Data Uncleaning Recipe


```
[
  {
    "op": "core/text-transform",
    "engineConfig": {
      "facets": [
        {
          "type": "list",
          "name": "Index",
          "expression": "jython:import random\na = [random.randint(1,
20000) for i in range(0,750)]\n\nfor value in range(1, 20001):\n if value
in a:\n   return \"True\"\n elif value not in a:\n  return \"False\"",
          "columnName": "Index",
          "invert": false,
          "omitBlank": false,
          "omitError": false,
          "selection": [
            {
              "v": {
                "v": "True",
                "l": "True"
              }
            }
          ],
          "selectBlank": false,
          "selectError": false
        }
      ],
      "mode": "row-based"
    },
    "columnName": "quality_dimension",
    "expression": "jython:return \"Validity\"",
    "onError": "keep-original",
    "repeat": false,
    "repeatCount": 10,
    "description": "Text transform on cells in column quality_dimension
using expression jython:return \"Validity\""
  },
  {
    "op": "core/text-transform",
    "engineConfig": {
      "facets": [
        {
          "type": "list",
          "name": "quality_dimension",
          "expression": "value",
          "columnName": "quality_dimension",
          "invert": false,
          "omitBlank": false,
          "omitError": false,
          "selection": [
            {
              "v": {
                "v": "Validity",
                "l": "Validity"
              }
            }
          ],
          "selectBlank": false,
```

```
          "selectError": false
        }
      ],
      "mode": "row-based"
    },
    "columnName": "release_date",
    "expression": "grel:toDate(value),\"dd/MM/yyyy\"",
    "onError": "keep-original",
    "repeat": false,
    "repeatCount": 10,
    "description": "Text transform on cells in column release_date using
expression grel:toDate(value),\"dd/MM/yyyy\""
  },
  {
    "op": "core/text-transform",
    "engineConfig": {
      "facets": [
        {
          "type": "list",
          "name": "Index",
          "expression": "jython:import random\na = [random.randint(1,
20000) for i in range(0,50)]\n\nfor value in range(1, 20001):\n if value
in a:\n   return \"True\"\n elif value not in a:\n  return \"False\"",
          "columnName": "Index",
          "invert": false,
          "omitBlank": false,
          "omitError": false,
          "selection": [
            {
              "v": {
                "v": "True",
                "l": "True"
              }
            }
          ],
          "selectBlank": false,
          "selectError": false
        }
      ],
      "mode": "row-based"
    },
    "columnName": "quality_dimension",
    "expression": "jython:return \"Validity1\"",
    "onError": "keep-original",
    "repeat": false,
    "repeatCount": 10,
    "description": "Text transform on cells in column quality_dimension
using expression jython:return \"Validity1\""
  },
  {
    "op": "core/text-transform",
    "engineConfig": {
      "facets": [
        {
          "type": "list",
          "name": "quality_dimension",
          "expression": "value",
          "columnName": "quality_dimension",
          "invert": false,
          "omitBlank": false,
```

```
          "omitError": false,
          "selection": [
            {
              "v": {
                "v": "Validity1",
                "l": "Validity1"
              }
            }
          ],
          "selectBlank": false,
          "selectError": false
        }
      ],
      "mode": "row-based"
    },
    "columnName": "release_date",
    "expression": "grel:toDate(value),\"dd/MM/yyyy\"",
    "onError": "keep-original",
    "repeat": false,
    "repeatCount": 10,
    "description": "Text transform on cells in column release_date using
expression grel:toDate(value),\"dd/MM/yyyy\""
  },
  {
    "op": "core/text-transform",
    "engineConfig": {
      "facets": [
        {
          "type": "list",
          "name": "quality_dimension",
          "expression": "value",
          "columnName": "quality_dimension",
          "invert": false,
          "omitBlank": false,
          "omitError": false,
          "selection": [
            {
              "v": {
                "v": "Validity1",
                "l": "Validity1"
              }
            }
          ],
          "selectBlank": false,
          "selectError": false
        }
      ],
      "mode": "row-based"
    },
    "columnName": "quality_dimension",
    "expression": "grel:\"Validity\"",
    "onError": "keep-original",
    "repeat": false,
    "repeatCount": 10,
    "description": "Text transform on cells in column quality_dimension
using expression grel:\"Validity\""
  },
  {
    "op": "core/text-transform",
    "engineConfig": {
```

```
      "facets": [
        {
          "type": "list",
          "name": "quality_dimension",
          "expression": "value",
          "columnName": "quality_dimension",
          "invert": false,
          "omitBlank": false,
          "omitError": false,
          "selection": [],
          "selectBlank": true,
          "selectError": false
        },
        {
          "type": "list",
          "name": "Index",
          "expression": "jython:import random\na = [random.randint(1,
20000) for i in range(0,850)]\n\nfor value in range(1, 20001):\n if value
in a:\n   return \"True\"\n elif value not in a:\n  return \"False\"",
          "columnName": "Index",
          "invert": false,
          "omitBlank": false,
          "omitError": false,
          "selection": [
            {
              "v": {
                "v": "True",
                "l": "True"
              }
            }
          ],
          "selectBlank": false,
          "selectError": false
        }
      ],
      "mode": "row-based"
    },
    "columnName": "quality_dimension",
    "expression": "jython:return \"Consistency\"",
    "onError": "keep-original",
    "repeat": false,
    "repeatCount": 10,
    "description": "Text transform on cells in column quality_dimension
using expression jython:return \"Consistency\""
  },
  {
    "op": "core/text-transform",
    "engineConfig": {
      "facets": [
        {
          "type": "list",
          "name": "quality_dimension",
          "expression": "value",
          "columnName": "quality_dimension",
          "invert": false,
          "omitBlank": false,
          "omitError": false,
          "selection": [
            {
              "v": {
```

```
                "v": "Consistency",
                "l": "Consistency"
              }
            }
          ],
          "selectBlank": false,
          "selectError": false
        }
      ],
      "mode": "row-based"
    },
    "columnName": "name",
    "expression": "value.toUppercase()",
    "onError": "keep-original",
    "repeat": false,
    "repeatCount": 10,
    "description": "Text transform on cells in column name using
expression value.toUppercase()"
  },
  {
    "op": "core/text-transform",
    "engineConfig": {
      "facets": [
        {
          "type": "list",
          "name": "quality_dimension",
          "expression": "value",
          "columnName": "quality_dimension",
          "invert": false,
          "omitBlank": false,
          "omitError": false,
          "selection": [],
          "selectBlank": true,
          "selectError": false
        },
        {
          "type": "list",
          "name": "quality_dimension",
          "expression": "jython:import random\na = [random.randint(1,
20000) for i in range(0,850)]\n\nfor value in range(1, 20001):\n if value
in a:\n   return \"True\"\n elif value not in a:\n  return \"False\"",
          "columnName": "quality_dimension",
          "invert": false,
          "omitBlank": false,
          "omitError": false,
          "selection": [
            {
              "v": {
                "v": "True",
                "l": "True"
              }
            }
          ],
          "selectBlank": false,
          "selectError": false
        }
      ],
      "mode": "row-based"
    },
    "columnName": "quality_dimension",
```

```
    "expression": "jython:return \"Accuracy\"",
    "onError": "keep-original",
    "repeat": false,
    "repeatCount": 10,
    "description": "Text transform on cells in column quality_dimension
using expression jython:return \"Accuracy\""
  },
  {
    "op": "core/text-transform",
    "engineConfig": {
      "facets": [
        {
          "type": "list",
          "name": "quality_dimension",
          "expression": "value",
          "columnName": "quality_dimension",
          "invert": false,
          "omitBlank": false,
          "omitError": false,
          "selection": [
            {
              "v": {
                "v": "Accuracy",
                "l": "Accuracy"
              }
            }
          ],
          "selectBlank": false,
          "selectError": false
        }
      ],
      "mode": "row-based"
    },
    "columnName": "album",
    "expression": "jython:return \"_\"+value+\"v\"",
    "onError": "keep-original",
    "repeat": false,
    "repeatCount": 10,
    "description": "Text transform on cells in column album using
expression jython:return \"_\"+value+\"v\""
  },
  {
    "op": "core/text-transform",
    "engineConfig": {
      "facets": [
        {
          "type": "list",
          "name": "quality_dimension",
          "expression": "value",
          "columnName": "quality_dimension",
          "invert": false,
          "omitBlank": false,
          "omitError": false,
          "selection": [],
          "selectBlank": true,
          "selectError": false
        },
        {
          "type": "list",
          "name": "quality_dimension",
```

```
          "expression": "jython:import random\na = [random.randint(1,
17723) for i in range(0,850)]\n\nfor value in range(1, 17724):\n if value
in a:\n   return \"True\"\n elif value not in a:\n  return \"False\"",
          "columnName": "quality_dimension",
          "invert": false,
          "omitBlank": false,
          "omitError": false,
          "selection": [
            {
              "v": {
                "v": "True",
                "l": "True"
              }
            }
          ],
          "selectBlank": false,
          "selectError": false
        }
      ],
      "mode": "row-based"
    },
    "columnName": "quality_dimension",
    "expression": "jython:return \"Completeness\"",
    "onError": "keep-original",
    "repeat": false,
    "repeatCount": 10,
    "description": "Text transform on cells in column quality_dimension
using expression jython:return \"Completeness\""
  },
  {
    "op": "core/text-transform",
    "engineConfig": {
      "facets": [
        {
          "type": "list",
          "name": "quality_dimension",
          "expression": "value",
          "columnName": "quality_dimension",
          "invert": false,
          "omitBlank": false,
          "omitError": false,
          "selection": [
            {
              "v": {
                "v": "Completeness",
                "l": "Completeness"
              }
            }
          ],
          "selectBlank": false,
          "selectError": false
        }
      ],
      "mode": "row-based"
    },
    "columnName": "year",
    "expression": "jython:return \"\"",
    "onError": "keep-original",
    "repeat": false,
    "repeatCount": 10,
```

```
    "description": "Text transform on cells in column year using
expression jython:return \"\""
  },
  {
    "op": "core/text-transform",
    "engineConfig": {
      "facets": [
        {
          "type": "list",
          "name": "quality_dimension",
          "expression": "value",
          "columnName": "quality_dimension",
          "invert": false,
          "omitBlank": false,
          "omitError": false,
          "selection": [],
          "selectBlank": true,
          "selectError": false
        },
        {
          "type": "list",
          "name": "quality_dimension",
          "expression": "jython:import random\na = [random.randint(1,
16872) for i in range(0,850)]\n\nfor value in range(1, 16873):\n if value
in a:\n   return \"True\"\n elif value not in a:\n  return \"False\"",
          "columnName": "quality_dimension",
          "invert": false,
          "omitBlank": false,
          "omitError": false,
          "selection": [
            {
              "v": {
                "v": "True",
                "l": "True"
              }
            }
          ],
          "selectBlank": false,
          "selectError": false
        }
      ],
      "mode": "row-based"
    },
    "columnName": "quality_dimension",
    "expression": "jython:return \"Confirmity\"",
    "onError": "keep-original",
    "repeat": false,
    "repeatCount": 10,
    "description": "Text transform on cells in column quality_dimension
using expression jython:return \"Confirmity\""
  },
  {
    "op": "core/text-transform",
    "engineConfig": {
      "facets": [
        {
          "type": "list",
          "name": "quality_dimension",
          "expression": "value",
          "columnName": "quality_dimension",
```

```
          "invert": false,
          "omitBlank": false,
          "omitError": false,
          "selection": [
            {
              "v": {
                "v": "Confirmity",
                "l": "Confirmity"
              }
            }
          ],
          "selectBlank": false,
          "selectError": false
        }
      ],
      "mode": "row-based"
    },
    "columnName": "quality_dimension",
    "expression": "jython:return \"Conformity\"",
    "onError": "keep-original",
    "repeat": false,
    "repeatCount": 10,
    "description": "Text transform on cells in column quality_dimension
using expression jython:return \"Conformity\""
  },
  {
    "op": "core/column-split",
    "engineConfig": {
      "facets": [],
      "mode": "row-based"
    },
    "columnName": "artists",
    "guessCellType": true,
    "removeOriginalColumn": true,
    "mode": "separator",
    "separator": "['",
    "regex": false,
    "maxColumns": 0,
    "description": "Split column artists by separator"
  },
  {
    "op": "core/column-removal",
    "columnName": "artists 1",
    "description": "Remove column artists 1"
  },
  {
    "op": "core/column-split",
    "engineConfig": {
      "facets": [],
      "mode": "row-based"
    },
    "columnName": "artists 2",
    "guessCellType": true,
    "removeOriginalColumn": true,
    "mode": "separator",
    "separator": "']",
    "regex": false,
    "maxColumns": 0,
    "description": "Split column artists 2 by separator"
  },
```

```
{
  "op": "core/column-removal",
  "columnName": "artists 2 2",
  "description": "Remove column artists 2 2"
},
{
  "op": "core/column-rename",
  "oldColumnName": "artists 2 1",
  "newColumnName": "artists",
  "description": "Rename column artists 2 1 to artists"
},
{
  "op": "core/text-transform",
  "engineConfig": {
    "facets": [
      {
        "type": "list",
        "name": "quality_dimension",
        "expression": "value",
        "columnName": "quality_dimension",
        "invert": false,
        "omitBlank": false,
        "omitError": false,
        "selection": [
          {
            "v": {
              "v": "Conformity",
              "l": "Conformity"
            }
          }
        ],
        "selectBlank": false,
        "selectError": false
      },
      {
        "type": "list",
        "name": "explicit",
        "expression": "value",
        "columnName": "explicit",
        "invert": false,
        "omitBlank": false,
        "omitError": false,
        "selection": [
          {
            "v": {
              "v": "FALSE",
              "l": "FALSE"
            }
          }
        ],
        "selectBlank": false,
        "selectError": false
      }
    ],
    "mode": "row-based"
  },
  "columnName": "explicit",
  "expression": "jython:return \"0\"",
  "onError": "keep-original",
  "repeat": false,
```

```
    "repeatCount": 10,
    "description": "Text transform on cells in column explicit using
expression jython:return \"0\""
  },
  {
    "op": "core/column-addition",
    "engineConfig": {
      "facets": [],
      "mode": "row-based"
    },
    "baseColumnName": "release_date",
    "expression": "jython:return \"16-03-2021\"",
    "onError": "set-to-blank",
    "newColumnName": "last_updated",
    "columnInsertIndex": 17,
    "description": "Create column last_updated at index 17 based on
column release_date using expression jython:return \"16-03-2021\""
  },
  {
    "op": "core/text-transform",
    "engineConfig": {
      "facets": [
        {
          "type": "list",
          "name": "quality_dimension",
          "expression": "value",
          "columnName": "quality_dimension",
          "invert": false,
          "omitBlank": false,
          "omitError": false,
          "selection": [
            {
              "v": {
                "v": "Conformity",
                "l": "Conformity"
              }
            }
          ],
          "selectBlank": false,
          "selectError": false
        },
        {
          "type": "list",
          "name": "explicit",
          "expression": "value",
          "columnName": "explicit",
          "invert": false,
          "omitBlank": false,
          "omitError": false,
          "selection": [
            {
              "v": {
                "v": "TRUE",
                "l": "TRUE"
              }
            }
          ],
          "selectBlank": false,
          "selectError": false
        }
```

```
      ],
      "mode": "row-based"
    },
    "columnName": "quality_dimension",
    "expression": "jython:return \"\"",
    "onError": "keep-original",
    "repeat": false,
    "repeatCount": 10,
    "description": "Text transform on cells in column quality_dimension
using expression jython:return \"\""
  },
  {
    "op": "core/text-transform",
    "engineConfig": {
      "facets": [
        {
          "type": "list",
          "name": "quality_dimension",
          "expression": "value",
          "columnName": "quality_dimension",
          "invert": false,
          "omitBlank": false,
          "omitError": false,
          "selection": [],
          "selectBlank": true,
          "selectError": false
        },
        {
          "type": "list",
          "name": "quality_dimension",
          "expression": "jython:import random\na = [random.randint(1,
16090) for i in range(0,850)]\n\nfor value in range(1, 16091):\n if value
in a:\n   return \"True\"\n elif value not in a:\n  return \"False\"",
          "columnName": "quality_dimension",
          "invert": false,
          "omitBlank": false,
          "omitError": false,
          "selection": [
            {
              "v": {
                "v": "True",
                "l": "True"
              }
            }
          ],
          "selectBlank": false,
          "selectError": false
        }
      ],
      "mode": "row-based"
    },
    "columnName": "quality_dimension",
    "expression": "jython:return \"Timliness\"",
    "onError": "keep-original",
    "repeat": false,
    "repeatCount": 10,
    "description": "Text transform on cells in column quality_dimension
using expression jython:return \"Timliness\""
  },
  {
```

```
    "op": "core/text-transform",
    "engineConfig": {
      "facets": [
        {
          "type": "list",
          "name": "quality_dimension",
          "expression": "value",
          "columnName": "quality_dimension",
          "invert": false,
          "omitBlank": false,
          "omitError": false,
          "selection": [
            {
              "v": {
                "v": "Timliness",
                "l": "Timliness"
              }
            }
          ],
          "selectBlank": false,
          "selectError": false
        }
      ],
      "mode": "row-based"
    },
    "columnName": "last_updated",
    "expression": "jython:return \"14-03-2021\"",
    "onError": "keep-original",
    "repeat": false,
    "repeatCount": 10,
    "description": "Text transform on cells in column last_updated using
expression jython:return \"14-03-2021\""
  },
  {
    "op": "core/text-transform",
    "engineConfig": {
      "facets": [
        {
          "type": "list",
          "name": "quality_dimension",
          "expression": "value",
          "columnName": "quality_dimension",
          "invert": false,
          "omitBlank": false,
          "omitError": false,
          "selection": [],
          "selectBlank": true,
          "selectError": false
        },
        {
          "type": "list",
          "name": "quality_dimension",
          "expression": "jython:import random\na = [random.randint(1,
15202) for i in range(0,850)]\n\nfor value in range(1, 15203):\n if value
in a:\n   return \"True\"\n elif value not in a:\n  return \"False\"",
          "columnName": "quality_dimension",
          "invert": false,
          "omitBlank": false,
          "omitError": false,
          "selection": [
```

```
            {
              "v": {
                "v": "True",
                "l": "True"
              }
            }
          ],
          "selectBlank": false,
          "selectError": false
        }
      ],
      "mode": "row-based"
    },
    "columnName": "quality_dimension",
    "expression": "jython:return \"Uniqueness\"",
    "onError": "keep-original",
    "repeat": false,
    "repeatCount": 10,
    "description": "Text transform on cells in column quality_dimension
using expression jython:return \"Uniqueness\""
  }
]
```