

Отчет по заданию «Визуализация данных»

Таскынов Ануар, 517 группа

Курс «Прикладные задачи анализа данных»

12 октября 2017 г.

Описание данных

Данные были получены при опросе студентов по математике и португальскому языку в средней школе. Данные содержат много интересной социальной, гендерной и учебной информации о школьниках. В школе всего три контрольные за год и нужно предсказать оценку на последней контрольной.

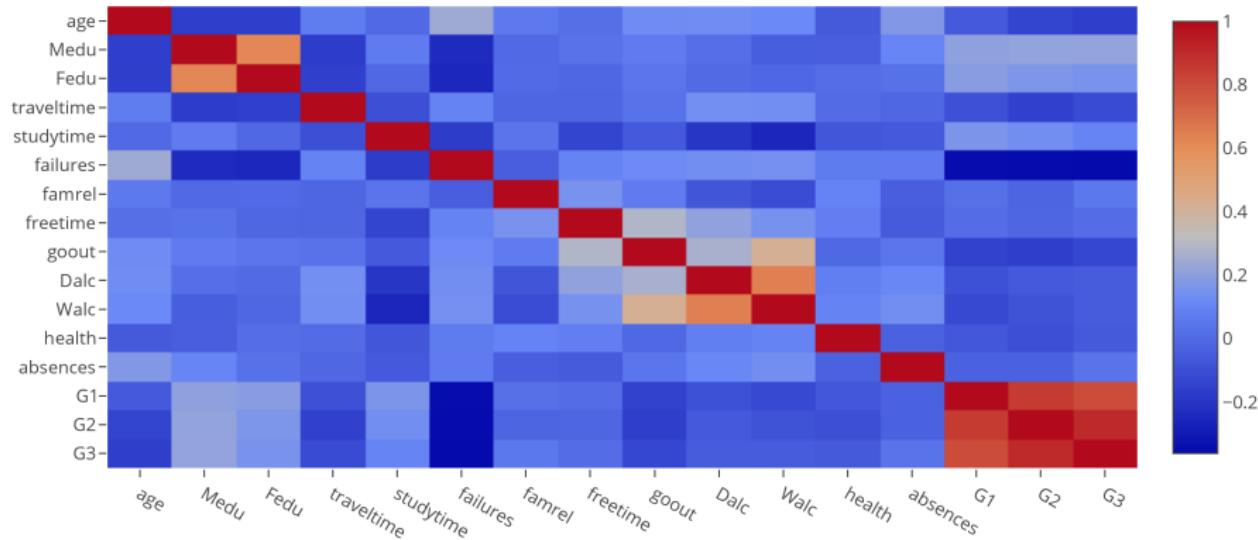
	school	sex	age	address	famsize	Pstatus	Medu	Fedu	Mjob	Fjob
0	GP	F	18	U	GT3	A	4	4	at_home	teacher
1	GP	F	17	U	GT3	T	1	1	at_home	other
2	GP	F	15	U	LE3	T	1	1	at_home	other
3	GP	F	15	U	GT3	T	4	2	health	services
4	GP	F	16	U	GT3	T	3	3	other	other

	reason	guardian	traveltime	studytime	failures	schoolsup	famsup	paid	activities
0	course	mother	2	2	0	yes	no	no	no
1	course	father	1	2	0	no	yes	no	no
2	other	mother	1	2	3	yes	no	yes	no
3	home	mother	1	3	0	no	yes	yes	yes
4	home	father	1	2	0	no	yes	yes	no

	nursery	higher	internet	romantic	famrel	freetime	goout	Dalc	Walc	health	absences	G1	G2	G3
0	yes	yes	no	no	4	3	4	1	1	3	6	5	6	6
1	no	yes	yes	no	5	3	3	1	1	3	4	5	5	6
2	yes	yes	yes	no	4	3	2	2	3	3	10	7	8	10
3	yes	yes	yes	yes	3	2	2	1	1	5	2	15	14	15
4	yes	yes	no	no	4	3	2	1	2	5	4	6	10	10

Корреляция между числовыми признаками.

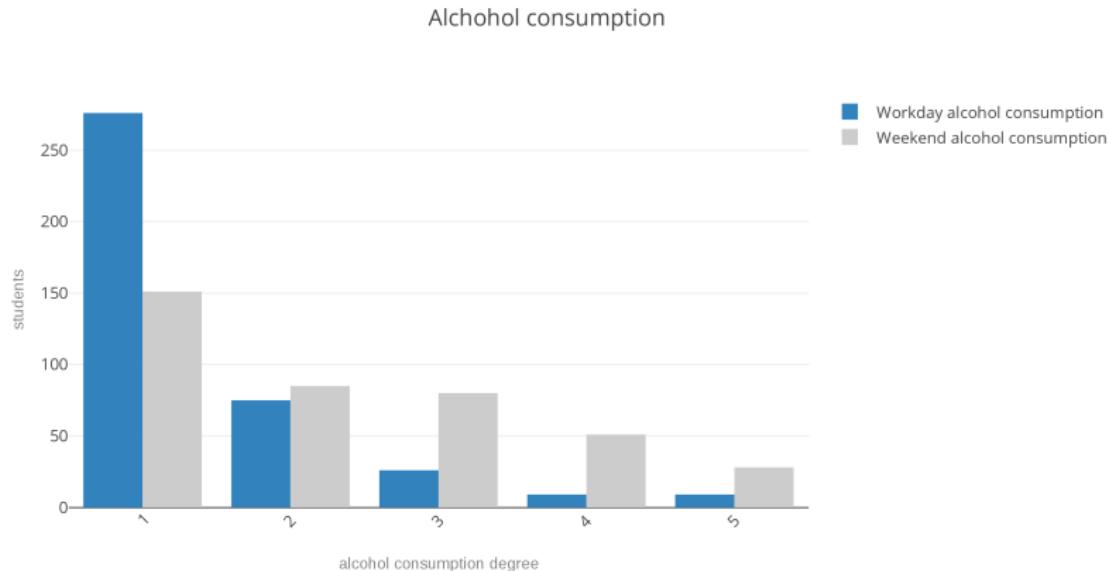
Для начала построим корреляционную карту между числовыми признаками.



Видно, что значение последующих оценок зависит строго от предыдущих оценок и не зависит от степени алкогольного потребления. Но так ли это на самом деле?

Распределение пьющих школьников.

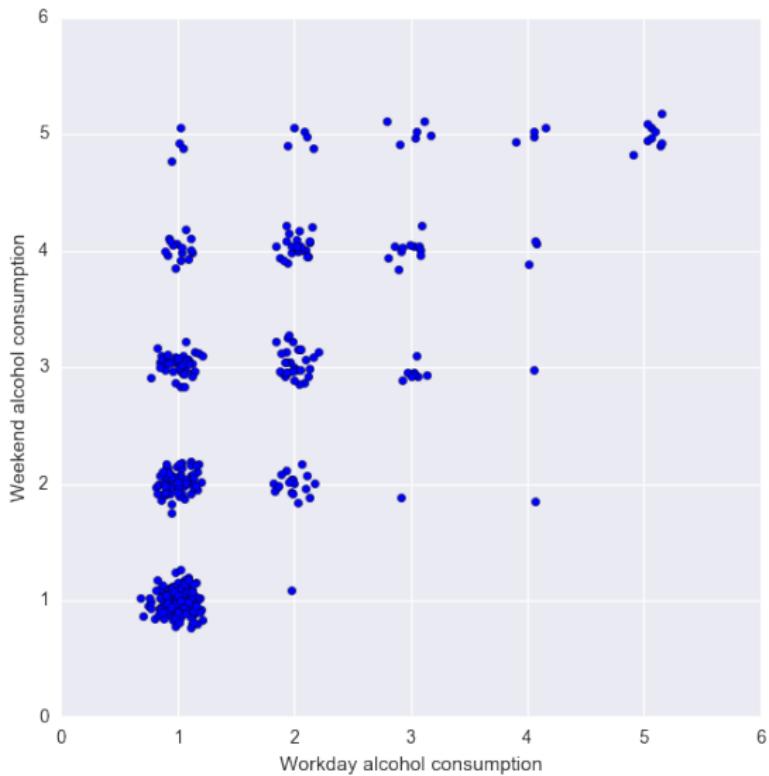
Рассмотрим распределение пьющих школьников по выходным дням и по будням.



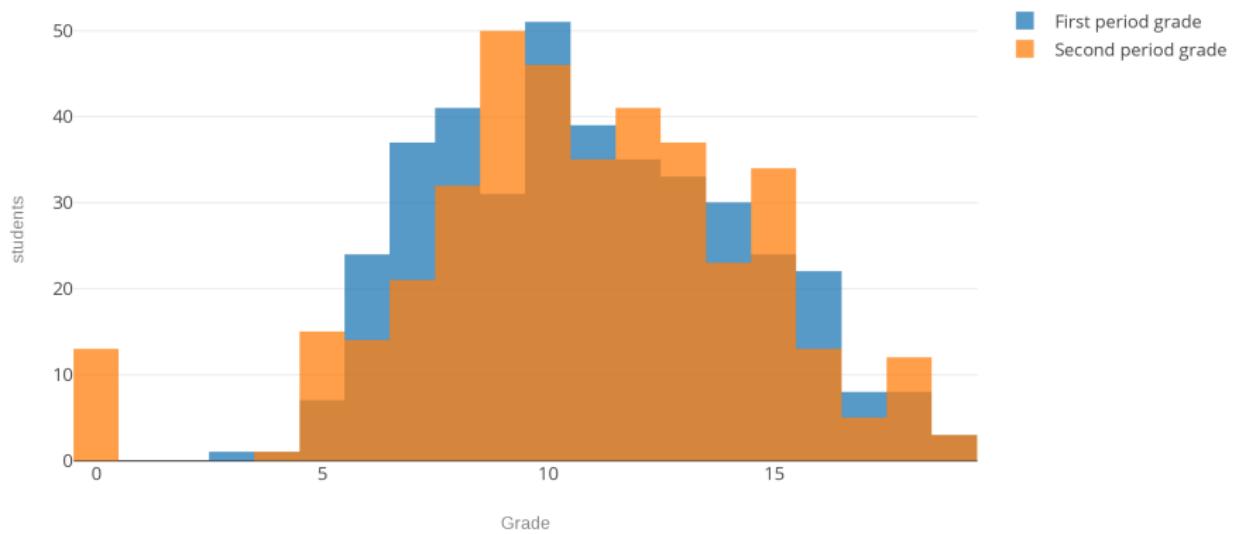
Количество студентов, пьющих в будние меньше, чем в выходные.

Распределение пьющих школьников.

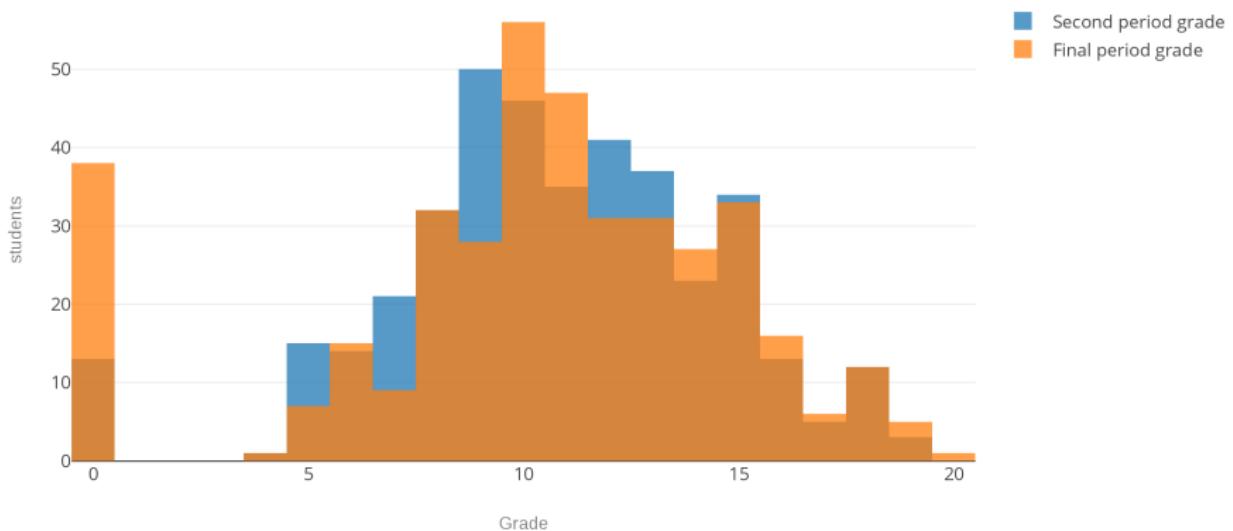
Не был учтен отдельно каждый школьник. По данной гистограмме невозможно это понять. Рассмотрим под другим углом.



Распределение по оценкам.

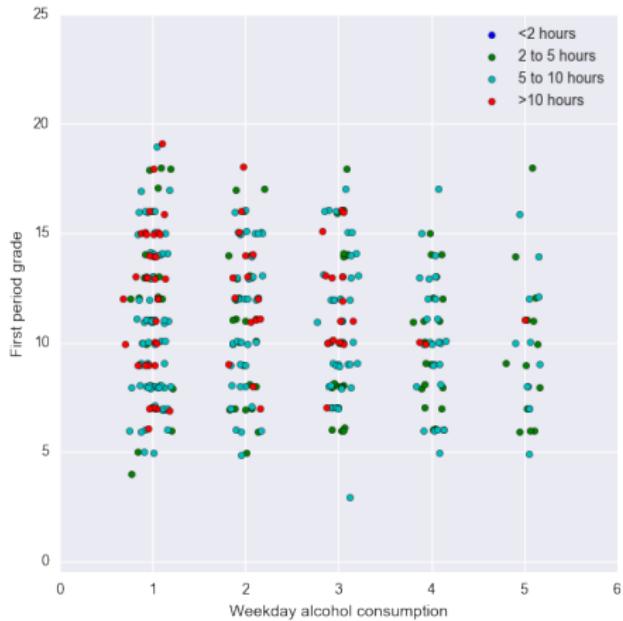
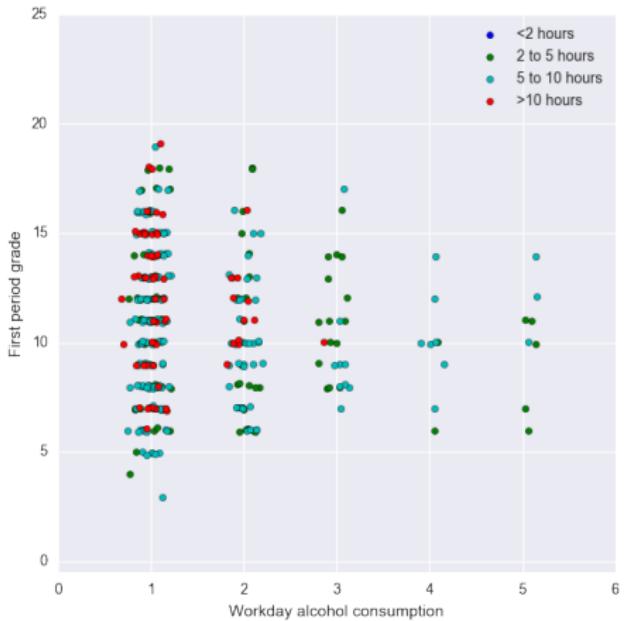


Распределение по оценкам.

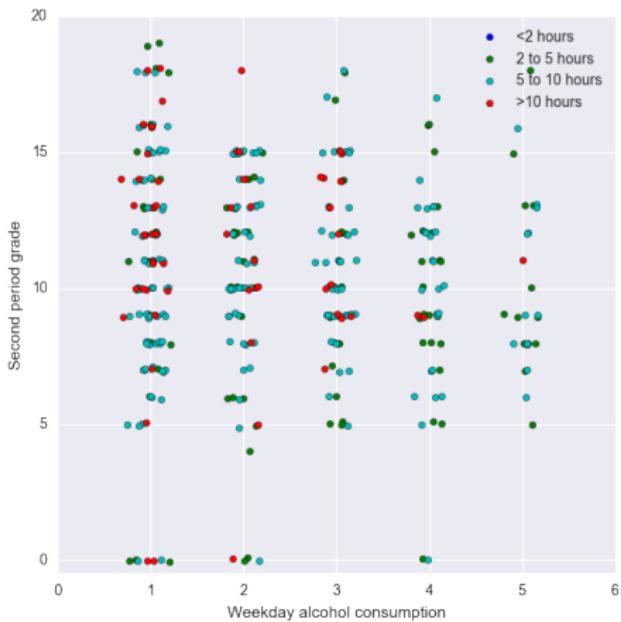
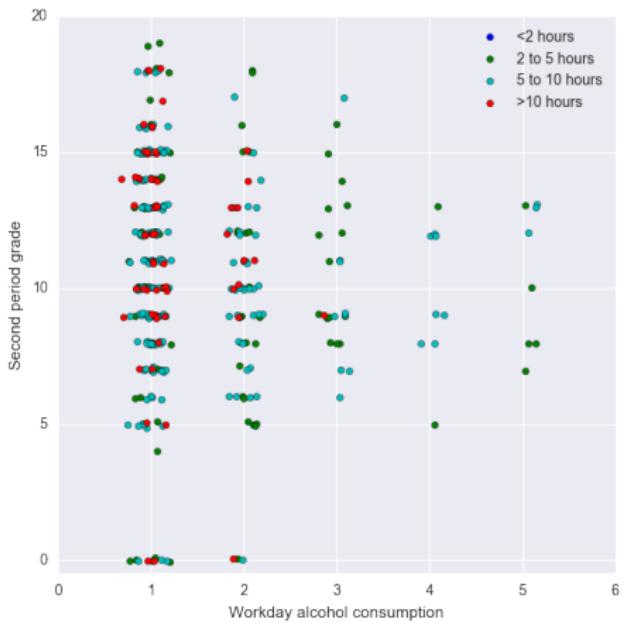


Все больше людей проваливают экзамены...

Сколько времени уделяется учебе?



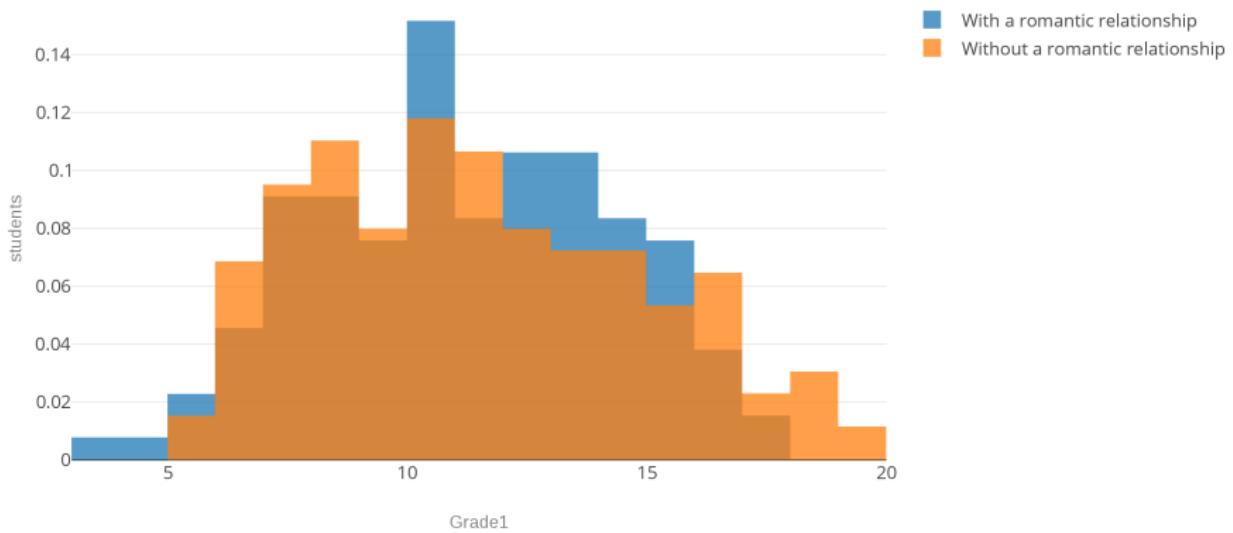
Сколько времени уделяется учебе?



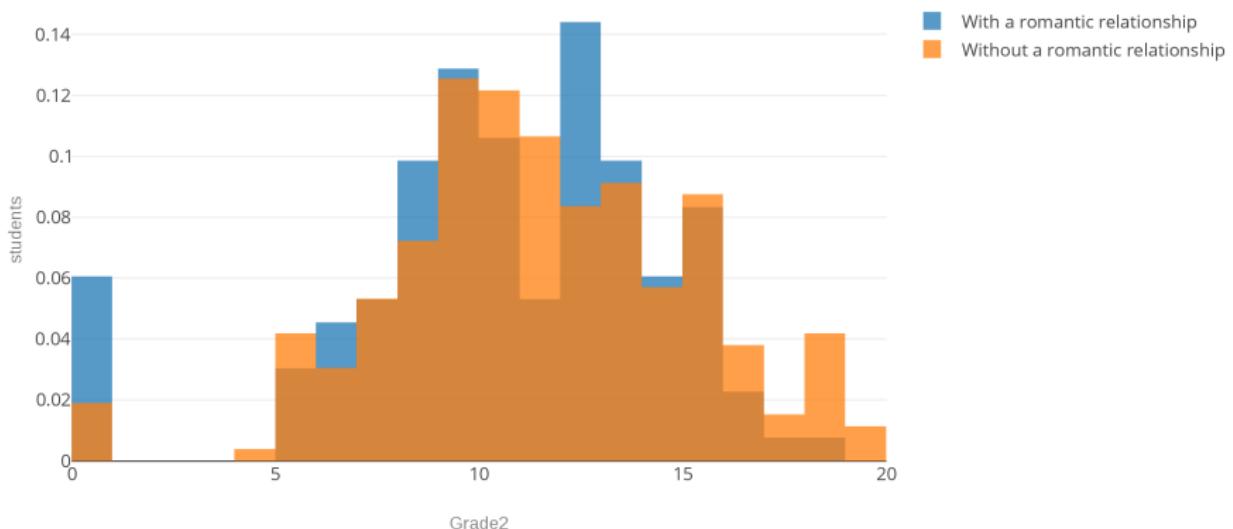
Люди, которые меньше всего готовятся — больше пьют и у них средняя оценка по экзаменам.

Почему люди, которые готовятся больше по времени стали хуже учиться?

Влияние отношений на учебу.



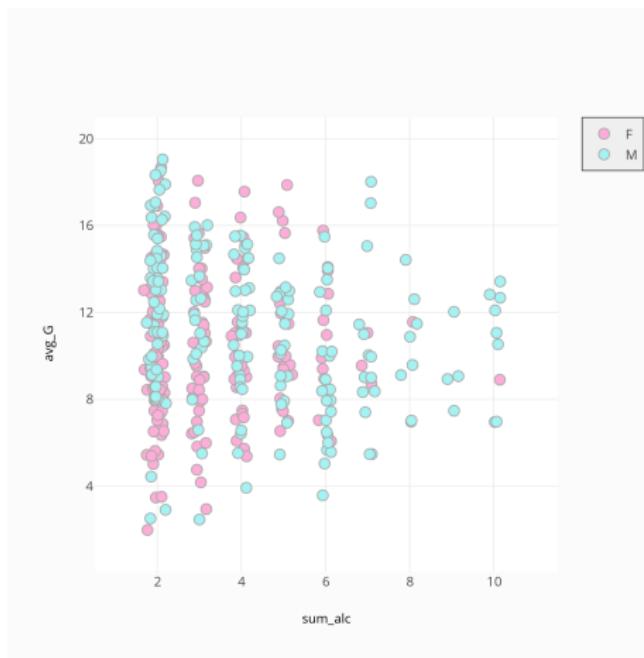
Влияние отношений на учебу.



Количество сдающих на 0 баллов увеличивается.

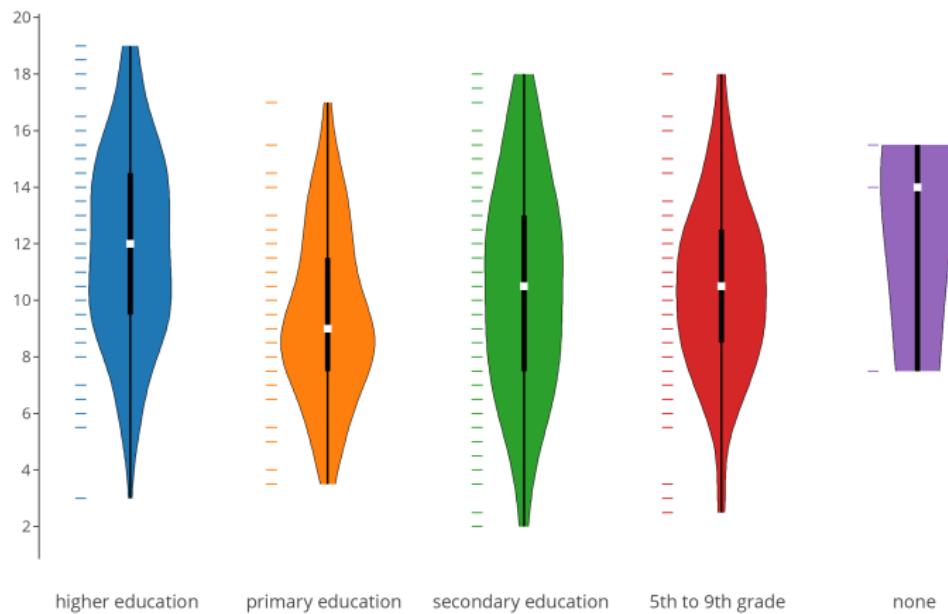
Пол и средняя оценка.

Введены признаки: суммарное количество алкоголя за неделю, средняя оценка за два зачета.

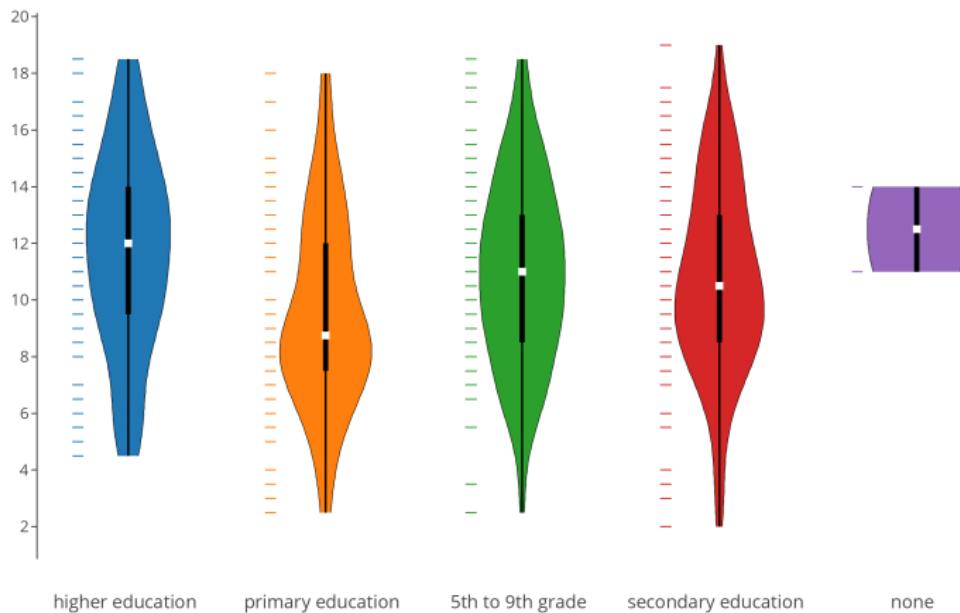


Больше пьющих мальчиков, чем девочек.

Relation between mother's education and average marks



Relation between father's education and average marks



Медиана средней оценки смещена вверх у тех, у кого мама или пapa имеют высшее образование.

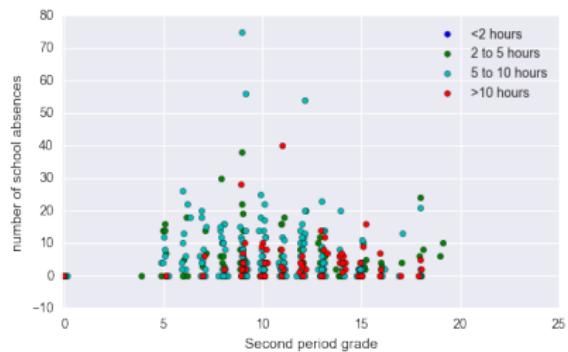
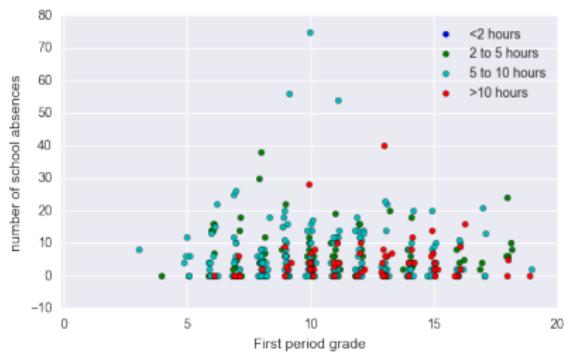
	Medu	Fedu	sum_alc	G1	G2	G3
127	0	1	2	7	8	9
249	0	2	6	13	15	15
324	0	2	5	16	15	15

Мама без образования, но у папы есть!

	Fedu	Medu	sum_alc	G1	G2	G3
76	0	4	2	11	11	10
171	0	1	2	13	15	16

Папа без образования, но у мамы есть!

Прогульщики и количество времени за домашними работами.



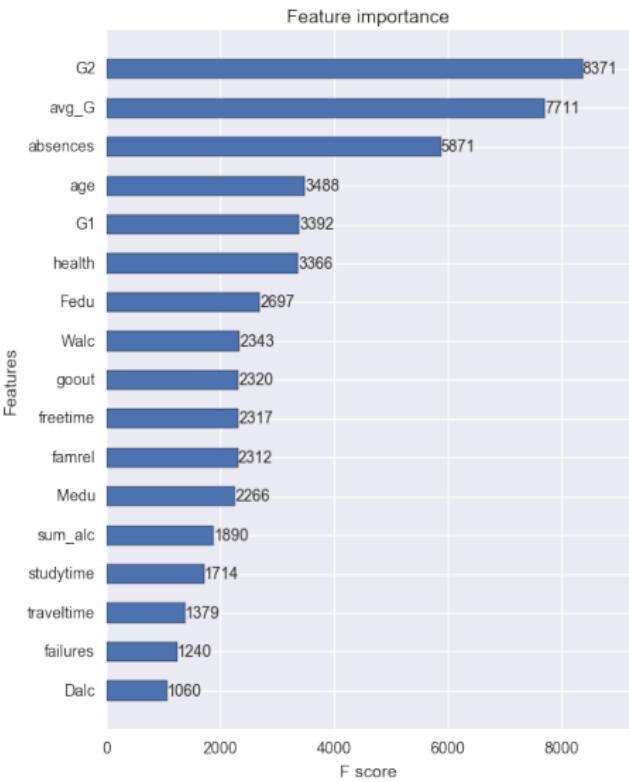
	romantic	absences	G1	G2	G3
276	yes	75	10	9	9

Наибольшее количество пропусков.

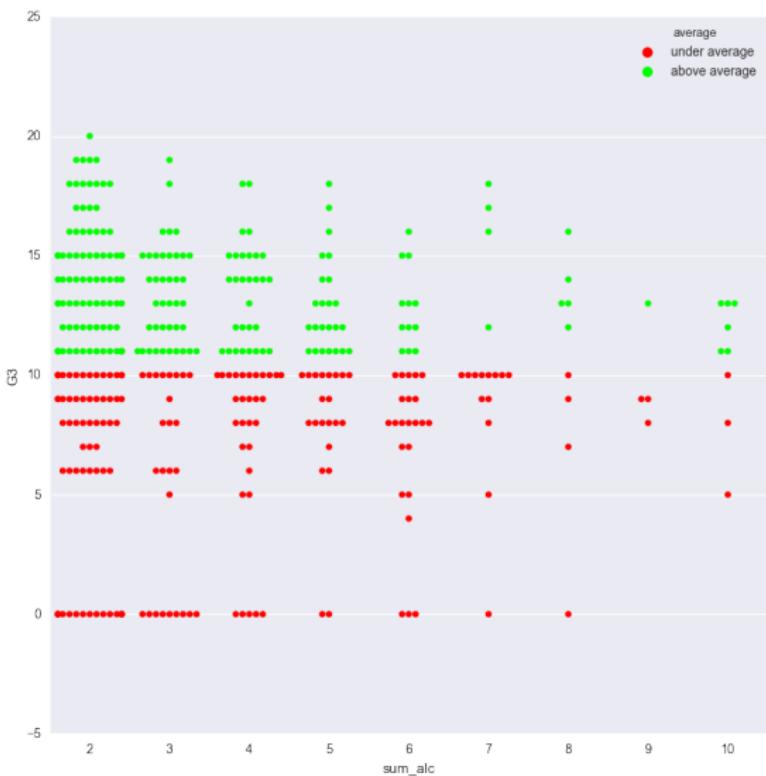
	romantic	absences	G1	G2	G3
130	yes	0	12	0	0
135	yes	0	11	0	0
244	yes	0	7	0	0

Без пропусков, однако ухудшили свои оценки.

Важность признаков по мнению xgboost.



Зависимость между суммой выпитого алкоголя и финальной оценкой G3.



По мере увеличения потребления алкоголя количество людей с оценкой выше среднего уменьшается.