Jayson Lima – Final Project SQL Queries and Excel Navigation

Data used: opensourcesport.com Men’s Basketball Database

Downloaded from: <http://opensourcesports.com/files/basketball/BasketballDB-20130121.zip>

I created a local database in SQL and called it ‘basketball’.

I took relevant Tables from the Men’s Basketball Database download and imported them into my SQL local database.

**Created ‘draft’ Table from basketball\_draft:**

CREATE TABLE draft(draftYear int, draftRound int, draftSelection int, draftOverall int, tmID text, firstName text, lastName text, suffixName text, playerID text, draftFrom text, lgID text);

**Created ‘players’ Table from basketball\_players:**

CREATE TABLE players(playerID text, year text, stint text, tmID text, lgID text, GP text, GS text, minutes text, points text, oRebound text, dRebound text, rebounds text, assists text, steals text, blocks text, turnovers text, PF text, fgAttempted text, fgMade text, ftAttempted text, ftMade text, threeAttempted text, threeMade text, postGP text, postGS text, postMinutes text, postPoints text, postoRebounds text, postdRebounds text, postRebounds text, postAssists text, postSteals text, postBlocks text, postTurnovers text, postPF text, postfgAttempted text, postfgMade text, postftAttempted text, postftMade text, postthreeAttempted text, postthreeMade text, note text);

**Created ‘awards’ Table from basketball\_awards\_players:**

CREATE TABLE awards(playerID text, award text, year text, lgID text, note text, pos text);

**Created ‘halloffame’ Table from basketball\_HoF:**

CREATE TABLE halloffame(year text, hofID text, name text, category text);

**Created ‘allstar’ from basketball\_player\_allstar:**

CREATE TABLE allstar(playerID text, lastName text, firstName text, seasonID text, conference text, leagueID text, gamesPlayed text, minutes text, points text, oRebounds text, dRebounds text, rebounds text, assists text, steals text, blocks text, turnovers text, personalFouls text, fgAttempted text, fgMade text, ftAttempted text, ftMade text, threeAttempted text, threeMade text);

**Altered ‘seasonid’ column name to ‘allstarseason’ in ‘allstar’ table:**

ALTER TABLE allstar RENAME COLUMN seasonid TO allstarseason

**Created ‘posteason’ from basketball\_series\_post:**

CREATE TABLE postseason(year text, round text, series text, tmIDwinner text, lgIDwinner text, tmIDloser text, lgIDloser text, w text, l text)

**Created ‘master’ from basketball\_master:**

CREATE TABLE master(bioID text, useFirst text, firstName text, middleName text, lastName text, nameGiven text, fullgivenName text, nameSuffix text, nameNick text, pos text, firstSeason text, lastSeason text, height text, weight text, college text, collegeOther text, birthDate text, birthCity text, birthState text, birthCountry text, highSchool text, hsCity text, hsState text, hsCountry text, deathDate text, race text);

**Brought Basketball Database into SQL:**

I then used the ‘Import/Export’ Tool to bring in the .csv files into the database. I selected the corresponding files, set to ‘Import’, clicked ‘Yes’ for Header, set the Delimiter to comma(,), then hit OK.

**THESE ARE THE SQL SYNTAX QUERIES USED TO CREATE TABLES FOR EXCEL:**

**Quick look at allstar data where year if 1985 or greater ordered in ascending order:**

SELECT \* FROM allstar WHERE CAST(seasonid AS int) >= 1985 ORDER BY seasonid ASC LIMIT 100

**COUNT OF NUMBER OF DISTINCT Player IDs IN Players:**

SELECT COUNT(DISTINCT playerid) FROM players

**Count of Distinct player IDs in Players who played in 1985 or greater:**

SELECT COUNT(DISTINCT playerid) FROM players WHERE CAST(year AS int) >= 1985

**Count of all Draft Picks since 1985:**

SELECT COUNT(DISTINCT playerid) FROM draft WHERE CAST(draftyear as int) >= 1985

**Merging draft table with master table:**

SELECT draft.draftyear, draft.draftround, draft.draftselection, draft.tmid, draft.firstname, draft.lastname, draft.playerid, master.pos, master.firstseason, master.lastseason, master.height, master.weight, master.college, master.birthdate, master.birthcity, master.birthstate, master.birthcountry, master.highschool, master.hscity, master.hsstate, master.hscountry FROM draft LEFT JOIN master ON draft.playerid = master.bioid WHERE CAST(draft.draftyear AS int) >= 1985

**Selecting players who played at least 70% (58) games since 1985:**

SELECT playerid, year, tmid, gp, minutes, points, orebound, drebound, rebounds, assists, steals, blocks, turnovers, pf, fgattempted, fgmade, ftattempted, ftmade, threeattempted, threemade FROM players WHERE CAST(year AS int) >= 1985 AND CAST(gp AS int) >= 58

**Created One Big Master Table by Combining 4 Tables:**

SELECT players.playerid, players.year, players.tmid, players.gp, players.minutes, players.points, players.orebound, players.drebound, players.rebounds, players.assists, players.steals, players.blocks, players.turnovers, players.pf, players.fgattempted, players.fgmade, players.ftattempted, players.ftmade, players.threeattempted, players.threemade, draft.draftround, draft.draftselection, draft.draftfrom, awards.award, awards.year, master.firstname, master.lastname, master.pos, master.height, master.weight, master.college, master.birthcity, master.birthstate, master.birthcountry, master.highschool, master.hscity, master.hsstate, master.hscountry, allstar.allstarseason

FROM players LEFT JOIN draft ON players.playerid = draft.playerid LEFT JOIN awards ON players.playerid = awards.playerid LEFT JOIN master ON players.playerid = master.bioid LEFT JOIN allstar ON players.playerid = allstar.playerid

WHERE CAST(players.year AS int) >= 1985 AND CAST(players.gp AS int) >= 58 AND CAST(awards.year AS int) >=1985 AND CAST(players.year AS int) = CAST(allstar.allstarseason AS int)

**All-NBA Stats Table since 1985:**

SELECT players.playerid, players.year, players.tmid, players.gp, players.minutes, players.points, players.orebound, players.drebound, players.rebounds, players.assists, players.steals, players.blocks, players.turnovers, players.pf, players.fgattempted, players.fgmade, players.ftattempted, players.ftmade, players.threeattempted, players.threemade, draft.draftround, draft.draftselection, draft.draftfrom, awards.award, awards.year, master.firstname, master.lastname, master.pos, master.height, master.weight FROM players LEFT JOIN draft ON players.playerid = draft.playerid LEFT JOIN awards ON players.playerid = awards.playerid LEFT JOIN master ON players.playerid = master.bioid WHERE CAST(players.year AS int) >= 1985 AND CAST(awards.year AS int) >=1985 AND awards.year = players.year AND awards.award LIKE 'All%';

**Table with allstars with 4 tables:**

SELECT players.playerid, players.year, allstar.allstarseason, players.tmid, players.gp, players.minutes, players.points, players.orebound, players.drebound, players.rebounds, players.assists, players.steals, players.blocks, players.turnovers, players.pf, players.fgattempted, players.fgmade, players.ftattempted, players.ftmade, players.threeattempted, players.threemade,

draft.draftround, draft.draftselection, draft.draftfrom, master.firstname, master.lastname, master.pos, master.height, master.weight

FROM players LEFT JOIN draft ON players.playerid = draft.playerid LEFT JOIN allstar ON players.playerid = allstar.playerid LEFT JOIN master ON players.playerid = master.bioid

WHERE CAST(players.year AS int) >= 1985 AND CAST(allstar.allstarseason AS int) >=1985 AND allstar.allstarseason = players.year ORDER BY players.year ASC;

**Bringing in Master table:**

SELECT bioid, firstname, lastname, pos, height, weight, college, birthdate, birthcity, birthstate, birthcountry, highschool, hscity, hsstate, hscountry FROM master

**THESE ARE OTHER QUERIES I TRIED USING BUT ULTIMATELY DID NOT USE DUE TO MULTIPLE DATA APPEARING:**

**Selecting Draftpicks from 1985 onward:**

SELECT draftyear, draftround, draftselection, playerid, draftfrom FROM draft WHERE CAST(draftyear as int) >= 1985

**Selecting Draft Picks who made Allstar Appearances since 1985, Joining draft and allstar tables:**

SELECT draft.draftyear, draft.draftround, draft.draftselection, draft.firstname, draft.lastname, allstar.playerid, draftfrom, allstar.seasonid FROM draft JOIN allstar ON draft.playerid = allstar.playerid WHERE CAST(draftyear as int) >= 1985 ORDER BY draftyear ASC

**Selecting all draft data after 1985, merging it with allstar and master and awards:**

SELECT \* FROM draft NATURAL JOIN allstar NATURAL JOIN master NATURAL JOIN awards WHERE CAST(draftyear AS int) >= 1985

**Adding draft columns to players table:**

SELECT players.playerid, year, players.tmid, gp, minutes, points, orebound, drebound, rebounds, assists, steals, blocks, turnovers, pf, fgattempted, fgmade, ftattempted, ftmade, threeattempted, threemade, draftround, draftselection, draftfrom FROM players LEFT JOIN draft ON players.playerid = draft.playerid WHERE CAST(year AS int) >= 1985 AND CAST(gp AS int) >= 58

**Adding award and award year to big table:**

SELECT players.playerid, players.year, players.tmid, gp, minutes, points, orebound, drebound, rebounds, assists, steals, blocks, turnovers, pf, fgattempted, fgmade, ftattempted, ftmade, threeattempted, threemade, draftround, draftselection, draftfrom, award, awards.year FROM players LEFT JOIN draft ON players.playerid = draft.playerid LEFT JOIN awards ON players.playerid = awards.playerid AND draft.playerid = awards.playerid WHERE CAST(players.year AS int) >= 1985 AND CAST(gp AS int) >= 58 AND CAST(awards.year AS int) >=1985

**Adding master columns to big table:**

SELECT players.playerid, players.year, players.tmid, gp, minutes, points, orebound, drebound, rebounds, assists, steals, blocks, turnovers, pf, fgattempted, fgmade, ftattempted, ftmade, threeattempted, threemade, draftround, draftselection, draftfrom, award, awards.year, master.firstname, master.lastname, master.pos, height, weight, college, birthcity, birthstate, birthcountry, highschool, hscity, hsstate, hscountry

FROM players LEFT JOIN draft ON players.playerid = draft.playerid LEFT JOIN awards ON players.playerid = awards.playerid LEFT JOIN master ON players.playerid = master.bioid WHERE CAST(players.year AS int) >= 1985 AND CAST(gp AS int) >= 58 AND CAST(awards.year AS int) >=1985

**Navigating My EXCEL WORKBOOK:**

Pages labeled as ‘imported’ (i.e.: all\_players\_imported, draft\_table\_imported) are the tables I brought in from SQL.

I then made copies of those and labeled them ‘clean’ (i.e.: allstars\_clean, all\_players\_clean1), then cleaned up those tables.

Then I made copies of the clean tables and called them ‘clean\_copy’ just taking their values (i.e.: allstars\_clean\_copy, draft\_table\_clean\_copy).

‘all\_players\_clean2’ is where I brought in allstar and allnba columns using INDEX/MATCH with Multiple Criteria. This may be making my excel workbook work slow.

I copied all that data and put it into ‘master\_table\_clean’ which is all my cleaned data and the dataset I used for Tableau.

‘pivot\_tables’ has all the pivot tables used.

‘sql\_queries’ lists the same queries I did here, but the page was required for the project in excel so I put it there too.

**Cleaning methods used in Excel:**

Removed columns with no data such as ‘first season’ or ‘last season’.

For players without a playerID in draft data replaced with ‘none’

Replaced blanks with ‘-‘ sign

Replace multiple position players with single position, either C, F, or G, based on the position listed first. Example, player pos labeled ‘C-F’ became ‘C’ because C listed first, indicating natural position, whereas F is an alternate position.

Calculated height from all inches to separate feet (quotient function) and remaining inches (mod function).

Created new columns on per game basis: minutes per game, points per game, rebounds per game, assists, steals, blocks, turnovers, fouls per game by dividing them by column ‘gp’.

Created Percentages for field goals (fg), free throws (ft), and three pointers.

Eliminated double entrees. Doubles occurred when there was multiple draft information (i.e. someone attended 2 colleges, or a foreign born player who also went to college in USA) Second entry was most recent schooling so eliminated the first. Or there was wrong draft pick information (no round 99, pick 0 is impossible), in both scenarios the second entry was the correct one.

Deleted duplicate rows

Got rid of incorrect calculations and left them as zeroes (0/0 just became 0)

Inserted relevant columns to create one master data table, including adding allstar and allnba info to the correct season where a playerid was awarded them (the INDEX/MATCH with Multiple Criteria).