Hash Function Performance Report

By: Mohamed Saifeldin Yahia

Introduction:
This report aims to use experimental results to explore the performance of three hash functions when applied to the same 3 supplementary datasets in terms of the average number of collisions per dataset and then in overall.

Analysis:
All the raw data can be seen within the following pages.

Dataset 1: flight-ticket1k.csv

Hash function 2 and 3 yielded the least collisions at 162 each, while hash function 1 yielded 254 collisions.

Dataset 2: flight-ticket10k.csv

Hash function 1 yielded the least collisions at 940, hash function 2 gave 2210 collisions and lastly hash function 3 gave 2214 collisions.

Dataset 3: flight-ticket100k.csv

Hash function 1 yielded the least collisions at 974, hash function 2 gave 2396 collisions and lastly hash function 3 gave 2400 collisions.

Hash function 1 overall performance:

Average collisions for hash function 1 is (254+940+974)/3 = 722.67 collisions

Hash function 2 overall performance:

Average collisions for hash function 2 is (162+2210+2396)/3 = 1589.33 collisions

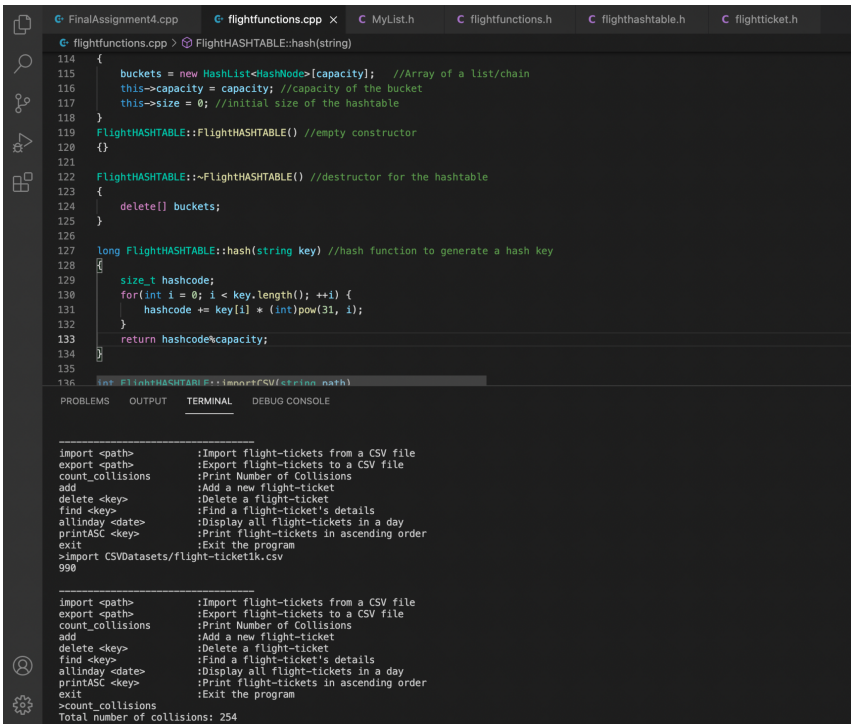
Hash function 3 overall performance:

Average collisions for hash function 3 is (162+2214+2400)/3 = 1592 collisions

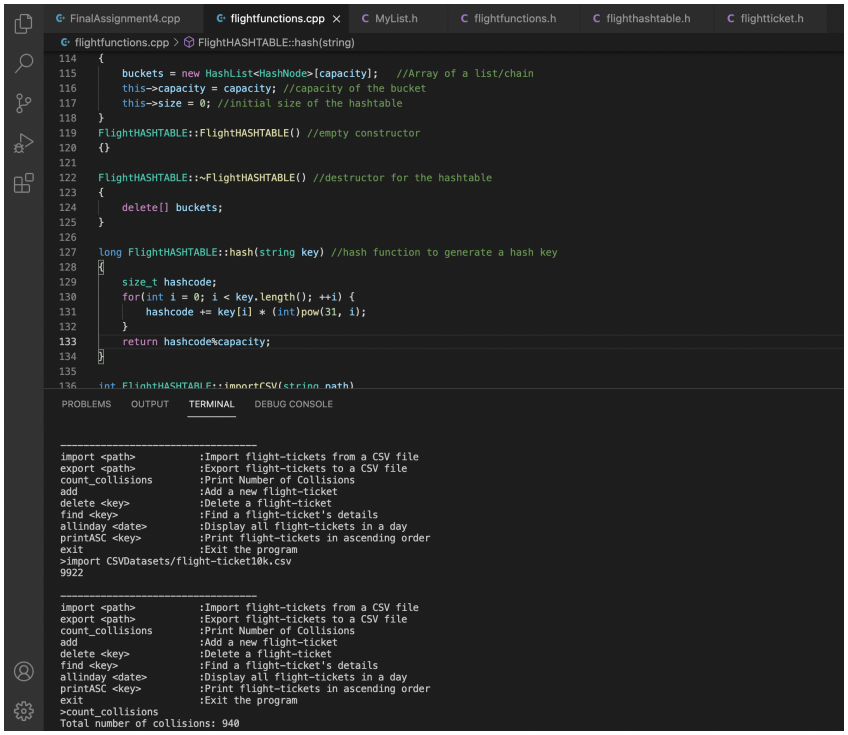
Conclusion:

Despite hash function 1 performing the worst for the first dataset, it yielded the least average collisions by a significant margin across all datasets at 722.67 collisions, hence it is the most effective and will be used for the project.

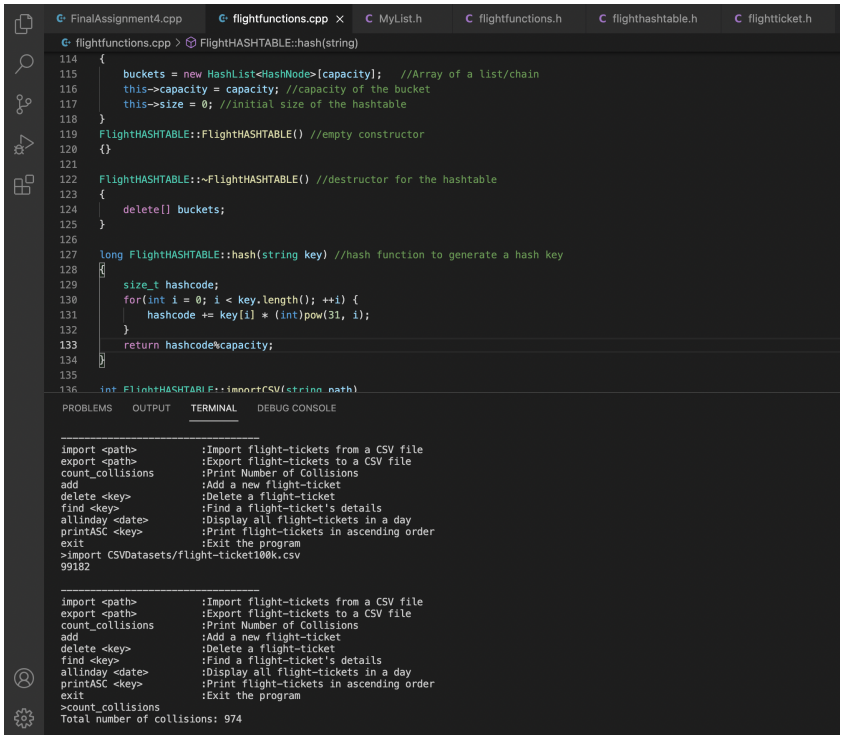
Hash Function 1 applied to dataset with 1k items:



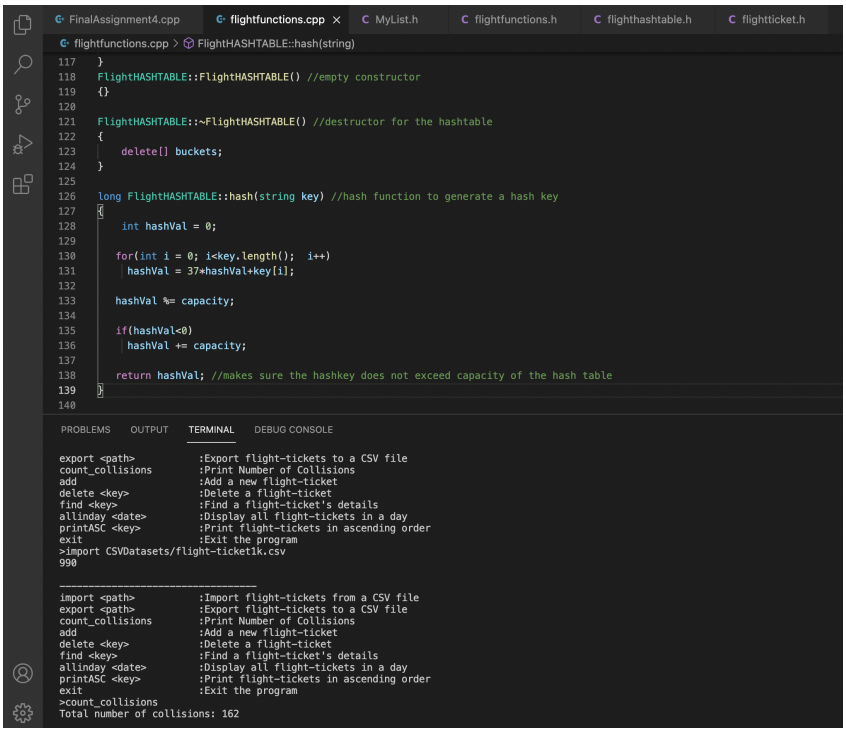
Hash function 1 applied to dataset with 10k items:



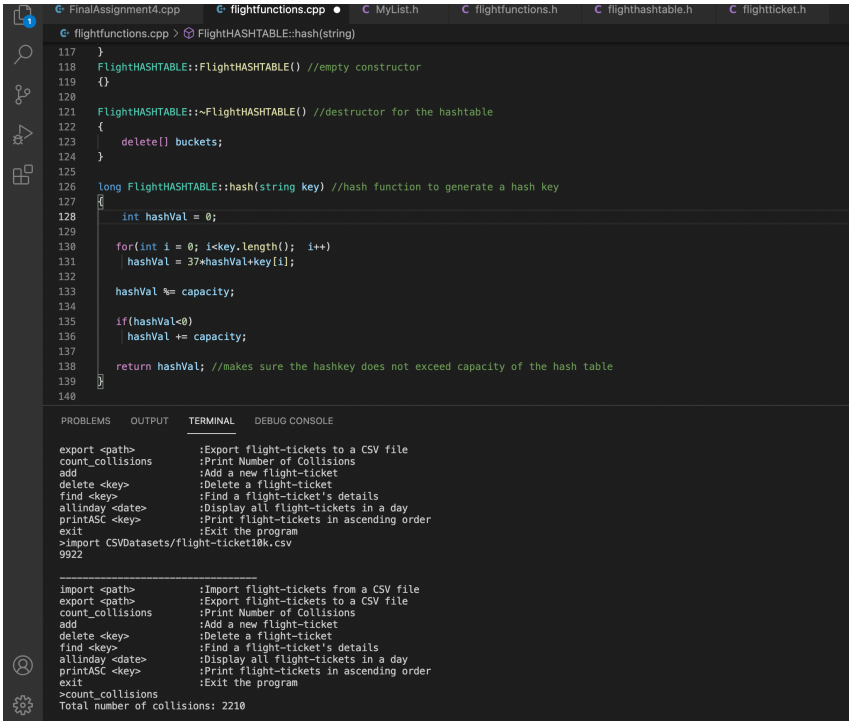
Hash function 1 applied to dataset with 100k items:



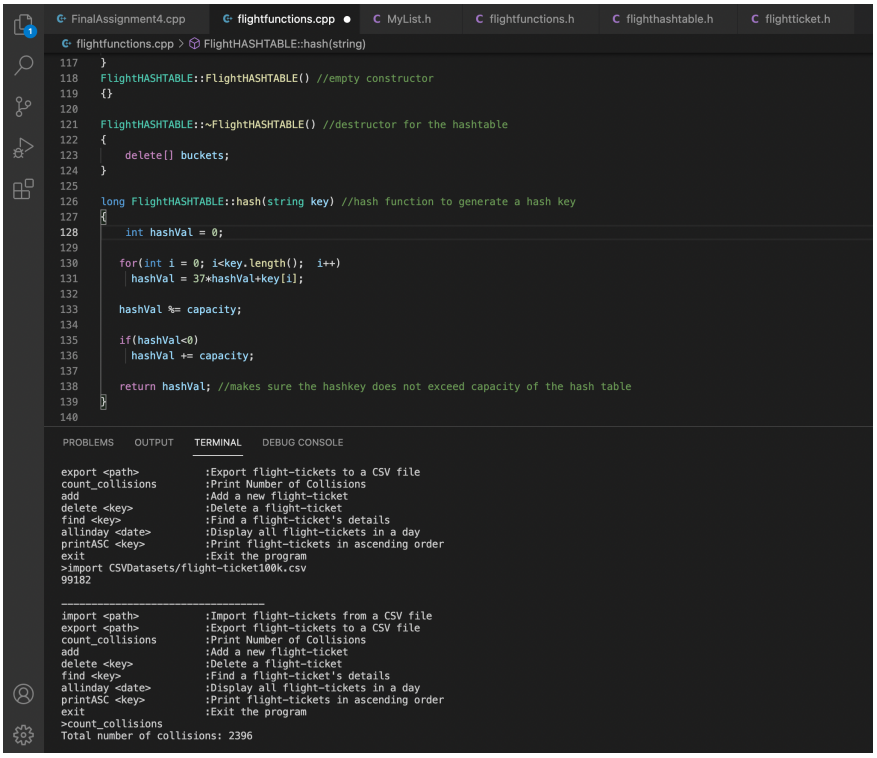
Hash function 2 applied to dataset with 1k items:



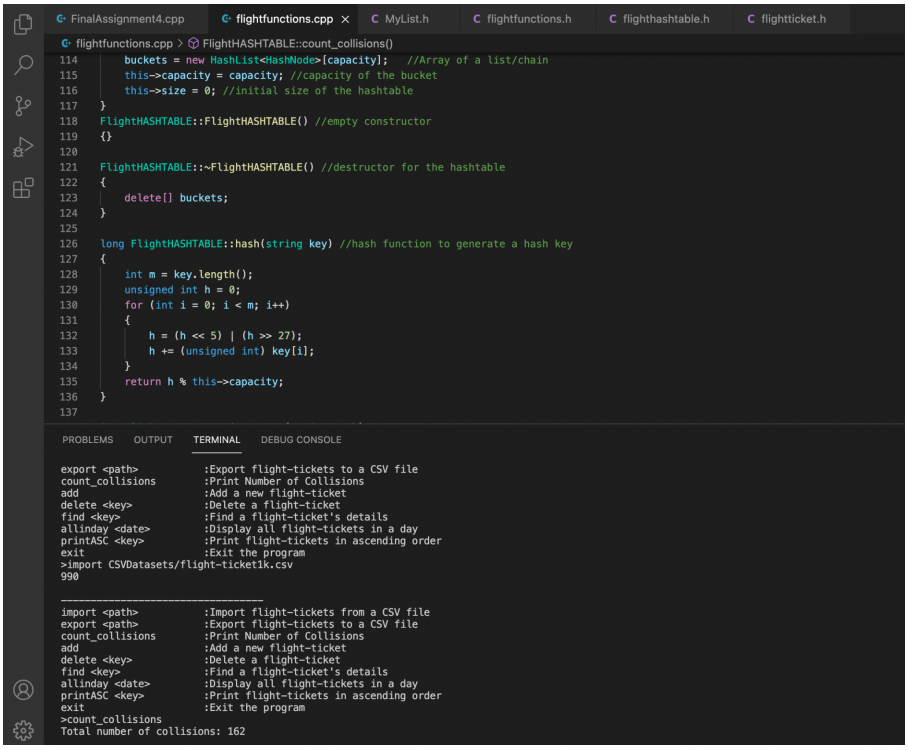
Hash function 2 applied to dataset with 10k items:



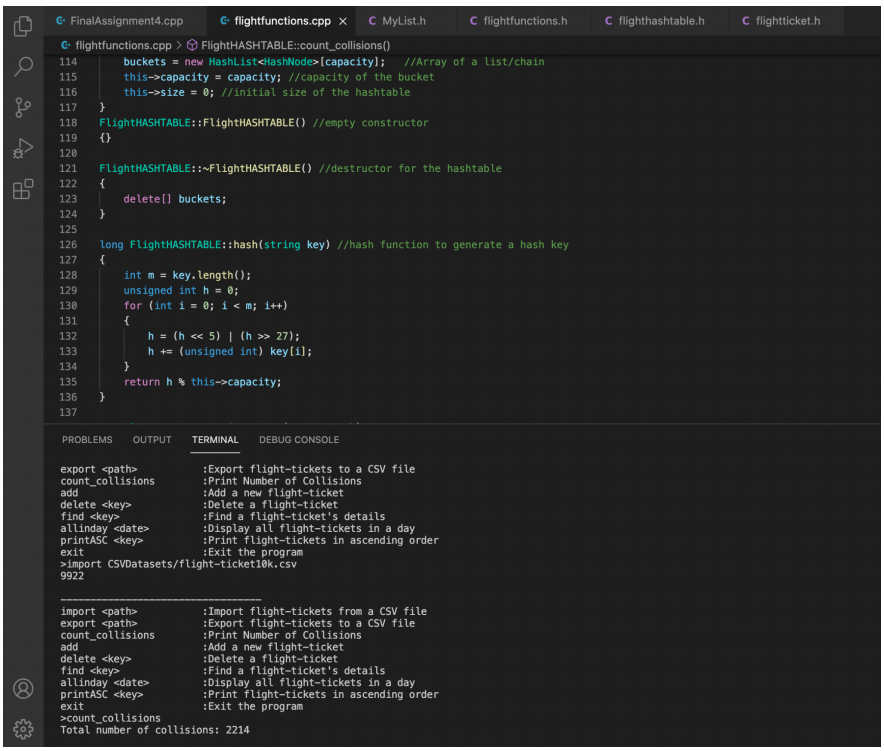
Hash function 2 applied to dataset with 100k items:



Hash function 3 applied to dataset with 1k items:



Hash function 3 applied to dataset with 10k items:



Hash function 3 applied to dataset with 100k items: