

**KANWAL SHEHZADI 02-131212-027**

**INTRODUCTION TO DATA SCIENCE**



**ASSIGNMENT 01 INTRODUCTION TO DATA SCIENCE**

For the text file provided with the assignment,construct regular expressions to extract the following :

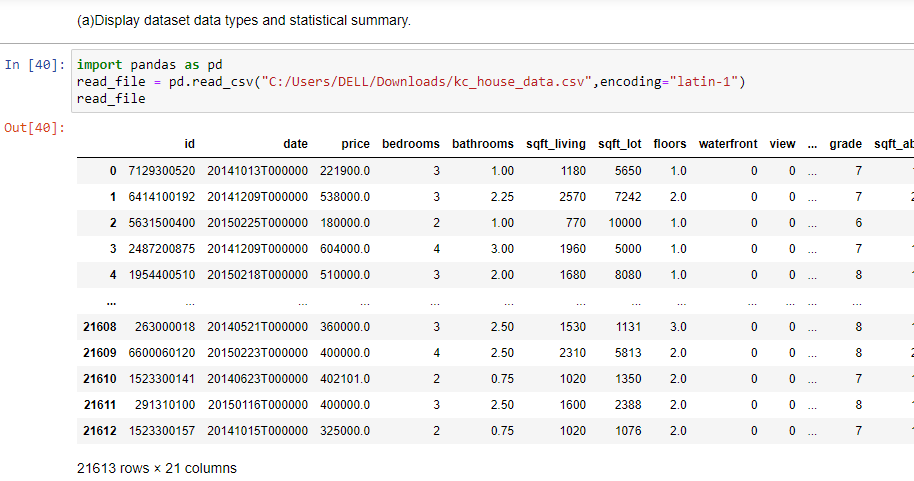
1. **All number :** \d+
2. **All e-mail ids :** b[\w.\_%+-][+@[\w.-]+\.[A-Za-z]{2,}\b](mailto:+@[\w.-%5d+\.%5bA-Za-z%5d%7b2,%7d\b)
3. **All URLs :** \bhttps|http\:\/\/www\.[\w.-\_ ]+\.[A-Za-z]{2,}\b

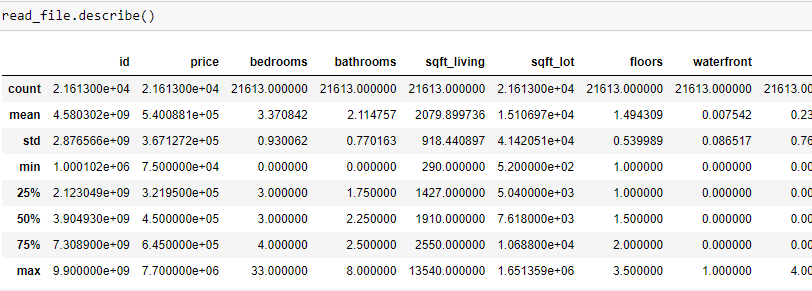
2. Download the dataset in csv format of house sales from the following website:

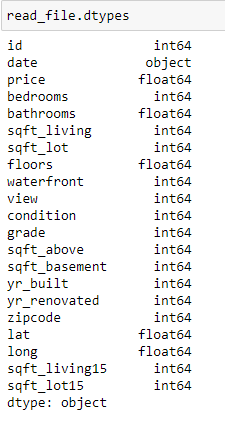
<https://www.kaggle.com/datasets/harlfoxem/housesalesprediction>

Write Python code for the following:

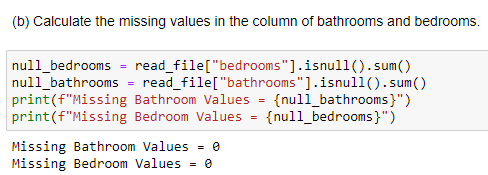
1. Display dataset data types and statistical summary.



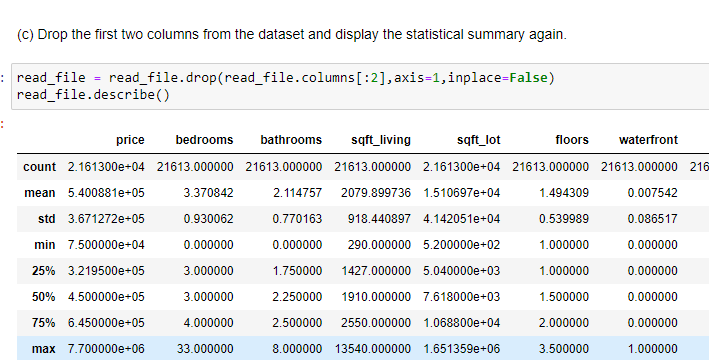




1. Calculate the missing values in the column of *bathrooms* and *bedrooms*.



1. Drop the first two columns from the dataset and display the statistical summary again.



(d) Count the number of houses with unique zip code and display the data in data frame.

