**Kobe Bryant Shot Selection**

**Team Lucky Shot:**

Jasmine Coleman, Linda Eliasen, and Jeff Leath

**Overview**

Kobe Bryant is undoubtably one of the most influential professional athletes of the modern day. Drafted into the NBA at the age of 17, Kobe earned the sport’s highest accolades throughout his long career. His professional career in the National Basketball Association was a remarkable 20 years. 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. Using 20 years of data on Kobe's swishes and misses, as provided by Kaggle, our goal is to build a model that most accurately predicts which shots made the basket given certain known parameters.

**Data**

Our dataset includes over 28 parameters for each of over 25,000 shots made by Kobe Bryant during his 20-year career. To assist in the model creation, we have also engineered an additional six variables. The explanatory data includes categorical and continuous features covering the dimensions of: Action, Distance, Game, Environment, and Time.



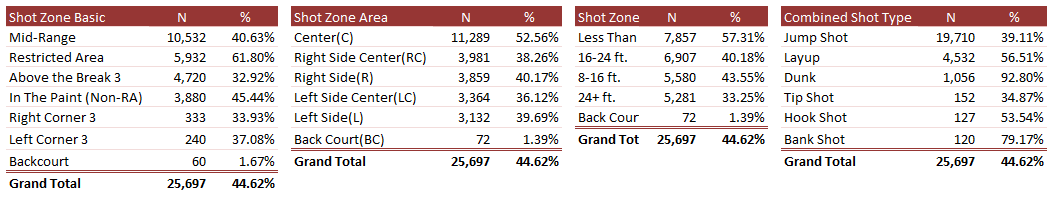
**Exploratory Data Analysis**

The response variable for this analysis is a binary response, where 1 = ‘shot-made’ and 0 = ‘shot-missed’. To understand the data, we first reviewed the frequency and distribution of certain key explanatory variables. (At this point, key explanatory variables have been identified by domain experts, pending completion of the analysis.)

Of the over 25,000 shots made at the basket, Kobe made 44.6% of them. This metric is referred to as the ‘Field Goal Percentage’ (FG%) in basketball.

Given 20 years of history, some clear observations can be made of Kobe’s shooting style:

* He prefers, and is most accurate, when shooting from the Center, as opposed to either Left or Right court.
* When he can’t shoot from the center, he prefers the Right side of the court, over the left.
* His most frequent Shot Type is the Jump Shot (76%), though his FG% here is below the overall mean.



**Shots Attempted**

The graphics below (Tableau Visualizations) display a density overlay of shots Kobe made (by various categories) on the ‘offensive’ side of the basketball court.

* There is a higher density of shots close to the basket and along the free throw line. It appears that Kobe prefers to take the shot when he is closest to the basket and frequency reduces the further the shot range is from the basket.

|  |  |
| --- | --- |
|  |  |
|  |  |

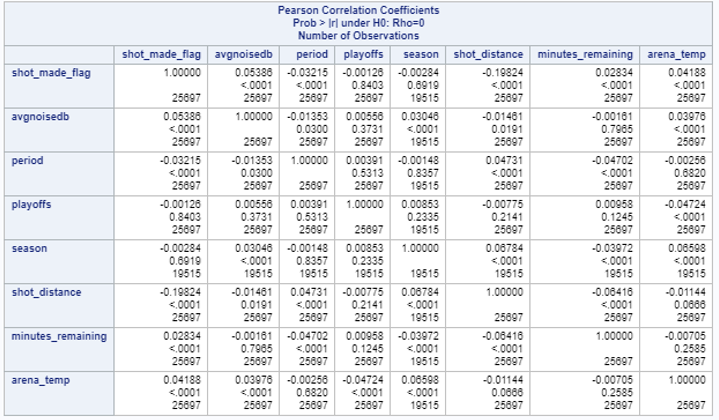
**Shot Accuracy**

Through our analysis, several patterns emerged:

* There is clearly a strong, negative linear relationship between FG% and ‘Distance’ from the basket.
* In terms of accuracy, the shorter the distance, the greater Kobe’s success.
* His most accurate shots were Layups and Dunks, both of which happen within eight feet of the goal.
* Following Kobe’s injury in 2013, it is evident that his median ratio of shots scored decreased drastically up until 2016 when he retired.
* The Lakers did not make the playoffs every season. With few exceptions, Kobe’s accuracy decreased during playoff games.

|  |  |
| --- | --- |
|  |  |
|  |  |

**Correlation Matrix**



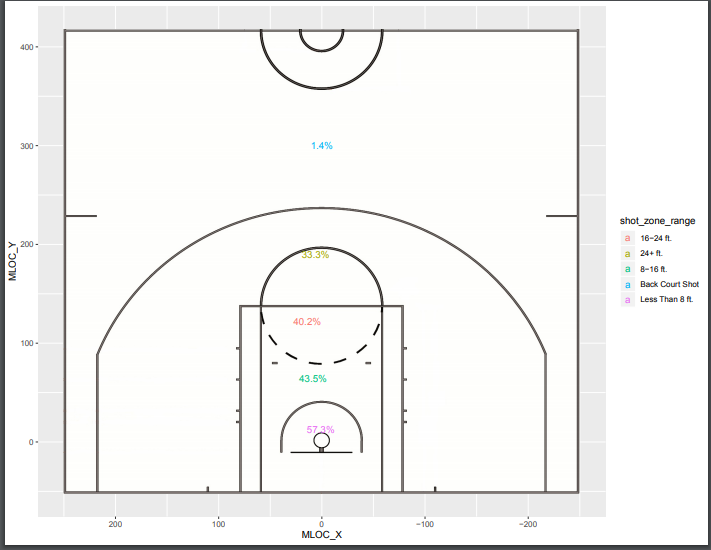
Upon reviewing the histograms and scatter plots (Appendix A - EXPLORATORY DATA ANALYSIS), it is evident that there is not normality across multiple features, such as shot distance, shot position and shot type. These elements have practical significance to the question of interest and no normalization is required.

Ten variables have been dropped from the dataset, as they are either distinct key identifiers or they had high multi-collinearity with existing variables. These variables are identified in the first graphic, where the ‘Use’ field equals ‘Drop’.

**Odds/Probability**

In our exploratory analysis, we found evidence to support the claim that Kobe was more successful at making shots the closer he was to the hoop.

**Accuracy by Shot Zone**



We used a Chi-Square test on a logistic model of shot distance on the shot made flag to test whether the odds of Kobe making a shot did indeed decrease with respect to the distance he was from the hoop. At a 95% significance level, we found shot distance to be statistically significant (*p* < 0.0001) as shown in the results below.

|  |  |
| --- | --- |
|  |  |

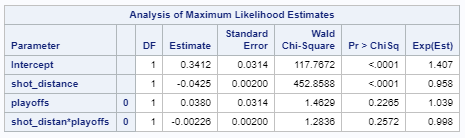
This means that for each one-unit increase in shot distance, the expected change in the log odds of a shot being made is -0.0441. Thus, we expect to see a 4.3% decrease in the odds of Kobe making a shot for each unit farther he was away from the hoop. A 95% confidence interval for this change in odds is (4.0%, 4.6%).

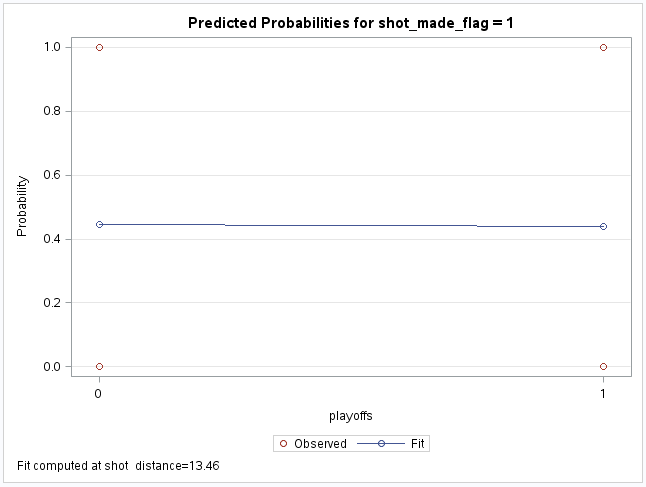
While we expected to find that the odds of Kobe making a shot decreased as his distance from the hoop increased, we wanted to see if this relationship was linear. The plot below illustrates the change in probability of Kobe making a shot with respect to the distance he was from the hoop.

|  |
| --- |
|  |

We can see that the probability of Kobe making a shot under the basket was ~60%. And, the probability appears to decrease linearly as distance increases until Kobe reached ~30ft. from the basket. After this we can see that Kobe didn’t make many successful shots, even though he made attempts all the way up until 80ft. away from the basket. As a result, we see that the probability of Kobe making a shot beyond 30ft. begins to curve and decrease exponentially.

Lastly, we wanted to investigate whether being in the playoffs had an impact on this relationship. We used logistic regression to model the relationship between Kobe’s accuracy and his distance from the hoop with an interaction on the playoff factor. The charts below show that neither the interaction between shot distance and playoffs (Chi-Square test *p* = 0.2572) nor being in the playoffs (Chi-Square test *p* = 0.2265) significantly impacted the odds of Kobe making a shot with respect to his distance from the hoop.





With this information in mind, our next task was to build a predictive model for classifying whether a shot was missed or made.

**Predictive Modeling**

The data has a binary outcome – either the shot made it in the basket, or it did not. Therefore, we focused on logistic regression and linear discriminant analysis (LDA) to develop our predictive model. The assumption is being made that all attempted shots are independent of each other; the success or failure of the previous shot does not have any effect on the next shot going in. This avoids time series issues, especially as 5,000 data points were removed from the submission file.

Logistic regression results were similar for forward, stepwise and our manual model. LDA had better success in predicting missed shots (12.3% error rate) versus shots that made it in (56% error rate)(Appendix A - PREDICTIVE MODELING). This seems to indicate that for misses, it is clear in which situations Kobe did not make the shot goal. However, for field goals, it seems that there may be more factors that influence success than what was provided in the data (i.e., guarded / unguarded).

When comparing the log-loss of each model, the best model was forward logistic regression with the lowest log-loss at 0.6108, followed closely by our manual model at 0.6112.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **MODEL COMPARISON** | | |  |  |  |
| **MODEL** | **AIC** | **SC** | **TRAIN ROC** | **TEST ROC** | **LOG-LOSS** |
| **Logistic** Manual | 3818.72 | 4066.54 | 0.7178 | 0.7089 | 0.61120 |
| Forward | 3816.9 | 4076.8 | 0.7104 | 0.7111 | 0.61085 |
| Backward | 4213.29 | 4461.11 | 0.7137 | 0.6275 | 0.67644 |
| Stepwise | 3829.62 | 4204.37 | 0.7137 | 0.7164 | 0.61292 |
| **LDA** |  |  |  |  | 3.68059 |

The forward model in our validation phase indicated that the variables in the chart below are most important to predicting the success or failure of achieving a field goal, holding all else constant.

|  |  |
| --- | --- |
| Forward Selection Results – Training data | Model Fit – Test Data |
|  |  |

Prior to applying this model to the prediction file, we trained the complete dataset on forward selection for confirmation. With the additional observations, loc\_y fell off and shot\_type\_pt, shot\_distance and LastSecShot made it into the model based on an alpha of 0.05.

**Final Variables for Predictive Model**

Shot Action Type Attendance Time Remaining Combined Shot Zone

Major Injury Period Arena Temperature Shot Type Points

Shot Distance Last Second Shot

**Kaggle Score for Predictive Model**



**Summary**

There are a multitude of factors that impact players on the court and their ability to perform well during a game. From the variables we have to examine and those we have deemed most relevant, we were able to produce a predictive model that accurately predicted over 60% of the shots Kobe made.

Although our linear discriminate model only misclassified 34.2% Kobe’s shots in total, it had low sensitivity in regards to correctly identifying the shots Kobe made. The LDA model only classified 43.4% of the shots Kobe did make accurately while only misclassifying 12.7% of the shots that Kobe missed.

This could be because we used the default threshold of 50% to classify shots as missed or made. However, Kobe’s overall accuracy was slightly lower at ~44%. Another reason for this could be due to the assumption of independence of each shot.

Future analysis should investigate whether the sequence of shots during a game or years of training and experience by the players have an effect on the accuracy and classification of shots which would then turn this into a time series analysis.

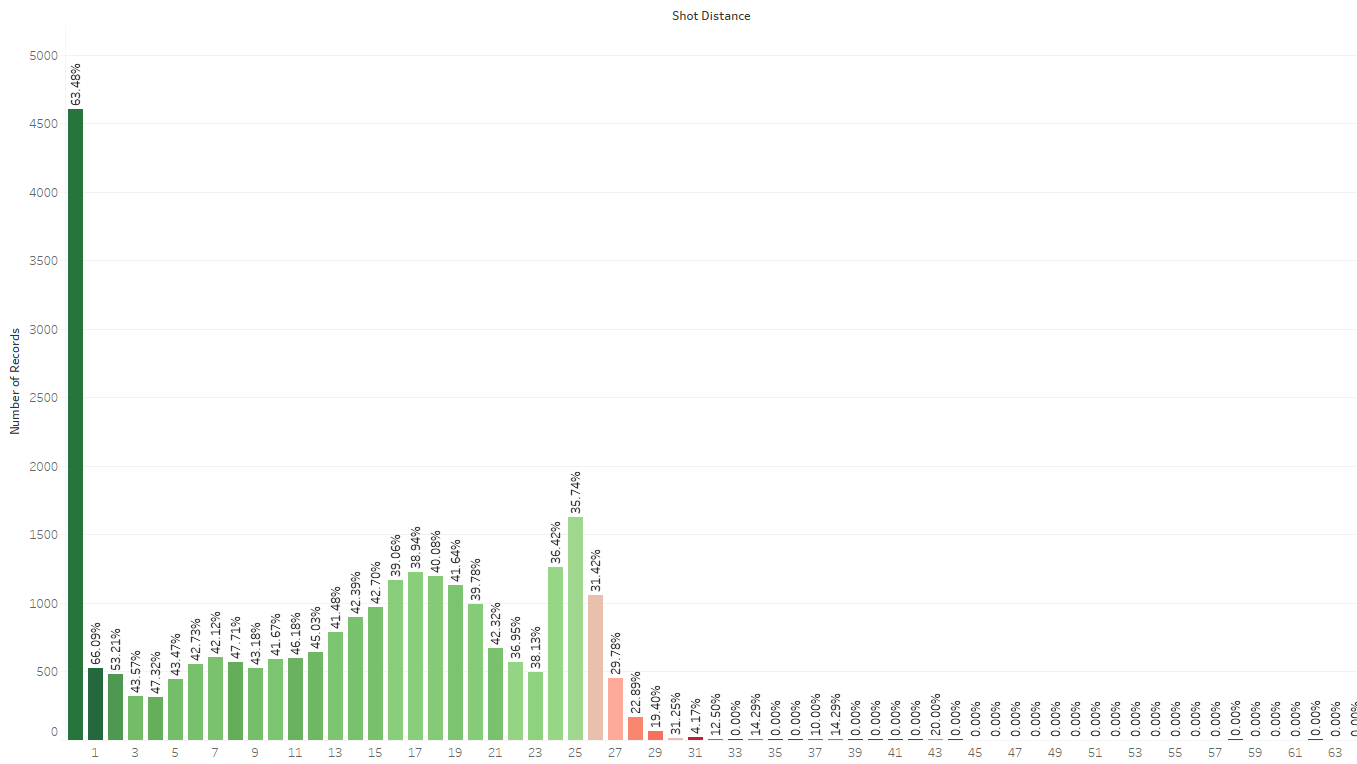
Nevertheless, from the information we have been given and the observations we have made, our best statistical advice to Kobe would have been to always make the shot within zero feet of the goal, preferably by dunking. Although this is obviously not practically feasible, as best said by Wayne Gretsky, “You miss 100% of the shots you don’t take”.

**Appendix A – Additional Charts**

**EXPLORATORY DATA ANALYSIS**

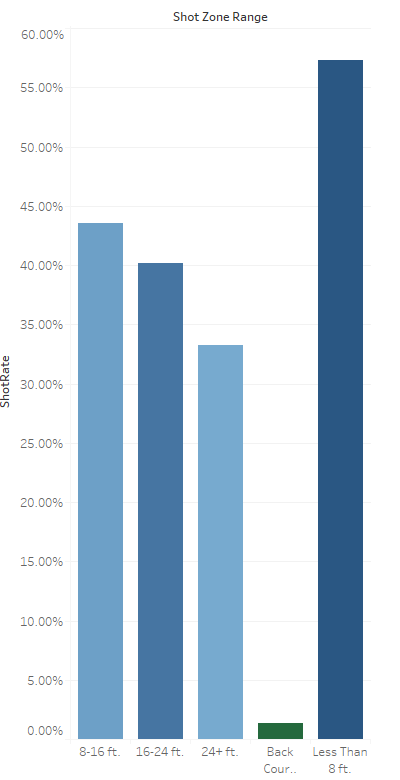
*Accuracy by Shot Distance (Tableau Visualization)*

Shot Distance: X=Shot Distance Y=Number of Shots: Accuracy% (green to red)

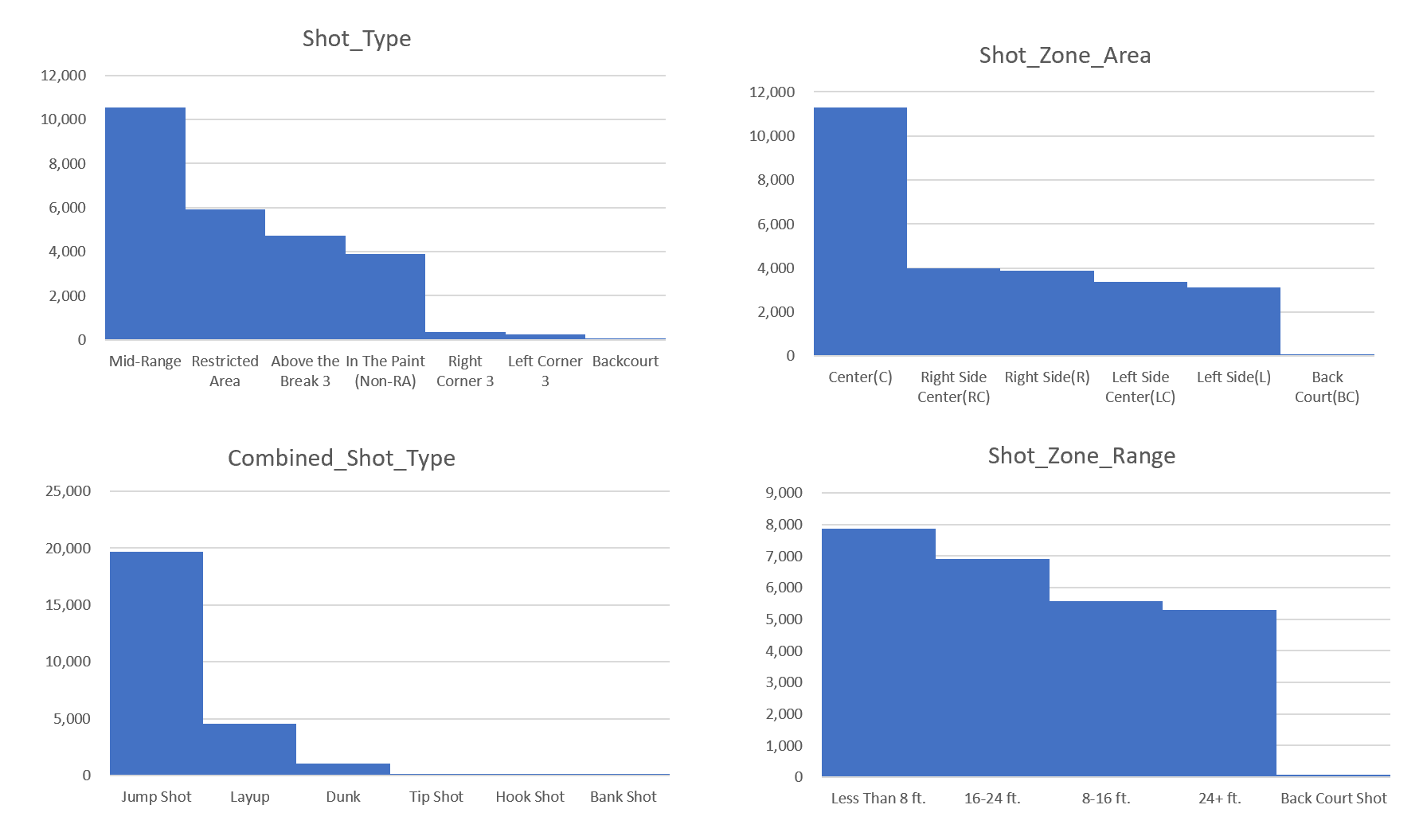


*Accuracy by Shot Zone Range (Tableau Visualization)*

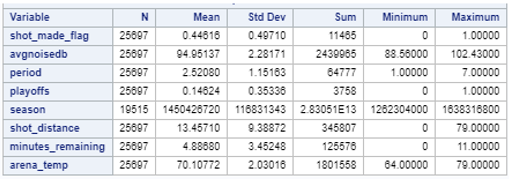
Shot Zone Range: X=Range Band, Y=Accuracy%; color-> Number of shots (dark blue to green)



*EDA Histograms*



*EDA Summary Statistics*



*Predictive Modeling*

**Logistic Regression**

All Variables Included

|  |  |
| --- | --- |
| Logistic Manual - Train | Logistic Manual - Test |
|  |  |
|  |  |
|  |  |

Forward Model

|  |  |
| --- | --- |
| Logistic Forward - Train | Logistic Forward - Test |
|  |  |
|  |  |
|  |  |

Backward Model

|  |  |
| --- | --- |
| Logistic Backward - Train | Logistic Backward - Test |
|  |  |
|  |  |
|  |  |

Stepwise Model

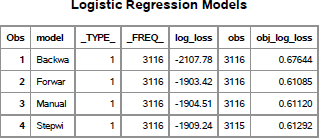
|  |  |
| --- | --- |
| Logistic Stepwise - Train | Logistic Stepwise - Test |
|  |  |
|  |  |
|  |  |

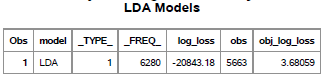
**LDA**

Cross-validation Summary using Quadratic Discriminant Function

|  |  |
| --- | --- |
| **Training Data** | **Test Data** |
|  |  |
|  |  |

**LOG-LOSS COMPARISON**





**Appendix B – Code**

*Exploratory Data Analysis*

/\* viewing the data \*/

proc univariate data=ds1;

histogram shot\_made\_flag;

qqplot shot\_made\_flag;

run;

/\*Historgram \*/

proc sgplot data = ds1;

histogram shot\_made\_flag;

density shot\_made\_flag;

title "ShotMade";

inset 'Inset Test'/ position = topright border;

run;

proc sgplot data = ds1;

histogram shot\_made\_flag;

density shot\_made\_flag;

title "ShotMade";

inset 'Inset Test'/ position = topright border;

run;

/\*Proc Corr - shot\_zone\_range\*/

proc corr data = ds1 plots(maxpoints=none)=matrix(histogram);

run;

/\* Correlation Matrix \*/

proc corr data=ds1;

var shot\_made\_flag avgnoisedb period playoffs season shot\_distance minutes\_remaining arena\_temp ;

run;

proc corr data=ds1;

var shot\_made\_flag shot\_distance;

run;

proc glm data = ds1;

model shot\_made\_flag = shot\_distance;

means shot\_distance / hovtest = bf;

run;

*Figure – Shot Accuracy by Shot Distance*

kobe %>%

group\_by(shot\_distance) %>%

summarise(Accuracy=mean(shot\_made\_flag)) %>%

ggplot(aes(x=shot\_distance, y=Accuracy)) +

geom\_line(aes(colour=Accuracy)) +

geom\_point(aes(colour=Accuracy), size=2) +

scale\_colour\_gradient(low="red", high="blue") +

labs(title="Accuracy by shot distance", x="Shot distance (ft.)") +

xlim(c(0,45)) +

theme\_bw() +

theme(legend.position="none",

plot.title=element\_text(hjust=0.5))

*Figure – Shot Accuracy by Shot Type*

kobe %>%

group\_by(action\_type) %>%

summarise(Accuracy=mean(shot\_made\_flag),

counts=n()) %>%

filter(counts>20) %>%

ggplot(aes(x=reorder(action\_type, Accuracy), y=Accuracy)) +

geom\_point(aes(colour=Accuracy), size=3) +

scale\_colour\_gradient(low="yellow", high="blue") +

labs(title="Accuracy by shot type") +

theme\_bw() +

theme(axis.title.y=element\_blank(),

legend.position="none",

plot.title=element\_text(hjust=0.5)) +

coord\_flip()

*Figure – Shot Accuracy by Season*

kobe %>%

group\_by(season) %>%

summarise(Accuracy=mean(shot\_made\_flag)) %>%

ggplot(aes(x=season, y=Accuracy, group=1)) +

geom\_line(aes(colour=Accuracy)) +

geom\_point(aes(colour=Accuracy), size=3) +

scale\_colour\_gradient(low="yellow", high="blue") +

labs(title="Accuracy by season", x="Season") +

theme\_bw() +

theme(legend.position="none",

plot.title=element\_text(hjust=0.5),

axis.text.x=element\_text(angle=45, hjust=1))

*Figure – Shot Accuracy by Season x Playoffs*

kobe %>%

group\_by(season) %>%

summarise(Playoff=mean(shot\_made\_flag[playoffs==1]),

RegularSeason=mean(shot\_made\_flag[playoffs==0])) %>%

ggplot(aes(x=season, group=1)) +

geom\_line(aes(y=Playoff, colour="Playoff")) +

geom\_line(aes(y=RegularSeason, colour="RegularSeason")) +

geom\_point(aes(y=Playoff, colour="Playoff"), size=3) +

geom\_point(aes(y=RegularSeason, colour="RegularSeason"), size=3) +

labs(title="Accuracy by season",

subtitle="Playoff and Regular Season",

x="Season", y="Accuracy") +

theme\_bw() +

theme(legend.title=element\_blank(),

plot.title=element\_text(hjust=0.5),

plot.subtitle=element\_text(hjust=0.5),

axis.text.x=element\_text(angle=45, hjust=1))

*Figure – Shots Attempted by Shot Zone Range*

kobe.shot\_made\_flag <- kobe$numeric\_shot\_made\_flag

# https://thedatagame.com.au/2015/09/27/how-to-create-nba-shot-charts-in-r/

# half court image

courtImg.URL <- "https://thedatagame.files.wordpress.com/2016/03/nba\_court.jpg"

court <- rasterGrob(readJPEG(getURLContent(courtImg.URL)),

width=unit(1,"npc"), height=unit(1,"npc"))

ggplot(kobe, aes(loc\_x, loc\_y)) +

annotation\_custom(court, -250, 250, -50, 420) +

geom\_point(aes(shape = shot\_made\_flag, color = shot\_zone\_range ))

shots <- ddply(kobe, .(shot\_zone\_range), summarize,

 SHOTS\_ATTEMPTED = length(shot\_made\_flag),

 SHOTS\_MADE = sum(as.numeric(as.character(shot\_made\_flag))),

 MLOC\_X = mean(loc\_x),

 MLOC\_Y = mean(loc\_y))

shots[which(shots$shot\_zone\_range=="Back Court Shot"),"MLOC\_X"] = 0

shots[which(shots$shot\_zone\_range=="Back Court Shot"),"MLOC\_Y"] = 300

shots$SHOT\_ACCURACY <- (shots$SHOTS\_MADE / shots$SHOTS\_ATTEMPTED)

shots$SHOT\_ACCURACY\_LAB <- paste(as.character(round(100 \* shots$SHOT\_ACCURACY, 1)), "%", sep="")

*Log Odds of Shot Made x Distance Code*

proc logistic data=class.KobeTrain;  
model shot\_made\_flag(event='1') = shot\_distance /lackfit ctable clparm=both;  
effectplot fit (x=shot\_distance);  
run;

*Log Odds of Shot Made x Distance Code x Playoffs*

proc logistic data=class.KobeTrain;

class playoffs;

model shot\_made\_flag(event='1') = shot\_distance | playoffs /lackfit ctable clparm=both expest ;

effectplot interaction (x=playoffs);

run;

*Modeling & Prediction Code*

data kobe;

infile '/folders/myfolders/Kobe data.csv' firstobs=2 dlm=',';

input recId Shot\_Act\_Type $ loc\_x loc\_y Time\_Remaining period playoffs season $ shot\_type\_Pt

Comb\_Shot\_Zone $ shot\_zone\_range $ shot\_distance shot\_made\_flag opponent $

shot\_id HomeCourt attendance arena\_temp avgnoisedb MajInj One LastSecShot ;

run;

proc sort data=kobe;

by recid;

data kobe2;

set kobe;

randNumber = ranuni(11);

run;

data train;

set kobe2;

if randNumber <=1/4 then delete;

run;

data test;

set kobe2;

if randNumber >1/4 then delete;

run;

/\*MODELING\*/

/\*Manual\*/

proc logistic descending data = train plots=ROC;

class Shot\_Act\_Type season Comb\_Shot\_Zone shot\_zone\_range opponent / param = ref;

model shot\_made\_flag = Shot\_Act\_Type loc\_x loc\_y Time\_Remaining period playoffs

season shot\_type\_Pt Comb\_Shot\_Zone shot\_distance opponent

HomeCourt attendance arena\_temp avgnoisedb MajInj LastSecShot;

run;

proc logistic descending data = test plots=ROC;

class Shot\_Act\_Type Comb\_Shot\_Zone / param = ref;

model shot\_made\_flag = shot\_act\_type time\_remaining period

comb\_shot\_zone attendance arena\_temp / ctable ;

output out=Manual predprobs=crossvalidate;

run;

data Manual;

set Manual;

model='Manual';

log\_loss = (shot\_made\_flag\*log(XP\_1) + (1 - shot\_made\_flag)\* log(1 - XP\_1));

run;

/\*Forward\*/

proc logistic descending data = train plots=roc;

class Shot\_Act\_Type season Comb\_Shot\_Zone shot\_zone\_range opponent / param = ref;

model shot\_made\_flag = Shot\_Act\_Type loc\_x loc\_y Time\_Remaining period playoffs

season shot\_type\_Pt Comb\_Shot\_Zone shot\_distance opponent

HomeCourt attendance arena\_temp avgnoisedb MajInj LastSecShot / selection=forward ;

run;

proc logistic descending data = test plots=ROC;

class Shot\_Act\_Type Comb\_Shot\_Zone / param = ref;

model shot\_made\_flag = Shot\_Act\_Type attendance Time\_Remaining Comb\_Shot\_Zone

MajInj period loc\_y arena\_temp ;

output out=Forward predprobs=crossvalidate;

run;

data Forward;

set Forward;

model='Forward';

log\_loss = (shot\_made\_flag\*log(XP\_1) + (1 - shot\_made\_flag)\* log(1 - XP\_1));

run;

/\*Backward\*/

proc logistic descending data = train plots=ROC;

class Shot\_Act\_Type season Comb\_Shot\_Zone shot\_zone\_range opponent / param = ref;

model shot\_made\_flag = Shot\_Act\_Type loc\_x loc\_y Time\_Remaining period playoffs

season shot\_type\_Pt Comb\_Shot\_Zone shot\_distance opponent

HomeCourt attendance arena\_temp avgnoisedb MajInj LastSecShot / selection=backward ;

run;

proc logistic descending data = test plots=ROC;

class opponent / param = ref;

model shot\_made\_flag = playoffs LastSecShot avgnoisedb majinj homecourt shot\_type\_pt

shot\_distance opponent loc\_x;

output out=Backward predprobs=crossvalidate;

run;

data Backward;

set Backward;

model='Backward';

log\_loss = (shot\_made\_flag\*log(XP\_1) + (1 - shot\_made\_flag)\* log(1 - XP\_1));

run;

/\*Stepwise\*/

proc logistic descending data = train plots=roc;

class Shot\_Act\_Type season Comb\_Shot\_Zone shot\_zone\_range opponent / param = ref;

model shot\_made\_flag = Shot\_Act\_Type loc\_x loc\_y Time\_Remaining period playoffs

season shot\_type\_Pt Comb\_Shot\_Zone shot\_distance opponent

HomeCourt attendance arena\_temp avgnoisedb MajInj LastSecShot / selection=stepwise

slentry=0.2 slstay=.2 details lackfit;

run;

proc logistic descending data = test plots=ROC;

class Shot\_Act\_Type Comb\_Shot\_Zone season/ param = ref;

model shot\_made\_flag = Shot\_Act\_Type attendance time\_remaining comb\_shot\_zone

majinj period loc\_y arena\_temp season ;

output out=Stepwise predprobs=crossvalidate;

run;

data Stepwise;

set Stepwise;

model='Stepwise';

log\_loss = (shot\_made\_flag\*log(XP\_1) + (1 - shot\_made\_flag)\* log(1 - XP\_1));

run;

/\*LOG LOSS SUMMARY\*/

data combine;

set Manual Forward Backward Stepwise;

keep model log\_loss;

run;

proc means data=combine sum;

class model;

var log\_loss;

output out=LogLoss sum=log\_loss n=obs;

run;

data logloss;

set logloss;

obj\_log\_loss = (-1\*(1/obs)\*log\_loss);

if \_TYPE\_ = 1;

run;

proc print data=logloss;

run;

**LDA**

data kobe;

infile '/folders/myfolders/Kobecodeddata.csv' firstobs=2 dlm=',';

input recId LayupLayup LayupDriving JumpTurnaround JumpFadeaway JumpRunning JumpPullup

JumpTurnaroundFadeaway DunkSlam LayupReverse JumpBank DunkDriving DunkDunk TipTip

LayupDrivingFingerRoll DunkAlleyOop JumpStepBack JumpFloating LayupDrivingReverse

HookHook LayupRunningFingerRoll LayupAlleyOop DunkReverse BankTurnaround LayupFingerRoll

BankRunning DunkDrivingSlam HookRunning JumpFadeawayBank JumpDriving JumpHook Other loc\_x loc\_y

TimeRemaining period playoffs Seas2 Seas3 Seas4 Seas5 Seas6 Seas7 Seas8 Seas9 Seas10 Seas11 Seas12 Seas13

Seas14 Seas15 Seas16 Seas17 Seas18 Seas19 Seas20 shot\_type\_Pt RightSideRMidRange CenterCInThePaintNonRA

LeftSideLMidRange RightSideCenterRCMidRange RightSideCenterRCAbovetheBreak3

LeftSideCenterLCAbovetheBreak3 LeftSideCenterLCMidRange CenterCMidRange

CenterCAbovetheBreak3 RightSideRInThePaintNonRA RightSideRRightCorner3

LeftSideLInThePaintNonRA LeftSideLLeftCorner3 BackCourtBCBackcourt BackCourtBCAbovetheBreak3

shot\_distance shot\_made\_flag BKN BOS CHA CHI CLE DAL DEN DET GSW HOU IND LAC MEM MIA MIL MIN NJN NOH

NOP NYK OKC ORL PHI PHX POR SAC SAS SEA TOR UTA VAN WAS HomeCourt attendance arena\_temp avgnoisedb

MajInj One LastSecShot;

run;

/\*LDA\*/

proc sort data=kobe;

by recid;

data kobe2;

set kobe;

randNumber = ranuni(11);

run;

data train;

set kobe2;

if randNumber <1/4 then delete;

run;

data test;

set kobe2;

if randNumber >1/4 then delete;

run;

proc discrim data= train pool = test crossvalidate testdata = test testout = ldares;

class shot\_made\_flag;

var LayupLayup LayupDriving JumpTurnaround JumpFadeaway JumpRunning JumpPullup

JumpTurnaroundFadeaway DunkSlam LayupReverse JumpBank DunkDriving DunkDunk TipTip

LayupDrivingFingerRoll DunkAlleyOop JumpStepBack JumpFloating LayupDrivingReverse

HookHook LayupRunningFingerRoll LayupAlleyOop DunkReverse BankTurnaround LayupFingerRoll

BankRunning DunkDrivingSlam HookRunning JumpFadeawayBank JumpDriving JumpHook Other loc\_x loc\_y

TimeRemaining period playoffs Seas2 Seas3 Seas4 Seas5 Seas6 Seas7 Seas8 Seas9 Seas10 Seas11 Seas12 Seas13

Seas14 Seas15 Seas16 Seas17 Seas18 Seas19 Seas20 shot\_type\_Pt RightSideRMidRange CenterCInThePaintNonRA

LeftSideLMidRange RightSideCenterRCMidRange RightSideCenterRCAbovetheBreak3

LeftSideCenterLCAbovetheBreak3 LeftSideCenterLCMidRange CenterCMidRange

CenterCAbovetheBreak3 RightSideRInThePaintNonRA RightSideRRightCorner3

LeftSideLInThePaintNonRA LeftSideLLeftCorner3 BackCourtBCBackcourt BackCourtBCAbovetheBreak3

shot\_distance BKN BOS CHA CHI CLE DAL DEN DET GSW HOU IND LAC MEM MIA MIL MIN NJN NOH

NOP NYK OKC ORL PHI PHX POR SAC SAS SEA TOR UTA VAN WAS HomeCourt attendance arena\_temp avgnoisedb

MajInj LastSecShot;

run;

data ldares;

set ldares;

model='LDA';

log\_loss = (shot\_made\_flag\*log(\_1) + (1 - shot\_made\_flag)\* log(1 - \_1));

run;

/\*LOG LOSS SUMMARY\*/

data LDASum;

set ldares;

keep model log\_loss;

run;

proc means data=ldares sum;

class model;

var log\_loss;

output out=LogLoss sum=log\_loss n=obs;

run;

data logloss;

set logloss;

obj\_log\_loss = (-1\*(1/obs)\*log\_loss);

if \_TYPE\_ = 1;

run;

proc print data=logloss;

run;

/\*FINAL MODEL - FORWARD\*/

data kobe;

infile '/folders/myfolders/Kobe data.csv' firstobs=2 dlm=',';

input recId Shot\_Act\_Type $ loc\_x loc\_y Time\_Remaining period playoffs season $ shot\_type\_Pt

Comb\_Shot\_Zone $ shot\_zone\_range $ shot\_distance shot\_made\_flag opponent $

shot\_id HomeCourt attendance arena\_temp avgnoisedb MajInj LastSecShot ;

run;

proc logistic descending data = kobe;

class Shot\_Act\_Type season Comb\_Shot\_Zone shot\_zone\_range opponent / param = ref;

model shot\_made\_flag = Shot\_Act\_Type loc\_x loc\_y Time\_Remaining period playoffs

season shot\_type\_Pt Comb\_Shot\_Zone shot\_distance opponent

HomeCourt attendance arena\_temp avgnoisedb MajInj LastSecShot /

selection=forward rsquare link=logit expb;

CODE file="/folders/myfolders/pprob.sas";

run;

data ForwardPredict;

infile '/folders/myfolders/Kobe predict.csv' firstobs=2 dlm=',';

input recId Shot\_Act\_Type $ loc\_x loc\_y Time\_Remaining period playoffs season $ shot\_type\_Pt

Comb\_Shot\_Zone $ shot\_zone\_range $ shot\_distance shot\_made\_flag opponent $

shot\_id HomeCourt attendance arena\_temp avgnoisedb MajInj LastSecShot ;

%include "/folders/myfolders/pprob.sas";

output = ForwardPredict;

run;

proc export data=ForwardPredict

outfile='/folders/myfolders/ForwardPredict.csv'

dbms=csv

replace;

run;