Movie Recommendation System

Aman Sharma

2023-06-13

Project Details.

This is a movie recommendation project made for **CodeClause**. Functions have been made instead of modelling for the predictions. The data set has been taken from *Kaggle*.

Loading the data and packages

```
library(tidyverse)
library(ggplot2)

data<- read.csv("C:/Users/91835/Desktop/CodeClause_data/movie_ratings.csv")
```

Exploring and cleaning the Data

Finding the mean of metascore and replacing *NA* with the mean value The mean is 74.27941, rounding off to 74

```
head(data)
```

```
##   ID               movie year imdb metascore   votes
## 1  0           Gladiator 2000  8.5        67 1187559
## 2  1             Memento 2000  8.5        80 1014199
## 3  2              Snatch 2000  8.3        74  707162
## 4  3 Requiem for a Dream 2000  8.3        68  688394
## 5  4               X-Men 2000  7.4        64  526411
## 6  5           Cast Away 2000  7.8        73  465125
```

```
dim(data)
```

```
## [1] 900   6
```

```
summary(data)
```

```
##        ID            movie                year           imdb
##  Min.   :   0.0   Length:900         Min.   :2000   Min.   :4.100
##  1st Qu.: 424.8   Class :character   1st Qu.:2004   1st Qu.:6.700
```

```
##  Median : 874.5   Mode  :character   Median :2008   Median :7.300
##  Mean   : 874.5                      Mean   :2008   Mean   :7.239
##  3rd Qu.:1324.2                      3rd Qu.:2013   3rd Qu.:7.800
##  Max.   :1749.0                      Max.   :2017   Max.   :9.900
##
##    metascore          votes
##  Min.   : 61.00   Min.   :  93428
##  1st Qu.: 67.00   1st Qu.: 184223
##  Median : 73.00   Median : 246739
##  Mean   : 74.28   Mean   : 317636
##  3rd Qu.: 81.00   3rd Qu.: 382583
##  Max.   :100.00   Max.   :2020298
##  NA's   :424
```

```
work_data<- data %>%
  replace_na(list(metascore=74))
```

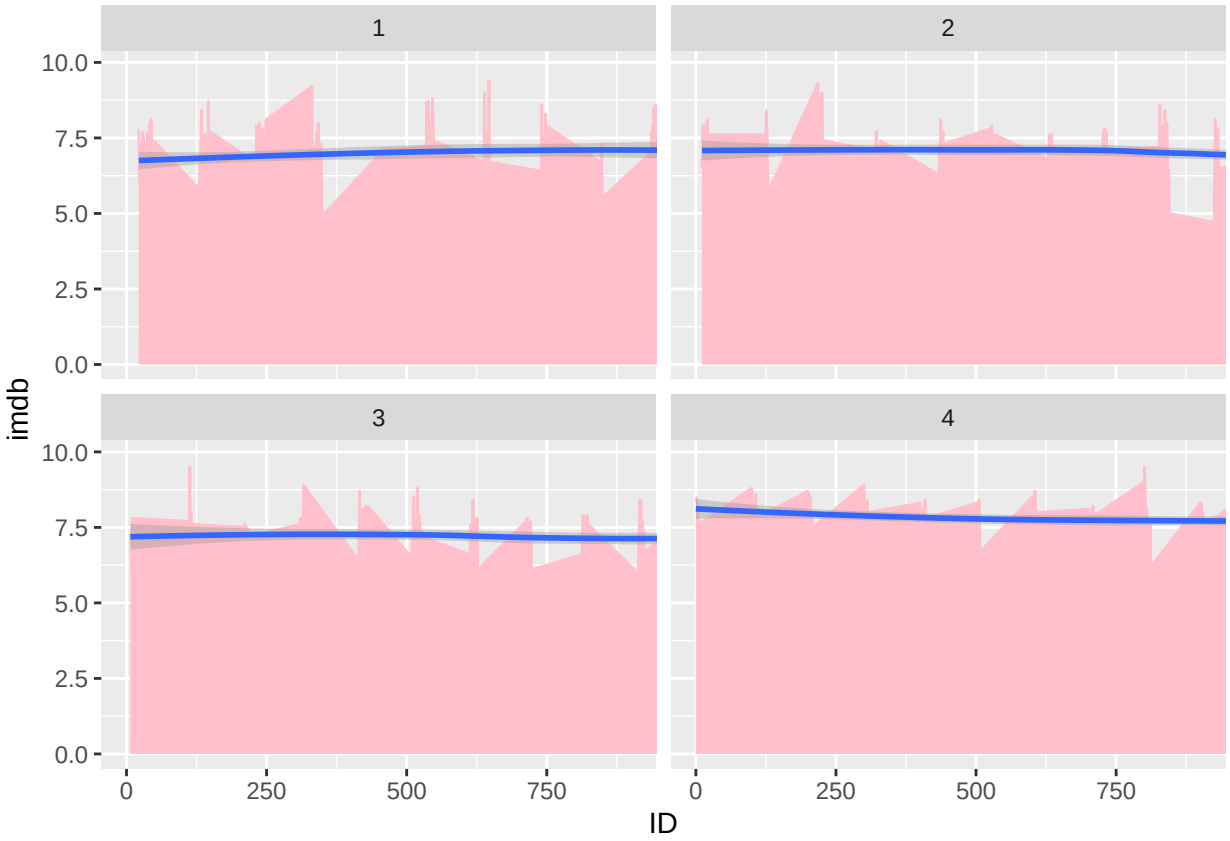
Binning the votes into 4 equal bins, 4 =Higher votes category, 1=Lower votes category.

```
work_data<-work_data %>%
  mutate(vote_category=ntile(votes,4))
```

Visualizing data

We observe that movies which are in vote category 4 tend to have a higher imdb rating.

```
ggplot(work_data, aes(x=ID, y=imdb))+
  geom_path(color="pink")+
  geom_area(fill="pink")+
  geom_smooth()+
  coord_cartesian(xlim = c(0,900))+
  facet_wrap(~vote_category)
```



Recommendations

Top rated movie functions

This function tells the top rated movies from the list. The criteria that it uses is : imdb rating >=8.5 8.5 Metascore >=85 Vote category 4 (Higher Votes)

```
top_rated<- function(){
for(i in 1:nrow(work_data))
{
  if(work_data$imdb[i]>=8.5 && work_data$metascore[i]>=85 && work_data$vote_category[i]==4)
  {
    print(work_data[i, "movie"])
  }
}
}
#Calling top_rated function
top_rated()
```

```
## [1] "The Lord of the Rings: The Fellowship of the Ring"
## [1] "Sen to Chihiro no kamikakushi"
## [1] "The Lord of the Rings: The Two Towers"
## [1] "The Pianist"
## [1] "The Lord of the Rings: The Return of the King"
```

```
## [1] "The Departed"
## [1] "Whiplash"
```

Recommendations by ID

This function will take in an argument which will be a numeric. The argument signifies the ID for whom the recommendations are to be generated. The way it works is that it reads the movie that has been seen by the user. It considers it's metascore and imdb ratings and recommends movies in that same range. The range has been set as: imdb to imdb+0.5 metascore to metascore+7 The function will consider the movie present at the ID as that user's preference and generate recommendations in that range only.

```
recommend<- function(x)
  {
  for(i in 1:nrow(work_data))
    {
    if(work_data$imdb[i]>=(work_data$imdb[x]) && work_data$imdb[i]<(work_data$imdb[x]+0.5))
    {
      if(work_data$metascore[i]>=work_data$metascore[x] && work_data$metascore[i]<(work_data$metascore[
         {

        print(work_data[i, "movie"])
      }
    }
  }
}

#Calling the function
recommend(2)
```

```
## [1] "Memento"
## [1] "The Pianist"
## [1] "The Departed"
```