Questions

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

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

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

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

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

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

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

8. Which department has the highest turnover rate?

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

10. How has the company's employee count changed over time based on hire and term dates?

1. What is the gender breakdown of employees in the company?
2. %%sql
3. SELECT gender, COUNT(\*) AS count
4. FROM hr
5. WHERE age >= 18
6. GROUP BY gender

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

%%sql

SELECT race, COUNT(\*) AS count

FROM hr

WHERE age >= 18

GROUP BY race

ORDER BY count DESC

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

%%sql

SELECT

  MIN(age) AS youngest,

  MAX(age) AS oldest

FROM hr

WHERE age >= 18

%%sql

SELECT FLOOR(age/10)\*10 AS age\_group, COUNT(\*) AS count

FROM hr

WHERE age >= 18

GROUP BY FLOOR(age/10)\*10;

%%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,

  COUNT(\*) AS count

FROM

  hr

WHERE

  age >= 18

GROUP BY age\_group

ORDER BY age\_group;

%%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 count

FROM

  hr

WHERE

  age >= 18

GROUP BY age\_group, gender

ORDER BY age\_group, gender;

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

%%sql

SELECT location, COUNT(\*) as count

FROM hr

WHERE age >= 18

GROUP BY location;

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

%%sql

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

FROM hr

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

%%sql

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

FROM hr

WHERE termdate <= CURDATE() AND age >= 18;

6. How does the gender distribution vary across departments?

%%sql

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

FROM hr

WHERE age >= 18

GROUP BY department, gender

ORDER BY department;

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

%%sql

SELECT jobtitle, COUNT(\*) as count

FROM hr

WHERE age >= 18

GROUP BY jobtitle

ORDER BY jobtitle DESC;

8. Which department has the highest turnover rate?

"Turnover rate" typically refers to the rate at which employees leave a company or department and need to be replaced. It can be calculated as the number of employees who leave over a given time period divided by the average number of employees in the company or department over that same time period.

%%sql

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

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

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

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

FROM hr

WHERE age >= 18

GROUP BY department

ORDER BY termination\_rate DESC;

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

%%sql

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

FROM hr

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?

This query groups the employees by the year of their hire date and calculates the total number of hires, terminations, and net change (the difference between hires and terminations) for each year. The results are sorted by year in ascending order.

%%sql

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,

    COUNT(\*) - SUM(CASE WHEN termdate <> '0000-00-00' AND termdate <= CURDATE() THEN 1 ELSE 0 END) AS net\_change,

    ROUND(((COUNT(\*) - SUM(CASE WHEN termdate <> '0000-00-00' AND termdate <= CURDATE() THEN 1 ELSE 0 END)) / COUNT(\*) \* 100),2) AS net\_change\_percent

FROM

    hr

WHERE age >= 18

GROUP BY

    YEAR(hire\_date)

ORDER BY

    YEAR(hire\_date) ASC;

In this modified query, a subquery is used to first calculate the terminations alias, which is then used in the calculation for the net\_change and net\_change\_percent column in the outer query.

%%sql

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)

) subquery

ORDER BY

    year ASC;

11. What is the tenure distribution for each department?

How long do employees work in each department before they leave or are made to leave?

%%sql

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

FROM hr

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

GROUP BY department

Summary of Findings

- There are more male employees

- White race is the most dominant while Native Hawaiian and American Indian are the least dominant.

- The youngest employee is 20 years old and the oldest is 57 years old

- 5 age groups were created (18-24, 25-34, 35-44, 45-54, 55-64). A large number of employees were between 25-34 followed by 35-44 while the smallest group was 55-64.

- A large number of employees work at the headquarters versus remotely.

- The average length of employment for terminated employees is around 7 years.

- The gender distribution across departments is fairly balanced but there are generally more male than female employees.

- The Marketing department has the highest turnover rate followed by Training. The least turn over rate are in the Research and development, Support and Legal departments.

- A large number of employees come from the state of Ohio.

- The net change in employees has increased over the years.

- The average tenure for each department is about 8 years with Legal and Auditing having the highest and Services, Sales and Marketing having the lowest.

Limitations

- Some records had negative ages and these were excluded during querying(967 records). Ages used were 18 years and above.

- Some termdates were far into the future and were not included in the analysis(1599 records). The only term dates used were those less than or equal to the current date.

**Data** - HR Data with over 22000 rows from the year 2000 to 2020.

**Data Cleaning & Analysis** - MySQL Workbench

**Data Visualization** - PowerBI