1. Различные способы обнаружения пропущенных данных включают:
   * Визуализация с использованием тепловых карт или графиков, показывающих распределение пропущенных значений.
   * Расчет общего количества пропущенных значений для каждого признака.
   * Использование методов, таких как **.isnull()** или **.isna()** для выявления пропущенных значений в каждой ячейке данных.
2. Тип данных каждого признака можно определить с помощью метода **.dtype** для каждого столбца в наборе данных. Например:

|  |
| --- |
| print(data.dtypes) |

1. Пример категориальных данных включает в себя данные, которые представляют собой определенные категории или классы, а не числовые значения. Например, типы автомобилей (седан, хэтчбек, кроссовер и т. д.), Цвета (красный, синий, зеленый) или типы продуктов (молоко, яйца, хлеб).
2. Категориальные данные можно закодировать различными способами:
   * One-Hot Encoding.
   * Label Encoding.
   * Ordinal Encoding.
   * Binary Encoding.
   * Frequency Encoding и другие.
3. One-Hot Encoding - это метод кодирования категориальных данных, который преобразует категориальные переменные в бинарные признаки, каждый из которых представляет отдельную категорию. Для каждой категории создается новый бинарный признак (столбец), в котором 1 обозначает принадлежность категории, а 0 - отсутствие принадлежности. Этот метод позволяет учесть информацию о принадлежности категории без введения порядка или предположений о числовых значениях.
4. Помимо пропусков и выбросов, ошибки данных могут включать:
   * Дубликаты записей.
   * Некорректные значения, например, отрицательные значения для неподходящих параметров.
   * Несогласованность данных между различными источниками или столбцами.
   * Неожиданные выбросы или аномалии, не связанные с реальными данными.
   * Некорректные типы данных для признаков.