Spark Version: 2.2.1
Scala Version: 2.11.0

Run with spark version 2.2.1 **else will produce weird warnings (Although, program still works).**

Run using same commands present in specifications document.

Analysis of Bloom Filtering:

Modified Bloom Filter to use size thousand bit array instead of million.

Intermediate Result:
Number of correct pred (So far): 195
Number of incorrect pred (So far): 2
False positives: 0.010152284

According to Formula:
P(1 FP) = (1 - e^(-D/T))^(Number hashes)

D = Number hash * Elements to insert = 197
T = Length of Filter = 3

P(1 FP) = 0.088
P(2 FP) = 0.07 (Or 0.088^2)

The theoretical value is bigger than the actual value. The result of the program doesn't conform to the formula. However, this is can be explained away by just the nature of probability and the limited samples (Formula assumes n -> infinity).