# Урок 24. Что такое данные и датасет?

7 класс · Урок 24 из 34 · 45 минут

Что такое датасет: урок для 7 класса о данных, строках и признаках таблицы, единицах измерения и создании набора ответов собственного опроса.

Число в таблице можно переписать без ошибки и всё равно понять неверно. Выясним, какие пояснения нужны к данным, и соберём ответы своего опроса так, чтобы другая группа смогла ими пользоваться.

## ▼ Содержание урока

1. Цель урока
2. Ход урока
3. Этап 1. Вход в тему: число без смысла
4. Этап 2. Как устроена таблица данных
5. Этап 3. Почему единицы меняют результат
6. Этап 4. Собираем данные, которые можно использовать
7. Этап 5. Итоги урока

## ▼ 01. Цель урока

Научиться собирать небольшой датасет так, чтобы другой человек мог правильно понять каждую запись.

К концу урока ученики смогут:

- Объяснять различие данных, их значения и человеческого знания.
- Различать объект наблюдения, строку и признак в таблице.
- Проверять единицы измерения и отличать пропуск от нуля.
- Создавать анонимный датасет реального группового опроса.

## ▼ 02. Ход урока

| Этап | Мин | Учитель | Ученики |
|---|---|---|---|
| 1. Вход в тему: число без смысла | 5 | Проводит опыт со своими числами, собирает разные трактовки. | Записывают число, отделяют запись от значения. |
| 2. Как устроена таблица данных | 10 | Объясняет строки, признаки и смысл значений. | Читают схему записи и различают типы данных. |
| 3. Почему единицы меняют результат | 9 | Объясняет перевод температур и разбирает историю единиц. | Сопоставляют число, единицу и правило пересчёта. |
| 4. Собираем данные, которые можно использовать | 15 | Организует добровольный опрос и обмен таблицами. | Создают датасет и исправляют неоднозначности описания. |
| 5. Итоги урока | 6 | Обсуждает общие выводы и ответы на итоговые вопросы. | Формулируют знания и объясняют ответы. |

## 03. Этап 1. Вход в тему: число без смысла 5 минут

### Говорим ученикам

Запишите на листе одно число, которое относится к вашему сегодняшнему дню, но пока не подписывайте его. Сосед предлагает две разные трактовки. Затем вы открываете смысл: минуты дороги, номер страницы или что-то другое. Почему одного числа оказалось недостаточно?

Данные — это зафиксированные сведения, с которыми можно работать: числа, текст, изображения, звук. Сама запись не объясняет, что измеряли. Наш опыт показывает: нужно знать, что записали, в каких единицах и при каких условиях. Сегодня выясним, что такое датасет и как сделать его понятным другим.

## 04. Этап 2. Как устроена таблица данных 10 минут

### Говорим ученикам

Датасет — это организованный набор данных для определённой задачи. Он может быть таблицей, коллекцией изображений с описаниями или набором звуковых записей. Форма разная, а требование одно: должно быть понятно, что входит в набор и как связаны его части.

В таблице строка обычно соответствует одному объекту или одному наблюдению. Если мы собираем ответы опроса, одна строка — один ответ. Столбцы — признаки, то есть характеристики, которые фиксируем одинаковым способом для всех строк. Название столбца и описание помогают понять его значения.

Числа, категории и тексты ведут себя по-разному. Минуты можно складывать; категории «бумажная книга» и «экран» — разные варианты, но их номера нельзя усреднять как длительность. Идентификатор записи помогает различать строки, а сам по себе не описывает важное свойство ответа.

Записи, которые получает компьютер, отличаются от знаний человека. Человек понимает вопрос, опирается на опыт и умеет уточнить смысл. Программа получает подготовленное представление. Если мы неверно подписали столбец, сложное вычисление может дать точный ответ на неверно заданный вопрос.

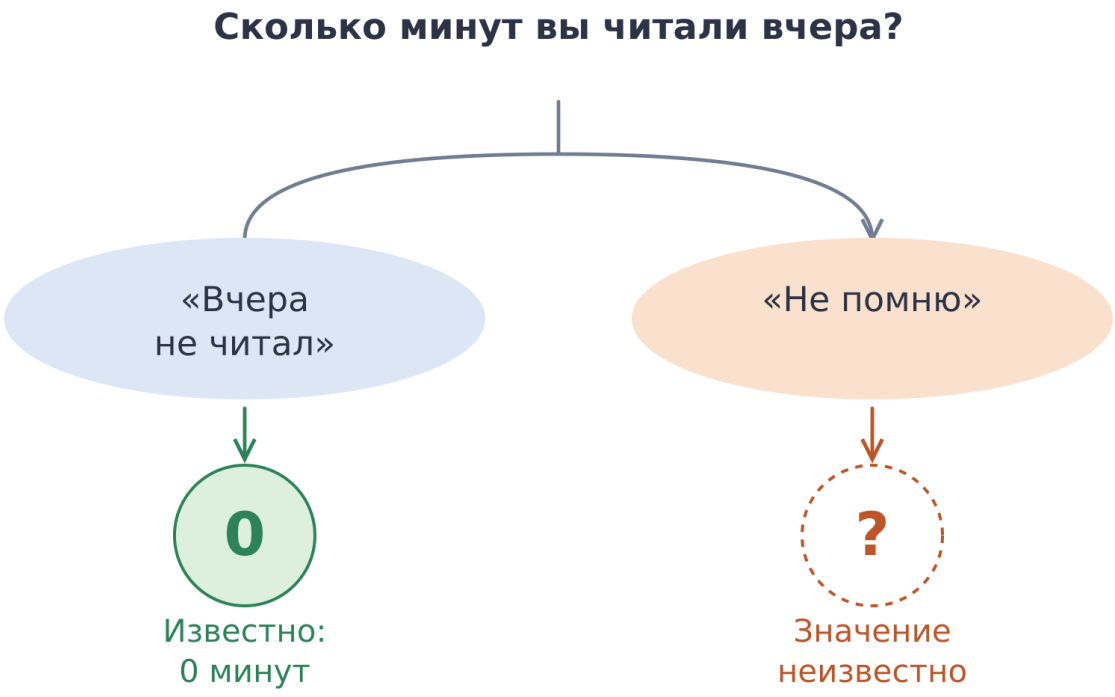
## 05. Этап 3. Почему единицы меняют результат

9 минут

### Говорим ученикам

Число 68 в столбце температуры не обязательно означает жару. В шкале Фаренгейта это та же температура, что 20 градусов Цельсия. Для перевода используют формулу C = (F − 32) × 5 / 9. Вычитаем 32, затем меняем масштаб. Одного переименования заголовка недостаточно: нужно пересчитать значения.

В одном столбце нельзя смешивать значения в разных единицах без пояснения и пересчёта. Если объединяем наборы, сначала читаем их описание. Пропуск тоже требует отдельного правила: ноль градусов — измерение, а пустое поле — отсутствие результата. Эти состояния говорят о разном.




Ноль — известная длительность, а пропуск означает, что длительность неизвестна. Заменив пропуск нулём, мы добавим в расчёт ответ, которого участник не сообщал.

Ошибка в смысле данных может возникнуть до всякого машинного обучения. Датасет нужен не ради аккуратной таблицы: он должен сохранять сведения так, чтобы расчёт относился к реальному вопросу.

### Реальный случай: аппарат потеряли из-за разных единиц

30 сентября 1999 года NASA JPL сообщила о предварительном расследовании потери аппарата Mars Climate Orbiter у Марса. Одна команда использовала английские единицы, а другая — метрические; несогласованность не обнаружили вовремя.

Это история обмена инженерными данными, а не пример работы нейросети. Она показывает общий принцип: число без согласованной единицы может привести к неверному действию, даже если передано без опечатки.

**Для учителя**

Короткая устная проверка: 32 °F = 0 °C, 50 °F = 10 °C. Различайте перевод шкал и сравнение записей. Объяснение не требует знания физики орбит; не переносите на урок неподтверждённые технические подробности происшествия.

## 06. Этап 4. Собираем данные, которые можно использовать 15 минут

### Говорим ученикам

Хороший опрос начинается с вопроса, ради которого ответы потом сравнят. Фраза «Сколько вы читаете?» допускает слишком разные толкования. Вопрос с одинаковым периодом и единицей помогает получить сопоставимые ответы. Мы не проверяем, кто «хорошо читает»: пробуем создать небольшой набор данных о привычках группы.

Каждый участник может не отвечать. Не запрашиваем имена, контакты и сведения, которые не нужны для исследования. Если ответ неизвестен, сохраняем пропуск. Маленькая группа даёт сведения только о своих участниках, а не обо всех семиклассниках.

### Задание. Опрос без потерянного смысла

Группа готовит небольшой опрос, полезный ей для выбора формы общего чтения или обсуждения: когда удобнее читать и какой формат комфортнее. Результат — таблица и понятное описание её столбцов.

1. Выберите один вопрос с числовым ответом и один с ответом-категорией. Уточните период, единицу и варианты категории.
2. Спросите добровольных участников группы. Одна строка — один анонимный ответ; присвойте номер записи.
3. Соберите таблицу в редакторе или на общем листе. Рядом опишите объект строки, смысл столбцов и обозначение пропуска.
4. Поменяйтесь таблицами с соседней группой. Без устных подсказок определите, что означает одна выбранная запись. Отметьте неоднозначность и исправьте описание.

Критерии: ответы сопоставимы; единицы названы; пропуски не заменены нулями; другой группе понятна таблица. Число участников не делает вывод представительным для всей школы.

**Для учителя**

Числовой вопрос может касаться минут чтения за вчерашний день, категориальный — предпочтительного формата. Это возможный вариант, не требуемая анкета. Если группа выберет другой период, он должен быть одинаковым для всех ответов. Не заставляйте придумывать числовой ответ вместо пропуска. Помогите ученикам внести ответы в таблицу. Сохраните их в том виде, в каком их дали участники.

## 07. Этап 5. Итоги урока 6 минут

### Говорим ученикам

Таблица полезна, когда сохраняет не только значения, но и их смысл. Датасет можно передать другому человеку или программе, если понятно, как получены записи, что означают признаки и какие ограничения есть у набора. На следующем уроке разберём, откуда берутся разные виды данных и почему больше — не всегда лучше.

### Главное за урок

- Датасет — организованный набор данных для определённой задачи.
- Строка таблицы описывает объект или наблюдение, а столбцы — его признаки.
- Значение нужно читать вместе с единицей, периодом и правилом получения.
- Пропуск и ноль означают разное; их смешение меняет выводы.

### Итоговые вопросы

Ответьте коротко. Затем обсудим объяснения двух-трёх участников.

1. Какие сведения нужны, чтобы понять число в незнакомой таблице?
2. Почему замена заголовка «°F» на «°C» не переводит температуру в другую шкалу?
3. Чем отличаются ответ «0 минут» и отсутствие ответа на вопрос о чтении?
4. Почему опрос одной группы не описывает автоматически всю школу?

**Для учителя**

1. Нужны смысл столбца, единица, объект наблюдения и условия сбора.

2. У шкал разные нулевые отметки и величина градуса; значения нужно пересчитать.
3. В первом случае участник сообщает об отсутствии чтения за заданный период; во втором данных нет.
4. Участники опроса могут отличаться от остальных. По ответам небольшой группы, которую оказалось проще опросить, трудно судить обо всех.

Принимаются ответы по смыслу. Важна связь с общим знанием, а не воспроизведение деталей реального случая.

**Для учителя: источники**

1. NASA JPL. Mars Climate Orbiter Team Finds Likely Cause of Loss. Дата материала: 30.09.1999. Материал старше 2024 года. Подтверждающая фраза: «one team used English units (e.g., inches, feet and pounds) while the other used metric units».