# **Documentation**

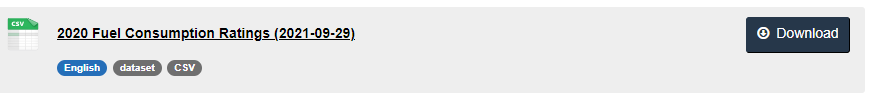
## **Introduction**

The project is about using machine learning to predict C02 emissions from cars based on their build using simple linear regression. Factors such as engine size and fuel consumption can determine how much CO2 a car is likely to emit. Consequently, this information gives a prospective car buyer the ability to contribute towards sustainability using this predictive model.

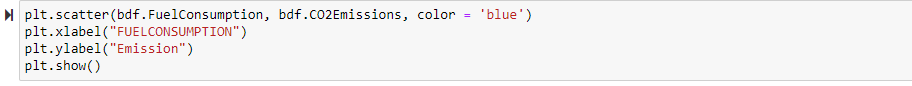
I used a tutorial from Edx titled ‘Machine Learning with Python: A Practical Introduction.

## **Data Collection and Modification**

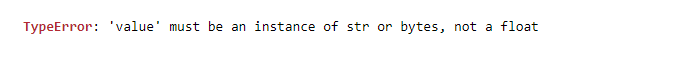
The dataset was sourced from [here](https://open.canada.ca/data/en/dataset/98f1a129-f628-4ce4-b24d-6f16bf24dd64?utm_medium=Exinfluencer&utm_source=Exinfluencer&utm_content=000026UJ&utm_term=10006555&utm_id=NA-SkillsNetwork-Channel-SkillsNetworkCoursesIBMDeveloperSkillsNetworkML0101ENSkillsNetwork20718538-2022-01-01). Specifically, I used the dataset from 2020.



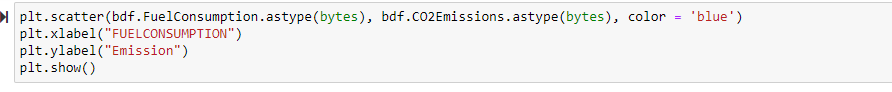
However, I modified this dataset due to errors I encountered along the way. Firstly, I followed the lab instructions and used these lines of code:



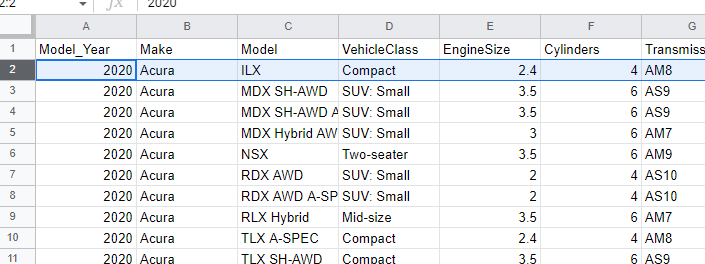
The output after running this cell was the following error message:



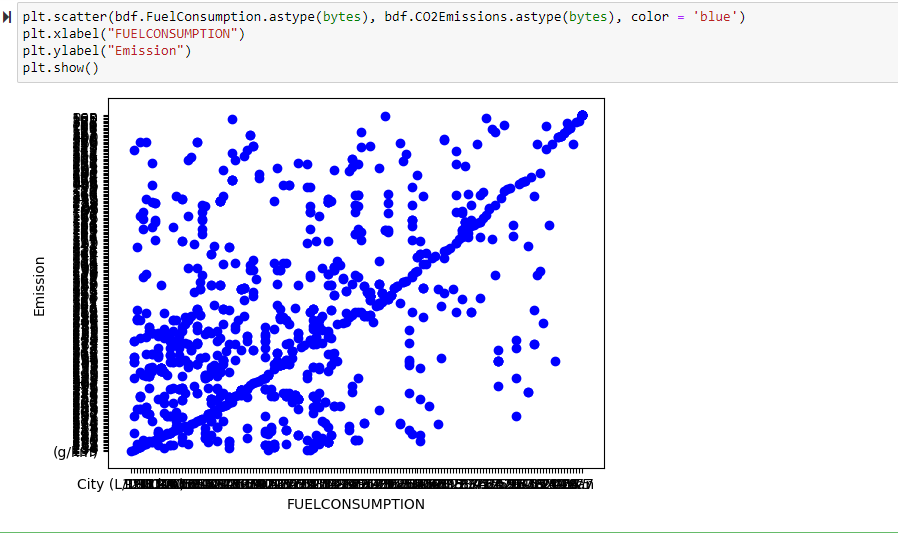
One might think that converting the values to type string or bytes might be a solution, as shown below.



Yes, this will execute the cell successfully, but this change will cause issues later on when training the model. The best solution is to edit the csv file itself. Ensure you delete content in the second row. Overall, it should look like this



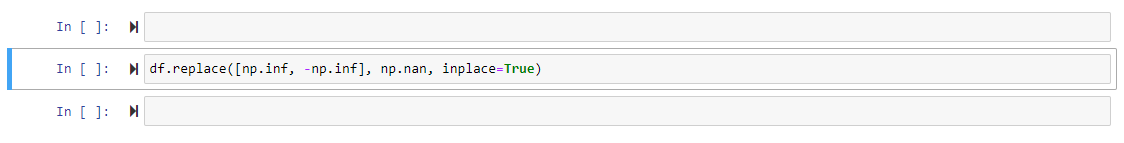
Secondly, another issue you might encounter are diagrams that look like this.



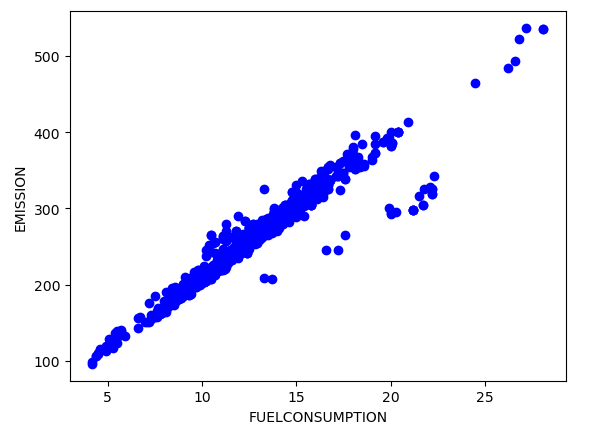
As you may have observed, the values on the axes do not read clearly. Deleting the first row in the csv file is one thing that solves this problem. Another solution is to delete the descriptive information written in the csv file at the very bottom.



Additionally, write the program below to eliminate any null values, because the Scikit library does not allow infinite, NaN or very large values.



Rember, there is no need to convert the types of the parameters to either bytes or string as this will cause problems later on. After doing all that, the graph should look like this:



## **Links**

Find below more important information about this project.

* [Github](https://github.com/PurpleChic/Prediction-using-ML-)
* [Edx course](https://learning.edx.org/course/course-v1:IBM+ML0101EN+2T2021/home)