SQL DATA ANALYSIS PROJECT

AVON HUMAN RESOURCE (HR)

MOHAMMED AMINU ABDUL WAHAB
22 August 2023

INTRODUCTION

THE SITUATION

AVON Corporation, has recently experienced increased turnover rates and is concerned about the overall composition and dynamics of its workforce. The HR department recognizes the importance of data-driven decisions and has collected extensive data on employee demographics, hiring, terminations, and departmental performance.

THE OBJECTIVE

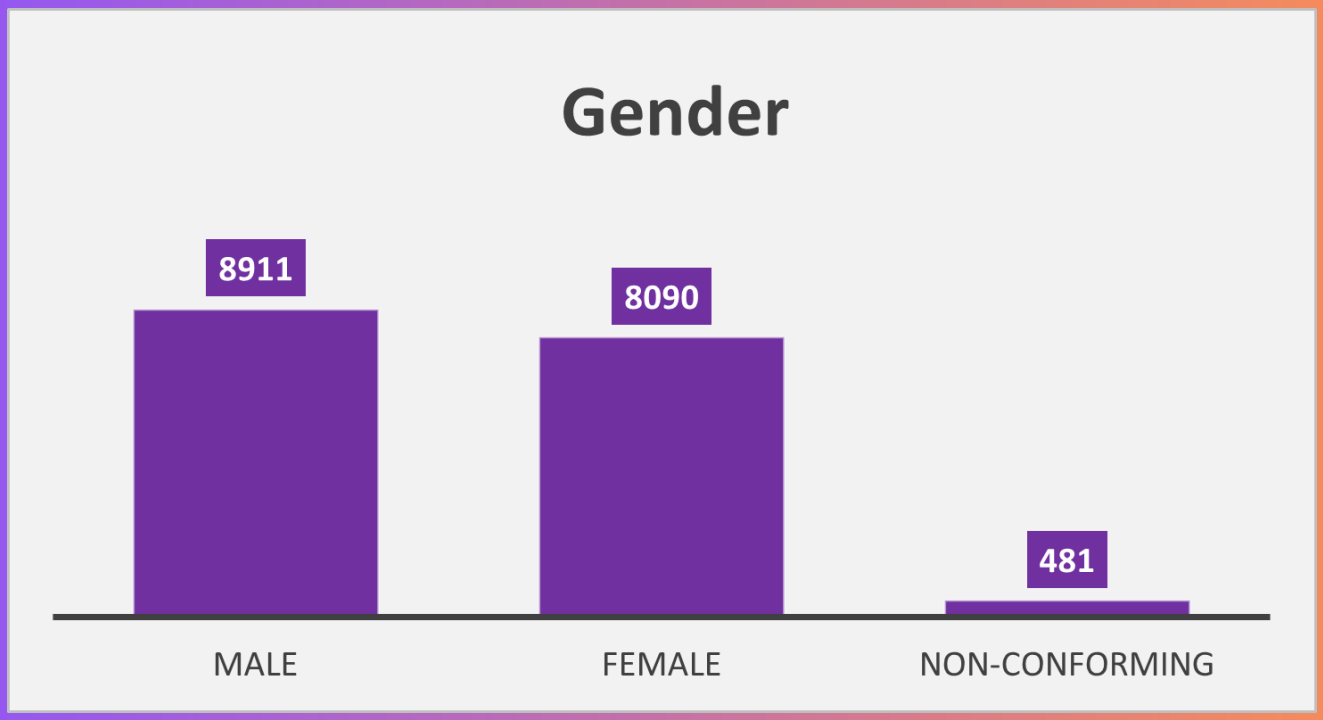
The overarching goal is to provide actionable insights to the HR department and organizational leadership, enabling them to make informed decisions to improve employee retention, diversity, and overall workforce performance.

What is the gender breakdown of employees in the company?

```sql
SELECT COUNT(*) AS employees_count, gender
FROM hr
WHERE termdate = '' OR termdate >= CURDATE()
GROUP BY gender
ORDER BY COUNT(*) DESC;
```

Result Grid | Filter Rows:

employees_count	gender
9652	Male
8804	Female
519	Non-Conforming



INSIGHT

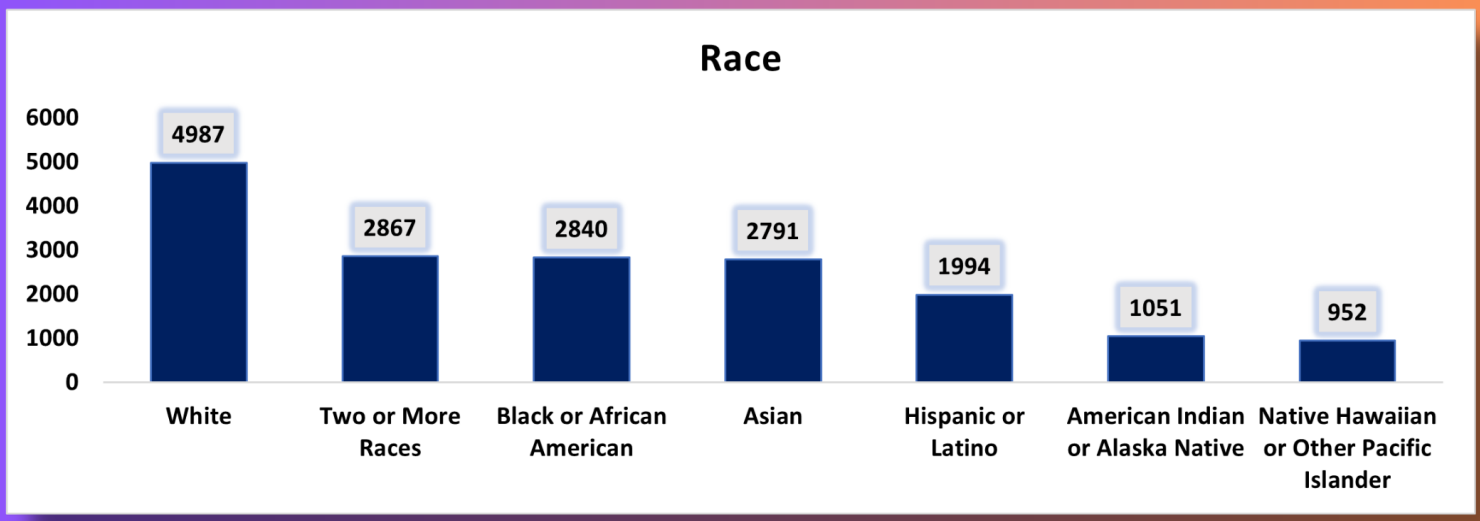
The gender distribution is skewed toward males, there are 821 more male than female and 8,430 more male than individuals who identify as Non-Conforming gender in the company. The presence of a Non-Conforming indicates that the company acknowledges diversity.

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

```sql
SELECT COUNT(*) AS employees_count, race
FROM hr
WHERE  termdate >= CURDATE() OR termdate = ''
GROUP BY race
ORDER BY  COUNT(*) DESC;
```

Result Grid | Filter Rows: | Export:

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



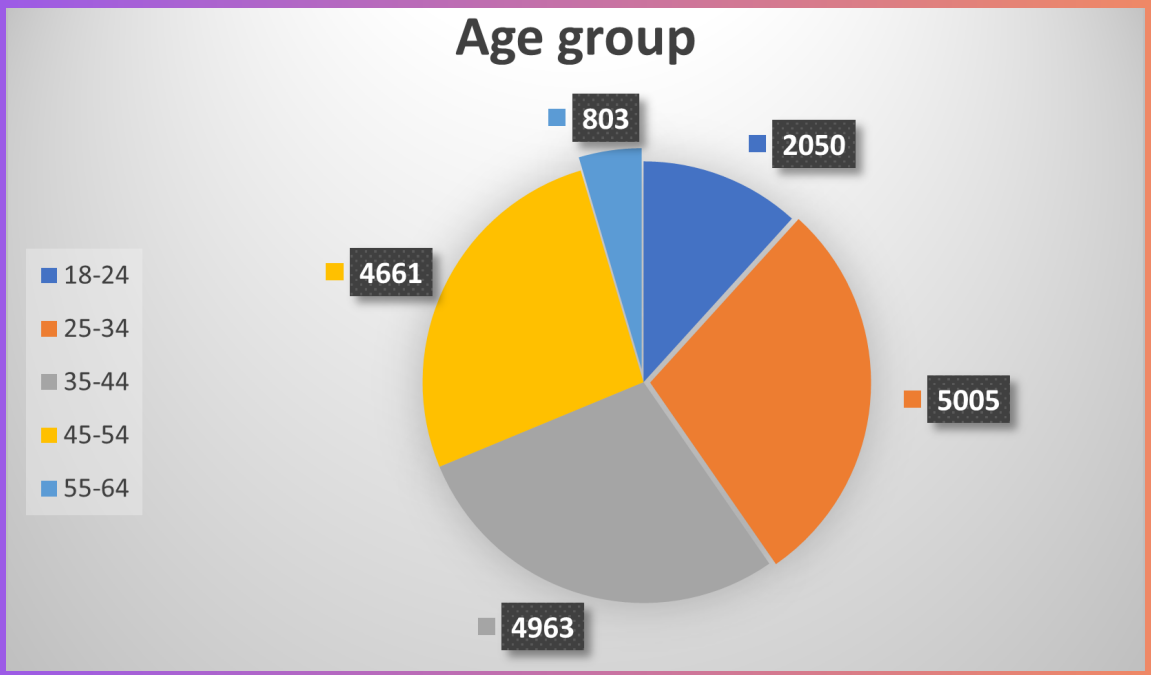
INSIGHT

The company's racial composition is primarily White, with 4,987 individuals, while Native Hawaiian or Other Pacific Islanders have the lowest representation at 952 individuals. Furthermore, there is a notable presence of individuals with multiracial backgrounds, totaling 2,867 in the "Two or More Races" category.

What is the age distribution of employees in the company?

```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 <= 54 THEN '45-54'
    WHEN age >= 55 AND age <= 64 THEN '55-64'
    ELSE '65+'
  END AS age_group,
  COUNT(*) AS count
FROM hr
WHERE  termdate >= CURDATE() OR termdate = ''
GROUP BY age_group
ORDER BY age_group;
```

age_group	count
18-24	2050
25-34	5005
35-44	4963
45-54	4661
55-64	803



INSIGHT

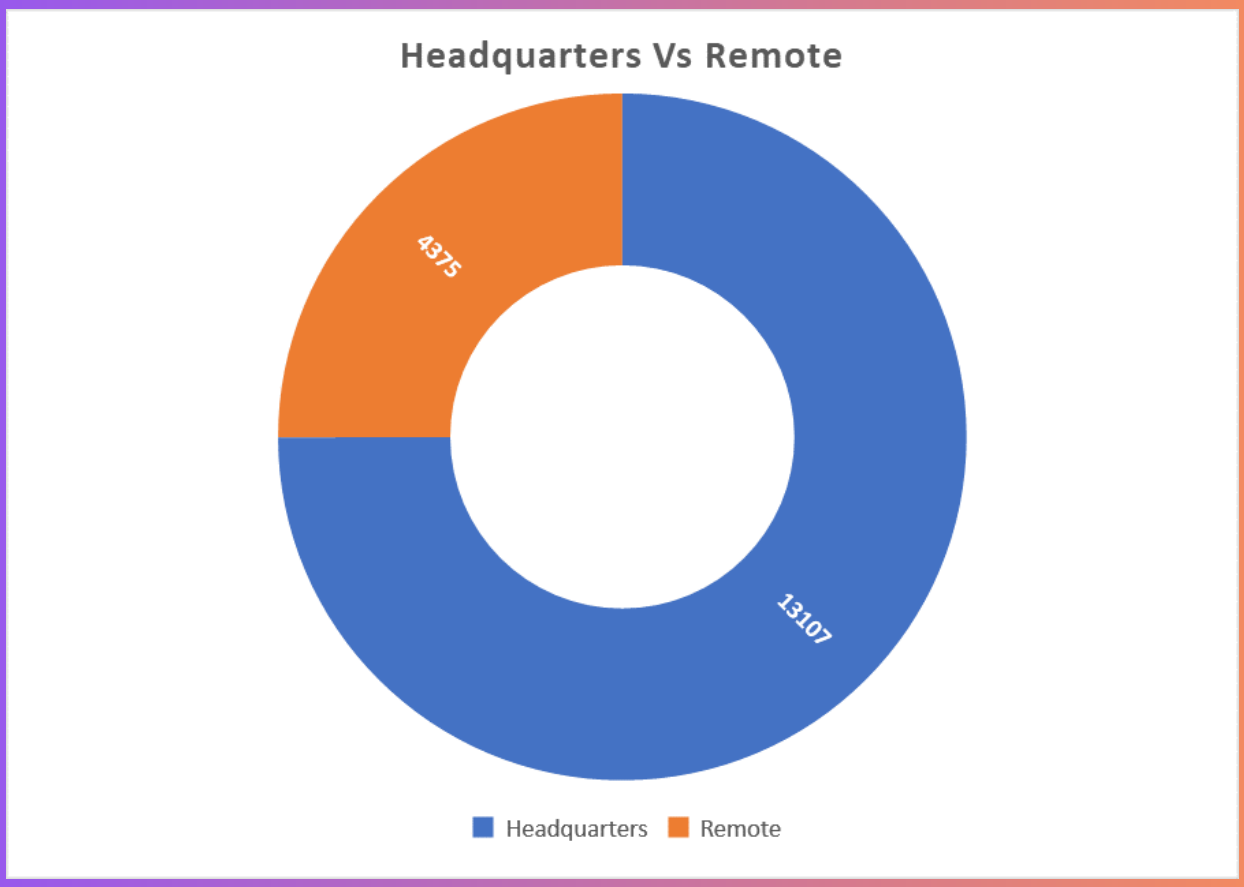
The most prevalent age group in the company is "25-34," consisting of 5,005 individuals, while the "55-64" age group has the lowest representation. Despite its smaller size, the latter group may offer valuable experience and expertise to the organization.

Headquarters Vs Remote

```sql
SELECT COUNT(*) AS Count, location
FROM hr
WHERE termdate = '' OR  termdate >= CURDATE
GROUP BY location;
```

Result Grid | Filter Rows:

	Count	location
▶	13107	Headquarters
	4375	Remote



INSIGHT

- The company's headquarters staff significantly outnumber remote workers, but the substantial remote workforce underscores the organization's commitment to flexible work arrangements.

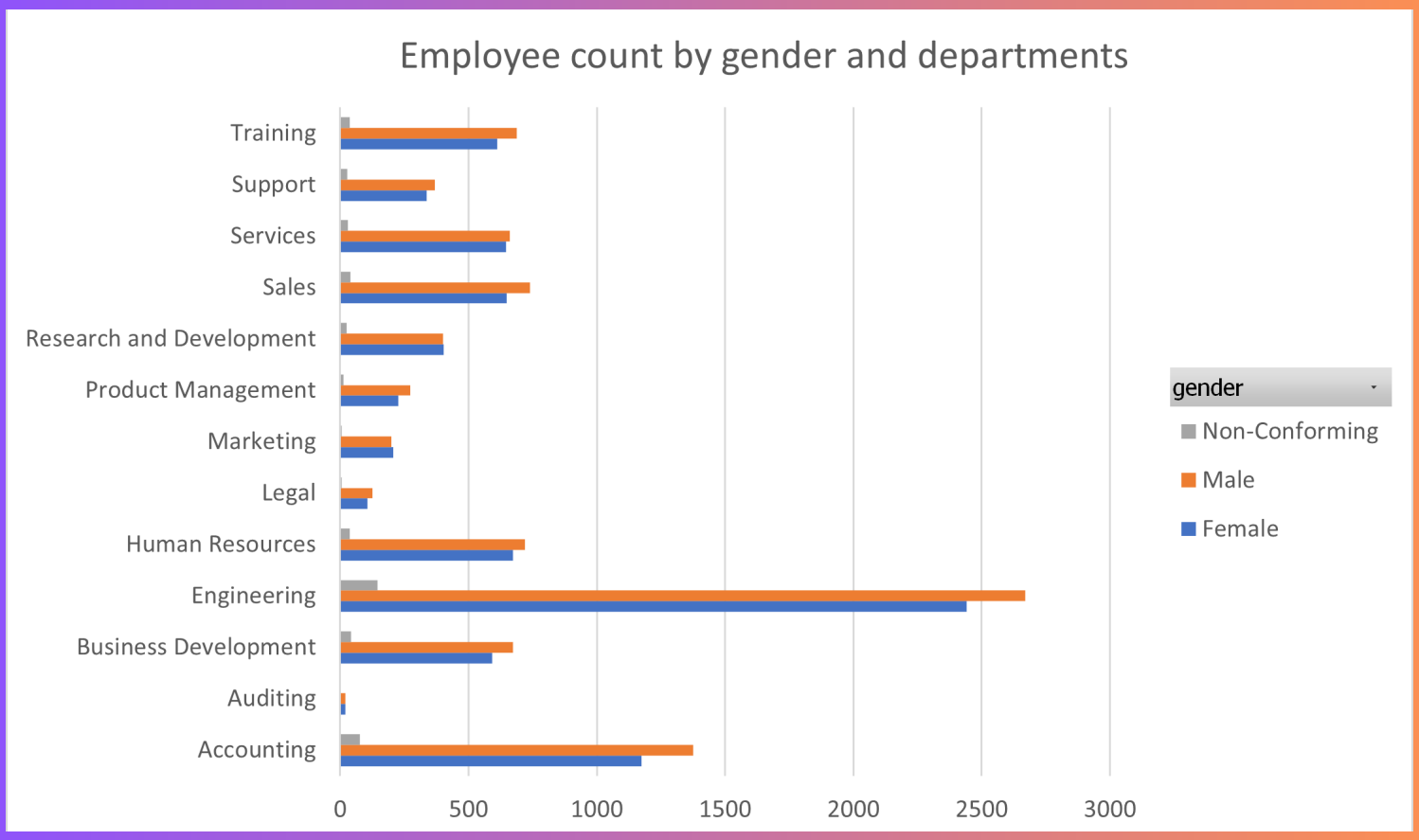
How does the gender distribution vary across departments?

```sql
SELECT COUNT(*) AS Count, gender, department
FROM hr
WHERE termdate = '' OR termdate >= CURDATE()
GROUP BY gender, department
ORDER BY department DESC;
```

Result Grid | Filter Rows:

Count	gender	department
613	Female	Training
688	Male	Training
37	Non-Conforming	Training
337	Female	Support
27	Non-Conforming	Support
368	Male	Support
647	Female	Services
661	Male	Services
29	Non-Conforming	Services
739	Male	Sales
39	Non-Conforming	Sales
648	Female	Sales
401	Male	Research ...
404	Female	Research ...
25	Non-Conforming	Research ...
272	Male	Product M...
227	Female	Product M...
13	Non-Conforming	Product M...
206	Female	Marketing
199	Male	Marketing
5	Non-Conforming	Marketing

Result 77



INSIGHT

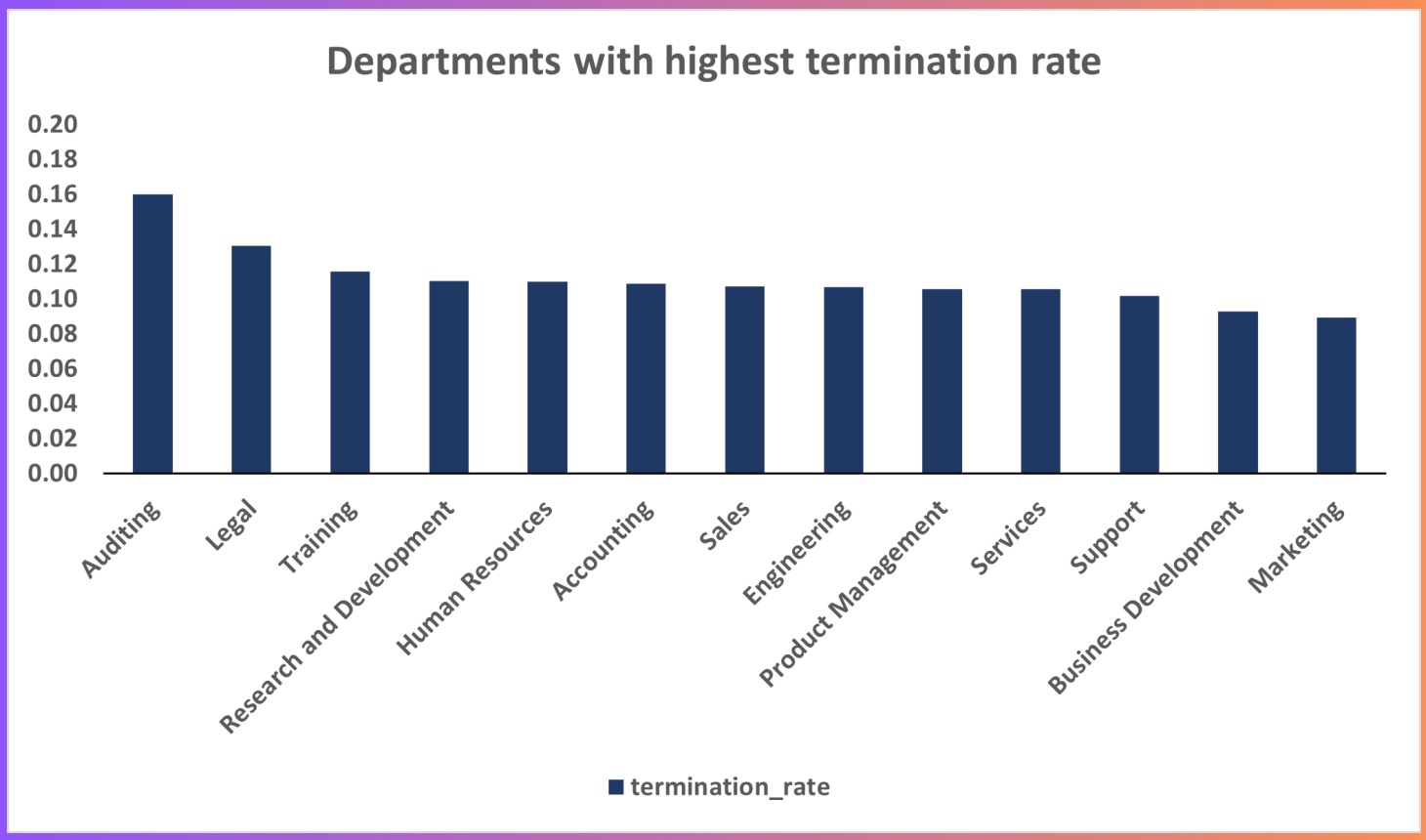
- Across the company, the Engineering department stands out with the highest representation of male, female, and non-conforming individuals, while the Auditing department has the lowest representation. Notably, the Auditing department maintains a balanced gender ratio with equal numbers of male and female employees. Additionally, the Product Management, Research and Development, Services, and Sports departments all exhibit relatively close male-female gender balances.

Which department has the highest turnover rate?

```sql
SELECT
department, total_count, termination_count,
termination_count/total_count AS termination_rate
FROM
( SELECT department,
COUNT(*) AS total_count,
SUM(CASE WHEN termdate <> '' AND termdate <=
CURDATE() THEN 1 ELSE 0 END) AS termination_count
FROM
 hr
GROUP BY
 department) AS subquery
ORDER BY
termination_rate DESC;
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

department	total_count	termination_count	termination_rate
Auditing	50	8	0.1600
Legal	299	39	0.1304
Training	1622	188	0.1159
Research and Development	1032	114	0.1105
Human Resources	1727	190	0.1100
Accounting	3192	347	0.1087
Sales	1745	187	0.1072
Engineering	6387	683	0.1069
Product Management	623	66	0.1059
Services	1618	171	0.1057
Support	903	92	0.1019
Business Development	1569	146	0.0931
Marketing	480	43	0.0896



INSIGHT

- The Auditing and Legal departments experience the highest turnover rates at 16% and 13%, respectively, while the Marketing and Business Development departments have the lowest turnover rates at 8.96% and 9.31%. Notably, the Engineering and Sales departments, both crucial functions within the organization, share a similar turnover rate of 10.66%.

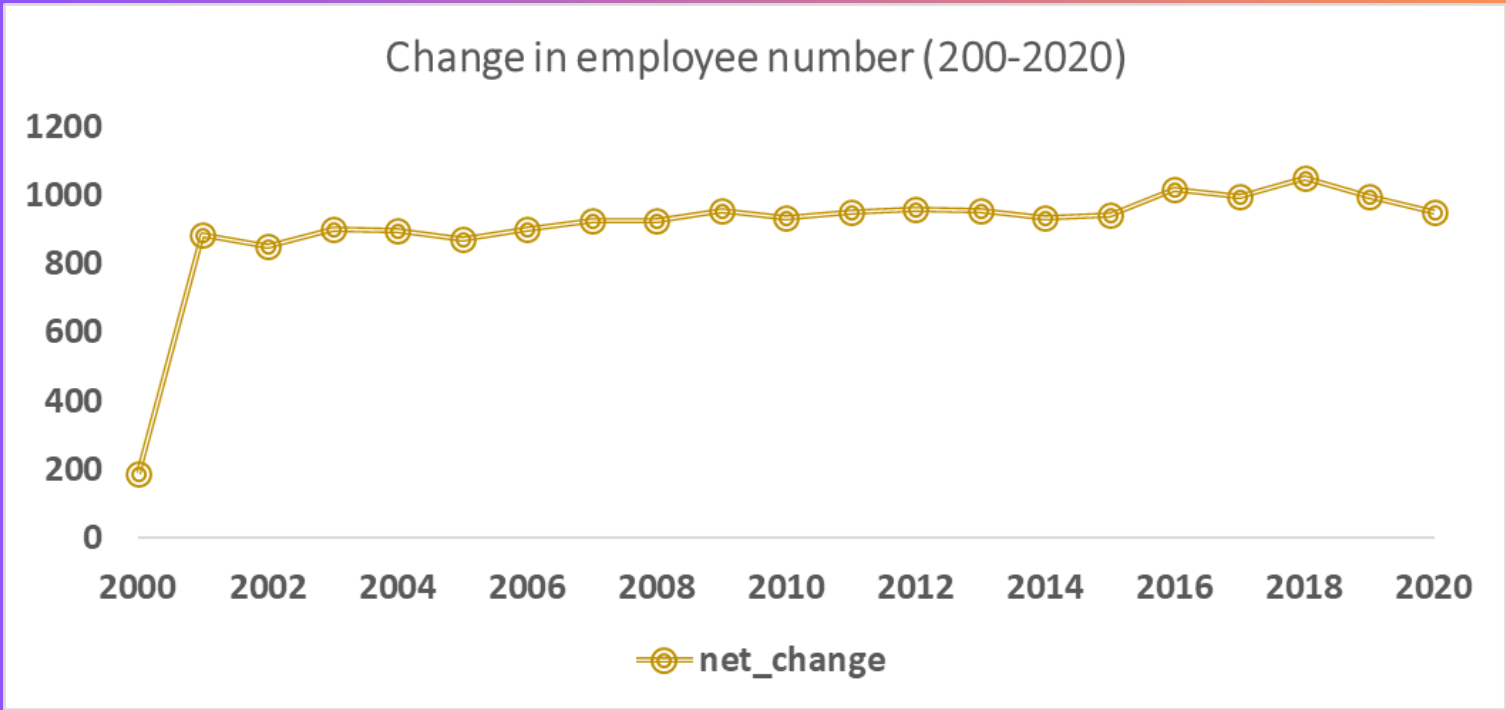
How has the company's employee count changed over time?

```
SELECT
year, hires, termination,
hires - termination AS net_change,
round((hires - termination)/hires*100,2) AS
net_change_percent
FROM
(SELECT YEAR(hire_date) AS year,
COUNT(*) AS hires,
SUM(CASE WHEN termdate <> '' AND termdate <=
CURDATE() THEN 1 ELSE 0 END) AS termination
FROM hr
GROUP BY year) AS aubquery
ORDER BY year;
```

Result Grid | Filter Rows: | Export: | Wrap

year	hires	termination	net_change	net_change_percent
2000	211	26	185	87.68
2001	1082	197	885	81.79
2002	1012	161	851	84.09
2003	1088	187	901	82.81
2004	1087	190	897	82.52
2005	1038	166	872	84.01
2006	1069	167	902	84.38
2007	1058	136	922	87.15
2008	1061	135	926	87.28
2009	1094	139	955	87.29
2010	1050	117	933	88.86
2011	1057	107	950	89.88
2012	1059	101	958	90.46
2013	1042	87	955	91.65
2014	1014	81	933	92.01
2015	1011	68	943	93.27
2016	1076	61	1015	94.33
2017	1043	47	996	95.49
2018	1090	41	1049	96.24
2019	1038	41	997	96.05
2020	967	19	948	98.04

Result 79



INSIGHT

- The organization's hiring pattern follows a cyclical trend, with fluctuations in hires ranging from a low of 967 in 2020 to a high of 1,094 in 2009. Generally, there's been consistent growth in the workforce over the years, except for a decrease of 26 employees in 2000. The years with the most significant net changes is 2020 with 98.04%.

How has the employee termination count changed over years?

```sql
SELECT YEAR(termdate) AS termination_year,
COUNT(*) AS total_employment_terminated
FROM hr
WHERE termdate <> '' AND termdate < CURDATE()
GROUP BY termination_year
ORDER BY total_employment_terminated;
```

Result Grid | Filter Rows:

termination_year	total_employment_terminated
2001	7
2002	10
2004	29
2003	31
2005	46
2007	49
2006	65
2009	70
2023	80
2011	86
2008	87
2013	105
2010	108
2014	111
2012	121
2015	128
2016	135
2017	146
2020	155
2018	160
2019	170
2022	178

Result 8

INSIGHT

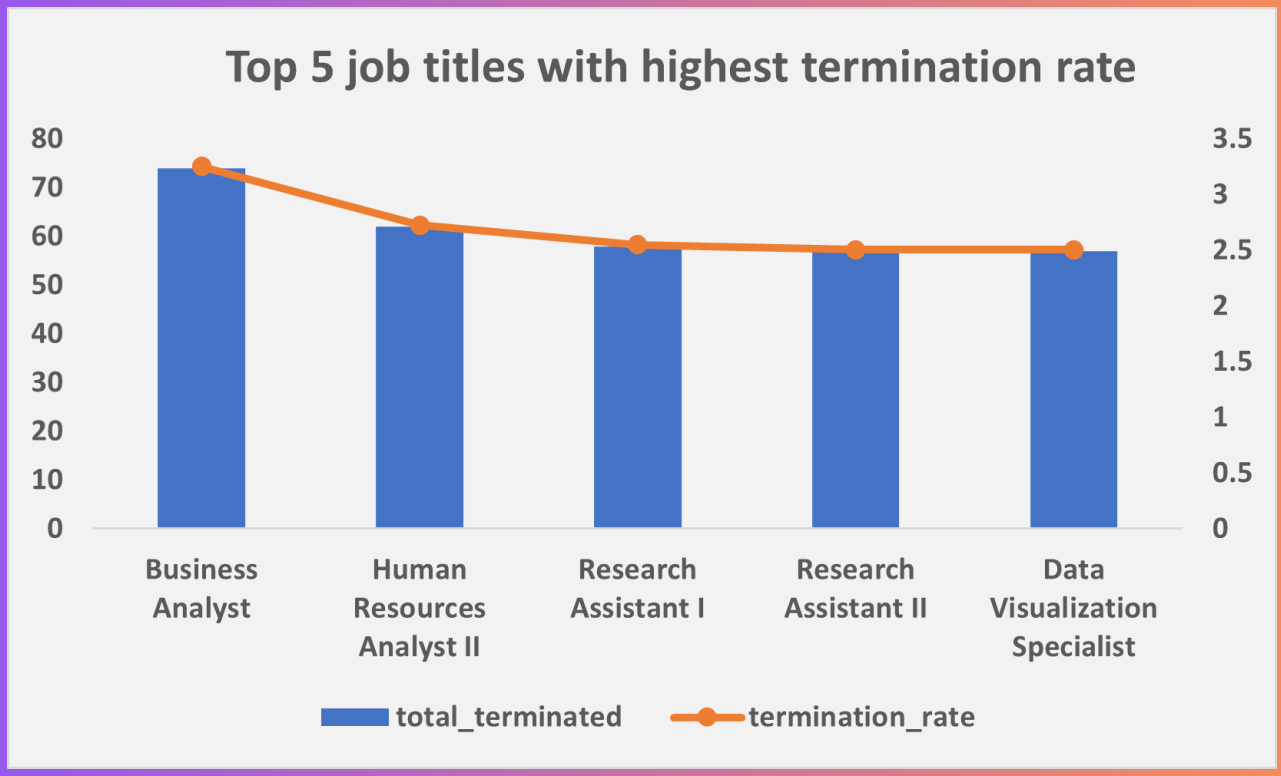
- Termination rates within the organization have fluctuated over the years, with no clear upward or downward trend. Some years, like 2001, saw minimal terminations (7), while 2021 had the highest number at 197. Periods of consecutive increases in terminations occurred in 2003-2005 and 2015-2020.

Top 5 job titles with highest termination rate?

```sql
SELECT
Jobtitle,
COUNT(*) AS total_terminated,
COUNT(*) / (SELECT COUNT(*) FROM hr WHERE termdate != '' AND
termdate <= CURDATE()) * 100 AS termination_rate
FROM
hr
WHERE
termdate != '' AND termdate <= CURDATE()
GROUP BY
Jobtitle
ORDER BY
termination_rate DESC
LIMIT 5;
```

Result Grid | Filter Rows: | Export:

Jobtitle	total_terminated	termination_rate
Business Analyst	74	3.2542
Human Resources Analyst II	62	2.7265
Research Assistant I	58	2.5506
Research Assistant II	57	2.5066
Data Visualization Specialist	57	2.5066



INSIGHT

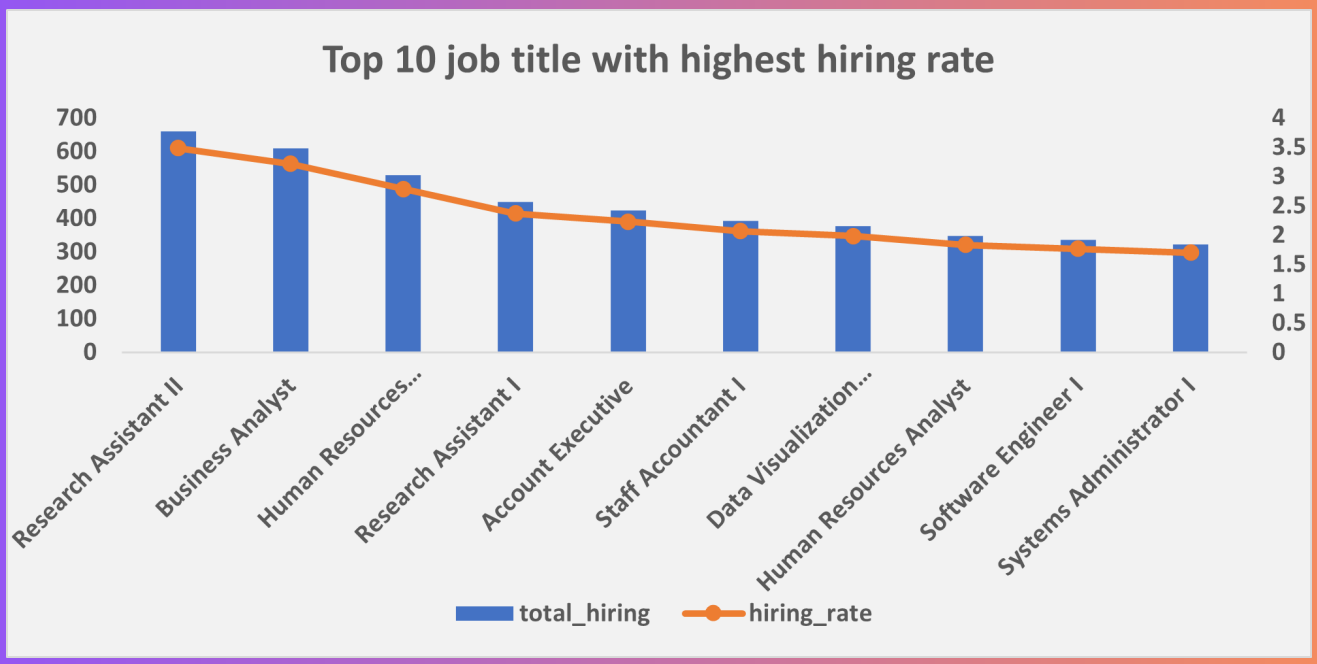
- Data Analyst job title recorded the highest termination rate in the company followed be research assistant II and I.

Top 10 job titles with highest hiring rate?


```sql
SELECT
Jobtitle, COUNT(*) AS total_hiring,COUNT(*) / (SELECT
COUNT(*) FROM hr WHERE termdate = '' OR termdate >=
CURDATE()) * 100 AS hiring_rate
FROM
hr
WHERE
termdate = '' OR termdate >= CURDATE()
GROUP BY
Jobtitle
ORDER BY
hiring_rate DESC
LIMIT 10;
```

Result Grid | Filter Rows: | Export:

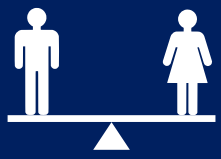
Jobtitle	total_hiring	hiring_rate
Research Assistant II	661	3.4839
Business Analyst	610	3.2151
Human Resources Analyst II	529	2.7882
Research Assistant I	449	2.3665
Account Executive	423	2.2295
Staff Accountant I	392	2.0661
Data Visualization Specialist	377	1.9870
Human Resources Analyst	348	1.8342
Software Engineer I	335	1.7657
Systems Administrator I	323	1.7024



INSIGHT

- The job title with the highest hiring rate in the company is Research Assistant II followed by Business Analyst

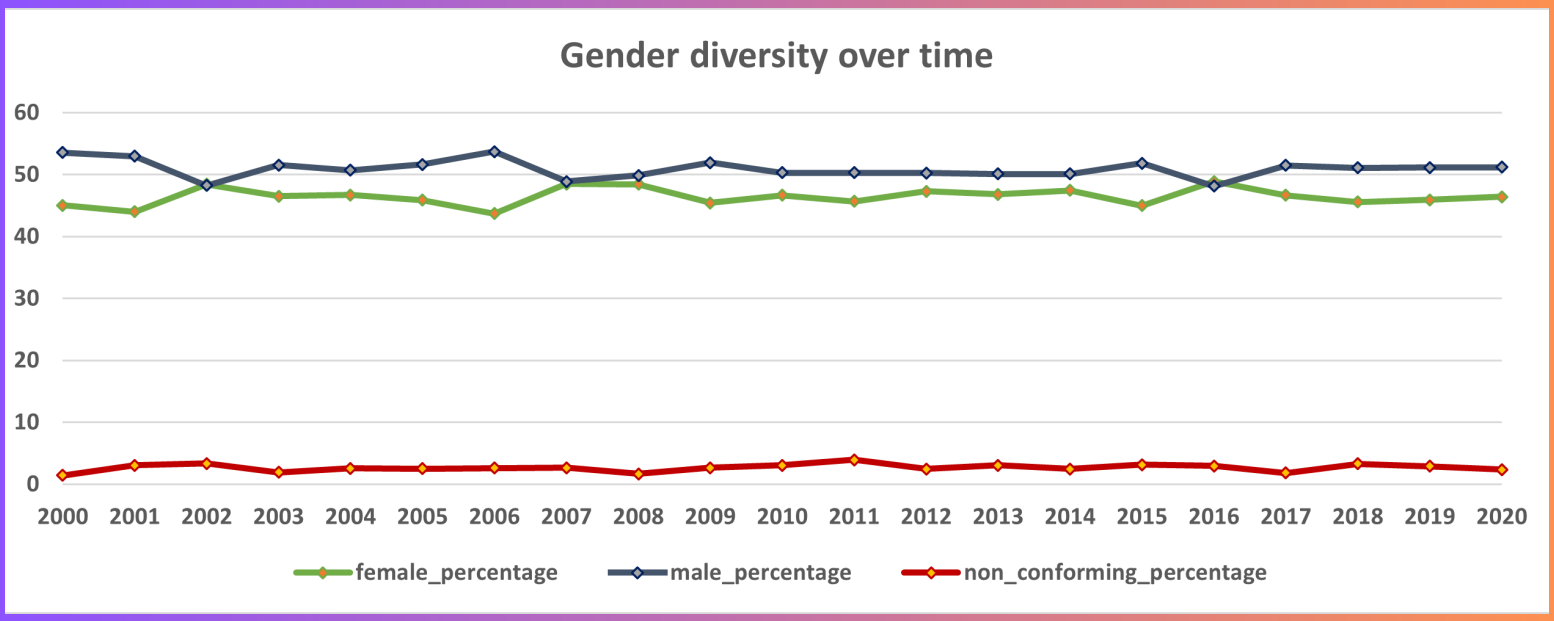
Has the workforce changed over time in terms of diversity?



```sql
SELECT YEAR(hire_date) AS hire_year, COUNT(*) AS total_employees,
      ROUND( SUM(CASE WHEN gender = 'Female' THEN 1 ELSE 0 END) /
COUNT(*)  * 100,2) AS female_percentage,

     ROUND(SUM(CASE WHEN gender = 'Male' THEN 1 ELSE 0 END) /
COUNT(*) * 100,2) AS male_percentage,

     ROUND(SUM(CASE WHEN gender = 'Non-Conforming' THEN 1 ELSE 0 END)
COUNT(*)  * 100,2) AS non_conforming_percentage

FROM hr
GROUP BY hire_year
ORDER BY hire_year;
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

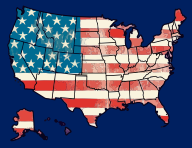
hire_year	total_employees	female_percentage	male_percentage	non_conforming_percentage
2001	1082	43.99	52.96	3.05
2002	1012	48.42	48.22	3.36
2003	1088	46.51	51.56	1.93
2004	1087	46.73	50.69	2.58
2005	1038	45.86	51.64	2.50
2006	1069	43.69	53.70	2.62
2007	1058	48.49	48.87	2.65
2008	1061	48.44	49.86	1.70
2009	1094	45.43	51.92	2.65
2010	1050	46.67	50.29	3.05
2011	1057	45.70	50.33	3.97
2012	1059	47.31	50.24	2.46
2013	1042	46.83	50.10	3.07
2014	1014	47.44	50.10	2.47
2015	1011	45.00	51.83	3.17
2016	1076	48.88	48.14	2.97
2017	1043	46.69	51.49	1.82
2018	1090	45.60	51.10	3.30
2019	1038	45.95	51.16	2.89
2020	967	46.43	51.19	2.38



INSIGHT

- Over the two-decades, the workforce has shown a modest increase in gender diversity, with the female percentage rising slightly from 45.02% to 46.43%. However, the male percentage remained relatively stable. Throughout this time frame, the non-conforming group's representation remained consistent within a low range.

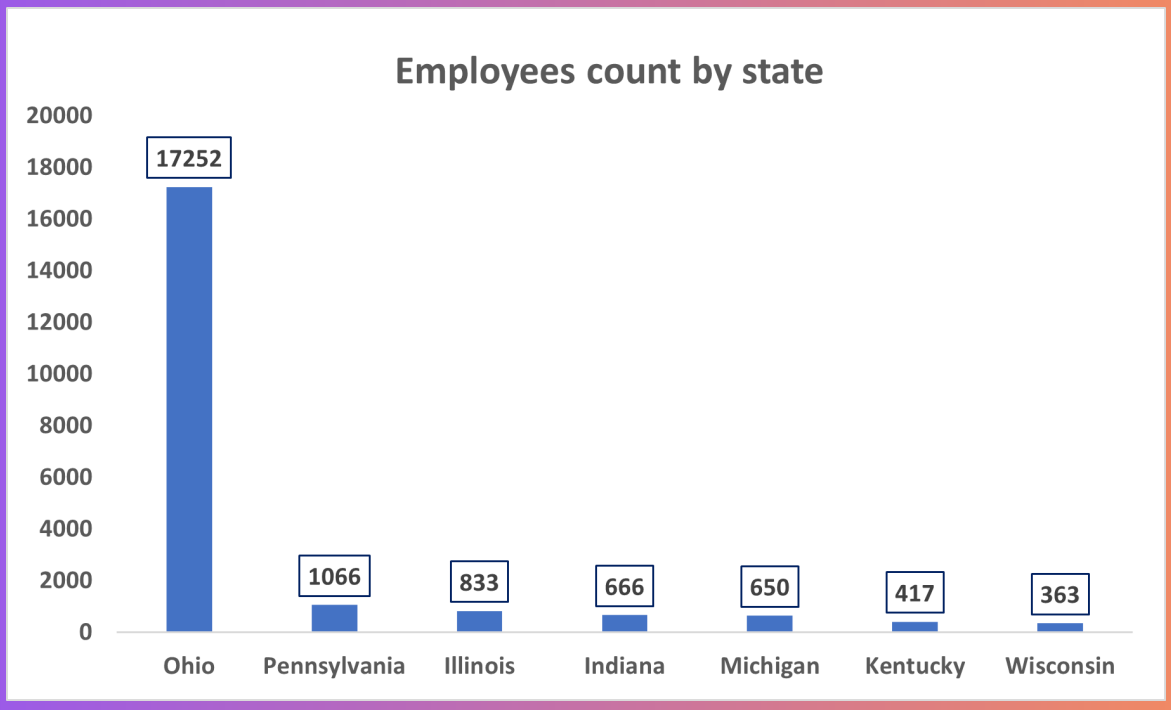
What is the distribution of employees across locations by city and state?



```sql
SELECT COUNT(*) AS count, location_state AS state
FROM hr
GROUP BY location_state
ORDER BY  COUNT(*) DESC;
```

Result Grid | Filter

count	state
17252	Ohio
1066	Pennsylvania
833	Illinois
666	Indiana
650	Michigan
417	Kentucky
363	Wisconsin



INSIGHT

Ohio has the highest employees. Ohio has more than twice the number of employees compared to the combined total of the other states.