HR REPORT

This project dives deep into the realm of data analysis using SQL and Power BI to uncover important human resource insights that can greatly benefit the company.

```
-- Create database

CREATE DATABASE hr_DATA;

-- After loading DB

USE hr_DATA;

-- Explore the loaded data into hr_data

SELECT *
FROM hr_data;


-- Explore table structure

SELECT COLUMN_NAME, DATA_TYPE, CHARACTER_MAXIMUM_LENGTH
FROM INFORMATION_SCHEMA.COLUMNS
WHERE TABLE_NAME = 'hr_data';

-- Fix column "termdate" formatting

-- format termdate datetime UTC values

-- Update date/time to date

UPDATE hr_data
SET termdate =FORMAT(CONVERT(DATETIME,LEFT(termdate,19),120),'yyyy-mm-dd');


-- Update from nvachar to date

-- First, add a new date column

ALTER TABLE hr_data
ADD new_termdate DATE;

-- Update the new date column with the converted values

UPDATE hr_data
SET new_termdate=CASE
   WHEN termdate IS NOT NULL AND ISDATE(termdate)=1 THEN CAST(termdate As DATETIME)ELSE NULL END;


SELECT termdate
FROM hr_data
ORDER BY termdate DESC


-- Create new column "age"

ALTER TABLE hr_data
ADD age nvarchar(50);


--Populate new column with age
UPDATE hr_data
SET age=DATEDIFF(YEAR,birthdate,GETDATE());

SELECT age
FROM hr_data;

--QUESTIONS TO ANSWER FROM THE DATA

--1.WHATS THE AGE DISTRIBUTION IN THE COMPANY?
```

```sql
--Age distribution

SELECT
MIN(age) AS youngest,
MAX(age) AS oldest
FROM hr_data;

--Age group by gender
SELECT age
FROM hr_data
ORDER BY age;


SELECT age_group,
count(*) AS count
FROM
(SELECT
    CASE
         WHEN age<=21 AND age<=30 THEN '21 to 30'
          WHEN age<=31 AND age<=40 THEN '31 to 40'
          WHEN age<=41 AND age<=50 THEN '41 to 50'
          ELSE '50+'
           END AS age_group
           FRom hr_data
           WHERE new_termdate IS NULL
           ) AS subquery
           GROUP BY age_group
           ORDER BY age_group;




--Age group by Gender
SELECT age_group,gender,count(*) AS count
FROM
   (SELECT
     CASE
         WHEN age<=21 AND age<=30 THEN '21 to 30'
         WHEN age<=31 AND age<=40 THEN '31 to 40'
         WHEN age <= 41 AND age<=50 THEN '41 to 50'
         ELSE '50+'
         END AS age_group,
         gender
         FROM hr_data
         WHERE new_termdate IS NULL
         ) AS subquery
         GROUP BY age_group,gender
         ORDER BY age_group,gender;

-- 2.What's the gender Breakdown in the company?
SELECT gender,count(gender) AS count
FROM hr_data
WHERE new_termdate IS NULL
GROUP BY gender
ORDER BY gender ASC;

--3. How does gender vary across departments and job titles?
SELECT department,gender,count(gender) AS count
FROM hr_data
WHERE new_termdate IS NULL
GROUP BY department,gender
ORDER BY department,gender ASC;

--job titles
 SELECT department,gender,jobtitle,count(gender) AS count
FROM hr_data
WHERE new_termdate IS NULL
GROUP BY department,gender,jobtitle
ORDER BY department,gender,jobtitle ASC;

--4.what's the race distributon in the company?
SELECT race,count(*) AS count
```

```
FROM hr_data
WHERE new_termdate IS NULL
GROUP BY race
ORDER BY count DESC;
```

--5.What's the average length of employement in the company ?

```
SELECT
AVG(DATEDIFF(YEAR,hire_date,new_termdate)) AS tenure
FROM hr_data
WHERE new_termdate IS NOT NULL AND new_termdate<= GETDATE();
```

--6.Which department has the highest turnover rate?
--get total count
--get terminated count
--terminated count/ total count

```
SELECT
  department,
  count(*) AS total_count,
  SUM(CASE
      WHEN new_termdate IS NOT NULL AND new_termdate<=GETDATE() THEN 1 ELSE 0
        END
        ) as terminated_count
        FROM hr_data
        GROUP BY department;

SELECT
department,
total_count,
terminated_count,
ROUND(CAST(terminated_count AS FLOAT)/total_count),2)*100 AS turnover_rate
FROM(
SELECT
  department,
  count(*) AS total_count,
  SUM(CASE
      WHEN new_termdate IS NOT NULL AND new_termdate<=GETDATE() THEN 1 ELSE 0
        END
        ) as terminated_count
        FROM hr_data
        GROUP BY department
        ) AS subquery
        ORDER BY turnover_rate DESC;
```

--7.What is the tenure distribution for each department?

```
SELECT
department,
AVG(DATEDIFF(YEAR,hire_date,new_termdate)) AS tenure
FROM hr_data
WHERE new_termdate IS NOT NULL AND new_termdate<= GETDATE()
GROUP BY department
ORDER BY tenure DESC;
```

--8.How many employees work remotely for each department?

```
 SELECT
  location,
  count(*) as count
  FROM hr_data
  WHERE new_termdate IS NULL
  GROUP BY location;
```

--9.What's the distribution of employess across different states?

```
SELECT
  location_state,
  count(*) AS count
  FROM hr_data
```

```
   WHERE new_termdate IS NULL
   GROUP BY location_state
   ORDER BY count DESC;
```

```
--10.How are job titles distributed in the company?
SELECT
jobtitle,count(*) AS count
FROM hr_data
WHERE new_termdate IS NULL
GROUP BY jobtitle
ORDER BY count DESC;
```

```
--11. How have employee hire counts Varied over time?
--calculate hires
--calculate terminations
--(hires-terminations)/hires percent hire Change

SELECT
hire_year,
hires,
terminations,
hires-terminations AS net_change,
ROUND(CAST(hires-terminations AS FLOAT)/hires,2)*100 AS percent_hire_change
FROM
(SELECT
YEAR(hire_date) AS hire_year,
count(*)
AS hires,
SUM(CASE
      WHEN new_termdate IS NOT NULL AND new_termdate<=GETDATE() THEN 1 ELSE 0
        END
        ) AS terminations
        FROM hr_data
        GROUP By YEAR (hire_date)
        ) AS subquery
        ORDER BY percent_hire_change ASC;
```