

# Who I am as a Teacher

An analysis of student outcomes  
in my classroom for the past four  
years as a teacher at

MT. SCOTT LEARNING CENTER

By Matt Nelson



Initial wonderings

# STATE OF EDUCATION



EDUCATION

Oregon's abysmal chronic attendance problem grew even worse last year, state reveals

Updated: Oct. 31, 2023, 10:21 a.m. | Published: Oct. 26, 2023, 6:00 a.m.



Bloomberg

Opinion | Editorial Board

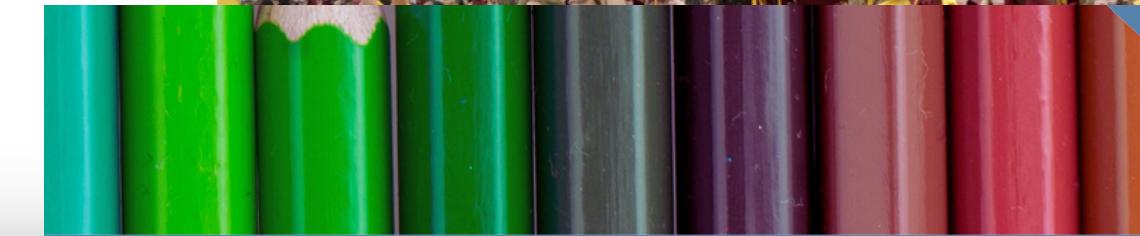
**America's High Schoolers Are Running Out of Time**

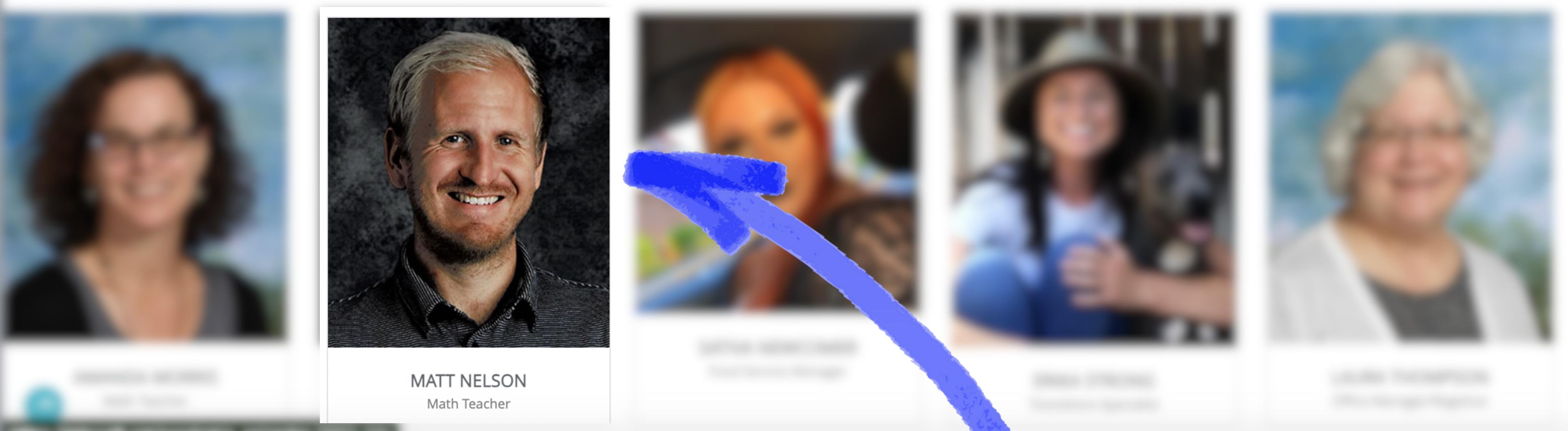
Millions of students are nearing graduation unprepared for college or the workforce. Policymakers should intervene before it's too late.

November 30, 2023 at 5:00 AM PST

By The Editorial Board

The Editorial Board publishes the views of the editors across a range of national and global affairs.



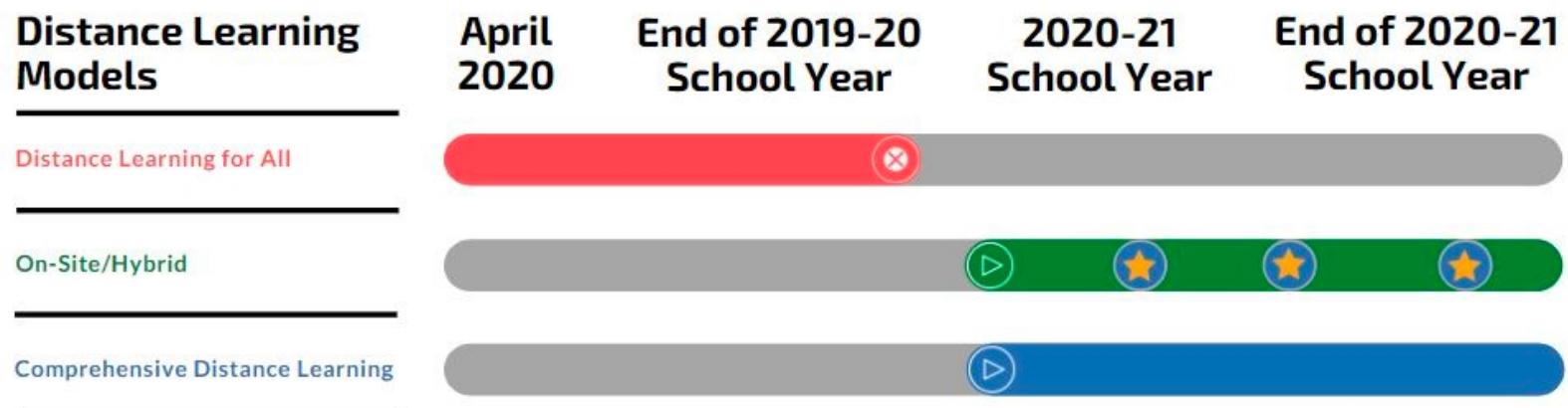


MATT NELSON  
Math Teacher



2023-2024 will be my fifth year  
teaching at Mt. Scott Learning  
Center.

# SHIFT TO COMPREHENSIVE DISTANCE LEARNING



2019-2020

2020-2021

2021-2022

2022-2023

MY  
FIRST  
YEAR

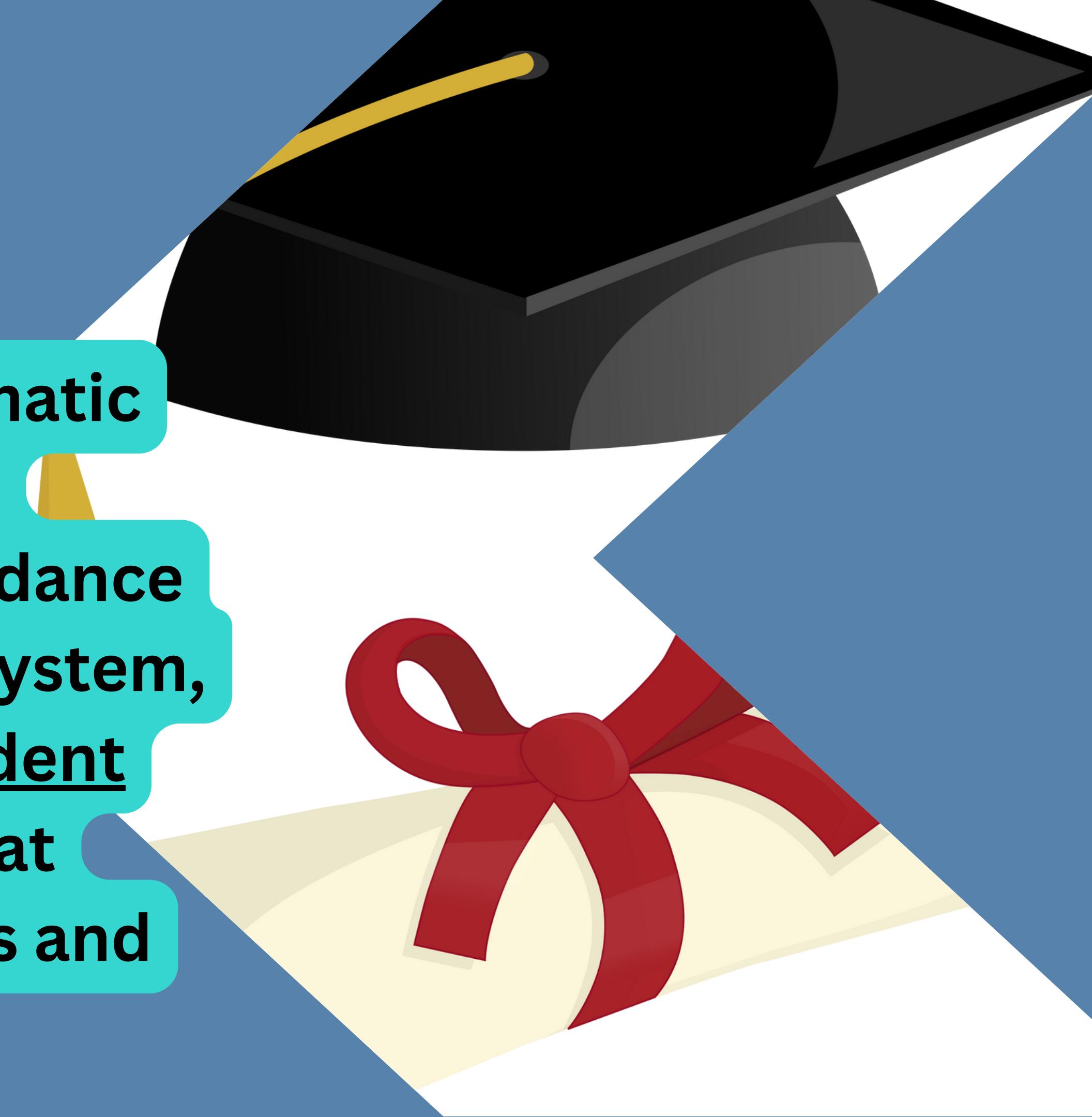
LOCKDOWN

LEARNING  
DISTANCE



# Problem

In an era of extreme systematic transition for both content delivery and student attendance rates in the US education system, what are the drivers of student outcomes. Specifically, what helps students pass classes and succeed in my room?



# Criteria

Due to my experience teaching the specific demographic of Mt. Scott, and based on school focus, I decided to limit my analysis to a few features

## FEATURES

School Year

Quarter

Period

Race/Ethnicity

Zip Code

Course

Days Absent

Absent Ratio

Letter Grade

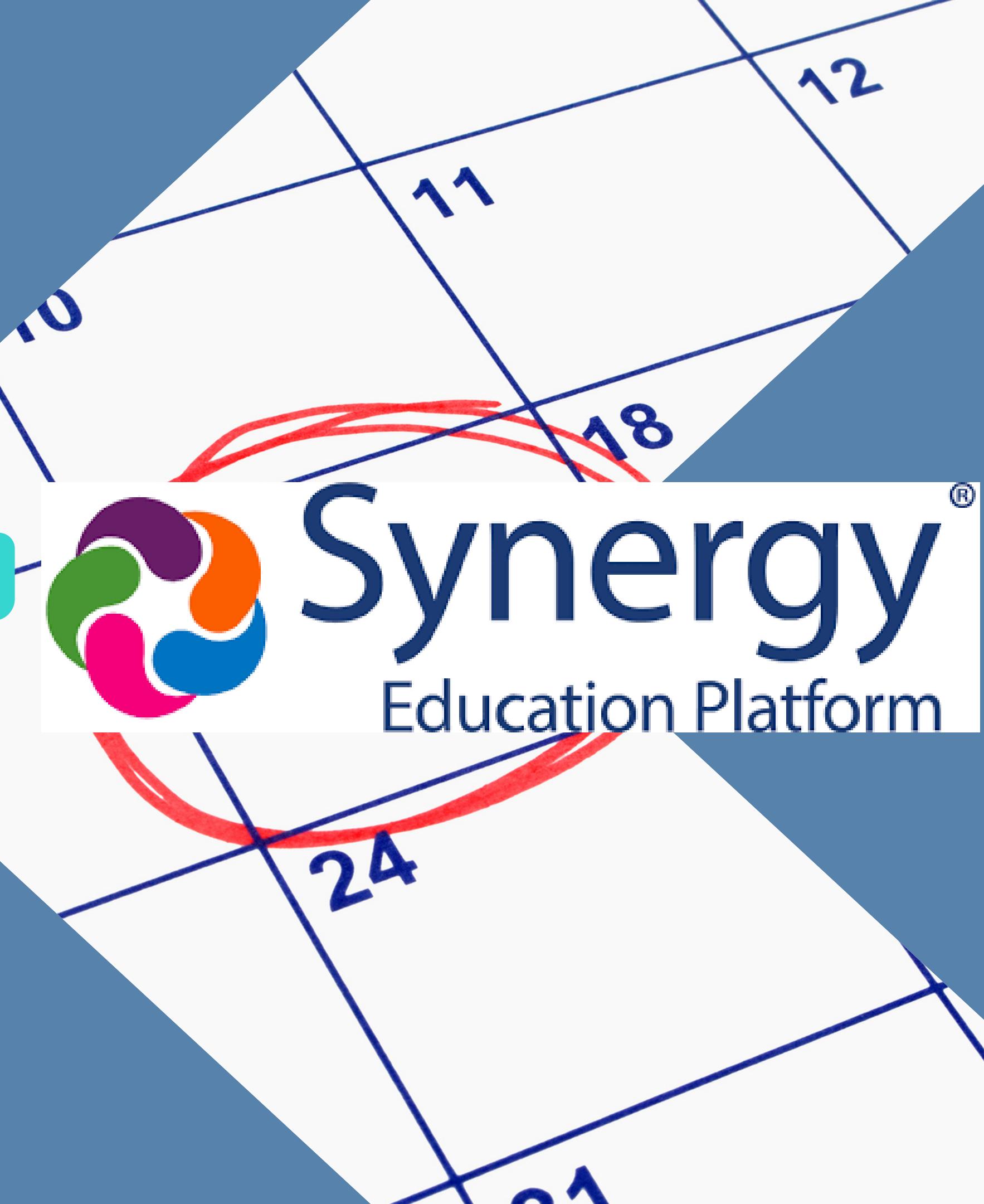
Percent Grade

Pass/Fail



# Datasets

To create the datasets, I've had to do multiple rounds of data acquisition through partnering with our office manager/registrar. They were able to give me access to a hardcopy dataset of student race/ethnicity and a digital table of student zipcodes per year. They also pulled data of student attendance and letter grade per class per quarter per year. For grade percentages, I downloaded my gradebooks on a system called Synergy for each math class for each quarter for each year. I then did a lot of merging



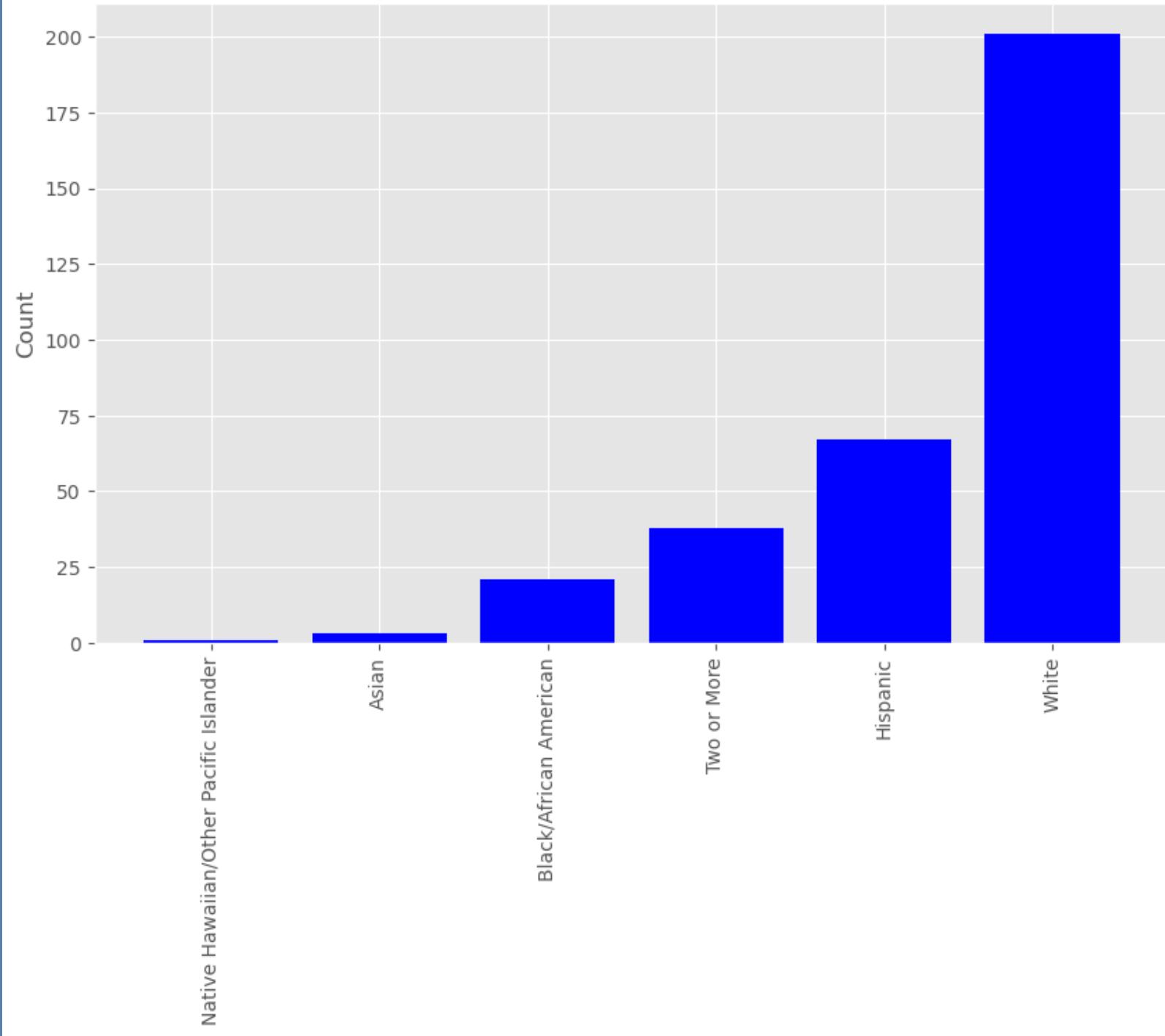
# Early Findings

**My first goal was to explore Mt. Scott's student body, to see what are the racial demographics, what are the grade counts for various factors, and where do students live.**

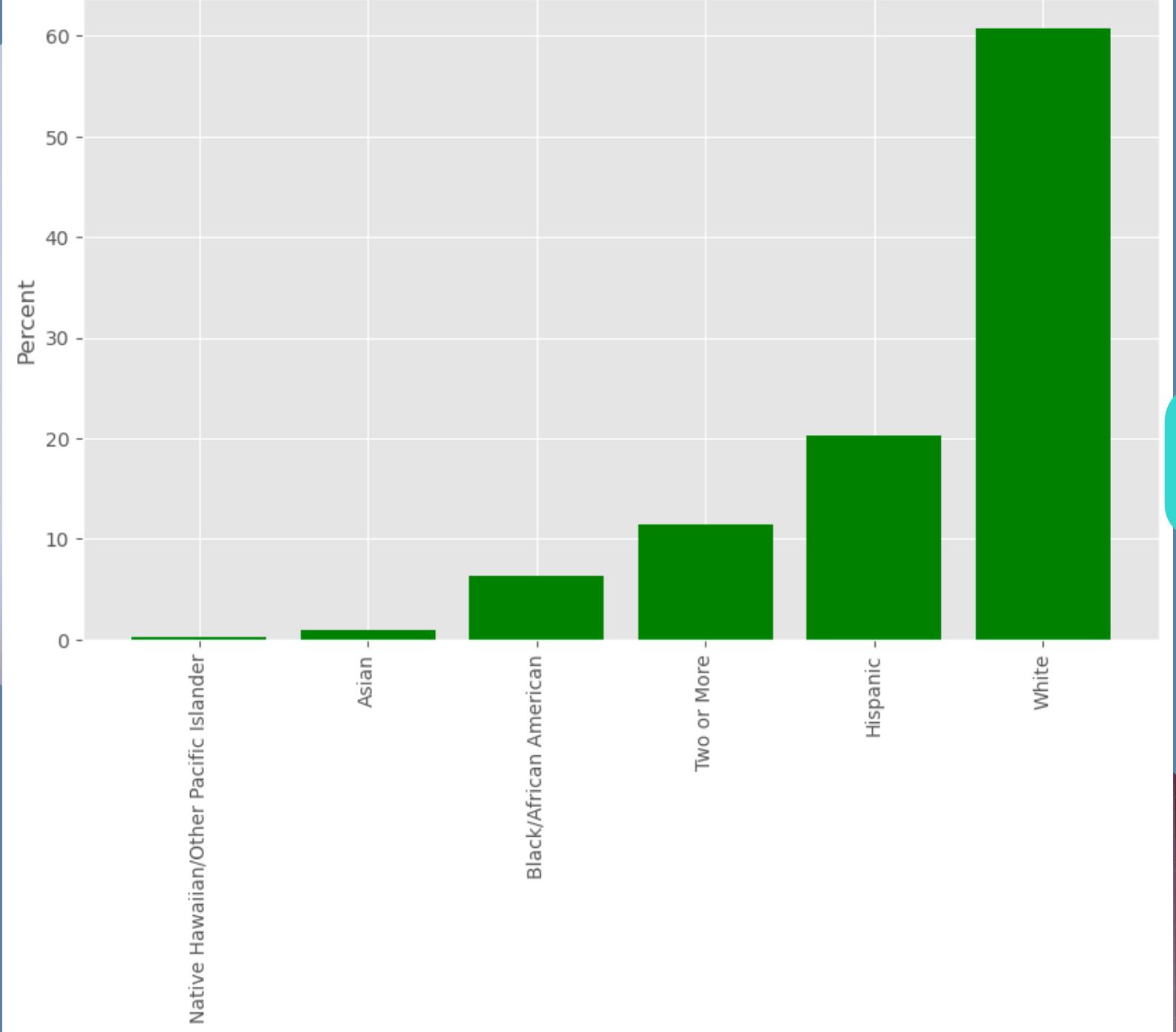


# Racial Data

Total Number of Students Throughout Career Taught by Race



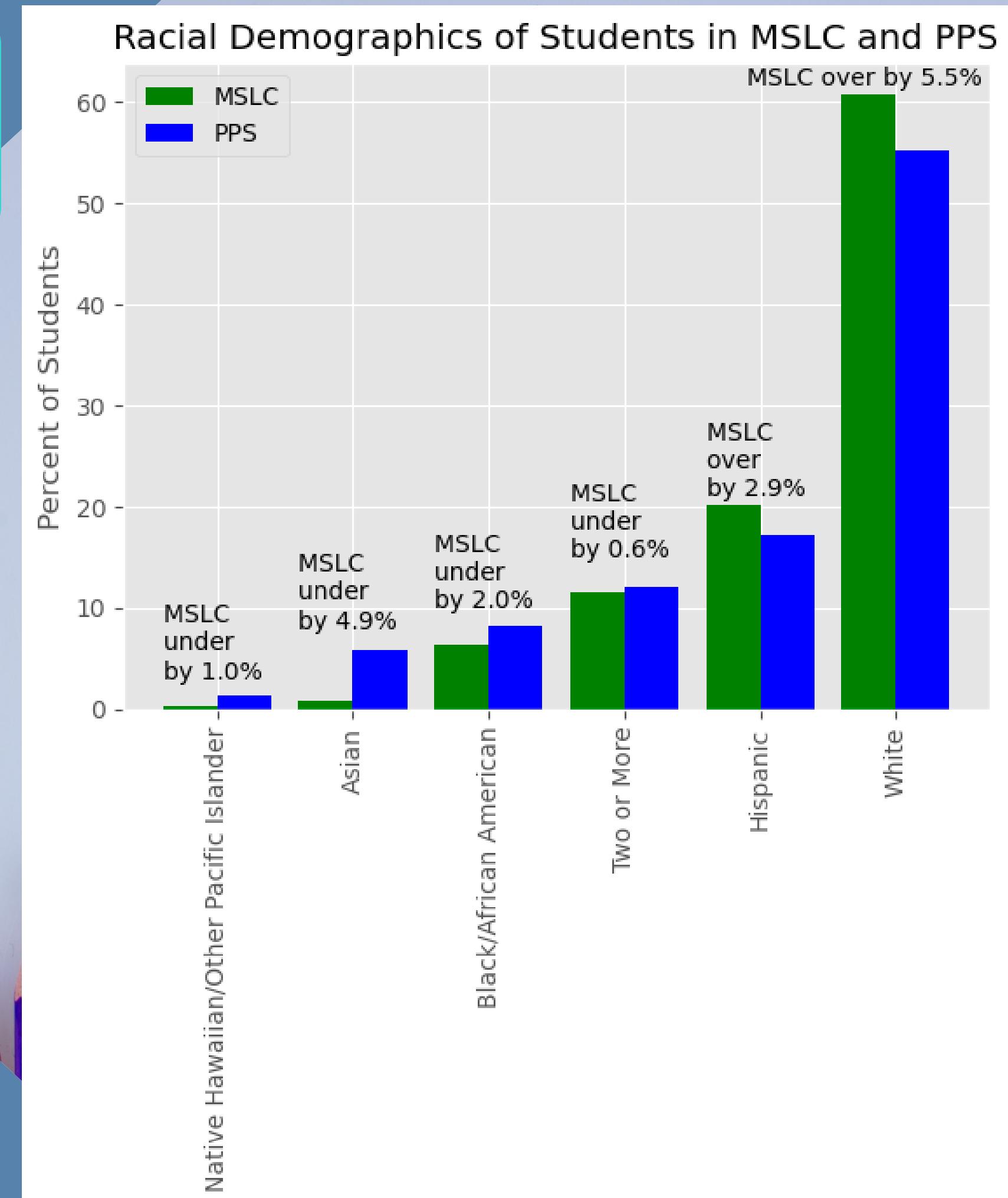
Racial Breakdown of Students



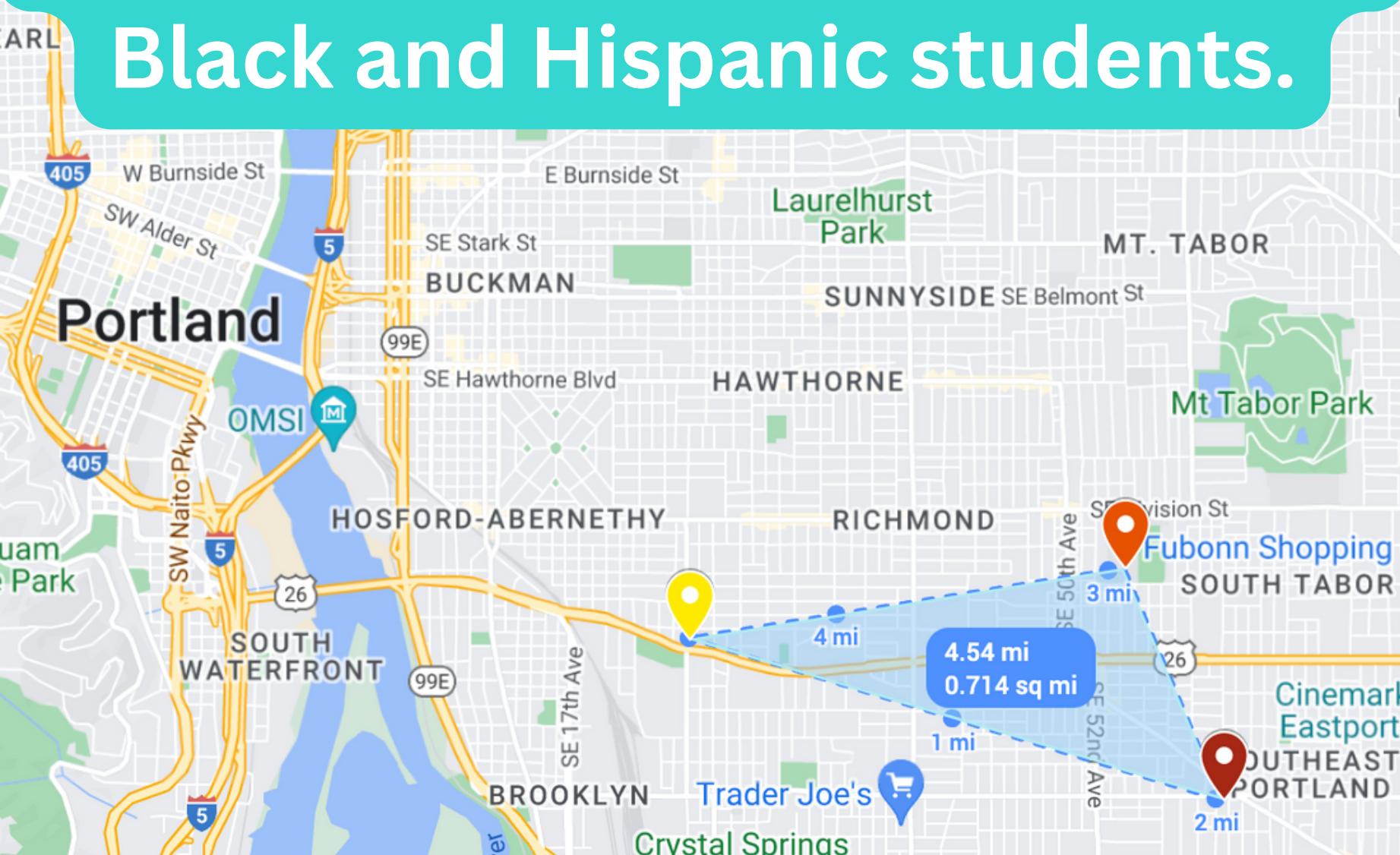
Over  
60% of  
my  
students  
have  
been  
White

**Looking at the demographic data for the whole district, Mt Scott's students are more White and more Hispanic, and less every other racial category**

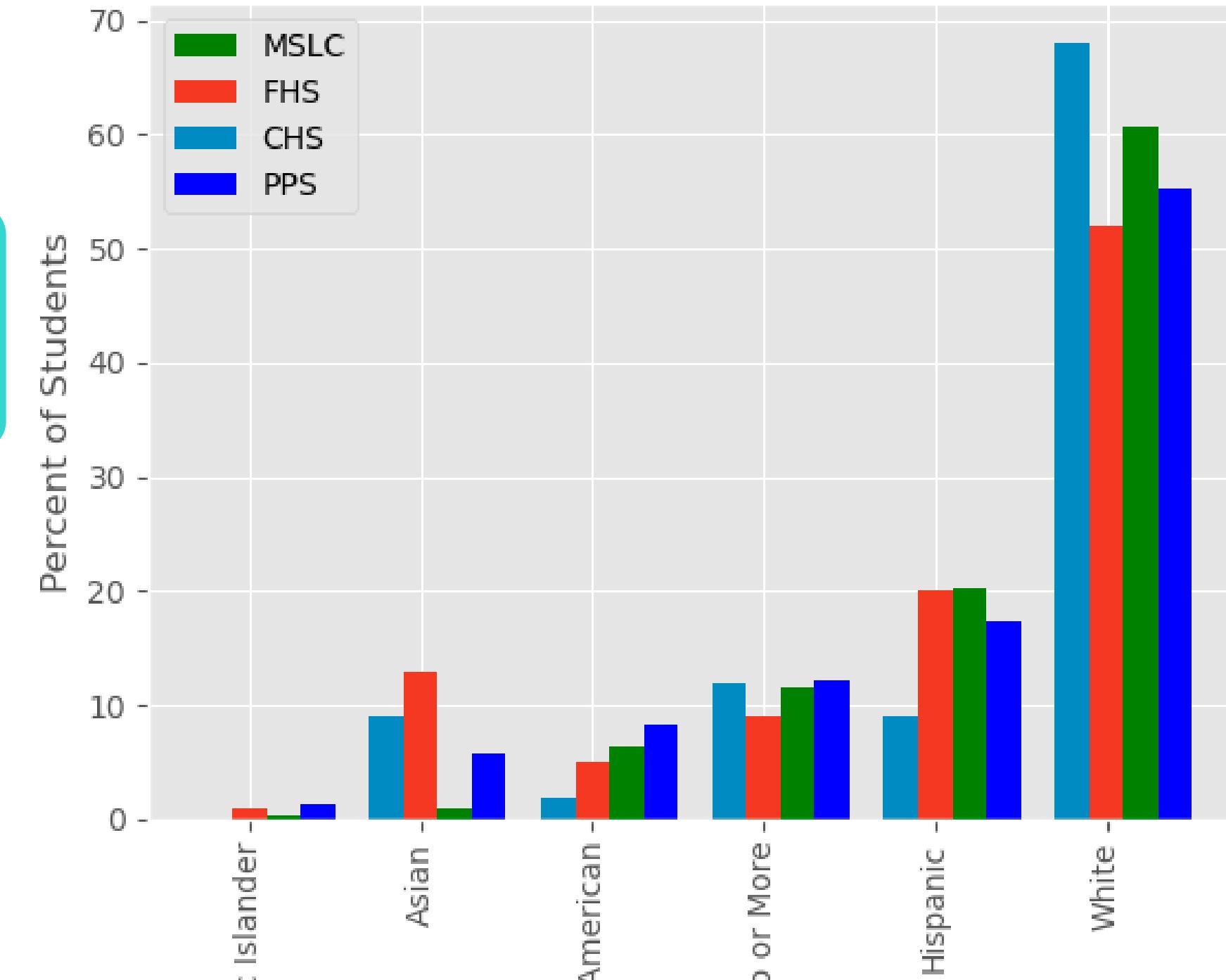
Facts & Figures 2022-23	
Portland Public Schools enrollment and demographic data	
Student Enrollment	Figures
District Schools	43,023
Community Based Programs	683
Special Services Programs	456
Public Charter Schools	1,294
<b>Total</b>	<b>45,456</b>
Student Information	
Hispanic/Latino	17.3%
Black	8.3%
Asian	5.8%
Native American/Alaskan Native	0.5%
Pacific Islander	0.8%
White	55.2%
Multi-Racial - Asian/White	5.4%
Multi-Racial - Other Ancestries	6.7%



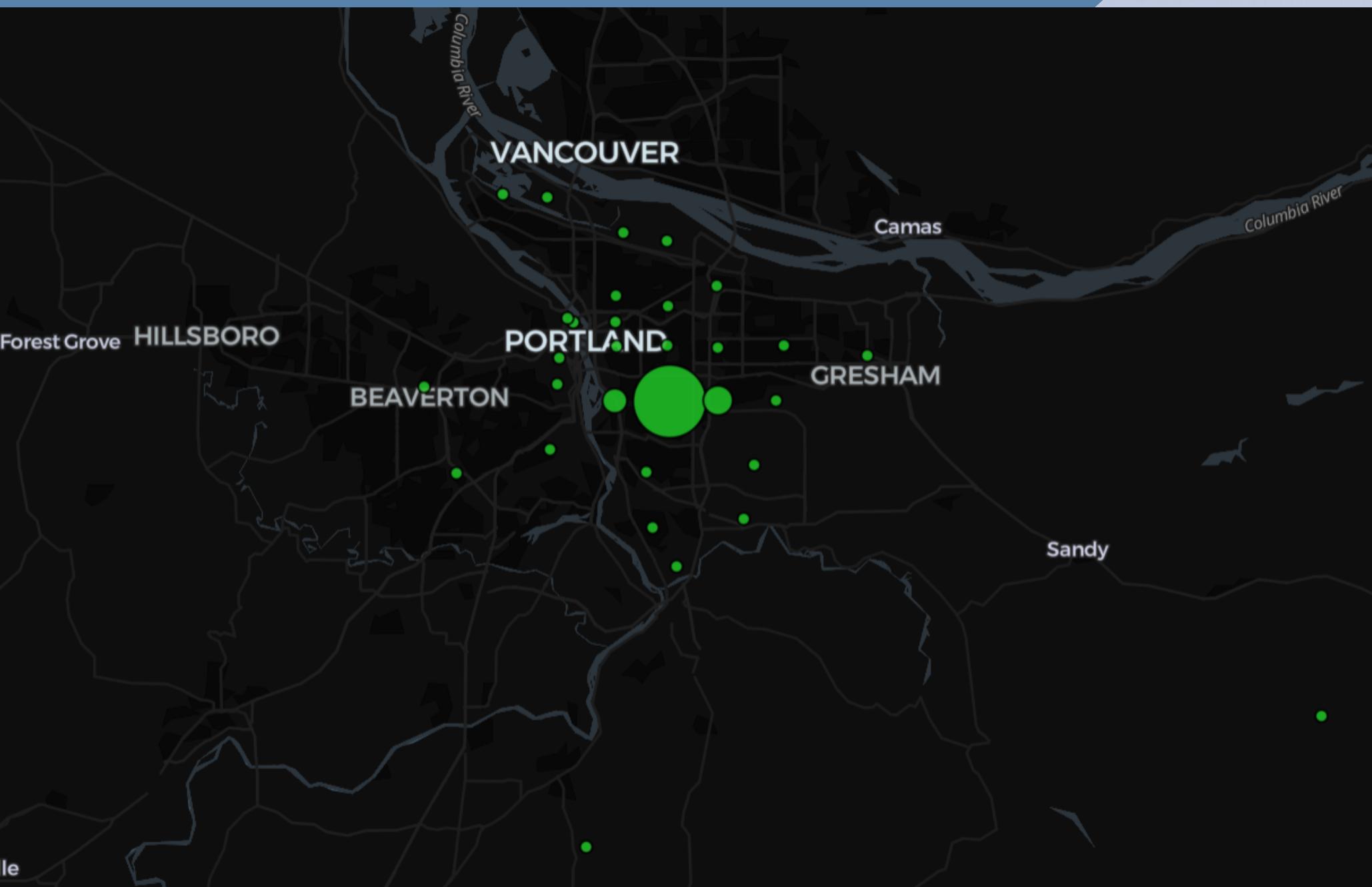
In terms of the location of the school, I wanted to check how those demographics were for our two closest feeder high schools, Cleveland HS and Franklin HS. MSLC has a higher percentage of Black and Hispanic students.



Racial Demographics of Students at FHS, CHS, MSLC and PPS

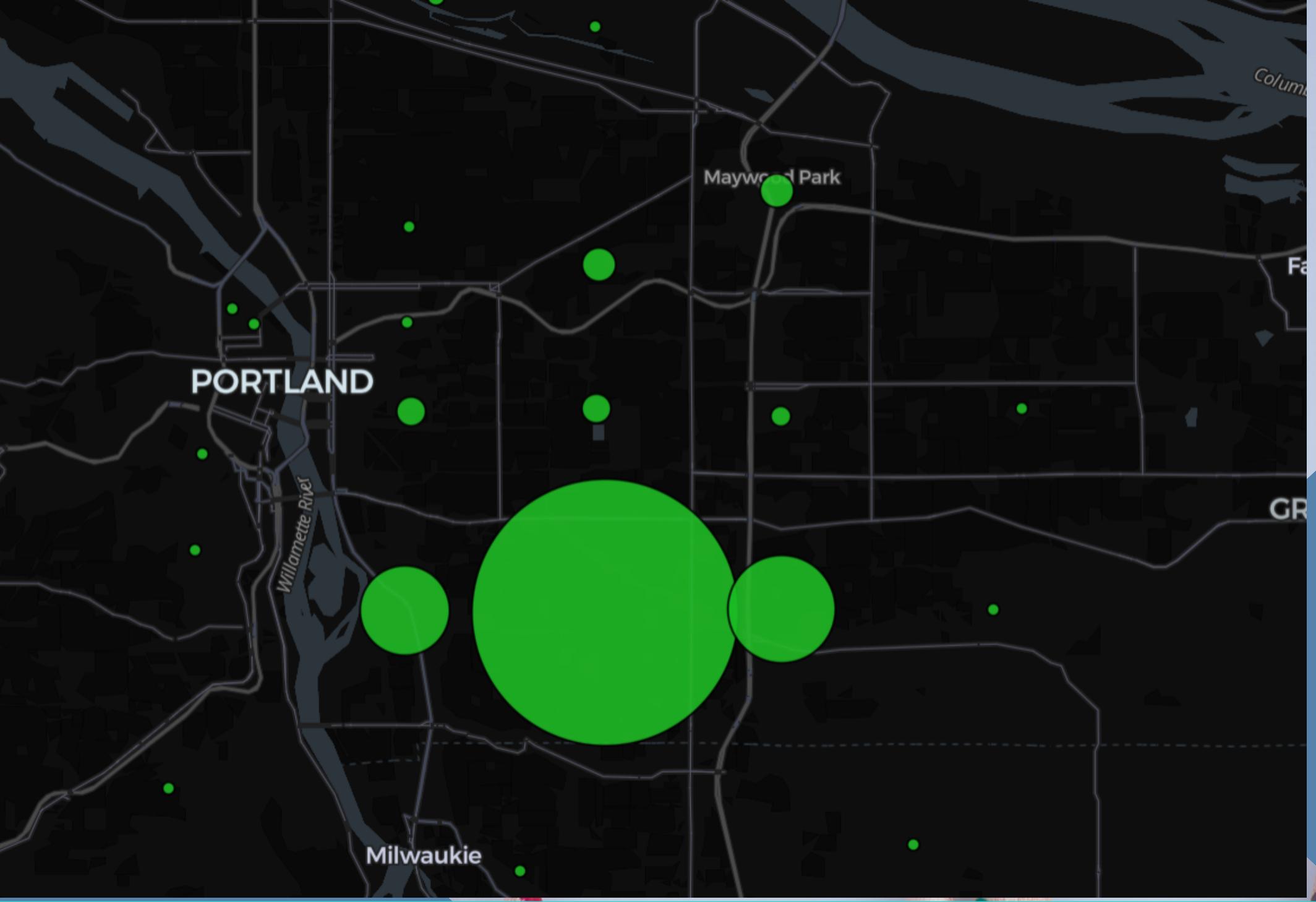


# Student Zip Codes



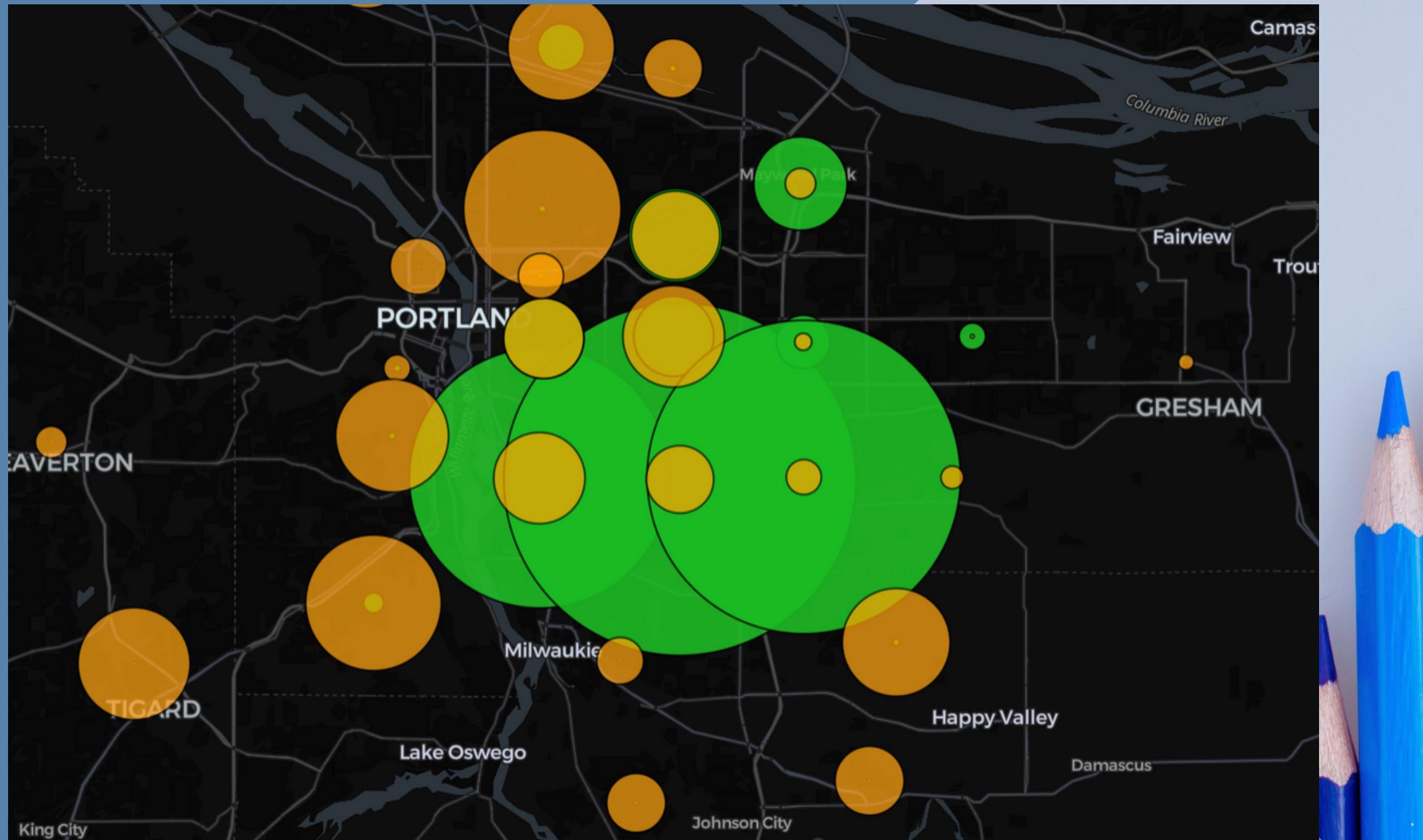
I was also very interested in where our students come from so I created a heat map of where students live by counts of zipcodes





Most students are coming from the East side, with the highest numbers coming from the same zip code as the school, 97206, then the eastern adjacent zip code, 97266, and then the western adjacent zip code, 97202.

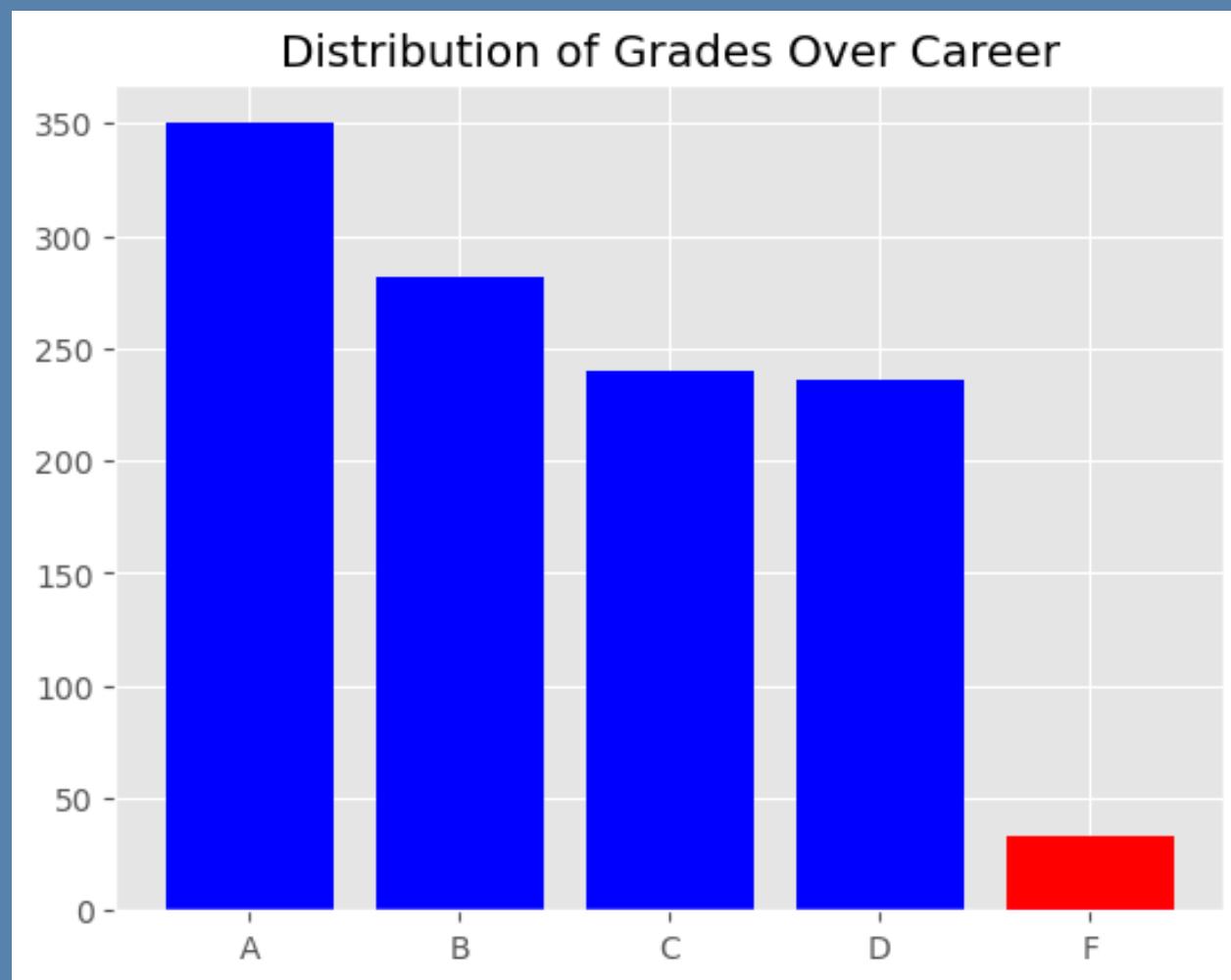
But zip code can also be an indicator of many other factors. One of which is income. I created another dataset from Census information on median incomes of zip codes for 2021. This map shows the scaled data for student count in green and scaled data for median income for each zip code in green.



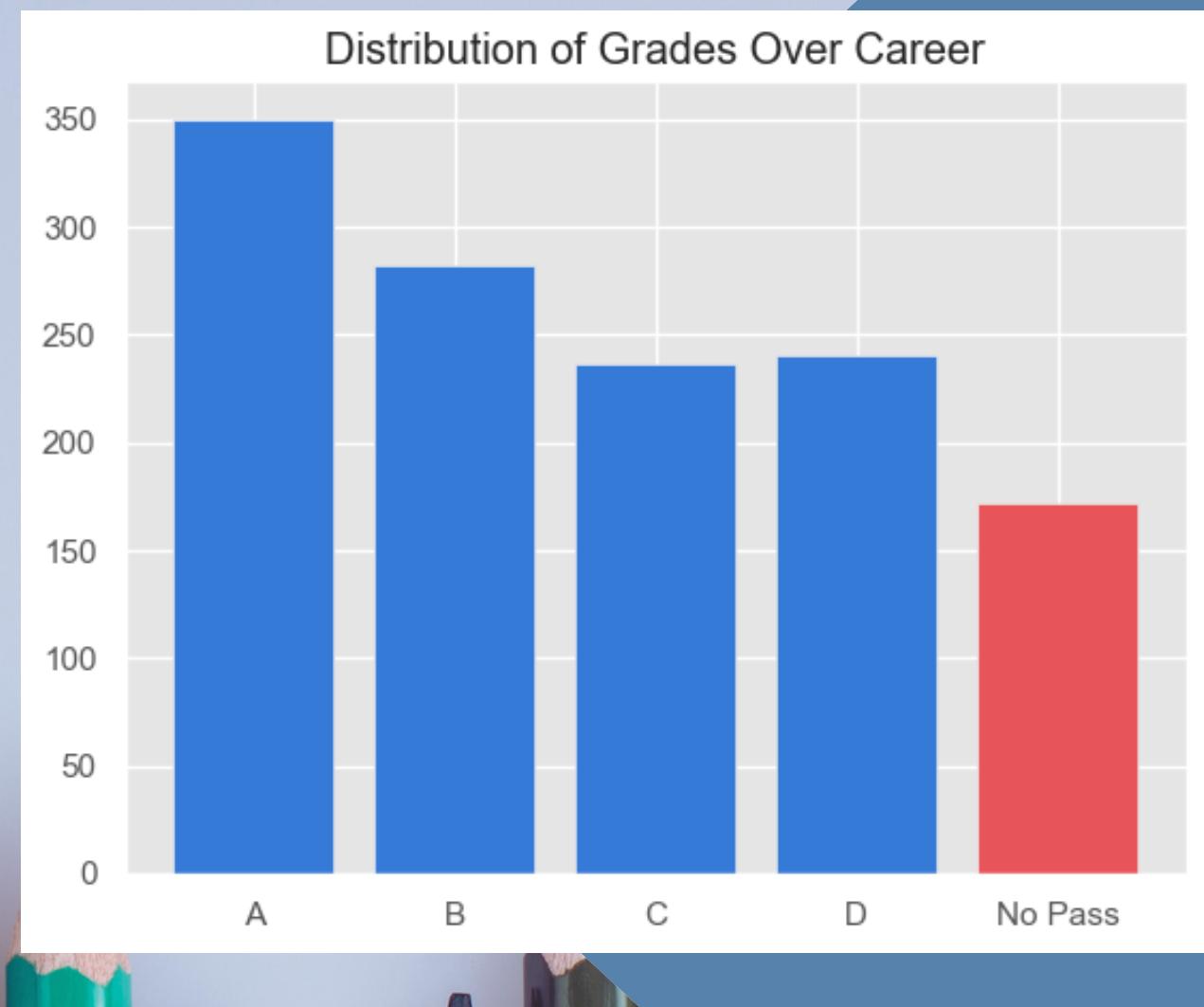
### Things to note

Zip Code	Count Rank	Income Rank
97206	1st	15th
97266	2nd	23rd
97202	3rd	9th
97213	t.4th	11th
97220	t.4th	25th

# Total Grade Distribution

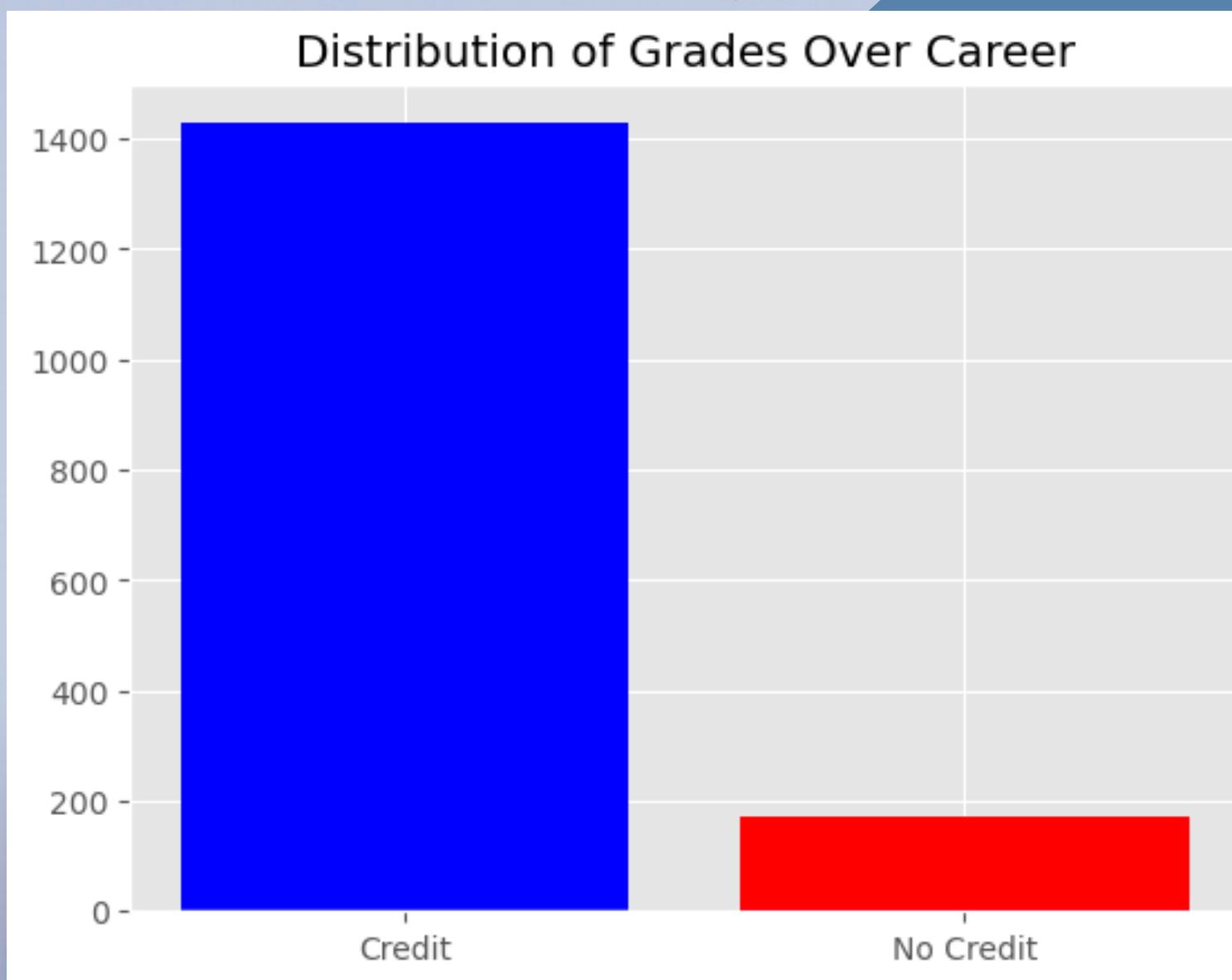
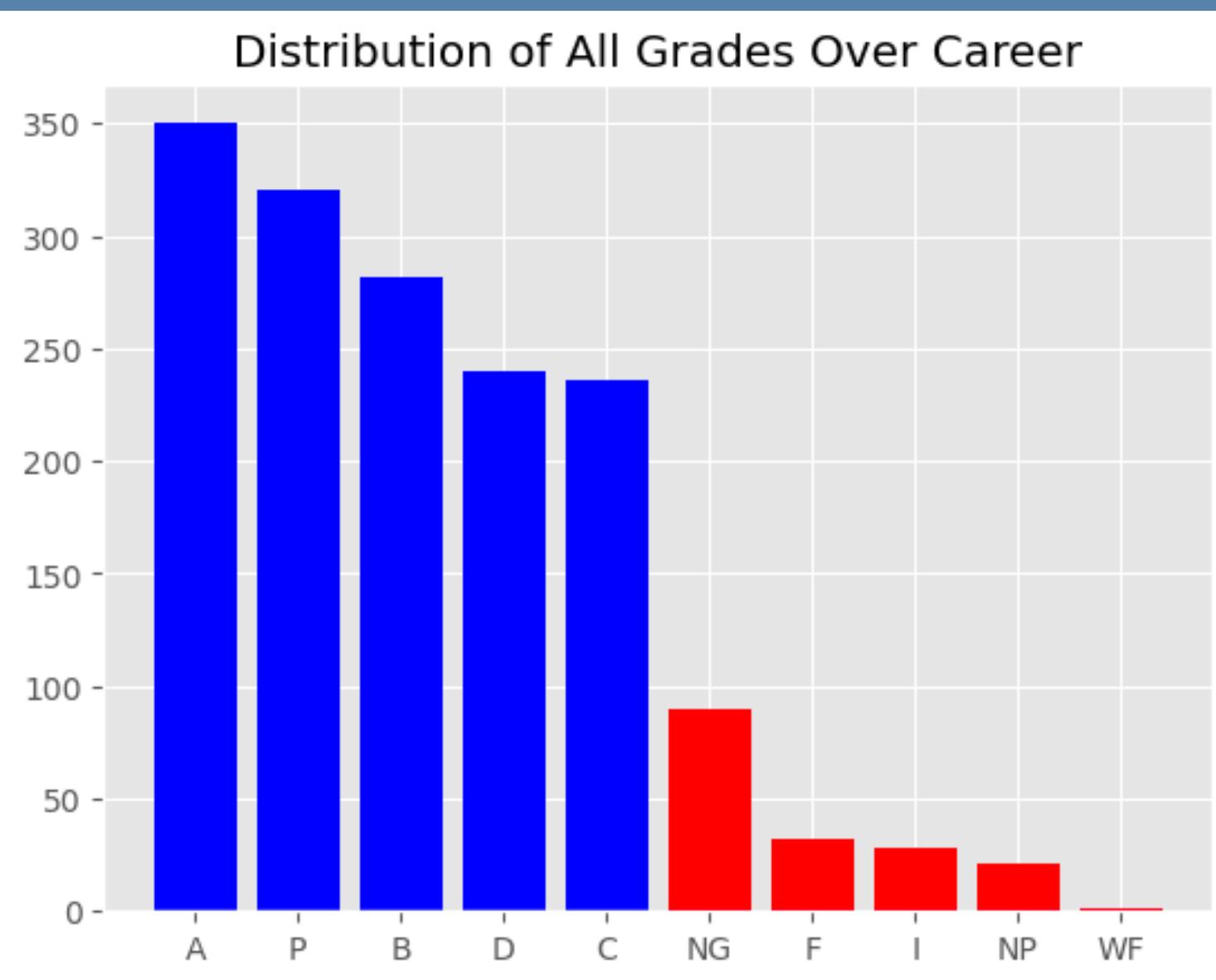


From the graph on the left, it looks like I've given out practically no Fs.



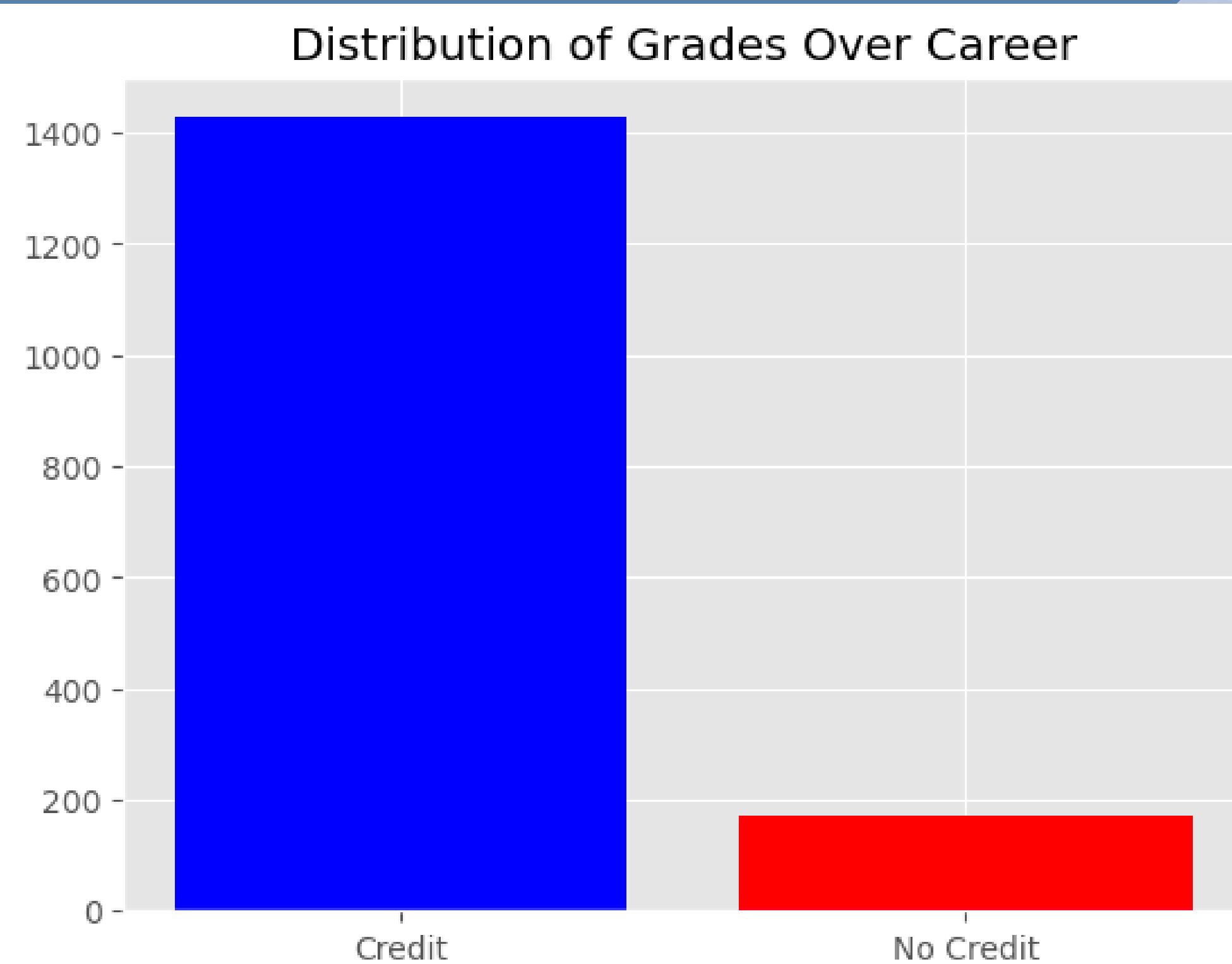
The reason why there are such low numbers is because the school stopped using Fs as a grading letter after lockdown. Students received either NPs or NGs which indicated that students did not receive credit, but the grade would not affect their GPA like an F does.

# Total Grade Distribution



Here are all the grade codes over the years on the right.  
And the left shows A-D combined in Credit, and everything  
else combined in No Credit

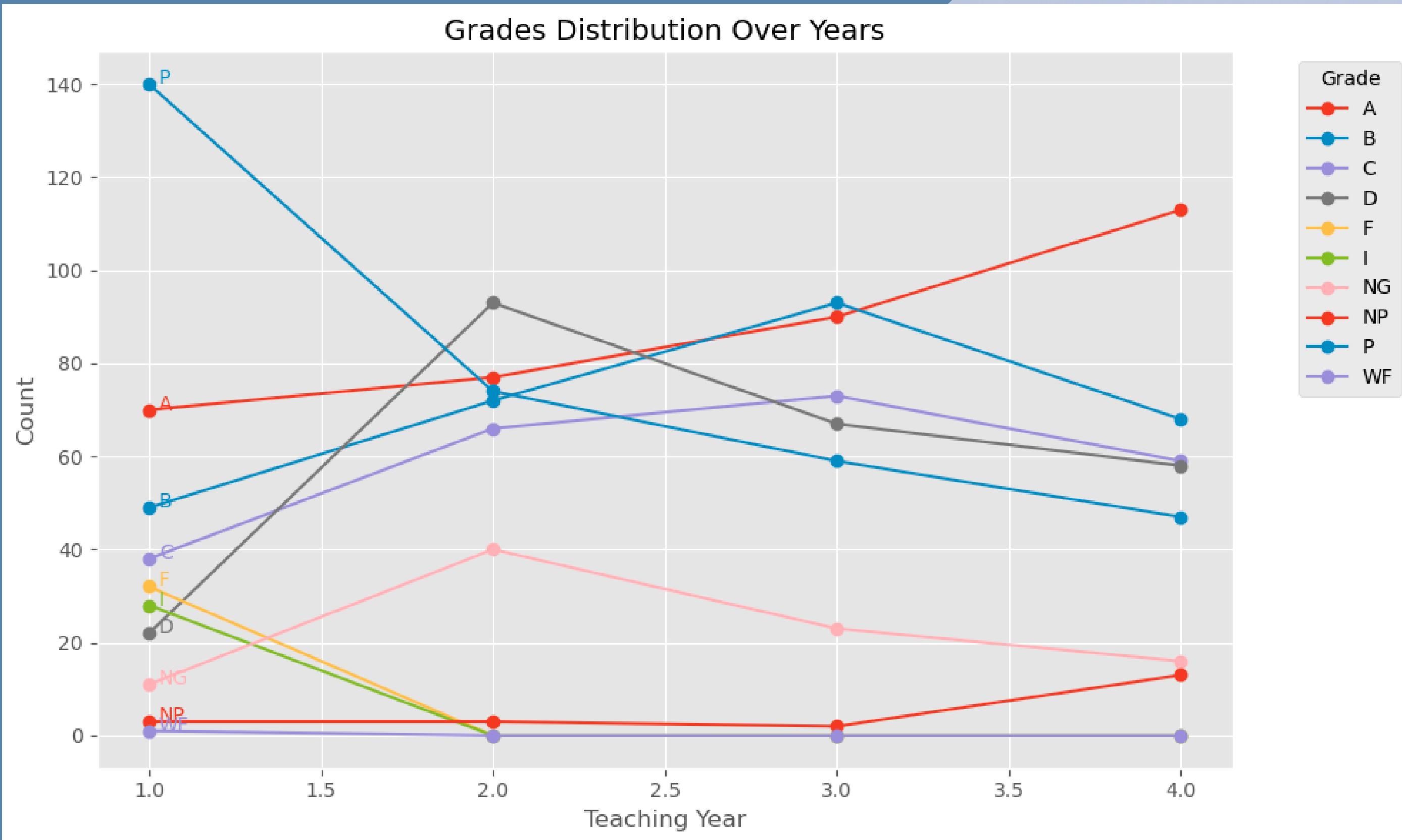
# Total Grade Distribution



Over the years, students have earned credit in my classes at a rate of

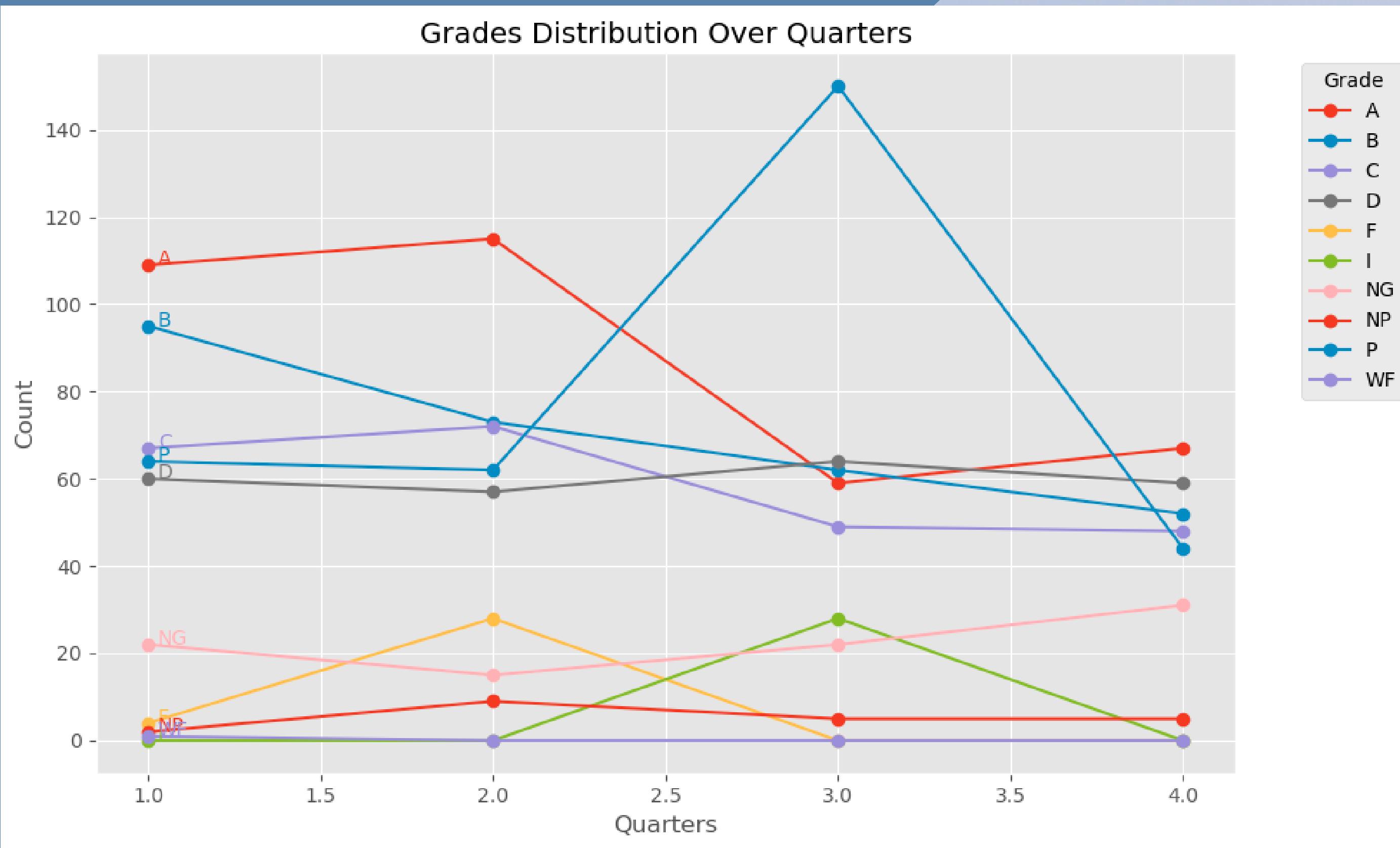
7:1

# Yearly Grade Distribution



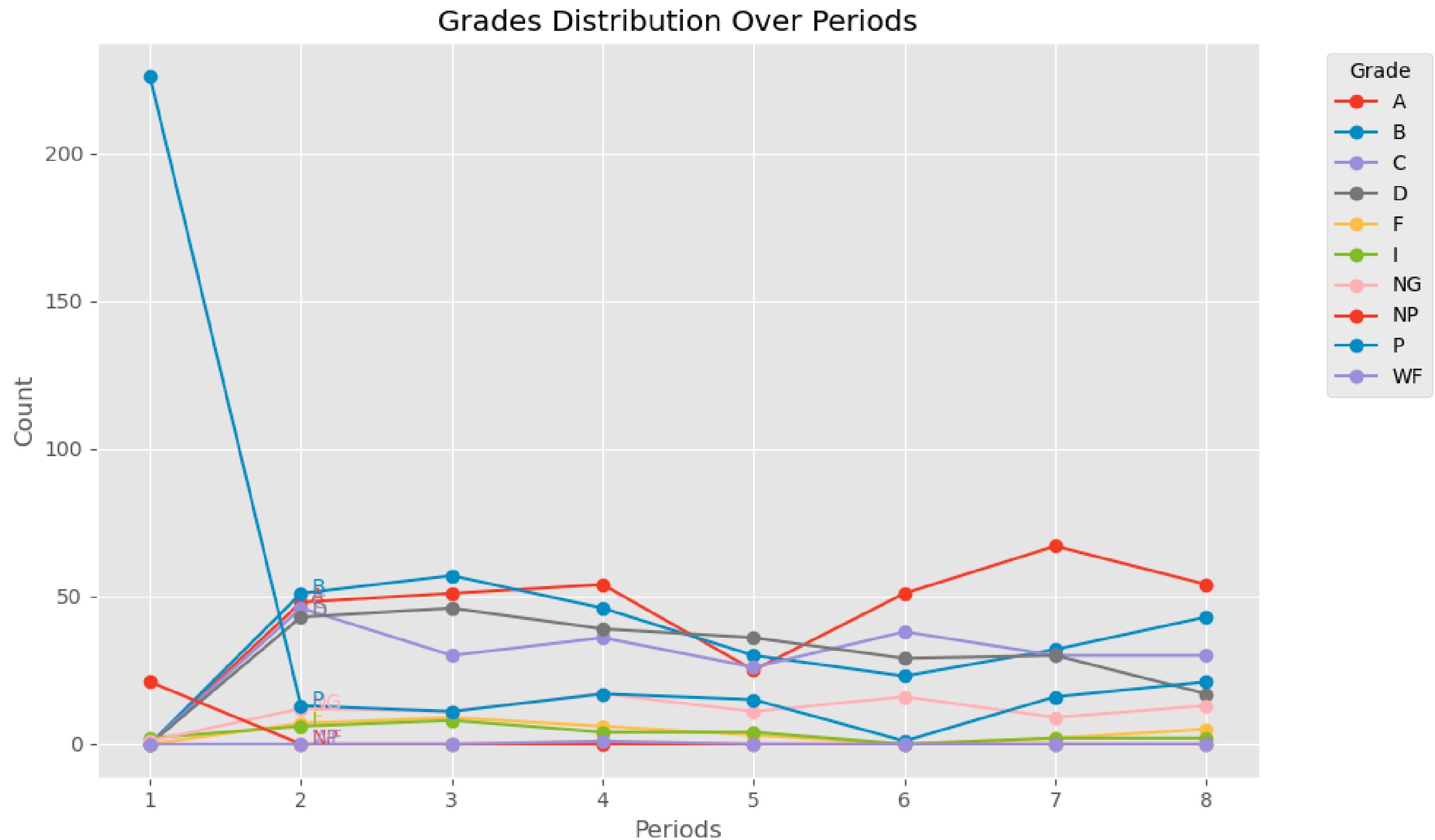
As have been  
on the  
increase over  
the years  
while NGs  
have almost  
gotten back to  
the same state  
as before  
Covid.

# Quarterly Grade Distribution



Grades really take a hit third quarter from As and Bs. Also, there seems to be the most NGs 4th quarter.

# Period Grade Distribution

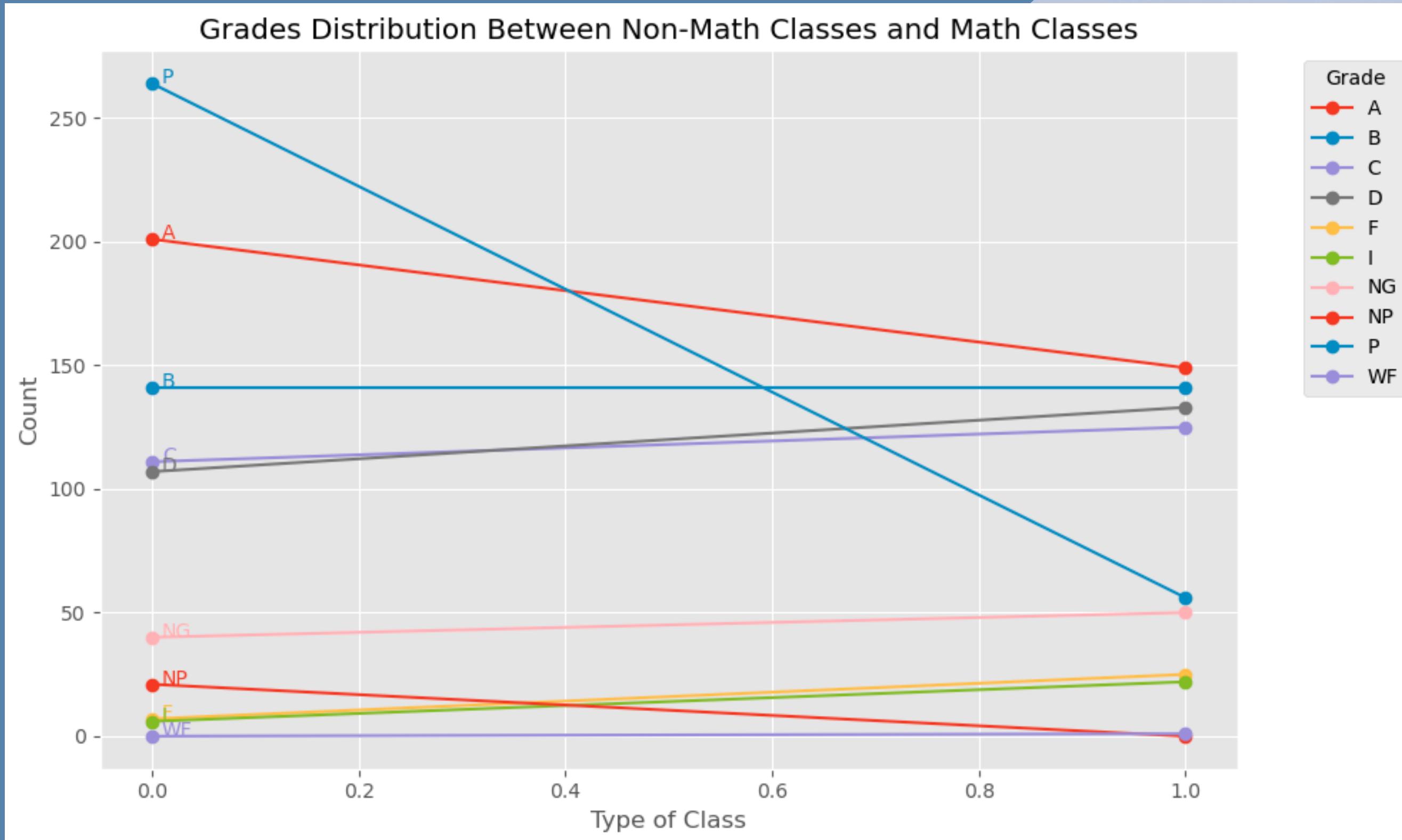


5th Period, the period right before lunch has the least amount of As.

6th Period, the period after lunch for half the students, has the most NGs

NGs

# Math/Non-Math Distribution



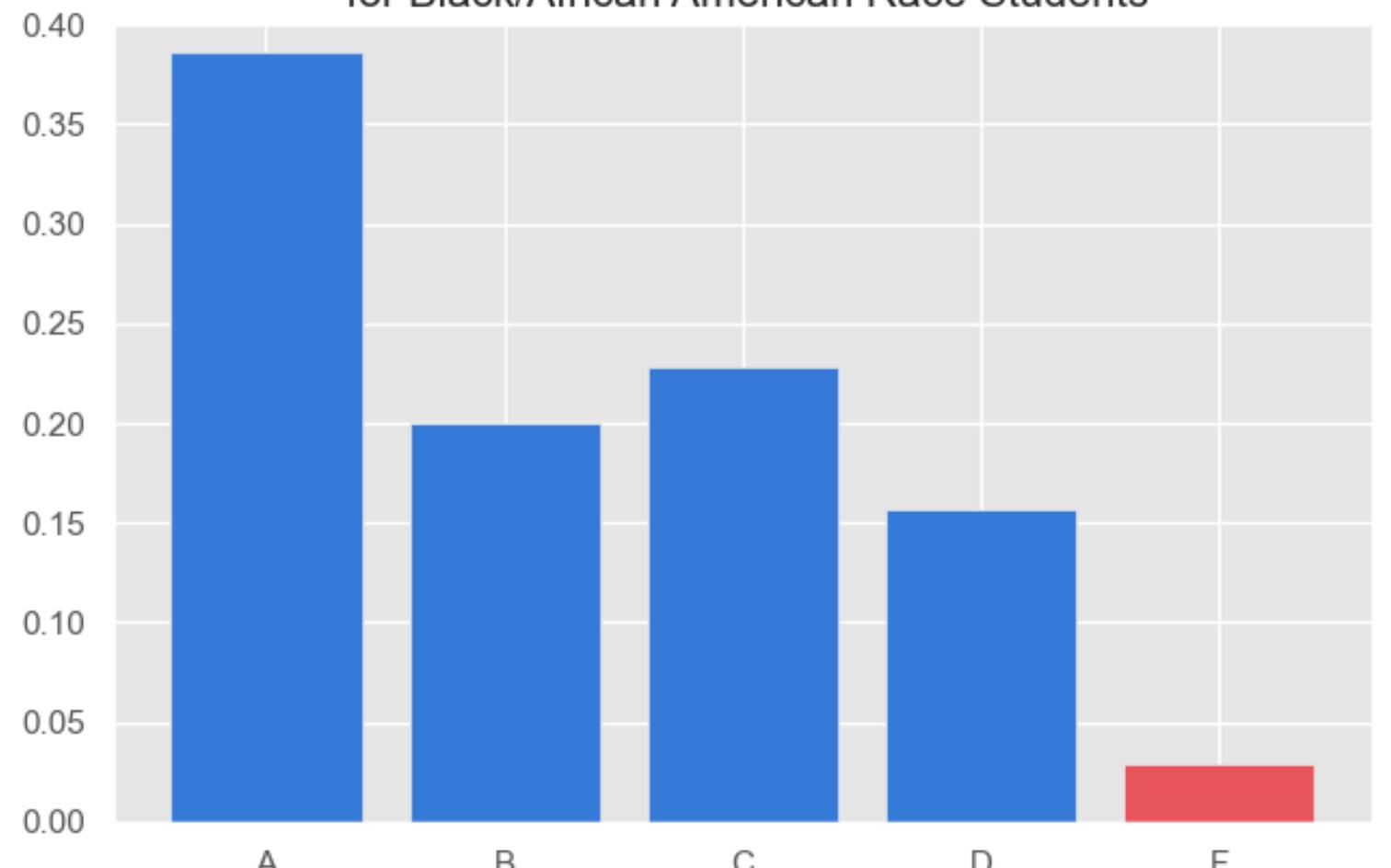
Non-math  
classes have  
more As and Bs  
and less NGs.

Math classes  
have more Cs  
and Ds.

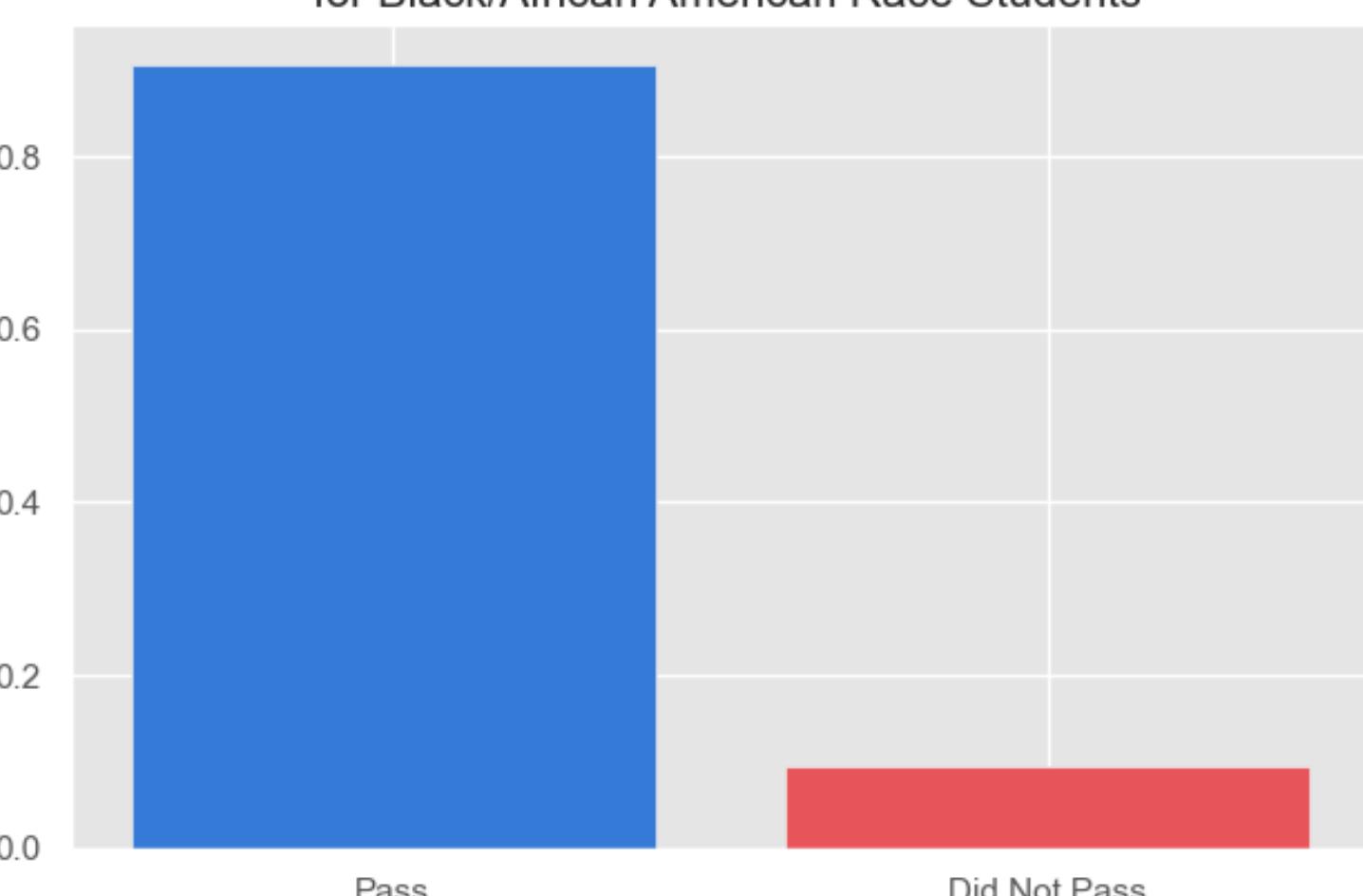
# Racial Grade Distribution

I'm very interested in the racial breakdown of grades in my classes. Since there are so few percentages of students who identify as Asian or Native Hawaiian/Pacific Islander, I've chosen to focus on Black, Hispanic, Two or More Races, and White

Grade Distribution  
for Black/African American Race Students



Credit Aquisition Percent  
for Black/African American Race Students



**Black**

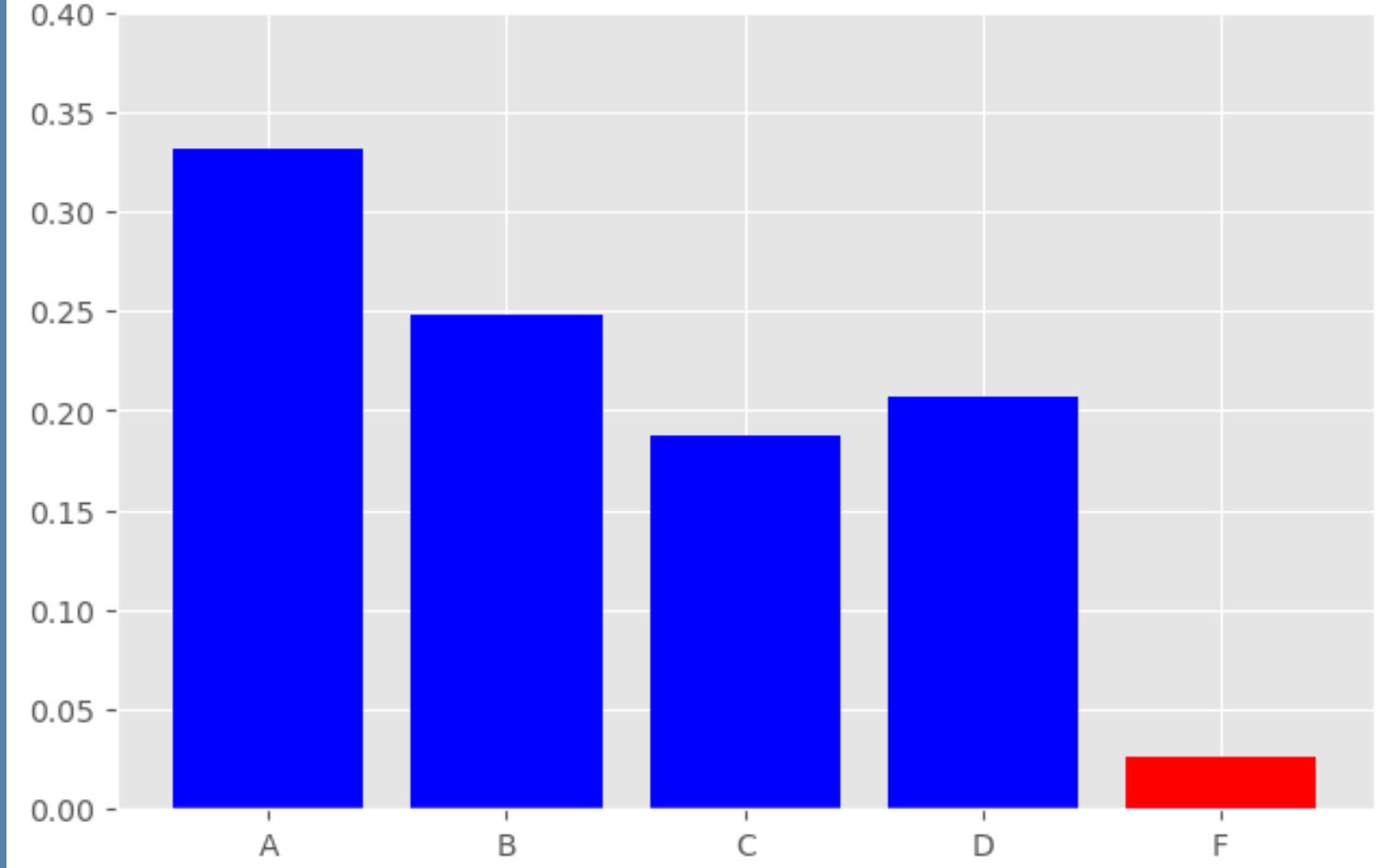
**Pass Rate:**  
**90.6%**

**Ranking of**

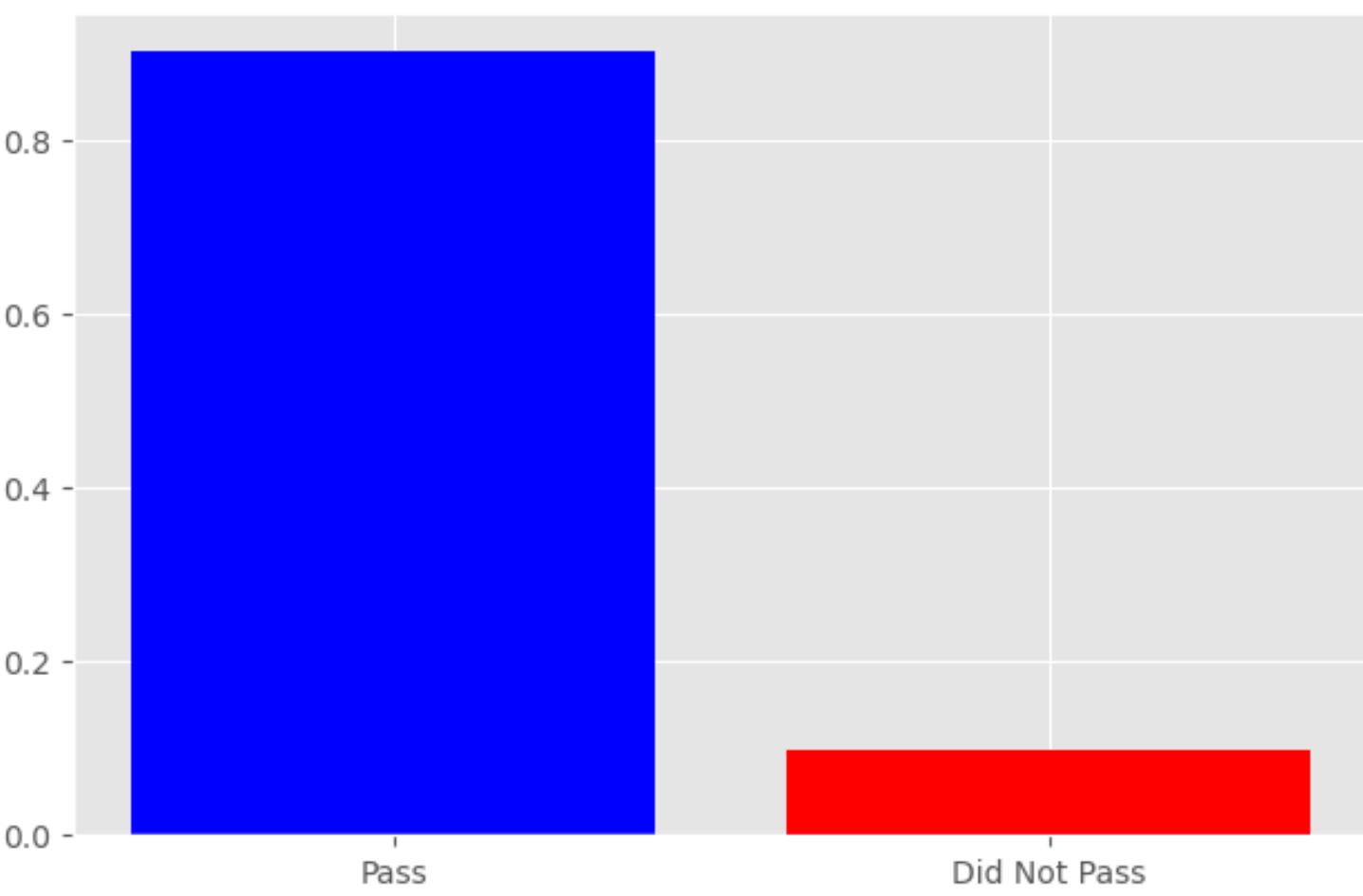
**A-B:**

**A, C, B, D**

Grade Distribution  
for White Race Students



Credit Aquisition Percent  
for White Race Students



**White**

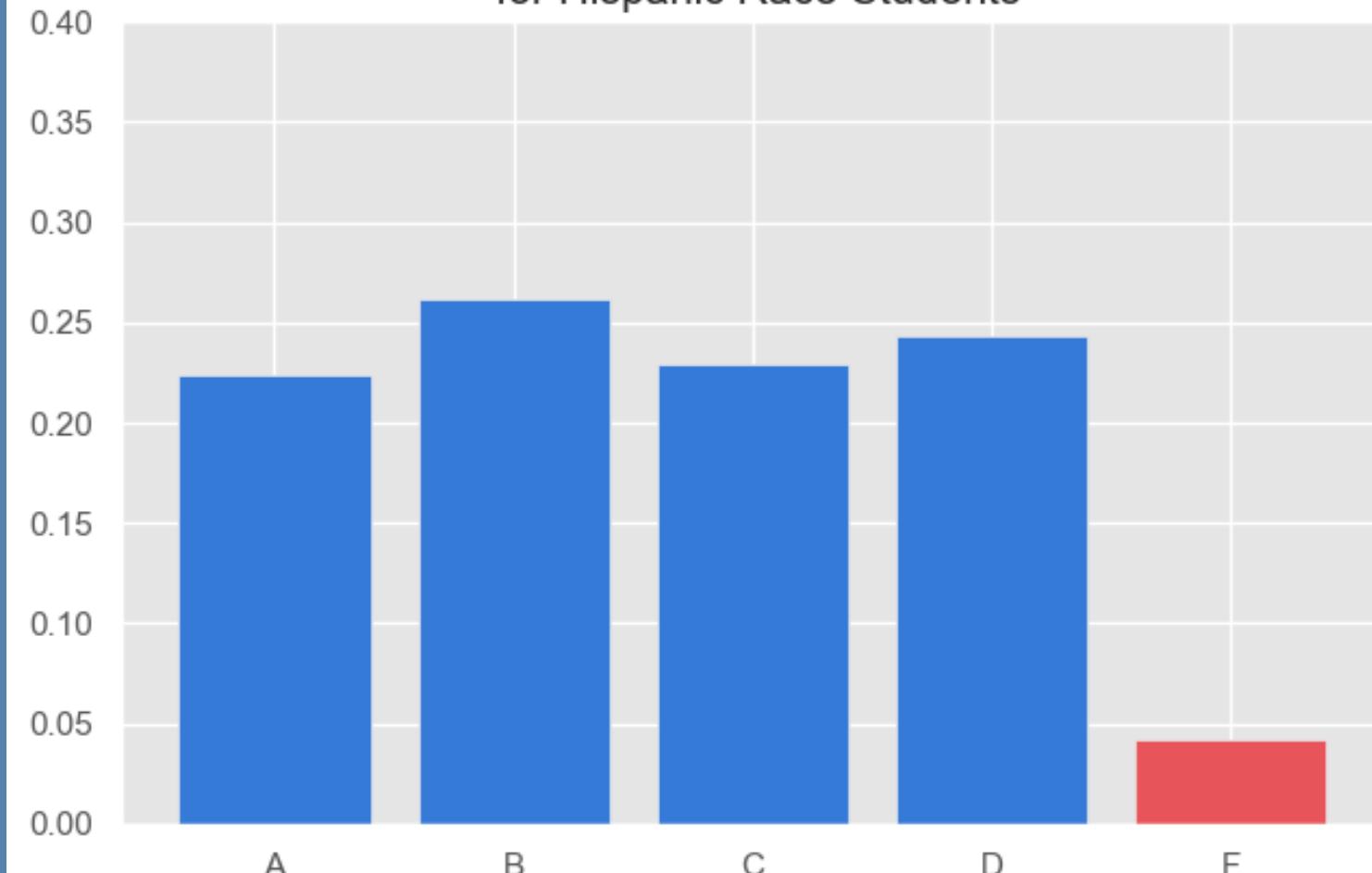
**Pass Rate:**  
**90.1%**

**Ranking of**

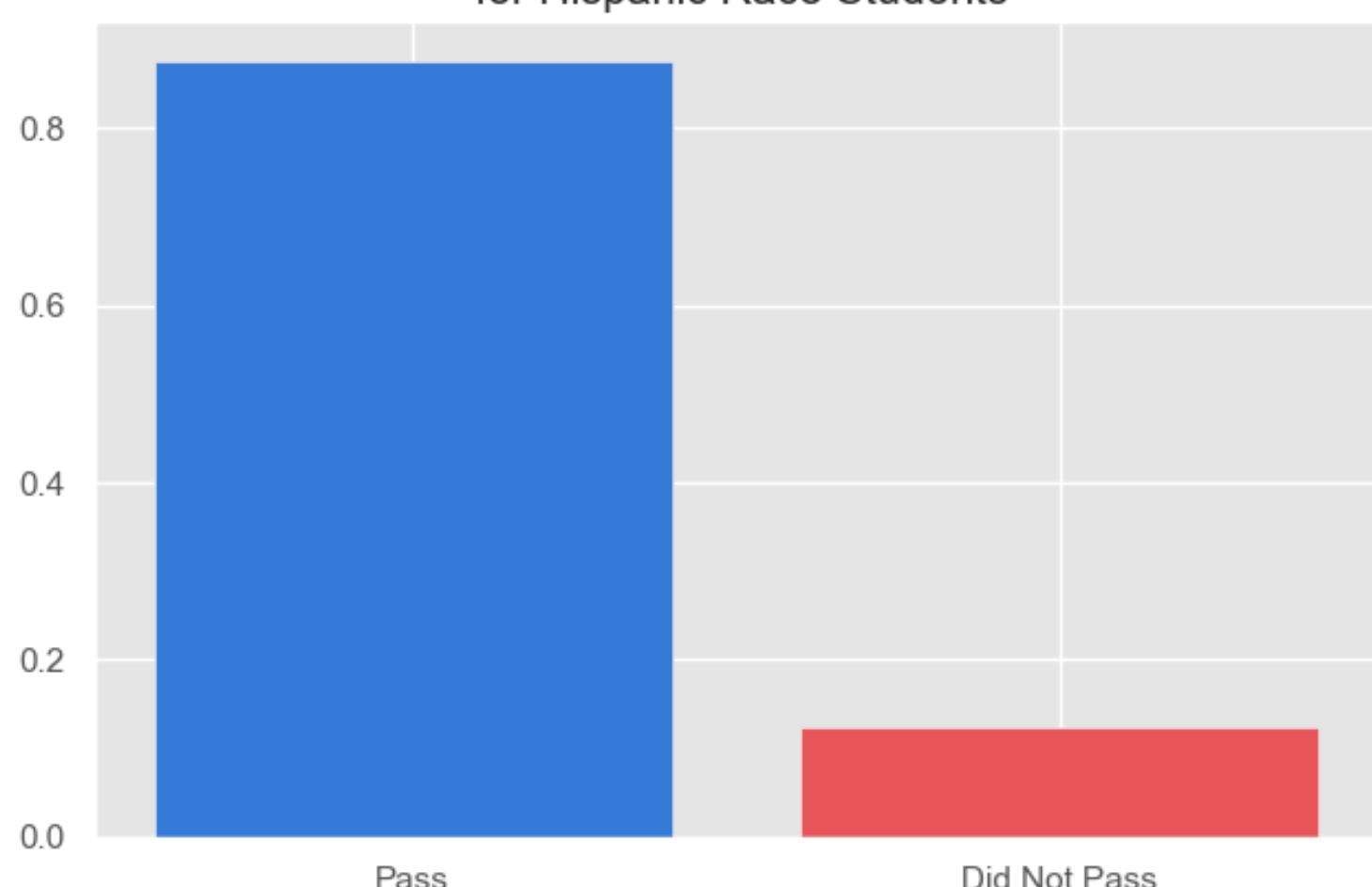
**A-B:**

**A, B, D, C**

Grade Distribution  
for Hispanic Race Students

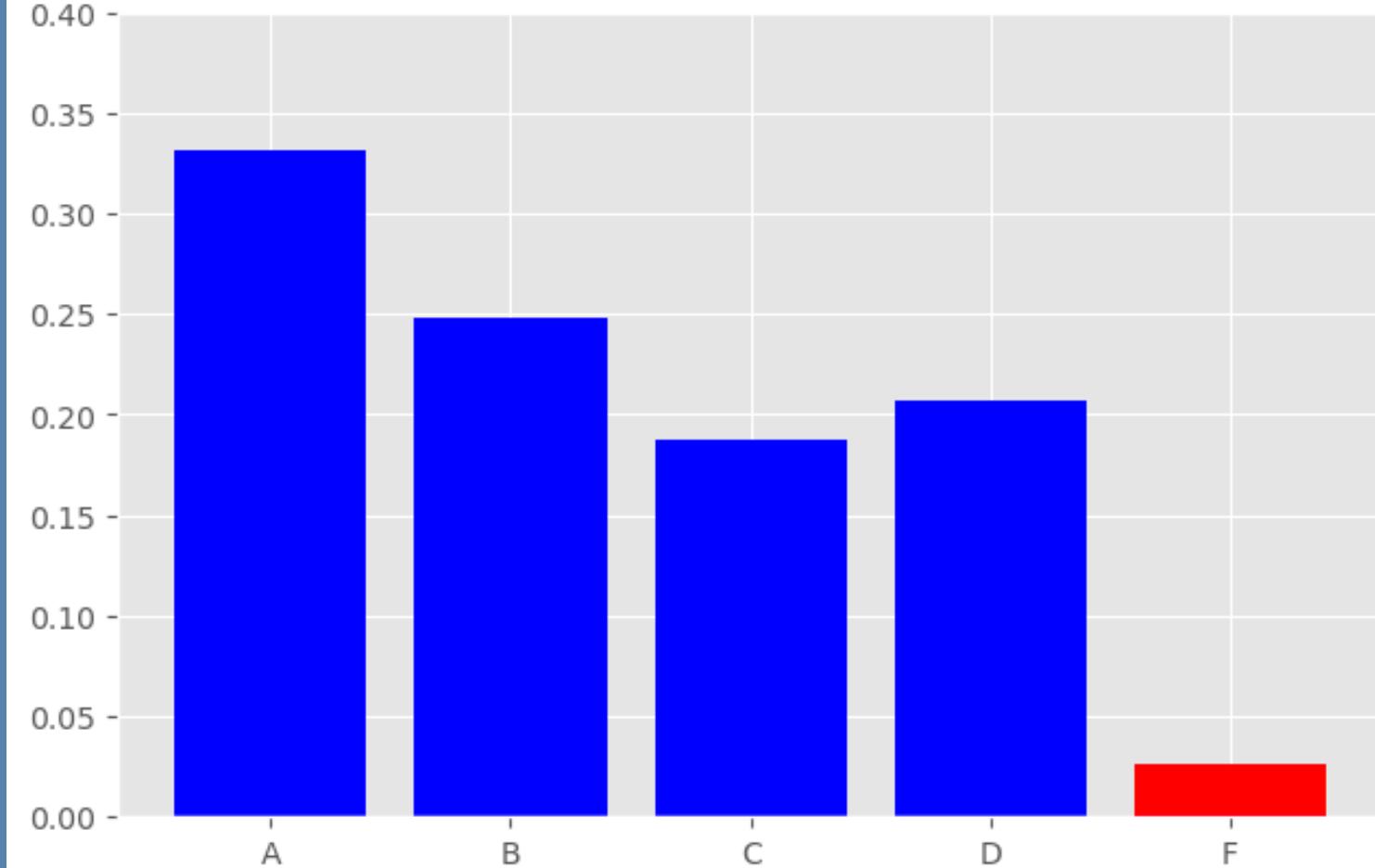


Credit Aquisition Percent  
for Hispanic Race Students

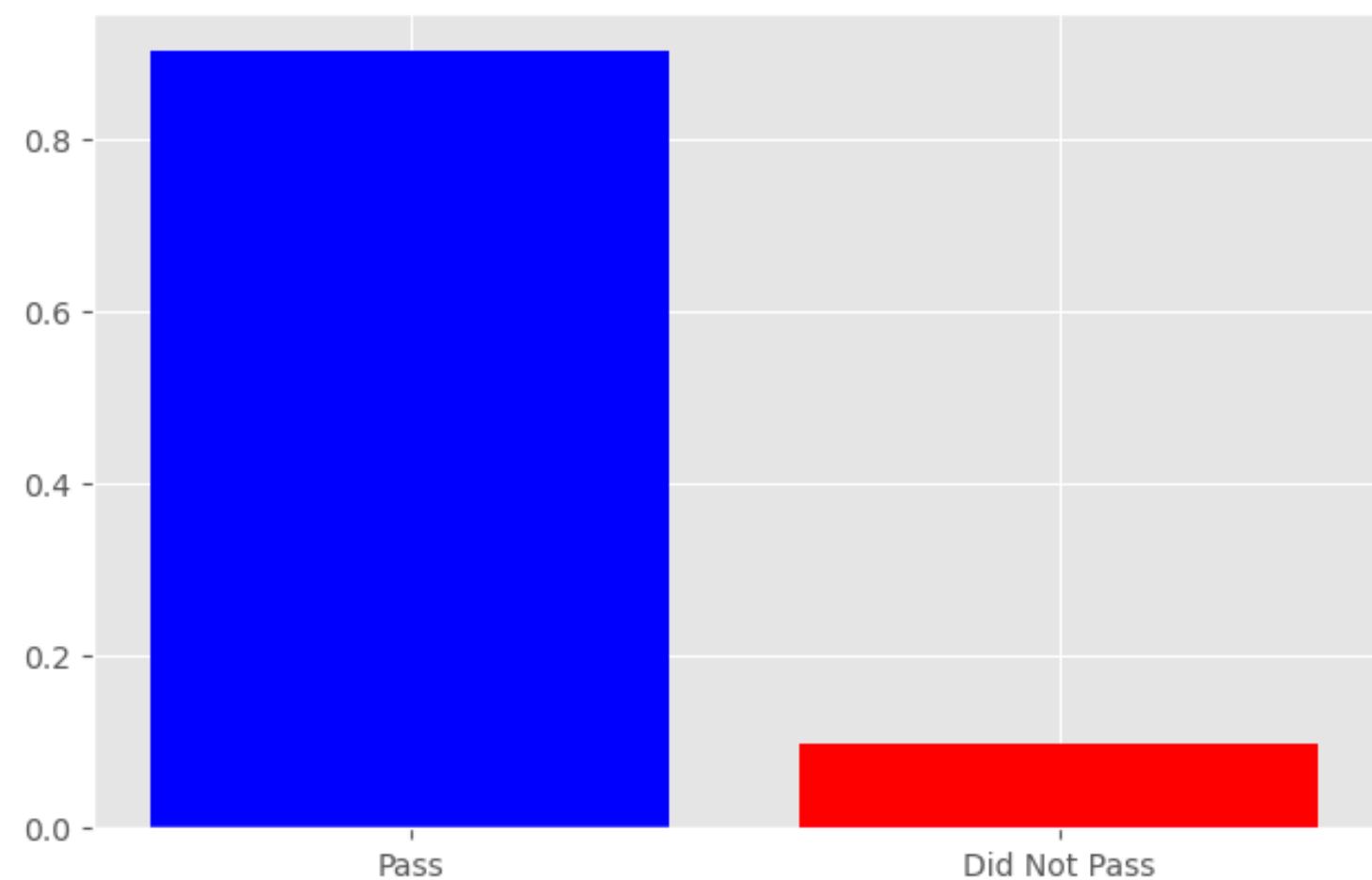


**Hispanic**  
Pass Rate:  
87.5%  
Ranking of  
A-B:  
B, D, C, A

Grade Distribution  
for White Race Students

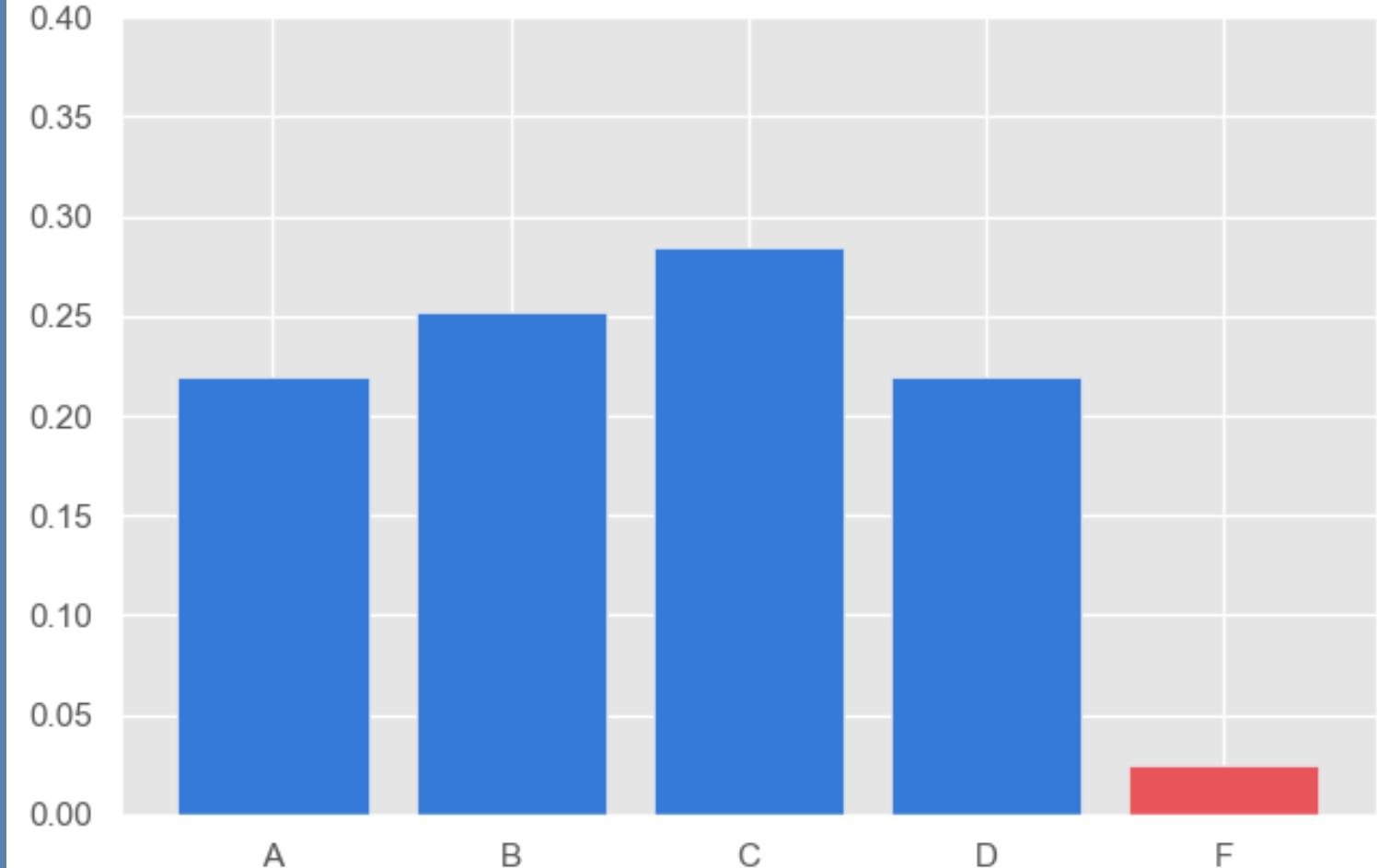


Credit Aquisition Percent  
for White Race Students

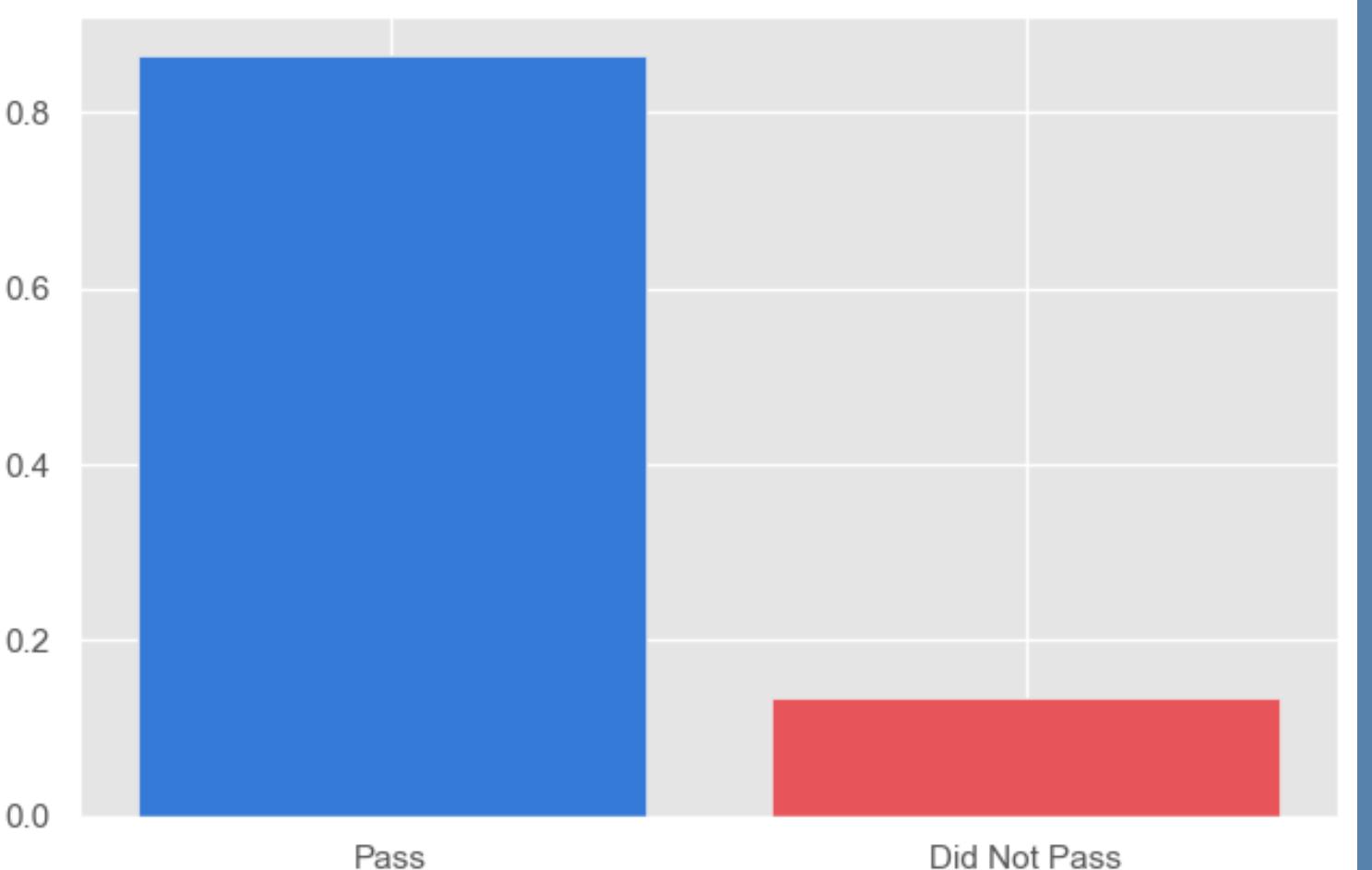


**White**  
Pass Rate:  
90.1%  
Ranking of  
A-B:  
A, B, D, C

Grade Distribution  
for Two or More Race Students

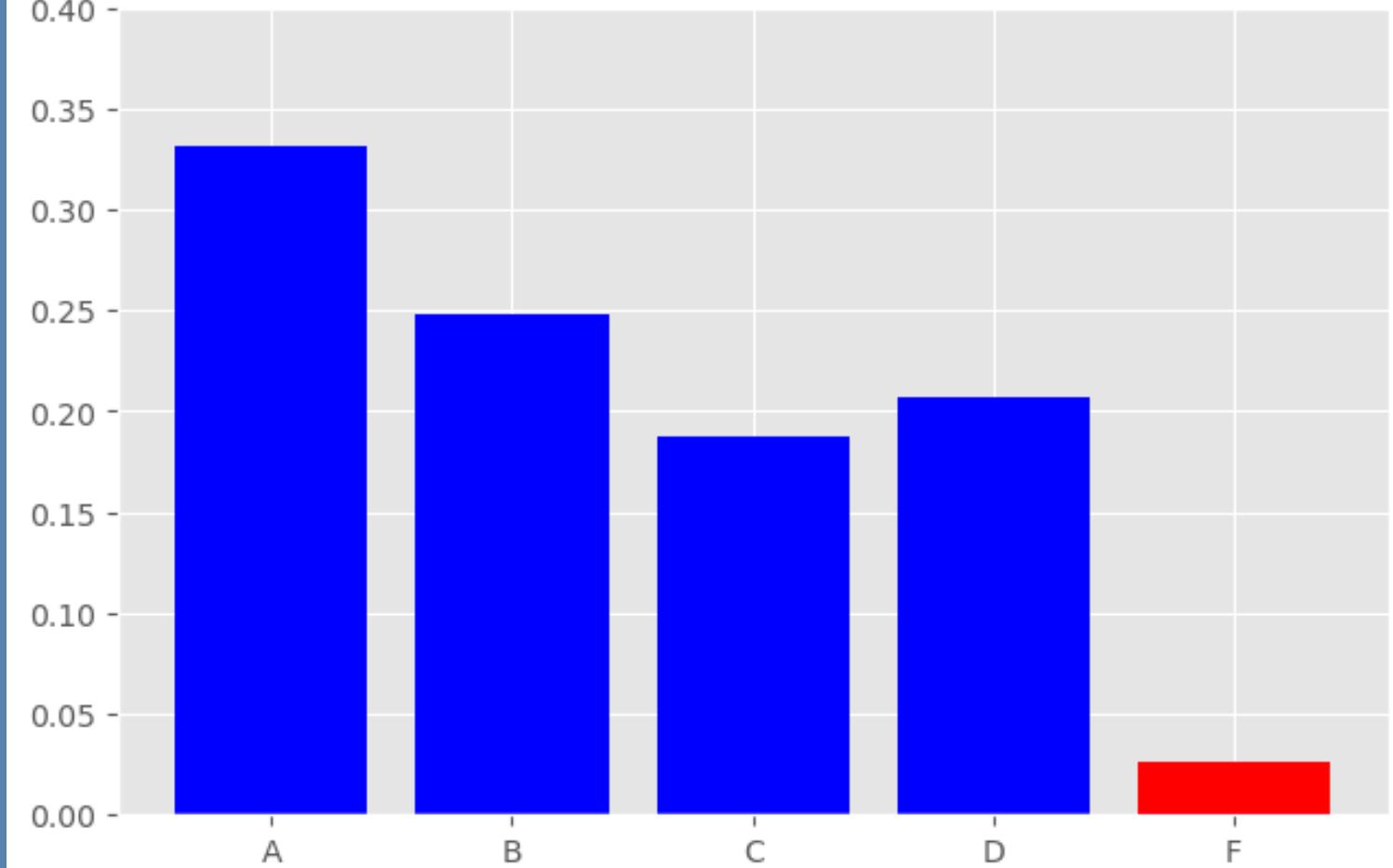


Credit Aquisition Percent  
for Two or More Race Students

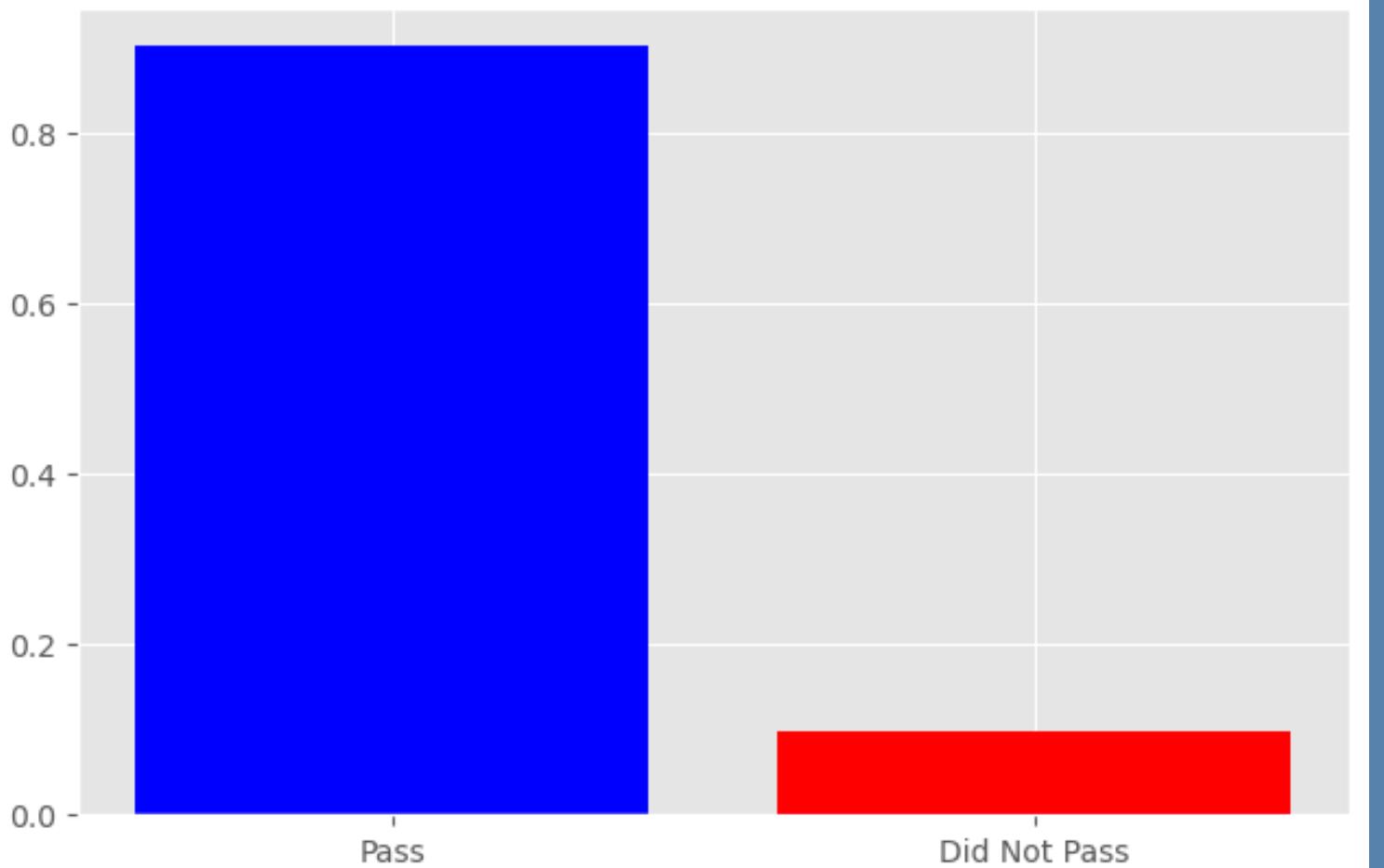


**Hispanic**  
Pass Rate:  
86.5%  
Ranking of  
A-B:  
C, B, A & D

Grade Distribution  
for White Race Students



Credit Aquisition Percent  
for White Race Students



**White**  
Pass Rate:  
90.1%  
Ranking of  
A-B:  
A, B, D, C

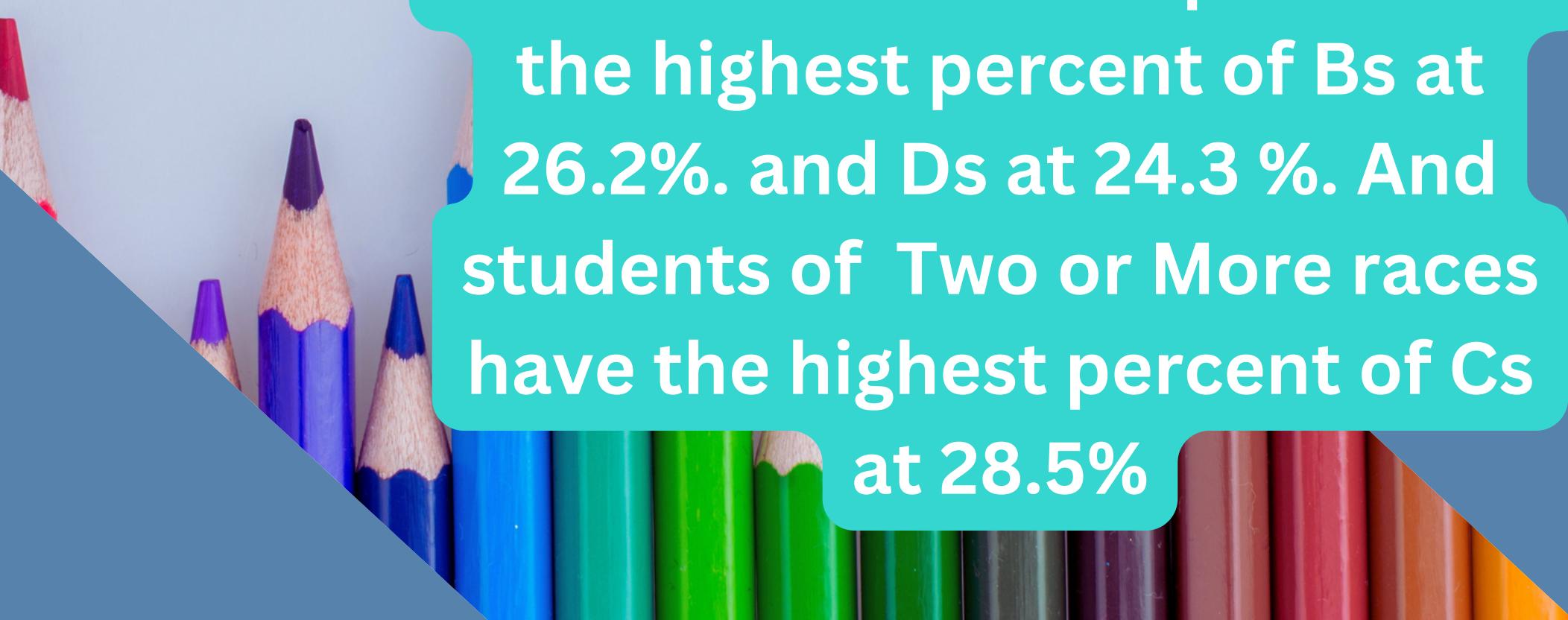
# Racial Grade Distribution

## Pass/No Pass

Based on pass rate and grade distribution, students who are Black pass the highest percentage of classes, then White, then Hispanic, then students of Two or More races.

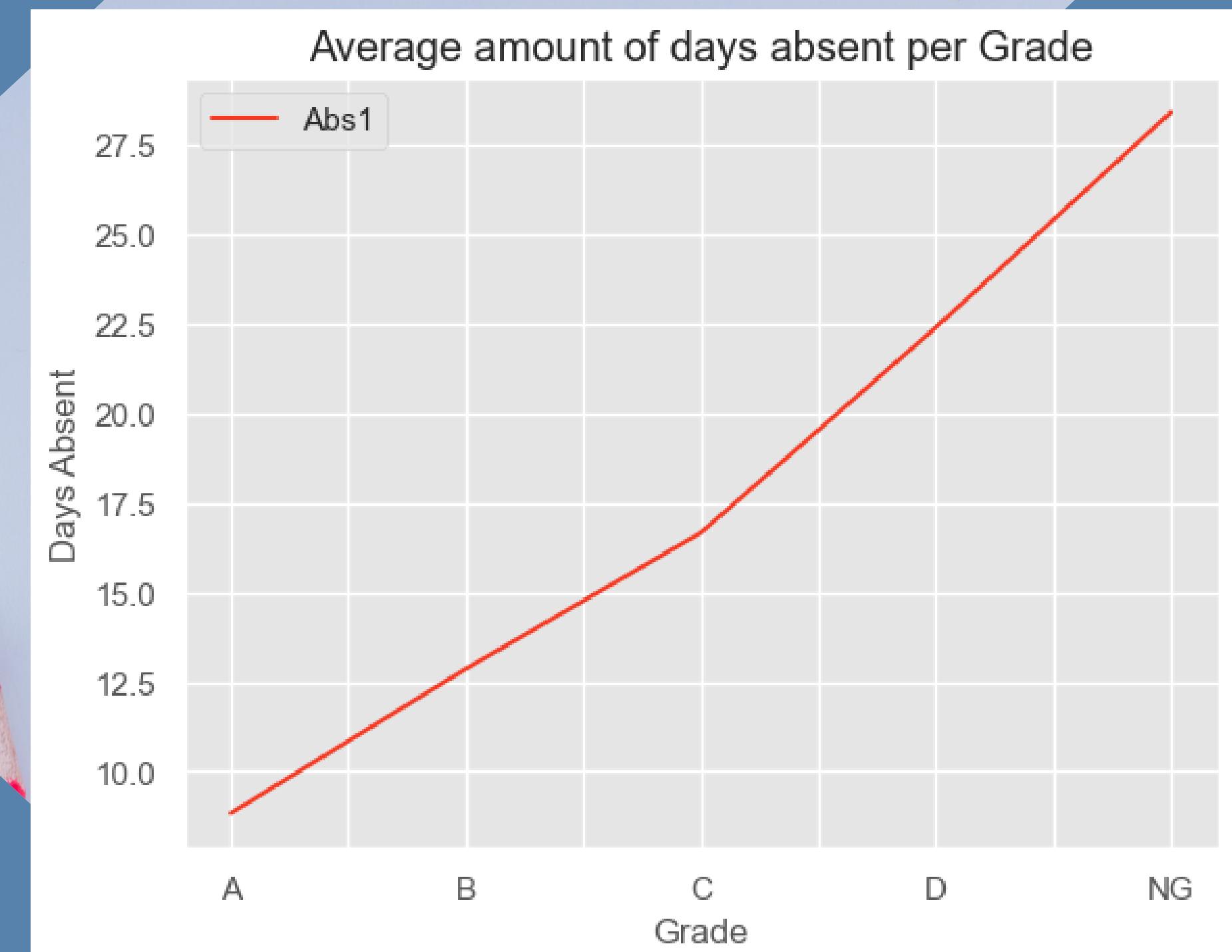
## Grades

Students who are White have the highest percent of any grade, with 33.1% getting As. Students who are Hispanic have the highest percent of Bs at 26.2%. and Ds at 24.3 %. And students of Two or More races have the highest percent of Cs at 28.5%



# Attendance

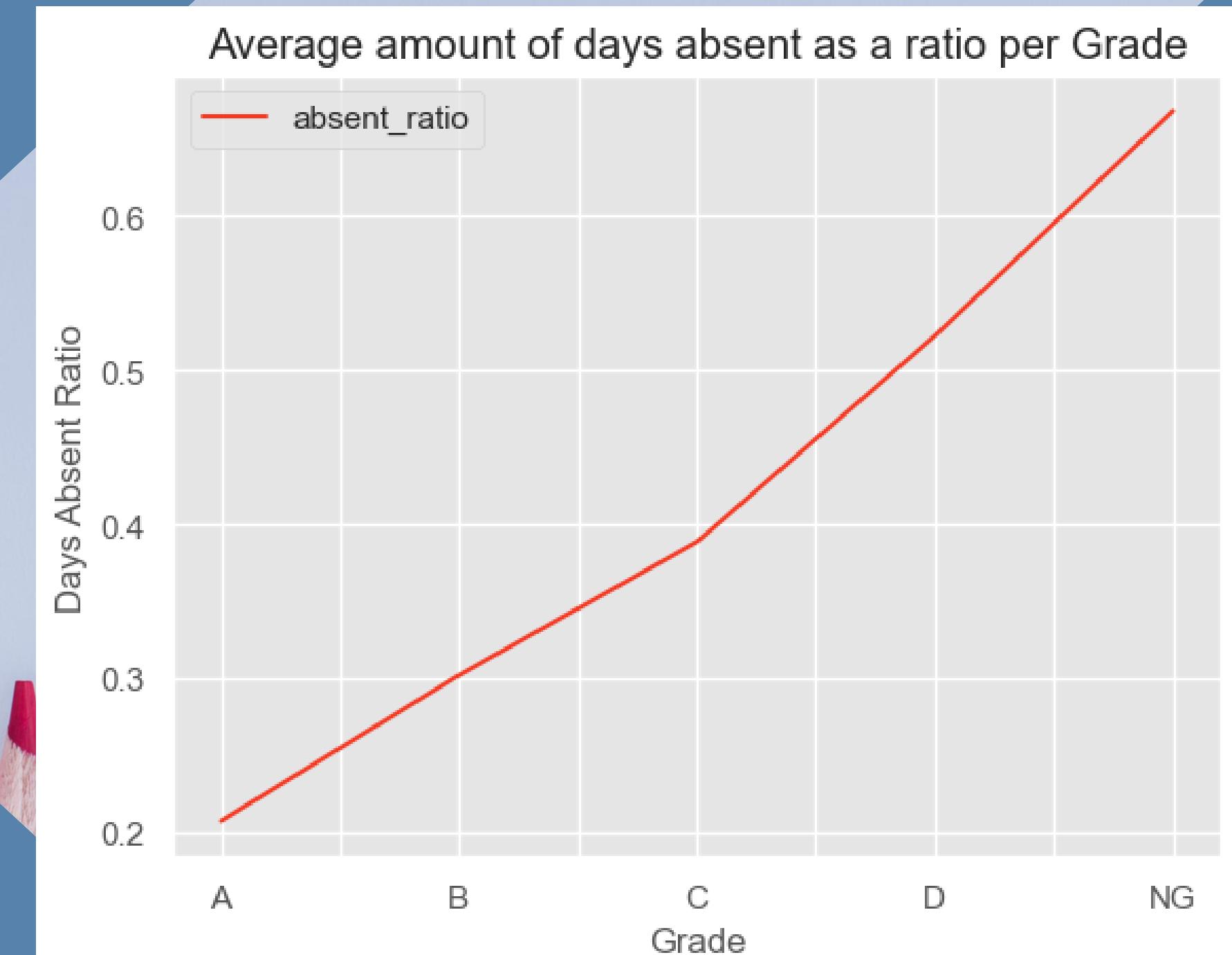
Maybe the most important feature for our school is attendance. Many of our students come to Mt Scott because they've been pushed out of their home school due to attendance issues. Here is a graph of the average days absent per grade with a breakdown below.



A: 9, B: 13, C: 17, D: 22, NG: 28 days

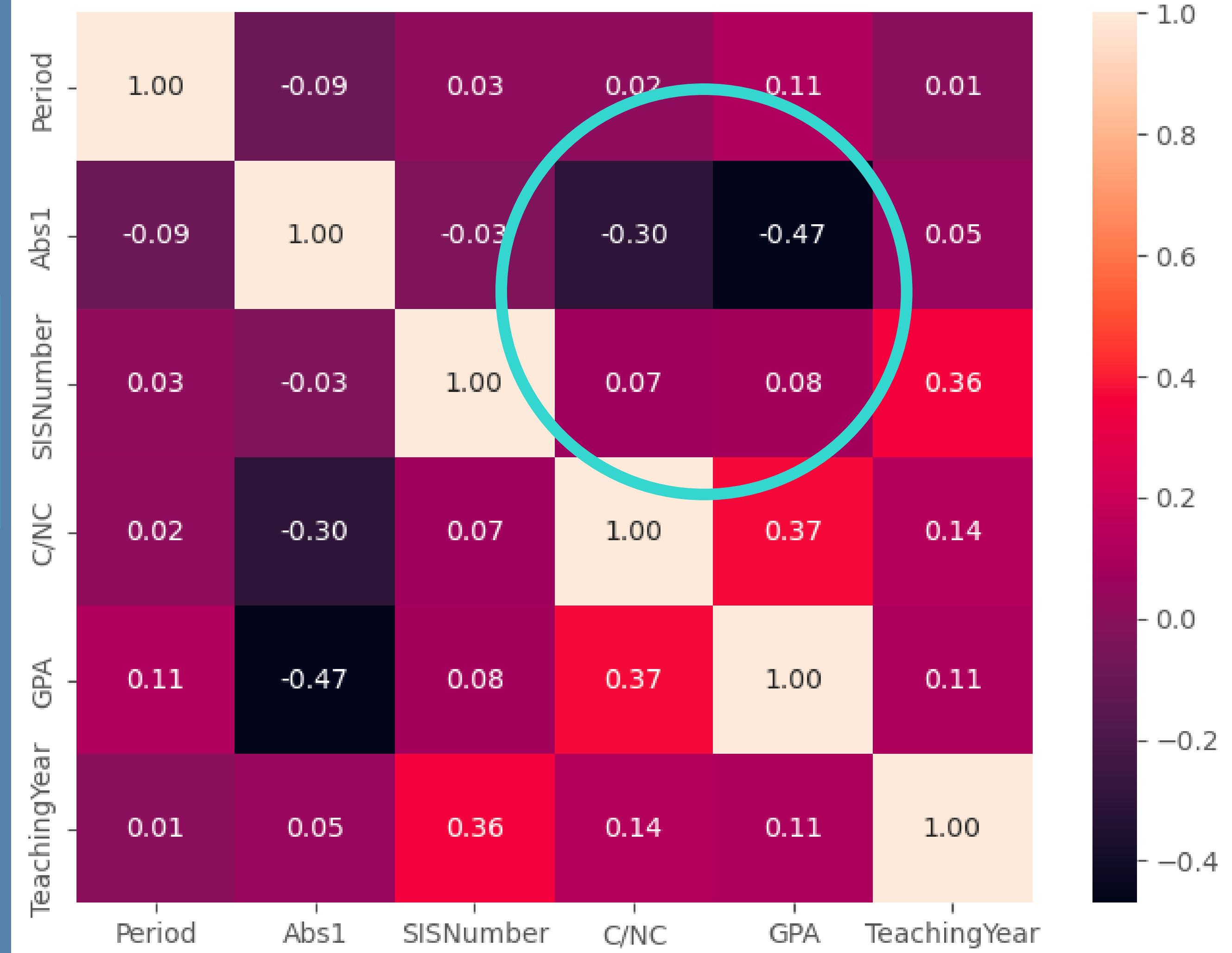
# Attendance

Here is a graph of the average ratio of days absent per grade with a breakdown below.



A: 20.8%, B: 30.2%, C: 38.9%, D: 52.3%, NG: 66.9%

Looking at a correlation heat map, the number of days absent has a -0.3 correlation to gaining credit, and a -0.47 correlation to GPA. The biggest correlation for early features

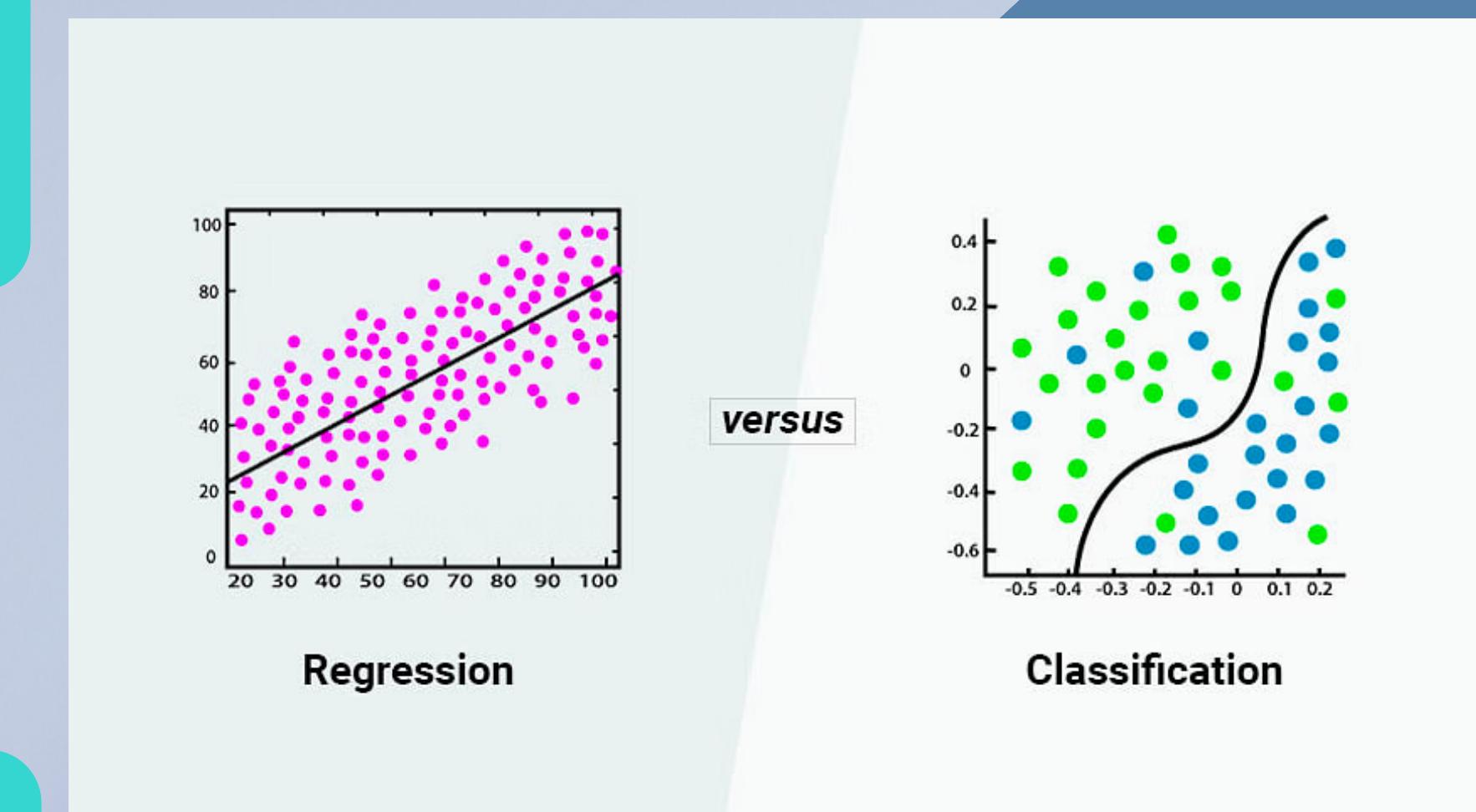


Looking at the Correlation Matrix for Race and Days Absent, both Hispanic and Black races have a positive correlation to days absent meaning they are more likely to be absent than White or Two or More races.



# Modeling

I wanted to try two different types of machine learning, both supervised. The first was a round of binary classification models for Pass/Fail features, and the second, a regression on Grade Percentages.



# Classification Modeling

I started with a SciKit Learn's logistic regression and got these accuracy numbers

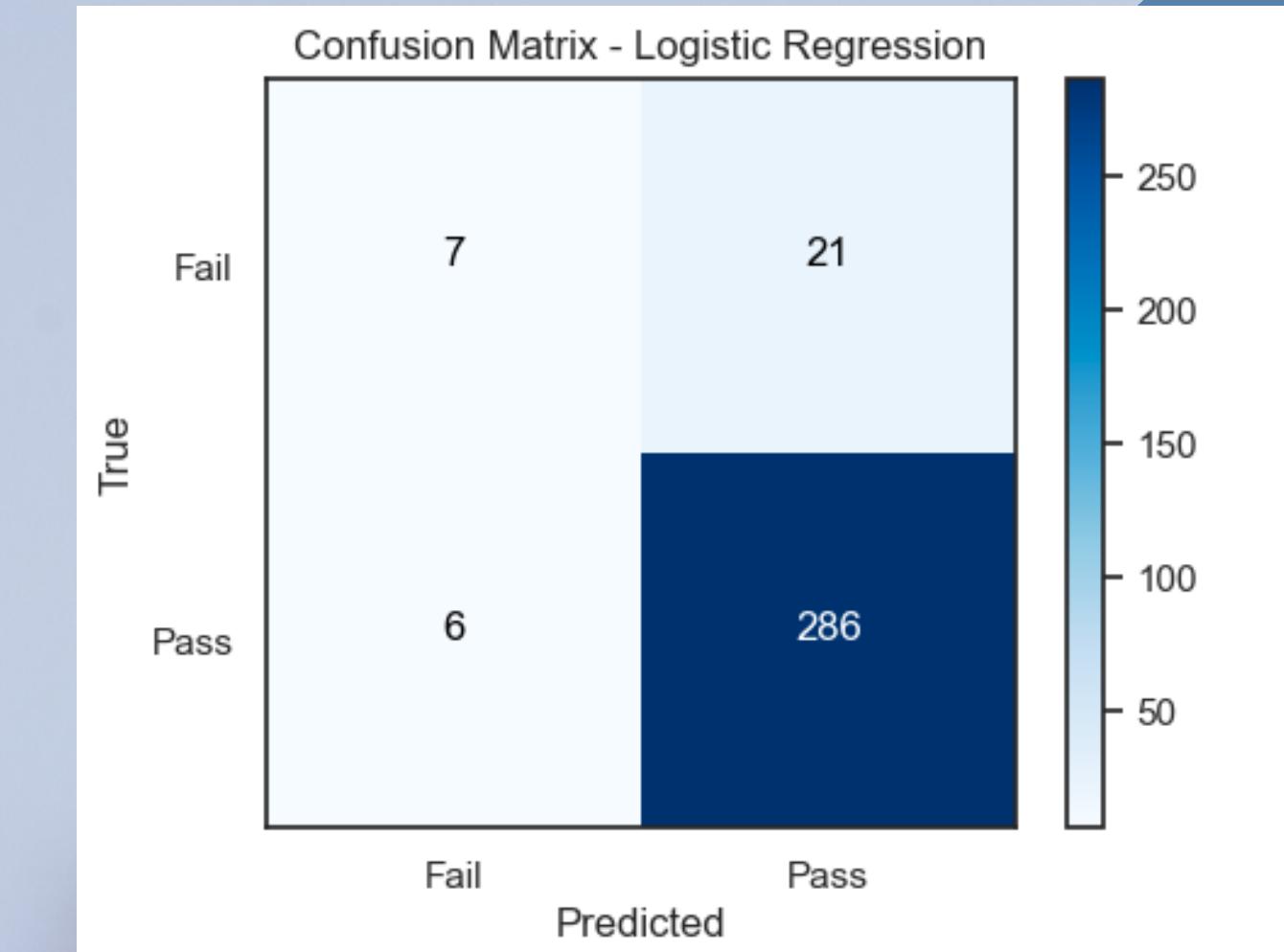
Accurary: 92.5%

Pass Precision: 94%

Fail Precsion: 67%

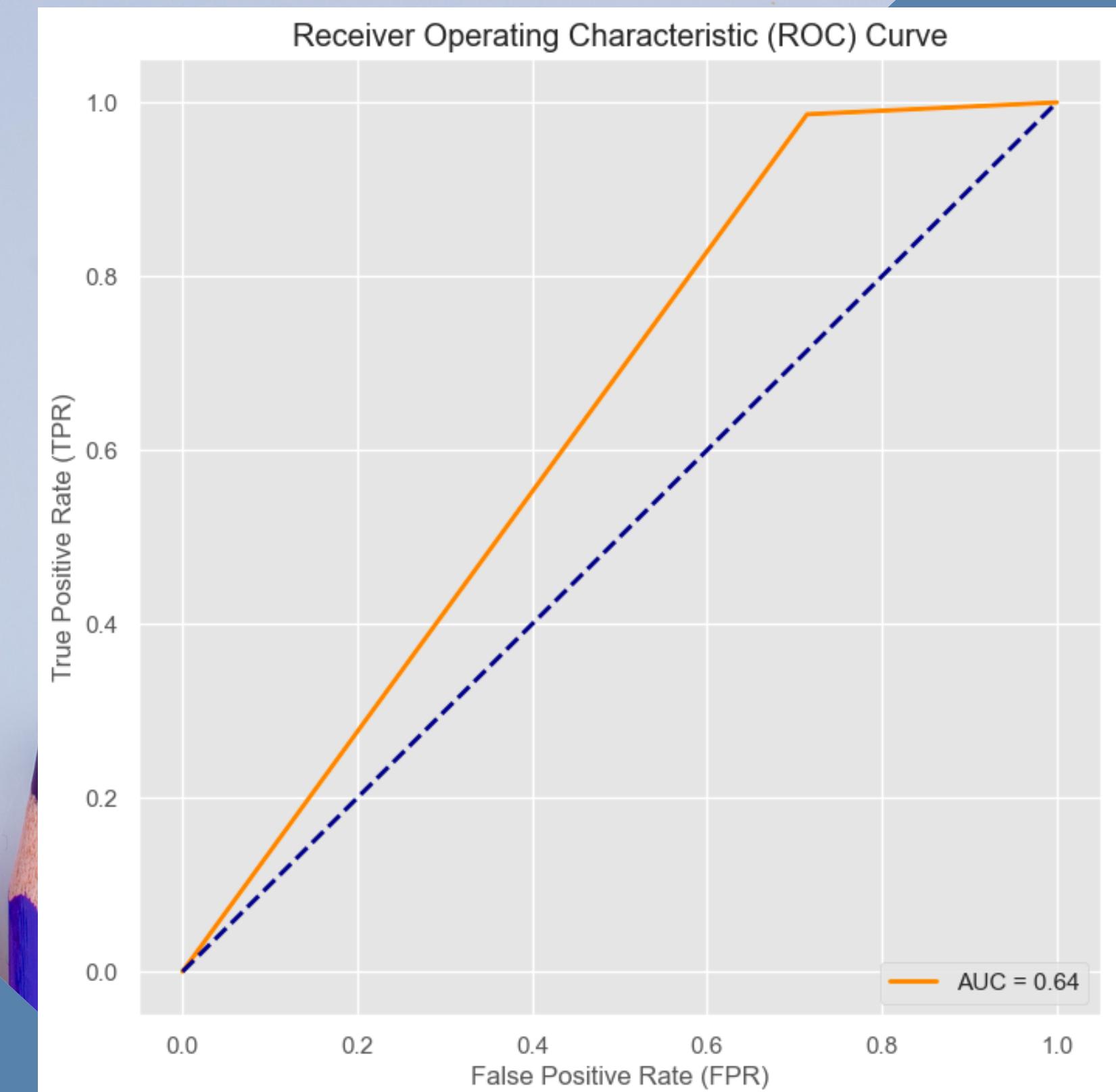
Pass Recall: 99%

Fail Precision: 29%



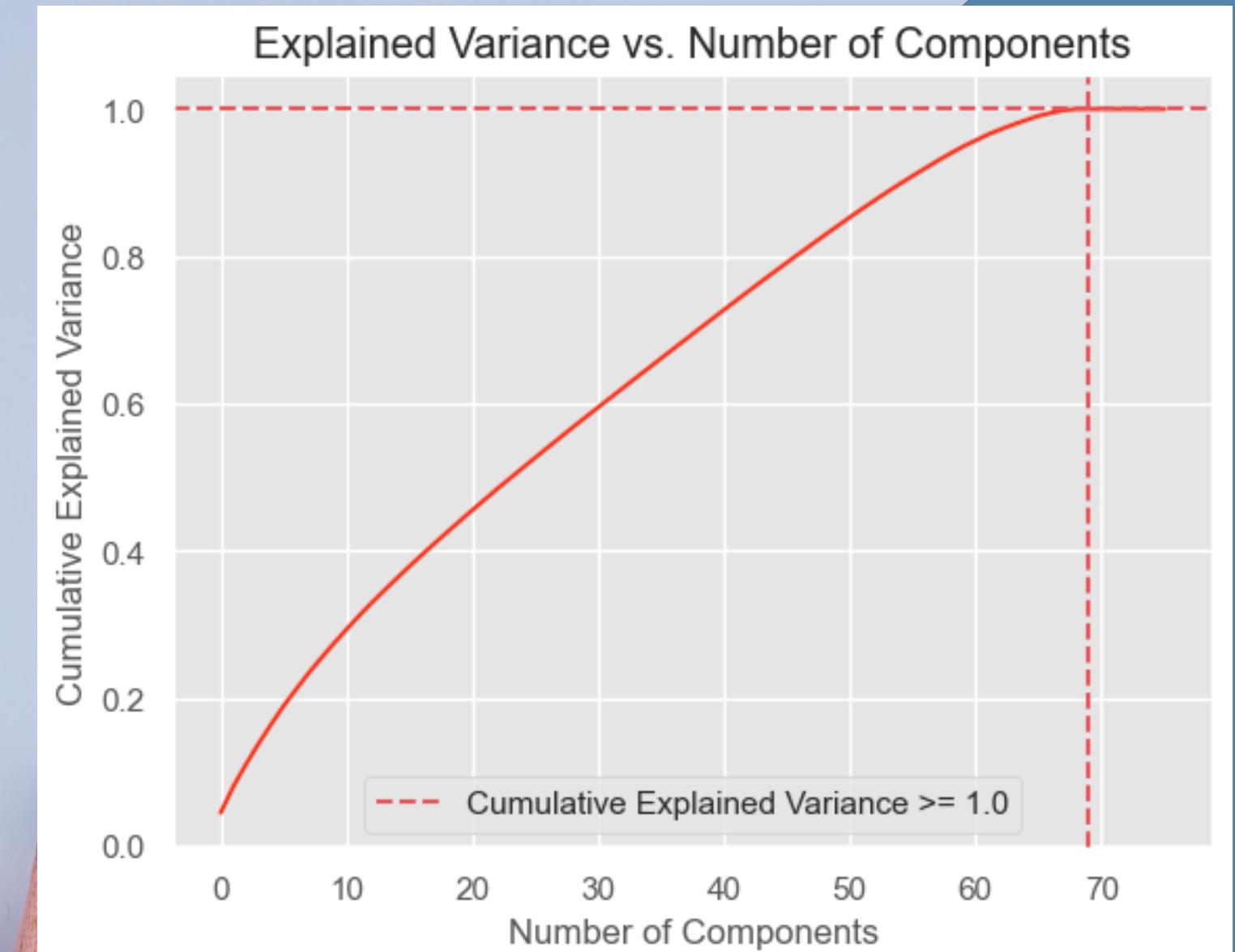
# Classification Modeling

I checked the AUC ROC and it wasn't too amazing with an AUC of 0.64, only 14% better than a random model.



# Classification Modeling

I also did a PCA to see if I could reduce factors, but it turned out that 69 of the features got to 100% of the variance.



# Classification Modeling

Here is a graph of the feature importances for the Logistic Regression

Regression

Zip 97233: -2.03

2020-2021 School Year: + 1.22

Zip 97211: + 1.04

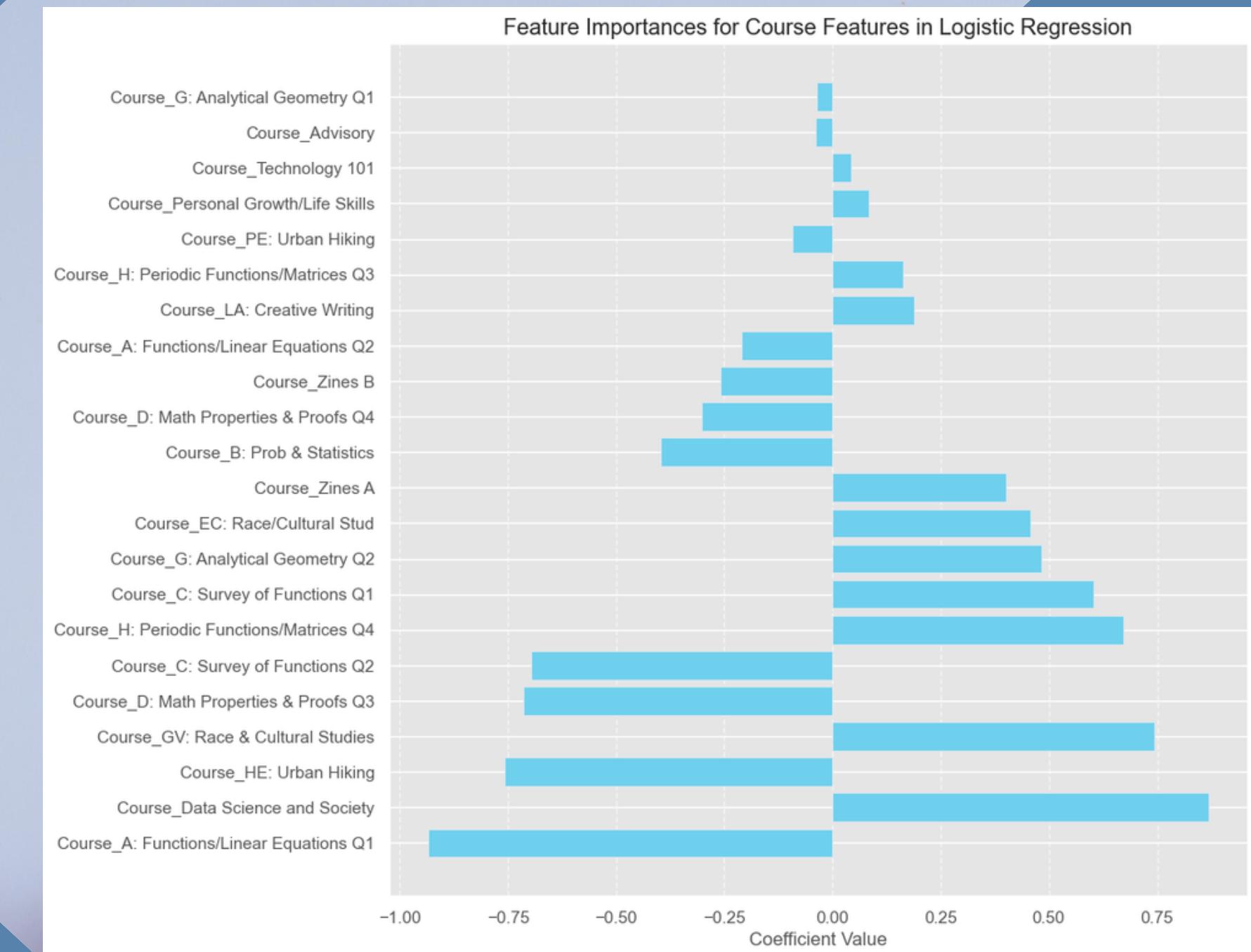
Zip 97220: +1.03

Math A: Functions and Linear Equations Q1: -0.93



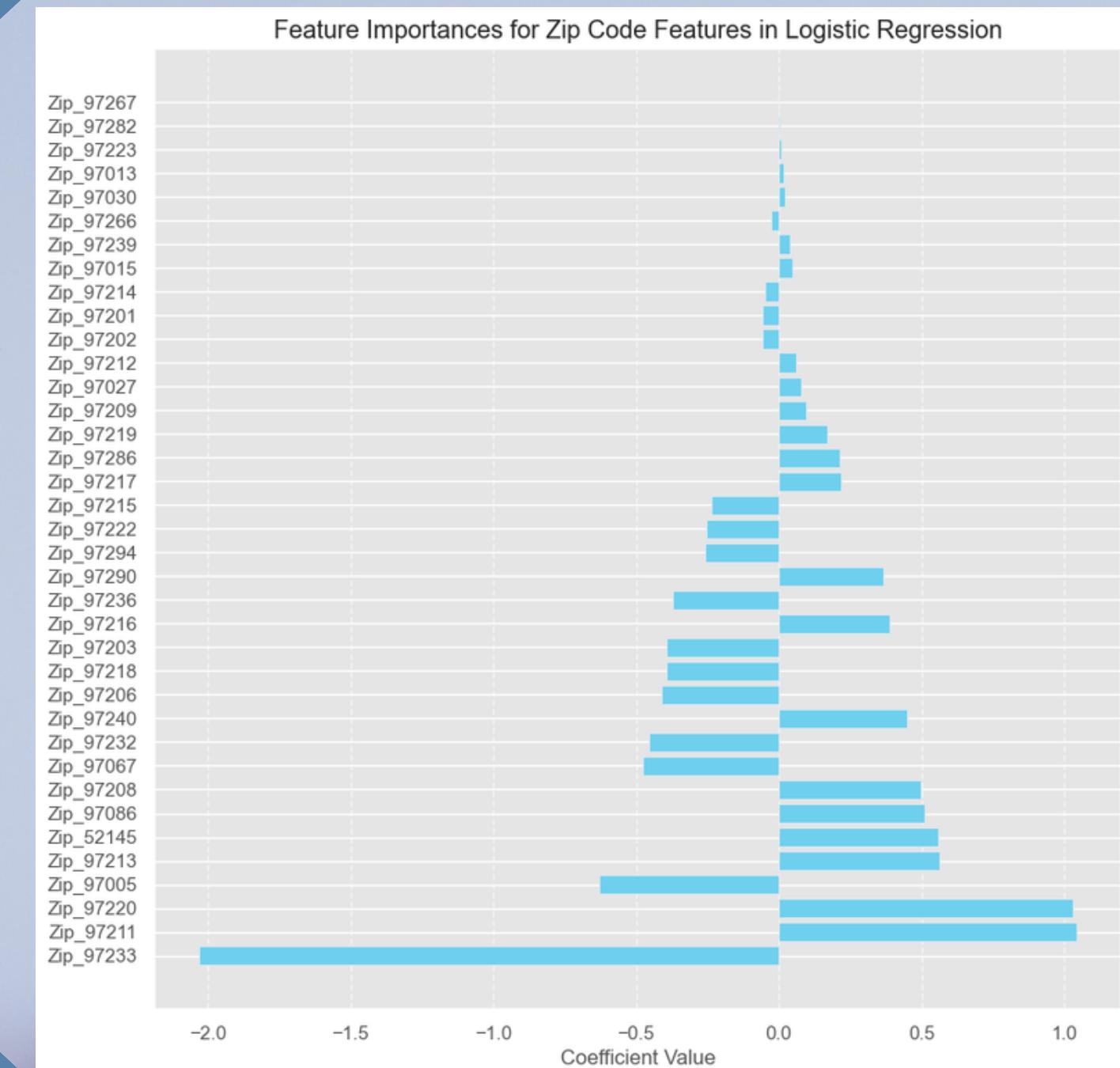
# Classification Modeling

Here is a graph of  
the feature  
importances for  
Courses



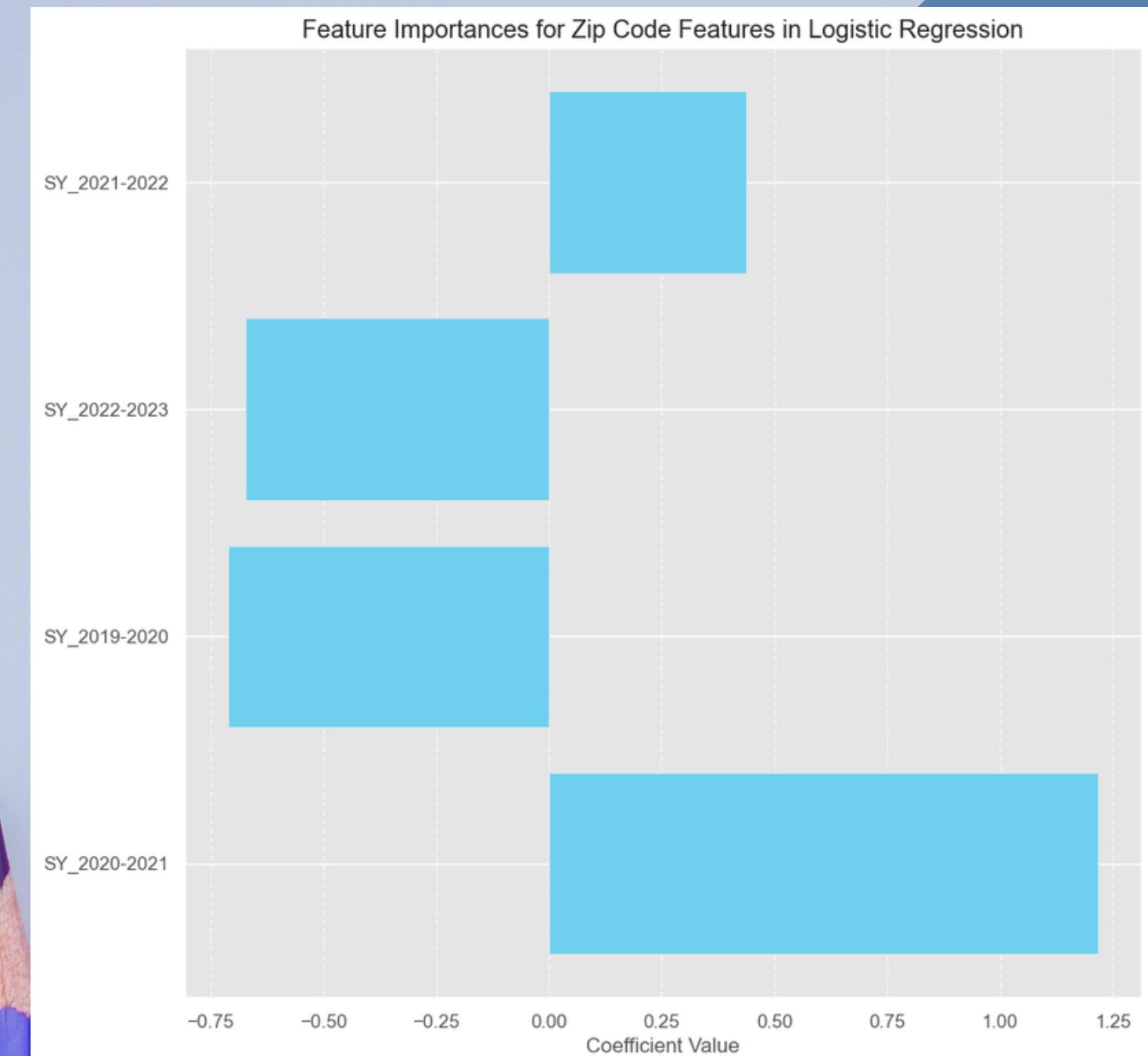
# Classification Modeling

Here is a graph of  
the feature  
importances for Zip  
Codes

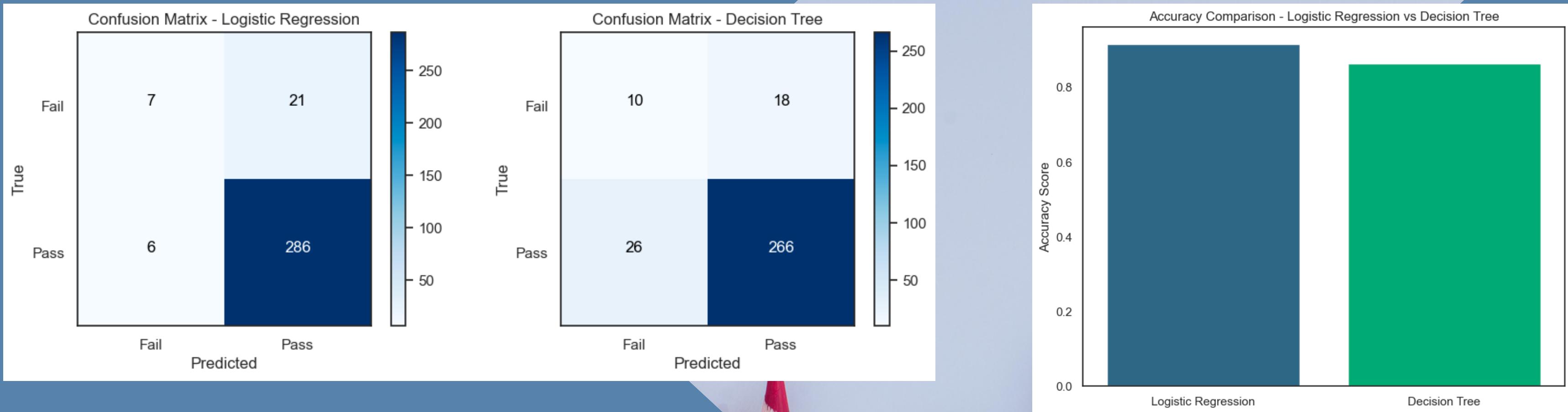


# Classification Modeling

Here is a graph of the feature importances for School Year for the Logistic Regression



# Classification Modeling



Then I decided to try a Decision tree but it performed worse than the logistic regression.

# Classification Modeling

I also tried a Grid Search on both a Random Forest and XGBoost Classifier models and here were my results.

## Random Forest

Accuracy: 91.6%

Pass Precision: 93%

Fail Precision: 55%

Pass Recall: 98%

Fail Precision: 21%

## XGBoost

Accuracy: 91.9%

Pass Precision: 93%

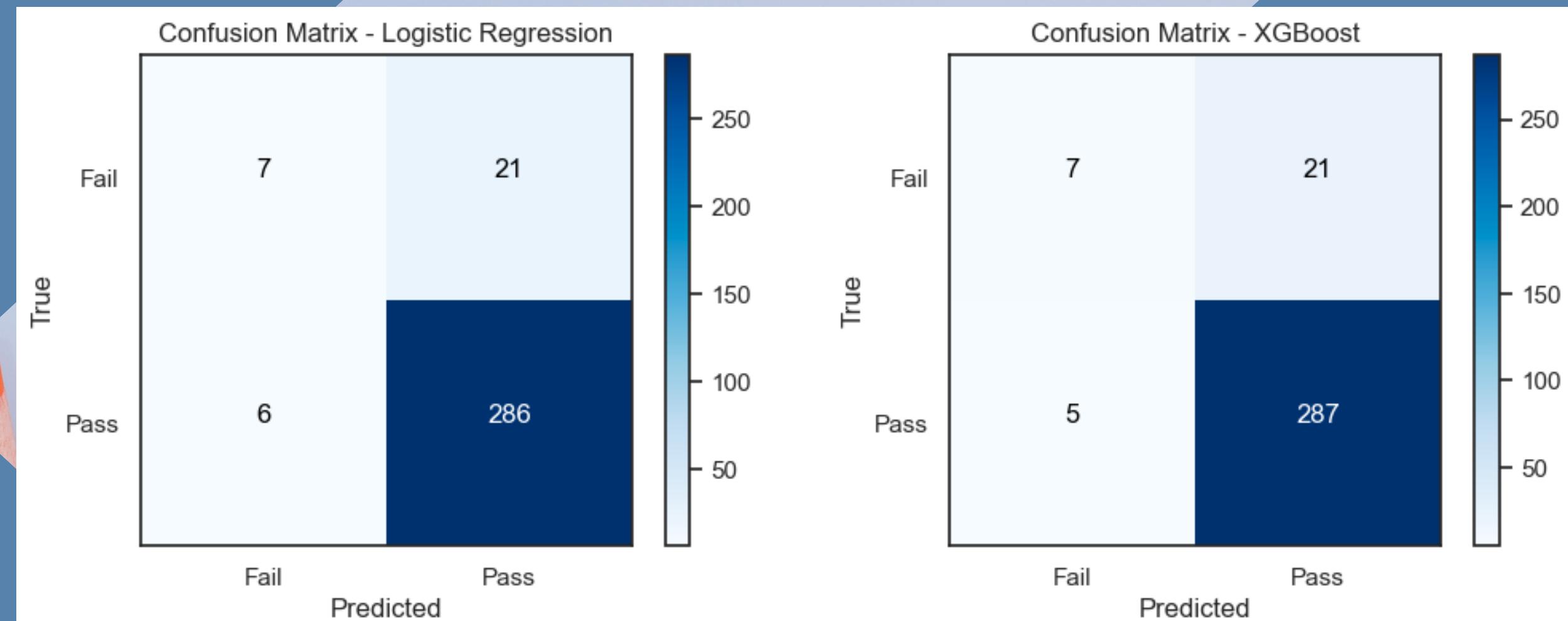
Fail Precision: 58%

Pass Recall: 96%

Fail Precision: 25%

# Classification Modeling

Comparing the confusion matrices of the LR and the XGBoost model, the XGB gets one more true Pass!



# Classification Modeling

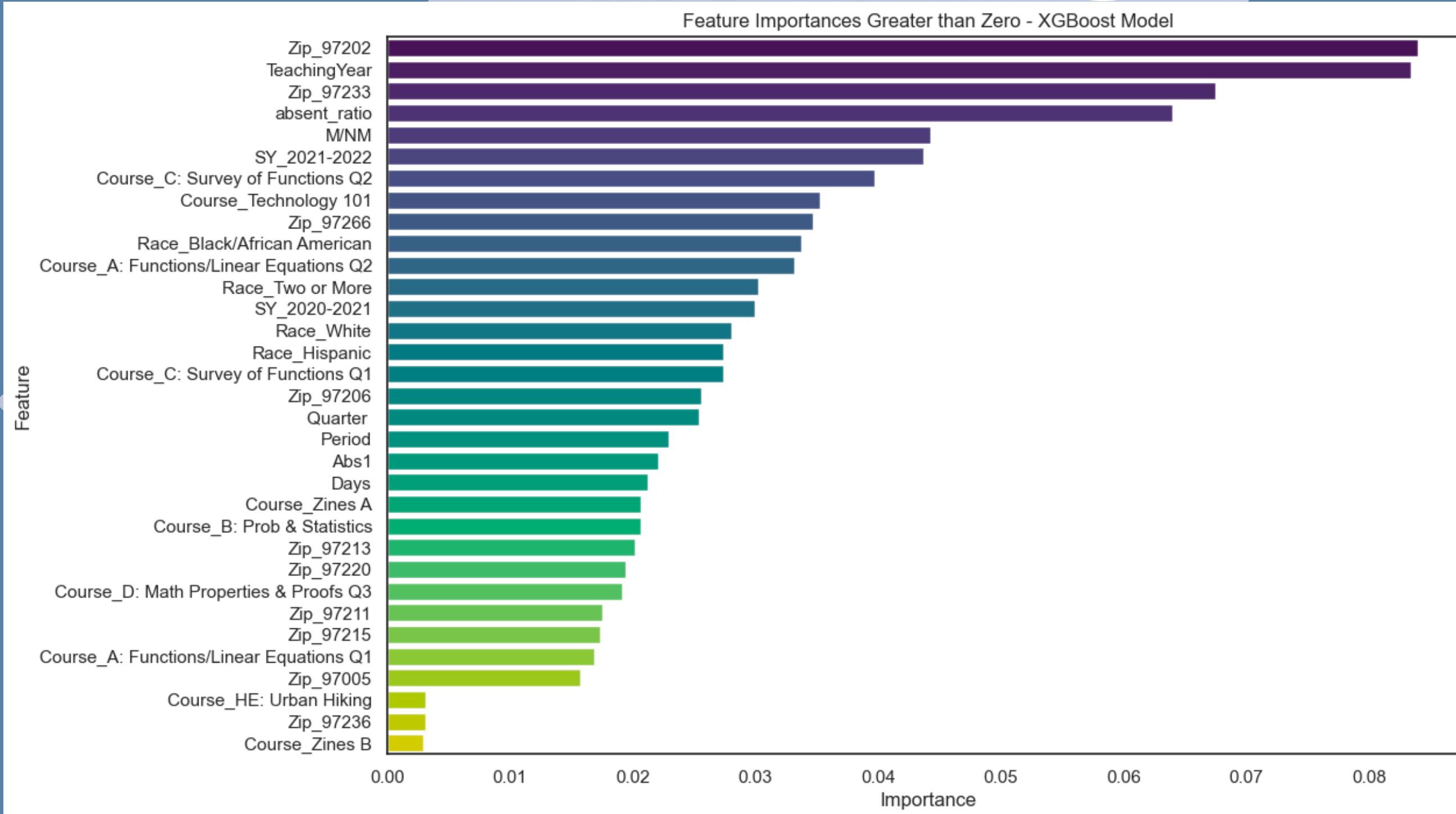
Here are the  
feature  
importances  
on the  
XGBoost.

97202: 0.08

Teaching Year: 0.08

97233: 0.07

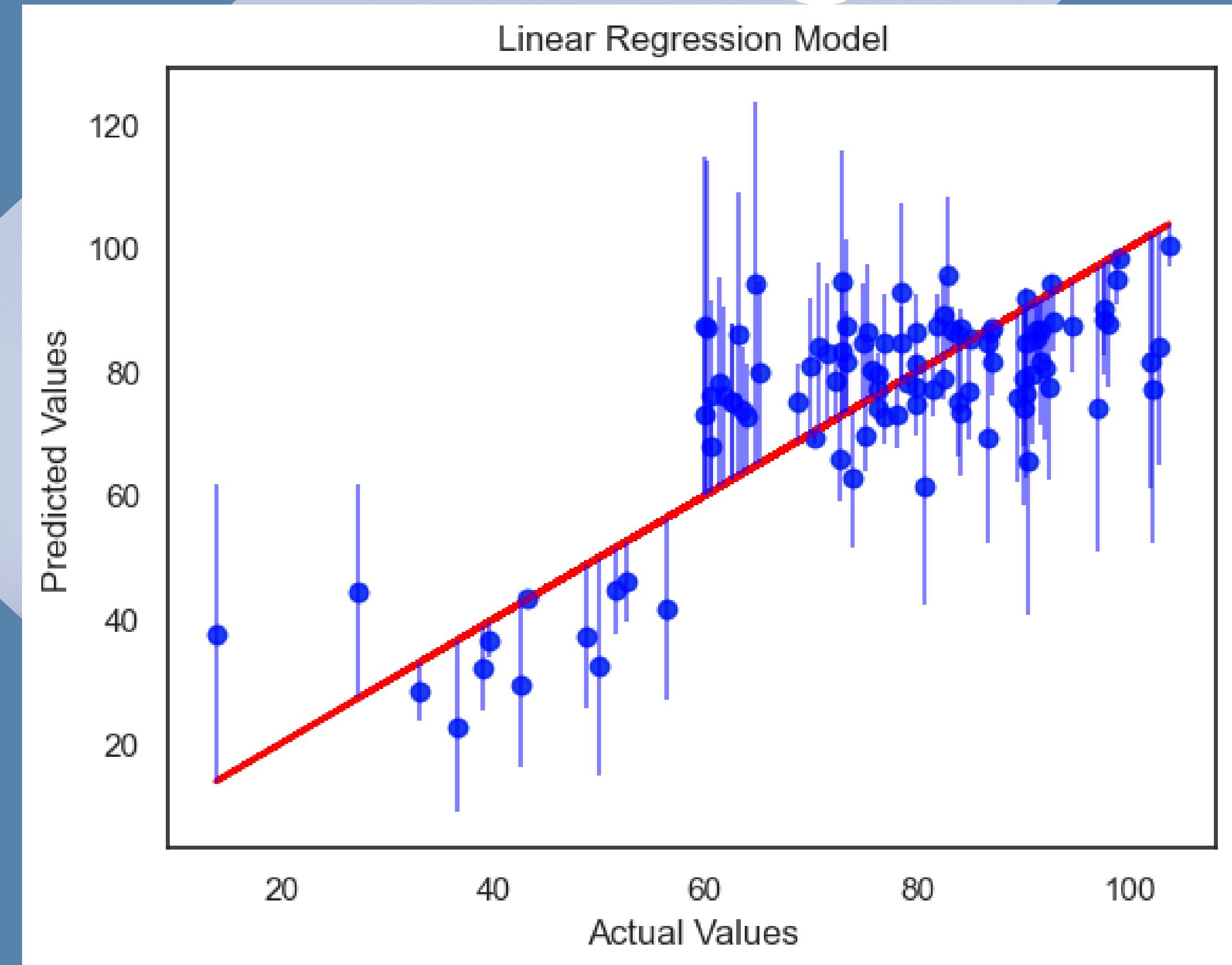
Absent Ratio: 0.06



# Regression Modeling

I only made one regression model, SciKit Learn's Linear Regressor and it performed adequately.

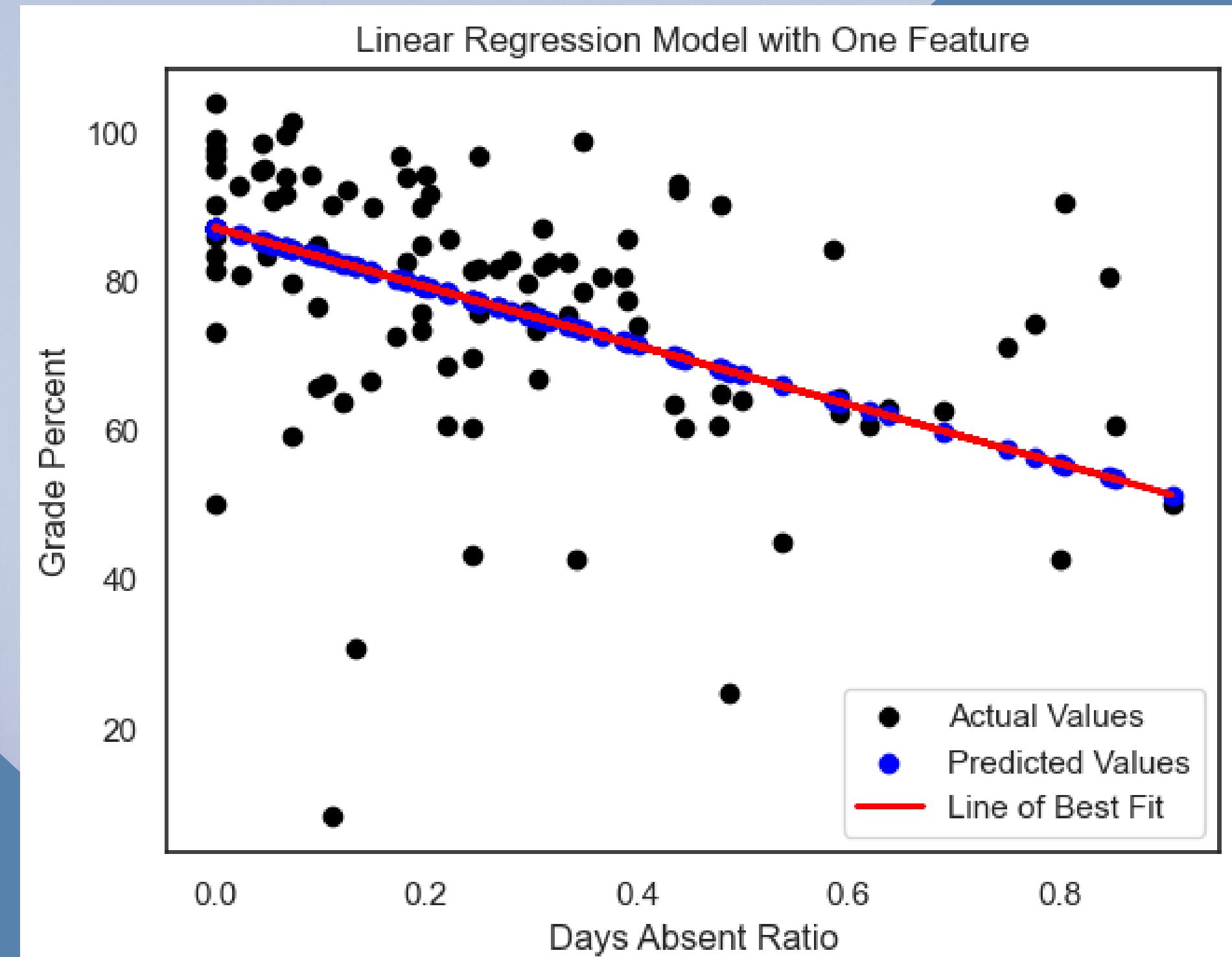
Mean Squared Error: 142.69384484942597  
R-squared: 0.544168796999436



# Regression Modeling

I modified the model to only account for Days Absent Ratio and also, the results were subpar.

Mean Squared Error: 268.07101714100884  
R-squared: 0.10789677243773821



# Future

There is one big thing I think would help improve the modeling: more data. If I had all the teachers' gradebooks, then the models would have more to work with. Same goes for more of my own grade data for all of my non-math classes.

i	i	i	Submission 1	i	i	i	i	i	Workshop 1						
Free Write: 1 Min MAX:8.00 PTS:1.00 11/9/2023	Free Write: 2 Mins MAX:8.00 PTS:1.00 11/17/2023	Free Write: 3 Mins MAX:20.00 PTS:10.00 11/21/2023	Free Write: 4 Mins MAX:1.00 PTS:1.00 11/21/2023	Free Write: 5 Mins MAX:5.00 PTS:1.00 12/1/2023	Free Write: 6 Mins MAX:5.00 PTS:1.00 12/8/2023	Free Writes: 6 Mins MAX:4.00 PTS:1.00 12/14/2023	Classwork (DO NO S)								
8	2	10	0	0	0	0									
6	4	10	1	2											
4	7	20	1	2.5											
5	6	20	1	1.5											
6	4	10	1	2.5											
8	8	20	0.5	2.5											
A	8	20	1	5											
8	4	20	1	3											
4	4	10	0.5	2											
5	6	10	1	3											
5	3	20	1	1.5											
8	6	20	0	5											
8	5	20	0	3											
Free Write: 1 Min		Free Write: 2 Mins		Submission 1		Free Write: 3 Mins		Free Write: 4 Mins		Free Write: 5 Mins		Free Writes: 6 Mins		Workshop 1	
Avg:79.8%		Avg:64.4%		Avg:80.8%		Avg:69.2%		Avg:51.5%							

# Future

But for now, the initial data exploration around the average days absent is a graspable unit that can be used in conversations with students and their families.

FOR EVERY

**4 days  
absent**

**YOUR GRADE GOES DOWN  
A WHOLE LETTER**

# Future

I'd also like to do some feature exploration of the zipcodes and see if there were any correlations between zipcodes and racial demographics and/or the distance between the zipcode and the school's zipcode.

Finally, to test my models, I'd like to make a dataset of my last quarter's students' features and targets.

Unfortunately, I was not able to get the attendance data in time for this round.