**GROUP NUMBER:** 023

**GROUP MEMBERS:** Andrea Felline, Marco Sadocco, Mattia Toffanin

**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 | 0.0604 | 0.3269 | 0.3241 |
| [1,10000] | 10000 | 0.0272 | 0.5055 | 0.5051 |
| [1,5000] | 5000 | 0.0218 | 0.4683 | 0.4675 |
| [1,1000] | 1000 | 0.0245 | 0.5947 | 0.5947 |

**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 | 0.3281 | 0.0638 | 0.1394 |
| 50 | 0.5189 | 0.3910 | 0.3405 |
| 20 | 0.7440 | 3.4264 | 2.1304 |
| 15 | 1.2936 | 2.4236 | 3.5258 |

**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 | 0.0138 | 0.0125 | 0.0147 |
| 50 | 0.3670 | 0.1686 | 0.2879 |
| 100 | 0.7040 | 1.3702 | 2.0359 |
| 200 | 2.3278 | 2.5409 | 2.1286 |