Step 1: the number of rows that return True [true](https://ultra.content.blackboardcdn.com/ultra/uiv3900.104.0-rel.17_98a0bf5)6327

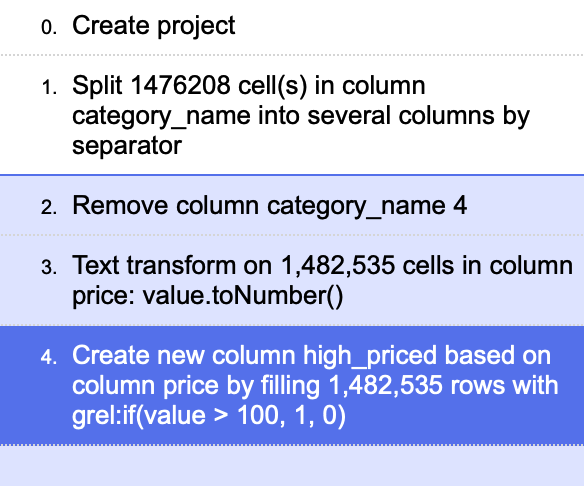
Step 2: the number of columns that are created : 5 ; the columns removed: 4,5

Step 3: the number of clusters found for each collision method:

|  |  |
| --- | --- |
| Method​ | ​nbr of clusters |
| ​Fingerprint | ​110 |
| n-Gram fingerprint | ​75 |
| metaphone3 | 14000​ |
| cologne phonetic | 230​ |
| Daitch-Mokotoff | 22000​ |
| Beider-Morse | 140 |

Step4: if(isNull(value), 'Unbranded', value)

Step5:first I changed the column to number, add column based on column, if(value > 100, 1, 0)



Step 6:

if(contains(value, 'discount'), 1, if(contains(value, 'offer'), 1, if(contains(value, 'sale'), 1, 0)))

