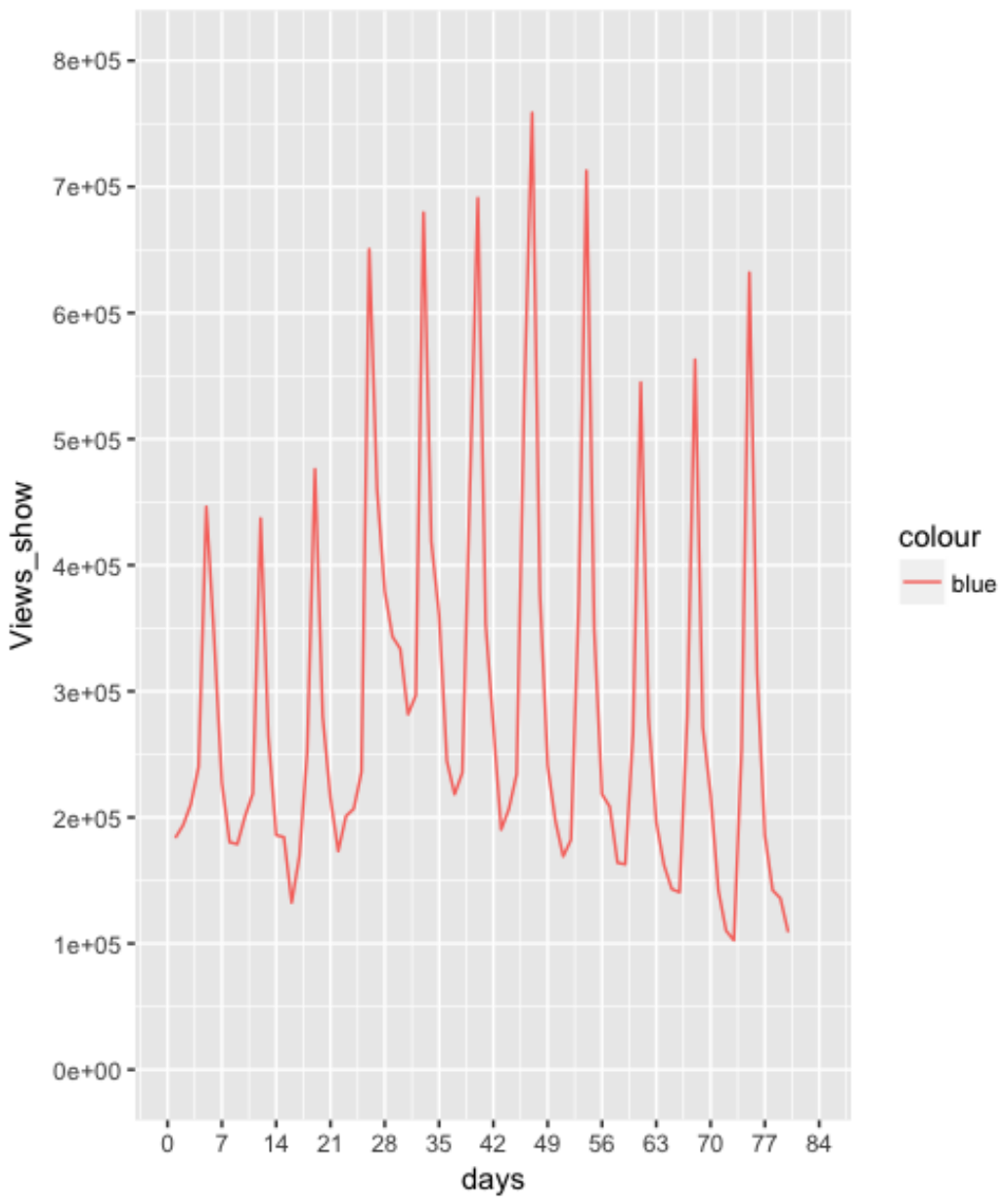
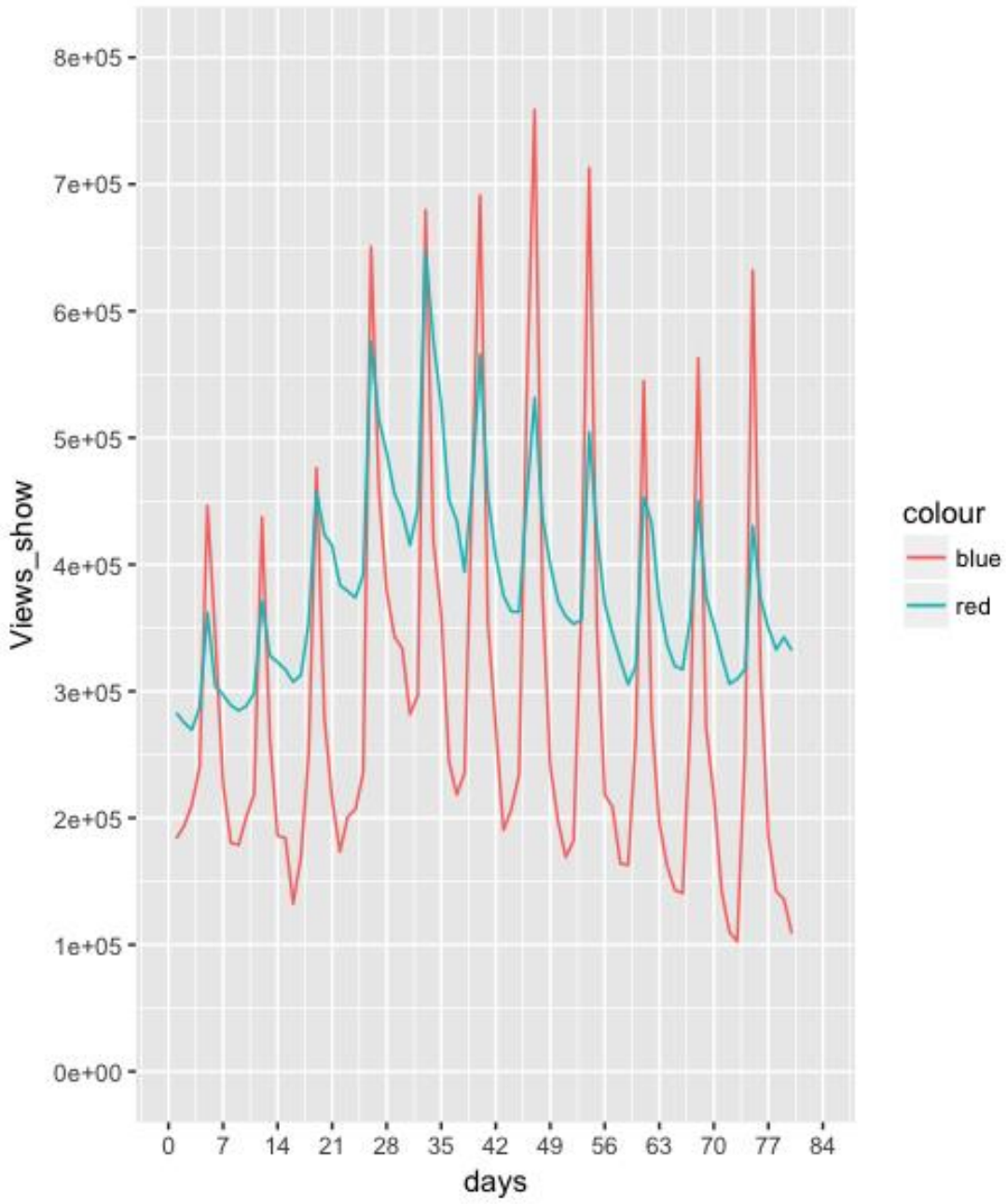
```
# Load essential libraries
library(ggplot2)
library(lubridate)


# load the media company data
media <- read.csv("mediacompany.csv", stringsAsFactors = F)

#checking the structure of the dataset
str(media)
#'data.frame': 80 obs. of  8 variables:
# $ Date             : chr  "3/1/2017" "3/2/2017" "3/3/2017" "3/4/2017" ...
#$ Views_show        : int  183738 193763 210479 240061 446314 342997 227313 180313 178800 201180 ...
#$ Visitors          : int  1260228 1270561 1248183 1492913 1594712 1330689 1363579 1353483 1345428 1344540 ...
#$ Views_platform    : int  1706478 1690727 1726157 1855353 2041418 1742540 1821742 1795456 1794157 1834967 ...
#$ Ad_impression     : num  1.06e+09 1.03e+09 1.01e+09 1.08e+09 1.36e+09 ...
#$ Cricket_match_india: int  0 0 0 1 0 0 0 0 0 0 ...
#$ Character_A       : int  0 0 0 0 0 0 0 0 0 0 ...
#$ X                : logi  NA NA NA NA NA NA ...

# convert the date to a standart date format
media$date <- as.Date(media$Date, "%m/%d/%Y")

# Deriving "days since the show started"
media$day<- difftime(media$date,as.Date("2017-02-28"), units="days")
media$day <- as.numeric(media$day)

#plot1
# Scatter Plot (days vs Views_show)
ggplot(media, aes(day, Views_show)) + geom_line(aes(colour = "blue" ))
+ scale_x_continuous(name = "days", breaks = seq(0,84,7), limits = c(0,84)) +
  scale_y_continuous(name = "Views_show", breaks = seq(0,800000,100000),
                     limits = c(0,800000))
# There is a spike in number of viewiers in the weekend than that of weekdays
```

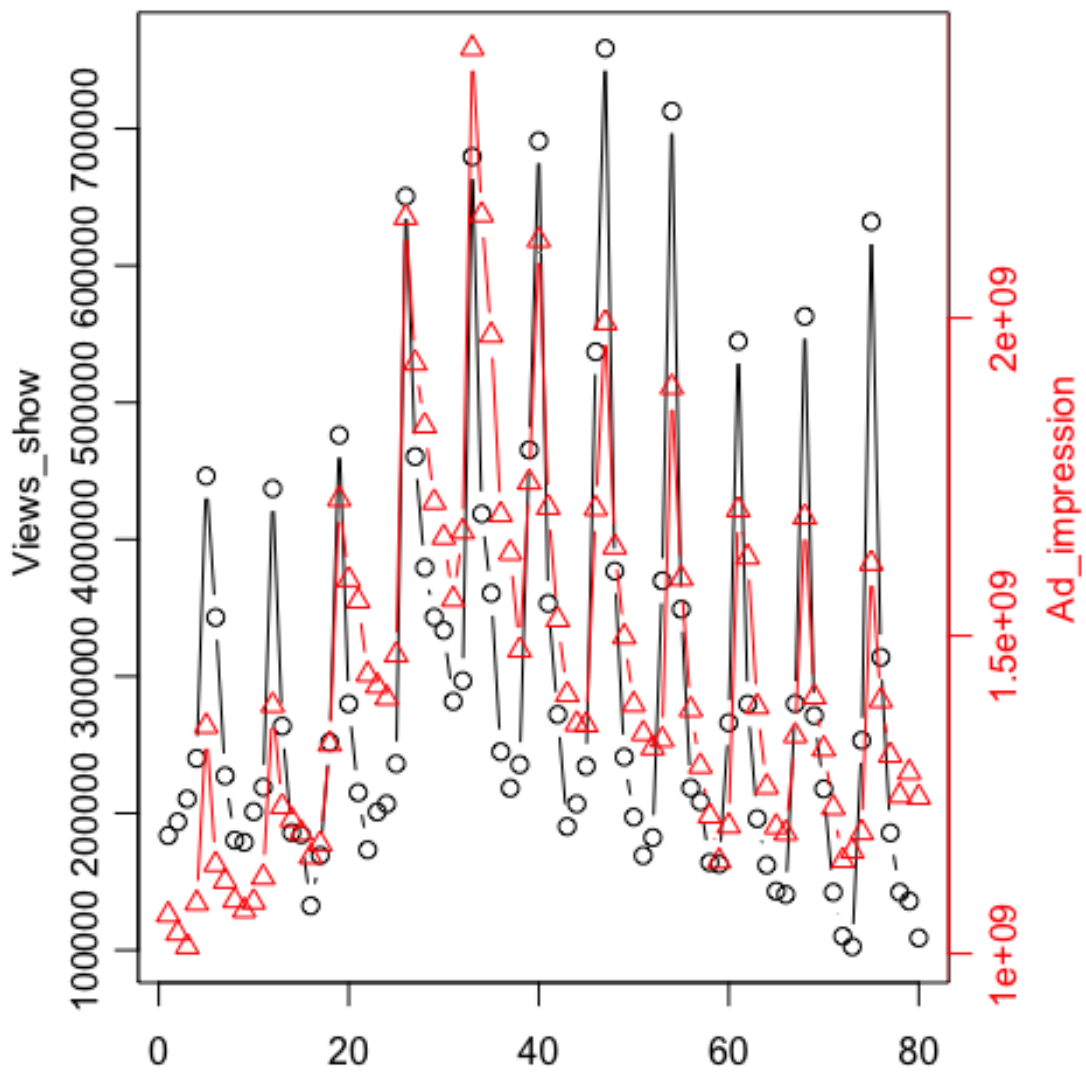


#Plot2

```
# Scatter Plot (days vs Views_show and days vs Ad_impressions)
# bringing the Ad_impression to the scale of Views_show. Think how this could be done.
ggplot(media, aes(day, Views_show)) + geom_line(aes(colour = "blue" )) +
  scale_x_continuous(name = "days", breaks = seq(0,84,7), limits = c(0,84))
+ scale_y_continuous(name = "Views_show", breaks = seq(0,800000,100000),limits = c(0,800000))
+ geom_line(aes(x=day, y=Ad_impression*8/30000, colour="red"))
# Although much better graphs could have been produced in Tableu, it serves the purpose of visualising variables
#Ad_impessions are highly correlated with the views_show
```



```
# Alternate- Using a library to generate a plot with 2 Y axis
#install.packages("plotrix")
#library(plotrix)
#plot3
twoord.plot( lx=media$day,ly=media$Views_show,ry=media$Ad_impression,
          rx = media$day, ylab="Views_show",rylab="Ad_impression")
# You could try and find more ways to do the same
```



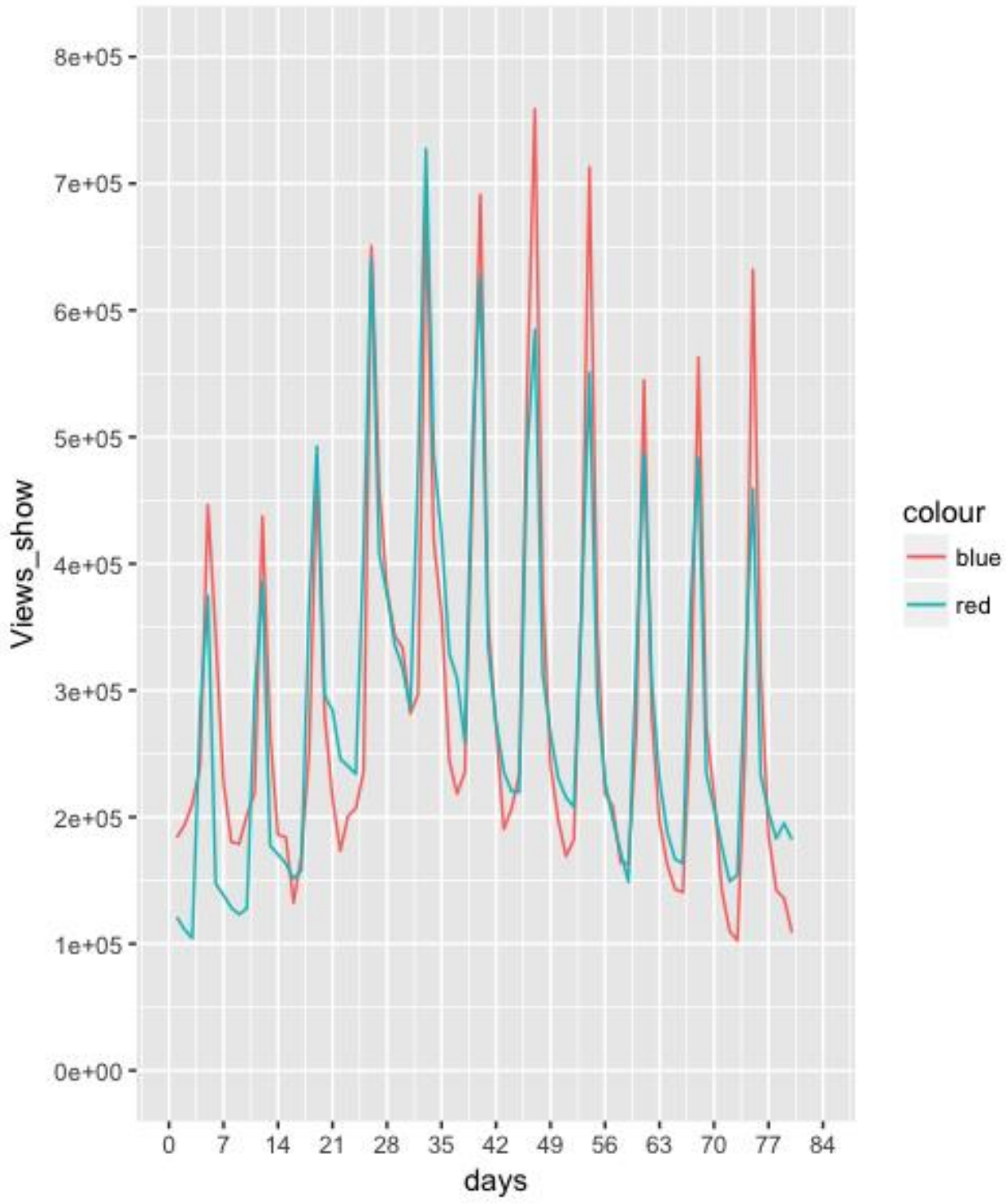

```
# Derived Metrics
# Weekdays are taken such that 1 corresponds to Sunday and 7 to Saturday
# Generate the weekday variable
media$Weekday <- (media$day+3)%%7
media$Weekday[which(media$Weekday==0)] <- 7
#View(media)
```

```
#model 1.1 Y vs Weekday & visitors
model_1.1 <- lm(formula = Views_show~Visitors + Weekday, data = media)
summary(model_1.1)
#Coefficients:
#  Estimate Std. Error t value Pr(>|t|)
#(Intercept) -3.862e+04  1.072e+05  -0.360     0.72
#Visitors     2.787e-01  5.675e-02   4.911 4.97e-06 ***
#  Weekday     -3.591e+04  6.591e+03  -5.448 5.90e-07 ***
  ---
#  Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

#Residual standard error: 112600 on 77 degrees of freedom
#Multiple R-squared:  0.485,  Adjusted R-squared:  0.4717
#F-statistic: 36.26 on 2 and 77 DF,  p-value: 8.009e-12


# create Weekend variable, with value 1 at weekends and 0 at weekdays

media$Weekend <- ifelse((media$day%%7==5)|(media$day%%7==4),1,0)
```

```
#model 1.2 Y vs Weekend & visitors
model_1.2 <- lm(formula = Views_show~Visitors + Weekend, data = media)
summary(model_1.2)
#Coefficients:
#  Estimate Std. Error t value Pr(>|t|)
#(Intercept) -8.833e+04  1.010e+05  -0.875  0.38438
#Visitors     1.934e-01  6.120e-02   3.160  0.00226 **
#  Weekend      1.807e+05  3.148e+04   5.740 1.79e-07 ***
  ---
#  Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

#Residual standard error: 110900 on 77 degrees of freedom
#Multiple R-squared:  0.5003,           Adjusted R-squared:  0.4873
#F-statistic: 38.55 on 2 and 77 DF,  p-value: 2.511e-12


#model 2 Y vs Weekend, Character A & visitors
model_2 <- lm(formula = Views_show~Visitors + Weekend + Character_A, data = media)
summary(model_2)
#Coefficients:
#  Estimate Std. Error t value Pr(>|t|)
#(Intercept) -4.722e+04  9.309e+04  -0.507 0.613450
#Visitors     1.480e-01  5.723e-02   2.586 0.011613 *
#  Weekend      1.812e+05  2.885e+04   6.281 1.92e-08 ***
#  Character_A  9.542e+04  2.408e+04   3.963 0.000166 ***
  ---
#  Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

#Residual standard error: 101600 on 76 degrees of freedom
#Multiple R-squared:  0.5859,           Adjusted R-squared:  0.5695
#F-statistic: 35.84 on 3 and 76 DF,  p-value: 1.533e-14



# Create lag variable
media$Lag_Views <- c(0, media$Views_show[seq_along(media$Views_show) -1])
View(media)


#Model 3 Y vs Weekend, Character A, LagViews & visitors
model_3 <- lm(formula = Views_show~Visitors + Lag_Views + Weekend +
              Character_A, data = media)
summary(model_3)
#Coefficients:
#  Estimate Std. Error t value Pr(>|t|)
#(Intercept) -2.980e+04  7.425e+04  -0.401  0.68929
#Visitors     6.587e-02  4.725e-02   1.394  0.16743
#Lag_Views    4.317e-01  6.463e-02   6.679 3.69e-09 ***
#  Weekend      2.273e+05  2.401e+04   9.467 1.93e-14 ***
#  Character_A  5.527e+04  2.012e+04   2.748  0.00751 **
  ---
#  Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

#Residual standard error: 81020 on 75 degrees of freedom
#Multiple R-squared:  0.7403,           Adjusted R-squared:  0.7265
#F-statistic: 53.46 on 4 and 75 DF,  p-value: < 2.2e-16


#model 4 Y vs Weekend, Character A & Views_platform
model_4 <- lm(formula = Views_show~ Weekend + Character_A +
              Views_platform, data = media)
summary(model_4)

#Coefficients:
#  Estimate Std. Error t value Pr(>|t|)
#(Intercept)    -1.205e+05  9.971e+04  -1.208  0.23074
#Weekend         1.781e+05  2.779e+04   6.410 1.11e-08 ***
#  Character_A     7.062e+04  2.599e+04   2.717  0.00815 **
#  Views_platform  1.507e-01  4.780e-02   3.152  0.00232 **
  ---
```

```
#  Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

#Residual standard error: 99710 on 76 degrees of freedom
#Multiple R-squared:  0.6015,           Adjusted R-squared:  0.5858
#F-statistic: 38.24 on 3 and 76 DF,  p-value: 3.589e-15


#model 5 Y vs Weekend, Character A & Visitors
model_5 <- lm(formula = Views_show~ Weekend + Character_A +
              Visitors, data = media)
summary(model_5)
#Coefficients:
#  Estimate Std. Error t value Pr(>|t|)
#(Intercept) -4.722e+04  9.309e+04  -0.507 0.613450
#Weekend      1.812e+05  2.885e+04   6.281 1.92e-08 ***
#  Character_A  9.542e+04  2.408e+04   3.963 0.000166 ***
#  Visitors     1.480e-01  5.723e-02   2.586 0.011613 *
  ---
#  Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

#Residual standard error: 101600 on 76 degrees of freedom
#Multiple R-squared:  0.5859,           Adjusted R-squared:  0.5695
#F-statistic: 35.84 on 3 and 76 DF,  p-value: 1.533e-14


#model 6 Y vs Weekend, Character A, Visitors and Ad_Impression
model_6 <- lm(formula = Views_show~ Weekend + Character_A + Visitors +
              Ad_impression, data = media)
summary(model_6)
#Coefficients:
#  Estimate Std. Error t value Pr(>|t|)
#(Intercept)   -2.834e+05  6.967e+04  -4.067 0.000116 ***
#  Weekend       1.485e+05  2.035e+04   7.296 2.57e-10 ***
#  Character_A   -2.934e+04  2.163e+04  -1.356 0.179071
#Visitors      1.439e-02  4.238e-02   0.340 0.735104
#Ad_impression  3.598e-04  3.958e-05   9.090 1.01e-13 ***
#  ---
#  Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

#Residual standard error: 70580 on 75 degrees of freedom
#Multiple R-squared:  0.8029,           Adjusted R-squared:  0.7924
#F-statistic:  76.4 on 4 and 75 DF,  p-value: < 2.2e-16


#model 7 Y vs Weekend, Character A and Ad_Impression
model_7 <- lm(formula = Views_show~ Weekend + Character_A +
              Ad_impression, data = media)
summary(model_7)
#Coefficients:
#  Estimate Std. Error t value Pr(>|t|)
#(Intercept)   -2.661e+05  4.745e+04  -5.609 3.15e-07 ***
#  Weekend       1.510e+05  1.884e+04   8.019 1.01e-11 ***
#  Character_A   -2.990e+04  2.145e+04  -1.394    0.167
#Ad_impression  3.645e-04  3.691e-05   9.875 2.87e-15 ***
  ---
#  Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

#Residual standard error: 70170 on 76 degrees of freedom
#Multiple R-squared:  0.8026,           Adjusted R-squared:  0.7949
#F-statistic:   103 on 3 and 76 DF,  p-value: < 2.2e-16


#Ad impression in million
media$ad_impression_million <- media$Ad_impression/1000000


#model 8 Y vs Weekend, Character A, Ad_Impression and Cricket_match_india
model_8 <- lm(formula = Views_show~ Weekend + Character_A +
                ad_impression_million + Cricket_match_india, data = media)
summary(model_8)
```

```
#Estimate Std. Error t value Pr(>|t|)
#(Intercept)          -263272.33   48005.06  -5.484 5.37e-07 ***
#  Weekend              152110.50   19045.06   7.987 1.26e-11 ***
#  Character_A           -31963.29   21930.22  -1.457    0.149
#ad_impression_million     363.79      37.11   9.802 4.50e-15 ***
#  Cricket_match_india    -13959.14   27369.53  -0.510    0.612
#---
#  Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

#Residual standard error: 70510 on 75 degrees of freedom
#Multiple R-squared:  0.8033,           Adjusted R-squared:  0.7928
#F-statistic: 76.59 on 4 and 75 DF,  p-value: < 2.2e-16



#model 9 Y vs Weekend, Character A, Ad_Impression and Cricket_match_india
model_9 <- lm(formula = Views_show~ Weekend + ad_impression_million, data = media)
summary(model_9)

#Coefficients:
#  Estimate Std. Error t value Pr(>|t|)
#(Intercept)          -230169.8    40068.4  -5.744 1.75e-07 ***
#  Weekend              155082.5    18724.4   8.282 2.91e-12 ***
#  ad_impression_million     331.0       28.2  11.736  < 2e-16 ***
#  ---
#  Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

#Residual standard error: 70600 on 77 degrees of freedom
#Multiple R-squared:  0.7976,           Adjusted R-squared:  0.7923
#F-statistic: 151.7 on 2 and 77 DF,  p-value: < 2.2e-16



#predicted shows
media$Predicted_views <- predict(model_9, media)
media$error <-  media$Views_show - media$Predicted_views

#plot5
# Plot - Actual vs Predicted Views Model9
ggplot(media, aes(day, Views_show)) + geom_line(aes(colour = "blue" )) + scale_x_continuous(name = "days", breaks = seq(0,84,7), limits =
c(0,84)) + scale_y_continuous(name = "Views_show", breaks = seq(0,800000,100000), limits = c(0,800000)) + geom_line(aes(x=day,
y=Predicted_views, colour="red"))
```




```
# Plot Model9 errors
ggplot(media, aes(day, error)) + geom_point() + scale_x_continuous(name = "days", breaks = seq(0,90,10), limits = c(0,84)) +
scale_y_continuous(name = "Error", breaks = seq(-250000,250000,50000), limits = c(-250000,250000)) + geom_hline(yintercept = 0)


# Predicting values and calculating error
media$Predicted_views_model5 <- predict(model_5, media)
media$error_model5 <-  media$Views_show - media$Predicted_views_model5


# Actual vs predicted values from model5
ggplot(media, aes(day, Views_show)) + geom_line(aes(colour = "blue" )) +
  scale_x_continuous(name = "days", breaks = seq(0,84,7), limits = c(0,84)) +
  scale_y_continuous(name = "Views_show", breaks = seq(0,800000,100000), limits = c(0,800000)) + geom_line(aes(x=day,
y=Predicted_views_model5, colour="red"))
```

#plot9
Plot model5 errors
ggplot(media, aes(day, error_model5)) + geom_line() +
  scale_x_continuous(name = "days", breaks = seq(0,90,10), limits = c(0,84)) +
  scale_y_continuous(name = "Error", breaks = seq(-250000,250000,50000), limits = c(-250000,250000)) + geom_hline(yintercept = 0)

we can see that the error doesnot follow any patter, so we can say that we have the best model and the error is just a white noise and can goahed wit the model.