**TOPIC: MULTIPLE REGRESSION ANALYSIS**

* Predict sales of the computer

Dummy variables:

Created dummies as there are categorical values in the given dataset.

Exploratory data analysis:

price speed hd ram screen cd multi premium

Min. : 949 Min. : 25.00 Min. : 80.0 Min. : 2.000 Min. :14.00 no :3351 no :5386 no : 612

1st Qu.:1794 1st Qu.: 33.00 1st Qu.: 214.0 1st Qu.: 4.000 1st Qu.:14.00 yes:2908 yes: 873 yes:5647

Median :2144 Median : 50.00 Median : 340.0 Median : 8.000 Median :14.00

Mean :2220 Mean : 52.01 Mean : 416.6 Mean : 8.287 Mean :14.61

3rd Qu.:2595 3rd Qu.: 66.00 3rd Qu.: 528.0 3rd Qu.: 8.000 3rd Qu.:15.00

Max. :5399 Max. :100.00 Max. :2100.0 Max. :32.000 Max. :17.00

ads trend

Min. : 39.0 Min. : 1.00

1st Qu.:162.5 1st Qu.:10.00

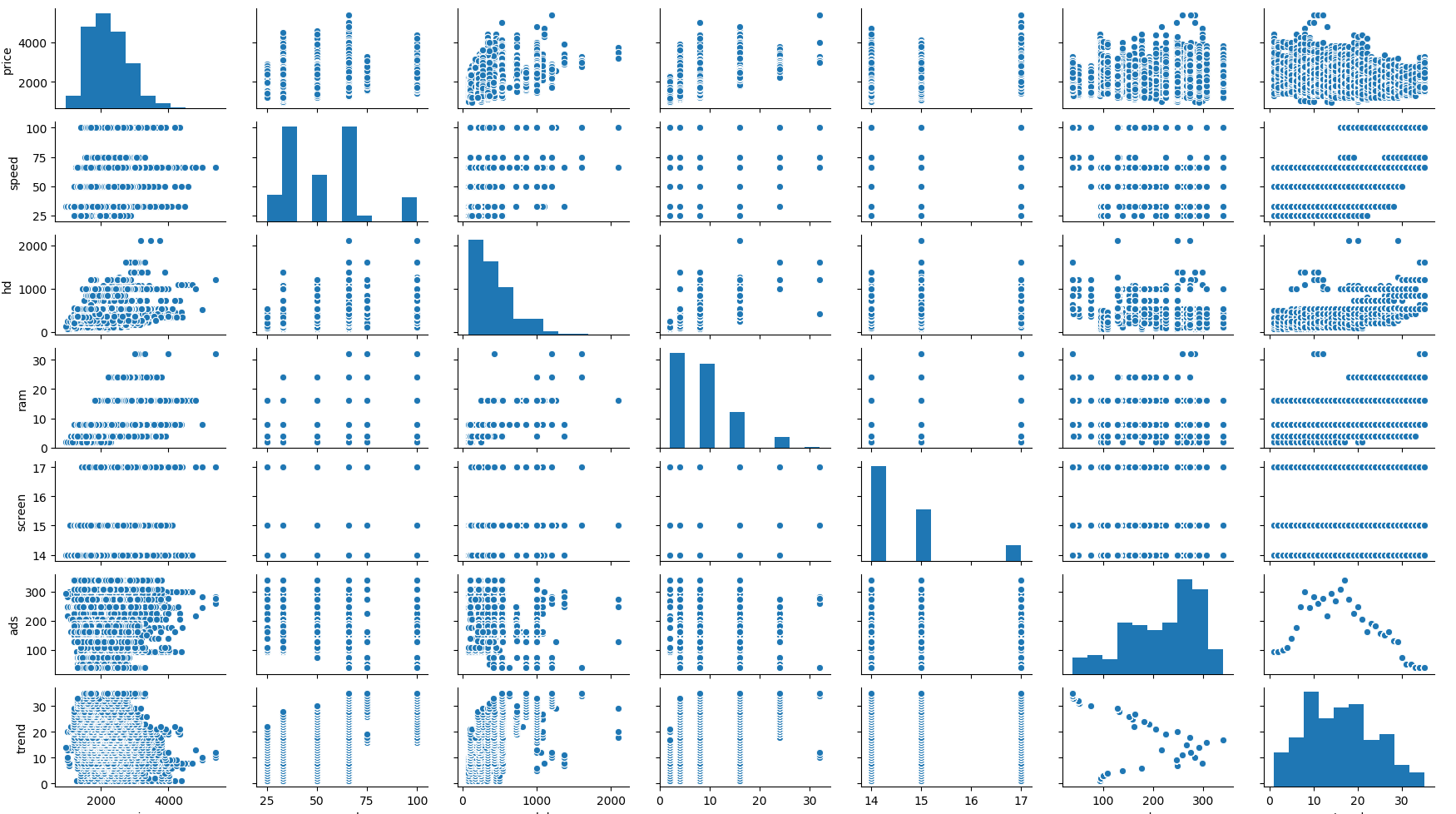
Median :246.0 Median :16.00

Mean :221.3 Mean :15.93

3rd Qu.:275.0 3rd Qu.:21.50

Max. :339.0 Max. :35.00

Scatter plot:



**Coefficient of determination:**

#Multiple R-squared: 0.7756<0.8=>moderate correlation

#p-value: < 2.2e-16<0.05= overall model is good

**Vif(variance inflation factor) values:**

Variables VIF

0 speed 1.265364

1 hd 4.207395

2 ram 2.974628

3 screen 1.081644

4 ads 1.217218

5 trend 2.022790

6 cd\_yes 1.859370

7 multi\_yes 1.290568

8 premium\_yes 1.109388

=>split data into 30 % test data and 70% train data

ADDED VARIABLES PLOT:

