

# Assignment 1

---

Course: *Big Data*  
Due date: *March 10th, 2023*

Create the database Employees stored in PostgreSQL/TimescaleDB as instructed on Učilnica.

Use Jupyter notebook to write and run your code, and report the results.

Your Jupyter notebook file should serve also as a report so be careful to include everything that should be included in the report. Display only the code results that are relevant. Your submitted file should look like a report.

The “report” has to be structured as follows: Problem description, Results (including code and visualizations) and Conclusions.

Using appropriate Python or ODBC drivers (e. g., psycopg2, odbc-posgresql or another) connect to the database Employees and perform the following tasks with SQL (driver connection) and the most adequate one with ORM (SQLAlchemy). Some queries are too complex and not adequate for ORM.

The general principle of your solution should be to appropriately split the work between the server (SQL) and the client (Python). Try to create reusable queries (use CTE-s when possible).

Rank the employee titles according to the average salary for each department and for the whole company. Present results in a bar chart.

The company active pursues gender equality. Prepare an analysis based on salaries and gender distribution by departments, managers, and titles. Present results in a chart (or several charts) of your choice.

You can choose the visualization tool you like (e.g. matplotlib, ggplot).

The same as above (gender equality) but prepare your analysis also on a yearly basis (the same charts by year).

Check if some employees earn more than their managers. Split the result by year, gender and department, if such employees exist.

Find the most successful department (with highest mean salaries) and chart its characteristics (distribution of titles, salaries, genders, ...). Compare this chart with charts from other departments and hypothesize on reasons for success.

Submit your homework on Učilnica including as one .ipynb file and name it My-Surname\_assignment1.ipynb

Please do not submit .zip files.

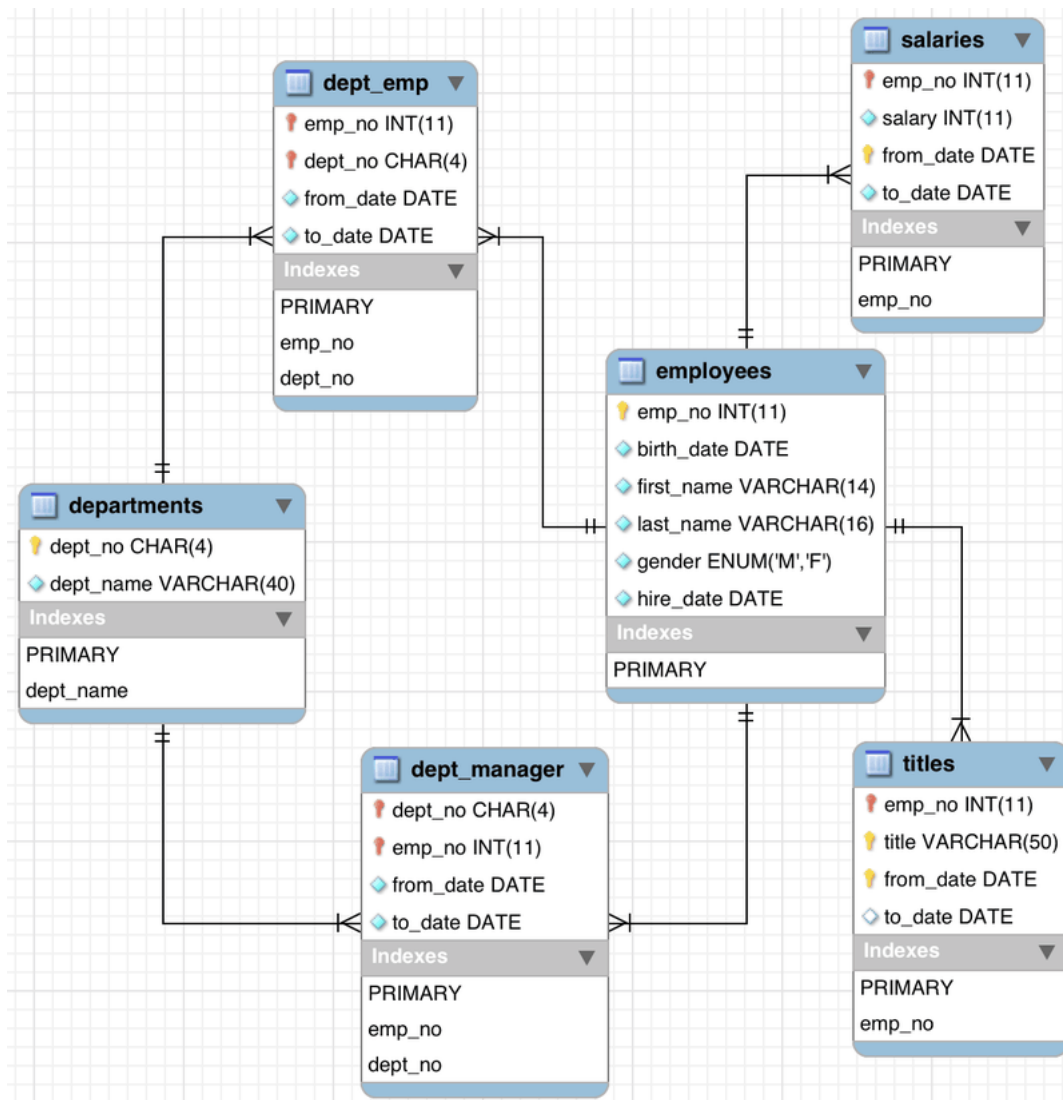


Figure 1: The employees database structure.