Департамент образования и науки города Москвы

Государственное автономное образовательное учреждение высшего образования города Москвы

«Московский городской педагогический университет»

Институт цифрового образования

Департамент информатики, управления и технологий

ДИСЦИПЛИНА:

Проектный практикум по разработке ETL-решений

**Вебинар 14-03-2025**

**Практическая работа на вебинаре**

Выполнила: Шведова С.С., группа: АДЭУ-211

Преподаватель: Босенко Т.М.

Москва

2025

**Цель**: Заполнить таблицу person в базе данных PostgreSQL фейковыми данными не менее 100 записей и провести анализ этих данных с использованием SQL. Также создать визуализации для полученных результатов.

#### **Пункты задания:**

1. **Заполнение таблицы фейковыми данными.**
   * Используйте библиотеку Faker для генерации фейковых данных.
   * Вставьте сгенерированные данные в таблицу person в базе данных PostgreSQL.
2. **Анализ возраста.**
   * Найдите средний, минимальный и максимальный возраст людей в таблице person.
3. **Анализ распределения по городам.**
   * Определите топ-5 городов, в которых проживает наибольшее количество людей.
4. **Анализ регистрации.**
   * Найдите количество регистраций в каждом месяце за последний год.
5. **Визуализация данных.**
   * Создайте графики для визуализации результатов анализа:
     + Гистограмма распределения возраста.
     + Диаграмма топ-5 городов по количеству проживающих.
     + Линейный график количества регистраций по месяцам.

Для начала надо загрузить контейнеры, это показано на рисунке 1.

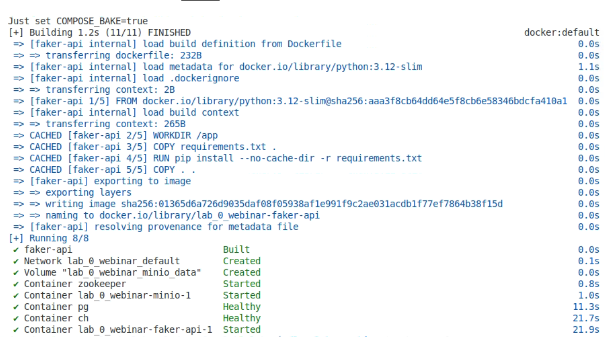


Рисунок 1. Загрузка контейнеров

На рисунке 2 показано загрузка аирфлоу

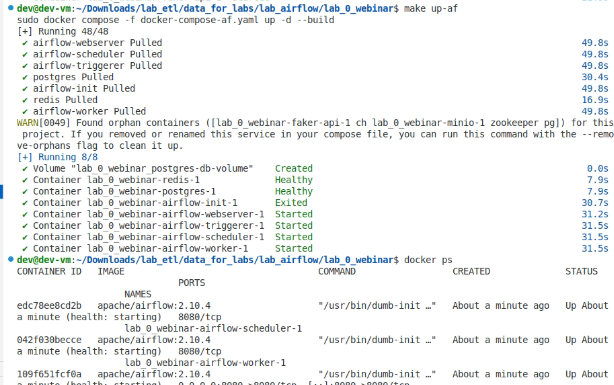


Рисунок 2. Загрузка аирфлоу

На рисунке 3 показан airflow

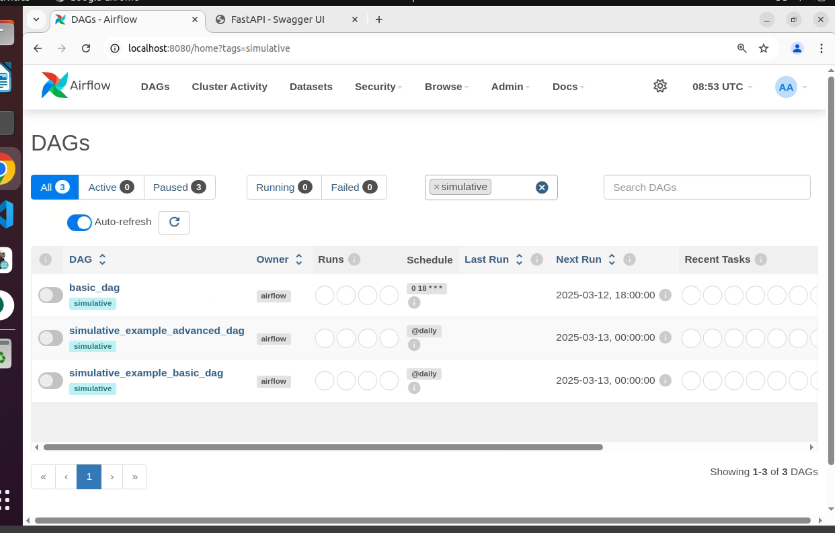


Рисунок 3. Airflow

На рисунке 4 показан dag basic\_dag

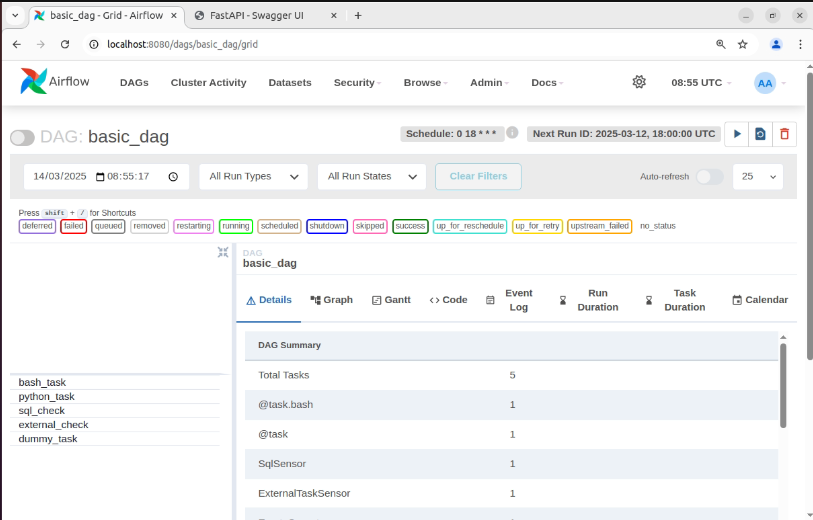


Рисунок 4. Basic\_dag

На рисунке 5 показан запуск дага

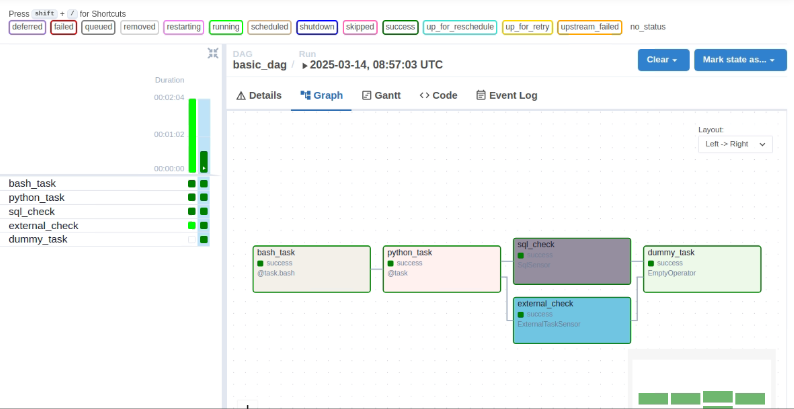


Рисунок 5. Запуск дага

На рисунке 6 показан запуск следующего дага

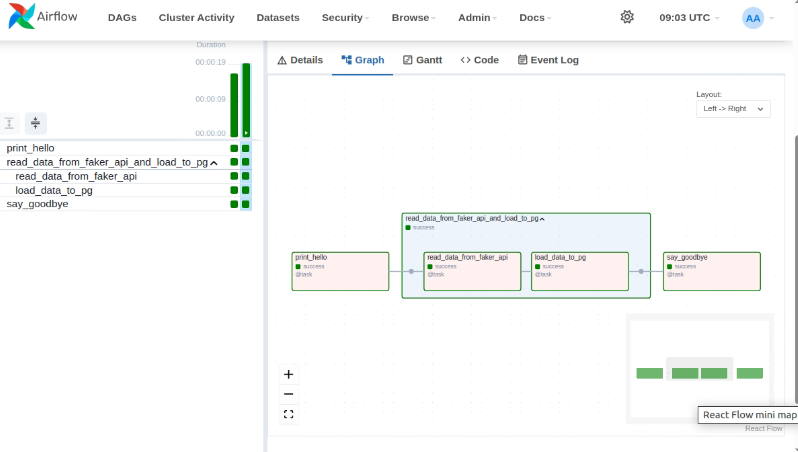


Рисунок 6. Запуск следующего дага

На рисунке 7 показан запуск последнего дага: simulative\_example\_advanced\_dag

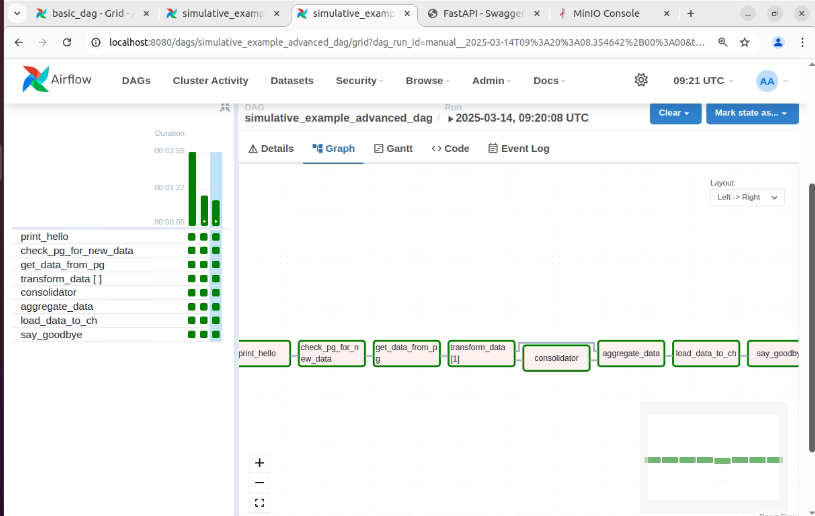


Рисунок 7. simulative\_example\_advanced\_dag

На рисунке 8 видна заполненность таблицы person.

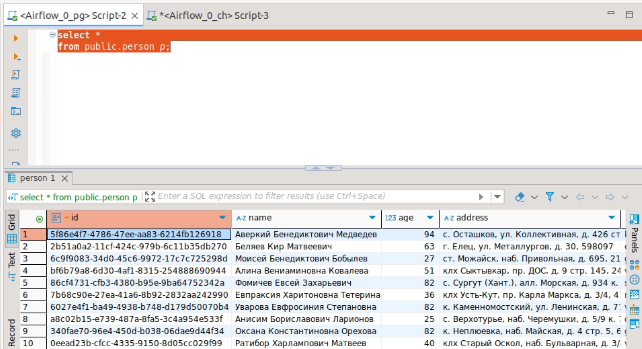


Рисунок 8. Заполненность таблицы person

На рисунке 9 видна заполненность таблицы person count by city

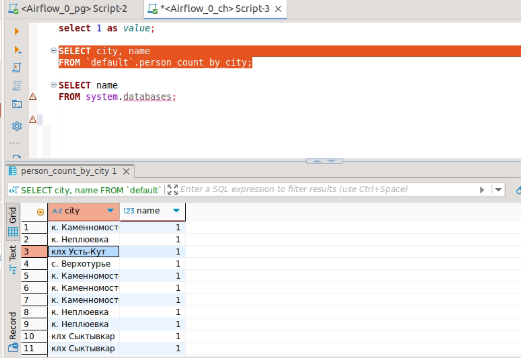


Рисунок 9. Заполненность таблицы person count by city

На рисунке 10 видно минимальный возраст, максимальный возраст и средний возраст

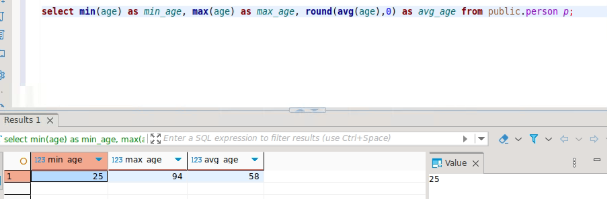


Рисунок 10. Минимальный, максимальный и средний возраст

На рисунке 11 группировка топ 5 городов по количеству людей

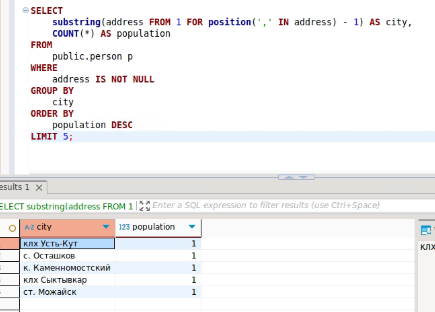


Рисунок 11. Топ 5 городов по количеству людей

На рисунке 12 видно количество регистраций в каждом из городов

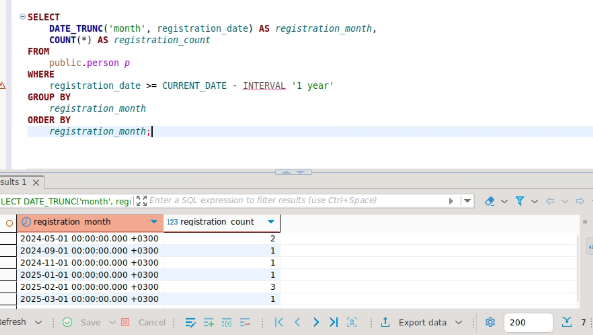


Рисунок 12. Количество регистраций в каждом из городов

На рисунке 13 на столбчатой диаграмме видно, что больше всего людей проживают в Можайске

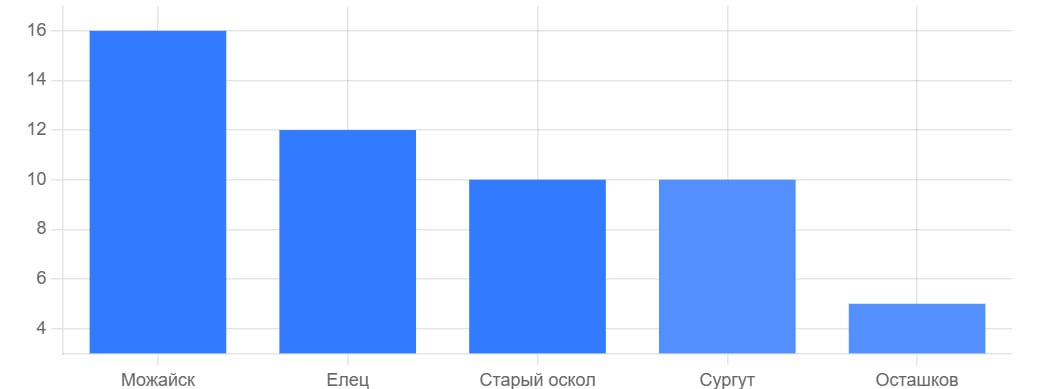


Рисунок 13. Диаграмма топ-5 городов по количеству проживающих

На рисунке 14 построен линейный график количества регистраций по месяцам. Больше всего регистраций было в июле

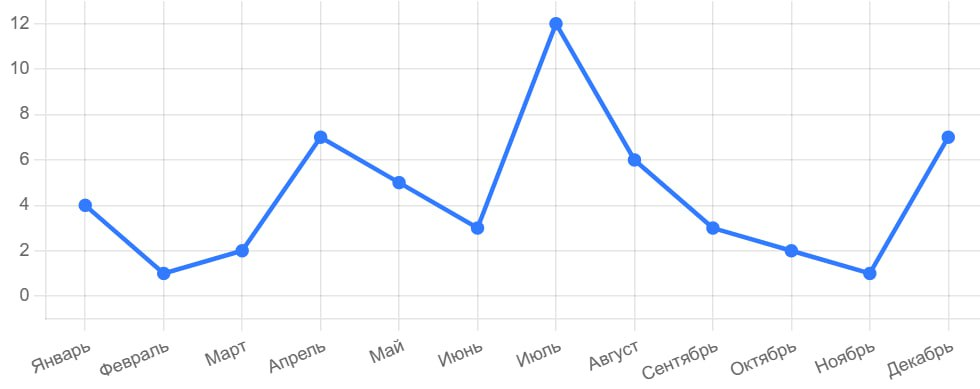


Рисунок 14. Количества регистраций по месяцам

На рисунке 15 показано распределение по возрасту.



Рисунок 15. Распределение по возрасту

Выводы:

1. Был проведен анализ количества людей по городам

2. Был проведен анализ распределения количества людей по городам

3. Был проведен анализ городов с минимальным и максимальным количеством людей