**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** | 0.6216440865719908 | 303314 | 0.5174593501216961 | 13843 |
| **Uber\_3\_large.csv** | **3** | **2000** | 0.47704697455441925 | 76648 | 0.4769492060725125 | 1410 |
| **Uber\_3\_small.csv** | **3** | **40** | 0.9058177270732621 | 165.66 | 0.8147901654880533 | 18 |
| **Uber\_3\_small.csv** | **3** | **80** | 0.9043919922822139 | 168.33 | 0.8204116300836526 | 29.33 |
| **Uber\_3\_small.csv** | **3** | **1000** | 0.914819489130646 | 589.33 | 0.9148194891306453 | 256 |

Run 1

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Dataset** | **k** | **t** | **approxSilhFull** | **Time (ms) to compute approxSilhFull** | **exactSilhSample** | **Time (ms) to compute exactSilhSample** |
| **Uber\_3\_small.csv** | **3** | **40** | 0.8885295392995511 | 167 | 0.7897131101542643 | 23 |
| **Uber\_3\_small.csv** | **3** | **80** | 0.9077878689384418 | 173 | 0.7997831430828228 | 30 |
| **Uber\_3\_small.csv** | **3** | **1000** | 0.914819489130646 | 580 | 0.9148194891306453 | 194 |

Run 2

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Dataset** | **k** | **t** | **approxSilhFull** | **Time (ms) to compute approxSilhFull** | **exactSilhSample** | **Time (ms) to compute exactSilhSample** |
| **Uber\_3\_small.csv** | **3** | **40** | 0.916793068700049 | 180 | 0.8192254284275061 | 17 |
| **Uber\_3\_small.csv** | **3** | **80** | 0.9045253900666286 | 166 | 0.829717330613579 | 33 |
| **Uber\_3\_small.csv** | **3** | **1000** | 0.914819489130646 | 508 | 0.9148194891306453 | 287 |

Run 3

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Dataset** | **k** | **t** | **approxSilhFull** | **Time (ms) to compute approxSilhFull** | **exactSilhSample** | **Time (ms) to compute exactSilhSample** |
| **Uber\_3\_small.csv** | **3** | **40** | 0.9121305732201861 | 150 | 0.8354319578823894 | 14 |
| **Uber\_3\_small.csv** | **3** | **80** | 0.9008627178415712 | 166 | 0.8317344165545559 | 25 |
| **Uber\_3\_small.csv** | **3** | **1000** | 0.914819489130646 | 680 | 0.9148194891306453 | 287 |