What's behind the sticker pr$ce of private college tuition and fees?

Final Project for Metis' Linear Regression and Web Scraping Module

Chloe Bergsma-Safar

Winter 2022

Introduction

- **Objective** - Understand what features of private non-profit institutions most influence the published price of their tuition and fees

- **Goal** - Help those involved in the decision-making process about higher education choices make more informed decisions
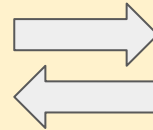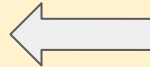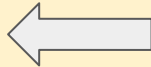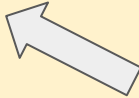
Methodology

Data Collection

Data Preparation

Modeling

Scraped data about **private non-profit colleges/universities offering Bachelor's degrees** from the National Center for Education Statistics (NCES)
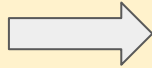
Cleaned and performed exploratory data analysis on scraped data

Ended up with **1322 schools** and **21 features** of each school

Round 1 - *without* financial aid as a feature
Round 2 - *with* financial aid as a feature

Models:
- Linear regression
- LassoCV regression
- Polynomial regression

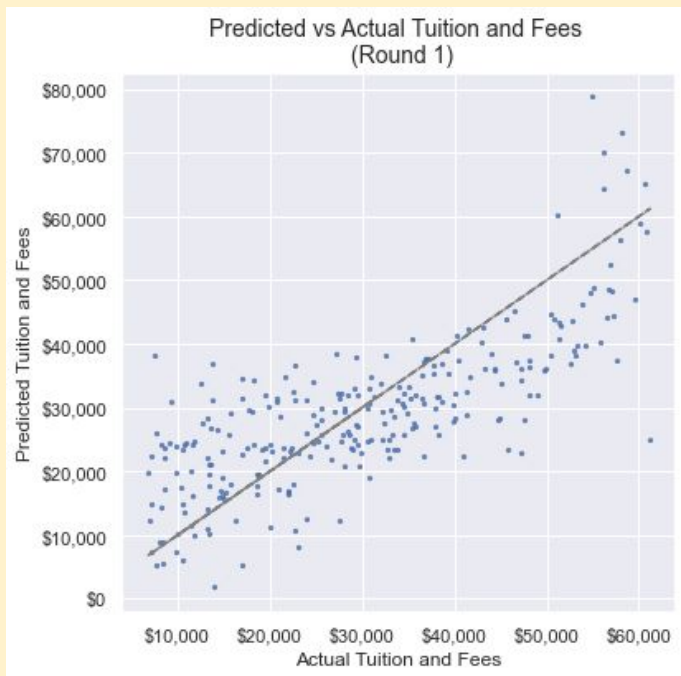Validation: cross-validation
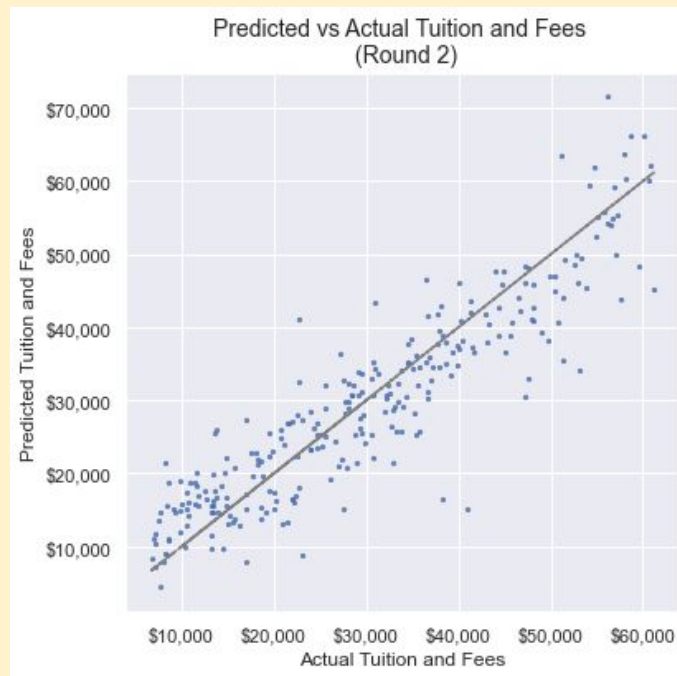
Evaluation metric: RMSE

Results

Round 1 - Modeling *without* financial aid

RMSE of LassoCV model: 10,048.67

Round 2 - Modeling *with* financial aid

RMSE of LassoCV model: 6,505.23

Results (cont.)

Round 1 - Modeling *without* financial aid

Final features (in order of importance)	Coefficient
Graduation rate	6,609
SAT/ACT scores are considered for admissions	5,485
# of applications	3,803
Acceptance rat	1,939
Price of books a	1,035
Enrollment	961
Student-to-faculty ratio	656
# of full-time faculty	492
# of part-time faculty	24

Round 2 - Modeling *with* financial aid

Final features (in order of importance)	Coefficient
Average amount of financial aid per student	11,051
Graduation rate	2,409
Price of books and supplies	1,602
# of applications	1,375
# of full-time faculty	1,114
SAT/ACT scores are co s	1,013
# of part-time faculty	789
Enrollment	376
Student-to-faculty ratio	176
Acceptance rate	55
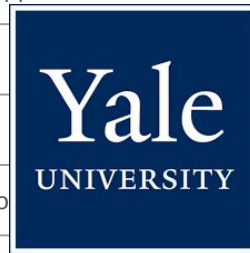
Conclusions

- Don't succumb to sticker shock before checking out a school's financial aid offerings.

- Consider a school's outcomes, like graduation rate, when evaluating the monetary investment of a school!

- These models could use some more tweaking
  - Scrape and add new features
  - Do more data cleaning (i.e., remove outliers from some features)