**BIG DATA COMPUTING 2020/21 – HOMEWORK 2**

Do the tests with the parameters indicated in the first 3 columns of the following table. Fill the table reporting, for each test, 4 values: the approximate average silhouette coefficient of the input clustering (**approxSilhFull**); the time to compute approxSilhFull; the exact silhouette coefficient of the clustering sample (**exactSilhSample**); the time to compute exactSilhSample. The running times must not include the time to load the input file and to compute the sample. The values in the table for Uber\_3\_small.csv must be the average of 3 runs of the same experiment.

**Table to be filled**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Dataset** | **k** | **t** | **approxSilhFull** | **Time (ms) to compute approxSilhFull** | **exactSilhSample** | **Time (ms) to compute exactSilhSample** |
| **Uber\_10\_large.csv** | **10** | **2000** | 0.6283685 | 274989.3333333 | 0.5329840 | 18765.3333333 |
| **Uber\_3\_large.csv** | **3** | **2000** | 0.4913480 | 86607.3333333 | 0.4793789 | 1869.3333333 |
| **Uber\_3\_small.csv** | **3** | **40** | 0.9086760 | 146 | 0.7794451 | 8.3333333 |
| **Uber\_3\_small.csv** | **3** | **80** | 0.9190534 | 182.3333333 | 0.8235615 | 10.6666666 |
| **Uber\_3\_small.csv** | **3** | **1000** | 0.9148195 | 418 | 0.9148198 | 128.3333333 |