```sql
CREATE DATABASE hur;
USE hur;


SELECT *
FROM hr_data;

SELECT termdate
FROM hr_data
ORDER BY termdate DESC;

UPDATE hr_data
SET termdate = FORMAT(CONVERT(DATETIME, LEFT(termdate, 19), 120), 'yyyy-MM-
  dd');


ALTER TABLE hr_data
ADD new_termdate DATE;

-- copy converted time values from termdate to new_termdate

UPDATE hr_data
SET new_termdate = CASE
 WHEN termdate IS NOT NULL AND ISDATE(termdate) = 1 THEN CAST(termdate AS
   DATETIME) ELSE NULL END;

 -- create a new column called "age"

ALTER TABLE hr_data
ADD age nvarchar(50);

-- populate age column
UPDATE hr_data
SET age=DATEDIFF(YEAR, birthdate, GETDATE());


-- Exploratory Data Analysis

-- QUESTIONS TO BE ANSWERED FROM THE DATA:

-- 1.  What's the age distribution in the company?

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

--SELECT age
--FROM hr_data
--ORDER BY age; (This query is used to see ages of employees)

SELECT age_group,
count(*) AS count
```

```sql
FROM
(SELECT
 CASE
  WHEN age <= 22 AND age <=30 THEN '22 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 <= 22 AND age <=30 THEN '22 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 is 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;

-- 3.  How does gender vary across departments and job titles?

      -- gender by department
SELECT
department,
gender,
COUNT(gender) AS count
FROM hr_data
```

```
WHERE new_termdate IS NULL
GROUP BY department, gender
ORDER BY department, gender;

      -- gender by 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;

-- 4.  What is 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 tota 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
ORDER BY tenure DESC;


-- 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?

SELECT
hire_yr,
hires,
terminations,
hires - terminations AS net_change,
```

```
(hires - terminations)/hires AS percent_hire_change
FROM
  (SELECT
  YEAR(hire_date) AS hire_yr,
  count(*) as hires,
  SUM(CASE WHEN new_termdate IS NOT NULL AND new_termdate <= GETDATE() THEN 1
    ELSE 0 END) terminations
  FROM hr_data
  GROUP BY year(hire_date)
  ) AS subquery
ORDER BY percent_hire_change ASC;

-- fixes zero values from the above query
SELECT
    hire_yr,
    hires,
    terminations,
    hires - terminations AS net_change,
    (round(CAST(hires - terminations AS FLOAT) / NULLIF(hires, 0), 2)) *100 AS
      percent_hire_change
FROM
    (SELECT
        YEAR(hire_date) AS hire_yr,
        COUNT(*) AS hires,
        SUM(CASE WHEN new_termdate IS NOT NULL AND new_termdate <= GETDATE()
          THEN 1 ELSE 0 END) terminations
    FROM hr_data
    GROUP BY YEAR(hire_date)
    ) AS subquery
ORDER BY hire_yr ASC;
```