**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** | 0.620502 | 5158 | 0.519944 | 8434 |
| **Uber\_3\_large.csv** | **3** | **2000** | 0.468020 | 2686 | 0.474616 | 896 |
| **Uber\_3\_small.csv** | **3** | **40** | 0.918921 | 253 | 0.796539 | 1 |
| **Uber\_3\_small.csv** | **3** | **80** | 0.913898 | 272 | 0.825127 | 2 |
| **Uber\_3\_small.csv** | **3** | **1000** | 0.914819 | 252 | 0.914819 | 34 |