**Applied Statistics - Lab 6**

**Date of performance:** 23-09-2021

**Date of Submission:** 28-09-2021

**SAP Id:** 500083382

**Name:** Anurag Singh

**Batch:** AI&ML B2

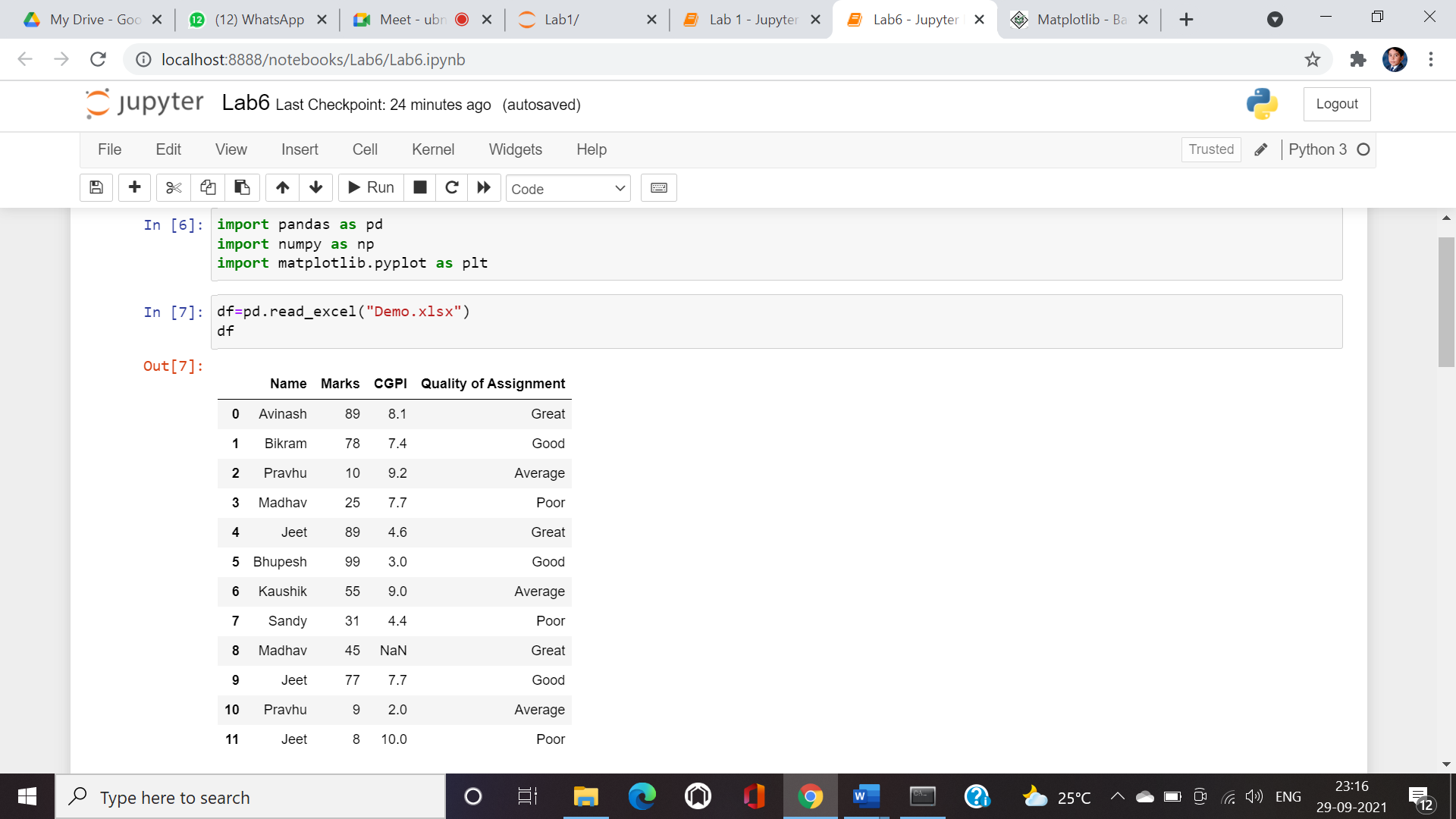
**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**Topic: Covariance and Correlation**

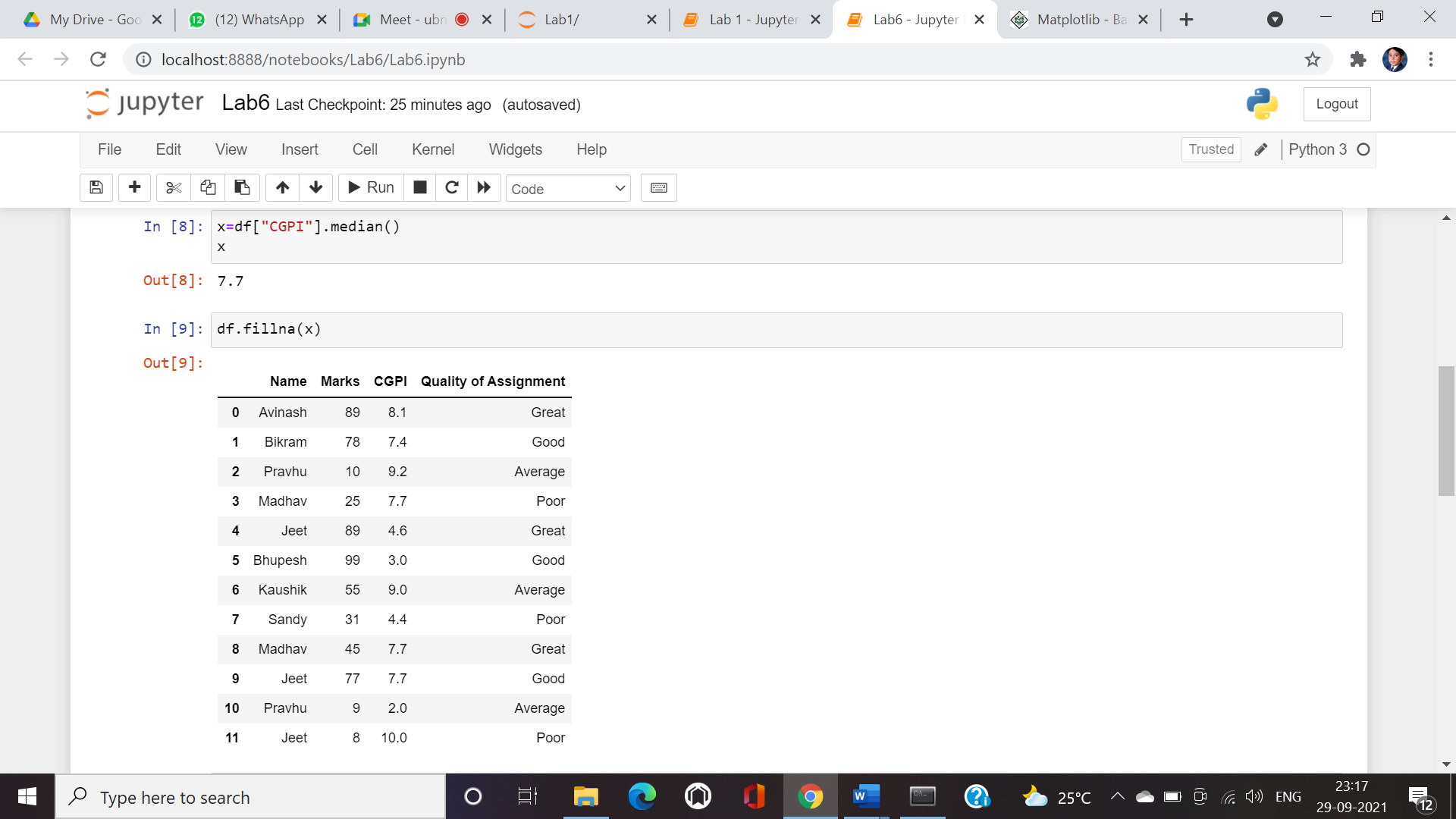
**Problem:**

In Lab 1 we discussed a dataset having variables 'Marks' and 'CGPI'. Plot both of them as two different axes. Discuss the covariance and correlation coefficient.

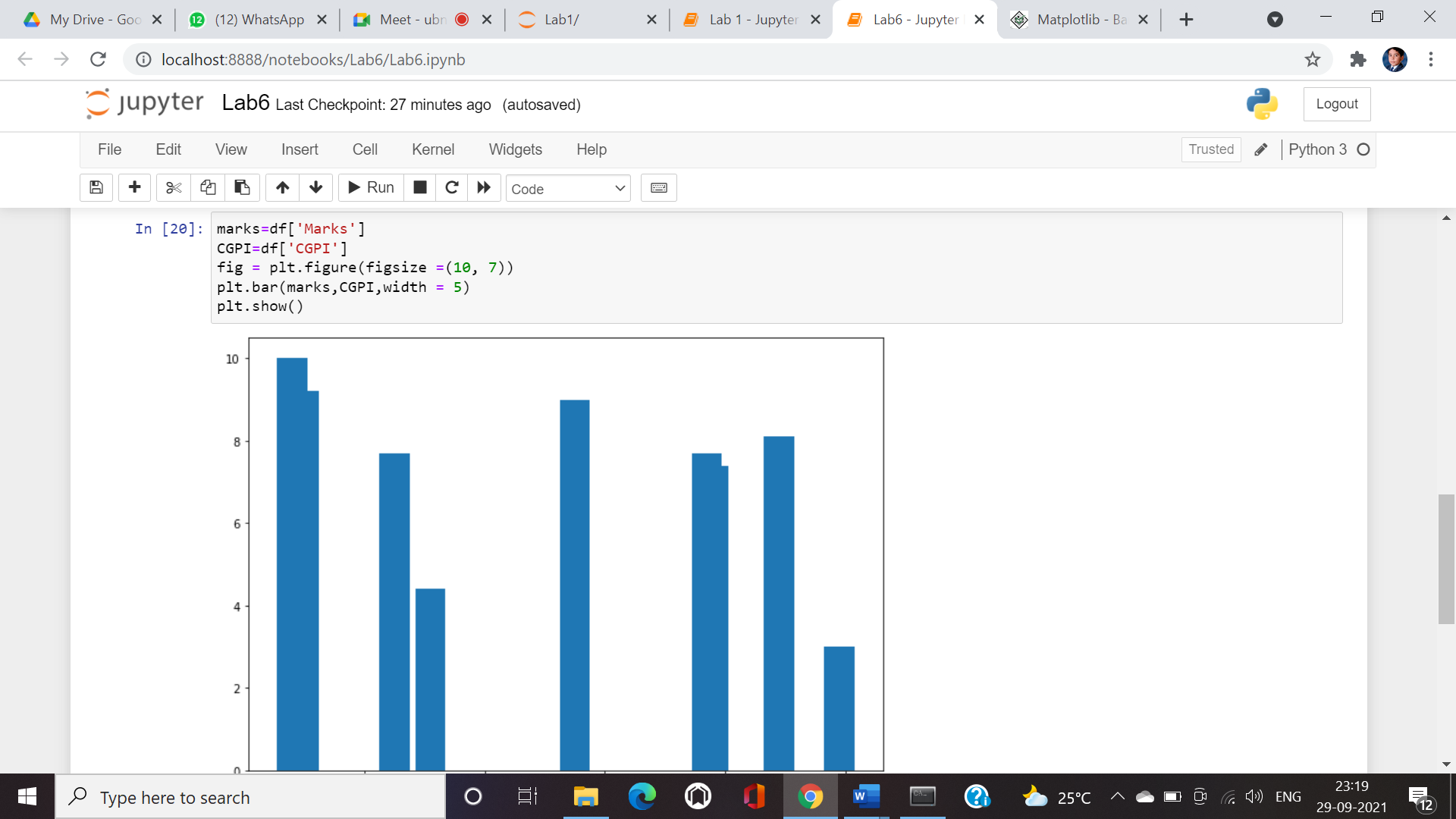
**Code to read the file as a data frame in Python:**



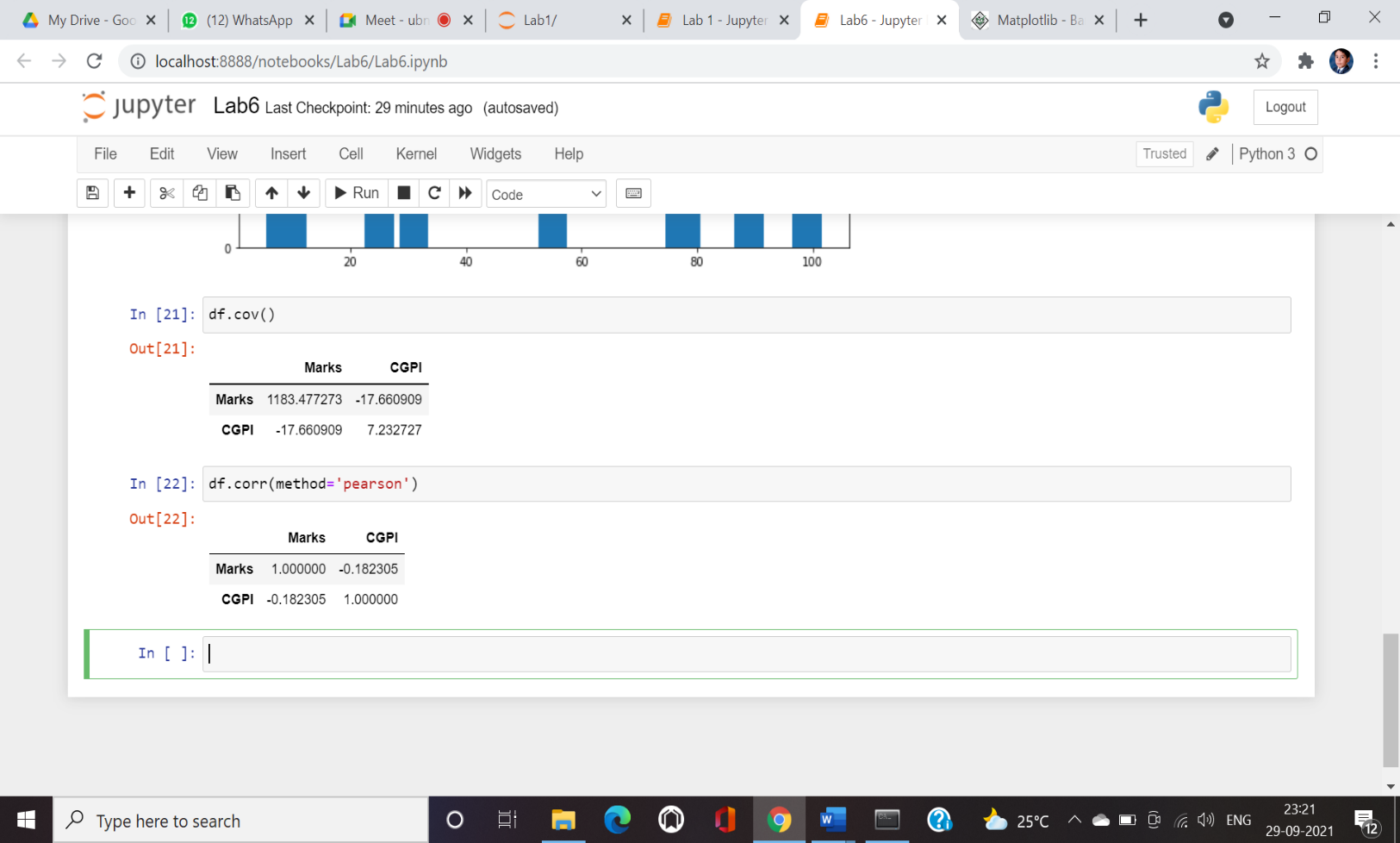
**Filling the Nan:**



**Plotting both of them as two different axes: X-axis and Y-axis**



**Covariance and Correlation:**



**Relation between Correlation and Covariance:** Both covariance and correlation measure the relationship and the dependency between two variables. Covariance indicates the direction of the linear relationship between variables while correlation measures both the strength and direction of the linear relationship between two variables. Correlation is a function of covariance. What sets these two concepts apart is the fact that correlation values are standardized whereas covariance values are not. We can obtain the correlation coefficient of two variables by dividing the covariance of these variables by the product of the standard deviations of the same values.