Joining and Grouping Data  
50 Points

Answer the questions and provide the SELECT in red below the question. Turn in this document unzipped.

1. Perform an inner join on Players and Managers on PlayerID. Include all fields in Players and all fields in Managers but PlayerID on PlayerID. How many records are there? Look over the data. Change the join to a left join. Look over the data results. How many records are there? What is the difference? Change the join back to an inner join. (3)

Inner Join: 3,404

Left Join: 21,555, includes all records in player columns regardless of values.

SELECT Players.birthDate, Players.birthCountry, Players.birthState, Players.birthCity, Players.deathDate, Players.deathCountry, Players.deathState, Players.deathCity, Players.nameFirst, Players.nameLast,

Players.nameGiven, Players.weight, Players.height, Players.bats, Players.throws, Players.debutDate, Players.finalGameDate, Managers.yearID, Managers.teamID, Managers.lgID, Managers.inseason,

Managers.G, Managers.W, Managers.L, Managers.rank, Managers.plyrMgr, Managers.bPlyrMgr, Managers.playerID, Managers.ManagersIDpk

FROM Players INNER JOIN

Managers ON Players.playerIDpk = Managers.playerID

1. Save the view as vwPlayerManagers. With an inner join we will get only parents (players) with children (managers) – in other words, only managers. Text the view (2)

select \* from vwPlayerManagers

1. Develop a SELECT to show each manager’s name, team and the number of games they managed that team overall. (5)

select nameGiven, teamID, sum(G) as 'Games Managed' from vwPlayerManagers

group by nameGiven, teamID

order by [Games Managed]

1. Provide the manager name, total wins and total losses for only managers with an overall winning record (total wins > total losses) (4)

select nameGiven, sum(W) as 'Total Wins', sum(L) as 'Total Losses' from vwPlayerManagers

group by nameGiven

having sum(W) > sum(L)

1. In a single SELECT, which manager has the greatest number of wins? (4)

You looked at it on Friday but couldn’t get it to work as a Select \* from (select \*from) type query

select Top 1 nameGiven, sum(W) as 'Wins' from vwPlayerManagers

group by nameGiven

order by Wins desc

1. Write a SELECT to compare Wins of players that were also manager versus those that were never managers (you can use either the plyMgr or bPlyMger fields). Does player experience seem to matter once you become a manager? (5)

It would appear that player managers won less games on average

select Avg(W)\*1.0/(select Avg(W) as 'NonMGRWins' from vwPlayerManagers where plyrMgr = 'N' ) as 'PlyrMgr Wins +-' from vwPlayerManagers

where plyrMgr = 'Y'

1. Who managed a single team the most seasons? Provide a single SELECT to demonstrate. (3)

Cornelius Alexander at 50

select top 1 nameGiven, count(teamID) as 'Times managed' from vwPlayerManagers

group by nameGiven, TeamID

order by [Times managed] desc

1. Who managed overall the most? Provide a single SELECT to demonstrate. (3)

You didn’t clarify if you meant most games or most seasons/years

For total games is was Cornelius Alexander (Connie Mack)

select top 1 nameGiven, sum(G) as 'Total managed games' from vwPlayerManagers

group by nameGiven

order by [Total managed games] desc

1. Which team had the most managers? (3)

CHN

Select top 1 teamID, count(distinct playerID) as 'Unique Managers' from vwPlayerManagers

group by teamID

order by [Unique Managers] desc

1. Who was the shortest manager? (3)

John A.

Select top 1 nameGiven, min(height) as 'Shortest Manager' from vwPlayerManagers

where height is not null

group by nameGiven

order by [Shortest Manager]

1. How many managers were foreign born? (3)

46

select count(distinct playerID) as '# of Foreign Managers' from vwPlayerManagers

where birthCountry != 'USA'

1. Inner Join Players to Batting. Create a view vwPlayerBatting. (3)

create view [vwPlayerBatting] as

SELECT Players.playerIDpk, Players.birthDate, Players.birthCountry, Players.birthState, Players.birthCity, Players.deathDate, Players.deathCountry, Players.deathState, Players.deathCity, Players.nameFirst,

Players.nameLast, Players.nameGiven, Players.weight, Players.height, Players.bats, Players.throws, Players.debutDate, Players.finalGameDate, Batting.yearID, Batting.stint, Batting.lgID, Batting.teamID,

Batting.G, Batting.G\_batting, Batting.AB, Batting.R, Batting.H, Batting.B2, Batting.B3, Batting.HR, Batting.RBI, Batting.SB, Batting.CS, Batting.BB, Batting.SO, Batting.IBB, Batting.HBP, Batting.SH, Batting.SF,

Batting.GIDP

FROM Players INNER JOIN

Batting ON Players.playerIDpk = Batting.playerID

1. Who hit the most home runs in history (single SELECT statement) (3)

It says Henry Louis, but my aggregate function cannot be functioning correctly, I don’t know what I’m doing here. It says he had 1248 homeruns. That’s not correct. I looked up all the rows belonging to him, and added them up manually to double check, and it adds up. So something in the data must be off, or my table is just screwed.

select top 1 nameGiven, sum(HR) as "Homeruns" from vwPlayerBatting

group by nameGiven

order by Homeruns desc

1. Provide a SELECT listing the most consistent home run hitters (average) that played (having) in at least 20 games in a season. (3)

select nameGiven, G as 'Games', avg(HR) as "Home Run Average" from vwPlayerBatting

group by nameGiven, G

having G > 20

order by [Home Run Average] desc

1. Who had the most triples (B3) in history (single SELECT statement). (3)

John Joseph

select top 1 nameGiven, sum(B3) as "Triples" from vwPlayerBatting

group by nameGiven

order by [Triples] desc