**HR analytics using Mysql**

Creating a database with the name Human Resources   
importing data from data.gov   
  
Data cleaning:   
Errors found   
1. Invalid column name ID   
changing the invalid column name ï»¿id to emp\_id   
Query:  
alter table hr

change column ï»¿id emp\_id varchar(30) null;

2. Changing the Date format for the birthdate column   
date the format in the database is mismatched so I will use the functions **date\_format** and **str\_to\_date** to update the table.  
**Query:  
UPDATE hr**

**SET birthdate = CASE**

**WHEN birthdate LIKE '%/%' THEN date\_format(str\_to\_date(birthdate, '%m/%d/%Y'), '%Y-%m-%d')**

**WHEN birthdate LIKE '%-%' THEN date\_format(str\_to\_date(birthdate, '%m-%d-%Y'), '%Y-%m-%d')**

**ELSE NULL**

**END;**explanation:   
update the table using the case when statement I have used the like operator to determine the format in which was the date using the **str\_to\_date** function to convert the string to the date and **date\_format** to put the date format correct.

3. Modifying the birthdate column into date format:  
Query:  
ALTER TABLE hr

MODIFY COLUMN birthdate DATE;

4. Changing the Date format for the hire date column:  
 **Query:**  
UPDATE hr

SET hire\_date = CASE

WHEN hire\_date LIKE '%/%' THEN date\_format(str\_to\_date(hire\_date, '%m/%d/%Y'), '%Y-%m-%d')

WHEN hire\_date LIKE '%-%' THEN date\_format(str\_to\_date(hire\_date, '%m-%d-%Y'), '%Y-%m-%d')

ELSE NULL

END;  
5. Modifying Term date to remove irrelevancy:  
  
In the term date column data has both date and timestamp where the time stamp is irrelevant and blank data on existing employees to clean this data  
**Query:**  
UPDATE hr

SET termdate = date(str\_to\_date(termdate, '%Y-%m-%d %H:%i:%s UTC'))

WHERE termdate IS NOT NULL AND termdate != ' ' AND termdate != '';   
// to add to add "0000-00-00" in the blank rows   
UPDATE hr

SET termdate = '0000-00-00'

WHERE termdate IS NULL OR termdate = '';

6. Adding a new column age:  
In this, we are going to add a new column called age using **timestampdiff** function by calculating age by birthdate and current date.

ALTER TABLE HR

ADD column age int;

UPDATE hr

set age = timestampdiff(year,birthdate,CURDATE());

**Business Questions to Answer**

**1. What is the gender breakdown of employees in the company**we use the count and group to retrieve the data.  
Query:  
select gender,

count(gender) as count

from hr

where age > 18

group by gender;

**2. What is the race/ethnicity breakdown of employees in the company?**

Query   
select race, count(race) as race

from hr

where age >18

group by race;

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

Query

SELECT age,

CASE

WHEN age BETWEEN 18 AND 30 THEN 'young adult'

WHEN age between 31 and 60 THEN 'senior'

ELSE 'other'

end as age\_distribution

from hr;

**4. How many employees work at headquarters versus remote locations?**

Query

select location,

count(location) as location\_distribution

from hr

group by location;

**5. What is the average length of employment for employees who have been terminated?**

Query:

select round(avg(datediff(termdate,hire\_date))/365,0) as avg\_length\_of\_employment

from hr

where termdate is not null

and termdate <> 000-00-00 and termdate <= curdate() and age >= 18;

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

Query

select

jobtitle, department, gender , count(\*) as count

from hr

group by gender, department , jobtitle

order by department;

**7. What is the distribution of job titles across the company?**

Query

select jobtitle , count(jobtitle) as Count

from hr

group by jobtitle;

1. **Which department has the highest turnover rate?**

**Turnover rate" typically refers to the rate at which employees leave a company or department and need to be replaced. It can be calculated as the number of employees who leave over a given time period divided by the average number of employees in the company or department over that same time period.**

Query:

SELECT department, COUNT(\*) as total\_count,

SUM(CASE WHEN termdate <= CURDATE() AND termdate <> 0000-00-00 THEN 1 ELSE 0 END) as terminated\_count,

SUM(CASE WHEN termdate = 0000-00-00 THEN 1 ELSE 0 END) as active\_count,

(SUM(CASE WHEN termdate <= CURDATE() THEN 1 ELSE 0 END) / COUNT(\*)) as termination\_rate

FROM hr

WHERE age >= 18

GROUP BY department

ORDER BY termination\_rate DESC;

**9. What is the distribution of employees across locations by state?**

Query:

select location\_state , count(\*) as cnt

from hr

group by location\_state

order by cnt desc;

### 10.How has the company's employee count changed over time based on hire and term dates?

Query:

SELECT

YEAR(hire\_date) AS year,

COUNT(\*) AS hires,

SUM(CASE WHEN termdate <> 0000-00-00 AND termdate <= CURDATE() THEN 1 ELSE 0 END) AS terminations,

COUNT(\*) - SUM(CASE WHEN termdate <> 0000-00-00 AND termdate <= CURDATE() THEN 1 ELSE 0 END) AS net\_change,

ROUND(((COUNT(\*) - SUM(CASE WHEN termdate <> 0000-00-00 AND termdate <= CURDATE() THEN 1 ELSE 0 END)) / COUNT(\*) \* 100),2) AS net\_change\_percent

FROM

hr

WHERE age >= 18

GROUP BY

YEAR(hire\_date)

ORDER BY

YEAR(hire\_date) ASC;

**Analysis**

**Age and Gender Demographics:**

The bar chart shows the distribution of employees across different age groups segmented by gender. The 35-44 age group is the most populous, indicating that the company's workforce is relatively mature and experienced. This could suggest that the company is successful in retaining employees through their mid-career stages. The presence of employees in the 55-64 age group indicates there are opportunities and roles for more senior professionals, which may contribute to knowledge retention within the company.

The small percentage of non-conforming gender individuals may reflect the inclusivity of the company's hiring practices or may indicate an area for growth in diversity and inclusivity initiatives.

**Job Titles and Tenure:**

The list of job titles with the sum of count indicates that certain roles, like 'Account Executive' and 'Accounting Assistant I,' are more common within the organization. This could suggest that these are entry-level or high-turnover positions, or possibly that these roles are critical to the company's operations and thus have a larger number of employees.

The 'Tenure by Department' chart shows uniformity across departments, which implies that the company has consistent retention rates regardless of department. This can be a sign of stable departmental cultures and equitable management practices across different functions of the organization.

**Location and Race Distribution:**

A large proportion of employees work remotely, which could indicate a flexible work policy or a geographically dispersed talent pool. The company could be leveraging remote work to attract and retain talent without geographical constraints.

The race distribution bar chart presents a workforce predominantly composed of white employees, followed by other racial groups. The distribution suggests that while there is some racial diversity within the company, there might be room to improve representation of minority groups.

**Turnover and Employment Statistics:**

The turnover trend line, which is increasing, could be interpreted in multiple ways. An increasing trend might be a sign of a growing company that regularly brings in new talent. Conversely, it could indicate rising turnover rates, which may be a concern if the company is losing valuable expertise or incurring high replacement costs.

The 'Average length of Employment in years' being 8 years is notable. It suggests that once employees join the company, they tend to stay for a significant period. This can be indicative of a positive work environment, good employee engagement, and effective retention strategies.

The state count table shows that a vast number of employees are based in Ohio. This suggests that Ohio may be a strategic location for the company, possibly housing the headquarters or key operations. It could also imply that the company's culture and employment opportunities are well aligned with the labor market in that state.

**Overall** **Conclusion**: The company appears to have a stable and mature workforce with a strong tenure and a significant remote employee base. There is evidence of good retention strategies, as seen by the average employment duration and the uniform tenure across departments. However, there seems to be an opportunity to enhance diversity initiatives, especially in terms of gender and race. The increasing trend in turnover warrants a closer examination to ensure it is due to positive growth rather than dissatisfaction or competitive job markets. Comparing these metrics to industry and regional benchmarks could provide further insights into the company's position relative to the market and highlight areas for strategic improvement.

Top of Form