INTRODUCTION

Algebra I is the most failed class in the country. As the foundational course to all secondary and higher-level mathematics and sciences, lack of mastery of Algebra can then be the indicator for lack of future involvement and/or achievement in STEM fields. As a former BIPOC teacher of BIPOC students, I sought out to determine if teacher race had any impact on predicting student performance on the Texas Algebra I STAAR.
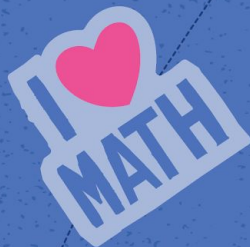
The concept of <u>racial hierarchy of mathematical ability</u>—a term coined by Danny Martin, education professor at the University of Illinois at Chicago—basically says constantly reading and hearing about underperforming black, Latino, and indigenous students begins to embed itself into how math teachers view these students, attributing achievement differences to their innate ability to succeed in math.

— THE ATLANTIC, "*How Does Race Affect a Student's Math Education?*" (April 2017)
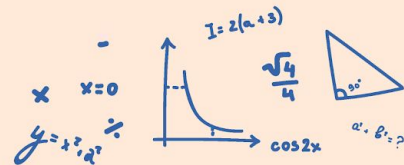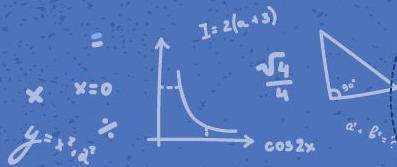
PROJECT GOALS

1) To create a web scraper and parser to scrape data from the Texas Education Agency for every campus in Texas
2) To predict the percentage of students of marginalized demographic on any given Texas public school campus who will score MASTERS on the End-of-Course Algebra I STAAR based on the percentage of BIPOC teachers they've been exposed to

TECH STACK

WEB SCRAPING

PARSING

pandas

NumPy

DATA

Seaborn

VISUALIZATION

scikit **learn**

MACHINE LEARNING

PROCESS

$$\begin{cases} x = \dfrac{1}{1+y} \\ x = y + 3 \end{cases}$$

$$x = \frac{1}{1+y}$$

$$y + 3 = \frac{1}{1+y}$$

Step 1

Web scraping + parsing two pages per school; merge dataframes, clean + cast data into usable form

Step 2

EDA!

Step 3

Modeling, then rinse and repeat
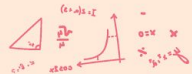
FEATURES OF THE SCRAPED DATA
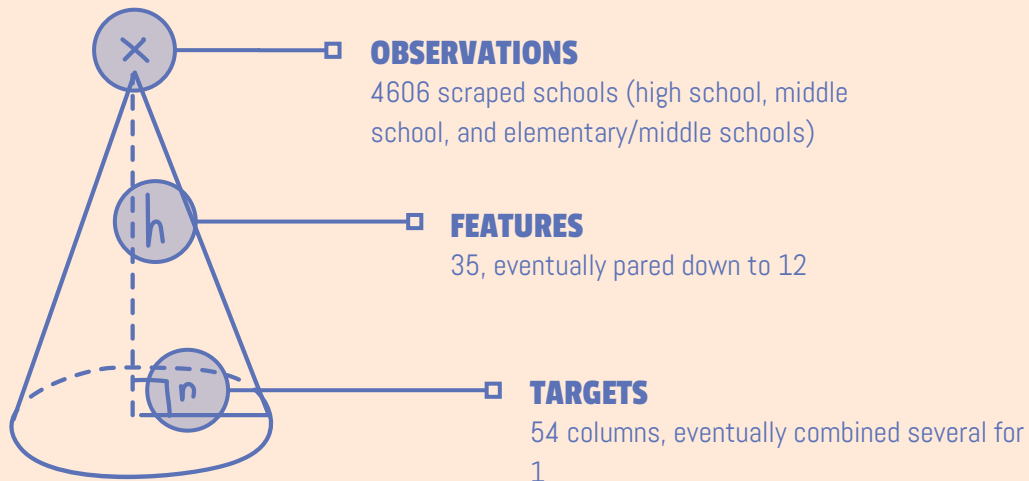
OBSERVATIONS

4606 scraped schools (high school, middle school, and elementary/middle schools)

FEATURES

35, eventually pared down to 12

TARGETS

54 columns, eventually combined several for 1

THE ORIGINAL DF

```
---   ------                                  --------------  -----
0     Campus_African_American                 127 non-null    float64
1     District_African_American               127 non-null    float64
2     State_African_American                  127 non-null    float64
3     Campus_Hispanic                         127 non-null    float64
4     District_Hispanic                       127 non-null    float64
5     State_Hispanic                          127 non-null    float64
6     Campus_White                            127 non-null    float64
7     District_White                          127 non-null    float64
8     State_White                             127 non-null    float64
9     Campus_American_Indian                  127 non-null    float64
10    District_American_Indian                127 non-null    float64
11    State_American_Indian                   127 non-null    float64
12    Campus_Asian                            127 non-null    float64
13    District_Asian                          127 non-null    float64
14    State_Asian                             127 non-null    float64
15    Campus_Pacific_Islander                 127 non-null    float64
16    District_Pacific_Islander               127 non-null    float64
17    State_Pacific_Islander                  127 non-null    float64
18    Campus_Two_or_More_Races                127 non-null    float64
19    District_Two_or_More_Races              127 non-null    float64
20    State_Two_or_More_Races                 127 non-null    float64
21    State_Approaching                       127 non-null    float64
22    District_Approaching                    127 non-null    float64
23    Campus_Approaching                      127 non-null    float64
24    African_American_Approaching            127 non-null    float64
25    Hispanic_Approaching                    127 non-null    float64
26    White_Approaching                       127 non-null    float64
27    American_Indian_Approaching             127 non-null    float64
28    Asian_Approaching                       127 non-null    float64
29    Pacifc_Islander_Approaching             127 non-null    float64
30    Two_or_More_Races_Approaching           127 non-null    float64
31    Special_Ed_Approaching                  127 non-null    float64
32    Econ_Disadv_Approaching                 127 non-null    float64
33    EL_Approaching                          127 non-null    float64
34    State_Meets                             127 non-null    float64
35    District_Meets                          127 non-null    float64
36    Campus_Meets                            127 non-null    float64
37    African_American_Meets                  127 non-null    float64
38    Hispanic_Meets                          127 non-null    float64
39    White_Meets                             127 non-null    float64
40    American_Indian_Meets                   127 non-null    float64
41    Asian_Meets                             127 non-null    float64
42    Pacifc_Islander_Meets                   127 non-null    float64
43    Two_or_More_Races_Meets                 127 non-null    float64
44    Special_Ed_Meets                        127 non-null    float64
45    Econ_Disadv_Meets                       127 non-null    float64
46    EL_Meets                                127 non-null    float64
47    State_Masters                           127 non-null    float64
48    District_Masters                        127 non-null    float64
49    Campus_Masters                          127 non-null    float64
50    African_American_Masters                127 non-null    float64
51    Hispanic_Masters                        127 non-null    float64
52    White_Masters                           127 non-null    float64
53    American_Indian_Masters                 127 non-null    float64
54    Asian_Masters                           127 non-null    float64
55    Pacifc_Islander_Masters                 127 non-null    float64
56    Two_or_More_Races_Masters               127 non-null    float64
57    Special_Ed_Masters                      127 non-null    float64
58    Econ_Disadv_Masters                     127 non-null    float64
59    EL_Masters                              127 non-null    float64
60    BIPOC_Masters                           127 non-null    float64
61    Difference_White_BIPOC_Masters          127 non-null    float64
62    BIPOC_Masters_no_Asian                  127 non-null    float64
63    Difference_White_BIPOC_Masters_no_Asian 127 non-null    float64
64    BIPOC_Campus                            127 non-null    float64
```

DF_FEATURES

Campus_ID	Campus_African_American	District_African_American	Campus_Hispanic	District_Hispanic	Campus_American_Indian	District_American_Indian
'001902001	5.2	7.8	4.3	2.0	0.0	0.0
'001906002	0.0	3.0	0.0	0.0	0.0	0.0
'003801001	5.6	10.8	4.3	6.2	0.0	0.0
'003904001	0.0	0.8	5.1	1.5	0.0	0.8
'004901041	1.1	0.9	16.5	17.7	2.1	0.5

```
<class 'pandas.core.frame.DataFrame'>
Index: 127 entries, '001902001 to '020902004
Data columns (total 13 columns):
 #   Column                     Non-Null Count   Dtype
---  ------                     --------------   -----
 0   Campus_African_American    127 non-null     float64
 1   District_African_American  127 non-null     float64
 2   Campus_Hispanic            127 non-null     float64
 3   District_Hispanic          127 non-null     float64
 4   Campus_American_Indian     127 non-null     float64
 5   District_American_Indian   127 non-null     float64
 6   Campus_Asian               127 non-null     float64
 7   District_Asian             127 non-null     float64
 8   Campus_Pacific_Islander    127 non-null     float64
 9   District_Pacific_Islander  127 non-null     float64
 10  Campus_Two_or_More_Races   127 non-null     float64
 11  District_Two_or_More_Races 127 non-null     float64
 12  BIPOC_Campus               127 non-null     float64
dtypes: float64(13)
memory usage: 18.9+ KB
```

OPTIONS FOR TARGETS

```
21  State_Approaching                           127 non-null    float64
22  District_Approaching                        127 non-null    float64
23  Campus_Approaching                          127 non-null    float64
24  African_American_Approaching                127 non-null    float64
25  Hispanic_Approaching                        127 non-null    float64
26  White_Approaching                           127 non-null    float64
27  American_Indian_Approaching                 127 non-null    float64
28  Asian_Approaching                           127 non-null    float64
29  Pacifc_Islander_Approaching                 127 non-null    float64
30  Two_or_More_Races_Approaching               127 non-null    float64
31  Special_Ed_Approaching                      127 non-null    float64
32  Econ_Disadv_Approaching                     127 non-null    float64
33  EL_Approaching                              127 non-null    float64
34  State_Meets                                 127 non-null    float64
35  District_Meets                              127 non-null    float64
36  Campus_Meets                                127 non-null    float64
37  African_American_Meets                      127 non-null    float64
38  Hispanic_Meets                              127 non-null    float64
39  White_Meets                                 127 non-null    float64
40  American_Indian_Meets                       127 non-null    float64
41  Asian_Meets                                 127 non-null    float64
42  Pacifc_Islander_Meets                       127 non-null    float64
43  Two_or_More_Races_Meets                     127 non-null    float64
44  Special_Ed_Meets                            127 non-null    float64
45  Econ_Disadv_Meets                           127 non-null    float64
46  EL_Meets                                    127 non-null    float64
47  State_Masters                               127 non-null    float64
48  District_Masters                            127 non-null    float64
49  Campus_Masters                              127 non-null    float64
50  African_American_Masters                    127 non-null    float64
51  Hispanic_Masters                            127 non-null    float64
52  White_Masters                               127 non-null    float64
53  American_Indian_Masters                     127 non-null    float64
54  Asian_Masters                               127 non-null    float64
55  Pacifc_Islander_Masters                     127 non-null    float64
56  Two_or_More_Races_Masters                   127 non-null    float64
57  Special_Ed_Masters                          127 non-null    float64
58  Econ_Disadv_Masters                         127 non-null    float64
59  EL_Masters                                  127 non-null    float64
60  BIPOC_Masters                               127 non-null    float64
61  Difference_White_BIPOC_Masters              127 non-null    float64
62  BIPOC_Masters_no_Asian                      127 non-null    float64
63  Difference_White_BIPOC_Masters_no_Asian     127 non-null    float64
64  BIPOC_Campus                                127 non-null    float64
```

In the current iteration of the End-of-Course Algebra I STAAR, students can score in one of four ways:

**Approaching** grade level (roughly 38% correct)
Meets grade level (roughly 55% correct)
Masters grade level (roughly 73% correct)

```
1  total_df['BIPOC_Masters'] = total_df[['African_American_Masters', 'Hispanic_Masters',
2                                       'American_Indian_Masters','Asian_Masters', 'Pacifc_Islander_Masters',
3                                       'Two_or_More_Races_Masters']].mean(axis=1)
4  total_df['Difference_White_BIPOC_Masters'] = total_df['White_Masters'] - total_df['BIPOC_Masters']
```

Ultimately created a new target column based on a combination of others that represents the difference between the percent of white students that score masters and the average of BIPOC students that score masters.
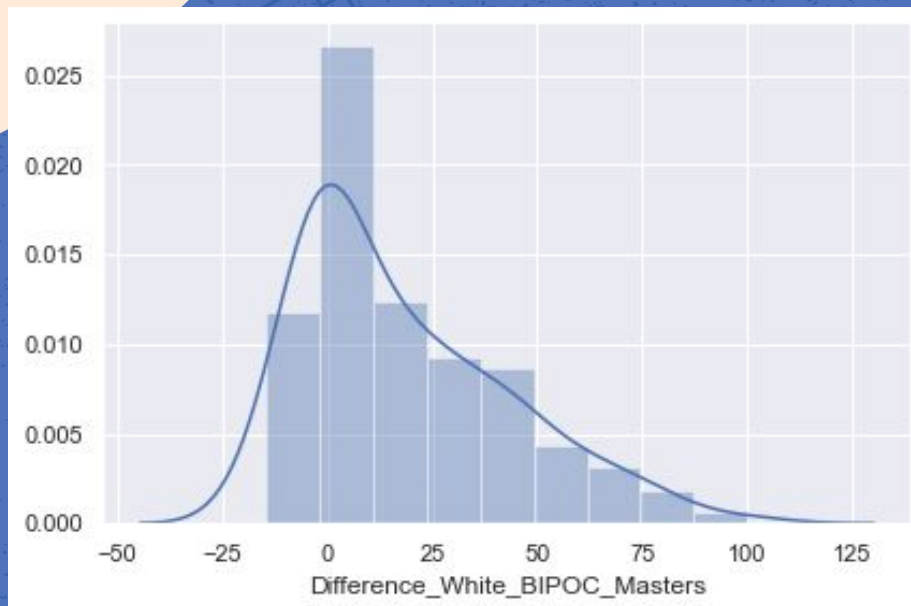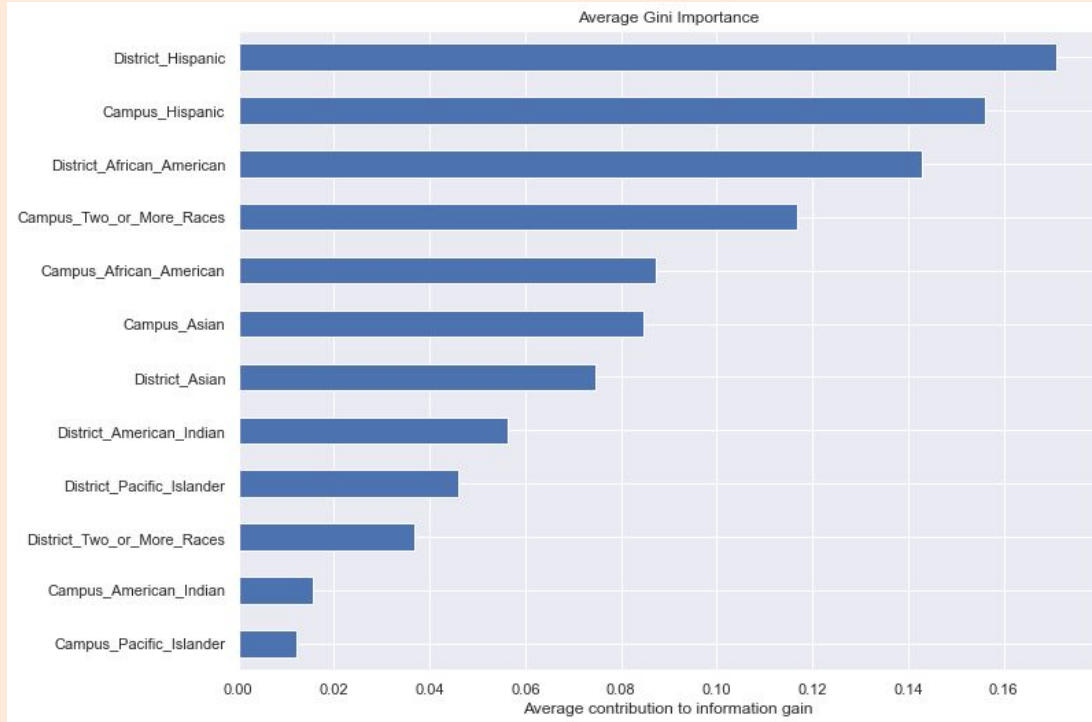
EDA
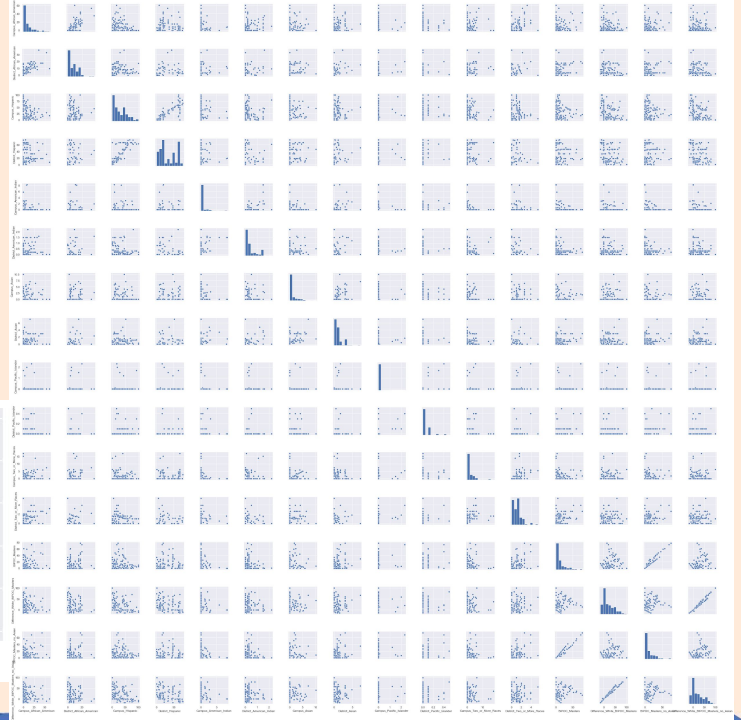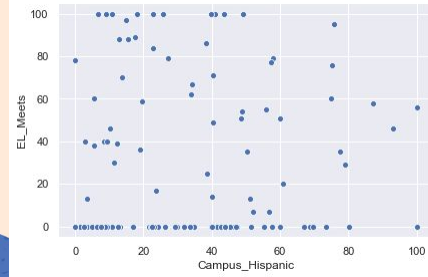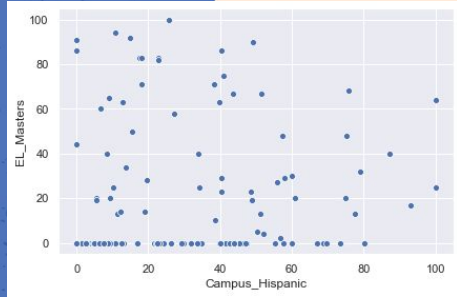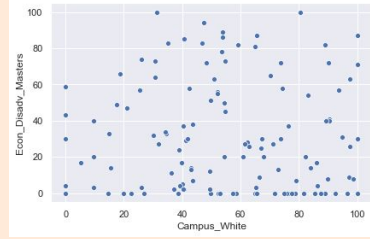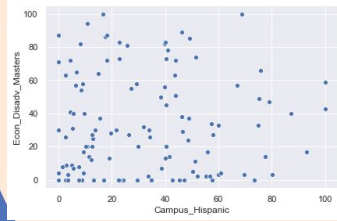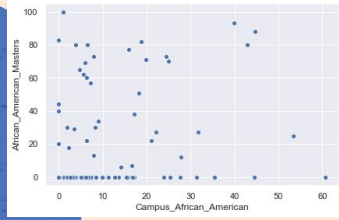
18.5%
mean

−14.5%, 100%
minimum, maximum

MODEL: LINEAR REGRESSION

RMSE = 16.05

RANDOM FOREST

Average Gini Importance

FUTURE STEPS

Do some feature engineering to make the features less numeric

Drop outliers/observations with negative target values

Is there any data on teacher demographics by subject?

Deeper EDA on total_df, including investigating loss of certain data points

THANKS FOR SWINGING BY!

Does anyone have any questions?

You can find me at:

rakhiagrawal.17@gmail.com
www.linkedin.com/in/rakhira
www.github.com/rakhira
At this desk, in this specific chair, for the rest of the apocalypse