1) The problem statement is to predict the insurance charges based on the several parameters.

2) 6 columns and 1339 rows

3) Categorical Column – Nominal – One Hot Encoding

4)

1. **MULTIPLE LINEAR REGRESSION (R^2 value)=** 0.7894

2. **SUPPORT VECTOR MACHINE:**

SNO	HYPER PARAMETER	LINEAR (R VALUE)	RBF (NON LINEAR)(R VALUE)	POLY (R VALUE)	SIGMOID (R VALUE)
1	C10	0.4624	-0.0322	0.0387	0.0393
2	C100	0.6288	0.3200	0.6179	0.5276
3	C500	0.7631	0.6642	0.8263	0.4446
4	C1000	0.7649	0.8102	0.8102	0.2874
5	C2000	0.7440	0.8547	0.8605	-0.5939
6	C3000	0.7414	0.8663	0.8598	-2.1244

The **SVM Regression** R^2 value (Non linear (Rbf) and hyper parameter (C3000))= 0.8663

3. DECISION TREE:

SL.NO	CRITERION	MAX_FEATURES	SPLITTER	R VALUE
1	squared_error	auto	best	0.6906
2	squared_error	auto	random	0.6981
3	squared_error	sqrt	best	0.7119
4	squared_error	sqrt	random	0.7232
5	squared_error	log2	best	0.7028
6	squared_error	log2	random	0.7060
7	friedman_mse	auto	best	0.7138
8	friedman_mse	auto	random	0.6950
9	friedman_mse	sqrt	best	0.7285
10	friedman_mse	sqrt	random	0.6260
11	friedman_mse	log2	best	0.6491
12	friedman_mse	log2	random	0.6868
13	absolute_error	auto	best	0.6697
14	absolute_error	auto	random	0.7586
15	absolute_error	sqrt	best	0.7162
16	absolute_error	sqrt	random	0.6761
17	absolute_error	log2	best	0.6561
18	absolute_error	log2	random	0.7210

The **Decision Tree** use R^2 value (Criterion (absolute error), Max features (auto), Splitter(random))= 0.7586

4. RANDOM FOREST:

SL.NO	CRITERION	MAX_FEATURES	N_ESTIMATORS	R VALUE
1	squared_error	auto	10	0.9252
2	squared_error	auto	100	0.9460
3	squared_error	sqrt	10	0.5191
4	squared_error	sqrt	100	0.7591
5	squared_error	log2	10	0.5191
6	squared_error	log2	100	0.7591
7	absolute_error	auto	10	0.9281
8	absolute_error	auto	100	0.9459
9	absolute_error	sqrt	10	0.7210
10	absolute_error	sqrt	100	0.7857
11	absolute_error	log2	10	0.7210
12	absolute_error	log2	100	0.7857
13	friedman_mse	auto	10	0.9206
14	friedman_mse	auto	100	0.9412
15	friedman_mse	sqrt	10	0.5272
16	friedman_mse	sqrt	100	0.7608
17	friedman_mse	log2	10	0.5272
18	friedman_mse	log2	100	0.7608
19	poisson	auto	10	0.9304
20	poisson	auto	100	0.9413
21	poisson	sqrt	10	0.7520
22	poisson	sqrt	100	0.7717
23	poisson	log2	10	0.7520
24	poisson	log2	100	0.7717

The **Random Forest** use R^2 value (Criterion (absolute_error), Max_Features (auto), n_estimators (100))= 0.9459

5) My final model is **Random Forest** because the value is closer to 1 and is the highest among all the machine learning regression models.