Environmental Protection Agency (EPA)

Research Proposal Presentation

By: Efren Lopez, Clarin Sunarjo, Andrea Lopez, Henry Lissner

March 09, 2024

Meet Our Team

Proponents

Efren Lopez

Clarin Sunarjo

Andrea Lopez

Henry Lissner

Agenda

01 Problem Statement

02 Stakeholders

03 Data

04 Findings

05 Recommendation

problem to solve

Over the last three years, how have EPA contract funds been distributed across industries, and which have gone up or down? What trends exist in this funding? How can we use this information to help small, women, and minority-owned businesses get more EPA contracts?

Stakeholders

01 Small Women-owned, and Minority-owned Businesses

02 Environmental Protection Agency (EPA) + Government Entities

03 Environmental Advocacy Organizations

04 Policy Makers + General Public

Data

The dataset used for analysis was sourced from USAspending.gov, focusing on contract spanning the years 2021 to 2023.

The dataset was compiled by combining files using Python,

- We start with, 39,385 rows and 297 columns

We are using **potential_total_value_of_awards** as **our target variable**

Approach

The data cleaning process involved several steps to ensure the dataset's quality and integrity:

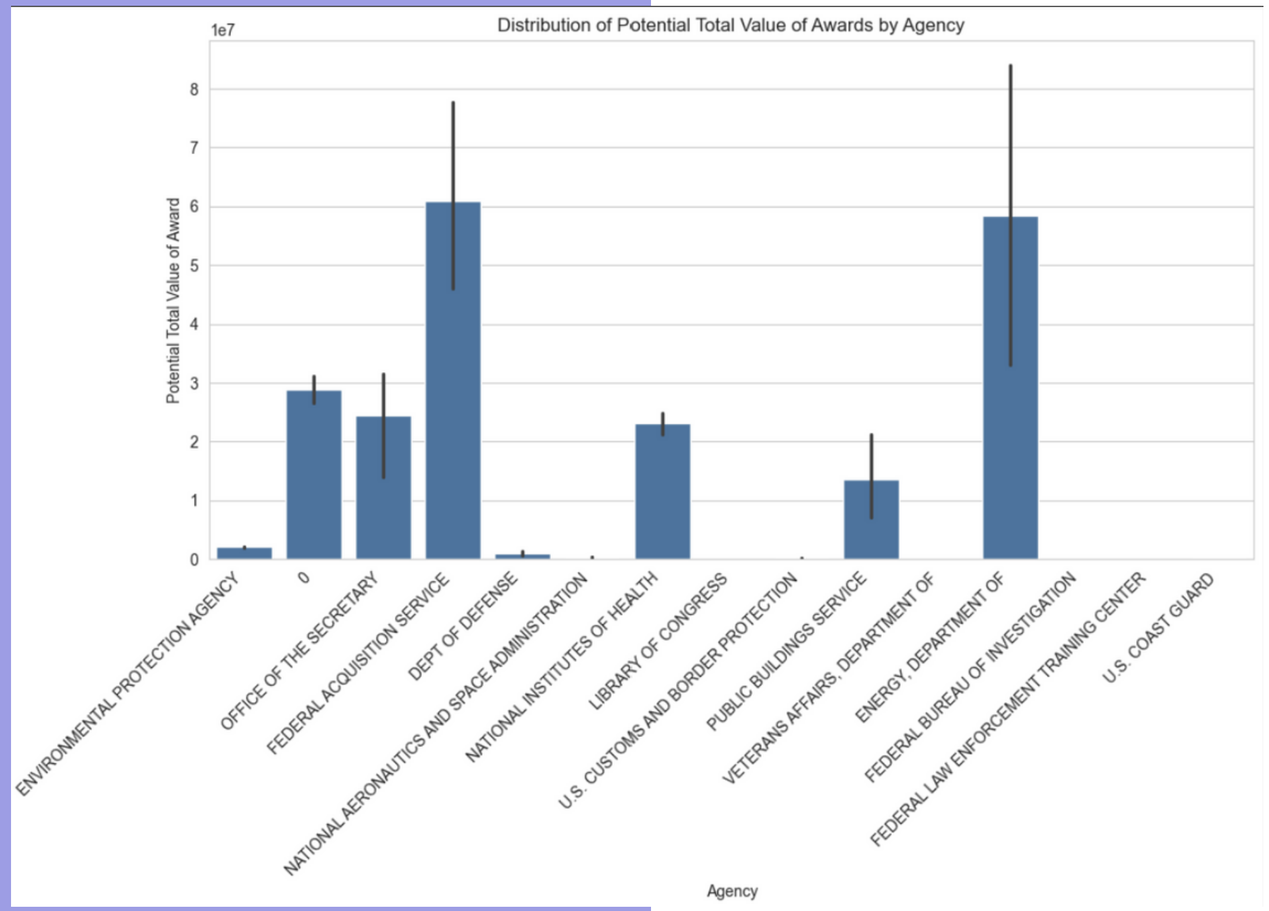
- Info and Shapes: Initial examination of the dataset's structure using the info() function to understand the data types and the shape attribute to determine the number of rows and columns.
- Dropping Missing Values: Rows or columns containing missing values were dropped to prevent bias and ensure the reliability of the analysis.
- Dropping Unimportant Columns: Columns deemed irrelevant or redundant for the analysis were dropped to streamline the dataset and focus on essential variables.

Data Preparation for Predictive Modeling

After cleaning and summarizing the dataset, it was prepared for predictive modeling using a **Random Forest Regressor:**

- Feature Selection: Relevant features were identified based on their importance for predicting the target variable.
- Feature Engineering: New features were created and existing features were transformed to enhance the predictive power of the model.
- Data Splitting: The dataset was split into training and testing sets to evaluate the model's performance on unseen data.
- Model Training: A Random Forest Regressor model was trained using the training dataset to predict the target variable.

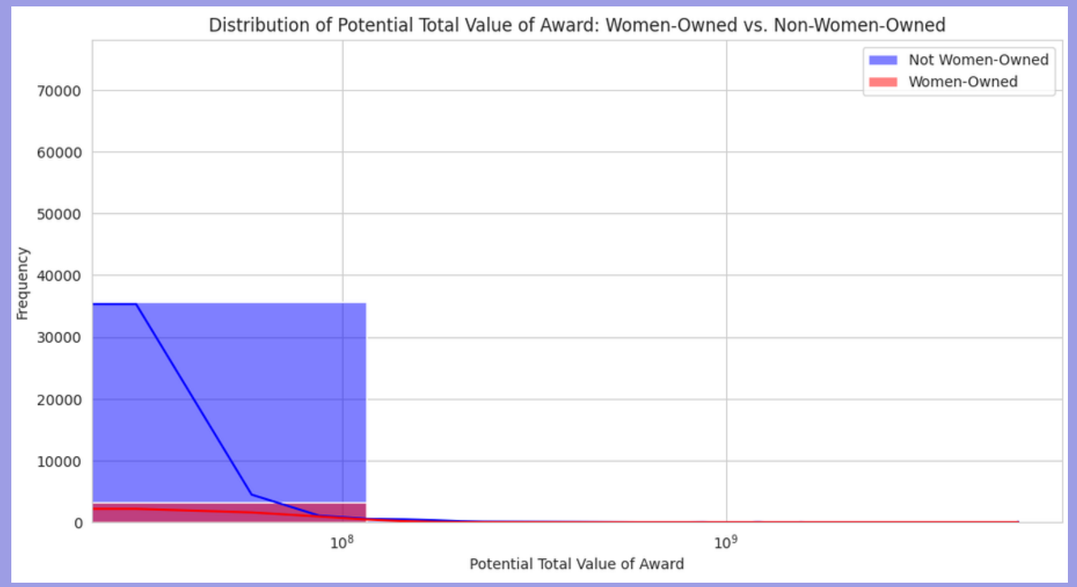
Distribution of Potential Total Value of Awards by Agency



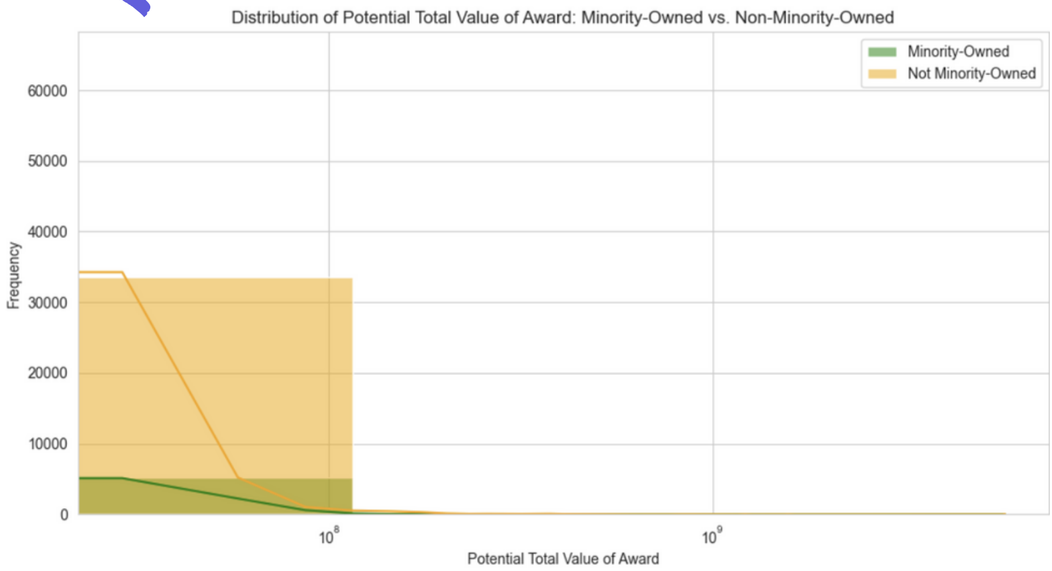
Details of Chart

- Taller bars indicate more funds available for contracts at specific agencies.
- Lines on the bars represent fluctuations in contract values, suggesting variability in funding.
- Target agencies with larger funding pools for better chances at securing contracts.

Distribution of Potential Total Value



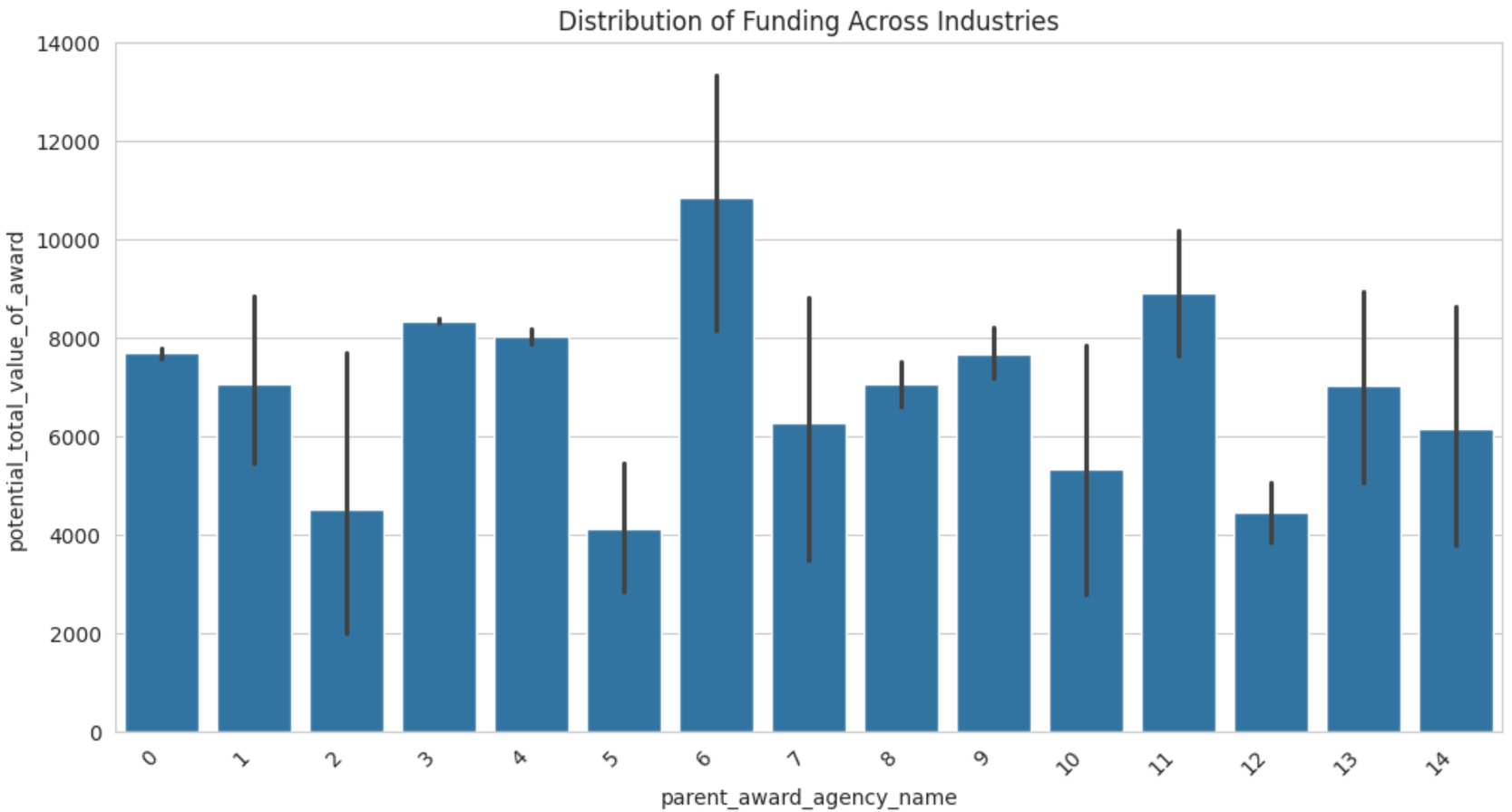
- Women-owned businesses tend to receive lower total value awards compared to non-women-owned counterparts.
- There's a significant disparity in award distribution, with non-women-owned entities occupying a larger share.
- Opportunities for growth in award allocation to women-owned businesses are evident.



- Minority-owned businesses show a concentration in lower-value award categories.
- Non-minority-owned businesses have a broader distribution across various award values.
- Indicates a potential gap in high-value award distributions to minority-owned businesses.

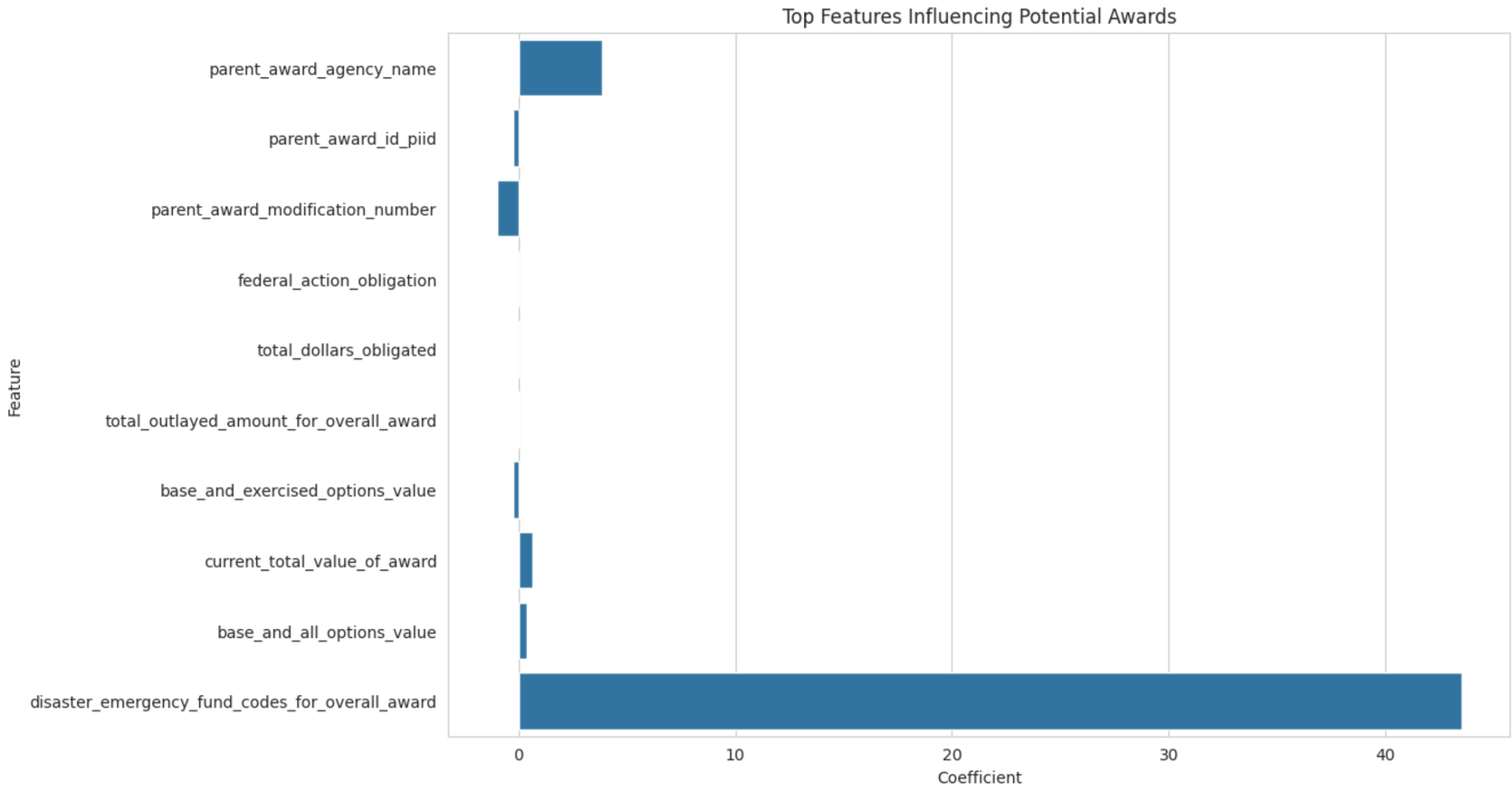
Random Forest Regressor

- Provides a comparative analysis of funding distribution among different agencies.
- Enables strategic targeting of agencies with the highest funding levels to enhance the chances of award reception for target businesses.
- Facilitates informed decision-making regarding agency partnerships to optimize award acquisition opportunities for target businesses.



Feature Importances

- Visualizes key factors influencing potential awards in a feature importance plot.
- Critical for understanding drivers behind award outcomes.
- Prioritizes influential features to optimize resource allocation and strategic planning.



Recommendations

Networking and Partnerships: Encourage collaboration with established firms in high-funding sectors to improve the chances of securing contracts.

Policy Advocacy: If disparities are found, advocate for policy adjustments to ensure equitable access to EPA contracts for underserved businesses.

T**raining and Resources:** Provide specialized training for high-opportunity industries identified in the trend analysis.

New Policy: Breaking down large contracts into smaller, more accessible for small businesses

Certification: to enhance the chance for getting a contract and increase visibility

Thank You