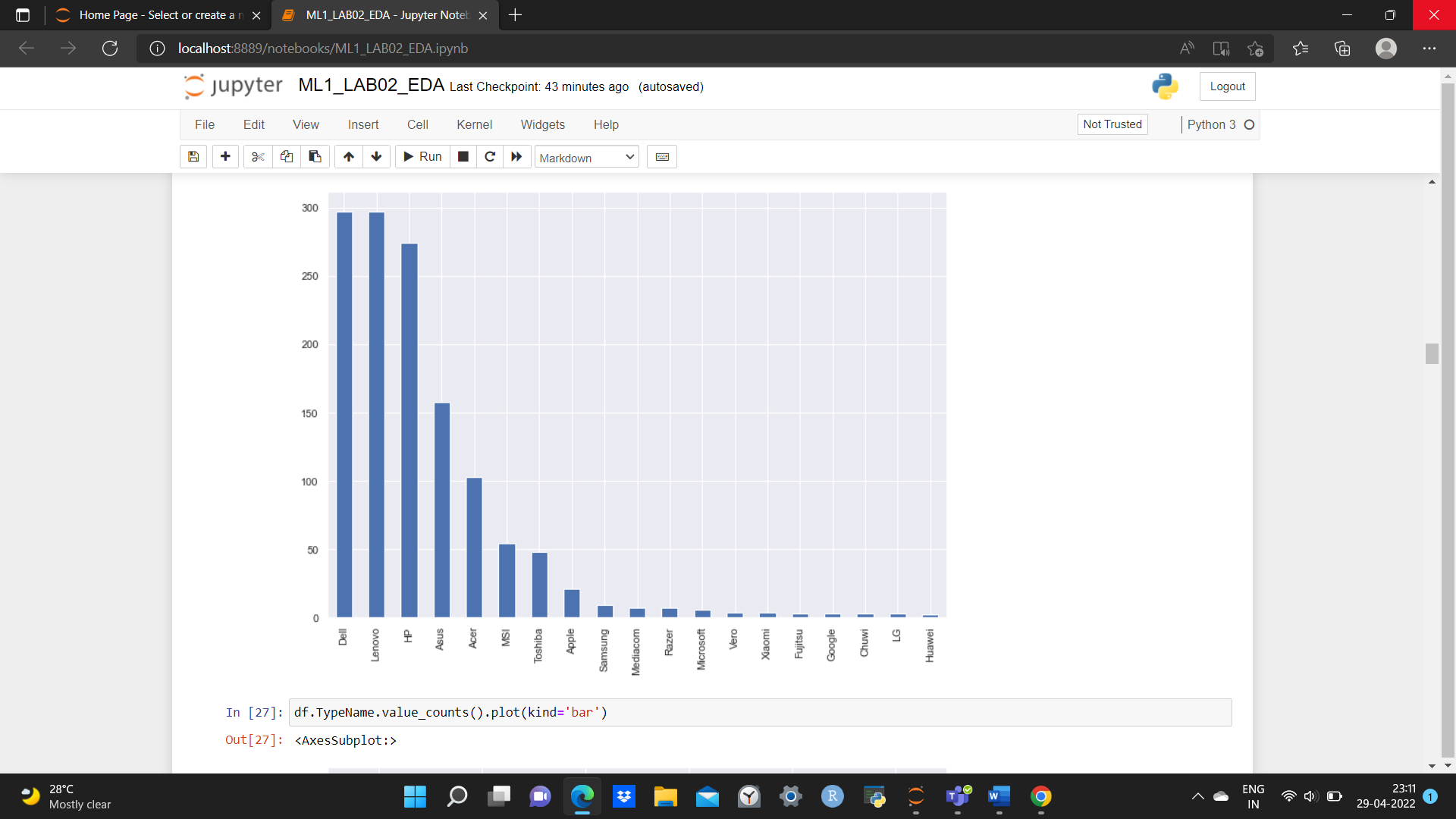
**ML-LAB02**

**ASSIGNMENT-02**

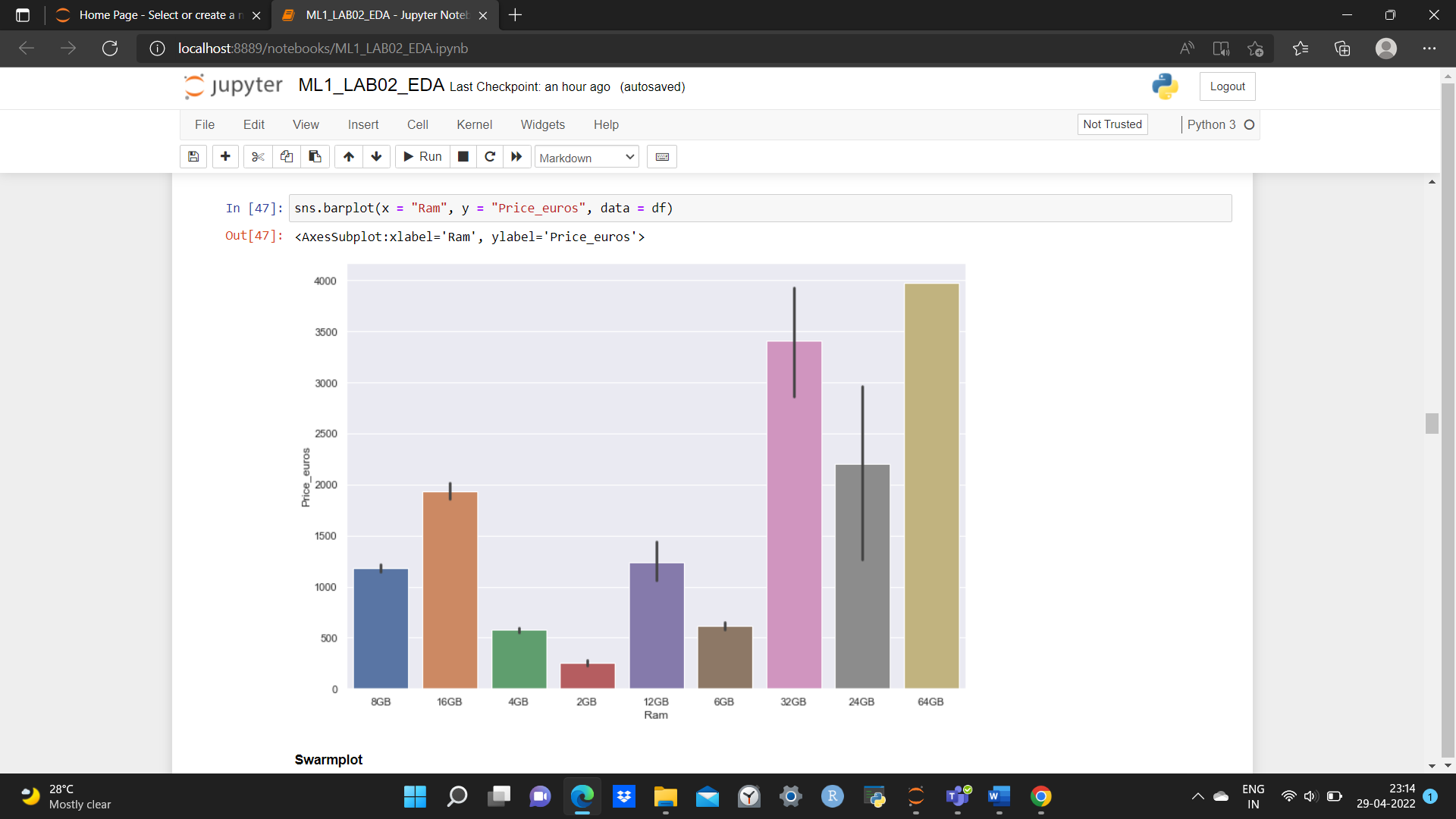
**Aparna k**

**21BDA24**

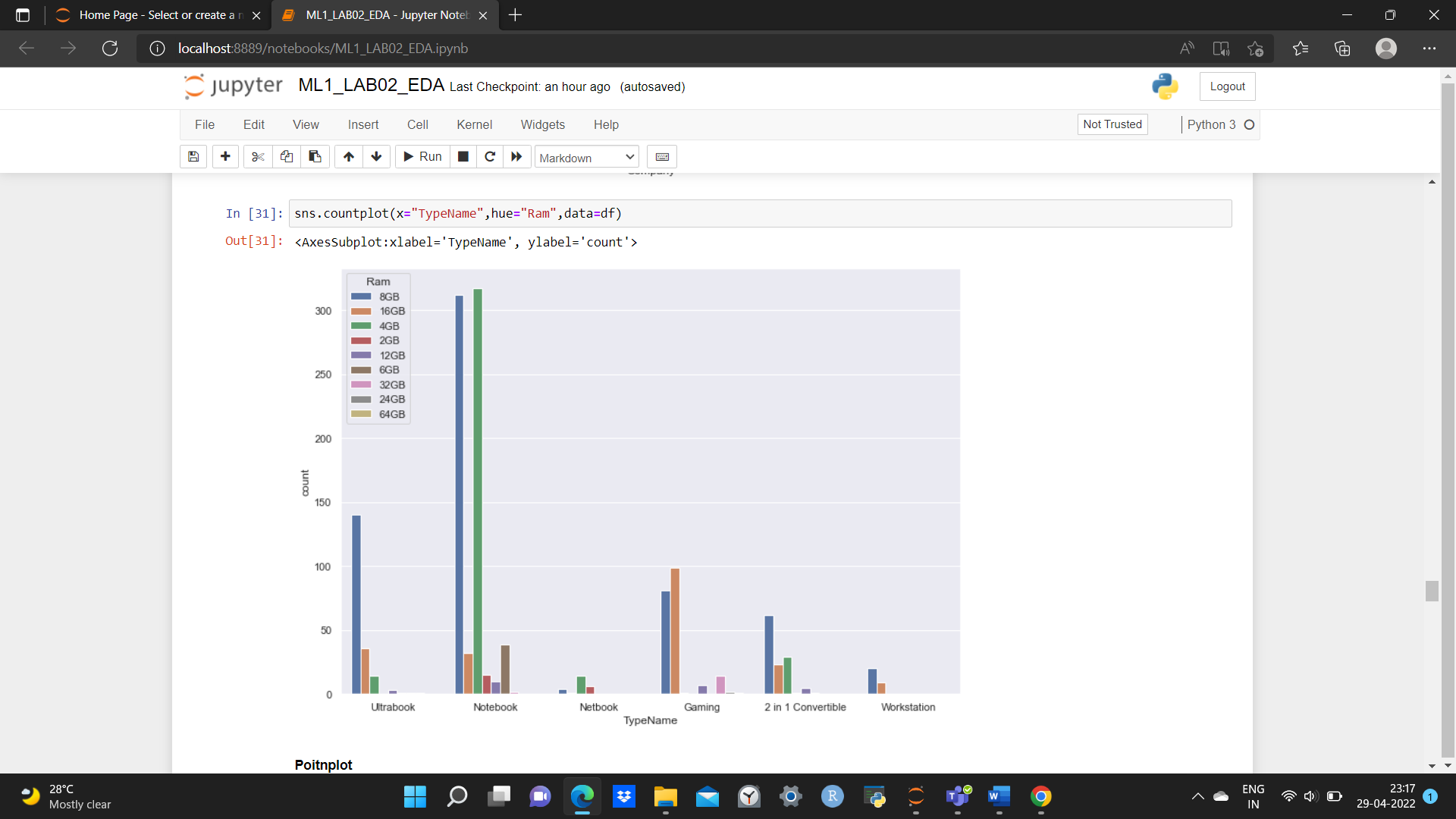
* **Insights**



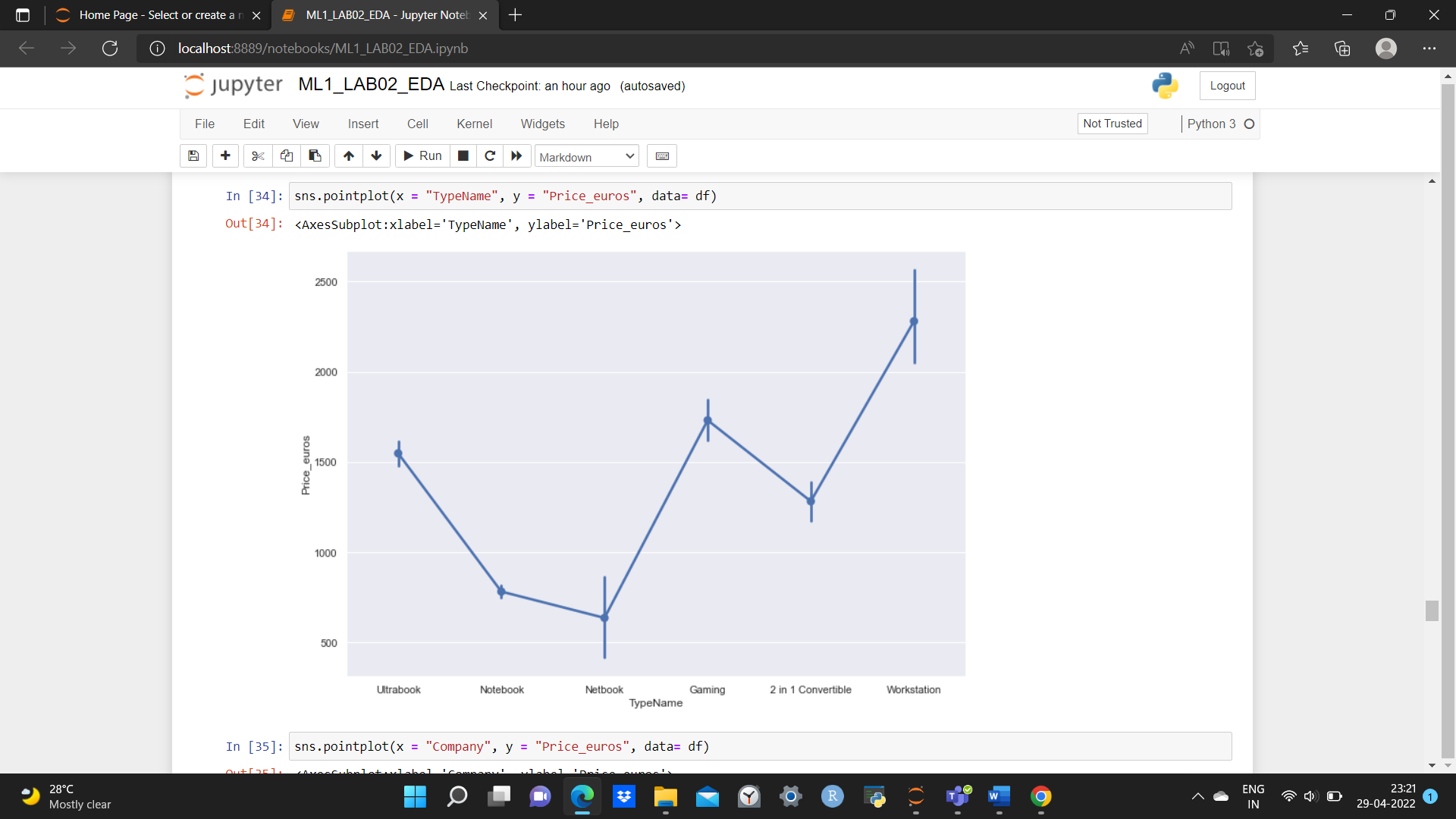
Dell is the most popular laptop company



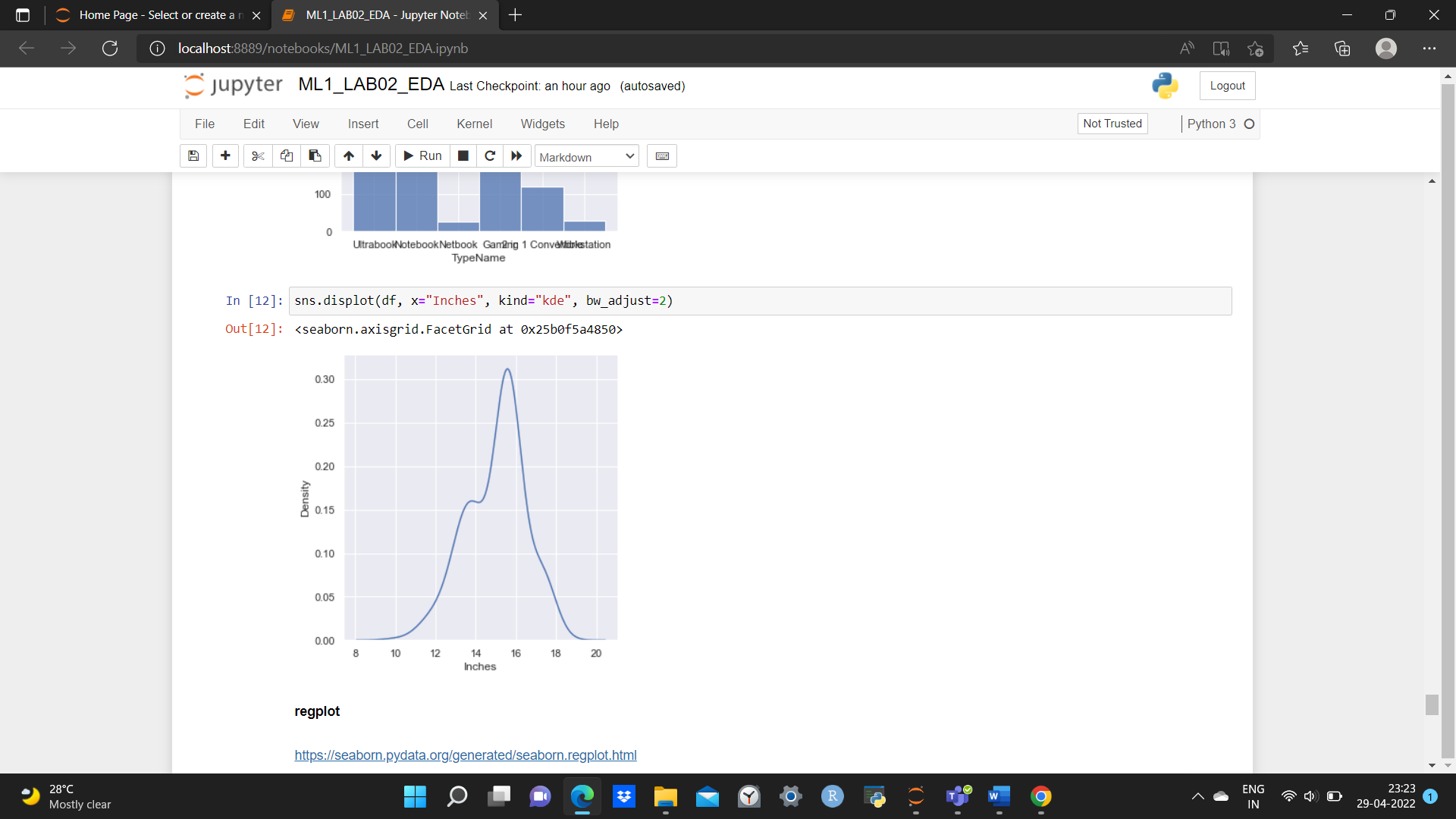
Laptops with 64GB RAM costs more than others



Notebook with 4GB and 8GB RAM are more in number, indicating that they are mostly bought.



Workstation is the expensive laptop type.



14–16-inch laptops are more in number, indicating that people prefer them .

1. What are the assumptions of linear regression?

* The Two Variables Should be in a Linear Relationship
* All the Variables Should be Multivariate Normal
* There Should be No Multicollinearity in the Data
* There Should be No Autocorrelation in the Data
* There Should be Homoscedasticity Among the Data

1. How can we evaluate a Regression model? Define each metric and its interpretation.

The performance of a regression model can be understood by knowing the error rate of the predictions made by the model. To measure the performance of your regression model,  some statistical metrics are used. Some of them are:

* Mean Absolute Error(MAE)
* Root Mean Square  Error(RMSE)
* Coefficient of determination or R2
* Adjusted R2

1.Mean Absolute Error: This is the simplest of all the metrics. It is measured by taking the average of the absolute difference between actual values and the predictions.

2. Root Mean Square  Error(RMSE): The Root Mean Square Error is measured by taking the square root of the average of the squared difference between the prediction and the actual value. It represents the sample standard deviation of the differences between predicted values and observed values(also called residuals).

3. Coefficient of Determination or R2: It measures how well the actual outcomes are replicated by the regression line. It helps you to understand how well the independent variable adjusted with the variance in your model. That means how good is your model for a dataset.

4.Adjusted R-squared: The adjusted R-squared is a modified version of R-squared that accounts for predictors that are not significant in a regression model. In other words, the adjusted R-squared shows whether adding additional predictors improve a regression model or not.

1. Can R squared be negative?

It is possible to get a negative R-square for equations that do not contain a constant term. Because R-square is defined as the proportion of variance explained by the fit, if the fit is actually worse than just fitting a horizontal line then R-square is negative.

1. what is dummy variable trap?

**when the number of dummy variables created is equal to the number of values the categorical value can take on**. this leads to multicollinearity, which causes incorrect calculations of regression coefficients and p-values.

1. Is One Hot Encoding different from Dummy Variables?

Dummy variables: each category is converted to its own column and the value 0 or 1 indicates if that category is present for each record  
one-hot-encoding: similar to dummy variables, but one column is dropped, as its value can be derived from the other columns. This is to prevent multicollinearity and the [dummy variable trap](https://en.wikipedia.org/wiki/Dummy_variable_(statistics)).

1. How is polynomial regression different from linear regression?

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.