Name: Allan Rodrigues

Class: TE IT A

Roll no: 59

Pid:191104

St. Francis Institute of Technology, Mumbai-400 103

**Department of Information Technology**

A.Y. 2021-2022

Class: TE-ITA/B, Semester: VI

Subject: **Data Science Lab**

**Experiment – 3: To implement Data Modelling.**

1. **Aim:** To implement Data Modelling 
2. **Objectives:** After study of this experiment, the student will be able to

* Understand how data to be re-scaled
* Understand how data partitioning in done using sklearn

1. **Outcomes:** After study of this experiment, the student will be able to

* Understand data rescaling and data partitioning using sklearn

1. **Prerequisite:** Fundamentals of Python Programming and Database Management System.
2. **Requirements:** Python Installation,Personal Computer, Windows operating system, Internet Connection, Microsoft Word.
3. **Pre-Experiment Exercise:**

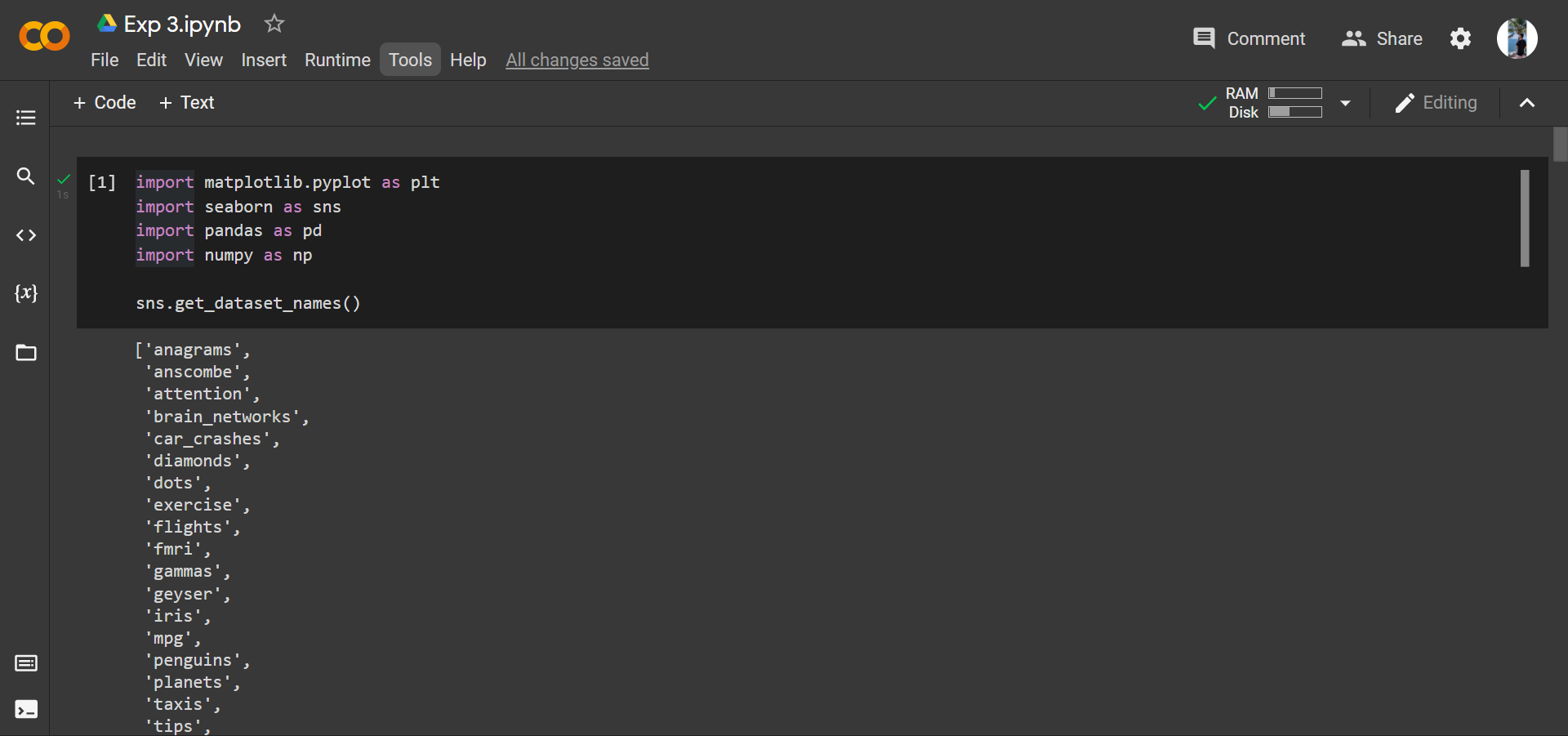
**Brief Theory:**

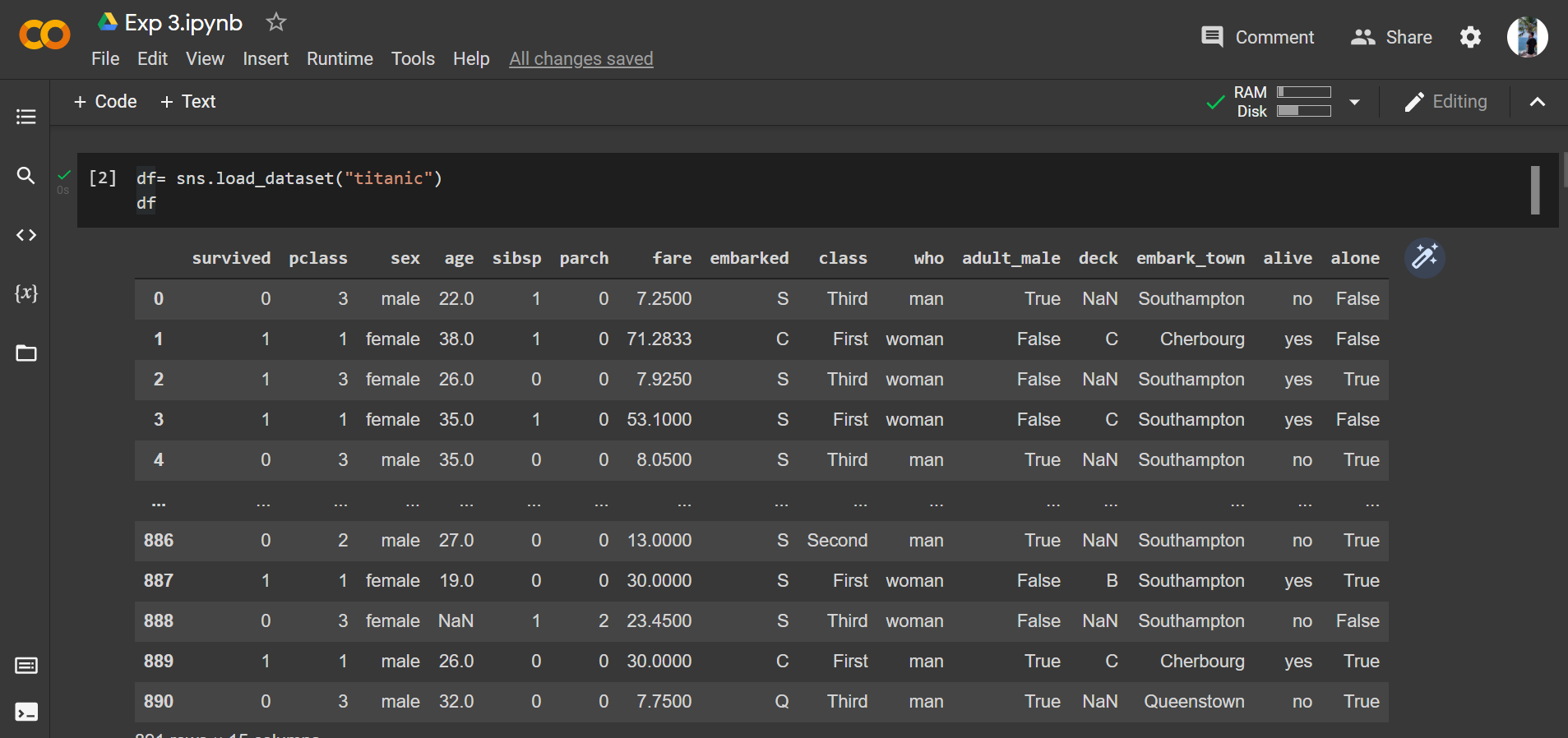
* Concept of sklearn package for machine learning**.**

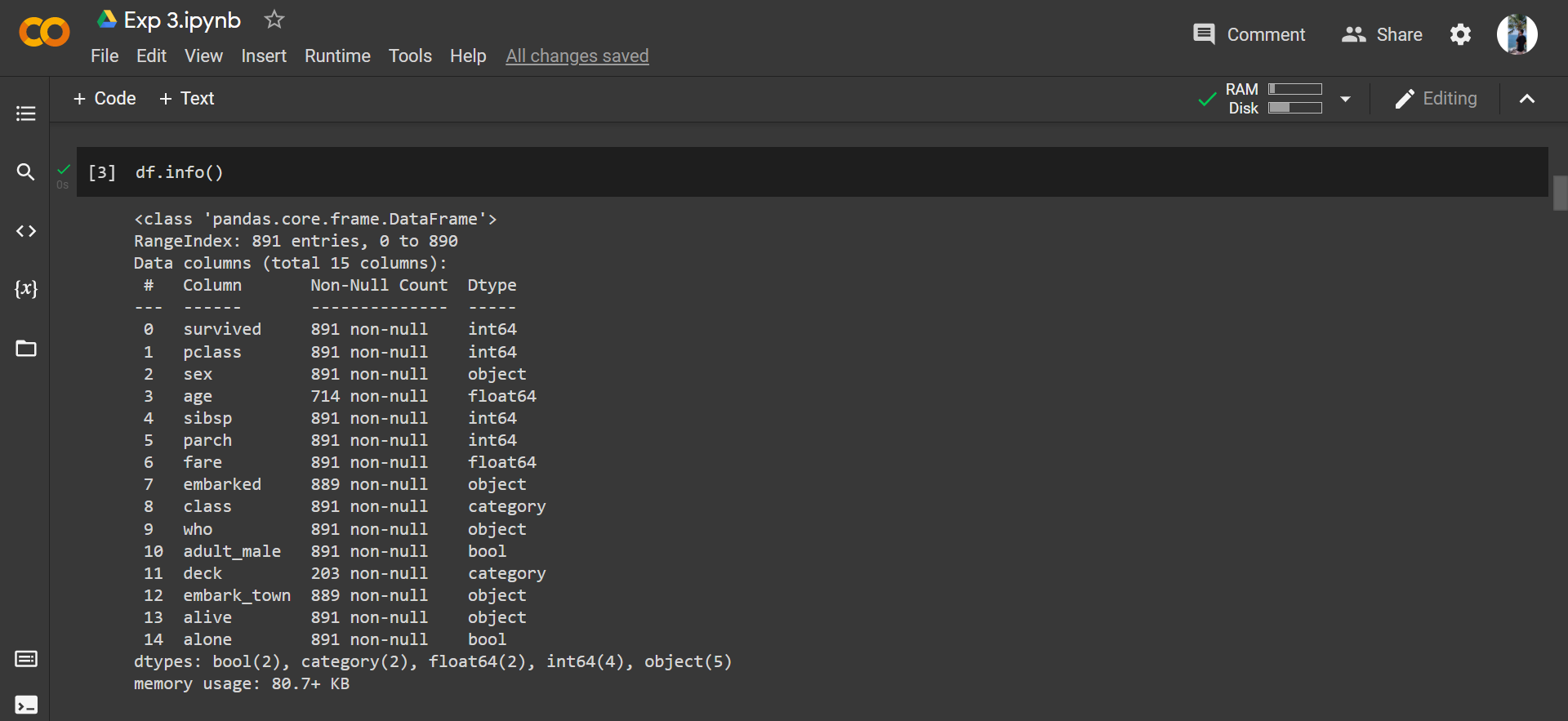
Scikit-learn provides a range of supervised and unsupervised learning algorithms via a consistent interface in Python. Extensions or modules for SciPy care conventionally named SciKits. As such, the module provides learning algorithms and is named scikit-learn. The library is focused on modeling data. It is not focused on loading, manipulating and summarizing data. For these features, refer to NumPy and Pandas.

**Laboratory Exercise**

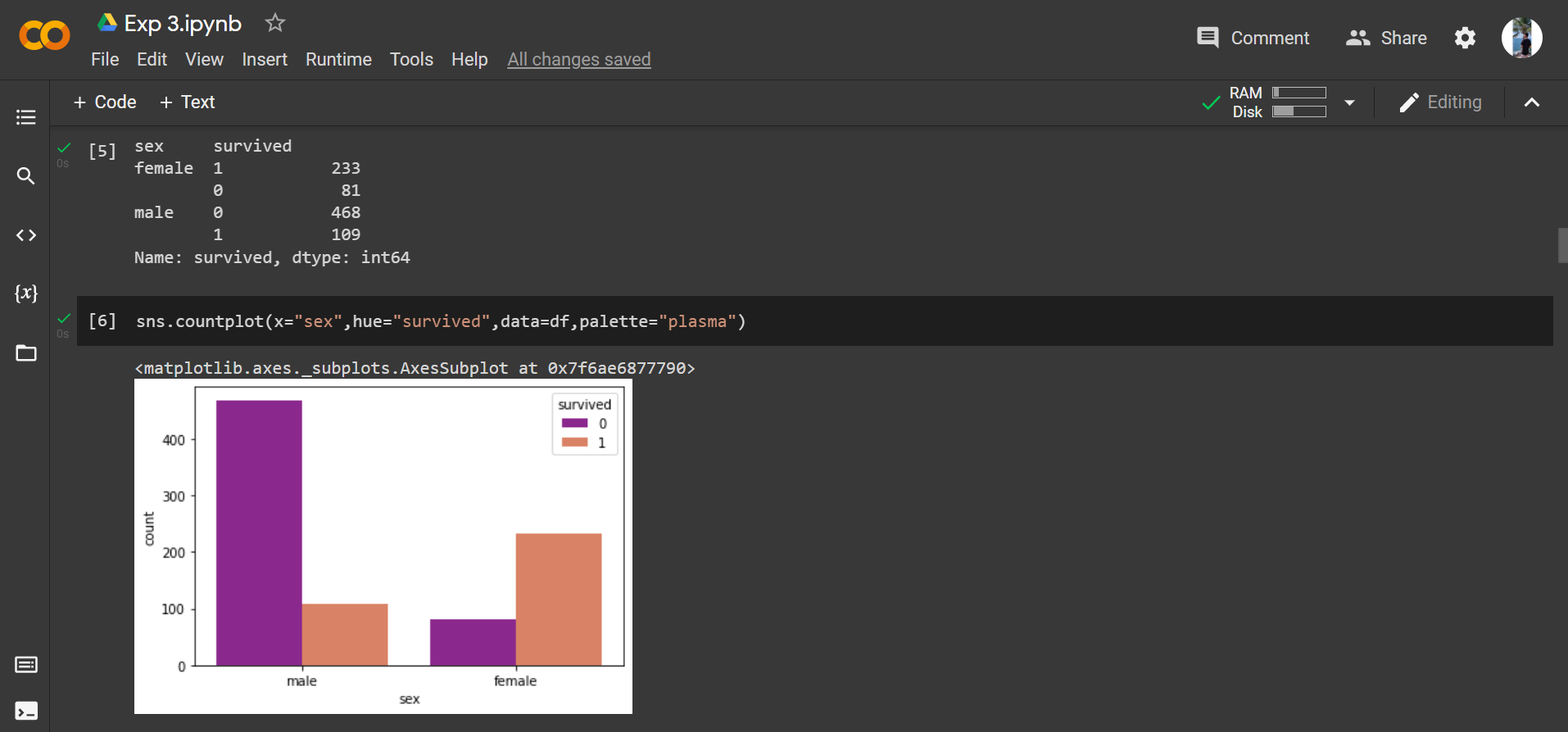
* + - 1. **Procedure:** (the sheet for commands in attached with the file)
      2. Paste Screenshots of above commands.

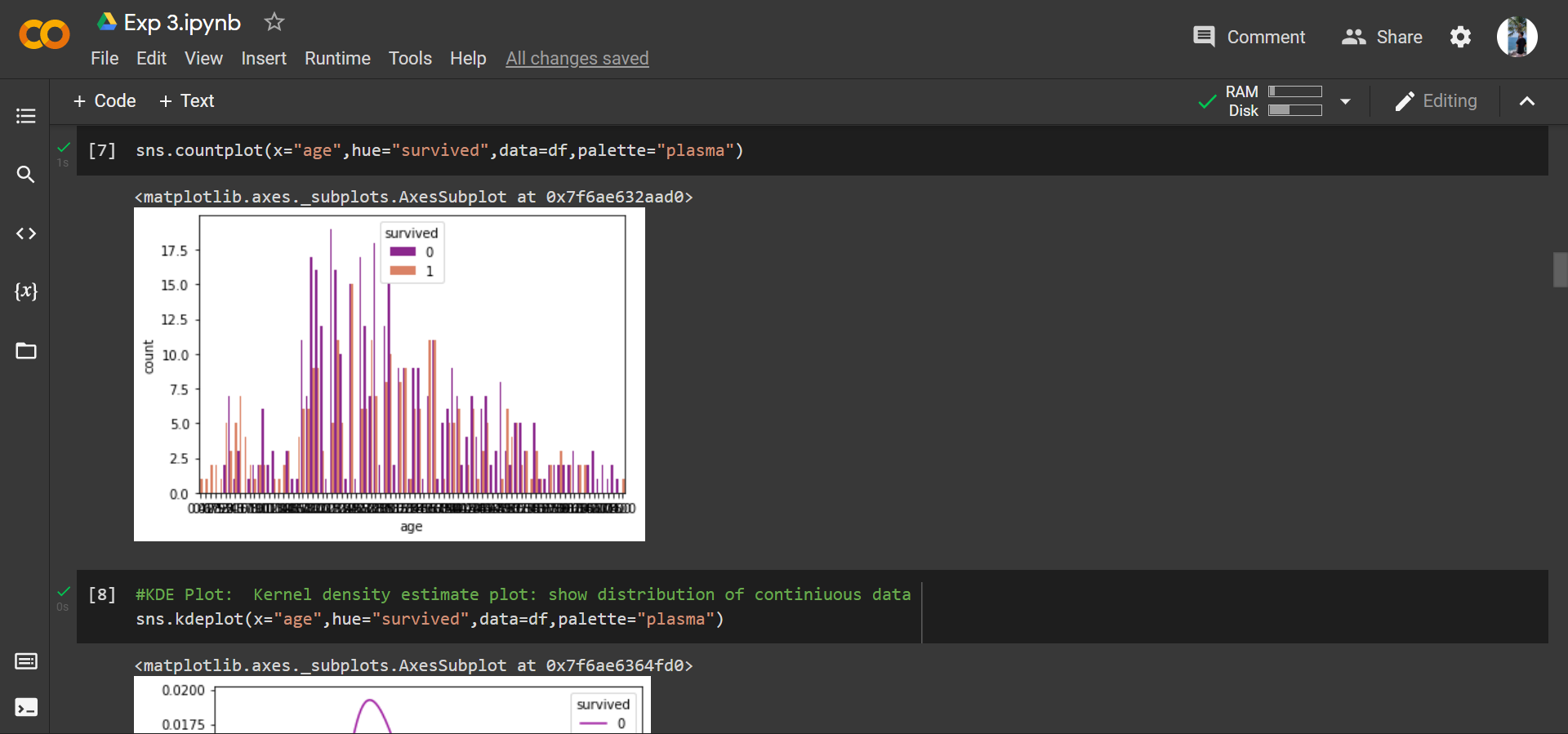


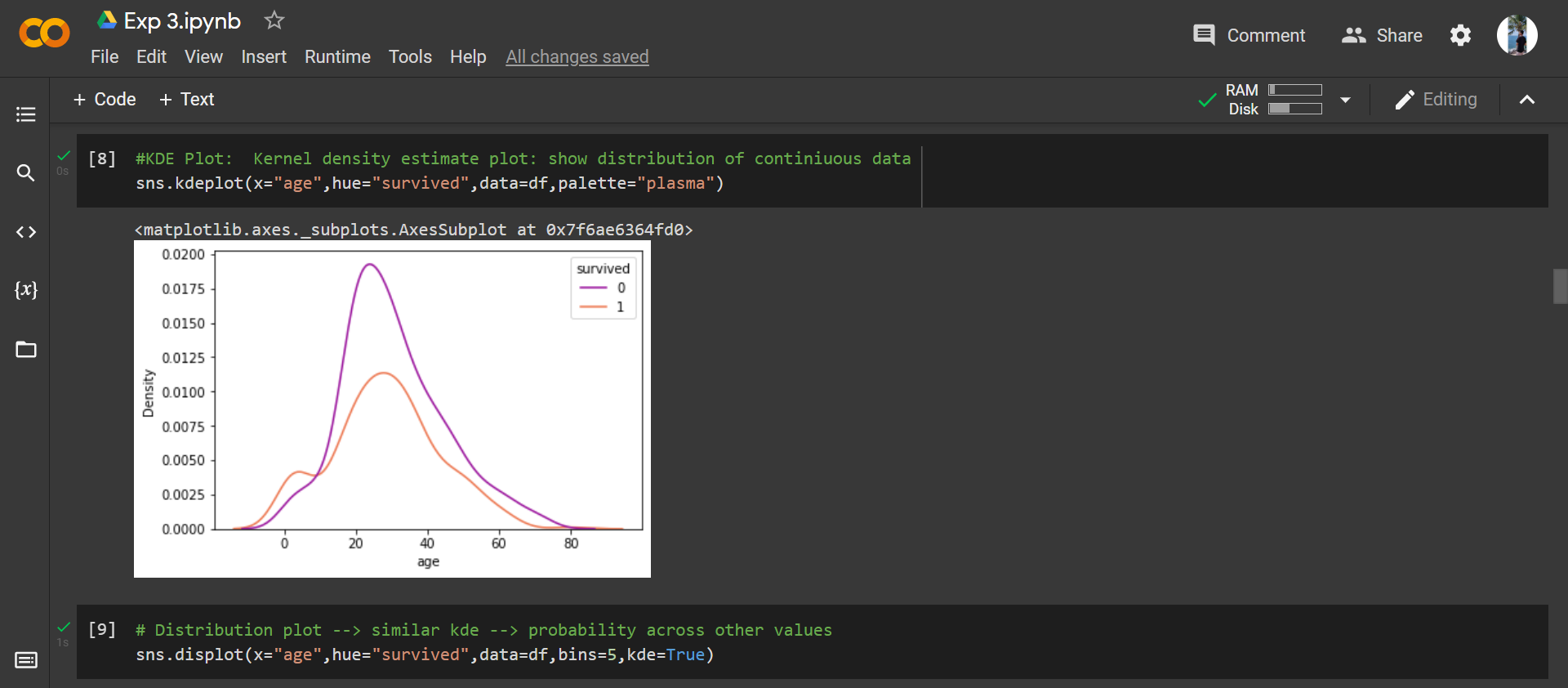


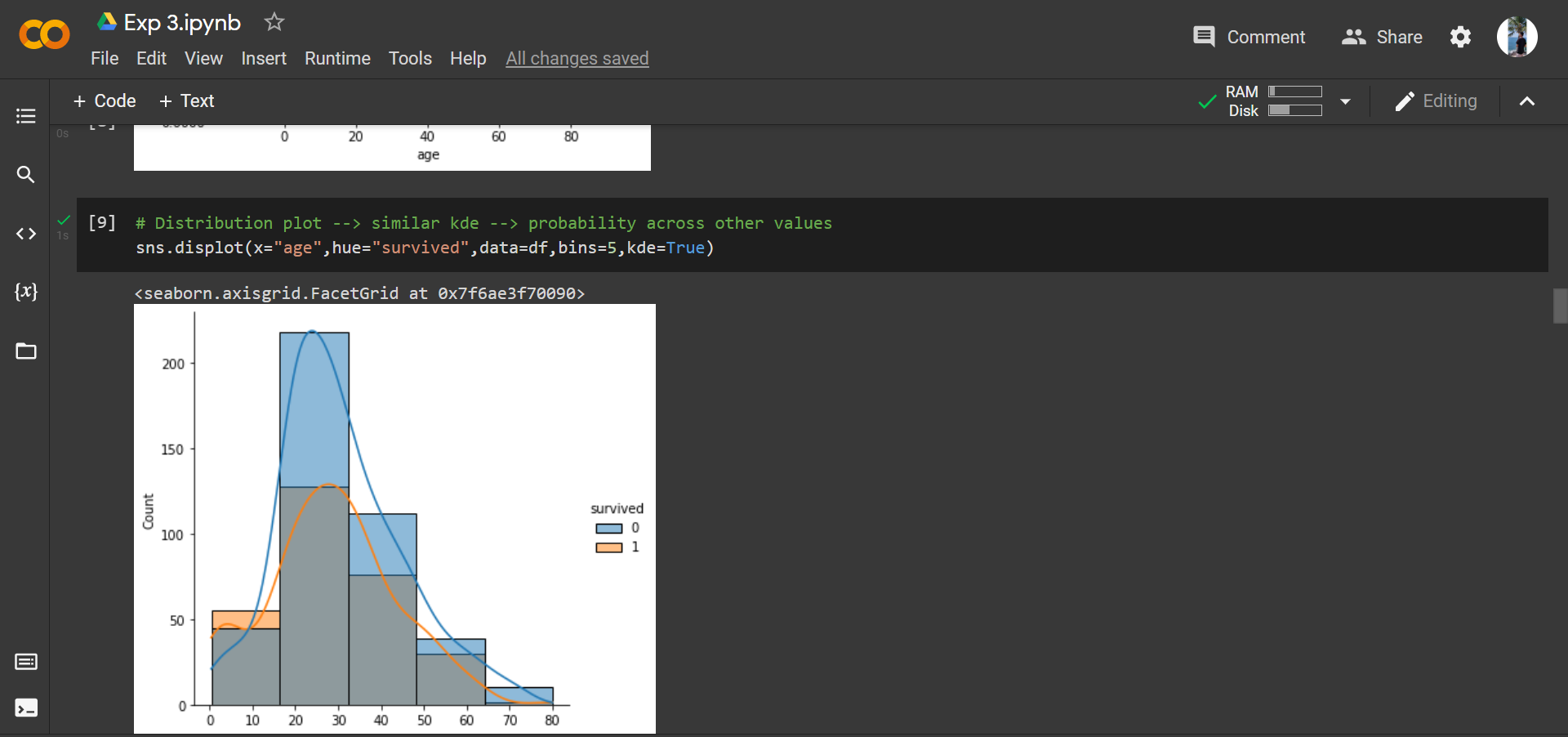


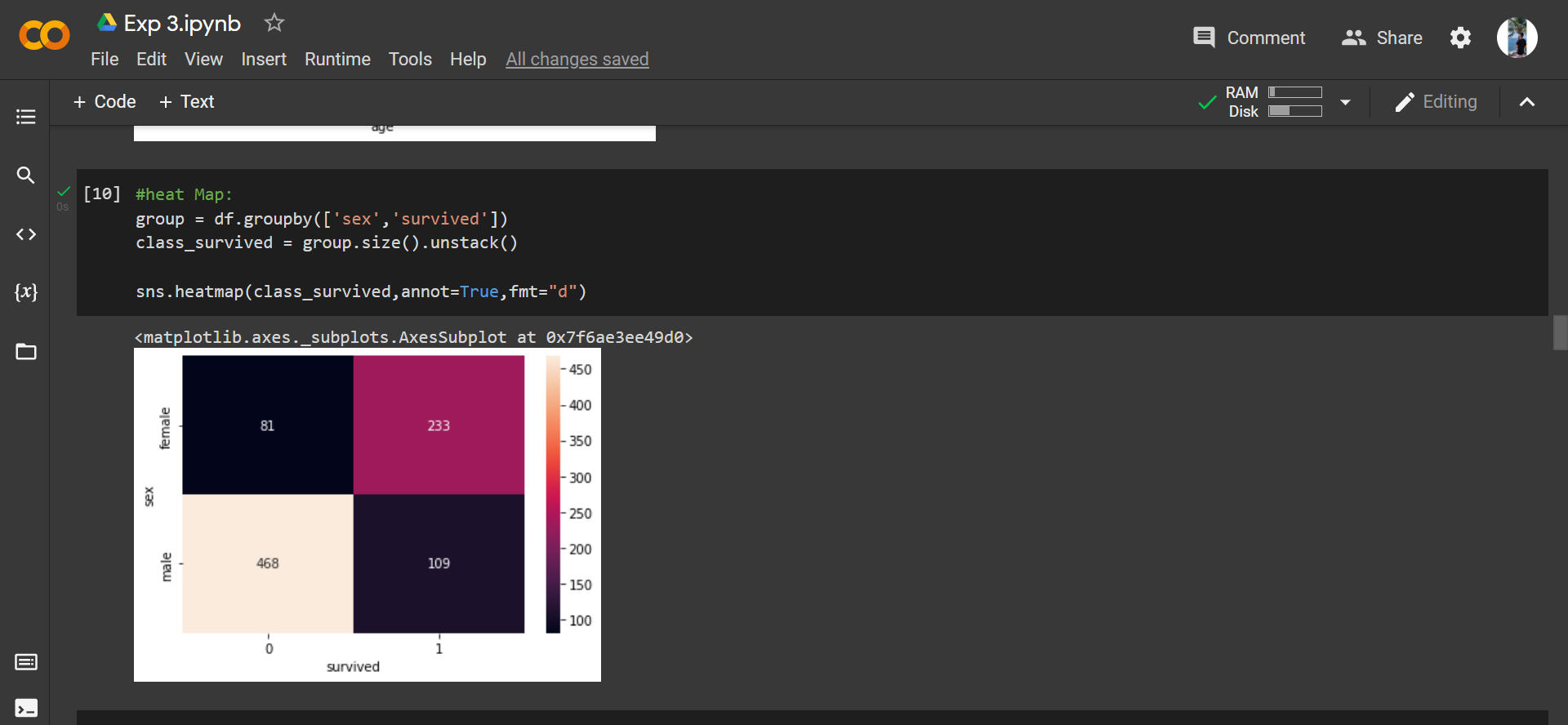


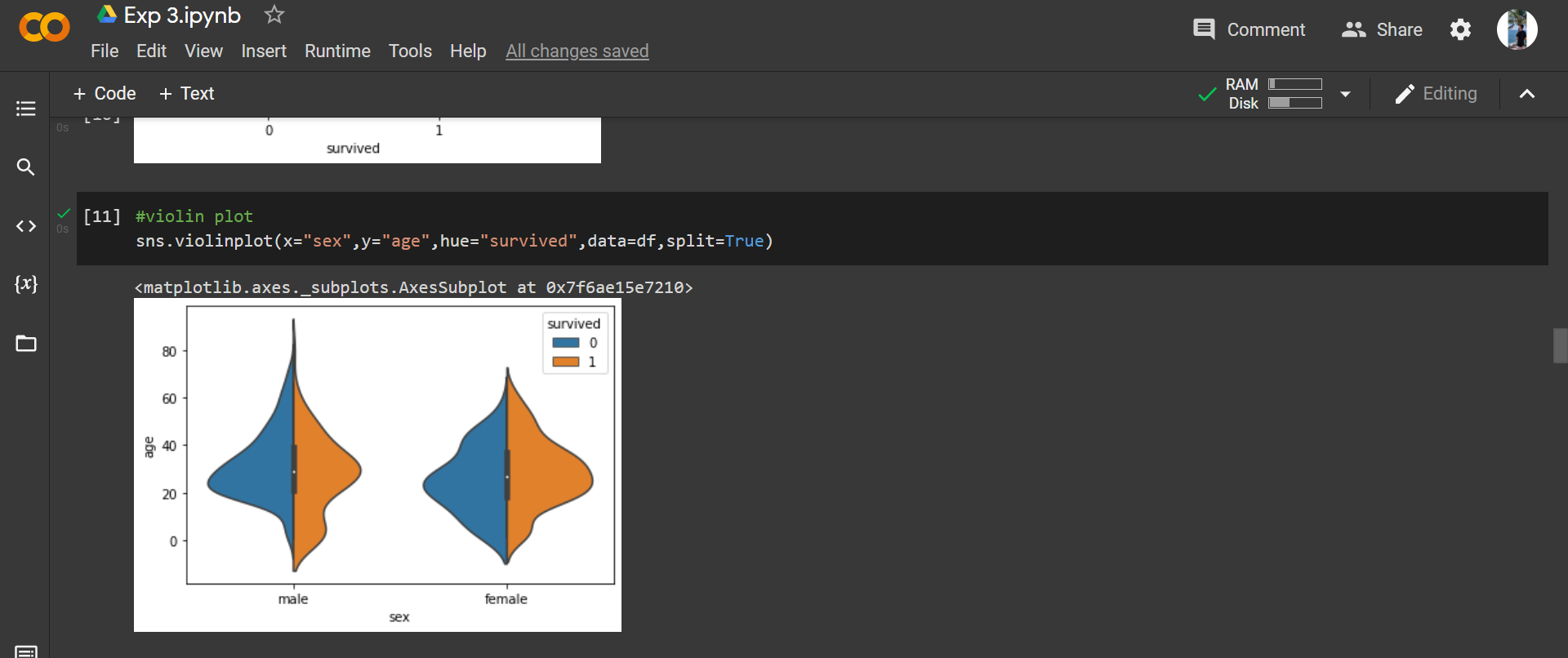


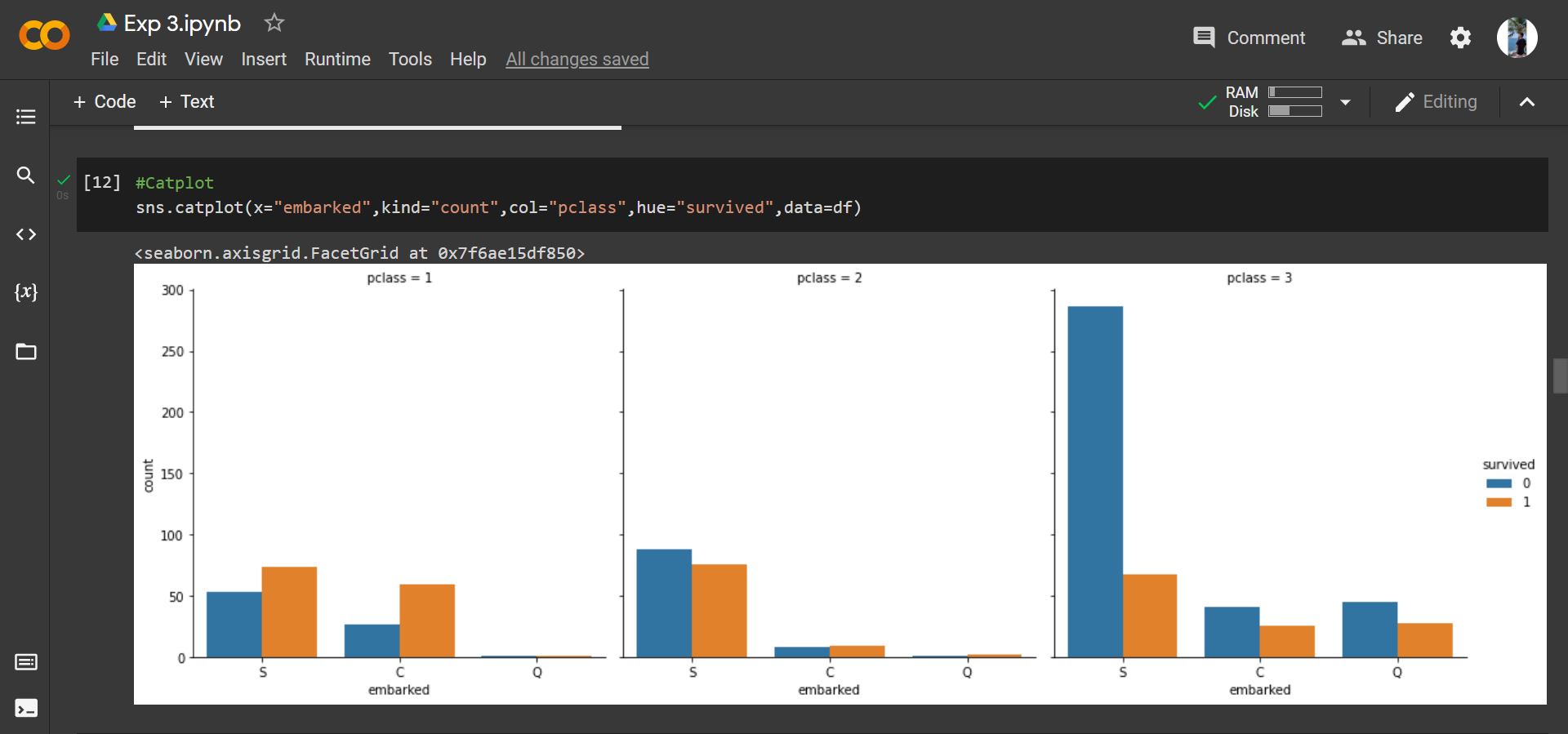


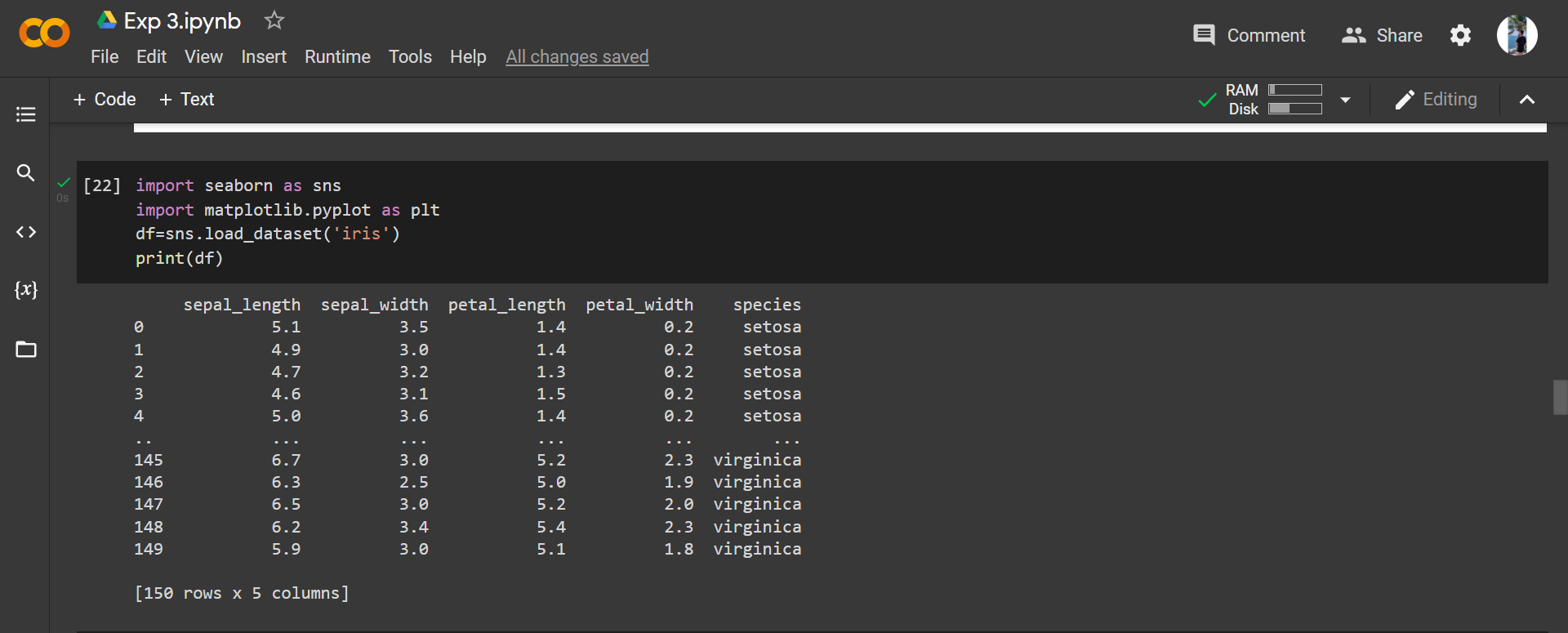


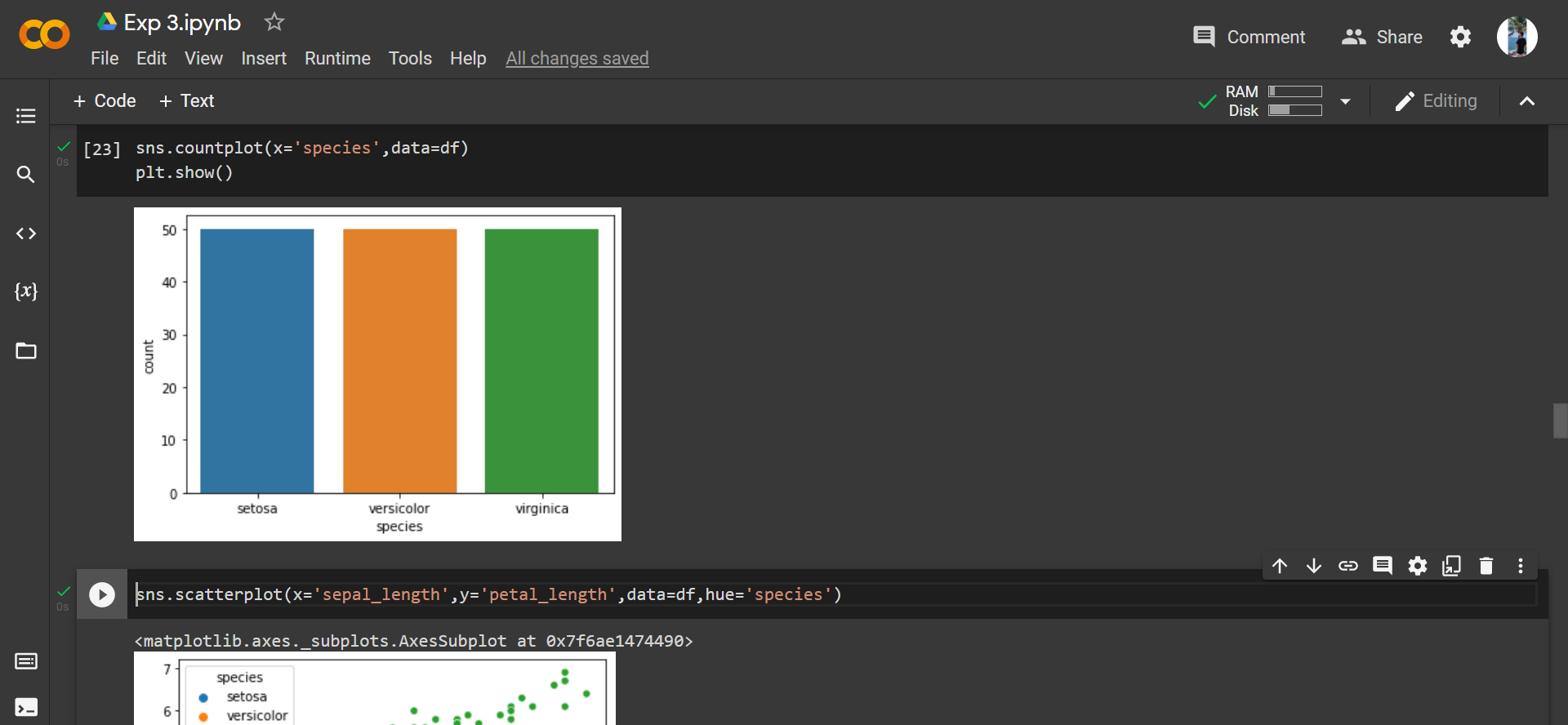


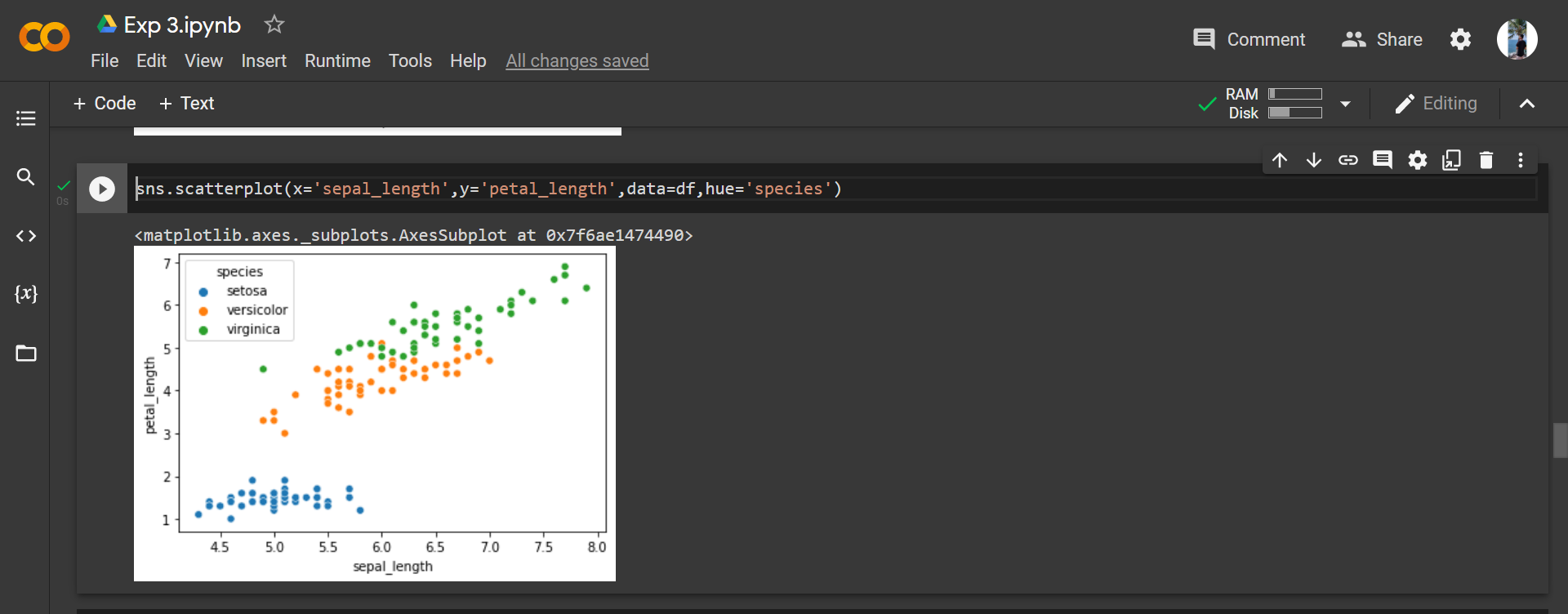




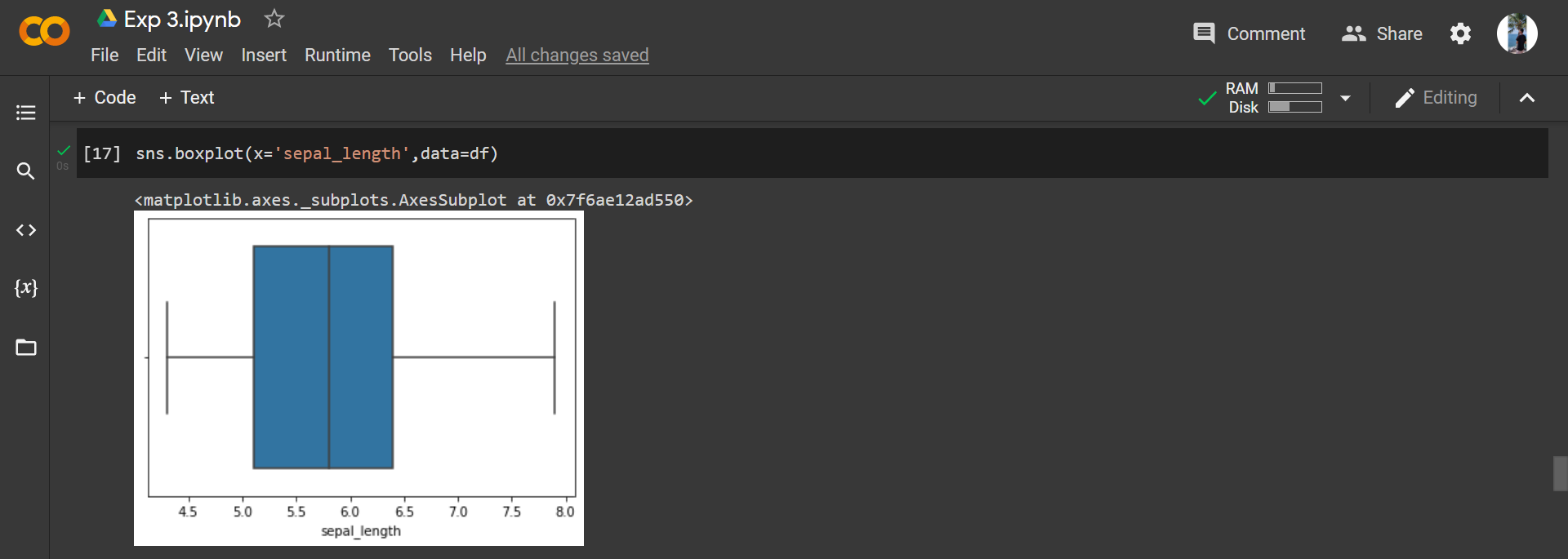


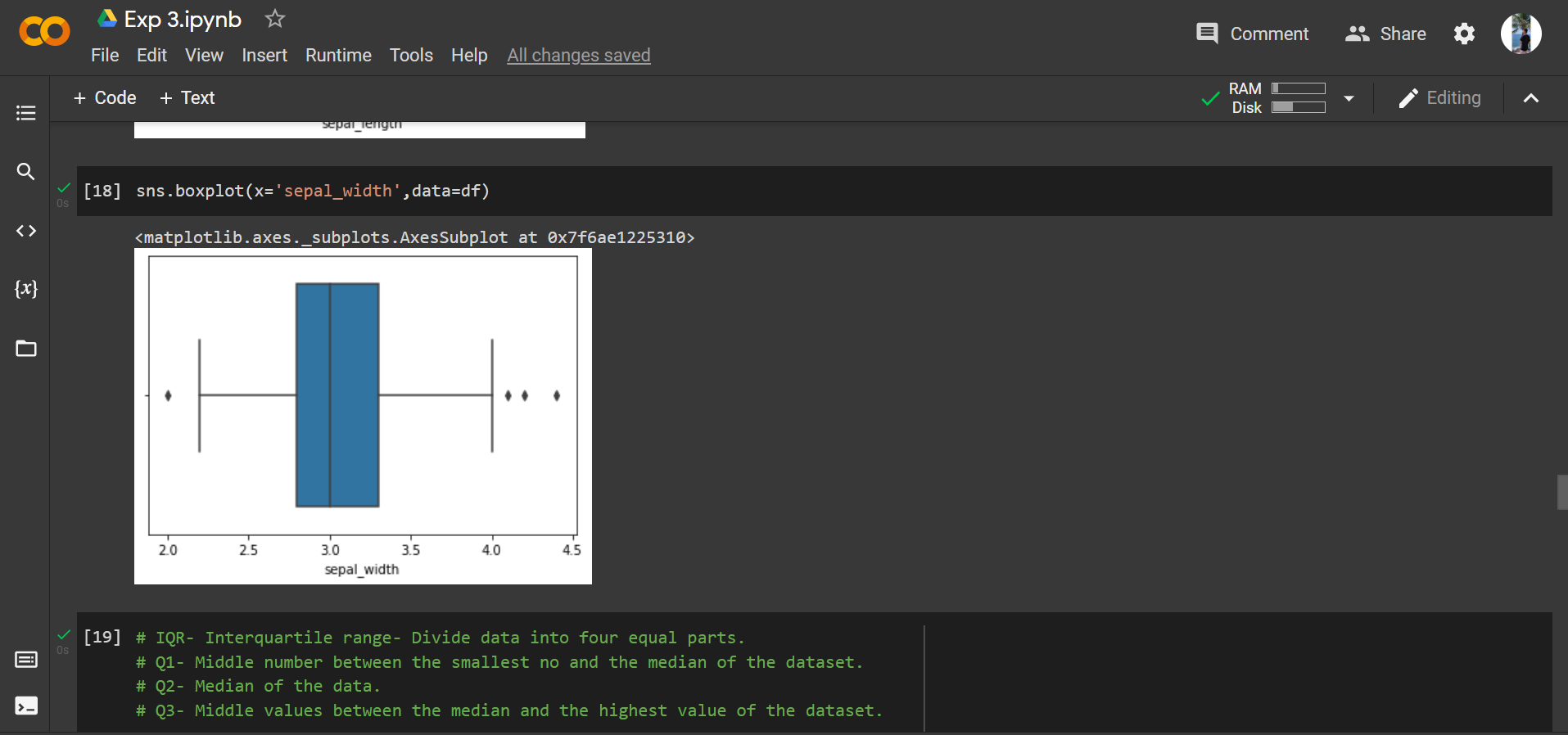


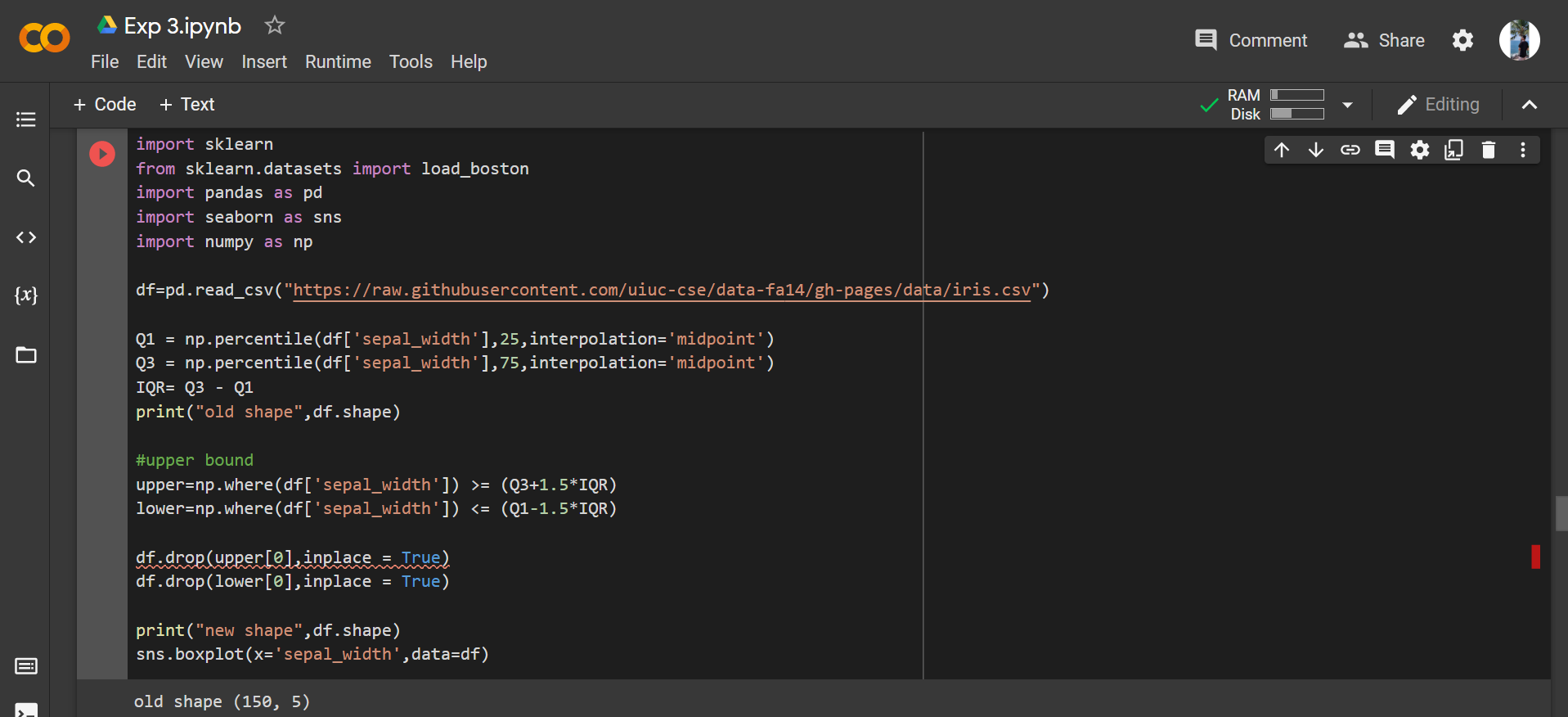


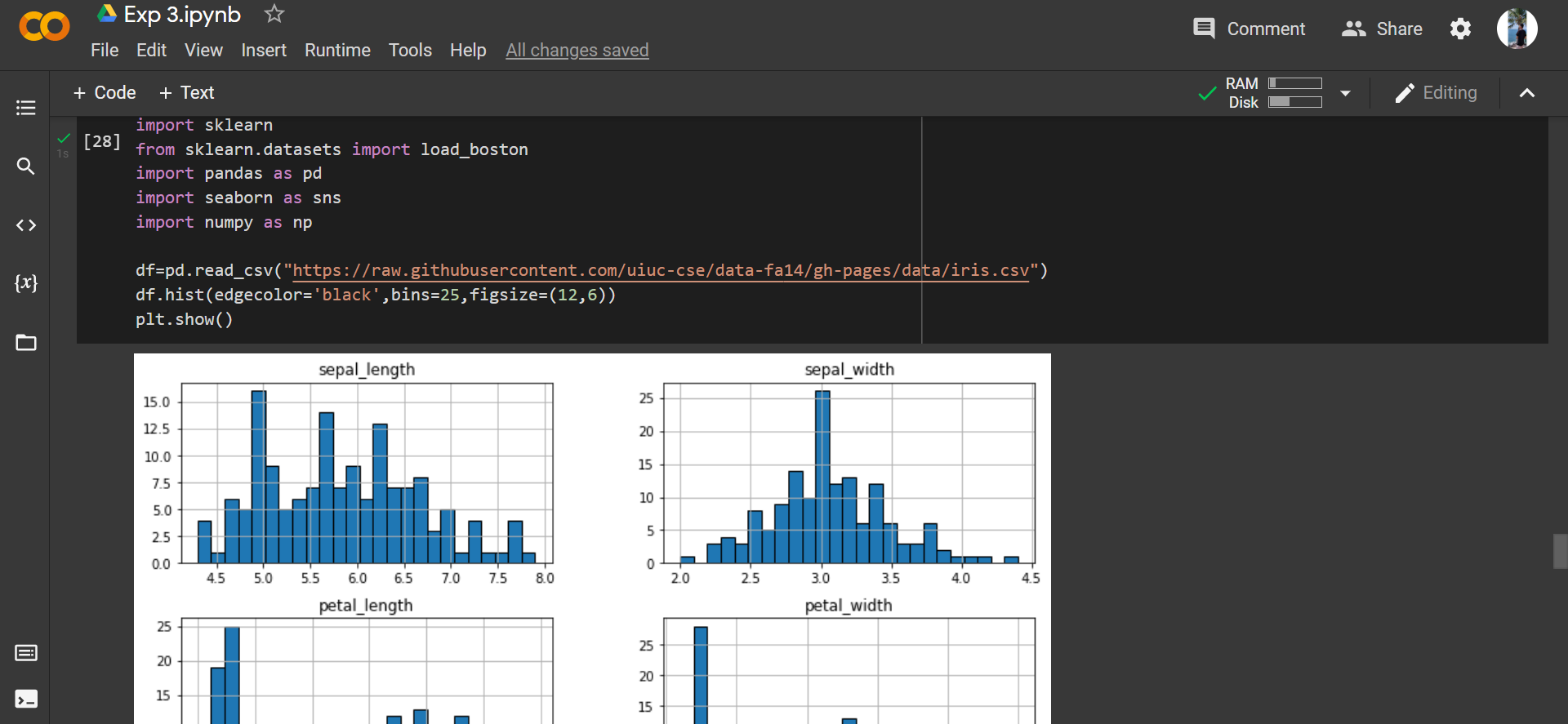


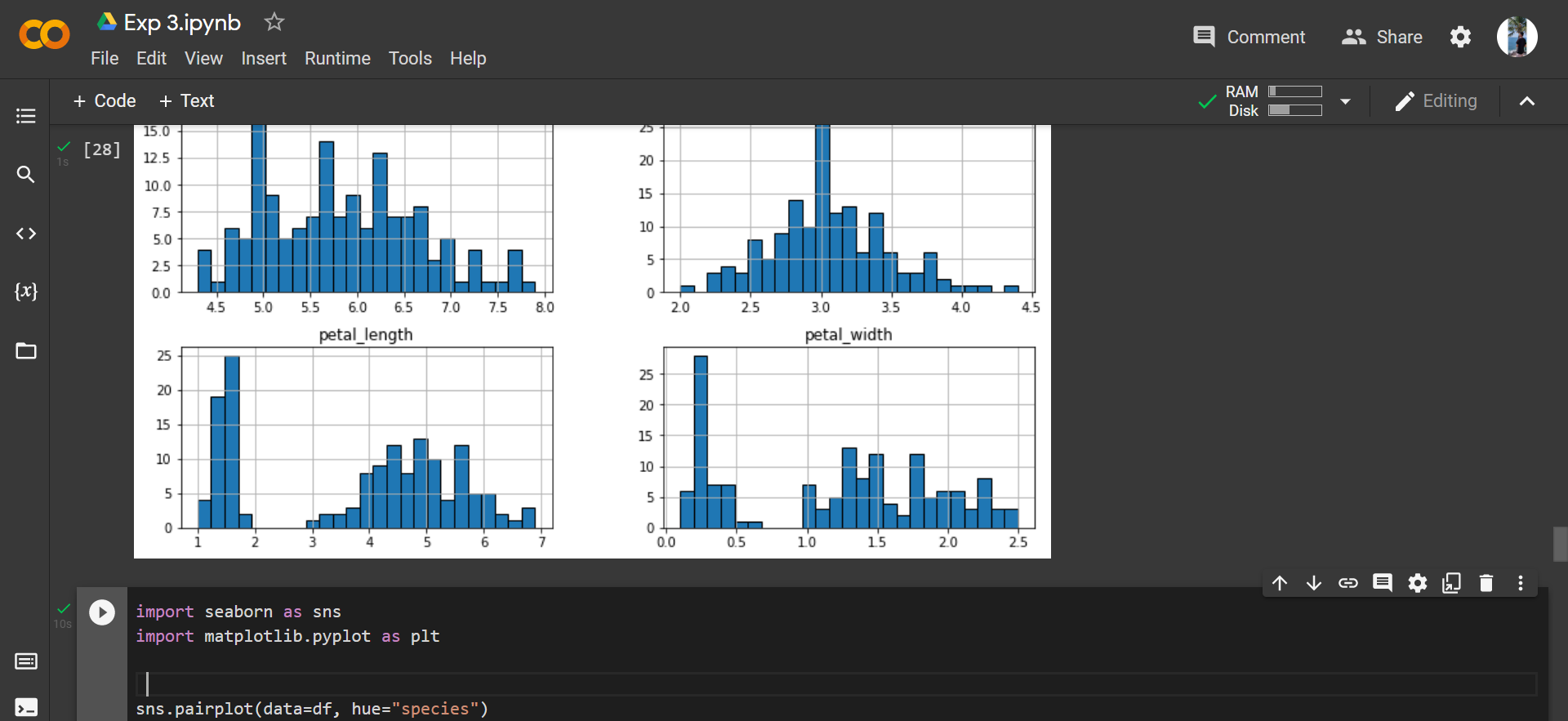


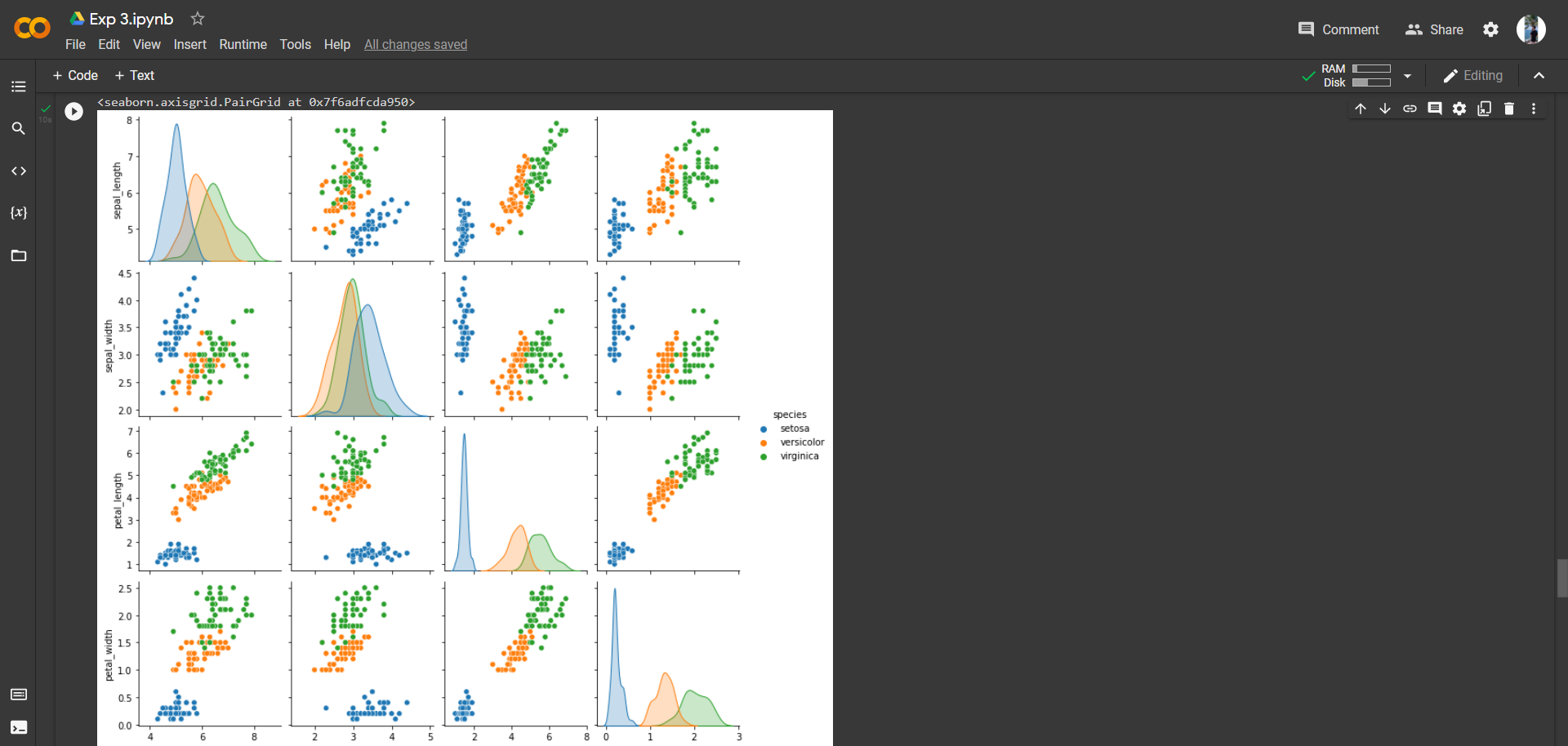


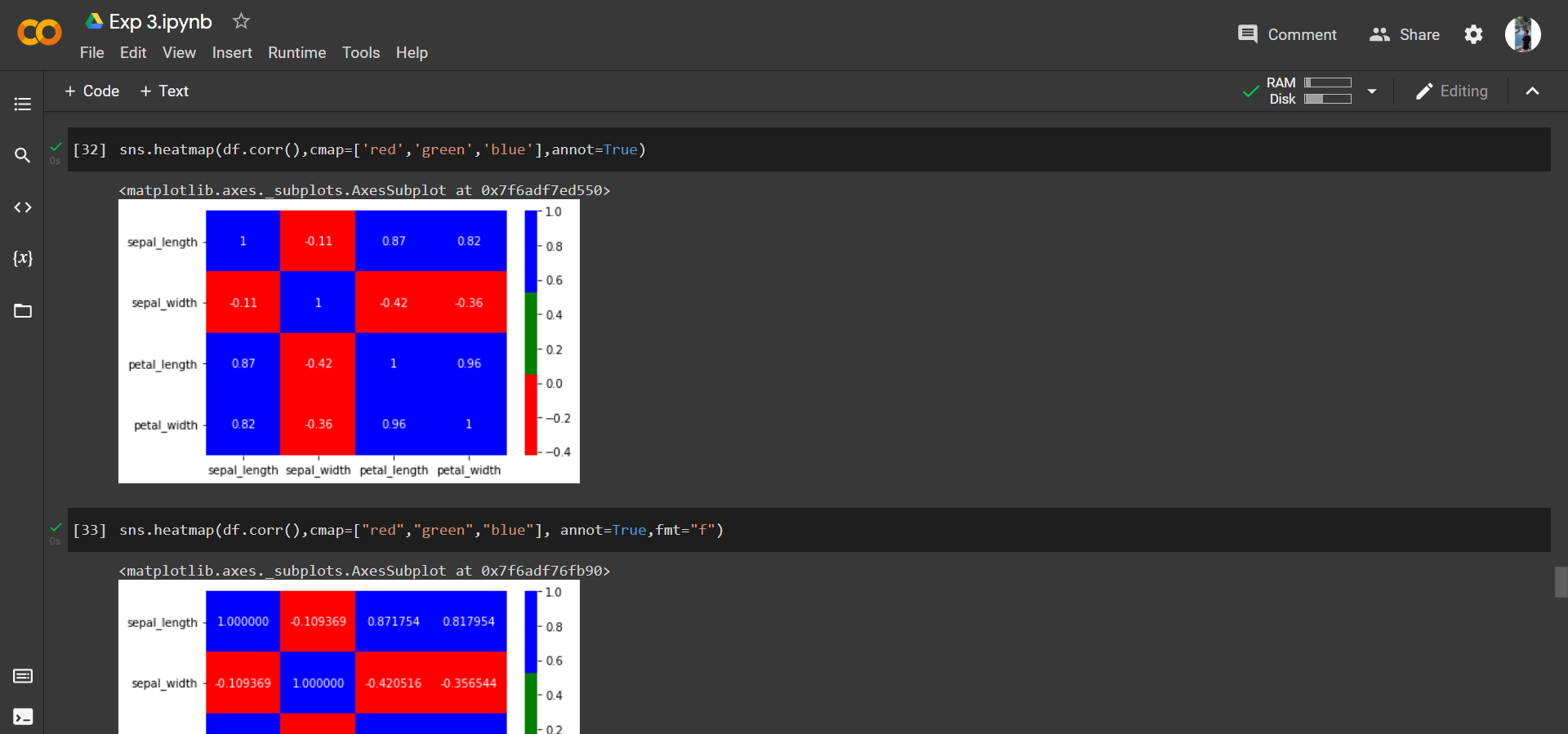




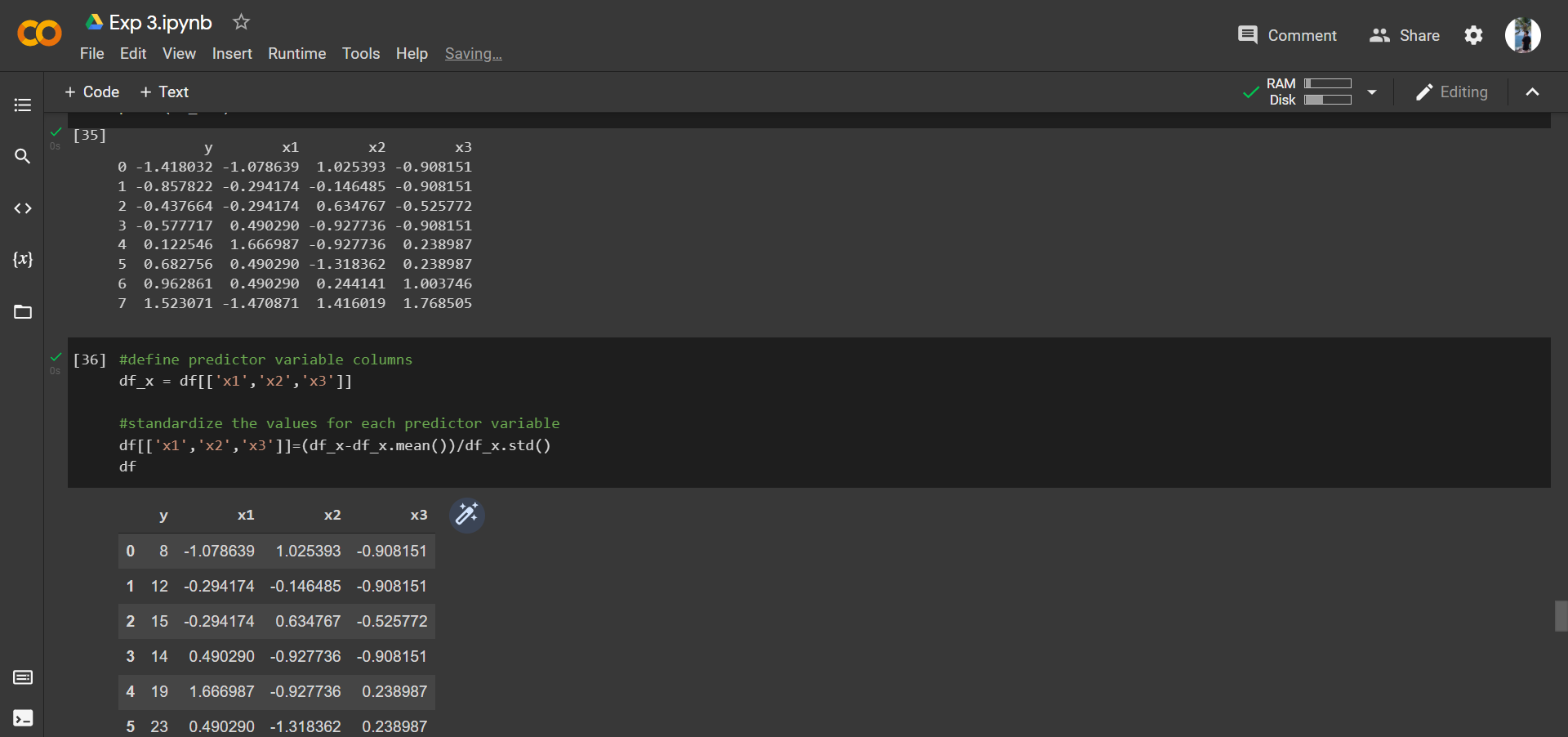


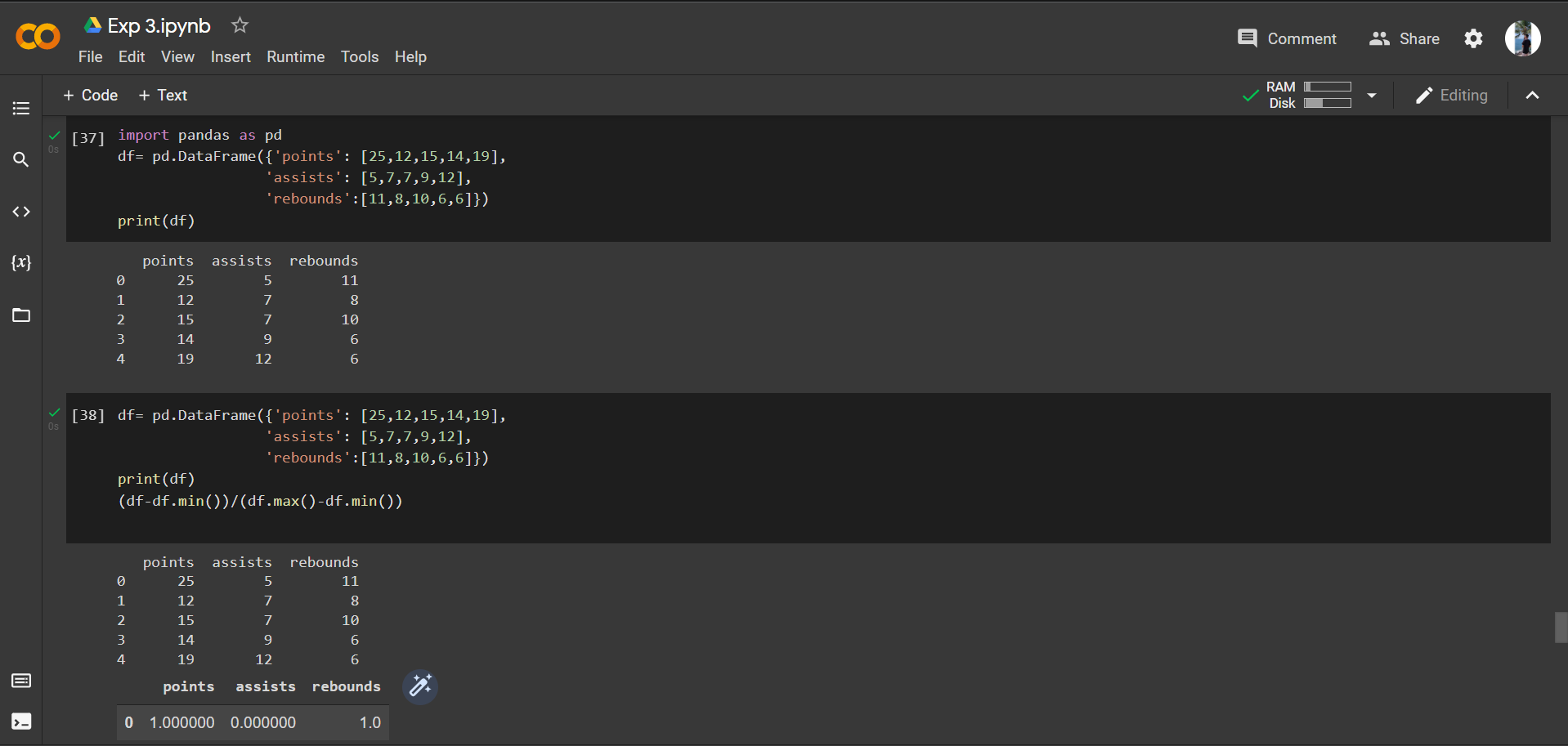


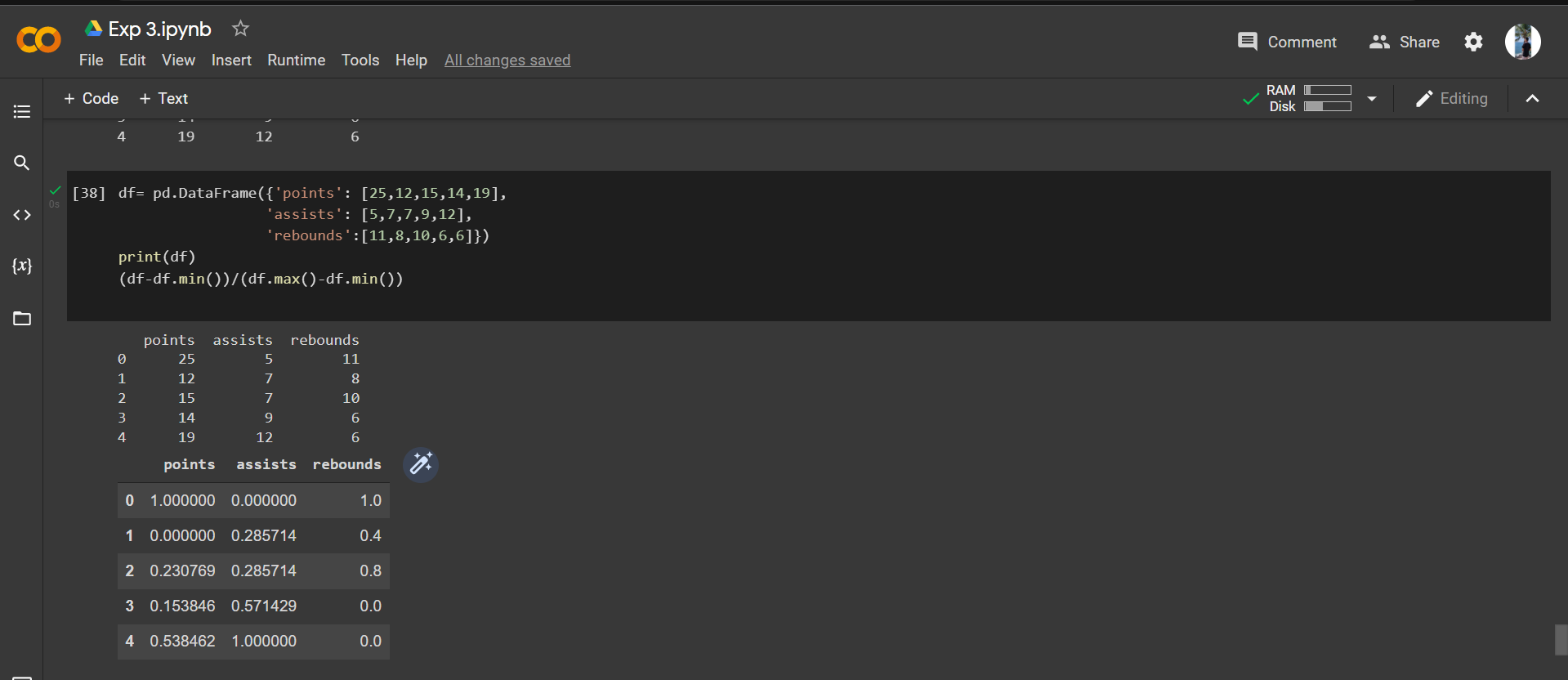


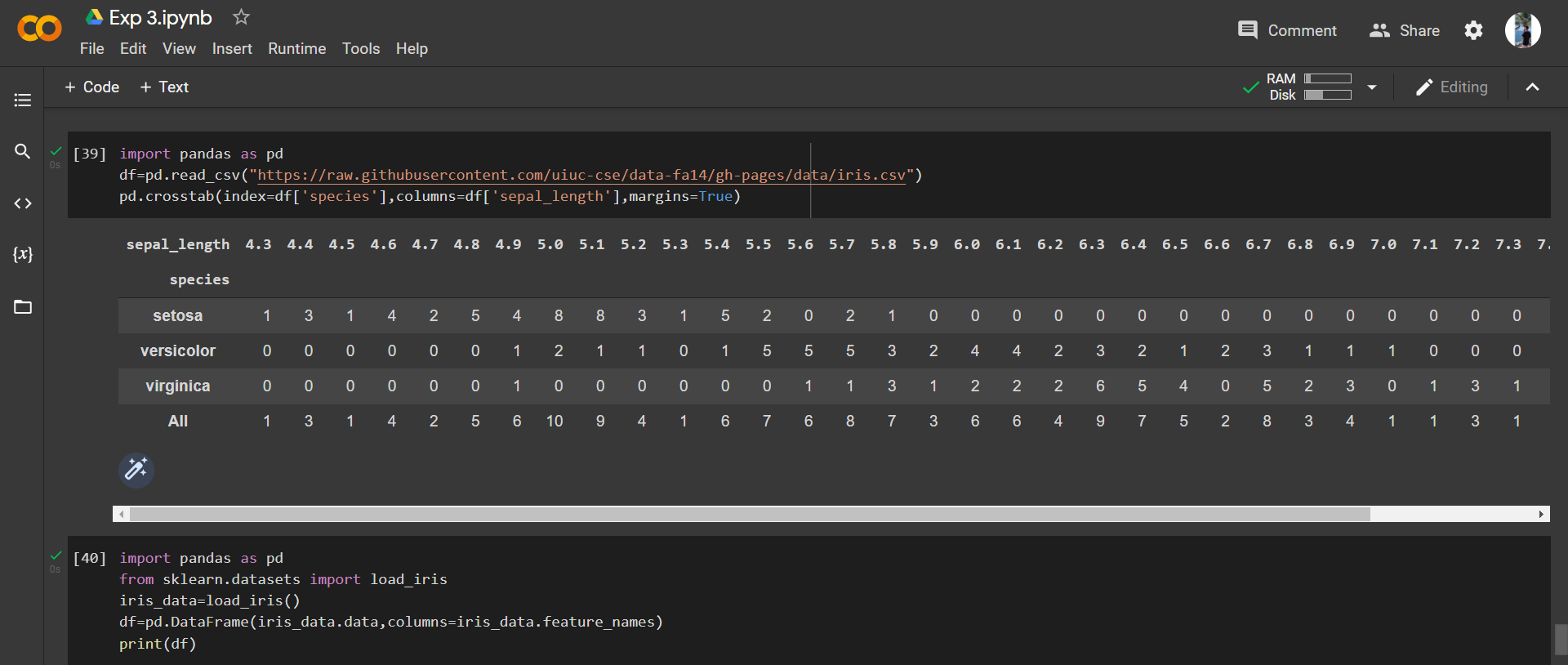


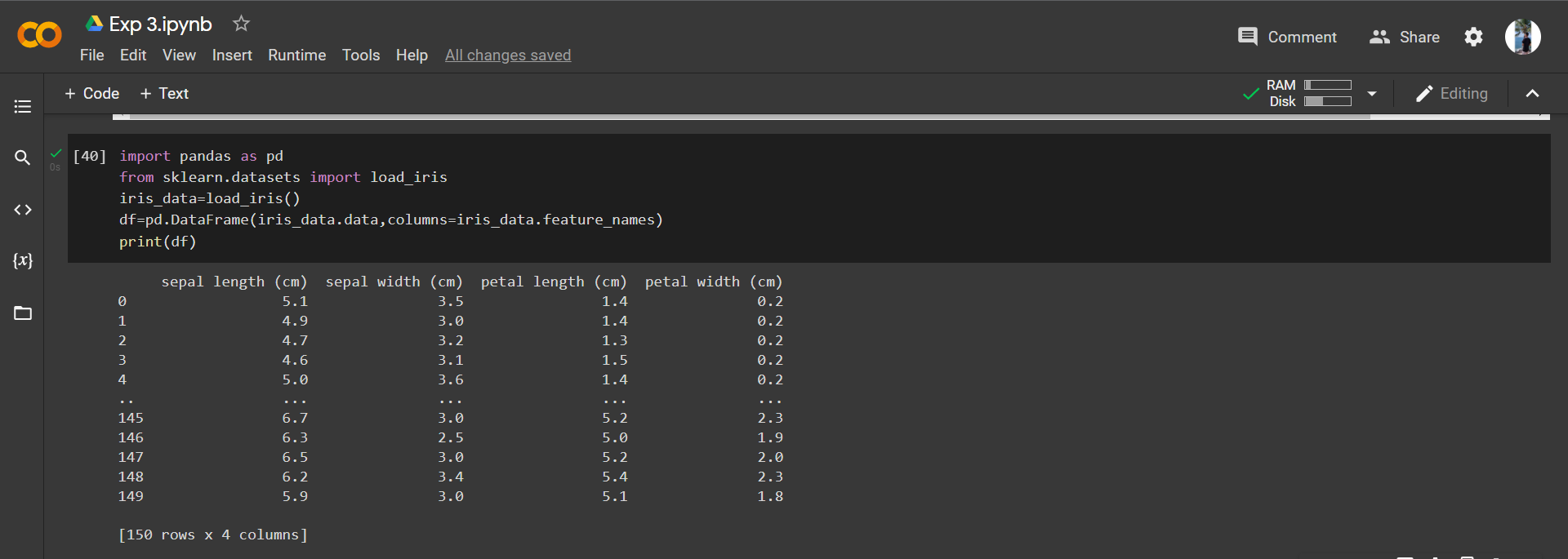


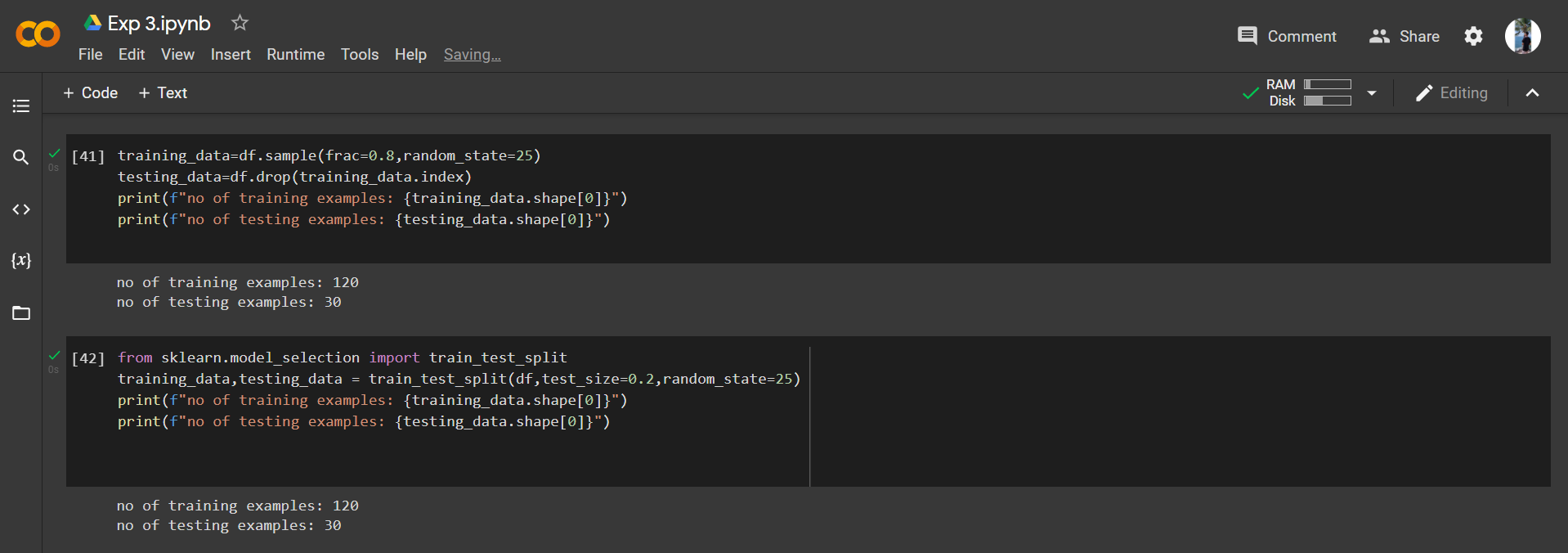


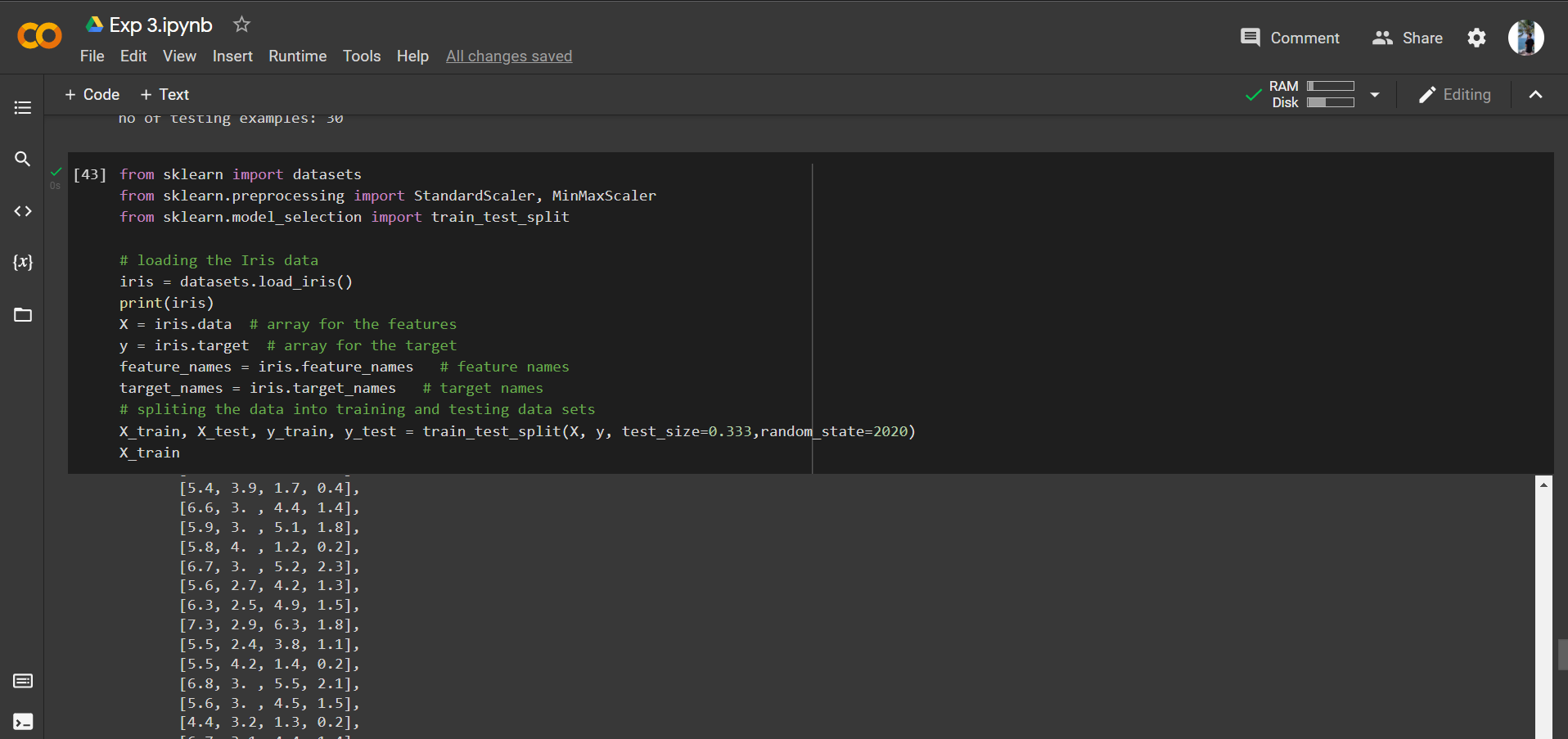








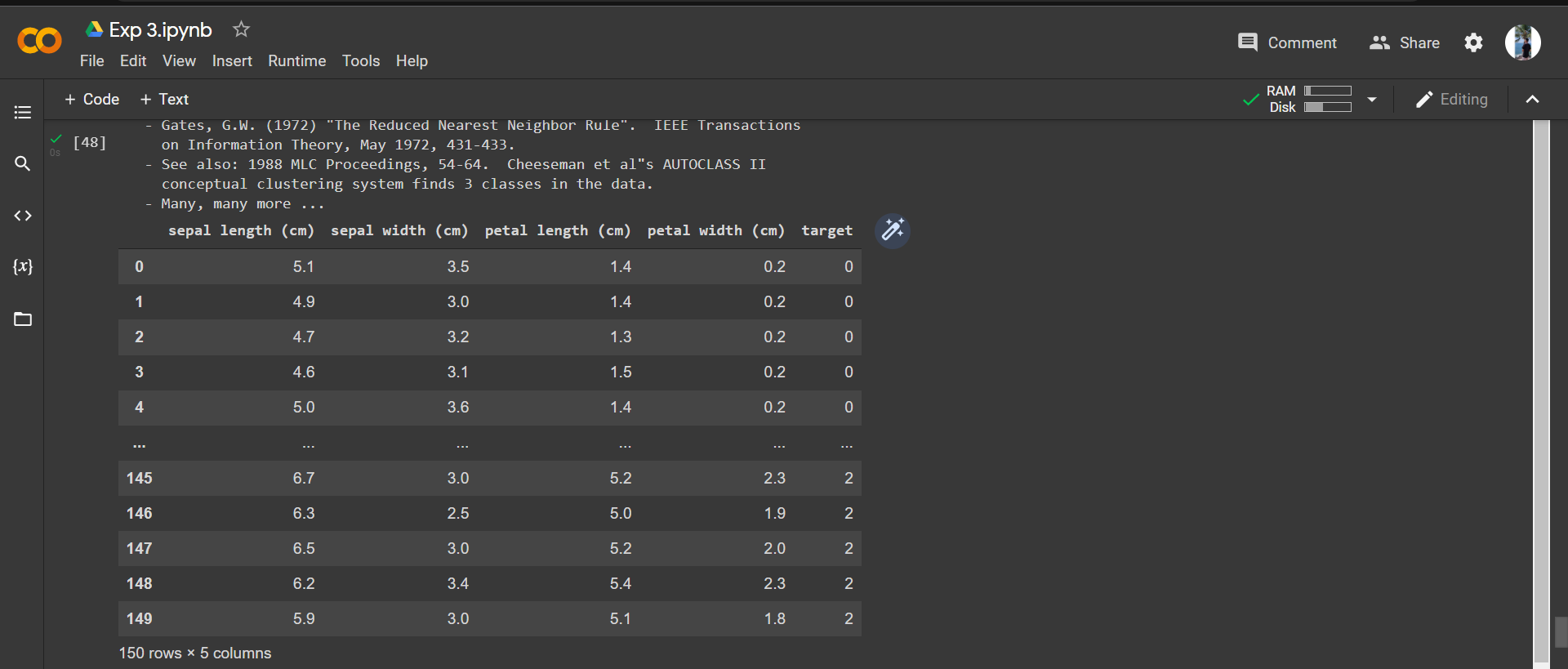


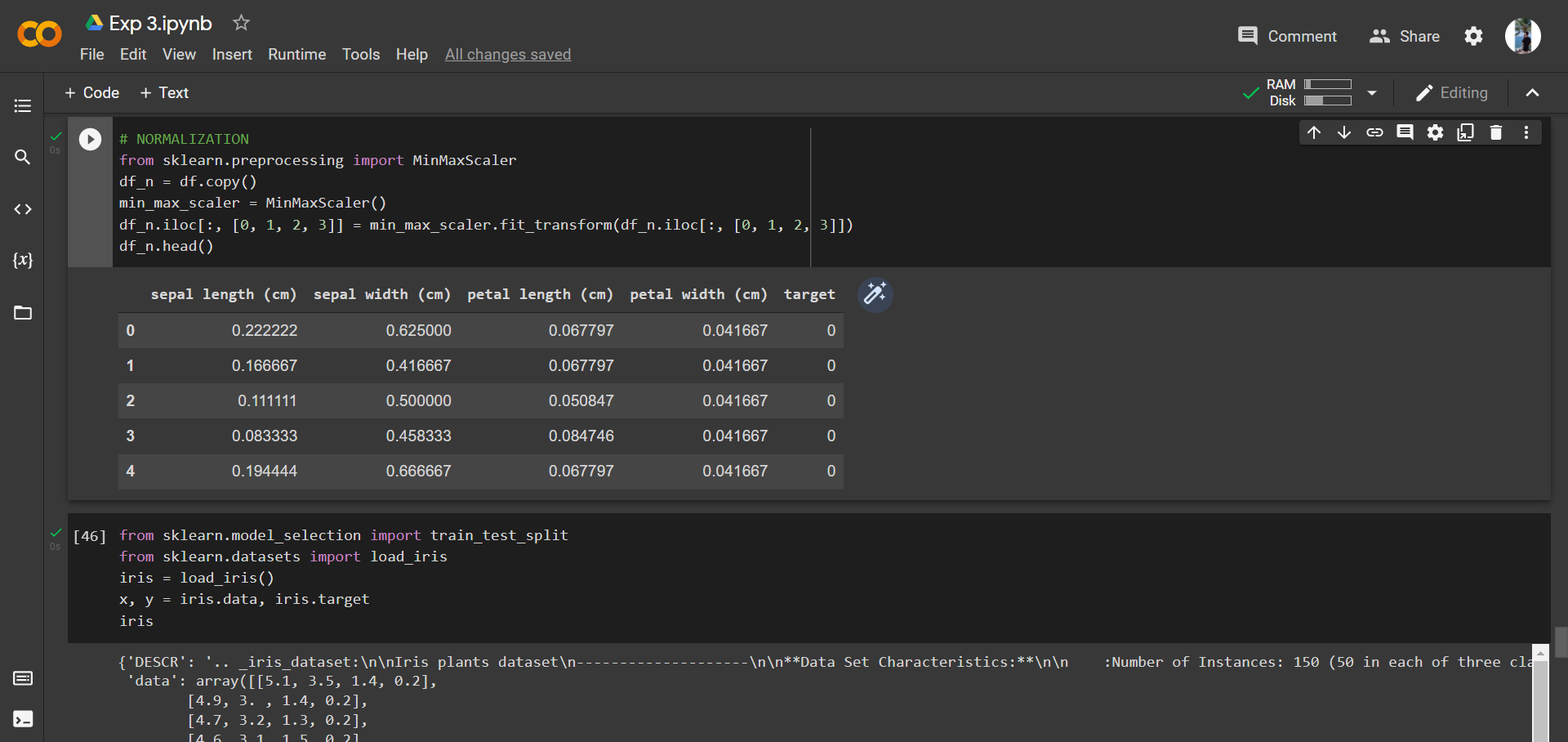


**8. Post-Experiments Exercise**

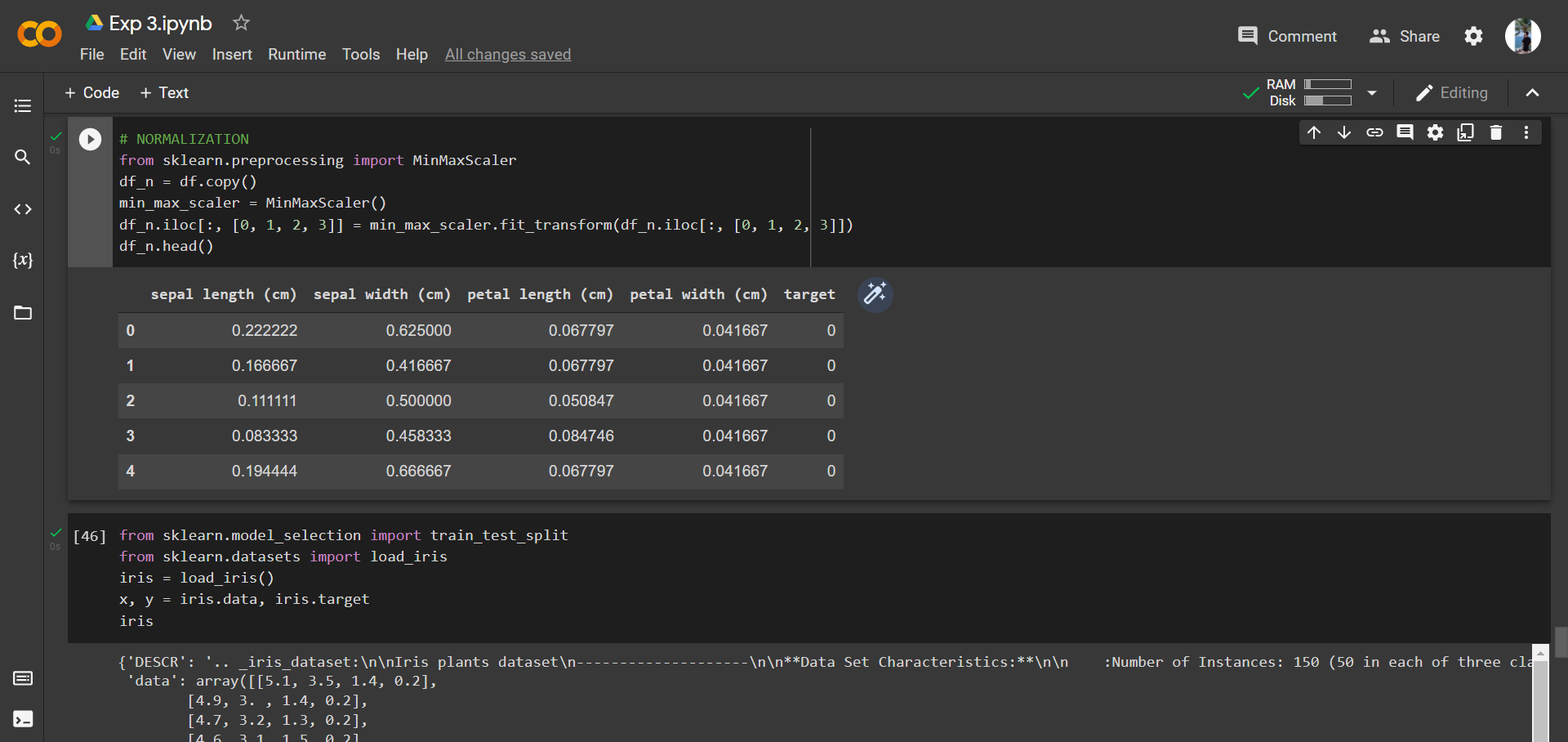
**A. Extended Theory: (Soft Copy)**

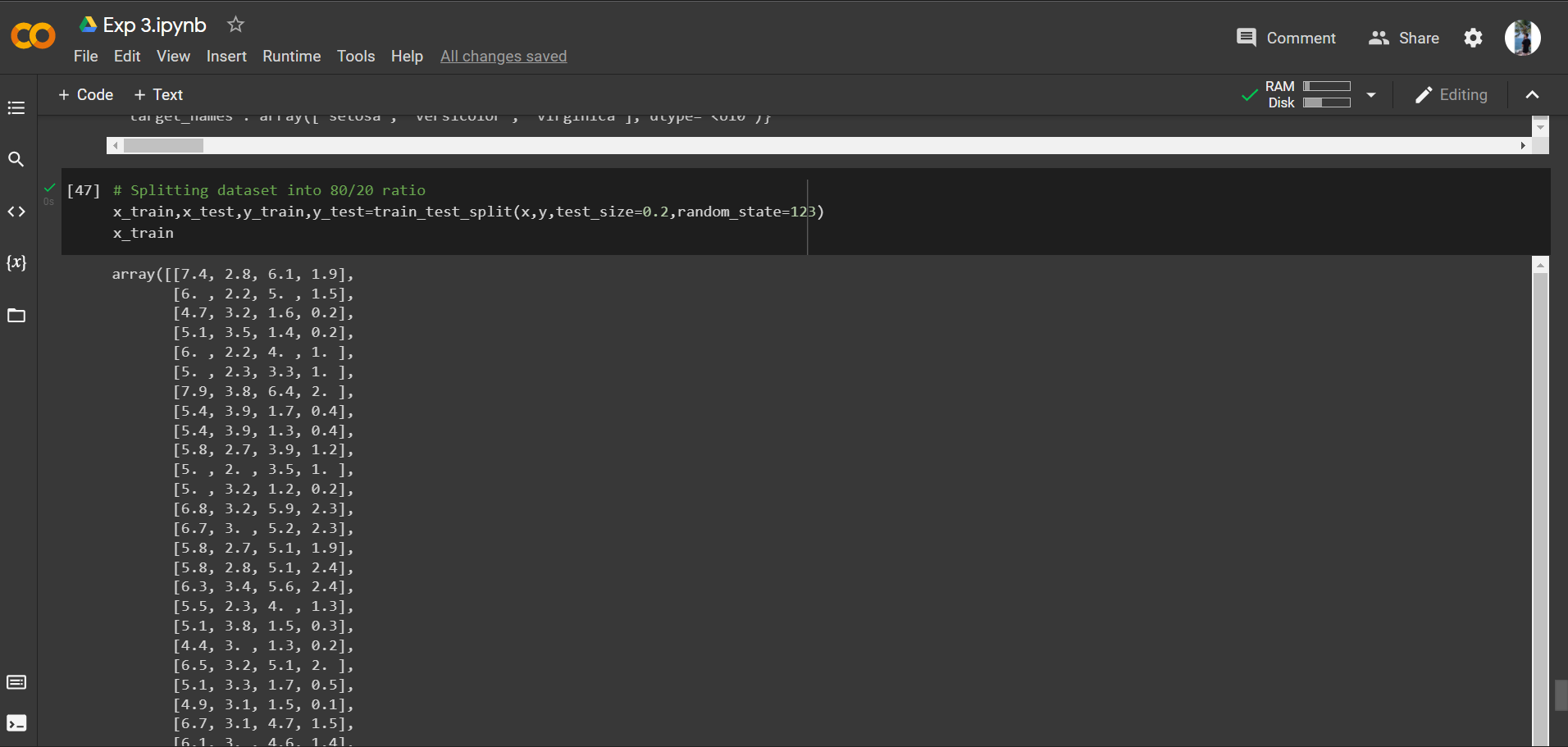
* Use iris dataset and perform rescaling using sklearn package using normalization





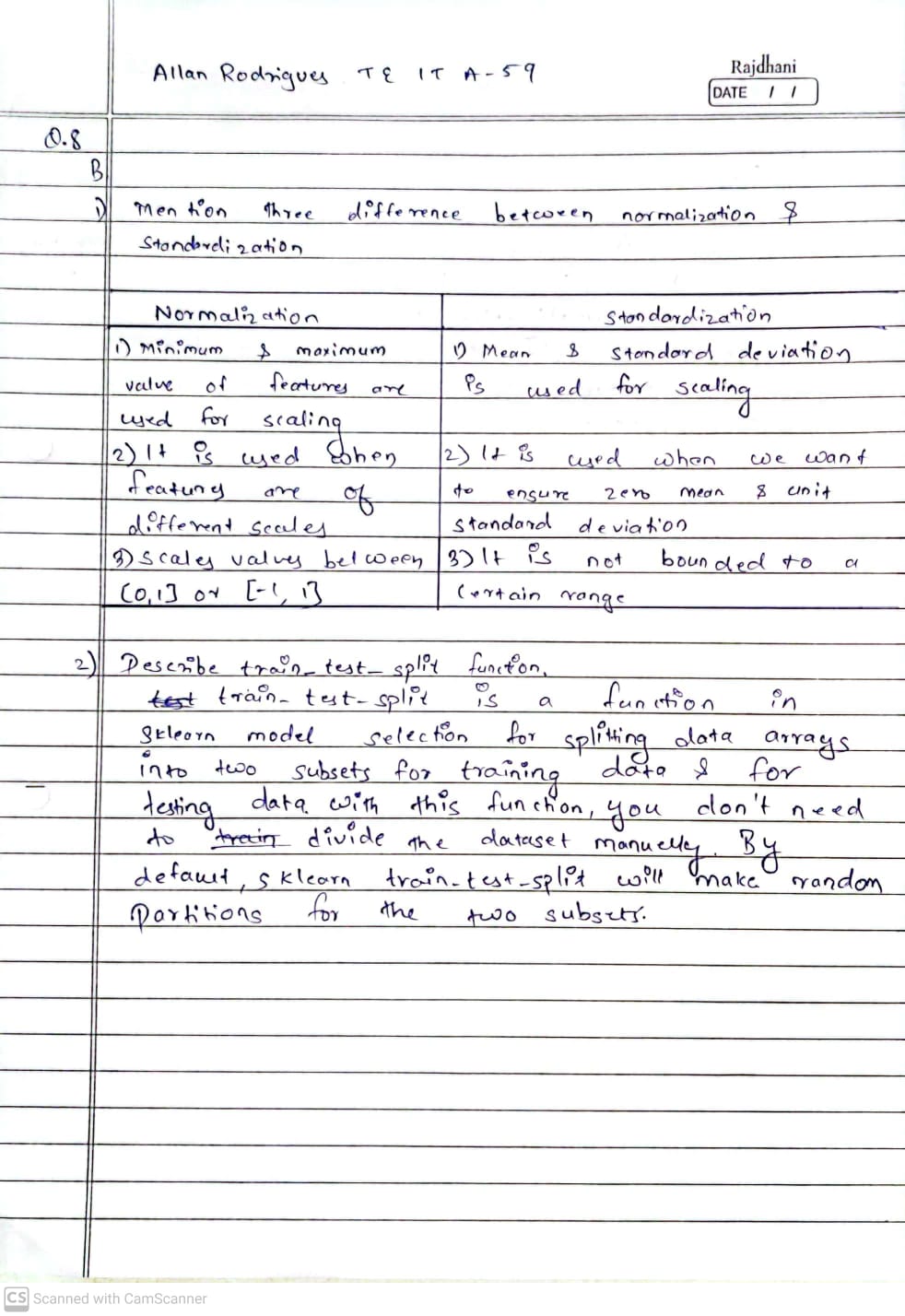
* Partition the iris dataset such that 80% data to be taken for training purpose





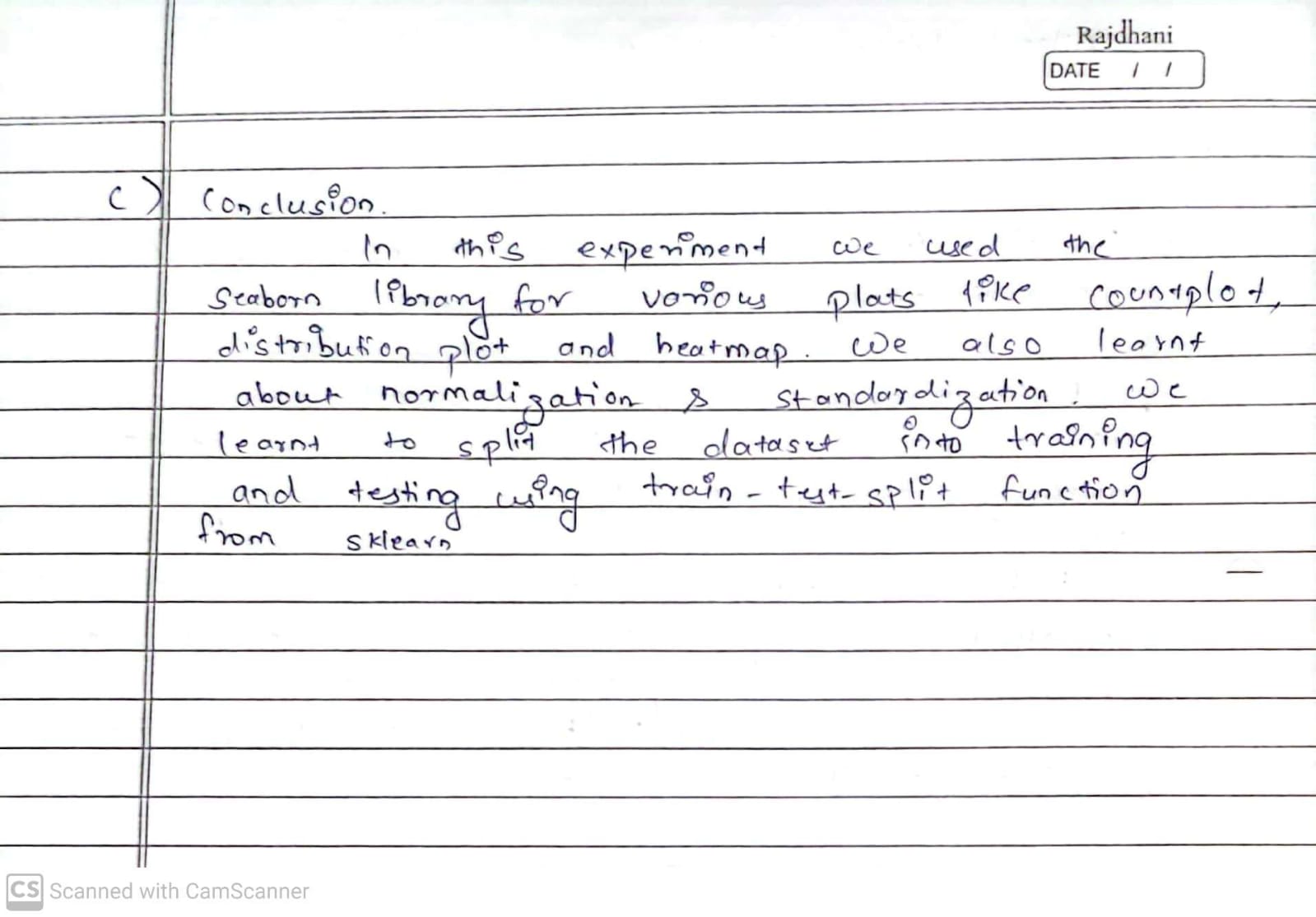
**B. Questions:**

* + - Mention three differences between normalization and standardization.
    - Describe train\_test\_split function



**C. Conclusion:**

Write the significance of the topic studied in the experiment.



**D. References:**

1. <https://www.geeksforgeeks.org/exploratory-data-analysis-on-iris-dataset/>

2. <https://www.statology.org/normalize-columns-pandas-dataframe/Normalization>

3. <https://www.datascienceguide.org/python-code-snippets.html>

4. <https://machinelearningmastery.com/standardscaler-and-minmaxscaler-transforms-in-python/>

**---------------------------------**