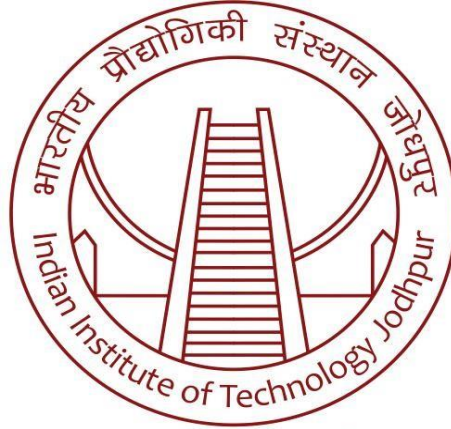# Analytics Lab(Python)

॥ त्वं ज्ञानमयो विज्ञानमयोऽसि ॥

Multi-Linear Regression

1-Checking model is statistically significant or not and along with variables also by p-value less than 0.05

Values	Dependent Variable	Independent Variable-1	Independent Variable-2	Independent Variable-3
	Charges	age(Beta1)	bmi(Beta2)	children(Beta3)
Coefficient		239.9945	332.0834	542.8647
p-value		0.000	0.000	0.036

Dependent Variable is Charges

Independent Variable is these three and there

coefficient age(Beta1)=239.9945 bmi(Beta2)=332.0834

children(Beta3)=542.8647

Beta3>Beta2>Beta1 Children have more effect on the charges

➔ **If we take all variables constant except children then for unit 1 year increase in age of children will increase charges by 542.8647 rupees**
➔ **If we take all variables constant except bmi then for 1 unit increase in bmi will increase charges by 332.0834 rupees**
➔ **If we take all variables constant except age then for 1 unit increase in age will increase charges by 239.9945 rupees**

R-squared: 0.120 12% variability in our model

Prob (F-statistic): 8.80e-37 this value is less than 0.05 so **model is statistically significant** age(p-value): 0.000

bmi(p-value): 0.000   **All variables are also statistically significant bcoz p-value**

children(p-value): 0.036   **is than 0.05**

```
                           OLS Regression Results
==============================================================================
Dep. Variable:                 charges   R-squared:                       0.120
Model:                             OLS   Adj. R-squared:                  0.118
Method:                  Least Squares   F-statistic:                     60.69
Date:                 Wed, 28 Sep 2022   Prob (F-statistic):           8.80e-37
Time:                         11:50:48   Log-Likelihood:                -14392.
No. Observations:                 1338   AIC:                         2.879e+04
Df Residuals:                     1334   BIC:                         2.881e+04
Df Model:                            3
Covariance Type:             nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const      -6916.2433   1757.480     -3.935      0.000   -1.04e+04   -3468.518
age          239.9945     22.289     10.767      0.000     196.269     283.720
bmi          332.0834     51.310      6.472      0.000     231.425     432.741
children     542.8647    258.241      2.102      0.036      36.261    1049.468
==============================================================================
Omnibus:                      325.395   Durbin-Watson:                   2.012
Prob(Omnibus):                  0.000   Jarque-Bera (JB):              603.372
Skew:                           1.520   Prob(JB):                    9.54e-132
Kurtosis:                       4.255   Cond. No.                         290.
==============================================================================
```

2- Stepwise Regression

Finally we gets statistically significant model with these independent variables

age: 0.000 bmi: 0.000 children: 0.001
smoker_yes: 0.000

all are having p-value less than 0.05

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                        Y    R-squared:                       0.750
Model:                              OLS    Adj. R-squared:                  0.749
Method:                   Least Squares    F-statistic:                     801.0
Date:                 Wed, 28 Sep 2022    Prob (F-statistic):               0.00
Time:                         15:38:44    Log-Likelihood:                 -13549.
No. Observations:                 1338    AIC:                         2.711e+04
Df Residuals:                     1332    BIC:                         2.714e+04
Df Model:                            5
Covariance Type:             nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
Intercept    -1.251e+04    963.375    -12.988      0.000   -1.44e+04   -1.06e+04
X[0]           257.4058     11.885     21.657      0.000     234.090     280.722
X[1]           329.4629     27.616     11.930      0.000     275.287     383.639
X[2]           479.5142    137.674      3.483      0.001     209.432     749.596
X[3]          2.381e+04    410.773     57.959      0.000     2.3e+04     2.46e+04
X[4]           773.9462    390.709      1.981      0.048       7.473    1540.419
==============================================================================
Omnibus:                       302.023    Durbin-Watson:                   2.083
Prob(Omnibus):                   0.000    Jarque-Bera (JB):              727.655
Skew:                            1.215    Prob(JB):                    9.81e-159
Kurtosis:                        5.674    Cond. No.                         300.
==============================================================================
```
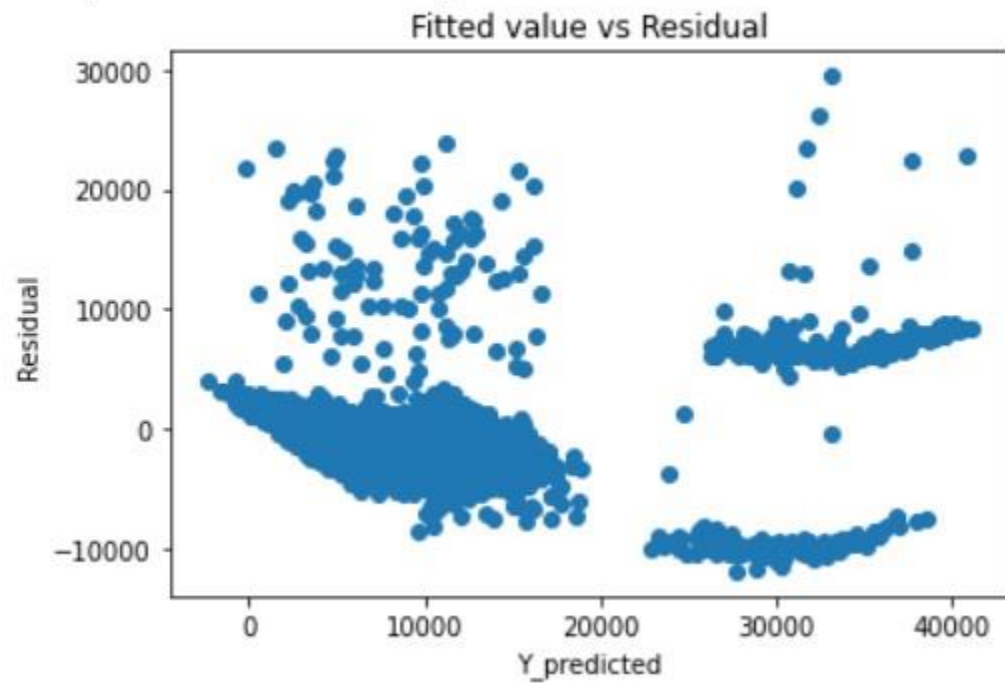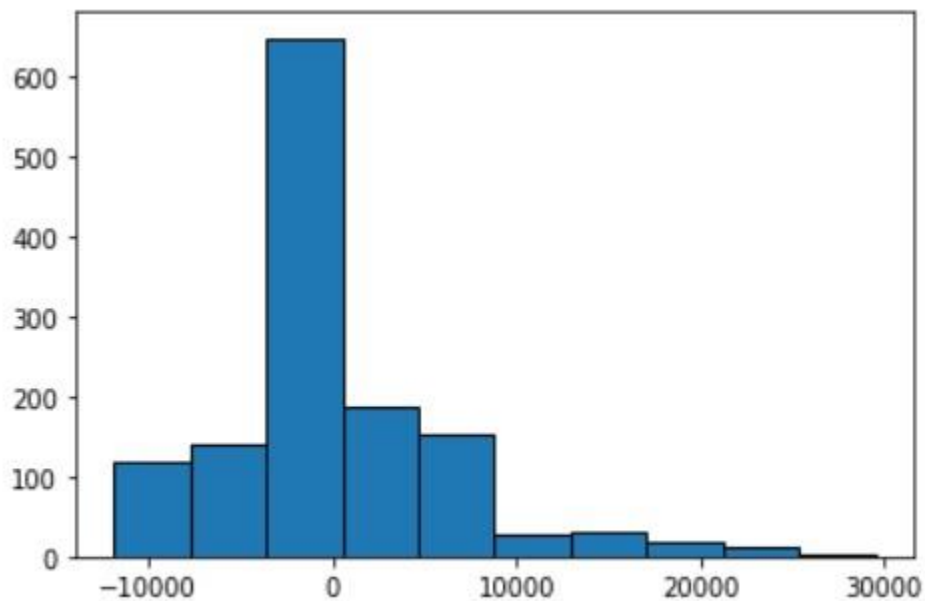
3-

(i)Checking Linearity b/w residual and predicted value of our model

`Text(0, 0.5, 'Residual')`

Fitted value vs Residual

There is no linearity between residual and predicted value

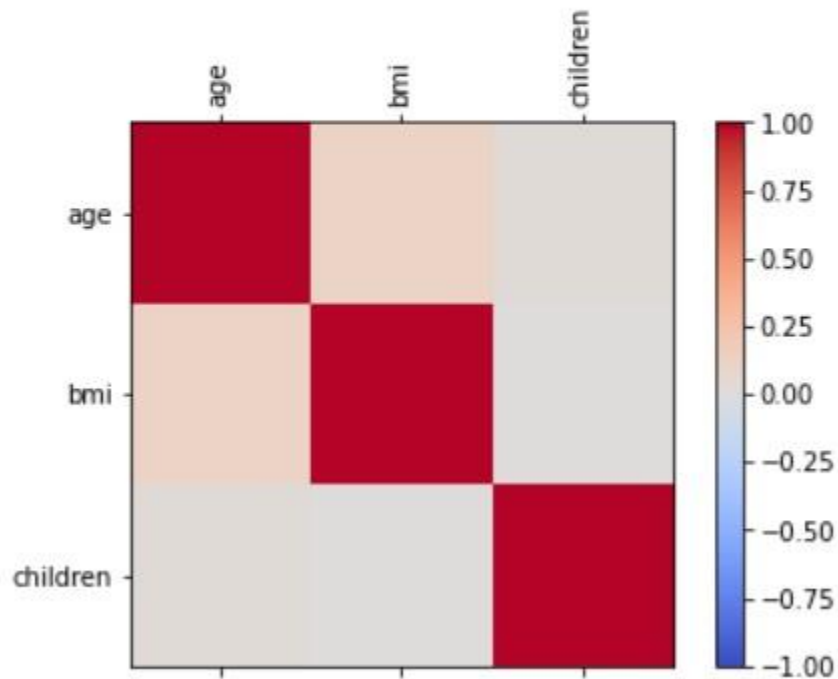## ii)Checking data is normally distributed or not

It is not properly normally distributed.

iii)Checking Multicollinearity is present in data or not

Covariance of variables is less than 0.5 so there is no problem

```
                age        bmi   children
age        1.000000   0.122240   0.025965
bmi        0.122240   1.000000   0.003858
children   0.025965   0.003858   1.000000
```


Variance inflation factor(VIF) is also less than 10 there is no problem

```
        feature         VIF
0           age    7.583324
1           bmi    7.924975
2      children    1.761431
```

(iv)Checking for Homoscedasticity is there or not

Residual Plot
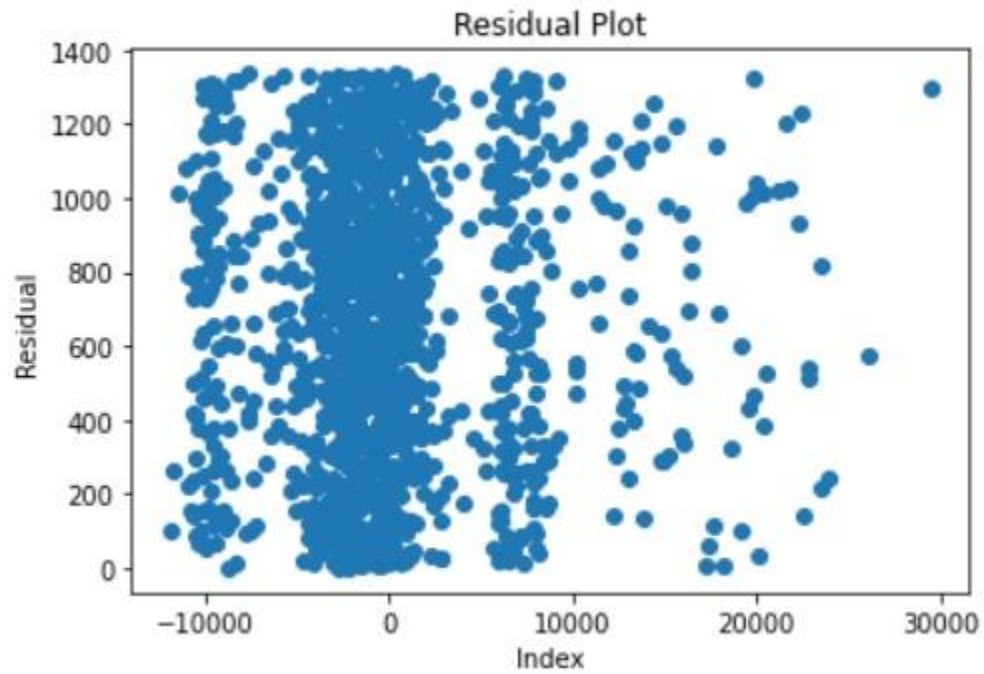
By this scatter plot we can see that homoscedasticity is not there.