**Dataset**:-COMPUTER

**Problem statement**:-Predict Price of the computer

Dependent variable:- Prices

Independent Variable:- speed, hd, ram, screen ,cd, multi ,premium, ads, trend

Now,lets make EDA of the given data:-

**Business Moment 1:**

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **X** | **PRICE** | **SPEED** | **HD** | **RAM** | **SCREEN** | **CD** | **PREMIUM** | **MULTI** | **ADS** | **TREND** |
| MEAN | **3130** | **2220** | **52.01** | **416.6** | **8.287** | **14.00** | **-** | **-** | **-** | **221.3** | **15.93** |
| MEDIAN | **3130** | **2144** | **50** | **340.0** | **8.000** | **14.61** | **-** | **-** | **-** | **246.0** | **16.00** |

**Business Moment 2:**

Variance SD RANGE

X: 3.265112e+06 1806.9619992 1-6259

Price: 3.373332e+05 580.8039557 949-5399

Speed: 4.476498e+02 21.1577354 25-100

Hd: 6.684730e+04 258.5484452 80-2100

Ram : 3.170928e+01 5.6310989 2-32

Screen 8.192336e-01 0.9051152 14-17

Cd NA NA NO ,YES

Multi:- NA NA NO,YES

Premium:- NA NA NO,YES

Ads:- 5.600320e+03 74.8352840 39-339

Trend:- 6.199962e+01 7.8739838 1-35

**Business moment 3 :**

SKEWNESS VALUES:-

PRICE: 0.7115542

SPEED: 0.6568505

HD:- 1.377689

RAM:- 1.38587

SCREEN: 1.633616

|  |
| --- |
| ADS: -0.5531955  TRENDS: 0.2366127 |
|  |
| OBSERVATION:-   * price,speed is moderately right skewed * hd,ram,screen is heavily right skewed * ads is moderately left skewed * trend is fairly symmetrical!! |

**Business Moment 4:-**

|  |  |
| --- | --- |
| PRICE | 3.728875 |
| SPEED | 2.723809 |
| ADS | 2.459629 |
| TREND | 2.325446 |
| HD | 5.449539 |
| RAM | 4.460124 |
| SCREEN | 4.849387 |

Speed,ads,trend have lighter peaks while others have heavy peaks!!

**Linear Model:-**

R-squared value is 0.712

Probability-value is less than 0.05

**Equation for data model with coefficients:-**

y=-246.675+8.893\*x1+0.708\*x2+47.387\*x3+126.702\*x4+0.969\*x5+(-47.081)\*x6

Now,since Speed ADS and Trend have higher peaks lets predict basis of them:-

1. Prediction based on Speed only:-

Multiple R-squared: 0.09059,

Adjusted R-squared: 0.09044

p-value: < 2.2e-16

1. Prediction based on Ads only:-

Multiple R-squared: 0.002975,

Adjusted R-squared: 0.002815

p-value: 1.58e-05

1. Prediction based on Trend only:-

Multiple R-squared: 0.03999,

Adjusted R-squared: 0.03984

p-value: < 2.2e-16

1. Prediction based on all 3:-

Multiple R-squared: 0.2157

Adjusted R-squared: 0.2153

p-value: < 2.2e-16

In all the cases it is significant!!

{since predicted r square value is less than the actual r squared value}