Данные из нескольких таблиц

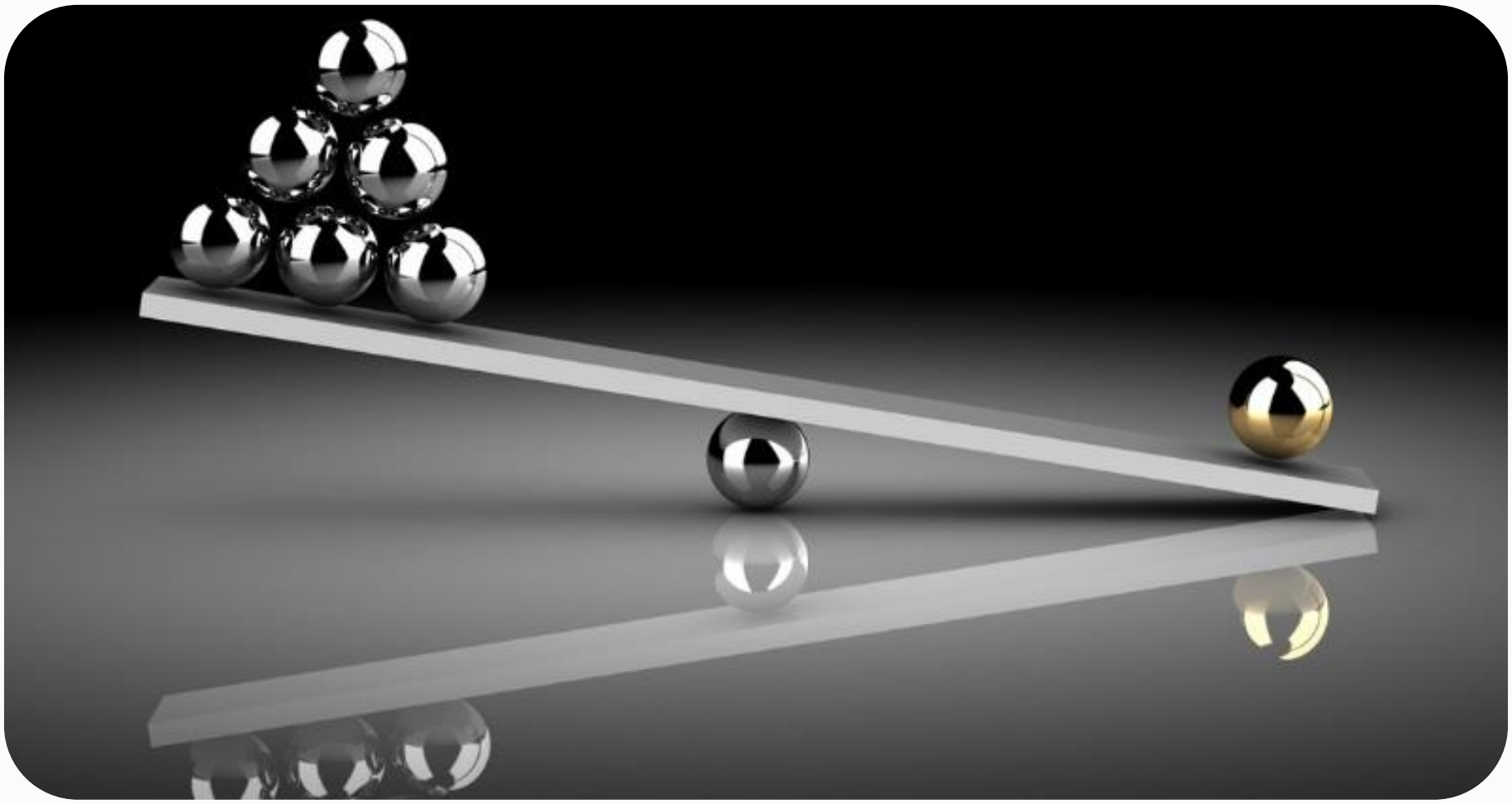
Семейный статус должников

married	secondary	22	no	no	yes
married	secondary	1466	yes	no	yes
married	tertiary	473	no	no	yes
married	primary	0	no	no	yes
married	secondary	521	yes	no	yes
married	primary	2103	yes	no	yes
married	secondary	986	no	yes	yes
married	secondary	-157	no	yes	yes
married	primary	512	yes	yes	yes
married	tertiary	1612	no	no	yes
married	secondary	-233	yes	no	yes
married	secondary	71	no	no	yes
married	unknown	1	no	no	yes
married	secondary	387	yes	no	yes
married	tertiary	0	yes	yes	yes
married	secondary	380	no	no	yes
married	secondary	-757	yes	yes	yes
married	secondary	23495	no	no	yes
married	tertiary	3634	no	no	yes
married	secondary	6	no	no	yes
married	secondary	2	yes	no	yes
married	primary	0	no	no	yes
married	primary	3777	yes	no	yes
married	secondary	614	no	no	yes

Признаки используются для предсказания целевой переменной.

Данные из нескольких таблиц

Дисбаланс классов



Что же делать?

1. Заменить на значение, которое встречается чаще остальных

Что же делать?

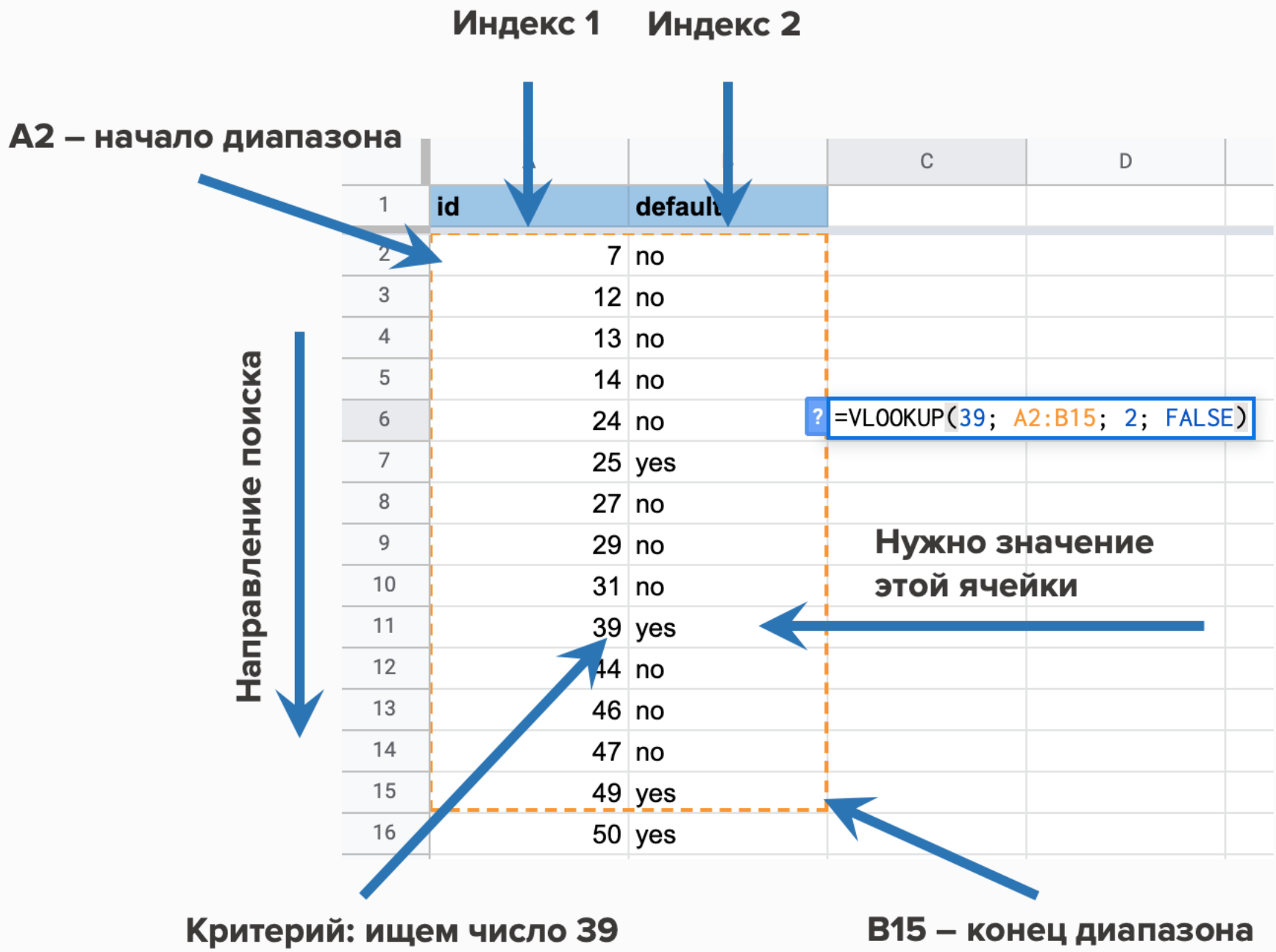
1. Заменить на значение, которое встречается чаще остальных

2. Удалить строки с пропусками

Что же делать?

1. Заменить на значение, которое встречается чаще остальных
2. Удалить строки с пропусками
3. Найти дополнительный источник данных

VLOOKUP


Индекс 1
Индекс 2
A2 – начало диапазона
Направление поиска
=VLOOKUP(39; A2:B15; 2; FALSE)
Нужно значение этой ячейки
Критерий: ищем число 39
B15 – конец диапазона