**Dataset**:-ToyotaCorollary

**Problem statement**:- prediction model for predicting Price.

Dependent variable:- Price

Independent Variable:- Age\_08\_04,KM,HP,cc,Doors,Gears,Quarterly\_Tax,Weight

Now,lets make EDA of the given data:-

**Business Moment 1:**

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **Price** | **AGE\_08\_04** | **KM** | **HP** | **CC** | **DOORS** | **GEARS** | **QUATERLY\_TAX** | **WEIGHT** |
| MEAN | **10731** | **55.95** | **68533** | **101.5** | **1577** | **4.033** | **5.026** | **87.12** | **1072** |
| MEDIAN | **9900** | **61** | **63390** | **110** | **1600** | **4** | **5** | **85** | **1070** |

**Business Moment 2:**

|  |  |  |  |
| --- | --- | --- | --- |
|  | **VARIANCE** | **SD** | **RANGE** |
| PRICE | 1.315487e+07 | 3.626965e+03 | 4350-32500 |
| AGE\_08\_04 | 3.459596e+02 | 1.859999e+01 | 1-80 |
| KM | 1.406734e+09 | 3.750645e+04 | 1-243000 |
| HP | 2.244327e+02 | 1.498108e+01 | 69-192 |
| CC | 1.801041e+05 | 4.243868e+02 | 1300-16000 |
| DOORS | 9.075927e-01 | 9.526766e-01 | 2-5 |
| GEARS | 3.553619e-02 | 1.885104e-01 | 3-6 |
| QUATERLY\_TAX | 1.691563e+03 | 4.112861e+01 | 19-283 |
| WEIGHT | 2.771088e+03 | 5.264112e+01 | 1000-1615 |

**Business moment 3 :**

SKEWNESS VALUES:-

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| |  |  | | --- | --- | | PRICE | 1.70210464 | | AGE\_08\_04 | -0.8258308 | | KM | 1.01485070 | | HP | 0.95483692 | | CC | 27.40313010 | | DOORS | -0.07631517 | | GEARS | 2.28157329 | | QUATERLY\_TAX | 1.99175056 | | WEIGHT | 3.10539127 | |
| OBSERVATION:-   * price is heavily right skewed * Age\_08\_04 is moderately left skewed * KM,HP,CC is heavily right skewed * Doors is fairly symmetrical * Gears,Quarterly\_Tax,Weight is heavily right skewed |

**Business Moment 4:-(kurtiosis)**

|  |  |
| --- | --- |
| PRICE | 1.70210464 |
| AGE\_08\_04 | -0.82583808 |
| KM | 4.675020 |
| HP | 11.801521 |
| CC | 930.469571 |
| DOORS | 1.127581 |
| GEARS | 40.568149 |
| QUATERLY\_TAX | 7.279218 |
| WEIGHT | 22.291372 |

Only age\_08\_04 and doors are lighter peaks while others are heavy peaks

**Linear Model:-**

R-squared value-0.863

P-value is less than 0.05

**Now,lets take prediction analysis at high peaks as per business decision 4**

1. **Prediction based on KM**

Multiple R-squared: 0.3249,

Adjusted R-squared: 0.3244

p-value: < 2.2e-16

1. **Prediction based on HP**

Multiple R-squared: 0.09922,

Adjusted R-squared: 0.09859

p-value: < 2.2e-16

1. **Prediction based on CC**

Multiple R-squared: 0.01597

Adjusted R-squared: 0.01529

P-value: 1.551e-06

1. **Prediction based on Gears**

Multiple R-squared: 0.003982,

Adjusted R-squared: 0.003288

p-value: 0.01678

1. **Prediction based on quarterly\_tax**

Multiple R-squared: 0.04805,

Adjusted R-squared: 0.04738

p-value: < 2.2e-16

1. **Prediction based on weight**

Multiple R-squared: 0.3378,

Adjusted R-squared: 0.3373

p-value: < 2.2e-16