

## АНАЛИЗ НАСЕЛЕНИЯ МИРА С ИСПОЛЬЗОВАНИЕМ ПРОГРАММНЫХ СРЕДСТВ

При анализе больших объемов информации используются различные базы данных. Существуют разные способы работы с базами данных. Рассмотрим интерактивную среду разработки Jupyter Notebook, которая имеет широкий набор инструментов для работы с базами данных. Она представляет собой веб-приложение, которое позволяет создавать и редактировать документы, содержащие код, текст, графику и другие элементы, которые могут быть выполнены и отображены в браузере.

Использование Jupyter Notebook изучим на примере анализа населения мира. Статистические данные обо всех странах находятся в свободном доступе в сети Интернет и могут быть представлены в базе данных SQLite.

Набор данных довольно прост. Он состоит из 11 столбцов и 262 строк: *id* – идентификационный номер, присвоенный стране; *code* – код страны, составленный из двух первых букв страны; *area* – общая площадь страны; *population* – численность населения страны; *population\_growth* – коэффициент прироста населения; *birth\_rate* – коэффициент рождаемости; *death\_rate* – уровень смертности; *migration\_rate* – соотношение людей, выехавших из одной страны в другую.

Для подключения Jupyter Notebook к базе данных используются специальные расширения.

`%%capture` – команда IPython, которая перенаправляет вывод стандартного потока в переменную.

`%load_ext sql` – команда IPython, которая загружает расширение `sql` для использования SQL-запросов в ноутбуке Jupyter.

`%sql sqlite:///factbook.db` – команда IPython, которая устанавливает соединение с базой данных `factbook.db`.

Например, для определения максимума и минимума численности населения и его прироста во всех странах следует выполнить SQL-запрос:

```
%sql
SELECT MIN(population) AS min_pop, MAX(population) AS
max_pop, MIN(population_growth) AS min_pop_growth,
MIN(population_growth) AS max_pop_growth FROM facts;
```

В результате выполнения запроса можно убедиться, что в Китае проживает максимальная численность населения.

Чтобы найти наиболее густонаселенную страну надо исполнить SQL-запрос:

```
%sql
SELECT name, population, area FROM facts
WHERE population > (SELECT AVG(population) FROM facts)
AND area < ((SELECT AVG(area) FROM facts) ORDER BY 2
DESC;
```

Данный SQL-запрос выполняет выборку из таблицы *facts* базы данных, содержащую среднее значение столбцов *population* и *area* для всех строк таблицы *facts*.

Анализ результата выполнения запроса показывает, что Бангладеш является самой густонаселенной страной на небольшой территории. Затем следуют Япония и Филиппины.

Можно также определить ТОП-10 стран с самым высоким уровнем рождаемости с помощью запроса:

```
%sql
SELECT name, birth_rate FROM facts
ORDER BY birth_rate DESC LIMIT 10;
```

SQL-запрос выполняет выборку из таблицы *facts* базы данных, содержащую названия стран и их показатели рождаемости (*birth\_rate*), отсортированные по убыванию показателя рождаемости. Функция *LIMIT* используется для ограничения количества выдаваемых строк до 10. На первом месте оказывается Нигерия, затем Мали, Уганда. В десятку входят страны Африки.

Таким образом, Jupyter Notebook является удобным инструментом для работы с базами данных. Он позволяет подключаться к базам данных, выполнять запросы и обрабатывать результаты, что позволяет быстро и эффективно проводить анализ данных. С его помощью можно быстро создавать и изменять код, он предоставляет широкий набор инструментов для работы с данными, поддерживает разные языки программирования и позволяет создавать отчеты в формате HTML.

Областью применения данного приложения является также машинное обучение, нейросети, визуализация данных и статистика. Ещё такую среду часто используют для поэтапной разработки, когда нужно по шагам проверять работу разных фрагментов кода. Дело в том, что код хранится в независимых ячейках и его можно запускать в любом порядке или поодиночке. Это позволяет быстро экспериментировать с алгоритмами и находить оптимальное решение.