

Исследование данных с использованием SQL

Целью проекта является исследование данных из открытых источников.

Работа с базой данных ведётся на языке SQL. Предварительная обработка осуществляется на Python.

1. Источник данных

Анализ проводился на данных о процессе обучения в Университете Генуи.

Студенты обучались цифровой электронике в программном эмуляторе (e-learning).

Ссылка на данные:

<https://academictorrents.com/details/e24e083cc337695bb84a2b68707695579c0ab4d8>

2. Постановка задачи

В рамках экспериментального проекта хотелось проверить несколько показателей:

- 1) Общее количество студентов,
Количество записей в основной таблице;
- 2) Период обучения в датасете;
- 3) Топ-10 студентов по успеваемости;
- 4) Статистика по успеваемости:
Среднее, мода, медиана, дисперсия;
- 5) Количество законченных сессий для каждого студента;
- 6) Время выполнения подзадний (activity), время выполнения заданий (exercise);
- 7) Максимальное и минимальное время выполнения подзадания.

3. Первичная обработка данных

Подготовка данных перед загрузкой в базу проводилась с помощью Python и библиотеки pandas.

Краткий список преобразований:

- 1) Переименование столбцов в «snake_case»;
- 2) Добавление несуществующих столбцов;
- 3) Очистка от пробелов;
- 4) Преобразование типов данных, *str* → *float*;
- 5) Приведение формата дата-время к ISO. Добавление «Ведущих нулей», преобразование даты для функций SQLite.

4. Загрузка данных в базу

Обработанные данные загружались из Python-скриптов в файл базы SQLite. Файл с готовой базой загружался в Dbeaver, в котором велась дальнейшая работа, - просмотр таблиц и создание SQL-запросов.

В скриптах Python использовались библиотеки:

- pandas — обработка табличных данных;
- sqlite3 — для обращения к базе SQLite;
- os — для работы с директориями и пакетной загрузки таблиц *processes*.

5. Исследование данных с помощью SQL

Решались 7 задач из раздела «2. Постановка задачи».

Главная особенность исследования:

мода, медиана и дисперсия рассчитаны на SQL.

Файл с SQL-запросами Dbeaver:

EPM_SQL_queries.txt

Снимки (скриншоты) с результатами запросов приведены в Приложении:

Приложение - результаты SQL запросов_02.pdf