

INFOUSA Employment Data

13 September 2011

Sean Sullivan
GIS Analyst
MWCOG



INFOUSA Presentation Outline

This presentation will include the following:

- The steps taken to geocode the INFOUSA Employment file.
- The results of the geocoding process (Va, Md and DC)
- Summarize INFOUSA vs. Round 8.0 results by jurisdiction
- FYI Tables have been prepared showing INFOUSA vs Round 8.0 results for each TAZ

The Geocoding Process

- Used NAVTEQ Streets dataset
- Used a COG Built Address Locator:
 - Based on US Streets with Zone
 - Spelling Sensitivity – 80
 - Minimum candidate score – 10
 - Minimum match score – 60
 - Side offset – 30 feet

The Geocoding Process Continued

- Removed the following records prior to geocoding:

- Records with Blank Addresses
- Records with PO Box Addresses
- Records with RR Addresses
- Records with "Route" or US Addresses
- Records with Street Name only Addresses (Md)

The above records were separated into tables that were sent to each jurisdiction.

- After the geocoder was run, COG Staff manually geocoded unmatched records for employers with 25 or more employees.

Virginia INFOUSA DATA SUMMARY

Original Dataset Name	InfoUSA Data TPB MODEL VIRGINIA & WEST VIRGINIA April 2010.csv
Original # of Records	100478
# Matched	90502
% Matched	90%
# Tied	634
% Tied	1%
# Unmatched	9342
% Unmatched	9%
Altered Dataset Name	Va_WVA_InfoUSA_Altered
# of Records in Altered Dataset	93894
# Matched	90468
% Matched	96%
# Matched w/score of 100	88172
% of matched with 100 Score	97%
# Matched w/score of > 80	89562
% of matched with > 80 Score	99%
# Matched w/score of > 60	90468
% of matched with > 60 Score	100%
# Tied	460
% Tied	0%
Unmatched	2966
% Unmatched	3%

Maryland INFOUSA DATA SUMMARY

Original Dataset Name	InfoUSA Data TPB MODEL
Original # of Records	MARYLAND April 2010
# Matched	92443
% Matched	81156
# Tied	88%
% Tied	281
# Unmatched	0%
% Unmatched	11006
	12%
Altered Dataset Name	InfoUSA_TPБ_MODEL_MARYLAND_
# of Records in Altered Dataset	April_2010_Altered
# Matched	85123
% Matched	81245
# Matched w/score of 100	95%
% of matched with 100 Score	79928
# Matched w/score of > 80	98%
% of matched with > 80 Score	80757
# Matched w/score of > 60	99%
% of matched with > 60 Score	81245
# Tied	100%
% Tied	133
Unmatched	0%
% Unmatched	3745
	4%

DC INFOUSA DATA SUMMARY

Original Dataset Name	InfoUSA Data TPB MODEL District of Columbia April 2010.csv	Percentage
Original # of Records	42309	
# Matched	38052	90%
# Tied	5	0%
# Unmatched	2240	5%
Altered Dataset Name	District of Columbia	
# of Records in Altered Dataset	40292	
# Matched	38052	94%
# Matched w/score of 100	37513	93%
# Matched w/score of > 80	37817	94%
# Matched w/score of > 60	38051	94%
# Tied	5	0%
Unmatched	2240	6%

Jurisdiction	No_of_TAZ	Employers	INFUSA_Emp	ROUND80Emp	PCTINFOUSA	PCTROUND80
Arlington County	133	7856	124117	202098	61.41%	162.83%
Charles County	111	4086	47897	62105	77.12%	129.66%
Alexandria City	64	7111	88864	108845	81.64%	122.48%
Fairfax City	15	2876	27329	28073	97.35%	102.72%
Falls Church	9	1250	12249	11401	107.44%	93.08%
Manassas City	13	2155	28083	26646	105.39%	94.88%
Manassas Park City	4	409	4302	5052	85.15%	117.43%
Fairfax County	526	36579	519381	645695	80.44%	124.32%
Loudoun County	267	9863	127053	143402	88.60%	112.87%
Prince William County	351	10332	122997	127799	96.24%	103.90%
Va_Totals	1493	82517	1102272	1361116	80.98%	123.48%
Gaithersburg City	33	2910	42861	60845	70.44%	141.96%
Rockville City	36	4629	53380	79906	66.80%	149.69%
Frederick County	121	7606	97502	138504	70.40%	142.05%
Montgomery County	377	39588	479041	507520	94.39%	105.95%
Prince George's County	616	26280	352262	359334	98.03%	102.01%
Md_Totals	1183	81013	1025046	1146109	89.44%	111.81%
District of Columbia	387	38052	591222	786223	75.20%	132.98%

Conclusion

- All data has been sent to the respective jurisdictions
- COG strongly encourages the jurisdictions to share the results of their geocoding with our GIS staff.
- Questions/comments