



Kirinyaga University

SPE 2403: Data Analytics

PRACTICAL EXERCISE 2

DATE: 17th Oct 2022

Deadline: 18th Oct 2022 at 1200hrs

Consider the *Cinema.csv* data-set. This data-set contains data of over 46000 movies from 1874 to 2016 with their title, year of release, rating, runtime, genre, imdb Rating, and country.

- Read this data into Python Pandas as a DataFrame.
- Check column information for the data.
- Convert the unit of the budget and gross columns from \$ to million \$ first by dividing by 1,000,000.
- Rename the budget and gross columns to budget (Million \$) and gross (Million \$).
- Create a new column called profit which contains the difference of the two columns: gross (Million \$) and budget (Million \$).
- Sort the data-frame using the profit column as reference.
- Extract the top ten profiting movies in descending order and store them in a new data-frame called top10.
- Plot a scatter or a joint plot between the columns budget and profit. Make any observations on the scatter graph.
- Extract the movies with a negative profit and store them in a new data-frame called Neg_profit.
- Plot a bar graph of genre_1 movies.