

Фиктивное значение (Dummy value)

Синонимы: Фейковые данные, Недействительное значение, Мнимое значение

В анализе данных фиктивным является значение, которое помещается в ячейку таблицы в случае, когда фактическое значение отсутствует или не вызывает доверия. Например, если клиент забыл занести в анкету номер карточки социального страхования, то оператор, который переносит данные из анкеты в OLTP-систему, может ввести значение типа 999-999-999, если не указан возраст, то может быть введено значение 0, и т.д. Также фиктивное значение может быть указано, если запись неразборчивая.

При выборе фиктивного значения важно, чтобы другие пользователи не приняли его за реальное, т.е. оно должно быть «странным».

Наличие в наборах данных фиктивных значений может стать серьезной проблемой при их анализе, поскольку они могут оказаться аномальными, породить противоречия (например, два одинаковых номера карточки социального страхования), дубликаты и просто исказить сведения.

Поэтому в процессе очистки данных фиктивные значения должны обрабатываться. Например, они могут заменяться на наиболее вероятные значения, на медиану или среднее значение по столбцу. Если используемые при анализе алгоритмы могут корректно работать с пропусками, то фиктивные значения могут быть заменены на пустые.