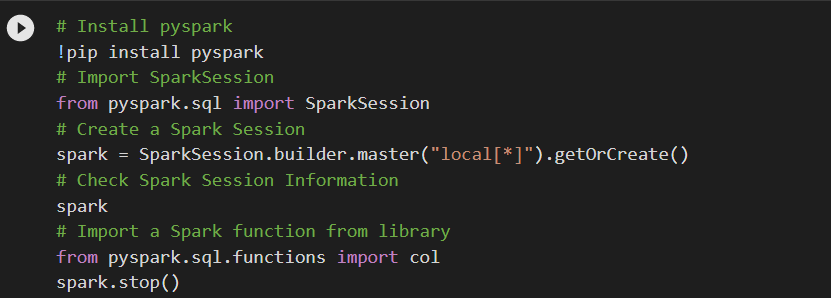
The given file is a google colab file

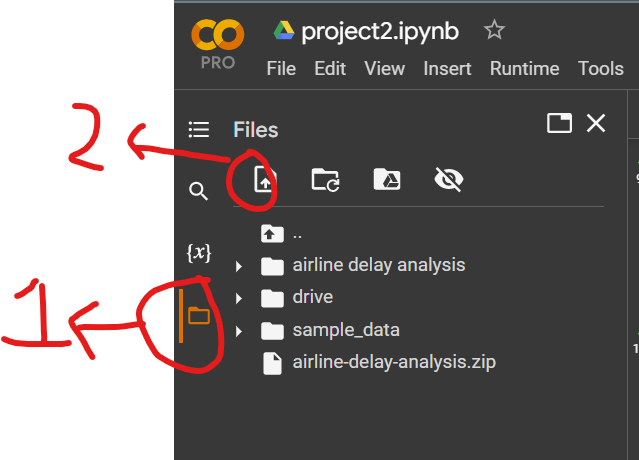
1. First, open the file in google colab and execute first cell to install pyspark library



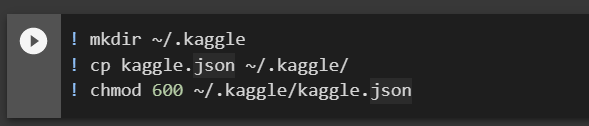
1. Then install Kaggle library by executing second cell given below



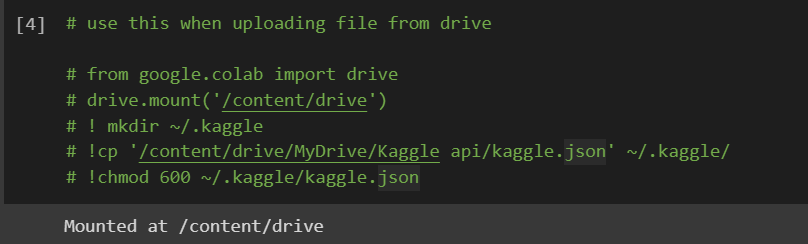
1. After installing the library, upload Kaggle.json file to the colab storage. This file contains the Kaggle api credentials to download the dataset



1. After uploading the file run the third cell to copy the file to a directory and give owner access to read and write over the file.

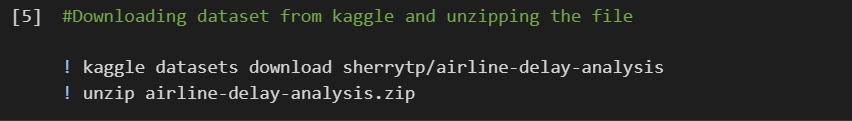


If the file is uploaded in drive, instead of uploading the file to colab, we can skip executing above above cell and run the cell after that to access google drive

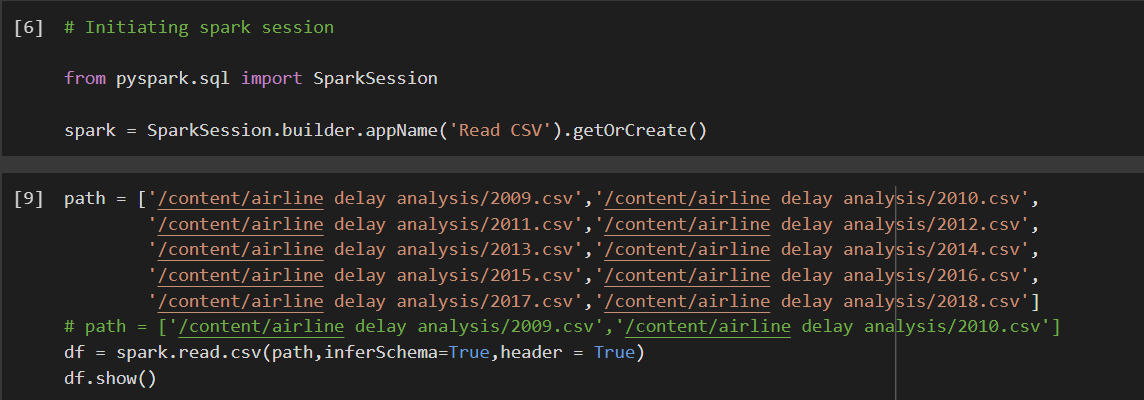


Give the path of the file in drive at 4th line and execute the cell

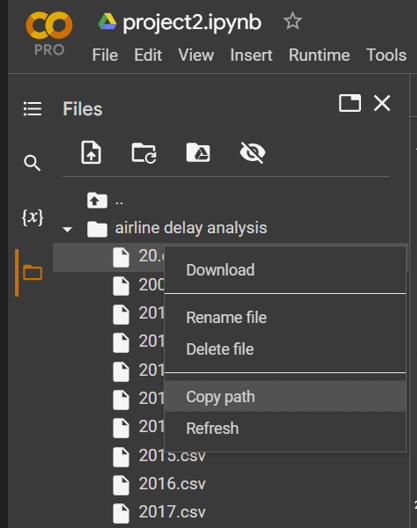
1. The last step before stating the code is to download data, it is downloaded by executing the cell after the above one.



1. Later we can create spark session and create spark dataframe as below



If there is any error when creating the dataframe, please check the path of the files



Go to the folder where files are unzipped and copy path