Thresholds Summary (Final Logic)

This document summarizes the threshold-based deduplication results.  
We define:  
 - Full Analysis: no repeated entries + maximum unique\_count (tie → highest threshold)  
 - Core-Level Analysis: no repeated entries + row\_count ≤ 9 (tie → highest threshold)  
All other thresholds are 'partial analysis'.  
The combined plot no longer includes any 'Row=Unique' annotations.  
Vertical lines for Full/Core thresholds are drawn only from y=0 to the maximum of the three metrics at that threshold, rather than the entire axis.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| threshold | row\_count | unique\_count | repeated\_count | no\_repeats\_bool |
| 0.0 | 1 | 1 | 0 | True |
| 0.05 | 1 | 1 | 0 | True |
| 0.1 | 1 | 1 | 0 | True |
| 0.15 | 1 | 1 | 0 | True |
| 0.2 | 1 | 1 | 0 | True |
| 0.25 | 1 | 1 | 0 | True |
| 0.3 | 2 | 2 | 0 | True |
| 0.35 | 2 | 2 | 0 | True |
| 0.4 | 4 | 4 | 0 | True |
| 0.45 | 5 | 5 | 0 | True |
| 0.5 | 6 | 6 | 0 | True |
| 0.55 | 6 | 6 | 0 | True |
| 0.6 | 6 | 6 | 0 | True |
| 0.65 | 7 | 7 | 0 | True |
| 0.7 | 7 | 7 | 0 | True |
| 0.75 | 8 | 8 | 0 | True |
| 0.8 | 10 | 10 | 0 | True |
| 0.85 | 10 | 10 | 0 | True |
| 0.9 | 10 | 10 | 0 | True |
| 0.95 | 11 | 10 | 2 | False |
| 1.0 | 11 | 10 | 2 | False |

Chosen thresholds:

- Full Analysis threshold = 0.90.

- Core-Level Analysis threshold = 0.75.

All other thresholds that don't meet either condition are 'partial analysis'.  
If Full and Core turned out the same, we unify them with one line labeled 'Full+Core'.