DATA ANALYSIS REPORT

THE IMPACT OF FOREIGN AID CUTOFF ON KENYA'S SOCIO-ECONOMIC SECTORS

Authors:

1. *Norman Mwapea, Lead Analyst*
2. *Pauline Karuiki, Assistant Team Lead*
3. *Angela Chesire*
4. *Carlton Ogolla*
5. *Emmanuel Chol*

Date: *November 5th, 2025*

Executive Summary

Between 2010 and 2025, the United States—through USAID—has been a cornerstone of Kenya's development, investing billions of dollars across crucial sectors such as health, education, agriculture, and economic growth. These investments have significantly contributed toward Kenya's Vision 2030 and the Sustainable Development Goals (SDGs).

In 2025, however, a U.S. foreign aid freeze triggered severe disruptions across these sectors. Health programs suffered staffing shortages, agricultural initiatives stalled, and educational projects were delayed. This event exposed Kenya's high dependency on U.S. aid and raised pressing concerns regarding the nation's fiscal resilience and sustainability.

This project undertakes a comprehensive data-driven assessment to measure the magnitude and implications of these funding cuts. Using data analytics, regression modeling, and forecasting techniques, the report quantifies impacts across sectors and models future

outcomes should aid reductions persist. The findings are intended to guide policymakers, donors, and development partners toward sustainable funding strategies and policy resilience.

Overview

The United States has long been one of Kenya's **primary development partners**, contributing billions of dollars annually through USAID to enhance healthcare, education, food security, and youth empowerment.
When the **2025 U.S. aid suspension** occurred, it disrupted vital programs and revealed structural vulnerabilities in Kenya's development financing framework.

The project's goal is to **analyze the scale, distribution, and impact** of USAID funding across all development sectors from **2000 to 2025**, and to forecast the likely economic and sectoral consequences if such funding constraints continue.

Problem Statement

The Foreign Aid Cut-Off Project was developed to address the challenge of unpredictable and uneven foreign aid flows to Kenya. Fluctuations and cuts in funding across sectors and agencies make it difficult for policymakers to plan effectively and allocate resources where they are most needed. By applying data-driven modeling and forecasting, the project aims to transform raw aid data into clear, actionable insights that support more stable, transparent, and equitable funding decisions.

Business Objectives

The analysis is structured around four core business questions:

1. **How dependent is Kenya's development on U.S. foreign aid?**
   The report assesses sectoral and agency-level funding flows to quantify this reliance.
2. **What are the short- and long-term effects of funding disruptions?**
   Through regression and time-series models, the analysis estimates how funding cuts affect sectoral performance and national development outcomes.
3. **Which sectors are most vulnerable?**
   Exploratory data analysis (EDA) to identify sectors most at risk under continued aid withdrawal.

Success Metrics

The success of this project was measured through both technical performance and practical impact:

- **Model Accuracy:** Achieving an **R^2 of 80%** (R-squared means how much data the model explains. The higher this value, the better the model**)** in regression modeling and strong predictive accuracy across forecasting models, ensuring reliable performance. We also evaluate our models using RMSE (Root Mean Squared Error) and MAE (Mean Absolute Error)- the closer to 0 these values are, the better the model.
- **Data Quality:** Produced a clean, structured, and reproducible dataset covering multiple sectors and agencies, enabling consistent analysis and validation.
- **Model Interpretability:** Established clear explanations for key drivers of funding patterns, including sector volatility and agency influence.
- **Actionable Insights:** Delivered insights that highlight major funding contributors, underfunded sectors, and potential areas for strategic realignment.
- **Future Scalability:** Designed the modeling framework to be modular and expandable, allowing integration with new data sources and real-time monitoring systems.

Data Source

This project's data was sourced from ForeignAssistance.gov, which provides comprehensive records of U.S. foreign aid disbursements to countries around the world.

The full dataset contains over 3 million records, covering aid flows from 1954 to 2025. For our analysis, we filtered the dataset to focus exclusively on Kenya and limited the time range to 2000–2025, reducing the data to 80,072 records across 56 columns. This dataset includes detailed information on aid flows, sectors, agencies, and funding amounts, making it highly suitable for analyzing patterns in U.S. assistance to Kenya.

Constraints

The dataset represents historical foreign aid data up to the most recent available records; any changes or commitments after the dataset's last update are not captured. Additionally, while comprehensive, the dataset may have occasional missing values or inconsistencies typical of large administrative datasets.

Data Understanding

Dataset Overview

Data Attributes:

Country and Regional Information

- **country_id** – Numeric identifier assigned to Kenya (404).
- **country_code** – ISO 3-letter country code (KEN).
- **country_name** – Recipient country name ("Kenya").
- **region_id** – Numeric code for geographic region (e.g., 5 for Sub-Saharan Africa).

- **region_name** – Regional classification of the country.

Income Classification

- **income_group_id** – Numeric classification based on World Bank standards.
- **income_group_name** – Income classification name (e.g. "Lower Middle Income Country").
- **income_group_acronym** – Abbreviated income category (e.g. LMIC).

Managing Agency Information

- **managing_agency_id** – Unique ID for the U.S. managing agency.
- **managing_agency_acronym** – Acronym for the managing agency (e.g. USAID, DFC, EPA).
- **managing_agency_name** – Full name of the managing agency.

Sub-Agency / Bureau

- **managing_subagency_or_bureau_id** – Numeric code for the sub-agency/bureau.
- **managing_subagency_or_bureau_acronym** – Short code of the bureau (e.g. AFR).
- **managing_subagency_or_bureau_name** – Full bureau or sub-agency name (e.g. Bureau for Africa).

Implementing Partner Information

- **implementing_partner_category_id** – Numeric ID representing partner type.
- **implementing_partner_category_name** – Broad category (e.g. "Enterprises," "NGOs").
- **implementing_partner_subcategory_id** – Numeric ID for the sub-category.
- **implementing_partner_subcategory_name** – Specific sub-category (e.g. "Enterprises – Non-U.S.").
- **implementing_partner_id** – Unique identifier for each implementing partner.
- **implementing_partner_name** – Official name of the partner organization.

International Classification

- **international_category_id** – Numeric code grouping projects by theme (e.g. Health, Education).
- **international_category_name** – Category name (e.g. "Health and Population").
- **international_sector_code** – Numeric code representing the specific sector.
- **international_sector_name** – Name of the sector (e.g. "HIV/AIDS").
- **international_purpose_code** – OECD DAC code for project purpose.
- **international_purpose_name** – Description of the purpose (e.g. "STD Control Including HIV/AIDS").

U.S. Classification

- **us_category_id** – Numeric identifier for the U.S. reporting category.
- **us_category_name** – U.S. aid classification (e.g., “Health,” “Economic Growth”).
- **us_sector_id** – Numeric code for the sector under the U.S. system.
- **us_sector_name** – Sector name (e.g. “Direct Administrative Costs,” “HIV/AIDS”).

Funding and Accounts

- **funding_account_id** – Identifier for the funding account (e.g. 19x1031).
- **funding_account_name** – Account name (e.g. “Dept. of State – Global Health Programs”).
- **funding_agency_id** – Numeric identifier for the funding agency.
- **funding_agency_name** – Name of the funding agency (e.g., “U.S. Agency for International Development”).
- **funding_agency_acronym** – Short form of the funding agency (e.g., “USAID”).

Program Objective

- **foreign_assistance_objective_id** – ID representing aid objective type.
- **foreign_assistance_objective_name** – Broad objective category (e.g. “Economic,” “Humanitarian”).

Aid Type

- **aid_type_group_id** – Numeric grouping of aid delivery methods.
- **aid_type_group_name** – Type of aid (e.g. “Project-Type,” “Technical Assistance”).
- **aid_type_id** – Unique code for specific aid intervention.
- **aid_type_name** – Description of the aid type (e.g. “Project-Type Interventions – Not Investment Related”).

Activity Information

- **activity_id** – Unique identifier for each activity or project.=
- **submission_id** – Internal reference for data submission batch.
- **activity_name** – Project title or short description.
- **activity_description** – Brief narrative description of the activity.
- **activity_project_number** – Project or contract reference number.
- **activity_start_date** – Start date of the activity.
- **activity_end_date** – Expected or actual end date.
- **submission_activity_id** – Tracking number used for project submission.

Transaction and Financials

- **transaction_type_id** – Numeric identifier for transaction type.
- **transaction_type_name** – Type of transaction (e.g. "Obligation," "Disbursement").
- **fiscal_year** – U.S. government fiscal year of the transaction.
- **transaction_date** – Exact date of transaction execution.
- **current_dollar_amount** – Nominal (non-adjusted) transaction amount in U.S. dollars.
- **constant_dollar_amount** – Inflation-adjusted amount standardized to a base year.
- **activity_budget_amount** – Total planned or approved budget for the activity.

Data Preparation

The data preparation phase transformed the raw **ForeignAssistance.gov** dataset into a clean, structured, and analytically usable format suitable for modeling and forecasting.
Given that the dataset spanned over seven decades (1954–2025) and included more than **80,000 records** across 56 attributes, extensive preprocessing was required to ensure integrity, consistency, and reproducibility.

Data Cleaning

Filtering by Time Frame

Since the study focuses on Kenya's recent aid landscape and the potential effects of the **2025 USAID aid freeze**, the dataset was filtered to only include transactions from **2000 to 2025**.
This ensured relevance to current policy and development contexts while maintaining adequate historical depth for trend analysis.

Dropping Non-Analytical and Redundant Fields

Several columns were excluded as they provided minimal analytical value or represented unique identifiers.
Examples include:

- All columns containing "_id" (e.g. country_id, funding_agency_id),
- Redundant textual fields (e.g. country_code, region_name, funding_account_name),
- Repetitive categorical aliases (income_group_acronym, managing_agency_acronym, etc.).

Handling Duplicates

- Duplicate records were identified and removed, reducing six redundant entries.
- Final dataset size after deduplication: **79,488 rows × 56 columns**.

Handling Missing Values

Missing values were concentrated in date-related and descriptive columns:

- transaction_date contained null entries due to incomplete reporting.
- Missing transaction_date values were imputed using the corresponding **fiscal_year**, assigning a synthetic mid-year date to maintain temporal continuity.
- Columns with excessive missing or redacted text (activity_description, activity_project_number, activity_budget_amount) were dropped.

After refinement, the dataset retained **14 key analytical columns**, such as: managing_agency_name, us_sector_name, funding_agency_name, implementing_partner_name, transaction_type_name, fiscal_year, and constant_dollar_amount.

Data Transformation

Column Standardization

To ensure uniformity and ease of reference:

- All column names were converted to lowercase.
- Spaces were replaced with underscores.
- Special characters were stripped using regular expressions.
  This produced a consistent naming convention ideal for pandas and modeling pipelines.

Data Type Conversion

Several variables were typecast into their appropriate formats:

- fiscal_year → Integer
- transaction_date, activity_start_date, activity_end_date → Datetime
- activity_budget_amount, current_dollar_amount, constant_dollar_amount → Float

These conversions ensured accurate numerical computation and time-series analysis.

Temporal Validation

The fiscal_year and transaction_date columns were cross-checked for logical alignment, ensuring no transaction occurred outside its stated fiscal period.
Where mismatches occurred, transactions were corrected or excluded.

Feature Engineering

To enable deeper analytical and predictive modeling, several derived features were created:

1. **Temporal Features:**
   Extracted year, month, and rolling averages to capture trends and seasonality.
2. **Aid Volatility:**
   Computed 3-year rolling means and standard deviations for each sector and agency to measure volatility in aid disbursement.
3. **Aggregate Features:**
   Summed total aid per sector, agency, and fiscal_year to support comparative analysis.
4. **Growth Metrics:**
   Calculated annual growth rates in aid per sector and agency.
5. **Dependency Ratios:**
   Derived measures of each sector's share of total aid to assess dependency and exposure to funding shocks.
6. **Normalized Ratios:**
   Created ratios of sectoral aid / total aid and agency aid / total aid to standardize across fiscal years.
7. **Diversity Indices:**
   Introduced an **Aid Diversity Index** using Shannon Entropy to evaluate funding concentration across partners and sectors.

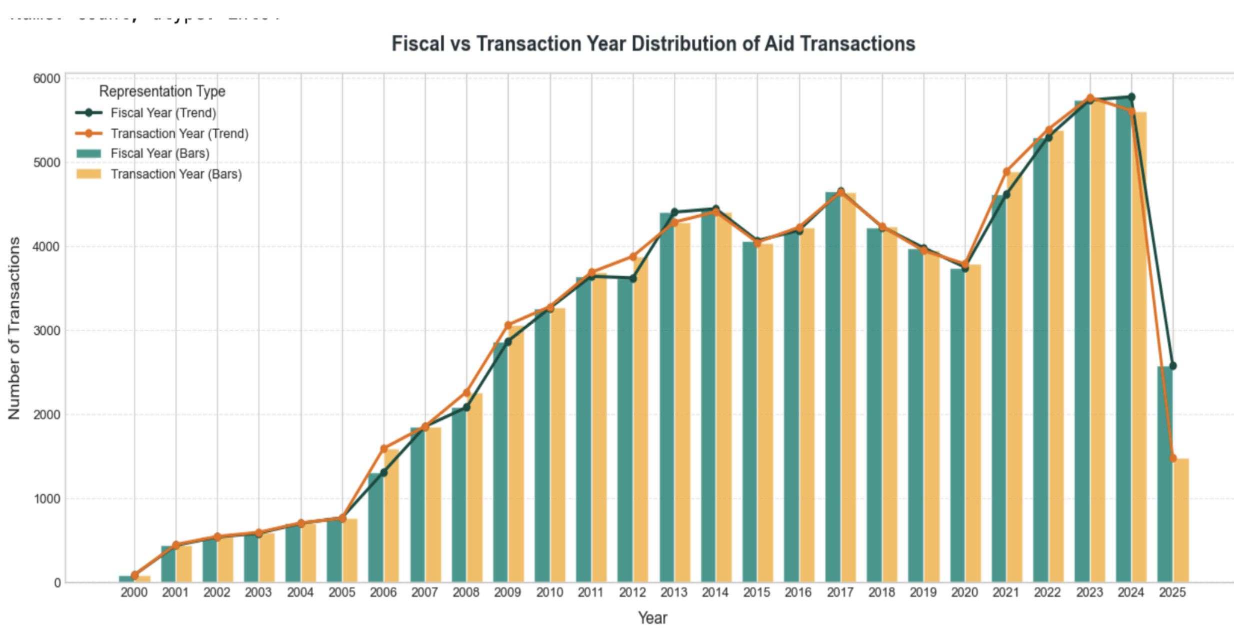
Final Prepared Dataset

After cleaning, transformation, and feature engineering:

- The dataset was reduced to approximately **79,000 clean, well-structured records**.
- All key variables were standardized, normalized, and enriched with temporal and volatility metrics.
- The dataset was then split into **training and testing subsets** for modeling and forecasting stages.

Exploratory Data Analysis

1. Fiscal vs Transaction Year of Aid Transaction

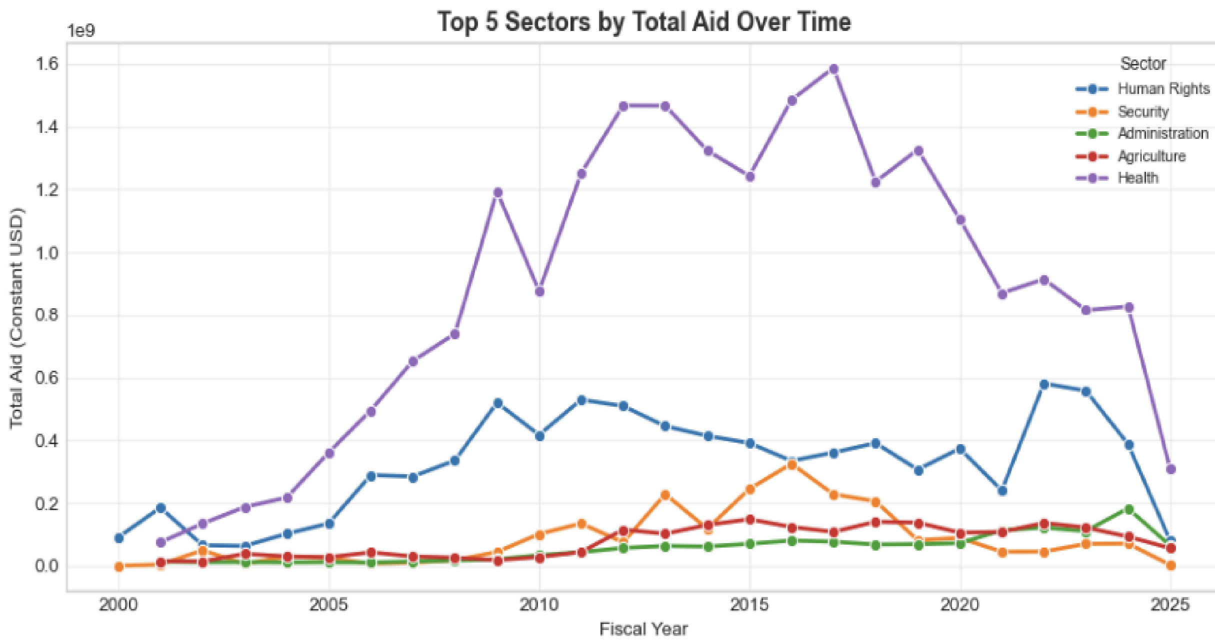


The figure above shows the trend difference in fiscal year and transaction year

Observation:

- The fiscal and transaction year trends align closely overall, but small gaps appear after 2006, suggesting delays between planned and actual disbursements, with both dropping sharply in 2025 due to funding reductions.

2. Top five sectors by total aid over time

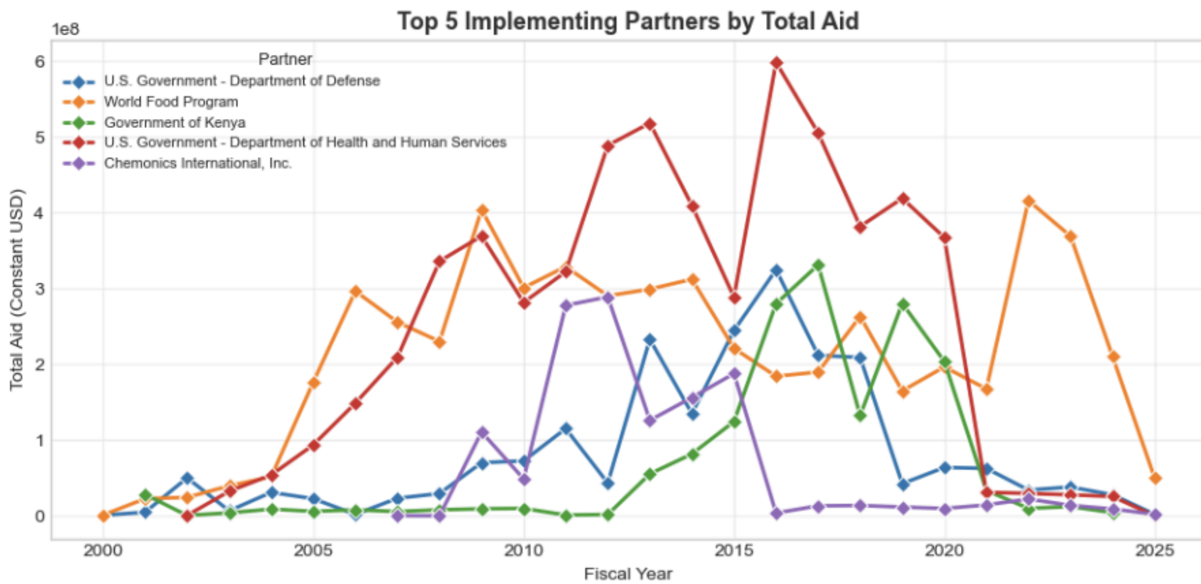


The figure above shows total aid growth trends over time for the top 5 socio-economic sectors

Observation:

- Health dominates aid allocations, with rising Human Rights and Security funding after 2010, while Agriculture and Administration remain steadily lower-funded.

3. Top Implementing Partners by Total Aid

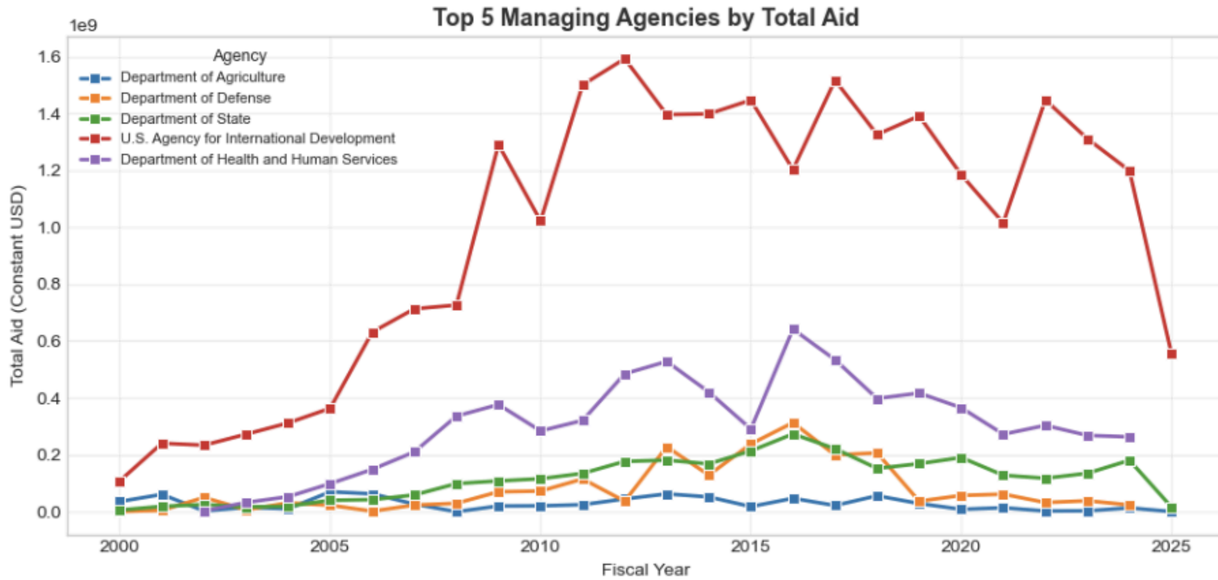


The figure above shows the top project implementation partners over time

Observation:

- Aid implementation is dominated by international partners like the World Food Program and Chemonics International, reflecting Kenya's continued reliance on external delivery capacity.

4. Top Managing Agencies By Total Aid

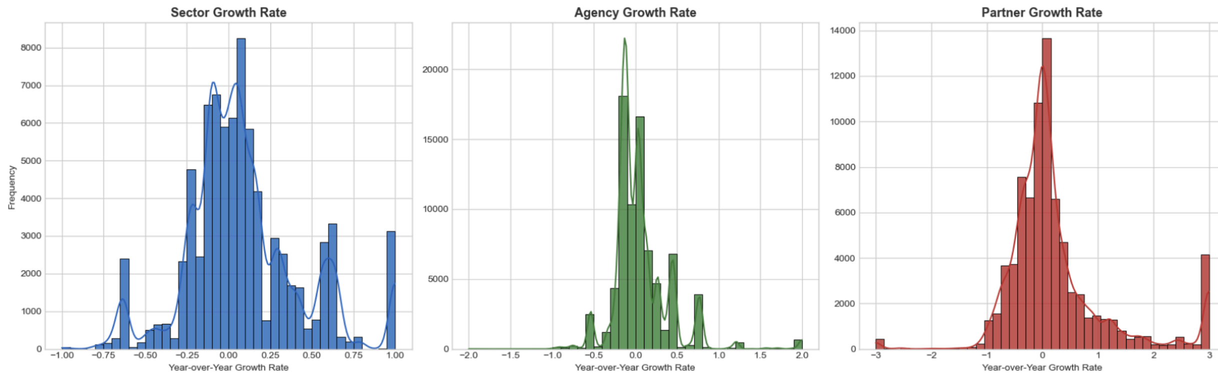


The figure above shows the top projecting management agencies by total aid over time.

Observation:

- USAID consistently leads as the top managing agency for total aid, accounting for the majority of disbursements across all years.
- Aid volumes rose sharply after 2005, indicating a period of major funding expansion.
- USAID aid peaked between 2010 and 2012, reaching over USD 1.5 billion before experiencing periodic declines thereafter.
- Department of State and Department of Health and Human Services show moderate but steady growth, maintaining aid levels between USD 0.4–0.6 billion post-2010.
- Department of Defense and Department of Agriculture remain minor contributors, reflecting narrower or more specialized aid mandates.
- Overall trend shows high volatility in aid flows for USAID, suggesting sensitivity to shifts in U.S. policy priorities or global events**.**

5. Distribution of Growth Rates

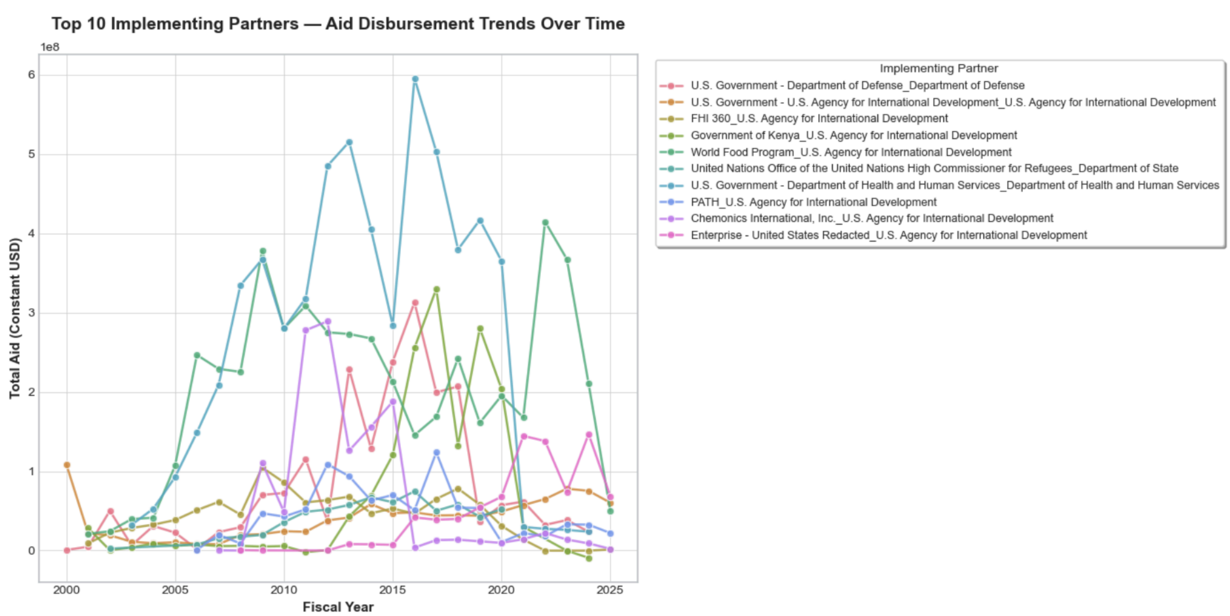


The figure above shows Kenya's aid growth rate distribution.

Observation:

- Sector, agency, and partner growth rates cluster around zero, indicating overall stability with moderate year-to-year fluctuations; however, a few extreme spikes suggest occasional sharp increases or declines in specific funding streams.

6. Aid Disbursement Trend Over Time For The Top Project Implementing Partners

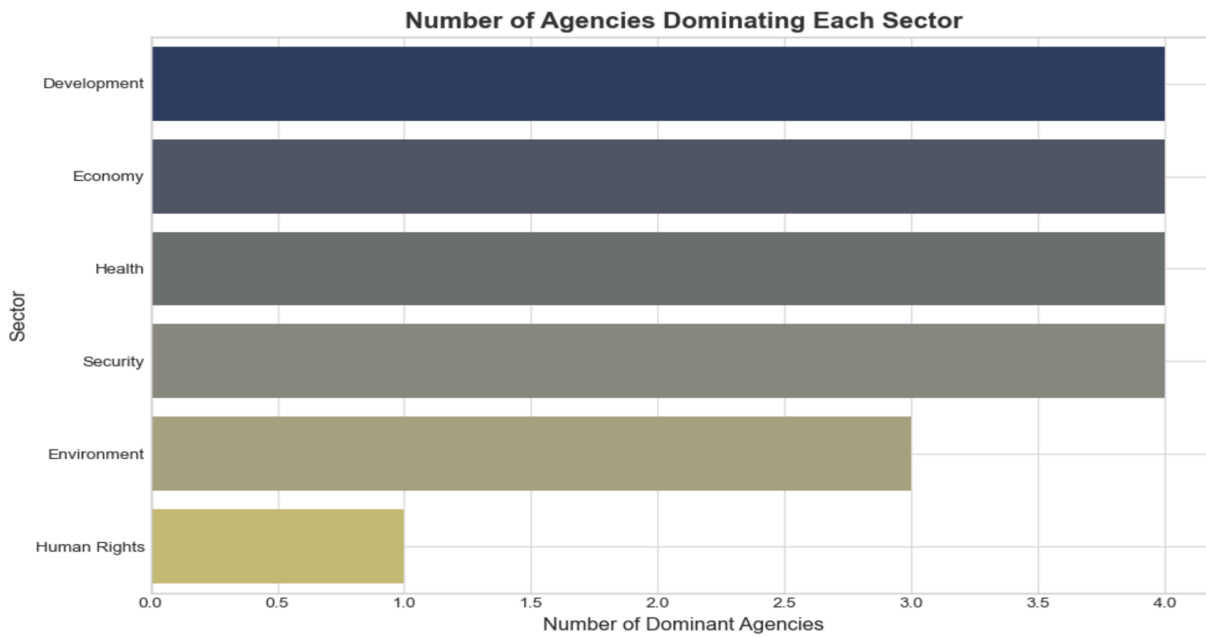


The figure above represents the aid disbursement trend over time for the top project implementing partners

Observation:

- Aid is dominated by a few major partners, notably the UNHCR and the Government of Kenya via USAID, with sharp year-to-year fluctuations between 2010 and 2020, followed by moderate stabilization after 2020 but continued dependency on key contributors.

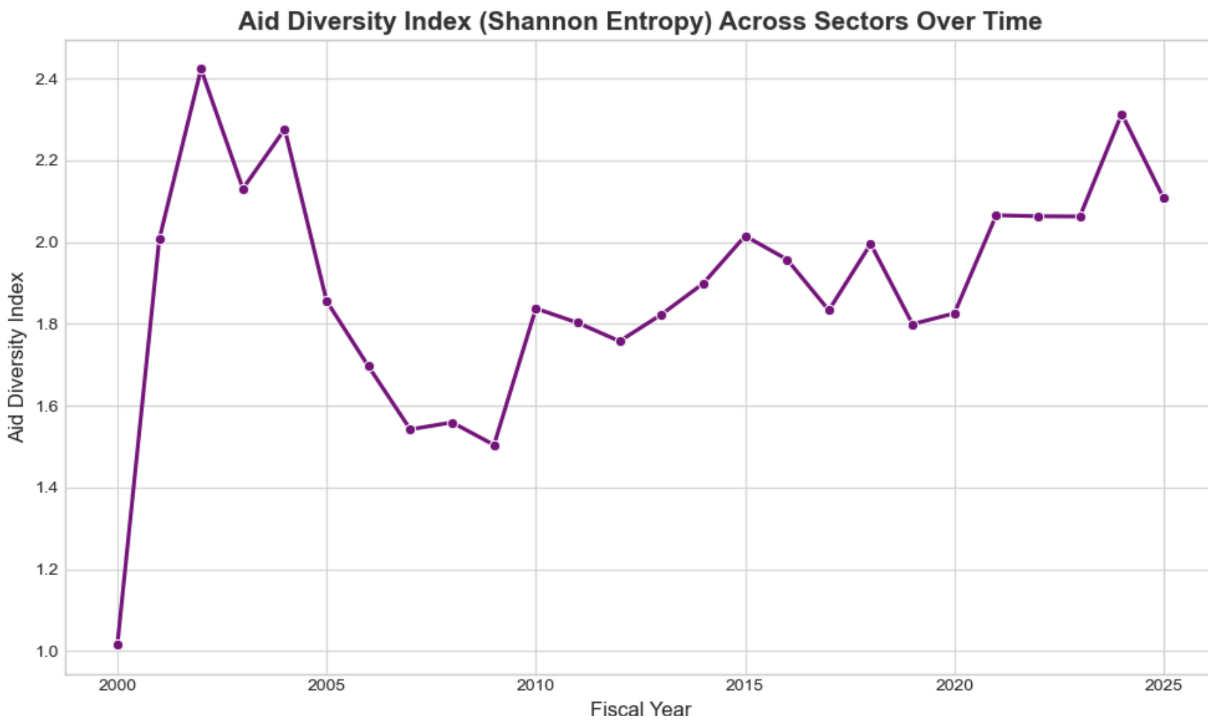
7. What sectors are dominated by which Agencies



The figure above shows the sector-agency share in Kenya

- Development, Economy, Health, and Security have the most number of agency concentration.
- Meanwhile, Environmental initiatives are led by three agencies.
- Human Rights, dominated by a single agency (the Department of State).

8. Aid Diversity Index Across Sectors Over Time

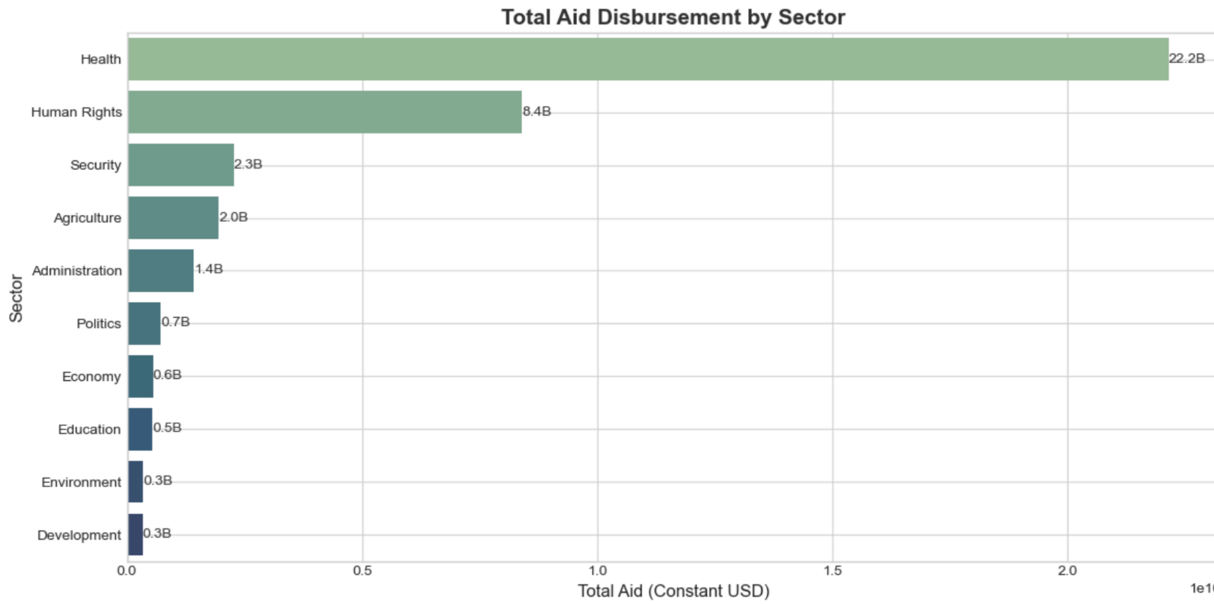


The aid diversity index across different sectors over the years is as depicted above

Observation:

- The Aid Diversity Index shows an initial surge in the early 2000s as funding broadened across multiple sectors, followed by mid-decade concentration in key areas like health and governance.
- From 2010 onward, diversity steadily recovered and stabilized, indicating a more balanced, mature aid distribution where no single sector dominates.

9. Total Aid Distribution by Sector

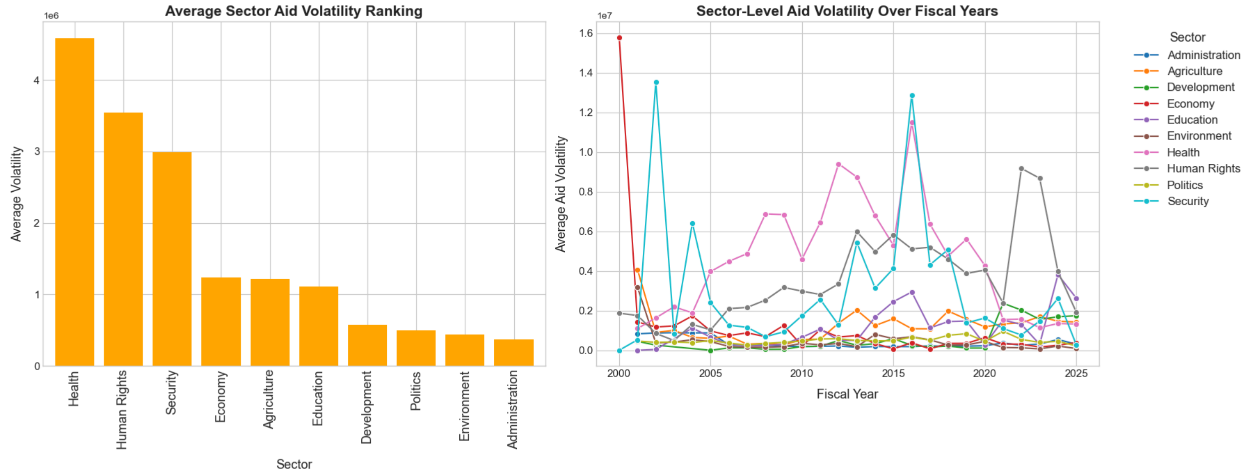


The figure above shows the total aid distribution for all sectors

Observation:

- Aid allocation is heavily concentrated in the health sector, which alone accounts for approximately $22.2 billion, underscoring a strong donor emphasis on public health and humanitarian priorities.
- Other sectors, such as Human Rights ($8.4B) and Security ($2.3B), receive moderate support, reflecting secondary focus on governance and stability. In contrast, areas like Education, Environment, and Development remain significantly underfunded — each receiving less than $0.6B — suggesting limited investment in long-term, sustainable growth initiatives.
- Overall, Kenya's aid profile is oriented toward immediate health and crisis needs over diversified development programming.

10. Sector Aid Volatility Analysis



The figure above shows the average average sectoral aid volatility ranking and sector level aid volatility trends.

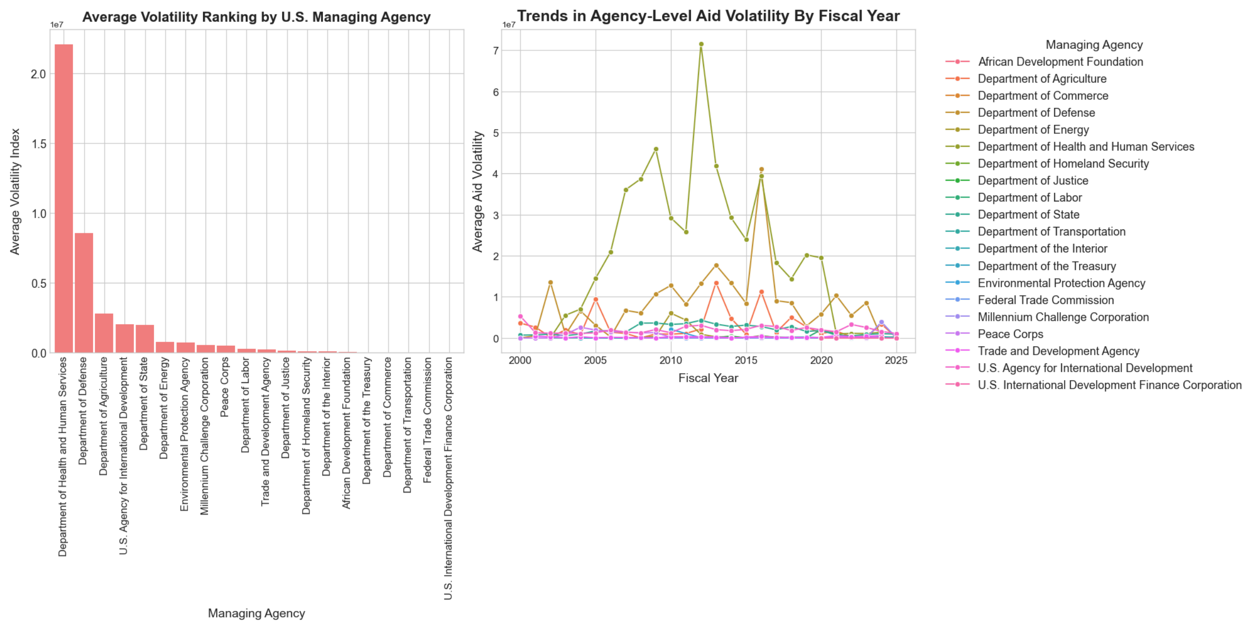
Observation:

- **Health sector exhibits the highest aid volatility**, indicating frequent and large fluctuations in funding levels over time.
- **Human Rights and Security sectors** follow closely, showing substantial instability that may reflect responsiveness to crises or shifting geopolitical priorities.
- **Economy, Agriculture, and Education sectors** show moderate volatility, suggesting relatively steadier but still variable funding patterns.
- **Development, Politics, Environment, and Administration sectors** experience the **lowest volatility**, implying more consistent aid flows and predictable funding.
- Temporal trends show **recurrent spikes** in volatility around **2005, 2010, and 2015**, especially in Health, Security, and Human Rights sectors.

Insight:

- The observed volatility patterns suggest that **aid allocation is highly reactive** to global events—such as health emergencies, conflict, or governance reforms—rather than following stable long-term commitments.

11. Agency Volatility Analysis



The figure above shows the agency volatility rankings

Observation:

- The **Department of Health and Human Services (HHS)** shows the **highest aid volatility** (≈ USD 22 million), indicating strong fluctuations likely tied to health emergencies or shifting global health priorities.
- The **Department of Defense** ranks second (≈ USD 8.6 million), reflecting reactive funding patterns related to security or humanitarian interventions.
- **Department of Agriculture**, **USAID**, and **Department of State** exhibit **moderate volatility** (≈ USD 2–3 million), suggesting periodic funding changes but more consistent overall trends.
- Agencies like the **Department of Energy**, **EPA**, and **Millennium Challenge Corporation** demonstrate **lower volatility**, pointing to steadier, project-based funding structures.
- The **Peace Corps** and **Department of Labor** also show **relatively stable patterns**, likely due to sustained programmatic commitments.
- **Minimal volatility** is observed among agencies such as the **Department of the Treasury**, **Commerce**, and **Transportation**, indicating predictable and limited foreign aid involvement.
- The **U.S. International Development Finance Corporation** records **no volatility**, consistent with either negligible or non-aid-related funding in the dataset period.

Modeling

The modeling phase aimed to analyze how foreign aid has fluctuated across Kenya's development sectors, and to predict potential future outcomes under continued funding reductions using available historical data.

The main modeling approach used was **regression analysis and predictive analytics** to quantify relationships between aid and sectoral dynamics.

Modeling Objective

To determine how sectoral funding patterns, agency allocations, and aid volatility correlate with the total constant-dollar amounts received, thereby identifying which variables significantly influence Kenya's funding trends.

Approach

Several algorithms were trained and tuned:

- Baseline models: Multiple Linear Regression and Random Forest Regressor.
- Ensemble models (main focus): XGBoost, CatBoost and Stacked Ensemble models.

Two strategies were applied:

1. Vanilla training - models with base parameters.
2. Hyperparameter optimization.

For ensemble models, hyperparameter optimization was performed using both RandomizedSearchCV and Optuna with TimeSeriesSplit Cross-Validation (TSCV).

Evaluation Metrics

The models were evaluated using:

- R-Squared- the total explained variance.
- Root Mean Squared Error (RMSE)
- Mean Absolute Error (MAE)
- Residuals plots

Results

1. **Multiple Linear Regression**

The analysis aimed to identify which sectors in Kenya are underfunded or heavily aid-dependent, and which U.S. agencies the country relies on most. All comparisons are made relative to defined baselines:

- Sector baseline: *Administration*
- Agency baseline: *African Development Foundation (ADF) / USIDFC*

Model Performance

All regression variants - including Ordinary Least Squares (OLS), Robust OLS, Clustered OLS, Weighted Least Squares (WLS), and ElasticNet - produced consistent and stable results, with minimal gaps between training and test performance.

Model	R^2 (Train)	R^2 (Test)	RMSE (Train)	RMSE (Test)	MAE (Test)
OLS	**0.7468**	**0.7063**	**1.4340**	**1.4566**	**1.0845**
Robust OLS	**0.7468**	**0.7063**	**1.4340**	**1.4566**	**1.0845**
Clustered OLS	**0.7468**	**0.7063**	**1.4340**	**1.4566**	**1.0845**
Weighted LS	**0.7433**	**0.7048**	**1.4439**	**1.4603**	**1.0796**
ElasticNet	**0.7427**	**0.7031**	**1.4456**	**1.4644**	**1.0960**

Interpretation:

- The models explain roughly 70% of the variation in U.S. aid amounts to Kenya - indicating strong predictive performance.
- OLS and its robust variants are both interpretable and stable.
- ElasticNet confirms the same structural relationships while penalizing weaker effects, validating the model’s generalization capacity.

Sector-Level Findings (Relative to *Administration*)

The *Administration* sector serves as the baseline; negative coefficients indicate sectors receiving less funding, while positive coefficients suggest higher funding relative to Administration.

Consistently Underfunded Sectors

Sectors that show persistently negative coefficients across all model types:

- **Education** — OLS ≈ *−1.97* ; ElasticNet ≈ *−0.30*
  → Strong evidence of chronic underfunding relative to Administration.
- **Environmen**t — OLS ≈ *−1.79* ; ElasticNet ≈ *−0.37*
  → Indicates a systemic gap in climate and resilience-related funding.
- **Economy / Economic Development** — OLS ≈ *−1.27* ; ElasticNet ≈ *−0.27*
  → Suggests limited investment in productivity and growth-oriented initiatives.
- **Development** — OLS ≈ *−1.13*
- **Security / Politics** — OLS ≈ *−1.12* and *−1.16* respectively.

These patterns persist even after regularization, signaling a structural bias in U.S. aid priorities - favoring governance and administrative sectors over long-term development areas.

Sectors at or Above Baseline

- **Administration** — Reference category.
- **Health** — OLS ≈ *−0.32* (statistically near baseline), suggesting funding levels comparable to Administration.
- **Human Rights** — Slightly positive in some model variants, though not statistically strong.

Implication: Kenya receives disproportionately less aid for education, environment, and economic development, reflecting a focus on short-term governance and health over long-term capacity building.

Agency-Level Findings (Relative to *ADF / USIDFC*)

The *African Development Foundation (ADF)* and *USIDFC* serve as baselines; positive coefficients denote agencies providing more aid than these reference points.

Primary Agencies Kenya Relies On

- **U.S. Agency for International Development (USAID)** — OLS ≈ *+2.96* ; ElasticNet ≈ *+1.32*
  → The principal and most stable funding source for Kenya.
- **Department of the Interior** — OLS ≈ *+2.03* ; ElasticNet ≈ *+0.29*
- **Department of Health & Human Services (HHS)** — OLS ≈ *+1.89* ; ElasticNet ≈ *+0.28*
- **Department of State** — OLS ≈ *+1.06* ; ElasticNet ≈ *+0.33*

These agencies represent Kenya's core U.S. aid partners, particularly in governance, health, and infrastructure programs.

Agencies Contributing Less than Baseline

- **Federal Trade Commission (FTC) —** OLS ≈ *−2.15*
- **Millennium Challenge Corporation (MCC) —** OLS ≈ *−1.23*

These entities are more specialized or selective in their interventions, resulting in lower aid volumes within the dataset.

Cross-Cutting Insight: Volatility and Dependency

The volatility coefficient (log_agency_rolling_std_3yr) is both large and positive (OLS ≈ *+0.77* ; ElasticNet ≈ *+1.52*).

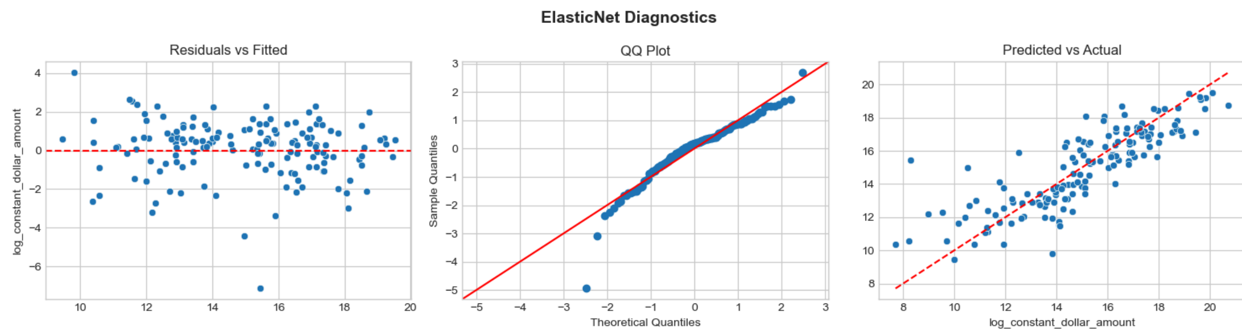
This suggests that agencies with higher past disbursement volatility tend to allocate larger amounts of aid in the present.

Interpretation: Kenya's aid inflows are heavily influenced by project-driven, variable agencies, exposing the country to funding uncertainty and making long-term planning more challenging.

Strategic Implications

- **Rebalance sectoral focus** — Encourage sustained funding for education, environment, and economic development to reduce long-term aid dependency.
- **Stabilize agency partnerships** — Strengthen engagement with high-volume but volatile agencies to achieve more predictable disbursement cycles.
- **Leverage health and governance stability** — Build from these relatively well-funded areas to channel support toward resilience and growth-oriented sectors.

Regression Plot Analysis



The figure above show residual plots for the different regression models tried

Residuals vs Fitted

- The residuals appear randomly scattered around zero, indicating that the model captures the main signal without evident patterns of heteroscedasticity or systematic bias.
- While a few extreme residuals are visible, suggesting the presence of outliers or unmodeled effects, the overall distribution is consistent with a well-specified model.

Q-Q Plot

- Residuals largely follow the 45-degree reference line, implying that they are approximately normally distributed.
- Minor deviations in the tails suggest the existence of a few atypical observations, but these do not materially affect the overall model performance.

Predicted vs Actual

- Predicted values closely align with observed values along the 1:1 line.
- This alignment demonstrates strong predictive accuracy and supports the model's ability to generalize across the range of aid amounts.
- Some dispersion at the extremes is observed but remains within acceptable limits for regression diagnostics.

Note: Among all the regression algorithms tested, the **ElasticNet model** emerged as the most statistically grounded. Its dual regularization framework - blending **L1 (Lasso)** and **L2 (Ridge)** penalties - provides a balanced mechanism for addressing both **overfitting** and **multicollinearity**, common challenges in complex socioeconomic data. By constraining coefficient magnitudes and selectively shrinking less informative predictors, ElasticNet preserves model interpretability while enhancing generalization to unseen data.

2. Ensemble Regression

Iteration	Model	MAE	RMSE	R^2	Notes
Vanilla	RandomForest	73,195.86	2,320,156.61	0.6446	Baseline ensemble model
Vanilla	XGBoost	93,990.67	1,793,715.74	0.7876	Baseline ensemble model
Vanilla	LightGBM	98,255.71	2,222,033.51	0.6741	Baseline ensemble model
Vanilla	CatBoost	66,005.40	1,282,695.68	0.8914	Baseline ensemble model
Vanilla	Stacked Ensemble	75,383.22	1,060,092.20	0.8646	Combines multiple base learners
Tuned (RandomizedSearchCV)	Stacked Ensemble (XGBoost + LGBM)	71,921.93	1,073,122.79	0.8612	RandomizedSearchCV applied on stacked model components
Tuned (RandomizedSearchCV + TimeSeriesSplit)	**XGBoost**	**16,884.69**	**181,031.84**	**0.9922**	**Exceptional accuracy and generalization**

Tuned (RandomizedSearchCV + TimeSeriesSplit)	CatBoost	33,513.60	510,427.57	0.9380	Stable model, high R^2 with moderate error
Optuna Optimization	XGBoost	38,174.08	688,506.65	0.8872	Optuna-based Bayesian tuning approach
Optuna Optimization	CatBoost	35,645.03	510,692.07	0.9380	Optuna results consistent with TimeSeriesSplit tuning

Among the ensemble regression models that we trained, including stacked ensembles, XGBoost with hyperparameter optimization using RandomizedSearchCV emerged the best on it's performance on test data with the following results:

- MAE: 16,884.69
- RMSE: 181,031.84
- R^2: 0.9922

CatBoost was a close second but it produced higher errors:

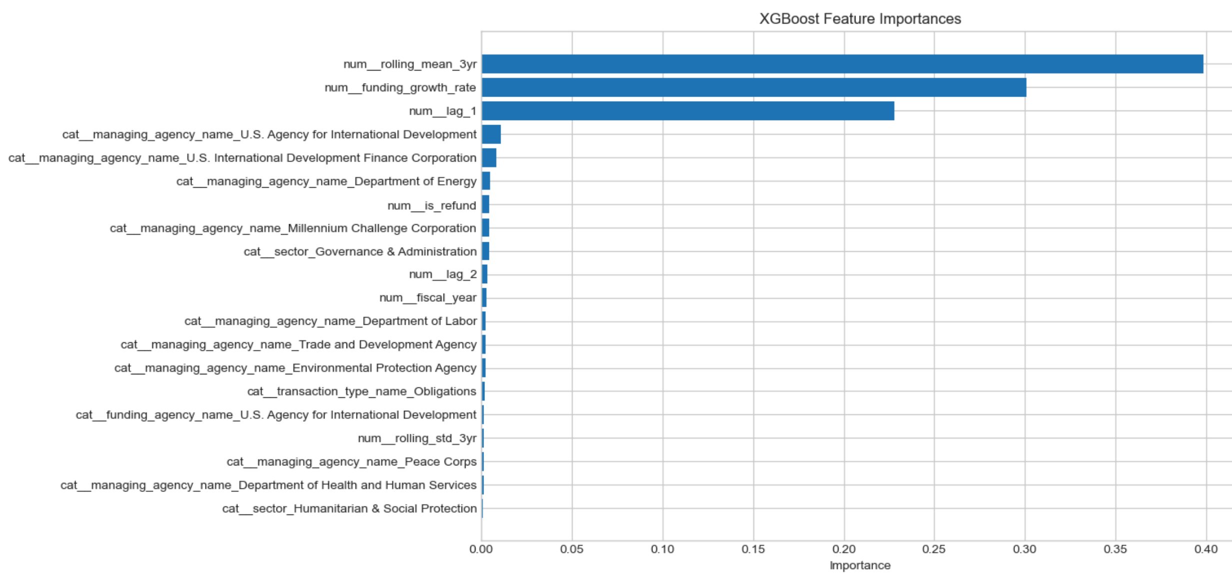
- MAE: 33,513.60
- RMSE: 510,427.57
- R^2: 0.9380

Feature Importance for our best model

Analysis of the XGBoost feature importances reveals that historical funding behavior is the dominant driver of future allocations, while agency and sector characteristics contribute minimally.

- Rolling 3-Year Average of Aid (rolling_mean_3yr, 0.416):
  The model places the highest weight on recent funding trends, suggesting that patterns in past allocations persist over time.

- Funding Growth Rate (funding_growth_rate, 0.295):
  Significant shifts in recent funding—either upward or downward—strongly influence expectations for future allocations.
- Previous Year's Funding (lag_1, 0.221):
  Funding in the prior year directly shapes the next year's budget, highlighting institutional momentum.
- Agency Indicators (0.004–0.009):
  While the specific agency handling funds has some influence, its effect is small relative to historical trends.
- Refund Transactions (is_refund, 0.007) and Fiscal Year (0.003):
  Minor adjustments and time-specific effects have limited predictive power.
- Sector Indicators (~0.001) and 3-Year Funding Variability (rolling_std_3yr, 0.001):
  Sector-level differences and short-term volatility have negligible impact on future allocations.



Conclusions

1. **Aid cuts threaten Vision 2030.**
USAID has been the primary driver of Kenya's aid agenda — setting priorities, shaping program design, and influencing national progress toward Vision 2030 through its funding decisions. The 2025 USAID funding freeze exposed critical gaps in local resilience and sustainability across major development programs.
2. **Health, Human Rights, and Security are most vulnerable.**
These sectors collectively receive over 80% of U.S. assistance, meaning even small disruptions cause wide-reaching effects—clinic closures and halted community programs.
3. **Funding is highly volatile.**
USAID aid levels fluctuate sharply in response to global and political shifts, making long-term planning difficult for Kenya's ministries and NGOs.
4. **Dependency outweighs diversification.**
Despite a broad base of implementing partners (≈230 annually), most aid is concentrated in a few large organizations, showing limited diversification and local ownership.
5. **Predictive analytics can guide policy.**
Regression and forecasting models ($R^2 \approx 0.99$ for tuned XGBoost) demonstrated strong potential to anticipate the effects of aid shocks — offering an early-warning system for policy planning.

Model Limitations

1. **Data Constraints:**
Analysis relied solely on USAID's ForeignAssistance.gov data (2000–2025). Local outcome data (e.g., school enrollment, yields, or employment) was unavailable, limiting causal interpretation.
2. **Volatility vs. Causality:**
Models capture correlations, not causation. High volatility may coincide with global crises, not necessarily funding decisions alone.
3. **Temporal Granularity:**
Annual fiscal data smooths short-term fluctuations. More granular (monthly/quarterly) data could improve forecasts.
4. **Feature Dependence:**
XGBoost heavily relied on past funding trends (rolling averages, lag features). Future models could integrate socio-economic indicators for more robust projections.

Recommendations

1. **Diversify Kenya's donor portfolio.**
   Strengthen engagement with non-U.S. partners (e.g., EU, AfDB, China) to mitigate shocks from single-donor dependencies.
2. **Mobilize domestic financing.**
   Build national fiscal buffers and local funding mechanisms to maintain service continuity during external aid disruptions.
3. **Stabilize partnerships with volatile agencies.**
   Negotiate multi-year funding frameworks with USAID and others to ensure predictability.
4. **Invest in data systems.**
   Develop Kenya's open data infrastructure to track development funding in real time — enhancing transparency and modeling precision.
5. **Leverage predictive modeling in planning.**
   Integrate machine learning forecasts into national development planning (e.g., Vision 2030 Secretariat, Ministry of Finance) to anticipate sectoral risks and allocate resources proactively.