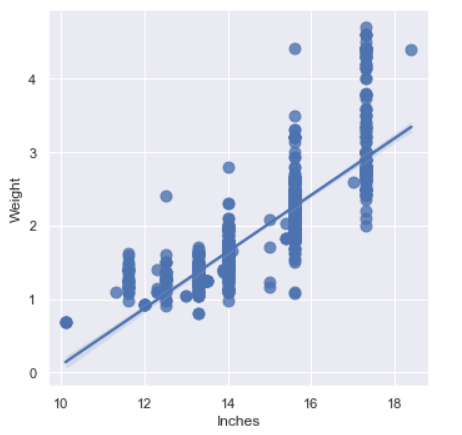
Lakshita Sethi

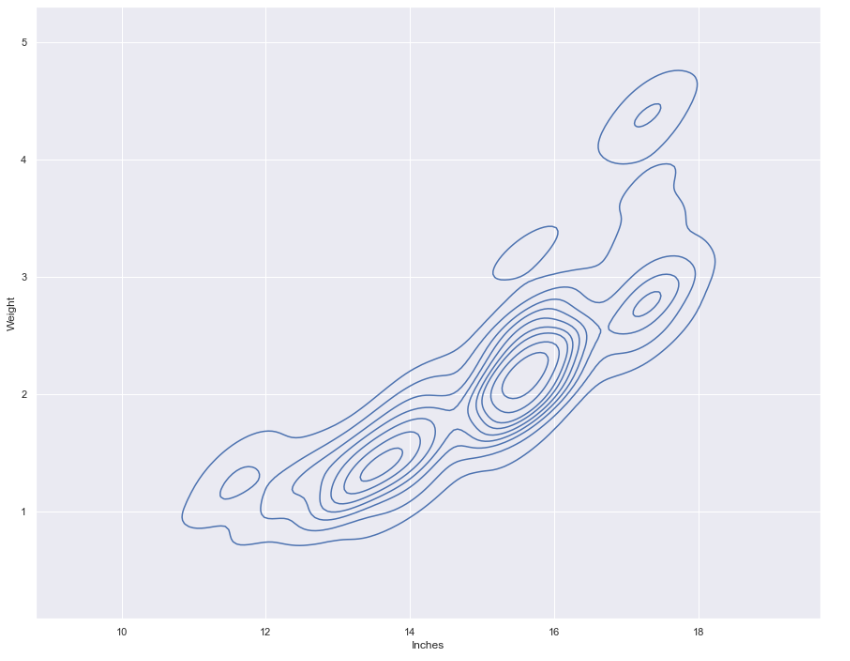
21BDA54

Question 2

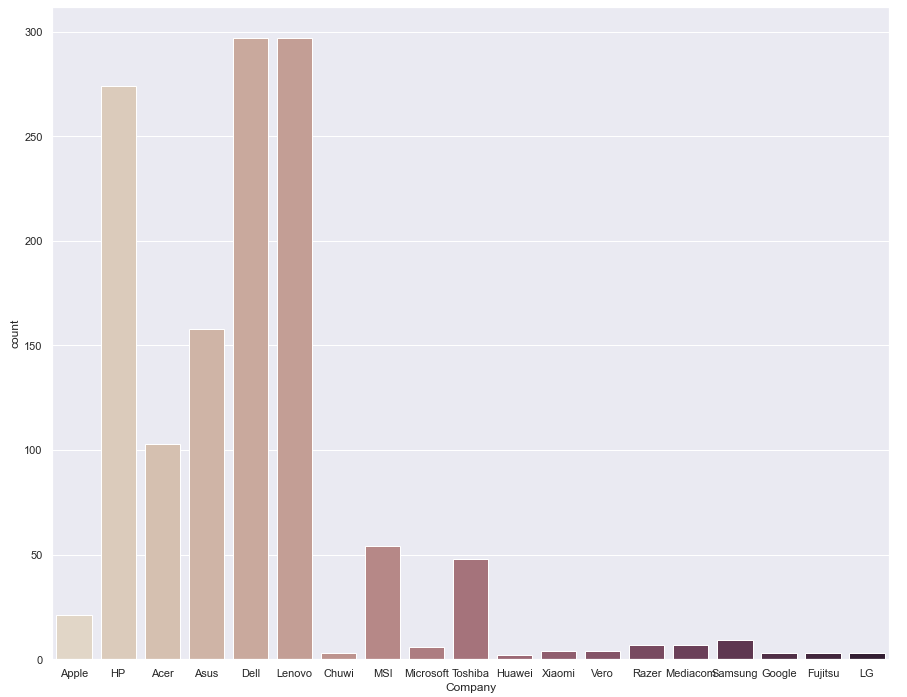
Key Insights from the data.

* The more the size of the laptop is, the more weight it'll carry.

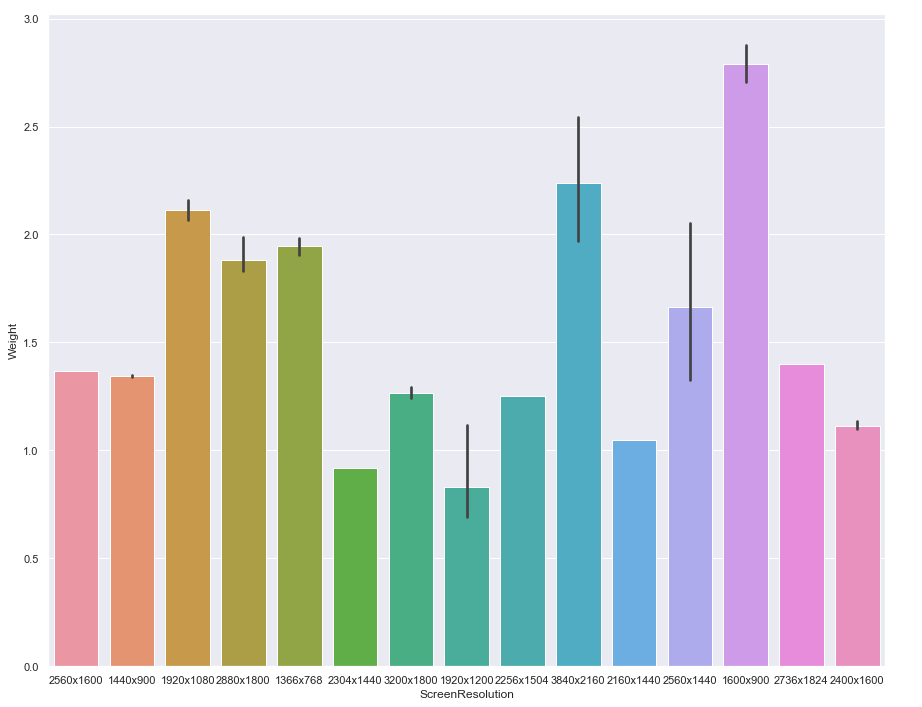




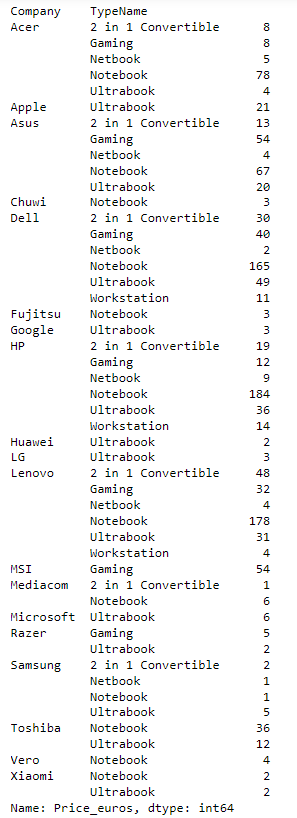
* Dell and Lenovo has the highest number of sales.



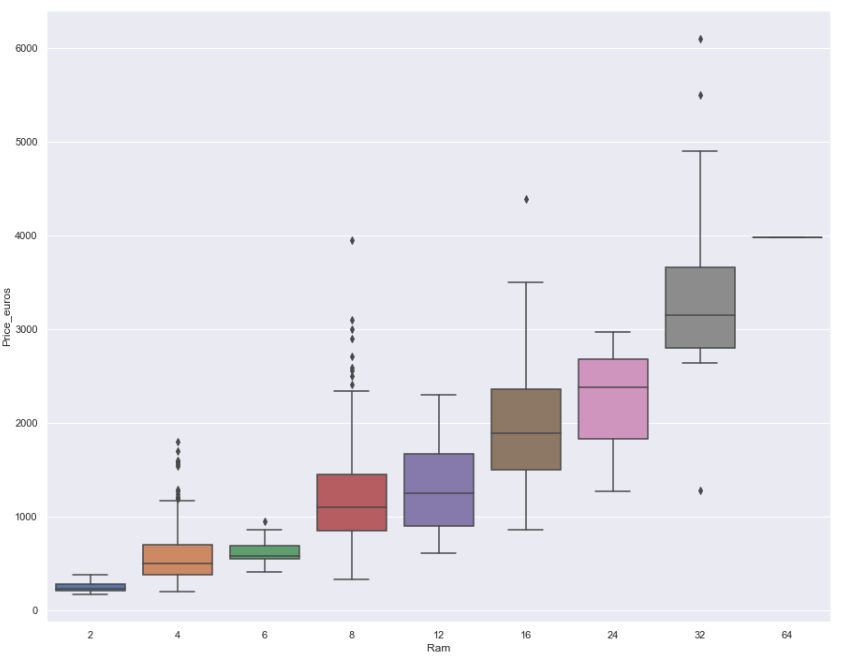
* The most preferred resolution is "1920x1080".



* The maximum number of "Gaming" laptops are being offered by two main companies "Asus" & "MSI", where "MSI" has only Gaming laptops.



* There are many outliers in 8GB RAM laptops. The most costly ones are the 32GB RAM one's.



2) Answer the following questions.

- 1 What are the assumptions of Linear Regression?

Ans. [**Linear regression**](https://www.statisticssolutions.com/free-resources/directory-of-statistical-analyses/what-is-linear-regression/) is an analysis that assesses whether one or more predictor variables explain the dependent (criterion) variable.  The regression has five key assumptions:

* Linearity – There must exist a linear relationship b/w the variables.
* Normality – The distribution should tend towards a normal distribution.
* No or less multi Collinearity – Since multicollinearity causes many problems.
* No auto correlation – A variable should not have a relation with it’s own self.
* Homoscedasity – Constant variance b/w the error terms.

-2 Can we evaluate a Regression model? Define each metric and its interpretation.

Ans. R-squared : R Square is calculated by the sum of squared of prediction error divided by the total sum of the square which replaces the calculated prediction with mean.

Adjusted R-squared: When we have more number of variables, then R-squared isn’t a good measure of accuracy. On the other hand, Adjusted R-squared balances with the number of variables present.

MSE: MSE is calculated by the sum of square of prediction error which is real output minus predicted output and then divide by the number of data points.

RMSE: Root Mean Square Error(RMSE) is the square root of MSE. It is used more commonly than MSE because sometimes MSE value can be too big to compare easily. And square root brings it to the same level of prediction.

MAE: Mean Absolute Error takes the sum of the absolute value of error.

-3 Can R squared be negative?

R2 is negative only when the chosen model does not follow the trend of the data, so fits worse than a horizontal line.

-4 What is a Dummy Variable?

Ans. Dummy variable is a variable that just takes the value of just 0 or 1.

-5 Is One Hot Encoding different from Dummy Variables?

In one-hot encoding, we create a new set of dummy (binary) variables that is equal to the number of categories (k) in the variable.

-6 How is polynomial regression different from linear regression?

Ans. Polynomial regression is a form of Linear regression where only due to the Non-linear relationship between dependent and independent variables we add some polynomial terms to linear regression to convert it into Polynomial regression.

-7 Interpret the screenshot below from the notebook we discussed in class today.

Ans. Since, model.score gives us the coefficient of determination.

And also since we are calculating it between the variable and our predicted values of the same variable through our model, it comes 1.0 . That means our model fits perfectly fine.

-8 Bonus: We saw Sweetviz as an Automated EDA option. What are the other options? Try a few of them and share which one did you find the best.

Ans. Some other options are: autoviz, pandas profiling, dtale.