

Hospital Visits Analysis Report

Reporting Period: Jan 2023 to Dec 2025.

Prepared by: Sharon Nyabuto

Note: This analysis uses a synthetic dataset. It should not be used to draw conclusions about any actual facility, patient, or health system.

Summary

The dataset covers ~84,000 visits by ~58,000 patients from 2023 to 2025, with demographics, vitals, diagnosis, admission details, billing, insurance, and discharge outcome recorded per visit.

Admission and mortality were more strongly associated with clinical factors than with demographic or access-related factors. Diagnosis category showed the strongest association with admission (Cramér's $V = 0.60$), while triage level had the strongest association with mortality ($V = 0.24$). Insurance status, distance to the hospital, payment method, and gender showed negligible associations with cost or patient outcomes. Age was also weakly associated with mortality ($r = -0.04$).

Visit volume rose sharply from June 2024 and stayed elevated, matching the period of unrest around Kenya's Finance Bill protests, with a corresponding rise in trauma admissions and mortality rates. A small set of chronic conditions (led by HIV/AIDS at a 38% repeat-visit rate) accounts for most of the hospital's repeat care.

1. Data and Method

Data quality issues corrected included: inconsistent date formats, outlier vital signs, inconsistent labels and missing values. Missingness in admission dates, discharge dates, and ward information was highest among outpatient visits (~71%) and was considered systematic rather than a data quality issue, as these fields are not applicable to most outpatient visits.

Bill amount, length of stay, and SHA-covered amount are all heavily right-skewed by a small number of very costly or very long visits. Weight is the only variable close to a normal distribution. Because of this skew, the analysis uses methods suited to non-normal data throughout: Spearman rather than Pearson correlation, and Mann-Whitney U and Kruskal-Wallis tests rather than t-tests and ANOVA. Categorical variable associations are explored using the Chi-Square test of independence (χ^2).

Given the large sample size, p-values were not interpreted in isolation, as even small associations may reach statistical significance. Effect sizes, including Cramér's $V(V)$, rank-biserial $r(r_{rb})$, and epsilon-squared (ϵ^2) were therefore reported alongside p-values to assess the practical significance of observed associations. For correlation analyses, Spearman's $\rho(\rho)$ was used, with the squared correlation coefficient (R^2) reported as an indication of the proportion of variation associated with the relationship. Terms and symbols are defined in [Appendix A](#). Full statistics for every test are in [Appendix B](#).

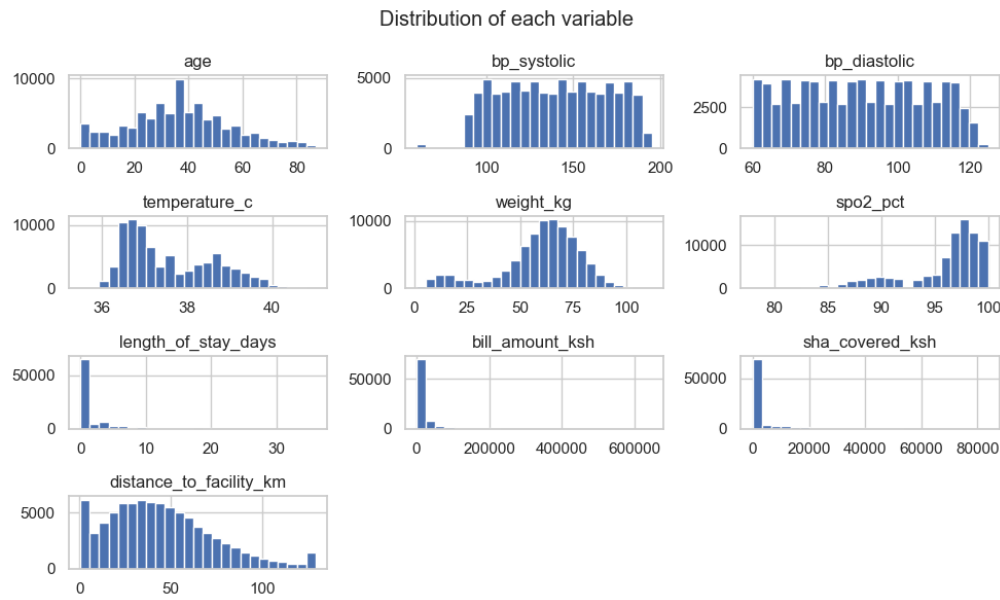


Figure 1: Numeric variable Distributions

2. Admission and Mortality Track Severity

Diagnosis category is the strongest predictor of admission status found in this analysis, by a wide margin ($V = 0.60$, a large effect). This relationship is partly structural, as some diagnoses occur exclusively among inpatients or outpatients, but it still reflects the expected pattern of more severe conditions requiring admission.

Triage level showed a moderate association with mortality ($V = 0.24$), while department had a weaker association with an effect size of 0.13. This reflects that some departments simply handle more critical cases than others.

Age does not predict mortality in any useful way. Median age was 37 for patients who died and 36 for those who survived, a gap too small to be practically meaningful, even though it is statistically significant. Overall, clinical severity at presentation was much more strongly associated with admission and mortality than age.

3. Cost and Coverage Follow the Bill, Not the Stay

Length of stay and bill amount are strongly correlated (Spearman $r = 0.70$): more days admitted means more charges, as expected. Bill amount and SHA-covered amount are moderately correlated ($r = 0.42$), since SHA coverage is calculated from the bill. Length of stay and SHA coverage on their own show a weaker correlation ($r = 0.30$), but this is a side effect, not a direct link. Controlling for bill amount, the partial correlation between length of stay and SHA coverage drops to 0.018, close to zero. Coverage is set by cost, not by how long a patient stays.

4. Access and Payment Show Almost No Effect

Four separate tests, on insurance, distance, payment method, and gender proved none to meaningfully predict cost or outcome once the obvious structural effects are removed. Insured and uninsured patients are billed similar amounts (median KSh 2,476 vs KSh 2,199; $r = -0.02$).

The gap is statistically detectable because of the large sample size, but is not meaningful in practice. Distance to the hospital has no relationship with mortality at all. Payment method, insurance-covered versus out-of-pocket, has no bearing on whether a patient is admitted or treated as an outpatient.

At first glance, gender is associated with the department a patient visits, albeit minimally ($V = 0.23$). Further inspection reveals the association comes almost entirely from Maternity, a department that is female-only by definition. Once it is removed, the effect drops further, to negligible ($V = 0.07$). The rest of the hospital is balanced by gender.

5. Repeat Care Concentrates in Chronic Disease

HIV/AIDS has a repeat-patient rate of 38.4%, roughly six times higher than any other diagnosis. The next tier, all between 5 and 7 percent, is entirely chronic or high-acuity disease.

Diagnosis	Repeat-patient rate
HIV/AIDS	38.36%
Chronic Kidney Disease Stage 5	6.79%
Chronic Kidney Disease Stage 3	6.36%
Pulmonary Tuberculosis	6.04%
Severe Pre-Eclampsia	6.04%

Repeat visits are observed in long-term disease management and not spread evenly across the hospital, which makes clinical sense.

6. Visit Volume Surge from Mid-2024/2025

Monthly visits held steady through 2023 and early 2024, at 1,600 to 1,850 a month. From June 2024, volume jumped past 4,100, the year's peak, and stayed above baseline through the rest of the year. A smaller spike repeated in June 2025. This period matches the civil unrest around Kenya's national protests in 2024 and 2025.

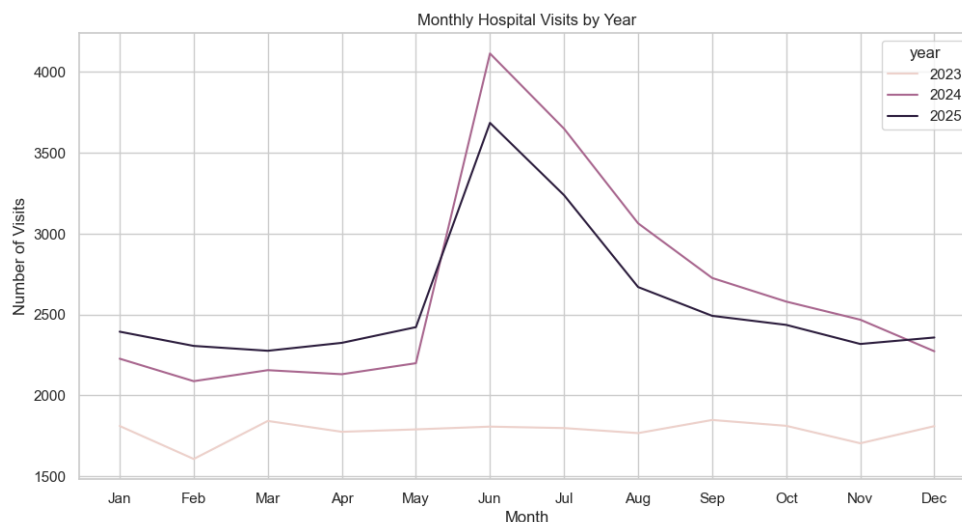


Figure 1.2: Visit distribution across reporting period

7. Doctor Workload and the Mid-2024 Surge

Doctor staffing stayed fixed while visit volume surged, and the impact was not even across departments. General Practice carries the heaviest overall workload (about 3,593 visits per doctor across the full period), but the departments hit hardest by the June 2024 surge were Surgery, ICU/HDU, and Emergency and Casualty, each seeing their per-doctor caseload rise 4.4 to 5.6 times over a typical 2023 month, compared with roughly 2.5 to 2.7 times for General Practice and TB/HIV Clinic.

8. Recommendations: What to do with this information

- Admission and mortality are driven by clinical need, not by insurance, distance, payment method, or gender. If a future audit shows this pattern shifting, for example insured patients starting to receive different care than uninsured patients with the same diagnosis, that is an early warning sign that would deserve attention.
- HIV/AIDS alone drives more repeat visits than the next several chronic conditions combined. Routing these patients through a dedicated chronic-care pathway, separate from general outpatient flow, would likely reduce congestion in General Practice and give HIV patients more consistent follow-up.
- The mid-2024 surge shows the hospital can face a near-doubling of monthly visits, with a heavier share of critical trauma cases, on short notice. A staffing and supply surge plan, tied to early warning signs of civil unrest and disease outbreaks would reduce the risk of the hospital being caught understaffed during the next such event.
- Because SHA coverage is driven by bill size and not length of stay, patients and staff should be told plainly that a longer admission does not by itself mean better coverage. This avoids confusion at discharge and billing.
- Build a targeted surge roster for Surgery, ICU/HDU, and Emergency and Casualty, whose per-doctor workload rose 4.4 to 5.6 times over baseline in June 2024, well above General Practice or TB/HIV Clinic. This would address the sharpest strain rather than spreading surge resources evenly.
- The features reported are sufficient to build a mortality risk model that would let triage staff flag high-risk patients earlier than triage level alone allows.

Appendix A: Terms and Symbols

Symbol / Term	Meaning
chi-square (χ^2)	Test statistic for comparing two categorical variables. Larger values mean a stronger deviation from what independence would predict.
df	Degrees of freedom. Depends on the number of categories being compared; used alongside the test statistic to judge significance.
p-value	The probability of seeing a result this strong if there were truly no relationship. Below 0.05 is conventionally treated as statistically significant.
V (Cramer's V)	Effect size for chi-square tests. Ranges 0 to 1. Below 0.1 is treated as negligible in this report, 0.1 to 0.3 small, 0.3 to 0.5 moderate, 0.5 and above large.
U (Mann-Whitney U)	Test statistic comparing two independent groups on a variable that is not normally distributed.
r (rank-biserial correlation)	Effect size for Mann-Whitney U tests. Ranges -1 to 1. Below 0.1 is treated as negligible in this report, 0.1 to 0.3 small, 0.3 to 0.5 moderate, 0.5 and above large.
H (Kruskal-Wallis H)	Test statistic comparing three or more independent groups on a variable that is not normally distributed.
epsilon-squared (ϵ^2)	Effect size for Kruskal-Wallis tests. Below 0.01 is treated as negligible in this report, 0.01 to 0.06 small, 0.06 to 0.14 moderate, 0.14 and above large.
Spearman r	Correlation coefficient for non-normally distributed or ranked data. Ranges -1 to 1.
R ²	Squared correlation coefficient, reported alongside Spearman r as an effect size. Indicates the proportion of shared variation between two variables.
Partial correlation	Correlation between two variables after removing the shared influence of a third variable.
SHA	Social Health Authority, Kenya's national health coverage scheme.
KSh	Kenyan Shilling.

Appendix B: Full Statistical Results

Test	Variables	Statistic	p-value	Effect size	Result
Chi-square	Department × Died	$\chi^2 = 1372.07$ (df=10)	< 0.001	V = 0.13	No practical significance
Chi-square	Triage level × Died	$\chi^2 = 4940.03$ (df=4)	< 0.001	V = 0.24	Statistically significant, effect size small
Chi-square	Diagnosis category × Visit class	$\chi^2 = 30150.53$ (df=15)	< 0.001	V = 0.60	Statistically significant
Chi-square	Gender × Department (incl. Maternity)	$\chi^2 = 4585.26$	< 0.001	V = 0.23	Little practical importance
Chi-square	Gender × Department (excl. Maternity)	$\chi^2 = 336.28$	< 0.001	V = 0.07	Negligible effect size
Chi-square	Visit class × Payment group	$\chi^2 = 1.75 / 0.05$	0.42 / 0.82	V = 0.00	No statistical significance
Mann-Whitney U	Age × Died	U = 121,312,203	< 0.001	r = -0.04	Statistically significant, but effect size negligible
Kruskal-Wallis	Length of stay × Department	$\epsilon^2 = 0.62$	< 0.001	$\epsilon^2 = 0.62$	Statistically significant, effect size large
Kruskal-Wallis	Bill amount × Triage level	H = 1392.42	< 0.001	$\epsilon^2 = 0.017$	Statistically significant, effect size small
Mann-Whitney U	Bill amount × Insured/Uninsured	U = 792,552,287.5	< 0.001	r = -0.02	Statistically significant, but effect size negligible
Mann-Whitney U	Distance × Died	U = 115,338,311.5	0.31	r = 0.01	No statistical significance
Spearman correlation	Length of stay × Bill amount	r = 0.70	< 0.001	R ² = 0.49	Strong positive relationship

Report- KLSRH Visits Analysis

Spearman correlation	Bill amount × SHA covered	$r = 0.42$	< 0.001	$R^2 = 0.18$	Moderate positive relationship
Spearman correlation	Length of stay × SHA covered	$r = 0.30$	< 0.001	$R^2 = 0.09$	No practical significance
Partial correlation	Length of stay × SHA covered, controlling for bill	$r = 0.018$	< 0.001	$R^2 = 0.00$	Negligible effect size (confirms mediation)
Spearman correlation	Age × Weight	$r = 0.30$	< 0.001	$R^2 = 0.09$	Not practically significant