Query 1 | mavenconsulting_vApril2022 | 2020 | 2020 | 2019

Limit to 1000 rows

```
1   -- 1st task: change column name
2   ALTER TABLE hr
3   CHANGE COLUMN id emp_id VARCHAR (20) NULL
4   -- I could have also run this query to change columnname in MYSQL
5   ALTER TABLE hr
6   RENAME COLUMN emp_id TO id
```

Limit to 1000 rows

```
7   /* 2nd task: DESCRIBE tablename to see all the data types in our table. I start with using the DATE FORMAT function
8   to convert the birthdate and hire_date columns into the default MYSQL date format which is YYYY-MM-DD, and then I change the data type of the
9   birthdate and hire_date columns from TEXT to DATE using ALTER TABLE hr MODIFY COLUMN columnname DATE*/
10  UPDATE hr
11  SET birthdate = CASE
12  WHEN birthdate LIKE '%/%' THEN date_format(birthdate, '%Y-%m-%d')
13  WHEN birthdate LIKE '%-%' THEN date_format(birthdate, '%Y-%m-%d')
14  ELSE NULL
15  END
16
17  ALTER TABLE hr
18  MODIFY COLUMN birthdate DATE
19  DESCRIBE hr
20
21  UPDATE hr
22  SET hire_date = CASE
23  WHEN hire_date LIKE '%/%' THEN date_format(birthdate, '%Y-%m-%d')
24  WHEN hire_date LIKE '%-%' THEN date_format(birthdate, '%Y-%m-%d')
25  ELSE NULL
26  END
27
28  ALTER TABLE hr
29  MODIFY COLUMN hire_date DATE
30  DESCRIBE hr
```

Limit to 1000 rows

```
32  /* below, I am going to convert the termination date (termdate) column into the default date format. I do not need the TIMESTAMP
33  or DATETIME, so I will run the following query:*/
34  UPDATE hr
35  SET termdate = IF(termdate IS NOT NULL AND termdate != ' ', date(date_format(termdate, '%Y-%m-%d %H:%i:%s UTC')),
36  '0000-00-00') WHERE true;
37  SELECT termdate FROM hr
38
39  ALTER TABLE hr
40  MODIFY COLUMN termdate DATE
41  DESCRIBE hr
42
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content: | Fetch rows:

termdate
0000-00-00
0000-00-00
0000-00-00
2018-07-01
2027-02-01

Result Grid

```sql
-- add another column named AGE
ALTER TABLE hr
ADD COLUMN age INTEGER
-- the column age has no values, so I am going to CALCULATE the AGE
UPDATE hr
SET age = timestampdiff(YEAR, birthdate, CURDATE())
      birthdate, age      hr

```

Result Grid | Filter Rows: | Export: | Wrap Cell Content: | Fetch rows:

birthdate	age
1991-06-04	32
1984-06-29	39
1989-07-29	34
1982-09-14	41
1994-04-11	29

```sql
-- QUESTIONS
-- 1. What is the gender breakdown of employees in the company?
select COUNT(emp_id) as number_of_employees, gender from hr
where age >= 18 and termdate = '0000-00-00'
group by gender

```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

number_of_employees	gender
9328	Male
8455	Female
502	Non-Conforming

```sql
-- 2. What is the race/ethnicity breakdown of employees in the company?
select count(*) AS number_of_employees, race from hr
where age >= 18 and termdate ='0000-00-00'
group by race
order by count(*) DESC

```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

number_of_employees	race
5214	White
2989	Two or More Races
2983	Black or African American
2936	Asian
2074	Hispanic or Latino

```sql
-- 3. What is the age distribution of employees in the company? First, I calculate the youngest and oldest employee, then
-- using COUNT CASE WHEN  calculate the number of employees for each age_group. It is worth noting that the CASE WHEN
-- corresponding function in Excel and Power BI is SWITCH TRUE()
select MIN(age) AS youngest_employee, MAX(age) AS oldest_employee from hr
where age >= 18 and termdate = '0000-00-00'
```

```sql
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 number_of_employees
FROM hr
WHERE age >=18 AND termdate = '0000-00-00'
GROUP BY age_group, gender
ORDER BY age_group, gender
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

age_group	gender	number_of_employees
18-24	Female	894
18-24	Male	1028
18-24	Non-Conforming	50
25-34	Female	2364
25-34	Male	2489

```sql
-- I COULD HAVE ANSWERED THIS QUESTION TOO WITH THE SUPER-POWERFUL 'COUNT CASE WHEN' function, absolutely my favourite one
select gender,
count(case when age >=18 AND age <=24 then emp_id else null end) AS number_of_emp_id_for_age_group_18_24,
count(case when age >=25 AND age <=34 then emp_id else null end) AS number_of_emp_id_for_age_group_25_34,
count(case when age >=35 AND age <=44 then emp_id else null end) AS number_of_emp_id_for_age_group_35_44,
count(case when age >=45 AND age <=54 then emp_id else null end) AS number_of_emp_id_for_age_group_45_54,
count(case when age >=55 AND age <=64 then emp_id else null end) AS number_of_emp_id_for_age_group_55_64
from hr
where age >=18 AND termdate = '0000-00-00'
group by gender
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

gender	number_of_emp_id_for_age_group_18_24	number_of_emp_id_for_age_group_25_34	number_of_emp_id_for_age_group_35_44	number_of_emp_id_for_ag
Male	1028	2489	2620	2493
Female	894	2364	2226	2308
Non-Conforming	50	135	139	138

```sql
-- 4. How many employees work at headquarters versus remote locations?
select location, count(emp_id) from hr
where age >=18 and termdate = '0000-00-00'
group by location
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

location	count(emp_id)
Headquarters	13710
Remote	4575

```sql
-- 5. What is the average length of employment for employees who have been terminated? I used the datediff function to
-- calculate the difference between 2 dates, in this case termdate and hire_date, and then divide the output by 365 to be able
-- get the number of years. In addition, because I want employees who have been terminated I will then filter out the termdate
-- column that has to be <= the current date
select round(avg(datediff(termdate, hire_date)) / 365,0) AS avg_length_employment  from hr
where age >=18 and termdate <= curdate() and termdate != '0000-00-00'
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

avg_length_employment
31

```sql
-- 6. How does the gender distribution vary across departments and job titles? I have found two ways to answer this query
-- first solution
select department, jobtitle,
count(case when gender = 'Female' THEN gender ELSE NULL END) AS number_of_female_employees,
count(case when gender = 'Male' THEN gender ELSE NULL END) AS number_of_male_employees,
count(case when gender = 'Non-Conforming' THEN gender ELSE NULL END) AS number_of_non_conforming_employees
from hr
where age >=18 and termdate = '0000-00-00'
group by department, jobtitle
order by department
-- second solution
select department, gender, count(*) AS number_of_employees from hr
where age >=18 and termdate = '0000-00-00'
group by department, gender
order by department
```

```sql
-- 7. What is the distribution of job titles across the company?
select jobtitle, count(*) AS number_of_employees from hr
where age >=18 and termdate = '0000-00-00'
group by jobtitle
order by jobtitle
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

jobtitle	number_of_employees
Account Coordinator	2
Account Executive	409
Account Manager	193
Accountant I	65
Accountant II	71

```sql
-- 8. What is the distribution of employees across locations by city and state?
select count(emp_id) AS number_of_employees, location_state, location_city from hr
where age >=18 and termdate = '0000-00-00'
group by location_state, location_city
order by count(emp_id) desc
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

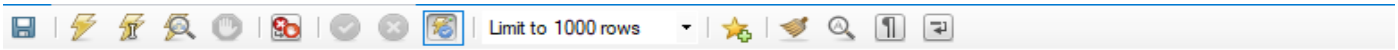
number_of_employees	location_state	location_city
13841	Ohio	Cleveland
294	Illinois	Chicago
278	Pennsylvania	Philadelphia
238	Pennsylvania	Pittsburgh
227	Ohio	Cincinnati

```sql
-- 09. What is the tenure distribution for each department? (How long employees stay in each department before they quit
-- or are sacked)?
select department, round(avg(datediff(termdate, hire_date) / 365),0) AS average_length_of_employment
from hr
where termdate <= curdate() and termdate != '0000-00-00' and age >=18
group by department
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

department	average_tenure
Engineering	31
Services	32
Human Resources	30
Business Development	30
Sales	32


```sql
-- 10. Which department has the highest turnover rate?
select department, number_of_employees, count_of_terminations,
count_of_terminations / number_of_employees AS termination_rate
from (
select department, count(*) AS number_of_employees,
sum(case when termdate != '0000-00-00' and termdate <=curdate() then 1 else 0 end) AS count_of_terminations
from hr
where age >= 18
group by department
) AS subquery
order by termination_rate desc
```




```sql
-- 11. 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 <= curdate() AND termdate <> '0000-00-00' THEN 1 ELSE 0 END) AS terminations
FROM hr
WHERE age >= 18
GROUP BY YEAR(HIRE_DATE)
) AS subquery
order by year ASC
```