

February, 2021

What Affects High School Graduation Rates?

**Yang Xu (a.k.a. Van)
Luke Heeringa
Anel Akiyanova**



AGENDA

- Problem Statement and Use Cases
- Data Collection and Cleaning
- EDA
- Feature Selection
- Model Comparison
- Summary and Conclusion
- App Presentation
- Recommendations



GOAL AND USE CASES

Goal

Determine what features are most important in predicting high school graduation rates

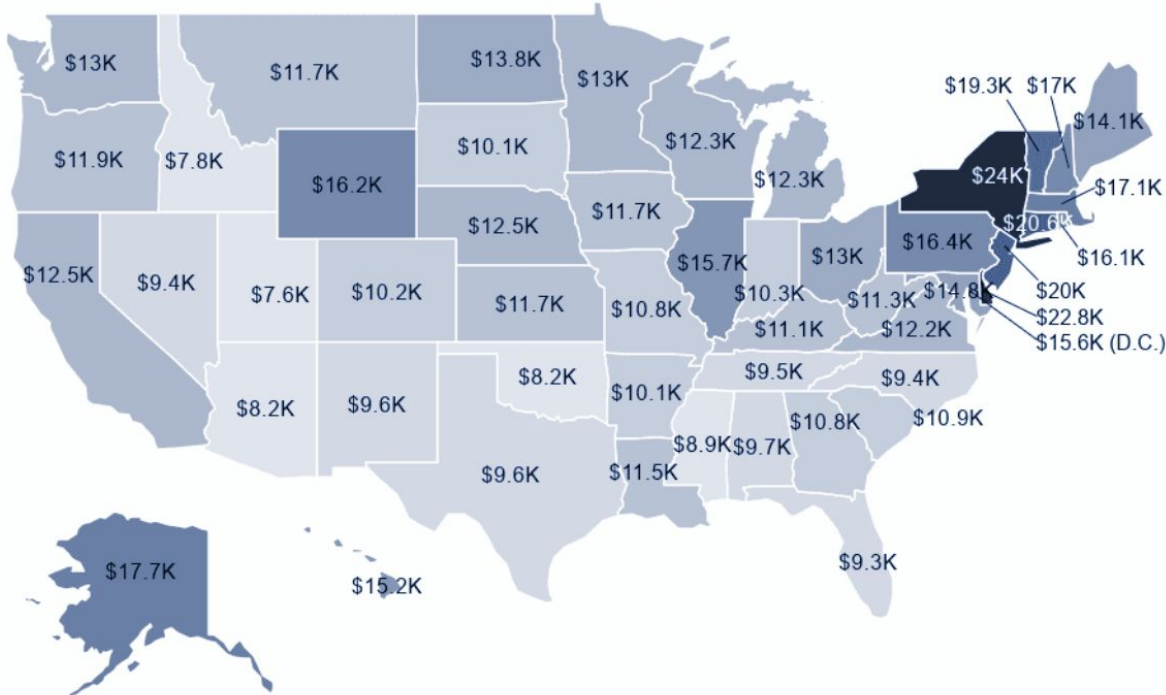
Uses

- Department of Education
- Federal, State, Local Governments
- Education Nonprofit organizations
- General public use



SPENDING PER STUDENT

Public School Spending Per Student

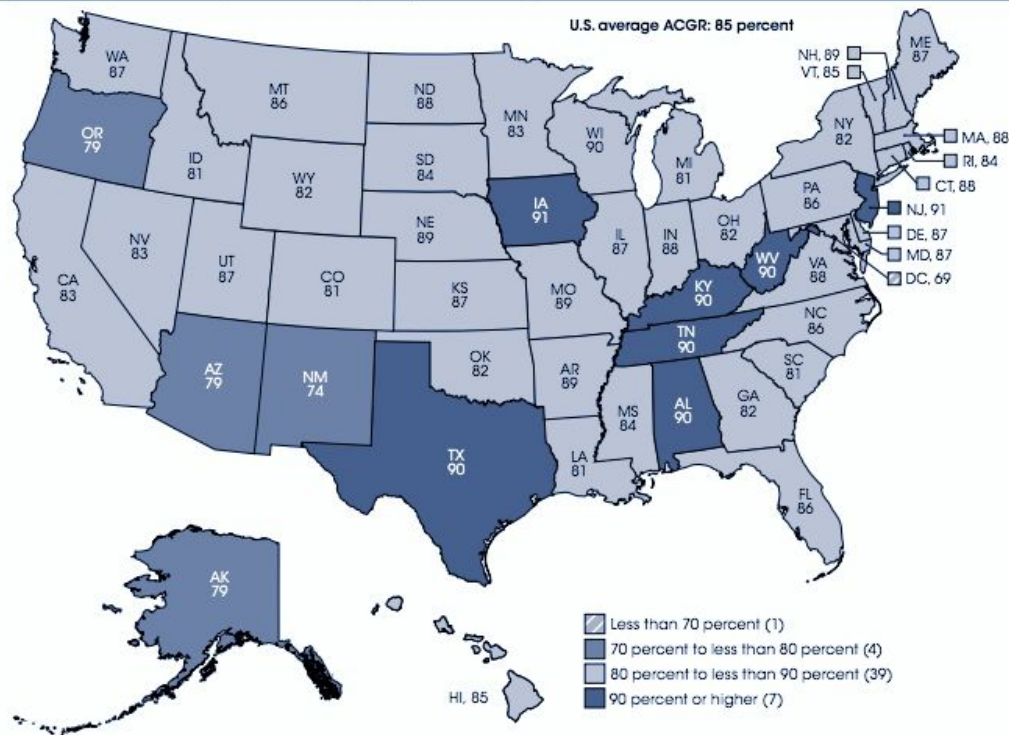


Source: Census Bureau

- **HIGHEST:** New York, \$24,040
- **LOWEST:** Utah, \$7,628
- Funding breakdown:
 - ▷ Federal 7.7%
 - ▷ State 46.7%
 - ▷ Local 45.6%

GRADUATION RATE BY STATE

Figure 1. Adjusted cohort graduation rate (ACGR) for public high school students, by state: 2017-18



- **HIGHEST:** Iowa (91%)
- **LOWEST:** D.C. (69%)
- 3 out of 4 of States between 80 and 90%

NATIONAL CENTER FOR
EDUCATION STATISTICS

DATA COLLECTION
DATA CLEANING
EDA



DATA COLLECTION

- **National Center for Education Statistics**
(NCES): 2017-2018 Financial Data



- District Graduation Rates: 49 States & D.C.
 - ▷ MT Data currently unpublished

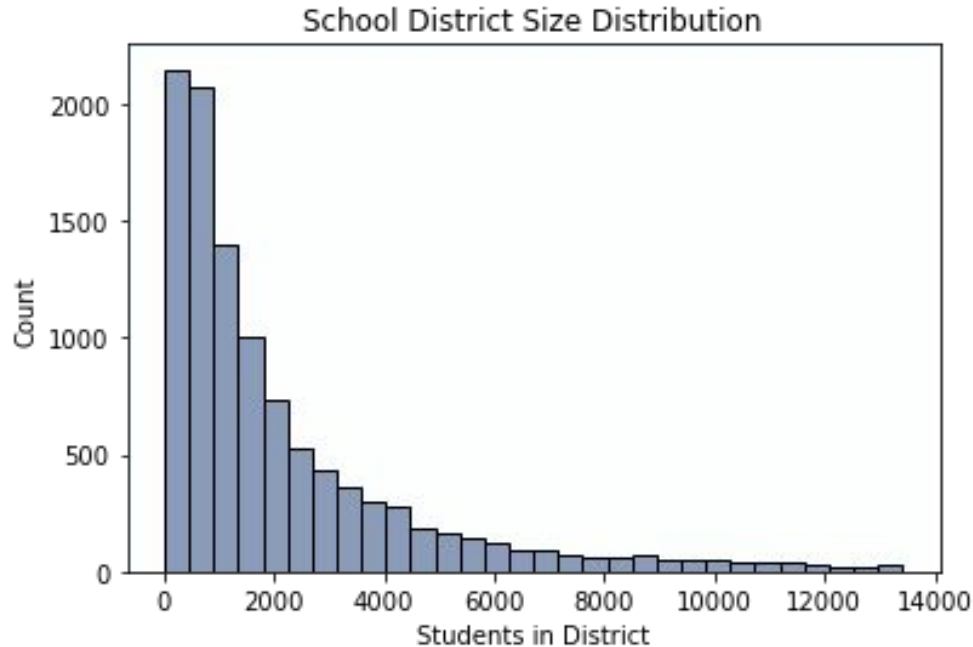


DATA CLEANING

- Merged data by School District
- Target: 4-year Graduation Rate, Full Cohort
 - ▷ Set between 5% and 95%
- Missing financial values mean imputed



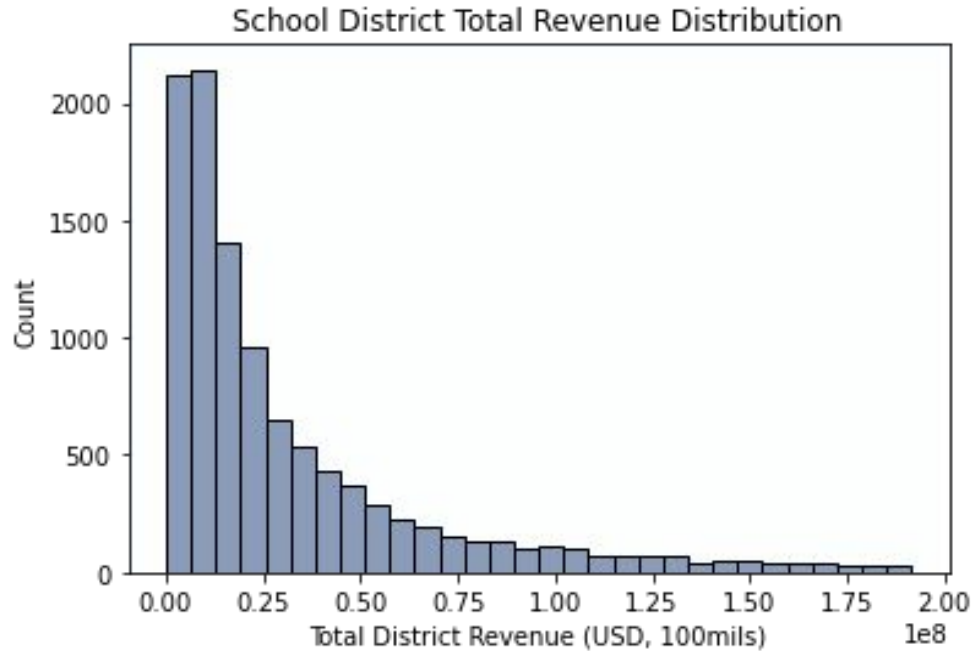
EXPLORATORY DATA ANALYSIS



- **Maximum:**
Los Angeles Unified, CA
621,414 students
- **Minimum:**
Plush S.D. 18, OR
8 students



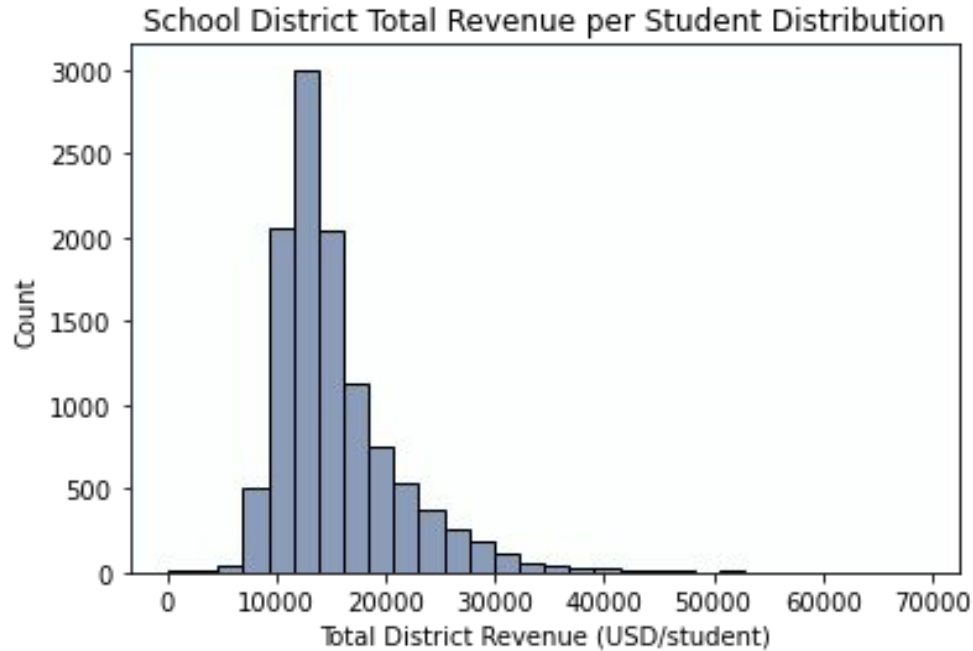
EXPLORATORY DATA ANALYSIS



■ Initial Distribution



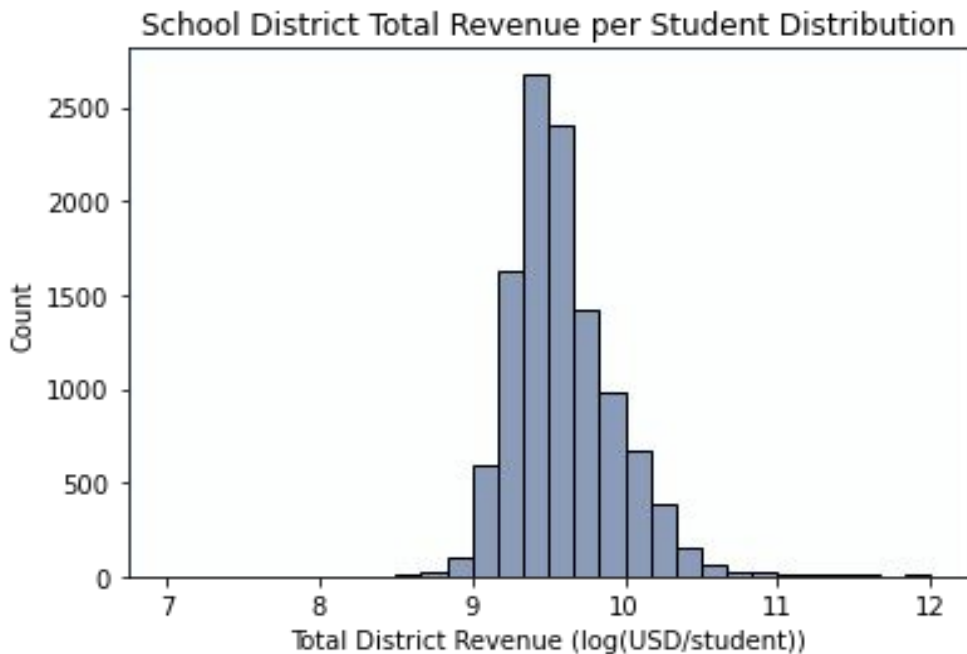
EXPLORATORY DATA ANALYSIS



- Per student distribution



EXPLORATORY DATA ANALYSIS



- Per student distribution log-scaled



EDA: AVERAGE GRADUATION RATES

**Charter-only
Districts:**

72%

**Charter/Non-charter
Mixed Districts:**

80%

**Non-charter
Districts:**

90%

FEATURE SELECTION MODEL COMPARISON



FEATURE SELECTION

Location

State

District

Revenue

Federal

State

Local

Expenditures

Support
Services

Utilities

Textbooks

Technology

Capital Outlay

Retired debt

Supplies

Salaries

Regular programs

Special Ed

Support Services

Administrative

Transportation

Student Support

Business

Food Services

Benefits

Instructional staff

Student Support

Administrative

Transportation

Business

Food Services

Assets

Sinking Fund

Bonds

Other

Debt

Long-term debt

Short-term
debt



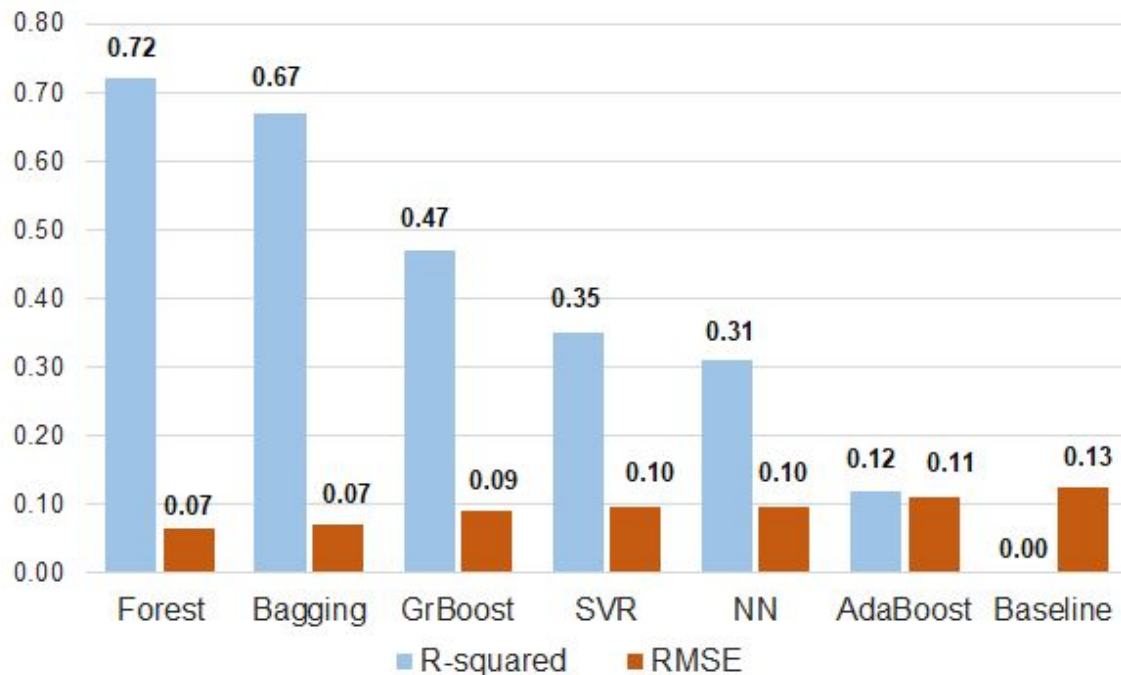
FEATURE SELECTION

- Looked at features that correlate with **graduation rates**
- Noticed multicollinearity between variables
- Dummified categorical features
- Final modeling: 33 predictors



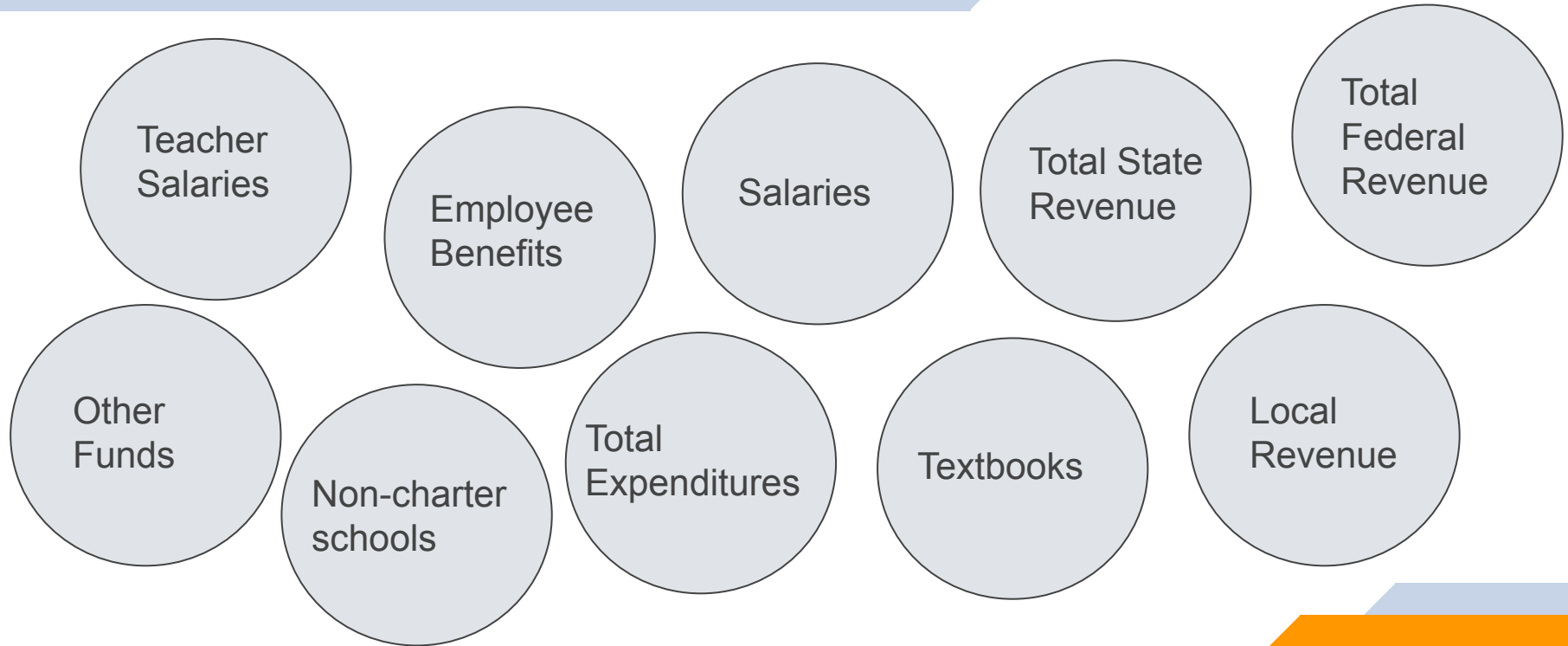
MODEL COMPARISON

- **Random Forest**
performed best
- **Baseline Model**
R2 score 0 RMSE 0.15
- Avoided Linear models
sensitive to feature
collinearity



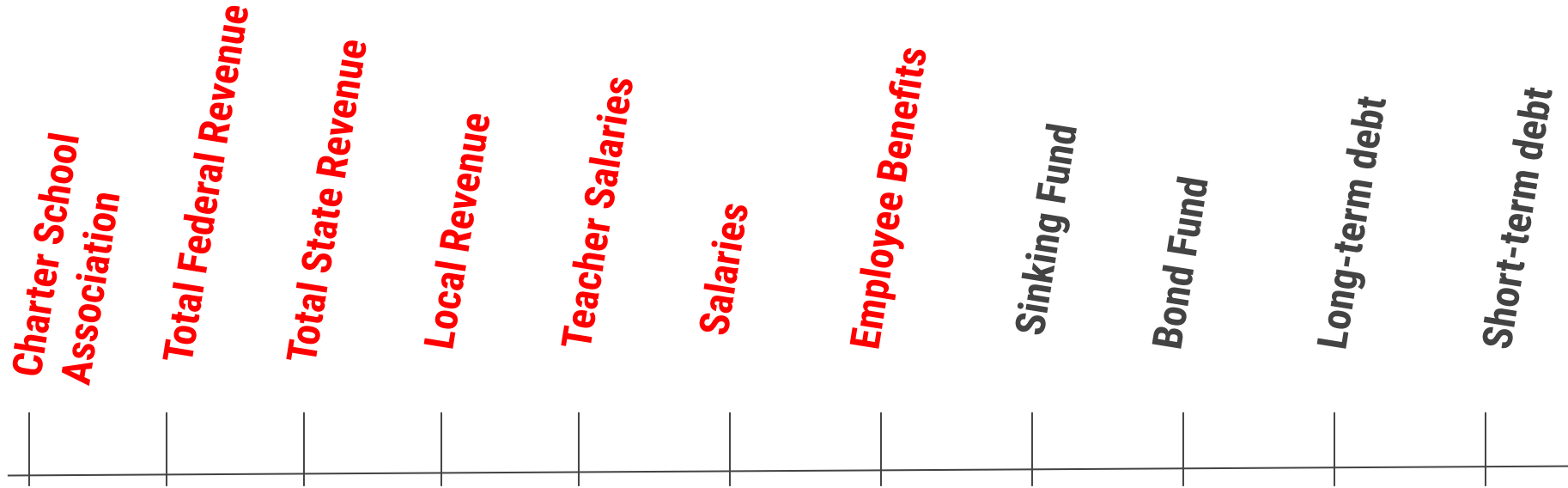


RECURSIVE FEATURE ELIMINATION





PRINCIPAL COMPONENT ANALYSIS





CONCLUSION AND RECOMMENDATION

Based on the feature analysis, **we recommend focusing** on these factors that consistently proven to be reliable predictors of graduation rates:

- Total Federal Revenue
- Total State Revenue
- Local Revenue - District Activity
- Teacher Salaries - support services - instructional staff
- Salaries - School Admin
- Salaries - business/ other
- Employee benefits - support services - school admin
- School Funds

APP PREDICTING GRADUATION RATE



THANK YOU!

questions?