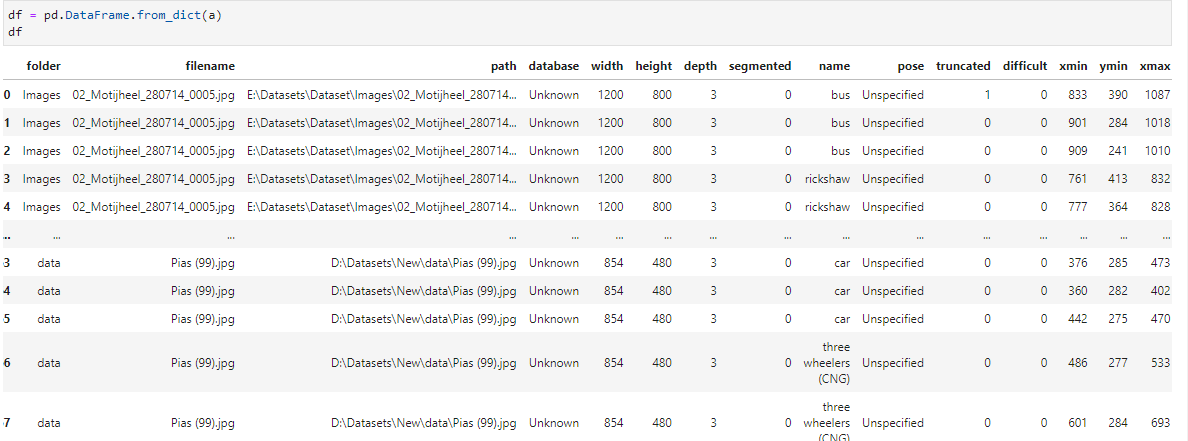
**Analysis**

1) All the xml files in the training dataset were read to analyze about the image

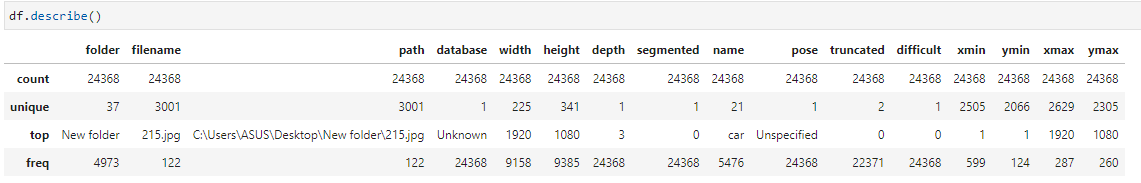
2) In the analysis ‘231.xml’ was found to be not in the required format and hence it was removed

3) Other xml files information was stored in the pandas data frame

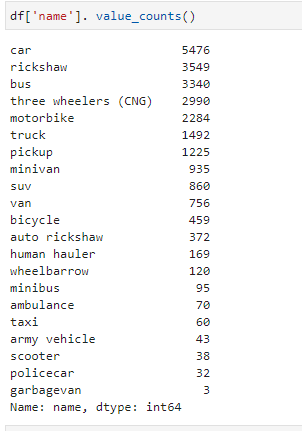


4) The data frame helps to get an idea about the images we are dealing

5) Using describe() we can get an overview of the data distribution



6) Further analysis on number of objects from each class in training dataset was identified



**Training:**

Data Preparation:

1) All the xml files were moved to a separate a separate directory

2) The rest of the images in the directory was spitted as train and test data in the ratio 80:20

3) Now the corresponding xml files is moved back to train and test data folders

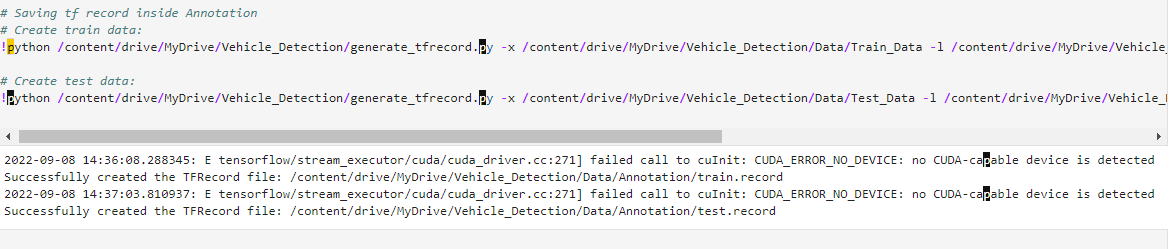
4) A **problem in the dataset** was filename inside the xml file and the image filenames were not matching. So, all the xml files have to be modified.



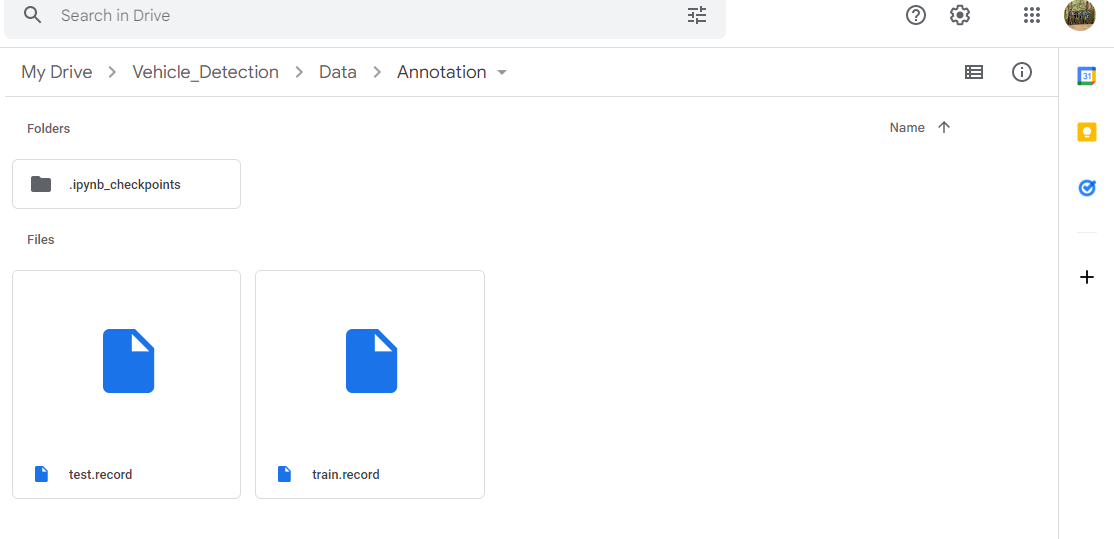
5) TensorFlow repository was cloned and the API was successfully installed.

6) label\_map.pbtxt file was created

7) tfrecord files were created for train and test dataset



Unable to upload the tfrecord file due to its large size.



Further training are pending because of insufficient system requirements and the Google Colab GPU support was also exhausted.

The folder with dataset containing correct xml files, test data, train data and tfrecord can be found in [https://drive.google.com/drive/folders/1iDso3F1XDDdyYgWhw-e963su1nbo9Fe3?usp=sharing](https://drive.google.com/drive/folders/1iDso3F1XDDdyYgWhw-e963su1nbo9Fe3?usp=sharing%20)