DSE 2141– Data Analytics Lab
Lab 4 – Date: 23rd August 2023

EXERCISE 1: Data Analysis using CEREALS dataset

Use the prepared or preprocessed data to answer the following:

1. Cross tabulate the type of cereal (hot vs cold) against the manufacturer
2. Which is the cereal with the best rating, worst rating?
3. Plot a side-by-side boxplot comparing the consumer rating of hot vs. cold cereals.
4. Is there a relation between sugars, calories, carbs, and fat?
5. Which manufacturers produce cereal with highest calories?
6. Use correlation tests and visualization to identify if the two variables calories and consumer rating associated?
7. Use correlation tests and visualization to identify if the two variables shelf and consumer rating associated?
8. Is there a relation between manufacturer and rating?
9. Which nutrients are essential for a good rating for a cereal?
10. Design a Linear regression model to predict the rating of a cereal based on top 3 related nutrients. Tabulate the accuracy of the model using an 80, 20 split.

EXERCISE 2 – Descriptive Analytics and Visualization using Matplotlib, Seaborn: (Cross tabulation, distributions, Multi-variate analysis, Various Plots)

Use the IPL datasets and answer the following:

1) Count the total number of matches conducted in the year 2008
2) Find the city name where maximum and minimum number of matches conducted.
3) Find total count of matches city wise.
4) Find the Team which is maximum and minimum toss winner.
5) Check the toss decision that the team has taken.
6) Count the total number of normal and tie matches.
7) Find the team names where the match result is tie.
8) Find the team name who won the match by highest runs.
9) Find the team name who won the match by lowest runs.
10) Find the players who was awarded "Player of the match" more than 3 times.
11) Find the player who was awarded as player of the match maximum times.
12) Find the Venue where the team won the match by highest runs.

13) Find the Venue where the team won the match by lowest runs.
14) Find the Umpires who did umpiring maximum times.
15) Find the Total matches played in each season
16) Find the Total runs in each season
17) No. of tosses won by each team
18) Visualize the Toss decision across seasons
19) Find the Dismissal Kind and Visualize using best fit graph
20) Find the Top 10 run scorers in IPL and Visualize using best fit graph
21) Visualize the Highest MOM award winners
22) Find Total Number of Played Matches by each team
23) Compare Total Played Matches vs Winning Matches vs Win Rate
24) Find the Distribution of Won the Matches
25) Ratio between Total Matches and Win Matches
26) What is the choice of each team after winning the toss?