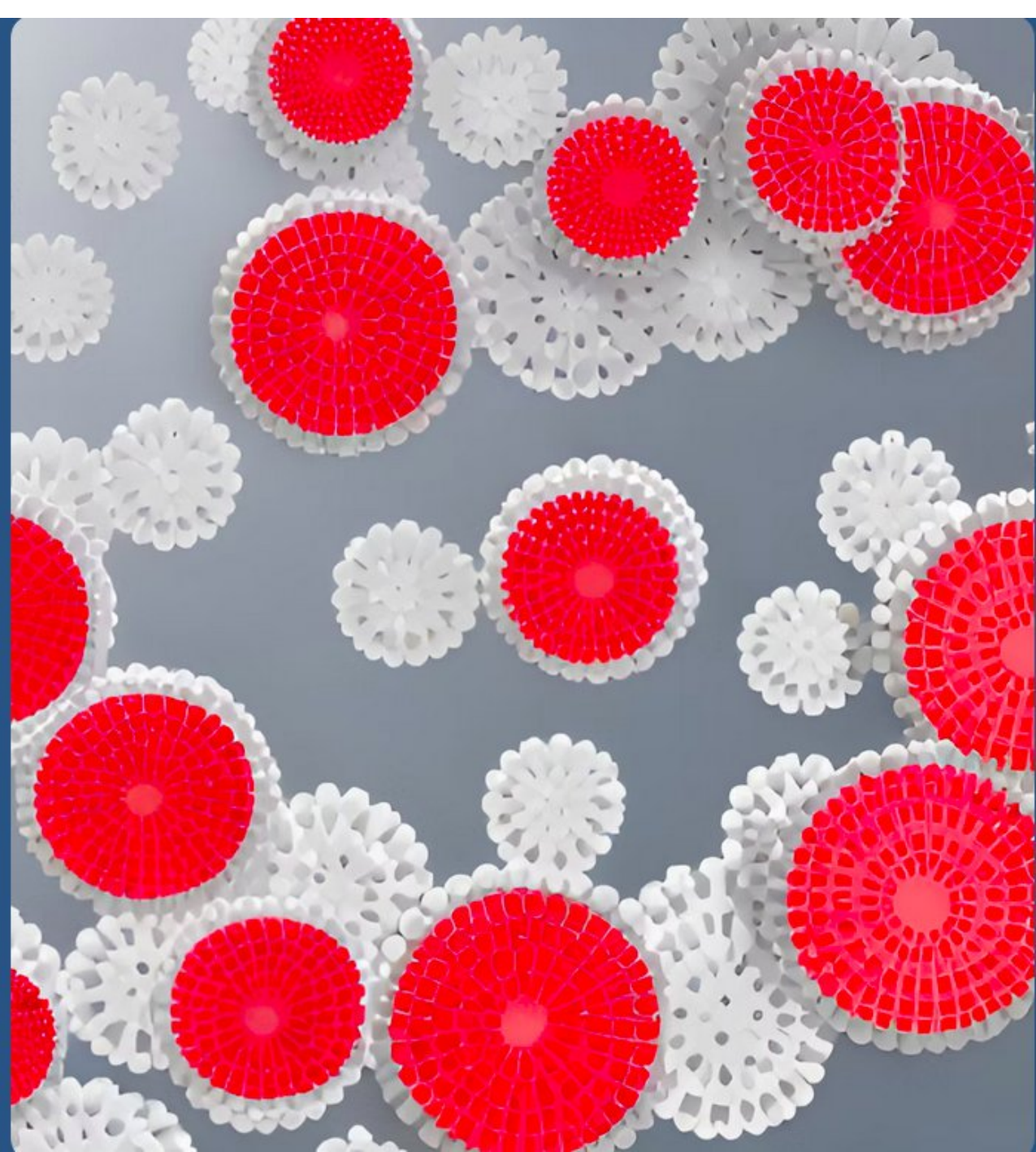# Data Integration:
2. Integration

before Integrating Data our records had some Problems:

1. Duplicate Records

2. No Identifiable Keysets

3. Hetrogenity of Number formats

4. unterschiedliche Zeichensätze (ASCII vs. UTF-8)

Processing Pipeline

1. Reading Files

2. Creating native Java Objects to Model Data sets
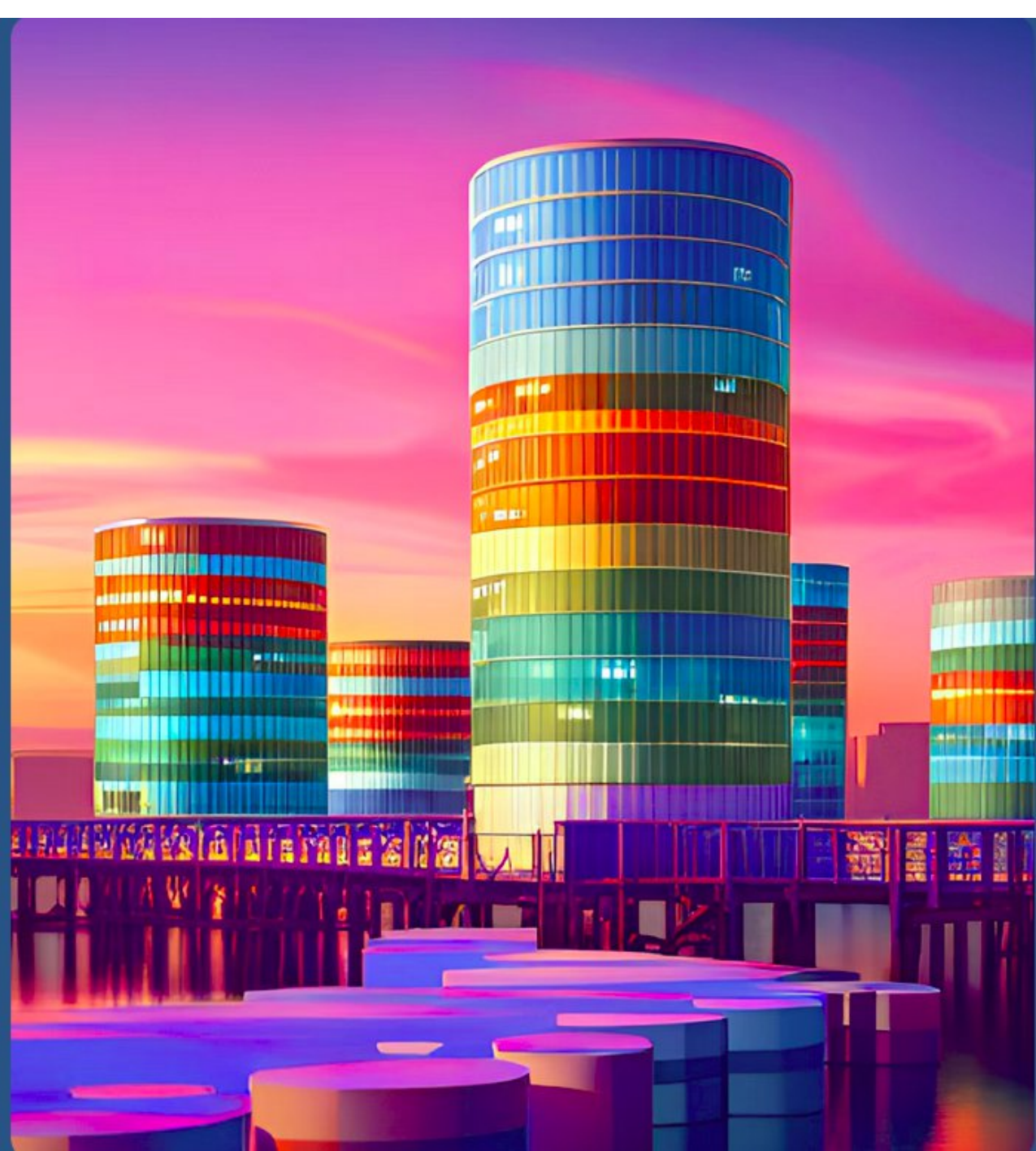
3. Simple Data Cleaning

4. Exploring Data Sets

1. Identiying Candidates for Key Attribute sets Using Bottom Up Algorithm

2. finding Candidates for Attribute Matching

5. Creating Data Base and Populating it with the Processed Records

After Integration

Brasilien	1.75	80.7	26.3	1.62	70.3	26.8
Bulgarien	1.73	81.8	27.2	1.64	69.6	25.9
Burundi	1.67	60.5	21.8	1.55	51.5	21.5
Chile	1.73	82.8	27.8	1.59	71.5	28.2
China	1.75	73.5	24.1	1.63	62.2	23.5
Cookinseln	1.78	103.7	32.7	1.67	92.8	33.3
Costa Rica	1.74	80.9	26.8	1.6	71.7	28
Dänemark	1.82	86.8	26.3	1.69	70.2	24.6
Deutschland	1.8	88.7	27.4	1.66	71.7	26

All Records are now stored in the Integrated Database