20/02/2025

В базе данных 1259 записей с двумя столбцами:

* username: Телеграм-юзернейм.
* status: Статус юзернейма (например, "Занято", "Доступно для покупки" и другие).
*  "Свободно": 765 записей
*  "Занято": 296 записей
*  "Доступно для покупки": 138 записей
*  "Невозможно определить": 26 записей
*  "Продано": 23 записи

**🧠 1. Описательный анализ (Descriptive Analysis)**

Я выделил различные характеристики юзернеймов:

* Длина имени.
* Наличие цифр.
* Наличие символа подчёркивания (\_).
* Наличие заглавных и строчных букв.
* Алфавитно-цифровой ли юзернейм.

Далее я просто посчитал средние значения этих характеристик в зависимости от статуса юзернейма ("Свободно", "Занято", "Продано" и т.д.). Это позволило увидеть, какие паттерны (шаблоны) чаще встречаются в каждой категории.

выводы из анализа паттернов юзернеймов по их статусам:

1. Длина юзернейма

"Свободно": Средняя длина юзернеймов — 14.7 символов. Длинные имена чаще остаются свободными.

"Занято": Средняя длина — 11.7 символов. Занятые имена короче, возможно, потому что короткие имена более привлекательны или легко запоминаются.

"Доступно для покупки": Длина около 10.4 символов. Имена, которые предлагаются к продаже, ещё короче. Возможно, продавцы выбирают короткие имена, так как они более ценные.

"Продано": Средняя длина — 6-7 символов. Это подтверждает, что самые короткие и, вероятно, лаконичные имена действительно пользуются спросом.

2. Наличие цифр

В "Продано" и "Доступно для покупки" почти нет цифр. Это подтверждает, что "чистые" текстовые имена ценятся выше.

В "Свободно" и "Невозможно определить" цифры встречаются чаще. Возможно, имена с цифрами менее популярны.

3. Символ подчеркивания (\_)

Чаще всего встречается в статусе "Свободно" (56%) и "Невозможно определить" (54%).

В "Занято" — 38%, а в "Продано" и вовсе 0%. Имена с подчёркиванием менее востребованы.

5. Алфавитно-цифровые имена (без специальных символов)

"Продано" — 100% алфавитно-цифровые. Это подчёркивает тренд на "чистые" имена.

"Доступно для покупки" — около 90% тоже алфавитно-цифровые.

"Свободно" и "Невозможно определить" имеют значительно меньше алфавитно-цифровых имен, что указывает на наличие специальных символов или сложных комбинаций.

Основные гипотезы:

Самые короткие и "чистые" имена (без символов и цифр) ценятся больше всего.

Подчёркивания и цифры снижают вероятность того, что имя будет продано или востребовано.

Длинные имена, особенно со специальными символами, остаются свободными.