DSCI 318 - Final Project Analysis Report

Russell Whitworth

2022-12-08

Objective

I want to create the best model to predict salary using contract length, team under contract for, and primary position played by a player. I will use backwards selection to create 9 models: full model, model with all three 2 way interactions, three models with two 2-way interactions, three models with a singular two way interaction, and the basic model; and determine the best model using adjusted r-squared.

Data Summary

The dataset used in this project, Major League Baseball Salaries 2021, was collected from USA Today Sports. This dataset details MLB salaries based upon each team's opening day roster at the beginning of the 2021 MLB season. Detailed within the source dataset are the following:

- Name: player name
- Team: team of player
- POS: primary position of player
- Salary: average annual value (AAV) of the contract in USD
- Years: length of current contract in years and the years the contract spans
- Total.Value: total value of the contract in USD

I made the following changes to the dataset outside of R:

- Created Years_Length: extracts the length of current contract in years from Years column
- Changed Salary and Total.Value from strings to numeric values by removing commas

Reading Dataset into R

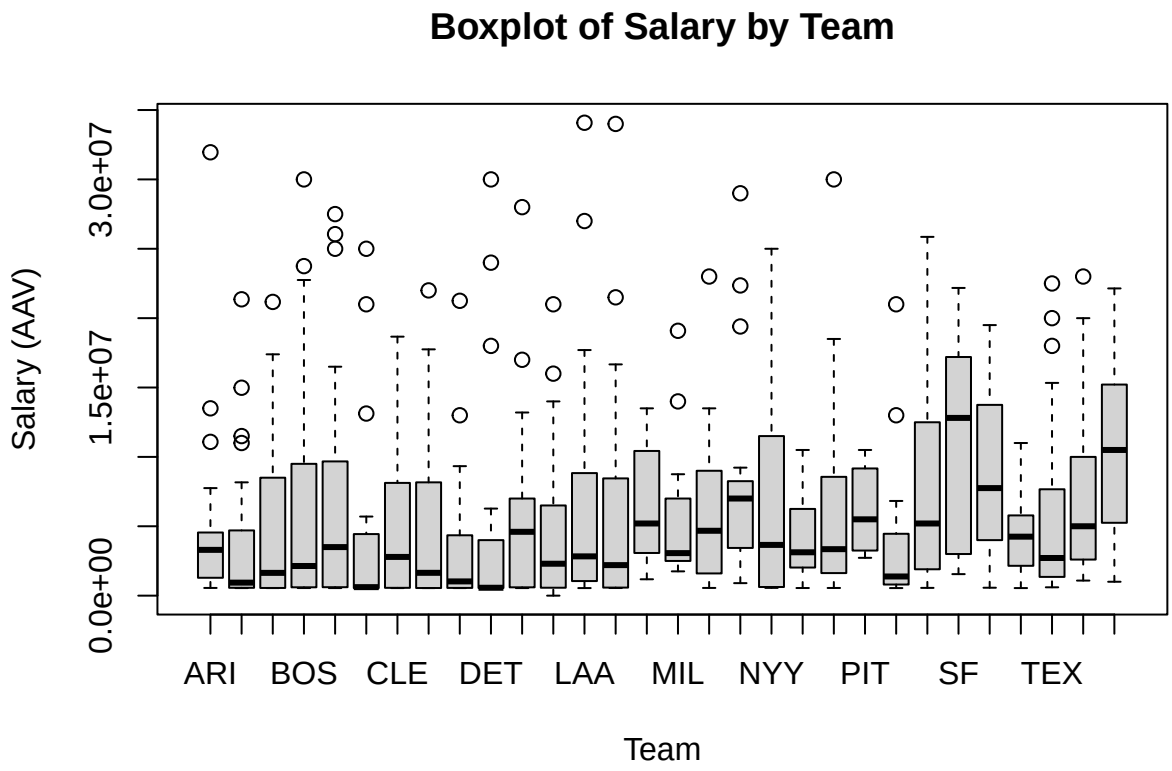
```
salary <- read.csv("salary.csv")

salary <- salary[,c("Team", "POS", "Years_Length", "Salary")]
salary$Team <- as.factor(salary$Team)
salary$POS <- as.factor(salary$POS)
head(salary)
```

```
##   Team POS Years_Length   Salary
## 1  LAA  OF            6 34083333
```

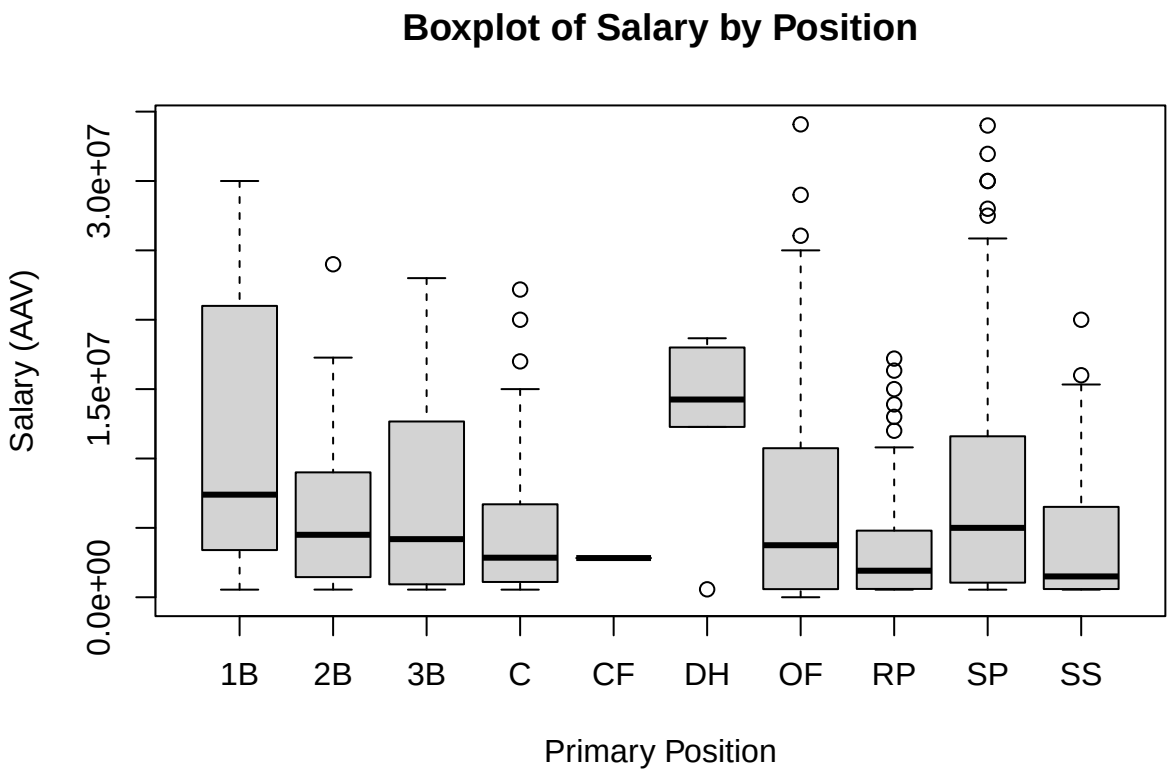
```
## 2  LAD  SP             7 34000000
## 3  ARI  SP             6 31954483
## 4  DET  1B            10 30000000
## 5  BOS  SP             7 30000000
## 6  PHI  SP             3 30000000
```

Visualizing the Predictors

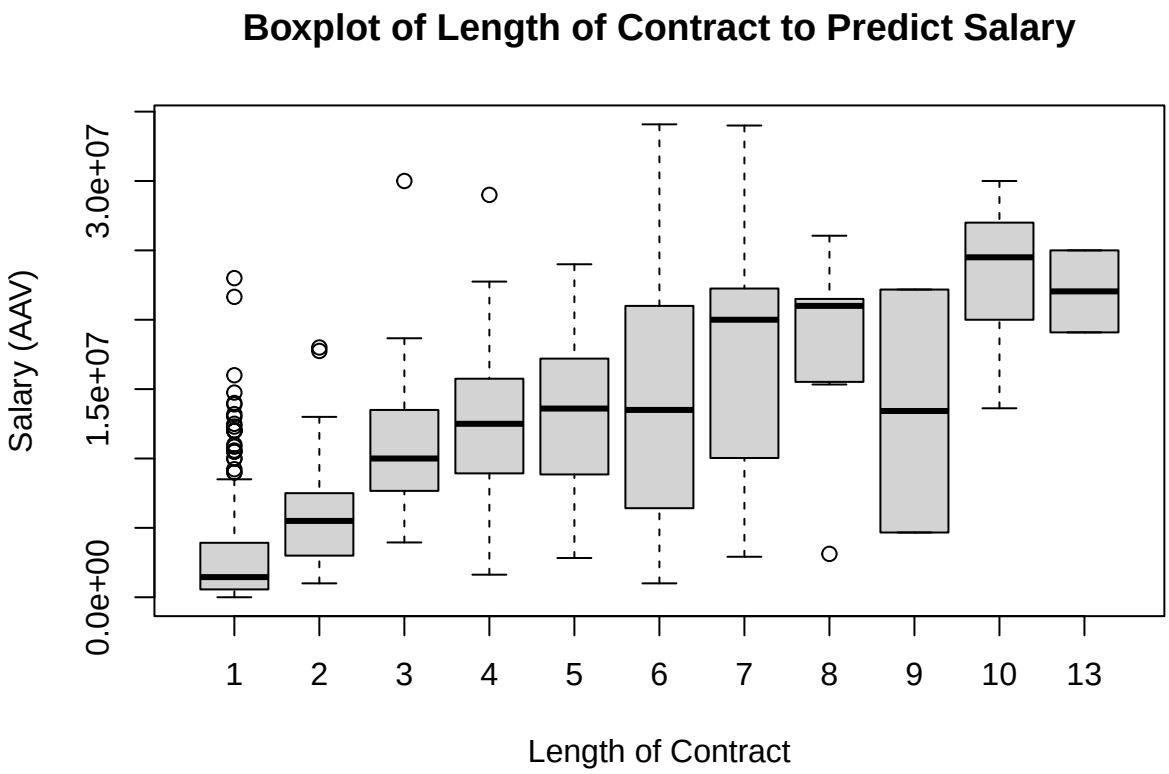
```
boxplot(Salary~Team,data=salary, main="Boxplot of Salary by Team",
   xlab="Team", ylab="Salary (AAV)")
```




```
boxplot(Salary~POS,data=salary, main="Boxplot of Salary by Position",
   xlab="Primary Position", ylab="Salary (AAV)")
```




```
boxplot(Salary~Years_Length,data=salary, main="Boxplot of Length of Contract to Predict Salary",
   xlab="Length of Contract", ylab="Salary (AAV)")
```



Models and Diagnosis

Full Model

```
fullmodel.lm <- lm(Salary ~ POS * Team * Years_Length, salary)
fullmodelr2 <- summary(fullmodel.lm)$r.squared
fullmodeladjustedr2 <- summary(fullmodel.lm)$adj.r.squared

# Printing out the first 10 coefficients and r-squared metrics
coef(fullmodel.lm)[0:10]
```

```
## (Intercept)       POS2B       POS3B        POSC       POSCF       POSDH
##  -270833.25  7869914.12 21200104.65  -129166.75  2254700.00 12786914.55
##       POSOF       POSRP       POSSP       POSSS
##  2775620.48   -51866.75  -899055.35  -248933.33
```

```
cat("R-Squared:", fullmodelr2, "\nAdjusted R-Squared:", fullmodeladjustedr2)
```

```
## R-Squared: 0.8655051
## Adjusted R-Squared: 0.7179656
```

ANOVA Analysis of Full Model

```
salaryfull.aov <- aov(Salary ~ Team * POS * Years_Length, salary)
summary(salaryfull.aov)
```

```
##                          Df    Sum Sq   Mean Sq F value   Pr(>F)
## Team                     29 1.962e+15 6.767e+13   5.271 1.54e-14 ***
## POS                       9 3.193e+15 3.548e+14  27.638  < 2e-16 ***
## Years_Length              1 1.162e+16 1.162e+16 905.045  < 2e-16 ***
## Team:POS                194 4.531e+15 2.336e+13   1.819 1.58e-06 ***
## Team:Years_Length        29 1.261e+15 4.349e+13   3.388 5.85e-08 ***
## POS:Years_Length          6 3.842e+14 6.403e+13   4.987 6.91e-05 ***
## Team:POS:Years_Length    60 1.751e+15 2.918e+13   2.273 3.31e-06 ***
## Residuals               299 3.838e+15 1.284e+13
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
summaryanova <- summary(salaryfull.aov)[[1]]["Df"]
df <- summary(salaryfull.aov)[[1]]["F value"]
colnames(df) <- gsub(" ", ".", colnames(df))

df$FCrit <- with(summaryanova, qf(c(0.05),Df,299,lower.tail=FALSE))
df$IsSignificant <- with(df, ifelse(F.value > FCrit, 1, 0))
df
```

```
##                         F.value  FCrit IsSignificant
## Team                       5.27 1.5055             1
## POS                       27.64 1.9113             1
## Years_Length             905.05 3.8727             1
## Team:POS                   1.82 1.2362             1
## Team:Years_Length          3.39 1.5055             1
## POS:Years_Length           4.99 2.1290             1
## Team:POS:Years_Length      2.27 1.3636             1
## Residuals                       1.2099
```

Using alpha=0.05, all of the main effects and interactions are significant.

Model with interaction between Team and POS (MODEL 2)

```
salary2.lm <- lm(Salary ~ Team * POS + Years_Length, salary)
model2r2 <- summary(salary2.lm)$r.squared
model2adjustedr2 <- summary(salary2.lm)$adj.r.squared

# Printing out the first 10 coefficients and r-squared metrics
coef(salary2.lm)[0:10]
```

```
## (Intercept)     TeamATL     TeamBAL     TeamBOS     TeamCHC     TeamCIN
##   -957564.3   2013533.8   2043252.9   8083333.5  -9254926.9    960128.3
##     TeamCLE     TeamCOL     TeamCWS     TeamDET
##   3458077.1   4731833.5  11457820.7   5960128.3
```

```
cat("R-Squared:", model2r2, "\nAdjusted R-Squared:", model2adjustedr2)
```

```
## R-Squared: 0.7465167
## Adjusted R-Squared: 0.5966141
```

ANOVA Analysis of Model 2

```
salary2.aov <- aov(Salary ~ Team * POS + Years_Length, salary)
summary(salary2.aov)
```

```
##                Df    Sum Sq   Mean Sq F value   Pr(>F)
## Team           29 1.962e+15 6.767e+13   3.685 2.26e-09 ***
## POS             9 3.193e+15 3.548e+14  19.323  < 2e-16 ***
## Years_Length    1 1.162e+16 1.162e+16 632.779  < 2e-16 ***
## Team:POS      194 4.531e+15 2.336e+13   1.272    0.024 *
## Residuals     394 7.234e+15 1.836e+13
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
summaryanova <- summary(salary2.aov)[[1]]["Df"]
df <- summary(salary2.aov)[[1]]["F value"]
colnames(df) <- gsub(" ", ".", colnames(df))

df$FCrit <- with(summaryanova, qf(c(0.05),Df,394,lower.tail=FALSE))
df$IsSignificant <- with(df, ifelse(F.value > FCrit, 1, 0))
df
```

```
##              F.value  FCrit IsSignificant
## Team            3.69 1.4964             1
## POS            19.32 1.9037             1
## Years_Length  632.78 3.8652             1
## Team:POS        1.27 1.2218             1
## Residuals            1.1805
```

Using alpha=0.05, all of the main effects and interactions are significant.

Model with interaction between Team and Years_Length (MODEL 3)

```
salary3.lm <- lm(Salary ~ Team * Years_Length + POS, salary)
model3r2 <- summary(salary3.lm)$r.squared
model3adjustedr2 <- summary(salary3.lm)$adj.r.squared

# Printing out the first 10 coefficients and r-squared metrics
coef(salary3.lm)[0:10]
```

```
## (Intercept)     TeamATL     TeamBAL     TeamBOS     TeamCHC     TeamCIN
##   -572783.1    528833.7   1248098.7   1221269.0    645046.4   -393581.0
##     TeamCLE     TeamCOL     TeamCWS     TeamDET
##   2830320.6  -1541323.2   1139552.2   -555593.5
```

```
cat("R-Squared:", model3r2, "\nAdjusted R-Squared:", model3adjustedr2)
```

```
## R-Squared: 0.641368
## Adjusted R-Squared: 0.597742
```

ANOVA Analysis of Model 3

```
salary3.aov <- aov(Salary ~ Team * Years_Length + POS, salary)
summary(salary3.aov)
```

```
##                     Df    Sum Sq   Mean Sq F value   Pr(>F)
## Team                29 1.962e+15 6.767e+13   3.696 8.52e-10 ***
## Years_Length         1 1.399e+16 1.399e+16 764.234  < 2e-16 ***
## POS                  9 8.188e+14 9.097e+13   4.968 1.91e-06 ***
## Team:Years_Length   29 1.530e+15 5.277e+13   2.882 1.27e-06 ***
## Residuals          559 1.024e+16 1.831e+13
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
summaryanova <- summary(salary3.aov)[[1]]["Df"]
df <- summary(salary3.aov)[[1]]["F value"]
colnames(df) <- gsub(" ", ".", colnames(df))

df$FCrit <- with(summaryanova, qf(c(0.05),Df,559,lower.tail=FALSE))
df$IsSignificant <- with(df, ifelse(F.value > FCrit, 1, 0))
df
```

```
##                    F.value  FCrit IsSignificant
## Team                  3.70 1.4879             1
## Years_Length        764.23 3.8581             1
## POS                   4.97 1.8966             1
## Team:Years_Length     2.88 1.4879             1
## Residuals                  1.1494
```

Using alpha=0.05, all of the main effects and interactions are significant.

Model with interaction between POS and Years_Length (MODEL 4)

```
salary4.lm <- lm(Salary ~ Team + Years_Length * POS, salary)
model4r2 <- summary(salary4.lm)$r.squared
model4adjustedr2 <- summary(salary4.lm)$adj.r.squared

# Printing out the first 10 coefficients and r-squared metrics
coef(salary4.lm)[0:10]
```

```
## (Intercept)     TeamATL     TeamBAL     TeamBOS     TeamCHC     TeamCIN
##    437074.7  -1631698.7    437395.4   1588329.1    179413.7  -2386122.1
##     TeamCLE     TeamCOL     TeamCWS     TeamDET
##   -934227.2    766913.4   -828010.6   -679425.6
```

```
cat("R-Squared:", model4r2, "\nAdjusted R-Squared:", model4adjustedr2)
```

```
## R-Squared: 0.6262689
## Adjusted R-Squared: 0.5959838
```

ANOVA Analysis of Model 4

```
salary4.aov <- aov(Salary ~ Team + Years_Length * POS, salary)
summary(salary4.aov)
```

```
##                    Df    Sum Sq   Mean Sq F value   Pr(>F)
## Team               29 1.962e+15 6.767e+13   3.680 9.11e-10 ***
## Years_Length        1 1.399e+16 1.399e+16 760.908  < 2e-16 ***
## POS                 9 8.188e+14 9.097e+13   4.947 2.02e-06 ***
## Years_Length:POS    8 1.099e+15 1.374e+14   7.472 1.66e-09 ***
## Residuals         580 1.067e+16 1.839e+13
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
summaryanova <- summary(salary4.aov)[[1]]["Df"]
df <- summary(salary4.aov)[[1]]["F value"]
colnames(df) <- gsub(" ", ".", colnames(df))

df$FCrit <- with(summaryanova, qf(c(0.05),Df,580,lower.tail=FALSE))
df$IsSignificant <- with(df, ifelse(F.value > FCrit, 1, 0))
df
```

```
##                   F.value  FCrit IsSignificant
## Team                 3.68 1.4871             1
## Years_Length       760.91 3.8575             1
## POS                  4.95 1.8960             1
## Years_Length:POS     7.47 1.9544             1
## Residuals                 1.1465
```

Using alpha=0.05, all of the main effects and interactions are significant.

Model with interactions between Team and POS, Team and Years_Length (MODEL 5)

```
salary5.lm <- lm(Salary ~ Team * Years_Length + Team * POS, salary)
model5r2 <- summary(salary5.lm)$r.squared
model5adjustedr2 <- summary(salary5.lm)$adj.r.squared

# Printing out the first 10 coefficients and r-squared metrics
coef(salary5.lm)[0:10]
```

```
## (Intercept)     TeamATL     TeamBAL     TeamBOS     TeamCHC     TeamCIN
##    -3751571     7085938     2288649     8052445   -17208027     6962419
##     TeamCLE     TeamCOL     TeamCWS     TeamDET
##     7685680     1753559    11825143   -19287679
```

```
cat("R-Squared:", model5r2, "\nAdjusted R-Squared:", model5adjustedr2)
```

```
## R-Squared: 0.7907075
## Adjusted R-Squared: 0.6404756
```

ANOVA Analysis of Model 5

```
salary5.aov <- aov(Salary ~ Team * Years_Length + Team * POS, salary)
summary(salary5.aov)
```

```
##                     Df    Sum Sq   Mean Sq F value   Pr(>F)
## Team                29 1.962e+15 6.767e+13   4.135 6.18e-11 ***
## Years_Length         1 1.399e+16 1.399e+16 855.072  < 2e-16 ***
## POS                  9 8.188e+14 9.097e+13   5.559 3.41e-07 ***
## Team:Years_Length   29 1.530e+15 5.277e+13   3.224 1.36e-07 ***
## Team:POS           194 4.262e+15 2.197e+13   1.342  0.00851 **
## Residuals          365 5.973e+15 1.636e+13
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
summaryanova <- summary(salary5.aov)[[1]]["Df"]
df <- summary(salary5.aov)[[1]]["F value"]
colnames(df) <- gsub(" ", ".", colnames(df))

df$FCrit <- with(summaryanova, qf(c(0.05),Df,365,lower.tail=FALSE))
df$IsSignificant <- with(df, ifelse(F.value > FCrit, 1, 0))
df
```

```
##                   F.value  FCrit IsSignificant
## Team                 4.14 1.4987             1
## Years_Length       855.07 3.8671             1
## POS                  5.56 1.9056             1
## Team:Years_Length    3.22 1.4987             1
## Team:POS             1.34 1.2254             1
## Residuals                 1.1882
```

Using alpha=0.05, all of the main effects and interactions are significant.

Model with interactions between Team and POS, POS and Years_Length (MODEL 6)

```
salary6.lm <- lm(Salary ~ Team * POS + POS * Years_Length, salary)
model6r2 <- summary(salary6.lm)$r.squared
model6adjustedr2 <- summary(salary6.lm)$adj.r.squared
```

```r
# Printing out the first 10 coefficients and r-squared metrics
coef(salary6.lm)[0:10]
```

```
## (Intercept)     TeamATL     TeamBAL     TeamBOS     TeamCHC     TeamCIN
##    -1967894     1339980     1706476     8083333   -10602034    -1397309
##     TeamCLE     TeamCOL     TeamCWS     TeamDET
##     3794854     4731833    12131374     3602691
```

```r
cat("R-Squared:", model6r2, "\nAdjusted R-Squared:", model6adjustedr2)
```

```
## R-Squared: 0.7584971
## Adjusted R-Squared: 0.6097363
```

ANOVA Analysis of Model 6

```r
salary6.aov <- aov(Salary ~ Team * POS + POS * Years_Length, salary)
summary(salary6.aov)
```

```
##                   Df    Sum Sq   Mean Sq F value   Pr(>F)
## Team              29 1.962e+15 6.767e+13   3.809 8.14e-10 ***
## POS                9 3.193e+15 3.548e+14  19.973  < 2e-16 ***
## Years_Length       1 1.162e+16 1.162e+16 654.055  < 2e-16 ***
## Team:POS         194 4.531e+15 2.336e+13   1.315  0.01243 *
## POS:Years_Length   6 3.419e+14 5.699e+13   3.208  0.00437 **
## Residuals        388 6.892e+15 1.776e+13
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
summaryanova <- summary(salary6.aov)[[1]]["Df"]
df <- summary(salary6.aov)[[1]]["F value"]
colnames(df) <- gsub(" ", ".", colnames(df))

df$FCrit <- with(summaryanova, qf(c(0.05),Df,388,lower.tail=FALSE))
df$IsSignificant <- with(df, ifelse(F.value > FCrit, 1, 0))
df
```

```
##                  F.value  FCrit IsSignificant
## Team                3.81 1.4968             1
## POS                19.97 1.9040             1
## Years_Length      654.06 3.8655             1
## Team:POS            1.31 1.2225             1
## POS:Years_Length    3.21 2.1220             1
## Residuals                1.1820
```

Using alpha=0.05, all of the main effects and interactions are significant.

Model with interactions between Team and Years_Length, POS and Years_Length (MODEL 7)

```
salary7.lm <- lm(Salary ~ Team * Years_Length + POS * Years_Length, salary)
model7r2 <- summary(salary7.lm)$r.squared
model7adjustedr2 <- summary(salary7.lm)$adj.r.squared

# Printing out the first 10 coefficients and r-squared metrics
coef(salary7.lm)[0:10]
```

```
## (Intercept)     TeamATL     TeamBAL     TeamBOS     TeamCHC     TeamCIN
##   -489867.1    262843.2    980122.6    692698.5    312504.1   -530829.3
##     TeamCLE     TeamCOL     TeamCWS     TeamDET
##   2434797.5  -1720381.6    140141.7   -778122.5
```

```
cat("R-Squared:", model7r2, "\nAdjusted R-Squared:", model7adjustedr2)
```

```
## R-Squared: 0.6813752
## Adjusted R-Squared: 0.6374269
```

ANOVA Analysis of Model 7

```
salary7.aov <- aov(Salary ~ Team * Years_Length + POS * Years_Length, salary)
summary(salary7.aov)
```

```
##                     Df    Sum Sq   Mean Sq F value   Pr(>F)
## Team                29 1.962e+15 6.767e+13   4.100 2.10e-11 ***
## Years_Length         1 1.399e+16 1.399e+16 847.882  < 2e-16 ***
## POS                  9 8.188e+14 9.097e+13   5.512 2.78e-07 ***
## Team:Years_Length   29 1.530e+15 5.277e+13   3.197 8.02e-08 ***
## Years_Length:POS     8 1.142e+15 1.427e+14   8.648 3.80e-11 ***
## Residuals          551 9.093e+15 1.650e+13
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
summaryanova <- summary(salary7.aov)[[1]]["Df"]
df <- summary(salary7.aov)[[1]]["F value"]
colnames(df) <- gsub(" ", ".", colnames(df))

df$FCrit <- with(summaryanova, qf(c(0.05),Df,551,lower.tail=FALSE))
df$IsSignificant <- with(df, ifelse(F.value > FCrit, 1, 0))
df
```

```
##                   F.value  FCrit IsSignificant
## Team                 4.10 1.4882             1
## Years_Length       847.88 3.8584             1
## POS                  5.51 1.8969             1
## Team:Years_Length    3.20 1.4882             1
## Years_Length:POS     8.65 1.9552             1
## Residuals                 1.1506
```

Using alpha=0.05, all of the main effects and interactions are significant.

Model with all three 2-way interactions (MODEL 8)

```
salary8.lm <- lm(Salary ~ Team * Years_Length + Team * POS + POS * Years_Length, salary)
model8r2 <- summary(salary8.lm)$r.squared
model8adjustedr2 <- summary(salary8.lm)$adj.r.squared

# Printing out the first 10 coefficients and r-squared metrics
coef(salary8.lm)[0:10]
```

```
## (Intercept)     TeamATL     TeamBAL     TeamBOS     TeamCHC     TeamCIN
##    -3642549     7319114     1265031     5750247   -19345669     7516342
##     TeamCLE     TeamCOL     TeamCWS     TeamDET
##     8080056     1035075    11899449   -14106547
```

```
cat("R-Squared:", model8r2, "\nAdjusted R-Squared:", model8adjustedr2)
```

```
## R-Squared: 0.8041679
## Adjusted R-Squared: 0.6579757
```

ANOVA Analysis of Model 8

```
salary8.aov <- aov(Salary ~ Team * Years_Length + Team * POS + POS * Years_Length, salary)
summary(salary8.aov)
```

```
##                    Df    Sum Sq   Mean Sq F value   Pr(>F)
## Team               29 1.962e+15 6.767e+13   4.347 1.09e-11 ***
## Years_Length        1 1.399e+16 1.399e+16 898.823  < 2e-16 ***
## POS                 9 8.188e+14 9.097e+13   5.843 1.31e-07 ***
## Team:Years_Length  29 1.530e+15 5.277e+13   3.389 3.57e-08 ***
## Team:POS          194 4.262e+15 2.197e+13   1.411 0.002683 **
## Years_Length:POS    6 3.842e+14 6.403e+13   4.113 0.000523 ***
## Residuals         359 5.589e+15 1.557e+13
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
summaryanova <- summary(salary8.aov)[[1]]["Df"]
df <- summary(salary8.aov)[[1]]["F value"]
colnames(df) <- gsub(" ", ".", colnames(df))

df$FCrit <- with(summaryanova, qf(c(0.05),Df,359,lower.tail=FALSE))
df$IsSignificant <- with(df, ifelse(F.value > FCrit, 1, 0))
df
```

```
##                   F.value  FCrit IsSignificant
## Team                 4.35 1.4992             1
## Years_Length       898.82 3.8675             1
## POS                  5.84 1.9060             1
## Team:Years_Length    3.39 1.4992             1
## Team:POS             1.41 1.2263             1
## Years_Length:POS     4.11 2.1239             1
## Residuals                 1.1899
```

Using alpha=0.05, all of the main effects and interactions are significant.

Basic Model (Model with no interaction)

```
basic.lm <- lm(Salary ~ Team + Years_Length + POS, salary)
basicr2 <- summary(basic.lm)$r.squared
basicmodeladjustedr2 <- summary(basic.lm)$adj.r.squared

# Printing out the first 10 coefficients and r-squared metrics
coef(basic.lm)[0:10]
```

```
## (Intercept)     TeamATL     TeamBAL     TeamBOS     TeamCHC     TeamCIN
##   2262781.8  -1380793.3    292592.4   1558287.4    585835.9  -2540558.9
##     TeamCLE     TeamCOL     TeamCWS     TeamDET
##  -1107425.8    633490.6  -1212937.8   -410530.2
```

```
cat("R-Squared:", basicr2, "\nAdjusted R-Squared:", basicmodeladjustedr2)
```

```
## R-Squared: 0.587751
## Adjusted R-Squared: 0.560408
```

ANOVA Analysis of Basic Model

```
salarybasic.aov <- aov(Salary ~ Team + Years_Length + POS, salary)
summary(salarybasic.aov)
```

```
##                Df    Sum Sq   Mean Sq F value   Pr(>F)
## Team           29 1.962e+15 6.767e+13   3.382 1.36e-08 ***
## Years_Length    1 1.399e+16 1.399e+16 699.329  < 2e-16 ***
## POS             9 8.188e+14 9.097e+13   4.546 8.29e-06 ***
## Residuals     588 1.177e+16 2.001e+13
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
summaryanova <- summary(salarybasic.aov)[[1]]["Df"]
df <- summary(salarybasic.aov)[[1]]["F value"]
colnames(df) <- gsub(" ", ".", colnames(df))

df$FCrit <- with(summaryanova, qf(c(0.05),Df,588,lower.tail=FALSE))
df$IsSignificant <- with(df, ifelse(F.value > FCrit, 1, 0))
df
```

```
##               F.value  FCrit IsSignificant
## Team             3.38 1.4869             1
## Years_Length   699.33 3.8573             1
## POS              4.55 1.8958             1
## Residuals             1.1454
```

Using alpha=0.05, all of the main effects are significant.

Model Selection

Creating a dataframe to compare model performance

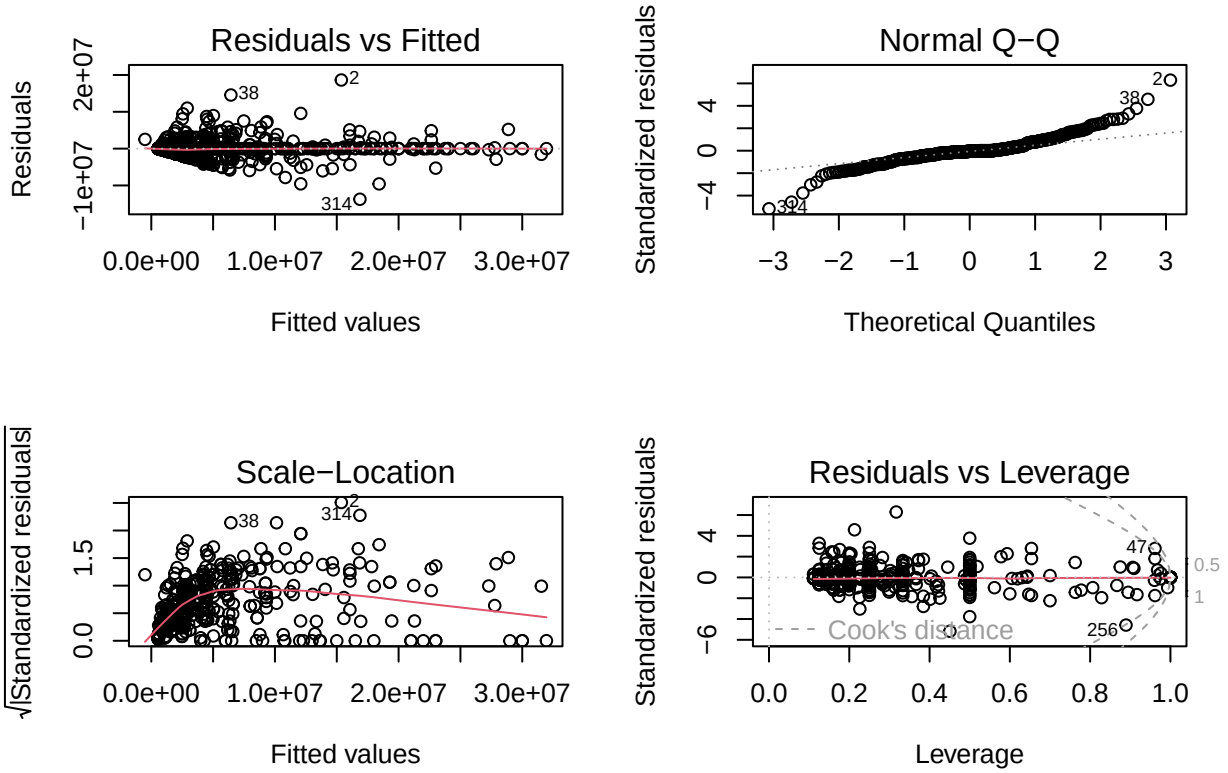
```
Model.Type <- c("Full Model", "Model 2", "Model 3", "Model 4",
                "Model 5", "Model 6", "Model 7", "Model 8", "Basic Model")
AdjustedRSquared <- c(fullmodeladjustedr2, model2adjustedr2, model3adjustedr2,
                      model4adjustedr2, model5adjustedr2, model6adjustedr2,
                      model7adjustedr2, model8adjustedr2, basicmodeladjustedr2)
AdjustedRSquared <- round(AdjustedRSquared, digits = 4)

comparisondf <- data.frame(Model.Type, AdjustedRSquared)
comparisondf <- comparisondf[order(-AdjustedRSquared),]
comparisondf
```

```
##    Model.Type AdjustedRSquared
## 1  Full Model           0.7180
## 8     Model 8           0.6580
## 5     Model 5           0.6405
## 7     Model 7           0.6374
## 6     Model 6           0.6097
## 3     Model 3           0.5977
## 2     Model 2           0.5966
## 4     Model 4           0.5960
## 9 Basic Model           0.5604
```

The Full Model is the best model as it produces the greatest adjusted r-squared value.

Model Adequacy Checking for Selected Model

```
library(ggplot2)
opar <- par(mfrow=c(2,2),cex=.8)
plot(salaryfull.aov)
```



There are no indications of model inadequecy.

Checking Factor Importance

```
noPOS.lm <- lm(Salary ~ Team * Years_Length, salary)
noPOSr2 <- summary(noPOS.lm)$r.squared
noPOSadjustedr2 <- summary(noPOS.lm)$adj.r.squared

cat("R-Squared:", noPOSr2, "\nAdjusted R-Squared:", noPOSadjustedr2)
```

```
## R-Squared: 0.6149549
## Adjusted R-Squared: 0.574959
```

```
noTeam.lm <- lm(Salary ~ POS * Years_Length, salary)
noTeamr2 <- summary(noTeam.lm)$r.squared
noTeamadjustedr2 <- summary(noTeam.lm)$adj.r.squared

cat("R-Squared:", noTeamr2, "\nAdjusted R-Squared:", noTeamadjustedr2)
```

```
## R-Squared: 0.5871265
## Adjusted R-Squared: 0.5749233
```

```r
noYears.lm <- lm(Salary ~ Team * POS, salary)
noYearsr2 <- summary(noYears.lm)$r.squared
noYearsadjustedr2 <- summary(noYears.lm)$adj.r.squared

cat("R-Squared:", noYearsr2, "\nAdjusted R-Squared:", noYearsadjustedr2)
```

```
## R-Squared: 0.4631009
## Adjusted R-Squared: 0.1477577
```

Models omitting on factor produce worse adjusted r-squared values than the full model: factors are important.

Conclusion

The best model to predict salary with the predictors Team, POS, and Years_Length is the full model. It produces the best adjusted r-squared value with all of its main effects and interactions significant. Furthermore, I've confirmed that all factors in my full model are important by comparing the adjusted r-squared values of my full model with separate models which omitted one variable.

The full model can be defined by:
$Salary = \beta_0 + YearsLength + Team + POS + YearsLength * Team + YearsLength * POS + Team * POS + YearsLength * Team * POS$