# Hot Shot: A Statistical Analysis of Duke Women's Basketball Offensive Field Goal Efficiency

Emma Schmidt

Spring 2023

## 1 Abstract

The purpose of this study is to assess the field goal efficiency of the Duke Women's Basketball offense with data collected on possession by possession breakdowns. More specifically, the aim is to understand what features make the most efficient possessions, and how a coaching staff can utilize this feedback to optimize the offense. Studies like this are robust in the NBA, where data and resources are nearly unlimited. However, in the women's college basketball space, this kind of work doesn't publicly exist. This is in part due to a lack of available data, as well as, limited resources at the collegiate level. Because data science work is underused in women's college basketball, a study like this could offer competitive advantage to programs who are able to leverage statistical tools to gain further statistical insight of themselves and their opponents before resources become more widely available in the sport. In order to proceed with this study, all of the data was hand collected from game film. Specifically, this analysis is run using data collected on offensive possessions from the 2022-23 Duke Women's Basketball Season. The possessions of specific interest were those in which a field goal attempt was made. Possession features in the data set include: location of the shot attempt, player that took the shot, number of passes in the half court, number of paint touches, number of times the ball changed sides of the floor, possession type (half court, transition, baseline out-of-bounds, sideline out-of-bounds), numbers advantage (3 on 2, 2 on 1, etc), shot clock at time of attempt, whether or not the possession was the result of an offensive rebound, and whether or not the field goal attempt was a make or miss. Using this data, I begin my investigation of offensive efficiency, first executing feature selection and then proceeding with model construction. Finally, I assess my model results, highlight key possession features, and offer offensive strategy modifications that look to improve overall field goal efficiency.

## 2 Methodology

### 2.1 Feature Selection

Beginning the process of feature selection, I start by hypothesising what characteristics of a possession offer the offense an advantage, and further whether or not these characteristics can be manipulated to create a more efficient offense. Two major concepts stand out: shot selection and ball movement.

First exploring shot selection, two features of the data offer insight into this metric, location of the shot attempt and whether or not the shot attempt came from a pass, off of the dribble, or from a put back. The first pinpoints where on the court the shot went up with respect to the basket, and the second explains how the shot was created. While both data points offer critical possession information, predetermining exactly where a shot will occur in a possession is hard to do, and as a result, location of the shot is not a good feature for the model. On the other hand, it is plausible to design offensive strategy around different types of shots (catch and shoot = from a pass, drives = off the dribble, etc). This makes the type of shot a sound feature for the model. Next, breaking down ball movement, features of the data describing this metric include: number of passes per possession, number of times the ball changes sides of the floor per possession, and paint touches per possession. Ball movement is critical to an offense because it can cause unnatural shifts and breakdowns to a defense that create an offensive advantage. Number of passes per possession is a simple but effective metric for determining if the ball is moving. Number of times the ball changed sides of the floor offers an even deeper look into ball movement, showing how many times the defense was potentially forced to shift from one side of the floor to the other. Finally, paint touches per possession demonstrates how many times the ball went inside the paint (the area that most closely surrounds the basket), an area that is most crucial for the defense to protect. All three of these features can be easily incorporated into potential offensive strategy, and are thus strong features for the model. A potential concern when considering three separate features that all relate to ball movement is high collinearity. A correlation analysis of the features reveals that this is not a serious concern. Paint touches vs. passes and side changes vs. paint touches demonstrate low correlation, with correlation coefficients of .13 and .16, respectively. Side changes vs. passes are moderately correlated with a correlation coefficient of .60. Given the context of the data this is not surprising, but is worth keeping in mind when exploring results. In summary, the four selected model features are shot type, number of passes, number of times the ball changed sides of the floor, and number of paint touches. These features were selected above others because they offer direct insight into shot selection and/or ball movement, and additionally, are controllable when designing offensive strategy. Figure 1 offers bar plot breakdowns for each of the selected features.

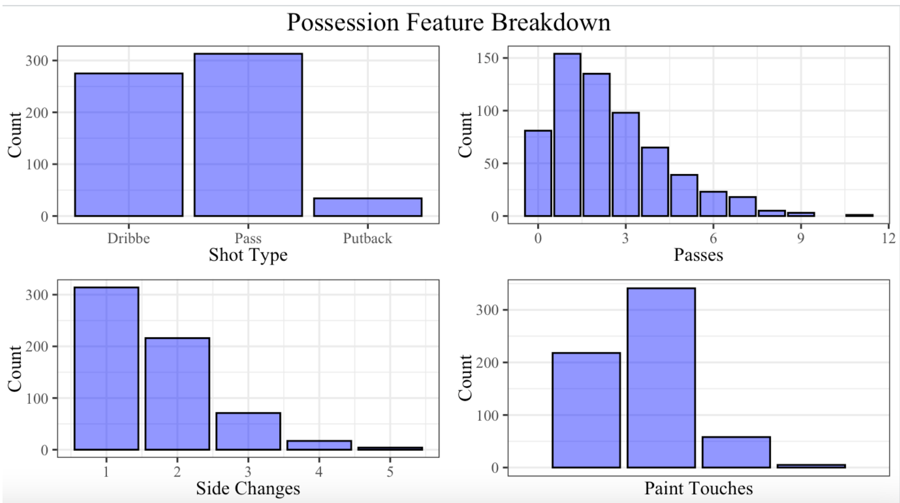


Figure 1: Model Feature Bar Plots

## 2.2 Model Selection

After determining which hypothesized modeling features will be indicative of an efficient offensive possession, I now decide how I will measure and model offensive efficiency. With the data at hand the obvious choice in defining success for a possession is whether or not the shot attempt was made or missed. With the selected model features and a binary response variable, I move forward with a logistic regression model. This is a sound modeling choice because it offers highly interpretable results that can be easily explained to a coaching staff. The model is specified below:

$$logit(p(x)) = \beta_0 + \beta_1 I(Shot = Pass) + \beta_2 I(Shot = Putback) + \beta_3 x_{passes} + \beta_4 x_{sides} + \beta_5 x_{paint\ touches}$$

Beginning the modeling process, I execute a random 80:20 split of the data into a train and test set. From there I utilize the training set to construct the model, and the test set to evaluate the success of the model. In evaluating the the model I consider both in-sample and out-of-sample prediction success rate. Prediction responses with probabilities greater than .50 will be considered made baskets (i.e. successful possessions).

# 3 Results

The model summary yields interesting results, in some cases supporting my hypothesis about a feature and in other cases is contradicting. The first major call outs are the statistically significant features: shots off of a pass, number of paint touches in possession, and side changes with significance at $\alpha = .05$, $\alpha = .01$, and $\alpha = .10$, respectively. Both shots off of a pass and paint touches offer positive estimates, suggesting that possessions in which the shot attempt was the result of a pass and possessions with paint touches are expected to have higher log odds of success than possessions that do not. Interestingly, both passes per possession and side changes per possession have negative estimates. I believe this occurs for a number of reasons. The first reason is that transition possessions (possessions in which the field goal attempt is within the first eight seconds of the shot clock), tend to have fewer passes and side changes, but higher field goal efficiency. To assess this hypothesis I run a second model, this time including an interaction term between passes and possession type (half court, transition, baseline-out-of-bounds, and sideline-out-of-bounds), and finally a third model this time including an interaction term between side changes and possession type. In both scenarios the estimates for passes and side changes were still negative regardless of possession type. This leads me to a less measurable theory that the team is oversharing the ball and potentially turning down a high efficiency shot in exchange for an ultimately less efficient shot. Additionally, in some scenarios a possession goes until near the end of the shot clock, in which there would be a high volume of passes and side changes, but low field goal efficiency. To better understand the effects of late shot clock field goals on overall efficiency future work could be done. The current analysis is focused on controllable features of a possession, and as a result, further shot clock analysis falls out of the scope of the study. Field goal attempts that are a result of a putback have a positive estimate, however the volume of putback shots in comparison to shots off the dribble and from a pass is very low. Moving to prediction accuracy, the model was able to successfully predict the outcome of in-sample possessions with 63.81%

accuracy and of out-of-sample possessions with 64.12% accuracy. Table 1, seen below, offers a full model summary.

Table 1: Model Summary

| | *Dependent variable:* |
|---|---|
| | Field Goal Estimate |
| Shot off Pass | 0.415** |
| Shot Putback | 0.369 |
| Paint Touches | 0.509*** |
| Side Changes | −0.248* |
| Passes | −0.089 |
| Intercept | −0.137 |
| Observations | 525 |
| Log Likelihood | −348.758 |
| Akaike Inf. Crit. | 709.517 |
| *Note:* | *p<0.1; **p<0.05; ***p<0.01 |
| In-Sample Prediction Rate | 63.81% |
| Out-of-Sample Prediction Rate | 64.12% |

## 4 Discussion

While the model accuracy may not seem strong, with regards to the data it is actually quite interesting. Of the possessions in the data a miss occurred 54.87% of the time, meaning on any given possession a field goal outcome can successfully be predicted roughly 55% of the time. With the model the number increases to approximately 64% accuracy. To further understand how an optimized possession compares to an average possession, Figure 2 offers heat maps displaying the field goal percentages across pre-specified zones. The average possession heat map contains information on every single possession in the data. The optimized possession includes data on possessions where either the shot was a result of the pass or there was at least one paint touch, and the possession had below the mean of side changes per possession, or some combination of the three. The heat map of the optimized possession is generally more red, meaning that field goal efficiency is improved in these possessions. Using this information I recommend that the coach puts high emphasis on getting the ball into the paint throughout the flow of the offense. Offensive game plan centered on getting the ball inside can stem from hard post ups from the centers, as well as from aggressive drives off of on ball screens by guards. Additionally, I would design plays that create shots off of the pass rather than off of the dribble. These shots can be created by a number of different off ball screening strategies including flex screens, push screens, stagger screens, and flares. With enough volume in the data one could potentially explore the efficiency of specific plays cross referenced with different lineups on the floor. This type of analysis is beyond the breadth of the current data. An additional consideration to obtain further accuracy and understanding could be done to examine how the heat maps are impacted by varying shot volumes in each of the specified zones.

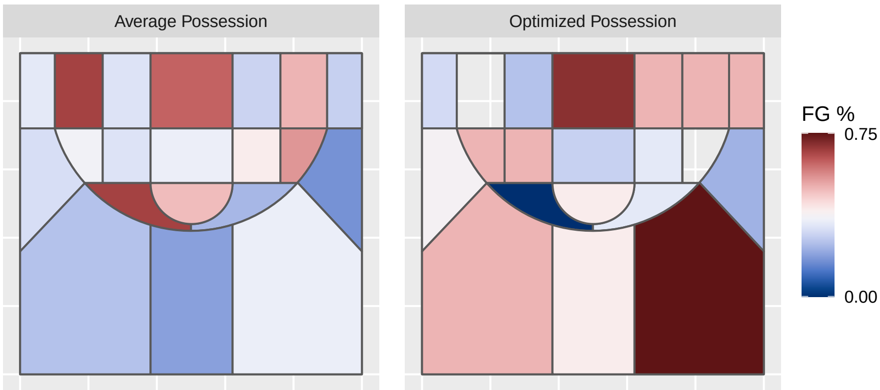


Figure 2: Possession Comparison