DDM-Spark (following the paper)

Read data from csv →

Create tubles of (value, colume name)
Convert columns to sets → Convert to rdd → Zip col + flatMap → Reduce by key (col names) → Convert to DS

→ Partition globally with union

Create rdd → Build attribute sets (reduce by key) → Build inclusion Lists [Map over col names / Filter this col name] → Intersect and group By key → Remove empty values → Convert to string → Collect → Sort lines and print