



Coder's Corner: Using the Power of Proc Geocode to enhance your analytics

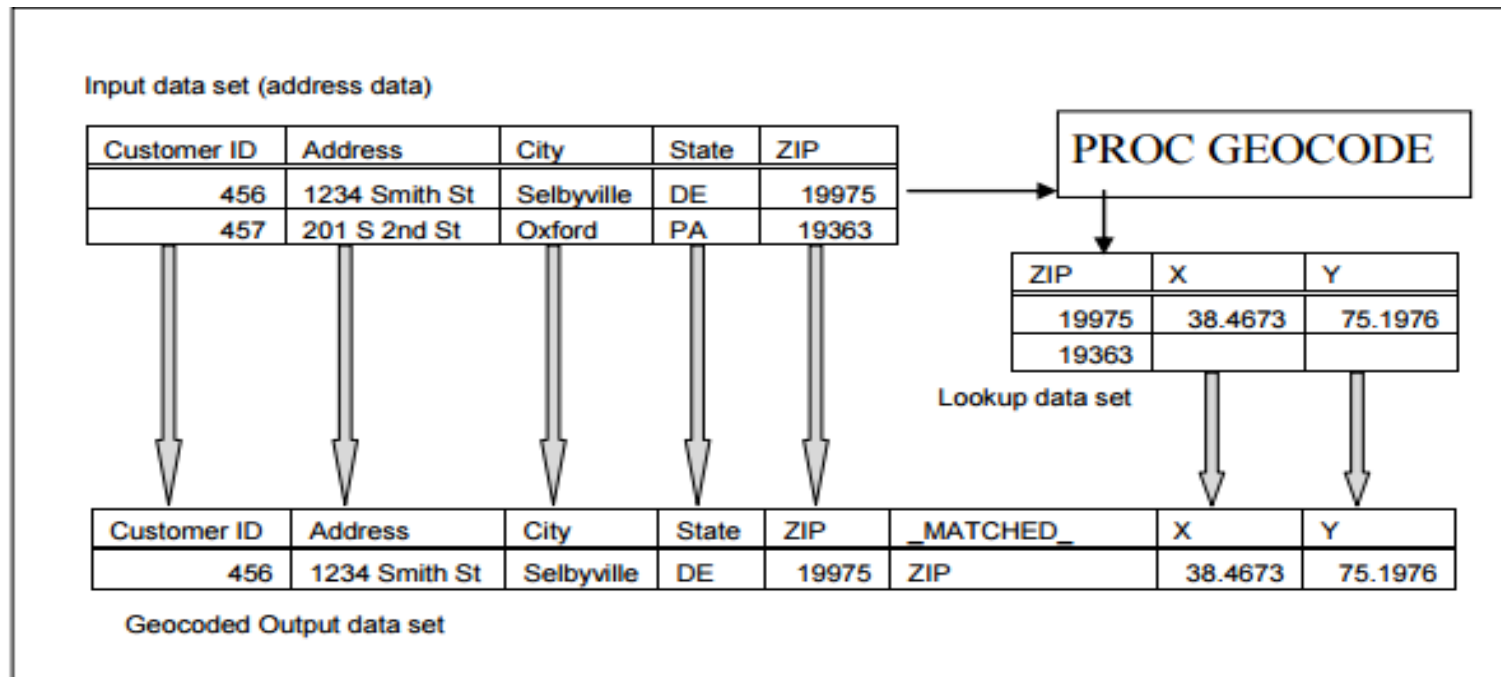
Kieran Shah

with the help of a number of SAS
bloggers, SAS gurus, and just SAS
addicts

March 2016



How PROC GEOCODE Works*



US-wide Lookup dataset: usm.sas7bdat

US-wide Geocoded Output dataset:

* Massengill, Darrell and Ed Odom. PROC GEOCODE: Now with Street-Level Geocoding."

SAS Global Forum 2010 paper and example source code download. SAS Institute Inc.

<http://support.sas.com/rnd/papers/>

Raw Data



Address	name	rating
[500 Technology Sq, Kendall Square/MIT, Cambridge, MA 02139]	Area Four	4
[280 Green St, Central Square, Cambridge, MA 02139]	Green Street	4
[853 Main St, Cambridge, MA 02139]	Craigie On Main	4.5
[190 Massachusetts Ave, Kendall Square/MIT, Cambridge, MA 02139]	Flour Bakery + Caf��	4
[73 Ames St, Kendall Square/MIT, Cambridge, MA 02142]	Study	4.5
[795 Main St, Kendall Square/MIT, Cambridge, MA 02139]	Cuchi Cuchi	4
[513 Putnam Ave, Cambridge, MA 02139]	Ampersand	5
[403 Massachusetts Ave, Central Square, Cambridge, MA 02139]	Caf�� Luna	4
[24 Pearl St, Central Square, Cambridge, MA 02139]	Rangzen Tibetan Restaurant	4
[567 Massachusetts Ave, Central Square, Cambridge, MA 02139]	Central Kitchen	4
[134 Hampshire St, Inman Square, Cambridge, MA 02139]	Oleana Restaurant	4.5
[9 Brookline St, Central Square, Cambridge, MA 02139]	Brookline Lunch	4
[765 Mass Ave, Central Square, Cambridge, MA 02139]	Life Alive	4.5
[64 Sidney St, Ste 1, Cambridge, MA 02139]	Bytes At University Park	4.5
[233 River St, Cambridge, MA 02139]	The Coast Cafe	4

Processed Data for Geocoding



name	rating	address	city	state	zip_code
EVOO Restaurant	3.5	Kendall Square	Kendall Square/MIT	Cambridge	.
75 Chestnut	4	75 Chestnut St	Boston	MA	2108
Jm Curley	4	21 Temple Pl	Boston	MA	2111
The Paramount	4	44 Charles St	Boston	MA	2114
Luke's Lobster	4.5	75 Exeter Street	Boston	MA	2116
The Gallows	4	1395 Washington St	Boston	MA	2118
Tavern at the End of the W	4	108 Cambridge St	Charlestown	MA	2129
Giulia	4.5	1682 Massachusetts Ave	Cambridge	MA	2138
Area Four	4	500 Technology Sq	Cambridge	MA	2139
Caf� Mami	4	1815 Massachusetts Ave	Cambridge	MA	2140
East Side Bar & Grille	4.5	561 Cambridge St	Cambridge	MA	2141
Study	4.5	73 Ames St	Cambridge	MA	2142
Machu Chicken	4.5	25 Union Sq	Somerville	MA	2143
Tenoch Mexican	5	382 Highland Ave	Somerville	MA	2144
Sarma	4.5	249 Pearl St	Somerville	MA	2145

PROC GEOCODE Code



PROC GEOCODE

```
METHOD = street  
DATA = work.yelp_DS_2  
OUT = work.yelp_DS_3  
LOOKUPSTREET = trct.USM  
ADDRESSVAR = address  
ADDRESSZIPVAR = zip_code  
ADDRESSCITYVAR = city  
ADDRESSSTATEVAR = state  
ATTRIBUTE_VAR = (TRACT, BLOCK); run;
```

```
quit;
```

This must be numeric

Optional output variables
that proc geocode will
produce.

Log for PROC GEOCODE



Geocoding Summary		Geocoding Progress				
		Completed		Time		
		Percent	Obs	Current	Remaining	Completion
Address data:	WORK.YELP_DS_2	0%	0	12:30:43	--	--
Output data:	WORK.YELP_DS_3	1%	9,736	12:42:55	20:07:33	08:50:28
STREET lookup data:	TRCT.USM	2%	19,472	12:50:36	16:14:25	05:05:01
CITY lookup data:	MAPSGFK.USCITY_ALL	3%	29,208	12:58:30	14:58:33	03:57:04
ZIP lookup data:	SASHELP.ZIPCODE	4%	38,944	13:10:18	15:50:10	05:00:29
Geocoding method:	Street level	5%	48,680	13:19:55	15:34:52	04:54:47
Run date:	15Feb2016	6%	58,416	13:28:57	15:12:18	04:41:16
Obs processed:	100	7%	68,152	13:42:23	15:52:15	05:34:39
Elapsed time:	00:00:02	8%	77,888	13:52:46	15:43:37	05:36:23
Obs per minute:	2,589	9%	87,624	14:08:06	16:24:43	06:32:49
Street matches:	94	10%	97,360	14:26:13	17:19:27	07:45:40
ZIP matches:	5	20%	194,720	16:40:37	16:39:36	09:20:13
City matches:	0	30%	292,080	18:45:10	14:33:43	09:18:54
Not matched:	1	40%	389,440	19:44:06	10:50:04	06:34:10
		50%	486,800	20:49:44	08:19:01	05:08:46
		60%	584,160	21:55:57	06:16:49	04:12:46
		70%	681,520	22:47:50	04:24:28	03:12:19
		80%	778,880	23:38:18	02:46:53	02:25:12
NOTE: PROCEDURE GEOCODE used (Total process time):		The SAS system				
real time	2.54 seconds	90%	876,240	00:35:00	01:20:28	01:55:29
cpu time	0.71 seconds	100%	973,599	01:27:32	00:00:00	--

Results and Output for Geocoded Data



address	M_ADDR	city	M_CITY	zip_code	M_ZIP	_MATCHED_	_STATUS_
Kendall Square		Kendall Square/MIT		.	.	None	
75 Chestnut St	75 Chestnut St	Boston	Boston	2108	2108	Street	Found
2 Winter Pl	2 Winter Pl	Boston	Boston	2108	2108	Street	Found
75 Chestnut St	75 Chestnut St	Boston	Boston	2108	2108	Street	Found
One Beacon St		Boston		2108	2108	ZIP	ZIP match
21 Temple Pl	21 Temple Pl	Boston	Boston	2111	2111	Street	Found
42 Beach St	42 Beach St	Boston	Boston	2111	2111	Street	Found
44 Charles St	44 Charles St	Boston	Boston	2114	2114	Street	Found
42 Charles St	42 Charles St	Boston	Boston	2114	2114	Street	Found
44 Charles St	44 Charles St	Boston	Boston	2114	2114	Street	Found
42 Charles St	42 Charles St	Boston	Boston	2114	2114	Street	Found

Debugging of Geocode



ERROR: The length of variable id in dataset WORK.HOMELESS36 is 1056.

This is greater than the maximum allowed length, 1000.

Data Set Name	WORK.HOMELESS36	Observations	1784
Member Type	DATA	Variables	7
Engine	V9	Indexes	0
Created	02/15/2016 12:36:11	Observation Length	1048
Last Modified	02/15/2016 12:36:11	Deleted Observations	0
Protection		Compressed	NO
Data Set Type		Sorted	NO
Label			
Data Representation	WINDOWS_64		
Encoding	wlatin1 Western (Windows)		

This cell addresses the error. Not the actual length of any variable.

```
New_variable=TRANSLATE(old_variable,' ','09'x);  
New_variable=translate(New_variable,' ','0A'x);
```

Using the new Geocoded Data with PROC GEOMAP (1/2)



```
PROC MAPIMPORT
```

```
DATAFILE="&shp.\CENSUS2010TRACTS_POLY.shp"
```

```
OUT=mass_tracts;
```

```
RUN;
```

```
PROC SQL;
```

```
CREATE TABLE mass_tracts_2 AS
```

```
SELECT a.*
```

```
FROM mass_tracts a
```

```
INNER JOIN yelp_ds_4 b
```

```
ON a.tractcel10
```

Using the new Geocoded Data with PROC GEOMAP (2/2)



```
PROC GMAP MAP=mass_tracts_2
DATA=yelp_ds_6 ;
TITLE "Yelp Map of Ratings in
Cambridge/Boston/Somerville";
    ID tractce10;
    CHORO rating_group / COUTLINE=black
NAME="Yelp Map of Ratings in
Cambridge/Boston/Somerville by Census
Tract" ;
RUN;
QUIT;
```

Final Results



Yelp Map of Ratings in Cambridge/Boston/Somerville



Rating Group

One or Fewer Restaurants with 4 or more stars
Three Restaurants with 4 or more stars

Two Restaurants with 4 or more stars
Four or More Restaurants with 4 or more stars

Alternative Analysis



PROC SQL;

CREATE TABLE MassTractsDemogRatings **AS**

SELECT a.*, b.*

FROM MassTractsDemog a

INNER JOIN yelp_ds_5 b

ON a.tractce10 = b.tractce10;

QUIT;

TRACTCE10	Population	Rating Group
353102	487	Two Restaurants with 4 or more stars
352800	237	One or Fewer Restaurants with 4 or more stars
350200	211	One or Fewer Restaurants with 4 or more stars
351000	170	Two Restaurants with 4 or more stars
351300	163	Four or More Restaurants with 4 or more stars

References



Papers:

1. Massengill, Darrell and Ed Odom. PROC GEOCODE: Now with Street-Level Geocoding." SAS Global Forum 2010 paper and example source code download. SAS Institute Inc. <http://support.sas.com/rnd/papers/>
2. Hadden, Louise, Mike Zdeb and Robert Allison. Wow! You Did That Map with SAS/GRAPH®? SAS Global Forum 2009.

Useful SAS datasets and Resources:

1. <http://support.sas.com/documentation/cdl/en/graphref/67881/HTML/default/viewer.htm#p0volumqexbvwn1kcpke78nqc7.htm>
2. <http://support.sas.com/rnd/datavisualization/mapsonline/index.html>

Kieran Shah, Abt Associates, Kieran.Shah@abtassociates.com, 617-520-2678, 55 Wheeler St, Cambridge, MA, 02139.