

DAN 

Формула стресса

Команда №5 "Абонент"



Исследовательский вопрос:
Что в большей степени влияет на показатель стресса?

Гипотеза

- Увеличенная академическая нагрузка приводит к меньшей успеваемости

Механизм:

- Большая нагрузка ведет к большому стрессу

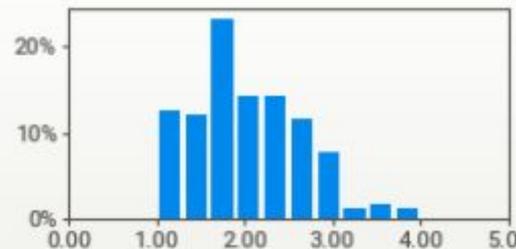
Dashboard

gpa

VALUES: 713 (46%)
MISSING: 846 (54%)
DISTINCT: 57 (4%)
ZEREOES: —

MAX 4.00
95% 3.00
Q3 2.40
AVG 2.03
MEDIAN 2.00
Q1 1.60
5% 1.20
MIN 1.00

RANGE 3.00
IQR 0.800
STD 0.607
VAR 0.369
KURT. 9.10e-4
SKEW 0.606
SUM 1,446

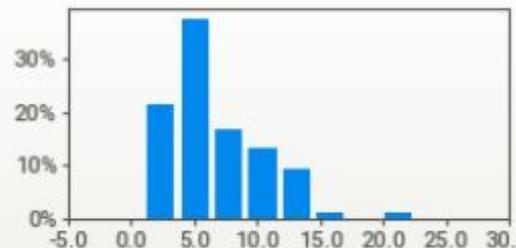


semester

VALUES: 1,300 (83%)
MISSING: 259 (17%)
DISTINCT: 22 (1%)
ZEREOES: —

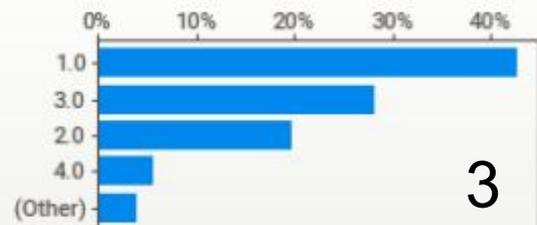
MAX 28.0
95% 13.0
Q3 9.0
AVG 6.5
MEDIAN 6.0
Q1 4.0
5% 2.0
MIN 1.0

RANGE 27.0
IQR 5.00
STD 3.88
VAR 15.1
KURT. 1.39
SKEW 0.963
SUM 8,490



degree

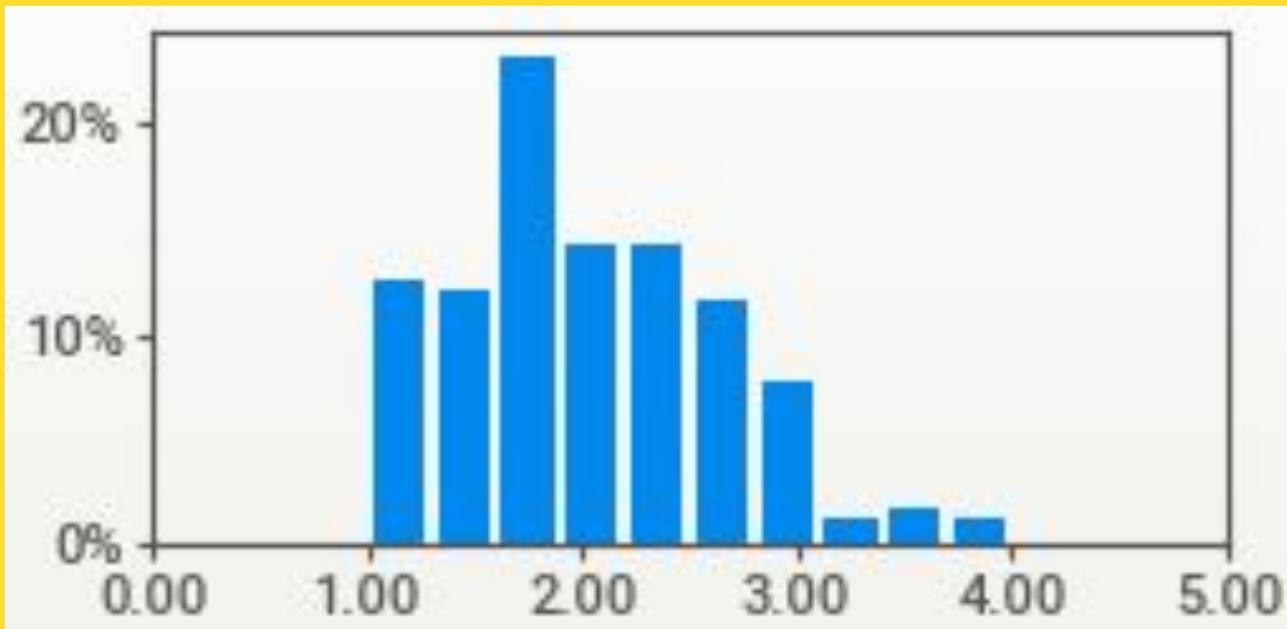
VALUES: 1,314 (84%)
MISSING: 245 (16%)
DISTINCT: 6 (<1%)



Изначальный GPA

VALUES: 713 (46%)
MISSING: 846 (54%)
DISTINCT: 57 (4%)

AVG = 2.03
MED = 2
Q1 = 1.6
Q3 = 2.4
SKEW = 0.6



Изначальный Week1

VALUES:	1,342	(86%)
MISSING:	217	(14%)
DISTINCT:	69	(4%)
ZEROES:	87	(6%)

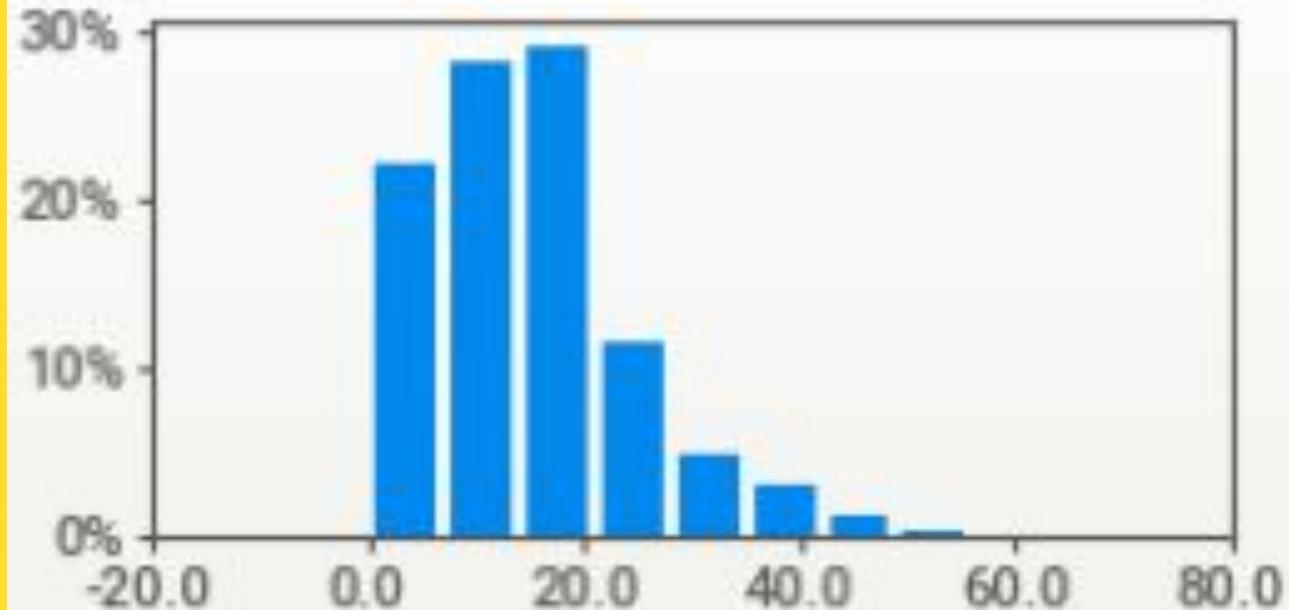
AVG = 14.6

MED = 13.8

Q1 = 8

Q3 = 20

SKEW = 0.995



Изначальный age

VALUES: 1,307 (84%)
MISSING: 252 (16%)
DISTINCT: 33 (2%)

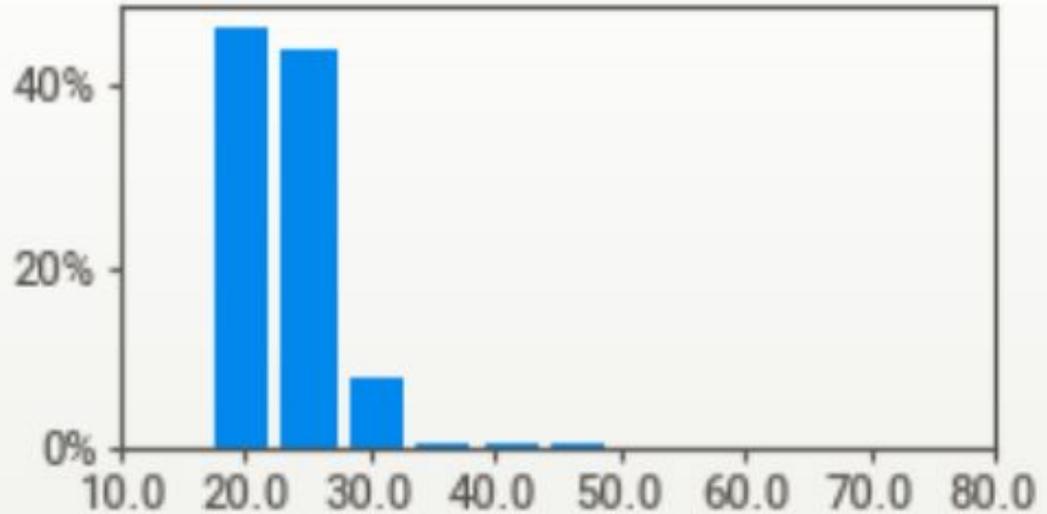
AVG = 23.4

MED = 23

Q1 = 21

Q3 = 25

SKEW = 3.71



8

Сон

Удовлетворенность
проведенным временем

degree

Age

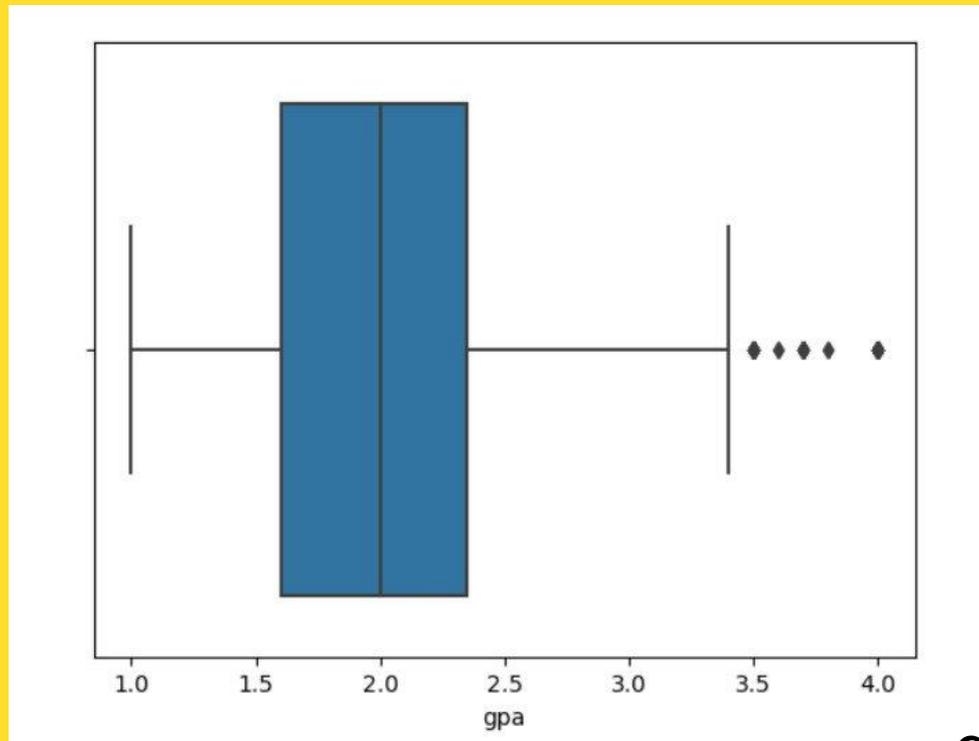


Преоброессинг

Заполнение пустых значений функцией Backfill.
Удаление выбросов используя BoxPlot..

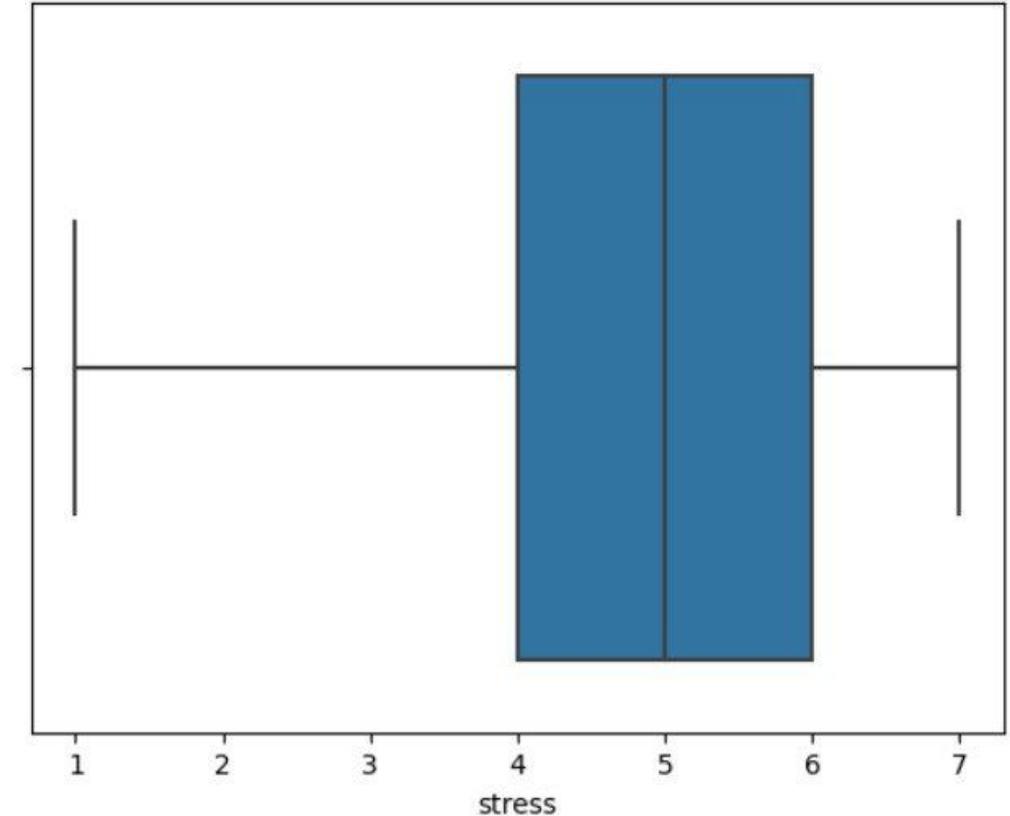
Редактируем данные переменных:

- week1
- week2
- stress
- gpa

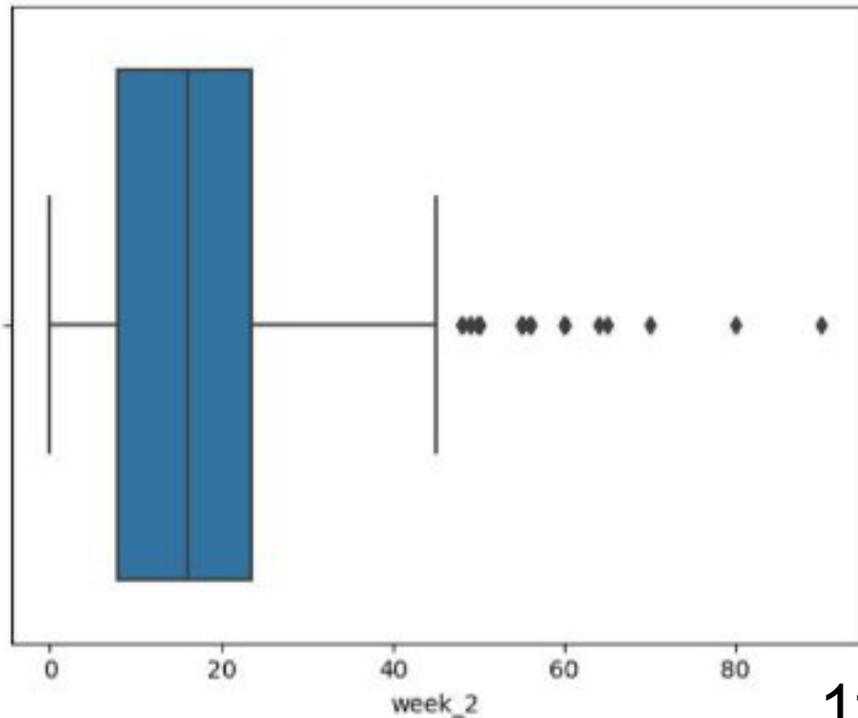
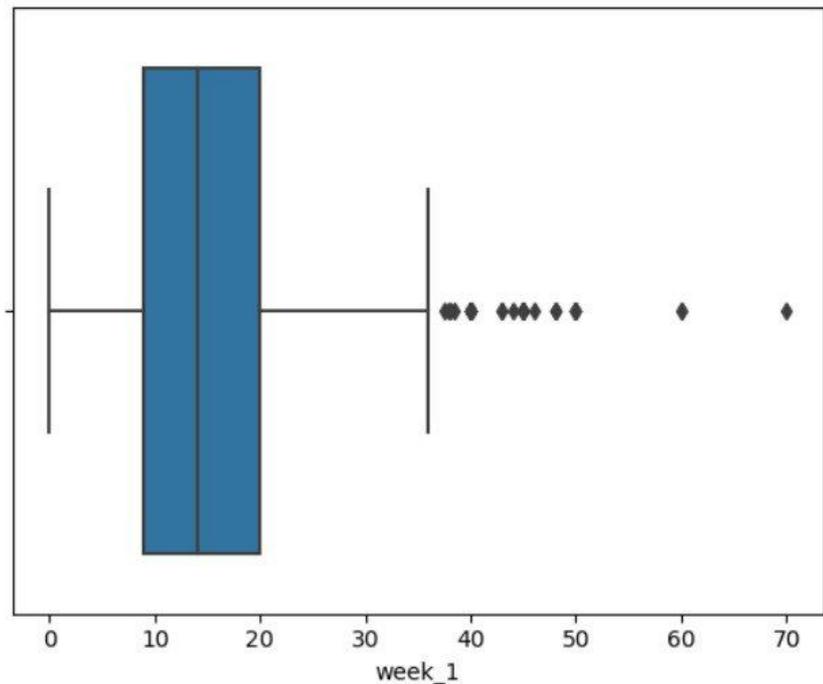


Препроцессинг

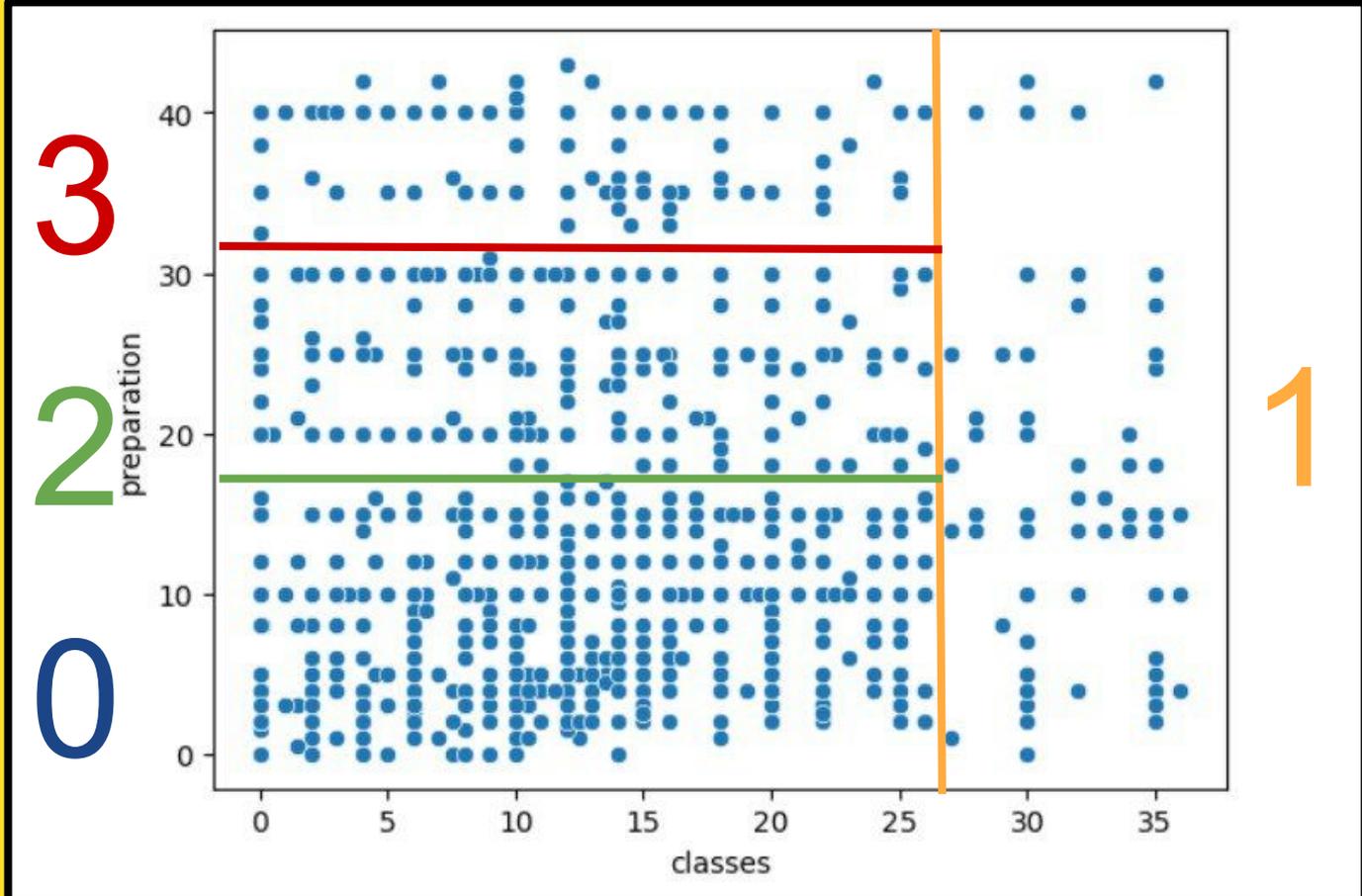
Метод
полуторного
межквартильного
размаха



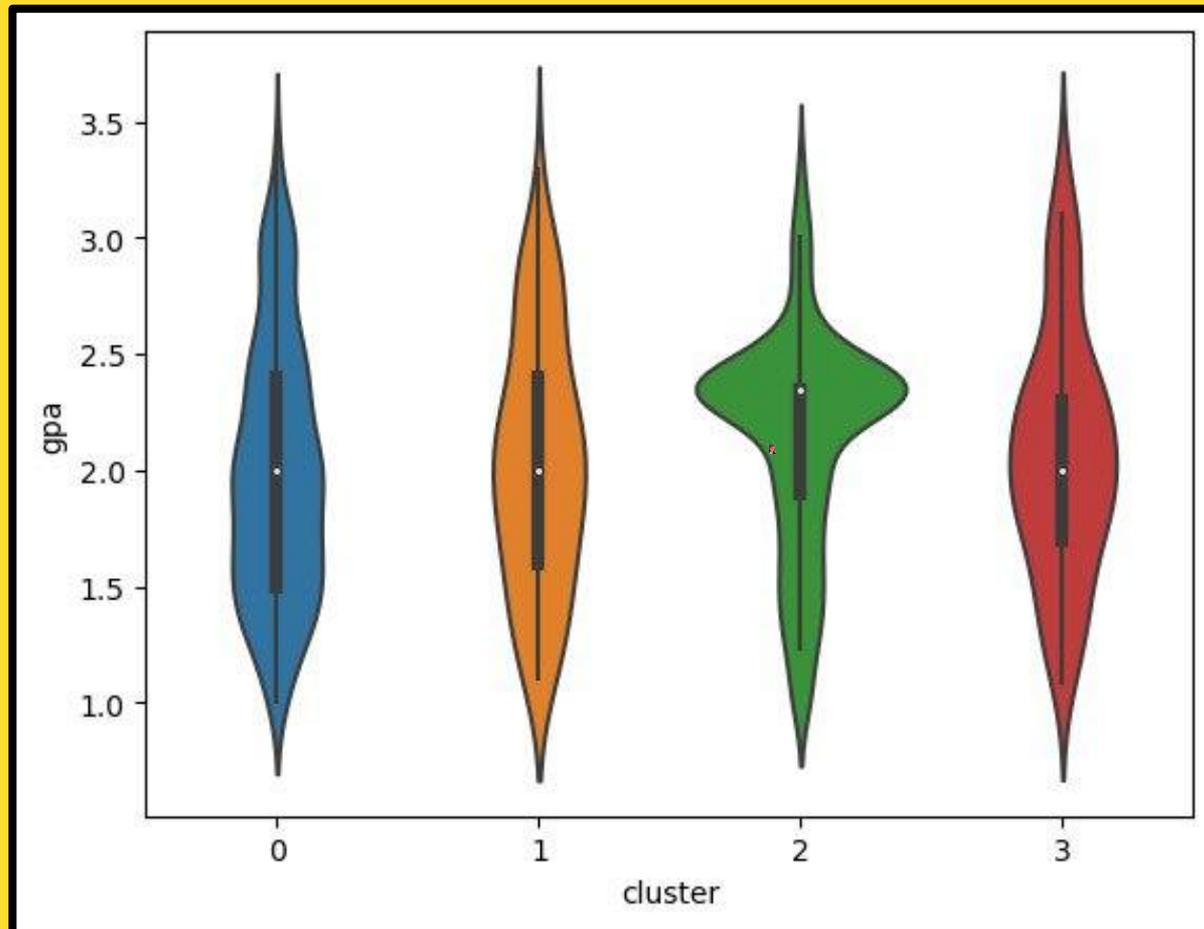
Препроцессинг



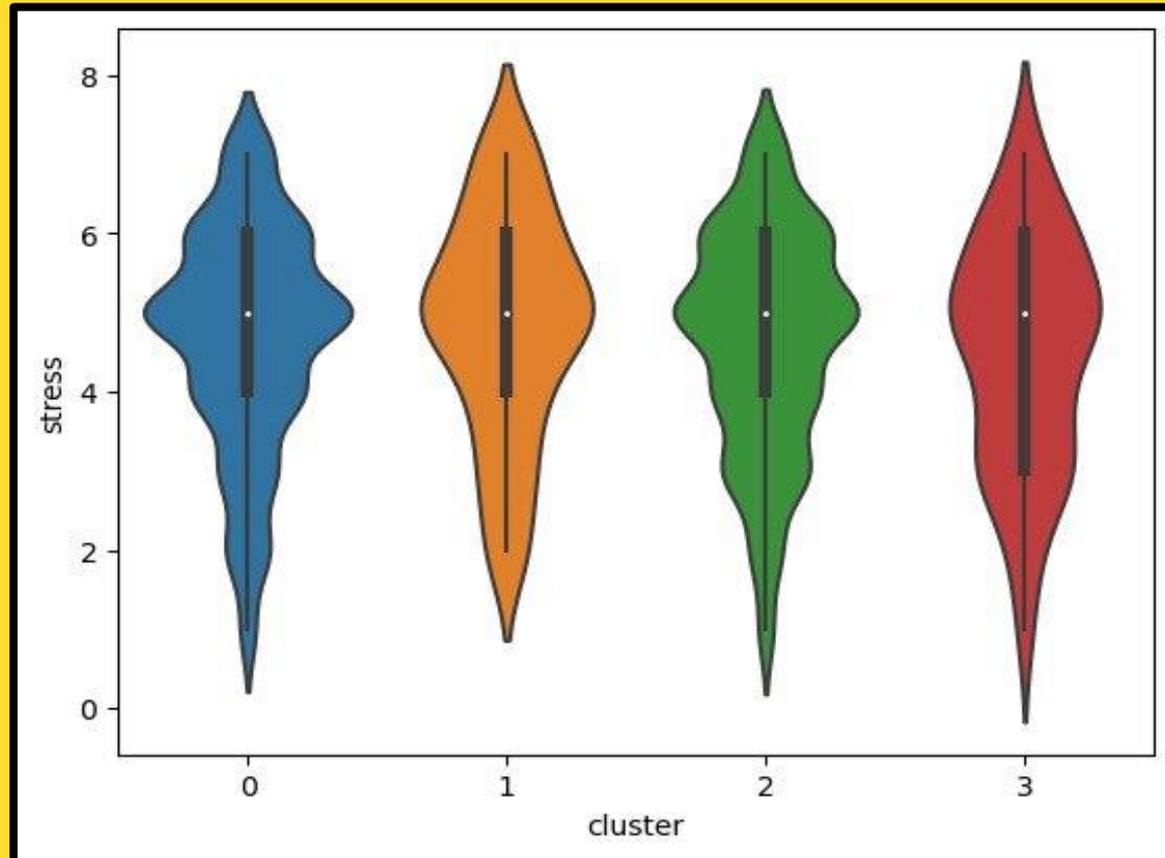
Деление на кластеры



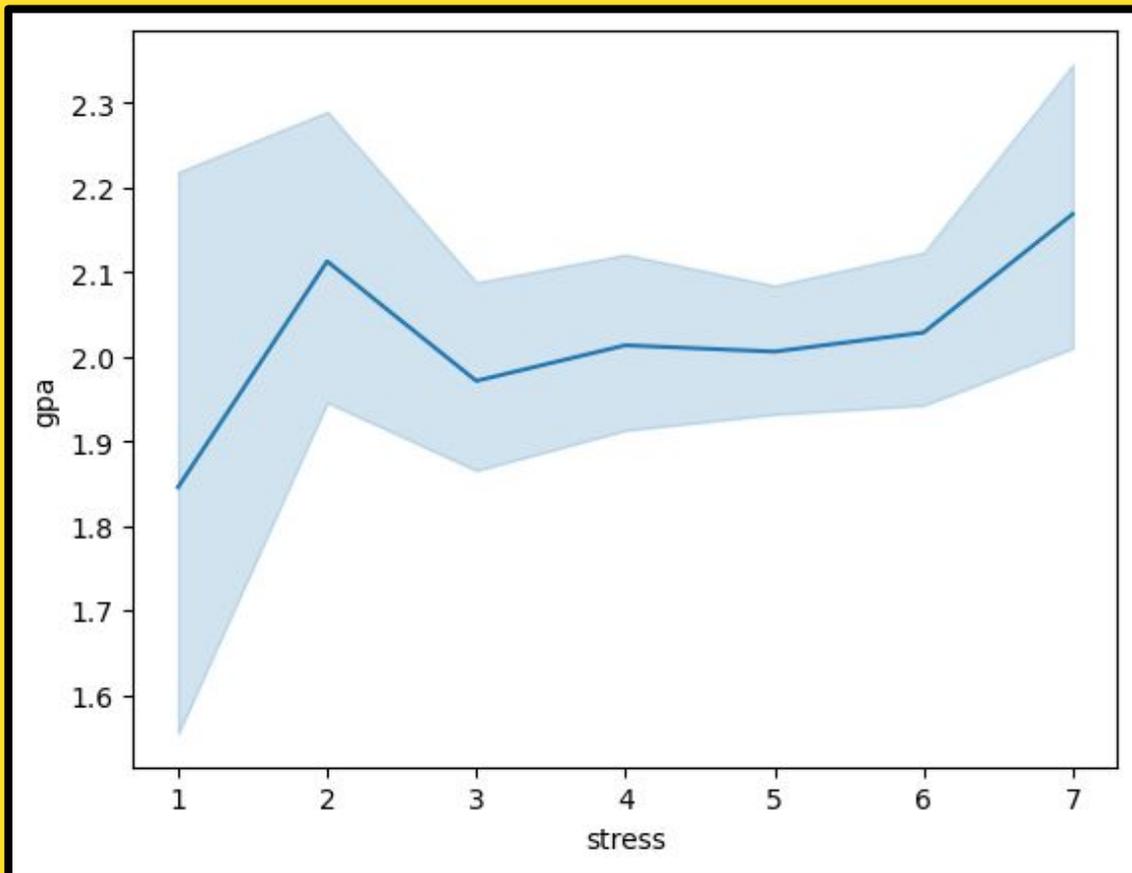
Зависимость GPA от Week1 и Week2



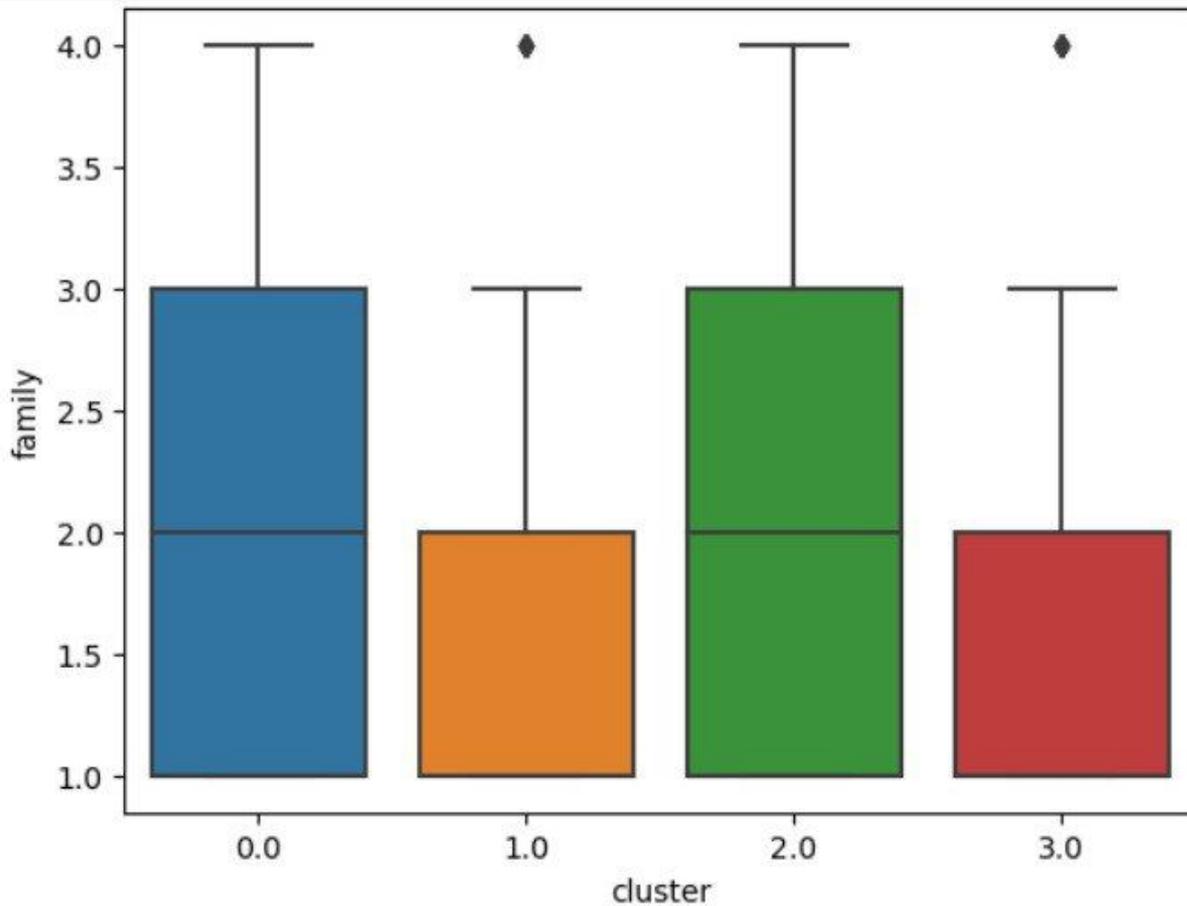
Зависимость Stress от Week_1 и Week_2



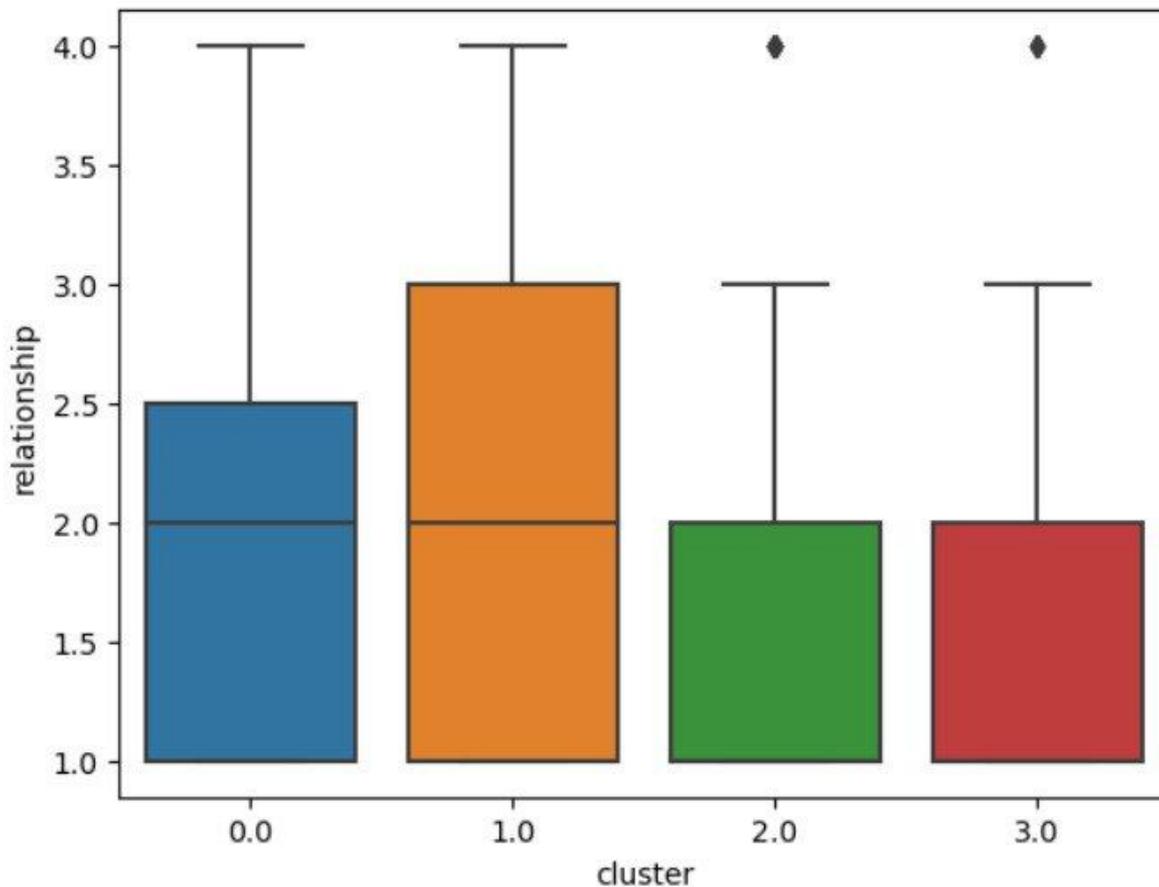
Точка перелома в графике успеваемости от стресса



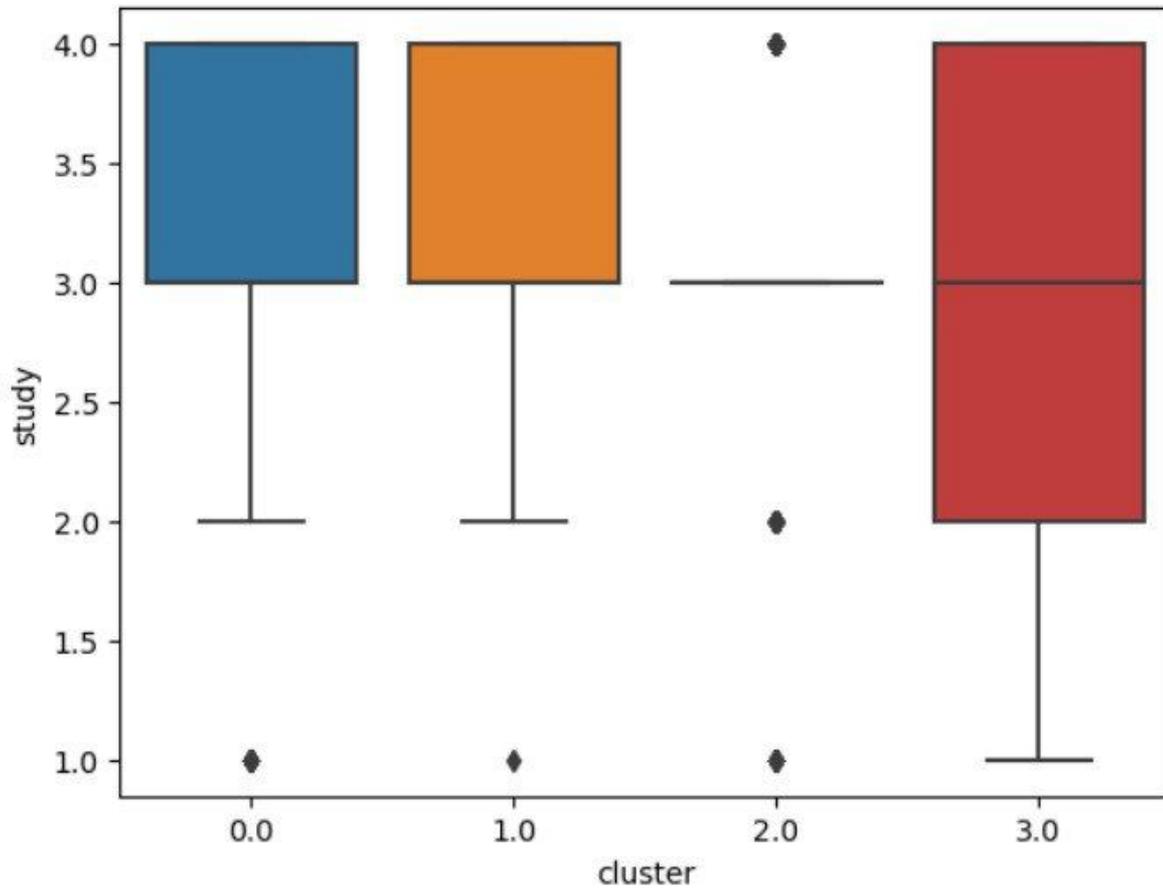
Источники стресса у разных категорий студентов



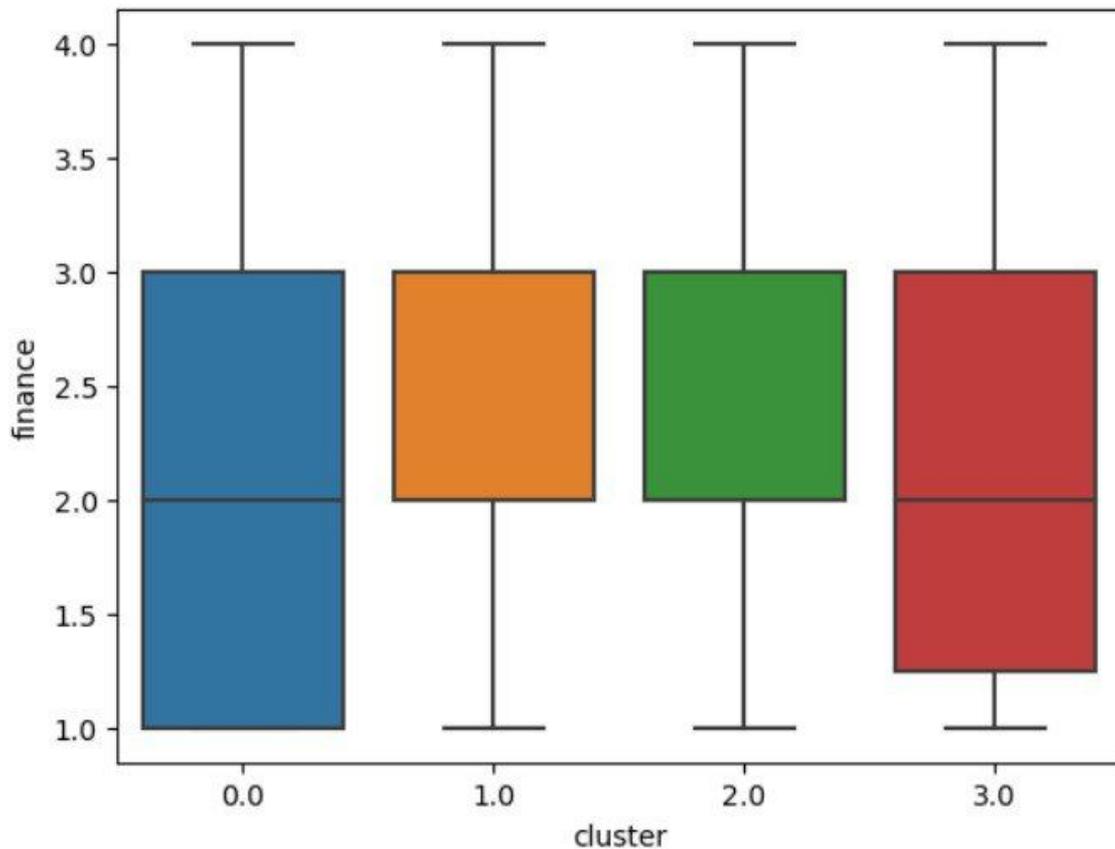
Источники стресса у разных категорий студентов



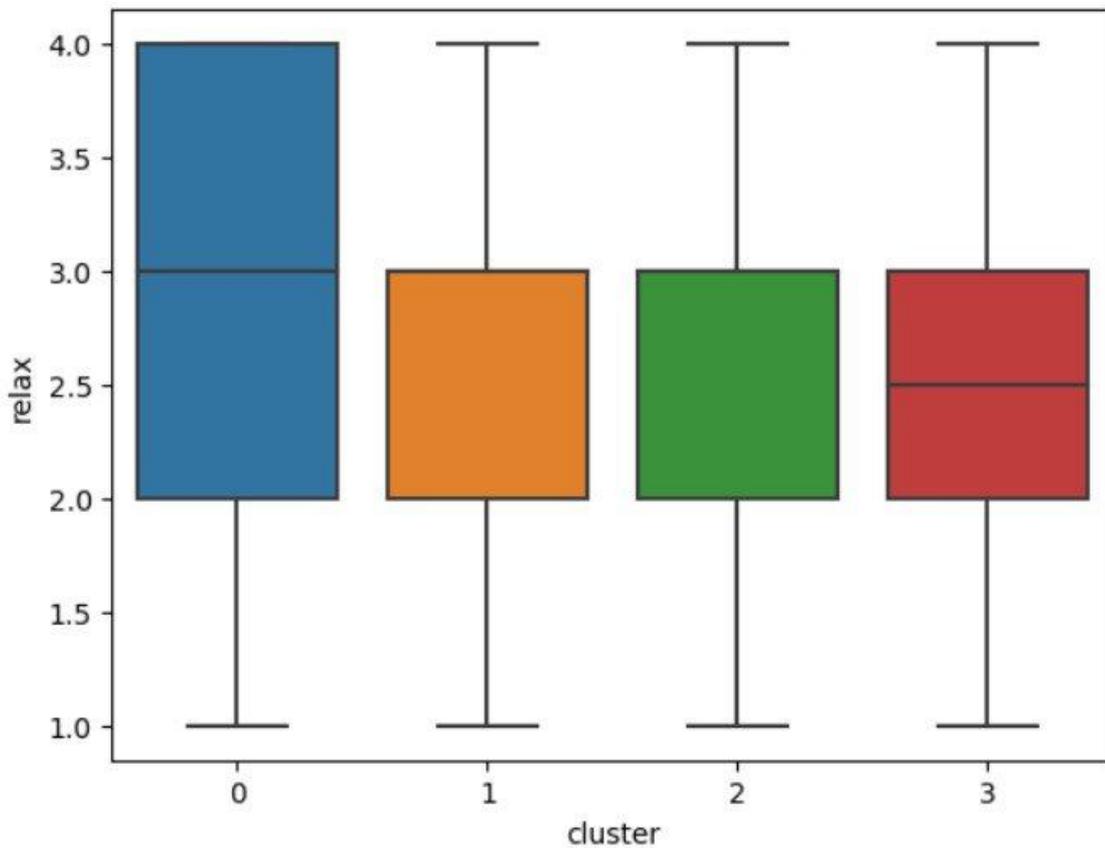
Источники стресса у разных категорий студентов



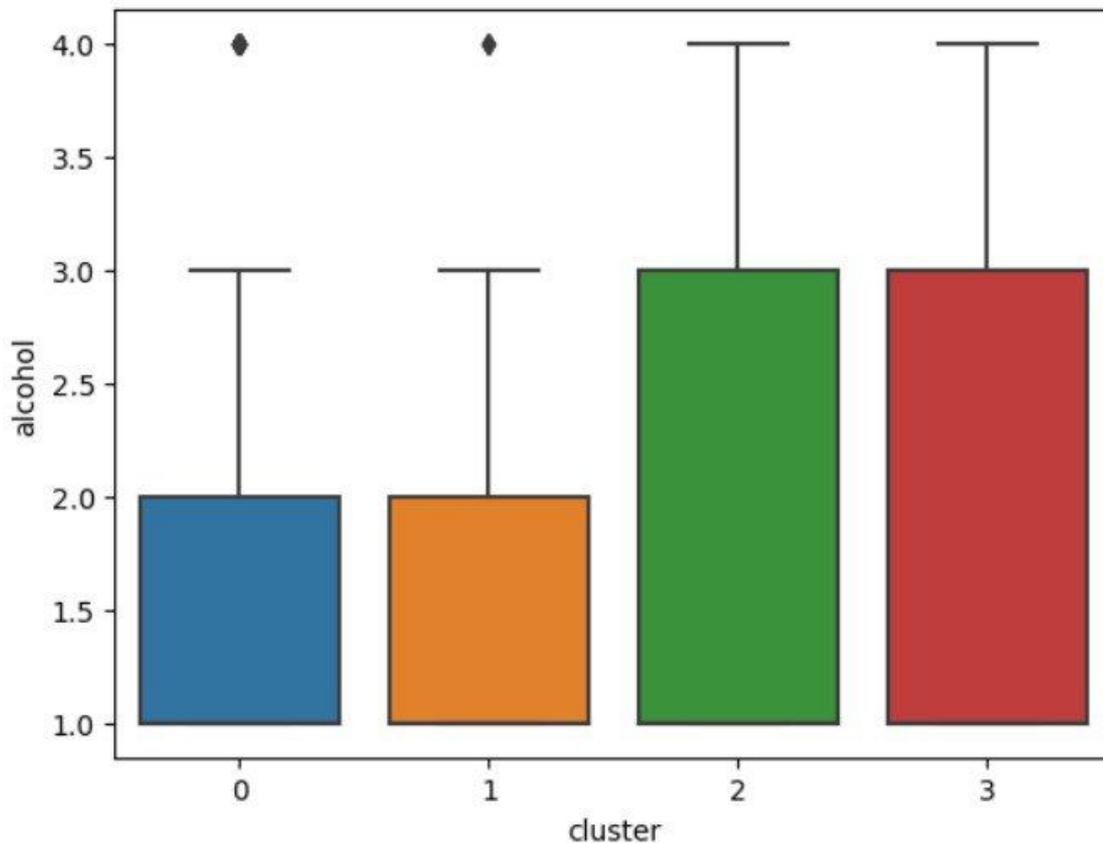
Источники стресса у разных категорий студентов



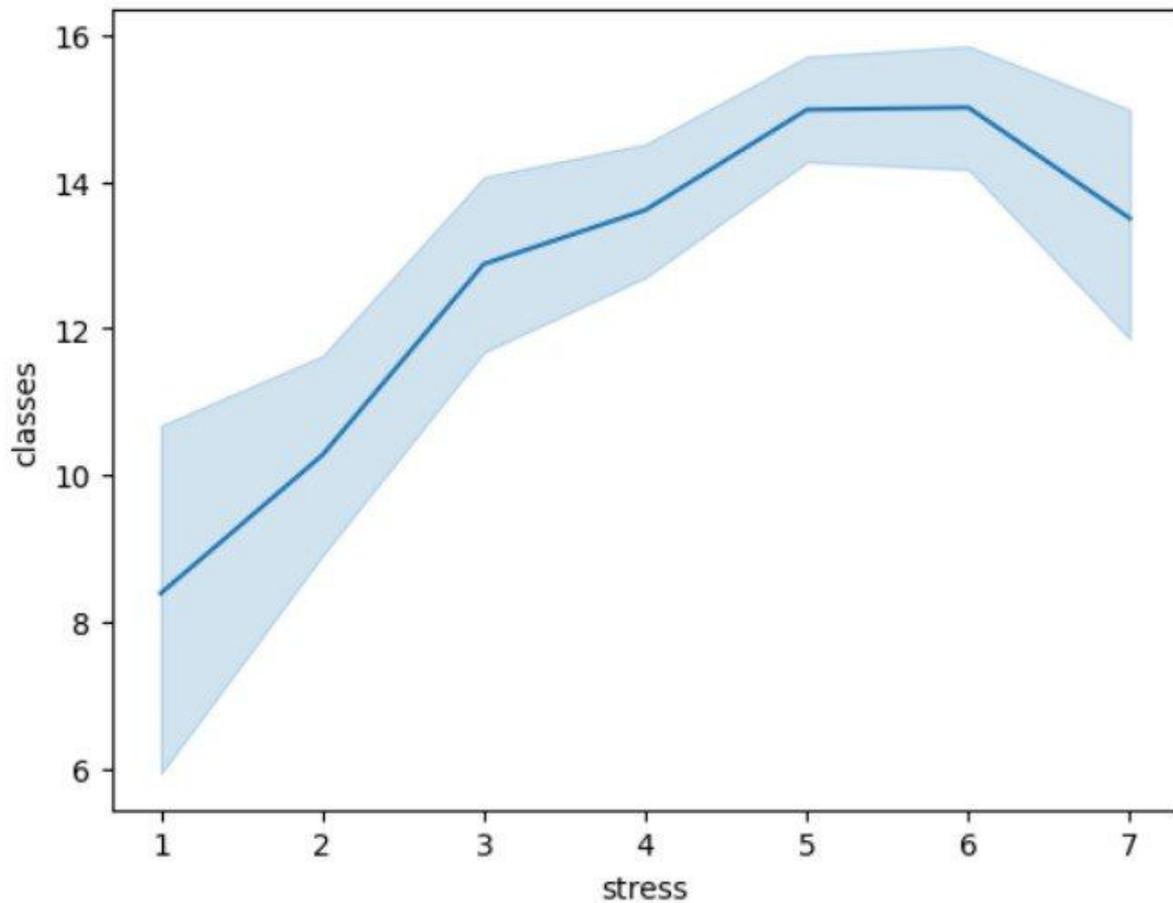
Способы отвлечься у разных категорий студентов



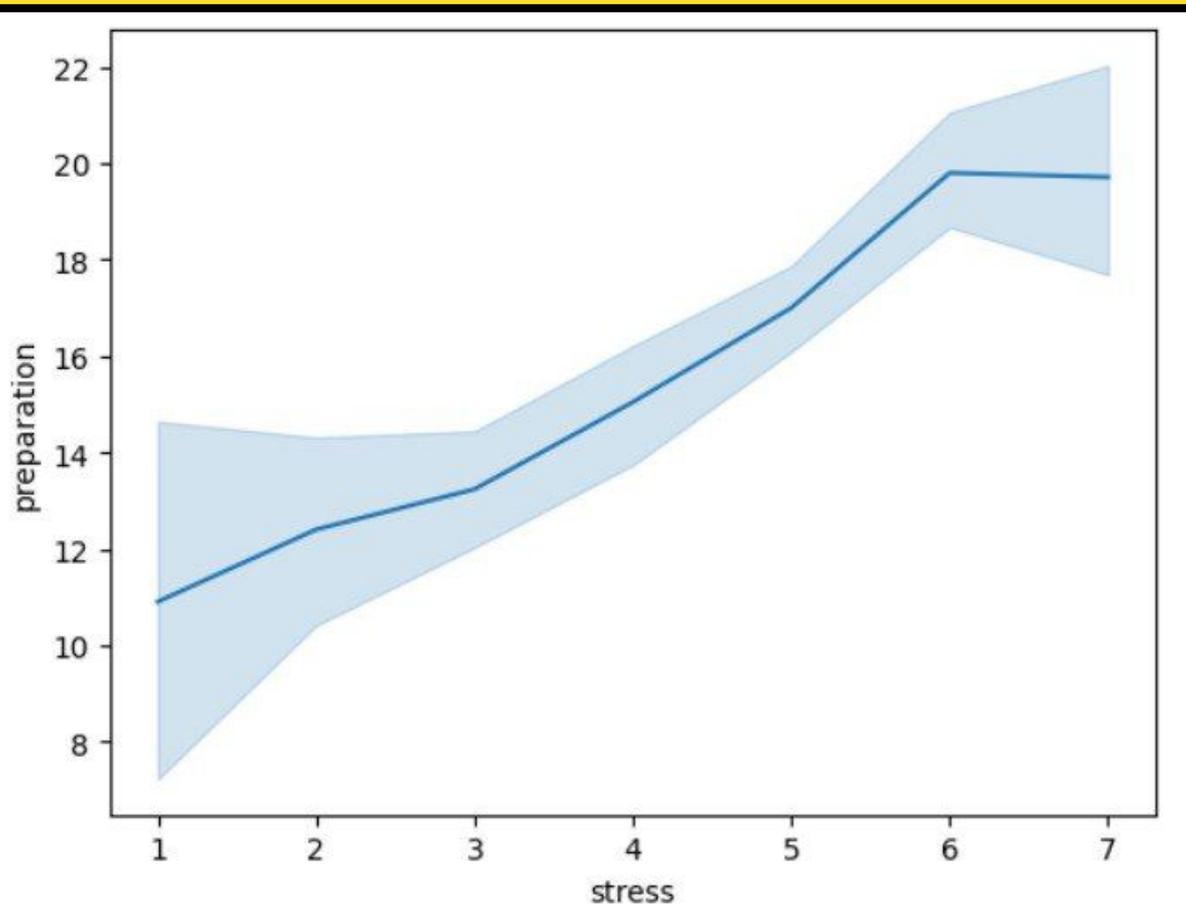
Способы отвлечься у разных категорий студентов



Посещение занятий в зависимости от стресса



Домашняя подготовка в зависимости от стресса



Обучение математической модели

Модель классификации - логистическая регрессия, поскольку нам необходимо определить уровень стресса.

```
clf = LogisticRegression().fit(X_train, y_train)
```

```
clf.score(X_test, y_test)
```

```
0.317016317016317
```

Связав средний балл и нагрузку, мы получили точность модели 0.31

Итоги

1. Мы хотели выяснить первопричины стресса у студентов, но получили низкую точность исследования. Это можно связать с тем, что факторов стресса у студентов больше чем было представлено в датасете
2. Большую точность исследования можно получить с динамическими данными, когда один студент переходит с одного курса на другой, а так же с большим кол-вом переменных, например - заработная плата и помощь родителей

Практическая польза

Получив влияющие на уровень стресса переменные (сон, время учебы и тд), нужно будет стараться избегать их для уменьшения стресса и повышения успеваемости

Рекомендации для студентов

1. Рекомендуем спать большее кол-во часов в неделю.
2. Найти для себя баланс между учебой в вузе и домашней подготовкой.
3. Спокойнее относитесь к своей успеваемости, не дайте плохим эмоциям повисить ваш уровень стресса.
4. Уделяйте больше времени хобби и досугу.

Спасибо за внимание!

Использованные инструменты:

python

excel

jupyter notebook

pandas

seaborn

sweetviz

LogisticRegression, sklearn

Доска с маркерами