SELECT * FROM Projects.HR

A-z id ▼	A-z first_name 🔻	A-z last_name 🔻	A-z birthdate 🔻	A-Z gender ▼	A-Z race ▼	A-Z department ▼	A-Z jobtitle	A-2
00-0037846	Kimmy	Walczynski	06-04-91	Male	Hispanic or Latino	Engineering	Programmer Analyst I	Нє
00-0041533	Ignatius	Springett	6/29/1984	Male	White	Business Development	Business Analyst	Нє
00-0045747	Corbie	Bittlestone	7/29/1989	Male	Black or African American	Sales	Solutions Engineer Manager	Hε
00-0055274	Baxy	Matton	9/14/1982	Female	White	Services	Service Tech	Нє
00-0076100	Terrell	Suff	04-11-94	Female	Two or More Races	Product Management	Business Analyst	Re
00-0116166	Kacie	Offiler	1/18/1971	Male	Asian	Engineering	Developer III	Нε
00-0363185	Sandro	Admans	11/19/1979	Male	Two or More Races	Product Management	Quality Engineer	Нє
00-0380704	Eugene	Lehrahan	10/14/1988	Female	Black or African American	Engineering	Developer I	Нє
00-0381660	Wainwright	Corfield	12/13/1996	Male	Asian	Engineering	Business Systems Development Analyst	Н€

USE PROJECTS

select birthdate, hiredate from hr

A-Z birthdate	•	A-z hire_date	•
06-04-91		1/20/2002	
6/29/1984		04-08-19	
7/29/1989		10-12-10	
9/14/1982		04-10-05	
04-11-94		9/29/2010	
1/18/1971		09-01-18	
11/19/1979		11-08-12	
10/14/1988		6/27/2007	
12/13/1996		2/20/2001	
3/27/1980		1/27/2005	

-- Standardize dates in the birthdate and hiredate column

```
UPDATE HR
SET birthdate = CASE
      WHEN birthdate LIKE '%/%' THEN DATE_FORMAT(STR_TO_DATE(birthdate, '%m/%d/%Y'),
      WHEN birthdate LIKE '%-%' THEN DATE_FORMAT(STR_TO_DATE(birthdate, '%m-%d-%Y'),
'%Y-%m-%d')
      ELSE NULL
END
SELECT hire_date FROM hr
ALTER TABLE PROJECTS.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'),
      WHEN hire_date LIKE '%-%' THEN DATE_FORMAT(STR_TO_DATE(hire_date , '%m-%d-%Y'),
'%Y-%m-%d')
      ELSE NULL
END
ALTER TABLE PROJECTS.HR MODIFY COLUMN hire_date DATE NULL;
select birthdate, hire_date from HR
```



-- Standardize termdate column

```
update hr
set termdate = date(str_to_date(termdate, '%Y-%m-%d %H:%i:%s UTC'))
where termdate is not null and termdate != ''

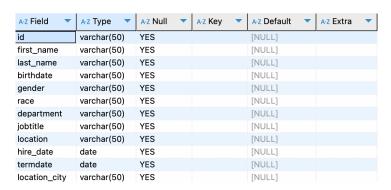
update hr
set termdate = null
where termdate = '';

select birthdate, termdate from hr
where termdate != null;
```



```
alter table projects.HR
modify column termdate date null;
alter table projects.HR
modify column hire_date date null;
```

describe hr



```
-- Add age column

ALTER TABLE HR

ADD COLUMN Age INT

UPDATE HR

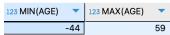
SET AGE = timestampdiff(YEAR, birthdate, CURDATE())

SELECT AGE, BIRTHDATE FROM HR
```



-- Cleaning ages and birthyears

SELECT MIN(AGE), MAX(AGE) FROM HR



SELECT COUNT (AGE) FROM HR WHERE AGE < 18



-- ANALYSIS

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

SELECT GENDER, COUNT(GENDER), (COUNT(GENDER))/ (SELECT COUNT(GENDER) FROM HR WHERE AGE >=18 AND TERMDATE IS NULL)*100) AS PROPORTION FROM HR

WHERE AGE>=18 AND TERMDATE IS NULL GROUP BY GENDER

A-Z GENDER ▼	123 COUNT(GENDER)	123 PROPORTION T
Male	8,911	50.9724
Female	8,090	46.2762
Non-Conforming	481	2.7514

-- What is the race/ethnicity breakdown of the employees?

SELECT race, COUNT(*) FROM projects.HR
where age > 18 and termdate is null
GROUP BY race
order by count(*) desc

A-z race	123 COUNT(*) ▼
White	4,987
Two or More Races	2,867
Black or African American	2,840
Asian	2,791
Hispanic or Latino	1,994
American Indian or Alaska Native	1,051
Native Hawaiian or Other Pacific Islander	952

— What is the age distribution of employees in the company SELECT

```
CASE
```

```
WHEN AGE >= 18 AND AGE <25 THEN '18-24'
WHEN AGE >= 25 AND AGE <35 THEN '25-34'
WHEN AGE >= 35 AND AGE <45 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(*)
```

FROM HR
WHERE AGE >= 18 AND TERMDATE IS NULL
GROUP BY AGE_GROUP, gender
ORDER BY AGE_GROUP

A-z age_group T	A-z gender ▼	123 count(*)
18-24	Female	635
18-24	Male	715
18-24	Non-Conforming	32
25-34	Female	2,297
25-34	Male	2,462
25-34	Non-Conforming	135
35-44	Female	2,271
35-44	Male	2,630
35-44	Non-Conforming	136
45-54	Female	2,046
45-54	Male	2,209

-- How many employees WFH vs at the office
select location, count(*) from hr

where age >=18 and termdate is NULL
GROUP BY LOCATION

A-z location	123 count(*)
Headquarters	13,107
Remote	4,375

-- What is the average length of employment for employees who have been terminated
SELECT avg(DATEDIFF(termdate, hire_date))/365 as avg_years FROM HR
where termdate is not null and termdate < year(CURDATE()) and age >= 18



— How does gender distribution vary across departments and job title SELECT DEPARTMENT, GENDER, COUNT(*) FROM HR WHERE AGE >=18 AND TERMDATE IS NULL GROUP BY DEPARTMENT, GENDER ORDER BY DEPARTMENT

A-Z DEPARTMENT ▼	A-Z GENDER ▼	123 COUNT(*)
Accounting	Female	1,175
Accounting	Male	1,375
Accounting	Non-Conforming	76
Auditing	Female	19
Auditing	Male	19
Business Development	Female	593
Business Development	Male	672
Business Development	Non-Conforming	42
Engineering	Female	2,442
Engineering	Male	2,671
Engineering	Non-Conforming	146

— What is the distribution of job titles across the company? SELECT JOBTITLE, COUNT(*) FROM HR WHERE AGE >=18 AND TERMDATE IS NULL GROUP BY JOBTITLE ORDER BY JOBTITLE DESC

A-Z JOBTITLE ▼	123 COUNT(*)
Web Developer IV	58
Web Developer III	53
Web Developer II	66
Web Developer I	79
Web Designer IV	5
Web Designer III	10
Web Designer II	3
Web Designer I	27
VP Sales	5
VP Quality Control	34
VP Product Management	31

A-Z department	123 total_count	123 terminated_count	123 termination_rate
Auditing	50	9	0.18
Legal	299	43	0.1438
Training	1,622	207	0.1276
Human Resources	1,727	211	0.1222
Research and Development	1,032	126	0.1221
Engineering	6,387	768	0.1202
Accounting	3,192	376	0.1178
Sales	1,745	205	0.1175
Support	903	106	0.1174
Services	1,618	185	0.1143
Product Management	623	71	0.114

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

```
select location_state, count(*)
from hr
where age >=18 and termdate is NULL
group by location_state
order by count(*) desc
```

```
A-Z location_state
                        123 count(*)
Ohio
                                  14,144
Pennsylvania
                                    892
Illinois
                                    698
Michigan
                                    550
Indiana
                                    545
Kentucky
                                    347
Wisconsin
                                    306
```

-- How has the company's employee count changed over time based on hire and term dates
select yearofoperation, hires, fired, hires-fired as net_change, round((hiresfired)/hires*100, 2) as net_change_percent
from (
select year(hire_date) as yearofoperation,
count(year(hire_date)) as hires,
SUM(CASE WHEN termdate is not null and termdate <= curdate() then 1 else 0 end) as
fired
from hr
where age >= 18
group by yearofoperation
order by yearofoperation
) as subquery

Subquei y				
123 yearofoperation 🔻	123 hires 🔻	123 fired	123 net_change	123 net_change_percent
2,000	211	26	185	87.68
2,001	1,082	197	885	81.79
2,002	1,012	162	850	83.99
2,003	1,088	198	890	81.8
2,004	1,087	200	887	81.6
2,005	1,038	188	850	81.89
2,006	1,069	182	887	82.97
2,007	1,058	152	906	85.63
2,008	1,061	145	916	86.33
2,009	1,094	153	941	86.01
2,010	1,050	135	915	87.14

-- What is the tenure distribution for each department?

SELECT department, **round**(**avg**(averagetenure), 0)

from(

SELECT department, DATEDIFF(termdate, hire_date)/365 as averagetenure
from hr

where termdate is null or termdate <=curdate() and age >= 18

) **as** subquery

GROUP BY department

ORDER BY avg(averagetenure) desc

A-Z department ▼	123 round(avg(averagetenure), 0)	•
Sales		9
Marketing		8
Engineering		8
Services		8
Accounting		8
Business Development		8
Research and Development		8
Auditing		8
Human Resources		8
Training		8
Support		7