**Data Mining Lab**

**Assignment 2**

**Name** – Neelangshu Roy

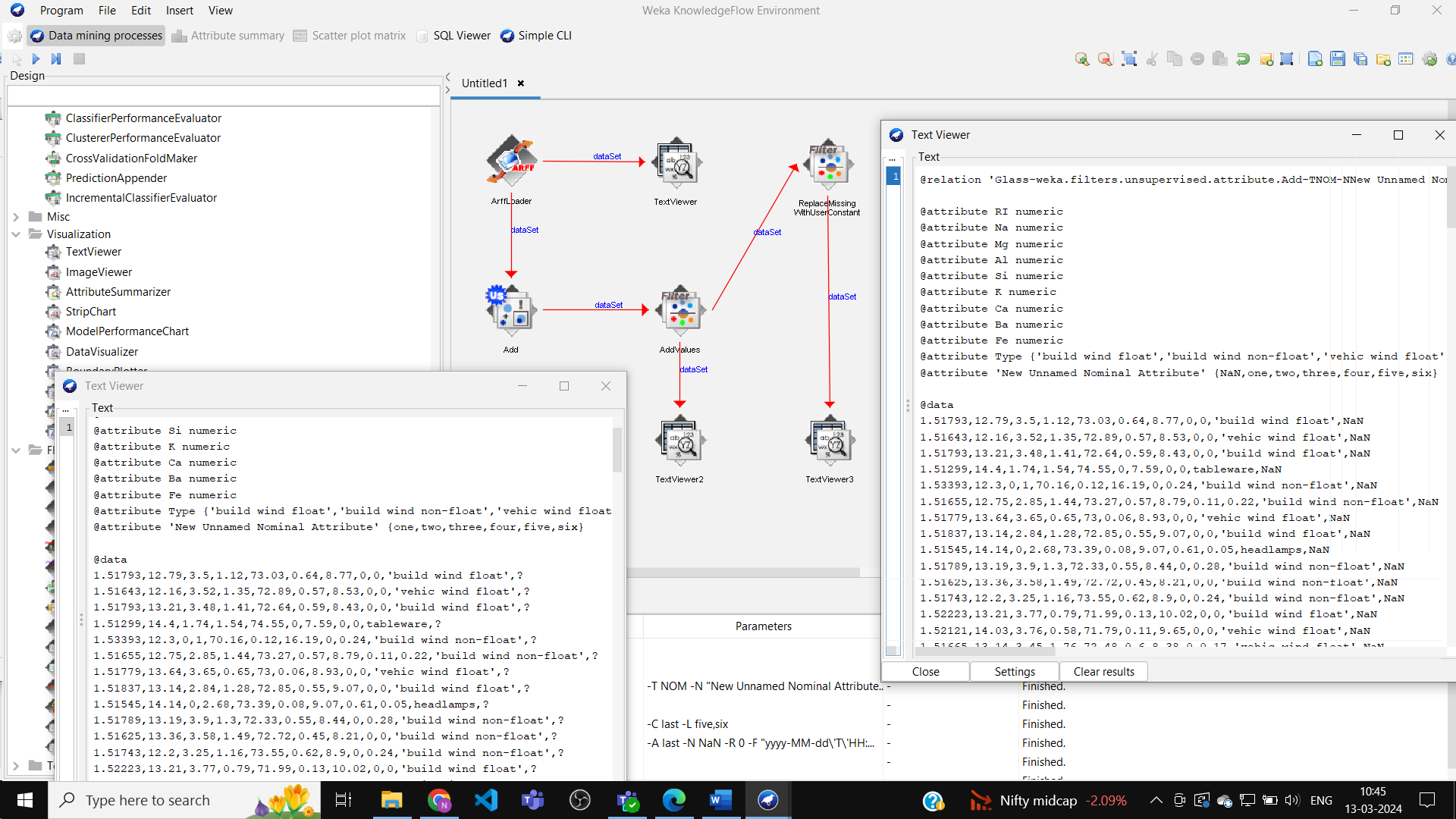
**Reg No** - 20214060

**Group – C**

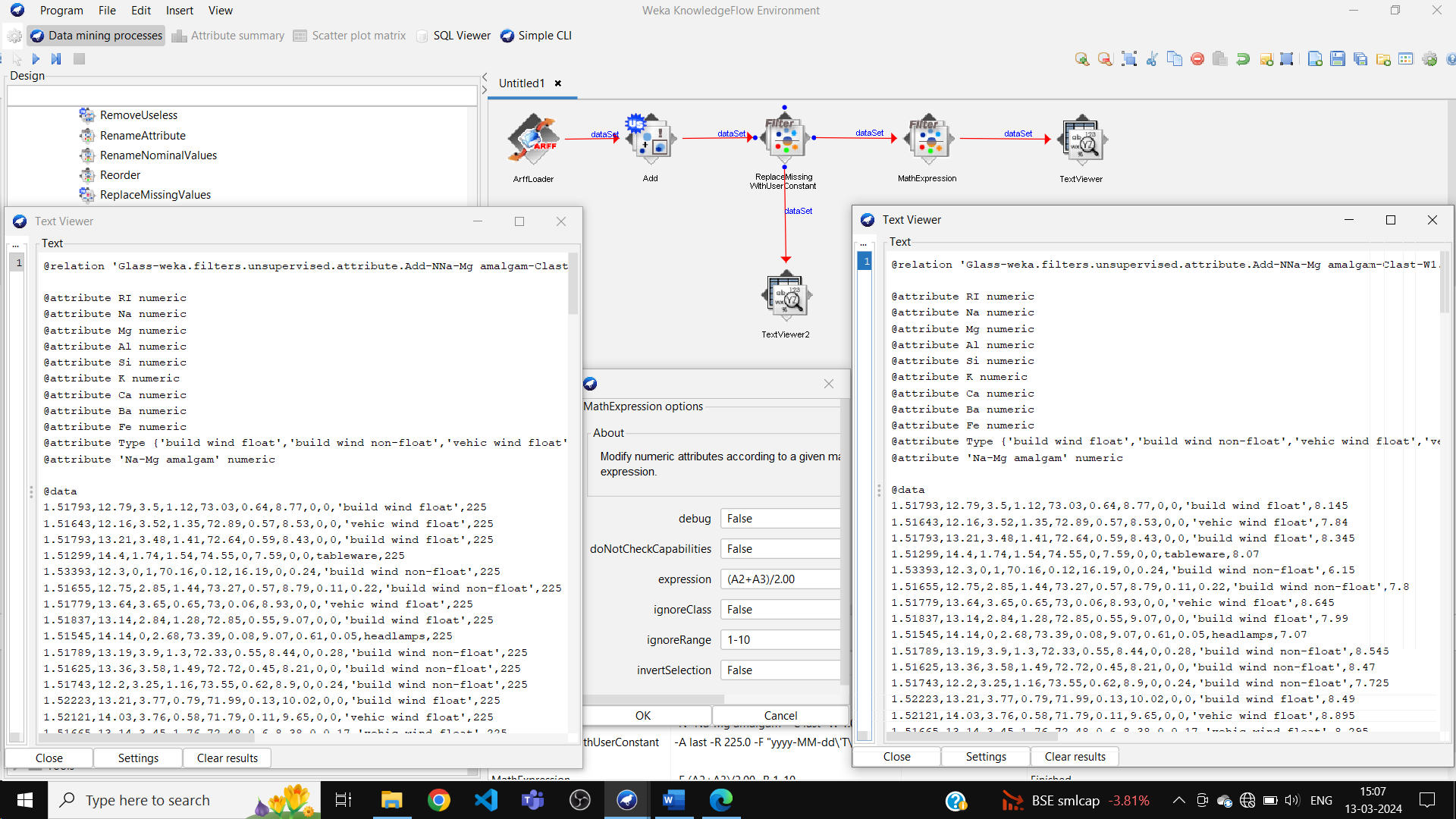
Q1. Load glass dataset :

Use KnowledgeFlow only –

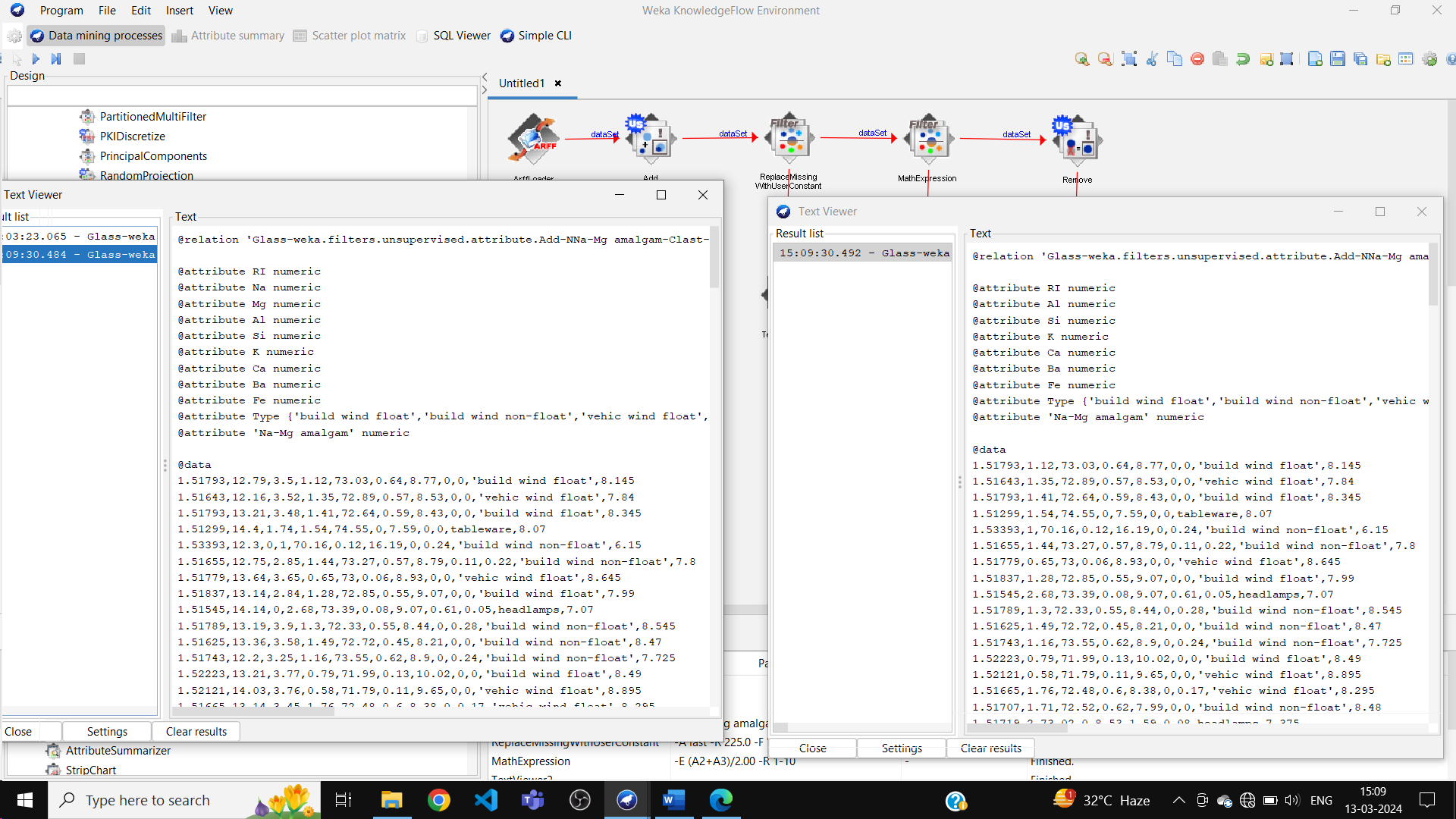
1. Mark missing values with NaN.



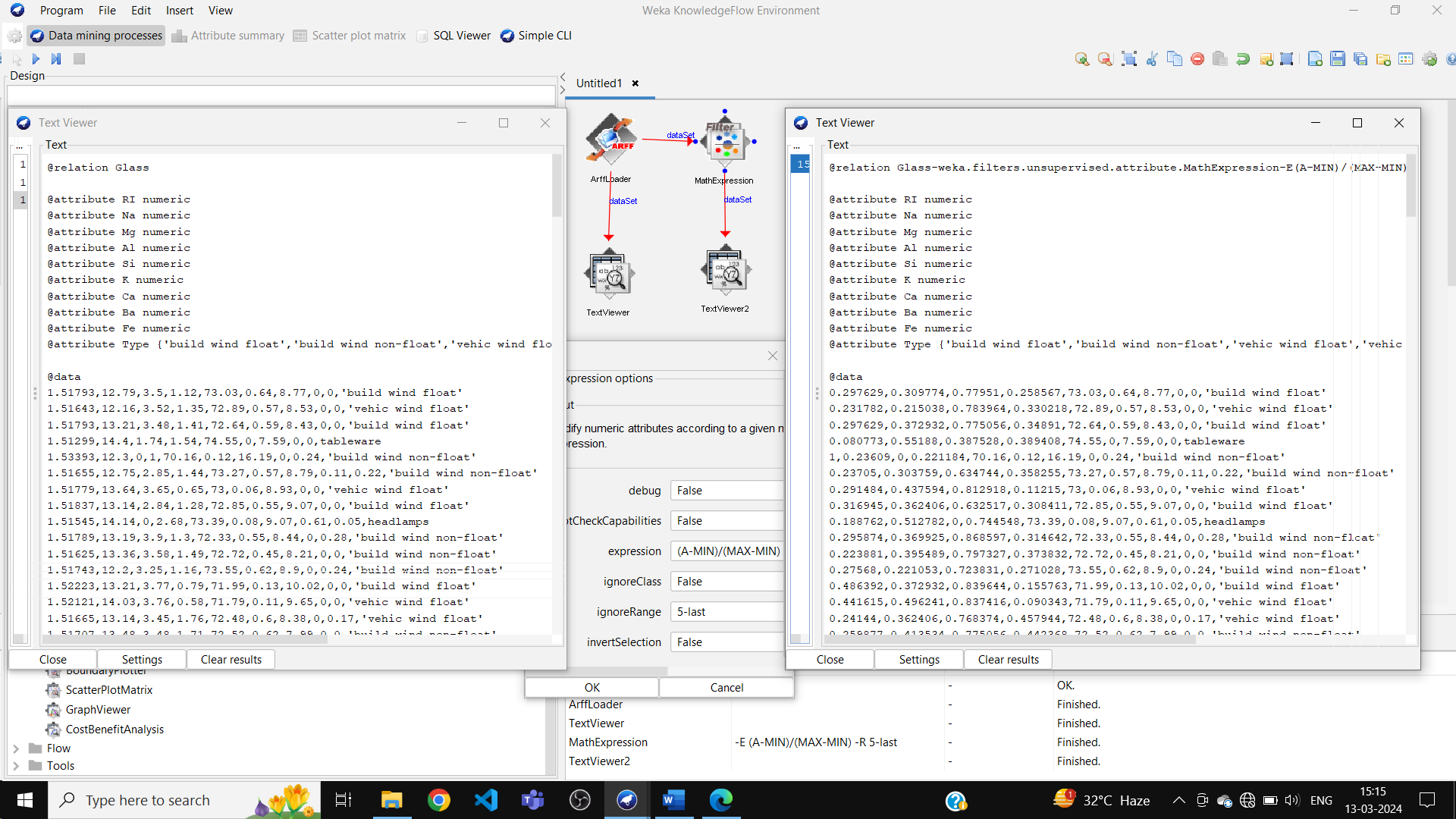
1. Merge any 2 numeric attributes to any new attribute.



After removing Na and Mg attributes –

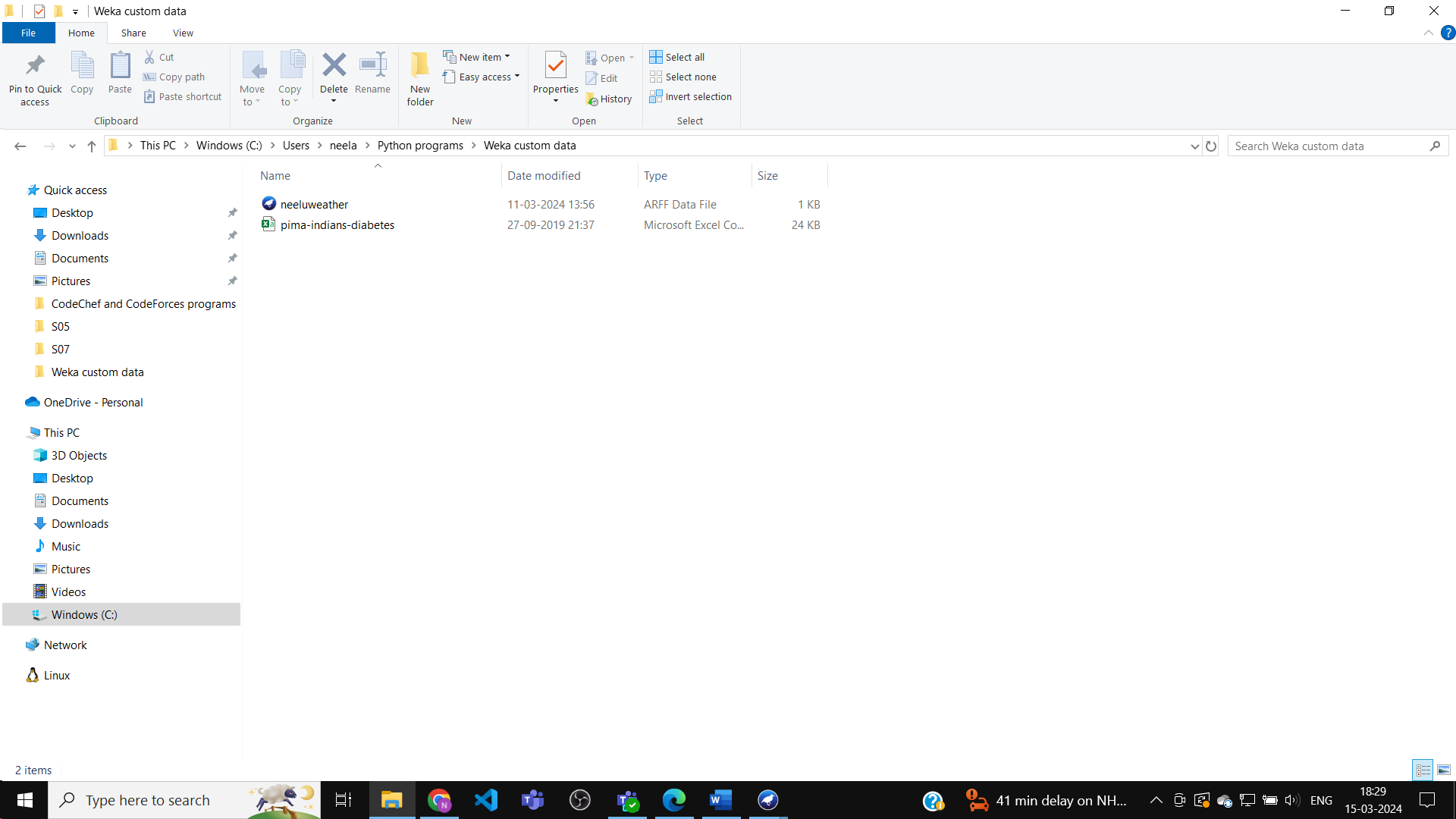


1. Select some numeric attributes and normalize them between (min-max) ranges.



Q2. Do following operations on diabetes data downloaded from Kaggle in .csv format and save all results in separate files using data sink .csv format.

Before –



Operations –

a. Normalize all numeric attributes

b. Remove all misclassified value by J48 class filter

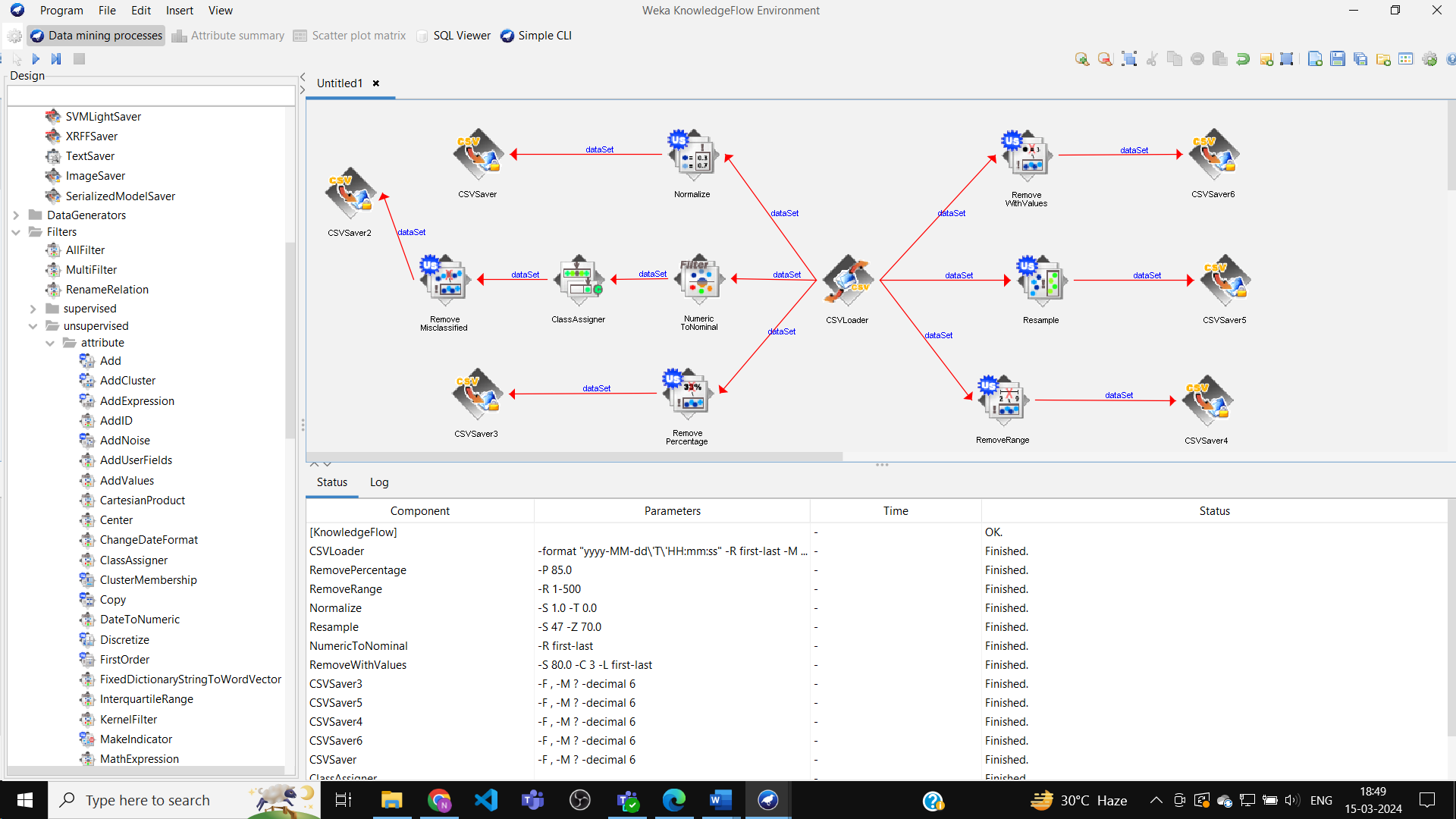
c. Remove Percent

d. Remove Range

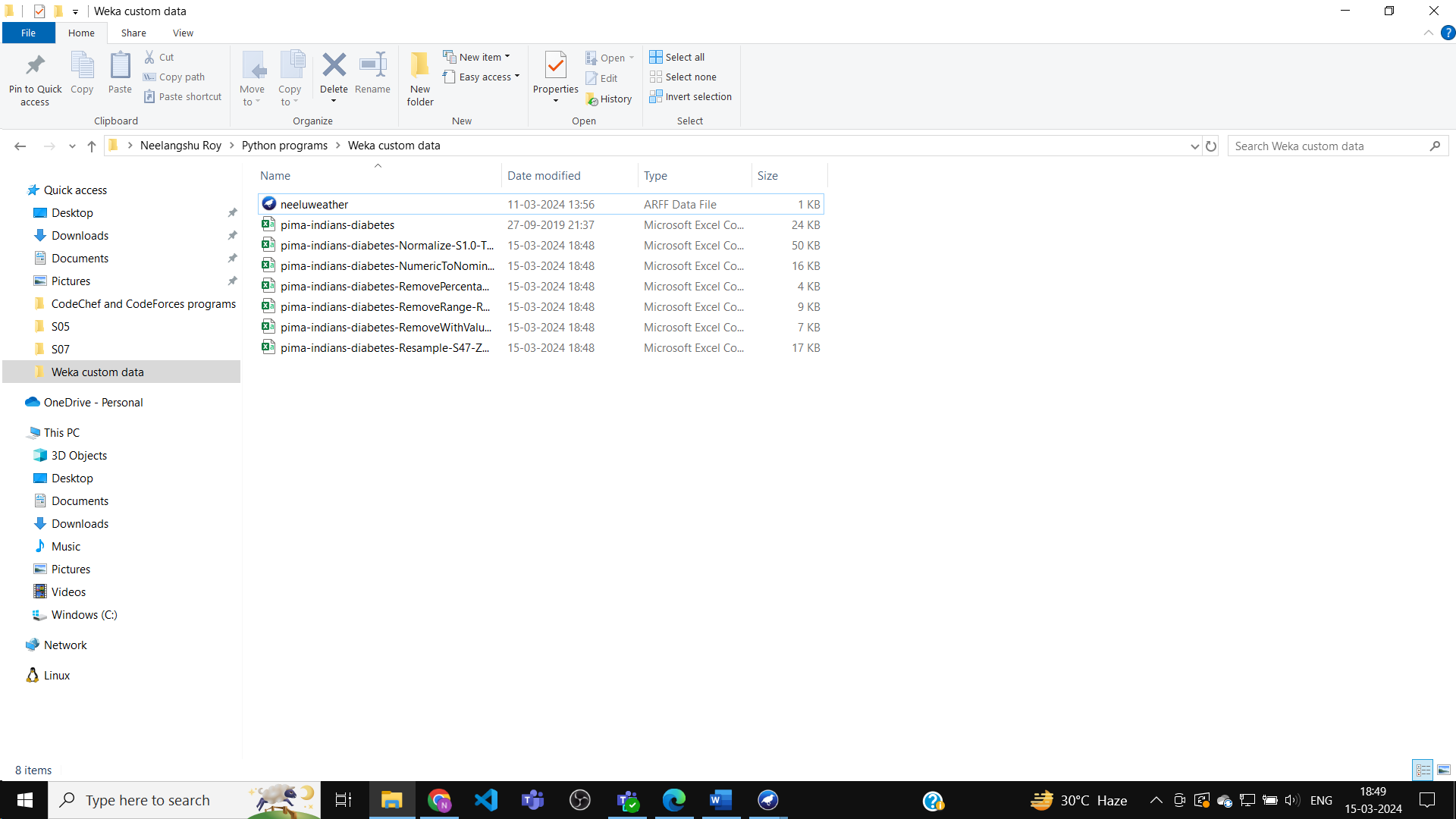
e. Resample

f. Remove With Values

After –



File directory –



Q3. Load weather.arff dataset and do the following operations –

a. Make play attribute as Class label.

b. Split dataset into training (80%) and testing(20%) part and save both splits in

different folds in directory location

c. Prepare 5 fold cross validation testing data and filter all such folds which having

all given class label nominal values.