MYSQL QUERIES

➢ *DATA CLEANING*

CREATE DATABASE human_resource;

USE human_resource;

SELECT * FROM HR;

ALTER TABLE HR

CHANGE COLUMN ï»¿id emp_id VARCHAR(100) NULL;

DESCRIBE HR;

SELECT birthdate FROM hr;

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;


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;

ALTER TABLE HR
CHANGE COLUMN hire_date hiredate 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());
```

➢ QUERIES FOR ANALYSIS

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

```
use human_resource;

set sql_mode='allow_invalid_dates';

select gender,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(*)
 from HR
 where age>=18 and termdate='0000-00-00'
 group by race
order by race;
```

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';

// Run this get a idea of ages you are working with then perform the query beneath

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 total

from HR

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

group by age_group

order by age_group;

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

select location, gender,count(*) as total

from HR

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

group by gender, location;

5. What is the average employment period for terminated employees?

select round(avg(datediff(termdate,hiredate))/365,0) as avg_employement_duration

from HR

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

6. How does the gender distribution vary across department?

select department, gender, count(*) as total

from HR

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

group by department, gender;

7. Show the distribution of jobs titles across the company?

Select jobtitles, gender, count(*) as total

from HR

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

group by jobtitles, gender

order by jobtitles;

8. Which department has the highest retention rate?

select department, total_count, terminated_count,
terminated_count/total_count as termination_rate,

(total_count-terminated_count)/total_count as retention_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

order by termination_rate;

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

select location_state,location_city, count(*) as total

from HR

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

group by location_state,location_city

order by location_state;

10. How does the company's employee count vary over time with respect to hire date and term date?

select year, hires,terminations, hires-terminations as net_change, round((hires-terminations)/hires 100,2) as net_change_percentage*

from (select

year(hiredate) 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(hiredate)) as sub

order by year;

11. What is the tenure distribution for each department?

Select department,round(avg(datediff(termdate,hiredate)/365),0) as avg_tenure

From HR

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

Group by department;

Side notes:

set sql_mode='allow_invalid_dates';

Set sql_safe_updates=0;

 The 'I' is used for minutes as m signifies month