

DRILLDOWN METHODOLOGY

Background

Beginning with the premise that a significant reason for inner-city disinvestment is lack of good market information, Social Compact developed the Neighborhood Market DrillDown to address key barriers to private investment in and around inner-city neighborhoods. Social Compact offers its neighborhood market analyses to local governments, community organizations and businesses looking to attract investment or to invest in inner cities. These analyses provide alternative assessments of population, income and housing that do not rely on outdated and potentially inaccurate decennial census data.

The DrillDown was established to provide up-to-date profiles of market strength, stability and opportunity for small, dense, and rapidly changing urban geographies. These analyses provide an alternative assessment of population, income and housing that do not rely on dated, and perhaps inaccurate, decennial census data. The DrillDown builds on current, finely sieved market information drawn from a wide spectrum of commercial, proprietary and local government sources (e.g. tax assessor, building permit, and commercial credit companies). Rather than relying on any one information set, DrillDown findings surface from a combined body of data. These findings, tested against supplemental data and the intuitive knowledge of local market leaders, serve as the foundation for an objective, systematic analysis of business attributes.

Based on this data, Social Compact has developed indicators of market size, strength and stability at the blockgroup level.

Basic Steps

- ~ Enumerate each residential address in the study area.
- ~ Estimate vacancy rate.
- ~ Estimate average household size.
- ~ Match owner address to property address to estimate homeownership rate.
- ~ Construct income distribution for median household income and informal economy.
- ~ Use Consumer Expenditure survey (CE) to convert distribution into aggregate income and expenditures.
- ~ Construct table with the percentage of households without a corresponding credit record (underbanked).
- ~ Calculate median home value using home sales, aggregate to census tract or neighborhood.
- ~ Create business list with business type, revenue or square footage, number of employees.
- ~ Count number of predatory lenders for informal economy estimate.
- ~ Calculate the informal economy and apply to aggregate income.
- ~ Gather crime statistics (trend and per capita) and code as violent, property or community.

Definitions:

Housing Unit: a house, an apartment, a mobile home, a group of rooms, or a single room that is occupied (or if vacant, is intended for occupancy) as separate living quarters.

Household: equivalent to an occupied housing unit; includes all the people who occupy a housing unit as their usual place of residence.

Population in households: the number of people living within an occupied housing unit.

Group quarters population: people living in college residence halls, residential treatment centers, skilled nursing facilities, group homes, military barracks, correctional facilities, workers' dormitories, and facilities for people the homeless.

Total population: population in households plus the group quarters population.

Housing Units:

HOUSING UNITS = TOTAL UNITS CONTAINED IN RESIDENTIAL BUILDINGS LIST

The complete and current residential address list is compiled from multiple data sources including: municipal tax assessment records, municipal building permit records, municipal home sales records, municipal residential utility data (ideally 12 months of data), proprietary real-estate data, credit bureau records, and other sources that may be available in particular cities. Some of these records will be building level and some will be unit level (household level records such as credit bureau information will be unit level, tax assessor will be a mix in cases where there are condos).

Individually these data sources are not adequately reliable or complete, but when blended together the results are generally robust.

Attributes of typical DATASETS:

Tax Assessor is often the most complete listing of residential addresses and number of units.

Dataset should include: PROPERTY ADDRESS, OWNER ADDRESS, NUMBER OF UNITS, a LAND USE CODE and a reliable distinction between residential and non-residential property.

Only those properties with a residential LAND USE CODE or some other residential distinction should be included.

The LAND USE CODE provides an additional source of unit information.

The number of units in multi-unit buildings is not always correct or defined. It may be problematic if a large number of properties have undefined units.

Due to the nature of the dataset the property address is not always a valid mailing address.

Home Sales data often contains similar information to the tax assessor, but may fill gaps, identify problems or simply provide further evidence.

Dataset should include: PROPERTY ADDRESS, OWNER ADDRESS, NUMBER OF UNITS, a LAND USE CODE and HOME SALE VALUE.

Only those properties with a residential LAND USE CODE should be included.

The LAND USE CODE provides an additional source of unit information.

Sales data often goes back several decades; older data is more suspect than newer data, as the housing stock may have changed.

Proprietary home sales data can either be very good or very bad. In other words, the quality is always suspect.

Building Permits data will provide information on new housing stock, demolitions, conversions resulting in a change in the number of units, and residential addresses otherwise missed by the above sources.

Dataset should include: PROPERTY ADDRESS, PERMIT TYPE (demolition, new construction or rehab), NUMBER OF UNITS, LAND USE CODE or other residential distinction. Sometimes these fields will have to be mined from actual descriptions.

Only those properties with a residential LAND USE CODE should be included.

Multiple permits for a single address will exist; the data should be reduced to the last permit for an address.

Utility hook-up records are sometimes more complete than even the tax assessor records and provide information about vacancy.

Dataset should include: PROPERTY ADDRESS, MONTHLY CONSUMPTION, NUMBER OF UNITS, some residential distinction.

Sometimes addresses are billing rather than property mailing addresses, which can create duplicate records in the final buildings table.

Household-level data such as credit bureau records, DMV records and voter registration records provide information on frequently missed households. For example, these records will often identify housing units when single-family homes include a separate apartment.

Dataset should include: PROPERTY ADDRESS

These addresses sometimes include commercial addresses in cases where individuals use a business as their personal mailing address.

Street-confirmed records are the most trusted and least numerous of the address records. Ideally buildings that are questionable after aggregation (discussed later) and new development projects will be street-checked and included.

Aggregating the records into a standard ADDRESS table:

Standardize the addresses within the various datasets. The addresses need to be standardized so that each housing unit or building is defined by a unique address (i.e. avoiding duplicates such as 192 Martin Luther King and 192 MLK, 1716 7th St Apt 1 and 1716 7th St # 1). Addresses should be marked as a building or apartment (ideally the apartment number/letter is a separate column). Each address needs latitude and longitude that uniquely identifies that building (not the zip+4 centroid).

Identify the number of units. Each record will have an associated UNITS field. Municipal data with land use codes should include a UNITS column and a LANDUSE UNITS column. The land use units will be the minimum number of units defined by the land use code. For example, a 5-10 unit apartment building will have LANDUSE UNITS=5.

Associate the address to a building permit. An address should be marked as DEMOLITION or NEW CONSTRUCTION where applicable, as determined by the last/most recent building permit for an address.

Determine the owner occupancy of an address. Owner occupancy is defined by matching the owner address to the property address where an owner address is given. Sometimes in large multi-unit buildings the owner address will be the building address (no apartment number) – *these need to be excluded when aggregating to the building level.*

Determine if the address has a verification of occupancy. Some data sources give clues as to whether an address is occupied. These datasets include: credit bureau record, water usage, voter registration record, etc.

Aggregating address records into the BUILDINGS table:

A building is defined as a rooftop (latitude/longitude). The buildings table will ultimately need the following fields: LAT, LON, ADDRESS, STFID, UNITS, DEMOLITION, NEW CONSTRUCTION, VERIFIED OCCUPIED UNITS, NUMBER OF OWNERS, and NUMBER OF CREDIT RECORDS, SOURCE INFORMATION (tax assessor record, credit records, etc.), and perhaps WATER CONSUMPTION.

The UNITS field for a building must be defined using assumptions about the trustworthiness of different data sources. These assumptions vary project to project, but there are some good rules of thumb:

- ~ The number of individually identified apartments is usually a good minimum number of units
- ~ Land use units is the minimum number of units
- ~ First American is the least trustworthy data source, only used when other information is absent
- ~ In some cases the utility company has better information than the city, but not always
- ~ If the number of units is much higher than the number of credit records that should be checked

This estimation process could become quite sophisticated comparing the consistency of information from different sources.

VERIFIED OCCUPIED UNITS are the number of addresses in that building with a confirmation of occupancy (as discussed above).

Remove non-residential and suspect buildings:

In some cases the data will contain non-residential addresses, abandoned buildings or invalid addresses – these errors should be minimized by cross-checking data sources. In order to assure the best possible data quality it is important to be conservative about the validity of each dataset. Here are few checks that will reduce the likelihood of error:

- ~ All final records should trace back to at least one source that validates the address as valid and residential.
- ~ Typically this means that the building has a tax assessor, building permit, public housing or utility data source. This distinction allows us to include records that may ultimately be excluded such as credit and voter records, which may be reflected as an office address. Similarly, there may be records with questionable land use codes or older home sales records that should be validated with another data source.
- ~ Some records are sufficiently confirmed by a single VERIFIED OCCUPIED UNIT. These records may include those with questionable land use codes or old home sales records.
- ~ In some cases a number of VERIFIED OCCUPIED UNITS implies that a building is residential despite its absence in other data sources. In cases where buildings have a large number of credit records (say 10+), it is reasonable to assume that the building is actually residential.
- ~ Some records require confirmation that they are *active* residential properties, reducing the likelihood of including buildings that are unoccupied, demolished or under-construction:

Generally, buildings marked NEW CONSTRUCTION, DEMOLITION or large multi-unit buildings with 0 water consumption require evidence that they are currently occupied (VERIFIED OCCUPIED UNIT). DEMOLITIONS are not automatically thrown-away because the demolition may not have occurred yet (or at all), sometimes permits are not correctly coded, a new construction not recorded, etc. Similarly zero water consumption could simply be an error in the data.

General cautions about the residential address list:

- ~ Addresses that span (1020-1022 First St NW) must be discarded or matched to a single mailing address, as they do not geocode well.
- ~ Address of a main apartment building office may be different than mailing address of residents – this will create duplicates in the data.
- ~ Examine inconsistencies between the datasets, improve the base assumptions and try again.

Housing Units is sum of the units in this final list.

Households & Vacancy/Occupancy Rate

HOUSEHOLDS = HOUSING UNITS * OCCUPANCY RATE

Vacancy Rate is calculated at the blockgroup level from the sample of records with vacancy information. Utility consumption is the best indicator of occupancy. Because utility data cannot generally handle vacancy rates for multi-unit buildings, the occupancy rate will be based on single unit buildings; therefore, the housing stock in the study area should be considered first. Utility records should contain unique addresses with information for each month (so it would be any address with use within the last year). Vacancy can then be calculated each month (or each quarter if data is not taken each month) as the percentage of units without sufficient consumption (average use is 6,000 gallons per month).

An alternative to utility data is using other data sources that imply a unit is occupied. Unit occupation can be estimated using credit bureau records, water consumption data, tax assessment information on homeownership, and Census vacancy rates. Addresses with credit bureau records are considered occupied, as are addresses with recent water usage and owner addresses matching the property address. Units with no water usage are assumed to be vacant. As credit bureau records and the water consumption data are not complete, Census vacancy rates are also applied when necessary.

Population & Average Household Size

POPULATION = (HOUSEHOLDS * AVERAGE HOUSEHOLD SIZE) + GROUP QUARTERS POPULATION

Average Household Size is calculated at the blockgroup level. Ideally this estimate is not Census derived. However, there are few ways to achieve a better estimate, so the Census generally provides the baseline. Other methods could potentially include household surveys, adjusting for the number of adults in a household using DMV records or using average water consumption per person, etc.

Group Quarters Population can be derived from combining the military population and university students in dormitories with either self-constructed estimates of the institutionalized population or estimates provided by the census. In cases where the group quarters underestimation and/or change is thought to be insignificant, the Census count or other estimate may confidently be used.

Underbanked

Social Compact approximates the percentage of underbanked households as the number of households without a corresponding credit record, or the difference between the number of households and the number of credit records in the buildings list.

Homeownership

Homeownership is calculated at the blockgroup level from the sample of records with owner information. A unit is considered owner occupied if the property tax bill is received at the dwelling address. In an urban area, it is important to examine owner occupancy by building in addition to owner occupancy by unit. Unlike suburban neighborhoods, a high percentage of the housing stock in an urban area is multi-unit or multi-family. For example, a resident may own a three-unit building, live in one unit and rent out the other two units, showing 100% owner occupancy at the building level, but only 33% at the unit level. As such, Social Compact assesses owner occupancy by building and in so doing reveals more owners in the neighborhood who have a stake in the stability of the market. A building is considered owner-occupied if there is at least one owner in a 5 unit or smaller building, or if at least 50% of the units are owner-occupied.

HOMEOWNERSHIP BY UNIT is the percent of units occupied by the owner (the sum of the units in buildings with owner information compared to the sum of owners).

HOMEOWNERSHIP BY BUILDING is the percent of buildings where, for one to five unit buildings, at least one unit is occupied by the owner and, for buildings of more than five units, at least half of units are occupied by the owner.

Median Home Value

Social Compact analyzes the home sales values from property transaction data extracted from municipal property sales information and/or proprietary real estate data for a period of 12 to 18 months. Generally, there are not enough sales to calculate a median value at the blockgroup level, so the estimate is derived for a census tract or neighborhood.

Base Income Estimates

The Income distribution is estimated for each blockgroup from credit bureau individual-level income estimates. The credit bureaus provide an estimated income range for each individual in their records. We assume that the credit bureau records provide a representative sample of the actual income distribution. Once aggregated this sample of records provides the percentage of households within each income bracket at the blockgroup level. The final income distribution table should contain the number of people in each Consumer Expenditure Survey (CE) income bracket at the blockgroup level. In cases where the credit bureau income ranges differ from the CE ranges, a conversion must be made assuming households are uniformly distributed within each range.

Median income is then trivially defined as:

$Lower + Width * (TotHhs/2 - BelowHhs) / IntervalHhs$, where

Lower = lower limit of the interval containing the median

Width = width of the interval containing the median

TotHhs = total number of households

BelowHhs = cumulative frequency corresponding to the lower limit

IntervalHhs = number of households in the interval containing the median

Aggregate Income is determined by aggregating the average incomes of households within each income bracket. The Bureau of Labor and Statistics Consumer Expenditure Survey can be used to determine the average and aggregate income from the income distribution. The distribution of individuals within each income bracket is a national estimate determined by the Bureau using surveys. Using this distribution is preferable to assuming the distribution is normal and taking the midpoint of the range as the average.

SocialCompact

738 7th Street, SE | Washington DC 20003 | 202-547-2581 (main) | 202-547-2560 (fax) | www.socialcompact.org

Aggregate income is then calculated as the sum of the average incomes multiplied by the number of households for each income bracket. The combined, or “study area,” average income is then the aggregate income divided by the total number of households for that area.

Aggregate Expenditures are similarly derived, but instead of using the CE average income of each income range, the average expenditure is used.

Informal Economy

Social Compact’s model uses proxies (measurable at the neighborhood level) of informal activity to estimate the income generated by that activity that may have been missed in census and census upgrade profiling. The model’s proxies have expanded over the eight years it has been employed but the methodology has essentially remained the same. This indicator is the only DrillDown measure that cannot be calculated at a geography smaller than a neighborhood. Currently, Social Compact measures eight proxies, using a combination of publicly and privately available data, to estimate local informal economy activity and associated income, these include:

- Percentage of households with a total income of less than \$30,000
- Ratio of household expenditures over income
- Percentage of households with no banking relationships or credit histories
- Percentage of utility payments made in cash
- The prevalence of check-casher operations per acre in the profiled neighborhood
- The prevalence of check-casher operations per household in the profiled neighborhood
- Modeled versus actual housing costs
- Percentage of the neighborhood’s population that is foreign born

Once collected, these eight proxies are then weighted using a points scoring system based on the results of Social Compact’s original work in four diverse Chicago neighborhoods. Social Compact identified the proxies through surveys of local residents and interviews with local community development practitioners in Chicago. The scale was weighted such that the Chicago neighborhood that exhibited the highest prevalence of a particular proxy for a particular indicator was considered the top value (represented 100% of possible points). The points scoring system amounts to a total of 10 possible points, with the proxies accorded the following values in the system:

- Percentage of households with a total income of less than \$30,000 (1 point): top value 45%
- Ratio of household expenditures over income (1 point): top value 120%
- Percentage of households with no banking relationships or credit histories (2 points): top value 75%
- Percentage of utility payments made in cash (1 point): top value 100%
- The prevalence of check-casher operations per acre in the profiled neighborhood (0.5 points): top value is a sliding scale between 1 check casher per 100 acres (100%) and 1 check casher per 250 acres (0%)
- The prevalence of check-casher operations per household in the profiled neighborhood (0.5 points): top value is a sliding scale between 1 check casher per 1000 households (100%) and 1 check casher per 2500 households (0%)
- Modeled versus actual housing costs (2 points): top value 100%
- Percentage of the neighborhood’s population that is foreign born (2 points): top value 80%

Businesses and Expenditures

Aggregate sales are estimated by an analysis of business-level data. The list must include a geocoded address, square footage or revenue, and business type. The businesses are classified by North American Industry Classification System (NAICS) codes, such as grocery, apparel, retail, etc. To classify businesses into types it is often necessary to individually scrutinize businesses as they are often not well placed in a NAICS category.

Grocery stores should be divided into full- and limited-service, where full-service can be initially be distinguished by having more than 20 employees or as being a nationally recognized grocery provider (i.e. Safeway even if it shows less than 20 employees). The list should be verified by calling the stores or finding information online when the stores are not known supermarkets (Albertsons, Safeway, etc.). Where possible, it is preferable to have a trusted, local level partner review and confirm the final list of full service grocers.

Additional indicators include the number of employees, business density, small business density, new businesses, and average distances to grocery stores, banks and predatory lenders (pawnshops, payday lenders, and check cashers). These indicators should be contextualized by the city and/or metro level indicators.

The informal economy estimate requires a list of predatory lenders. These lenders can generally be distinguished by NAICS codes. In cases when check-cashing occurs in liquor stores, the list will require another data source (lending permits or something similar) in order to differentiate establishments.

Aggregate Retail Float/Leakage

Social Compact determines the retail float/leakage through analysis of either aggregate square footage or aggregate revenue in comparison to aggregate expenditures. Expenditures can be converted into square footage demand using International Council of Shopping Centers (ICSC) average sales per square foot estimates. Unfortunately many NAICS categories cannot easily be converted into CE expenditure categories. Grocery, apparel, total retail, restaurant and home improvement can all be analyzed for leakage without too many assumptions. Additionally, leakage can nicely be mapped as leakage within a one-mile or two-mile ring for each individual blockgroup.

Crime

Social Compact examines crime data at the block group level, as part of its neighborhood risk and stability analysis. Crime statistics, obtained from local police records, are gathered longitudinally to assess trends in documented criminal activity over time. In order to maintain consistency across comparisons of time and geography, Social Compact uses population and household counts from Census 2000 and not current DrillDown estimates; and employs the Federal Bureau of Investigation's (FBI) Uniform Criminal Reporting (UCR) coding to define property and violent crime (see below). Unless otherwise stated, trends are reported for the start of the earliest year through the end of the latest year for which data was obtained (i.e. "2002 – 2006" is equivalent to January 1 2002 – December 31 2006). VIOLENT CRIME includes reported incidents of murder, forcible rape, robbery, and aggravated assault (in accordance with FBI UCR coding). PROPERTY CRIME includes reported incidents of arson, burglary, larceny/theft, and motor vehicle theft (in accordance with FBI UCR coding).

In addition, Social Compact designed an indicator of community crime to assess visible incidents of crime that, although not typically included in many official crime statistics, may be of particular interest to the business community. Community crimes include reported incidents of blighted properties, disorderly conduct, littering, loitering, prostitution, trespassing and vandalism.

Other Market Indicators

HMDA New Homeowner Income	(2003-2007; tract)
HMDA Number of Home Purchase Loans	(2003-2007; tract)
USPS Commercial + Residential Counts	(2006-2009; tract)
USPS Commercial + Residential Count Change	(2006-2009; tract)
IRS Change in Adjusted Gross Income	(1998-2005; zipcode)
Change in Student Enrollment	
Change in the Number of Utility Hook-Ups	