Introduction:

Kobe Bryant marked his retirement from basketball by scoring 60 points in his final game as a member of the Los Angeles Laker team on Wednesday, April 12, 2016. Starting to play professional basketball at the age of 17, Kobe earned the sport’s highest accolades throughout his long career. Using 20 years of data on Kobe's shots made and shots missed.

EDA:

During Exploratory Data Analysis, we are assuming following features/variables are of no effect on the model.

game\_event\_id

recId

game\_id

season

game\_date

team\_id

team\_name

Also, we assume the multi collinearity of following variables

* action\_type combined\_shot\_type --> So, we had gone with combined\_shot\_type
* lat, loc\_x, loc\_y and lon --> So, we had gone with lat, lon
* matchup, opponent --> So, we had gone with opponent

Outlier:

Based on Cooks’s D data and plain data analysis , we are seeing no outlier present in the selected variables. Here is the first 10 observations with highest cook’s D value. Since these values are less than 3, we assume that there are no outliers

| **Obs** | **recId** | **cooks** |
| --- | --- | --- |
| **1** | 2694 | .000776686 |
| **2** | 27990 | .000524152 |
| **3** | 29122 | .000450069 |
| **4** | 1074 | .000356229 |
| **5** | 17734 | .000351587 |
| **6** | 28077 | .000343475 |
| **7** | 13341 | .000336435 |
| **8** | 19447 | .000336255 |
| **9** | 17359 | .000332707 |
| **10** | 28422 | .000331336 |

Upon inspection of these results, one is quickly drawn to the negative correlation coefficient of lat and

shot\_distance which values at a whopping 0.8810. We definitely have a case for further collinearity

investigation here. This is further supported in our review of the parameter estimates results for VIF and

Tol: