**MACHINE LEARNING REPORT**

Objective:  
the dataset consists of 1000 observations with 16 variables (movie data)

We are trying find out the main factors influencing the rating of a movie, like what factors influence the most.

Does Rating Impact on Gross

Does cast impact on Gross Value.

We used boxplots, scatterplots to know how they are related to each other on a scale.

And we used K mean algorithm to know the cluster value and its characteristics.

A picture containing scatter chart

Description automatically generated

There are total 16 variables

Poster\_links: links address of every movie poster.

Series\_Title: Is the movie name which is character.

Released\_year: The movie release date.

Certificate: Type of movie based on the content (A, U/A,)

Runtime: Total time taken for each movie.

Genre: Each movie has its own genre like action, horror, drama …

IMDB ratings – rating for each movie by IMDB

Overview- is the summary of each movie

Meta score -Scores are assigned to movie's reviews of large group of the world's most respected critics, and weighted average are applied to summarize their opinions range.

Director- director names

Star1, star2, star3, star4- no of stars for each movie count

No\_of\_votes- votes for each movie

Gross- gross for each movie produced

Chart, histogram

Description automatically generated

Numbers of people who votes more as per rating, like people voted more in between 7.6 to 7.8

And people who gave rating 8.7 to 9 are very less which is less than 20 people.

Chart, pie chart

Description automatically generated

U classification type movie = 24%

A classification type movie = 22% followed by

UA classification type movie = 17%

So majority of the movies are from the above categories.

Chart, histogram

Description automatically generated

The frequency of rating is more left and less in the right

Boxplot of IMDb ratings

Chart, box and whisker chart

Description automatically generated

Chart, histogram

Description automatically generated

Histogram between meta score and no of people revied.

ALGORITHAM USED:

K means uses a two-phase iterative algorithm to minimize the sum of point-to-centroid distances, summed over all k clusters  
Chart, line chart

Description automatically generated

Using elbow method, the k value is 2

Chart, line chart

Description automatically generated

The k value is 2 in silhouette method, In both the methods the value is 2

Text, letter

Description automatically generated

Effecting the IMDB rating, meta score and no of votes, as you can see the no of votes are more, and meta score is more so the rating is high.

A screenshot of a computer

Description automatically generated with medium confidence

Cluster 2 is near to origin and have similar characteristics

Chart, scatter chart

Description automatically generated