Great Learning’s

INFERENTIAL

STATISTICS PROJECT REPORT

By

Raghvendra Singh

Email: [raghavsingh0027@gmail.com](mailto:raghavsingh0027@gmail.com)

Phone: +91-8130670022

Table of Contents

[Data & Dictionary 3](#_Toc158580733)

[Problem 3 Data Dictionary 3](#_Toc158580734)

[Problem 4 Data Dictionary 3](#_Toc158580735)

[List of Figures 4](#_Toc158580736)

[PROBLEM 1 5](#_Toc158580737)

[1.1 What is the probability that a randomly chosen player would suffer an injury? 5](#_Toc158580738)

[1.2 What is the probability that a player is a forward or a winger? 5](#_Toc158580739)

[1.3 What is the probability that a randomly chosen player plays in a striker position and has a foot injury? 5](#_Toc158580740)

[1.4 What is the probability that a randomly chosen injured player is a striker? 6](#_Toc158580741)

[PROBLEM 2 7](#_Toc158580742)

[2.1 What proportion of the gunny bags have a breaking strength of less than 3.17 kg per sq cm? 7](#_Toc158580743)

[2.2 What proportion of the gunny bags have a breaking strength of at least 3.6 kg per sq cm.? 8](#_Toc158580744)

[2.3 What proportion of the gunny bags have a breaking strength between 5 and 5.5 kg per sq cm.? 9](#_Toc158580745)

[2.4 What proportion of the gunny bags have a breaking strength NOT between 3 and 7.5 kg per sq cm.? 9](#_Toc158580746)

[PROBLEM 3 10](#_Toc158580747)

[Inferences: 12](#_Toc158580748)

[3.1 Zingaro has reason to believe that the unpolished stones may not be suitable for printing. Do you think Zingaro is justified in thinking so? 13](#_Toc158580749)

[Test Selection: 13](#_Toc158580750)

[Level of significance: 13](#_Toc158580751)

[Hypothesis: 13](#_Toc158580752)

[Conducting Ztest and calculating P value: 13](#_Toc158580753)

[Decision making: 13](#_Toc158580754)

[Conclusion 14](#_Toc158580755)

[3.2 Is the mean hardness of the polished and unpolished stones the same? 14](#_Toc158580756)

[Test Selection: 14](#_Toc158580757)

[Level of significance: 14](#_Toc158580758)

[Hypothesis: 14](#_Toc158580759)

[Conducting Independent 2 sample T-test and calculating P value: 14](#_Toc158580760)

[Conclusion 14](#_Toc158580761)

[Problem 4 15](#_Toc158580762)

[4.1 How does the hardness of implants vary depending on dentists? 17](#_Toc158580763)

[Infernece: 18](#_Toc158580764)

[HYPOTHESIS: 18](#_Toc158580765)

[TEST SELECTION: 18](#_Toc158580766)

[Shapiro-Wilk’s Test 18](#_Toc158580767)

[Levene Test 18](#_Toc158580768)

[ONE WAY ANOVA TEST 19](#_Toc158580769)

[Conclusion 19](#_Toc158580770)

[4.2 How does the hardness of implants vary depending on methods? 19](#_Toc158580771)

[Hypothesis 19](#_Toc158580772)

[Conclusion: 20](#_Toc158580773)

[Identifyng Different value: 20](#_Toc158580774)

[4.3 What is the interaction effect between the dentist and method on the hardness of dental implants for each type of alloy? 20](#_Toc158580775)

[Hypothesis 20](#_Toc158580776)

[Conclusion: 21](#_Toc158580777)

[Observation for Alloy 1: 21](#_Toc158580778)

[Hypothesis 21](#_Toc158580779)

[Conclusion: 21](#_Toc158580780)

[Observations for alloy 2: 22](#_Toc158580781)

[4.4 How does the hardness of implants vary depending on dentists and methods together? 22](#_Toc158580782)

Data & Dictionary

## Problem 3 Data Dictionary

1. Unpolished: Brinell's hardness index of unpolished stones (Numerical)
2. Treated and polished: Brinell's hardness index of treated and polished stones (Numerical)

## Problem 4 Data Dictionary

1. Dentist: Categorizes different dentist numerically
2. Method: Categorizes different method numerically
3. Alloy: Categorizes different alloys numerically
4. Temprature: Temprature of treatment of implant in Kelvin as numerical value
5. Response: Overall hardness of the implant as numerical value

List of Figures

[Figure 1. Breaking stregth < 3.17 7](#_Toc158580668)

[Figure 2. Breaking stregth atleast 3.6 8](#_Toc158580669)

[Figure 3. Breaking stregth between 5 & 5.5 9](#_Toc158580670)

[Figure 4: Breaking stregth not between 3 & 7.5 9](#_Toc158580671)

[Figure 5: Zingaro data info 10](#_Toc158580672)

[Figure 6: Boxplot Unpolished & Treated and Unpolished 11](#_Toc158580673)

[Figure 7: Boxplot after outlier treatement 11](#_Toc158580674)

[Figure 8: Unpolished and Treated Description (No outliers) 12](#_Toc158580675)

[Figure 9: Dental implat data Description 15](#_Toc158580676)

[Figure 10: Dental implat data info 16](#_Toc158580677)

[Figure 11: Boxplot: Dentist vs Response 16](#_Toc158580678)

[Figure 12Multiple comparision for different Method 19](#_Toc158580679)

[Figure 13: Interaction for Alloy 1: Dentist & method 20](#_Toc158580680)

[Figure 14: Interaction for Alloy 2: Dentist & Method 21](#_Toc158580681)

[Figure 15: Combined interaction: Dentist & Method 22](#_Toc158580682)

**PROBLEM 1**

A physiotherapist with a male football team is interested in studying the relationship between foot injuries and the positions at which the players play from the data collected.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Striker | Forward | Attacking Midfielder | Winger | **Total** |
| Players Injured | 45 | 56 | 24 | 20 | **145** |
| Players Not Injured | 32 | 38 | 11 | 9 | **90** |
| **Total** | **77** | **94** | **35** | **29** | **235** |

Based on the above data, answer the following questions.

## What is the probability that a randomly chosen player would suffer an injury?

Probability that a randomly chosen player will suffer injury can be calculated as follows,

P=No. of injured players/No. of total players

**P=145/235 = 0.61**

**Thus, we can say that probability that a randomly chosen player would suffer injury is 61%.**

## What is the probability that a player is a forward or a winger?

Probability that a player is forward or a winger can be calculated as follows,

P=(No. of Forward +No. of winger)/Total Players

P=(94+29)/235

**P=.523**

**Thus, we can say that probability that a player is forward or a winger is 52.3%.**

## What is the probability that a randomly chosen player plays in a striker position and has a foot injury?

**P = P(A & B) =P(A)\*P(B)**

In this case,

P(A)= Probability that chosen player is striker= 77/235= 0.3276

P(B)= Probability that a chosen player has foot injury= 45/77=.5844

**P= 0.3276 \* 0.5844= 0.1914**

**Alternate:**

P(A&B)= No. of strikers who are injured/No. of total players

**P = 45/235 =0.1914**

**Therefore, we can say that probability that a randomly chosen player plays in striker position and has a foot injury is 19.14%.**

## What is the probability that a randomly chosen injured player is a striker?

P = P(A & B) =P(A)\*P(B)

In this case, P= No. of injured players who are striker/Total players injured

**P= 45/77 =.5844**

**Therefore, we conclude that probability that a randomly chosen injured player is a striker is 58.44%.**

**PROBLEM 2**

The breaking strength of gunny bags used for packaging cement is normally distributed with a mean of 5 kg per sq. centimetre and a standard deviation of 1.5 kg per sq. centimetre. The quality team of the cement company wants to know the following about the packaging material to better understand wastage or pilferage within the supply chain; Answer the questions below based on the given information; **(Provide an appropriate visual representation of your answers, without which marks will be deducted)**

**Info Provided:**

* **Breaking strength for cement is normally distributed**
* **Mean = 5 kg/sqcm.**
* **Sigma= 1.5 kg/sqcm**
* In this question sample size in not mentioned.
* Population mean is provided. Therefore, we will use Z-test for our analysis.

## 2.1 What proportion of the gunny bags have a breaking strength of less than 3.17 kg per sq cm?

The proportion of gunny bags having breaking strength less than 3.17 kg/sq cm is **11.12%**. As per the below figure blue shaded region represents the proportion of gunny bags with breaking strength less than 3.17 kg/sqcm.

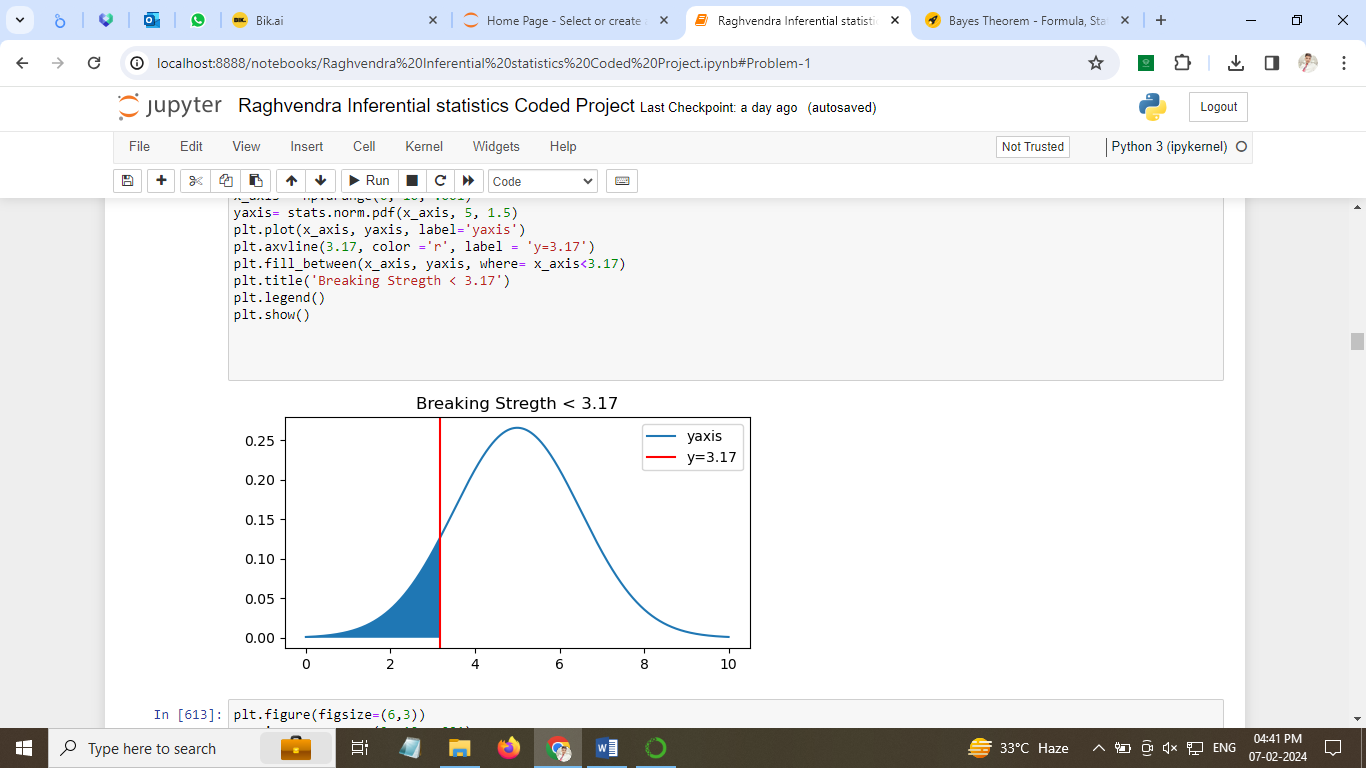
**N**

Figure 1. Breaking stregth < 3.17

## 2.2 What proportion of the gunny bags have a breaking strength of at least 3.6 kg per sq cm.?

The proportion of gunny bags having breaking strength at least 3.6kg/sq cm is **82.46%**. As per the below figure blue shaded region represents the proportion of gunny bags with breaking strength of at least 3.6 kg/sqcm.

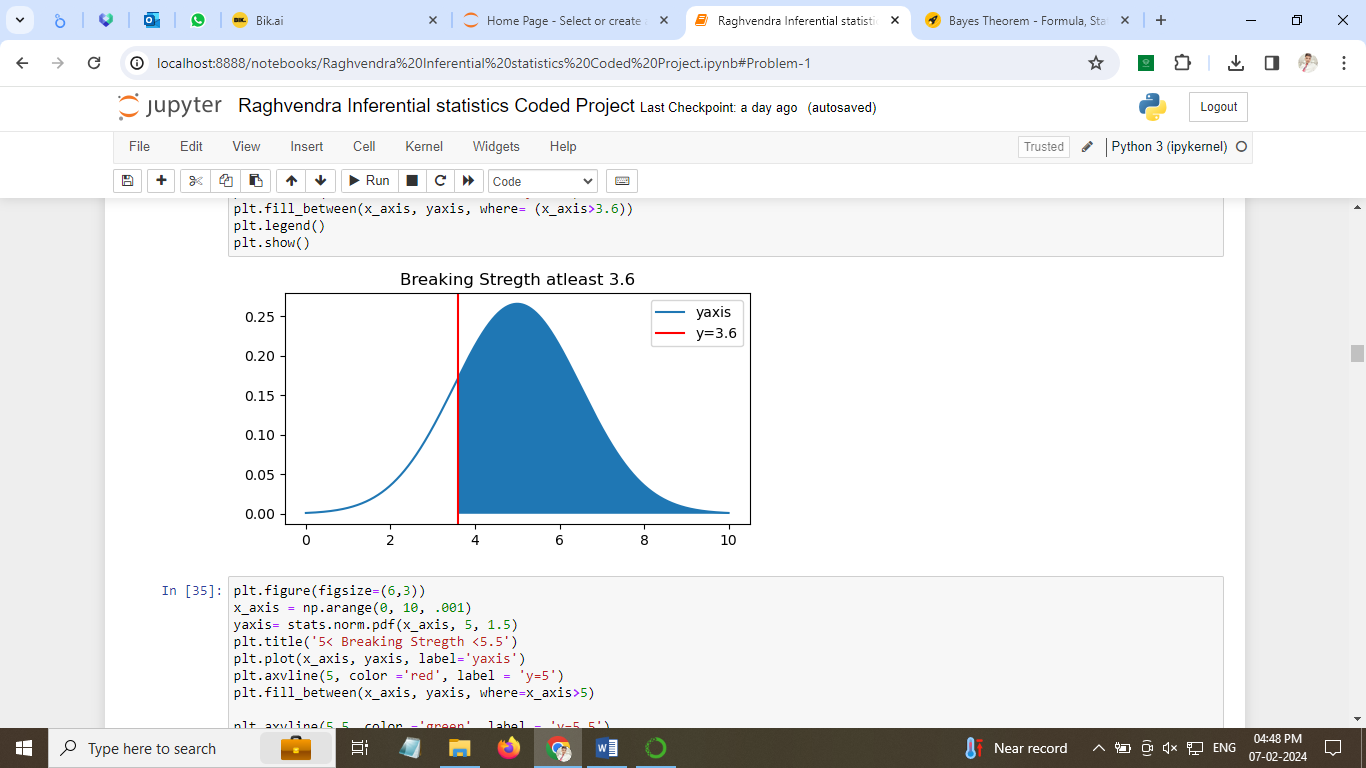


Figure 2. Breaking stregth atleast 3.6

## 2.3 What proportion of the gunny bags have a breaking strength between 5 and 5.5 kg per sq cm.?

The proportion of gunny bags having breaking strength between 5 and 5.5 kg/sq cm is **13.05%**. As per the below figure blue shaded region represents the proportion of gunny bags with breaking strength between 5 and 5.5 kg/sqcm.

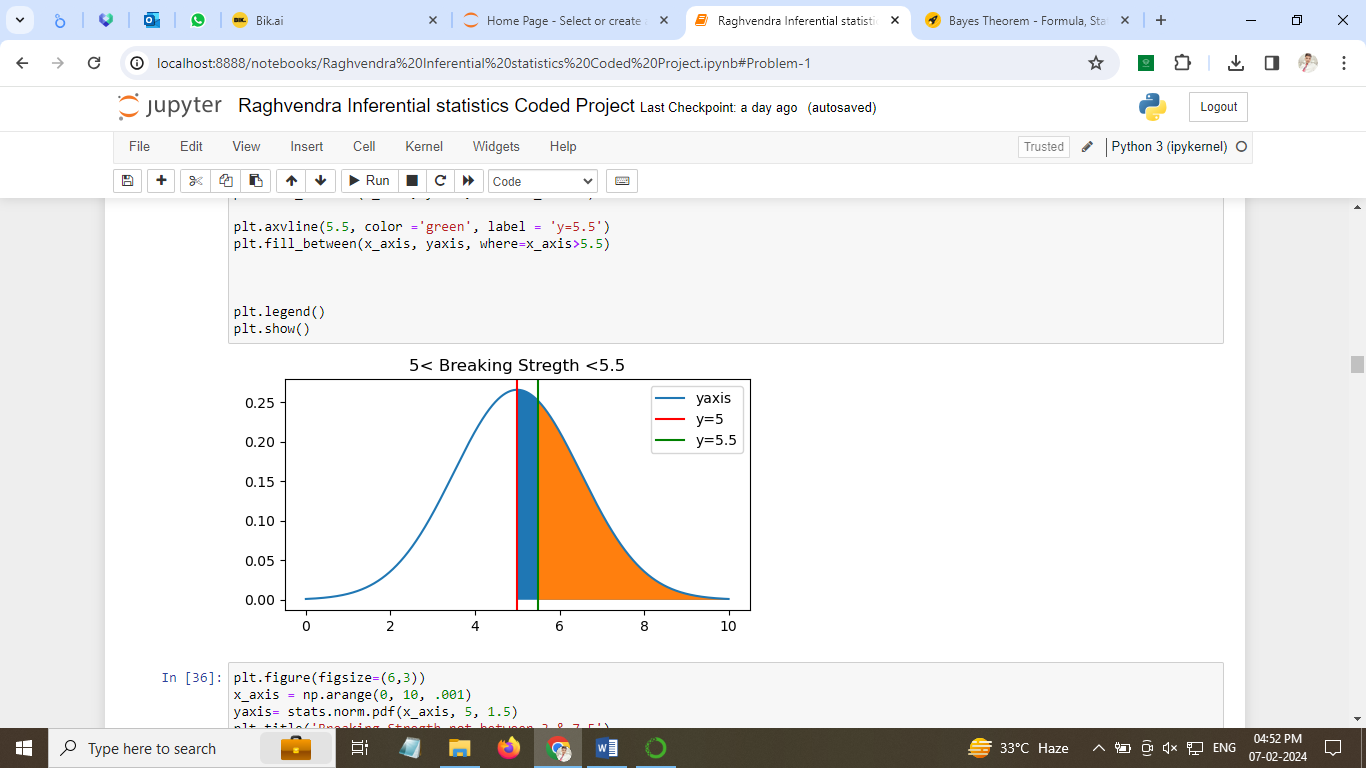


Figure 3. Breaking stregth between 5 & 5.5

## 2.4 What proportion of the gunny bags have a breaking strength NOT between 3 and 7.5 kg per sq cm.?

The proportion of gunny bags having breaking strength between 5 and 5.5 kg/sq cm is **13.90%**. As per the below figure blue & orange shaded region represents the proportion of gunny bags with breaking strength between 5 and 5.5 kg/sqcm.

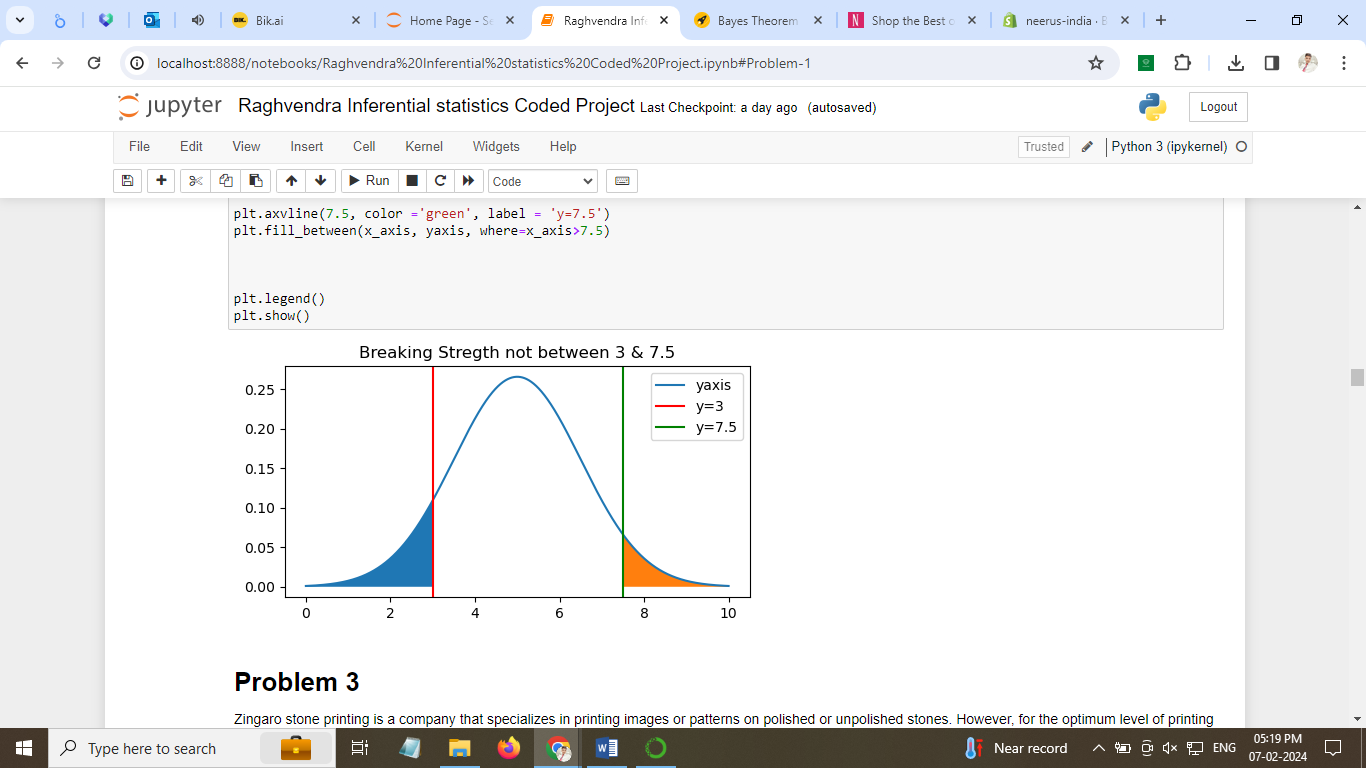


Figure 4: Breaking stregth not between 3 & 7.5

**PROBLEM 3**

Zingaro stone printing is a company that specializes in printing images or patterns on polished or unpolished stones. However, for the optimum level of printing of the image, the stone surface has to have a Brinell's hardness index of at least 150. Recently, Zingaro has received a batch of polished and unpolished stones from its clients. Use the data provided to answer the following (assuming a 5% significance level).

We have checked the data info as follows and found that there are 75 rows and 2 columns

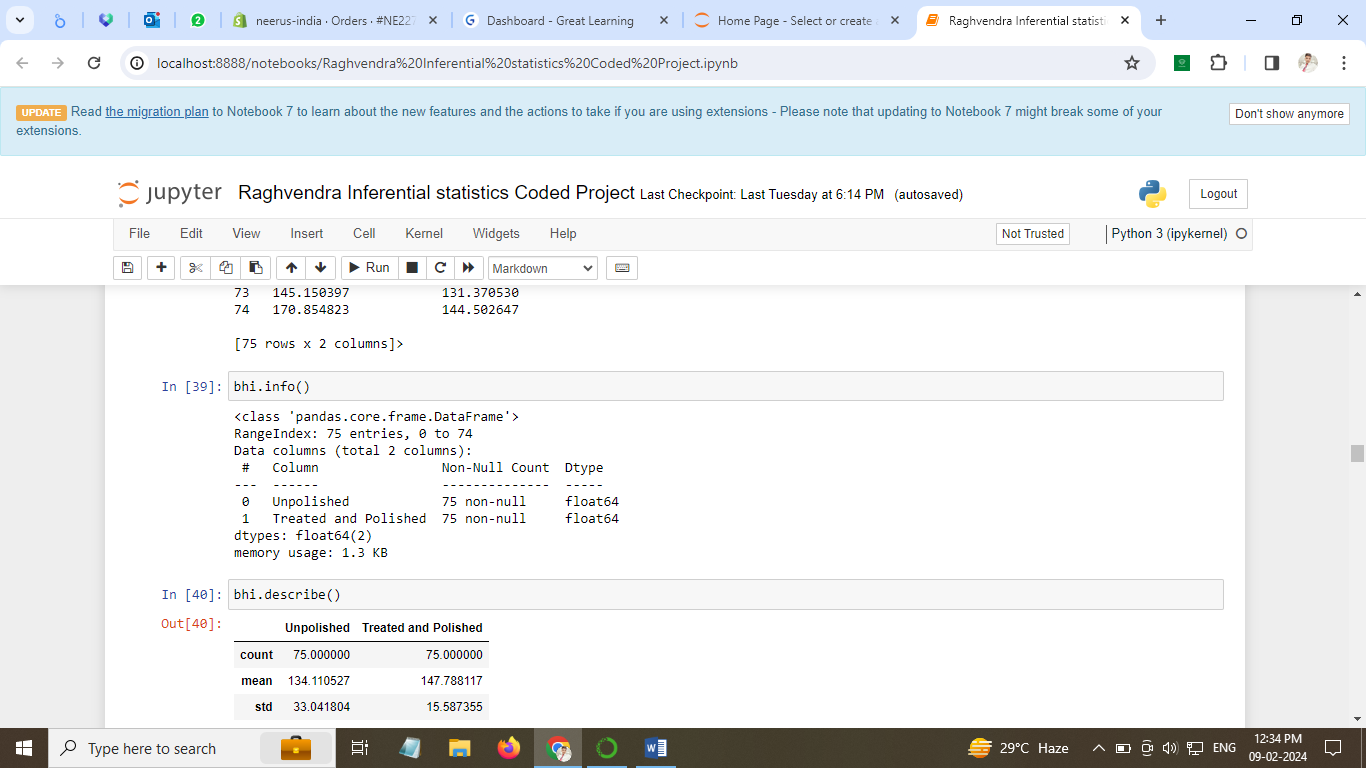
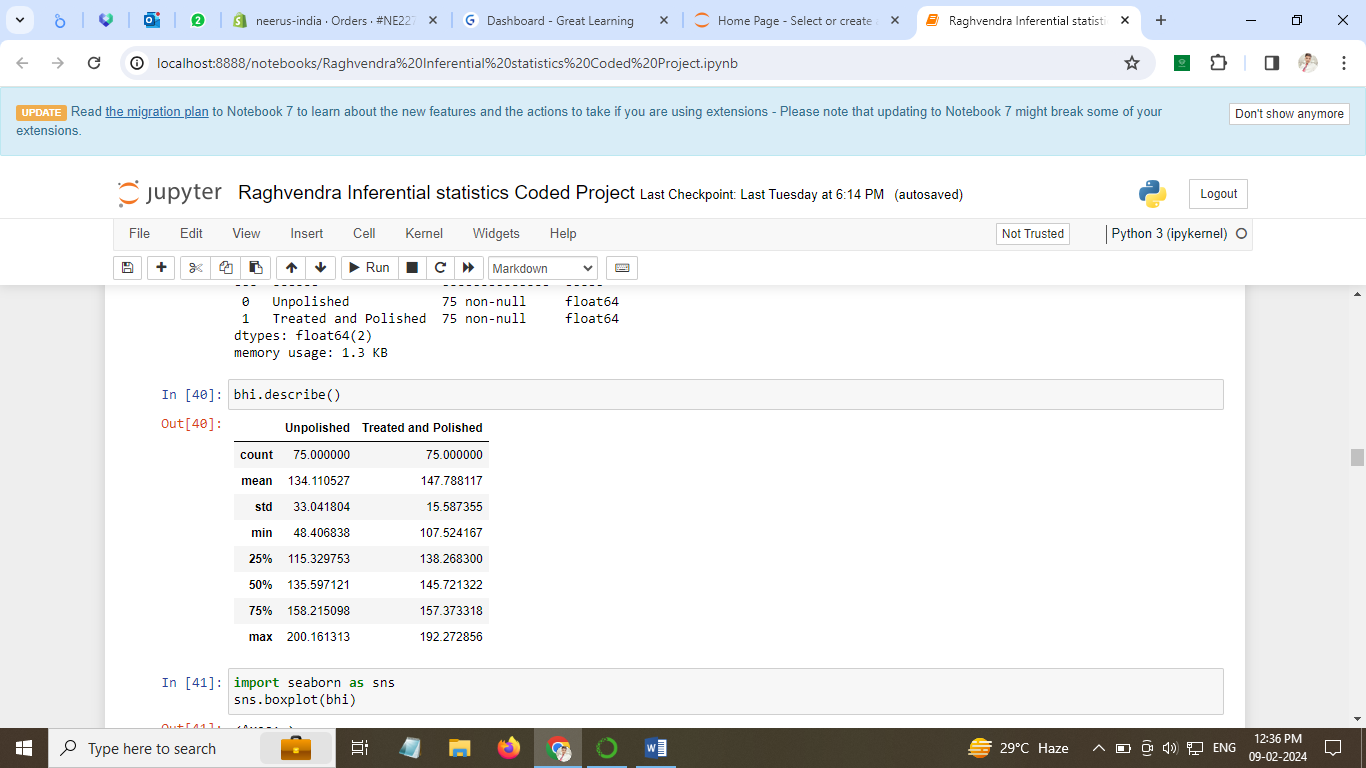


Figure 5: Zingaro data info

We have checked the data description and found below output



We check the visual representation for any outliers present by creating a boxplot.

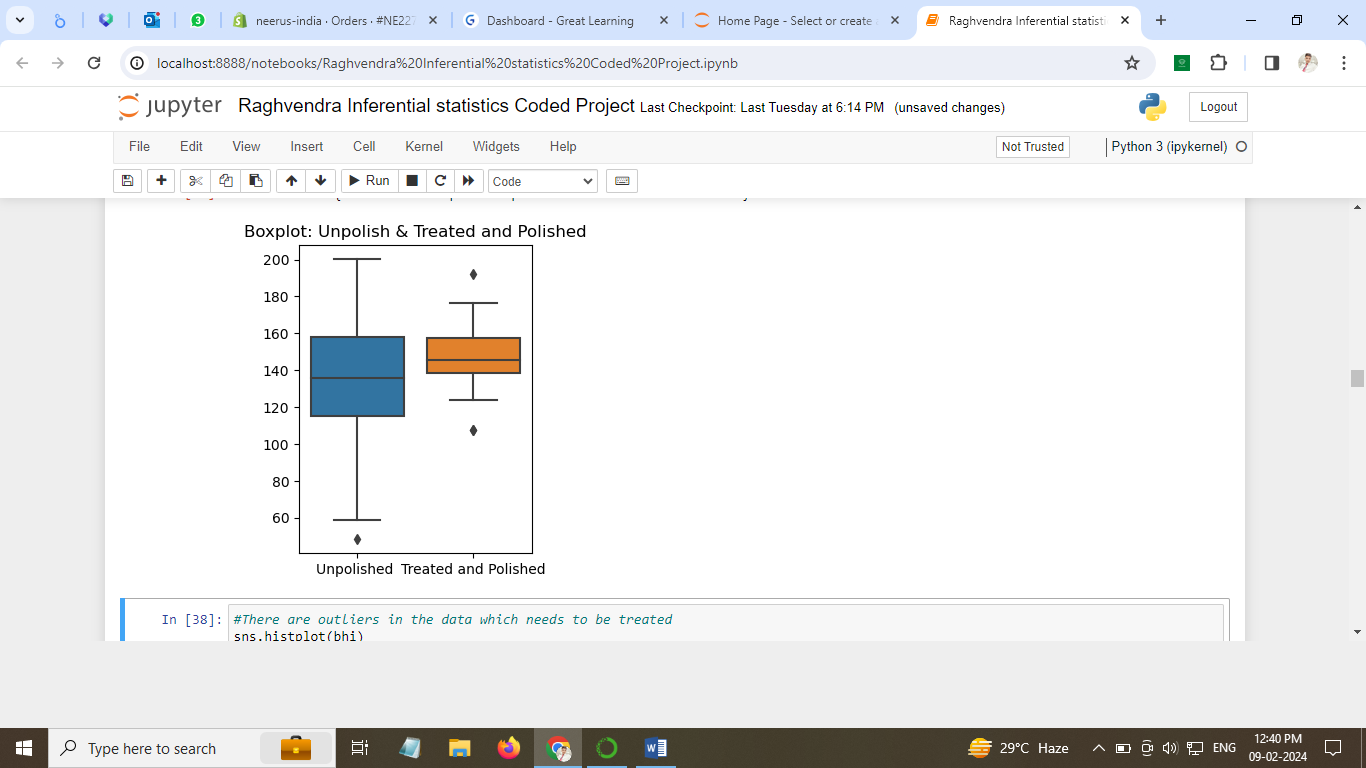


Figure 6: Boxplot Unpolished & Treated and Unpolished

After outlier treatment we are ready for our analysis.

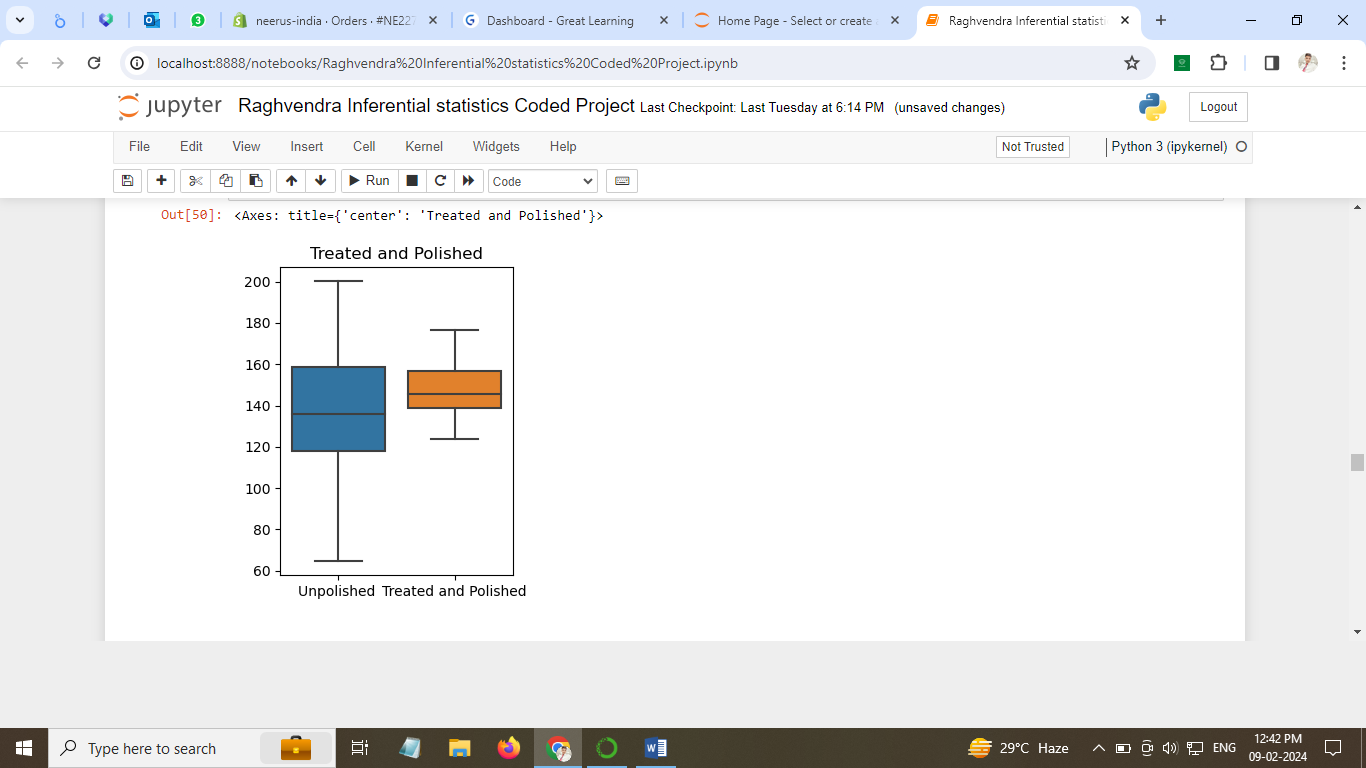


Figure 7: Boxplot after outlier treatement

* Range of the unpolished stones is large compared to Treated and polished stones.
* Unpolished stones lie in range of 60 to 200
* Treated and polished stones varies between 120 to 180

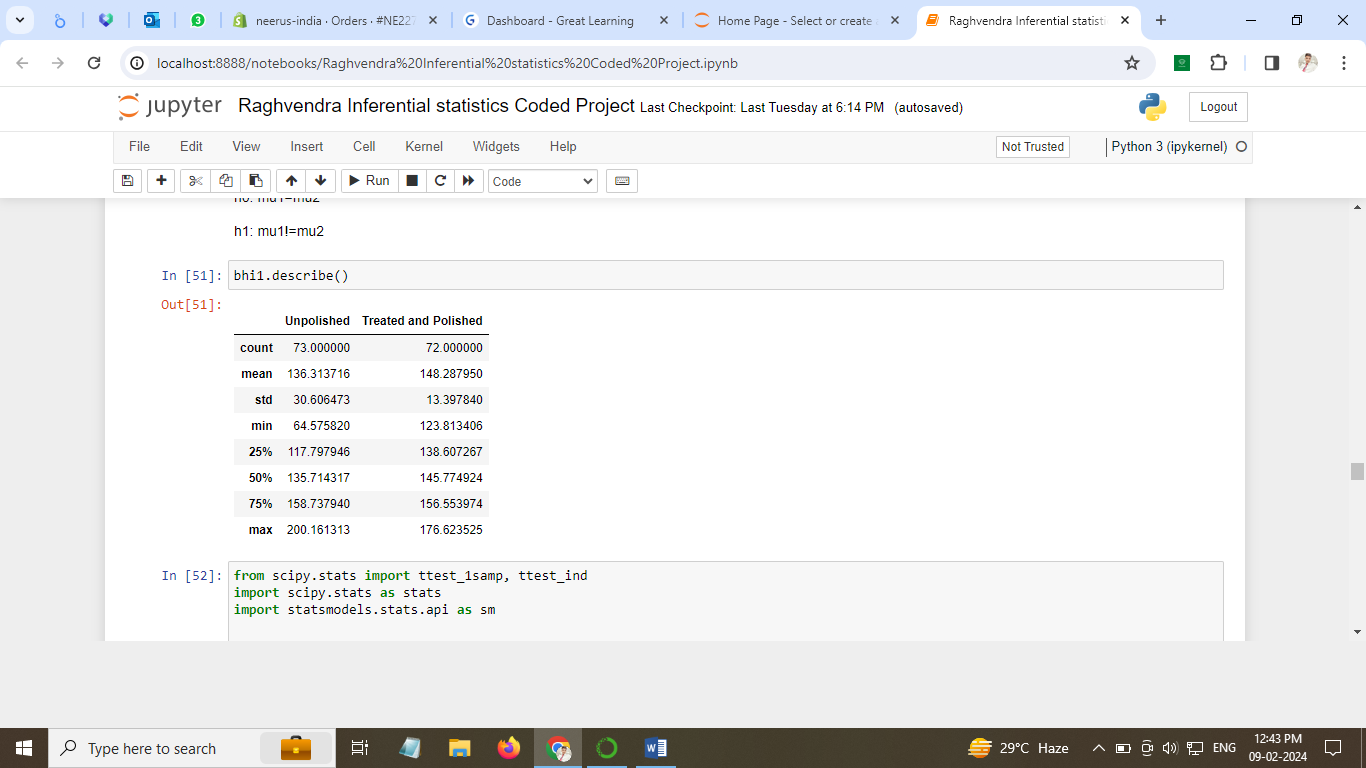


Figure 8: Unpolished and Treated Description (No outliers)

### Inferences:

We have below information after outlier treatment:

* Mean for Unpolished is 136.31
* Mean for Treated and Unpolished is 148.28
* Standard Deviation for Unpolished is 30.6
* Standard Deviation for Treated and Polished is 13.39
* Sample size is greater than 30

## 3.1 Zingaro has reason to believe that the unpolished stones may not be suitable for printing. Do you think Zingaro is justified in thinking so?

### Test Selection:

Population mean is given and sample size is greater than 30. Therefore we will use Z test for testing hypothesis.

### Level of significance:

The value of Alpha (level of significance) is predetermined at .05

### Hypothesis:

Given: Minimum hardness required for printing is 150

#### **H0 (null hypothesis): Mu(Unpolished) >=150**

#### **Ha (alternate hypothesis): Mu(Unpolished) < 150**

* Null hypothesis states that mean hardness of Unpolished stone surface is less than or equal to 150
* Alternate hypothesis states that mean hardness of Unpolished stone surface is greater than 150
* This is a case of one tailed Ztest.

### Conducting Ztest and calculating P value:

Z score is calculated using below formula:

Z= Xbar-Mu(population)/(standard deviation/square root of sample size)

* After conductiong Z test we find z value to be **-3.8206**
* P-value is **0.0001**
* P-Value is less than level of significance **P<Alpha(.05)**
* Therefore, we have enough evidence to reject the null Hypothesis and conclude that mean

Hardness is less than 150 for unpolished stones. We accept the alternate hypothesis.

### Decision making:

For left tail hypothesis test, If P < .05 : Reject H0 for left tail test

For right tail hypothesis test, If P > .05: Reject H0 for right tail test

Conclusion:

We have enough evidence to reject null hypothesis and conclude that mean hardness for of unpolished stones is less than 150. Zingaro is justified in believing that unpolished stones may not be fit for printing.

## 3.2 Is the mean hardness of the polished and unpolished stones the same?

### Test Selection:

Population mean is given and sample size is greater than 30. Therefore we will use Z test for testing hypothesis.

### Level of significance:

The value of Alpha (level of significance) is predetermined at .05

### Hypothesis:

#### **H0 (null hypothesis): Mu (Unpolished) = Mu (treated & unpolished)**

#### **Ha (alternate hypothesis): Mu (Unpolished) != Mu (treated & unpolished)**

* Null hypothesis states that mean hardness of Unpolished stone is equal to mean hardness of Treated and Unpolished stone.
* Alternate hypothesis states that mean hardness of Unpolished stone is not equal to mean hardness of Treated and Unpolished stone.
* We will conduct a independent **two sample T test** since we need to compare the mean of two different dataset.

### Conducting Independent 2 sample T-test and calculating P value:

We conduct the test by importing ttest\_1samp, ttest\_ind libraries to find the values of t\_statistic, p\_value.

* After conducting T test we find **Tstat = -3.0444**
* **P Value = 0.0028**
* P-Value is less than level of significance **P<Alpha(.05)**
* Therefore, we have enough evidence to reject the null Hypothesis and accept we Alternate

Hypotheis.

Conclusion:

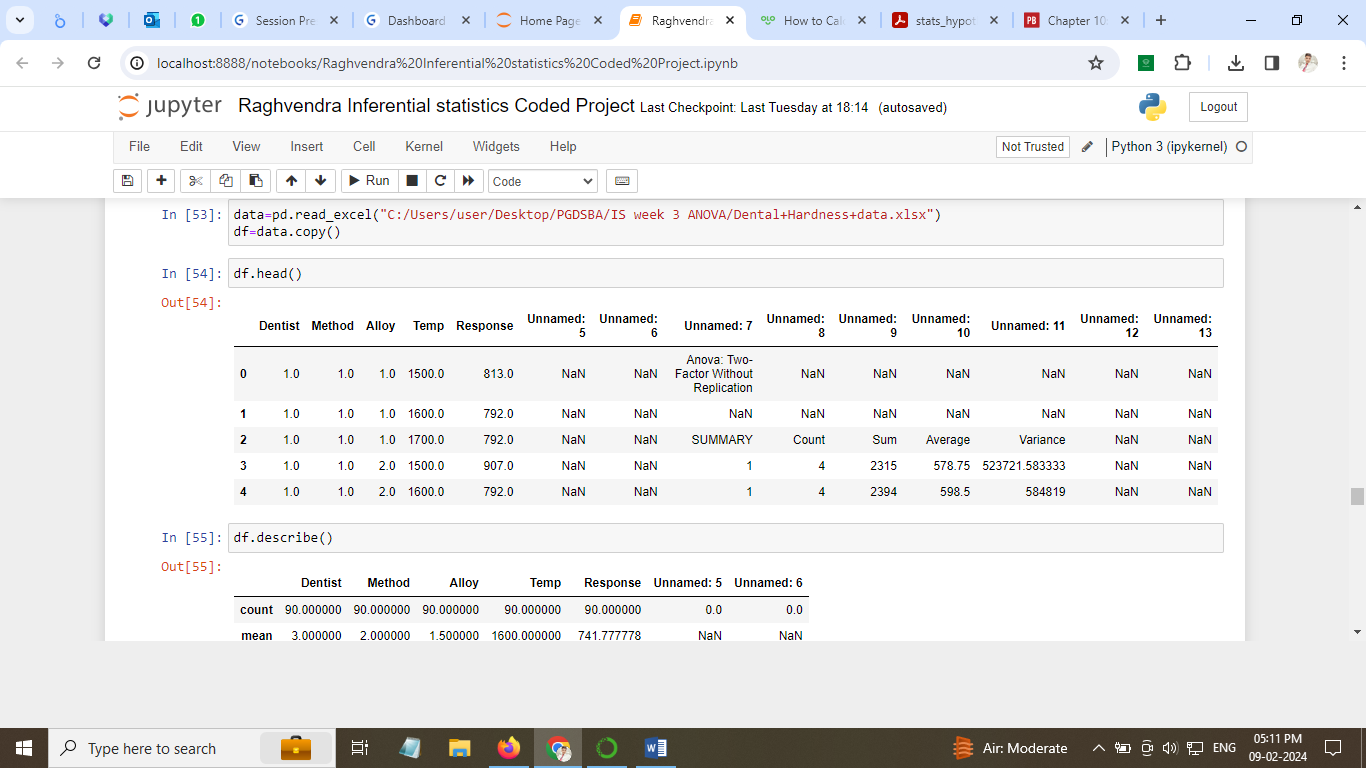
We conclude that the mean hardness of Unpolished stones is not equal to mean hardness of Treated and Polished stone.

**Problem 4**

Dental implant data: The hardness of metal implants in dental cavities depends on multiple factors, such as the method of implant, the temperature at which the metal is treated, the alloy used as well as the dentists who may favor one method above another and may work better in his/her favorite method. The response is the variable of interest.

Data Preparation:

Head:



We see that first 5 rows are only significant for our analysis. Therefore we will consider them only.

Describe:

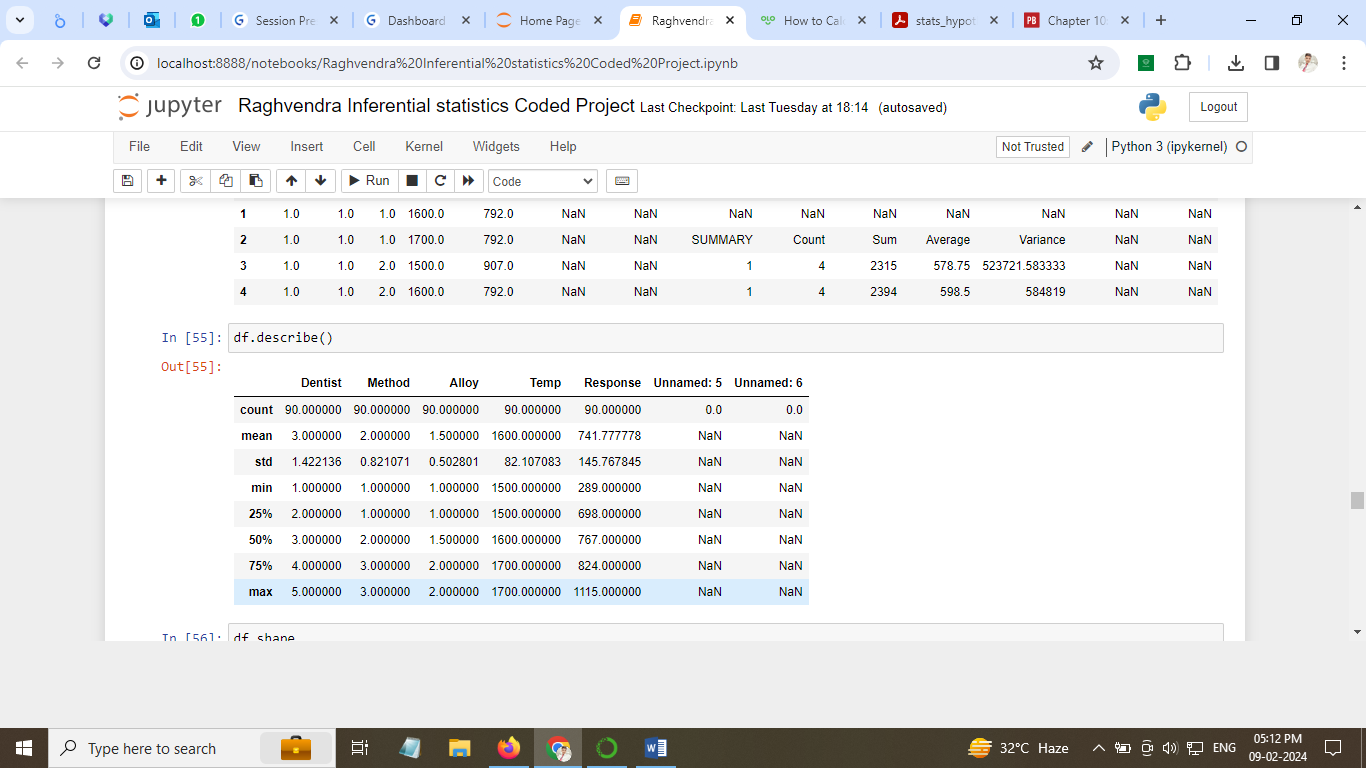


Figure 9: Dental implat data Description

* There are total 90 counts of Dentist, Method, Alloy, Temperature and Response
* Maximum harness is 1115 and minimum hardness is 289. Mean hardness is 289
* There are 5 Dentists, 3 Methods and 2 Alloys
* Maximum temperature is 1700 k and minimum temperature is 1500 K

Info:

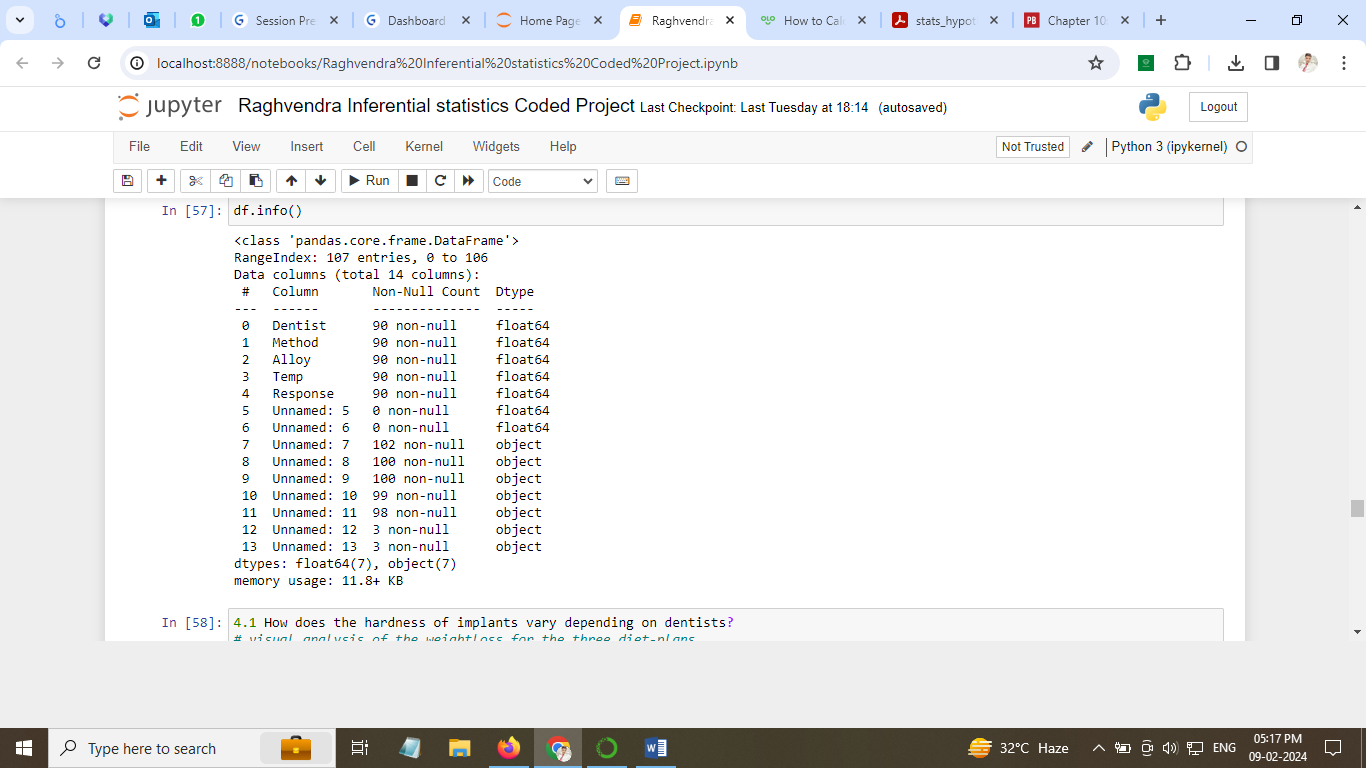


Figure 10: Dental implat data info

* Data has 14 columns
* Dentist, Method, Alloy, Temp and Response are to be considered for our analysis. They have 90 non null values.
* There are 7 float and 7 objects

## 4.1 How does the hardness of implants vary depending on dentists?

Lets visualize the data available for different doctors and response (hardness of implants)

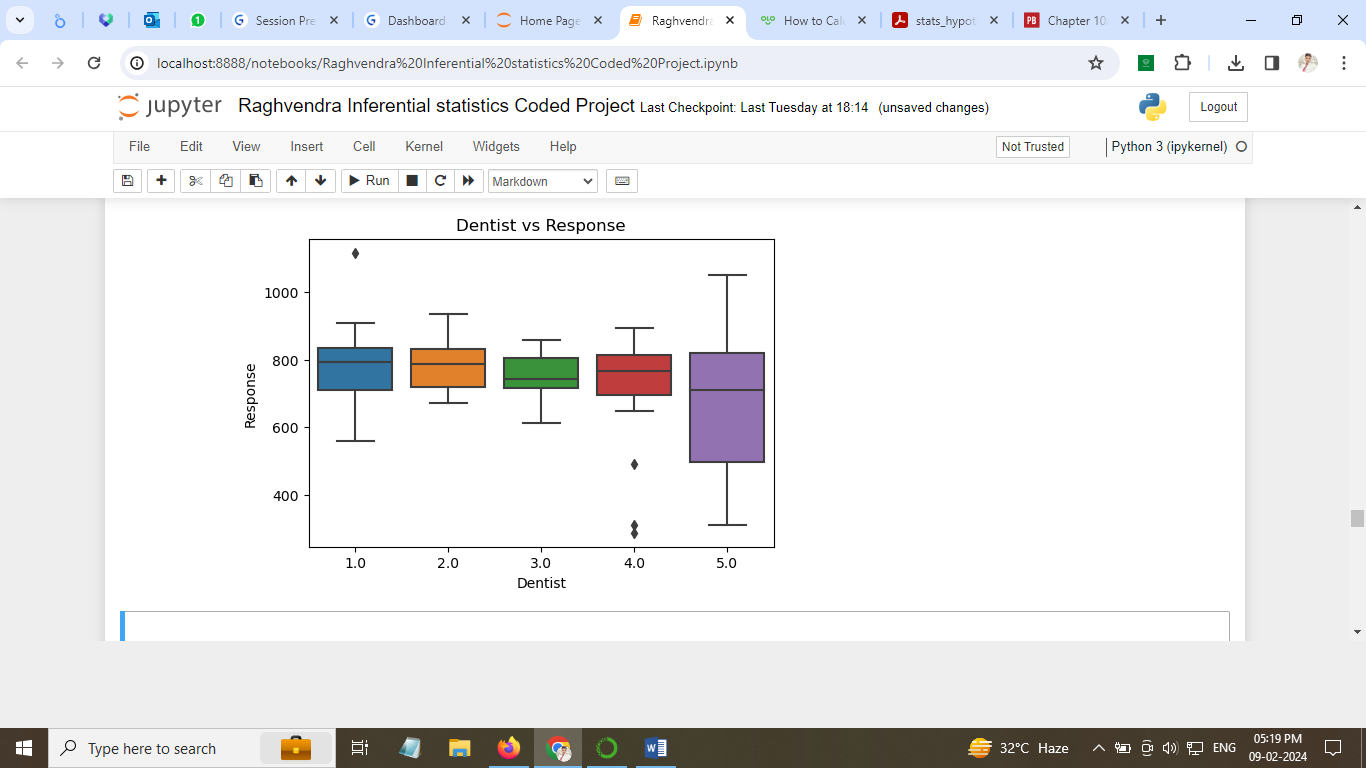


Figure 11: Boxplot: Dentist vs Response

### Infernece:

* The mean hardness for all doctors lies between 600 to 800.
* Denstist 5 has the maximum range of hardness among all doctors.
* Minimum hardness of implants is above 600 for all except dentist 5
* There are outliers present in dentist 1 and dentist 4 data
* Dentist 3 has the smallest range of hardnes of implants among all doctors.

### HYPOTHESIS:

The null and alternative hypotheses can be formulated as:

**H0 : The mean hardness is same for all dentists.**

mu1 = mu2 = mu3 = mu4 = mu5

**Ha : At least one of the mean hardness is different for the dentist.**

mu1 = mu2 = mu3 = mu4 != mu5

### TEST SELECTION:

The test involves comparing the mean of 3 or more groups belonging to each dentist. We will use ONE-WAY ANOVA test to compare the mean of each group.

However, we need to check below assumptions:

1. All population should have normal distribution. (Shapiro-Wilk’s test)
2. All population should have equal variances (Levene test)
3. Observations should be independent of each other

### Shapiro-Wilk’s Test

We will test for Normality of the population distribution,

**H0: The hardness follow a normal distribution**

**Ha: The hardness do not not follow a normal distribution**

Level of significance (alpha) =.05

After conducting Shapiro test on Response we find,

**P value = 1**

**P > Alpha**

Therefore, we fail to reject the Null hypothesis and conclude that response follows Normal distribution.

### Levene Test

We will test equality of variance using Levene test

**H0: All population variances are equal**

**Ha: Atleast one variance is different from others**

Level of significance (alpha) =.05

After conducting Levene test on Response we find,

**P value= 0.0079**

Therefore, we reject the Null hypothesis for homogeneity of variances. Conclude that atleast one population variances is different from other.

### ONE WAY ANOVA TEST

Let us apply One way Anova test to check the below Hypothesis:

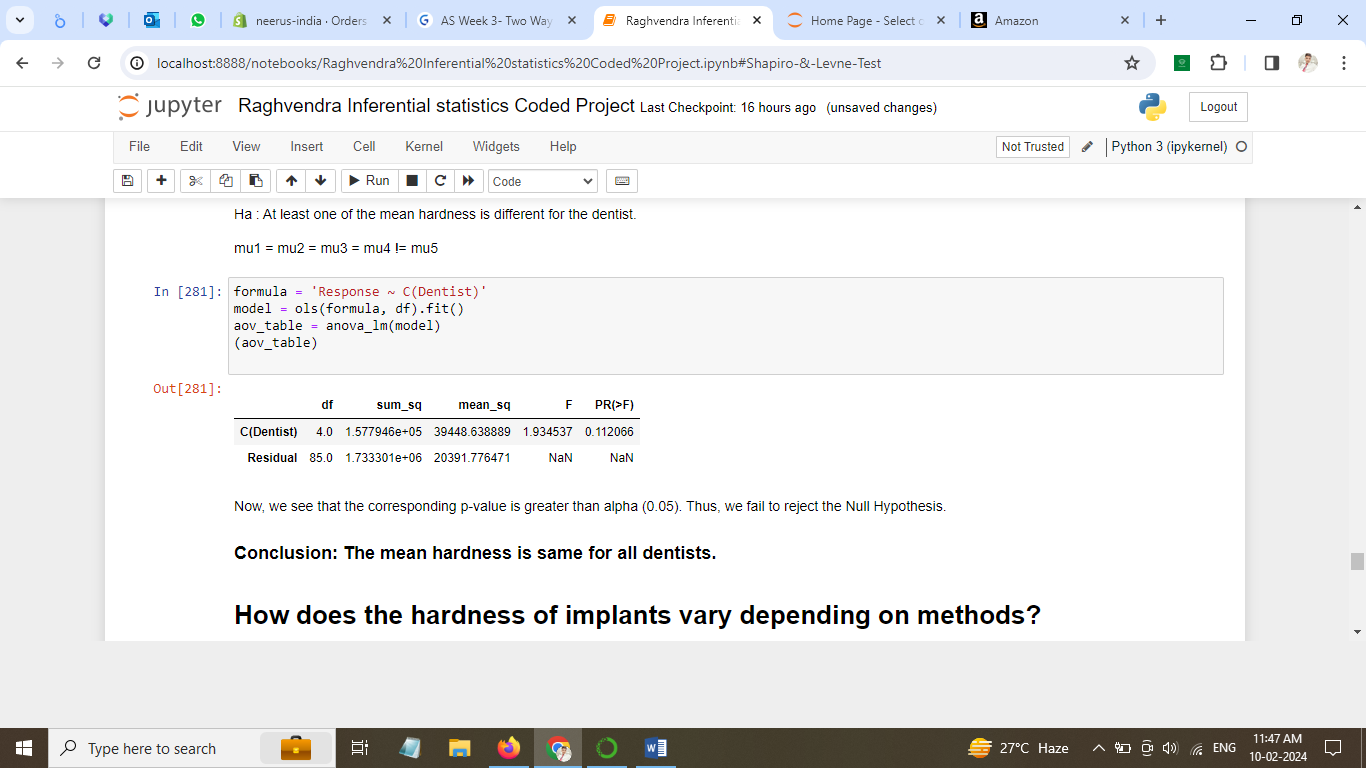
**H0 : The mean hardness is same for all dentists.**

mu1 = mu2 = mu3 = mu4 = mu5

**Ha : At least one of the mean hardness is different for the dentist.**

mu1 = mu2 = mu3 = mu4 != mu5

After conducting One Way ANOVA test we find that,



**P value = 0.112066**

**P> .05**

### Conclusion

**We fail to reject the Null Hypothesis and conclude that mean hardness is same for all dentists.**

## 4.2 How does the hardness of implants vary depending on methods?

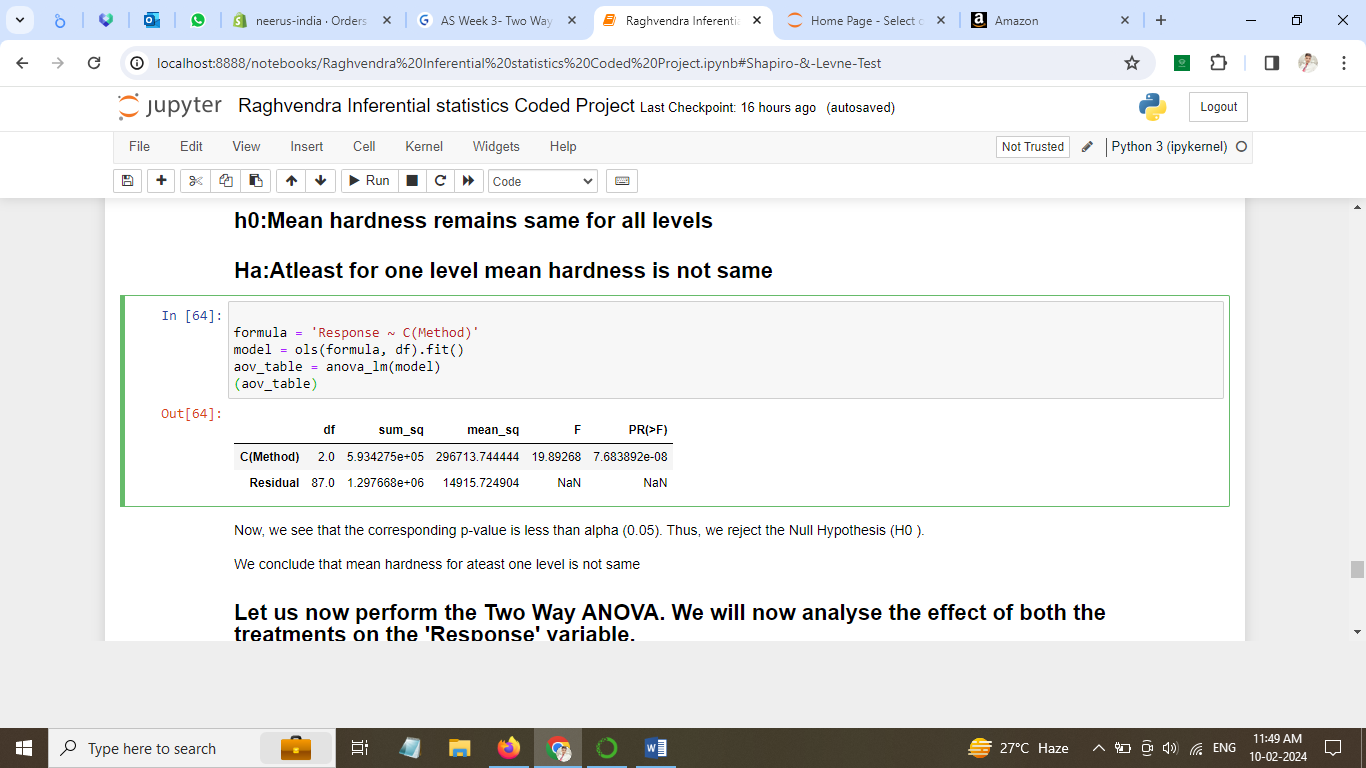
We need to perform One way anova and test the hypothesis. The hypothesis can be framed as follows.

### Hypothesis

**H0: Mean hardness remains same for all methods**

**Ha: At least one mean hardness is different for a method.**

After conducting one way ANOVA test we find,



**P value < 0.05**

The P-Value is less than level of significance. We have enough evidence to reject the null Hypothesis.

### Conclusion:

**Mean hardness of dental implants for atleast one method is different from others.**

### Identifyng Different value:

We will also identify the different mean value for Method by doing Tuckey comparision.

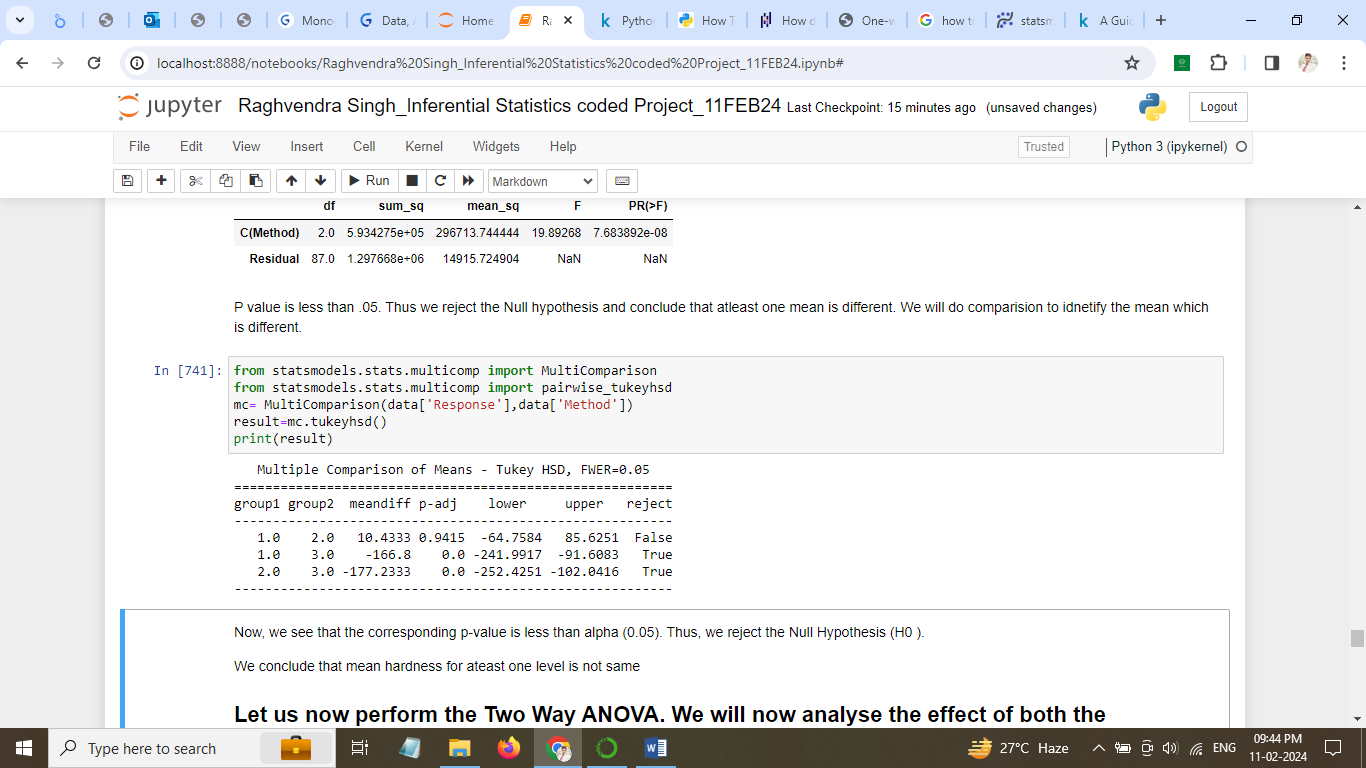
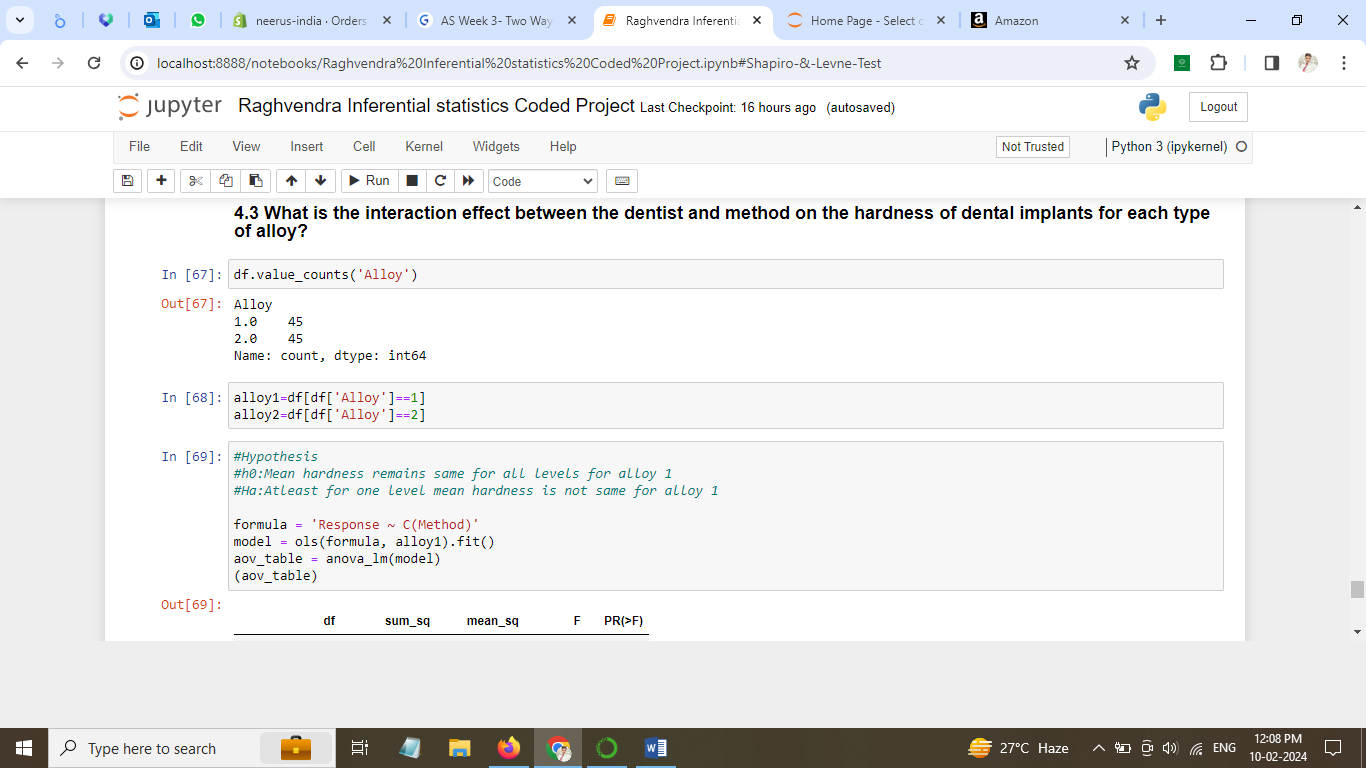


Figure 12Multiple comparision for different Method

* P value for for method 1 and method 2 is significantly different as P value is greater than .05
* We conclude that the mean difference for Method 1 and Method 2 is different.

## 4.3 What is the interaction effect between the dentist and method on the hardness of dental implants for each type of alloy?

We see that there are two type of alloy. We will create different subset of alloy and conduct ANOVA test for each alloy and study the interaction effect between dentist and method.

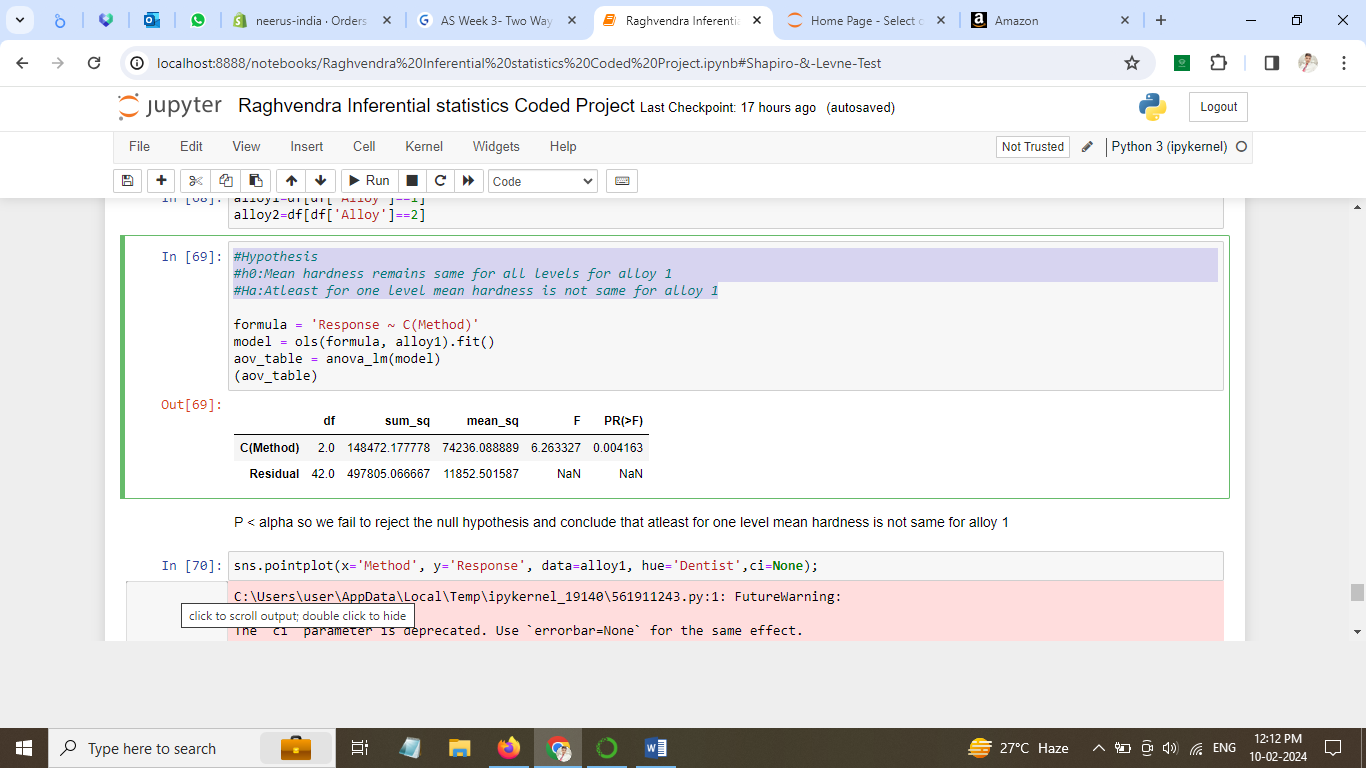


### Hypothesis

**H0: Mean hardness remains same for alloy 1**

**Ha: Atleast for one mean hardness is different for alloy 1**

#### One Way ANOVA TEST:



We can see that P Value is less than level of significance of .05. Therefore, we reject the null hypothesis.

### Conclusion:

**Atleast one Mean hardness is different for Alloy 1.**

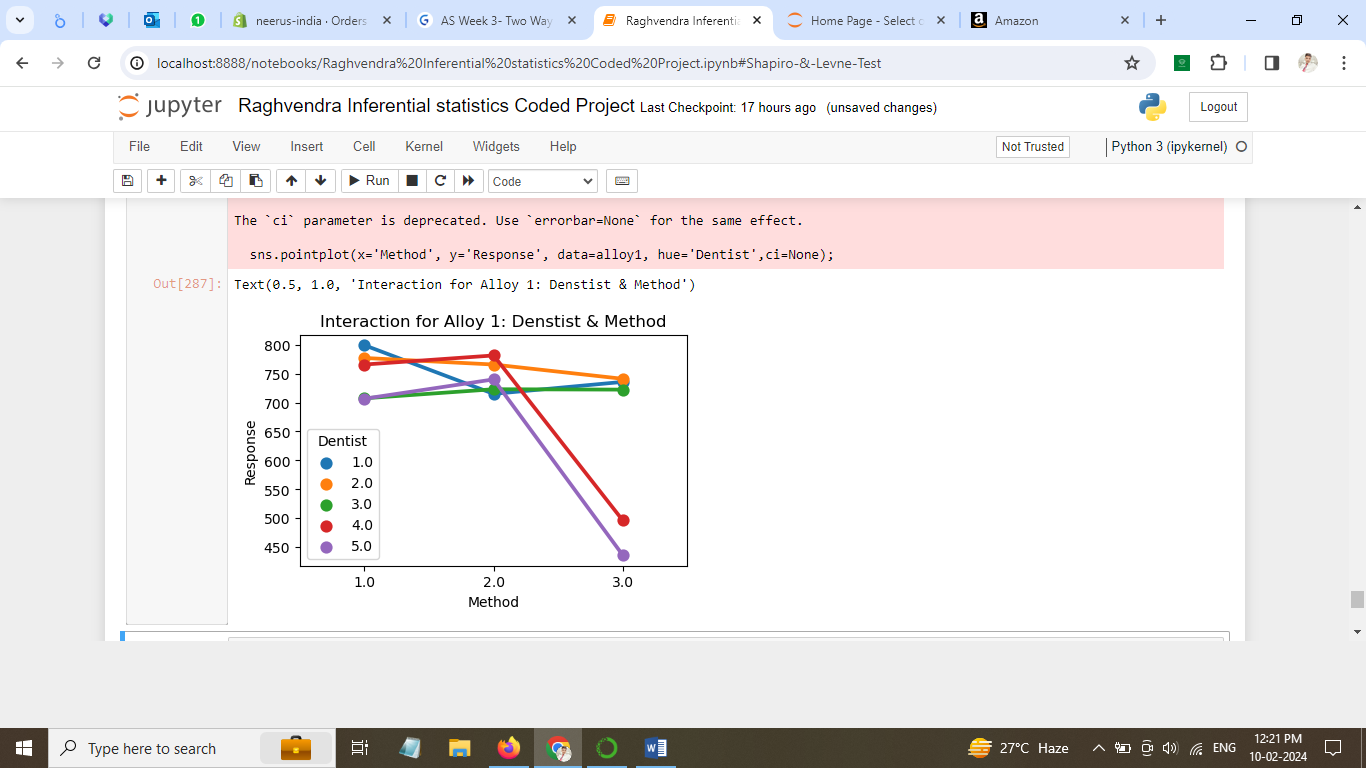


Figure 13: Interaction for Alloy 1: Dentist & method

### Observation for Alloy 1:

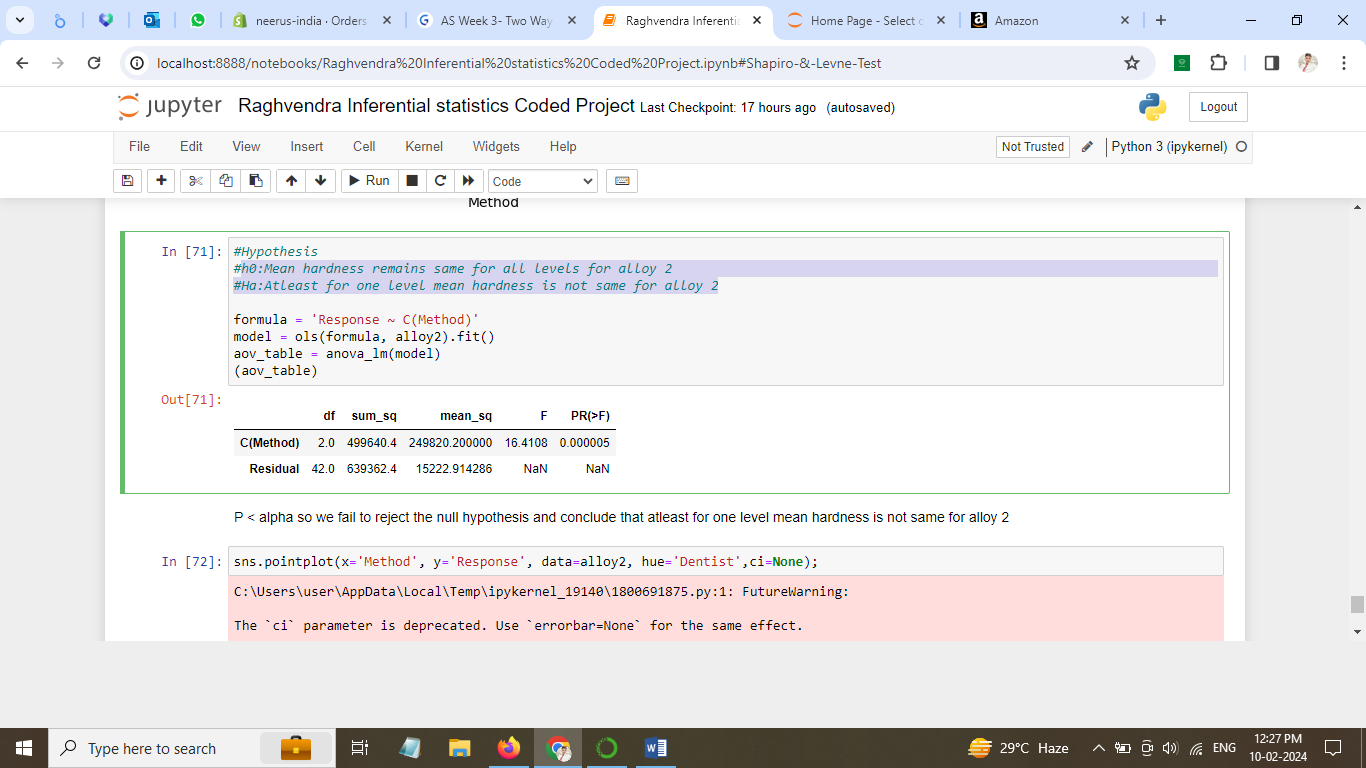
1. Method 2 gives minimum variation for harness for alloy 1
2. Method 3 gives very large variation for hardness for dentist 5 and dentist 4. They should not rely on method 3
3. Method 1 gives the highest harness lever for dentist 1, 2 and 3 and should be the preferred method for treatment
4. Dentist 2 can prefer any of the 3 methods as these is very less variation in hardness
5. Dentist 3 must use method 3 for best harness results

### Hypothesis

**H0:Mean hardness remains same for all levels for alloy 2**

**Ha:Atleast for one level mean hardness is not same for alloy 2**

**One Way ANOVA TEST:**



We can see that P Value is less than level of significance of .05. Therefore, we reject the null hypothesis.

### Conclusion:

**Atleast for one level mean hardness is not same for alloy 2**

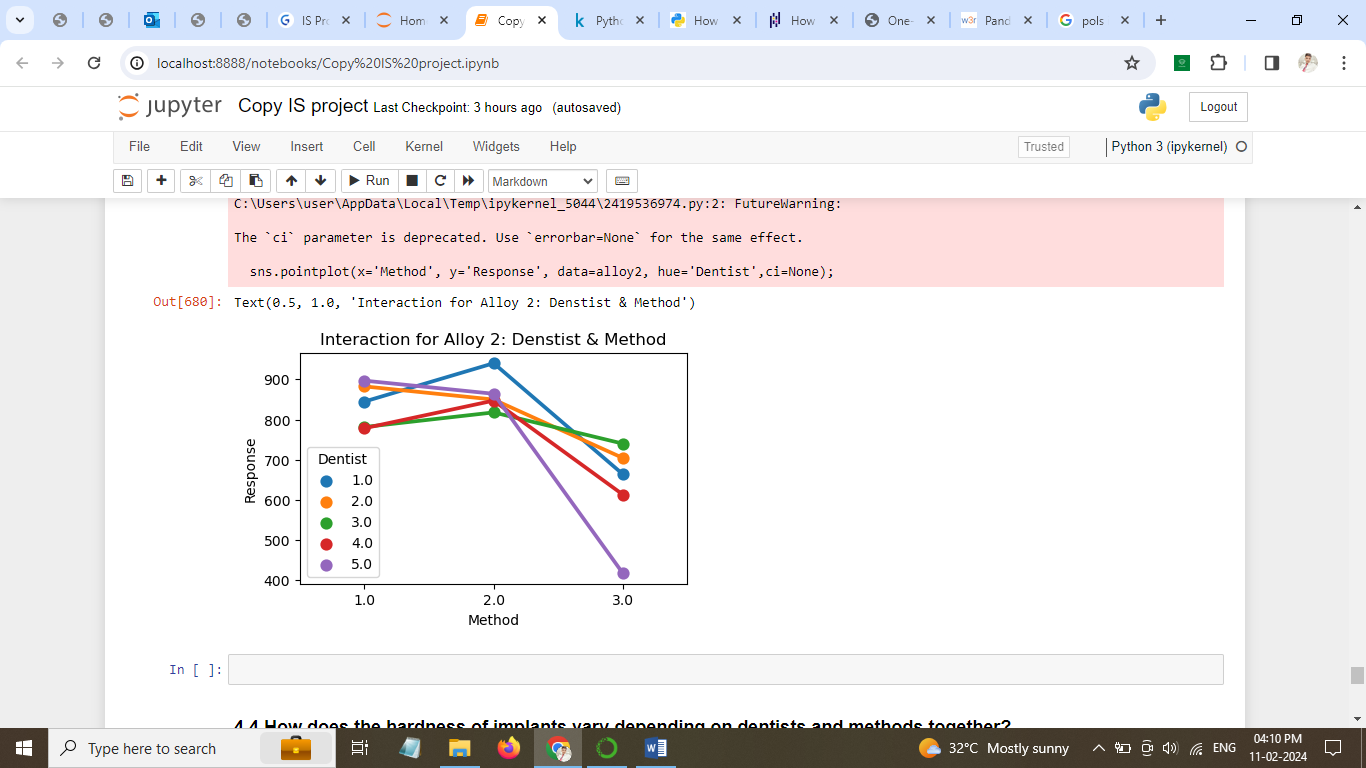


Figure 14: Interaction for Alloy 2: Dentist & Method

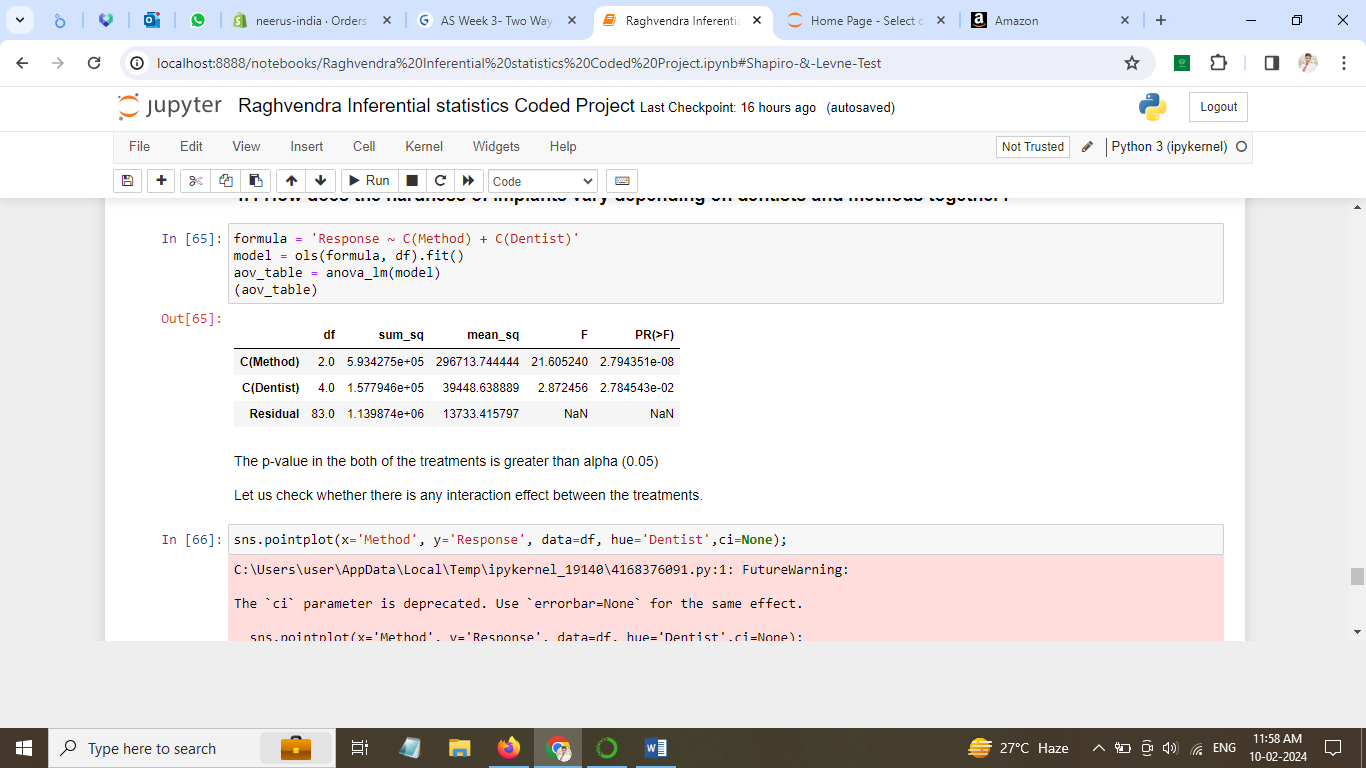
### Observations for alloy 2:

* Dentist 2, Dentist 3 and Dentist 4 have best hardness results with method 2 among other 2 methods.
* Dentist 5 has least harness response for Alloy 2 and Alloy 1 both.
* Method 3 gives large variation of harness among dentists
* Dentist 5 and Dentist 2 must prefer method 1 over other methods.
* Method 1 give similar outcomes for dentist 4 & 5

## 4.4 How does the hardness of implants vary depending on dentists and methods together?

We have two factors Dentist and Method therefore we will use TWO-WAY ANOVA for checking the interaction effect on Response.

After conducting Two way ANOVA test we found,



P value is < .05 for Dentist and Method both. We will check if there is any interaction between

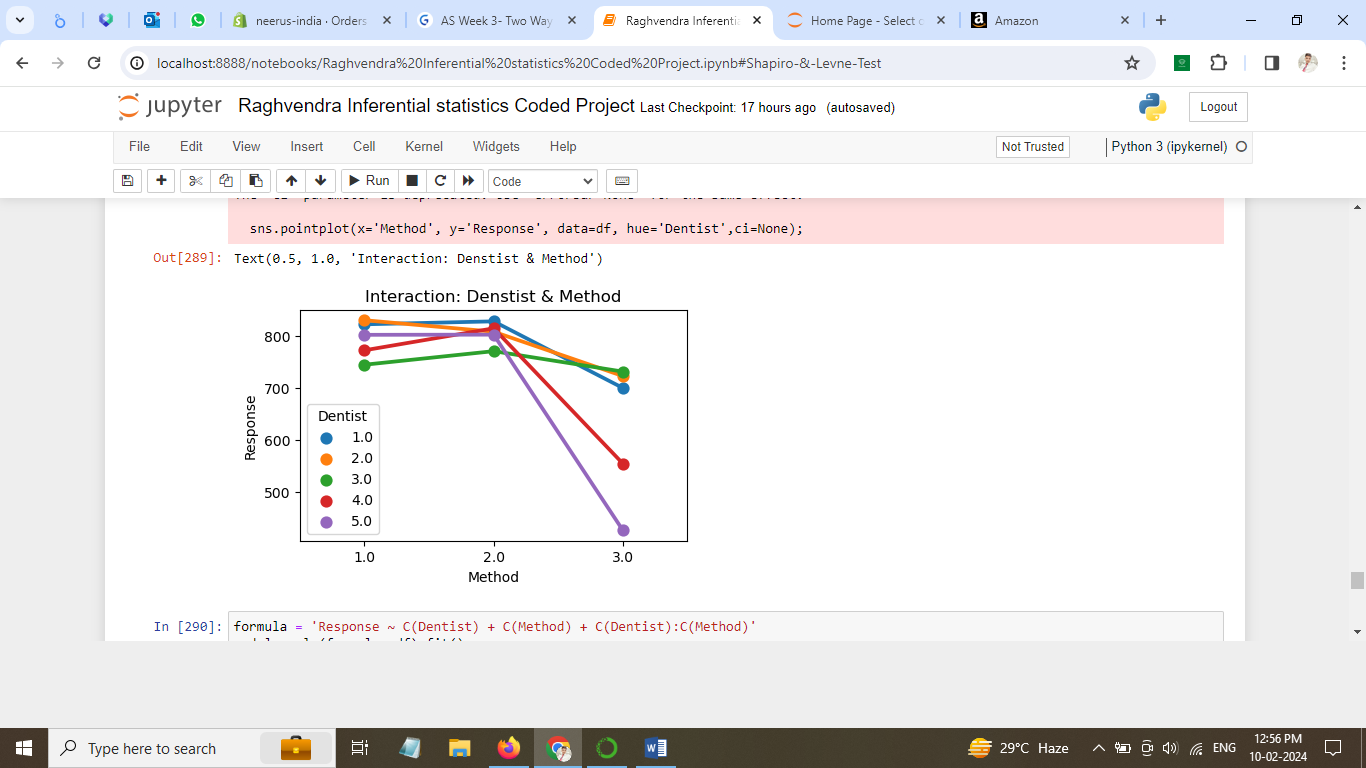
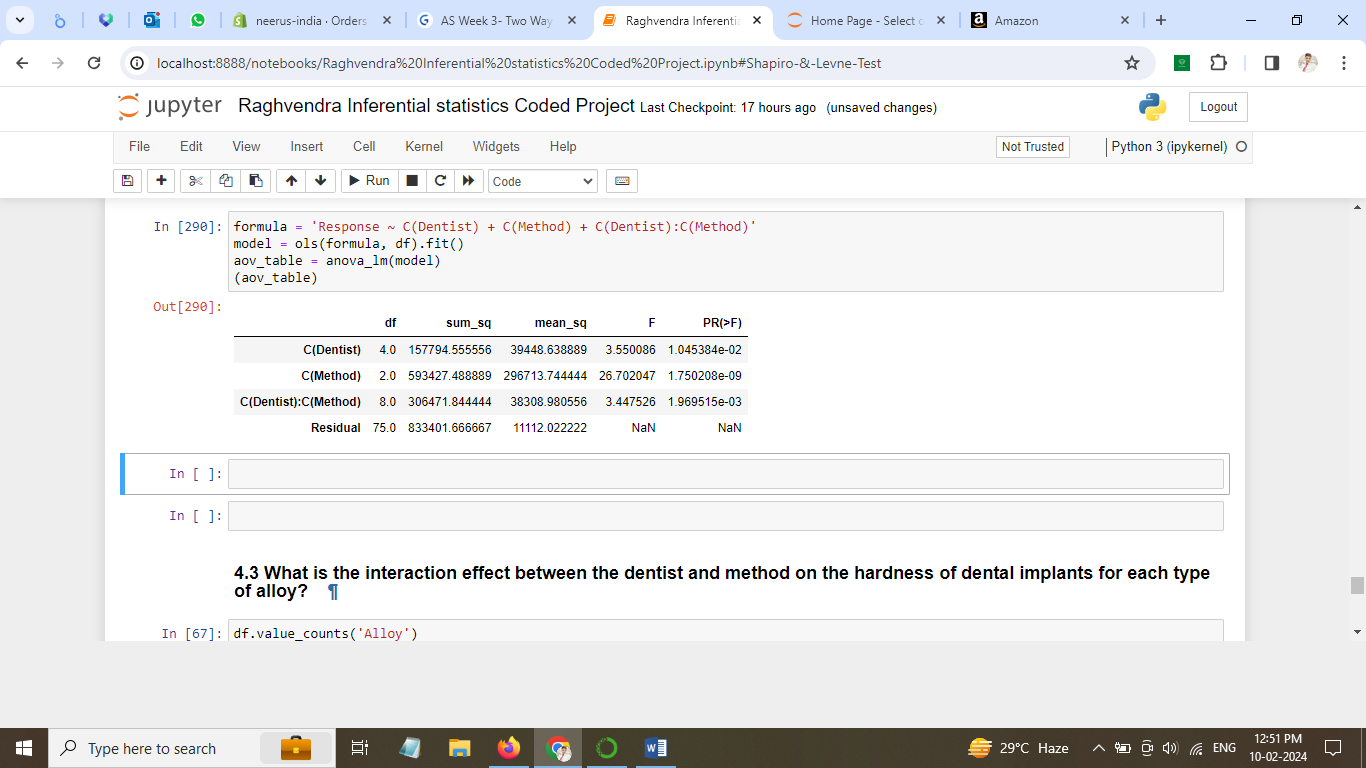


Figure 15: Combined interaction: Dentist & Method

We can see that there is some interaction between Dentist and Method. We will introduce new term while performing two way anova.



We see that p value is less than .05 for all interactions.

* **After inclusion of interaction effect term we see change in P value of the first two treatments as compared to Two-Way ANOVA without interaction terms.**
* **P-value of the interaction effect term of 'Dentist’ and 'Method' suggests that the Null Hypothesis is rejected in this case.**