SQL Query Script

* Data Cleaning --

**-- changing column name from id to emp id**

alter table hr

change column id emp\_id VARCHAR(20) null;

**-- Checking datatypes**

describe hr;

select birthdate from hr;

**-- turning off safemode in order to update data**

set sql\_safe\_updates = 0;

**-- cleaning inconsistent birthdates data**

update hr

set birthdate = case

when birthdate like '%/%' then date\_format(str\_to\_date(birthdate,'%m/%d/%Y'), '%Y-%m-%d')

when birthdate like '%-%' then date\_format(str\_to\_date(birthdate,'%m-%d- %Y'), '%Y-%m-%d')

else null

end;

**-- changing birthdate datatype from text to date**

alter table hr

modify column birthdate DATE;

**-- cleaning inconsistent hiredates data**

update hr

set hire\_date = case

when hire\_date like '%/%' then date\_format(str\_to\_date(hire\_date,'%m/%d/%Y'), '%Y-%m-%d')

when hire\_date like '%-%' then date\_format(str\_to\_date(hire\_date,'%m-%d- %Y'), '%Y-%m-%d')

else null

end;

**-- changing hire date datatype from text to date**

alter table hr

modify column hire\_date DATE;

**-- cleaning termdate column by keeping only date and removing time**

UPDATE hr

SET termdate = IF(termdate IS NOT NULL AND termdate != '', date(str\_to\_date(termdate, '%Y-%m-%d %H:%i:%s UTC')), '0000-00-00')

WHERE true;

SELECT termdate from hr;

**SET sql\_mode = 'ALLOW\_INVALID\_DATES';**

**-- changing termdate datatype from text to date**

alter table hr

modify column termdate DATE;

**-- adding a new collumn 'age' to hr table**

alter table hr add column age INT;

update hr

set age = timestampdiff(year,birthdate,curdate());

**-- checking for min & max age**

select min(age) as youngest, max(age) as oldest from hr

**-- removing employee details for emp with age below 18**

select count(\*) from hr where age <18;

delete from hr

where age <18;

* Questions --

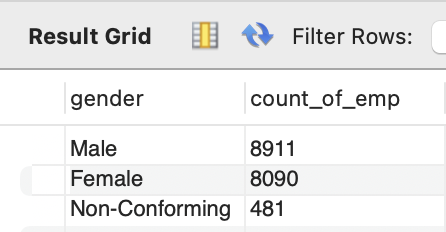
1. **What is the gender breakdown of employees in the company?**

select gender,count(emp\_id) as count\_of\_emp

from hr

where age >=18 and termdate = 0000-00-00

group by gender;



1. **What is the race/ethnicity breakdown of employees in the company?**

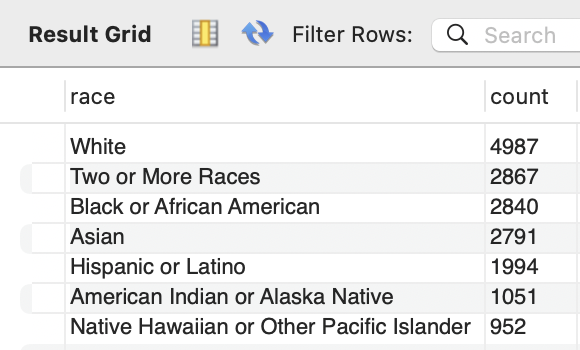
select race,count(\*) as count

from hr

where age >=18 and termdate = 0000-00-00

group by race

order by count(\*) desc;



1. **What is the age distribution of employees in the company?**

**(Age distribution)**

select case

when age >= 18 and age <= 24 then '18-24'

when age >= 25 and age <= 34 then '25-34'

when age >= 35 and age <= 44 then '35-44'

when age >= 45 and age <= 54 then '45-54'

when age >= 55 and age <= 64 then '55-64'

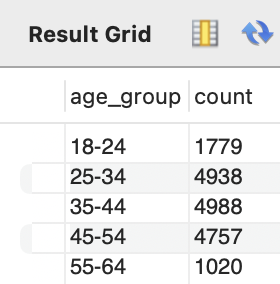
else '65+' end as age\_group, count(\*) as count

from hr

where age >=18 and termdate = 0000-00-00

group by age\_group

order by age\_group



**(Age distribution by Gender)**

select case

when age >= 18 and age <= 24 then '18-24'

when age >= 25 and age <= 34 then '25-34'

when age >= 35 and age <= 44 then '35-44'

when age >= 45 and age <= 54 then '45-54'

when age >= 55 and age <= 64 then '55-64'

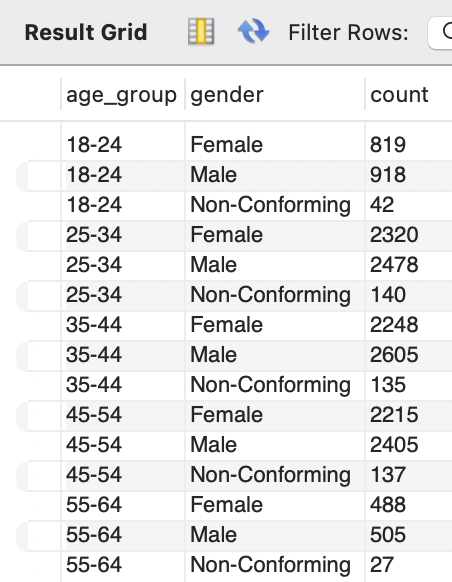
else '65+' end as age\_group,gender, count(\*) as count

from hr

where age >=18 and termdate = 0000-00-00

group by age\_group, gender

order by age\_group, gender



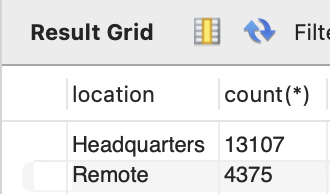
1. **How many employees work at headquarters versus remote locations?**

select location, count(\*)

from hr

where age >=18 and termdate = 0000-00-00

group by location

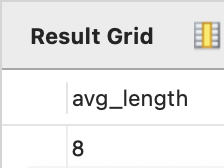


1. **What is the average length of employment for employees who have been terminated?**

select round(avg(datediff(termdate,hire\_date))/365,0) as avg\_length

from hr

where age >=18 and termdate != 0000-00-00 and termdate <= curdate()



1. **How does the gender distribution vary across departments and job titles?**

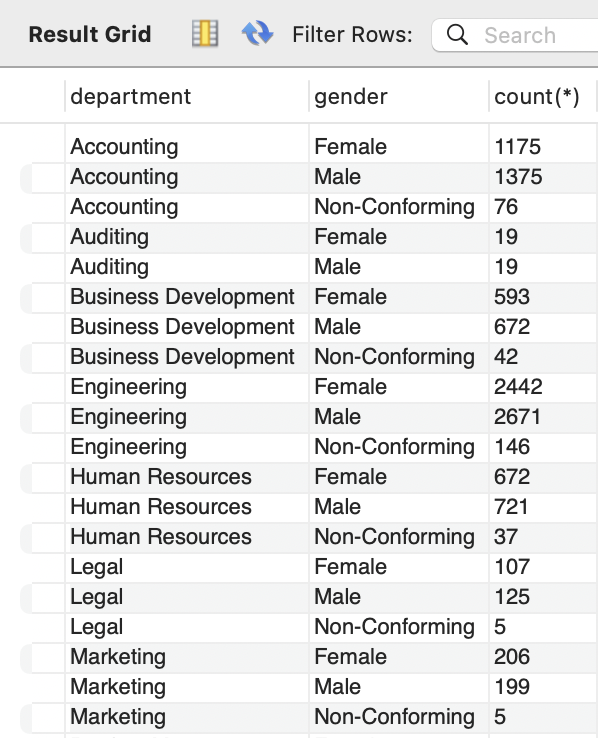
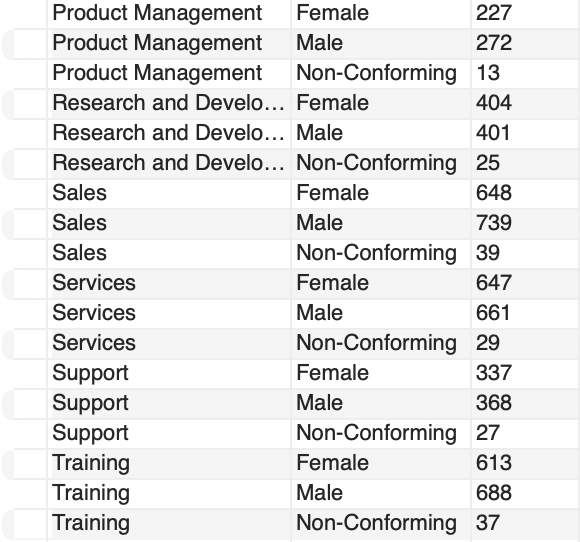
select department, gender, count(\*)

from hr

where age >=18 and termdate = 0000-00-00

group by department, gender

order by department, gender

1. **What is the distribution of job titles across the company?**

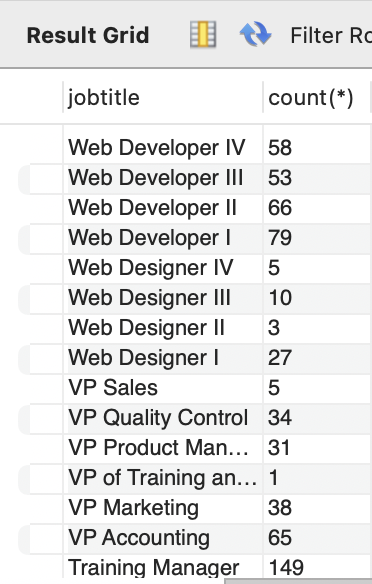
select jobtitle, count(\*)

from hr

where age >=18 and termdate = 0000-00-00

group by jobtitle

order by jobtitle desc



1. **Which department has the highest turnover rate?**

select department, total\_count, terminated\_count,

(terminated\_count/total\_count) as termination\_rate

from (

select department, count(\*) as total\_count,

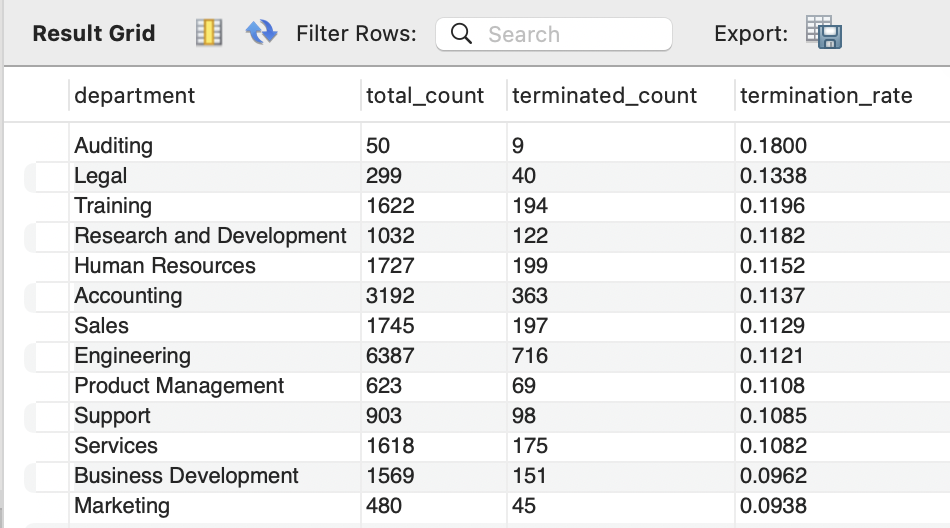
sum(case when termdate != '0000-00-00' and termdate <= curdate() then 1 else 0 end) as terminated\_count

from hr

where age >= 18

group by department) as subquery

order by termination\_rate desc



1. **What is the distribution of employees across locations by city and state?**

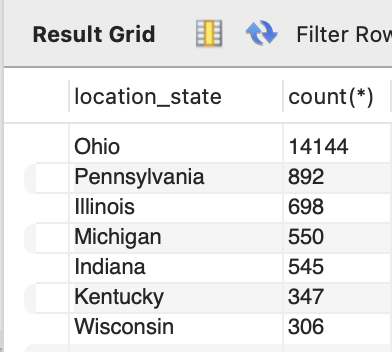
select location\_state, count(\*)

from hr

where age >=18 and termdate = 0000-00-00

group by location\_state

order by count(\*) desc



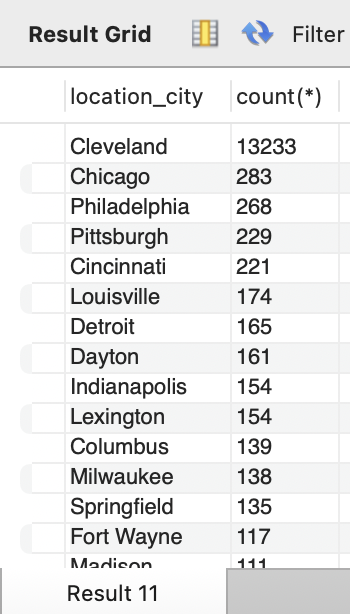
select location\_city, count(\*)

from hr

where age >=18 and termdate = 0000-00-00

group by location\_city

order by count(\*) desc



1. **How has the company's employee count changed over time based on hire and term dates?**

select year, hires, terminations, (hires-terminations) as net\_change,

round(((hires-terminations)/hires)\*100,2) as net\_change\_percent

from (

select year(hire\_date) as year,

count(\*)as hires,

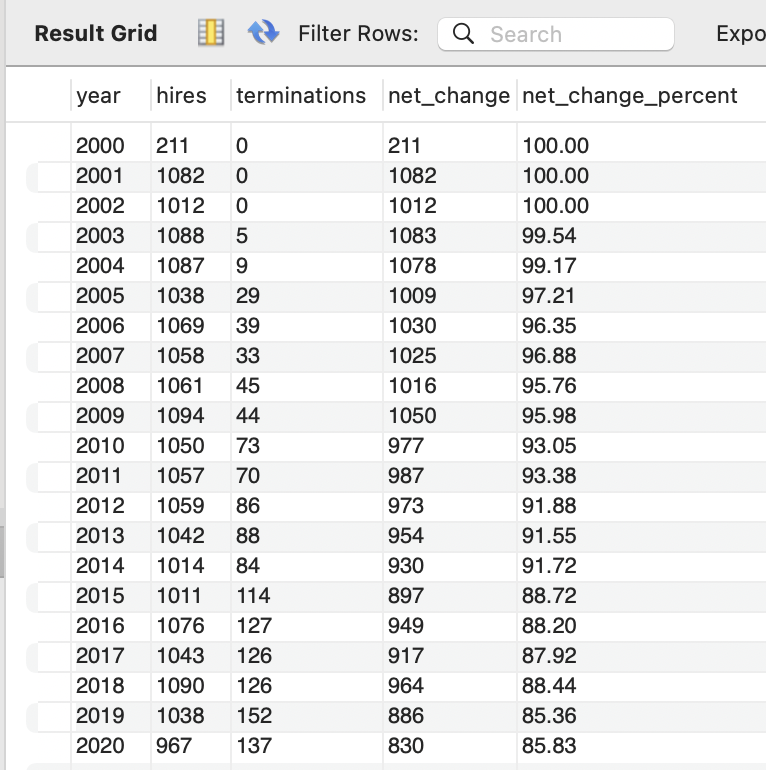
sum(case when termdate != 0000-00-00 and termdate >= curdate() then 1 else 0 end )as terminations

from hr

where age >= 18

group by year(hire\_date))as subquery

order by year asc



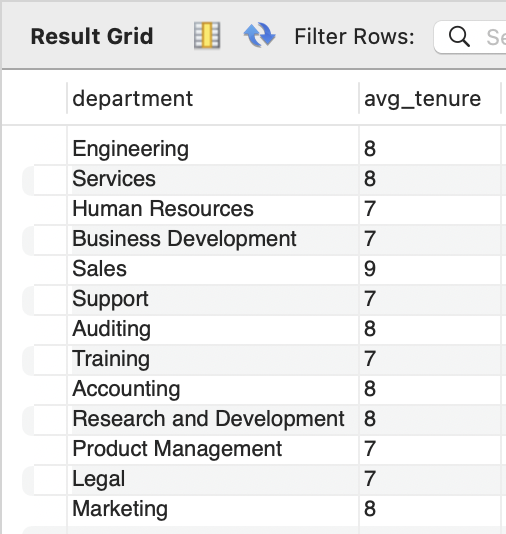
1. **What is the tenure distribution for each department?**

select department, round(avg(datediff(termdate,hire\_date))/365,0) as avg\_tenure

from hr

where age >=18 and termdate <= curdate() and termdate != '0000-00-00'

group by department

****