**GROUP NUMBER: 008**

**GROUP MEMBERS: Riccardo Bellan, Davide Cosma, Riccardo Forzan**

**TEST 1:** The goal is to assess the accuracy of the count-sketch estimations as the number of distinct items (regulated by the interval [left,right]) varies. The values D, W, and K are fixed. You must fill in the following table.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **ACCURACY WITH RESPECT TO NUMBER OF DISTINCT ITEMS, USING D=9, W=30, K=10**  Use 4 decimal digits for floating points | | | | |
| **[left,right]** | **Number of received distinct items in [left,right]** | **Average relative error for items with top-K frequencies** | **True normalized F2** | **Approximate normalized F2** |
| [1,15000] | 15000 | 0.0233 | 0.4439 | 0.4438 |
| [1,10000] | 10000 | 0.0099 | 0.3824 | 0.3816 |
| [1,5000] | 5000 | 0.0130 | 0.6135 | 0.6135 |
| [1,1000] | 1000 | 0.0088 | 0.4921 | 0.4924 |

**TEST 2:** The goal is to assess the accuracy of the count-sketch estimations as the number W of columns of the sketch varies. The values D, K and the interval are fixed. Repeat each experiment 3 times. You must fill in the following table.

|  |  |  |  |
| --- | --- | --- | --- |
| **ACCURACY WITH RESPECT TO NUMBER OF COLUMNS W, USING D=9, K=30, [left,right]=[1,10000]**  Use 4 decimal digits for floating points and report averages over 3 runs | | | |
| **W** | **Average relative error for items with top-K frequencies. RUN 1** | **Average relative error for items with top-K frequencies. RUN 2** | **Average relative error for items with top-K frequencies. RUN 3** |
| 100 | 0.0672 | 0.1411 | 0.0419 |
| 50 | 0.1391 | 0.0803 | 0.0453 |
| 20 | 0.3473 | 0.9172 | 1.6620 |
| 15 | 3.8338 | 3.5459 | 3.3866 |

**TEST 3:** The goal is to assess the accuracy of the count-sketch estimations as K varies. The values D, W and the interval are fixed. Repeat each experiment 3 times. You must fill in the following table.

|  |  |  |  |
| --- | --- | --- | --- |
| **ACCURACY WITH RESPECT TO K, USING D=9, W=100, [left,right]=[1,10000]**  Use 4 decimal digits for floating points and report averages over 3 runs | | | |
| **K** | **Average relative error for items with top-K frequencies. RUN 1** | **Average relative error for items with top-K frequencies. RUN 2** | **Average relative error for items with top-K frequencies. RUN 3** |
| 10 | 0.0033 | 0.0040 | 0.0036 |
| 50 | 0.4177 | 0.1452 | 0.4004 |
| 100 | 0.4719 | 0.8749 | 0.3297 |
| 200 | 1.6111 | 1.4278 | 1.9449 |