

Machie Learning Lab
IV Sem, DSE, Feb 2023

Week - 02

mtcars.csv dataset

1. Tabulate number of cylinders in the dataset.
2. Find the five number summary of milespergallon
- ~~3. Draw a histogram for hp~~
- ~~4. Box plot the miles per gallon~~
5. Find the avg weight of all cars
6. Find the car with the minimum displacement
7. Find the car with the maximum qsec.
8. Find the median horse power and find the car with the highest fuel efficiency
9. Find the car with the lowest fuel efficiency
10. Find the car with the best hp
11. Tabulate mpg for different number of gears
- ~~12. Do side by side boxplot of mpg for cars with am (automatic transmission) Vs cars with Manual transmission.~~

Cereals1 data set

- 1) Tabulate the following attributes: mfr, and type of cereals
- 2) Display the 5 number summary for all nutritional attributes starting from calories to vitamins. *(Note: A value of -1 in nutrients indicates a missing observation.)*
- 3) For missing values find and replace with arithmetic mean of the attributes.
- 4) Find and replace outliers with median.
- 5) Compare the 5 number summary before and after preprocessing.
- ~~6) Draw side by side box plots of Calories of Hot Vs cold cereals.~~
- 7) Are the attributes calories and consumer rating correlated?
- 8) Are the attributes mfr and consumer rating correlated?
- 9) Which is the best Vs worst cereal in terms of user rating?
- 10) Which is the best Vs worst cereal in terms of calories?
- 11) Rate the top 5 cereals in terms of user rating?
- 12) Identify the cereal with the highest sodium.
- 13) Identify the cereal with the lowest carbohydrate.
- 14) Are the variable shelf and sugar correlated?
- 15) Identify the manufacturer of the cereal with the highest sugar content.
- 16) Replace missing values with mean
- 17) Find outliers using $1.5 * IQR$