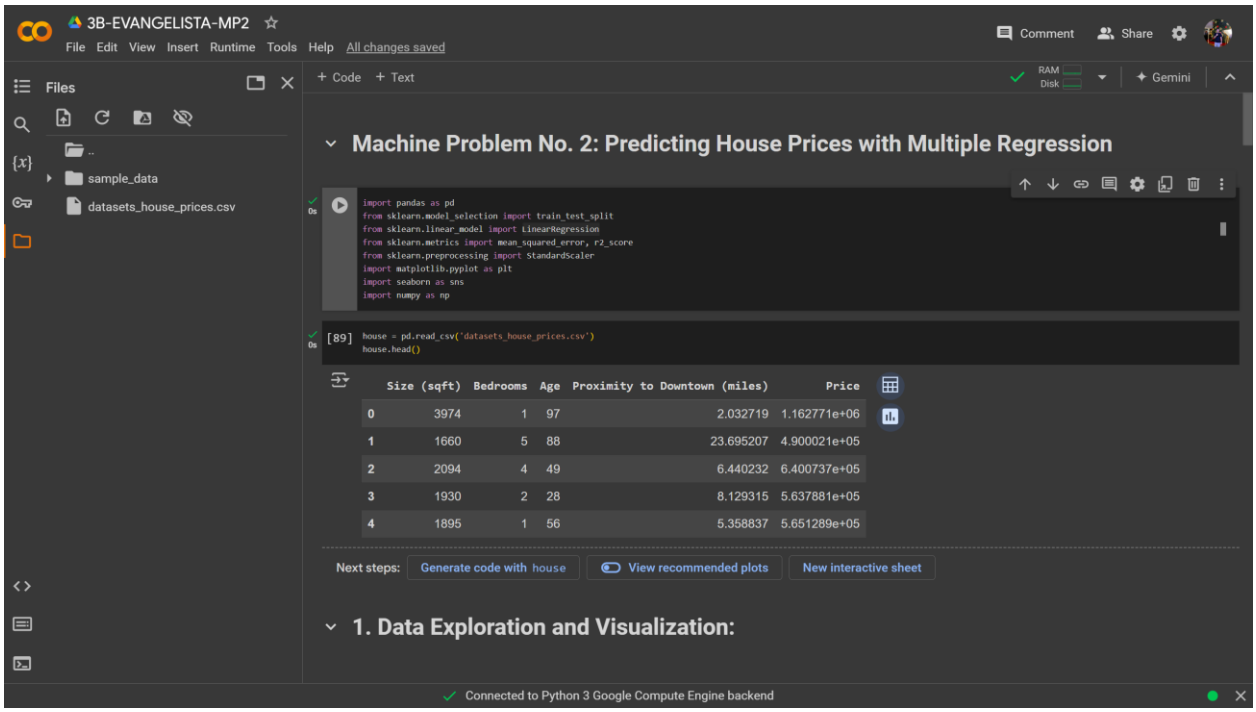
John Kervin D. Evangelista
BSCS-3B-IS

CSST102(Basic Machine Learning)

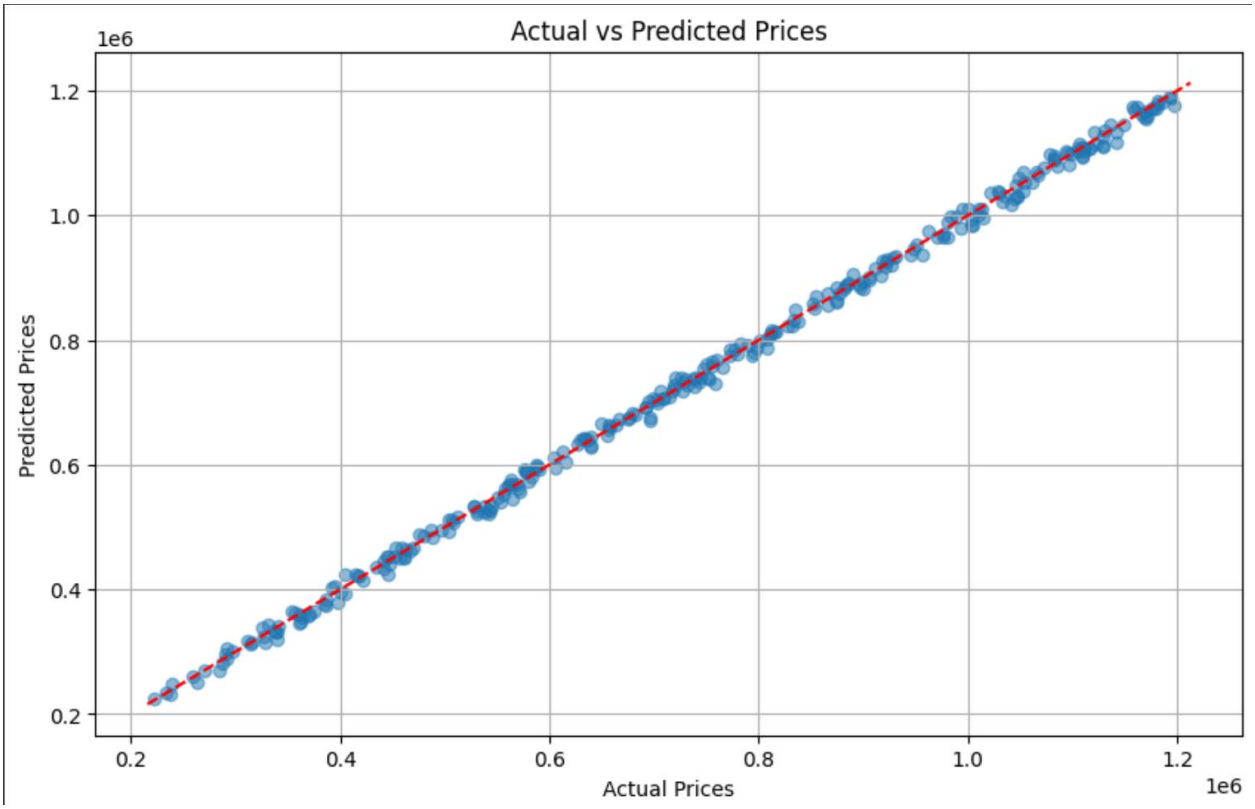
Machine Problem No. 2



The process starts as I get the data and then select what libraries to import that I will use throughout the model building. The dataset was provided by the instructor, it has a name “datasets_house_prices” and it is on a csv file. The data then was processed with the instructions given.

The model creation was aided by the help of my past works which becomes the base of what model I created, also I seek help on Gpt such as Chatgpt and perplexity to be more familiarize with the code and to ensure that what I’m doing is correct. The ai help was also implemented for me to improve my model.

The challenges that I’ve faced when I am creating the model is the unfamiliarity. It’s not that I don’t know what I’m doing, its just that I already forgot everything that has been taught to me so that’s why I overlook my past works to analyze what strategies I need to use to construct the model. With the challenges I face, I also use Gpt’s to seek for help especially when things are too complex and not that easy to understand I use ai to explain in depth to me in order for me to finish the model. On the model itself, the challenges that I’ve faced is the data provided was not that much so I can see that it is no that accurate and not have the score that I will expect.



This plot shows the high accuracy of the model created. The line and the scattered data show how near they are to each other which shows the models consistency. The model also gives scores in performance metrics such as Mean Squared Error (MSE): 100214724.631287, Root Mean Squared Error: 10010.73, R-squared: 0.998631, and Adjusted R-squared: 0.9986128876702134. The metrics indicate that the linear regression model fits the data exceptionally well, as evidenced by high R-squared (0.998631) and Adjusted R-squared (0.998612) values, which suggest that nearly all variability in the dependent variable is explained by the independent variables. Although the Mean Squared Error (100,214,724.63) and Root Mean Squared Error (10,010.73) indicate some prediction error, this may be acceptable depending on the context. Furthermore, the closeness of Adjusted R-squared to R-squared indicates that the inclusion of predictors has not led to overfitting, indicating that all variables are likely relevant for predicting outcomes. Overall, these metrics suggest a robust model with strong explanatory power and relatively low prediction error.

The model is suitable in different life scenarios as it can be use in different fields, it can predict different values depending on the given dataset, the model also can be used to evaluate different things from small to big but it can be sometimes not accurate if the data given is not enough. With that we can not say that the model is applicable in every dataset as it focuses on linear regression which have a wide variety of functions but still not applicable on other data.