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

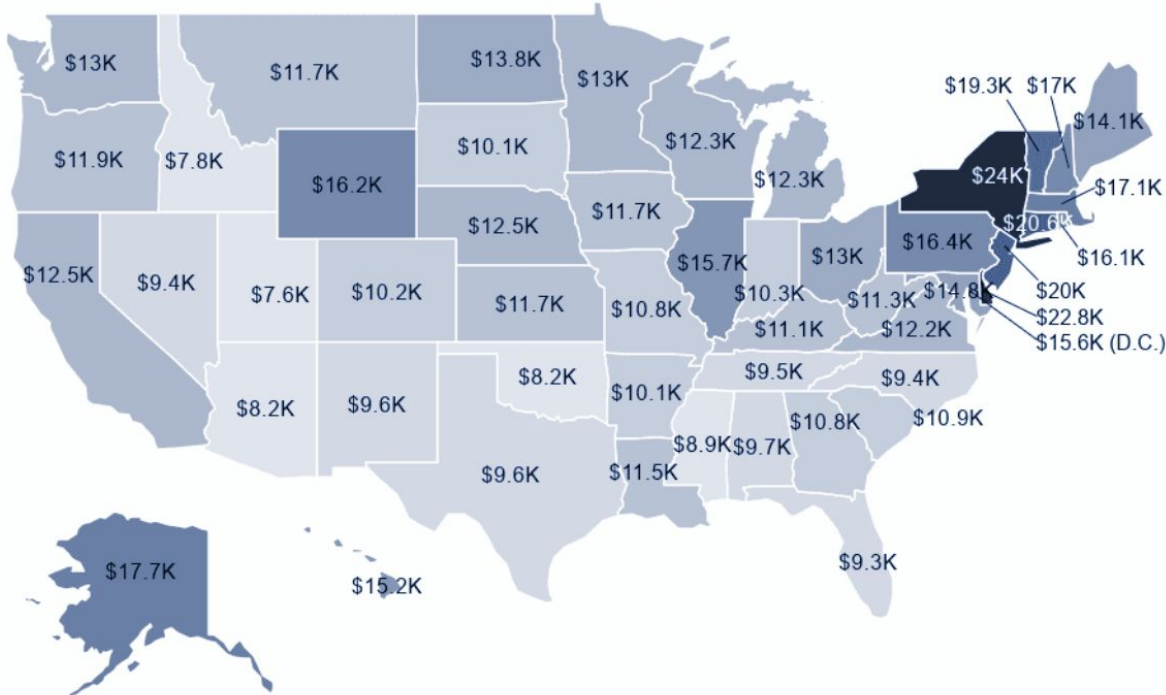Determine what features are most important in predicting high school graduation rates

Uses

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

SPENDING PER STUDENT

Public School Spending Per Student

$13K, $11.7K, $13.8K, $13K, $19.3K, $17K, $14.1K
$11.9K, $7.8K, $16.2K, $10.1K, $12.3K, $24K, $17.1K
$12.5K, $11.7K, $12.3K, $20.6K, $16.1K
$12.5K, $9.4K, $7.6K, $10.2K, $11.7K, $12.5K, $15.7K, $13K, $16.4K
$10.8K, $10.3K, $11.3K, $14.8K, $20K
$8.2K, $10.1K, $11.1K, $12.2K, $22.8K
$8.2K, $9.6K, $9.5K, $9.4K, $15.6K (D.C.)
$8.9K $9.7K, $10.8K, $10.9K
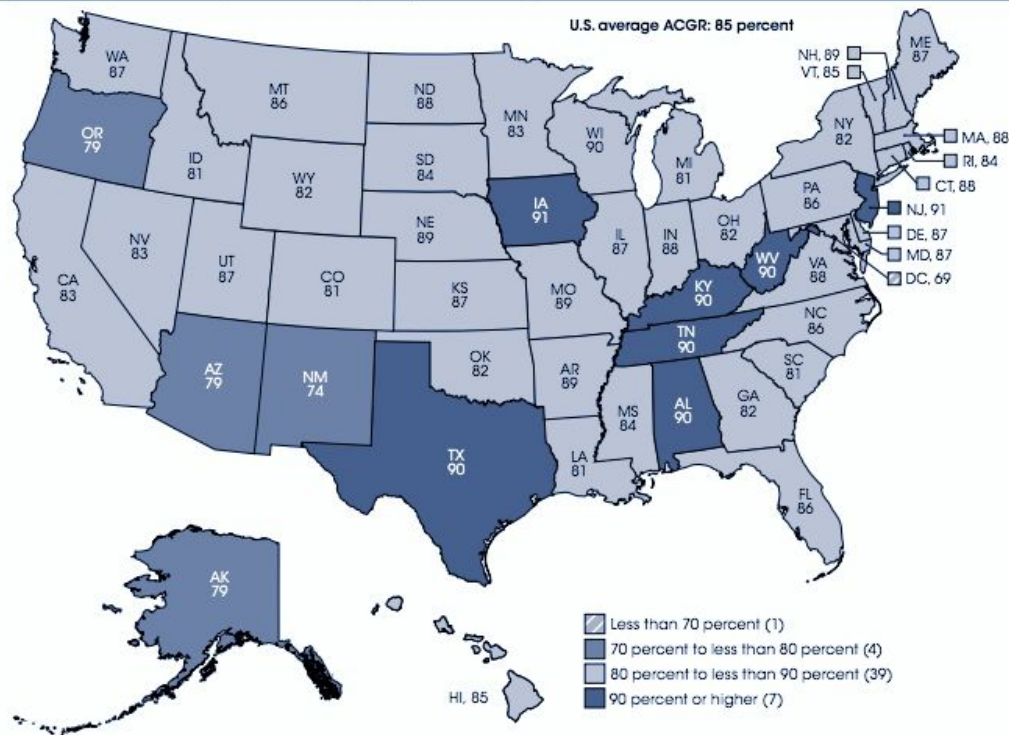$9.6K, $11.5K
$9.3K
$17.7K, $15.2K

Source: Census Bureau

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

4

GRADUATION RATE BY STATE

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

U.S. average ACGR: 85 percent

HIGHEST: Iowa (91%)
LOWEST: D.C. (69%)

3 out of 4 of States between 80 and 90%

DATA COLLECTION
DATA CLEANING
EDA

DATA COLLECTION

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

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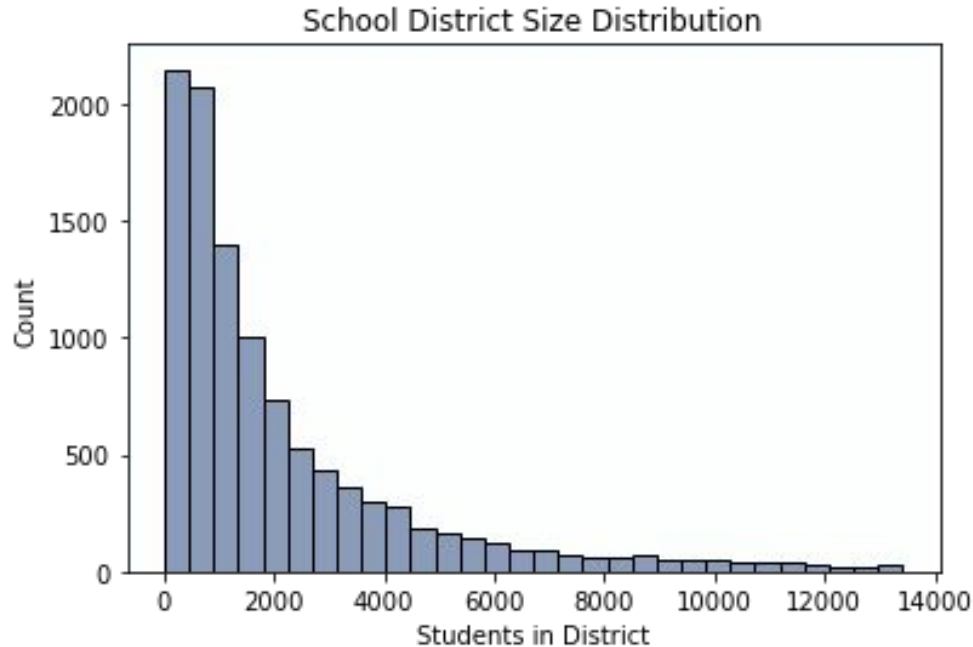 - ▷ MT Data currently unpublished

DATA CLEANING

- Merged data by School District

- Target: 4-year Graduation Rate, Full Cohort

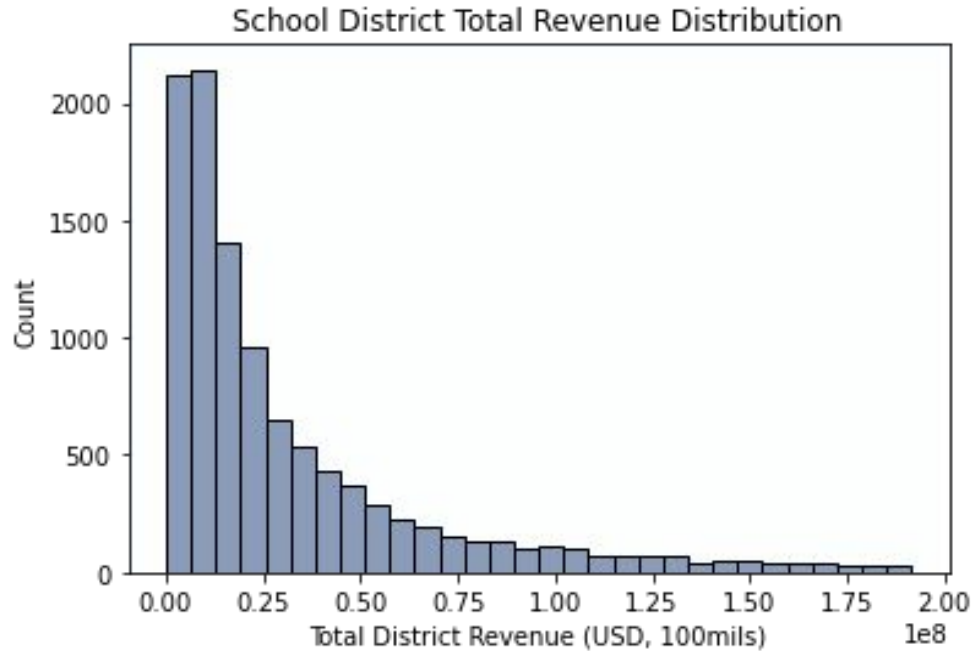 ▷ Set between 5% and 95%

- Missing financial values mean imputed

School District Size Distribution

Maximum:
Los Angeles Unified, CA
621,414 students

Minimum:
Plush S.D. 18, OR
8 students

School District Total Revenue Distribution

- Initial Distribution

School District Total Revenue per Student Distribution

- Per student distribution

EXPLORATORY DATA ANALYSIS

School District Total Revenue per Student Distribution
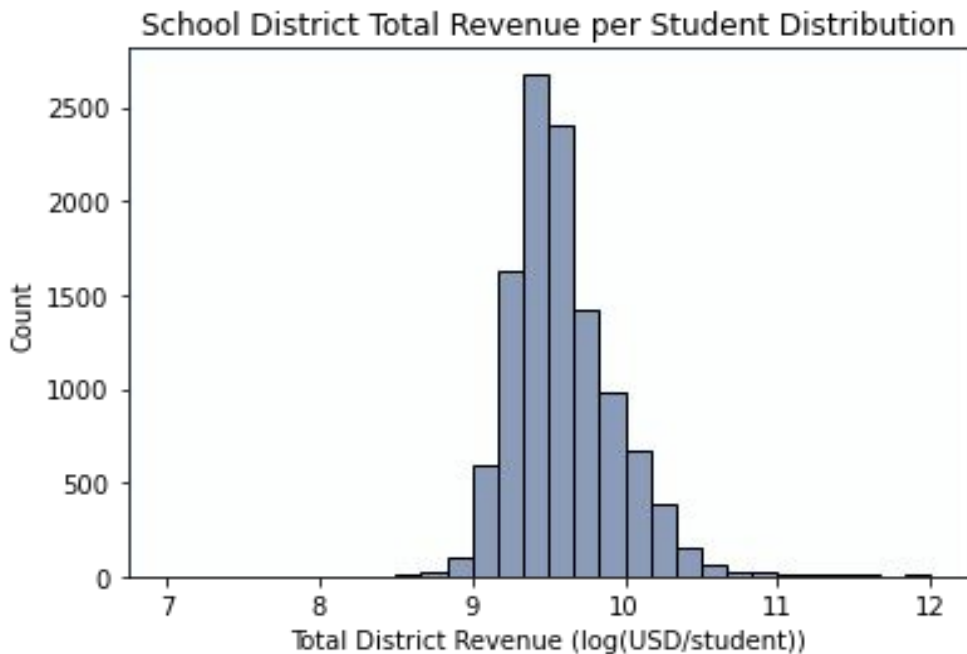
- Per student distribution log-scaled

EDA: AVERAGE GRADUATION RATES

Charter-only Districts:	Charter/Non-charter Mixed Districts:	Non-charter Districts:
72%	**80%**	**90%**

FEATURE SELECTION MODEL COMPARISON

FEATURE SELECTION

Location	Revenue	Expenditures	Salaries	Benefits	Assets	Debt
State	Federal	Support Services	Regular programs	Instructional staff	Sinking Fund	Long-term debt
District	State	Utilities	Special Ed	Student Support	Bonds	Short-term debt
	Local	Textbooks	Support Services	Administrative	Other	
		Technology	Administrative	Transportation		
		Capital Outlay	Transportation	Business		
		Retired debt	Student Support	Food Services		
		Supplies	Business			
			Food Services			

FEATURE SELECTION

- Looked at features that correlate with **graduation rates**

- Noticed multicollinearity between variables

- Dummified categorical features

- Final modeling: 33 predictors

MODEL COMPARISON

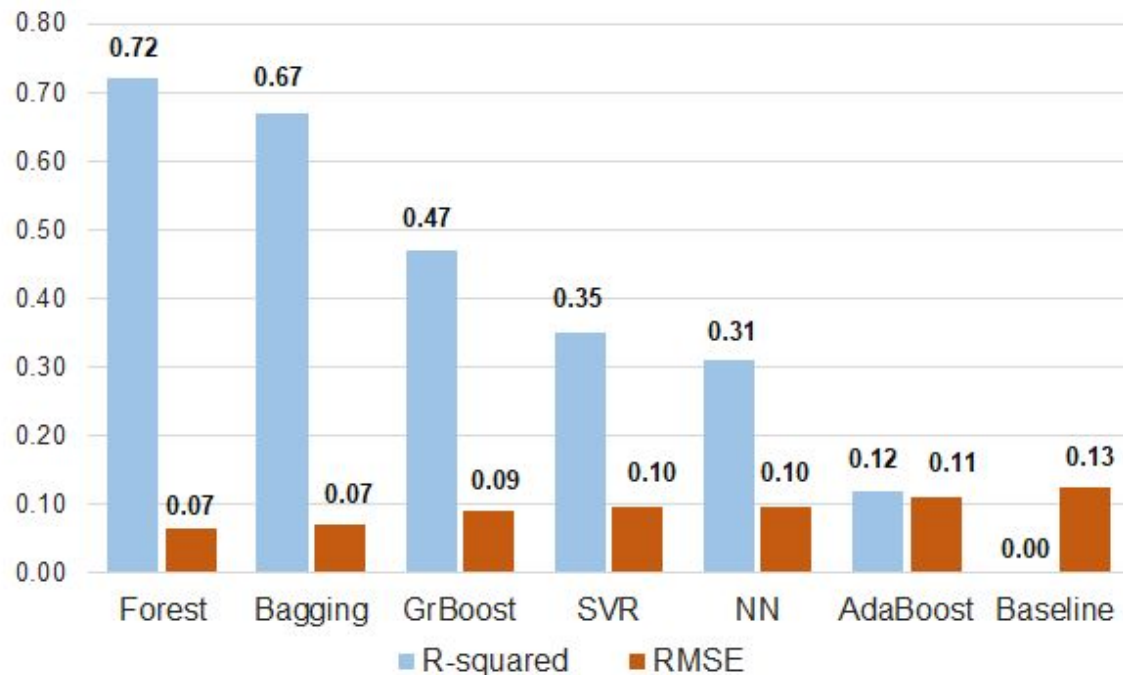- **Top performing model:**
  Random Forest
  R2 score 0.72
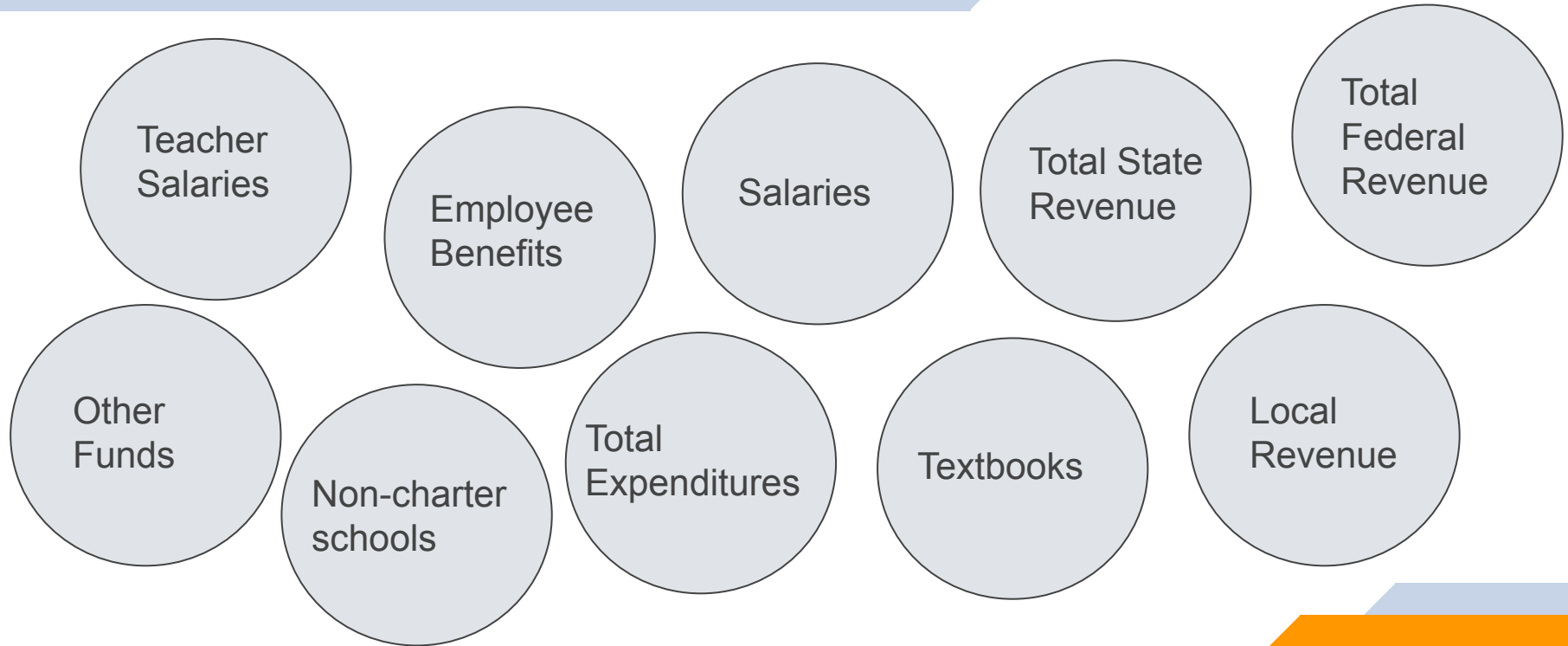 RMSE 0.07

- **Baseline Model**
  R2 score 0, RMSE 0.13

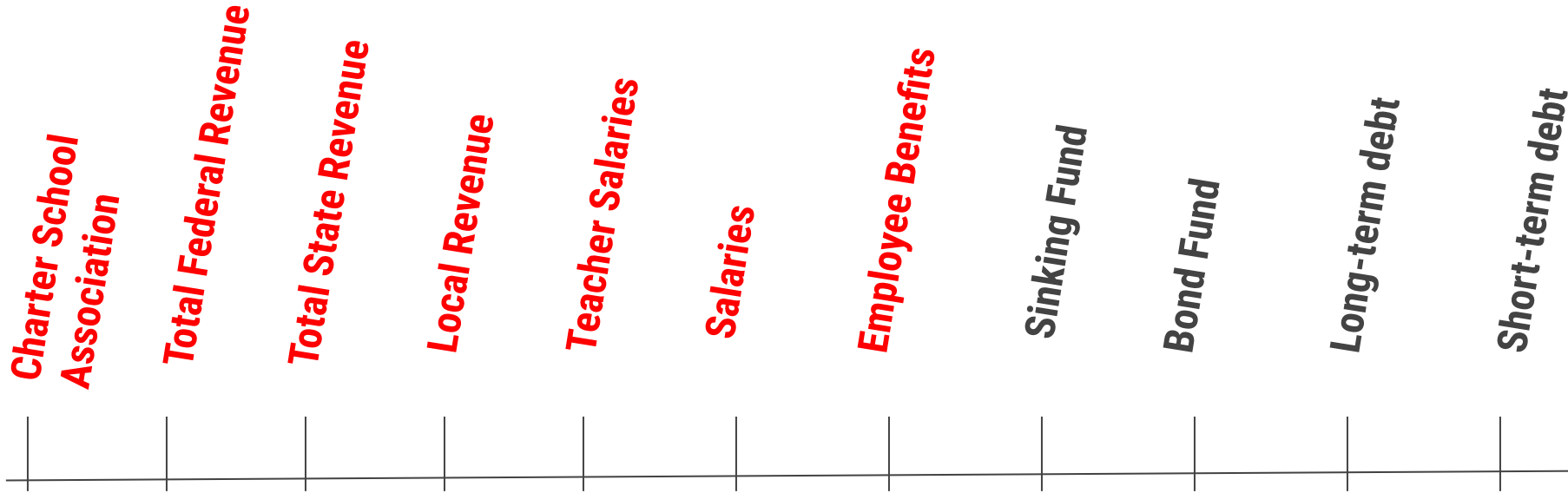- Avoided Linear models sensitive to feature collinearity

RECURSIVE FEATURE ELIMINATION

Teacher Salaries

Employee Benefits

Salaries

Total State Revenue

Total Federal Revenue

Other Funds

Non-charter schools

Total Expenditures

Textbooks

Local Revenue

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?