**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** | 0.6290736822851846 | 116994 | 0.531309310256696 | 4166 |
| **Uber\_3\_large.csv** | **3** | **2000** | 0.47048619792143037 | 15343 | 0.4707803206444292 | 461 |
| **Uber\_3\_small.csv** | **3** | **40** | 0.9163122319033653 | 53 | 0.8034587642940736 | 2 |
| **Uber\_3\_small.csv** | **3** | **80** | 0.9137061256337747 | 64 | 0.8601300196640022 | 5 |
| **Uber\_3\_small.csv** | **3** | **1000** | 0.9148194891306463 | 105 | 0.9148194891306457 | 69 |