Rudra Gandhi

Data 200

April 29, 2021

**Homework 5**

Graphical user interface, text, application

Description automatically generated

Graphical user interface, application, table, Excel

Description automatically generated

Graphical user interface, application, table

Description automatically generated

Graphical user interface, text, application

Description automatically generated

Graphical user interface, text, application

Description automatically generated

Graphical user interface, application, Teams

Description automatically generated

**Link to Recipe:** <https://us-east-2.console.aws.amazon.com/databrew/home?region=us-east-2#recipes>

**Link to Github:** <https://github.com/rudragandhi98/AWS-Data-Cleaning/blob/main/DataCleaning-recipe.json>

**Recipe in JSON format:**

[

{

"Action": {

"Operation": "DELETE",

"Parameters": {

"sourceColumns": "[\"Corporate Author\",\"Corporate Contributors\",\"Former owner\",\"Engraver\"]"

}

}

},

{

"Action": {

"Operation": "REMOVE\_COMBINED",

"Parameters": {

"collapseConsecutiveWhitespace": "false",

"removeAllPunctuation": "false",

"removeAllQuotes": "false",

"removeAllWhitespace": "false",

"removeCustomCharacters": "false",

"removeCustomValue": "false",

"removeLeadingAndTrailingPunctuation": "false",

"removeLeadingAndTrailingQuotes": "false",

"removeLeadingAndTrailingWhitespace": "false",

"removeLetters": "false",

"removeNumbers": "true",

"removeSpecialCharacters": "true",

"sourceColumn": "Place of Publication"

}

}

},

{

"Action": {

"Operation": "YEAR",

"Parameters": {

"functionStepType": "YEAR",

"sourceColumn": "Date of Publication",

"targetColumn": "Date of Publication\_extracted",

"units": "year"

}

}

},

{

"Action": {

"Operation": "DELETE",

"Parameters": {

"sourceColumns": "[\"Date of Publication\"]"

}

}

},

{

"Action": {

"Operation": "RENAME",

"Parameters": {

"sourceColumn": "Date of Publication\_extracted",

"targetColumn": "Date of Publication"

}

}

},

{

"Action": {

"Operation": "EXTRACT\_PATTERN",

"Parameters": {

"pattern": "London",

"sourceColumn": "Place of Publication",

"targetColumn": "Place of Publication\_extracted"

}

}

},

{

"Action": {

"Operation": "REMOVE\_VALUES",

"Parameters": {

"sourceColumn": "Place of Publication\_extracted"

}

},

"ConditionExpressions": [

{

"Condition": "IS\_MISSING",

"Value": "string",

"TargetColumn": "Place of Publication\_extracted"

}

]

},

{

"Action": {

"Operation": "DELETE",

"Parameters": {

"sourceColumns": "[\"Place of Publication\"]"

}

}

},

{

"Action": {

"Operation": "RENAME",

"Parameters": {

"sourceColumn": "Place of Publication\_extracted",

"targetColumn": "Place of Publication"

}

}

},

{

"Action": {

"Operation": "REMOVE\_COMBINED",

"Parameters": {

"collapseConsecutiveWhitespace": "false",

"customCharacters": "[]",

"removeAllPunctuation": "false",

"removeAllQuotes": "false",

"removeAllWhitespace": "false",

"removeCustomCharacters": "true",

"removeCustomValue": "false",

"removeLeadingAndTrailingPunctuation": "false",

"removeLeadingAndTrailingQuotes": "false",

"removeLeadingAndTrailingWhitespace": "false",

"removeLetters": "false",

"removeNumbers": "false",

"removeSpecialCharacters": "false",

"sourceColumn": "Edition Statement"

}

}

},

{

"Action": {

"Operation": "FILL\_WITH\_EMPTY",

"Parameters": {

"sourceColumn": "Edition Statement"

}

}

}

]