Here the prediction is for insurance charges.

There are total rows 1338

Nominal data are sex and smoker column columns.

So the below stages will be,

Machine Learning -> supervised -> regression.

**Multiple Linear Regression**

R Score = 0.7894790349867009

**Support Vector Machine**

| Hyper Parameter (C) | linear | poly | rbf | sigmoid | precomputed |
| --- | --- | --- | --- | --- | --- |
| 0.1 | -0.08095996843 | -0.08830237655 | -0.08907451521 | -0.0882699145 | Not Applicable |
| 1 | -0.01010266532 | -0.07569965571 | -0.08338238594 | -0.07542924281 | Not Applicable |
| 10 | 0.4624684142 | 0.03871622276 | -0.03227329391 | 0.03930714378 | Not Applicable |
| 100 | 0.6288792857 | 0.6179569624 | 0.3200317832 | 0.5276103547 | Not Applicable |
| 1000 | 0.7649311739 | 0.8566487676 | 0.8102064852 | 0.2874706949 | Not Applicable |
| 10000 | 0.7414230132 | 0.8591715095 | 0.8779952401 | 34.15153598 | Not Applicable |
| 100000 | 0.7414188971 | 0.8577881129 | 0.872498445 | -3465.953574 | Not Applicable |

Highest R score = 0.877995240144991

**Decision Tree**

|  | best | random |
| --- | --- | --- |
| squared\_error | 0.7060102409 | 0.7319722536 |
| friedman\_mse | 0.6765449176 | 0.6807612699 |
| absolute\_error | 0.6949280907 | 0.7034725968 |
| poisson | 0.7420367832 | 0.7162495396 |

Highest R score = 0.742036783150447

**Random Forest**

| ***squared\_error*** | 0.8498329315 |
| --- | --- |
| ***absolute\_error*** | 0.8526655994 |
| ***friedman\_mse*** | 0.8500716139 |
| ***poisson*** | 0.8491075958 |

Highest R score = 0.850071613933229

From all the observation we will select the SVM algorithm with kernel as rbf and hyper parameter as 10000