The attributes for these data tables are relatively easy to match since the team names are unique and the typically the same in each table. Also, the date attribute follows the same format, making matching the date simple. As for the scores, all you need is an exact match. Because of the simplicity, I added another table to concatenate as well called "MLBStadiums.csv". This table has the team name that plays there, the stadium name, stadium location, seating capacity, and the year it opened up.

Matching Strategies:

- I used exact matching when verifying the information of when the game was played and the scores of the game. All the information should be the exact same since we are dealing with dates and scores.
- I used fuzzy string matching when matching the team names. This is mainly a safety thing. The team names should be the same, but if there happens to be a space or missing letter, then it would still match the team names. There are also a few instances where team names have slightly changed (i.e. LA Angels of Anaheim to Los Angeles Angels → matches at 0.63). The matcher will still match these team names relatively highly. This is what I want since they are the same franchise, just different names.

Sizes of Tables:

- mlbGames → 36,270 tuples
- mlbGames2 → 30,441 tuples
- MLBStadiums → 125 tuples
- C → 29,930 tuples

So table C has these attributes:

- ID
- ltable_ID
- ltable_Date
- ltable_HomeTeam
- ltable_HomeScore
- ltable_AwayTeam
- ltable_AwayScore
- rtable_ID
- rtable_Date
- rtable_HomeTeam

- rtable_HomeScore
- rtable_AwayTeam
- rtable_AwayScore
- stable_ID
- stable_TeamName
- stable_StadiumName
- stable_StadiumLocation
- stable_Capacity
- stable_OpeningYear

Problems:

- Assigning the stadiums to game games was slightly difficult. The were multiple teams that have moved stadiums since 2010. So, I needed to gather information on past stadiums and assign them based on the date the game was played on and if the stadium was still being used.
- There are 2,430 games played each season, so 32,490 played since 2010 given there weren't any strikes or shortened seasons. So, I am missing around 2,560 games in my concatenated table.