

SQL HOMEWORK

EMPLOYEE DATABASE: A Mystery in Two Parts

INTRODUCTION:

The homework seeks to highlight the understanding of PostgreSQL by applying the skills developed in:

- Data Modelling and Engineering
- Data Analysis
- As a bonus: Test the credibility of dataset

1. Data Modelling and Engineering

From the csv files provided, a sketch of an ERD (Entity Relationship Diagram) is carried using [<https://www.quickdatabasediagrams.com>] as a tool which is saved as QuickDBD-SQL_Home_Work.png. Based on the data primary keys are assigned to unique columns in order to identify a row and importantly is to import the data based on the same order that the tables were created by accounting for headers. Table schema, saved as Schemas.sql.

2. Data Analysis

From the schema created, a database (SQL_homework) was created in order to be able to carry out the analysis on the following:

1. List the following details of each employee: employee number, last name, first name, sex, and salary.
2. List first name, last name, and hire date for employees who were hired in 1986.
3. List the manager of each department with the following information: department number, department name, the manager's employee number, last name, first name.
4. List the department of each employee with the following information: employee number, last name, first name, and department name.

5. List first name, last name, and sex for employees whose first name is "Hercules" and last names begin with "B."
6. List all employees in the Sales department, including their employee number, last name, first name, and department name.
7. List all employees in the Sales and Development departments, including their employee number, last name, first name, and department name.
8. In descending order, list the frequency count of employee last names, i.e., how many employees share each last name.

The queries to the above questions are saved as Queries.sql.

3. As a bonus: Test the credibility of dataset

The requirement is to clear every suspicion in the data by generating some visualization of the data.

Solution:

To achieve this the database, SQL_homework was exported into Jupyter Notebook by importing the following dependencies:

```
import os
import pandas as pd
import psycopg2
from matplotlib import pyplot as plt
import sqlalchemy
from sqlalchemy.ext.automap import automap_base
from sqlalchemy.orm import Session
from sqlalchemy import create_engine, inspect
%matplotlib notebook
```

From the Analysis a histogram visualization is created figure 1.0 and 1.1 showing the distribution of salary ranges for all employees. It can be observed that the average salary for all employees is around 52970, with the 25th percentile at 40000, 75th percentile at 61758 and a maximum salary standing at 129493.

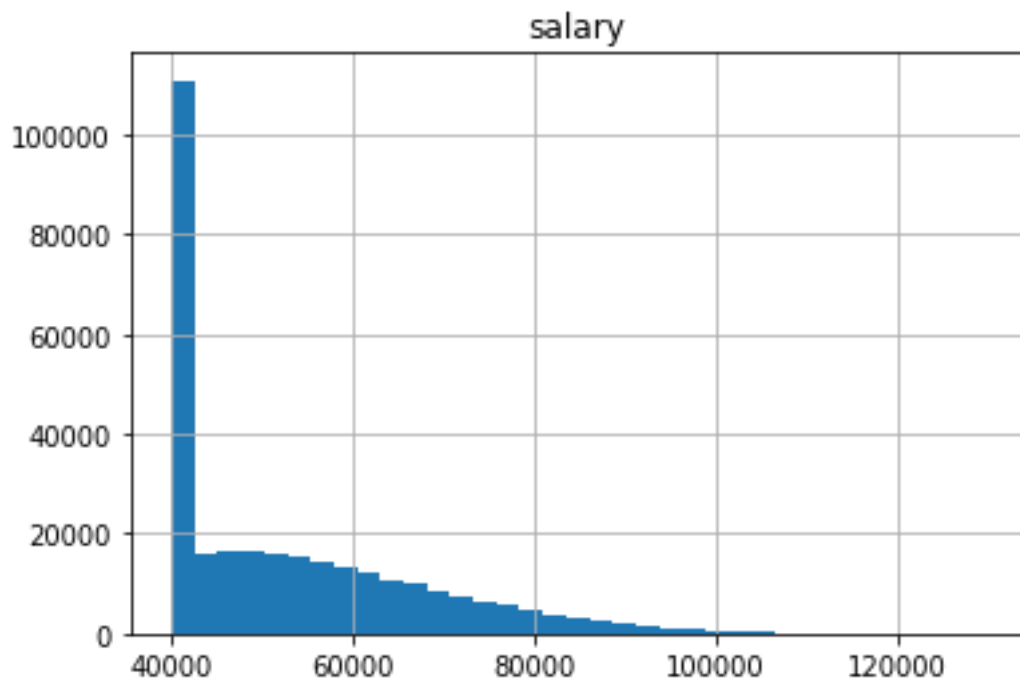


Fig. 1.0: A histogram plot for salary ranges of employees

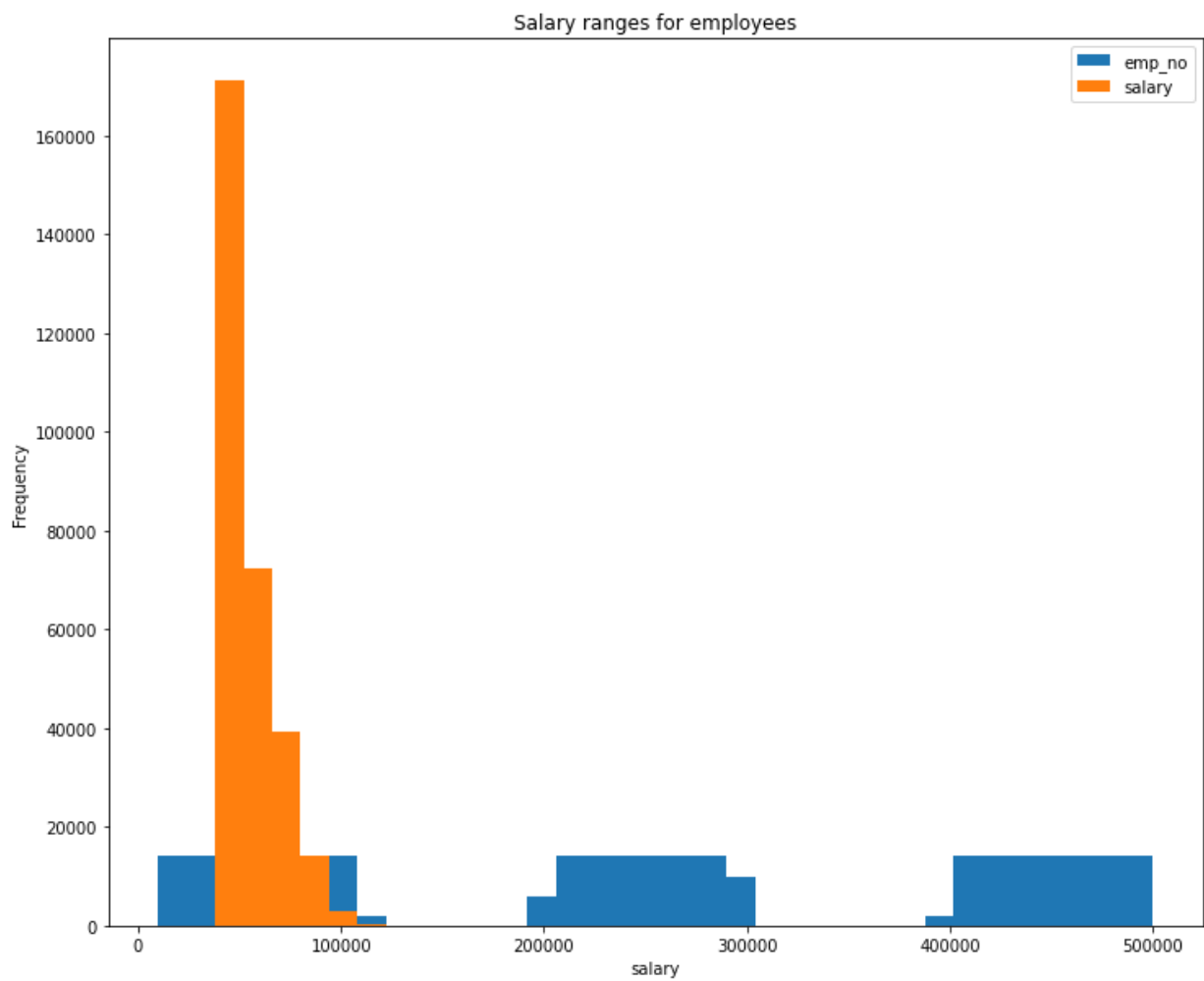


Fig.1.1: Showing salary ranges for employees with label

Finally, a bar chart is plotted based on the average salary by title. Table 1.0 and Figure 1.3 as shown below.

Table 1.0: Average salary by Title of employees

Title	Average Salary
Staff	58465.382850
Senior Staff	58550.172704
Assistant Engineer	48564.434447
Engineer	48535.336511
Senior Engineer	48506.799871
Technique Leader	48582.896092
Manager	51531.041667

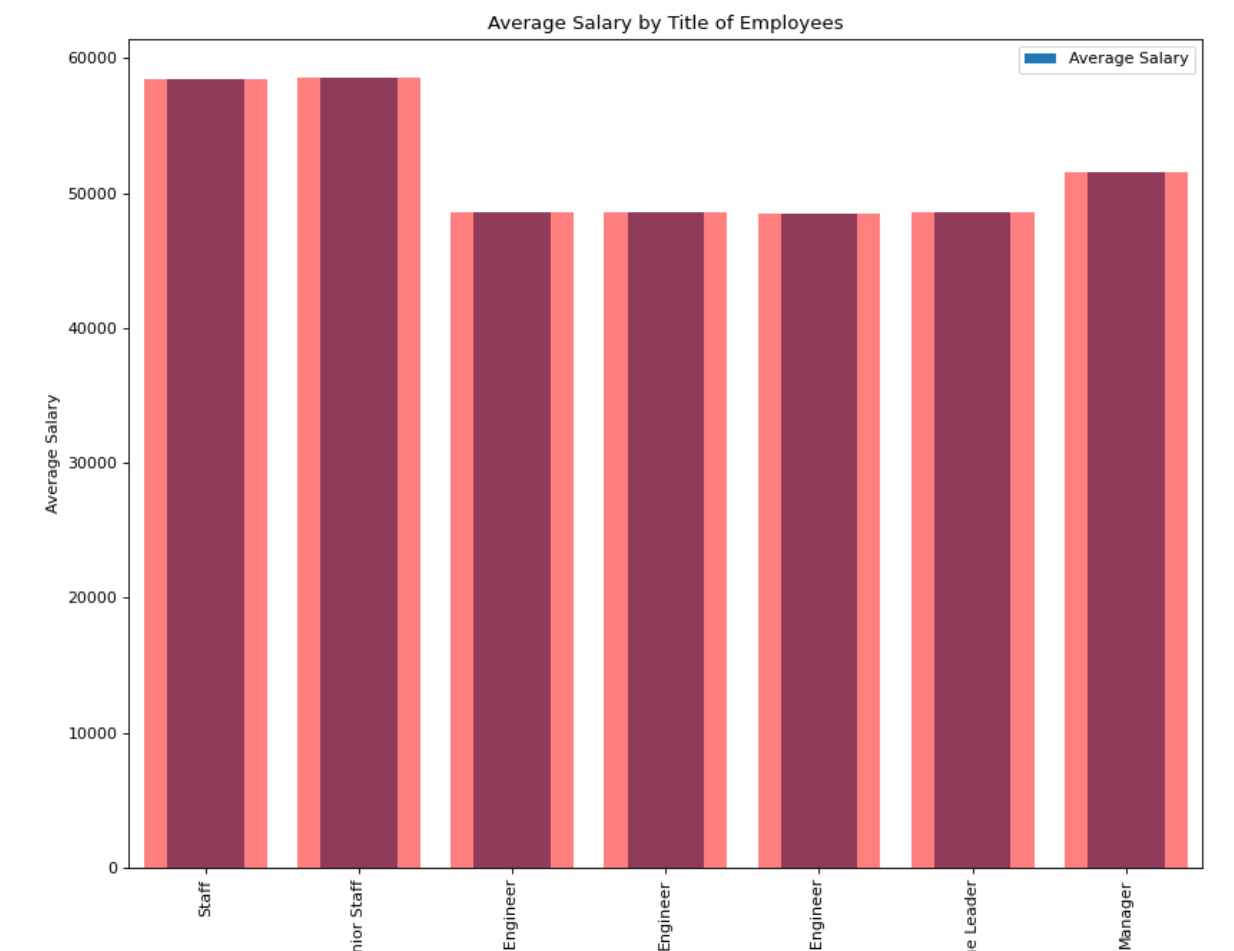


Fig. 1.3: Bar chart of Average salaries by Title of Employees

CONCLUSION:

The sql-challenge was an experienced with the use of many different tools. The major one was getting the database into Jupyter notebook. I had to dig deep as it took majority of this assignment but was at the end a success. Critically, the visualization shows the very import details.