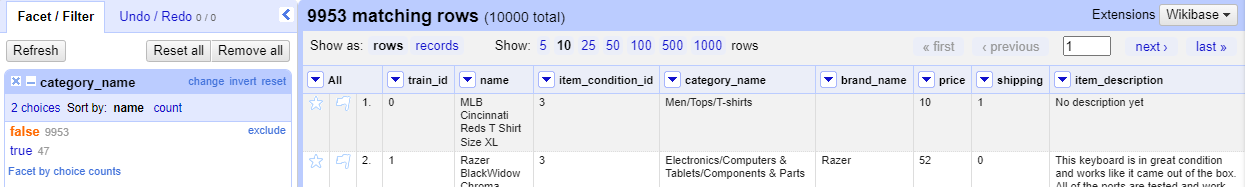
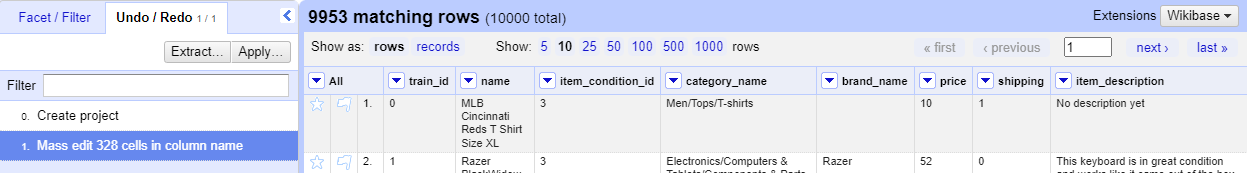
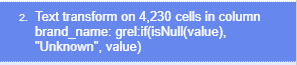
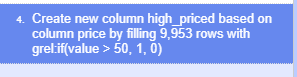
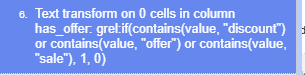
3. Clean/Refine the data. Take a screenshot at the end of each exercise. NOTE: OpenRefine maintains a log of all changes. You can undo changes by the "Undo/Redo" button at the upper left corner. Follow the exact output format specified in every part below

a) Select the category\_name column and choose "Facet by Blank" (Facet → Customized Facets → Facet by Blank) to filter out the records that have blank values in this column. Exclude these rows (hint: choose include of an appropriate Boolean variable in the left-side output panel). How many rows left?  
  


b) Select the column name, apply cluster by selecting the Edit Cells → Cluster. This opens a window where you can choose different “methods” and “keying functions” to use while clustering. Choose the keying function that produces the largest number of clusters under the “Key Collision” method. Click ‘Select All’ and ‘Merge Selected & Close’. Provide the name of the keying function and the number of clusters produced.  
  


c) Replace the null values in the brand\_name column with the text “Unknown” (Edit Cells → Transform). Hint: read the [GREL instruction](https://openrefine.org/docs/manual/grelfunctions) and lean how to use the "if" function. Provide your GREL code. Click "OK" when the results in preview window are correct. Note: the same as with Excel, you should learn how to find the appropriate function and apply it.  
  


d) Go to the "price" column, choose “Edit cells”, select “Common transformations.”, and choose number as the desired format. Then, use an appropriate function under "Edit Column" to create a new column high\_priced with the values 0 or 1 based on the “price” column with the following conditions: if the price is greater than 50, high\_priced should be set as 1, else 0. Provide your GREL code. Click "OK" when the results in preview window are correct.  
  


e) Create a new column has\_offer with the values 0 or 1 based on the item\_description column with the following conditions: If it contains the text “discount” or “offer” or “sale”, then set the value in has\_offer as 1, else 0. Hint: you can use both if() and or() functions.  
  


4. Summarize and reflect on what you learned throughout this lab assignment. Note that to get the points, you must mention some details.   
  
I learned how to set up and use OpenRefine for data cleaning, including clustering data and utilizing GREL(**General Refine Expression Language)** instructions for tasks like creating new columns and editing cells. Initially, navigating OpenRefine was challenging due to my unfamiliarity with the software, but over time, I became more comfortable with its features. This lab experience has equipped me with valuable skills that I plan to incorporate into my future career and add to my skill set.