

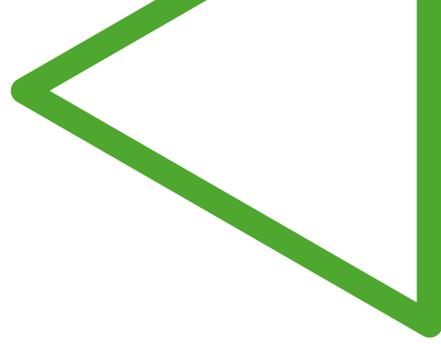
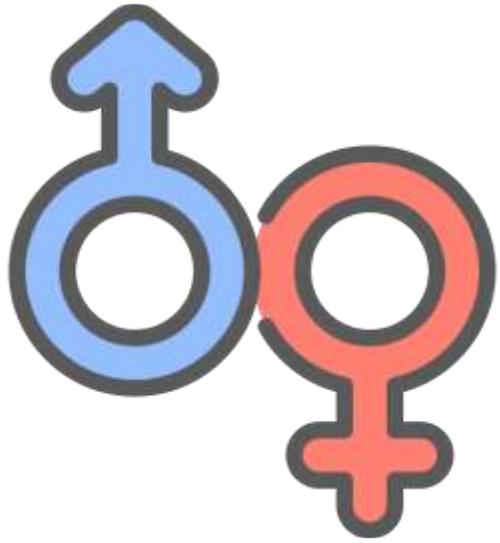


Predicting Academic Outcome Based on Data

Milo, Parker, Zach

Introducing the Data Set

- Our project focuses on a data set which used more than 37 values and 4,500 subjects to try and predict the academic outcomes of the subjects. The data was collected from many disjointed higher education databases. The data uses many personal information variables as well as grades from the first few terms of education.
- We'll be examining if this database holds any biases, whether it be from the variables included or the conclusion of success. The nature of how these variables decide on academic outcome will show bias, especially within marginalized groups present. This data study seeks to understand these statistics as well as the reasons for bias.

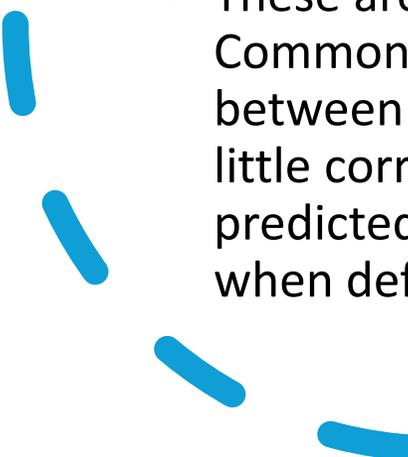


Nationality/Gender





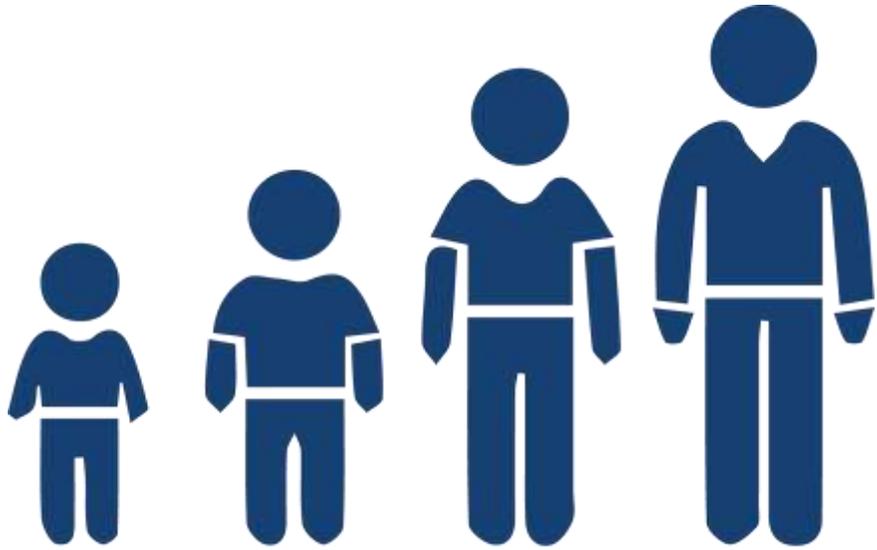
Nationality and Gender

- Nationality and gender were both chosen as a variable in this study and groups were given outcome based on these factors. These were self-reported among students and used to judge resources available based on racial or gender factors.
 - Minority communities may struggle disproportionately, and creating resources to assist and cause equity is important. There is also study to show differences in academics based on gender, which should be considered by educators.
 - There is extreme worry in bias for this section, particularly with assumptions based on these factors or deciding weight of target outcome.
 - These are both very personal, defining characteristics that aren't controlled by a person. Commonly, researchers include statistics like this to try and track whether outcomes differ between groups. How do they factor into the conclusion? The data shows that nationality had little correlation to the prediction, but looking at gender, we can see that males more often were predicted as "Dropout" than "Graduate". This may implicate some sort of bias or assumption when deferring the conclusion.
- 

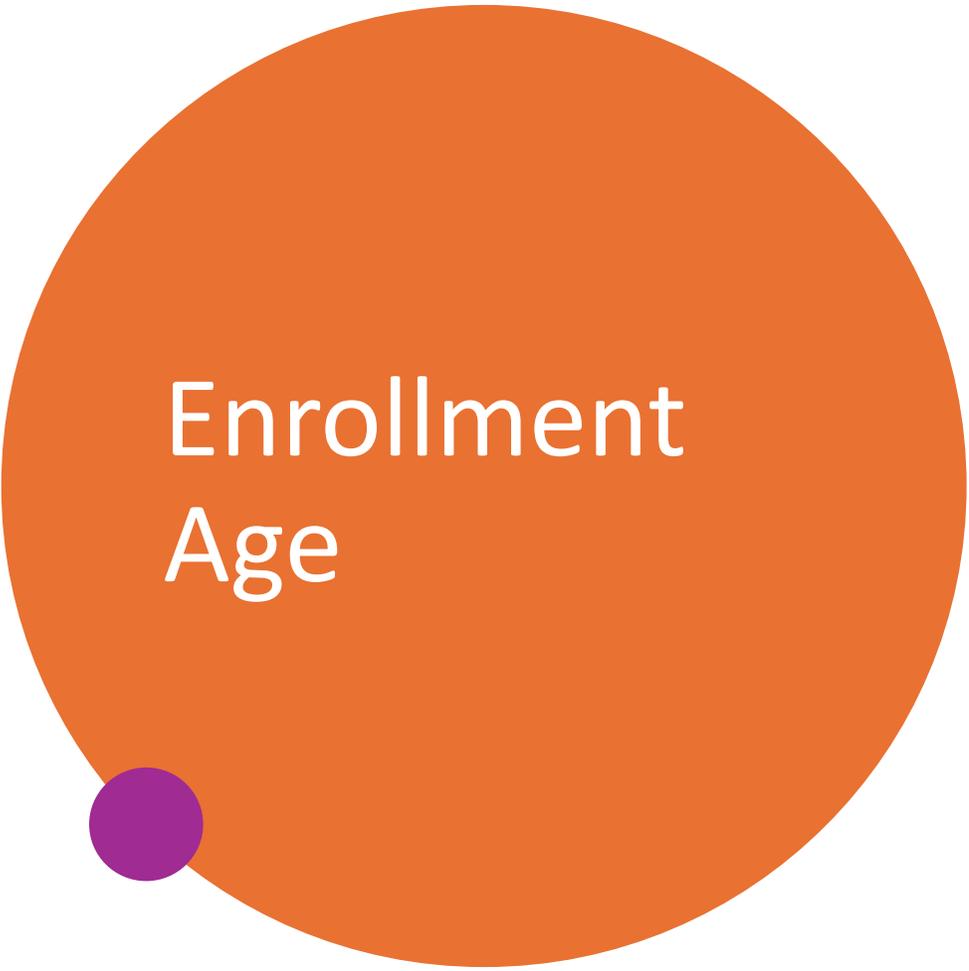
Charts

Nationality	# of Outcome Type	% of Outcome Type	# of Outcome Type	% of Outcome Type	# of Outcome Type	% of Outcome Type
Portuguese	1389	32.20%	770	17.85%	2155	49.95%
German		0.00%		0.00%	2	100.00%
Spanish	4	30.77%	5	38.46%	4	30.77%
Italian		0.00%		0.00%	3	100.00%
Dutch		0.00%		0.00%	1	100.00%
English		0.00%		0.00%	1	100.00%
Lithuanian	1	100.00%		0.00%		0.00%
Angolan	1	50.00%	1	50.00%		0.00%
Cape Verdean	4	30.77%	1	7.69%	8	61.54%
Guinean	1	20.00%		0.00%	4	80.00%
Mozambican		0.00%	1	50.00%	1	50.00%
Santomean	1	7.14%	5	35.71%	8	57.14%
Turkish		0.00%	1	100.00%		0.00%
Brazilian	14	36.84%	6	15.79%	18	47.37%
Romanian		0.00%	1	50.00%	1	50.00%
Moldovan	2	66.67%	1	33.33%		0.00%
Mexican	1	50.00%		0.00%	1	50.00%
Ukrainian	1	33.33%		0.00%	2	66.67%
Russian	1	50.00%	1	50.00%		0.00%
Cuban		0.00%	1	100.00%		0.00%
Colombian	1	100.00%		0.00%		0.00%

	Dropout		Enrolled		Graduate	
Gender	# of Outcome Type	% of Outcome Type	# of Outcome Type	% of Outcome Type	# of Outcome Type	% of Outcome Type
Female	720	25.10%	487	16.98%	1661	57.91%
Male	701	45.05%	307	19.73%	548	35.22%



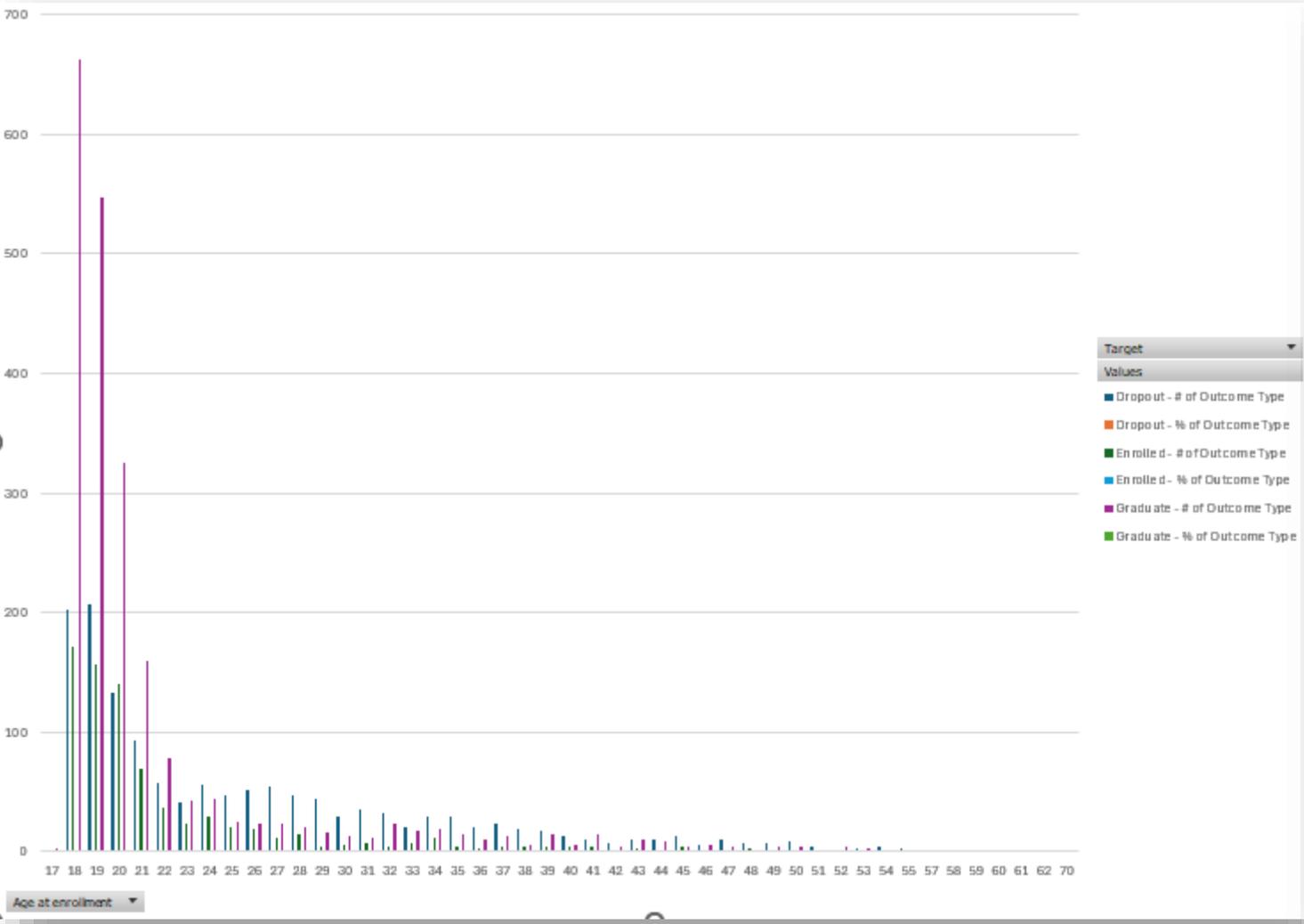
Enrollment Age



Enrollment Age

- There is a significant difference in overall outcome throughout ages
- Age directly effects other variables and often causes learning differences based on speed of education
- Most commonly the younger students are predicted to drop out more often when compared to those within the older side of the spectrum(30 and on)
- We can see throughout the predictions that age not only affects drop-out rate, but also GDP, financial support, and other important impactful data points

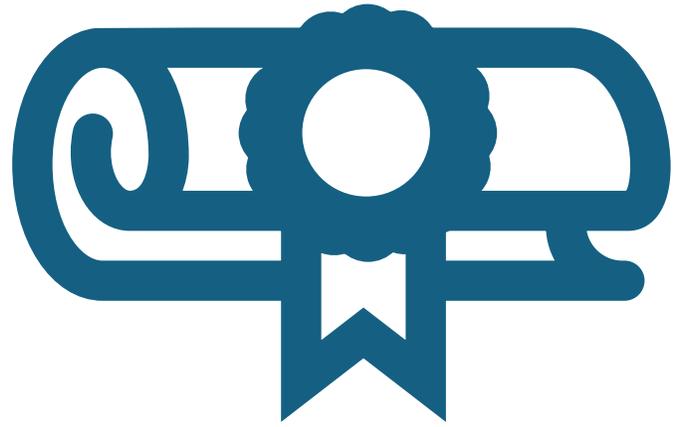
Graphical Representation



Age at enroll	Dropout		Enrolled		Graduate	
	# of Outcome Type	% of Outcome Type	# of Outcome Type	% of Outcome Type	# of Outcome Type	% of Outcome Type
17		0.00%	2	40.00%	3	60.00%
18	202	19.50%	172	16.60%	662	63.90%
19	207	22.72%	157	17.23%	547	60.04%
20	133	22.20%	140	23.37%	326	54.42%
21	93	28.88%	70	21.74%	159	49.38%
22	58	33.33%	37	21.26%	79	45.40%
23	41	37.96%	24	22.22%	43	39.81%
24	56	42.75%	30	22.90%	45	34.35%
25	47	50.54%	21	22.58%	25	26.88%
26	52	55.32%	19	20.21%	23	24.47%
27	55	60.44%	12	13.19%	24	26.37%
28	47	56.63%	15	18.07%	21	25.30%
29	45	68.18%	5	7.58%	16	24.24%
30	30	61.22%	6	12.24%	13	26.53%
31	36	65.45%	7	12.73%	12	21.82%
32	33	54.10%	5	8.20%	23	37.70%
33	20	44.44%	8	17.78%	17	37.78%
34	29	48.33%	12	20.00%	19	31.67%
35	29	59.18%	5	10.20%	15	30.61%
36	21	60.00%	3	8.57%	11	31.43%
37	23	54.76%	5	11.90%	14	33.33%
38	19	65.52%	4	13.79%	6	20.69%
39	18	47.37%	5	13.16%	15	39.47%
40	13	56.52%	4	17.39%	6	26.09%
41	11	35.48%	5	16.13%	15	48.39%
42	7	53.85%	2	15.38%	4	30.77%
43	11	44.00%	3	12.00%	11	44.00%
44	10	47.62%	2	9.52%	9	42.86%
45	13	59.09%	4	18.18%	5	22.73%
46	6	50.00%		0.00%	6	50.00%
47	11	61.11%	2	11.11%	5	27.78%
48	7	63.64%	3	27.27%	1	9.09%
49	7	53.85%	1	7.69%	5	38.46%
50	9	56.25%	2	12.50%	5	31.25%
51	5	71.43%		0.00%	2	28.57%
52		0.00%		0.00%	4	100.00%
53	3	42.86%	1	14.29%	3	42.86%
54	4	57.14%	1	14.29%	2	28.57%
55	3	60.00%		0.00%	2	40.00%
57	1	50.00%		0.00%	1	50.00%
58	2	66.67%		0.00%	1	33.33%
59	2	66.67%		0.00%	1	33.33%
60		0.00%		0.00%	2	100.00%
61	1	100.00%		0.00%		0.00%
62		0.00%		0.00%	1	100.00%
70	1	100.00%		0.00%		0.00%

Displacement



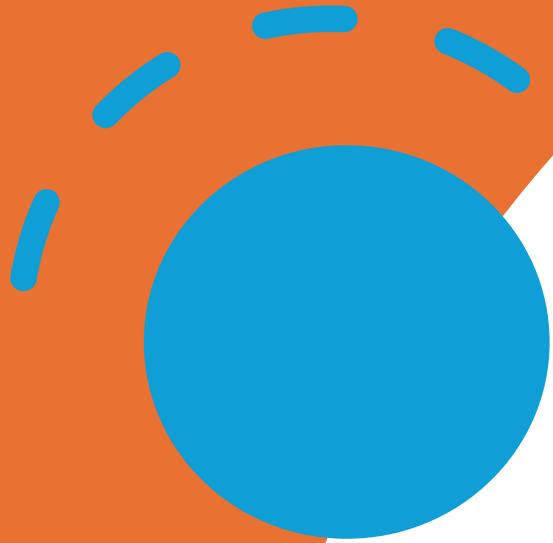


Displacement

- Someone labeled as "displaced", is either a refugee, not born within the country or financially unstable
- Displacement is very important in academic related fields, as one's entire quality of life is altered and may lead to a negative academic outcome.
- The data itself does not show a significant difference of outcome based on this variable, though the effects of displacement are important to consider in the prediction itself. Displacement is also most correlated with factors such as nationality causing bias through two related factors.

Chart

Displaced <input type="checkbox"/>	Dropout		Enrolled		Graduate	
	# of Outcome Type	% of Outcome Type	# of Outcome Type	% of Outcome Type	# of Outcome Type	% of Outcome Type
0	752	37.64%	361	18.07%	885	44.29%
1	669	27.58%	433	17.85%	1324	54.58%



Predictions & Exclusions



Predictions

- Data is collected from disjoint sources from the Polytechnic Institute of Portalegre database
- Predictions are made from a machine learning model based on previous students' academic records & outcomes.
- We performed linear regression on the data to see how heavily variables were weighted on the outcome, examining R and F values.
- Personal Characteristic data has a smaller correlation number, but even smaller values can reflect some assumptions in the predictions.



Exclusions

- We also examined bias based on some variables that were not included in the data
- Students' income or their family's income may have been value for examining background and ability, which could factor into access to academic resources and graduation.
- The absence of these contextual variables may lead to predictive models to miss some key details of students' lives and make an inaccurate or biased decision.

Conclusion

- Our goal was to examine whether the dataset contains bias based on the variables included and how academic success is defined.
- We found that demographic variables such as nationality and gender showed low correlation, while age showed slightly stronger correlation with academic outcomes.
- This may imply even with low statistical influence, variables like displacement and demographics may still reflect underlying social factors or potential bias in the predictive model.