Problem Statement or Requirement:

A client's requirement is, he wants to predict the insurance charges based on the several parameters. The Client has provided the dataset of the same. As a data scientist, you must develop a model which will predict the insurance charges.

Subset of the given Dataset:

age	sex	bmi	children	smoker	charges
19	female	27.9	0	yes	16884.92
18	male	33.77	1	no	1725.552
28	male	33	3	no	4449.462
33	male	22.705	0	no	21984.47

Solution:

The Client's expectation is to predict the insurance charges based on the given data. So it is clear that the column 'charges' will be the output and the remaining columns will be the input.

With that in mind different models are created and the output was verified with r2 value.

1. The Linear Regression and Multiple Linear Regression is created and the output is verified against R2 value.
   Linear Regression R2 Value =
   Multiple Linear Regression R2 Value =
2. SVM

S.No	Kernel	C	R2 Value
1	linear	0.2	-0.124139887
2	rbf	0.2	-0.089444508
3	sigmoid	0.2	-0.08977804
4	poly	0.2	-0.084311029
5	linear	1	-0.111661287
6	linear	0.1	-0.122076684
7	linear	0.02	-0.084456245
8	linear	0.00002	-0.089670876

3. Decision Tree

S.No	Criterion	MAX_Features	Split	max_depth	R2 Value
1	absolute_error		best	7	0.8462391
2	absolute_error		best	None	0.6813335
3	absolute_error		random	7	0.8479104
4	absolute_error		random	None	0.720914
5	squared_error		best	7	0.7710478

6	squared_error		random	7	0.8560644
7	squared_error		best	None	0.7021979
8	squared_error		random	None	0.7042655
9	friedman_mse		best	7	0.7907647
10	friedman_mse		best	None	0.6903859
11	friedman_mse		random	7	0.8142985
12	friedman_mse		random	None	0.69696
13	poisson		best	7	0.8222267
14	poisson		best	None	0.7177371
15	poisson		random	7	0.862218
16	poisson		random	None	0.623141
17	squared_error	sqrt	random	None	0.7175588
18	squared_error	log2	best	None	0.6932822
19	absolute_error	sqrt	best	None	0.7277527
20	absolute_error	log2	best	None	0.7009231
21	friedman_mse	sqrt	best	None	0.7453327
22	friedman_mse	log2	best	None	0.749795

4. Random Forest

S.No	n_estimators	random_state	criterion	max_features	R2 Value
1	20	0	absolute_error	sqrt	0.86454664
2	1	0	absolute_error	sqrt	0.74823322
3	45	0	absolute_error	sqrt	0.87099647
4	45	0	squared_error	sqrt	0.86901172
5	20	0	friedman_mse	sqrt	0.8594757
6	40	0	friedman_mse	log2	0.86916225
7	40	0	poisson	log2	0.86407311

As Random Forest gives the nearest R2 value 0.86916225, we can choose Random Forest as the best model for our project “Insurance Premium Evaluation”