1. Problem statement – To predict the outcome of insurance charges based on input parameters of age, sex, bmi, children, smoker, charges

2. Data set has a total of 1338 rows × 6 columns

3. Pre-processing steps
   Out of 6 parameters, age, sex, bmi, children, smoker, charges – the sex column and smoker column is categorical so I will be converting it into numerical value / nominal data
4. Trying out different models:

A - Random Forest

S.No	N_estimators	Random state	R value
1.	50	0	0.8498329315421834
2.	100	0	0.8538307913484513
3.	1000	0	0.8541778123151671
4.	10	0	0.83303041340085

B - Decision Tree

S.No	Criterion	Splitter	R value
1	friedman_mse	best	0.6959731595948448
2	friedman_mse	random	0.7410817834865593
3	squared_error	best	0.6843657056447416
4	squared_error	random	0.7537133297740599
5	absolute_error	best	0.6749612314344458
6	absolute_error	random	0.7460729701361937
7	poisson	best	0.7284245497681144
8	poisson	random	0.749804147521383

C – Multiple linear regression R score is 0.7894790349867009

D – Support vector machine

S.No	Hyperparameter	Linear (r value)	RBF (non linear)	Poly (r value)	Sigmoid (r value)
1.	C10	-0.111661287	-0.08842	-0.06429258	-0.089

5. I have chosen Random forest as it gives the highest R score