Data 612 - Project 1

Joseph Simone

2/11/2020

```
library(dplyr)
library(tidyr)
library(caTools)
```

Background

For this project, the recommender system I would like to put into effevt would be one that would recommend Films, more specifically the Star Wars Skywalker Film Franchise.

This is an industry, where the Ratings or Cristics Reviews for a specific Film could make or break a companies fiscal year. Therefore, there is an abundancy of data to be collected on these various Films.

Data-Set

First, I wanted to do was created my own smaller Data-Set of the Star Wars Films using a simple scaling metric to interpert and the same Critics spanning from 1977-2019. Therefore, I decided to use the website Meta Critic.

Since this website includes various Critics that use the same numerical ratings system of a 0-100 scale. In addition, this gave me the opportunity to scrap real Reviews from the time of the Film's release, instead of using arbitrary values.

Data Import

```
star_wars_data <- read.csv("https://raw.githubusercontent.com/josephsimone/Data-612/master/project_1/sta
colnames(star_wars_data) <- gsub("ï..Critics", "Critics", colnames(star_wars_data))
star_wars_data
```

```
##                 Critics The.Phantom.Menace Attack.of.the.Clones
## 1       Chicago Tribune                 88                  100
## 2 Entertainment Weekly                  67                   88
## 3    The New York Times                 80                   40
## 4    The Globe and Mail                 63                   38
## 5  The Washington Post                  40                   NA
## 6         San Fransico                  50                   50
## 7           Reel Views                  88                   83
## 8           Daily News                  63                   91
## 9     Los Angeles Times                 NA                   50
##    Revenge.of.the.Sith A.New.Hope Empire.Strikes.Back Return.of.the.Jedi
## 1                  100         88                 100                100
## 2                   75         NA                 100                 83
## 3                   93         80                  NA                 20
## 4                   50        100                  50                 NA
## 5                   90        100                  60                 70
```

```
## 6                   50        100             100             75
## 7                   NA        100              80             75
## 8                   90        100              90            100
## 9                   70        100             100             90
##    The.Force.Awakens The.Last.Jedi Rise.of.Skywalker
## 1                 75            NA                75
## 2                 83            83                50
## 3                 90            90                50
## 4                 63            50                50
## 5                 75            75                50
## 6                 NA            75                50
## 7                 75            75                50
## 8                 80            90                NA
## 9                 60           100                50
```

Splitting Data into Training/Testing Sets

In this seection, the Data-Frame was first converted to long format.

Then split into training and testing sets based on 0.75 split ratio.

```
split_star_wars_data <- star_wars_data %>% gather(key = Film, value = Review, -Critics)
```

```
set.seed(50)
split <- sample.split(split_star_wars_data$Review, SplitRatio = 0.75)
```

```
train_data_set <- split_star_wars_data
train_data_set$Review[!split] <- NA
head(train_data_set)
```

```
##                 Critics              Film Review
## 1      Chicago Tribune The.Phantom.Menace     88
## 2 Entertainment Weekly The.Phantom.Menace     67
## 3   The New York Times The.Phantom.Menace     80
## 4   The Globe and Mail The.Phantom.Menace     63
## 5  The Washington Post The.Phantom.Menace     40
## 6         San Fransico The.Phantom.Menace     50
```

```
test_data_set <- split_star_wars_data
test_data_set$Review[split] <- NA
```

```
head(test_data_set)
```

```
##                 Critics              Film Review
## 1      Chicago Tribune The.Phantom.Menace     NA
## 2 Entertainment Weekly The.Phantom.Menace     NA
## 3   The New York Times The.Phantom.Menace     NA
## 4   The Globe and Mail The.Phantom.Menace     NA
## 5  The Washington Post The.Phantom.Menace     NA
## 6         San Fransico The.Phantom.Menace     NA
```

Since, there are now two different dataset randomly selected, it is time to move onto the RMSE calculations.

```r
raw_avg <- sum(train_data_set$Review, na.rm = TRUE) / length(which(!is.na(train_data_set$Review)))

rmse_raw_train <- sqrt(sum((train_data_set$Review[!is.na(train_data_set$Review)] - raw_avg)^2) /
                       length(which(!is.na(train_data_set$Review)))))
rmse_raw_train
```

```
## [1] 20.39405
```

```r
rmse_raw_test <- sqrt(sum((test_data_set$Review[!is.na(test_data_set$Review)] - raw_avg)^2) /
                      length(which(!is.na(test_data_set$Review)))))
rmse_raw_test
```

```
## [1] 19.18502
```

We can observe that RMSE values are significantly larger than expected in a smaller sample space.

Baseline Predictors

```r
Critics_bias <- train_data_set %>% filter(!is.na(Review)) %>%
  group_by(Critics) %>%
  summarise(sum = sum(Review), count = n()) %>%
  mutate(bias = sum/count-raw_avg) %>%
  select(Critics, CriticsBias = bias)
CriticsBias<-Critics_bias$CriticsBias
```

```r
Film_bias <- train_data_set %>% filter(!is.na(Review)) %>%
  group_by(Film) %>%
  summarise(sum = sum(Review), count = n()) %>%
  mutate(bias = sum/count-raw_avg) %>%
  select(Film, FilmBias = bias)
FilmBias<-Film_bias$FilmBias
```

```r
train_data_set <- train_data_set %>% left_join(Critics_bias, by = "Critics") %>%
  left_join(Film_bias, by = "Film") %>%
  mutate(RawAvg = raw_avg) %>%
  mutate(Baseline = RawAvg + CriticsBias + FilmBias)
train_data_set
```

```
##                   Critics                  Film Review CriticsBias
## 1         Chicago Tribune    The.Phantom.Menace     88  16.5357143
## 2    Entertainment Weekly    The.Phantom.Menace     67  -0.4071429
## 3       The New York Times    The.Phantom.Menace     80  -5.7738095
## 4       The Globe and Mail    The.Phantom.Menace     63 -15.4642857
## 5      The Washington Post    The.Phantom.Menace     40  -4.6071429
## 6             San Fransico    The.Phantom.Menace     50  -3.7738095
## 7               Reel Views    The.Phantom.Menace     88   0.5357143
## 8               Daily News    The.Phantom.Menace     NA  15.6428571
## 9        Los Angeles Times    The.Phantom.Menace     NA   3.7261905
## 10        Chicago Tribune Attack.of.the.Clones    100  16.5357143
```

```
## 11 Entertainment Weekly Attack.of.the.Clones      88  -0.4071429
## 12   The New York Times Attack.of.the.Clones      40  -5.7738095
## 13   The Globe and Mail Attack.of.the.Clones      38 -15.4642857
## 14 The Washington Post Attack.of.the.Clones        NA  -4.6071429
## 15          San Fransico Attack.of.the.Clones      NA  -3.7738095
## 16            Reel Views Attack.of.the.Clones      83   0.5357143
## 17            Daily News Attack.of.the.Clones      91  15.6428571
## 18     Los Angeles Times Attack.of.the.Clones      NA   3.7261905
## 19        Chicago Tribune  Revenge.of.the.Sith     100  16.5357143
## 20 Entertainment Weekly  Revenge.of.the.Sith      NA  -0.4071429
## 21    The New York Times  Revenge.of.the.Sith      93  -5.7738095
## 22    The Globe and Mail  Revenge.of.the.Sith      50 -15.4642857
## 23  The Washington Post  Revenge.of.the.Sith      90  -4.6071429
## 24          San Fransico  Revenge.of.the.Sith      50  -3.7738095
## 25            Reel Views  Revenge.of.the.Sith      NA   0.5357143
## 26            Daily News  Revenge.of.the.Sith      NA  15.6428571
## 27     Los Angeles Times  Revenge.of.the.Sith      70   3.7261905
## 28        Chicago Tribune          A.New.Hope      NA  16.5357143
## 29 Entertainment Weekly          A.New.Hope      NA  -0.4071429
## 30    The New York Times          A.New.Hope      NA  -5.7738095
## 31    The Globe and Mail          A.New.Hope     100 -15.4642857
## 32  The Washington Post          A.New.Hope     100  -4.6071429
## 33          San Fransico          A.New.Hope     100  -3.7738095
## 34            Reel Views          A.New.Hope      NA   0.5357143
## 35            Daily News          A.New.Hope     100  15.6428571
## 36     Los Angeles Times          A.New.Hope      NA   3.7261905
## 37        Chicago Tribune  Empire.Strikes.Back     100  16.5357143
## 38 Entertainment Weekly  Empire.Strikes.Back      NA  -0.4071429
## 39    The New York Times  Empire.Strikes.Back      NA  -5.7738095
## 40    The Globe and Mail  Empire.Strikes.Back      50 -15.4642857
## 41  The Washington Post  Empire.Strikes.Back      60  -4.6071429
## 42          San Fransico  Empire.Strikes.Back     100  -3.7738095
## 43            Reel Views  Empire.Strikes.Back      80   0.5357143
## 44            Daily News  Empire.Strikes.Back      90  15.6428571
## 45     Los Angeles Times  Empire.Strikes.Back     100   3.7261905
## 46        Chicago Tribune  Return.of.the.Jedi     100  16.5357143
## 47 Entertainment Weekly  Return.of.the.Jedi      83  -0.4071429
## 48    The New York Times  Return.of.the.Jedi      20  -5.7738095
## 49    The Globe and Mail  Return.of.the.Jedi      NA -15.4642857
## 50  The Washington Post  Return.of.the.Jedi      70  -4.6071429
## 51          San Fransico  Return.of.the.Jedi      NA  -3.7738095
## 52            Reel Views  Return.of.the.Jedi      75   0.5357143
## 53            Daily News  Return.of.the.Jedi      NA  15.6428571
## 54     Los Angeles Times  Return.of.the.Jedi      90   3.7261905
## 55        Chicago Tribune   The.Force.Awakens      75  16.5357143
## 56 Entertainment Weekly   The.Force.Awakens      NA  -0.4071429
## 57    The New York Times   The.Force.Awakens      90  -5.7738095
## 58    The Globe and Mail   The.Force.Awakens      63 -15.4642857
## 59  The Washington Post   The.Force.Awakens      75  -4.6071429
## 60          San Fransico   The.Force.Awakens      NA  -3.7738095
## 61            Reel Views   The.Force.Awakens      75   0.5357143
## 62            Daily News   The.Force.Awakens      80  15.6428571
## 63     Los Angeles Times   The.Force.Awakens      60   3.7261905
## 64        Chicago Tribune      The.Last.Jedi      NA  16.5357143
```

```
## 65 Entertainment Weekly        The.Last.Jedi     83  -0.4071429
## 66    The New York Times        The.Last.Jedi     90  -5.7738095
## 67    The Globe and Mail        The.Last.Jedi     50 -15.4642857
## 68 The Washington Post          The.Last.Jedi     75  -4.6071429
## 69           San Fransico       The.Last.Jedi     75  -3.7738095
## 70             Reel Views       The.Last.Jedi     75   0.5357143
## 71             Daily News       The.Last.Jedi     NA  15.6428571
## 72      Los Angeles Times       The.Last.Jedi    100   3.7261905
## 73         Chicago Tribune   Rise.of.Skywalker    75  16.5357143
## 74 Entertainment Weekly     Rise.of.Skywalker    50  -0.4071429
## 75    The New York Times    Rise.of.Skywalker    NA  -5.7738095
## 76    The Globe and Mail    Rise.of.Skywalker    NA -15.4642857
## 77 The Washington Post      Rise.of.Skywalker    50  -4.6071429
## 78           San Fransico   Rise.of.Skywalker    50  -3.7738095
## 79             Reel Views   Rise.of.Skywalker    50   0.5357143
## 80             Daily News   Rise.of.Skywalker    NA  15.6428571
## 81      Los Angeles Times   Rise.of.Skywalker    50   3.7261905
##       FilmBias    RawAvg  Baseline
## 1   -6.6071429 74.60714  84.53571
## 2   -6.6071429 74.60714  67.59286
## 3   -6.6071429 74.60714  62.22619
## 4   -6.6071429 74.60714  52.53571
## 5   -6.6071429 74.60714  63.39286
## 6   -6.6071429 74.60714  64.22619
## 7   -6.6071429 74.60714  68.53571
## 8   -6.6071429 74.60714  83.64286
## 9   -6.6071429 74.60714  71.72619
## 10  -1.2738095 74.60714  89.86905
## 11  -1.2738095 74.60714  72.92619
## 12  -1.2738095 74.60714  67.55952
## 13  -1.2738095 74.60714  57.86905
## 14  -1.2738095 74.60714  68.72619
## 15  -1.2738095 74.60714  69.55952
## 16  -1.2738095 74.60714  73.86905
## 17  -1.2738095 74.60714  88.97619
## 18  -1.2738095 74.60714  77.05952
## 19   0.8928571 74.60714  92.03571
## 20   0.8928571 74.60714  75.09286
## 21   0.8928571 74.60714  69.72619
## 22   0.8928571 74.60714  60.03571
## 23   0.8928571 74.60714  70.89286
## 24   0.8928571 74.60714  71.72619
## 25   0.8928571 74.60714  76.03571
## 26   0.8928571 74.60714  91.14286
## 27   0.8928571 74.60714  79.22619
## 28  25.3928571 74.60714 116.53571
## 29  25.3928571 74.60714  99.59286
## 30  25.3928571 74.60714  94.22619
## 31  25.3928571 74.60714  84.53571
## 32  25.3928571 74.60714  95.39286
## 33  25.3928571 74.60714  96.22619
## 34  25.3928571 74.60714 100.53571
## 35  25.3928571 74.60714 115.64286
## 36  25.3928571 74.60714 103.72619
```

```
## 37   8.2500000 74.60714  99.39286
## 38   8.2500000 74.60714  82.45000
## 39   8.2500000 74.60714  77.08333
## 40   8.2500000 74.60714  67.39286
## 41   8.2500000 74.60714  78.25000
## 42   8.2500000 74.60714  79.08333
## 43   8.2500000 74.60714  83.39286
## 44   8.2500000 74.60714  98.50000
## 45   8.2500000 74.60714  86.58333
## 46  -1.6071429 74.60714  89.53571
## 47  -1.6071429 74.60714  72.59286
## 48  -1.6071429 74.60714  67.22619
## 49  -1.6071429 74.60714  57.53571
## 50  -1.6071429 74.60714  68.39286
## 51  -1.6071429 74.60714  69.22619
## 52  -1.6071429 74.60714  73.53571
## 53  -1.6071429 74.60714  88.64286
## 54  -1.6071429 74.60714  76.72619
## 55  -0.6071429 74.60714  90.53571
## 56  -0.6071429 74.60714  73.59286
## 57  -0.6071429 74.60714  68.22619
## 58  -0.6071429 74.60714  58.53571
## 59  -0.6071429 74.60714  69.39286
## 60  -0.6071429 74.60714  70.22619
## 61  -0.6071429 74.60714  74.53571
## 62  -0.6071429 74.60714  89.64286
## 63  -0.6071429 74.60714  77.72619
## 64   3.6785714 74.60714  94.82143
## 65   3.6785714 74.60714  77.87857
## 66   3.6785714 74.60714  72.51190
## 67   3.6785714 74.60714  62.82143
## 68   3.6785714 74.60714  73.67857
## 69   3.6785714 74.60714  74.51190
## 70   3.6785714 74.60714  78.82143
## 71   3.6785714 74.60714  93.92857
## 72   3.6785714 74.60714  82.01190
## 73 -20.4404762 74.60714  70.70238
## 74 -20.4404762 74.60714  53.75952
## 75 -20.4404762 74.60714  48.39286
## 76 -20.4404762 74.60714  38.70238
## 77 -20.4404762 74.60714  49.55952
## 78 -20.4404762 74.60714  50.39286
## 79 -20.4404762 74.60714  54.70238
## 80 -20.4404762 74.60714  69.80952
## 81 -20.4404762 74.60714  57.89286
```

```r
test_data_set <- test_data_set %>% left_join(Critics_bias, by = "Critics") %>%
  left_join(Film_bias, by = "Film") %>%
  mutate(RawAvg = raw_avg) %>%
  mutate(Baseline = RawAvg + CriticsBias + FilmBias)
test_data_set
```

```
##              Critics              Film Review CriticsBias
## 1    Chicago Tribune  The.Phantom.Menace     NA  16.5357143
```

```
## 2   Entertainment Weekly   The.Phantom.Menace      NA  -0.4071429
## 3    The New York Times     The.Phantom.Menace      NA  -5.7738095
## 4    The Globe and Mail     The.Phantom.Menace      NA -15.4642857
## 5   The Washington Post     The.Phantom.Menace      NA  -4.6071429
## 6          San Fransico     The.Phantom.Menace      NA  -3.7738095
## 7            Reel Views     The.Phantom.Menace      NA   0.5357143
## 8            Daily News     The.Phantom.Menace      63  15.6428571
## 9      Los Angeles Times   The.Phantom.Menace      NA   3.7261905
## 10     Chicago Tribune Attack.of.the.Clones        NA  16.5357143
## 11 Entertainment Weekly Attack.of.the.Clones        NA  -0.4071429
## 12   The New York Times Attack.of.the.Clones        NA  -5.7738095
## 13   The Globe and Mail Attack.of.the.Clones        NA -15.4642857
## 14  The Washington Post Attack.of.the.Clones        NA  -4.6071429
## 15         San Fransico Attack.of.the.Clones        50  -3.7738095
## 16           Reel Views Attack.of.the.Clones        NA   0.5357143
## 17           Daily News Attack.of.the.Clones        NA  15.6428571
## 18     Los Angeles Times Attack.of.the.Clones       50   3.7261905
## 19     Chicago Tribune  Revenge.of.the.Sith        NA  16.5357143
## 20 Entertainment Weekly  Revenge.of.the.Sith        75  -0.4071429
## 21   The New York Times  Revenge.of.the.Sith        NA  -5.7738095
## 22   The Globe and Mail  Revenge.of.the.Sith        NA -15.4642857
## 23  The Washington Post  Revenge.of.the.Sith        NA  -4.6071429
## 24         San Fransico  Revenge.of.the.Sith        NA  -3.7738095
## 25           Reel Views  Revenge.of.the.Sith        NA   0.5357143
## 26           Daily News  Revenge.of.the.Sith        90  15.6428571
## 27     Los Angeles Times  Revenge.of.the.Sith       NA   3.7261905
## 28     Chicago Tribune          A.New.Hope        88  16.5357143
## 29 Entertainment Weekly          A.New.Hope        NA  -0.4071429
## 30   The New York Times          A.New.Hope        80  -5.7738095
## 31   The Globe and Mail          A.New.Hope        NA -15.4642857
## 32  The Washington Post          A.New.Hope        NA  -4.6071429
## 33         San Fransico          A.New.Hope        NA  -3.7738095
## 34           Reel Views          A.New.Hope       100   0.5357143
## 35           Daily News          A.New.Hope        NA  15.6428571
## 36     Los Angeles Times          A.New.Hope       100   3.7261905
## 37     Chicago Tribune  Empire.Strikes.Back        NA  16.5357143
## 38 Entertainment Weekly  Empire.Strikes.Back       100  -0.4071429
## 39   The New York Times  Empire.Strikes.Back        NA  -5.7738095
## 40   The Globe and Mail  Empire.Strikes.Back        NA -15.4642857
## 41  The Washington Post  Empire.Strikes.Back        NA  -4.6071429
## 42         San Fransico  Empire.Strikes.Back        NA  -3.7738095
## 43           Reel Views  Empire.Strikes.Back        NA   0.5357143
## 44           Daily News  Empire.Strikes.Back        NA  15.6428571
## 45     Los Angeles Times  Empire.Strikes.Back       NA   3.7261905
## 46     Chicago Tribune   Return.of.the.Jedi        NA  16.5357143
## 47 Entertainment Weekly   Return.of.the.Jedi        NA  -0.4071429
## 48   The New York Times   Return.of.the.Jedi        NA  -5.7738095
## 49   The Globe and Mail   Return.of.the.Jedi        NA -15.4642857
## 50  The Washington Post   Return.of.the.Jedi        NA  -4.6071429
## 51         San Fransico   Return.of.the.Jedi        75  -3.7738095
## 52           Reel Views   Return.of.the.Jedi        NA   0.5357143
## 53           Daily News   Return.of.the.Jedi       100  15.6428571
## 54     Los Angeles Times   Return.of.the.Jedi       NA   3.7261905
## 55     Chicago Tribune    The.Force.Awakens        NA  16.5357143
```

```
## 56 Entertainment Weekly    The.Force.Awakens    83  -0.4071429
## 57   The New York Times    The.Force.Awakens    NA  -5.7738095
## 58   The Globe and Mail    The.Force.Awakens    NA -15.4642857
## 59 The Washington Post     The.Force.Awakens    NA  -4.6071429
## 60         San Fransico    The.Force.Awakens    NA  -3.7738095
## 61            Reel Views   The.Force.Awakens    NA   0.5357143
## 62            Daily News   The.Force.Awakens    NA  15.6428571
## 63    Los Angeles Times    The.Force.Awakens    NA   3.7261905
## 64       Chicago Tribune       The.Last.Jedi    NA  16.5357143
## 65 Entertainment Weekly       The.Last.Jedi    NA  -0.4071429
## 66   The New York Times       The.Last.Jedi    NA  -5.7738095
## 67   The Globe and Mail       The.Last.Jedi    NA -15.4642857
## 68 The Washington Post        The.Last.Jedi    NA  -4.6071429
## 69         San Fransico       The.Last.Jedi    NA  -3.7738095
## 70            Reel Views      The.Last.Jedi    NA   0.5357143
## 71            Daily News      The.Last.Jedi    90  15.6428571
## 72    Los Angeles Times       The.Last.Jedi    NA   3.7261905
## 73       Chicago Tribune  Rise.of.Skywalker    NA  16.5357143
## 74 Entertainment Weekly   Rise.of.Skywalker    NA  -0.4071429
## 75   The New York Times   Rise.of.Skywalker    50  -5.7738095
## 76   The Globe and Mail   Rise.of.Skywalker    50 -15.4642857
## 77 The Washington Post    Rise.of.Skywalker    NA  -4.6071429
## 78         San Fransico   Rise.of.Skywalker    NA  -3.7738095
## 79            Reel Views   Rise.of.Skywalker    NA   0.5357143
## 80            Daily News   Rise.of.Skywalker    NA  15.6428571
## 81    Los Angeles Times   Rise.of.Skywalker    NA   3.7261905
##        FilmBias    RawAvg  Baseline
## 1   -6.6071429 74.60714  84.53571
## 2   -6.6071429 74.60714  67.59286
## 3   -6.6071429 74.60714  62.22619
## 4   -6.6071429 74.60714  52.53571
## 5   -6.6071429 74.60714  63.39286
## 6   -6.6071429 74.60714  64.22619
## 7   -6.6071429 74.60714  68.53571
## 8   -6.6071429 74.60714  83.64286
## 9   -6.6071429 74.60714  71.72619
## 10  -1.2738095 74.60714  89.86905
## 11  -1.2738095 74.60714  72.92619
## 12  -1.2738095 74.60714  67.55952
## 13  -1.2738095 74.60714  57.86905
## 14  -1.2738095 74.60714  68.72619
## 15  -1.2738095 74.60714  69.55952
## 16  -1.2738095 74.60714  73.86905
## 17  -1.2738095 74.60714  88.97619
## 18  -1.2738095 74.60714  77.05952
## 19   0.8928571 74.60714  92.03571
## 20   0.8928571 74.60714  75.09286
## 21   0.8928571 74.60714  69.72619
## 22   0.8928571 74.60714  60.03571
## 23   0.8928571 74.60714  70.89286
## 24   0.8928571 74.60714  71.72619
## 25   0.8928571 74.60714  76.03571
## 26   0.8928571 74.60714  91.14286
## 27   0.8928571 74.60714  79.22619
```

```
## 28   25.3928571 74.60714 116.53571
## 29   25.3928571 74.60714  99.59286
## 30   25.3928571 74.60714  94.22619
## 31   25.3928571 74.60714  84.53571
## 32   25.3928571 74.60714  95.39286
## 33   25.3928571 74.60714  96.22619
## 34   25.3928571 74.60714 100.53571
## 35   25.3928571 74.60714 115.64286
## 36   25.3928571 74.60714 103.72619
## 37    8.2500000 74.60714  99.39286
## 38    8.2500000 74.60714  82.45000
## 39    8.2500000 74.60714  77.08333
## 40    8.2500000 74.60714  67.39286
## 41    8.2500000 74.60714  78.25000
## 42    8.2500000 74.60714  79.08333
## 43    8.2500000 74.60714  83.39286
## 44    8.2500000 74.60714  98.50000
## 45    8.2500000 74.60714  86.58333
## 46   -1.6071429 74.60714  89.53571
## 47   -1.6071429 74.60714  72.59286
## 48   -1.6071429 74.60714  67.22619
## 49   -1.6071429 74.60714  57.53571
## 50   -1.6071429 74.60714  68.39286
## 51   -1.6071429 74.60714  69.22619
## 52   -1.6071429 74.60714  73.53571
## 53   -1.6071429 74.60714  88.64286
## 54   -1.6071429 74.60714  76.72619
## 55   -0.6071429 74.60714  90.53571
## 56   -0.6071429 74.60714  73.59286
## 57   -0.6071429 74.60714  68.22619
## 58   -0.6071429 74.60714  58.53571
## 59   -0.6071429 74.60714  69.39286
## 60   -0.6071429 74.60714  70.22619
## 61   -0.6071429 74.60714  74.53571
## 62   -0.6071429 74.60714  89.64286
## 63   -0.6071429 74.60714  77.72619
## 64    3.6785714 74.60714  94.82143
## 65    3.6785714 74.60714  77.87857
## 66    3.6785714 74.60714  72.51190
## 67    3.6785714 74.60714  62.82143
## 68    3.6785714 74.60714  73.67857
## 69    3.6785714 74.60714  74.51190
## 70    3.6785714 74.60714  78.82143
## 71    3.6785714 74.60714  93.92857
## 72    3.6785714 74.60714  82.01190
## 73  -20.4404762 74.60714  70.70238
## 74  -20.4404762 74.60714  53.75952
## 75  -20.4404762 74.60714  48.39286
## 76  -20.4404762 74.60714  38.70238
## 77  -20.4404762 74.60714  49.55952
## 78  -20.4404762 74.60714  50.39286
## 79  -20.4404762 74.60714  54.70238
## 80  -20.4404762 74.60714  69.80952
## 81  -20.4404762 74.60714  57.89286
```

```
rmse_base_train <- sqrt(sum((train_data_set$Review[!is.na(train_data_set$Review)] -
                             train_data_set$Baseline[!is.na(train_data_set$Review)])^2) /
                       length(which(!is.na(train_data_set$Review))))
rmse_base_test <- sqrt(sum((test_data_set$Review[!is.na(test_data_set$Review)] -
                            test_data_set$Baseline[!is.na(test_data_set$Review)])^2) /
                      length(which(!is.na(test_data_set$Review))))
```

RMSE

The Largest Bias are determined by the NA Critic's Reviews that did not review the Film.

`rmse_base_train`

```
## [1] 14.25514
```

`rmse_base_test`

```
## [1] 14.31247
```

The table below represents the RMSE values for both the Training and Testing Sets and the Raw Average and Baseline Predictors.

```
## Warning in kableExtra::kable_styling(., bootstrap_options = c("striped", :
## Please specify format in kable. kableExtra can customize either HTML or
## LaTeX outputs. See https://haozhu233.github.io/kableExtra/ for details.
```

	RMSE
Training: Raw Average	20.39405
Training: Baseline Predictor	14.25514
Testing: Raw Average	19.18502
Testing: Baseline Predictor	14.31247

Summary

The RMSE values improved the Baseline Predictors in both the Training and Test Sets.

Even with a smaller Data-Set that included incomplete values, this was enough information to be able to visualize and apply specific Films and Critic bias into our model.