

Predicting Health Center Program Funding

Capstone Project

Katia Lopes-Gilbert

HRSA's Health Center Program

The Health Resources and Services Administration (HRSA) Health Center Program funds nearly 1,400 health centers operating over 15,500 service delivery sites nationwide. **Health centers deliver affordable, accessible, high- quality, and cost-effective primary health care services to medically underserved communities in the U.S., U.S. territories, and the District of Columbia.**



Problem Statement

Becoming an FQHC is a long and complex process that requires certain procedures, organizational structures, and government funding experience as they receive a significant federal grant to become an FQHC.



How much HCP funding could a new entity based on their patients, service area, and organization?



Value Proposition

Help nonprofit health systems to



Improve decision-making
regarding becoming an
FQHC



Advocate for increased
Health Center Program
funding

Data Science Methodology

Source

Clean &
Wrangle

Exploratory
Data Analysis
(EDA)

Preprocess

Model
Development

Methodology

Source

Import data
from HRSA
Reading Room

Decide on
critical tables
and features

Clean &
Wrangle

Deal with
missing data

Create new
summary data

Consolidate 10
tables into 1
table

Exploratory
Data Analysis
(EDA)

Understand
feature
distributions &
correlations

Visualize the
data

Test predictive
power of
features

Preprocess

Impute
missing values

Scale the data

Encode
categorical
data

Create various
train/test
splits

Model
Development

Train models

Make
predictions

Evaluate and
report

Methodology

Source

Import data
from HRSA
Reading Room

Decide on
critical tables
and features

Clean &
Wrangle

Deal with
missing data

Create new
summary data

Consolidate 10
tables into 1
table

Exploratory
Data Analysis
(EDA)

Understand
feature
distributions &
correlations

Visualize the
data

Test predictive
power of
features

Preprocess

Impute
missing values

Scale the data

Encode
categorical
data

Create various
train/test
splits

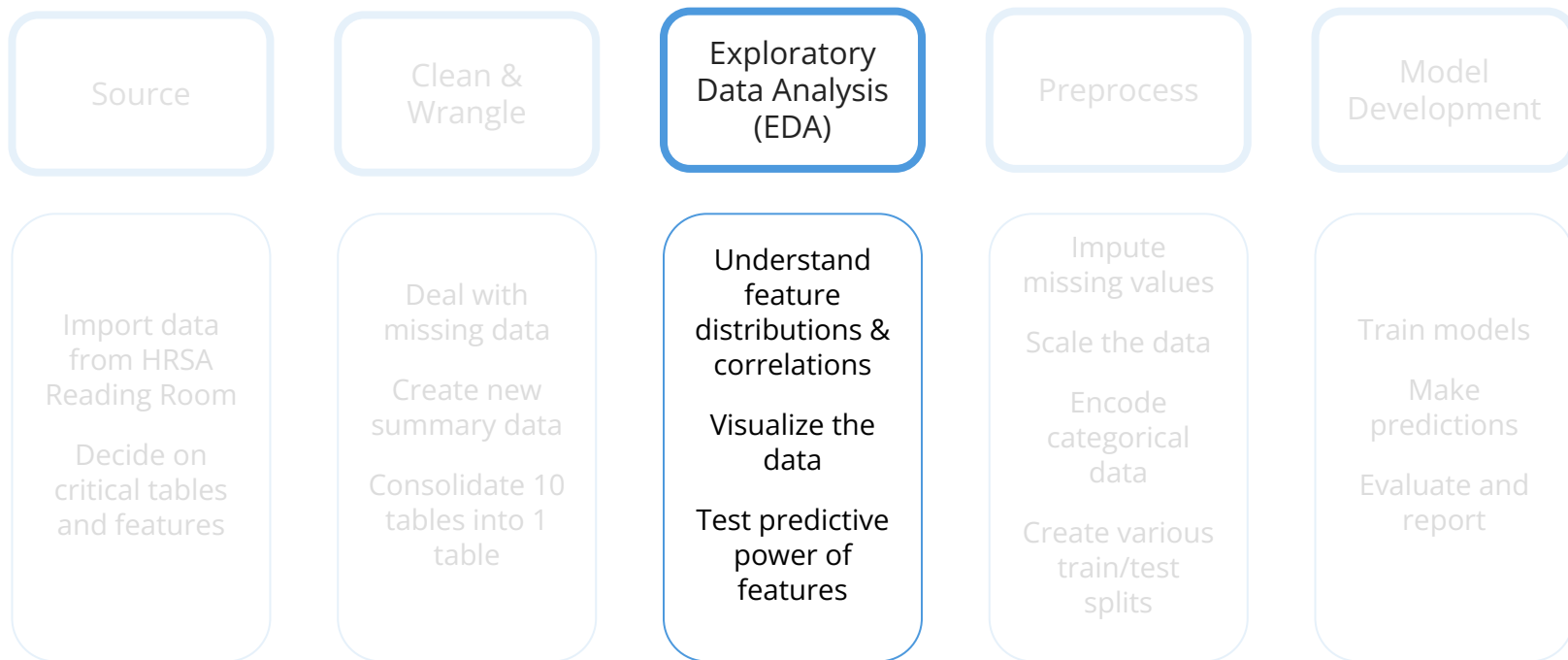
Model
Development

Train models

Make
predictions

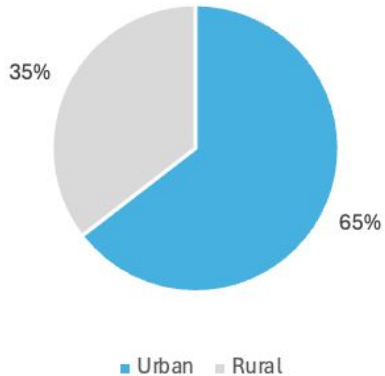
Evaluate and
report

Methodology



Do urban providers receive more HCP funding than rural providers after accounting for total patients served?

Total HCP Funding Distribution between Urban and Rural Providers

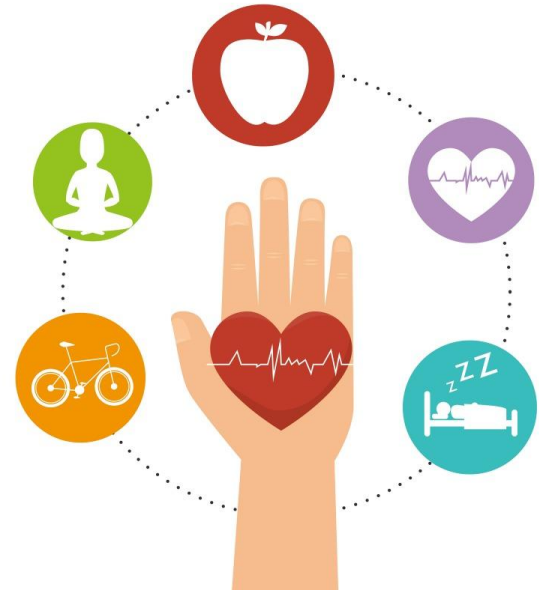


- Urban providers receive more total funding
- Not statistically significantly more after accounting for total number of patients served
- Funding is distributed equitably between urban and rural entities

Do entities that serve larger proportions of patients facing various SDOH receive more HCP funding?

Looked at 15 SDOH factors in the following categories

- Race and ethnicity
- Financial status
- Insurance status and type
- Limited English Proficiency
- Special populations

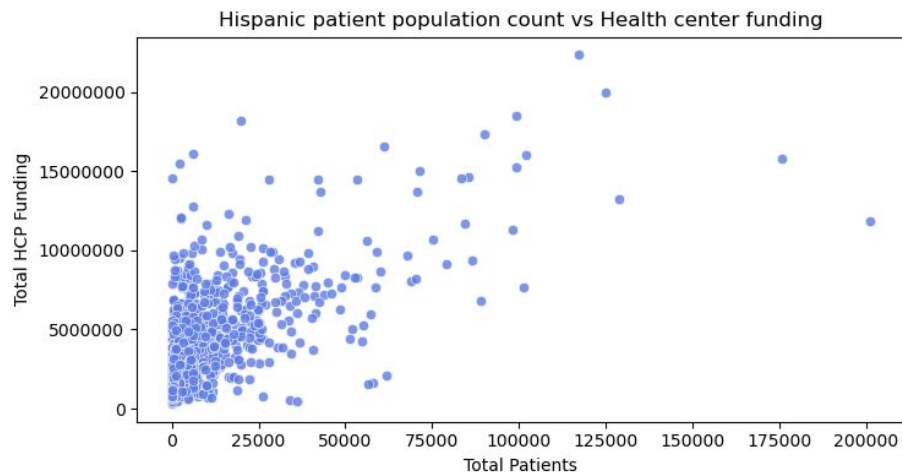


Statistical Tests Methodology

- Compared total HCP funding correlation to total numbers vs proportions
- Created groupings based on proportion quantiles (0-25th, 25th-50th, 50th-75th, 75th-100th)
- Compared average funding received between groups

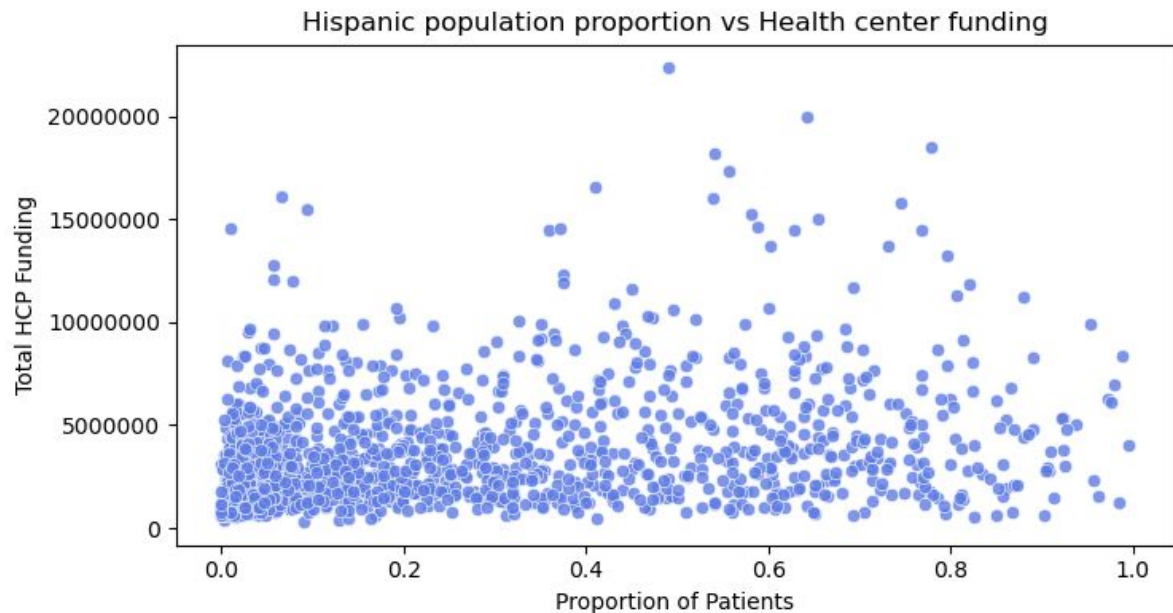


Do entities that serve larger proportions of patients facing various SDOH receive more HCP funding?



Some positive correlation (correlation coefficient of 0.63) between HCP funding and number of Hispanic patients served.

Do entities that serve larger proportions of patients facing various SDOH receive more HCP funding?



Looking at **proportions**, the **correlation** coefficient **decreases** to 0.25 indicating a **weak positive relationship**.

Statistical Test Results

- Higher Need = More Funding
- SDOH measures only explain 0.5% to 4% of HCP funding variations
- Size of operation in terms of # of patients served more important than SDOH factors



Methodology

Source

Import data
from HRSA
Reading Room

Decide on
critical tables
and features

Clean &
Wrangle

Deal with
missing data

Create new
summary data

Consolidate 10
tables into 1
table

Exploratory
Data Analysis
(EDA)

Understand
feature
distributions &
correlations

Visualize the
data

Test predictive
power of
features

Preprocess

Impute
missing values

Scale the data

Encode
categorical
data

Create various
train/test
splits

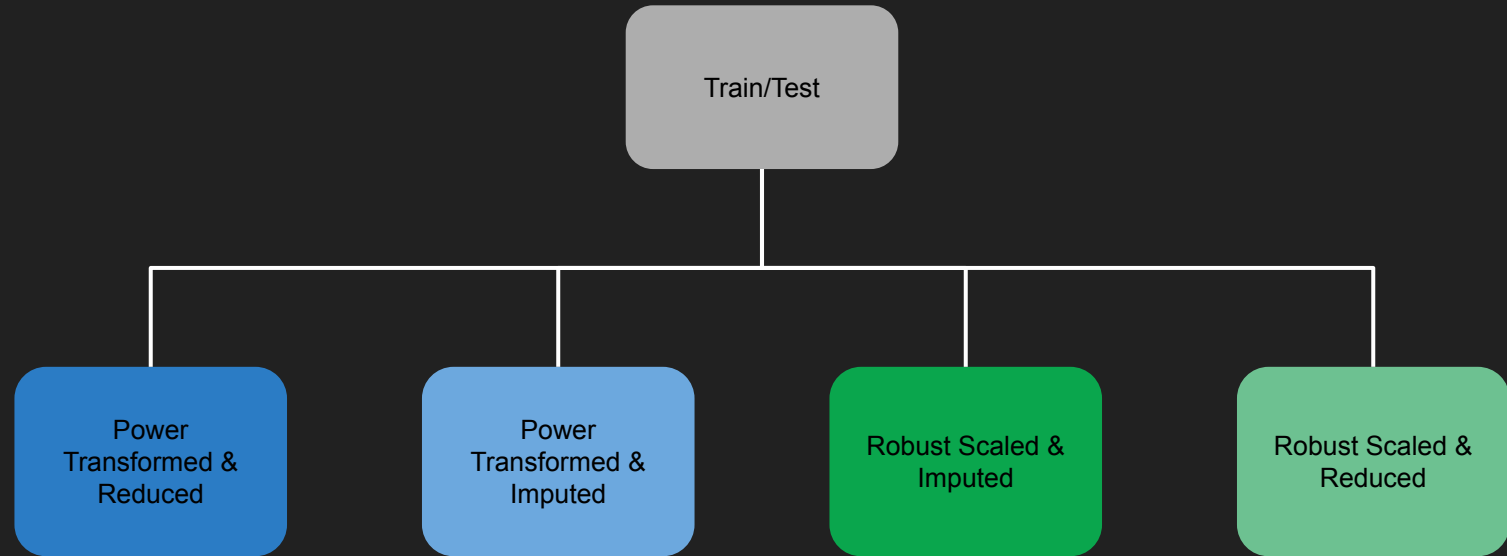
Model
Development

Train models

Make
predictions

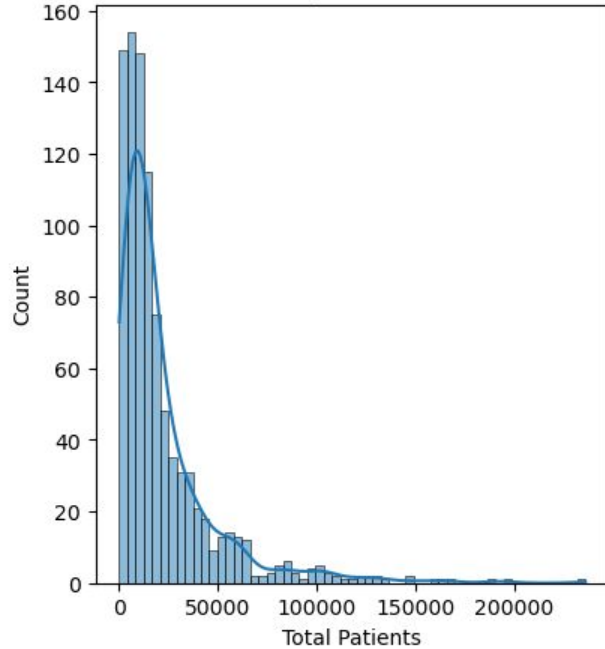
Evaluate and
report

Train / Test Splits: Features

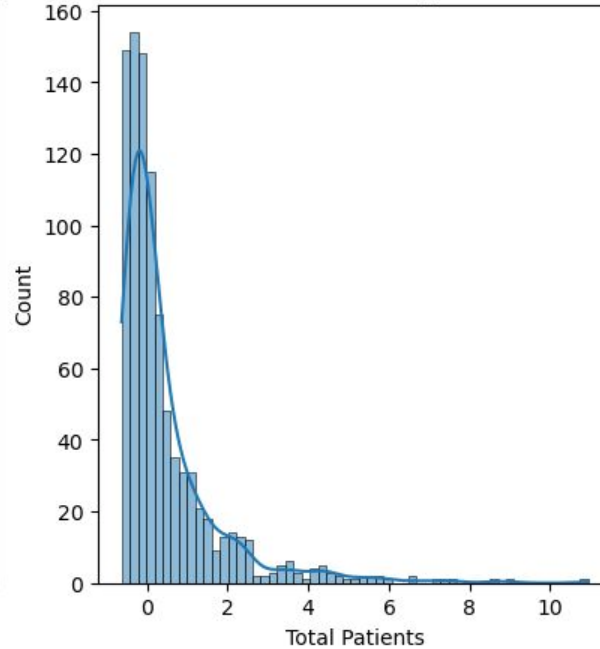


Effect of different scaling techniques on the data

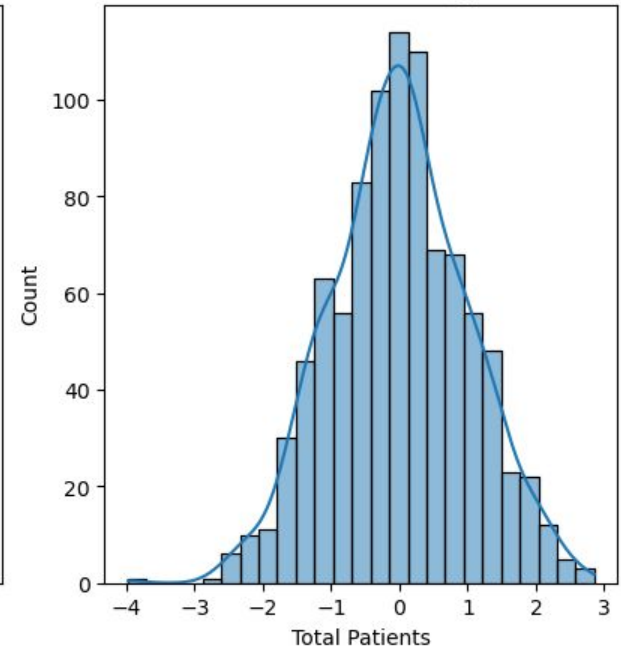
Original Data



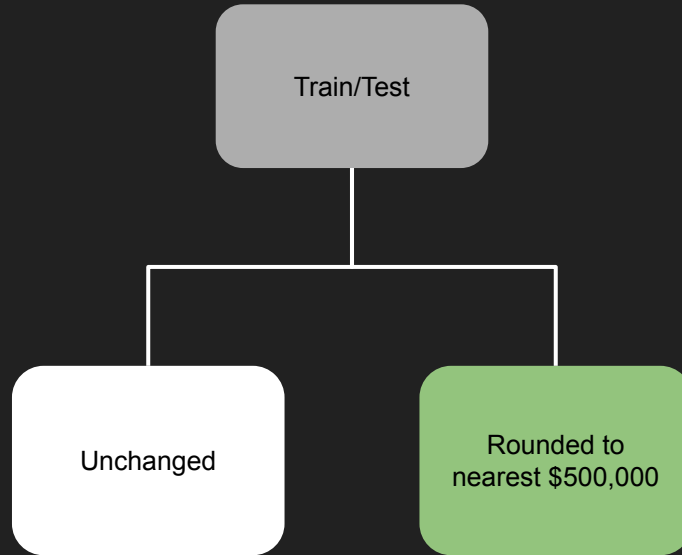
Robust Scaler Applied



Power Transformer Applied



Train / Test Splits: Target Variable



Methodology

Source

Import data
from HRSA
Reading Room

Decide on
critical tables
and features

Clean &
Wrangle

Deal with
missing data

Create new
summary data

Consolidate 10
tables into 1
table

Exploratory
Data Analysis
(EDA)

Understand
feature
distributions &
correlations

Visualize the
data

Test predictive
power of
features

Preprocess

Impute
missing values

Scale the data

Encode
categorical
data

Create various
train/test
splits

Model
Development

Train models

Make
predictions

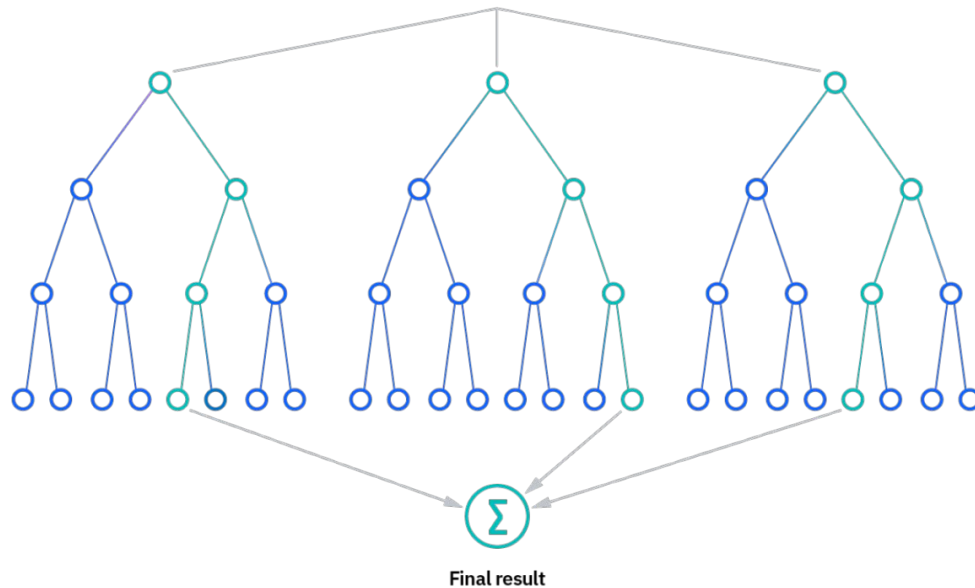
Evaluate and
report

Results

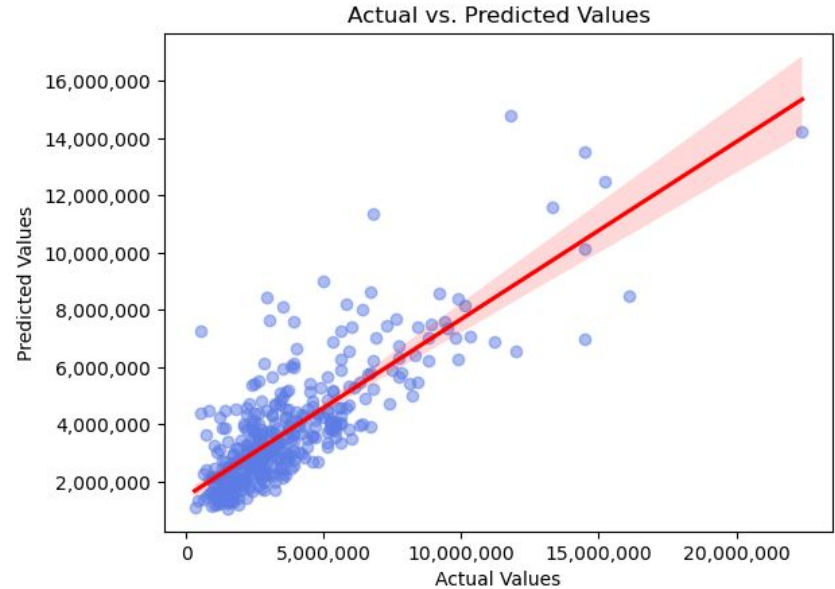
Best Model: **Random Forest Regressor**

Best feature split: **Robust Scaled and Reduced Dataset**

Best target split: **Rounded health center funding**



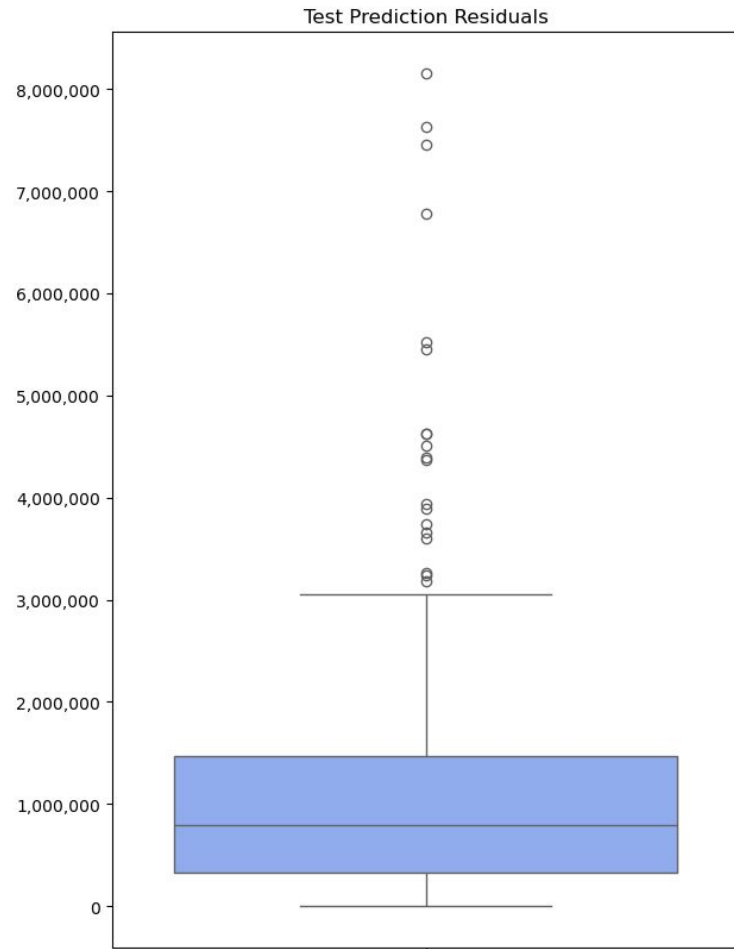
65% effective at
predicting HCP
funding



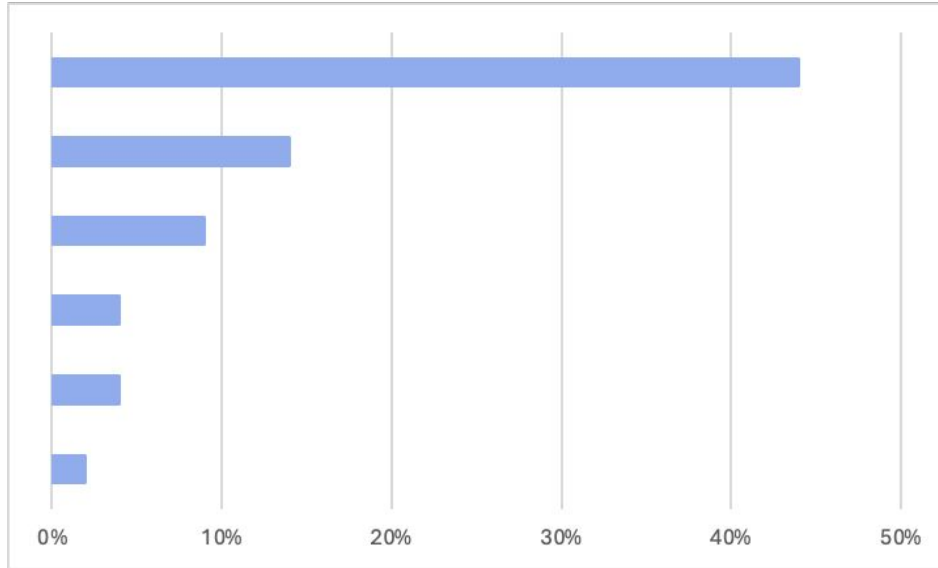
Mean Absolute Error

Error Summary:

- 50% of results were off by \$797K or less
- Lowest error: \$631
- Highest error: \$8M



Most Important Features



44% Total Patients

14% Weekly Hours of Operation

9% Other Grant Funding

4% Number of Sites

4% Proportion of Uninsured Patients

2% Proportion of Racially Diverse Patients

Future Improvements

State-Level Data: Ensuring representative state proportions in each split may improve future model performance.

New Entity Data: Include additional data about entities (site categories, regions, location types, entity age, organization type, proximity to U.S.-Mexico border).

Clustering: Attempt unsupervised clustering methods to discover new data groupings.

Classification: Turn this problem into a classification problem and predict a funding group (ex: between \$500K to \$1M, \$1M-\$2M, etc).

Other Expertise: Meet with other experts who are knowledgeable about Health Center Program funding and get their thoughts on the biggest factors that influence the funding an entity receives to incorporate other data for modeling.

Questions?

