

Проект 2. Разведывательный анализ данных.

Цель проекта:

Цель проекта проследить влияние условий жизни учащихся на их успеваемость по математике, чтобы на ранней стадии выявлять студентов, находящихся в группе риска.

Задачи проекта:

Задача данного проекта состоит в том, чтобы проанализировать имеющиеся данные и найти и посчитать количество пустных строк, удалить выбросы и аномалии в данных. После этого необходимо провести корреляционный анализ и предложить данные, которые будут использованы для дальнейшего построения модели.

Информация о данных:

Данные представлены в Эксель таблице и содержат информацию о 395 респондентах -учениках от 15 до 22 лет, их родителях, школе, жизненной ситуации и оценках по математике.

Данные содержали мало выбросов и аномалий и могут считаться достаточно чистыми.

- 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 класса, 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 - 5-10 часов, 4 - >10 часов)
- 15 failures — количество внеучебных неудач (n, если $1 \leq n \leq 3$, 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 nursery — посещал детский сад (yes или no)
- 21 higher — хочет получить высшее образование (yes или no)
- 22 internet — наличие интернета дома (yes или no)
- 23 romantic — в романтических отношениях (yes или no)
- 24 famrel — семейные отношения (от 1 - очень плохо до 5 - очень хорошо)
- 25 freetime — свободное время после школы (от 1 - очень мало до 5 - очень много)
- 26 goout — проведение времени с друзьями (от 1 - очень мало до 5 - очень много)
- 27 health — текущее состояние здоровья (от 1 - очень плохо до 5 - очень хорошо)
- 28 absences — количество пропущенных занятий
- 29 score — баллы по госэкзамену по математике

Этапы работы:

1. Очистка и анализ данных
2. Обработка пустых значений и удаление аномальных значений
3. Промежуточные выводы
4. Анализ корреляций числовых значений
5. Анализ номинативных значений
6. Выводы и предложение по использованию данных в модели

Саморефлексия:

1. Какова была ваша роль в команде?
Моя задача была проанализировать данных на предмет выбросов и пустых значений, провести корреляционный анализ и сформировать предложение по дальнейшему использованию данных
2. Какой частью своей работы вы остались особенно довольны?
Корреляционным анализом числовых переменных
3. Что не получилось сделать так, как хотелось? Над чем ещё стоит поработать?
Хотелось бы использовать больше функций для нахождения и изменения значений и отклонений в данных
4. Что интересного и полезного вы узнали в этом модуле?
Мне было интересно узнать про статистические понятия и формулы
5. Что является вашим главным результатом при прохождении этого проекта?
Сдача проекта
6. Какие навыки вы уже можете применить в текущей деятельности?
Пока никаких
7. Планируете ли вы дополнительно изучать материалы по теме проекта?
Я планирую пересмотреть модуль с самого начала и прочитать книгу по статистике.