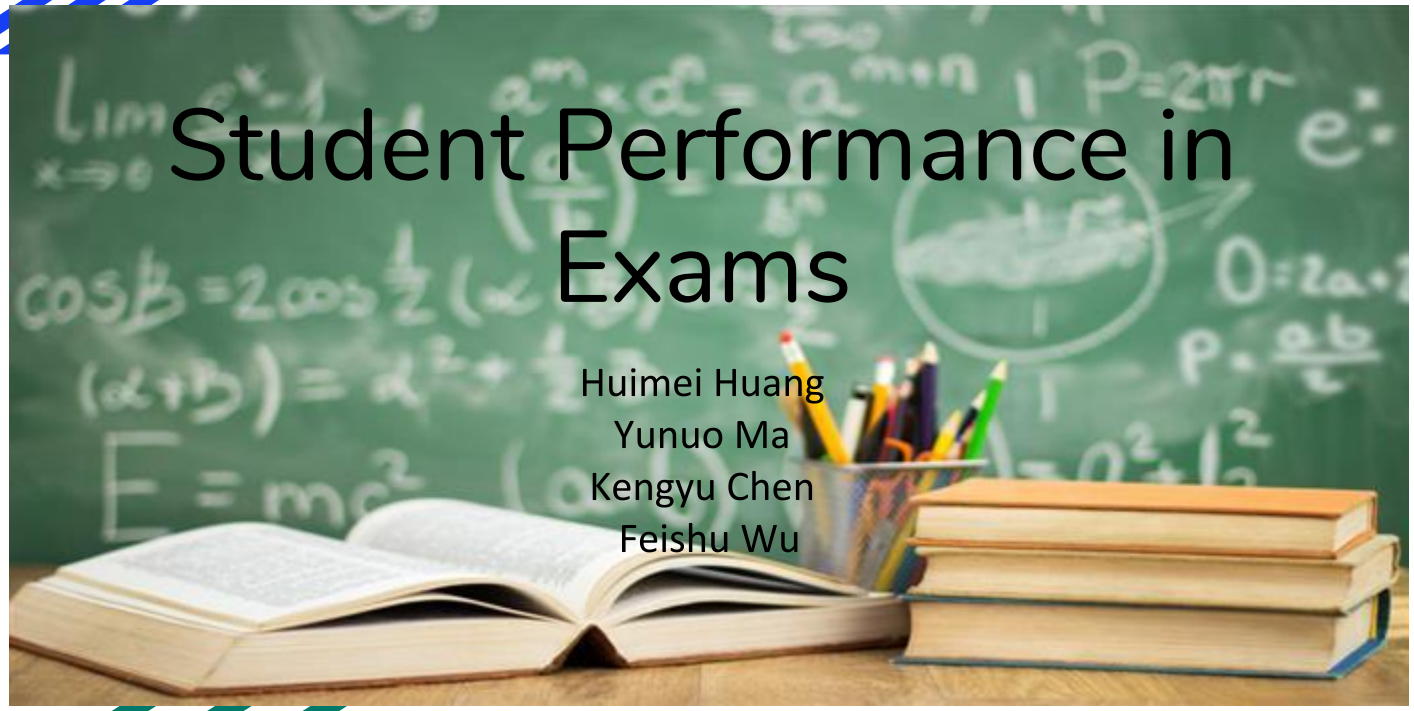
Student Performance in Exams



Huimei Huang
Yunuo Ma
Kengyu Chen
Feishu Wu

Introduction and field of study

Discuss the certain factors that influence and affect each student's performance in exams

–*Intrinsic factors*

–*Extrinsic factors*

–*Miscellaneous factors*

–*Personal factors*

Find out the influence of the parents' background, test preparation etc. on students' performance.

Data Background information

Name: Student Performance in Exams

Source: Kaggle.com

Size: 1000 rows and 8 columns

1000 rows represents 1000 students

8 columns are gender, race/ethnic, lunch, parental level of education, test preparation course, Math score, Writing score, and reading score.

Data Description

The columns following the Math, writing, reading scores estimate the extent to which each of five factors:

Gender: Male(52%) and Female(48%)

race/ethnicity: Group A(9%), B(19%), C(32%), D(26%), E(24%)

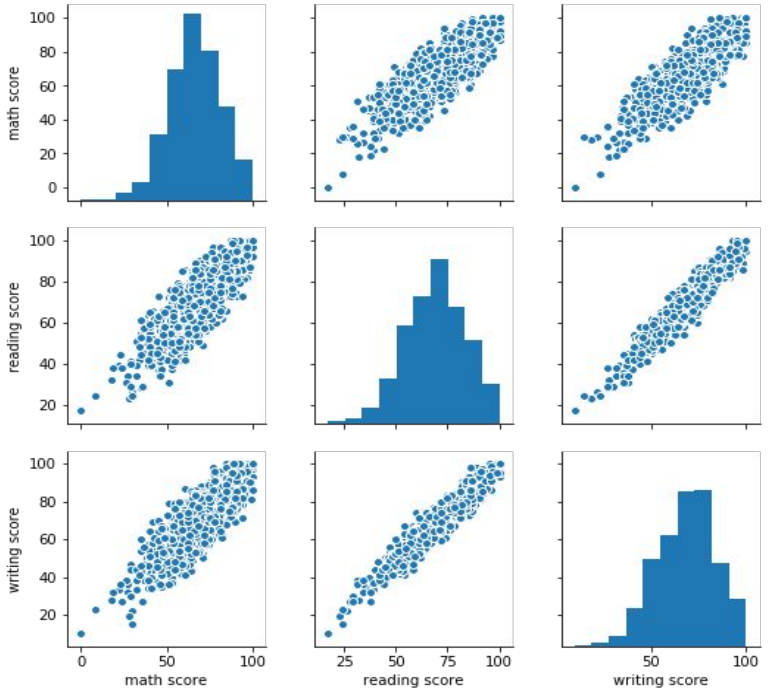
Parental level of education: Some college (23%), associated degree (22%), high school (20%), some high school (18%) bachelor's degree (11%) and master's degree (7%)

Lunch: standard (65%) and free/reduced (36%)

Test preparation Course: None(64%) and Completed(36%)

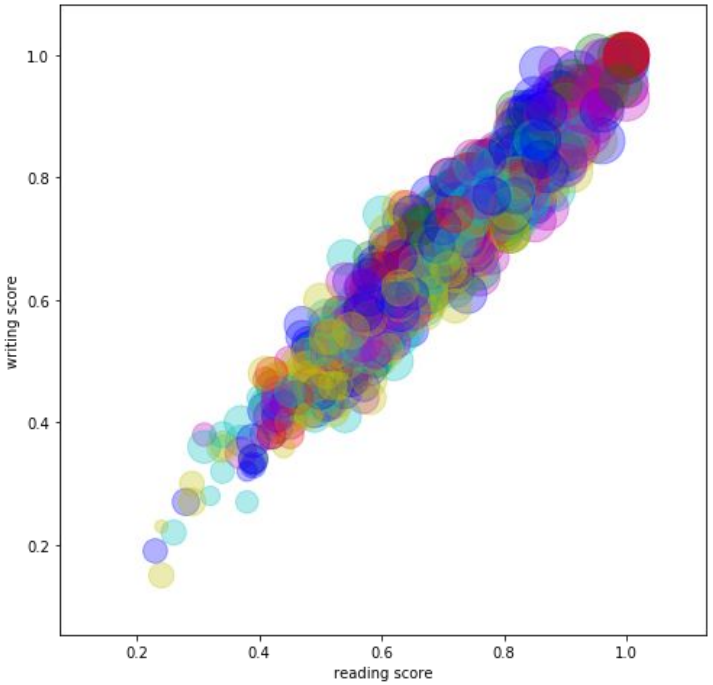
What is the correlation of these 3 scores?

Math score, reading score and writing score are all positive relatives and 3 distribution of scores are approximately normalized distribution.



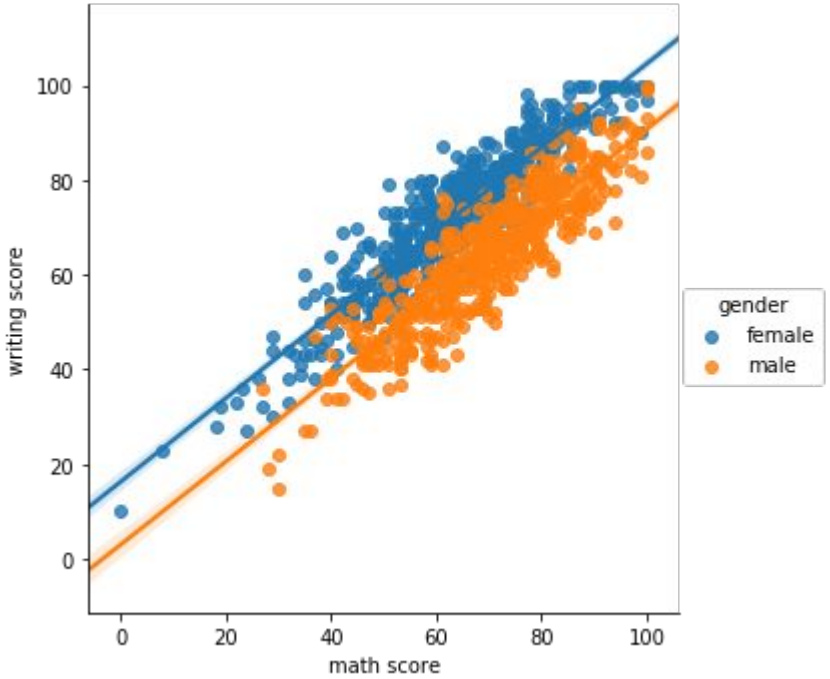
The relation between math score, writing score, reading score and parental level of education.

The parental level of education has effect on students all kind of score. If the education level of students' parents are high school or some high school, their score will be lower than students whose parents has some college, bachelor's degree, master's degree and associate's degree.



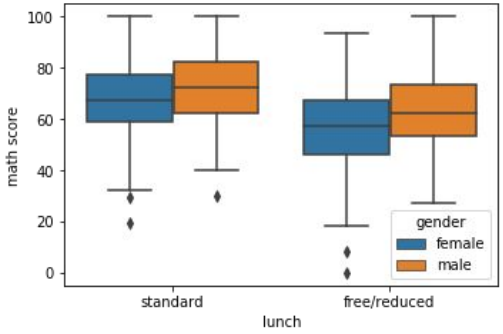
Is there has any relationship between writing score, math score and gender?

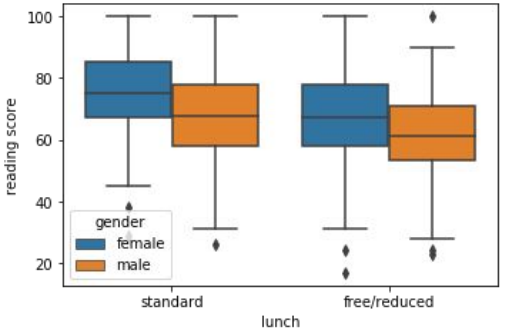
The plot shows female has higher writing score than male. And there has no that much difference between female and male in math score.

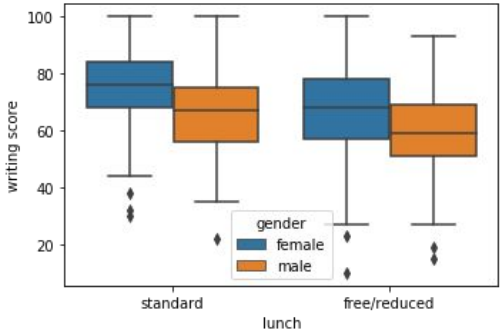


Does free lunch for different gender affect the score?

female have higher score in writing and reading, male have higher score in math. Also, we see that students who have free/reduced lunch have lower score than standard.

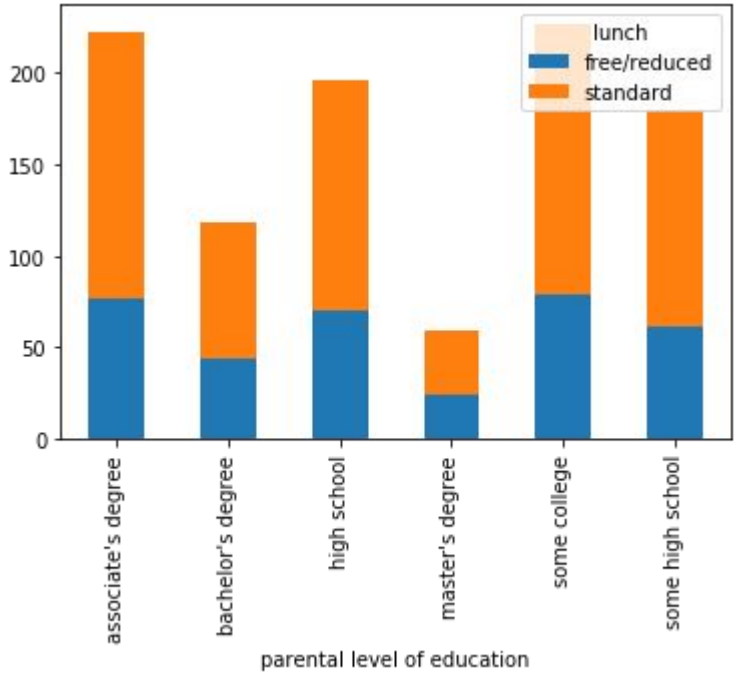






Does the parents' education level closely related to student's lunch eating habits?

There might exist some positive relationship between parent's education level and the proportion of free/reduced lunch. It is interesting that for parents' who have a higher education level with a specific degree, their children might have a higher possibility to have free/reduced lunch. However, this not very obvious. This might result in family eating habits and the ways of parenting.



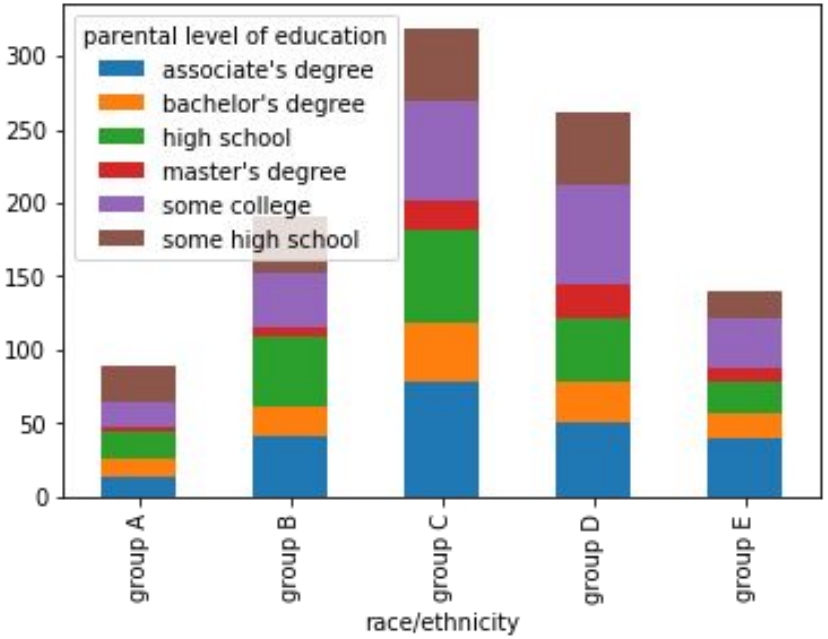
Which race has higher score ?

Group E has highest score in all three scores(both male and female), and group A has lowest score in all three score (both male and female).

	math score		reading score		writing score	
gender	female	male	female	male	female	male
race/ethnicity						
group A	58.527778	63.735849	69.000000	61.735849	67.861111	59.150943
group B	61.403846	65.930233	71.076923	62.848837	70.048077	60.220930
group C	62.033333	67.611511	71.944444	65.424460	71.777778	62.712230
group D	65.248062	69.413534	74.046512	66.135338	75.023256	65.413534
group E	70.811594	76.746479	75.840580	70.295775	75.536232	67.394366

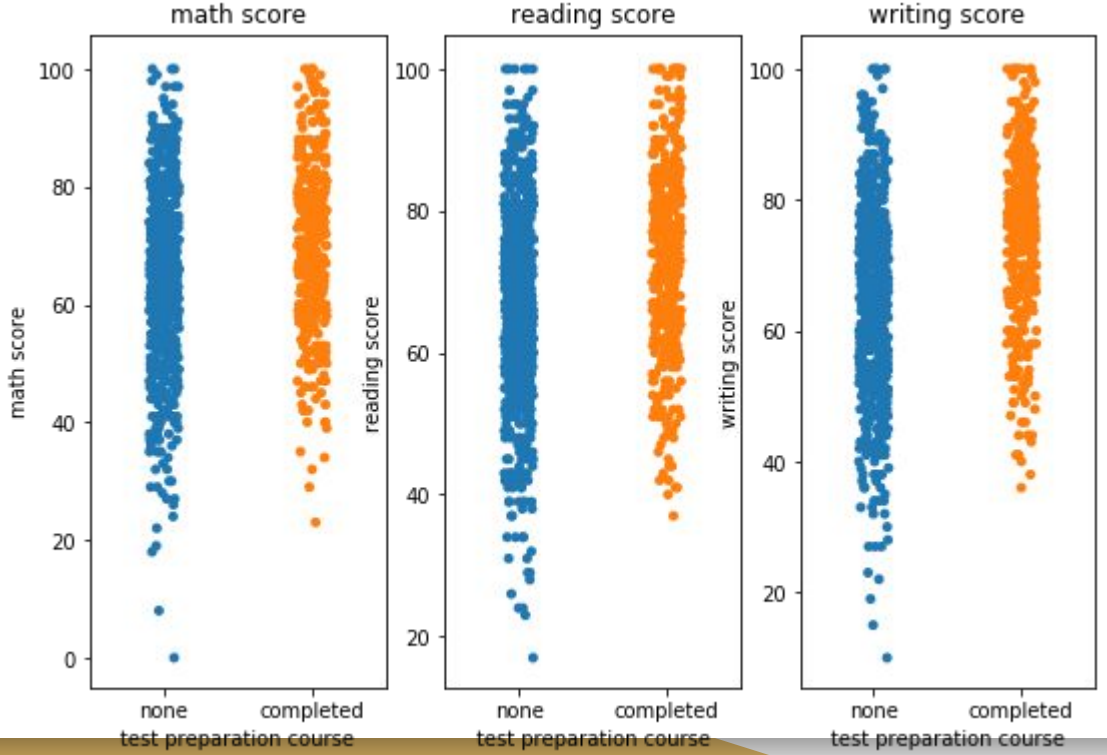
Group E has highest score, does it relation to their parental level of education?

The education level of most of the students' parents in group E are associate's degree and some college, and there only small number on some high school and high school.



How does test preparation affect students' average reading score?

The plots shows students who had completed the test preparation course have higher score in all three subject.



Modeling

1. Linear Regression Model
2. Logistic regression model
3. Classification logistic regression with grid search cross validation
4. Random forest
5. Randomized search cross validation with decision tree

Change category variables to dummy variables

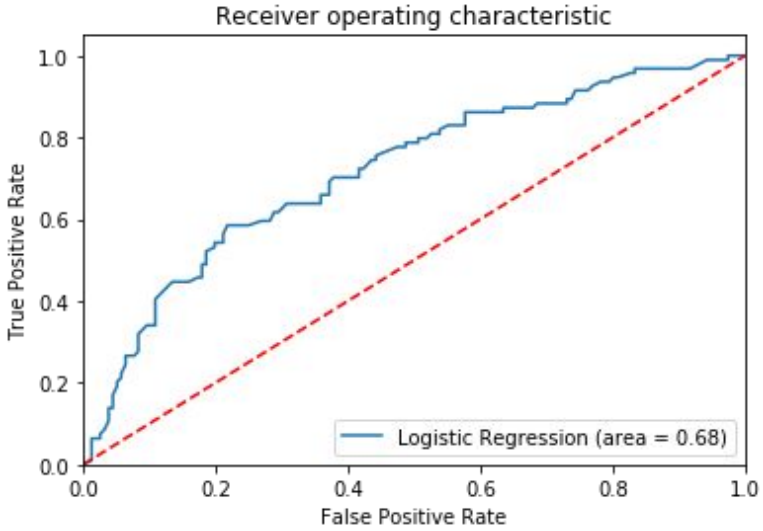
female	male	group A	group B	group C	group D	group E	associate's degree	bachelor's degree	high school	master's degree	some college	some high school	free/reduced	standard	completed	none
1	0	0	1	0	0	0	0	1	0	0	0	0	0	1	0	1
1	0	0	0	1	0	0	0	0	0	0	1	0	0	1	1	0
1	0	0	1	0	0	0	0	0	0	1	0	0	0	1	0	1
0	1	1	0	0	0	0	1	0	0	0	0	0	1	0	0	1
0	1	0	0	1	0	0	0	0	0	0	1	0	0	1	0	1

Change all three score to P/NP

Pass > = 70	1 as pass
No pass < 69	0 as no pass

simple Logistic Regression Model

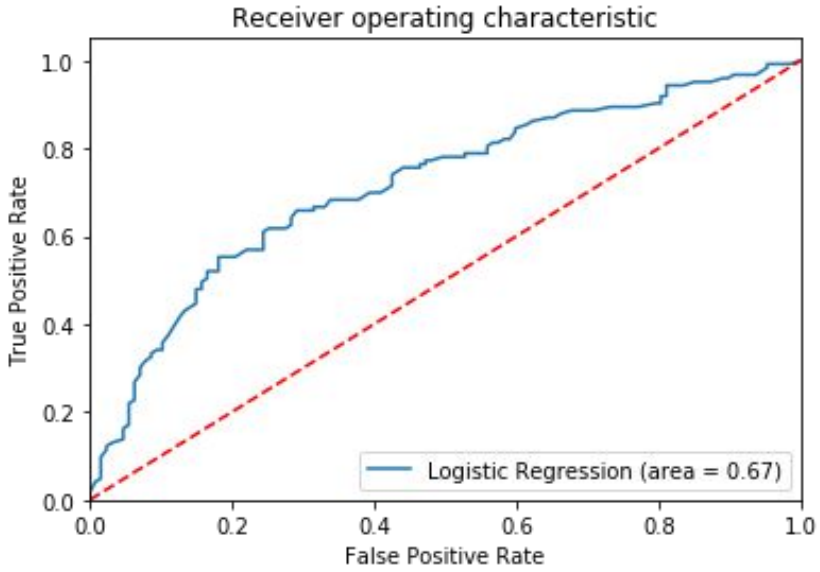
Math



	predict: P	predict: NP
actual: P	123	33
actual: NP	41	53

training accuracy:	0.66
testing accuracy:	0.7

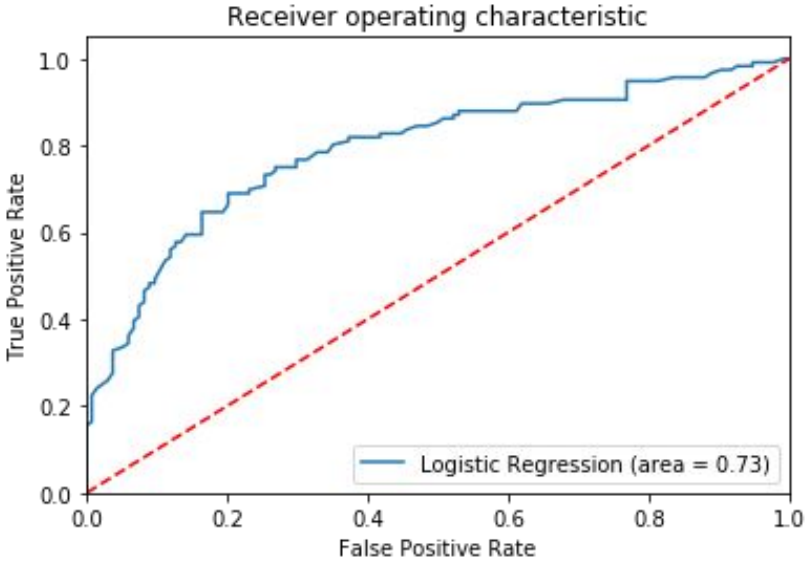
Reading



	predict: P	predict: NP
actual: P	85	42
actual: NP	41	82

training accuracy:	0.67
testing accuracy:	0.67

Writing



	predict: P	predict: NP
actual: P	103	31
actual: NP	35	81

training accuracy:	0.69
testing accuracy:	0.74

Classification Logistic Regression with Grid Search Cross Validation

Math: Tuned Logistic Regression Accuracy: 0.6516666666666666

Reading: Tuned Logistic Regression Accuracy: 0.6583333333333333

Writing: Tuned Logistic Regression Accuracy: 0.695

Randomized Search CV with Decision Tree

Math: Best Score is 0.643

Reading: Best score is 0.632

Writing: Best score is 0.682

Accuracy comparative

	Train_accu			test_accu		
model	LogisRegres w GridSearch CV	Logistic Regression	Random Forest	LogisRegres w GridSearch CV	Logistic Regression	Random Forest
subject						
Math	NA	0.66	0.741	0.652	0.70	0.620
Reading	NA	0.67	0.743	0.658	0.67	0.584
Writing	NA	0.69	0.760	0.695	0.74	0.608

Conclusion

After exploring the data of student performance on the exam, we get a better understanding about what factors contribute the most to the exam scores. We also examined how different variables affect and interact with one another. We are now about 70% confident to predict students' exam scores in the future with the given information in the data using logistic regression.¶