Unpacking Kenya's National Budget.

Analysis of MDAs – Ministries, Departments & Agencies

Data Insights to Power the Bottom-Up Transformation

The Big Numbers……

Kenya's 2024/25 Budget at a Glance
Ksh 3.9 Trillion – Total Budget Size

36%

Share allocated to debt servicing

14%

Share allocated to Education (Largest sectoral allocation)

12%

Allocation to infrastructure

7%

Allocation to health

Big Numbers, Bigger Questions

Kenya's national budget allocations are large on paper, but execution gaps remain. Recurring discrepancies between approved budgets and actual expenditures raise concerns about efficiency, governance, and value for money.

Budget Watch for 2024-25 (Parliament of Kenya) — states a Ksh 3.9 trillion size of the budget. Parliament of Kenya
How Govt plans to spend Ksh.3.9 trillion in 2024/25 budget" — confirms the 3.9 trillion figure. Citizen Digital

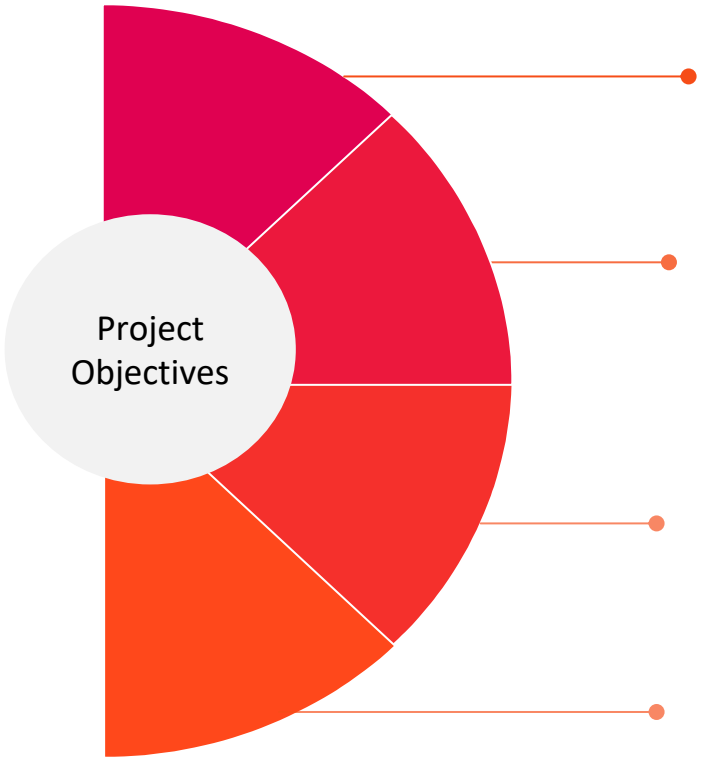

HARAMBEE

Problem

Big Allocations, Weak Execution

Kenya's public funds must be allocated and utilized transparently to ensure accountability, efficiency, and value for money. Single-year audits often miss deeper patterns such as persistent underspending, overspending, or repeated audit concerns.
This raises critical questions:
Are recurring discrepancies between approved budgets and actual expenditures signs of inefficiencies, weak execution, or potential audit risks?

This project investigates four financial years (2020/21 – 2023/24) of budget and expenditure data across national ministries, departments, and agencies (MDAs) to:



- Detect recurring gaps between allocations and expenditures
- Highlight systematic inefficiencies and weak budget executions
- Flag potential audit concerns and governance risks.
- Provide actionable insights for policymakers, oversight bodies, and civil society.

Budget Watch for 2024-25 (Parliament of Kenya) — states a Ksh 3.9 trillion size of the budget. Parliament of Kenya
How Govt plans to spend Ksh.3.9 trillion in 2024/25 budget" — confirms the 3.9 trillion figure. Citizen Digital

Data Understanding and Preparation

- This project draws on official Auditor-General reports, which provide the authoritative record of Kenya's national government budgets, expenditures, and audit observations.

Data Sources

- Auditor-General Reports (2020/21–2023/24): authoritative, audited actuals and observations.
- National Govt Budget "Blue Book" (2021/22): approved estimates at vote/program level: baselines for approved budgets across MDAS.
- Kenya_National_Govt_Budget_2021_2024.csv– Structured, machine-readable dataset (2021–2024).– Enables stats, joins, and year-to-year checks.
- MDAs Report 2023/24 (PDF)– Detailed vote-level audit opinions.– Includes "Budget vs Actual" statements and systemic issues (pending bills, late releases, control weaknesses).

Why Suitable

- Government-issued, official, and comprehensive
- Directly aligned with the problem: budget allocation vs execution

Data Understanding and Preparation

Data Preparation & Structure

Planned Extraction

- Used Python PDF tools (pdfplumber) to extract budget vs actuals, appropriation summaries, and control/performance notes.
- Normalized MDA names across years to account for mergers/renaming.

Build Unified Dataset

- Key fields: MDA, Financial year, approved budget, actual expenditure, variance, % variance, audit observations.
- Data Size: Several hundred rows (all MDAs × 4 years).

Descriptive Stats Computed

- Approved vs actual totals, averages, variance distributions.
- Frequency of underspending, overspending, or within tolerance.
- Audit observations categorized (procurement, unsupported spending, late releases).

Limitations & Mitigation

- Format inconsistencies: Hybrid extraction + manual cleaning.
- Name changes: Canonical mapping maintained.
- Accounting/reporting differences: Documented assumptions.
- Qualitative audit notes: Structured tags + confidence flags.

Data Cleaning for Financial Analysis and NLP

From Raw PDFS to clean Data

Raw Data Challenges

- Budget reports came in PDFs with messy tables & text.
- Inconsistencies in ministry names (mergers, abbreviations, duplicates).
- Budgets had commas, spaces, and text fragments instead of numbers.
- Noise rows (like “TOTAL” or “BALANCE”) distorted calculations.

Cleaning Actions Taken

- Standardized and normalized MDA names across years.
- Cleaned and converted budget figures into usable numeric format.
- Applied NLP to parse unstructured audit text, extract ministries, years, and amounts, and tag recurring audit concerns.
- Removed noise rows and impossible values (e.g., < KSh 1000).
- Aggregated duplicates to ensure one row per MDA per year.
- Created variance and % variance columns to enable comparability.

Outcome

- Built a unified, machine-readable dataset covering 4 years and 100+ MDAs.
- Dataset is now accurate, consistent, and ready for analysis.
- Allowed computation of core features (utilization rate, spending status, audit flag).

Additionally, we enhanced the dataset with new features such as Utilization Rate, Spending Status, and Audit Flags — enabling deeper insights into inefficiencies and risks

EDA

Budget Execution Patterns Across MDAs

- Majority of government entities are not effectively executing their budgets as planned. The significant underspending (50.5%) suggests potential inefficiencies in budget implementation, delayed projects, or poor financial planning. Overspending (33%) indicates possible budget miscalculations or unexpected expenditures.

On Budget:

16.4%

very few MDAs achieve execution discipline.

Overspent:

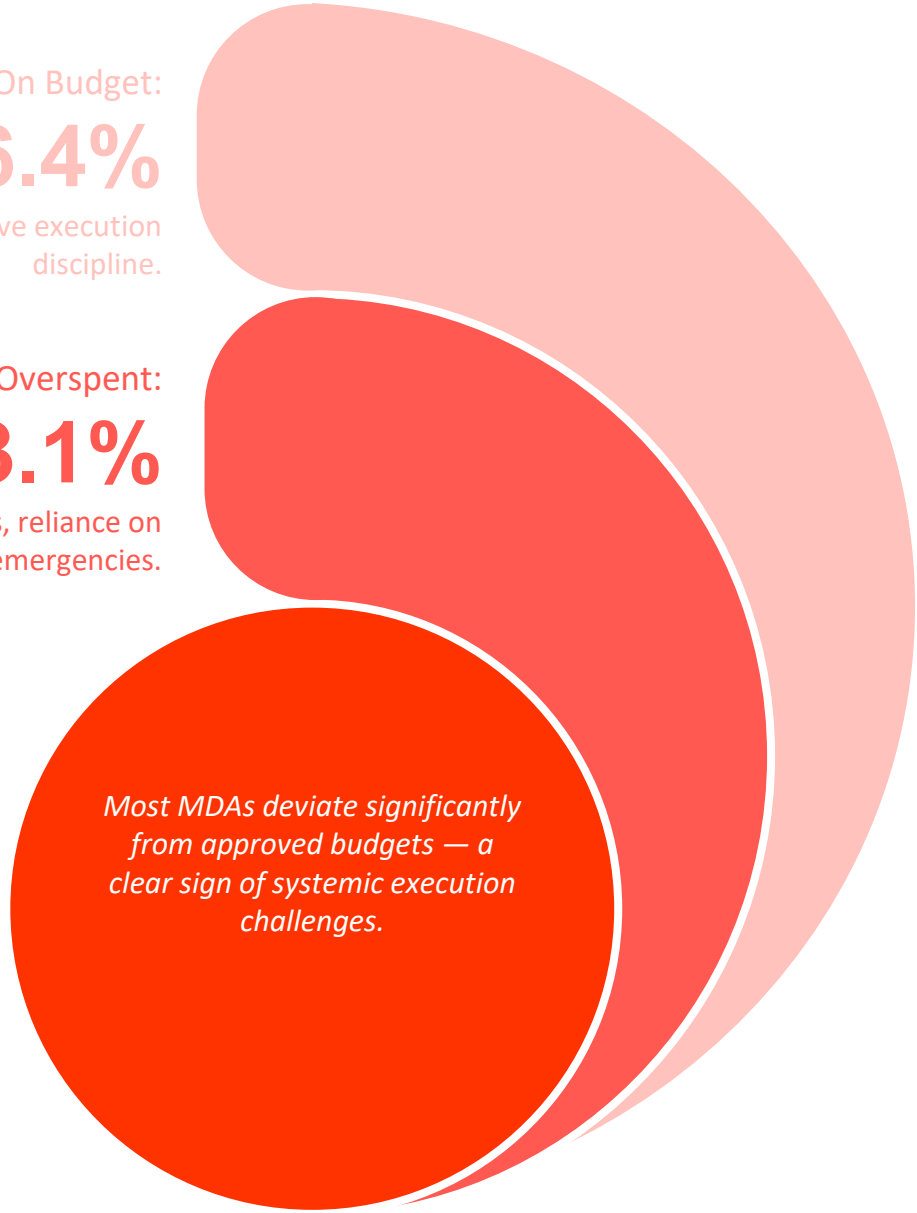
33.1%

possible budget miscalculations, reliance on supplementary budgets, or unplanned emergencies.

Underspent:

50.5%

inefficiencies in budget absorption, delayed projects, or weak financial planning.

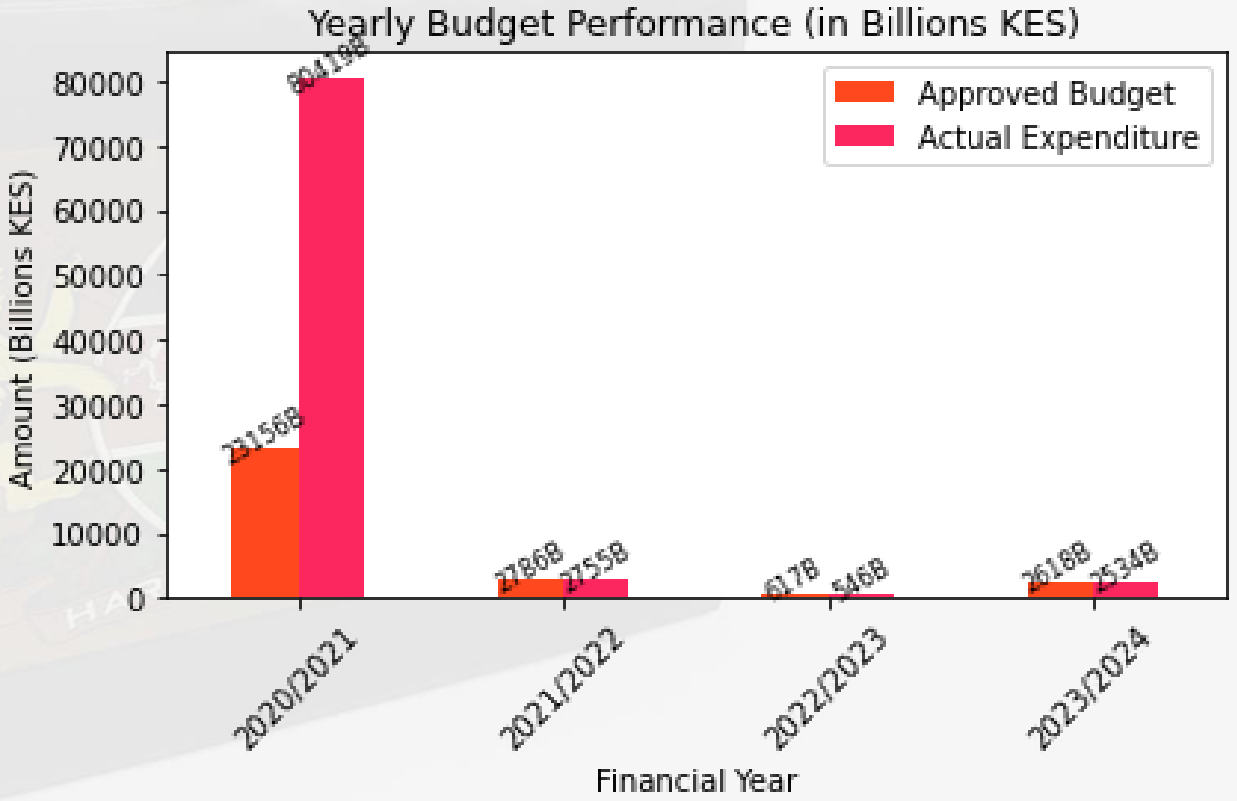


Yearly Budget Performance

- The persistent gap between approved and actual expenditures suggests systemic budget execution issues rather than one-time anomalies. The improvement in later years might indicate better budget management practices being implemented, though significant variances still exist.

Note:

- The National Treasury reported that the 2021/22 budget rose to KSh 3.03T from KSh 2.89T in 2020/21, largely for economic recovery and the Big Four Agenda.
- Development budget as % of GDP was projected to decline.
- The Ministry of Health received a significant increase in 2020/21 to address COVID-19 priorities.



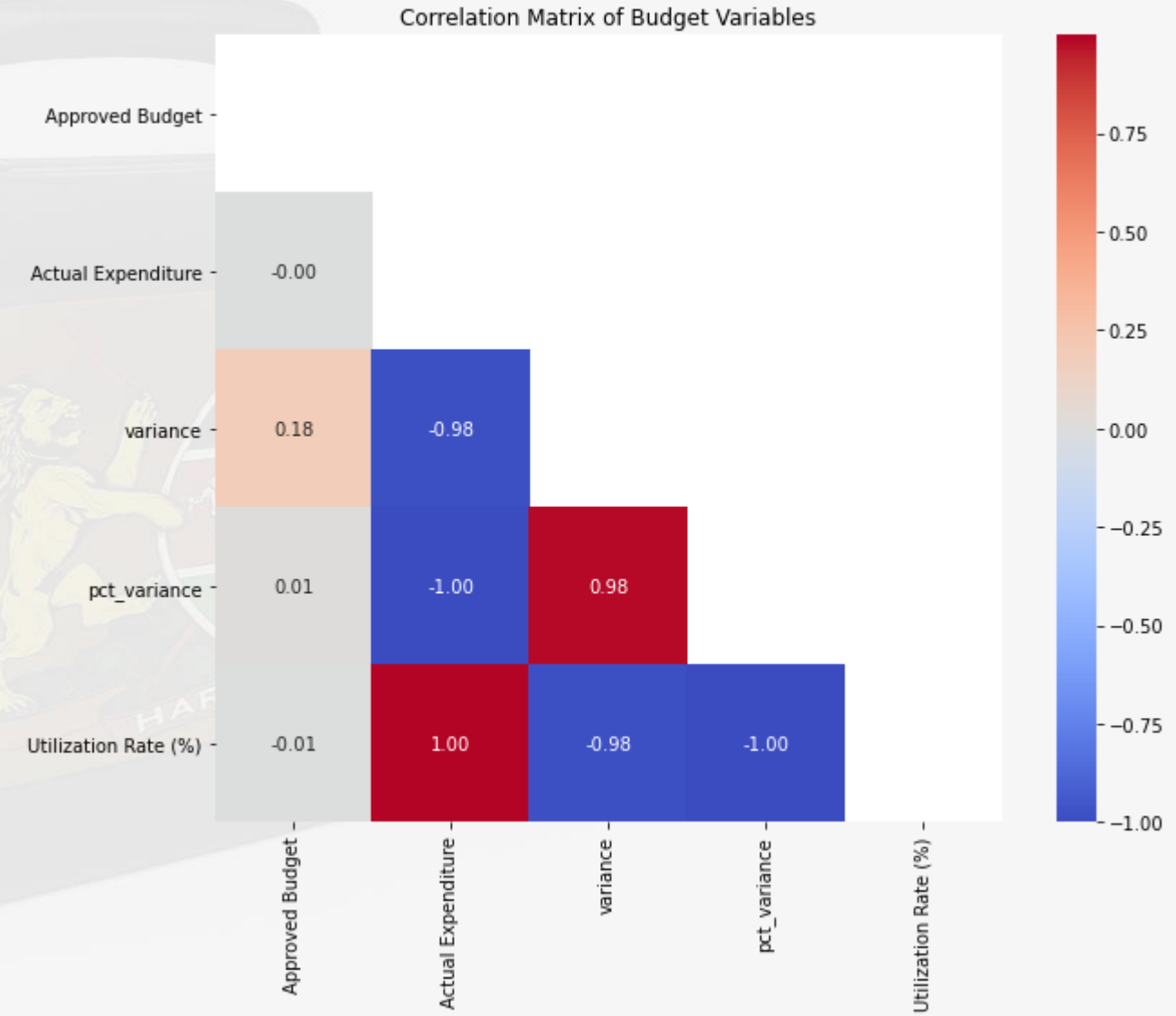
Correlation Between Budget Variables

Key Insights:

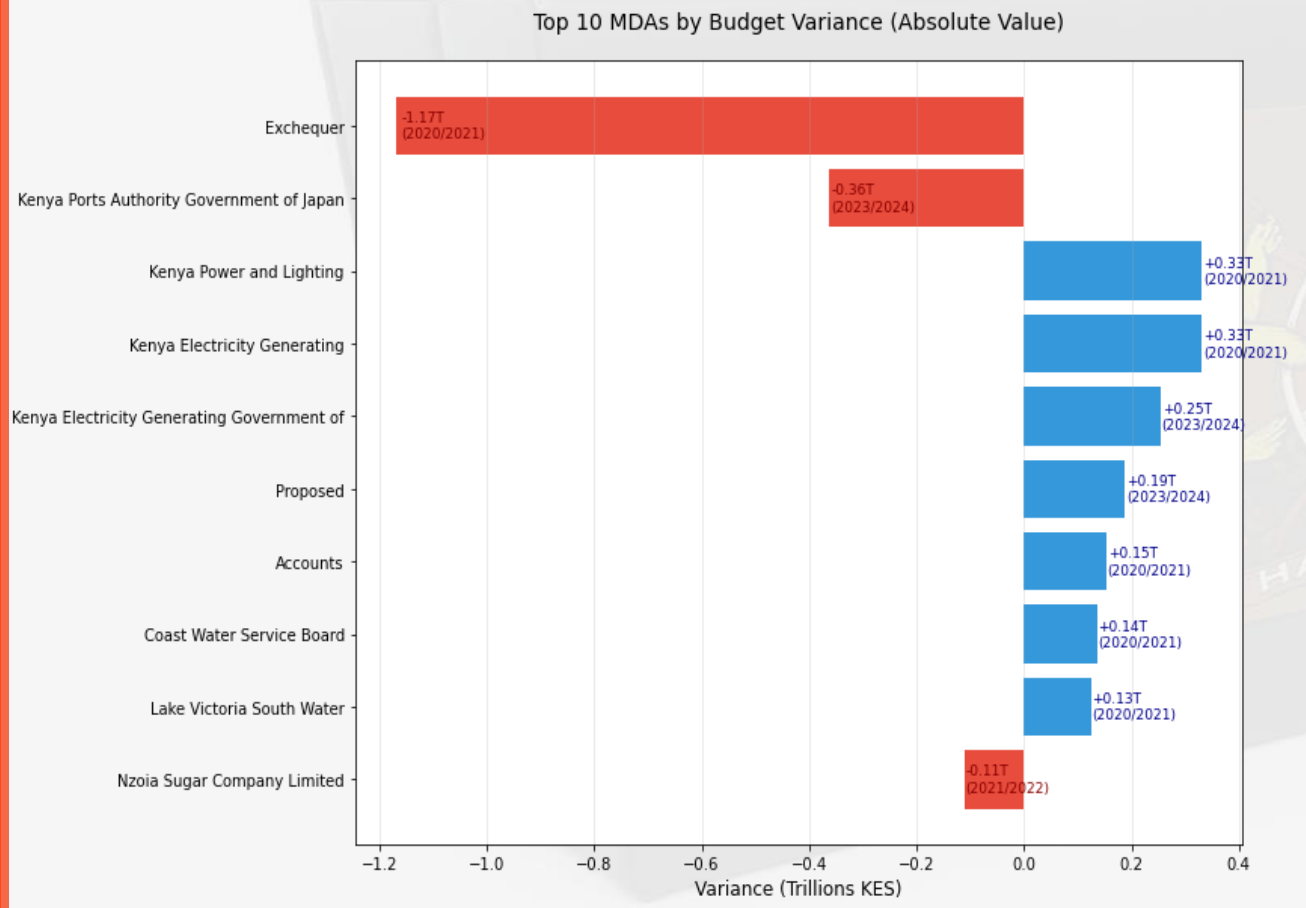
- Approved vs Actual Expenditure: –0.00
- Budget Size vs % Variance: 0.01
- Utilization vs Actual Expenditure: 1.00

Key Takeaways:

- No alignment between approved and actual spending.→ MDAs spend without regard to their approved budgets, signaling weak planning or mismanagement.
- Budget size doesn't matter.→ Both small and large MDAs struggle equally, proving this is a systemic issue.
- Utilization is perfectly tied to expenditure.→ This reflects a mathematical link, not behavior, and confirms that execution challenges are in allocations vs spending, not in calculations.



Top 10 MDAs by Budget Variance (2020–2024)



- The top 10 MDAs account for the largest budget deviations, either through massive underspending or extreme overspending. Together, these agencies explain the bulk of inefficiencies in Kenya's national budget execution

Key Insights:

- **Total underspending**: +2.12T went unused.

Suggests inefficiencies in budget absorption (projects not implemented, delayed, or over-allocated).

- **Total overspending:** -2.05T spent beyond approved budgets.

Indicates weak fiscal discipline and reliance on supplementary budgets

- **Largest underspending**: 0.33T (≈ 330B). The single biggest unutilized allocation by an MDA.

Likely a flagship ministry/project that stalled.

- **Largest overspending:** -1.17T (≈ 1.2T)

A single MDA overspent by almost one-third of Kenya's entire annual budget.

This is an extreme outlier → a must-flag for audit review.

Key Takeaway:

Kenya has a systemic problem:

Almost equal under- and overspending, but both at the trillion-shilling scale.

Very few MDAs hit their budget exactly.

Hypothesis Testing:

Spending Patterns Across MDAs

	Observed Distribution	Expected if balanced (~94 each):
Overspent	93	94
Underspent	142	94
On Budget	46	94

p-value: $2.07 \times 10^{-11} \rightarrow$ Reject H_0

Key Insights

- Since p-value $\ll 0.05$, we reject H_0.

This means the distribution of MDAs across Overspent / Underspent / On Budget is not equal. This reveals that there's a systematic bias:

- Most MDAs are underspending (142 vs expected ~94).
- Very few MDAs are On Budget (46 vs expected ~94).
- Overspent MDAs (93) are close to expected but still part of the imbalance.

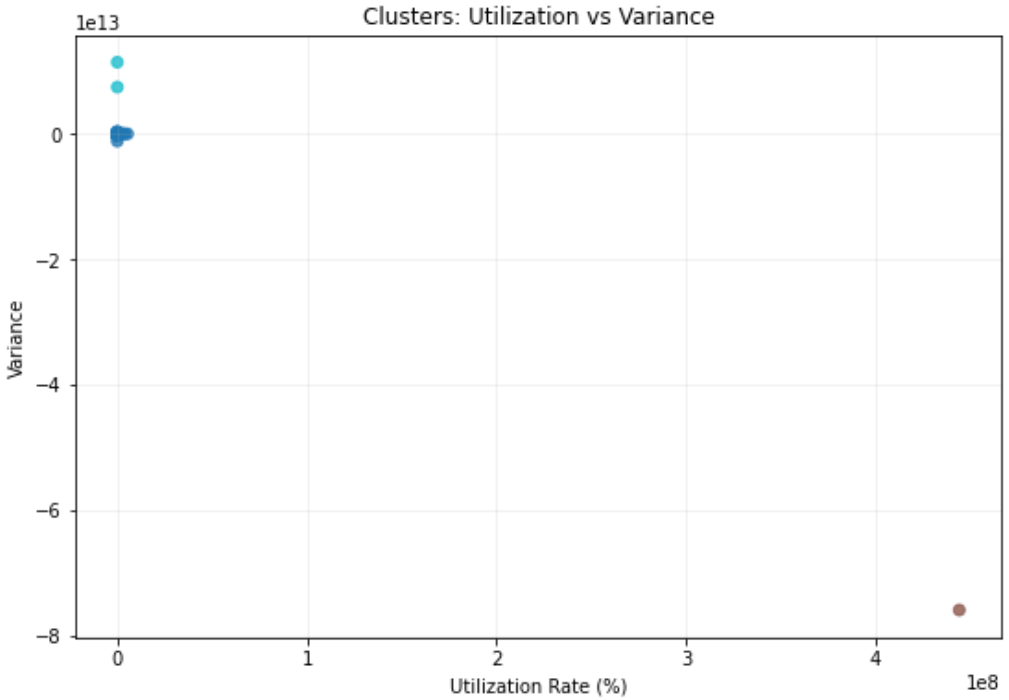
Policy / Audit Implication

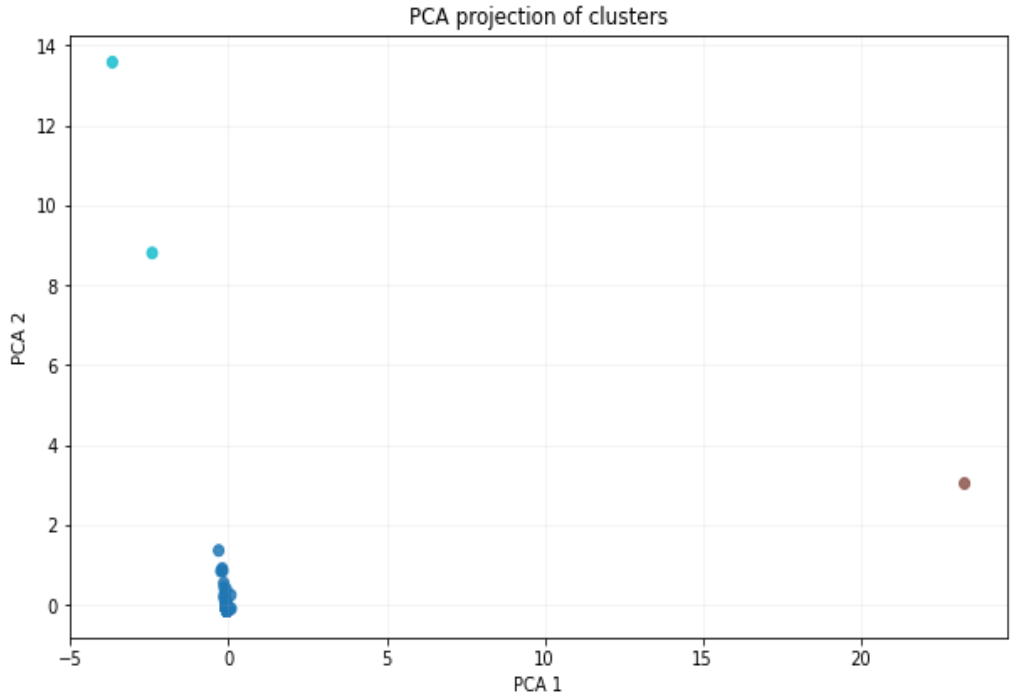
The scarcity of MDAs executing budgets as approved highlights deep inefficiencies in planning and absorption. Underspending points to:

- Delayed fund disbursement
- Poor absorption capacity
- Overestimation during budget approval

Clustering Analysis

Spending Behavior Clusters Across MDAs



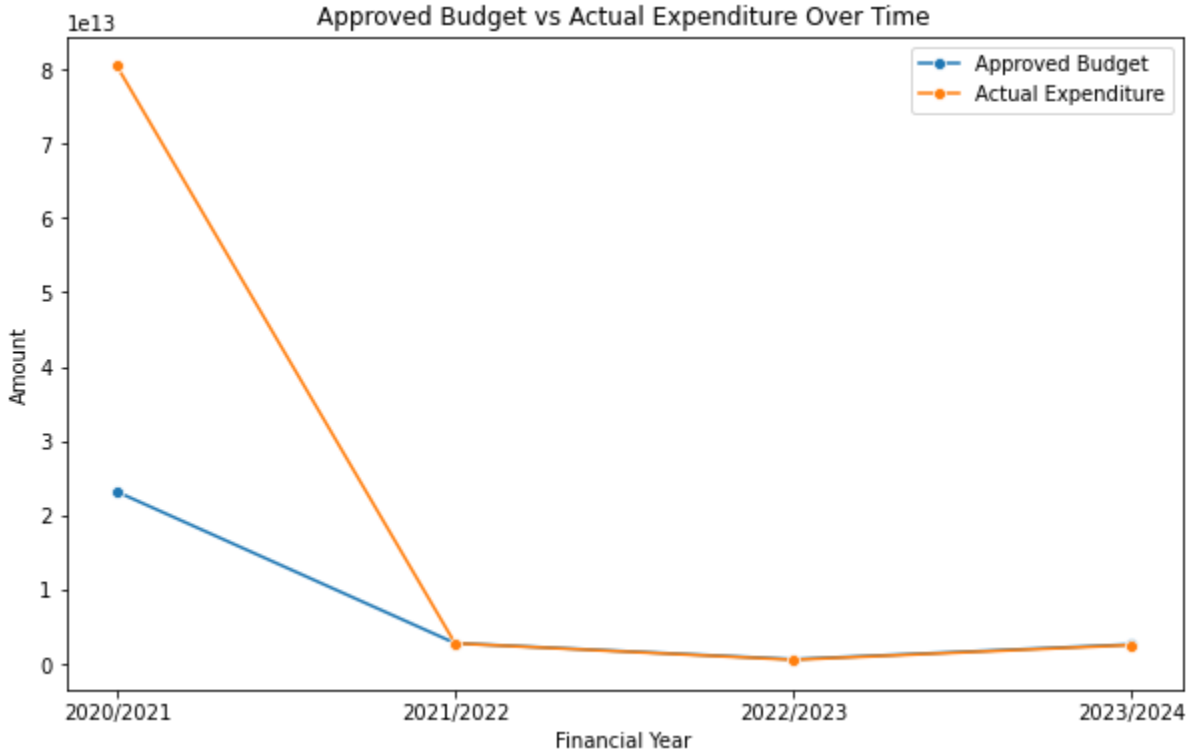


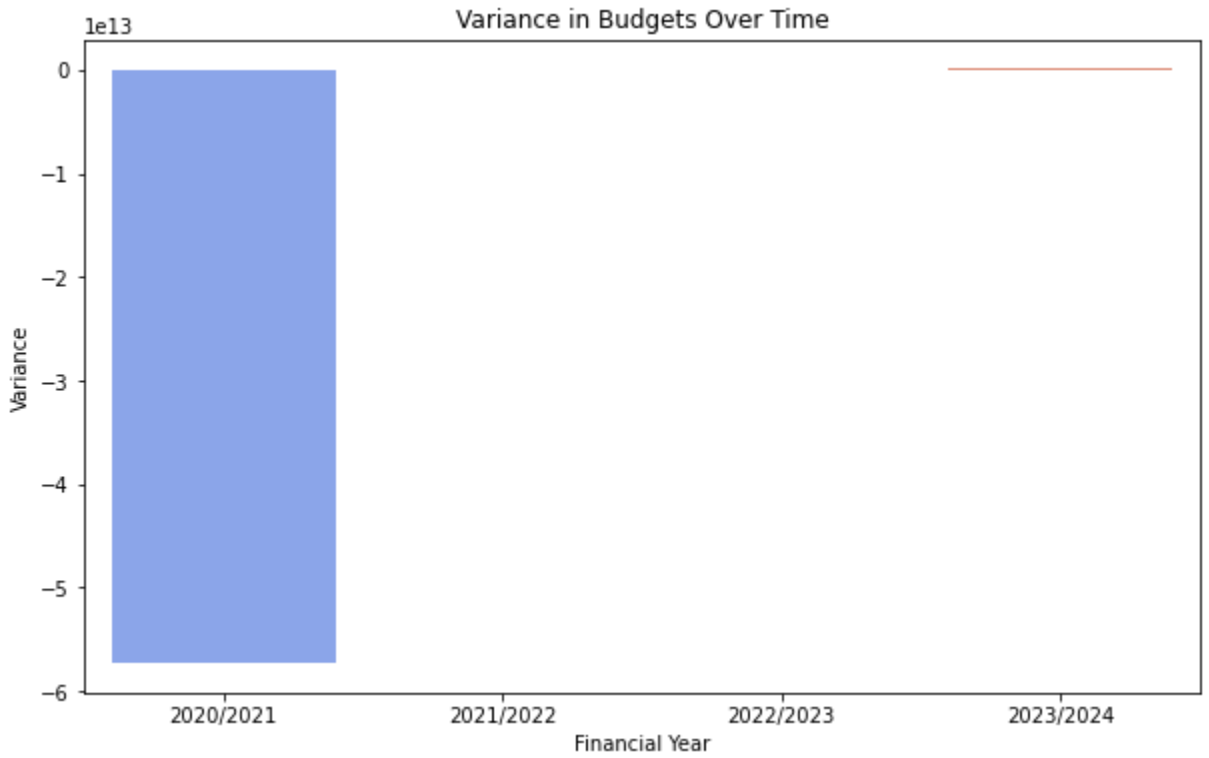
We used **KMeans clustering** to group MDAs based on budget, utilization, and variance.

Three clusters emerged:
Cluster 0 (Efficient Spenders): High utilization, low variance.
Cluster 1 (Underspenders): low utilization, large positive variances.
Cluster 2 (Overspenders): Overspenders Negative variances, spent more than allocated.

Most MDAs fall in Cluster 0, but the existence of risky outliers (Clusters 1 & 2) signals red flags for audits.

Budget and Expenditure Trends Over Time





Key Insights

Budget Growth: Approved budgets have generally increased across years.
Expenditure Gaps: In some years, expenditures lag behind allocations, creating under-utilization.
Variance Trends: Variance spikes in certain years suggest inefficiencies or unspent funds.

These findings can guide fiscal policy and highlight years where budget execution needs scrutiny..

Feature Engineering & Modeling Outcomes

Model Performance (Tuned XGBoost)

RMSE: 2.28×10^{11}

R^2: 0.48

Key Findings:

- RMSE dropped dramatically R^2 jumped from negative to 0.48, meaning the model now explains about 48% of the variance in Approved Budget. This is a substantial improvement — the model is capturing real patterns in the data. The fact that XGBoost can predict them reasonably well suggests consistent patterns across years, ministries, and text descriptions.

Limitations:

~52% of variance remains unexplained.
Likely due to **political, policy, or external drivers** not in the dataset.

Baseline Model (TF-IDF + Linear Regression)
Performed poorly.
Text alone in a linear framework cannot explain allocations.

Insights and recommendations

Financial & Policy Recommendations

1. Address Chronic Underspending. Several MDAs consistently underspend their approved budgets. This points to absorption challenges and inefficiencies in project execution. Action: Review procurement and implementation bottlenecks, and revise budget ceilings for recurrent underspenders to better align resources with capacity.
2. Tighten Oversight on Overspending MDAs. Overspending cases pose significant financial risks. Action: Establish a red-flag threshold (e.g., >120% budget utilization) that automatically triggers audits or closer monitoring.
3. Audit Prioritization Using Clusters. With nearly 70% of MDAs flagged, resources for auditing should be allocated strategically. Action: Prioritize overspenders and extreme underspenders, while maintaining light-touch monitoring of efficient MDAs.
4. Improve Budget Forecasting & Planning. Persistent variance between approved budgets and actuals suggests weak forecasting mechanisms. Action: Require evidence-based justifications for allocations,and apply historical performance trends when setting future ceilings.
5. Capacity Building for Low-Performing MDAs. Low utilization rates often reflect weak planning and reporting capacity. Action: Provide training and technical support, simplify fund disbursement procedures, and strengthen accountability frameworks.
6. Strengthen Budget Formulation. Require MDAs to justify allocations with detailed implementation capacity assessments (staffing, procurement readiness, past absorption rates). Adopt evidence-based budgeting where future allocations are tied to past budget utilization trends.
7. Introduce Performance Contracts. Tie senior management performance appraisals to budget absorption KPIs (e.g., ≥90% execution rate). Use dashboards to flag anomalies in real time (based on cluster membership).
8. Capacity Building & Digitalization. Train MDA finance departments on planning, reporting, and monitoring. Enhance the Integrated Financial Management Information System (IFMIS) to automatically track variance and trigger alerts.

Model Deployment

Link to the model:
Link to the Tableau Dashboard:

Thank You!

Contact Information:

erastus@gmail.com
lucianandanu9@gmail.com
irenekibengo@gmail.com
bensonmwihia@gmail.com
danakwabi@gmail.com
sydneywere563@gmail.com