1. **Introduction**

One of most important subject in school is math. What I want to do is check whether other variables influences student's math score and if is possible predict the student math score.

2. **Data**

I will use a kaggle dataset (https://www.kaggle.com/spscientist/students-performance-in-exams).

This dataset have 8 columns:
- gender
- race/ethnicity
- parental level of education
- lunch
- test preparation course
- math score
- reading score
- writing score

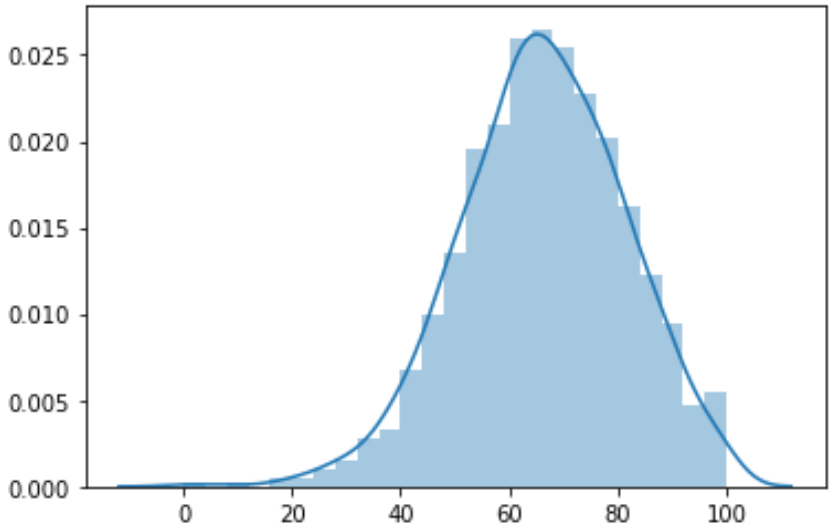
3. **Methodology**

The first step is the data exploration to understand the data, checking missing data and data balance. Next I will verify if exists any relationship among features and math score.

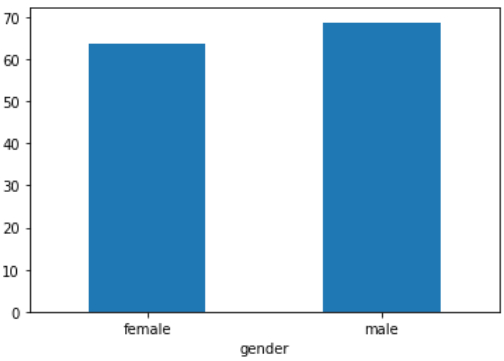
Then I will try predictive models to identify a student's math score and identify features that more influences the math score.

4. **Results and discussion**

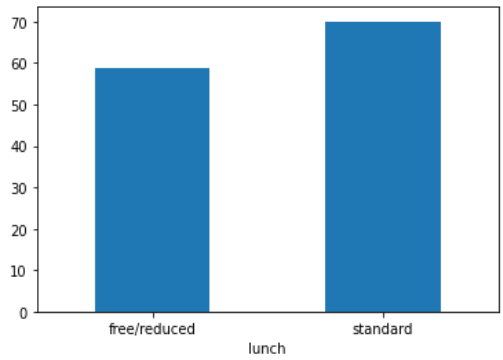
After data analysis I detect the data has no missing data and the balance is OK (math score).



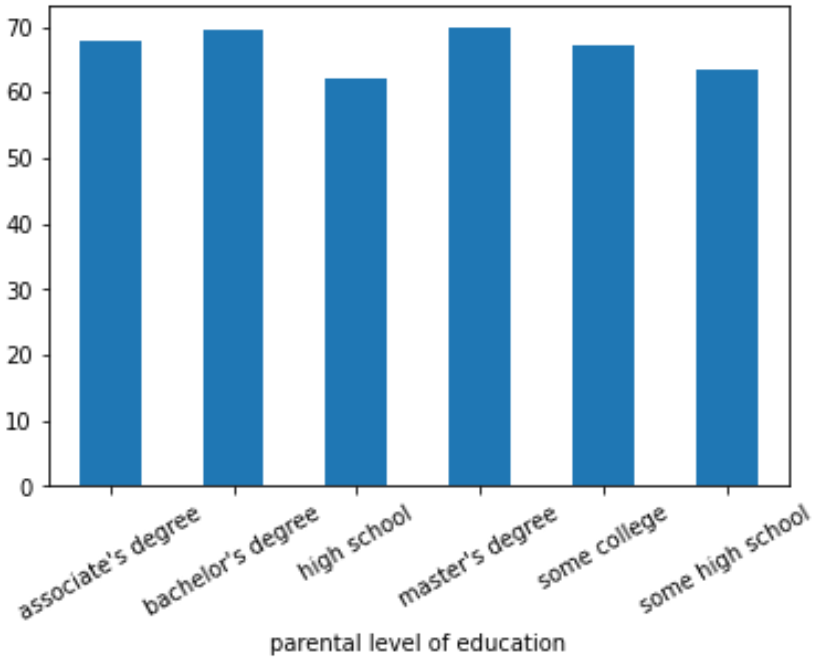
Verifying the relation of math score and other features I can see that exists a correlation, like:



gender x math score mean



lunch x math score mean



Parental level education x math score mea

I use Label Encoder to encode some features like gender and lunch and use get dummies to features with many values possible.

Then I try to predict the student math score with KNN, Decision Tree, Extra Regression Tree and Randon Forest.

5. **Conclusion**

The regression's accuracy table is:

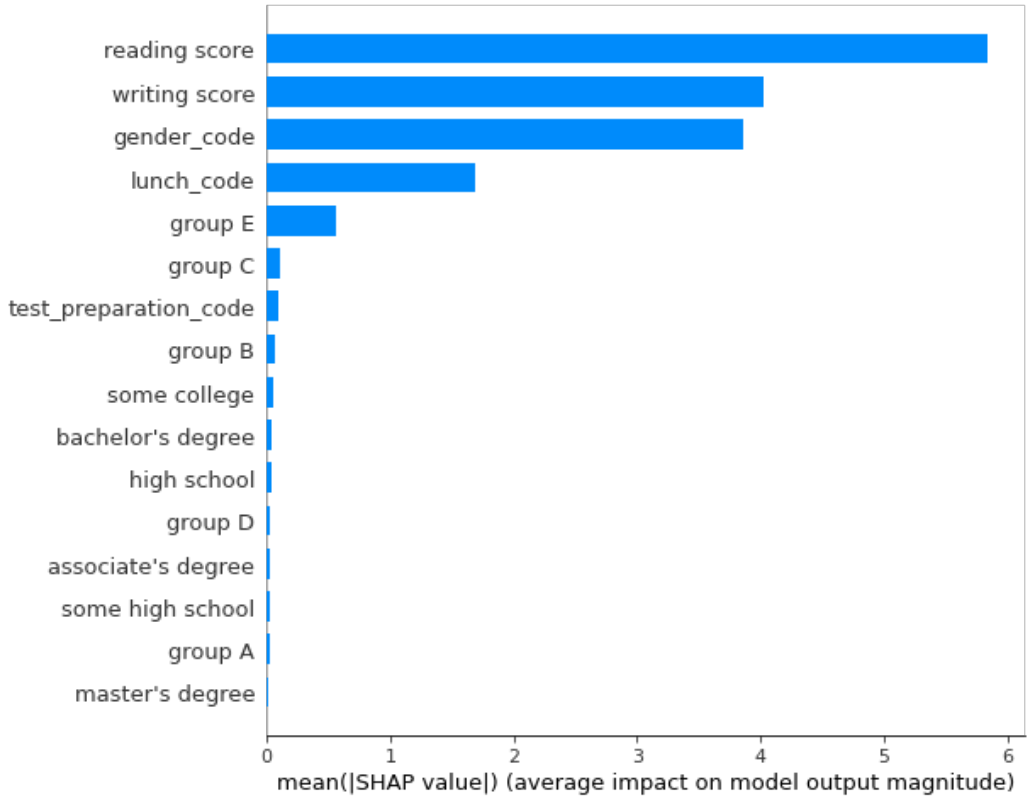
Model	MAE
KNN Regressor (k=6)	7.934166666666665
Decicion Tree Regressor (n=5)	4.931523210662771
Extra Tree Regression	5.770071595729278
Random Forest Regression	4.884995948853916

With Random Forest Regression I have better results and using SHAP I can demonstrate that some features influences the math score.

I understand the reading / wrinting score are importants, because are basics, so without it you can not good math score.

Then, lunch is a important feature. A good nutrition is essential for any activity.

In the next I show the features importance in Random Forest Regression and the features correlation



	gender_code	lunch_code	test_preparation_code	math score	reading score	writing score
gender_code	1.000000	0.021372	-0.006028	0.167982	-0.244313	-0.301225
lunch_code	0.021372	1.000000	0.017044	0.350877	0.229560	0.245769
test_preparation_code	-0.006028	0.017044	1.000000	-0.177702	-0.241780	-0.312946
math score	0.167982	0.350877	-0.177702	1.000000	0.817580	0.802642
reading score	-0.244313	0.229560	-0.241780	0.817580	1.000000	0.954598
writing score	-0.301225	0.245769	-0.312946	0.802642	0.954598	1.000000