| MINI PROJECT  BOSTON HOUSE PRICE PREDICTION |  |
| --- | --- |

REPORT

BY

MOHAMMED YOONUS P

RAJALAKSHMI ENGINEERING COLLEGE

CHENNAI

INTRODUCTION:

Mohammed yoonus P

Data Science

Project is about the Boston price prediction

DATA DESCRIPTION:

The Boston Housing Dataset The Boston Housing Dataset is a derived from information collected by the U.S. Census Service concerning housing in the area of Boston MA. The following describes the dataset columns:

● CRIM - per capita crime rate by town

● ZN - proportion of residential land zoned for lots over 25,000 sq. ft.

● INDUS - proportion of non-retail business acres per town.

● CHAS - Charles River dummy variable (1 if tract bounds river; 0 otherwise)

● NOX - nitric oxides concentration (parts per 10 million)

● RM - average number of rooms per dwelling

● AGE- proportion of owner-occupied units built prior to 1940

● DIS - weighted distances to five Boston employment centers

● RAD - index of accessibility to radial highways

● TAX - full-value property-tax rate per $10,000

● PTRATIO - pupil-teacher ratio by town

● B - 1000(Bk - 0.63)^2 where Bk is the proportion of blacks by town

● LSTAT - % lower status of the population

● MEDV - Median value of owner-occupied homes in $1000's

APPROACH:

The way of approaching this report by classifying it in following topics

Visualization

Information

description

co-relation

Rad values

Chas features

Chas data

Age

Crime rate

Number of rooms in the house

ALGORITHM

I have used various classification models on this dataset and they have different accuracy and other performance measures. I have used the following machine learning algorithms on this dataset.

• numpy as np

• pandas as pd

• csv

• matplotlib. pyplot as plt

• seaborn as sns

• sklearn.feature\_selection – SelectKBest, chi2

• sklearn.ensemble - ExtraTreesClassifier

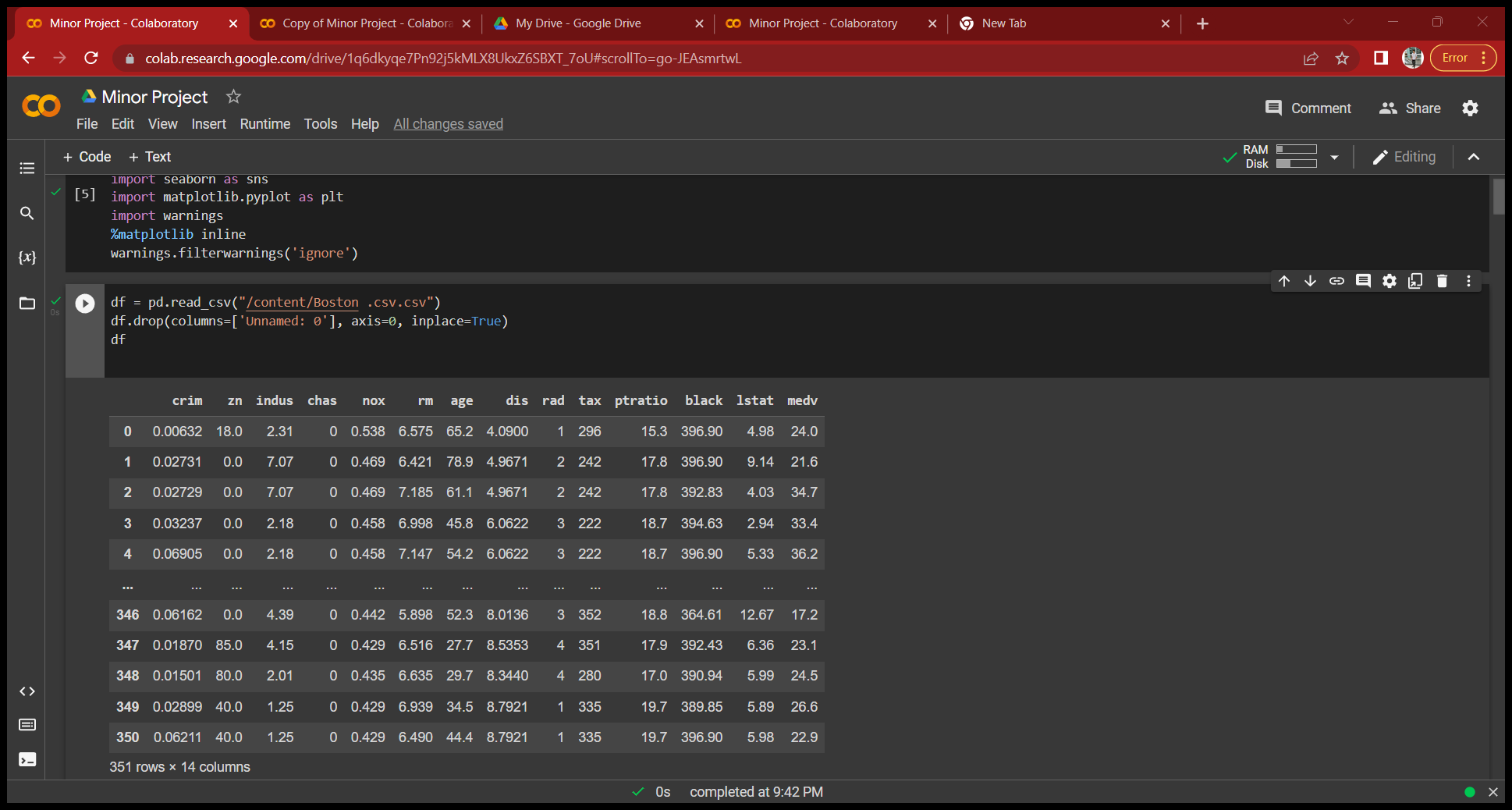
• sklearn.model\_selection - train\_test\_split

• sklearn.linear\_model – LinearRegression

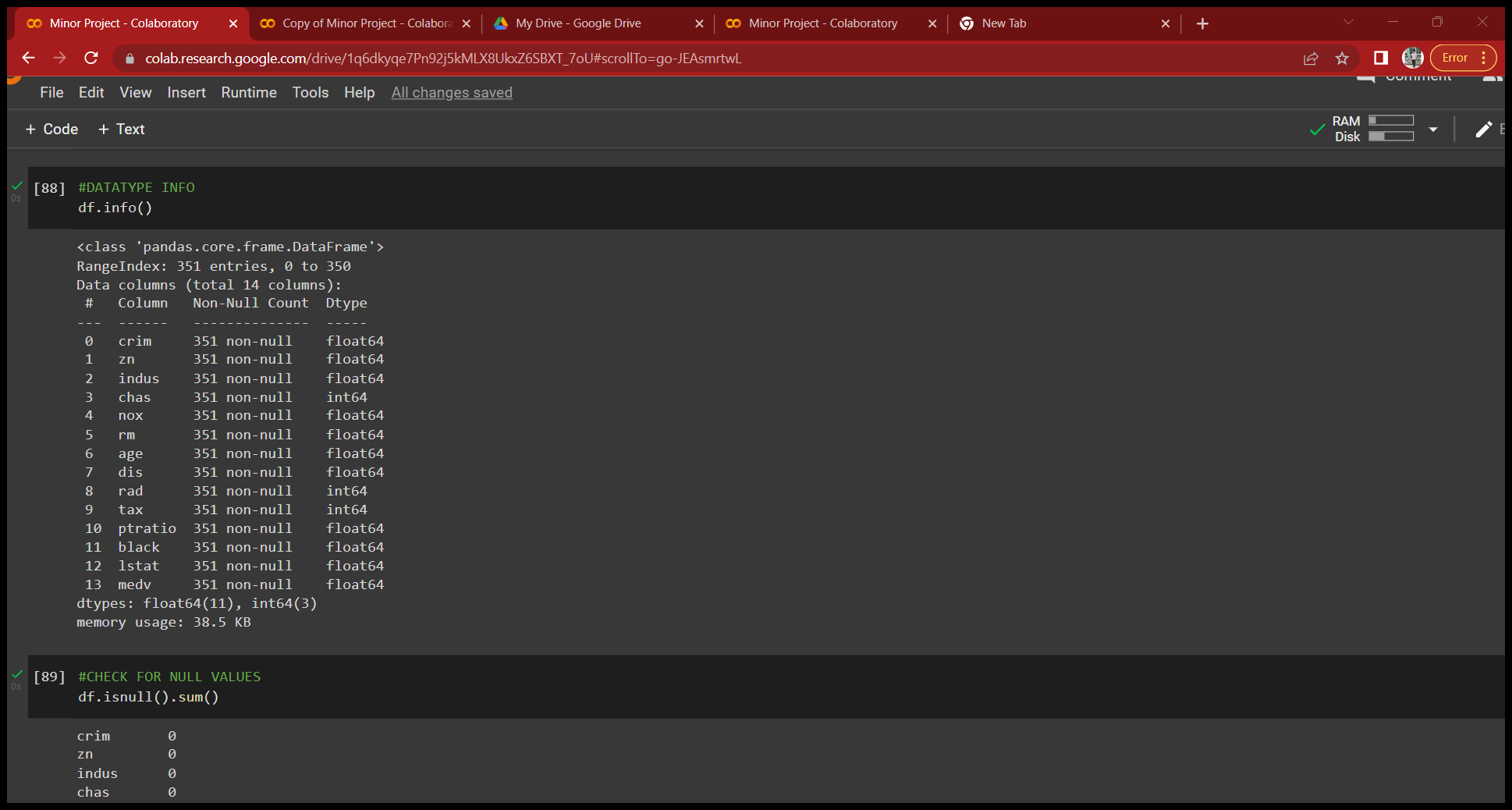
• sklearn.metrics - mean\_squared\_error,r2\_score

VISUALIZATION

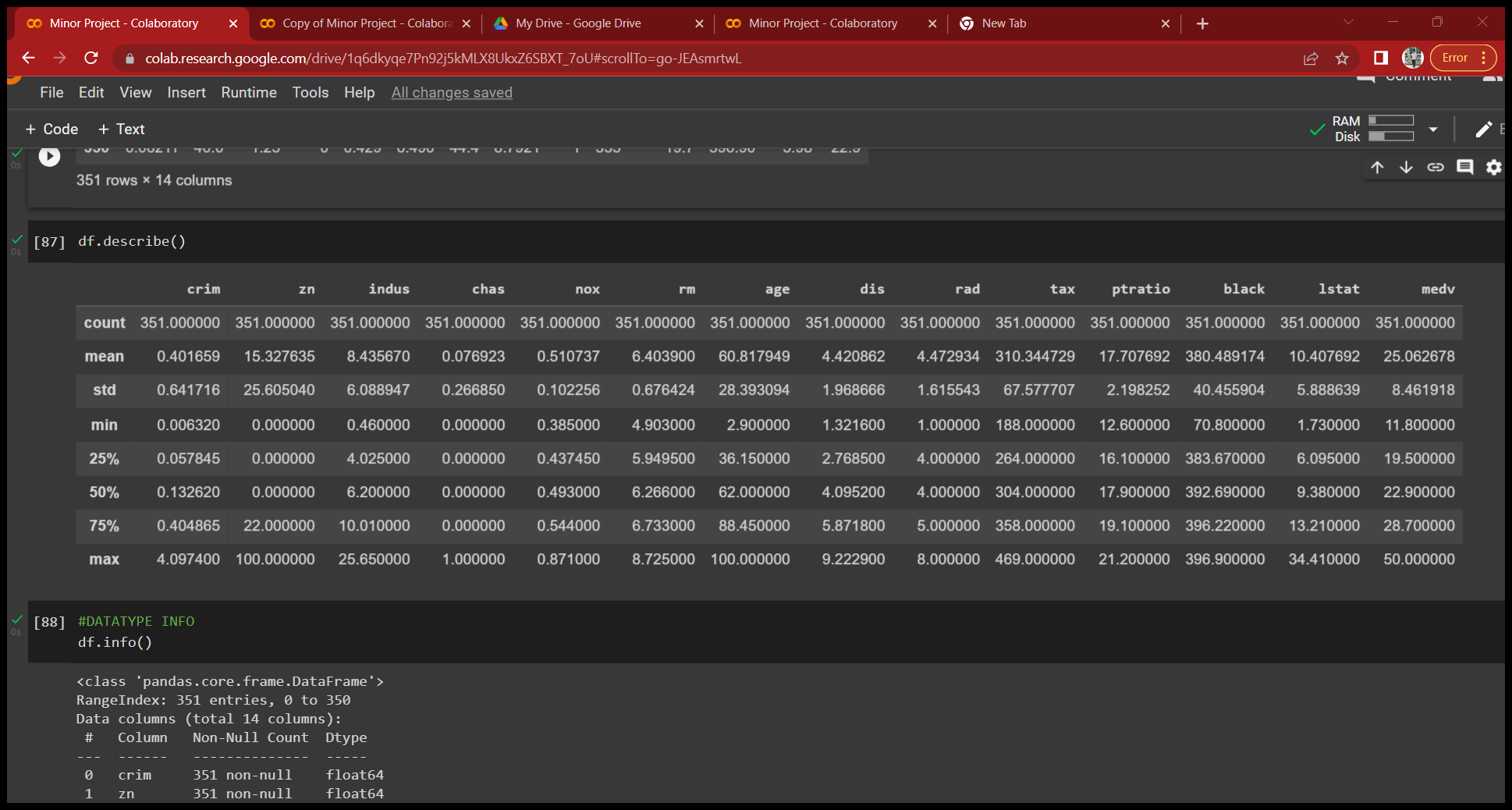
BOSTON TABLE



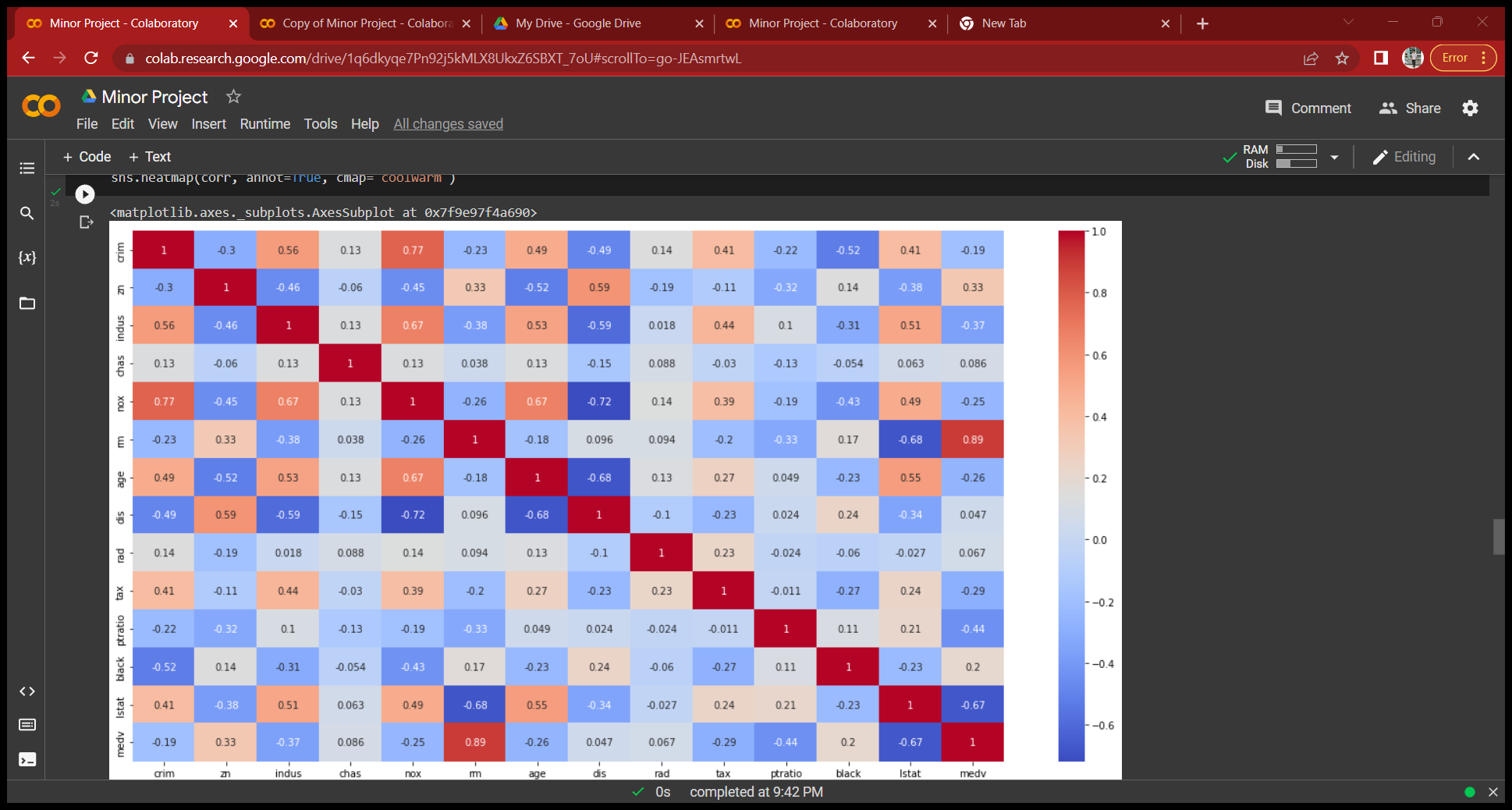
INFORMATION



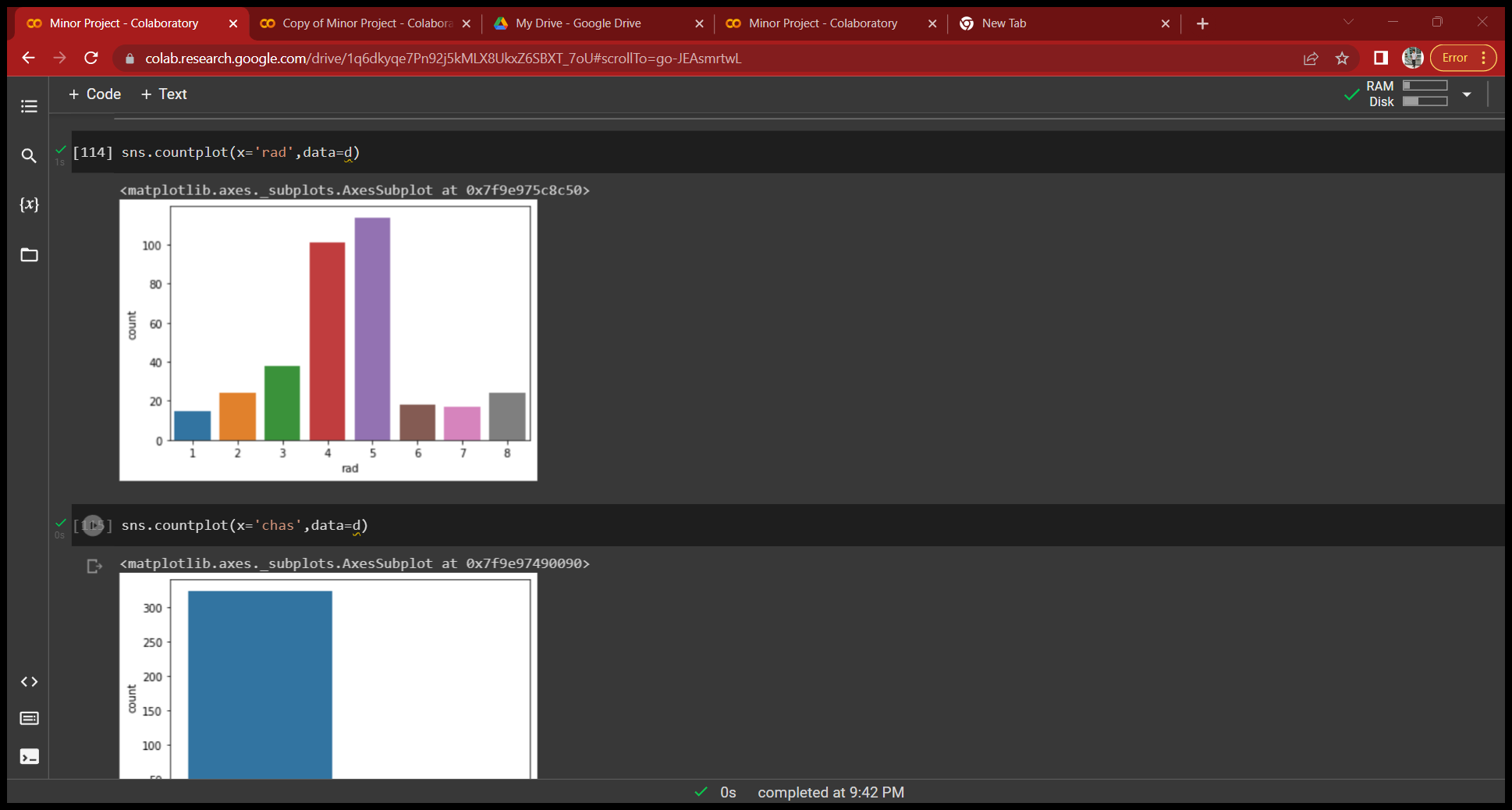
DESCRIPTION



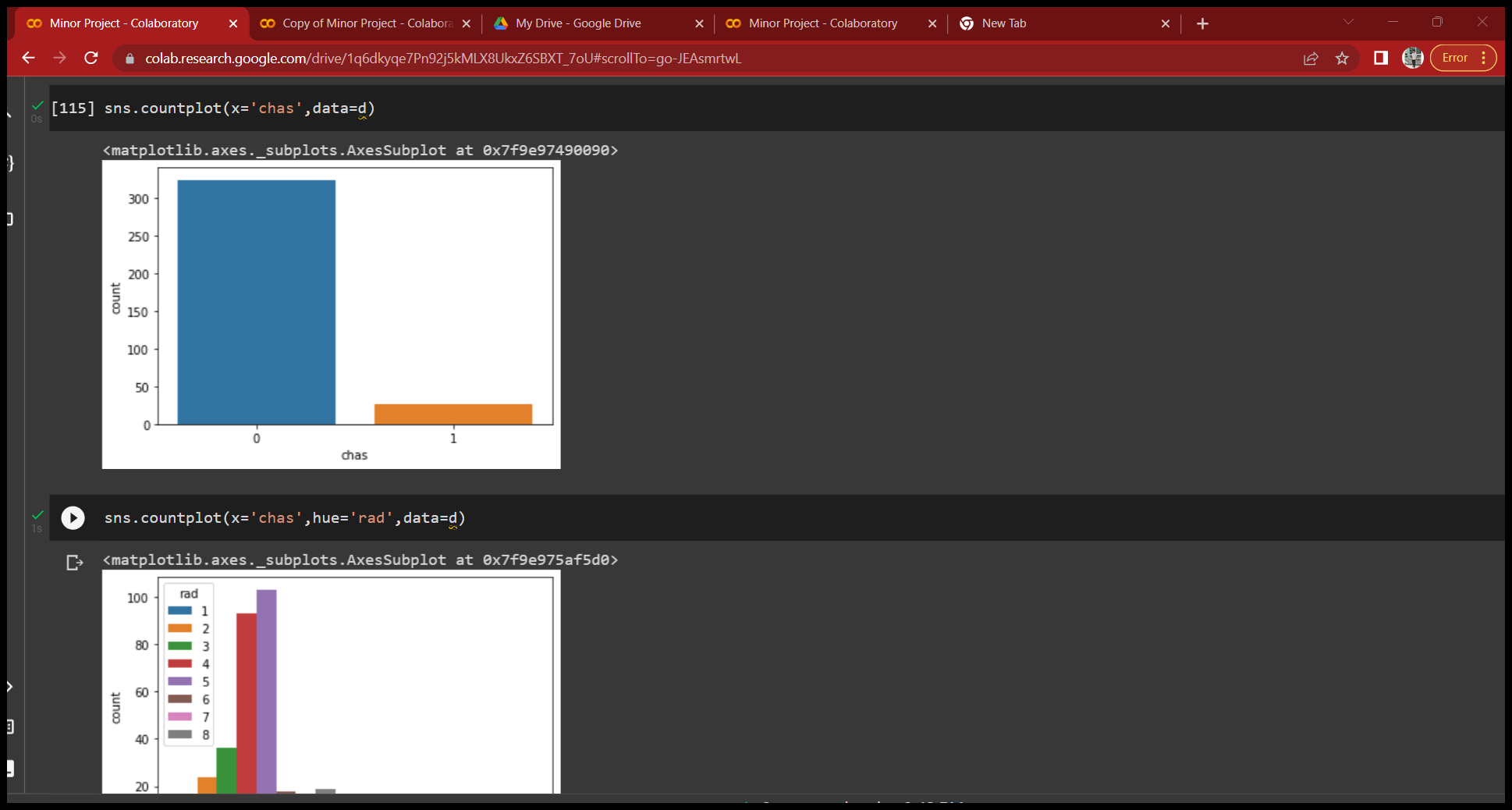
CO-REALTION



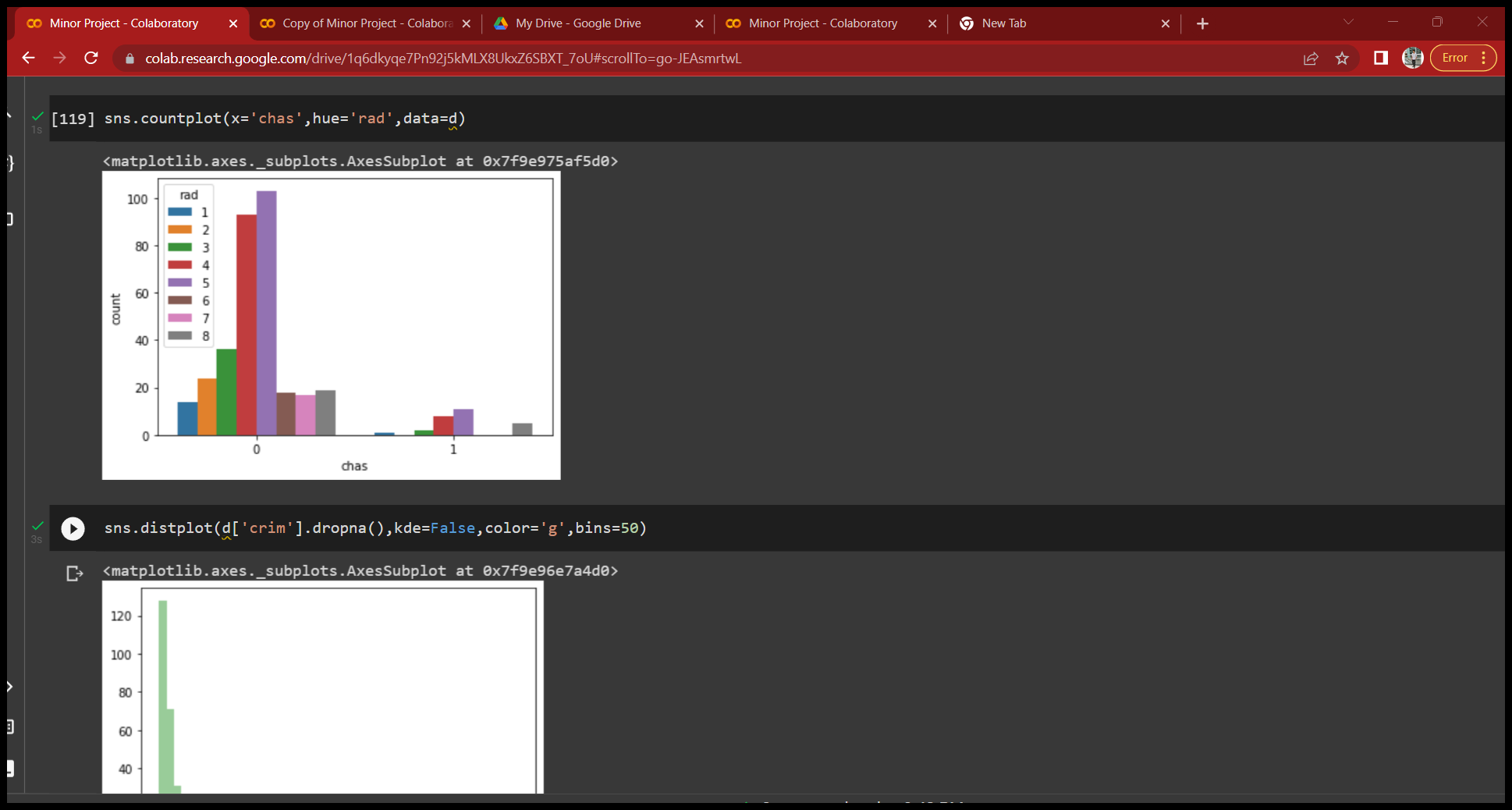
RAD VALUES



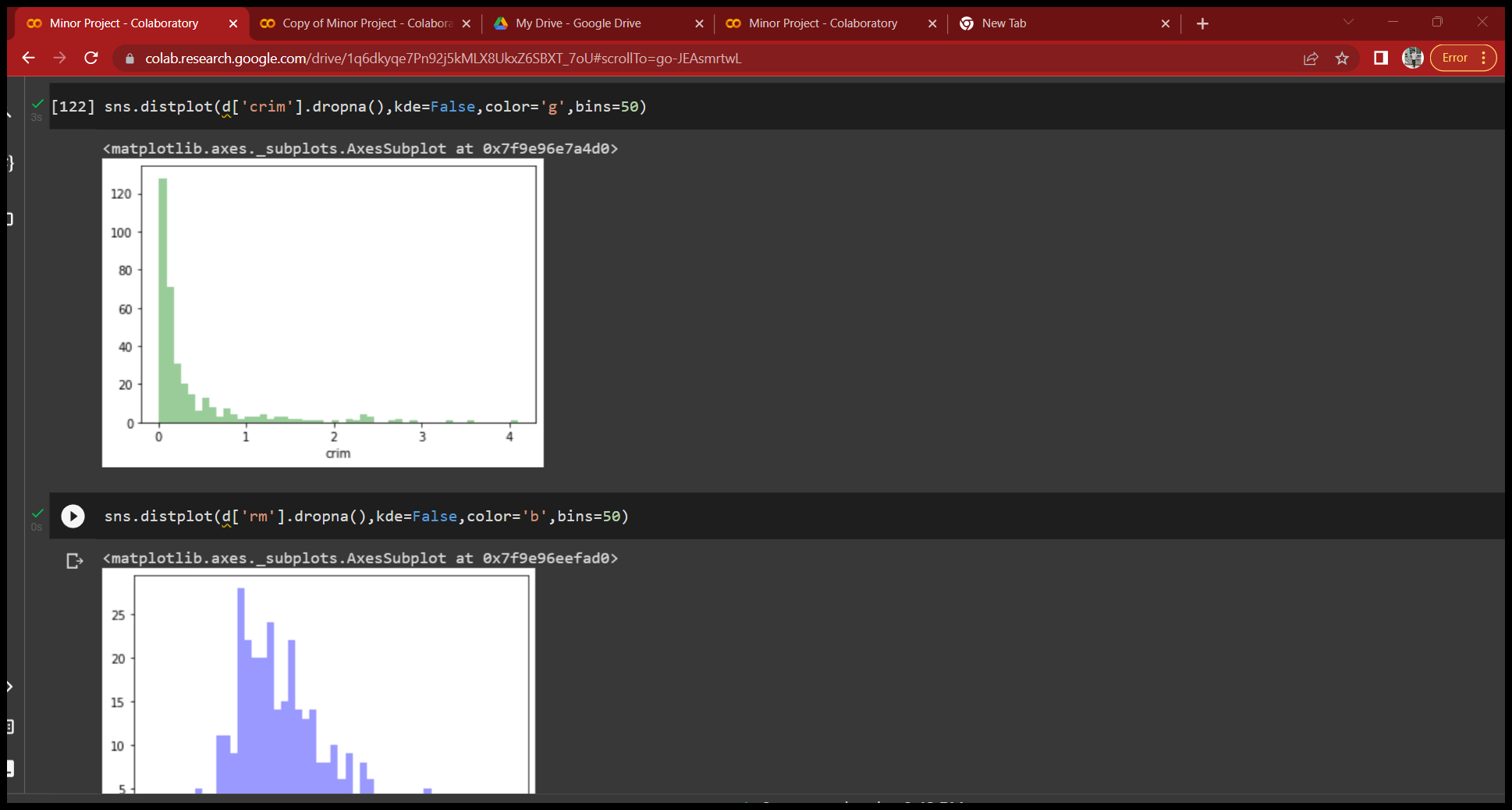
CHAS FEATURES



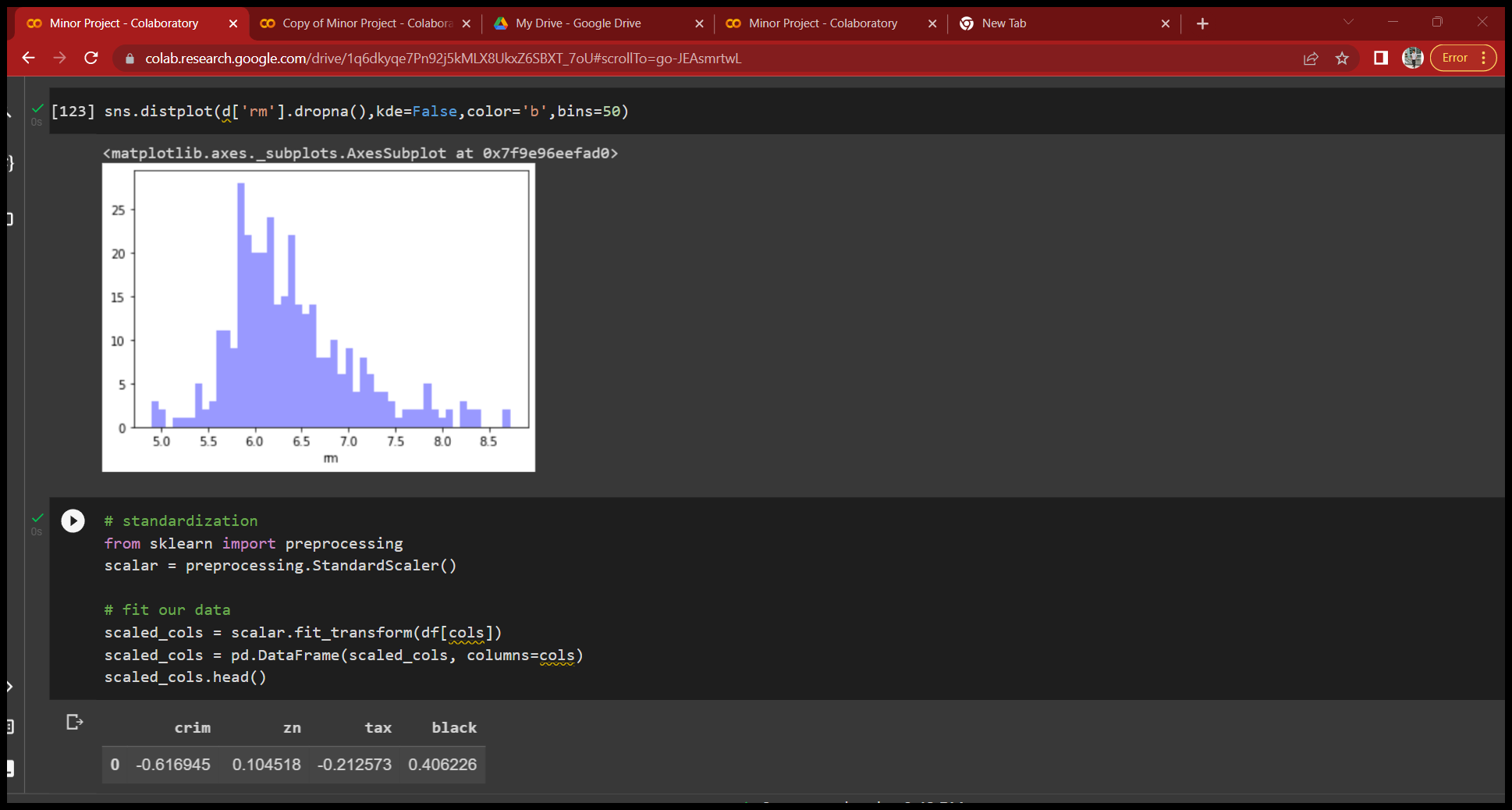
CHAS DATA



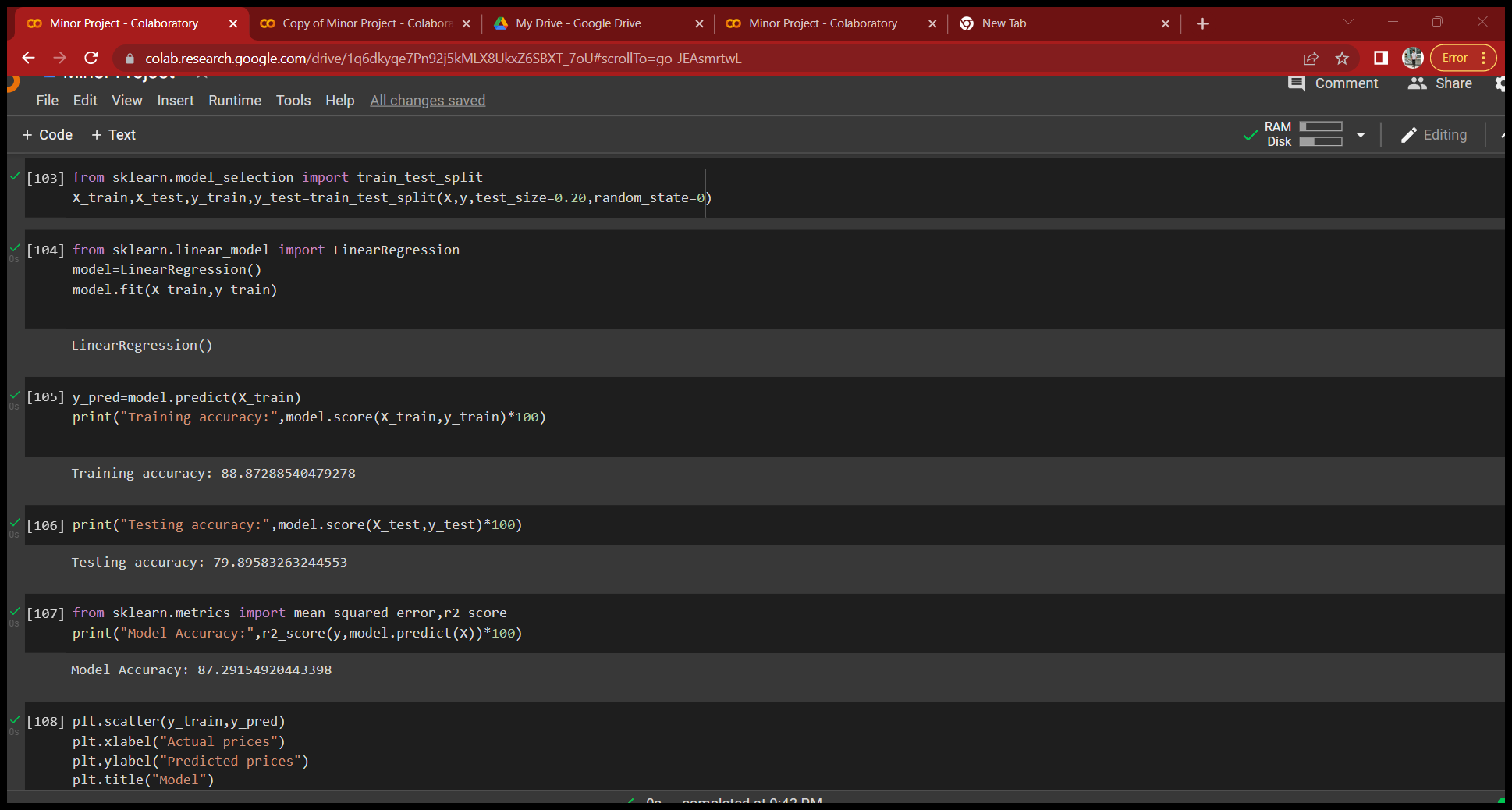
CRIME RATE

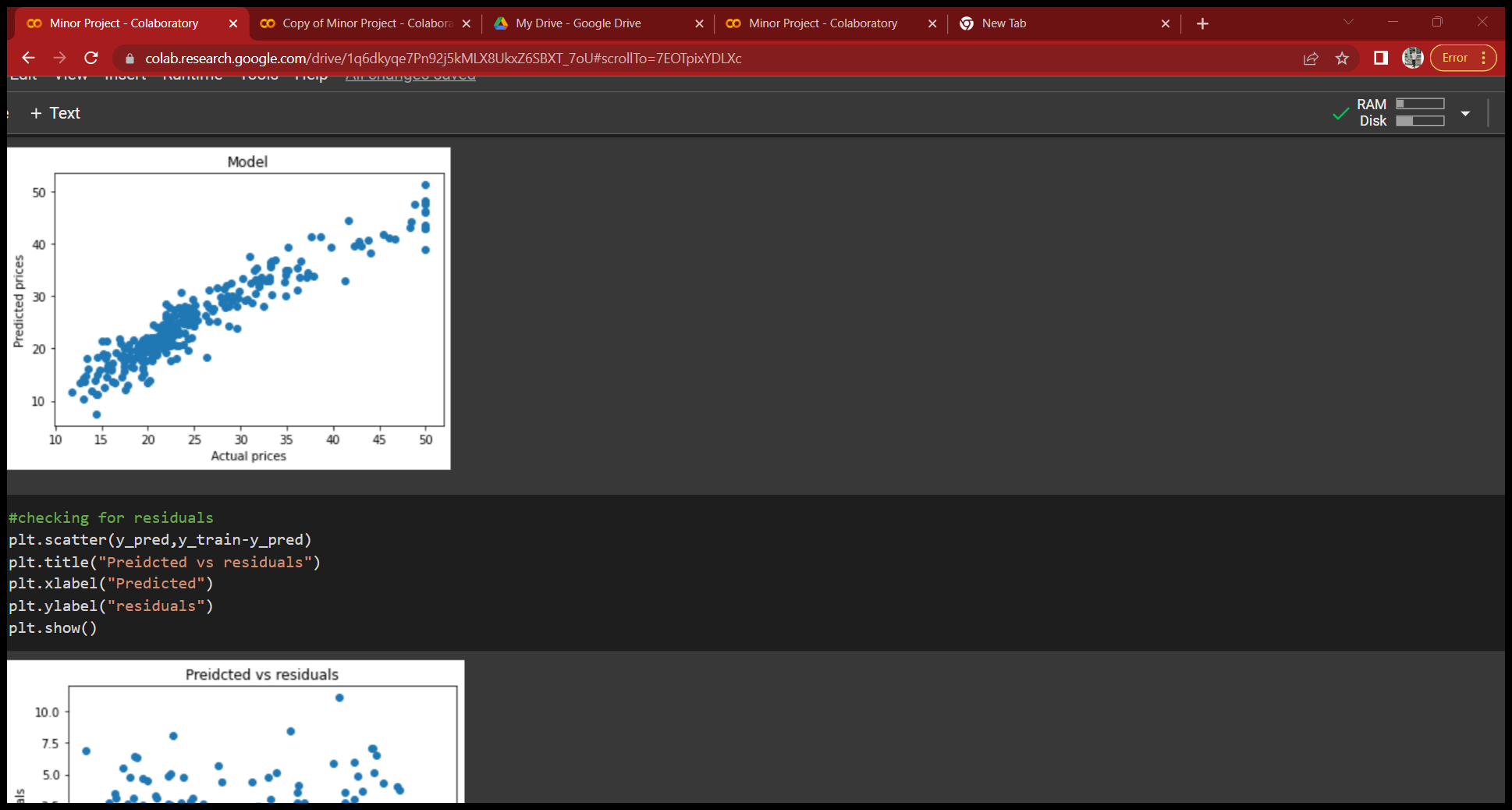


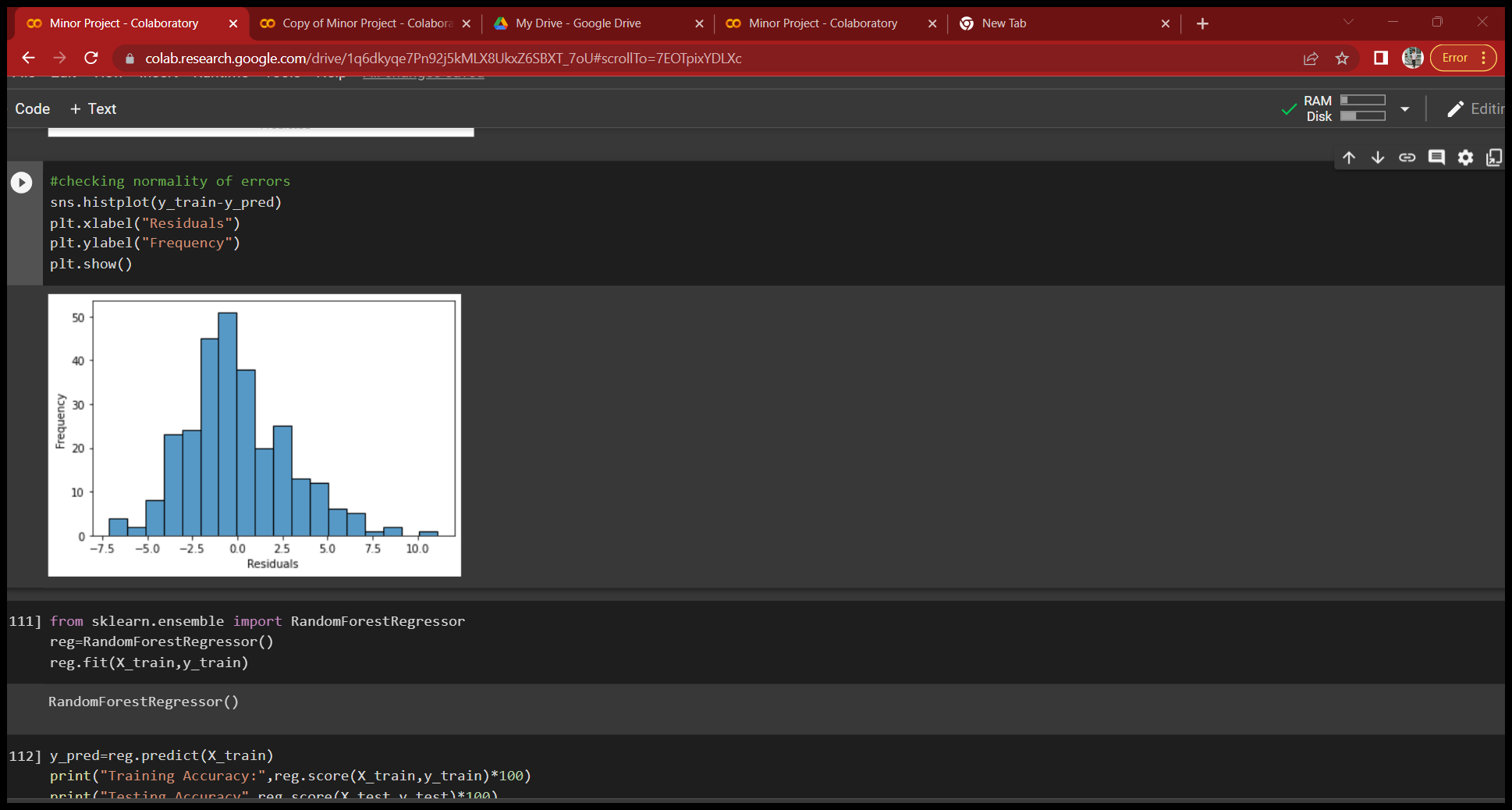
NUMBER OF ROOM IN THE HOUSE

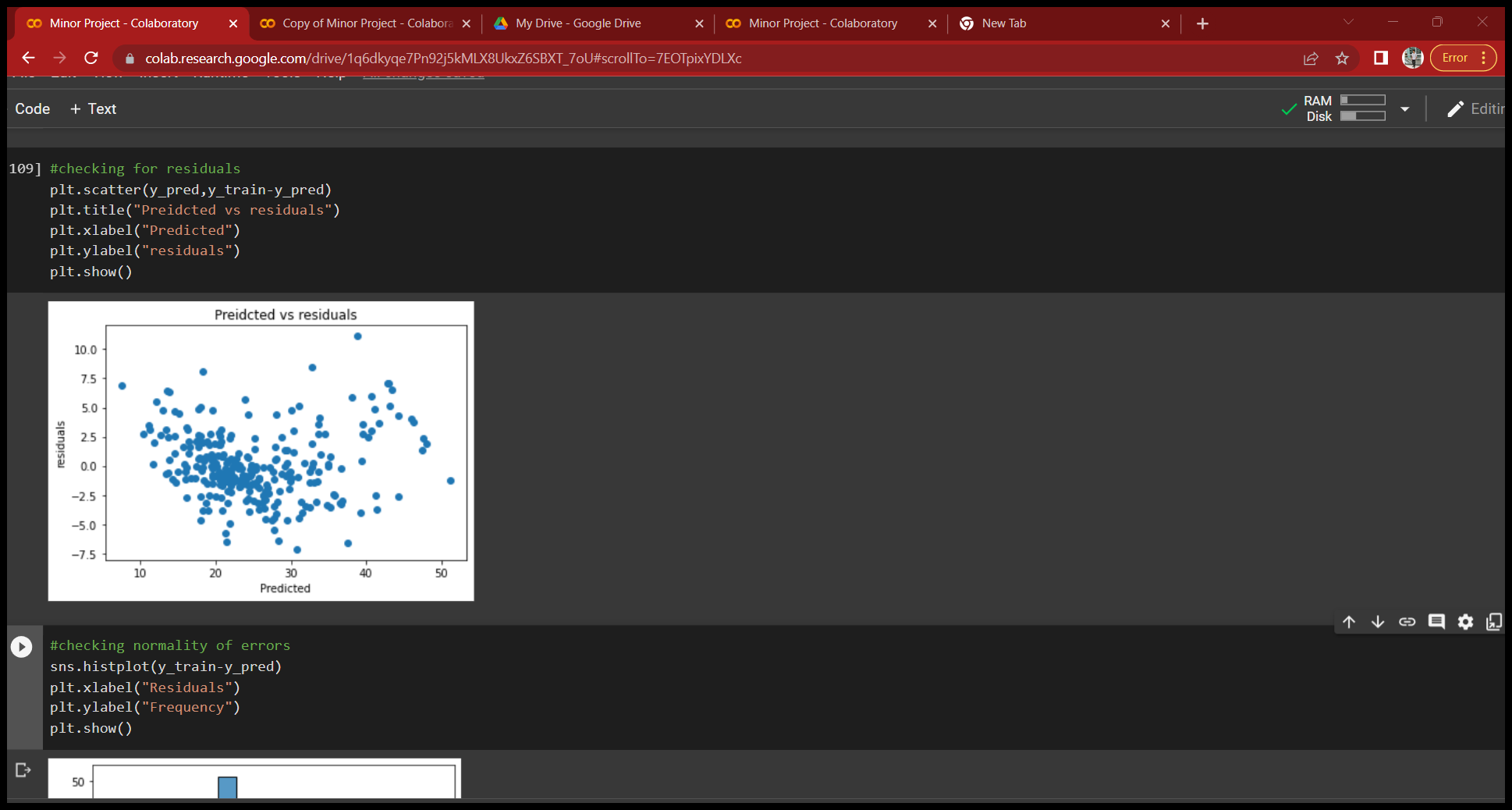


EVALUATION AND COMPARISION

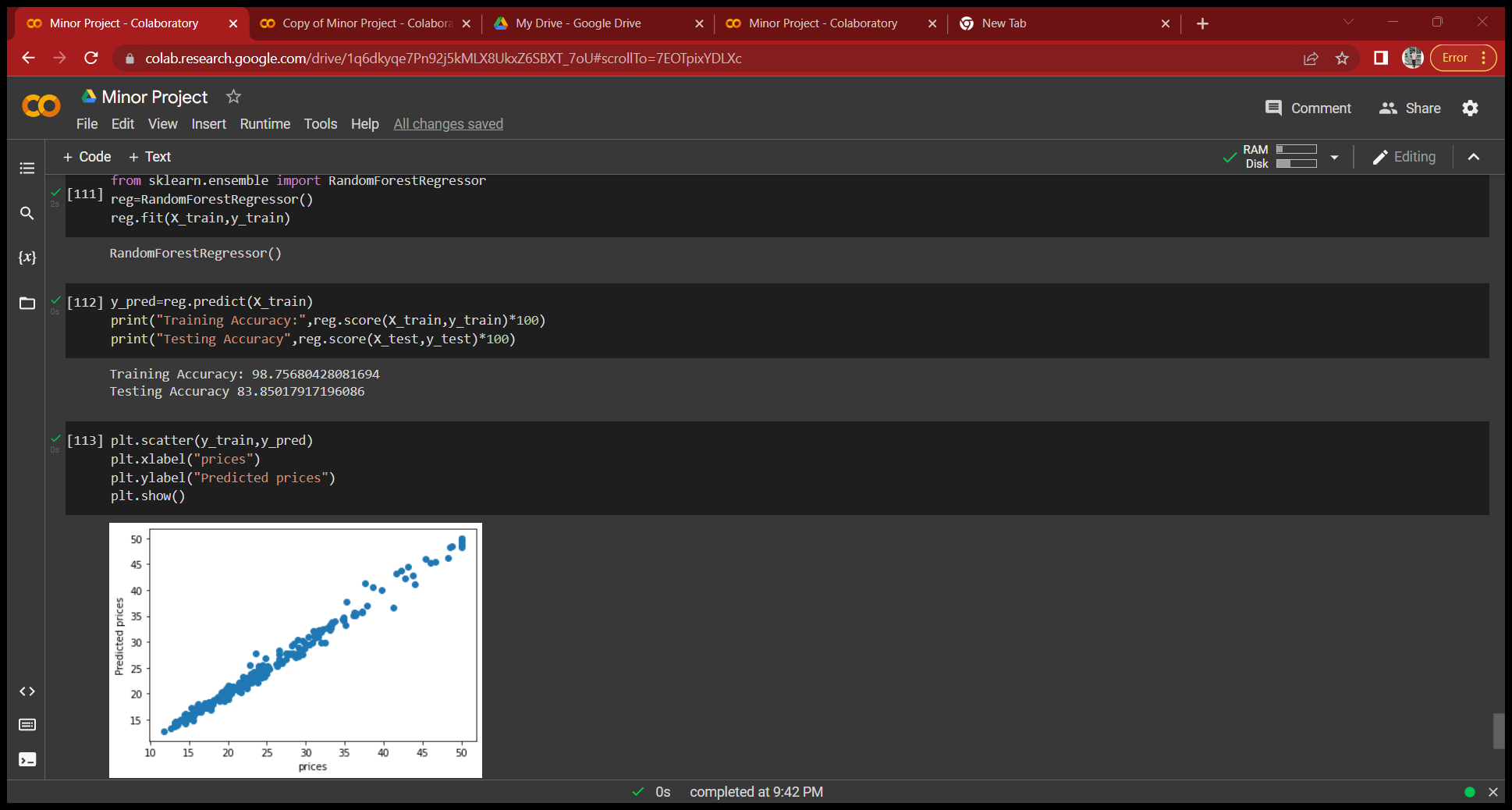








RESULT



LINEAR REGRESSION

Model Accuracy: 87.2%

Training Accuracy: 88.9%

Testing Accuracy: 78.4%

Random Forest Regressor

Training Accuracy: 98.7% Accuracy

Testing Accuracy: 83.8% Accuracy

CONCLUSION

From the Exploratory Information Investigation, we could create knowledge from the information. How every one of the elements connects with the objective. Additionally, it very well may be seen from the assessment of three models that Random Forest Regressor performed better compared to Linear Regression.

References

1.sklearn

2.stackoverflow

3.Tutorialspint

4.Javapoint

5.ge