

Задание по анализу данных

Вам будет предоставлен доступ к базе данных с одной таблицей: id обращения, datetime обращения, бренд и тип источника (органика или другое). Также будет предоставлен файл csv с количеством показов списка запросов по брендам: дата, бренд, id поисковой фразы и количество ее показов. Нужно написать скрипт на языках python/R и SQL, который определит зависимость ежедневного количества обращений с каналов органики от спроса (частотности запросов в вордстате).

Результатом работы будут являться три файла:

1. Файл самого скрипта .py/.ipynb/.r.
2. Файл requirements.txt со списком пакетов и их версий, использованных при работе (для воспроизводимости результата) или его аналог для языка R.
3. Пояснительная записка к результату (опционально, может быть написана в самом блокноте рядом с кодом).

Подробности по выполнению задания:

1. Вам на почту или другим способом высылается пара логин-пароль.
2. Сервер PostgreSQL расположен по адресу analytics.maximum-auto.ru:15432. Подключиться нужно к базе данных data.
3. В БД лежит таблица wordstat.communications, в которой хранится информация обо всех обращениях пользователей, произошедших за определенный период, в разбивке по дням и брендам.
4. На почту отправим файл в формате csv с выгрузкой по частотности запросов в разбивке по брендам и дням за тот же период. Значение в столбце shows показывает сумму количества показов за последний месяц. Подробности см. в описании поля SearchedWith в [справке Yandex.Direct API](#).
5. Напишите скрипт, который решает задачу:
 - 5.1. Формирует общий датасет, соединяя данные из таблицы обращений на сервере PostgreSQL с данными из файла csv.
 - 5.2. Проверяет, существует ли взаимосвязь между количеством обращений по бренду и количеством показов соответствующих поисковых фраз за последний месяц.
6. Напишите пояснительную записку, в которой:
 - 6.1. Кратко опишите подход к решению задачи.
 - 6.2. Опишите результаты анализа: существует связь или нет, какие-то дополнительные ее характеристики (сила, статистическая значимость и подобное)
 - 6.3. (Опционально) Опишите дальнейшие шаги по улучшению модели или предобработке данных, которые можно было бы предпринять в реальной, а не тестовой задаче.