MLB-Player-Analytics-Sean-Lahman-Baseball-DB

In each decade, how many schools were there that produced players?

```
with years as (select round(yearID,-1) as new_year,count( DISTINCT schoolID) as count_of_schools
               from schools
               group by 1
               order by new_year asc)
select * from years;
```

Results:

new_year	count_of_scho...
1860	1
1870	8
1880	18
1890	61
1900	119
1910	176
1920	193
1930	178
1940	159
1950	147
1960	230
1970	366
1980	441
1990	500
2000	467
2010	196

What are the names of the top 5 schools that produced the most players?

```
with school_details as (select s.playerID,s.schoolID,sd.name_full as school_name
                        from schools as s
                        left join school_details as sd
                        on s.schoolID = sd.schoolID
                        )
select school_name,count(distinct playerID) as count_of_players
from school_details
group by school_name
order by count_of_players desc
limit 5;
```

Results:

school_name	count_of_play...
University of Texas at Austin	107
University of Southern California	105
Arizona State University	101
Stanford University	86
University of Michigan	76

For each decade, what were the names of the top 3 schools that produced the most players?

```
with years as (select round(yearID,-1) as new_year,
                s.schoolID as s_id,
                sd.name_full,
                s.playerID
                from schools as s
                left join school_details as sd
                on s.schoolID = sd.schoolID
                order by yearID asc),
player_cnt as (select new_year,name_full,count(distinct playerID) total_players
                from years
                group by 1,2
                ),
college_wise as(select new_year,name_full,sum(total_players)over(PARTITION BY new_year,name_full )as total_cnt
                    from player_cnt
                    order by new_year),

rank_wise as(select new_year,name_full,total_cnt,dense_rank()over(PARTITION BY new_year order by total_cnt desc) as rnk
from college_wise
order by new_year)

select *
from rank_wise
where rnk<=3;
```

Results:

new_year	name_full	total_cnt	rnk
1860	Fordham University	1	1
1870	Fordham University	2	1
1870	Dartmouth College	1	2
1870	Harvard University	1	2
1870	Seton Hall University	1	2
1870	University of Michigan	1	2
1870	Villanova University	1	2
1870	Washington and Lee University	1	2
1870	Yale University	1	2
1880	Brown University	6	1
1880	Yale University	6	1
1880	University of Michigan	3	2
1880	Case Western Reserve Unive...	2	3
1880	Cornell University	2	3
1880	Pennsylvania State University	2	3
1880	Princeton University	2	3
1880	St. Mary's College of California	2	3
1880	Union College	2	3
1880	University of Rochester	2	3
1890	College of the Holy Cross	11	1
1890	University of Pennsylvania	9	2

For each team, show the cumulative sum of spending over the years

```
with team_year as(select teamID,yearID,sum(salary) as total_spending
                  from salaries
                  group by 1,2
                  order by 1,2)
select teamID,yearID,total_spending,sum(total_spending)over(order by teamID,yearID) as cum_sum
from team_year;
```

Results:

teamID	yearID	total_spendi...	cum_sum
ANA	1997	31135472	31135472
ANA	1998	41281000	72416472
ANA	1999	55388166	127804638
ANA	2000	51464167	179268805
ANA	2001	47535167	226803972
ANA	2002	61721667	288525639
ANA	2003	79031667	367557306
ANA	2004	100534667	468091973
ARI	1998	32347000	500438973
ARI	1999	68703999	569142972
ARI	2000	81027833	650170805
ARI	2001	85082999	735253804
ARI	2002	102819999	838073803
ARI	2003	80657000	918730803
ARI	2004	69780750	988511553
ARI	2005	62329166	10508407...
ARI	2006	596842226	1110524945
ARI	2007	52067546	1162592491
ARI	2008	66202712	12287952...
ARI	2009	73115666	13019108...
ARI	2010	60718166	13626290...

Return the first year that each team's cumulative spending surpassed 1 billion

```sql
with team_year as(select teamID,yearID,sum(salary) as total_spending
                    from salaries
                    group by 1,2
                    order by 1,2),
sum_sum as      (select teamID,yearID,total_spending,
                sum(total_spending)over(order by teamID,yearID) as cum_sum
                from team_year),
rnk_wise as (select teamID,yearID,row_number()over(partition by teamID order by yearID) as rn
from sum_sum
where cum_sum>=1000000000)

select teamID,yearID from rnk_wise
where rn=1;
```

Results:

teamID	yearID
ARI	2005
ATL	1985
BAL	1985
BOS	1985
CAL	1985
CHA	1985
CHN	1985
CIN	1985
CLE	1985
COL	1993
DET	1985
FLO	1993
HOU	1985
KCA	1985
LAA	2005
LAN	1985
MIA	2012
MIL	1998
MIN	1985
ML4	1985
MON	1985
NYA	1985
NYM	2014
NYN	1985
OAK	1985
PHI	1985
PIT	1985
SDN	1985

For each player, calculate their age at their first game, their last game, and their career length (all in years). Sort from longest career to shortest career.

```
with playerdetails as (select playerID,nameGiven,
                        cast(concat(birthYear,"-",birthMonth,"-",birthDay) as DATE) as date_birth,
                        debut,finalGame
                       from players ),
carrer as (select playerID,nameGiven,
        TIMESTAMPDIFF(year,date_birth,debut) as first_game_age,
        TIMESTAMPDIFF(year,date_birth,finalGame) as final_game_age
from playerdetails)

select playerID,nameGiven,
        first_game_age,
        final_game_age,
        (final_game_age-first_game_age) as len_carrer
from carrer
order by len_carrer desc;
```

Results:

playerID	nameGiven	first_game_a...	final_game_age	len_carrer
altroni01	Nicholas	21	57	36
orourji01	James Henry	21	54	33
minosmi01	Saturnino Orestes Armas	23	54	31
olearch01	Charles Timothy	28	58	30
lathaar01	Walter Arlington	20	49	29
mcguide01	James Thomas	20	48	28
jennihu01	Hugh Ambrose	22	49	27
eversjo01	John Joseph	21	48	27
streega01	Charles Evard	21	48	27
ryanno01	Lynn Nolan	19	46	27
moyerja01	Jamie	23	49	26
francju01	Julio Cesar	23	49	26
ansonca01	Adrian Constantine	19	45	26
johnto01	Thomas Edward	20	46	26
broutda01	Dennis Joseph	21	46	25
oroscje01	Jesse Russell	21	46	25
quinnja01	John Picus	25	50	25
morgami01	Michael Thomas	18	42	24
newsobo01	Louis Norman	22	46	24
oconnja01	John Joseph	20	44	24

What team did each player play on for their starting and ending years?

```
with player_team as (select p.playerID,p.nameGiven,
                        p.debut as starting_year,p.finalGame as ending_year,s.yearID as start_year_id,
                        s.teamID as starting_team,e.yearID as end_year_id,e.teamID as ending_team
                    from           players as p
                           inner JOIN salaries as s
                           on p.playerID = s.playerID
                           and year(p.debut) = s.yearId
                           inner join salaries as e
                           on p.playerID=e.playerID and year(p.finalGame)=e.yearID)

select playerID,nameGiven,starting_year,starting_team,ending_year,ending_team
from player_team;
```

Results:

playerID	nameGiven	starting_ye...	starting_te...	ending_year	ending_team
mccaski01	Kirk Edward	1985-05-01	CAL	1996-07-20	CHA
duncama01	Mariano	1985-04-09	LAN	1997-09-17	NYA
higuete01	Teodoro Valenzuela	1985-04-23	ML4	1994-08-09	ML4
mcdowro01	Roger Alan	1985-04-11	NYN	1996-08-14	BAL
birtsti01	Timothy Dean	1985-05-03	OAK	1990-10-03	CIN
colemvi01	Vincent Maurice	1985-04-18	SLN	1997-04-14	DET
assenpa01	Paul Andre	1986-04-12	ATL	1999-10-03	CLE
olwined01	Edward R.	1986-06-02	ATL	1988-10-02	ATL
speckcl01	Robert Clifford	1986-07-30	ATL	1986-09-25	ATL
finlech01	Charles Edward	1986-05-29	CAL	2002-09-28	CLE
joynewa01	Wallace Keith	1986-04-08	CAL	2001-06-14	ANA
bonilbo01	Roberto Martin An...	1986-04-09	CHA	2001-10-07	SLN
karkoro01	Ronald Joseph	1986-08-17	CHA	1997-09-26	CHA
mckeojo01	Joel Jacob	1986-05-06	CHA	1987-07-10	CHA
thigpbo01	Robert Thomas	1986-08-06	CHA	1994-04-27	SEA
martimi01	Joseph Michael	1986-08-15	CHN	1986-09-27	CHN
moyerja01	Jamie	1986-06-16	CHN	2012-05-27	COL
danieka01	Kalvoski	1986-04-09	CIN	1992-09-22	LAN
jonestr01	Tracy Donald	1986-04-07	CIN	1991-10-06	SEA
larkiba01	Barry Louis	1986-08-13	CIN	2004-10-03	CIN

How many players started and ended on the same team and also played for over a decade?

```
with player_team as (select p.playerID,p.nameGiven,
                        p.debut as starting_year,p.finalGame as ending_year,s.yearID as start_year_id,
                        s.teamID as starting_team,e.yearID as end_year_id,e.teamID as ending_team
                    from           players as p
                           inner JOIN salaries as s
                           on p.playerID = s.playerID
                           and year(p.debut) = s.yearId
                           inner join salaries as e
                           on p.playerID=e.playerID and year(p.finalGame)=e.yearID),

year_cal as (select playerID,nameGiven,starting_year,starting_team,ending_year,ending_team,
                    TIMESTAMPDIFF(year,starting_year,ending_year)as total_year_played
                from player_team
                where starting_team=ending_team)
SELECT playerID,nameGiven,total_year_played
FROM year_cal
where total_year_played>=10
order by total_year_played desc;
```

Results:

playerID	nameGiven	total_year_play...
griffke02	George Kenneth	21
glavito02	Thomas Michael	20
jonesch06	Larry Wayne	19
pettian01	Andrew Eugene	18
larkiba01	Barry Louis	18
sosasa01	Samuel Peralta	18
riverma01	Mariano	18
burksel01	Ellis Rena	17
heltoto01	Todd Lynn	16
willibe02	Bernabe	15
woodke02	Kerry Lee	14
lankfra01	Raymond Lewis	14
aurilri01	Richard Santo	14
hollida01	David Michael	12
hentgpa01	Patrick George	12
karkoro01	Ronald Joseph	11
radkebr01	Brad William	11
pagnoto01	Thomas Alan	11
utleych01	Chase Cameron	11
perezed02	Eduardo Rafael	10
higgibo02	Robert Leigh	10
pulidca01	Juan Carlos	10
dreifda01	Darren James	10
taubeed01	Edward Kenneth	10
thompro01	Robert Randall	10
mauerjo01	Joseph Patrick	10

Which players have the same birthday?

```
with player_details as (select playerID,nameGiven,
                        cast(concat(birthYear,"-",birthMonth,"-",birthDay) as DATE) as date_birth
                        from players)
select date_birth,GROUP_CONCAT(nameGiven SEPARATOR ", ") as total_people,count(nameGiven) as cnt
FROM player_details
where date_birth is not null and year(date_birth) between 1980 and 1990
group by 1
having cnt>1
order by 1;
```

Results:

date_birth	total_people	cnt
1980-01-15	Jeffrey Darrin, Matthew Thomas	2
1980-01-16	Brooks Litchfield, Jose Alberto	2
1980-01-17	Thomas Joseph, Michael Gregory	2
1980-01-20	Franklyn Miguel, Luis	2
1980-01-26	Brandon Edward, Antonio Miguel	2
1980-02-04	Stephen John, Douglas	2
1980-03-11	Christopher Allen, Richard Joseph, Daniel Cooley	3
1980-04-25	Michael Gregory, Kazuhito	2
1980-05-18	Juan Ramon, Luis Enrique	2
1980-05-22	Ruddy Joraider, Chad Austin	2
1980-07-02	Nyjer Jamid, Jermaine Russell	2
1980-07-15	Reginald Damascus, Jung Keun, Christopher A…	4
1980-07-21	Kyuji, Carsten Charles	2
1980-07-25	Santiago, Shawn Willis	2
1980-08-08	Craig Andrew, Joseph Buren	2
1980-08-16	Ryan Michael, Benjamin Anthony	2
1980-08-17	Brett Allen, Michael Patrick, Jeffrey Allen, Christ…	4
1980-08-23	Marcus Andre, Patrick Martin	2
1980-08-28	Theodore Lester, Ryan Michael	2
1980-08-30	Russ Moore, Roberto	2
1980-09-12	Sean Patrick, Maicer, Kevin Grant	3
1980-09-19	Ryan Alan, Raymond Lee	2
1980-09-28	Christopher Neil, Francisco Alberto	2
1980-09-29	Miguel, Amos Dewon	2
1980-10-19	Jose Antonio, Rajai Lavae	2

Create a summary table that shows for each team, what percent of players bat right, left and both

```
WITH team_details as (select s.teamid,p.playerID,p.bats
                      from salaries as s
                      join players as p
                      on s.playerID = p.playerID)
select teamID,
     round( sum( case when bats="R" then 1 else 0  end)/count(playerID)*100 ,1)as righty,
       round(sum(case when bats="L" then 1 else 0 end)/count(playerID)*100,1)as lefty,
      round( sum(case when bats ="B" THEN 1 ELSE 0 END)/count(playerID)*100,1) AS both_b_l
from team_details
group by 1;
```

Results:

teamid	righty	lefty	both_b_l
ATL	61.8	29.2	8.9
BAL	61.8	29.6	8.6
BOS	62.0	29.4	8.6
CAL	60.6	29.4	10.1
CHA	59.7	33.5	6.8
CHN	63.8	28.5	7.7
CIN	62.6	29.4	8.0
CLE	59.6	29.7	10.8
DET	60.8	28.6	10.6
HOU	62.3	23.9	13.8
KCA	64.3	27.2	8.5
LAN	63.0	27.8	9.2
MIN	60.9	26.7	12.4
ML4	59.6	29.4	11.0
MON	63.8	24.1	12.1
NYA	58.8	30.7	10.5
NYN	56.1	30.2	13.7
OAK	62.7	27.5	9.9
PHI	58.5	31.4	10.2
PIT	64.4	27.4	8.2
SDN	61.5	28.9	9.6
SEA	61.7	28.9	9.4
SFN	61.1	27.5	11.3
SLN	61.9	26.5	11.6
TEX	63.6	26.4	9.9
TOR	64.0	26.6	9.4
COL	63.7	27.8	8.5
FLO	66.3	24.3	9.4
ANA	61.1	31.6	7.3

How have average height and weight at debut game changed over the years, and what's the decade-over-decade difference?

```
with decade_wise as(select round(year(debut),-1) AS decade,avg(weight) as avg_weight,avg(height) as avg_height
                    from players
                    group by 1
                    order by 1),
over_cal as (select decade,avg_weight,
                    lag(avg_weight)over(order by decade) as next_decade_weight,
                    avg_height,
                    lag(avg_height)over(order by decade) as next_decade_height
                    from decade_wise)

select decade,(avg_weight-next_decade_weight) as weight_diff,
       (avg_height-next_decade_height) as height_diff
from over_cal
where decade is not null;
```

Results:

decade	weight_diff	height_diff
1870	-23.3981	-2.5384
1880	9.3491	0.8704
1890	0.9088	0.4080
1900	2.5512	0.3518
1910	1.2430	0.5936
1920	-1.0963	0.0611
1930	3.6059	0.3865
1940	5.2398	0.6969
1950	1.9556	0.3154
1960	2.3327	0.4755
1970	0.7502	0.2414
1980	1.2865	0.3134
1990	2.3135	0.1728
2000	9.5709	0.1449
2010	8.9710	0.0763