

Варіант №1, пошук найкращих водіїв

Встановлюємо спарк, завантажуюємо бібліотеки pyspark, geopy, faker, numpy.

Завантажуємо London.csv та генеруємо дані за допомогою generator.py

```
93000 codes
94000 codes
95000 codes
96000 codes
97000 codes
98000 codes
99000 codes
100000 codes
101000 codes
102000 codes
103000 codes
104000 codes
105000 codes
106000 codes
107000 codes
108000 codes
109000 codes
110000 codes
111000 codes
112000 codes
113000 codes
114000 codes
115000 codes
```

Активаци  
Чтобы активи

Після цього викликаємо main.py за допомогою spark-submit

```
20/12/16 16:14:19 INFO MapOutputTrackerMasterEndpoint: MapOutputTrackerMasterEndpoint stopped!
20/12/16 16:14:19 INFO MemoryStore: MemoryStore cleared
20/12/16 16:14:19 INFO BlockManager: BlockManager stopped
20/12/16 16:14:19 INFO BlockManagerMaster: BlockManagerMaster stopped
20/12/16 16:14:19 INFO OutputCommitCoordinator$OutputCommitCoordinatorEndpoint: OutputCommitCoordinator stopped!
20/12/16 16:14:19 INFO SparkContext: Successfully stopped SparkContext
20/12/16 16:14:19 INFO ShutdownHookManager: Shutdown hook called
20/12/16 16:14:19 INFO ShutdownHookManager: Deleting directory /tmp/spark-535f2d43-f27a-4948-ba15-d661d913a285/pyspark-e01c9cd6-54f8-40ba-b72e-a158341b708c
20/12/16 16:14:19 INFO ShutdownHookManager: Deleting directory /tmp/spark-535f2d43-f27a-4948-ba15-d661d913a285
20/12/16 16:14:19 INFO ShutdownHookManager: Deleting directory /tmp/spark-4aa6b7de-eb61-4eaf-b596-e90f51e6346a
root@23052bbaa7cd:/home/university/spark# ls -lh results/
total 100K
-rw-r--r-- 1 root root 3.9K Dec 16 14:47 top_drivers.json
```