**UPenn Data Analysis Bootcamp – Fall 2020 Cohort**

**SQL Challenge Homework**

**Jeremy Strauch**

**March 2, 2021**

# **Summary**

Our task was to conduct basic data analysis on employees from a corporation spanning the 1980s and 1990s. The analysis was based on six CSV files specifying employee information, departments within the organization, department managers, employee salaries, employee’s assigned departments and employee titles. Our tasks were:

1. Inspect the CSVs and sketch out an ERD of the tables
2. Use the information you have to create a table schema for each of the six CSV files. Remember to specify data types, primary keys, foreign keys, and other constraints
3. Import each CSV file into the corresponding SQL table and then answer the following:
4. List the following details of each employee: employee number, last name, first name, sex, and salary.
5. List first name, last name, and hire date for employees who were hired in 1986.
6. List the manager of each department with the following information: department number, department name, the manager's employee number, last name, first name.
7. List the department of each employee with the following information: employee number, last name, first name, and department name.
8. List first name, last name, and sex for employees whose first name is "Hercules" and last names begin with "B."
9. List all employees in the Sales department, including their employee number, last name, first name, and department name.
10. List all employees in the Sales and Development departments, including their employee number, last name, first name, and department name.
11. In descending order, list the frequency count of employee last names, i.e., how many employees share each last name.
12. Import the SQL database into Pandas.
13. Create a histogram to visualize the most common salary ranges for employees.
14. Create a bar chart of average salary by title.

# **Observations**

1. There are 300,024 employees amongst nine departments.
2. Twenty-four managers led the departments across the lifespan of the data set.
3. 3,121 employees were hired in 1986.
4. There were an amazing twenty people with the first name ‘Hercules.’
5. There were 226 people who shared the last name ‘Baba.’
6. The most common salary, as well as the base salary, is $40,000.
7. 93,000+ employees made $40,000 per year, making the salary histogram skewed to the right
8. The highest salary is $129,492
9. The average salary for most of the departments is approximately $50,000; finance and marketing departments average $60,000; the sales department has the highest average salary at $70,000
10. The database analysis was apparently a test based on employee 499942 being named April Foolsday!!!!!!

# **Appendix**

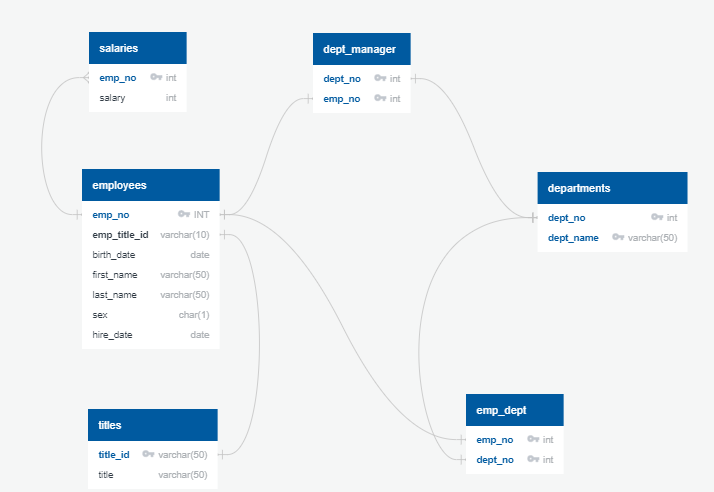


Figure . ERD for the six tables used in the data analysis

Table . Employee information (appended)

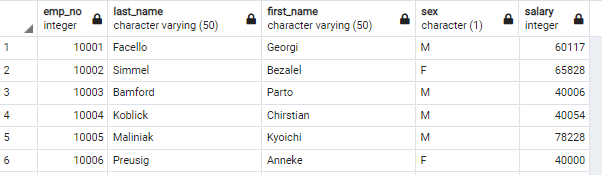


Table . Employees hired in 1986 (appended)

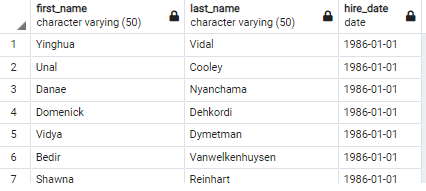


Table . Department managers (appended)

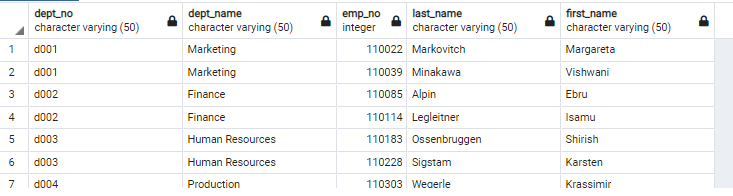


Table . Employees and their assigned departments (appended)

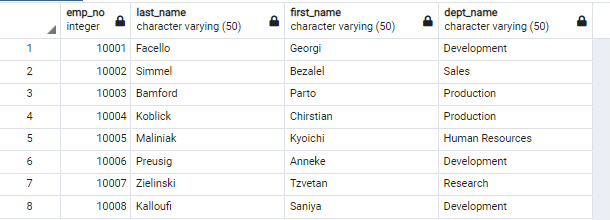


Table . Employees with the first name 'Hercules' (appended)

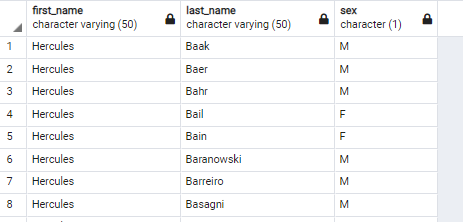


Table . Employees in Sales Department (appended)

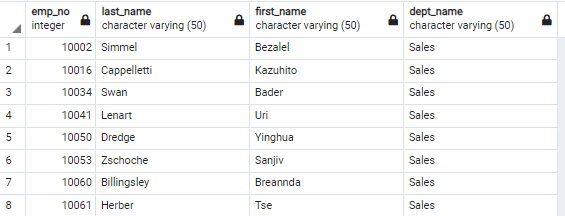


Table . Employees in Sales and Development Departments (appended)

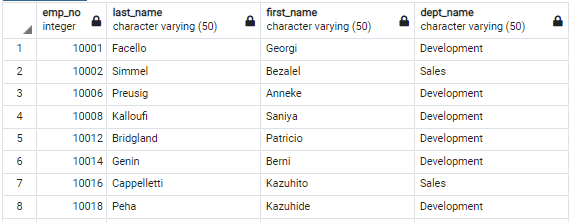
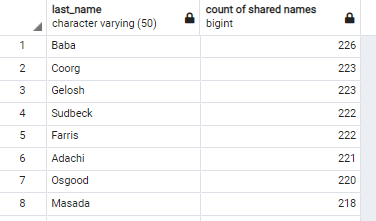


Table . Count of repeated last names, descending order (appended).



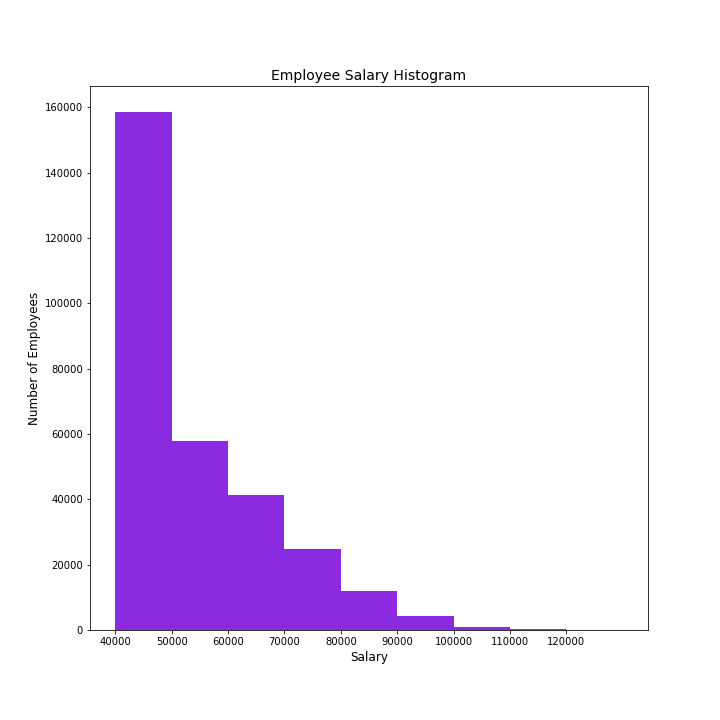


Figure . Employee salary histogram

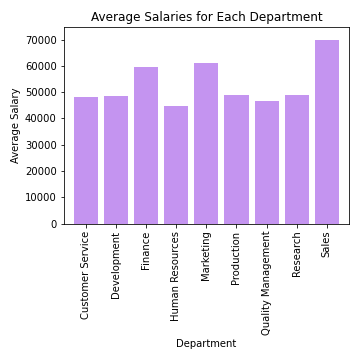


Figure . Average salaries for employees across each department.