Linear Model

```r
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```r
df <- read.csv("training_df.csv")
df$homeTeam <- factor(df$homeTeam)
df$awayTeam <- factor(df$awayTeam)
head(df)
```

```
##   X index            homeTeam              awayTeam  price ticket_count
## 1 0     0 Los Angeles Dodgers        Texas Rangers  58.62            0
## 2 1    22 Los Angeles Dodgers       Houston Astros 110.02            0
## 3 2    21 Los Angeles Dodgers     Colorado Rockies  33.92            0
## 4 3    20 Los Angeles Dodgers     Colorado Rockies  38.83            0
## 5 4    19 Los Angeles Dodgers     Colorado Rockies  32.69            0
## 6 5    17 Los Angeles Dodgers San Francisco Giants  33.92            0
##   days_to_game
## 1            3
## 2           56
## 3           47
## 4           46
## 5           45
## 6           43
```

```r
model <- lm(price~ homeTeam + awayTeam + days_to_game, data = df)
model
```

```
##
## Call:
## lm(formula = price ~ homeTeam + awayTeam + days_to_game, data = df)
##
## Coefficients:
##               (Intercept)    homeTeamLos Angeles Dodgers
##                  -7.49100                       30.44315
```

```
##   homeTeamSan Francisco Giants         awayTeamAtlanta Braves
##                     23.22931                       12.14109
##      awayTeamBaltimore Orioles          awayTeamBoston Red Sox
##                     24.10661                       28.56885
##           awayTeamChicago Cubs       awayTeamColorado Rockies
##                     17.93451                        9.65903
##         awayTeamDetroit Tigers          awayTeamHouston Astros
##                     26.14418                       23.64203
##     awayTeamLos Angeles Angels    awayTeamLos Angeles Dodgers
##                     25.83376                       44.73083
##       awayTeamMilwaukee Brewers         awayTeamNew York Mets
##                      3.40345                        6.97358
##        awayTeamNew York Yankees      awayTeamOakland Athletics
##                     35.08037                       22.86542
## awayTeamPhiladelphia Phillies     awayTeamPittsburgh Pirates
##                     26.68442                       17.40879
##        awayTeamSan Diego Padres   awayTeamSan Francisco Giants
##                     17.34711                       13.73749
##         awayTeamSeattle Mariners    awayTeamSt. Louis Cardinals
##                     19.28820                       13.14638
##          awayTeamTexas Rangers        awayTeamToronto Blue Jays
##                     20.24148                       24.11333
##   awayTeamWashington Nationals                   days_to_game
##                     26.66362                        0.08501
```

summary(model)

```
##
## Call:
## lm(formula = price ~ homeTeam + awayTeam + days_to_game, data = df)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -28.327  -5.840  -1.585   4.079 124.431
##
## Coefficients:
##                               Estimate Std. Error t value Pr(>|t|)
## (Intercept)                  -7.491003   0.435427 -17.204   <2e-16 ***
## homeTeamLos Angeles Dodgers  30.443148   0.341329  89.190   <2e-16 ***
## homeTeamSan Francisco Giants 23.229315   0.337459  68.836   <2e-16 ***
## awayTeamAtlanta Braves       12.141087   0.407107  29.823   <2e-16 ***
## awayTeamBaltimore Orioles    24.106614   0.968206  24.898   <2e-16 ***
## awayTeamBoston Red Sox       28.568849   0.924849  30.890   <2e-16 ***
## awayTeamChicago Cubs         17.934508   0.948891  18.900   <2e-16 ***
## awayTeamColorado Rockies      9.659035   0.374775  25.773   <2e-16 ***
## awayTeamDetroit Tigers       26.144181   1.239931  21.085   <2e-16 ***
## awayTeamHouston Astros       23.642026   0.421167  56.135   <2e-16 ***
## awayTeamLos Angeles Angels   25.833757   0.624221  41.386   <2e-16 ***
## awayTeamLos Angeles Dodgers  44.730827   0.456765  97.930   <2e-16 ***
## awayTeamMilwaukee Brewers     3.403453   0.383656   8.871   <2e-16 ***
## awayTeamNew York Mets         6.973582   0.420203  16.596   <2e-16 ***
## awayTeamNew York Yankees     35.080373   0.627695  55.888   <2e-16 ***
## awayTeamOakland Athletics    22.865419   0.499502  45.776   <2e-16 ***
## awayTeamPhiladelphia Phillies 26.684425   1.062796  25.108   <2e-16 ***
```

```
## awayTeamPittsburgh Pirates    17.408794   0.476794   36.512    <2e-16 ***
## awayTeamSan Diego Padres       17.347113   0.320457   54.132    <2e-16 ***
## awayTeamSan Francisco Giants   13.737488   0.561932   24.447    <2e-16 ***
## awayTeamSeattle Mariners       19.288198   0.543546   35.486    <2e-16 ***
## awayTeamSt. Louis Cardinals    13.146379   0.869461   15.120    <2e-16 ***
## awayTeamTexas Rangers          20.241476   0.563062   35.949    <2e-16 ***
## awayTeamToronto Blue Jays      24.113333   0.679801   35.471    <2e-16 ***
## awayTeamWashington Nationals   26.663620   0.816957   32.638    <2e-16 ***
## days_to_game                    0.085007   0.002992   28.408    <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 
## Residual standard error: 10.66 on 18804 degrees of freedom
## Multiple R-squared:  0.5908, Adjusted R-squared:  0.5903
## F-statistic:  1086 on 25 and 18804 DF,  p-value: < 2.2e-16
```

Predictions

```r
predict_df <- data.frame("homeTeam" = "Los Angeles Dodgers",
                         "awayTeam" = "New York Yankees",
                         # "ticket_count" = 2000,
                         "days_to_game" = 0)
predict(model, predict_df)
```

```
##        1
## 58.03252
```

```r
predict_df <- data.frame("homeTeam" = "Los Angeles Dodgers",
                         "awayTeam" = "New York Yankees",
                         # "ticket_count" = 2000,
                         "days_to_game" = 70)
predict(model, predict_df)
```

```
##        1
## 63.98304
```