Steps to execute the following commands.

1. Setting up the environment variables to run PySpark on a cluster node.

export PYTHONHOME=/opt/python-2.7.11

export PYTHONPATH=/opt/python-2.7.11

export PATH=$PYTHONHOME/bin:$PATH

1. Cd $SPARK\_HOME- loads the spark home.
2. ./bin/spark-submit /home/013/s/sx/sxa151231/filter.py ---to run the file filter.py
3. Copy the files in the local home or you can upload to hdfs and use the absolute path from there while running the python script.