Project Report of Human Resources Data Analysis and Visualization

For

Technology in Business

Submitted To: - Shashank Gaikwad Sir

Submitted By – Akash Kumar

PRN – 23030141003

Swarn Prabha

PRN – 23030141069

1. Introduction

The project involved analyzing HR data to gain insights into employee demographics, departmental distributions, and termination rates. The process was executed in three major steps: data cleaning and transformation using MySQL Workbench, exporting the cleaned data into CSV files, and finally creating visualizations using PowerBI.

2. **Data Cleaning and Transformation Using MySQL**

The initial dataset was imported into MySQL Workbench for data cleaning. Several queries were written to address missing values, inconsistent data formats, and other anomalies within the dataset. Below are the major steps and queries used:

- **Snapshot of the Data:**

id	first_nam	last_name	birthdate	gender	race	departme	jobtitle	location	hire_date	termdate	location_c	location_state
00-003784	Kimmy	Walczynsk	06-04-1991	Male	Hispanic o	Engineerir	Programm	Headquar	1/20/2002		Cleveland	Ohio
00-004153	Ignatius	Springett	6/29/1984	Male	White	Business [	Business A	Headquar	04-08-2019		Cleveland	Ohio
00-004574	Corbie	Bittleston	7/29/1989	Male	Black or A	Sales	Solutions	Headquar	10-12-2010		Cleveland	Ohio
00-005527	Baxy	Matton	9/14/1982	Female	White	Services	Service Te	Headquar	04-10-2005		Cleveland	Ohio
00-00761(	Terrell	Suff	04-11-1994	Female	Two or M	Product M	Business A	Remote	9/29/2010	2029-10-2	Flint	Michigan
00-011616	Kacie	Offiler	1/18/1971	Male	Asian	Engineerir	Developer	Headquar	09-01-2018		Cleveland	Ohio
00-036318	Sandro	Admans	11/19/1979	Male	Two or M	Product M	Quality En	Headquar	11-08-2012		Cleveland	Ohio
00-03807(	Eugene	Lehrahan	10/14/1988	Female	Black or A	Engineerir	Developer	Headquar	6/27/2007		Cleveland	Ohio
00-038166	Wainwrigh	Corfield	12/13/1996	Male	Asian	Engineerir	Business S	Headquar	2/20/2001	2008-12-0	Cleveland	Ohio
00-04192(	Dyann	Isoldi	3/27/1980	Male	Two or M	Engineerir	Web Deve	Headquar	1/27/2005		Cleveland	Ohio
00-047228	Grantley	Oret	09-06-1975	Male	Two or M	Services	Service Te	Headquar	11-01-2004		Cleveland	Ohio
00-047283	Elmore	Worner	01-07-1966	Female	White	Engineerir	Business S	Headquar	12-05-2000		Cleveland	Ohio
00-056638	Dud	Brain	3/17/1984	Male	Two or M	Business [	Business A	Headquar	9/17/2008		Cleveland	Ohio
00-057107	Aguie	Conford	11-02-1971	Male	White	Business [	Research /	Headquar	11/25/2015		Cleveland	Ohio
00-062418	Katerina	Rosborou	8/20/1967	Male	Hispanic o	Engineerir	Analyst Pr	Headquar	5/17/2019		Cleveland	Ohio
00-071521	Alida	Longley	1/28/1973	Female	American	Accountin	Staff Acco	Headquar	02-04-2002		Cleveland	Ohio
00-075564	Laraine	Petre	05-11-1967	Male	White	Engineerir	Software I	Headquar	9/30/2010		Cleveland	Ohio

- **Handling Missing Values**: Rows with missing or null data in critical fields like employee age, department, or gender were identified and handled.

```
SELECT * FROM hr WHERE age IS NULL;
```

- **Standardizing Data Formats**: Inconsistencies in data formats (e.g., gender categories, date formats) were cleaned using

queries. We ensured that data such as 'Male', 'Female', and 'non-conforming' were consistently represented.

```
SELECT birthdate FROM hr;
SET sql_safe_updates = 0;
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;

ALTER TABLE hr
MODIFY COLUMN birthdate DATE;
```

- **Filtering Relevant Data**: The dataset was filtered to focus on specific age groups, departments, and termination rates to create meaningful insights.

```
SET age = timestampdiff(YEAR, birthdate, CURDATE());

SELECT
    min(age) AS youngest,
    max(age) AS oldest
FROM hr;
```

- **Filtering Relevant Data (Questions):**

```
-- QUESTIONS

-- 1. What is the gender breakdown of employees in the company?
-- 2. What is the race/ethnicity breakdown of employees in the company?
-- 3. What is the age distribution of employees 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 city and state?
- -- 10. How has the company's employee count changed over time based on hire and term dates?
- -- 11. What is the tenure distribution for each department?

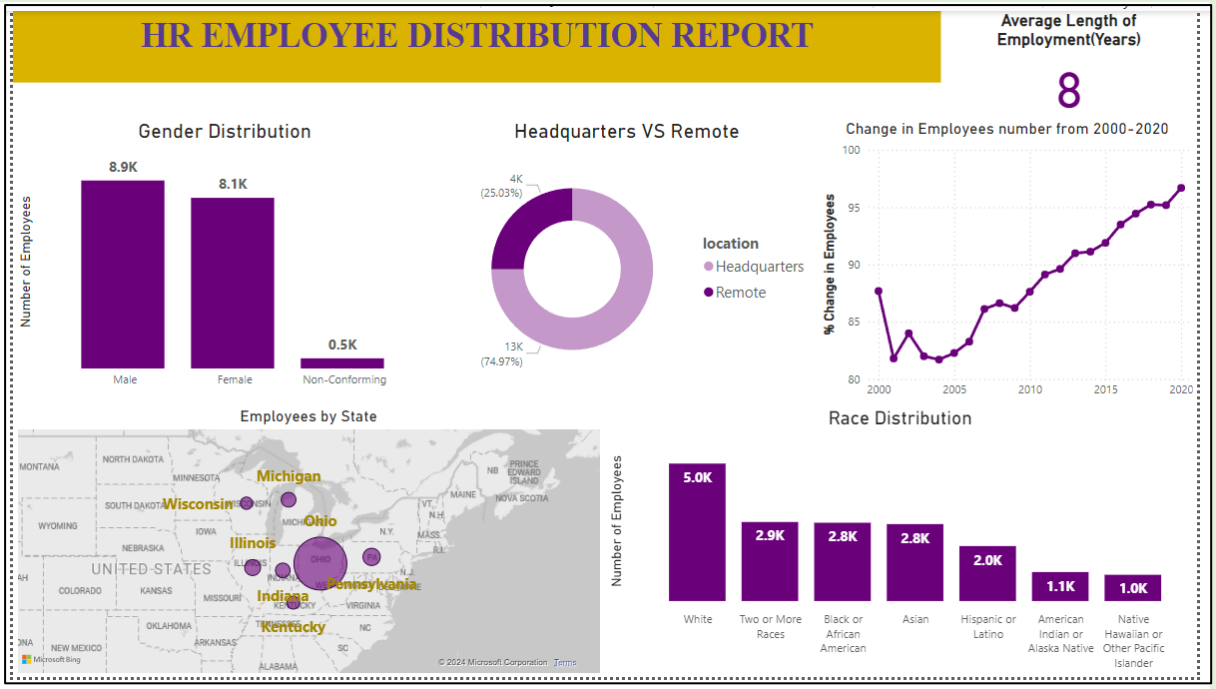
Sample SQL Queries for the questions

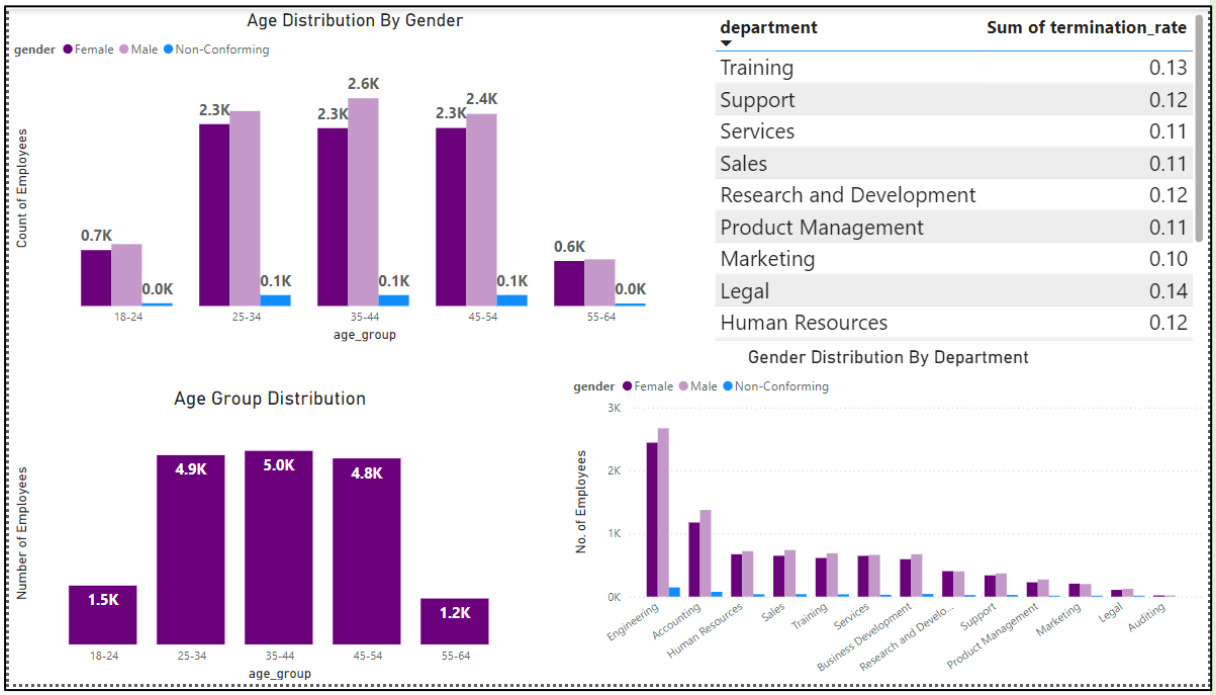
```sql
-- 1. What is the gender breakdown of employees in the company?--
select gender, count(*) as count
from hr
where age>=18 and termdate IS NULL
group by gender;

-- 2. What is the race/ethnicity breakdown of employees in the company?
select race, count(*) as count
from hr
where age>=18 and termdate IS NULL
group by race
order by count(*) desc;

-- 3. What is the age distribution of employees in the company?
select
    min(age) as youngest,
    max(age) as oldest
from hr
where age>=18 and termdate IS NULL;
```

- **Exporting Cleaned Data**: Once the data was cleaned and formatted, it was exported into CSV files. These files were later used for visualization in PowerBI.

3. Data Visualization Using PowerBI

Once the data was exported to CSV files, PowerBI was used to create visual representations of the cleaned data. The key areas of analysis include employee distribution by age, gender, department, and termination rates.



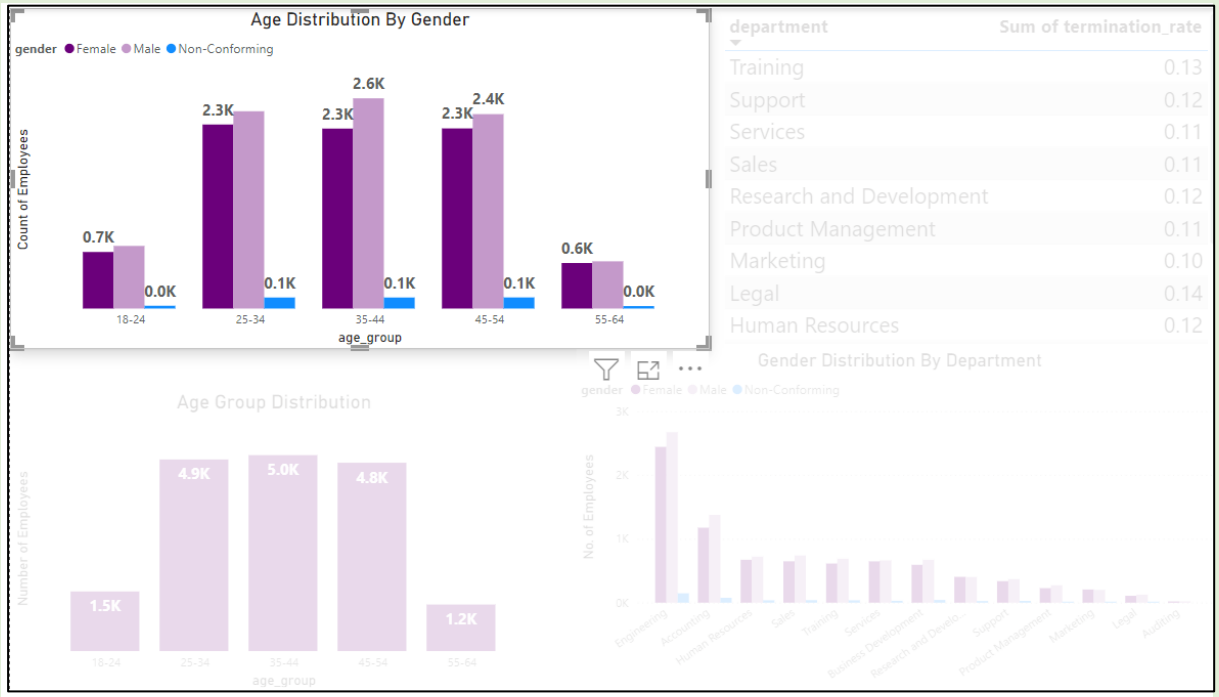


a. Age Distribution by Gender

One of the main visualizations created was an age distribution by gender. It shows the number of employees in different age groups (18-24, 25-34, 35-44, 45-54, 55-64) divided by gender (Male, Female, Non-Conforming).

- **Key Insights**:

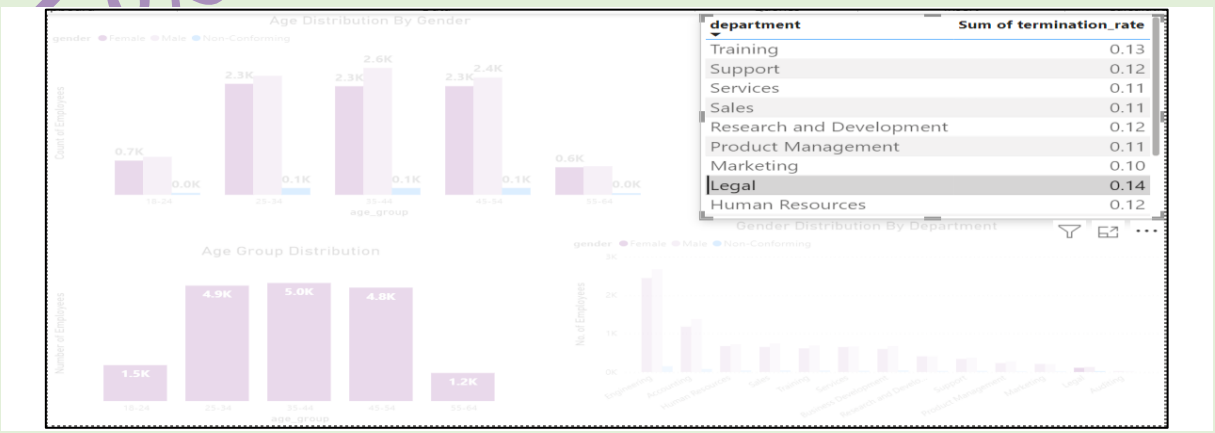
- The highest number of employees are in the 25-34 and 35-44 age groups.
- Gender distribution shows a fairly balanced representation between males and females across most age groups, with a minor representation for non-conforming individuals.



b. Termination Rate by Department

A table was created to show the termination rates across different departments. Departments like Legal, Training, and Support had higher termination rates compared to others.

- **Key Insights**:
  - **Legal Department** shows the highest termination rate (0.14), which could indicate possible issues related to work culture or job dissatisfaction in this department.
  - Departments like **Marketing** have lower termination rates, suggesting a more stable workforce.



4. Conclusion

The HR data analysis revealed several insights into the workforce demographics and departmental trends. The age group distribution indicates a younger workforce with a peak in the 25-44 age range. Gender representation appears balanced across most age groups, though some departments may require more attention in terms of diversity. Finally, termination rates vary across departments, with certain areas needing further investigation to improve employee retention.

The project successfully demonstrated how SQL-based data cleaning and PowerBI visualizations can provide actionable insights into HR data.

5. Recommendations

- **Focus on Retention Strategies**: For departments with high termination rates, particularly Legal and Training, it may be beneficial to investigate the underlying causes of employee turnover and implement retention strategies.
- **Promote Gender Diversity**: Departments with uneven gender distribution can benefit from diversity and inclusion programs to ensure equal representation across the organization.
- **Maintain a Balanced Workforce**: The current age distribution shows a balanced workforce in the 25-44 range, but attention should be given to succession planning for employees in the 45+ age group as they approach retirement.

6. Appendix

- **SQL Queries Used for Data Cleaning**

SQL_Data_Cleaning.sql | Simple Linear Regression.ipynb | HR Data Questions.sql | -- 1. What is the gender breakdow

C: > Users > khila > OneDrive > Documents > Technology in Business > Project > SQL_Data_Cleaning.sql

```sql
CREATE DATABASE projects;

USE projects;

SELECT * FROM hr;

ALTER TABLE hr
CHANGE COLUMN i»¿id emp_id VARCHAR(20) NULL;

DESCRIBE hr;

SELECT birthdate FROM hr;

SET sql_safe_updates = 0;

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;

ALTER TABLE hr
MODIFY COLUMN birthdate DATE;

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')
```

- **CSV Files Used for PowerBI Visualization**

C: > Users > khila > OneDrive > Documents > Technology in Business > Project > HR-Dashboard-MySQL-PowerBI-main > HR Data Questions.sql

```
-- QUESTIONS

-- 1. What is the gender breakdown of employees in the company?
-- 2. What is the race/ethnicity breakdown of employees in the company?
-- 3. What is the age distribution of employees 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 city and state?
-- 10. How has the company's employee count changed over time based on hire and term dates?
-- 11. What is the tenure distribution for each department?

```

Name	Status	Date modified	Type	Size
age_group	✓	02-09-2024 13:53	Microsoft Excel Com...	1 KB
age_group_gender	✓	02-09-2024 13:57	Microsoft Excel Com...	1 KB
avg_length_employment	✓	02-09-2024 14:14	Microsoft Excel Com...	1 KB
avg_tenure	✓	02-09-2024 20:57	Microsoft Excel Com...	1 KB
employee_change	✓	02-09-2024 20:54	Microsoft Excel Com...	1 KB
gender	✓	02-09-2024 13:39	Microsoft Excel Com...	1 KB
gender_department	✓	02-09-2024 20:34	Microsoft Excel Com...	1 KB
jobtitle	✓	02-09-2024 20:35	Microsoft Excel Com...	5 KB
location	✓	02-09-2024 13:59	Microsoft Excel Com...	1 KB
race	✓	02-09-2024 13:45	Microsoft Excel Com...	1 KB
state	✓	02-09-2024 20:44	Microsoft Excel Com...	1 KB
turnover_rate	✓	02-09-2024 20:42	Microsoft Excel Com...	1 KB