Heinrich Heine
University Düsseldorf

Institut für Informatik
Datenbanken und Informationssysteme
Prof. Dr. Stefan Conrad
Thomas Germer

WiSe 24/25
Assignment 0
Due October 15, 8:00 AM

Relational Databases and Data Analysis

- Welcome to Relational Databases and Data Analysis! Before you get started, please take some time to read the following notes.
- To be admitted to the oral exam, you have to
  1. get 80 of 120 points (12 assignments with 10 points each, copied assignments give 0 points)
  2. (correctly) explain your solutions to Thomas. You can book an appointment starting Jan 24 in ILIAS. Unfortunately, this requirement has become necessary because we have found that many students no longer understand "their" code since ChatGPT was released.
- Your solutions must pass automated tests with specific Python library versions. You can run the tests with the command `python3 zipme.py`, which will create a `*.tar.gz` file that must be uploaded in ILIAS.
- This 0th assignment is an introductory assignment and does not give any points, but you should upload it anyway for testing.

Exercise 1 *Install Python and useful libraries* (0 Points)

The exercises will make use of Python 3.10 and the libraries listed in `data/requirements.txt`. On Debian-based operating systems, you can install the libraries with the following commands:

```
sudo apt install python3-pip
pip3 install -r data/requirements.txt
```

After the installation is finished, solve the task in `exercise_1.py`, run `text_exercise_1.py`, then run `zipme.py` and upload the generated `*.tar.gz` archive in ILIAS.

We will test all assignments with the Docker image created with `data/Dockerfile`. Make sure that your code works in that environment or you wont get any points.

A few things to watch out for:

- Do **not** use absolute file paths like `C:\Users\DonaldDuck\assignment-0\data\data.csv`. Use relative file paths instead, e.g. `data/data.csv`.
- File paths on Linux are case sensitive.
- By default, Python on Windows might not use UTF-8 text encoding. It can be enabled by setting the environment variable `PYTHONUTF8=1` or passing `-Xutf8` to the `python` interpreter.
- We will set `psyopg`'s username etc. with environment variables. Do not set them in Python!

Exercise 2 *PostgreSQL and Docker* (0 Points)

Some exercises will require a PostgreSQL database. Setting up such a database with Docker is relatively easy. Note that you only need Docker *Engine* (not Docker Desktop) on [U/X/K/L]buntu.

We will use PostgreSQL version 17.0.

```
docker run \
    --name mydb \
    --rm \
    -p 5432:5432 \
    -e POSTGRES_PASSWORD="chooseGoodPassword" \
    postgres:17.0
```

Make sure to choose a good password, because anyone can now connect to your database on port 5432 (if you are not running a firewall, which Windows does by default).

Once Postgres started, open another terminal and connect to the database. When connecting from inside the docker container, a password is not necessary.

```
docker exec -ti mydb bash
psql -U postgres
```

If you want to test whether Postgres works or are feeling lonely, you can run the following query to get a date.

```
SELECT NOW();
```

In future exercises, we will interact with the database using the `psycopg` library. The function `psycopg.connect()` will read the username, password and host from the following environment variables:

- `PGUSER` - should have the value `postgres`
- `PGHOST` - can be found in the output of `docker inspect mydb`
- `PGPASSWORD` - you chose that one yourself

Set those environment variables and then run the command `python3 psycopg_example.py`. It should print "`Success!`".

In the exercise group on Thursday, October 17th (Room 25.12.03.33), we will discuss any installation issues you might have.