**Examples of duplicates in test set**

> table(counts)

counts

1 2 3 4 5 6 7 8

57513 9931 3695 1004 218 30 5 1

8 case

|  |  |  |
| --- | --- | --- |
| filename | filesize | md5 |
| 10084 | 1648 | 44ea3b609860c3f538f721be0b54ecfb |
| 117336 | 1648 | 44ea3b609860c3f538f721be0b54ecfb |
| 120878 | 1648 | 44ea3b609860c3f538f721be0b54ecfb |
| 137310 | 1648 | 44ea3b609860c3f538f721be0b54ecfb |
| 159251 | 1648 | 44ea3b609860c3f538f721be0b54ecfb |
| 3105 | 1648 | 44ea3b609860c3f538f721be0b54ecfb |
| 39527 | 1648 | 44ea3b609860c3f538f721be0b54ecfb |
| 56867 | 1648 | 44ea3b609860c3f538f721be0b54ecfb |

|  |  |  |  |
| --- | --- | --- | --- |
|  |  |  |  |
|  |  |  |  |

One of 7 cases

|  |  |  |
| --- | --- | --- |
| filename | filesize | md5 |
| 11574 | 916 | 1852ad677e8bf3556199c165e2a80a60 |
| 12828 | 916 | 1852ad677e8bf3556199c165e2a80a60 |
| 19976 | 916 | 1852ad677e8bf3556199c165e2a80a60 |
| 36521 | 916 | 1852ad677e8bf3556199c165e2a80a60 |
| 4847 | 916 | 1852ad677e8bf3556199c165e2a80a60 |
| 49518 | 916 | 1852ad677e8bf3556199c165e2a80a60 |
| 59124 | 916 | 1852ad677e8bf3556199c165e2a80a60 |

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| |  |  |  |  | | --- | --- | --- | --- | |  |  |  |  | |  |  |  |  | |  |