Linear Regression on Medical Insurance Dataset

Objective:

- To understand the structure of the dataset
- To implement simple linear regression and predict values
- To implement multiple linear regression & predict values

Medical Insurance Dataset:

The details regarding this 'Insurance' dataset are present in the data dictionary

	age	sex	bmi	children	smoker	region	charges
0	19	female	27.900	0	yes	southwest	16884.92400
1	18	male	33.770	1	no	southeast	1725.55230
2	28	male	33.000	3	no	southeast	4449.46200
3	33	male	22.705	0	no	northwest	21984.47061
4	32	male	28.880	0	no	northwest	3866.85520
5	31	female	25.740	0	no	southeast	3756.62160
6	46	female	33.440	1	no	southeast	8240.58960
7	37	female	27.740	3	no	northwest	7281.50560
8	37	male	29.830	2	no	northeast	6406.41070
9	60	female	25.840	0	no	northwest	28923.13692
10	25	male	26.220	0	no	northeast	2721.32080
11	62	female	26.290	0	yes	southeast	27808.72510
12	23	male	34.400	0	no	southwest	1826.84300

Lab Environment: Jupyter Notebook

Domain: Medical

Tasks to be performed:

- Read the .csv file and understand the structure of the dataset
  - Using the seaborn library make a pair plot for all of the columns
- Building simple linear models:
  - Divide the dataset into train & test sets in 80:20 ratio.
  - Build a simple linear model where the dependent variable is 'charges' & the independent variable is 'age'. Predict the values on the test set and find the root mean square error
- Build multiple linear models:
  - Build a multiple linear model where the dependent variable is 'charges' & the independent variables are 'children' & 'bmi'. Predict the values on the test set and find the root mean square error.