**NBA Hot Hand Theory - Milestone #3**

Group: Zainab Doctor, Hai-Tsang Huang, Zecai Liang, Bailey Liao

**1. Project Approach and Cleaning the Data**

The variables we are interested in using are Game\_ID, shot\_number, period or game\_clock, shot\_clock, shot\_distance, shot\_result, closest\_defender\_ID, close\_defender\_distance, FGM, points and player\_ID.

We cleaned the data by eliminating shots that meet certain criteria which as a result cannot or should not be included in the analysis. Shots where the shot\_clock is NaN or under 1 second and where the game\_clock is under 1 second were excluded, with the assumption that these may be last-second attempts during which the ball is thrown up. The results of this filtering are shown in **Figure 1.**



**Figure 1.** Result of filtering shot log data based on the criteria detailed above.

**2.**

**3.**

**4.**

**5. Next Steps and Future Directions**

Other features we would like to look at are whether players are more likely to get “hot” when playing at home versus away, with the hypothesis that they know their home court better. Additionally, if players are stratified based on their average points per game across a season (ie 0-10, 10-20, 20-30+ points per game), are higher tier players more likely to get hot? Are there mid-tier players that get "hot" by having nights that look like those of upper tier players? Another aspect that would be interesting to explore is using Defender\_ID to investigate whether certain defenders are better at stopping or breaking streaks.