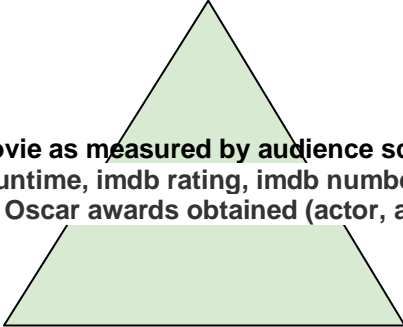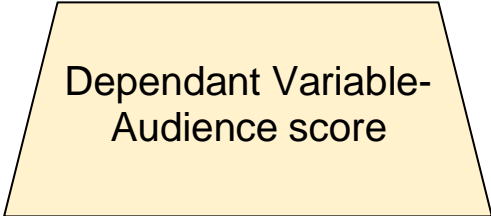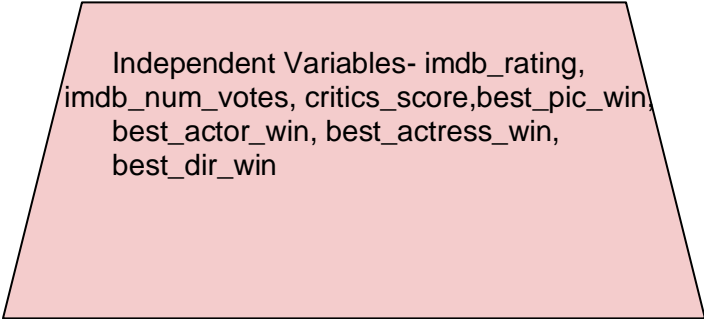**Structural Pyramid Analysis Plan**

Predict the popularity of a movie as measured by audience score related to the attributes like to the type of movie, genre, runtime, imdb rating, imdb number of votes, critics rating, critics score, audience rating, Oscar awards obtained (actor, actress, director and picture)

The SMART goal outlined above clearly pertains to audience score as a metric. What is less clear is how that parameter should be computed. Is a movie's popularity, as measured by audience score, related to the type of movie, genre, runtime, imdb rating, imdb number of votes, critics rating, critics score, audience rating, Oscar awards obtained (actor, actress, director and picture)? Being able to answer this question will help us to predict a movie's popularity.
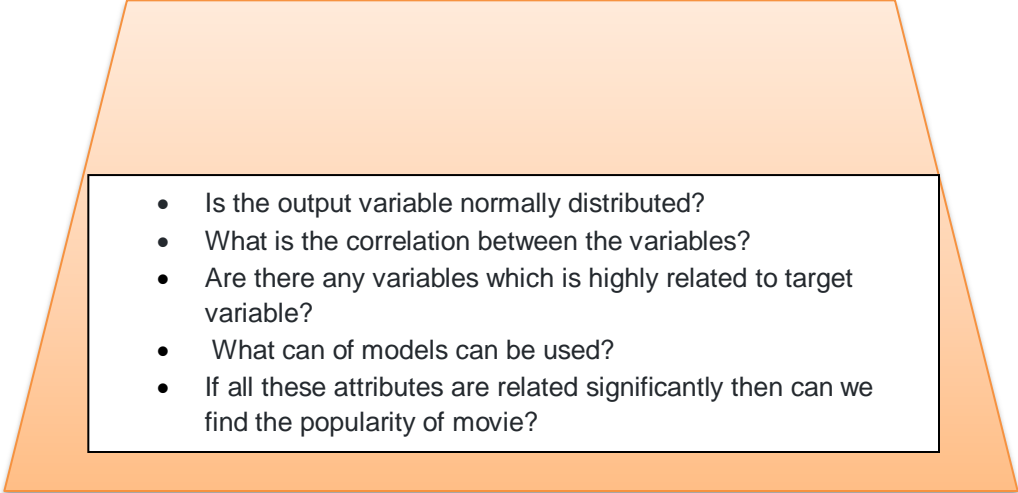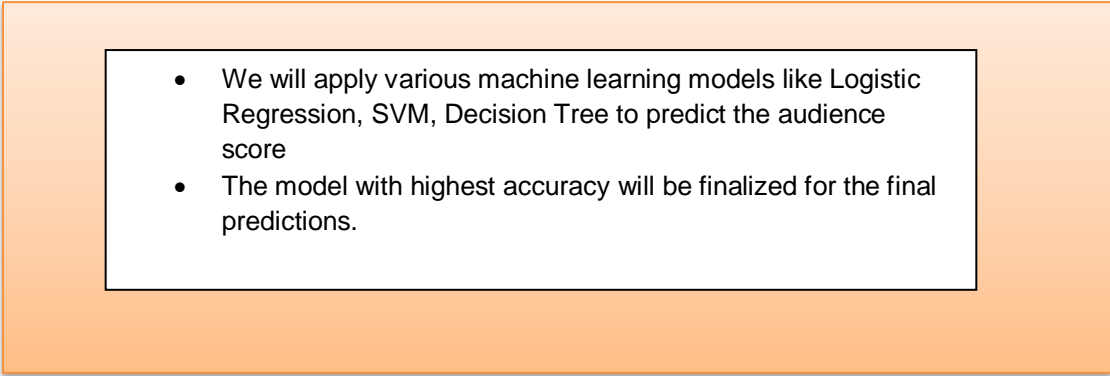
Dependant Variable-
Audience score

Next Step is to analyze the independent variables categories to be considered to predict the final audience score. Below are a few variables we will be considering for our analysis. We will be applying PCA to determine the significant variables that impact the audience score.

Independent Variables- imdb_rating, imdb_num_votes, critics_score,best_pic_win, best_actor_win, best_actress_win, best_dir_win

The next stage involves raising specific questions about possible interactions between the types of data.

- Is the output variable normally distributed?
- What is the correlation between the variables?
- Are there any variables which is highly related to target variable?
- What can of models can be used?
- If all these attributes are related significantly then can we find the popularity of movie?

- We will apply various machine learning models like Logistic Regression, SVM, Decision Tree to predict the audience score
- The model with highest accuracy will be finalized for the final predictions.