**MISSING VALUE**

OLS Regression Results

==============================================================================

Dep. Variable: y R-squared: 0.040

Model: OLS Adj. R-squared: 0.040

Method: Least Squares F-statistic: 439.1

Date: Sun, 10 Mar 2019 Prob (F-statistic): 0.00

Time: 20:32:50 Log-Likelihood: -2.5681e+05

No. Observations: 52584 AIC: 5.136e+05

Df Residuals: 52578 BIC: 5.137e+05

Df Model: 5

Covariance Type: nonrobust

==============================================================================

coef std err t P>|t| [0.025 0.975]

------------------------------------------------------------------------------

Intercept 464.0122 34.090 13.611 0.000 397.195 530.830

x[0] 0.3174 0.167 1.903 0.057 -0.010 0.644

x[1] -0.2104 0.043 -4.889 0.000 -0.295 -0.126

x[2] -0.3678 0.032 -11.409 0.000 -0.431 -0.305

x[3] -1.1108 0.164 -6.761 0.000 -1.433 -0.789

x[4] -0.0505 0.002 -28.279 0.000 -0.054 -0.047

==============================================================================

Omnibus: 44117.629 Durbin-Watson: 0.116

Prob(Omnibus): 0.000 Jarque-Bera (JB): 2177761.812

Skew: 3.763 Prob(JB): 0.00

Kurtosis: 33.616 Cond. No. 2.49e+05

==============================================================================

Berdasarkan hasil regresi di atas, diperoleh nilai R-squared untuk model regresi pertama adalah sebesar 0,040, artinya variabel Y dapat dijelaskan oleh sekelompok variabel independen x1, x2, x3, x4, dan x5 secara serentak atau simultan sebesar 4% sedangkan sisanya 96% dijelaskan oleh variabel lain di luar model yang tidak diteliti.

**MISSING VALUE, OUTLIER**

OLS Regression Results

==============================================================================

Dep. Variable: y2 R-squared: 0.047

Model: OLS Adj. R-squared: 0.047

Method: Least Squares F-statistic: 485.7

Date: Sun, 10 Mar 2019 Prob (F-statistic): 0.00

Time: 20:34:50 Log-Likelihood: -2.2293e+05

No. Observations: 49515 AIC: 4.459e+05

Df Residuals: 49509 BIC: 4.459e+05

Df Model: 5

Covariance Type: nonrobust

==============================================================================

coef std err t P>|t| [0.025 0.975]

------------------------------------------------------------------------------

Intercept 487.8774 24.086 20.256 0.000 440.669 535.086

x2[0] -0.2716 0.122 -2.217 0.027 -0.512 -0.032

x2[1] -0.0559 0.032 -1.773 0.076 -0.118 0.006

x2[2] -0.4110 0.023 -18.055 0.000 -0.456 -0.366

x2[3] -0.3524 0.120 -2.928 0.003 -0.588 -0.116

x2[4] -0.0644 0.002 -35.876 0.000 -0.068 -0.061

==============================================================================

Omnibus: 10431.939 Durbin-Watson: 0.166

Prob(Omnibus): 0.000 Jarque-Bera (JB): 33801.960

Skew: 1.069 Prob(JB): 0.00

Kurtosis: 6.437 Cond. No. 2.50e+05

==============================================================================

Berdasarkan hasil regresi di atas, diperoleh nilai R-squared untuk model regresi kedua adalah sebesar 0,047, artinya variabel Y dapat dijelaskan oleh sekelompok variabel independen x1, x2, x3, x4, dan x5 secara serentak atau simultan sebesar 4,7% sedangkan sisanya 95,3% dijelaskan oleh variabel lain di luar model yang tidak diteliti.

**MISSING VALUE, TRANSFORMASI**

OLS Regression Results

==============================================================================

Dep. Variable: y3 R-squared: 0.040

Model: OLS Adj. R-squared: 0.040

Method: Least Squares F-statistic: 439.1

Date: Sun, 10 Mar 2019 Prob (F-statistic): 0.00

Time: 20:39:08 Log-Likelihood: -2.5681e+05

No. Observations: 52584 AIC: 5.136e+05

Df Residuals: 52578 BIC: 5.137e+05

Df Model: 5

Covariance Type: nonrobust

==============================================================================

coef std err t P>|t| [0.025 0.975]

------------------------------------------------------------------------------

Intercept 464.0122 34.090 13.611 0.000 397.195 530.830

x3[0] 0.3174 0.167 1.903 0.057 -0.010 0.644

x3[1] -0.2104 0.043 -4.889 0.000 -0.295 -0.126

x3[2] -0.3678 0.032 -11.409 0.000 -0.431 -0.305

x3[3] -1.1108 0.164 -6.761 0.000 -1.433 -0.789

x3[4] -0.0505 0.002 -28.279 0.000 -0.054 -0.047

==============================================================================

Omnibus: 44117.629 Durbin-Watson: 0.116

Prob(Omnibus): 0.000 Jarque-Bera (JB): 2177761.812

Skew: 3.763 Prob(JB): 0.00

Kurtosis: 33.616 Cond. No. 2.49e+05

==============================================================================

Berdasarkan hasil regresi di atas, diperoleh nilai R-squared untuk model regresi ketiga adalah sebesar 0,040, artinya variabel Y dapat dijelaskan oleh sekelompok variabel independen x1, x2, x3, x4, dan x5 secara serentak atau simultan sebesar 4% sedangkan sisanya 96% dijelaskan oleh variabel lain di luar model yang tidak diteliti.

**MISSING VALUE, OUTLIER, TRANSFORMASI**

OLS Regression Results

==============================================================================

Dep. Variable: y4 R-squared: 0.391

Model: OLS Adj. R-squared: 0.391

Method: Least Squares F-statistic: 6363.

Date: Sun, 10 Mar 2019 Prob (F-statistic): 0.00

Time: 20:41:16 Log-Likelihood: 1.3160e+05

No. Observations: 49515 AIC: -2.632e+05

Df Residuals: 49509 BIC: -2.631e+05

Df Model: 5

Covariance Type: nonrobust

==============================================================================

coef std err t P>|t| [0.025 0.975]

------------------------------------------------------------------------------

Intercept 5.2726 0.031 170.559 0.000 5.212 5.333

x4[0] 0.8273 0.096 8.657 0.000 0.640 1.015

x4[1] -0.5819 0.024 -23.976 0.000 -0.629 -0.534

x4[2] -5.1835 0.031 -168.624 0.000 -5.244 -5.123

x4[3] -1.0546 0.090 -11.754 0.000 -1.230 -0.879

x4[4] -0.5354 0.003 -172.032 0.000 -0.541 -0.529

==============================================================================

Omnibus: 2023.254 Durbin-Watson: 0.172

Prob(Omnibus): 0.000 Jarque-Bera (JB): 4132.340

Skew: -0.292 Prob(JB): 0.00

Kurtosis: 4.289 Cond. No. 2.47e+03

==============================================================================

Berdasarkan hasil regresi di atas, diperoleh nilai R-squared untuk model regresi keempat adalah sebesar 0,391, artinya variabel Y dapat dijelaskan oleh sekelompok variabel independen x1, x2, x3, x4, dan x5 secara serentak atau simultan sebesar 39,1% sedangkan sisanya 60,9% dijelaskan oleh variabel lain di luar model yang tidak diteliti.

**KESIMPULAN:**

Dari keempat model regresi, didapatkan nilai R-squared terbesar dari keseluruhan model adalah 0,391. R-squared tersebut diperoleh dari data yang telah melalui tahap pre-processing (identifikasi missing value, outlier, dan transformasi), sehingga dapat disimpulkan bahwa untuk mendapatkan model regresi terbaik, data sebaiknya terlebih dahulu bebas dari missing value dan outlier serta dilakukan transformasi.