1.) Identify the problem statement

- Dataset is very clear and provided input and output data,so this is supervised learning
- We need to predict the insurance charges,so this will come under regression
- There are so many input so we will not use simple linear regression

2.) About the dataset

- Dataset have 1337 rows and 6 column

3.) pre-processing the dataset

- There are 2 string columns that need to be converting into numeric
- Columns Sex and smoker are non orderable so it will come under nominal data
- For nominal data we need to use one-hot coding method
- Input columns are 'age', 'bmi', 'children', 'sex_male', 'smoker_yes'
- Output column is 'charges'

R2 Value:

- Multiple linear regression we got R2 value is 0.78
- SVM Module R2 value tuning using below parameter

Finding Best SVM Module using R2 value

kernel	C	coef0	degree	R2 value
linear	1000			0.76
poly	1000			0.85
rbf	1000			0.21
sigmoid	1000			0.85
sigmoid	1000	0.5		-1.62
poly	1000	0.5		0.87
poly	1000	0.5	3	0.87
poly	1000	0.5	1	0.63
poly	1000		1	0.87
poly	1000	155.5		0.86
linear	2000			0.74

We got high r2 value =0.87 for kernel=poly ,C=1000 parameter and coef0=0.5,so this will be the best model

Finding Best Decision Tree Module using R2 value

R2 will change depends up on various hyper tuning parameter

criterion	splitter	min_samples_split	ccp_alpha	random_state	R2 value
squared_error					0.67
squared_error	best	2	0.0	None	0.67
squared_error	***random***				0.7
squared_error	***random***	14	0.5		0.84
squared_error	***random***	5	0.5	10	0.85
friedman_mse					0.69
friedman_mse	best	2	0.0	None	0.69
friedman_mse	***random***				0.73
friedman_mse	***random***	14			0.84
friedman_mse	***random***	14	0.5		0.85
friedman_mse	***random***	14	0.5	10	0.85
absolute_error					0.69
absolute_error	best	2	0.0	None	0.69
absolute_error	***random***				0.73

absolute_error	***random***	14	0.5	10	0.85
absolute_error	***random***	14			0.87
absolute_error	***random***	14	0.5		0.88
poisson					0.67
poisson	best	2	0.0	None	0.67
poisson	***random***				0.71
poisson	***random***	14			0.82
poisson	***random***	14	0.5		0.86
poisson	***random***	14	0.5	10	0.85

We got high r2 value =0.88 for the hyper tuning parameter criterion=absolute_error,splitter =random,min_samples_split=14,ccp_alpha=0.5

Finding Best Random forest Module using R2 value

R2 will change depends up on various hyper tuning parameter

n_estimators	criterion	random_state	R2 value
	squared_error		0.85
50	***squared_error***		0.84
50	***squared_error***	None	0.84
61	***squared_error***	18	0.84
	friedman_mse		0.85
50	***friedman_mse***		0.85
61	***friedman_mse***	None	0.85
61	***friedman_mse***	18	0.85
	absolute_error		0.85
50	***absolute_error***		0.85
50	***absolute_error***	18	0.85
61	***absolute_error***	18	0.85
	poisson		0.84
50	***poisson***		0.84
61	***poisson***	None	0.84

50	***poisson***	18	0.84

We got high r2 value =0.85 for the hyper tuning parameter criterion=absolute_error ,*friedman_mse, squared_error*

4. I have created many models using machine learning algorithm like
   - Multiple linear regression
   - Support vector machine
   - Decision tree
   - Random forest

   We can select the best model having a high R2 value.

   Compared to all other values, the Decision Tree R2 value 0.88 is high,so we can finalize that.