SQL + PowerBI

This project Analyze the dataset of HR Professionals in various departments such as Auditing ,Accounting, Engineering etc. The dataset is in CSV format and it is imported to SQL where first preprocessing part of the data is done thereafter all the queries are executed in SQL and the outputs are exported In the form of CSV files.

In Second Part of this Project I have Imported all the output files from the SQL results in the powerBI respectively and then Data Visualization is done using suitable PowerBI tools.

Part 1:SQL

Step1:-In the first step I will be creating database named as **resources** and importing data in the table using Import Wizard of MySQL(here one can also used DBeaver if the import wizard is slow while importing).

Step2:- In the this step I will use database resources and query all fields in the Table "**human resources**"

Query:-

```
use resources;
SELECT * FROM resources.`human resources` ;
```

Output:-

ïˆüid	first_name	last_name	birthdate	gender	race	department	jobtitle
00-0037846	Kimmy	Walczynski	06-04-91	Male	Hispanic or Latino	Engineering	Programmer Analyst I
00-0041533	Ignatius	Springett	6/29/1984	Male	White	Business Development	Business Analyst
00-0045747	Corbie	Bittlestone	7/29/1989	Male	Black or African American	Sales	Solutions Engineer Manag
00-0055274	Baxy	Matton	9/14/1982	Female	White	Services	Service Tech
00-0076100	Terrell	Suff	04-11-94	Female	Two or More Races	Product Management	Business Analyst
00-0116166	Kacie	Offiler	1/18/1971	Male	Asian	Engineering	Developer III

clearly first column name is very weird and birthdate and hire date column don't have any fix and consistent format .

Step3:- In this step I will be changing name of column and converting both date columns in a fix format.

Query for changing the column name.

```
ALTER TABLE `human resources`
CHANGE COLUMN `ï»¿id` `emp_id` VARCHAR(20) NULL;
```

Query for changing the birthdate column to the format of '%Y-%m-%d'

```
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;
```

Query for Changing the hire date column ti the format of '%Y-%m-%d'

```
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;
```

Now to check for the data type of all the columns run below query

- `describe pranita.`human resources`;`

Output for below:-

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

Field	Type	Null	Key	Default	Extra
birthdate	text	YES		NULL	
gender	text	YES		NULL	
race	text	YES		NULL	
department	text	YES		NULL	
jobtitle	text	YES		NULL	
location	text	YES		NULL	
hire_date	text	YES		NULL	

Result 2

Clearly ,birthdate ,hire date ,termdate column should be of date type but they are of text type so data type of these columns should be changed to date type respectively .

Step:- In this step data types will be changed.

```
ALTER TABLE `human resources`
CHANGE COLUMN birthdate birthdate DATE;
```

```
• ALTER TABLE `human resources`
  CHANGE COLUMN hire_date hire_date  DATE;
```

Step:- In termdate column there is a time stamp component in the date but for our Analysis we only need date format so here we will be doing all this as below:-

```
ALTER TABLE `human resources`
ADD COLUMN new_termdate DATE;
```

In new column add the new formatted data.

```
UPDATE `human resources`
SET new_termdate = CASE
    WHEN termdate != ' ' THEN STR_TO_DATE(termdate, '%Y-%m-%d')
    ELSE NULL
END;
```

Previous column termdate is dropped and new termdate is renamed as termdate and its data type is assigned as DATE.

```
ALTER TABLE `human resources`
DROP COLUMN termdate;
```

```
ALTER TABLE `human resources`
CHANGE COLUMN new_termdate termdate DATE;
```

Add new column of age int and populate the values using timestampdiff function.

```
ALTER TABLE `human resources` ADD COLUMN age INT;

UPDATE `human resources`
SET age = TIMESTAMPDIFF(YEAR, birthdate, CURDATE());
```

All the preprocessing steps are done till now and now I will enter into the use cases part.

Case1:-What is the gender breakdown of the employees in the company?

Query:-

```
72
73 ●  select gender,count(*) as count from `human resources` where age>=18 and termdate is null
74     group by gender;
```

gender	count
Male	7429
Female	6683
Non-Conforming	411

Case2:-What is the race/ethnicity breakdown of employees in the company.?

Query:-

```
75
76 ●  select race,count(*) as count from `human resources` where age>=18 and termdate is null
77     group by race order by count desc ;
```
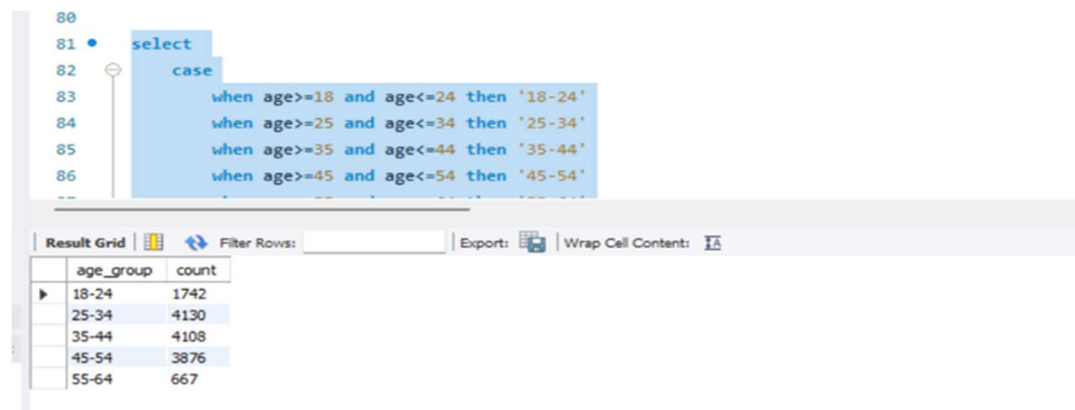
race	count
White	4115
Two or More Races	2359
Asian	2344
Black or African American	2338
Hispanic or Latino	1700
American Indian or Alaska Native	875

Case3:-What is the age distribution of employees in the company.?

Query:-

```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,gender,
    count(*) as count
from `human resources`
where age>=18 and termdate is null
group by age_group,gender
order by age_group,gender;
```
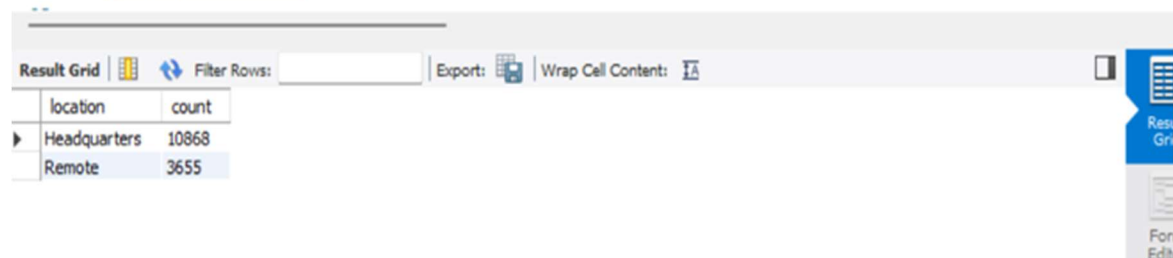
```
80
81 •    select
82   ⊖       case
83                 when age>=18 and age<=24 then '18-24'
84                 when age>=25 and age<=34 then '25-34'
85                 when age>=35 and age<=44 then '35-44'
86                 when age>=45 and age<=54 then '45-54'
```

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

age_group	count
18-24	1742
25-34	4130
35-44	4108
45-54	3876
55-64	667

Case4:- How many employees work at headquarters versus remote locations?

Query:-

```
95
96 •    select location,count(*) as count  from `human resources`  where age>=18 and termdate is null
97        group by location ;
```

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

location	count
Headquarters	10868
Remote	3655

Case5:-How does the gender distribution vary across departments and job titles?

Query:-

```
102
103 •   select department,gender,count(*) as count from `human resources`
104     where  age>=18 and termdate is null
105     group by department,gender
106     order by department;
```

department	gender	count
Accounting	Female	972
Accounting	Male	1162
Accounting	Non-Conforming	63
Auditing	F Female	15
Auditing	Male	18
Business Development	Female	487

Result 48 ✕

Case6:-What is the average length of employment for employees who have been terminated?

Query:-

```
98
99 •   select avg(datediff(termdate,hire_date))/365 as avg_length_employment
100    from `human resources`
101    where termdate is not null and termdate<=curdate() and age>=18;
```

avg_length_employment
7.49149574

Case7:- What is the distribution of job titles across the company?

Query:-

```
107
108 •   select jobtitle,count(*) from `human resources`
109    where  age>=18 and termdate is null
110    group by jobtitle
111    order by jobtitle desc;
```

jobtitle	count(*)
Web Developer IV	46
Web Developer III	45
Web Developer II	56
Web Developer I	70
Web Designer IV	5
Web Designer III	8

Result 50 ✕

Case8:- Which department has the highest turnover rate?

Query:-

```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 IS NOT NULL AND termdate <= CURDATE() THEN 1 ELSE 0 END) AS terminated_count
    FROM `human resources`
    WHERE age >= 18
    GROUP BY department
) AS subquery
ORDER BY termination_rate desc;
```

department	total_count	terminated_count	termination_rate
Auditing	44	8	0.1818
Legal	250	34	0.1360
Training	1346	155	0.1152
Research and Development	851	97	0.1140
Human Resources	1461	159	0.1088
Product Management	518	56	0.1081

Case9:- What is the distribution of employees across locations by city and state?

Query:-

```sql
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	11750
Pennsylvania	740
Illinois	585
Michigan	460
Indiana	438
Kentucky	294

Case 10:- How has the company's employee count changed over time based on hire and term date?

Query:-

```
144        order by year asc;
145
146  •    SELECT year, hires, terminations, hires - terminations AS net_change,
147               ROUND((hires - terminations) / hires * 100, 2) AS net_change_percent
148    ⊖  FROM (
149            SELECT YEAR(hire_date) AS year, COUNT(*) AS hires,
150                   SUM(CASE WHEN termdate IS NOT NULL AND termdate <= CURDATE() THEN 1 ELSE 0 END) AS terminations
151            FROM `human resources`
152            WHERE age >= 18
153            GROUP BY YEAR(hire_date)
154        ) AS subquery
155        ORDER BY year ASC;
156
157
```

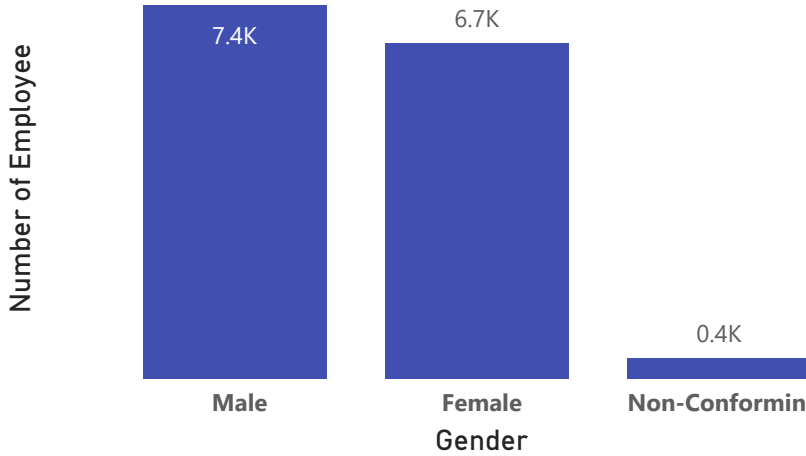| year	hires	terminations	net_change	net_change_percent
2000	170	19	151	88.82
2001	897	155	742	82.72
2002	845	134	711	84.14
2003	892	153	739	82.85
2004	909	145	764	84.05
2005	865	135	730	84.39
2006	884	141	743	84.05

Result 55 ✕

Output

Action Output

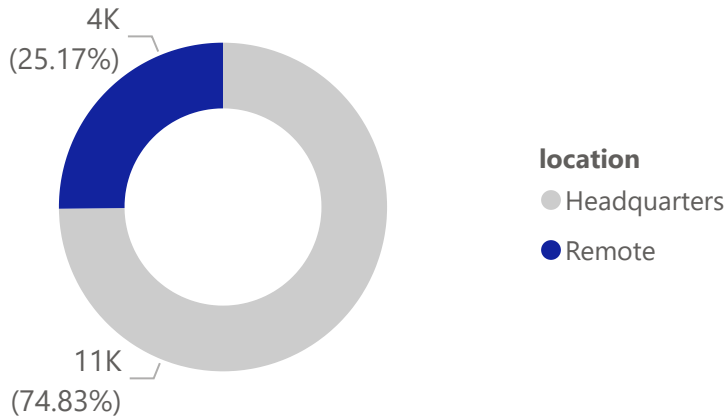# DATA VISUALIZATION BY POWERBI

HR Employee Distribution Reports

Average Length of Employment(Years)
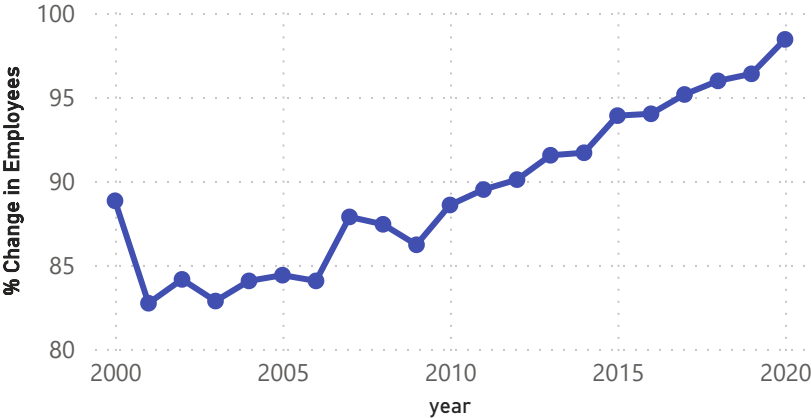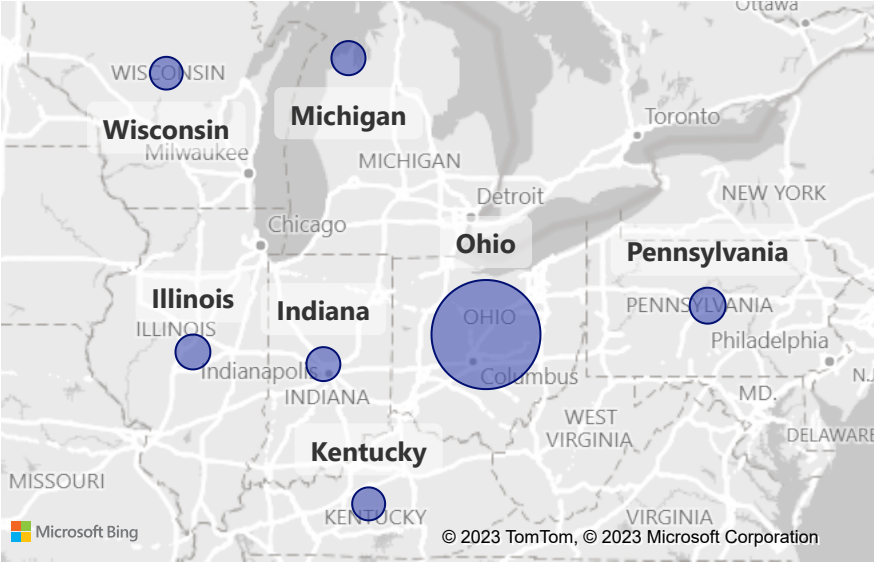
7.49

Gender Distribution

Number of Employee (y-axis)
Gender (x-axis): Male 7.4K, Female 6.7K, Non-Conformin 0.4K

HeadQuarters Vs Remote

4K (25.17%) — Remote
11K (74.83%) — Headquarters

location
- Headquarters
- Remote

Change in Employee Number 2000-2020

% Change in Employees (y-axis: 80 to 100)
year (x-axis: 2000 to 2020)

Employees By State

Wisconsin, Michigan, Illinois, Indiana, Ohio, Pennsylvania, Kentucky

Microsoft Bing
© 2023 TomTom, © 2023 Microsoft Corporation

Race Distribution

Number of Employees (y-axis)
- White 4.1K
- Two or More Races 2.4K
- Asian 2.3K
- Black or African American 2.3K
- Hispanic or Latino 1.7K
- American Indian or Alaska Native 0.9K
- Native Hawaiian or Other Pacific Islander 0.8K
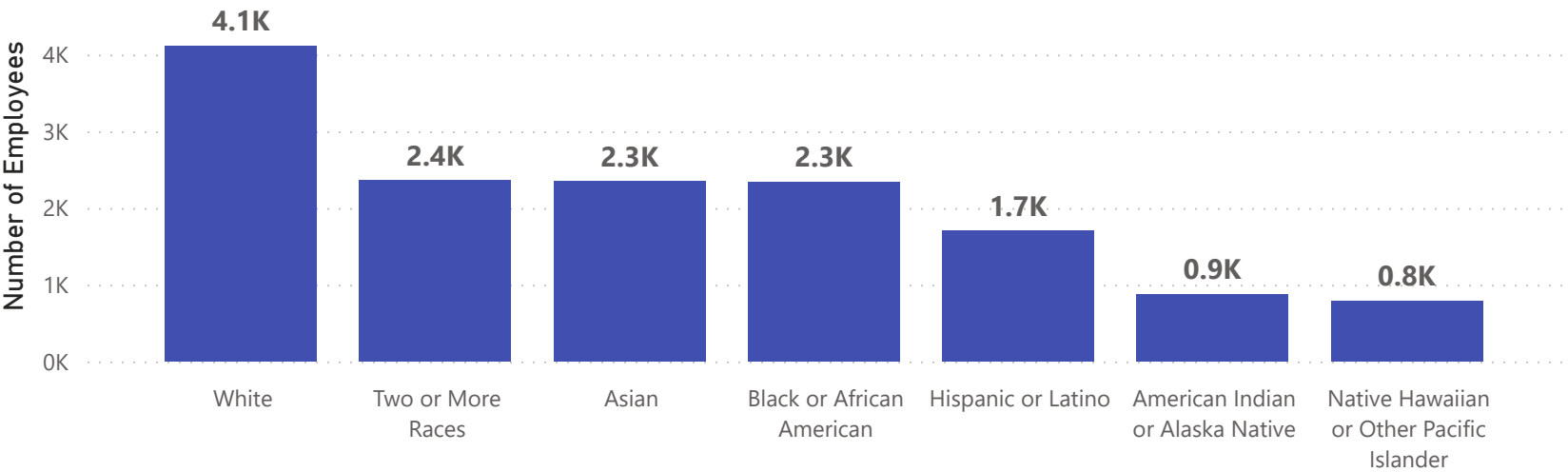
Count of count by age_group and gender

gender ● Female ● Male ● Non-Conforming
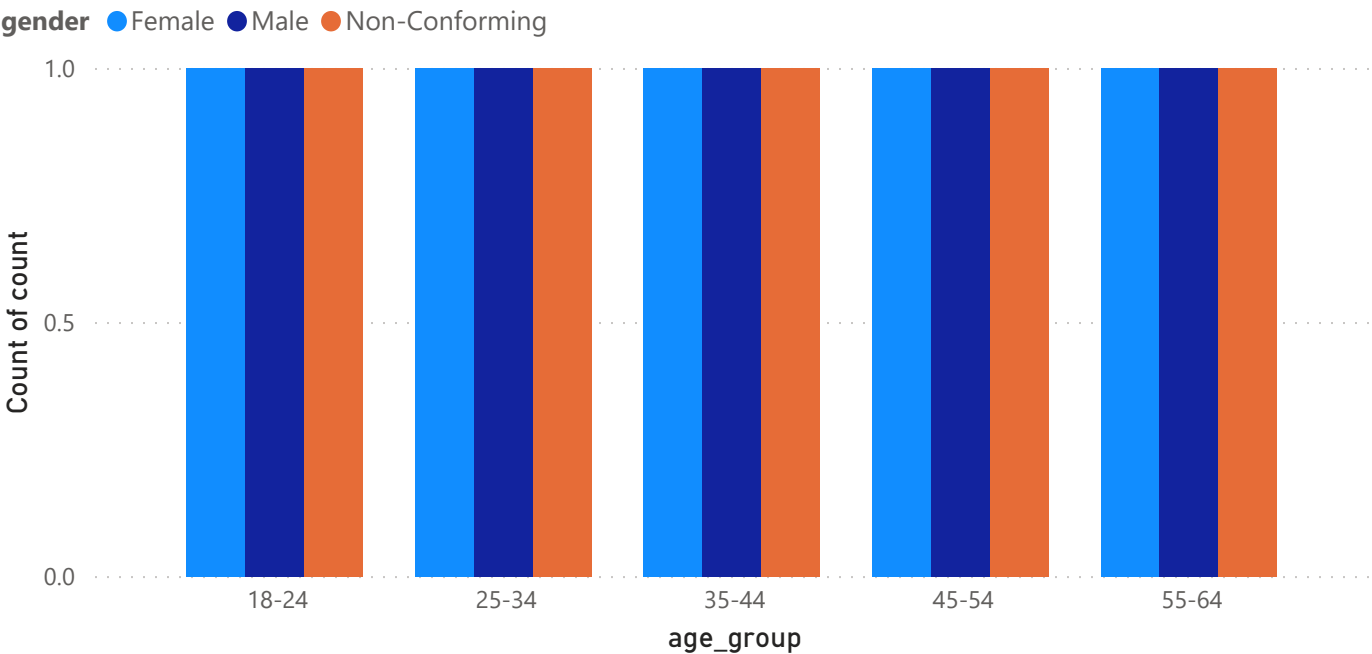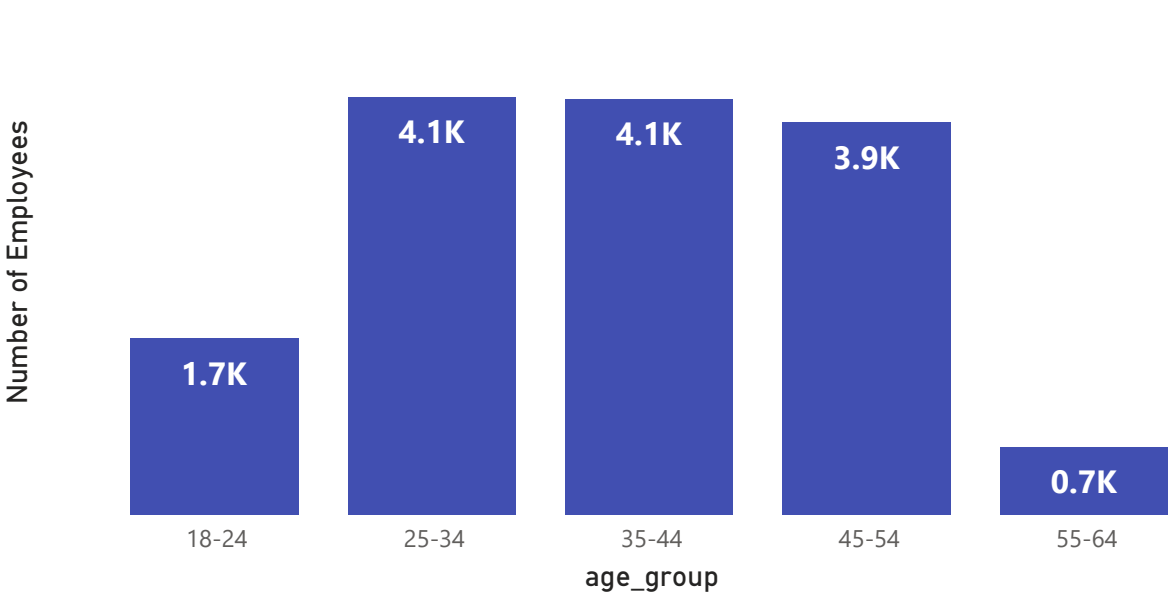
department	Sum of termination_rate
Product Management	0.11
Engineering	0.11
Services	0.11
Accounting	0.10
Sales	0.10
Support	0.10
Business Development	0.10
Marketing	0.08

Age Group Distribution

Gender Distribution By Department

gender ● Female ● Male ● Non-Conforming