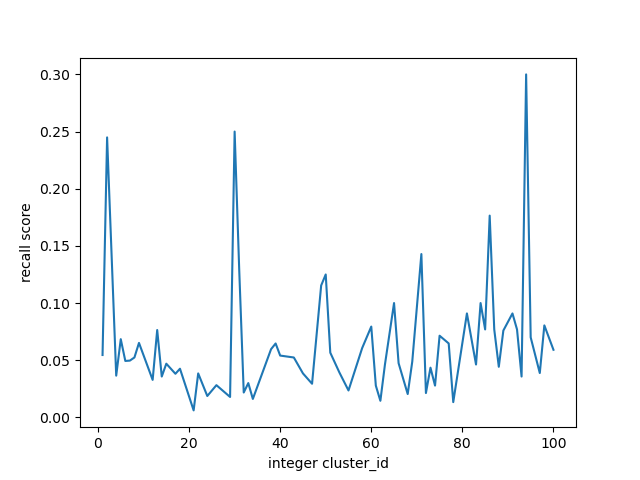
*Unfortunately, due to a lack of computing power, we were not able to answer the questions of the ‘third end product’ for the entire dataset. In order to still give an answer, we therefore took a sample of 1.000 entries from the patstat\_golden\_set table and answered the questions for this sample.*

a)

Chart

Description automatically generatedChart, histogram

Description automatically generated



b)

|  |  |
| --- | --- |
| Average precision | 0.84 |
| Average f1 | 0.11 |
| Average recall | 0.06 |

c)

A cluster with really good (the best so far actually) is cluster 94 with an f1\_measure of 0.46

A cluster with really poor (the worst so far actually) is cluster 21 with an f1\_measure of 0.01

The table score\_cluster contains all the data on precision, f1 and recall for all the clusters