

Data warehousing report

Group:

Jiyang Chen(A20490982)

Yang Li(A20478590)

Mingyuan Wang(A20501460)

Contents:

- [Introduce]
- [Data cleaning]
- [Data integration]
- [Data extraction]

Introduce

building Table Columns Schema

Column Name	Data Type	Description
Building_ID	INTEGER	The unique identifier for the building
Address	VARCHAR(255)	The street address of the building
Building_Type	VARCHAR(100)	The classification of the building (e.g., residential, commercial)
Build_Year	INTEGER	The year the building was constructed
Zip	VARCHAR(10)	The postal code for the building
State	VARCHAR(2)	The U.S. state abbreviation where the building is located
City	VARCHAR(100)	The city where the building is located
Community_Name	VARCHAR(255)	The name of the community or neighborhood
Community_Number	INTEGER	The numeric identifier for the community
Latitude	DECIMAL(9,6)	The latitude coordinate of the building
Longitude	DECIMAL(9,6)	The longitude coordinate of the building
Location	VARCHAR(255)	The combined latitude and longitude information

The following constraints have been defined for the dataset:

Functional Dependencies

Building_ID → Address

Zip → State

Zip → City

Address → Location

Latitude, Longitude → Location

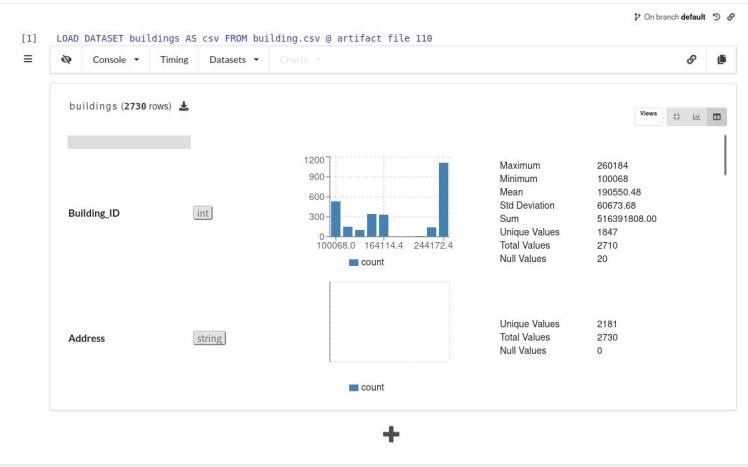
Location, Community_Number → Community_Name

🔑 Data cleaning

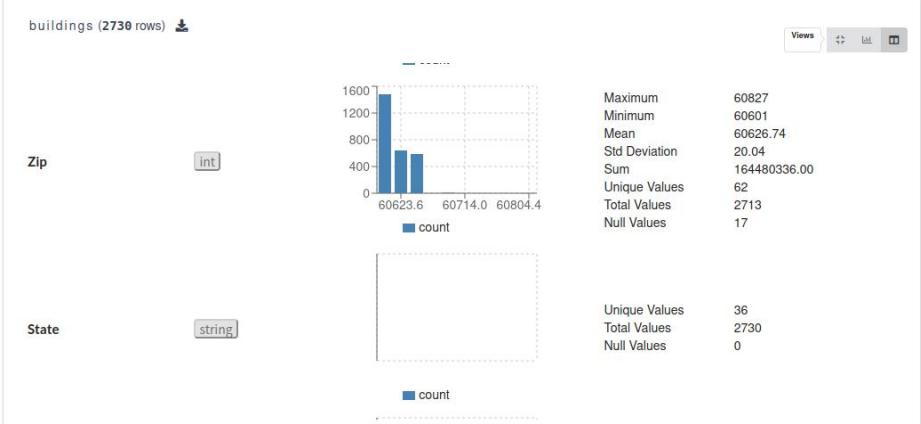
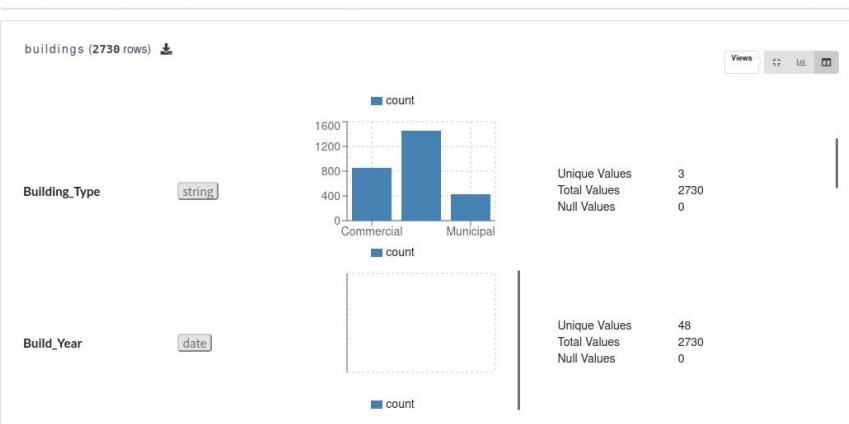
Data import into VizierDB:

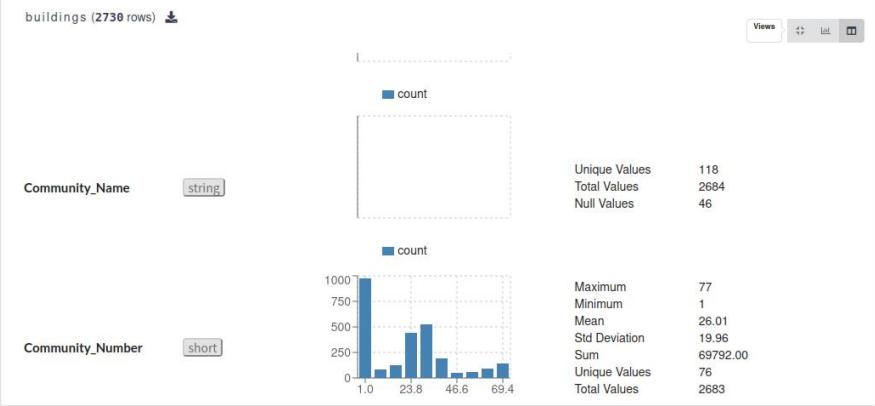
	Building_ID (int)	Address (string)	Building_Type (string)	Build_Year (date)	Zip (int)	State (string)	City (string)	Community_Name (string)	Community_Number (int)
0	101717	55 E Monroe St	Commercial	2010-02-01	60603	IL	Chicago	LOOP	32
1	101717	55 E Monroe St	Commercial	2010-02-01	60603	IL	Chicago	DOUGLAS	35
2	101717	55 E Monroe St	Commercial	2010-02-01	60603	IL	Chicago	ARMOUR SQUARE	34
3	101717	55 E Monroe St	Commercial	2010-02-01	60603	IL	Chicago	LINCOLN PARK	7
4	101717	55 E Monroe St, "	Commercial	2008-02-01	60603	IL	Chicago	LOOP	32
5	172786	619 645 E WOODLAND PARK	Residential	2008-02-01	60616	IL	Chicago	DOUGLAS	35
6	172786	619 645 E WOODLAND PARK	Residential	2008-02-01	60616	IL	Chicago	LINCOLN PARK	7
7	172786	619 645 E WOODLAND PARK	Residential	2008-02-01	60616	IL	Chicago, "	UPTOWN	3
8	172786	619 645 E WOODLAND PARK	Residential	2008-02-01	60616	IL	Chicago	WEST GARFIELD PARK	26
9	172786	619 645 E WOODLAND PARK, "	Residential	2008-02-01	60616	IL	Chicago	DOUGLAS	35
10	252251	300 W 23rd St	Residential	2008-03-01	60616	IL	Chicago	ARMOUR SQUARE	34
11	252251	300 W 23rd St	Residential	2008-03-01	60616	ILLINOIS	Aurora	DOUGLAS	35
12	252251	300 W 23rd St	Residential	2008-03-01	60616	ILLINOIS	Aurora, "	HYDE PARK	41
13	252251	300 W 23rd St	Residential	2008-03-01	60616	ILLINOIS	Aurora	LINCOLN PARK	7
14	252251	300 W 23rd St, "	Residential	2008-03-01	60616	ILLINOIS	Aurora	ARMOUR SQUARE	34
15	251331	2312 N Clifton Ave	Commercial	2008-03-01	60614	ILLINOIS	Aurora	LINCOLN PARK	7
16	251331	2312 N Clifton Ave	Commercial	2008-03-01	60614	ILLINOIS	Aurora	LINCOLN PARK	7
17	251331	2312 N Clifton Ave	Commercial	2008-03-01	60614	ILLINOIS, "	Aurora	LOOP	32
18	251331	2312 N Clifton Ave	Commercial	2008-03-01	60614	ILLINOIS	Aurora	NEAR WEST SIDE	28
19	251331	2312 N Clifton Ave, "	Commercial	2008-03-01	60614	ILLINOIS	Aurora	LINCOLN PARK	7

Viewing individual column information, You can see a lot of nulls and duplicates by the information in the columns:

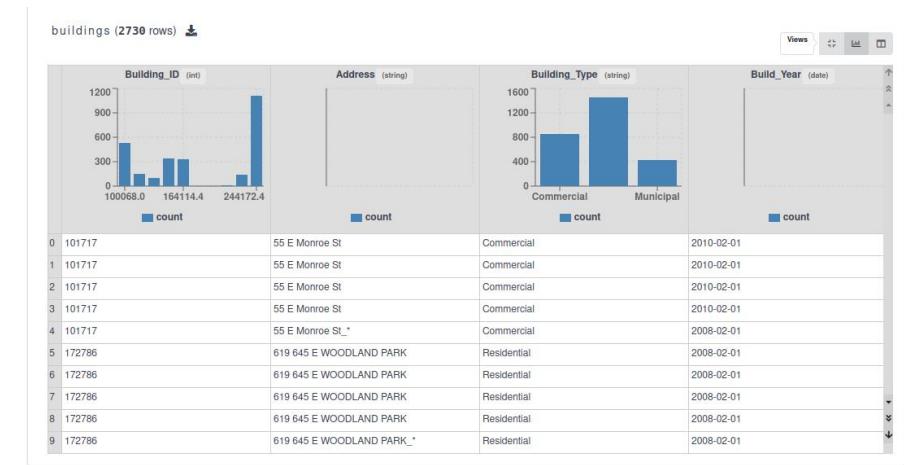


Connected to vizier @ http://localhost:5001/vizier-db/api/v1/





Viewing the distribution of column values:



Detecting Constraint Violations:

[2]

```
SELECT Building_ID,COUNT(DISTINCT Address) AS num_addresses
FROM buildings
GROUP BY Building_ID
HAVING COUNT(DISTINCT Address) > 1;;
```

e0_check (232 rows)

	Building_ID (int)	num_addresses (long)
0	251742	2
1	254564	2
2	251529	2
3	152061	2
4	165664	2
5	256568	2
6	103933	2
7	102809	2
8	251535	2
9	101887	2
10	172307	2
11	250146	2
12	172795	2
13	254362	2
14	239042	2
15	160349	2
16	251581	2
17	251408	2
18	251382	2
19	251469	2

[3]

```
SELECT a.Building_ID, a.Address AS Address1, b.Address AS Address2
FROM buildings a, buildings b
WHERE a.Building_ID = b.Building_ID AND a.Address <> b.Address;
```

e0_pairs_check (1848 rows)

	Building_ID (int)	Address1 (string)	Address2 (string)
0	101717	55 E Monroe St_*	55 E Monroe St
1	101717	55 E Monroe St_*	55 E Monroe St
2	101717	55 E Monroe St_*	55 E Monroe St
3	101717	55 E Monroe St_*	55 E Monroe St
4	101717	55 E Monroe St	55 E Monroe St_*
5	101717	55 E Monroe St	55 E Monroe St_*
6	101717	55 E Monroe St	55 E Monroe St_*
7	101717	55 E Monroe St	55 E Monroe St_*
8	172786	619 645 E WOODLAND PARK_*	619 645 E WOODLAND PARK
9	172786	619 645 E WOODLAND PARK_*	619 645 E WOODLAND PARK
10	172786	619 645 E WOODLAND PARK_*	619 645 E WOODLAND PARK
11	172786	619 645 E WOODLAND PARK_*	619 645 E WOODLAND PARK
12	172786	619 645 E WOODLAND PARK	619 645 E WOODLAND PARK_*
13	172786	619 645 E WOODLAND PARK	619 645 E WOODLAND PARK_*
14	172786	619 645 E WOODLAND PARK	619 645 E WOODLAND PARK_*
15	172786	619 645 E WOODLAND PARK	619 645 E WOODLAND PARK_*
16	252251	300 W 23rd St_*	300 W 23rd St
17	252251	300 W 23rd St_*	300 W 23rd St
18	252251	300 W 23rd St_*	300 W 23rd St
19	252251	300 W 23rd St_*	300 W 23rd St

Checking for data with a null ID:

[4]

```
SELECT * FROM buildings WHERE Building_ID IS NULL OR Building_ID = '';
```

empty_ids (28 rows)

Building_ID	Address	Building_Type	Build_Year	Zip	State	City	Community_Name	Community_Nu
0	2754 N HAMPDEN CT	Residential	1925-02-01	60614	NY	Fishers Island	LINCOLN PARK	7
1	223 W Erie St	Commercial	1925-02-01	60654	NY	Fishers Island	NEAR NORTH SIDE	8
2	2021 W Chicago Ave	Commercial	1925-02-01	60622	NY	Fishers Island	WEST TOWN	24
3	4626 N Knox Ave	Commercial	1925-02-01	60630	NY	Fishers Island	IRVING PARK	16
4	2600 N Southport Ave	Residential	1925-02-01	60614	NY	Fishers Island	LINCOLN PARK	7
5	30 E Lake St	Municipal	1925-02-01	60601	NY	Fishers Island	LOOP	32
6	6166 N SHERIDAN RD	Residential	1925-02-01	60660	NY	Fishers Island	EDGEWATER	77
7	3630 N Clark St	Residential	1925-02-01	60613	NY	Fishers Island	LAKE VIEW	6
8	139 S Parkside Ave	Municipal	1925-02-01	60644	NY	Fishers Island	AUSTIN	25
9	648 E 91st Pl	Municipal	1925-02-01	60619	NY	Fishers Island	CHATHAM	44
10	7200 S SOUTH SHORE DR	Residential	1925-02-01	60649	NY	Fishers Island	SOUTH SHORE	43
11	1040 N Keeler Ave	Municipal	1925-02-01	60651	NY	Fishers Island	HUMBOLDT PARK	23
12	1244 North Clybourn	Residential	1925-02-01	60610	NY	Fishers Island	Near North Side	8
13	6345 S Stewart	Residential	1925-02-01	60621	NY	Fishers Island	ENGLEWOOD	68
14	3800 N CALIFORNIA AVE	Residential	1925-02-01	60618	NY	Fishers Island	IRVING PARK	16
15	6165 N WINTHROP AVE	Residential	1925-02-01	60660	NY	Fishers Island	EDGEWATER	77
16	2720 2730 N Greenview Ave	Residential	1925-02-01	60614	NY	Fishers Island	LINCOLN PARK	7
17	1744 W Pryor Ave	Municipal	1925-02-01	60643	NY	Fishers Island	MORGAN PARK	75
18	3300 3306 W DIVERSEY AVE	Residential	1925-02-01	60647	NY	Fishers Island	AVONDALE	21
19	125 S Clark St	Commercial	1925-02-01	60603	NY	Fishers Island	LOOP	32

+

19	125 S Clark St	Commercial	1925-02-01	60603	NY	Fishers Island	LOOP	32
----	----------------	------------	------------	-------	----	----------------	------	----

[5]

```
ds = vizierdb.get_dataset('buildings')

for index, row in enumerate(ds.rows):
    if row['Building_ID'] is None:
        print(f"Index: {index}, Row: {row}")
```

Index: 2710, Row: <None, 2754 N HAMPDEN CT, Residential, 1925-02-01, 60614, NY, Fishers Island, LINCOLN PARK, 7, 41.93215942
Index: 2711, Row: <None, 223 W Erie St, Commercial, 1925-02-01, 60654, NY, Fishers Island, NEAR NORTH SIDE, 8, 41.8937454223
Index: 2712, Row: <None, 2021 W Chicago Ave, Commercial, 1925-02-01, 60622, NY, Fishers Island, WEST TOWN, 24, 41.895268859
Index: 2713, Row: <None, 4626 N Knox Ave, Commercial, 1925-02-01, 60630, NY, Fishers Island, IRVING PARK, 16, 41.96533203125
Index: 2714, Row: <None, 2600 N Southport Ave, Residential, 1925-02-01, 60614, NY, Fishers Island, LINCOLN PARK, 7, 41.92905
Index: 2715, Row: <None, 30 E Lake St, Municipal, 1925-02-01, 60601, NY, Fishers Island, LOOP, 32, 41.8866643676758, -87.62
Index: 2716, Row: <None, 6166 N SHERIDAN RD, Residential, 1925-02-01, 60660, NY, Fishers Island, EDGEWATER, 77, 41.994316101
Index: 2717, Row: <None, 3630 N Clark St, Residential, 1925-02-01, 60613, NY, Fishers Island, LAKE VIEW, 6, 41.9548110961914
Index: 2718, Row: <None, 139 S Parkside Ave, Municipal, 1925-02-01, 60644, NY, Fishers Island, AUSTIN, 25, 41.87814712524414
Index: 2719, Row: <None, 648 E 91st Pl, Municipal, 1925-02-01, 60619, NY, Fishers Island, CHATHAM, 44, 41.72842025756836, -8
Index: 2720, Row: <None, 7200 S SOUTH SHORE DR, Residential, 1925-02-01, 60649, NY, Fishers Island, SOUTH SHORE, 43, 41.762
Index: 2721, Row: <None, 1040 N Keeler Ave, Municipal, 1925-02-01, 60651, NY, Fishers Island, HUMBOLDT PARK, 23, 41.90030288
Index: 2722, Row: <None, 1244 North Clybourn, Residential, 1925-02-01, 60610, NY, Fishers Island, Near North Side, 8, 41.994
Index: 2723, Row: <None, 6345 S Stewart, Residential, 1925-02-01, 60621, NY, Fishers Island, ENGLEWOOD, 68, 41.7785148620605
Index: 2724, Row: <None, 3800 N CALIFORNIA AVE, Residential, 1925-02-01, 60618, NY, Fishers Island, IRVING PARK, 16, 41.9506
Index: 2725, Row: <None, 6165 N WINTHROP AVE, Residential, 1925-02-01, 60660, NY, Fishers Island, EDGEWATER, 77, 41.9909019
Index: 2726, Row: <None, 2720 2730 N Greenview Ave, Residential, 1925-02-01, 60614, NY, Fishers Island, LINCOLN PARK, 7, 41.
Index: 2727, Row: <None, 1744 W Pryor Ave, Municipal, 1925-02-01, 60643, NY, Fishers Island, MORGAN PARK, 75, 41.69164657592
Index: 2728, Row: <None, 3300 3306 W DIVERSEY AVE, Residential, 1925-02-01, 60647, NY, Fishers Island, AVONDALE, 21, 41.9287
Index: 2729, Row: <None, 125 S Clark St, Commercial, 1925-02-01, 60603, NY, Fishers Island, LOOP, 32, 41.8799934387207, -87.

161

fixes data with empty ID, fix ID to the maximum ID value plus 1:

The screenshot shows a Jupyter Notebook environment. Cell [6] contains Python code to update a dataset named 'buildings'. It reads the dataset, finds the maximum Building_ID, and updates rows where Building_ID is None. Cell [7] shows the resulting temporary dataset with 0 rows.

```
[6]
ds = vizierdb.get_dataset('buildings')
max_id = max([row['Building_ID'] for row in ds.rows if row['Building_ID'] is not None])
for row in ds.rows:
    if row['Building_ID'] is None:
        max_id += 1
        row['Building_ID'] = max_id

# Update existing dataset with given name.
vizierdb.update_dataset('buildings', ds)

[7]
SELECT * FROM buildings WHERE Building_ID IS NULL OR Building_ID = '';
```

temporary_dataset (0 rows)

Connected to vizier @ http://localhost:5001/vizier-db/api/v1/

Now start fixing the e0 constraint data: the program uses "Update Equivalent Right Side" to fix the violation, using the frequency of each class this strategy: select the value with the highest number of occurrences in each of the right hand side attributes of the equivalence class, and select the value with the highest number of occurrences of Addresses under the Building_ID category Update:

The screenshot shows a Jupyter Notebook environment. Cell [11] contains Python code to fix e0 constraint data. It iterates through rows, updates violations dictionary, and prints updated rows. Cell [34] shows the final output of the print statements.

```
[11]
from collections import defaultdict
# e0_check
e0_check = vizierdb.get_dataset('e0_check')
# buildings
buildings = vizierdb.get_dataset('buildings')
violations = defaultdict(lambda: defaultdict(int))

# e0_check
for row in e0_check.rows:
    violations[row['Building_ID']] # dict

for row in buildings.rows:
    building_id = row['Building_ID']
    if building_id in violations:
        violations[building_id][row['Address']] += 1

updated_rows = []
for row in buildings.rows:
    building_id = row['Building_ID']
    if building_id in violations:
        most_common_address = max(violations[building_id], key=violations[building_id].get)
        if row['Address'] != most_common_address:
            row['Address'] = most_common_address
            updated_rows.append((building_id, most_common_address))

for building_id, address in updated_rows:
    print(f"Building ID {building_id} updated to address: {address}")

vizierdb.update_dataset('buildings', buildings)
```

Building ID 101717 updated to address: 55 E Monroe St
 Building ID 172786 updated to address: 619 645 E WOODLAND PARK
 Building ID 252251 updated to address: 300 W 23rd St
 Building ID 251331 updated to address: 2312 N Clifton Ave
 Building ID 251254 updated to address: 720 W Gordon Ter

Checking after fixing,The e0 constraint data has been successfully fixed:

[12]

```
Building_ID 242661 undated_to address: 2921 S MCHTGAN AVE
```

```
SELECT Building_ID,COUNT(DISTINCT Address) AS num_address
FROM buildings
GROUP BY Building_ID
HAVING COUNT(DISTINCT Address) > 1;
```

[13]

```
temporary_dataset (0 rows)
```

```
SELECT a.Building_ID, a.Address AS Address1, b.Address AS Address2
FROM buildings a, buildings b
WHERE a.Building_ID = b.Building_ID AND a.Address <> b.Address;
```

Export the fixed data and import the exported data into openRefine:

OpenRefine dataset 246 csv Permalink

Facet / Filter Undo / Redo 0 / 0

Using facets and filters

Use facets and filters to select subsets of your data to act on. Choose facet and filter methods from the menus at the top of each data column.

Not sure how to get started?
Watch these screencasts

2730 rows

All	Building_ID	Address	Building_Type	Built_Year	Zip	State	City	Community_Name	Community_Number	Latitude	Longitude	Location	
1	101717	55 E Monroe St	Commercial	2010-03-01	60603	IL	Chicago	LOOP	32	41.880177	-87.6594	(41.8801573, -87.6564433)	
2	101717	55 E Monroe St	Commercial	2010-03-01	60603	IL	Chicago	DOUGLAS	35	41.845505	-87.6305	(41.845505, -87.63056693)	
3	101717	55 E Monroe St	Commercial	2010-03-01	60603	IL	Chicago	ARMOUR SQUARE	34	41.85050	-87.6351	(41.8505033, -87.63518007)	
4	101717	55 E Monroe St	Commercial	2010-03-01	60603	IL	Chicago	LINCOLN PARK	7	41.92946	-87.6579	(41.9294692, -87.6579559)	
5	101717	55 E Monroe St	Commercial	2008-03-01	60603	IL	Chicago	LOOP	32	41.880177	-87.6594	(41.8801573, -87.6564433)	
6	172786	610 W 445 E WOODLAND PARK	Residential	2008-03-01	60616	IL	Chicago	DOUGLAS	35	41.845505	-87.6305	(41.845505, -87.63056693)	
7	172786	610 W 445 E WOODLAND PARK	Residential	2008-03-01	60616	IL	Chicago	WOODLAND PARK	7	41.92946	-87.6579	(41.9294692, -87.6579559)	
8	172786	610 W 445 E WOODLAND PARK	Residential	2008-03-01	60616	IL	Chicago	UPTOWN	3	41.95817	-87.6483	(41.9581769, -87.64866912)	
9	172786	610 W 445 E WOODLAND PARK	Residential	2008-03-01	60616	IL	Chicago	WEST GARFIELD PARK	28	41.878708	-87.3570	(41.878708, -87.3579)	
10	172786	610 W 445 E WOODLAND PARK	Residential	2008-03-01	60616	IL	Chicago	DOUGLAS	35	41.845505	-87.6305	(41.845505, -87.63056693)	
11	252251	300 W 23rd St	Residential	2008-04-01	60616	IL	Chicago	ARMOUR SQUARE	34	41.85050	-87.6351	(41.8505033, -87.63518007)	
12	252251	300 W 23rd St	Residential	2008-04-01	60616	ILLINOIS	Aurora	DOUGLAS	35	41.82021	-87.61246	(41.8202138, -87.6124687)	
13	252251	300 W 23rd St	Residential	2008-04-01	60616	ILLINOIS	Aurora	HYDE PARK	41	41.787815	-87.98927	(41.7878147, -87.98927692)	
14	252251	300 W 23rd St	Residential	2008-04-01	60616	ILLINOIS	Aurora	LINCOLN PARK	7	41.922912	-87.6462	(41.922912051, -87.64621957)	
15	252251	300 W 23rd St	Residential	2008-04-01	60616	ILLINOIS	Aurora	ARMOUR SQUARE	34	41.85050	-87.6351	(41.8505033, -87.63518007)	
16	251331	2312 N Coflin Ave	Commercial	2008-04-01	60614	ILLINOIS	Aurora	LINCOLN PARK	7	41.92946	-87.6579	(41.9294692, -87.6579559)	
17	251331	2312 N Coflin Ave	Commercial	2008-04-01	60614	ILLINOIS	Aurora	UPTOWN	3	41.93385	-87.6984	(41.9338537, -87.6984196)	
18	251331	2312 N Coflin Ave	Commercial	2008-04-01	60614	ILLINOIS	Aurora	LOOP	32	41.871124	-87.62471	(41.8711248, -87.6247096)	
19	251331	2312 N Coflin Ave	Commercial	2008-04-01	60614	ILLINOIS	Aurora	NEAR WEST SIDE	28	41.881123	-87.69945	(41.88112401, -87.69945854)	
20	251331	2312 N Coflin Ave	Commercial	2008-04-01	60614	ILLINOIS	Aurora	LINCOLN PARK	7	41.92946	-87.6579	(41.9294692, -87.6579559)	
21	251254	720 W Gordon Ter	Residential	2008-04-01	60613	ILLINOIS	Aurora	UPTOWN	3	41.95817	-87.6487	(41.9581769, -87.64866912)	
22	251254	720 W Gordon Ter	Residential	2008-04-01	60613	ILLINOIS	Aurora	NEAR NORTH SIDE	8	41.89313	-87.62306	(41.89313049, -87.62306158)	
23	251254	720 W Gordon Ter	Residential	2008-04-01	60613	ILLINOIS	Aurora	AUSTIN	25	41.912785	-87.78313	(41.912785, -87.78313)	
24	251254	720 W Gordon Ter	Residential	2008-04-01	60613	ILLINOIS	Aurora	NEAR NORTH SIDE	8	41.898914	-87.62306	(41.898914, -87.62306158)	
25	251405	720 W Gordon Ter	Residential	2008-04-01	60613	ILLINOIS	Aurora	UPTOWN	3	41.95817	-87.6487	(41.9581769, -87.64866912)	
26	251405	4409 W Wilcox St	Municipal	2008-04-01	60624	OH	Pottery Addition	HIGH GLENFIELD PARK	26	41.878708	-87.3579	(41.878708, -87.3579)	
27	251405	4409 W Wilcox St	Municipal	2008-04-01	60624	OH	Pottery Addition	GRAND GLENFIELD PARK	38	41.81195	-87.65513	(41.8119513, -87.65512794)	
28	251405	4409 W Wilcox St	Municipal	2008-04-01	60624	OH	Pottery Addition	HYDE PARK	41	41.79888	-87.50293	(41.7988799, -87.50290228)	
29	251405	4409 W Wilcox St	Municipal	2008-04-01	60624	OH	Pottery Addition	LAKEVIEW	6	41.93984	-87.6594	(41.9398511, -87.65944752)	
30	251405	4409 W Wilcox St	Municipal	2008-04-01	60624	OH	Pottery Addition	WEST GLENFIELD PARK	26	41.878708	-87.3579	(41.878708, -87.3579)	
31	120518	3500 S GILES AVE	Commercial	2008-03-01	60653	MS	Marion	DOUGLAS	35	41.82021	-87.612465	(41.8202138, -87.6124687)	
32	120518	3500 S GILES AVE	Commercial	2008-03-01	60653	MS	Marion	NEAR WEST SIDE	28	41.880884	-87.68008	(41.8808807, -87.68008143)	
33	120518	3500 S GILES AVE	Commercial	2008-03-01	60653	MS	Marion	ARMOUR SQUARE	32	41.88183	-87.6267	(41.8818292, -87.6269992)	
34	120518	3500 S GILES AVE	Commercial	2008-03-01	60653	MS	Marion	LOOP	32	41.87393	-87.6383	(41.8739316, -87.6383678)	
35	120518	3500 S GILES AVE	Commercial	2008-03-01	60653	MS	Marion	DOUGLAS	35	41.88203	-87.614945	(41.8820304, -87.614945)	
36	252057	1126 E 59th St	Commercial	2008-04-01	60637	TX	Converse	HYDE PARK	41	41.787815	-87.598274	(41.7878147, -87.59827692)	
37	252057	1126 E 59th St	Commercial	2008-04-01	60637	TX	Converse	LAKEVIEW	6	41.935238	-87.39725	(41.9352387, -87.3972431)	
38	252057	1126 E 59th St	Commercial	2008-04-01	60637	TX	Converse	Converse	Loop	32	41.8833	-87.62033	(41.8833008, -87.62033043)
39	252057	1126 E 59th St	Commercial	2008-04-01	60637	TX	Converse	NEAR NORTH SIDE	8	41.90029	-87.62906	(41.9002893, -87.6290631)	
40	252057	1126 E 59th St	Commercial	2008-04-01	60637	TX	Converse	HYDE PARK	41	41.787815	-87.598274	(41.7878147, -87.59827692)	
41	173718	420 W WRIGHTWOOD AVE	Residential	2008-03-01	60614	ID	Wardner	LINCOLN PARK	7	41.922912	-87.64642	(41.922912, -87.64642)	
42	173718	420 W WRIGHTWOOD AVE	Residential	2008-03-01	60614	ID	Wardner	SOUTH DEERFIELD	51	41.72029	-87.5829	(41.720298, -87.5828921)	

View Number facet under Building_ID column,Check for non-numeric,You can see that there are no non-numeric:

OpenRefine dataset 246 csv [Permalink](#)

Facet / Filter Undo / Redo 0 / 0

2730 rows

Show: 5 10 25 50 100 500 1000 rows

Refresh Reset all Remove all Building_ID change reset No numeric value present.

All	Building_ID	Address	Building_Type	Build_Year	Zip	State	City	Community_Name	Community_Number	Latitude	Longitude	Location
1.	101717	55 E Monroe St	Commercial	2010-03-01	60603	IL	Chicago	LOOP	32	41.880177	-87.62564	(41.880177, -87.62564)
2.	101717	55 E Monroe St	Commercial	2010-03-01	60603	IL	Chicago	DOUGLAS	35	41.845505	-87.830554	(41.845505, -87.830554)
3.	101717	55 E Monroe St	Commercial	2010-03-01	60603	IL	Chicago	ARMOUR SQUARE	34	41.8509	-87.62564	(41.8509, -87.62564)
4.	101717	55 E Monroe St	Commercial	2010-03-01	60603	IL	Chicago	LINCOLN PARK	7	41.923945	-87.65159	(41.923945, -87.65159)
5.	101717	55 E Monroe St	Commercial	2008-03-01	60603	IL	Chicago	LOOP	32	41.880177	-87.62564	(41.880177, -87.62564)
6.	177286	619 645 E WOODLAND PARK	Residential	2008-03-01	60616	IL	Chicago	DOUGLAS	35	41.845505	-87.830554	(41.845505, -87.830554)
7.	177286	619 645 E WOODLAND PARK	Residential	2008-03-01	60616	IL	Chicago	LINCOLN PARK	7	41.923945	-87.65759	(41.923945, -87.65759)
8.	177286	619 645 E WOODLAND PARK	Residential	2008-03-01	60616	IL	Chicago	UPTOWN	3	41.85817	-87.64857	(41.85817, -87.64857)
9.	177286	619 645 E WOODLAND PARK	Residential	2008-03-01	60616	IL	Chicago	WEST GARFIELD PARK	26	41.878708	-87.73579	(41.878708, -87.73579)
10.	177286	619 645 E WOODLAND PARK	Residential	2008-03-01	60616	IL	Chicago	DOUGLAS	35	41.8509	-87.62564	(41.8509, -87.62564)
11.	252251	300 W 23rd St	Residential	2008-04-01	60616	IL	Chicago	ARMOUR SQUARE	34	41.8509	-87.630319	(41.8509, -87.630319)
12.	252251	300 W 23rd St	Residential	2008-04-01	60616	IL	Chicago	DOUGLAS	35	41.92021	-87.612465	(41.92021, -87.612465)
13.	252251	300 W 23rd St	Residential	2008-04-01	60616	IL	ILLINOIS	Aurora	41	41.787815	-87.598274	(41.787815, -87.598274)
14.	252251	300 W 23rd St	Residential	2008-04-01	60616	IL	ILLINOIS	UPTOWN	3	41.922012	-87.65462	(41.922012, -87.65462)
15.	252251	300 W 23rd St	Residential	2008-04-01	60616	IL	ILLINOIS	ARMOUR SQUARE	34	41.8509	-87.63518	(41.8509, -87.63518)
16.	251331	2312 N C offline Ave	Commercial	2008-04-01	60614	IL	ILLINOIS	AURORA	7	41.923945	-87.65759	(41.923945, -87.65759)
17.	251331	2312 N C offline Ave	Commercial	2008-04-01	60614	IL	ILLINOIS	DOUGLAS	35	41.87124	-87.62471	(41.87124, -87.62471)
18.	251331	2312 N C offline Ave	Commercial	2008-04-01	60614	IL	ILLINOIS	NEAR WEST SIDE	28	41.911123	-87.698485	(41.911123, -87.698485)
19.	251331	2312 N C offline Ave	Commercial	2008-04-01	60614	IL	ILLINOIS	UPTOWN	3	41.923945	-87.65759	(41.923945, -87.65759)
20.	251331	2312 N C offline Ave	Commercial	2008-04-01	60614	IL	ILLINOIS	ARMOUR SQUARE	34	41.85817	-87.64857	(41.85817, -87.64857)
21.	252154	720 W Gordon Ter	Residential	2008-04-01	60613	IL	ILLINOIS	AURORA	41	41.78959	-87.598274	(41.78959, -87.598274)
22.	252154	720 W Gordon Ter	Residential	2008-04-01	60613	IL	ILLINOIS	NEAR NORTH SIDE	8	41.89313	-87.62306	(41.89313, -87.62306)
23.	252154	720 W Gordon Ter	Residential	2008-04-01	60613	IL	ILLINOIS	AUSTIN	25	41.922012	-87.78313	(41.922012, -87.78313)
24.	252154	720 W Gordon Ter	Residential	2008-04-01	60613	IL	ILLINOIS	NEAR NORTH SIDE	8	41.88994	-87.626236	(41.88994, -87.626236)
25.	252154	720 W Gordon Ter	Residential	2008-04-01	60613	IL	ILLINOIS	NEAR NORTH SIDE	8	41.88994	-87.626236	(41.88994, -87.626236)
26.	252154	720 W Gordon Ter	Residential	2008-04-01	60613	IL	ILLINOIS	DOUGLAS	35	41.85817	-87.64857	(41.85817, -87.64857)
27.	252154	720 W Gordon Ter	Residential	2008-04-01	60613	IL	ILLINOIS	UPTOWN	3	41.923945	-87.65759	(41.923945, -87.65759)
28.	252154	720 W Gordon Ter	Residential	2008-04-01	60613	IL	ILLINOIS	WEST GARFIELD PARK	26	41.878708	-87.73579	(41.878708, -87.73579)
29.	252154	720 W Gordon Ter	Residential	2008-04-01	60613	IL	ILLINOIS	DOUGLAS	35	41.845505	-87.630554	(41.845505, -87.630554)
30.	252154	720 W Gordon Ter	Residential	2008-04-01	60613	IL	ILLINOIS	ARMOUR SQUARE	34	41.8509	-87.65759	(41.8509, -87.65759)
31.	252154	720 W Gordon Ter	Residential	2008-04-01	60613	IL	ILLINOIS	HYDE PARK	41	41.92021	-87.612465	(41.92021, -87.612465)
32.	252154	720 W Gordon Ter	Residential	2008-04-01	60613	IL	ILLINOIS	NEAR NORTH SIDE	8	41.922012	-87.598274	(41.922012, -87.598274)
33.	252154	720 W Gordon Ter	Residential	2008-04-01	60613	IL	ILLINOIS	AUSTIN	41	41.787815	-87.598274	(41.787815, -87.598274)
34.	252154	720 W Gordon Ter	Residential	2008-04-01	60613	IL	ILLINOIS	NEAR NORTH SIDE	8	41.89313	-87.62306	(41.89313, -87.62306)
35.	252154	720 W Gordon Ter	Residential	2008-04-01	60613	IL	ILLINOIS	DOUGLAS	35	41.85817	-87.64857	(41.85817, -87.64857)
36.	252154	720 W Gordon Ter	Residential	2008-04-01	60613	IL	ILLINOIS	UPTOWN	3	41.923945	-87.65759	(41.923945, -87.65759)
37.	252154	720 W Gordon Ter	Residential	2008-04-01	60613	IL	ILLINOIS	WEST GARFIELD PARK	26	41.878708	-87.73579	(41.878708, -87.73579)
38.	252154	720 W Gordon Ter	Residential	2008-04-01	60613	IL	ILLINOIS	DOUGLAS	35	41.85817	-87.64857	(41.85817, -87.64857)
39.	252154	720 W Gordon Ter	Residential	2008-04-01	60613	IL	ILLINOIS	ARMOUR SQUARE	34	41.8509	-87.65759	(41.8509, -87.65759)
40.	252154	720 W Gordon Ter	Residential	2008-04-01	60613	IL	ILLINOIS	HYDE PARK	41	41.787815	-87.598274	(41.787815, -87.598274)
41.	173718	420 W WRIGHTWOOD AVE	Residential	2008-03-01	60614	ID	Warder	LINCOLN PARK	7	41.922012	-87.598274	(41.922012, -87.598274)
42.	173718	420 W WRIGHTWOOD AVE	Residential	2008-03-01	60614	ID	Warder	SOUTH DEERING	51	41.72029	-87.5829	(41.72029, -87.5829)

Convert all values under the Build_ID column to numeric types: use Commons transforms To Number, You can see that all are converted to numeric types with no null or other values:

OpenRefine dataset 246 csv [Permalink](#)

Facet / Filter Undo / Redo 1 / 1

2730 rows

Show: 5 10 25 50 100 500 1000 rows

Refresh Reset all Remove all Building_ID change reset

Text transform on 2,730 cells in column Building_ID:
value.toNumber() Undo

All	Building_ID	Address	BUILDING_TYPE	Build_Year	Zip	State	City	Community_Name	Community_Number	Latitude	Longitude	Location
1.	101717	55 E Monroe St	Commercial	2010-03-01	60603	IL	Chicago	LOOP	32	41.880177	-87.62564	(41.880177, -87.62564)
2.	101717	55 E Monroe St	Commercial	2010-03-01	60603	IL	Chicago	DOUGLAS	35	41.845505	-87.830554	(41.845505, -87.830554)
3.	101717	55 E Monroe St	Commercial	2010-03-01	60603	IL	Chicago	ARMOUR SQUARE	34	41.8509	-87.62564	(41.8509, -87.62564)
4.	101717	55 E Monroe St	Commercial	2010-03-01	60603	IL	Chicago	LINCOLN PARK	7	41.923945	-87.65159	(41.923945, -87.65159)
5.	101717	55 E Monroe St	Commercial	2008-03-01	60603	IL	Chicago	LOOP	32	41.880177	-87.62564	(41.880177, -87.62564)
6.	177286	619 645 E WOODLAND PARK	Residential	2008-03-01	60616	IL	Chicago	DOUGLAS	35	41.845505	-87.830554	(41.845505, -87.830554)
7.	177286	619 645 E WOODLAND PARK	Residential	2008-03-01	60616	IL	Chicago	UPTOWN	3	41.85817	-87.64857	(41.85817, -87.64857)
8.	177286	619 645 E WOODLAND PARK	Residential	2008-03-01	60616	IL	Chicago	WEST GARFIELD PARK	26	41.878708	-87.73579	(41.878708, -87.73579)
9.	177286	619 645 E WOODLAND PARK	Residential	2008-03-01	60616	IL	Chicago	DOUGLAS	35	41.8509	-87.62564	(41.8509, -87.62564)
10.	251331	2312 N C offline Ave	Commercial	2008-04-01	60614	IL	ILLINOIS	AURORA	7	41.923945	-87.65759	(41.923945, -87.65759)
11.	251331	2312 N C offline Ave	Commercial	2008-04-01	60614	IL	ILLINOIS	DOUGLAS	35	41.87124	-87.62471	(41.87124, -87.62471)
12.	251331	2312 N C offline Ave	Commercial	2008-04-01	60614	IL	ILLINOIS	NEAR WEST SIDE	28	41.911123	-87.698485	(41.911123, -87.698485)
13.	251331	2312 N C offline Ave	Commercial	2008-04-01	60614	IL	ILLINOIS	UPTOWN	3	41.923945	-87.65759	(41.923945, -87.65759)
14.	251331	2312 N C offline Ave	Commercial	2008-04-01	60614	IL	ILLINOIS	ARMOUR SQUARE	34	41.85817	-87.64857	(41.85817, -87.64857)
15.	252251	300 W 23rd St	Residential	2008-04-01	60616	IL	Chicago	DOUGLAS	35	41.923945	-87.65759	(41.923945, -87.65759)
16.	252251	300 W 23rd St	Residential	2008-04-01	60616	IL	Chicago	ARMOUR SQUARE	34	41.85817	-87.64857	(41.85817, -87.64857)
17.	252251	300 W 23rd St	Residential	2008-04-01	60616	IL	Chicago	LINCOLN PARK	7	41.923945	-87.65759	(41.923945, -87.65759)
18.	252251	300 W 23rd St	Residential	2008-04-01	60616	IL	Chicago	DOUGLAS	35	41.85817	-87.64857	(41.85817, -87.64857)
19.	252251	300 W 23rd St	Residential	2008-04-01	60616	IL	Chicago	ARMOUR SQUARE	34	41.85817	-87.64857	(41.85817, -87.64857)
20.	252251	300 W 23rd St	Residential	2008-04-01	60616	IL	Chicago	HYDE PARK	41	41.787815	-87.598274	(41.787815, -87.598274)
21.	252154	720 W Gordon Ter	Residential	2008-04-01	60613	IL	ILLINOIS	UPTOWN	3	41.923945	-87.65759	(41.923945, -87.65759)
22.	252154	720 W Gordon Ter	Residential	2008-04-01	60613	IL	ILLINOIS	NEAR NORTH SIDE	8	41.89313	-87.62306	(41.89313, -87.62306)
23.	252154	720 W Gordon Ter	Residential	2008-04-01	60613	IL	ILLINOIS	AUSTIN	41	41.917815	-87.78313	(41.917815, -87.78313)
24.	252154	720 W Gordon Ter	Residential	2008-04-01	60613	IL	ILLINOIS	NEAR NORTH SIDE	8	41.88994	-87.626236	(41.88994, -87.626236)
25.	252154	720 W Gordon Ter	Residential	2008-04-01	60613	IL	ILLINOIS	UPTOWN	3	41.95817	-87.64857	(41.95817, -87.64857)
26.	251405	4409 W Wilcox St	Municipal	2008-04-01	60624	OH	Pottery Addition	GRAND BOULEVARD	38	41.81195	-87.65513	(41.81195, -87.65513)
27.	251405	4409 W Wilcox St	Municipal	2008-04-01	60624	OH	Pottery Addition	HYDE PARK	41	41.798708	-87.65024	(41.798708, -87.65024)
28.	251405	4409 W Wilcox St	Municipal	2008-04-01	60624	OH	Pottery Addition	LAKE VIEW	6	41.8581		

OpenRefine dataset 246 csv [Permalink](#)

Facet / Filter [Undo / Redo 1 / 1](#) [Reorder rows](#) [Undo](#)

Show as: rows records Show: 5 10 25 50 100 500 1000 rows

Building_ID Address Building_Type

Sort by Building_ID

Community_Number Latitude Longitude Location

Drag and drop to re-order

OK Cancel

Sort cell values as Position blanks and errors

Valid values Errors Blanks

① smallest first ② largest first

100,000 — 270,000

Data after permanent sorting:

OpenRefine dataset 246 csv [Permalink](#)

Facet / Filter [Undo / Redo 2 / 2](#) [Reorder rows](#) [Undo](#)

Show as: rows records Show: 5 10 25 50 100 500 1000 rows

Building_ID Address Building_Type Build_Year Zip State City Community_Name Community_Number Latitude Longitude Location

Using facets and filters

Use facets and filters to select subsets of your data to act on. Choose facet and filter from the dropdown menus at the top of each data column.

Not sure how to get started? Watch these screencasts

Build_ID column values are checked for duplicates adjacent to each other using Blank down:

127.0.0.1:3333/project?project=269886161935

fine dataset 246 csv Permalink Undo / Redo 3 / 3 2730 rows Show as: rows records Show: 5 10 25 50 100 500 1000 rows

ts and filters

id filters to select subsets
o act on. Choose facet
hds from the menus at
21 data column.

I'm get started?
screenshots

All Building_ID Address Building_Type Build_Year Zip State City Community_Name Community_Number Latitude Longitude Location

1.	100008	555 W Roosevelt Rd	Commercial	2017-03-01	60607	IN	Goodland	NEAR WEST SIDE	32	41.8666867	-87.64145	(41.8666867, -87.64144957)	
2.	100026	4975 S Archer Ave	Municipal	1996-03-01	60632	NE	Cambridge	ARCHER HEIGHTS	57	41.75392	-87.601974	(41.802759, -87.722671)	
3.	100030	4039-4051 N LAPORTE AVE	Residential	1936-03-01	60641	MN	Russell	PORTAGE PARK	15	41.94615	-87.74708	(41.94615, -87.74707694)	
4.	100048	2609 W BELMONT AVE	Residential	1954-03-01	60618	VA	Leesburg	AVONDALE	21	41.94627	-87.70319	(41.9467055, -87.70319673)	
5.	100072	2701 W 68th St	Commercial	1967-03-01	60629	IL	Rockbridge	CHICAGO LAWN	66	41.770004	-87.89084	(41.770004, -87.890841)	
6.	100081	7701 S Cicero Ave	Commercial	1967-03-01	60652	IA	Mingo	WEST LAWN	65	41.920395	-87.756165	(41.75882921, -87.73962958)	
7.	100086	3939 W 79th St	Municipal	1966-03-01	60652	NC	Northchase	DOUGLAS	35	41.74946	-87.72046	(41.74946, -87.720464)	
8.	100091	875 N Michigan Ave	Commercial	1967-03-01	60611	IL	Rockbridge	NEAR NORTH SIDE	8	41.898804	-87.62298	(41.898804, -87.62297591)	
9.	100433	2233 W Division St	Commercial	2017-03-01	60622	IN	Goodland	WEST TOWN	5	41.902515	-87.68336	(41.90251573, -87.6833696)	
10.	100447	211 E Chicago Ave	Commercial	1936-03-01	60611	MI	Russell	NEAR NORTH SIDE	8	41.89645	-87.62527	(41.89645, -87.62525575)	
11.	100448	198 E Delaware Pl	Residential	1914-03-01	60611	OH	Northgate	NEAR NORTH SIDE	8	41.883194	-87.82865	(41.899532, -87.62215183)	
12.	100476	30 W Chicago	Residential	1900-03-01	60654	VA	Ebony	NEAR NORTH SIDE	8	41.87781	-87.65405	(41.896703, -87.639373)	
13.	100498	4620 4528 S GREEKL BLVD	Residential	2017-03-01	60653	IN	Goodland	KENWOOD	3	41.82021	-87.612465	(41.82021138, -87.6124687)	
14.	100507	4950 4952 S PRAIRIE AVE	Residential	1936-03-01	60615	MN	Russell	GRAND BOULEVARD	38	41.802242	-87.80333	(41.80224043, -87.6032678)	
15.	100531	600 S Stevens Ave	Municipal	1964-03-01	60621	VA	Leesburg	ENGLEWOOD	68	41.78205	-87.83427	(41.78205, -87.83427631)	
16.	100571	811 S Dearborn St	Commercial	1967-03-01	60620	IL	Rockbridge	BROOKLYNSHAM	71	41.74503	-87.65152	(41.74503, -87.6515079)	
17.	100623	1345-1350 W ESTES AVE	Residential	1967-03-01	60628	IL	Whitmore Village	ROGERS PARK	1	42.01036	-87.65028	(42.01036, -87.6502165)	
18.	100624	1345-1350 W ESTES AVE	Residential	1967-03-01	60628	IL	Whitmore Village	NEAR SOUTH SIDE	33	41.84157	-87.629265	(41.84157, -87.6291845)	
19.	100625	1345-1350 W ESTES AVE	Residential	1967-03-01	60628	IL	Whitmore Village	WASHINGTON HEIGHTS	73	41.72019	-87.65915	(41.72019, -87.6591245)	
20.	100626	1345-1350 W ESTES AVE	Residential	1967-03-01	60628	IL	Whitmore Village	UPTOWN	3	41.97983	-87.65495	(41.97983, -87.6549564)	
21.	100627	1345-1350 W ESTES AVE	Residential	1967-03-01	60628	IL	Whitmore Village	ROGERS PARK	1	42.01036	-87.6583	(42.01036, -87.65839165)	
22.	100628	1345-1350 W ESTES AVE	Residential	1967-03-01	60628	IL	Whitmore Village	LAKEVIEW	6	41.89217	-87.54913	(41.89217, -87.54912018)	
23.	100687	510 EMW & ADDISON ST	Residential	1960-03-01	60613	VA	Ebony	EAST GARFIELD PARK	27	41.89029	-87.51815	(41.89029, -87.518153)	
24.	100691	3245 W ARTHURTON ST	Commercial	1914-03-01	60624	OH	Northgate	NORTH LAWNDALE	29	41.870007	-87.70832	(41.8700049, -87.70831909)	
25.	100674	3301 W Arthington St	Residential	2012-03-01	60624	IL	KY	Moordland	WEST TOWN	24	41.89894	-87.65764	(41.89894, -87.657614)
26.	100675	3301 W Arthington St	Residential	2012-03-01	60624	IL	KY	Moordland	ROSELAND	49	41.696182	-87.64217	(41.696182, -87.64217052)
27.	100676	3301 W Arthington St	Residential	2012-03-01	60624	IL	KY	Moordland	HUMBOLDT PARK	23	41.910187	-87.726316	(41.910187, -87.72631778)
28.	100677	3301 W Arthington St	Residential	2012-03-01	60624	IL	KY	Moordland	NORTH LAWNDALE	29	41.870007	-87.70832	(41.8700049, -87.70831909)
29.	100678	3301 W Arthington St	Residential	2012-03-01	60624	IL	KY	Moordland	NEAR SOUTH SIDE	33	41.77613	-87.40406	(41.77613, -87.404094)
30.	100683	7000 S PARINELL AVE	Residential	1970-03-01	60621	OK	Snake Creek	NEAR WEST SIDE	33	41.89029	-87.65483	(41.89029, -87.6548121)	
31.	100729	801 S Canal St	Commercial	1900-03-01	60607	VA	Ebony	NEAR WEST SIDE	28	42.01036	-87.6883	(42.01036, -87.68839278)	
32.	100733	840 S CANAL ST	Commercial	1996-03-01	60607	NE	Cambridge	NEAR WEST SIDE	28	41.89029	-87.6883	(41.89029, -87.688374)	
33.	100861	600 W Chicago	Commercial	1925-03-01	60654	IL	Big Foot Prairie	NEAR NORTH SIDE	8	41.8965	-87.64286	(41.8965, -87.64282)	
34.	100867	1653 W Congress Pkwy	Commercial	1996-03-01	60612	IL	Big Foot Prairie	NEAR WEST SIDE	28	41.845505	-87.30554	(41.875121, -87.680005)	
35.	100874	1521 1523 W SHERWIN AVE	Residential	2010-03-01	60626	MN	East Gut Lake	ROGERS PARK	1	42.01036	-87.6883	(42.01036, -87.68829795)	
36.	100946	6425 S Lowe	Residential	1907-03-01	60621	IA	Mingo	ENGLEWOOD	68	41.780464	-87.60531	(41.772263, -87.60533938)	
37.	101013	7320 7332 N DAMEN AVE	Residential	1969-03-01	60645	CO	Crisman	ROGERS PARK	30	42.00884	-87.69503	(42.00884, -87.69502812)	
38.	101038	3750 S COTTAGE GROVE AVE	Residential	1957-03-01	60653	HI	Whitmore Village	OAKLAND	36	41.82021	-87.612465	(41.82021138, -87.6124687)	
39.	101039	3750 S COTTAGE GROVE AVE	Residential	1957-03-01	60653	HI	Whitmore Village	WEST ENGLEWOOD	67	41.872066	-87.72212	(41.87206597, -87.7220934)	
40.	101040	3750 S COTTAGE GROVE AVE	Residential	1957-03-01	60653	HI	Whitmore Village	GAGE PARK	63	41.793297	-87.68838	(41.79329756, -87.68837684)	
41.	101041	3750 S COTTAGE GROVE AVE	Residential	1957-03-01	60653	HI	Whitmore Village	GREATER GRAND CROSSING	69	41.779964	-87.51674	(41.7799687, -87.51673927)	
42.	101042	3750 S COTTAGE GROVE AVE	Residential	1957-03-01	60653	HI	Whitmore Village	OAKLAND	36	41.82021	-87.612465	(41.82021138, -87.6124687)	
43.	101023	5110 S DR MARTIN LUTHER KING JR DR	Residential	1954-03-01	60615	VA	Leesburg	WASHINGTON PARK	40	41.802242	-87.60333	(41.80224043, -87.6032678)	

To detect if a Building_ID column value has a null character, use Facet by blank:

Facet / Filter Undo / Redo 4 / 4 933 matching rows (2730 total) Extensions Wikibase ▾

Reset all Remove all Show as: rows records Show: 5 10 25 50 100 500 1000 rows

Building_ID change invert reset 2 choices Sort by: name count false 175 true 933 exclude

Facet by choice counts

All Building_ID Address Building_Type Build_Year Zip State City Community_Name Community_Number Latitude Longitude Location

19.	1345-1350 W ESTES AVE	Residential	1957-03-01	60628	IL	Whitmore Village	NEAR SOUTH SIDE	33	41.84157	-87.627015	(41.84157, -87.62701562)	
20.	1345-1350 W ESTES AVE	Residential	1957-03-01	60628	IL	Whitmore Village	WASHINGTON HEIGHTS	73	41.72019	-87.65675	(41.72019, -87.65675165)	
21.	1345-1350 W ESTES AVE	Residential	1957-03-01	60628	IL	Whitmore Village	UPTOWN	3	41.97983	-87.65465	(41.97983, -87.6546594)	
22.	1345-1350 W ESTES AVE	Residential	1957-03-01	60628	IL	Whitmore Village	ROGERS PARK	1	42.01036	-87.6683	(42.01036, -87.66839295)	
23.	3301 W Arthington St	Residential	2012-03-01	60624	IL	KY	Moordland	WEST TOWN	24	41.89883	-87.65767	(41.89883, -87.6576414)
24.	3301 W Arthington St	Residential	2012-03-01	60624	IL	KY	Moordland	ROSELAND	49	41.696182	-87.64217	(41.696182, -87.64217052)
25.	3301 W Arthington St	Residential	2012-03-01	60624	IL	KY	Moordland	NORTH LAWNDALE	29	41.870007	-87.70832	(41.8700049, -87.70831909)
26.	3301 W Arthington St	Residential	2012-03-01	60624	IL	KY	Moordland	NEAR SOUTH SIDE	67	41.793297	-87.68838	(41.79329756, -87.68837894)
27.	3301 W Arthington St	Residential	2012-03-01	60624	IL	KY	Moordland	GAGE PARK	83	41.779964	-87.68838	(41.7799687, -87.68837894)
28.	3301 W Arthington St	Residential	2012-03-01	60624	IL	KY	Moordland	GREATER GRAND CROSSING	89	41.779964	-87.61974	(41.7799687, -87.61973927)
29.	55 E Monroe St	Commercial	2013-03-01	60603	IL	Chicago	Douglas	35	41.845505	-87.63054	(41.845505, -87.63055693)	
30.	55 E Monroe St	Commercial	2013-03-01	60603	IL	Chicago	ARMOUR SQUARE	34	41.85996	-87.63518	(41.859963, -87.63518087)	
31.	55 E Monroe St	Commercial	2013-03-01	60603	IL	Chicago	LINCOLN PARK	7	41.923946	-87.65759	(41.923946, -87.65759559)	
32.	55 E Monroe St	Commercial	2013-03-01	60603	IL	Chicago	LOOP	32	41.880177	-87.62564	(41.8801753, -87.62564433)	
33.	500 W Monroe	Commercial	1957-03-01	60661	IL	Moordland	LAKEVIEW	6	41.89842	-87.644745	(41.89842, -87.64474572)	
34.	500 W Monroe	Commercial	1957-03-01	60661	IL	Moordland	NEAR NORTH SIDE	33	41.857285	-87.61016	(41.857285, -87.6101697)	
35.	500 W Monroe	Commercial	1957-03-01	60661	IL	Moordland	NEAR WEST SIDE	28	41.88008	-87.63079	(41.88008, -87.630778)	
36.	101 N State St E	Commercial	1957-03-01	60602	IL	Whitmore Village	NEAR WEST SIDE	28	41.879112	-87.646515	(41.87911201, -87.64651787)	
37.	101 N State St E	Commercial	1957-03-01	60602	IL	Whitmore Village	Hyde Park	41	41.80094	-87.58985	(41.8009428, -87.58984522)	
38.	101 N State St E	Commercial	1957-03-01	60602	IL	Whitmore Village	KENWOOD	39	41.80917	-87.59393	(41.80917058, -87.59395443)	
39.	101 N State St E	Commercial	1957-03-01	60602	IL	Whitmore Village	LOOP	32	41.88343	-87.62875	(41.8834304, -87.6287545)	
40.	35 W Wacker Dr	Commercial	2012-03-01	60601	IL	KY	Moordland	NEAR NORTH SIDE	8	41.895382	-87.62414	(41.895382, -87.62412413)
41.	35 W Wacker Dr	Commercial	2012-03-01	60601	IL	KY	Moordland	LINCOLN PARK	7	41.925022	-87.645935	(41.92502369, -87.64593626)
42.	35 W Wacker Dr	Commercial	2012-03-01	60601	IL	KY	Moordland	NEAR NORTH SIDE	8	41.89842	-87.62063	(41.89842141, -87.62063001)
43.	225 W Randolph	Commercial	2012-03-01	60606	IL	KY	Moordland	NEAR NORTH SIDE	28	41.88425	-87.63245	(41.88425, -87.632454)
44.	225 W Randolph	Commercial	2012-03-01	60606	IL							

OpenRefine dataset 246 csv Permalink

Remove 933 rows Undo

Facet / Filter Undo / Redo 5 / 5 Extensions Wikibase

0 matching rows (1797 total)

Show as: rows records Show: 5 10 25 50 100 500 1000 rows

All Building_ID Address Building_Type Build_Year Zip State City Community_Name Community_Number Latitude Longitude Location

Building_ID Address Building_Type Build_Year Zip State City Community_Name Community_Number Latitude Longitude Location

2 choices Sort by: name count

false 1797
true 0 exclude

Facet by choice counts

Checking the e0 constraints again, you can see that they are all satisfied:

Start detecting the e1 constraint: Zip → State

First look at the State column value distribution

In order to standardise the State nomenclature, change the non-uniform nomenclature ILLINOIS to IL.

View the distribution of State column values after modification:

Filter / Filter	Undo / Redo	6 / 6	1979 rows														Extensions	Wikibase			
Refresh	Reset all	Remove all	Show as:	rows	records	Show	5	10	25	50	100	500	1000	rows	< first	< previous	1	next	> last		
State	change																				
24 choices	Sort by: name	count	Cluster																		
IL	200			1	100098 555 W Willow Rd	Commercial	2017-03-01	00007	00010	00010	00010	00010	00010	00010	NEAR WEST SIDE	32	41.988687	87.641445	(41.988687, 87.641445)		
VA	200			2	100256 407 S Archer Ave	Municipal	1998-03-01	00012	00012	00012	00012	00012	00012	00012	CAMBRIDGE	57	41.73032	87.601074	(41.73032, 87.601074)		
MN	120			3	100320 4039 4051 N LAPORTE AVE	Residential	1938-03-01	00041	00051	00051	00051	00051	00051	00051	PORTAGE PARK	15	41.94915	87.74708	(41.9491536, 87.74707954)		
FL	100			4	100348 2609 W BELMONT AVE	Residential	1954-03-01	00058	00018	00018	00018	00018	00018	00018	LEEDSBURG	21	41.94872	87.70319	(41.9487295, 87.7019673)		
OH	100			5	100372 2701 W 88th St	Commercial	1957-03-01	00059	00110	00110	00110	00110	00110	00110	ROCKIDGE	66	41.77004	87.69984	(41.77004, 87.69984)		
CO	100			6	100371 7601 S Cicero Ave	Commercial	1907-03-01	00052	00152	00152	00152	00152	00152	00152	WEST LAWN	65	41.920395	87.759165	(41.920395, 87.759165)		
IA	100			7	100373 7740 S Westmore Ave	Commercial	1957-03-01	00053	00153	00153	00153	00153	00153	00153	ASHBURN	70	41.75293	87.68297	(41.75293, 87.6829748)		
IN	100			8	100397 3639 W 79th St	Municipal	1965-03-01	00055	00152	00152	00152	00152	00152	00152	NORTHCHASE	35	41.74946	87.720146	(41.74946, 87.720146)		
KY	100			9	100405 875 N University Ave	Commercial	1957-03-01	00056	00156	00156	00156	00156	00156	00156	ROCKIDGE	8	41.988687	87.62239	(41.988687, 87.62239)		
NC	100			10	100432 1001 W Division St	Commercial	1957-03-01	00062	00162	00162	00162	00162	00162	00162	WEST TOWN	6	41.80219	87.63349	(41.80219, 87.63349)		
NE	100			11	100447 211 E Chicago Ave	Commercial	1903-03-01	00061	00111	00111	00111	00111	00111	00111	RUSSELL	NEAR NORTH SIDE	8	41.99845	87.62557	(41.9984496, 87.6255765)	
OK	100			12	100448 198 E Delaware Pl	Residential	1914-03-01	00061	00111	00111	00111	00111	00111	00111	NORTHgate	NEAR NORTH SIDE	8	41.83194	87.63688	(41.83194, 87.621165)	
PA	100			13	100470 30 W Chicago	Residential	1900-03-01	00054	00154	00154	00154	00154	00154	00154	Ebony	NEAR NORTH SIDE	8	41.87781	87.65493	(41.87781, 87.65493)	
AZ	100			14	100489 4520 S DREXEL BLVD	Residential	2017-03-01	00053	00153	00153	00153	00153	00153	00153	KENWOOD	3	41.92021	87.612465	(41.92021, 87.612465)		
TX	100			15	100507 4950 4952 S PRAIRIE AVE	Residential	1938-03-01	00015	00155	00155	00155	00155	00155	00155	GRAND BOULEVARD	38	41.802242	87.603325	(41.802242, 87.603328)		
ND	20			16	100533 6201 S Stewart Ave	Municipal	1954-03-01	00021	00121	00121	00121	00121	00121	00121	ENGLEWOOD	68	41.720555	87.63427	(41.720555, 87.6342731)		
NY	20			17	100572 813 N May St	Commercial	1957-03-01	00020	00120	00120	00120	00120	00120	00120	ROCKIDGE	71	41.74538	87.61552	(41.7453815, 87.6152078)		
PA	20			18	100646 1345 1359 N ESTES AVE	Residential	1957-03-01	00050	00150	00150	00150	00150	00150	00150	Whitney	ROGERS PARK	1	41.201038	87.66668	(41.201038, 87.66929165)	
MS	1			19	100657 510 534 W ADDISON ST	Residential	1900-03-01	00013	00113	00113	00113	00113	00113	00113	VIA	LAKE VIEW	6	41.89217	87.63413	(41.89217, 87.63413)	
TX	1			20	100674 3245 W ARTHINGTON ST	Commercial	1914-03-01	00024	00124	00124	00124	00124	00124	00124	NORTHEAST	GARFIELD PARK	27	41.89502	87.61816	(41.870493, 87.704382)	
WI	1			21	100674 3301 N KELLOGG ST	Residential	1917-03-01	00024	00124	00124	00124	00124	00124	00124	MOORLAND	29	41.870007	87.70523	(41.870004, 87.70523)		
Facet by choice counts				22	100679 3000 S 3rd AVE	Residential	1978-03-01	00024	00124	00124	00124	00124	00124	00124	ROSEDALE	Creek	NEAR SOUTHWEST	28	41.799501	87.61700	(41.799501, 87.61700)
				23	100709 801 S Canal St	Commercial	1900-03-01	00007	00017	00017	00017	00017	00017	00017	EBONI	NEAR WEST SIDE	28	41.01038	87.68683	(41.01038, 87.68683)	
				24	100733 840 S CANA ST	Commercial	1998-03-01	00007	00017	00017	00017	00017	00017	00017	NEAR WEST SIDE	28	41.99029	87.61619	(41.970377, 87.6412731)		
				25	100766 600 W Chicago	Commercial	1925-03-01	00054	00154	00154	00154	00154	00154	00154	Hill	Big Front Porch	NEAR NORTH SIDE	8	41.9965	87.64288	(41.996501, 87.64288)
				26	100987 1653 W Congress Pkwy	Commercial	1998-03-01	00012	00012	00012	00012	00012	00012	00012	CAMBRIDGE	NEAR WEST SIDE	28	41.945605	87.630554	(41.975121, 87.688005)	

Because State update and Zip, City columns are associated, taking into account the cost of updating, in order to State find duplicate matches, we use distance measure, select threshold $\beta = 5$, using Python distance measure state column distribution and the distribution of the above columns is the same, proving that our threshold β is the correct choice, will be merged to match the merged data to be re-saved:

(1795, 1796)]	
State	
IL	205
VA	200
MN	120
FL	109
IN	100
NE	100
CO	100
OK	100
KY	100
HI	100
OH	100
NC	100
IA	99
TN	99
MT	49
MD	46
PA	20
ND	20
NY	20

Name: count, dtype: int64

Import the re-saved data into openRefine and then revisit the distribution of State values corresponding to each Zip:

Facet / Filter Undo / Redo / 0 / 0

Reset all Remove all

Zip change invert reset
59 choices Sort by: name count Cluster

60611.0 119 exclude

60614.0 100

60618.0 98

60657.0 92

60697.0 73

60654.0 69

60650.0 68

60616.0 68

60615.0 59

60613.0 54

60612.0 54

60632.0 40

All 119 matching rows (1787 total)

Show as: rows records Show 5 10 25 50 100 500 1000 rows

#	Building_ID	Address	Building_Type	Build_Year	Zip	State	City	Community_Name	Community_Number	Latitude	Longitude	Location	
9	100425	275 N Michigan Ave	Commercial	1987-03-01	60611.0	IL	Rockridge	NEAR NORTH SIDE	8.0	41.898284	-87.925203	(41.898284,-87.925203)	
11	100477	211 E Chicago Ave	Commercial	1936-03-01	60611.0	IL	Russell	NEAR NORTH SIDE	8.0	41.898445	-87.82237	(41.898445,-87.82237)	
12	100445	198 E Delaware Pl	Residential	1914-03-01	60611.0	IL	Northgate	NEAR NORTH SIDE	8.0	41.893194	-87.829085	(41.893194,-87.829085)	
88	101755	541 N Farbush	Commercial	2014-03-01	60611.0	IL	TN	Sale Creek	NEAR NORTH SIDE	8.0	41.759883	-87.787238	(41.759883,-87.787238)
90	101804	401 N Michigan Ave	Commercial	1990-03-01	60611.0	IL	Vineyards	NEAR NORTH SIDE	8.0	41.889814	-87.822836	(41.889814,-87.822836)	
91	101816	43 E Ohio St	Commercial	1900-03-01	60611.0	IL	VA	Ebony	NEAR NORTH SIDE	8.0	41.709604	-87.847064	(41.709604,-87.847064)
90	101840	710 N Lakeshore Dr	Commercial	1985-03-01	60611.0	IL	Mt	Reed Point	NEAR NORTH SIDE	8.0	41.89534	-87.816946	(41.89534,-87.816946)
97	101852	301 N Water St	Residential	2014-03-01	60611.0	IL	TN	Sale Creek	NEAR NORTH SIDE	8.0	41.97223	-87.8646	(41.97223,-87.8646)
103	101874	444 N Michigan Ave	Commercial	2012-03-01	60611.0	IL	KY	Moordland	NEAR NORTH SIDE	8.0	41.890694	-87.82478	(41.890694,-87.82478)
106	101883	676 N Michigan Ave	Residential	1914-03-01	60611.0	IL	OH	Northgate	NEAR NORTH SIDE	8.0	41.77034	-87.83724	(41.77034,-87.83724)
109	101890	345 E Superior St	Commercial	1996-03-01	60611.0	IL	NC	Northgate	ASHBURN	7.0	41.89545	-87.81832	(41.89545,-87.81832)
110	101920	410 N MICHIGAN AVE	Commercial	1957-03-01	60611.0	IL	HI	Whitmore Village	NEAR NORTH SIDE	8.0	41.890064	-87.82515	(41.890064,-87.82515)
120	101930	500 N Michigan Ave	Commercial	2017-03-01	60611.0	IL	IN	Woodland	NEAR NORTH SIDE	8.0	41.89117	-87.824695	(41.89117,-87.824695)
127	101971	25 E Peoria St	Commercial	2013-03-01	60611.0	IL	PA	Sale Creek	NEAR NORTH SIDE	8.0	41.940508	-87.819215	(41.940508,-87.819215)
128	101972	220 N LAKE SHORE DR	Residential	1987-03-01	60611.0	IL	IN	Wingspan Village	NEAR NORTH SIDE	8.0	41.89715	-87.82268	(41.89715,-87.82268)
171	103115	230 E CANTARIO ST	Residential	1907-03-01	60611.0	IL	PA	Mannette	NEAR NORTH SIDE	8.0	41.891704	-87.82248	(41.891704,-87.82248)
172	103117	379 E CEDAR ST	Residential	2014-03-01	60611.0	IL	TN	Sale Creek	NEAR NORTH SIDE	8.0	41.780464	-87.80303	(41.780464,-87.80303)
176	103594	455 N Clybourn Plaza Dr	Commercial	1921-03-01	60611.0	IL	FL	Campbellton	NEAR NORTH SIDE	8.0	41.890007	-87.82158	(41.890007,-87.82158)
196	103632	251 E Huron St	Commercial	1914-03-01	60611.0	IL	OH	Northgate	NEAR NORTH SIDE	8.0	41.876954	-87.84968	(41.876954,-87.84968)
194	103661	880 N Michigan Ave	Commercial	1917-03-01	60611.0	IL	MD	Manchester	NEAR NORTH SIDE	8.0	41.93384	-87.81421	(41.93384,-87.81421)
205	103748	710 N Garbars Ct	Commercial	2012-03-01	60611.0	IL	KY	Moordland	NEAR NORTH SIDE	8.0	41.89531	-87.82038	(41.89531,-87.82038)
210	103762	136 148 E Huron St	Residential	1900-03-01	60611.0	IL	VA	Ebony	NEAR NORTH SIDE	8.0	42.009603	-87.85729	(42.009603,-87.85729)
211	103768	844 W Rush St	Commercial	1914-03-01	60611.0	IL	OH	Northgate	NEAR NORTH SIDE	8.0	41.795812	-87.85892	(41.795812,-87.85892)
213	103816	330 N Wabash Av	Commercial	1970-03-01	60611.0	IL	KY	Snake Creek	HEGEWISCH	55.0	41.888638	-87.82752	(41.888638,-87.82752)
255	105709	103 109 E DELAWARE DR	Residential	1985-03-01	60611.0	IL	MD	Reed Point	NEAR NORTH SIDE	8.0	41.895794	-87.82248	(41.895794,-87.82248)
267	105749	211 217 E CHESTNUT ST	Residential	1966-03-01	60611.0	IL	NC	Northgate	LOOP	32.0	41.895794	-87.82424	(41.895794,-87.82424)
272	105864	600 E Grand Ave	Commercial	1954-03-01	60611.0	IL	VA	Leesburg	NEAR NORTH SIDE	8.0	41.891743	-87.804706	(41.891743,-87.804706)
277	106014	211 213 E DELAWARE PL	Residential	2012-03-01	60611.0	IL	CO	Crisman	NEAR NORTH SIDE	8.0	41.895794	-87.82428	(41.895794,-87.82428)
280	106015	201 N Michigan Ave	Residential	2014-03-01	60611.0	IL	MD	Leesburg	NEAR NORTH SIDE	8.0	41.895794	-87.82428	(41.895794,-87.82428)
304	110368	600 N Lake Shore Dr	Residential	1995-03-01	60611.0	IL	NC	Northgate	LOOP	32.0	41.892301	-87.81458	(41.892301,-87.81458)
335	113632	1130 N LAKE SHORE DR	Residential	1938-03-01	60611.0	IL	MD	Russell	NEAR NORTH SIDE	8.0	41.89778	-87.82247	(41.89778,-87.82247)
336	113668	229 235 E LAKE SHORE DR	Residential	1957-03-01	60611.0	IL	HI	Whitehouse Village	NEAR NORTH SIDE	8.0	41.895704	-87.82248	(41.895704,-87.82248)
337	113677	221 E WALTON PL	Residential	1964-03-01	60611.0	IL	VA	Leesburg	NEAR NORTH SIDE	8.0	41.895794	-87.82248	(41.895794,-87.82248)
338	113690	20 E CEDAR ST	Residential	1996-03-01	60611.0	IL	NE	Cambridge	NEAR NORTH SIDE	8.0	41.895805	-87.82355	(41.895805,-87.82355)
339	113700	233 E ERIE ST	Residential	2017-03-01	60611.0	IL	MD	Leesburg	NEAR NORTH SIDE	31.0	41.895794	-87.82248	(41.895794,-87.82248)
340	113726	227 239 E DELAWARE PL	Residential	2012-03-01	60611.0	IL	KY	Moordland	NEAR NORTH SIDE	8.0	41.895794	-87.82248	(41.895794,-87.82248)
343	114191	228 240 E WALTON PL	Residential	1998-03-01	60611.0	IL	NC	Russell	NEAR NORTH SIDE	8.0	41.895794	-87.82248	(41.895794,-87.82248)
437	124236	830 N Michigan Ave	Commercial	1996-03-01	60611.0	IL	NE	Cambridge	NEAR NORTH SIDE	8.0	41.802303	-87.80205	(41.802303,-87.80205)
447	125859	730 N Michigan Ave	Commercial	1936-03-01	60611.0	IL	MD	Russell	NEAR NORTH SIDE	8.0	41.89603	-87.824096	(41.89603,-87.824096)
450	127047	669 N Michigan Ave	Commercial	2012-03-01	60611.0	IL	KY	Moordland	NEAR NORTH SIDE	8.0	41.894463	-87.82327	(41.894463,-87.82327)

Facet / Filter Undo / Redo / 0 / 0

Reset all Remove all

Zip change invert reset
44 matching rows (1787 total)

Show as: rows records Show 5 10 25 50 100 500 1000 rows

#	Building_ID	Address	Building_Type	Build_Year	Zip	State	City	Community_Name	Community_Number	Latitude	Longitude	Location	
62	101707	111 E Wacker Dr	Commercial	1970-03-01	60601.0	OK	EBony	LOOP	32.0	41.88759	-87.82356	(41.88759,-87.82356)	
65	101713	200 E Randolph St	Commercial	1969-03-01	60601.0	CO	Crisman	LOOP	32.0	41.885292	-87.82152	(41.885292,-87.82152)	
76	101719	300 E RANDOLPH ST	Commercial	1970-03-01	60601.0	IL	FL	Campbellton	LOOP	32.0	41.884846	-87.81968	(41.884846,-87.81968)
100	101750	36 W Wacker Dr	Commercial	2017-03-01	60601.0	KY	MD	Reed Point	LOOP	32.0	41.890261	-87.804541	(41.890261,-87.804541)
86	101776	151 E Wacker Dr	Residential	1907-03-01	60601.0	IA	Mingo	LOOP	32.0	41.890085	-87.810465	(41.890085,-87.810465)	
108	101883	333 N Michigan Ave	Commercial	2012-03-01	60601.0	KY	Moordland	LOOP	32.0	41.878394	-87.824235	(41.878394,-87.824235)	
113	101901	151 171 W Randolph St	Residential	1996-03-01	60601.0	NC	Northgate	WEST TOWN	24.0	41.88411	-87.83305	(41.88411,-87.83305)	
115	101919	160 N LaSalle St	Commercial	1996-03-01	60601.0	NE	Cambridge	LOOP	32.0	41.947426	-87.82428	(41.947426,-87.82428)	
129	101980	30 E Lake St	Municipal	1954-03-01	60601.0	VA	Leesburg	LOOP	32.0	41.889066	-87.826854	(41.889066,-87.826854)	
179	103594	222 N LaSalle St	Commercial	1970-03-01	60601.0	IL	SD	Snake Creek	Lower West Side	31.0	41.88571	-87.83252	(41.88571,-87.83252)
187	103630	221 N LaSalle St	Commercial	1914-03-01	60601.0	OH	Northgate	LOOP	32.0	41.96442	-87.80000	(41.96442,-87.80000)	
188	103643	201 N LaSalle St	Commercial	1970-03-01	60601.0	MD	Manchester	LOOP	32.0	41.89534	-87.82304	(41.89534,-87.82304)	
190	103645	201 N LaSalle St	Commercial	1912-03-01	60601.0	IL	FL	Floridian	LOOP	32.0	41.890063	-87.81919	(41.890063,-87.81919)
191	103674	75 N Michigan Ave	Residential	1996-03-01	60601.0	SD	Russell	LOOP	32.0	41.895021	-87.819841	(41.895021,-87.819841)	
192	103675	71 E Wacker Dr	Residential	2017-03-01	60601.0	IN	Goostland	LOOP	1.0	41.804047	-87.81546	(41.804047,-87.81546)	
104	104360	400 E Randolph St	Residential	2017-04-01	60601.0	IN	Goostland	LOOP	13.0	41.884865	-87.81635	(41.884865,-87.81635)	
691	100279	151 - 155 N Michigan Ave	Residential	1907-03-01	60601.0	IA	Mingo	LOOP	32.0	41.759933	-87.676702	(41.759933,-87.676702)	
697	100299	111 W WACKER DR	Residential	1900-03-01	60601.0	VA	Ebony	LOOP	32.0	41.947426	-87.824056	(41.947426,-87.824056)	
699	100303	360 E South Water St	Residential	1936-03-01	60601.0	MN	Russell	LOOP	32.0	41.886734	-87.618187	(41.886734,-87.618187)	
105	104474	201 N Westshore Dr	Residential	1914-04-01	60601.0	OH	Northgate	LOOP	32.0	41.845505	-87.80554	(41.845505,-87.80554)	
107	104479	450 E Waterside Dr	Residential	1936-04-01	60601.0	MN	Russell	LOOP	32.0	41.88716	-87.815304	(41.88716,-87.815304)	
108	104519	75 E Wacker Dr	Residential	1900-04-01	60601.0	VA	Ebony	LOOP	32.0	41.885653	-87.774748	(41.885653,-87.774748)	
109	104520	177 N Michigan Ave	Commercial	1914-04-01	60601.0	OH	Northgate	LOOP	32.0	41.900005	-87.69661	(41.900005,-87.69661)	
110	104515	180 N Wabash Ave	Commercial	1914-04-01	60601.0	OH	Northgate	LOOP	32.0	41.703765	-87.692676	(41.703765,-87.692676)	
111	105137	225 N Wabash	Residential	1968-04-01	60601.0	CO	Crisman	LOOP	24.0				

You can see that although the use of distance measure to remove some of the duplicate data, but still can not meet the e1 constraints, in order to meet the e1 constraints, we select the Zip under the State most frequent values, will be under the State class value update to the most frequent values, for the most frequent value of the same value for more than one State value, the use of weight algorithms, the weight based on the State column, the frequency of the appearance of State divided by the total number of values in the column of the State calculation, the City columns are also used in the same way to fix the repair, the use of Python merge repair, the completion of the repair, reread after importing the OpenRefine.

Distribution of Zip,State,City columns after fixing:

Building_ID	Address	Building_Type	Build_Year	Zip	State	City	Community_Name	Community_Number	Latitude	Longitude	Location
100429	875 N Michigan Ave	Commercial	1960-01-01	60611.0	VA	Northgate	NEAR NORTH SIDE	8.0	41.889804	-87.622958	(41.889804,-87.622958)
100447	211 E Chicago Ave	Commercial	1960-01-01	60611.0	VA	Northgate	NEAR NORTH SIDE	8.0	41.889804	-87.622958	(41.889804,-87.622958)
100465	100 E 133rd Pl	Residential	1914-03-01	60611.0	VA	Northgate	NEAR NORTH SIDE	8.0	41.889194	-87.626085	(41.889194,-87.626085)
101795	540 N Fairbanks Ct	Commercial	2014-03-01	60611.0	VA	Northgate	NEAR NORTH SIDE	8.0	41.759893	-87.670268	(41.759893,-87.670268)
101804	401 N Michigan Ave	Commercial	1960-03-01	60611.0	VA	Northgate	NEAR NORTH SIDE	8.0	41.889114	-87.622658	(41.889114,-87.622658)
101816	43 E Ohio St	Commercial	1960-03-01	60611.0	VA	Northgate	NEAR NORTH SIDE	8.0	41.706804	-87.470484	(41.706804,-87.470484)
101840	710 N Lakeshore Dr	Commercial	1965-03-01	60611.0	VA	Northgate	NEAR NORTH SIDE	8.0	41.895334	-87.169646	(41.895334,-87.169646)
101852	301 N Water St	Residential	2014-03-01	60611.0	VA	Northgate	NEAR NORTH SIDE	8.0	41.97223	-87.9545	(41.97223,-87.9545)
101874	444 N Michigan Ave	Commercial	2012-03-01	60611.0	VA	Northgate	NEAR NORTH SIDE	8.0	41.890094	-87.624748	(41.890094,-87.624748)
101883	678 N Michigan Ave	Commercial	1914-03-01	60611.0	VA	Northgate	NEAR NORTH SIDE	8.0	41.77034	-87.63724	(41.77034,-87.63724)
101890	345 E Superior St	Commercial	1965-03-01	60611.0	VA	Northgate	ASHBURN	7.0	41.895465	-87.61832	(41.895465,-87.61832)
101920	410 N MICHIGAN AVE	Commercial	1957-03-01	60611.0	VA	Northgate	NEAR NORTH SIDE	8.0	41.890064	-87.62518	(41.890064,-87.62518)
101935	500 N Michigan Ave	Commercial	2017-03-01	60611.0	VA	Northgate	NEAR NORTH SIDE	8.0	41.89117	-87.624695	(41.89117,-87.624695)
101971	25 E Pearson St	Commercial	2014-03-01	60611.0	VA	Northgate	NEAR NORTH SIDE	8.0	41.840508	-87.6162	(41.840508,-87.6162)
103132	199 E LAKE SHORE DR	Residential	1957-03-01	60611.0	VA	Northgate	NEAR NORTH SIDE	8.0	41.895794	-87.62248	(41.895794,-87.62248)
103315	230 E ONTARIO ST	Residential	1907-03-01	60611.0	VA	Northgate	NEAR NORTH SIDE	8.0	41.895794	-87.62248	(41.895794,-87.62248)
103352	201 N CEDAR ST	Residential	1907-03-01	60611.0	VA	Northgate	NEAR NORTH SIDE	8.0	41.895794	-87.62248	(41.895794,-87.62248)
103594	455 N Michigan Ave	Commercial	1923-03-01	60611.0	VA	Northgate	NEAR NORTH SIDE	8.0	41.80007	-87.621658	(41.80007,-87.621658)
106182	291 E Huron St	Commercial	1914-03-01	60611.0	VA	Northgate	NEAR NORTH SIDE	8.0	41.870854	-87.649685	(41.870854,-87.649685)
106361	980 N Michigan Ave	Commercial	1917-03-01	60611.0	VA	Northgate	NEAR NORTH SIDE	8.0	41.939384	-87.614261	(41.939384,-87.614261)
20512749	710 N Fairbanks Ct	Commercial	2012-03-01	60611.0	VA	Northgate	NEAR NORTH SIDE	8.0	41.89531	-87.620265	(41.89531,-87.620265)
21013782	138 E Huron St	Residential	1960-03-01	60611.0	VA	Northgate	NEAR NORTH SIDE	8.0	42.009603	-87.66729	(42.009603,-87.66729)
21113798	844 W Rush St	Commercial	1914-03-01	60611.0	VA	Northgate	NEAR NORTH SIDE	8.0	41.795612	-87.59862	(41.795612,-87.59862)
21310316	330 N Wabash Av	Commercial	1970-03-01	60611.0	VA	Northgate	HEGEWISCH	55.0	41.888838	-87.62752	(41.888838,-87.62752)
26510579	103 10E DELAWARE PL	Commercial	1965-03-01	60611.0	VA	Northgate	NEAR NORTH SIDE	8.0	41.895794	-87.62248	(41.895794,-87.62248)
267105749	211 217 E CHESTNUT ST	Residential	1966-03-01	60611.0	VA	Northgate	LOOP	32.0	41.895794	-87.62248	(41.895794,-87.62248)
272105864	600 E Grand Blvd	Commercial	1954-03-01	60611.0	VA	Northgate	NEAR NORTH SIDE	8.0	41.891743	-87.604706	(41.891743,-87.604706)
277106014	211 213 E DELAWARE PL	Residential	1969-03-01	60611.0	VA	Northgate	NEAR NORTH SIDE	8.0	41.895794	-87.62248	(41.895794,-87.62248)
286105046	102 E OHIO ST	Residential	2014-03-01	60611.0	VA	Northgate	NEAR NORTH SIDE	8.0	41.871033	-87.7334	(41.871033,-87.7334)
304111038	607 N Lake Shore Dr	Residential	1966-03-01	60611.0	VA	Northgate	LOOP	32.0	41.89281	-87.61456	(41.89281,-87.61456)
305111032	1100 N Michigan Ave	Commercial	1989-03-01	60611.0	VA	Northgate	NEAR NORTH SIDE	8.0	41.887786	-87.624587	(41.887786,-87.624587)
335112986	229 255 E LAKE SHORE DR	Residential	1957-03-01	60611.0	VA	Northgate	NEAR NORTH SIDE	8.0	41.895794	-87.62248	(41.895794,-87.62248)
337113077	291 E WALTON PL	Residential	1954-03-01	60611.0	VA	Northgate	NEAR NORTH SIDE	8.0	41.895794	-87.62248	(41.895794,-87.62248)
338113899	20 E CEDAR ST	Residential	1968-03-01	60611.0	VA	Northgate	NEAR NORTH SIDE	8.0	41.888698	-87.623055	(41.888698,-87.623055)
339113706	233 E ERIE ST	Residential	2017-03-01	60611.0	VA	Northgate	NEAR NORTH SIDE	31.0	41.895794	-87.62248	(41.895794,-87.62248)
340113726	227 239 E DELAWARE PL	Residential	2012-03-01	60611.0	VA	Northgate	NEAR NORTH SIDE	8.0	41.895794	-87.62248	(41.895794,-87.62248)
343114191	228 240 E WALTON ST	Residential	1936-03-01	60611.0	VA	Northgate	NEAR NORTH SIDE	8.0	41.895794	-87.62248	(41.895794,-87.62248)
437124236	830 N Michigan Ave	Commercial	1966-03-01	60611.0	VA	Northgate	NEAR NORTH SIDE	8.0	41.802303	-87.6025	(41.802303,-87.6025)
447125859	730 N Michigan Ave	Commercial	1935-03-01	60611.0	VA	Northgate	NEAR NORTH SIDE	8.0	41.896003	-87.624506	(41.896003,-87.624506)
450127047	609 N Michigan Ave	Commercial	2012-03-01	60611.0	VA	Northgate	NEAR NORTH SIDE	8.0	41.894463	-87.623272	(41.894463,-87.623272)
458128800	155 E Superior St	Commercial	1966-03-01	60611.0	VA	Northgate	NEAR NORTH SIDE	8.0	41.780464	-87.60531	(41.780464,-87.60531)
464129374	444 N WMA#454 AVE	Commercial	1946-03-01	60611.0	VA	Northgate	NEAR NORTH SIDE	24.0	41.899114	-87.69698	(41.899114,-87.69698)

Distribution of Zip,State,City columns after fixing:

Building_ID	Address	Building_Type	Build_Year	Zip	State	City	Community_Name	Community_Number	Latitude	Longitude	Location
102587	1310 S Carver Dr	Municipal	2017-03-01	60827.0	IN	Cambridge	RIVERDALE	68.0	41.65756	-87.59212	(41.65756,-87.59212)
250039	801 E 133rd Pl	Municipal	1996-04-01	60827.0	IN	Cambridge	RIVERDALE	54.0	41.895477	-87.600355	(41.895477,-87.600355)

Distribution of Zip,State,City columns after fixing, You can see that e2,e3 have now satisfied the constraints:

OpenRefine building V5 csv Permalink

Facet / Filter Undo / Redo 0 / 0 2 matching rows (1787 total)

Show as: rows records Show 5 10 25 50 100 500 1000 rows

All	Building_ID	Address	Building_Type	Build_Year	Zip	State	City	Community_Name	Community_Number	Latitude	Longitude	Location
156	102587	13100 S Carver Dr	Municipal	2017-03-01	60827 0	IN	Cambridge	RIVERDALE	68.0	41.65756	-87.59212	(41.65756, -87.59212)
1093	250039	801 E 133rd Pl	Municipal	1998-04-01	60827 0	IN	Cambridge	RIVERDALE	54.0	41.895477	-87.660355	(41.895478, -87.660355)

x first < previous 1 next > last *

Extensions Wikibase Help

Facet by choice counts

Zip change invert reset
59 choices. Sort by: name count Cluster
60645 0 10
60656 0 8
60638 0 8
60636 0 5
60707 0 5
60633 0 2
60634 0 2
60627 0 2
60654 0 1
(blank) 13
Exclude

State change
1 choice. Sort by: name count Cluster
IN 2
Facet by choice counts

City change
1 choice. Sort by: name count Cluster
Cambridge 2
Facet by choice counts

Now start fixing e4, first check the distribution of e4 constraint columns:

OpenRefine building V5 csv Permalink

Facet / Filter Undo / Redo 0 / 0 30 matching rows (1787 total)

Show as: rows records Show 5 10 25 50 100 500 1000 rows

All	Building_ID	Address	Building_Type	Build_Year	Zip	State	City	Community_Name	Community_Number	Latitude	Longitude	Location
216	103933	316 N CAMBRIDGE AVE	Residential	1957-03-01	60827 0	MN	Russell	LAKE VIEW	6.0	41.93985	-87.65845	(41.939851, -87.65844752)
217	103962	2828 N CANTONIAN AVE	Residential	1970-03-01	60827 0	MN	Russell	NEAR WEST SIDE	28.0	41.93985	-87.65845	(41.939851, -87.65844752)
218	104295	440 W MELROSE AVE	Residential	1960-03-01	60827 0	MN	Russell	LAKE VIEW	6.0	41.93985	-87.65845	(41.939851, -87.65844752)
219	105070	424 W GARDALE AVE	Residential	1954-03-01	60827 0	MN	Russell	LAKE VIEW	6.0	41.93985	-87.65845	(41.939851, -87.65844752)
220	104228	608 620 W CORNELIA AVE	Residential	1964-03-01	60827 0	MN	Russell	EAST MELROSE PARK	27.0	41.93985	-87.65845	(41.939851, -87.65844752)
221	104239	517 525 W BROMPTON AVE	Residential	1985-03-01	60827 0	MN	Russell	LAKE VIEW	6.0	41.93985	-87.65845	(41.939851, -87.65844752)
222	104250	544 W MELROSE ST	Residential	1969-03-01	60827 0	MN	Russell	LAKE VIEW	6.0	41.93985	-87.65845	(41.939851, -87.65844752)
223	104295	529 534 W ALDINE AVE	Residential	1921-03-01	60827 0	MN	Russell	LAKE VIEW	6.0	41.93985	-87.65845	(41.939851, -87.65844752)
224	104296	424 W BARRY AVE	Residential	1936-03-01	60827 0	MN	Russell	LAKE VIEW	6.0	41.93985	-87.65845	(41.939851, -87.65844752)
225	104396	426 W SURF ST	Residential	1957-03-01	60827 0	MN	Russell	LAKE VIEW	6.0	41.93985	-87.65845	(41.939851, -87.65844752)
226	105915	421 W BARRY AVE	Residential	1966-03-01	60827 0	MN	Russell	LINCOLN PARK	7.0	41.93985	-87.65845	(41.939851, -87.65844752)
261	105917	441 W GARDALE AVE	Residential	1925-03-01	60827 0	MN	Russell	LAKE VIEW	39.0	41.93985	-87.65845	(41.939851, -87.65844752)
276	106500	433 W MELROSE ST	Residential	1923-03-01	60827 0	MN	Russell	LAKE VIEW	6.0	41.93985	-87.65845	(41.939851, -87.65844752)
285	106502	344 350 W GARDALE AVE	Residential	2010-03-01	60827 0	MN	Russell	LAKE VIEW	50.0	41.93985	-87.65845	(41.939851, -87.65844752)
295	109413	550 555 W CRAFT ST	Residential	1957-03-01	60827 0	MN	Russell	LAKE VIEW	6.0	41.93985	-87.65845	(41.939851, -87.65844752)
317	112427	415 421 W ALDINE AVE	Residential	1960-03-01	60827 0	MN	Russell	LAKE VIEW	6.0	41.93985	-87.65845	(41.939851, -87.65844752)
364	115942	507 W SURF ST	Municipal	1968-03-01	60827 0	MN	Russell	NEAR NORTH SIDE	8.0	41.93985	-87.65845	(41.939851, -87.65844752)
387	117179	3318 N LAKE SHORE DR	Residential	1995-03-01	60827 0	MN	Russell	NEAR NORTH SIDE	8.0	41.93985	-87.65845	(41.939851, -87.65844752)
800	126215	2843 N LINCOLN AVE	Residential	1936-03-01	60827 0	MN	Russell	LAKE VIEW	6.0	41.93985	-87.65845	(41.939851, -87.65844752)
838	127352	415 421 W MELROSE ST	Residential	1957-03-01	60827 0	MN	Russell	LAKE VIEW	6.0	41.93985	-87.65845	(41.939851, -87.65844752)
840	127354	422 424 W MELROSE ST	Residential	2017-03-01	60827 0	MN	Russell	LAKE VIEW	6.0	41.93985	-87.65845	(41.939851, -87.65844752)
842	127363	425 427 W ROCOCO ST	Residential	2010-03-01	60827 0	MN	Russell	LAKE VIEW	6.0	41.93985	-87.65845	(41.939851, -87.65844752)
845	127374	2816 2828 N BURLING ST	Residential	1957-03-01	60827 0	MN	Russell	LAKE VIEW	6.0	41.93985	-87.65845	(41.939851, -87.65844752)
855	127375	301 303 W MELROSE AVE	Residential	1950-03-01	60827 0	MN	Russell	LAKE VIEW	6.0	41.93985	-87.65845	(41.939851, -87.65844752)
895	175265	420 W MELROSE ST	Residential	1921-03-01	60827 0	MN	Russell	LAKE VIEW	6.0	41.93985	-87.65845	(41.939851, -87.65844752)
900	175294	540 W SURF ST	Residential	1968-03-01	60827 0	MN	Russell	WEST TOWN	24.0	41.93985	-87.65845	(41.939851, -87.65844752)
917	175492	3158 3194 N CLARK ST	Residential	1907-03-01	60827 0	MN	Russell	LAKE VIEW	6.0	41.93985	-87.65845	(41.939851, -87.65844752)
928	175698	1001 W BELMONT AVE	Residential	2017-03-01	60827 0	MN	Russell	LAKE VIEW	6.0	41.93985	-87.65845	(41.939851, -87.65844752)
1338	251640	6515 S Champlain Ave	Municipal	1987-04-01	60827 0	IL	Whitmore Village	WOODLAWN	42.0	41.93985	-87.65845	(41.775182, -87.6090298)
1497	253805	420 W Surf St	Residential	2012-04-01	60827 0	MN	Russell	LAKE VIEW	6.0	41.93985	-87.65845	(41.939851, -87.65844752)

x first < previous 1 next > last *

Extensions Wikibase Help

Facet by choice counts

Latitude change invert reset
989 choices. Sort by: name count Cluster
41.93985 30
41.84550 37
41.91675 25
41.90833 25
41.92425 24
41.82242 22
41.92201 22
42.01038 22
41.95402 21
41.91781 20
41.93973 17
41.67739 16
Facet by choice counts

Longitude change
1 choice. Sort by: name count Cluster
-87.65845 30
Facet by choice counts

Location change
2 choices. Sort by: name count Cluster
(41.775182, -87.6090298) 1
(41.939851, -87.65844752) 29
Facet by choice counts

In order to meet the e4 constraints, we choose Latitude, Longitude under the Location of the most frequent values, will be updated to the most frequent values under the Location category, for the most frequent values are equal to the value of multiple Locations, the default direct return to the first to find the maximum value of the use of Python merge fix, merge all the Locations will be set to a combination of Latitude and Longitude after the completion of the fix to re-import the OpenRefine

OpenRefine building V7 csv [Permalink](#)

Facet / Filter Undo / Redo 0 / 0 < 30 matching rows (1787 total)

Show as: rows records Show: 5 10 25 50 100 500 1000 rows

Latitude change invert reset
989 choices Sort by: name count Cluster
41.93985 30 exclude

Longitude change
1 choice Sort by: name count Cluster
-87.65845 30 Facet by choice counts

Location change
1 choice Sort by: name count Cluster
(41.93985, -87.65845) 30 Facet by choice counts

All Building_ID Address Building_Type Build_Year Zip State City Community_Name Community_Number Latitude Longitude Location

216	103933	3161 N CAMBRIDGE AVE	Residential	1957-03-01	60657.0	MN	Russell	LAKE VIEW	6.0	41.93985	-87.65845	(41.93985, -87.65845)
217	103962	2828 N CAMBRIDGE AVE	Residential	1970-03-01	60657.0	MN	Russell	NEAR WEST SIDE	28.0	41.93985	-87.65845	(41.93985, -87.65845)
218	103965	447 457 W WELLINGTON AVE	Residential	1968-03-01	60657.0	MN	Russell	LAKE VIEW	6.0	41.93985	-87.65845	(41.93985, -87.65845)
219	103970	424 W OAKDALE AVE	Residential	1954-03-01	60657.0	MN	Russell	LAKE VIEW	6.0	41.93985	-87.65845	(41.93985, -87.65845)
221	104226	606 620 W CORNELIA AVE	Residential	1966-03-01	60657.0	MN	Russell	EAST GARFIELD PARK	27.0	41.93985	-87.65845	(41.93985, -87.65845)
223	104239	517 525 W BROMPTON AVE	Residential	1985-03-01	60657.0	MN	Russell	LAKE VIEW	6.0	41.93985	-87.65845	(41.93985, -87.65845)
224	104250	544 W MELROSE ST	Residential	1989-03-01	60657.0	MN	Russell	LAKE VIEW	6.0	41.93985	-87.65845	(41.93985, -87.65845)
227	104285	528 534 W ALDINE AVE	Residential	1921-03-01	60657.0	MN	Russell	LAKE VIEW	6.0	41.93985	-87.65845	(41.93985, -87.65845)
236	104835	445 W BARRY AVE	Residential	1936-03-01	60657.0	MN	Russell	LAKE VIEW	6.0	41.93985	-87.65845	(41.93985, -87.65845)
237	104838	421 W BARRY AVE	Residential	1957-03-01	60657.0	MN	Russell	LAKE VIEW	6.0	41.93985	-87.65845	(41.93985, -87.65845)
260	105915	428 W SURF ST	Residential	1995-03-01	60657.0	MN	Russell	LINCOLN PARK	7.0	41.93985	-87.65845	(41.93985, -87.65845)
261	105917	441 W OAKDALE AVE	Residential	1925-03-01	60657.0	MN	Russell	LAKE VIEW	39.0	41.93985	-87.65845	(41.93985, -87.65845)
279	106930	433 W BRIAR PL	Residential	2012-03-01	60657.0	MN	Russell	LAKE VIEW	6.0	41.93985	-87.65845	(41.93985, -87.65845)
285	106922	344 358 W OAKDALE AVE	Residential	2017-03-01	60657.0	MN	Russell	LAKE VIEW	50.0	41.93985	-87.65845	(41.93985, -87.65845)
295	109413	550 558 W SURF ST	Residential	1957-03-01	60657.0	MN	Russell	LAKE VIEW	6.0	41.93985	-87.65845	(41.93985, -87.65845)
317	112437	415 421 W ALDINE AVE	Residential	1990-03-01	60657.0	MN	Russell	LAKE VIEW	6.0	41.93985	-87.65845	(41.93985, -87.65845)
364	115942	501 W SURF ST	Municipal	1996-03-01	60657.0	MN	Russell	NEAR NORTH SIDE	8.0	41.93985	-87.65845	(41.93985, -87.65845)
387	117179	3318 N LAKE SHORE DR	Residential	1996-03-01	60657.0	MN	Russell	NEAR NORTH SIDE	8.0	41.93985	-87.65845	(41.93985, -87.65845)
800	172615	2843 N LINCOLN AVE	Residential	1996-03-01	60657.0	MN	Russell	LAKE VIEW	6.0	41.93985	-87.65845	(41.93985, -87.65845)
839	173752	415 421 W MELROSE ST	Residential	1957-03-01	60657.0	MN	Russell	LAKE VIEW	6.0	41.93985	-87.65845	(41.93985, -87.65845)
840	173754	422 424 W MELROSE ST	Residential	2017-03-01	60657.0	MN	Russell	LAKE VIEW	6.0	41.93985	-87.65845	(41.93985, -87.65845)
842	173763	425 427 W ROSSCOE ST	Residential	2010-03-01	60657.0	MN	Russell	LAKE VIEW	6.0	41.93985	-87.65845	(41.93985, -87.65845)
845	173776	2818 2828 W BURLING ST	Residential	1957-03-01	60657.0	MN	Russell	LAKE VIEW	6.0	41.93985	-87.65845	(41.93985, -87.65845)
848	173779	311 S CANTON AVE	Residential	1980-03-01	60657.0	MN	Russell	LAKE VIEW	6.0	41.93985	-87.65845	(41.93985, -87.65845)
856	176255	600 W MELROSE ST	Residential	1921-03-01	60657.0	MN	Russell	LAKE VIEW	6.0	41.93985	-87.65845	(41.93985, -87.65845)
900	176264	640 W SURF ST	Residential	1996-03-01	60657.0	MN	Russell	WEST TOWN	24.0	41.93985	-87.65845	(41.93985, -87.65845)
917	176402	3158 319 N CLARK ST	Residential	1907-03-01	60657.0	MN	Russell	LAKE VIEW	6.0	41.93985	-87.65845	(41.93985, -87.65845)
926	176569	1001 W BELMONT AVE	Residential	2017-03-01	60657.0	MN	Russell	LAKE VIEW	6.0	41.93985	-87.65845	(41.93985, -87.65845)
1336	251640	6543 S Champaign Ave	Municipal	1987-04-01	60657.0	IL	Whitemore Village	WIGGOLAWN	42.0	41.93985	-87.65845	(41.93985, -87.65845)
1497	253905	420 W Surf St	Residential	2012-04-01	60657.0	MN	Russell	LAKE VIEW	6.0	41.93985	-87.65845	(41.93985, -87.65845)

Detect column value distribution after fixing e4:

OpenRefine building V7 csv [Permalink](#)

Facet / Filter Undo / Redo 0 / 0 < 4 matching rows (1787 total)

Show as: rows records Show: 5 10 25 50 100 500 1000 rows

Latitude change invert reset
989 choices Sort by: name count Cluster
41.93985 30 exclude

Longitude change
1 choice Sort by: name count Cluster
-87.640686 4 Facet by choice counts

Location change
1 choice Sort by: name count Cluster
(41.77613, -87.640686) 4 Facet by choice counts

All Building_ID Address Building_Type Build_Year Zip State City Community_Name Community_Number Latitude Longitude Location

22	100683	7000 S PARINELL AVE	Residential	1970-03-01	60621.0	FL	Northshore	NEAR SOUTH SIDE	33.0	41.77613	-87.640686	(41.77613, -87.640686)
543	152408	735 745 W GARFIELD BLVD	Residential	1969-03-01	60621.0	FL	Northshore	ENGLEWOOD	22.0	41.77613	-87.640686	(41.77613, -87.640686)
638	159660	1730 N Clark St	Residential	1917-03-01	60614.0	VA	Leesburg	LINCOLN PARK	7.0	41.77613	-87.640686	(41.77613, -87.640686)
847	173783	4840 N LUNDER AVE	Residential	1917-03-01	60630.0	VA	Leesburg	JEFFERSON PARK	11.0	41.77613	-87.640686	(41.77613, -87.640686)

Detect column value distribution after fixing e4, You can see that e4 now satisfies the constraint:

Screenshot of OpenRefine interface showing a search results page for "1 matching rows (1787 total)".

The search results table includes columns: Building_ID, Address, Building_Type, Build_Year, Zip, State, City, Community_Name, Community_Number, Latitude, Longitude, and Location.

Facets on the left side show:

- Latitude**: 899 choices, Sort by name, count, Cluster. Includes 41.670017, 41.689095, 41.685665, 41.696396, 41.691947, 41.692963 (highlighted in red), 41.692986, 41.693764, 41.693988, 41.694427, 41.695263.
- Longitude**: 1 choice, Sort by name, count, Cluster. Includes -87.60058.
- Location**: 1 choice, Sort by name, count, Cluster. Includes (41.692963, -87.60058).

Start to fix e3, first view e3 constraints column value distribution, you can see that the existing data is to meet the e3 constraints, the use of Python code also did not detect a violation of the constraints e3 situation.

Screenshot of OpenRefine interface showing a search results page for "1 matching rows (1787 total)".

The search results table includes columns: Building_ID, Address, Building_Type, Build_Year, Zip, State, City, Community_Name, Community_Number, Latitude, Longitude, and Location.

Facets on the left side show:

- Address**: 1778 choices, Sort by name, count, Cluster. Includes 950 E 61st St, 9501 S Martin Luther King Dr, 9510 S CONSTANCE AVE, 9525 S Halsted St, 954 W Washington Blvd, 9730 N Paulina, 980 N Michigan Ave, 9811 S Lowe Ave, 984 N MILWAUKEE AVE (highlighted in red), 9912 S Avenue H.
- Location**: 1 choice, Sort by name, count, Cluster. Includes (41.899406, -87.66143).

Start fixing e5 to see the distribution of e5 constraint column values

OpenRefine building V7 csv [Permalink](#)

Facet / Filter Undo / Redo 0 / 0

Refresh Reset all Remove all

1787 rows

Show as: rows records Show: 5 10 25 50 100 500 1000 rows

All	Location	Building_ID	Address	Building_Type	Build_Year	Zip	State	City	Community_Name	Community_Number	Latitude	Longitude	Location
1	100088	655 W Roosevelt Rd	Commercial	2017-03-01	60607.0	IL	Cambridge	NEAR WEST SIDE	32.0	41.999397	-87.64165	(41.999397, -87.64165)	
2	100256	407 S Archer Ave	Municipal	1996-03-01	60620.0	KY	Moordland	ARCHER HEIGHTS	67.0	41.75302	-87.001974	(41.75302, -87.001974)	
3	100320	4039-4051 N LAPORTE AVE	Residential	1996-03-01	60641.0	IL	Snake Creek	PORTRIDGE PARK	15.0	41.94915	-87.74708	(41.94915, -87.74708)	
4	100346	2600 W BELMONT AVE	Residential	1994-03-01	60610.0	IL	Crooked	AVONDALE	21.0	41.94872	-87.70219	(41.94872, -87.70219)	
5	100372	2701 W 68th St	Commercial	1997-03-01	60620.0	IL	Rockridge	CHICAGO LAWN	66.0	41.77004	-87.69084	(41.77004, -87.69084)	
6	100383	7601 S Cicero Ave	Commercial	1997-03-01	60622.0	IL	Northgate	WEST LAWN	65.0	41.920395	-87.756165	(41.920395, -87.756165)	
7	100396	7740 S Western Ave	Commercial	1997-03-01	60620.0	IL	Leesburg	ASHBURN	70.0	41.75293	-87.68267	(41.75293, -87.68267)	
8	100397	3039 W 70th St	Municipal	1996-03-01	60620.0	IL	Northgate	DOUGLAS	35.0	41.74945	-87.720146	(41.74945, -87.720146)	
9	100429	875 N Michigan Ave	Commercial	1997-03-01	60611.0	VA	Northgate	NEAR NORTH SIDE	8.0	41.89880	-87.82288	(41.89880, -87.82288)	
10	100433	2233 W Division St	Commercial	2017-03-01	60622.0	IL	Leesburg	WEST TOWN	5.0	41.902515	-87.83336	(41.902515, -87.83336)	
11	100447	211 E Chicago Ave	Commercial	1996-03-01	60611.0	VA	Northgate	NEAR NORTH SIDE	8.0	41.89645	-87.82257	(41.89645, -87.82257)	
12	100448	198 E Delaware Pl	Residential	1994-03-01	60611.0	VA	Northgate	NEAR NORTH SIDE	8.0	41.883194	-87.828685	(41.883194, -87.828685)	
13	100470	30 W Chicago	Residential	1990-03-01	60654.0	VA	Sale Creek	NEAR NORTH SIDE	8.0	41.87781	-87.65493	(41.87781, -87.65493)	
14	100489	4520 4528 S DREXEL BLVD	Residential	2017-03-01	60653.0	VA	Cambridge	KENWOOD	3.0	41.82021	-87.612465	(41.82021, -87.612465)	
15	100507	4501 4952 S PRAIRIE AVE	Residential	1998-03-01	60615.0	VA	Crisman	GRAND BOULEVARD	38.0	41.902242	-87.80325	(41.902242, -87.80325)	
16	100533	1621 S Stewart Ave	Municipal	1994-03-01	60621.0	FL	Northchase	ENGLEWOOD	68.0	41.762055	-87.63427	(41.762055, -87.63427)	
17	100572	813 S May St	Commercial	1997-03-01	60611.0	VA	Northgate	AUBURN GRESHAM	71.0	41.74538	-87.65152	(41.74538, -87.65152)	
18	100623	1345 1359 W ESTES AVE	Residential	1997-03-01	60626.0	IL	Rockridge	ROGERS PARK	1.0	42.01036	-87.6684	(42.01036, -87.6683)	
19	100637	510 534 W ADDISON ST	Residential	1990-03-01	60613.0	VA	Ebony	LAKE VIEW	6.0	41.89217	-87.63413	(41.89217, -87.63413)	
20	100666	3245 W ARTHINGTON ST	Commercial	1994-03-01	60624.0	OH	Northgate	EAST GARFIELD PARK	1.0	41.89029	-87.61818	(41.89029, -87.61818)	
21	100670	2012 S Armitage St	Residential	2012-03-01	60624.0	OH	Northgate	NEAR EAST SIDE	24.0	41.890007	-87.63332	(41.890007, -87.63332)	
22	100693	7001 S KEDRICK AVE	Residential	1991-03-01	60621.0	IL	Northgate	NEAR SOUTH SIDE	33.0	41.777803	-87.771813	(41.777803, -87.771813)	
23	100729	801 S Canal St	Commercial	1990-03-01	60607.0	IL	Canbridge	NEAR WEST SIDE	28.0	42.03038	-87.68933	(42.03038, -87.68933)	
24	100733	940 S CANAL ST	Commercial	1998-03-01	60610.0	IL	Canbridge	NEAR WEST SIDE	28.0	41.98029	-87.611916	(41.98029, -87.611916)	
25	100881	600 W Chicago	Commercial	1995-03-01	60654.0	VA	Sale Creek	NEAR NORTH SIDE	8.0	41.8996	-87.64288	(41.8996, -87.64288)	
26	100897	1653 W Congress Pkwy	Commercial	1996-03-01	60612.0	IL	Northchase	NEAR WEST SIDE	28.0	41.945505	-87.630554	(41.945505, -87.630554)	
27	100978	1521 1523 N W SHERWIN AVE	Residential	2010-03-01	60626.0	IL	Rockridge	ROGERS PARK	1.0	42.01036	-87.66833	(42.01036, -87.66833)	
28	100984	8425 S Lowe	Residential	1997-03-01	60621.0	FL	Northgate	ENGLEWOOD	68.0	41.780464	-87.60531	(41.780464, -87.60531)	
29	101013	7320 7323 N DAMEN AVE	Residential	1999-03-01	60654.0	IL	Mingo	ROGERS PARK	3.0	42.0088	-87.69503	(42.0088, -87.69503)	
30	101030	3750 S COTTAGE GROVE AVE	Residential	1997-03-01	60653.0	VA	Cambridge	OAKLAND	36.0	41.82021	-87.612465	(41.82021, -87.612465)	
31	101032	5110 S DR MARTIN LUTHER KING JR DR	Residential	1994-03-01	60615.0	VA	Crisman	WASHINGTON PARK	40.0	41.802242	-87.60332	(41.802242, -87.60332)	
32	101039	3250 W Adams St	Municipal	1970-03-01	60624.0	OH	Northgate	NEAR NORTH SIDE	8.0	41.879097	-87.707405	(41.879097, -87.707405)	
33	101043	2835 W Polk St	Municipal	1914-03-01	60612.0	IL	Northchase	EAST GARFIELD PARK	27.0	41.883114	-87.64848	(41.883114, -87.64848)	
34	101086	1016 W HOLLYWOOD AVE	Residential	1914-03-01	60600.0	IL	Northgate	EDGEWATER	77.0	41.77034	-87.63724	(41.77034, -87.63724)	
35	101090	4554 S DREXEL BLVD	Residential	1990-03-01	60653.0	VA	Cambridge	KENWOOD	39.0	41.88966	-87.620305	(41.88966, -87.620305)	
36	101185	350 E Germak	Commercial	1925-03-01	60616.0	IL	Big Foot Prairie	NEAR SOUTH SIDE	8.0	41.853867	-87.61836	(41.853867, -87.61836)	
37	101199	4934 S Wabash Ave	Municipal	2017-03-01	60615.0	VA	Crisman	GRAND BOULEVARD	43.0	41.804024	-87.62487	(41.804024, -87.62487)	

Merge to fix some of the Community_Name column data using the openRefine clustering algorithm.

OpenRefine building V7 csv [Permalink](#)

Facet / Filter Undo / Redo 0 / 0

Refresh Reset all Remove all

1787 rows

Show as: rows records Show: 5 10 25 50 100 500 1000 rows

Cluster and edit column "Community_Name"

Find groups of different cell values that might be other representations of the same thing. For example, "New York" and "new york" likely refer to the same concept and just differ by capitalization, and "Gödel" and "Godel" probably refer to the same person. Find out more...

Method: Key collision Keying function: Fingerprint

41 clusters found

Cluster size	Row count	Values in cluster	Merge? New cell value
2	42	HYDE PARK (39 rows) Hyde Park (3 rows)	<input type="checkbox"/> HYDE PARK
2	8	AVALON PARK (7 rows) Avalon Park (2 rows)	<input type="checkbox"/> AVALON PARK
2	4	GARFIELD RIDGE (3 rows) Garfield Ridge	<input type="checkbox"/> GARFIELD RIDGE
2	10	AVONDALE (8 rows) Avondale (2 rows)	<input type="checkbox"/> AVONDALE
2	181	LOOP (174 rows) Loop (7 rows)	<input type="checkbox"/> LOOP
2	12	OHARE (11 rows) O'Hare	<input type="checkbox"/> OHARE
2	2	CALUMET HEIGHTS Calumet Heights	<input type="checkbox"/> CALUMET HEIGHTS

Rows in cluster

Average length of choices

Length variance of choices

Select all Deselect all Export clusters Merge selected & re-cluster Merge selected & Close Close

Longitude	Location
87.61445	(41.989897, -87.654145)
87.601974	(41.75302, -87.601974)
87.747078	(41.94615, -87.747078)
87.70319	(41.94673, -87.70319)
87.69904	(41.77004, -87.69904)
87.756165	(41.920395, -87.756165)
87.692097	(41.75293, -87.692097)
87.720148	(41.74946, -87.720148)
87.622998	(41.89804, -87.622998)
87.63336	(41.902515, -87.63336)
87.62597	(41.99445, -87.62597)
87.638914	(41.893194, -87.638914)
87.654903	(41.97781, -87.654903)
87.612465	(41.82021, -87.612465)
87.603332	(41.902242, -87.60332)
87.634207	(41.78205, -87.634207)
87.65152	(41.74538, -87.65152)
87.69883	(42.01096, -87.69883)
87.61816	(41.90209, -87.61816)
87.70832	(41.570007, -87.70832)
87.640993	(41.77613, -87.640993)
87.69983	(42.01096, -87.69983)
87.61816	(41.90209, -87.61816)
87.64298	(41.99695, -87.64298)
87.630554	(41.845505, -87.630554)
87.66883	(42.01096, -87.66883)
87.60531	(41.780464, -87.60531)
87.69903	(42.0088, -87.69903)
87.612465	(41.82021, -87.612465)
87.60332	(41.902242, -87.60332)
87.637045	(41.787097, -87.637045)
87.61816	(41.90209, -87.61816)
87.63724	(41.77034, -87.63724)
87.60531	(41.780464, -87.60531)

Clustering algorithms merge the number of fix rows:

OpenRefine building V7 csv

1787 rows

Show as: rows records Show: 5 10 25 50 100 500 1000 rows

Mass edit 1,536 cells in column Community_Name Undo

facet / Filter Undo / Redo 1 / 1

refresh Reset all Remove all

Location change

06 choices Sort by: name count Cluster

1 651405, -87.54583 | 1
1 657314, -87.55272 | 1
1 670017, -87.62316 | 1
1 680508, -87.66246 | 1
1 680894, -87.53555 | 1
1 684685, -87.65477 | 1
1 685665, -87.62176 | 1
1 686398, -87.63336 | 1
1 691647, -87.66579 | 1
1 692963, -87.60058 | 1
1 696948, -87.67073 | 1

Community_Name change

1 choices Sort by: name count Cluster

BANK PARK 4
CHER HEIGHTS 2
MOURN SQUARE 15
SHAW 10
JBURN GRESHAM 7
JSTIN 30
ALON PARK 8
ONDALE 10
ELMONT CRAGIN 8
NDIGOPEST 5
RIGHTON PARK 6
IDKSPINE 1

All Building_ID Address Building_Type Build_Year Zip State City Community_Name Community_Number Latitude Longitude Location

1 100068 555 W Roosevelt Rd Commercial 2017-03-01 60607 0 IL Cambridge NEAR WEST SIDE 32.0 41.866867 -87.64145 (41.866867, -87.64145)
2 100256 4975 S Archer Ave Municipal 1996-03-01 60632 0 KY Mooriand ARCHER HEIGHTS 57.0 41.75392 -87.601974 (41.75392, -87.601974)
3 100320 4039 4051 N LAPORTE AVE Residential 1936-03-01 60641 0 OK Snake Creek PORTAGE PARK 15.0 41.94615 -87.74708 (41.94615, -87.74708)
4 100346 2609 W BELMONT AVE Residential 1954-03-01 60618 0 IL Crisman AVONDALE 21.0 41.94672 -87.70319 (41.94672, -87.70319)
5 100372 2701 W 68th St Commercial 1957-03-01 60629 0 IL Rockbridge CHICAGO LAWN 66.0 41.77004 -87.69084 (41.77004, -87.69084)
6 100393 7601 S Cicero AVE Commercial 1907-03-01 60652 0 IL Northgate WEST LAWN 65.0 41.920395 -87.75615 (41.920395, -87.75615)
7 100396 7740 S Western Ave Commercial 1957-03-01 60620 0 IL Leesburg ASHBURN 70.0 41.75262 -87.68297 (41.75262, -87.68297)
8 100397 3939 W 70th St Municipal 1966-03-01 60652 0 IL Northgate DOUGLAS 35.0 41.74946 -87.720146 (41.74946, -87.720146)
9 100429 875 N Michigan Ave Commercial 1957-03-01 60611 0 VA Northgate NEAR NORTH SIDE 8.0 41.898804 -87.62298 (41.898804, -87.62298)
10 100433 2233 W Division St Commercial 2017-03-01 60622 0 IL Leesburg WEST TOWN 5.0 41.902515 -87.68336 (41.902515, -87.68336)
11 100447 211 E Chicago Ave Commercial 1936-03-01 60611 0 VA Northgate NEAR NORTH SIDE 8.0 41.89645 -87.62257 (41.89645, -87.62257)
12 100448 198 E Delaware Pl Residential 1914-03-01 60611 0 VA Northgate NEAR NORTH SIDE 8.0 41.883194 -87.628685 (41.883194, -87.628685)
13 100470 30 W Chicago Residential 1900-03-01 60654 0 VA Sale Creek NEAR NORTH SIDE 8.0 41.87781 -87.65493 (41.87781, -87.65493)
14 100489 4520 4520 S DREXEL BLVD Residential 2017-03-01 60653 0 VA Cramen GRAND BOULEVARD 38.0 41.80242 -87.603325 (41.80242, -87.603325)
15 100507 4950 4952 S PRAIRIE AVE Residential 1936-03-01 60615 0 VA Cramen KENWOOD 3.0 41.802242 -87.612465 (41.802242, -87.612465)
16 100533 6201 S Street Ave Municipal 1954-03-01 60621 0 FL Northchase ENGLEWOOD 68.0 41.780255 -87.63427 (41.780255, -87.63427)
17 100572 8131 S May St Commercial 1957-03-01 60620 0 IL Leesburg AUBURN GRESHAM 71.0 41.74538 -87.65152 (41.74538, -87.65152)
18 100623 1345 1359 W ESTES AVE Residential 1957-03-01 60626 0 IL Rockbridge ROGERS PARK 1.0 42.01036 -87.6886 (42.01036, -87.6886)
19 100637 510 534 W ADDISON ST Residential 1900-03-01 60613 0 VA Ebony LAKE VIEW 6.0 41.89217 -87.63413 (41.89217, -87.63413)
20 100668 3245 W ARTHINGTON ST Commercial 1914-03-01 60624 0 OH Northgate EAST GARFIELD PARK 27.0 41.89029 -87.61816 (41.89029, -87.61816)
21 100674 3301 W Arthington St Residential 2012-03-01 60624 0 OH Northgate NORTH LAWNDALE 29.0 41.870007 -87.70832 (41.870007, -87.70832)
22 100683 7000 S PARNELL AVE Residential 1970-03-01 60621 0 FL Northchase NEAR SOUTH SIDE 33.0 41.77613 -87.640686 (41.77613, -87.640686)
23 100729 801 S Canal St Commercial 1900-03-01 60607 0 IL Cambridge NEAR WEST SIDE 28.0 42.01036 -87.6686 (42.01036, -87.6686)
24 100733 840 S CANAL ST Commercial 1996-03-01 60607 0 IL Cambridge NEAR WEST SIDE 28.0 41.89029 -87.61816 (41.89029, -87.61816)
25 100851 600 W Chicago Commercial 1925-03-01 60654 0 VA Sale Creek NEAR NORTH SIDE 8.0 41.8965 -87.64288 (41.8965, -87.64288)
26 100867 1653 W Congress Pkwy Commercial 1996-03-01 60612 0 IL Northchase NEAR WEST SIDE 28.0 41.845505 -87.630554 (41.845505, -87.630554)
27 100878 1521 1523 W SHERWIN AVE Residential 2010-03-01 60626 0 IL Rockbridge ROGERS PARK 1.0 42.01036 -87.6686 (42.01036, -87.6686)
28 100946 6425 S Lowe Residential 1907-03-01 60621 0 FL Northchase ENGLEWOOD 68.0 41.780464 -87.60531 (41.780464, -87.60531)
29 101013 7320 7332 N DAMEN AVE Residential 1969-03-01 60645 0 IA Mingo ROGERS PARK 30.0 42.0088 -87.69503 (42.0088, -87.69503)
30 101030 3750 S COTTAGE GROVE AVE Residential 1957-03-01 60653 0 VA Cambridge OAKLAND 36.0 41.82021 -87.612465 (41.82021, -87.612465)
31 101032 5110 S DR MARTIN LUTHER KING JR DR Residential 1954-03-01 60615 0 VA Cramen WASHINGTON PARK 40.0 41.80242 -87.603325 (41.80242, -87.603325)
32 101038 2350 W Adams St Municipal 1970-03-01 60624 0 OH Northgate NEAR NORTH SIDE 8.0 41.879097 -87.707405 (41.879097, -87.707405)
33 101043 2935 W Polk St Municipal 1914-03-01 60612 0 IL Northchase EAST GARFIELD PARK 27.0 41.883114 -87.64488 (41.883114, -87.64488)
34 101086 1016 W HOLLYWOOD AVE Residential 1914-03-01 60660 0 IL Northgate EDGEWATER 77.0 41.77034 -87.63724 (41.77034, -87.63724)
35 101090 4564 S DREXEL BLVD Residential 1900-03-01 60653 0 VA Cambridge KENWOOD 39.0 41.88698 -87.628055 (41.88698, -87.628055)
36 101185 360 F. Cormac Commercial 1905-03-01 60611 0 IL Rin Front Prairie NEAR NORTH SIDE 8.0 41.85987 -87.61896 (41.85987, -87.61896)

After the merge, the e5 constraints are checked again, and you can see that some of them still do not satisfy the e5 constraints.

OpenRefine building V8 csv

3 matching rows (1787 total)

Show as: rows records Show: 5 10 25 50 100 500 1000 rows

Extensions: Wikibase Open Export Help

facet / Filter Undo / Redo 0 / 0

refresh Reset all Remove all

Location change invert reset

113 choices Sort by: name count Cluster

(41.873993, -87.65160) 4
(41.877644, -87.67413) 4
(41.89029, -87.61816) 4
(41.845505, -87.630554) 3 exclude
(41.874294, -87.65018) 3
(41.97525, -87.69722) 3
(41.71459, -87.68148) 2
(41.75392, -87.601974) 2
(41.79329, -87.630838) 2
(41.84615, -87.67062) 2
(41.877457, -87.635735) 2
(41.88711, -87.64537) 2 exclude

Community_Number change invert reset

14 choices Sort by: name count Cluster

12.0 1
22.0 1
24.0 1
28.0 3
30.0 1
32.0 1
33.0 4
34.0 4
35.0 4
40.0 1
42.0 2
6.0 1

Community_Name change

2 choices Sort by: name count Cluster

ARMOUR SQUARE 1
NEAR WEST SIDE 2

Facet by choice count

All Building_ID Address Building_Type Build_Year Zip State City Community_Name Community_Number Latitude Longitude Location

26 100867 1653 W Congress Pkwy Commercial 1996-03-01 60612 0 IL Northchase NEAR WEST SIDE 28.0 41.845505 -87.630554 (41.845505, -87.630554)
938 175754 120 N Green St Residential 1996-03-01 60607 0 IL Cambridge NEAR WEST SIDE 28.0 41.845505 -87.630554 (41.845505, -87.630554)
1765 260135 309 W 23RD ST Residential 2017-04-01 60616 0 IL Big Foot Prairie ARMOUR SQUARE 28.0 41.845505 -87.630554 (41.845505, -87.630554)

Use the same fix method as e1 to fix e5, and re-import openRefine after the fixed, You can see that the fixed data satisfy the constraint e5.

OpenRefine building V9 csv Permalink

Facet / Filter Undo / Redo 0 / 0 Extensions Wikibase Help

3 matching rows (1787 total)

Location	Building_ID	Address	Building_Type	Build_Year	Zip	State	City	Community_Name	Community_Number	Latitude	Longitude	Location
113 choices. Sort by: name count Cluster	26	100695	1055 W Congress Pkwy	Commercial	1996-03-01	60612.0	IL	Northchase	NEAR WEST SIDE	28.0	-87.636554	(41.945505, -87.636554)
(41.95544, -87.634006) 1	936	175754	120 N Green St	Residential	1996-03-01	60607.0	IL	Cambridge	NEAR WEST SIDE	28.0	-87.635554	(41.945505, -87.635554)
(41.96422, -87.66098) 1	1785	260135	309 W 23RD ST	Residential	2017-04-01	60616.0	IL	Big Foot Prairie	NEAR WEST SIDE	28.0	-87.635554	(41.945505, -87.635554)

Facet by choice counts

14 choices. Sort by: name count Cluster

- 12.0 1
- 22.0 1
- 24.0 1
- 26.0 3
- 3.0 1
- 32.0 1
- 33.0 4
- 34.0 4
- 35.0 4
- 40.0 1
- 42.0 2
- n.a. 2

Facet by choice counts

1 choice. Sort by: name count Cluster

NEAR WEST SIDE 3

Facet by choice counts

Detect Zip as empty, set Zip to State, City common Zip, as setting to the most common Zip does not violate the constraints of the previous fixes.

OpenRefine building v9.csv remains

Facet / Filter Undo / Redo 0 / 0 Extensions Wikibase Help

46 matching rows (1787 total)

State	Building_ID	Address	Building_Type	Build_Year	Zip	State	City	Community_Name	Community_Number	Latitude	Longitude	Location
1 choice. Sort by: name count Cluster	80	101755	115 S LaSalle St	Commercial	1914-03-01	60603.0	IL	Goodland	LOOP	32.0	-87.970306	(41.970306, -87.95058)
IL 46	123	101940	5 S Wabash Ave	Commercial	1957-03-01	60603.0	IL	Goodland	LOOP	32.0	-87.88186	(41.88186, -87.62567)
Facet by choice counts	125	101961	30 W Monroe	Commercial	2017-03-01	60603.0	IL	Goodland	LOOP	31.0	-87.88109	(41.88109, -87.62558)
131	101987	20 S Clark St	Commercial	1996-03-01	60603.0	IL	Goodland	LOOP	32.0	-87.7398	(41.7398, -87.65349)	
191	103647	135 S LaSalle St	Commercial	1936-03-01	60603.0	IL	Goodland	LOOP	32.0	-87.87972	(41.87972, -87.63155)	
198	103687	1 S State St	Commercial	1907-03-01	60603.0	IL	Goodland	LOOP	32.0	-87.81013	(41.81013, -87.64727)	
201	103716	17 E Monroe St	Residential	1921-03-01	60603.0	IL	Goodland	LOOP	32.0	-87.88032	(41.88032, -87.62695)	
492	134783	37 S Wabash Ave	Commercial	1957-03-01	60603.0	IL	Goodland	LOOP	32.0	-87.88102	(41.88102, -87.62564)	
531	147895	2472 W FOSTER	Residential	2017-03-01	60625.0	IL	Goodland	LINCOLN SQUARE	1.0	-87.97284	(41.97284, -87.70235)	
567	157767	3436 NW Wilson	Municipal	1957-03-01	60625.0	IL	Goodland	ALBANY PARK	14.0	-87.95501	(41.95501, -87.714516)	
577	158121	10 S Dearborn St	Commercial	1921-03-01	60603.0	IL	Goodland	LOOP	32.0	-87.88164	(41.88164, -87.630104)	
581	158184	33 W Monroe St	Commercial	1936-03-01	60603.0	IL	Goodland	LOOP	32.0	-87.88038	(41.88038, -87.62874)	
592	158402	172 W Adams St	Residential	1970-03-01	60603.0	IL	Goodland	LOOP	32.0	-87.87913	(41.87913, -87.63224)	
602	158472	8 W Monroe St	Residential	1957-03-01	60603.0	IL	Goodland	LOOP	32.0	-87.88096	(41.88096, -87.62778)	
678	160196	111 S Michigan Ave	Commercial	1925-03-01	60603.0	IL	Goodland	LOOP	8.0	-87.88045	(41.88045, -87.62423)	
746	164680	5400 N Louis Ave	Commercial	1957-03-01	60625.0	IL	Goodland	NORTH PARK	13.0	-87.97805	(41.97805, -87.71769)	
752	165264	4740 N Western Ave	Commercial	2014-03-01	60625.0	IL	Goodland	LINCOLN SQUARE	4.0	-87.88028	(41.968117, -87.62328)	
894	181640	30 32 E ADAMS ST	Commercial	1985-03-01	60603.0	IL	Goodland	LOOP	32.0	-87.87654	(41.87654, -87.62649)	
1024	251841	100 N Dearborn St	Residential	2017-04-01	60625.0	IL	Goodland	LOOP	32.0	-87.88083	(41.88083, -87.62501)	
1227	251939	4540 N Campbell Ave	Municipal	2012-04-01	60625.0	IL	Goodland	LINCOLN SQUARE	4.0	-87.964208	(41.964208, -87.61190)	
1290	251506	5600 N O'RISTIANA	Municipal	1914-03-01	60625.0	IL	Goodland	NORTH PARK	13.0	-87.92298	(41.92298, -87.692375)	
1375	251746	5501 N Kedzie Ave	Municipal	2014-03-01	60625.0	IL	Goodland	NORTH PARK	13.0	-87.889732	(41.889732, -87.63912)	
1453	252068	22 W Monroe St	Residential	1907-03-01	60603.0	IL	Goodland	LOOP	32.0	-87.86025	(41.86025, -87.6205)	
1515	254001	4434 4441 N Damen Ave	Residential	1917-03-01	60625.0	IL	Goodland	NORTH CENTER	5.0	-87.834167	(41.834167, -87.62294)	
1555	254157	5114 N Christiana Ave	Commercial	2017-04-01	60625.0	IL	Goodland	NORTH CENTER	5.0	-87.88425	(41.88425, -87.62245)	
1556	254159	3205 W Cammen Ave	Commercial	1925-03-01	60625.0	IL	Goodland	WEST TOWN	24.0	-87.88425	(41.88425, -87.623245)	
1636	255927	4447 N Kedzie Ave	Residential	1914-03-01	60625.0	IL	Goodland	ALBANY PARK	14.0	-87.930915	(41.930915, -87.76908)	
1696	256947	2356 East 70th Street	Residential	1969-04-01	IL	Goodland	LOOP	43.0	-87.88425	(41.88425, -87.63245)		
1697	256953	3212 West Division Ave	Residential	1957-04-01	IL	Goodland	LOOP	43.0	-87.88425	(41.88425, -87.63245)		
1698	256956	2800 North Kedzie Ave	Residential	1957-04-01	IL	Goodland	LOOP	43.0	-87.88425	(41.88425, -87.63245)		
1699	256967	6547 N Harlem Avenue	Residential	1936-04-01	IL	Goodland	LOOP	43.0	-87.88425	(41.88425, -87.63245)		
1700	256968	900-10 W Winona Street	Residential	1925-04-01	IL	Goodland	LOOP	25.0	-87.88425	(41.88425, -87.63245)		
1701	256969	4011 N Francisco Avenue	Residential	1996-04-01	IL	Goodland	LOOP	25.0	-87.88425	(41.88425, -87.63245)		
1702	256971	1061 W 16th Street	Residential	1925-04-01	IL	Goodland	LOOP	25.0	-87.88425	(41.88425, -87.63245)		
1703	256975	1750 N Wolcott Avenue	Residential	1936-04-01	IL	Goodland	LOOP	43.0	-87.88425	(41.88425, -87.63245)		
1705	256977	1220 W Jackson Blvd	Residential	1970-04-01	IL	Goodland	WASHINGTON HEIGHTS	73.0	-87.88425	(41.88425, -87.63245)		
1733	256938	5841 S Maryland Ave MC0985	Commercial	1969-04-01	IL	Goodland	NORTH CENTER	5.0	-87.88425	(41.88425, -87.63245)		
1734	256842	2017 Mendell St	Commercial	1925-04-01	IL	Goodland	LOOP	22.0	-87.88425	(41.88425, -87.63245)		
1738	256845	4900 N Bernard	Commercial	1996-04-01	IL	Goodland	LOOP	22.0	-87.88425	(41.88425, -87.63245)		
1738	256856	2201 Lakeside Blvd	Residential	1970-04-01	IL	Goodland	EAST GARFIELD PARK	27.0	-87.88425	(41.88425, -87.63245)		
1764	260127	18 28 S Michigan Ave	Commercial	1900-04-01	60603.0	IL	Goodland	LOOP	32.0	-87.76739	(41.76739, -87.56693)	
1797	260128	125 S Clark St	Commercial	1900-04-01	60603.0	IL	Goodland	LOOP	22.0	-87.767005	(41.767005, -87.567005)	

Facet by choice counts

2 choices. Sort by: name count Cluster

- 6003.0 1
- 60025.0 12
- (blank) 13

Facet by choice counts

Fix data after Zip column is empty:

OpenRefine building V9 csv Permalink

46 matching rows (1787 total)

Show as: rows records Show: 5 10 25 50 100 500 1000 rows

	All	Building_ID	Address	Building_Type	Build_Year	Zip	State	City	Community_Name	Community_Number	Latitude	Longitude	Location
44	101507	5148 N KEDZIE AVE	Commercial	1985-03-01	60625.0	IL	Goodland	NORTH PARK	13.0	41.97508	-87.7087	(41.97508, -87.7087)	
55	101717	55 E Monroe St	Commercial	2010-03-01	60603.0	IL	Goodland	LOOP	32.0	41.880177	-87.62564	(41.880177, -87.62564)	
69	101720	131 S Dearborn	Commercial	1907-03-01	60603.0	IL	Goodland	LOOP	32.0	41.87981	-87.62855	(41.87981, -87.62855)	
77	101739	1 S Dearborn	Commercial	2017-03-01	60603.0	IL	Goodland	LOOP	32.0	41.881653	-87.628845	(41.881653, -87.628845)	
80	101755	115 S LaSalle St	Commercial	1914-03-01	60603.0	IL	Goodland	LOOP	32.0	41.970306	-87.65058	(41.970306, -87.65058)	
123	101940	5 S Wabash Ave	Commercial	1957-03-01	60603.0	IL	Goodland	LOOP	32.0	41.88186	-87.62567	(41.88186, -87.62567)	
125	101961	30 W Monroe	Commercial	2017-03-01	60603.0	IL	Goodland	LOOP	31.0	41.88109	-87.620586	(41.88109, -87.620586)	
131	101987	20 S Clark St	Commercial	1996-03-01	60603.0	IL	Goodland	LOOP	32.0	41.7398	-87.653495	(41.7398, -87.653495)	
191	103647	135 S LaSalle St	Commercial	1936-03-01	60603.0	IL	Goodland	LOOP	32.0	41.879772	-87.63155	(41.879772, -87.63155)	
198	103687	1 S State St	Commercial	1907-03-01	60603.0	IL	Goodland	LOOP	32.0	41.810013	-87.64727	(41.810013, -87.64727)	
201	103716	17 E Monroe St	Residential	1921-03-01	60603.0	IL	Goodland	LOOP	32.0	41.88032	-87.62695	(41.88032, -87.62695)	
492	134783	37 S Wabash Ave	Commercial	1957-03-01	60603.0	IL	Goodland	LOOP	32.0	41.881012	-87.62564	(41.881012, -87.62564)	
531	147895	2472 W FOSTER	Residential	2017-03-01	60625.0	IL	Goodland	LINCOLN SQUARE	1.0	41.972084	-87.70235	(41.972084, -87.70235)	
567	157767	3436 W Wilson Ave	Municipal	1957-03-01	60625.0	IL	Goodland	LINCOLN PARK	14.0	41.96501	-87.714516	(41.96501, -87.714516)	
577	158121	10 S Dearborn St	Commercial	1921-03-01	60603.0	IL	Goodland	LOOP	32.0	41.88164	-87.630104	(41.88164, -87.630104)	
581	158184	33 W Monroe St	Commercial	1936-03-01	60603.0	IL	Goodland	LOOP	32.0	41.880398	-87.628784	(41.880398, -87.628784)	
592	158402	172 W Adams St	Residential	1970-03-01	60603.0	IL	Goodland	LOOP	32.0	41.879913	-87.633224	(41.879913, -87.633224)	
602	158472	8 W Monroe St	Residential	1957-03-01	60603.0	IL	Goodland	LOOP	32.0	41.88096	-87.6278	(41.88096, -87.6278)	
678	160196	111 S Michigan Ave	Commercial	1925-03-01	60603.0	IL	Goodland	LOOP	8.0	41.88045	-87.62423	(41.88045, -87.62423)	
748	164980	5400 N St Louis Ave	Commercial	1957-03-01	60625.0	IL	Goodland	NORTH PARK	13.0	41.979805	-87.71769	(41.979805, -87.71769)	
752	165664	4740 N Western Ave	Commercial	2012-03-01	60625.0	IL	Goodland	LINCOLN SQUARE	4.0	41.968117	-87.68928	(41.968117, -87.68928)	
967	181940	30 32 E ADAMS ST	Commercial	1985-03-01	60603.0	IL	Goodland	LOOP	32.0	41.87954	-87.62649	(41.87954, -87.62649)	
1204	251341	60 E Monroe St	Residential	2017-04-01	60603.0	IL	Goodland	LOOP	32.0	41.88082	-87.62572	(41.88082, -87.62572)	
1227	251393	4540 N Campbell Ave	Municipal	2012-04-01	60625.0	IL	Goodland	LINCOLN SQUARE	4.0	41.964225	-87.69196	(41.964225, -87.69196)	
1280	251506	5500 N CHRISTIANA	Municipal	1914-04-01	60625.0	IL	Goodland	NORTH PARK	13.0	41.92366	-87.693275	(41.92366, -87.693275)	
1376	251746	5501 N Kedzie Ave	Municipal	2014-04-01	60625.0	IL	Goodland	NORTH PARK	13.0	41.889732	-87.63612	(41.889732, -87.63612)	
1453	252068	22 W Monroe St	Residential	1907-04-01	60603.0	IL	Goodland	LOOP	32.0	41.869255	-87.62625	(41.869255, -87.62625)	
1515	254001	4434 4444 N Damen Ave	Residential	1917-04-01	60625.0	IL	Goodland	NORTH CENTER	5.0	41.834167	-87.62294	(41.834167, -87.62294)	
1558	254157	5114 N Christiana Ave	Commercial	2017-04-01	60625.0	IL	Goodland	NORTH CENTER	5.0	41.88425	-87.62425	(41.88425, -87.62425)	
1558	254159	3205 W Carmen Ave	Commercial	1925-04-01	60625.0	IL	Goodland	WEST TOWN	24.0	41.88425	-87.63245	(41.88425, -87.63245)	
1636	255927	4447 N Kedzie Ave	Residential	1914-04-01	60625.0	IL	Goodland	ALBANY PARK	14.0	41.938915	-87.76808	(41.938915, -87.76808)	
1696	256047	2356 East 70th Street	Residential	1969-04-01	60603.0	IL	Goodland	LOOP	43.0	41.88425	-87.62425	(41.88425, -87.62425)	
1697	256653	3212 West Diversey Ave	Residential	1957-04-01	60603.0	IL	Goodland	IL	44.9897	-87.63245	(41.88425, -87.62425)		
1698	256656	2600 North Kedzie Ave	Residential	1957-04-01	60603.0	IL	Goodland	LOOP	44.9897	-87.63245	(41.88425, -87.62425)		
1699	256667	6547 N Harlem Avenue	Residential	1935-04-01	60603.0	IL	Goodland	IL	44.9897	-87.63245	(41.88425, -87.62425)		
1700	256668	9010-W Winona Street	Residential	1925-04-01	60603.0	IL	Goodland	IL	44.9897	-87.63245	(41.88425, -87.62425)		
1701	256699	4011 N Francisco Avenue	Residential	1996-04-01	60603.0	IL	Goodland	IL	42.01036	-87.6683	(42.01036, -87.6683)		
1702	256761	1061 W 16th Street	Residential	1925-04-01	60603.0	IL	Goodland	IL	41.88425	-87.63245	(41.88425, -87.63245)		
1703	256675	1750 N Wolcott Avenue	Residential	1936-04-01	60603.0	IL	Goodland	IL	41.88425	-87.63245	(41.88425, -87.63245)		

Completes the case where Community_Name is not null and Community_Number is null, and Community_Number is set to a value common to Community_Name:

OpenRefine building V9 csv Permalink

31 matching rows (1787 total)

Edit single cell on row 253, column Community_Number

Show as: rows records Show: 5 10 25 50 100 500 1000 rows

	All	Building_ID	Address	Building_Type	Build_Year	Zip	State	City	Community_Name	Community_Number	Latitude	Longitude	Location
253	105478	6230 6238 N KENMORE AVE	Residential	1925-03-01	60600.0	IL	Northgate	EDGEWATER	77.0	41.990902	-87.66861	(41.990902, -87.66861)	
277	107044	2112 1/2 N DELAWARE PL	Residential	1965-03-01	60601.0	VA	Northgate	NEAR NORTH SIDE	77.0	41.895794	-87.62249	(41.895794, -87.62249)	
543	107670	2102 1/2 N DELAWARE PL	Commercial	1920-03-01	60612.0	IL	Northgate	NEAR NORTH SIDE	77.0	41.89053	-87.62193	(41.89053, -87.62193)	
548	159883	1825 W Jackson Blvd	Commercial	1870-03-01	60612.0	IL	Northgate	NEAR NORTH SIDE	77.0	41.88945	-87.62045	(41.88945, -87.62045)	
558	157159	1109 S Canal St	Commercial	1925-03-01	60607.0	IL	Cambridge	NEAR WEST SIDE	77.0	41.8897	-87.63985	(41.8897, -87.63985)	
562	157280	1185 N MILWAUKEE AVE	Residential	1993-03-01	60642.0	NC	Northgate	WEST TOWN	77.0	41.913208	-87.681404	(41.913208, -87.681404)	
592	158402	1721 W Adams St	Residential	1976-03-01	60603.0	IL	Goodland	IL	41.879193	-87.63224	(41.879193, -87.63224)		
598	159807	5650 N HIRSHMAN RD	Residential	1970-03-01	60600.0	IL	Northgate	IL	41.89449	-87.65573	(41.89449, -87.65573)		
840	173754	422 424 N MELROSE ST	Residential	2017-03-01	60657.0	MN	Russia	LAKEVIEW	77.0	41.93985	-87.65945	(41.93985, -87.65945)	
948	175845	2451 W HOWARD ST	Residential	1970-03-01	60645.0	IA	Milng	LAKEVIEW	77.0	42.0089	-87.69503	(42.0089, -87.69503)	
985	232062	1717 S Prairie Ave	Residential	1966-04-01	60616.0	IL	Big Foot Prairie	LAKEVIEW	77.0	41.858475	-87.61985	(41.858475, -87.61985)	
126	250114	116 W Elm St	Municipal	1970-04-01	60610.0	VA	Ebony	LAKEVIEW	77.0	41.903164	-87.61383	(41.903164, -87.61383)	
1491	252318	7800 N Bowerton	Residential	1970-04-01	60626.0	IL	Rockbridge	LAKEVIEW	77.0	42.01945	-87.66651	(42.01945, -87.66651)	
1605	254993	400 W HIRON ST	Residential	1965-04-01	60654.0	VA	Sale Creek	LAKEVIEW	77.0	41.88474	-87.63874	(41.88474, -87.63874)	
1658	256426	35 W 33rd Street	Commercial	1965-04-01	60616.0	IL	Big Foot Prairie	LAKEVIEW	77.0	41.842323	-87.62715	(41.842323, -87.62715)	
1694	256839	619 S LaSalle	Commercial	2012-04-01	60605.0	IL	Leesburg	LAKEVIEW	77.0	41.86949	-87.625854	(41.86949, -87.625854)	
1697	256653	3212 West Diversey Ave	Residential	1957-04-01	60603.0	IL	Goodland	IL	41.88425	-87.63245	(41.88425, -87.63245)		
1698	256656	2600 North Kedzie Ave	Residential	1957-04-01	60603.0	IL	Goodland	IL	41.88425	-87.63245	(41.88425, -87.63245)		
1699	256667	6547 N Harlem Avenue	Residential	1936-04-01	60603.0	IL	Goodland	IL	41.88425	-87.63245	(41.88425, -87.63245)		
1701	256669	4011 N Francisco Avenue	Residential	1996-04-01	60603.0	IL	Goodland	IL	42.01036	-87.6683	(42.01036, -87.6683)		
1703	256675	1750 N Wolcott Avenue	Residential	1936-04-01	60603.0	IL	Goodland	IL	41.88425	-87.63245	(41.88425, -87.63245)		
1704	256679	1511-1525 W Cullom Ave	Residential	1921-04-01	60613.0	VA	Ebony	IL	41.954025	-87.670265	(41.954025, -87.670265)		
1707	256729	6018 S Karlov Avenue	Commercial	1921-04-01	60629.0	IL	Rockbridge	IL	41.782684	-87.69868	(41.782684, -87.69868)		
1709	256731	6448 S Tripp Avenue	Commercial	2010-04-01	60629.0	IL	Rockbridge	IL	41.782684	-87.69868	(41.782684, -87.69868)		
1709	256733	5329 S Oak Park Avenue	Commercial	1921-04-01	60656.0	NE	Cambridge	IL	41.975525	-87.8018	(41.975525, -87.8018)		
1710	256746	6835 N Normal Blvd	Commercial	1921-04-01	60621.0	FL	Northchase	IL	41.769844	-87.64316	(41.769844, -87.64316)		
1728	256827	6950 N Hawley Avenue	Commercial	1957-04-01	60646.0	IL	Rockbridge	IL	41.99103	-87.74164	(41.99103, -87.74164)		
1728	256829	368 W Superior	Residential	1997-04-01	60654.0	VA	Sale Creek	IL	41.892376	-87.63491	(41.892376, -87.63491)		
1730	256831	2111 S Wabash Ave	Residential	1996-04-01	60654.0	VA	Sale Creek	IL	41.809326	-87.59248	(41.809326, -87.59248)		
1730	256835	1744 Western Ave	Residential	1990-04-01	60645.0	IA	Mingo	IL	42.012226	-87.69301	(42.012226, -87.69301)		
1735	256845	4900 N Bernard	Commercial	1996-04-01	60603.0	IL	Goodland	IL	41.792213	-87.59995	(41.792213, -87.59995)		

Completes the case where Community_Number is not null, Community_Name is null, and Community_Name is set to a value common to Community_Name:

OpenRefine building V9 csv [Permalink](#)

Facet / Filter Undo / Redo 9 / 9

Show as: rows records Show: 5 10 25 50 100 500 1000 rows

30 matching rows (1787 total)

Community_Name Building_ID Address Building_Type Build_Year Zip State City Community_Name Community_Number Latitude Longitude Location

514	137570	2145 W Grand Ave	Commercial	1970-03-01	60010.0	IL	Northchase	41.88093	-87.88113	(41.88093, -87.88113)
548	156993	1945 W Jackson Blvd	Commercial	1970-03-01	60010.0	IL	Northchase	41.87727	-87.69845	(41.87727, -87.69845)
592	158402	172 W Adams St	Residential	1970-03-01	60030.0	IL	Goodland	41.879913	-87.63224	(41.879913, -87.63224)
659	156907	6650 N SHERIDAN RD	Residential	1970-03-01	60060.0	IL	Northgate	41.88449	-87.65473	(41.88449, -87.65473)
948	175845	2451 W HOWARD ST	Residential	1970-03-01	60045.0	IA	Mingo	42.0088	-87.89503	(42.0088, -87.89503)
985	232062	1717 S Prairie Ave	Residential	1966-04-01	60016.0	IL	Big Foot Prairie	41.858475	-87.81085	(41.858475, -87.81085)
1128	252014	116 W Elm St	Municipal	1970-04-01	60010.0	VA	Ebony	41.903184	-87.53183	(41.903184, -87.53183)
1491	252318	7600 N Roswell	Residential	1970-04-01	60020.0	IL	Rockbridge	42.01945	-87.66951	(42.01945, -87.66951)
1805	254916	400 W HURON ST	Residential	1968-04-01	60054.0	VA	Sale Creek	41.89474	-87.63874	(41.89474, -87.63874)
1858	258421	35 W 3rd Street	Commercial	1968-04-01	60016.0	IL	Big Foot Prairie	41.84223	-87.62715	(41.84223, -87.62715)
1894	256639	619 S LaSalle	Commercial	2012-04-01	60065.0	IL	Leesburg	41.86949	-87.625854	(41.86949, -87.625854)
1896	256647	2350 East 70th Street	Residential	1968-04-01	60063.0	IL	Goodland	SOUTH SHORE	43.0	(41.88425, -87.63245)
1897	256653	3212 West Diversity Ave	Residential	1957-04-01	60063.0	IL	Goodland			
1898	256656	2600 North Kedzie Ave	Residential	1957-04-01	60063.0	IL	Goodland			
1899	256667	6547 N Harlem Avenue	Residential	1936-04-01	60063.0	IL	Goodland			
1700	256668	900-10 W Winona Street	Residential	1925-04-01	60063.0	IL	Goodland	25.0	43.0	(41.88425, -87.63245)
1701	256669	4011 N Francisco Avenue	Residential	1996-04-01	60063.0	IL	Goodland			
1702	256671	1061 W 16th Street	Residential	1925-04-01	60063.0	IL	Goodland	25.0	43.0	(41.88425, -87.63245)
1703	256675	1750 N Wolcott Avenue	Residential	1936-04-01	60063.0	IL	Goodland			
1704	256679	1511-1625 W Culver Avenue	Residential	1921-04-01	60016.0	VA	Ebony	41.954025	-87.670295	(41.954025, -87.670295)
1707	256729	6018 S Karlov Avenue	Commercial	1921-04-01	60020.0	IL	Rockbridge	41.782684	-87.69868	(41.782684, -87.69868)
1708	256743	6446 S Tripp Avenue	Commercial	2010-04-01	60020.0	IL	Rockbridge	41.782684	-87.69868	(41.782684, -87.69868)
1709	256753	5225 N Oak Park Avenue	Commercial	1921-04-01	60020.0	IL	Rockbridge	41.782684	-87.69868	(41.782684, -87.69868)
1710	256754	6001 S Paulina Blvd	Commercial	1921-04-01	60020.0	IL	Rockbridge	41.782684	-87.69868	(41.782684, -87.69868)
1728	256807	6900 N Hawatha Avenue	Commercial	1937-04-01	60040.0	IL	Rockbridge	41.89103	-87.74184	(41.89103, -87.74184)
1729	256809	395 W Superior	Residential	1937-04-01	60054.0	VA	Sale Creek	41.882076	-87.63491	(41.882076, -87.63491)
1730	256831	2111 S Wakash Avenue	Residential	1996-04-01	60054.0	VA	Sale Creek	41.890326	-87.59248	(41.890326, -87.59248)
1722	256835	1744 Western Avenue	Residential	1990-04-01	60045.0	IA	Mingo	42.012226	-87.69301	(42.012226, -87.69301)
1734	256842	2017 Mendell St.	Residential	1925-04-01	60063.0	IL	Goodland	22.0	43.0	(41.88425, -87.63245)
1735	256845	4900 N Bernard	Commercial	1996-04-01	60063.0	IL	Goodland			

Because we also modified the Community_Number, Community_Name for the null case, will lead to the original repair e5 data will have a violation of the e5 situation, re-fix data.

OpenRefine building V11 csv [Permalink](#)

Facet / Filter Undo / Redo 0 / 0

Show as: rows records Show: 5 10 25 50 100 500 1000 rows

2 matching rows (1787 total)

Location Building_ID Address Building_Type Build_Year Zip State City Community_Name Community_Number Latitude Longitude Location

1554	254156	5025 N Paulina St	Commercial	1969-04-01	60040.0	IL	Northchase	UPTOWN	43.0	41.88425	-87.63245
1696	256647	2356 East 70th Street	Residential	1969-04-01	60063.0	IL	Goodland	UPTOWN	43.0	41.88425	-87.63245

After fixing all the constraint data, import it into VizierDb to view it.

Console ▾ Timing Datasets ▾ Charts ▾ Views

buildings_new (1787 rows)

Building_ID (int)	Address (string)	Building_Type (string)	Build_Year (date)	Zip (float)	State (string)	City (string)	Community_Name (string)	Community_Nr↑
0	100068	555 W Roosevelt Rd	Commercial	2017-02-01	60607	IL	Cambridge	NEAR WEST SIDE
1	100256	4975 S Archer Ave	Municipal	1996-02-01	60632	KY	Mooreland	ARCHER HEIGHTS
2	100320	4039 4051 N LAPORTE AVE	Residential	1936-02-01	60641	OK	Snake Creek	PORTAGE PARK
3	100346	2609 W BELMONT AVE	Residential	1954-02-01	60618	IL	Crisman	AVONDALE
4	100372	2701 W 68th St	Commercial	1957-02-01	60629	IL	Rockbridge	CHICAGO LAWN
5	100393	7601 S Cicero Ave	Commercial	1907-02-01	60652	IL	Northgate	WEST LAWN
6	100396	7740 S Western Ave	Commercial	1957-02-01	60620	IL	Leesburg	ASHBURN
7	100397	3939 W 79th St	Municipal	1966-02-01	60652	IL	Northgate	DOUGLAS
8	100429	875 N Michigan Ave	Commercial	1957-02-01	60611	VA	Northgate	NEAR NORTH SIDE
9	100433	2233 W Division St	Commercial	2017-02-01	60622	IL	Leesburg	WEST TOWN
10	100447	211 E Chicago Ave	Commercial	1936-02-01	60611	VA	Northgate	NEAR NORTH SIDE
11	100448	198 E Delaware Pl	Residential	1914-02-01	60611	VA	Northgate	NEAR NORTH SIDE
12	100470	30 W Chicago	Residential	1900-02-01	60654	VA	Sale Creek	NEAR NORTH SIDE
13	100489	4520 4528 S DREXEL BLVD	Residential	2017-02-01	60653	VA	Cambridge	KENWOOD
14	100507	4950 4952 S PRAIRIE AVE	Residential	1936-02-01	60615	VA	Crisman	GRAND BOULEVARD
15	100533	6201 S Stewart Ave	Municipal	1954-02-01	60621	FL	Northchase	ENGLEWOOD
16	100572	8131 S May St	Commercial	1957-02-01	60620	IL	Leesburg	AUBURN GRESHAM
17	100623	1345 1359 W ESTES AVE	Residential	1957-02-01	60626	IL	Rockbridge	ROGERS PARK
18	100637	510 534 W ADDISON ST	Residential	1900-02-01	60613	VA	Ebony	LAKE VIEW
19	100668	3245 W ARTHINGTON ST	Commercial	1914-02-01	60624	OH	Northgate	EAST GARFIELD PARK

[19]

```
SELECT Building_ID,COUNT(DISTINCT Address) AS num_addresses
FROM buildings new
GROUP BY Building_ID
HAVING COUNT(DISTINCT Address) > 1;
```

Console ▾ Timing Datasets ▾ Charts ▾ Views

Recheck e0 constraints.

19 100668 3245 W ARTHINGTON ST Commercial 1914-02-01 60624 OH Northgate EAST GARFIELD PARK 27

[19]

```
SELECT Building_ID,COUNT(DISTINCT Address) AS num_addresses
FROM buildings_new
GROUP BY Building_ID
HAVING COUNT(DISTINCT Address) > 1;
```

e0_check_v2 (0 rows)

Building_ID (int)	num_addresses (long)

[20]

```
SELECT a.Building_ID, a.Address AS Address1, b.Address AS Address2
FROM buildings_new a, buildings_new b
WHERE a.Building_ID = b.Building_ID AND a.Address <> b.Address;
```

e0_pairs_check_v2 (0 rows)

Building_ID (int)	Address1 (string)	Address2 (string)

Recheck e1 constraints.

[21]

```
SELECT Zip,COUNT(DISTINCT State) AS num_states
FROM buildings_new
GROUP BY Zip
HAVING COUNT(DISTINCT State) > 1;
```

e1_check_v1 (0 rows)

Zip	num_states

[22]

```
SELECT a.Zip, a.State AS state1, b.State AS state2
FROM buildings_new a, buildings_new b
WHERE a.Zip = b.Zip AND a.State <> b.State;
```

e1_pairs_check_v1 (0 rows)

Zip	state1	state2

Recheck e2 constraints.

[23]

```
SELECT Zip,COUNT(DISTINCT City) AS num_cities
FROM buildings_new
GROUP BY Zip
HAVING COUNT(DISTINCT City) > 1;
```

e2_check_v1 (0 rows)

Zip	num_cities

[24]

```
SELECT a.Zip, a.City AS city1, b.City AS city2
FROM buildings_new a, buildings_new b
WHERE a.Zip = b.Zip AND a.City <> b.City;
```

e2_pairs_check_v1 (0 rows)

Zip	city1	city2

Recheck e3 constraints.

[25]

```
SELECT Address, COUNT(DISTINCT Location) AS num_locations
FROM buildings_new
GROUP BY Address
HAVING COUNT(DISTINCT Location) > 1;
```

e3_check_v1 (0 rows)

Address	num_locations

[26]

```
SELECT A1.Address, A1.Location AS location1, A2.Location AS location2
FROM buildings_new A1, buildings_new A2
WHERE A1.Address = A2.Address AND A1.Location <> A2.Location;
```

e3_pairs_check_v1 (0 rows)

Address	location1	location2

Recheck e4 constraints.

```
[27]   SELECT Latitude, Longitude, COUNT(DISTINCT Location) AS num_locations  
  FROM buildings_new  
 GROUP BY Latitude, Longitude  
 HAVING COUNT(DISTINCT Location) > 1;
```

Console ▾ Timing Datasets ▾ Charts ▾

e4_check_v1 (0 rows)			Views
	Latitude (float)	Longitude (float)	num_locations (long)

```
[28]   SELECT A1.Latitude, A1.Longitude, A1.Location AS location1, A2.Location AS location2  
FROM buildings_new A1, buildings_new A2  
WHERE A1.Latitude = A2.Latitude AND A1.Longitude = A2.Longitude AND A1.Location > A2.Location
```

Console ▾ Timing Datasets ▾ Charts ▾

e4_pairs_check_v1 (0 rows)				Views			
	Latitude (m.s)	Longitude (m.s)	location1 (lat,lon)	location2 (lat,lon)			

Recheck e5 constraints.

```
[29]   SELECT Location, Community_Number, COUNT(DISTINCT Community_Name) AS num_community_names  
FROM buildings_new  
GROUP BY Location, Community_Number  
HAVING COUNT(DISTINCT Community_Name) > 1;
```

Console ▾ Timing Datasets ▾ Charts ▾

e5_check_v1 (0 rows) 			Views	Up	Down	Left	Right	
	Location (string)	Community_Number (float)	num_community_names (long)					

```
[30]   SELECT A1.Location, A1.Community_Number, A1.Community_Name AS communityName1, A2.Community_Name AS communityName2  
FROM buildings_new A1, buildings_new A2  
WHERE A1.Location = A2.Location  
AND A1.Community_Number = A2.Community_Number  
AND A1.Community_Name <> A2.Community_Name;
```

Console ▾ Timing Datasets ▾ Charts ▾

Location (string)	Community_Number (float)	communityName1 (string)	communityName2 (string)
-------------------	--------------------------	-------------------------	-------------------------



Data integration

Using Talend to do data integration, and extraction tasks.

source table:

building Table Columns Schema

Column Name	Data Type	Description
Building_ID	INTEGER	The unique identifier for the building
Address	VARCHAR(255)	The street address of the building
Building_Type	VARCHAR(100)	The classification of the building (e.g., residential, commercial)
Build_Year	INTEGER	The year the building was constructed
Zip	VARCHAR(10)	The postal code for the building
State	VARCHAR(2)	The U.S. state abbreviation where the building is located
City	VARCHAR(100)	The city where the building is located
Community_Name	VARCHAR(255)	The name of the community or neighborhood
Community_Number	INTEGER	The numeric identifier for the community
Latitude	DECIMAL(9,6)	The latitude coordinate of the building
Longitude	DECIMAL(9,6)	The longitude coordinate of the building
Location	VARCHAR(255)	The combined latitude and longitude information

source table:

building_energy Table Schema

Column Name	Data Type	Description
Data_Year	YEAR	The year the data was reported
ID	INTEGER	The unique identifier for the property
Property_Name	VARCHAR(255)	The name of the property
Address	VARCHAR(255)	The street address of the property
Energy_Rating	VARCHAR(50)	The energy efficiency rating of the property
Property_type	VARCHAR(100)	The type of property
Floor_Area	DECIMAL(10,2)	The total floor area of the property
Water_Use	DECIMAL(10,3)	The annual water usage in kilogallons
Electricity_Use	DECIMAL(10,3)	The annual electricity usage in kilo British thermal units
Natural_Gas_Use	DECIMAL(10,3)	The annual natural gas usage in kilo British thermal units
Total_GHG_Emissions	DECIMAL(10,2)	The total greenhouse gas emissions in metric tons of CO2 equivalent
GHG_Intensity	DECIMAL(10,4)	The greenhouse gas intensity in kilograms CO2 equivalent per square foot
Latitude	DECIMAL(9,6)	The latitude coordinate of the property
Longitude	DECIMAL(9,6)	The longitude coordinate of the property
Location	POINT	The geographical point location of the property
Wards	INTEGER	The ward number the property is located in
Community_Area	VARCHAR(255)	The name of the community area the property is located in

source table:

building_community Table Schema

Column Name	Data Type	Description
Building_ID	INTEGER	The unique identifier for the building
Community_ID	INTEGER	The unique identifier for the community
Community_Name	VARCHAR(255)	The Name of the community
Description	TEXT	The description of the community
Zip	VARCHAR(10)	Postal code
State	VARCHAR(2)	State abbreviation
Country	VARCHAR(50)	Country name
Area	DECIMAL(10,2)	Total area of the community
Population	INTEGER	Total population of the community
EmploymentRate	DECIMAL(5,2)	Employment rate within the community
CrimeRate	DECIMAL(5,2)	Crime rate within the community
Amenities	TEXT	Amenities available in the community
PublicTransportAccess	DECIMAL(5,2)	Access to public transport
QualityOfLifeIndex	DECIMAL(5,2)	Quality of life index
HealthCareIndex	DECIMAL(5,2)	Health care index
CostOfLivingIndex	DECIMAL(5,2)	Cost of living index
PollutionIndex	DECIMAL(5,2)	Pollution index
ClimateIndex	DECIMAL(5,2)	Climate index

source table:

building_street Table Schema

Column Name	Data Type	Description
Street_ID	INTEGER	The unique identifier for the street
Street_Name	VARCHAR(255)	The name of the street
Description	TEXT	The description of the street
LocationType	VARCHAR(100)	Type of location (e.g., residential, commercial)
Zip	VARCHAR(10)	Postal code for the street
State	VARCHAR(2)	State abbreviation
Country	VARCHAR(50)	Country name
Address	VARCHAR(255)	Complete address
Latitude	DECIMAL(9,6)	Latitude coordinate
Longitude	DECIMAL(9,6)	Longitude coordinate
Location	POINT	Geographical point location
Elevation	DECIMAL(7,2)	Elevation of the street above sea level
GeoPoint	POINT	Geospatial point data
Region	VARCHAR(255)	Region in which the street is located
Population	INTEGER	Population in the street area
LandUse	VARCHAR(100)	Land use type
Accessibility	VARCHAR(100)	Accessibility rating
Amenities	TEXT	Amenities available on the street
SafetyIndex	DECIMAL(5,2)	Safety index for the street
EnvironmentalQuality	DECIMAL(5,2)	Environmental quality index
DevelopmentIndex	DECIMAL(5,2)	Development index for the area
GeospatialData	GEOMETRY	Geospatial data for the street

source table:

community_service Table Schema

Column Name	Data Type	Description
Service_ID	INTEGER	Unique identifier for the service
Community_ID	INTEGER	Identifier for the community the service belongs to
ServiceName	VARCHAR(255)	Name of the service
ServiceType	VARCHAR(100)	Type of service provided
ProviderName	VARCHAR(255)	Name of the service provider
Address	VARCHAR(255)	Address where the service is provided
Contact	VARCHAR(100)	Contact information for the service
OperatingHours	VARCHAR(100)	Hours during which the service operates
Capacity	INTEGER	Capacity of the service (e.g., number of people it can serve)
Specialties	TEXT	Specialties offered by the service
InsuranceAccepted	BOOLEAN	Whether insurance is accepted
QualityRating	DECIMAL(3,2)	Quality rating of the service
Accessibility	VARCHAR(100)	Accessibility information
EmergencyServices	BOOLEAN	Whether emergency services are provided
AverageWaitTime	INTEGER	Average wait time to receive service
PatientSatisfactionScore	DECIMAL(3,2)	Patient satisfaction score
Accreditation	VARCHAR(255)	Accreditation status
Website	VARCHAR(255)	Website for the service

target table:

Building_Covered Table Schema

Column Name	Data Type	Description
Building_ID	INTEGER	The unique identifier for the building
Building_Address	VARCHAR(255)	Address of the building
Building_Type	VARCHAR(100)	Type of the building (e.g., residential, commercial)
Build_Year	INTEGER	Year when the building was constructed
Zip	VARCHAR(10)	Postal code where the building is located
State	VARCHAR(2)	State where the building is located
City	VARCHAR(100)	City where the building is located
Community_ID	INTEGER	Identifier for the community
Community_Name	VARCHAR(255)	Name of the community or neighborhood
Latitude	DECIMAL(9,6)	Latitude coordinate for the building
Longitude	DECIMAL(9,6)	Longitude coordinate for the building
Location	POINT	Geographical location of the building
Energy_Rating	INTEGER	Energy efficiency rating of the building
Property_Type	VARCHAR(100)	Type of property
Floor_Area	DECIMAL	Total floor area of the building
Water_Use	DECIMAL	Annual water usage
Electricity_Use	DECIMAL	Annual electricity usage
Natural_Gas_Use	DECIMAL	Annual natural gas usage
Total_GHG_Emissions	DECIMAL	Total greenhouse gas emissions
GHG_Intensity	DECIMAL	Greenhouse gas emissions intensity
Wards	VARCHAR(50)	Ward information for the building
Community_Area	VARCHAR(100)	Community area where the building is located

target table:

Community Table Schema

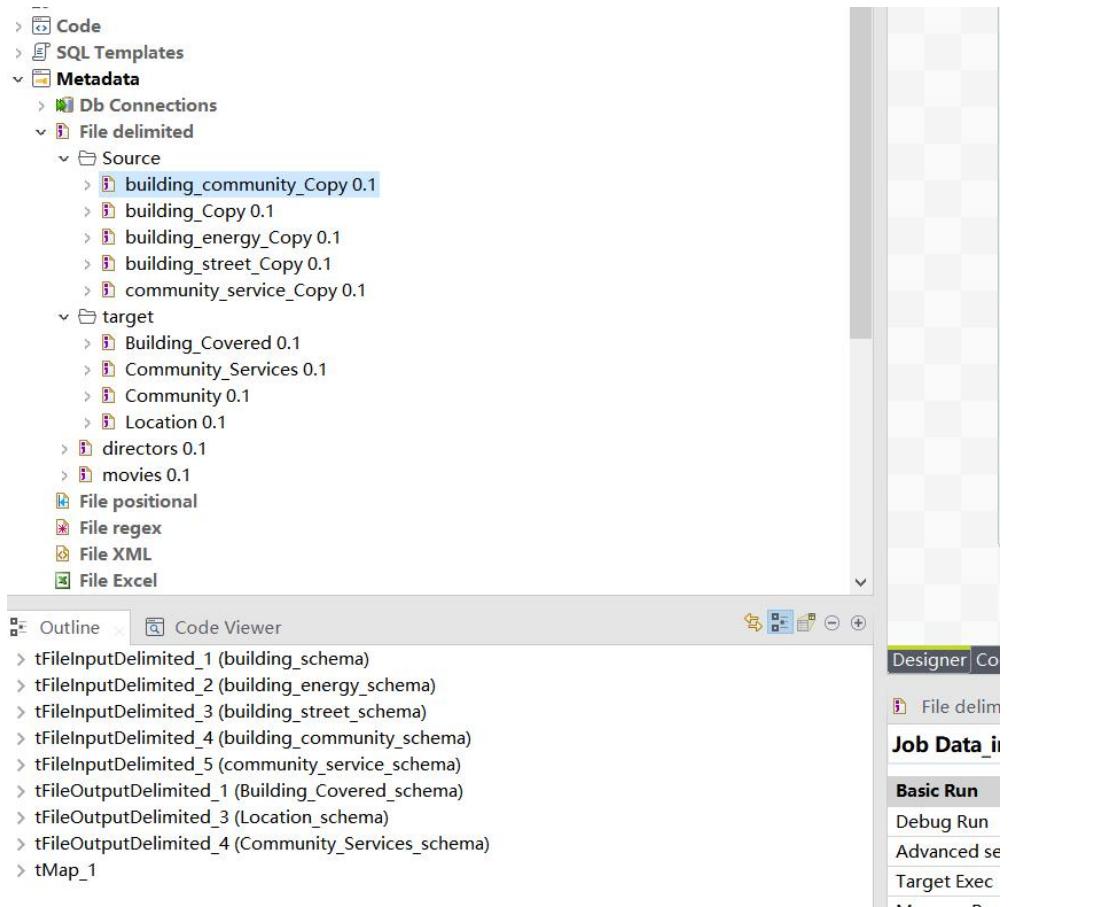
Column Name	Data Type	Description
Community_ID	INTEGER	Unique identifier for the community
Community_Name	VARCHAR(255)	Name of the community or neighborhood
Zip	VARCHAR(10)	Postal code for the community area
State	VARCHAR(2)	State where the community is located
Country	VARCHAR(50)	Country where the community is located
Area	DECIMAL	Total area of the community
Population	INTEGER	Population living in the community
Employment_Rate	DECIMAL(5,2)	Employment rate within the community
Crime_Rate	DECIMAL(5,2)	Crime rate within the community
Amenities	TEXT	Amenities available in the community
Transport	TEXT	Transport options available in the community
Life_Index	DECIMAL(5,2)	Quality of life index
Health_Index	DECIMAL(5,2)	Health care index
Living_Index	DECIMAL(5,2)	Cost of living index
Environment_Index	DECIMAL(5,2)	Environmental quality index
Climate_Index	DECIMAL(5,2)	Climate index for the community

target table:

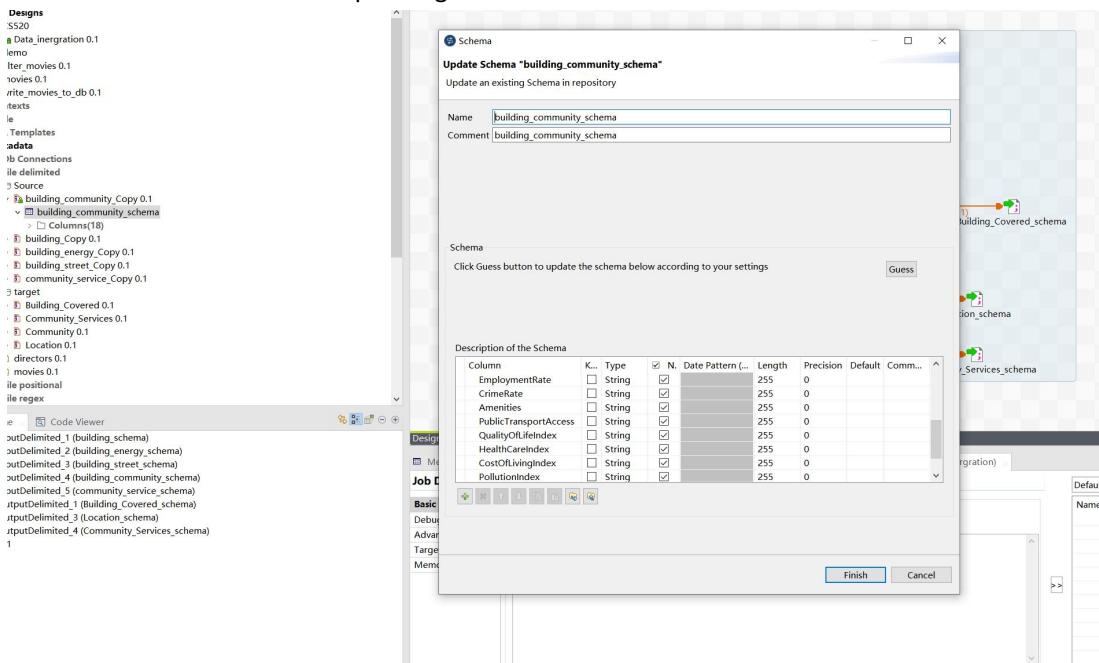
Location Table Schema

Column Name	Data Type	Description
L_ID	INTEGER	Unique identifier for the location
L_Name	VARCHAR(255)	Name of the location
Description	TEXT	Description of the location
L_Type	VARCHAR(50)	Type of the location (e.g., urban, rural)
Zip	VARCHAR(10)	Postal code for the location
State	VARCHAR(2)	State where the location is situated
Country	VARCHAR(50)	Country where the location is situated
L_Address	VARCHAR(255)	Address of the location
Latitude	DECIMAL(9,6)	Latitude coordinate of the location
Longitude	DECIMAL(9,6)	Longitude coordinate of the location
Location	POINT	Geographic point combining latitude and longitude
Elevation	DECIMAL	Elevation above sea level
Region	VARCHAR(100)	Region where the location is situated
Population	INTEGER	Population in the location
Safe_Index	DECIMAL(5,2)	Safety index of the location
Environment_Index	DECIMAL(5,2)	Environmental quality index of the location
Development_Index	DECIMAL(5,2)	Development index of the location
Geo_Data	JSON	Geospatial data in JSON format

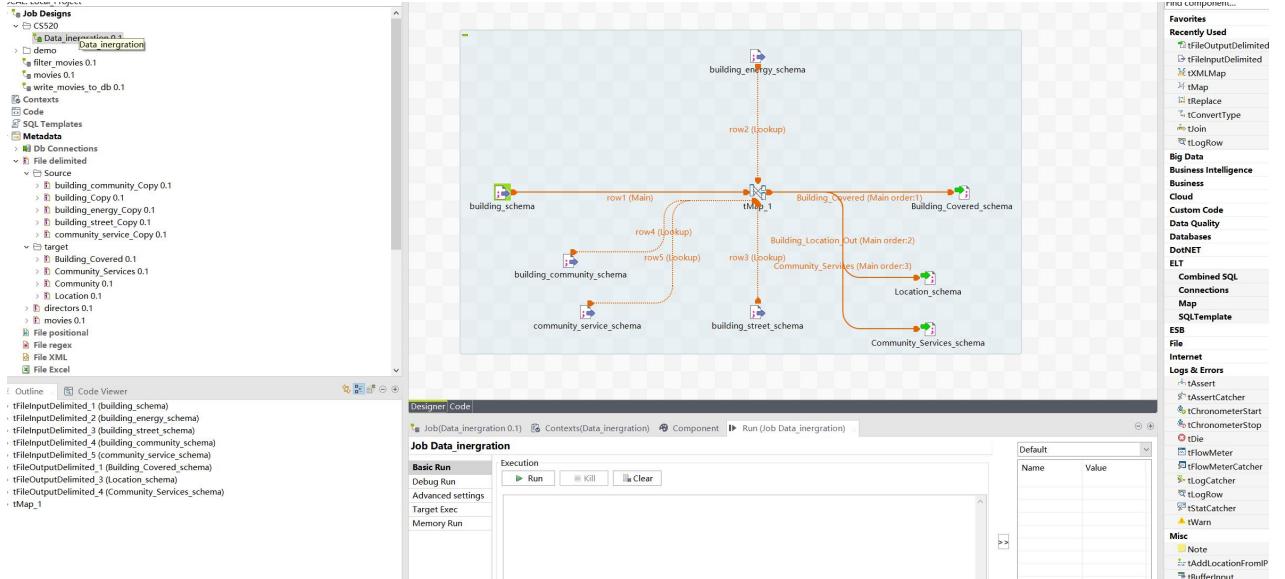
Create source and target files delimited.



Create a schema for the corresponding table



Create Data Integration Tasks.



Creating schema mappings.

Screenshot showing the schema mapping configuration for the "Building_Covered" component.

Schema Editor:

Column	Expression	Type	Length	Precision	Default	Comment
Building_ID	row1.Building_ID	String	255	0		
Address		String	255	0		
Building_Type		String	255	0		
Build_Year		String	255	0		
Zip		String	255	0		
State		String	255	0		
City		String	255	0		
Community_Name		String	255	0		
Community_Number		String	255	0		
Latitude		String	255	0		
Longitude		String	255	0		
Location		String	255	0		

Expression Editor:

Column	Expression	Type	Length	Precision	Default	Comment
Building_ID	row1.Building_ID	String	255	0		
Building_Address	row1.Address	String	255	0		
Building_Type	row1.Building_Type	String	255	0		
Build_Year	row1.Build_Year	String	255	0		
Zip	row1.Zip	String	255	0		
State	row1.State	String	255	0		
City	row1.City	String	255	0		
Community_ID	row1.Community_Number	String	255	0		
Community_Name	row1.Community_Name	String	255	0		
Latitude	row1.Latitude	String	255	0		
Longitude	row1.Longitude	String	255	0		
Location	row1.Location	String	255	0		

Mapping View:

The mapping view shows the relationships between columns from different schemas. Yellow lines indicate the mapping paths between the source columns and the target columns in the "Building_Covered" schema.

Building_Covered Schema Definition:

Column	K... Type	N	Date Pattern (Ctrl+Space)	Length	Precision	Default	Comment
Building_ID	String	1		255	0		
Building_Address	String	2		255	0		
Building_Type	String	3		255	0		
Build_Year	String	4		255	0		
Zip	String	5		255	0		
State	String	6		255	0		
City	String	7		255	0		
Community_ID	String	8		255	0		
Community_Name	String	9		255	0		
Latitude	String	10		255	0		
Longitude	String	11		255	0		

This screenshot shows a data mapping interface with two rows being mapped from a source schema to a target schema.

Source Schema (Left):

- row3:** Expr. key: Column
 - row1.Latitude
 - row1.Longitude
 - row1.Location
- row4:** Expr. key: Column
 - row1.Building_ID
 - row1.Community_Number

Target Schema (Right):

- row2.Property_Type:** Column
 - L_ID
 - L_Name
 - Description
 - L_Type
 - Zip
 - State
 - Country
 - L_Address
 - Latitude
 - Longitude
 - Elevation
 - Region
 - Population
 - Safe_Index
 - Environment_Index
 - Development_Index
 - Geo_Data
- Building_Location_Out:** Expression
 - row3.Street_ID
 - row3.Street_Name
 - row3.Description
 - row3.LocationType
 - row3.Built_Year
 - row3.Zip
 - row3.State
 - row3.Country
 - row3.Address
 - row3.Latitude
 - row3.Longitude
 - row3.Elevation
 - row3.Region
 - row3.Population
 - row3.SafetyIndex
 - row3.EnvironmentalQuality
 - row3.DevelopmentIndex
 - row3.GeospatialData
- Community_Services:** Expression
 - ID
 - Community_ID
 - S_Name
- Building_Covered:** Column
 - Building_ID
 - Building_Address
 - Building_Type
 - Built_Year
 - Zip
 - State
 - City
 - Community_ID
 - Community_Name
 - Latitude
 - Longitude

This screenshot shows a data mapping interface with two rows being mapped from a source schema to a target schema, with many yellow lines indicating mappings between specific fields.

Source Schema (Left):

- row1.assingning_u:** Community_ID, Community_Name, Description, Zip, State, Country, Area, Population, EmploymentRate, CrimeRate, Amenities, PublicTransportAccess, QualityOfLifeIndex, HealthCareIndex, CostOfLivingIndex, PollutionIndex, ClimateIndex
- row5:** Expr. key: Column
 - row4.Community_ID

Target Schema (Right):

- row3.Property_Type:** Column
 - Description
 - L_Type
 - Zip
 - State
 - Country
 - L_Address
 - Latitude
 - Longitude
 - Elevation
 - Region
 - Population
 - Safe_Index
 - Environment_Index
 - Development_Index
 - Geo_Data
- Community_Services:** Expression
 - ID
 - Community_ID
 - S_Name
 - S_Type
 - P_Name
 - S_Address
 - Contact
 - OperatingHours
 - Capacity
 - Specialties
 - InsuranceAccepted
 - EmergencyServices
 - AverageWaitTime
 - Qualification
 - Website
- Building_Covered:** Column
 - Building_ID
 - Building_Address
 - Building_Type
 - Built_Year
 - Zip
 - State
 - City
 - Community_ID
 - Community_Name
 - Latitude
 - Longitude

🔑 Data extraction

Configure the schema matching policy

row1

Column
Building_ID
Address
Building_Type
Build_Year
Zip
State
City
Community_Name
Community_Number
Latitude
Longitude
Location

row2

Property	Value
Lookup Model	Load once
Match Model	Unique match
Join Model	Inner Join
Store temp data	false

Expr. key	Column
row1.Building_ID	Data_Year
	ID
	Property_Name
	Address
	Energy_Rating
	Property_Type
	Floor_Area
	Water_Use_kGal
	Electricity_Use_kBtu
	Natural_Gas_Use_kBtu
	Total_GHG_Emissions_Metric_To...
	GHG_Intensity_kg_CO2e_sq_ft
	Latitude

Schema editor Expression editor

row2

Building_Covered

row1

Column
Electricity_Use_kBtu
Natural_Gas_Use_kBtu
Total_GHG_Emissions_Metric_To...
GHG_Intensity_kg_CO2e_sq_ft
Latitude
Longitude
Location
Wards
Community_Area

row3

Property	Value
Lookup Model	Load once
Match Model	Unique match
Join Model	Inner Join
Store temp data	false

Expr. key	Column
row1.Latitude	Street_ID
row1.Longitude	Street_Name
row1.Location	Description
	LocationType
	Zip
	State
	Country
	Address
	Latitude
	Longitude
	Location
	Elevation
	GeoPoint
	Region
	Population
	LandUse
	Accessibility
	Amenities

Schema editor Expression editor

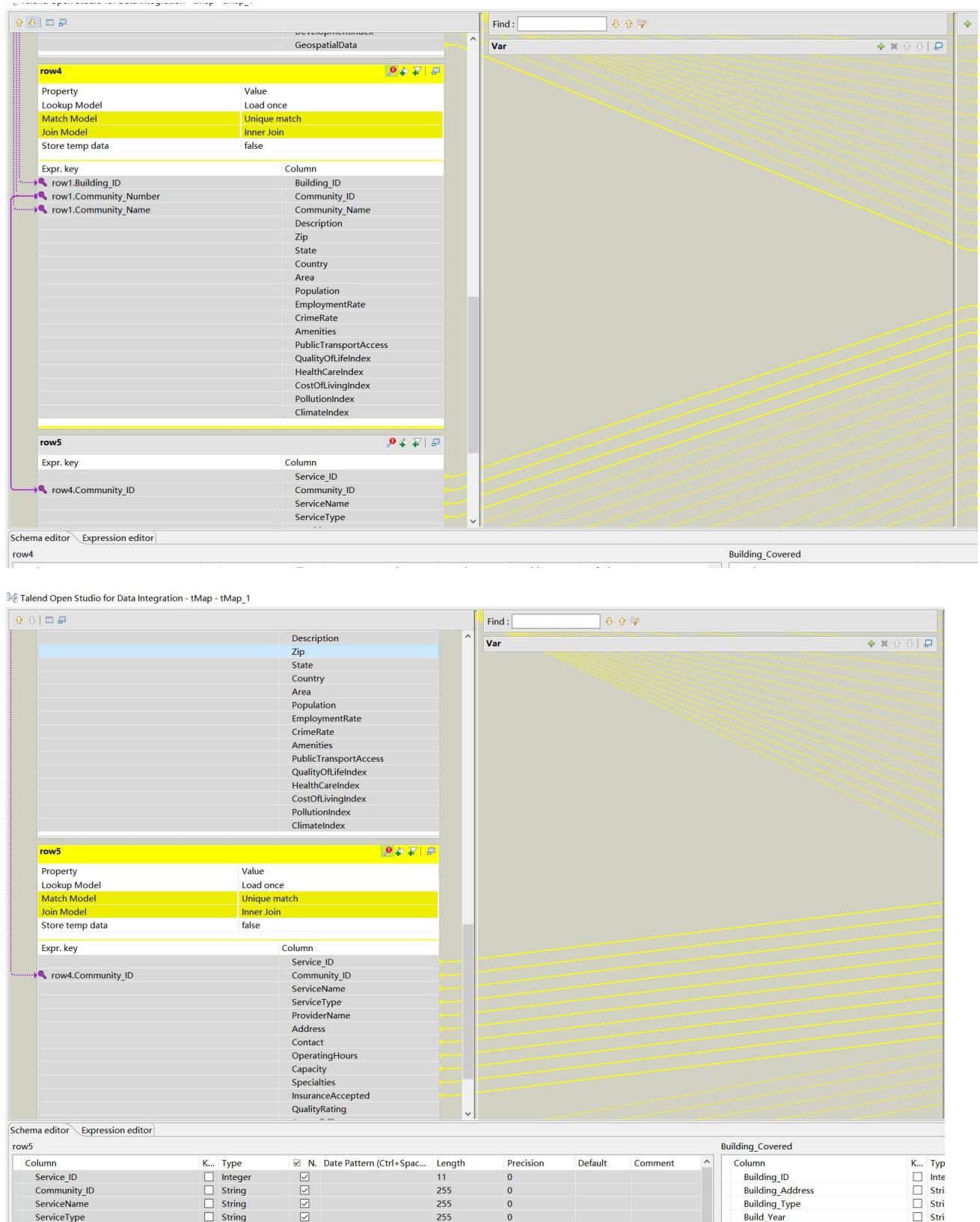
row3

Building_Covered

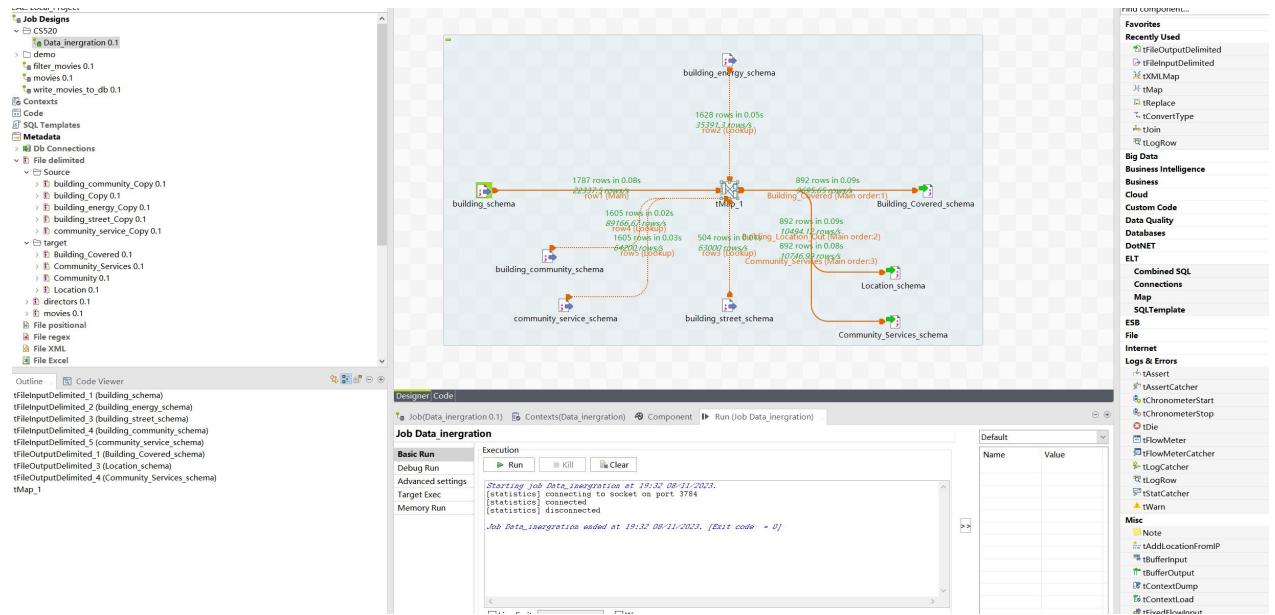
Column	K..	Type	N.	Date Pattern (Ctrl+Space)	Length	Precision	Default	Comment
Street_ID	<input type="checkbox"/>	Integer	<input checked="" type="checkbox"/>		11	0		
Street_Name	<input type="checkbox"/>	String	<input checked="" type="checkbox"/>		255	0		
Description	<input type="checkbox"/>	String	<input checked="" type="checkbox"/>		255	0		
LocationType	<input type="checkbox"/>	String	<input checked="" type="checkbox"/>		255	0		
Zip	<input type="checkbox"/>	String	<input checked="" type="checkbox"/>		255	0		
State	<input type="checkbox"/>	String	<input checked="" type="checkbox"/>		255	0		
Country	<input type="checkbox"/>	String	<input checked="" type="checkbox"/>		255	0		
Address	<input type="checkbox"/>	String	<input checked="" type="checkbox"/>		255	0		
Latitude	<input type="checkbox"/>	String	<input checked="" type="checkbox"/>		255	0		

Building_Covered

Column
Building_ID
Building_Address
Building_Type
Build_Year
Zip
State
City
Community_ID
Community_Name



Running the acquisition task, you can see the exact amount of quantities to be acquired and the amount of data to be output.

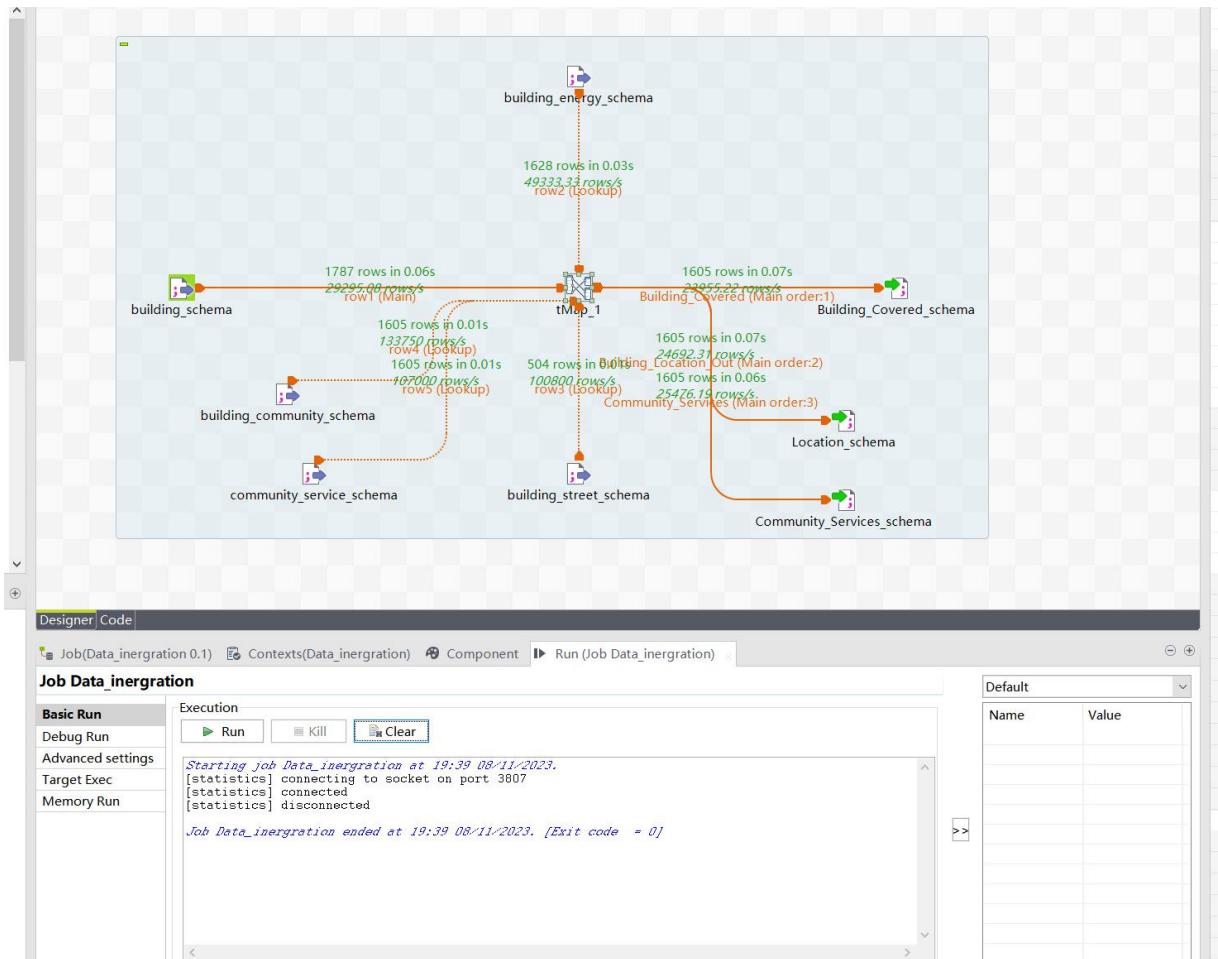


Toggles the matching method:

row1	Var
Column	
Building_ID	
Address	
Building_Type	
Build_Year	
Zip	
State	
City	
Community_Name	
Community_Number	
Latitude	
Longitude	
Location	

row2	Var
Property	Value
Lookup Model	Load once
Match Model	Unique match
Join Model	Left Outer Join
Store temp data	false
Expr. key	Column
row1.Building_ID	Data_Year
	ID
	Property_Name
	Address
	Energy_Rating
	Property_Type
	Floor_Area
	Water_Use_kGal
	Electricity_Use_kBtu
	Natural_Gas_Use_kBtu
	Total_GHG_Emissions_Metric_To...
	GHG_Intensity_kg_CO2e_sq_ft

Changes in the amount of data after switching the matching method:



the resulting csv file:

Building_Covered.csv	2023/11/8 19:39	XLS 工作表	379 KB
Community_Service.csv	2023/11/8 19:39	XLS 工作表	407 KB
Location.csv	2023/11/8 19:39	XLS 工作表	236 KB