hdfs dfs -rm /CA2/dataset.csv

hdfs dfs -rm -r -f -skipTrash /CA2

hdfs dfs -mkdir /CA2

hdfs dfs -copyFromLocal /home/hduser/CA2/dataset.csv /CA2

hdfs dfs -rm -r -f -skipTrash /Output/Output1

hdfs dfs -copyToLocal /Output /home/hduser/CA2

cd mysql-connector-java-5.1.45/

mysql -u root -p

service mysqld start

pwd

cd /home/hduser/spark/bin/

ls

chmod 777 \*.\*

start-all.sh

jps

nano pythonScript.py

spark-submit --packages org.mongodb.spark:mongo-spark-connector\_2.12:3.0.0 --jars /home/hduser/spark/bin/mysql-connector-java-5.1.45/mysql-connector-java-5.1.45-bin.jar pythonScript.py

http://insideairbnb.com/get-the-data.html

http://insideairbnb.com/dublin/?neighbourhood=&filterEntireHomes=false&filterHighlyAvailable=false&filterRecentReviews=false&filterMultiListings=false