GROUP NUMBER: 095

GROUP MEMBERS: Dejan Dichoski, Marija Cveevska

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	0.0284	0.4102	0.4099
[1,10000]	10000	0.0365	0.5421	0.5420
[1,5000]	5000	0.0217	0.4462	0.4464
[1,1000]	1000	0.0283	0.5042	0.5043

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	0.1798	0.0832	0.0829
50	1.3895	0.3560	0.4869
20	0.4188	2.1384	3.1475
15	1.2049	0.7139	2.3110

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	0.0090	0.0137	0.0031
50	0.6890	0.8446	0.1013
100	0.2569	0.7079	0.3890
200	0.1556	0.2068	0.1193