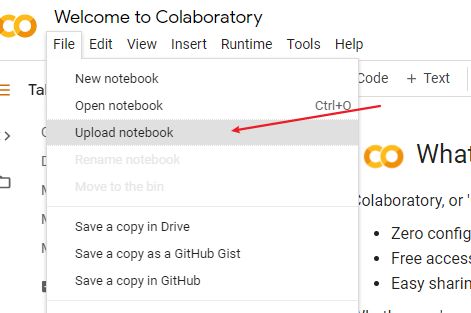
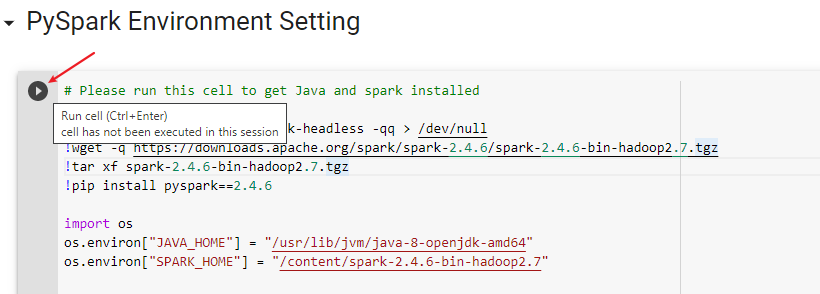
Using Google Colab for Task 3

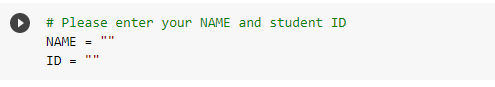
1. Visit and log in to Google Colab site: https://colab.research.google.com/
2. Download the template from Blackboard and then upload it to Colab



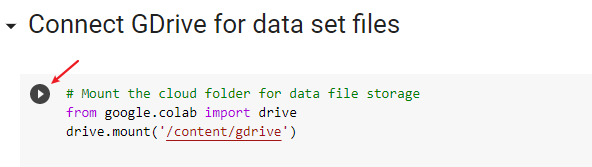
1. Run “PySpark Environment Setting” cell to get spark and pySpark installed.

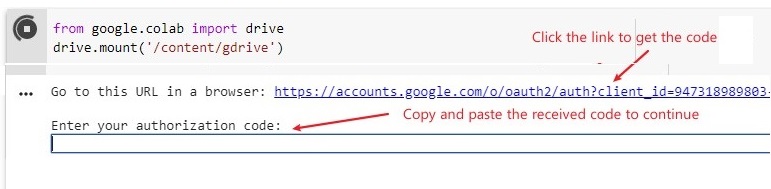


1. Type in your name and id in the first coding cell.

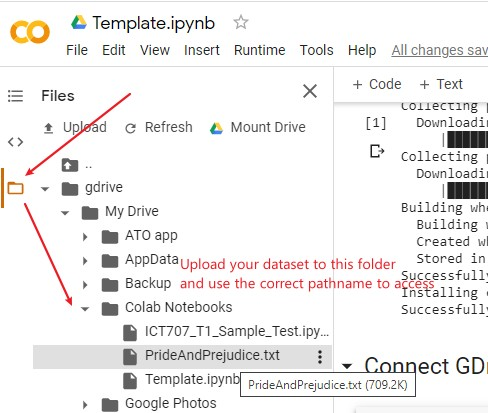


1. Run the first cell of “Connect GDrive for data set files” to mount GDrive as the storage of data files. Follow the instruction to complete the authorization of using GDrive.

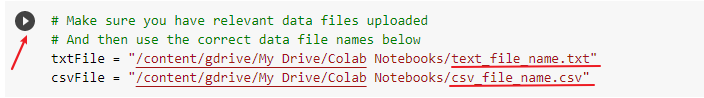




1. Upload data files to GDrive.



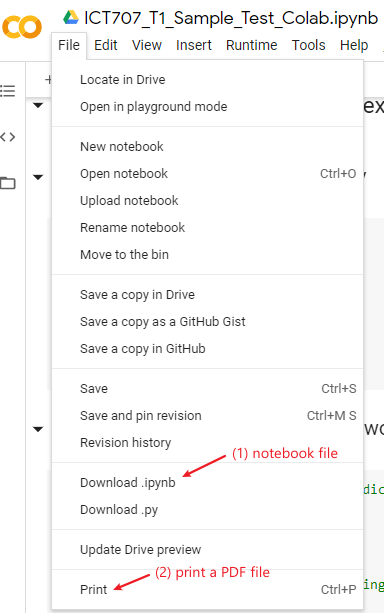
1. Type in the correct data file names and run this cell to get correct file references.



1. Start the coding of your assignment. You can add as many text cells and code cells as you need.



1. After completing the coding and execution, (1) download a notebook file and (2) print a PDF file. Zip these two files for task submission.



Use Google to search for a fix you have any trouble with the usage of Google Colab.