**Use the Bollywood Dataset to Answer Questions 1 to 12.**

The data file *bollywood.csv* contains box office collection and social media promotion information about movies released in 2013−2015 period.

Following are the columns and their descriptions.

|  |  |  |
| --- | --- | --- |
|  | **Feature** | **Description** |
| 1 | SlNo |  |
| 2 | Release Date | Date of release |
| 3 | MovieName | Name of the movie |
| 4 | ReleaseTime | Mentions special time of release. LW (Long weekend), FS (Festive Season), HS  (Holiday Season), N (Normal) |
| 5 | Genre | Genre of the film such as Romance, Thriller, Action, Comedy, etc. |
| 6 | Budget | Movie creation budget |
| 7 | BoxOfficeCollection | Box office collection |
| 8 | YoutubeViews | Number of views of the YouTube trailers |
| 9 | YoutubeLikes | Number of likes of the YouTube trailers |
| 10 | YoutubeDislikes | Number of dislikes of the YouTube trailers |

1. How many records are present in the dataset? Print the metadata information of the dataset.
2. How many movies got released in each genre? Which genre had highest number of releases? Sort number of releases in each genre in descending order.
3. How many movies in each genre got released in different release times like long weekend, festive season, etc. (Note: Do a cross tabulation between *Genre* and *ReleaseTime.*)
4. Which month of the year, maximum number movie releases are seen? (Note: Extract a new column called month from *ReleaseDate* column.)
5. Which month of the year typically sees most releases of high budgeted movies, that is, movies with budget of 22.5 crore or more?
6. Which are the top 5 movies with maximum return on investment (ROI)? Calculate return on investment (ROI) as (*BoxOfficeCollection – Budget) / Budget.*
7. Do the movies have higher ROI if they get released on festive seasons or long weekend? Calculate the average ROI for different release times.
8. Draw a histogram and a distribution plot to find out the distribution of movie budgets. Interpret the plot to conclude if the most movies are high or low budgeted movies.
9. Compare the distribution of ROIs between movies with comedy genre and drama.

Which genre typically sees higher ROIs?

1. Is there a correlation between box office collection and YouTube likes? Is the correlation positive or negative?
2. Which genre of movies typically sees more YouTube likes? Draw boxplots for each genre of movies to compare.
3. Which of the variables among *Budget, BoxOfficeCollection, YoutubeView, YoutubeLikes, YoutubeDislikes* are highly correlated? Note: Draw pair plot or heatmap.

Email: ashok.21aug@gmail.com