**Google Open-Refine**

I have taken the same dataset (Bitcoin Historical Data) from Kaggle for cleaning it using the OpenRefine.

I have squeezed down the dataset to 3379 rows as OpenRefine couldn’t load a dataset with more than 1.5 lacs values.

Following are the screenshot of experimentation I did on the dataset using openrefine.

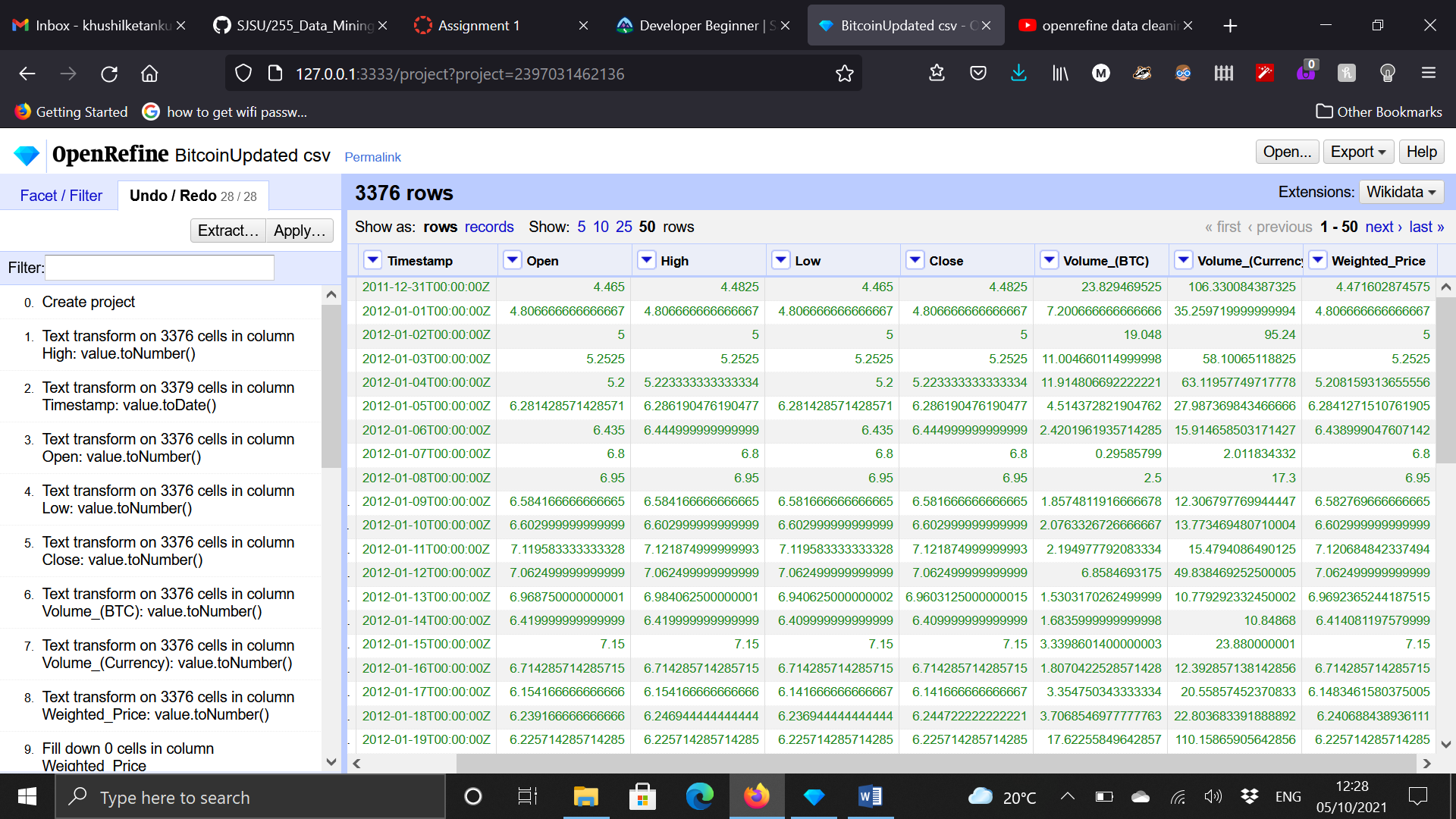


Fig.1

The first step was I created the project and after that I used timeline facet on Timestamp and converted Timestamp to Timeline Facet and similarly I used Number facet on all the other columns and edit row cells to number and we get the following output visible on the left side of the Fig.2 and Fig.3

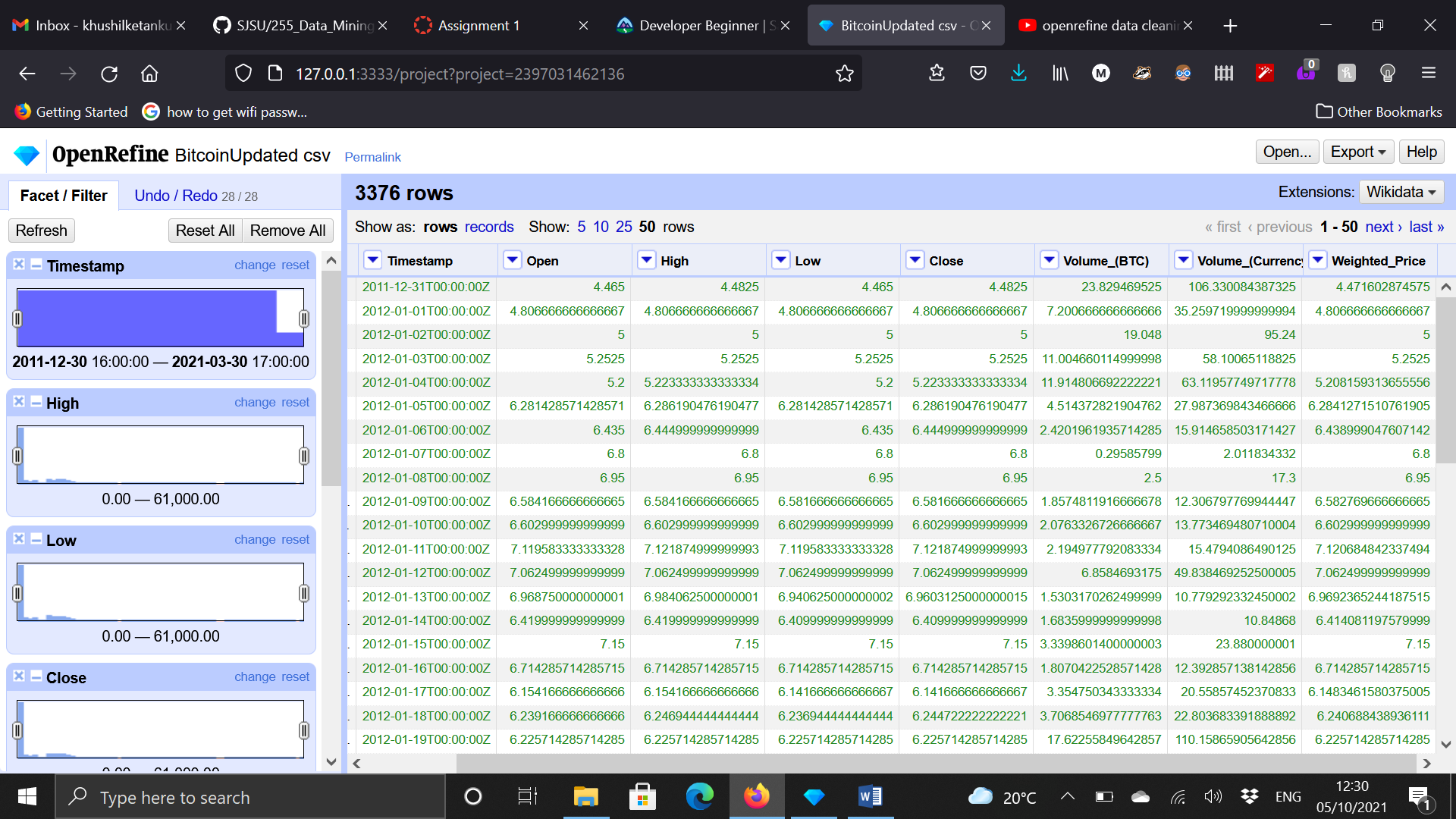


Fig. 2

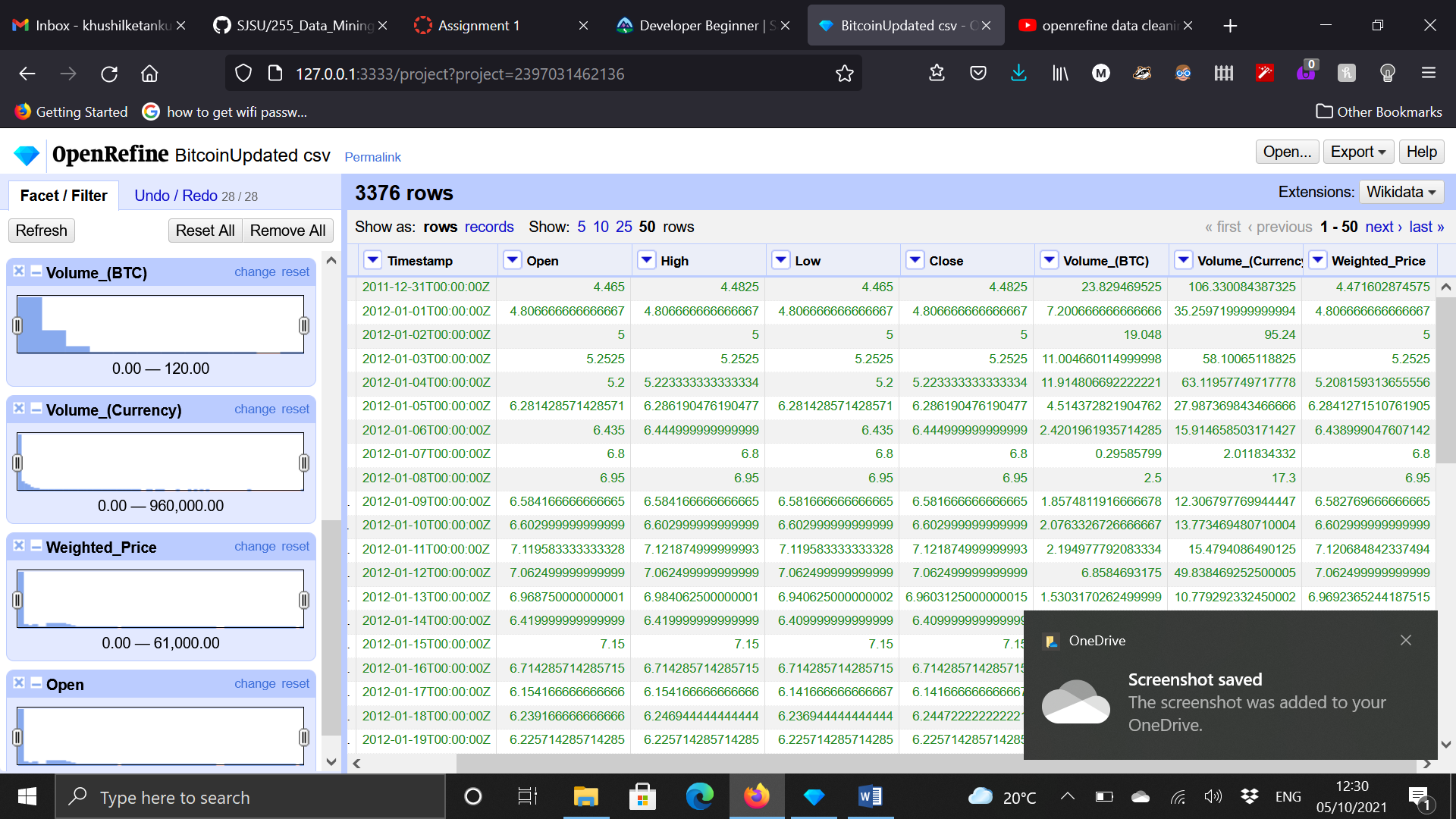


Fig. 3

From the facets you can change the limits of your dataset and can manipulate it accordingly.

From the numeric facets we were able to find out that there were empty rows in our dataset and we were able to remove it by flagging them and then selecting option “remove similar kind of rows” shown in Fig.4

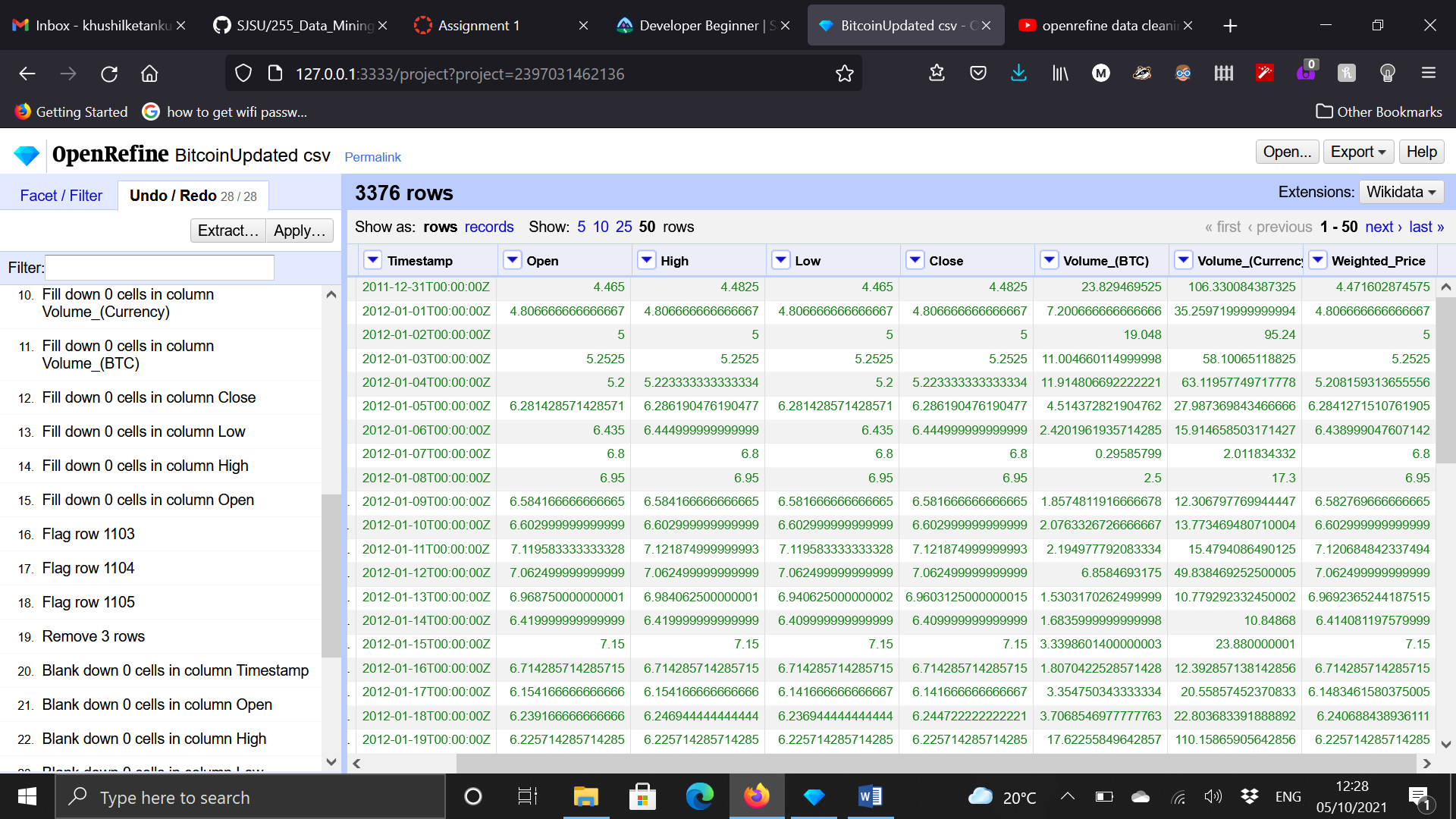


Fig. 4

After removing the empty data we checked for data redundancy and duplicate data inside our dataset using sort function and blank down function as shown in Fig 5.

I tried multiple things with my dataset and different dataset which included text, number, id, etc. to get a hang of Open Refine more efficiently.

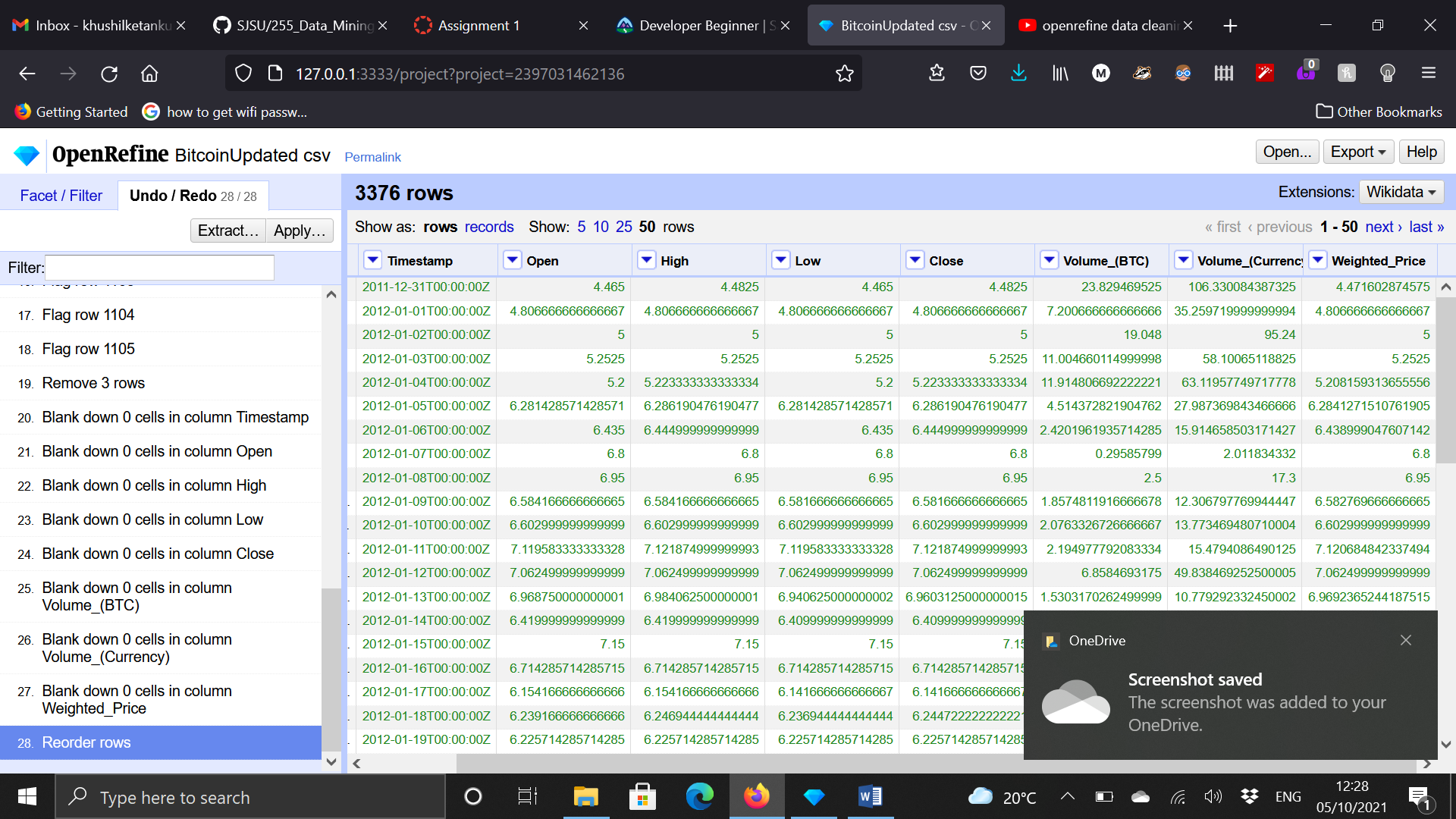


Fig. 5

Thus this is where I felt that the dataset is mostly cleaned as it didn’t contained any redundant, empty, NaN and duplicate data.

**References:**

<https://www.youtube.com/watch?v=jyUlT8ohlG4>

<https://www.youtube.com/watch?v=xZlz4ISgNBc>

<https://www.youtube.com/watch?v=nORS7STbLyk>