PRIVATE OR PUBLIC COLLEGE?

CLASSIFICATION PROBLEM

OUR TEAM

NAZMI BUNJAKU

SHUBHAM DAVE

TABLE OF CONTENTS

1

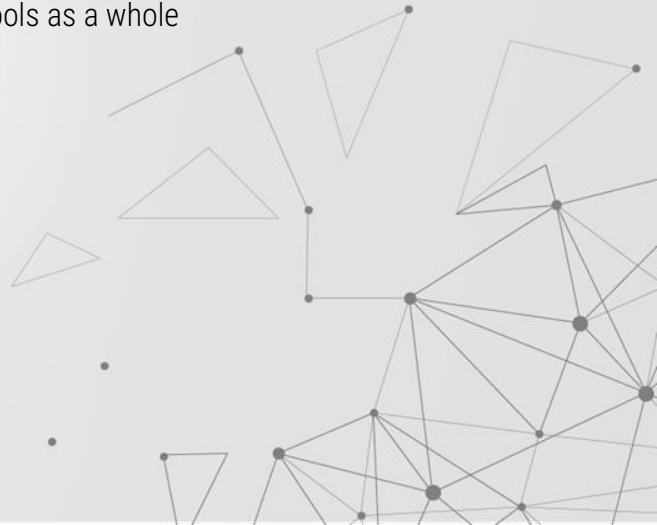
PROBLEM DESCRIPTION

BACKGROUND

- Public colleges are government-funded, while private schools rely more on tuition and endowments.
- Though often more costs are associated to them, private schools may offer generous financial aid.
- Many public universities boast a wider array of program offerings.
- Private and public universities offer distinct campus and residential experiences.

WHY SOLVE THIS ISSUE?

- Using advanced ML modeling we can uncover trends, outliers, and patterns
- Determine how private and public universities differ more, and where they relate
- Pinpoint weaknesses for both private and public colleges to better enhance their schools as a whole
- Where are private colleges advancing more than public colleges, and vice versa



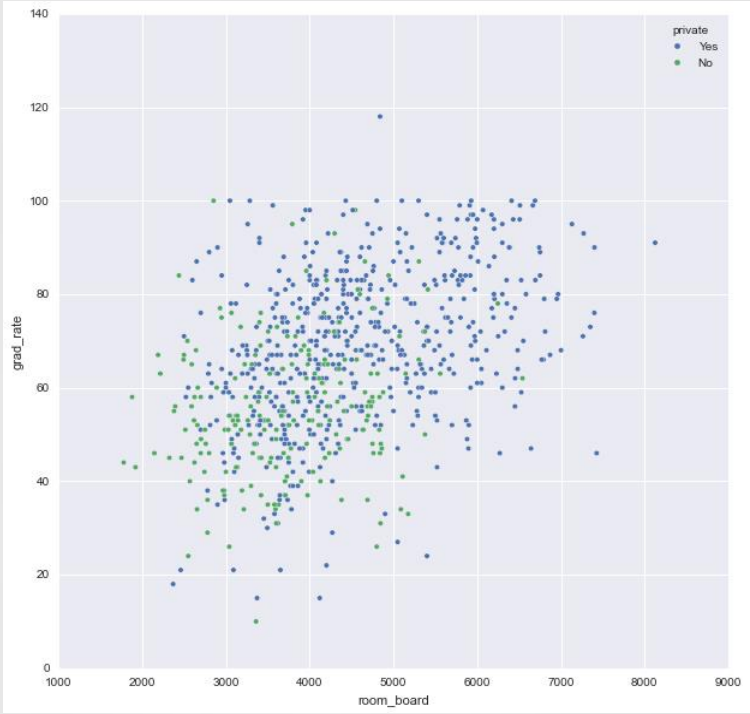
2

DATA UNDERSTANDING

EDA #1

SCATTERPLOT

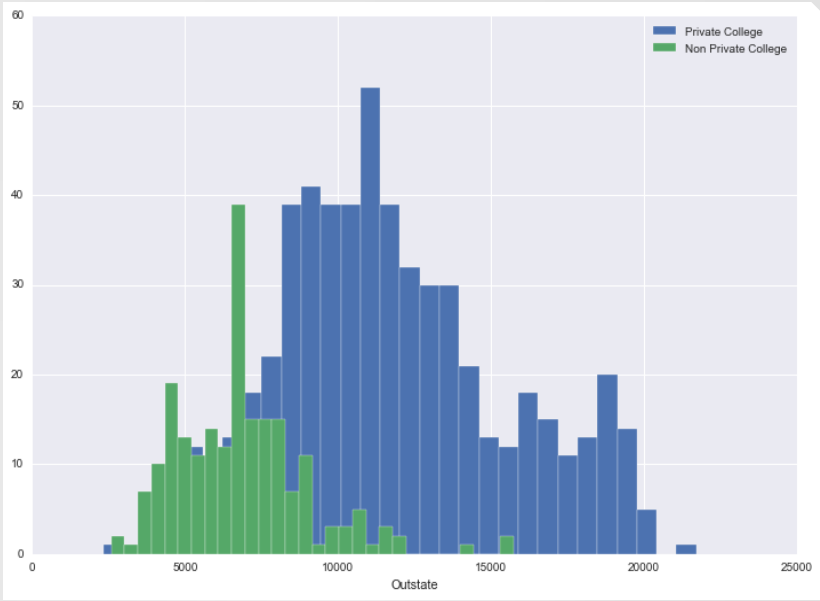
Room and Board Costs vs Graduation Rate of both private and non-private colleges



EDA #2

HISTOGRAM

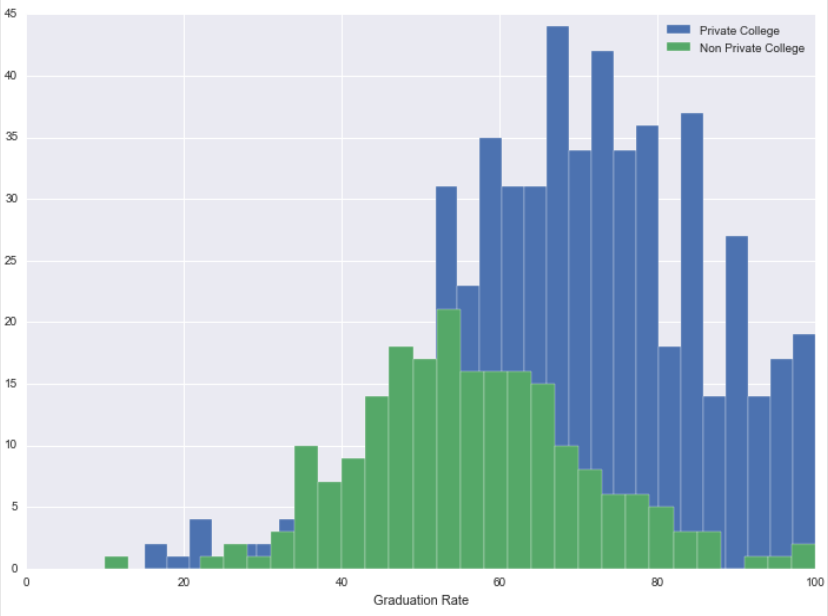
Out of state tuition cost for both private and non-private colleges



EDA #3

HISTOGRAM

Graduation Rate for both private and non-private colleges



3

APPROACH

K-MEANS CLUSTERING

- K-Means clustering is an unsupervised machine learning algorithm
- In contrast to traditional supervised ML algorithms, K-Means attempts to classify data without having first been trained with labeled data
- Once the algorithm has been run and the groups are defined, any new data can be easily assigned to the most relevant group
- For our model, we will attempt to use K-Means Clustering to cluster Universities into to two groups, Private and Public

CHALLENGES + SOLUTIONS

- Implementing K-Means algorithm from scratch
    - Solution: Utilize class slides and online resource for help
- Create a confusion matrix and classification report to see how well the K-Means clustering worked without being given any labels
    - Solution: Import K-Means function from the Sci-Kit Learn library for easy comparison

4

RESULTS

RESULTS

Before Scaling:

	PUBLIC	PRIVATE
PRECISION	0.69	0.79
RECALL	0.35	0.94
F1-SCORE	0.46	0.86

Accuracy = 78%

After Scaling:

	PUBLIC	PRIVATE
PRECISION	0.30	0.77
RECALL	0.69	0.40
F1-SCORE	0.42	0.53

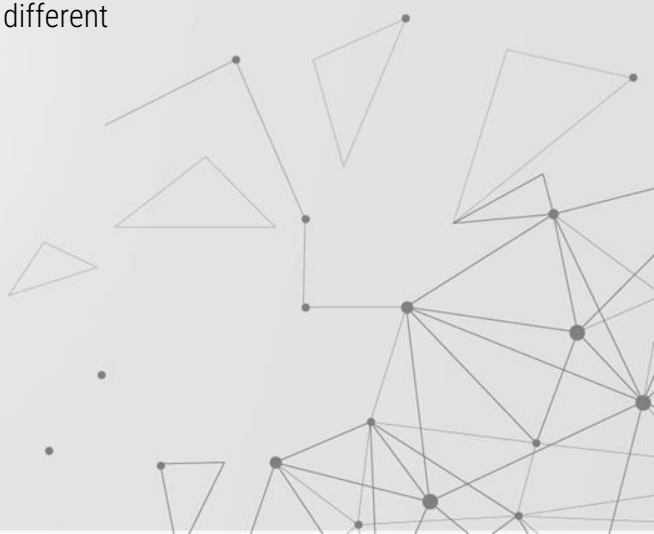
Accuracy = 48%

5

CONCLUSION

CONCLUSION

- From our project we can see how K-Means is useful for clustering un-labeled data
- We have visually seen how private colleges are excelling in more areas such as graduation rate than public colleges are
- However, public colleges are not that far apart especially with tuition being drastically different
- Our model achieved efficient accuracy with an unstructured dataset



THANKS

Does anyone have any questions?

n.bunjaku@vikes.csuohio.edu
s.dave13@vikes.csuohio.edu