



AuditsMadeSimple

An Interactive Platform for Exploring Government Data

AGA Datathon

Meet the Team

The Fiscal Patriots



Khaled Alkurd
Team Lead



Pranavi Doodala
Project Manager



Mariam Debas
Visualization Lead



**Nikita
Chandrasing**
Product Manager



Andy Yaro
Product Developer

The Mission



Create a **supplemental resource**
promoting literacy in audit terminology
and data to make government engagement
accessible for the general public.

The Solution

AuditsMadeSimple - a launchpad for understanding federal spending and audit data

The screenshot shows a presentation slide from the **fiscal patriots** website. The slide is titled "AGA . DATATHON 2026" and features the main heading "Where does **federal financial assistance** go?". Below the heading, there is a small navigation bar with arrows and the text "1 / 4". A call-to-action button labeled "PART 1" is visible. To the right, a "New here?" pop-up window provides information about the site's mission to make government spending transparent and includes buttons for "Take the Tour" and "Got it".

Features

01

Exploratory
Visualizations

02

Glossary of Auditing
Terms

03

Dataset Guide

04

Case Studies

Let's Take a Look →

01

Exploratory Visualizations

- Interactive and downloadable summaries of Federal Spending and Audit data
- Enable quick understanding of high-level data
- Highlight patterns in funding concentration and change over time

PART 1

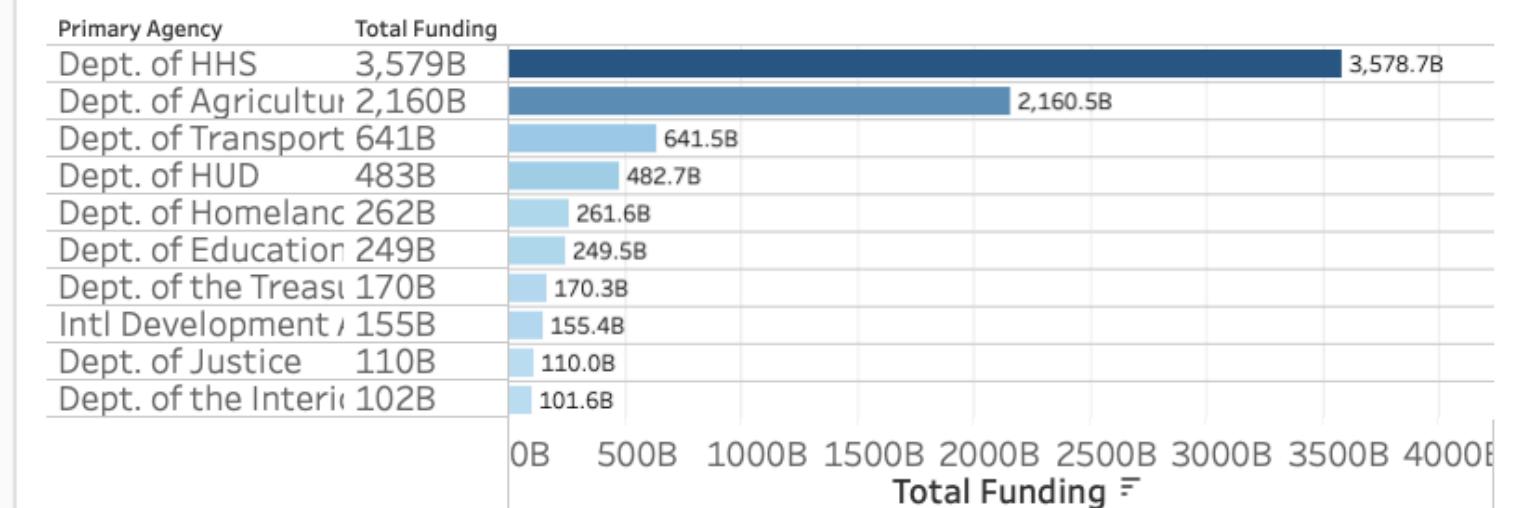
Funding by State

What you are seeing: Total obligation amounts aggregated by recipient state. This helps spot geographic concentration.

[Top 10 States \(50.9%\)](#)
[Grant spending map](#)
[Top 10 Agencies](#)

See which federal agencies are distributing the most funds.

Top 10 Federal Agencies by Funding


[View on Tableau...](#)


Note: High spending does not imply inefficiency.

02

Glossary for Financial Literacy

- Interactive, searchable cards featuring essential terms
- Prepares users to dive into and interpret data without a technical background
- Making government accountability language accessible for everyday citizens

Financial Literacy & Context

Government data is complex. Use this glossary to understand key terms like "Material Weakness" or "Obligation."

Search terms (e.g., 'audit', 'risk')...

All Basics Findings Outcomes Oversight Standards
 Finance Controls Performance Risk Data

Risk points	Audit	Auditor	Audit Report
BASICS	BASICS	BASICS	BASICS
Audit Scope	Documents, data, and analyses used by auditors to support findings and conclusions. <i>Why it matters: The proof backing up the audit's claims.</i>	Finding	Recommendation
BASICS	FINDINGS	FINDINGS	FINDINGS
Management Response	Corrective Action Plan	Implementation Status	GAO
OUTCOMES	OUTCOMES	OUTCOMES	OVERSIGHT

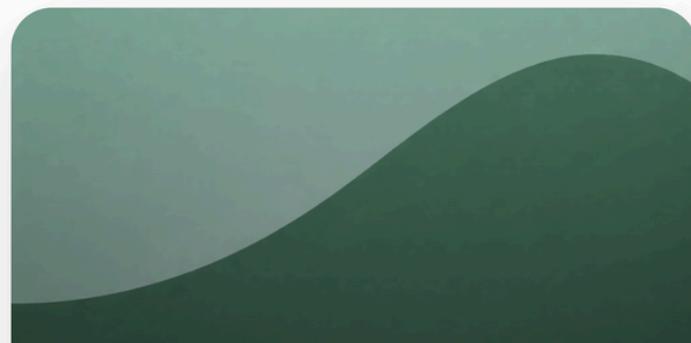
03

Dataset Guide

- Descriptions and links to cleaned datasets to help users navigate government data sources
- Assists users in identifying which sources can assist them in answering their own research questions

Datasets

We combine three official datasets to build this view. Understanding what each system tracks, and misses, is critical for analysis.



USAspending.gov

This is the official open data source for federal spending information, answering who received the money, how much was obligated, and which agency funded it. It does not show whether the money was used effectively.

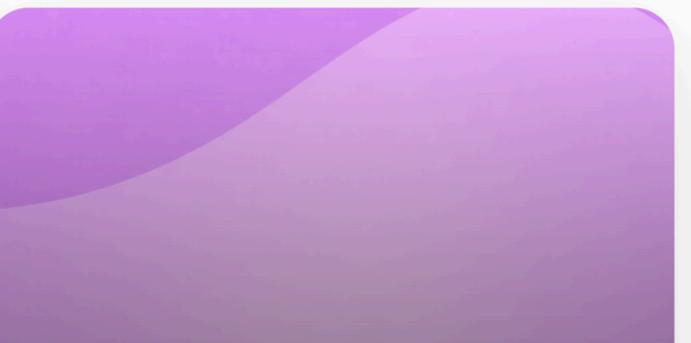
View Dataset



FAC.gov

This is the central repository for Single Audit reports filed by recipients of federal funds, useful for checking recent findings, material weaknesses, and the auditor's opinion. It does not provide real-time daily spending.

View Dataset



SAM.gov

This is the primary database for entity registration and debarments (exclusions), helping you check if an entity is debarred, its UEI (Unique Entity Identifier, a government-wide identifier for entities), and whether registration is active. It does not provide transaction history.

View Dataset

04

What's next?

- Next Steps
 - For the novice user
- Use Cases
- Case Studies
 - Exploratory exemplars of further analysis

EASY CHECKLIST

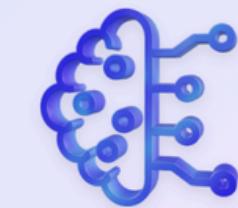
3 Things You Can Do Next

What can I do with government spending and audit data?

Government spending and audit data (from sources like USA Spending.gov and FAC.gov) can be useful for many types of users - from curious citizens to students, educators, community leaders, journalists, researchers, and nonprofit or local government staff. Here are a few practical ways people use this information:

Case Studies

Discover how transparency and data analysis have uncovered insights, improved oversight, and driven accountability in federal spending.



PREDICTIVE ANALYTICS

Predicting Next-Year Audit Findings

View how Machine Learning can be used on data to predict what it looks like in the future.

[Read Full Study →](#)



RISK ASSESSMENT

Audit Health Score

Creating innovative methodology from data.

[Read Full Study →](#)



The Process

Data Cleaning

USASpending.gov

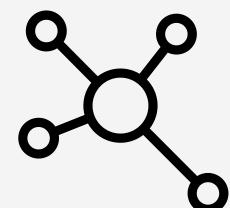
- Federal financial assistance (grants, loans, subsidies)
- 74M records FY2019-2024
- Tracks who receives federal funds
- Primary spending data source

FAC.gov

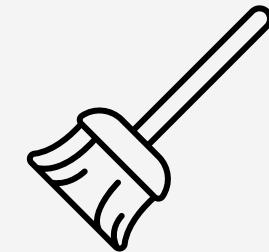
- Single audit findings and compliance
- 57,448 entities FY2016-2024
- Required for entities receiving \$750k+

SAM.gov

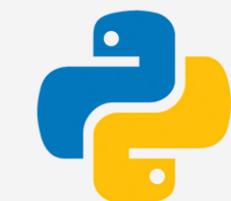
- Excluded & debarred entities
- Governmentwide exclusions
- 167,000 records
- 33,780 entities with UEI
- Public Extract V2 (Dec 27, 2025)



Connected by UEI - Unique Entity Identifier



Cleaned with Alteryx and Python for ML



The Process

Data Integration Findings

FAC + SAM

- Prince Hall Village Charitable Trust
 - Excluded: July 1997 (HUD)
 - In FAC audit data: 2022
 - Gap: 25+ years
- Finding
 - Intra-agency gap: HUD excluded the entity, yet HUD continued funding it for 25 years without detection

SAM + USA Spending

- Amerihost Services, LLC
 - Excluded: September 2021 (EPA)
 - HUD awards: FY2022-2024
 - Amount: \$759,584 (15 awards)
 - Gap: 3+ years
- Finding
 - Cross-agency gap: EPA banned the entity, yet HUD payments continued. Exclusion data doesn't flow between agencies.

The Process

Data Visualization

- Developed research questions
- Created visualizations in Tableau
- Summarized insights and notable patterns



How is funding concentrated across states?

- Top 10 states, led by California, New York, and Texas, collectively receive more than half of all federal grant dollars

Which Federal Agencies are distributing the most funding?

- Dept of Health & Human Services (HHS) distributes \$3,579B in total funding, leading in funding distribution

How has funding changed over time?

- Approximately a 151% increase in funding between 2019 and 2021 for pandemic relief, which then decreased by 29.8% by 2022 and then stabilized.

The Process

Predictive Audit Findings Model

Model

Algorithm: HistGradientBoosting (HGB) classifier

Unit of Prediction: Entity-year (recipient in year t → predicts findings in t+1)

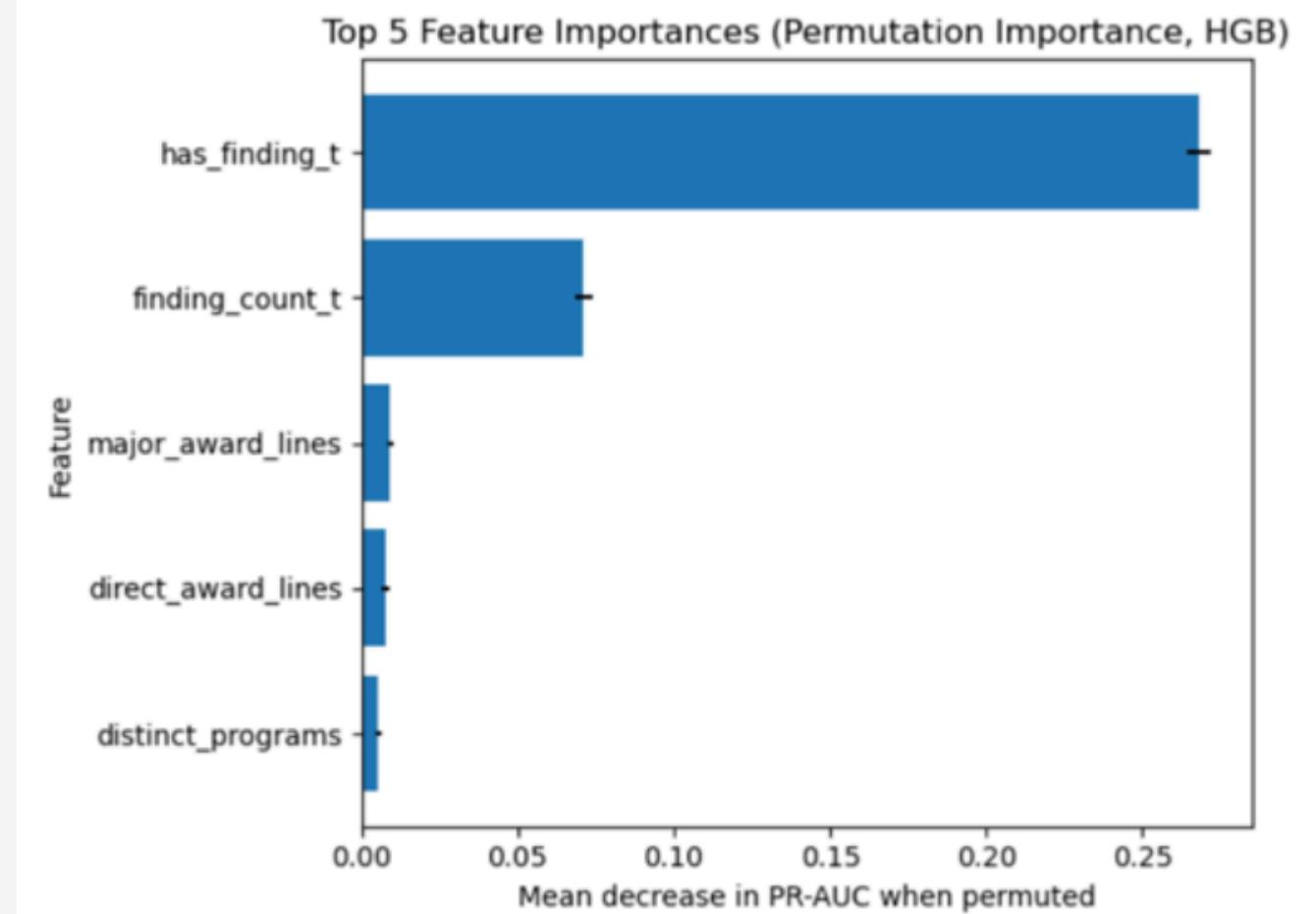
Inputs: prior findings indicators/counts, award line counts, funding totals/mix, program & agency diversity

Performance

ROC-AUC: 0.7656
PR-AUC: 0.5439

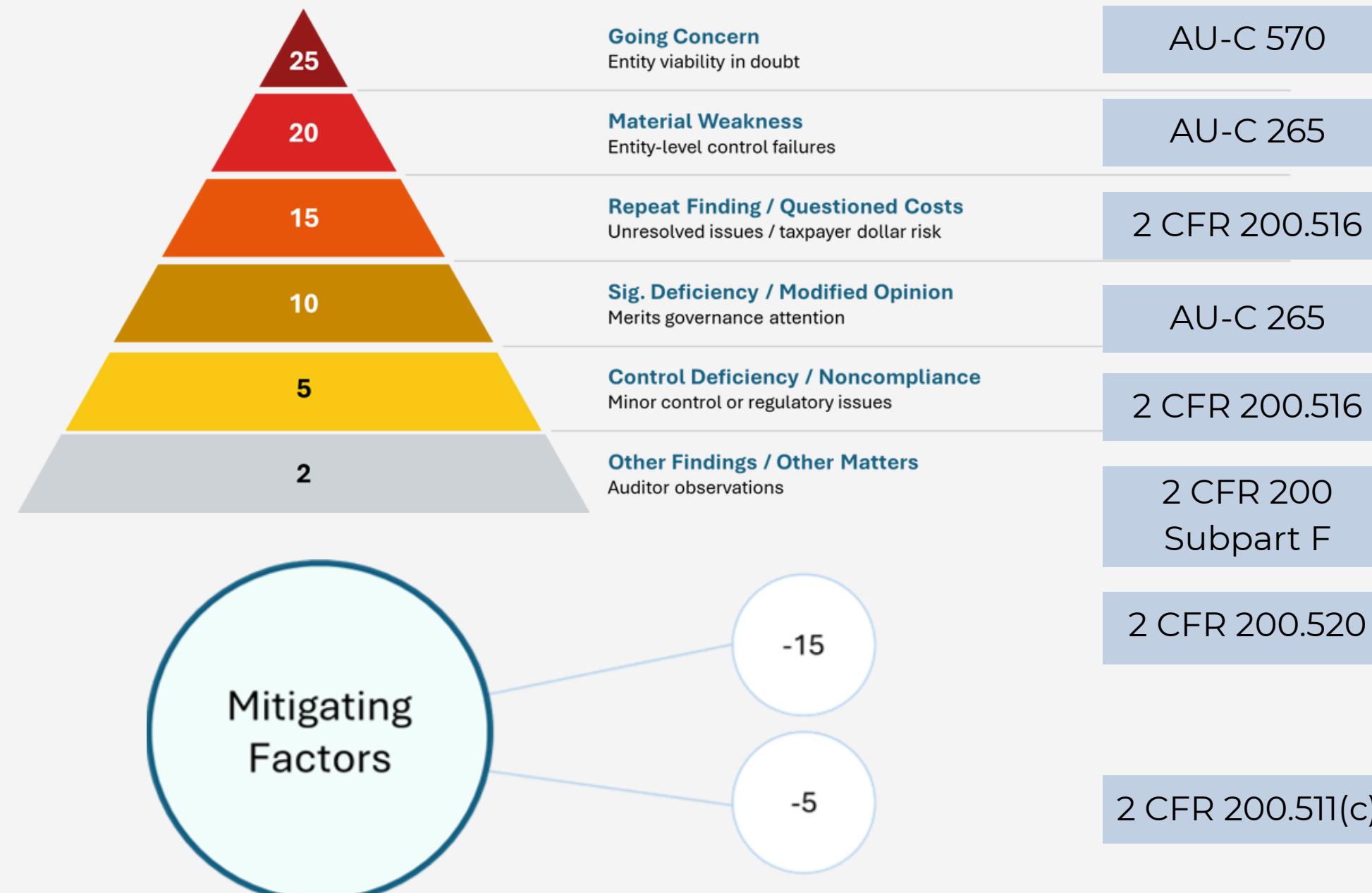
Key Signals

- Prior finding flag / finding count
- Award-line volume (# of direct award lines)
- Breadth of funding (distinct programs/agencies)
- Scale of funding (program totals / major award lines)

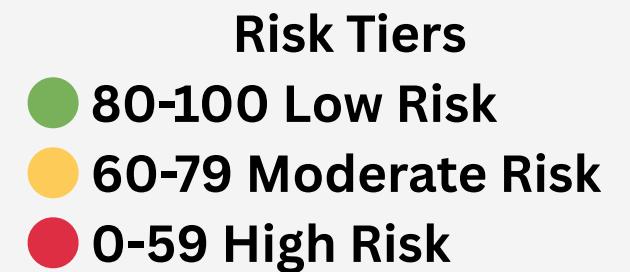


The Process

Audit Health Score



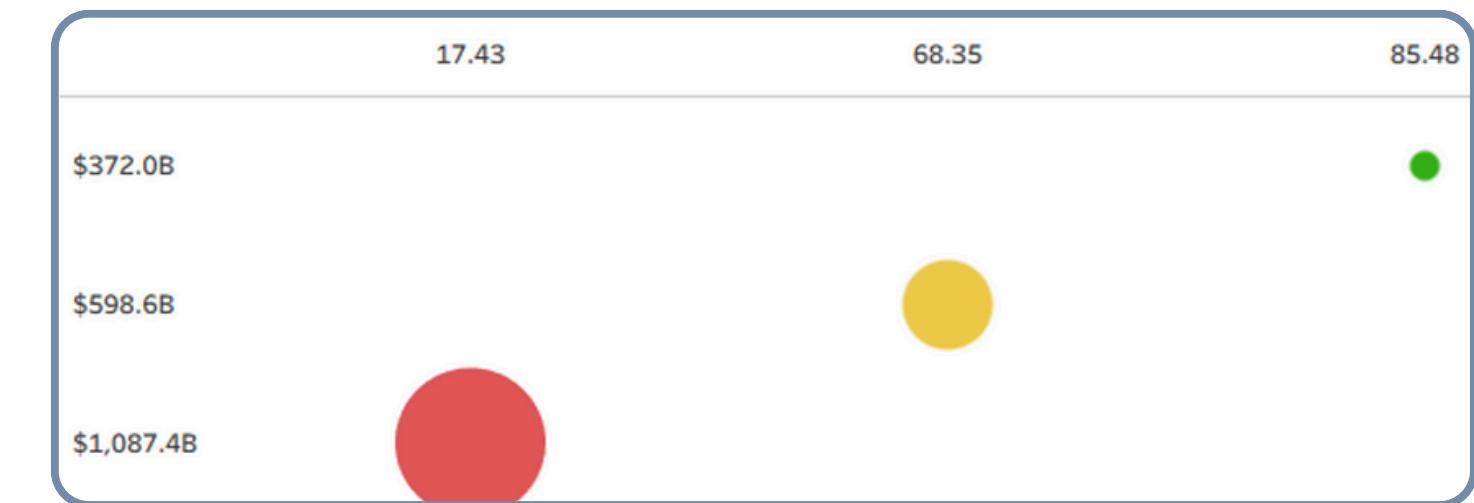
**Score = 100 - Risk Points
(Capped 0-100)**



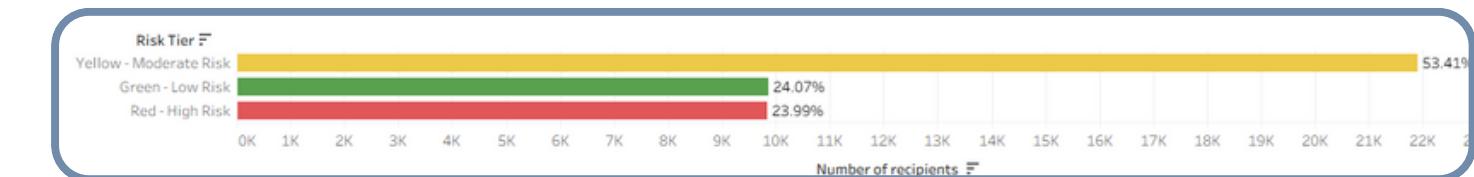
Risk Tiers

- \$1.09 trillion in taxpayer dollars went to recipients that auditors repeatedly flagged.

Federal Funding by Audit Health Score



Recipient Distribution by Risk Tier



The Process

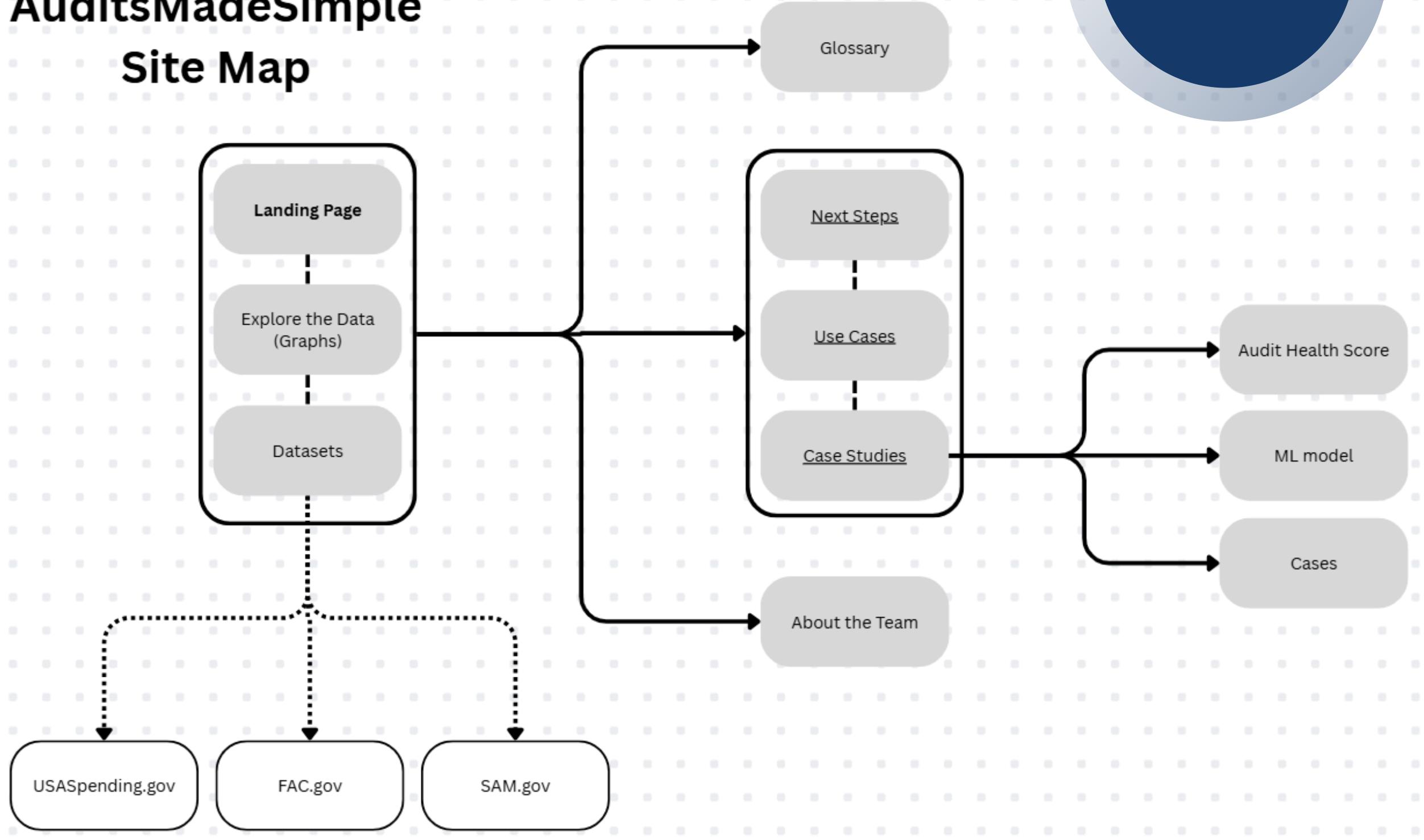
Website Development

- Outlined project scope and requirements
- Developed, tested, and iterated on features
- Mapped user flow and site navigation
- Organized educational content

Tech Stack

- Code Generation Assistance: Google Gemini
- Programming Languages: html, CSS
- Website Hosting: AWS

AuditsMadeSimple Site Map



The Process

Educational Resources

Glossary Terms

- compiled relevant terms from data fields, government websites, and audit reports
- researched definitions and created glossary cards

Data Source Guide

- created data source descriptions
- outlined what each dataset can help answer to guide users in navigating federal websites
- linked to each website for accessibility

User Stories

Primary User: The Non-Technical, Curious Civilian

Adjacent Users:

- Community Advocates and Nonprofit Leaders
- Educators and Students
- Journalists and Policy Researchers

“I’m concerned about a non-profit serving my community and I want to take a look at an audit but I don’t know where to find it.”

“I’m doing research on Federal Assistance spending for a project, but the data is overwhelming. I want to get a quick picture of funding concentration.”

“I’m exploring single audits in FAC.gov, but I don’t understand the terminology, I’d like to familiarize myself with audit terms before diving in.”

Expected Outcomes



Improved **literacy** through dataset visualization, audit terminology, and background information.



Increased **engagement** by providing a navigation channel among government websites.



Enhanced **accessibility** for a non-technical audience by contextualizing data with case studies and real audit findings.

Conclusion

This project is a scalable, centralized platform that improves public understanding of audit data, lowering barriers to government engagement.

However, this is just the tip of the iceberg, with public datasets the possibilities are endless!

[Explore the website →](#)

Appendix



Data sources

- [USAspending.gov: API docs + data dictionary](#)
- [Federal Audit Clearinghouse \(FAC\): dataset overview + files used](#)
- [SAM.gov Exclusions: public extract documentation \(UEI coverage limitation\)](#)
- ["Allowable Data Sources " \(competition rules screenshot / doc\)](#)
- [Data dictionaries used in this repo \(FAC, USAspending, Merged outputs\)](#)



Audit Health Score

- [Variable Definitions \(fields used in scoring\)](#)
- [Weighting rationale \(risk factors vs mitigating factors\)](#)
- [Score formula and tier thresholds](#)
- ["Problem Entity" rule \(logic and flags\)](#)



Regulatory citations

- [GAO-05-479 \(2005\) reference used](#)
- [GAO-09-174 \(2009\) reference used](#)
- [2 CFR 200 Subpart F references\(audit requirements\)](#)

Appendix



Methodology

- Pipeline overview (Clean --> Score --> Merge) visual
- Join logic: UEI mapping (auditee uei = recipient uei) explanation
- Alteryx workflows (screenshots + the .yxmd files)
- Data acquisition notes (what we downloaded, and why)



ML Model

- ML training dataset build workflow
- Train/test split methodology
- Feature list + feature importance chart (or placeholder if not done)
- Model validation metrics
- Final ML-ready outputs



Case Studies

- Full audit findings for flagged entities
- Supporting documentation: screenshots/tables used in the write-up