IMDB\_Movie\_dataset

Approach/Steps:

1. Extract the data from given source
2. Loaded into python pandas dataframe (movie\_df)
3. Understand the attributes of the data
4. Identified and replaced the null and blank values with zero (e.g.:- gross, budget) using fillna() function
5. Calculated Profit1, profit2, Gross profit from the following formula by using lambda function
6. Calculations are loaded into new column named *profit1 , profit2 , gross\_profit*
7. Exported the required data from master dataframe into *top\_ten\_genre*
8. Produced the result of top 10 genres based on the profitability
9. Exported the required data from master dataframe into *top\_ten\_directors* and *top\_ten\_actors*
10. Produced the result of top 10 actors and directors based on the profitability
11. Exported the required data from master dataframe into *best\_pair*
12. Eliminated the values of empty or blank values of director/actor by using *dropna()*
13. Produced the result of best pair actor and directors based on the imdb\_ratings
14. Created the result of best pair actor and directors average ratings

Library used: pandas

Used pandas’ functions:

dropna()

illna()

sort\_values()

mean()

head()

sum()

groupby()

apply()

lambda()

read\_csv()