Data Used

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

**Data Cleaning & analysis** – MySQL, sqlmagic (Jupyter Notebook)

**Data Visualization** – MS Excel

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?

11. What is the tenure distribution for each department?

Summary of Findings

* There are more male employees
* White race is the more dominating at this work place while Hawaiian and Indians are the least dominant.
* The youngest employee is 21 years old and the oldest is 58.
* Three age groups were created (18-25,26-35, 35+). A large number of employees were older than 35+
* A large quantity of employees work at the office rather than from home.
* Average length of employment for terminating is around 8 years.
* The gender distribution across the departments is fairly balanced but there more male than females.
* The marketing department has the highest turnover, followed by Training. The least turnover rate is recorded as Research and Development, support and Legal departments.
* A large number of employees come from Ohio.
* The net change in employees increases as the years pass.
* The average tenure for each department is about 8 years with Legal and Auditing having the highest and Sales and Marketing hang the lowest

Limitations

* Some records had negative ages and these were excluded from the queries which was roughly 700+. Ages used were 18 years and above.
* Some term\_dates were far into the future and were not included in the analysis (around 1,550 records). The term\_dates used were those less than or equal to the current dates.
* Distribution across states would’ve been better represented if maps was used instead of Tree Map.