Introduction

In this evolving landscape of human resources,data_driven insights are the key to optimizing workforce management. The project harnesses the power of SQL workbench for meticulous data cleaning and analysis ,paired with the dynamic visual capabilities of power bi.By transforming raw HR data into meaningful visual stories, we uncover critical trends in employee turnover ,performance and demographics.This project not only highlights the synergy between SQL and Power BI but also demonstrates their impact in driving strategic HR decisions.

DATA USED

DATA - HR Data with over 22,000 rows from the year 2000 to 2200

Data Cleaning & Analysis - MYSQL Workbanch

Data Visualization - PowerBI

QUESTIONS

1.What is the gender breakdown of employees in the country?

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

3.What is the age distribution of employee 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 state?

10.How has the company’s employee count changed over time based on hire and termdates?

11.What is the tenure distribution for each department?

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

```
SELECT gender ,count(*) as count
FROM hr
WHERE age>= 18 AND termdate = '0000-00-00'
GROUP BY gender;
```

gender	count
Non-Conforming	481
Male	8911
Female	8090

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

```
SELECT race ,count(*) as count
FROM hr
WHERE age>= 18 AND termdate = '0000-00-00'
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

What is the age distribution of employees in the company?

```sql
SELECT
  min(age) AS youngest,
  max(age) AS oldest
FROM hr
WHERE age >= 18 AND termdate = '0000-00-00';
```

youngest	oldest
21	58

```sql
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 <= 55 THEN '45-55'
      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 = '0000-00-00'
GROUP BY age_group
ORDER BY age_group;
```

age_group	count
18-24	1572
25-34	4911
35-44	5027
45-55	5144
55-64	828

```sql
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 <= 55 THEN '45-55'
      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 = '0000-00-00'
GROUP BY age_group,gender
ORDER BY age_group,gender;
```

age_group	gender	count
18-24	Female	732
18-24	Male	805
18-24	Non-Conforming	35
25-34	Female	2303
25-34	Male	2467
25-34	Non-Conforming	141
35-44	Female	2257
35-44	Male	2634
35-44	Non-Conforming	136
45-55	Female	2392
45-55	Male	2605
45-55	Non-Conforming	147
55-64	Female	406
55-64	Male	400
55-64	Non-Conforming	22

How many employees work at headquaters versus remote locations?

```sql
SELECT location,count(*) AS count
FROM hr
WHERE age>=18 AND termdate = '0000-00-00'
GROUP BY location;
```

location	count
Headquarters	13107
Remote	4375

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

```sql
SELECT round(avg(datediff(termdate,hire_date))/365,0) AS avg_length_employment
FROM hr
WHERE termdate <= curdate() AND termdate <> '0000-00-00' AND age >= 18;
```

avg_length_employment
8

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

```sql
SELECT department,gender,count(*) AS count
FROM hr
WHERE age >= 18 AND termdate = '0000-00-00'
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
Business Development	Non-Conforming	42
Engineering	Female	2442
Engineering	Male	2671
Engineering	Non-Conforming	146
Human Resources	Female	672
Human Resources	Male	721
Human Resources	Non-Conforming	37
Legal	Female	107
Legal	Male	125
Legal	Non-Conforming	5
Marketing	Female	206
Marketing	Male	199
Marketing	Non-Conforming	5
Product Management	Female	227

sult 14 ×

What is the distribution of job titles across the country?

```sql
SELECT jobtitle, count(*) as count
FROM hr
WHERE age >= 18 AND termdate = '0000-00-00'
GROUP BY jobtitle
ORDER BY jobtitle 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
Human Resources Analyst	324
Software Engineer I	308
Systems Administrator I	302
Analyst Programmer	290
Data Coordiator	284
Senior Developer	281
Business Systems Develop...	272
Project Manager	269
Service Tech III	268
Service Manager	267
Help Desk Technician	266
Service Tech II	265
Software Consultant	262
Service Coordinator	259

Which department has the highest turnover rate?

```sql
SELECT department,
total_count,
terminated_count,
terminated_count/total_count AS termination_rate
FROM(
    SELECT department,
    count(*) AS total_count,
    sum(CASE WHEN termdate <> '0000-00-00'
    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;
```

department	total_count	terminated_count	termination_rat
Auditing	50	9	0.1800
Legal	299	42	0.1405
Training	1622	202	0.1245
Research and Development	1032	124	0.1202
Human Resources	1727	204	0.1181
Support	903	105	0.1163
Engineering	6387	737	0.1154
Accounting	3192	368	0.1153
Sales	1745	200	0.1146
Product Management	623	71	0.1140
Services	1618	181	0.1119
Business Development	1569	159	0.1013
Marketing	480	45	0.0938

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

```sql
SELECT location_state ,count(*) AS count
FROM hr
WHERE age >=18 AND termdate='0000-00-00'
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

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

```sql
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<>'0000-00-00'
         AND termdate <= curdate()
         THEN 1
         ELSE 0
         END) AS terminations
      FROM hr
      WHERE age>= 18
      GROUP BY YEAR(hire_date)
      ) AS subquery
ORDER BY year ASC;
```

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	196	892	81.99
2004	1087	199	888	81.69
2005	1038	181	857	82.56
2006	1069	177	892	83.44
2007	1058	146	912	86.20
2008	1061	141	920	86.71
2009	1094	151	943	86.20
2010	1050	130	920	87.62
2011	1057	115	942	89.12
2012	1059	109	950	89.71
2013	1042	93	949	91.07
2014	1014	89	925	91.22
2015	1011	81	930	91.99
2016	1076	69	1007	93.59
2017	1043	55	988	94.73
2018	1090	50	1040	95.41
2019	1038	50	988	95.18
2020	967	30	937	96.90

sult 9 ×

What is the tenure distribution for each department?

```sql
SELECT department,round(avg(datediff(termdate,hire_date)/365),0) AS avg_tenure
FROM hr
WHERE termdate <= curdate() AND termdate<>'0000-00-00' 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
Training	7
Accounting	8
Research and Development	8
Product Management	7
Legal	7
Marketing	8

Summary of Findings

- There are more male employees.
- White race is the most dominant 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 by35-44 while the smallest group was 55-64.
- A large number of employees work at the headquaters 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 turnover 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. Ages used were 18 years and above.
- Some term dates were far into the future and were not included in the analysis.The only termdates used were those less than or equal to the current date.