CREATE DATABASE IF NOT EXISTS projects;

SELECT \* FROM projects.hr;

-- COLUMN NAME CHANGE

ALTER TABLE hr

CHANGE COLUMN ï»¿id emp\_id VARCHAR(20) NULL ;

DESCRIBE hr;

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;

SET sql\_safe\_updates = 0;

ALTER TABLE hr

MODIFY COLUMN birthdate DATE;

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;

ALTER TABLE hr

MODIFY COLUMN hire\_date DATE;

UPDATE hr

SET termdate = date(str\_to\_date(termdate, "%Y-%m-%d %H:%i:%s UTC"))

WHERE termdate IS NOT NULL AND termdate != ' ';

SELECT termdate

FROM hr;

UPDATE hr

SET termdate = date(str\_to\_date(termdate,'%Y-%m-%d %H:%i:%s UTC'))

WHERE termdate IS NOT NULL AND termdate != ' ';

rollback;

UPDATE hr

SET termdate = date(str\_to\_date(termdate, '%Y-%m-%d %H:%i:%s UTC'), '0000-00-00')

WHERE termdate IS NOT NULL AND termdate != ' ';

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 ;

SET SQL\_MODE = ' ';

ALTER TABLE hr

MODIFY COLUMN termdate DATE;

ALTER TABLE hr

ADD COLUMN age INT;

SELECT \* FROM projects.hr;

UPDATE hr

SET age = timestampdiff(YEAR,birthdate, CURDATE());

select birthdate, age FROM hr;

SELECT

MIN(age) AS youngest,

MAX(age) AS oldest

FROM hr;

SELECT COUNT(\*) FROM hr WHERE age < 18;

-- QUESTIONS

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

SELECT gender, COUNT(\*) AS count

FROM hr

WHERE age >= 18 AND termdate = 0000-00-00

GROUP BY gender;

-- 2. 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;

-- 3. What is the age distribution of employees in the company?

SELECT

MIN(age) AS youngest,

MAX(age) AS oldest

FROM hr

WHERE age >= 18 AND termdate = 0000-00-00;

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;

-- 4. How many employees work at headquarters versus remote locations?

SELECT location, COUNT(\*) AS count

FROM hr

WHERE age >= 18 AND termdate = 0000-00-00

GROUP BY location

ORDER BY count;

-- 5. What is the average length of employment for employees who have been terminated?

SELECT

ROUND(

AVG(DATEDIFF(termdate,hire\_date) ) / 365) AS avg\_lengh\_employment

FROM hr

WHERE termdate <= curdate() AND termdate <> 0000-00-00 AND age >= 18;

-- 6. How does the gender distribution vary across departments and job titles?

SELECT department,gender,COUNT(\*) AS count

FROM hr

WHERE age >= 18 AND termdate = 0000-00-00

GROUP BY department, gender

ORDER BY department;

-- 7. What is the distribution of job titles across the company?

SELECT jobtitle, COUNT(\*) as COUNT

FROM hr

WHERE age >= 18 AND termdate = 0000-00-00

GROUP BY jobtitle

ORDER BY jobtitle DESC;

-- 8. 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 sub\_query

ORDER BY termination\_rate;

-- 9. What is the distribution of employees across locations by city and state?

SELECT

location\_state,COUNT(\*) AS count

FROM hr

WHERE age >= 18 AND termdate = 0000-00-00

GROUP BY location\_state

ORDER BY count DESC;

-- 10. 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 sub\_query

ORDER BY year ASC;

-- 11. What is the tenure distribution for each department?

SELECT department, round(avg(datediff(termdate,hire\_date)/365),0) AS avg\_tenure

FROM hr

WHERE termdate <= curdate() AND termdate <> 0000-00-00 AND age >= 18

GROUP BY department;