MCA 5141 – Machine Learning Lab
Week – 3

EXERCISE 1: Data Analysis using mtcars

1. Find the car with the best mpg
2. Find the car with the worst mpg
3. Find the car with the best horsepower
4. Find 5 number summary of displacement
5. Find median horse power
6. What is average mpg for manual vs. automatic cars
7. Draw a histogram of miles per gallon
8. Boxplot of mpg for each cylinder type
9. Create a crosstab displaying count of automatic vs. manual cars
10. Create a crosstab displaying count of "am vs cyl"
11. What is the correlation between the weight of the car and mpg

EXERCISE 2: Descriptive Analytics and Visualization

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:

- SlNo
- 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

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.