**GROUP NUMBER:** 003

**GROUP MEMBERS:** Leonardo Bellin, Maria Teresa Pepaj, Mariafiore Tognon

**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 | 0.0186 | 0.3364 | 0.3368 |
| [1,10000] | 10000 | 0.0111 | 0.3595 | 0.3600 |
| [1,5000] | 5000 | 0.0086 | 0.5569 | 0.5569 |
| [1,1000] | 1000 | 0.0067 | 0.5877 | 0.5875 |

**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 | 0.0521 | 0.0800 | 0.0445 |
| 50 | 0.0948 | 0.0431 | 0.0768 |
| 20 | 0.5716 | 0.3699 | 0.3562 |
| 15 | 0.3435 | 0.4908 | 0.5149 |

**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 | 0.0031 | 0.0046 | 0.0051 |
| 50 | 0.0809 | 0.1047 | 0.1189 |
| 100 | 0.4080 | 0.6408 | 0.3877 |
| 200 | 0.8472 | 0.7280 | 0.5042 |