**GROUP NUMBER: 070**

**GROUP MEMBERS: Alessio Cocco, Andrea Valentinuzzi, Giovanni Brejc**

**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 | 0.0107 | 0.3901 | 0.3898 |
| [1,10000] | 10000 | 0.0074 | 0.4903 | 0.4900 |
| [1,5000] | 5000 | 0.0123 | 0.4047 | 0.4038 |
| [1,1000] | 1000 | 0.0103 | 0.5640 | 0.5642 |

**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 | 0.0462 | 0.0311 | 0.0404 |
| 50 | 0.0996 | 0.0693 | 0.0432 |
| 20 | 0.6986 | 0.3539 | 0.6792 |
| 15 | 1.4153 | 0.8776 | 1.6100 |

**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 | 0.0056 | 0.0033 | 0.0083 |
| 50 | 0.2167 | 0.2155 | 0.1397 |
| 100 | 0.6052 | 0.6539 | 0.4362 |
| 200 | 1.3686 | 1.1461 | 1.0184 |