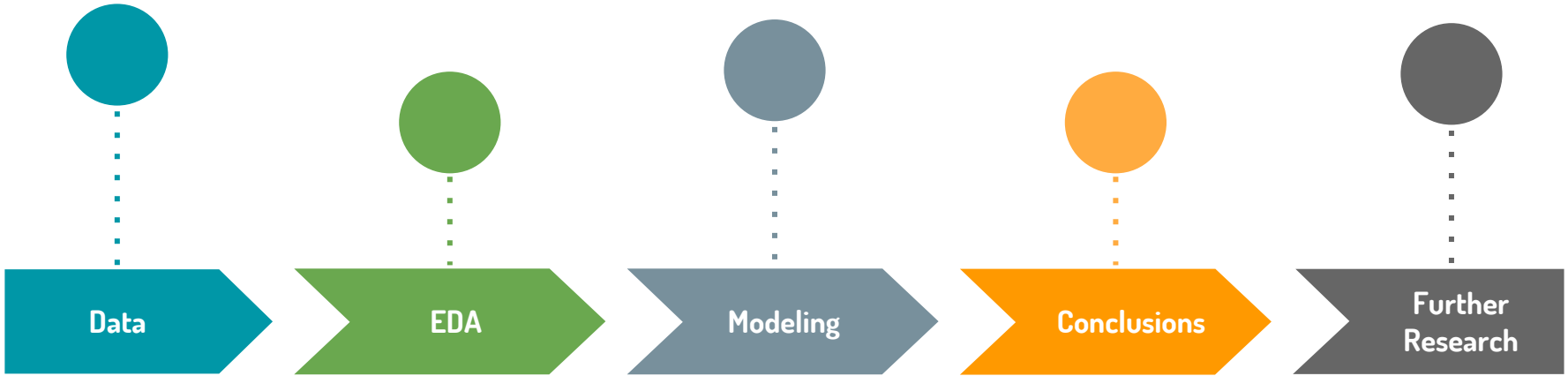
I'm here to School you

Rowan Langford

Agenda

Problem Statement

There is a glut of students who are encouraged to enter college, but never graduate, leaving them with lots of student debt and nothing to show for it. I analyzed a 2002 longitudinal study from the National Center for Education Statistics to predict which factors cause a student to drop out.



FGST

freegoogleslidestemplates.com

Data

Problems and Benefits

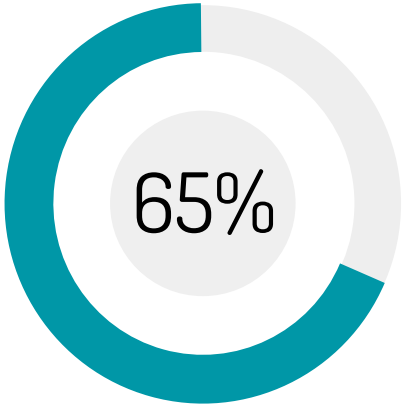
1. Looked at over 16,000 students from 9th grade-post college
2. 4,012 variables, at least 1000 of which were suppressed/restricted
3. About 9,000 of these students ever attended college
4. Most of the data was categorical

FGST

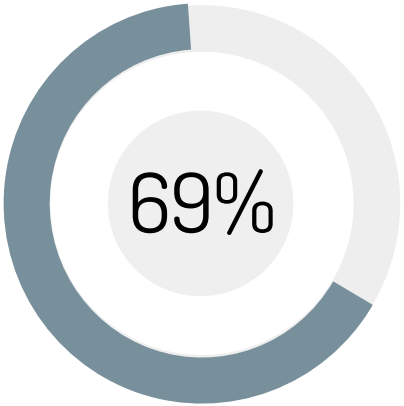
freegoogleslidestemplates.com

Education Statistics

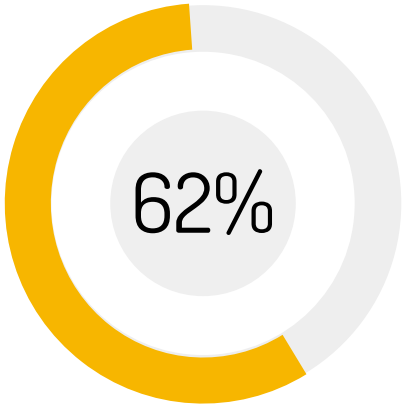
Statistics from the BLS/NCE



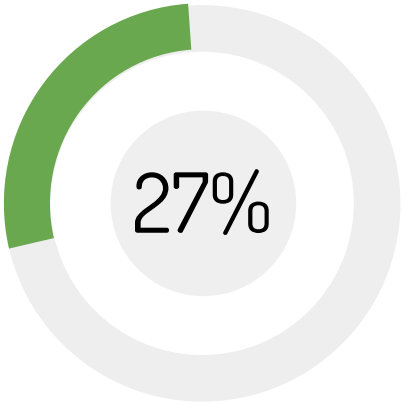
% of HS graduates who enter college



% of Students in study who graduate in 6 years



% Students who graduate from non-profit colleges



% Students who graduate from for-profit colleges

FGST

freegoogleslidestemplates.com

EDA

Correlation with target

Found highly correlated variables

Correlation with each other

Removed variables which were highly correlated with each other or had lots of missing values

Removing Variables

Removing variables from 3rd follow up study

Interaction Terms

Make Interaction Terms for highly corr variables

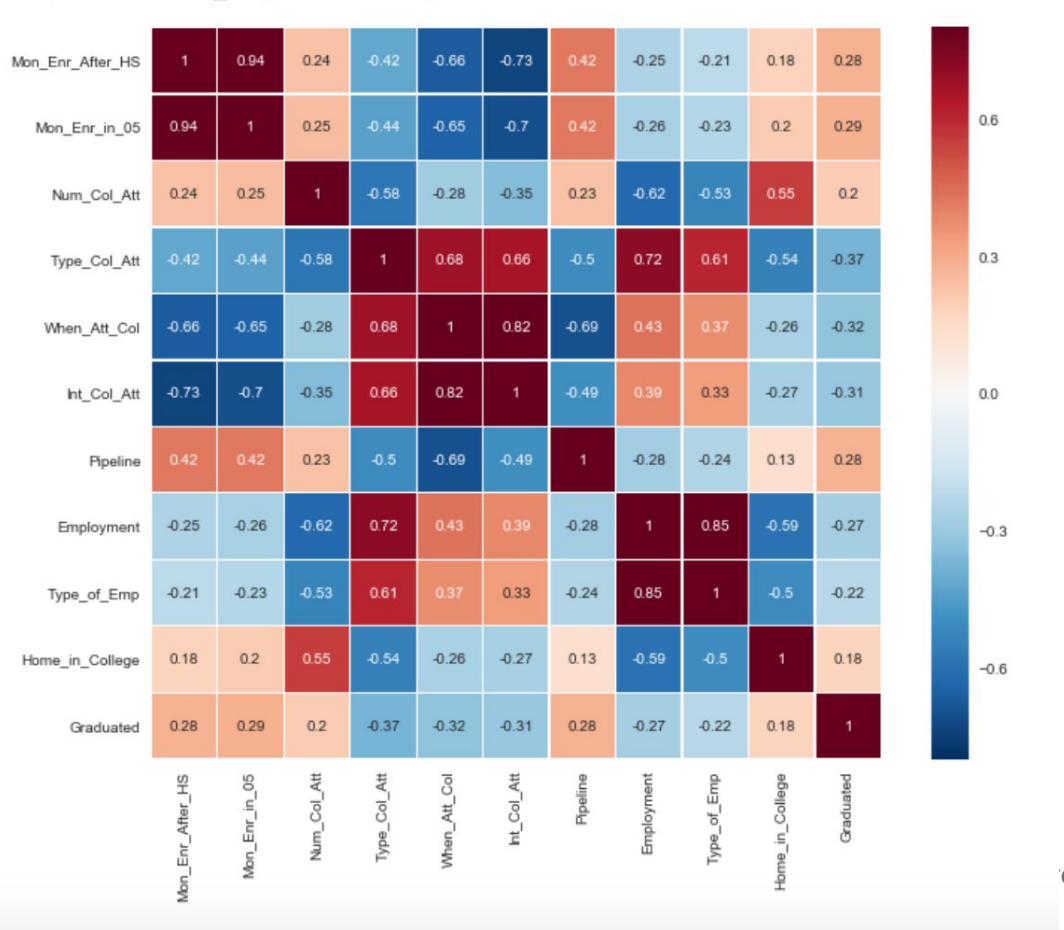
Codebook

Determine what the variables are and rename them

FGST

freegoogleslidestemplates.com

Correlation Heat Map



FGST

Chosen X Variables

Months Enrolled after HS and in 05/06 School Year

These variables were highly correlated so I included them separately and with an interaction term

Number of Colleges Attended

Integers of the number of colleges attended

Type of College Attended

This includes profit/not-for-profit and <2year, 2 year and 4 year colleges.

When Attended College

Their pathway to college

Intensity of Enrollment

Whether they were full/part time students and at what kind of school

Employment/Type of Employment

Whether or not they are employed and kind of employment. Included separately and with an interaction term.

Where they lived while in college

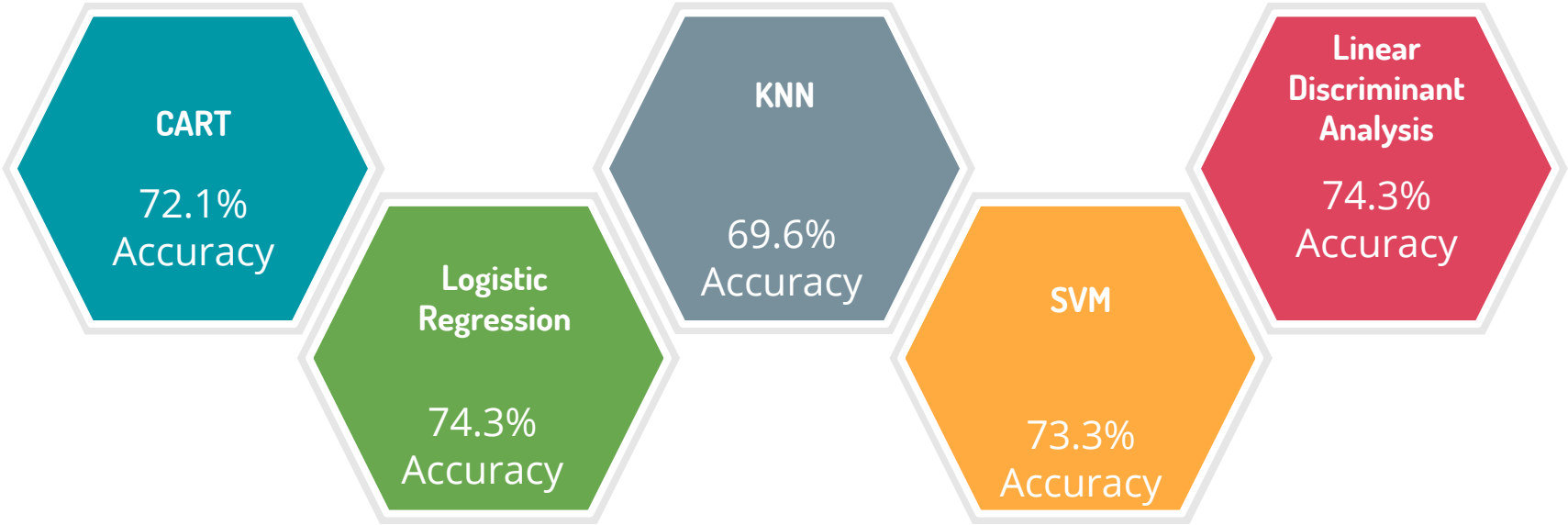
Whether they lived at home, on campus housing, or somewhere else off campus during college.

Pipeline

Post-secondary education pipeline.

FGST

Models Built



Feature Importances

	0
Intercept	1.263097
Mon_Enr_After_HS	1.016707
Mon_Enr_in_05	1.021323
Num_Col_Att	0.923786
Type_Col_Att	0.809374
When_Att_Col	0.955550
Int_Col_Att	0.836649
Pipeline	1.114810
Employment	0.539497
Type_of_Emp	1.093026
Home_in_College	0.959082
Mon_Enr_After_HS:Mon_Enr_in_05	1.006612
Employment:Type_of_Emp	1.093026
Int_Col_Att:When_Att_Col	1.024171

- Chose Logistic Regression as it was more white-box that LDA even though they had the same score
- Most important feature was pipeline
- This was followed by Type of Employment, which includes things like military

freegoogleslidestemplates.com

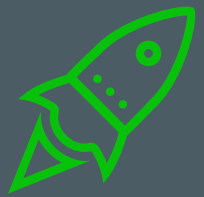
Areas for Further Research

- Request the full data as demographic information is probably important
- Methodology of Study
- Look at the more recent longitudinal studies put out

References

- https://www.washingtonpost.com/wp-stat/graphics/business/who-finishes-college/index.html
- https://www.bls.gov/news.release/archives/hsgec_06252003.pdf
- https://nces.ed.gov/fastfacts/display.asp?id=40
- https://nces.ed.gov/datatools/

@googleslides



Questions?