The Bloom filter

The Bloom filter is another famous probabilistic data structure, used for checking membership in a set. Please watch this video about the Bloom filter:
https://www.youtube.com/watch?v=kfFacplFY4Y.
Try to think about the benefits and hinderbacks of this algorithm, and write a few insights about it. No need for a complete or long analysis, just writing a few bulletpoints is sufficient, but please try to be creative and critical.
This is not the main use of the Bloom filter, but to practice similarity search, try the bloom filter you can find in the pybloom_live library. Estimate the cardinalty of the set similarly than we did with hyperloglog. If you feel enthusiastic about the topic, you can implement your own version of the Bloom filter instead of using the one in the library.
After you received the similarity matrix, run the clustering algorithm as we did during practice.
Finally write a few (2-3) sentences about your experiences with Bloom filter.
To be submitted: (until the 20th of March to Balázs via email)

- A document including the insights about Bloom filter, written before trying it out, and the insights after testing it.
- A jupyter notebook with the code you wrote, and the results you got[1]

2025.03.13