# Obfuscation of pyspark UDF code with pyarmor

## Steps of obfuscation

1. To install pyarmor, execute following command in command-line

pip install pyarmor

1. Preview of sample directory before obfuscation

A picture containing background pattern

Description automatically generated

1. Prefix UDF methods with `lambda\_`, to avoid error from pyspark `serializers.py`
2. From root\_dir (*obfuscation/* in above example), run following command

pyarmor obfuscate --restrict=0 --recursive .

1. Preview of sample directory after obfuscation



1. Ship only contents of dist/ directory

## Points to remember

1. Pyarmor is both python-version-dependent and platform-dependent. Obfuscation should be done in the right environment with desired python version and target platform.
2. If python files are shipped to Spark workers, following additional steps are to be performed
   * Put *\_pytransform* application bundle (*.dll* or *.so* – as per the platform) in a specific path in spark workers before-hand
   * Add following two lines in the first lines of all *py* files shipped to workers (If they don’t have them already)

from pytransform import pyarmor\_runtime

pyarmor\_runtime("<pre-defined absolute path of dir in worker containing *\_pytransform* app bundle >")

* + Note that pre-defined absolute path of *\_pytransform* application bundle to be mentioned in the second line calling pyarmor\_runtime above.
  + Also, ship the pytransform directory (obtained after the obfuscation step above) to the workers.