BRACT’s

**Vishwakarma Institute of Information Technology**

**Department of Engineering and Applied Sciences**



**SCE Report**

For

**Optimizing agriculture production**

By

|  |  |  |
| --- | --- | --- |
| Name: | Gr number | Roll number |
| Vaishnavi Bhavsar | 21910446 | 321008 |
| Deepika Gujar | 21911076 | 321023 |
| Atharva Borawke | 21911120 | 321011 |

**Index**

**1 . Introduction**

**2 . Data**

* Dataset Introduction
* Attributes and Types

**3 . Importing Libraries and Data**

**4 . Data Cleaning**

**5 . Data Analysis and Labeling**

* Value counting
* Statistic
* Average (mean)
* Min & max threshold
* Comparing values requirement
* Finding Some interesting patterns in dataset
* Comparing the crops which require above and below-average conditions
* Clustering analysis
* Elbow method

1. **. Models**

Logistic Regression Model

* Checking accuracy
* Report for accuracy

Decision Tree Model

* Checking accuracy
* Report for accuracy

**7. Result**

Comparing results

**8. Conclusion**

**9 . Research Papers**

**Problem statement**: Build a predictive model so as to suggest the most suitable crops to grow based on the available climatic and soil conditions.

**Goal**: Achieve precision Farming by Optimizing Agriculture production.

* This project is intended on Precision Farming.
* We have to optimize productivity
* By understanding requirements of climatic and soil conditions for crops.
* Helps us to cope up with weather Unpredictability.

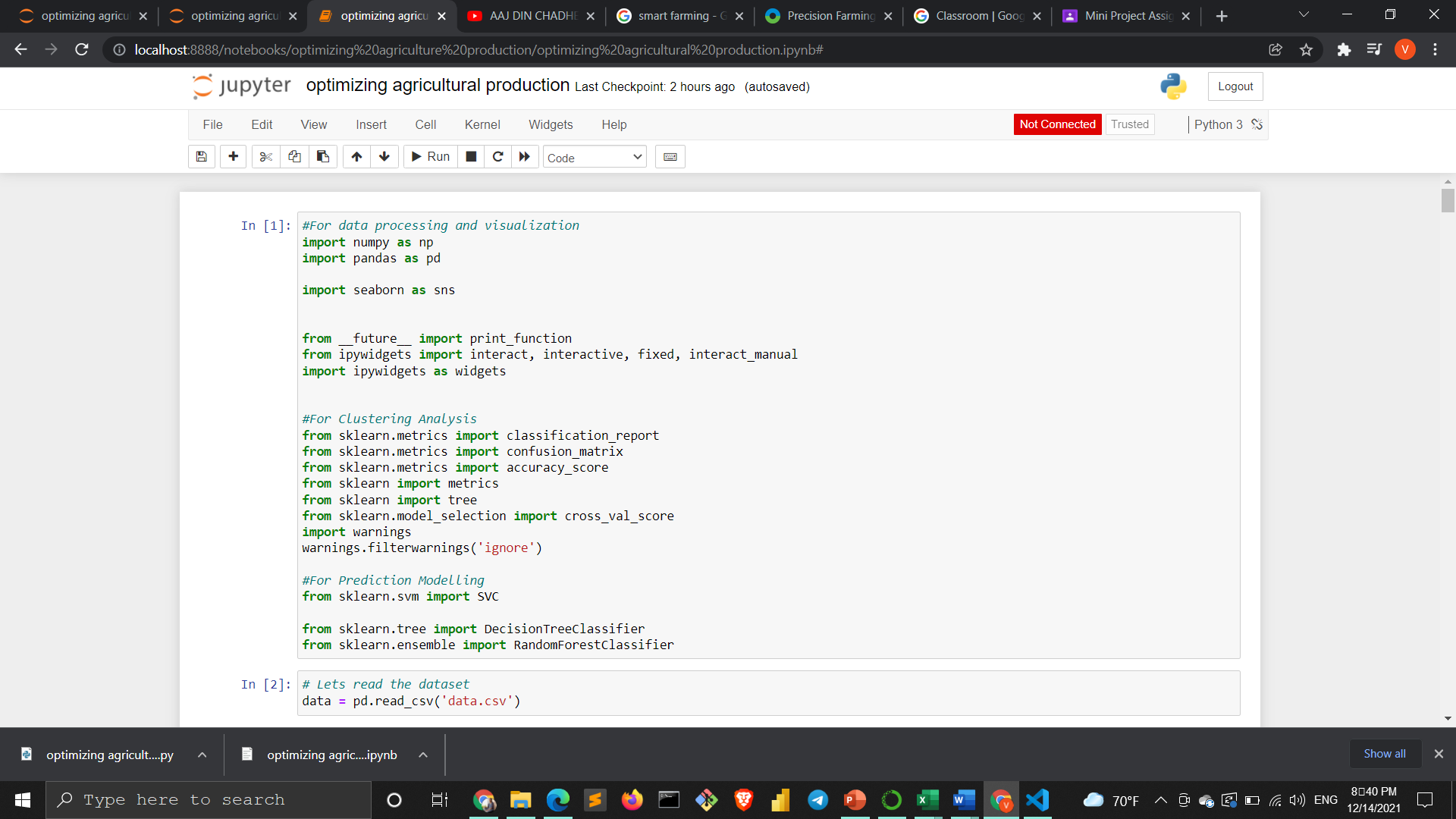
**Explanation of dataset:**

|  |  |  |
| --- | --- | --- |
| **SR NO** | **Attributes** | **Description** |
| 1. | Source of the dataset | Kaggle |
| 2. | Number of attributes | 8 |
| 3. | Attributes name | Nitrogen, Phosphorus, Potassium, Temperature , Humidity, PH, Rainfall,Label |
| 4. | No. of samples | 2201 |
| 5. | Missing values | No missing values |

**Attributes and types:**

|  |  |  |
| --- | --- | --- |
| **Attribute name** | **Types (Values)** | **Description** |
| Nitrogen | Integer | Amount of nitrogen in soil |
| Phosphorus | Integer | Amount of nitrogen in soil |
| Potassium | Integer | Amount of nitrogen in soil |
| Temperature | Float point number | Temperature of atmosphere |
| Humidity | Float point number | Humidity of atmosphere |
| PH | Float point number | Ph of soil |
| Rainfall | Float point number | Rainfall during crop season |
| Label | String Literal(rice, maize, chickpea, kidney beans,  Pigeon peas, moth beans, mungbean, black gram, lentil, pomegranate, banana, mango, grapes, watermelon, muskmelon, apple, orange, papaya, coconut, cotton, jute, coffee) | Tells crops name during range of environmental factors. |

Libraries and Data imported:



Numpy :

Seaborn:

Pandas:

Ipywidgets:

Sklearn:

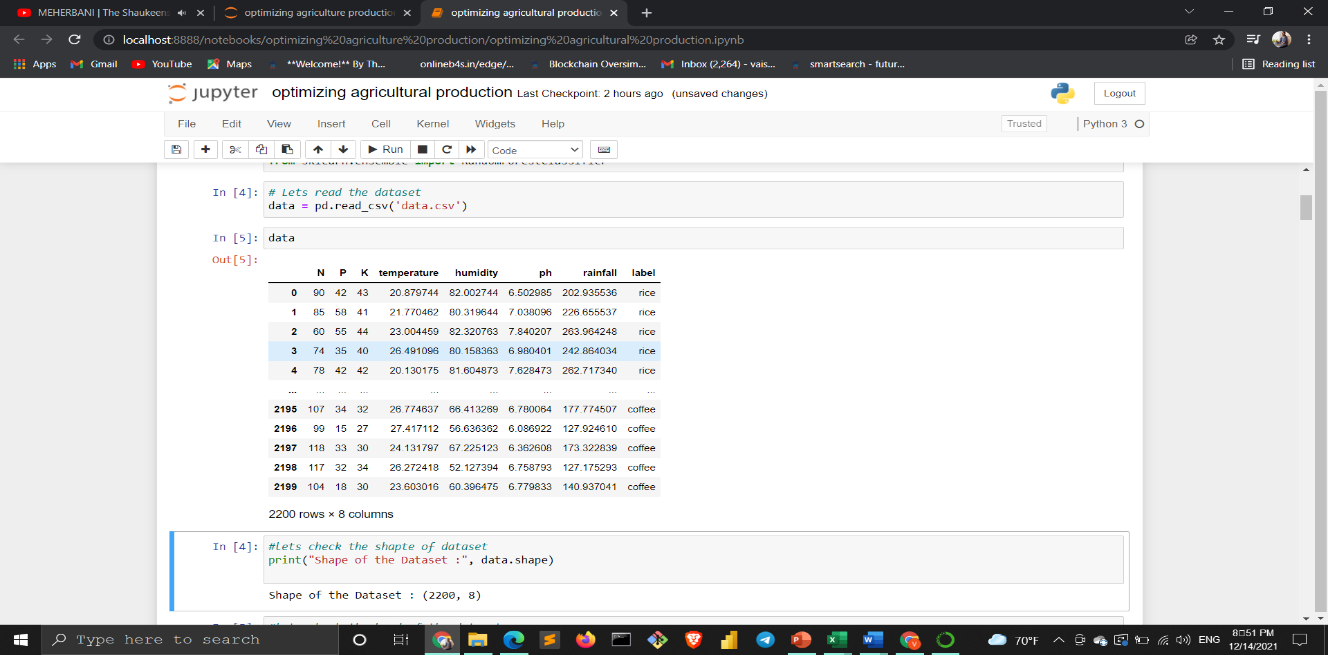
DecisionTreeClassifer:

**Reading data**

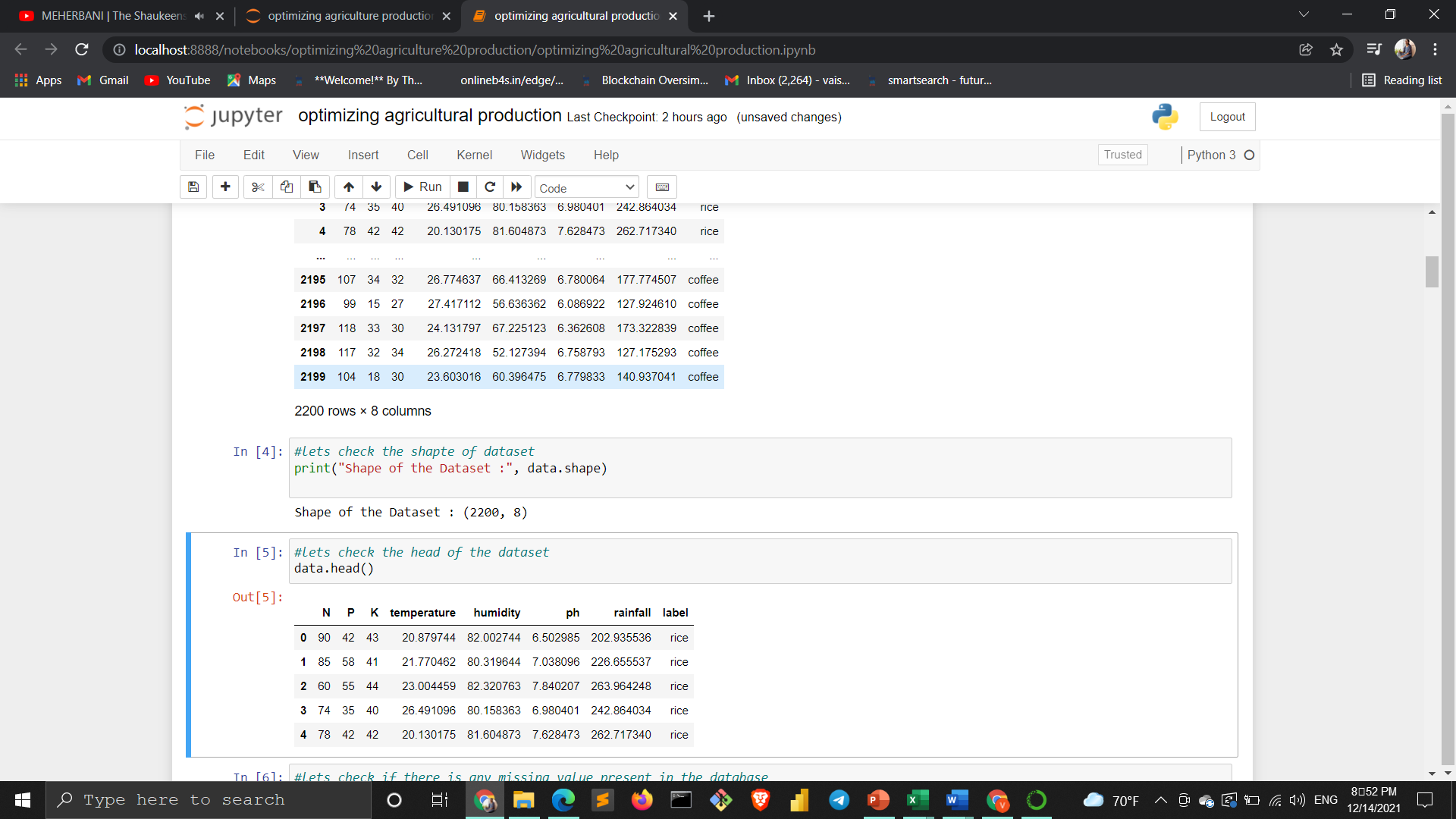
**CSV**: With help of this CSV, we create a prediction model by understanding the pattern in the data set

Data = pd.read\_csv(‘data.csv’)

\*main dataset stored in variable named data



Checking size of dataset Print(“shape of the dataset :”,data. shape)

2200 number of rows and 8 number of columns in data set

**Head function: shows first n rows of the dataset**

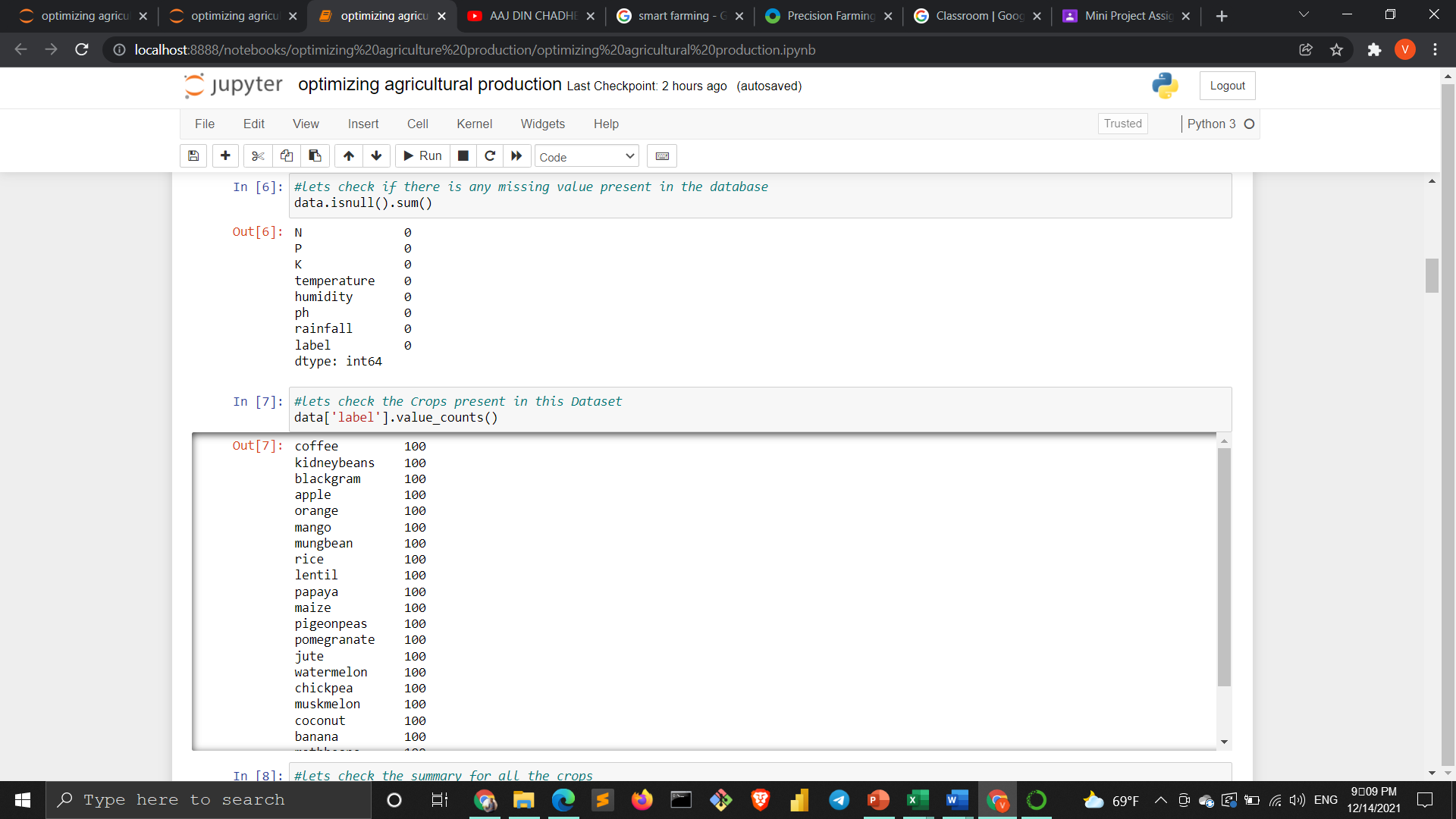
Data.head()

**Data cleaning:**

**Dealing with null values:**

**Data.isnull().sum()**

**0 indicates no missing values in dataset**



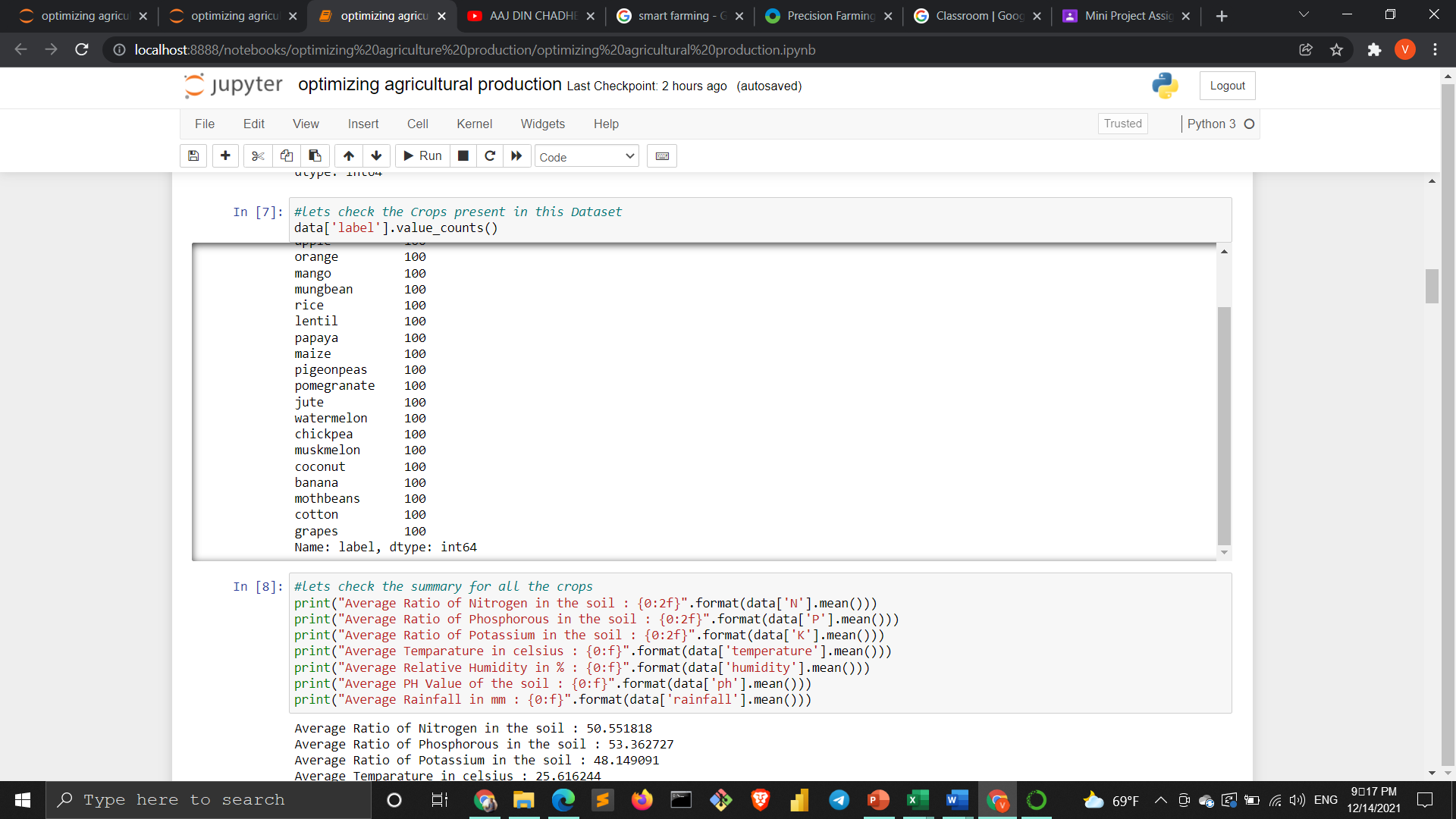
* **Fill-NA function is used to replace these missing values such as mean, mode or median**
* **Na means Not Available.**
* **Pandas have functions like fill-NA to treat missing values**

**No unwanted column (attribute) in data set to drop out**

**Data analysis**

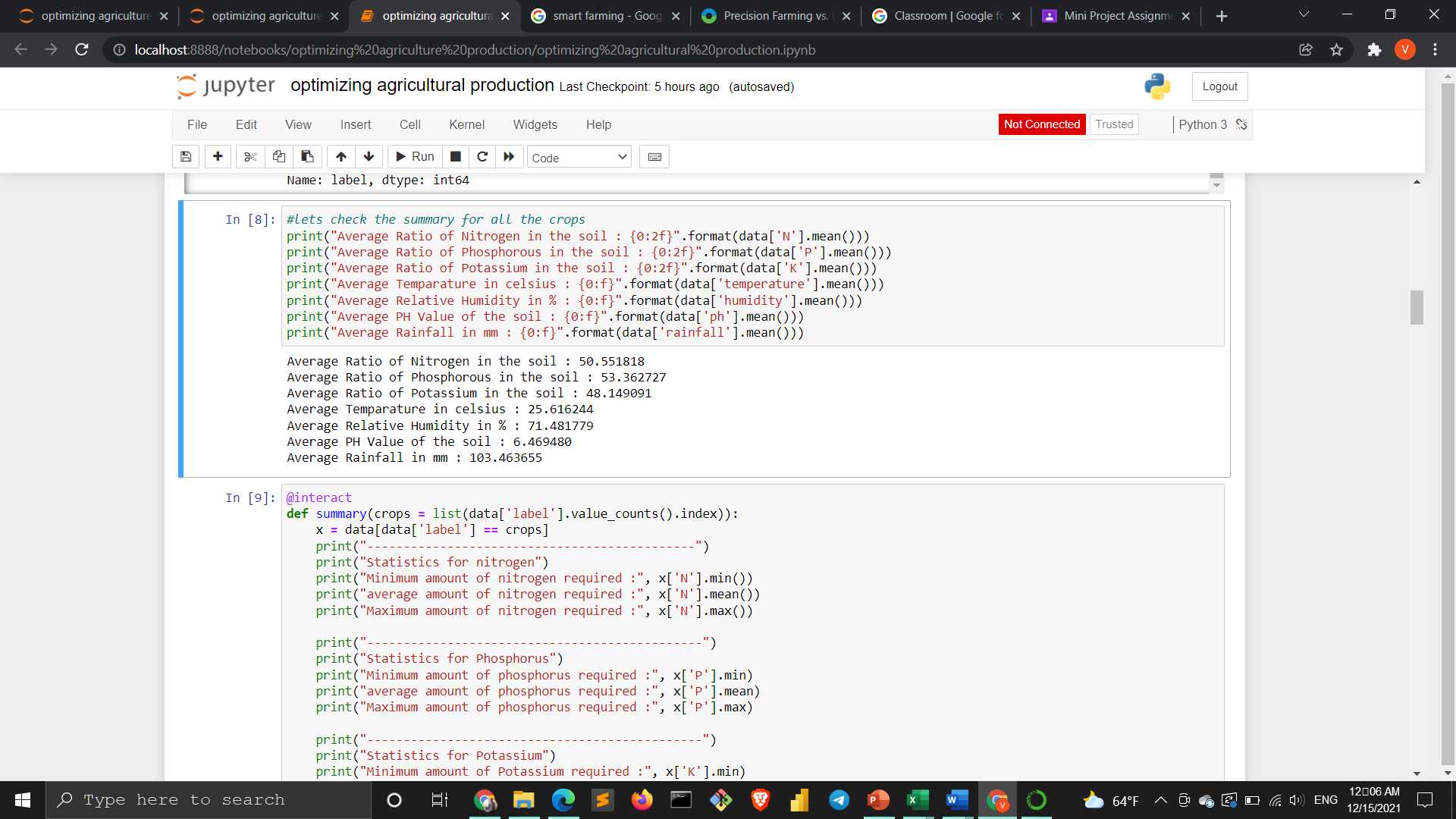
**Value counting: it tells us the occurrence of every unique value of the column**

**It tells the number of records of each crop in dataset**

****

**Statistics:**

**Average climatic and soil requirements:**

****

**Results:**

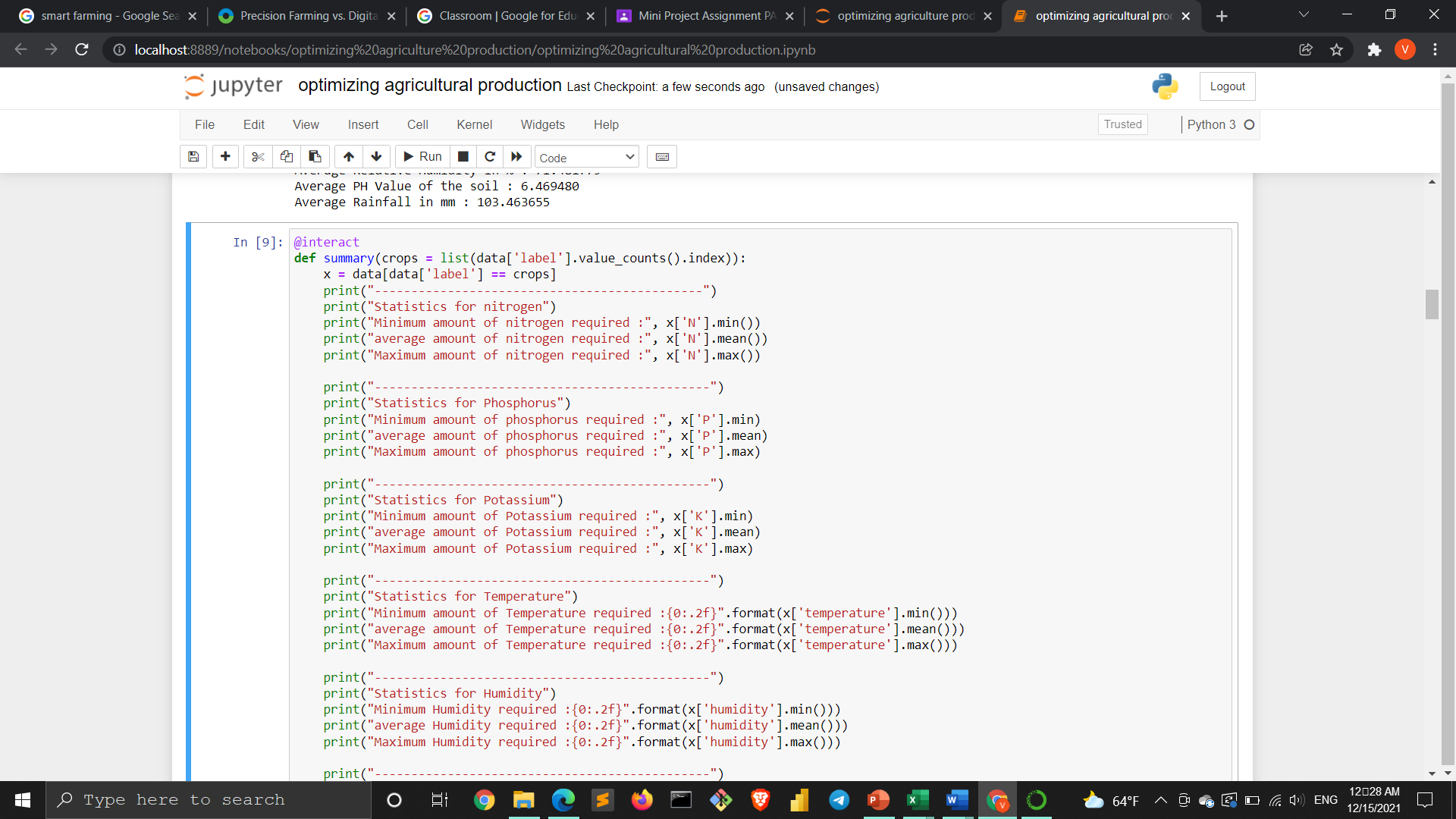
**Averagely content of nitrogen, potassium, and phosphorous in the soil is 50%, 53%and 48% respectively**

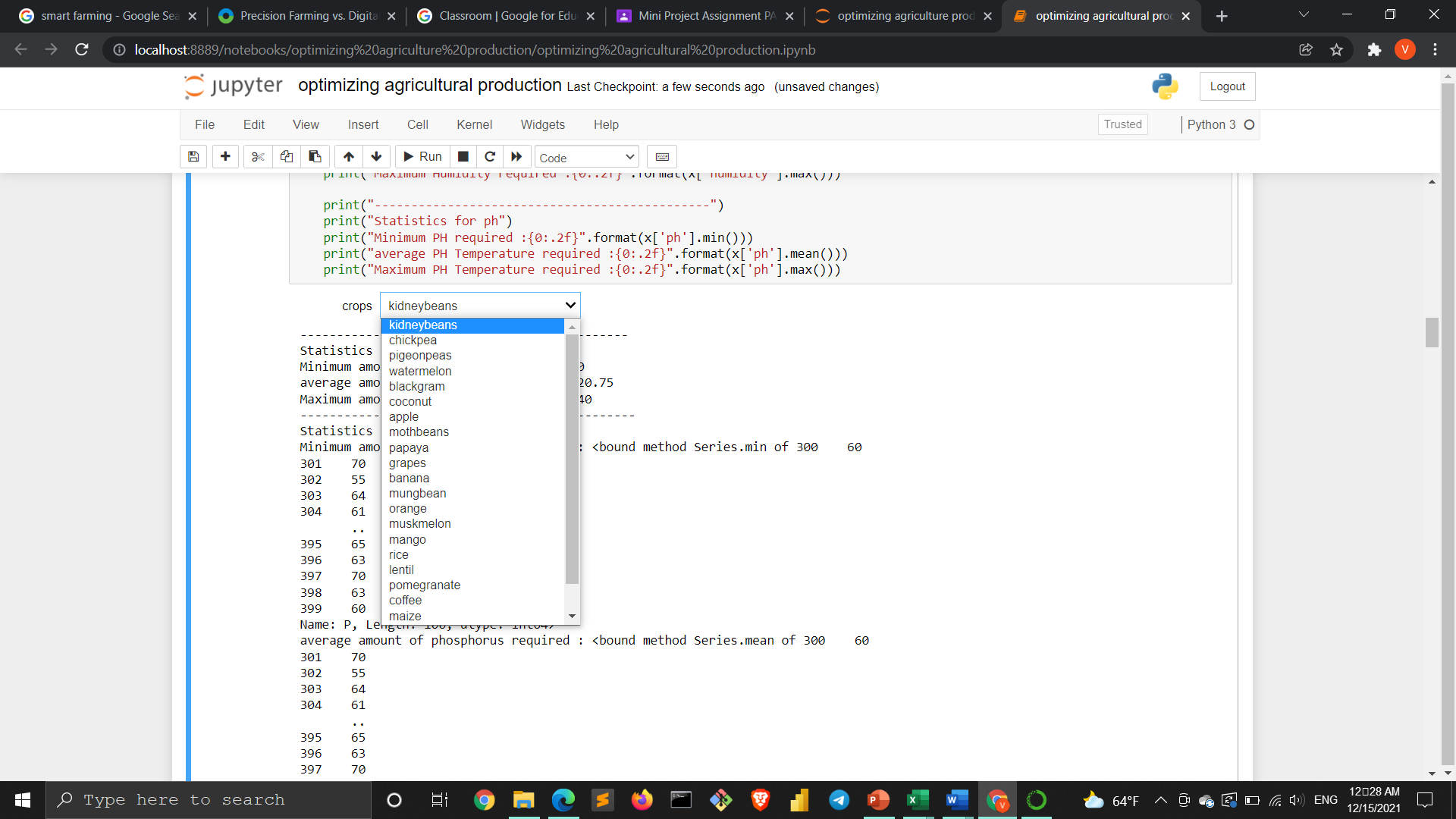
**The average temperature requirement is 25 degrees Celcius similarly humidity 71% average PH required is 6.47**

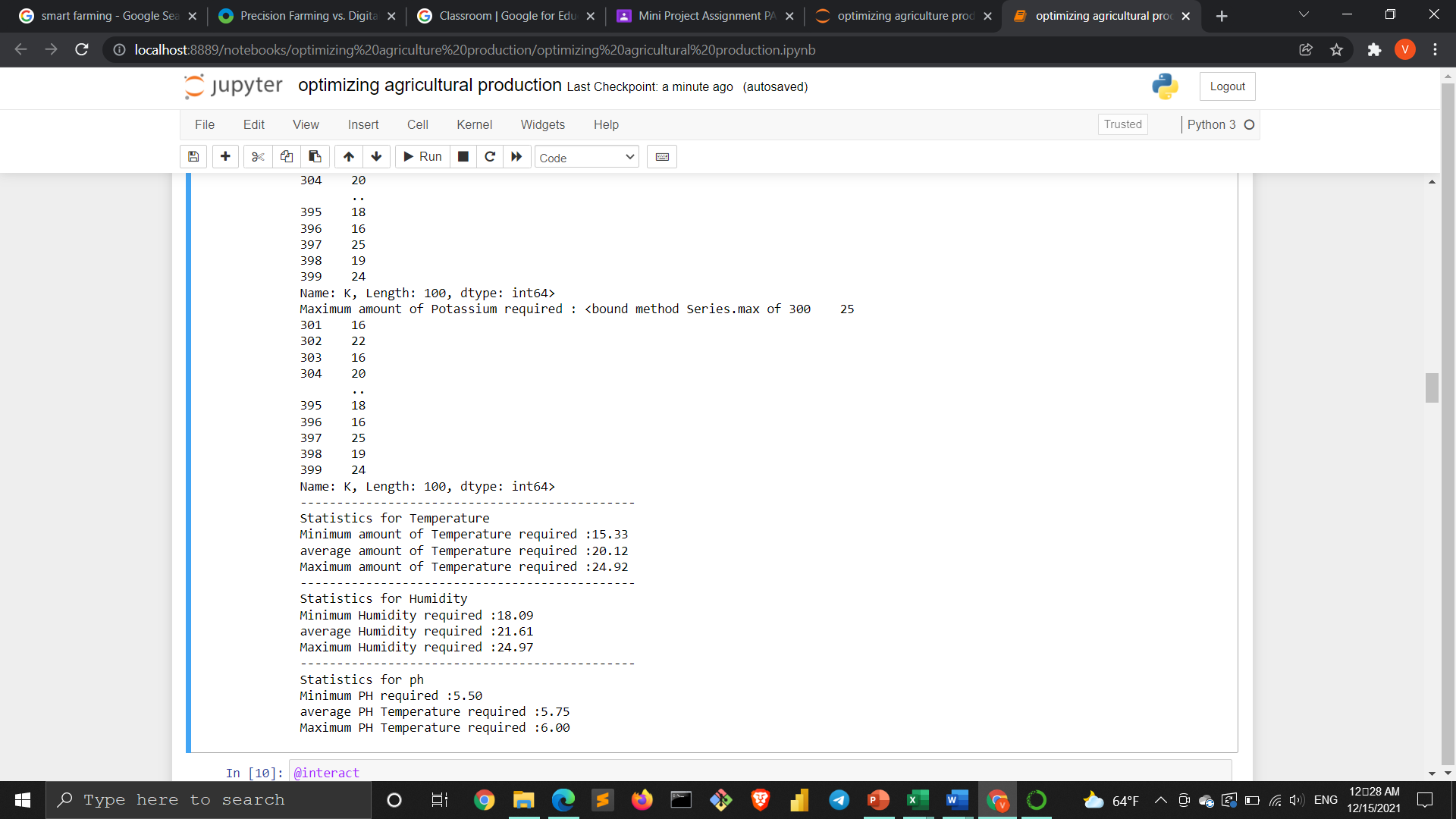
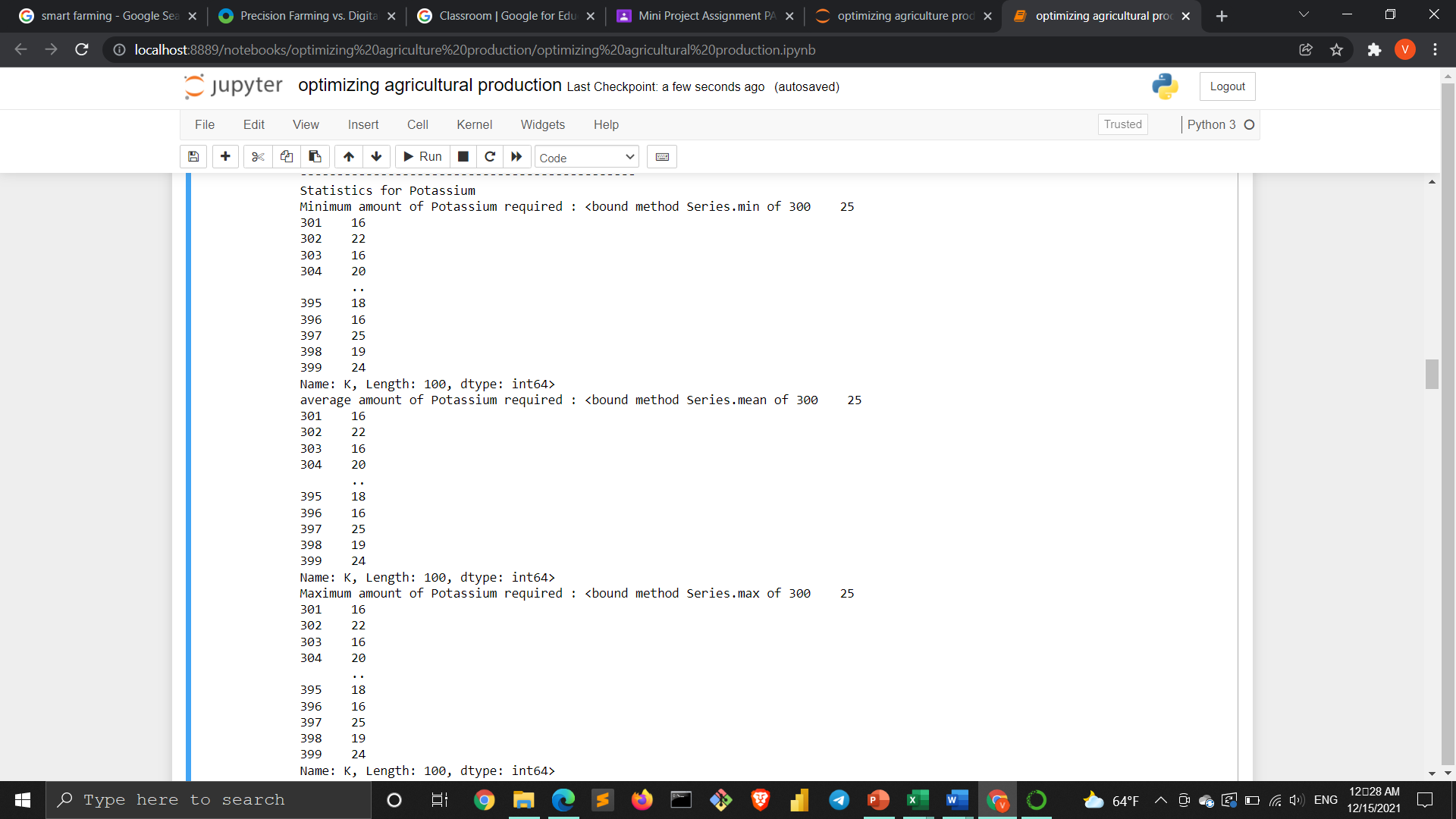
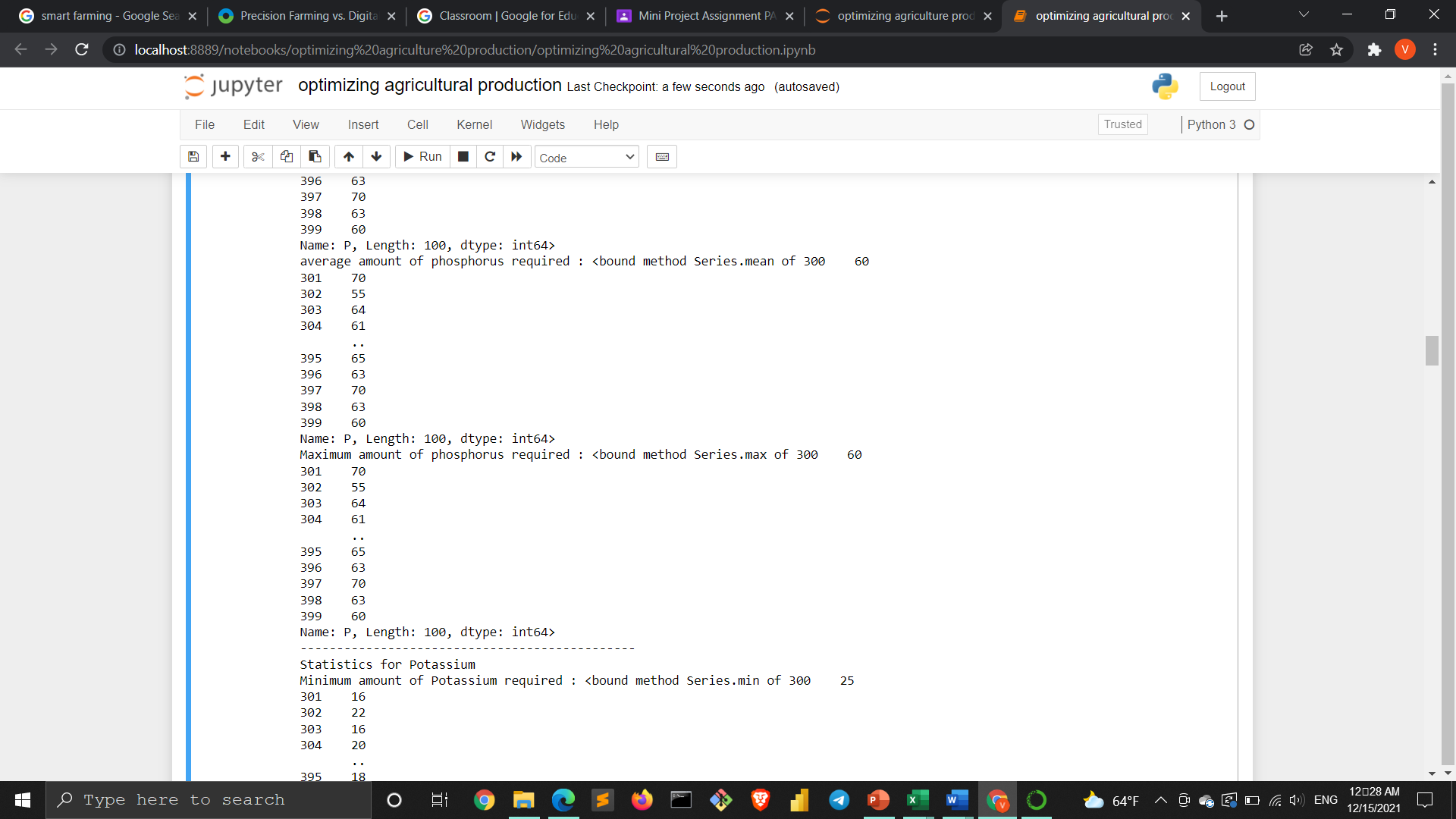
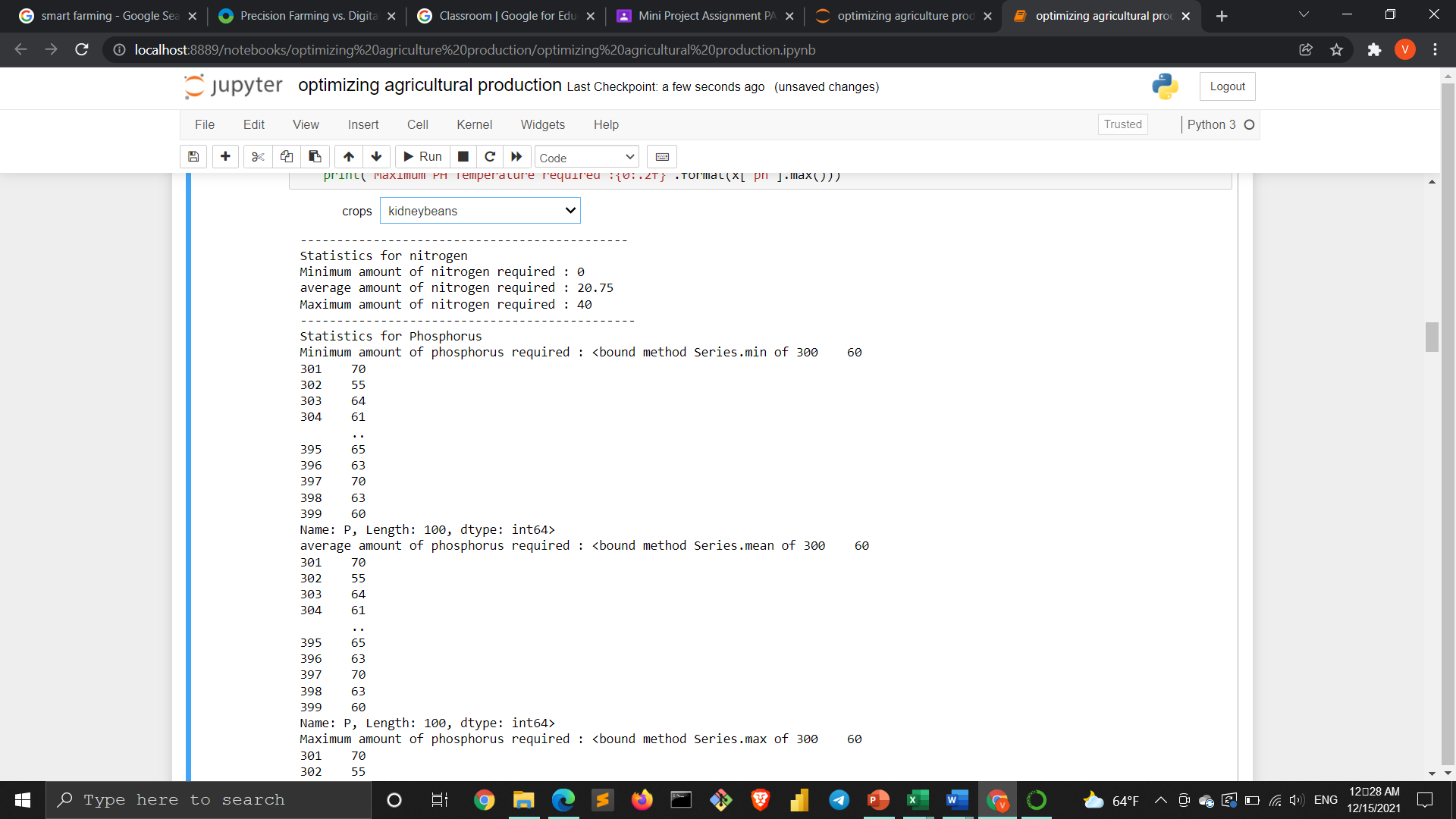
**And average rainfall is 103 mm**

**Minimum and maximum requirements of soil and climatic conditions for every crop**

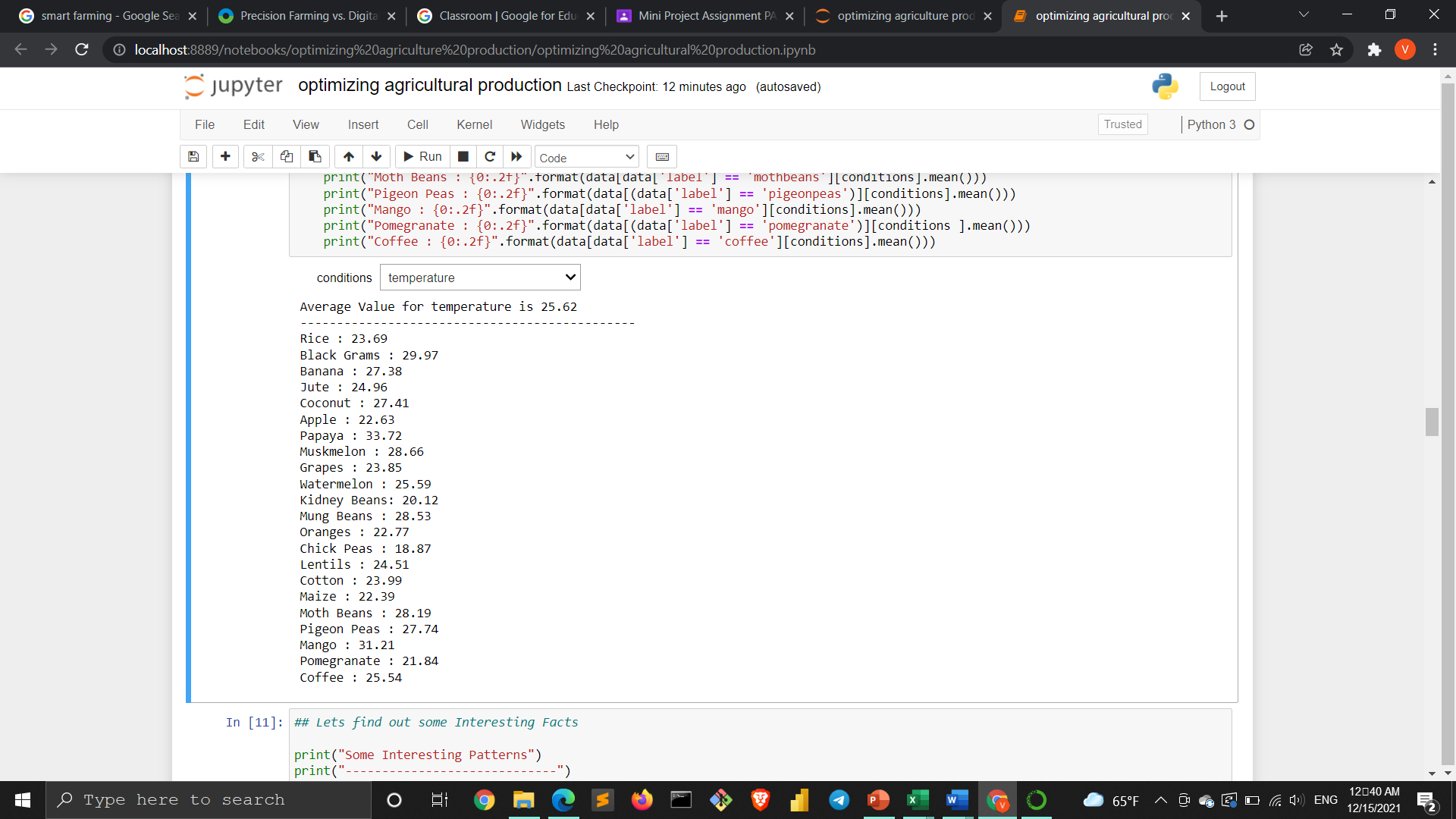
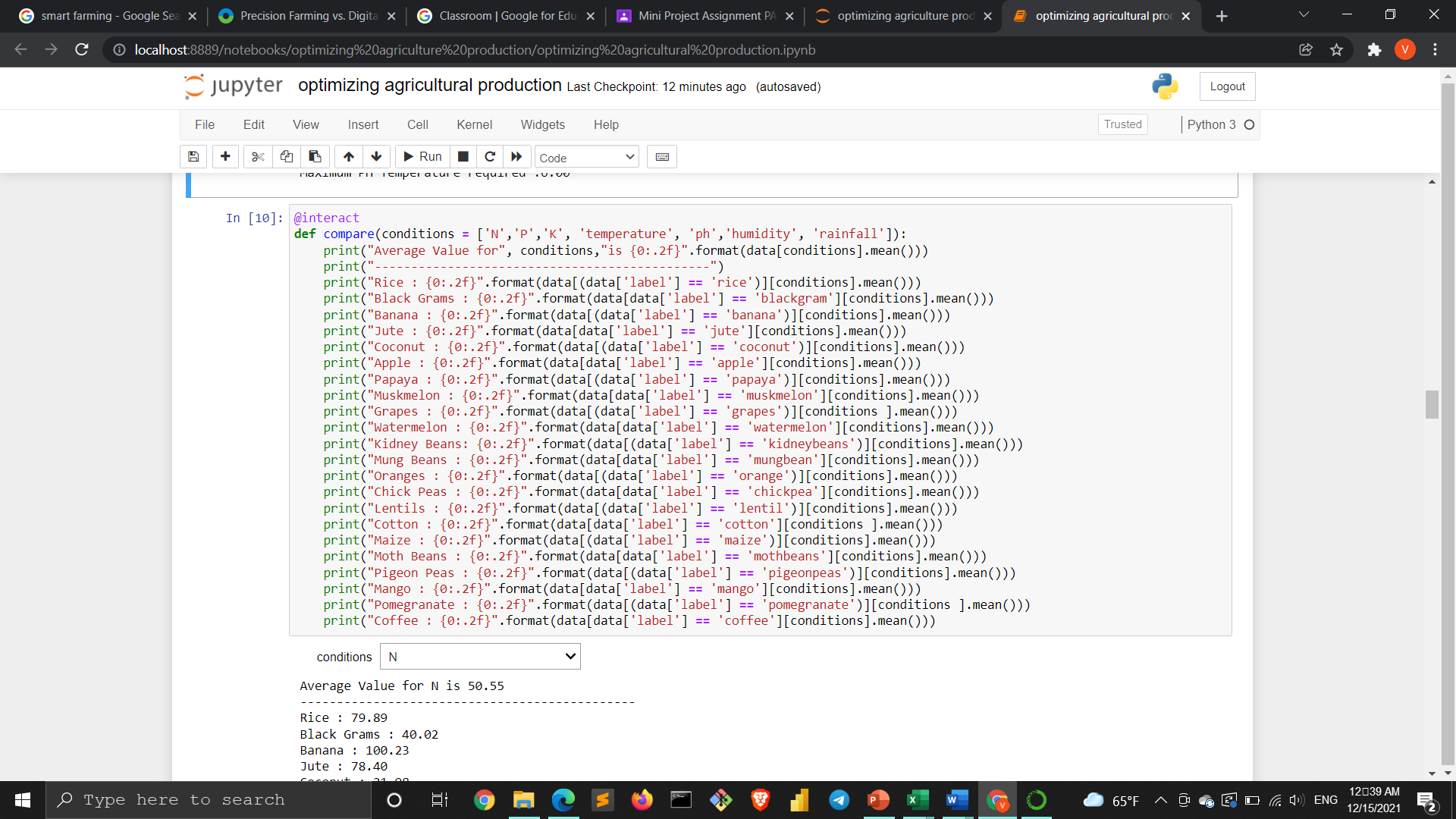
**Here we used the @interact keyword this create a drop-down this can be only used with the ipywidgets library**

****

****

****

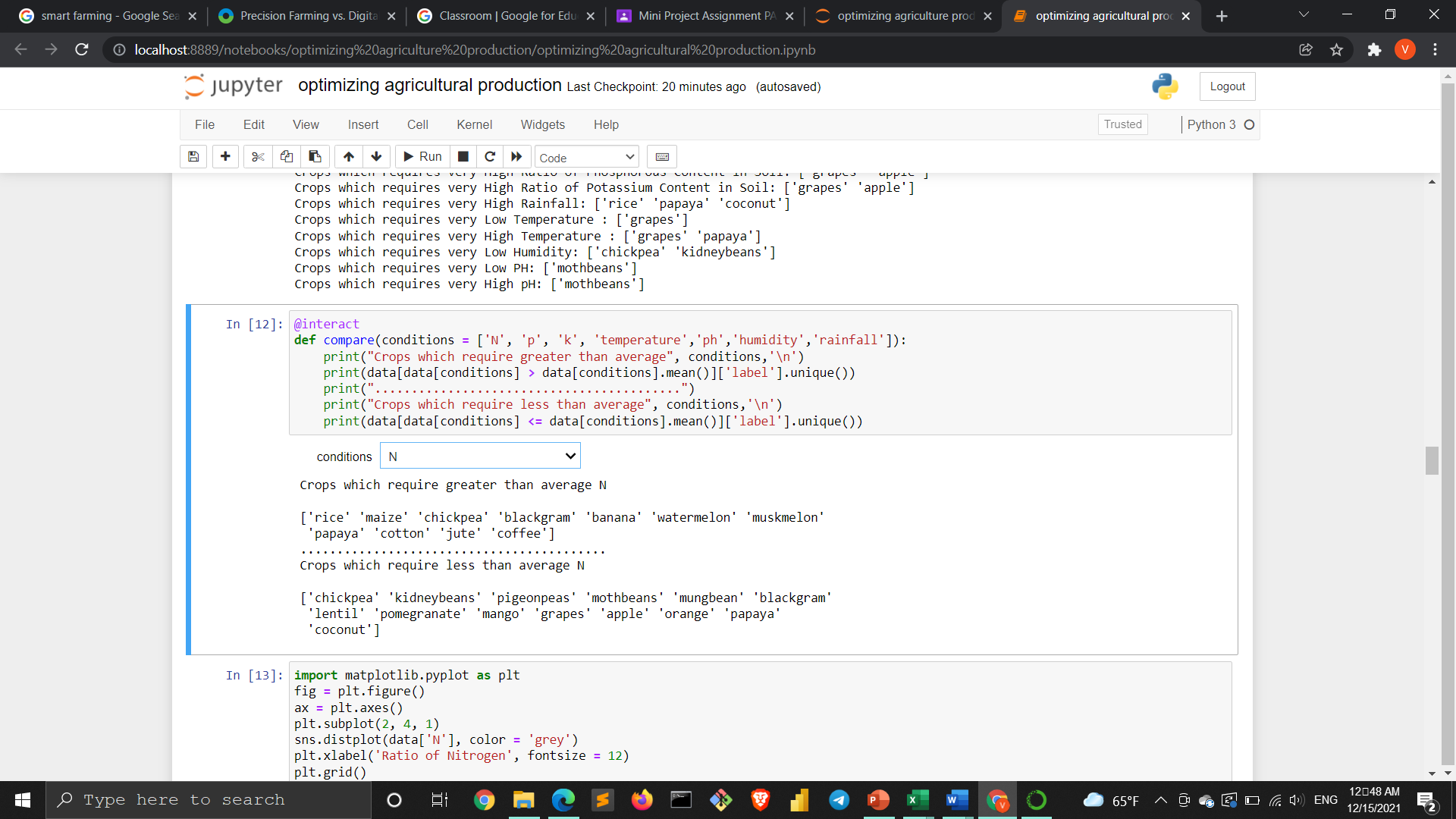
**Comparing values of the requirement for every crop:**

****

**Conclusion: from the above temperature comparison we come to know the crop which requires very high temperature is a summer crop and which require low are a winter crop**

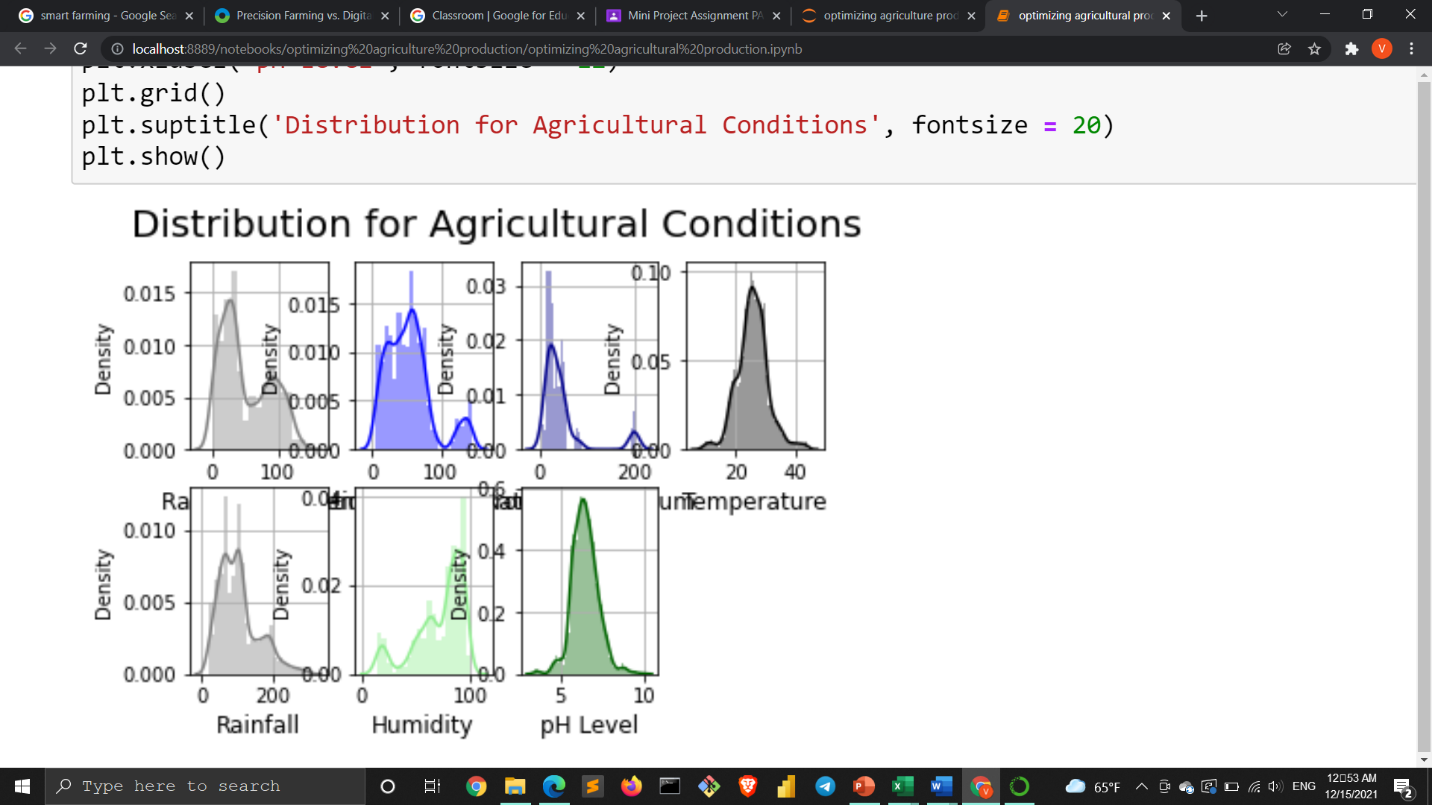
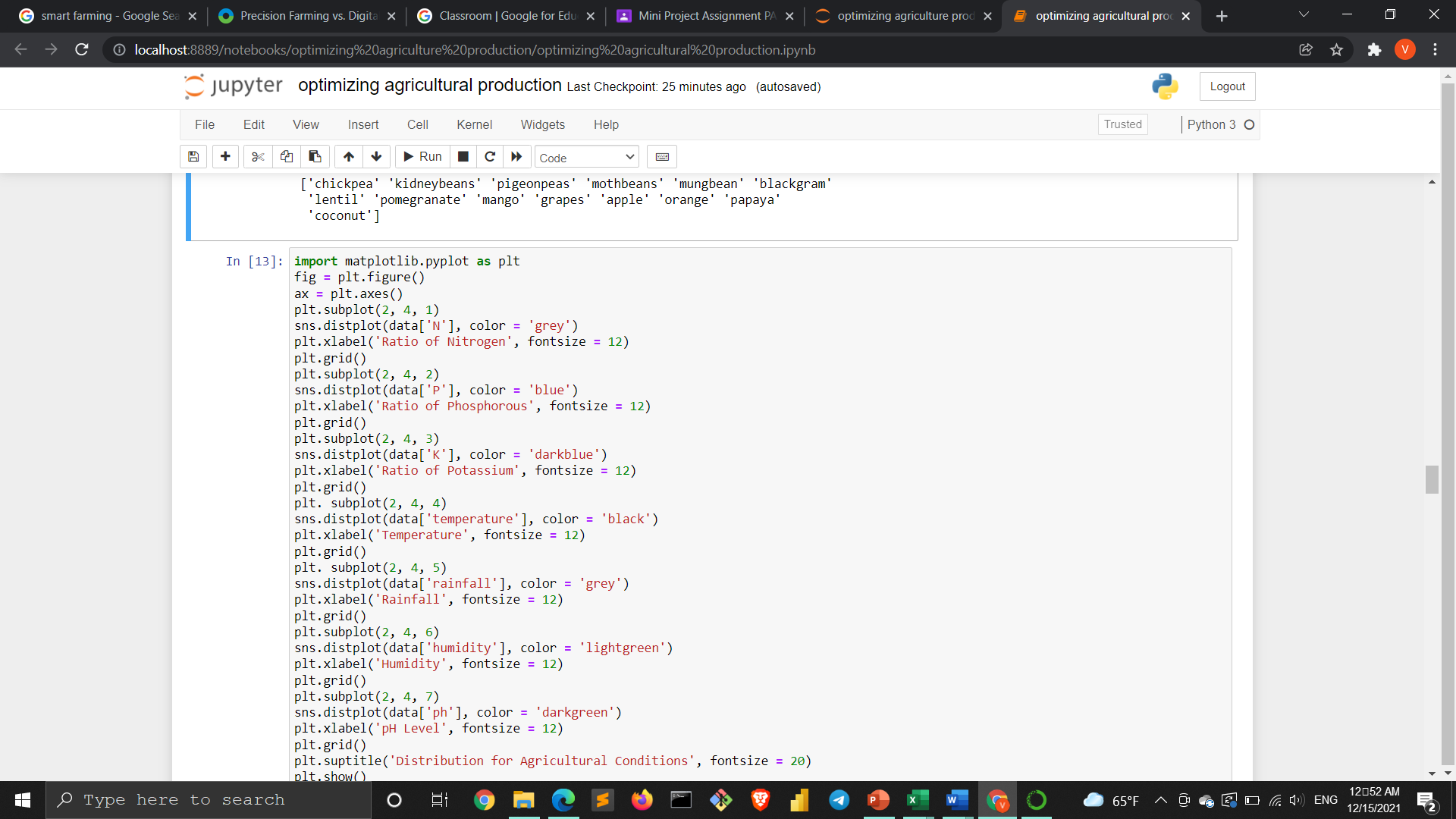
**Some interesting patterns:**

**Comparing the crops which require above and below-average conditions**

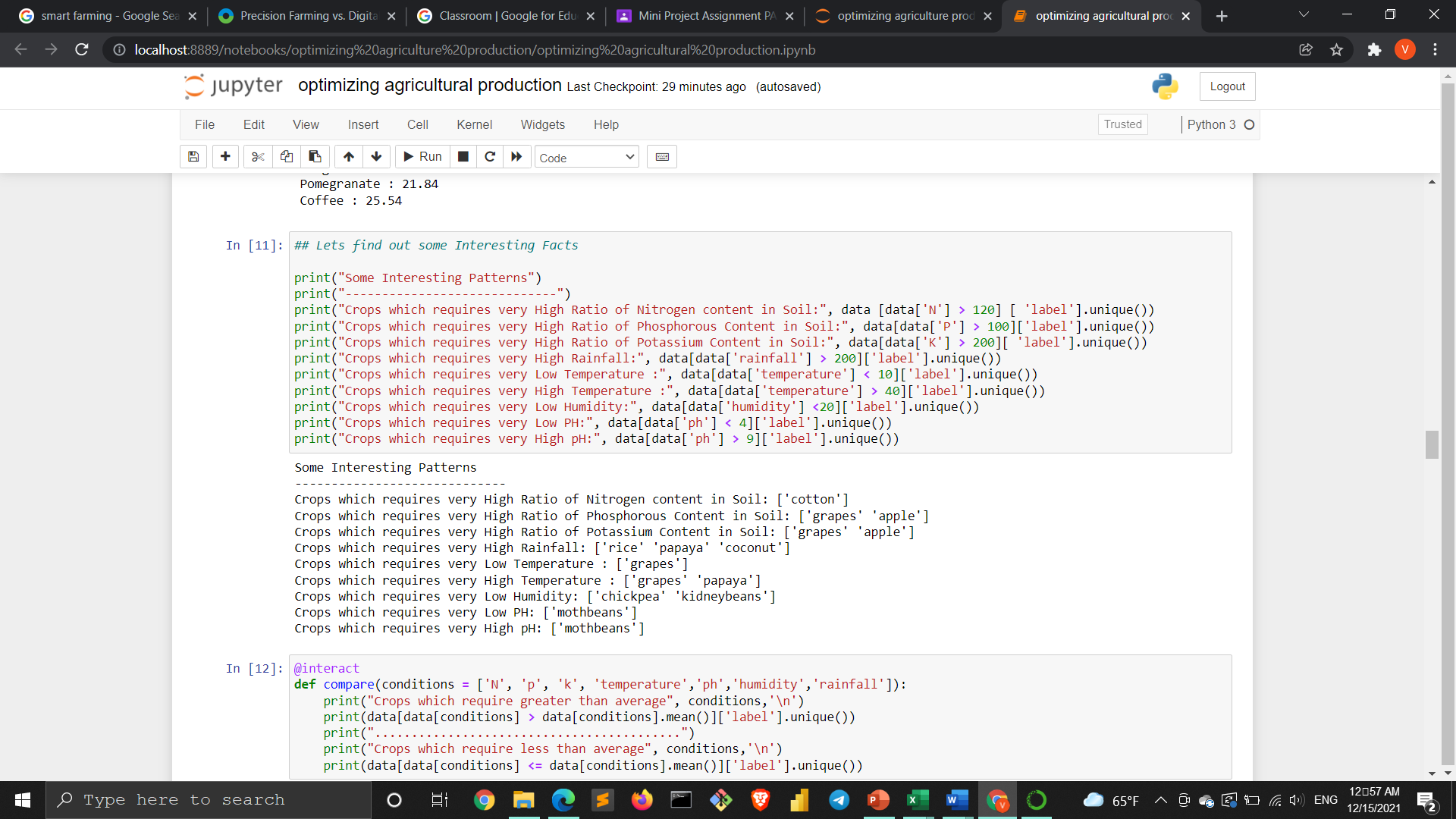
****

**Checking distribution of columns for understanding patterns:**

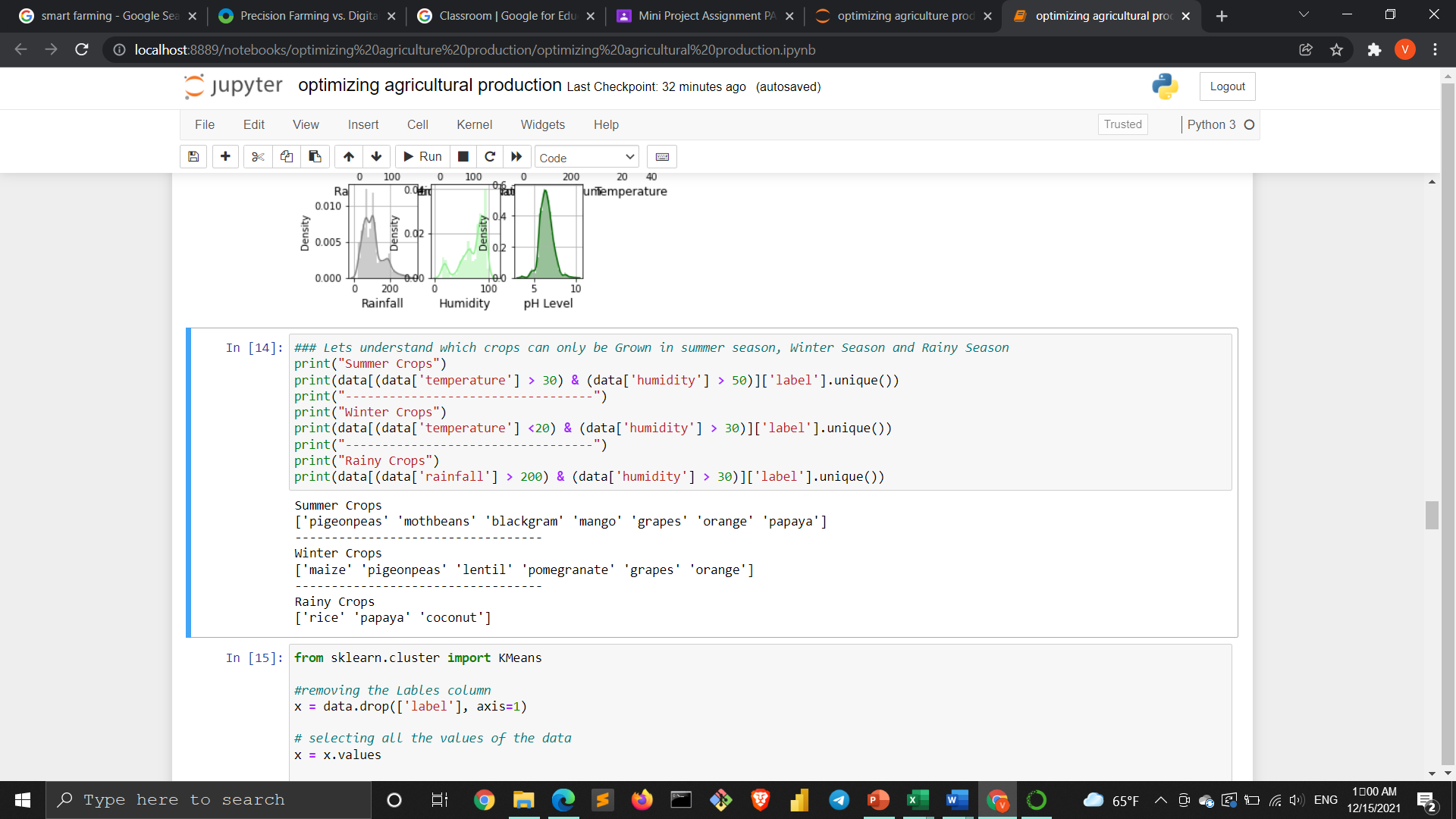
**Distribution plots:**

****

**Finding minimum and maximum thresholds (crops that have unusual requirements)**

****

**Season ideal crops:**

****

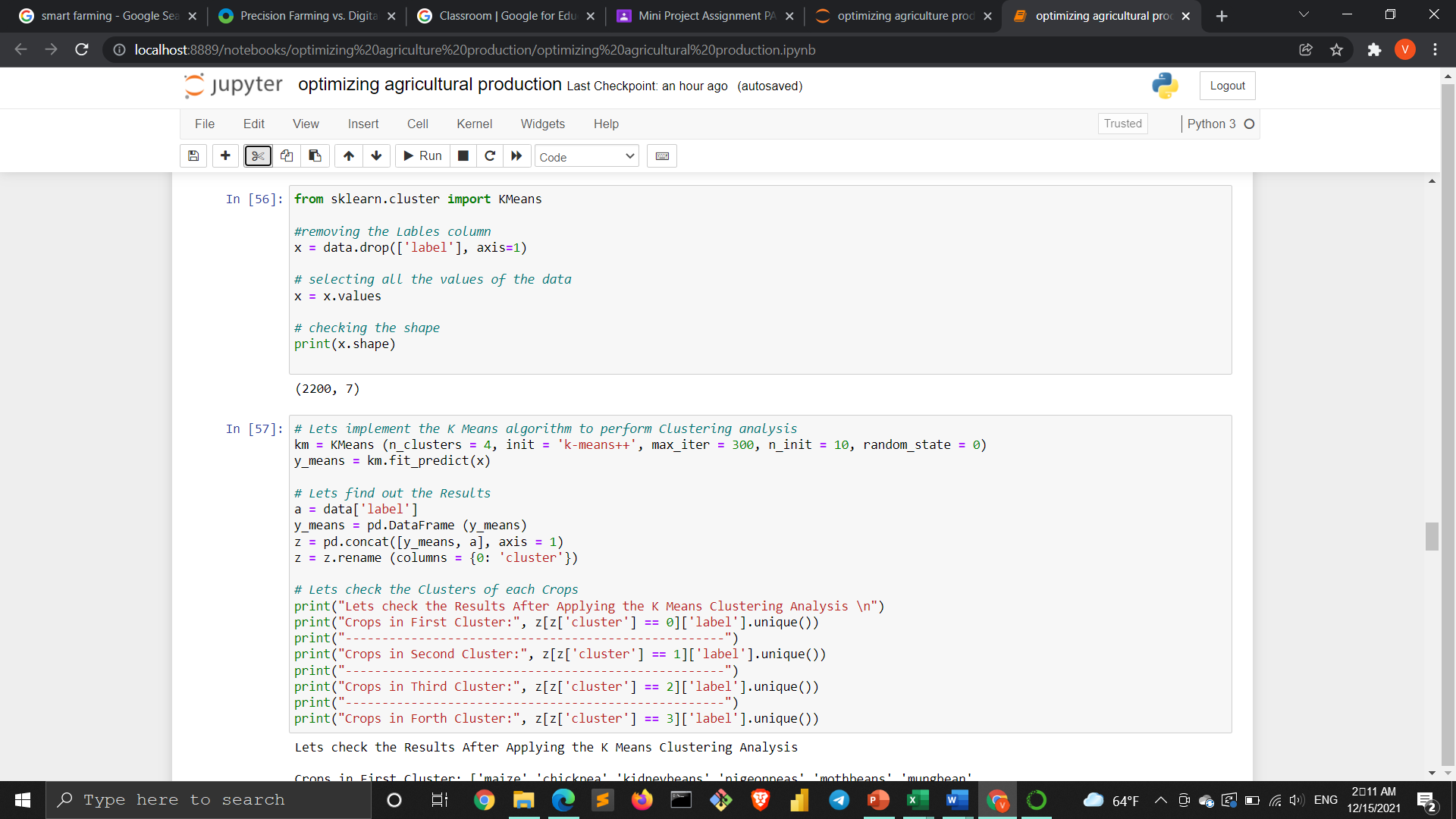
**Clustering analysis: it is a technique used to classify the data points into relative groups are called clusters.**

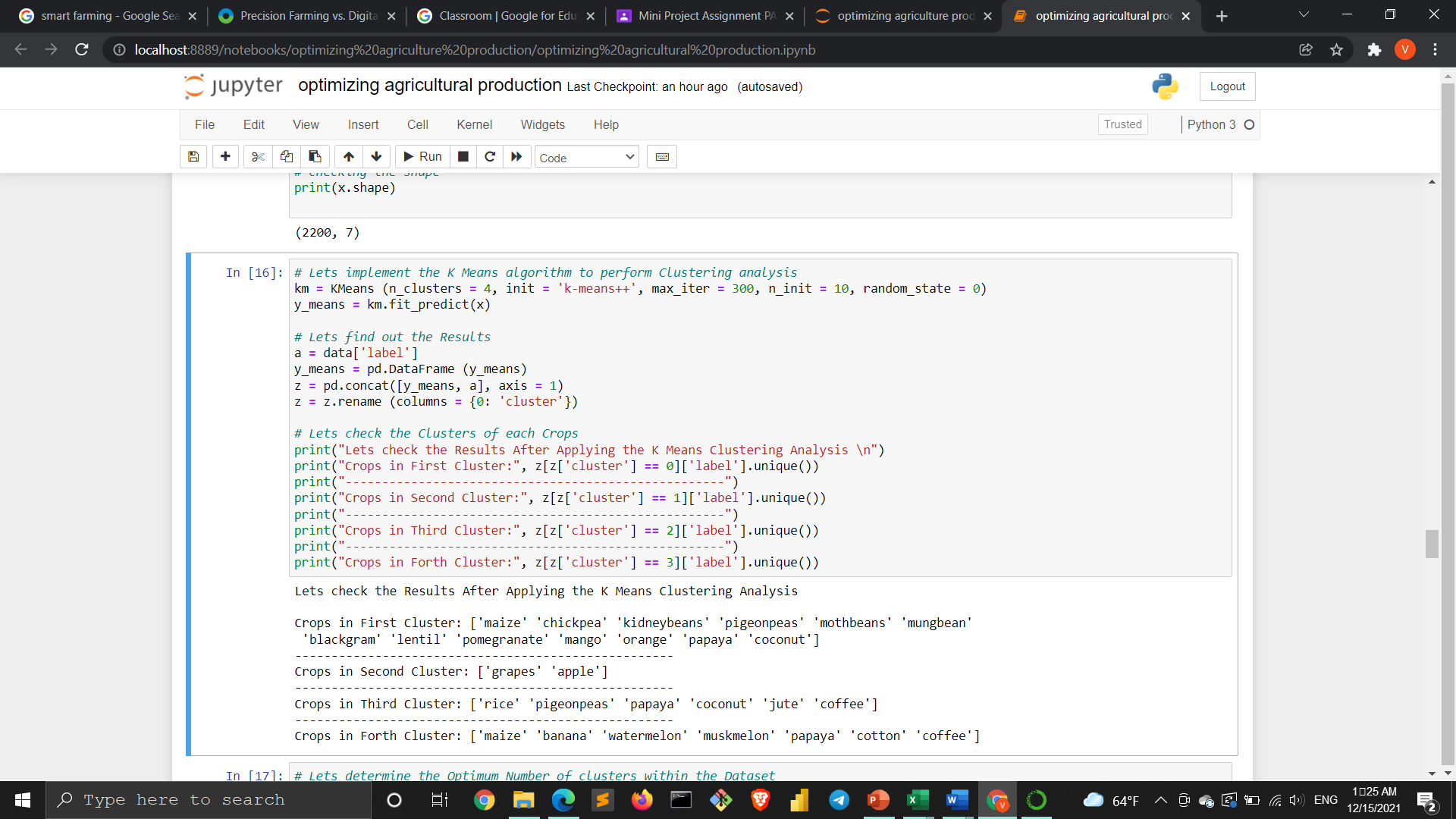
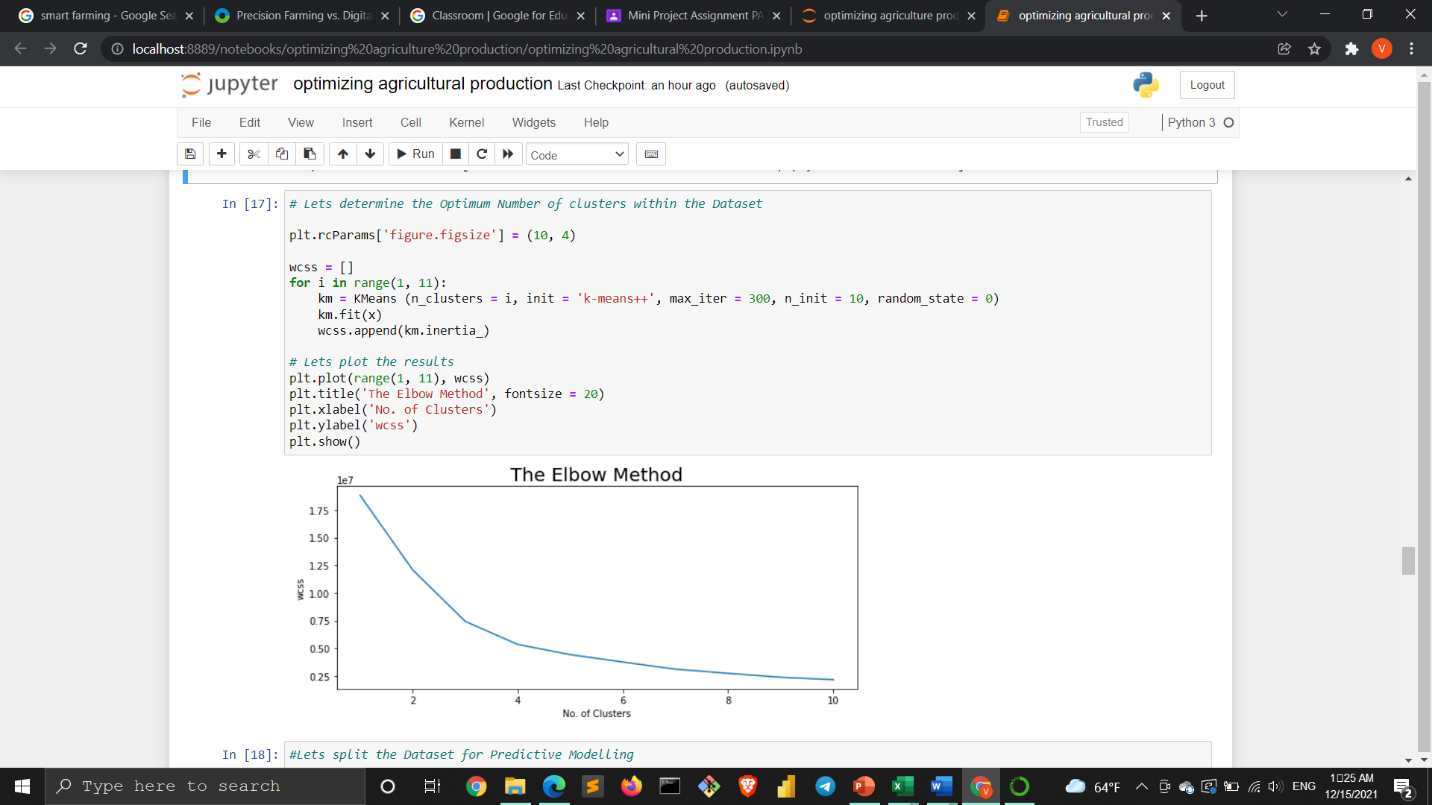
**It comes under the unsupervised learning category**

**Clustering analysis used by importing sklearn library**

**And removing labels by using the drop function**

**To know the number of clusters has to apply the elbow method**

****

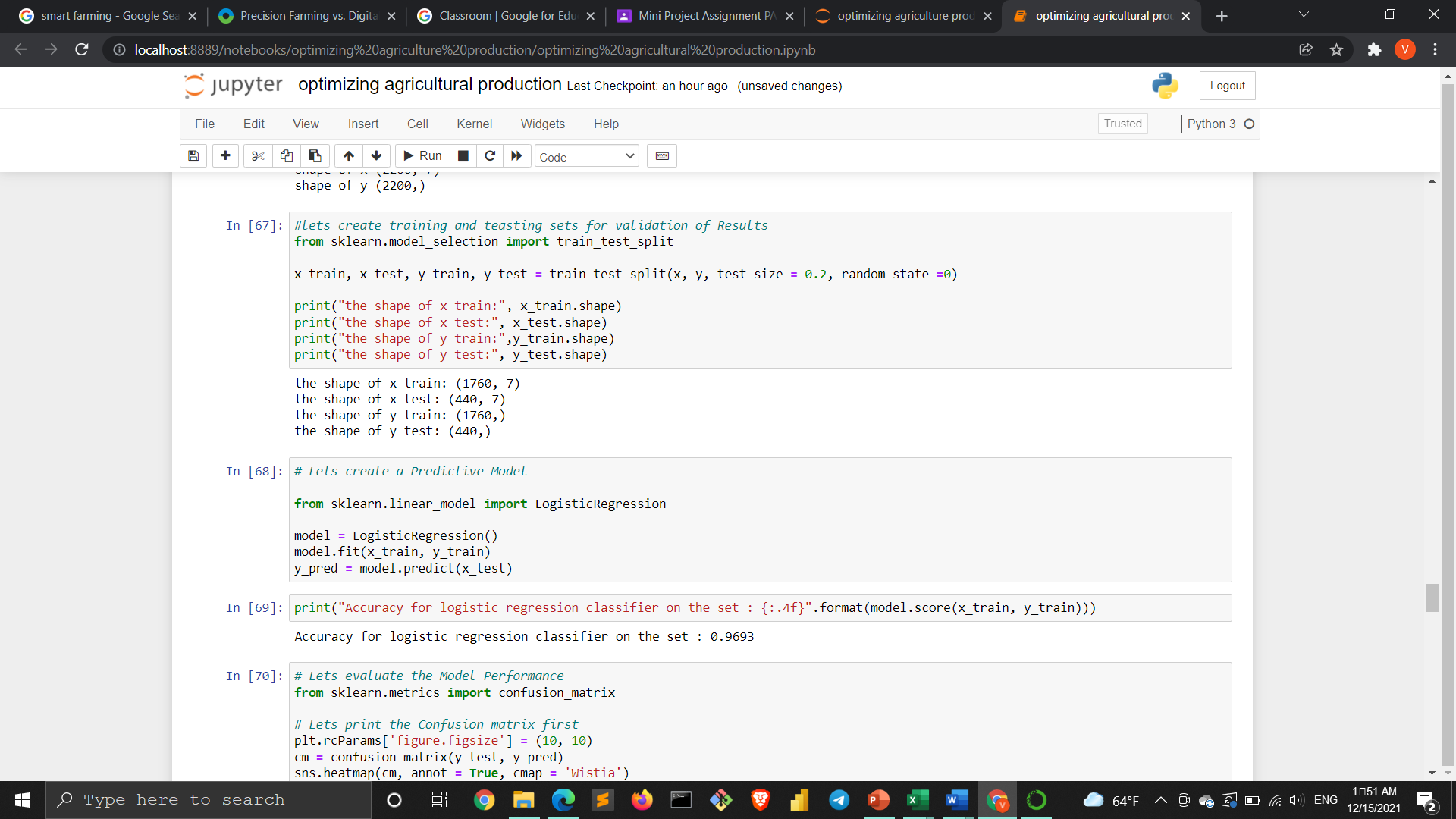
****

**Machine learning predictive model:-**

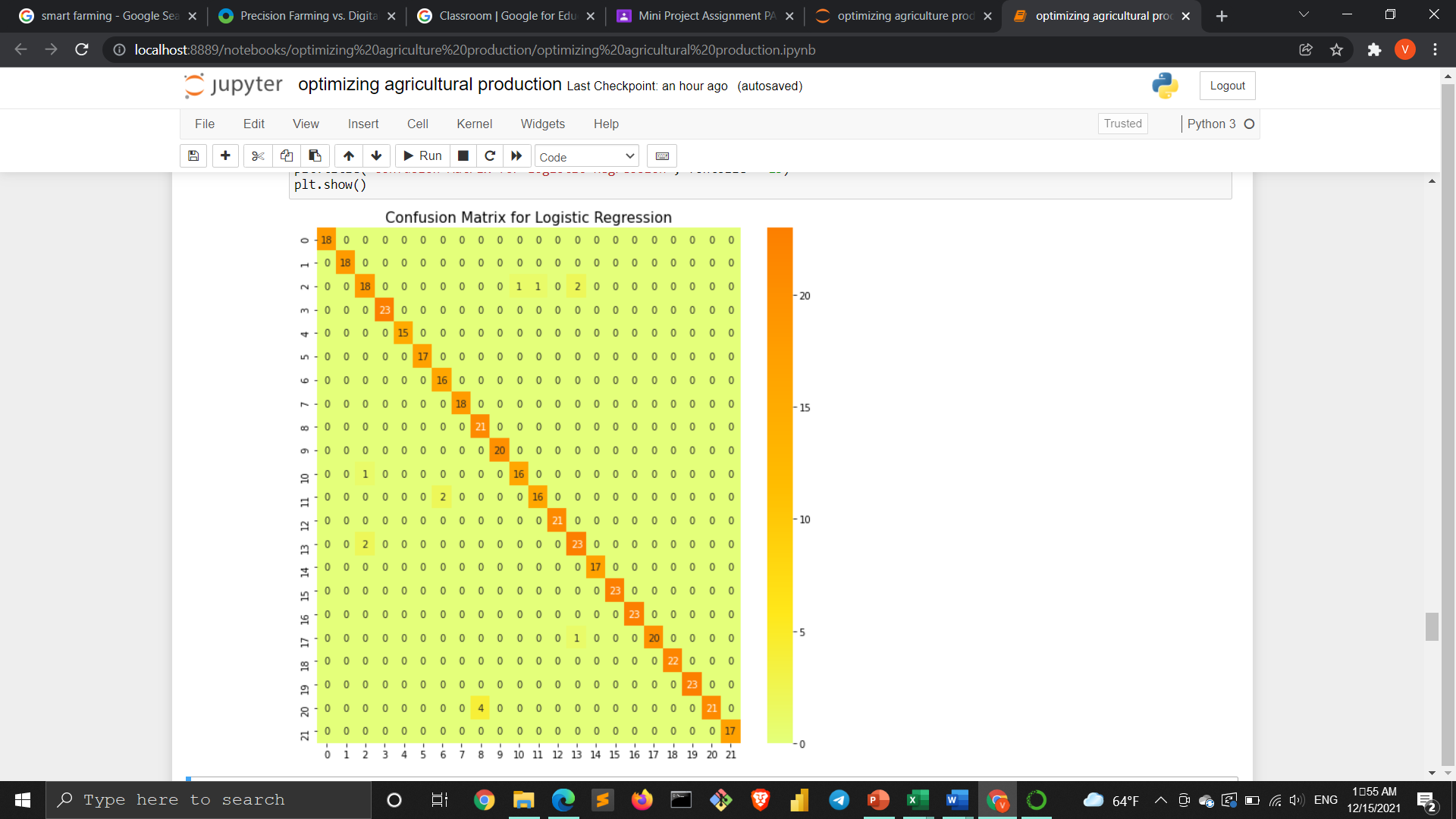
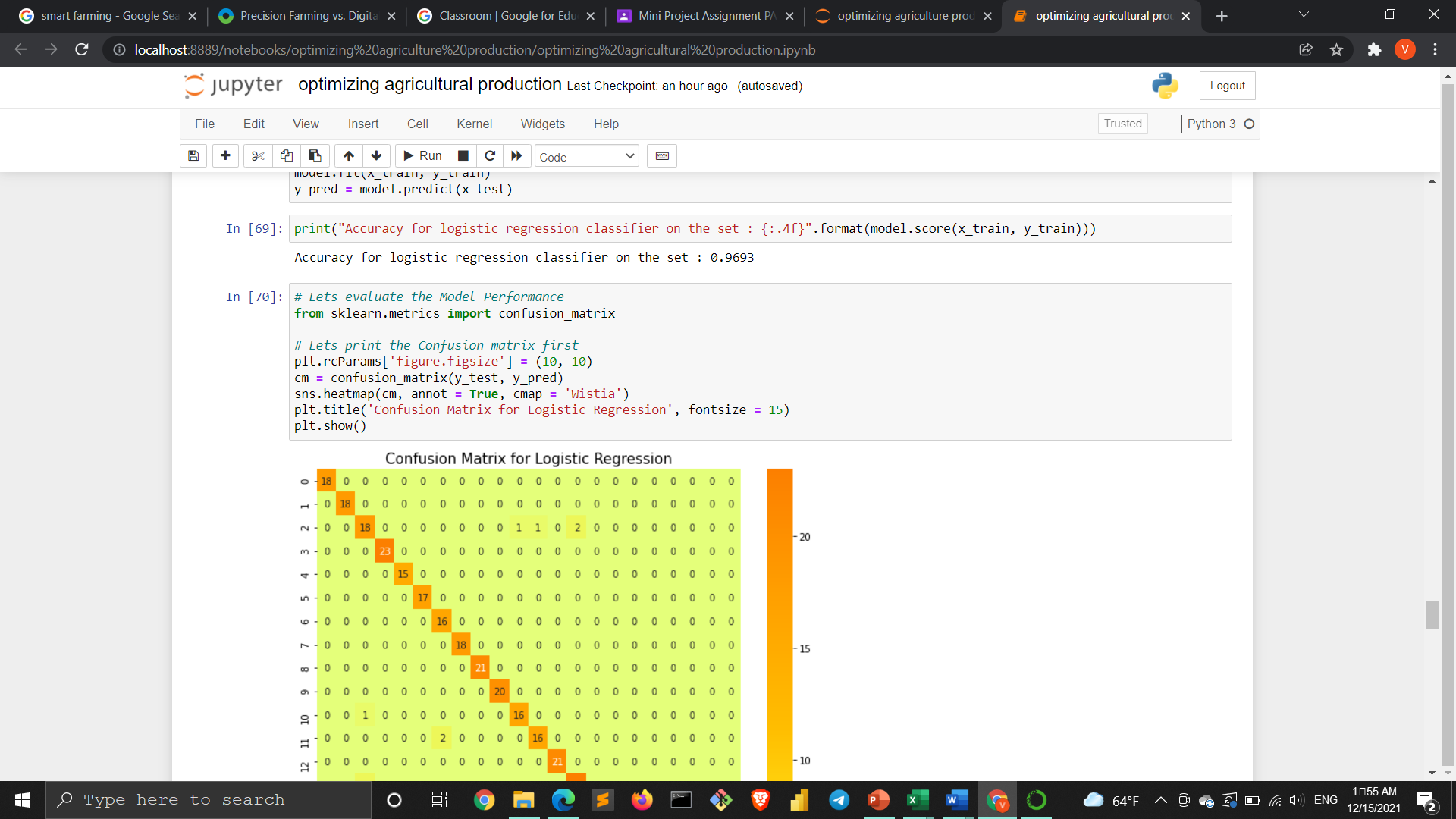
**Removing labels from the dataset :**

**And data divide into two parts training sets and testing set into 80:20 ratio**

**Applying logistic regression :**

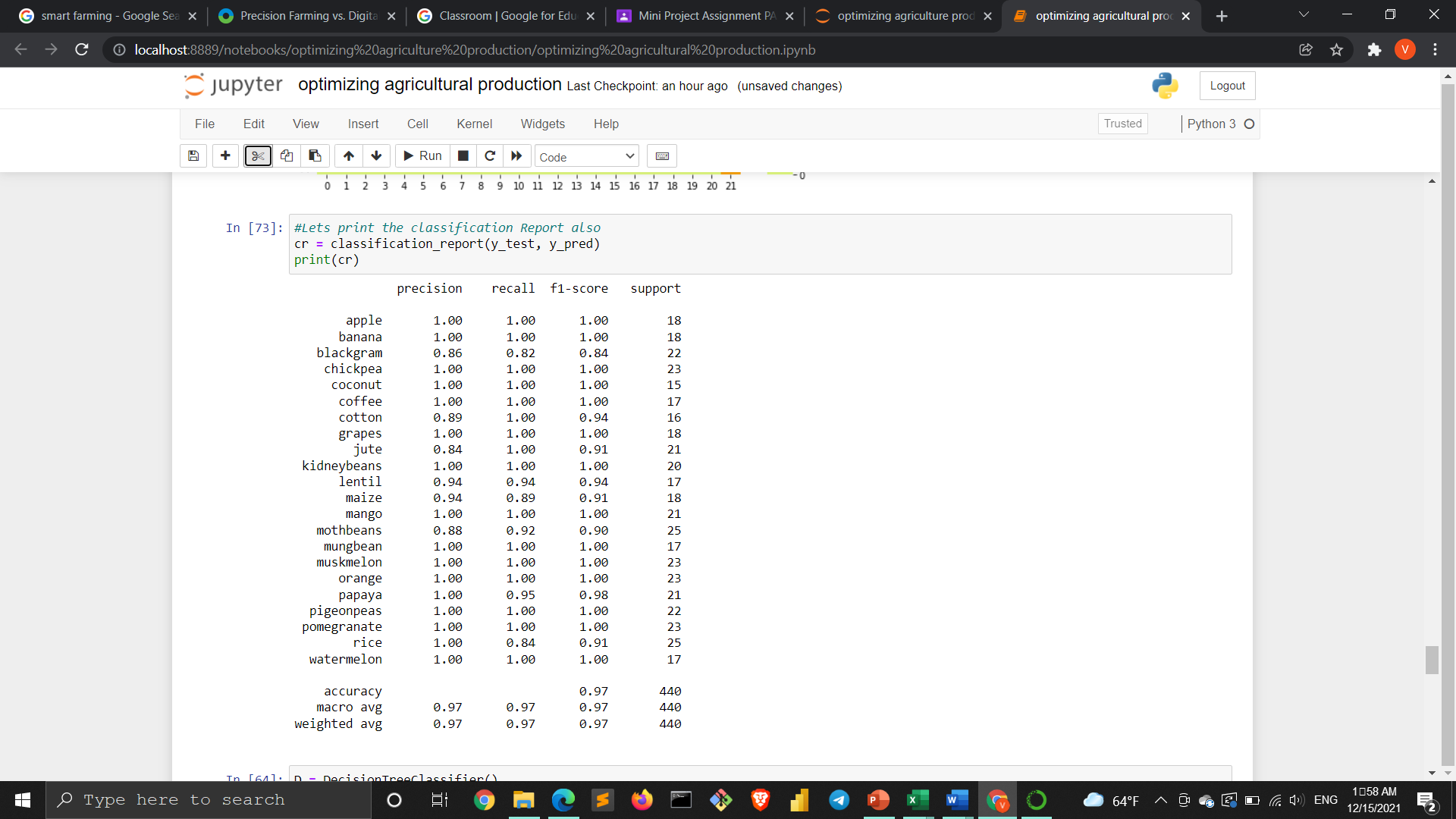


**Classification report:(for checking accuracy)**

****

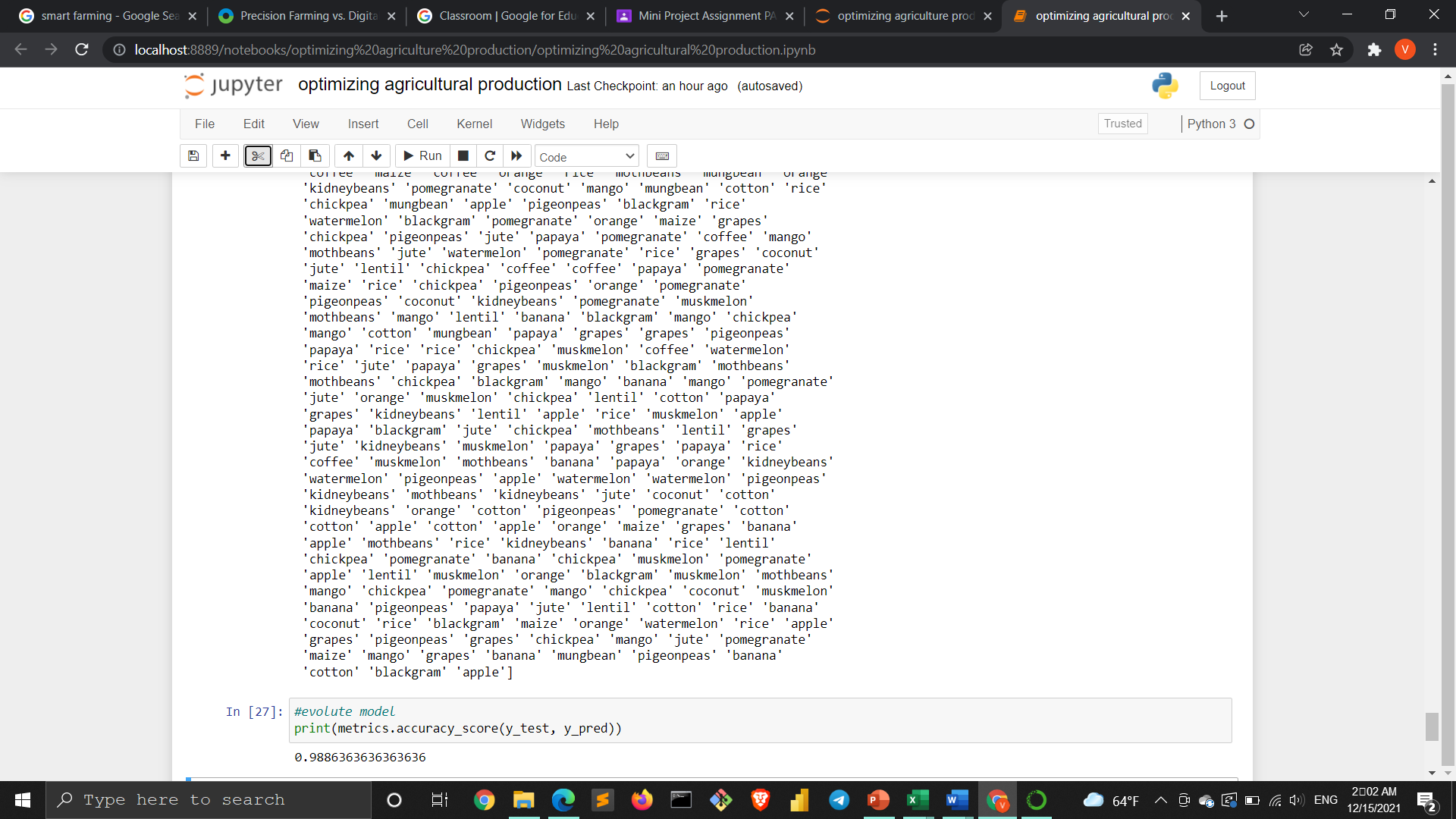
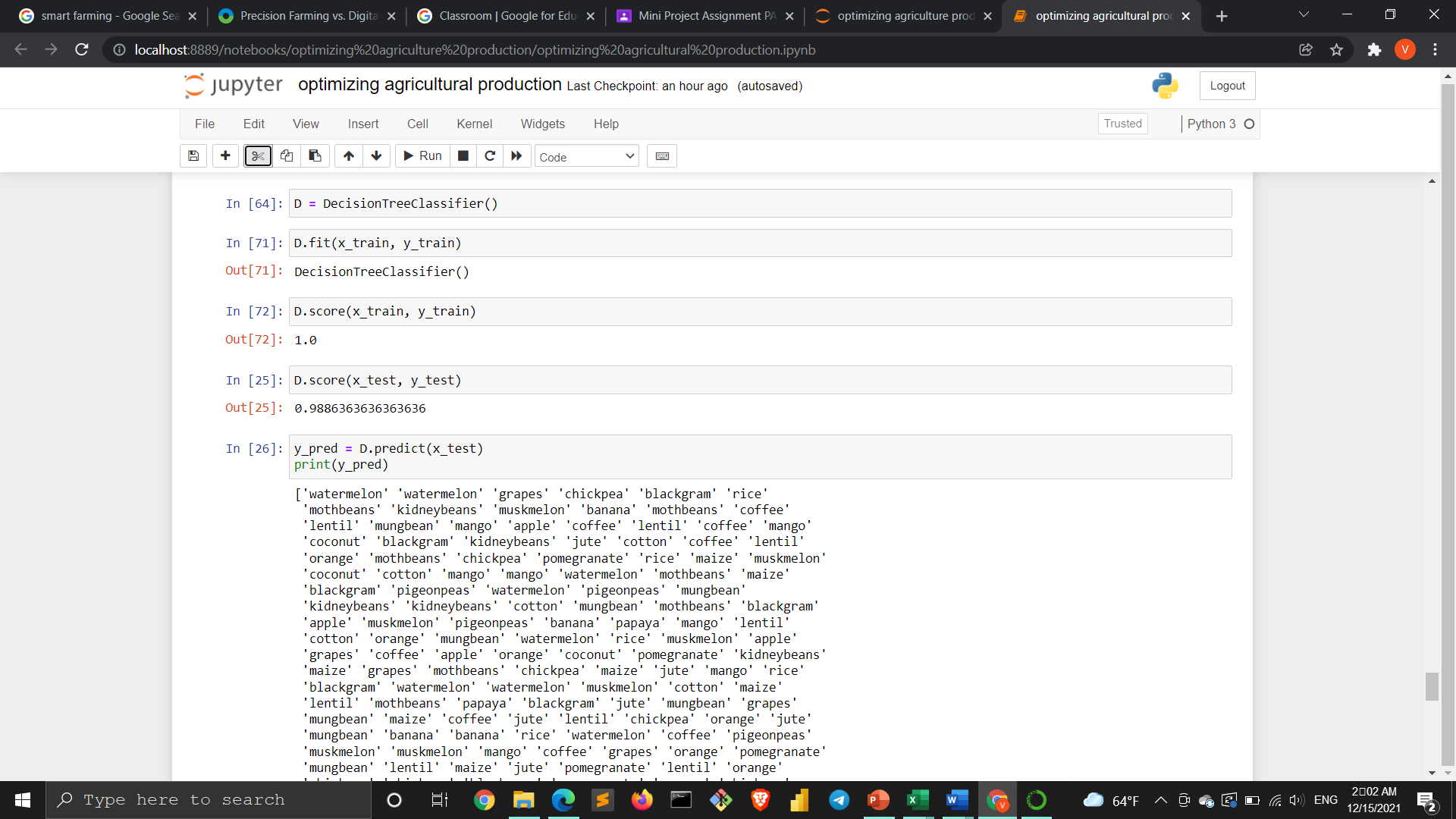
**We come to know precision and recall values from this**

**Report of this classification:**

****

**From the above report, we get precision and recall value for each crop is almost 1**

**Decision tree model:**

****

**The accuracy of this model is 0.9886**

**Checking the final result by taking rice values from dataset**

****

**Result and discussion:**

**Accuracy of a classification model and decision tree model**

|  |  |
| --- | --- |
|  |  |
|  |  |

|  |  |
| --- | --- |
| Precision recall f1-score support  apple 1.00 1.00 1.00 18  banana 1.00 1.00 1.00 18  blackgram 0.86 0.82 0.84 22  chickpea 1.00 1.00 1.00 23  coconut 1.00 1.00 1.00 15  coffee 1.00 1.00 1.00 17  cotton 0.89 1.00 0.94 16  grapes 1.00 1.00 1.00 18  jute 0.84 1.00 0.91 21  kidneybeans1.00 1.00 1.00 20  lentil 0.94 0.94 0.94 17  maize 0.94 0.89 0.91 18  mango 1.00 1.00 1.00 21  mothbeans 0.88 0.92 0.90 25  mungbean 1.00 1.00 1.00 17  muskmelon 1.00 1.00 1.00 23  orange 1.00 1.00 1.00 23  papaya 1.00 0.95 0.98 21  pigeonpeas1.00 1.00 1.00 22  pomegranate1.00 1.00 1.00 23  rice 1.00 0.84 0.91 25  watermelon 1.00 1.00 1.00 17  accuracy 0.97 440  macro avg 0.97 0.97 0.97 440  weighted avg 0.97 0.97 0.97 440 | Precision recall f1-score support  1.00 1.00 1.00 18  1.00 1.00 1.00 18  0.95 1.00 0.98 21  1.00 0.92 0.96 25  1.00 1.00 1.00 15  1.00 1.00 1.00 17  1.00 1.00 1.00 16  1.00 1.00 1.00 18  1.00 0.95 0.98 22  0.90 1.00 0.95 18  1.00 1.00 1.00 17  1.00 1.00 1.00 18  1.00 1.00 1.00 21  1.00 0.96 0.98 26  1.00 1.00 1.00 17  1.00 1.00 1.00 23  1.00 1.00 1.00 23  1.00 1.00 1.00 21  1.00 1.00 1.00 22  1.00 1.00 1.00 23  0.96 1.00 0.98 24  1.00 1.00 1.00 17  0.99 440  0.99 0.99 0.99 440  0.99 0.99 0.99 440 |

Conclusion: This will help them in improving their crop production both qualitatively and quantitatively

This will also help them to maintain the quality and nutritional contents of the soil.

The system is to plan and grow a recommendation model to create the recommendations for crops relied on geological and climatic attributes using machine learning procedures

Future scope

As concerning future scores when the farmers sow a particular crop, there might face some issues or diseases in the crop before harvesting. In that case, they can upload photographs of the crop and the soil report. Then the AI model can identify the problems and provide them with probable solutions

We can also provide IOT solutions through APIs or virtual agents which can help the farmers connect with raw material dealers, who can provide them with the materials required for instance seeds and fertilizers according to the crop recommended to them by the model

Advantages from optimizing crop production

* This will help farmers to earn more profit
* It gives the knowledge and idea about the soil fertility
* This model could predict the crops which we can get on a particular farm just by knowing the attributes or parameters in the dataset
* This model could predict crops season wize or according to climatic conditions
* It makes prediction very easy and saves the effort of farmer in other direction

Research paper: 