HR REPORT

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

```
--CREATE DATABASE

CREATE DATABASE hr_DATA;

--AFTER LOADING DB

USE hr_DATA;

--EXPLOARE THE LOADED DATA

SELECT *
FROM hr_data;
```

	id	first_name	last_name	birthdate	gender	race	department	jobtitle	location	hire_
1	00-0037846	Kimmy	Walczynski	1991-06-04	Male	Hispanic or Latino	Engineering	Programmer Analyst I	Headquarters	200
2	00-0041533	Ignatius	Springett	1984-06-29	Male	White	Business Development	Business Analyst	Headquarters	201
3	00-0045747	Corbie	Bittlestone	1989-07-29	Male	Black or African American	Sales	Solutions Engineer Manager	Headquarters	201
4	00-0055274	Baxy	Matton	1982-09-14	Female	White	Services	Service Tech	Headquarters	200
5	00-0076100	Terrell	Suff	1994-04-11	Female	Two or More Races	Product Management	Business Analyst	Remote	201
6	00-0116166	Kacie	Offiler	1971-01-18	Male	Asian	Engineering	Developer III	Headquarters	201
7	00-0363185	Sandro	Admans	1979-11-19	Male	Two or More Races	Product Management	Quality Engineer	Headquarters	201
8	00-0380704	Eugene	Lehrahan	1988-10-14	Female	Black or African American	Engineering	Developer I	Headquarters	200
9	00-0381660	Wainwright	Corfield	1996-12-13	Male	Asian	Engineering	Business Systems Development Analyst	Headquarters	200
10	00-0419202	Dyann	Isoldi	1980-03-27	Male	Two or More Races	Engineering	Web Developer I	Headquarters	200
11	00-0472287	Grantley	Oret	1975-09-06	Male	Two or More Races	Services	Service Tech II	Headquarters	200
12	00-0472832	Elmore	Worner	1966-01-07	Female	White	Engineering	Business Systems Development Analyst	Headquarters	200
13	00-0566380	Dud	Brain	1984-03-17	Male	Two or More Races	Business Development	Business Analyst	Headquarters	200
14	00-0571075	Aguie	Conford	1971-11-02	Male	White	Business Development	Research Assistant II	Headquarters	201
15	00-0624189	Katerina	Rosborough	1967-08-20	Female	Hispanic or Latino	Engineering	Analyst Programmer	Headquarters	201
16	00-0715313	Alida	Langley	1972-01-28	Female	American Indian or Alaska Native	Accounting	Staff Accountant III	Headquarters	200

		department	jobtitle	location	hire_date	termdate	location_city	location_state	new_termdate	age
1	or Latino	Engineering	Programmer Analyst I	Headquarters	2002-01-20	NULL	Cleveland	Ohio	NULL	33
2		Business Development	Business Analyst	Headquarters	2019-04-08	NULL	Cleveland	Ohio	NULL	40
3	African American	Sales	Solutions Engineer Manager	Headquarters	2010-10-12	NULL	Cleveland	Ohio	NULL	35
4		Services	Service Tech	Headquarters	2005-04-10	NULL	Cleveland	Ohio	NULL	42
5	lore Races	Product Management	Business Analyst	Remote	2010-09-29	2029-09-29	Flint	Michigan	2029-09-29	30
6		Engineering	Developer III	Headquarters	2018-09-01	NULL	Cleveland	Ohio	NULL	53
7	lore Races	Product Management	Quality Engineer	Headquarters	2012-11-08	NULL	Cleveland	Ohio	NULL	45
8	African American	Engineering	Developer I	Headquarters	2007-06-27	NULL	Cleveland	Ohio	NULL	36
9		Engineering	Business Systems Development Analyst	Headquarters	2001-02-20	2008-21-05	Cleveland	Ohio	NULL	28
10	lore Races	Engineering	Web Developer I	Headquarters	2005-01-27	NULL	Cleveland	Ohio	NULL	44
11	lore Races	Services	Service Tech II	Headquarters	2004-11-01	NULL	Cleveland	Ohio	NULL	49
12		Engineering	Business Systems Development Analyst	Headquarters	2000-12-05	NULL	Cleveland	Ohio	NULL	58
13	lore Races	Business Development	Business Analyst	Headquarters	2008-09-17	NULL	Cleveland	Ohio	NULL	40
14		Business Development	Research Assistant II	Headquarters	2015-11-25	NULL	Cleveland	Ohio	NULL	53
15	or Latino	Engineering	Analyst Programmer	Headquarters	2019-05-17	NULL	Cleveland	Ohio	NULL	57
16	Indian or Alaska Native	Accounting	Staff Accountant III	Headquarters	2002-02-04	NULL	Cleveland	Ohio	NULL	51

```
--EXPLORE THE TABLE STRUCTURE

SELECT COLUMN_NAME, DATA_TYPE, CHARACTER_MAXIMUM_LENGTH
FROM INFORMATION_SCHEMA.COLUMNS
WHERE TABLE_NAME = 'hr_data'; SELECT termdate
FROM hr_data
ORDER BY termdate DESC
```

	COLUMN_NAME	DATA_TYPE	CHARACTER_MAXIMUM_LENGTH
1	id	nvarchar	50
2	first_name	nvarchar	50
3	last_name	nvarchar	50
4	birthdate	date	NULL
5	gender	nvarchar	50
6	race	nvarchar	50
7	department	nvarchar	50
8	jobtitle	nvarchar	50
9	location	nvarchar	50
10	hire_date	date	NULL
11	termdate	nvarchar	50
12	location_city	nvarchar	50
13	location_state	nvarchar	50
14	new_termdate	date	NULL
15	age	nvarchar	50

```sql
--termdate formatting
--convert dates to yyyy-mm-dd
--create new column new_termdate
--copy converted time values from to new_termdate

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



ALTER TABLE hr_data
ADD new_termdate DATE;


-- copy converted time values from termdate to new-termdate

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


--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?
```

```
--age distribution

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

	youngest	oldest
1	22	59

```
--agr group by gender
SELECT age
FROM hr_data
ORDER BY age;
```

	age
1	22
2	22
3	22
4	22
5	22
6	22
7	22
8	22
9	22
10	22
11	22
12	22
13	22
14	22
15	22
16	22

```
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	count
1	31 to 40	5633
2	41 to 50	5928
3	50+	9823

```
--Age group by Gender
SELECT age_group,gender,count(*) AS count
```

```sql
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;
```

▦ Results ▤ Messages

	age_group	gender	count
1	31 to 40	Female	2617
2	31 to 40	Male	2875
3	31 to 40	Non-Conforming	141
4	41 to 50	Female	2716
5	41 to 50	Male	3045
6	41 to 50	Non-Conforming	167
7	50+	Female	4609
8	50+	Male	4944
9	50+	Non-Conforming	270

```sql
-- 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;
```

▦ Results ▤ Messages

	gender	count
1	Female	9942
2	Male	10864
3	Non-Conforming	578

```sql
--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;
```

	department	gender	count
1	Accounting	Female	1484
2	Accounting	Male	1652
3	Accounting	Non-Conforming	86
4	Auditing	Female	24
5	Auditing	Male	26
6	Business Development	Female	735
7	Business Development	Male	805
8	Business Development	Non-Conforming	47
9	Engineering	Female	2997
10	Engineering	Male	3253
11	Engineering	Non-Conforming	185
12	Human Resources	Female	829
13	Human Resources	Male	873
14	Human Resources	Non-Conforming	40
15	Legal	Female	136
16	Legal	Male	157
17	Legal	Non-Conforming	8

```sql
--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;
```

	department	gender	jobtitle	count
1	Accounting	Female	Accountant I	29
2	Accounting	Female	Accountant II	47
3	Accounting	Female	Accountant III	43
4	Accounting	Female	Accountant IV	38
5	Accounting	Female	Accounting Assistant I	42
6	Accounting	Female	Accounting Assistant II	31
7	Accounting	Female	Accounting Assistant III	33
8	Accounting	Female	Accounting Assistant IV	49
9	Accounting	Female	Actuary	13
10	Accounting	Female	Administrative Assistant I	4
11	Accounting	Female	Administrative Assistant II	4
12	Accounting	Female	Administrative Assistant III	6
13	Accounting	Female	Administrative Assistant IV	3
14	Accounting	Female	Administrative Officer	11
15	Accounting	Female	Budget/Accounting Analyst I	79
16	Accounting	Female	Budget/Accounting Analyst II	67
17	Accounting	Female	Budget/Accounting Analyst III	77

```sql
--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;
```

	race	count
1	White	6091
2	Two or More Races	3526
3	Black or African American	3478
4	Asian	3419
5	Hispanic or Latino	2416
6	American Indian or Alaska Native	1277
7	Native Hawaiian or Other Pacific Islander	1177

```sql
--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();
```

	tenure
1	8

```sql
--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;
```

Results | Messages

	department	total_count	terminated_count
1	Product Management	641	15
2	Auditing	52	2
3	Marketing	494	11
4	Legal	311	8
5	Sales	1832	52
6	Services	1686	49
7	Engineering	6686	174
8	Training	1692	55
9	Support	954	32
10	Human Resources	1807	42
11	Research and Development	1084	35
12	Accounting	3333	77
13	Business Development	1642	36

```sql
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?

```sql
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;
```

	department	tenure
1	Marketing	10
2	Legal	9
3	Support	9
4	Training	8
5	Sales	8
6	Services	8
7	Product Management	8
8	Auditing	8
9	Business Development	8
10	Engineering	8
11	Human Resources	8
12	Research and Development	7
13	Accounting	7

```sql
--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;
```

	location	count
1	Headquarters	16080
2	Remote	5304

```sql
 --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;
```

	location_state	count
1	Ohio	17330
2	Pennsylvania	1077
3	Illinois	838
4	Indiana	677
5	Michigan	654
6	Kentucky	433
7	Wisconsin	375

```sql
--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;
```

	jobtitle	count
1	Research Assistant II	732
2	Business Analyst	680
3	Human Resources Analyst II	589
4	Research Assistant I	512
5	Account Executive	482
6	Data Visualization Specialist	437
7	Staff Accountant I	429
8	Human Resources Analyst	393
9	Software Engineer I	378
10	Analyst Programmer	363
11	Systems Administrator I	362
12	Senior Developer	349
13	Data Coordinator	343
14	Software Consultant	326
15	Project Manager	326
16	Business Systems Development Analyst	325
17	Service Tech III	324

```sql
--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
```

```sql
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;
```

⊞ Results 📄 Messages

	hire_year	hires	terminations	net_change	percent_hire_change
1	2004	1135	41	1094	96
2	2001	1122	42	1080	96
3	2005	1097	49	1048	96
4	2003	1142	45	1097	96
5	2006	1118	46	1072	96
6	2012	1103	31	1072	97
7	2009	1140	39	1101	97
8	2000	220	6	214	97
9	2002	1067	36	1031	97
10	2008	1108	38	1070	97
11	2007	1090	33	1057	97
12	2016	1122	18	1104	98
13	2013	1105	22	1083	98
14	2010	1099	18	1081	98
15	2014	1053	18	1035	98
16	2011	1101	27	1074	98