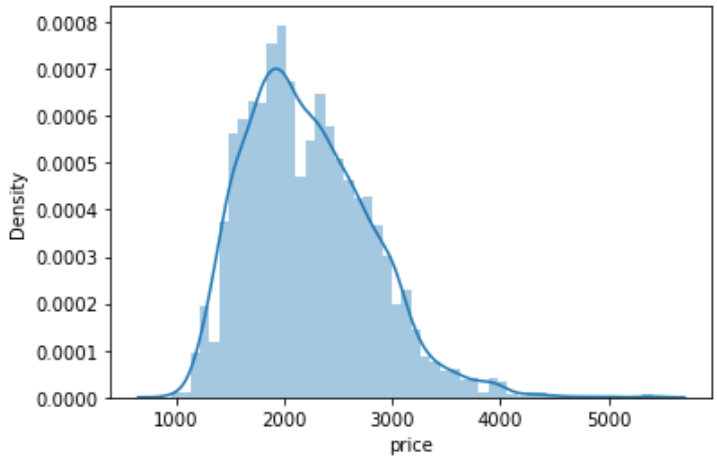
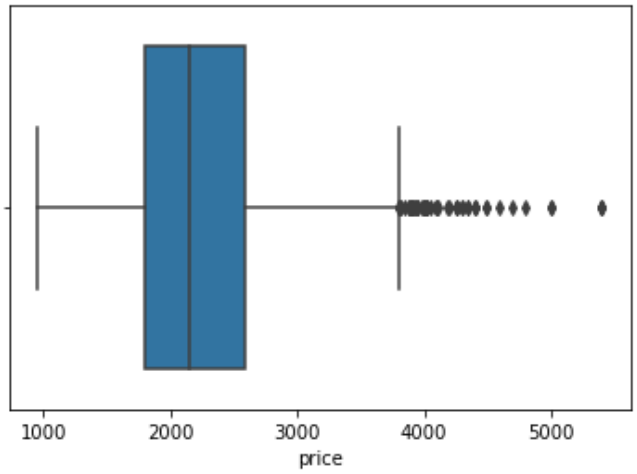
1. Variables:

- Numerical variable: price, hd, screen, ads, trend

- Categorical variable: cd, multi, premium, screen, ram

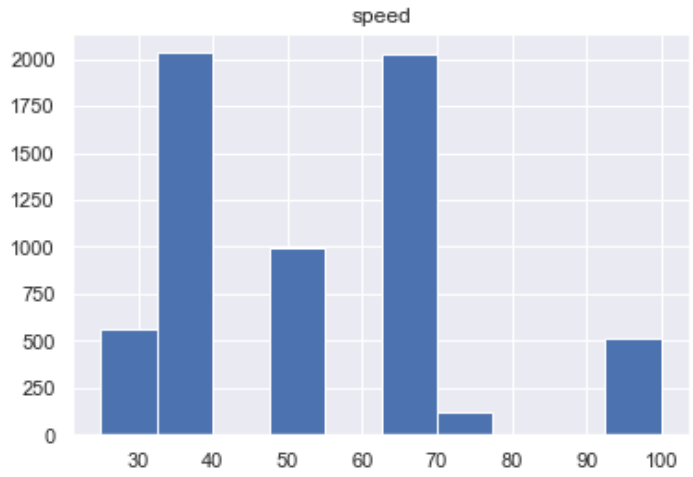
2. Distribution of Price

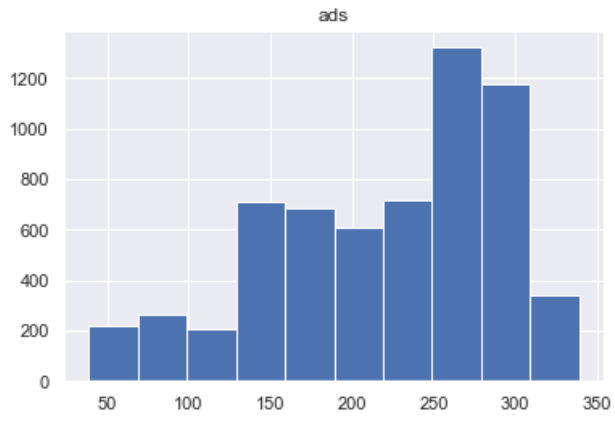


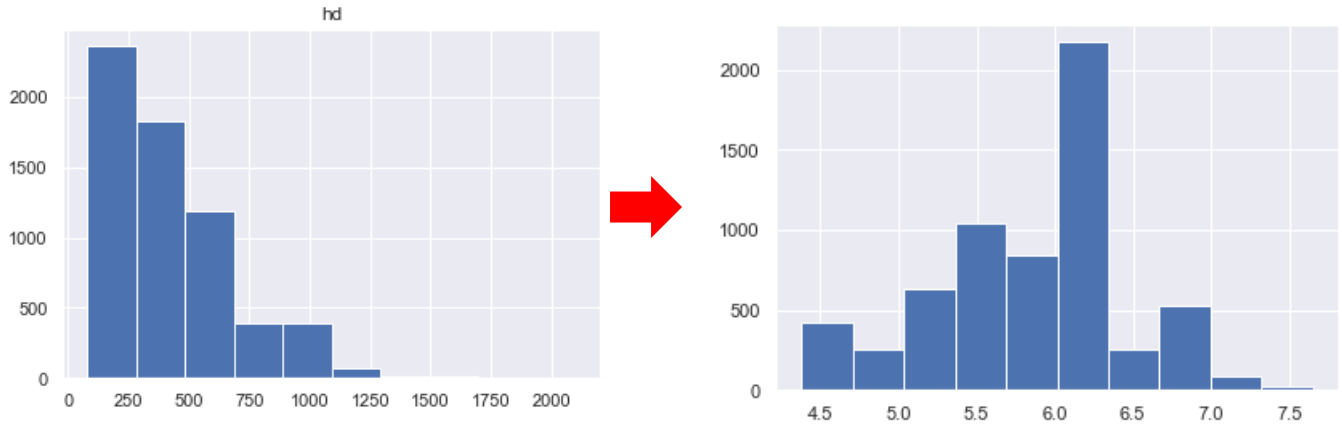


Distribution slightly skewed to the right, with outliers

3. Distribution of numerical variables







'HD' is normalized by applying log transformation.

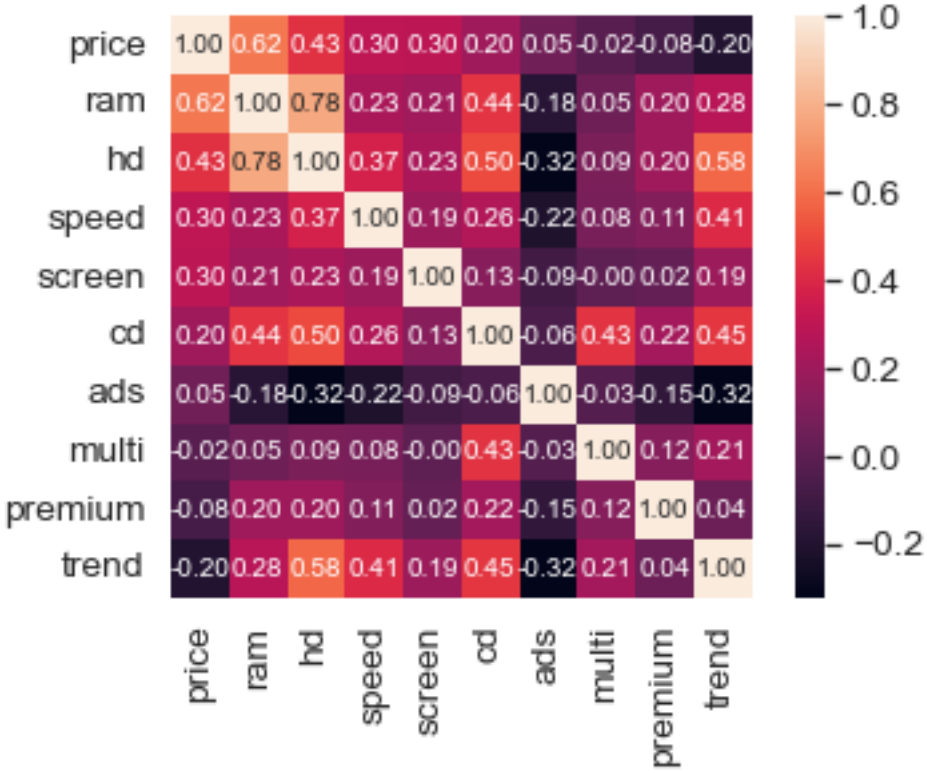
4. Check for multicollinearity between variables based on heatmap:

Price is highly correlated with RAM (0.62) and HD (0.43)

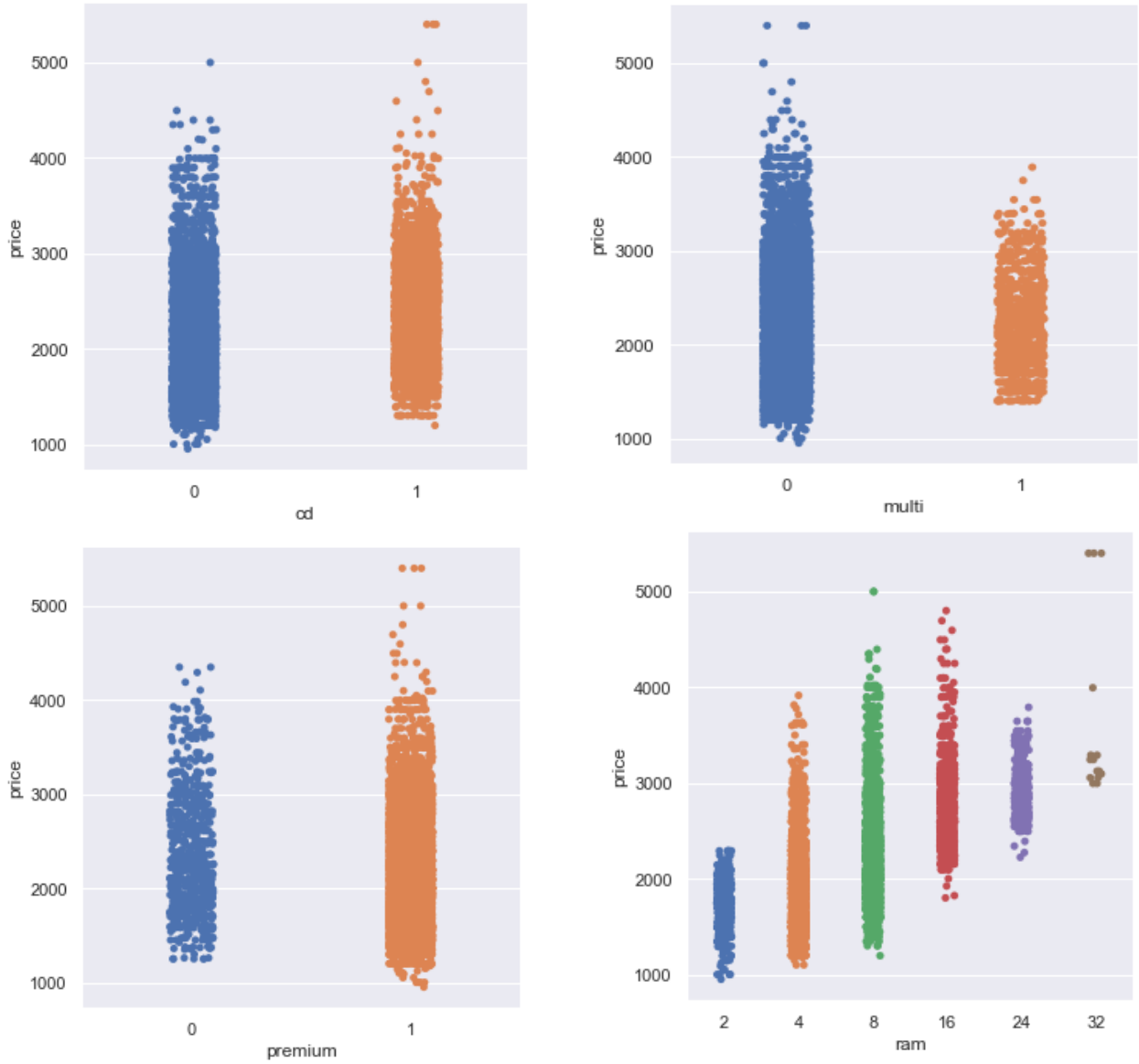
Correlation between HD and RAM is 0.78

Correlation between CD and RAM is 0.44

Correlation between HD and TREND is 0.58



5. Compare Price across categorical variables

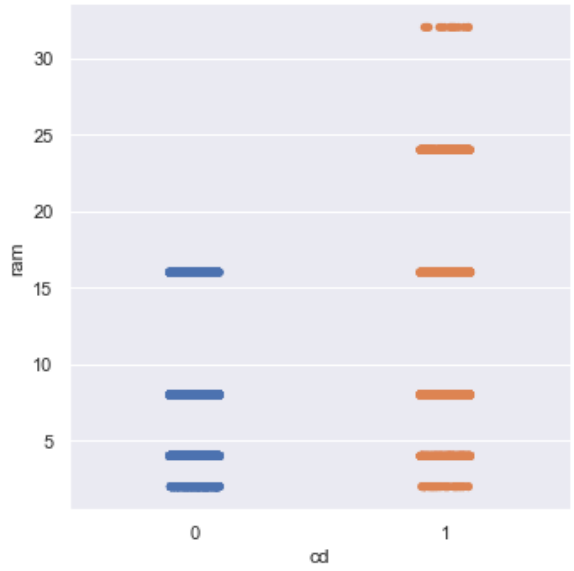


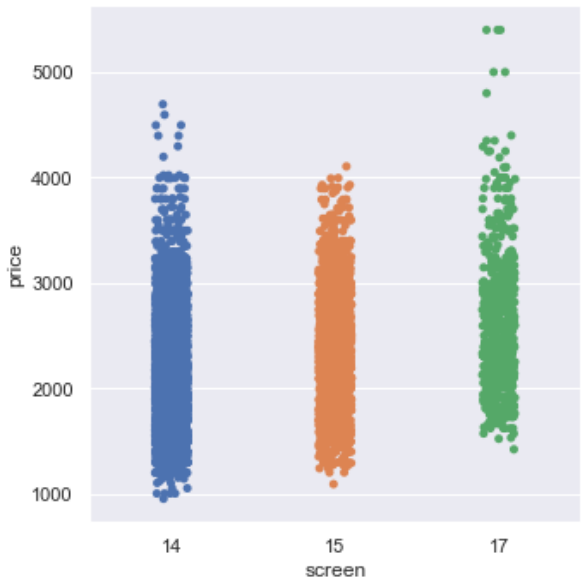
i) Multi and Premium show high correlation with Price.

ii) Computer with no Multi have higher Price

iii) Computer with Premium have higher Price.

iv) Computer with higher RAM have higher Price.





Computer with CD have higher RAM, so higher RAM will cause higher Price. Larger Screen have higher Price.

6. VIF:

	Variables	VIF
0	price	4.455684
1	speed	1.778995
2	hd	4.747040
3	ram	3.949940
4	screen	1.245590
5	cd	1.871565
6	multi	1.098532
7	premium	1.377403
8	ads	1.215409
9	trend	4.030482

Based on observation, correlation between independent variables are low. All features do have correlation with dependent variable (Price).

R^2 (before log transformation on HD) = 0.776

R^2 (after log transformation on HD) = 0.795