

Лабораторная работа №2

Пакет pandas

Рассмотрим датасет с данными, собранными благодаря опросу студентов математического курса средней школы в Португалии (возраст - от 15 до 22 лет). ("**math_students.csv**"). Целевой переменной является итоговая оценка студента за курс.

Подробная расшифровка признаков:

- **school** - тип школы ("GP" - Gabriel Pereira или "MS" - Mousinho da Silveira)
- **sex** - пол ("F" - female или "M" - male)
- **age** - возраст (от 15 до 22)
- **address** - откуда студент ("U" - urban или "R" - rural)
- **famsize** - размер семьи ("LE3" - меньше или равно 3 или "GT3" - больше 3)
- **Pstatus** - в каких отношениях родители ("T" - живут вместе "A" - раздельно)
- **Medu** - образование матери (0 - никакого, 1 - начальное образование (4 класса), 2 - от 5 до 9 классов, 3 - среднеспециальное или 4 - высшее)
- **Fedu** - образование отца (0 - никакого, 1 - начальное образование (4 класса), 2 - от 5 до 9 классов, 3 - среднеспециальное или 4 - высшее)
- **Mjob** - работа матери ("teacher", "health" care related, civil "services" (e.g. administrative or police), "at_home" or "other")
- **Fjob** - работа отца ("teacher", "health" care related, civil "services" (e.g. administrative or police), "at_home" or "other")
- **reason** - причина выбора школы (близко к дому — "home", репутация школы — "reputation", предпочтение некоторым предметам - "course" или "other")
- **guardian** - опекун ("mother", "father" или "other")
- **traveltime** - время от дома до школы (1 - меньше 15 мин., 2 - 15 до 30 мин., 3 - 30 мин. до 1 часа, или 4 - больше 1 часа)
- **studytime** - количество часов обучения в неделю (1 - меньше 2 часов, 2 - от 2 до 5 часов, 3 - от 5 до 10 часов, или 4 - больше 10 часов)
- **failures** - количество ранее не сданных предметов (n if $1 \leq n < 3$, else 4)
- **schoolsup** - дополнительные занятия (yes or no)
- **famsup** - помощь от семьи при выполнении заданий (yes or no)
- **paid** - дополнительные платные занятия (yes or no)
- **activities** - внеклассная деятельность (yes or no)
- **nursery** - посещал детский сад (yes or no)
- **higher** - желание высшего образования (yes or no)
- **internet** - домашний интернет (yes or no)
- **romantic** - состоит в романтических отношениях (yes or no)
- **famrel** - насколько хороши отношения в семье (от 1 - очень плохие до 5 - превосходные)
- **freetime** - наличие свободного времени после школы (от 1 - очень мало до 5 - очень много)
- **goout** - гуляет с друзьями (от 1 - редко до 5 - очень часто)
- **Dalc** - употребление алкоголя в будние дни (от 1 - очень редко до 5 - очень часто)
- **Walc** - употребление алкоголя в выходные (от 1 - очень редко до 5 - очень часто)
- **health** - текущее состояние здоровья (от 1 - очень плохое до 5 - очень хорошее)
- **absences** - количество школьных пропусков (от 0 до 93)
- **G1** - оценка за первый семестр (от 0 до 20)
- **G2** - оценка за второй семестр (от 0 до 20)
- **G3** - итоговая оценка (от 0 до 20)

- 1) Загрузить данные (`read_csv()`)
- 2) Вывести первые и последние 10 строк таблицы (`head()`, `tail()`)
- 3) Вывести число объектов и их характеристик
- 4) Вывести названий всех колонок
- 5) Есть ли в данных пропуски
- 6) Вывести статистику по значениям признаков
- 7) Вывести более подробное описание значений признаков (количество непустых значений, типов столбцов и объема занимаемой памяти)
- 8) Какие значения принимает признак один из признаков (например, `guardian`)? (какие именно значения, сколько уникальных, сколько значений каждого типа)
- 9) Вывести только тех студентов, у которых опекуном является мать, и которая работает учителем или на дому:
- 10) Создать признак `"alc"`, который будет отражать общее употребление алкоголя в течение недели
- 11) Вывести новый размер, новые колонки
- 12) Какая причина выбора школы была самой частой? В качестве ответа приведите соответствующее значение признака.
- 13) Найдите количество студентов, у родителей которых нет никакого образования. Вывести информацию об этих студентах
- 14) Найдите минимальный возраст учащегося школы `Mousinho da Silveira`. Вывести информацию об этих студентах
- 15) Найдите количество студентов, имеющих нечетное число пропусков. Вывести информацию об этих студентах
- 16) Найдите разность между средними итоговыми оценками студентов, состоящих и не состоящих в романтических отношениях. В качестве ответа приведите число, округленное до двух значащих цифр после запятой.
- 17) Создать признак `"study_time_ratio"`, который будет отражать соотношение времени, затраченного на учебу (`study_time`), к общему времени, затраченному на учебу и досуг (`study_time + leisure_time`).
- 18) Вывести наиболее распространенное количество несданных предметов

19) Найдите количество студентов, чьи мать и отец работают. Вывести информацию об этих студентах.

20) Найдите максимальный возраст студентов, у которых оба родителя работают в сфере услуг (полиции). Вывести информацию об этих студентах.

21) Найдите количество студентов, имеющих оценку за первый семестр выше среднего балла. Вывести информацию об этих студентах.

22) Разделите студентов на две группы: те, у которых матери имеют высшее образование, и те, у которых матери не имеют высшего образования. Сравните средние итоговые оценки между этими двумя группами.