Assignment

Dataset

- Consider the dataset in the file "*HousingPrice.csv*"
- The dataset contains the following field;
  - Area
  - Bedrooms
  - Bathrooms
  - Stories
  - Main Road (Y/N)
  - Basement (Y/N)
  - Hotwaterheating(Y/N)
  - Airconditioning (Y/N)
  - Parking
  - Furnishing Status (furnished / semi-furnished /unfurnished)
  - Price

Assignment - II

- Read the data using pandas
- Show the top 10 rows (hint: *head* function)
- Drop the following columns
  - Guestroom
  - Basement
  - Hotwaterheating
  - Airconditioning

Assignment - III

- Some columns have missing values
  - Identify them
  - Hint : Missing values are represented by ?
- How will you handle missing values?
  - You can replace them by calculating the average of the remaining value of that column.
  - Will this be a good strategy?
- How will you deal with missing values in "*price*" column?
  - Drop the whole row?
- How will you deal with columns with Y/N or categorical data?
  - MainRoad (Y/N)
  - Furnishing Status (furnished / semi-furnished /unfurnished)

Assignment - IV

- Try some visualization (matplotlib)
  - Line Charts
    - X-Axis = Area, Y-Axis = Price
    - X-Axis = Bedrooms, Y-Axis = Price
    - X-Axis = Bedrooms, Y-Axis = Area
    - X-Axis = Bedrooms, Y-Axis = Stories
    - X-Axis = Area, Y-Axis = Stories
  - Do you see any trends?
- Develop a linear regression model. Use all remaining columns. Identify the error.