

D A N



ВЫСШАЯ ШКОЛА
ЭКОНОМИКИ



ТИНЬКОФФ



Уральский
федеральный
университет

ЗДРАВСТВУЙТЕ, УВАЖАЕМЫЕ УЧАСТНИКИ!

Сегодня мы предлагаем вам поближе познакомиться с миром анализа данных и поработать над реальной задачей в команде с такими же юными исследователями.

Исследование — это сложный процесс, поэтому для того, чтобы вы могли справиться со всеми возникающими трудностями и вопросами, к каждой команде будет прикреплен ментор.



Кто такой ментор?

Это человек, который имеет опыт работы над исследовательским проектом, не раз сталкивался со всеми трудностями при решении аналитических задач, а теперь готов помочь вам сделать качественный проект и победить! Не стесняйтесь обсуждать с ментором любые свои идеи, предлагать ему различные способы решения поставленной задачи и спрашивать совета в ходе работы. Каждый ментор много знает и умеет, не стоит игнорировать его помощь, однако не надейтесь, что он сделает всю работу за вас! :)

ВВЕДЕНИЕ

Обязательное школьное образование в Португалии состоит из трёхциклового (1-4 классы, 5-6 классы, 7-9 классы) начального образования и средней школы, обучение в которой длится 3 года. Обычно школьники начинают обучение в школе в возрасте 6-ти лет, в 15 переходя в старшие классы, где обучаются до достижения 18-ти лет. Разумеется, встречаются и исключения.

Успеваемость студентов оценивается три раза за учебный год. В конце последней трети учебного года студенты получают итоговые оценки по изучаемым в этом году дисциплинам. Успеваемость оценивается по двадцатибалльной шкале от 0 (самая низкая оценка из возможных) до 20 (наивысший балл). При этом оценки ниже 10 считаются неудовлетворительными.

Данные содержат результаты опроса 370 учеников двух школ в Португалии: Габриель Перейра (Gabriel Pereira, GP) и Моузино да Силвейра (Mousinho da Silveira, MS). Опрос проводился в 2005-2006 учебном году. Оценки по математике и португальскому языку и данные о пропусках предоставлены школами.

Описание переменных

1. school — школа ('GP' — Gabriel Pereira или 'MS' — Mousinho da Silveira)
2. sex — пол школьника ('F' — женский или 'M' — мужской)
3. age — возраст школьника (число: от 15 до 22)
4. address — адрес проживания школьника ('U' — городской или 'R' — сельская местность)
5. famsize — размер семьи ('LE3' — меньше или равен 3 чел. или 'GT3' — больше 3-х чел.)
6. Pstatus — совместное проживание родителей ('T' — вместе или 'A' — отдельно)
7. Medu — образование матери (число: 0 — нет, 1—4 класса, 2 — 5-9 классов, 3 — средняя школа, 4 — высшее образование)
8. Fedu — образование отца (число: 0 — нет, 1 — 4 класса, 2 — 5-9 классов, 3 — средняя школа, 4 — высшее образование)
9. Mjob — работа матери (категория: 'teacher' — учитель, 'health' — здравоохранение, 'services' гос. служба (вкл. полицию), 'at_home' или 'other')

10. Fjob — работа отца (категория: 'teacher' — учитель, 'health' — здравоохранение, 'services' гос. служба (вкл. полицию), 'at_home' или 'other')
11. reason — выбор школы (категория: 'home' — близость к дому, 'reputation' — репутация, 'course' — обр. программа или 'other')
12. studytime — учёба в неделю (порядков.: 1 — <2 часов, 2 — 2-5 часов, 3 — 5-10 часов, или 4 — >10 часов)
13. failures_MATH и failures_PORTUGESE — кол-во несдач дисциплины в прошлом (число: n если от 1 до 3-х, в противном случае 4)
14. schoolsup — помогают учиться в школе (yes или no)
15. famsup — помогают учиться дома (yes или no)
16. paid_MATH и paid_PORTUGESE — платные занятия с репетитором по дисциплине (yes или no)
17. activities — внеклассная активность (yes или no)
18. nursery — посещал детский сад (пре-школа) (yes или no)
19. higher — хочет получать высшее обр. (yes или no)
20. internet — есть интернет дома (yes или no)
21. romantic — в романтических отношениях (yes или no)
22. famrel — оценка взаимоотношений в семье (число: от 1 — очень плохие до 5 — отличные)
23. freetime — свободное время после учёбы (число: от 1 — оч. мало до 5 — оч. много)
24. goout — время с друзьями (число: от 1 — оч. мало до 5 — оч. много)
25. health — оценка здоровья (число: от 1 — оч. плохое до 5 — оч. хорошее)
26. absences_MATH и absences_PORTUGESE — число пропущенных уроков в течение года по дисциплинам (число: от 0 до 93)
27. finalGrade_MATH и finalGrade_PORTUGESE — итоговая оценка по дисциплинам (число: от 0 до 20)

ЗАДАНИЕ

Вам предлагается поработать над ответом на следующий исследовательский вопрос: как использование интернета связано с успеваемостью школьника по математике и языкам?

Для ответа на данный вопрос предлагается выполнить следующие шаги:

1. Объясните, стоит ли ожидать, что использование интернета школьниками как-либо связано с их успеваемостью? Для всех ли дисциплин ожидается такой эффект? Почему?
2. Ознакомьтесь с предложенным набором данных. Какие переменные можно отнести к категориальным, а какие — к количественным?
3. Что можно сказать о выборке школьников, с которой вы работаете? Подготовьте для имеющегося набора данных описательные статистики: распределения категориальных переменных; сводные статистические характеристики количественных величин. Есть ли в данных пропущенные значения? Представьте описательные статистики в виде таблицы.
4. Какая доля учеников имеет доступ к интернету дома?
5. Используя коэффициент корреляции Пирсона и/или графическое представление данных, проверьте, наблюдается ли связь между переменными, отражающими успеваемость по математике и португальскому языку? Почему так?
6. Среди имеющихся переменных выберите те, которые характеризуют социально-демографический статус школьника.
7. Проанализировав частоты релевантных признаков и распределения переменных, как бы вы описали среднестатистического ученика каждой из школ? Насколько однородны эти подвыборки студентов по социально-демографическому статусу?
8. Проверьте, есть ли связь между наличием интернета дома у школьника и успеваемостью по математике и языку. Стоит ли ожидать, что у школьников разного соц. дем. статуса это влияние будет проявляться по-разному? Почему? Проверьте.
9. Какой вывод можно сделать? Проявилась ли взаимосвязь, которую вы ожидали изначально?

10. Были ли у вашего подхода недостатки? Как можно улучшить ваше исследование для более убедительного ответа на поставленный вопрос?
11. Если бы стояла цель повторить сбор данных для подобного исследования в 2022-м году, чтобы это исследование актуализировать, можно ли предлагать школьникам ответить на вопрос о наличии доступа к интернету дома? Почему? Как бы вы измеряли использование интернета сейчас?
12. Стоит ли ожидать других результатов, нежели в 2005 году? Почему?