1) Stage 1: Machine Learning (Inputs are in Number)

Stage 2: Supervised Learning (Inputs and Output are well known)

Stage 3: Regression (Output is in Number)

2) 1338 rows

6 columns

3) One Hot Encoding

4) To find the following machine learning regression method using R^2 value

1. MULTIPLE LINEAR REGRESSION (R^2 value) = 0.7895

2. SUPPORT VECTOR MACHINE:

S.NO	HYPER PARAMETER	LINEAR (r^2 value)	RBF (NON LINEAR) (r^2 value)	POLY (r^2 value)	SIGMOID (r^2 value)
1	C10	0.4624	-0.0323	0.0387	0.0393
2	C100	0.6289	0.3200	0.6180	0.5276
3	C500	0.7631	0.6643	0.8264	0.4446
4	C1000	0.7649	0.8102	0.8566	0.2874
5	C2000	0.7440	0.8548	0.8606	-0.5940
6	C3000	0.7414	0.8663	0.8599	-2.1244

The SVM Regression use R^2 value (rbf and hyper parameter (C3000)) = 0.8663

3. DECISION TREE:

S. NO	CRITERION	MAX FEATURES	SPLITTER	R^2 VALUE
1	mse	auto	best	0.6952
2	mse	auto	random	0.6892
3	mse	sqrt	best	0.6974
4	mse	sqrt	random	0.6942
5	mse	log2	best	0.6802
6	mse	log2	random	0.6540
7	mae	auto	best	0.6739
8	mae	auto	random	0.7553
9	mae	sqrt	best	0.7291
10	mae	sqrt	random	0.7025
11	mae	log2	best	0.7487
12	mae	log2	random	0.7314
13	friedman_mse	auto	best	0.6798
14	friedman_mse	auto	random	0.7045
15	friedman_mse	sqrt	best	0.7082
16	friedman_mse	sqrt	random	0.6639
17	friedman_mse	log2	best	0.7390
18	friedman_mse	log2	random	0.6306

The Decision Tree Regression use R^2 value (mae, auto, random) = 0.7553

4. RANDOM FOREST:

S. NO	CRITERION	MAX FEATURES	n_estimators	R^2 VALUE
1	mse	auto	50	0.8498
2	mse	auto	100	0.8538
3	mse	sqrt	50	0.8696
4	mse	sqrt	100	0.8701
5	mse	log2	50	0.8696
6	mse	log2	100	0.8701
7	mae	auto	50	0.8527
8	mae	auto	100	0.8520
9	mae	sqrt	50	0.8708
10	mae	sqrt	100	0.8711068
11	mae	log2	50	0.8708
12	mae	log2	100	0.8711068
13	friedman_mse	auto	50	0.8501
14	friedman_mse	auto	100	0.8541
15	friedman_mse	sqrt	50	0.8702
16	friedman_mse	sqrt	100	0.871105
17	friedman_mse	log2	50	0.8702
18	friedman_mse	log2	100	0.871105

The Random Forest Regression use R^2 value (mae, sqrt, n=100) = 0.8711068

R^2 value (mae, log2, n=100) = 0.8711068

6) I am choosing Random Forest Regression as my final model because of it's R^2 value is larger than any other models.