# 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

# Methodology

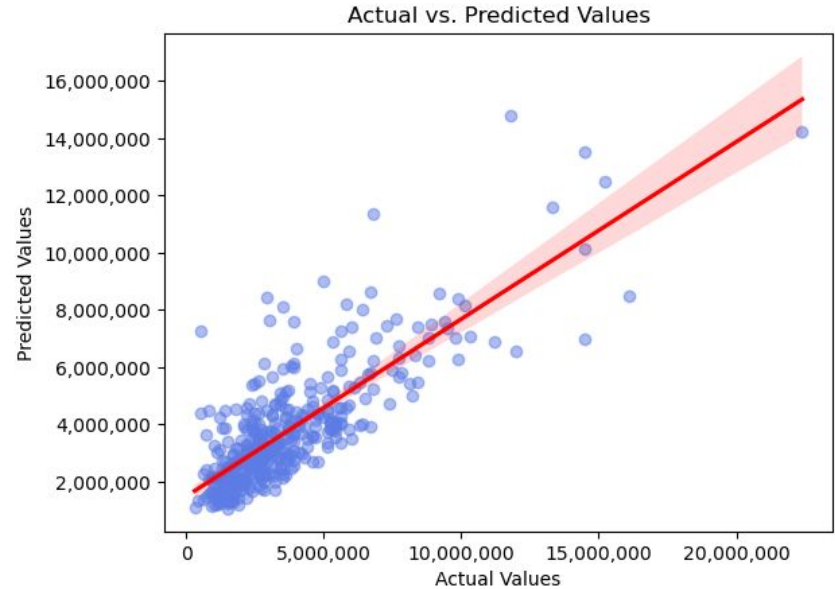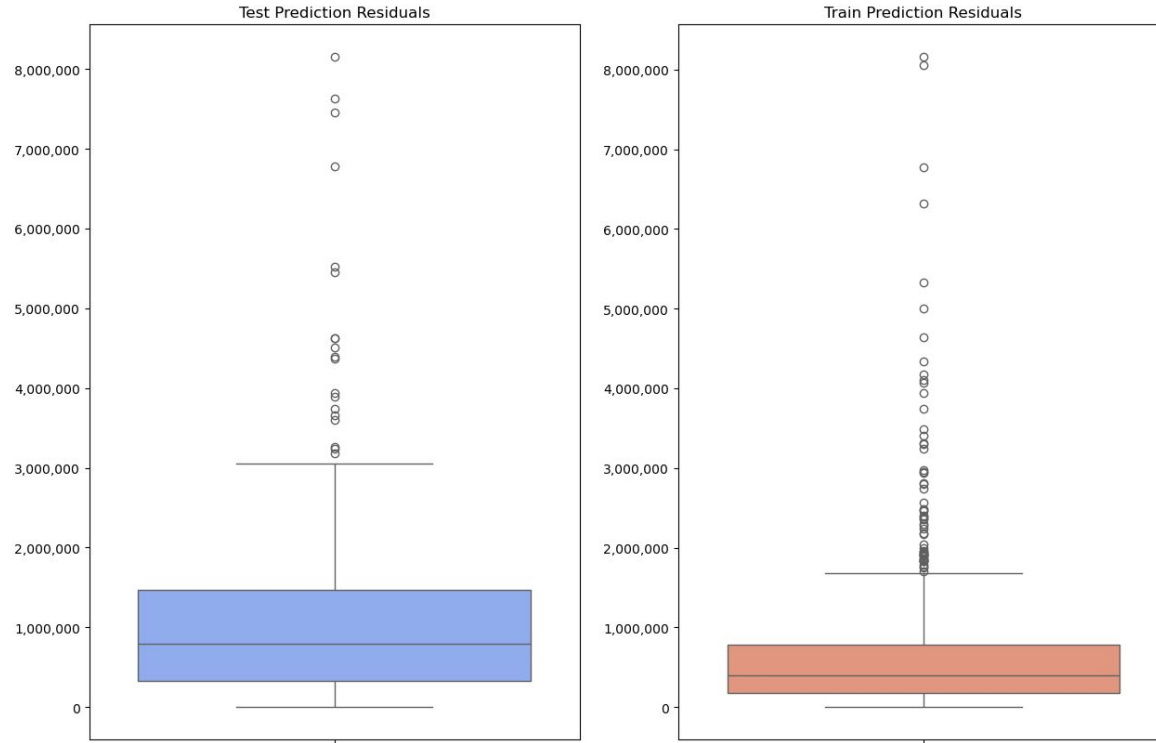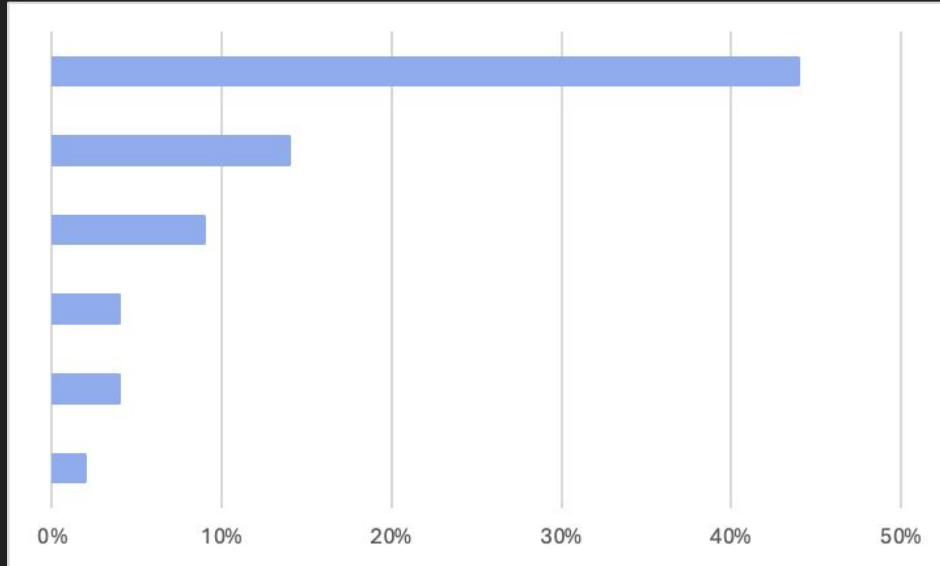| Source | Clean & Wrangle | Exploratory Data Analysis (EDA) | Preprocess | Model Development |
|---|---|---|---|---|
| Import data from HRSA Reading Room<br><br>Decide on critical tables and features | Deal with missing data<br><br>Create new summary data<br><br>Consolidate 10 tables into 1 table | Understand feature distributions & correlations<br><br>Visualize the data<br><br>Test predictive power of features | Impute missing values<br><br>Scale the data<br><br>Encode categorical data<br><br>Create various train/test splits | Train models<br><br>Make predictions<br><br>Evaluate and report |

# Results

# 65% effective at predicting HCP funding



Actual vs. Predicted Values

# Mean Absolute Error

# 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%     Proportions 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 additional data groupings.

**Classification:** Turn this problem into a classification problem.

**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.

# Questions?