SMU

Data modeling, engineering, and analysis

SQL DATABASE

Raymond Bell

November 29, 2022

The table below was generated from a SQL database query that extracted employee number, last name, first name, and sex from the employee table. Then we had to execute an inner join to salaries, by the employee number, to extract the salary data. As shown in the database diagram, the selected columns are contained within these two tables.

Dashboard Properties SQL Statistics Dependencies Dependents Processes schema.sql BellSchema.sql public.departm... department23_code_bellr.sql*

department23_db/postgres@PostgreSQL 15

No limit

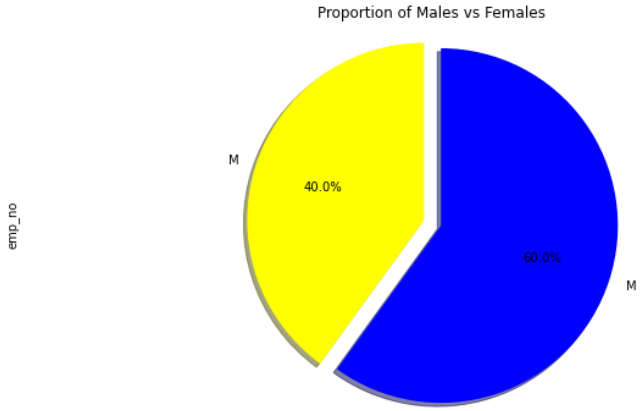
Query Query History Scratch Pad

```
1  -- 1. List the following details of each employee: employee number, last name, first name, sex, and salary
2  SELECT
3      Employees.emp_no,
4      Employees.last_name,
5      Employees.first_name,
6      Employees.sex,
7      Salaries.salary
8  FROM Employees
9  INNER JOIN Salaries ON Employees.emp_no=Salaries.emp_no;
10
```

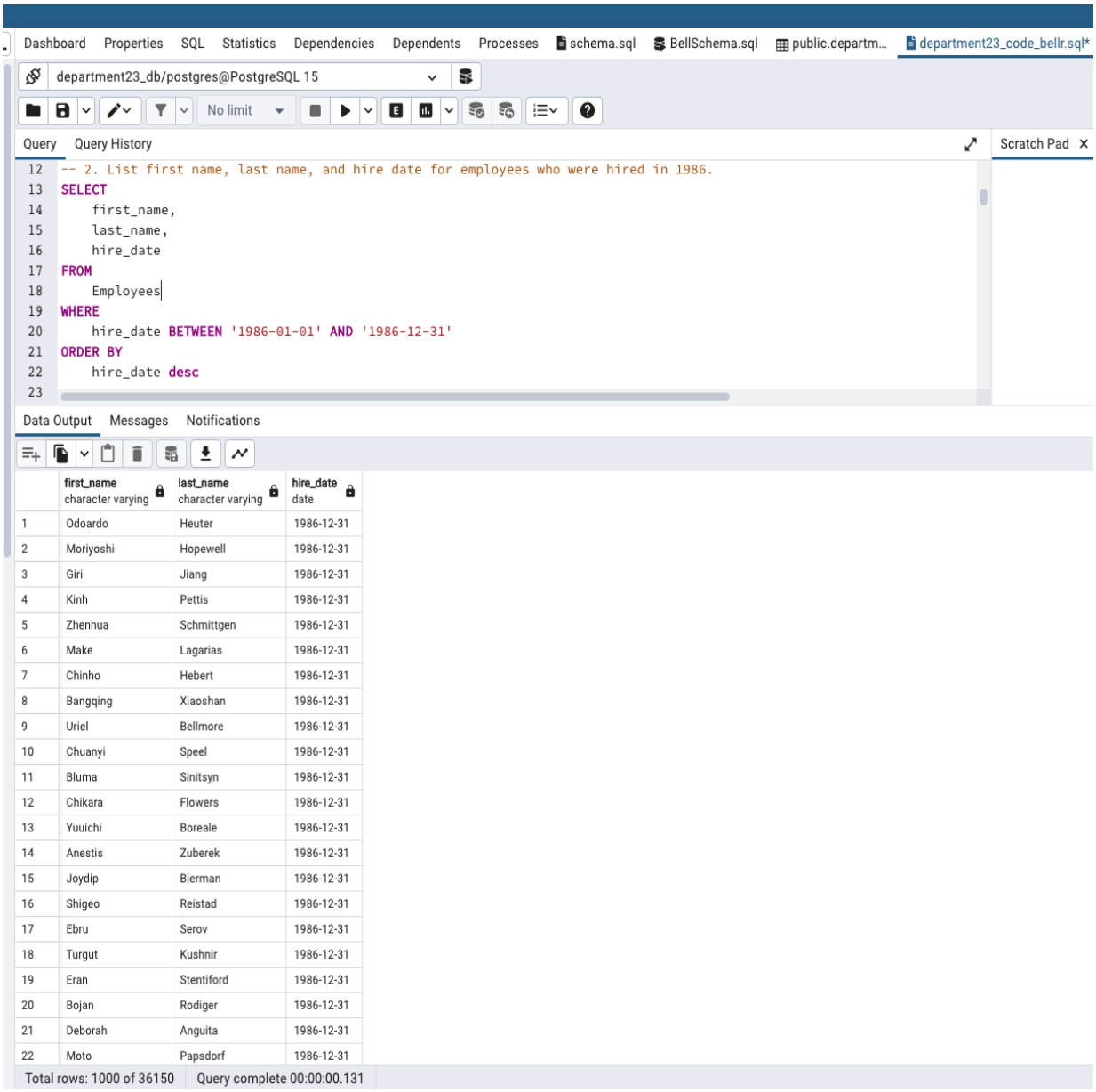
Data Output Messages Notifications

	emp_no integer	last_name character varying	first_name character varying	sex character varying	salary integer
1	57444	Babb	Berry	F	48973
2	263976	Cusworth	Eran	M	40000
3	461591	Samarati	Xudong	M	40000
4	477657	Magliocco	Lihong	M	54816
5	29920	Tyugu	Shuichi	F	40000
6	13616	Lorho	Perry	F	40000
7	246449	Bultermann	Subbu	F	87084
8	21529	Zallocco	Bojan	M	40000
9	17934	Wuwongse	Bilhanan	M	48795
10	48085	Gilg	Venkatesan	M	63016
11	240129	Karnin	Roddy	M	43548
12	427958	Poujol	Feixiong	F	63377
13	280408	Perl	Elliott	M	75963
14	104639	Escriba	Theron	F	40000
15	214555	Copas	Nechama	F	40000
16	229010	Cappelli	Mark	M	40000
17	35916	Talmon	Wonhee	F	40000
18	67799	Cools	Pantung	M	48111
19	411412	Schaar	Hairong	M	42348
20	105597	Kabayashi	Christfried	M	43841
21	244689	Saoudi	Urs	M	40000
22	48885	Antonisse	Phuoc	M	55736
23	244492	Pettit	Shaunak	M	63685

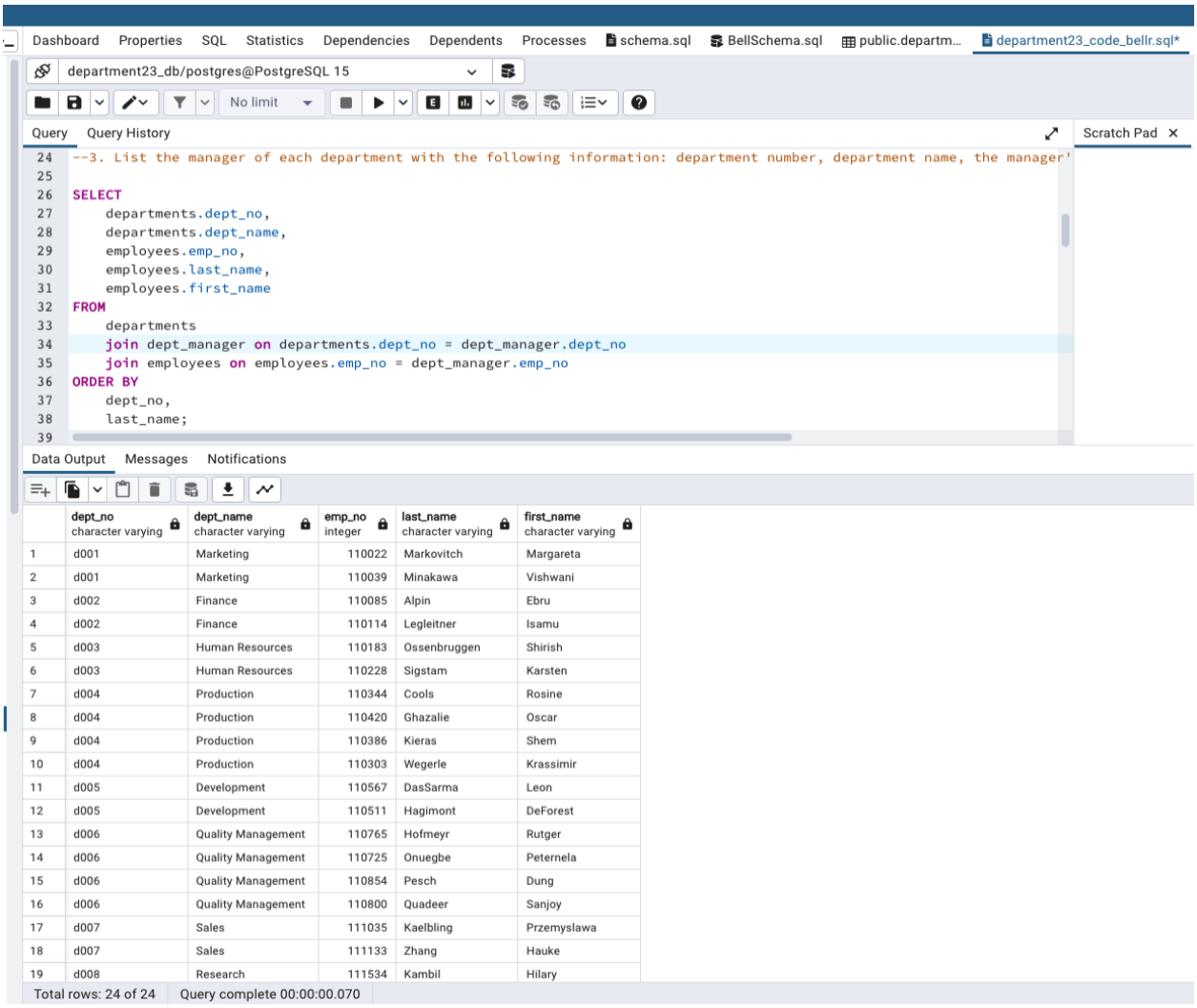
Total rows: 1000 of 300024 Query complete 00:00:00.738



Further, we can see the proportion of men versus women in company.

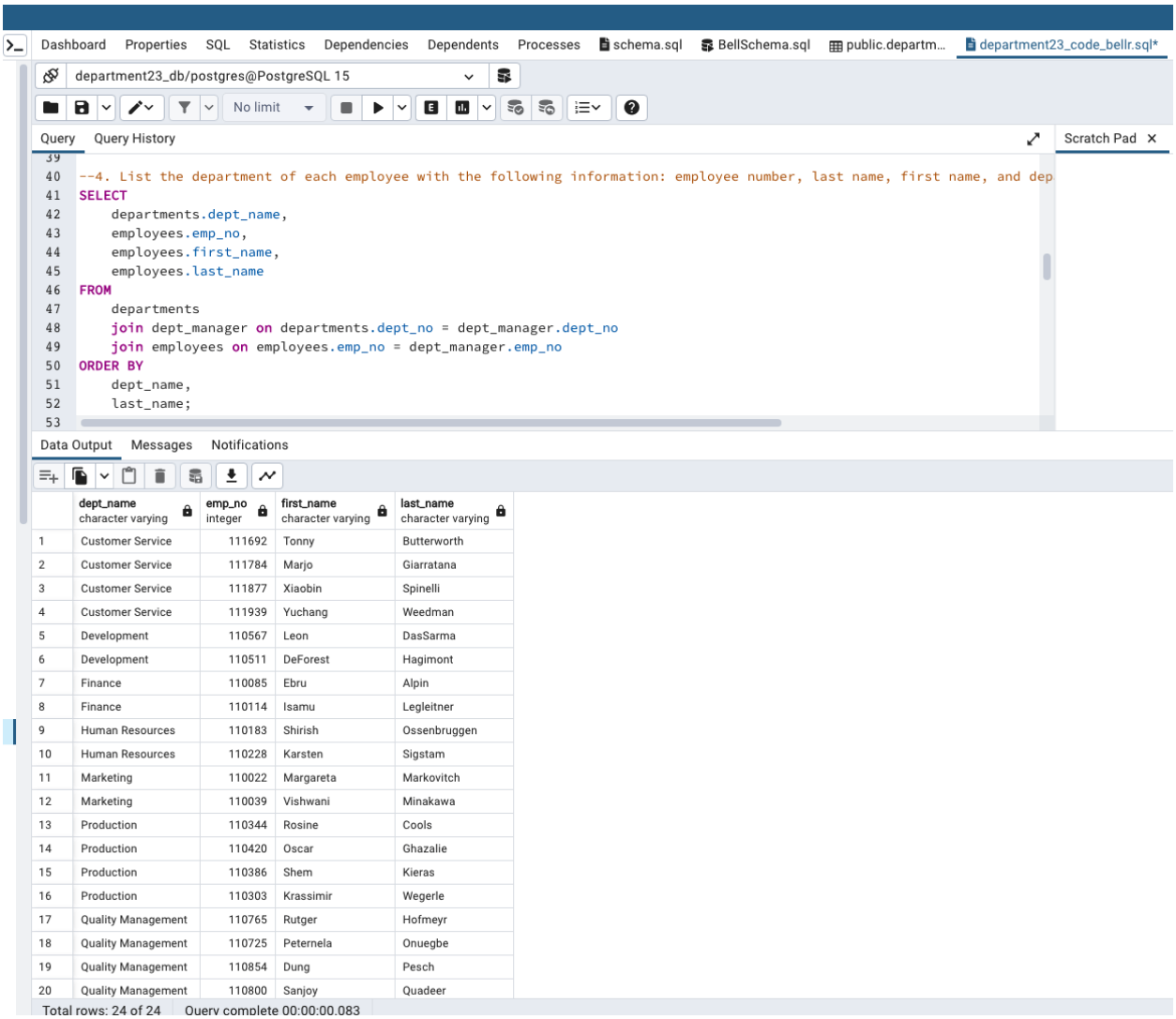


Given the table above we can determine the names of the employees hired in 1986. This information could be used to determine eligibility for rewards and benefits, according to years of service. There are a number of other scenarios where a query such as this might be used.



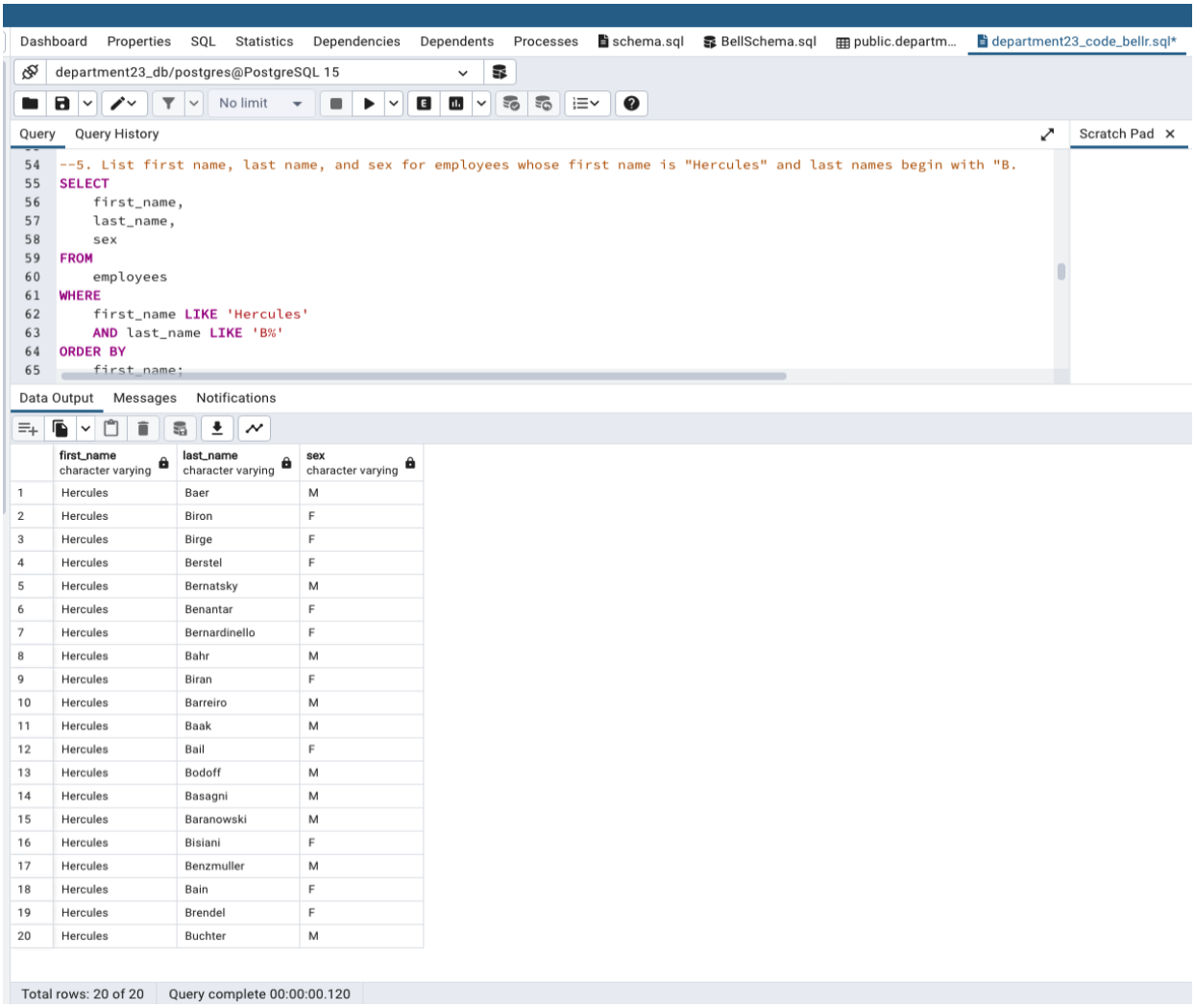
	dept_no character varying	dept_name character varying	emp_no integer	last_name character varying	first_name character varying
1	d001	Marketing	110022	Markovitch	Margareta
2	d001	Marketing	110039	Minakawa	Vishwani
3	d002	Finance	110085	Alpin	Ebru
4	d002	Finance	110114	Legleitner	Isamu
5	d003	Human Resources	110183	Ossenbruggen	Shirish
6	d003	Human Resources	110228	Sigstam	Karsten
7	d004	Production	110344	Cools	Rosine
8	d004	Production	110420	Ghazalie	Oscar
9	d004	Production	110386	Kieras	Shem
10	d004	Production	110303	Wegerle	Krassimir
11	d005	Development	110567	DasSarma	Leon
12	d005	Development	110511	Hagimont	DeForest
13	d006	Quality Management	110765	Hofmeyr	Rutger
14	d006	Quality Management	110725	Onuegbe	Peternela
15	d006	Quality Management	110854	Pesch	Dung
16	d006	Quality Management	110800	Quadeer	Sanjoy
17	d007	Sales	111035	Kaelbling	Przemyslawa
18	d007	Sales	111133	Zhang	Hauke
19	d008	Research	111534	Kambil	Hilary

The table above is a way to identify which employees work within which departments.



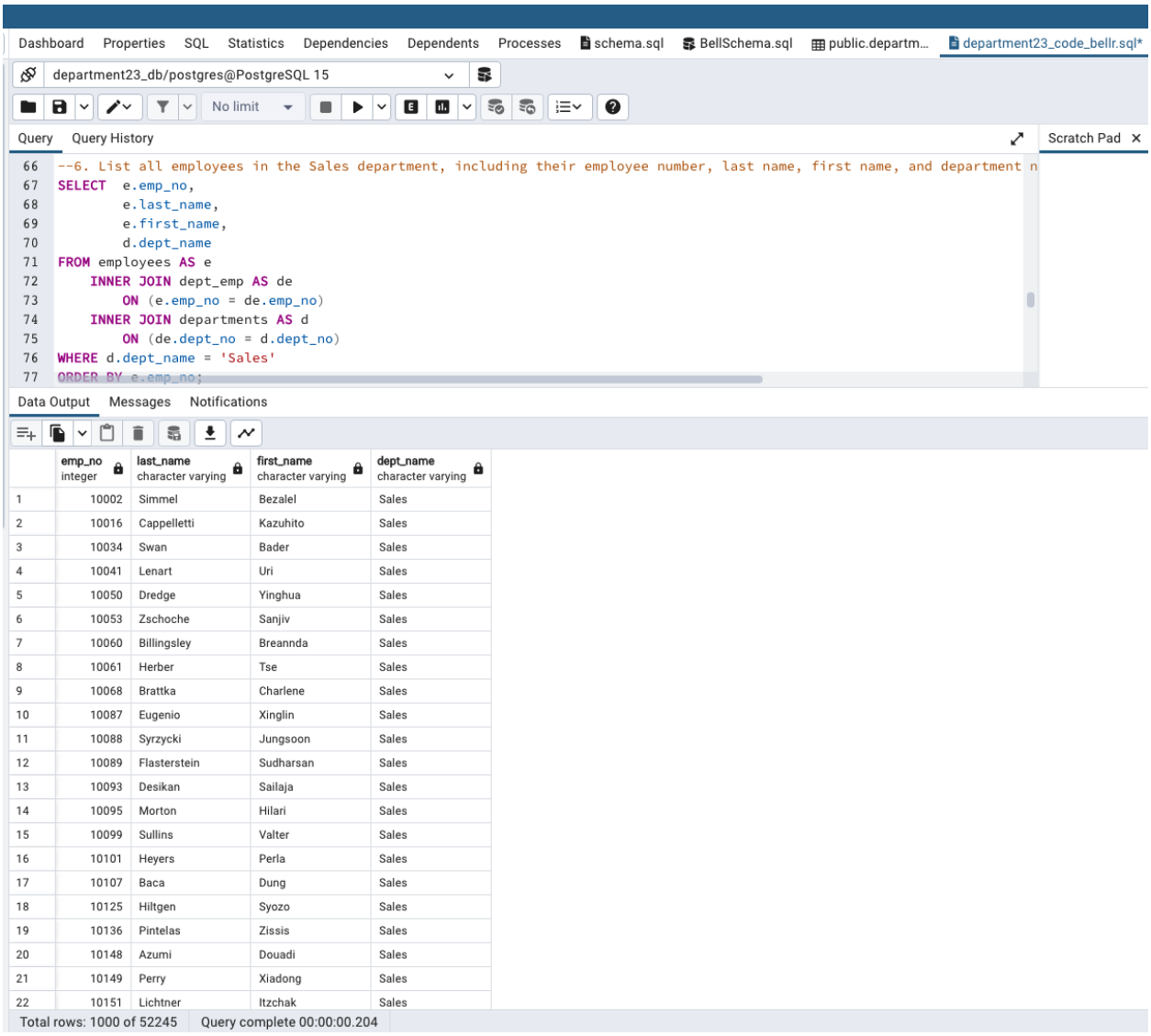
	dept_name character varying	emp_no integer	first_name character varying	last_name character varying
1	Customer Service	111692	Tonny	Butterworth
2	Customer Service	111784	Marjo	Giarratana
3	Customer Service	111877	Xiaobin	Spinelli
4	Customer Service	111939	Yuchang	Weedman
5	Development	110567	Leon	DasSarma
6	Development	110511	DeForest	Hagimont
7	Finance	110085	Ebru	Alpin
8	Finance	110114	Isamu	Legleitner
9	Human Resources	110183	Shirish	Ossenbruggen
10	Human Resources	110228	Karsten	Sigstam
11	Marketing	110022	Margareta	Markovitch
12	Marketing	110039	Vishwani	Minakawa
13	Production	110344	Rosine	Cools
14	Production	110420	Oscar	Ghazalie
15	Production	110386	Shem	Kieras
16	Production	110303	Krassimir	Wegerle
17	Quality Management	110765	Rutger	Hofmeyr
18	Quality Management	110725	Peternela	Onuegbe
19	Quality Management	110854	Dung	Pesch
20	Quality Management	110800	Sanjoy	Quadeer

This output is similar to the previous table minus the column listing the respective department numbers.



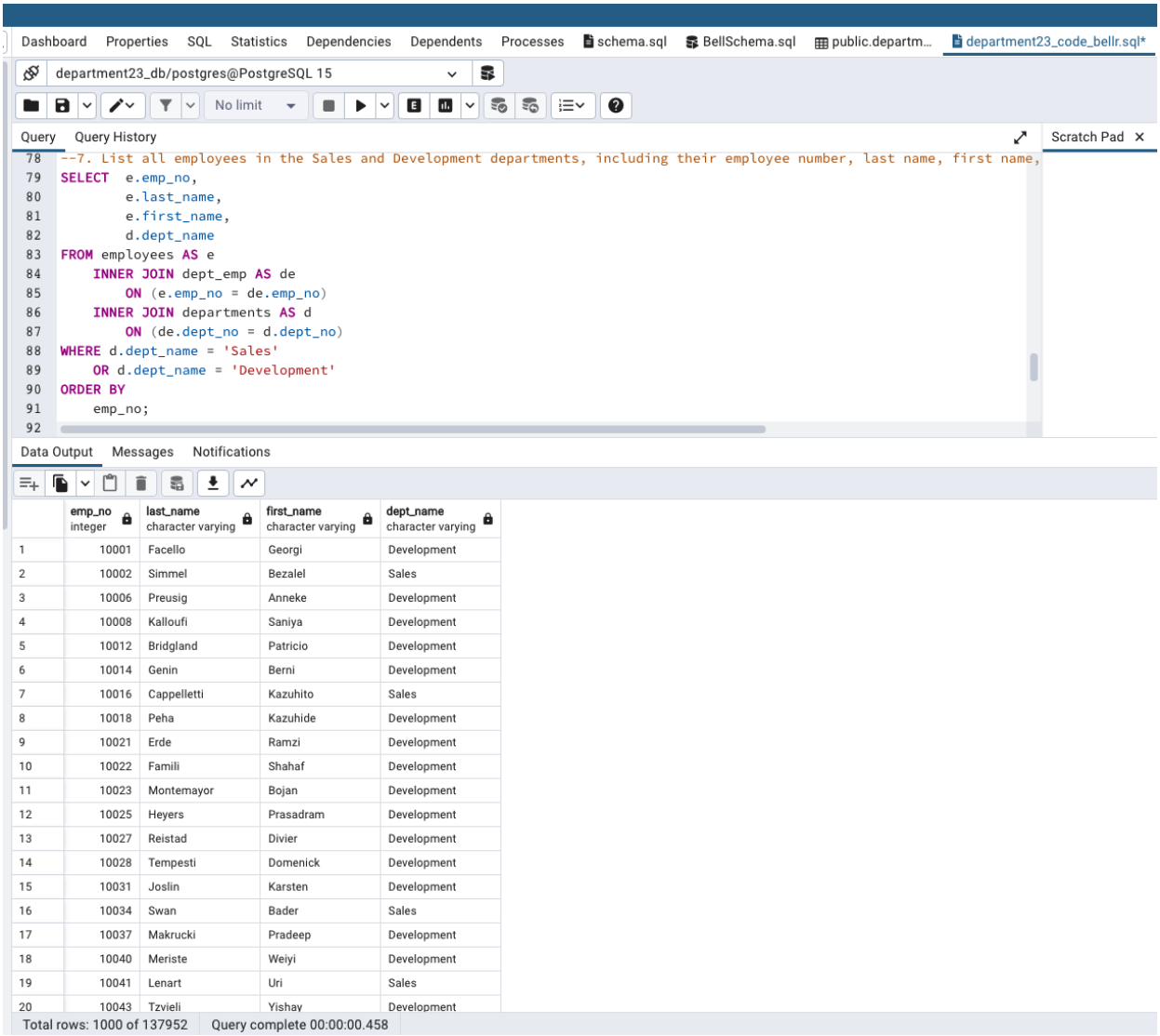
	first_name character varying	last_name character varying	sex character varying
1	Hercules	Baer	M
2	Hercules	Biron	F
3	Hercules	Birge	F
4	Hercules	Berstel	F
5	Hercules	Bernatsky	M
6	Hercules	Benantar	F
7	Hercules	Bernardinello	F
8	Hercules	Bahr	M
9	Hercules	Biran	F
10	Hercules	Barreiro	M
11	Hercules	Baak	M
12	Hercules	Bail	F
13	Hercules	Bodoff	M
14	Hercules	Basagni	M
15	Hercules	Baranowski	M
16	Hercules	Bisiani	F
17	Hercules	Benzmuller	M
18	Hercules	Bain	F
19	Hercules	Brendel	F
20	Hercules	Buchter	M

Here, we can see that there are twenty people with “Hercules” as a first name. Each of them has last names that start with the letter “B.”



	emp_no integer	last_name character varying	first_name character varying	dept_name character varying
1	10002	Simmel	Bezalel	Sales
2	10016	Cappelletti	Kazuhito	Sales
3	10034	Swan	Bader	Sales
4	10041	Lenart	Uri	Sales
5	10050	Dredge	Yinghua	Sales
6	10053	Zschoche	Sanjiv	Sales
7	10060	Billingsley	Breannda	Sales
8	10061	Herber	Tse	Sales
9	10068	Brattka	Charlene	Sales
10	10087	Eugenio	Xinglin	Sales
11	10088	Syrzycki	Jungsoon	Sales
12	10089	Flasterstein	Sudharsan	Sales
13	10093	Desikan	Sailaja	Sales
14	10095	Morton	Hilari	Sales
15	10099	Sullins	Valter	Sales
16	10101	Heyers	Perla	Sales
17	10107	Baca	Dung	Sales
18	10125	Hiltgen	Syozo	Sales
19	10136	Pintelas	Zissis	Sales
20	10148	Azumi	Douadi	Sales
21	10149	Perry	Xiadong	Sales
22	10151	Lichtner	Itzchak	Sales

This table simply captures a list of all of the employees within the Sales department.



	emp_no integer	last_name character varying	first_name character varying	dept_name character varying
1	10001	Facello	Georgi	Development
2	10002	Simmel	Bezalel	Sales
3	10006	Preusig	Anneke	Development
4	10008	Kalloufi	Saniya	Development
5	10012	Bridgland	Patricio	Development
6	10014	Genin	Berni	Development
7	10016	Cappelletti	Kazuhito	Sales
8	10018	Peha	Kazuhide	Development
9	10021	Erde	Ramzi	Development
10	10022	Famili	Shahaf	Development
11	10023	Montemayor	Bojan	Development
12	10025	Heyers	Prasadram	Development
13	10027	Reistad	Divier	Development
14	10028	Tempesti	Domenick	Development
15	10031	Joslin	Karsten	Development
16	10034	Swan	Bader	Sales
17	10037	Makrucki	Pradeep	Development
18	10040	Meriste	Weiyi	Development
19	10041	Lenart	Uri	Sales
20	10043	Tzvieli	Yishay	Development

In the table above, we went from departments to dept-employee to employees to extract this data. Inner joins were conducted on dept_no and on emp_no to produce this table.

```sql
--8. List the frequency count of employee last names (i.e., how many employees share each last name) in descending order.
SELECT
    last_name,
    count(*) as emp_counts
FROM
    employees
GROUP BY
    last_name
order by
    emp_counts desc
limit 20;
```

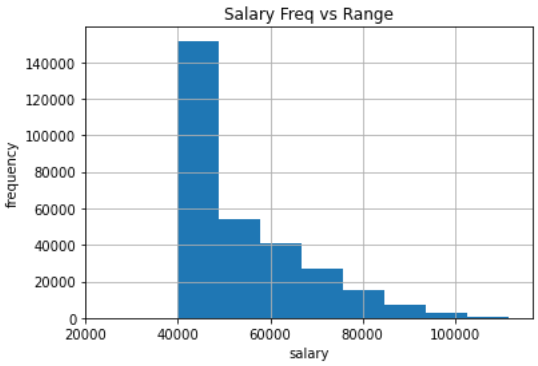
Data Output | Messages | Notifications

	last_name character varying	emp_counts bigint
1	Baba	226
2	Gelosh	223
3	Coorg	223
4	Sudbeck	222
5	Farris	222
6	Adachi	221
7	Osgood	220
8	Neiman	218
9	Masada	218
10	Mandell	218
11	Wendorf	217
12	Boudaillier	217
13	Mahnke	216
14	Cummings	216
15	Solares	216
16	Pettis	216
17	Maksimenko	215
18	Kulisch	215
19	Collette	215
20	Emmart	215

Total rows: 20 of 20 | Query complete 00:00:00.231

The table above shows a list of twenty last names, including the number of employees with a particular last name. According to this fictitious data, there are 226 employees whose last name is Baba.

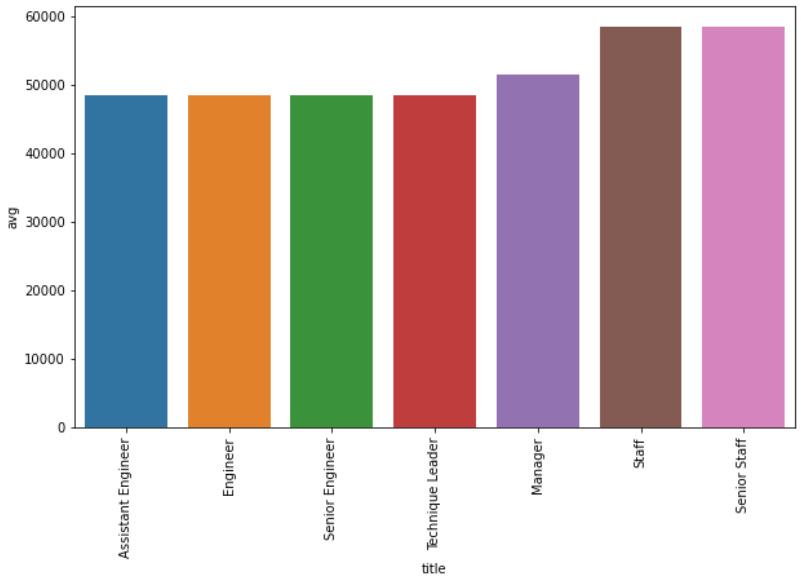
BONUS:
Create a histogram to visualize the most common salary ranges for employees




```
count    300024.000000
mean      52970.732451
std       14301.478491
min       40000.000000
25%       40000.000000
50%       48681.000000
75%       61758.000000
max      129492.000000
Name: salary, dtype: float64
```

The graph and table above show the minimum salary as $40,000, while the maximum salary is $129,492. Further, the table shows the other descriptive statistics. Notice that the majority of the salaries are between $40,000 and $48,000. Fewer salaries are above $100,000. We can do additional queries to determine the number of employees earning this amount and above, including the corresponding roles or titles. Based on the data, it appears that these salary ranges are attributed to executive level management. The bar chart below gives a snapshot shot of job titles, including salary ranges. Notice that the senior staff members are the top earners. This would suggest that executive level management are not included in this pool, representing the outliers seen in the histogram above.

Create a bar chart of average salary by title



The graph above shows the different job titles along with the average salaries across the Sales and Development departments.