

Emeric Rochford
3-9-2022
Capstone 1

Phase One

NBA Stats (1947 – Present): <https://www.kaggle.com/sumitrodatta/nba-aba-baa-stats>

1. The useability score of this dataset is 9.8, so it is above the 7.0 threshold.
 2. The dataset has 5488 rows of data, which is above the 3000 row estimate for what would be a suitable size.
 3. There are 35 columns in the dataset, which is more than the required 7 columns.
 4. Although the dataset is mostly numerical, there is a good mix of both numerical and categorical data.
-
- a. This dataset is something I'm very interested in, and always have been interested in. I love comparing players stats throughout the years, as well as finding ways to do this through different decades. I'm a big fan of basketball and sports in general, so I will thoroughly enjoy the opportunity to visualize this data.
 - b. This dataset is at least 90% ready for me to make clear visualizations. Some of the column names will need matching and renaming, and potentially some players will be removed. However, these won't be major overhauls and shouldn't take too long.
 - c. There is a decent mix of categorical and numerical data in this dataset.