Step 1, Data Preprocessing and Exploratory Data Analysis:

We import the 'movie_metadata.csv' data.

There are 28 columns. We need to select the most useful features to do the visualization and prediction.

First, keep the numerical columns, drop the non-numerical and non-related columns, or we should assume the value to those columns, for example color (0/1), language (0/1), 'plot_keywords' and country.

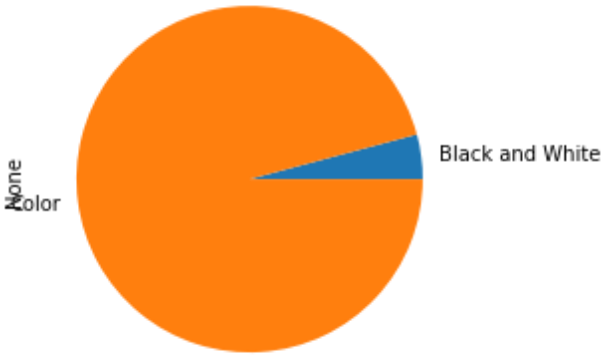
Second, some columns like director_name, actor_2_name, actor_1_name are either difficult to get the patterns or related to other columns like 'director_facebook_likes', 'actor_2_facebook_likes' and 'actor_1_facebook_likes'. So we could remove those colums.

Third, the movie_imdb_link, movie_title should be removed for the purpose of predicting the movie rating.

Finally, some columns are numerical but useless in prediction, like 'duration', 'facenumber_in_poster', 'title_year'.

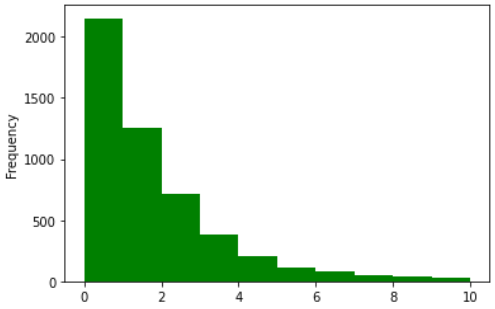
Step 2, visualization:

Pie chart for the color:

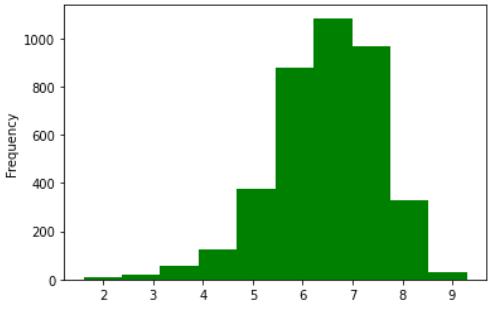


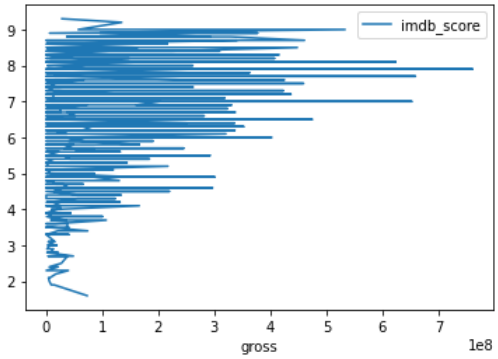
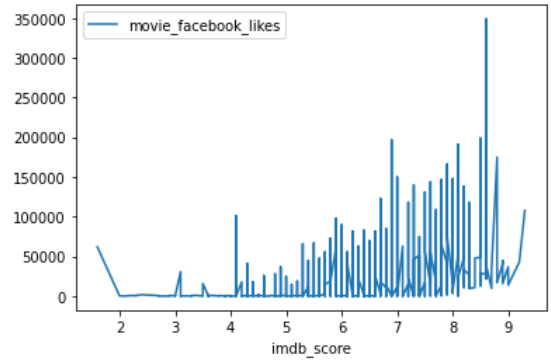
Pie chart for language:

Histogram for facenumber in poster:

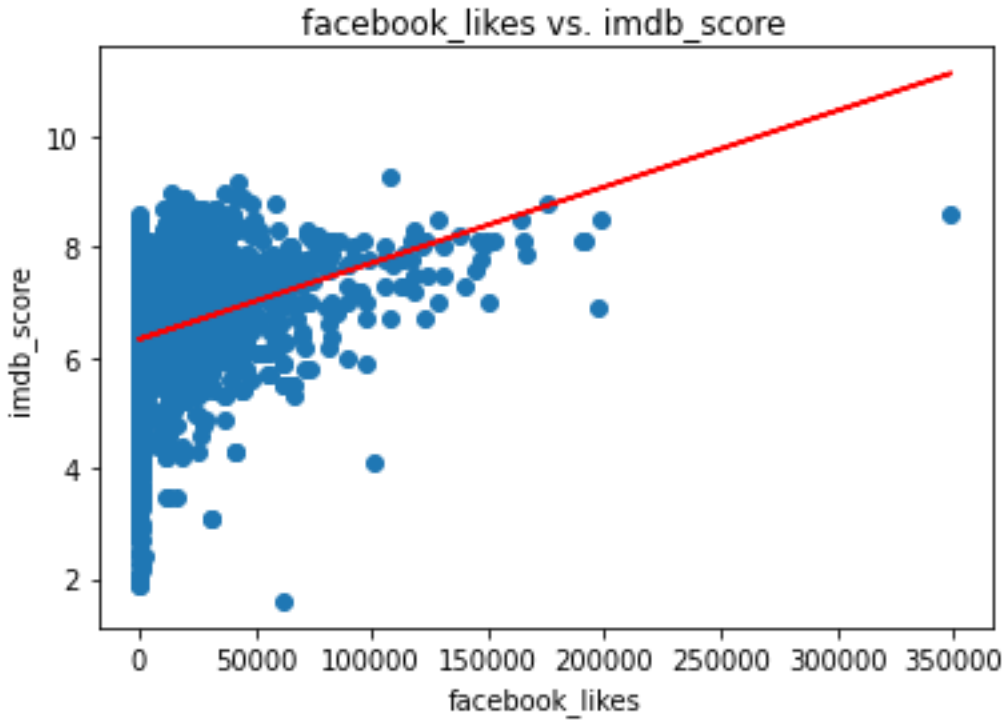


Histogram for the imdb score

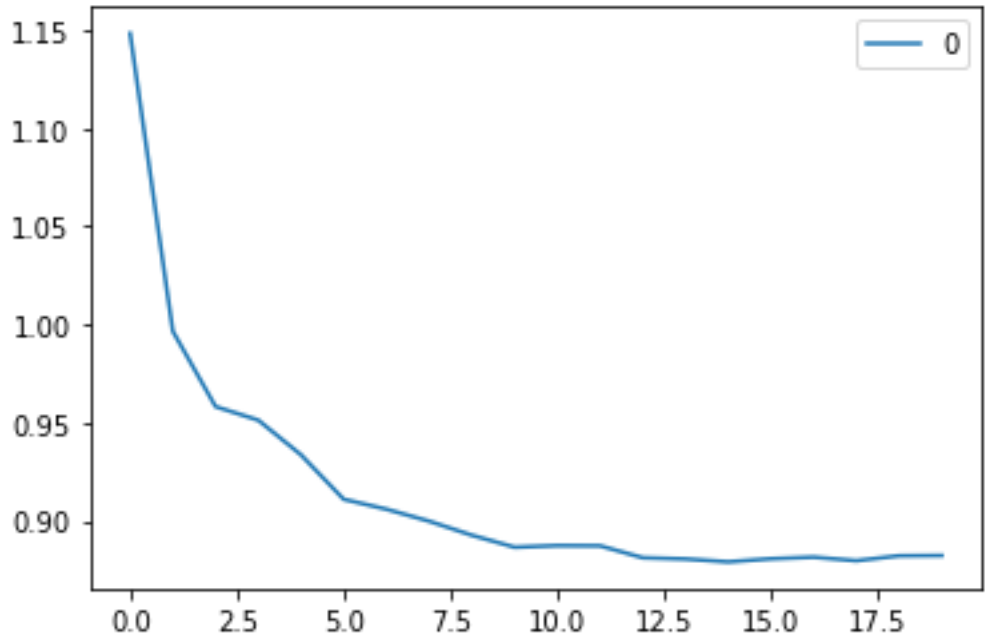


Line chart 'gross' and 'imdb_score'	Line chart 'facebook_likes' and 'imdb_score'
	

Step 3, Simple Linear Regression:



Step 4, K Nearest Neighbor:



So K = 9 is a better choice.

Step 5, Neural Network:

	precision	recall	f1-score	support
0	0.84	0.91	0.87	689
1	0.71	0.58	0.64	281
accuracy			0.81	970
macro avg	0.78	0.74	0.75	970
weighted avg	0.80	0.81	0.80	970

The precision, recall and f1-score seem good, Neural Network is a good model for prediction.