SELECT \* FROM hr.human\_resources;

----- Go to setting and change column name 'ï»¿id' into 'emp\_id' and 'termdate' into terminate\_date -----

DESCRIBE human\_resources;

SET sql\_safe\_updates=0;

UPDATE human\_resources

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;

ALTER TABLE human\_resources

MODIFY COLUMN birthdate DATE;

UPDATE human\_resources

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 human\_resources

MODIFY COLUMN hire\_date DATE;

UPDATE human\_resources

SET terminate\_date = DATE(STR\_TO\_DATE(terminate\_date, '%Y-%m-%d %H:%i:%s UTC'))

WHERE terminate\_date IS NOT NULL AND terminate\_date != '';

UPDATE human\_resources

SET terminate\_date = '0000-00-00'

WHERE terminate\_date = '';

ALTER TABLE human\_resources ADD COLUMN age INT;

UPDATE human\_resources SET age = TIMESTAMPDIFF(YEAR, birthdate, CURDATE());

-- QUESTIONS (Problem Statement)

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

SELECT gender, COUNT(\*) AS count

FROM human\_resources

WHERE age >= 18

GROUP BY gender;

-- 2. What is the race/ethnicity breakdown of employees in the company?

SELECT race, COUNT(\*) AS count

FROM human\_resources

WHERE age >= 18

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 human\_resources

WHERE age >= 18;

-- 3.1 age\_group

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

human\_resources

WHERE

age >= 18

GROUP BY age\_group

ORDER BY age\_group;

-- 3.2 age\_group\_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

human\_resources

WHERE

age >= 18

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 human\_resources

WHERE age >= 18

GROUP BY location;

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

SELECT ROUND(AVG(DATEDIFF(terminate\_date, hire\_date))/365,0) AS avg\_length\_of\_employment

FROM human\_resources

WHERE terminate\_date <> '0000-00-00' AND terminate\_date <= CURDATE() AND age >= 18;

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

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

FROM human\_resources

WHERE age >= 18

GROUP BY department, gender

ORDER BY department;

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

SELECT jobtitle, COUNT(\*) as count

FROM human\_resources

WHERE age >= 18

GROUP BY jobtitle

ORDER BY jobtitle DESC;

-- 8. Which department has the highest turnover rate?

SELECT department, COUNT(\*) as total\_count,

SUM(CASE WHEN terminate\_date <= CURDATE() AND terminate\_date <> '0000-00-00' THEN 1 ELSE 0 END) as terminated\_count,

SUM(CASE WHEN terminate\_date = '0000-00-00' THEN 1 ELSE 0 END) as active\_count,

(SUM(CASE WHEN terminate\_date <= CURDATE() THEN 1 ELSE 0 END) / COUNT(\*)) as termination\_rate

FROM human\_resources

WHERE age >= 18

GROUP BY department

ORDER BY termination\_rate DESC;

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

SELECT location\_state, COUNT(\*) as count

FROM human\_resources

WHERE age >= 18

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 terminate\_date <> '0000-00-00' AND terminate\_date <= CURDATE() THEN 1 ELSE 0 END) AS terminations

FROM

human\_resources

WHERE age >= 18

GROUP BY

YEAR(hire\_date)

) subquery

ORDER BY

year ASC;

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

SELECT department, ROUND(AVG(DATEDIFF(CURDATE(), terminate\_date)/365),0) as avg\_tenure

FROM human\_resources

WHERE terminate\_date <= CURDATE() AND terminate\_date <> '0000-00-00' AND age >= 18

GROUP BY department