FinalProj720.R

poisonaidee

2024-05-14

### Set Working Directory  
setwd("/Users/poisonaidee/Downloads/Marketing Analytics")  
  
### Read in Data Set  
Spotify <-read.csv("updated\_spotify\_review.csv")  
  
### Change to Factor  
Spotify$Predicted.Sentiment <- as.factor(Spotify$Predicted.Sentiment)  
  
Pos\_Likes <- Spotify$Total\_thumbsup[Spotify$Predicted.Sentiment=='Positive']  
Neg\_Likes <- Spotify$Total\_thumbsup[Spotify$Predicted.Sentiment=='Negative']  
  
Rating\_PosSentiment <- Spotify$Rating[Spotify$Predicted.Sentiment=='Positive']  
Rating\_NegSentiment <- Spotify$Rating[Spotify$Predicted.Sentiment=='Negative']  
  
t\_test <- t.test(Pos\_Likes,Neg\_Likes)  
print(t\_test)

##   
## Welch Two Sample t-test  
##   
## data: Pos\_Likes and Neg\_Likes  
## t = -8.8706, df = 58873, p-value < 2.2e-16  
## alternative hypothesis: true difference in means is not equal to 0  
## 95 percent confidence interval:  
## -7.465067 -4.763180  
## sample estimates:  
## mean of x mean of y   
## 4.200356 10.314480

t\_test1 <- t.test(Rating\_PosSentiment,Rating\_NegSentiment)  
print(t\_test1)

##   
## Welch Two Sample t-test  
##   
## data: Rating\_PosSentiment and Rating\_NegSentiment  
## t = 228.67, df = 60070, p-value < 2.2e-16  
## alternative hypothesis: true difference in means is not equal to 0  
## 95 percent confidence interval:  
## 2.239781 2.278509  
## sample estimates:  
## mean of x mean of y   
## 4.470639 2.211494

### ANOVA  
  
anova\_model <- aov(Rating ~ Predicted.Sentiment + Total\_thumbsup, data = Spotify)  
summary(anova\_model)

## Df Sum Sq Mean Sq F value Pr(>F)   
## Predicted.Sentiment 1 74988 74988 48149.46 <2e-16 \*\*\*  
## Total\_thumbsup 1 0 0 0.29 0.59   
## Residuals 60383 94040 2   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

anova\_model1 <- aov(Rating ~ Predicted.Sentiment + Predicted.Sentiment\*Total\_thumbsup, data = Spotify)  
summary(anova\_model1)

## Df Sum Sq Mean Sq F value Pr(>F)   
## Predicted.Sentiment 1 74988 74988 48164.94 < 2e-16 \*\*\*  
## Total\_thumbsup 1 0 0 0.29 0.59   
## Predicted.Sentiment:Total\_thumbsup 1 32 32 20.41 6.27e-06 \*\*\*  
## Residuals 60382 94008 2   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Spotify$Predicted.Sentiment <- relevel(Spotify$Predicted.Sentiment, ref = "Positive")  
summary(linear\_model <- lm(Rating ~ Predicted.Sentiment + Total\_thumbsup, data = Spotify))

##   
## Call:  
## lm(formula = Rating ~ Predicted.Sentiment + Total\_thumbsup, data = Spotify)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -3.4708 -1.2118 0.5292 0.5293 2.9333   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 4.471e+00 7.856e-03 569.061 <2e-16 \*\*\*  
## Predicted.SentimentNegative -2.259e+00 1.030e-02 -219.289 <2e-16 \*\*\*  
## Total\_thumbsup -3.037e-05 5.636e-05 -0.539 0.59   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 1.248 on 60383 degrees of freedom  
## Multiple R-squared: 0.4436, Adjusted R-squared: 0.4436   
## F-statistic: 2.407e+04 on 2 and 60383 DF, p-value: < 2.2e-16

# Convert Predicted.Sentiment to a factor if not already done  
Spotify$Predicted.Sentiment <- as.factor(Spotify$Predicted.Sentiment)  
sentiment\_table <- table(Spotify$Predicted.Sentiment)  
print(sentiment\_table)

##   
## Positive Negative   
## 25255 35131

# Perform the chi-squared test  
chi\_squared\_test <- chisq.test(sentiment\_table)  
  
# Display the results of the chi-squared test  
print(chi\_squared\_test)

##   
## Chi-squared test for given probabilities  
##   
## data: sentiment\_table  
## X-squared = 1615.2, df = 1, p-value < 2.2e-16

summary(linear\_model <- lm(Rating ~ Predicted.Sentiment + Predicted.Sentiment\*Total\_thumbsup, data = Spotify))

##   
## Call:  
## lm(formula = Rating ~ Predicted.Sentiment + Predicted.Sentiment \*   
## Total\_thumbsup, data = Spotify)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -3.4728 -1.2105 0.5272 0.5277 2.7895   
##   
## Coefficients:  
## Estimate Std. Error t value  
## (Intercept) 4.4728433 0.0078686 568.445  
## Predicted.SentimentNegative -2.2623883 0.0103276 -219.063  
## Total\_thumbsup -0.0005247 0.0001231 -4.263  
## Predicted.SentimentNegative:Total\_thumbsup 0.0006254 0.0001384 4.517  
## Pr(>|t|)   
## (Intercept) < 2e-16 \*\*\*  
## Predicted.SentimentNegative < 2e-16 \*\*\*  
## Total\_thumbsup 2.02e-05 \*\*\*  
## Predicted.SentimentNegative:Total\_thumbsup 6.27e-06 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 1.248 on 60382 degrees of freedom  
## Multiple R-squared: 0.4438, Adjusted R-squared: 0.4438   
## F-statistic: 1.606e+04 on 3 and 60382 DF, p-value: < 2.2e-16