```
[>>> exit()
[(base) [mc7787@hlog-2 ~]$ PYTHONSTARTUP=team3/etl_code/cleaning2.py pyspark --deploy-mode client
Python 3.7.9 (default, Feb  5 2021, 09:29:06)
[GCC 4.8.5 20150623 (Red Hat 4.8.5-39)] on linux
Type "help", "copyright", "credits" or "license" for more information.
Setting default log level to "WARN".
To adjust logging level use sc.setLogLevel(newLevel). For SparkR, use setLogLevel(newLevel).
Welcome to
      ____              __
     / __/__  ___ _____/ /__
    _\ \/ _ \/ _ `/ __/  '_/
   /__ / .__/\_,_/_/ /_/\_\   version 2.4.0-cdh6.3.4
      /_/

Using Python version 3.7.9 (default, Feb  5 2021 09:29:06)
SparkSession available as 'spark'.
>>>
```

```
airbnb.csv  airbnb_police.csv  cleaning2.py  cleanings3.py  cleaning4.py  data.csv  X_price.csv  X_review.csv  y_price_real.csv  y_review_bi.c
(base) [mc7787@hlog-2 etl_code]$ cd ..
(base) [mc7787@hlog-2 team3]$ cd ..
(base) [mc7787@hlog-2 ~]$ PYTHONSTARTUP=team3/etl_code/cleaning3.py pyspark --deploy-mode client
Python 3.7.9 (default, Feb  5 2021, 09:29:06)
[GCC 4.8.5 20150623 (Red Hat 4.8.5-39)] on linux
Type "help", "copyright", "credits" or "license" for more information.
Setting default log level to "WARN".
To adjust logging level use sc.setLogLevel(newLevel). For SparkR, use setLogLevel(newLevel).
Welcome to
      ____              __
     / __/__  ___ _____/ /__
    _\ \/ _ \/ _ `/ __/  '_/
   /__ / .__/\_,_/_/ /_/\_\   version 2.4.0-cdh6.3.4
      /_/

Using Python version 3.7.9 (default, Feb  5 2021 09:29:06)
SparkSession available as 'spark'.
[[1.         0.04461898]
 [0.04461898 1.        ]]
>>>
```

```
.35.150:9866,DS-47da2445-0ed0-4739-9157-09dbbf8d3db0,DISK]].
[(base) [mc7787@hlog-2 ~]$ PYTHONSTARTUP=team3/etl_code/cleaning4.py pyspark --deploy-mode client
Python 3.7.9 (default, Feb  5 2021, 09:29:06)
[GCC 4.8.5 20150623 (Red Hat 4.8.5-39)] on linux
Type "help", "copyright", "credits" or "license" for more information.
Setting default log level to "WARN".
To adjust logging level use sc.setLogLevel(newLevel). For SparkR, use setLogLevel(newLevel).
Welcome to
      ____              __
     / __/__  ___ _____/ /__
    _\ \/ _ \/ _ `/ __/  '_/
   /__ / .__/\_,_/_/ /_/\_\   version 2.4.0-cdh6.3.4
      /_/

Using Python version 3.7.9 (default, Feb  5 2021 09:29:06)
SparkSession available as 'spark'.
22/04/26 23:46:45 WARN util.Utils: Truncated the string representation of a plan since it was too large. This behavior can be adjusted by setting 'spark.debug.maxToStr
ingFields' in SparkEnv.conf.
>>>
```