QUESTIONS TO ANSWER FROM THE DATA

1) What's the age distribution in the company?

age distribution

```
SELECT
 MIN(age) AS youngest,
 MAX(age) AS OLDEST
FROM hr_data;
```

age group count

```
SELECT age_group,
count(*) AS count
FROM
(SELECT
 CASE
  WHEN age <= 21 AND age <= 30 THEN '21 to 30'
  WHEN age <= 31 AND age <= 40 THEN '31 to 40'
  WHEN age <= 41 AND age <= 50 THEN '41 to 50'
  ELSE '50+'
  END AS age_group
 FROM hr_data
 WHERE new_termdate IS NULL
 ) AS subquery
GROUP BY age_group
ORDER BY age_group;
```

Age group by gender

```
SELECT age_group,
gender,
count(*) AS count
FROM
(SELECT
 CASE
  WHEN age <= 21 AND age <= 30 THEN '24 to 30'
```

```
 WHEN age <= 31 AND age <= 40 THEN '31 to 40'
 WHEN age <= 41 AND age <= 50 THEN '41 to 50'
 ELSE '50+'
 END AS age_group,
 gender
FROM hr_data
WHERE new_termdate IS NULL
) AS subquery
GROUP BY age_group, gender
ORDER BY age_group, gender;
```

2) What's the gender breakdown in the company?

```
SELECT
gender,
 COUNT(gender) AS count
FROM hr_data
WHERE new_termdate IS NULL
GROUP BY gender
ORDER BY gender ASC;
```

3) How does gender vary across departments and job titles?

```
SELECT
department,
gender,
count(gender) AS count
FROM hr_data
WHERE new_termdate IS NULL
GROUP BY department, gender,
ORDER BY department, gender ASC;
```

job titles

```
SELECT
department, jobtitle,
gender,
count(gender) AS count
FROM hr_data
WHERE new_termdate IS NULL
GROUP BY department, jobtitle, gender
ORDER BY department, jobtitle, gender ASC;
```

4) What's the race distribution in the company?

```
SELECT
race,
count(*) AS count
FROM
hr_data
WHERE new_termdate IS NULL
GROUP BY race
ORDER BY count DESC;
```

5) What's the average length of employment in the company?

```
SELECT
AVG(DATEDIFF(year, hire_date, new_termdate)) AS tenure
FROM hr_data
WHERE new_termdate IS NOT NULL AND new_termdate <= GETDATE();
```

6) Which department has the highest turnover rate?

- get total count
- get terminated count
- terminated count/total count

```
SELECT
 department,
 total_count,
 terminated_count,
 (round((CAST(terminated_count AS FLOAT)/total_count), 2)) * 100 AS turnover_rate
 FROM
        (SELECT
         department,
         count(*) AS total_count,
         SUM(CASE
                WHEN new_termdate IS NOT NULL AND new_termdate <= GETDATE() THEN 1 ELSE 0
                END
                ) AS terminated_count
        FROM hr_data
        GROUP BY department
        ) AS subquery
ORDER BY turnover_rate DESC;
```

7) What is the tenure distribution for each department?

```
SELECT
   department,
   AVG(DATEDIFF(year, hire_date, new_termdate)) AS tenure
FROM
   hr_data
WHERE
   new_termdate IS NOT NULL
   AND new_termdate <= GETDATE()
GROUP BY
   department;
```

8) How many employees work remotely for each department?

```
SELECT
 location,
 count(*) as count
FROM hr_data
WHERE new_termdate IS NULL
GROUP BY location;
```

9) What's the distribution of employees across different states?

```
SELECT
 location_state,
 count(*) AS count
FROM hr_data
WHERE new_termdate IS NULL
GROUP BY location_state
ORDER BY count DESC;
```

10) How are job titles distributed in the company?

```
SELECT
 jobtitle,
 count(*) AS count
 FROM hr_data
 WHERE new_termdate IS NULL
 GROUP BY jobtitle
 ORDER BY count DESC;
```

11) How have employee hire counts varied over time?

- calculate hires

- calculate terminations
- (hires-terminations)/hires percent hire change

```
SELECT
 hire_year,
 hires,
 terminations,
 hires - terminations AS net_change,
 (round(CAST(hires-terminations AS FLOAT)/hires, 2)) * 100 AS percent_hire_change
 FROM
        (SELECT
         YEAR(hire_date) AS hire_year,
         count(*) AS hires,
         SUM(CASE
                WHEN new_termdate is not null and new_termdate <= GETDATE() THEN 1 ELSE 0
                        END
                        ) AS terminations
        FROM hr_data
        GROUP BY YEAR(hire_date)
        ) AS subquery
ORDER BY percent_hire_change ASC;
```