

Movie Rating EDA

Nnaemeka Newman Okereafor

About The Dataset

Original Dataset; source: <https://www.kaggle.com/rounakbanik/the-movies-dataset>

```
## Rows: 44,922
## Columns: 5
## $ userid      <int> 1, 1, 1, 1, 1, 1, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, ~
## $ movieId     <int> 1371, 1405, 2105, 2193, 2294, 2455, 17, 62, 110, 144, 150, 1~
## $ Title       <chr> "Rocky III", "Greed", "American Pie", "My Tutor", "Jay and S~
## $ rating      <int> 3, 1, 4, 2, 2, 3, 5, 3, 4, 3, 5, 4, 3, 3, 3, 3, 5, 1, 3, ~
## $ Timestamp   <int> 1260759135, 1260759203, 1260759139, 1260759198, 1260759108, ~
```

Derived Dataset

```
## Rows: 44,922
## Columns: 8
## $ userid      <int> 1, 1, 1, 1, 1, 1, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2~
## $ movieId     <int> 1371, 1405, 2105, 2193, 2294, 2455, 17, 62, 110, 144, 150, ~
## $ Title       <chr> "Rocky III", "Greed", "American Pie", "My Tutor", "Jay and~
## $ rating      <int> 3, 1, 4, 2, 2, 3, 5, 3, 4, 3, 5, 4, 3, 3, 3, 3, 5, 1, 3~
## $ Rating_date <dtm> 2009-12-14 02:52:15, 2009-12-14 02:53:23, 2009-12-14 02:5~
## $ Weekday     <chr> "Monday", "Monday", "Monday", "Monday", "Monday", "Monday"~
## $ Year        <dbl> 2009, 2009, 2009, 2009, 2009, 2009, 1996, 1996, 1996, 1996~
## $ Hour        <int> 2, 2, 2, 2, 2, 2, 11, 11, 11, 11, 11, 11, 11, 11, 11, 11, ~
```

Rating Summary

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##      1.000   3.000   4.000   3.671   4.000   5.000
```

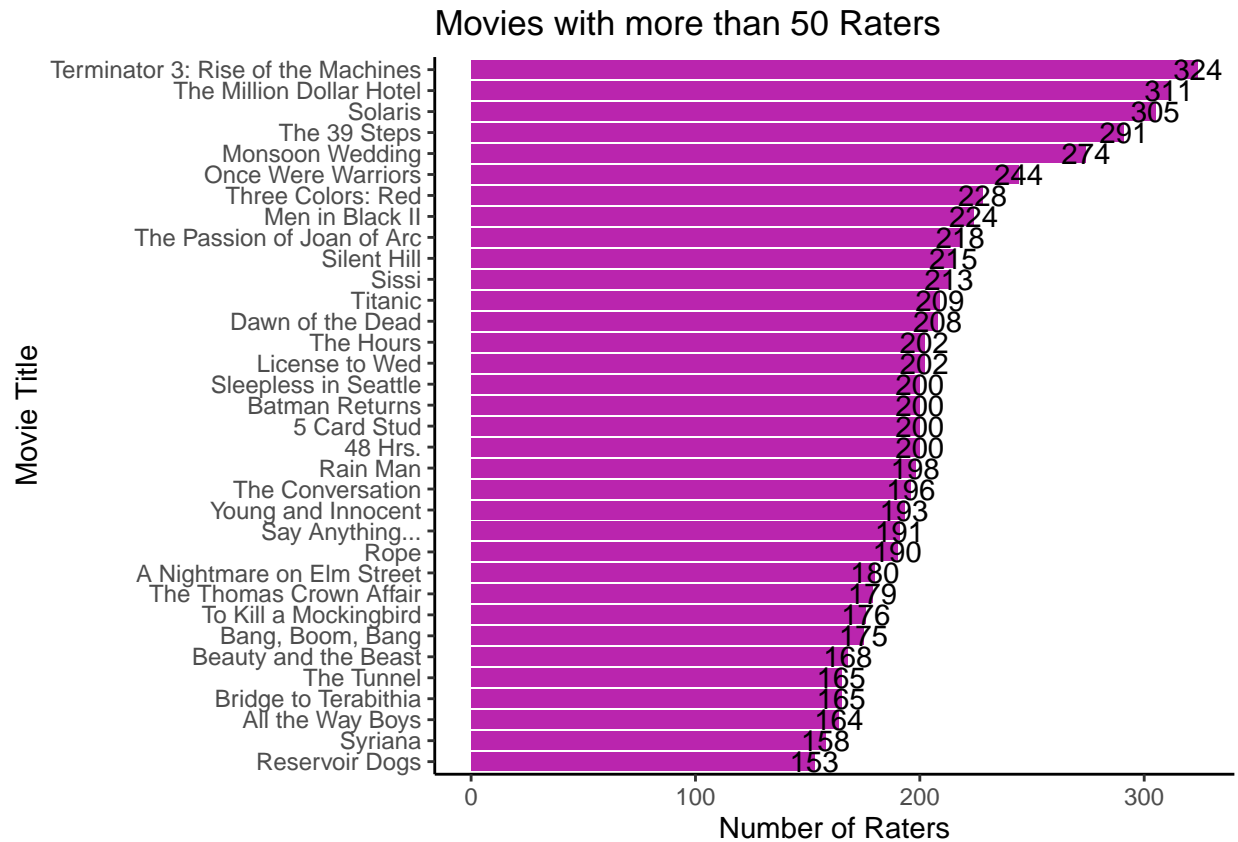
Number of Raters

```
## # A tibble: 1 x 1
##       n
##   <int>
## 1    671
```

Number of movies

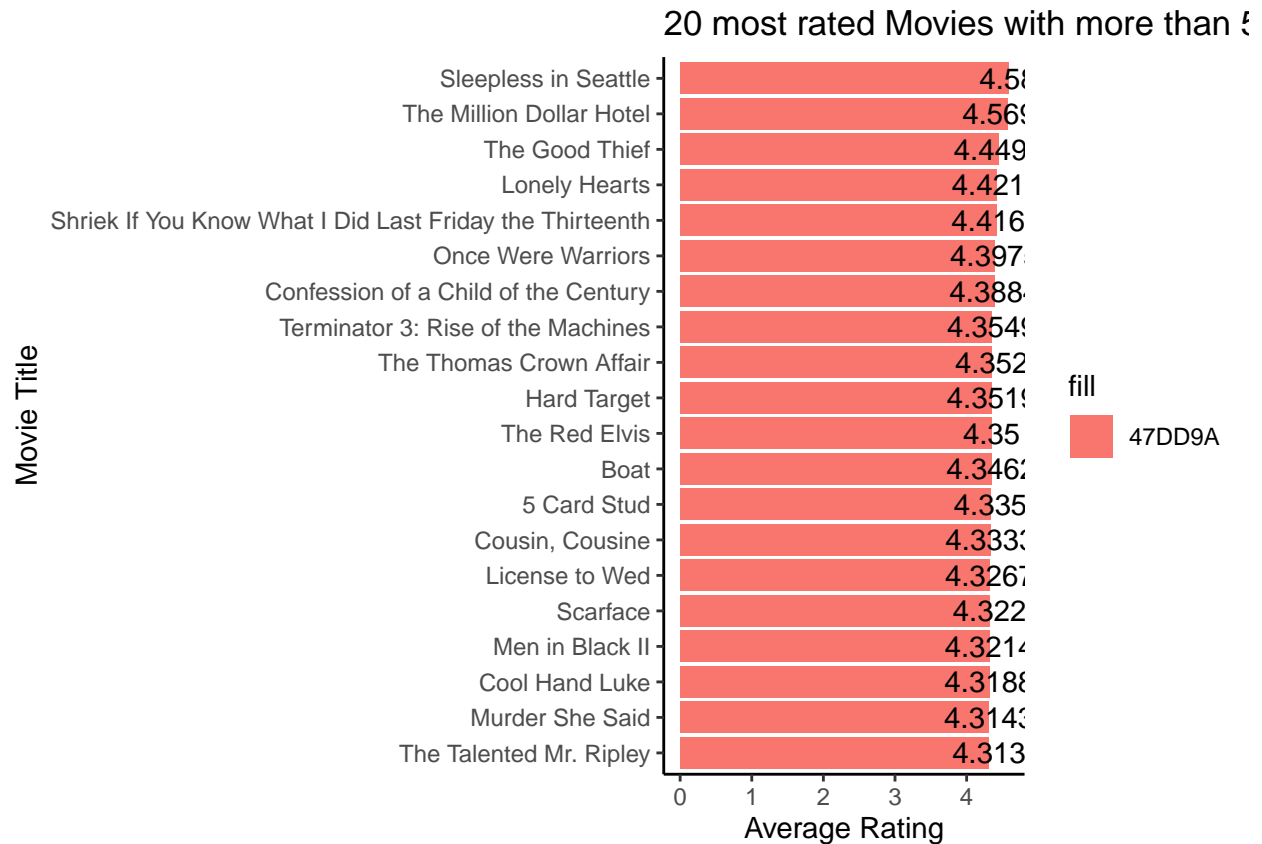
```
## # A tibble: 1 x 1
##       n
##   <int>
## 1   2785
```

What are the movies that have at least 50 Raters?



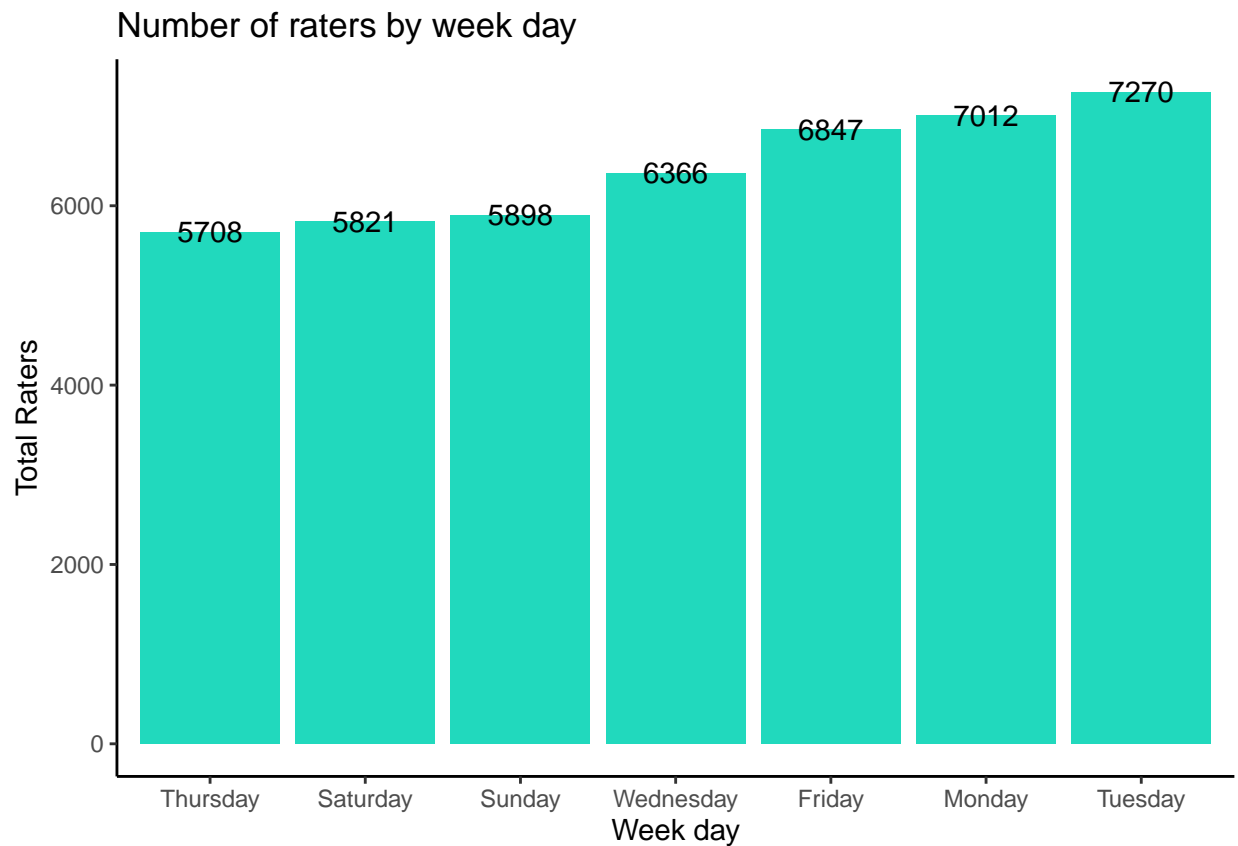
The display of first 35th movies that have a minimum of 50 raters sorted according to the number of raters of each movie. It revealed that Terminator 3: Rise of the Machines as the movie with the highest number of raters.

What is the most rated movie?



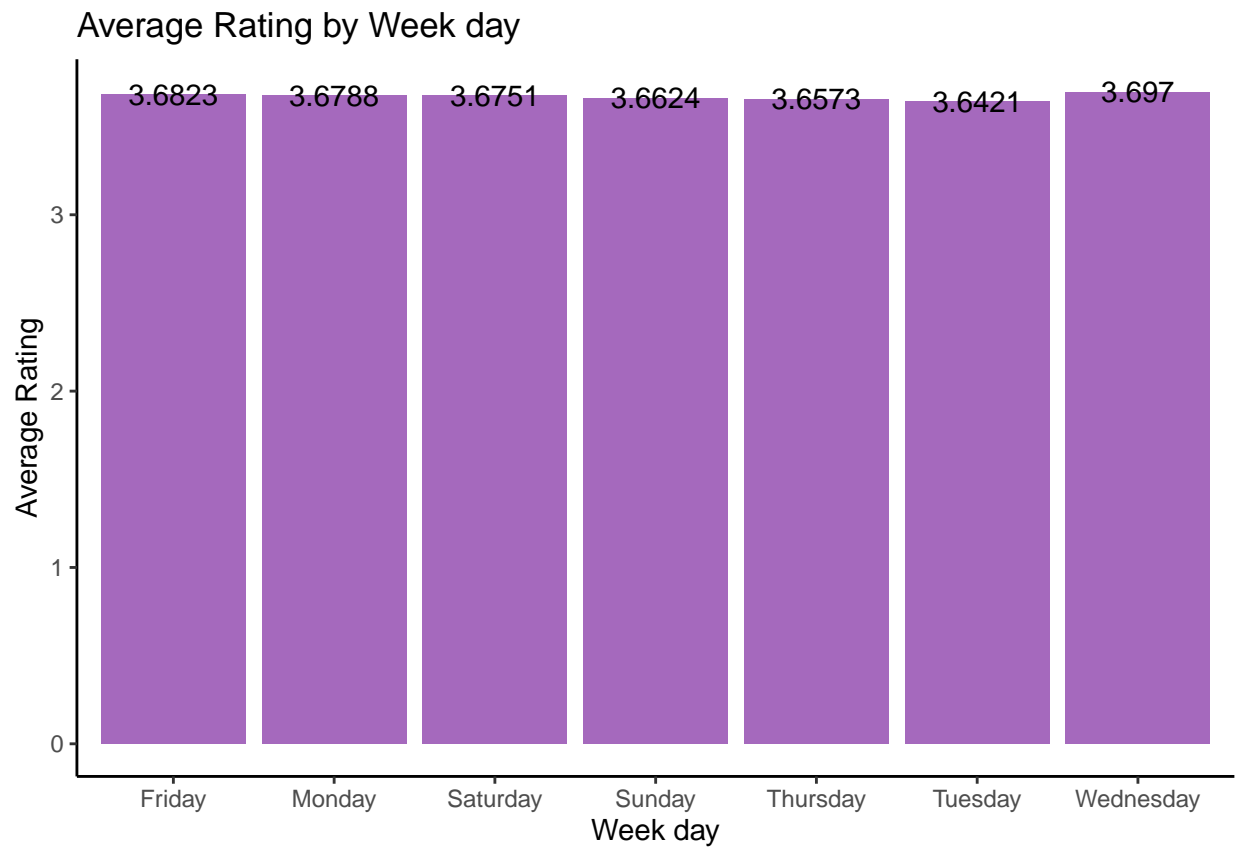
The average rating of each movie indicates that Sleepless in Seattle has the highest average rating. However, this average was taken for only the movies that have not less than 50 raters. I chose to select only movies with a minimum of 50 raters to reduce bias in the analysis. It is interesting to know that only 245 movies out of 2785 movies have 50 Raters and above.

What day of the week do People rate movies the most?



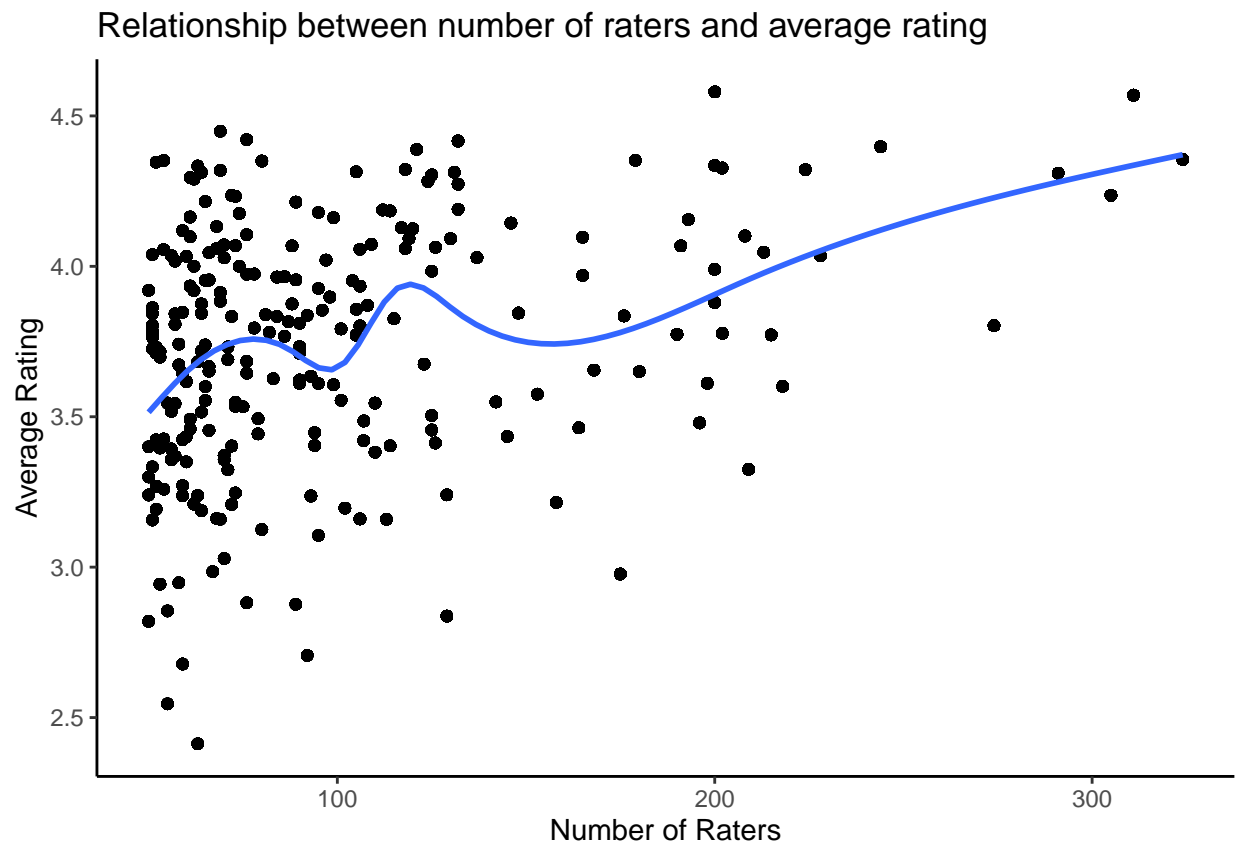
The count of the total number of raters that rated the movies shows that the highest number of rating was carried out on Tuesday. The lowest number of ratings was on Thursday. So, the range of number of raters between Tuesday and Thursday is 1562 raters. The range is a significant value. It makes sense to say that Tuesdays are the most favored days for movie rating. Meanwhile, I'm by no way saying that movie rating is mostly carried out by people on Tuesdays as this is on the realm of speculation until it is hypothesized and proven to be true.

What day of the week did People give highest rating?



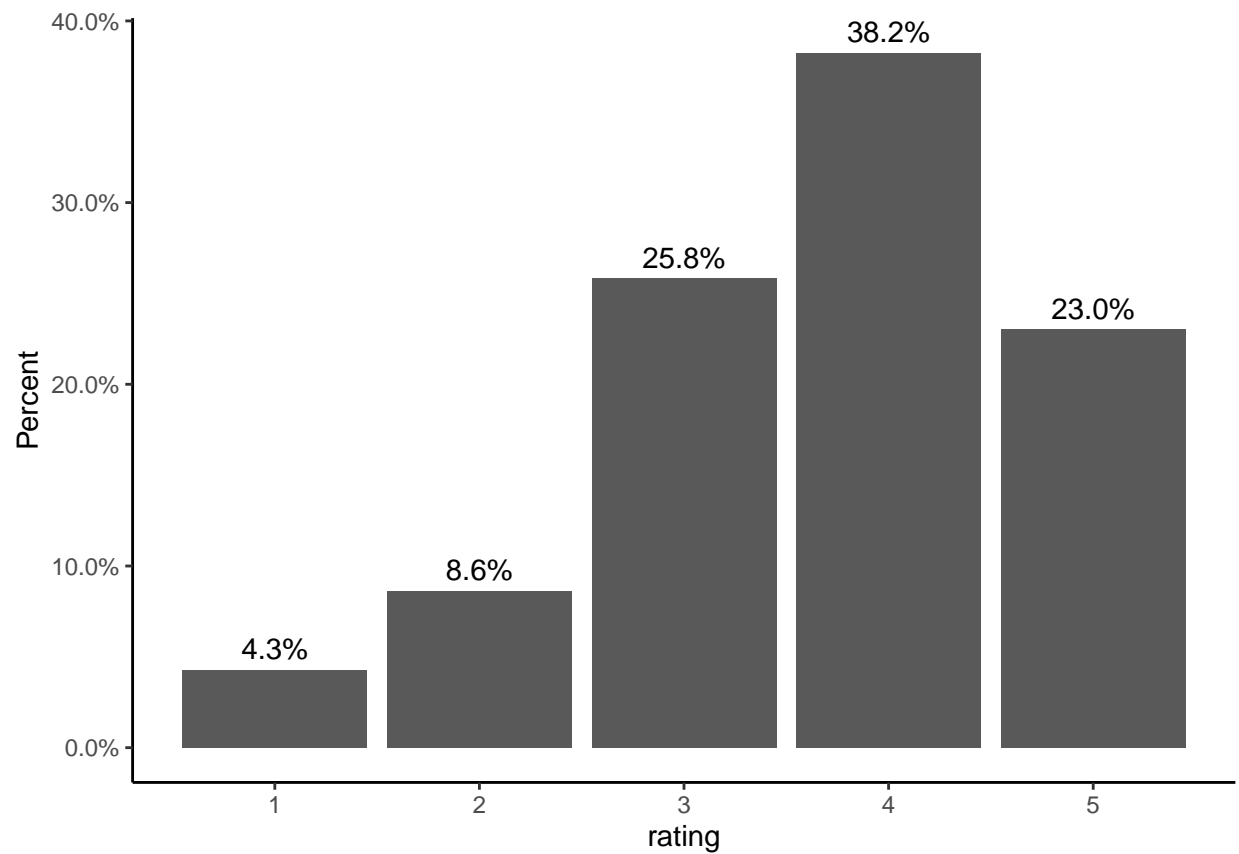
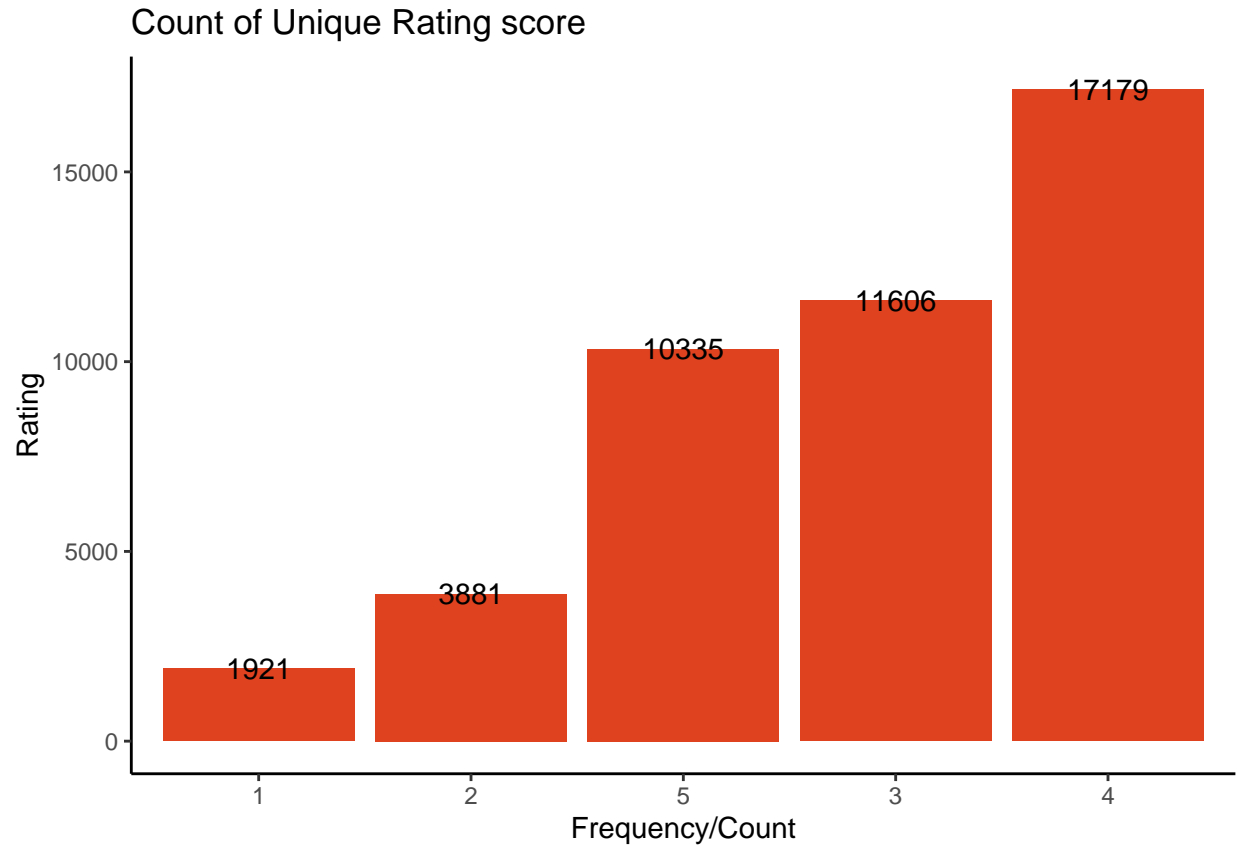
The average week day rating indicates that weekday does not play any role on the value of rating given by the raters. This is so because the average rating across board can be approximated to be same.

What is the relationship between rating score and number of Raters?



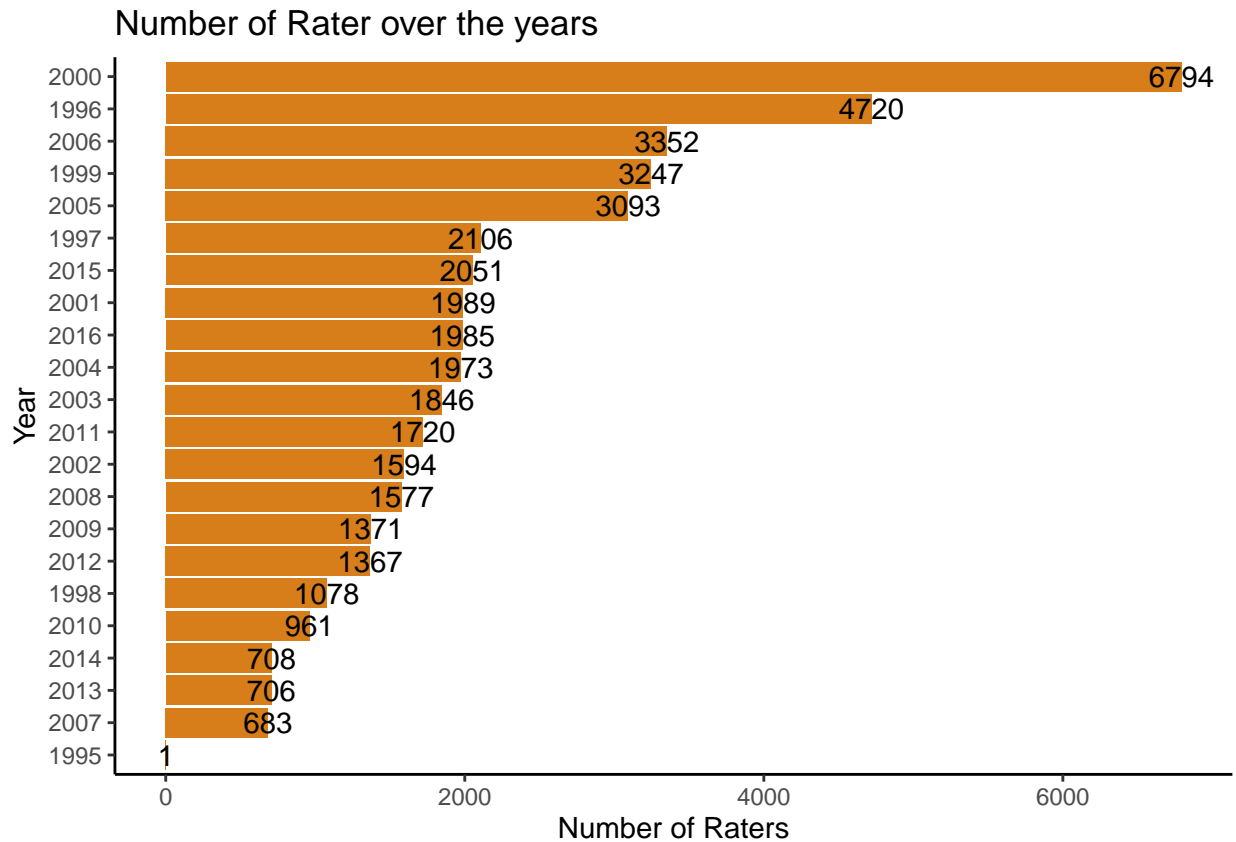
It is pertinent to view the relationship between the number of movie Raters and average rating. The relationship is positive. However, it will be naive to conclude based on this analysis that movie rating increases as number of Raters increase. Hypothesis testing is needed for such conclusion to be drawn.

What rating score is most prevalent?



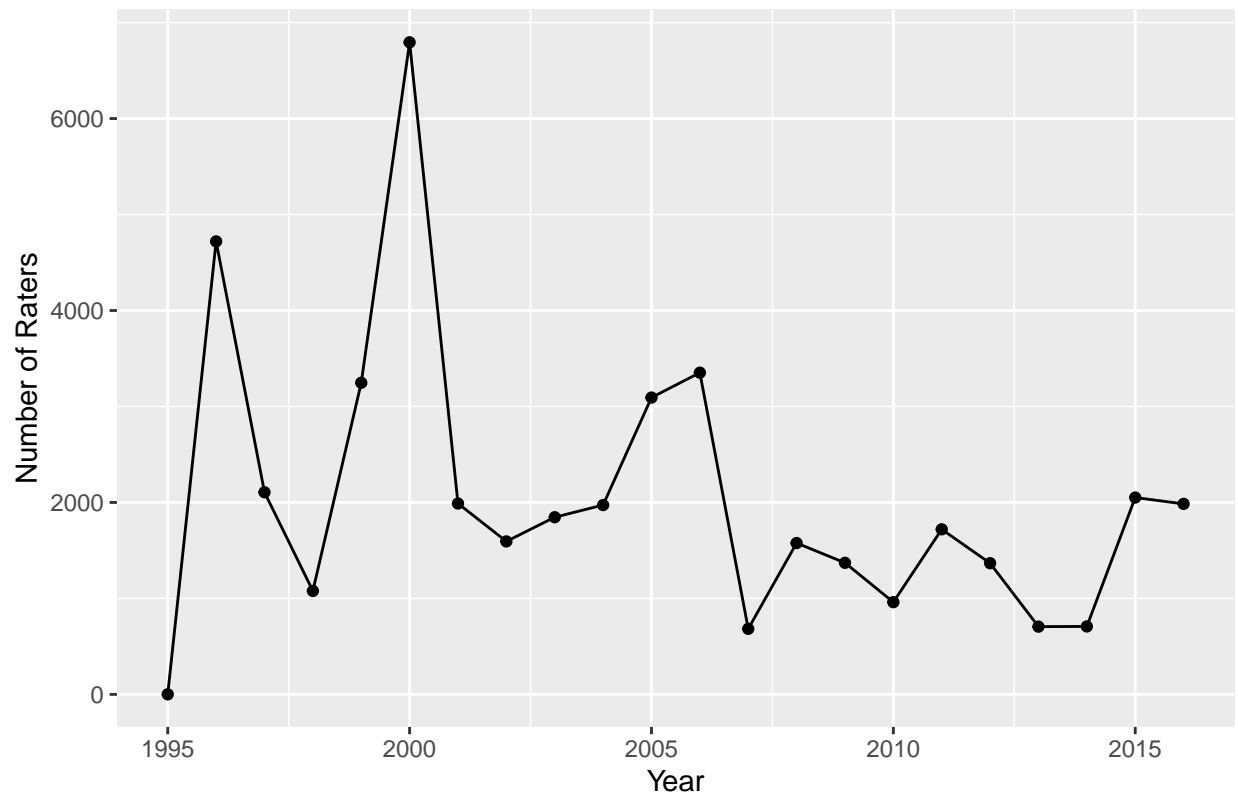
Score 4 is the most prevalent rating followed by 3. Score 4 is approximately 38% i.e 17179 of the total rating. Having the median of 4, mean of 3.671 and the 3rd quartile as 4 are testaments that majority of movies are rated 4. The lowest frequent movies rating is 1. This is 4.3% of the entire rating. However, a substantial number of movies received 5 rating. A total of about 10335 i.e 23% of the total movies surveyed.

What year did People rate movies the most?



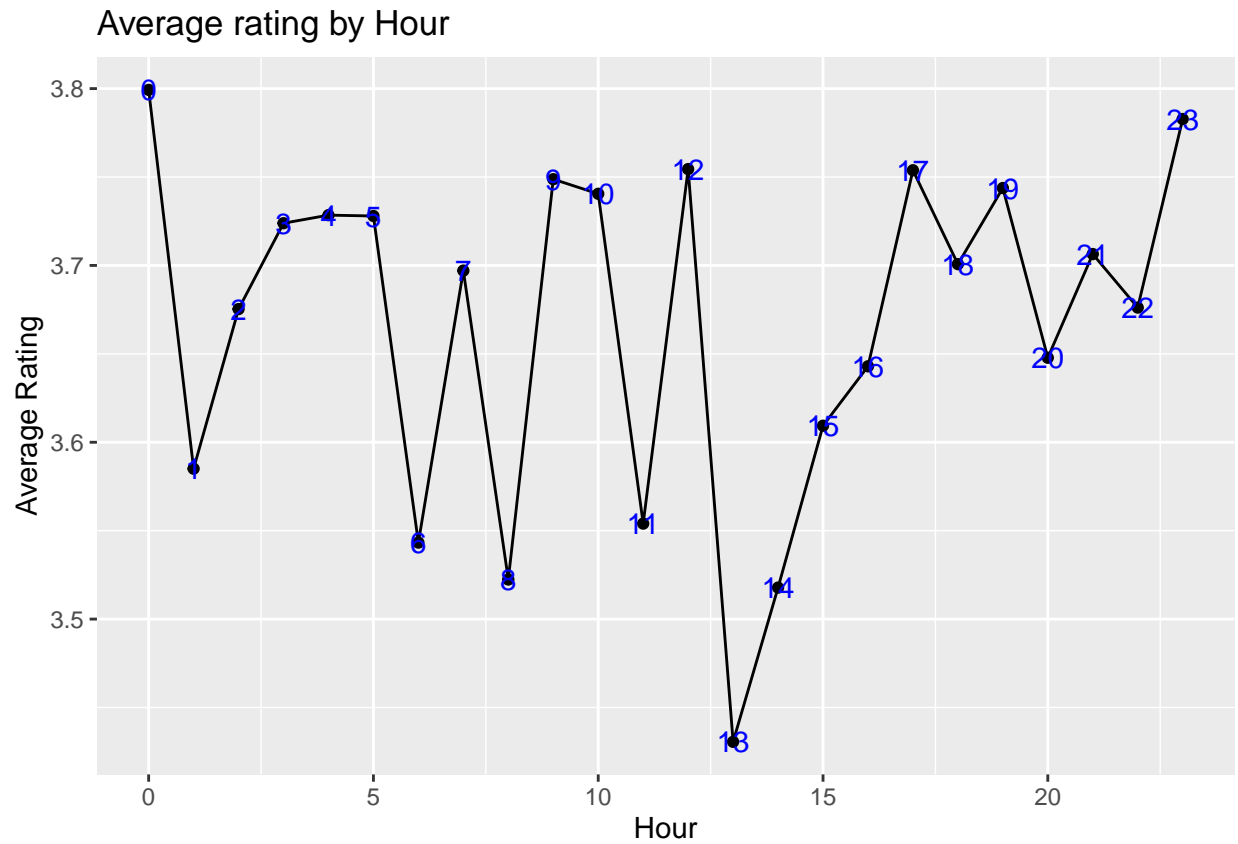
```
##  movieId      Title rating      Rating_date
## 1      21 The Endless Summer      3 1995-01-09 11:46:49
```


Distribution of Number of Rater over the years



Majority of the rating was given in year 2000 followed by year 1996. The total number of rating in year 2000 is 6794 which is approximately 15% of all ratings over the years. 1996 has 4720 number of rating which is about 11% of all rating that took place. However, 1995 has 1 rating for the movie The Endless Summer.

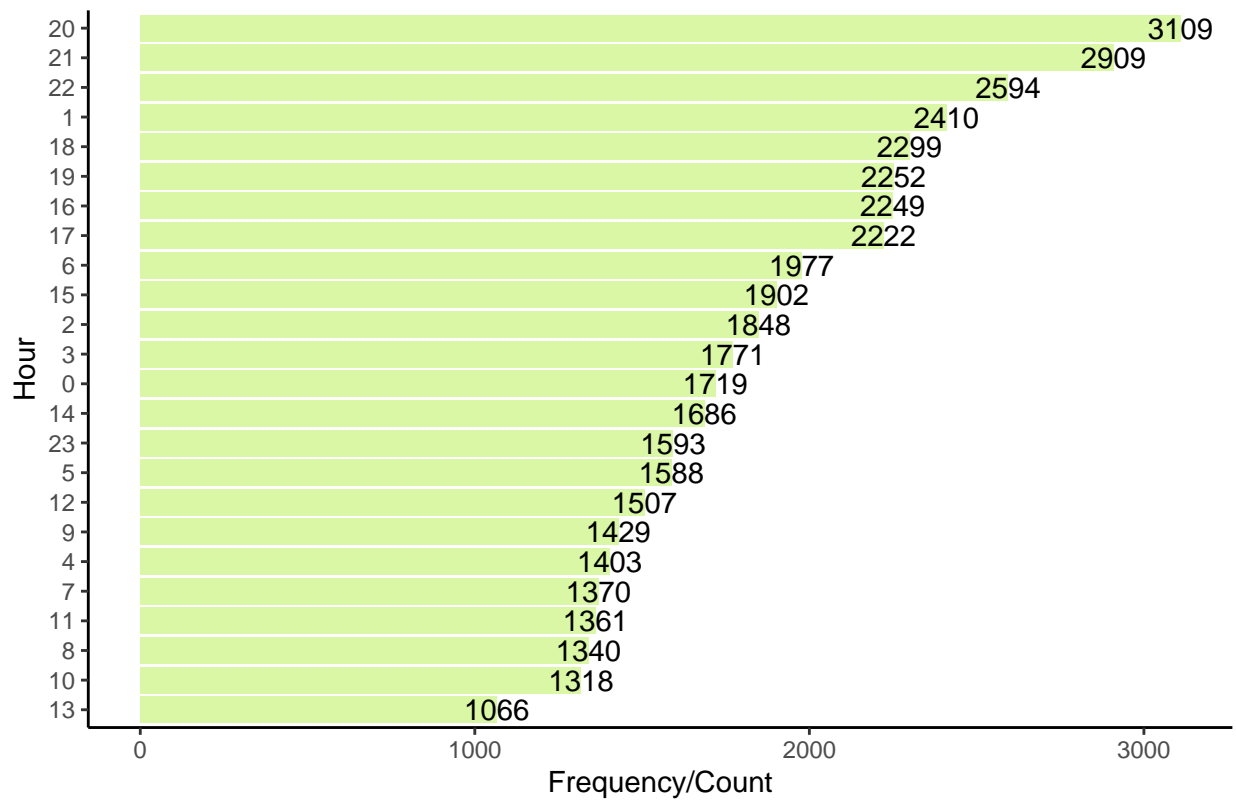
What Time of the day do People give highest rating score?



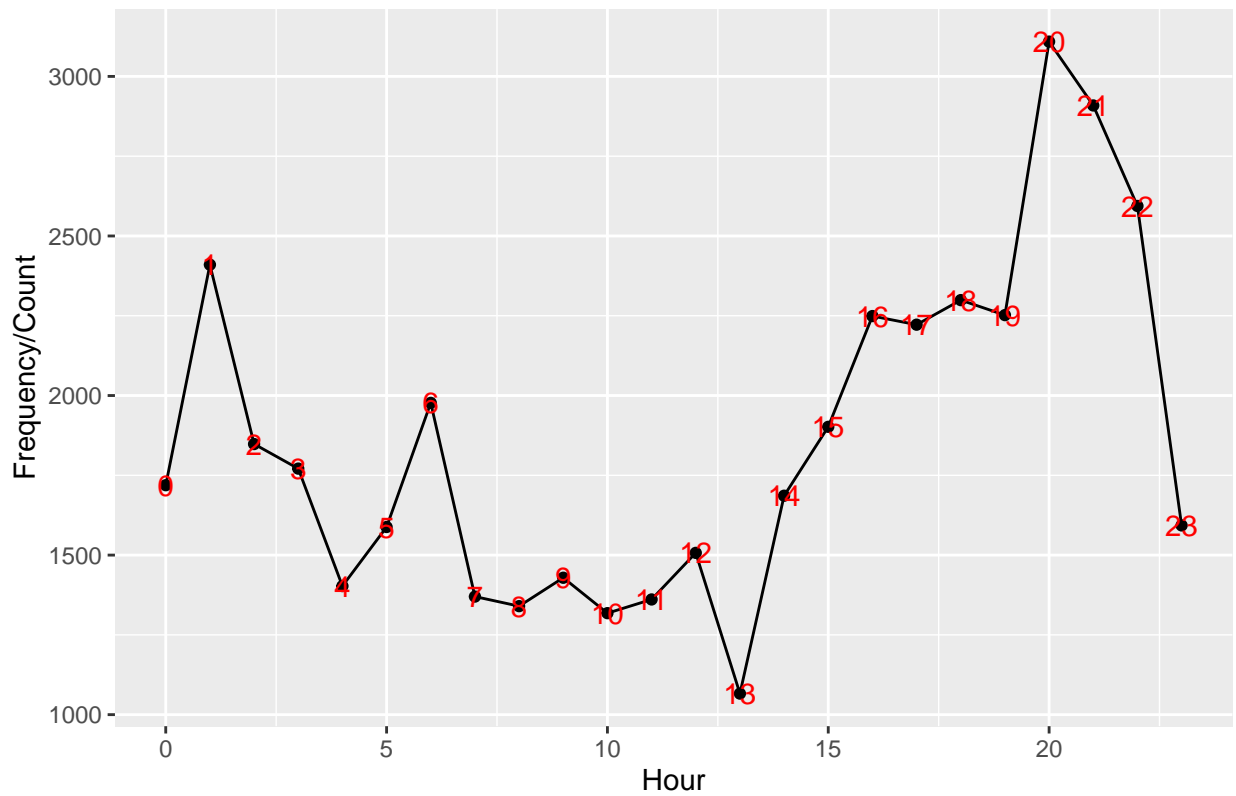
From the graph above, the raters who rated the movies between the hours of 23:00 to 00:00 ie 11pm to 12mid night gave the best rating score. The highest average rating of approximately 4 was given at 12 mid night. The lowest rating score was given at 13:00 (1pm). The average rating score given at 1pm is 3.43.

What Time of the day do People rate movies the most(frequency)?

Count of Raters by Hours



Count of Raters by Hours



Most Raters rated the movies at 20:00 to 22:00 (8pm-10pm) and 01:00am. The rare hour that raters rated any movie was at 13:00(1pm).

Note that the `echo = FALSE` parameter was added to the code chunk to prevent printing of the R code that generated the plot.