

## Part - 1

| Model details  | R2-score            | MSE-score          |
|--|---------------------|--------------------|
| Linear regression with raw data; features ['age', 'children', ...] | 0.04225304958507703 | 141012174.4510918  |
| With Pandas get.dummies()  | 0.7807756903022633  | 32277102.620504204 |
| Introduction of OneHotEncoding                                     | 0.7807756903022633  | 32277102.620504208 |
| Introduction of new feature <b>BMI categories</b>                  | 0.7742644682532225  | 33235770.85648383  |
| Introduction of new feature <b>Life stage categories</b>           | 0.7744913606954102  | 33202364.749969427 |
| Introduction of new feature <b>Risk factor</b>                     | 0.811462866057491   | 27758930.6164651   |

## Part - 2

| Model details  | R2-score            | MSE-score         |
|--|---------------------|-------------------|
| Linear regression with raw data; features ['age', 'children', ...] | 0.04225304958507703 | 141012174.4510918 |
| With Pandas get.dummies()  | 0.7807316546587777  | 32283586.13041105 |
| Introduction of OneHotEncoding                                     | 0.7807316546587777  | 32283586.13041105 |
| Introduction of new feature <b>BMI categories</b>                  | 0.7741877656397048  | 33247064.02981477 |
| Introduction of new feature <b>Life stage categories</b>           | 0.774413283752124   | 33213860.27030756 |
| Introduction of new feature <b>Risk factor</b>                     | 0.8114456631711873  | 27761463.53365635 |

Here missing BMI values are filled with the mean values derived from the datasets.