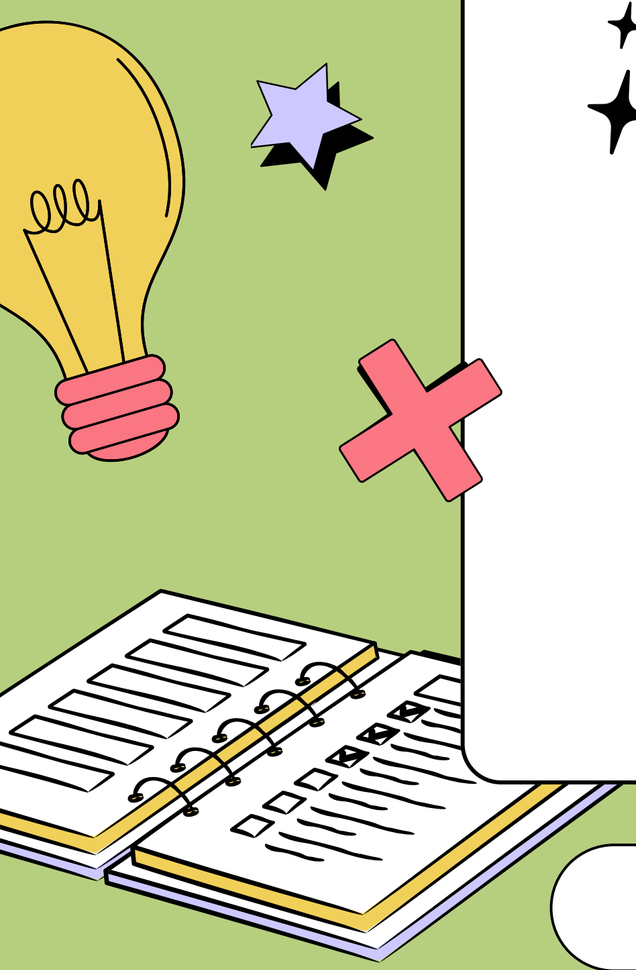
DATA 4382 - CAPSTONE PROJECT 2

Boosted Reteach Planning:

Optimizing Student Learning Outcomes



Zehra Erden

Contents:

- [ ] Background
- [ ] Data Understanding
- [ ] Data Preprocessing
- [ ] Data Transformation
- [ ] Model Selection
- [ ] Model Evaluation
- [ ] Conclusion

Background

No Child Left Behind (NCLB) 2001

The federal government of the United States authorized the No Child Left Behind Act (NCLB) in 2001 to ensure that schools are held accountable for every student's performance and to provide more opportunities for students in need. Until 2015, local states were required to conduct yearly assessments to demonstrate their students' improvement.

Every Student Succeeds Act (ESSA) 2015

In 2015, the federal government signed Every Student Succeeds Act (ESSA), replacing the NCLB. ESSA grants more flexibility to states and, once again, requires every state to assess the performance of their students in reading, math, and science.

STAAR

STAAR tests cover the yearly curriculum to measure students' performance and readiness for the next academic year's curriculum. Every year, students from the 3rd to the 12th grade take STAAR tests in core subjects, including Reading and Language Arts (RLA), math, science, and social studies.

Data Understanding

Source:

Harmony Public Schools Administration & Teachers

Type:

Reteach Plans - pdf
Student Academic Performance - csv

4
8th Grade
Math
Teachers

6 reteach
plans

249
students

Data Understanding

Reteach Plans

> each teacher created one per class/course that they teach

> reteach period:
begins after fall interim assessment
ends before spring interim assessment

> 5 reteach plans from 2021-22
1 reteach plan from 2022-2023

TEKS & Percent mastery % (Can be copied from TAMS report when you hover over the standard	**Identifying Error/Misconception:**	**Plan for Re-Teach Instruction:**	**When will you reteach**	**What reassessment/ questions will you use**	**Reassessment Result after reteach**	**Reflection**
TEKS: 8.2.D The students are expected to order a set of real numbers arising from mathematical and real-world contexts. **Percent Mastery:** 42% ***Assessed on CCBA-1?(Yes/No)***	*What misunderstandings are revealed in the data? Why do you think students failed to reach mastery? What gaps in the instruction of the standard contributed to these misunderstandings?* *Item 19 - The students may forget that the order of*	*What adjustments in instructional strategies and teaching will you implement to improve students' mastery?* ** I will have my students review the **IA** questions they missed by:* *Reteach* ** I will model the following question with students:* *- STAAR Question*	*Choose a date considering your lab times and other available times in the next six weeks* ***(complete re-teach by February 18th)*** 1/10 - 1/14	*Choose similar questions from previously released questions or test prep materials, use Schoology or Quizizz to get immediate results.* *Reteach* *- STAAR Question Bank (#8)* *- STAAR Question Bank (#12)*	*Jot down your class average after the reteach and compare it with the original percentage (17% for this sample Teks/question)* *(To be updated after reteach)*	*Celebrate if the reassessment is higher or reflect and plan further clean up if the desired improvement is not observed.* *(To be updated after reteach)*

Data Understanding

Student Academic Performance Data

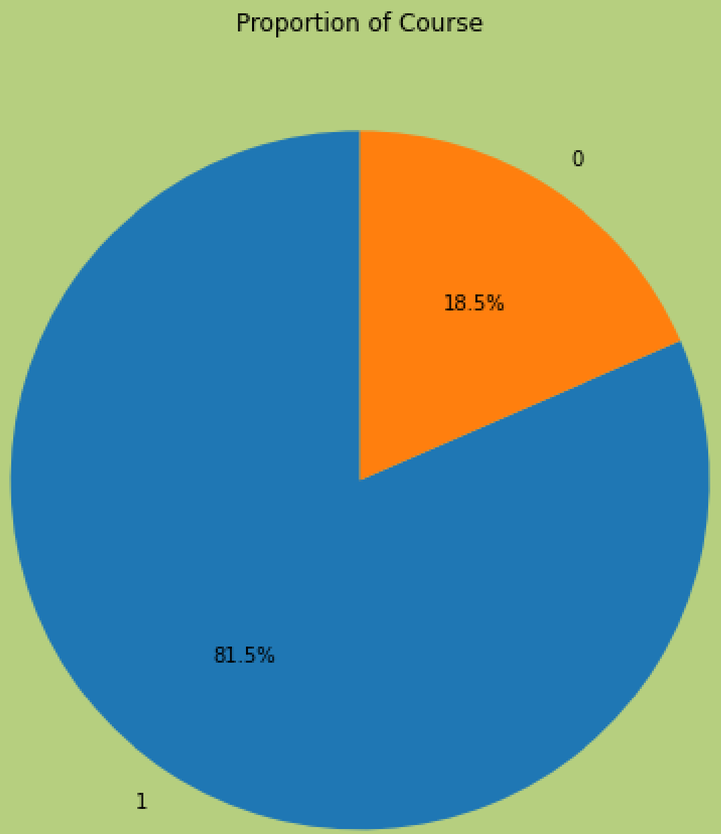
> Teacher
> Student_ID
> Course | Math-8 and Algebra I
> Percent Score (FA & SP)
> Scale Score (FA & SP)
> Mastery Projection (FA & SP)
> Approach Probability (FA & SP)
> Meet Probability (FA & SP)
> Master Probability (FA & SP)
> Projected Tier (FA & SP)

Data Understanding

Course Distribution

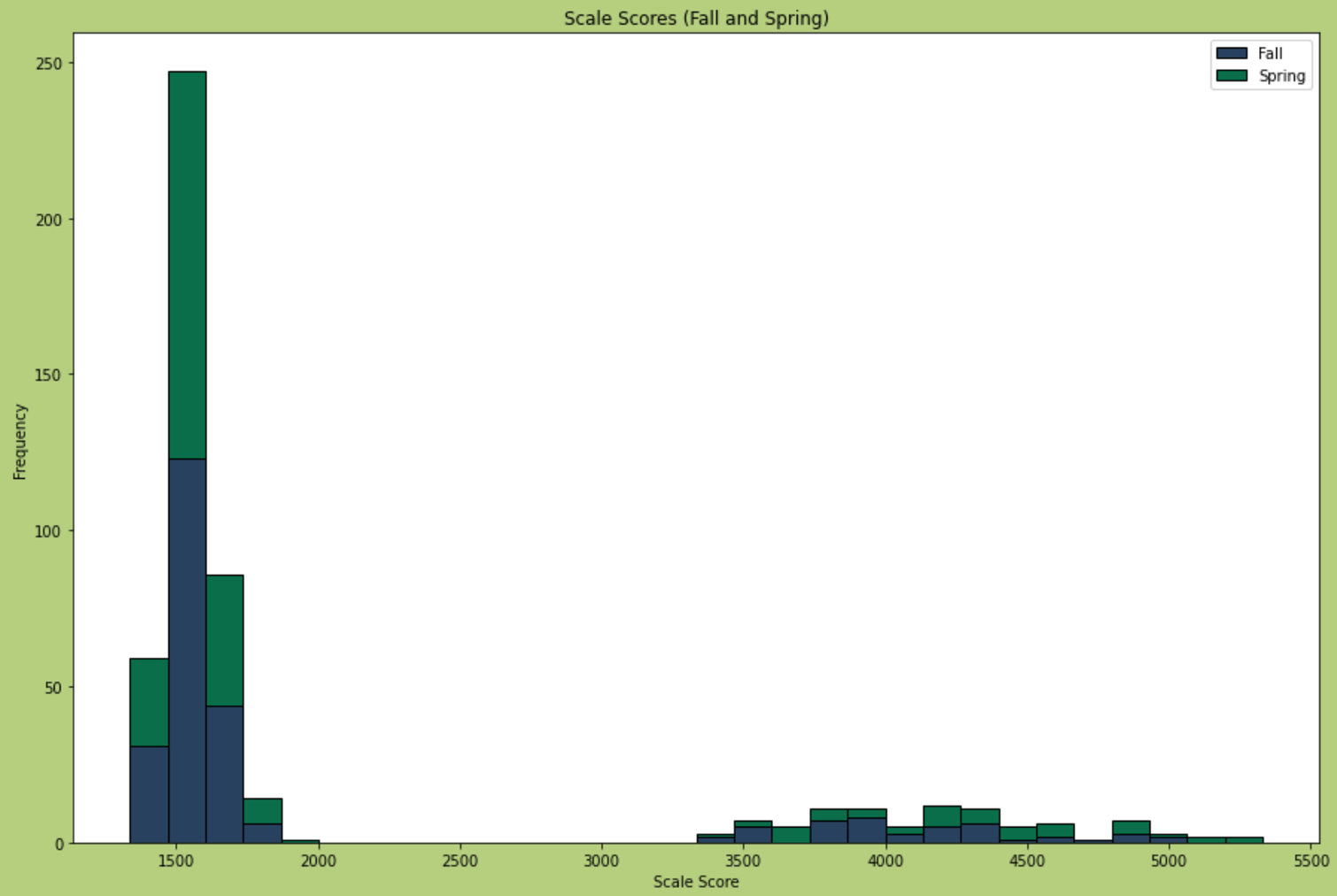
> '1' = 8th grade Math

> '0' = Algebra



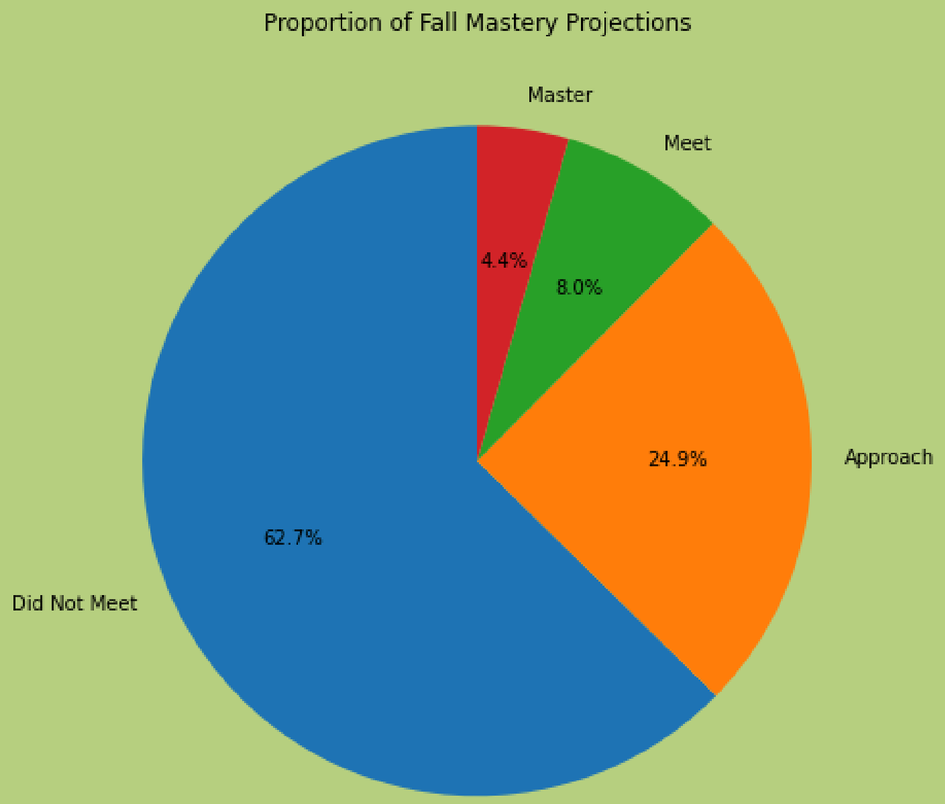
Data Understanding

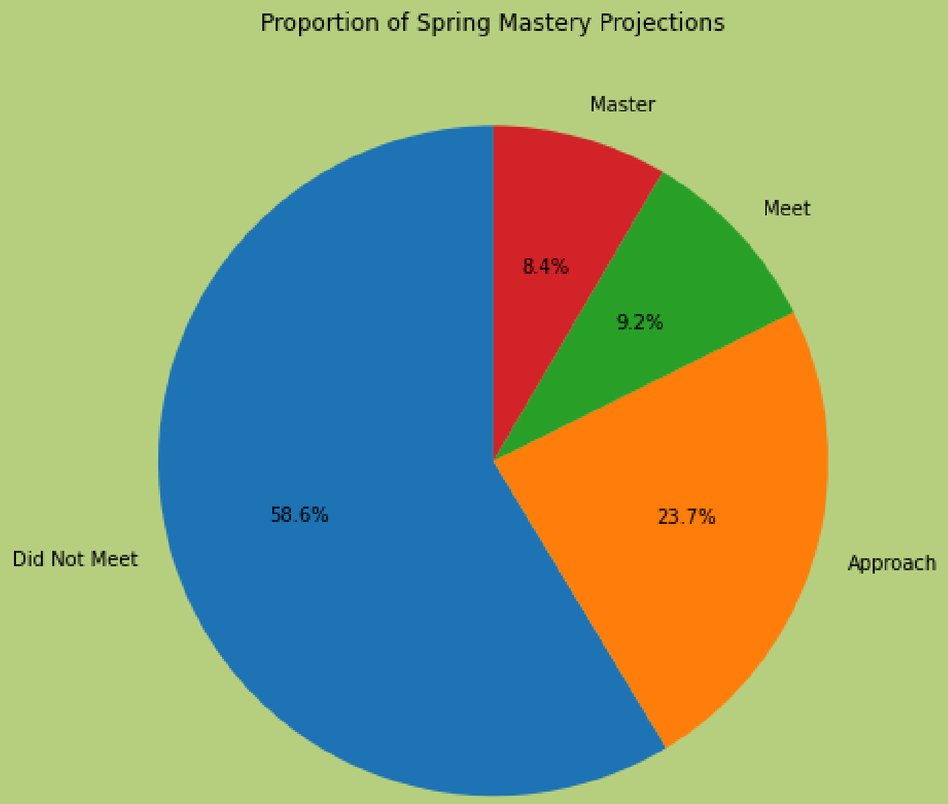
Scale Scores (Fall vs Spring)



Data Understanding

Mastery Projections (Fall vs Spring)





Data Understanding

Percent Scores and THE problem

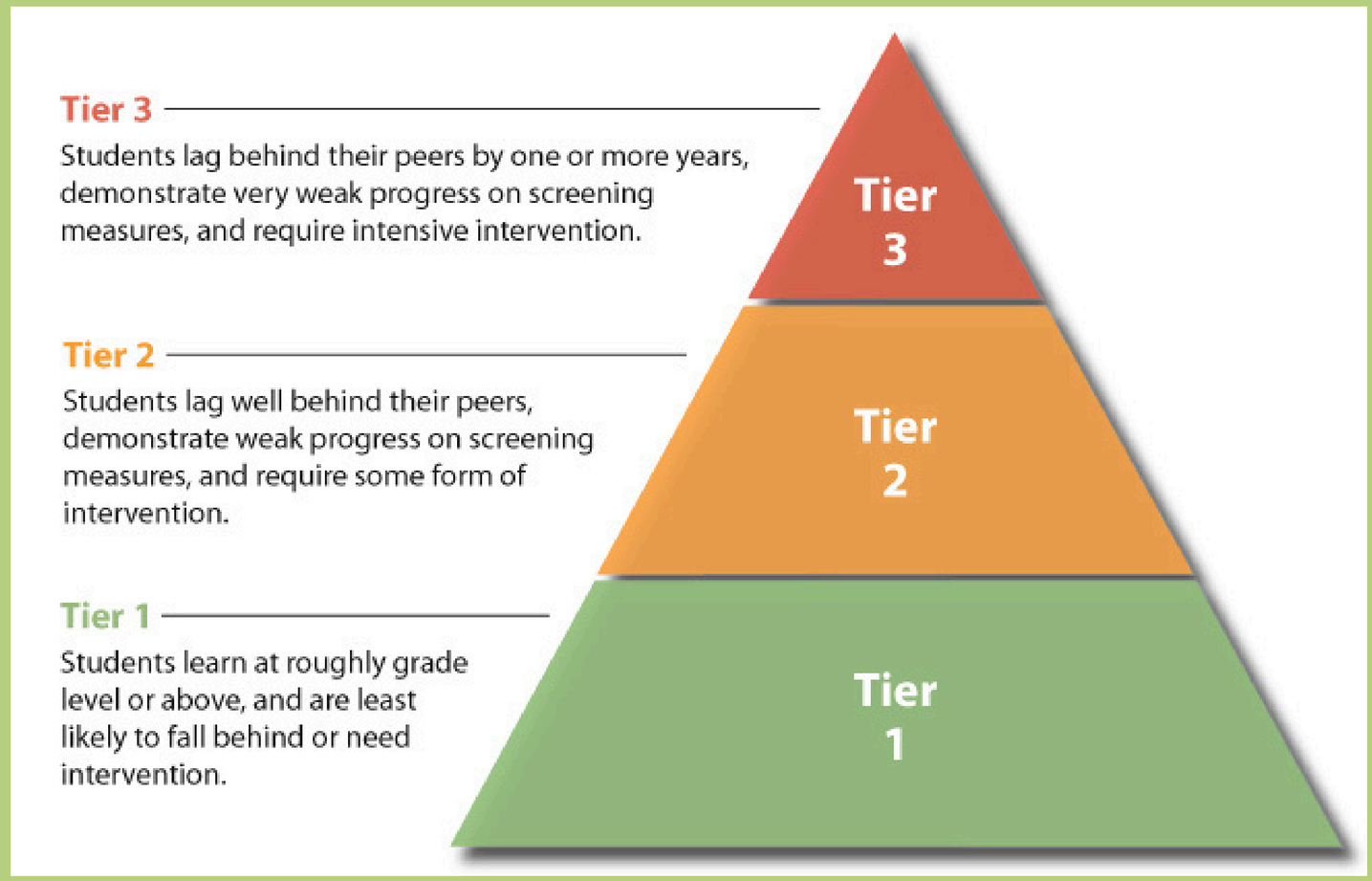
A percent score indicates the rank of the student compared to others in same grade and campus.

For example, a percent score of 0.75 indicates that 75% of the students who took the same standardized test received the same score or lower.

58%	4333	Master	99%	99%	57%	Tier 1
56%	4310	Meet	99%	97%	44%	Tier 1
56%	4101	Meet	99%	74%	7%	Tier 1
18%	1429	Did Not Meet	1%	1%	1%	Tier 3 (RTI)
32%	1528	Did Not Meet	8%	1%	1%	Tier 3 (RTI)
29%	1578	Did Not Meet	52%	4%	1%	Tier 2
41%	1647	Approach	90%	15%	1%	Tier 1
12%	1375	Did Not Meet	1%	1%	1%	Tier 3 (RTI)
65%	1822	Meet	99%	99%	62%	Tier 1

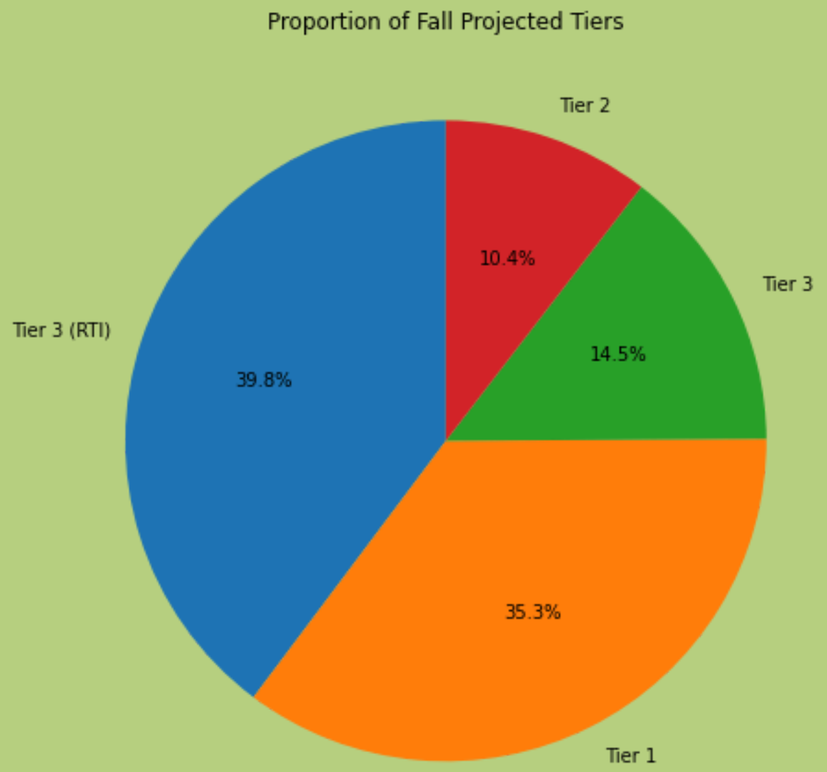
Data Understanding

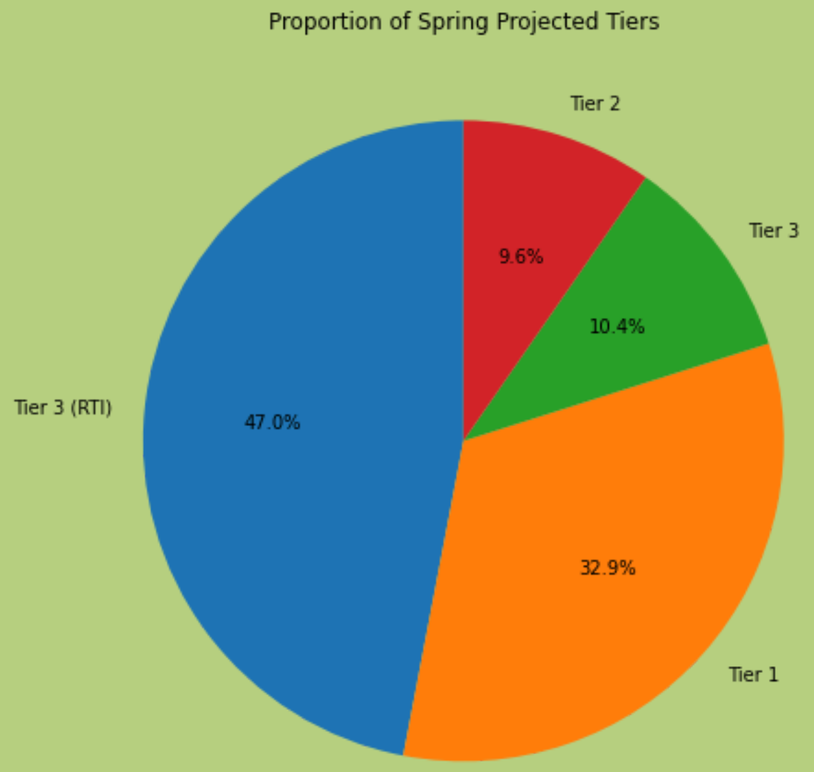
Projected Tiers and THE problem



Data Understanding

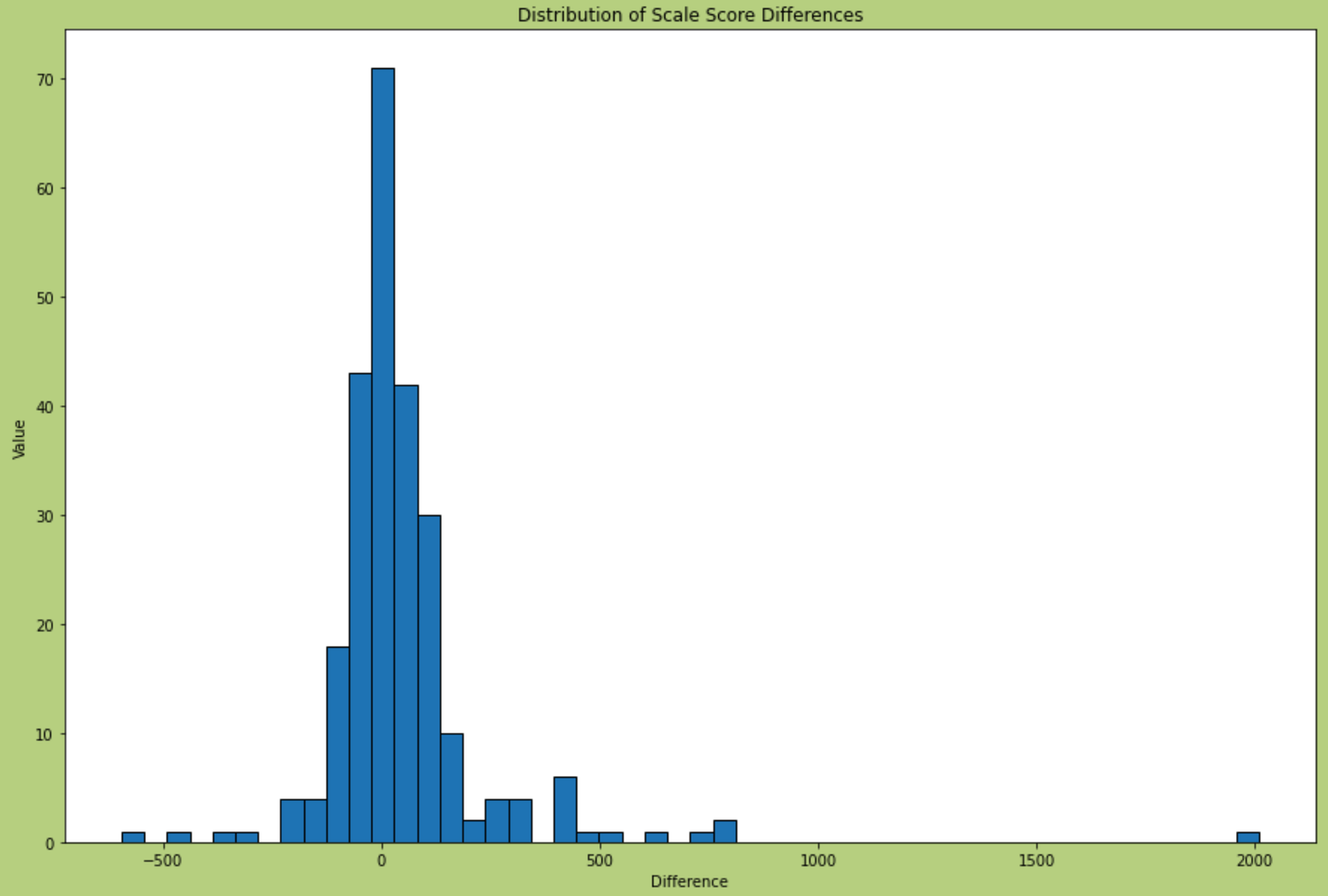
Projected Tiers and THE problem





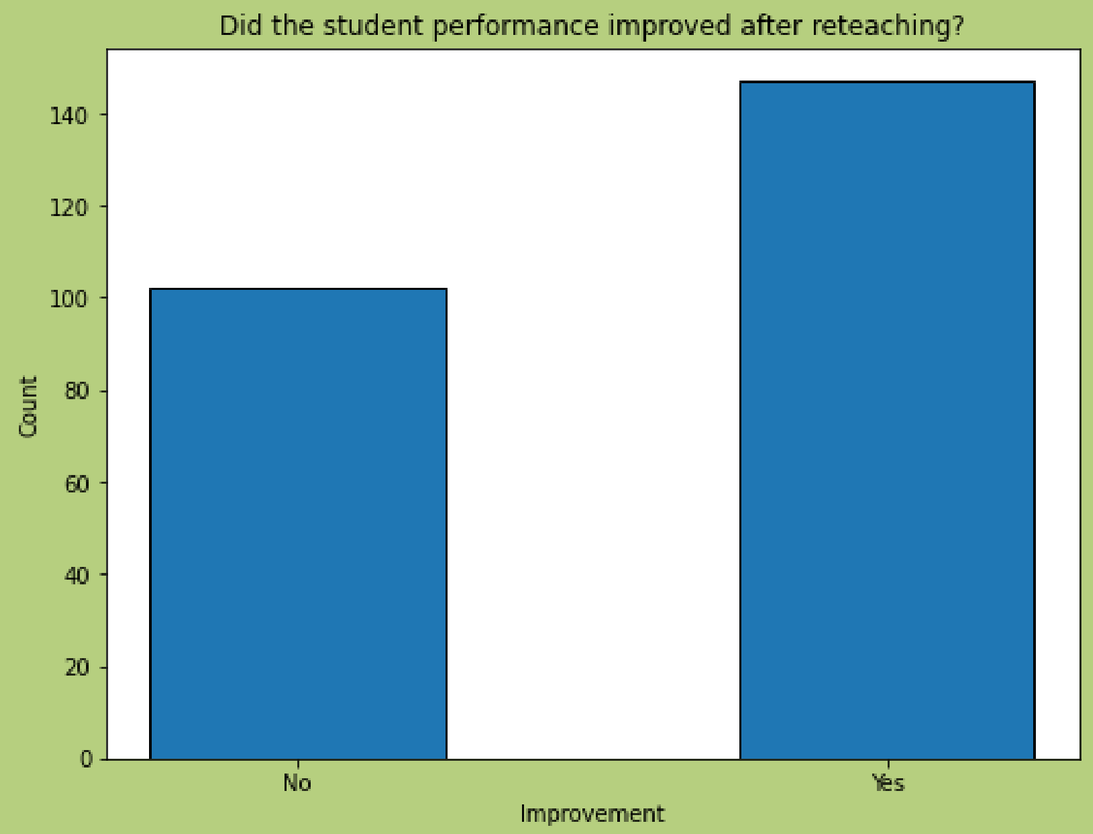
Feature Engineering

‘Difference’



Feature Engineering

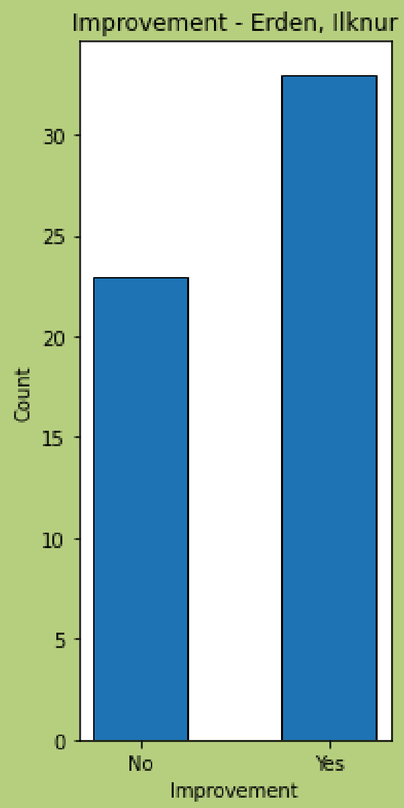
‘Improvement’

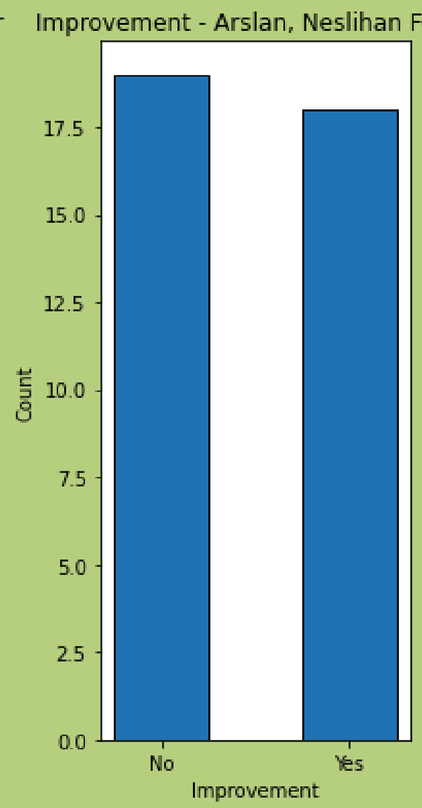


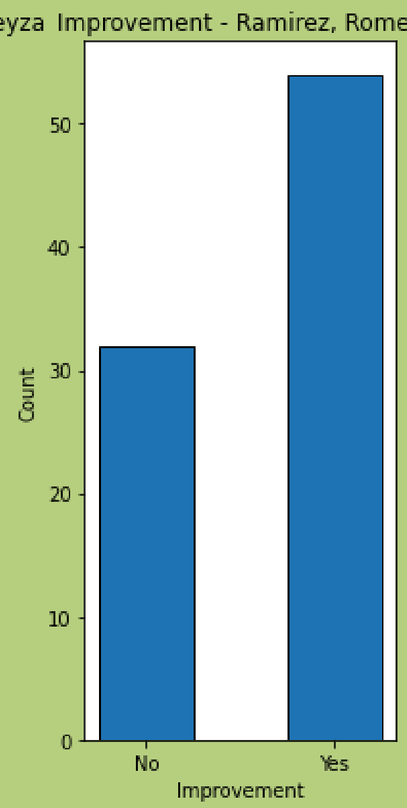
> there are more student improved but the count of students who did not do better in spring are also significant

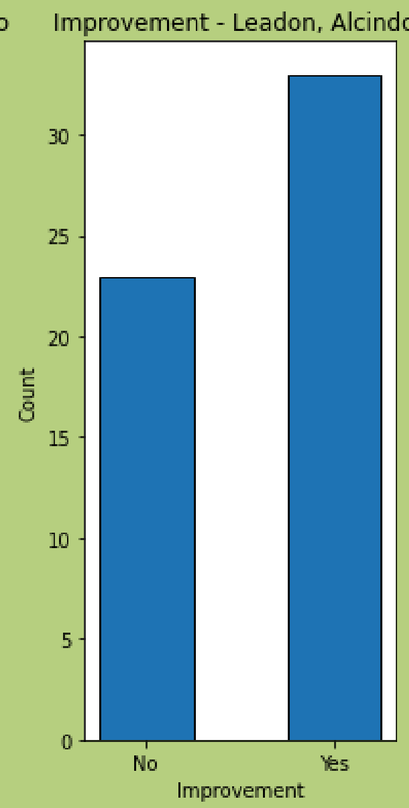
Data Understanding

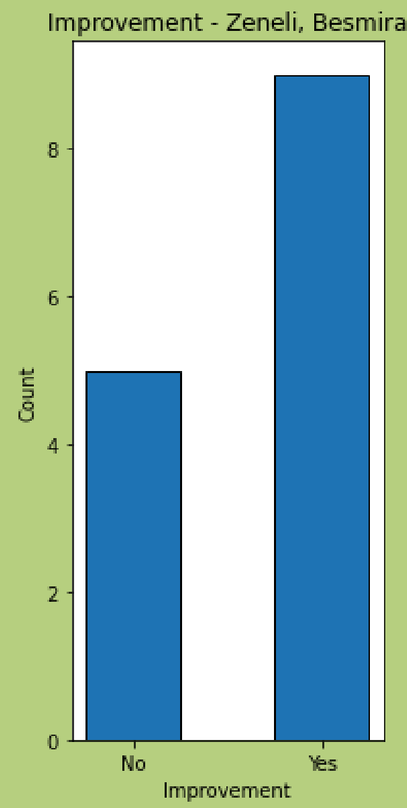
Improvement x Teacher





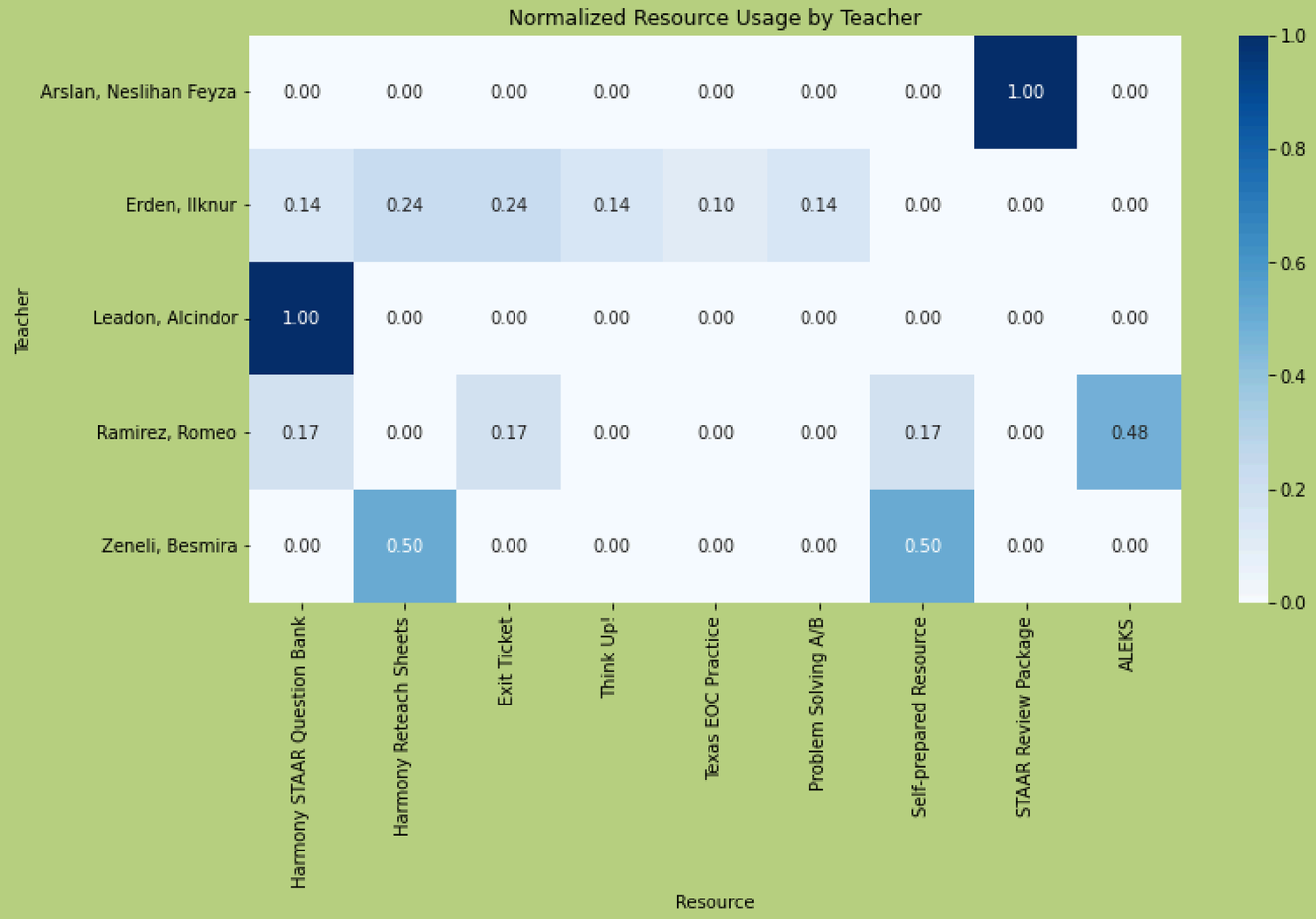






Data Understanding

Resource x Teacher



Data Understanding

One-Way ANOVA (Course v. FA Scale Score

```
   Multiple Comparison of Means - Tukey HSD, FWER=0.05
=====================================================
group1 group2  meandiff  p-adj   lower      upper    reject
-----------------------------------------------------
     0      1 -2502.6848   0.0 -2582.3714 -2422.9983   True
-----------------------------------------------------
ANOVA F-value: 3826.5269095179806
ANOVA p-value: 2.428445311522122e-152
```

> the mean of group2 (Math-8) is approximately 2502.6848 units lower than the mean of group1 (Algebra I).

Conclusion: Null hypothesis of no difference between the groups can be rejected based on the adjusted p-value. "True," means the null hypothesis is rejected, suggesting that there is a significant difference between the groups.

Data Understanding

Two-Way ANOVA (Teacher + Course v. Difference

	sum_sq	df	F	PR(>F)
C(Teacher)	1.827545e+06	4.0	13.165178	5.430736e-08
C(Course)	8.921356e+05	1.0	25.706882	7.904064e-07
C(Teacher):C(Course)	5.700087e+05	4.0	4.106200	3.077971e-03
Residual	8.398405e+06	242.0	NaN	NaN

> "Teacher" factor has a significant effect on the "Difference" variable, as indicated by the very low p-value (close to 0).

> "Course" factor also has a significant effect on the "Difference" variable, again indicated by the very low p-value.

> Interaction between "Teacher" and "Course" is also significant, as indicated by the low p-value.

Data Preprocessing

Getting data ready.

> Merged the datasets.

> No missing values.

> Outliers kept.

Feature Engineering

“Scaled_Score_Category”

```
data['Scaled_Score_Category'] = pd.cut(data['FA Scale Score'], bins=3, labels=['Low', 'Medium', 'High'])
```

> After removing Percent Score and Projected Tiers, I decided to create a “Projected Tier” like feature to robust the model.

Data Transformation

Label Encoding

> FA Mastery Projection

```
Label Mapping: {'Did Not Meet': 0, 'Approach': 1, 'Meet': 2, 'Master': 3}
```

> Scaled_Score_Category

```
Label Mapping: {'Low': 0, 'Medium': 1, 'High': 2}
```

Data Transformation

Feature Scaling

> Min-Max Scaling

$$x' = \frac{x - \mu}{\max(x) - \min(x)}$$

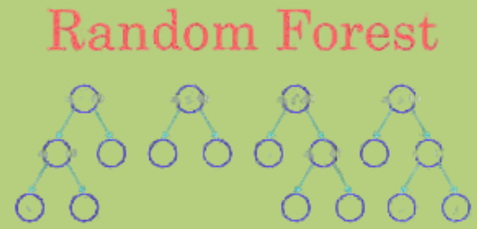
- scaling all values 0 to 1.
- not too sensitive to outliers like standardization

> Scaled features:

+ FA Scale Score
+ Difference

Model Selection

Random Forest



> Task: Regression

> Predicting ‘Difference’

> Advantages: Feature Importance and better accuracy most of the time. Robust to overfitting.

Model Selection

Logistic Regression

> Task: Classification

> Predicting ‘Improvement‘

> Advantages: makes it easy to interpret the effect of each predictor on the outcome. provides probabilistic predictions.

> Disadvantage: sensitive to outliers

Model Selection

XGBoost

XGBoost

> Task: Classification

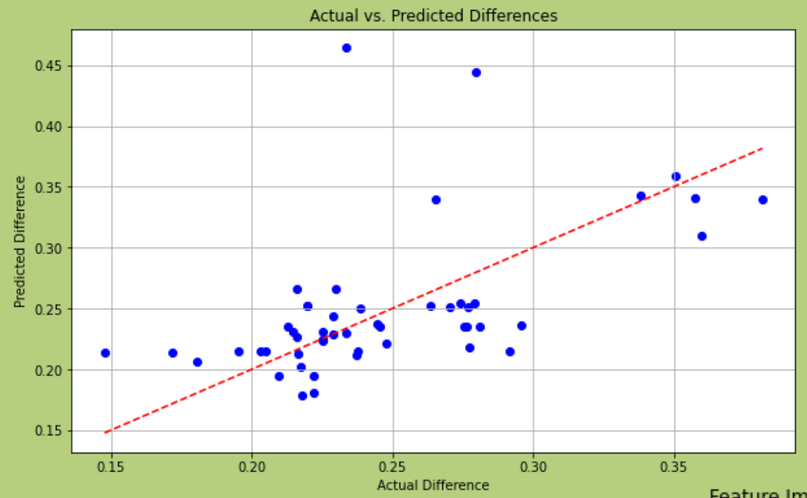
> Predicting ‘Improvement‘

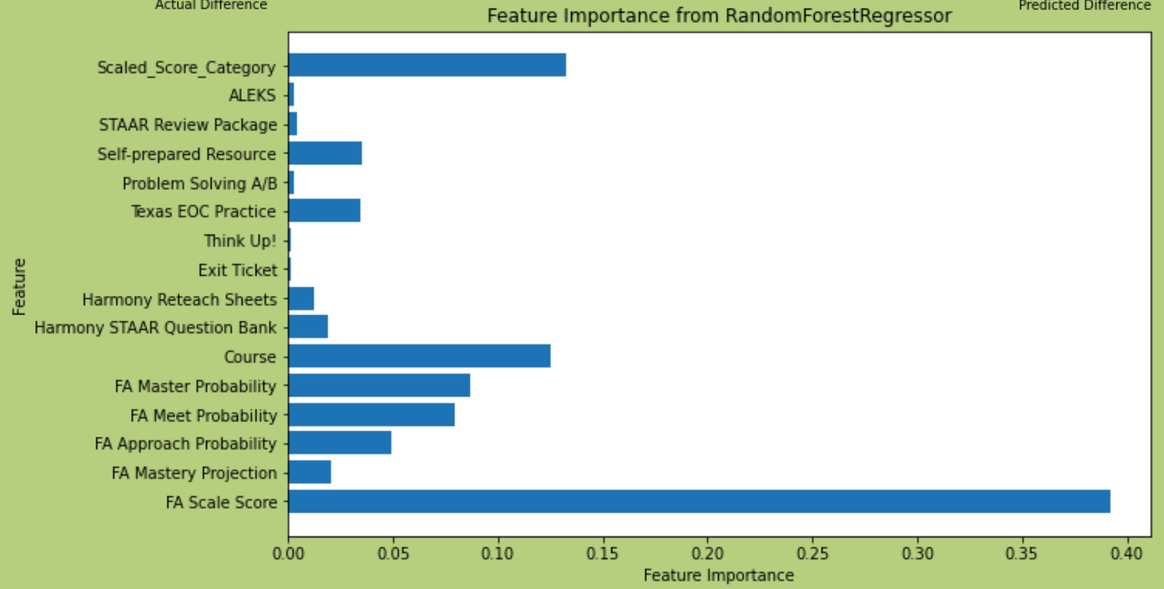
> Advantages: Feature Importance, high performance, prevents overfitting.

Model Evaluation

Random Forest

Mean Absolute Error: 0.03319865935322941
Mean Squared Error: 0.0026355245484846563
Root Mean Squared Error: 0.05133736016279622
R-squared: -0.14371070760874693

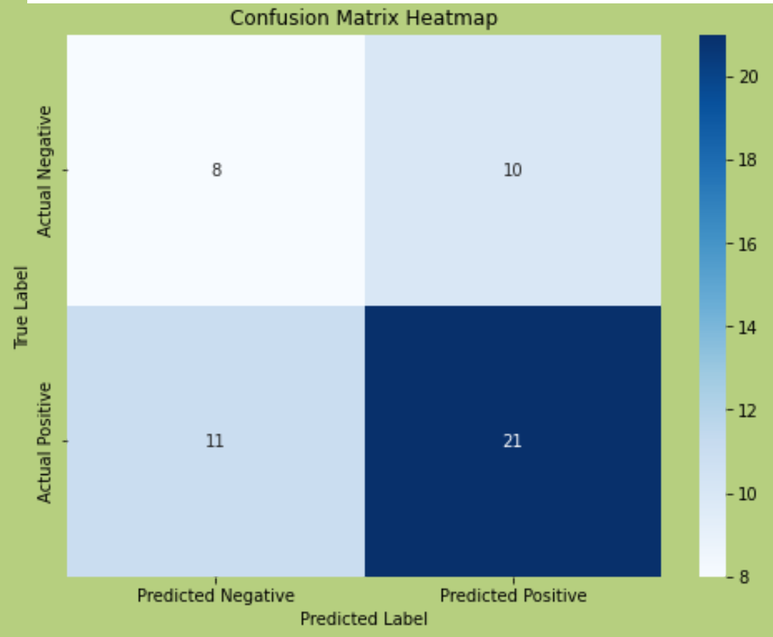




Model Evaluation

Logistic Regression

	coef	std err	t	P>|t|	[0.025	0.975]
Harmony STAAR Question Bank	0.5893	0.065	9.005	0.000	0.460	0.718
Harmony Reteach Sheets	0.4239	0.092	4.618	0.000	0.243	0.605
Exit Ticket	-0.0257	0.095	-0.269	0.788	-0.214	0.162
Think Up!	-0.2513	0.047	-5.389	0.000	-0.343	-0.159
Texas EOC Practice	0.3409	0.066	5.144	0.000	0.210	0.471
Problem Solving A/B	-0.2513	0.047	-5.389	0.000	-0.343	-0.159
Self-prepared Resource	0.2190	0.081	2.711	0.007	0.060	0.378
STAAR Review Package	0.4865	0.081	6.043	0.000	0.328	0.645
ALEKS	0.5714	0.062	9.262	0.000	0.450	0.693

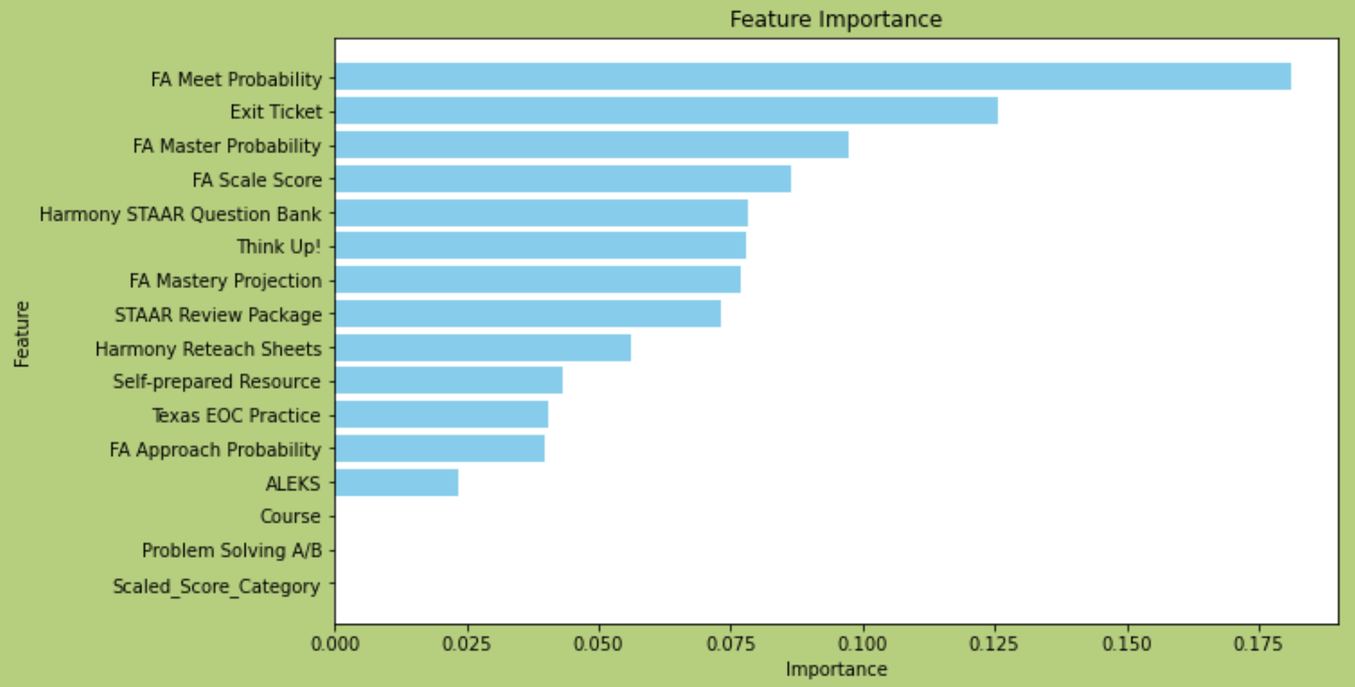


Accuracy: 0.58
Classification Report:

	precision	recall	f1-score	support
0	0.42	0.44	0.43	18
1	0.68	0.66	0.67	32
accuracy			0.58	50
macro avg	0.55	0.55	0.55	50
weighted avg	0.59	0.58	0.58	50

Model Evaluation

XGBoost



Accuracy: 0.64

Precision: 0.7333333333333333

Recall: 0.6875

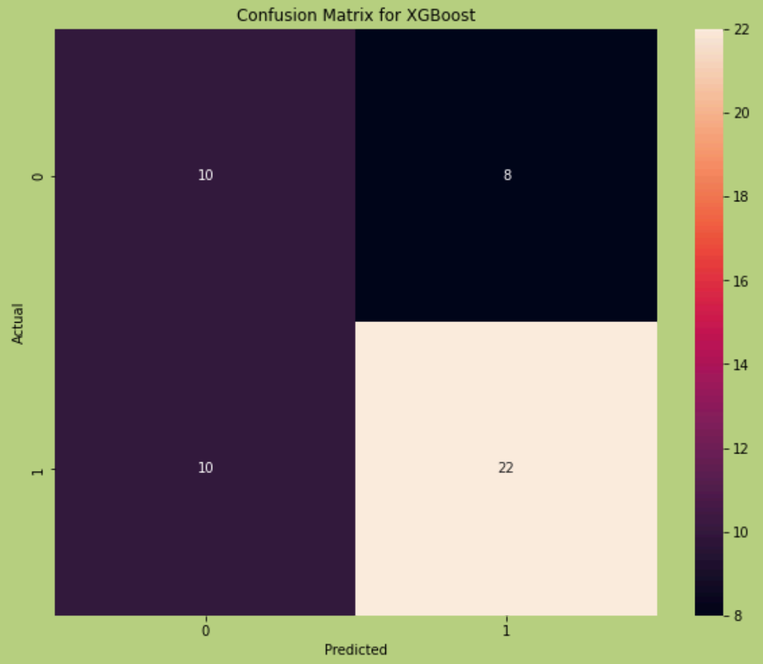
F1 Score: 0.7096774193548386

AUC-ROC: 0.640625

MAE: 0.3600

MSE: 0.3600

Log Loss: 0.9788



Conclusion

Limitations

> Small and imbalanced dataset.

> Could be better with more features including student demographics. log data etc.

Resource Effectiveness

> Think Up! and Problem Solving A/B negatively affects the reteaching process.

> Harmony STAAR Question Bank, Harmony Reteach Sheets, Texas EOC Practice, Self -prepared resources, STAAR Review Package, and ALEKS positively affects the student performance.