This shows the collisions that would occur when hashing this amount of data into the limited available bins. It is clear that the collisions follow a standard bell curve distribution centered at a mean of around 75. This shows and is consistent with the fact that a lot of last names are similar in length and have similar characters so there index value calculates to about the same.