**HR data project SQL queries**

**Data cleaning**

CREATE DATABASE projects;

USE projects;

SELECT \* FROM hr;

ALTER TABLE hr

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

DESCRIBE hr;

-- changing birth day to date format, making all dates look the same

SET sql\_safe\_updates = 0;

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;

Select birthdate from hr;

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;

Select termdate from hr;

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 != '';

ALTER TABLE hr

MODIFY COLUMN termdate DATE;

ALTER TABLE hr

ADD COLUMN age INT;

UPDATE hr

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

Select birthdate,age from hr;

SELECT

MIN(age) as youngest,

MAX(age) as oldest

FROM hr;

**QUESTIONS**

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

SELECT gender, count(\*) AS Count

FROM hr

WHERE age >= 18 AND termdate = ''

GROUP BY gender

ORDER BY count

;

A screenshot of a computer

Description automatically generated

-- 2) what is race/ethnicity breakdown?

SELECT race, count(\*) AS Count

FROM hr

WHERE age >= 18 AND termdate = ''

GROUP BY race

ORDER BY count DESC

;

A screenshot of a computer

Description automatically generated

-- 3) what is the age distirbution of employees in the company?

SELECT

min(age) AS youngest,

max(age) AS oldest

FROM hr

WHERE age >= 18 AND termdate = ''

;

A screenshot of a computer

Description automatically generated

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'

ELSE '55+'

END AS age\_group,

COUNT(\*) AS COUNT

FROM hr

WHERE age >= 18 AND termdate = ''

GROUP BY age\_group

ORDER BY age\_group

;

A screenshot of a computer

Description automatically generated

-- adding 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'

ELSE '55+'

END AS age\_group,

gender,

COUNT(\*) AS COUNT

FROM hr

WHERE age >= 18 AND termdate = ''

GROUP BY age\_group, gender

ORDER BY age\_group, gender

;

A screenshot of a computer

Description automatically generated

-- 4) how many employees work at HQ and remote?

SELECT

location, count(\*) as count

FROM hr

WHERE age >= 18 AND termdate = ''

GROUP BY location

;

A screenshot of a computer

Description automatically generated

-- 5) AVG length of emplyment for employees who terminated?

SELECT

AVG(datediff(termdate, hire\_date))/365 AS avg\_length\_employment\_in\_years

FROM hr

WHERE termdate <= curdate() AND termdate <> ''''

;

A screenshot of a computer

Description automatically generated

-- 6) How does the gender distributon vary across departments?

SELECT department,

gender,

count(\*) AS Count

FROM hr

WHERE age >= 18 AND termdate = ''

GROUP BY gender, department

ORDER BY department

;

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

-- 7) How does the gender distributon vary across jobtitle?

SELECT jobtitle,

gender,

count(\*) AS Count

FROM hr

WHERE age >= 18 AND termdate = ''

GROUP BY gender, jobtitle

ORDER BY jobtitle

;

A screenshot of a computer

Description automatically generated

-- 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<>'' 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

;

A screenshot of a computer

Description automatically generated

-- 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 = ''

GROUP BY location\_state

ORDER BY COUNT DESC

;

A screenshot of a computer

Description automatically generated

-- 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 <> '' 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;

A screenshot of a data

Description automatically generated

-- 11) how many years employees stay in 1 department?

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

FROM hr

WHERE termdate <= curdate() AND termdate <> '' and age >= 18

GROUP BY department;

A screenshot of a computer

Description automatically generated