Analysis:

Relationships between statistics:

Conflict count is the count of how many times an insertion reached a conflict, this would increase as the number of elements in the table increased. The table is rehashed when it is half full, so the estimated conflict count for a table filled till rehash would be about $1/5^{th}$ of the table size. Probe total would also increase with the number of elements, for a similar case where the table is filled till rehashing i would expect a probe total of 5x the table size as the chain length increases as the table fills resulting in progressively longer probes required. The probe max is the length of the longest chain, if the table was filled to half, this would result in long chains forming, the expected value would be around 1/10th of the table size. The rehash count would only increase when the table is rehashed, so if filled to half capacity the rehash count would be 1.

Fake data:

When feeding in the data from the file aust_cities.txt, provided by Saksham Nagpal, a data set with approximately 1000 elements, into a table of size 2000, the following statistics were returned: (321, 576, 13, 1). The conflict count is near the expected 1/5th but the probe total as well as the max probe are much lower than anticipated. This implies that chains of significant length are forming much less frequently then i expected. The rehash count is 1 as expected.

Analysis of data is preformed in: if __name__ == "__main__" of hash_table.py