**NAME:** SYED MOHTASHIM ALI

**SEAT:** B17101113

**SEC: B (morning)**

**BSCS-515 ARTIFICIAL INTELLIGENCE**

**FINAL LAB ASSIGNMENT -1.**

**FINAL LAB REPORT**

GUIDE TO MY CODE:

In my project you will overcome a functional approach i.e. I feel comfortable enclosing my solutions on a function based approach. Most of the code will not be hard-coded.

>WARM-UP EXERCISE:

In this exercise, firstly I created a pandas series using pd.Series() function that is a built-in pandas function for Series (Columns ).

The content of this series will be provided by a function and some open code;

alphabets ( n ) here ‘n’ is the number of alphabets we want.

I wrote a code to generate ‘10’ prime numbers.

Secondly, I re-created a pandas data-frame from the given diagram.To create a data-frame I used pandas function pd.DataFrame() . Set the index using data-frame.set\_index() function.

**(THIS ENDS UP THE WARM-UP EXERCISE)**

**FINAL-LAB-1 TASKS:**

1. Downloaded the data-set from UCI website in a (.csv) format.
2. Read the .csv file from pandas function pandas.read\_csv().
3. Identified the length of the data or (number of rows) by len(data-frame)
4. Identified the shape of the data-frame form the data-frame.shape function
5. Identified the column names by data-frame.columns
6. Identified the different species by data-frame.unique() function.
7. Performed some data-cleaning.
8. Created two new columns ’ petal-ratio ’ and ‘ sepal-ratio ’ in my data-frame which consist of the ratio between ‘petal-length ’ and ‘ petal-width ’.
9. I saved the new data-frame with added columns as ‘ iris\_corrected.csv ’ using a pandas function new-data-frame.to\_csv().
10. This is the data-visualization task in which I took the following steps;
11. Created a histogram using pandas data-frame.hist()
12. Created a Function graph( ) that generates sea-born scatter-plots.
13. Created a scatter-matrix using using pandas.plotting.scatter\_matrix( ).
14. Used train\_test\_split from sci-kit learn.
15. Used Standard-Scalar for feature-scaling.
16. The algorithms I used are ‘ K-nearest neighbour ’ and ‘ Support vector clustering(SVC)’.
17. I used accuracy-score form sci-kit learn to measure the percentage of correctness of my model.

DATA-CLEANING:

Data cleaning is one of the most tricky (sometimes) and important machine learning task.Without it any machine learning algorithm will not perform the it was suppose to.

Here , in this exercise , we were ask to replace the data-values at index 34 and 37(35 and 38)

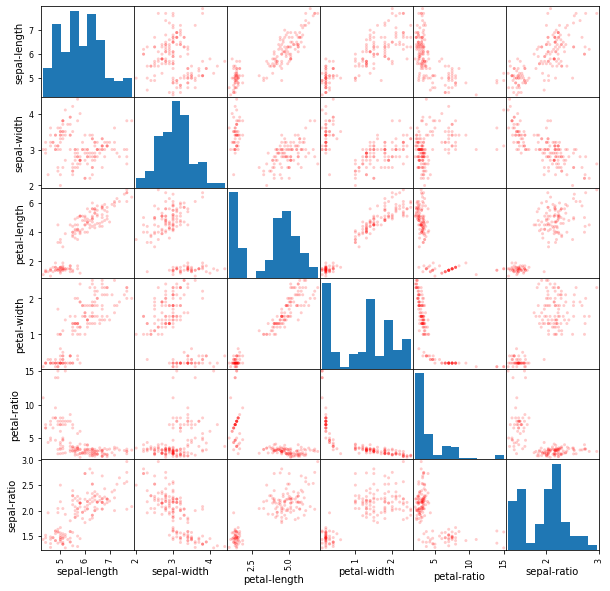
respectively. I used data-frame.iloc[ index ] function to confirm the data-points and replaced the required improper data-point by the correct one using data-frame.at [index , column]

Function. This ended the data-cleaning process.

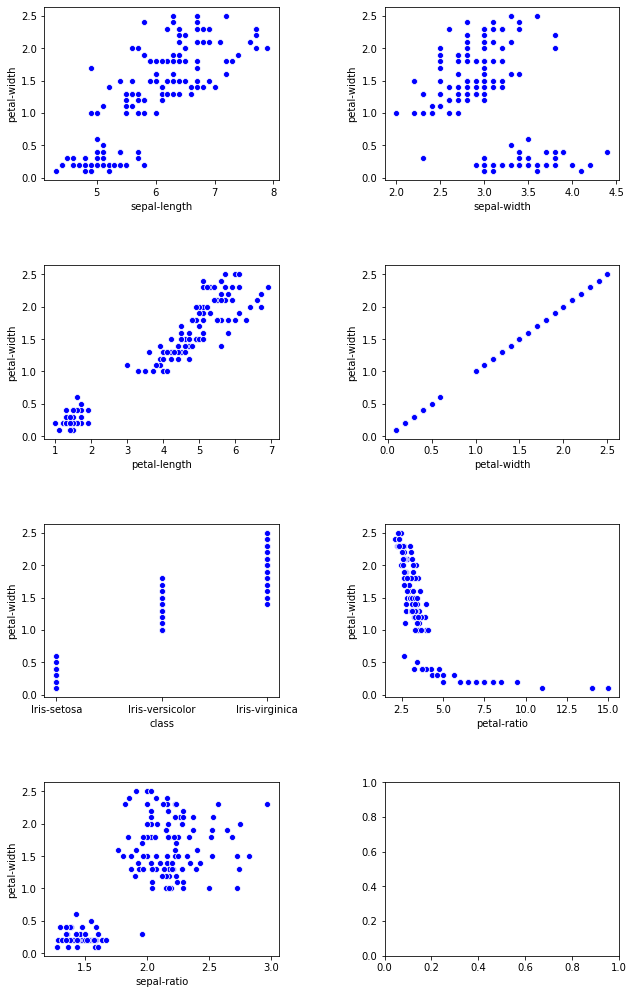
FEATURE-SCALING:

Another important aspect in machine learning (for numeric-data) is Feature-scaling.It is a technique used frequently to normalize the data and it’s rightfully used during the data preprocessing step.It normalizes the data within a particular range.It helps the algorithm to speed up the calculations. I used Standard-Scalar from sci-kit learn for this purpose.

DATA-VISUALIZATION:



**(Scatter-Matrix)**

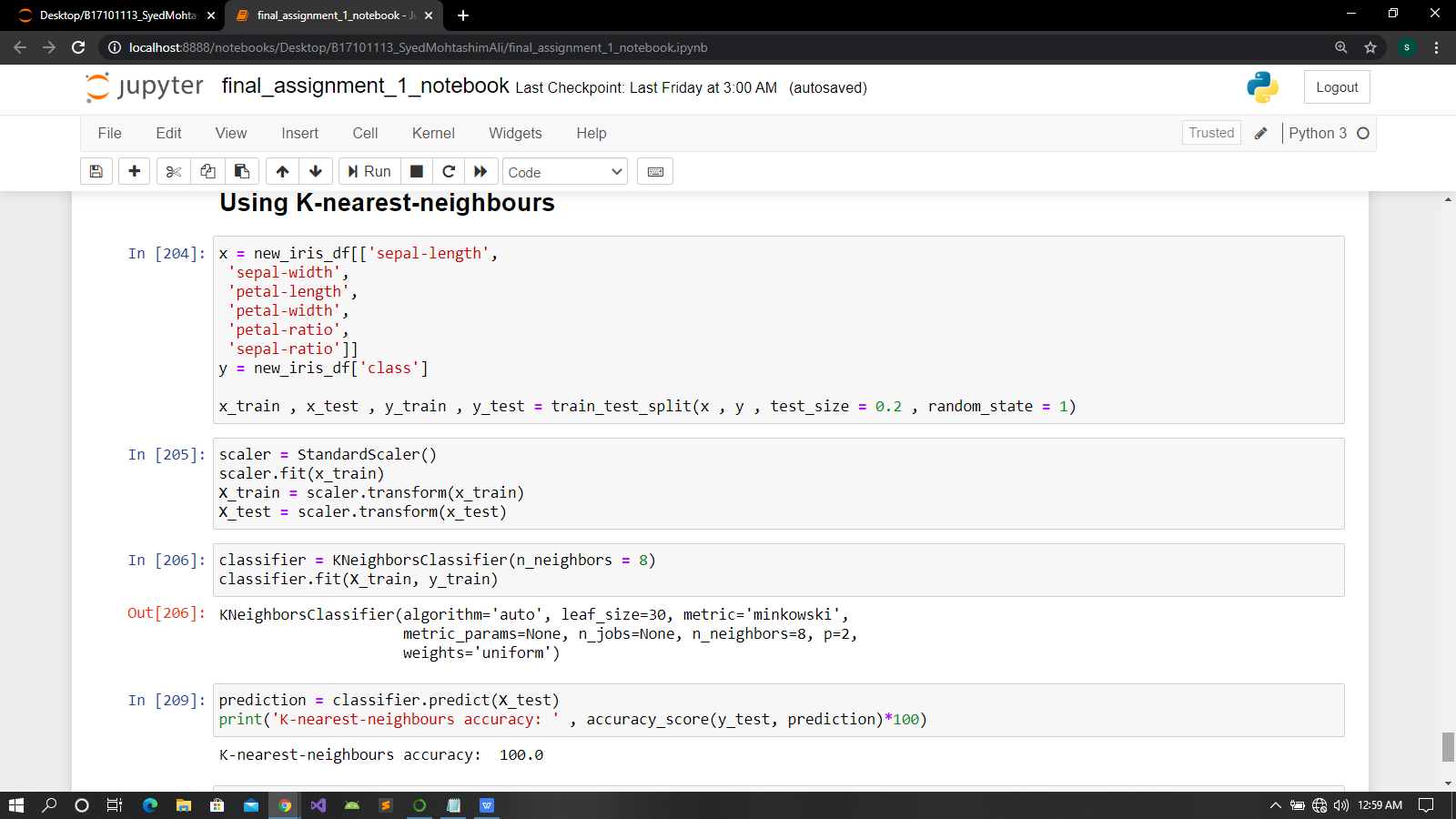


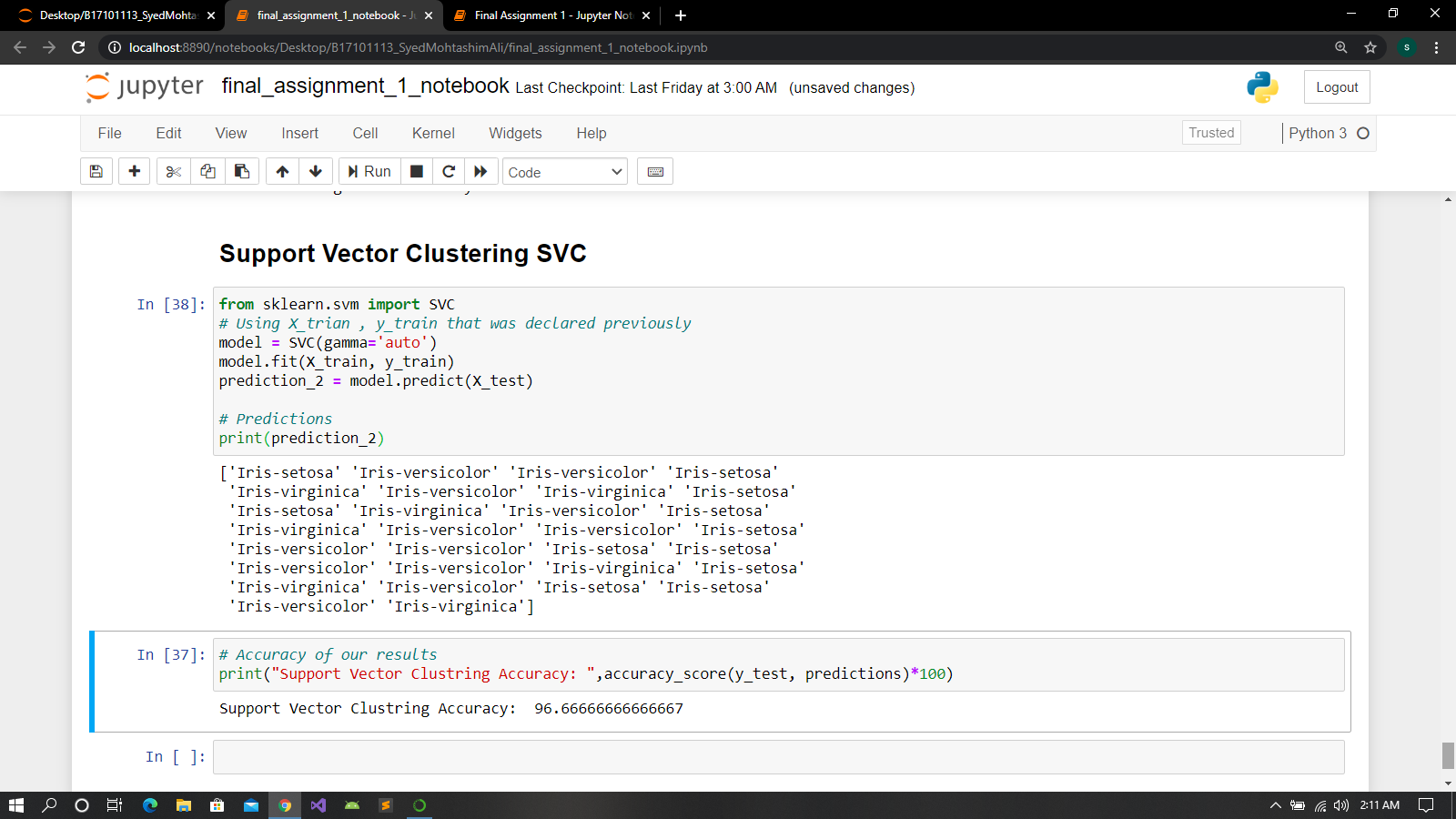
**(Some scatter-plots)**

CODE-SNIPPETS:

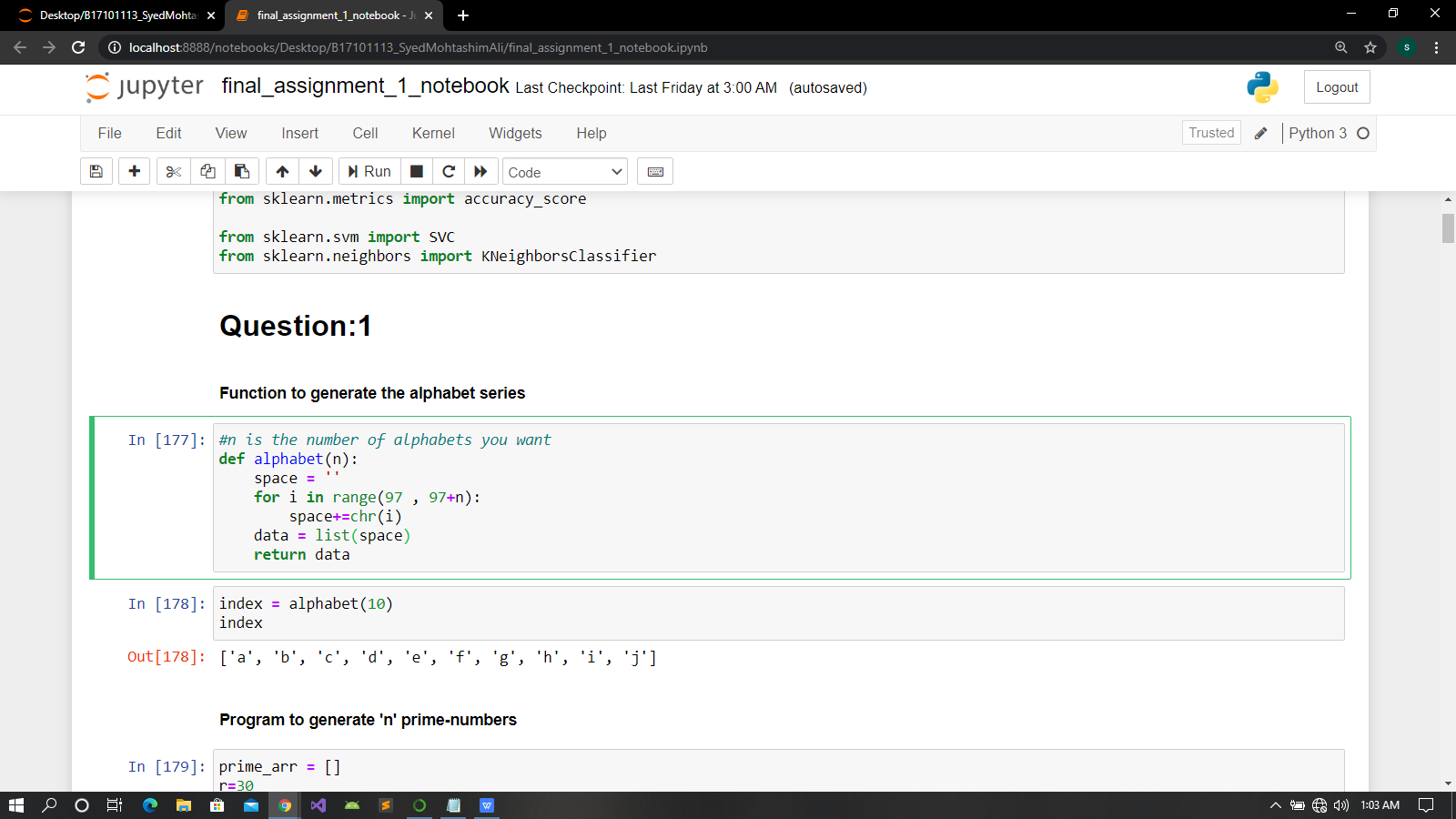
Some important code snippets are as follows

1. K-NEAREST NEIGHBOUR and S.V.C implementation

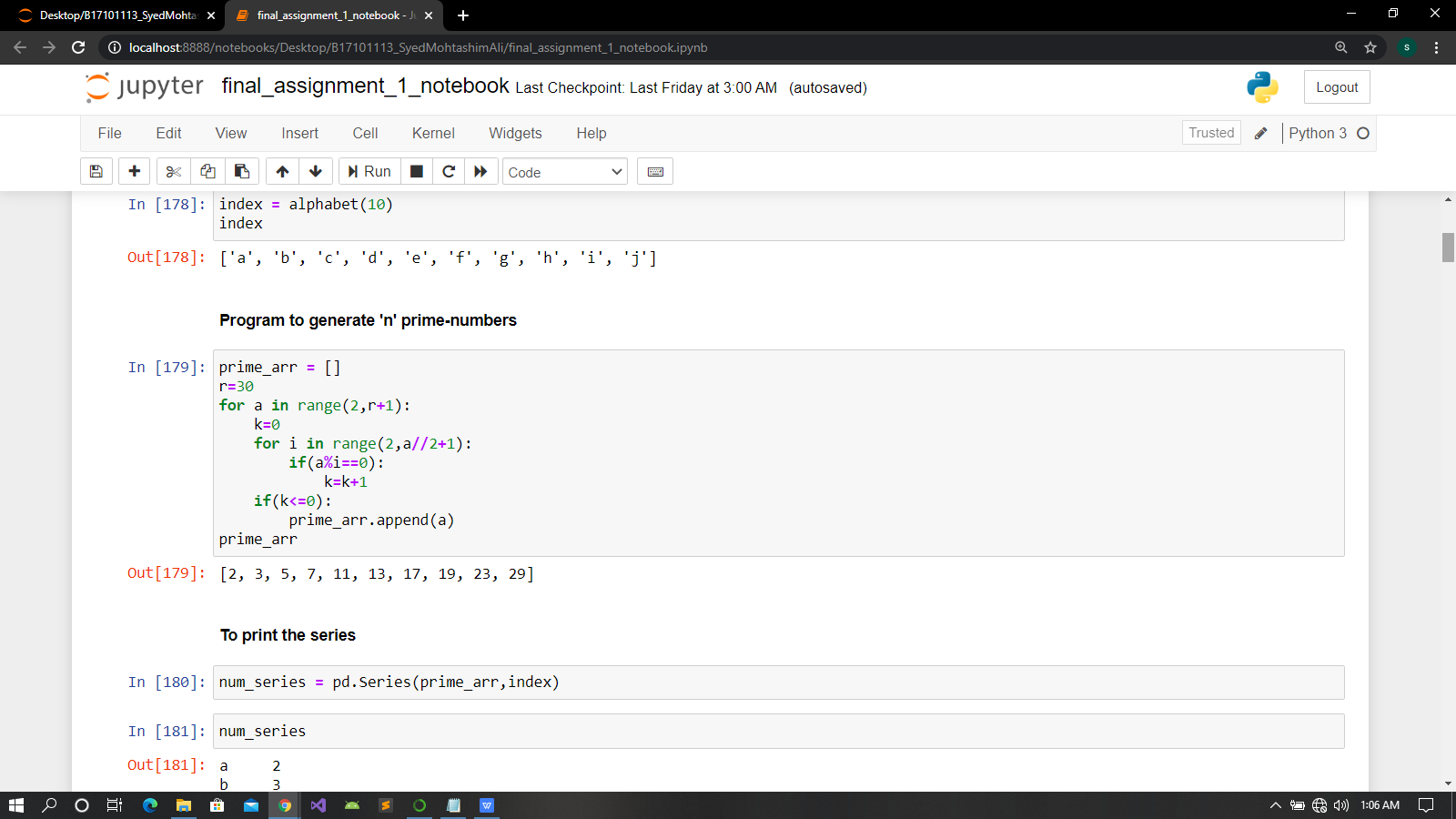




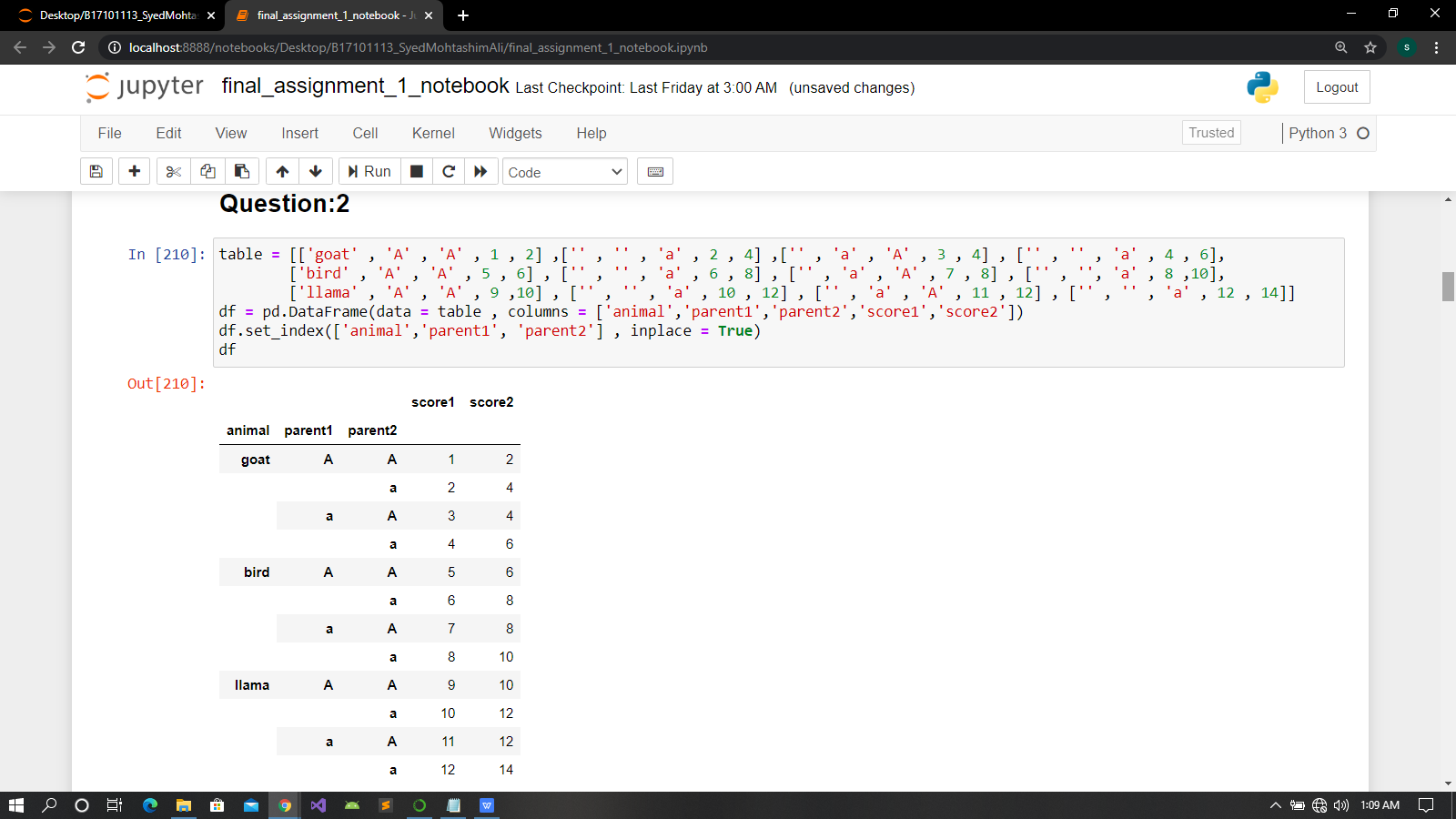
1. Function that generates alphabets



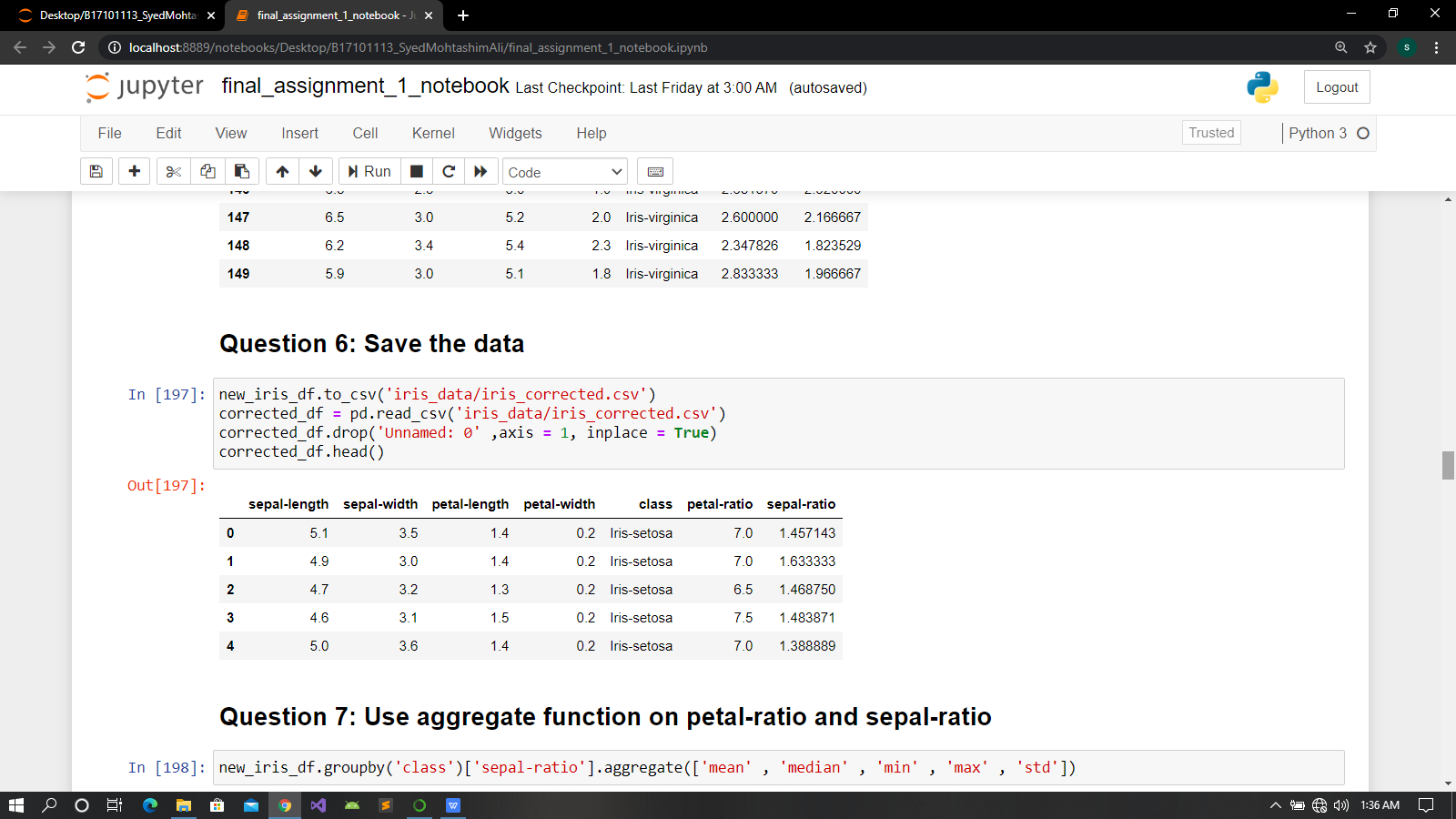
1. Program to generate 10 prime-numbers



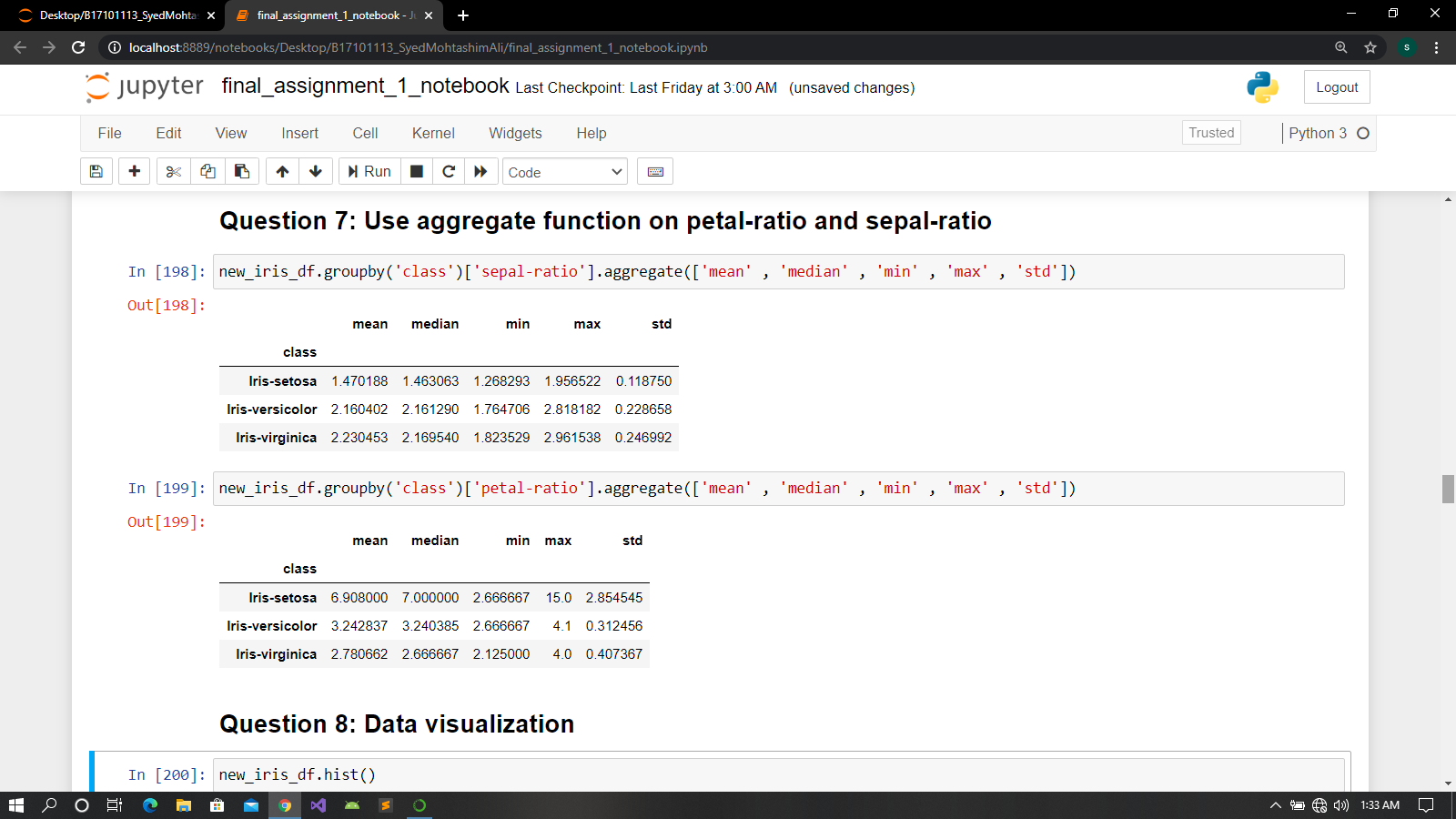
1. To write code for a series



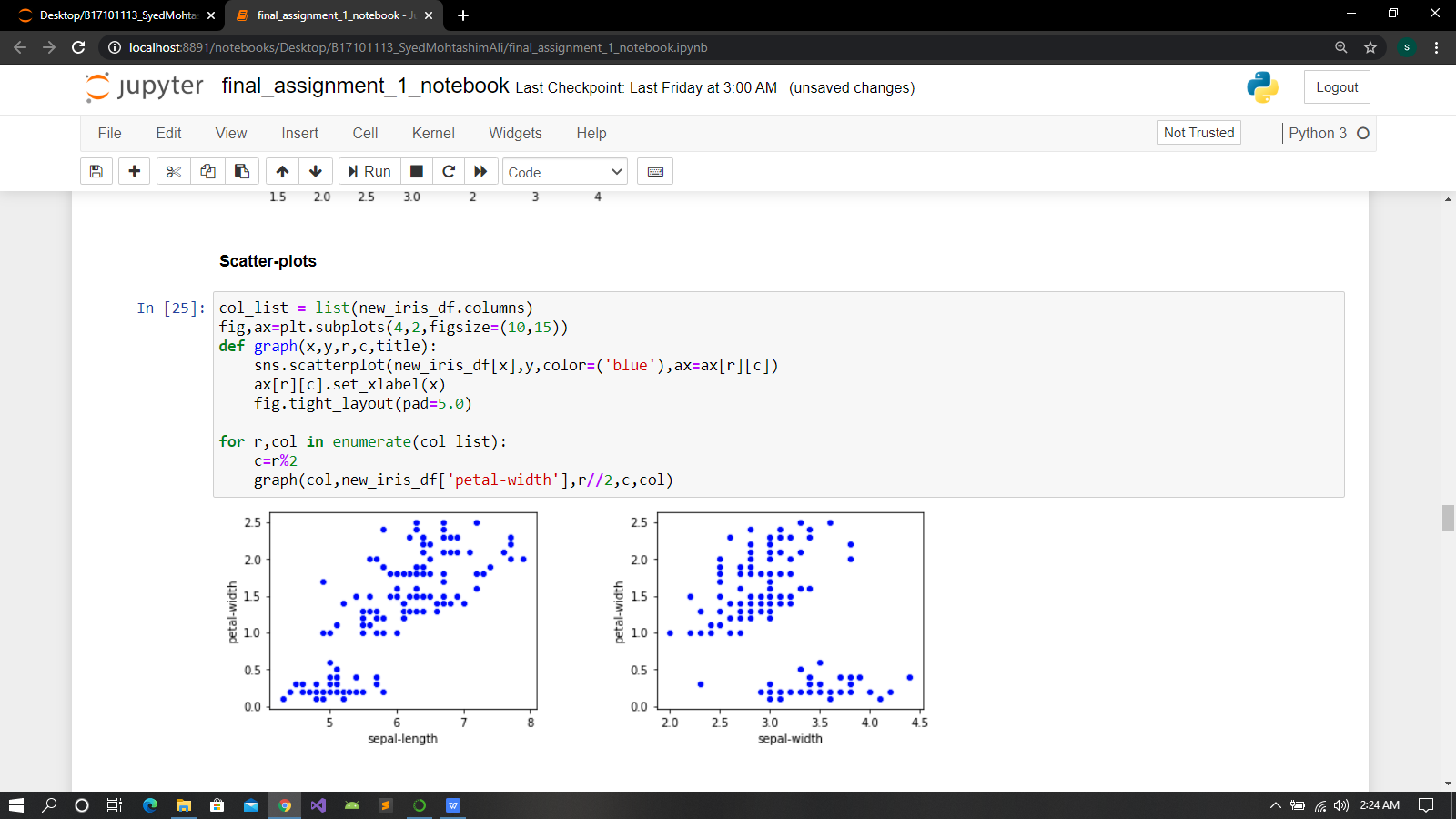
1. save the data



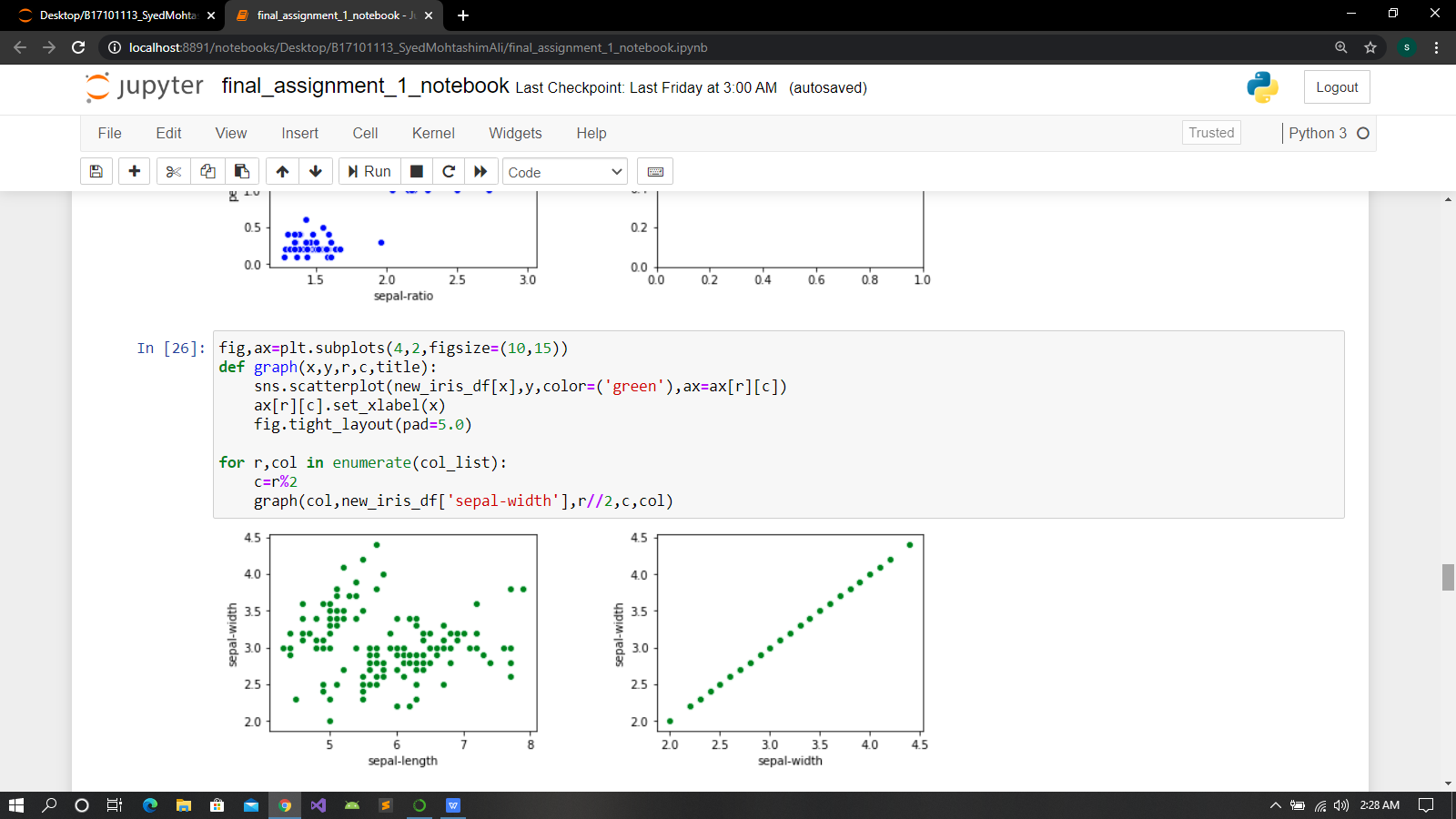
vi)To use aggregate function



1. Graph function I used to generate separate scatter plots

A)

(I took these function from a kaggle notebook while I was working on a closed kaggle competition “House Price Prediction using Advanced regressions techniques ”.)

b)

**REFRENCES:**

Here I will list all the links and a book I found useful while completing this assignment and some of my side-projects.

1. <https://stackoverflow.com/questions/31569384/set-value-for-particular-cell-in-pandas-dataframe-with-iloc>
2. <https://www.geeksforgeeks.org/python-pandas-dataframe-aggregate/>
3. <https://www.marsja.se/pandas-scatter-matrix-pair-plot/>
4. <https://www.tutorialspoint.com/machine_learning_with_python/machine_learning_with_python_knn_algorithm_finding_nearest_neighbors.htm>

**Books I read**:

1. 2-Aurélien-Géron-Hands-On-Machine-Learning-with-Scikit-Learn-Keras-and-Tensorflow\_- Concepts-Tools-and-Techniques-to-Build-Intelligent-Systems-O’Reilly-Media-2019.
2. Python-for-Data-Analysis.
3. learning-the-pandas-library-python-tools-for-data-munging-analysis-and-visual.
4. A Smarter Way to Learn Python.