Lab 3 – SECTION A, BATCH 2 Date:10th Sept. 2022

Exer 1 : Perform data analysis and visualization on the movies.csv data set and answer the given questions.

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. :

1. SlNo
2. Release Date
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

Use Python code to answer the following questions:

1. How many records are present in the dataset?
2. How many movies got released in each genre? Sort number of releases in each genre in descending order.
3. Which genre had highest number of releases?
4. 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.)
5. Which month of the year, maximum number movie releases are seen? (Note: Extract a new column called month from ReleaseDate column.)
6. Which month of the year typically sees most releases of high budgeted movies, that is, movies with budget of 25 crore or more?
7. Which are the top 10 movies with maximum return on investment (ROI)? Calculate return on investment (ROI) as (BoxOfficeCollection – Budget) / Budget.
8. Do the movies have higher ROI if they get released on festive seasons or long weekend? Calculate the average ROI for different release times.
9. Is there a correlation between box office collection and YouTube likes? Is the correlation positive or negative?
10. Which genre of movies typically sees more YouTube likes? Draw boxplots for each genre of movies to compare.
11. Which of the variables among Budget, BoxOfficeCollection, YoutubeView, YoutubeLikes, YoutubeDislikes are highly correlated? Note: Draw pair plot or heatmap.
12. During 2013−2015 period, highlight the genre of movies and their box office collection? Visualize with best fit graph.
13. Visualize the Budget and Box office collection based on Genre.
14. Find the distribution of movie budget for every Genre.
15. During 2013−2015, find the number of movies released in every year. Also, visualize with best fit graph.

Exer 2: Business Analytics

Perform analysis on the NORTHWIND (COMBINED) data set using the pivot tables and charts in MS Excel.

1. Identify the top 5 and bottom 5 selling products in the company.
2. Identify the top 5 selling products and the salesmen who sell them.
3. Tabulate the total sales of each product, ship country wise.
4. Tabulate the total sales of "Boston Crab Meat" , customer wise.
5. Tabulate the customer's region wise sales of products in each category.
6. Visualize the customer's region wise sales of products in each category using an appropriate chart.
7. Visualize the total sales of each product, employee wise with an appropriate chart.
8. Tabulate the total sales of each product, category-wise as a percentage of the entire sales.
9. Visualize the total sales of each product, category-wise as a percentage of the entire sales.
10. Summarize the sales for each product, year wise and visualize the same in an appropriate chart.