

Meeting Time:

- 02/07/2025, 13:00 EST

Meeting Reason:

- Dataset Approval with Dr. Mallavarapu

Discussion:

- How to introduce null values
 - One column vs multiple columns
 - How to add null values using code
 - Introduce null values after merge
- Which columns to merge
 - Left, Right, Inner, or Outer?
 - Which columns are used to merge the two datasets?
- **Players.csv (3922 rows and 8 columns)**
 - <https://www.kaggle.com/drgilermo/nba-players-stats?select=Players.csv>
 - **KEEP:**
 - Player (name)
 - Height
 - Weight
 - College
- **Seasons_Stats.csv (24,691 rows and 53 columns)**
 - https://www.kaggle.com/drgilermo/nba-players-stats?select=Seasons_Stats.csv
 - **KEEP:**
 - Year (end of season year),
 - Player (name),
 - Position,
 - Age,
 - Games,
 - Minutes Played,
 - Player Efficiency Rating,
 - True Shooting (TS) %,
 - Usage%,
 - Win Shares (WS),
 - Box Plus/Minus (BPM)
 - Field Goal Percentage (FG%)
 - 3 Point Percentage (3P%)
 - 2 Point Percentage (2P%)
 - Effective Field Goal Percentage (eFG%)
 - Free Throw Percentage (FT%)
 - Total Rebounds
 - Assists
 - Steals

- Blocks
- Turnovers
- Points
- **DISCARD:**
 - Team,
 - Games Started,
 - OWS,
 - DWS,
 - WS/48,
 - blanl,
 - blank2,
 - 3PAr,
 - FTr,
 - ORB%,
 - DRB%,
 - TRB%,
 - AST%,
 - STL%,
 - BLK%,
 - TOV%,
 - OBPM,
 - DBPM,
 - Value Over Replacement (VORP) (UPDATE: KEEP VORP)
 - Field Goals
 - Field Goal Attempts
 - 3 Point Goals
 - 3 Point Attempts
 - 2 Point Goals
 - 2 Point Attempts
 - Free throw attempts
 - Free throws
 - Offensive rebounds
 - Defensive rebounds
 - Personal Fouls
- **CLEAN (before merging):** [Games,

```
ex1 = pd.DataFrame()
ex1['name'] = ['harsh', 'minji', 'aaron']
ex1['major'] = ['csc', 'csc', 'sts']
ex1
```

✓ 0.0s

	name	major
0	harsh	csc
1	minji	csc
2	aaron	sts

```
ex2 = pd.DataFrame()
ex2['name'] = ['aaron', 'harsh', 'minji', 'csc442']
ex2['colors'] = ['blue', 'green', 'yellow', 'purple']
ex2
```

✓ 0.0s

	name	colors
0	aaron	blue
1	harsh	green
2	minji	yellow
3	csc442	purple

```
mergeddf = pd.merge(ex1, ex2, on='name', how='inner')
mergeddf
```

✓ 0.0s

	name	major	colors
0	harsh	csc	green
1	minji	csc	yellow
2	aaron	sts	blue

```
mergeddf = pd.merge(ex1, ex2, on='name', how='right')  
mergeddf
```

✓ 0.0s

	name	major	colors
0	aaron	sts	blue
1	harsh	csc	green
2	minji	csc	yellow
3	csc442	NaN	purple