1) Problem statement

Predict a customer's medical insurance **charges** from available demographic/health attributes.

---

2) Dataset basic info

- insurance_pre.csv
- Total rows: **1338**
- Total columns: **6**
- Columns found: age, sex, bmi, children, smoker, charges
- Missing values: **none**

---

3) Pre-processing

- Detected categorical columns: sex, smoker (and any other object cols).
- Converted categorical variables to dummy/one-hot variables via pd.get_dummies(..., drop_first=True).
- Standardized numeric features (StandardScaler) prior to model training.
- Train/test split: 70% train / 30% test (random_state=0).

---

4) Machine Learning Algorithms:

(i)Multiple Linear Regression

Best R^2 Score = 0.7894790349867009

(ii)Support Vector Machine

S. No	C	Linear R^2	Rbf R^2	Poly R^2	Sigmoid R^2
1	10	-0.001617632488647213	-0.08196910396420853	-0.09311615532848516	-0.09078319814614
2	100	0.5432818196692804	-0.12480367775039669	-0.09976172333666167	-0.11814554828411405
3	1000	0.634036931263208	-0.11749092439183229	-0.055505937517909665	-1.665908131553306 4
4	3000	0.759089036774 1108	-0.09621285083097098	0.04892896389686563	-12.019048105850 73
5	5000	0.764893815014 5156	-0.07310107353590722	0.14622378659405622	-31.568282071990 63

Best R^2 Score:

Linear kernel at C = 5000:**R^2 = 0.7649**

(iii)Decision Tree

Sl.No	Criterion	Splitter	R^2 Value
1	squared_error	best	0.682050774394739
2	squared_error	random	0.7194937388580465
3	friedman_mse	best	0.6939741374705168
4	friedman_mse	random	0.7890426027556149
5	absolute_error	best	0.6686785651066409
6	absolute_error	random	0.7092140339803142
7	poisson	best	0.7167036501017701
8	poisson	random	0.718530210384976

Best R^2 Score: friedman_mse with random splitter: **R^2 = 0.7890**

(iv)Random Forest

Sl.No	n_estimators	criterion	random_state	R^2 Value
1	100	squared_error	0	0.8538307913484513
2	50	squared_error	0	0.8498329315421834
3	10	squared_error	0	0.83303041340085
4	100	absolute_error	0	0.8520093621081837
5	50	absolute_error	0	0.8526655993519747
6	10	absolute_error	0	0.835063555313752
7	100	friedman_mse	0	0.8540518935149612
8	50	friedman_mse	0	0.8500716139332296
9	10	friedman_mse	0	0.8331662678473348
10	100	poisson	0	0.8526334258892607
11	50	poisson	0	0.8491075958392151
12	10	poisson	0	0.8313991040134341

Best R^2 Score: friedman_mse with 100 estimators: **R^2 = 0.8541**

---

5)Best Model:

Final Model: Random Forest Regressor

Justification: It achieved the highest R^2 score (0.8541) using friedman_mse with 100 estimators, showing superior accuracy, stability, and generalization compared to other models.