Part A Summary of Preprocessing

The preprocessing of data was performed in Phase 2 of the project. The steps to do so were as follows:

1. **Removing duplicate columns**: Some columns that are found in multiple dataframes were dropped to prepare for merging
2. **Renaming columns**: Remaned to identify the type of statistic they represent (ex:_PERGAME, _TOTAL, etc)
3. **Merging dataframes** 'per_game', 'total', and 'advanced' frames merged based on the common columns 'Player' and 'Tm' using a left join.
4. **Mapping Team Abbreviations**: A dictionary mapping NBA team names to their corresponding abbreviations was created and then used to create a dataframe.
5. **Merging Standings Data**: Standings data, including team records, is merged with the team abbreviation DataFrame based on the 'Team' column.
6. **Calculating Win Percentage**: Win percentage ('PCT') is calculated from the 'Record' column in the standings DataFrame.
7. **Merging MVP Data**: MVP data is merged into the main DataFrame based on player names.
8. **Cleaning Player Names**: Player names are cleaned by removing unnecessary characters.
9. **Removing Duplicate Players**: Duplicate player entries are removed, keeping only the first occurrence.
10. **Filtering MVP Candidates**: Two sets of filtering criteria, labeled as 'TIGHT FILTER' and 'LOOSE FILTER', are applied to identify MVP-caliber players based on various statistical thresholds and MVP vote shares.
11. **Removing Unnecessary Columns**: Columns related to team information ('Tm', 'Team', 'Record') are dropped from the DataFrame.
12. **Generating Surrogate Keys**: Sequential surrogate keys starting from 1 are generated and set as the first column of the DataFrame.