304 final project

Abstract


```
   The various of sports have become indispensable part of human’s life. Especially t
here are millions of people pay attention to footballs and basketballs. There are hug
e potentials and profits on them. According to report, some of footballs such as Mess
i has earned billions of money. This report use different methods analysis the factor
s which will influence on the wages of football players.
```


Introduction

```
   I find the wages of superstars can not only depend on their potentials, ages. Ther
efore in this report I pick a group of dataset on Kaggle about the football player on
2019 which include their name, age, wages and different kinds of abilities. I want to
analysis which factors impact on the wages of football players
     In my report, I prefer to make a Multiple Linear regression model to verify whic
h variables have a good prediction to the play’s wages. Just one dataset will be appl
ied on my test. However, there are so many variables on this dataset. Of course, I de
cide to pick some of them by the common sense firstly. Obviously, the variables like
ID are not related to wages at all. After that, according to the variables that I pic
ked I will make the conclusion of my report.
```

```
data <- read.csv("/Users/yuweisun/Desktop/data.csv")
data$Wage <- as.numeric(data$Wage)
data$Value <- as.numeric(data$Value)
```

```
model1 <- lm(Wage ~  Age+Potential+Value+Crossing+Finishing+HeadingAccuracy+Volleys,
data = data)
summary(model1)
```

```
##
## Call:
## lm(formula = Wage ~ Age + Potential + Value + Crossing + Finishing +
##     HeadingAccuracy + Volleys, data = data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -75.099 -23.537  -1.904  21.699  96.725
##
## Coefficients:
##                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)     -21.594294   3.739650  -5.774 7.85e-09 ***
## Age               1.050291   0.058186  18.050  < 2e-16 ***
## Potential         0.619727   0.045105  13.740  < 2e-16 ***
## Value             0.019684   0.003587   5.487 4.15e-08 ***
## Crossing          0.060588   0.019599   3.091 0.001995 **
## Finishing         0.037143   0.027988   1.327 0.184488
## HeadingAccuracy   0.058943   0.017295   3.408 0.000656 ***
## Volleys          -0.002343   0.032699  -0.072 0.942871
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 33.95 on 18151 degrees of freedom
##   (48 observations deleted due to missingness)
## Multiple R-squared:  0.03761,    Adjusted R-squared:  0.03724
## F-statistic: 101.3 on 7 and 18151 DF,  p-value: < 2.2e-16
```

```
model2 <- lm(Wage ~  Age+Potential+Value+Crossing+HeadingAccuracy, data = data)
summary(model2)
```

```
##
## Call:
## lm(formula = Wage ~ Age + Potential + Value + Crossing + HeadingAccuracy,
##     data = data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -74.657 -23.581  -2.012  21.683  96.495
##
## Coefficients:
##                   Estimate Std. Error t value Pr(>|t|)
## (Intercept)     -21.842070   3.708468  -5.890 3.94e-09 ***
## Age               1.047115   0.057478  18.218  < 2e-16 ***
## Potential         0.626590   0.044787  13.991  < 2e-16 ***
## Value             0.019650   0.003588   5.477 4.38e-08 ***
## Crossing          0.080832   0.015918   5.078 3.85e-07 ***
## HeadingAccuracy   0.067093   0.016657   4.028 5.65e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 33.95 on 18153 degrees of freedom
##   (48 observations deleted due to missingness)
## Multiple R-squared:  0.0374, Adjusted R-squared:  0.03713
## F-statistic: 141.1 on 5 and 18153 DF,  p-value: < 2.2e-16
```

data

The Kaggle FIFA 19 complete player dataset will use on this report, which include the information of all the FIFA football players. There are 18207 observation in my target population. I will create a regression model based on, the age, the value, release clause

Result

```
    According to the common sense, We pick several variables include Age, Value, pot
ential, Crossing, Finishing, Heading Accuracy, Volleys and we get the model equation,
beta0 equal to
```

-21.594294 beta1 equal to 1.050291, beta2=0.619727, beta3=0.019684, beta4=0.060588, beta5= 0.037143, beta6=0.058943, beta7=-0.002343. However, when I get the summary table of them, I find the p-value of finishing and Volley is large than 0.05, Therefore, they are not significant and we need to delete them As a result, we create a new model which only exist Age, Value, potential, Crossing, Heading Accuracy and beta0 equal to -21.594294 beta1 equal to 1.050291, beta2=0.619727, beta3=0.019684, beta4= 0.037143, beta5=0.058943.

Discussion

```
     In our report, we are interested in what will impact on the wage of football pl
ayer’s. After we get the results. Obviously, the young people will be easy to get a h
igher wage, because they have more potentials. In addition, the players who have the
higher ability to get the score include the Crossing speed, Heading Accuracy will hav
e more wage, because the winning of the match is depend on the score of the games.
```

Reference


```
     Gadiya, K. (2018, December 21). FIFA 19 complete player dataset. Retrieved Decem
ber 21, 2020, from https://www.kaggle.com/karangadiya/fifa19
```