**Yes Workflow:**

# @begin DataCleaning\_Project @desc Data Cleaning Project WorkFlow

# @in f1 @as input\_file @file csv

# @out f6 @file yes\_workflow @as workflow\_created

# @begin Load\_Dataset\_OpenRefine @desc Load datset csv input file in Open Refine

# @in f1 @as input\_file

# @out f2 @file csv @as file\_loaded\_OpenRefine

# @end Load\_File

# @begin Clean\_Dataset\_OpenRefine @desc Clean datset in Open Refine

# @out f3 @as cleaned\_file\_OpenRefine @file csv

# @in f2 @file csv @as file\_loaded\_OpenRefine

# @end Clean\_Dataset\_OpenRefine

# @begin Load\_Dataset\_SQL @desc Load cleaned datset in SQL

# @in f3 @as cleaned\_file\_OpenRefine @file csv

# @out f4 @file sql\_table @as table\_loaded\_SQL

# @end Load\_Dataset\_SQL

# @begin Check\_Constraints\_SQL @desc Apply constraints in SQL table

# @out f5 @as violations\_diaplyed\_in\_table @file sql\_table

# @in f4 @file sql\_table @as table\_loaded\_SQL

# @end Constraints\_SQL

# @begin Create\_yes\_workflow @desc Make Yes Workflow for the project

# @in f5 @as violations\_diaplyed\_in\_table @file sql\_table

# @out f6 @file yes\_workflow @as workflow\_created

# @end Create\_yes\_workflow

# @end DataCleaning\_Project



# @begin DataCleaning\_Using\_OpenRefine @desc Data Cleaning Project

# @in f1 @as input\_file @file csv

# @out f9 @file csv @as file8

# @begin load\_File @desc Load csv input file in Open Refine

# @in f1 @as input\_file

# @out f2 @file csv @as file1\_loaded

# @end load\_File

# @begin trim\_spaces @desc Trim spaces of the columns

# @in f2 @file csv @as file1\_loaded

# @out f3 @file csv @as file2

# @end trim\_spaces

# @begin numeric\_columns\_facet Facet @desc Convert columns to numeric facets

# @in f3 @file csv @as file2

# @out f4 @file csv @as file3

# @end numeric\_columns\_facet

# @begin replace\_special\_characters Facet @desc Replace special characters by blank value

# @out f5 @file csv @as file4

# @in f4 @file csv @as file3

# @end numeric\_columns\_facet

# @begin cluster\_columns Facet @desc Cluster the data in columns

# @in f5 @file csv @as file4

# @out f6 @file csv @as file5

# @end numeric\_columns\_facet

# @begin convert\_date\_isoformat @desc Change date to ISO format

# @in f6 @file csv @as file5

# @out f7 @file csv @as file6

# @end convert\_date\_isoformat

# @begin convert\_datetime\_to\_date @desc Add a new column to display only date part

# @in f7 @file csv @as file6

# @out f8 @file csv @as file7

# @end datetime\_to\_date

# @begin replace\_blank\_values @desc Replace blank values by unknown

# @in f8 @file csv @as file7

# @out f9 @file csv @as file8

# @end replace\_blank\_values

# @end DataCleaning\_Using\_OpenRefine



# @begin DataCleaningUsingSQL @desc Data Cleaning Project

# @in f1 @as input\_cleaned\_file\_from\_openrefine @file csv

# @out f8 @file stream:stdout @as file8

# @out f11 @as violation\_checked\_file @file sql\_table

# @begin Load\_File @desc Load cleaned csv file in SQL

# @in f1 @as input\_cleaned\_file\_from\_openrefine

# @out f2 @file sql\_table @as file1\_loaded

# @end Load\_File

# @begin check\_constarints\_in\_sql @desc Apply constraints in SQL table

# @in f2 @file sql\_table @as file1\_loaded

# @out f3 @as check\_primary\_key\_constraints

# @out f4 @as date\_check\_constarint

# @out f5 @as status\_check

# @out f6 @as page\_count

# @out f7 @as dish\_count

# @end check\_constarints\_in\_sql

# @begin constraints\_checked\_file @desc Display violations in table

# @in f3 @as check\_primary\_key\_constraints

# @in f4 @as date\_check\_constarint

# @in f5 @as status\_check

# @in f6 @as page\_count

# @in f7 @as dish\_count

# @out f11 @as violation\_checked\_file @file sql\_table

# @end constraints\_checked\_file

# @end DataCleaningUsingSQL

