```
--2. Find number of rows, unique names, and years

select
count(*) as num_rows,
count(distinct name) as num_names,
count(distinct years) as num_years
from names;
```

	num_rows	num_names	num_years
1	2052781	101338	142

```
--3. Are there any null values?

select
count(*) - count(name) as name_nulls,
count(*) - count(years) as years_nulls,
count(*) - count(gender) as gender_nulls,
count(*) - count(births) as births_nulls
from names;
```

	name_nulls	years_nulls	gender_nulls	births_nulls
1	0	0	0	0

```
--4. Display all births per year (regardless of gender)

select years, sum(births) as total_births
from names
group by years
order by years asc;
```

	years	total_births
1	1880	201484
2	1881	192690
3	1882	221533
4	1883	216944
5	1884	243461
6	1885	240854
7	1886	255317
8	1887	247393
9	1888	299473
10	1889	288946
11	1890	301401
12	1891	286671

```
--5. Display all births per year by gender

select years, [M] as male_births, [F] as female_births
from
(select births, gender, years from names) as tab1
pivot
(
sum(births)
for gender in ([M], [F])) as tab2
order by years asc;
```

	years	male_births	female_births
1	1880	110490	90994
2	1881	100737	91953
3	1882	113686	107847
4	1883	104625	112319
5	1884	114442	129019
6	1885	107799	133055
7	1886	110784	144533
8	1887	101411	145982
9	1888	120851	178622
10	1889	110580	178366
11	1890	111024	190377
12	1891	101191	185480
13	1892	122035	212339
14	1893	112319	212905
15	1894	115769	222922

```
--6. Find number of years which have more male births than female

select count([years]) as num_years
from
(select births, gender, [years] from names) as tab1
pivot
(
sum(births)
for gender in (M, F)) as tab2
where M > F;
```

	num_years
1	88

```
--7. Display births for all names in year range 2000 - 2003

select name, [2000], [2001], [2002], [2003]
from (select distinct name, births, years from names) as tab1
pivot
(
sum(births)
for years in ([2000], [2001], [2002], [2003])) as tab2
order by name asc;
```

	name	2000	2001	2002	2003
1	Aaban	NULL	NULL	NULL	NULL
2	Aabha	NULL	NULL	NULL	NULL
3	Aabid	NULL	NULL	NULL	5
4	Aabidah	NULL	NULL	NULL	NULL
5	Aabir	NULL	NULL	NULL	NULL
6	Aabriella	NULL	NULL	NULL	NULL
7	Aada	NULL	NULL	NULL	NULL
8	Aadam	6	NULL	11	10
9	Aadan	NULL	NULL	NULL	5
10	Aadarsh	5	7	10	10
11	Aadav	NULL	NULL	NULL	NULL
12	Aadaya	NULL	NULL	NULL	NULL
13	Aaden	NULL	7	13	20

```
--8. Find longest name(s) and its length

select name, len(name) as name_length
from names
where len(name) = (select max(len(name)) from names);
```

	name	name_length
1	Christophermich	15
2	Christopherjohn	15
3	Christopherjohn	15
4	Mariadelosangel	15
5	Mariadelrosario	15
6	Johnchristopher	15
7	Christophermich	15
8	Michaelchristop	15
9	Johnchristopher	15
10	Christopherjohn	15
11	Christophermich	15
12	Christopherryan	15

```
--9. Display the longest name for each year (regardless of gender)

select name, years from names as tab1
where len(name) >= (select max(len(name)) from names
where years=tab1.years) and not exists
(select name from names
where years=tab1.years and len(name) = len(tab1.name) and name < tab1.name)
order by years asc;
```

	name	years
1	Bartholomew	1880
2	Bartholomew	1881
3	Bartholomew	1882
4	Francisquita	1883
5	Bartholomew	1884
6	Bartholomew	1885

```
--10. Find unique number of names for each gender

select count(distinct(name)) as count_names, 'M' as gender from names
where gender = 'M'
union all
select count(distinct(name)), 'F' from names
where gender = 'F';
```

	count_names	gender
1	43093	M
2	69527	F

```
--11. Display most popular name by year for each gender (two rows per year)

select years, name, gender, births from names as tab1
where not exists (
select name from names as tab2
where tab2.years = tab1.years
and tab2.gender = tab1.gender
and tab2.births > tab1.births
)
order by years asc, births asc;
```

	years	name	gender	births
1	1880	Mary	F	7065
2	1880	John	M	9655
3	1881	Mary	F	6919
4	1881	John	M	8769
5	1882	Mary	F	8148
6	1882	John	M	9557
7	1883	Mary	F	8012
8	1883	John	M	8894
9	1884	Mary	F	9217
10	1884	John	M	9388

```
--12. Display unisex names (those who have both genders)

select name, count(distinct gender) as num_gender
from names
group by name
having count(distinct gender) = 2;
```

	name	num_gender
1	Meyers	2
2	Nautica	2
3	Simi	2
4	Marle	2
5	Dianna	2
6	Kunta	2
7	Tateum	2
8	Paisley	2
9	Charon	2
10	Eman	2

```
--13. Find 5 least popular female names for last decade (2010s)

select top (5) name, floor(years/10)*10 as decade
from names
where floor(years/10)*10 = 2010 and gender = 'F'
order by births asc;
```

	name	decade
1	Aadrika	2010
2	Lanaysha	2010
3	Aaditri	2010
4	Addelin	2010
5	Aahna	2010

```
--14. Display total births (regardless of gender) for each decade.

select floor(years/10)*10 as decade, sum(births) as total_births
from names
group by floor(years/10)*10
order by decade asc;
```

	decade	total_births
1	1880	2408095
2	1890	3362511
3	1900	4285149
4	1910	14831519
5	1920	22972286
6	1930	21228928
7	1940	29370324
8	1950	39448617
9	1960	37527745
10	1970	31975361
11	1980	35636435
12	1990	37480291
13	2000	38422442
14	2010	36249736
15	2020	6688794

```
--15. Display total births for each decade by gender.

select decade, [M] as male_births, [F] as female_births
from (select births, gender, floor(years/10)*10 as decade from names) as tab1
pivot
(
sum(births)
for gender in ([M], [F])) as tab2
order by decade asc;
```

	decade	male_births	female_births
1	1880	1095405	1312690
2	1890	1140951	2221560
3	1900	1357580	2927569
4	1910	6675237	8156282
5	1920	11019557	11952729
6	1930	10566090	10662838
7	1940	14885862	14484462
8	1950	20214902	19233715
9	1960	19263608	18264137
10	1970	16522910	15452451
11	1980	18461358	17175077
12	1990	19476869	18003422
13	2000	19952908	18469534
14	2010	18774716	17475020
15	2020	3452525	3236269

```
--16. Baby name researcher Laura Wattenberg pointed out on her website that the distribution
--of boy names by final letter has changed significantly over the last century. Display total
--births for male names ending in each letter throughout the 20th century (for example in
--years 1900, 1950, and 2000).

select tab1.last_letter, births_1900, births_1950, births_2000
from

(select distinct(right(name, 1)) as last_letter, sum(births) as births_1900
from names
where years = 1900 and gender = 'M'
group by right(name, 1)) as tab1
```

```
left join

(select distinct(right(name, 1)) as last_letter, sum(births) as births_1950
from names
where years = 1950 and gender = 'M'
group by right(name, 1)) as tab2
on tab1.last_letter=tab2.last_letter

left join

(select distinct(right(name, 1)) as last_letter, sum(births) as births_2000
from names
where years = 2000 and gender = 'M'
group by right(name, 1)) as tab3
on tab2.last_letter=tab3.last_letter

order by births_1900 desc;
```

	last_letter	births_1900	births_1950	births_2000
1	e	22724	168700	148967
2	n	20708	251851	614660
3	s	20049	241839	146040
4	d	15483	263732	64366
5	y	12742	253265	143373
6	m	10803	68866	41485
7	r	10286	61092	169611
8	t	10256	122985	50712
9	l	9308	172057	155472
10	h	6234	67954	85130

```
--17. Find number of female names ending with 'e' for last 30 years from today.

with tab1 as (
select *, right(name, 1) as letter_e
from names
where gender = 'F' and years >= year(getdate()) - 30 and right(name, 1) = 'e'
)
select years, count(letter_e) as ends_with_e
from tab1
group by years
order by years desc;
```

	years	ends_with_e
1	2021	2708
2	2020	2686
3	2019	2799
4	2018	2854
5	2017	2932
6	2016	2954
7	2015	3095
8	2014	3119

```
--18. Display name diversity over time (number of unique baby names per year).

select years, [M] AS male_names, [F] AS female_names
from
(select distinct name, gender, years from names) as ps
pivot
(count(name)
for gender in ([M], [F])) as pvt
order by years asc;
```

	years	male_names	female_names
1	1880	1058	942
2	1881	996	938
3	1882	1099	1028
4	1883	1030	1054
5	1884	1125	1172
6	1885	1097	1197
7	1886	1110	1282
8	1887	1067	1306

```
--19. Find number of given names by first letter regardless od sex and year.

select distinct(left(name, 1)) as last_letter, sum(births) as total_births
from names
group by left(name, 1)
order by total_births desc;
```

	last_letter	total_births
1	J	45798659
2	M	33979277
3	A	30722526
4	C	26435262
5	D	24765319
6	R	24330850
7	S	22133847

```
--20. Find 15 most popular male names of all time.

select top (15) name, sum(births) as total_births
from names
where gender = 'M'
group by name
order by total_births desc;
```

	name	total_births
1	James	5202714
2	John	5150510
3	Robert	4834094
4	Michael	4392696
5	William	4156142
6	David	3646903
7	Joseph	2639396
8	Richard	2571082
9	Charles	2411608
10	Thomas	2331794
11	Christopher	2048886
12	Daniel	1948350
13	Matthew	1625606
14	George	1476135
15	Anthony	1462540

```
--21. Some names changed their gender over years (for example: Jean, Donnie, Leslie, Lauren).
--Display an example throughout time.

select years, name, [M] - [F] as gender_diff
from names
pivot
(sum(births)
for gender in ([M], [F])) as pvt
where name = 'Jean'
order by years asc;
```

	years	name	gender_diff
1	1880	Jean	-47
2	1881	Jean	-35
3	1882	Jean	-48
4	1883	Jean	-64
5	1884	Jean	-78
6	1885	Jean	-82
7	1886	Jean	-113
8	1887	Jean	-91

```
--22. Find overall percentage of each gender.

select
count(gender) * 100/(select count(*) from names) as male_percent,
100 - count(gender) * 100/(select count(*) from names) as female_percent
from names
where gender = 'M';
```

	male_percent	female_percent
1	41	59

```
--23. Find years with advantage of female births upon male births. This can be described as
--the number of males per 100 females (gender ratio).

select years, [M]*100 / [F] as gender_ratio
from
(select births, gender, years from names) as tab1
pivot
(
sum(births)
for gender in ([M], [F])) as tab2
where [M] * 100 / [F] < 100
order by years asc;
```

	years	gender_ratio
1	1883	93
2	1884	88
3	1885	81
4	1886	76
5	1887	69
6	1888	67
7	1889	61
8	1890	58
9	1891	54
10	1892	57
11	1893	52