

Изучение данных, размещенных на платформе StackOverflow

Статус проекта: *завершен*

Использованные

библиотеки: pandas, matplotlib, numpy, sys, sqlalchemy, datetime, seaborn, warnings

Описание проекта:

Мы будем работать с базой данных StackOverflow — сервиса вопросов и ответов о программировании. StackOverflow похож на социальную сеть — пользователи сервиса задают вопросы, отвечают на посты, оставляют комментарии и ставят оценки другим ответам. Работа будет проходить с версией базы, где хранятся данные о постах за 2008 год, но в таблицах будет информация и о более поздних оценках, которые эти посты получили.

Наша задача произвести различные выгрузки данных с помощью SQL-запросов и проанализировать результаты.

Исследование пройдет в три этапа:

1. Изучение схемы (ER-диаграммы).
2. Обзор данных.
3. Выгрузки данных с помощью SQL-запросов и анализ.