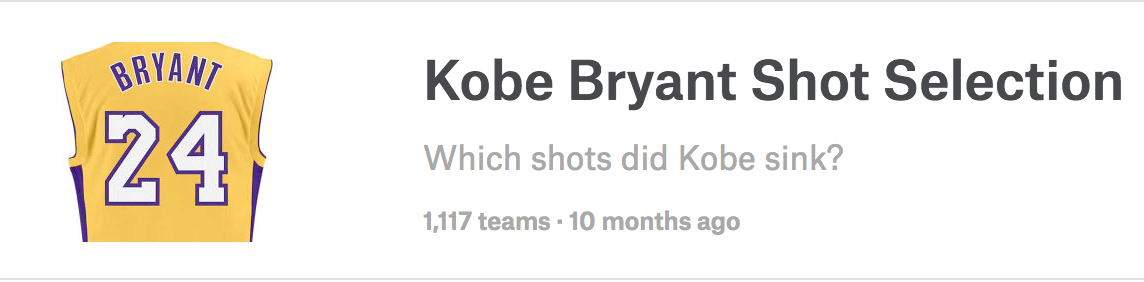
PROJECT 3



**KAGGLE LINK:** [**https://www.kaggle.com/c/kobe-bryant-shot-selection**](https://www.kaggle.com/c/kobe-bryant-shot-selection)

**OVERVIEW:**

Kobe Bryant marked his retirement from the NBA by scoring 60 points in his final game as a Los Angeles Laker on Wednesday, April 12, 2016. Drafted into the NBA at the age of 17, Kobe earned the sport’s highest accolades throughout his [long career](http://www.npr.org/sections/thetwo-way/2016/04/13/474107238/kobe-bryants-life-and-career-by-the-numbers).

Using 20 years of data on Kobe's swishes and misses, can you predict which shots will find the bottom of the net? This competition is well suited for practicing classification basics, feature engineering, and time series analysis. Practice got Kobe an eight-figure contract and 5 championship rings. What will it get you? [ANSWER: ENLIGHTENMENT! > Eight Figures!]

**DATA:**

This data contains the location and circumstances of every field goal attempted by Kobe Bryant took during his 20-year career. Your task is to predict whether the basket went in (shot\_made\_flag).

We have removed 5000 of the shot\_made\_flags (represented as missing values in the csv file). These are the test set shots for which you must submit a prediction. You are provided a sample submission file with the correct shot\_ids needed for a valid prediction.

To avoid [leakage](https://www.kaggle.com/wiki/Leakage), your method should only train on events that occurred prior to the shot for which you are predicting! Since this is a playground competition with public answers, it's up to you to abide by this rule. This is from Kaggle, I will relax this requirement, although it may make sense depending on your analysis!

The field names are self explanatory and contain the following attributes:

|  |  |
| --- | --- |
| * action\_type * combined\_shot\_type * game\_event\_id * game\_id * lat * loc\_x * loc\_y * lon * minutes\_remaining * period * playoffs * season * seconds\_remaining | * shot\_distance * shot\_made\_flag (this is what you are predicting) * shot\_type * shot\_zone\_area * shot\_zone\_basic * shot\_zone\_range * team\_id * team\_name * game\_date * matchup * opponent * shot\_id |

**DELIVERABLES:**

1 Paper (8 page limit with up to 5 page Appendix … 13 page max.)

Intro

Data Description

Exploratory Data Analysis

Interpretation Models

Build models to provide arguments and evidence for or against each proposition:

1. Kobe’s shooting percentage is subject to a home field advantage. That is, Kobe’s shooting percentage is better or worse at home than when he is away.
2. The odds of Kobe making a shot decrease with respect to the distance he is from the hoop. If there is evidence of this, quantify this relationship. (CIs, plots, etc.)
3. The probability of Kobe making a shot decreases linearly with respect to the distance he is from the hoop. If there is evidence of this, quantify this relationship. (CIs, plots, etc.)
4. The relationship between the distance Kobe is from the hoop and the odds of him making the shot is different if they are in the playoffs. If there is evidence of this, quantify this relationship. (CIs, plots, etc.)
5. With respect to question 4, is there evidence of a difference after accounting for a home field advantage? That is, does the answer to question 4 depend on a home filed advantage? If there is evidence of this, quantify this relationship. (CIs, plots, etc.)
6. Is Kobe clutch? After accounting for the distance of the shot, does Kobe’s shooting percentage increase when he is taking a shot in the last 30 seconds of a period?

Predictive Model

Build a model to win the Kaggle competition. In contrast to Project 1, you only need to present your best model (rather than your top 3.)

**ASSESSMENT / EVALUATION:**

Notice that there are not a lot of specific requirements for the number or type of methods used in the analysis nor are their specific requirement for the presentation of the analysis. It is up to you and your group to provide the best, most complete and most appropriate analysis possible. The grading scale will be based relative to the quality of the projects that come in. This means that the best paper will “set the curve” and all other papers will be judged as their “distance” from that paper.

I will say that good papers:

1. Are presented in an organized, neat and consistent fashion. (Labeled plots, figures and tables, consistently formatted, indented and labeled headers and sub headers, etc.)
2. Have no typos, misspelled words, grammatical mistakes, etc.
3. Use a variety of methods.
4. Usually have a creative use of a method or methods.
5. Have been contributed two by all (both) group members and over time so that iterative changes can be made (as opposed to all at once … example: the night before.)

Don’t let this less structured grading scheme make you too nervous. If your group puts strong and consistent work into this project, you will do fine (B+ or better). Also, it is mathematically possible that everyone could get an A … it’s not that kind of “curve”. It all depends on the “distance” between the papers.

**SOFTWARE AND METHODS:**

For the above, you must use SAS and must use only the methods we have studied thus far in your coursework. That being said, you can use innovative techniques inside of those methods like the polynomial coding scheme for ordinal variables. If you have any questions about this please let me know and we can discuss your idea.

**BONUSES:**

1. An extra point will be given to the Kaggle winner: the group with the best public Kaggle score.
2. Up to an extra 3pts will be given to teams that solve the last Kaggle prediction problem with a method we have not studied this semester. In addition, for this problem, you may use any language you wish (SAS, R, Python, C++, etc.) Also, you may use an additional 1 page to present this analysis.
3. (3pts) Model Kobe’s shooting percentage over time. Does he appear to get better over time? Use your knowledge of the methods we have studied so far to answer this question the best way possible. (Again, you may have 1 additional page to answer this question.)