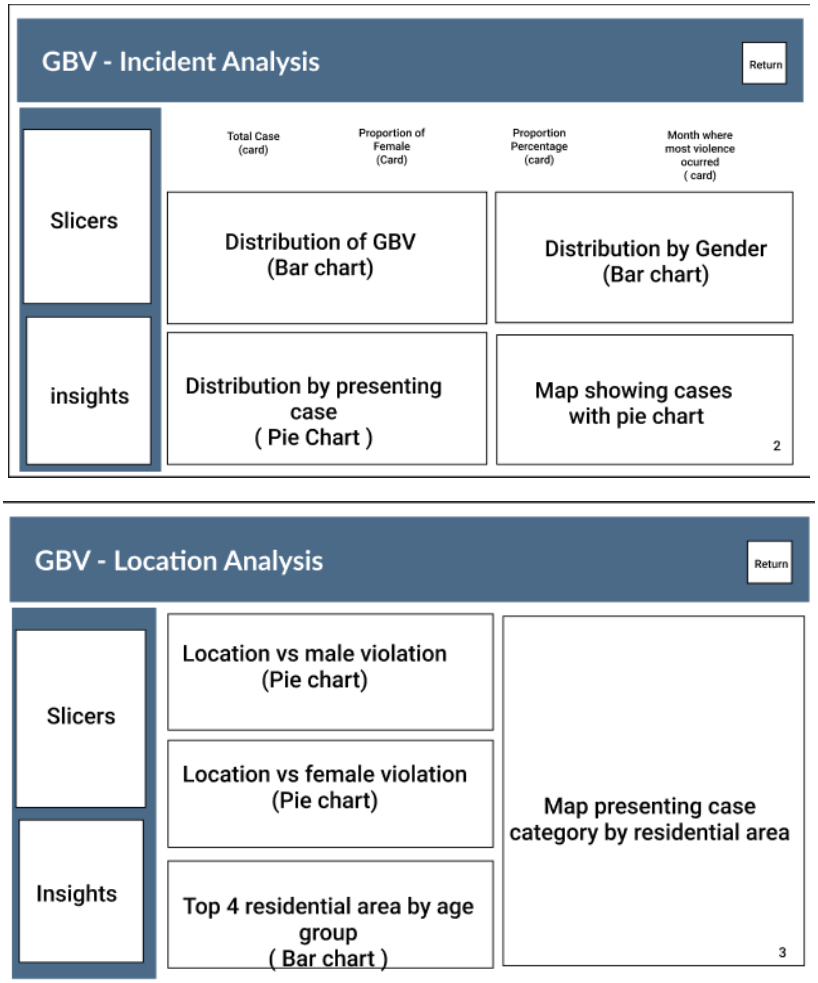
Primary Font: Segoe
Secondary Font: Times New Roman

Segoe was selected as the primary font due to its clean and modern appearance, which enhances readability and user interface clarity. Times New Roman was chosen as the secondary font to complement Segoe in detailed descriptions where enhanced readability for longer reads is essential. These fonts together provide visual hierarchy allowing quicker scan and information gain from the readers when utilised with boldness.

Based on the Design decisions that were developed in the earlier section and communication with teammates, the wireframe was developed using Figma as shown.



At this part of the prototype, navigation was streamlined with a clear, intuitive single-coloured layout facilitating quick access to different sections of the dashboard with the use of buttons. This allows users to navigate without too much information at once.



To maintain simplicity across the dashboard, pages 2, 3, and 4 share a similar layout. This design choice simplifies navigation and improves usability by providing a familiar interface as users switch between these pages.

Insights into prioritisation, data story, and features for future iterations

To prioritise features and functionalities of the dashboard, we highlighted features that directly contribute to understanding, analysing and addressing GBV incidents. The selected dashboard segments 'incident analysis, location analysis and reporting analysis' directly support this goal. Additionally, our team considered user feedback and requirements during prioritisation. Functionalities, such as slicers and buttons for page navigation, enhance the user experience by enabling interactivity and data exploration for users. Secondly, features were aligned with KenyaPulse's strategic goals in combating GBV. Specifically, we prioritised features that facilitated data-driven decision-making and strategic planning.

Aligning with KenyaPulse's strategic goals, the **overview page** contributes by providing key insights and trends related to GBV cases, raising awareness about the prevalence and consequences of GBV in Kenya. The **incident analysis page** delves deeper into specific GBV incidents, providing data-driven insights useful for advocacy and policy changes at both local and national levels. The **location analysis page** helps identify hotspots of GBV incidents, hence facilitating targeted interventions and ensuring individuals affected by GBV have access to essential support services. The **reporting analyses page** provides comprehensive insights into reporting dynamics, demographics involved and patterns in reporting methods and witness statements, supporting advocacy efforts and policy changes. The **functional slicers** help users filter data dynamically, enabling them to focus on specific subsets of information and derive actionable insights efficiently. The **buttons** help enable seamless navigation between report pages, facilitating the exploration of insights and analyses. The **clear filter** allows users to remove all applied filters from the report easily. This feature is important for new users to "de-select" or reset all applied filters. Users may also double-click on the visualisation (e.g.: a column) to reset the applied filters. However, the "clear filter" is a user-friendly option for achieving the same goal. This ensures an intuitive yet dynamic view of the GBV data. The **cards** that display a single numeric value allow users to quickly grasp key metrics and trends related to GBV incidents. A **filter of the month with top cases according to the year** enables easy identification and analysis of months with the highest number of reported GBV cases, supporting targeted interventions and resource allocation.

Effort estimation and prioritisation of epics are covered with the RICE Method, which can be seen in Appendix D. Following this, we built the approximate key timelines to ensure progress can be tracked (refer to Appendix E).

Considering emerging trends and stakeholder requirements, some potential enhancements for future iterations may include incident severity metrics, which can help quantify the severity of GBV incidents. For instance, including the psychological trauma experienced, the level of physical harm inflicted and the duration of abuse. We can also explore opportunities to integrate the GBV dataset with healthcare records such as hospital admissions and mental health consultations to better aid stakeholders in understanding the medical consequences on victims. Next, we discovered from the user feedback that the data is incomplete as the intake date is only until August 2017, and until June 2018. We can engage with relevant stakeholders, such as healthcare providers to obtain supplementary data sources to reconstruct missing data for periods with incomplete records. From the end-user point of view, a preview button is also essential for reviewing the entire dashboard. This button prevents disruption during testing, as interactions between elements, such as buttons and text boxes, can be thoroughly

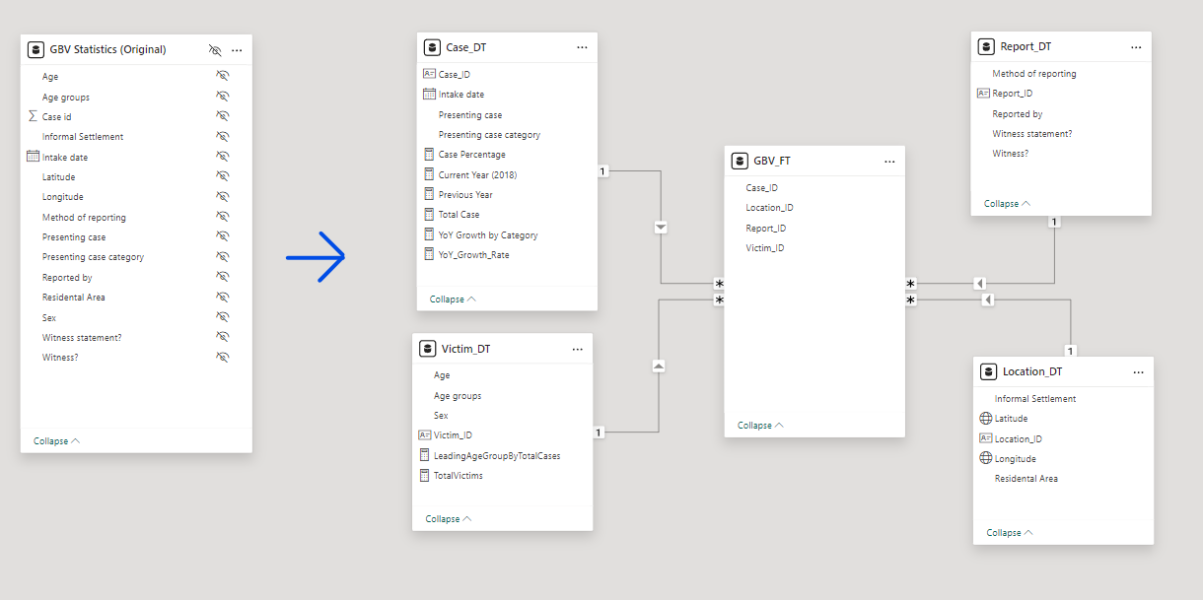
examined. For future iterations, we can develop and integrate a dedicated preview button directly into the dashboard interface, allowing the end-users to easily access the preview mode without disrupting the testing process.

Overview of data collation, ETL processes, and alignment with testing requirements

The data collation process involved the primary data sourced by KenyaPulse through surveys. Investigation of the location metrics revealed that the data was gathered from informal settlements within Nairobi, Kenya. Specifically, the geographical data was predominantly composed of Mathare and Kibera reports. Although seemingly biased, the data is consistent in following KenyaPulse's goal to target urban slums for improvement, as both settlements are densely populated slums with inadequate living conditions contributing to the increased rates of violence. For further analysis, our team located secondary data sources online to extract, such as from the Humanitarian Data Exchange and verified research reports. However, due to confounding variables, limited data context and raw data, we deemed the databases and website sources unusable to develop an unbiased and accurate analysis.

The flat data Excel file provided by KenyaPulse was loaded into the Power BI workspace by the 'Get Data' tool. Through the Power Query Editor from "Transform data", the data was cleaned and prepared for analysis. 'Column quality', 'Column distribution' and 'Column profile' were enabled to view initially to identify potential issues quickly and provide an overview of the data. Additionally, filters and appropriate data types were applied to identify structural errors and outliers in the cleaning process. The main issues with the messy data were inconsistent, duplicate and missing values. To maintain data integrity, we aimed to remove as little data as possible (See *Appendix B* for a detailed list of applied methods to address each error).

Finally, after data cleaning, the flat data was normalised and separated into tables following the star schema format. Normalised data ensures data quality and efficiency for analysis through reduction of redundant data and one-to-many relationship cardinalities. Primary Key IDs were generated for each dimensions table through the 'Index Column' in Power Query to maintain data integrity of unique values. Additionally, the centre fact table contains foreign keys to easily link information between tables for efficient analysis and visualisation loading. Overall, data quality and integrity were verified at every step of the ETL process by data profiling, data imputations, and understandable relationship modelling.



Testing procedure is an important part in completing the dashboard. In this instance, the testing procedure is divided into two categories; 1) data accuracy and 2) user interface testing. The first part tests the accuracy and reliability of the dashboard data while the second part puts the dashboard performance and satisfaction to test *(See Appendix C for Test Case Process).*

To test the data accuracy and reliability, we import the data from PowerBI into Microsoft Excel, and we mimicked the procedures to replicate the results. The tests include verifying that the legends, labels, numbers, dates, and strings entries match in both Excel and PowerBI. We put all visualisation to test and along the way, have met some data errors and discrepancies. The Test Analyst is responsible to alert the Product Manager and Data Analyst and work together with them to correct the mistakes. Then, the Test Analyst will document the process and compile all documentations. Therefore, data accuracy testing procedure is essential to prevent misinformation and uphold our dashboard's data integrity.

To test the dashboard performance and satisfaction, we ran a user interface and acceptance test. This test mimicked users' experience in navigating and using the dashboard. The tests include ease of navigation, clarity of visualisation, interactivity and responsiveness, accessibility, consistency in design and dashboard performance. As expected, we found several parts of the dashboard that can use some improvements to enhance user's experience. We noticed that the clarity of the visuals in terms of colours and orientation are a bit problematic. To address this, we use less colours to remove distractions and rearrange the visualisation using some of Gestalt's principles of visual perception (proximity and continuity). It is now easier to engage users and simultaneously capture their attention when the right aesthetics and affordances. Also, we also ensure that people from all walks of life can access and gain insights from the dashboard. This includes people of colour blindness, non-native English speakers, people of any age and technological proficiency. To summarize, user interface and acceptance test demonstrates a commitment to enhancing user satisfaction and engagement with the dashboard.

Bibliography

Human Rights Watch 2021, *Kenya: Survivors of Gender-Based Violence Lack Help,* accessed 7 April 2024, <https://www.hrw.org/news/2021/09/21/kenya-survivors-gender-based-violence-lack-help>

National Crime Research Centre 2014, *Gender Based Violence in Kenya,* accessed 25 Merch 2024, <https://crimeresearch.go.ke/wp-content/uploads/2018/02/wwwroot_publications_Gender-Based-Violence-in-Kenya.pdf>

Lawal, S 2024, *Femicide in Kenya: What's causing an epidemic of violence against women?*, Aljazeera, 27 January 2024, accessed 7 April 2024, <https://www.aljazeera.com/news/2024/1/27/femicide-in-kenya-whats-causing-the-frequent-murders-of-women>

United Nations Development Group 2017, *Data Privacy, Ethics and Protection: Guidance Note on Big Data for Achievement of the 2030 Agenda,* accessed 12 April 2024, <https://unsdg.un.org/resources/data-privacy-ethics-and-protection-guidance-note-big-data-achievement-2030-agenda>

World Health Organization 2007, *WHO ethical and safety recommendations for researching, documenting and monitoring sexual violence in emergencies*, accessed 14 April 2024, <https://www.who.int/publications/i/item/9789241595681>

People In Need 2022, *Rapid Guide to Collecting Data on Gender Based Violence,* IndiKit, accessed 14 April 2024, <https://www.indikit.net/userfiles/files/Rapid%20Guide%20to%20Collecting%20Data%20on%20Gender%20Based%20Violence_FINAL.pdf>

Appendix

Appendix A: DAX Measures

Following is a table indicating DAX measures:

Table Name	Measure Name	Description	Measure Expression
Case_DT	Case Percentage	This measure calculates the percentage of cases in the GBV_FT table compared to all cases in the GBV_FT table, disregarding any filters applied on the GBV_FT table.	DIVIDE(CALCULATE(COUNT(GBV_FT[Case_ID])), CALCULATE(COUNT(GBV_FT[Case_ID]), ALL(GBV_FT)))
Case_DT	Current Year (2018)	This measure calculates the count of distinct case IDs in the 'Case_DT' table for the year 2018 based on the intake date.	CALCULATE(DISTINCTCOUNT('Case_DT'[Case_ID]), FILTER('Case_DT', YEAR('Case_DT'[Intake date]) = 2018))
Case_DT	Previous Year	This measure calculates the count of distinct case IDs in the 'Case_DT' table for the year preceding the current year, based on the intake date.	CALCULATE(DISTINCTCOUNT('Case_DT'[Case_ID]), DATEADD('Case_DT'[Intake date].[Date],-1,YEAR))
Case_DT	Total Case	This measure calculates the total count of case IDs in the GBV_FT table.	CALCULATE(COUNT(GBV_FT[Case_ID]))
Case_DT	YoY Growth by Category	This measure computes the year-over-year growth rate by category. It compares the count of cases in the current year (2018) to the count of cases in the previous year (2017) for each presenting case category.	VAR CurrentYear = YEAR(MAX('Case_DT'[Intake Date])) VAR CurrentYearCounts = SUMMARIZE(FILTER('Case_DT', YEAR('Case_DT'[Intake Date]) = CurrentYear), 'Case_DT'[Presenting Case Category], "CurrentYearCount", DISTINCTCOUNT('Case_DT'[Case_ID])) VAR PreviousYearCounts = SUMMARIZE(FILTER('Case_DT', YEAR('Case_DT'[Intake Date]) = CurrentYear - 1), 'Case_DT'[Presenting Case Category], "PreviousYearCount",

			DISTINCTCOUNT('Case_DT'[Case_ID])) RETURN DIVIDE(SUMX(CurrentYearCounts, [CurrentYearCount]) - SUMX(PreviousYearCounts, [PreviousYearCount]), SUMX(PreviousYearCounts, [PreviousYearCount]))
Case_DT	YoY_Growth_Rate	This measure calculates the year-over-year growth rate by subtracting the count of cases in the previous year (2017) from the count of cases in the current year (2018) and then dividing by the count of cases in the previous year (2017)	DIVIDE([Current Year (2018)] - [Previous Year], [Previous Year])
Victim_DT	LeadingAgeGroupByTotalCases	This measure identifies the age group with the highest total number of cases in the Victim_DT table and returns it as the leading age group.	VAR AgeGroupCounts = SUMMARIZE ('Victim_DT', 'Victim_DT'[Age groups], "TotalCases", [Total Case]) VAR MaxCases = MAXX(AgeGroupCounts, [TotalCases]) RETURN SELECTCOLUMNS (FILTER (AgeGroupCounts, [TotalCases] = MaxCases), "Leading Age Group", 'Victim_DT'[Age groups])
Victim_DT	TotalVictims	This measure calculates the total count of non-blank victim IDs in the Victim_DT table, indicating the total number of victims	CALCULATE (COUNTROWS(Victim_DT), FILTER (Victim_DT, NOT(ISBLANK(Victim_DT[Victim_ID]))))

Appendix B: Data ETL Process

Following is a table indicating the categorical errors found in the dataset and the rationale for each method of handling to support data integrity. Contextual information was inferred through data dictionary assumptions and research. Power BI functions like 'Promote Headers' and sorting were also used to organise and profile the data.

Issues	Action	Rationale
Duplicate Values	• It would be removed if the entire row was duplicated • For unique categories like Case_ID, the column was replaced.	The data engineer thoroughly checked if there were duplicate rows. However, the only important duplicated value was in the Case_ID column, where there were two case 53s. All the data in the rows were identical to each other except for the age. Our team discussed and assessed if the rows were caused by a potential data entry error in age or case_ID, or from a database issue. However, due to limited background information, we concluded the importance of maintaining the data for analysis quality and assumed they were valid data. However, since the Case ID was intended to be a primary key in the data modelling process and as it did not contribute important relational information, the column was removed. A key component of primary keys is unique identifiers, otherwise, the data modelling would become complex and inefficient for Power BI through many-to-many relationships. Hence, the column was replaced by an index column incrementing from 1 to uniquely assign each case in preparation for normalisation
Empty Rows	• Removed	Rows that were empty for all columns except one, e.g. Case_ID, were eradicated due to redundancy as they contained no valuable information.
Missing Values	• Replace null data for 'Text' data type columns through the 'Replace Value' option in Power Query Editor, with what was contextually appropriate on a category-by-category basis.	As we aimed to remove as little data as possible, the null text values were replaced with values that suited the survey context from data collation. For example, null values in; • Presenting Case & Presenting Case Category, was replaced with *'Others'* • Reported by, Method of Reporting, Witness, Witness statement, Residential Area & Informal Settlement, was replaced with *'Unknown'*

	• Missing values for numerical data type columns were addressed through forward imputation or imputation with median value where necessary.	• Age, Sex & Age Groups, was replaced by *'Prefer not to say'* Forward imputation was generally used through the 'Fill Up/Down' function for non-text data to maintain data quality and maximise the amount of data usable, E.g. Intake Date. Median imputation was not used with age, despite it previously being a numerical data type, due to concerns of skewed data from outlier impacts. For longitude and Latitude, if the associated informal settlement was *'Unknown'* then the space was also imputed with *'Unknown'*. Otherwise, it was filled with the median coordinates for Mathare as Kibera had no missing coordinate data.

Appendix C – Testing Table

Page	Section	Test Case	Dashboard Output	Excel Output	Outcome
Overview	Total Case card	Verify that the total cases value is correct.	1116 cases in total.	1116 cases in total.	PASS
	Growth Rate in GBV Cases (2017-2018) card	Verify that the growth rate in GBV cases is correct.	50%	50%	PASS
	Month with Top Cases card	Verify that the month with the most cases is correct.	January – 137 January (2017) – 84 June (2018) - 61	January – 137 January (2017) – 84 June (2018) - 61	PASS
	Total Witness card	Verify that the total witness value is correct.	31	31	PASS
	Leading Age Group by Total Cases	Verify that the age group is correct	25-34	25-43	PASS
	Leading Gender card	Verify that the most affected gender is female and the value displayed is correct.	Female-identifying victims make up 73.84% of the total cases.	Female-identifying victims make up 73.84% of the total cases.	PASS
	Leading Presenting Case Category horizontal bar chart	Verify that the leading presenting case category and the values are correct.	Domestic abuse (Female: 340, Male: 34)	Domestic abuse (Female: 340, Male: 34)	PASS
	Leading Residental Area horizontal bar chart	Verify that the leading residental area and the values are correct.	Gatwekerra (Female: 274, Male: 96)	Gatwekerra (Female: 274, Male: 96)	PASS
	Top 5 Presenting Case by Gender bar chart	Verify that the top 5 presenting case and the values of corresponding genders are correct.	Domestic violence (Female: 317, Male: 27), physical assault (Female: 170, Male: 95), negligence (Female: 136, Male: 112), defilement (Female: 111, Male: 0), juvenile (Female: 16, Male: 31).	Domestic violence (Female: 317, Male: 27), physical assault (Female: 170, Male: 95), negligence (Female: 136, Male: 112), defilement (Female: 111, Male: 0), juvenile (Female: 16, Male: 31).	PASS

Page	Section	Test Case	Dashboard Output	Excel Output	Outcome
Incident Analysis	Total Case card	Verify that the total cases value is correct.	1116	1116	PASS
	Proportion of Female in GBV card	Verify that the proportion is correct.	73.84%	73.84%	PASS
	Proportion of Male in GBV card	Verify that the proportion is correct.	25.81%	25.81%	PASS
	Month with Top Cases	Verify that the month and the corresponding value are correct.	January – 137 January (2017) – 84 June (2018) - 61	January – 137 January (2017) – 84 June (2018) - 61	PASS
	Distribution of GBV Cases by Age Group	Verify that the bar chart shows the right legends.	1-10, 11-17, 18-24, 25-34, 35-44, 45-59, 60+	1-10, 11-17, 18-24, 25-34, 35-44, 45-59, 60+	PASS
	Distribution of GBV Cases by Age Group	Verify that the bar chart shows the correct values.	1-10 (Male: 16, Female: 20), 11-17 (Male: 7, Female: 20), 18-24 (Male: 1, Female: 25), 25-34 (Male: 3, Female: 16), 35-44 (Male: 3, Female: 7), 45-59 (Male: 1, Female: 2)	1-10 (Male: 16, Female: 20), 11-17 (Male: 7, Female: 20), 18-24 (Male: 1, Female: 25), 25-34 (Male: 3, Female: 16), 35-44 (Male: 3, Female: 7), 45-59 (Male: 1, Female: 2)	PASS
	Distribution of Presenting Case Category by Gender	Verify that the bar chart shows the right legends.	Domestic abuse, child abuse, physical assault	Domestic abuse, child abuse, physical assault	PASS
	Distribution of Presenting Case Category by Gender	Verify that the bar chart shows the correct values.	Domestic abuse (Male: 17, Female: 190), child abuse (Male: 104, Female: 95), physical assault (Male: 60, Female: 104)	Domestic abuse (Male: 17, Female: 190), child abuse (Male: 104, Female: 95), physical assault (Male: 60, Female: 104)	PASS
	YoY Growth Presenting Case Category	Verify that the bar chart shows the right legends.	Child abuse, sexual abuse, physical assault, legal counsel, domestic abuse	Child abuse, sexual abuse, physical assault, legal counsel, domestic abuse	PASS
	YoY Growth Presenting Case Category	Verify that the bar chart shows the correct values.	Child abuse (87%), sexual abuse (79%), physical assault (49%), legal counsel (20%), domestic abuse (17%)	Child abuse (87%), sexual abuse (79%), physical assault (49%), legal counsel (20%), domestic abuse (17%)	PASS
	Distribution of Presenting Case by Gender	Verify that the bar chart shows the right legends.	Negligence, domestic violence, physical assault, defilement, rap, seek legal counsel, sexual violence	Negligence, domestic violence, physical assault, defilement, rap, seek legal counsel, sexual violence	PASS
	Distribution of Presenting Case Category by Gender	Verify that the bar chart shows the correct values.	Negligence (Male: 18, Female: 22), domestic violence (Male: 2, Female: 32), physical assault (Male: 10, Female: 20), defilement (Male: 0, Female: 12), rape (Male: 0, Female: 3), seek legal counsel (Male: 1, Female: 1), sexual violence (Male: 0, Female: 2)	Negligence (Male: 18, Female: 22), domestic violence (Male: 2, Female: 32), physical assault (Male: 10, Female: 20), defilement (Male: 0, Female: 12), rape (Male: 0, Female: 3), seek legal counsel (Male: 1, Female: 1), sexual violence (Male: 0, Female: 2)	PASS

Page	Section	Test Case	Dashboard Output	Excel Output	Outcome
Location Analysis	Top 4 GBV Cases Among Males by Residental Area	Verify that the donut chart shows the right legends.	Gatwekera, Mathare 4B, Kisumu Ndogo, Kosovo	Gatwekera, Mathare 4B, Kisumu Ndogo, Kosovo	PASS
	Top 4 GBV Cases Among Males by Residental Area	Verify that the donut chart shows the correct values.	Gatwekera (59.63%), Mathare 4B (19.25%), Kisumu Ndogo (10.56%), Kosovo (10.56%)	Gatwekera (59.63%), Mathare 4B (19.25%), Kisumu Ndogo (10.56%), Kosovo (10.56%)	PASS
	Top 4 GBV Cases Among Females by Residental Area	Verify that the donut chart shows the right legends.	Gatwekera, Mathare 4B, Kosovo, Lindi	Gatwekera, Mathare 4B, Kosovo, Lindi	PASS
	Top 4 GBV Cases Among Females by Residental Area	Verify that the donut chart shows the correct values.	Gatwekera (62.99%), Mathare 4B (15.17%), Kosovo (11.03%), Lindi (10.08%)	Gatwekera (62.99%), Mathare 4B (15.17%), Kosovo (11.03%), Lindi (10.08%)	PASS
	Total Case in Top 3 Residental Area by Age Groups	Verify that the bar chart shows the right legends.	Gatwekera, Mathare 4B, Kosovo	Gatwekera, Mathare 4B, Kosovo	PASS
	Total Case in Top 3 Residental Area by Age Groups	Verify that the bar chart shows the correct values.	Gatwekera (1-10: 73, 11-17: 59, 18-24: 64, 25-34: 100, 35-44: 31, 45-59: 15, 60+: 1), Mathare 4B (1-10: 24, 11-17: 19, 18-24: 27, 25-34: 22, 35-44: 4), Kosovo (1-10: 17, 11-17: 15, 18-24: 13, 25-34: 18, 35-44: 2)	Gatwekera (1-10: 73, 11-17: 59, 18-24: 64, 25-34: 100, 35-44: 31, 45-59: 15, 60+: 1), Mathare 4B (1-10: 24, 11-17: 19, 18-24: 27, 25-34: 22, 35-44: 4), Kosovo (1-10: 17, 11-17: 15, 18-24: 13, 25-34: 18, 35-44: 2)	PASS

Page	Section	Test Case	Dashboard Output	Excel Output	Outcome
Reporting Insights	Distribution of Reporting Method	Verify that the donut chart shows the right legends.	Walk-in, screening, hotline, champion	Walk-in, screening, hotline, champion	PASS
	Distribution of Reporting Method	Verify that the donut chart shows the correct values.	Walk-in (52.08%), screening (22.26%), hotline (20.66%), champion (4.9%)	Walk-in (52.08%), screening (22.26%), hotline (20.66%), champion (4.9%)	PASS
	Presence of Witness	Verify that the donut chart shows the right legends.	No, Yes	No, Yes	PASS
	Presence of Witness	Verify that the donut chart shows the correct values.	No (97.19%), Yes (2.81%)	No (97.19%), Yes (2.81%)	PASS
	Distribution of Gender When Witness is Present based on Top 4 Presenting Case Category	Verify that the bar chart shows the right legends.	Child abuse, physical assault, domestic abuse, sexual abuse	Child abuse, physical assault, domestic abuse, sexual abuse	PASS
	Distribution of Gender When Witness is Present based on Top 4 Presenting Case Category	Verify that the bar chart shows the correct values.	Child abuse (Male: 4, Female: 8), physical assault (Male: 5, Female: 5), domestic abuse (Male: 0, Female: 5), sexual abuse (Male: 0, Female: 3)	Child abuse (Male: 4, Female: 8), physical assault (Male: 5, Female: 5), domestic abuse (Male: 0, Female: 5), sexual abuse (Male: 0, Female: 3)	PASS
	Report Entity by Top 4 Presenting Case Category When Witness is Present	Verify that the bar chart shows the right legends.	Domestic abuse, child abuse, physical assault, sexual abuse	Domestic abuse, child abuse, physical assault, sexual abuse	PASS
	Report Entity by Top 4 Presenting Case Category When Witness is Present	Verify that the bar chart shows the correct values.	Domestic abuse (Victim: 317, Well-Wisher: 33), child abuse (Victim: 59, Well-Wisher: 218), physical assault (Victim: 219, Well-Wisher: 28), sexual abuse (Victim: 61, Well-Wisher: 75)	Domestic abuse (Victim: 317, Well-Wisher: 33), child abuse (Victim: 59, Well-Wisher: 218), physical assault (Victim: 219, Well-Wisher: 28), sexual abuse (Victim: 61, Well-Wisher: 75)	PASS
	Distribution of Age Groups based on Top 3 Method of Reporting	Verify that the bar chart shows the right legends.	Walk-in, screening, hotline	Walk-in, screening, hotline	PASS
	Distribution of Age Groups based on Top 3 Method of Reporting	Verify that the bar chart shows the correct values.	**Walk-in** (1-10: 82, 11-17: 84, 18-24: 111, 25-34: 136, 35-44: 44, 45-59: 16, 60+: 1), **screening** (1-10: 23, 11-17: 22, 18-24: 59, 25-34: 70, 35-44: 20, 45-59: 8, 60+: 3), **hotline** (1-10: 63, 11-17: 46, 18-24: 35, 25-34: 27, 35-44: 3, 45-59: 1)	**Walk-in** (1-10: 82, 11-17: 84, 18-24: 111, 25-34: 136, 35-44: 44, 45-59: 16, 60+: 1), **screening** (1-10: 23, 11-17: 22, 18-24: 59, 25-34: 70, 35-44: 20, 45-59: 8, 60+: 3), **hotline** (1-10: 63, 11-17: 46, 18-24: 35, 25-34: 27, 35-44: 3, 45-59: 1)	PASS
	Distribution of Method of Reporting based on Top 3 Residental Areas	Verify that the bar chart shows the right legends.	Gatwekera, Kisumu Ndogo, Lindi	Gatwekera, Kisumu Ndogo, Lindi	PASS
	Distribution of Method of Reporting based on Top 3 Residental Areas	Verify that the bar chart shows the correct values.	Gatwekera (Champion: 22, Clinic screening: 56, Hotline: 65, Walk-in: 204), Kisumu Ndogo (Champion: 0, Clinic screening: 14, Hotline: 6, Walk-in: 36), Lindi (Champion: 1, Clinic screening: 20, Hotline: 5, Walk-in: 30)	Gatwekera (Champion: 22, Clinic screening: 56, Hotline: 65, Walk-in: 204), Kisumu Ndogo (Champion: 0, Clinic screening: 14, Hotline: 6, Walk-in: 36), Lindi (Champion: 1, Clinic screening: 20, Hotline: 5, Walk-in: 30)	PASS

Page	Test Case	**Test Case**	**Comment**	**Outcome**
Overall Dashboard	Ease of Navigation	Test whether users can easily navigate to different sections of the dashboard using navigation menus or buttons. "Given that I want to go to other pages, when I click on the navigator, then I expect it to take me there."	On the left-hand side of the Overview page, there are four-page navigators labelled 'Overview,' 'Incident Analysis,' 'Location Analysis' and 'Reporting Insights.' Contrasting colours are used to help users navigate easier towards the page navigators. The layout is also very intuitive even for new users.	PASS
	Clarity of Visual Representation	Assess the clarity and effectiveness of the visual elements used in the dashboard, such as charts, graphs, and maps. Ensure that the data is presented in a clear and understandable manner, with appropriate use of colours, labels, and legends. "Given that I want to see key information, when I look at the dashboard, then I am able to tell the important information with little visual processing."	Cards are used to display key information and capture users attention. Selection of bar charts are also used to represent key distributions and values. Map is included to get users to locate cases and connect with the demography. Contrasting colours are used throughout the dashboard. White canvas is used to accentuate the visualisation and navy blue is used around its borders. Labels and legends are used on every visualisation to ensure clear presentation of information.	PASS
	Interactivity and Responsiveness	Test how well users can interact with the dashboard elements, such as filtering data, drilling down into details, and exploring different views. "Given that I want to filter the data, when I click on the slicer, then I expect it to filter the visualisation accordingly."	Slicers are provided for users to interact with the visualisation in real-time. The slicer will interact dynamically and give users new insightful and detailed information.	PASS
	Accessibility	Check the accessibility of the dashboard for users with low technological ability. Ensure that the dashboard is compatible, and that it adheres to accessibility standards, such as WCAG (Web Content Accessibility Guidelines). "Given that I am new to the dashboard with little technological proficiency, when I use the dashboard, then I expect that I will be able to gain as many benefits as other more technologically proficient and experienced users."	The dashboard adheres to the components emphasised in the Web Content Accessibility Guidelines. 1) Perceivable: With the assistance of page navigators, the dashboard becomes easy to navigate for someone who is new to the website. 2) Operable: For people of all background and ability. Help button is also provided to allow users to get assistance. 3) Understandable: Clear language is used and dashboard glossary is also included.	PASS
	Consistency in Design	Evaluate the consistency of design elements and layout across different parts of the dashboard. Check if there is consistency in the use of colours, fonts, icons, and spacing to create a cohesive visual experience. "Given that I am new to the concept of dashboards, when I use this dashboard, I expect to the information to flow consistently."	A consistent layout is used throughout the whole dashboard. The 'Overview' page is slightly different to give variety to the dashboard. The colour palette and fonts used are highly consistent throughout the dashboard.	PASS
	Performance	Measure the performance of the dashboard in terms of loading times and responsiveness. "Given that I am in an urgent need of the data, when I use the dashboard, I expect it to quickly process the information and not crash."	The visualisation loads properly and is tested multiple times across multiple devices that are Power BI compatible.	PASS

Appendix D: Epics and RICE Method Prioritisation for Features and Functionalities

- Overview page (to set the context for the entire dashboard)

Epic: As the CEO of KenyaPulse, I want to have an overview page on our Power BI dashboard so that I can quickly grasp the key insights and trends related to gender-based violence (GBV) cases in Kenya.
Prioritisation using RICE Method: Reach: High reach since it directly impacts the CEO of KenyaPulse Impact: Significant impact Confidence: High confidence Effort Estimation: Moderate to High

- Incident Analysis page

Epic: As the CEO of KenyaPulse, I want to have an incident analysis page on our Power BI dashboard so that I can delve deeper into the specifics of gender-based violence (GBV) incidents, understand the trends over time, and identify areas for targeted interventions.
Prioritisation using RICE Method: Reach: High reach since it directly impacts the CEO of KenyaPulse Impact: Significant impact Confidence: High confidence Effort Estimation: High

- Location Analysis page

Epic: As the CEO of KenyaPulse, I want to have a location analysis page on our Power BI dashboard so that I can gain insights into the geographical distribution of gender-based violence (GBV) incidents, identify hotspots, and tailor our interventions to specific areas effectively.
Prioritisation using RICE Method: Reach: High reach since it directly impacts the CEO of KenyaPulse Impact: Significant impact Confidence: High confidence Effort Estimation: High

- Reporting Analysis page

Epic: As the CEO of KenyaPulse, I want to have a reporting analysis page on our Power BI dashboard so that I can gain comprehensive insights into the reporting dynamics of gender-based violence (GBV) incidents, understand the demographics involved, and

identify patterns in reporting methods and witness statements for informed decision-making and strategic planning.
Prioritisation using RICE Method: Reach: High reach since it directly impacts the CEO of KenyaPulse Impact: Significant impact Confidence: High confidence Effort Estimation: High

- Functional Slicers

Epic: As the CEO of KenyaPulse, I want to have functional slicers on our Power BI dashboard so that users can filter the data dynamically, enabling them to focus on specific subsets of information and derive actionable insights efficiently.
Prioritisation using RICE Method: Reach: Moderately high Impact: Significant impact Confidence: High confidence Effort Estimation: Moderate

- Buttons

Epic: As the CEO of KenyaPulse, I want to have buttons on our Power BI dashboard so that users can easily navigate between different report pages, facilitating seamless exploration of various insights and analyses related to gender-based violence (GBV) incidents and our organization's efforts to address them effectively.
Prioritisation using RICE Method: Reach: Moderately high Impact: Significant impact Confidence: High confidence Effort Estimation: Low to moderate

- Clear filter

Epic: As the CEO of KenyaPulse, I want to have a "Clear Filter" functionality on our Power BI dashboard so that users can remove all applied filters from the report easily, ensuring a comprehensive view of gender-based violence (GBV) data and allowing for more accurate analysis and decision-making.
Prioritisation using RICE Method: Reach: Moderately high Impact: Significant impact

Confidence: High confidence Effort Estimation: Low to moderate

- Cards

Epic: As the CEO of KenyaPulse, I want to utilize cards in our Power BI dashboard to display single numeric values or simple text descriptions so that I can quickly grasp key metrics and trends related to gender-based violence (GBV) incidents in Kenya.
Prioritisation using RICE Method: Reach: High reach since it directly impacts the CEO of KenyaPulse Impact: Significant impact Confidence: High confidence Effort Estimation: Moderate

- Filter of Month with top cases according to year

Epic: As the CEO of KenyaPulse, I want to have a filter for the month with top cases according to the selected year on our Power BI dashboard so that I can easily identify and analyze the months with the highest number of reported gender-based violence (GBV) cases for each year, enabling targeted interventions and resource allocation.
Prioritisation using RICE Method: Reach: Moderate Impact: Significant impact Confidence: High confidence Effort Estimation: Moderate

Appendix E: Markers (Rough Key Timelines)

	Feb-24	Mar-24	Apr-24
Overview page		Visualisation Generation	
			Data Accuracy Testing, UI Testing
Incident Analysis page	Visualisation Generation		
			Data Accuracy Testing, UI Testing
Location Analysis page	Visualisation Generation		
			Data Accuracy Testing, UI Testing
Reporting Analysis page	Visualisation Generation		
			Data Accuracy Testing, UI Testing
Functional Slicers		Adding slicers and testing	
Buttons		Adding buttons and testing	
Clear filter		Adding clear filters and testing	
Cards		Generate measures and testing	
Filter of Month with top cases according to year		Adding filters and testing	