Оглавление

[**Введение:** 1](#_Toc560707)

[**SOLUTION 1:** 1](#_Toc560708)

[**SOLUTION 2:** 2](#_Toc560709)

[**SOLUTION 3:** 2](#_Toc560710)

# **Введение**

*Добрый день!*

*Спасибо за интерес к нашей вакансии..*

*Мы хотели бы предложить вам выполнить тестовое задание, состоящее из трех пунктов.*

*Ограничение по срокам выполнения:* ***2 дня****.*

*Если у вас есть какие-то вопросы - задавайте, постараюсь на них ответить.*

*Выполненное задание направляйте на почту:* [*grekova@sports.ru*](mailto:grekova@sports.ru)

*Обратную связь по выполненным заданиям мы даём в течение 3 рабочих дней с момента получения.*

1. ***Даны две таблицы:*** [***http://sqlfiddle.com/#!17/5d361***](http://sqlfiddle.com/#!17/5d361)***. При помощи SQL-запросов получить: топ-5 тегов (название, число упоминаний), которые чаще всего встречаются в постах; пост с наибольшим количеством тегов. Результаты представить в виде текстового файла либо ссылки на sqlfiddle.com.***

# **SOLUTION 1**

* Топ-5 тегов (название, число упоминаний), которые чаще всего встречаются в постах

-- Разобьем задачу на несколько малых. Первая: получим таблицу тегов, которые

-- чаще всего встречались в постах. Вторая: сделаем JOIN таблицы с названиями и

-- таблицы полученной в первой задаче. Третья: выведем на экран нужное значение

-- топ-X (где X=5) тегов

[LINK](http://sqlfiddle.com/#!17/5d361/4465/0)

SELECT sports\_tags.name, table\_2.ct FROM sports\_tags

LEFT JOIN (SELECT DISTINCT unnest(tags) as a1, COUNT(unnest(tags)) as ct

FROM sports\_posts

GROUP BY a1) as table\_2 on sports\_tags.tag\_id = table\_2.a1

ORDER BY table\_2.ct DESC

LIMIT 5

* Пост с наибольшим количеством тегов

-- Можно сделать, например, двумя способами.

-- Выбрать все массивы тегов и агрегировать их длины (cardinality). После:

-- 1. Отсортировать по убиванию/возрастанию и выбрать первый или последний

-- (*опасно, так как могут быть посты с одинаковым количеством тегов*)

-- 2. Сделать подзапрос (sub query / nested query) на максимальную длину массива.

[LINK](http://sqlfiddle.com/#!17/5d361/4463/0)

SELECT post\_id, icount(tags) FROM sports\_posts

WHERE icount(tags) IN (SELECT MAX(icount(tags)) FROM sports\_posts)

1. ***Проанализировать*** [***массив данных***](https://drive.google.com/file/d/1f0oO6ZXrEog-3dULKn_QUPfAliKUxsCo/view?usp=sharing) ***при помощи языка Python (допускается и рекомендуется использование дополнительных библиотек): вычисление среднего, максимального/минимального значений, медианы, моды числовых значений как для всего массива в целом, так и для каждого типа контента (столбец Type) в отдельности. Найти самый популярный объект в выборке, объяснить почему. Решение предоставить в виде .py/.ipynb файла на github.***

# **SOLUTION 2**

* [LINK](https://github.com/Witold1/sports_ru_test/blob/master/notebooks/task_2_python.ipynb)

1. ***Установите*** [***приложение Sports.ru***](https://redirect.appmetrica.yandex.com/serve/1178879059539581405)***. Назовите по 3 ключевых показателя, которые должны учитывать product-менеджеры, редакция и маркетинг.***

# **SOLUTION 3**

**Первым делом поймем, что есть продукт***: кроссплатформенное приложение, публикующее информацию на спортивную тематику, с встроенной рекламой и подпиской на отключение рекламы*

**Теперь попробуем выбрать фреймворк** для решения: *«1 показатель по аудитории + 1 показатель по команде + 1 показатель по деньгам + представление целей + описание показателей»*, *-* и работать в нём*.*

**Product-менеджеры**: Retention, настроение команды, LTV.

**Цели:** важно понимать проблему и «боль», от которых избавляет продукт.

Важно качество продукта (совокупность работы редакции, разработчиков, маркетинга), доходы и среда в командах, связанных с продуктом.

Важно представление о месте продукта среди конкурентов, его плюсах и минусах

**Описание показателей:**

*Retention* – какая доля пользователей у нас в период T+1 по отношению к периоду T;

*Настроение команды* – результаты внутренних мониторингов, личные ощущения;

*LTV* – какую ценность принесёт пользователь за период пользования продуктом;

**Редакция (editoral)**: MAU, «Editorial calendar completed rates», Total interactions.

**Цели:** важно качество контента для аудитории (сроки публикаций, качество публикаций, количество опечаток и количество просмотров, количество уникальных пользователей, показатели сессии – время, удовлетворённость, активность) и характеристики аудитории (кто, что и как хотят получить).

**Описание показателей:**

*Month active users* – сколько пользователей было за период времени равный месяцу;

*Выполненность календаря публикаций* – были ли значимые просрочки, были ли не сданные задания, статистики по задачам, удовлетворённость команды своей работой;

*Total interactions* – сколько действий с продуктом сделал пользователь;

**Маркетинг (marketing)**: Churn, «достаточное понимание аудитории», ROI.

**Цели:** важна жизнь и восприятие продукта на рынке: приток пользователей, доходы от пользователей, характеристики (репутация, узнаваемость) продукта.

**Описание показателей:**

*Churn* – сколько новых пользователей перестают быть пользователями в период времени T+1 (при действиях по привлечению);

*«достаточное понимание аудитории»* – выделение кластеров с высоким качеством, описание кластеров, достаточное использование этой информации при действиях по привлечению;

*LTV и ROI* – ценность, приносимая пользователем и возврат инвестиций от привлечения пользователя;