Data Analysis

- **Selecting a current Database.**

```
USE PROJECT;
```

#What is the gender breakdown of employees in the company?

```
SELECT GENDER, COUNT(*) AS COUNT
FROM HR
WHERE AGE >= 18 AND TERMDATE IS NULL
GROUP BY GENDER;
```

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

```
SELECT
	RACE, COUNT(*) AS COUNT
FROM
	HR
WHERE
	AGE >= 18 AND TERMDATE IS NULL
GROUP BY RACE
ORDER BY COUNT DESC;
```

#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 IS NULL;
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 AND TERMDATE IS NULL
GROUP BY AGE_GROUP
ORDER BY AGE_GROUP;

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 AND TERMDATE IS NULL
GROUP BY AGE_GROUP, GENDER
ORDER BY AGE_GROUP, GENDER;
```

 # How many employees work at headquarters versus remote locations?

```
SELECT LOCATION, COUNT(*) AS COUNT
FROM HR
WHERE AGE >= 18 AND TERMDATE IS NULL
GROUP BY LOCATION;
```

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

```
SELECT ROUND(AVG(DATEDIFF(TERMDATE, HIRE_DATE))/365,0) AS AVG_LENGTH_OF_EMPLOYMENT
FROM HR
WHERE TERMDATE IS NOT NULL    AND TERMDATE <= CURDATE() AND AGE >= 18;
```

#How does the gender distribution vary across departments?

```
 SELECT DEPARTMENT, GENDER, COUNT(*) AS COUNT
FROM HR
WHERE AGE >= 18 AND TERMDATE IS NULL
GROUP BY DEPARTMENT, GENDER
ORDER BY DEPARTMENT;
```

#What is the distribution of job titles across the company?

```
 SELECT JOBTITLE, COUNT(*) AS COUNT
FROM HR
WHERE AGE >= 18 AND TERMDATE IS NULL
GROUP BY JOBTITLE
ORDER BY JOBTITLE DESC;
```

#Which department has the highest turnover rate?

```
SELECT DEPARTMENT, TOTAL_COUNT,
TERMINATED_COUNT,
TERMINATED_COUNT/TOTAL_COUNT AS TERMINATION_RATE
FROM (
        SELECT DEPARTMENT,
        COUNT(*)AS TOTAL_COUNT,
        SUM(CASE WHEN TERMDATE IS NOT NULL AND TERMDATE<=CURDATE() THEN 1 ELSE 0 END)
AS TERMINATED_COUNT
        FROM HR
          WHERE AGE>=18
         GROUP BY DEPARTMENT
         ) AS SUBQUERY
ORDER BY TERMINATION_RATE DESC;
```

#What is the distribution of employees across locations by state?

```
SELECT LOCATION_STATE, COUNT(*) AS COUNT
FROM HR
WHERE AGE >= 18 AND TERMDATE IS NULL
GROUP BY LOCATION_STATE
ORDER BY COUNT DESC;
```

#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 IS NOT NULL 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;
```

#What is the tenure distribution for each department?

```
 SELECT DEPARTMENT, ROUND(AVG(DATEDIFF(CURDATE(), TERMDATE)/365),0) AS AVG_TENURE
FROM HR
WHERE TERMDATE <= CURDATE() AND TERMDATE IS NOT NULL AND AGE >= 18
GROUP BY DEPARTMENT;
```