HR Analytics

Objective:

To analyze employee demographic and workforce data to uncover trends in hiring, diversity, tenure, and attrition using SQL and Power BI.

Data Overview:

- We have created ‘hr’ Database using MySQL
- Imported ‘Human_resources.csv’ file and named the table as ‘human_resources’
- The csv file consists of 22214 rows and 13 columns.

Data Cleaning:

Purpose:

To standardize and clean the HR dataset for analysis in Power BI.

- Renamed a column with a formatting issue and adjusted datatype of employee_id to ensure consistency.

```
ALTER TABLE human_resources
CHANGE COLUMN ï»¿id employee_id VARCHAR(20) NOT NULL;
ALTER TABLE human_resources
MODIFY COLUMN employee_id VARCHAR(20) NULL;
DESCRIBE human_resources;
```

- Changed the data format and datatype of birthdate column

```
UPDATE human_resources
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;
ALTER TABLE human_resources
MODIFY COLUMN birthdate DATE;
```

- Changed the data format and datatype of hire_date column

```
UPDATE human_resources
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;
ALTER TABLE human_resources
MODIFY COLUMN hire_date DATE;
```

- Changed the data format and datatype of termdate column

```
UPDATE human_resources
SET termdate = CASE
        WHEN termdate LIKE '% UTC' THEN LEFT(termdate, 10)
        WHEN termdate LIKE '%/%' THEN date_format(STR_TO_DATE(termdate, '%m/%d/%Y'),'%Y-%m-%d')
        WHEN termdate LIKE '%-%' THEN date_format(str_to_date(hire_date, '%m-%d-%Y'), '%Y-%m-%d')
        ELSE NULL
    END;
ALTER TABLE human_resources
MODIFY COLUMN termdate DATE;
```

- Created age column

```
ALTER TABLE human_resources ADD COLUMN age INT;
UPDATE human_resources
SET age = timestampdiff(YEAR, birthdate, CURDATE());
```

Key metrics:

- Total hires

  SELECT COUNT(*) AS total_hires FROM human_resources WHERE hire_date IS NOT NULL AND hire_date<=curdate();

	total_hires
▶	22214

- Total terminations

  SELECT COUNT(*) AS total_terminations FROM human_resources WHERE termdate IS NOT NULL AND termdate<=curdate();

	total_terminations
▶	2689

- Turnover rate

  SELECT COUNT(*) AS total_count,(sum(CASE WHEN termdate IS NOT NULL AND termdate <=curdate() THEN 1 END) * 1.0 / COUNT(*))AS turnover_rate FROM human_resources;

	total_count	turnover_rate
▶	22214	0.12105

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

  SELECT round(avg(year(termdate)-year(hire_date)),0) AS avg_length_employment FROM human_resources WHERE termdate<=curdate() AND termdate IS NOT NULL AND age>=18;

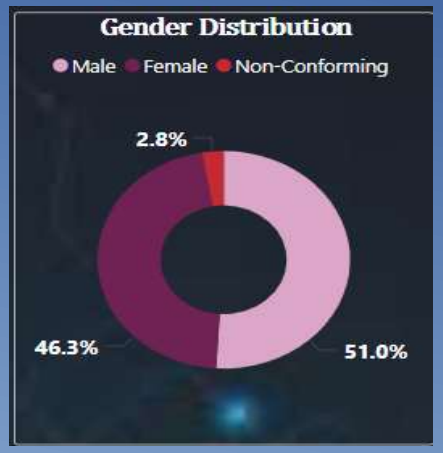
	avg_length_employment
▶	8

SQL queries and dashboard views:

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

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

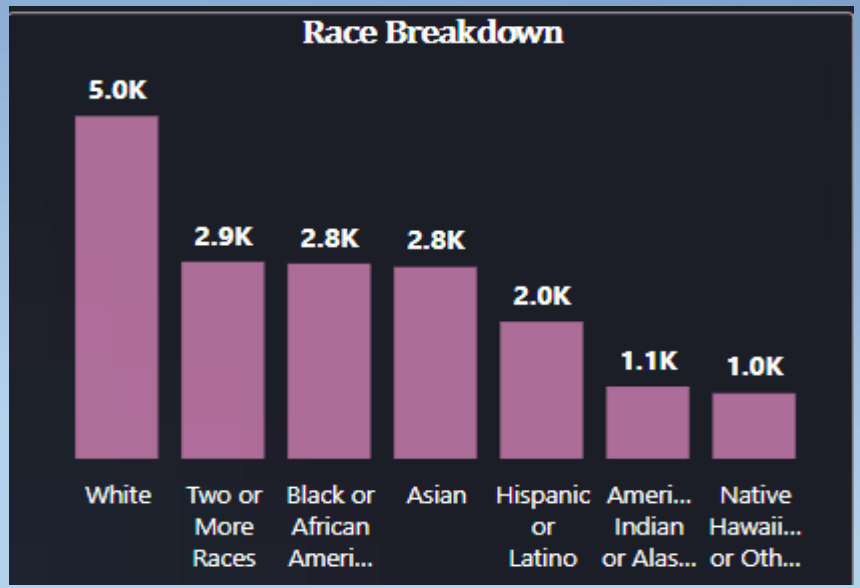
gender	count
Male	8911
Female	8090
Non-Conforming	481



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

   - SELECT race, COUNT(*) AS count FROM human_resources WHERE age >= 18 AND termdate IS NULL GROUP BY race ORDER BY count DESC;

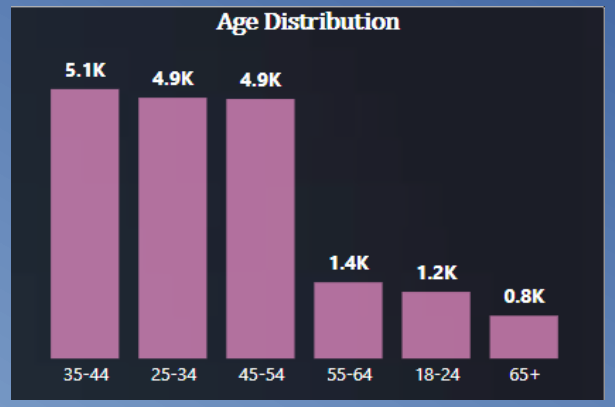
race	count
White	4987
Two or More Races	2867
Black or African American	2840
Asian	2791
Hispanic or Latino	1994
American Indian or Alaska Native	1051
Native Hawaiian or Other Pacific Islander	952



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

- 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 human_resources WHERE termdate IS NULL GROUP BY age_group ORDER BY age_group;

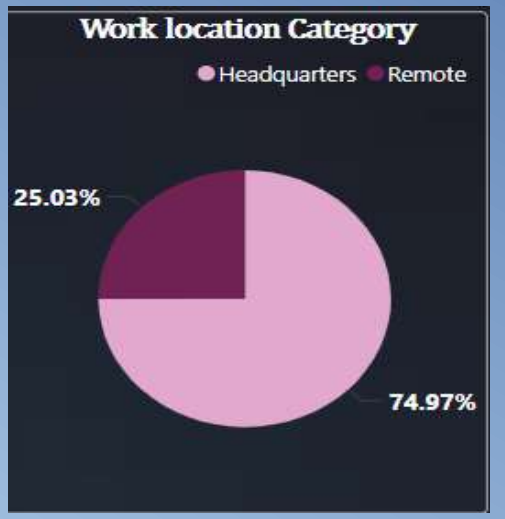
age_group	count
18-24	1249
25-34	4888
35-44	5051
45-54	4865
55-64	1429
65+	803



4. How many employees work at headquarters versus remote locations

- SELECT location, COUNT(*) AS countFROM human_resources WHERE age >= 18 AND termdate IS NULL GROUP BY location;

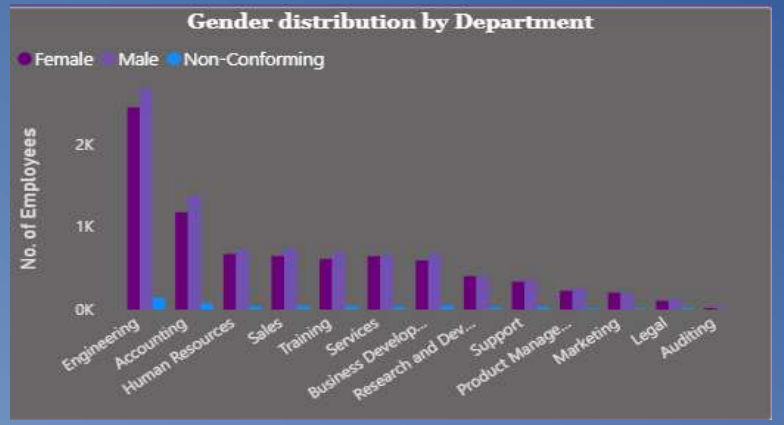
location	count
Headquarters	13107
Remote	4375



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

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

department	gender	count
Accounting	Female	1175
Accounting	Male	1375
Accounting	Non-Conforming	76
Auditing	Female	19
Auditing	Male	19
Business Development	Female	593
Business Development	Male	672



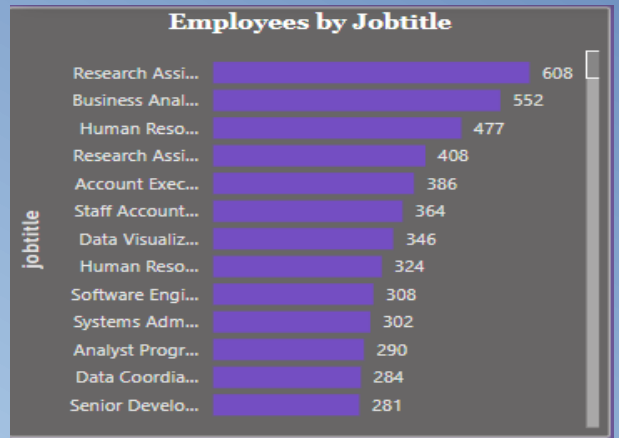
- SELECT jobtitle, gender, COUNT(*) AS count FROM human_resources WHERE age >= 18 AND termdate IS NULL GROUP BY jobtitle, genderORDER BY jobtitle, gender;

jobtitle	gender	count
Account Coordinator	Male	2
Account Executive	Female	195
Account Executive	Male	183
Account Executive	Non-Conforming	8
Account Manager	Female	91
Account Manager	Male	93
Account Manager	Non-Conforming	4

6. What is the distribution of job titles across the company?

- SELECT jobtitle, COUNT(*) AS count FROM human_resources WHERE age >= 18 AND termdate IS NULL GROUP BY jobtitle ORDER BY count DESC;

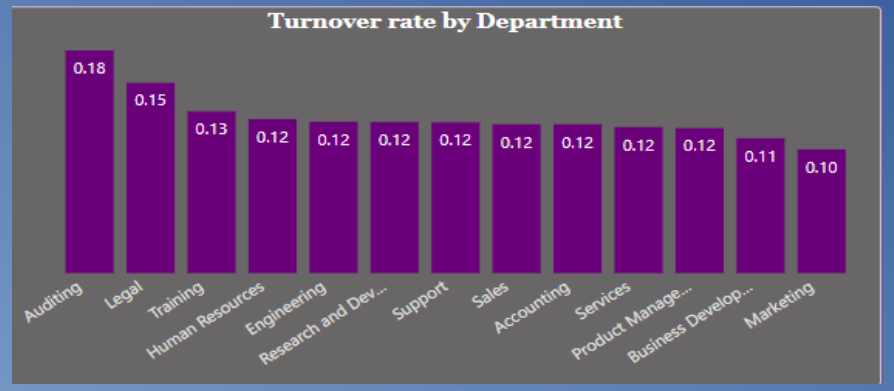
jobtitle	count
Research Assistant II	608
Business Analyst	552
Human Resources Analyst II	477
Research Assistant I	408
Account Executive	386
Staff Accountant I	364
Data Visualization Specialist	346



7. Which department has the highest turnover rate?

- SELECT department, COUNT(*) AS total_count, COUNT(CASE WHEN termdate IS NOT NULL AND termdate <=curdate() THEN 1 END) * 1.0 / COUNT(*) AS turnover_rate FROM human_resources WHERE age>=18 GROUP BY department ORDER BY turnover_rate DESC;

department	total_count	turnover_rate
Auditing	50	0.18000
Legal	299	0.15385
Training	1622	0.13132
Human Resources	1727	0.12449
Engineering	6387	0.12228
Research and Development	1032	0.12209
Support	903	0.12182



8. What is the distribution of employees across locations by city and state?

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

location_state	count
Ohio	14144
Pennsylvania	892
Illinois	698
Michigan	550
Indiana	545
Kentucky	347
Wisconsin	306

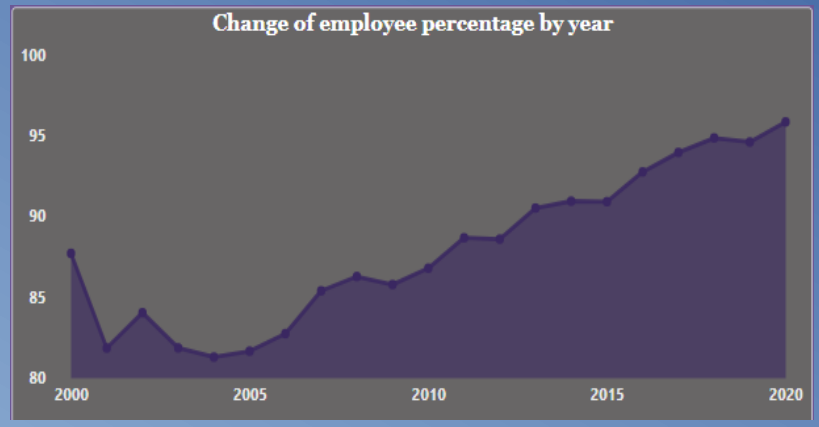
- SELECT location_city, COUNT(*) AS count FROM human_resources WHERE age>=18 AND termdate IS NULL GROUP BY location_city ORDER BY count DESC;

location_city	count
Cleveland	13233
Chicago	283
Philadelphia	268
Pittsburgh	229
Cincinnati	221
Louisville	174
Detroit	165

9. 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 human_resources WHERE age>=18 GROUP BY YEAR(hire_date))AS sub_query ORDER BY year;

year	hires	terminations	net_change	net_change_percent
2000	211	26	185	87.68
2001	1082	197	885	81.79
2002	1012	162	850	83.99
2003	1088	198	890	81.80
2004	1087	204	883	81.23
2005	1038	191	847	81.60
2006	1069	185	884	82.69



10. What is the tenure distribution for each department?

- SELECT department, round(avg(year(termdate)-year(hire_date)),0) AS avg_tenure FROM human_resources WHERE termdate<=curdate() AND termdate IS NOT NULL AND age>=18 GROUP BY department;

department	avg_tenure
Engineering	8
Services	8
Human Resources	8
Business Development	8
Sales	9
Support	8
Auditing	8

Departments by average tenure

department	avg_tenure
Marketing	9
Sales	9
Accounting	8
Auditing	8
Business Development	8
Engineering	8
Human Resources	8
Research and Development	8
Services	8
Support	8
Training	8

Key Insights:

- There are more male employees in the company.
- Employees with White ethnicity is the most dominant while Native Hawaiian and American Indian are the least dominant.
- 6 age groups were created (18-24, 25-34, 35-44, 45-54, 55-64, 65+). A large number of employees were between 35-44 followed by 25-34 while the smallest group was 65+.
- A large number of employees work at the headquarters.
- The average length of employment for terminated employees is around 8 years.
- The gender distribution across departments is fairly balanced but there are generally more male than female employees.
- The Auditing department has the highest turnover rate followed by Legal. The least turn over rate is for the Marketing department.
- 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 Marketing and sales having the highest.