

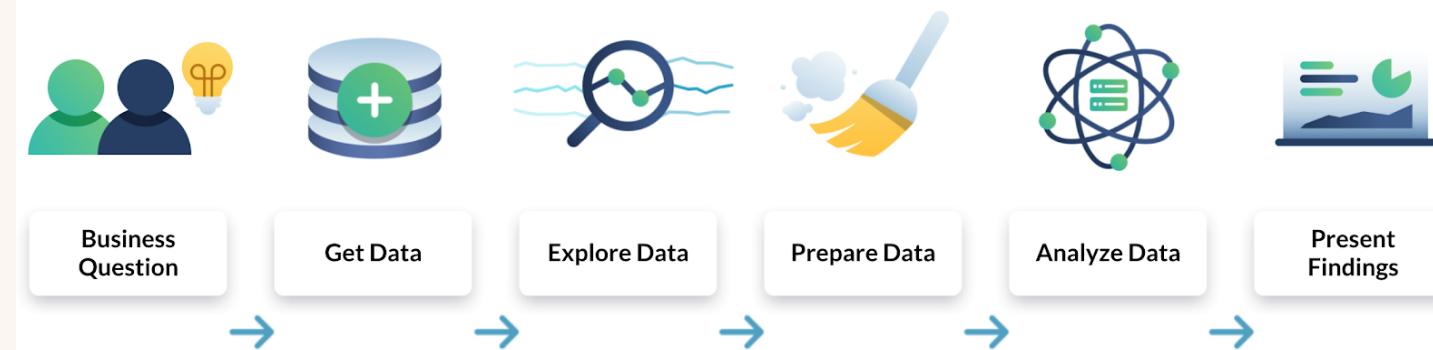
Presented by Manoj Patil

SQL + POWER BI

Data Analysis Project

<https://www.linkedin.com/in/manojpatil26>

The Data Analysis Process



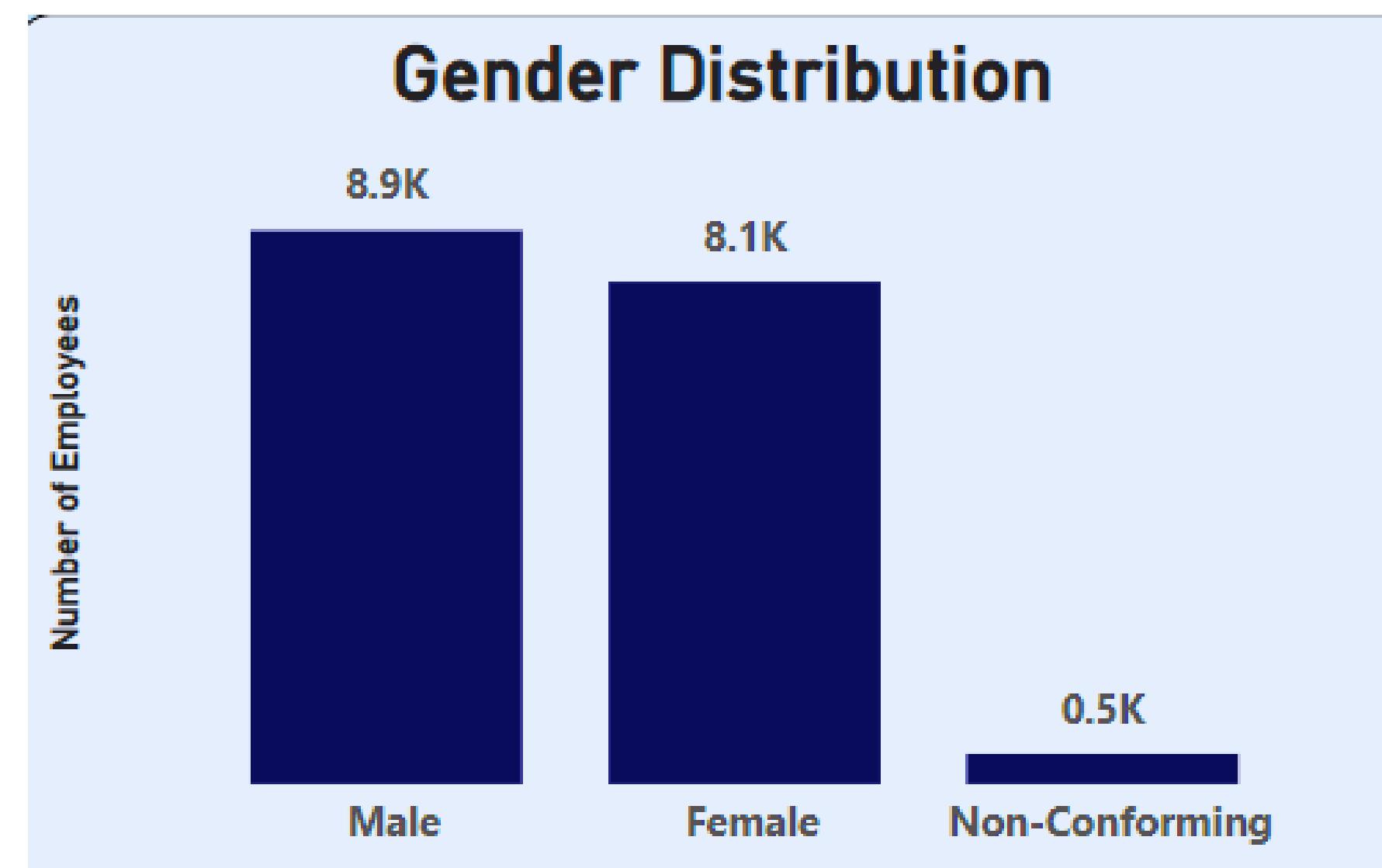
ABOUT PROJECT

The dataset contains extensive HR data spanning two decades, from 2000 to 2020, and includes over 22,000 rows. This robust dataset offers a wealth of information for analyzing long-term trends and patterns in various HR metrics such as employee demographics, performance, retention, promotions, and turnover rates.

By examining this data, one can gain valuable insights into workforce dynamics, the effectiveness of HR policies over time, and the impact of external factors on employee behavior. Such comprehensive analysis can aid in strategic planning, improving HR practices, and making data-driven decisions to enhance organizational effectiveness.

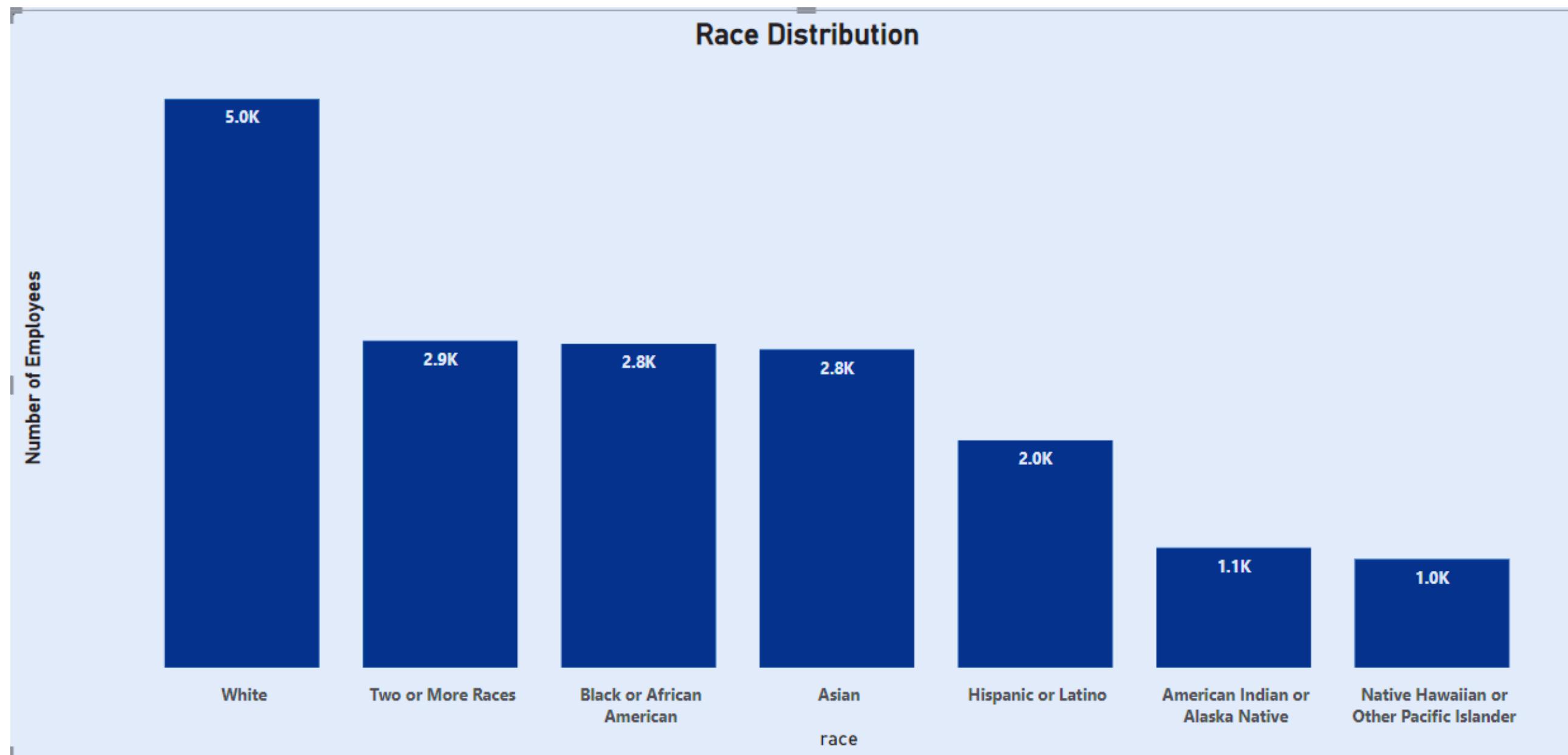
-- 1. What is the gender breakdown of employees in the company?

```
SELECT gender, count(*) AS count  
FROM hr  
WHERE age >= 18 AND termdate IS NULL  
GROUP BY gender;
```



-- 2. 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;
```

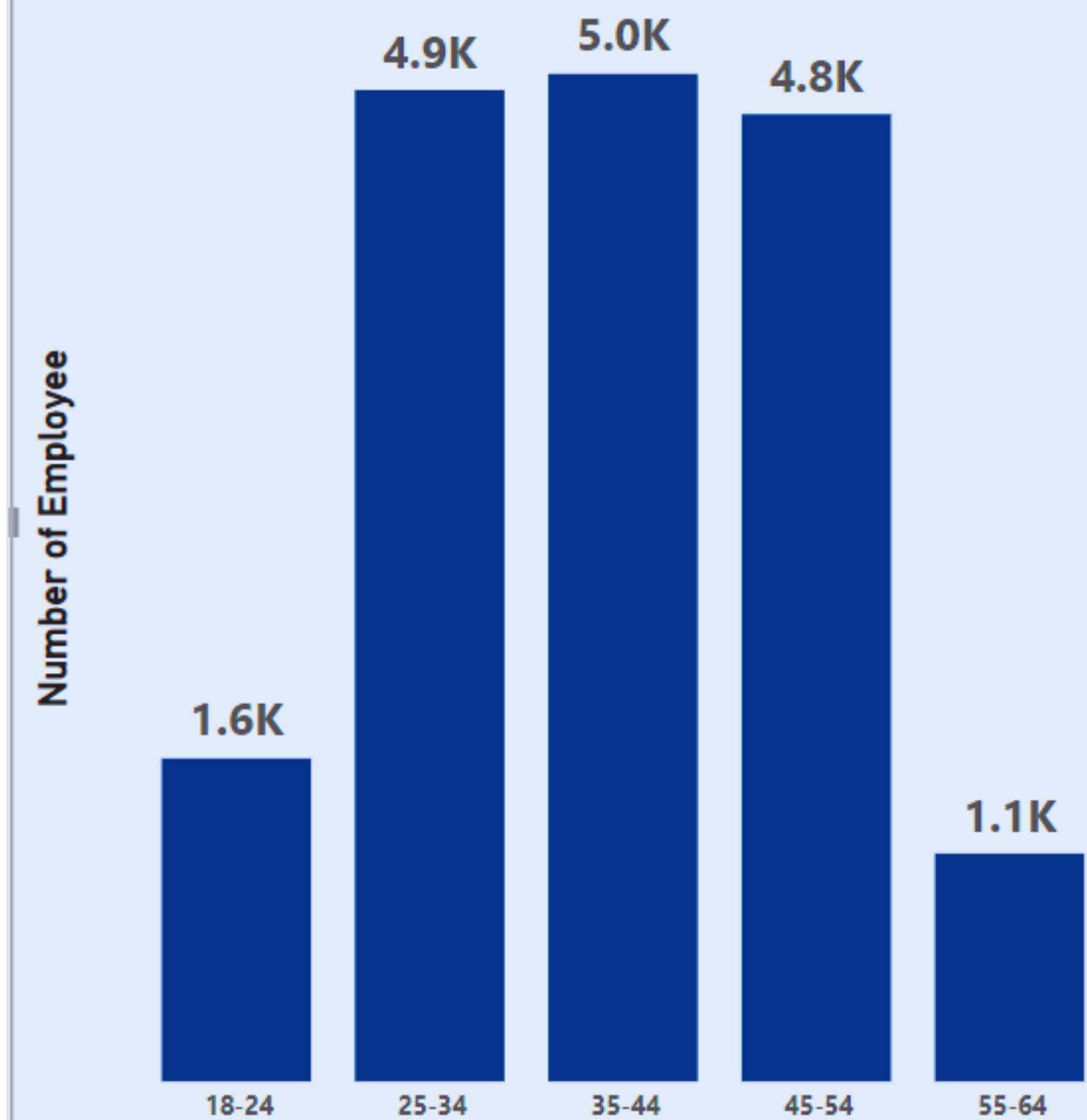


HR Project

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

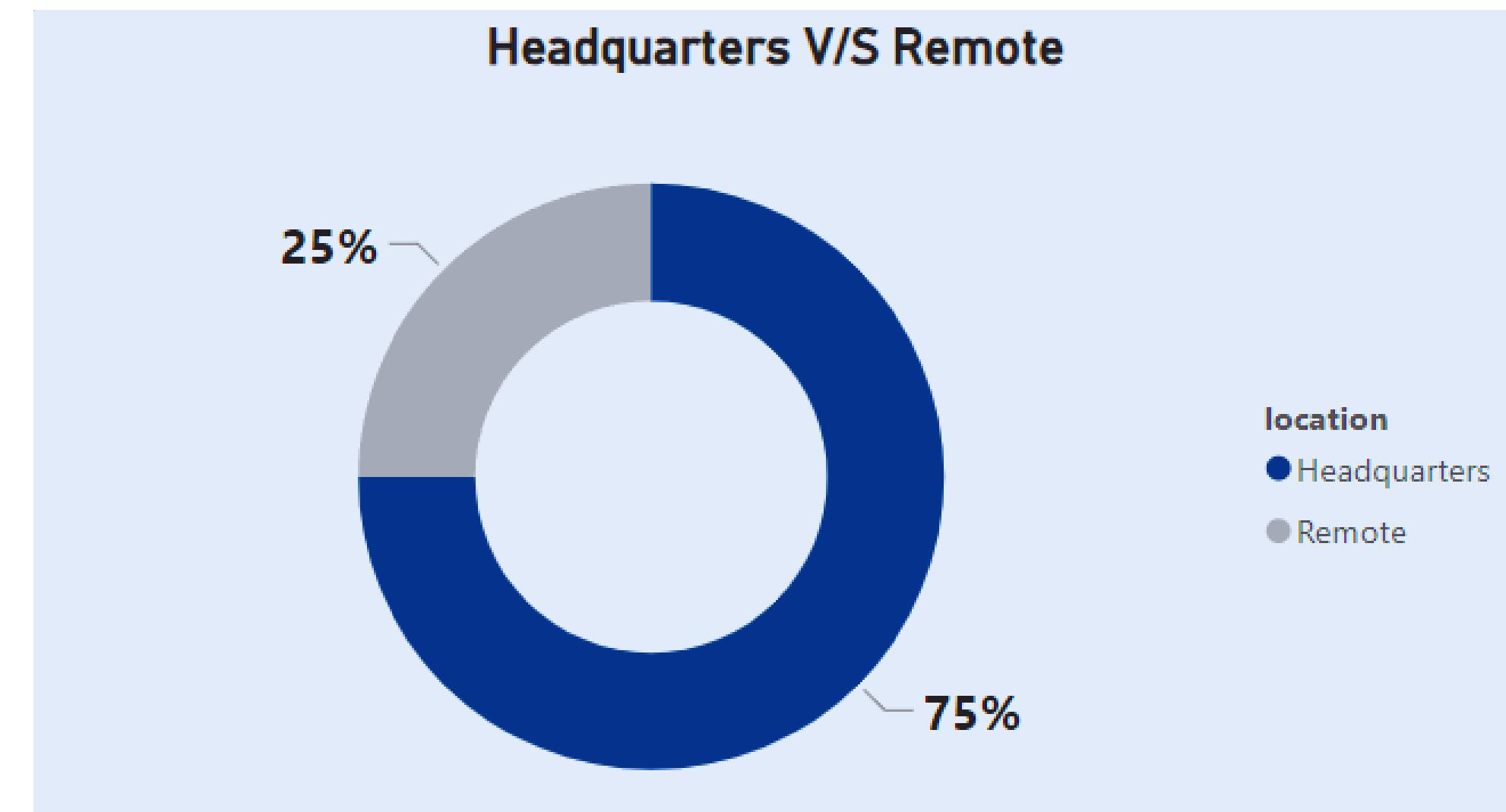
```
SELECT  
    min(age),  
    max(age)  
FROM hr  
WHERE age>=18 AND termdate IS NULL;  
  
SELECT CASE  
    WHEN age>= 18 and age<=24 THEN '18-24'  
    WHEN age>= 25 and age<=34 THEN '25-34'  
    WHEN age>= 35 and age<=44 THEN '35-44'  
    WHEN age>= 45 and age<=54 THEN '45-54'  
    WHEN age>= 55 and age<=64 THEN '55-64'  
    ELSE '65+'  
END AS age_group,  
    count(*) AS count,gender  
FROM hr  
WHERE age>=18 AND termdate IS NULL  
GROUP BY age_group,gender  
ORDER BY age_group,gender;
```

Age Group Distribution



-- 4. 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;
```





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

```
SELECT  
    round(avg(datediff(termdate,hire_date))/365,0) AS avg_length_employment  
FROM hr  
Where termdate <= curdate() AND termdate IS NOT NULL AND age>= 18;
```

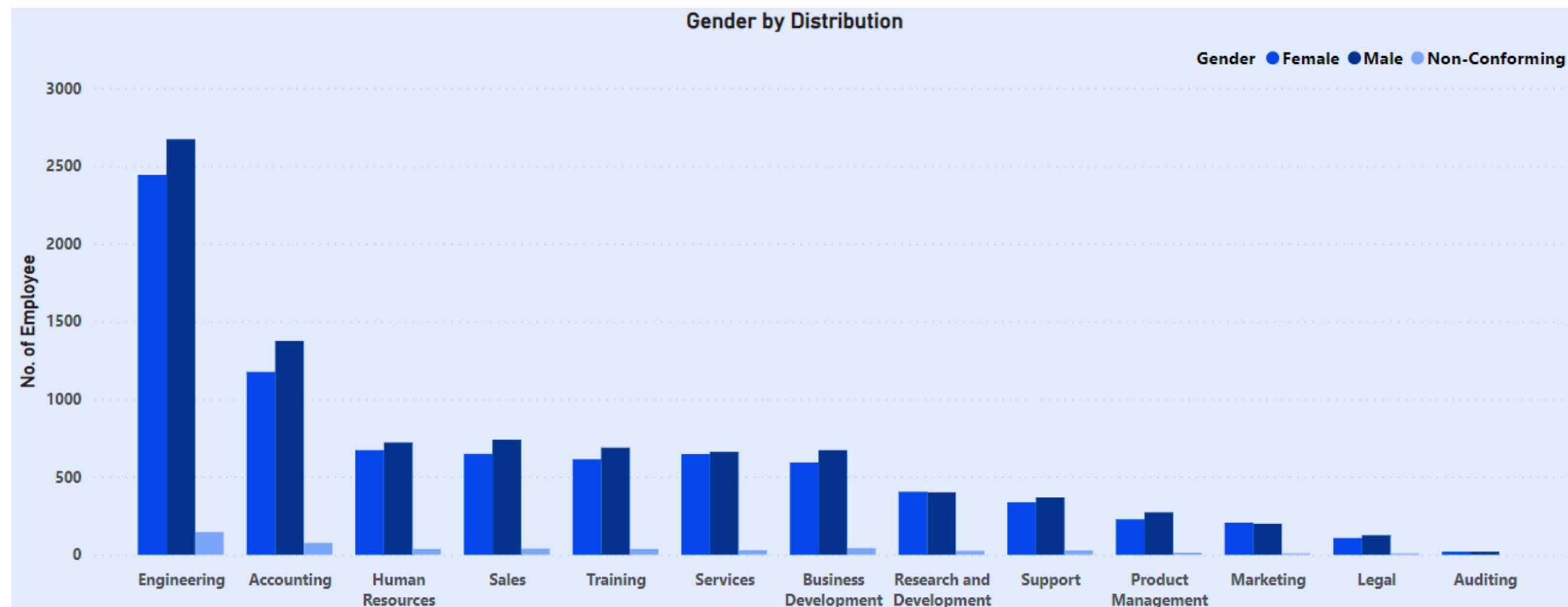
Average Length of Employment(YEARS)

8



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

```
SELECT department,gender,COUNT(*) AS count
FROM hr
WHERE age>=18 AND termdate IS NULL
GROUP BY department,gender
ORDER BY department;
```





-- 7. Which department has the higher turnover rate?

```
SELECT department,
       total_count,
       termination_count,
       termination_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
termination_count
    FROM hr
   WHERE age >=18
  GROUP BY department
) AS subquery
ORDER BY termination_rate DESC;
```

department	Sum of termination_rate
Accounting	0.11
Auditing	0.18
Business Development	0.10
Engineering	0.11
Human Resources	0.12
Legal	0.14
Marketing	0.09
Product Management	0.11
Research and Development	0.12
Sales	0.11
Services	0.11
Support	0.12
Training	0.12

```
SELECT location_state,COUNT(*) AS count
FROM hr
WHERE age>=18 AND termdate IS NULL
GROUP BY location_state
ORDER BY count DESC;
```

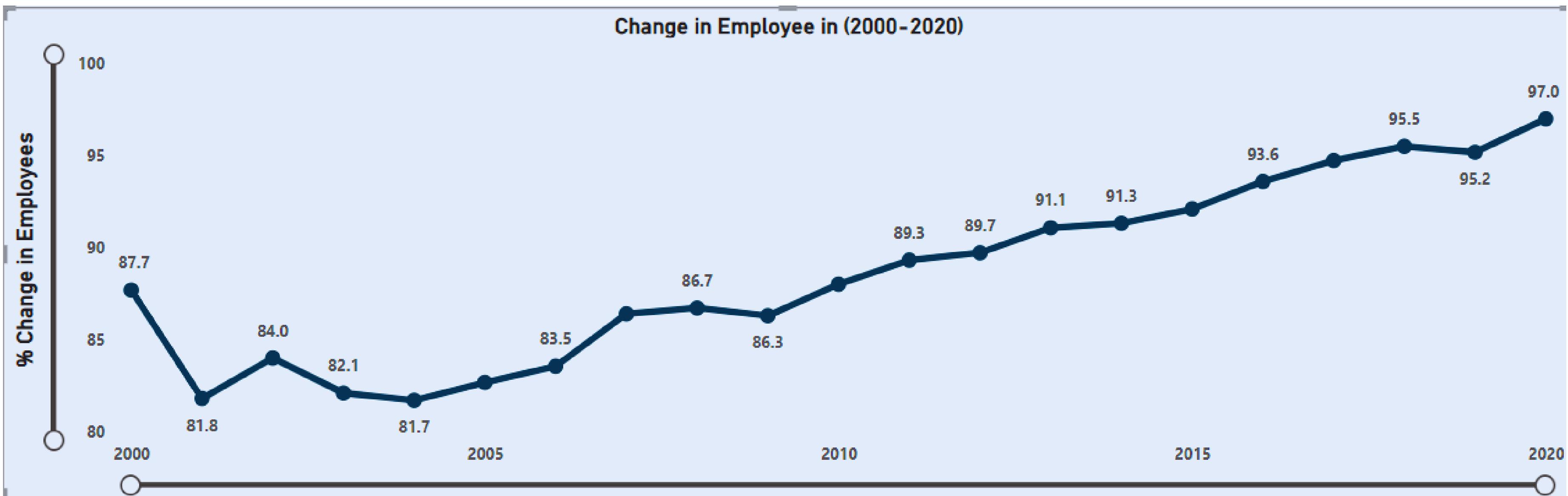




-- 9. How has the company's employees count changed over time based on hire and term dates?

```
SELECT
    year,
    hires,
    termination,
    hires-termination AS net_change,
    round((hires-termination)/hires*100,2) AS net_change_percentage
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 termination
    FROM hr
    WHERE age >= 18
    GROUP BY YEAR(hire_date)
) AS subquery
ORDER BY year;
```

Change in Employee in (2000-2020)

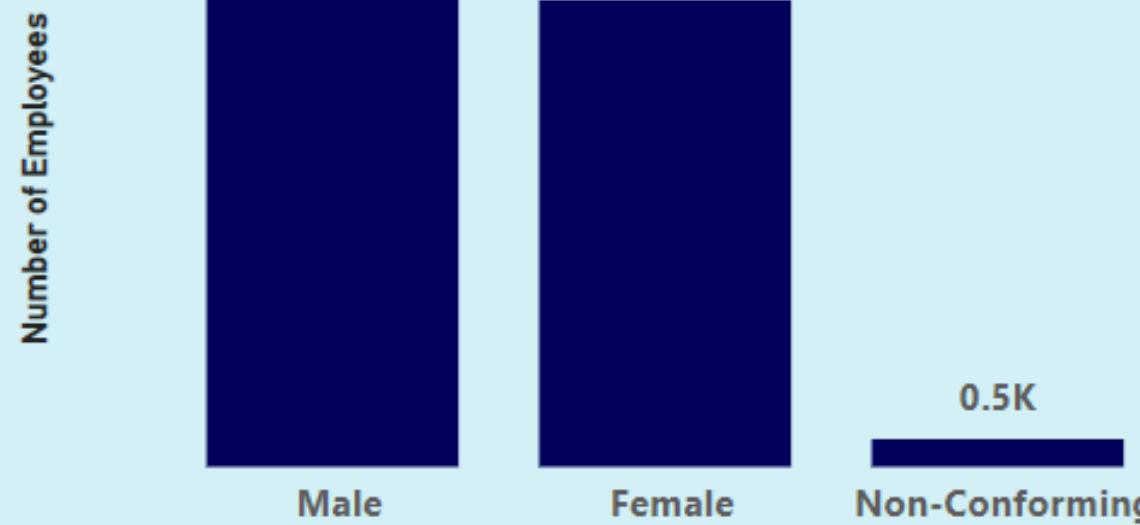


HR EMPLOYEE DISTRIBUTION REPORT

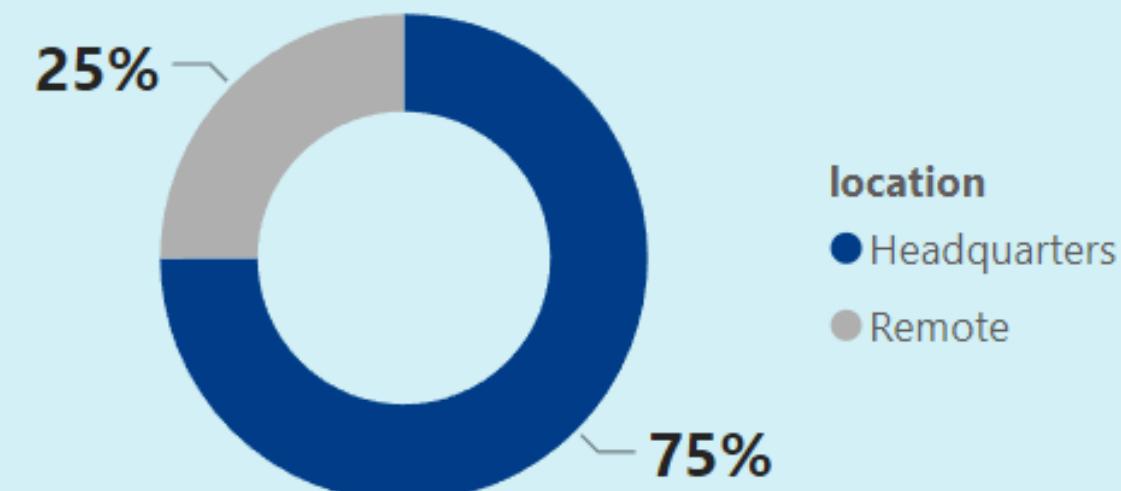
Average Length of Employment(YEARS)

8

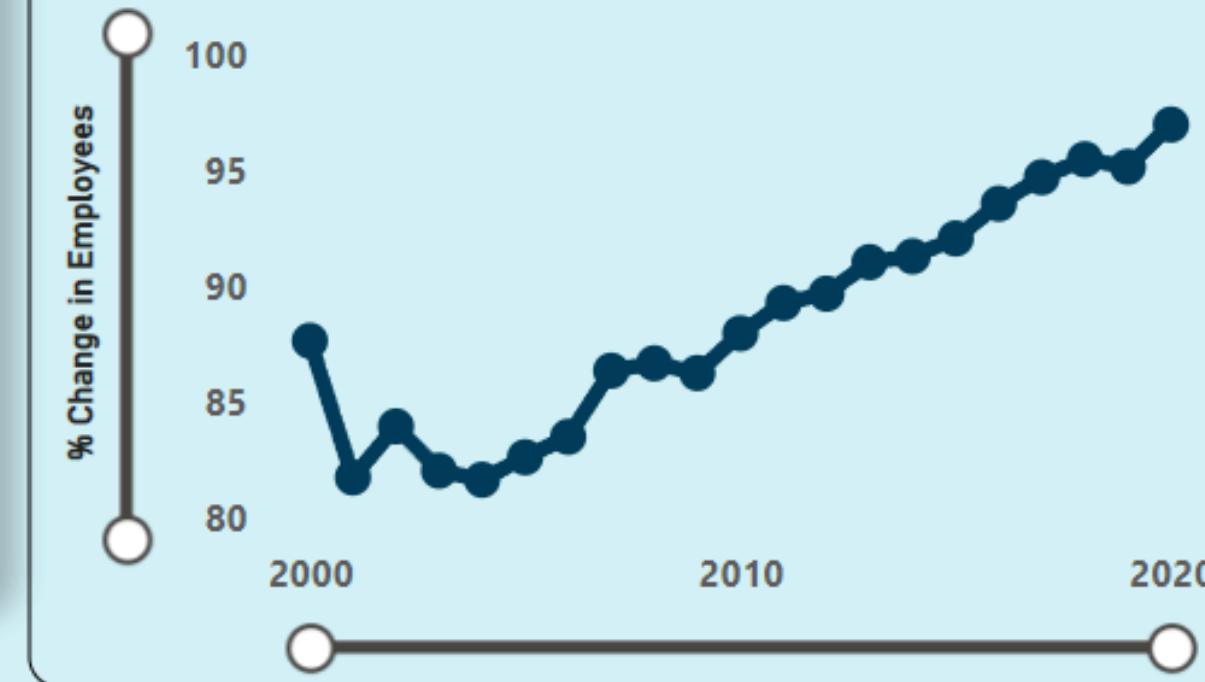
Gender Distribution



Headquarters V/S Remote



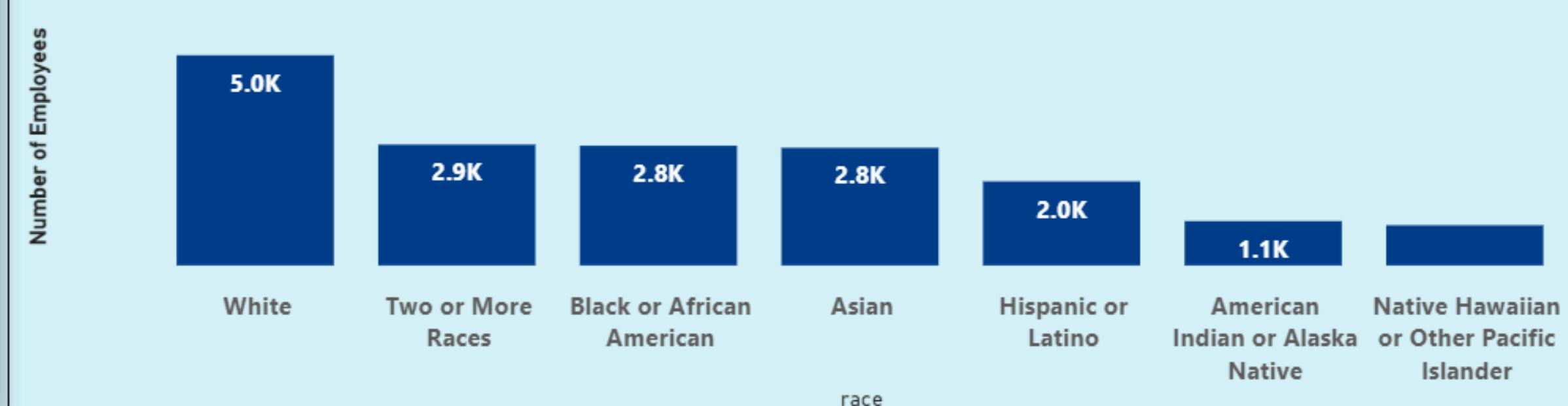
Change in Employee in (2000-2020)



Employee By State



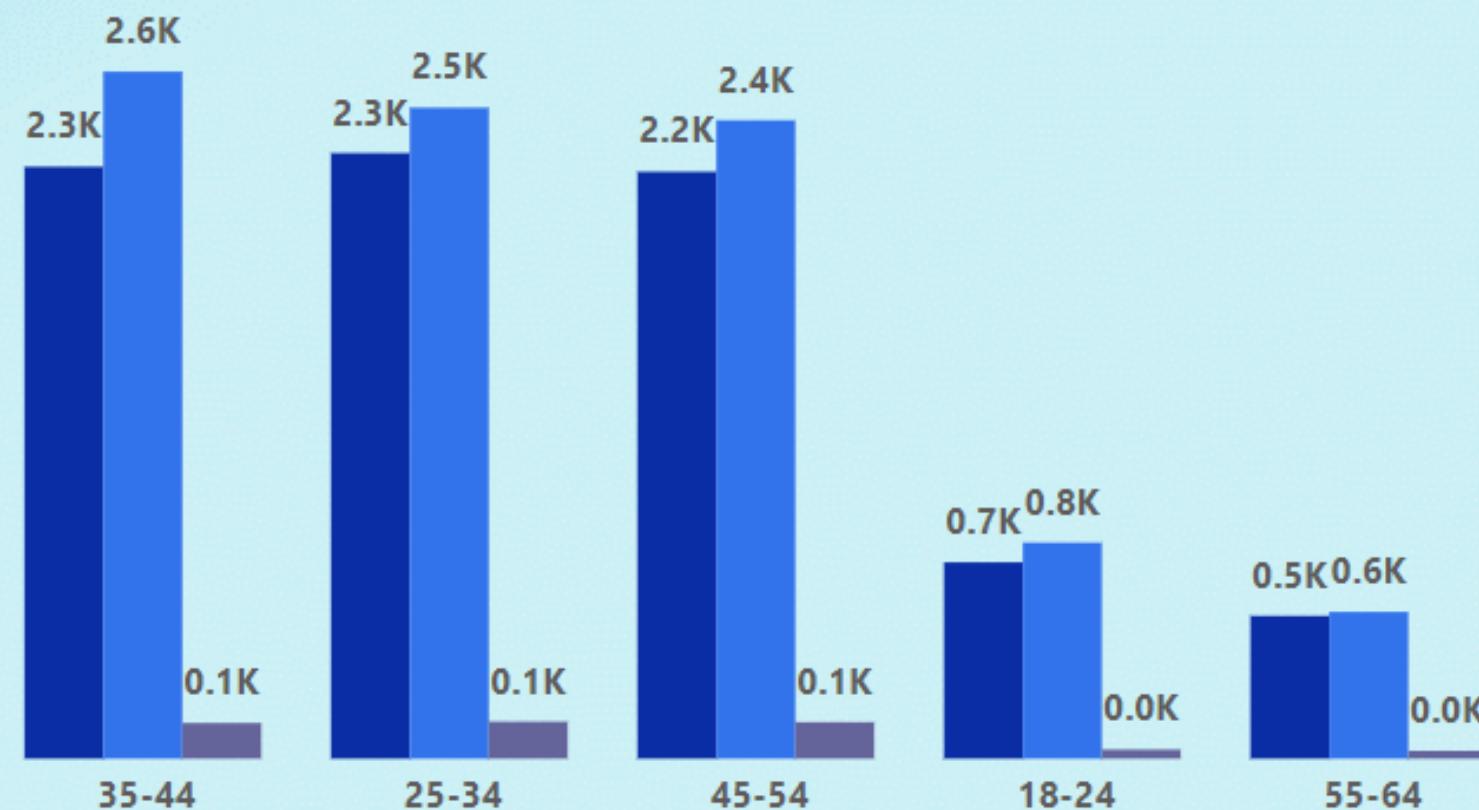
Race Distribution



Age Distribution by Gender

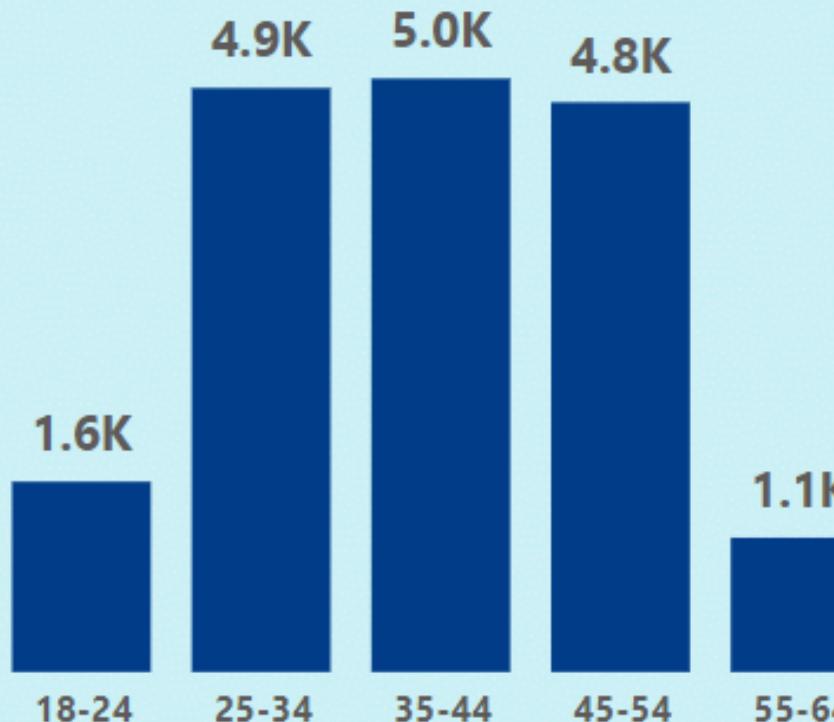
Gender ● Female ● Male ● Non-Conforming

Count of Employees



Age Group Distribution

Number of Employee



department

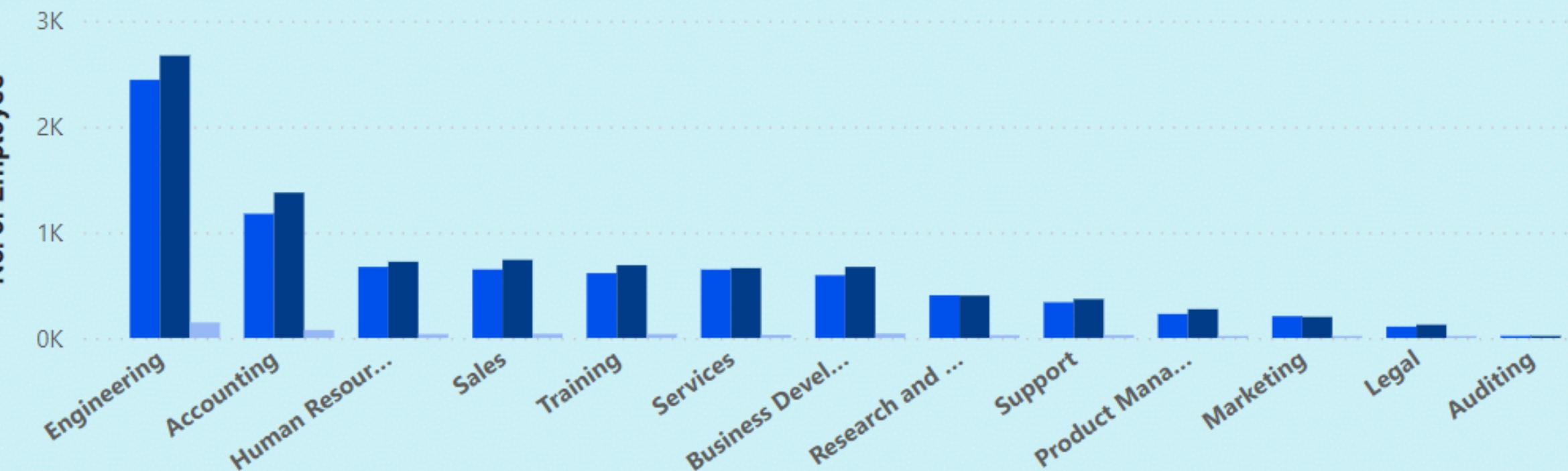
Sum of termination_rate

department	Sum of termination_rate
Auditing	0.18
Legal	0.14
Training	0.12
Research and Development	0.12
Human Resources	0.12
Support	0.12
Engineering	0.11
Accounting	0.11

Gender by Distribution

Gender ● Female ● Male ● Non-Conforming

No. of Employee



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.

**THANK
YOU !**

BY MANOJ PATIL