**Задача проекта:**

Отследить влияние условий жизни учащихся в возрасте от 15 до 22 лет на их успеваемость, чтобы определиться с параметрами будущей модели, которая предсказывала бы результаты госэкзамена по математике для каждого ученика школы.

**Этапы работы над проектом:**

* Необходимо проверить гипотезу о связи различных переменных с успеваемостью по математике 'scores'
* Проверить качество данных и очистить их (пустые строки, выбросы);
* Построить гистограмму распределений для количественных переменных
* Провести корреляционный анализ и анализ номинативных переменных
* Определить параметры для дальнейшего построения модели

**Информация о данных:**

1. school — аббревиатура школы, в которой учится ученик
2. sex — пол ученика ('F' - женский, 'M' - мужской)
3. age — возраст ученика (от 15 до 22)
4. address — тип адреса ученика ('U' - городской, 'R' - за городом)
5. famsize — размер семьи('LE3' <= 3, 'GT3' >3)
6. Pstatus — статус совместного жилья родителей ('T' - живут вместе 'A' - раздельно)
7. Medu — образование матери (0 - нет, 1 - 4 клаcса, 2 - 5-9 классы, 3 - среднее специальное или 11 классов, 4 - высшее)
8. Fedu — образование отца (0 - нет, 1 - 4 класса, 2 - 5-9 классы, 3 - среднее специальное или 11 классов, 4 - высшее)
9. Mjob — работа матери ('teacher' - учитель, 'health' - сфера здравоохранения, 'services' - гос служба, 'at\_home' - не работает, 'other' - другое)
10. Fjob — работа отца ('teacher' - учитель, 'health' - сфера здравоохранения, 'services' - гос служба, 'at\_home' - не работает, 'other' - другое)
11. reason — причина выбора школы ('home' - близость к дому, 'reputation' - репутация школы, 'course' - образовательная программа, 'other' - другое)
12. guardian — опекун ('mother' - мать, 'father' - отец, 'other' - другое)
13. traveltime — время в пути до школы (1 - <15 мин., 2 - 15-30 мин., 3 - 30-60 мин., 4 - >60 мин.)
14. studytime — время на учёбу помимо школы в неделю (1 - <2 часов, 2 - 2-5 часов, 3 - 10 часов, 4 - >10 часов)
15. failures — количество внеучебных неудач (n, если 1<=n<=3, иначе 0)
16. schoolsup — дополнительная образовательная поддержка (yes или no)
17. famsup — семейная образовательная поддержка (yes или no)
18. paid — дополнительные платные занятия по математике (yes или no)
19. activities — дополнительные внеучебные занятия (yes или no)
20. studytime, granular - ?
21. nursery — посещал детский сад (yes или no)
22. higher — хочет получить высшее образование (yes или no)
23. internet — наличие интернета дома (yes или no)
24. romantic — в романтических отношениях (yes или no)
25. famrel — семейные отношения (от 1 - очень плохо до 5 - очень хорошо)
26. freetime — свободное время после школы (от 1 - очень мало до 5 - очень много)
27. goout — проведение времени с друзьями (от 1 - очень мало до 5 - очень много)
28. health — текущее состояние здоровья (от 1 - очень плохо до 5 - очень хорошо)
29. absences — количество пропущенных занятий
30. таргетная величина:
31. score — баллы по госэкзамену по математике

**Ответы на вопросы саморефлексии:**

1. Какова была ваша роль в команде?  
   Проект выполнялся самостоятельно.
2. Какой частью своей работы вы остались особенно довольны?  
   Анализом числовых признаков.
3. Что не получилось сделать так, как хотелось? Над чем ещё стоит поработать?  
   Заполнение пропусков.
4. Что интересного и полезного вы узнали в этом модуле?  
   Статистические методы, в общем раздел со статистикой был крайне полезен.
5. Что является вашим главным результатом при прохождении этого проекта?  
   Получилось сделать выводы, отобрать признаки для построения модели и обосновать этот выбор.
6. Планируете ли вы менять стратегию изучения последующих модулей?  
   Лучше конспектировать теорию, вести библиотеку.

\*Дополнительно: оставьте комментарий по своей работе для ментора.

Огорчает обратная связь и поддержка менторов в этом модуле, особенно в разделе статистики.