**Hypothesis Testing Exercise**

**Q1**

**A F&B manager wants to determine whether there is any significant difference in the diameter of the cutlet between two units. A randomly selected sample of cutlets was collected from both units and measured? Analyze the data and draw inferences at 5% significance level. Please state the assumptions and tests that you carried out to check validity of the assumptions.**

**Ans =>**

Let’s consider,H

H0: There is no difference between the diameter of cutlets .

H1: There is difference between the diameter of cutlets.

Lets consider diameter d1,d2 be the respective cutlet diameter of 2 units ,unit1 and unit2 Therefore

Thus difference between them be d1-d2= [θ](https://www.greeksymbols.net/theta-symbol)

[Therefore,](https://www.greeksymbols.net/theta-symbol)

[H](https://www.greeksymbols.net/theta-symbol)[0](https://www.greeksymbols.net/theta-symbol)[: θ=0 (There is no difference between the diameter of cutlets .)](https://www.greeksymbols.net/theta-symbol)

[H](https://www.greeksymbols.net/theta-symbol)[1](https://www.greeksymbols.net/theta-symbol)[: θ≠0 (There is difference between the diameter of cutlets.)](https://www.greeksymbols.net/theta-symbol)

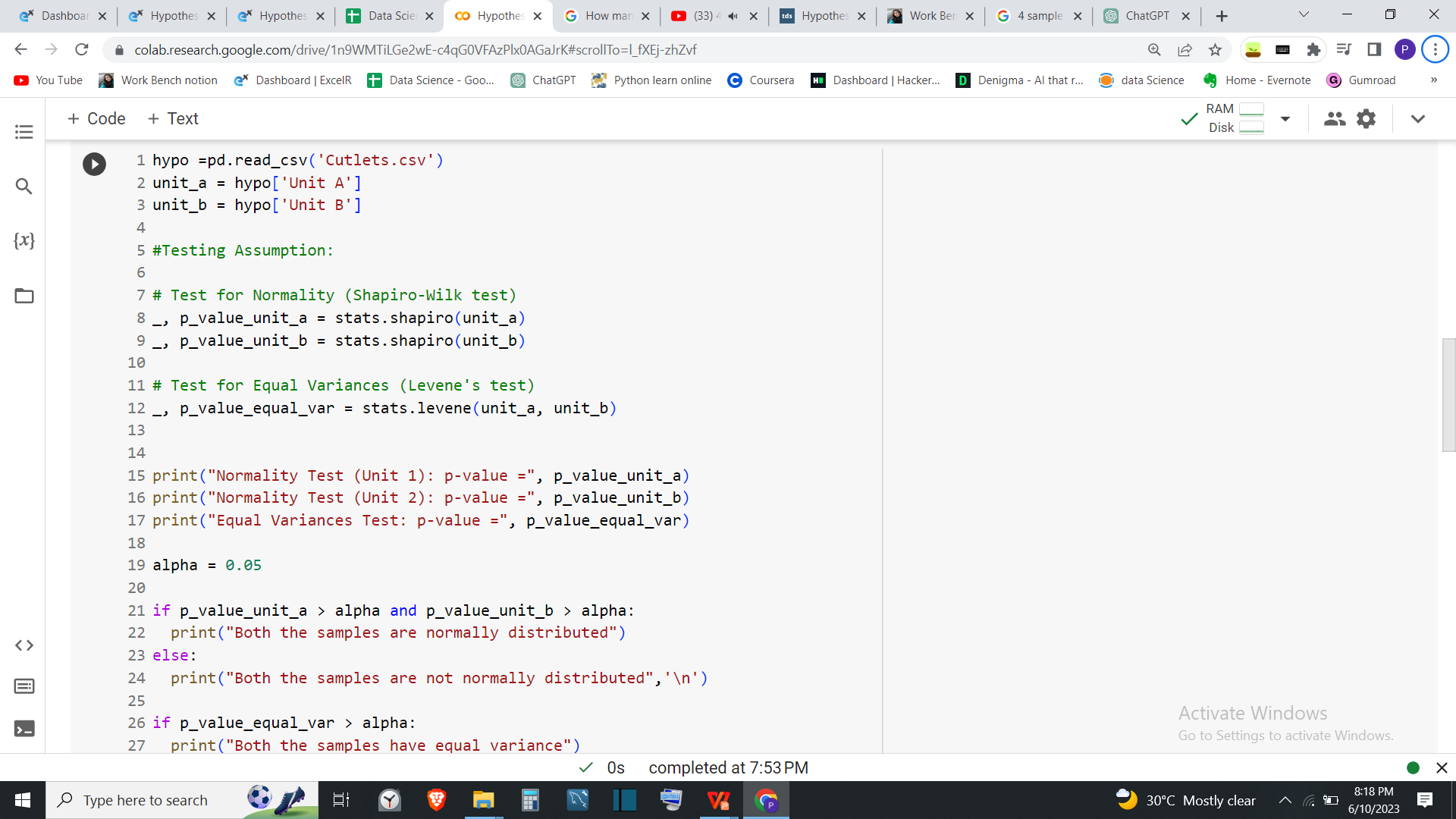
**[Assumptions](https://www.greeksymbols.net/theta-symbol)** [:](https://www.greeksymbols.net/theta-symbol)

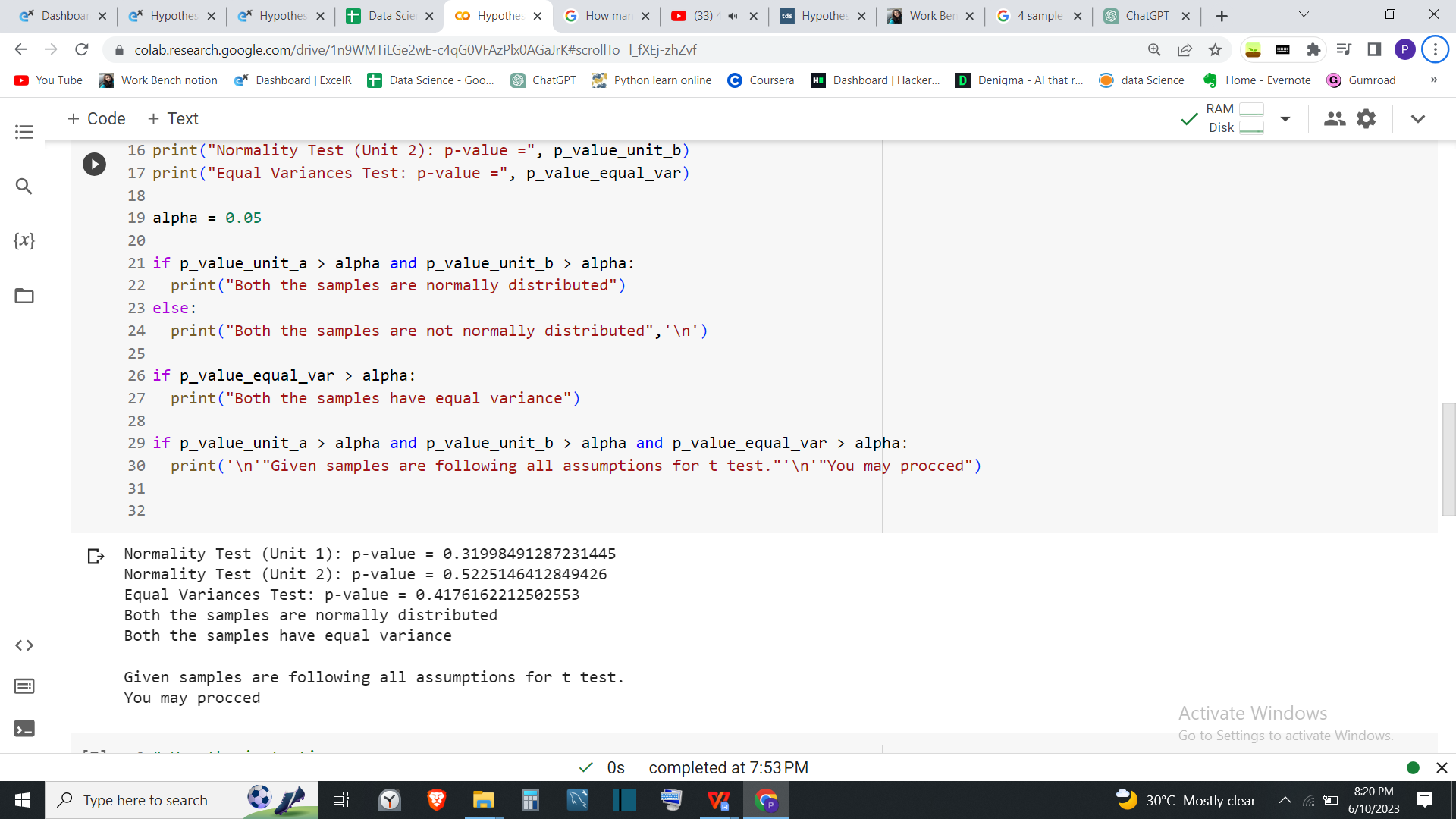
1. The samples collected from both units are random and representative of their respective populations.
2. The distribution of the diameter measurements within each unit is approximately normal.
3. The variances of the diameter measurements in both units are equal.

We are going to use 2 Independent sample t test for this example.

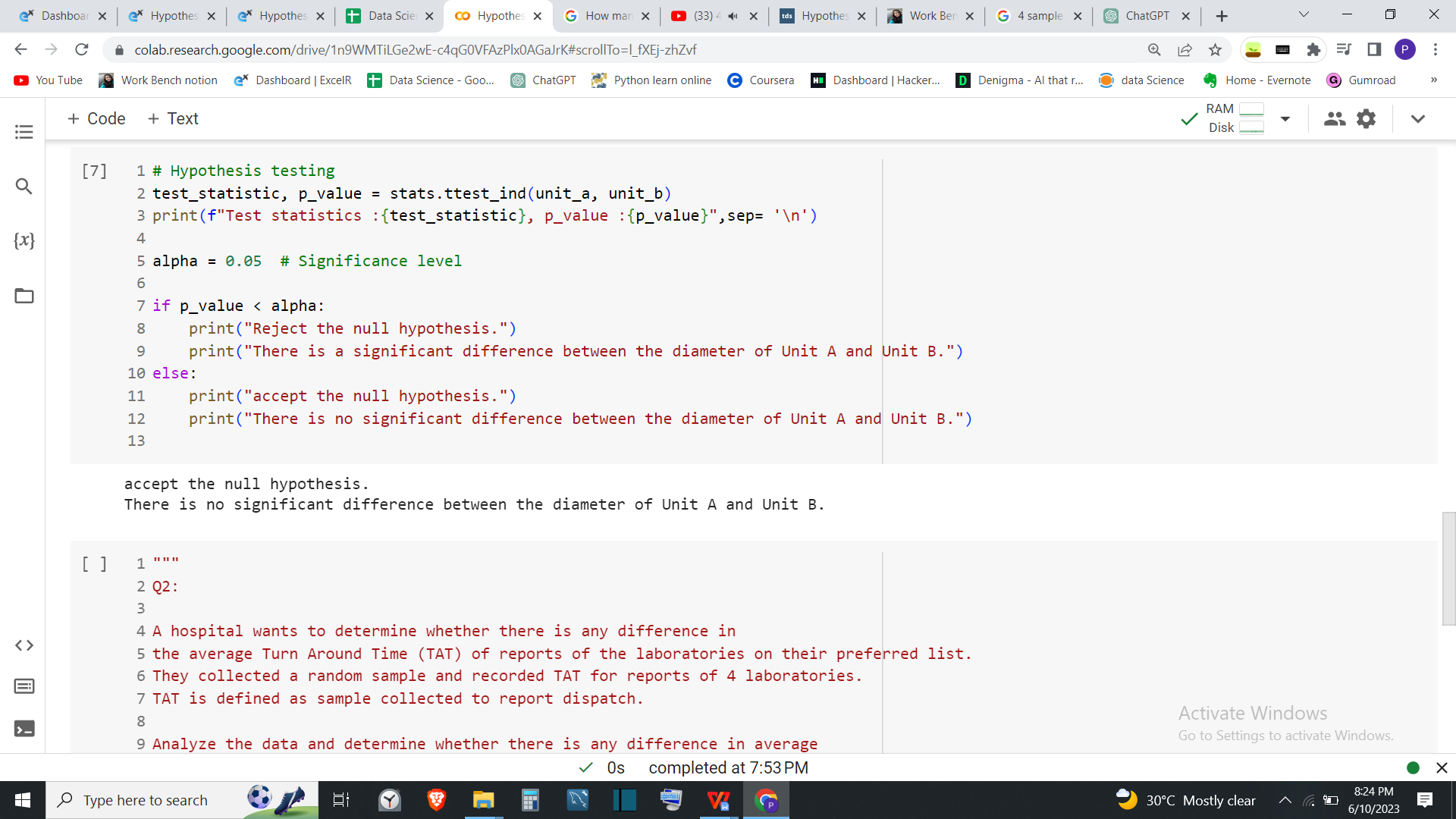
Python codes and result:

**Testing assumptions:**





Now test of hypothesis:



**Result:**

With 5% of significance level we can say that, there is significant difference in the diameter of the cutlet between two units.

**Q.2**

**A hospital wants to determine whether there is any difference in the average Turn Around Time (TAT) of reports of the laboratories on their preferred list. They collected a random sample and recorded TAT for reports of 4 laboratories. TAT is defined as sample collected to report dispatch.**

**Analyze the data and determine whether there is any difference in average TAT among the different laboratories at 5% significance level.**

**Ans =>**

Let’s consider,

H0: There is **no difference** between average Turn Around Time of reports of 4 laboratories .

H1: There **is difference** between average Turn Around Time of reports of 4 laboratories .

**[Assumptions](https://www.greeksymbols.net/theta-symbol)** [:](https://www.greeksymbols.net/theta-symbol)

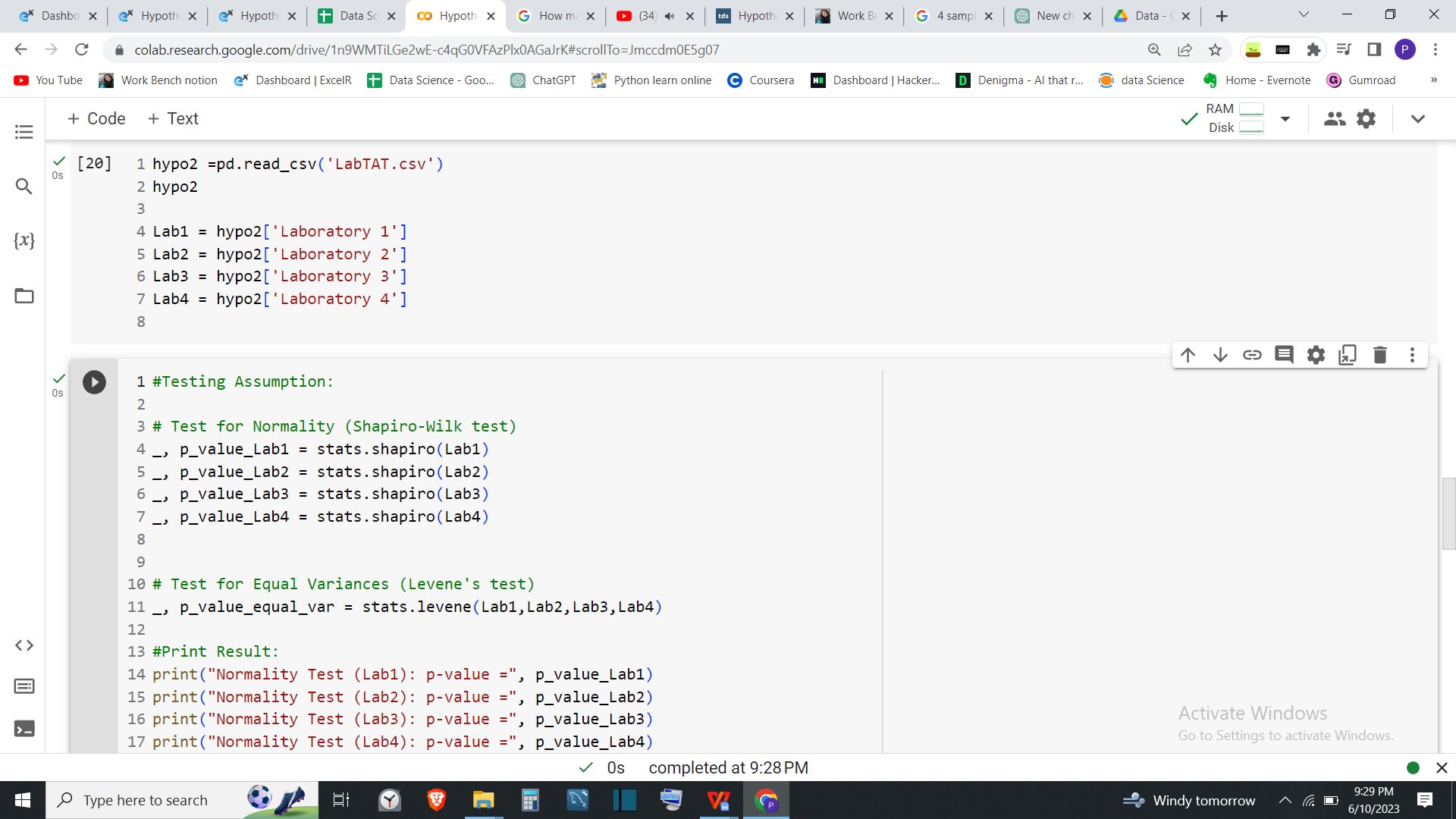
1.The samples collected are random and representative of their respective populations.

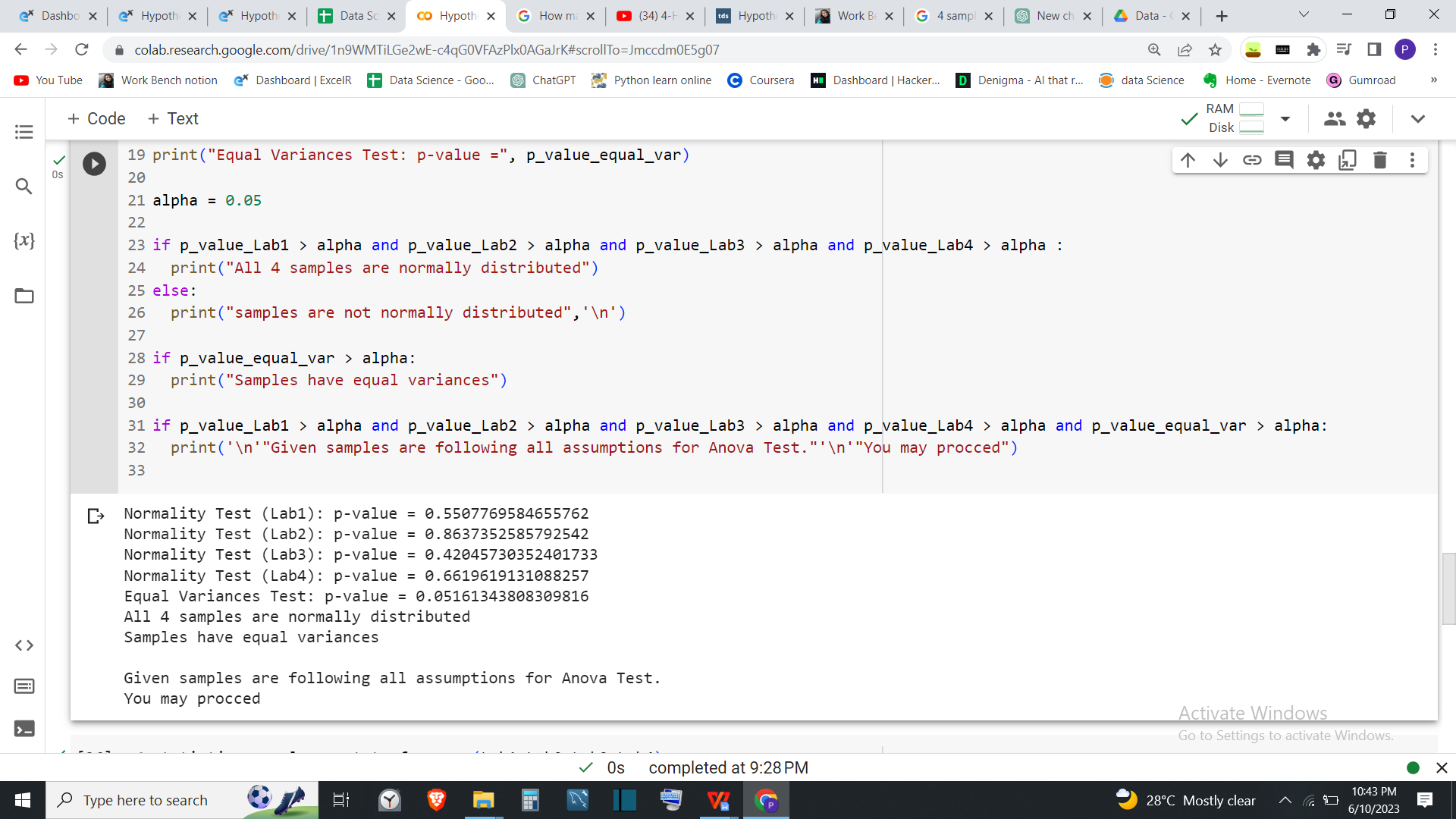
1. The distribution of the sample data is approximately normal.
2. The variances of the all the sample data are equal.

We are going to use one way Anova test because there are more than 2 samples in given problem and there is one independent variable in the problem.

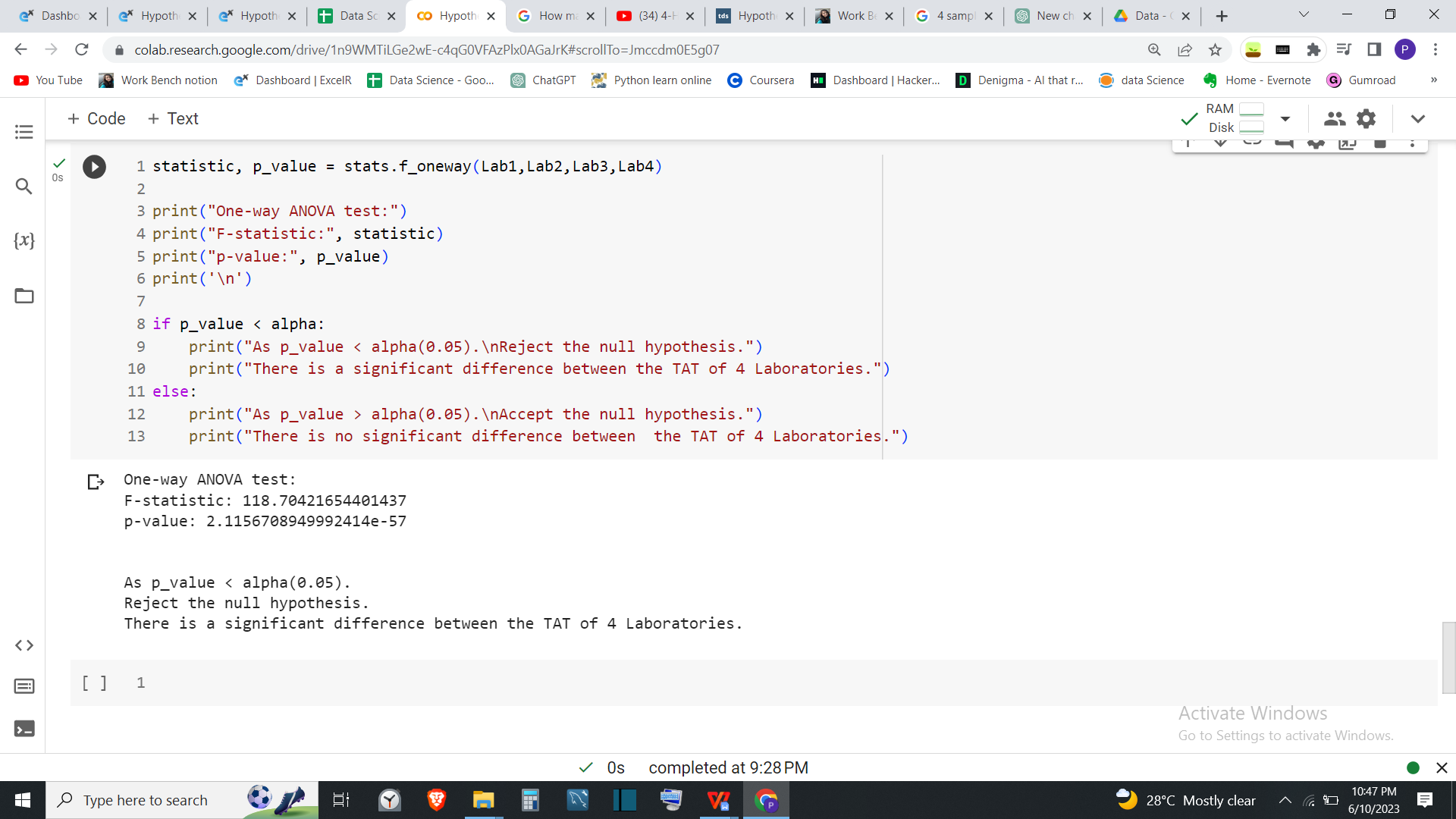
Python codes and result:

**Testing assumptions:**





**Testing Hypothesis:**



**Result:**

With 5% of significance level we can say that ,there **is a difference** between the average Turn Around Time of reports of 4 laboratories.

**Q3**

**Sales of products in four different regions is tabulated for males and females. Find if male-female buyer rations are similar across regions**.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **East** | **West** | **North** | **South** |
| Males | 50 | 142 | 131 | 70 |
| Females | 550 | 351 | 480 | 350 |

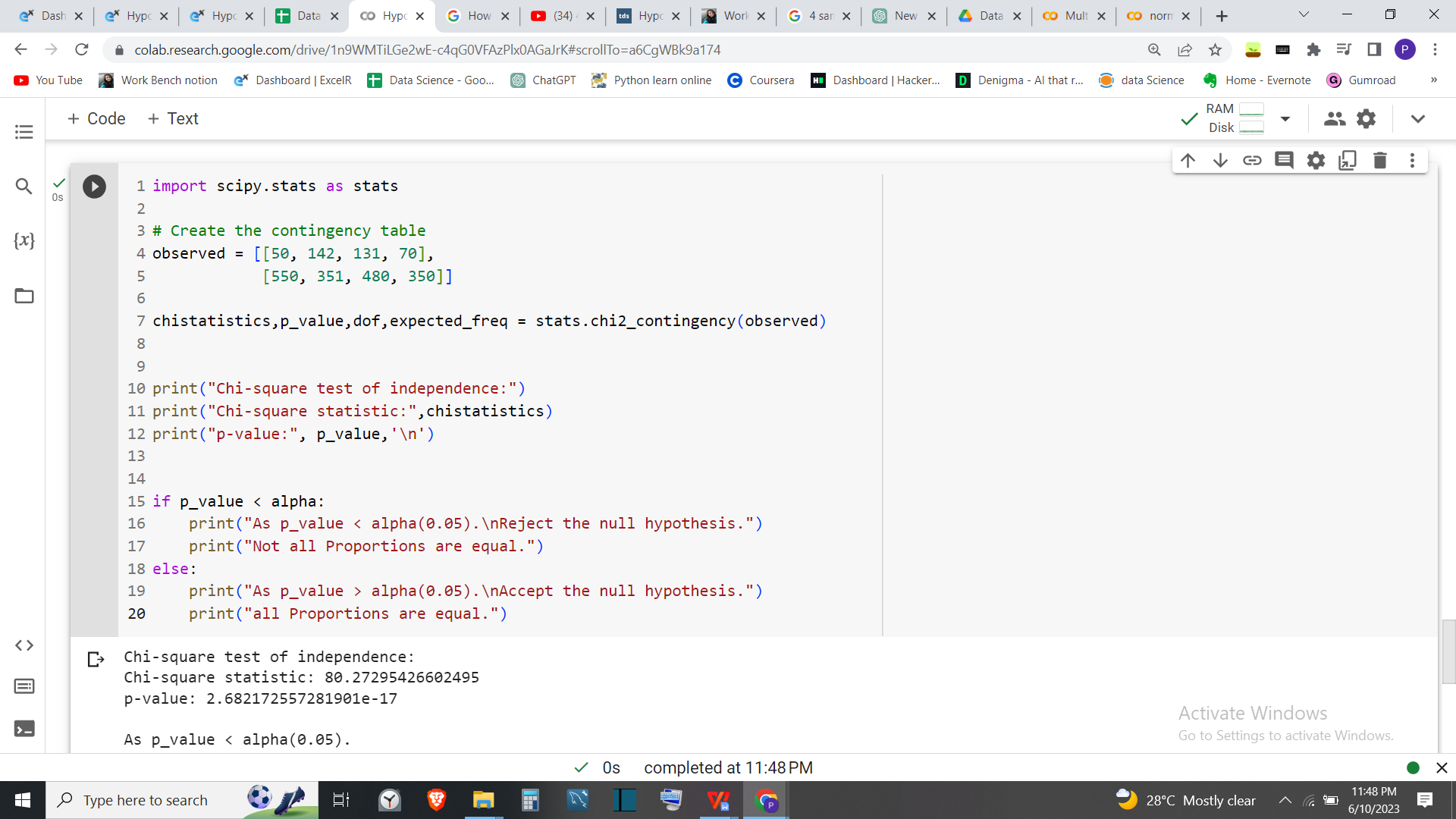
**Ans =>**

Let’s consider,

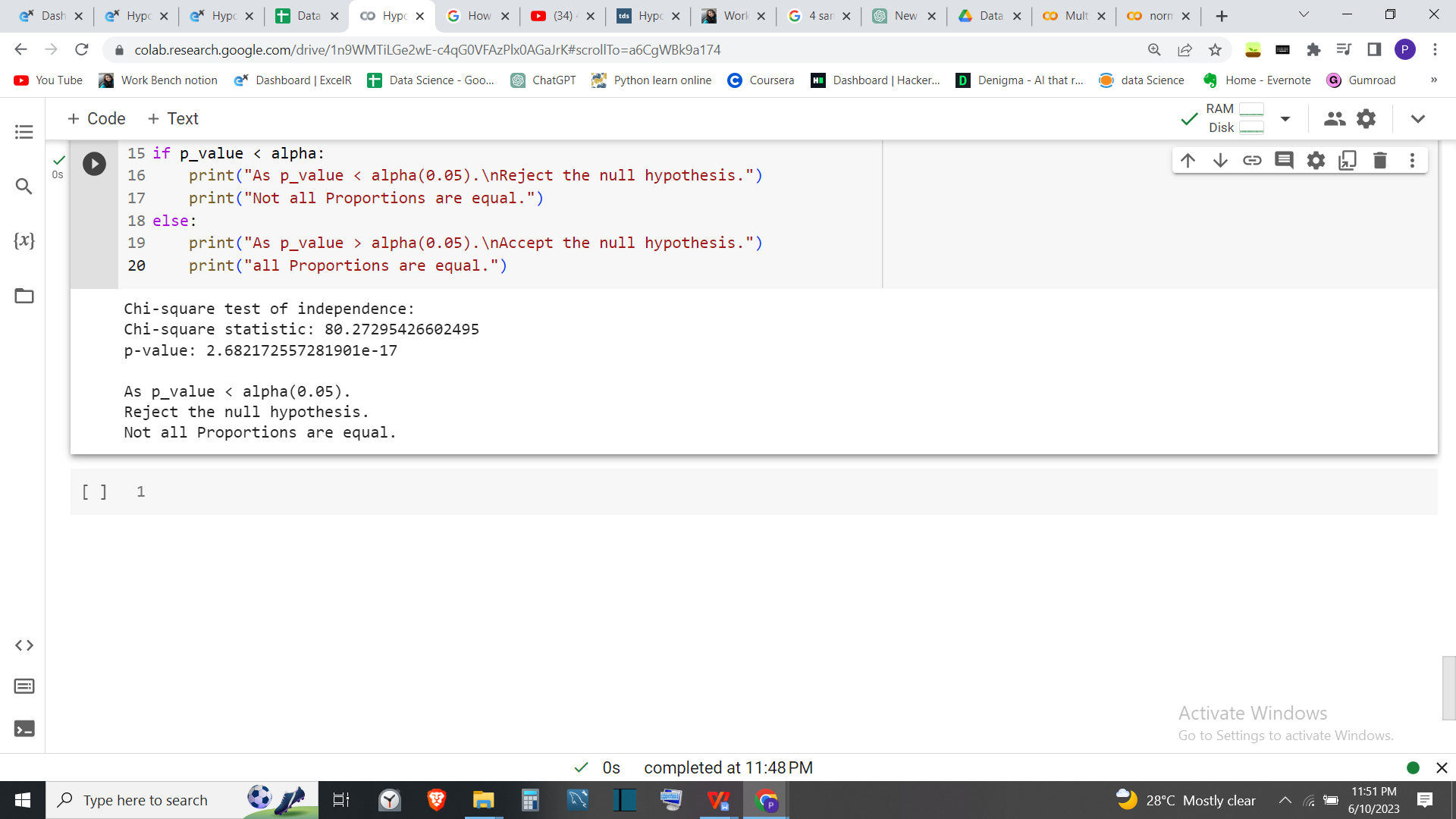
H0: All proportions are equal.

H1: Not all Proportions are equal.

**Testing Hypothesis:**



**Result**:



With 5% significance level we can say that, there is significant difference between the proportions of product sale between male and female across the regions.

**Q4:**

**TeleCall uses 4 centers around the globe to process customerorder forms. They audit a certain % of the customer order forms. Any error in order form renders it defective and has to be reworked before processing. The manager wants to check whether the defective % varies by centre. Please analyze the data at 5% significance level and help the manager draw appropriate inferences**

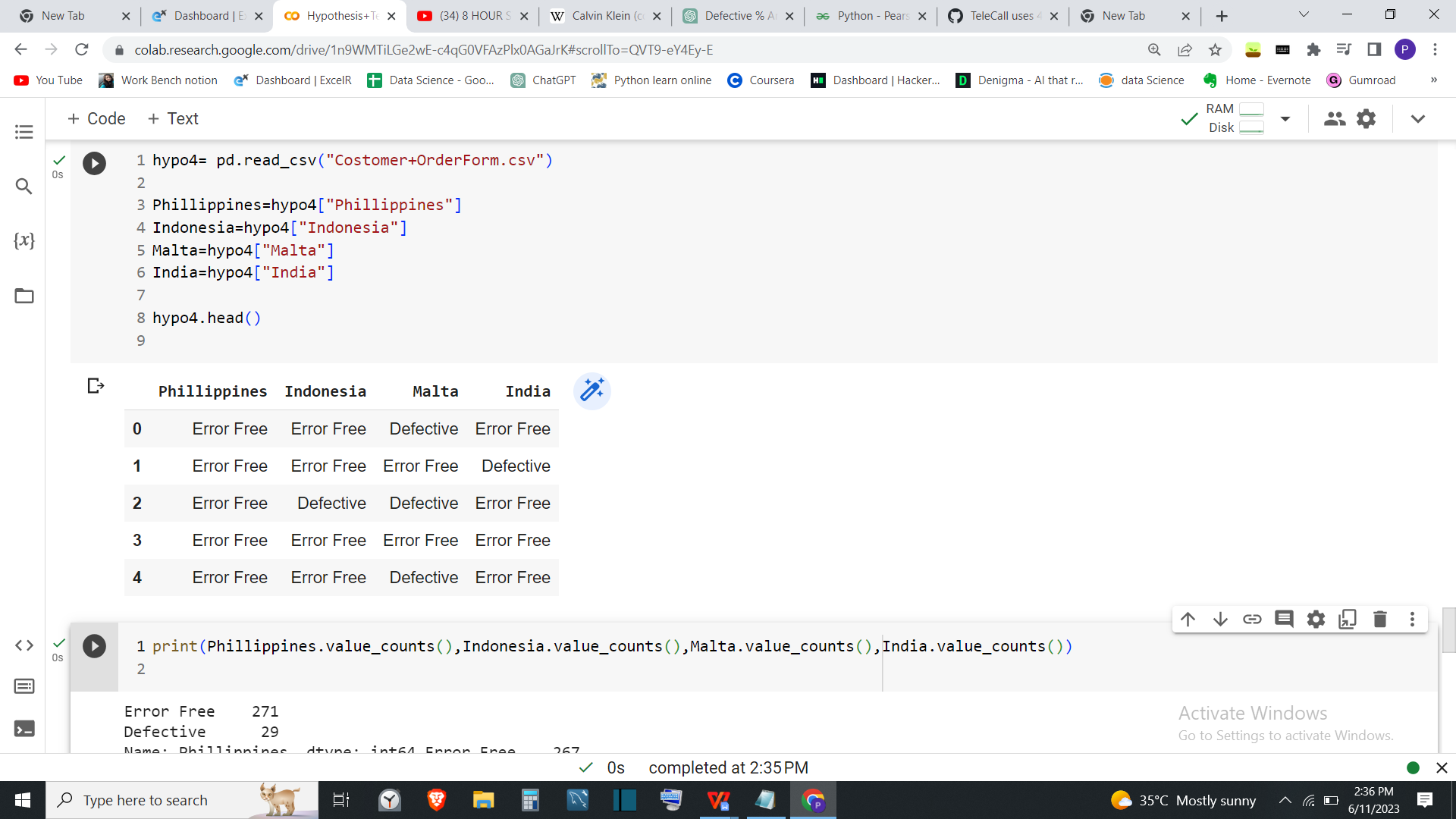
**Ans =>**

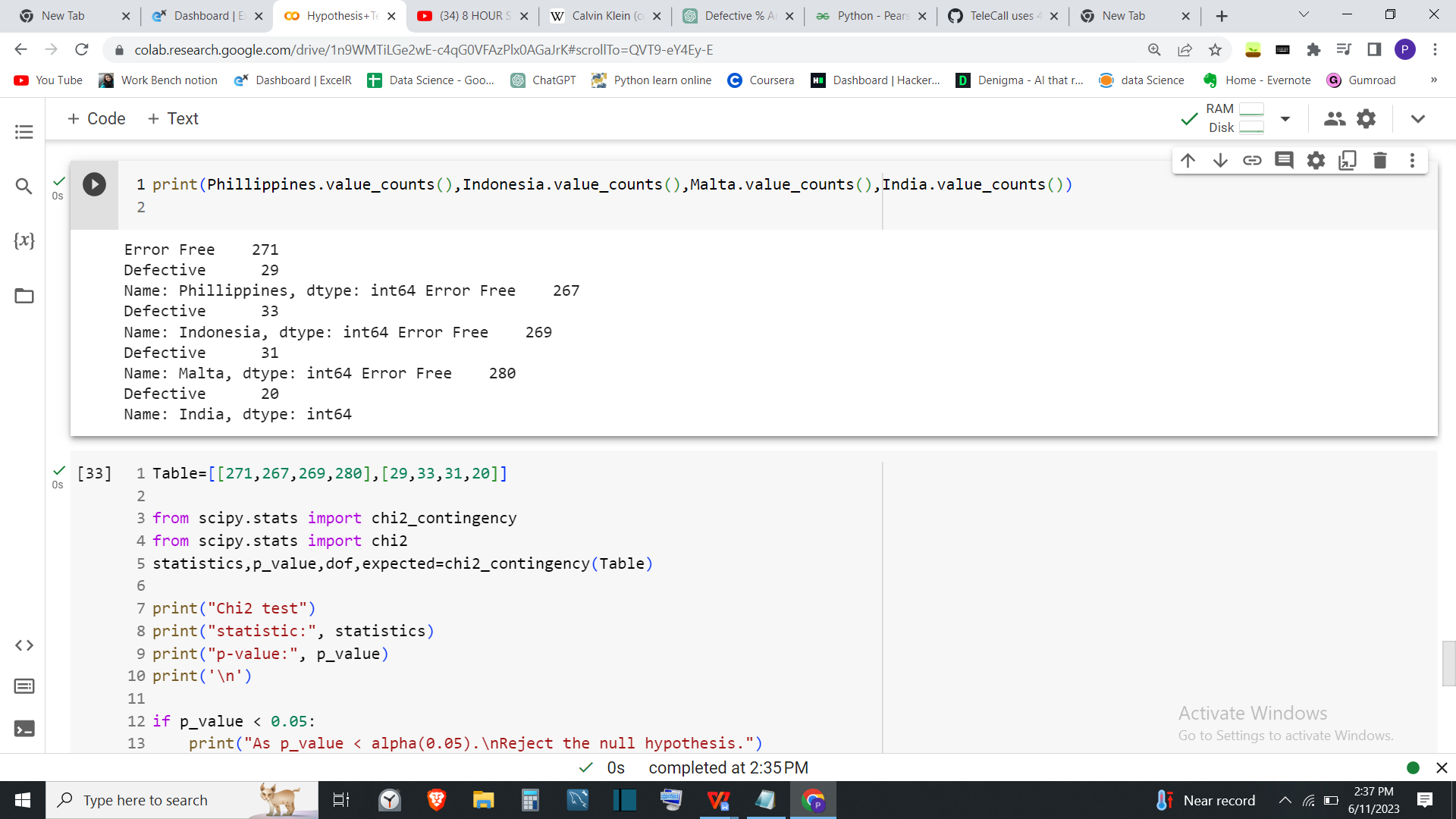
Let’s consider,

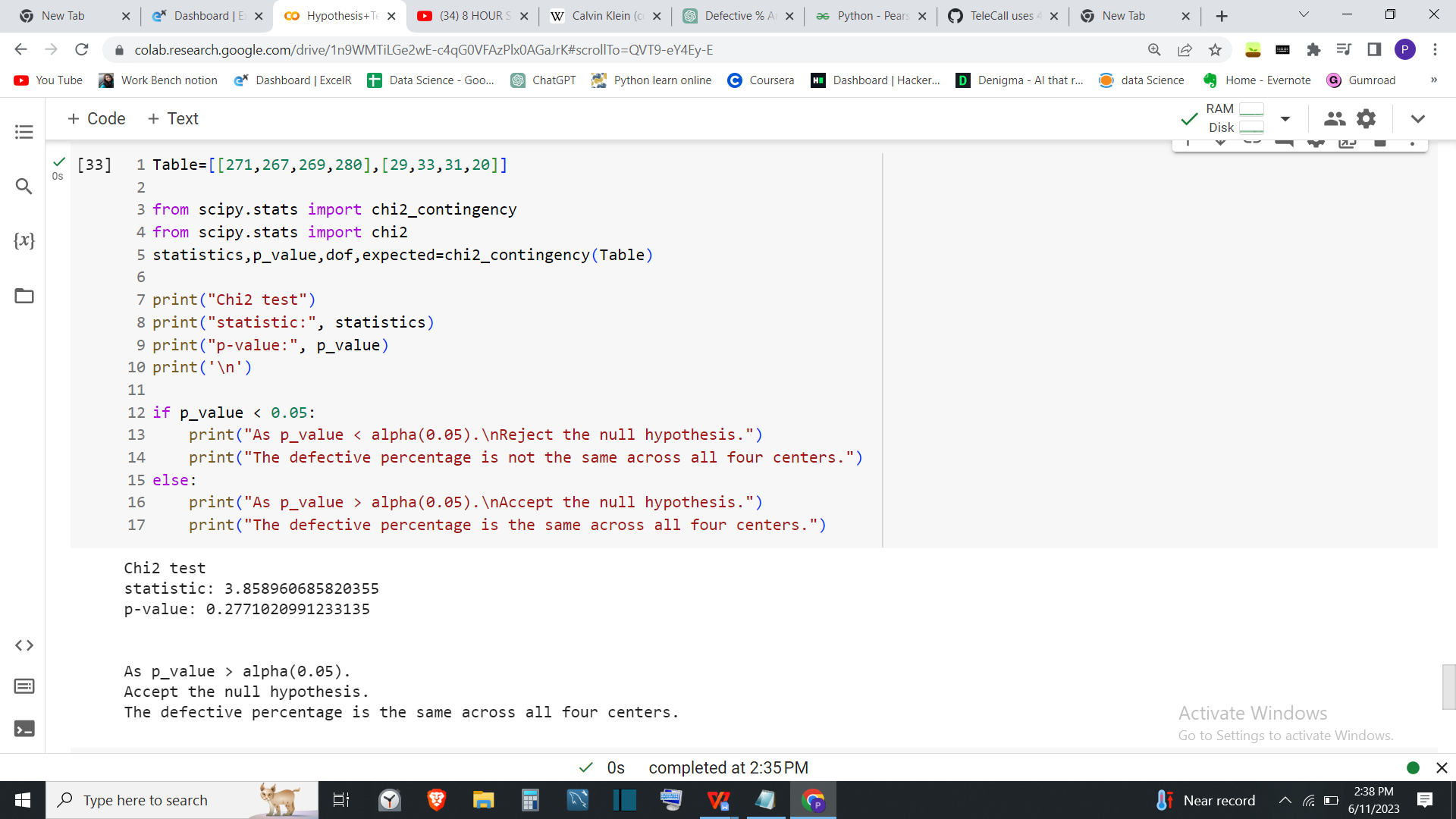
H0: The defective percentage is the same across all four centers.

H1: The defective percentage is not the same across all four centers.

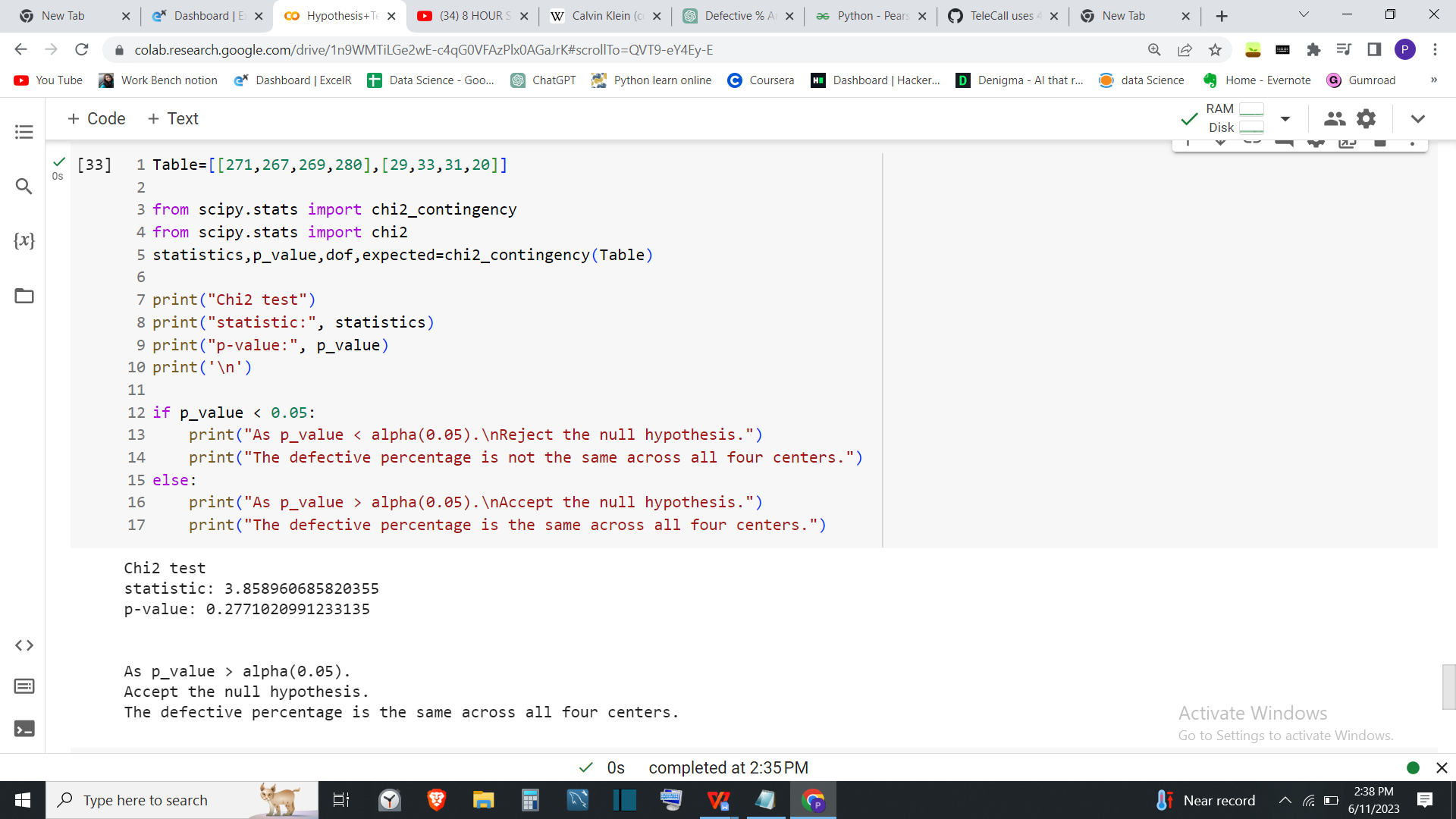
**Testing Hypothesis**







**Result**:



**Link for colab**:

https://colab.research.google.com/drive/1n9WMTiLGe2wE-c4qG0VFAzPlx0AGaJrK#scrollTo=hh1m7E7suAYk