Project III report
Due: 3/19
Gabriel Gorospe
Cruzid: ggorospe
SID: 1696580

Compile:
python3 bloom_filter.py

a)

```
gabrielgorospe@gabemg525 Project III data and code % python3 bloom_filter.py
Adding malicious URLS to Bloom Filter
(a) Computing False Positive Rate (FPR) on 10000 Unseen URLs
The false positive is: 0.0359
Please include your Solution for problem (b) in the submitted pdf file.
gabrielgorospe@gabemg525 Project III data and code %
```

b) Storing 1 million URLs in a set would require allocating memory for each URL, which on average is 25 bytes, resulting in a total space requirement of 25,000,000 bytes or 25 MB.

On the other hand, if we use a Bloom filter with k=10 hash functions and m=800,000 buckets to store the same URLs with a 1% false positive rate, the number of bits required per URL can be calculated using the formula: b = -(m / k) * log2(p), where p is the desired false positive rate.

For our scenario, the number of bits required per URL turns out to be approximately 9585.74 bits. Converting this to bytes and then to MB gives us a space requirement of approximately 1.20 MB.