Menu.CSV:

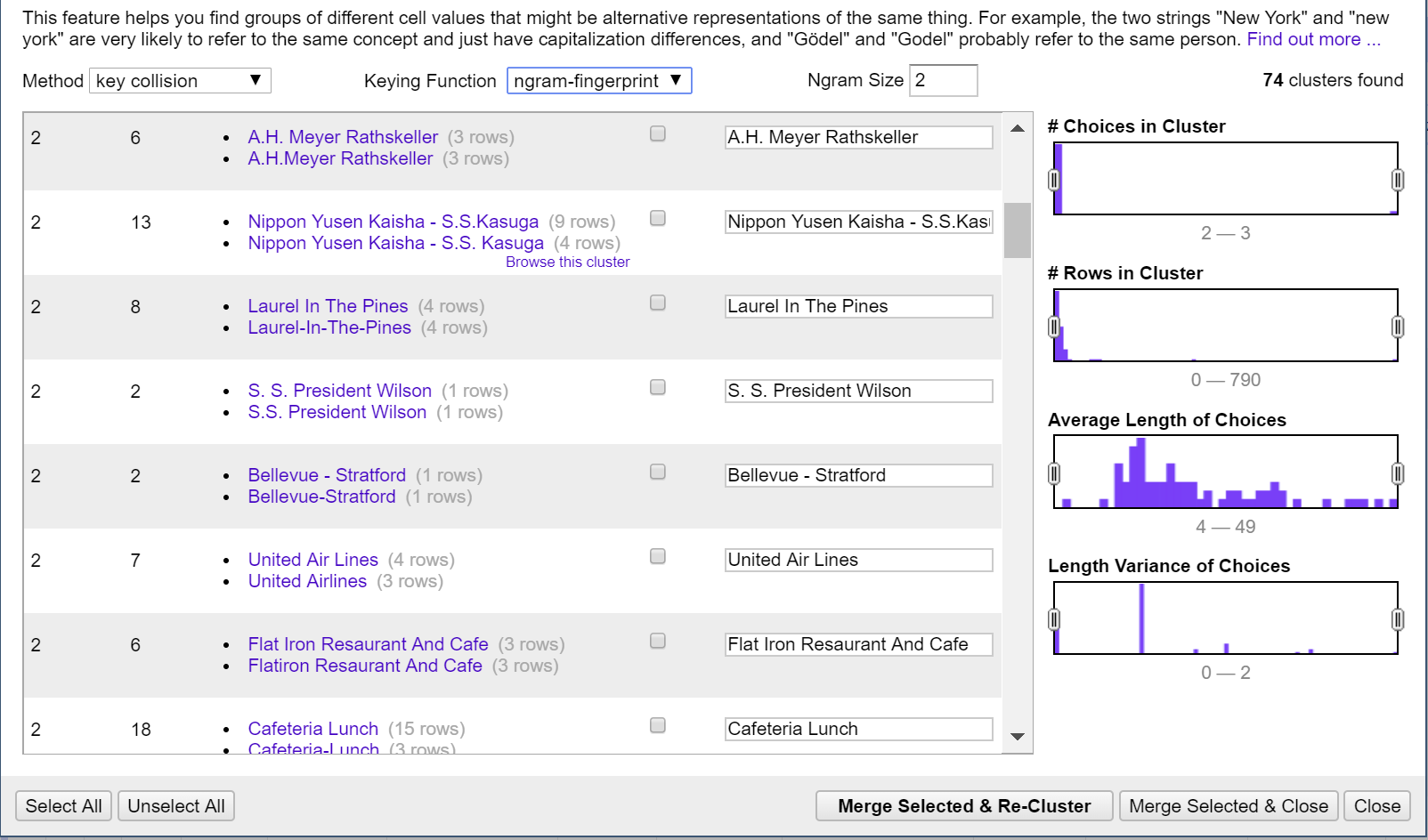
Step 1: Converted Date Field to Date

Step 2: Analyzed Date field outliers and based on that removed outliers

Step 3: Used Clustering from open refine on column “Sponsor”. Following methods were used:

1. Method: Key-Collision, Function: Fingerprint
2. Method: Key-Collision, Function: n-gram fingerprint (n=2)
3. Method: Key-Collision, Function: metaphone3
4. Method: Key-Collision, Function: cologne-phonetic
5. Method: Nearest Neighbor, Function: PPM (Radius: 1, Block Chars: 6)
6. Method: Nearest Neighbor, Function: Levenshtein (Radius: 1, Block Chars: 6)
7. Group updates

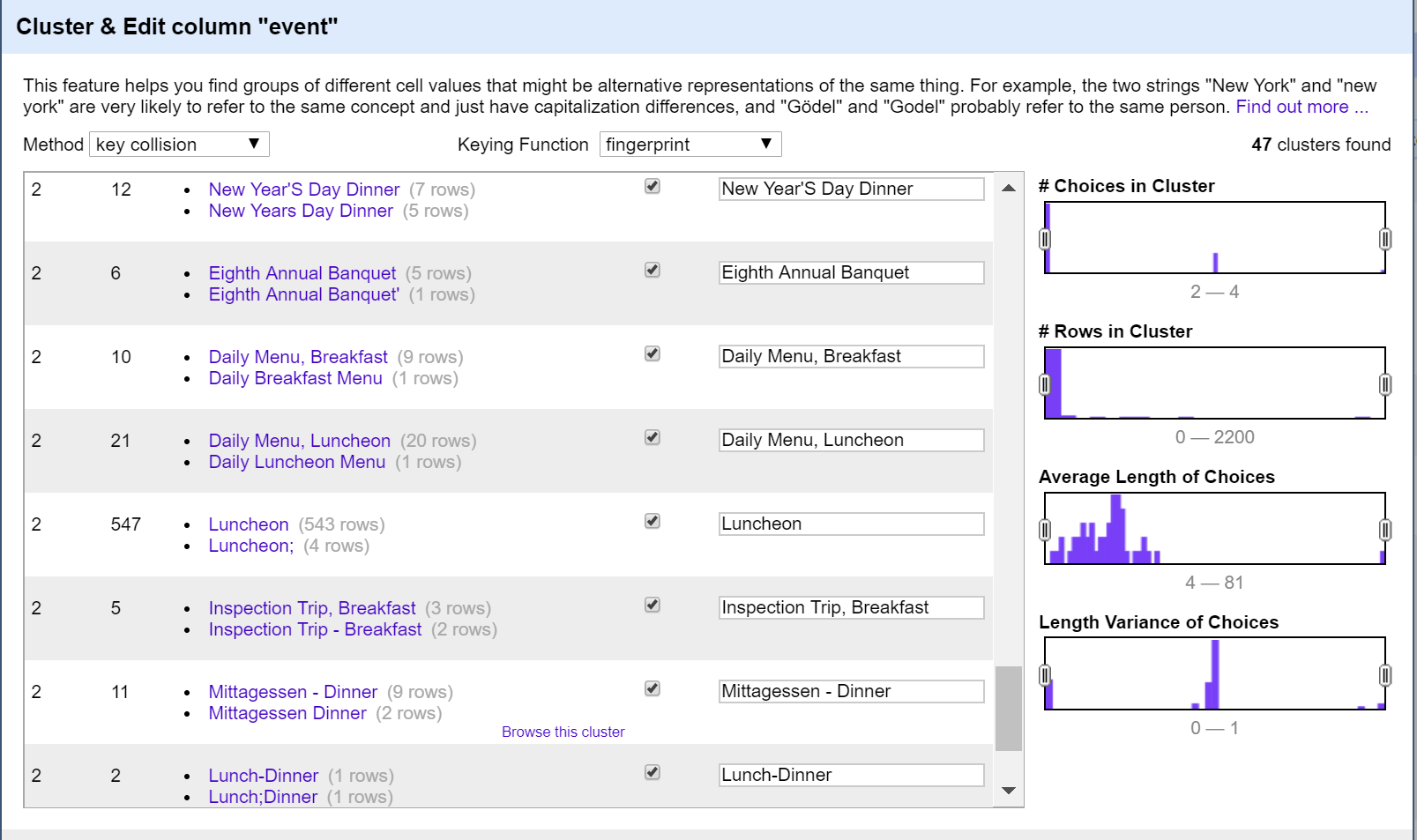
Screenshot:



Step 4: Used Clustering from open refine on column “Event”. Following methods were used:

1. Method: Key-Collision, Function: Fingerprint
2. Method: Key-Collision, Function: n-gram fingerprint (n=2)
3. Method: Key-Collision, Function: metaphone3
4. Method: Key-Collision, Function: cologne-phonetic
5. Method: Nearest Neighbor, Function: PPM (Radius: 1, Block Chars: 6)
6. Method: Nearest Neighbor, Function: Levenshtein (Radius: 1, Block Chars: 6)
7. Group updates

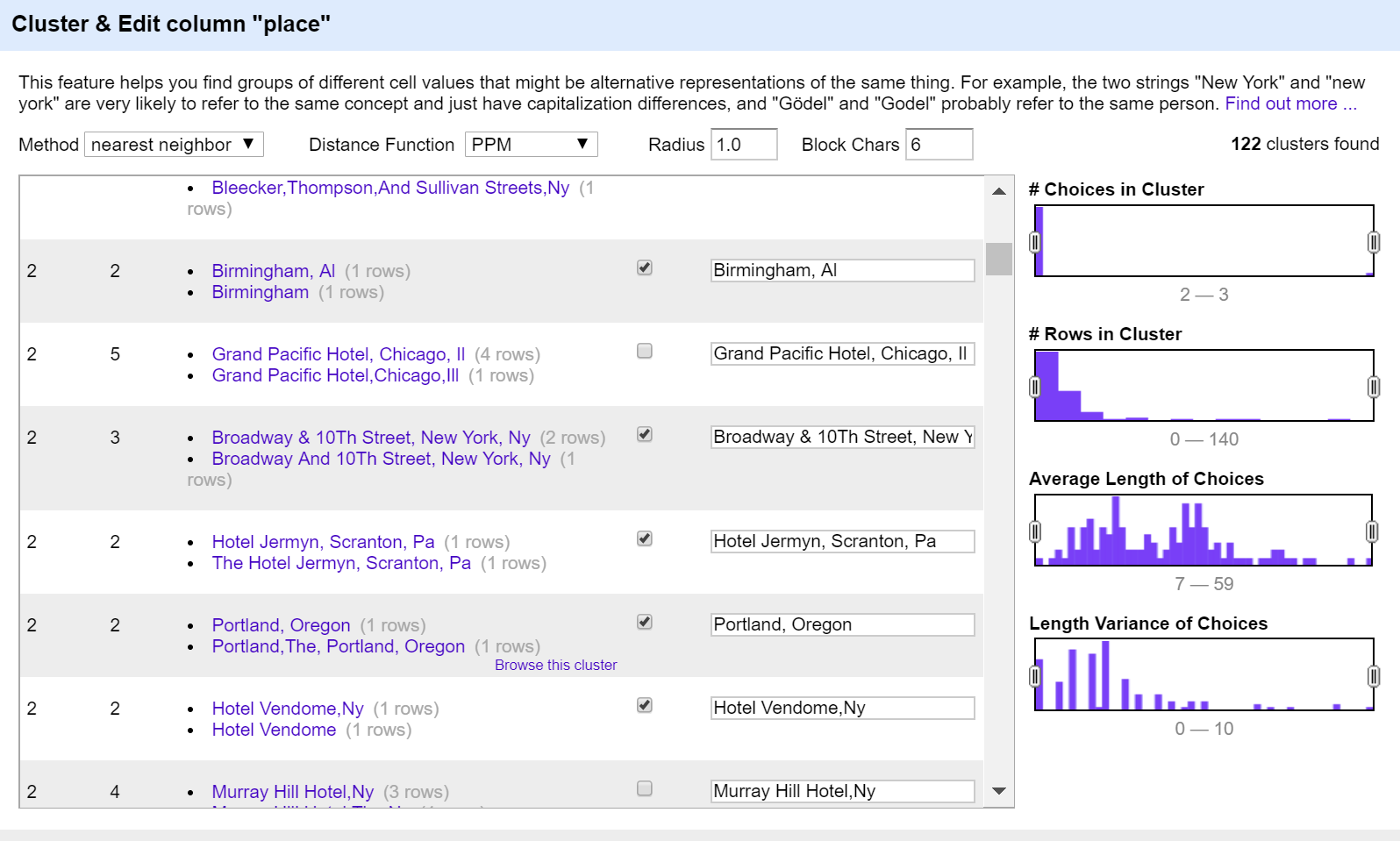
Screenshot:



Step 5: Used Clustering from open refine on column “Place”. Following methods were used:

1. Method: Key-Collision, Function: Fingerprint
2. Method: Key-Collision, Function: n-gram fingerprint (n=2)
3. Method: Key-Collision, Function: metaphone3
4. Method: Key-Collision, Function: cologne-phonetic
5. Method: Nearest Neighbor, Function: PPM (Radius: 1, Block Chars: 6)
6. Method: Nearest Neighbor, Function: Levenshtein (Radius: 1, Block Chars: 6)
7. Group updates

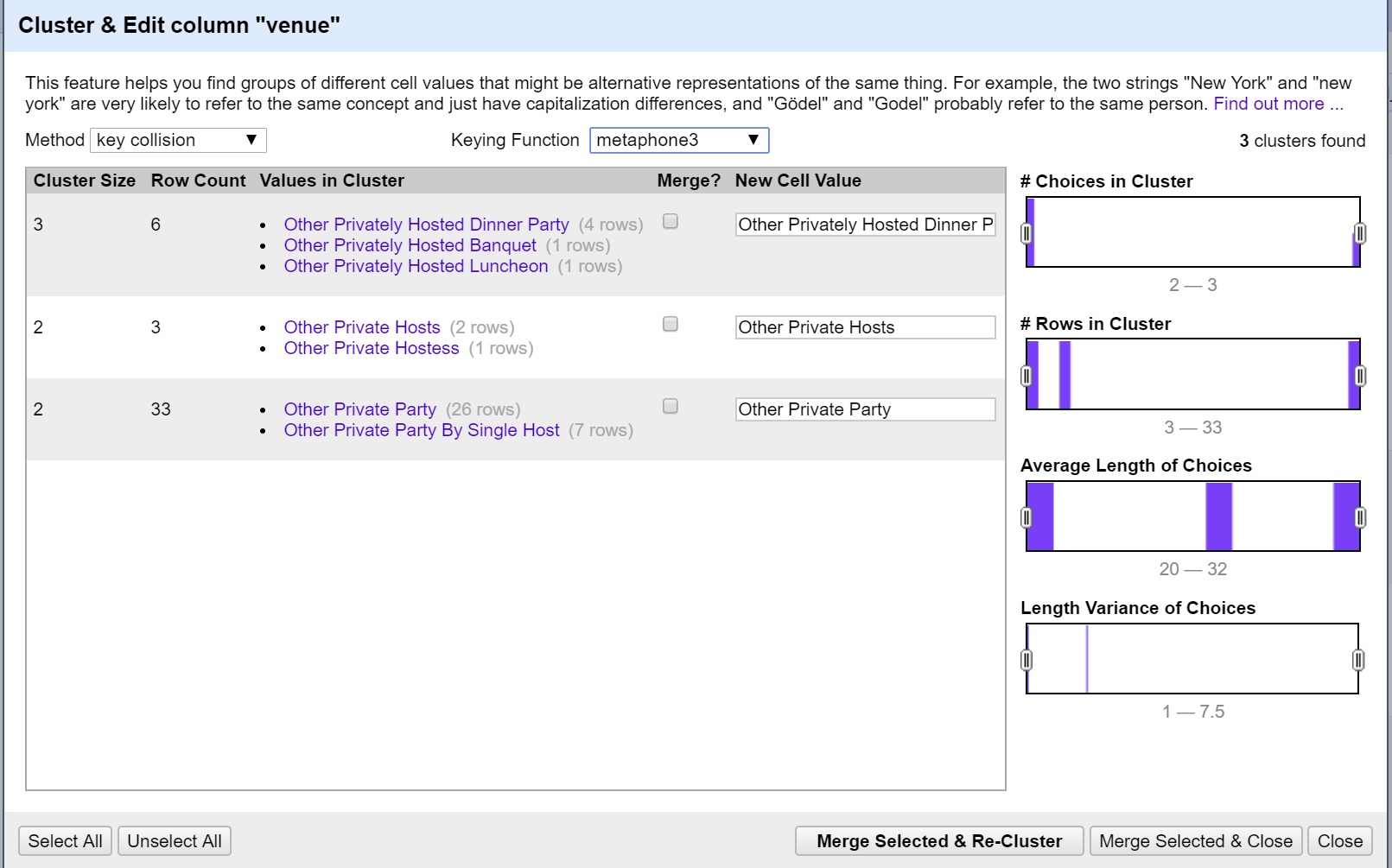
Screenshot:



Step 6: Used Clustering from open refine on column “Venue”. Following methods were used:

1. Method: Key-Collision, Function: Fingerprint
2. Method: Key-Collision, Function: n-gram fingerprint (n=2)
3. Method: Key-Collision, Function: metaphone3
4. Method: Key-Collision, Function: cologne-phonetic
5. Method: Nearest Neighbor, Function: PPM (Radius: 1, Block Chars: 6)
6. Method: Nearest Neighbor, Function: Levenshtein (Radius: 1, Block Chars: 6)
7. Group updates

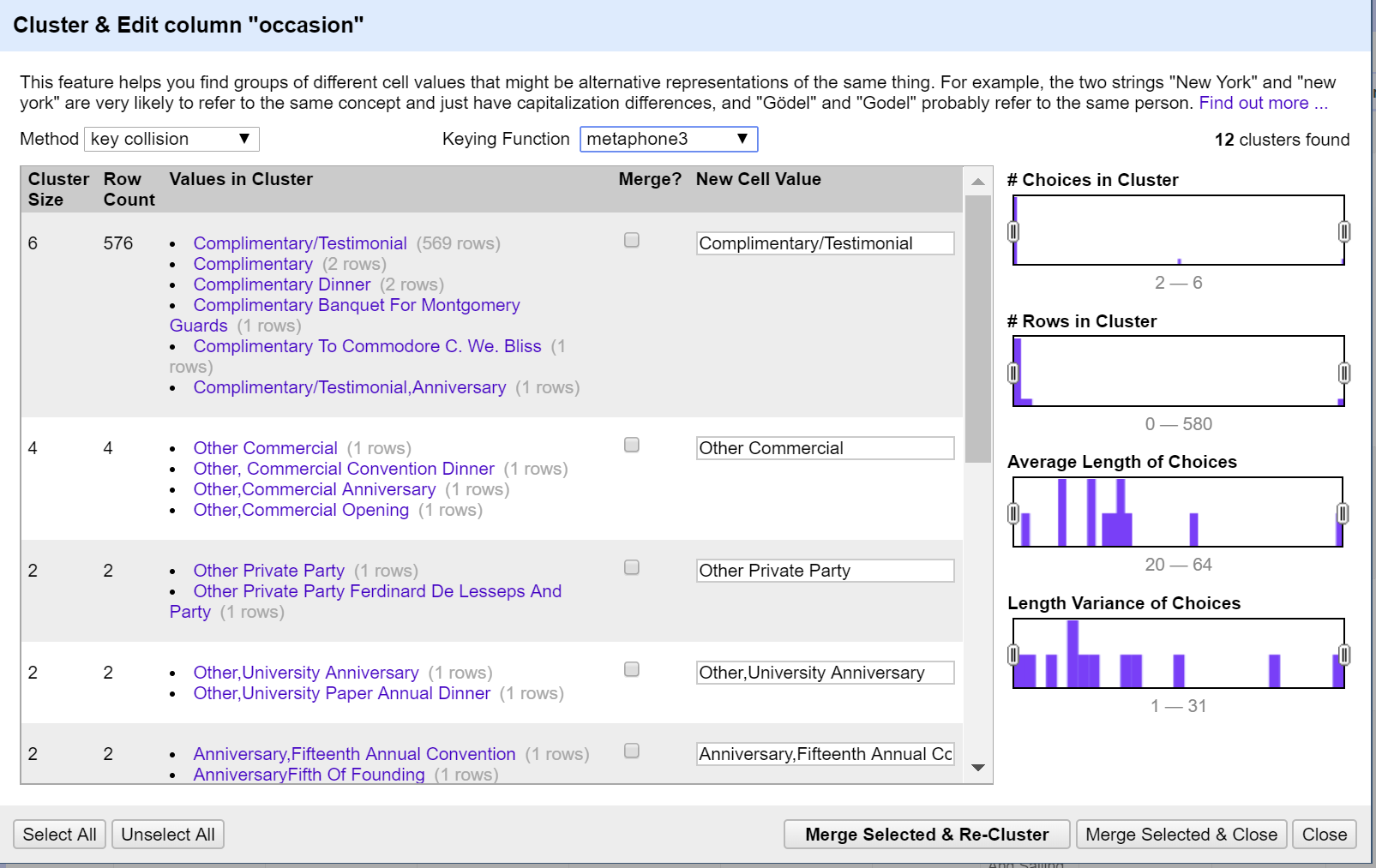
Screenshot:



Step 7: Used Clustering from open refine on column “Occasion”. Following methods were used:

1. Method: Key-Collision, Function: Fingerprint
2. Method: Key-Collision, Function: n-gram fingerprint (n=2)
3. Method: Key-Collision, Function: metaphone3
4. Method: Key-Collision, Function: cologne-phonetic
5. Method: Nearest Neighbor, Function: PPM (Radius: 1, Block Chars: 6)
6. Method: Nearest Neighbor, Function: Levenshtein (Radius: 1, Block Chars: 6)
7. Group updates

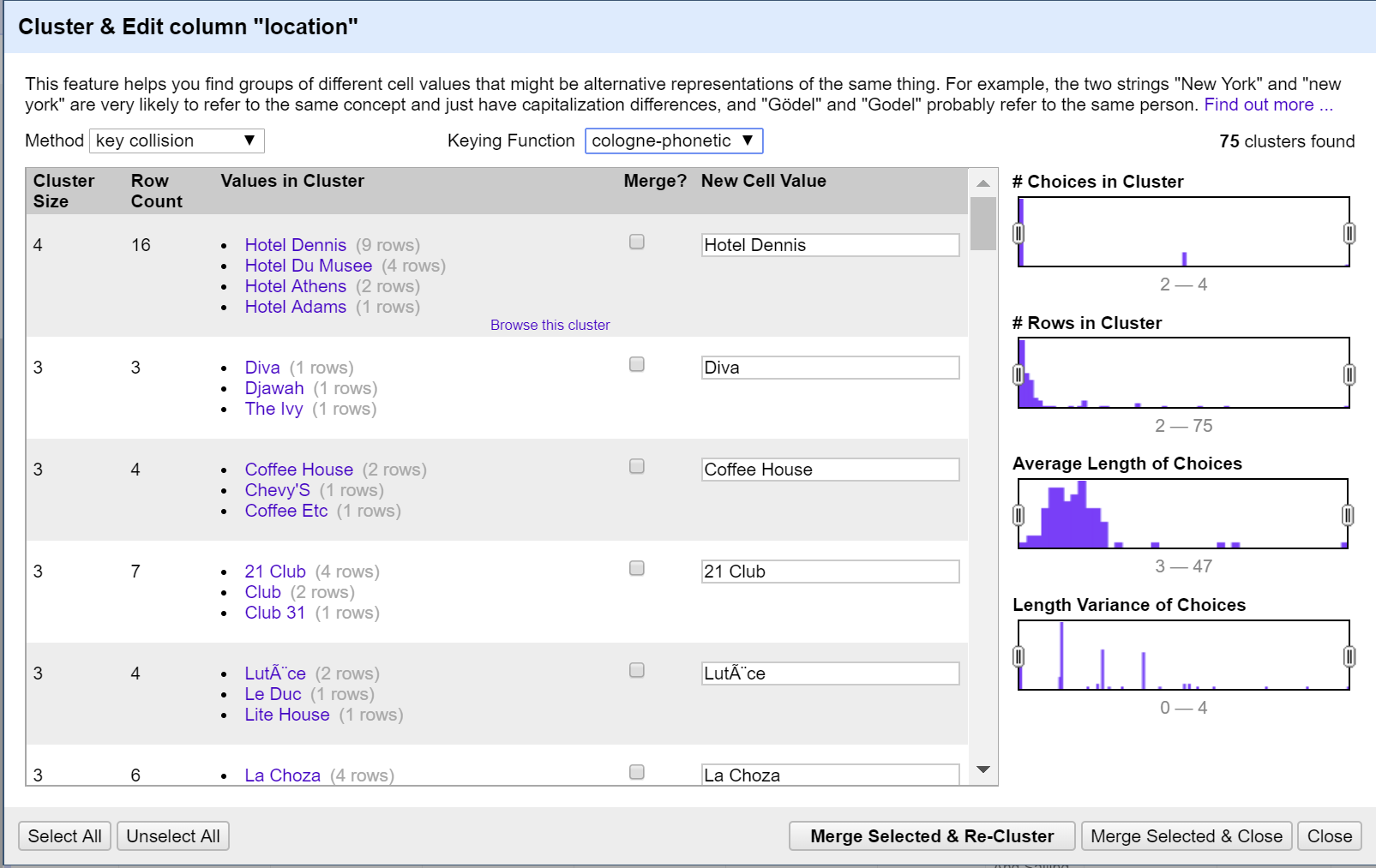
Screenshot:



Step 8: Used Clustering from open refine on column “Location”. Following methods were used:

1. Method: Key-Collision, Function: Fingerprint
2. Method: Key-Collision, Function: n-gram fingerprint (n=2)
3. Method: Key-Collision, Function: metaphone3
4. Method: Key-Collision, Function: cologne-phonetic
5. Method: Nearest Neighbor, Function: PPM (Radius: 1, Block Chars: 6)
6. Method: Nearest Neighbor, Function: Levenshtein (Radius: 1, Block Chars: 6)
7. Group updates

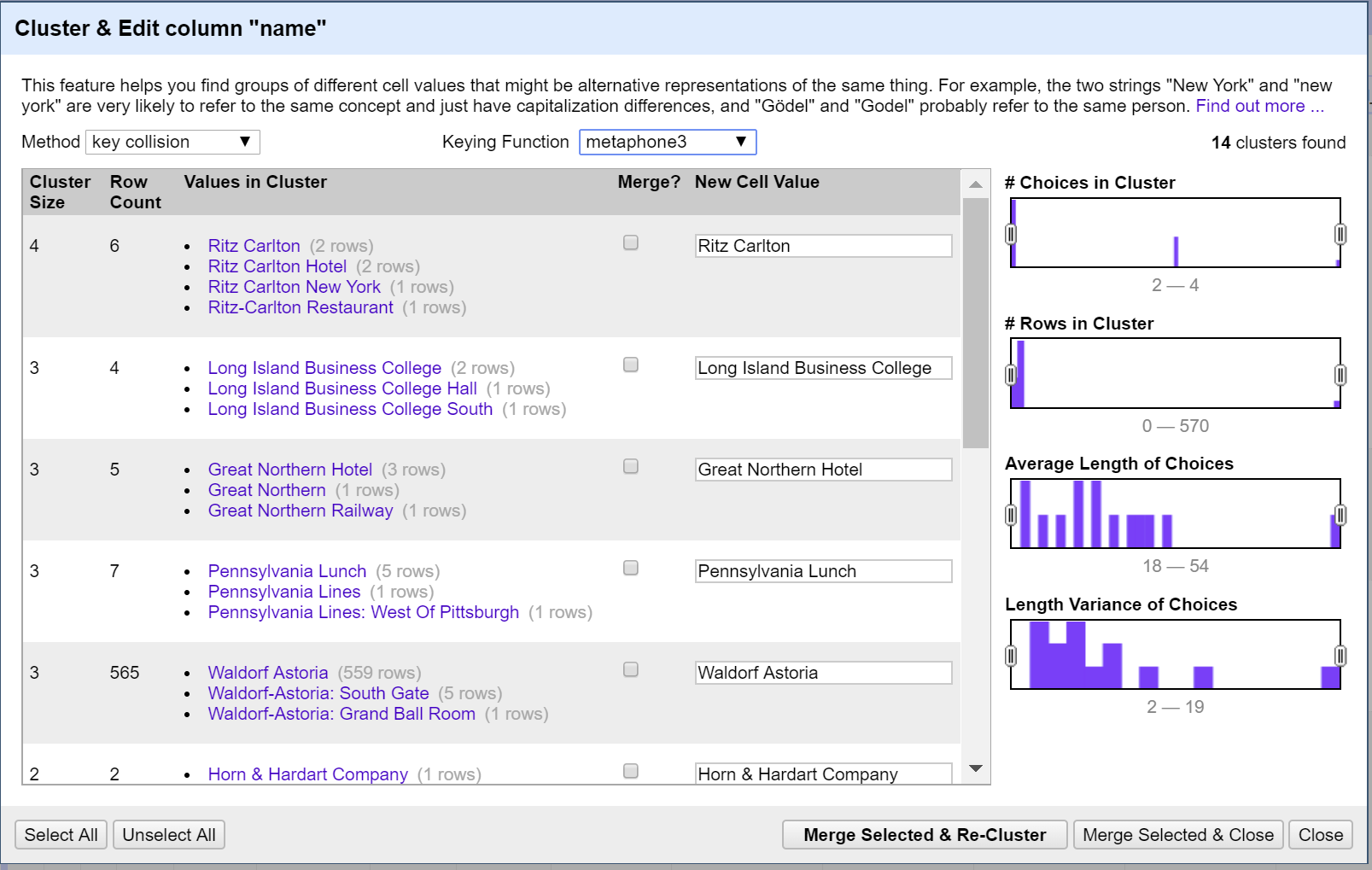
Screenshot:



Step 9: Used Clustering from open refine on column “Name”. Following methods were used:

1. Method: Key-Collision, Function: Fingerprint
2. Method: Key-Collision, Function: n-gram fingerprint (n=2)
3. Method: Key-Collision, Function: metaphone3
4. Method: Key-Collision, Function: cologne-phonetic
5. Method: Nearest Neighbor, Function: PPM (Radius: 1, Block Chars: 6)
6. Method: Nearest Neighbor, Function: Levenshtein (Radius: 1, Block Chars: 6)
7. Group updates

Screenshot:



Step 10: Split the physical\_description column. It resulted in 7 new columns:

physical\_description 1

physical\_description 2

physical\_description 3

physical\_description 4

physical\_description 5

physical\_description 6

physical\_description 7

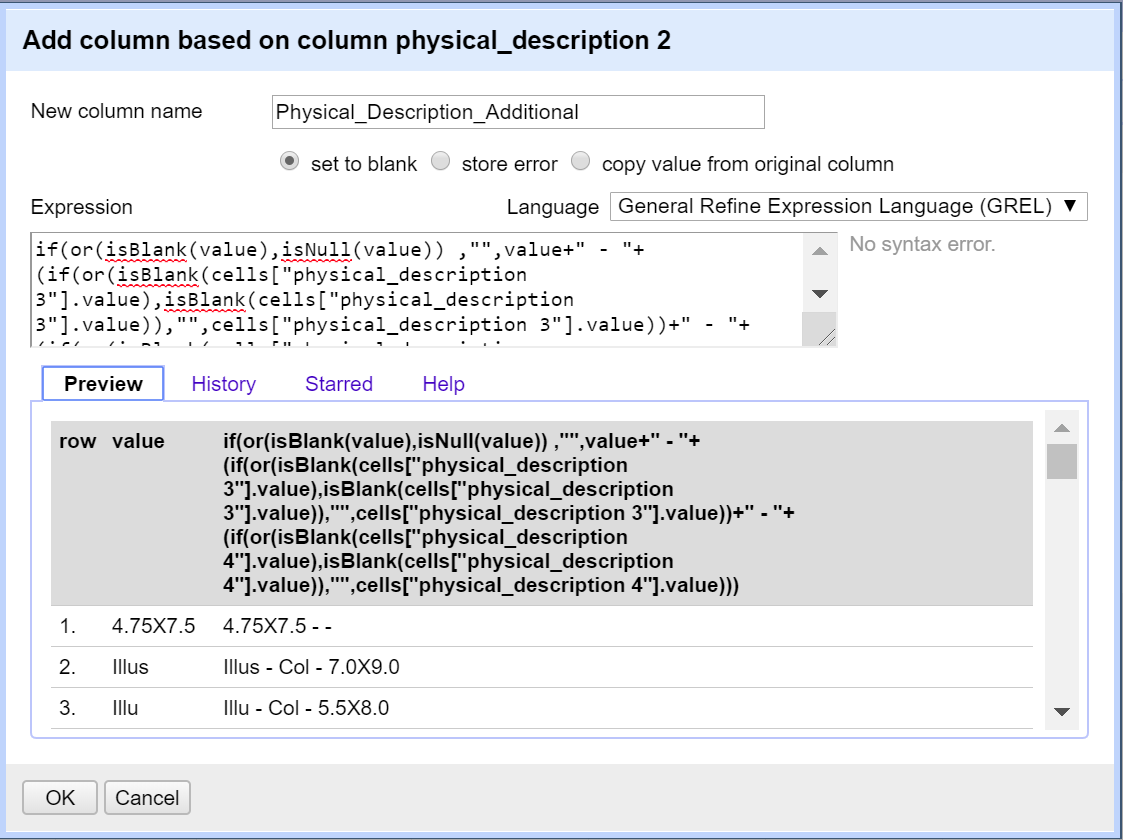
Updated the column physical\_description 1 to physical description type and then created a new column Physical\_description\_additional by merging following columns:

physical\_description 2

physical\_description 3

physical\_description 4

Screenshot:



Dish.csv:

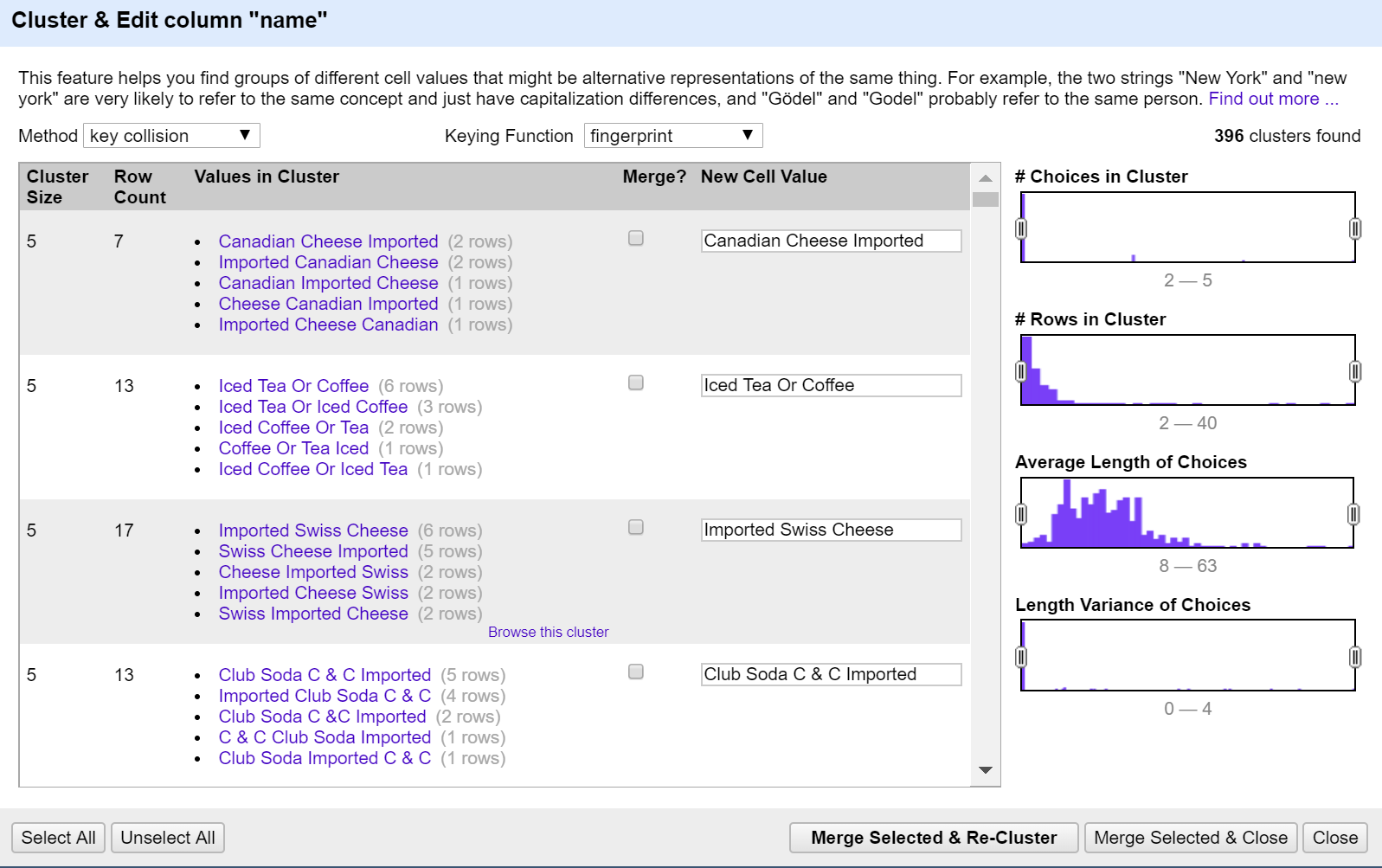
Step 11: Transform data to remove unwanted characters. Formula:

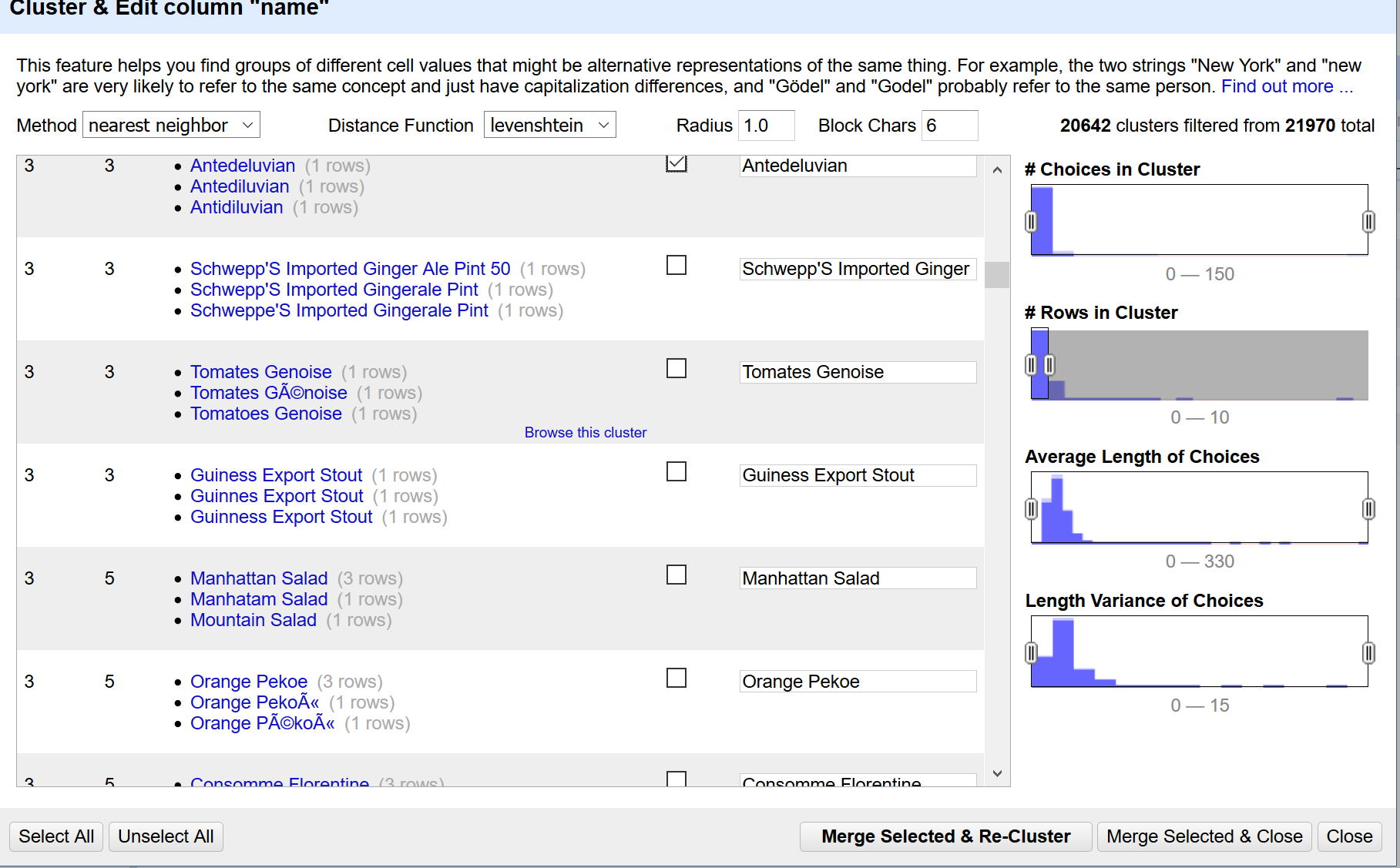
value.replace(/[>:<%#@!\\()\[\]\?\"\-\\*,\.\+]/, " ").replace(/\s+/," ").trim()

Step 12: Used Clustering from open refine on column “Name”. Following methods were used:

1. Method: Key-Collision, Function: Fingerprint
2. Method: Key-Collision, Function: n-gram fingerprint (n=2)
3. Method: Key-Collision, Function: metaphone3
4. Method: Key-Collision, Function: cologne-phonetic
5. Method: Nearest Neighbor, Function: PPM (Radius: 1, Block Chars: 6)
6. Method: Nearest Neighbor, Function: Levenshtein (Radius: 1, Block Chars: 6)
7. Group updates

Screenshot:





MenuItem.csv:

Step 12: Transform created\_at field to date by creating a new filed created\_date

Step 13: Transformed updated\_at field to date by creating a new filed updated\_date.

(Note: both created\_at and created\_date field exists in the csv file. Similarly, both updated \_at and updated \_date field exists in the database.)

Step 14: Transform xpos field to number

Step 15: Transformed ypos field to number

MenuPage.cs

No transformation were performed for this csv in open refine.