Derek Zhou, Hiu Tung Ngan, Yilei Fu

Info 257 Assignment 2b

We are going to create a database for the last 4 World Cups and users can search for information regarding to matches, players, and tournaments. Soccer enthusiasts can search through game history to learn interesting details including which referees handed out the most penalties, which countries score the most, or which stadiums host the most matches. One main issue is that we are very dependent on the Player table, which we reference as a foreign key in several other tables. If it is altered incorrectly our entire database will not work.

We plan to create 10 CSV files containing information about players, matches, group stages, formations, stadiums, referees, countries, and more. Datasources include FIFA documents, Wikipedia records, and different soccer fan websites. We manually entered and sorted the data we found into Excel files.

A preliminary data dictionary for our entities is as follows:

For each player, we assign a unique player id to each player as primary key. It is used as a foreign key in the datasets Lineup, Goal, and Disciplines so users can check player statistics. Players each have varchars Name, Position, Club, and integers Birthday\_year, Birthday\_month, Birthday\_day, and Caps.

Each lineup in Lineups contains foreign keys Match\_ID, PlayerID, Integers Pos, Sub\_Mins, and a boolean Starter. Pos represents position, and Sub\_Mins means substitute minutes.

Each discipline in Disciplines contains foreign keys Match\_ID, Player\_ID, varchar Discipline\_Type, and integer Mins (minutes played).

For each match in Matches, we assign an integer primary key Match\_ID, varchar Day, Team1\_country, Team2\_country, Stage, date Date, time Time, and integers Arena\_ID, Team1\_Score, Team2\_Score, Team1\_Shootout, Team2\_Shootout, Stadium\_ID. Match\_ID is used as a foreign key in the datasets Lineups and Disciplines. Team1\_country and Team2\_country are names that are found in the dataset Country. Stage is a number to be found in Group\_Stage.

For each country in Countries, we assign an integer primary key Country\_ID and varchars Country\_Name, Abbreviation, and Confederation. We may plan to add a dataset of the various soccer clubs. We currently have data for such, but this data does not link each club to their country, and we would have to manually add that information in.

For each stadium in Stadiums, we assign an integer primary key Stadium\_ID, varchars Stadium\_Name, and integers Capacity and Opened (opening year). Capacity represents the total amount of seats located in the stadium. A possible future addition could be to add how crowded a stadium was during any match.

For each referee in Referees, we assign varchars Name, Country, date Born, and integers Yellow\_Cards, Yellow\_to\_Red\_Cards and Red\_Cards. Country references a Country\_Name in the dataset Countries.

For each group\_stage in Group\_Stage, we assign varchar Group\_Name, Team, and integers Pos, PId, W, D, L GF, GA, GD, and Pts.

