Unveiling Twitch Trends

Loka Adarsh Dronamraju

2023-11-27

R Markdown

This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see http://rmarkdown.rstudio.com (http://rmarkdown.rstudio.com).

When you click the **Knit** button a document will be generated that includes both content as well as the output of any embedded R code chunks within the document. You can embed an R code chunk like this:

```
data <- read_csv("C:\\Users\\dlada\\OneDrive\\Documents\\twitchdata-update.csv")
```

```
## Rows: 1000 Columns: 11
## ── Column specification ────────────────────────────────────────────────────────
## Delimiter: ","
## chr (2): Channel, Language
## dbl (7): Watch time(Minutes), Stream time(minutes), Peak viewers, Average vi...
## lgl (2): Partnered, Mature
##
## ℹ Use `spec()` to retrieve the full column specification for this data.
## ℹ Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
dim(data)
```

```
## [1] 1000   11
```

```
str(data)
```

```
## spc_tbl_ [1,000 × 11] (S3: spec_tbl_df/tbl_df/tbl/data.frame)
##  $ Channel             : chr [1:1000] "xQcOW" "summit1g" "Gaules" "ESL_CSGO" ...
##  $ Watch time(Minutes) : num [1:1000] 6196161750 6091677300 5644590915 3970318140 3671000070
...
##  $ Stream time(minutes): num [1:1000] 215250 211845 515280 517740 123660 ...
##  $ Peak viewers        : num [1:1000] 222720 310998 387315 300575 285644 ...
##  $ Average viewers     : num [1:1000] 27716 25610 10976 7714 29602 ...
##  $ Followers           : num [1:1000] 3246298 5310163 1767635 3944850 8938903 ...
##  $ Followers gained    : num [1:1000] 1734810 1370184 1023779 703986 2068424 ...
##  $ Views gained        : num [1:1000] 93036735 89705964 102611607 106546942 78998587 ...
##  $ Partnered           : logi [1:1000] TRUE TRUE TRUE TRUE TRUE TRUE ...
##  $ Mature              : logi [1:1000] FALSE FALSE TRUE FALSE FALSE FALSE ...
##  $ Language            : chr [1:1000] "English" "English" "Portuguese" "English" ...
##  - attr(*, "spec")=
##   .. cols(
##   ..   Channel = col_character(),
##   ..   `Watch time(Minutes)` = col_double(),
##   ..   `Stream time(minutes)` = col_double(),
##   ..   `Peak viewers` = col_double(),
##   ..   `Average viewers` = col_double(),
##   ..   Followers = col_double(),
##   ..   `Followers gained` = col_double(),
##   ..   `Views gained` = col_double(),
##   ..   Partnered = col_logical(),
##   ..   Mature = col_logical(),
##   ..   Language = col_character()
##   .. )
##  - attr(*, "problems")=<externalptr>
```

```
summary(data)
```

```
##    Channel          Watch time(Minutes)  Stream time(minutes)  Peak viewers
##  Length:1000        Min.   : 122192850   Min.   :  3465       Min.   :   496
##  Class :character   1st Qu.: 163189894   1st Qu.: 73759       1st Qu.:  9114
##  Mode  :character   Median : 234990788   Median :108240       Median : 16676
##                     Mean   : 418427930   Mean   :120515       Mean   : 37065
##                     3rd Qu.: 433739918   3rd Qu.:141844       3rd Qu.: 37570
##                     Max.   :6196161750   Max.   :521445       Max.   :639375
##  Average viewers    Followers       Followers gained   Views gained
##  Min.   :   235   Min.   :   3660   Min.   : -15772   Min.   :   175788
##  1st Qu.:  1458   1st Qu.: 170546   1st Qu.:  43758   1st Qu.:  3880602
##  Median :  2425   Median : 318063   Median :  98352   Median :  6456324
##  Mean   :  4781   Mean   : 570054   Mean   : 205519   Mean   : 11668166
##  3rd Qu.:  4786   3rd Qu.: 624332   3rd Qu.: 236131   3rd Qu.: 12196762
##  Max.   :147643   Max.   :8938903   Max.   :3966525   Max.   :670137548
##  Partnered         Mature          Language
##  Mode :logical   Mode :logical   Length:1000
##  FALSE:22        FALSE:770       Class :character
##  TRUE :978       TRUE :230       Mode  :character
##
##
##
```

```
head(data)
```

```
## # A tibble: 6 × 11
##   Channel   `Watch time(Minutes)` `Stream time(minutes)` `Peak viewers`
##   <chr>                     <dbl>                  <dbl>          <dbl>
## 1 xQcOW                6196161750                 215250         222720
## 2 summit1g             6091677300                 211845         310998
## 3 Gaules               5644590915                 515280         387315
## 4 ESL_CSGO             3970318140                 517740         300575
## 5 Tfue                 3671000070                 123660         285644
## 6 Asmongold            3668799075                  82260         263720
## # ℹ 7 more variables: `Average viewers` <dbl>, Followers <dbl>,
## #   `Followers gained` <dbl>, `Views gained` <dbl>, Partnered <lgl>,
## #   Mature <lgl>, Language <chr>
```

```
sum_of_duplicates <- sum(duplicated(data))
print(sum_of_duplicates)
```

```
## [1] 0
```

```
sum(is.na(data))
```

```
## [1] 0
```

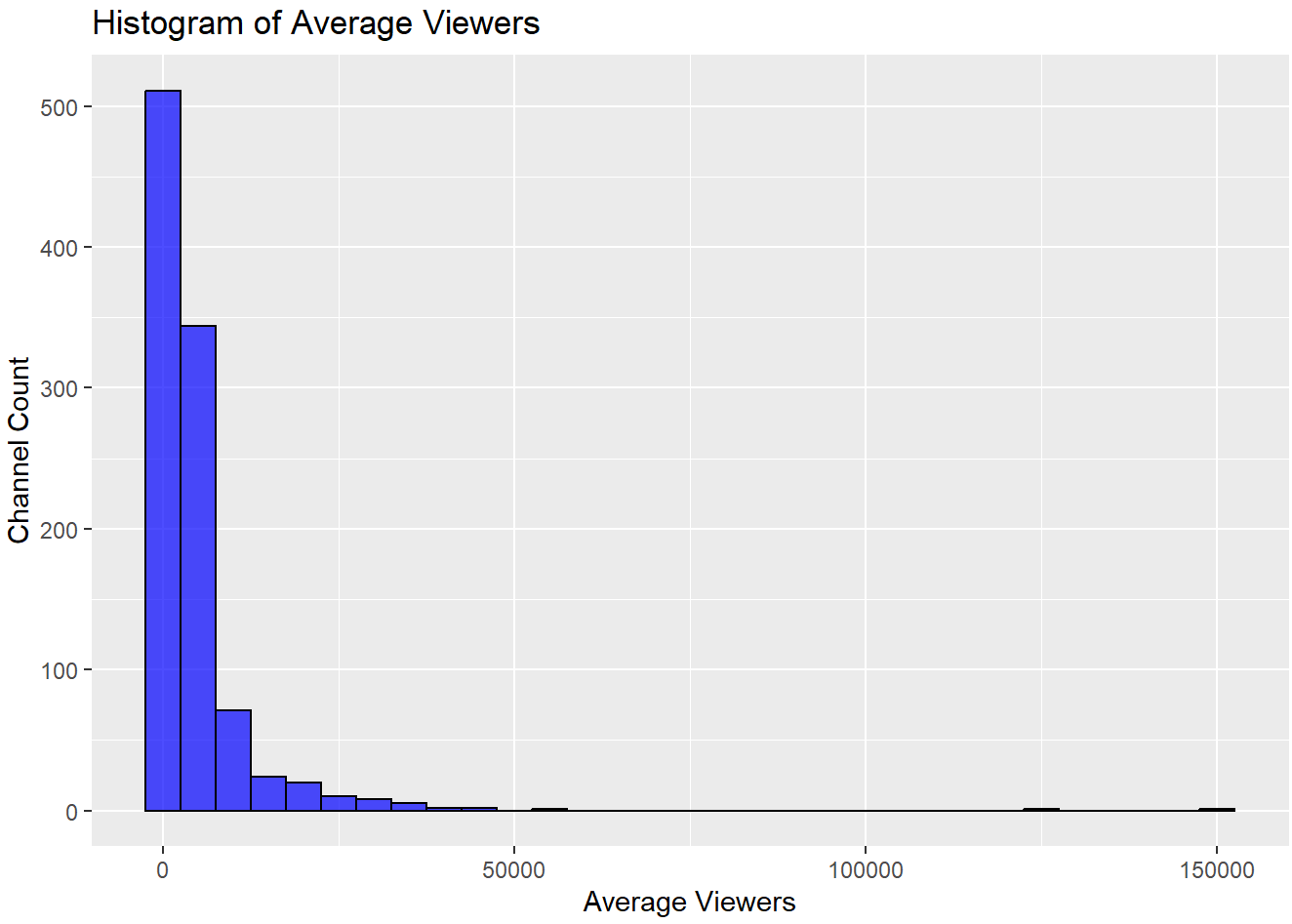
```
null_counts <- colSums(is.na(data))
null_counts
```

```
##              Channel  Watch time(Minutes) Stream time(minutes)
##                    0                    0                    0
##         Peak viewers      Average viewers            Followers
##                    0                    0                    0
##     Followers gained         Views gained            Partnered
##                    0                    0                    0
##               Mature             Language
##                    0                    0
```

EXPLORATORY DATA ANALYSIS

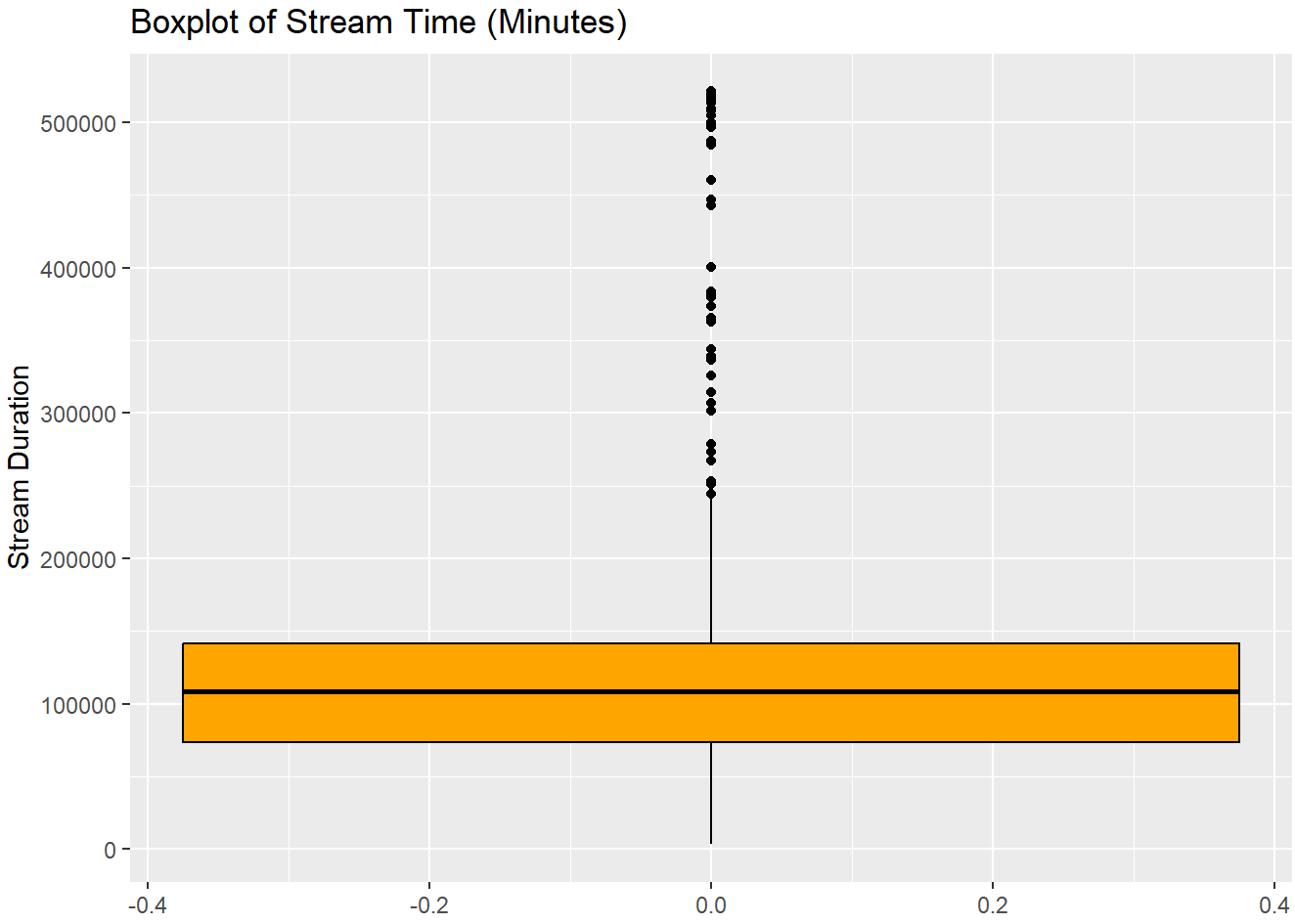
i. Histogram

```
ggplot(data, aes(x = `Average viewers`)) +
  geom_histogram(binwidth = 5000, fill = "blue", color = "black", alpha = 0.7) +
  labs(title = "Histogram of Average Viewers",
       x = "Average Viewers",
       y = "Channel Count")
```



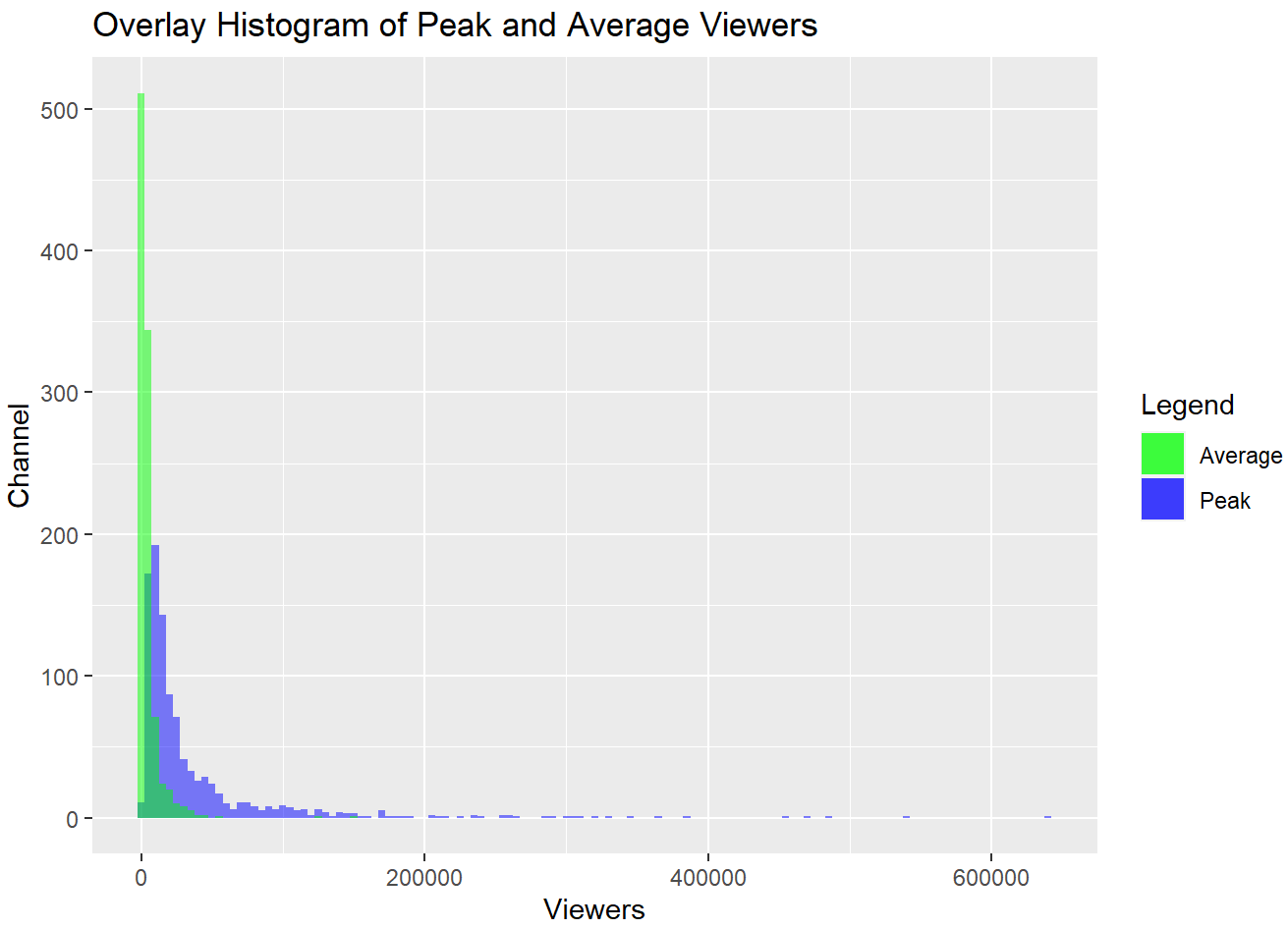
ii. Boxplot

```
ggplot(data, aes(y = `Stream time(minutes)`)) +
  geom_boxplot(fill = "orange", color = "black") +
  labs(title = "Boxplot of Stream Time (Minutes)",
       y = "Stream Duration")
```



iii. Overlay Histogram

```
ggplot(data, aes(x = `Peak viewers`, fill = "Peak")) +
  geom_histogram(binwidth = 5000, alpha = 0.5) +
  geom_histogram(aes(x = `Average viewers`, fill = "Average"), binwidth = 5000, alpha = 0.5) +
  labs(title = "Overlay Histogram of Peak and Average Viewers",
       x = "Viewers",
       y = "Channel",
       fill = "Legend") +
  scale_fill_manual(values = c("Peak" = "blue", "Average" = "green"))
```



#T-test

```
unique(data$Language)
```

```
##  [1] "English"    "Portuguese" "Spanish"    "German"     "Korean"
##  [6] "French"     "Russian"    "Japanese"   "Chinese"    "Czech"
## [11] "Turkish"    "Italian"    "Polish"     "Thai"       "Arabic"
## [16] "Slovak"     "Other"      "Hungarian"  "Greek"      "Finnish"
## [21] "Swedish"
```

```
english_data <- subset(data, Language == "English")
other_language_data <- subset(data, Language != "English")
str(english_data)
```

```
## tibble [485 × 11] (S3: tbl_df/tbl/data.frame)
##  $ Channel             : chr [1:485] "xQcOW" "summit1g" "ESL_CSGO" "Tfue" ...
##  $ Watch time(Minutes) : num [1:485] 6196161750 6091677300 3970318140 3671000070 3668799075
...
##  $ Stream time(minutes): num [1:485] 215250 211845 517740 123660 82260 ...
##  $ Peak viewers        : num [1:485] 222720 310998 300575 285644 263720 ...
##  $ Average viewers     : num [1:485] 27716 25610 7714 29602 42414 ...
##  $ Followers           : num [1:485] 3246298 5310163 3944850 8938903 1563438 ...
##  $ Followers gained    : num [1:485] 1734810 1370184 703986 2068424 554201 ...
##  $ Views gained        : num [1:485] 93036735 89705964 106546942 78998587 61715781 ...
##  $ Partnered           : logi [1:485] TRUE TRUE TRUE TRUE TRUE TRUE ...
##  $ Mature              : logi [1:485] FALSE FALSE FALSE FALSE FALSE FALSE ...
##  $ Language            : chr [1:485] "English" "English" "English" "English" ...
```

```
str(other_language_data)
```

```
## tibble [515 × 11] (S3: tbl_df/tbl/data.frame)
##  $ Channel             : chr [1:515] "Gaules" "Rubius" "auronplay" "MontanaBlack88" ...
##  $ Watch time(Minutes) : num [1:515] 5644590915 2588632635 2410022550 2408460990 2186662470
...
##  $ Stream time(minutes): num [1:515] 515280 58275 40575 67740 181230 ...
##  $ Peak viewers        : num [1:515] 387315 240096 170115 181600 26999 ...
##  $ Average viewers     : num [1:515] 10976 42948 53986 33514 12201 ...
##  $ Followers           : num [1:515] 1767635 5751354 3983847 2911316 494445 ...
##  $ Followers gained    : num [1:515] 1023779 3820532 3966525 1101093 92205 ...
##  $ Views gained        : num [1:515] 102611607 58599449 41514854 37189666 34405975 ...
##  $ Partnered           : logi [1:515] TRUE TRUE TRUE TRUE TRUE TRUE ...
##  $ Mature              : logi [1:515] TRUE FALSE FALSE TRUE FALSE FALSE ...
##  $ Language            : chr [1:515] "Portuguese" "Spanish" "Spanish" "German" ...
```

```
t_test_result <- t.test(english_data$'Followers gained', other_language_data$'Followers gained')
print(t_test_result)
```

```
##
##  Welch Two Sample t-test
##
## data:  english_data$"Followers gained" and other_language_data$"Followers gained"
## t = -0.20156, df = 991.54, p-value = 0.8403
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -46367.50  37729.77
## sample estimates:
## mean of x mean of y
##  203294.3  207613.2
```

Multiple Linear Regression

```
colnames(data)[colnames(data) == "Followers gained"] <- "Followers_gained"
colnames(data)[colnames(data) == "Average viewers"] <- "Average_Viewers"
```

```
set.seed(123)

indices <- sample(1:nrow(data), size = 0.8 * nrow(data))
train_data <- data[indices, ]
test_data <- data[-indices, ]

linear_model <- lm(Followers_gained ~ . - Channel, data = data)


summary(linear_model)
```

```
##
## Call:
## lm(formula = Followers_gained ~ . - Channel, data = data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1221321   -54191    -3012    51663  2284897
##
## Coefficients:
##                               Estimate       Std. Error t value
## (Intercept)             274542.16834908  108911.02919714   2.521
## `Watch time(Minutes)`        0.00008583       0.00002069   4.148
## `Stream time(minutes)`      -0.28496672       0.09133387  -3.120
## `Peak viewers`               0.02810984       0.17500627   0.161
## Average_Viewers              3.13532424       1.17241787   2.674
## Followers                    0.24106374       0.01186905  20.310
## `Views gained`               0.00004209       0.00032958   0.128
## PartneredTRUE           -17299.20281964   48503.96954586  -0.357
## MatureTRUE              -10610.96383803   16993.92125160  -0.624
## LanguageChinese        -271861.03039598  105503.03109044  -2.577
## LanguageCzech          -209690.03028155  131921.33808843  -1.590
## LanguageEnglish        -244167.37406756   97862.56697203  -2.495
## LanguageFinnish        -241528.28365114  238486.31705338  -1.013
## LanguageFrench         -186951.47610850  100972.00697397  -1.852
## LanguageGerman         -213223.32635225  102181.71310419  -2.087
## LanguageGreek          -217058.73542288  237713.64739120  -0.913
## LanguageHungarian      -239856.59194792  181693.77647239  -1.320
## LanguageItalian        -134278.41973639  110638.17678532  -1.214
## LanguageJapanese       -247445.21673065  119368.37083875  -2.073
## LanguageKorean         -243897.76218930  100558.42919467  -2.425
## LanguageOther          -218426.98009409  243415.25850703  -0.897
## LanguagePolish         -237819.04773617  115774.05370317  -2.054
## LanguagePortuguese     -148528.05889364  101433.60757643  -1.464
## LanguageRussian        -231510.70596097  100624.22313183  -2.301
## LanguageSlovak         -271311.61583593  237737.44720971  -1.141
## LanguageSpanish          92952.21754011  100551.23731161   0.924
## LanguageSwedish        -223000.48846495  238232.68363488  -0.936
## LanguageThai           -226323.54518412  117562.76948290  -1.925
## LanguageTurkish        -188686.11260178  107629.64847790  -1.753
##                                  Pr(>|t|)
## (Intercept)                       0.01187 *
## `Watch time(Minutes)`            0.0000366 ***
## `Stream time(minutes)`            0.00186 **
## `Peak viewers`                    0.87242
## Average_Viewers                   0.00762 **
## Followers               < 0.0000000000000002 ***
## `Views gained`                    0.89840
## PartneredTRUE                     0.72143
## MatureTRUE                        0.53251
## LanguageChinese                   0.01012 *
## LanguageCzech                     0.11227
## LanguageEnglish                   0.01276 *
```

```
## LanguageFinnish                    0.31143
## LanguageFrench                     0.06440 .
## LanguageGerman                     0.03717 *
## LanguageGreek                      0.36141
## LanguageHungarian                  0.18711
## LanguageItalian                    0.22517
## LanguageJapanese                   0.03844 *
## LanguageKorean                     0.01547 *
## LanguageOther                      0.36976
## LanguagePolish                     0.04023 *
## LanguagePortuguese                 0.14344
## LanguageRussian                    0.02162 *
## LanguageSlovak                     0.25406
## LanguageSpanish                    0.35549
## LanguageSwedish                    0.34947
## LanguageThai                       0.05451 .
## LanguageTurkish                    0.07990 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 216800 on 971 degrees of freedom
## Multiple R-squared:  0.6047, Adjusted R-squared:  0.5933
## F-statistic: 53.04 on 28 and 971 DF,  p-value: < 0.00000000000000022
```

```
  test_predictions_linear <- predict(linear_model, newdata = test_data)
  summary(test_predictions_linear)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## -108063   50768  119372  195519  238021 2054908
```

```
  test_actual_values <- test_data$Followers_gained
  summary(test_actual_values)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    1328   48286  106064  193720  253521 1734810
```

```
  mae_linear <- mean(abs(test_actual_values - test_predictions_linear))
  cat("Mean Absolute Error (Linear Regression):", mae_linear, "\n")
```

```
## Mean Absolute Error (Linear Regression): 102774.5
```

```
  mse_linear <- sqrt(mean((test_actual_values - test_predictions_linear)^2))
  cat("Mean Squared Error (Linear Regression):", mse_linear, "\n")
```

```
## Mean Squared Error (Linear Regression): 173098.1
```

#forward stepwise Regression

```
 set.seed(123)

  indices <- sample(1:nrow(data), size = 0.8 * nrow(data))
  train_data <- data[indices, ]
  test_data <- data[-indices, ]
init_model <- lm(Followers ~ 1, data = test_data)
all_model <- lm(Followers ~ . - Channel, data = test_data)
forwardmodel <- stepAIC(init_model, direction = "forward", scope = formula(all_model))
```

```
## Start:  AIC=5427.09
## Followers ~ 1
##
##                          Df      Sum of Sq             RSS    AIC
## + Followers_gained        1 54692306248715  65941919438779 5308.3
## + Average_Viewers         1 47618780133952  73015445553542 5328.7
## + `Peak viewers`          1 42355122078119  78279103609375 5342.6
## + `Watch time(Minutes)`   1 37949013344153  82685212343341 5353.5
## + `Views gained`          1 22934777970615  97699447716879 5386.9
## <none>                                     120634225687494 5427.1
## + Mature                  1   973928175412 119660297512081 5427.5
## + `Stream time(minutes)`  1   926592306535 119707633380959 5427.6
## + Partnered               1    64042647045 120570183040449 5429.0
## + Language               13  6271400710589 114362824976904 5442.4
##
## Step:  AIC=5308.3
## Followers ~ Followers_gained
##
##                          Df      Sum of Sq            RSS    AIC
## + Average_Viewers         1 10733385202770 55208534236009 5274.8
## + `Peak viewers`          1  9281691791299 56660227647480 5280.0
## + `Watch time(Minutes)`   1  2699245996164 63242673442615 5301.9
## + `Views gained`          1  2049395011980 63892524426799 5304.0
## + Partnered               1  1203562558686 64738356880093 5306.6
## <none>                                    65941919438779 5308.3
## + Mature                  1   582747095122 65359172343657 5308.5
## + `Stream time(minutes)`  1   183543903936 65758375534843 5309.7
## + Language               13  4441687739680 61500231699099 5320.3
##
## Step:  AIC=5274.76
## Followers ~ Followers_gained + Average_Viewers
##
##                          Df     Sum of Sq            RSS    AIC
## + `Peak viewers`          1 2817390857390 52391143378619 5266.3
## + Partnered               1  817320945701 54391213290309 5273.8
## <none>                                    55208534236009 5274.8
## + Mature                  1  214519701271 54994014534738 5276.0
## + `Stream time(minutes)`  1  144406119946 55064128116064 5276.2
## + `Watch time(Minutes)`   1   81070087904 55127464148105 5276.5
## + `Views gained`          1    6202455209 55202331780801 5276.7
## + Language               13 3196028664176 52012505571833 5288.8
##
## Step:  AIC=5266.29
## Followers ~ Followers_gained + Average_Viewers + `Peak viewers`
##
##                          Df     Sum of Sq            RSS    AIC
## + Partnered               1  620184147749 51770959230870 5265.9
## <none>                                    52391143378619 5266.3
## + Mature                  1  295139015293 52096004363326 5267.2
## + `Views gained`          1  181695467229 52209447911390 5267.6
## + `Watch time(Minutes)`   1   18546996148 52372596382471 5268.2
## + `Stream time(minutes)`  1     930622600 52390212756019 5268.3
```

```
## + Language               13 2734137628255 49657005750364 5281.6
##
## Step:  AIC=5265.91
## Followers ~ Followers_gained + Average_Viewers + `Peak viewers` +
##     Partnered
##
##                          Df     Sum of Sq            RSS    AIC
## <none>                                    51770959230870 5265.9
## + `Views gained`          1  238393208726 51532566022144 5267.0
## + Mature                  1  202787944301 51568171286569 5267.1
## + `Watch time(Minutes)`   1   30127684282 51740831546588 5267.8
## + `Stream time(minutes)`  1   10138219083 51760821011787 5267.9
## + Language               13 2664094893297 49106864337573 5281.3
```

```
summary(forwardmodel)
```

```
##
## Call:
## lm(formula = Followers ~ Followers_gained + Average_Viewers +
##     `Peak viewers` + Partnered, data = test_data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1771091  -129387   -23885    83516  4439199
##
## Coefficients:
##                      Estimate   Std. Error t value        Pr(>|t|)
## (Intercept)      -353528.1493  237441.6130  -1.489         0.13813
## Followers_gained       1.3840       0.1929   7.173 0.0000000000149 ***
## Average_Viewers       33.1210       8.3190   3.981 0.0000966507899 ***
## `Peak viewers`         2.4686       0.7858   3.142         0.00194 **
## PartneredTRUE     360516.2434  235879.5587   1.528         0.12804
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 515300 on 195 degrees of freedom
## Multiple R-squared:  0.5708, Adjusted R-squared:  0.562
## F-statistic: 64.84 on 4 and 195 DF,  p-value: < 0.00000000000000022
```

```
forwardmodel$anova
```

```
## Stepwise Model Path
## Analysis of Deviance Table
##
## Initial Model:
## Followers ~ 1
##
## Final Model:
## Followers ~ Followers_gained + Average_Viewers + `Peak viewers` +
##     Partnered
##
##
##                 Step Df       Deviance Resid. Df      Resid. Dev      AIC
## 1                                          199 120634225687494 5427.093
## 2 + Followers_gained  1 54692306248715       198  65941919438779 5308.296
## 3  + Average_Viewers  1 10733385202770       197  55208534236009 5274.764
## 4   + `Peak viewers`  1  2817390857390       196  52391143378619 5266.288
## 5        + Partnered  1   620184147749       195  51770959230870 5265.907
```

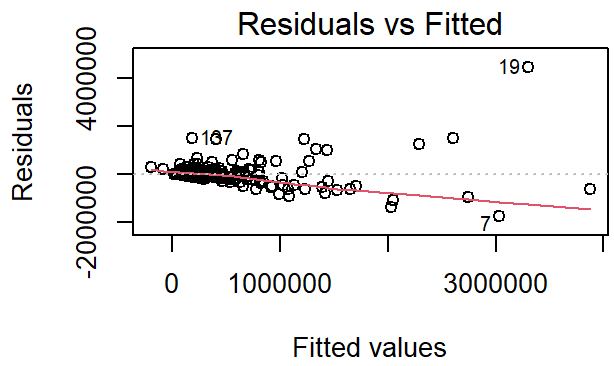
```
forward_pred <-predict(object = forwardmodel, newdata = test_data)
MAE(test_data$Followers_gained, forward_pred)
```

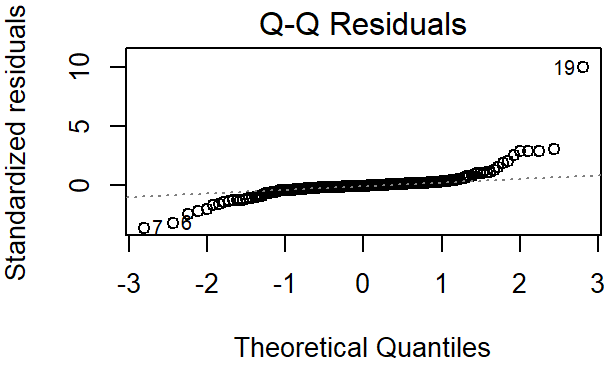
```
## [1] 324993.8
```

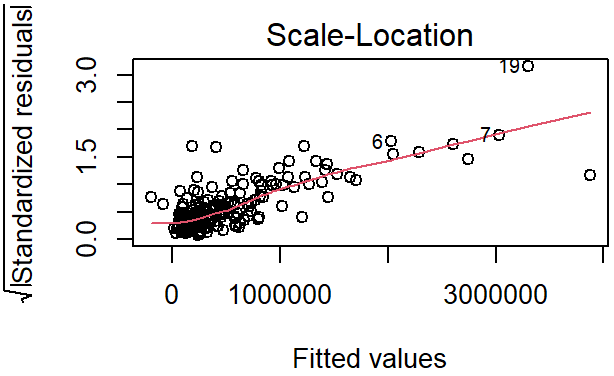
```
MSE(test_data$Followers_gained, forward_pred)
```

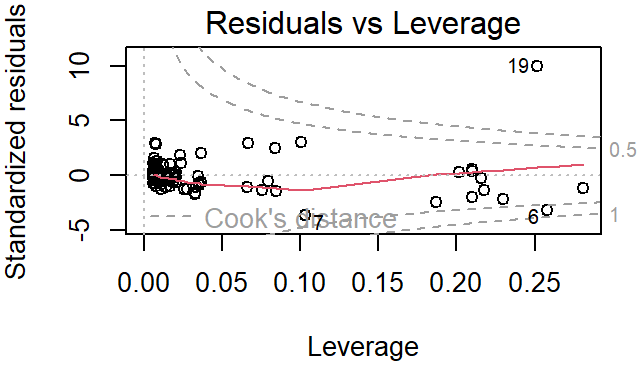
```
## [1] 252532314951
```

```
par(mfrow=c(2,2))
plot(forwardmodel)
```










```
backwardmodel <- stepAIC(all_model, direction = "backward")
```

```
## Start:  AIC=5285.46
## Followers ~ (Channel + `Watch time(Minutes)` + `Stream time(minutes)` +
##     `Peak viewers` + Average_Viewers + Followers_gained + `Views gained` +
##     Partnered + Mature + Language) - Channel
##
##                          Df     Sum of Sq            RSS    AIC
## - Language               13 3019247272142 51182837196118 5271.6
## - `Watch time(Minutes)`   1     230221259 48163820145236 5283.5
## - `Stream time(minutes)`  1    1204887264 48164794811240 5283.5
## - `Views gained`          1  161791312329 48325381236306 5284.1
## - Partnered               1  412349007692 48575938931669 5285.2
## <none>                                   48163589923977 5285.5
## - Mature                  1  770983163584 48934573087561 5286.6
## - `Peak viewers`          1 2463885428939 50627475352916 5293.4
## - Average_Viewers         1 2924643733900 51088233657876 5295.3
## - Followers_gained        1 9458773000578 57622362924554 5319.3
##
## Step:  AIC=5271.62
## Followers ~ `Watch time(Minutes)` + `Stream time(minutes)` +
##     `Peak viewers` + Average_Viewers + Followers_gained + `Views gained` +
##     Partnered + Mature
##
##                          Df      Sum of Sq            RSS    AIC
## - `Watch time(Minutes)`   1     5535620723 51188372816842 5269.6
## - `Stream time(minutes)`  1    55898723455 51238735919574 5269.8
## - Mature                  1   269636724833 51452473920952 5270.7
## - `Views gained`          1   323492410061 51506329606179 5270.9
## <none>                                   51182837196118 5271.6
## - Partnered               1   609833577345 51792670773463 5272.0
## - `Peak viewers`          1  2813027081362 53995864277480 5280.3
## - Average_Viewers         1  3532777022933 54715614219051 5283.0
## - Followers_gained        1 10773139788775 61955976984894 5307.8
##
## Step:  AIC=5269.64
## Followers ~ `Stream time(minutes)` + `Peak viewers` + Average_Viewers +
##     Followers_gained + `Views gained` + Partnered + Mature
##
##                          Df      Sum of Sq            RSS    AIC
## - `Stream time(minutes)`  1    82958528236 51271331345078 5268.0
## - Mature                  1   264217076378 51452589893220 5268.7
## - `Views gained`          1   372207386439 51560580203280 5269.1
## <none>                                   51188372816842 5269.6
## - Partnered               1   621047135950 51809419952791 5270.1
## - `Peak viewers`          1  2830687537815 54019060354657 5278.4
## - Average_Viewers         1  4056710432778 55245083249619 5282.9
## - Followers_gained        1 14013994255462 65202367072304 5316.0
##
## Step:  AIC=5267.97
## Followers ~ `Peak viewers` + Average_Viewers + Followers_gained +
##     `Views gained` + Partnered + Mature
##
##                    Df      Sum of Sq            RSS    AIC
```

```
## - Mature            1   261234677066 51532566022144 5267.0
## - `Views gained`    1   296839941491 51568171286569 5267.1
## <none>                               51271331345078 5268.0
## - Partnered         1   576315205650 51847646550728 5268.2
## - `Peak viewers`    1  2989673235995 54261004581073 5277.3
## - Average_Viewers   1  4275179746880 55546511091957 5282.0
## - Followers_gained  1 13943718729873 65215050074951 5314.1
##
## Step:  AIC=5266.98
## Followers ~ `Peak viewers` + Average_Viewers + Followers_gained +
##     `Views gained` + Partnered
##
##                    Df      Sum of Sq            RSS    AIC
## - `Views gained`    1   238393208726 51770959230870 5265.9
## <none>                               51532566022144 5267.0
## - Partnered         1   676881889246 52209447911390 5267.6
## - `Peak viewers`    1  2858609571596 54391175593740 5275.8
## - Average_Viewers   1  4442273126940 55974839149084 5281.5
## - Followers_gained  1 13887331478433 65419897500576 5312.7
##
## Step:  AIC=5265.91
## Followers ~ `Peak viewers` + Average_Viewers + Followers_gained +
##     Partnered
##
##                    Df      Sum of Sq            RSS    AIC
## <none>                               51770959230870 5265.9
## - Partnered         1   620184147749 52391143378619 5266.3
## - `Peak viewers`    1  2620254059439 54391213290309 5273.8
## - Average_Viewers   1  4208351915269 55979311146139 5279.5
## - Followers_gained  1 13661417440058 65432376670927 5310.7
```

```
summary(backwardmodel)
```

```
##
## Call:
## lm(formula = Followers ~ `Peak viewers` + Average_Viewers + Followers_gained +
##     Partnered, data = test_data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1771091  -129387   -23885    83516  4439199
##
## Coefficients:
##                      Estimate   Std. Error t value        Pr(>|t|)
## (Intercept)      -353528.1493  237441.6130  -1.489         0.13813
## `Peak viewers`          2.4686       0.7858   3.142         0.00194 **
## Average_Viewers        33.1210       8.3190   3.981 0.0000966507899 ***
## Followers_gained        1.3840       0.1929   7.173 0.0000000000149 ***
## PartneredTRUE      360516.2434  235879.5587   1.528         0.12804
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 515300 on 195 degrees of freedom
## Multiple R-squared:  0.5708, Adjusted R-squared:  0.562
## F-statistic: 64.84 on 4 and 195 DF,  p-value: < 0.00000000000000022
```

```
backwardmodel$anova
```

```
## Stepwise Model Path
## Analysis of Deviance Table
##
## Initial Model:
## Followers ~ (Channel + `Watch time(Minutes)` + `Stream time(minutes)` +
##     `Peak viewers` + Average_Viewers + Followers_gained + `Views gained` +
##     Partnered + Mature + Language) - Channel
##
## Final Model:
## Followers ~ `Peak viewers` + Average_Viewers + Followers_gained +
##     Partnered
##
##
##                       Step Df      Deviance Resid. Df     Resid. Dev      AIC
## 1                                                 178 48163589923977 5285.461
## 2               - Language 13 3019247272142       191 51182837196118 5271.622
## 3  - `Watch time(Minutes)`  1    5535620723       192 51188372816842 5269.643
## 4 - `Stream time(minutes)`  1   82958528236       193 51271331345078 5267.967
## 5                 - Mature  1  261234677066       194 51532566022144 5266.984
## 6         - `Views gained`  1  238393208726       195 51770959230870 5265.907
```

```
backward_pred <-predict(object = backwardmodel, newdata = test_data)
MAE(test_data$Followers_gained, backward_pred)
```

```
## [1] 324993.8
```

```
MSE(test_data$Followers_gained, backward_pred)
```

```
## [1] 252532314951
```

```
par(mfrow=c(2,2))
plot(backwardmodel)
```