IDS 702 Final Project: US TV Series Renewal

Jiayue (JY) Xu

12/10/2019

Summary

This study uses logistic regression to investigate the factors affecting the odds of renewal of US tv shows and whether or not the odds of tv show renewal differs across different network or genres of the series. It is found that seasons with more episodes, higher 18 to 49 demographic ratings, premiered on a different day from its regular timeslot, and from series with higher IMDB rating tend to have higher odds of renewal, while seasons with higher season numbers and from series with longer episode runtime tend to have lower odds of renewal. It is also found that the odds of renewal differ across network that the show is being broadcasted, but do not differs across different genres of the series. Lastly, it can be concluded that the effect of 18 to 49 demographic rating on tv show renewal rates differ across different tv networks.

1. Introduction

This study aims to analyze the factors affecting the odds of renewal of US tv shows using data obtained from various online tv show databases to see which shows were renewed and which shows were not renewed for all shows aired from the tv seasons 2013/14 to 2018/2019. In addition, the odds of tv show renewal across different networks and genres are analyzed. This study will also further explore how the tv show renewal rates differ across different tv network for different show ratings.

Section 2 describes the data sources, data transformation and the exploratory data analysis to understand the data based upon individual factors. In Section 3, the process of final model building is described, and the final model is presented. The evaluation of the final model and the statistical significance of individual parameters are discussed. Results that were obtained from the model selection are also summarized. Section 4 presents the conclusions that can be infered from the relationship and limitations of the model and the dataset.

2. Data

Data Preparation and Cleaning

The data used for this analysis consists 2 major categories: season and series and compiled in 2 stages. Firstly, season data were obtianed from spoilertv.com and tvseriesfinale.com by extraction of published tables on their websites and merged using Microsoft Excel. This includes 609 seasons across 295 series from 5 major tv networks: ABC, CBS, CW, FOX, NBC. There are no missing fields in the records. Dataset is partitioned into training and test set based on 80:20 split.

Season Features include: - Status: status of the show in the season (Renewed, Renewed/Ending, Ended, Cancelled, Cancelled/Revived) - Show: name of series - TVSeason: staring year of tv season (2013 to 2018) - Season: season number of the series in the season - Episode: number of episodes in the season - AirDay: day of the week that show is aired in the season, multiple entries if day changed - AirTime: timeslot that show is aired in the season, multiple entries if time changed - Debut.on.Different.Day: 1 for shows that premiered on a different day from its regular timeslot, 0 for otherwise - PremiereDay: date of premiere of show in the season - 18-49 key demo rating for every episode of the season (1 to 25)

Next, series data were obtained from IMBD.com via web-scrapping.

Series Features include: - Runtime: episode runtime of the series in mins - Certificate: rating category of the series - ReleaseYr: release year of the series - IMDBrating: IMDB rating of the series - Genre: IMDB genre of the series (string consists of 1 to 3 genre categories separated by comma)

Data Transformation

The process of data transformation include transforming existing variables as well as adding new features.

New Season Features include: - TVSeason2013: transform TVSeason to number of years from 2013, calculated from TVSeason - Season1: number of additional season from minimum 1 season, calculated from Season - FirstSeason: 1 for season with season number = 1, 0 for otherwise, calculated from Season - Episode1: number of episodes in the season from minimum 1 episode, calculated from TVSeason - AirDayInitial: day of the week that show is initially aired in the season, first entry from AirDay - AirTimeInitial: timslot that show is initially aired in the season, first entry from AirTime - AirDayTimeInitial: day and timeslot that show is initially aired in the season, calculated from AirDayInitial and AirDayTimeInitial - ChangeAirDayTime: 1 for shows that change air time or day in the season, 0 for otherwise, calculated from AirDay and AirTime - DaysfromTVSeasonPremiere: number of days premiere date is from the first day of tv season, calculated from PremiereDay - X18to49_Min: minimum rating across all episodes in the season - X18to49_YrtoYrChange: % change in average rating from previous season of the series - X18to49_MinDropPremiere: % decrease in minimum rating from rating of premiere episode
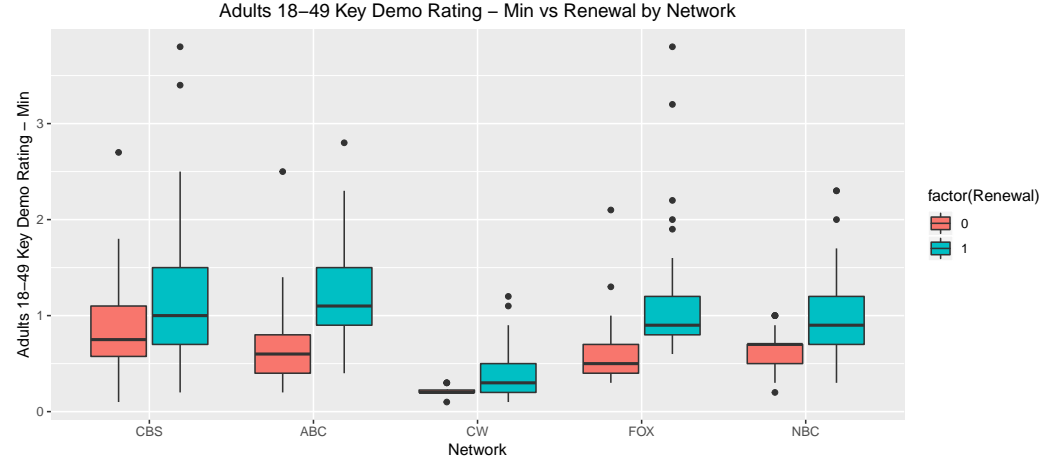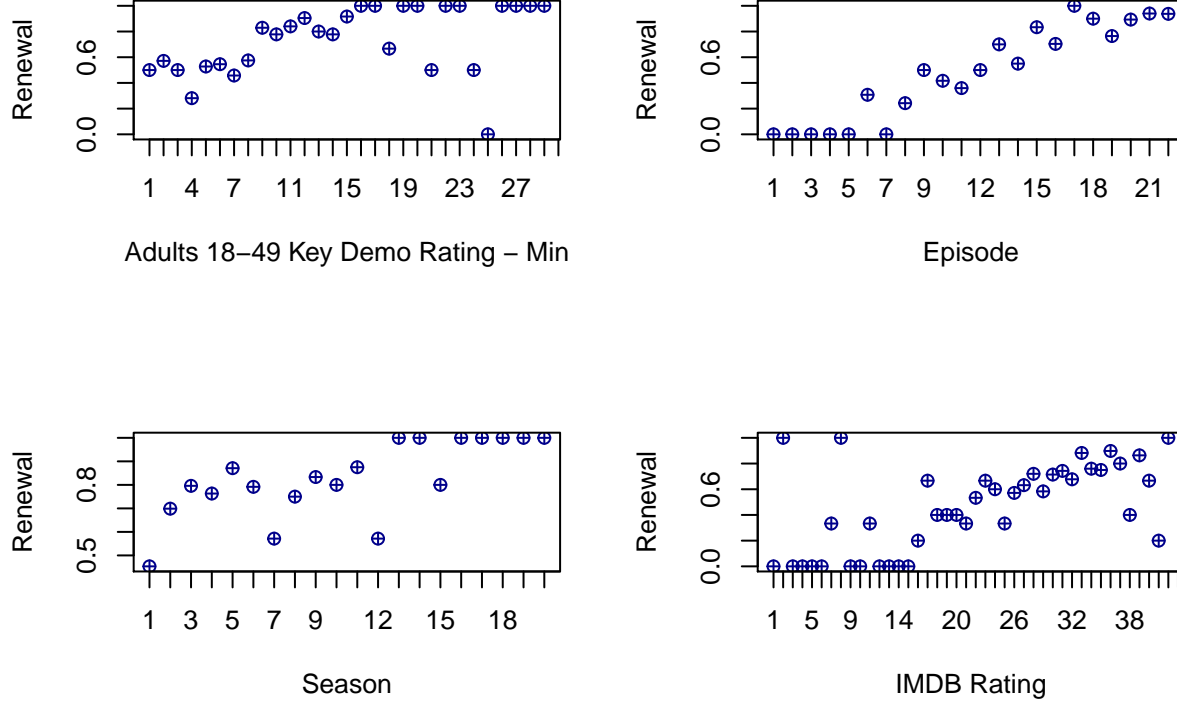
New Series Features include: - ReleaseYr1967: release year of the series - 12 Genre: IMDB genre of the series (string consists of 1 to 3 genre categories separated by comma)

Exploratory Data Analysis

With the binned scatter plots of all the continuous variables, four of which appeared to have a positive relationship with probabilty of renewal: X18to49_Min, Season1, Episode1, IMDBratings. This suggests that the odds of renewal of shows appears to be increasing with increases in the 18-49 key demographic ratings of the season, the number of seasons the show has been running, the number of episode in the season and the IMDB rating of the series.

By conducting Chisq tests on the categorical variables, three of which were tested to be significant: Network, Genre_Count, Debut.on.Different.Day. This suggests that the odds of renewal are likely to differ across different network that the show is airing, across different categories of genre counts of series (i.e series with 1 genre may have a different renewal rate that series with 2 or 3 genre categories), and between shows that premiered on a different day from its regular timeslot or not. All the 12 dummy variables for each genre were tested to have insignificant association with renewal.

From the boxplot of 18 to 49 demographic rating for renewed shows against non-renewed shows across different networks, it is observed that there are sifgnificant differences of the effects of 18 to 49 demographic rating on renewal across different networks. Hence this interaction term may be significant in determining the odds of renewal of the show and will be included in the model building process.

Adults 18–49 Key Demo Rating – Min

Episode

Season

IMDB Rating

Adults 18–49 Key Demo Rating – Min vs Renewal by Network

3. Model

By including all variables and the interaction term, the final model selected through selection using AIC and AUC value as the model selection criteria, is as below:

$$logit(Pr[Renewal_i = 1]) = \beta_0 + \beta_1 Episode1_i + \beta_2 X18to49\_Min_i + \beta_3 Network_i + \beta_4 TVSeason2013_i + \beta_5 DaysfromTVSeason$$

The predictors variables selected include nine season variables (Episode1, 18to29 Min, Network, TVSeason2013, DaysfromTVSeasonPremiere, Debut.on.Different.DayYes, Season1, X18to49_MinDropPremiere), of which two were insignificant at 95% confidence level (DaysfromTVSeasonPremiere, X18to49_MinDropPremiere), three series variables (IMDBrating, Genre_Count, Runtime), one interaction term (X18to49_Min:NetworkABC), all of which were consistent with the EDA findings of each predictor variables.

Model Building Process

The methodology adopted in the model building was initialized with the stepwise selection of all season variables and interaction term between network and 18to49 rating using AIC and AUC values as the judgement criteria. The intermediate model was then supplied as the starting model for the next stepwise selection after including all the series variables except genre using using AIC and AUC values as the judgement criteria. Again the intermediate model was then supplied as the starting model for the next stepwise selection after including the 12 genre dummy variables using using AIC and AUC values as the judgement criteria.

Model Assessment and Validation

The scatter plot between residual and each of the numeric predictor variables shows only random scatter and no discernable pattern. The spread of the points appear to be constant across the scatter plot between residual and fitted values where it shows no discernable pattern, the model assumptions are met and can be used to answer the inferential questions.

The confusion matrix is defined for the model for the training set. The cut-off value was reset according to the ROC curve's best threshold of 0.712. The final model can predict the true positives and true negatives proportionally with an overall accuracy of 85.63%. The optimal sensitivity and specificity suggest that the model is 84.22% correct at identifying true positives (i.e. actual renewals), and 88.24% correct identifying true negatives (i.e. non renewals). The estimated AUC based on the ROC curve is 0.926, which suggests the model is 92.6% well at successfully predicting whether a show will be renewed for any given season based on the given season and series statistics of the show.

Further, the test set performance is also relatively decent, through decreased significantly from training set performance, probably due to the small size of the data. Using the same threshold from training set of 0.712, the overall accuracy decreased to 72.95%, with sensitivity and specificity of 69.23% and 79.55% respectively.

Results

Based on the logistic regression output, hold all other factors constant: - Every additional year after 2013, the odds of renewal of the show increases by 65%. - Every additional increase in the number of episode in the season, the odds of renewal of the show increases by 38%. - Every additional increase in the season number, the odds of renewal of the show decreases by 15%. - Shows that premiered on a different day from its regular timeslot were 80% less likely to be renewed compared to shows that did not. - Compared to CBS shows, ABC shows are 93% less likely to be renewed, NBC and FOX shows are 95% less likely to be renewed. - For CBS shows, every 1% increase in key demographic rating for adult age 18 to 49, the odds of renewal of the show increases by 2.80 times. - For ABC shows, every 1% increase in key demographic rating for adult age 18 to 49, the odds of renewal of the show increases by 55.93 times. - For FOX shows, every 1% increase in key demographic rating for adult age 18 to 49, the odds of renewal of the show increases by 123.60 times. - For NBC shows, every 1% increase in key demographic rating for adult age 18 to 49, the odds of renewal of the show increases by 341.96 times. - Every additional unit increase in the IMBD rating of the series, the odds of renewal of the show increases by 167%. - Every addtional min increase in the episode runtime of the series, the odds of renewal of the show decreases by 2%.

```
kable(model_table[, c(1:2, 4:7)], caption="Logistic Regression Results for Final Model", digits = 5, fo
```

	Estimate	Std. Error	Pr(>|z|)	Exp Coefficient	Exp CI 2.5%	Exp CI 97.5%
(Intercept)	- 14.62457	2.16138	0.00000	0.00000	0.00000	3.000000e-05
Episode1	0.32545	0.04574	0.00000	1.38465	1.27043	1.520790e+00

	Estimate	Std. Error	Pr(>|z|)	Exp Coefficient	Exp CI 2.5%	Exp CI 97.5%
X18to49_Min	1.33566	0.69460	0.05449	3.80250	1.08557	1.635032e+01
NetworkABC	-2.68911	1.10054	0.01455	0.06794	0.00728	5.561900e-01
NetworkCW	-1.01858	1.56717	0.51573	0.36111	0.01213	6.028740e+00
NetworkFOX	-2.91726	1.17939	0.01338	0.05408	0.00466	4.932000e-01
NetworkNBC	-3.06597	1.24329	0.01366	0.04661	0.00368	4.977400e-01
TVSeason2013	0.50310	0.10925	0.00000	1.65384	1.34331	2.064320e+00
DaysfromTVSeasonPremiere	0.00498	0.00278	0.07298	1.00500	0.99961	1.010610e+00
Debut.on.Different.DayYes	-1.56995	0.64538	0.01499	0.20806	0.05662	7.123200e-01
Season1	-0.16621	0.04999	0.00088	0.84687	0.76724	9.350600e-01
IMDBrating	0.98303	0.22355	0.00001	2.67254	1.74093	4.193590e+00
X18to49_MinDropPremiere	1.58715	0.93798	0.09063	4.88978	0.78816	3.154981e+01
Genre_Count2	0.07322	0.47844	0.87837	1.07597	0.41941	2.754870e+00
Genre_Count3	0.81991	0.37808	0.03011	2.27029	1.08647	4.805630e+00
Runtime	-0.01136	0.00504	0.02434	0.98871	0.97885	9.984700e-01
X18to49_Min:NetworkABC	2.70610	1.13773	0.01738	14.97070	1.74359	1.558842e+02
X18to49_Min:NetworkCW	10.12394	6.24796	0.10516	24932.77595	1.89096	8.333887e+10
X18to49_Min:NetworkFOX	3.48946	1.32149	0.00828	32.76811	2.95678	5.602967e+02
X18to49_Min:NetworkNBC	4.50195	1.52629	0.00318	90.19295	5.26508	2.226315e+03

5. Conclusion

Limitations

As data is compiled form various sources published by online third-party services, hence the accuracy of the data may be questionable. Also, as the actual date when renewal decision is announced is unknown, it is possible that the length of the season and/or the air day and time of the season is affected by the renewal/cancellation decision.

Conclusions

This study concludes that there are various factors affecting the odds of renewal of US tv shows, including number of episode in the season, how many seasons the show has been running, year of tv season, min 18 to 49 demographic rating, whether or not the season premiered on a different day from its regular timeslot, IMDB rating of the series, genre count of series, episode runtime of the series. Also, it is found that the odds of tv show renewal differs significantly across different network the show is being broadcasted, but do not differs across different genres of the series. Lastly, it can be concluded that the effect of 18 to 49 demographic rating on tv show renewal rates differ across different tv networks.

6. Appendix

Link to Github: https://github.com/jy-xu/IDS702-Final-Project

```r
tvseries <- read.csv('Data.csv')
#tvseries$TV.Season <- factor(tvseries$TV.Season)
tvseries$FirstSeason <- 0
tvseries$FirstSeason[tvseries$Season == 1] <- 1
tvseries$Network = relevel(tvseries$Network, "NBC")
tvseries$AirDayInitial = relevel(tvseries$AirDayInitial, "Sun")

#web scrap
library(rvest)

for (show in tvseries$Show) {
  showlink = gsub(" ", "+", show)
  url <- paste('http://www.imdb.com/search/title/?title=', showlink, '&title_type=TVMini-Series&tvseries
  webpage <- read_html(url)
  genre_data_html <- html_nodes(webpage,'.genre')
  genre_data <- html_text(genre_data_html)[1]
  genre_data <- trimws(genre_data, which = c("both", "left", "right"), whitespace = "[ \t\r\n]")
  tvseries$Genre[tvseries$Show == show] <- genre_data

  runtime_data_html <- html_nodes(webpage,'.runtime')
  runtime_data <- html_text(runtime_data_html)[1]
  runtime_data <- trimws(runtime_data, which = c("both", "left", "right"), whitespace = "[ \t\r\n]")
  runtime_data <- strsplit(runtime_data, " ")[[1]][1]
  tvseries$Runtime[tvseries$Show == show] <- runtime_data

  certificate_data_html <- html_nodes(webpage,'.certificate')
  certificate_data <- html_text(certificate_data_html)[1]
  certificate_data <- trimws(certificate_data, which = c("both", "left", "right"), whitespace = "[ \t\r\
  tvseries$Certificate[tvseries$Show == show] <- certificate_data

  releaseyr_data_html <- html_nodes(webpage,'.unbold')
  releaseyr_data <- html_text(releaseyr_data_html)[2]
  releaseyr_data <- gsub("[^0-9.-]", " ", releaseyr_data)
  releaseyr_data <- trimws(releaseyr_data, which = c("both", "left", "right"), whitespace = "[ \t\r\n]")
  releaseyr_data <- strsplit(releaseyr_data, " ")
  tvseries$ReleaseYr[tvseries$Show == show] <- releaseyr_data[[1]][1]
  tvseries$EndYr[tvseries$Show == show] <- releaseyr_data[[1]][2]

  imdbrating_data_html <- html_nodes(webpage,'strong')
  imdbrating_data <- html_text(imdbrating_data_html)[3]
  imdbrating_data <- trimws(imdbrating_data, which = c("both", "left", "right"), whitespace = "[ \t\r\n]
  tvseries$IMDBrating[tvseries$Show == show] <- imdbrating_data
}

tvseries$Certificate[is.na(tvseries$Certificate)] <- "Not Rated"
tvseries$EndYr[is.na(tvseries$EndYr)] <- "On-Going"
tvseries$EndedSeries <- 1
tvseries$EndedSeries[tvseries$EndYr == "On-Going"] <- 0

genrelist <- unique(unlist(strsplit(tvseries$Genre, split = ", ")))
for (genre in genrelist) {
  tvseries[genre] <- 0
  tvseries[genre][tvseries$Genre %like% genre,] <- 1
```

```r
  if (genre == "Music") {
    tvseries[genre][tvseries$Genre %like% "Musical",] <- 0
  }
}

tvseries$Genre_Count <- rowSums(tvseries[genrelist])

# genre aggregate counts
genreagg <- data.frame(unlist(strsplit(tvseries$Genre, split = ", ")))
colnames(genreagg) <- c("genre")
genre_count <- genreagg %>%
  group_by(genre) %>%
  summarise(number = n())

#tvseries$GenreMain <- do.call(rbind, strsplit(tvseries$Genre, split = ", "))[,1]
GenreList = c("Action","Comedy", "Drama", "Crime")
tvseries$GenreMain <- ""
tvseries$GenreMain[tvseries$Action == 1] <- "Action"
tvseries$GenreMain[tvseries$Comedy == 1] <- paste(tvseries$GenreMain, "Comedy", sep = "")
tvseries$GenreMain[tvseries$Drama == 1] <- paste(tvseries$GenreMain, "Drama", sep = "")
tvseries$GenreMain[tvseries$Crime == 1] <- paste(tvseries$GenreMain, "Crime", sep = "")
tvseries$GenreMain[tvseries$GenreMain == ""] <- "Other"

write.csv(tvseries, "tvseries.csv")
```

```r
# data cleaning
tvseries <- read.csv('tvseries.csv')
tvseries <- tvseries[, 2:122]
genrelist <- colnames(tvseries[, 98:120])
tvseries_clean <- tvseries[,c(1, 4:33, 90:121)]
cols <- c("ChangeAirDayTime", "FirstSeason", "Genre", "Certificate", genrelist, "EndedSeries", "Genre_C
tvseries_clean[cols] <- lapply(tvseries_clean[cols], as.factor)
tvseries_clean$Premier.Date <- as.numeric(as.Date(tvseries_clean$Premier.Date, format="%m/%d/%y"))
tvseries_clean$Finale.Date <- as.numeric(as.Date(tvseries_clean$Finale.Date, format="%m/%d/%y"))
tvseries_clean$Runtime <- as.numeric(tvseries_clean$Runtime)
tvseries_clean$ReleaseYr <- as.numeric(tvseries_clean$ReleaseYr)
tvseries_clean$ReleaseYr[is.na(tvseries$ReleaseYr)] <- 2014
tvseries_clean$EndYr <- as.numeric(tvseries_clean$EndYr)
tvseries_clean$EndYr[tvseries$EndYr == "On-Going"] <- 0
tvseries_clean$IMDBrating <- as.numeric(tvseries_clean$IMDBrating)
tvseries_clean$X18to49_MinDropPremiere <- (tvseries_clean$X18to49_Premiere - tvseries_clean$X18to49_Min]
tvseries_clean$Total_MinDropPremiere <- (tvseries_clean$Total_Premiere - tvseries_clean$Total_Min)/tvse:

tvseries_clean$TVSeason2013 <- tvseries_clean$TVSeason - min(tvseries_clean$TVSeason)
tvseries_clean$Season1 <- tvseries_clean$Season - 1
tvseries_clean$Episode1 <- tvseries_clean$Episode - 1
tvseries_clean$ReleaseYr1967 <- tvseries_clean$ReleaseYr - min(tvseries_clean$ReleaseYr)
tvseries_clean$Network <- relevel(tvseries_clean$Network, "CBS")
summary(tvseries_clean)
```

```
##             Show        Renewal        TVSeason       Network
##   Arrow       :  6   Min.   :0.0000   Min.   :2013   CBS:150
##   Blue Bloods :  6   1st Qu.:0.0000   1st Qu.:2014   ABC:144
```

```
##   Bob's Burgers      :  6    Median :1.0000    Median :2016    CW : 82
##   Brooklyn Nine-Nine:  6    Mean   :0.6486    Mean   :2016    FOX:115
##   Chicago Fire       :  6    3rd Qu.:1.0000    3rd Qu.:2017    NBC:118
##   Chicago P.D.       :  6    Max.   :1.0000    Max.   :2018
##   (Other)           :573
##       Season          Episode          AirDay              AirTime
##   Min.   : 1.000   Min.   : 2.00   Tue    :121   9:00 PM        :181
##   1st Qu.: 1.000   1st Qu.:13.00   Thu    :111   8:00 PM        :140
##   Median : 2.000   Median :19.00   Wed    : 97   10:00 PM       :128
##   Mean   : 3.805   Mean   :17.55   Mon    : 87   8:30 PM        : 46
##   3rd Qu.: 5.000   3rd Qu.:22.00   Sun    : 82   9:30 PM        : 39
##   Max.   :30.000   Max.   :24.00   Fri    : 63   9:00 PM/8:00 PM: 13
##                                    (Other): 48   (Other)        : 62
##   AirDayInitial   AirTimeInitial    AirDayTimeInitial ChangeAirDayTime
##   Fri: 66         10:00 PM:137    Tue 9:00 PM : 47       0:518
##   Mon: 98         10:30 PM:  2    Thu 9:00 PM : 38       1: 91
##   Sat:  3         7:30 PM :  4    Wed 9:00 PM : 36
##   Sun: 86         8:00 PM :150    Thu 10:00 PM: 31
##   Thu:127         8:30 PM : 63    Thu 8:00 PM : 29
##   Tue:124         9:00 PM :201    Fri 9:00 PM : 28
##   Wed:105         9:30 PM : 52    (Other)     :400
##   Debut.on.Different.Day  Premier.Date    Finale.Date
##   No :571                 Min.   :15964   Min.   :15979
##   Yes: 38                 1st Qu.:16353   1st Qu.:16567
##                           Median :17064   Median :17196
##                           Mean   :16949   Mean   :17115
##                           3rd Qu.:17442   3rd Qu.:17667
##                           Max.   :18018   Max.   :18114
##
##   DaysfromTVSeasonPremiere   SeasonDays     X18to49_Avg      X18to49_Premiere
##   Min.   :  0.00           Min.   :  0.0   Min.   :0.120    Min.   :0.100
##   1st Qu.: 15.00           1st Qu.: 98.0   1st Qu.:0.720    1st Qu.:0.900
##   Median : 28.00           Median :190.0   Median :1.030    Median :1.300
##   Mean   : 61.16           Mean   :165.7   Mean   :1.157    Mean   :1.547
##   3rd Qu.:112.00           3rd Qu.:231.0   3rd Qu.:1.420    3rd Qu.:1.900
##   Max.   :249.00           Max.   :357.0   Max.   :5.090    Max.   :6.700
##
##    X18to49_Max      X18to49_Min      X18to49_YrtoYrChange
##   Min.   : 0.200   Min.   :0.1000   Min.   :-0.6990
##   1st Qu.: 1.000   1st Qu.:0.5000   1st Qu.:-0.2387
##   Median : 1.400   Median :0.8000   Median :-0.0812
##   Mean   : 1.683   Mean   :0.8808   Mean   :-0.1284
##   3rd Qu.: 2.000   3rd Qu.:1.1000   3rd Qu.: 0.0000
##   Max.   :11.400   Max.   :3.8000   Max.   : 0.5505
##
##   X18to49_LossGainPremiere X18to49_.DropPremiere   Total_Avg
##   Min.   :-4.7000          Min.   :-0.3395        Min.   :  -0.540
##   1st Qu.:-0.5300          1st Qu.: 0.1250        1st Qu.:   3.510
##   Median :-0.2700          Median : 0.2188        Median :   5.280
##   Mean   :-0.3902          Mean   : 0.2220        Mean   :   7.818
##   3rd Qu.:-0.1200          3rd Qu.: 0.3167        3rd Qu.:   6.920
##   Max.   : 1.2900          Max.   : 0.7705        Max.   :1446.070
##
##   Total_Premiere     Total_Max        Total_Min      Total_YrtoYrChange
```

```
##  Min.   : 0.400   Min.   :    0.50   Min.   : 0.310   Min.   :-0.69340
##  1st Qu.: 3.290   1st Qu.:    3.83   1st Qu.: 1.940   1st Qu.:-0.17830
##  Median : 5.760   Median :    6.12   Median : 3.400   Median :-0.02350
##  Mean   : 6.208   Mean   :   65.86   Mean   : 3.985   Mean   :-0.09223
##  3rd Qu.: 8.300   3rd Qu.:    8.82   3rd Qu.: 5.390   3rd Qu.: 0.00000
##  Max.   :20.020   Max.   :35977.00   Max.   :15.880   Max.   : 0.70910
##
##  Total_LossGainPremiere Total_.DropPremiere FirstSeason
##  Min.   :  -8.600       Min.   :-184.8702   0:374
##  1st Qu.:  -1.340       1st Qu.:  -0.0054   1:235
##  Median :  -0.570       Median :   0.1003
##  Mean   :   1.609       Mean   :  -0.2640
##  3rd Qu.:   0.020       3rd Qu.:   0.1749
##  Max.   :1438.290       Max.   :   2.3500
##
##                         Genre        Runtime          Certificate
##  Comedy                   :127   Min.   : 21.00   TV-14    :336
##  Crime, Drama, Mystery    : 67   1st Qu.: 30.00   TV-PG    :147
##  Action, Crime, Drama     : 52   Median : 43.00   R        : 41
##  Drama                    : 32   Mean   : 49.33   Not Rated: 28
##  Action, Adventure, Drama : 27   3rd Qu.: 45.00   PG-13    : 28
##  Drama, Fantasy, Horror   : 21   Max.   :170.00   TV-MA    : 11
##  (Other)                  :283                    (Other)  : 18
##    ReleaseYr        EndYr          IMDBrating     EndedSeries Comedy
##  Min.   :1967   Min.   : 0.000   Min.   :3.500   0:381       0:338
##  1st Qu.:2010   1st Qu.: 0.000   1st Qu.:6.900   1:228       1:271
##  Median :2014   Median : 0.000   Median :7.500
##  Mean   :2011   Mean   : 3.048   Mean   :7.365
##  3rd Qu.:2016   3rd Qu.: 8.000   3rd Qu.:7.900
##  Max.   :2020   Max.   :10.000   Max.   :8.700
##
##  Crime    Action   Drama   Mystery Romance Adventure Sci.Fi  Sport
##  0:421    0:467    0:253   0:515   0:556   0:544     0:583   0:608
##  1:188    1:142    1:356   1: 94   1: 53   1: 65     1: 26   1:  1
##
##
##
##
##
##  Biography Animation Thriller Fantasy Reality.TV Family  Musical Music
##  0:601     0:583     0:588    0:545   0:607      0:604   0:601   0:602
##  1:  8     1: 26     1: 21    1: 64   1:  2      1:  5   1:  8   1:  7
##
##
##
##
##
##  Horror   Documentary History Western Short   War    Genre_Count
##  0:572    0:607       0:608   0:608   0:607   0:606   1:166
##  1: 37    1:  2       1:  1   1:  1   1:  2   1:  3   2:112
##                                                       3:331
##
##
##
```

```
##
##  X18to49_MinDropPremiere Total_MinDropPremiere  TVSeason2013
##  Min.   :0.0000          Min.   :0.0000          Min.   :0.000
##  1st Qu.:0.3000          1st Qu.:0.2339          1st Qu.:1.000
##  Median :0.3947          Median :0.3478          Median :3.000
##  Mean   :0.4047          Mean   :0.3570          Mean   :2.547
##  3rd Qu.:0.5000          3rd Qu.:0.4659          3rd Qu.:4.000
##  Max.   :0.9048          Max.   :0.8727          Max.   :5.000
##
##     Season1          Episode1        ReleaseYr1967
##  Min.   : 0.000   Min.   : 1.00   Min.   : 0.00
##  1st Qu.: 0.000   1st Qu.:12.00   1st Qu.:43.00
##  Median : 1.000   Median :18.00   Median :47.00
##  Mean   : 2.805   Mean   :16.55   Mean   :44.44
##  3rd Qu.: 4.000   3rd Qu.:21.00   3rd Qu.:49.00
##  Max.   :29.000   Max.   :23.00   Max.   :53.00
##
```

```r
roc(test$Renewal,predict(glm, newdata = test, "response"),plot=T,print.thres="best",legacy.axes=T,
    print.auc =T,col="red3")
```

```
## Setting levels: control = 0, case = 1
```

```
## Setting direction: controls < cases
```
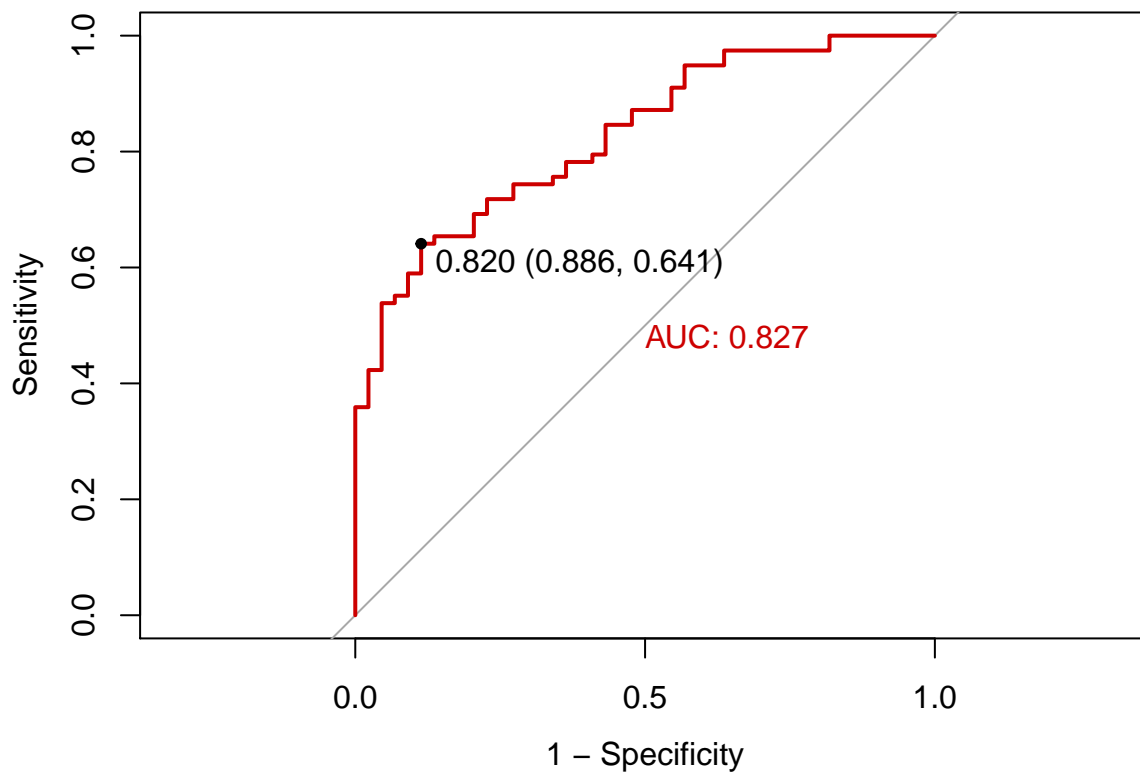


```
##
## Call:
```

```
## roc.default(response = test$Renewal, predictor = predict(glm,    newdata = test, "response"), plot =
##
## Data: predict(glm, newdata = test, "response") in 44 controls (test$Renewal 0) < 78 cases (test$Rene
## Area under the curve: 0.8272

conf_mat <- confusionMatrix(as.factor(ifelse(predict(glm, newdata = test, "response") >= 0.712, "1","0"]
                            as.factor(test$Renewal),positive = "1")
conf_mat$table
```

```
##          Reference
## Prediction  0  1
##         0 35 24
##         1  9 54
```

```
conf_mat$overall["Accuracy"];
```

```
##  Accuracy
## 0.7295082
```

```
conf_mat$byClass[c("Sensitivity","Specificity")]
```

```
## Sensitivity Specificity
##   0.6923077   0.7954545
```