# Missing Value Imputation Summary

Input Dataset: ->e.csv.

Defination of Missing Value: empty cells, characters, zeros, negative values, values less than 500.

Features Removal: TRUE. 10 compounds with more than 50% of missing values were removed before the missing value imputation. These compounds are marked as excluded = Yes in the Table 2.

Missing Value Summary: Table 2.

| label | empty.cells | zeros | negative.values | characters | values.less.than.500 | total | excluded |
| --- | --- | --- | --- | --- | --- | --- | --- |
| n-acetyl-d-hexosamine | 16 | 0 | 0 | 0 | 0 | 16 | Yes |
| melibiose | 16 | 0 | 0 | 0 | 0 | 16 | Yes |
| glutaric acid | 16 | 0 | 0 | 0 | 0 | 16 | Yes |
| cysteine-glycine | 16 | 0 | 0 | 0 | 0 | 16 | Yes |
| p-hydroxylphenyllactic acid | 13 | 0 | 0 | 0 | 0 | 13 | Yes |
| 1-kestose | 13 | 0 | 0 | 0 | 0 | 13 | Yes |
| galactitol | 10 | 0 | 0 | 0 | 0 | 10 | Yes |
| 1,2,4-benzenetriol | 10 | 0 | 0 | 0 | 0 | 10 | Yes |
| ribonic acid | 8 | 0 | 0 | 0 | 0 | 8 | Yes |
| conduritol-beta-epoxide | 8 | 0 | 0 | 0 | 0 | 8 | Yes |
| sucrose | 7 | 0 | 0 | 0 | 0 | 7 | No |
| maltotriose | 6 | 0 | 0 | 0 | 0 | 6 | No |
| guanosine | 6 | 0 | 0 | 0 | 0 | 6 | No |
| isohexonic acid | 5 | 0 | 0 | 0 | 0 | 5 | No |
| adenosine-5-monophosphate | 4 | 0 | 0 | 0 | 0 | 4 | No |
| UDP GlcNAc | 3 | 0 | 0 | 0 | 0 | 3 | No |
| glucose-6-phosphate | 3 | 0 | 0 | 0 | 0 | 3 | No |
| 3,6-anhydro-D-galactose | 3 | 0 | 0 | 0 | 0 | 3 | No |
| fructose-6-phosphate | 2 | 0 | 0 | 0 | 0 | 2 | No |
| alpha-ketoglutarate | 2 | 0 | 0 | 0 | 0 | 2 | No |
| uric acid | 1 | 0 | 0 | 0 | 0 | 1 | No |
| thymine | 1 | 0 | 0 | 0 | 0 | 1 | No |
| tagatose | 1 | 0 | 0 | 0 | 0 | 1 | No |
| ribose | 1 | 0 | 0 | 0 | 0 | 1 | No |
| oxalic acid | 1 | 0 | 0 | 0 | 0 | 1 | No |
| homoserine | 1 | 0 | 0 | 0 | 0 | 1 | No |
| hexose-6-phosphate | 1 | 0 | 0 | 0 | 0 | 1 | No |
| glyceric acid | 1 | 0 | 0 | 0 | 0 | 1 | No |
| fructose | 1 | 0 | 0 | 0 | 0 | 1 | No |
| arachidic acid | 1 | 0 | 0 | 0 | 0 | 1 | No |
| 2-picolinic acid | 1 | 0 | 0 | 0 | 0 | 1 | No |

Table 2: missing value summary (->Missing Value Imputation->result summary.csv)

Missing Value Imputation Method: replace by half minimum. For each compound, the 1/2 of the smallest non-missing value was used to impute the missing values. A small disturbance (random values from a normal distribution with a standard deviation of 1/10 of the smallest non-missing value) was added to each imputed value to prevent constant value that may cause trouble in statistical modelings.

Output Dataset: ->Missing Value Imputation->result dataset.csv.