

Анализ базы данных сервиса для чтения книг по подписке

Статус проекта: *завершен*

Использованные библиотеки: pandas, matplotlib, numpy, sys, sqlalchemy

Описание проекта:

Наша компания решила купить крупный сервис для чтения книг по подписке: из-за коронавируса жители городов стали меньше выходить на улицу, посещать кафе, кинотеатры, торговые центры и другие общественные места. Однако теперь стало больше времени для книг, что подтолкнуло многих стартаперов создать приложения для тех, кто любит читать.

Наша первая задача — проанализировать базу данных купленного сервиса. В ней — информация о книгах, издательствах, авторах, а также пользовательские обзоры книг. Эти данные помогут сформулировать ценностное предложение для нового продукта.

Исследование пройдет в три этапа:

1. Исследуем таблицы базы данных.
2. С помощью SQL-запросов посмотрим:
 - посчитаем, сколько книг вышло после 1 января 2000 года;
 - для каждой книги посчитаем количество обзоров и среднюю оценку;
 - определим издательство, которое выпустило наибольшее число книг толще 50 страниц — так мы исключим из анализа брошюры;
 - определим автора с самой высокой и самой низкой средней оценкой книг — учитывая только книги с 50 и более оценками;
 - посчитаем среднее количество обзоров от пользователей, которые поставили больше 50 оценок.
3. Опишем выводы по результатам знакомства с базой данных.