**MOVIEPROJECT:-**

1. Datasets from all three years are grouped together in one large database.
2. A sample of random instances or records are drawn for each model such that clarity is maintained for 70 percent of the complete database.
3. The output variable is regarded to be the Movie ID which will act as a determining factor to assess what movie is the best using a classification algorithm.
4. Below are summarized scatter plots for the variables taken into account. Variables such as movie name, ZipCode, UserID, Hours are removed for simplicity. The Correlation to the movies can be made through the MovieIDs.



