**1. Explore the data to obtain an understanding of users, movies and how users have rated movies.**

**a) OVERALL DISTRIBUTION OF RATINGS:**



***Figure 1.1***

Movies have been rated from values 1 to 5 by users where 1 stands for a low rating and 5 stands for a high rating.

From the above figure 1.1 , we can see that the distribution of ratings is slightly skewed to the left. This means that most movies have been rated high by users. Nearly 35% of the movies have a rating 4 and more than 25% have a rating of 3.

The average movie rating is 3.524.

**b) AVERAGE MOVIE RATINGS AND USER RATINGS:**

****

***Figure 1.2***

We can see from the above plot that a user has given an average rating of 3.6 to movies. However, a movie has received an average rating of only about 3.1. Most movies have lower average ratings. However, there is no particular trend visible in the average ratings given by users.

**c) NUMBER OF MOVIES RATED BY USERS:**

****

***Figure 1.3***

On an average users rate about 95 movies. The median of this distribution is 55. This means that most users rate a lower number of movies compared to the average. The range of the number of movies that users rate ,however, is very large. The minimum number of movies rated by a user is 10 and the maximum is 727.

**d)** **NUMBER OF RATINGS FOR MOVIES:**



***Figure 1.4***

The distribution of the number of ratings that movies get is highly right skewed. This means that most movies get a lower count of ratings. A movie get around 50 ratings on an average. However, the median of the distribution is much less at approximately 25.

The range of the number of ratings a movie gets ranges from 1 to 495.

**e) DISTRIBUTION OF RATING LEVEL AND TREND IN RATINGS:**

****

***Figure 1.5***

From the above graph we can infer the following :

* Less number of users have chosen to give lower rating to a movie.
* Most users have chosen to give an average or high rating to a movie.
* Most movies have received a rating of 3. A slightly lower number of movies have a rating of 2 and 4.
* The number of movies being rated low is much higher than the number of movies being rated high.
* Also, as is evident from the above figure and also from figure 1.1, the number of low ratings is lesser than high ratings.