Regression Assignment

1.) **Problem Statement:**
A client wants to predict the insurance charges based on the several parameters.
Stage 1: Machine Learning
Stage 2: Supervised Learning
Stage 3: Regression

2.) **Total number of rows:** 1338
Total number of columns: 6

3.) Since it is a nominal data, we are using **ONE-HOT ENCODING** as a pre-processing method.

4.) **Final model used:** Support Vector Machine with C-value:5000 and kernel used:**rbf**

R_score: 0.874777
5.) **r_score values:**

Multiple Linear Regression - 0.7894790349867009

Support Vector Machine

S.No.	**Penalty (C)**	**Kernel**			
		Linear	**rbf**	**Poly**	**Sigmoid**
1	100	0.628879	0.320031	0.617956	0.527610
2	1000	0.764931	0.810206	0.856648	0.287470
3	5000	0.741417	0.874777	0.859565	-7.530043

Decision tree

S.No.	**Criterion**	**Splitter**	**R score**
1	squared_error	best	0.6833247504174381
	squared_error	random	0.728775182120118
2	friedman_mse	best	0.6804108541827755
	friedman_mse	random	0.6869117093862669
3	absolute_error	best	0.6502539814719586
	absolute_error	random	0.6716193204403982
4	poisson	best	0.7237932999215247
	poisson	random	0.7237932999215247

Random forest:

n_estimators	**r_score**
50	0.8498329315421834
100	0.8538307913484513
1000	0.8541778123151671

6.) **Final model used:** Support Vector Machine with C-value:5000 and kernel used:**rbf**

Because this model has more r_score value, compared to other models.