

2nd Semester B. Tech
DSE 1271 Data Visualization [1 0 3 2]
Week 10 & 11 Assignment:
Marks(10)

Task: For the dataset given (SET 5), answer the questions with a report that contains a story, visuals, and data summaries to provide actionable insights into the data.

SET 5: Bollywood.csv:

The data file contains box office collection and social media promotion information about movies released in 2013–2015 period. Following are the columns and their descriptions.

- *SINo* – Release Date
- *MovieName* – Name of the movie
- *ReleaseTime* – Mentions special time of release. LW (Long weekend), FS (Festive Season), HS
 - (Holiday Season), N (Normal)
- *Genre* – Genre of the film such as Romance, Thriller, Action, Comedy, etc.
- *Budget* – Movie creation budget
- *BoxOfficeCollection* – Box office collection
- *YoutubeViews* – Number of views of the YouTube trailers
- *YoutubeLikes* – Number of likes of the YouTube trailers
- *YoutubeDislikes* – Number of dislikes of the YouTube trailers

Question:

1. Perform data exploration of Products, product categories, Customers, and Orders and describe the dataset.
(2 M)
2. Perform the required data cleaning for missing or inconsistent data. **(2 M)**
3. Use Python and libraries like Pandas, Matplotlib and Seaborn to create insightful tabulations and visualization for the following:

(4 M for analysis and 2 M for report and notebook submission)
(6 M)

1. Movies & Genres :

Example Insights: How many movies got released in each genre? Which genre had highest number of releases? How many movies in each genre got released in different release times ?

2. Box Office :

Example Insights: Which month of the year the maximum number of movie releases are seen? Which month of the year typically sees most releases of high-budgeted movies, that is, movies with a budget of 25 crore or more? 6. Which are the top 10 movies with maximum return on investment (ROI)? Calculate return on investment (ROI) as

$(\text{BoxOfficeCollection} - \text{Budget}) / \text{Budget}$. 7. Do movies have higher ROI if released on festive seasons or long weekends? Calculate the average ROI for different release times.

3. Social Media:

Example Insights: Is there a correlation between box office collection and YouTube likes? Is the correlation positive or negative? Which genre of movies typically sees more YouTube likes? Draw boxplots for each genre of films to compare. Which variables among Budget, BoxOfficeCollection, YoutubeView, YoutubeLikes, YoutubeDislikes are highly correlated?

Submission on LMS

- Python notebook to showcase the process of exploring, cleaning, and analysis of the dataset. From the data summaries, charts/plots and analysis of the dataset, gather the insights and write as markdown (in text) in the notebook.
- A report (in pdf) containing a narrative considering the audience is the Sales manager. You are presenting the findings and insights from the data analysis of the Northwind dataset.
- The report should contain sections on
 1. Objective of the analysis
 2. Data Exploration
 3. Data Cleaning
 4. Analysis and Insights
 5. Conclusion